Correlation between vocal tract length, body height, formant frequencies, and pitch frequency for the five Japanese vowels uttered by fifteen male speakers

Hiroaki Hatano[1], *Tatsuya Kitamura*[2], *Hironori Takemoto*[3], *Parham Mokhtari*[3], *Kiyoshi Honda*[4], *Shinobu Masaki*[5]

[1,2]Faculty of Intelligence and Informatics, Konan University, Kobe, Hyogo, Japan
[3]Universal Communication Research Institute, NICT, Keihanna Science City, Kyoto, Japan
[4]University of Paris III, Paris, France
[5]ATR/ATR-Promotions, Keihanna Science City, Kyoto, Japan
[1]hatano.hiroaki@gmail.com, [2]t-kitamu@konan-u.ac.jp

Abstract

We conducted quantitative analyses of a magnetic resonance imaging (MRI) database to examine the correlation between physical measures (vocal tract length and body height) and acoustic parameters (pitch and formant frequencies) of vowels. The vocal tract length was measured from MRI data for the five Japanese vowels produced by fifteen male Japanese speakers between the ages of 24 and 55. The acoustic features were computed from vowel sounds recorded during scan. The vocal tract length showed a weak positive correlation with the speakers' age (correlation coefficient r = 0.51) but not with the speaker body height (r = 0.08). There were only weaker correlations between the vocal tract length and the first four formant frequencies except that F1 and F2 of the vowel /e/ show negative correlations with the vocal tract length (F1: r = -0.65, F2: r = -0.56). The result suggests that the vocal tract length is one of the dominant factors causing individual differences in the formant frequencies for the vowel /e/, produced by not forming a strong constriction. Furthermore, the pitch frequency was negatively correlated with the body height (r = -0.61).
Index Terms: magnetic resonance imaging (MRI), vocal tract length, body height, formant frequencies, pitch frequency

1. Introduction

It is widely believed that the vocal tract length reflects the speaker's body size and, as a result, the body size is also related to the formant frequencies of speech sounds. The vocal tract length is clearly different between adults and children and between adult males and females, and these differences cause variations in the formant frequencies. However, it has not been established whether the vocal tract length is correlated with the body size and formant frequencies within a group of adult males or females. In the present study, we thus investigate these relationships using a magnetic resonance imaging (MRI) database.

Several studies have reported the measurement of vocal tract lengths of children and adults, mainly from the viewpoint of developmental changes in the vocal tract. Yang and Kasuya [1] described non-uniform dimensional differences between vocal tract area functions extracted from the MRI data of a boy, an adult female, and an adult male. They compared methods to normalize the vocal tract area functions on the basis of non-uniform and uniform scalings of the vocal tract length. Fitch and Giedd [2] measured the vocal tract geometry from MRI data of 129 participants between the ages of 2 and 25, and demonstrated the changes with development, the relationship between the vocal tract length and the body size, and gender-related differences. Vorperian *et al.* [3] showed developmental changes in the speech apparatus of 327 males and 278 females between the ages birth to 19 years. They showed growth curves of physical parameters with respect to the subjects' age. Despite these previous studies, however, the correlation between the vocal tract length measured during speaking and the formant frequencies of the speech sounds has remained unclear.

Recently, ATR-Promotions released a midsagittal MRI database of the craniofacial region of fifteen adult male speakers obtained during the production of Japanese vowels. It includes speech sounds recorded during the MRI scan, the vocal tract length extracted from the MRI data by the authors, and the speakers' body height. In this study, we analyze the MRI database to explore whether the vocal tract length and body height are the dominant factors for individual variations of the acoustic features of the five Japanese vowels.

2. Materials and methods

2.1. Speakers

Fifteen native Japanese male speakers, listed in Table 1, participated in the experiment. Their ages ranged from 24 to 55 years with a mean and standard deviation of 37

Table 1: *Speakers' age and body height in cm.*

ID	Age	Body height	ID	Age	Body height
M01	29	169	M09	52	165
M02	29	175	M10	55	168
M03	30	171	M11	24	175
M04	34	178	M12	40	162
M05	36	176	M13	27	184
M06	38	177	M14	44	169
M07	38	175	M15	35	175
M08	47	175			

years and 9 years, respectively. Their self-reported body heights were from 162 cm to 184 cm with a mean and standard deviation of 173 cm and 6 cm, respectively.

