Automated Facial Expressions Analysis in Schizophrenia: a Continuous Dynamic Approach

Talia Tron[1,4], Abraham Peled[2,3], Alexander Grinsphoon[2,3], and Daphna Weinshall[4]

[1] Hebrew University, ELSC center for Brain Science, Jerusalem 91904, Israel
[2] Sha'ar Menashe Mental Health Center, Sha'ar Menashe 38706, Israel
[3] Technion Institute of Technology, Rappaport Faculty of Medicine, Haifa 3200003, Israel.
[4] Hebrew University, School of Computer Science and Engineering, Jerusalem 91904, Israel
Talia.Tron@mail.huji.ac.il

Abstract. Facial expressions play a major role in psychiatric diagnosis, monitoring and treatment adjustment. We recorded 34 schizophrenia patients and matched controls during a clinical interview, and extracted the activity level of 23 facial Action Units (AUs), using 3D structured light cameras and dedicated software. By defining dynamic and intensity AUs activation characteristic features, we found evidence for blunted affect and reduced positive emotional expressions in patients. Further, we designed learning algorithms which achieved up to 85% correct schizophrenia classification rate, and significant correlation with negative symptoms severity. Our results emphasize the clinical importance of facial dynamics, and illustrate the possible advantages of employing affective computing tools in clinical settings.

Key words: Schizophrenia, Machine learning, Mental health, Facial expressions, 3D cameras, FACS

1 Introduction

Both clinical observations and computational studies suggest that facial activity plays a major role in signaling people's emotional and mental state [14, 8, 13]. Accordingly, several mental disorders are manifested by reduced or altered facial activity, and facial observations are an integral part of psychiatric diagnosis. To date, there are no objective, quantitative methods to measure these alterations, and no clear relation between them and the underlying brain disturbances. This causes multiple interpretations of phenomenology and results in low reliability and validity of psychiatric diagnosis [2].

Schizophrenia is one of the most severe mental disorders, with lifetime prevalence of about 1% worldwide. The disorder is characterized by negative symptoms, which involve the loss of functions and abilities (e.g. blunted affect), and by positive symptoms, which are pathological functions not present in healthy individuals (e.g. hallucinations). Studies have found that patients with schizophre-

nia demonstrate less positive emotions than controls [10], and lower congruity of emotional response [1]. Furthermore, there has been evidence for reduced upper facial activity [3] and reduced overall facial expressivity [12, 5, 7]. Nonetheless, these studies use a limited set of facial activity characteristic features, not necessarily ecologically relevant, and ignore information regarding facial dynamics and variability. An extensive use of computational methods together with clinical intuition is needed in order to obtain a more comprehensive description of patients behavior.

Our study combines descriptive methods with data-driven analysis. We use machine learning tools and cutting edge technology, in order to study a wide range of facial activity characteristic features, the relation between them, and the way they are manifested in clinical setting.

2 Materials and Methods

2.1 Study Design

Participants The study was done in collaboration with Sha'ar Menashe mental health center. Participants were 34 patients and 33 control subjects. All patients were diagnosed as suffering from schizophrenia according to DSM-5, and the course of illness in these patients varied from 1.5 years up to 37 years, with mean of 16.9 years. All patients but one were under stable drug treatment (mood stabilizer, antidepressant, antipsychotic and/or sedatives). Informed consent was obtained from all individual participants included in the study.

Psychiatric Evaluation Participants were evaluated by a trained psychiatrist using the *Positive and Negative Symptoms Scale* (PANSS), a 30 item scale especially designed to asses the severity of both negative and positive symptoms in schizophrenia [9]. The majority of patients suffered from post-psychotic residual negative signs (Type II) schizophrenia, namely, they showed severe negative symptoms (higher than 5 in the PANSS scale), while severe positive and general symptoms were rather rare (less than 10% of patients). 16 of the symptoms did not vary enough for statistical analysis and learning; therefore, the analysis focused on the remaining symptoms: 3 positive symptoms (Delusions, Conceptual disorganization and Grandiosity), 2 general symptoms (Motor retardation and Poor attention) and 7 negative symptoms (Blunted affect, Emotional withdrawal, Poor rapport, Passive/apathetic social withdrawal, Difficulty in abstract thinking, Lack of spontaneity and flow of conversation and Stereotyped thinking). To test for diagnosis consistency, the PANSS evaluation was repeated independently by a second trained psychiatrist who watched the interview videos. Inter-rater reliability was calculated separately for each PANSS symptom using Pearson correlation test.

