Features Derived From Behavioral Experiments To Distinguish Mental Healthy People From Depressed People

Wenyu Li, Huimin Ma, Xiang Wang, Dazhao Shi
Department of Electronic Engineering, Tsinghua University
Room 6-203, Rohm Electronic Engineering Building, No.1 Qinghua Yuan, Haidian district, Peking, China
mhmpub@tsinghua.edu.cn

ABSTRACT

How to distinguish mental healthy people from people suffered from depression objectively is an unsolved question in both pattern recognition and clinical psychology. Traditional diagnosis of depression was drawn by doctors according to the results of questionnaire analysis or scale tests, which were prone to be subjective. In this paper, we selected negative and positive scenes from IAPS as the background, positive and negative faces from Taiwanese Facial Expression Image Database as foreground. Based on the competing effect and the priming effect of background scenes, we designed two experiments by controlling the relative appearing orders of background scenes and foreground faces. By recording subjects' response times at key-strokes, we quantified the identification differences towards different combinations of background and foreground between depressed people and mental healthy people, and these quantified differences can be used as criteria to derive features distinguishing two kinds of people objectively.

KEY WORDS

Depression, psychology, behavioral experiment, pattern recognition, feature

1. Introduction

Nowadays, depression is becoming an Epidemic disease around the world. Traditional diagnosis of depression was drawn by doctors referring to the results of questionnaire analysis or scale tests, such as the Depression Anxiety Stress Scale [1], Cognitive Emotion Regulation Questionnaire [2] and Mini International Neuropsychiatric Interview [3]. Obviously, this kind of diagnosis might be influenced by experiences and status of doctors, which were prone to be subjective.

With the development of information technology and computer science, more and more tech means were applied to obtain people's physiological indicators for psychological studies, among which keyboard-based behavioral experiment was one of the most popular. In a keyboard-based behavioral experiment, participants are first showed some stimuli, such as simple words, faces and images [4-6], then they are asked to response to detectors [4,7] (mostly a dot) or identification objects [8-10] (mostly emotional faces) by pressing keyboard. And their reaction time which is regarded as a physiological indicator is recorded by computers.

In theory, using the reaction time as features, it is not difficult to design classifiers to distinguish different kinds of people. And meanwhile, the results of classification are more objectively. But unfortunately, to achieve this goal, it needs large amount of data which is hard to get in reality, especially for depressed people. To solve this problem, we must find another approach.

Statistical analysis is a good tool which is commonly used in behavioral experiments and widely accepted by authority in psychology. We could introduce it in our study, by what cognitive differences that might result from attentional bias [5,11,12] could be used as criteria to derive features for classification.

2. Behavioral Experiments Design

2.1 Participants

140 participants were recruited from certain university, and their age range was from 18 to 28. 135 of them were right-handed and asked to take part in two different clinical measures before experiments. First, they were interviewed by Mini International Neuropsychiatric Interview [3], after which the patient group was made up of 75 participants with a primary diagnosis of MDD (major depressive disorder). Then they completed the Beck Depression Inventory-II-NL [13], a reliable and valid self-report measure of intensity of depression. After that the MDD group was consisted of 58 participants scoring above 8, and the normal control (NC) group was consisted of 58 participants scoring below 4.

24 subjects from the MDD group (17 males and 7 females, mean age 21.8 years, S.D.=2.64) and 24 subjects from the NC group (15 males and 9 females, mean age 20.9 years, S.D.=2.72) participated in Experiment 1; 24 subjects from the MDD group (19 males and 5 females, mean age 21.2 years, S.D.=2.88) and 24 subjects from the NC group (13 males and 11 females, mean age 20.8 years, S.D.=2.66) participated in Experiment 2. None of them participated in two experiments. Their experimental data were first used to analyze the differences between the MDD group and the NC group in Section 3, and then used as training samples in Section 4.

