A mental health assessment method based on emotion level derived from voice

Shuji Shinohara (shinohara@bioeng.t.u-tokyo.ac.jp)
The University of Tokyo https://orcid.org/0000-0001-8442-836X
Mitsuteru Nakamura
The University of Texas
Yasuhiro Omiya
PST Inc
Naoki Hagiwara
AGI Inc.
Shunji Mitsuyoshi
The University of Tokyo
Hiroyuki Toda
National Defense Medical College
Taku Saito
National Defense Medical College
Masaaki Tanichi
National Defense Medical College
Aihide Yoshino
National Defense Medical College
Shinich Tokuno
The University Tokyo

Research article

Keywords: mental health assessment, vitality, mental activity, voice emotion analysis, non-invasiveness

Posted Date: December 7th, 2019

DOI: https://doi.org/10.21203/rs.2.18354/v1

Title

A mental health assessment method based on emotion level derived from voice

Authors

Shuji Shinohara[a], Mitsuteru Nakamura[b], Yasuhiro Omiya[c], Naoki Hagiwara[d], Shunji Mitsuyoshi[a], Hiroyuki Toda[e], Taku Saito[e], Masaaki Tanichi[e], Aihide Yoshino[e], and Shinich Tokuno[f]

Affiliations

[a]Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku 113-8656, Japan. Emails: shinohara@bioeng.t.u-tokyo.ac.jp (SS); mitsuyoshi@bioeng.t.u-tokyo.ac.jp (SM)

[b]Health Science Center at San Antonio, The University of Texas, San Antonio, TX, USA. Emails: m-nakamura@m.u-tokyo.ac.jp (MN)

[c]PST Inc., Yamashita-cho 2, Naka-ku, Yokohama, Japan. Email: omiya@medical-pst.com (YO)

[d]AGI Inc., 1-9-1-8F, Higashi-shimbashi, Minato-ku, Tokyo, Japan. Email: hagiwara@agi-web.co.jp (NH)

[e]Department of Psychiatry, National Defense Medical College, Namiki 3-2, Tokorozawa, Japan. Emails: toda1973@gmail.com (HT); t.saito3025@gmail.com, (TS); mtanichi@gmail.com (MT); aihide@ndmc.ac.jp (AY)

[f]Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan. Emails: tokuno@m.u-tokyo.ac.jp (ST)

Corresponding author

Shuji Shinohara

Tel: +81-3-5841-0439

Fax: +81-3-5841-7798

E-mail: shinohara@bioeng.t.u-tokyo.ac.jp

Abstract

Background: In many developed countries, mental health disorders have become a problem, and the economic loss due to treatment costs and interference with work is immeasurable. Therefore, a simple technique must be developed to determine individuals' depressive state and stress levels. Voice analysis using smartphones is not only noninvasive, it does not require a dedicated device; thus, it can be performed conveniently and remotely. Consequently, we developed a method to assess individuals' mental health levels using emotional components contained in the human voice.

Methods: We proposed two indices of mental health: a short-term index (vitality) and mental activity calculated from long-term trends in vitality. We used the voices of healthy individuals (men: n = 10, M_{age} = 42.7 ± 6.0 years; women: n = 4, M_{age} = 35.0 ± 14.4 years) and patients with major depression (men: n = 19, M_{age} = 43.7 ± 11.0 years; women: n = 11, M_{age} = 53.9 ± 8.2 years). For patients, simultaneously with voice collection, specialists assessed current depression severity using the Hamilton Rating Scale for Depression (HAM-D).

Results: A significant negative correlation existed between the vitality extracted from voice and HAM-D score ($r = -0.33$, $p < .05$). We could discriminate the voice data of healthy individuals and patients with depression (judged as moderate or severe by the

specialists) with high accuracy using vitality ($p = .0085$, the area under the curve (AUC) of the receiver operating characteristic curve = 0.87). However, there was no significant difference between the vitality of the healthy individuals and the patients judged to be the "no depression group with almost no depressive symptoms," even if they were outpatients with depression ($p > .1$, AUC = 0.64).

Conclusions: We developed a method to estimate stress through emotion instead of analyzing stress directly from voice data. By daily monitoring of vitality using smartphones, we can encourage hospital visits for people before they become depressed or during the early stages of depression. This may lead to reduced economic loss due to treatment costs and interference with work.

Trial registration: Not applicable.

Keywords: mental health assessment, vitality, mental activity, voice emotion analysis, non-invasiveness

Background

In many developed countries, mental health disorders have become a problem [1], and the economic loss due to treatment costs and interference with work is

63 immeasurable [2]. Therefore, a simple technique to determine individuals' depressive
64 state and stress level is desired.

65 Self-administered psychological tests, such as the General Health
66 Questionnaire (GHQ) [3,4] and Beck Depression Inventory (BDI) [5,6], can be used as
67 screening methods for patients with mental health disorders. In addition, a stress-check
68 method that uses biomarkers in saliva [7] and blood has been proposed [8]. Although
69 self-administered psychological tests are useful for early detection and as diagnostic
70 aids, there is a problem with reporting bias—in which specific information such as
71 smoking history and medical history are selectively suppressed or expressed by
72 participants [9]. Stress-check methods that use biomarkers also have problems such as
73 the cost of the test and the burden on the participants during specimen collection; i.e.,
74 they are not convenient.

75 On the other hand, with the recent widespread use of smartphones, pathological
76 analysis using voice data has become popular [10-12]. Voice analysis using
77 smartphones is not only noninvasive, it does not require a dedicated device; thus, it can
78 be performed conveniently and remotely.

79 The relationship between mental illness and voice has been observed in
80 previous studies; e.g., studies regarding the speaking rate of patients with depression

[13-15], studies on switching pause and percent pause of patients with depression [15,16], etc. There are also studies in which Lyapunov exponents and Kolmogorov entropy for the voice of patients with depression were measured using chaos analysis [17]. A study that used frequency analysis showed that the shimmer and jitter in vowels as voiced by patients with depression were higher than those of healthy people, and the first and second formant frequency were low [18]. Zhou and colleagues proposed a new feature derived from the Teager energy operator for the classification of voices under stress [19]. In another study, a method was proposed to assess mental health from envelope information for pitch and speech waveforms [20].