Experimental Paradigm All subjects were individually recorded using a 3D structured light camera (carmine 1.09), during a 15 minute long interview con-

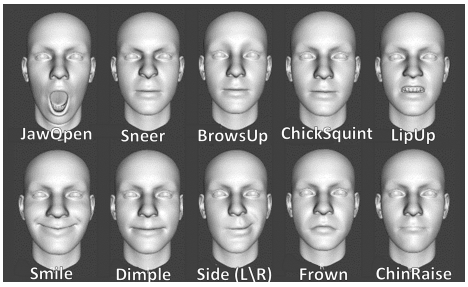
ducted by a trained psychiatrist. The interview was constructed out of one general question ('Tell me about yourself'), and three emotionally evocative questions regarding subject's current mood and recent emotional events. The camera was placed on the table between subject and interviewer, in a way that did not interfere with eye contact and none of the subjects reported discomfort from being recorded. All procedures performed in the study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

2.2 Facial Activity Features

The Facial Action Coding System (FACS) scores the activity of 46 individual facial muscles called Action Units (AUs) based on their intensity level and temporal segments [4]. Scoring is traditionally done manually, one frame at a time, by certified FACS coders, and automated FACS coding poses a major challenge in the field of affective computing. The advantage of the coding system is that it does not interpret the emotional value of specific features, and allows for a continuous and dynamic facial activity analysis.

Facial Activity Extraction For AUs activity extraction we used the *Faceshift*© commercial software which provides real time 3D face and head tracking, and which is typically used for animating avatars in film and game industry (`www.faceshift.com`). The software automatically analyzes data from 3D cameras based on structured light technology. These cameras capture facial surface data, which is less sensitive to head pose and to lightning conditions than 2D data, and yields a better recognition rate of AUs [11]. *Faceshift* outputs the intensity level over time for 48 AUs. The output was manually evaluated for tracking sensitivity and noise level. Subsequently, 23 *Faceshift* Action Units (AUs) were selected for further analysis and learning, including Brows-up (center, left and right), Mouth-side (left or right), Jaw-open, Lips-up, Lips-Funnel, Eye-In-Right (looking left), Chin-raise, Sneer and both sides (left and right) of Blink, Smile, Frown, Dimple, Lips-Stretch, and Chick-squint (see Fig. 1)

Fig. 1: Illustration of *Faceshift* facial Action Units (AUs) used for learning.



Characteristic Features Computation In order to obtain a detailed characterization of facial behavior, which captures both the dynamics and intensity of the activity in a clinically relevant manner, we calculated 5 characteristic features separately for each AU. First, the raw *Faceshift* signal was quantized using k-means (k=4) clustering. Then a transition matrix was generated, measuring the number of transitions between quantization levels. 5 facial activity characteristic features were then computed:

1. *Activation Ratio* - Fraction of segment during which the AU was activated
2. *Activation Level* - Mean intensity of AU activation
3. *Activation Length* - Number of frames that the AU activation lasted
4. *Change Ratio* - Fraction of the period of AU activation when there was a change in activity level
5. *Fast Change Ratio* - Fraction of fast changes (> 1) in activation level

Activation Level and *Change Ratio* were calculated using frames with non-zero activity only, so that they will not overlap with the *Activation Ratio*. For *Fast Change Ratio*, we normalized the number of fast changes frames by the total number of frames with activity change.

3 Analysis and Learning

The first part of our analysis was descriptive, and was aimed to obtain detailed characterization of facial activity in patients in comparison with controls. In the second part, we applied machine learning tools to generate predictions. We tested whether facial features have predictive power for patients vs. control classification, and for evaluating symptoms severity. To exclude possible confounds such as gender, education level, age and religion, we performed one-way ANOVA; a variable that was found to be different between groups, was further investigated for its effect on facial activity within groups.