The rest of 10 subjects from the MDD group (6 males and 4 females, mean age 21.7 years, S.D.=2.98) and 10 subjects from the NC group (5 males and 5 females, mean age 22.3 years, S.D.=2.75) participated in both experiments. Their experimental data were used to value the result of our feature description in Section 4.

2.2 Image selection

We used emotional images from IAPS [14] as the background scenes. Considering that all of participates were from certain university, we made an emotional attributes re-scoring for all the emotional images that were selected. According to the result of re-scoring, 50 positive and 50 negative images that showed clear discrimination were used as our experimental background.

Emotional faces working as the foreground were from Taiwanese Facial Expression Image Database. Meanwhile, in order to eliminate differences of race, gender and identity, as well as the affections of background and accessories, we performed the following steps:

- We chose 160 faces (80 positive and 80 negative) from the database.
- We balanced posture, took the major face area interception, did face alignment and changed the color image into gray degree for each face.
- We randomly picked 40 positive faces, equally divided them into 10 groups and merged each group into average face. We used the same way to get 10 average negative faces.
- We used PCA algorithm (the first 50 dimensions were chosen) to verify the effect of our average faces with the rest 80 faces as training samples and the 20 average faces as test samples.
- 16 average faces (8 positive and 8 negative) were correctly classified and worked as the foreground stimuli.

2.3 Procedure of experiment

According to the relationship between foreground and background, we designed two experiments. Experiment 1 was designed according to Kenichi Ito's study, which presented the foreground and the background simultaneously with the background as a distractor. Experiment 2 was referred to the affectively prim research with the background as an initiator. Meanwhile, to make sure that the information of images can be fully transferred to participants, we set the priming time to 1000ms-1500ms randomly. The key-pressing response times of identification for emotional faces were collected and analyzed to find the cognitive differences between mental healthy people and depressed people, by what we got the features to classify them.

The procedure of two experiments was almost the same with a few differences. Before started, participates were asked to read the on-screen prompts which informed them that: 1) they should focus on the center of the screen during the whole experiment; 2) each trial was composed by a background scene and a foreground emotional face (emotional faces are in the center of scenes); 3) when a face appeared, half of them should make positive and negative attributes of the face judgment as fast as possible without sacrificing accuracy by pressing left "Ctrl" for the positive face and right "Ctrl" for the negative face, while the other half should make the judgment on the contrary (All of the participates were right-handed); 4) the scene and face disappeared after pressing followed with an interval of 1.5s black background and then next trial arrived. Specific procedure of two experiments was shown in Fig .1& Fig .2. First, each participate was given the opportunity to practice 10 trials, then they would complete 80 formal experimental trials. Response time and accuracy for each trial were recorded.

Experiment 1 focused on the competing effect of the background. We presented the background and the foreground simultaneously.

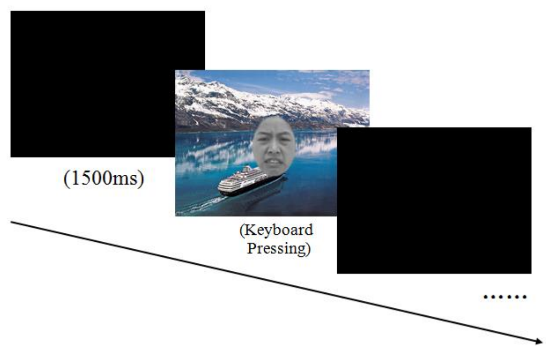

Fig .1 Specific procedure of Experiment 1

Experiment 2 focused on the priming effect of the background. The experimental procedure was similar to Experiment 1, except that the background was presented first following with the foreground after the background disappeared.

