258 Perception of American English sarcasm by Japanese listeners

O Donna Erickson, Seiko Hayashi, Yukiko Hose, Mayuko Suzuki, Yuka Ueno (Gifu City Women's College)
and Kikuo Maekawa (National Language Research Institute)

1. Introduction

Being able to correctly assess a speaker's mental attitude is crucial for good human interactions. We have all had the experience of mistaking a companion's mental attitude which has consequently led to misunderstandings and feelings of uncomfortableness, among other things. When we meet a friend, we can understand the mental attitude by both the facial and vocal expressions, as well as the general situation. If we mistake the intended mental attitude, we may correct our impressions from visual cues. But not all human interactions occur in face to face situations; sometimes only vocal cues are available, e.g., when talking on the telephone. To share ideas and feelings with a companion, it is often necessary to understand the mental attitude conveyed by only the voice.

Recent studies have shown that it is possible to perceive various mental attitudes from a single word or short phrase. Hayashi [1] reported that Japanese listeners can identify different mental attitudes, just by listening to the interjection "eh". A similar study showed that both British listeners and Japanese listeners can identify different mental attitudes spoken on the interjection "ah" [2,3]. A primary cue is the shape of the F0 contour. In the latter study, it was also reported that speakers of different language backgrounds may use different acoustic cues, specifically, different F0 patterns for signaling the same mental attitude. In addition, they may have different sensitivities to different aspects of the F0 contour. It also was reported [4] that F0 contours may play a stronger role in conveying mental attitudes for Japanese speakers than for American English speakers.

Thus, vocal expressions of mental attitude may be different even if people are using the same language. In this age of internationalization, being able to communicate well with people from different language backgrounds, including accurately understanding their words as well as mental attitudes, is extremely important.

In this study, we wish to further explore how mental attitude is conveyed through the voice, and specifically, how vocal expression of mental attitude may vary across different language backgrounds.

2. Methods

An American college student (f, 21 yrs old, midwestern dialect) was asked to record the sentence, "That's wonderful" with different mental attitudes—*admiration*, *anger*, *disappointment*, *sarcasm*, and *suspicion*. The mental attitudes were elicited by asking the speaker to first read a short scenario ending with the sentence "That's wonderful." For instance, the dialogue to elicit *sarcasm* was as follows:

"You learn that your best friend is getting married to someone that you really loved, and had hoped to marry. And you say, *That's wonderful.*"

Ten repetitions of each attitude were recorded with a DAT recorder in a sound proof booth at the National Language Research Institute, Tokyo. Perception tests were done with the Psyscope software on a Macintosh G3 Labtop computer. The listeners were 20 Japanese college students (female, 19-20 yrs old) from Gifu City Women's College and 10 English speakers (5 males, 5 females, 22-48 yrs old; 8 Americans, 1 British, and 1 Australian).

Two tokens of five of the 10 repetitions of each utterance type were used for a total of 50 possible responses. Each subject heard the same token of each utterance twice before entering a number from a set of forced choices to stand for the mental attitude heard. For the English listeners the choices were *admiration*, *anger*, *disappointment*, *sarcasm* or *suspicion*. For the Japanese, they were kanshin 感心, ikari 怒り, gakkari がっかり, hiniku 皮肉 and utagai 疑い. A short practice session preceded the test to familiarize the subject with the range of stimuli.

3. Results

F0 analysis

A token F0 contour (for the utterance best perceived by the English listeners) of each of the mental attitudes is shown in the second column of Fig. 1. The F0 contours for *admiration*, *anger*, and *suspicion* all have a rise-fall pattern on "wonderful", with *admiration* having the highest pitch (484.5 Hz), then anger (302.4 Hz) and then *suspicion* (162.7 Hz). Notice that for *suspicion*, there is a substantial lengthening of the final /s/ of "that's". Anger tended to be noticeably louder than the others. The F0 contour on *disappointment* is relatively flat; that of *sarcasm* is a rise on "that's" followed by a drop to a flat level F0 on "wonderful."

Perception test results

The averaged results of the perception tests are shown in the right columns of Fig. 1, with the results of the Japanese listeners in the upper left sections of each column, and those of the English, in the lower right sections. Japanese listeners perceived *anger* and *disappointment* at 95% and 94%, respectively, while English listeners perceived four of the five mental attitudes at 80% or above, but not *suspicion*. With respect to *sarcasm*, about half of the Japanese listeners could understand *sarcasm* (51%) while almost all English could (81%). Interestingly, *suspicion* was difficult to perceive for both English and Japanese listeners.

4. Discussion

In general, the English listeners were able to perceive the mental attitudes intended by the speaker. The one strong exception was *suspicion*, which was also least well perceived by Japanese listeners This may be because the F0 contour of *suspicion* was on a falling contour, rather than a rising contour which is the one often expected for *suspicion* [4]. When the speaker was asked after the experiment why she spoke *sarcasm* with a falling contour, she replied that she thought all the utterances were to be spoken as statements, not questions. The fact that the speaker felt constrained to produce *suspicion* on a falling contour suggests that pitch shape (i.e., falling *vs.* rising) is not necessarily an overriding factor in expressing mental attitude for American English speakers. Although neither Japanese nor English listeners perceived *suspicion* well, English listeners did better than Japanese listeners (50% *vs.* 11%, respectively), further supporting the notion that F0 contour *per se* may be more important to Japanese listeners than English listeners for conveying mental attitudes.