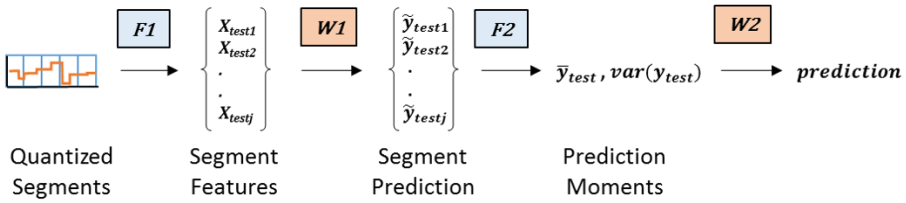
Descriptive Data Analysis In the descriptive part of the analysis, we explored how the facial activity is altered in different parts of the face, paying special attention to smiles. This was done using two tail student's t-tests on the *Activity Level* of each AU separately. For smiles, we further analyzed the difference in all characteristic features, using separate t-test for each feature type. The AU activity was given an emotional interpretation (e.g. high smile level indicates positive emotion), based on the *Emotional Facial Action Coding System* (EMFACS) developed by Paul Ekman, which systematically categorizes combination of AUs to specific emotional categories [6].

To study the way blunted affect is manifested in patients, we performed a regularized ridge regression between symptom severity and all features over all AUs. Feature selection (n=10) was done using f-regression, based on d' scores. Regression results were evaluated by $Pearson's\ R$, and the output regression weights were used for further feature type analysis.

Machine Learning Tools To test the predictive power of our features we trained a learner on train data and evaluated its performance on one test patient at a time, following the Leave-One-Out (LOO) procedure. The basic learning algorithm we used was Support Vector Machine (SVM) for patients vs. control classification, and ridge regression for symptom severity prediction. Before the regression, principle component analysis (PCA) was performed on train data separately for each feature type, resulting in a mixture of AUs. Feature selection was performed based on train data using f-regression (for SVM), or by selecting the highest PCA components (for regression).

To increase learning robustness, we employed a two step prediction algorithm, where each stage is learned separately from train data (see Fig. 2). Interview data of each individual subject was divided into 30 seconds long segments, and 5 representative features were computed separately for each segment ($F1$). In step 1, a learner was trained on the segments of all train subjects, giving as output the first model weights ($W1$) and a prediction for each segment. In step 2, prediction mean and standard deviation over all segments were calculated for each subject ($F2$), and a second learner was trained to predict a participant's label from these moments ($W2$).

Fig. 2: Illustration of the 2-step algorithm.



Performance evaluation was done between-subjects, namely, all segments of one subjects were left out for testing the algorithm. The SVM classifier was evaluated by the area under the Receiver Operator Curve (AUC), a combined measure for the learner's sensitivity (true positive rate) and specificity (true negative rate) with 1 signaling perfect separation and 0.5 signaling chance. Regression results were evaluated by $Pearson's\ R$ between the psychiatrist score and the algorithm prediction, separately for each PANSS symptom.

