

The discriminant analysis of the voice expression of emotion - Focus on the nursing experience -

Yahan. Gao, Yuko. Ohno, Feng. Qian, ZhaoYan. Hu, ZhiXiong. Wang

Abstract—the characteristics of emotional expression in everyday life vary significantly depending on one's life experiences, role in society, and other factors, especially special occupation experiences. We focus on the nursing experience. The purpose of this study is to establish a method to discriminate voice of nurse's emotion. We performed the pronunciation experiment at first and through the sound analysis, extracted the effective speech parameters. For make clear of the differences of these parameters between emotions and discriminate emotions, we performed the multiple comparisons and discriminant analysis. In the result, found that the speech parameters different between emotions and these affected parameters are different at two groups based on nursing experience. In four emotions of our study, the Sadness is easy discriminant emotion; the Joy is most difficult to discriminate. Through the discriminant analysis, found two stable speech parameters.

Keywords: emotion, nursing experience, discriminant analysis, voice, speech parameter, healthcare information, nursing care management

I. INTRODUCTION

The emotion is a mystery since ancient time, and a lot of problems still not clear yet. Although there is a lot of discussion about the emotion, but it is still not unified in the opinions and definition. It is caused by the differences in the research standpoints and methods. Some researchers put forward that the emotion has two aspects: state and trait. 1-2) It is differentiated from the duration of emotion. The state is short time experience. The trait is personal stable property that if easy to experience some emotion of effected by long time experience or not. This study focused on the voice feature of how to performing emotion to other person and outside world, analyze the short time a transient state. About the emotion category, we adopts the six basic emotion categories (Joy, Surprise, Angry, Sadness, Disgust and Fear) proposed by Ekman 3). That six basic emotions is considered common to all mankind, has been widely used in many studies. But due to the nurse's professional characteristics, they are rarely performed disgust and fear to their patient. So, we considered four basic emotional states: Joy, Sadness, Angry, and Surprise in our study, and added neutral as a control state.

The characteristics of emotional expression in everyday life vary significantly depending on one's life experiences, role in society, and other factors. 4) This suggests to us that take into

account of occupational or societal roles of the participants are important when analyzing the features of speech parameters in emotional expression. Given the nature of their occupation, nurses must interact with all kinds of people from various walks of life and in different states of health, so they must be skilled communicators. In other words, it is an occupational requirement for nurses to exercise discipline in their speech representation. Here we noticed on the relationship between emotional expression and speech representation by focusing specifically on the practical work experience of nurses.

Connections between speech and emotion have been extensively studied in recent years, and previous research has shown that the fundamental frequency F0 is affected by changes in emotion 5), and the fundamental frequency F0 and sound pressure differ for different emotions 6). While many of these studies have compared differences in the fundamental frequency for different emotions in speech, no one has investigated speech parameters of emotions varying from a neutral state. We assume that the change from neutral for each emotion an individual feels is related to the person's ability to express emotion, and this study will analyze the change in the fundamental frequency from neutral for a range of emotions.

The fantasy of human about robots can be traced back to 3000 years ago, but the arrival of the first industry robot was just a few years ago (1959). The development of information technology has brought many possibilities to robot technology. From complete some "fixed, simple" operation in the previous industry, the robot technology is used in many fields at now. Such as military, service, entertainment, healthcare, etc. However, the technology in the recognition and expression of emotion is still in exploring stage. Emotional expressions of nurses have their own features, so we consider this in our study, aimed to build a communication system of robot in hospital that could take a communication with patient, understanding patient's emotion and expressing the emotions like a real nurse.

In this study, we first performed the pronunciation experiment, and through the sound analysis extracted the effective speech parameters. Then, performed the multiple comparisons on each speech parameters across all emotions before and then performed the discriminant analysis between emotions.

II. MATERIALS AND METHOD

A. Subjects

The purpose of the experiment was illustrated before this. After being briefed on the understanding that the participants would not be identified or compared, and other experimental conditions, 22 candidates approved of the study's objectives

Yahan Gao is with the Shanghai Medical Instrumentation College, China (phone: 65485551-3377; e-mail: gaoyh@smic.edu.cn).

Yuko Ohno is with Osaka university University, Japan.(e-mail: ohno@sahs.med.osaka-u.ac.jp).

Feng Qian, ZhaoYan Hu, ZhiXiong Wang is with the Shanghai Medical Instrumentation College, China(phone:65485551-3377;e-mail:qianf@smic.edu.cn ,huz@smic.edu.cn ,wangzx@smic.edu.cn).

and volunteered to participate. The speech parameters for men and for women are quite different 7), so only women subjects were used in this study. We collected basic data on the participants prior to the experiment including nursing work experience, age, and family structure.

