ICIC Express Letters
Part B: Applications
Volume **7**, Number **6**, June **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 1201–1205

COMPARISON OF VOCAL TRACT AREA FUNCTION ESTIMATED FROM AUDIO SIGNAL AND MRI ANALYSIS

Megumi Saito[1], Shunsuke Ishimitsu[1], Kazutaka Kasai[2], Kaori Ishii[2]
Izumi Nishio[2], Kimiko Yamashita[2] and Satoshi Horihata[2]

[1]Graduate School of Information Science
Hiroshima City University
3-4-1 Ozukahigashi, Hiroshima Asa-Minami-Ku, Hiroshima 731-3194, Japan
ishimitu@hiroshima-cu.ac.jp

[2]School of Dentistry at Matsudo
Nihon University
2-870-1 Sakaecho, Matsudo, Chiba 271-0062, Japan

Received November 2015; accepted February 2016

Abstract. *Malocclusion caused by the tongue position makes trouble for communication and health. To cure oral habits that underlie such issues, oral myofunctional therapy plays an important role. Unfortunately, physical examinations using X-rays or devices installed in the patient's mouth should be involved as the side effects currently. Therefore, the objective tongue position evaluation without burden to the patients have been proposed and studied for a long while. Here, we employ an evaluation technique using vocal tract area function that was estimated from the partial auto-correlation (PARCOR) coefficient to calculate intraoral information while reducing the burden on patients. The research results showed that as the difference between the measured value and the estimated vocal tract area function was significant, it should be reduced with a new method.*
Keywords: PARCOR analysis, Speech analysis, Vocal tract, MRI

1. **Introduction.** Malocclusion by the tongue produces various health problems such as temporomandibular disorder, and a solution to this issue has long been sought. However, such occlusion has a very wide range of occurring positions and causes. Oral habits, such as the tongue-thrust swallowing, contribute to malocclusion. If these oral habits are not corrected, orthodontic treatment cannot proceed smoothly and the positions of the teeth could well return to their former state even after the completion of treatment. Oral myofunctional therapy (MFT), which improves the coordination of the oral muscles, is generally used to improve such oral habits [1]. Its effects can be evaluated by examining mastication, deglutition, and pronunciation. More conventional methods, for instance, cineradiography and palatograms, have been used as analysis methods to find the tongue position. However, problems exist with these techniques: cineradiography involves radiation exposure, and palatograms require the patient to wear a sensor in their mouth and thus their natural movements cannot be evaluated.

Therefore, an objective analysis of tongue function that does not burden the patient during speech is necessary. Formant frequency analysis is a useful way to determine the position of the tongue [2-4]. The first formant frequency (F1) is changed by the height of the tongue and is high in the low tongue sound. The second formant frequency (F2) is changed by the anteroposterior position of the tongue and is high in the retraction of the tongue. However, this analysis can evaluate only the tip of the tongue.

In a previous study, we used articulatory feature analysis to evaluate tongue function from the viewpoint of the basic units of phonological structure. We first examined speech while a transpalatal arch was attached, and then analyzed the speech of participants with the tongue-thrust swallowing habit. In the tongue-thrust swallowing habit, the tongue is projected forward from the opening between the upper and lower front teeth when swallowing. The teeth position of the patient suffering from the tongue-thrust swallowing habit might be moved by the pressure of the tongue, as long as their oral habits still remain.

The articulatory feature refers to a property of speech based on its voicing or on its place or manner of articulation in the vocal tract. It is possible to evaluate each voice in detail by investigating each distinctive articulatory feature. Articulatory feature analysis can analyze the transmission rate of the articulatory feature defined. As a result, we confirmed its validity in the evaluation of tongue position. However, since the voices of healthy people did not differ from those with tongue-thrust swallowing habits in terms of auditory impression at all, a significant difference was not observed in transmission rates. Because of individual differences between participants, a standard analysis method has not been established.

Thus, in this study we focused on vocal tract area function estimated using the partial auto-correlation (PARCOR) coefficient to evaluate tongue position and confirmed the validity of the proposed method as a new diagnostic technique by comparing with the measured values from magnetic resonance imaging (MRI).

2. Experiment.

2.1. Measuring the vocal tract area by MRI. In the present study, while participants spoke each Japanese vowel for 20 seconds, we imaged the coronal section area of the head and neck using MRI. The MRI apparatus was a Philips Intera Achieva 1.5 T Nova (Amsterdam, Nederland). However, an MRI cannot be extracted from hard tissue. Therefore, to photograph the teeth, we took a teeth mould of the participant and prepared mouthpiece-type trays with spacers of about 1 mm. The mouthpiece-type trays were injected with vegetable oil to create a contrast effect, and was then attached while taking the MRI image. Figure 1 shows the created mouthpiece. Figure 2 shows an example of the vocal tract in a coronal section MRI image when a participant was speaking the Japanese vowel /a/. We divided the vocal tract from the lip to the vocal chords in the coronal section MRI image into 10 parts and measured each of the vocal tract' areas in "ImageJ [5]". Vocal tract area was a plane perpendicular to the vocal tract center line.

