SPEAKER INDIVIDUALITIES OF VOCAL TRACT SHAPES OF JAPANESE VOWELS MEASURED BY MAGNETIC RESONANCE IMAGES

Chang-Sheng Yang and Hideki Kasuya

Faculty of Engineering, Utsunomiya University,

2753 Ishii-machi, Utsunomiya 321, Japan

E-mail: yang@utsunomiya-u.ac.jp

ABSTRACT

Three dimensional vocal tract shapes of three males and three females were measured from the magnetic resonance images that were taken during sustained phonation of the five Japanese vowels. Dimensional differences in the vocal tract length of the subjects were quantitatively measured by dividing the entire vocal tract into the oral, the pharyngeal and the laryngeal sections. To investigate main factors that contribute to the differences of the formant patters, uniform and non-uniform normalization methods were applied to the vocal tract shapes. The formant frequencies were also computed from the normalized area functions and compared.

The results show that the physiological dimensions of the vocal tract continuously distributed from females to males. The normalization experiments suggest that the non-uniform scaling of the vocal tract was not significant and back/ front cavity volume in addition to the vocal tract length is an important factor to describe speaker individualities.

1. INTRODUCTION

There are a large difference of formant patterns of an identical vowel between males and females [1-3]. This difference has been ascribed partly to group characteristics of the anatomical structure of the vocal tract [2]. It is also well known that differences exist even in the same gender group [1-3]. These phenomena have made the problem much more difficult to get an obvious relation between the auditory perception of the same vowels and the acoustic parameters of formant frequencies phonated by different speakers. In order to find a reasonable solution to the problem of invariant nature of the vowel, more systematic data are needed at the levels of articulation and formant frequencies.

In the meantime, for the purpose of vowel recognition, Kasuya, et al.[4] and Fujisaki and Nakamura[5] established coordinate systems to represent the auditory space of vowel perception based on uniform scaling of the formant frequencies in terms of the vocal tract (VT) length. Fant[2] and Chiba and Kajiyama [6] observed that the ratio of the pharyngeal length to the mouth cavity length is greater for males than for females and that laryngeal cavities are more developed in males. Fant explained that non-uniformity in the formant patterns between males and females is the result of non-uniformity in these VT dimensions. Nordstrom[7] used data reported by Chiba and Kajiyama[6] and Fant[8] to simulate a number of different articulations by using uniform and non-uniform scaling methods. He found that the anatomical differences between men and women/children only explained part of the formant differences. Our experiment[9] based on the VT shapes measured from the MR images of a male, a female and a child also showed nearly the same result as Nordstrom.

This work tries to gather detailed data and knowledge about the VT shapes of individual speakers and to find important factors of the VT that contribute to differences of the formant frequencies.

2. MEASUREMENT OF THE VOCAL TRACT SHAPES

We have developed a new method to measure a three dimensional VT shape from magnetic resonance (MR) images that were taken during sustained phonation of a vowel [10]. The VT shape of the five Japanese vowels /a, i, u, e, and o/ of three males (MA, MB and MC) and three females (FA, FB and FC) were measured by using the method. A General Electric SIGNA machine was used to acquire the MR image. For each of the vowels, a mid-sagittal image, 23 slices of coronal images for the oral cavity and 26 slices of axial images for the pharyngeal and laryngeal regions were recorded with a slice thickness of 5 mm. The mid-sagittal image was used to measure a VT length and as a reference to reconstruct three dimensional VT shapes.

2.1. VT Dimensions Along the Center Line

To see dimensional individualities in the VT of the subjects, the VT length was divided into three sections: the oral section (from the lips to the top of the uvula), the pharyngeal section (from the top of the uvula to the top of the epiglottis) and the laryngeal section (from the top of the epiglottis to the glottis). A length of each of the three sections was measured from the mid-sagittal MR image. Then a percentage of each section to the whole VT length was calculated. Figure 1 gives a schematic illustration to measure the lengths.

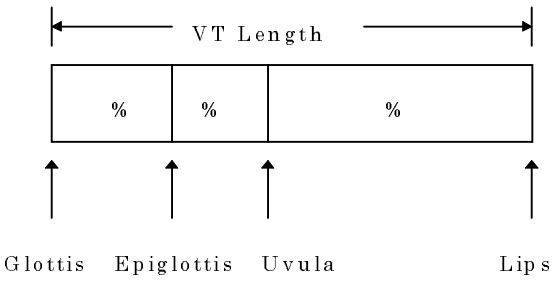

Figure 1: Measurement of vocal tract length dimensions.

