Acoust. Sci. & Tech. **31**, 6 (2010)

PAPER

Comparison of emotion perception among different cultures

Jianwu Dang[1,2,*], Aijun Li[3,†], Donna Erickson[4,‡], Atsuo Suemitsu[1,§], Masato Akagi[1,¶], Kyoko Sakuraba[5], Nobuaki Minematsu[6] and Keikichi Hirose[6]

[1]*Japan Advanced Institute of Science and Technology, Japan*
[2]*Tianjin University, China*
[3]*Institute of Linguistics Chinese Academy of Social Sciences, China*
[4]*Showa Academia Musicae, Japan*
[5]*Dokkyo Medical University Koshigaya Hospital, Japan*
[6]*The University of Tokyo, Japan*

(*Received 4 September 2009, Accepted for publication 2 June 2010*)

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 had no experience living abroad. An emotional speech database without linguistic information was used in this study and evaluated using three- and/or six-emotional dimensions. Principal component analysis (PCA) indicates that the common factors could explain about 60% variance of the data among the three cultures by using a three-emotion description and about 50% variance between Japanese and Chinese cultures by using a six-emotion description. The effects of the emotion categories on perception results were investigated. The emotions of anger, joy and sadness (group 1) have consistent structures in PCA-based spaces when switching from three-emotion categories to six-emotion categories. Disgust, surprise, and fear (group 2) appeared as paired counterparts of anger, joy and sadness, respectively. When investigating the subspaces constructed by these two groups, the similarity between the two emotion groups was found to be fairly high in the two-dimensional space. The similarity becomes lower in 3- or higher dimensional spaces, but not significantly different. The results from this study suggest that a wide range of human emotions might fall into a small subspace of basic emotions.

Keywords: Emotional speech, Emotion cognition, Multiple cultures, Basic emotion, PCA analysis

PACS number: 43.71.Hw, 43.71.Bp, 43.71.An [doi:10.1250/ast.31.394]

1. INTRODUCTION

Speech communication in daily life is for conveying not only linguistic information but also paralinguistic information and nonlinguistic information. The first type of information is discrete categorical information explicitly represented by the written language or uniquely inferred from context. Paralinguistic information can be discrete and continuous information added by the speaker to modify or supplement the linguistic information, while nonlinguistic information is the component that generally cannot be controlled by the speaker, such as the speaker's emotion, gender, and age (cf. [1]). In daily conversation, we have experiences in which we can successfully perceive emotions via speech even if we cannot understand the linguistic meaning, but misunderstandings also occur even when we are confident we understand the emotion. The production and perception of emotional speech are affected, to some extent, by nonlinguistic factors such as language and cultural backgrounds. In this study, we investigate the common factors and differences involved in emotion perception among different languages.

Among studies of the cultural effects on emotion perception, Abelin and Allwood recorded utterances with different expressive emotions from a Swedish speaker, and asked subjects from five countries to judge the emotions [2]. The results showed that emotions were interpreted with different degrees of success depending on the mother tongue of the listeners; native listeners were the most successful. Scherer *et al.* conducted a cross-linguistic study with listeners from nine countries, and reported that effects

*e-mail: jdang@jaist.ac.jp
†e-mail: liaj@cass.org.cn
‡e-mail: ericksondonna2000@gmail.com
§e-mail: sue@jaist.ac.jp
¶e-mail: akagi@jaist.ac.jp

on vocal expression may be motivated in part by universal psychobiological mechanisms, and in part by the segmental and suprasegmental aspects of the particular language [3] and also by cultural differences [4]. Sawamura *et al.* reported that some common loading patterns were observed in their principal component analysis (PCA) of emotion perception for subjects with different cultural backgrounds [5]. Huang reported that the role nonlinguistic information plays in the perception of expressive speech categories was common to listeners of different cultural backgrounds [6].

Emotion in speech is the component related to nonlinguistic information. As mentioned earlier, nonlinguistic information cannot be manipulated consciously. Most existing emotional speech databases are expressive ones performed by actors/actresses. To characterize each emotion, the emotional speech database is constructed by choosing some exaggerated "emotional speech" utterances that have supposedly been uttered intentionally with a certain emotion. However, there are few speech sounds that have only one pure emotion in real-life communication [5,7,8]. Also, as pointed out in previous studies, speech-based emotion cognition is affected by differences among the cultures of the speakers and listeners [9,10]. It has been shown that the identification rate for certain intended emotions may be higher for speakers and listeners who have the same language and cultural background [9,10]. However, there is no answer as to what the common factors are in emotion identification and whether or not there are idiosyncratic differences in listeners with the same cultural background.

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 were asked to evaluate each speech utterance according to three or six emotions, independent of which emotion had been intended by the speaker. Section 2 describes these experiments in detail. In Section 3, the evaluation of the perception results among the different cultures is given. In Section 4, the analysis of the common factors in emotion perception for people with different cultural backgrounds is described. The similarities among the emotion perceptual spaces for the different cultures are investigated in Section 5. In Section 6, a summary and implications of this study are presented.

