Perception by Japanese, Korean and American listeners to a Korean speaker's recollection of past emotional events: Some acoustic cues

Donna Erickson

Showa Music University, Kawasaki City, Japan,
EricksonDonna2000@gmail.com

Abstract

Acoustic and perceptual analyses of spontaneous Korean were made of a Korean woman recalling past emotional events in her life. A subset of 20 single word utterances and 20 isolated vowels were presented to Japanese, American and Korean listeners who were asked to (1) rate the intensity of the perceived emotion and (2) identify the perceived emotion. Listeners could rate intensity and identify emotion of even short utterances, including only vowels. There are differences in identification of emotion by Japanese, American and Korean listeners, presumably due to different modes of processing --native speakers seem to approach the task linguistically while non-speakers of the language, approach it non-linguistically. Also, we found cross-linguistic differences in interpretation of acoustic cues in the speech signal.

1. Introduction

This is the fourth in a series of studies about production/perception of spontaneous emotions by speakers/listeners of different languages. The first study examined American English utterances [1, 2], the second, Japanese [1,3], the third, Chinese [4], and this fourth one, Korean. The previous studies differ from the current study in that the previous studies (1) included articulatory EMA data (2) did not include *angry* speech, and (3) the speakers were actively engaged in experiencing the emotion, due to the recentness of the emotional experience or to the intensity of the emotional experience. In this current study of Korean, the speaker is reminiscing about past emotional *angry*, *sad* or *happy* events in her life, and given the lack of "recentness", perhaps there is less intensity in her expression of emotions, compared with those of the previous studies. A fine line may exist between spontaneous expressions of intense emotions, recollection of past emotions and acted emotions. The acoustic characteristics of the recalled emotional speech may be more similar to that of acted speech than intense spontaneous emotional speech. Our previous studies on spontaneous expressions of emotional speech suggest that active *sad* grieving speech (American English, Japanese) was characterized by high F0, a tendency for changed F1, together with changed voice quality, e.g., change in glottal opening cycle characteristics/ spectral tilt [1]. For Chinese *sad* speech, voice quality changes were increased H1-H2 & H1-A3, both possible indicators of increased breathiness; *happy* Chinese had high F0 as well as increased intensity [4]. In addition, non-speakers of Chinese identified *happy* and *sad* speech sounds from the prosody of the foreign language without knowing the lexical meaning, even for very short syllables. A native-language effect was observed in that the non-speakers of Chinese tended to show a pattern of identification different from that of the Chinese speakers.

2. Methods

2. 1. Data collection

Acoustic data were recorded for a Korean female speaker, using a technique similar to that reported in previous research by the first author [1, 2]. An informal spontaneous dialogue with another speaker in the same room was conducted with her Korean friend who sat next to the subject. The friend guided the conversation so as to evoke *happiness*, *sadness* and *anger*.

2.2. Perception tests

From this data, a set of 20 utterances, 9 lexical words, which sounded *angry, sad,* or *neutral* were selected by the subject's conversation partner in the experiment. The utterances selected are shown in Table 1 below. Two perception tests were administered: (1) words (as shown in Table 1) and (2) vowels taken from the words (the vowels that are capitalized in Table 1.) Table 1 also shows the emotions of each of the words, as identified by the conversation partner, where S indicates *sad*, A indicates *angry*, and N indicates *neutral*. There were a total of 3 each of /e/, i/, and /u/ vowels and 11 /a/ vowels.

Table 1. Emotional utterances. S indicates *sad*, A, *angry* and N, *neutral*.

Word	**Emotion**
nund**E** (**'**present participle**')**	S,A,N
wor**I** ***(***'we'***)***	S,A,N
gur**U**go ('then')	S,A,N
ajj**A**tdun ('anyhow')	S,N
app**A** ('daddy')	S,N
guny**A**ng ('just the way')	S,N
eomm**A** ('mommy')	A,N
jinjj**A** ('really')	A,N
h**A**yeoton ('anyway')	A

For Test 1, 20 words and for Test 2, 20 vowels were presented (each with 3 randomizations of each utterance). Each test was preceded by a practice test of 5 utterances, and presented through HDA200 Sennheiser headphones in a quiet room, using a Windows-based computer software from Runtime Revolution. The listeners (13 Japanese university students in the greater Kanto area, and 12 Koreans (11 for Test 2 with vowels), also living in the greater Kanto area), responded to two questions: (1) rate each word according to the perceived

degree of emotion on a 5 point scale, with "5" most emotional; (2) identify the perceived emotion—(1) *angry*, (2) *sad*, (3) no emotion, (4) other emotion (5) unknown. The questions were framed to not bias the listeners' perception to a single particular emotion.

