# STRESS EVALUATION BY VOICE: FROM PREVENTION TO TREATMENT IN MENTAL HEALTH CARE

Shinichi_Tokuno

Verbal Analysus of Pathophysiology, Graduate School of Medicine, The University of Tokyo,
e-mail: tokuno@m.u-tokyo.ac.jp

Abstract. *The implementation of large-scale mental health care requires low-cost, high-accuracy screening methods. Reporting bias cannot be eliminated from the self-administered screening methods that are currently in general use, and screening using biomarkers, which has seen remarkable developments in recent years, remains costly.Mental disorders resulting from stress alter the expression of emotion and cause changes in certain voice qualities. We have therefore attempted to assess stress intensity using vocal emotion recognition technology through a computer programme. In a prior verification comparing our programme with assessments through self-administered questionnaires (GHQ-30) and interview assessments, the proposed technology obtained almost the same detection sensitivity as GHQ-30.The main advantages of the present technology are its ease of implementation and low cost. Its application in preventative medicine is therefore promising, and, if used in combination with the various biomarker-based diagnostic techniques currently in development, better induction to health care specialists will be possible.*

*Keywords: emotion recognition, screening system, mental health, mental stress, depression*
PACS numbers: 87.19.X-, 87.15.ad, 07.64.+z

## 1. INTRODUCTION

Many victims and relief workers involved in the Great East Japan Earthquake are said to be affected by post-traumatic stress disorder (PTSD) or depression. In implementing large-scale mental health care for PTSD and depression at times of disasters, it is not possible for specialists to conduct interviews with all subjects. This is similar for mental health care in large corporations. Therefore, a screening technique which is inexpensive, simple, and highly accurate is highly necessary.

Generally, mental health care for major disorders in a population involves counselling or treatment after filtering based on self-administered tests. However, reporting bias (mainly underestimation due to self-consciousness) cannot be eliminated from these self-administered tests. Specifically, it has been reported that, in hierarchical organisations such as the fire services, police, and armed forces, detection rates are significantly reduced [1], [2], [3]. The reasons for this underestimation in self-administered screening are thought to include resistance, prejudice, and discrimination with regards to mental health problems, as well as concern over adverse effect on one's career [1].

To prevent reporting bias, the tests should include indices that subjects are either unaware of or cannot control. In recent years, screening and auxiliary diagnostic techniques based on biomarkers have undergone remarkable developments as a tool for objective evaluations in the field of psychiatry. As shown in Fig. 1, large numbers of biological substances and physiological indicators are studied as biomarker candidates. Each has its own characteristics, and at this point in time, none have been found that can replace self-administered screening in large-scale mental health care. For instance, cortisol, amylase, chromogranin A, and catecholamine have short response times with respect to stress, which makes the timing of examinations difficult, and individual abnormalities cannot be discerned in most markers if they are not compared with the 'normal' baseline. Moreover, as indicators such as catecholamine, cortisol, heart rate variability, brain waves (alpha waves), and the acceleration of pulse waveforms also react to factors other than mental stress, it is essential that these are used in conjunction with other findings. Sometimes, special testing equipment is needed, such as for near-infrared spectroscopy (NIRS). Most tests require samples to be obtained, which necessitates an enormous amount of time and manpower in large-scale screening. When the

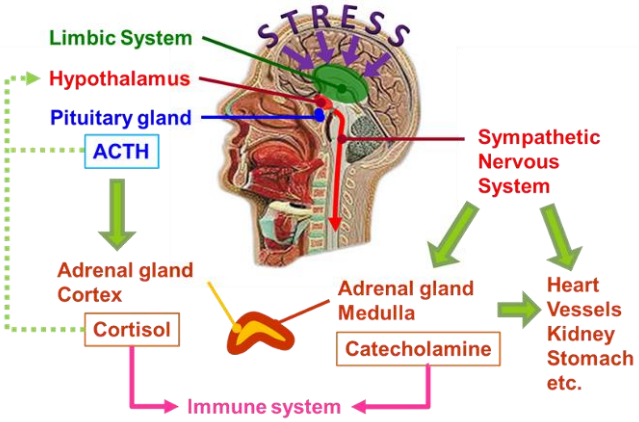


Fig. 1

procurement of reagents and equipment is taken into account, large-scale screening becomes unfeasible from a cost perspective.

Obtaining samples for biomarkers for the purpose of large-scale screening should be easy and require no large equipment or measurement-tailored reagents. We have therefore focused on the physiological indicators among the biomarkers shown in Fig. 1. Most physiological indicators change in reaction to factors other than mental stress, but only speech possesses a qualitative element. Thus, we have explored the potential of using speech in the psychiatric field and for mental stress screening.

