岐阜市立女子短期大学研究紀要第５２輯（２００３年３月）

Perception by native and non－native listeners of vocal emotion in a bilingual movie

Ayako Nakamichi, Aki Jogan, Michie Usami, and Donna Erickson

Abstract

When we make conversation with other people, we usually use both our eyes and our ears. Can people understand feeling only by listening to the voice? We researched this question by using a popular movie. We extracted 5 sentences that expressed 5 feelings. We recorded both the Japanese and English versions onto an MD. For the Japanese version, 19 Japanese listeners and 15 American English listeners were asked to listen to each of the 5 words and to choose one of the following: "surprise," "anger," "doubt," "anxious," and "sarcasm." For the English version, 8 Japanese listeners listened to the perception test using the same procedure as described above. In general, listeners when listening to their native language were able to identify emotions by listening to the voice of the speaker only. It is more difficult for listeners to identify emotions when they are listening to a language other than their own. Acoustic characteristics, such as loudness or final pitch rise, affected the identification of emotion.

1. Introduction

When we make conversation with other people, we usually use both our eyes and our ears. We understand how our companion is thinking by the expression of his or hers words, but also by such things as body movements, facial expressions, changes in the voice, and so on. However, in certain situations, such as when we talk on the telephone, we only use our voice.

Recent research has shown that people can often identify emotion, just by listening to the speaker's voice (e.g., Hayashi, 1998a,b). Acoustic cues are very important for identifying emotions. For instance, for "surprise" often the pitch is high and rising (e.g., Hayashi, 1998a, b). For "doubt" the pitch also rises at the end, similar to the question intonation (e.g., Hayashi, 1998a, b; Maekawa, 1998; and Erickson and Maekawa, 2001). For "anger" often the voice is loud and long on a high pitch (e.g., Williams and Stevens, 1972). "Sarcasm" often ends with a high flat pitch in Japanese (Yanagida, 2001) but a low flat pitch in English (Erickson *et al.*, 2002).

What about understanding emotions in a language other than one's native language? It is difficult especially when one is talking on the telephone, and can only hear the other speaker's voice. Recent research, in fact, has shown that listeners may identify emotions differently depending on their native language. For instance, when Japanese and British listeners listened to "ah" spoken by a Japanese speaker, Japanese speakers identified the "ah" spoken with "surprise" 82% of the time, while British listeners, only 40% of the time (Tomita *et al*, 2001). Another study (Erickson and Maekawa, 2001) showed that both Japanese and English listeners could identify "admiration" in English almost equally well if the pitch contour had a falling pattern. But it the pitch contour was rising, American listeners could still identify it as admiration 100% of the time, yet Japanese listeners only 63% of the time. This research suggests that if the pitch contour is the same as in their native language, listeners can identify emotions well, but if the pitch contour is different, they can not identify them as well.

Today in Japan, we have many things from foreign countries, for example, Western music, foreign films and so on. Often we have a choice of watching the same movie in more than one language, such as Japanese or English or some other language. If we watch the same scene in different languages, we sometimes have a different feeling about the meaning and the emotion, depending on which language we watch the movie in. For example, even the same line may be different depending on whether it is spoken in English or Japanese, because actors may express themselves differently in terms of loudness of voice or whether the pitch is high and low. When we see a movie, we understand the state and emotion of the actor not only by the voice but also by the image on the screen. Is it possible to identify the emotion in the movie by listening to the voice only? And is it possible to identify the emotion from only the voice if one listens to the movie in a language that is not one's native language? Specifically, we want to know if American listeners can understand the emotions only by listening to the vocal portion of a movie in Japanese; also, can Japanese listeners understand the emotions by listening to the same vocal portions in the movie, but in English?

For instance, since the pitch contour for "sarcasm" is different for Japanese and English, we might expect English listeners will have difficulty identifying "sarcasm" when the sentence is spoken in Japanese and similarly, Japanese listeners may have difficulty identifying "sarcasm" when it is spoken in English. However since the pitch contours for "doubt" are similar in both languages, we expect identification will be relatively easy.

2. Methods

In order to answer these questions, we decided to look at the movie, "Harry Potter and the philosopher's stone." The reason we selected this movie is because it is an adventure story, filled with emotions. We think the lines in the movie contain especially rich and clear emotions because they were performed by skilled British actors and actress, and the Japanese voice dubbers were also very skilled.