2.3. Acoustic analysis

Duration, average intensity, average F0, F1 and F2, F3, F4, H1-H2 and H1-A3 (made at the acoustic steady-state center of the syllable) were measured using Wavesurfer. High values of H1-H2 and H1-A3 (voice quality measurements related to glottal opening and speed of glottal closing, respectively) indicate a more open/breathy quality vs. low values, which indicate a more closed/pressed quality.

3. Results

3.1. Acoustic Measurements

Words. As shown in Table 2, average F0 is highest for *angry*, then *sad*, and then *neutral*, and intensity is loudest for *angry*, then *neutral*, then *sad*. However, ANOVA showed no significant differences.

Table 2. Average Acoustic Measurements Words

Emo	Av F0 (Hz)	Dur (sec)	Intensity (dB)
A	227.1	0.030	52.06
S	201.6	0.039	44.04
N	194.8	0.032	46.28

Vowels. As shown in Table 3, average F0 is highest for *angry*, then *sad*, and then *neutral*; intensity is loudest for *angry*, then *neutral*, then *sad*. H1, H1-H2, & H1-H3 are smallest for *angry*, then *neutral*, then *sad*. H3 is smallest for *angry*, then *sad*, then *neutral*. Also, F3 and F4 values are lowest for *sad*, then *angry*, and then *neutral*. ANOVA showed significant differences only for H1-H2.

Table 3.Average Acoustic Measurements Vowels. F0, F3, F4 values are in Hz, duration in seconds, and the remainder in dB.

E	AvF0	Dur	Int	H1-H2	H1-A3	A3	F3	F4
A	222.4	0.07	58	-6.5	18.2	-42	2654	3684
S	202.8	0.07	50	0.8	25.9	-53	2593	3434
N	197.0	0.08	52	-2.7	22.2	-40	2677	3747

3.2. Perception Results

Emotional Intensity Ratings. Figs. 1, 2 & 3 show the average emotional intensity ratings by Japanese, Korean and American listeners, respectively. For all language groups, (1) the average rating of the utterances was never "emotional" (a rating of 3 meant the utterance was heard as emotional), (2) ratings of words were generally higher than that of vowels alone, and (3) *angry* utterances were heard as more emotional than *sad*, regardless whether the utterances were words or short vowels extracted from the words. The exception was with American listeners who rated *sad* utterances as most emotional. Japanese and American listeners showed a similar pattern of intensity rating for both words and vowels; however, American listeners showed a marked increase in intensity rating for *sad* words, not seen with Japanese or Korean listeners. Korean listeners in general showed significantly lower emotional ratings for the vowels. This suggests that maybe Korean listeners were processing the words linguistically, and unable either to process the vowels linguistically, or switch to a non-linguistic mode of processing.

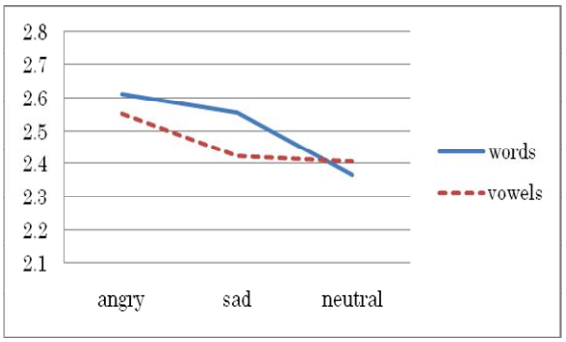

Figure 1*: Average intensity ratings by 13 Japanese listeners for words and vowels.*

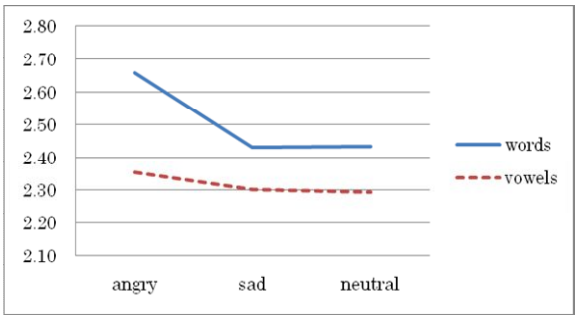

Figure 2. *Average intensity ratings by 11 (for vowels) and 12 (for words) Korean listeners.*