2.2. MRI and speech data acquisition

The vocal tract shapes of the speakers on the midsagittal plane were scanned by a 1.5 T Shimadzu-Marconi ECLIPSE 1.5T Power Drive 250 MRI scanner installed at ATR-Promotions Brain Activity Imaging Center. The imaging parameters used in the MRI scans were as follows: the repetition time (TR) was 15.0 ms, the echo time (TE) was 3.9 ms, the flip angle (FA) was 10°, the field of view (FOV) was 256 × 256 mm, the matrix size was 512 × 512, the slice thickness was 5 mm, and the number of averaging was 2. Each scan took approximately 5 s. The speakers lay supine in the MRI gantry and were asked to sustain a phonation of the five Japanese vowels (/a/, /e/, /i/, /o/, and /u/) during the scan. Each speaker produced the vowels three times each.

Twelve of the fifteen speakers' voices were recorded during the scan at a sampling rate of 16 kHz with 16-bit resolution through a fiber optic microphone (Optoacoustics, FOMRI). Three utterances were recorded for each vowel, with the exception of speaker M09, for whom two utterances were recorded for the vowel /i/ owing to an error in the measurement. Voices of the remaining other three speakers were missing.

2.3. Extraction of vocal tract length

In the present study, the vocal tract length is defined by the glottis-to-lips length of the vocal tract midline with a 2.5 mm resolution obtained by the method proposed by Takemoto *et al.* [4]. In the method, the vocal fold line, which corresponds to the longitudinal axis of the vocal folds, is first identified manually on the midsagittal image. Next, a contour map with respect to the distance from the vocal fold line is computed for the vocal tract region. The centroids of each contour line are then calculated, and the spline curve passing through the centroids

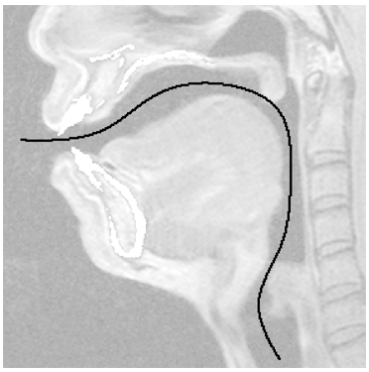

Figure 1: *Example of vocal tract midline drawn by method of Takemoto et al. [4].*

is defined as the vocal tract midline. Figure 1 depicts an example of a midline.

2.4. Extraction of formant frequencies and pitch frequency

Because the vowel sounds during the scan were masked by the loud scanning noise, the formant frequencies and pitch frequency were calculated from vowel segments of approximately 1 s prior to the scan.

The first four formant frequencies (F1, F2, F3, and F4) were measured from the log spectral envelopes of the vowel segments calculated by the unbiased log spectral estimation [5] and averaged with respect to the frames. The frame length was 64 ms, the frame period was 16 ms, the order of the cepstrum was 60, and the number of iterations was 3. The log spectral envelopes were then averaged with respect to the frames, and the formants were identified by a peak-picking method.

The pitch frequency was extracted using the Pitch Contour function of WaveSurfer with its default parameters; the method used was ESPS, the frame length was 7.5 ms, and the frame period was 10 ms. The obtained pitch frequency was corrected manually and averaged with respect to the frames.

3. Results

3.1. Correlation between vocal tract length, body height, and age

Table 2 lists the vocal tract lengths of the speakers for the five vowels and the mean length for the five vowels. The mean and standard deviation of the data are 17.0 cm and 0.8 cm, respectively.

For most of the speakers, the vocal tract length is shortest for the vowel /e/, which has no narrow section in the vocal tract, and longest for the vowel /o/, the only rounded vowel among the Japanese vowels. The shortest and longest vocal tract lengths in the table are 15.3 cm for

Table 2: *Vocal tract lengths of fifteen speakers for five Japanese vowels in cm. Each number is the mean of three utterances.*

ID	/a/	/e/	/i/	/o/	/u/	Mean
M01	16.3	15.5	16.0	17.2	16.7	16.3
M02	16.2	15.3	15.8	17.3	17.2	16.4
M03	16.5	15.8	16.5	17.3	17.0	16.6
M04	17.3	16.2	17.3	18.5	17.9	17.4
M05	17.7	16.9	17.6	18.4	18.8	17.9
M06	16.9	16.3	16.8	17.3	17.0	16.9
M07	16.8	16.1	16.3	18.4	17.2	17.0
M08	17.1	16.4	16.8	17.3	17.0	16.9
M09	16.7	16.0	16.2	17.5	17.0	16.7
M10	18.3	17.5	18.0	19.3	18.7	18.4
M11	15.8	15.6	16.1	16.8	16.8	16.2
M12	16.8	16.3	16.6	17.8	16.8	16.9
M13	17.0	16.6	16.9	18.0	18.0	17.3
M14	16.8	16.4	16.3	17.4	17.9	17.0
M15	16.1	15.5	15.8	17.7	17.4	16.5
Mean	16.8	16.2	16.6	17.7	17.4	17.0