4 Results

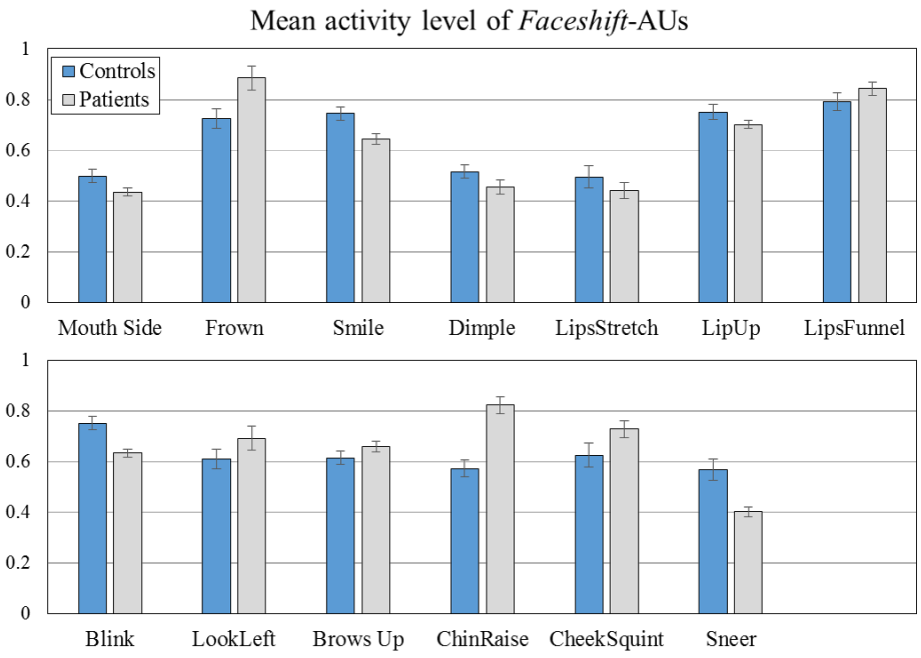
4.1 Inter-rater Reliability

All negative symptoms scores were at high agreement between raters (with an average of $R = 0.850$, $p << 0.01$), and so was 3 positive symptoms ($R = 0.630$, $p = 0.021$ for Delusions, $R = 0.880$, $p << 0.01$ for Conceptual disorganization) and one general symptoms ($R = 0.671$, $p << 0.01$ for Motor Retardation). Poor Attention and Grandiosity were not significantly correlated between raters.

4.2 Facial Activity, Descriptive Analysis

Facial Parts Analysis We found a significant difference in the *Activation Level* of 16 out of 23 *Facefhit*-AUs (see Fig. 3). Specifically, patients demonstrated lower level of activity in Smile, Dimple, Lip-stretch and Lip-up ($p << 0.01$), AUs which are typically in correspondence with positive emotional state. Frowns, Brows-Up and Chin-raise, on the other hand, were at much higher level in patients than in controls, which may indicate the presence of negative valance emotions (sadness, surprise and fear). Although those facial expressions were more intense, they changed more slowly, with reduced *Change Ratio* ($p = 0.004$ for Chin-raise) and *Fast Change Ratio* ($p << 0.01$ for both Chin-raise and Frowns). Blink *Activation Level* was reduced in patients, which in the *Faceshift* framework could mean that they closed their eyes less than controls. Sneer *Activation Level* was also significantly reduced. The level of Cheek-Squint activation was surprisingly enhanced in patients.

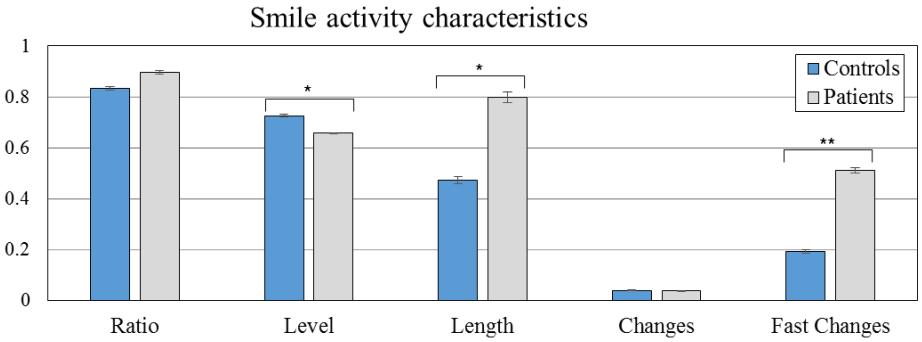
Fig. 3: Mean *Activation level* of facial Action Units in patients and controls. Only significantly different results are presented ($p < 0.05$ in student's t-test).



Smiles Analysis A closer look at smile activation (Fig 4) reveals that in comparison with controls, smile *Activation Level* was reduced, while *Activation Length* and *Fast Change Ratio* were significantly enhanced in patients. These

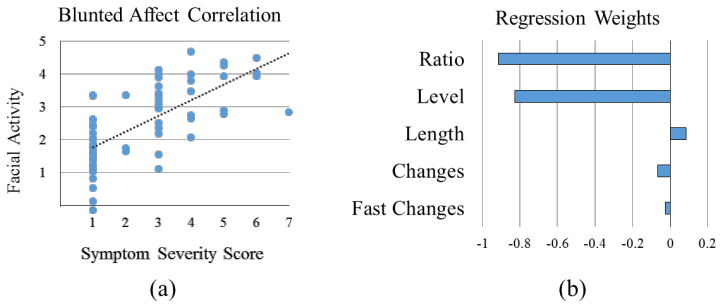
results suggest that in clinical settings, patients may not necessarily smile less, but rather their smiles are at lower intensity, longer, and with faster onset and offset (aka frozen or fake smiles).

