CHANGES OF VOCAL TRACT SHAPE AND AREA FUNCTION BY F0 SHIFT

T. Kitamura[1], P. Mokhtari [1], H. Takemoto[1]

[1]Department of Biophysical Imaging, ATR Human Information Science Laboratories, Kyoto, Japan

Articulatory variation due to the production of vowels at five pitch frequency (F0) levels (110 Hz, 123 Hz, 130 Hz, 146 Hz, and 164 Hz) was analyzed by volumetric magnetic resonance imaging (MRI). Three Japanese male subjects produced sustained Japanese vowels /a/ and /i/. Observation of vocal tract area functions extracted from the MRI data revealed that F0 shift in vowel production affects not only the length of the vocal tract but also its shape. Analysis employing coefficient of variation for length-normalized area functions revealed that the shape of the vocal tract does not change proportionately by F0 shift and that each subject adopt different strategies for controlling F0 while maintaining the phonetic identity of the vowel.

I. INTRODUCTION

The larynx and the supra-laryngeal articulators are connected mechanically and interact with each other to produce speech sounds [1]. Vocal tract shape is thus affected by F0 change. Except for a few previous studies [2][3], however, effects of F0 shift on vocal tract shape have not been studied. In addition, differences of the effects among individuals and their acoustic manifestation have not been reported. The present study therefore aims to investigate possible effects of F0 shift on vocal tract shape and area function by examining individual variations of the interaction and their corresponding acoustic effects.

Effects on the shape of the vocal tract by changing F0 have been measured using several imaging systems. For instance, Hirai *et al.* [2] described differences in vocal tract shape during production of the Japanese vowel /a/ associated with 1.5-octave F0 falling by using magnetic resonance imaging (MRI). They also investigated mechanisms of F0 control in detail and proposed a physiological articulatory model with tongue-larynx coupling mechanism. Tom *et al.* [3] reported differences in vocal tract area function during production of the vowel /a/ under two registers, five F0 levels, and two loudness levels by using electron-beam computed tomography (EBCT).

However, those studies reported the results only for a single vowel of a single subject. Because each vowel has a different constriction location, effects on vocal tract shape may be different among vowels. Also, each speaker may adopt different strategies to control F0. In this study, we thus investigated changes in articulation of a front and a back vowel at different F0 levels for three male subjects. Magnetic resonance images were acquired during producing the Japanese vowels /a/ and /i/ at five F0 levels, and analyzed for the effects on vocal tract shapes and area functions, as well as for the corresponding acoustic effects using a transmission line model.

II. MRI DATA ACQUISITION

Magnetic resonance images of three Japanese male subjects were obtained during sustained production of Japanese vowels /a/ and /i/ with a Shimadzu-Marconi ECLIPSE 1.5T Power Drive 250 at the ATR Brain Activity Imaging Center. The subjects are denoted below as A, B, and C. The imaging sequence was a sagittal Fourier Acquired Steady State (FAST) series with 3.0-mm slice thickness, no slice gap, a 256 × 256 mm field of view (FOV), a 512 × 512 pixel image size, 18 slices, 90° flip angle (FA), 9-ms echo time (TE), and 4,900-ms repetition time (TR). The total acquisition time was approximately 15 sec. These parameters were selected to complete data acquisition in a single breath.

Each subject was positioned to lie supine on the platform of the MRI unit and put on non-magnetic intra-aural headphones. Harmonic complex tones whose fundamental frequency was 110, 123, 130, 146, or 164 Hz were presented through the headphones during scanning. The subjects were instructed to adjust their F0 to the fundamental frequency of the harmonic complex tone while maintaining steady phonation during scanning. Each subject's voice during the scan was recorded through an optical microphone (phone-or FOMRI). After the scan, each utterance was examined to confirm whether the subjects adjusted their F0 as instructed. Any MRI data outside a margin of F0 error of ± 5 Hz was excluded from further analysis. The data for the lowest F0, from subject B, were excluded on this basis.

