EVALUATION OF THE RELATION BETWEEN EMOTIONAL CONCEPTS AND EMOTIONAL PARAMETERS IN SPEECH

Tsuyoshi MORIYAMA, Hideo SAITO, Shinji OZAWA

Department of Information and Computer Science, Keio Univ.,
3-14-1 Hiyoshi Kouhoku-ku, Yokohama-shi, Kanagawa, 223-8522 Japan
moriyama@sak.iis.u-tokyo.ac.jp

ABSTRACT

In this paper, the relation model which relates the physical changes caused from the speech including emotion to the emotional content perceived from the speech is proposed. Using the statistical bases extracted from each physical parameters and emotion words instead of themselves, the model enables to relate them independently on how to choose those variables. Also in this model, emotion which are shared among the listeners, so to speak the emotion stereotype, is considered as the standard in stead of the emotion intended by the speaker, accordingly the emotion can be supposed to be observable and reproducible. In this study, first, the physical parameters are calculated and emotional contents of several speeches are obtained through a psychological experiment, and they are processed by the statistical methods to get the orthogonal bases respectively. Second, these bases are related linearly by the multiple regression analysis. Consequently, the relation information which enables transformation between physical parameters and emotional content of the speech is obtained.

1. INTRODUCTION

Recently, the remarkable progress of the information communication infrastructure and the various development of the multimedia technology enables to provide many levels and kinds of information communication services to consumers. The "objects" in such services not only carry terrible and routine tasks out, but extend its role to the partner for the mental metabolism in the communication. In order to provide a kind of comfortability of a life through human machine communication, it is supposed to be necessary of the objects to understand, so to speak, the human factors such as emotion and also to imitate them. In the past studies, many researches concerned with emotion information carried by speech have been conducted aimed to such human-friendly technologies [3, 4, 5, 8, 9]. Almost of them are to formulate the relation between the physical parameters and the emotional content, and to synthesize emotional expression on the neutral speech, but not enough results have been obtained. Because the psychological quantity of emotion have the decision factors such as cultural background of the participants of the communication, situational context, vocal personality and meanings of words and so on, so a kind of the limitations or the conditions are required in order to treat it engineeringly and the differences among the viewpoints of the studies are reflected in such constraints. In the past studies, such constraints have rarely been investigated, accordingly it is supposed to have been difficult to clarify the relation among results of the studies.

The constraints which should be considered are that 1) the definition of the emotion which is appropriate to the engineering approach, 2) how to choose the observed variables, it means that what kinds of emotions are related to what kinds of physical features of speech?, and 3) the method of modeling the relation between emotions included in the speech and physical quantities of the speech.

In this paper, the solution about these constraints is proposed and the computational model which relates the physical changes of speech caused from including emotion to the emotion perceived from the speech is implemented, in which the relation between the psychological and the physical aspects of the speech which carries emotional content is defined as the relation information.

In section 2, the proposed model is briefly explained. In section 3, the statistical methods to obtain the relation information are shown, the model is implemented in section 4. In section 5, the results of the statistical methods are shown and the efficiency of the model is discussed. The last section concludes this study.

2. THEORY

2.1. Relation model

In communication, it is supposed that two types of emotion exist; one is the subjective one which is intended by the speakers, and the other is the emotion stereotypes which is rearranged or shared among those who observe the emotional expression objectively. The former, which is affected

by the outer factors such as the environment, the inner factors such as physiological conditions and also the previous emotional state, is difficult to observe because it is impossible to determine the causality among those decision factors. Meanwhile many of the past studies put the emotion standard on the former emotion. To the contrast, the latter is supposed to be observable through the statistical methods such as the subjective experiments [2]. Accordingly, it is supposed to be effective to evaluate the emotional content of the speech from the viewpoints of the listeners, especially for the purpose of measuring the psychological quantities of emotion. This study regards the latter one as the standard of emotion.

It is in fact necessary to select the limited number of the observed objects of physical parameters and emotion words for the experiments, and the selected objects mean the representatives of the object space. In past researches, the observed emotion words are selected based on so called the fundamental emotion [1, 10, 11] which had been proposed by the specified psychologist, while there have never been the studies which based on the emotion stereotypes. In this study, the emotion words from the past references and the results of the preliminary experiments are all gathered and used in the subjective evaluation experiments and a set of emotion words which are approximately independent on each other is obtained as the representatives of the stereotypes of the emotion. About the physical parameters, the physical features of the speech which include emotions have been investigated in the past studies [1], and in the results of them, it has been said that the prosodic components of the emotional speech is the carrier of para-linguistic information such as the intention of the speaker and non-linguistic information such as emotion. Accordingly also in this study, the prosodic components of speech is analyzed in order to measure physical aspects of an emotional speech.

