EYE MOVEMENT ANALYSIS FOR DEPRESSION DETECTION

Sharifa Alghowinem[1,5], Roland Goecke[2,1], Michael Wagner[2,1],
Gordon Parker[3], Michael Breakspear[4,3]

[1]Australian National University, Canberra, Australia, [2]University of Canberra, Canberra, Australia,
[3]University of New South Wales, Sydney, Australia, [4]Queensland Institute of Medical Research, Brisbane, Australia,
[5]Ministry of Higher Education: Kingdom of Saudi Arabia
sharifa.alghowinem@anu.edu.au, roland.goecke@ieee.org, michael.wagner@canberra.edu.au,
g.parker@blackdog.org.au, mjbreaks@gmail.com

ABSTRACT

Depression is a common and disabling mental health disorder, which impacts not only on the sufferer but also on their families, friends and the economy overall. Despite its high prevalence, current diagnosis relies almost exclusively on patient self-report and clinical opinion, leading to a number of subjective biases. Our aim is to develop an objective affective sensing system that supports clinicians in their diagnosis and monitoring of clinical depression. In this paper, we analyse the performance of eye movement features extracted from face videos using Active Appearance Models for a binary classification task (depressed vs. non-depressed). We find that eye movement low-level features gave 70% accuracy using a hybrid classifier of Gaussian Mixture Models and Support Vector Machines, and 75% accuracy when using statistical measures with SVM classifiers over the entire interview. We also investigate differences while expressing positive and negative emotions, as well as the classification performance in gender-dependent versus gender-independent modes. Interestingly, even though the blinking rate was not significantly different between depressed and healthy controls, we find that the average distance between the eyelids ('eye opening') was significantly smaller and the average duration of blinks significantly longer in depressed subjects, which might be an indication of fatigue or eye contact avoidance.

Index Terms— Eye movement, active appearance model, shape analysis, affective sensing

1. INTRODUCTION

Unlike mood fluctuation, clinical depression is a common mental disorder that lasts longer and causes disability and reduced functionality. Moreover, at its most severe level, it might lead to suicide. A recent World Health Organization (WHO [1]) report estimated that 350 million people worldwide are affected by depression. It causes more than two-thirds of suicides each year [2]. The suicide risk is more than 30 times higher among depressed patients than that of the population without these disorders [3]. Although treatment of depression disorders has proven to be effective in most cases [4], misdiagnosing depressed patients is a common barrier [5]. Based on the WHO report, the barriers to effective diagnosis of depression include a lack of resources and trained health care providers. Moreover, the assessment methods of diagnosing depression rely almost exclusively on patient-reported or clinical judgments of symptom severity [6], risking a range of subjective biases. Our goal here is to investigate the general characteristics of depression, which we

hope will lead to an objective affective sensing system that assists clinicians in their diagnosis and monitoring of clinical depression. Ultimately, we hope to assist patients with depression to monitor the progress of their illness in a similar way that a patient with diabetes monitors their blood sugar levels with a small portable device.

In this study, we investigate whether eye movement and blinking holds discriminative power for detecting depression. We extract eye movement features from face videos using user-specific Active Appearance Models (AAM), trained with 74 points located in the eye region. We examine the performance of these features using Gaussian Mixture Models (GMM) and Support Vector Machines (SVM) for classification on a real-world clinically validated dataset of 30 patients with severe depression and 30 healthy control subjects. We also analyse the statistical significance of these features between the two groups. In this paper, we investigate the differences in the eye movement while expressing positive and negative emotions, as well as differences in males and females from both groups.