III. METHOD

A. Morphological analysis

The effects of F0 on vowel articulation were analyzed with reference to the rigid structures. When the subjects' head position in the MR images was different across F0 levels, the MR images were aligned with reference to the line connecting the anterior nasal spine and the posterior margin of the foramen magnum using an affine transformation. Following the alignment, outlines of the vocal tract, hyoid bone, and mandible were traced manually on the mid-sagittal plane to be superimposed together for each vowel.

B. Analysis of vocal tract area function

Cross-sectional areas along the mid-line of the vocal tract were extracted at 2.5-mm intervals from the MRI data set to obtain the area function. Intra-speaker variations of the vocal tract with respect to F0 were examined using the coefficient of variation as an index. Each vocal tract area function was resampled by cubic-spline interpolation in 44 equal-length sections [4], and the coefficient of variation for each section $cv(\mathbf{x})$ was obtained by

$$cv(\mathbf{x}) = \frac{s(\mathbf{x})}{\bar{A}(\mathbf{x})}, \qquad (1)$$

$$\bar{A}(\mathbf{x}) = \frac{1}{N}\sum_{f} A(\mathbf{x}, f), \qquad (2)$$

where $A(\mathbf{x}, f)$ is an interpolated vocal tract area functions for a given F0, $\mathbf{x}$ is the index vector [1, 2, …, 44], N is the number of F0 level, and $s(\mathbf{x})$ is the standard deviation of $A(\mathbf{x}, f)$.

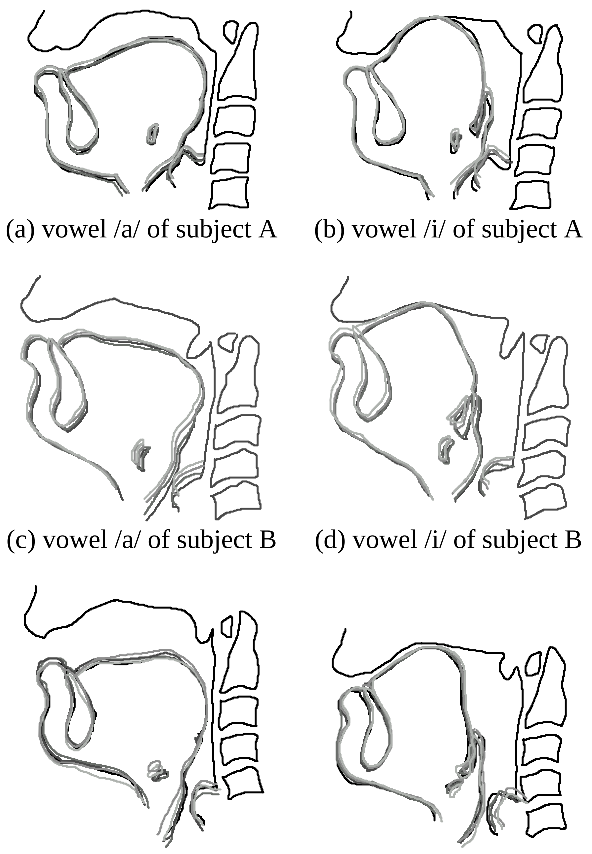
C. Simulation using transmission line model

In order to estimate the acoustic effects of the changes in area function due to F0 shift, the first two formant frequencies were calculated by using a transmission line model. Calculations of the velocity-to-velocity transfer functions of the vocal tract were performed for the frequency region up to 4 kHz. The first (F1), second (F2), and third (F3) formant frequencies were then identified from the transfer functions using a peak-picking method.

IV. RESULTS AND DISCUSSIONS

A. Morphological analysis

Figure 1 shows all the tracings to depict the systematic change in the positions of the speech organs with F0 shift. The changes of the vocal tract shape on the mid-sagittal plane were considerably smaller than those in the previous studies [2][5]. The larynx tended to rise with F0 while the shape of the laryngeal cavity was almost constant for subjects A and B. In contrast, subject C did not exhibit obvious changes in larynx height, rather showing expanding laryngeal cavity with rising F0.



