A Measurement of Emotional Content in Speech and Its Application to Cyber Commerce

Tsuyoshi MORIYAMA Shinji OZAWA

November 10 1997 SERI @ Taejon.Korean

1 Introduction

We propose cyber commerce system which is able to learn customer's vague retrieval key and by which customers can obtain desired goods efficiently and fast. In this study, we introduce two main technologies in it: the first, a measurement method of emotional content from speech(such as 'anger', 'sorrow', and 'pleased' especially in cyber commerce), and the second, an image retrieval system not using any words as keys, which learns customer's vague image.

In the past studies, the image retrieval algorithms that treat user's sensibility are based on statistical methods. They, beforehand, determine the relationship between images(physical features) and their impressions(measured by psychological evaluation experiments), and it is fixed thereafter. Moreover, it is theoretically impossible to adapt individuals suitably. On the other hand, there are common sense that the retrieval keys are words. But it is, in fact, difficult for users to express their images by words, especially in the case that the object includes artistic factors like pictures and clothes. So a new image retrieval algorithm is required, which accepts vague keys which are difficult to be expressed by words, and adapts to individuals by learning mechanism.

On the other hand, the man-machine interface is very important component in such a interactive condition. Speech interface is very effective way to realize naturalness and to communicate contents which are difficult to express by words mutually. Such contents, so to say paralinguistic information, are conveyed by speech with linguistic information. So it is very effective to use speech interface in this case which needs to communicate vague contents as retrieval key.

In the following sections, we propose the cyber commerce system which consists of mentioned technologies. In the 2nd section, we show the outline of the system, and explain the modules in it after that.

2 Cyber Commerce System

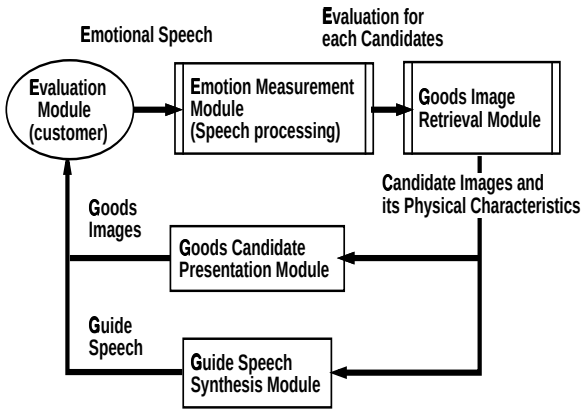

Fig. 1 Cyber Commerce System

The Cyber Commerce System is shown in Fig.1.

It has simple iterative mechanism. At first the system presents candidates of goods at random, and then the customer evaluates them by speech which includes some emotional content. The system analyzes the speech parameters, and extracts psychological evaluation values from them. Based on them, the system determines next candidates.

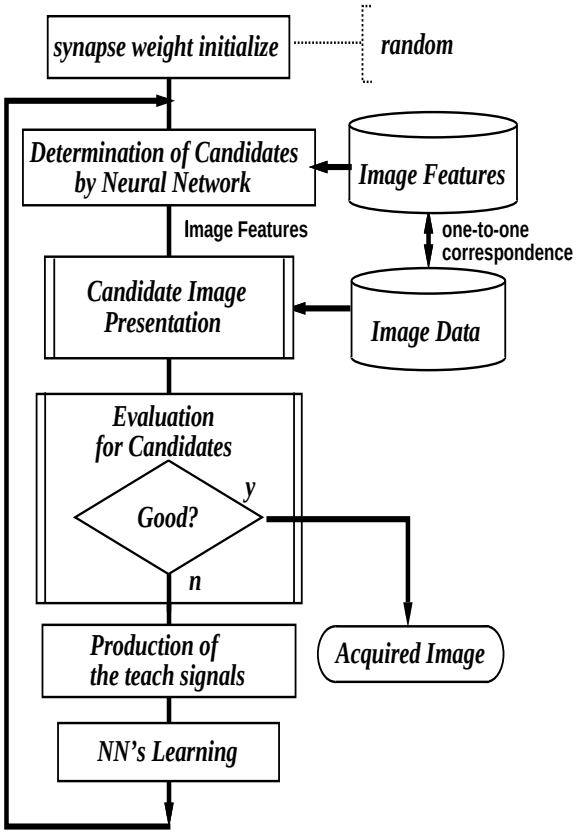

Fig. 2 Goods image retrieval module

3 Goods Image Retrieval Module

This is the kernel module of all. It is shown in Fig.2. It includes the interface(to present candidates and obtain the customer's evaluation) and the emotion measurement module in fact. It consists of iterative mechanism as indicated in Fig.2, which is the data flow of it.