2. BACKGROUND

Research into potential bio-markers of central nervous system disorders such as affective disorders have explored subtle changes in eye movements as possible physiologically based indicators of disease progression, severity or treatment efficacy [7]. Most eye movements analysed in the literature include: (1) saccades, which are the rapid movements of eyes while scanning a visual scene (e.g. reading); (2) pursuit movements, which are the eye movements while tracking a moving object; and (3) rapid eye movements (R.E.M.) sleep. The eye movements are usually measured using an electrooculograph, a pair of electrodes placed close to the subject's eyes, which produce an electrooculogram signal. Lipton *et al.* [8] found abnormal horizontal pursuit eye movements in depressed patients compared to healthy controls. This was also confirmed in [9], finding that the pursuit and saccade eye movement rates correlate strongly in controls but are reduced or absent in affective disorder patients, concluding abnormality in patients' ocular motor systems as a form of psycho-motor retardation. Crawford *et al.* [10] found the same abnormal saccades in patients even when not under medication. Furthermore, the results in [11] identify not only significant eye motor system disturbances in depression but also significant disturbances in their cognitive performance. R.E.M. sleep was also analysed in depressives, finding that R.E.M. latencies are shortened in depressives and suggesting that altered R.E.M. latency can be regarded as a single measure of the severity of illness [12]. Eye blink rate was investigated showing elevated blink rates, which return to normal

levels as the depressed patient's condition improves [13].

Computer vision offers a large variety of techniques that are employed for eye movement measurement including eye-marker cameras, head-mounted corneal reflex illumination, the contact lens method, etc. [14]. We are particularly interested in non-obtrusive techniques. Therefore, in terms of analysing eye movements in a video, techniques for locating the head first, then finding the eye feature points are used. Next, eye tracking image processing is commonly used. It is often the case that the best feature to track is the contour between the iris and the sclera known as the limbus [15]. Moreover, semi-automatic approaches such as Active Appearance Models – a generative method for modelling deformable visual objects – could be trained to model and track the eyes [16]. In the AAM approach, selected images of different eye variation (e.g. open, half open, closed eye) have to be annotated, before an eye AAM model can be trained. That model is then used to detect and track the eyes in the entire video.

To the best of our knowledge, this study based on automated computer vision techniques is the first attempt of using eye movement analysis for automatic depression analysis. In this paper, we use subject-specific AAMs to model and track the eyes in order to analyse the differences between depressed patients and healthy controls in eye movement behaviour in a recorded subject interview. Beside analysing the general differences in eye movements, we specifically investigate the differences in the eye movements while subjects express positive and negative emotions. In addition, we investigate the potential effect of gender-dependent classification of depression. After extracting features from each frame in the video recordings, we use a hybrid classifier based on a Gaussian Mixture Model for each subject and a Support Vector Machine for classification. We also analyse the eye movement features statistically in terms of functionals to identify how eye movements may differ between both groups.

3. METHOD

3.1. Real-World Clinically Validated Data

For the experimental validation, we used real-world data collected in an ongoing study at the Black Dog Institute, a clinical research facility in Sydney, Australia, offering specialist expertise in depression and bipolar disorder. Subjects included healthy controls as well as patients who had been diagnosed with depression. To date, data from over 40 depressed subjects with over 40 controls (both females and males) have been collected. Data were acquired after obtaining informed consent from the participants in accordance with approval from the local institutional ethics committee.

In this paper, a gender-balanced subset of 30 depressed subjects and 30 healthy controls was analysed. We acknowledge that the amount of data used here is relatively small, but this is a common problem in similar studies. As we continue to collect more data, future studies will be able to report on a larger dataset. The audio-video experimental paradigm contains several parts, including an interview with the subjects. The interview was conducted by asking specific open questions (in 8 question groups) to describe events that had aroused significant emotions. In this paper, the interview part with all 8 question groups is used for analysing spontaneous eye movements. We also compare eye movements while expressing positive and negative emotions by analysing two related questions from the interview: "Can you recall some recent *good news* you had and how did that make you feel?" and "Can you recall news of *bad or negative nature* and how did you feel about it?" For simplicity, these two questions will be referred to as "Good News" and "Bad News", resp.

Table 1. Duration of the interviews (in min) of depressed and control subjects

	All gender		Male Only		Female Only	
Questions	Depressed	Control	Depressed	Control	Depressed	Control
All 8 questions	309.2	199.8	126.6	112.0	182.5	87.8
"Good News"	34.7	21.3	14.2	9.8	20.5	11.5
"Bad News"	37.6	28.7	16.5	17.9	21.0	10.8

(a) 74 points AAM for Open Eyes

(b) 74 points AAM for Closed Eyes (*eyelids points overlapped*)

