Comparison of Emotion Perception among Different Cultures

Jianwu Dang[1], Aijun Li[2], Donna Erickson[3], Atsuo Suemitsu[1], Masato Akagi[1]
Kyoko Sakuraba[4], Nobuaki Minematsu[5], and Keikichi Hirose[5]
[1]Japan Advanced Institute of Science and Technology, Japan
[2]Institute of Linguistics Chinese Academy of Social Sciences, China,
[3]Showa Academia Musicale, Japan
[4]Kiyose-shi Welfare Center for the Handicapped, [5]The University of Tokyo, Japan
[jdang, sue, akagi]@jaist.ac.jp; liaj@cass.org.cn; ericksondonna2000@gmail.com

Abstract- In this study, we conducted a comparative experiment on emotion perception among different cultures. Emotional components were perceived by subjects from Japan, the United States and China, all of whom have no experience living abroad. An emotional speech database sans linguistic information was used in this study and evaluated using three- and/or six-emotional dimensions. It was found that most speech materials were perceived to have multiple emotional components, even though the speakers intended to express a single emotion in the data collection. Based on the principle component analysis (PCA), the common factors could explain about 67% variance of the data among the three cultures by using a three-emotion description, and could explain about 53% variance between Japanese and Chinese cultures by using a six-emotion description. The emotions *anger, joy* and *sad* have the same structure in PCA based low dimensional spaces derived from the three-emotion and six-emotion descriptions, while the emotions *Disgust*, *surprise*, and *fear* appeared in the six-emotion derived low dimension space as paired counterparts of *anger, joy* and *sad*, respectively. The similarity of subspaces consisting of these two emotion groups was around 0.9. This indicates that the emotions *anger, joy* and *sad* can be considered as basic emotions.

Keywords: Emotional speech, emotion cognition, multiple cultures, basis emotion, PCA analysis.

1. INTRODUCTION

Although speech understanding is a language-dependent activity, perception of non-linguistic information such as emotions is to some extent independent of cultural backgrounds. In daily conversation, we have experiences in which we can successfully perceive the emotions via speech even if we cannot understand the linguistic meaning, but misunderstandings also occur even when we are confident we understand the emotion. This study investigates what the common factors are involved in emotion perception among different languages, and what the differences are.

To characterize each emotion, many studies tries to extract a set of physical parameters from emotional speech. In most such researches, the emotional speech database is constructed by choosing a strongly emotional speech utterance supposedly uttered with an intended emotion. However, there are few speech sounds that have only one pure emotion in daily communication [1-3]. As pointed out in previous studies, speech-based emotion cognition is affected by differences among cultures of the speakers and listeners [4, 5]. It has been shown that the identification rate for certain intended emotions is higher for speakers and listeners who have the same cultural background. However, there is no answer as to what the common factors are in emotion identification and whether there are idiosyncratic differences in listeners with the same cultural background. This study is designed to answer these questions.

In this study, listeners from Japan, the United States, and China participated in experiments where a Japanese emotional speech database was employed for emotion evaluation. Subjects could freely evaluate any speech material using three emotions or six emotions, independent of which emotion had been intended by the speakers.

2. EMOTION PERCEPTION EXPERIMENTS

The purpose of this study is to clarify the common factors in emotion perception. We conducted perception experiments on the same database for subjects with different cultural backgrounds. The details of the experiments are described below.

2.1. Emotional speech database

Since linguistic information may affect the perception of emotions, the emotional speech database should be devoid of linguistic (i.e., lexical/semantic) information, especially for cross-language experiments. Based on such a consideration, we chose the database built up by Sakuraba et al. [6] for this study.

In constructing the database, 15 Japanese children ranging from 4 to 10 years old were asked to produce the voice of "*Pikachu*" when they saw an emotional picture of the character Pikachu which was selected from the famous animation of "Pocket Monster".

