# RSISE PhD Notes

**Gordon McIntyre**

**August 2007**

**The Department of Information Engineering**
**The Research School of Information Sciences and Engineering**
**The Australian National University**

# 1 Log

see Articles section for summary of articles

**13 November 2007** Spent a week installing QT, Eclipse, analysing Jason's code. Re-compiled FETA to check that all was working. Have now launched into designing iteratively building a prototype for the Anxiety System (NXS).

**7 November 2007** Cought up with Roland.

**24 October 2007** Met with Jason to discuss getting latest copy of his code. He will provide something this afternoon.

**19 September - 23 October 2007** Building models of anxious faces, making last minute changes references within Springer paper, submitted abstract for Summerfest and accepted, had meeting with Bruce and just need to tidy up study plan.

**10 October 2007** Email from Jinni Harrigan

> What an enterprising idea you have for coding! I think you've got the right Aus for fear/anxiety. Check out some other sources for Paul's work and also our book, The New Handbook of Methods in Nonverbal Behavioral Research. There's a lot of info on facial coding and other types of coding. Good luck to you in your work and please keep me informed of your progress.

–

**23-30 September 2007** Various

Sent Summerfest abstract

Sent final version of Springer paper.

**19 September 2007** various

Downloaded eye blink software (source code) from MPT

Corresponded with Professor Jinni HArrigan from California State University regarding detection of anxiety. Comments as follows:

> finally able to have a minute to respond to your query. there are no studies that i am aware of that examine AUs and anxiety. i used fear because as Paul Ekman and i have frequently debated, i believe in the modern world anxiety is more often experienced than fear. anxiety should show some of the same movements as fear (mouth, eyes). i don't know if i would recommend learning the FACS. it's a lot of time to learn it and without a partner to help with reliability it's not very useful, unless you plan to do a lot of facial evaluations. i'd like to know more about what you mean about machine recognition of anxiety. is it a program you are designing? hope my comments are helpful.

**18 September 2007** Met with Jason to do a reality check on the use of AAMs

Following that, resurrected Arne's application properly and it seems to run pretty well. Next step is to use the fear samples from FGNET but it looks like I have to annotate the points.

**17 September 2007** Various

Wrote abstract for Summerfest 2007 and will submit it 24 September.
Met with Jason to discuss the viability of using AAMs to recognise AUs.

**07 September 2007** Meeting

Met with Richard O'Kearney of Psychology School. He had some good suggestions and steered me towards the social phobia, arachnophobia, vertigo, etc. and then maybe look at more general anxiety. He also mentioned work by a colleague to look at mirror neuron systems and recognition.

**06 September 2007** Various

Attended "Science Communication" course from 4-6 September
Rescheduled meeting with Richard O'Kearney of Psychology School for Friday 7 September. He had to cancel previous meeting due to illness.

Talked with Jason who agreed to package his AAM modules by 14 Sep 2007

Download Cohn-Kanade database

**30 August 2007** Various

Arranged meeting with Richard O'Kearney.

**29 August 2007** Various

Caught up with Roland and discussed:

- **StudyPlan** - I will incorporate suggestions and forward to Bruce and Antonio.
- **Image processing technique to model motion as in FACS AUs.** - Discussed Cohn's paper where they use multiple approaches such as optical flow and Gabor wavelets, and compared to Matthew's paper using AAM to model FACS AUs. **Decided to evaluate the use of AAM to model FACS but incorporate time domain.** With regard to boundary detection (onset peak offset of AUs), I will check the approach of researcher at RMIT. I will also investigate use of Histograms.
- **Choice of classifier** - FETA uses ANNs and although FANN is a nice piece of software, a multimodal system would likely be better served using HMMs and DBN.
- **OZCHI** - Decided not to submit and to concentrate on system development.
- **Jason** - Roland has asked Jason to package up his AAM
- **Eye tracking** - Roland suggested that there may be a solution that I could use, within a few weeks.