(c) Fitting AMM Example for Open Eyes

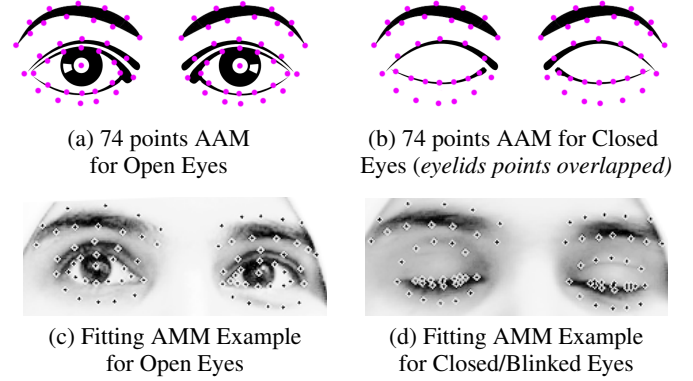(d) Fitting AMM Example for Closed/Blinked Eyes

Fig. 1. AAM with 74 points for both eyes including eyebrows (*showing the overlapped points between upper and lower eyelids including iris centre for closed eyes*)

We assume that these questions elicit the emotions, even though the answers were not validated for certain emotions. The interview was manually labelled to separate each question, where the total duration was 509min for all 60 subjects used in this study. The duration of the interviews of depressed and control subjects for all 8 questions, the "Good News" question and the "Bad News" question used in this paper is shown in Table 1.

3.2. Data Preparation

On average, 45 images per subject were manually selected having different eye status (e.g. open, half open, closed eye) and head position variation. Inspired by [16], these images were annotated using 74 points for both eyes including eyebrows and iris centres, using the distance the between eyelid points to determine the different eye status (open/closed) (see Figure 1 (a)&(b)). These annotated points and images are used to build subject-specific eye AAM, using linear parameters to update the model in an iterative framework as a discriminative fitting method [17]. The points of the trained model are initialised in a semi-automated manner, i.e. face detection [18] is performed in the first frame, then the rough eyes' location is estimated using the top third of the detected face. Finally, a fitting function is called. If the fitting function is not accurate enough, the steps of changing the model location and the calls for the fitting function are manually repeated before the tracking is started for the entire video. The trained model fits on and tracks the subject's eyes for the entire interview, producing the position of the 74 points in each frame. An example of the fitted AAM model is shown in Figure 1 (c)&(d).

3.3. Feature Extraction

For each eye, three features were extracted: horizontal, vertical, and eyelid movement. The horizontal movement is measured by the length of the line connecting the inner corner of the eye and the iris
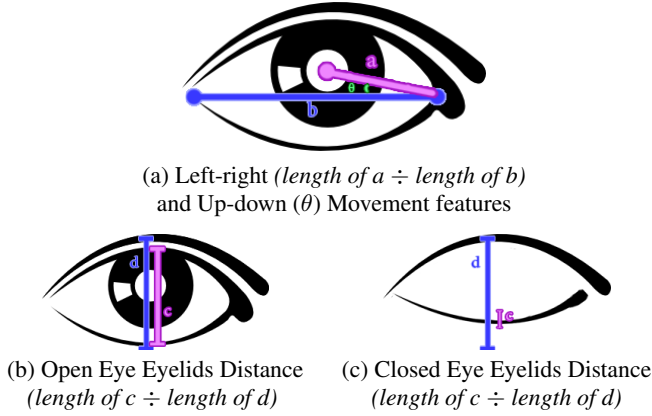
(a) Left-right *(length of a ÷ length of b)*
and Up-down *(θ)* Movement features

(b) Open Eye Eyelids Distance
(length of c ÷ length of d)

(c) Closed Eye Eyelids Distance
(length of c ÷ length of d)

Fig. 2. Extracting and normalising eye movement features

centre, which has been normalized based on the line connecting the eye corners (see Figure 2 (a)). The longer the line, the further the iris is from the inner eye corner and vise versa. The vertical movement is measured by the angle between the previous two lines (see Figure 2 (a)). A larger angle indicates a higher iris position and vise versa. On the other hand, distances between the eyelids are measured by the length of the line connecting the centre of both eyelids, normalised by the line connecting the centre of the eye fold and the centre of the outer border of the lower eyelid, which also apply for closed eyes (see Figure 2 (b)&(c)). These three features of each eye are extracted for each frame (25 frames per second), then their velocity and acceleration are extracted to give a total 18 features per frame. Outlier frames, which are caused by erroneous fitting of the AAM, or the absence of the eyes in a frame, are detected using Grubbs' test for outliers [19]. The detected outlier frames are skipped.