Since the children were familiar with the animation, it is expected that they learned the voice by means of understanding the emotions of the Pikachu character. Thus, the children said *pikachu* in the way they felt appropriate to express the emotion of the emotional picture of Pikachu. Such utterances did not have linguistic information regarding emotion. This database consisted of the four intended emotions: *anger*, *joy*, *sad*, and *surprise*. The number of speech utterances was 27 for *anger*, 28 for *joy*, 30 for *sad*, and 28 for *surprise*. The emotion of the speech defined in the database is referred to hereafter as the *intended emotion* to distinguish it from the perceived emotion obtained by the evaluations of this study.

2.2. Setup of the experiments

In this study, two experiments were designed. In the first experiment, we asked the subjects to evaluate the speech materials using only the three emotions of *anger*, *joy*, and *sad,* no matter what the intended emotion was in the database. The speech materials were the three emotions *anger*, *joy* and *sad* out of the database, which are referred to as *dataset 1.* For each emotion component, the evaluation score ranged from 1 to 5. For any specific component, score 5 means "emotion strongly perceived", 4 is "emotion perceived", 3 is "emotion perceived somewhat", 2 is "emotion not clear", and 1 is "no emotion perceived".

The subjects who participated in experiment 1 were from three countries, Japan, the United States, and China. Japanese subjects were 17 male graduate students living in Ishikawa prefecture, Japan. American subjects were 11 male and 4 females living in South Dakota, United States; and Chinese were 13 male undergraduate students living in Beijing, China. None had experience living abroad.

In Experiment 2, the database (comprised of the four intended emotions), referred to as *dataset 2,* was evaluated using six emotions of *anger*, *joy*, *sad*, *fear*, *surprise*, and *disgust*. The evaluation method for each emotion was the same as in the first experiment. Exp 2 was conducted with Chinese and Japanese subjects, where the Chinese subjects were the same as that participating in Exp 1. Japanese subjects were 13 male graduate students living in Ishikawa prefecture, Japan, who are different from the subjects who served in Exp 1. To guarantee the comparability when switching the category number from three to six, both Exp 1 and Exp 2 were carried out on these 13 Japanese subjects.

3. ANALYSES OF PERCEPTION

The speech materials of the database were evaluated by one (emotion) dimension evaluation [6]. To avoid potential artifacts caused by the forced selection, we evaluated the speech materials in three and six emotions and compared the results obtained from the different evaluation conditions. The common factors were investigated using the principle component analysis (PCA).

3.1. Evaluation of intended emotion in multiple emotion dimensions

Before answering what the common factors are in emotion perception, we first clarify the difference between one dimension evaluation (ODE) and multiple dimension evaluation (MDE) of emotions using the paradigm of Exp. 1. Figure 1 shows evaluation results for the intended emotional speech of *anger*, *joy*, and *sad*. One can see that the evaluations have quite large variations. To evaluate the identification rate, we suppose that the intended emotion is identified if an utterance has an evaluation score of 4 or 5 on the intended emotion. The results show that for Japanese, about 66% of the intended *sad* utterances were identified, while about 40% of *anger* and *joy* were identified. The identification rate was less than 40% for American and Chinese subjects for all three intended emotions. The identification rate is slightly higher for native-language listeners than for the non-native ones. This tendency is similar to that pointed out by Shigeno [4] and Nakamichi et al [7] . In MDE, however, the difference is not significant between the native and non-native subjects.