The images are stored in image database and physical features database, and each images and its features correspond one-to-one, so they are treated in the form of physical feature vectors.

We explain about the initial candidates and effective retrieval method using neural networks, and change the section about the motion measurement module.

3.1 Physical Features

The physical features are calculated as shown in Fig.3. These are not specialized to

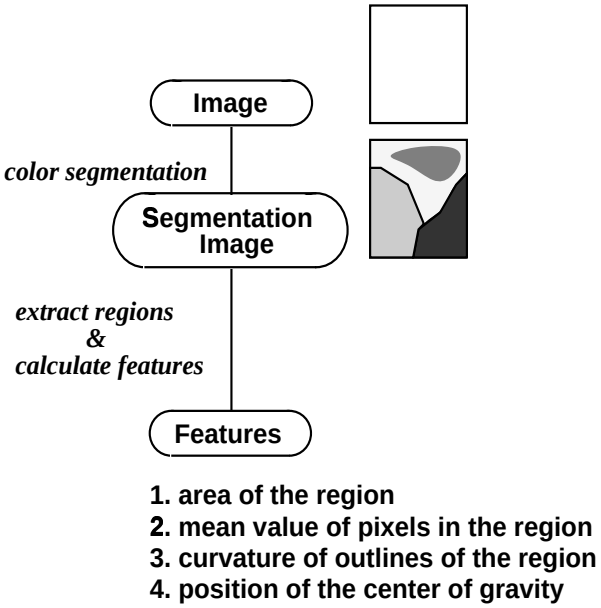

Fig. 3 Calculation of physical features

specific conditions or the kinds of retrieval keys, but fair to every kinds of images.

3.2 Initial Candidates

The initial candidates of goods are determined at random. It is achieved by setting the weights of each synapses of neural networks(Fig.4) to random values, and sorting the output values for each images when the physical features are input to input layer. The top 4 images are the initial candidates.

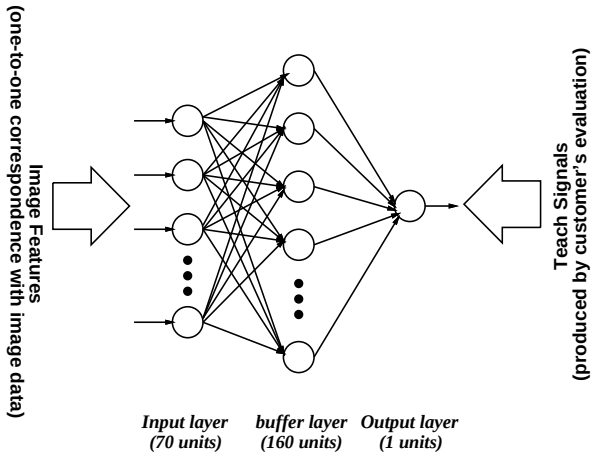

Fig. 4 The structure of neural networks

3.3 Image Retrieval Method

Neural networks learns the customer's request in the form of the weights of synapses. Networks presents the images which match up to the customer's request by means as follows:

1. The customer evaluates the candidates.
2. Networks minimizes the RMS between the customer's evaluation values and the output values using current weights of synapses by updating the weights(**BP Algorithm**).
3. The top 4 images are the next candidates using new weights of synapses.
4. Go to 1.

4 Emotion Measurement Module

This module measures the customer's evaluation for each candidates. Fig.5 shows the flow of obtaining psychological evaluation values from speech parameters. As indicated in Fig.5, it is necessary to get the relative information to transform physical quantities to psychological quantities. The procedure to get the relative information is shown in Fig.6. In this study, we use statistical methods which assume linearlity among observed variables(speech parameters and emotion words (questionaire) used in the subjective experiment).