Table 3: *Correlation coefficient (r) between vocal tract length and formant frequencies.*

	/a/	/e/	/i/	/o/	/u/
F1	-0.16	-0.65	-0.27	-0.48	-0.46
F2	-0.15	-0.56	-0.07	-0.50	0.05
F3	-0.21	-0.43	-0.20	-0.41	0.22
F4	-0.24	-0.58	0.38	-0.28	0.23

the vowel /e/ of speaker M02 and 19.3 cm for the vowel /o/ of speaker M10.

The upper panel of Fig. 2 shows the correlation between the mean vocal tract length and body height. The correlation coefficient (r) between them is 0.08, indicating that the vocal tract length does not reflect the body height.

The lower panel of Fig. 2 shows the correlation between the mean vocal tract length and age. Contrary to the result of the body height, the correlation coefficient between the vocal tract length and the age is 0.51, showing a weak positive correlation between them.

3.2. Correlation between vocal tract and formant frequencies

The correlation coefficients (r) between the vocal tract length and the first four formant frequencies for the Japanese vowels are listed in Table 3. Except for F1, F2, and F4 for the vowel /e/, there is no correlation between the vocal tract length and the formant frequencies. Figure 3 shows the correlations between the vocal tract

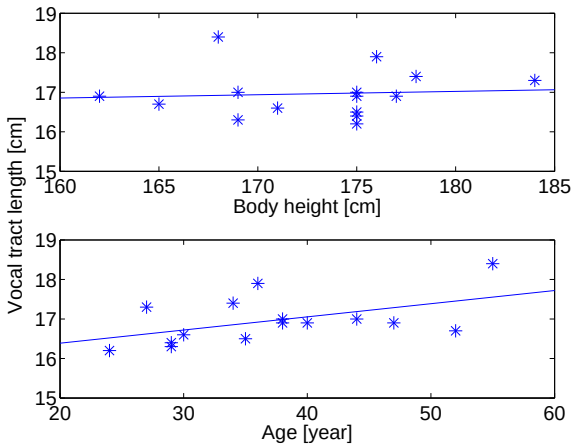

Figure 2: *Upper panel: correlation between mean mean vocal tract length and body height. Lower panel: correlation between mean vocal tract length and age.*

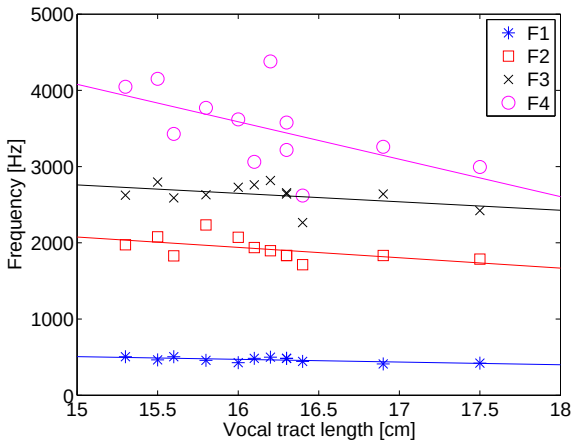

Figure 3: *Correlations between vocal tract length in Japanese vowel /e/ and first (F1), second (F2), third (F3), and fourth (F4) formant frequencies for vowel /e/.*

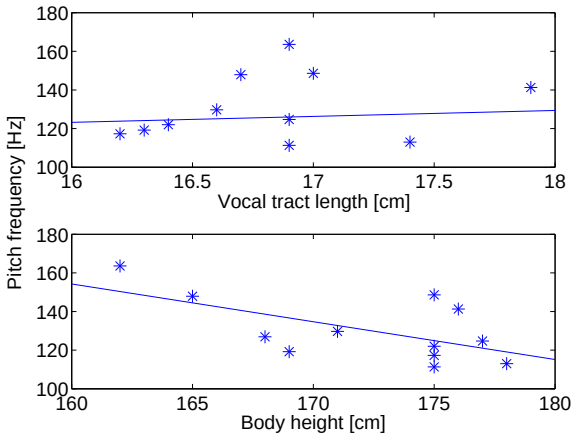

Figure 4: *Upper panel: correlation between vocal tract length and pitch frequency. Lower panel: correlation between body height and pitch frequency.*

length and the formant frequencies for the vowel /e/.

