EVALUATION OF THE RELATIONSHIP BETWEEN EMOTIONAL CONCEPTS AND EMOTIONAL PARAMETERS ON SPEECH

Tsuyoshi MORIYAMA *Hideo SAITO* *Shinji OZAWA*

Department of Electrical Engineering, Keio Univ., 3–14–1
Hiyoshi Kouhoku-ku, Yokohama, Kanagawa, 223 Japan
moriyama@ozawa.elec.keio.ac.jp

ABSTRACT

In this paper, we propose the linear model of the relationship between the physical changes in speech and perceived emotional concepts. We make use of orthogonal bases in spite of emotional words and physical parameters themselves in order to avoid dependence on the method of selecting words and parameters. Furthermore we regard the emotions that listeners perceive from speech as the standard of emotional concepts because the emotions that speakers intented rely on personality and temporary psychological state [1]. Evaluation for relative information indicates that the proposed linear model is representable for the relationship between physical quantities and psychological quantities in speech.

1. INTRODUCTION

Recently, the informations called human factors such as emotion or KANSEI have been much important in many fields [2] such as HI. To achieve friendly man-machine communication, machines should have the structure of emotional evaluation for partner's facial and/or vocal expression. Then it is necessary to obtain knowledges about relationship between observed physical information and emotional content.

In the past researches, it is assumed that the emotion belongs to the speaker. However the emotion that the speaker intended is not always communicated to the listener. Moreover, regarding speaker's emotion as the standard contains some uncertain factors such as speaker's personality and speaking situation. Then obtained relative information between the emotional content and physical parameters of speech lacks of generality and reproducibility. On the other hand, the number and the kind of parameters and emotional words are actually limited for observation, it has dominant effect on the confidence of obtained relative information. In the past studies, a few emotional words and parameters are used, and they are chosen by the worker's taste or citation from "primary emotions" [2, 3, 4]. The primary emotions are also extracted in different modality of communication, they are not necessarily suitable for representing vocal emotion [5, 6].

In this paper, the solution for mentioned problems is proposed. In stead of the speaker's emotion, the listener's emotion, so to speak, rearranged emotion is regarded as the standard [1]. And relative information is provided not between emotional words and speech parameters themselves but between orthogonal bases extracted from them respectively. Using orthogonal bases, relative information is independent of the way and the kind of choose emotional words and speech parameters. We propose linear model that associates psychological bases with physical bases statistically.

In the following chapters, we show the relationship model, and the methods of obtaining relative informations in the model are presented.

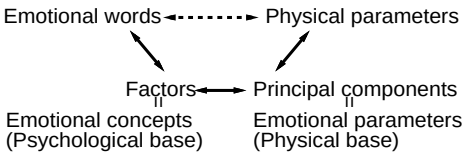

Figure 1. The proposed model

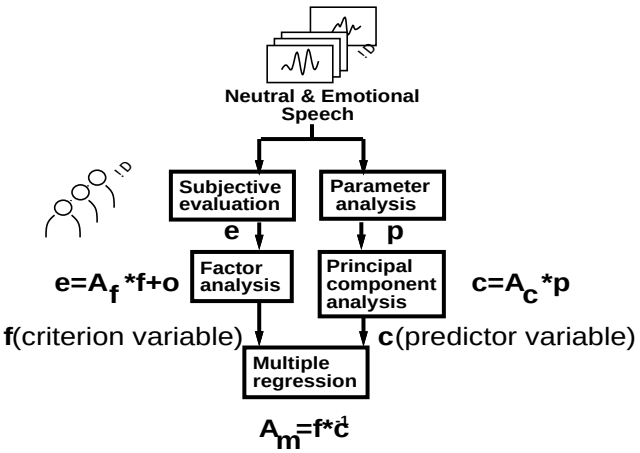

Figure 2. Obtaining the relative informations

2. RELATIONSHIP MODEL

Fig.1 shows the relationship model. It makes use of extracted orthogonal bases in spite of the observation objects of emotional words and speech parameters.

3. CONSTRUCTION OF THE MODEL

Fig.2 shows method for extracting the orthogonal bases. Factor analysis is conducted for the evaluation values **e** ob-

Table 1. Speech data for system construction.

#	speaker	word
1	A	mi-ro-yo
2	A	son-na
3	A	son-na
4	B	ha-ya-ku
5	B	mi-ro-yo
6	B	mi-ro-yo
7	C	mi-ro-yo
8	C	o-ma-e
9	D	mi-ro-yo
10	D	o-ma-e
11	D	o-ma-e
12	D	son-na

Table 2. Questionnaires

1	anger	24	scorn
2	joy	25	delight
3	disgust	26	cynicism
4	contempt	27	indifference
5	funny	28	praise
6	worry	29	pride
7	kind	30	love
8	relief	31	grief
9	outrage	32	flattery
10	shame	33	satisfaction
11	calm	34	boring
12	admiration	35	painful
13	irritation	36	hope
14	complaint	37	happy
15	longing	38	favorite
16	pitiful	39	dislike
17	tolerance	40	displeased
18	gloat	41	discouragement
19	disappointment	42	criticism
20	scolding	43	anxiety
21	sad	44	surprise
22	fear	45	flurried
23	hateful	46	amazed

tained by subjective experiment. And extracted factors **f** are called "emotional concepts". Principal component analysis is done for the physical changes **p** of speech including emotions. And extracted principal components **c** are called "emotional parameters". The multiple regression analysis relates both bases **c** and **f**.

