2017 RISP International Workshop on Nonlinear Circuits,
Communications and Signal Processing (NCSP'17)
Guam, USA, February 28th to March 3rd, 2017

Articulatory Characteristics of Expressive Speech in Activation-Evaluation Space

Takuya Asai[1], Atsuo Suemitsu[2] and Masato Akagi[1]

[1] Japan Advanced Institute of Science and Technology
1–1 Asahidai, Nomi, Ishikawa 923–1292 Japan
Phone/FAX:+81-761-51-11110
E-mail: s1510002@jaist.ac.jp

[2] Sapporo University of Health Sciences
2–1–15 Nakanuma Nishi 4jou, Higashi-ku, Sapporo, Hokkaido,
007–08994 Japan
Phone/FAX:+81-011-792-3350
E-mail: sue@sapporo-hokeniryou-u.ac.jp

Abstract

In this study, influences of articulatory movement patterns related Activation and Evaluation dimensions of emotions are investigated.

Two professional actors are asked to utter five types of emotional speech with three degree. These emotional types were selected by Activation-Evaluation space. Articulatory data and sound waves were collected using the electromagnetic articulatograph (EMA) simultaneously.

The results show that the jaw is raised for Joy or Anger, and, lips are tinned for Joy. In addition, tongue dorsum goes to back for joy. These results suggest that the jaw position affects Activation dimension, and lips and/or tongue positions affect Evaluation dimension.

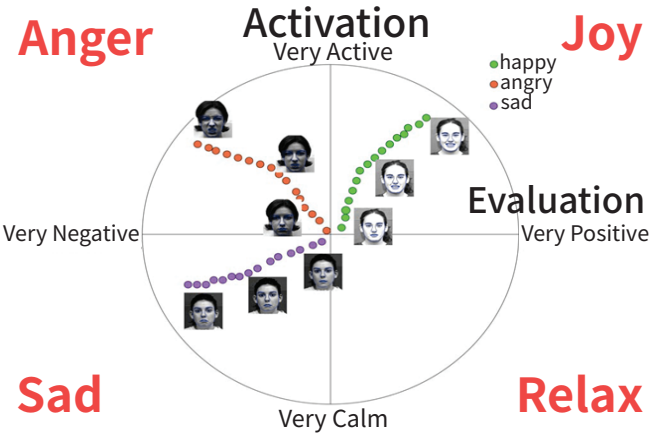


Figure 1: Activation-Evaluation Space.

1. Introduction

Currently, production mechanisms for expressive speech are spotlighted in speech science [1].

As a study on production for expressive speech, Maekawa and Kagomiya [2] examined movements of the tongue, jaw and lips of a Japanese speaker, when uttering voewl /a/ in "sasadaga". They found that the tongue dorsum moved to front for *suspicion* and moved to back for *admiration*. As a similar type of EMA experiment, Erickson examined articulatory movements of two American English speakers [3]. She found that raised tongue dorsum for *suspicion* and backed tongue dorsum for *admiration*. In addition, she found jaw lowering for *anger*.

Although these findings are valuable, these emotional expressions are limited. Relationships between emotional expressions and articulatory movement patterns have not been explained systematically.

In this paper, articulatory movements of five types of emotional speech were investigated. These emotional types were selected based on Activation and Evaluation dimensions [4] (Figure 1).

2. Data Base

To investigate relationships between emotion types and articulation movements, JAIST EMA Emotion Speech Database (JEESD) was created. Speech sound and articulatory data were recorded with EMA system simultaneously. The speakers are two Japanese professional female actors. They were asked to utter five types of emotional speech (anger, joy neutral, relax, sad) with three degree (little, nomal, very).

In this study, vowels /a/ in "dakara mana ha atama ga sarasara da (So, Mana has smooth hair)" were selected for analysis. In order to segment the sentence, the segmental labelings based on "Corpus of Spontaneous Japanese" manual [5] were applied to all recorded sounds. Vowels with ambiguous boundaries were excluded from the analyses. In addition, the vowels at the end of the phrase were excluded from the analyses since the durations are longer than typical vowels.

3. Articulatory data

The articulatory data are collected using EMA (Carstens

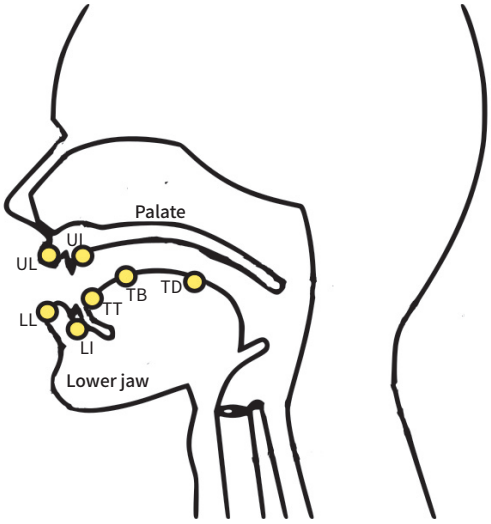


Figure 2: Placements of markers.