On the other hand, the correlation among the prosodic parameters or emotion words has rarely been considered in the past studies. From the results of the preliminary statistical experiments here, however, it has clarified that there are strong correlation among prosodic parameters, moreover among emotion words. It indicates that the physical parameters and emotion words used for relating themselves include redundancy and they are likely to be partial in the physical and psychological space respectively. So in this research, these observed parameters are regarded as the points in the coordinate system of the orthogonal bases. It is supposed to be efficient to summarize physical parameters to the several basic components because it becomes possible to get relation information which is independent on how to choose the observed variables. Fig.1 shows the proposed model, the relation information is obtained by the process indicated as the solid line part.

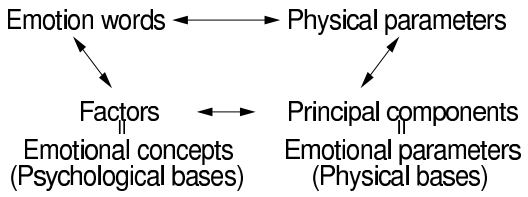

Figure 1: Proposed relation model

2.2. Relation information

The past studies have mattered the evaluation values $\boldsymbol{e}$ in the subjective evaluation experiments and the physical parameters $\boldsymbol{p}$ themselves when they are related, while in this study, $\boldsymbol{e}$ is analyzed by the factor analysis to extract the factors which are basic vectors of the psychological space, and the factor scores $\boldsymbol{f}$, which is obtained by mapping $\boldsymbol{e}$ to the basic space, and the factor loadings matrix $\boldsymbol{A_f}$, by which $\boldsymbol{e}$ can be transformed to $\boldsymbol{f}$, are both estimated. And $\boldsymbol{p}$ is analyzed by the principal component analysis to extract the principal components which are basic vectors of the physical space, and the principal component scores $\boldsymbol{c}$, which is obtained by mapping $\boldsymbol{p}$ to the basic space, and the eigen vectors $\boldsymbol{A_c}$, by which $\boldsymbol{p}$ can be transformed to $\boldsymbol{c}$, are both calculated. Then the multiple regression analysis where $\boldsymbol{c}$ and $\boldsymbol{f}$ are regarded as the predictor variables and the criterion variables respectively is conducted to obtain the partial regression coefficients matrix $\boldsymbol{A_m}$ as the relation information (Fig.1). By using the relation information, it becomes possible to calculate the emotional content of the speech from the prosodic parameters and to obtain the physical changes given to the neutral speech to synthesize the desired emotional speech. In the following section, the statistical methods used to obtain the relation information are shown.

2.3. Statistical methods

2.3.1. Factor analysis

In this study, it is assumed that the emotional concepts described by the emotion words consist of the linear combinations of the basic factors. Accordingly the factor analysis which is called 'analytic analysis' is applied in order to extract the basic factors. Eq.(1) describes the factor analysis model.

$$e_i = \sum_j \alpha_{ij} f_j + o_i, \tag{1}$$

where $\boldsymbol{e} = \{e_i | i = 1, \cdots, s\}$ stands for observed variables which is described by the linear combination of the common factors $\boldsymbol{f} = \{f_j | j = 1, \cdots, m; m \leq s\}$ and the original factors $\boldsymbol{o}$. α_{ij} denotes a component of the factor loading matrix

$\boldsymbol{A_f}(s \times m)$, where s, m stands for the number of variables and factors respectively. The purpose of factor analysis is to estimate this factor loading matrix $\boldsymbol{A_f}$ and the common factor scores $\boldsymbol{f}$.

Observed variables here are the subjective evaluation values for emotional content of speech. Extracted common factors f_j is called emotional concepts as the basic concepts of the stereotype shared among the listeners of emotional speech. In this study, the main factor method and the regression estimation method are applied to estimate common factors and factor scores respectively.

