Development of a Chinese Depressed Speech Corpus Based on The Disturbed Effect of Self-Processing

Xiaoyong Lu[1,4,6], *Yanqin Li*[1,4], *Haizhen An*[1,4], *Tao Pan*[3], *Renjun Li*[1,4], *Yanbin Hu*[2,6], *Aibao Zhou*[2,6], *and Hongwu Yang*[1,4,5]

[1] School of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, China

[2] School of Psychology, Northwest Normal University, Lanzhou, China

[3] Lanzhou Resources and Environment Voc-Tech College, Lanzhou, China

[4] Engineering Research Center of Gansu Province for Intelligent Information Technology and Application, Lanzhou, China

[5] National and Provincial Joint Engineering Laboratory of Learning Analysis Technology in Online Education, Lanzhou, China

[6] Key Laboratory of behavioral and Mental Health, Lanzhou, China

E-mail: luxy@nwnu.edu.cn

***Abstract*— Depression has long been recognized as one of the leading causes of disability and burden worldwide. In psychology, it is well known that the self is not only the cognitive subject, but also the core of personality. And the high incidence of suicide and pervasive hopelessness in depressed individuals suggested that the self might be abnormal among them. In order to expand the application of the depression detection, we employ classical scientific psychology paradigms on abnormalities of self-related processing in depressed individuals to develop a Chinese depressed speech corpus. Eleven depressed individuals and ten healthy subjects, who are gender-balanced and age-balanced, were recruited to participate in this study. Currently we have preliminarily collected 6 and 2.5 hours of speech data respectively, with the results of preliminary analysis indicating that there exist abnormalities in the depressed speech. The study results will provide a new perspective and strategy for further study on the building and application of Chinese speech corpus in depression.**

I. INTRODUCTION

With the rapid development of Chinese social economy, the pace of life is accelerated gradually. So with the mental pressure increasing, there are more and more depressed patients in China. There is evidence for depressed individuals and families bear the high socio-economic costs, and depression finally places an individual at higher risk of engaging in suicidal behaviours 0. Current diagnostic methods still rely on interview style assessments [2]. However, lacking an objective predictor of depression, these tests are subjective and partial in nature with a low accuracy. Therefore, to enhance current diagnostic methods, an objective screening mechanism, based on physiological and behavioral signals, is needed.

In recent years, research into the automatic and objective classification of depression, using various ways has gained popularity. Learning from previous relevant research and preliminary experimental results, a range of acoustic features have already been identified for suitability in the classification of depression [3] [4] [5]. Therefore, how to build speech database is an important question and is one of hotspot problems in diagnostic method using speech analysis. Currently, compared to the field of depressed speech abroad, to the best of our knowledge, there have been no publicly available depression speech corpus yet in China and corpus data for Chinese depressive patients are seriously lacking.

Directing toward the design of depressed speech corpus, among various methods to realize this, this paper focuses on the influence of psychological activities of depressed individuals on speech production. American psychologist William James [6] believed the self is not only the cognitive subject, but also the core of personality; and pervades virtually every aspect of mental activity. Klein et al. [7] based on data from neuropsychology, divided complex and unified self-system into six components, and suggested the patients with brain injury may lack one or two of the above-mentioned components. According to research approaches to brain imaging on self-knowledge and self-conscious at different depths, Chinese psychologist Zhu Ying [8] separated it to three parts. In addition, experiments were made and the results showed that depressed individuals might take on level of self-abnormalities in comparison to healthy participants by Jia Xiaohong et al. [9]. Drawing on these psychology findings, from psychological perspective, we adopt classical scientific psychology paradigms on abnormalities of self-processing in depressed individuals to build the corpus.

Our paper is organized as follows. In the next section, we will briefly give an overview of currently available depressed speech corpora. In section 3, we will focus on the disturbed effect of self-processing in depressed individuals. In section 4, we will discuss the design and the collection process of the speech corpus we have developed. Some preliminary analysis are also conducted in section 4 to show the results of our approach. Finally, we will bring out our conclusion and perspectives.