On the other hand, stress is known to have an impact on emotions [21], and a method is being developed to estimate stress through emotion instead of analyzing stress directly from voice data [22-24]. Mitsuyoshi and colleagues [22] proposed an algorithm that estimates the expression of emotion from emotion components of the voice—the vocal affect display. In addition, they experimentally analyzed the relationship between this index and stress and estimated individuals' stress level from their voice. In this study, sensibility technology (ST) that analyzes emotion in speakers' voices was used [25-27]. The present study proposes a method to assess the mental

health of a speaker from the emotional components in his or her voice using ST with a focus on the relationship between mental health and emotions.

Methods

Acquisition of Voices

In this study, we collected voice data in two categories—healthy individuals and outpatients with depression. All participants provided written consent. Voice acquisition of the patient group was performed intermittently from August 2013 to October 2014 with outpatients at Kitahara rehabilitation hospital in Japan. Voices were recorded during patients' conversations with physicians during examination. All data were then confirmed audibly; overlaps with other speakers and background noises were removed manually.

Voices of healthy people were acquired from February to mid-May 2015. During the acquisition period, participants worked normally at their jobs without visiting medical facilities for a mental illness. Voice acquisition was continuously performed once every several days; at each time, 14 types of fixed phrases were read aloud twice. Voices were recorded in a quiet environment with little background noise.

Voices were recorded by a gun microphone (AT9944: audio-technica, Tokyo, Japan) placed approximately 100 cm from participants, or by a pin microphone

(ME52W: OLYMPUS, Tokyo, Japan) attached to the chest at approximately 15 cm from participants' mouth. The recording device was MS-PST1 (NORITSU KOKI, Wakayama, Japan; not commercially available).

Table 1 shows participants' information per group. It should be noted that the number of participants and the number of data differed because data may have been collected multiple times from the same participant on different days. The average number of data collected per healthy person were 24.4 ± 33.3 for men and 6.3 ± 6.1 for women. For patients with depression, they were 6.0 ± 2.9 for men and 6.8 ± 3.2 for women. These collected data were used to create algorithms to calculate vitality and mental activity.

Table 1. Experimental participant information for algorithm preparation

Group	Sex	Number of participants	Mean age	Number of data
Healthy	Male	9	42.9 ± 5.6	220
	Female	4	33.3 ± 15.4	25
Major depression	Male	4	54.0 ± 12.0	24
	Female	5	49.4 ± 15.4	34

Regarding the above-described recorded voice, a healthy person's voice is a fixed-phrase utterance. On the other hand, a patient's voice is a free speech in the form of dialogue with a doctor, and the type of speech differed between a healthy person and a patient. Further, the recording location differed. To unify both speech types and

131 recording environments, data for algorithm verification were collected at the National
132 Defense Medical College Hospital in Japan with participants' consent. Participants were
133 informed that the anonymity and confidentiality of their data were guaranteed, and that
134 they were free to withdraw at any time. Participants were not rewarded for their
135 participation.

136 First, from December 2015 to June 2016, fixed-phrase reading voices were
137 collected from outpatients with major depression. Table 2 shows 17 types of Japanese
138 fixed phrases that were used for recording. At the time of voice collection, specialists
139 evaluated patients' depression severity using the Hamilton Rating Scale for Depression
140 (HAM-D) [28]. The HAM-D is not a self-assessment-type psychological test; rather,
141 experts such as doctors evaluate the characteristic items of depression symptoms. The
142 purpose of the HAM-D is for a professional to objectively quantify an individual's
143 depressive state. On the other hand, for voices of healthy individuals, in mid-December
144 2016, the same fixed-phrase reading voices as the patients were recorded in the same
145 examination room as the patients. However, for healthy people, severity assessment
146 using the HAM-D was not conducted.

147 **Table 2. Seventeen phrases used for recording**

No.	Phrase in Japanese	Purpose (meaning)

1	I-ro-ha-ni-ho-he-to	Non-emotional (no means like “a-b-c”)
2	Honjitsu ha seiten nari	Non-emotional (It is fine today)
3	Tsurezurenaru mama ni	Non-emotional (Having nothing to do)
4	Wagahai ha neko dearu	Non-emotional (I am a cat)
5	Mukashi mukashi aru tokoro ni	Non-emotional (Once upon a time, there lived)
6	a-i-u-e-o	Check pronunciation of vowel sounds (no means like “a-b-c”)
7	Ga-gi-gu-ge-go	Check sonant pronunciation (no means like “a-b-c”)
8	Ra-ri-ru-re-ro	Check liquid sound pronunciation (no means like “a-b-c”)
9	Pa-pi-pu-pe-po	Check p-sound pronunciation (no means like “a-b-c”)
10	Omoeba tooku he kita monda	Non-emotional (When thinking, I have come to the far place)
11	Garapagosu shotou	Check pronunciation (Galápagos Islands)
12	Tsukarete guttari shiteimasu.	Emotional (I am tired/dead tired)
13	Totemo genki desu	Emotional (I am very cheerful)
14	Kinou ha yoku nemuremashita	Emotional (I was able to sleep well yesterday)
15	Shokuyoku ga arimasu	Emotional (I have an appetite)
16	Okorippoi desu	Emotional (I am irritable)
17	Kokoroga odayaka desu	Emotional (My heart is calm)

These voices were recorded by a pin microphone ME52W (OLYMPUS, Tokyo, Japan) attached to the chest about 15 cm from participants’ mouth. The recording device used was Portable Recorder R-26 (Roland, Shizuoka, Japan). Table 3 shows participants’ information for algorithm verification. The number of healthy individuals for verification and the number of their voice data were the same because

they were collected only once from each healthy participant. Regarding patients, some participants performed multiple data acquisitions. Seven, three, and one performed data acquisition twice, three, and four times, respectively. Data were acquired only once from the remaining 19 people. The recording format of the voices was linear PCM, the sampling frequency was 11025 Hz, and the number of quantization bits was 16 bits.

