DEVELOPMENT OF VERBAL ANALYSIS OF PATHOPHYSIOLOGY

Shunji_Mitsuyoshi

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

Abstract. *Since 1999, the author has been engaged in the research and development of a quantitative measurement of emotion and stress. He has built an emotion model derived from emotions and biological substances in physiological tests. He has analysed emotional speech based on fMRI and people's subjective views (their own/a third person's), and has built a system that visualises the type and intensity of emotion in real time based on voice data. This is known as Sensibility Technology (ST). This article outlines the structure of the emotion measurement technique in ST, and compares the measurements with emotion-related brain activity, as well as with "voice, emotion, and stress" on the basis of recent reporting on stress detection.*

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

1. INTRODUCTION

After the Great East Japan Earthquake, there was an increase in mental illnesses such as anxiety disorder and depression, caused by mental and physical fatigue and stress. The prevalence of suicide as a result of these mental illness issues also increased. I was requested by the Self-Defense Forces to propose countermeasures for this. A technique was sought which allowed for the simple, quantitative measurement of stress and emotional change. I therefore studied the possibility of simply measuring stress and emotional states using vocal emotion recognition technology [1].

2. VOCAL EMOTION RECOFNITION

I examined simple techniques for measuring a person's emotions. This revealed that it is relatively difficult for people with inborn impaired hearing to understand emotions based solely on facial expressions and textual information. Moreover, I also found that Japanese people have a strong tendency to suppress emotions, and that in Japan's cultural environment emotions are particularly unlikely to be apparent in a person's attitude or expression. Therefore, I decided to study techniques to identify emotions from voice, and to create language-independent vocal emotion recognition. First, the colour representation in the aforementioned emotion model (green, red, blue, yellow, orange) was used to collect people's subjectively coded speech, eliminating as much language dependency and language impact as possible. Using the collected voice samples, I built a software tool that can classify emotions automatically and in real time, similar to people's subjective views (with colour denoting the attribute, and quantity scored on a scale of 1–10). This is the vocal emotion analysis technique, known as Sensibility Technology (ST). Furthermore, I compared this against emotion-related activity in the brain and physiological responses.

2.1. Structure of emotional utterances

I studied how emotion-related activity in the brain affects vocal utterances. The following general theory is important here: [1] Processing results for emotion-related information in the amygdala are first transmitted to the hypothalamus, which is the centre for autonomic nervous function and hormone secretion, causing an autonomic nervous response that increases the heartbeat and alters gastrointestinal motion. [2] When subjected to an impulse such as fear, information is simultaneously transmitted from the amygdala to the midbrain, causing freezing behaviour. [3] Furthermore, the amygdala transmits stimuli to the limbic system including the cingulate gyrus and the hippocampus in the brain, significantly impacting long-term memory [2]. The following general theory on the causes of vocal cord disorders due to recurrent laryngeal nerve palsy in the relationship between vocal cord disorders and the recurrent laryngeal nerve is also important: [4] Disorders due to a ruptured vagus nerve to the vocal folds or the branched recurrent laryngeal nerve in traffic accidents or after surgery. [5] In nerve compression, disorders when the recurrent laryngeal nerve, which descends to near the heart then ascends, is also subject to compression through the enlargement of the heart and the nearby aorta (e.g. raised heartbeat). Disorders when the nerve is subject to compression due to e.g. malignant tumours or after intubation during general anaesthesia. [6] Disorders due to acute infections, medication, or neurological disease (including emotional disorders). [7] Other disorders for which the cause is unknown, termed "sudden (idiopathic) recurrent laryngeal nerve palsy". Unlike [4], [5], and [6], this occurs suddenly, and sometimes also heals spontaneously [3].

Based on these two general theories, namely "the emotional mechanisms of brain and nerves [1], [2], [3]" and "the link between nerves and vocal cord disorders [4], [5], [6], [7]", I hypothesised that "the vocal folds are impacted by heartbeat changes and nerve status associated with recurrent laryngeal nerve activity which is affected by emotion-related brain activity". This relation is outlined as follows.