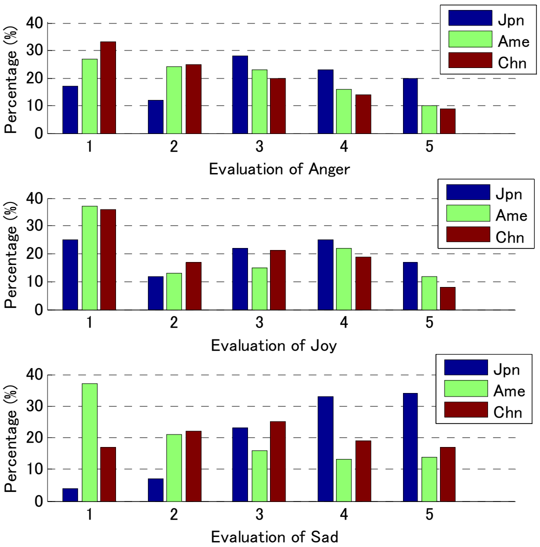

Figure 1: Evaluation results of each intended emotion based on multiple dimension evaluation (MDE).

3.2. Cross evaluation between the cultures

To clarify tendencies of the emotion perception differences between intended and perceived emotions, we quantify the distribution of the intended emotion against the perceived emotions using equation (1). Here, we exemplify the quantification of the relation between the intended emotion and the perceived emotions using the intended emotion, indicated by *I*.

$$D_I(k) = \frac{\sum_{i=1}^{5} i \cdot m_{P/I}(i,k)]}{\sum_{i=1}^{5} m_{P/I}(i,k)} \qquad (1)$$

where $P=\{A, J, S\}$, $I=\{A, J, S\}$, and $P\neq I$. A, J, and S stand for *anger*, *joy*, and *sad*. $m_{P/I}(i,k)$ is the number of the subjects who perceived an utterance with intended emotion of I as P, and give score i, while the utterance is given score k as the intended emotion I. $D_I(k)$ is the average score for the counterpart emotions when the intended emotion I is scored as k. The ratio $R_I(k)$ of the number of perceived emotion different from the intended one to the utterance number is calculated by the following equation, where N_I is the utterance number for a given intended emotion I.

$$R_I(k) = \frac{1}{N_I} \sum_{\substack{P=\{A,J,S\} \\ P\neq I}} \sum_{i=2}^{5} m_{P/I}(i,k), \ \ I=\{A,J,S\} \qquad (2)$$

Figure 2 shows the tendencies of the three countries' subjects, where $D_I(k)$ is plotted by the lines and $R_I(k)$ is illustrated by the bars. One can see that averagely about 30% of the utterances was perceived to have no component of the intended emotion. As the evaluation score of intended emotions gets lower, the average score for the counterpart emotions increases. This tendency is common for the three cultures in general. Japanese subjects have less divergence in perceiving the intended *sad* utterances, while American subjects perceived about 40% of the intended *sad* utterances as other strong emotions. Many utterances with the intended emotions of *Anger* and *Sad* are perceived as *Joy*, indicated by a bundle of sold lines and dotted lines in upper and lower panels, respectively. For the intended emotion of *Joy*, however, there is no dominance shown in the perception between the counterpart emotions.

The results show that when the intended emotion is strongly perceived, in general, the utterances will not be perceived as other emotions. For Japanese subjects, however, about 10% utterances with a score of 5 were perceived as other emotions. One possibility for this phenomenon is that Japanese have more detailed emotional categories for these utterances than the non-native subjects.

Sakuraba et al. evaluated this database using American and Japanese subjects in ODE [6]. Their results showed that the identification rate was about 80% for both American and Japanese subjects in the forced selection. The results obtained using MDE are much lower than the identification in ODE. This implied that even for most of the intended single-emotion speech utterances, they probably included other emotion components. These results suggest the necessity for emotion researchers to be aware that emotion perception may involve multiple components, even though the intended emotion may be only one.