(a) vowel /a/ of subject A (b) vowel /i/ of subject A

(c) vowel /a/ of subject B (d) vowel /i/ of subject B

(e) vowel /a/ of subject C (f) vowel /i/ of subject C

Figure 1: *Superimposed mid-sagittal tracings for the Japanese vowels /a/ and /i/ obtained from three male Japanese subjects. F0 level corresponds to the degree of line saturation of the tracings: the black lines show outlines for the lowest F0 (110 Hz) and the lightest gray lines show those for the highest F0 (164 Hz). The anterior direction is to the left.*

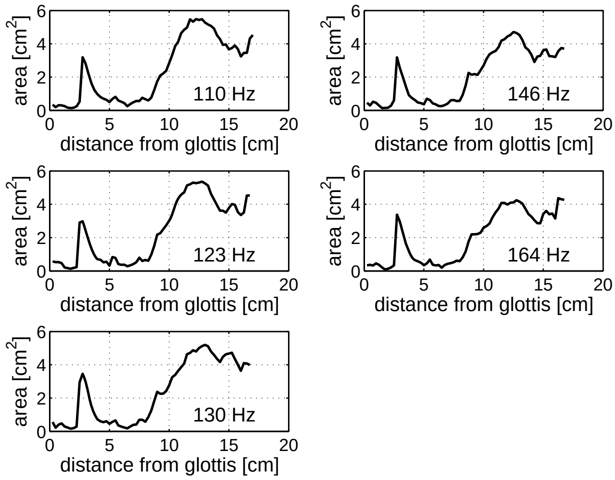


Figure 2: *Vocal tract area functions at all F0 levels for the vowel /a/ of subject A.*

Table 1: *Vocal tract length [cm] associated with variations in F0 and Pearson's correlation coefficient r between them.*

F0	Subject A /a/	Subject A /i/	Subject B /a/	Subject B /i/	Subject C /a/	Subject C /i/
110 Hz	16.6	16.8	---	---	17.2	16.0
123 Hz	16.4	16.7	17.3	16.1	17.6	16.2
130 Hz	16.4	16.7	17.1	16.1	17.6	16.0
146 Hz	16.2	16.6	16.8	15.8	17.2	15.8
164 Hz	16.0	16.2	16.5	15.7	17.4	16.3
r	-0.99	-0.93	-1.00	-0.96	-0.04	0.26

B. Analysis of vocal tract area function

Figure 2 depicts vocal tract area functions for the vowel /a/ of subject A indicating that the F0 shift during vowel production affects not only the length of the vocal tract but also its shape. The areas of the oral cavity of the subject tended to decrease with rising F0 for the subject, although the changes of the vocal tract shape on the mid-sagittal plane were considerably small. This tendency was also found for the other subjects.

Figure 3 depicts the length-normalized mean area functions and their coefficients of variation (CVs) for each section. Non-uniform CV patterns demonstrate that the shape of the vocal tract does not vary proportionately with F0 shift, and sharp peaks of the CVs indicate large changes of the cross-sectional area at the sections among the data. The peak of the CV at the seventh section from the glottis for the vowels of subject A indicates that the junction between the lower pharyngeal and laryngeal cavities varies in location with F0 shift. The peak near the junction can also be found for the vowel /i/ of subject B. The peak of the CV at the 23rd section for the vowel /a/ of subject B indicates that the ratio of oral and the pharyngeal cavity lengths altered with F0 shift. Additionally, the sharp peak near the 42nd section for the vowel /i/ of all the subjects corresponds with movements of the lips with F0 shift.

The CVs at constricted sections are relatively smaller than those at non-constricted sections for the vowel /i/. Because vowel acoustics are relatively sensitive to changes in constriction area [7], this strategy contributes to preserving vowel features regardless of the F0 level.

In contrast to the local change of the shape of the vocal tract for subjects A and B, the lower pharyngeal and the laryngeal cavities (from the first section to the 15th section) of subject C varied widely with F0 changes. Thus, inter-speaker differences of the CV pattern indicate that the strategy to control F0 and vowel articulation varies from subject to subject.

