2018 RISP International Workshop on Nonlinear Circuits,
Communications and Signal Processing (NCSP2018)
Honolulu, Hawaii, USA, March 4-7, 2018

Study on differences between perceptions of Japanese and Chinese emotional speech by Japanese and Chinese listeners

Chenyi Zhang and Masato Akagi

Graduate School of Advanced Science and Technology,
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 932-1292, JAPAN
Email: {s1610126, akagi}@jaist.ac.jp

Abstract

Without understanding of one language, emotional contents of a voice can still be judged by human beings. However, it is reported that differences occur in emotion perception among listeners with different mother languages. Investigating reasons that differences occur may provide a systematic method to the discussions of emotion perception in a cross-language scenario. Therefore, this study discusses commonalities and differences between emotion perception of Japanese and Chinese listeners focusing on semantic primitives (the second layer) in a three-layered model. The results suggest that Japanese and Chinese listeners have commonalities and differences in choosing semantic primitives, and the different usage of semantic primitives may explain the reason why differences occur in the emotion perception of listeners with different mother languages.

1. Introduction

In conversations, linguistic information is very important. However, even without understanding of one language, expressive contents of a voice can still be judged. Therefore, emotional information is one of the most important cues embedded in speech as well as linguistic information. Due to emotion perception in general varied among listeners, particularly for whom speak different mother languages, affective communication is more difficult for humans from different countries [1]. An interesting research topic on investigating differences in emotion perception among humans with different languages has been introduced. Existing studies suggested that differences occur in emotion perception among listeners with mother language and non-mother one. Clarifying reasons for occurring differences may contribute to provide a systematic method to discuss emotion perception in a cross-language scenario.

Differences in emotion perception among listeners using different mother languages have been reported in the emotion dimensional space by previous studies [2] [3]. However, it is still not clear what reasons are contributed to this conclusion. Since human-emotion perception is a multi-layer process [4], a three-layered model has been proposed to compare the human-emotion perception of Japanese and Chinese in terms of Japanese emotional speech [5]. Toward clarify the reasons for differences of emotion perception of listeners with different mother languages, comparisons should be carried out not only within one language but also across another.

In order to solve these problems, this study focuses on the second layer (semantic primitives) in the three-layered model to discuss commonalities and differences between perceptions of Japanese and Chinese emotional speech by Japanese and Chinese listeners and then explains the reasons why the differences occur using multi-dimensional scaling (MDS) and multiple regression analysis (MRA).

2. Experiments

In this paper, to find out what semantic primitives are used by Chinese and Japanese listeners in the perception of expressive speech, two experiments are conducted using speech databases in Chinese and Japanese.

2.1 Databases

Two emotional speech databases in Chinese and Japanese are selected for the listening tests. Chinese database is CASIA database and the Japanese one is FUJITSU database. The utterances used in Chinese have covered four basic emotions, neutral, joy, anger and sad, for each emotion four utterances with different degrees are included. Those in Japanese are covered five basic emotions, neutral, joy, cold anger, sad, and hot anger, and each emotion includes three utterances with different degrees. The total number of utterances are 31.

2.2 Subjects

10 Chinese, whose Japanese level is below JLPT N2, and 10 Japanese, who cannot speak Chinese, graduate students from 23 to 29 years old are asked to participate the experiments.

2.3 Experiment 1

Experiment 1 was conducted to construct a perceptual space of utterances in different emotion categories, and results were analyzed by MDS. The perceptual space constructed was then used in the next experiment to select suitable semantic primitives.

2.3.1 Procedure

This experiment was divided into 2 sessions. Session 1 is the evaluation for 16 × 15 = 240 Chinese utterances pairs and session 2 is for 15 × 14 = 210 Japanese ones. Similarities of emotional speech in Japanese and Chinese are evaluated by Japanese and Chinese listeners. Stimuli were 15 Chinese utterances and 16 Japanese utterances chosen from the databases. The subjects were asked to rate each of the utterances pair on a 5-point-scale according to how similar they perceived them to be. The 5-point-scale is from -2 to 2, including 0, -2 means totally different and 2 means extremely similar. The utterances pairs were randomly presented followed by a pause of 2s through a binaural headphone (STAX SPM-a/MK-2) at a comfortable sound pressure level in a soundproof room. The SSPS MDS ALSCAL procedure was applied to analyze the experimental results.

