ISCA Archive
http://www.isca-speech.org/archive

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

SOUND PERCEPTION BETWEEN TWO LANGUAGES BASED ON ANALYSES OF ONOMATOPOEIC EXPRESSION

Manabu Kotani[1] and Haruya Matsumoto[2]

[1]Faculty of Engineering, Kobe University, 1-1 Rokkodai, Nada, Kobe 657, JAPAN
[2]Osaka Institute of Technology, 5-16-1 Omiya, Asahi, Osaka 535, JAPAN

10.21437/Eurospeech.1995-354

ABSTRACT

This paper describes the study on the difference of sound perception between language speakers through the comparison of onomatopoeic utterances on the analytical basis. Onomatopoeic expressions are the words whose sounds relate to the meaning. They, therefore, may form the key by which the relationship between the sound perception and the sound production is clarified. We pay attention to three kinds of psychological attributes of the sound perception and compare them for the Japanese and Chinese languages for original sounds. The result shows that the manner of the perception of acoustic features differs in some extent between Japanese and Chinese speakers. The investigation and the comparison of the onomatopoeic expressions will be effective for the study of the nature of sound perception of each language speaker.

1. INTRODUCTION

The expressions of sounds such as those of animal cries, which are essentially independent of the language, vary with the language. For Example, the horse cry is generally expressed as "hi hîn" in Japanese, but as "hui hui" in Chinese and "neigh" in English. It is considered that these different expressions are originated from the different sound perceptions depending on the language. Onomatopoeic expressions[1] are the words whose sounds relate to the meaning. They, therefore, may form the key by which the relationship between sound perception and sound production is clarified.

The purpose of this paper is to investigate by analyzing onomatopoeic utterances how the sound perceptions depend on the language. In order to pay attention to three kinds of psychological attributes of sound perception, we examine the physical measures corresponding to the psychological attributes. The measures are the power, the pitch and the spectrum envelope. Here, Japanese and Chinese utterances are compared with the original sounds on the basis of these measures. The LPC (Linear Predictive Coefficients) method[2] is used to extract the feature of the utterances.

2. SOUND DATA

The sound data used here are the onomatopoeic utterances of animal cries and synthesizer's sounds. Table 1 shows the sound data. Animal cries are of a dog, a cat, a horse, a cow and a rooster. These animal cries are widely used as onomatopes. The synthesizer's sounds used here have no meanings, and would be just like the tapping sounds on a table. Three tones of pitches are used in the synthesizer's sounds, high, medium and low.

The data are collected from 10 persons, 3 females and 7 males, for each language. They are native speakers whose ages are from 20 to 40. The procedure for the collection is as follows:

1. A speaker listens to the original sound from the tape recorder.
2. The speaker utters onomatopoeic sounds in their languages, not the mimic ones, for the animal cries, and utters the perceived sound as it is, for the synthesizer's sounds.
3. Return to step 1 until the speaker utters all sound data.
4. Return to step 1 until the speaker utters 5 times.

The utterances are recorded in the digital audio tape-deck at 48 kHz sampling frequency.

3. ANALYZING METHODS

The collected data are down-sampled at 12kHz frequency and pre-emphasized. The Linear Prediction[LP] analysis[2] is carried out every 10.0 ms over Hanning-windowed frames with 21.3 ms width. In order to normalize the period of utterance, we apply the Dynamic Programming algorithm[5].

The physical measures are calculated as follows:

- Power : The amplitude of the logarithmic power is normalized in each utterance.
- Pitch : The pitch is calculated from LPC results. (Refer to [6])

Table 1: Sound data

(a) Animal Cries

Animal	Chinese	Japanese
Dog	Wang Wang	Wan Wan
Cat	Miao	Niao
Horse	Hui Hui	Hi Hîn
Cow	Mou	Mô
Rooster	Wo Wo Wo	Ko Ke Ko Kkô

(b) Synthesizer's Sounds

Symbol	Pitch
X	High
Y	Medium
Z	Low

- Spectrum envelope : The spectrum envelope is also calculated from LPC results. (Refer to [7])

We examine various features to compare the physical measures. The adopted features include the distance (measure of similarity) and the fluctuation of the power, the distance and the fluctuation of the pitch, and the distance of the spectrum envelope.