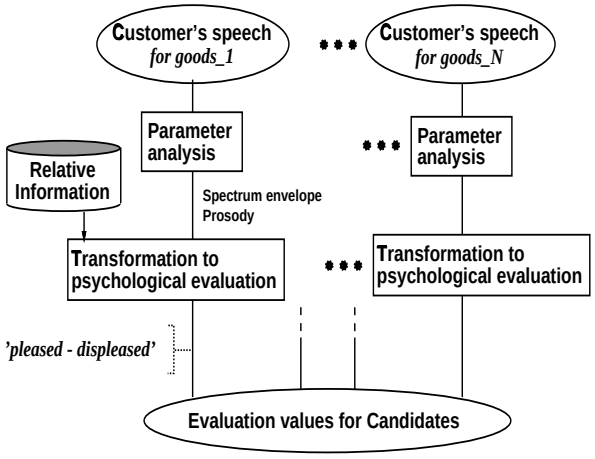

Fig. 5 Emotion analysis

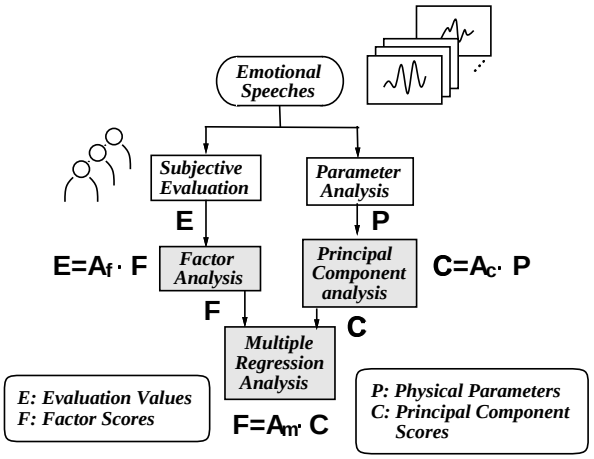

Fig. 6 Obtaining the relative informations

5 Experiments

5.1 Image Retrieval System

In this experiment, the purpose is examining the efficiency of the image retrieval system itself without the emotion measurement module, so evaluation values, here, are given manually in 7 steps for rating.

5.1.1 Condition

Image DB ··· 300 Scenery Images
Subjects ··· 4 students
Target sensibility words
··· "refreshing"
··· "simple"

Subjects are indicated to evaluate candidates for 7 steps at 5 times.

5.1.2 Results

Tab.1 and Tab.2 show the transition of the evaluation values for candidates in two cases. For the initial candidates, the evaluation values range over 1 to 7 in both cases, then the number of times growing, they also increased.

This results indicates that the image retrieval system learns user's implicit retrieval key and looks corresponding images up effectively.

Tab. 1 Transition of evaluation values for candidates (the case of "refreshing")

	Candidates			
times	No.1	No.2	No.3	No.4
1	6	1	3	5
2	4	3	2	6
3	5	2	6	5
4	5	3	6	4
5	6	7	7	6

Tab. 2 Transition of evaluation values for candidates (the case of "simple")

	Candidates			
times	No.1	No.2	No.3	No.4
1	5	2	1	5
2	6	6	6	2
3	6	4	5	6
4	1	2	4	6
5	5	5	7	6

Tab. 3 Accuracy of multiple regression

Factor#	multiple correlation coefficient	deterministic coefficient
1	0.57	0.33
2	0.30	0.09
3	0.24	0.06

5.2 Relative Information

Accuracy of measurement of psychological evaluation values (emotional content) from speech parameters is dependent on accuracy of the relative informations. We therefore examine the criterion which indicates accuracy of multiple regression analysis(Tab.3).

In this experiments, three factors are extracted by factor analysis. The positioning of questionaires(emotion words) to factor axes about each factors are shown in Fig.7.

Judging from the words positioned on the 2nd factor axis, it is possible to regard this factor as "pleased – displeased(to candidate goods)" axis, so accuracy only about 2nd factor in Fig.3

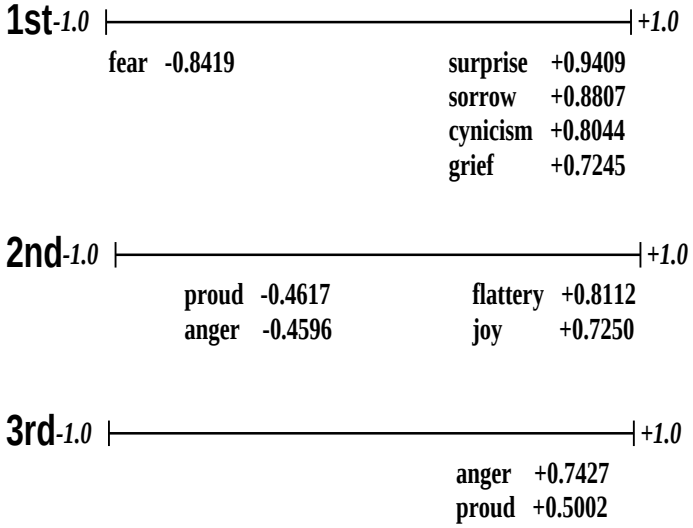

Fig. 7 Positioning of emotion words on factor axes

is important.