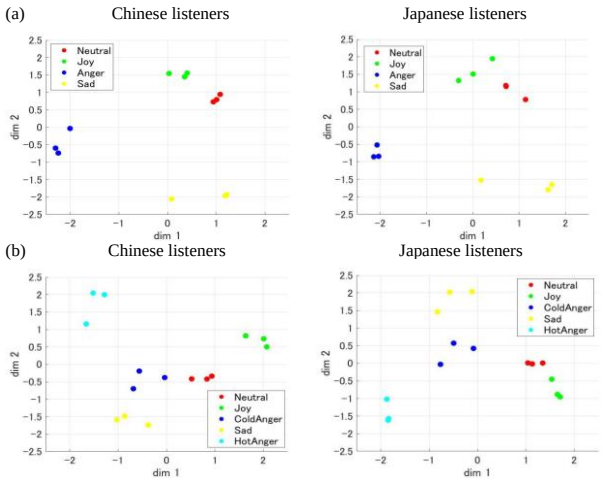

Fig1: Perceptual space of utterances in different categories of emotional speech

2.3.2 Results and discussion

Fig. 1 shows the distribution of utterances in the resulting two-dimensional perceptual space (STRESS value was 20%). Perceived utterances in Chinese are (a) and in Japanese are (b). All categories of emotional speech are separated clearly, and utterances belong to the same category are almost close to each other. The distribution in perceptual space suggests that, considering similarities among utterances, this can be used to determine the semantic primitives suitable for the next experiment.

2.4 Experiment 2

Experiment 2 was carried out to determine suitable semantic primitives for the perceptual space. Semantic primitives are adjectives that appropriate to describe emotional speech according to the previous study. In order to pick up adjectives related to selected utterances from many possible adjectives, a pre-experiment was carried out.

2.4.1 Pre-experiment to select semantic primitives

50 adjectives from the previous studies were selected as candidates. 34 of them were from the previous study by Huang and Akagi, which used to compare Japanese expressive speech perception by Japanese and Chinese listeners [5]. Another 16 adjectives were selected from the previous study of Ueda [6]. In his work, 50 adjectives provided are often used to describe music, therefore adjectives not appropriate for describing emotional speech and ones with similar meanings were removed. In the pre-experiment, stimuli and subjects were the same as Experiment 1. Subjects were asked to hear 15 Chinese utterances and 16 Japanese utterances respectively and to select the adjectives appropriate to describe the emotion of each utterance they heard. The 28 adjectives in Table. 1 had the larger counts than others selected by Chinese and Japanese listeners, therefore those adjectives were chosen for the Experiment 2.

2.4.2 Procedure

This experiment also had 2 sessions. Session 1 is for Chinese utterances and session 2 is for Japanese ones. The stimuli and subjects were the same as Experiment 1. The stimuli were randomly presented to each subject through a binaural headphone at a comfortable sound pressure level in a soundproof room. Subjects were asked to rate each of the adjectives on a 4-point-scale when they heard each utterance, according to how appropriate the adjective was for describing the utterances they heard. About the 4-point-scale, it is from 0 to 3, 0 means very appropriate and 3 means not very appropriate. In order to investigate which adjectives were more appropriate to describe emotional speech and how each adjective was related to each category