3.1. Speech data

Four male speakers(one actor included) recorded words in Table 1 at a sampling rate of 44.1 kHz with 16-bit resolution. Before uttering words, the speakers were taught the questionnaires in Table 2 about which subjects evaluate speech. Without being specified emotions, the speakers uttered freely(about 10 kinds for each 8 words, 445 speech data in all and 12 speeches were picked up here).

3.2. Emotional concepts

3.2.1. Subjective experiment

The 18 listeners (16 males, 2 females) were graduate students who were instructed to fill in the evaluation on questionnaires about impressed emotion on each speech. Fig.3

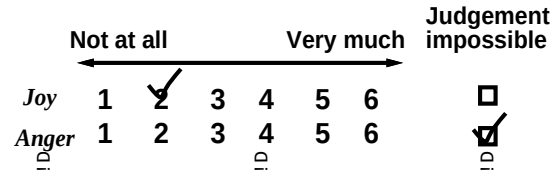

Figure 3. Questionnaire

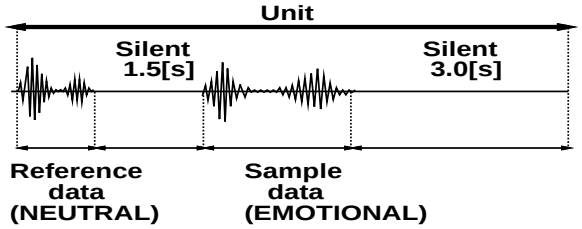

Figure 4. Data structure

is one of the questionnaires which are based on the seven steps measure for rating.

Speech was structured indicated in Fig.4 of pairing off neutral speech and emotional speech with silence between them. Each speech was repeated until all subjects finish to answer for all terms in Table 2. The order of 46 words is randomized and 18 subjects is divided into three groups and the order of 12 speeches is randomized among the groups in order to avoid distortion caused by the order.

3.2.2. Factor analysis

$$\mathbf{e} = \mathbf{A_f}\mathbf{f} + \mathbf{o} \tag{1}$$

Factor analysis estimates factor loading matrix $\mathbf{A_f}$ and extracts factors(factor scores) **f** in Eq.(1) simultaneously. **e** indicates evaluation values for 46 usual emotional words in Table 2. $\mathbf{A_f}$ is the relative information between the emotional basic concepts and usual emotional words.

It is assumed in this research that usual emotional concepts are consists of linear combination of several basic concepts, therefore it is suitable for it to apply the factor analysis model.

3.3. Emotional parameters

3.3.1. Physical parameters

Table 3 shows the physical parameters used here. Each character in Table 3 means as follows:

P Short-time average power
F Pitch frequency extracted by FFT cepstrum
L Length
D First order differentiation
A Average
R Dynamic range
T Accent-part in each word
E End-part in each word

ex.1 TFA means an average pitch frequency for accent-part
ex.2 FR means dynamic range of pitch frequency for total-part

Table 3. Physical parameters

#	parameter	#	parameter
1	FA	15	FDA
2	FR	16	FDR
3	FDA	17	TFA
4	FDR	18	TFR
5	TFA	19	TPDA
6	TFR	20	TPDR
7	TFDA	21	EPA
8	TFDR	22	EPR
9	EFA	23	EPDA
10	EFR	24	EPDR
11	EFDA	25	L
12	EFDR	26	TL
13	PA	27	EL
14	PR		

3.3.2. Principal component analysis

$$\mathbf{c} = \mathbf{Acp} \tag{2}$$

Principal component analysis estimates eigenvectors $\mathbf{Ac}$ and extracts principal components $\mathbf{c}$ in Eq.(2). $\mathbf{p}$ means phsysical changes of the speech including emotions in Table 3. $\mathbf{Ac}$ is the relative information between the emotional basic parameters and physical parameters' changes.

It is assumed in this research that several physical parameters should be contributed to the emotional parameters which explain physical changes caused by the speech including emotions, it is, therefore, suitable for it to apply the principal component analysis model.

3.4. The relationship between emotional concepts and emotional parameters

3.4.1. Multiple regression analysis

$$\mathbf{f} = \mathbf{Amc} \tag{3}$$

In the present research, the predictor variables is the emotional parameters $\mathbf{c}$ and the criterion variables is the emotional concepts $\mathbf{f}$. Multiple regression analysis estimates the partial regression coefficient matrix $\mathbf{Am}$ in Eq.(3).

It is assumed that several physical changes are integrated to contribute to convey the psychological quantity of emotion. It is, therefore, suitable to apply the multiple regression analysis model for prediction of emotional concepts by emotional parameters.