2. EMOTION PERCEPTION EXPERIMENTS

The purpose of this study is to clarify the common factors and differences among various cultures 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, particularly for cross-language experiments. Because of such a consideration, we chose the database constructed by Sakuraba *et al.* [11].

In the database, 15 Japanese children ranging from 4 to 10 years old were asked to produce the voice of "Pikachu" upon watching an emotional picture of the "Pocket Monster" animation character *Pikachu*. In the animation, Pikachu only says "Pikachu," but if Pikachu is happy, Pikachu says "Pikachu" with a happy voice. If Pikachu is sad, Pikachu says "Pikachu" with a sad voice, etc. Since the children are familiar with the animation, it is expected that they learned the voice by understanding the emotions of the Pikachu character. Thus, the children said "Pikachu" in a way they felt to be appropriate to express the emotional state of Pikachu. Such utterances did not have linguistic information regarding emotion, since the only thing Pikachu says is "Pikachu." This database consisted of the four intended emotions: anger, joy, sadness, and surprise. The numbers of speech utterances were 27 for anger, 28 for joy, 30 for sadness, 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 from the evaluations by the listeners in this study.

2.2. Setup of Experiments

In this study, two experiments were designed. In the first experiment, the subjects were asked to evaluate each of the speech materials according to what extent (using a 1–5 scale, explained below) of anger, joy, and sadness they heard in each speech sound, regardless of the intended emotion in the database. The speech materials comprised the three emotions (anger, joy, sadness) from the database, and are referred to as *dataset 1*. The evaluation score ranged from 1 to 5, where a score 5 meant "emotion strongly perceived," 4 meant "emotion perceived," 3 meant "emotion perceived somewhat," 2 meant "emotion not clear," and 1 meant "no emotion perceived."

The subjects who participated in Experiment 1 (Exp. 1) were from three countries: Japan, the United States, and China. Japanese subjects were 17 male graduate students in their 20s to 30s, living in Ishikawa Prefecture, Japan. American subjects were 11 male and 4 female undergraduate students in their 20s, living in South Dakota, the United States, and Chinese subjects were 6 male and 7 female researchers in their 20s to 40s, living in Beijing, China. None of them had any experience of living abroad.

In Experiment 2 (Exp. 2), the database comprised four intended emotions (anger, joy, sadness, surprise) and is

referred to as *dataset 2*. These four emotions were evaluated in a manner similar to that in Exp. 1, but six emotions (anger, joy, sadness, fear, surprise, and disgust) were used. Experiment 2 was conducted with Chinese and Japanese subjects, where the Chinese subjects were the same as those participating in Exp. 1. The Japanese subjects were 13 male graduate students living in Ishikawa Prefecture, Japan, different from those in Exp. 1. To guarantee the consistency between the results of the 3-category evaluation (Exp. 1) and the 6-category evaluation (Exp. 2), for comparison purposes, we re-ran Exp. 1 with the 13 Japanese subjects who participated in Exp. 2.

3. EVALUATION OF EMOTION PERCEPTION

One of the aims of this study is to investigate the effects of different emotional categories on emotion perception. Even for one intended emotion, it is possible to have multiple emotions. Therefore, a simple forced selection in emotion perception may give rise to some artifacts in the results. It is necessary to investigate the difference in emotion perception caused by the evaluation category and also to analyze the effects on evaluation across different cultures.

3.1. Evaluation of Intended Emotion in Multiple Emotion Dimensions

The database used in this study was evaluated using a one (emotion)-dimensional evaluation [11]. In Exp. 1, we investigate the cultural effects on emotional perception in a one-dimensional evaluation (ODE) and a multidimensional evaluation (MDE). Figure 1 shows perception results for the intended emotional speech of anger, joy, and sadness. Misperception rates are not displayed here. The evaluations show large variations in identifying the emotions. To evaluate the identification rate, we assume that the intended emotion is identified if an utterance is evaluated with the score of 4 or 5 for the intended emotion. As a result, about 66% of the intended sad utterances were identified by Japanese subjects, and about 40% for anger and joy. The identification rate was less than 40% for American and Chinese subjects for all three intended emotions. As shown in Fig. 1, the identification rate is somewhat higher for native-language listeners than for non-native ones, similar to that reported by Shigeno [9] and Nakamichi *et al.* [12]. However, the difference between native and non-native subjects is smaller than that in those reports [9,12]. The lower identification rate in our study may be because the subjects had more choices in MDE than in ODE, which was used in the past studies.