B. Experimental Conditions

1) Speech Samples

Most past studies have used actual words for speech samples 5), but the associative implications of real words can affect the emotional content 8), so we sought to eliminate this effect to make sure that we were measuring the true expression of emotion. To eliminate any potential bias from the associative implications of an actual word, we created a nonsensical word to compare parameters among the emotions studied. To create the speech sample, we arranged 50 sounds in random order to create eight syllables word, then created the meaningless words and selected "ko-ne-yo-chi-yu-e-ho-te" that were evaluated as the easiest to pronounce by two examinees.

Note that most past studies have used actual words for speech samples 5), but the associative implications of real words can affect the emotional content. If you use an actual word or phrase, it is really impossible to tell if the expressed effect derives from the cadence of word or from the meaning of the word 9), and this concern led us to adopt a meaningless word for the speech sample in this study.

2) Measurement

Recording was done in a quiet room used an ICR-PS1000M digital sound recorder (Sanyo Audio), and the microphone maintained at a distances of 4-6 cm from the subject's mouths for compare sound pressure. The recording procedure was as follows. First we showed the subjects a visual cue card with a facial expression depicting an emotional state and a written character corresponding to the emotional state. Next, we presented a card showing the meaningless speech sample word "ko-ne-yo-chi-yu-e-ho-te," and had the subjects read the word (taking as much time as they wanted) expressing the emotion presented while we recorded the speech. Each emotion was recorded three times, and the emotions were presented in the following order: neutral(3times), Joy(3times), neutral(1times), Sadness(3times), neutral(1times), Angry(3times), neutral(1times), and Surprise(3times). In order to reset or return to a normal state between emotions, we had the subjects read the made-up speech sample word one time while showing them the neutral emotion cue card. Because it is necessary to acquire the sound data of the neutral, we record this three times in the first of the measurement and used only the third one. This is similar to the other emotions. We concluded that the third recorded reading was a more stable expression of emotion than the previous two, so we discarded the first two samples and used only the third recorded speech samples for the emotion analysis to follow.

3) Sound Analysis

Features of speech emotion are said to derive from prosodic changes¹⁰), so here we defined four speech feature parameters — speech duration change, frequency change, sound pressure change, and speech variations—and calculate the values of each parameter. For the speech duration change we measured the total time (TT) needed to speak the

eight-syllable sample word for each emotion. In pronouncing the word, a natural break occurs after the first four syllables, so we measured the first half time (FT) to say the first four syllables and the last half time (LT) to speak the last four syllables.

Connections between speech and emotion have been extensively studied in recent years. Previous research has shown that the fundamental frequency F0 is affected by changes in emotion 5), and the fundamental frequency F0 and sound pressure differ for different emotions 7). In this study, the fundamental frequency Hz (F0) is thought to accurately reflect the distinctive features of an individual's speech 9), so we included F0 as one of the feature parameters in the study. We investigated the max sound pressure frequency in Hz (Fmax) to determine how the frequency changes when the voice was raised to its highest level in a state of emotion. For the sound pressure parameter, we measured the average sound pressure in dB (AvedB) for each emotion, and the max sound pressure in dB (MaxdB) when the voice was raised to its highest level in a state of emotion.

Although many of previous studies have compared differences in the fundamental frequency for different emotions in speech, no one has investigated speech parameters of emotions while varying from a neutral state. It is supposed that the change from neutral for each emotion of an individual feels is related to the person's ability to express that emotion. So this study analyzed the change in the fundamental frequency from neutral for a range of emotions. In order to measure the speech variations parameter for each emotion, we took the value of speech parameter for the neutral phonation as standard, calculated the ratio of each emotion to the standard for the frequency and phonation time, and then calculated the differences of each emotion from the standard for the sound pressure. The analysis of the speech samples was done using Audacity 1.3.5.

III. STATISTIC ANALYSIS AND RESULT

We divided the participants into two groups based on the whether they have nursing experience. We performed multiple comparisons on each speech parameters across all emotions before, then performed the discriminant analysis between emotions and compared the result of discriminant analysis in two groups.