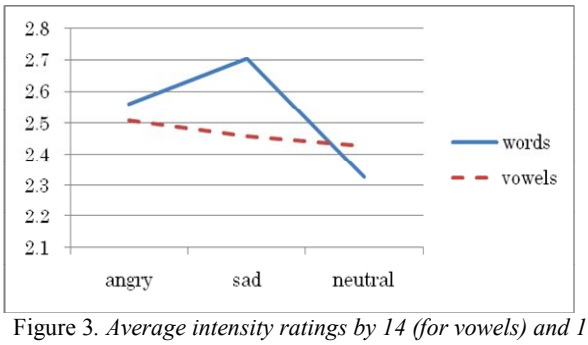

Figure 3. *Average intensity ratings by 14 (for vowels) and 15 (for words) American listeners.*

Emotional Identification. Figures 4 & 5 show the emotional identification by Japanese listeners of the words and vowels, respectively. Three points can be made: (1) the average identification of *angry* and *sad* utterances was above chance (chance level 20%), (2) identification of emotions of words were generally higher than that of vowels alone, and (3) *sad* words had a higher rate of identification (58%) than *angry* words (42%) whereas *sad* and *angry* vowels were identified about the same rate (36% and 35%, respectively).

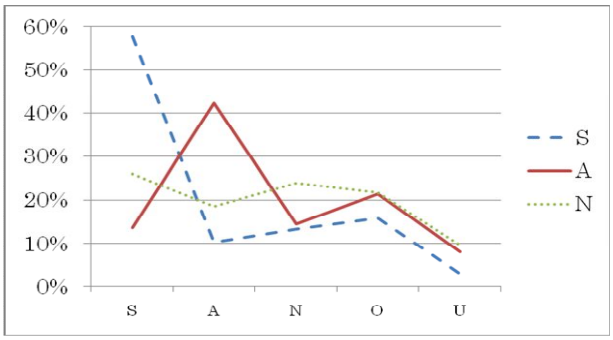

Fig.4. *Japanese listeners ID of emotion (Words)*

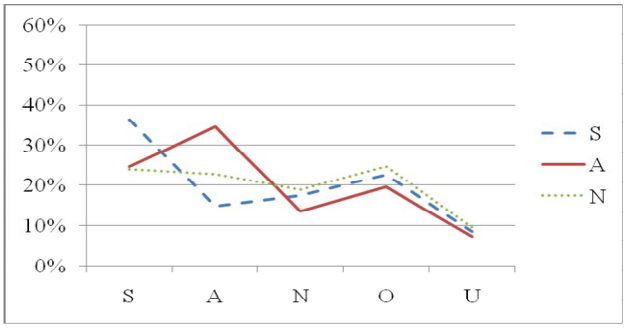

Fig. 5. *Japanese listeners ID of emotion (Vowels)*

Figures 6 & 7 show the emotional identification of the words and vowels, respectively, by Korean listeners. The identification of words was better than for vowels. For words, the identification was highest for *angry* (39%), then *sad* (34%) and then *neutral* (30%), with a confusion between *sad* and *neutral* for neutral utterances. For *angry* vowels, there was a confusion with *angry* and *neutral* (28% & 25%, respectively), and *sad* vowels were identified barely above chance (21%). Also, the overall identification of emotions was higher for the Japanese listeners than the Korean listeners, both for words and vowels. These results further support that Koreans used a linguistic processing mode for the identification tasks. They did relatively well with the word task, but not the vowel task. There also was a significant difference in the number of identifications of each emotion according to language group: Japanese listeners identified more utterances as *sad*, or *neutral*, while Koreans, more utterances as *unknown*.

Figures 8 & 9 show the emotional identification of the words and vowels, respectively, by American listeners, and the results are similar to those seen for the Japanese listeners: (1) the average identifications of *angry* and *sad* utterances were above chance, (2) identification of emotions of words was slightly higher than that of vowels alone, (3) and also *sad* words had a higher rate of identification (59%) than *angry* words (35%) but in addition, *sad* vowels had a slightly higher rate of identification (42%) than *angry* vowels (31%).