FIGURE 1. Mouthpiece

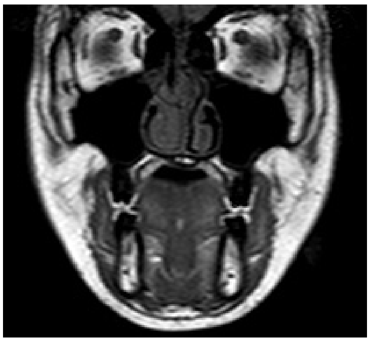

FIGURE 2. MRI image /a/ of coronal section

2.2. **Recording of speech data.** The recording of speech data is necessary to compare the vocal tract area function estimated from PARCOR coefficient and the vocal tract area measured from MRI image. When MRI image has not been performed, loud noise exists because inside of the MRI apparatus is always working the cooling equipment and air-conditioning. Even during measuring MRI image, very large noise occurs. Therefore, it is not possible to perform MRI imaging and audio recording simultaneously. In this study, speech was recorded while attaching mouthpiece in supine state before or after the MRI experiment. It was also recorded using a microphone placed 20 cm away from their mouth. Four healthy people (three females) and a patient with the tongue-thrust swallowing habit (one female), all aged 20-30 years old, participated in this study after giving their informed consent. Participants spoke each of the Japanese vowels /a/, /i/, /u/, /e/, /o/, and repeated this three times. All participants spoke Japanese as their first language and had normal hearing.

2.3. **Estimation of vocal tract area function using the PARCOR analysis.** In this study, we calculated the vocal tract area using the PARCOR coefficient. The vocal tract area function, which was divided into 10 parts from the lips to the vocal cords, was analyzed using this method. Figure 3 shows the flow chart of vocal tract area function estimation using the PARCOR analysis. Here, vocal tract area function is based on the speech production model developed by Kelly et al. [6]. It is a set of cross-sectional areas of each section that describes the vocal tract using a cascade connection of micro-acoustic tubes. The cross-sectional area is constant; see Figure 4. In Figure 4, l is the length of the micro-acoustic tube, A_n is the cross-sectional area of section n, and κ_n is the reflection coefficient between section n and section $n+1$.

First, because a voice waveform contains the characteristics of the sound source as well as the radiation, it is necessary to remove these properties before obtaining vocal tract area function. After removal of the characteristics of the sound source and the radiation, we determined the PARCOR coefficient using an autocorrelation function of the speech waveform. Here, the reflection coefficient κ_n can be expressed by Equation (1).

$$\kappa_n = \frac{A_n - A_{n+1}}{A_n + A_{n+1}} \tag{1}$$

Reflection coefficient κ_n and cross-sectional area A_n are related via Equation (2).

$$A_n = \frac{\kappa_n - \kappa_{n+1}}{\kappa_n + \kappa_{n+1}} A_{n+1} \tag{2}$$

In addition, because the PARCOR coefficient k_n corresponds to the reflection coefficient κ_n, vocal tract area can be determined from the PARCOR coefficient k_n via Equation (3):

$$A_n = \frac{k_n + k_{n+1}}{k_n - k_{n+1}} A_{n+1} \qquad (n = p, p-1, \cdots, 1) \tag{3}$$

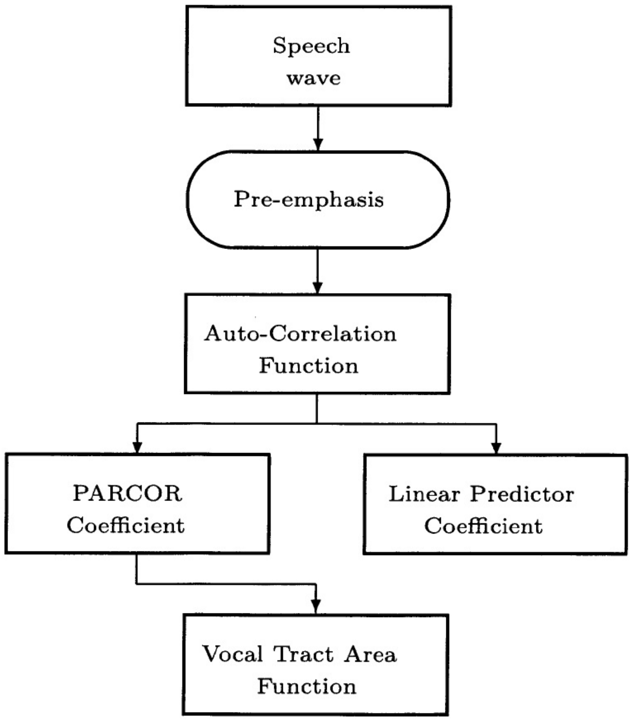

FIGURE 3. Flow of the vocal tract area function estimation from the speech waveform [7]