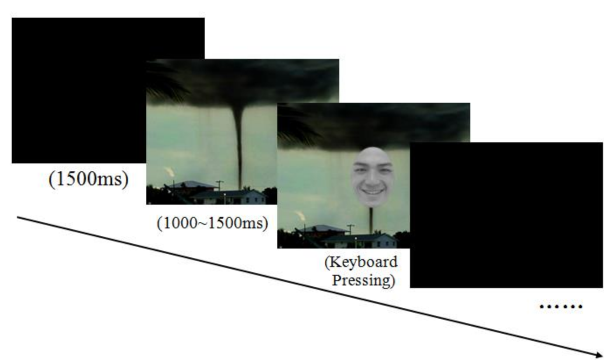

Fig .2 Specific procedure of Experiment 2

3. Significant Difference Analysis

We used repeated measures ANOVA to analyze the main effect and the interaction effect, independent samples t-test to analyze the differences between subjects, and repeated sample t-test to analyze the differences within subjects.

3.1 Difference under competing effect (CE)

3.1.1 Main effect and interaction effect under CE

We conducted 2 (Group: MDD vs. NC) × 2 (Background: positive scene VS. negative scene) × 2 (Foreground: positive face VS. negative face) repeated measures ANOVA, with participates' keyboard response time being the dependent variable (we also considered accuracy, but the accuracy of each participate was generally above 95%, so taking into account the ceiling effect, it was not used).

We found a significant main effect of group ($F=42.177$, $p<0.001$), which indicated the differences between subjects. A main effect of foreground or background was not significant, ps > 0.2.The two-way and three-way interactions were not significant, ps > 0.5.

3.1.2 Differences between subjects under CE

The following independent samples t-test showed that in four different combinations of foreground and background, the identification speed of the MDD group was smaller than that of NC group (Fig .3, $t_{ps\&pf}$ =3.290, p =.002, $t_{ps\&nf}$=3.047, p =.004, $t_{ns\&pf}$=3.064, p=.004, $t_{ns\&nf}$=3.570, p=.001, Table 1).

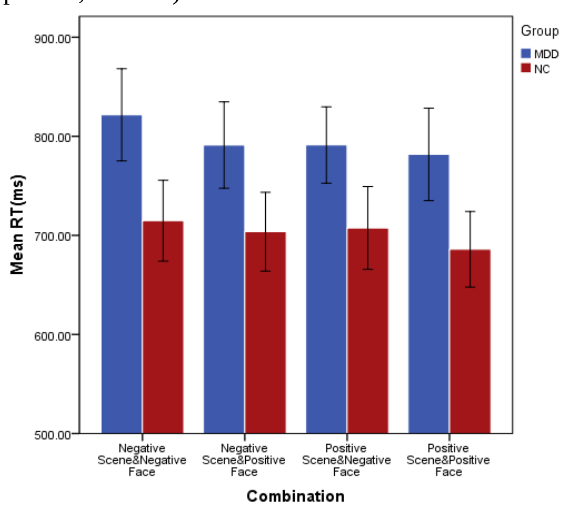

Fig .3 Differences between subjects under four combinations of Experiment 1

Table 1 Mean RTs and SDs of Experiment 1

	Group			
	MDD		NC	
	Mean RT	S.D.	Mean RT	S.D.
PS&PF	781.77	110.54	685.95	90.23
PS&NF	791.20	91.15	704.47	99.07
NS&PF	791.16	103.38	703.69	94.16
NS&NF	821.63	110.22	714.81	96.64

3.1.3 Differences within subjects under CE

Subsequent repeated samples t-test showed that under the influence of negative scenes, The MDD group identified positive faces faster than negative faces ($t=2.811$, $p=.010$), meanwhile, they identified negative faces under the influence of positive scenes faster than under the influence of negative scenes ($t=3.541$, $p=.002$). The NC group was just to the opposite. Their identification speed of negative faces under different positive-negative background scenes was not significant and their identification speed of different faces under the influence of negative scenes was not significant, either. Meanwhile, under the influence of positive scenes, they identified positive faces faster than negative faces ($t=3.206$, $p=.004$). And they identified positive faces under the influence of positive scenes faster than under the influence of negative scenes ($t=2.831$, $p=.009$).