## 2. SPEECH AND EMOTION

In a setting of constant diagnoses, it is an oft-repeated experience that the expression of emotion alters when a patient has a mental disorder brought about by stress. People are said to recognise the emotions of another person by judging their expression, voice, or gestures. However, based on the fact that people can also infer another person's emotions over the telephone, it is thought that the voice plays the greatest role in emotion recognition.

When a person is under stress, the limbic system stimulates the hypothalamus, and causes various reactions in the body through the parasympathetic nerves. Increased heart rate, rising blood pressure, and tensioning of muscles are typical, but changes in voice quality are also evident. This is because the vocal folds are innervated by the recurrent laryngeal nerve, a parasympathetic nerve which is susceptible to psychological impact, and a change in heartbeat is similarly caused by the cardiac branches of the recurrent laryngeal nerve (Fig. 2). Changes in voice quality due to psychological impact include, for instance, the voice rising when a person becomes tense, or lowering when they are sad. In interpersonal communication, such changes

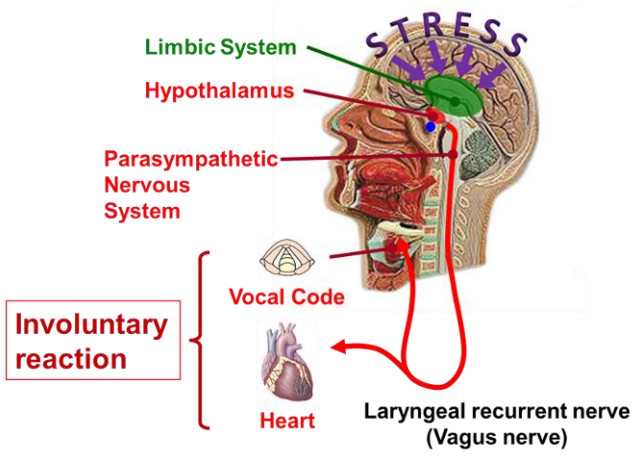


Fig. 2

can be recognised, allowing another person's emotions to be estimated. Specialist professionals such as psychiatrists and psychological counsellors can often infer the existence of psychological disorders based on past experiences. Therefore, if people can recognise qualities in the voice so as to determine emotions, they will be able to infer stress levels.

Speech contains a linguistic component that is generated by moving the pharynx and the mouth. This is an involuntary component that can be altered consciously using voluntary muscles, and an involuntary component which cannot be controlled by the individual, depending on the innervation mentioned above (Fig. 2). To avoid the reporting bias outlined in the Introduction, we must focus on the involuntary component which individuals have no awareness of and cannot control.

## 3. VOCAL EMOTION RECOGNITION TECHNOLOGY

Many human emotion recognition systems have been developed in the field of engineering [4], [5]. People use linguistic elements (spoken content and vocabulary) and prosody elements (cadence) to recognise emotion in speech. In the past, many studies have reported emotion recognition systems based on linguistic elements that use recognition dictionaries, as in speech recognition. However, because of the difficulty of supporting massive dictionaries, emotion recognition systems based on prosodic information began to be studied. Several studies have been conducted in Japan, and stress and fatigue have been evaluated using speech [6].

The problem in studying vocal emotion recognition is how to define "emotion". To date, most studies have used classification/analysis based on various concepts. These can be broadly divided into the following categories:

1. Methods that screen levels of pleasure/displeasure [7].
2. Methods that classify emotions into a number of feelings, e.g. sadness, anger, surprise, fear, disgust, contempt, joy [8].

Mitsuyoshi suggested there is a relationship between these feelings of pleasure/displeasure and various emotions, and proposed a system to indicate anger, happiness, sadness, and calm, as well as the degree of pleasure/displeasure, which signifies emotional intensity [9]. A technology to identify human emotions from prosodic information in speech has been established by Mitsuyoshi et al.