Based on the link between voice and emotion, ST argues that "emotion-related brain activity will at least be apparent in the voice", and "through this, the state of emotion-related activity in the body and the brain is communicated through the circuitry connected to the vocal folds". A system was built by comparing ST with subjective voice data evaluation results and brain function activity, and its accuracy was verified.

matter	excite	stress	anxiety	disgust	agonism	fear	depression	Pleasure/un	stability	euphoria	expectation	hart	pupill	sympathetic	swetat	temp	bloodpress	cycle	immune
CRH	arousal	○ACTH○	◎		◎		◎				○	◎					◎	biorhythms	
NPY	calm	○CRH◎	○×	NA×															
Cortisol		○																	
VP		○homeos			◎	◎								active			◎	biorhythms	
ACTH		○			◎×											×			
CCK-4	F◎		○ ◎S			○ ◎strong	○								○				
CCK-8	A◎SF								○										
Melatonin		○×						hap	○			×		forget				season	active
endorphin	calm		NA×			NA×		pleasure		○				motion-ple					NKact
βEnd		CRH×										◎					◎		
ACh		CRH◎					◎												
NA	◎	CRH◎△	◎	◎	◎	◎	◎					◎	expan	Tension memo					
Adrenaline		CRH◎△	◎		◎	J◎						◎	expan	tention					
DA	◎	CRH◎			◎		low◎							memory					
5-HT		CRH◎	◎×		low◎	◎	◎												
Ang-		CRH◎																	
Garanin		CRH◎	×																
SRIF		CRH×																	
α−MSH		CRH×														×			
GABA		CRH×	NA×			NA×													
BZD		×	◎																
Diazepam			NA×			NA×				○									
Ethanol			NA×			NA×													
cnk																			
β-carboline			○◎																
Isoprenaline			◎																
Yohimbine			◎																
Fenfluramine			◎																
Sodium lactate			◎																
CO_2			◎strong			◎strong													
Caffeine	arousal		◎																
Galanin			× ?																
Oxytocin	×	×	× ?															matetnal	
FMRF Amide			×		◎ ?														
Testosterone			×		◎strong														
Androgen			×		◎														
Estrogen			×		×														
Progesterone					×														
Corticoid						○													

Table 1. The research of relation between Emotional response / Physical reaction and materials.
○ shows synthesis/secretion, ○× shows secretion and suppression, ◎ shows acceleration, CRH ◎ is CRH synthesis/secretion acceleration, front ◎ is acceleration in the prefrontal cortex, all ◎ is acceleration in the entire brain, front × is suppression in the prefrontal cortex, low ◎ is decrease then acceleration, × is suppression, ○ homeostatic is maintaining homeostasis through secretion, CRH × is CRH synthesis/secretion suppression, ◎× is control, △ is adjustment, NK is natural killer cells, ? is reported information, - is not indicated in literature [9]. As seen in the Table, the many '-' signs show that many of the relations among the mind, feelings, and secretory substances remain unknown. The CCK system, 5-HT, GABA, and DA interact in complex ways and seem to act to emotions. Additionally, sex hormones seem to have a close relation with attack. These secretions and substances are controlled by cranial nerve activity, and the limbic system, emotions, and memory work together intimately

