Some gender and cultural differences in perception of affective expressions

Donna Erickson

Department of International Cultural Studies
Gifu City Women's College, Gifu, Japan
erickson@gifu-cwc.ac.jp

Abstract

This study investigates whether people can understand vocal affective expression in a language that is not their native language, as well as whether there is a difference in the way males and females understand vocal affective expressions. We investigated the affectively-neutral Japanese word /banana/ as uttered with five different affective expressions: *anger*, *sad*, *surprised*, *suspicious*, and *happy*. The listeners were 20 American listeners, 9 Korean listeners, and 20 Japanese listeners who were asked to indicate which affect they heard. The results showed that the perception of affect differed according to the native language as well as to the gender of the listener.

1. Introduction

What a person is intending to say can be conveyed by the choice of word or by the expression of the voice (e.g., "affective (or 'emotional') prosody"). Acoustic cues of affective prosody include F0, intonation pattern, duration/speech rate, loudness, voice quality, and a combination of all of these. Voice quality is one of the most important but perhaps most difficult-to-analyze aspect of expressive speech (e.g., Campbell and Mokhtari, 2003). A definition of voice quality is given in ANSI (1973) as "the quality of a sound by which a listener can tell that two sounds of the same loudness and pitch are dissimilar". Changes in voice quality can signal both paralinguistic information in terms of changes in the speaker's affective state, mood, and attitude to the message and listener, and non-linguistic information in terms of the speaker's social or geographical background and personal characteristics related to the speaker's physique or health (Mokhtari, 2003). (See Fujisaki, 2004, for a more detailed description of paralinguistic vs. nonlinguistic information) A human's ability to "listen-between-the-lines" is heavily dependent on voice quality cues (e.g., Campbell, 2004). Voice quality can be signaled in the acoustic wave form by changes in energy in the spectrum. For instance, a boost-up of energy around 2-3.5 kHz is known to signal the "singing formant" in an opera singer's voice (e.g., Sundberg, 1974). Changes in spectral tilt, such as a steep drop-off of energy (steeper spectral tilt), tend to indicate a breathy voice.

Cross-linguistic studies show that vocal expression of affect may be motivated in part by universal psychobiological mechanisms, and in part by the segmental and suprasegmental aspects of the particular language (Scherer et al., 2001). Interpretation of acoustic characteristics of expressive speech, are affected by the language and culture background of the speaker or listener. For instance, vocal fry (e.g., creaky voice, pressed voice) in Japanese, may convey that the speaker is displaying (not necessarily experiencing) an attitude of being under high pressure, e.g., *suffering* or *admiring* (Sadanobu, 2004), whereas in British English, it tends to signal boredom (Laver, 1980). However, it can also signal *suspicion* in American English (Fujimura and Erickson, 2004) or Japanese (Maekawa, 1998) or *grief* in Russian laments (Mazo et al., 1995).

Gender also influences the production and perception of affective expressive speech. Imaizumi et al. (2004a,b, 2005) used functional MRI to look at brain activity during perception of warmhearted and coldhearted statements spoken with warmhearted and coldhearted prosody. The results showed that male listeners showed significantly stronger activation in the cortical area in the right frontomedian cortex, compared to females, as well as a longer response time. Schirmer and Kotz (2002) investigated processing of words with positive, neutral or negative meaning also spoken with either congruent or incongruent affective prosody. Participants judged the affective prosody while ignoring word meaning, or judged the affective word meaning while ignoring prosody. Event related potentials (ERPs) revealed an interaction of affective prosody and word meaning only in female participants. Male participants showed independent effects for word meaning and affective prosody, suggesting that men process both types of affective information independently. Schirmer et al. (2003) used fMRI to examine brain activity of male and female listeners while listening to semantically positive and negative words spoken with *happy* or *angry* prosody. Their findings of differences in activity in the inferior frontal gyrus (IFG) in male and female listeners are interpreted as evidence that processing of meaning is more influenced by affective prosody in women than in men.

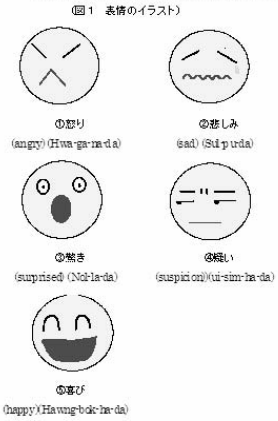
The questions we ask here are (1) does language and culture affect perception of vocal affect? and (2) does gender affect perception of vocally-expressed affect?