Table 3. Experimental participant information for algorithm verification

Group	Sex	Number of participants	Mean age	Number of data
Healthy	Male	10	42.7 ± 6.0	10
	Female	4	35.0 ± 14.4	4
Major depression	Male	19	43.7 ± 11.0	34
	Female	11	53.9 ± 8.2	12

Voice Emotion Analysis System

We used software ST Ver. 3.0 (AGI Inc., Tokyo, Japan) [25-27] to extract emotions from participants' voice. The categories of emotional elements detected by ST software are: "anger," "joy," "sorrow," "calmness" and "excitement." The strength of each emotion is represented as an integer value from 0 to 10. A value of 0 means that the input speech does not contain the emotion at all. A value of 10 means that the input speech contains the emotion most strongly. The unit of speech emotion analysis by ST software is "utterance." This is a part of continuous voice divided by breath. When a silent state changes to a speech state, it is considered that an utterance has started. When

the speech state continues for a certain period and changes to silence, it is considered that the utterance has ended. Whether the silent state or the speech state is determined from the volume using the threshold. The threshold was adjusted manually for each recording, as the volume of the audio is affected by the participant and the condition of the recording.

Algorithm

Vitality and Mental Activity

We proposed two scales—vitality and mental activity—as indices for the degree of mental health obtained through voice analysis. Generally, “vitality” can be defined in diverse ways; however, here, vitality refers to a scale that measures low for patients with illnesses such as depression and high for healthy people. The main difference between vitality and mental activity is the duration of the measurement. Vitality is calculated from the emotional components of voice (calm, anger, joy, sorrow, and excitement) based on short-term voice data such as a single phone conversation or a hospital visit.

On the other hand, mental activity is calculated based on vitality data accumulated over a certain period. Vitality changes based on the conditions at the time of measurement in the same manner that blood pressure changes between post workout

and at rest. As accurate identification of high blood pressure is possible through long-term monitoring, in this study, we aimed to accurately assess mental health by introducing mental activity.

Vivacity and Relaxation

To calculate vitality, we introduced two new indices: “relaxation” and “vivacity.” To define these indices, we used four out of five indices output by ST: calmness, joy, sorrow, and excitement.

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders describes the characteristics of a major depressive episode as a continuing depressive state with loss of interests and happiness and feeling sorrow and emptiness [29]. In contrast, if there is a component of joy more relative to sorrow in emotion, it is considered a good mental state. Consequently, vivacity for an utterance was defined as follows:

$$Vivacity = \frac{Joy}{Joy + Sorrow} \tag{1}$$

Stress and tension are major factors in mental health disorders. On the other hand, the relaxed state is mentally positive; thus, relaxation for an utterance was defined as follows:

$$Relaxation = \frac{Calm}{Calm + Excitement} \tag{2}$$

In other words, relaxation increases with the increasing calmness component of emotion and decreasing excitement. Each emotional value output by ST, as well as the excitement, are expressed with integers in the range of 0–10. Therefore, vivacity and relaxation become real numbers in the range of 0.0–1.0. Vivacity and relaxation as defined above were calculated for each utterance. Below, we define vivacity and relaxation for an acquired voice as the mean value for each utterance contained in the acquired voice.

Vitality Calculation Algorithm

Vitality was calculated as the weighted mean of vivacity and relaxation defined in the previous section. Fig. 1 shows a scatter plot of relaxation and vivacity as calculated from data for the algorithm preparation.

Fig. 1. Scatter plot of relaxation and vivacity. × and □ show the data of the healthy and patient groups for algorithm preparation. The straight line represents $0.60x + 0.40y = 0.52$.

The symbol × in the figure represents the data of the healthy group, while □ represents the data of the patient group. Data are plotted for each voice acquisition. There are 245

data for the 13 people in the healthy group, and 58 for the 9 people in the patient group. We added a straight line that separates the healthy group from the patient group (0.60X + 0.40Y = 0.52). Based on this line, the vitality for each acquired voice was defined as follows:

$$Vitality = 0.60 \times Vivacity + 0.40 \times Relaxation \quad (3)$$

Mental Activity Calculation Algorithm

Vitality was calculated from short-term voice data such as a single examination or consultation. Therefore, depending on participants' current mood, even healthy people might score low in vitality, while patients may score high. To compensate for such a weakness, mental activity was calculated from long-term trends in vitality. Specifically, to express long-term trend in vitality, we calculated the mean of accumulated vitality ($\overline{Vitality}$).

Furthermore, when vitality has little fluctuation and is stagnant at low values, it is determined to have low mental activity. To actualize such a determination, we introduce a new index: standard deviation ($VitalitySD$) that expresses variations in vitalities for utterances contained in acquired voice. Then, the mean of vitality standard deviation of the accumulated acquired voice ($\overline{VitalitySD}$) was calculated.

Fig. 2. Scatter plot of mean vitality and the mean (standard deviation) of vitality for each participant. × and □ indicate data for the healthy group and the patient group, respectively. The straight line represents $0.75x + 0.25y = 0.426$.

Fig. 2 is a scatter plot of the mean vitality and mean standard deviation of vitality for each participant calculated from the data for the algorithm preparation. The number of data plotted were 13 people for the healthy group and 9 people for the patient group (N = 22). When calculating the mean, we used all acquired voice data. In the figure, we added a straight line that separates the healthy group and the patient group (0.75X + 0.25Y = 0.426). Based on this line and using the mean and standard deviation of vitality, we define mental activity as follows:

$$MindActivity = 0.75 \times \overline{Vitality} + 0.25 \times \overline{VitalitySD} \quad (4)$$

Method of Analysis

According to the definition of Zimmerman and colleagues [30], the data of the patient group were divided into 3 groups by HAM-D score: no depression (≤ 7), mild (8–16), and moderate or severe (≥ 17).