II. CURRENT AVAILABLE DEPRESSED SPEECH CORPORA

As we all know, any research on speech is inseparable from the support of speech corpus. The quality and completeness of the collected speech data directly determines the performance of depression recognition models. In this paper, we give a brief overview of the representative three depression corpora abroad. Although these corpora may not cover most of speech resources, they all contain the high quality speech data and have wide effect in this filed.

This first is the Audio/Visual Emotion Challenge and workshop (AVEC) 2013 dataset [10]. All speakers in AEVC-2013 are native German speakers. This corpus consists of 150 recordings, divided into training, development and testing parts, each of 50 recordings with a mean file length of 14 min 52 sec, collected using a headset at a sampling rate of 44.1 kHz, with 16 bit encoding.

Another one was originally collected by Mundt et al. [11], which consisted of speech recordings of 35 depressive subjects under treatment over a 6-week assessment period. Speech recordings (sampled at 8 kHz) were collected at weeks 0,2,4, and 6 during assessment process, in which all the subjects speaking American English. The forms of data acquisition include responses to three questions on the patient's emotional and physical state and a reading passage used by speech pathologists to test fluency.

In addition to the above corpora, there are other publicly available depressed corpus, such as the Black Dog [12], Scherer et al. [13], Low et al. [14], and Alghowinem et al. [15].

III. THE DISTURBED EFFECT OF SELF-PROCESSING IN DEPRESSED INDIVIDUAL

In psychology, some papers illustrate the abnormality of self-related processing among depression can be summarized in three crucial characteristics [16], which can be conceptualized as increased self-focus, association of the self with negative emotions, and increased cognitive processing of one's own self, as shown in Figure 1.

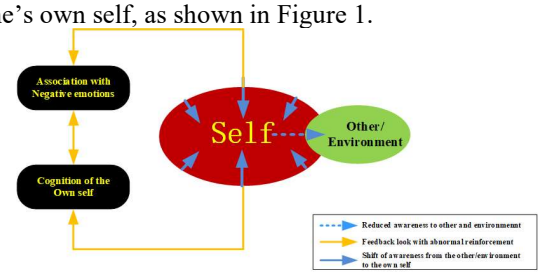

Fig. 1 Changes of self, others and body in depression individual

Specifically, the first characteristic of depression is taken by almost all depressed patients looking inward more than outward; that is, they are self-centered and have difficulty shifting attention to others. Another characteristic of depression is the tendency to attribute negative emotions to one's self. The self is associated with sadness, guilt, mistakes, inabilities, death, illness, and so on, which may ultimately result in paranoid delusions. Finally, people with depression typically suffer from increased cognitive processing of their own self; they think about themselves and their mood and desperately try to discover the reasons for their depression, but ending up with being stuck in the depressed mood, tighter and tighter.

We therefore introduce some novel methods such as different psychological self and psychological experimental paradigms: (1) different time dimension in self; (2) different evaluation perspective in self; (3) directed forgetting and self-reference. All stimulus programs are designed by E-prime 2.0.

IV. DESIGN OF DEPRESSED SPEECH CORPUS

This paper [17] just raised a design program, more work is need to perfect it. So the block diagram of the approach we proposed is shown in Figure 2, in which the three critical points that address the currently existing challenges are introduced.

- Matched-subject design between two groups.
- Task design based on the disturbed effect of self-processing.
- Corresponding speech data collection process.

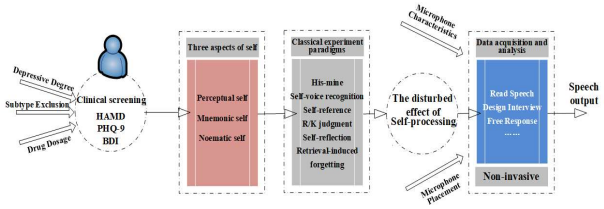

Fig. 2 Block diagram of the proposed approach.

A. Subjects

Our target corpus size is 100 depressed patients and 100 control healthy subjects. In order to ensure a certain degree of accuracy and reliability of the comparison with experimental results, we induct a psychological matched-subject design in the process of subjects selection. It is an important concept for studies in psychology, which aims for equating groups on some variables to reduce their effect on skewing the results. So we balanced gender, age and degree of education respectively. The distribution of the depressed patients and controls recorded to date is shown in Table I. All subjects who speak Mandarin completed all tasks.

Table I: The distribution of subjects.

