Facial Expressions and Flat Affect in Schizophrenia, Automatic Analysis from Depth Camera Data

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

***Abstract*— One of the prominent clinical manifestations of schizophrenia is flat or altered facial activity, and flattening of emotional expressiveness (*Flat Affect*). In this study we used a structured-light depth camera and dedicated software to automatically measure the facial activity of schizophrenia patients and healthy individuals during a short structured interview. Based on K-means clustering analysis, facial activity was characterized in terms of *Typicality*, *Richness* and *Distribution* of 7 facial-clusters. Thus we found patients' facial activity to be poorer, more typical, and characterized mainly by neutral (flat) expressions. The facial features defined in our study achieved up to 85% correct diagnosis classification rate in a SVM based two-step algorithm, and were in significant correlation with *Flat Affect* severity. Our results demonstrate how the use of assistive technology and data-driven computational tools allow for a comprehensive description of patients' facial behavior in clinical settings, and may contribute to the reliability and accuracy of psychiatric diagnosis.**

I. INTRODUCTION

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. lack of motivation, cognitive impairments), and by positive symptoms, which are pathological functions not present in healthy individuals (e.g. hallucinations and delusions). Both clinical observations and computational studies suggest that schizophrenia is manifested by reduced or altered facial activity, and by overall affective flattening [15], [14]. *Flat affect*, also known as blunted affect, is clinically defined as 'a severe reduction in emotional expressiveness', and may be expressed in diminished facial expressions, monotonic speech, lack of expressive gestures, and overall apathetic appearance [11]. It is a matter of debate whether the observed flattening is a result of motor or emotional deficits, nonetheless, there is evidence for high congruence between symptom severity, patients wellbeing and treatment outcome [1].

Facial activity is traditionally analyzed in terms of emotional 'prototype expressions' such as anger, fear, sadness, happiness, and disgust [4] in what is known as the *categorical approach* of emotions. Using this approach, it has been shown that patients with schizophrenia demonstrate less positive emotions than controls [12], and lower congruity of emotional response [3]. The downfall of the approach however, is that it uses exaggerated, static and posed facial expressions, while those presented in everyday life are dynamic, spontaneous and far more subtle [9], [2]. An alternative approach is to analyze the facial activity without interpreting its emotional state, which is commonly done using the *Facial Action Coding System* (FACS). This system scores the activity of roughly 46 individual facial muscles called *Action Units* (AUs), based on their intensity level and temporal segments. FACS has been mapped into prototype emotions using the *Emotional Facial Action Coding System* (EMFACS), which systematically categorizes combination of AUs to specific emotions [7] but it can also be used independently. Schizophrenia studies based on FACS has found evidence for reduced upper facial activity [5] and reduced overall facial expressivity [13], [6], [8]. Nonetheless, these studies use a limited set of facial activity characteristic features, not necessarily ecologically relevant, and ignore information regarding facial 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 suggests a new data-driven approach, combining FACS analysis with the assumption that typical universal emotions can be discovered in a bottom-up analysis. We combine cutting edge technology with data-driven analysis to define a set of 'prototype' facial expression clusters, and to characterize facial activity in terms of *Typicality*, *Richness* and *Distribution* of these clusters. This allow us to study a wide range of facial features, the relation between them, and the way they are manifested in clinical setting.

II. MATERIALS AND METHODS

A. Study Design

The study was done in collaboration with Sha'ar Menashe mental health center. Participants were 34 patients diagnosed as suffering from schizophrenia according to DSM-5 and 33 control subjects. The duration of illness in participating patients was $1.5-37$ years (mean=16.9), and all but one were under stable drug treatment. Informed consent was obtained from all individual participants included in the study.