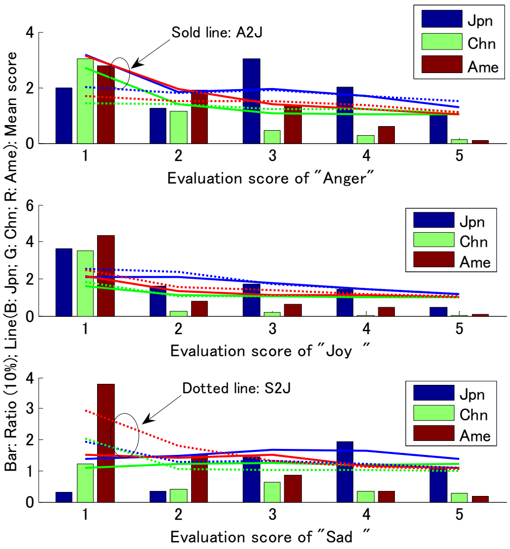

Figure 2: Cross evaluation: the average score of the other emotions *vs.* the intended emotion.

4. COMMON FACTORS IN EMOTION PERCEPTION

In this section, we examine the common factors in perception of emotion by investigating the component loading and emotion vectors in emotion spaces.

4.1. Loading pattern of the explanatory variables

Based on the perception results of *anger, joy* and *sadness* from Exp 1, we used PCA to measure the component loading for each language group. Nine explanatory variables, three emotions by three countries, were used in the PCA. The PCA analysis shows that the first five components can describe about 90% of the variance, while the first three can explain about 74% of the variance. Figure 3 shows the loading pattern of the explanatory variables in the first three components, where J-a, J-j, and J-s denote the explanatory variables for emotions of *anger*, *joy* and *sad* of Japanese subjects. Similarly, A-a, A-j, and A-s are for American subjects, and C-a, C-j, and C-s for Chinese. One can see that the loading patterns of the explanatory variables in the first two components are consistent among the three language groups. The patterns are different in the third principle component.

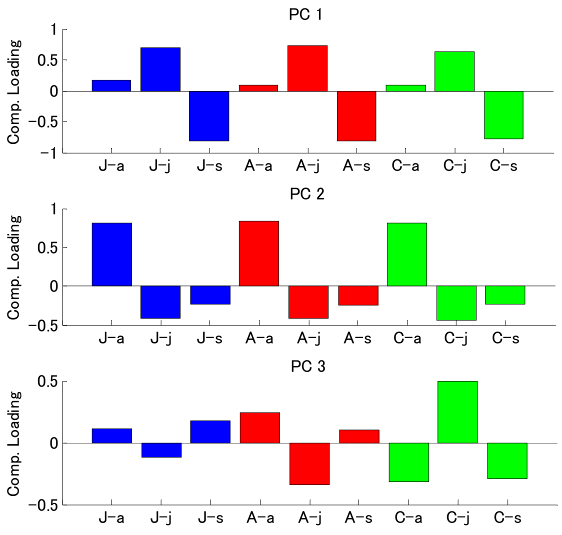

Figure 3: The component loading in the first three components in the three-emotion evaluation.

Focusing on the loading patterns among the countries, it is convenient to divide the nine explanatory variables as three vectors in each loading component according to the countries. We use the following equation to define a similarity between the countries.

$$S(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{\|x_i\| \cdot \|x_j\|}} \qquad (3)$$

where x_i and x_j $(i \neq j)$ represent one of the three vectors of [J-a, J-j, J-s]' [A-a, A-j, A-s]' , and [C-a, C-j, C-s]' , respectively. Table I shows the calculated similarities. As listed in the table, the similarity coefficients between any two countries are larger than 0.99 for PCA component 1 and 2. This implies that the loading patterns are common in the first two components of PCA for the three countries. In contrast, the similarity in component 3 is less than 0.5 between Japanese and American subjects, while they are close to zero between Chinese and the other countries. The difference indicates that the loading patterns in component 3 are independent among the three countries.

Table I. Similarities of the loading patterns between countries in PCA components

	Jpn&Ame	Ame&Chn	Chn&Jpn
PCA1	0.993	0.991	0.992
PCA2	0.999	0.998	0.999
PCA3	0.493	0.016	0.007

Since the first two components explained 67% of the variance, this result implies that about 67% of the emotion perception cues are shared among the three-country listeners in this evaluation. Based on this finding, we might generalize that humans can perceive emotions from only speech sounds, devoid of linguistic information, with about 60% accuracy.