A. The Multiple Comparisons

It is necessary to make clear of that each parameter is how different between the emotions prior to the discriminant analysis. We adopt the Correspondence Discriminant Analysis Holm Method, and Statistic analysis soft ware is R version 2.15.2. We performed this analysis in all data at first to make clear each parameter is how different between the emotions in common (women). The characteristics of emotional expression in everyday life vary significantly depending on one's life experiences, role in society, and other factors. 4) So, we performed this in two groups based on the whether they have nursing experience. Although in this way data will be reduced and some characteristics became not obvious, but we can confirm the feature of two groups more clearly. The following Table 1 to 3 show the result of these analyses, statistically significant items are presented.

TABLE I. THE MULTIPLE COMPARISONS ON EACH PARAMETER BETWEEN EMOTIONS

Emotion1: Emotion2	Speech Parameter	Mean Value		p-value	
		Emotion1	Emotion2		
Joy: Angry	MaxdB _X -AvedB _X	0.10±0.78	-0.30±0.87	2.61E-02	
Joy: Sadness	F0 _X /F0 _{neutral}	0.27±0.58	-0.77±0.39	2.50E-06	
Joy: Surprise	AvedB _X -AvedB _{neutral}	-0.25±0.58	0.37±1.25	4.50E-02	
Sadness: Angry	F0 _X /F0 _{neutral}	-0.77±0.39	-0.17±0.93	2.50E-02	
	F _X /F _{neutral}	-0.44±0.54	0.43±1.28	1.90E-02	
	F _X /F0 _X	-0.26±0.62	0.60±1.38	3.76E-02	
	F _X /F0 _{neutral}	-0.49±0.58	0.50±1.31	1.10E-02	
	FT _X /FT _{neutral}	0.56±0.76	-0.21±0.99	4.80E-04	
	AT _X /AT _{neutral}	0.37±0.73	-0.11±1.06	4.30E-02	
	TT _X /TT _{neutral}	0.48±0.73	-0.16±1.02	4.78E-03	
	MaxdB _X -MaxdB _{neutral}	0.41±1.32	-0.36±0.70	3.40E-03	
	MaxdB _X -AvedB _X	0.39±1.08	-0.30±0.87	5.60E-05	
Sadness: Surprise	MaxdB _X -AvedB _{neutral}	0.25±1.21	-0.23±0.95	3.40E-03	
	F0 _X /F0 _{neutral}	-0.77±0.39	0.67±1.26	8.60E-05	
	FT _X /FT _{neutral}	0.56±0.76	-0.44±0.96	1.10E-07	
	AT _X /AT _{neutral}	0.37±0.73	-0.19±1.12	2.10E-02	
	TT _X /TT _{neutral}	0.48±0.73	-0.30±1.07	1.90E-04	
	AvedB _X -AvedB _{neutral}	-0.27±0.93	0.37±1.25	3.90E-02	
	MaxdB _X -AvedB _X	0.39±1.08	-0.19±1.14	7.50E-03	
	Angry: Surprise	F0 _X /F0 _{neutral}	-0.17±0.93	0.67±1.26	2.50E-02
		F _X /F0 _X	0.60±1.38	-0.37±0.39	5.50E-03

The result of multiple comparisons on each parameter between four emotions, information from Table 1:

- Each speech parameter between emotions has obvious differences. F0 and sound pressure between many emotions appeared obvious differences.
- Joy and other emotion's differences are mainly appeared on the changes of sound pressure and frequency items.
- Sadness and other emotions are obviously different at speech speed.
- The change of F0: the voice of Joy is higher than Sadness. The voice of Surprise higher than Angry. But difference between Joy and Angry are not obvious. The change of F0 in word: Angry larger than Sadness and Surprise. The change of sound pressure: Surprise larger than Joy, Angry smaller than Angry, Surprise larger than Sadness. The change of sound pressure in word: Angry larger than Joy. Sadness larger than Angry. Sadness is larger than Surprise. This is different from our usual thought. It is due to the characteristics of our data, because part of the participants has nursing experience. So, we need to analyze this in two groups. The analysis results of two groups show to Table2 and 3.

TABLE II. THE MULTIPLE COMPARISONS ON EACH PARAMETER BETWEEN EMOTIONS, THE GROUP OF PARTICIPANTS WHO HAVE NURSING EXPERIENCE

Emotion1: Emotion2	Speech Parameter	Mean Value		P-value
		Emotion1	Emotion2	
Joy: Sadness	F0 _X /F0 _{neutral}	0.19±0.44	-0.66±0.39	0.0047
Sadness: Angry	MaxdB _X -MaxdB _{neutral}	0.64±1.22	-0.36±0.71	0.0057
	MaxdB _X -AvedB _X	0.59±1.15	-0.33±0.76	0.0020
	MaxdB _X -AvedB _{neutral}	0.38±1.10	-0.22±1.01	0.0057
Sadness: Surprise	FT _X /FT _{neutral}	0.48±0.67	-0.33±0.99	0.0005
	MaxdB _X -AvedB _X	0.59±1.15	-0.29±0.94	0.0350