253 The vitality of the four groups (i.e., these three and the healthy group) were compared
254 among each other. P-values from Tukey-Kramer tests, the area under the curve (AUC)
255 against Receiver Operating Characteristic (ROC), sensitivity, and specificity were used
256 to evaluate the classification accuracy of Vitality.
257
258 **Results**
259 ***HAM-D score***
260 The mean values of HAM-D score in each group are shown in Table 4. In
261 addition, the mild group and the moderate or severe group will be collectively referred
262 to as the depression group (HAM-D score ≥ 8). The mean HAM-D score for the
263 depression group was 16.1 ± 7.4 (n = 22). The number of participants in each group was
264 11 men and 8 women in the no depression group and 5 men and 3 women in the mild
265 group. All three participants in the moderate or severe group were men.
266 **Table 4. Average value of HAM-D score for each group**

Group	Number of participants	Number of data	HAM-D score
No depression (HAM-D ≤ 7)	19	24	3.1 ± 2.3
Mild (HAM-D = 8–16)	8	13	11.5 ± 3.2
Moderate or severe	3	9	22.8 ± 6.6

(HAM-D ≥ 17)

Performance evaluation of vitality

We evaluated the performance of vitality using the data for algorithm verification shown in Table 3. Fig. 3 shows the relationship between HAM-D score and vitality for 46 data obtained from the patient group. There was a significant negative correlation between the two ($r = -0.33$, $n = 46$, $p < .05$).

Fig. 3. Relationship between HAM-D score and vitality in the data of patient group for algorithm verification.

Fig. 4. Comparison of vitality for each group. (a) represents data distribution of a healthy group, the no depression group, and the depression group. (b) shows the data distribution when the depression group is divided into the mild group and the moderate or severe group.

Fig 4 shows the distribution of vitality scores of the healthy group, the no depression group, the mild group, the moderate or severe group, and the depression

group. The mean vitality in each group was 0.60 ± 0.10 (n = 14), 0.55 ± 0.10 (n = 24), 0.51 ± 0.13 (n = 13), 0.47 ± 0.07 (n = 9), and 0.49 ± 0.11 (n = 22), respectively.

The Tukey-Kramer test revealed significant differences between the healthy group and the depression group, and between the healthy group and the moderate or severe group (Ps = .0085 and .020, respectively). We used statistical analysis software R [31].

Next, to evaluate the discrimination performance of vitality, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, the sensitivity, and the specificity were used. Figure 5 shows the ROC curves when using vitality to identify whether the data for verification is for the healthy group or for each patient group. Here, the horizontal axis represents 1-specificity (false positive rate), and the vertical axis represents sensitivity (positive rate).

Fig. 5. Receiver operating characteristic curves when using vitality to identify groups

Table 5 shows the performance when the data of the healthy group and each group were distinguished using vitality. The AUC was 0.87, and the sensitivity and

specificity were 0.78 and 0.86, respectively regarding the discrimination performance between the healthy group and the moderate or severe group. On the other hand, both AUC were less than 0.7 regarding discrimination performance between the healthy group and the no depression group or mild group.

Table 5. Discrimination ability of vitality

Group	AUC	Sensitivity	Specificity
Health–Depression	0.76	0.55	0.93
Health–Moderate or severe	0.87	0.78	0.86
Health–Mild	0.69	0.46	0.93
Health–No depression	0.64	0.80	0.50

AUC: area under the curve of the receiver operating characteristic curve.

Discussion

In this study, we developed a method to measure mental health using emotional components contained in voice. Two indicators were proposed: vitality based on short-term voice data and mental activity calculated from long-term voice data. As shown in Fig 3, there was a significant negative correlation between vitality and HAM-D score (i.e., depression severity assessed by a physician). In addition, as shown in Fig. 4, the group with a higher severity of depression tended to have a lower mean vitality.

There was a significant difference between the healthy group and the depression group, and between the healthy group and the moderate or severe group in

316 vitality. On the other hand, there was no significant difference between the healthy
317 group and the no depression group with almost no depressive symptoms, even if they
318 were outpatients with depression. This suggests the possibility of measuring treatment
319 effects by vitality (i.e., voice). Moreover, as shown in Fig. 5 and Table 5, the voice data
320 of the healthy group and the voice data of the moderate or severe group could be
321 identified with high accuracy using vitality. This suggests the possibility of screening
322 for severe depression in individuals by using voice.

323 In our other study, we verified vitality with Romanian and Russian native
324 speakers [32]. In this verification, BDI tests were conducted simultaneously with voice
325 recordings. There was a significant difference between the mean vitality of the
326 depression high-risk group (BDI scores ≥ 17) and the mean vitality of the depression
327 low-risk group (BDI scores < 17; $p < .05$). Specifically, the scores for question 9—
328 concerning suicidal ideation—took a value that ranged 0–3. There was a significant
329 difference between the mean vitality of the suicide low-risk group (0 or 1 points) and
330 the mean vitality of the suicide high-risk group (2 or 3 points; $p < .01$). In the future, we
331 will examine the vitality of native speakers of other languages, such as English.

332 As a limitation of this research, only the fixed phrase read-out speech was used
333 for verification. To apply vitality to free speech such as a call, further verification is

required. Furthermore, in the verification data, the number of voices collected for each participant, the sex ratio, and the age were not unified between groups. These differences may be reflected in the features of voice. For example, all participants in the moderate or severe group were men, and the number of participants was as small as three. In the future, it is necessary to acquire a lot of voices of female patients, especially those with severe depression, and to evaluate the performance level of vitality.

Further, mental activity was not validated because continuous data could not be collected sufficiently for the same participants in both the healthy group and the patient group. However, comparing Figs. 1 and 2 showing data for algorithm preparation, there is a possibility that mental activity can more accurately identify the data as compared to vitality, which will be addressed in the future.

Vitality and the mental activity can be measured by only voice, and their advantages are that they are non-invasive and less expensive as compared to self-administered tests such as the GHQ-30 and BDI and stress-check methods using saliva and blood. Moreover, it is also possible to record day-to-day state changes easily by implementing them on a smartphone or the like.

We developed a smart phone application that implemented the algorithm for vitality and mental activity—the Mind Monitoring System (MIMOSYS). We are currently conducting world-wide demonstration experiments using the MIMOSYS [33]. In the future, we plan to verify the effectiveness of vitality and mental activity with such large-scale data.

Conclusions

In this study, we developed a method to measure mental health from voice. The algorithm to estimate stress through emotion instead of analyzing stress directly from voice data is novel. The MIMOSYS implemented the algorithm for vitality and mental activity, which is a cost-effective and convenient measurement device. If the correlation between HAM-D score and vitality can be further enhanced, it may be used to aid doctors' diagnoses in the future. By daily monitoring of vitality and mental activity using the MIMOSYS, we can encourage hospital visits for people before they become depressed or during the early stages of depression. This may lead to reduced economic loss due to treatment costs and interference with work.