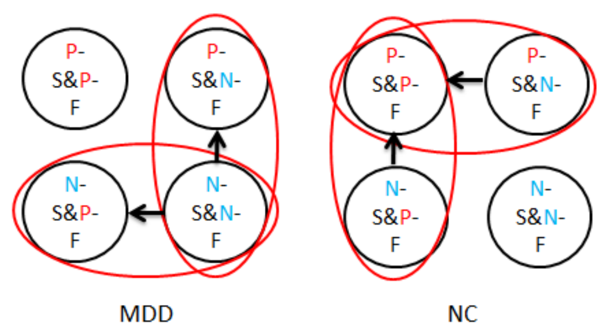

Fig .4 Existence of cognitive differences within subjects of Experiment 1

Then we made a drawing that showed the existence of cognitive differences within subjects (Fig .4), where "P-S" stood for positive scenes, "N-F" stood for negative faces and so on. The pairs of combinations highlighted by red circles indicated there were significant differences between them, and the one with an arrow pointing to was faster than the other.

3.2 Difference under priming effect (PE)

3.2.1 Main effect and interaction effect under PE

We also conducted 2 (Group: MDD vs. NC) × 2 (Background: positive scene VS. negative scene) × 2 (Foreground: positive face VS. negative face) repeated measures ANOVA, with participates' keyboard response time being the dependent variable.

We found a significant main effect of group (F=41.583, p<0.001), which indicated the differences between subjects. A main effect of foreground or background was not significant, ps> 0.7.The two-way and three-way interactions were not significant, ps > 0.4.

3.2.2 Differences between subjects under PE

The following independent samples t-test showed that in four different combinations of the foreground and the background, the identification speed of the MDD group was slower than that of NC group (Fig .5, $t_{ps\&pf}$=4.053, p<.001, $t_{ps\&nf}$=2.856, p=.006, $t_{ns\&pf}$=3.588, p=.001, $t_{ns\&nf}$=2.523, p=.015, Table 2).

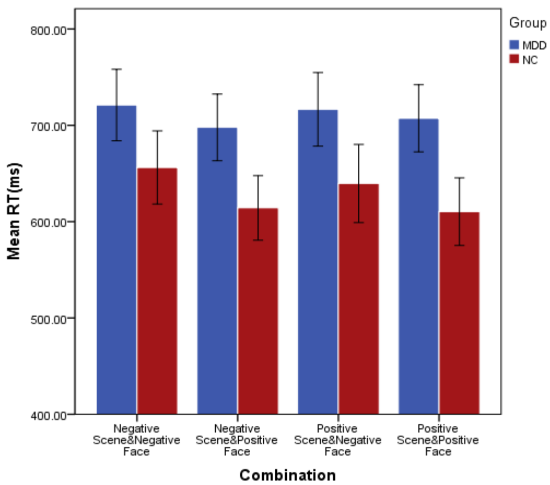

Fig .5 Differences between subjects under four combinations of Experiment 2

Table 2 Mean RTs and SDs of Experiment 2

	Group			
	MDD		NC	
	Mean RT	S.D.	Mean RT	S.D.
PS&PF	707.30	82.53	610.36	83.16
PS&NF	716.58	90.52	639.63	96.08
NS&PF	697.85	81.88	614.26	79.50
NS&NF	721.07	87.96	656.24	90.04

3.2.3 Differences within subjects under PE

Subsequent repeated samples t-test showed that the MDD group had no significant differences between any pair of four combinations. Meanwhile, the NC group identified positive faces faster than negative faces under both different positive-negative background scenes (t=3.096, p=.005 for positive background and t=3.357, p=.003 for negative background). Besides, when identifying negative faces, they responded faster under the influence of positive scenes than under the influence of negative scenes (t=2.753, p=.011).