4. EVALUATION OF THE RELATIONSHIP

4.1. Result of factor analysis

The primary factor had the large rate of contribution(33.70[%]) alone. Fig.5 shows positionings of some emotional words on the one-dimensional axis. The primary factor can be interpreted as describing "pleasant to another — unpleasant to another" from the kinds of the positioned words.

4.2. Result of principal component analysis

The primary and secondary principal component had the large rate of contributions(35.79[%] and 12.58[%], respectively). Fig.6 and Fig.7 shows positionings of some physical parameters on the one-dimensional axis for each compo-

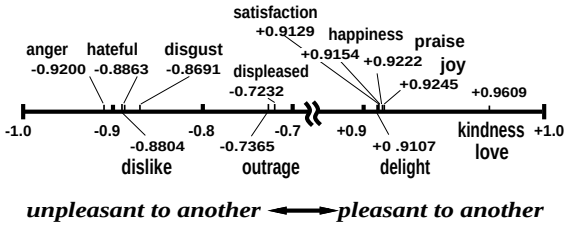

Figure 5. Primary emotional concept with positioning evaluation words.

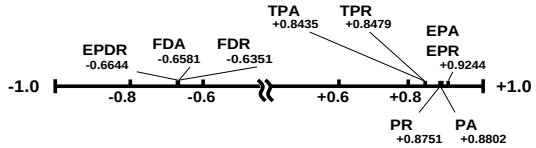

Figure 6. Primary emotional parameter with positioning physical parameters.

nents. The primary principal component can be interpreted as representing physical changes caused by increasing power on the speech, and we call it "loud" parameter. The secondary principal component can be interpreted as standing for physical changes caused by extending length of the speech, and we call it "extension" parameter.

4.3. Result of multiple regression analysis

Table 4 shows the result of multiple regression analysis for each emotional concepts. From the result of factor analysis, it can be expected that the primary factor has so large rate of contribution that its precision is dominant to total one, then in Table 4, the primary factor has large values.

5. CONCLUSION

In this paper, we proposed the linear relationship model between the emotions, which the listeners perceived from speech, and physical changes caused by the speech including emotions. The orthogonal bases in emotional space are extracted through factor analysis, and obtained relative information which translates the emotional bases from the emotional evaluations to usual words. Emotional parameters are extracted through principal component analysis, and obtained relative information which translates the emotional parameters from physical parameters. The relative information which translates emotional concepts from emotional parameters are obtained through multiple regression analysis. By applying these relative informations, emotional expression transformed from neutral speech, for example,

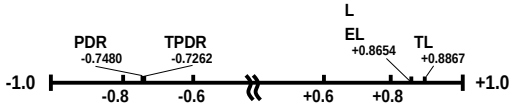

Figure 7. Secondary emotional parameter with positioning of physical parameters.

Table 4. Multiple correlation coefficients and coefficients of determination.

Factor#	multiple correlation coefficient	deterministic coefficient
1	0.89	0.56
2	0.82	0.26
3	0.81	0.26
4	0.84	0.34
5	0.76	0.08
6	0.85	0.39
7	0.72	-0.07
8	0.91	0.61
9	0.87	0.45
10	0.82	0.30
11	0.81	0.26
12	0.89	0.52
13	0.77	0.09
14	0.83	0.30

and evaluation for emotional content for arbitrary speech are available.

In the future, we also determine the neutral speech based on listeners' evaluations. It is expected that the model holds consistency. Then the emotional analysis/synthesis system will be constructed using proposed relationship model.

ACKNOWLEDGE

The authors acknowledge with gratitude the cooperation for recording speech data for Messrs. Masaaki Asonuma, Hirohito Horie and Norihiko Sumiyoshi.

REFERENCES

[1] L.A.Streeter, N.H.Macdonald, R.M.Krauss, W.Apple, K.M.Galotti, "Acoustic and perceptual indicators of emotional stress, " J.Acoust.Soc.Am., vol.73, No.4, pp.1354-1360, April 1983.

[2] H.Levin, W.Lord, "Speech Pitch Frequency as an Emotional State Indicator, " IEEE Trans. Systems, Man, and Cybernetics, vol.SMC-5, No.2, pp.259-273, March 1975.

[3] P.Lieberman, S.B.Michaels, "Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to the Emotional Content of Speech, " J.Acoust.Soc.Am., Vol.34, Number 7, pp.922-927, July 1962.

[4] C.E.Williams, K.N.Stevens, "Emotions and Speech: Some Acoustical Correlates, " J.Acoust.Soc.Am., Vol.52, Number 4, Part 2, pp.1238-1250, 1972.

[5] I.R.Murray, J.L.Arnott, "Toward the simulation of emotion in synthetic speech : A review of the literature on human vocal emotion, " J.Acoust.Soc.Am., vol.93, No.2, pp.1097-1108, February 1993.

[6] H.Schlosberg,
"The Psychological Review, " J.exp.Psychol., Vol.61, No.2, pp.81-88, March 1954.