- **Mapping of FACS AUs for Anxiety** - I will discuss this with Richard O'Kearney. code.

**27 August 2007** Admin literature review

Finished and sent draft schedule

Read several articles

**20-21 August 2007** Admin

Revamped log

Wrote draft schedule

Wrote draft Study Plan

Read several articles

**13-17 August 2007** Admin

Sent request for collaboration to ANU School of Psychiatry and Centre for Mental Health Research

Wrote ASSTA article

**01 - 8 August 2007** Various

Wrote first couple of pages of proposal and met Roland at Vanilla, John Curtin

**30 July 2007** Got back to work with clear idea of project.

**20-27 July 2007** Beijing HCI International conference 9 July 2007 Resurrected Arne Arnolds' code Started reading Statisical Models of Appearance for Computer Vision, TF Cootes, et. Al. 2001

**6 July 2007** Various

Following correspondence with:

Dr.-Ing. Frank Wallhoff Inst. for Human-Machine Communication Technische Universitt Mnchen Arcsisstr. 16, Room S2638 80333 Munich, Germany Phone: +49-89-289-28552 Fax: +49-89-289-28535 Email: wallhoff@ei.tum.de http://www.mmk.ei.tum.de/ waf

I download FGNET database - about 1GB of mpeg and jpeg files.

Starting designing system. 5. Picked up passport with visa from Chinese embassy

## 2 Articles

**26 September 2007 current readings**

"Bayesian Networks as ensemble of Classifiers*", "Automated Facail Image Analysis for Measurement of Emotion Expression" Jeffrey Cohn, "Emotion Elicitation Using Films", Jonathan Rottenberg, "Meticulously Detailed Eye Region Model and Its Application to Analsyis of Facial Images, IEEE, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions", Irfan A. Essa
and the book, "Eye Tracking Methodology: Theory and Practice", Duchowski.
(to be catalogued)

**26 September 2007 Eye Typing using Markov and Active Appearance Models** [4]
This article was co-authored by Mikkel Stegmann, owner of the site for the AAM-API. Some of the material describing AAMs has been taken directly from Stegmann's article, "Mutli-band modelling of appearance".

The motivation for the system is to enable disabled people, especiall those with Amyotrophic Lateral Sclerosis, to perform eye-typing.
They use a low cost, off-the-shelf, non-IR camera, positioned on the table left of the computer keyboard. They list three reasons for the placement:

1. the camera needs to be close to the subject because they will use the eye movements to perform eye-typing
2. they don't want to obscure the line of sight
3. images taken from below usually obtain a much better view of the eye because of the smaller movement of the lower eyelid (INTERESTING)

They contend that color information is useful for eye tracking and the shape and texture model is implemented in AAMs. They use mean-shift paradigm to first track greater shifts of the eyes.

An interesting statement, probably true given the co-author, is that an overlap of roughly 60 percent between the model and the object in the image is generally sufficient for accurate location when fitting.

They use an on-screen typing pad and first calibrate it with 12 on-screen points. Using AAM-API, only 28 images are used to train the non-person-specific AAMs.

**12 September 2007 HOW DO YOU LOOK WHEN FEELING ANXIOUS? FACIAL DISPLAYS OF ANXIETYcurrent readings** [5]
This is an important article for my project. The primary author was Professor Jinni HArrigan from California State University. They used FACS to measure induced anxiety in a number of undergraduates. They reasoned that although anxiety is a composite affect, the most heavily weighted affect would be fear. Muscle movements involved in fear include raising the eyebrows and drawing them together, and/or stretching the lips horizontally so that the lips form a rectangular motth shape. In more extreme fear, the lower eyelid is tensed, the upper eyelid raised, exposing the sclera. They also predicted an increase in eyeblinks,

more “non-enjoyment” smiles, and more head movement.
In summary, they hypothesised:

- more fear facail action units
- a greater eye blink rate
- more non-enjoyment than enjoyment smiles
- more total facial movement

The participants were 37 undergraduates - 18 females and 19 males. Ages ranges from 19-28 with one student aged 48. They were video-taped while describing feelings and reactions to the “most-anxious” events that they had experienced. Participants self-rated on three adjective scales uncomfortable, nervous, apprehensive, ranging from 1(“Not at all”0 to 9 (“Very). These are similar to the Spielberger’s State Anxiety Scale.

