Journal of Affective Disorders 225 (2018) 214–220

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Major depressive disorder discrimination using vocal acoustic features

Takaya Taguchi[a,b], Hirokazu Tachikawa[a,c,*], Kiyotaka Nemoto[a,c], Masayuki Suzuki[d], Toru Nagano[d], Ryuki Tachibana[d], Masafumi Nishimura[e], Tetsuaki Arai[a,c]

[a] Department of Psychiatry, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Japan
[b] University of Tsukuba Hospital, Japan
[c] Department of Psychiatry, Faculty of Medicine, University of Tsukuba, Japan
[d] IBM Japan, LTD., IBM Research, Tokyo, Japan
[e] Graduate School of Integrated Science and Technology, Shizuoka University, Japan

ARTICLE INFO

Keywords:
Voice analysis
Acoustic features
Depression
Biomarker
MFCC
Discriminant analysis

ABSTRACT

Background: The voice carries various information produced by vibrations of the vocal cords and the vocal tract. Though many studies have reported a relationship between vocal acoustic features and depression, including mel-frequency cepstrum coefficients (MFCCs) which applied to speech recognition, there have been few studies in which acoustic features allowed discrimination of patients with depressive disorder. Vocal acoustic features as biomarker of depression could make differential diagnosis of patients with depressive state. In order to achieve differential diagnosis of depression, in this preliminary study, we examined whether vocal acoustic features could allow discrimination between depressive patients and healthy controls.
Methods: Subjects were 36 patients who met the criteria for major depressive disorder and 36 healthy controls with no current or past psychiatric disorders. Voices of reading out digits before and after verbal fluency task were recorded. Voices were analyzed using OpenSMILE. The extracted acoustic features, including MFCCs, were used for group comparison and discriminant analysis between patients and controls.
Results: The second dimension of MFCC (MFCC 2) was significantly different between groups and allowed the discrimination between patients and controls with a sensitivity of 77.8% and a specificity of 86.1%. The difference in MFCC 2 between the two groups reflected an energy difference of frequency around 2000–3000 Hz.
Conclusions: The MFCC 2 was significantly different between depressive patients and controls. This feature could be a useful biomarker to detect major depressive disorder.
Limitations: Sample size was relatively small. Psychotropics could have a confounding effect on voice.

1. Introduction

The voice carries various information, including that beyond the verbal message, important for diagnosis or state evaluation of mental disorders. It is known that psychiatric symptoms are diagnosed not only from a patient's spoken communication, but also from emotions and non-verbal information such as intention, attitude, or physical state (Ladd, 1980). Clinical psychiatrists or psychologists have experiences having difficulties in understanding what patients with depressions say because of their muffled speech. The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) illustrates psychomotor retardation as “speech that is decreased in volume, inflection, amount, or variety of content, or muteness” and “it must be severe enough to be observable by others”. However, there are no objective criteria about these speeches of psychomotor retardation.

Acoustic features which assess voices include directly-relevant features how voices are heard directly (e.g., volume, speaking duration, pause duration, musical pitch (or fundamental frequency), and formants) and indirectly-relevant features (e.g., zero crossing rate, harmonics to noise ratio, and mel-frequency cepstrum coefficients). Zero crossing rate is the rate of sign-changes along a voice signal. Harmonics to noise ratio quantifies the amount of additive noise in the voice signal. Mel-frequency cepstrum coefficients (MFCCs) were introduced by Mermelstein in the 1970's (Davis and Mermelstein, 1980), which have been shown to reflect vocal tract changes (Yinghua Zhu et al., 2013) and have been widely used in speech recognition. The method for extracting MFCCs is as follows: 1) calculate Fast Fourier Transform (FFT) spectrum from frequency, 2) extract filterbank output allocated on a mel scale based upon human aural characteristics, and 3) get cepstrum coefficient from Discrete Cosine Transform (DCT). Mitrović et al. showed that the lowest MFCC (MFCC0) represents the average power of the spectrum, and the second MFCC (MFCC 1) approximates the broad

* Correspondence to: Department of Psychiatry, Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tennoudai Tsukuba, Ibaraki 305-8575, Japan.
E-mail address: tachikawa@md.tsukuba.ac.jp (H. Tachikawa).

http://dx.doi.org/10.1016/j.jad.2017.08.038
Received 15 January 2017; Received in revised form 1 June 2017; Accepted 14 August 2017
Available online 16 August 2017

shape of the spectrum. The higher-order coefficients, including MFCC 2, represent finer spectral details (Mitrović et al., 2010). Therefore, lower-order coefficients are hard to be affected by the voice pitch and coefficients except MFCC 0 can omit influence of volume.

Various studies have reported a relationship between parameters derived from the voice and depression (Tolkmitt and Scherer, 1986; Wittels et al., 2002). For example, Nilsonne and colleagues reported that the fundamental frequencies of the voices of patients with depression decreased and pauses between the interviewer's questions and the patients' answers lengthened (Nilsonne et al., 1988). Another study reported a correlation between the Hamilton Depression Rating Scale and speaking rate, pitch variability, and percent of pause time (Cannizzaro et al., 2004). Several studies have explored the possibility of utilizing various information gleaned from the voice as therapeutic markers. Mundt and colleagues reported that response to depression treatment was related to pitch variability, pauses while speaking, and speed of speaking (Mundt et al., 2007). They also reported a relationship between response to treatment and vocalization time, number of pauses, speaking rate, fundamental frequency, and formants in major depressive disorder patients (Mundt et al., 2012). Alpert et al. focused on prosody and reported that patients with depression spoke with reduced prosody compared to normal subjects and that treatment affected prosody (Alpert et al., 2001). However, directly-relevant acoustic features including vocalization time, number of pauses, and speaking rate make recording time long, and formants, fundamental frequency, and prosody are subject to be affected by the individual characteristics and hard to detect the differences between individuals. Moreover, these acoustic features only reflect how voice is heard.