2. Method

The utterance "banana" was recorded with 5 different affective expressions by a 20 year old female Japanese speaker, using a Sony MD MZ-R909 MD player with a Audio-technica AT810F headset microphone. The word "banana" was chosen for two reasons: (1) it is an affectively neutral word, and (2) it was originally spoken by the student upon opening the freezer in the research lab, finding a frozen banana, and exclaiming in a surprised tone of voice "banana!" This inspired her producing the Japanese word /banana/ using different types of voices: *angry*, *happy*, *sad*, *surprised* and *suspicious*. Acoustic

analysis consisted of examining F0 contours, F0 maximum and minimum, duration, loudness, and spectral tilt, using Wavesurfer software (www.wavesurfer.com). Perception tests, using a G4 Macintosh ibook computer, psyscope software, and Sennheiser HDA200 earphones, with 4 randomized repetitions of each utterance were done with 49 listeners: 20 Japanese (10 male, 10 female, 18-25 years old, Gifu, Japan); 20 Americans (11 male, 9 female, 22-29 years old, South Dakota); and 9 Koreans (3 male, 6 female, 27-36 years old, Seoul). The listeners were given instructions in their native language (see Fig. 1) and asked to push key 1 if they heard an *angry* tone of voice, key 2 if they heard a *happy* tone of voice, etc.

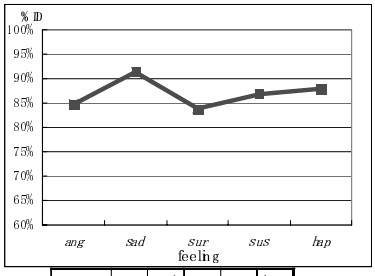
Fig. 1. Perception test instruction sheet



3. Results

3.1. Perception test results

The overall results (Fig. 2) indicate that listeners were able to identify the intended affect at a rate of 85% or better. The best identified was *sad*. There was some confusion between *angry/surprised*, *surprised/happy*, and *suspicious/sad* (as shown

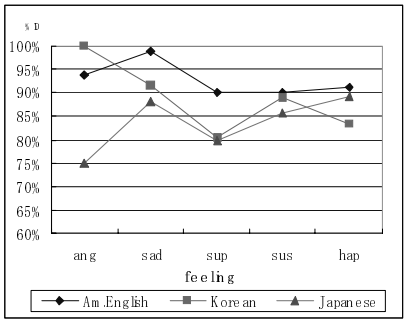


	ang	sad	sur	sus	hap
anger	85%	1%	13%	0%	2%
sad	3%	91%	1%	5%	1%
suprised	3%	1%	84%	2%	12%
suspicion	3%	8%	0%	87%	3%
happy	1%	1%	9%	2%	88%

Fig. 2. Overall identification of affective expressions

3.1.1 Effect of language and culture on identification of affective expression

Identification of vocal expressions was also affected by the language and culture of the listeners (fig. 3). The results show that generally Americans perceived the intended affective meaning best, Japanese, least well, while Koreans identified *anger b*est.

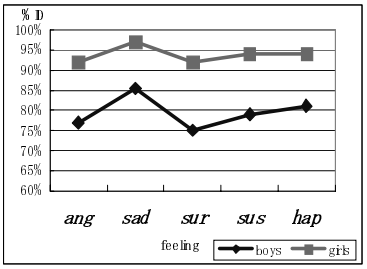


	ang	sad	sur	sus	hap
Am.English	94%	99%	90%	90%	91%
Korean	100%	92%	81%	89%	83%
Japanese	75%	88%	80%	86%	89%

Fig. 3. Identification of affective expression by American, Korean and Japanese listeners

3.1.2. Effect of gender of listeners on identification of affective expression

The gender of the listeners also affected the identification of the intended affect, with women showing a better rate of identification than men (fig. 4).



male

	ang	sad	sur	sus	hap
anger	77%	1%	18%	0%	4%
sad	3%	85%	1%	9%	1%
suprised	3%	1%	75%	2%	17%
suspicion	3%	10%	0%	79%	5%
happy	1%	1%	14%	3%	81%