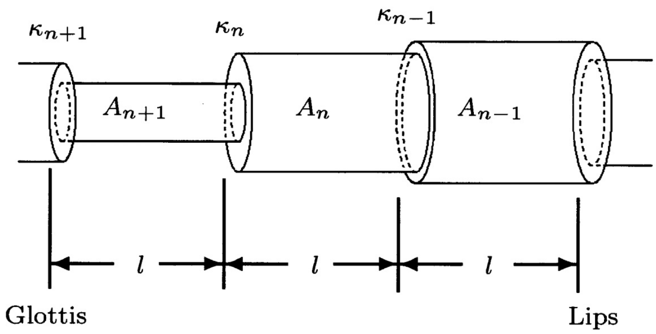

FIGURE 4. Vocal-tract model [7]

3. **Result.** We compared the vocal tract areas analyzed from the measured values in MRI images and the estimated values from PARCOR analysis. Figure 5 shows an example of vocal tract area function determined from the MRI image. Figure 6 shows an example of the vocal tract area function of the Japanese vowel /a/ estimated using PARCOR analysis. We can evaluate the pitch of the tongue objectively since the area is close to 0 when the tongue approaches the palate. The smaller order is the lips side on the horizontal axis. The larger order is the glottis side. The last glottal side is set to 0 as a reference, and the relative value of the vocal tract area is determined. The vowels /i/, /u/, /e/, /o/ were estimated similarly. From Figure 5 and Figure 6, there was little correlation seen between each of the vocal tract area functions. The relationship between measured and estimated values was also little for not only other participants but also other vowels. This indicates that the vocal tract shape is unable to be estimated directly using the PARCOR coefficients.

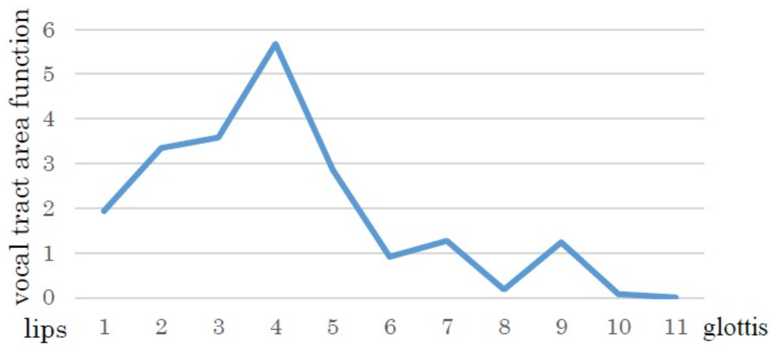

Figure 5. Vocal tract area function /a/ that was measured by the MRI image

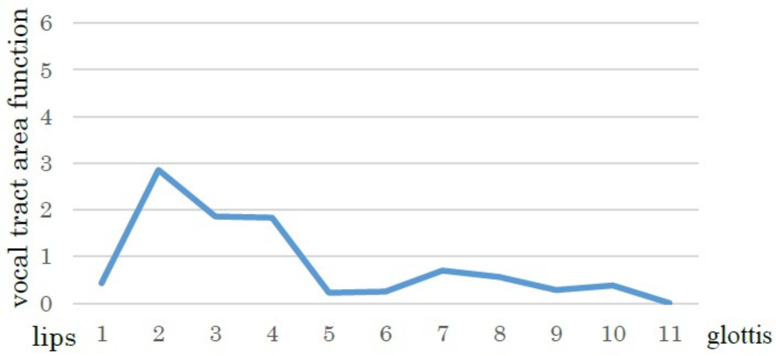

Figure 6. Vocal tract area function /a/ that was estimated by using the PARCOR coefficient

4. **Conclusions.** In this study, we examined and compared the measurement of vocal tract area function using MRI and estimating it from the audio signal. The results showed that the difference between the measured value and the estimated vocal tract area function was large. In addition, it was also similar across participants and vowel sounds. For this reason, we need to consider a new approach to reduce the difference from now on.

REFERENCES

[1] Yamaguchi et al., *Oral Myofunction Therapy to Learn from the Introduction to MFT Beginning*, Wakaba Publishing Inc., 2007.

[2] Ishii et al., Morphologic and acoustic characteristic of a chin face frame and the pharynx respiratory tract of the adult having a tongue prominence habit, *Orthodontic Waves*, vol.71, no.1, pp.33-41, 2012.

[3] Ishii et al., Clinical application of the sound analysis in the tongue usability test, *Orthodontic Waves*, vol.71, no.3, pp.170-177, 2012.

[4] Nishio et al., The long-term study about changed in hard and soft tissues and the position of the tongue at speech after postorthognathic surgery, *Orthodontic Waves*, vol.74, no.1, 2015.

[5] W. S. Rasband, *ImageJ*, U. S. National Institutes of Health, Bethesda, MD, USA, http://rsbweb.nih.gov/ij/.

[6] J. L. Kelly et al., Speech synthesis, *Proc. of the 4th International Congress on Acoustics*, pp.1-4, 1962.

[7] Y. Uchimura et al., Analysis of the vocal tract area function aimed at manipulation of voice quality, *IPSJ SIG Technical Report 2008*, vol.74, no.23, 2008.