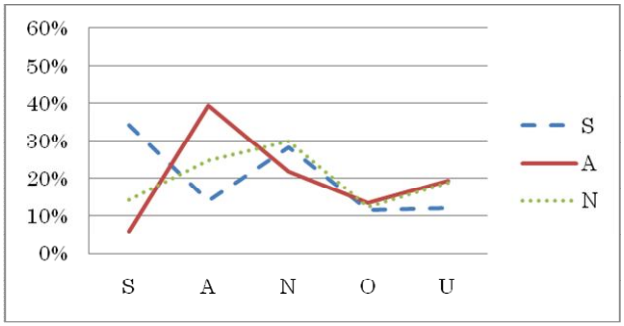

Fig.6. *Korean listeners ID of emotion (Words)*

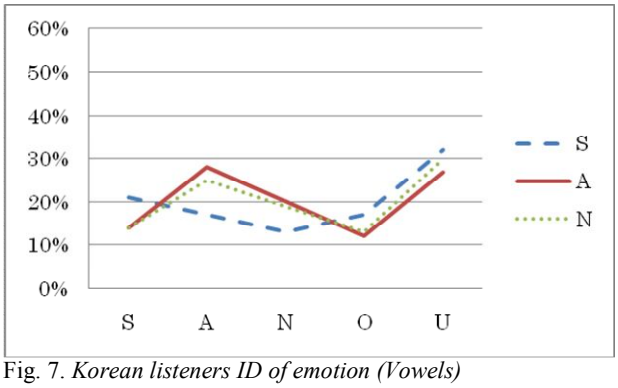

Fig. 7. *Korean listeners ID of emotion (Vowels)*

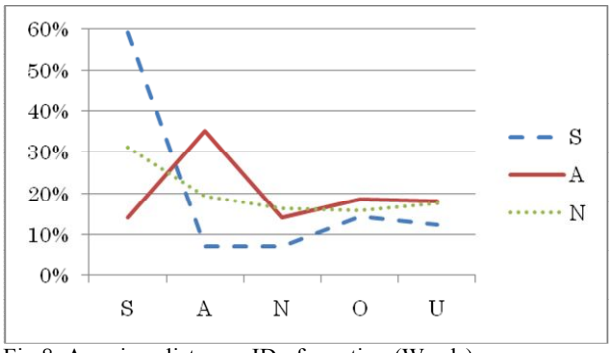

Fig.8. American listeners ID of emotion (Words)

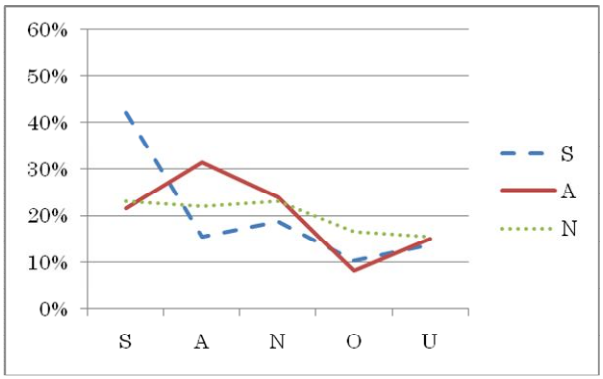

Fig.9. American listeners ID of emotion (Vowels)

3.3. Correlation between perception/acoustics

Words A Pearson product correlation analysis was done with words and acoustic measurements for the perceptions of *sad* and *angry*. Japanese and American listeners showed significant correlations between intensity and *angry* (p=0.012

& p=0.002, respectively); for Korean listeners, no significant correlations were found.

Vowels. A Pearson product correlation analysis was done with vowels and acoustic measurements for the perceptions of *sad* and *angry*. The results are shown in Tables 4 and 5 below.

Table 4. Significant correlations for perception of *sad* and acoustic measurements. N.S. indicates no significant differences.

Listeners	Intensity	H1	A3
Japanese	N.S.	0.011.	0.013
Korean	N.S.	N.S.	N.S.
American	N.S.	N.S.	N.S.

Table 5. Significant correlations for perception of *angry* and acoustic measures. N.S. indicates no significant differences.