Figure 3: Sensors of EMA.

AG500). Figure 2 and Figure 3 show the sampling points with EMA. These points were recorded at a sampling frequency of 200 Hz and subjected to a low pass filter of 50 Hz. Using the values of the nasion and right and left mastoid, the inclination of the face was normalized.

In this paper, jaw opening, lip rounding and tongue movements were measured for the articulatory movement patterns. As jaw opening, the distance between upper incisor and lower incisor was measured (Figure 4.1). Lip rounding is an index for the protrusion of the lips. In this paper, distance between lower lip and lower incisor was measured (Figure 4.2). As tongue movements, distances between incisors and tongue dorsum were measured (Figure 4.3). Distance between the upper incisor and the tongue dorsum shows the movement of the tongue in the vertical direction. Distance between the lower incisor and the tongue dorsum shows the movement of the tongue in the horizontal direction.

Figure 5 shows an example of the trajectory of the jaw opening uttering vowel /a/. The articulatory movement starts with the effects of the previous phoneme, coutirves to the structure of the target phoneme, then moves to the next phoneme. Since the articulatory movement is influenced not only by the emotional expression but also by the before and after phonemes, only the half of the voewl at center was analyzed.

4. Statistical analysis

Since each distance greatly varies depending on oral cavities of the speakers, analyses was performed for each speaker. The number of vowels used for analyses was 424 and 447 for each speaker.

Two-way ANOVAs were used. The dependent variable was the mean value within the distance between articulatory organs. The explanatory variables was emotional expressions and degrees. In order to confirm significant differences for emotional expressions, Tukey's range test (Tukey HSD) was adopted. The significance levels were .05 .

5. Jaw opening

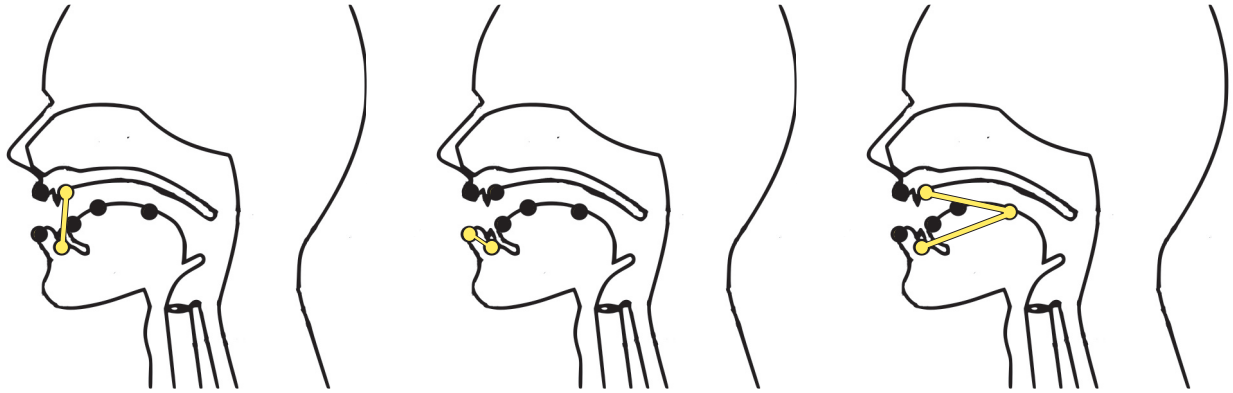
Figure 6.1 shows the relationships between expression of emotions and jaw opening for Speaker 1. When expressing Anger, Joy, and Neutral, the jaw opening was wideer than when expressing Relax and Sad. In addition, when expressing Anger, the jaw opening was to be wider depending on the degree.

There were significant differences between emotional expressions for both speakers (Speaker 1: $F(4, 423) = 34.74, p < .05$, Speaker 2: $F(4, 446) = 25.990, p < .05$). In addition, there were significant differences between degrees (Speaker 1: $F(2, 423) = 8.26, p < .05$, Speaker 2: $F(2, 446) = 4.74, p < .05$). In speaker 1, there was an interaction between emotional expressions and degrees (Speaker 1: $F(8, 423) = 3.99, p < .05$, Speaker 2: $F(8, 446) = 1.39, p > .05$).

