COMMON FACTORS IN EMOTION PERCEPTION AMONG DIFFERENT CULTURES

Kanae Sawamura[1], Jianwu Dang[1], Masato Akagi[1], Donna Erickson[2], Aijun Li[3], Kyoko Sakuraba[4], Nobuaki Minematsu [5], Keikichi Hirose[5]

[1]Japan Advanced Institute of Science and Technology, Japan , [2]Showa Academia Musicale, Japan
[3]Institute of Linguistics Chinese Academy of Social Sciences, China,
[4]Kiyose-shi Welfare Center for the Handicapped , [5]The University of Tokyo, Japan
[k-sawamu, jdang, akagi]@jaist.ac.jp, ericksondonna2000@gmail.com

ABSTRACT

There are may exist some common factors independent of languages and cultures in human perception of emotion via speech sounds. This study investigated the factors using 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 3- and 6-emotional dimensions. It was found that most speech materials were perceived to have multiple emotional components, even though a single emotion had been intended to be expressed by the speakers. As the listener's evaluation of the intended emotion gets lower, the components of the other emotions were perceived more strongly. This phenomenon is common across the three cultures. The principle component analysis showed that the loading pattern of the explanatory variables was consistent with one another for the three different cultures at about a 67% cover rate. Extending the evaluation dimension from three emotions to six emotions, it was found that *anger joy* and *sad* may constitute three basic emotions, while the other emotions converge to those basic emotions with about 60% accuracy.

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

1. INTRODUCTION

During daily conversation, we can perceive emotions via speech even if we cannot understand the linguistic meaning. Therefore, there may be some common factors in emotion perception among different language backgrounds. This study attempts to reveal the common factors by means of cognition experiments.

The main stream of studies on emotional speech is to find a set of physical parameters for characterizing each specific emotion. Note that the term "emotion" here is used generically to refer to expressive speech, without drawing a distinction between acted and non-acted emotions. In most of such research, the emotional speech database is constructed by choosing a strongly emotional speech sound supposedly having a single emotion evaluation. However, there are few speech sounds that have only one pure emotion in daily communication [1, 2]. As pointed out in previous studies, speech-based emotion cognition is affected by differences among cultures of the speakers and listeners [3, 6]. 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 about 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 conduct perception experiments on the same database for subjects with different cultural backgrounds. The details of the experiments are described in the 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. [4] 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 ways they felt were appropriate to express the emotion of the emotional picture of Pikachu. Such utterances did not have any linguistic information about 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 emotion obtained by the evaluations of this study.

2.2. Subjects

The subjects participated in this study 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.

2.3. 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. For each emotion component, the evaluation score ranged from 1 to 5. For any specific component, score 5 means "strongly perceived", 4 is "perceived", 3 is "perceived somewhat", 2 is "not clear", and 1 is "nothing perceived". All subjects participated in Experiment 1. The speech materials were the three emotions *anger*, *joy* and *sad* out of the database, which are referred to as *dataset 1*.

In Experiment 2, the subjects were asked to evaluate the whole database (comprised of four intended emotions), referred to as *dataset 2*, using the six emotions of *anger*, *joy*, *sadness*, *fear*, *surprise*, and *disgust*. The evaluation method for each emotion was the same as in the first experiment. Experiment 2 was conducted with only Chinese subjects, the same ones as in Exp 1.

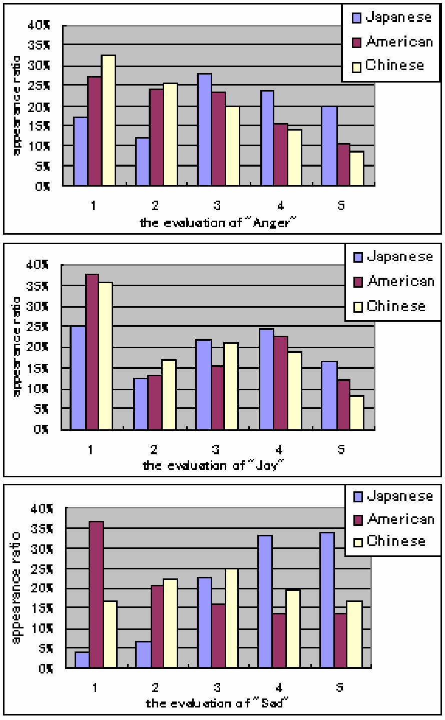
3. ANALYSES OF PERCEPTION

The speech materials of the database were evaluated by one (emotion) dimension evaluation (ODE) [4]. 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 a 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 clarify the difference between ODE and multiple dimension evaluation (MDE) of emotions. Figure 1 shows evaluation results for the intended emotional speech of *anger*, *joy*, and *sad*. Suppose that the intended emotion of the utterances with "score 5" and "score 4" are identified as single emotions, 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 emotions. The identification rate is slightly higher for native-language listeners than for the non-native ones. This is similar to that pointed by Shigeno [3] and Nakamichi et al [5].