4.2. Emotional vectors in 2D emotion space

We construct a two-dimensional (2D) emotion space using the first two PCA components that can explain 67% of the variance, and then project all the utterances of the dataset 1 into the emotion space. Figure 4 shows the distribution of the emotional speech materials in the 2D emotional space, where the big dots display the data with the maximum score and the small dots display the others. One can see that the basic distribution of the speech materials shaped as a three-pointed star and the speech utterances with a score of 5 are located in the area near the vertices; *anger* at the top, *joy* at the lower right, and *sad* at the lower left.

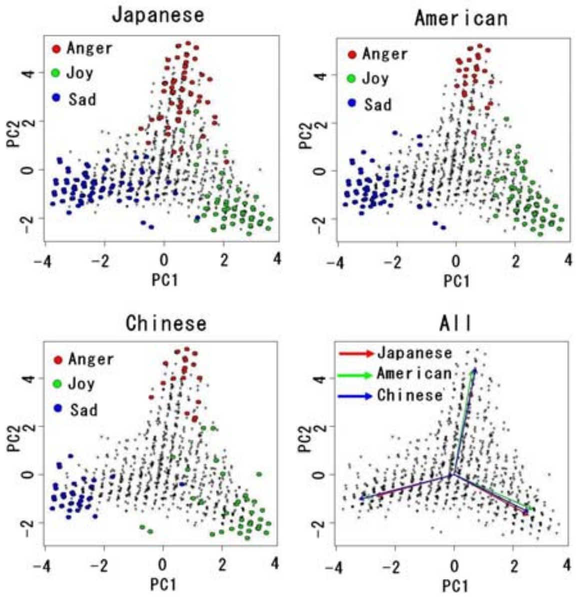

Figure 4: Component loading of the first and second components, (a) Japanese, (b) American, and (c) Chinese. (d) shows the vectors for the three countries.

For convenience, the utterances with the maximum evaluation score are referred to as *pure emotion speech*. The distribution demonstrates a general tendency that the purer the emotion of the utterances, the larger the absolute component loading, while many intended emotional utterances fell in the ambiguous area. Especially for Japanese subjects, some utterances with pure emotion are located in the centroid area, which in fact is where “neutral emotions” would be expected. In contrast, few utterances with pure emotion fell in the ambiguous area for American and Chinese subjects. Perhaps the Japanese listeners were highly attuned to the possible multiplicity of emotion perception when listening to their native language. This needs to be investigated further.

To evaluate the emotion in a low dimensional space, the centroids were calculated for each pure emotion area, and an emotion vector is defined as the vector from the origin of the PCA space to the centroid of each pure emotion. The emotion vectors are plotted in Figure 4 (d). The emotion vectors almost overlap for the three cultures. Table II shows the angles between the emotion vectors. One can see that the angles are consistent with one another for the three cultural backgrounds. This indicates that the structure of the 2D emotion space is about the same for the three cultures.

Table II. The angles between the emotion vectors

Angle (deg.)	Japanese	American	Chinese
Anger-Joy	114	111	113
Anger-Sad	119	116	117
Joy-Sad	127	133	130

5. INVESTIGATION USING HIGHER EVALUATION DIMENSION

For the basic emotions in speech, some researchers use the same three emotions as those used in Exp. 1, while others propose to use six emotions of *anger*, *joy*, *sadness*, *fear*, *surprise*, and *disgust* as the basic emotions. What is the relation between the six emotions and three emotions? For the same emotional data, what would happen if different emotional dimensions are used in the perception experiments? To answer these questions, we designed the second experiment, Exp 2, to investigate the emotion perception when the evaluating categories are extended from the three emotions to the six emotions. The evaluation method for each emotion was the same as in Exp. 1. Exp 2 was conducted with 13 Chinese and 13 Japanese subjects, respectively. (see Section 2.2 for the details)