Facial movements were coded with Accredited FACS coders AUs using FACS and EMFACS. In accordance with the FACS guidelines, two FACS trained coders annotated the facial movements and c

**07 September 2007 Facial expressions: What the mirror neuron system can and cannot tell uscurrent readings** [12]

Mirroring of facial expressions has recently generated a lot of attention. There has been evidence that the MRS has been involved in the observation of facial expressions in humans. The inferior frontal gyrus (IFG) and posterior parietal cortex have been considered to compose a mirror neuron system (MNS) for the motor components of facial expressions, while the amygdala and insula may represent an additional MNS for emotional states. In three separate event-related fMRI experiment, subjects had to (1) observe (2) discriminate and (3) imitate facial expressions.

Results of previous studies have shown that premotor and parietal cortex were both involved in facial expression *observation* and execution, supporting the idea of the involvement of the MNS in understanding facial expressions.

The experiment consisted of 3-second movie clips depicting the emotions happiness, disgust, fear or neutral. The aim was to use fMRI to detect activity when the subjects were asked to Observe, Discriminate, and Imitate.

The study found evidence that bilateral inferior frontal gyrus and inferior parietal cortex to be involved in both the observation and execution of facial expressions. IN addition to these ”classical” sites, the study also found other regions to be involved in the observation and execution of facial expressions.

**30 August 2007 Automatic segmentation of echocardiographic sequences by active appearance motion models** [2]

Ultrasound images present a difficult case in image recognition. There is no simple relation between pixel intensity and physical property of the tissue visualised; the image is highly anisotropic (not invariant from all directions) and person dependent; lots of artifacts (noise,

shadowing, echoes, reverberations). The object was to view the performance of the LV wall. The authors report that AAM matching is a promising segmentation technique because it can deal with some of the problems mentioned earlier.

Their AAMM approach models the shape and appearance of the heart in combination with the dynamics of the cardiac cycle (16 phases) by considering the time sequence as a stack of 2-D images (time frames). They consider a whole image sequence as a single shape/intensity sample.

One thing to note though is that in the intensity normalisation stage, they apply a non-linear normalisation before calculating the average image. This is because echocardiograms have a highly non-Gaussian intensity histogram.

All single-beat sequences are phase normalised into a fixed number of frames so that the end-diastolic (dilation of heart chambers allowing blood to enter the heart) and end-systolic (contraction driving blood through the aorta and pulmonary artery) frames map to the same frame number. The shape vectors for all time frames are concatenated in the order of their phase number and further treated as a single 2-D shape vector.

The AAMM matching procedure resembles conventional 2-D AAM matching. However, the rms error criterion and the parameter regression matrices for the appearance coefficients, pose, and global intensity are calculated for the full image sequences in AAMM.

The authors report very good results (97 percent successful matches). However, part of the setup involved using a special non-linear intensity normalisation process and when compared against the conventional linear normalisation, it performs much better (linear 73.4 percent).

**Discussion**
Could this approach be used to model FACS AUs? One of the nice things about heartbeats is that (you hope) they are reasonably regular. I guess you could say the same about individual AUs, that is, if you can do the boundary detection to find the start-onset and end-offset points.

**28 August 2007 Facial Action Coding Using Multiple Visual Cues and a Hierarchy of Particle Filters** [7]
Discusses approaches of optical flow and AAMs. Describes a real time system that uses a hierarchy of particle filters tied together in a DBN. Particle filtering is a convenient method for estimating the probability density of the current state of an object given the object's previous states as well as current and past observations.