Participants were individually recorded using a structured-light depth camera (carmine 1.09), during a short structured interview done by a trained psychiatrist which included four questions regarding their emotional state. They then underwent a psychiatric evaluation using the *Positive and*

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

[1]ELSC center for Brain Science, Hebrew University, Jerusalem 91904, Israel talia.tron@mail.huji.ac.il

[2]Rappaport Faculty of Medicine, Technion Institute of Technology, Haifa 3200003, Israel

[3]Sha'ar Menashe Mental Health Center, Sha'ar Menashe 38706, Israel

[4]Hebrew University, School of Computer Science and Engineering, Jerusalem 91904, Israel

978-1-5090-2455-1/16/$31.00 ©2016 IEEE

Negative Symptoms Scale (PANSS), a 30 item scale especially designed to assess the severity of both negative and positive symptoms in schizophrenia [10]. 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.

B. Facial Activity Extraction

Facial AUs extraction out of depth camera video was done using *Faceshift*© commercial software, which provides real time 3D face and head tracking (`www.faceshift.com`). The software automatically analyzes data from depth-cameras based on structured light technology. The output includes the intensity level over time for 48 facial Action Units (AUs), corresponding to the FACS AUs described in Section I. *Faceshift* output was manually evaluated for tracking sensitivity and noise level, and subsequently 23 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*.

C. Facial-Cluster Characterization

In order to find the most common combinations of facial-AUs activation in our data, the 23 dimensional vector returned by *Faceshift* was segmented using k-means clustering on data from all subjects simultaneously. Subsequently, each video frame was assigned a cluster label $i \in [k]$ representing its closest cluster centroid (c_i). The resulting facial-cluster centroids can be thought of as the data-driven facial 'prototypes', somewhat equivalent to the categorical expressions described in I, but with no theoretical assumptions regarding the nature of emotions.

The optimal number of clusters was determined using the "elbow criterion". Let V_k be the percent of data variance explained by k centroids. Then $\Delta V = V_k - V_{k-1}$ denotes the difference in the percent of reduced variance when adding one cluster. Under the assumption that ΔV is F distributed, we look for the highest k such that ΔV is statistically significant. In other words, adding more clusters will not significantly improve the ratio of variance explained.

The new vector representation was used to quantitatively describe facial activity in terms of *Richness* (how many prototype expressions appeared), *Typicality* (how similar they were to the prototype) and *Distribution* (which expressions were more prevalent). Facial features were calculated individually for each subject in the following manner:

1) Richness: Let n denote the number of clusters that appeared in a subject's video clip, and k the number of clusters used for the k-means algorithm:

$$Richness = \frac{n-1}{k-1} \tag{1}$$

This measure varies from 0 (only one cluster appeared in the video) to 1 (full richness, all clusters appeared); it can be thought of as a measure for the diversity in facial activity throughout the video.

2) Typicality: Let the Within-Cluster Sum of Squares ($WCSS$) be the sum of distances of each data point x in cluster C_i from its nearest cluster centroid (c_i), with an additional sum over clusters:

$$WCSS_k = \sum_{i=1}^{k} \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - c_i\|^2 \tag{2}$$

For $k = 1$, $WCSS_1$ is proportional to the data variance (the average sqaured distance of the raw data from its mean). For $k > 1$, we define *Typicality* as the percent of the general variance which remains after adding more clusters:

$$Typicality = 1 - \frac{WCSS_k}{WCSS_1} \tag{3}$$

In facial activity terms, we can think of $WCSS_k$ as measuring how similar the video-frame activation is to its assigned 'prototype' among the k facial-clusters. Thus *Cluster Typicality* with score close to 1 indicates that the subject's expressions are similar to the prototypes, while a score close to 0 indicates a significant variability around the prototypes.