5.1. Multiple emotions in a lower dimension space

The PCA is applied on the perception data obtained from the six-emotion evaluation. 12 explanatory variables of [J-a, J-j, J-s, J-p, J-f, J-d, C-a, C-j, C-s, C-p, C-f, C-d] were used in the principle component analysis, J and C denote Japanese and Chinese subjects, and a, j, s, p, f, and d represent Anger, Joy, Sad, Surprise, Fear, and Disgust, respectively. Figure 5 shows the loading patterns for the first four components in the six-emotion evaluation. One can see that Japanese and Chinese subjects show very similar loading patterns in the first two components, while different patterns are seen in the fourth component. We treat the patterns of Japanese and Chinese as two six-element vectors, and calculate their similarity using Eq. (3). The similarity between Japanese and Chinese subject are 0.803, 0.816, 0.550, and 0.024 for the components 1 to 4, respectively. This implies that Japanese and Chinese subjects show high similarity in the first two components and the similarity gets lower in the third component. When components are high than 3, it may say that the loading patterns are completely different since their similarities are about 0.05.

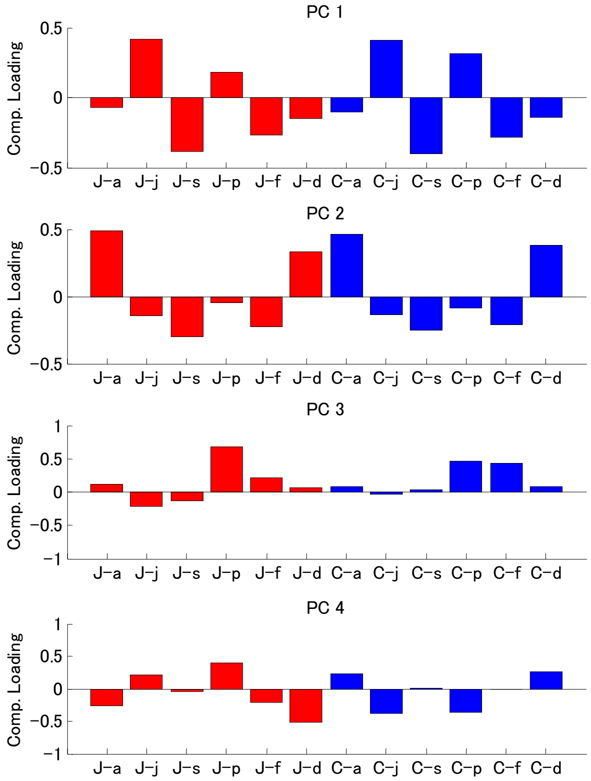

Figure 5: The component loading in the first four components in six-emotion evaluation.

The first four components explained about 61% of the variance, 53% for the three components, and 43% for the first two components. For easy understanding, we use the first two components to display the relation of the emotions in a 2D space. Figure 6 shows the 2D emotion space for Japanese and Chinese subjects, respectively, and projects all of the evaluation values into the space. The utterances with an evaluation score of 5, referred to as *pure emotion*, are plotted as the larger colored dots, while small dark dots display the utterances with other scores. The utterances with pure emotion are located in the extreme positions in the space for Chinese subjects and scattered in relative wide areas for Japanese subjects.