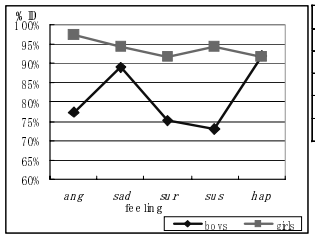
female

	ang	sad	sur	sus	hap
anger	92%	0%	8%	0%	0%
sad	3%	97%	0%	0%	0%
suprised	0%	0%	92%	1%	7%
suspicion	0%	5%	0%	94%	1%
happy	1%	0%	5%	0%	94%

Fig. 4. Identification of affective expression by male and female listeners

3.1.3. Effect of gender with American listeners

Looking at gender differences with American listeners (fig. 5), we see poorer identification of *anger*, *surprise*, and *suspicion* by male listeners than by female listeners.



male

	ang	sad	sur	sus	hap
anger	77%	2%	14%	0%	7%
sad	2%	89%	2%	7%	0%
suprised	2%	2%	75%	2%	18%
suspicion	3%	12%	0%	73%	12%
happy	0%	0%	8%	0%	92%

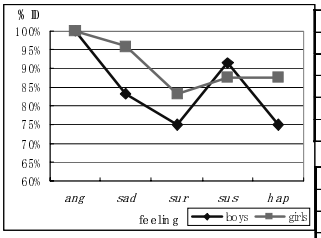
female

	ang	sad	sur	sus	hap
anger	97%	0%	3%	0%	0%
sad	6%	94%	0%	0%	0%
suprised	0%	0%	92%	3%	5%
suspicion	0%	6%	0%	94%	0%
happy	0%	0%	8%	0%	92%

Fig. 5. Gender differences in identification of affective meaning for American listeners

3.1.4. Effect of gender with Korean listeners

Looking at gender differences with Korean listeners (fig. 6), we see generally poorer identification of affective meaning by males than by females, except that *anger* was identified 100% by both male and female listeners, and *suspicion* was slightly better identified by males than females.



male

	ang	sad	sur	sus	hap
anger	100%	0%	0%	0%	0%
sad	0%	83%	0%	17%	0%
surprised	8%	0%	75%	0%	17%
suspicion	0%	8%	0%	92%	0%
happy	0%	0%	25%	0%	75%

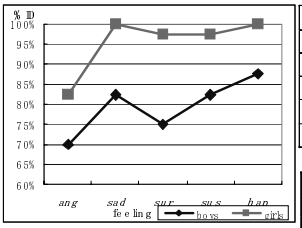
female

	ang	sad	sur	sus	hap
anger	100%	0%	0%	0%	0%
sad	4%	96%	0%	0%	0%
surprised	0%	0%	83%	0%	17%
suspicion	0%	8%	0%	88%	4%
happy	4%	0%	8%	0%	88%

Fig. 6. Gender differences in identification of affective meaning for Korean listeners

3.1.5. Effect of gender with Japanese listeners

Looking at gender differences with Japanese listeners (fig. 7), we see consistently poorer identification by males than by females.



male

	ang	sad	sur	sus	hap
anger	70%	0%	28%	0%	2%
sad	5%	83%	0%	10%	2%
surprised	8%	0%	75%	2%	15%
suspicion	7%	10%	0%	83%	0%
happy	2%	0%	10%	0%	88%

female

	ang	sad	sur	sus	hap
anger	83%	0%	17%	0%	0%
sad	0%	100%	0%	0%	0%
surprised	0%	0%	98%	0%	2%
suspicion	0%	2%	0%	98%	0%
happy	0%	0%	0%	0%	100%

Fig. 7. Gender differences in identification of affective meaning for Japanese listeners

3.2. Acoustic analysis

Fig. 8 shows the F0 contours of each of the affective utterances, along with the duration, F0 max and min, and loudness. The F0 contours for *angry*, *surprised*, and *happy* are somewhat similar, as is the loudness in terms of decibels, and the F0 minima for *sad* and *suspicious* are also similar in that they are much lower than the other utterances, as also is the softness (dB) of these utterances.