Furthermore, no significant relationship was found between the body height and formant frequencies.

3.3. Correlation between vocal tract length and pitch frequencies

Figure 4 shows the relationships between the pitch frequency and the vocal tract length and body height. The vocal tract lengths were averaged between the vowels. There was no correlation between the pitch frequency and vocal tract length (r = 0.10) and a weak correlation between the pitch frequency and body height (r = -0.61).

4. Discussion

Fitch and Giedd [2] found that the vocal tract length is strongly correlated with the body height (r = 0.926). In contrast to their findings, the present study showed that there is no relationship between them (r = 0.08). This is probably because the ranges of the participants' ages are considerably different between the two studies; the age range was 2 to 25 in the former study, which focused on the developmental changes in the vocal tract shape, whereas the age range was 24 to 55 in the present study. Although the number of speakers is small in this study, our results indicate that the vocal tract length is not correlated with the body height within adult male speakers, whose speech apparatus is already fully developed.

We verified our result on the basis of Fig. 5 of Fitch and Giedd [2] assuming the body height of adult males is over 170 cm. The vocal tract length and the body height were extracted manually from the figure. The correlation coefficient between them is approximately -0.04, supporting our finding.

The vocal tract length showed a negative correlation with the speakers' age. The larynx is lowered with increasing age, especially for elderly people. Thus, the age could be one of the factors for variability of the vocal tract length even for fully developed adults.

For the vowel /e/, the lower formant frequencies show negative correlations with the vocal tract length. In the production of the vowel /e/, the mouth opens moderately, the lips are not rounded, and the tongue does not deform so much and there is no strong constriction in the vocal tract. The vowel /e/ thus has the vocal tract shape closest to a uniform tube among the Japanese vowels. This is probably the reason for the correlations between the lower formant frequencies and the vocal tract length for the vowel. Furthermore, the result suggests that one of the dominant factors causing speaker-to-speaker differences in the formant frequencies is the length of the vocal tract for the vowel /e/.

For the other vowels, the formant frequencies are not correlated with the vocal tract length. This would be because the position and the cross-sectional area at the constriction much affect formant frequencies.

The pitch frequency was only weakly correlated with the body height in the present study. This result suggests that the length of the vocal fold may track the body height, even for a fully developed, adult population.

5. Conclusion

In this study, we investigated the correlation between the vocal tract length, the body height, the first four formant frequencies, and the pitch frequency using MRI data acquired during vowel utterances. The analyses of the data from the fifteen adult male speakers revealed that there was no correlation between the vocal tract length and the body height and that there was almost no correlation between the vocal tract length and the formant frequencies. The latter suggests that individual differences in the formant frequencies, which contribute to voice characteristics, might not be mainly caused by differences in the vocal tract length within the adult male group.

Future works that need to be carried out are to reinforce the results obtained in this study by increasing the number of speakers and to investigate the correlation between the physical and acoustic factors for adult female speakers.

6. Acknowledgements

This study was supported by JSPS KAKENHI (Nos. 21300071 and 21500184). The MRI data and recorded sound data used in this study are parts of the "ATR vocal tract MRI data for Japanese vowels" database that was acquired at Human Information Science Laboratories in Advanced Telecommunications Research Institute International (ATR) and released by ATR-Promotions Co. Ltd. The use of the data is under licensed agreement with ATR-Promotions Co. Ltd. We thank Dr. Sadao Hiroya from NTT Communication Science Laboratories for his fruitful comments.

7. References

[1] Yang, C.-S. and Kasuya, H., "Uniform and non-uniform normalization of vocal tracts measured by MRI across male, female and child subjects," IEICE trans. info. & syst. E78-D, 6, 732–737 (1995).

[2] Fitch, W. T. and Giedd, J., "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am., 106, 3, 1511–1522 (1999).

[3] Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J. and Gentry, L. R., "Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study," J. Acoust. Soc. Am., 125, 3, 1666–1678 (2009).

[4] Takemoto, H., Honda, K., Masaki, S., Shimada, Y. and Fujimoto, I., "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," J. Acoust. Soc. Am., 119, 2, 1037–1049 (2006).

[5] Imai, S. and Furuichi, C., "Unbiased estimator of log spectrum and its application to speech signal processing," Trans. IEICE, J70-A, 3, 471–480 (1987).