2.3.2. Principal component analysis

It is assumed that bases of prosodic changes caused from the speech including emotion consist of linear combination of several prosodic parameters, and principal component analysis is applied to extract the bases. The model of principal component analysis is described in Eq.(2).

$$c_k = \sum_l \beta_{kl} p_l, \tag{2}$$

where $\boldsymbol{p} = \{p_l | l = 1, \cdots, t\}$ is observed variables, and the principal component scores $\boldsymbol{c} = \{c_k | k = 1, \cdots, n; n \leq t\}$ is described by linear combination of p_l. β_{kl} stands for a component of eigen vector matrix $\boldsymbol{A_c}(n \times t)$, where t, n denotes the number of parameters and principal components respectively. The purpose of principal component analysis is to calculate eigen vector matrix $\boldsymbol{A_c}$.

Observed variables here is delta of prosodic parameters of speech from neutral state. Principal component scores c_k is called emotional parameters here as basic physical components which carry emotional information on speech. In this study, the main factor method is applied to estimate principal components.

2.3.3. Multiple regression analysis

It is assumed that several emotional parameters contribute to communicate emotional concept by speech here, and multiple regression analysis is conducted in which the former is regarded as the criterion variable and the latter is as the predictor variable. Obtained partial regression coefficients is considered as the relation information. The model of multiple regression analysis is described in Eq.(3).

$$f_j = \sum_k \gamma_{jk} c_k, \tag{3}$$

where γ_{jk} denotes a component of the partial regression coefficients matrix $\boldsymbol{A_m}(m \times n)$. The purpose of multiple regression analysis is to estimate the combination coefficient matrix $\boldsymbol{A_m}$.

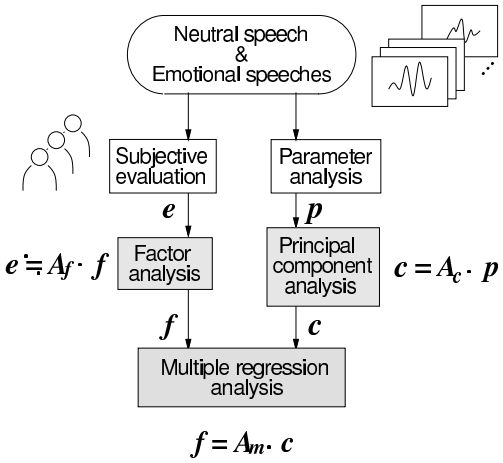

Figure 2: Obtaining the relation informations.

Table 1: Emotion words

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

3. OBTAINING RELATION INFORMATION

Fig.2 indicates the flow of obtaining the relation information. The left side of of the figure shows about extracting bases of emotion space, the right side about extracting basis of physical space and the lower side about relating those bases. For the purpose of getting the relation information, it is necessary to prepare the neutral speech and the emotional speeches which include various emotions.

3.1. Collecting emotional speech

Because any of emotional speech databases in japanese do not exist yet, it is necessary to collect emotional speech data for the experiments. In this study, 4 words of /omae/, /sonna/, /miroyo/ and /hayaku/ are spoken by 4 actors, who

are instructed to utter many kinds of emotional speech referring to the emotion words in Table 1. In this study, the words which consist of one syllable are used for avoiding the effect of the context, and which are used in usual life with various emotion expressions. 10 kinds of utterances per words, 445 speech data in all are stored. Each speech is sampled by 44100 Hz and quantized in 16 bit linearly.

3.2. Extraction of emotion words

3.2.1. Subjective evaluation experiment

Although it is necessary to obtain the evaluation values for more number of speech stimuli than the number of the evaluation terms, but about 2×10^4 evaluation (445 speeches and 46 emotion words) should cause in distortion of the fairness of the evaluation by tiredness of subjects. Accordingly, it is actually appropriate to decrease the number of emotion words to several significant number.

Then the subjective evaluation experiment is divided to two steps. At first, the several significant evaluation terms (emotion words) are selected by using all words in Table 1 in the preliminary experiment. Second, the subjective evaluation experiment is conducted using the selected words, and the obtained psychological data is analyzed by the factor analysis to extract the factor. The number of speech data is also decreased at the preliminary experiment.

