The goal of this research was to examine how speakers of different languages and cultures produce emotional speech. The languages examined were American English, Japanese, Korean, Chinese and Russian. A technique was developed to collect articulatory data while the speakers were experiencing emotions. One of the findings was that sad crying speech and happy speech are produced by similar changes in jaw, lip and tongue positions and characterized by high voice pitch, and increased breathiness.

1. **Research Background**

When a speaker is emotional, his or her voice changes. A number of studies have examined acoustic and perceptual characteristics of emotional speech and the differences due to different language background of the speakers and listeners. However, relatively few studies have been done on the articulation of emotional speech. This is to a large extent due to the difficulty in...
collecting articulatory data while a speaker is experiencing emotion.

2. **Research Goals**

The goal of this 36-month period research was to examine how speakers of different languages and cultures produce emotional speech. One of the aims of this research period was to devise a research method for collecting articulatory data during emotional speech.

An additional goal of this study was to examine and compare cross-linguistic differences in the production of emotional speech. The languages examined were American English, Japanese, Korean, Chinese and Russian. The emotions examined were happy (laughing and smiling) and sad crying speech. For Korean, angry speech was also examined.

3. **Methodology**

Six experiments were conducted/analyzed in the research period: 5 EMA (Electromagnetic Articulograph) experiments: (1) with American English sad crying and happy speech, (2) with Japanese sad crying speech, (3) with Chinese sad crying and happy speech, (4) with Korean sad and angry speech, (5) with Russian laments (sad crying speech), and the (6) sixth experiment was high speed laryngeal imaging of vocal folds during Russian laments (sad crying speech).

3.1 **EMA Experiments.**

For each of the experiments (except for Russian), informal spontaneous dialogues were conducted, where the subject sat in the EMA experimental room, and the conversation partner sat in a separate room. The two speakers carried on a conversation through an earphone/microphone set-up. The conversation partner asked the subject various unrehearsed questions based on a list of topics related to the subject’s personal life to evoke happiness, sadness or anger. Happy (laugh and smile) speech, but also sad crying speech, was well evoked, since the subjects were at the time grieving the loss of their mothers (for the American and Japanese subjects) or husband (in the case of the Chinese subject). The conversation usually began with low emotional intensity, but gradually the intensity increased, resulting in crying and sobbing at the same time the speaker was trying to talk. The dialogue continued in a natural manner, while EMA recordings were made within a window frame of 20 sec, with a break in recording of about 3 seconds between frames. Acoustic recording, however, was continuous.

From the approximately one hour’s worth of recording, phrases containing the same or similar words, spoken as happy (including while the speaker was either laughing or smiling), sad (including while the speaker was very sad, at times crying) and neutral, were selected from the dialogue by the speaker.

For the American speaker, the acoustic and articulatory characteristics of eight happy, sad, neutral triplet sets of words (total of 24 words) were examined. For the Chinese speaker, the acoustic and articulatory characteristics of 24 syllable utterances (4 syllable types, /ta1, wo3, jiu4, le, shi4/) which sounded happy or sad were examined. For the Korean speaker, only acoustic characteristics were examined, for 20 utterances. There were 9 lexical words which were spoken as angry, sad, or neutral. For the Japanese speaker, and also for the American speaker, we report on results of a previous study, in which the subjects were asked to imitate the phrasing and emotion of certain of the original spontaneous utterances. Comparisons were made of the articulatory characteristics of the “real” emotional utterance with those of the imitated utterances, spoken as the same lexical item.

For the EMA experiment, EMA receiver coils were
attached to various parts of the tongue, lips and teeth. Measurements were made for the x-y coil positions for the upper and lower lip (UL, LL), for the mandible (J), and the tongue (T1, T2, T3), using a MATLAB-based analysis program. Articulatory and acoustic measurements were made at the time of maximum jaw opening for each of the words. In all experiments, because this was natural, spontaneous conversation, the corpus contains an unbalanced mixture of vowel types. Average duration and, average F0 were calculated for all vowel types, since these characteristic vary marginally for different vowels. In addition, other acoustic characteristics which vary marginally across vowel types were measured, and these were measured at the time of maximum jaw opening: F0, F4, H1-H2, and H1-A3. However, articulatory analysis was done only for those words with vowel nuclei which had the largest number of occurrences. The reason for this is that articulation is highly vowel-quality dependent (since the shape of the vocal tract determines the particular vowel—jaw and tongue movement vary considerably). For the American English utterances, those utterances containing the vowel nucleus /o/ were analyzed (N=11). For the Chinese utterances, those utterances with the vowel nucleus /e/ and /iu/ were analyzed (N=9, N=2, respectively). For the Korean utterances, those utterances with the vowel nucleus /a/ (N=11) were analyzed. For the Russian data, we recorded a Russian ethnomusicologist, who has published widely on lament. Lament is a highly emotional process that combines singing, chanting, and crying, and includes sobbing, excited exclamations, speech interruptions, sighs, voiced breathing. As lamenting progresses, emotional tension intensifies, effecting acoustical changes, such as instability of pitch, intensity, and the voice quality, including vocal fry, diplophonia, etc. These changes do not exist in singing and speaking by the same woman with the same text.