3) Cluster Distribution: For each facial-cluster i separately, we counted the number of frames in which it appeared t_i, and normalized it by the length of the video clip T. This allowed for a specific comparison between subjects over the degree of activation of each cluster (or prototype) among the different facial-clusters.

$$Cluster\ Distribution_i = \frac{t_i}{T} \quad \forall i \in [k] \tag{4}$$

D. Data Analysis

Patients vs. controls differences were tested using two-tail student's t-test for *Cluster Typicality*, *Richness*, and *Cluster Distribution* (separately for each facial-cluster). Result significance was evaluated using the *Bonferroni correction*, a family-wise error rate (FWER) for multi-hypothesis testing. In order to allow comparison of our results with the *categorical approach* emotions, the k centroids returned by the clustering algorithm were also evaluated for their affective meaning based on EMFACS (see Section I).

The relation between facial-cluster features and the severity of the *Flat Affect* symptom was tested using regularized ridge regression. A regression model was built for each facial feature separately and for all features together, using a custom designed two-step algorithm (see II-E). *Pearson's* correlation coefficient between the algorithm prediction and the *Flat Affect* score was calculated on train and test data. Symptom's severity was also tested for correlation with all other clinical symptoms scores evaluated by the psychiatrist.

To test for diagnosis consistency, PANSS evaluation was repeated independently by a second trained psychiatrist who watched the interview videos. Inter-rater agreement for *Flat Affect* was tested using *Pearson's R*. Finally, to exclude possible confounds such as gender, education level, age and religion, one-way ANOVA was performed; a variable that was found to be different between groups, was further investigated for its effect on facial activity within groups.

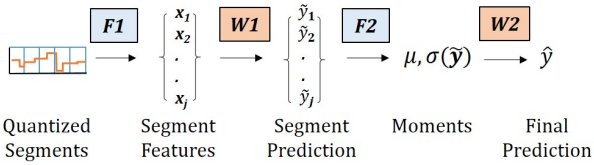

Fig. 1: Illustration of the 2-step algorithm used for learning. Interview data of each individual subject was divided into 30 seconds long segments, and features were calculated separately for each segment ($F1$). In step 1, a learner was trained on the segments of all train subjects, giving as output the first model weights ($W1$) and a prediction for each segment. In step 2 for each subject, prediction mean and standard deviation over all segments were calculated ($F2$) and a second learner was trained to predict subject's label from these moments ($W2$)

E. Learning

In order to evaluate the predictive power of facial-cluster features, we trained a support vector machine (SVM) for patients vs. control classification, and a regularized ridge regression model for *Flat Affect* severity prediction. To increase learning robustness, we employed a two step prediction algorithm, where each stage is learned separately from train data (Fig. 1). The algorithm was trained and tested separately for each feature, and using all features together, following a Leave-One-Out (LOO) procedure with f-regression feature selection (n=5).

Learning performance was evaluated by the Area Under the Receiver Operator Curve (AUC), a combined measure for the learning sensitivity (true positive rate) and specificity (true negative rate) with 1 signaling perfect separation and 0.5 signaling chance. *Pearson's R* was calculated between *Flat Affect* severity score and the algorithm's prediction.

III. RESULTS

A. Facial-Clusters Characteristics

Following the elbow method described in Section II-C, $k = 7$ was chosen for K-means clustering segmentation. Fig. 2 illustrates the centroids of 3 out of 7 facial-clusters returned by the clustering algorithm. The centroid of facial-cluster C_1 (c_1) is characterized by low intensity in all AUs, and may be interpreted as neutral or flat expression. In c_4 we see high intensity of *'ChinLowerRaise'* and *'LipsStretch'*, which correspond with negative valence emotions such as sadness, fear, or anger (according to EMFACS [7]). c_7, on the other hand, is characterized by high intensity smile and dimple, and by overall higher levels of AU activation corresponding to positive emotions such as happiness and content.