The subjects consist of 18 students (16 males and 2 females). They are instructed, "Please evaluate the speech about how extent it includes each emotion relative to the reference speech and check the correspondent number on the questionnaire". If the objective speech is presented only once before evaluation, it is difficult to evaluate all 46 terms equivalently. And if the objective speech is repeated to keep the impression, it is afraid the effect of paralysis occurs. So in this experiment, the set of the reference speech and the objective speech is presented repeatedly because it can be assumed that the subjects are initialized by listening to the reference speech every time.

On the other hand, in order to avoid the effect caused from the order of the evaluation terms on the questionnaire (Fig.3), it is randomized about all subjects, and the presented set of speech is also randomized about three groups of the subjects. The psychological data is summarized by calculating the medium values of the subjects.

3.2.2. Extraction of emotional concepts

The psychological data obtained in the subjective experiment is analyzed by factor analysis to extract factors (emotional concepts). The number of factors is determined by the cumulative contribution which exceeds 70% here.

	never included	very little included	little included	cannot judge	slightly included	almost included	quite included
joy	1	2	3	4	5	6	7
anger	1	2	3	4	5	6	7
⋮				⋮			

Figure 3: Sheet format for evaluation of emotion

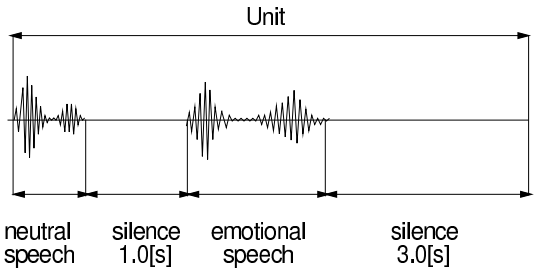

Figure 4: Structure of speech data

3.3. Extraction of emotional parameters

3.3.1. Prosodic parameters

The prosodic components of the speech means height, strength, rhythm and tempo, and in this study, F_0, short-time average power and velocity of an utterance are calculated as the prosodic parameters. F_0 and short-time average power are calculated in log scale for whole part of the speech, and the average value and the dynamic range of the temporal transition and its first differentiation are calculated. F_0 is estimated for the voiced part judged automatically using FFT cepstrum, where the frame length is 1024 points and the frame interval is 512 points.

F_0 and power amplitude is transformed to log scale based on Weber–Fechner's law.

3.3.2. Extraction of emotional parameters

The prosodic parameters of the emotional speech are normalized by those of the neutral speech, and analyzed by principal component analysis to extract the principal components (emotional parameters). The number of the principal components is also determined by the cumulative contribution which exceeds 70% here.

3.4. Relating bases

Based on the assumption that the extracted factors (emotional concepts) consist of the linear combination of the

Table 2: Selected emotion words

1.	anger
2.	joy
3.	cynicism
4.	fear
5.	sorrow
6.	surprise
7.	flattery
8.	calm
9.	funny

principal components (emotional parameters) which are the physical carrier of emotion, the multiple regression analysis is conducted where the factor scores are regarded as the predictor variables and the principal component scores are as the criterion variables.

4. RESULTS OF PROCESSING

4.1. Chosen emotional words and speech data

As the result of a subjective evaluation experiment in which all words in Table 1 are used, 9 words in Table 2 were chosen by means of that the correlation coefficients were calculated among all evaluation value vectors correspondent to each words in Table 1, and those which have large positive coefficients were interpreted as similar words and merged, while those which have large negative coefficients were interpreted as positioning opposite side of psychological space and remained.

Then more than 10 speeches is necessary to keep reliability of factors for 9 words statistically, accordingly speeches of 445 speeches in Table 3 which seem to include emotions easy to judge and different each other are selected in order to reduce the load for the subjects in the 2nd step of subjective evaluation experiment.

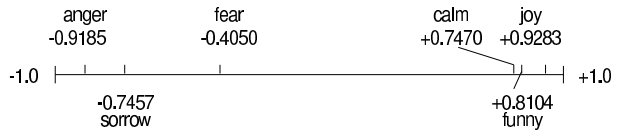

Figure 5: Positioning of evaluation terms for the 1st emotional concept

Table 3: Speech for obtaining relation information

#	Speaker	Word
1.	A	/miroyo/
2.	A	/sonna/
3.	A	/sonna/
4.	B	/hayaku/
5.	B	/miroyo/
6.	B	/miroyo/
7.	C	/miroyo/
8.	C	/omae/
9.	D	/miroyo/
10.	D	/omae/
11.	D	/omae/
12.	D	/sonna/

Table 4: Extracted factors (Emotional concepts)

factor#	C[%]	CC[%]
1	44.757	44.757
2	30.444	75.202
3	12.655	87.857

4.2. Result of extraction of emotional concept

As the result of factor analysis for psychological data obtained using 9 words in Table 2, the factors in Table 4 were extracted, where C and CC are abbreviations of contribution and cumulative contribution respectively. The positions of the factors which have relatively large factor loading on the factor loading axes and shown in Fig.5 and 6, where factor loading were estimated by varimax rotation.