List of abbreviations

MIMOSYS: Mind Monitoring System

GHQ: General Health Questionnaire

BDI: Beck Depression Inventory

HAM-D: Hamilton Rating Scale for Depression

ST: sensibility technology

ROC: receiver operating characteristic

AUC: area under the curve

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the National Defense Medical Collage Ethics Committee (no. 2248) and the Kitahara Rehabilitation Hospital Ethics Committee (no. 3).

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Center of Innovation Program from the Japan Science and Technology Agency and by JSPS KAKENHI [grant numbers JP16K01408 and JP15H03002]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

S. T. was responsible for the design of the clinical study. A. Y., H. T., T. S., and M. T. were responsible for the execution of the clinical study including patient recruitment and retention and data collection. S. S. conceived the algorithm, analyzed data, and wrote the manuscript. M. N., Y. O., N. H., S. M., and S. T. contributed to the interpretation of study findings. All authors participated in the editing and revision of the final version of the manuscript.

Acknowledgements

We thank Dr. Shinsuke Kondo for assistance with data collection and all participants for participating. We also thank Editage [http://www.editage.com] for English-language editing.

405 **References**

406 [1] World Health Organization. The global burden of disease: 2004 update. WHO Press,
407 Geneva, Switzerland; 2008. pp. 46-9.

408 [2] Kessler RC, Akiskal HS, Ames M, Birnbaum H, Greenberg P, Hirschfield, RM, et al.
409 Prevalence and effects of mood disorders on work performance in a nationally
410 representative sample of U.S. Workers. Am J Psychiatry. American Psychiatric
411 Association Publishing; 2006;163(9):1561-8.

412 [3] Goldberg DP, Blackwell B. Psychiatric illness in general practice: a detailed study
413 using a new method of case identification. Br Med J. 1970;2(5707):439-43.

414 [4] Goldberg D. Manual of the general health questionnaire. NFER Nelson; 1978.

415 [5] Beck AT. A systematic investigation of depression. Compr Psychiatry.
416 1961;2(3):163-70.

417 [6] Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring
418 depression. Arch Gen Psychiatry. 1961;4(6):561-71.

419 [7] Takai N, Yamaguchi M, Aragaki T, Eto K, Uchihashi K, Nishikawa Y. Effect of
420 psychological stress on the salivary cortisol and amylase levels in healthy young
421 adults. Arch Oral Biol. 2004;49(12):963-8.

[8] Suzuki G, Tokuno S, Nibuya M, Ishida T, Yamamoto T, Mukai Y, et al. Decreased plasma brain-derived neurotrophic factor and vascular endothelial growth factor concentrations during military training. PLoS One. 2014;9(2):e89455.

[9] Miquel P. A dictionary of epidemiology. 6th ed. Oxford University Press; 2014.

[10] Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. Parkinsonism Relat Disord. 2015;21(6):650-3.

[11] Rachuri KK, Musolesi M, Mascolo C, Rentfrow PJ, Longworth C, Aucinas A. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In Proceedings of the 12th ACM international conference on Ubiquitous computing 2010 Sep 26 (pp. 281-290). ACM.

[12] Lu H, Frauendorfer D, Rabbi M, Mast MS, Chittaranjan GT, Campbell AT, et al. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing 2012 Sep 5 (pp. 351-360). ACM.

[13] Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. Brain Cognition. 2004;56(1):30-5.

[14] Moore E, Clements M, Peifer J, Weisser L. Analysis of prosodic variation in speech for clinical depression. In Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439) 2003 Sep 17 (Vol. 3, pp. 2925-2928). IEEE.

[15] Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguistics. 2007;20(1):50-64.

[16] Yang Y, Fairbairn C, Cohn JF. Detecting depression severity from vocal prosody. IEEE Transactions on Affective Computing. 2012;4(2):142-50.

[17] Shimizu T, Furuse N, Yamazaki T, Ueta Y, Sato T, Nagata S. Chaos of vowel/a/in Japanese patients with depression: a preliminary study. J Occupat Health. 2005;47(3):267-9.

[18] Vicsi K, Sztahó D, Kiss G. Examination of the sensitivity of acoustic-phonetic parameters of speech to depression. In 2012 IEEE 3rd International Conference on Cognitive Info Communications (CogInfoCom) 2012 Dec 2 (pp. 511-515). IEEE.

[19] Zhou G, Hansen JH, Kaiser JF. Nonlinear feature based classification of speech under stress. IEEE Trans Speech Audio Process. 2001;9(3):201-16.

[20] Shinohara S, Mitsuyoshi S, Nakamura M, Omiya Y, Tsumatori G, Tokuno S. Validity of a voice-based evaluation method for effectiveness of behavioural therapy. In International Symposium on Pervasive Computing Paradigms for Mental Health 2015 Sep 24 (pp. 43-51). Springer, Cham.

[21] Lazarus RS. From psychological stress to the emotions: a history of changing outlooks. Ann Rev Psychol. 1993;44(1):1-22.

[22] Mitsuyoshi S, Nakamura M, Omiya Y, Shinohara S, Hagiwara N, Tokuno S. Mental status assessment of disaster relief personnel by vocal affect display based on voice emotion recognition. Disaster Mil Med. Springer Nature; 2017;3(1).

[23] Tokuno S, Mitsuyoshi S, Suzuki G, Tsumatori G. Stress evaluation using voice emotion recognition technology: a novel stress evaluation technology for disaster responders. In Proceedings of the XVI World Congress of Psychiatry 2014 Sep (Vol. 2, p. 301).

[24] Tokuno S, Tsumatori G, Shono S, Takei E, Yamamoto T, Suzuki G, et al. Usage of emotion recognition in military health care. 2011 Defense Science Research Conference and Expo. IEEE; 2011 Aug.

[25] Mitsuyoshi S, Ren F, Tanaka Y, Kuroiwa, S. Non-verbal voice emotion analysis system. Int J ICIC. 2006;2(4).

[26] Mitsuyoshi S, Shibasaki K, Tanaka Y, Kato M, Murata T, Minami T, et al. Emotion voice analysis system connected to the human brain. 2007 International Conference on Natural Language Processing and Knowledge Engineering. IEEE; 2007 Aug.