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



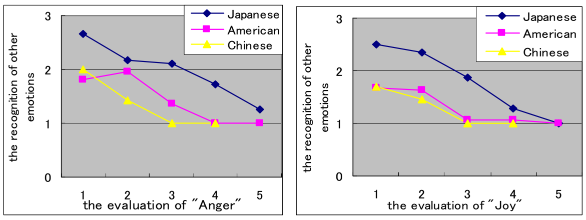
In MDE, however, the difference is not significant between the native and non-native subjects. To clarify tendencies of the emotion perception changes in the intended emotions, we quantify the distribution of the intended emotion against the other emotions using equation (1). Here, we exemplify the relation using the intended emotion *anger* (A)

$$D_A(k) = \frac{\sum_{i=1}^{5} i\cdot[m_{S/A}(i,k)+m_{J/A}(i,k)]}{\sum_{i=1}^{5}[m_{S/A}(i,k)+m_{J/A}(i,k)]} \qquad (1)$$

where $m_{S/A}(i,k)$ and $m_{J/A}(i,k)$ are the number of the subjects who perceived the utterance as *sad*(S) and *joy*(J), respectively, and give score *i* when the intended emotion *A* is evaluated as score *k*. $D_A(k)$ is the average score for the counterpart emotions when the intended emotion is scored as *k*. Figure 2 shows the tendencies of the three countries' subjects for the intended emotions of *anger* and *joy*. As the evaluation of a specific intended emotion decreases, the average score for the counterpart emotions increases. Such a tendency is common for the three cultures. This tendency shows that the emotion space is not a continuum. If the intended emotion is not expressed well it most likely degenerates to be other emotions rather than a neutral one. Japanese subjects showed a larger average score for the counterpart emotions than the other countries. One possibility for this phenomenon is that Japanese have clearer emotional categories for these utterances than non-native subjects.

Sakuraba et al. evaluated this database using American and Japanese subjects in ODE [4]. Their results showed that the identification rate was about 80% for both American and Japanese subjects. The results obtained using MDE were 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: Examples of the average score of the other emotions *vs.* the intended emotion.



3.2. Analysis of common factors

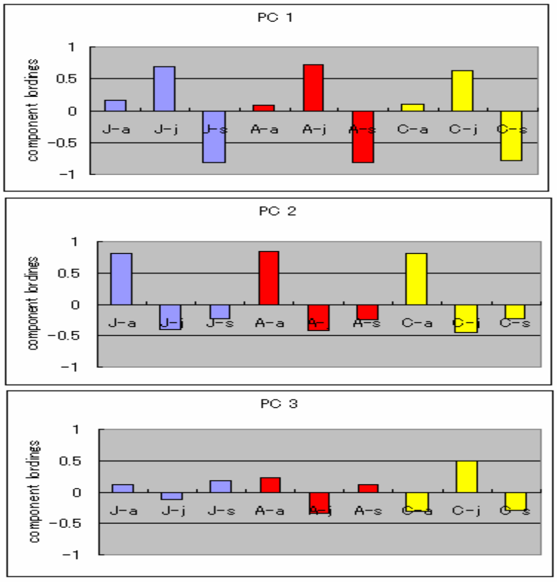
In this section, we try to determine the common factors in perception of emotion by investigating the component loading and emotion vectors in emotion spaces.

3.2.1. Loading pattern of the explanatory variables

Based on the perception 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 results show 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, A-s, C-a, C-j, and C-s are for American and Chinese subjects. 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.

The first two components explained 67% of the variance. This implies that about 67% of the emotion perception cues are shared among the three-country listeners. From this, we might generalize to say that humans can perceive emotions from only speech sound, devoid of linguistic information, with about 60% accuracy.

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



3.2.2. Emotional vectors in 2D emotion space

We construct a two-dimension (2D) emotion space using the first two principle 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 scatter of the emotional speech materials in the 2D emotional space, where the big dots indicate the data with the maximum score and the small dots display the others. One can see that the basic distribution of all the speech materials shaped as a three-pointed star and the pure emotion speech are located in the area near the vertices, with *anger* at the top, *joy* at the lower right, and *sad* at the lower left. For convenience, the utterances with the maximum evaluation score are referred to as *pure emotion speech*. This 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.