In Fig.5, 'joy', 'funny' and 'gentle' position at +1.0 side and 'anger' and 'sorrow' at -1.0 side, therefore the 1st factor axis can be identified as 'pleased-displeased' axis. And in Fig.6, 'cynical' and, though it has small absolute value, 'funny' exist at +1.0 side and 'fear' and 'surprise' at -1.0 side, therefore the 2nd factor axis can be interpreted as 'tension-slackness' axis. Moreover in Fig.7, emotions of 'fear' and 'surprise' and those of 'gentle' and 'funny' position at opposite side of the axis, so it can be interpreted as the 3rd axis 'attention-refusal' reported in past studies [1]. These factors extracted here are consistent with basic concepts reported in psychology [10].

4.3. Result of extraction of emotional parameters

Prosodic parameters mentioned in 3.3.1 are calculated for all speeches in Table 3, which are normalized by neutral one and analyzed by principal component analysis. The result is shown in Table 5, and the positioning of prosodic parameters which have relatively large principal component

Figure 6: Positioning of evaluation terms for the 2nd emotional concept

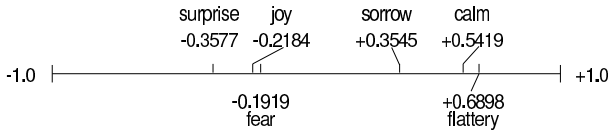

Figure 7: Positioning of evaluation terms for the 3rd emotional concept

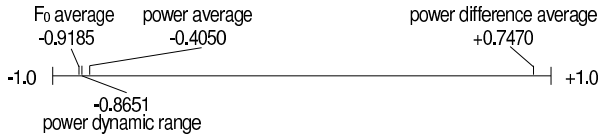

Figure 8: Positioning of prosodic parameters for the 1st emotional parameter

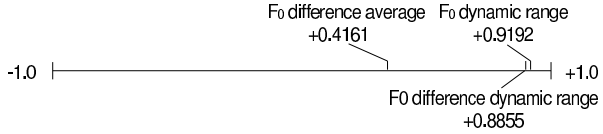

Figure 9: Positioning of prosodic parameters for the 2nd emotional parameter

loadings on the principal component axes are shown in Fig.8 and 9.

In Fig.8, average of F_0 and power negatively contribute to the 1st principal component and average of differential of power contour positively contributes to it, therefore the 1st principal component can be interpreted as 'loudness of whole speech' (sign inverted) axis. Also in Fig.9, the 2nd principal component doesn't have any parameters which have large loadings, but dynamic range and differentiation of F_0 contour strongly contribute to it on the positive side in stead, accordingly it can be interpreted as 'inflection' axis. The cumulative contribution exceeds 70%, so these two principal components are supposed to be dominant in physical space.

4.4. Result of relating bases

Factor scores and principal component scores were estimated for all speeches in Table 3 and these were analyzed by multiple regression analysis to obtain the partial regression coefficient matrix $\boldsymbol{A_m}$ shown in Table 6. Numbers in () are the standardised partial regression coefficients which indicate importance of each predictor variables. Examining partial regression coefficients in Table 6, it is found that the louder and the more inflective the speech changes, the more 'pleased' emotion is included, and the more silent and the calmer the speech changes, the less 'slackness' emotion is included. Moreover it can be said every factors (emotional concepts) weights on loudness of speech from the values of the standardised partial regression coefficients.

On the other hand, in order to evaluate the accuracy of multiple regression analysis, multiple correlation coefficients and coefficients of determination, which are modified in terms of the number of speech, are calculated as shown in Table 7. Multiple correlation coefficients means the correlative relation between the factor scores estimated from the prosodic parameters of speech and the factor loadings estimated from the subjective evaluation values, and as shown in Table 7, though it is a little insufficient in its accuracy for the 1st factor, almost good analysis result was obtained for the 2nd factor.