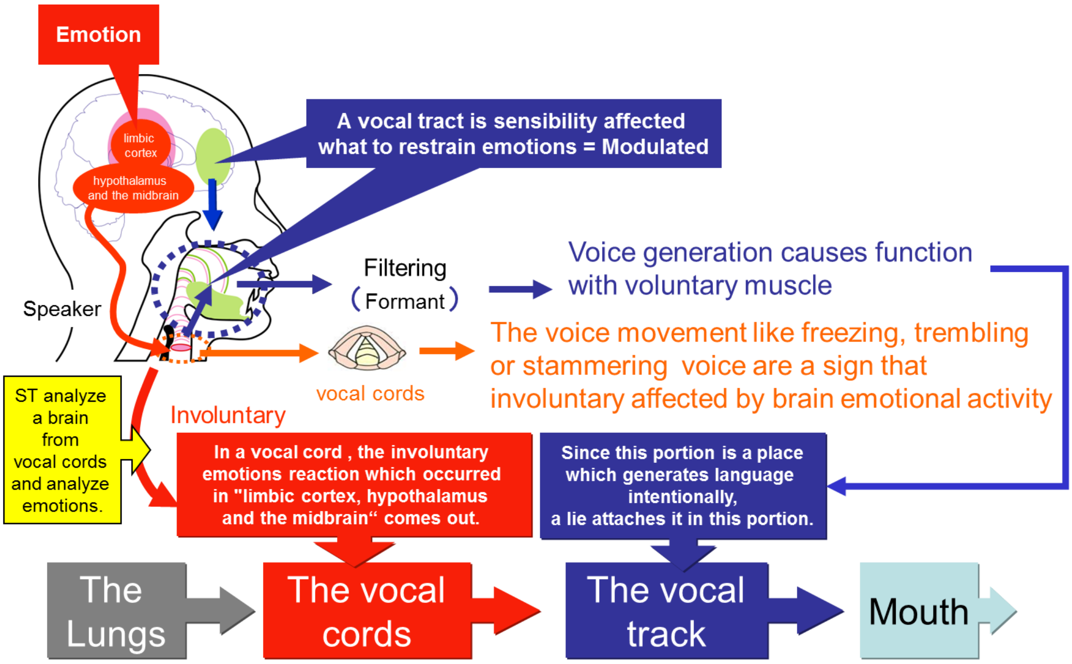

Fig. 1 Body structure: A short term effect on brain emotion activity. Human has a strong emotion, it is usually under control of intelligence. This is the human emotion. Usually human modulate emotions well on social life.

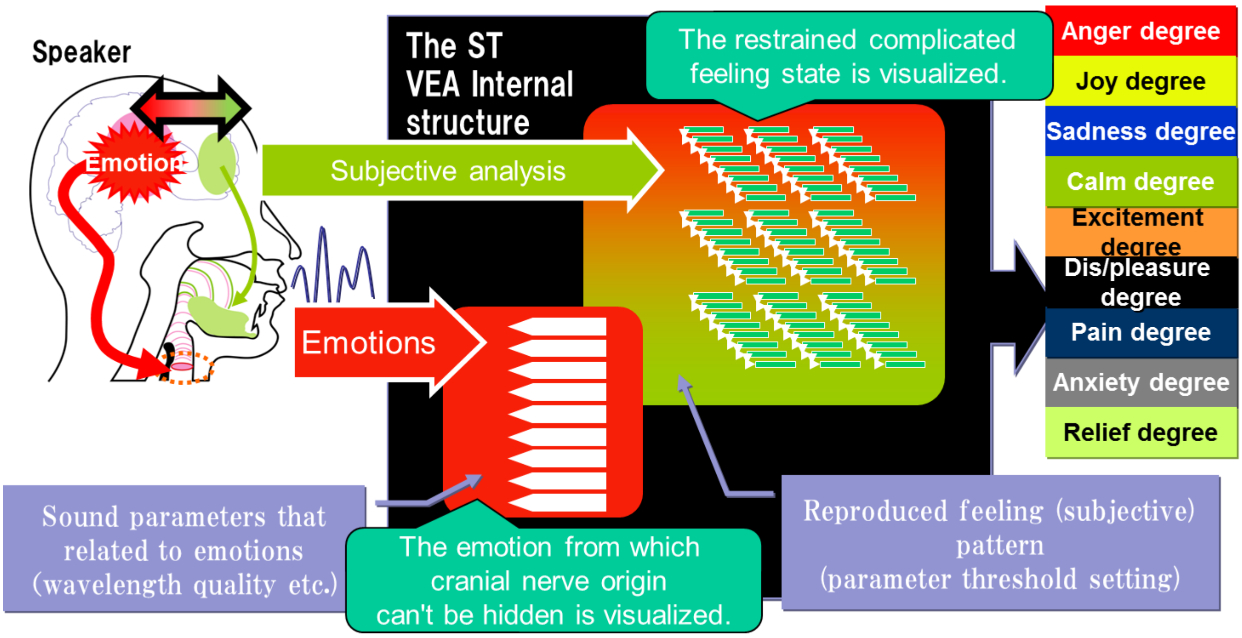

Fig. 2 ST structure and its output: Emotions and respiration are both related. VEA is enable to analyze emotions at utterance interval between a breath to a breath. The strongest emotion will be selected by majority decision on judgment logic and display as classified colors and strength. An other emotions will also list following the first chosen.