Table 1 shows vocal tract length measured for each condition and Pearson's correlation coefficients with F0. These results indicate that there are strong negative correlations between vocal tract length and F0 for subjects A and B, but not for subject C. The results are consistent with the observation that the larynx position rises with rising F0 for subjects A and B.

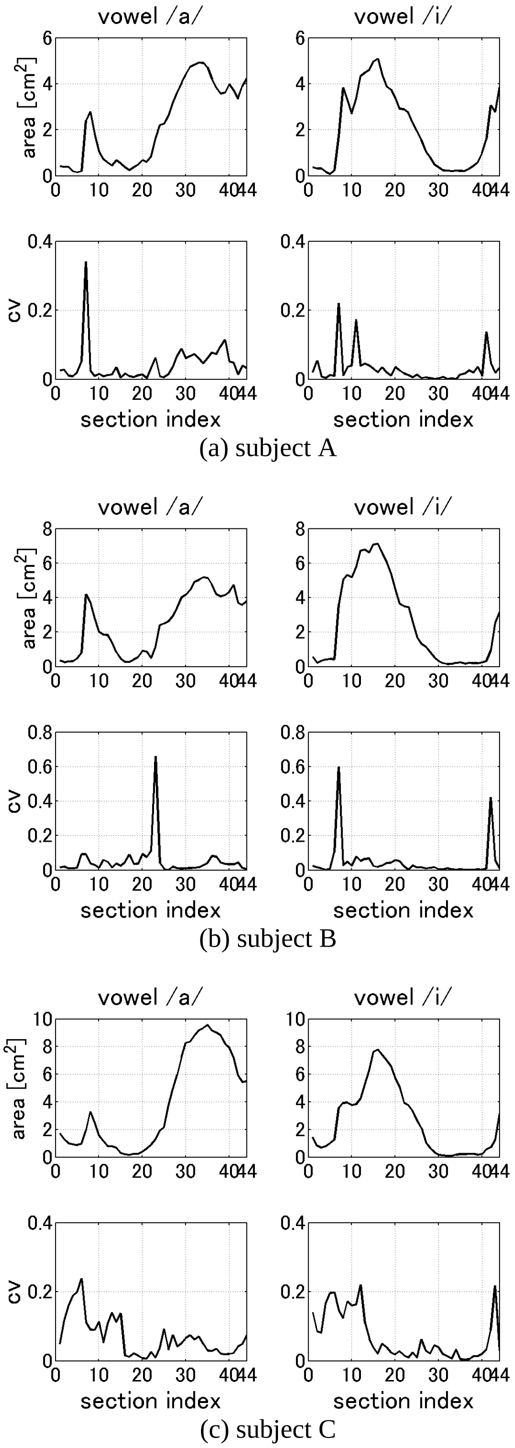


Figure 3: *Average and coefficient of variation (CV) of the length-normalized area function for three male subjects.*

C. Simulation using transmission line model

Figures 4 and 5 depict the frequencies of the lower three formant (F1, F2, and F3) obtained from calculated transfer functions for the vowels /a/ and /i/ for the subjects. The frequencies do not increase uniformly with rising F0, indicating that the shape of the vocal tract does not change proportionately by F0 shift.

There is a positive correlation between F0 and F2 for the vowel /a/ of all the subjects (r = 0.83 for subject A, r = 0.90 for subject B, and r = 0.56 for subject C). The positive correlations are caused by the decrease of the area of the oral cavity with rising F0 for the vowel /a/ mentioned above. In contrast to the vowel /a/, there is no common positive or negative correlation between F0 and the formant frequencies for the vowel /i/.