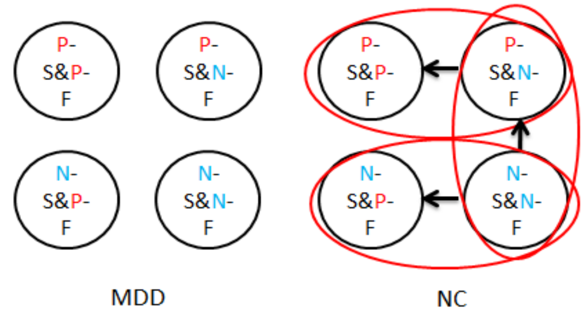

Fig .6 Existence of cognitive differences within subjects of Experiment 2

We also made a drawing that showed the existence of cognitive differences within subjects (Fig .6), where “P-S” stood for positive scenes, “N-F” stood for negative faces and so on. The pairs of combinations highlighted by red circles indicated there were significant differences between them, and the one with an arrow pointing to was faster than the other.

4. Results of Classification

According to the analysis, we could see that the differences between subjects were more stable and easy to describe, so we defined between-subject features of any participant as feature vector below:

$$F = [t_{11}, t_{12}, t_{13}, t_{14}, t_{21}, t_{22}, t_{23}, t_{24}]$$

In the feature vector, “t_{11}” stood for the mean response time of the participant under the combination of positive background and positive foreground in Experiment 1, and “t_{23}” stood for the mean response time of the participant under the combination of negative background and positive foreground in Experiment 2. The specific meanings of features were shown in Table 3.

Table 3 Specific meanings of between-subjects features (BSF)

BSF	t_{11}	t_{12}	t_{13}	t_{14}	t_{21}	t_{22}	t_{23}	t_{24}
Exp	1	1	1	1	2	2	2	2
Face	pos	neg	pos	neg	pos	neg	pos	neg
Scene	pos	pos	neg	neg	pos	pos	neg	neg

Then we got the training samples which were partly listed in Table 4. In the table, subjects numbered 1-24 did not participate in Experiment 2, so their last 4 features were marked by “N/A”; subjects numbered 25-48 did not participate in Experiment 1, so their first 4 features were marked by “N/A”.

Table 4 Between-subjects features of training samples

Group	No.	Feature Vector							
		t_{11}	t_{12}	t_{13}	t_{14}	t_{21}	t_{22}	t_{23}	t_{24}
MDD	1	804.64	789.46	822.63	923.50	N/A	N/A	N/A	N/A
	2	728.25	742.00	648.00	723.26	N/A	N/A	N/A	N/A
	……								
	24	914.10	916.41	886.92	965.70	N/A	N/A	N/A	N/A
	25	N/A	N/A	N/A	N/A	862.18	916.07	867.70	860.78
	26	N/A	N/A	N/A	N/A	759.00	753.21	712.00	761.52
	……								
	48	N/A	N/A	N/A	N/A	804.00	786.27	679.57	772.52
NC	1	695.47	753.92	733.22	774.80	N/A	N/A	N/A	N/A
	2	775.15	819.00	773.29	754.00	N/A	N/A	N/A	N/A
	……								
	24	588.94	566.31	568.78	557.90	N/A	N/A	N/A	N/A
	25	N/A	N/A	N/A	N/A	641.00	709.08	672.81	633.82
	26	N/A	N/A	N/A	N/A	616.82	619.72	577.17	617.11
	……								
	48	N/A	N/A	N/A	N/A	574.47	578.55	568.16	623.12

Using the training samples, we could design classifiers. First, we designed two 4-dimension Gauss classifiers (GC). GC1 was based on the training samples numbered 1-24. GC2 was based on based on the training samples numbered 25-48. The result of classification was shown in Table 5.

Table 5 Result of classification by GC

	GC1		GC2	
	correct	wrong	correct	wrong
MDD	8	2	7	3
NC	9	1	8	2
Accuracy	85%		75%	

Then, we designed two 4-dimension linear SVM classifiers (LSC). LSC1 was based on the training samples numbered 1-24. LSC2 was based on based on the training samples numbered 25-48. The result of classification was shown in Table 6.