2.2. Technique for measuring emotion in the voice

Based on the aforementioned structural link between emotion-related brain activity and utterances, I created ST, which was structured into acoustic parameters relating to emotion (acoustic elements such as prosody) and people's subjective reproduction logic (set of parameter thresholds for the assessment of emotion based on acoustic parameters, similarly to people's subjective views). The structural diagram for this is as Fig. 2.

Because two assessment techniques have been obtained, i.e. acoustic parameters and subjective reproduction logic, this technique allows for a comparison of biological emotional responses and neurological disorders against acoustic parameters, and enables long-term dispositions to be compared with psychogenic depressive states using subjective reproduction logic.

2.3. Technique for measuring emotion in the voice

An experiment was performed to ascertain whether ST is consistent with people's own subjective views.

Standard evaluation: The utterer listens to their speech after uttering, and their evaluation of their emotional state is set as the standard.

Experimental method: Subjects are asked to listen to the standard voice data and provide intuitive, subjective evaluations of utterance units in a randomised order. The concordance rates with the standard voice data evaluations are checked.

Subjects: a) Twenty-six utterers (re-evaluated after some time had elapsed); b) six third-person evaluators, also Japanese speakers; c) six third-person evaluators, foreigners who do not speak Japanese; d) ST evaluation results for which emotions were measured from voice data.

Results: a) 73%, b) 60%, c) 55%, and d) 70%.

As a) and d) are approximately the same, the results suggest that ST is broadly equivalent to the utterers' subjective views. However, since the utterers' re-evaluation of the standard speech after some time had elapsed resulted in close to 30% inconsistency, it is difficult to study emotion recognition rates in comparative experiments for subjective views alone. This is because Japanese speech has little undulation in its intonation, and is rather flat. Additionally, the subjective evaluation of speech samples which had been segmented into utterance units and presented in a randomised order was difficult for the speakers themselves.

2.4. ST and emotion-related brain activity

A performance evaluation of the reproduction of the subjective views of the utterers themselves using ST revealed a number of obstacles, such as changes in subjective views due to complex cognitive influences and the passing of time, and Japanese speech characteristics such as flatness. A performance evaluation exceeding 70% accuracy turned out to be impossible. It was therefore necessary to conduct a comparison with biological responses, which are the origin of emotions. The results from a comparative experiment with emotion-related brain activity are described below.

2.4.1. Study of emotion-related brain communication

This experiment used 3T fMRI (Siemens) to measure brain function. The provision of this machine and the experimental work formed part of a three-year scheme funded by the National Institute of Information and Communications Technology (NICT).

2.4.2. Measurements of emotion-related brain activity

It is well known that MRI creates 130 classes of digital noise when running, and it is impossible to obtain speech without noise inside an MRI gantry. Using a mouth-covering dust mask to improve the seal, I manufactured a mask using fibre-reinforced plastic (FRP). By plugging silicone of high specific gravity around the mouth, noise levels were cut and the noise problem was solved. Next, to prevent suffocation by this sealed mask, and to prevent muffling the sound in this narrow space, a thick sound-proof silicone hose was manufactured. By connecting this to the mask, I used the air convection (created by circulating fresh air using an air compressor and working the shape of the mask) to prevent muffled sound and echoes, thus allowing for clean MRI speech recordings. For the experiment, the utterer's head was secured in the MRI machine. A total of six subjects took part, and each had a conversation partner (a close relative) that was familiar with episodes that would evoke emotion in the subject. All conversations were recorded, and measurements of brain activity, heartbeat, eye blinking, body temperature, blood pressure, and eye movement were recorded. Apart from brain activity, no significant physiological differences were observed. Thus, this paper describes the comparison of ST with brain activity.