Similar to Figure 4, an emotion vector is defined as a vector from the origin to the centroid point of the scatter area of the pure emotion for each emotion. The emotion vectors are plotted in Figure 6. One can see that the six emotion vectors are bundled as three pairs: *anger* and *disgust; joy* and *surprise*; and *sad* and *fear*. The three pairs are located in the space with about equal intervals. The angles between pairs and within pairs are summarized in Table III, where Anger, Joy, and Sad are used to represent their pairs, respectively. The angles between Anger, Joy, and Sad obtained in the six-emotion evaluation are consistent with the ones obtained in the three-emotion evaluation, whose difference was 5 degrees on the average. The angles within each pair are equal to or less than 4 degrees for both cultures. Based on this analysis, roughly speaking, *anger*, *joy*, and *sad* may be considered as the basic emotions

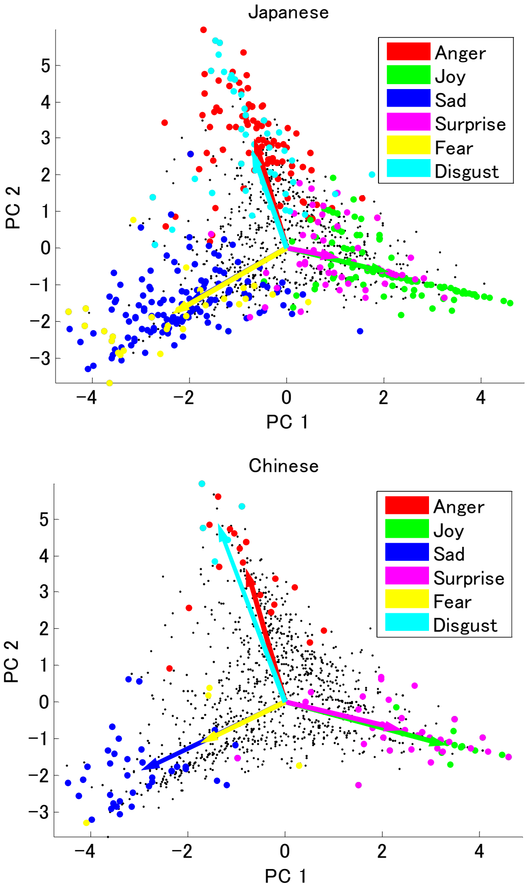

Figure 6: Scatter of the utterances in the emotion space consisting of the first and second principle components.

Table III. Angles between adjacent emotion vectors

	Ang-Joy	Joy-Sad	Sad-Ang	Disg-Ang	Fear-Sad	Surp-Joy
Jpn Sub.	120°	126°	114°	3°	1°	3°
Chn Sub.	122°	128°	110°	4°	1°	2°

To investigate the relations of the three basic emotions and the other emotions, we measure various similarities in the PCA-based spaces between the two emotion groups for the two cultures. The similarities are shown in Table IV. The first measurement is the similarity of PCA-based emotion spaces for six emotions between Japanese and Chinese subjects. The dimensions of the space were changed from one to 12 in the calculations, and the table includes those up to five dimensions. As seen in the table, the similarities are larger than 0.75 between two emotional spaces until 3-D, and 0.57 for 4-D. As the dimension increases, the similarity between the emotion spaces decreases. When the dimension is larger than 5, the similarity is ranged between 0.3 and 0.4. This implies that the two emotion spaces of the Japanese and Chinese subjects are highly consistent with each other.

Table IV. Similarities between the cultures and pairs

*	1-D	2-D	3-D	4-D	5-D
Jpn&Chn	0.803	0.81	0.752	0.574	0.491
**	C1-pair	C2-pair	C3-pair	C4-pair	C5-pair
Jpn	0.866	0.971	0.002	0.11	0.017
Chn	0.833	0.9	0.004	0.875	0.178
***	1-D-pair	2-D-pair	3-D-pair	4-D-pair	5-D-pair
Jpn	0.866	0.925	0.543	0.514	0.442
Chn	0.833	0.876	0.652	0.664	0.618

*Similarity between cultures up to five dimensions
**Similarity between pairs for each component
***Similarity between the spaces consisting of the pairs

As shown in Fig. 6, the six emotion vectors are merged into three pairs. Two indexes are used to investigate the stability of the relation between the basic emotions, *anger*, *joy*, and *sad,* and the additional emotions, *disgust*, *surprise*, and *fear*. The similarity of these two groups is measured in each PCA component for Japanese and Chinese. The results show that the similarity values are larger than 0.83 for the first two components for both two cultures. The similarities are about zero for the third component. For most of the higher components, the similarity is smaller than 0.1, although some larger values can be seen for Chinese subjects. This somehow implies that the third component plays crucial role in distinguishing these two groups.