Data recording

We watched the Japanese language version of the video "Harry Potter and the philosopher's stone" at the Gifu City Women's College library. We selected 5 phrases from the movie that expressed 5 different emotions. We recorded these phrases directly from the video onto an MD player. We also recorded the same sentences from the English version of the movie. However, we were not able to record the English sentence expressing "anxious," because the sentence was spoken too softly in the movie.

Tables 1 and 2 show which sentences were recorded in the Japanese and English versions, respectively. Table 3 describes what was happening in the movie when the recording was made. Based on discussions among ourselves, we decided the emotions expressed in these phrases were the following: "surprise," "anger," "doubt," "anxious," and "sarcasm."

Table 1. Sentences recorded from Japanese version

驚き（surprise)	体が消えた!!
疑問（doubt)	どうした、何が見える？
怒り(anger)	今すぐにここから出て行け!
不安(anxious)	どうしちゃったんだよ～
皮肉(sarcasm)	魔法をかけるの？やって見せて？

Table 2. Sentences recorded from English version

驚き（surprise)	My body is gone.
疑問（doubt)	Tell me, what do you see?
怒り(anger)	I demand that you leave at once.
皮肉(sarcasm)	Oh, are you doing magic? Let's see then.

Table 3. Description of event when we recorded the sentence

驚き (surprise)	Harry (hero of the story) said this when he put on a strange cloak to become an invisible man.
疑問 (doubt)	Voldemort (the arch-enemy of Harry) asked Harry this when Harry looked into the strange mirror (called "The Mirror of Erised"). When a person looked into this mirror, it showed his inward wish.
怒り (anger)	Mr.Duraley (Harry's uncle) said this when Hagrid (the keeper in woods) suddenly broke the door and angrily went into the hut.
不安 (anxious)	Ron (Harry's best friend) said this in the middle of the game (called "Quidditch", something like soccer) when Harry's broom didn't work
皮肉 (sarcasm)	Harmione (Harry's best friend) said when Ron cast a spell over a mouse named Scabbers in the train bound for Hogwarts School of Witchcraft and Wizardry.

We then recorded the phrases onto the Fujitsu Fmv-6400nu4/L computer, using the software WAVESURFER (www.speech.kth.se/wavesurfer), and edited out the background sound effects as much as possible. We then used the software Psyscope on the Macintosh ibook computer to make the perception tests.

Perception tests

For the Japanese version, 19 Japanese listeners and 15 American listeners were asked to listen to each of the 5 words and to choose one of the following 5 feelings: "surprise," anger," "doubt," "anxious," and "sarcasm." The Japanese listeners were all student volunteers at Gifu City Women's College and were given chocolate candy after the tests. The American listeners were students at Black Hills State University who were paid for their participation. The perception tests were given on the Macintosh computer and listeners used earphones. At first, the listeners were asked to listen to 6 practice questions. After that, they were asked to listen to 5 randomizations of the 5 sentences for a total of 25 sentences. They were asked to indicate which one of the 5 emotions they heard. If they heard "surprise," they were told to type 1, if "angry," type 2, if "doubt," type 3, if "anxious," type 4, if "sarcasm," type 5. Each word was repeated twice, and after the subject typed a number, the next word was automatically played.

For the English version, 8 Japanese listeners (Gifu City Women's College students) listened to the perception test using the same procedure as described above.

3. Results

The results are shown in Figures 1 and 2 below. The x-axis shows the "emotion" The y-axis shows how well each emotion was perceived by listeners. Figure 1 shows the results of the Japanese and American listeners when they listened to the Japanese version of the movie. The line connected with the diamonds shows how well the Japanese listeners identified the five emotions; the line connected with the triangles, shows how well the American listeners identified the emotions. Generally, Japanese listeners were able to identify the emotions expressed in the sentences at least 90% or better. The one exception to this was "doubt" which was identified 73% of the time. Generally, American listeners were able to identify the emotions expressed in the sentences only about 64%-75%. The one exception to this was "surprise" which was identified 87% of the time.