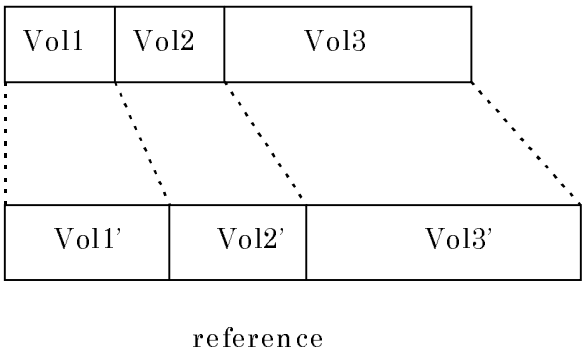

Figure 2: Area function normalization by vocal tract length scaling.

2.2. Area Functions and Formants

A three dimensional VT shape of a vowel was reconstructed from VT boundaries that were extracted from coronal (for the oral region) and axial (for the pharyngeal and the laryngeal regions) MR images by referring to the VT outline that was semi-automatically marked on the mid-sagittal MR image. Cross-sectional areas of the VT were obtained by estimating the VT boundary on the plane perpendicular to the VT center line from the lips to the glottis at an equi-interval of about 0.85 cm. The transfer function of the VT from the glottis to the lips was obtained from the area function by a method proposed by Sondhi and Schroeter[11]. Formant frequencies and bandwidths were computed from the transfer function.

3. NORMALIZATION OF AREA FUNCTIONS

To find the VT dimensions that primarily contribute to differences of the formant frequencies, experiments were performed on the normalization of the area function. Three normalization methods (see Figure 2) were applied with respect to the VT length for each of the vowels:

1. A VT length was uniformly scaled to that of the reference.
2. Lengths of the three VT sections were individually normalized to those of the reference using measured ratios of the corresponding sections.
3. A VT length was first uniformly scaled and the oral cavity volume was then adjusted so as to be the same as that of the reference.

Formant frequencies were computed from the normalized area functions and compared with those of the reference.

Sub.	$\frac{L1}{L}$(%)	$\frac{L1}{L}$(%)	$\frac{L1}{L}$(%)	L(mm)
MA	27.1	16.7	56.2	179
MB	25.3	16.1	58.6	159
MC	24.2	15.4	60.4	165
FA	22.9	16.8	60.3	142
FB	21.6	18.5	59.9	136
FC	23.2	14.6	62.2	140

Table 1: Averaged percentages of the three VT section length to the whole VT length of the subjects, in which L1: percentage of the laryngeal section, L2: percentage of the pharyngeal section, L3: percentage of the oral section, and L: the whole VT length.

4. RESULTS AND DISCUSSIONS

4.1. VT Dimensions

Percentages of each of the three VT section lengths to the whole VT length are shown for six subjects in Table 1, where each item represents the average for the five Japanese vowels. In Table 1, we see that variations of the percentage among the males are relatively larger than those of the females. The values of the oral section of the males MA and MB are relatively smaller in percentage and those of the laryngeal section are relatively larger as compared to the female subjects. The results support the observations of Chiba and Kajiyama[6] and of Fant[2]. But for MC, the values of the oral section are close to that of the female subjects. This suggests that non-uniform differences of the oral cavity length is not necessarily a significant feature to discriminate males from females.

In Table 1, variations of the percentage are relatively larger among the males than the females. The main reason for this is that there is a big difference of the laryngeal height among the male subjects. This can be observed from the mid-sagittal MR images.

Table 1 also suggests that the physiological difference of the oral section length or the laryngeal and pharyngeal section between males and females is non-uniform in the sense of group mean, but individual dimensions of the vocal tract may continuously distribute from females to males.

4.2. Area Functions and Formant Frequencies

Area functions of /a/, /i/ and /u/ are shown in Figure 3 for the males MA and MB. Table 2 lists for the same subjects the first three formant frequencies F1, F2 and F3 of the five Japanese vowels and their respective differences.