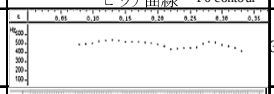
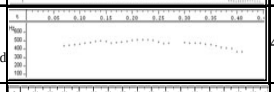
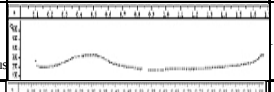
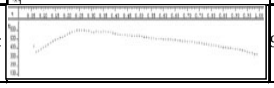
感情	ピッチ曲線 F0 contour	duration 長さ	F0 max F0 高さ	F0 min F0 低さ	loudness 大きさ
①怒り angry		345ms	540Hz	228Hz	60.17 dB
②悲しみ sad		835ms	221Hz	96Hz	31.53 dB
③驚き surprised		421ms	514Hz	370Hz	57.43 dB
④疑い suspicious		593ms	335Hz	87Hz	39.87 dB
⑤喜び happy		948ms	603Hz	356Hz	60.58 dB

Fig. 8. Acoustic characteristics of “banana” utterances

Some comments on voice quality of *anger* and *surprise*: Fig. 9 shows a spectral slice during the final syllable /na/ at about 230 Hz. Especially for the final syllable, *anger* (light line) sounds harsher, and has increased energy around 2-3.5 kHz; *surprise* (dark line) sounds more breathy has a slightly steeper spectral tilt-- -39 dB compared to -41dB for anger.

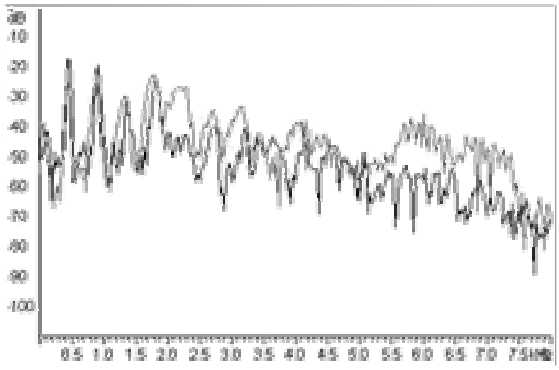


Fig. 9. Spectral slice at 230 Hz during final syllable /na/ for *anger* (light line) and *surprise* (dark line).

3. Discussion

A single word can convey different affective meanings, depending on its acoustic characteristics. The similarity in F0 contours and loudness for the *angry-surprised-happy* utterances may have contributed to listeners’ confusion among these utterances. For those listeners who did not confuse *angry-surprised*, perhaps the differences in voice qualities of the final syllables of these utterances contributed to listeners correctly identifying the intended affect. The similarities in F0 minimum and softness may have contributed to the confusion between *sadness* and *suspicion.* The interesting result of Korean listeners doing better than other listeners in identification of *anger* needs to be further explored. The somewhat surprising result that Americans did better than the other two groups in identifying the affective meaning of the utterances may be related to the fact that Americans did not

recognize the lexical meaning of the Japanese word "banana." The Japanese pronunciation of /banana/ has no prominence on the middle syllable, and no reduced initial or final vowels. Therefore perhaps American listeners did not sense an anomaly between lexical meaning and prosodic meaning. To test this hypothesis, we plan to do similar perception tests with recordings of English "banana" to see if Americans do less well in identifying the affective meaning of the utterance. Why did Japanese do most poorly? Perhaps it is because Japanese listeners recognized the lexical meaning of the word, so sensed an anomaly between the neutral lexical and expressive prosodic meaning, making it more difficult to identify the affective meaning. Why did women do better than men in identifying the intended affect? Perhaps because, according to the studies by Schirmer & Kotz (2002), Schirmer et al. (2004), and Imaizumi et al. (2004a, 2004b, and 2005), who reported that females had an interaction of affective prosody and word meaning, whereas males had independent effects for word meaning and affective prosody, we might surmise that men process both types of affective information independently. Therefore men may have been more sensitive to the anomaly between the neutral word meaning of "banana" and the various affective meanings of "banana", and consequently did less well in identifying the intended affective meaning. We also plan to record a male speaker producing "banana" with various affective expressions, in order to examine whether the gender of the speaker affects perception of affective meaning. Future work will in addition examine how culture, language, and gender of the listener affect perception of voice quality.

