

Determining the Relative Differences of Emotional Speech Using Vocal Tract Ratio

Jianglin Wang* · Cheolwoo Jo*

ABSTRACT

In this paper, our study focuses on obtaining the differences of emotional speech in three different vocal tract sections. The vocal tract area was computed from the area function of the emotional speech. The total vocal tract was divided into 3 sections (vocal fold section, middle section and lip section) to acquire the differences in each vocal tract section of emotional speech. The experiment data include 6 emotional speeches from 3 males and 3 females. The 6 emotions consist of neutral, happiness, anger, sadness, fear and boredom. The measured difference is computed by the ratio through comparing each emotional speech with the normal speech. The experimental results present that there is not a remarkable difference at lip section, but the fear and sadness have a great change at the vocal fold part.

Keywords: emotional speech, vocal tract ratio, voice quality

1. Introduction

Emotional speech is one of the complicated human voices. And the transmission of emotions in human speech communication is a topic that has been recently received considerable attention [1]. There are many aspects which are related to voice quality. Voice quality can be defined as verbal expressions such as breathy, whispery, creaky, falsetto, tense, harsh, modal, lax etc. [2] But such expressions are too subjective and do not give us objective ways for classifying different voice qualities. We need to know more about voice quality to investigate the characteristics of emotional voice.

Various recent studies have dealt with the characteristic of the emotional speech in recent years. The major purpose of such researches is to obtain the property of emotional speech and

* SASPL, School of Mechatronics, Changwon National University

to incorporate the expression of emotions into synthetic speech. Therefore, such kinds of research seem to have gained popularity in recent years.

2. Data collection

The emotional speech data was collected from the professional actors who were requested to utter the suggested emotional speeches. [3] The actor read ten given sentences with six sorts of different emotions. These six emotions were neutral (N), happiness (H), anger (A), sadness (S), fear (F) and boredom (B). They were chosen and adopted because they are widely used in the general life. The collected emotional data was evaluated by 20 judges using subjective opinion test. Overall coincidence rate was over 80% on the average.

Recording process was performed in semi-sound proof room, which has approximately -10 dB shield effect. [4] The data of the sampling rate was 48000 Hz and the resolution 16 bits. It is well known that the vowel /a/ is most widely used in the research of speech signal processing. So the vowel /a/ was segmented and extracted from each sentence using the voice analysis software PRAAT.

3. Method

The equations between the pressure and volume velocity in an acoustic tube are well known in acoustics as the momentum equation

$$\frac{\partial p_m(x,t)}{\partial x} = \frac{-\rho}{A_m(x)} \frac{\partial u_m(x,t)}{\partial t} \quad (1)$$

and the continuity of the mass equation

$$\frac{\partial u_m(x,t)}{\partial x} = \frac{-A_m(x)}{\rho c^2} \frac{\partial p_m(x,t)}{\partial t} \quad (2)$$

$p_m(x,t)$ — The pressure

$u_m(x,t)$ — The volume velocity

ρ — The air density

c — The velocity of the propagation

The reflection coefficient u_m is defined as

$$u_m = \frac{A_{m-1} - A_m}{A_{m-1} + A_m} \quad (3)$$

u_m — The reflection coefficient of the m^{th} segment

A_m — The vocal tract area function of the m^{th} segment

Depending on the above relationship between the reflection coefficient and the vocal tract area function, the vocal tract area function can be calculated from the reflection coefficient. [5][6][7] To obtain the differences of the vocal tract ratio, the vocal tract was divided into 10 segments. Vocal tract area was divided into 3 sections to measure the changes of each part of vocal tract to see the articulatory behavior according to each emotion. The first 3 segments are defined as vocal fold section, the middle 4 segments are defined as middle section and the last 3 segments are defined as lip section. It is shown in figure 1. For each segment the average area function was calculated. The vocal tract ratio of each emotion is compared between the average area function of the neutral emotion and other emotion.

The relative index of i -th vocal tract section is computed by the following expression.