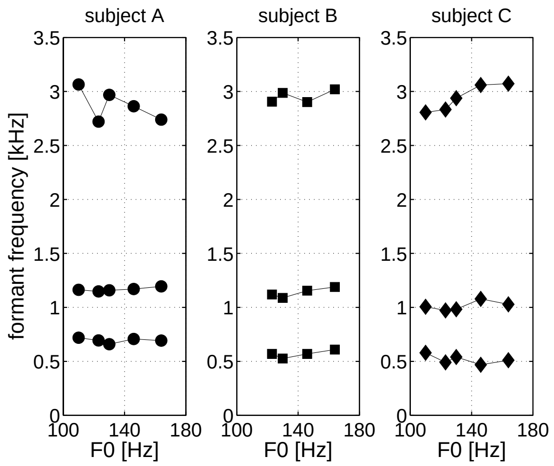


Figure 4: *First (F1), the second (F2), and the third (F3) formants of velocity-to-velocity transfer functions for the vowel /a/ associated with variations in F0.*

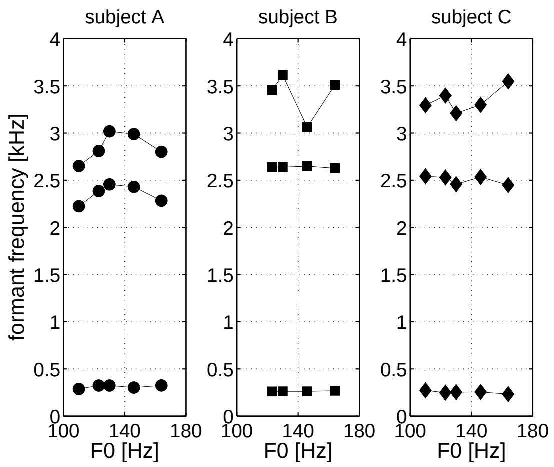


Figure 5: *First (F1), the second (F2), and the third (F3) formants of velocity-to-velocity transfer functions for the vowel /i/ associated with variations in F0.*

VI. CONCLUSIONS

Volumetric MRI was used to investigate changes in vocal tract configuration during vowel production with by F0 changes. The results for the Japanese vowels /a/ and /i/ of three male subjects demonstrated that F0 shift affects not only the length of the vocal tract but also the shape. The data also showed that the strategy for controlling F0 preserving vowel characteristics differs across individuals. The results of the analysis of intra-speaker variation of the vocal tract area functions indicated that the shape of the vocal tract changes non-uniformly with F0 and the regions of changes are different among vowel types and subjects. The results from the acoustical simulation indicated that the vowel /a/ tends to be neutralized with F0 rising while the vowel /i/ is kept constant over the F0 levels.

ACKNOWLEDGEMENTS

This research was supported in part by the Ministry of Internal Affairs and Communications on their Strategic Information and Communications R&D Programme and the National Institute of Information and Communications Technology.

REFERENCES

[1] K. Honda, "Relationship between pitch control and vowel articulation," in Bless D. M. and Abbs J. H. (eds.) *Vocal Fold Physiology*, San Diego, College-Hill Press, 1983, pp. 286-297.

[2] H. Hirai, J. Dang, and K. Honda, "A physiological model of speech organs incorporating tongue-larynx interaction," *J. Acoust. Soc. Jpn.*, vol. 51, pp. 918-928, 1995.

[3] K. Tom, I. Titze, E. A. Hoffman, and B. H. Story, "Three-dimensional vocal tract imaging and formant structure: Varying vocal register, pitch, and loudness," *J. Acoust. Soc. Amer.*, vol. 109, pp. 742-747, 2001.

[4] B. H. Story, and I. Titze, "Parameterization of vocal tract area functions by empirical orthogonal modes," *J. Phonetics*, vol. 26, pp. 223-260, 1998.

[5] H. Hirai, K. Honda, I. Fujimoto, and Y. Shimada, "Analysis of magnetic resonance images on the physiological mechanisms of fundamental frequency control," *J. Acoust. Soc. Jpn.*, vol. 50, pp. 296-304, 1994.

[6] K. Honda, "Formant frequency shift due to fundamental frequency change," IEICE Tech. Rep. SP86-122, 1986.

[7] G. Fant, "Vocal-tract area and length perturbations," *STL-QPSR* vol. 4, pp. 1-14, 1975.