Fig. 4: Smile activation characteristic features for patients and controls.



Blunted Affect Regression results (Fig. 5) suggest a significant correlation between AUs activation features and psychiatric evaluation of blunted affect severity ($R_{Pearson} = 0.686$, $p << 0.01$). Based on the regression weights, the two most discriminative AU features were *Activation Level* and *Activation Ratio*, which were in negative correlation with symptom's score. *Change Ratio* and *Fast Change Ratio* were also given negative weights, while *Activation Length* seemed to be positively correlated with the severity of the symptom. These Results are consistent with clinical observations.

Fig. 5: (a) Regression between blunted affect severity and facial activity features. (b) Weights given to each feature by the regression model.

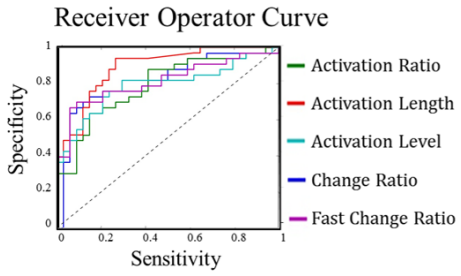


Possible Confounds One-way ANOVA on patients and controls data revealed significant difference between groups for gender ($F = 16.77$, $p << 0.01$) and education level ($F = 6.42$, $p = 0.014$). Neither of these variables was found to have a significant effect on facial activation characteristic within each group. The possible effect of neuroleptic drugs on observed facial activity could not be excluded, since all of our patients were under drug treatment, and additional control is needed.

4.3 Facial Expression Predictive Power

Patients vs. Controls Classification We employed the 2-step learning algorithm one feature type at a time, and using all features together. Each of the feature types was distinctive on its own on test data with AUC significantly better than chance (Fig 6). *Activation Length* gave out the best classification results ($AUC = 0.887$), followed by *Fast change ratio* ($AUC = 0.815$) and *Fast change ratio* ($AUC = 0.814$). This indicates the importance of looking at the the duration and dynamic of facial activity, rather than general intensity measures. The predictive power of using all features together was slightly lower ($AUC = 0.799$), most likely as a result of small sample and subsequent over-fitting.

Fig. 6: (a) ROC curves of each feature type for patients vs. control classification. (b) Classification results summarized as Area Under the ROC Curve (AUC).



(a)

Classification Results

	AUC
Activation Ratio	0.789
Activation Length	0.887
Activation Level	0.790
Change Ratio	0.815
Fast Change Ratio	0.814
All Features	0.799

(b)

PANSS Severity Regression For all negative symptoms, the prediction of the algorithm was significantly correlated with the score given by the psychiatrist ($R > 0.3$, $p \leqslant 0.01$). No such significance was found for any of the positive symptoms, which can be explained by the small variability of positive symptoms scores in our data. We got an unexpected result for general symptoms, with significant correlation only for Poor attention ($R = 0.292$,$p < 0.05$), which outperform the inter-rater correlation for this symptom. Train and test results are summarized in Table 1.

Code	PANSS symptom	Train R	p-val	Test R	p-val
G11	Motor retardation	0.463	1.023E-03	0.154	0.213
G7	Poor attention	0.566	9.35E-07	0.292	0.0166
N1	Blunted affect	0.686	8.27E-10	0.530	4.042E-06
N2	Emotional withdrawal	0.652	4.52E-09	0.510	1.045E-05
N3	Poor rapport	0.550	2.53E-06	0.315	0.00949
N4	Passive/apatheticsocial withdrawal	0.548	2.89E-06	0.368	0.00216
N5	Difficulty in abstract thinking	0.585	3.83E-07	0.369	0.00211
N6	Lack of spontaneity and conversation flow	0.555	1.58E-06	0.301	0.0133
N7	Stereotyped thinking	0.539	3.86E-06	0.369	0.00211
P1	Delusions	0.344	0.005	0.017	0.891
P2	Conceptual disorganization	0.332	0.007	0.065	0.600
P5	Hallucinations	0.306	0.013	0.055	0.660

Table 1: Summary of ridge regression results on train and test data, separately for each PANSS symptom. $Pearson's\ R$ was calculated between the algorithm prediction and symptom severity as scored by a trained psychiatrist.