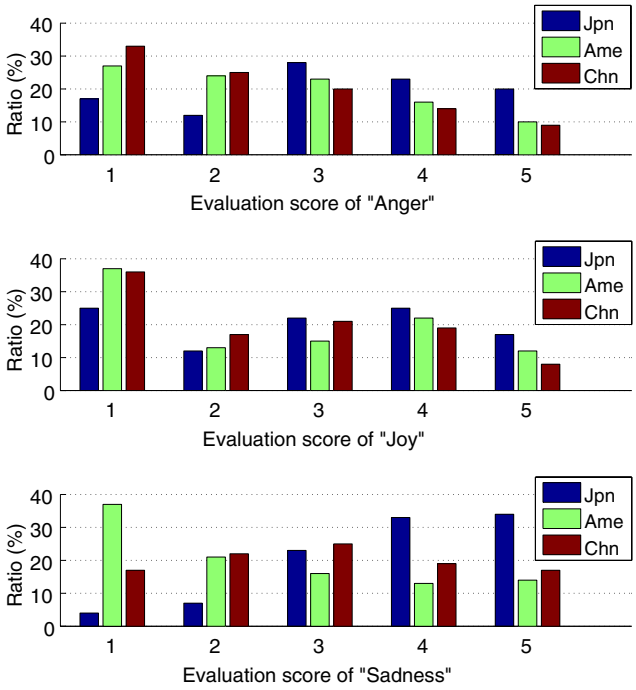

Fig. 1 Results of multiple dimension evaluation (MDE) for each intended emotion.

3.2. Evaluation Across Cultures

It is important to note that (1) the language and the culture of the listeners have a pronounced effect on the identification of emotions, and (2) there is no perfect match between intended and perceived emotion. In fact, as shown in Fig. 1, the match rate, which is the ratio of the number of utterances with a score of 5 to that of the intended emotion, is less than 40% while the unmatched rate is higher than 60%. Accordingly, the perception results can be separated into the matched group and unmatched group. To better understand the differences between the intended and perceived emotions, in this section, we focus on the unmatched group and its relationship to the matched group. The distribution of the unmatched utterances is quantified using Eq. (1). We exemplify the quantification of the relationship between the matched and unmatched groups using the intended emotion, indicated by I, and the perceived emotion, indicated by P. I and P each represent one of the three emotions, anger (A), joy (J), or sadness (S).

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

Here, $m_{P/I}(i,k)$ is the number of subjects who identified the intended emotion of I for a given utterance and rated score k. They also perceived the same utterance as emotion P $(P \neq I)$ and gave score i. Thus, $m_{P/I}(i,k)$ represents all of the unmatched cases. $D_I(k,P)$ is the average score of perceived emotion P for the intended emotion I with an evaluation score of k.

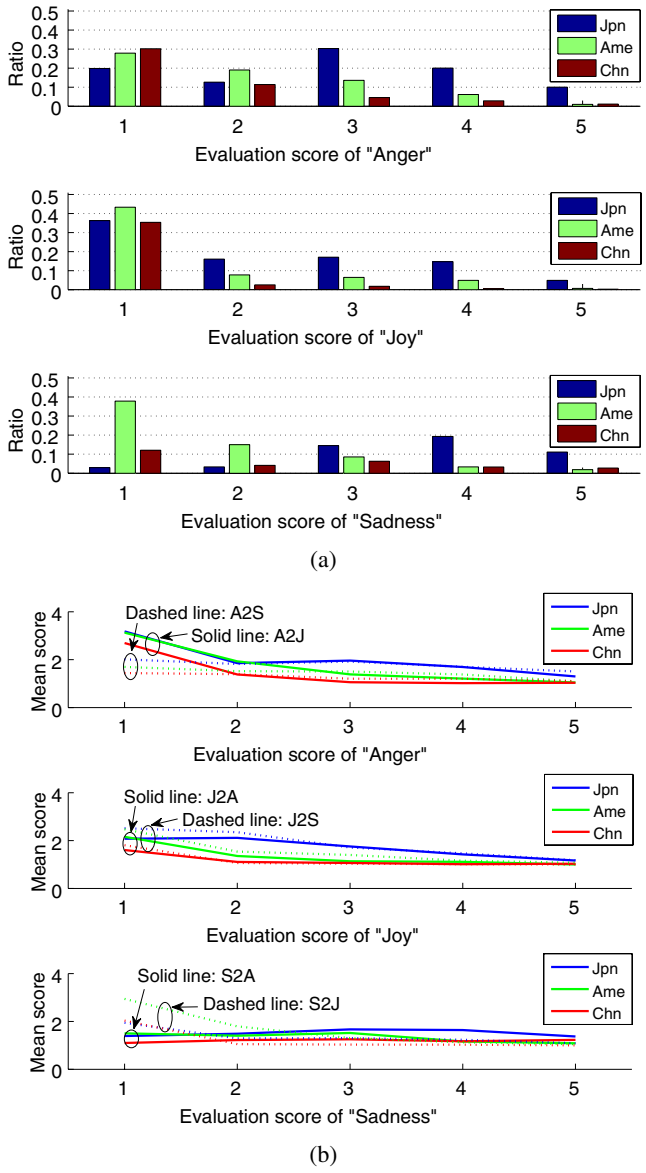

Fig. 2 Identified emotions *vs* unmatched emotions: (a) rate of unmatched utterances, and (b) average score of unmatched utterances. Horizontal axis: evaluation score of intended emotion.