2.4.3. Comparison with emotion-related brain activity

Experiment: using the aforementioned technique, I examined the difference between brain activity during conversation when ST could confirm negative emotion, and when it could not (using t-test, critical region for both 0.1%, non-corrected). The focus was on negative emotion-related brain activity, because at the time there were extremely few study outcomes for brain activity related to positive emotion.

Results: The most frequent components of conversation in the ST analysis during brain activity measurements were excitement and tranquillity. Conversation components were extracted in the fMRI analysis, firstly through block design.

Results showed activation of Brodmann areas 4, 6 (subject TO, $p < 0.001$), and 46 (subject TN, $p < 0.001$). Brodmann area 46 is the so-called Brodmann language area.

When the subjects were in conversation, activation was observed in the left prefrontal area (DLPFC, Brodmann areas 45, 47) and the base of the frontal lobe (Brodmann area 10) (subjects TO, SM). In subject SM, activity was also seen in the amygdala. In the ST analysis, times of excitement and times of tranquillity produced different activity in the left angular gyrus (Brodmann area 39, Wernicke's area) [4].

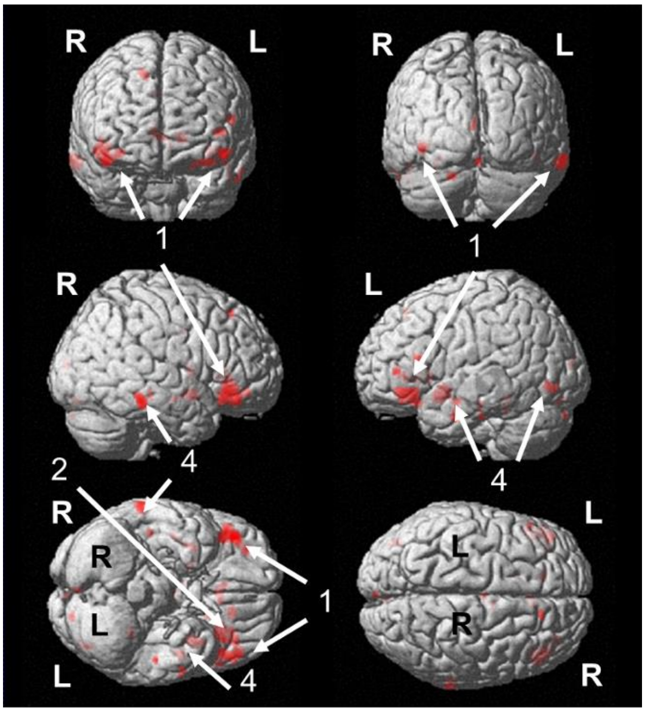

Fig.3 This is a proof which ST had judged to be a negative emotion only when brain activity was a negative emotion.

1. Left and right frontal BA44,45 - L:speech generation, R: sympathy
2. Left dorsofrontal BA12 - hormonal control, suppressive emotion
3. Left amygdalate complex - negative emotion
4. L/R inferotemporal - visual imagery, body image

3. ST AND STRESS

It is well known that mental and physical stress is related to depression and anxiety disorders. In collaboration with the National Defense Medical College, we verified whether ST was superior in detecting stress from states of emotional change.

3.1. Voice and stress comparison

Having established that the overseas deployment corps of the Self-Defense Forces suffers the maximum mental and physical stress, a comparative verification was conducted for ST and stress. Stress levels were established by the period of local deployment. A comparison of long-term and short-term deployment cohorts showed a tendency for the happiness component to decrease and the sadness component to increase. Furthermore, no changes were seen in the anger and tranquillity components. This suggests that it is possible to detect stress from states of emotional change in ST [5].