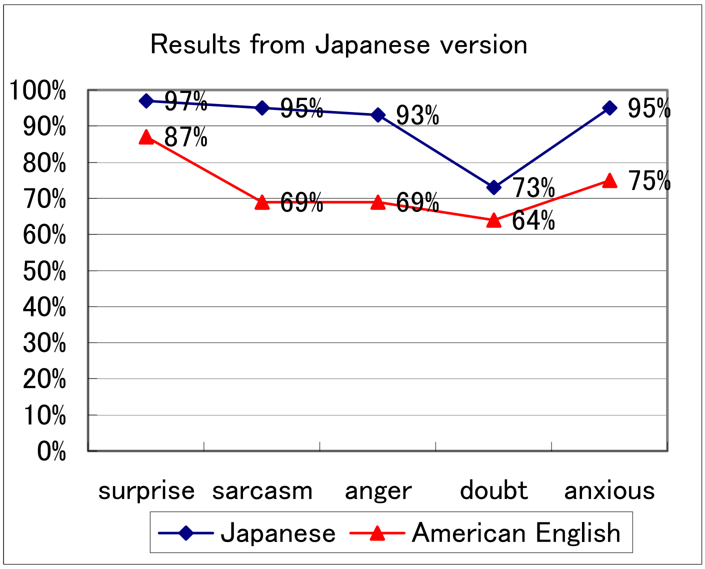

Figure 1. Results of Japanese and American English listeners to Japanese version of the video.

Let us now look at Figure 2 which shows the results of the Japanese listeners when they listened to the English version of the movie. Generally, Japanese listeners were able to identify the emotions expressed in the English sentences only about 65%-70% of the time, similar to how the American listeners did for the Japanese sentences. The one exception was "anger" which was identified 80% of the time.

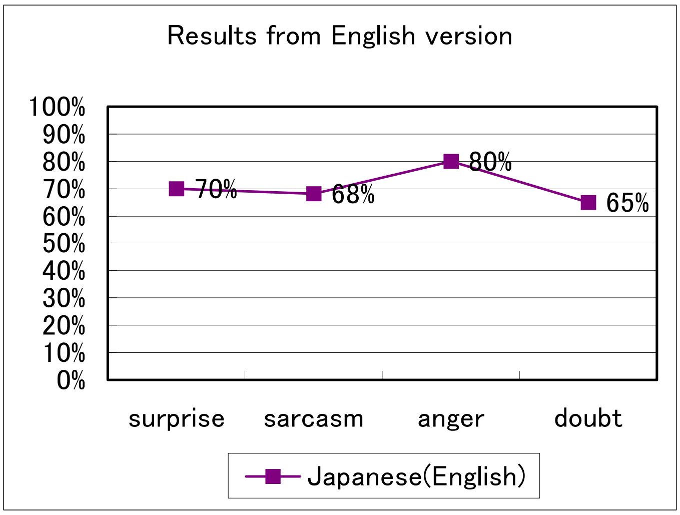

Figure 2. Results of Japanese listeners to English version of the video.

The next three tables are confusion matrixes, which show which emotions were difficult for the listeners to identify, and which emotions they confused them with. Table 4 shows in detail how Japanese listeners identified emotions when they listened to the Japanese version. From Table 4, we see that “doubt” which was identified by Japanese listeners 68% of the time as “doubt” was sometimes identified as “anger” 18% of the time.

Table 4. Japanese listeners-Japanese version

	sur	ang	dou	anx	sar
surprise	97%	0%	1%	2%	0%
anger	3%	93%	2%	1%	1%
doubt	5%	18%	68%	8%	0%
anxious	4%	0%	1%	95%	0%
sarcasm	0%	1%	5%	0%	94%

Table 5 shows how American listeners identified emotions when they listened to the Japanese version. Note that “anger” and “doubt” were often mistaken for each other. “Anger” was identified as “anger” 69% of the time, and as “doubt” 19% of the time. “Doubt” was identified as “doubt” 65% of the time, and as “anger” 20% of the time. “Anxious” was identified as “anxious” 75% of the time, and as “surprise”11% of the time. “Sarcasm” was identified as “sarcasm” 69% of the time, and as “surprise” 12% of the time, and as “doubt” 13% of the time.