A difference of $F(i)(i=1,2,3)$ is calculated as follows.

$$F_{Diff}(i) = \frac{F_{MB}(i) - F_{MA}(i)}{F_{MA(i)}} \times 100\% \qquad (1)$$

From Table 2 we find that MB who has a very short VT length (15.9 cm on average being close to a VT length of a female) shows similar values of the first three formant frequencies to MA who has a longer average VT length of 17.9 cm. MB produces very low F2s and F3s in most vowels and much lower F1s, although MB has shorter VT lengths than MA by 11.2% on average. This fact seems to imply that MB pronounces Japanese vowels of male-like formant frequencies by moving the place of the constriction slightly forward and by maneuvering to have a smaller front cavity volume.

V.	Sub.	F1 Hz	Diff. (%)	F2 Hz	Diff. (%)	F3 Hz	Diff. (%)	L (mm)
/a/	MA	686		1121		2501		177
	MB	677	-1.3	1205	7.5	2466	-1.4	161
/i/	MA	328		2098		2584		182
	MB	321	-2.1	2245	7.0	3206	24.1	154
/u/	MA	430		1313		2296		177
	MB	347	-19.3	1391	5.9	2408	4.9	162
/e/	MA	543		1724		2314		175
	MB	493	- 9.2	1954	13.3	2521	8.9	152
/o/	MA	522		874		2558		182
	MB	469	-10.2	865	-1.0	2524	-1.3	165

Table 2: The first three formant frequencies calculated from the measured area functions of MA and MB, and their respective differences.

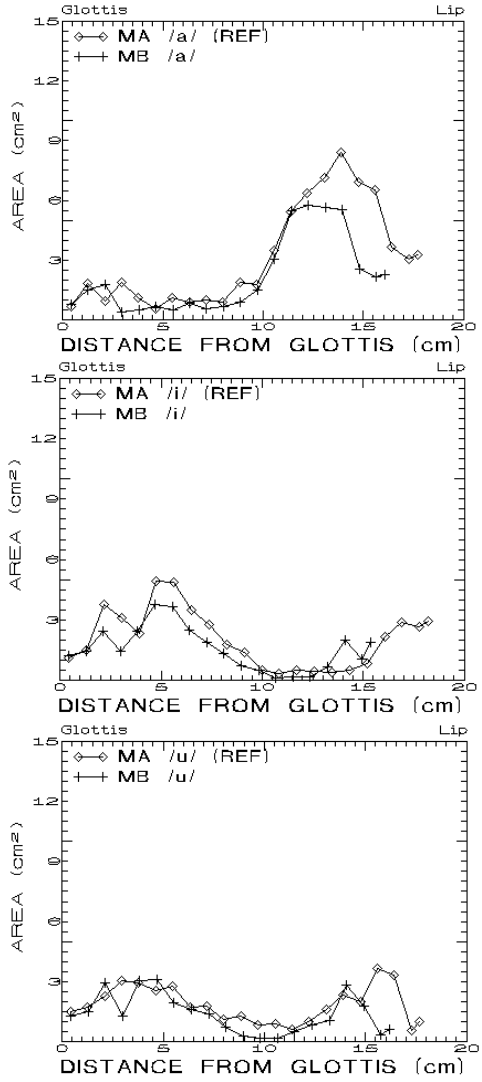

Figure 3: Area functions of Japanese /a/, /i/ and /u/ measured from the MR images of the subjects MA and MB.

4.3 Normalization of Area Functions

Area functions of MA were used as the references for the five Japanese vowels. Area functions of MB were normalized by the three method described earlier. The first three formant frequencies that were computed from the normalized area functions are shown in Table 3. Comparing Table 3 with Table 2, we see that the normalization methods (1) and (2) increase the formant differences. We observed that a small shift of the constriction contributed little to the differences. A large part of the formant differences is eliminated by the method (3) in which both the VT length and the oral cavity volume were normalized. Table 3 seems to suggest that MB produces Japanese vowels of male-like formant frequencies by mainly manipulating the oral cavity volume rather than the place of articulation.