5. DISCUSSION

5.1. Concerned with the result of relating processing

The reason why the accuracy of relating processing was insufficient is supposed that variance of the psychological data includes not only emotional component but personality

Table 5: Extracted principal components

principal component#	C [%]	CC [%]
1	44.63	44.63
2	25.58	70.21

Table 6: Partial regression coefficients

	1st prin. comp.	2nd prin. comp.
1st factor	−0.1717 (−0.3441)*	+0.1123 (+0.1704)
2nd factor	+0.3039 (+0.6091)	−0.0849 (−0.1288)

* numbers in () means standardised partial regression coefficient

Table 7: Evaluation of the multiple regression analysis

factor#	mult. corr. coeff.	coeff.of det.
1	0.38	0.15
2	0.62	0.39

or meanings effects, because speech were uttered by several speakers using several words. On the other hand, linear model has been assumed here, it also can be supposed to be one of the reasons to ignore the nonlinear relation among physical parameters.

Accordingly it is supposed to be effective to use the speech uttered by only one person using a meaningless word and to incorporate some nonlinear methods such as neural networks without changing the framework in this study which transform observable emotions and physical parameters to bases to relate them.

5.2. Validity of factor interpretation

To verify the validity of interpretations given to factors, emotion included in speech is evaluated again based on the interpretation of factors and the correlation with the factor scores. The 1st factor is examined here, because it was given an interpretation which seems to be easy to recognize for human. In the subjective experiment, the 5 subjectives other than those in 3.2.1 were instructed to evaluate speeches in Table 3 on the questionnaire with 7 steps for rating for 'pleased-displeased' measurement, which is the interpretation of the 1st factor. In Fig.10, average values among 5 evaluation values is on the horizontal axis and the normalized 1st factor scores estimated for each speech in 4.2 is on the vertical axis. The correlation coefficient between these axes is 0.6056, it can be said that the interpretation word 'pleased-displeased' describes this factor almost well. Also in order to examine the correlation relation between the subjective evaluation values and the factor scores calculated by the relation model implemented in this study, the points for the same horizontal axis of Fig.10 and the 1st factor scores calculated from physical parameters of speech based on the relation information on the vertical axis were plotted on Fig.11.

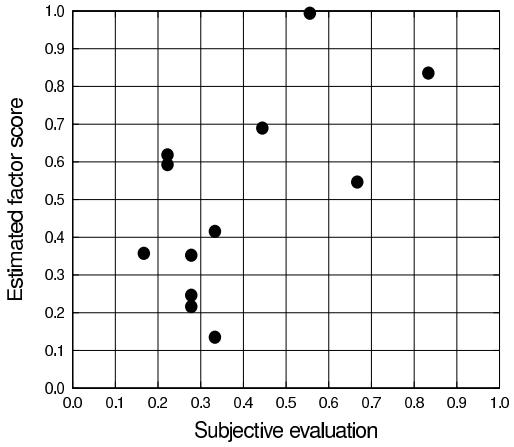

Figure 10: Relation between the evaluation values on 'pleased–unpleased' axis and the estimated 1st factor scores in sec.4.2

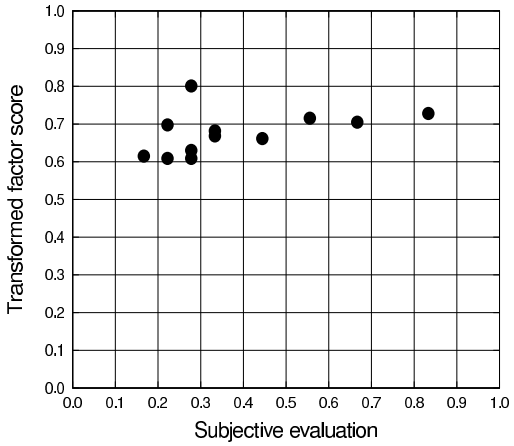

Figure 11: Relation between the evaluation values on 'pleased–unpleased' axis and the 1st factor scores

From Fig.11, positive correlation (correlation coefficient is 0.4434) can be seen between these axes though it is small. Accordingly it can be said that the validity of the interpretation word given for the 1st factor is verified.