Subjects	Gender	Age	Education	Total
Depression	Male	25-50	High school or above	6
	Female	29-50		5
Control	Male	25-50		5
	Female	29-50		5

B. Task Design

The stimuli we design are structured to include a range of speaking styles: single word, short sentence, paragraph and

conversation. The four tasks, which are formed by ten subtasks , performed by our all subjects are as follows:

- Task 1: Word-level stimuli. Task 1 adopted 562 self-personality adjectives formulated by Huang Xiting et al. [18]. Ten words of them are as shown in Table II.
- Task 2: Sentence-level stimuli. The prompts of Task 2 consist of short sentence. A set of 60 short sentences, with each category under self/others × positive/negative/neutral containing ten sentences, are designed according to different paradigms of self-related processing. Partial examples are shown in Table III.
- Task 3: Paragraph-level stimuli. The prompts of Task 3 are designed for a phonetically rich passage using the psychological autobiographical memory. The passage contains 160 Chinese characters.
- Task 4: Conversation stimuli.

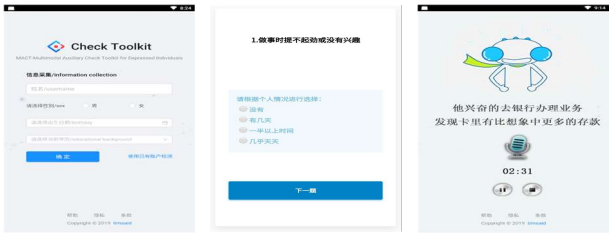

Fig. 3 Some interfaces of Android app.

Table II: Examples of word-level stimuli.

Words	Translation	Words	Translation
忠厚的	guileless	空虚的	void
朴实的	plain	脆弱的	vulnerable
机敏的	astute	厌倦的	bored
谦逊的	modest	憔悴的	haggard
英俊的	handsome	烦躁的	agitated

Table III: Examples of sentence-level stimuli.

Paragraph
我下班后，轻松地回到家，坐在沙发上有种温馨的感觉。 After work, care-free, I got back home, sitting in sofa, feeling cozy.
我下班回到家，坐在沙发上有种凉快的感觉。 After work, I got back home, sitting in sofa, feeling pleasantly cool.
我下班后，疲惫地回到家。坐在沙发上有种凄凉的感觉。 After work, exhausted, I got back home, feeling miserable.

According to doctors' clinical experience and suggestion, the average duration in subtask does not exceed twelve minutes. The total duration for all tasks is about 110 minutes.

C. *Data Collection*

Part of recording takes place in a recording studio, which is a special recording place built for us to fulfill specific acoustic requirements of recording environment. The other part takes place in a quiet meeting room of hospital department. In this environment, we use Roland R-26 portable recorder to capture real-time speech. The stereo audio is sampled at 44.1 kHz and quantized at 16 bits. Figure 3 illustrates the whole setup. In addition, we also have designed and developed an Android app to collect additional speech data sources. Figure 4 shows some interfaces of APP.

Fig. 4 The recording setup (Right). The subject sits in a professional recording studio.

A monitor is placed in front of the subject. During this whole process, we adhere to ethical principles such as patients' voluntariness, patient's inform consent, and confidentiality and patients' protection etc. So when all the conditions are met, we will also put a digital camera, Aoni A30, on the top of monitor.

For Task 1, subjects read aloud the prompts displayed randomly on the screen. one prompt is displayed each time. For Task 2 and Task 3, the font, the font size, and the lasting time of the experiment assignments are controlled by E-prime program. For Task 4, it takes the experimenter to induce the subjects to complete tasks by cooperation. In some specific experiments, subjects also need to put required content all down on paper.

Considering this effect and based on real situation, we allow depressive subjects to select only two subtasks to complete, preventing the duration of task from exceeding thirty minutes. Doing so can reduce subject fatigue and avoid the conflict, tired mood.

V. PRELIMINARY ANALYSIS

Previous works [4] [5] have found the acoustic abnormal characteristics of depressed speech. The sensitive parameters include the speaking rate, energy, and the pitch. We have studied the acoustic characteristics of our collected data to see whether the characteristics match the prior findings in the literatures.