Listeners	Intensity	A3	H1-A3
Japanese	0.001	0.045	N.S.
Korean	N.S.	N.S.	N.S.
American	0.000	0.012	0.035

Japanese listeners showed significant correlations for the identification of *angry* with intensity (p=0.001) and A3 (p=0.045) and of *sad*, also with intensity (p=0.000) & A3 (p=0.013). For Korean listeners, no significant correlations were seen. For American listeners, significant correlations were found for *angry* & intensity (p=0.000), A3 (p=0.012), and H1-A3 (p=0.035), but no significant correlations for *sad* Thus, for *angry*, Japanese and American listeners may cue into (1) overall increase of intensity and (2) the increased amplitude of the third formant (A3). In addition, American listeners cue into the tense voice quality brought about by the more rapid closing of the vocal folds (low H1-A3, compared with that for *sad* or *neutral*). For *sad*, Japanese may cue into the decrease in amplitude of the first harmonic (H1) and the third formant (A3). As for Americans, even though they showed a large number of *sad* identifications, there are no significant correlations with any of the acoustic measurements. This suggests they may be tuning into other acoustic cues, such as perhaps formant frequency changes. Due to the variety of vowels and small sampling size, it is not possible to do statistical analyses with formant frequencies. However, we note that F3 and F4, which are relatively independent of vowel category, are lowest for *sad* vowels. Previous studies [e.g., 5] have shown that American listeners may be more sensitive than Japanese or Korean listeners to lowered F4 as a cue to *sadness*.

4. Discussion & Summary

Definite conclusions cannot be drawn from this small data set; however, native listeners may be doing linguistic processing of emotional expressions, while listeners with no understanding of the language, do non-linguistic processing. Japanese, Korean and American listeners may cue into different acoustic information. For Japanese and American listeners *angry* may be cued by overall loudness plus increased loudness of F3. For American listeners, voice quality changes associated with increased tension (caused by decrease in H1-A3) may be important. For *sad* speech, Japanese cue into a quality of softness plus softness of F3. It is not clear from this small data set what American listeners cue into for *sadness*.

As for Korean listeners, the fact that they knew they were listening to Korean may have influenced the results. American and Japanese listeners were able to tune into only the prosody of the speech sound to identify and rate the perceived emotion, since they did not know Korean. For Korean listeners, however, since the lexical meaning of short vowels and phrases was not especially emotional, most Korean listeners reported a great deal of frustration in trying to assign an emotion to the utterances.

Future work is underway to present the vowel data to Korean listeners, without telling them the utterances are Korean. In this way, we might be able to ascertain what acoustic cues are salient for emotional identifications for Korean listeners and compare these results with Japanese and American listeners.

Acknowledgements

We thank Albert Rilliard (LIMSI-CNRS, France), Jianwu Dang (JAIST), Caroline Menezes (U. of Toledo), Jong Hye Han (Korea U.) and two Koreans who provided the data for this experiment, J. Jung and M. Kang.. Also we thank the Japanese, Korean and American listeners who participated in the experiment, and T. Oomae, Y. Tanaka, and V. Tan for their help with the perception tests. This work was supported by the Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007-2010):19520371 to the first author, and part of this work was supported by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan, and also by the Ministry of Education, Science, Sport and Culture, Grant-in-Aid for Scientific Research (A), 16202006, 19202013.

5. References

[1] Erickson, D., Yoshida, K., Menezes, C., Fujino, A., Mochida, T., and Shibuya, Y. (2006) Exploratory study of some acoustic and articulatory characteristics of *sad* speech. *Phonetica,* (63)1-25.

[2] Nguyen, B.P., Tokuda, I., and Erickson, D. (2008) Analysis of the roles of glottal features for emotion classification in spontaneous and acted emotional signals. *Proc. ASJ '2008 Fall Meeting*, 1-Q-20.

[3] Erickson, D., Yoshida, K., Mochida, T., and Shibuya, Y. (2004) Acoustic and articulatory analysis of sad Japanese speech. *Phonetic Society of Japan Fall Meeting*, Sept 25,26, 2004, pp. 113-118.

[4] Erickson, D., Huang, C-F., Shochi, T., Rilliard, A., Dang, J., Iwata, R., and Lu, X. (2008) Acoustic and articulatory cues for Taiwanese, Japanese and American listeners' perception of Chinese *happy* and *sa*d speech. *Proc. ASJ '2008 Fall Meeting,* 1-Q-14.

[5] Erickson, D., Rilliard, A., Shochi, T., Han, J., Kawahara, H., and Sakakibara, K. (2008) A cross-linguistic comparison of perception to formant frequency cues in emotional speech. *COCOSDA, Kyoto, Japan,* 163-167.