$$I_i = \frac{A_{ei}}{A_{Ni}} \quad (4)$$

A_{ei} : area of emotional state in i -th vocal tract section

A_{Ni} : area of normal state in i -th vocal tract section

I_i : relative index of i -th vocal tract section

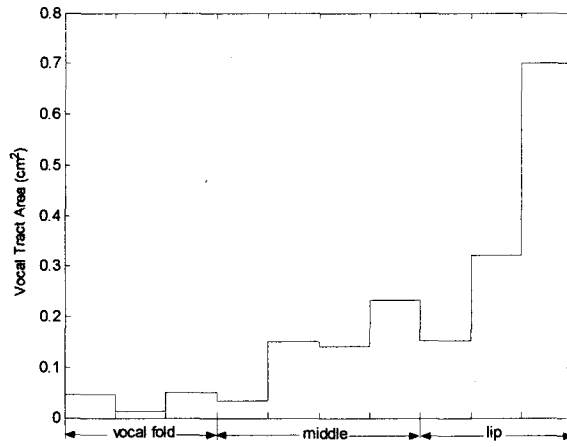


Figure 1. The vocal tract area function of the vowel /a/

4. Experimental results

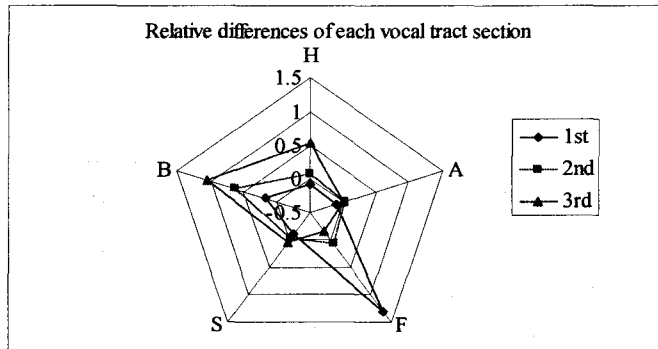


Figure 2. The relative differences of emotional speech of the female MYS

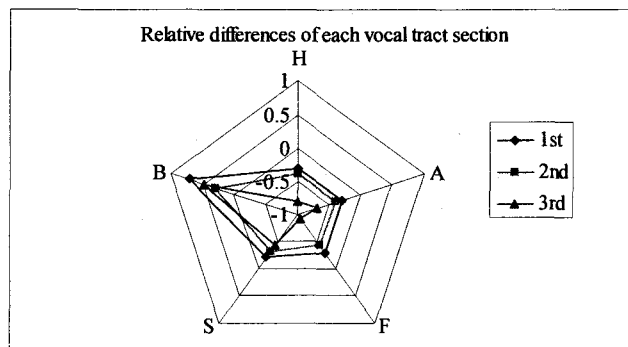


Figure 3. The relative differences of emotional speech of the female KKS

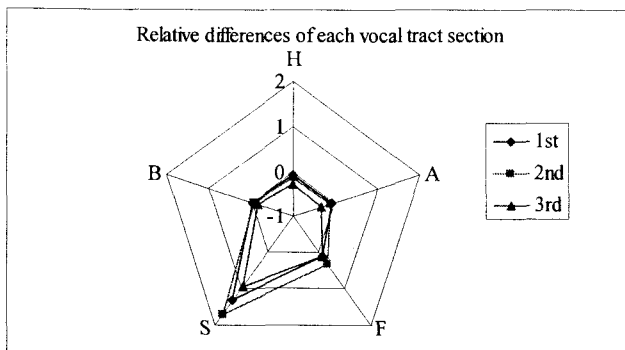


Figure 4. The relative differences of emotional speech of the female LHJ

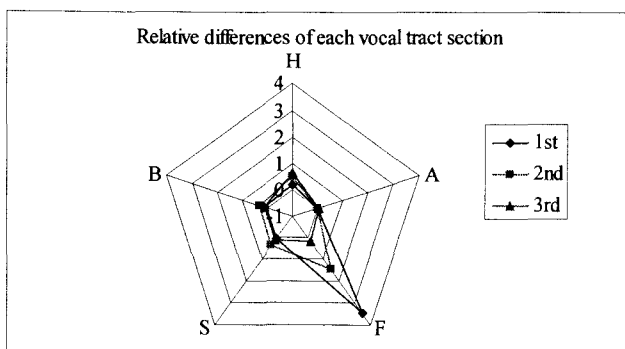


Figure 5. The relative differences of emotional speech of the male CWJ

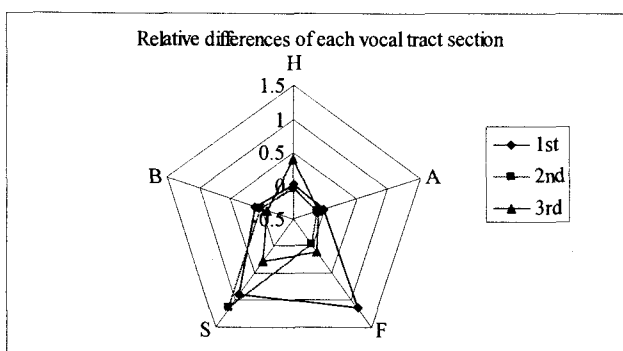


Figure 6. The relative differences of emotional speech of the male PYH

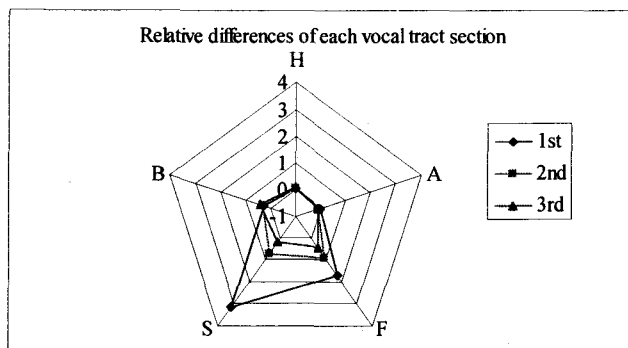


Figure 7. The relative differences of emotional speech of the male YSW

The relative differences of each vocal tract section are presented from figure 2 to figure 7. The computed index was drawn in star graphs to be compared. In those figures, 1st, 2nd, 3rd denote 3 different sections (vocal fold section, middle section and lip section). To speaker MYS, the fear emotion presents a great difference comparing with neutral in vocal fold section and the boring emotion also has a difference in lip section. To speaker KKS, the obvious difference is the boring emotion and the other emotions possess a little difference. To speaker LHJ, it is marked that the sadness has an obvious difference in middle section and lip section. To speaker CWJ, the fear emotion represents a big change in vocal tract section. To speaker PYH, the fear and sad emotion have a great difference. To speaker YSW, the biggest difference is the sadness in vocal tract section and the second is the fear.

It is much more convenient to make a table to compare the difference of emotional speech in different section. In table 1, the symbol “-” represents that the relative difference value is below -0.5, the symbol “+” represents that the relative difference is beyond 0.5, the symbol “++” shows the difference is beyond 1 and the symbol “+++” shows the difference is beyond 1.5. From the table 1, it is easily observed the fear emotion of most speakers represents the relative difference beyond 1.5 in vocal folds section (i.e. section 1) and the sadness has obvious differences in middle and lip section (i.e. section 1 and 2). The big difference of the other emotions can’t be observed from the limited emotional speech.

Table 1. Relative difference table in each section

| | Section | MYS | KKS | LHJ | CWJ | PYH | YSW |