A. *Speaking Rate*

We measure the speaking rate according to the audio files and calculate the average and maximum syllables per second. The results are shown in Figure 5. In Figure 5, compared with control subjects (illustrated with white bubbles with letters B and D), the depressed subjects (illustrated with white bubbles with letters A and C) utter lower number average syllables per second and the maximum syllables per second or their average syllables per second is significantly different from their maximum syllables per second. However, the range of average speaking rate and maximum speaking rate of different control subjects are relatively consistent and stable. In contrast, the range of average speaking rate and maximum speaking rate of

depressed subjects is more obvious and unstable. The size of the bubbles represents the difference between maximum and average syllables per second. A large bubble size indicates the greater increase of the subject's speaking rate. There were significant differences among subjects with different depressive tendencies.

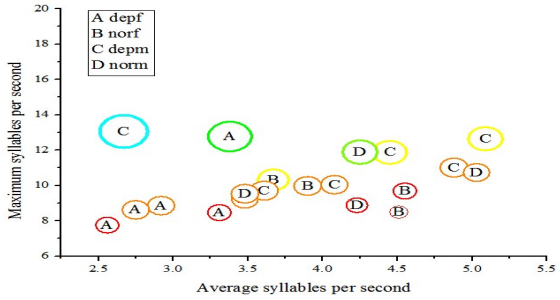

Fig. 5 The speaking rate of each subject is represented by a bubble.

B. Energy

In addition, we measure the intensity of speech produced by both the depressed subjects and the control subjects. The speech was produced by the subjects reading the corpus we built according to their usual speaking rate. Take the intensity value of 10 points for each syllable, and the intensity is extracted according to the maximum intensity of each syllable. Figure 6 shows the distribution of speech intensity in the depression group and the control group. We observed that the intensity of the control group changed within a small range, and the intensity of the control group was relatively concentrated, with a certain degree of stability. On the contrary, the intensity range of the depression tendency group was larger, and the intensity value was more discrete and unstable. The mean and median of intensities of the depressed subjects were also significantly lower than those of the control group. Most of the depressed subjects had a lower intensity than the control group.

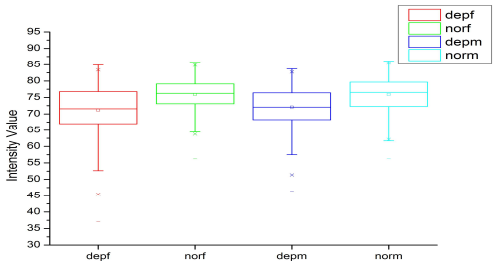

Fig. 6 The distribution of energy of the subjects.

C. Fundamental Frequency

In addition, we also count the fundamental frequency of the subjects. Take 10 fundamental frequencies out of each syllable. The fundamental frequency is extracted according to the maximum fundamental frequency of each syllable. Figure 7 shows the fundamental frequency distribution of both depression group and control group. We found that the variation range of the fundamental frequency in the control group was small and the fundamental frequency values were concentrated. Compared with the control group, the fundamental frequency distribution range of depressed subjects is relatively discrete and not stable enough.

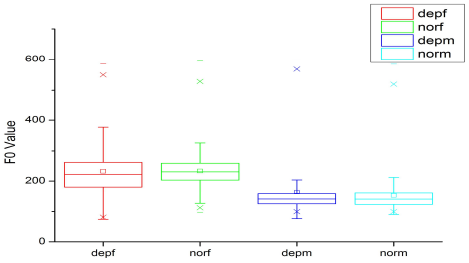

Fig. 7 The distribution of F0 change.

VI. Limitations and Future Work

In this work, it is next step to detect depression in a binary classification manner (i.e. depressed vs healthy controls) from the dataset. However, such a classification model also needs a large speech corpus with depression. It is found from the experimental results that we have not so far collected enough speech data. Furthermore, the current experiment only focus on preliminary analysis.

To the best of the authors' knowledge, even though it is a small study in academic research field, the work undertake by us is the first to investigate using classical scientific psychology paradigms on abnormalities of self-related processing in depressed individuals to develop a Chinese depressed speech corpus. We believe it has the significance of practical application and will provide a new strategy for enriching the diagnosis of depression. Meanwhile, we will continue to collect more depressed speech data to support our research. And we will further test the validity of method for building speech corpus based on the disturbed effect of self-processing.