Moreover, a total of 126 statistical features "functionals" were extracted, which are:

- Maximum, minimum, mean, variance, and standard deviation for all 18 low-level features mentioned earlier (5×18)

- Maximum, minimum, and average of: looking left, right, up and down duration, as well as the blinks duration for both eyes ($3 \times 5 \times 2$)

- Closed eye duration rate, and closed eye to open eye duration rate for both eyes (2×2)

- Blinking rate for both eyes (1×2)

A blink is detected when the normalised eyelid distance is lower than the average of the normalised eyelids distance minus half the standard deviation of that feature for each subject. The same applies for the viewing direction.

3.4. Classification and Evaluation

Classification methods and techniques can be divided into two categories: generative and discriminative models. Generative models, such as the Gaussian Mixture Model, learn to cover a subspace that belongs to one class. Discriminative models, such as the Support Vector Machine, learn boundaries between two classes. Different classifiers have been used in emotion recognition; in this paper, we use GMM for classification and a hybrid classifier that combines GMM with SVM for the frame-by-frame features. For the GMM, we use 16 mixtures empirically chosen and fixed to ensure consistency in the comparison, knowing that some features benefit more

from more detailed modelling. For the hybrid classifier, we create a GMM for each subject, then feed these models to an SVM classifier. The use of GMMs served as a dimensionality reduction technique as well as for obtaining the same number of features to be fed to the SVM regardless of the duration of the subject's interview. On the other hand, for the functionals, we use an SVM for classification.

The eye movement patterns were initially classified in a binary, subject-independent scenario (i.e. depressed / non-depressed). In addition, a manual gender separation was used in our study to test the overall accuracy for each gender group. To mitigate the effect of the limited amount of data, leave-one-out cross-validation was used without any overlap between training and testing data for all classification tasks. In order to evaluate the performance of the system, several statistical measures could be evaluated, such as recall or precision. In this study, the average recall (AR) was computed.

In order to analyse the statistical significance of any differences in the functionals, a two-tailed T-test for two samples was performed, assuming unequal variances and $p = 0.05$. The state of the T-test was calculated to identify the direction of effect.

4. RESULTS

Previous eye movement studies on depressed subjects showed significant changes in affective disorder patients. Therefore, in future studies, eye movement could be treated as a complementary cue that would be measured simultaneously with other cues (e.g. speech). In this paper, we aim to measure the performance of eye movement features by themselves for the classification of depression to get a baseline result. We also analyse the eye movement features statistically to identify significant differences and the direction of effect in depressives and controls.

4.1. Classification Results

Table 2 shows the classification results for both frame-by-frame features and functionals using different classifiers. As can be seen, the GMM alone gave lower results. We show the GMM results here as a baseline for comparison with the hybrid classifier. Although eye movements would most likely be used as a complementary cue in an affective sensing system for depression, rather than as the sole measure, they performed well on their own. Low-level features gave 70% AR using the hybrid classifier, while functionals gave 75% AR, which are both above chance level. These findings support previous studies, which concluded abnormality in patients' ocular motor system [8, 9, 10, 12].

We also investigated the influence of gender-dependent classification, which as shown in the table has no significant difference. A study [20] about gender differences in depression behaviour pointed out that depressed women are more likely to be detected than depressed men, which might be caused by women amplifying their moods, while men are more likely to engage in distracting behaviours that dampen their mood when depressed. This finding was supported by our previous study on speech cues to detect depression, where we found a significantly higher depression detection in depressed women than depressed men using the same subjects as in this study [21]. Even though our eye movement classification results conflict with the conclusions of gender differences of a previous study [20, 21], it might indicate a physical abnormality in eye movement rather than a behavioural one.