The result of multiple comparisons on each parameter between emotions in two groups, compare the Table 3 and 4:

- The group of participants who have nursing experience is mainly different at sound pressure items. In this group, sound pressures in word of the Sadness, Angry and Surprise are different. Obviously, when expressing the Angry and Surprise, they control the voice calm down, and this is affecting the analysis result of total data.
- The group of participants who don't have nursing experience is mainly different at speech speed.
- Relation of Joy and Sadness in two groups is same.

TABLE III. THE MULTIPLE COMPARISONS ON EACH PARAMETER BETWEEN EMOTIONS, THE GROUP OF PARTICIPANTS WHO DON'T HAVE NURSING EXPERIENCE

Emotion1: Emotion2	Speech Parameter	Mean Value		p-value
		Emotion1	Emotion2	
Joy: Sadness	F0 _X /F0 _{neutral}	0.40±0.71	-0.93±0.40	0.002
Joy: Angry	FT _X /FT _{neutral}	0.34±0.83	-0.41±0.67	0.013
Joy: Surprise	FT _X /FT _{neutral}	0.34±0.83	-0.64±0.96	0.015
Sadness: Angry	FT _X /FT _{neutral}	0.72±0.96	-0.41±0.67	0.015
Sadness: Surprise	F0 _X /F0 _{neutral}	-0.93±0.40	0.86±0.74	0.001
	FT _X /FT _{neutral}	0.72±0.96	-0.64±0.96	0.001
	AT _X /AT _{neutral}	0.48±0.80	-0.23±0.96	0.022
	TT _X /TT _{neutral}	0.65±0.88	-0.44±0.86	0.001
Angry: Surprise	F0 _X /F0 _{neutral}	-0.33±1.02	0.86±0.74	0.0093

From the above results, we know following information:

- The way of expressing emotions are different between the two groups.
- The group of participants who have nursing experience is control their voice's loudness, height and speech speed to expressing emotion.
- The group of participants who don't have nursing experience is control their voice's height and speech speed to expressing emotion.
- Due to the characteristics of the occupation, the nurses may be pressing the two kinds of emotion Angry and Surprise in her daily work.

B. The Discriminatory Analysis

The purpose of the discriminant analysis is to establish a classification rule that constituted by numerical indicators. Then apply this classification rule to unknown samples. We use step wise discriminant analysis, discriminant distance adopt the mahalanobis distance. Because the mahalanobis distance is considering the relationship between independent variables. This distance is used in general discriminant. Step wise method will eliminate the insignificant parameters, analyze the significant parameters, and then establish the discriminant function. Significance level is 0.05.

From the Table4, we through the discriminant analysis found the effective parameters of F0_X/F0_{neutral} and F_X/F0_X. Fundamental frequency items are more stable than other items. From the mahalanobis distance, we found the most similar emotions, these are Joy and Surprise; the most different emotions are Sadness and Surprise. When distance is far, the

error rate is become lower. The lowest error rate appears at discriminant analysis of Sadness and Surprise, the highest error rate appears at discriminant analysis of Joy and Surprise. All error rates are lower than their baseline of 50%; it means these functions are effective. But the error rates are still higher than we need.

TABLE IV. THE RESULT OF DISCRIMINANT ANALYSIS, ABOUT EFFECTIVE PARAMETERS, DISTANCE, AND EEROR RATE.

Emotion1: Emotion2	Standardized discriminant coefficient		Constant	Mahalanobis distance	Error rate
	$F0_x/F0_{neutral}$	$F_x/F0_x$			
Angry: Joy	-0.630	0.672	-0.177	0.821	34.1%
Angry: Sadness	0.764	0.930	0.200	1.122	28.7%
Angry: Surprise	-1.181	1.122	0.161	1.444	23.5%
Joy: Sadness	1.394	0.258	0.377	1.238	26.8%
Joy: Surprise	-0.550	0.450	0.338	0.626	37.7%
Sadness: Surprise	-1.944	0.192	-0.039	1.682	20%

We also get some interesting results, when compare the discriminant rate between groups:

