



Similarity of effects of emotions on the speech organ configuration with and without speaking

Tatsuya Kitamura

Faculty of Intelligence and Informatics, Konan University, Japan

t-kitamu@konan-u.ac.jp

Abstract

In this work we propose and verify a hypothesis on emotional speech production: emotions induce physical and physiological changes in the whole body including changes in the configuration and physical/mechanical properties of the speech organs, regardless of whether or not the person is speaking, and as a side effect, this changes the voice quality. To verify this hypothesis, we measured the configuration of the speech organs of professional actors simulating four emotions (neutral, hot anger, joy, and sadness) with and without speaking by magnetic resonance imaging. The results clearly showed that emotions affect the speech organ configuration, and the same tendency of changes in the speech organ configuration was found regardless of whether or not the person was speaking. We also measured electromagnetic articulography data while a participant watched a relaxation or horror movie, and the result implies that emotional changes can deform the speech organ configuration even if the participant does not speak. These results support our hypothesis.

Index Terms: emotions, speech production, speech organ configuration, MRI, EMA

1. Introduction

Emotions are linked to various physical and physiological reactions such as facial expression, respiration, heart rate, and the tension of muscles in the body. Such responses are induced by emotions regardless of whether or not the person is speaking and affect the configuration and physical/mechanical properties of the speech organs. These changes in the speech organs probably contribute to the voice quality when a speaker utters while experiencing a certain emotion. Although the articulatory characteristics of emotional speech have been examined to some extent [1], there have been almost no studies from the above viewpoint. In the present study, we thus measure the speech organ configuration of participants experiencing emotions and examine the similarity between the deformation of the speech organ configuration with and without speaking.

Articulatory studies of emotional speech production have been carried out using measurement techniques such as electromagnetic articulography (EMA) and magnetic resonance imaging (MRI), and the characteristics of articulation for several emotions have been reported. For example, Erickson *et al.* [2] measured EMA data for spontaneous sad speech and Lee *et al.* [3] reported the vocal tract shape during simulated emotional speech production utilizing a fast MRI system that enabled the measurement of the whole vocal tract shape on the midsagittal plane. These studies focused on the configuration and movement of the articulator during speaking; the present study, however, aims to explore effects of emotion on the speech organ

configuration when the articulator is not speaking and examine their contribution to articulation.

In this paper, we first propose our hypothesis on emotional speech production and then present the results of two experiments: in the first experiment, the deformation of the speech organ configuration while participants simulate four emotions is measured by MRI, and in the second experiment, the deformation while a participant watches movies to elicit spontaneous emotions is measured by an EMA system. Changes in the speech organ configuration are compared in the cases of with and without speaking in both experiments.

2. Side effect hypothesis of emotional speech production

In this paper, on the basis of the revolutionary hierarchical hypothesis of feeling proposed by Fukuda [4], we focus on the effects of emotion on articulation. In Fukuda's hypothesis, "affection" is divided into "emotion" and "feeling." The former involves primitive and basic emotions and the latter involves social and intellectual feelings. Primitive emotions are the most primeval and the other emotions have been acquired gradually during evolutionary history. Primitive emotions are composed of pleasure and displeasure, and basic emotions are composed of joy, anger, fear, disgust, and acceptance or affection. Social feelings, such as love and jealousy, are acquired through group living and intellectual feelings, peculiar to humans, involve affection associated with concepts such as culture and religion.

Emotional changes induce various physical and physiological responses. In this paper, we propose the hypothesis that the changes in the physical and physiological state due to a certain emotion are the direct cause of the changes in the voice quality. On the basis of this hypothesis, the voice quality of emotional speech is not produced consciously but is produced unconsciously owing to the uncontrollability of the physical state. In fact, we do not suppose that there is an articulation target for emotional speech production. For instance, a speaker has little control over the speech organs when experiencing anger, which induces physical and physiological changes in the body including the speech organs, and the speech produced in this state has the voice quality associated with anger. That is, we consider that the changes in the voice quality caused by an emotion are a side effect of the emotion in this hypothesis. In contrast with emotional speech production, the voice quality of speech associated with social or intellectual feelings can be consciously controlled.