The distance of the power, D_{power}, is defined as the Euclid distance between the onomatopes and the original sounds.

$$D_{power} = \left(\sum_n (P(n) - P'(n))^2 \right)^{\frac{1}{2}} \qquad (1)$$

where $P(n)$ and $P'(n)$ are the logarithmic power of the sample data and the original data at n-th frame, respectively. Equation (1) would be a measure of similarity of power variations over the period of onomatopoeic sounds. The fluctuation of the power, F_{power}, is defined as the following equation.

$$F_{power} = \frac{\max - \min}{\text{mean}} \qquad (2)$$

where max, mean and min are the maximum, mean and minimum values of the power in each utterance, respectively. The distance and the fluctuation of pitch, D_{pitch} and F_{pitch}, are the same definitions as those of the power.

We use two types of features about the distance of the spectrum envelope between the onomatopes and the original sounds, that is, the ratio of distance and the COSH distance averaged over the period of each onomatopes. The ratio of the distance is the ratio of COSH distance[3] to WLR distance[4]. The COSH distance measure is defined as the following equation.

$$COSH = \sum_n \int_{-\pi}^{\pi} \left(\frac{g(n,\lambda)}{f(n,\lambda)} + \frac{f(n,\lambda)}{g(n,\lambda)} - 2 \right) \frac{d\lambda}{2\pi} \qquad (3)$$

where $f(n, \lambda)$ is the original sound, $g(n, \lambda)$ is the sample sound and λ is the normalized angular frequency. On the other hand, WLR distance measure in which weights is put on the formant is defined as the following equation.

$$\begin{aligned} WLR &= \sum_n \int_{-\pi}^{\pi} \left(\frac{f(n,\lambda)}{u} - \frac{g(n,\lambda)}{v} \right) \\ &\times \left(\log f(n,\lambda) - \log g(n,\lambda) \right) \frac{d\lambda}{2\pi} \\ &+ \frac{v}{u} + \frac{u}{v} - 2 \end{aligned} \qquad (4)$$

where u and v are the powers of the original sound and the sample sound, respectively.

4. ANALYZED RESULTS

Since a speaker utters onomatopoeic sounds five times for each originals, we use the mean values of features for a speaker. The features for a language are the mean values for 8 persons to the exclusion of the maximum and minimum values. Concerning to the pitch information, there are no significant differences between two languages. So, we show the results of the power and the spectrum envelope.

4.1. Power

Figure 1 shows the distance of the power between the onomatopes and the original sounds. The distances of Japanese onomatopes are generally less than those of Chinese onomatopes.

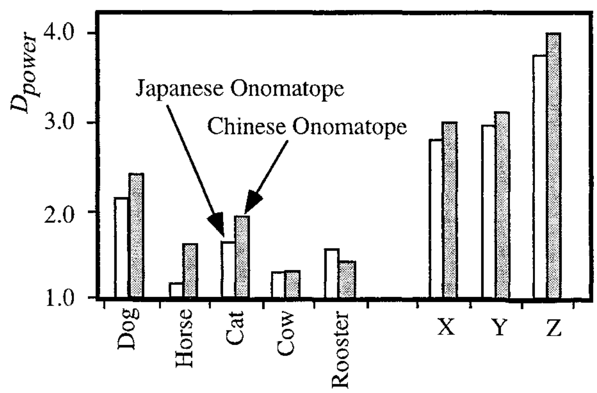


Figure 1: Distance of power between Onomatopes and original sounds

Figure 2 shows the fluctuation of the power. The fluctuations of Japanese onomatopes are closer to those of the original sounds for the animal cries. The fluctuations of Japanese onomatopes are less than those of Chinese onomatopes for the animal cries and the synthesizer's sounds. In other words, Japanese