Table 1: 28 adjectives chosen form the pre-experiment

ID	Adj. (Japanese)	Adj. (Chinese)	Adj. (English)
1	落ち着いた	沉着的	clam
2	抑揚のある	抑扬顿挫的	well-modulated
3	ゆっくり	缓慢的	slow
4	早い	快速的	fast
5	はっきりとした	清晰的	definite
6	細い	尖细的	thin
7	鋭い	尖锐的	sharp
8	太い	粗犷的	thick
9	声が低い	音调低的	low
10	うるさい	嘈杂的	noisy
11	丸みのある	圆润的	roundish
12	荒っぽい	粗暴的	violent
13	落ち着きのない	慌张的	unstable
14	流暢な	流畅顺畅的	fluent
15	迫力のある	有力的	powerful
16	暖かみのある	温和的	warm
17	きれいな	悦耳的	clean
18	柔らかい	柔和的	soft
19	声が高い	音调高的	high
20	重い	沉重的	heavy
21	明るい	爽朗的	bright
22	軽い	轻快的	light
23	弱い	纤弱的	weak
24	静かな	寂静的	quiet
25	暗い	阴沉忧郁的	dark
26	堅い	生硬的	hard
27	強い	强劲的	strong
28	単調な	单调的	monotonous

of emotional speech, the 28 adjectives were superimposed into the perceptual space built in Experiment 1 by the application of MRA. Since the scaling being used in the evaluation of Experiment 2 has two opposite trends (very appropriate and not very appropriate), for each semantic primitive, it exists a situation that listeners had a neutral point of view with regard to an utterance. Eq. (1) is the regression equation.

$$y = a_1 x_1 + a_2 x_2 \qquad (1)$$

where x_1 and x_2 are the positions (x_1, x_2) of one utterance in the two-dimensional perceptual space, and y is the rating of an adjective for an utterance. Regression coefficients a_1 and a_2 were calculated by performing a least square fit. The multiple correlation coefficients of each adjective were also calculated.

2.4.2 Results and discussion

Fig. 2 is an example of the resulting diagram that presents 28 adjectives plotted in two-dimensional perceptual space for Japanese utterances perceived by Chinese listeners. The utterances are represented by the same form as in Fig. 1 and 28 adjectives are represented by a line in the plot with the same ID numbers as in Table. 1. The directions of the arrowhead of the lines indicate that the adjectives are most related to the utterances. For example, the adjectives *powerful* (ID: 15) was more related to utterances in *hot anger* than to the utterances in other emotions.

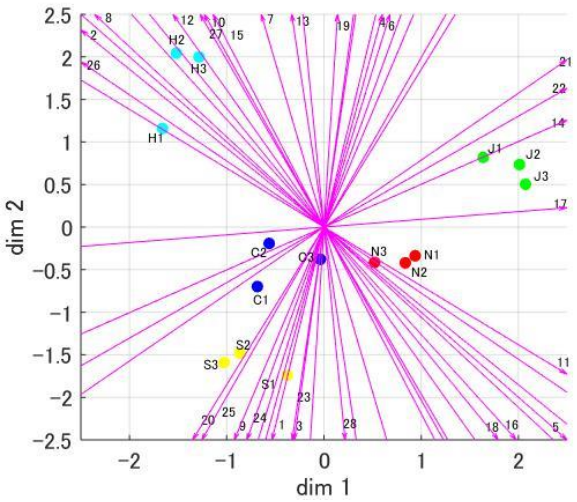

Fig2: Direction of adjectives in two-dimensional perceptual space for Japanese utterances perceived by Japanese listeners

Semantic primitives appropriate to the perceptual space were selected according to three criteria, the directions of 28 adjectives in the perceptual space, the angles between the adjectives, and the multiple correlation coefficient of each adjective. (1) By the directions of the adjectives, it is possible to know that what emotion category the adjectives was related. (2) By the angles of the adjectives, the similarity between each pair of adjectives can be known. The smaller the angle is, the more similar the adjectives are. (3) According to the multiple correlation coefficient of each adjective, it is possible to find the appropriateness of each adjective. The higher the multiple correlation coefficient is, the more appropriate the adjective is to describe the emotional speech. For example, in Fig. 2 *thick* (ID: 8) and *violent* (ID: 12) are both very related to the utterances in *hot anger*, and both of these two adjectives' multiple correlation coefficients are above 0.9. A regression line that represents the relationship of the three utterances in *hot anger* was calculated. Then the angle between *thick* and the regression line and the one between *violent* and the regression line was figured out. Since the angle between

thick and the regression line (1.8°) is smaller than the angle between *violent* and the regression line (9.6°), *thick* was selected in the final list of semantic primitives.