As a result of Tukey HSD, there were significant differences between Anger - Relax, Anger - Sad, Joy - Relax, Joy - Sad, Neutral - Relax, and Neutral - Sad for Speaker 1. On the other hand, there were significant differences between Anger - Neutral, Anger - Relax, Anger - Sad, Joy - Relax, Joy - Sad, Neutral - Relax, and Neutral - Sad for Speaker 2.

6. Lip rounding

Figure 6.2 shows the relationships between emotional expression and lip rounding for Speaker 2. The lips were to be thinner when expressing Joy, and to be thicker when expressing Anger.



(1) Jaw opening. (2) Lips rounding. (3) Tongue movement.

Figure 4: Measured points for Articulatory data by EMA.

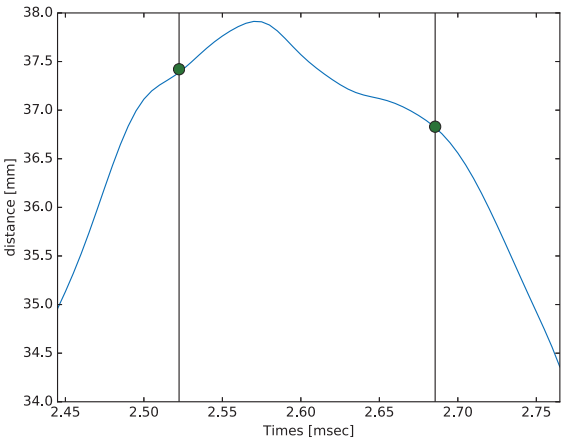


Figure 5: Trajectory of distance between upper and lower incisors in vowel /a/. The X-axis shows the sampling times and the Y-axis shows distance. The black lines indicate the period of the analysis portion.

There were significant differences between emotional expressions of both speakers (Speaker 1: $F(4, 423) = 173.75, p < .05$, Speaker 2: $F(4, 446) = 154.91, p < .05$). For Speaker 1, no significant difference was confirmed between degrees ($F(2, 423) = 1.72, p > .05$). On the other hand, there is a significant difference between degrees for speaker 2 ($F(2, 446) = 23.67, p < .05$).

As a result of Tukey HSD, there were significant differences between all emotions except Neutral - Relax for Speaker 1. There were significant differences among all emotions except Neutral - Anger for Speaker 2.

7. Tongue movement

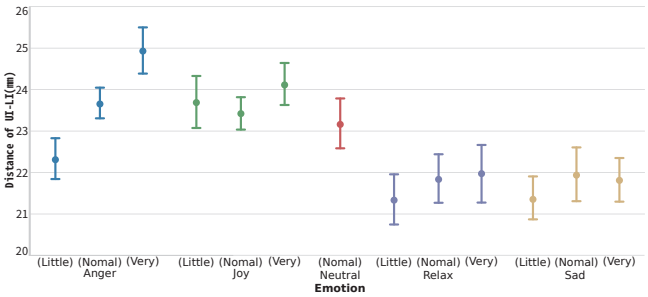
Figure 6.3 shows the relationships between emotional expression and the distance between the upper incisor and the tongue dorsum. Figure 6.4 shows the relationships between emotional expression and the distance between the lower incisor and the tongue dorsum. In the relationships between emotional expression and tongue movement, the tongue dorsum should go down to the back when expressing Anger, Relax and Sad.

There were significant diffearences between emotional expressions for both speakers with the distance of a upper incisor - a tongue dorsum. (Speaker 1: $F(4, 423) = 3.59, p < .05$, Speaker 2: $F(4, 446) = 8.27, p < .05$). There were no significant differences between degrees and distance for both speakers (Speaker 1: $F(2, 423) = 0.16, p > .05$, Speaker 2: $F(2, 446) = 1.24, p > .05$). As a result of Tukey HSD, there was a significant difference only between Anger and Joy for Speaker 1. There were a significant differences between Joy - Anger, Joy - Relax and Joy - Sad for Speaker 2.

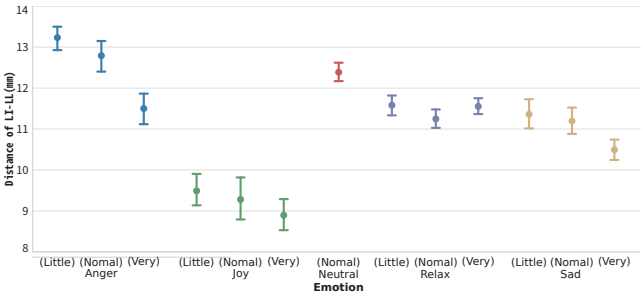
There were significant differences between emotional expressions for both speakers with the distance of the lower incisor - tongue dorsum. (Speaker 1: $F(4, 423) = 3.85, p < .05$, Speaker 2: $F(4, 446) = 2.85, p < .05$). There were no significant difference between the distance and degrees for both speakers (Speaker 1: $F(2, 423) = 0.04, p > .05$, Speaker 2: $F(2, 446) = 1.31, p > .05$). As a result of Tukey HSD, there was a significant difference between Joy - Anger and Joy - Relax for Speaker 1. There was no significant difference in all conditions for the Speaker 2.