Fujisaki [5] grouped information conveyed in speech into linguistic, paralinguistic, and nonlinguistic information. He classified emotion into nonlinguistic information, which cannot generally be controlled by the speaker. This emotion referred to

10.21437/Interspeech.2010-309

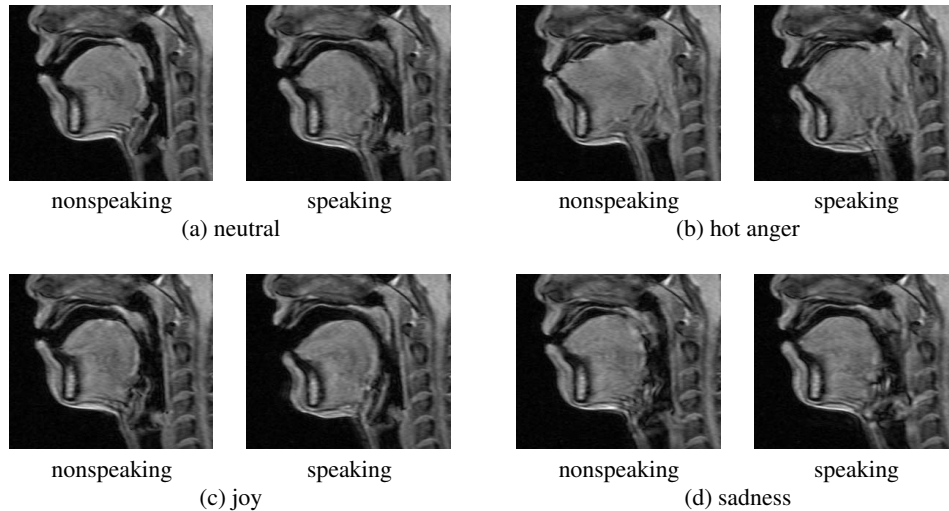


Figure 1: Midsagittal MRI data of the female participant exhibiting (a) neutral emotion, (b) hot anger, (c) joy, and (d) sadness. The left and right images show the nonspeaking and /eh/-speaking states, respectively.

by Fujisaki is the same as that in this paper.

Generally, emotions make one’s body become tense or relaxed. Tenseness and relaxation of the muscles of the speech organs affect voice quality [6], and the changes in the voice quality give listeners clues enabling them to perceive a speaker’s emotion. For example, the constriction of the larynx affects the glottal wave, and an increase in the mechanical impedance of the vocal tract wall emphasizes spectral peaks in higher-frequency region.

We attempt to verify the proposed hypothesis by two measurements of changes in the speech organ shape as a consequence of an emotion, as described in the following sections. If there are similar tendencies of the deformation with and without speaking, the results can be interpreted as indirectly supporting the hypothesis.

3. Experiment 1

Experiment 1 was designed to measure the deformation of the speech organ configuration due to simulated emotions with and without speaking by MRI.

3.1. Methods

MRI data were obtained with an MRI scanner (Shimadzu-Marconi MAGNEX ECLIPSE 1.5 Power Drive 250) at ATR Brain Activity Imaging Center from two Japanese professional actors (one male actor and one female actor). Prior to these measurements, we attempted to obtain MRI data from participants with no theatrical experience; however, they could not make emotional utterances while lying supine in the narrow bore of the MRI scanner. We thus employed the actors in this study.

The participants simulated four different emotions, a neutral emotion, hot anger, joy, and sadness. They were presented with the following dialogues between speakers A and B including /eh/, and were asked to take on the role of speaker B. Note that the sentences in the round brackets are instructions for the participants. /eh/ in Japanese has many meanings, acting as a filler, an expression of agreement, surprise, or anger, for instance. MRI data were acquired once while the actors remained silent and then twice while uttering /eh/ for each of the four

emotions. The participants were instructed not to prepare for the articulation while they expressed the emotions silently, and furthermore to show the emotions not only by speech but also by facial expression.