We also introduced one more index to describe the ratio of the number of unmatched utterances to the total number of utterances with a specific intended emotion. The ratio, $R_I(k)$, is calculated using

$$R_I(k) = \frac{1}{N_I} \sum_{\forall P(P \neq I)} \sum_{i=2}^{5} m_{P/I}(i,k), \tag{2}$$

where N_I is the number of utterances for a given intended emotion I. Because Score 1 means no specific emotion, it is excluded from the summation of Eq. (2).

Figure 2 shows the tendencies of the subjects from the three countries. $R_I(k)$ is illustrated in Fig. 2(a). One can see that about 30% of the utterances was perceived to have no component of the intended emotion. As the evaluation score of intended emotions decreases, the average score of the unmatched emotions increases. This tendency is common for the three cultures. For the matched group with scores of 4 and 5, American and Chinese have a lower unmatched rate than Japanese. In particular, in the case of sadness, the lower unmatched rate for matched utterances with the score of 1 indicates that for Japanese subjects, the intended sad emotion is perceived as including other emotions as well. In contrast, for American subjects, about 40% of the intended sad utterances have no "intended emotion," but rather, are perceived as a different emotion.

$D_I(k, P)$ describes the number of unmatched cases, and is plotted in Fig. 2(b), where the label $I2P$ means that the intended emotion I is perceived as emotion P, where I and P have the same definitions as above. It is interesting that a certain number of utterances with the intended emotions of anger and sadness are perceived as joy, as indicated by a bundle of solid lines in the upper panel and dashed lines in the lower panel, respectively. For the intended emotion of joy, however, there is no dominance shown in the perception between the counterpart emotions.

The results suggest that when the intended emotion is strongly perceived, the utterance will not be perceived as another emotion. For Japanese subjects, however, about 10% of utterances with a score of 5 were perceived as other emotions. One possibility for this phenomenon is that, compared with non-native listeners, Japanese have a larger number of categories for these emotions, possibly due to the fact that it is their native language, but possibly also due to the characteristics of the Japanese culture [4].

Sakuraba *et al.* evaluated this database using American and Japanese subjects in ODE [11]. Their results showed that the identification rate was about 70% for both American and Japanese subjects in forced selection. The results obtained using MDE are much lower than the identification rate in ODE. This implies that even for most of the intended single-emotion speech utterances, they probably include more than one emotion component. 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.

4. COMMON FACTORS IN EMOTION PERCEPTION

In this section, we examine the common factors in the perception of emotion, by investigating the eigenvectors and emotion vectors in emotion spaces by principal component analysis (PCA).

4.1. Eigenvectors for Explanatory Variables

From the evaluation experiment, we obtained nine combinations of three emotions from listeners from three

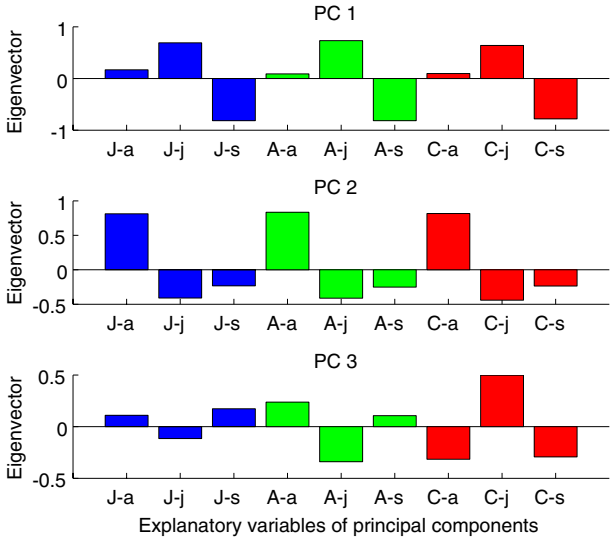

Fig. 3 Eigenvectors in the first three principal components of the evaluation of the three-emotion categories.

countries. PCA is applied on the nine combinations (*i.e.*, nine explanatory variables) to find out the eigenvectors in emotion space. PCA reveals that the first five principal components can describe about 90% of the variance, while the first three can explain about 74% of the variance. Figure 3 shows the eigenvectors for the explanatory variables in the first three principal components, where J-a, J-j, and J-s denote the explanatory variables for the emotions of anger, joy and sadness, for Japanese subjects. Similarly, A-a, A-j, and A-s are for American subjects, and C-a, C-j, and C-s are for Chinese subjects. One can see that the eigenvectors for the explanatory variables in the first two principal components are consistent among the three language groups. The patterns are different in the third principal component.

Focusing on the eigenvectors among the countries, we divide the nine explanatory variables into three vectors in each eigenvector according to the countries. The similarity between the countries is defined by

$$S(\boldsymbol{x}, \boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{\|\boldsymbol{x}\|\cdot\|\boldsymbol{y}\|}}, \tag{3}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ ($\boldsymbol{x} \neq \boldsymbol{y}$) 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 1 shows the calculated similarities. As listed in the table, the similarity coefficients between any two countries are larger than 0.99 for principal components 1 (PC1) and 2 (PC2). This implies that the eigenvectors are common in the first two principal components of PCA for the three countries. In contrast, the similarity in principal component 3 (PC3) is less than 0.5 between Japanese and American subjects, while they are close to zero between Chinese and the other countries. In PC3, the amplitude of the vector for Chinese is higher than those for American and Japanese. These results indicate that PC3 is independent among the three countries.