8. Discussion

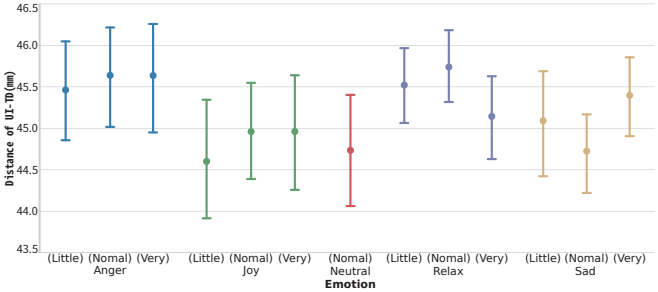
The jaw opening changed with emotional expressions. When expressing Anger, jaw opening became larger according to the degrees of emotion. The statistical analysis showed that there was two groups for jaw opening. Anger, Joy, and Neutral are one group, and Relax and Sad was another group. This result suggests that the jaw opening is related to the Ac-



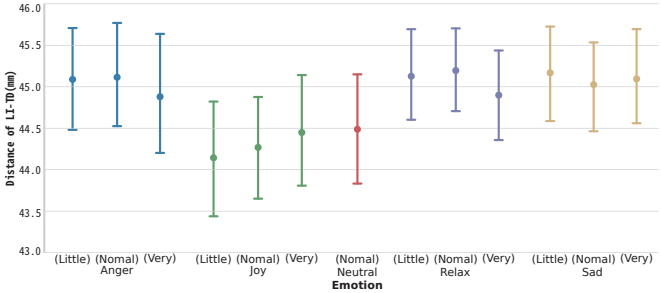
(1) Distance between upper incisor and lower incisor.



(2) Distance between lower incisor and lower lip.



(3) Distance between upper incisor and tongue dorsum.



(4) Distance between lower incisor and tongue dorsum.

Figure 6: Relationship between articulatory data and emotions.

tivation dimension of speech expression.

The lip rounding also changed according to emotional expression. The distance was the shortest when expressing Joy and the longest when expressing Anger. This result suggests that the lip rounding is related to Evaluation dimension.

Tongue movement also changed by emotional expressions. On upper incisors - tongue dorsum distance, there were significantly difference between Joy - Anger. On lower incisors - tongue dorsum distance, there were significant different between Joy - Anger, Joy-Relax and Joy-Sad. These result suggests that there is a relationship between tongue movement and Evaluation dimension. In addition, tongue movement showed big differences among speakers.

9. Conclusion

In this paper, articulatory movements of five types of emotional speech were investigated to clarify the influence of the articulatory differences on Activation and Evaluation dimensions. The results suggested that jaw opening affects Activation dimension, and lip or tongue positions affects Evaluation dimension.

The analysis in this paper was only for articulatory changes to emotional expression. For the future works, it is necessary to investigate the influence of acoustic features from articulatory movements. In regards to that, there is a need to discuss about articulatory movements, acoustic features and impression(Activation-Evaluation dimensions) comprehensively.

Acknowledgment

Part of this research was supported by Grant-in-Aid for Scientific Research (25240026).

References

[1] D. Erickson “Expressive speech: Production, perception and application to speech synthesis,” Acoustical Science and Technology, Vol. 26, pp. 317-325 (2005).

[2] K. Maekawa and T. Kagomiya, “Influence of paralinguistic information on segmental articulation,” Proc. Int. Conf. Spoken Language Processing, Vol. 2, pp. 349-352 (2000).

[3] D. Erickson, A. Abramson, K. Maekawa and T. Kaburagi, “Articulatory characteristics of emotional utterances in spoken English,” Proc. Int. Conf. Spoken Language Processing, Vol. 2, pp. 365-368 (2000).

[4] V. Thurid et al. “Automatic recognition of emotions from speech: a review of the literature and recommendations for practical realisation,“ Springer. Affect and emotion in human-computer interaction, pp. 75-91 (2008).

[5] K. Maekawa, H. Kikuchi, and W. Tsukahara. ”Corpus of Spontaneous Japanese: design, annotation and XML representation.“ Reproduction 16.16 (2004): 5-5.