We constructed subspaces by using these two groups, respectively, where the dimension of the space varied from one to 12. The similarity of the subspaces is shown in Table IV until five dimensions. The similarities of the 2D subspaces are larger than 0.85, while they are around 0.6 for 3D subspaces. For the subspaces with higher dimension more than 5, the similarity is larger than 0.4 for Chinese subjects, while it is around 0.3 for Japanese subjects. The results show that the dimension higher than 5 contributes to distinguishing the emotional pairs.

6. SUMMARIES

In this study, we investigated common factors involved in human emotion perception by comparing the evaluations of people with different language/cultural backgrounds. The common factor obtained from PCA implied that people can perceive emotion from speech sounds sans linguistic information with about 60% accuracy. In the emotion perception, there was a significant difference between single-emotion evaluation and multiple-emotion evaluation. However, there is no significant difference in the PCA-based emotion spaces when extending the evaluation dimension from three emotions to six emotions. It was found that the basic structure of *anger joy* and *sad* has no change, and the three additional emotions merged with those three emotions to be three pairs. The result of the six-emotion perception also showed that similarity of the emotion spaces between the Japanese and Chinese subjects was highly consistent with one another, which does not change even when manipulating dimensions of the spaces.

To further investigate the details, the six emotions were treated as two groups: the basic emotions and additional emotions. The results showed that these two groups have a high similarity in the first two components for both Japanese and Chinese, while the similarity gets much lower for the higher components. When investigating the subspaces constructed by these two groups, it was found that the similarity relation of these two emotion groups not only appeared in a two dimensional space, but also in a higher dimensional space. This implies that these three basic emotions may roughly represent a spread of human emotions. To describe the detailed emotions, of course, additional higher dimensions are required. However, the results from this study suggest that the higher dimension analysis has perhaps less significance compared with the lower dimensional analyses. The preliminary results from this study suggest the possibility that a wide range of human emotions can fall into a rather small subset of basic emotions. Further exploration into expression of human emotion in speech is needed in order to substantiate this finding.

Acknowledgment: This work was partially carried out by Kanae Sawamura in her master studies and was partially done by Xuemei Piao when she was a research student in JAIST. The authors would like to thank them for their contributions. This study was supported in part by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan, and also in part by the Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007-2010):19520371 to the third author.

7. REFERENCES

1. Plutick, R., *Emotions: A psychoevolutionary synthesis*. 1980, New York: Harper & Row.
2. Izard, C., *Human emotions*. 1977, New York: Plenum Press.
3. Sawamura, K., et al. *COMMON FACTORS IN EMOTION PERCEPTION AMONG DIFFERENT CULTURES*. in *International Conference of Phonetic Science*. 2007,8. Germany.
4. Shigeno, S., *Recognition of emotion transmitted by vocal and facial expression: Comparison between the Japanese and the American.* The AGU Journal of Psychology, 2003. **3**: p. 1-8.
5. Erickson, D. and K. Maekawa. *Perception of American English emotion by Japanese listeners*. in *Acoustical Society of Japan, Spring Meeting*. 2001.
6. Sakuraba, K., et al. *Phonetic Constrains of Japanese and English Emotional Expressions in Children: Acoustic Analysis of /pikachu/ in Japanese and English*. in *Technical Report of IEICE*. 2001.
7. Nakamichi, A., et al., *Perception by native and non-native listeners of vocal emotion in a bilingual movie.* Gifu City Women’s College Research Bulletin, 2002. **52**: p. 87-91.