In addition to these acoustic features, a recent development in voice analysis has enabled the use of a wider variety of indirectly-relevant acoustic features including MFCCs and zero-crossing rate. In the engineering field, the Interspeech 2009 Emotion Challenge (Schuller et al., 2009) and the Interspeech 2010 Paralinguistic Challenge (Schuller et al., 2010) revealed a relationship between various acoustic features and emotions or paralinguistic information (Akkaralaertsest and Yingthawornsuk, 2015; Cummins et al., 2015; Joshi et al., 2013; Low et al., 2011). Among these reports, a relationship between MFCCs and depression was frequently reported. However, most of the data used in these challenges were derived from normal subjects, and few studies explored the MFCCs derived from the voices of patients with depression (Akkaralaertsest and Yingthawornsuk, 2015; Cummins et al., 2011). Furthermore, most studies adopted approach which artificial intelligence discriminates depressive voice from healthy voice by multiple MFCCs and no studies investigated which MFCCs are related to depressive voice. If acoustic features were used as biomarker of depression, it could make differential diagnosis of depressive states. That means that voice of helpline call or interactive voice response system would be an asset for evaluation of depressive state or risk assessment of the person. Questionnaires have risk of dishonest responses. In that respect, assessment by voice could be an advantage. It could be also useful for people with dementia who have difficulty to answer questionnaires by themselves.

In order to achieve differential diagnosis of depression, as a first step, we investigated whether various vocal acoustic features, including acoustic features indirectly-relevant to how voice is heard, could allow discrimination between patients with depression and control subjects in this study. Based on previous reports, we hypothesized that some MFCCs could be descriminant factors, and thus investigated which voice properties affected MFCCs. Considering the possibility of clinical application, we evaluated short utterances. In addition, we compared voices before and after a verbal fluency task (VFT), which is known to activate the frontal lobe (Herrmann et al., 2003; Pu et al., 2012), to examine whether such tasks could influence acoustic features.

2. Methods and materials

2.1. Subjects

Thirty-eight patients with depression were enrolled from the department of psychiatry at the University of Tsukuba Hospital, Tsukuba University Health Center, Ibaraki Prefectural Medical Center of Psychiatry and Kurita Hospital. They all met the criteria for major depressive disorder (MDD), which were determined with clinical interviews based on the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Two patients whose voices were not recorded properly were excluded, and the remaining 36 patients were adopted for the study. The control group consisted of 66 subjects who did not have current or past psychiatric disorders and who were recruited from the local population. One subject whose voice was not recorded properly was excluded. Two participants were excluded because they had a history of major medical or neurological illness, significant head trauma, or a lifetime history of alcohol or drug dependence. In addition, we excluded 27 subjects whose scores for the Quick Inventory of Depressive Symptomatology - Self-Report, Japanese version (QIDS-SRJ) (Fujisawa, 2010; Rush et al., 2003) were higher than six (cut-off point). As a result, the subjects consisted of 36 patients with MDD (22 males and 14 females; age: 21–79 years; mean age ± standard deviation (SD): 44.0 ± 16.3), and 36 healthy control individuals (16 males and 20 females; age: 22–58 years; mean age ± SD: 38.0 ± 10.4). In depression patients, the mean duration of illness was 7.00 ± 5.73(SD) years and the median was 5.47 years. The equivalent doses (Inada and Inagaki, 2015) of psychotropic drugs (mean ± SD) were 111.0 ± 72.4 mg of imipramine in antidepressants, 54.1 ± 117.5 of chlorpromazine in antipsychotics, and 9.9 ± 11.0 of diazepam in anxiolytics. Only one patient was unmedicated. All the subjects were Japanese. Demographics for the subjects are summarized in Table 1. This study was approved by all of the ethical committees of the department of psychiatry at the University of Tsukuba Hospital, Tsukuba University Health Center, Ibaraki Prefectural Medical Center of Psychiatry, and Kurita Hospital, and written informed consent was obtained from each participant.

2.2. Voice recording

Each subject was asked to read out ten digits "012–345–6789" like a telephone number. Next, we administered a VFT in which subjects spoke aloud as many words as possible beginning with the vowels "a", "u", and "o" within thirty seconds. Afterwards, subjects were asked to again read out the number "012–345–6789". Each task was conducted in Japanese. Voices were recorded using a Google Nexus 7 (TM) tablet

Table 1
Demographics of subjects.

	Depression patients (mean ± SD)	Controls (mean ± SD)	*p*
n (male/female)	36 (22/14)	36 (16/20)	n.s.
Age (years)	44.0 ± 16.3	38.0 ± 10.4	n.s.
QIDS-SRJ	11.7 ± 6.2	2.47 ± 1.8	< 0.001
Disease duration (years)	7.00 ± 5.73		
Antidepressants (mg)	114.0 ± 72.4		
Antipsychotics (mg)	54.1 ± 117.5		
Anxiolytics (mg)	9.9 ± 11.0		

SD: standard deviation.
QIDS-SRJ: Quick Inventory of Depressive Symptomatology - Self-Report, Japanese version.
n.s.: not significant.
Psychotropic drugs were categorized as antipsychotics, antidepressants and anxiolytics, and the dose of each drugs was calculated as equivalent doses (antypsychotics: chlorpromazine equivalent dose, antidepressants: imipramine equivalent dose, anxiolytics: diazepam equivalent dose) (Inada and Inagaki, 2015).