3.2. High speech vocal fold imaging.

For the Russian laments, in addition, high speed image recordings were also made at a separate time at the University of Tokyo University Hospital.

4.1. 研究成果. Russian

Results of the EMA data suggest interesting differences in acoustics and articulation among the three modes—speaking, singing and lamenting. As the lament progresses, the subject becomes more emotional, and shows larger articulatory movements.

Results of the high speed imaging of the vocal folds during lamenting shows differences as well among the three modes, with vocal fold vibration during lamenting showing a great deal more instability compared with the other modes. Specifically, for lament, we see a large posterior chink, which accompanies a breathy voice. As the lament progresses, we can see tense constricted vocal folds, accompanying the increased F0. For speaking and singing, the vocal fold closing is complete, and the posterior chink not found significantly.

4.2. Happy and sad utterances (American English)

Some of the acoustic and articulatory results of the study with American English happy and sad speech are shown in Figures1-2.

Fig. 1. Bar graphs showing mean acoustic values of average F0 for smile, laugh, sad and neutral speech.
One observation is that the emotional utterances (smile, laugh and sad) have similar acoustic characteristics: compared to neutral; they are higher in F0 (voice pitch), and lower in F4 (4th resonant frequency of the vocal tract).

In terms of articulation, as shown in Figure 3, happy utterances have significantly more retracted and lower upper-lip position.

4.3. Happy and sad utterances (Chinese)

The acoustic and articulatory results of Chinese happy and sad speech are shown in Figures 4 and 5. Figure 4 shows the horizontal positions of the upper-lip, lower-lip, and tongue tip. We see a difference in articulation for happy and sad speech, with the upper and lower lips and tongue being more retracted for happy speech. Increased retraction of the upper-lip was also seen for American English happy speech.

Figure 4. lip/tongue positions for happy & sad Chinese

Fig. 5 shows that sad speech has lower lower-lip and jaw positions than happy for Chinese speech; however, we did not see this difference for American English sad speech.

Figure 5.lower lip/jaw positions happy & sad Chinese

4.4. Angry and sad utterances (Korean)

The acoustic results of the study with Korean angry and sad speech suggest that sad vowels have lower F0 (voice pitch) and are softer than angry vowel utterances. With regard to voice quality, H1-H2 and H1-A3 are larger for sad than angry (or neutral).

4.5. Summary and discussion

The articulatory results of sad crying speech, as measured from several languages are summarized in Table 1. More research is needed to better understand what these articulatory differences mean: specifically, are these vowel characteristics, language
characteristics, or speaker characteristics. The next focus of articulatory research will be on developing an algorithm for vowel normalization, in order to analyze the data more fully.

Table 1. Articulatory characteristics of sad speech. American English, Japanese, Chinese, and Russian.

<table>
<thead>
<tr>
<th>language</th>
<th>upper lip</th>
<th>lower lip</th>
<th>jaw</th>
<th>tongue tip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Am.Eng.</td>
<td>raised, protruded</td>
<td>raised</td>
<td>protruded</td>
<td>lowered</td>
</tr>
<tr>
<td>Japanese</td>
<td>retracted, lowered</td>
<td>raised, protruded</td>
<td>raised</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>protruded</td>
<td>lowered, protruded</td>
<td>lowered</td>
<td>protruded</td>
</tr>
<tr>
<td>Russian</td>
<td>retracted</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The acoustic results of sad crying speech as measured from several languages can be summarized as follows: For all languages F0 is high for crying sad speech. Also, there are changes in voice quality—especially a steeper spectral tilt (increased H1-A3) for sad crying speech, found in American English, Japanese, Chinese, and Korean. An additional interesting characteristic of sad crying speech reported for American English was lowered F4, which may be brought about a lowered larynx during sad crying speech.

Interestingly, the results from the happy speech for American English and Chinese show similar acoustic and articulatory characteristics to sad, crying speech: high F0, increased H1-A3 (increased breathiness), and changes in articulation. This suggests that perhaps the underlying mechanisms for happy and sad speech may be similar, and leads to questions about similarities in biophysiological motivations of these two supposedly different emotions.

As for angry utterances, only Korean utterances were examined. In general angry utterances were louder and had higher F0 (voice pitch) than sad utterances. With regard to voice quality, H1-H2 and H1-A3 were smaller for angry than sad (or neutral), thus indicating that angry voices were less breathy than sad ones.

The method and results of the studies reported here serve as guidelines for investigating various acoustic and articulatory characteristics of emotional speech in the future. These types of studies have not been done previously and as such, they are pioneering works.

5. 主な発表論文等 雑誌論文（計 3 件）

Erickson, D., Shochi, T., Menezes, C., and Dang, J. (conditional accept). Some preliminary Observations about articulatory, acoustic and perceptual characteristics of laugh and smile speech in comparison with sad and neutral speech. Mouton de Gruyter, Special Issue on Laughing.


Shochi, T., Erickson, D., Sekiyama, K., Rilliard, A.,