We have therefore used the vocal emotion recognition software Sensibility Technology (ST) Emotion by AGI Inc., which was developed by Mitsuyoshi et al. This device incorporates a solid fundamental frequency estimation technique and an if/then rule-base derived from a massive emotion-

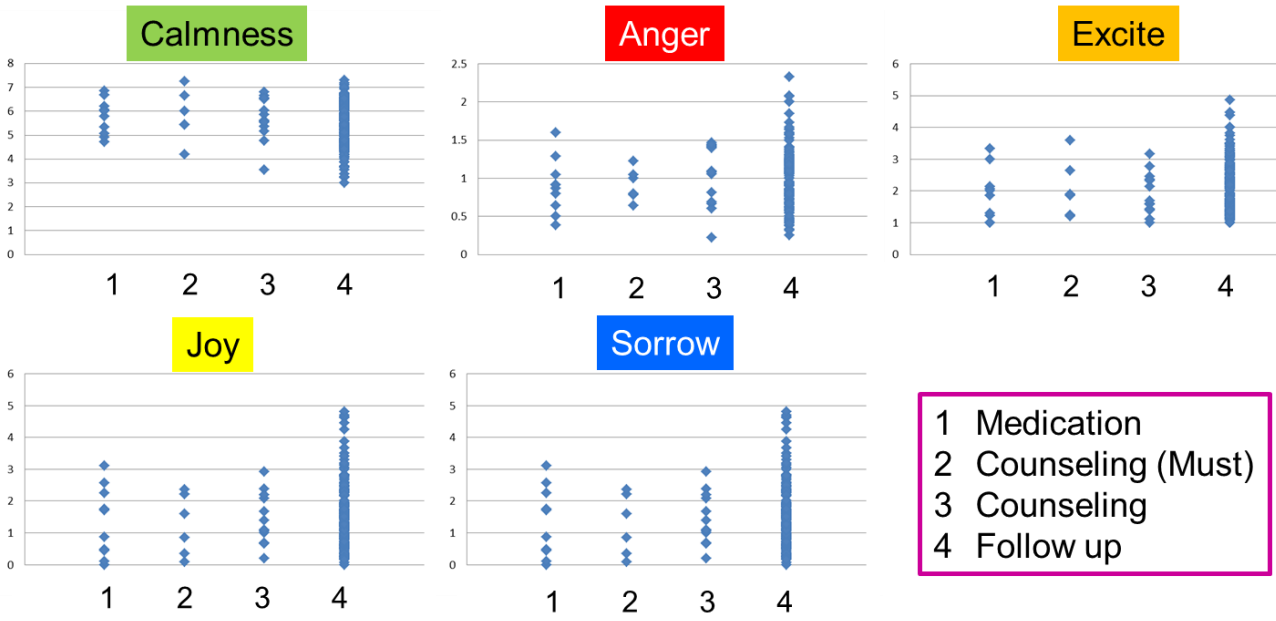


Fig. 3

labelled speech database. ST detects over 200 emotion-characteristic parameters from spoken natural utterances and, based on the utterance frequency for these parameters and their speech patterns, displays in real time the percentage of anger, happiness, sadness, and calm contained in a person's speech. Simultaneously, ST displays the emotional intensity of the speech.

Mitsuyoshi's advanced research confirms the link between the emotions obtained from speech and activity in the brain through fMRI [10]. That is, activity in the right amygdala, the prefrontal lower brain (near Brodmann area 12), and the prefrontal area, which are linked to emotion, can be confirmed by fMRI brain activity when the vocal speech recognition output is excitement or anger. Furthermore, simultaneous heartbeat measurements confirmed a causal link between brain, heartbeat, and vocal emotion recognition.

## 4. EMOTIONAL CHANGES DUE TO STRESS

In prior research, we successfully used the present technology to capture emotional changes due to stress [11]. For nine members of the Self-Defense Forces who were deployed in a disaster relief operation for the Haiti earthquake, we compared the ratio of emotions contained in their speech against the length of deployment periods. Compared to personnel who had been deployed for short periods, the speech of those who had been deployed for long periods tended to exhibit increased sadness components and decreased happiness components.

Motivated by these results, we then conducted a larger-scale survey. Self-administered mental tests (GHQ-30) and speech recordings were conducted for 444 members of the Self-Defense Forces on routine duty and 1,004 members of the Self-Defense Forces who were deployed in the Great East Japan Earthquake. Furthermore, interviews were conducted with 225 personnel. These included people who were considered to have a disorder based on the GHQ-30, and those who wished to or who were recommended by their superior. A comparison of 1,448 subjects with a cut-off score of 7 points in the GHQ-30 found no clear difference in emotions. However, increases in emotional intensity and sadness were observed in subjects deemed, based on their interview results, to have been under a level of stress that required counselling or medical intervention. For subjects who were deemed to have been suffering from stress up to a level that required medical intervention, emotional intensity was decreased, as were anger, happiness, and sadness (Fig. 3). However, although changes in emotion due to levels of stress could be captured, this survey failed to determine the intensity of stress on the basis of changes in emotion [12], [13].