V.	Method	F1 Hz	Diff (%)	F2 Hz	Diff (%)	F3 Hz	Diff (%)
/a/	(1)	631	-8.0	1116	-0.4	2329	-6.9
	(2)	621	-9.5	1137	1.4	2272	-9.2
	(3)	684	-0.3	1091	-2.7	2400	-4.0
/i/	(1)	298	-9.1	1929	-8.1	2742	6.1
	(2)	305	-7.0	1846	-12.0	2765	7.0
	(3)	317	-3.4	1873	-10.7	2747	6.3
/u/	(1)	335	-22.1	1324	0.8	2291	-0.2
	(2)	336	-21.9	1336	1.8	2264	-1.4
	(3)	394	-8.4	1317	0.3	2142	-6.7
/e/	(1)	444	-18.2	1772	2.8	2266	-2.1
	(2)	442	-18.6	1731	0.4	2288	-1.1
	(3)	510	-6.1	1703	-1.2	2274	-1.7
/o/	(1)	441	-15.5	811	-7.2	2381	-6.9
	(2)	437	-16.3	832	-4.8	2298	-10.2
	(3)	472	-9.6	805	-7.9	2401	-6.1

Table 3:Area function normalization results of MB by using area functions of MA as references.

The non-uniform normalization caused no essential effect on the formant frequency values as compared to the uniform ones. The same results were obtained for the other subjects.

These experiments indicate that not only the VT length but also factors such as the back and front cavity volumes are very important in considering speaker individualities of vowel articulation.

5. SUMMARY

Vocal tract shapes of three males and three female were measured from the MR images that were taken during sustained phonation of the five Japanese vowels. VT dimensions of the subjects were quantitatively measured.

We showed that the physiological dimensions of the vocal tract distributed continuously from females to males. This suggests that the non-uniform difference of the oral cavity length is not necessarily a significant feature in discriminating males from females.

The normalization experiment showed that the idea of non-uniform scaling of the VT length did not necessarily reflect the reality of articulatory phonetics. In addition to the VT length, back and front cavity volumes are important factors in describing speaker individualities.

Slight differences in the phonetic quality were observed between the male subjects MA and MB for all the vowel samples except for the vowel /a/. Some method must be devised to modify the area functions measured from the MRI so that the area functions result in the same phonetic quality, in order to make more accurate comparisons at the articulatory and formant levels.

6. ACKNOWLEDGEMENTS

The authors are very grateful to Dr. Shigeru Kano, National Hospital of Tochigi , Dr. Toshihiko Sato, Washiya Hospital, and the subjects who participated in the experimenrts. This work was partly supported by a grant from the Saneyoshi Scholarship Foundation.

7. REFERENCES

1. Peterson, G. E. and Barney. H. L., “Control methods used in a study of the vowels,” J. Acoust. Soc. Amer. Vol. 24, pp.175-194, 1952.
2. Fant, G., *Speech Sound and Features*, The MIT Press, 1973.
3. Kasuya, H., Suzuki, S. and Kido, K., “Changes in pitch and first three formant frequencies of five Japanese vowels with age and sex of speakers,” J. Acoust. Soc. Jpn., Vol.24, No.6, pp.355-364, 1968 (in Japanese).
4. Kasuya, H., Suzuki, S. and Kido, K., “On auditory model of vowel perception,” Proc. of International Congress on Acoustics, B-3-3, Tokyo, 1968.
5. Fujisaki, H. and Nakamura, N., “Normalization and recognition of vowels,” Annual Report of the Engineering Research Institute, Vol.28, pp.61-66,1969.
6. Chiba, T. , and Kajiyama, M., *The Vowels, Its Nature and Structure*, Tokyo-Kaiseikan, Tokyo, 1941.
7. Nordstrom, P.-E., “Female and infant vocal tracts simulated from male area functions,” J. of Phonetics, Vol.5, pp.81-92, 1977.
8. Fant, G., *Acoustic Theory of Speech Production*, Mouton, The Hague, The Netherlands, 1960.
9. Yang, C.-S. and Kasuya, H., “Uniform and non-uniform normalization of vocal tracts measured by MRI across male, female and child,” IEICE Trans. On Inf. & Syst., Vol.E78-D, No.6, pp.732-737, 1995.
10. Yang, C.-S. and Kasuya, H., “Accurate measurement of vocal tract shapes from magnetic resonance images of child, female and male subjects,” Proc. ICSLP94, pp.623-626, 1994.
11. Sondhi, M. M. and Schroeter, J., “A hybrid time-frequency domain articulatory speech synthesizer,” IEEE Trans. On Acoust. Speech Signal Process., Vol.35, No.7, pp.955-967, 1987.