5 Discussion

Our results are in excellent agreement with previous studies and reported clinical observations. We found clear evidence for clinically reported phenomenon such as blunted affect and lack of positive emotional expressions, and demonstrated how the disorder is manifested differently in different facial parts. Our findings highlight the importance of looking at dynamic characteristics of facial activity and may be employed in clinical settings.

The results give hope that real time automated facial analysis may one day be used for disease monitoring, drug adjustment and treatment outcome evaluation. To achieve these goals, future studies should include monitoring facial activity over time, studying Type-I (positive symptom) schizophrenia patients, and controlling subjects' drug usage. Other future directions include broadening facial activity research to other disorders such as depression and autism, and investigating the relation to neural mechanisms and cognitive performance.

Acknowledgements: This work was supported in part by the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI), and the Gatsby Charitable Foundations.

References

1. Giuseppe Bersani, Elisa Polli, and Giuseppe Valeriani. Facial expression in patients with bipolar disorder and schizophrenia in response to emotional stimuli: a partially shared cognitive and social deficit of the two. *Neuropsychiatric Disease and Treatment*, 9:1137–1144, 2013.

2. Jeffrey F. Cohn, Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai Nguyen, Margara Tejera Padilla, Feng Zhou, and Fernando De la Torre. Detecting depression from facial actions and vocal prosody. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–7, September 2009.
3. Paul Ekman and Wallace V Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Nonverbal communication, interaction, and gesture*, pages 57–106, 1981.
4. Paul Ekman and Erika L Rosenberg. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.
5. Irina Falkenberg, Mathias Bartels, and Barbara Wild. Keep smiling! *European archives of psychiatry and clinical neuroscience*, 258(4):245–253, 2008.
6. Wallace V Friesen and Paul Ekman. Emfacs-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco*, 2:36, 1983.
7. Wolfgang Gaebel and Wolfgang Wölwer. Facial expressivity in the course of schizophrenia and depression. *European archives of psychiatry and clinical neuroscience*, 254(5):335–342, 2004.
8. Hatice Gunes and Maja Pantic. Automatic, Dimensional and Continuous Emotion Recognition. *International Journal of Synthetic Emotions*, 1(1):68–99, 2010.
9. Stanley R Kay, Abraham Flszbein, and Lewis A Opfer. The positive and negative syndrome scale (panss) for schizophrenia. *Schizophrenia bulletin*, 13(2):261, 1987.
10. Annett Lotzin, Barbara Haack-Dees, Franz Resch, Georg Romer, and Brigitte Ramsauer. Facial emotional expression in schizophrenia adolescents during verbal interaction with a parent. *European archives of psychiatry and clinical neuroscience*, 263(6):529–36, 2013.
11. Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Lijun Yin. Static and dynamic 3d facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10):683–697, 2012.
12. Gwenda Simons, Johann Heinrich Ellgring, Katja Beck-Dossler, Wolfgang Gaebel, and Wolfgang Wölwer. Facial expression in male and female schizophrenia patients. *European archives of psychiatry and clinical neuroscience*, 260(3):267–76, 2010.
13. Fabien Trémeau, Dolores Malaspina, Fabrice Duval, Humberto Corrêa, Michaela Hager-Budny, Laura Coin-Bariou, Jean-Paul Macher, and Jack M Gorman. Facial expressiveness in patients with schizophrenia compared to depressed patients and nonpatient comparison subjects. *The American journal of psychiatry*, 162(1):92–101, 2005.
14. Michel Valstar, B Schuller, Kirsty Smith, and Florian Eyben. AVEC 2013 - The Continuous Audio/Visual Emotion and Depression Recognition Challenge. *cs.nott.ac.uk*, 2013.