## 5. DETERMINATION OF STRESS INTENSITY

Using these changes in emotion, we have now developed new software to determine stress intensity based on emotion recognition in speech. In developing this system, we divided the aforementioned data from domestically deployed personnel into two groups matched for age, gender,

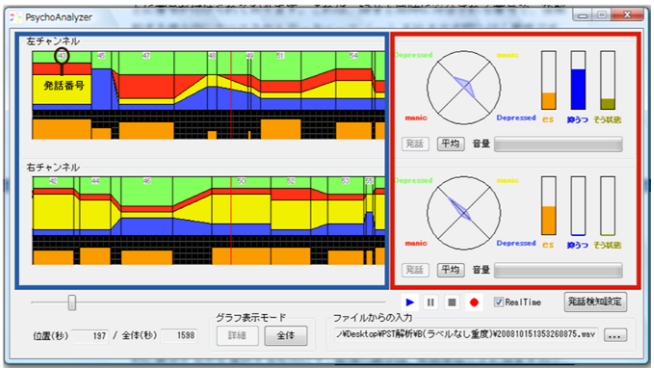


Fig. 4

and assessment. One dataset was used for training the software, and the other set was used for verification. Although an assessment rate of 85% was achieved, we did not attain an assessment accuracy that is satisfactory in a clinical setting in, for example, the high-stress cohort.

We therefore applied empirical coefficients to the parameters used in emotion recognition, and created experimental stress intensity parameters that are displayed on the PC screen in real time (Fig. 4). This was compared with the aforementioned interview results and the GHQ-30 scores. Compared to GHQ-30 screening, our software exhibited an almost satisfactory detection rate (sensitivity) for those requiring counselling or treatment, but its specificity left room for improvement. However, cases thought to have a reporting bias which could not be detected in GHQ-30 were all recognised, thus overcoming the problem of bias [12], [13] (Fig. 5). Because the parameters we created were experimental, more scientific and accurate parameters that can be used in statistical analyses need to be developed.

## 6. BUILDING A JOINT RESEARCH INFRASTRUCTURE

The development of high-accuracy parameters requires a great deal of speech data linked to medical data. For a single institution or company alone to accumulate such speech data would require an enormous amount of time, and is virtually infeasible. For instance, it would be theoretically possible to use the present technology for languages other than Japanese, but no evidence currently exists, and verification for other languages cannot reasonably be done in Japan. The same goes for applications to disorders other than stress, outlined below.

We are therefore currently building a joint research infrastructure using cloud technology to collect a variety of speech data from around the world. In summary, each institution participating in the joint research collects medical information and speech data in line with the various research topics. Once anonymised, these data are uploaded to the cloud. The team developing the parameters (i.e. us) uses these speech data from across the world to create parameters, and uploads the finished material to the cloud as an analytical tool. The joint research institutions use this analytical tool to analyse data in line with the various research topics, and produce various research results. The parameter development team accesses all speech data to create the parameters, but detailed analyses including any medical verification are performed by the respective joint research institutions. However, each institution can only access the speech data that they collected. Separate research agreements are required to be able to access speech data collected by other institutions.

We expect this technology to grow exponentially through this joint research infrastructure.

## 7. APPLICATION DEVELOPMENT

However good the developed technology, it is of no significance if it cannot be commonly used by companies. We have therefore created an application that runs on smartphones with the objective of popularising the system (Fig. 6). This application allows the level of vitality (the absence of depression) in someone's speech to be monitored without the user being particularly aware of it, simply by using the smartphone as a mobile phone.

We think that using the present application will enable users to become aware of changes in their mental state or condition at an early stage, thus encouraging changes in lifestyle or a visit to a health care facility. A large-scale field trial using this application is currently in preparation in Japan,

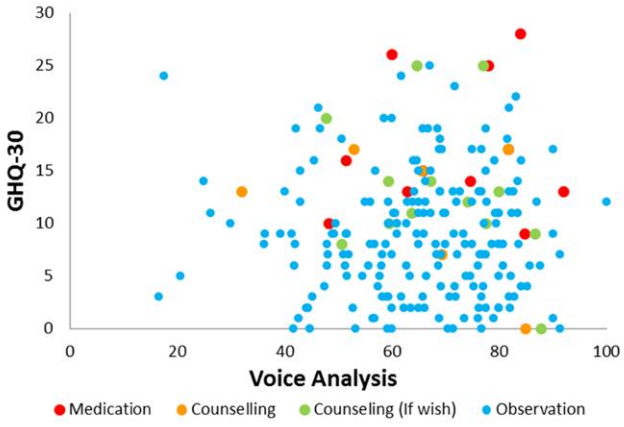


Fig. 5


Fig. 6

and we hope to commence this study in the summer of 2015.