[27] Mitsuyoshi S, inventor; Advanced Generation Interface Inc, assignee. Emotion recognizing method, sensibility creating method, device, and software. United States patent US 7,340,393. 2008 Mar 4.

[28] Hamilton MA. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol. 1967;6(4):278-96.

[29] America Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington: D.C.; 2013.

[30] Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton depression rating scale. J Affect Disord. 2013;150(2):384-8.

[31] https://cran.r-project.org/

[32] Uraguchi T, Shinohara S, Denis N A, Țaicu M, Săvoiu G, Omiya Y, et al. Evaluation of Mind Monitoring System (MIMOSYS) by subjects with Romanian and

Russian as their native language. 40th International Conference of the IEEE Engineering in Medicine and Biology Society; 2018 July.

[33] Shinohara S, Omiya Y, Hagiwara N, Nakamura M, Higuchi M, Kirita T, et al. Case studies of utilization of the mind monitoring system (MIMOSYS) using voice and its future prospects. ESMSJ. 2017;7(1):7-12.

Figures

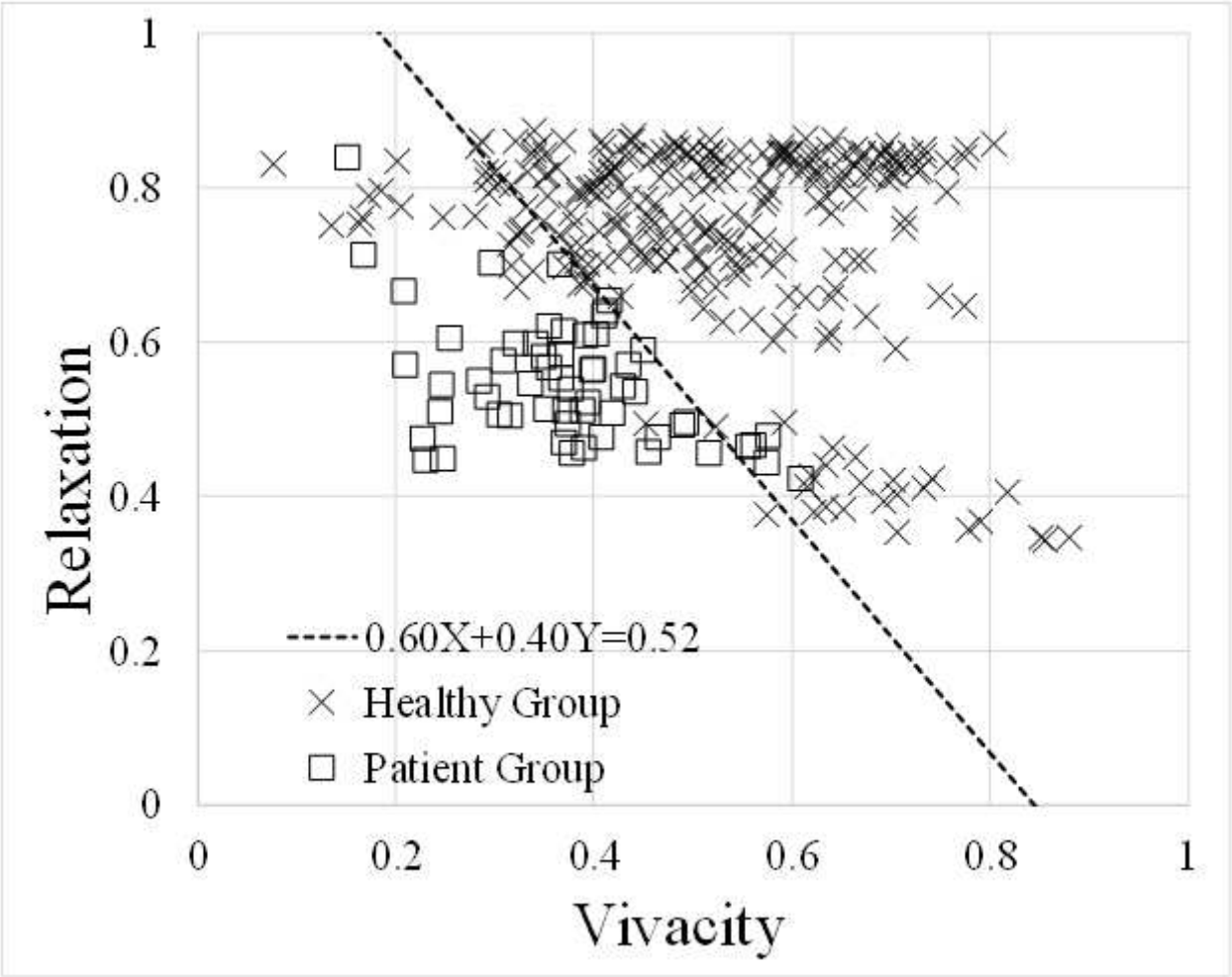

Figure 1

Scatter plot of relaxation and vivacity. × and □ show the data of the healthy and patient groups for algorithm preparation. The straight line represents 0.60x+0.40y=0.52.