The most interesting results of this study have to do with *sarcasm*. The Japanese listeners showed a poor perception rate for *sarcasm* (51%) while the English listeners showed a relatively good one (81%). This may be because of language and culture differences. English speakers tend to use *sarcasm* frequently where as Japanese tend not to so much and consequently perhaps are not accustomed to *sarcasm*.

However, it may be more complex, related to both the shape of the F0 contour itself, as well as semantic differences between the two languages. For the perception test, there were 5 repetitions of *sarcasm*. Four of these had similar F0 contours—a rising F0 on "that's", followed by a low flat F0 on "wonderful". (The final four utterances in Fig. 2.) These were

poorly perceived by Japanese listeners (45.6% on the average). However, the first repetition (#1) had a different F0 contour: a rise on "that", followed by another rise on "won", and falling on "derful". This particular pitch contour was perceived fairly well by Japanese listeners as *sarcasm* (72.5 %).

In an informal listening task of the two types of F0 contours for *sarcasm*, Japanese listeners felt that the utterances with the low flat F0 contour on "wonderful" conveyed *anger*, while that with the rise-fall F0, conveyed *envy*. American English *sarcasm* perhaps is difficult for Japanese listeners to understand because the connotation of *sarcasm* in the two languages may be different. Also, only when the F0 is a rise-fall pattern is it heard as *sarcasm* by Japanese listeners. However, both a rise-fall and a low flat F0 contour are heard as *sarcasm* by English listeners.

These findings about cross-language differences in the perception of *sarcasm* are new. They invite many more questions: What are the other cues for *sarcasm* in Japanese and English? Also, what are the other cues of *sarcasm* for English listeners, since not all the instances of *sarcasm* have the same F0 contours?

Acknowledgements. We wish to thank Yasuko Nagasaki for her help converting the sound files to Sound Edit 16 format, and Deirdre Smith, for recording the sentences. We also thank the students of the International Cultural Studies Department at Gifu City Women's College as well as the AET teachers in the Gifu area for their help with the perception tests.

References

[1] Tomita, K., Okuta, A., Arai, N., Ikeda, K., and Erickson, D. (in press). Perception of mental attitude in "Ah"": A cross language study. *Gif' City Women's College Bulletin.*

[2] Hayashi, Y. (.., 8). 感動詞「ええ」におけるピッチ曲線と感情認知. 社団法人電子情報通信学会 通学技法, SP98-43, pp. 65-72.

[3] Hayashi, Y. (1999). Recognition of vocal expression of mental attitudes in Japanese: Using the interjection "eh". Proceedings of International Congress of Phonetic Sciences, San Francisco, pp. 2355-2358.

[4] Erickson, D., and Maekawa, K. (2001).Perception of American English mental attitude by Japanese listeners. 日本音響学会議演論文集,春 pp333-4..

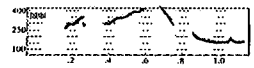
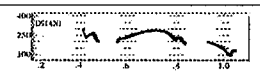
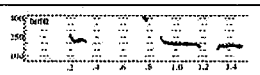
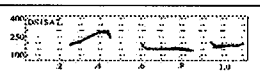
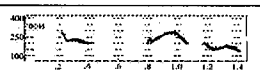
	F0 Contour on *That's wonderful*	ADM	ANG	DIS	SAR	SUS
ADM		64%	2%	12%	12%	11%
		88%	*0%*	*4%*	*3%*	*5%*
ANG		1%	95%	0%	2%	3%
		0%	*92%*	*2%*	*3%*	*3%*
DIS		1%	1%	94%	3%	2%
		1%	*0%*	*88%*	*0%*	*11%*
SAR		1%	10%	6%	51%	32%
		1%	*1%*	*5%*	*81%*	*12%*
SUS		1%	6%	72%	12%	11%
		1%	*0%*	*46%*	*3%*	*50%*

Figure 1. Averaged results (confusion matrix) of perception test with Japanese (upper left numbers) and English (lower right italic numbers). Typical F0 contour shown in column 2.

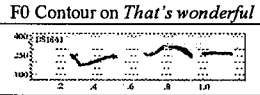
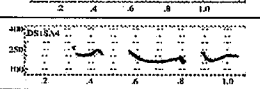
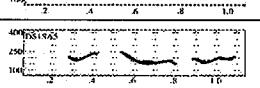
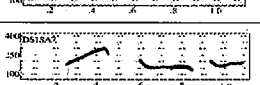
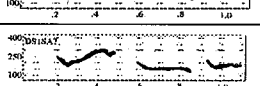
	F0 Contour on *That's wonderful*	ADM	ANG	DIS	SAR	SUS
SAR1		5%	10%	0%	72.5%	12.5%
		0%	*0%*	*5%*	*80%*	*15%*
SAR4		0%	7.5	7.5%	42.5%	42.5%
		0%	*0%*	*5%*	*70%*	*25%*
SAR5		0%	2.5%	15%	45%	37.5%
		0%	*5%*	*5%*	*85%*	*5%*
SAR7		0%	25%	2.5%	42.5%	30%
		5%	*0%*	*0%*	*95%*	*0%*
SAR9		0%	5%	5%	52.5%	37.5%
		0%	*0%*	*10%*	*75%*	*15%*

Figure 2. Perception results (confusion matrix) for each repetition of *sarcasm* by Japanese (upper left bold numbers) and English (lower right italic numbers) listeners. F0 contour shown in column 2.