Similar results were obtained when investigating expressions of positive and negative emotions. However, both gender-dependent

Table 2. Average recall in % for the classification results of different features

Classifier	Features per subject	All Questions Gender Independent	Males	Females	"Good News" Question Gender Independent	Males	Females	"Bad News" Question Gender Independent	Males	Females
GMM	18 features per frame for the duration of the interview	60	54	56	60	70	43	60	54	50
SVM	GMM model	70	75	68	63	70	68	73	70	68
SVM	126 statistical features	75	77	80	69	86	64	63	60	65

Table 3. Gender independent significant T-test results of eye movement features for different question parts

Eye	Feature	All Questions Direction	P-value	"Good News" Question Direction	P-value	"Bad News" Question Direction	P-value
Left	Right-Left Mean	D>C	0.02	NS	NS	D>C	0.003
	Eyelid Distance Mean	C>D	0.04	C>D	0.03	C>D	0.003
	Eyelid Distance Min	NS	NS	C>D	0.05	C>D	0.024
	Blink Duration Mean	D>C	0.05	NS	NS	NS	NS
	Blink Duration Max	D>C	0.01	D>C	0.04	NS	NS
Right	Eyelid Distance Mean	C>D	0.02	C>D	0.01	C>D	0.001
	Eyelid Distance Min	C>D	0.05	C>D	0.03	C>D	0.01
	Blink Duration Mean	D>C	0.02	D>C	0.05	D>C	0.05
	Blink Duration Max	D>C	0.01	D>C	0.002	NS	NS

NS: Not Significant

and emotion-dependent investigations gave relatively good recognition rates despite the huge reduction in the sample size. This result could be explained by the thin-slicing theory, i.e. when using different smaller parts of the interview, the performance is similar if not better than using the entire interview. The psychological thin-slicing theory states that using a brief observation, or thin slice, of behaviour can be used to predict physiological outcome at levels above that expected by chance [22].

4.2. Statistical Analysis

Table 3 shows only the significant T-test results of analysing the functionals extracted from different parts of the interviews. Given that the interviewer approximately stands in middle of the interviewee's field of view, the average distance between the left eye corner and the iris centre (i.e. left-right movement) is larger in depressed subjects, meaning more gazing to the left, which might be an indicator for avoiding eye contact with the interviewer. We could also see that this feature is not statisically significant while expressing positive emotions, although it is statisically significant while expressing negative emotions. On the other hand, the right eye has no significant result for the same feature. This might be due to, when a person looks to the left, the distance between the inner right eye corner and the iris centre being short, while the distance between the inner left eye corner to the iris centre is large. Interestingly, the results show that the average distance between the eyelids is significantly smaller in depressed subjects. Also, even though the blink rate was not significantly different between depressed and control subjects, the results show that the blink duration was significantly longer in depressed subjects. The latter findings suggest that, assuming that the physical need of a blink to lubricate the eye is similar in both groups, the longer duration of a blink and the smaller distance between the eyelids in depressed subjects might be an indication of fatigue, which is a common symptom of depression.

5. CONCLUSIONS AND FUTURE WORK

Our aim is to work towards an objective affective sensing system that supports clinicians in their diagnosis of clinical depression. To this end, we investigated eye movement patterns using Active Appearance Models for recognising depression from video data of subject interviews and whether initial manual gender splitting influences the recognition rate. Although eye movements are complementary features that could, in future studies, be fused with other cues (e.g. facial expressions, speech), they gave good results on their own (70% and 75% average recall for low-level features and functionals, resp.) as well. In general, we conclude that eye movement abnormality is a physical cue rather than behavioural one, which is in line with the psychology literature in that depression leads to psycho-motor retardation. Gender-dependent classification has no significant influence in recognising depression from eye movements.

Moreover, our investigation of the difference in expressing positive and negative emotions in depressed and control subjects resulted in no statistically significant difference being found. However, even with the reduction of the sample size in both gender-dependent and the positive and negative emotion expressing experiments, eye movement features extracted from face videos were still giving good recognition rates, which could be explained by the thin-slicing theory. Interestingly, we found that the average distance between the eyelids was significantly smaller and the average duration of blinks was significantly longer in depressed subjects, which might be an indication of fatigue and eye contact avoidance.

Fusing face, body, and voice [21] features with this current research will be a next step towards a multimodal system. A known limitation is the fairly small number of (depressed and control) subjects. This is a common problem in similar studies. As data collection is ongoing, we anticipate to report on a larger dataset in the future.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] C Mathers, J Boerma, and D Fat, *The Global Burden of Disease: 2004 Update*, WHO, Geneva, Switzerland, 2008.