Table 1 Similarities in eigenvectors between countries in the first three principal components.

	Jpn&Ame	Ame&Chn	Chn&Jpn
PC 1	0.993	0.991	0.992
PC 2	0.999	0.998	0.999
PC 3	0.493	0.016	0.007

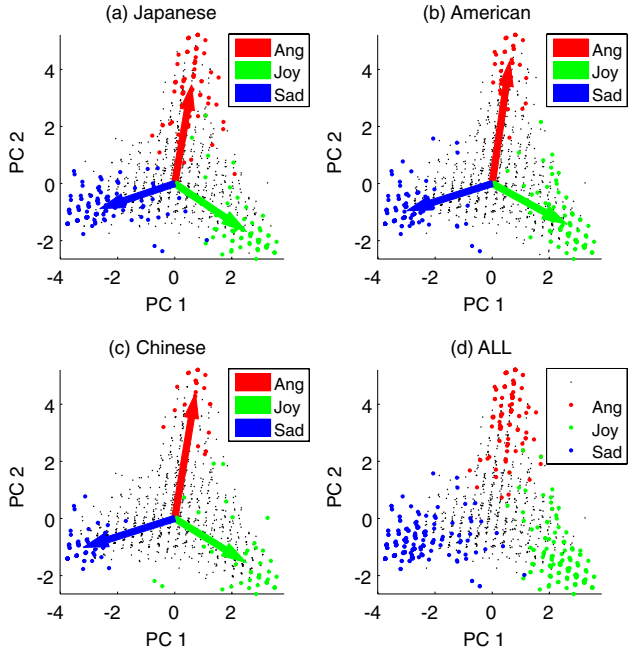

Fig. 4 Component scores for the first and second principal components, (a) Japanese, (b) American, and (c) Chinese. (d) Distribution for all three countries.

Since the first two principal components could explain 67% of the variance, it implies that about 60% to 70% of acoustic cues for the emotional expression of speech devoid of linguistic information is shared among subjects with different cultures.

4.2. Emotional Vectors in 2D Emotion Space

We construct a two-dimensional (2D) emotion space using the first two principal components that explained 67% of the variance, and then project the utterances of dataset 1 into the emotion space. Figure 4 shows the distribution of the emotional speech materials in the 2D emotion space. Panels (a), (b) and (c) show the data for Japanese, American and Chinese, respectively, and (d) is a plot of all data together. The big dots indicate the data with the maximum scores and the smaller dots indicate the others. One can see that the basic distribution of the speech materials resembles a three-pointed star, with the speech utterances having a score of 5 in the area near the vertices:

Table 2 Angles between emotion vectors.

	Japanese	American	Chinese
Anger-Joy	114°	111°	113°
Anger-Sad	119°	116°	117°
Joy-Sad	127°	133°	130°

anger is at the top, joy at the lower right, and sadness at the lower left.

For convenience, the utterances with the maximum evaluation score are referred to as *pure emotion speech*. The distribution demonstrates the general tendency that the purer the emotion of the utterances, the higher the amplitude of the components, although many matched emotional utterances also fell in the ambiguous area. Particularly for Japanese subjects, some utterances with pure emotion fall in the centroid area, which, in fact, is where "neutral emotion" is 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.

We propose a vector structure of emotion to measure the emotion distribution in a low-dimensional space. The emotion vector is defined as the vector from the origin of the PCA space to the centroid point of each pure emotion. The emotion vectors are plotted in Figs. 4(a), (b) and (c). Table 2 gives the angles between the emotion vectors for each cultural background. 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.

5. EFFECTS OF EMOTION CATEGORIES ON PERCEPTION

The number of emotions examined in previous studies varied, e.g., three [5,13], five [14], six [5,15] and ten [16] emotions. A question is what would happen if the number of emotional categories is changed in the perception test? How would this affect the results? To answer these questions, we designed the second experiment, Exp. 2, so that there were 6 forced choices (anger, joy, sad, fear, surprise, and disgust), and we compared the results with those of Exp. 1, where there were 3 forced choices (anger, joy, and sad). The evaluation method for each emotion was the same as that in Exp. 1. Experiment 2 was conducted with 13 Chinese and 13 Japanese subjects, respectively (see Section 2.2 for the details).

5.1. Perception Analysis in PCA-Based Space

PCA was applied to the perception data obtained from the six-emotion evaluation. The 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 PCA. J and C denote Japanese and Chinese subjects, and a, j, s, p, f, and d represent anger, joy, sadness, surprise, fear, and disgust, respectively. Figure 5 shows the eigenvectors for the first four principal components in the six-emotion evaluation. One can see that Japanese and Chinese subjects show very similar eigenvectors in the first two principal components, while different patterns are seen in the fourth principal component. We treat the patterns for Japanese and Chinese as two six-element vectors, and calculate their similarity using Eq. (3). The similarities between Japanese and Chinese subjects are 0.803, 0.816, 0.550, and 0.024 for principal components 1 to 4, respectively. This implies that Japanese and Chinese subjects show high similarity in the first two principal components. The similarity decreases in the third principal component. There are significant differences in principal component 4 and the above components since their similarities are small, about 0.05.