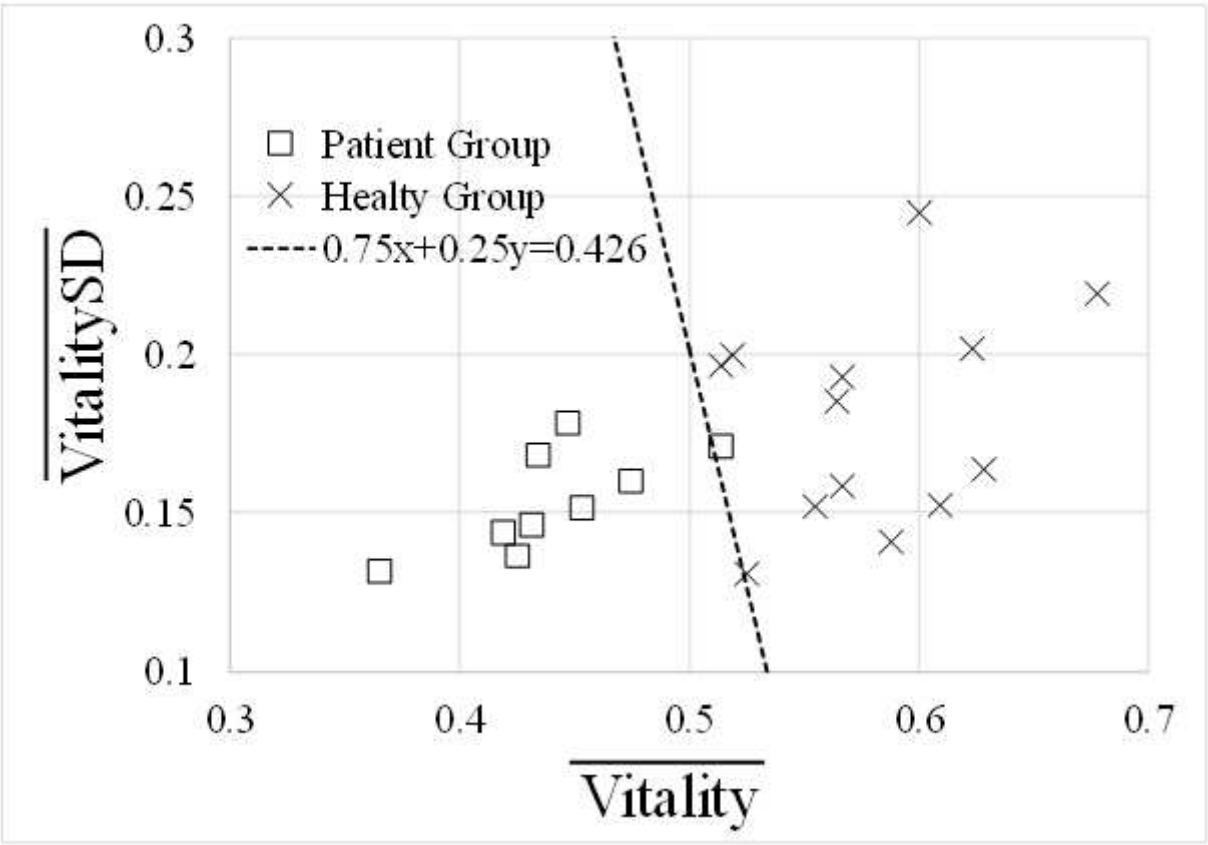

Figure 2

Scatter plot of mean vitality and the mean of standard deviation of vitality for each subject. × and □ indicate data for the healthy group and the patient group, respectively. The straight line represents 0.75x+0.25y=0.426.

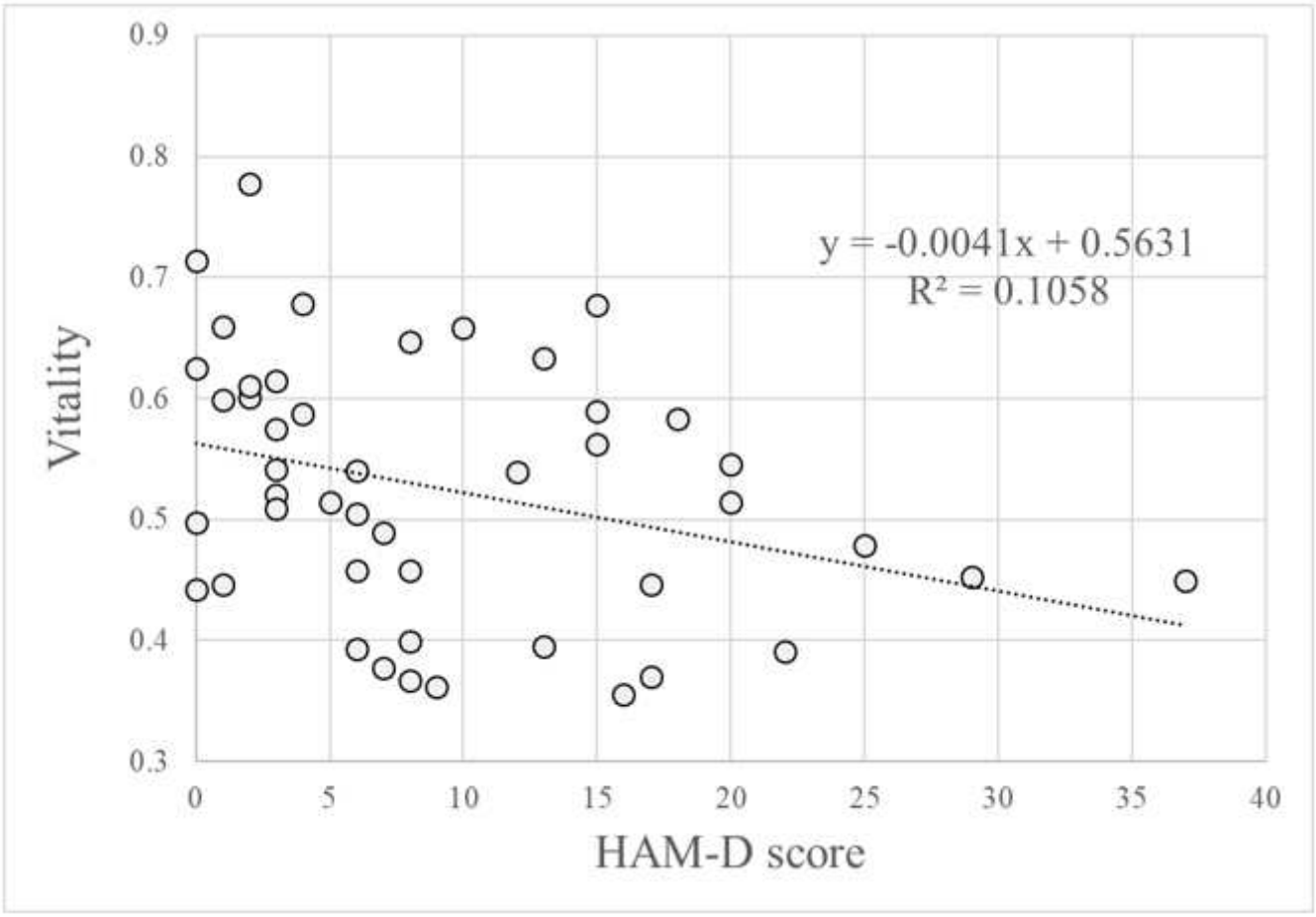

Figure 3

Relationship between HAM-D score and vitality in the data of patient group for algorithm verification.