They also mention the use of thin-plate splines, which is a physical analogy involving the bending of a thin sheet of metal. In the physical setting, the deflection is in the z direction, orthogonal to the plane. The tracked landmarks are used for control points in the TPS warping.

**28 August 2007 An empirical comparison of supervised machine learning techniques in bioinformatics** [10]

An issue that arises quite often in my mind is how do you select a classifier, i.e. I know that we can split them up between supervised and non-supervised, and NNs are a bit deficient in the lack of traceabiltiy etc. However, what heuristics are there?

This article is a fairly simple read and although it is pitched to bioinformatics, it is still useful. They perform an empirical comparison of rule-based learning systems (Decision trees, One Rule, Decision rules), statistical learning systems (Naive Bayes, Instance Based, SVM and NNs) and ensemble methods (stacking, bagging and boosting).
The exercise was against several datasets. Evaluation measurements were

$$Accuracy(Acc) = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$PositivePredictiveAccuracy(PPV) = \frac{TP}{TP + FP} \tag{2}$$

$$Sensitivity(S_n) = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity(S_p) = \frac{TN}{TN + FP} \tag{4}$$

**Rules-of-thumb**

**How does one choose which algorithm is best for their data set?**
*Ratio of the training data* - If the training TPs and TNs are almost equal in size, the algorithms tend to construct much better classifiers. If the size of the TP set is small compared to that of TN, most probably the classifier will overfit the positive examples and thus perform poorly in the cross validation stages.
*Attributes* - SVM and NNs tend to perform much better over multi-dimensions and continuous attributes. By contrast, rule-base systems tend to perform better in discrete/categorical attributes.
*Credibility vs. Comprehensibility*
Ask yourself what you want to discover from the data.
**Are combined methods better than a single approach?**
From their experiments, most of the combined methods berform better thatn the individual learner. This is because none of the individual methods can claim that they are superior to the others due to statistical, computational and representational reasons.
**How does one compare the effectiveness of a particular algorithm to the others?**
Predictive accuracy - most of the time, accuracy is no thte ultimate measurement when comparing the learner's credibility. For example, if the training data set has 95 TNs and 5 TPs, by classifying all the instances into a negative class, the classifier can still achieve 95

**22 August 2007 Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavour" [1]**

This article presents a comparison of machine learning methods applied to the problem of fully automatic recognition of facial expression and spontaneous expression using AUs.

Although a short paper it assumes an understanding of PCA, LDA, SVM (RBF), Gabor filters, and AdaBoost.
Learning was performed against DFAT-504. The system operates real-time at 24 frames/second on a 3 GHz Pentium IV for 320x240 images. They claim that they achieved 100 percent face detection with their software (available see below) and the located faces were rescaled to 48x48 pixels.
They claim that SVMs were well suited to the task because of the high dimensionality of the Gabor reprentation (using a bank of filters at 8 orientations and 9 spatial frequencies (2:32 pixels per cycle at 1/2 octave steps).
The results are a bit brief and preliminary. There is some freely available source code for face detection that they have written that implements Gentleboostat at http://kolmogorov.sourceforge.net.
They report that the best results are from selecting a subset of Gabor filters using AdaBoost and then training SVMs on the output of the filters selected by AdaBoost.

**21 August 2007 Real Time Inference of Complex Mental Sates from Facial Expressions and Head Gestures [6]**
This paper is quite similar to the one below but goes more into the reasons for the choice of HMM and DBN classifiers. FACS UAs are used to derive facial expressions (not archetypes). Underlying mental states of "agreement, concentrating, disagreement, thinking and unsure and interested" are derived from the expressions. Facial expressions are modelled as HMMs. Each mental state is modelled as a separate DBN that uses the HMMs.