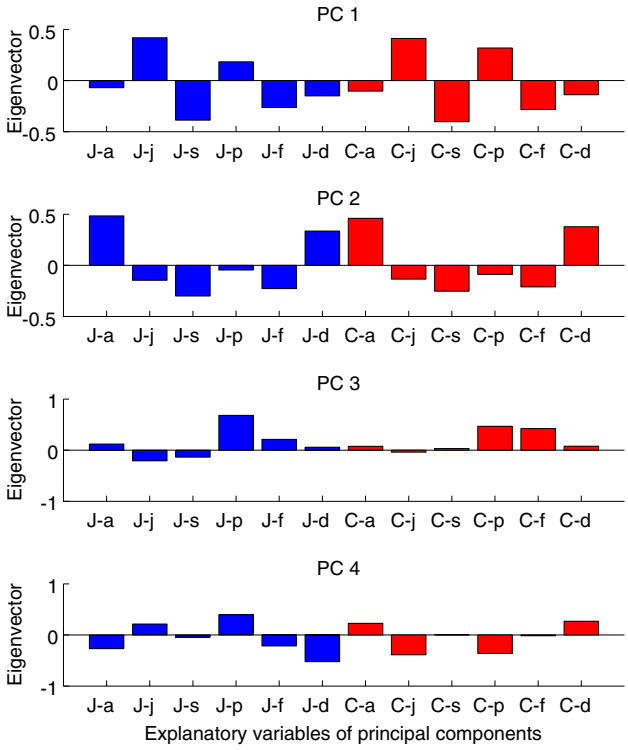

Fig. 5 Eigenvectors in the first four principal components in six-emotion evaluation.

The first four principal components can explain about 61% of the variance, 53% of the first three principal components, and 43% of the first two principal components. For easy understanding, we use the first two principal components to display the relationship of the emotions in a 2D space. Figure 6 shows the 2D emotion space for Japanese and Chinese subjects, and the evaluation

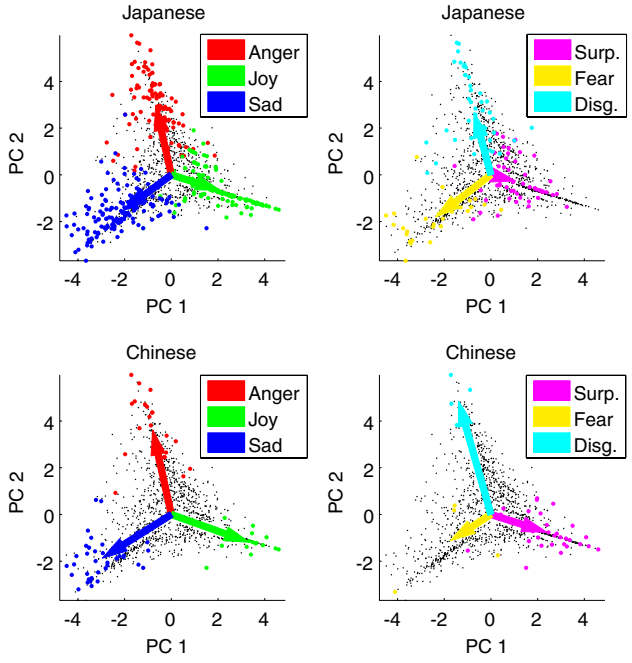

Fig. 6 Scatter of utterances in the emotion space consisting of first and second principal components.

values are projected into the space. One can see that the basic tendency of distribution for the pure emotions with a score of 5 is consistent with that in the three-emotion evaluation. The utterances with pure emotion are located in the extreme positions in the emotion space for Chinese subjects, while they are scattered over relatively wide areas for Japanese subjects. No clear distribution tendency can be seen for PC3 and above.

5.2. Relationships among Emotion Categories

As mentioned with reference to Fig. 4 (Section 4.2), an emotion vector is defined as a vector from the origin to the centroid point of the scatter area of each pure emotion. The emotion vectors are plotted in Fig. 6. One can see that the six emotion vectors are bundled as three pairs: anger and disgust; joy and surprise; and sadness and fear. The three pairs are located in the space at about equal intervals. The angles between pairs and within pairs are summarized in Table 3, where anger, joy, and sadness are used to represent the respective pairs. The angles between anger, joy, and sadness obtained by the six-emotion evaluation are consistent with those obtained by three-emotion evaluation; the difference was within 10 degrees compared with those in Table 2. The angles within each pair are equal to or less than 4 degrees for both cultures.