TABLE I: Patients vs. controls classification results

Feature Type	*AUC*
Richness and *Cluster Distribution*	0.85
Typicality	0.84
All Features	0.80

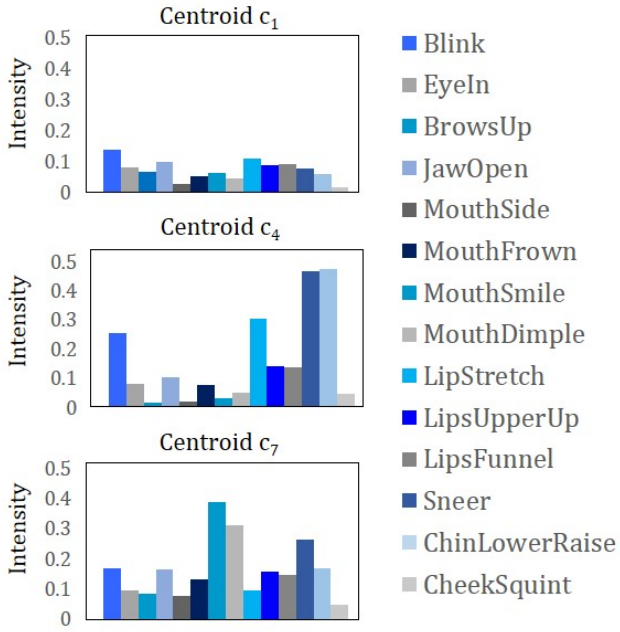

Fig. 2: The centroids of 3 facial-clusters returned by the K-means clustering algorithm (k=7)

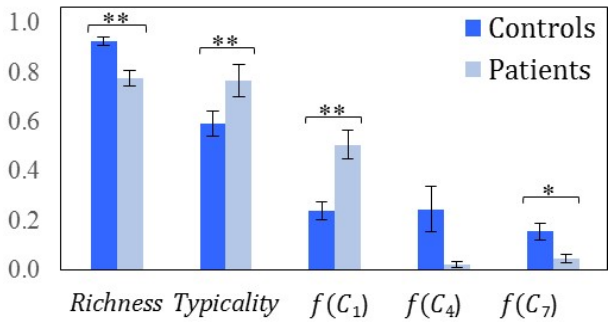

Fig. 3: Group difference for *Richness*, *Typicality*, and for the *Cluster Distribution* (f) of facial-clusters C_1, C_4 and C_7

B. Patients Vs. Control

Significant group differences were found in *Cluster Distribution* for facial-clusters C_1, C_4 and C_7. C_1 was significantly more frequent in patients in comparison with controls ($t = 4.14$, $p << 0.01$), while the frequency of C_4 and C_7 was reduced in patients ($t = 2.43$, $p = 0.018$, and $t = 2.84$, $p = 0.006$ respectively). The results for facial-cluster C_4 are not significant under the *Bonferroni correction*, and further investigation using a larger sample is needed to avoid type-I error. No significant difference was found for the remaining facial-clusters. *Richness* was significantly reduced in patients in comparison with controls ($t = 4.87$, $p << 0.01$), while *Typicality* was higher in patients ($t = -3.39$, $p << 0.01$). Results are summarized in Fig .3.

Learning results suggest that facial-cluster features are predictive for patients vs. control classification (Table I). A classifier (SVM) trained to discriminate between patients and controls, using as input *Richness* and *Cluster Distribution*, achieved the best results ($AUC = 0.85$). *Typicality* achieved

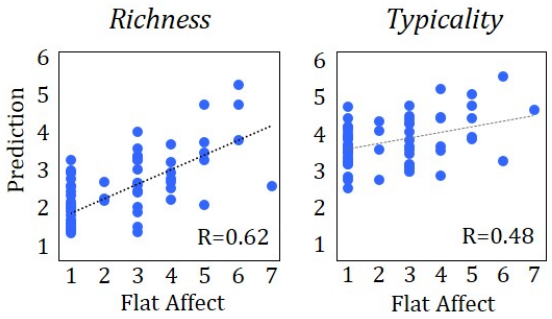

Fig. 4: Correlation between *Flat Affect* score and the prediction of the learning algorithm using the train data

the second best results ($AUC = 0.84$). Classification was not improved by letting the classifier use all the features, most likely due to the small sample limitation and subsequent over-fitting.