**21 August 2007 Mind Reading Maching: Automated Inference of Cognitive Mental States from Video [8]**
The relevance of this paper is that it uses FACS AUs in conjunction with DBNs. The head action units are extracted as follows. The head yaw is given by the rotation of left to right eye widths. Head roll is given by the orientation angle of the two inner eye corners. Head pitch is determined from the vertical displacement of the nose tip normalised against the distance between the two eye corners to account for the scale variations.
Facial AUs are identified from component-based facial features (Seems to be mouth only) from motion, shape and colour descriptors. So, an anchor point is first established, the mid-point between the two mouth corners when the mouth is at rest and is at a distance d from the line joining the two inner eye corners l. In subsequent frames the point is measured at a distance d from l, after accounting for head turns. On each frame the polar distance between each of the two mouth corners is derived. An change of more than 9 percent indicates a lip pull or pucker. Colour regions are tested to detect the aperture and teeth display.
The authors use HMMs to model spatio-temporal changes and to classify facial expressions. The head nod HMM is a 4 state, 3 symbol HMM (up, down, no action). Tilt and turn displays use a 2-state HMM with 7 observable symbols. ML is used to determine the parameters of the HMMs. So, this is how they derive facial expression.
They then combine the HMMs into a DBN in order to derive the underlying mental state of "agreement, concentrating, disagreement, thinking and unsure and interested". The mind-reading dataset used to train the system is one used in the study of autism (from Cambridge I think).

**21 August 2007 Y. Tian, T. Kanade, and J. Cohn, Recognizing Action Units for Facial Expression Analysis, tech. report CMU-RI-TR-99-40, Robotics Institute, Carnegie Mellon University, December, 1999.[11]**

This article is a few years old and raises a number of interesting points. It outlines a method for automating FACS recognition. Apart from the usual issues though, it raises raises robustness issues:

- Complex environment
- Various lighting conditions
- Occlusions
- Image resolution

It also raises the question as to whether geometric feature-based parametrization a good approach. Should they use:

- Geometric feature-based?
- Appearance-based: Gabors, Eigenfaces, ICA (and whether Holistic or local appearance)
- Motion: optical flow
- Hybrid

Also questions if a multilayer NN is the best approach. This is a question that comes up quite a bit.

**20 August 2007 Affective State Detection With Dynamic Bayesian Networks, Delft University of Technology [3]**

This is a literature survey of 17 papers with a peculiar scope. However, it does have an interesting discussion at the beginning on the merits of the probabilistic approach, e.g. HMM and DBN versus the NN approach. Obviously, the former win, hence the title. It does raise a good point though about the drawback with ANN is that it is difficult to know how the knowledge is distributed over the nodes of the NN. **This has implications for my work and I'd like to discuss it with someone. I think that I might know how to arrive at an answer.**

One of the papers surveyed in particular is interesting, especially as its to do with a system that has been built, is "Mind Reading Maching: Automated Inference of Cognitive Mental States from Video". This has some analog with my work. So, I went on to read it and another paper by one of the authors, "Real Time Inference of Complex Mental Sates from Facial Expressions and Head Gestures". It is reported above.

**17 August 2007 AAM Derived Face Representations for Robust Facial Action Recognition [9]** Explains a method of using 2d and 3d AAMs to map to FACS AUs (just AUs not emotions). It uses NN-LDA and SVM for the pattern recognition and the latter seems to perform better. The results of the 3d AAM were disappointing and the authors attribute this to noise associated with inferring the depth information of the face. Best results are achieved using 2d with removal of the similarity transform.