- | | |
|-----------|--|
| Hot anger | A: That was a complete lie. B: (In anger, almost hitting A) <i>Eh</i> , I can never forgive you. (Express not surprise but anger) |
| Joy | A: You’ve got a job offer? B: <i>Eh</i> , yes, I have. (Express not relief but joy) |
| Sadness | A: You broke that expensive dish? B: <i>Eh</i> , yes, I did. |

The imaging sequence was a sagittal RF-FAST series with 2.5 mm slice thickness, no slice gap, no averaging, a 256×256 mm² field of view, a 512×512-pixel image size, 13 slices, a 10° flip angle, 3.36 ms echo time, and 10.0 ms repetition time. These parameters were determined such that the data acquisition of the volume including the vocal tract was completed in approximately 15 s, that is, during an utterance.

Speech sounds were also recorded during scanning by an optical microphone (Phone-Or SOM) and a solid-state recorder (Marantz PMD-670). After the data collection, the author verified that the speech data sounded like the speakers exhibiting the specified emotions.

3.2. Results and discussion

Mid sagittal images of the nonspeaking and speaking states corresponding to the four emotions for the female and male participants are shown in Figs. 1 and 2, respectively. The similarity of the deformation of the speech organ configuration for the two states for each emotion suggests that an emotional change can give rise to deformation of the speech organ configuration regardless of whether or not the person is speaking. Specifically, in the images of the female actor exhibiting hot anger (Fig. 1(b)), the lower jaw appears to be pulled backward, the pharyngeal and laryngeal cavities are narrowed, and the laryngeal height is greater than that for the neutral emotion (Fig. 1(a)). This implies that hot anger can excite the pharyngeal and laryngeal muscles of the participant. In the case of joy (Fig. 1(c)), in contrast to hot anger, the lower jaw is pulled down, the pharyn-

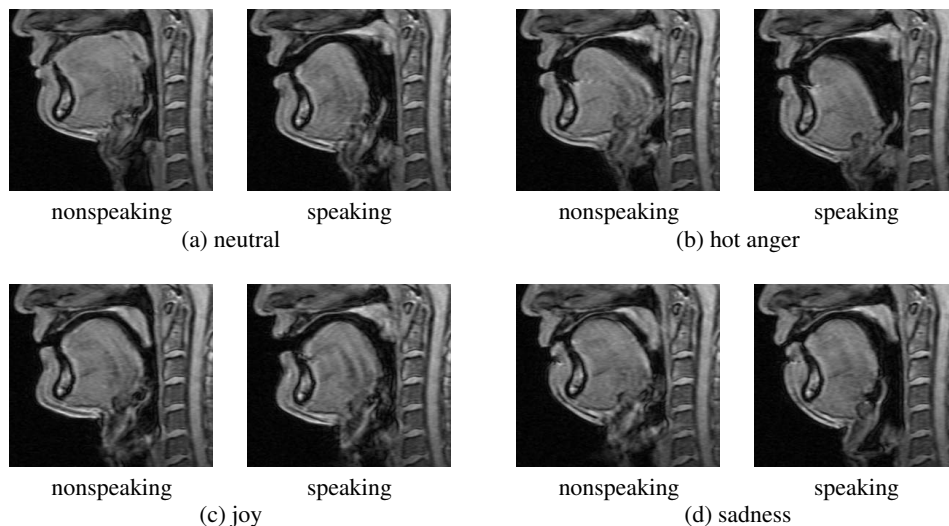


Figure 2: Midsagittal MRI data of the male participant exhibiting (a) neutral emotion, (b) hot anger, (c) joy, and (d) sadness. The left and right images show the nonspeaking and /eh/-speaking states, respectively.

geal and laryngeal cavities dilate, and the height of the larynx is lower than that for the neutral emotion. The results indicate that the pharynx and larynx probably relaxed when the participant exhibited joy. In the case of sadness emotion (Fig. 1(d)), the pharynx and larynx are blurred, indicating that these parts moved during scanning.

The deformation of the speech organ configuration can also be observed for the male participant, but the manner of deformation is different from that of the female participant. This suggests that the effects of emotion on the speech organ configuration differ from person to person. When expressing hot anger (Fig. 2(b)), the male participant pushed the front of the tongue dorsum up and forward, increasing the volume of the oral and upper pharyngeal cavities. Also, the laryngeal height is lowered than that for the neutral emotion. These results are different from those for the female participant. In contrast, the narrowing of the laryngeal cavity is common to both participants. In the case of joy (Fig. 2(c)), the lower pharyngeal and laryngeal cavities dilate, as for the female participant, but the protrusion of the lower jaw and the blurring of the images are different from those of the female participant. The speech organs of the male participant expressing sadness (Fig. 2(d)) underwent a different deformation from those of the female participant; for example, the lower jaw was pulled upward and backward both with and without speaking.