6. CONCLUSION

In this paper, the relation model is proposed, in which the emotion perceived by listeners from speech is related to the physical changes caused from the speech including emotion through orthogonal bases. This model gives a solution to the issue of relating emotion carried by speech to the physical

changes caused from it, which is unavoidable when emotion is engineeringly treated in emotion information processing.

As the result of implementing this model, the emotion stereotype shared among listeners are summarized in 9 words, and the basic concepts obtained from these words through factor analysis are 3 of 'pleased-displeased', 'tension-slackness' and 'refusal-attention' axes which are consistent with the basic emotions reported by Schlosberg [10], which was based on facial emotional expressions, so it is interesting that the same basic concepts were obtained beyond different media types. And it was also found that the components of 'loudness' and 'inflection' are significant in physical changes caused from the speech including emotion.

In the future, as mentioned in 5.1, it is necessary to improve the accuracy of the relation information and to construct the system [7] which is able to recognize and synthesize vocal emotion by means of using parameters suitable for speech synthesis.

Acknowledgment

We appreciate following persons for cooperating us to collect emotional speeches: M.Asonuma, H.Horie and N.Sumiyoshi.

7. REFERENCES

[1] 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.

[2] 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.

[3] Dellaert, F., Polzin, T., Waibel, A., "Recognizing emotion in speech," *Proc. ICSLP 96*, vol.3, pp.1970–1973, 1996.

[4] Sato, J., Morishima, S., "Emotion modeling in speech production using emotion space," *Proc. 5th IEEE International Workshop on Robot and Human Communication Ro-MAN'96*, pp.472–477, 1996.

[5] Galanis, D., Darsinos, V.,, Kokkinakis, G., "Investigating emotional speech parameters for speech synthesis," *Proc. of the 3rd IEEE ICECS'96*, vol.2, pp.1227–1230, 1996.

[6] T.Moriyama, S.Ozawa, "Emotion recognition and synthesis system on speech, " *Proc. IEEE International Conference on Multimedia Computing and Systems*, vol.1, pp.840–844, 1999.

[7] T.Moriyama, H.Saito, S.Ozawa, "Evaluation of the relation between emotional concepts and emotional parameters on speech," *Proc. of IEEE ICASSP*, SPCH7P, 1997.

[8] 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.

[9] 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.

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

[11] R.Plutchik, "Emotion, A psychoevolutionary synthesis, " *Harper&Row*, 1980.

[12] J.R.Davitz, J.R.Davitz, "Correlates of Accuracy in the Communication of Feelings, " *J. Commun. 9*, pp.110–117, 1959.

Tsuyoshi MORIYAMA

Tsuyoshi Moriyama received the B.E., M.E., and Ph.D. degrees in Electrical Engineering from Keio univ., Japan, in 1995, 1997, and 1999, respectively. In 1999, he began to work as a post doctor fellow in univ. of Tokyo. His current research interests include pattern recognition, multimedia processing, speech processing, human emotion and sensibitily. He is a member of IEEE, Institute of Electronics, Information and Communication Engineers, and Acoustical Society of Japan.

Hideo SAITO

Hideo Saito received the B.E., M.E., and Ph.D. degrees in Electrical Engineering from Keio University, Japan, in 1987, 1989, and 1992, respectively. In 1992, he began to work as a research associate in Keio University, where he is now an assistant professor in the Department of Information and Computer Science. From 1997 to 1999, he was a visiting scientist in Robotics Institute in Carnegie Mellon University, where he worked on Virtualized Reality Project with Prof. Takeo Kanade. His current research interests include pattern recognition, image processing, signal processing, and computer vision. He is a member of IEEE, Information Processing Society of Japan, and Virtual Reality Society of Japan.

Shinji OZAWA

Shinji Ozawa received the B.E., M.E., and Ph.D. degrees in Electrical Engineering from Keio University, Japan, in 1967, 1969, and 1972, respectively. In 1970, he began to work as a research associate in Keio University, where he is now a chair professor in the Department of Information and Computer Science. From 1984 to 1985, he was a visiting professor in University of Maryland, where he worked with Prof. A. Rosenfeld. His current research interests include pattern recognition, image processing, signal processing, and computer vision. He is a member of IEEE, Information Processing Society of Japan, and Institute of Electrical Engineers of Japan.