However, due to time limitation, the detailed information can not be presented in this paper, but probably shown in the near future.

Acknowledgment

This research is based upon work supported by the National Science Foundation of China (NSFC) under grant No. 31860285 and No. 31660281. Additionally, part of this work is performed in the Scientific Research Project in Higher Education Institutions of Gansu Province (Grant No. 2017A-165). The authors also thank Dr. Dong Qiangli from Lanzhou University Second Hospital for his contribution in data collection process.

References

[1] Porter, R. J., et al. "Neurocognitive impairment in drug-free patients with major depressive disorder. " British Journal of Psychiatry the Journal of Mental Science 182.5(2018):214.
[2] Leentjens, A. F., et al. "The validity of the Beck Depression Inventory as a screening and diagnostic instrument for depression in patients with Parkinson's disease. " Movement Disorders Official Journal of the Movement Disorder Society 15.6(2015):1221-1224.
[3] Morales, Michelle Renee, and R. Levitan. "Speech vs. text: A comparative analysis of features for depression detection systems." Spoken Language Technology Workshop 2017.
[4] Tan, Tze Ern Shannon, J. A. Dai, and S. S. Lan. "Speech analysis and depression." Signal & Information Processing Association Summit & Conference 2017.
[5] Nicholas Cummins, Stefan Scherer, Jarek Krajewski, Sebastian Schnieder, Julien Epps, Thomas F. Quatieri. "A review of depression and suicide risk assessment using speech analysis." Speech Communication 71(2015):10-49.
[6] Myers, Gerald E. The Divided Self of William James. The divided self of William James. 1999.
[7] Klein, Stanley B., K. Rozendal, and L. Cosmides. "A Social-Cognitive Neuroscience Analysis Of The Self." Social Cognition 20.2(2002):105-135.
[8] Ying, Zhu. "Neuroimaging studies of self-reflection." Progress in Natural Science Materials International 14.4(2004):296-302.
[9] Shanshan, et al. "Neuropsychological impairment: the disturbed effect of self-processing in patients with major depressive disorder." Chinese Science Bulletin 59.28(2014):3595.
[10] Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S.,Schnieder, S., Cowie, R., Pantic, M., 2013. AVEC 2013: The continuous audio/visual emotion and depression recognition challenge.In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge. Barcelona, Spain, pp. 3–10.
[11] Mundt, J.C., Snyder, P.J., Cannizzaro, M.S., Chappie, K., Geralts, D.S., 2007. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J. Neuroling. 20, 50–64.
[12] Mcintyre, Gordon , et al. "An approach for automatically measuring facial activity in depressed subjects." International Conference on Affective Computing & Intelligent Interaction & Workshops IEEE, 2009.
[13] Scherer, S., Stratou, G., Gratch, J., Morency, L., 2013b. investigating voice quality as a speaker-independent indicator of depression and PTSD. Proceedings of Interspeech. ISCA, Lyon, France, pp. 847–851.
[14] Low, L.S.A., Maddage, N.C., Lech, M., Allen, N., 2009. Mel frequency cepstral feature and Gaussian Mixtures for modeling clinical depression in adolescents. In: 8th IEEE International Conference on Cognitive Informatics, 2009 (ICCI ’09), pp. 346–350.
[15] Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., Parker, G., 2012. From joyous to clinically depressed: mood detection using spontaneous speech. In: Twenty-Fifth International FLAIRS Conference. Marco Island, Florida, pp. 141–146.
[16] G.Northoff, “How is our self altered in psychiatric disorders? A neurophenomenal approach to psychopathological symptoms,” Psychopathology, vol. 47, 2014, pp.365-376.
[17] Lu, X., H. An, Y. Li, R. Li and C. Pan, "Design of Chinese Depression Spoken Speech Corpus Based on Psychological Self-related Processing," 2018 International Conference on Orange Technologies (ICOT), Nusa Dua, BALI, Indonesia, 2018, pp. 1-5.
[18] Huang, X., S. Zhang, "Desirability, meaningfulness and familiarity ratings of 562 personlity-trit djectives" Psychological Science,5(1992):17-22.