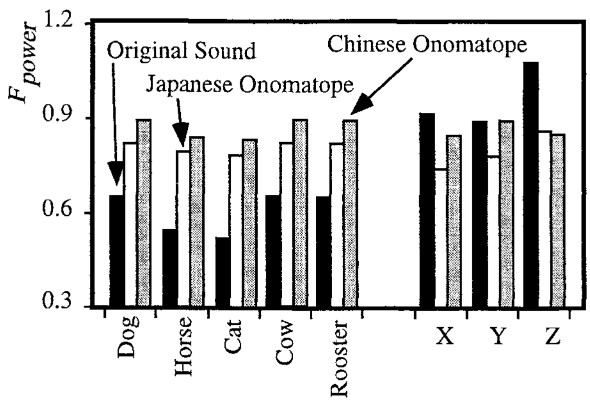


Figure 2: Fluctuation of power

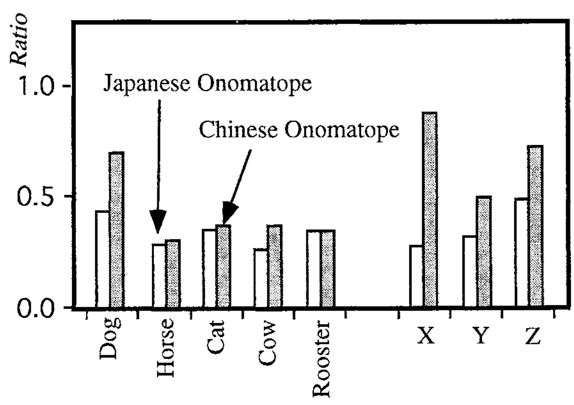


Figure 3: Ratio of COSH distance to WLR distance

utterances are more "flat" in the power variation compared with Chinese utterances.

4.2. Spectrum Envelope

Table 2 shows the distance between onomatopes and original sounds for the spectrum envelope.

Table 2: Distance between onomatopes and original sounds for spectrum envelope

(a) Animal Cries

Measure	Dog	Cat	Horse	Cow	Rooster
COSH	J	J	C	C	J
WLR	J	/	C	C	/

(b) Synthesizer's Sounds

Measure	X	Y	Z
COSH	C	/	J
WLR	J	J	J

"J" indicates that Japanese onomatope is closer to original sound than Chinese one and "C" indicates that Chinese onomatope is closer than Japanese one. "/" indicates that there are no meaningful differences between Japanese and Chinese onomatopes. This table shows that Japanese onomatopes are a little closer to original sounds than Chinese onomatopes, in particular, for the synthesizer's sounds with WLR distance measure.

Figure 3 shows the ratio of COSH distance to WLR distance. WLR measure is weighted in the spectrum peaks, that is, emphasis in comparison is on vowels in speech. When the speaker perceives mainly the spectrum peaks of the original sounds and utters according to the perception, the ratio comes to a small value. Since the distance ratios of Japanese onomatopes are less than those of Chinese onomatopes in Figure 3, it is considered that Japanese speakers attach importance to vowels. On the other hand, Chinese speakers, if anything, attach importance to consonants.

Figure 4 shows the example of the variations of the COSH distance over the period of each onomatopes. For the animal cries, the distances between onomatopes of both languages and the original sounds at the start and the end are less than the other period. For the synthesizer's sound, the distance at the end is less than the other period. Since the distance for the animal cries and the synthesizer's sound are small values at the end, it is considered that Japanese and Chinese speakers exactly perceive the features of the ending of the original sounds. On the other hand, the distances at the start are small values only for the animal cries. It might be said that the speakers can perceive the animal cries, but can't perceive the synthesizer's sounds exactly at the start. The reason is that the synthesizer's sounds are just the sounds, no meaning. It is considered that the differences in the sound perception lead to the result of Figure 4.

5. CONCLUSIONS

We have examined how the sound perceptions depend on the language on the basis of the analysis of onomatopoeic utterances. The following primary results have been obtained.

1. Power:
   Japanese onomatopes are closer a little to original sounds than Chinese onomatopes, and are more flat in the power fluctuations compared with Chinese ones.

2. Pitch:
   There are no meaningful differences between two languages.

3. Spectrum envelope:

(a) Japanese onomatopes are closer to original sounds than Chinese onomatopes for the synthesizer's sounds.

(b) Japanese speakers attach importance to vowels and Chinese speakers, to consonants.

The result indicates that the manner of the perception of acoustic features differs in some extent between Japanese speakers and Chinese speakers. The investigation and the comparison of the onomatopoeic expressions would be effective for the study of the nature of the sound perception of each language speaker.