As understood from Tab.3, accuracy is not sufficient in fact. This is caused from following factors:

- Lack of the number of speech used in subjective experiments (it causes lack of the significancy of factor scores)
- Insufficiency of indication to subjects which value is zero emotion(it causes invalid distribution among evaluation values)

To improve accuracy, it is necessary to resolve these problem in the future work.

On the other hand, Fig.8 indicates the relationship between estimated factor scores and emotional content in speech. Emotional content is, here, determined by listening experiment.

As shown in Fig.8, there is slight cross-correlation between them. Factor scores, however, don't range 0.0 to 1.0 and some bias can be seen. From this results, it could be supposed that it is successful to extract emotional content from speech parameters although accuracy of multiple regression analysis is not sufficient.

In the following section, we examine the total system using this relative information.

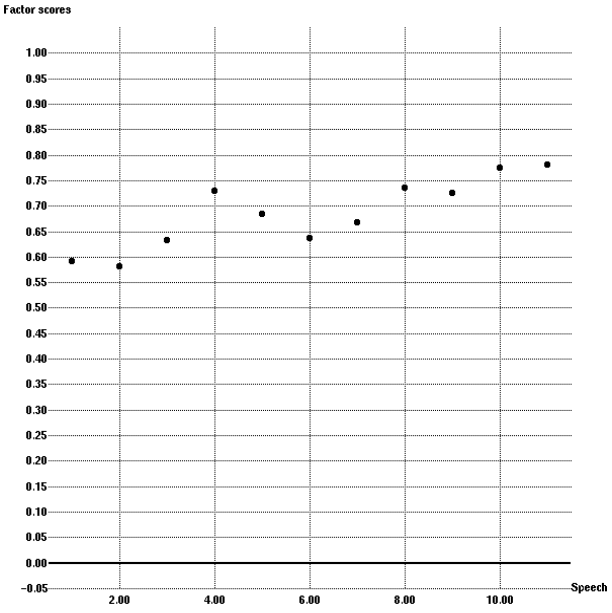

Fig. 8 Correlation between the 2nd factor scores and emotional content in speech

5.3 Total System Evaluation

To obtain psychological evaluation values, we derivate the 2nd factor scores in accordance with Fig.5 when a customer utters speech for evaluation. Factor scores are normalized through dividing by the root of eigen values(the standard deviation of factors).

The retrieval experiment is done under the same condition as section 5.1.1. And the result of it is shown in Tab.4 and Tab.5.

Like the result of section 5.1, evaluation values of the initial candidates range over 1 to 7 equally. Then in both cases, it can be seen that evaluation values against the 2nd candidates is high in general, and against the 3rd candidate, evaluation values are scattered again(subjects are forced to evaluate 6 times, so originally they will have acquired the image from the result of the 1st evaluation).

Tab. 4 Transition of evaluation values for candidates (the case of "refreshing")

	Candidates			
times	No.1	No.2	No.3	No.4
1	7	4	5	2
2	7	7	2	7
3	1	6	7	4
4	6	5	6	5
5	1	2	1	4
6	7	3	6	6

Tab. 5 Transition of evaluation values for candidates (the case of "simple")

	Candidates			
times	No.1	No.2	No.3	No.4
1	6	4	2	3
2	5	5	5	7
3	2	6	6	4
4	4	4	7	5
5	4	1	3	1
6	7	5	3	6

6 Conclusion

We construct an image retrieval system with speech interface which is able to measure emotional content in speech. It is aimed to cyber commerce service in the future.

In this paper, we proposed two main technologies: first, measurement method of emotional content in speech, second, image retrieval system which doesn't need any linguistic keys. It was confirmed that image retrieval system was successful to achieve efficient image retrieval. On the other hand, about the former, accuracy of the relative information was not sufficient, but estimated factor scores were correlated to emotional content in speech slightly, and from the result of total system evaluation, applying vocal emotion to evaluation mechanism in an image retrieval system would be significant.