# 3 Milestones

## 3.1 Overall bi-annual milestones of PhD program

– The 8 bi-annual milestones for the entire 3 years of PhD research (starting August 2005)

| **Bi-annual Period** | **Milestone description** | **Major outcomes** |
|---|---|---|
| 1 (PT) Aug 2005 | Literature Review | |
| 2 (PT) | Literature Review | |
| 3 (PT) | Literature Review | Conference papers submitted<br>Conferences attended |
| 4 (PT) | Literature Review | Conference papers accepted<br>Journal papers submitted<br>Conferences attended |
| 5 (FT) July 2007 | system development | Requirements Statement<br>Collaboration with other areas<br>System design specification |
| 6 (FT) | System proof of concepts trial | Results evaluation<br>Journal papers submitted<br>Conferences attended |
| 7 (FT) | Anxiety recognition system development<br>System proof of concepts trial | Conference papers submitted<br>Journal papers submitted<br>Conferences attended |
| 8 (FT) | Summarising researches and results<br>Writing Thesis | Conference papers submitted<br>Journal papers submitted<br>Thesis submitted |

### 3.2 Monthly outcome (from Jul 07 - Dec 07)

| Month | Plan Description | Major outcomes |
|---|---|---|
| Jul 07 | Arranging commencement of Full-time status<br>Prepare and ready presentation<br>for Beijing International Conference<br>Self learning Active Appearance Models | Commence Full-Time<br>Attend conference |
| Aug 07 | Port FETA system to Windows<br><br><br><br>Engage School of Psychiatry<br>and Centre for Mental Health Research<br>Prepare detailed timeline for Study Plan<br>Procure a copy of FACS?<br>Self learn FACS<br>Try to identify an anxiety expression database | Windows based system<br>for evaluation, enhancement,<br>and to simplify presentations<br>Collaboration<br><br>Enhance Plan |
| Sep 07 | Develop expression recognition system<br>Self learn FACS | |
| Oct 07 | Prepare article for<br>Language and Speech: Special Issue<br>on Audiovisual Prosody<br>Develop expression recognition system<br>Self learn FACS | |
| Nov 07 | Submit article<br>Develop expression recognition system<br>Review plans with collaborators | |
| Dec 07 | Develop expression recognition system | |

## References


1. M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: Machine learning and application to spontaneous behavior. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 568–573, Washington, DC, USA, 2005. IEEE Computer Society.
2. J. Bosch, S. Mitchell, B. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. Reiber. Automatic segmentation of echocardiographic sequences by active appearance motion models. 21(11):1374–1383, November 2002.
3. M. A. de Jongh. Affective state detection with dynamic bayesian networks. Technical report, Delft University of Technology, 2005.
4. D. W. Hansen, J. P. Hansen, M. Niels, and M. B. Stegmann. Eye typing using markov and active appearance models. page 132, 2002.
5. H. J.A. and O. D.M. How do you look when feeling anxious? facial displays of anxiety. *Personality and Individual Differences*, 21, August 1996.
6. R. Kaliouby and P. Robinson. Real-time inference of complex mental states from facial expressions and head gestures. In *CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 10*, page 154, Washington, DC, USA, 2004. IEEE Computer Society.
7. J. McCall and M. Trivedi. Facial action coding using multiple visual cues and a hierarchy of particle filters. In *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, page 150, Washington, DC, USA, 2006. IEEE Computer Society.
8. R.Kalioub and P. Robinson. Mind reading machines: automated inference of cognitive mental states from video. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 682–688. IEEE, Oct 2004.
9. S. Lucey and I. Matthews and C. Hu and Z. Ambadar and F. de la Torre and J. Cohn. Aam derived face representations for robust facial action recognition. In *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 155–162, Washington, DC, USA, 2006. IEEE Computer Society.
10. A. Tan and D.Gilbert. An empirical comparison of supervised machine learning techniques in bioinformatics. In *APBC '03: Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics 2003*, pages 219–222, Darlinghurst, Australia, Australia, 2003. Australian Computer Society, Inc.
11. Y.-L. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. Technical Report CMU-RI-TR-99-40, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 1999.
12. C. van der Gaag, R. Minderaa, and C. Keysers. Facial expressions: What the mirror neuron system can and cannot tell us. *Social Neuroscience*, 2(3):179–222, 2007.