C. Correlation with Flat Affect

The evaluation of *Flat Affect* severity was at high agreement between raters ($R = 0.910$, $p << 0.01$), and was found to be significantly correlated with 3 negative symptoms, including *Emotional withdrawal* ($R = 0.907, p << 0$), *Lack of spontaneity and conversation flow* ($R = 0.818, p << 0.01$) and *Difficulty in abstract thinking* ($R = 0.764, p = 0.0014$).

Fig. 4 illustrates the correlation between *Flat Affect* score, and the prediction given by the algorithm based on different feature types. The most highly correlated feature was *Richness*, followed by *Typicality*. Correlation was also significant on test data, ruling out the possibility of mere over-fitting.

Note that the positive correlation is not between symptom severity and feature score, rather it is the correlation with the prediction of the algorithm when learning is *based* on the specific feature. Specifically, the average regression weights ($\bar{w}$) of *Richness* in the first regression ($W1$ in Fig. 1) are negative ($\bar{w} = -0.62$), while *Typicality* is given a positive weight ($w = 0.36$) as expected.

Train and test results are summarize in Table II.

TABLE II: *Pearson* correlation between *Flat Affect* score and algorithm prediction on train and test data

	R-train	p-value	R-test	p-value
All Features	0.647	5.82E-09	0.431	2.72E-04
Richness	0.618	4.18E-08	0.420	3.98E-04
Typicality	0.480	3.912E-05	0.354	0.003
Cluster Distribution	0.472	6.95E-05	0.172	0.163

D. 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 cluster-facial features. 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.

IV. CONCLUSIONS

Our results are in excellent agreement with clinical findings, and suggest that in clinical settings schizophrenia patients demonstrate a smaller range of expression, characterized mainly by reduced overall facial activity. In contrast to other studies [12], we found a reduction in both positive and negative emotional expressions. Another interesting finding is that *Typicality* is higher in patients. This may indicate that they don't have a different set of basic facial expressions, but rather that their expressivity is less diverse and more repetitive. Finally, we found that information embedded in facial activity is sensitive enough for symptom severity evaluation, and for automatic patient vs. control seperation; this may be one day beneficial for diagnosis, monitoring and treatment.

REFERENCES

[1] Murray Alpert, Stanley D Rosenberg, Enrique R Pouget, and Richard J Shaw. Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry research*, 97(2):107–118, 2000.

[2] Hillel Aviezer, Shlomo Bentin, Veronica Dudarev, and Ran R Hassin. The automaticity of emotional face-context integration. *Emotion (Washington, D.C.)*, 11(6):1406–14, December 2011.

[3] 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.

[4] Paul Ekman. Universal and cultural differences in facial expression of emotion. In *Nebraska symposium on motivation*, volume 19, pages 207–284. University of Nebraska Press Lincoln, 1972.

[5] 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.

[6] Irina Falkenberg, Mathias Bartels, and Barbara Wild. Keep smiling! *European archives of psychiatry and clinical neuroscience*, 258(4):245–253, 2008.

[7] Wallace V Friesen and Paul Ekman. Emfacs-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco*, 2:36, 1983.

[8] 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.

[9] Hatice Gunes and Maja Pantic. Automatic, Dimensional and Continuous Emotion Recognition. *International Journal of Synthetic Emotions*, 1(1):68–99, 2010.

[10] Stanley R Kay, Abraham Flszbein, and Lewis A Opfer. The positive and negative syndrome scale (panss) for schizophrenia. *Schizophrenia bulletin*, 13(2):261, 1987.

[11] Peter F Liddle. Schizophrenia: the clinical picture. In *Seminars in general adult psychiatry*, pages 167–186, 2007.

[12] 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.

[13] 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.

[14] 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.

[15] 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.