[2] US Department of Health and Human Services, *Healthy People 2010: Understanding and improving health*, vol. 2, US Government Printing Office, Washington, DC, 2000.

[3] S B Guze and E Robins, "Suicide and Primary Affective Disorders," *The British Journal of Psychiatry*, vol. 117, no. 539, pp. 437–438, Oct. 1970.

[4] L G Kiloh, G Andrews, and M Neilson, "The long-term outcome of depressive illness," *The British Journal of Psychiatry*, vol. 153, pp. 752–757, Dec. 1988.

[5] N Niedermaier, E Bohrer, K Schulte, P Schlattmann, and I Heuser, "Misdiagnosed patients with bipolar disorder: comorbidities, treatment patterns, and direct treatment costs," *The Journal of clinical psychiatry*, vol. 66, no. 11, pp. 1619–1623, 2004.

[6] J C Mundt, P J Snyder, M S Cannizzaro, K Chappie, and D S Geralts, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response technology," *Journal of Neurolinguistics*, vol. 20, no. 1, pp. 50–64, 2007.

[7] N Kathmann, A Hochrein, R Uwer, and B Bondy, "Deficits in gain of smooth pursuit eye movements in schizophrenia and affective disorder patients and their unaffected relatives," *The American Journal of Psychiatry*, vol. 160, no. 4, pp. 696–702, 2003.

[8] R B Lipton, S Levin, and P S Holzman, "Horizontal and vertical pursuit eye movements, the oculocephalic reflex, and the functional psychoses," *Psychiatry Research*, vol. 3, no. 2, pp. 193–203, 1980.

[9] L A Abel, L Friedman, J Jesberger, A Malki, and H Y Meltzer, "Quantitative assessment of smooth pursuit gain and catch-up saccades in schizophrenia and affective disorders," *Biological Psychiatry*, vol. 29, no. 11, pp. 1063–1072, 1991.

[10] T J Crawford, B Haeger, C Kennard, M A Reveley, and L Henderson, "Saccadic abnormalities in psychotic patients. I. Neuroleptic-free psychotic patients," *Psychological Medicine*, vol. 25, no. 3, pp. 461–471, 1995.

[11] J A Sweeney, M H Strojwas, J J Mann, and M E Thase, "Prefrontal and cerebellar abnormalities in major depression: evidence from oculomotor studies," *Biological Psychiatry*, vol. 43, no. 8, pp. 584–594, 1998.

[12] D J Kupfer and F G Foster, "Interval Between Onset of Sleep and Rapid-eye-movement Sleep as an Indicator of Depression," *The Lancet*, vol. 300, no. 7779, pp. 684–686, Sept. 1972.

[13] J H Mackintosh, R Kumar, and T Kitamura, "Blink rate in psychiatric illness," *The British Journal of Psychiatry*, vol. 143, no. 1, pp. 55–57, 1983.

[14] L R Young and D Sheena, "Survey of eye movement recording methods," *Behavior Research Methods & Instrumentation*, vol. 7, no. 5, pp. 397–429, Sept. 1975.

[15] D Winfield and D J Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*. vol. 3, pp. 79–79, IEEE.

[16] Ioana Bacivarov, Mircea Ionita, and Peter Corcoran, "Statistical models of appearance for eye tracking and eye-blink detection and measurement," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1312–1320, 2008.

[17] J Saragih and R Goecke, "Iterative Error Bound Minimisation for AAM Alignment," in *ICPR2006*, Aug. 2006, vol. 2, pp. 1192–1195.

[18] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2001*. IEEE, 2001, vol. 1, pp. I–511.

[19] F E Grubbs, "Procedures for Detecting Outlying Observations in Samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[20] Susan Nolen-Hoeksema, "Sex differences in unipolar depression: Evidence and theory," *Psychol*, , no. 101, pp. 259–282, 1987.

[21] S Alghowinem, R Goecke, M Wagner, J Epps, M Breakspear, and G Parker, "From Joyous to Clinically Depressed: Mood Detection Using Spontaneous Speech," in *Proc. FLAIRS-25*, 2012, pp. 141–146.

[22] N Ambady and R Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis," *Psychological Bulletin*, vol. 111, no. 2, pp. 256, 1992.