The results suggest that emotional arousal affects the speech organ configuration even if the participants do not speak and that the tendency of the deformation is almost the same for the nonspeaking and speaking states; that is, the characteristics of the deformation due to emotion in the nonspeaking state were retained during speech. This finding supports our hypothesis described in the previous section.

4. Experiment 2

Experiment 2 was designed to investigate the effects of spontaneous emotion on the speech organ configuration with and without speaking. We measured articulatory data while a participant watched a relaxation or horror movie. To obtain articulatory data for emotions that were as spontaneous as possible, we employed an EMA system, in which the participant could sit in an

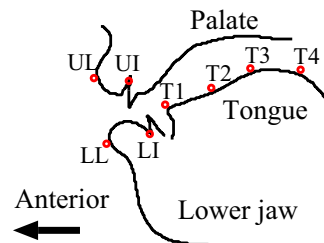


Figure 3: Schematic illustration of midsagittal eight-sensor placement constellation. (UL : upper lip, UI : upper incisor, LL : lower lip, LI : lower incisor, T_n : tongue no. n)

upright posture.

4.1. Methods

An EMA system (Carstens AG500) at JAIST was used to track the positions of eight sensors adhered using a dental glue at the positions shown in Fig. 3. The positions of two reference sensors were also tracked to compensate for head movement. The EMA system sampled articulatory data at 200 Hz and acoustic data at 16 kHz.

A Japanese male with no theatrical experience participated in the measurement. The participant sat with his head and neck positioned inside the EMA Cube. An LCD monitor and a stereo speaker were set 0.7 m in front of the participant, and a relaxation movie showing forests (Condition 1) and a horror movie (Condition 2) were presented. The former was shown with the aim of relaxing the participant; the latter, on the other hand, was shown with the expectation that it would generate fear and tension. The horror movie did not contain any particularly gruesome scenes in consideration of ethical guidelines on human experiments.

While the participant was absorbed in each movie, an experimenter started to ask casual questions about the participant's interests and so forth. The experimenter stood out of sight of the participant and his questions were presented from another speaker set in front of the participant so that the par-

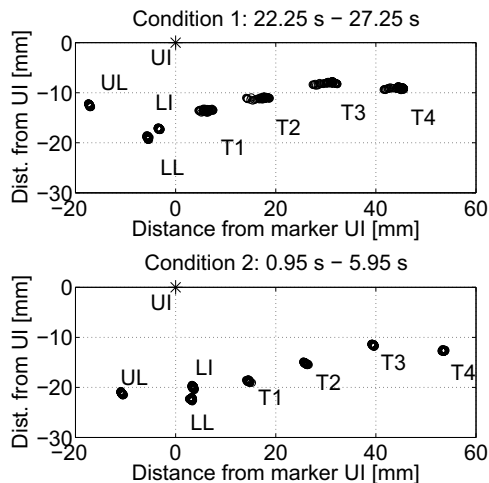


Figure 4: Relative positions of the eight markers in 5-s non-speech segment prior to uttering /eh/. The upper and lower panels show results for the Conditions 1 (relaxation movie) and 2 (horror movie), respectively.

participant's eyes remained fixed on the monitor and that he could maintain his attention on the movie. The participant was asked to add a filler /ehQto/ at the beginning of each answer. The articulatory and acoustic data were measured for a few minutes during the question-and-answer conversations.

4.2. Results and discussion

In this study, we analyzed the articulatory data for the 5 s non-speech segment prior to the first /ehQto/ in the conversation and the /eh/ segment of the first /ehQto/, because the first segments were expected to be the most strongly affected by the movies.