4. MEASUREING STRESS THROUGH VOICE

A joint research project with the Faculty of Engineering, University of Tokyo [6], has seen the construction of a subjective reproduction logic, and tests have been performed for large-scale subjective evaluations of the basic feelings of pleasure/displeasure and six emotions. The results suggest that, as stated in Section 2.4, it is not possible to achieve accuracy in excess of 70–80%, which is the limit of subjective evaluation by the utterers themselves. This stands to reason, as it is the limit of people's subjective view. However, the addition of "pleasure/displeasure, pain, anxiety, tranquillity" to "anger, happiness, sadness, excitement, tranquillity" makes the ST structure robust. On the other hand, solving the noise issue enabled the successful acquisition of voice data in the difficult fMRI gantry environment. Additionally, results concerning the consistency of emotion-related brain activity and ST showed that solid brain activity measurements can be obtained for at least the excitatory response, and brain function mapping for individual emotional responses (e.g. anger) in line with subjects' perceived intensity has also become possible. Furthermore, in collaborative research with the National Defense Medical College, a comparison of ST and stress response voice data suggested the possibility of detecting stress from states of emotional change in ST, and voice data samples could be collected for the purpose of larger-scale verification and improving accuracy. A collaborative project involving a large-scale field trial with members of the general public is scheduled to take place, whereby the mental status of utterers will be routinely estimated. Deviations in the frequency of emotions over the long term will be detected, with those trends taken to denote a person's level of health. For this, a robust ST will be compared with people's subjective views, and changes in emotion, which are an important element of one's mental state, will be analysed based on voice data. This has the potential to play a supporting role in psychoanalysis and diagnosis by physicians, based on early detection of pathology.

5. CONCLUSION

ST has become the current medical technology of PST (Pathologic condition analysis and Sensibility Technology), and a Social Collaboration Lecture titled "Verbal analysis of pathophysiology" has been established at the Faculty of Medicine, University of Tokyo, by MAZDA and MKI. Softbank's emotional robot Pepper is also used. Regarding preventative medicine, Finland, Romania, the US armed forces, and Japanese medical institutions, as well as the Japanese government, have expressed strong interest. In Japan, it has been designated as a National Strategic Special Zone technology, and the University of Tokyo and the Ministry of Education, Culture, Sports, Science and Technology have designated it as advanced and innovative research. However, even if the significance of stress in emotional change has become apparent, classifying its causes (psychogenetic, neurogenic, physical) is difficult based on subjective views. Therefore, if, in physician-led intervention experiments, the link between acoustic parameters and neural/biological responses can be derived, and if neurogenic and physical symptoms and pathology can be defined, there is a possibility of pathological analysis and simple and routine visualization of brain and neural activity in real time, as "X-rays of the mind".

Acknowledgements

For this study, the author has combined a doctoral thesis and research outcomes from AGI Inc. I received guidance throughout the latter part of my Doctoral program from Professor Fuji Ren at Tokushima University. Furthermore, I was given the opportunity to collaborate with Associate Professor Shinichi Tokuno at the University of Tokyo, which contributed greatly to the collection of voice data from members of the Self-Defense Forces. Emeritus Professor Schein of MIT in the US also provided guidance on psychology, and offered me the opportunity to study with various scientists at Stanford University, deepening my understanding of emotion.

Part of the present study and the pathology analysis study was funded by the National Institute of Information and Communications Technology (NICT), and by a grant for the promotion of defense medical research by the Ministry of Defense, a Grant-in-aid for Scientific Research, and a MAZDA and MKI Social Collaboration Lecture.

Terminology

Utterance unit: One utterance unit is continued speech uttered in one breath, and indicates speech uttered between consecutive breaths.

6. REFERENCES

[1] 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.

[2] http://www.brain-mind.jp/newsletter/04/story.html

[3] http://homepage1.nifty.com/jibiaka50/seitaimahi.htm

[4] 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.

[5] 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.

[6]Shuzo M., Yamamoto Y., Shimura M., Monnma, F., Mitsuyoshi S., Yamada I, Construction of Natural Voice Database for Analysis of Emotion and Feeling. IPSJ Journal. 52(3).2011:1185-1194.[Japanese]