In general, the emotions of anger, joy and sadness have been employed as essential emotions in most studies. Accordingly, these three emotions are referred to as *basic emotions* hereafter, while the emotions of disgust, surprise and fear are referred to as *additional emotions*. Thus, the six-emotion vector can be separated into two subvectors of the basic emotions and additional emotions. The similarities are investigated under three conditions: similarity of the subspaces consisting of the six-emotion vectors between cultures, similarity between two subvectors of the basic emotions and additional emotions in each principal component, and similarity between the subspaces consisting of the subvectors within each culture. The results for the similarities are shown in Table 4.

Table 3 Angles between adjacent emotion vectors.

	Ang-Joy	Joy-Sad	Sad-Ang	Dis-Ang	Fea-Sad	Sur-Joy
Jpn	120°	126°	114°	3°	1°	3°
Chn	122°	128°	110°	4°	1°	2°

Table 4 Similarities between cultures, between subvectors (SV), and between subspaces consisting of subvectors (SS) of basic emotions and additional emotions.

*	1D	2D	3D	4D	5D
Jpn&Chn	0.803	0.81	0.752	0.574	0.491
**	P1-SV	P2-SV	P3-SV	P4-SV	P5-SV
Jpn	0.866	0.971	0.002	0.11	0.017
Chn	0.833	0.9	0.004	0.875	0.178
***	1D-SS	2D-SS	3D-SS	4D-SS	5D-SS
Jpn	0.866	0.925	0.543	0.514	0.442
Chn	0.833	0.876	0.652	0.664	0.618

*Similarity of subspaces between cultures.
**Similarity of subvectors for each principal component.
***Similarity between subspaces consisting of subvectors.

First, to investigate the relationship between Japanese and Chinese cultures in PCA-based emotion spaces, we calculated the similarity between $(J^1, J^2, \ldots, J^i)$ and $(C^1, C^2, \ldots, C^i)$ using Eq. (3), where J^i = (J-a^i, J-j^i, J-s^i, J-p^i, J-f^i, J-d^i) and C^i = (C-a^i, C-j^i, C-s^i, C-p^i, C-f^i, C-d^i) were obtained from the eigenvector of the i-th principal component, and J, C, a, j, s, p, f, and d are the same as in Section 5.1. The spaces were constructed and investigated from 1 to 12 dimensions (1D∼12D), and the results up to five dimensions are shown in Table 4. Looking at the part marked with * in the table (first row), we can see that the similarities are larger than 0.75 between the two cultures in the emotion spaces up to 3D, and then gradually decrease to 0.491 for the fifth-dimensional space. When the dimension of the spaces is larger than 5, the similarity ranges between 0.3 and 0.4. This implies that up to the 3D emotion space, the perceptions of Japanese and Chinese subjects are highly consistent with each other.

As shown in Fig. 6, the six emotion vectors merge into three pairs. To clarify the relationship between the basic emotions and additional emotions, we began with the investigation of the similarities of the subvectors in each principal component within each culture. For this reason, we constructed the subvectors $Z^i_{\mathrm{ajs}} = (\text{Z-a}^i, \text{Z-j}^i, \text{Z-s}^i)$ and $Z^i_{\mathrm{pfd}} = (\text{Z-p}^i, \text{Z-f}^i, \text{Z-d}^i)$ in i-th principal component, where Z represents either Japanese (J) or Chinese (C), and calculated the similarity of Pi-SV ($i = 1, \ldots, 12$) between Z^i_{ajs} and Z^i_{pfd} using Eq. (3). The results are indicated by ** in Table 4. The similarity is larger than 0.83 in the first two principal components for both cultures, where the second principal component has higher similarity than the first one. The similarities are near zero for PC3. Beyond PC3, the similarity is on the order of 0.1 for both cultural backgrounds, except for a few similarities that have a larger value for Chinese subjects. This implies that PC3 plays a crucial role in distinguishing the basic emotions from the additional ones.

In order to evaluate the contribution of each component to the discrimination of the basic emotions from the additional ones, we constructed two kinds of subspaces $(Z^1_{\mathrm{ajs}}, Z^2_{\mathrm{ajs}}, \ldots, Z^i_{\mathrm{ajs}})$ and $(Z^1_{\mathrm{pfd}}, Z^2_{\mathrm{pfd}}, \ldots, Z^i_{\mathrm{pfd}})$ by adding the subvectors one by one, namely, iD-SS ($i = 1, \ldots, 12$), and investigated the similarity between these two subspaces using Eq. (3). The similarity of the subspaces up to five dimensions is shown in the part marked with *** in Table 4. The similarities of 2D-SS are larger than 0.85, but around 0.6 for 3D-SS. For the subspaces with higher dimensions above 5, the similarity is larger than 0.4 for Chinese subjects, whereas it is around 0.3 for Japanese subjects. The results suggest that the higher dimensions (above 5) do not contribute to distinguishing between the emotional pairs in the subspaces. This means that listeners, both Chinese and Japanese, are sensitive to subtle differences in acoustic information, depending on the given task, e.g., selection in three emotional categories *vs* six emotional categories. However, as evidenced by the high ratings of similarity in the lower principal components, one can see that when listeners are given a single data set and asked to make a three-category or six-category decision, they are able to make broad associations of acoustic cues in order to group the emotional categories into paired emotional categories.