The relative positions of the eight markers for the non-speech segment for Conditions 1 and 2 are plotted in the sagittal plane in Fig. 4. The contour of the hard palate is not drawn, because we were unable to measure the contours in the experiment. Relative to the result for Condition 1, the positions of the markers on the tongue dorsum and lower jaw are displaced approximately 10 mm backward and 3 to 5 mm downward, implying that the pharyngeal and laryngeal cavities were narrowed under Condition 2. In contrast, the articulatory data for the /eh/ segments depicted in Fig. 5 showed that there are no significant differences of the positions of the markers between the conditions.

The results for the nonspeaking state suggest that emotion can affect the shape of the speech organ configuration regardless of whether the person is speaking or not. This is consistent with the results of Experiment 1. The pharyngeal and laryngeal cavities were narrowed in the case of hot anger in Experiment 1 and Condition 2 of Experiment 2. It is likely that the two situations had similar effects on the tensive properties of emotion.

The results for the speaking state, in contrast, indicate that differences in the experimental conditions did not affect the configuration of the speech organs. When the author listened to the /ehtto/ utterances recorded under the two conditions, he could not identify the participant's emotion. This is possibly because the participant had recovered from the emotional impact of the movies when the experimenter asked the first question. Improvements in the experimental design enabling participants to retain a specific emotion throughout the measurement need to be explored in future studies.

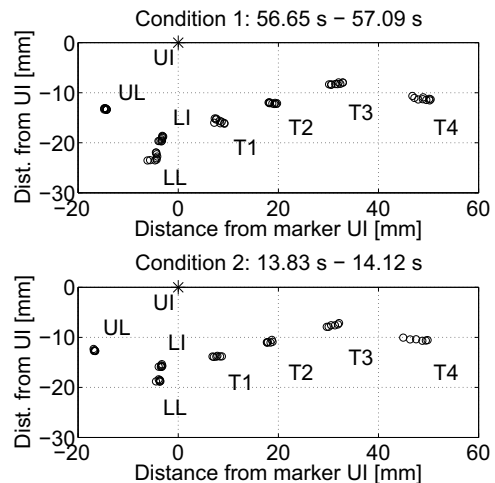


Figure 5: Relative positions of the eight markers during the production of /eh/. The upper and lower panels show the results for the Conditions 1 (relaxation movie) and 2 (horror movie), respectively.

5. Conclusions

In this paper, we proposed the side-effect hypothesis of emotional speech production that the voice quality of emotional speech is produced by speaking with speech organs that have been affected by the emotion experienced by the speakers. We conducted two experiments to verify the hypothesis, in which the speech organ configuration was measured under certain emotions with and without the participant speaking. The results suggest that emotions affect the speech organ configuration regardless of whether or not the person is speaking, and the effects on the speech organ configuration in the speaking and nonspeaking states are similar for certain emotions. These findings support our hypothesis.

6. Acknowledgements

This study was supported by SCOPE (071705001) of the Ministry of Internal Affairs and Communications, Japan, and JSPS KAKENHI (21300071). The author also wishes to thank Professor Jianwu Dang, Drs. Atsuo Suemitsu and Wei Jianguo, and Mr. Kazuya Fujii of JAIST for their kind help in the EMA experiment.

7. References

- [1] D. Erickson, Expressive speech: Production, perception and application to speech synthesis, *Acoust. Sci. & Tech.*, 26(4), 317–325, 2005.
- [2] D. Erickson, K. Yoshida, C. Menezes, A. Fujino, T. Mochida, and Y. Shibuya, Exploratory study of some acoustic and articulatory characteristics of sad speech, *Phonetica*, 63, 1–25, 2006.
- [3] S. Lee, E. Bresch, J. Adams, A. Kazemzadeh, and S. Narayanan, A study of emotional speech articulation using a fast magnetic resonance imaging technique, *Proc. ICSLP2006*, 2006.
- [4] M. Fukuda, Hierarchical hypothesis of feeling based on evolution of the brain: Positioning of social feeling within evolution, *Japanese J. Res. Emotions*, 16(1), 25–35, 2008.
- [5] H. Fujisaki, Prosody, models, and spontaneous speech, In *Computing Prosody* (Y. Sagisaka, N. Campbell, and N. Higuchi, eds.), Springer-Verlag, 27–42, 1996.
- [6] J. Laver, *The phonetic description of voice quality*, Chap. 4, Cambridge Univ. Press, 1980.