4. Conclusion

A single word can convey different meanings, depending on its acoustic characteristics. If utterances have similar F0-characteristics, their affective meanings may be less well-identified. The results from this study also show there are gender and cultural differences in perception of vocally-expressed affect. There is an interesting interaction between lexical and prosodic meaning--which also seems to be affected by culture and gender of listener. The results of this study suggest that if the listeners perceive the lexical and prosodic affective meaning to be anomalous, their perception of the affective meaning of the utterance is less good than those listeners who do not perceive any anomaly. Specifically, Japanese listeners (who understood the lexical meaning of "banana") did less well identifying the affect than did the American listeners (who did NOT know the lexical meaning of the utterance). Also, male listeners (who presumably process word meaning and affective prosody independently), did less well identifying the affect than female listeners. More work needs to be done in this area of perception of affect in speech, and especially we need to explore the "anomaly hypothesis" introduced here as a way of explaining the experimental results.

5. Acknowledgements

I wish to thank my students Mayuko Ohnishi and Haruka Kurihara for their work on an earlier version of this paper (Ohnishi et al., 2004), as well as the student who produced the various "banana" utterances. I thank the students of Gifu City Women's College for their participation in the Japanese perception tests, Ms. JeanHee Jung and her friends for the Korean perception test, and Megan Wyett and students at Black Hills State University, Spearfish, S.D. for the American perception tests. This study is supported by a Grants-in-Aid for Scientific Research, Japanese Ministry of Education, Culture, Sports, Science and Technology (2002-5): 14510636.

6. References

ANSI (1973). Psychoacoustical terminology. *Technical Report, S.3.30, American National Standard Report.*

Campbell, N. (2004). Accounting for voice-quality variation. *Proceedings of Speech Prosody 2004, Nara*, 217-220.

Campbell, N. and Mokhtari, P. (2003) Voice quality: the 4th prosodic parameter. *Proc. 15th International Congress of Phonetic Sciences*, 2417-2420.

Fujimura, O. & Erickson, D. (2004) The C/D Model for prosodic representation of expressive speech in English. *Acoust. Soc. Japan, Fall Meeting*, p. 271-2.

Fujisaki, H. (2004) Prosody, information, and modeling—with emphasis on tonal features of speech. *Proceedings of Speech Prosody 2004, Nara*, pp. 1-10.

Imaizumi, S., Honma, M., Ozawa, Y., Maruishi, M. Muranaka, H. (2004a) Gender differences in the function af organization of the brain for emotional prosody processing. *Proceedings of Speech Prosody 2004, Nara*, 605-608.

Imaizumi, S., Homma, M., Ozawa, Y., Maruishi, M., and Muranaka, H. (2004b) Gender differences in emotional prosody processing—an fMRI study. *Psychologia*, **47**, 113-124.

Imaizumi, S., Homma, M., Ozawa, Y., Yamasaki, K., Maruishi, M., Muranaka, H. (2005) Organization and development of the brain mechanism for understanding speakers' real intentions. *Humanity and Science* **5.1**, 21-29.

Laver, J. (1980) T*he Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Maekawa, K. (1998) Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *Proc. Internt'al Conf on Spoken Lang Proc*., 635-8.

Mazo, M., Erickson, D., and Harvey, T. (1995) Emotion and expression: Temporal data on voice quality in Russian lament. *The Eighth Vocal Fold Physiology Conference, Kurume, Japan*, 173-187.

Mokhtari, P. (2003) Mokhtari, P. (2003a). Parameterization and control of laryngeal voice quality by principal components of glottal waveforms. *J. Phonetic Soc. Japan*, **7 (3)**, 40-54.

Ohnishi, M., Kurihara, H., and Erickson, D. (2003) Difference in perception of vocal emotion by men and women (in Japanese). *Gifu City Women's College Research Bulletin* **53**, 85-90.

Sadanobu, T. (2004). A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, **8**, 29-44.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, **32(1)***, 76-92.

Schirmer, A., and Kotz, S. A. (2002) Sex differentiates the STROOP-effect in emotional speech: ERP evidence. *Proceedings of Speech Prosody 2002*, 631-634.

Schirmer, A., Zysset, S. Kotz, S. A., von Cramon, D. Y. (2004) Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, **21**, 1114-1123.

Sundberg, J. (1974) Articulatory interpretation of the singing formants. *J. Acoust. Soc. Am*., **97**, 3112-3124.