We calculated the centroids for each pure emotion area as well as for all of the utterances. Emotion vectors are formed for the centroid of all utterances to that of each pure emotion, and are plotted in Figure 4 (d). The emotion vectors almost overlap for the three cultures. Table 1 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.

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.

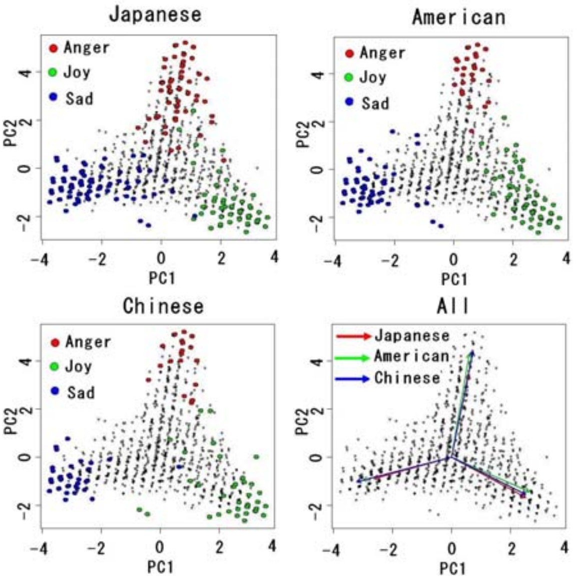


Table 1: The angles between the emotion vectors

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

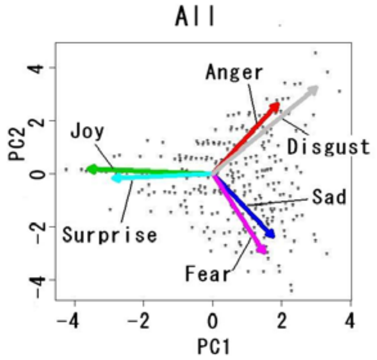
3.3. Multiple emotions in a lower dimension space

For the basic emotions in speech, some researchers use the same three emotions as those in Table 1, while others propose to use the 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? Can the three emotions represent the six emotions, at least to some extent?

To answer these questions, we designed the second experiment using the same Chinese subjects as in Exp. 1, where evaluation emotions are extended from the three emotions to the six emotions.

Using PCA, the explanatory variables are the six evaluation emotions. The first three components explained about 73% of the variance, while the first two covered 58% of the variance. Accordingly, we use the 2D space to display the relation of the emotions. Figure 5 shows the component loading and the emotion vectors for the six emotions. The basic pattern of the component loading does not change with this extension. One can see that with an accuracy of about 60% the six emotions are merged into three directions. *Disgust* is similar to *anger*, *surprise* is close to *joy*, and *fear* to *sad*. In this sense, *anger*, *joy*, and *sad* may be considered as basic emotions. The other three emotions have very high correlation with their counterparts in the 2D dimension. This relation seems to be relatively independent of the data set and evaluation emotions. It implies that these three basic emotions may roughly represent a larger spread of human emotions. To describe the details of the emotions, of course, requires more dimensions.

Figure 5: Scatter of the utterances in the emotion space consisting of the first and second principle components.



4. SUMMARIES

In this study, we investigated common factors involved in human emotion perception via speech sounds for–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. There was a significant difference between one-single evaluation (ODE) and multiple dimensions (MDE) using three emotions. However, extending the evaluation dimension from three emotions to six emotions showed no significant difference. It would seem that *anger joy* and *sad* constitute three basic emotions that encompass a yet undetermined number of other emotions.

Acknowledgment: This study was supported in part by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

5. REFERENCES

[1] C. Izard, Human emotions, New York: Plenum Press. 1977.

[2] R. Plutick, Emotions: A psychoevolutionary synthesis, New York:Harper & Row.1980.

[3] J. Shigeno, “Recognition of emotion transmitted by vocal and facial expression: Comparison between the Japanese and the American”, The AGU Journal of Psychology, vol3., 1-8, 2003

[4] K. Sakuraba, S. Imaizumi, K. Kakehi, D. Erickson, "Phonetic Constrains of Japanese and English Emotional Expressions in Children: Acoustic Analysis of /pikachu/ in Japanese and English" , Technical Report of IEICE, SP 2000-16,47-54,2001.

[5] Nakamichi, A., Jogan, A., Usami, M. and Erickson, D. . Perception by native and non-native listeners of vocal emotion in a bilingual movie. *Gifu City Women’s College Research Bulletin*, **52**, 87-91 (2002).

[6] Erickson, D. and Maekawa, K. (2001). Perception of American English emotion by Japanese listeners. *Acoustical Society of Japan, Spring Meeting*, 333-334.