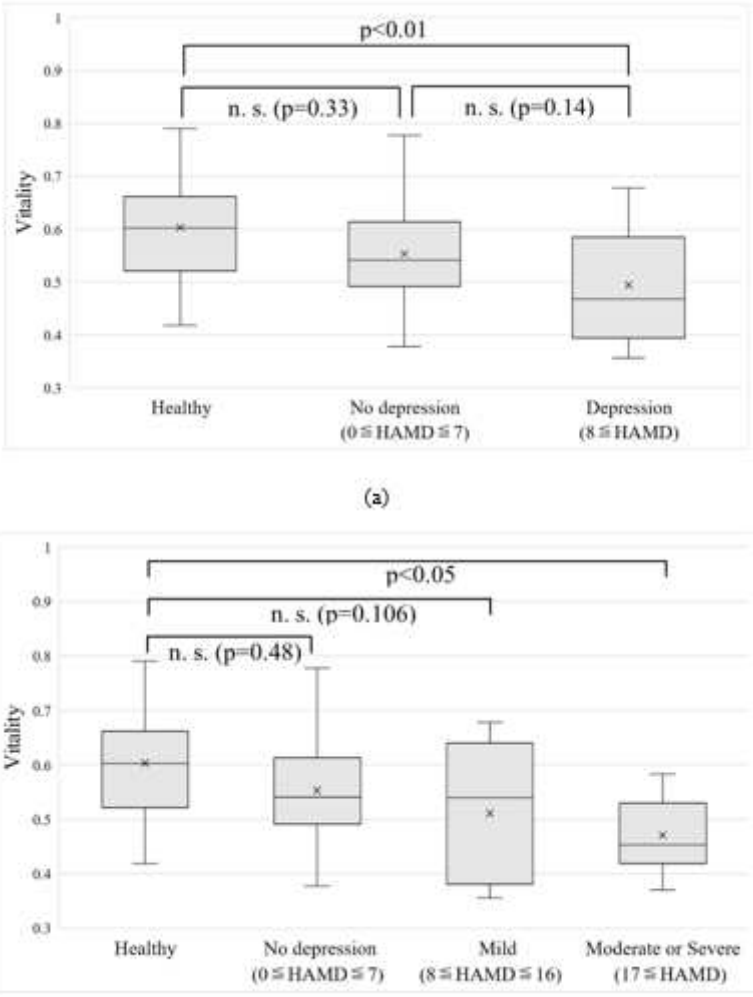

Figure 4

Comparison of the vitality for each group. (a) represents data distribution of a healthy group, a no depression group, and a depression group. (b) shows the data distribution when the depression group is divided into the mild group and the moderate or severe group.

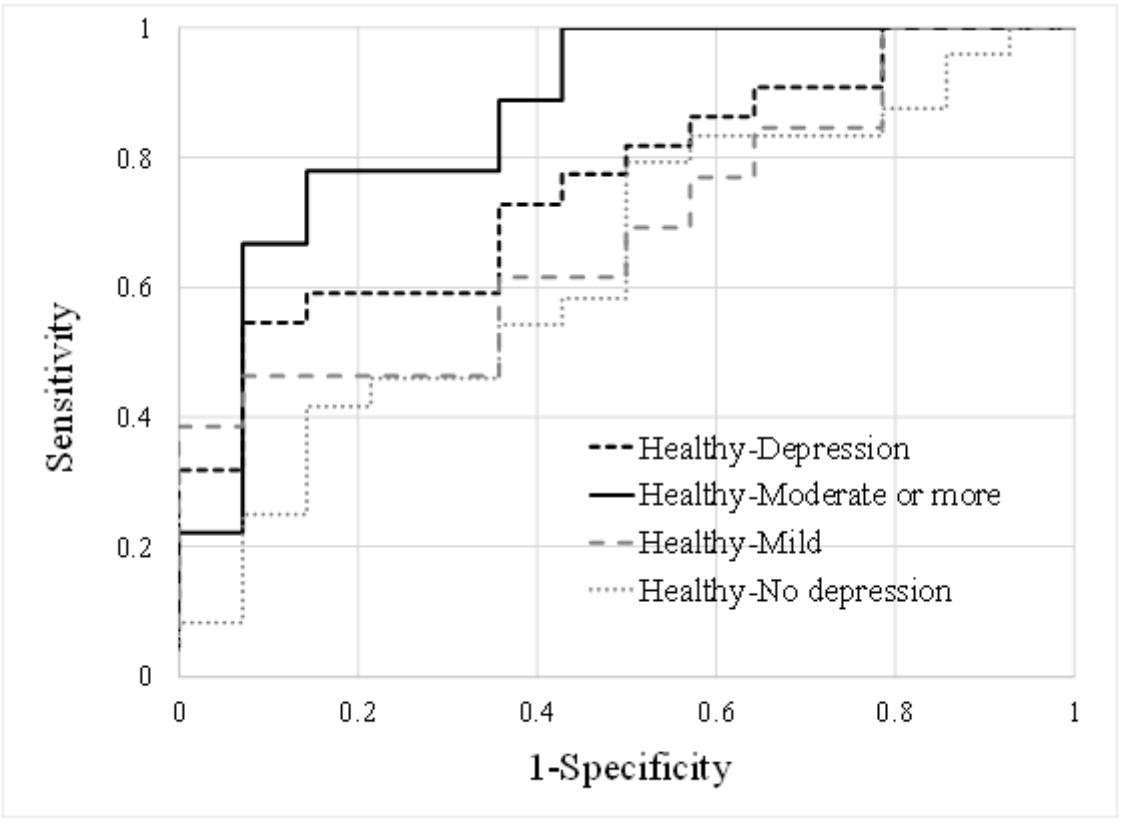

Figure 5

ROC curves when using the vitality to identify whether the data for verification is for the healthy group or for each patient group