Table 6 Result of classification by LSC

	LSC1		LSC2	
	correct	wrong	correct	wrong
MDD	8	2	6	4
NC	8	2	9	1
Accuracy	80%		75%	

As we can see, the result was remarkable. Thus, the way we derived features was acceptable.

5. Conclusion

Through the two experiments, we combined pattern recognition with psychology research. By analyzing the keyboard response time of identification for emotional faces, we found the cognitive differences between mental healthy people and depressed people, and then we used these quantified differences as criteria to derive features to distinguish two kinds of people objectively.

The most advantage of the way we derived the features is that it overcomes the difficulty of lacking of data. Together with the support of psychological theories, it could make the features more robust. We got a remarkable result here, but we still need more data to verify its effectiveness. As the within-subjects features were different in two experiments, and they were not easy to describe, we used the between-subjects features only. Maybe the within-subjects features could also make contribution to the result of classification. Meanwhile, the sample expansion and the introduction of new paradigms could make the features more plenty and stable. All of above are new directions to improve our study.

Our research prompts the relation between pattern recognition and psychology. And it introduces a new way to distinguish mental healthy people from depressed people objectively, which works as a reference to traditional diagnosis and is potential to be a new approach to prediction of depression in future.

Acknowledgement

This research is sponsored by the National Natural Science Foundation of China (NSFC61171113).

References

[1] S.H. Lovibond & P.F. Lovibond. *Manual for the depression anxiety stress scales.* (Psychology Foundation of Australia, 1996).
[2] N. Garnefski, V. Kraaij &P. Spinhoven. Negative life events, cognitive emotion regulation and emotional problems. *Personality and Individual differences, 30*(8), 2001, 1311-1327.
[3] N.R. Pinninti, H. Madison, E. Musser, et al. MINI International Neuropsychiatric Schedule: clinical utility

and patient acceptance. *European Psychiatry*, *18*(7), 2003, 361-364.
[4] H.T. Neshat-Doost, A.R. Moradi, M.R. Taghavi, et al. Lack of attentional bias for emotional information in clinically depressed children and adolescents on the dot probe task. *Journal of Child Psychology and Psychiatry, 41*(3), 2001, 363-368.
[5] L. Leyman, R. De Raedt, R. Schacht, et al. Attentional biases for angry faces in unipolar depression. *Psychological Medicine, 37*(3), 2007, 393-402.
[6] R. Righart & B. De Gelder. Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience, 8*(3), 2008, 264-272.
[7] B.P. Bradley, K. Mogg, J. White, et al. Attentional bias for emotional faces in generalized anxiety disorder. *British Journal of Clinical Psychology, 38*(3), 1999, 267-278.
[8] M.J. Fenske & J.D. Eastwood. Modulation of focused attention by faces expressing emotion: evidence from flanker tasks. *Emotion, 3*(4), 2003, 327.
[9] J.D. Eastwood, D. Smilek & P.M. Merikle. Negative facial expression captures attention and disrupts performance. *Perception & Psychophysics, 65*(3), 2003, 352-358.
[10] K. Ito, T. Masuda & K. Hioki. Affective Information in Context and Judgment of Facial Expression Cultural Similarities and Variations in Context Effects Between North Americans and East Asians. *Journal of Cross-Cultural Psychology*, *43*(3), 2012, 429-445.
[11] K. Mogg & B.P. Bradley. Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive therapy and research, 29*(1), 2005, 29-45.
[12] I.H. Gotlib, E. Krasnoperova, D.N. Yue, et al. Attentional biases for negative interpersonal stimuli in clinical depression. *Journal of abnormal psychology, 113*(1), 2004, 127.
[13] A.T. Beck, C.H. Ward, M. Mendelson, et al. An inventory for measuring depression. *Archives of general psychiatry*, 4(6), 1961, 561.
[14] P.J. Lang, M.M. Bradley & B.N. Cuthbert. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (NIMH, Center for the Study of Emotion & Attention, 2005)