## 8. FUTURE PROSPECTS

The present technology can be incorporated in various devices and, since its main advantages are ease of implementation and low cost, its application in preventative medicine is promising. Using the accumulated technologies to date, new speech analysis algorithms can be developed for stress evaluation or screening for mental disorders, allowing for the construction of ICT-based screening systems. This could prevent non-essential and non-urgent counselling or visits to psychiatric specialists, as well as contributing to specialist health care and prevention for clients requiring intervention in the past.

The present technology has a wide range of applications. Besides stress assessment and screening for mental disorders such as depression, we believe it could also be applied in the screening of mental disorders such as PTSD (Post Traumatic Stress Disorder) and schizophrenia, neurological disorders such as cerebral infarctions or Parkinson's disease, dementia including Alzheimer's, sleep apnoea syndrome, COPD (Chronic Obstructive Pulmonary Disease), respiratory diseases such as asthma, diabetic autonomic neuropathy, and combined disorders such as post-heart attack depression. Moreover, as it allows for continuous, real-time, and objective evaluation, we also think that speech-based emotion recognition can be used to assess treatment results for the disorders outlined above and conduct monitoring at home. We believe that our system may offer objective indicators for the development of drugs for a number of disorders which, to date, have relied on subjective evaluation, allowing for more effective drug development.

## 9. CONCLUSION

We have introduced a system for stress intensity assessment using emotion recognition technology through speech. This represents an inexpensive, high-accuracy screening method for large-scale mental health care.

A prior verification compared it with assessments using self-administered questionnaires (GHQ-30) and interview-based assessments. This showed that almost equal examination sensitivity was obtained as for the GHQ-30, and that the problem of reporting bias in self-administered surveys was overcome. However, this is still a developing technology, and further improvement is required.

The main advantages of the present technology are its ease of implementation and low cost. Its application in preventative medicine is therefore promising, and, if combined with diagnostic technologies using various biomarkers, better induction to health care specialists will be possible.

## 10. REFERENCES

[1] Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. New England Journal of Medicine, 351(1), 13-22.

[2] Perrin, M., DiGrande, L., Wheeler, K., Thorpe, L., Farfel, M., & Brackbill, R. (2007).

Differences in PTSD prevalence and associated risk factors among World Trade Center disaster rescue and recovery workers. American Journal of Psychiatry, 164(9), 1385-1394.

[3] McLay, R. N., Deal, W. E., Murphy, J. A., Center, K. B., Kolkow, T. T., & Grieger, T. A. (2008). On-the-record screenings versus anonymous surveys in reporting PTSD. The American journal of psychiatry, 165(6), 775-776.

[4] Nwe, T. L., Foo, S. W., & De Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. Speech communication, 41(4), 603-623.

[5] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. Signal Processing Magazine, IEEE, 18(1), 32-80.

[6] Shiomi, K. (2008). Voice processing technique for human cerebral activity measurement. In Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on (pp. 3343-3347). IEEE.

[7] Russell J. A., (1980) "A circumplex model of affect," Journal of Personality and Social Psychology, Vol. 39, pp. 1161–1178, 1980.

[8] Eckman P., (2003), Emotions revealed: Understanding faces and feelings, Weidenfeld & Nicolson, London, England.

[9] Mitsuyoshi S., (2006), Research on the phonetic recognition of feelings and a system for emotional physiological brain signal analysis, Ph.D. thesis, The University of Tokushima.

[10] Mitsuyoshi, S., Monnma, F., Tanaka, Y., Minami, T., Kato, M., & Murata, T. (2011). Identifying neural components of emotion in free conversation with fMRI. In Defense Science Research Conference and Expo (DSR), 1-4. IEEE.

[11] Tokuno, S., Tsumatori, G., Shono, S., Takei, E., Suzuki, G., Yamamoto, T. & Shimura, M. (2011). Usage of emotion recognition in military health care. In Defense Science Research Conference and Expo (DSR), 1-5. IEEE.

[12] Tokuno S., Mitsuyoshi S., Suzuki G., Tsumatori G. (2014). [Proc] STRESS EVALUATION BY VOICE: a novel stress evaluation technology, 9th International Conference on Early Psychosis (Tokyo).

[13] Tokuno S., Mitsuyoshi S., Suzuki G., Tsumatori G. (2014). [Proc] Stress Evaluation Using Voice Emotion Recognition Technology: A Novel Stress Evaluation Technology for Disaster Responders. XVI World Congress of Psychiatry (Madrid).

[14] Tokuno S. et.al, (2013). [Proc] Usage of Emotion Recognition in Stress Resilience Program. 40th ICMM World Congress on Military Medicine (Saudi Arabia).