Table 2: The adjectives appropriate to the perceptual space, (a), (b), (c) and (d) represent the uttered language (Chinese or Japanese) / listeners (Japanese or Chinese)

(a) C/JL	(b) C/CL	(c) J/JL	(d) J/CL
clam	clam	clam	clam
heavy	dark	light	fluent
light	light	low	heavy
noisy	monotonous	noisy	noisy
powerful	noisy	powerful	powerful
quiet	powerful	quiet	quiet
sharp	quiet	sharp	roundish
slow	sharp	slow	sharp
soft	slow	strong	slow
strong	strong	thick	strong
thick	thick	unstable	thick
unstable	unstable	violent	unstable
violent	violent	warm	violent
weak	weak	weak	weak

Table. 2 shows the adjectives appropriate to the perceptual space. The adjectives appropriate to the perceptual space for Chinese utterances perceived by Japanese listeners is (a), for Chinese utterances perceived by Japanese listeners is (b), for Japanese utterances perceived by Japanese listeners is (c), and for Japanese utterances perceived by Chinese listeners is (d). According to Table. 2, Chinese and Japanese listeners selected adjectives *clam*, *noisy*, *powerful*, *quiet*, *sharp*, *slow*, *strong*, *thick*, *unstable*, *violent*, *weak* in common, and some adjectives are selected differently, e.g., *soft*, *dark*, *low*, etc..

3. General discussion

Comparison of the results from the above experiments, it is possible to find there are some semantic primitives in common. However, there are also some semantic primitives were used differently between different listeners and different languages. For example, for Chinese listeners, when perceiving Japanese utterances, they described *joy* with *fluent*, when perceiving Chinese utterances, *light* was used to describe it. While for perceiving Japanese utterances, to describe *cold anger*, Chinese listeners used *thick*, Japanese listeners picked up *heavy*. It also happened when Japanese and Chinese listeners perceived their non-mother languages. These differences account for why the differences in emotion perception occur between Japanese and Chinese.

4. Conclusions

In order to clarify the reason that why the differences occur in emotion perception of listeners with different mother languages, the commonalities and differences between perceptions of Japanese and Chinese emotional speech by Japanese and Chinese listeners were discussed. The similarity of each utterances pair and what semantic primitives used by Chinese and Japanese listeners in the perception of Chinese and Japanese emotional speech were investigated by three experiments. According to the evaluation results of similarity of each utterances pair, a two-dimensional perceptual space was built by MDS to determine appropriate semantic primitives. Experiment 2 finds out the usage of semantic primitives by Chinese and Japanese listeners. The results suggest that there are some semantic primitives in common and there are some others were used differently between both listeners and languages. The differences in using semantic primitives could account for the reasons that differences occur in emotion perception of Chinese and Japanese listeners.

References

[1] T. Satou, Y. Endou, M. Kaneko, A. Takeuchi, H. Fujimoto, “Research on Recognition of the Feeling Information Included in Speech,” IEICE technical report, HIP2001-60, pp, 25-29, Dec. 2001.

[2] J. Dang, A. Li, D. Erickson, A. Suemitsu, M. Akagi, K. Sakuraba, N. Minematsu and K. Hirose, “Comparison of emotion perception among different cultures,” Acoust. Sci. & Tech. Vol. 31, no. 6, pp. 394-402, 2010.

[3] X. Han, R. Elbarougy, M. Akagi, J.-F. Li, T.-D. Ngo and T.-D. Bui “A study on perception of emotional states in multiple languages on Valence-Activation approach,” NCSP’2015, pp, 86-89, 2015.

[4] C. Huang, D. Erickson, M. Akagi, “Comparison of Japanese expressive speech perception by Japanese and Taiwanese listeners,” Proc. Acoustics 08 Paris, pp. 2317-2322, 2008.

[5] E. Brunswik, “Historical and thematic relations of psychology to other science,” Scientific Monthly, Vol.83, pp. 151-161, 1956.

[6] K. Ueda, “Should we assume a hierarchical structure for adjectives describing timbre,” Acoust. Sci. & Tech. Vol. 44, no. 2, pp. 102-107, 1988.