Table 5. American English listeners-Japanese version

	sur	ang	dou	anx	sar
surprise	86%	4%	1%	8%	0%
anger	4%	69%	19%	1%	7%
doubt	11%	20%	65%	1%	3%
anxious	11%	7%	3%	75%	5%
sarcasm	12%	3%	13%	3%	69%

Table 6 shows how Japanese listeners identified emotions when they listened to the English version. “Surprise” was identified as “surprise” 75% of the time, and as “anger” 25% of the time. “Anger” was identified as “anger” 80% of the time, and as “surprise” 20% of the time. “Doubt” was identified as “doubt” 65% of the time, and as “surprise” 13% of the time, and as “sarcasm” 15% of the time. “Sarcasm” was identified as “sarcasm” 68% of the time, and as “doubt” 21% of the time.

Table 6. Japanese listeners-English version

	sur	ang	dou	sar
surprise	70%	25%	3%	3%
anger	20%	80%	0%	0%
doubt	13%	8%	65%	15%
sarcasm	8%	3%	21%	68%

4. Discussion

It is interesting that for the Japanese version, both Japanese and American listeners perceived “surprise” best. Maybe this was because “surprise” was spoken on a very high pitch. According to Hayashi (1998a, b) high pitch is an acoustic characteristic of “surprise”. Our results suggest that for English also, then high pitch is a characteristic of “surprise”. We also note that there was a breathless quality toward the end of the utterance which also may have contributed to the impression of “surprise”.

Another interesting observation was that both Japanese listeners did not identify “doubt” very well when they listened to the Japanese version. They often mistook “doubt” for “anger”. This may be because this sentence was spoken in a loud voice, similar to the loud voice characteristic of “anger.” Also, Japanese listeners tended to mistake “doubt” for “sarcasm” and “sarcasm” for “doubt.” This may be because “doubt” and “sarcasm” both ended with a rising pitch. This suggests that acoustic characteristics of the voice play a role in identifying the emotion of an utterance.

It is interesting to note that contrary to our hypothesis, perception of “sarcasm” in the Japanese version of the movie was not more poorly perceived by American listeners than the other emotions; also, perception of “anger” and “doubt” were not better perceived. All emotions expressed in Japanese, except for “surprise,” were equally poorly identified by American listeners.

5. Summary

The results of this research can be summarized as follows. In general, listeners when listening to their native language were able to identify emotions by listening to the voice of the speaker only. It is more difficult for listeners to identify emotions when they are listening to a language other than their own. Acoustic characteristics, such as loudness or final pitch rise, affected the identification of emotion.

Acknowledgements

We thank Kikuo Maekawa for providing the funds to pay the American English listeners. We also thank the students at Gifu City Women College for their kind cooperation in taking the perception test.

References

Erickson, D. and Maekawa, K. (2001). Perception of American English emotion by Japanese listeners. 日本音響学会講演論文集 pp.333-334

Erickson, D., Hayashi, S., Hosoe, Y., Suzuki, M., Ueno, Y., and Maekawa, K.. (2002). Perception of American English sarcasm by Japanese listeners. 日本音響学会講演論文集 pp.277-278

Hayashi, Y. (1998a). F0 contour and recognition of vocal expression of feeling: Using the interjectory word 'eh'. Technical Report of IEICE SP98, 43 pp65-72.

Hayashi, Y. (1998b)「音声に含まれる感性的情報とピッチ曲線　感動詞『ええ』を利用して」－日本音響学会講演論文集　pp.381-382.

Maekawa, K. (1998). Phonetic and phonological characteristics of paralinguistic information in spoken Japanese, Proc. ICSLP98 (CD-ROM), Paper #0997

Tomita, K., Okuta, A., Arai, N., Ikeda, K., and Erickson, D. (2001). Perception of emotion in "Ah": A cross language study. 岐阜市立女子短期大学研究紀要, vol. 51, pp.99-104

Williams, C. and Stevens, K. (1972). Emotions and speech: Some acoustical correlates. Journal of the Acoustical Society of America, vol 52, pp.1238-1250.

Yanagida, M. (2002). Discriminating Ironies from Praising － Acoustic Parameters vs. Prosodic Parameters. Proceedings for 2001 2nd Plenary Meeting and Symposium on Prosody and Speech Processing, pp.143-146

(提出期日　2003 年 3 月 5 日)