|---|---------|-----|-----|-----|-----|-----|-----|
| H | 1 | | | | | | |
| | 2 | | | | | | |
| | 3 | + | - | | | + | |
| A | 1 | | | | | | |
| | 2 | | | | | | |
| | 3 | | - | | | | |
| F | 1 | +++ | | | +++ | +++ | +++ |
| | 2 | | | | + | | ++ |
| | 3 | | - | | | | |
| S | 1 | | | ++ | | ++ | +++ |
| | 2 | | | +++ | | +++ | ++ |
| | 3 | | | + | | | |
| B | 1 | | + | | | | |
| | 2 | + | | | | | |
| | 3 | ++ | | | | | |

-: diff<-0.5, +: diff>0.5, ++: diff>1, +++: diff>1.5

In the experiment we failed to find any distinctive results from statistical analysis. Two reasons can be considered. One is the limit of size in emotional data and the other is lack of definite control to coincide the emotional status of the actors.

5. Conclusions

In this experiment, the differences of emotional speech were observed using the vocal tract ratio in the 3 vocal tract sections. From the vowel /a/ experimental results, the most distinctive

characteristic is that the fear and sadness have great differences in lip section and the sadness also possesses relative difference in middle section. The other emotions represent a little difference of vocal tract ratio. The experimental results can become a research reference of emotional speech.

The total amount of emotion speech is still not enough to generalize the characteristic. Therefore, In the future works much more emotional speech data collection is required. Also the control method to normalize the emotional states need to be investigated.

References

- [1] Gobl, C. & Chasaide, A. N. 2003. "The role of the voice quality in communicating emotions, mood and attitude", *Speech Communication*, vol. 40, pp. 189-212.
- [2] The Institute for Natural Language Processing (IMS), *EGG & Voice Quality*, <http://www.ims.uni-stuttgart.de/phonetik/EGG/frmst1.htm>, the University of Stuttgart, Germany.
- [3] SiTEC, *Korean Emotional Speech Database*, 2004.
- [4] Jo, Cheolwoo. & Il-suh Bak. 2004. "Collection of Korean Emotional Speech Database from Actors", Proceedings of ASK conference, vol.23, No.1, pp.45-48.
- [5] Kent, R. D. & Ball, M. J. 2000. "Voice Quality Measurement", Singular.
- [6] Rabiner, L. R. & Schafer, R. W. 1978. "Digital Processing of Speech Signals", Prentice-Hall.
- [7] Marker, J. D. & Gray, R. M. 1980. "Linear Prediction of Speech", Springer-Verlag.

received: January 30, 2006

accepted: March 13, 2006

▲ Jianglin Wang

SASPL, School of Mechatronics, Changwon National University
9 Sarim-dong, Changwon, Kyungnam, 641-773, Korea
Tel: +82-55-279-7559 Fax: +82-55-262-5064
E-mail: xiaowangyc@hotmail.com

▲ Cheolwoo Jo

SASPL, School of Mechatronics, Changwon National University
9 Sarim-dong, Changwon, Kyungnam, 641-773, Korea
Tel: +82-55-279-7552 Fax: +82-55-262-5064
E-mail: cwjo@sarim.changwon.ac.kr