6. SUMMARY

We investigated some common factors involved in the perception of emotion among humans, by comparing the evaluations of listeners with different language/cultural backgrounds. The common factor obtained from PCA implied that people can perceive emotion from speech sounds without linguistic information with about a 60% accuracy in a three-emotion evaluation and about 50% in a six-emotion evaluation. In emotion perception, there was a significant difference between single-emotion evaluation and multiple-emotion evaluation.

When extending the evaluation dimension from three emotions to six emotions, the basic structure of anger, joy and sadness was maintained. Three additional emotions were merged with these three basic emotions to form three pairs in 2D space. The perception results with the six-emotion categories also showed a higher similarity between the Japanese and Chinese emotion spaces.

Moreover, when the six emotion categories were treated as two groups (the three basic emotions and the three additional ones), Japanese and Chinese showed a high similarity for both in the first two principal components, but the similarity became much lower for the higher principal components. For the subspaces constructed with these two groups, the similarity of these two emotion groups was high in the 2D space for both countries. The similarity became lower when adding PC3 in the subspace, and gradually decreased as higher principal components were added one by one. This implies that PC3 plays a crucial role in distinguishing between the pairs in these two groups.

In this study, the perception experiments were carried out only on male Japanese subjects. To examine the effects of gender on emotion perception, we conducted the same experiment on 15 Japanese female subjects whose age ranged between 25–50. The results obtained from these female subjects were consistent with those presented above. This implies that the conclusion obtained in this study is not significantly affected by the gender of the subjects.

The preliminary results of this study suggest the possibility that a wide range of human emotions can fall into a rather small subspace of basic emotions. To describe the emotions in detail, however, more dimensions are required; however, there will still be some ambiguity between the two groups since they also share basic acoustic properties. Further exploration of the expression of human emotion in speech is needed to substantiate this finding.

ACKNOWLEDGMENT

This work was partially carried out by Kanae Sawamura in her master studies and also partially by Xuemei Piao when she was a research student at 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, Culture, Sport, Science and Technology Grant-in-Aid for Scientific Research (C) (2007-2010): 19520371 to the third author.

REFERENCES

[1] H. Fujisaki, "Information, prosody, and modeling: With emphasis on tonal features of speech," *Proc. Speech Prosody 2004*, pp. 1–10 (2004).

[2] A. Abelin and J. Allwood, "Cross linguistic interpretation of emotional prosody," *Proc. ISCA Workshop on Speech and Emotion*, pp. 110–113 (2000).

[3] K. R. Scherer, R. Banse and H. G. Wallbott, "Emotion inferences from vocal expression correlate across languages and cultures," *J. Cross-Cult. Psychol.*, **32**, 76–92 (2001).

[4] K. R. Scherer and T. Brosch, "Culture-specific appraisal biases contribute to emotion dispositions," *Eur. J. Pers.*, **23**, 265–288 (2009).

[5] K. Sawamura, J. Dang, M. Akagi, D. Erickson, A. Li, K. Sakuraba, N. Minematsu and K. Hirose, "Common factors in emotion perception among different cultures," *Proc. Conf. Phonetic Science*, pp. 2113–2116 (2007).

[6] C. Huang, "A study on a three-layer model for the perception of expressive speech," Ph.D thesis at JAIST (2008).

[7] R. Plutick, *Emotions, A Psychoevolutionary Synthesis* (Harper & Row, New York, 1980).

[8] C. Izard, *Human Emotions* (Plenum Press, New York, 1977).

[9] S. Shigeno, "Recognition of emotion transmitted by vocal and facial expression: Comparison between the Japanese and the American," *AGU J. Psychol.*, **3**, 1–8 (2003).

[10] D. Erickson and K. Maekawa, "Perception of American English emotion by Japanese listeners," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 333–334 (2001).

[11] K. Sakuraba, S. Imaizumi and K. Kakehi, "Emotional expression in "pikachuu"," *J. Phonet. Soc. Jpn.*, **8**, pp. 77–84 (2004) (in Japanese).

[12] A. Nakamichi, A. Jogan, M. Usami and D. Erickson, "Perception by native and non-native listeners of vocal emotion in a bilingual movie," *Gifu City Women's Coll. Res. Bull.*, **52**, 87–91 (2002).

[13] Y. Hashizawa, S. Takeda, M. D. Hamzah and G. Ohyama, "On the differences in prosodic features of emotional expressions in Japanese speech according to the degree of emotion," *Proc. Speech Prosody 2004*, pp. 655–658 (2004).

[14] K. R. Scherer, R. Banse, H. G. Wallbott and T. Goldbeck, "Vocal cues in emotion encoding and decoding," *Motiv. Emotion*, **15**, 123–148 (1991).

[15] A. Paeschke, "Global trend of fundamental frequency in emotional speech," *Proc. Speech Prosody 2004*, pp. 671–674 (2004).

[16] L. Leinonen, T. Hiltunen, I. Linnanakoski and M. L. Laakso, "Expression of emotional-motivational connotations with a one-word utterance," *J. Acoust. Soc. Am.*, **102**, 1853–1863 (1997).