- Sadness belongs to easy to discriminant emotion. Expressing of this emotion in two groups is same. (The discriminant rate is 81.8%)
- Secondly easy to discriminated emotion is Surprise (The discriminant rate is 59.1%). But this is easy to be mistaken for Sadness in group of have nursing experience (The mistaken rate is 41.6%). Group of who didn't have nursing experience is good discriminated than group of who have nursing experience (The discriminant rates are 70%:50%). Next one is Angry (The discriminant rate is 40.9%); this emotion is same to Surprise, easy to be mistaken for Sadness in group of who have nursing experience (The mistaken rate is 41.6%).
- Most bad discriminated emotion is Joy (The discriminant rate is 9.1%); in group of who have nursing experiences is easy to mistaken for Sadness and Angry (The mistaken rate is 33.3%).. In group of who didn't have nursing experience is easy to mistaken for Surprise (The mistaken rate is 60%).

IV. CONCLUSION

In our study, we set the speech parameters that is didn't use in previous study yet. At first, we performed the multiple comparisons between the emotions: found the speech parameters different between emotions and these effective parameters are different at two groups based on nursing experience. The two groups are very different at expressing emotion through voice. For instance, the people who have nursing experience is mainly control their loudness of the voice (sound pressure) to expressing their emotions, the people who didn't have nursing experience is mainly control their speech speed to expressing their emotions. Then we performed the discriminant analysis to classificate the emotions: found Sadness is easy discriminant emotion, Joy is most difficult to discriminate. The way of express Sadness is similar between two groups. The way of express Joy, Angry

and Surprise is different between two groups. In the group who has nursing experience, these three emotions are easy to mistaken for Sadness. We think this result is due to they are used to control the emotions in her work, void to stimulate patient.

In this study, we investigated speech parameters of emotions varying from a neutral state. We assume that the change from neutral for each emotion an individual feels is related to the person's ability to express emotion, and analyzed the change in each speech parameter from neutral for a range of emotions and trough the discriminant analysis found more stable parameters: $F0_x/F0_{neutral}$ and $F_x/F0_x$. Although the error rates of discriminant analysis are 20%-37.7%, lower from their baselines, it means these functions are effective, but still need to reduce this. We need to consider the method to improve efficiency, enhance the effective number of parameters. In this study we considered a lot of parameters, but in establishing functions, many have been eliminated. Although the fundamental frequency items are more stable parameters, but the sound pressure and speech speed items showed more personal performance characteristics, all these speech parameters are important to our study. We need to consider a method to reflect all these parameters effectively in our subsequent research.

Through this study, we know that emotional expressions of who have nursing experiences are different with who didn't have nursing experiences. And these differences could using at our study for future goal of robot communication system in hospital.

REFERENCES

- [1] Scherer, K.R. "Vocal affect expression: a review and model for future research. ", *Psychological Bulletin*, Vol.99, pp.143-165. (1986).
- [2] Scherer, K.R. "Vocal communication of emotion: A review of research paradigms." *Speech Communication*, Vol.40, pp.227-256. 2003.
- [3] Ekman, P. "An argument for basic emotion", *Cognition and Emotion*, Vol6 (3/4), pp.169-200.1992.
- [4] Yahan Gao, Yuko Ohno, Shima Okada. "Extraction of speech parameters relating to the characteristics of emotional expression: focus on the attentiveness-related parameters such as the nursing experience." *Japanese Journal of Applied IT Healthcare*. Vol.5, No.2, pp.114-123. 2010.
- [5] Wendt, B., & Scheich, H. "The Magdeburger Prosodie-Korpus." In: *Proceedings of the Speech Prosody*. April, 699-701, 2002.
- [6] Pereira, C., "Dimensions of emotional meaning in speech", *ISCA Workshop on Speech and Emotion*. Vol1, 25-28, 2000.
- [7] Shoichi Takeda, Ayaka Tochitani, Yasuki Hashizawa et al. "Analysis of prosodic feature of guidance speech with emotional expressions". *The Acoustical Society of Japan*. Vol60 (11), 629-638, 2004.
- [8] Pihan H, Tabert M, Assuras S, Borod J. "Unattended emotional intonations modulate linguistic prosody processing". *Brain and language*. Vol105 (2), 141-7, 2007.
- [9] Alpert, M., Pouget, E. R., Silva, R. R., "Reflections of depression in acoustic measures of the patients speech." *J. Affective Disorders*. Vol66, 59-69, 2001.
- [10] Rahrkar, M., Hansen, J. H. L., "Frequency band analysis for stress detection using a Teager energy operator based feature". In: *Proc. Int. Conf. Spoken Language Processing (ICSLP ' 02)*. Vol3, 2021-2024, 2002.