6. REFERENCES

[1] H. Kakehi : Onomatopoeic Expressions in Japanese and English, Proc. of the XIII-th International Congress of Linguists, 913 - 918 (1983)

[2] J. D. Markel and A. H. Gray, Jr : Linear Prediction of Speech, Springer-Verlag (1976)

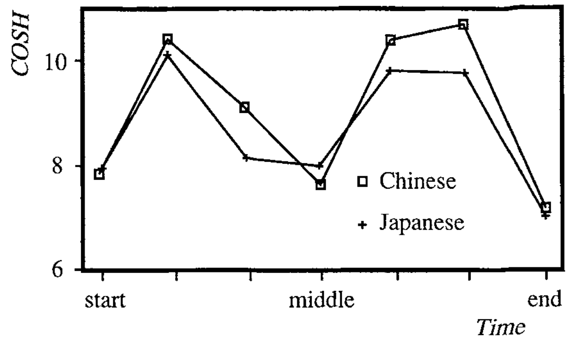
[3] A. H. Gray, Jr and J. D. Markel : Distance Measures for Speech Processing, IEEE Trans. on ASSP, Vol. 24, No. 5, 380 - 391 (1976)

[4] M. Sugiyama and K. Shikano : Frequency Weighted LPC Spectral Matching Measures, IECE Trans., Vol. J65A, No. 9, 965 - 972 (1982)

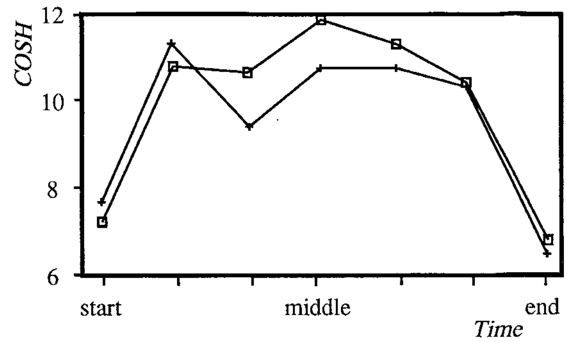
[5] H. Sakoe and S. Chiba : Dynamic Programing Algorithm Optimization for Spoken Word Recognition, IEEE Trans. on ASSP, Vol. 26, No. 1, 43 - 49 (1978)

[6] J. D. Markel : The SIFT Algorithm for Fundamental Frequency Estimation, IEEE Trans. on Audio and Electroacoustics, Vol. 29, No. 5, 367 - 377 (1972)

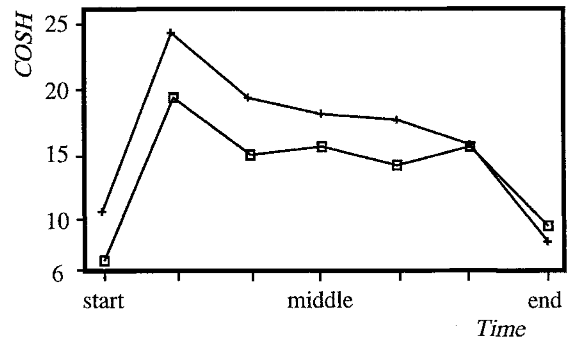
[7] J. Makhoul : Spectral Linear Prediction : Properties and Applications, IEEE Trans. on ASSP, Vol. 23, No. 3, 43 - 49 (1975)



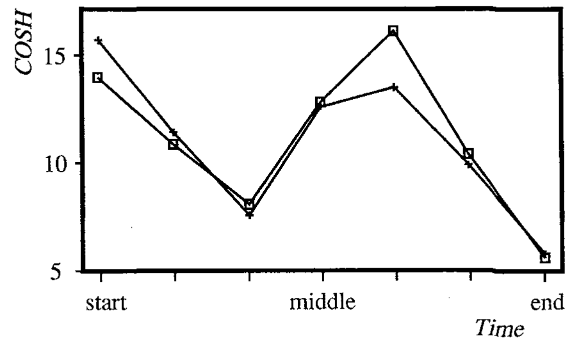
(a) Dog



(b) Cat



(c) Horse



(d) Synthesizer's sound (high pitch)

Figure 4: COSH distance for a period of time each onomatopes