Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

# Observation of voiced consonants using MRI movie

**Takayoshi Nakai(1), Kotaro Hirayama(1) and Keizo Ishida (1)**

(1) Facuty of Enginerring, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, 432-8561, JAPAN

**PACS:** 43.70.BK

## ABSTRACT

We found that the sound is radiated from the nostrils during pronounce of the buzz of voiced consonants. However, how to produce voiced consonants is not known well. Therefore, movies on mid-sagittal plane of the head were made using f-MRI. Speech materials are /cVcV/, where /c/ is one of voiced consonants /b/, /d/ and /g/, and V is one of 5 vowels. There are 3 male speakers. The frame period is 16.7 milli-second. The length whose velum touches the back vocal tract wall, and the distance between the upper part of the velum and the back vocal tract wall were measured on every frame by visual judgment. At the beginning of the initial buzz of a word, the velum does not touch the back vocal tract and has a 1mm distance from the back vocal tract. But for medial buzz of a word, the velum completely contacts the back vocal tract. It is the same as for vowels. For simulation of this phenomenon, we tried the analysis using the electrical equivalent circuit mode of the tube with loss. From the result, it is seen that at the beginning of the beginning buzz of a word, if the distance between the velum and the back vocal tract is 1mm, the sound can be radiated from the nostrils and is similar to one for the nasal consonants.

## INTRODUCTION

We found that the sound is radiated from the nostrils during pronounce of the buzz of voiced consonants. However, how to produce voiced consonants is not known well. In order to clarify how to produce voiced consonants, movies on mid-sagittal plane of the head were made using f-MRI. Speech materials are /cVcV/. The length whose velum touches the back vocal tract wall, and the distance between the upper part of the velum and the back vocal tract wall were measured on every frame by visual judgment. The results of measuring are shown. Buzz production will be estimated from the simulation of equivalent electrical ciruits of the tube based the results.

## MRI MOVIES

Movies on mid-sagittal plane of the head were made using f-MRI at ATR Brain Activity Imaging Center, Kyoto, Japan.. Speech materials are /cVcV/, where /c/ is one of voiced consonants /b/, /d/ and /g/, and V is one of 5 vowels. There are 3 male speakers. The image element is 1 mm wide and 1 mm high. The frame size is 256 pixel wide and 256 pixel high. The frame period is 16.7 milli-seconds.

## MEASUREMENT OF MOVEMENT OF VELUM

It is difficult to measure the movement of the velum, since it is very fast and complex.

The thicknessof the velum was measured [1]. The results showed as follows. (1) The thickness is changed for vowels. (2) For /b/and /g/ consonants the thickness is changed due to the following vowels, but the thickness of the consonant and its following vowel is not changed.

This time, we measured the distance between the upper part of the velum and the back vocal tract wall as shown in Figure 1, and the contact length whose velum touches the back vocal tract wall as shown in Figure 2.

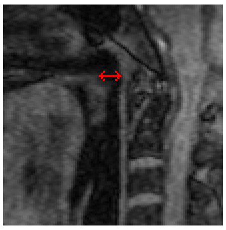

**Figure 1** Distance between the velum and the vocal tract (red arrows).

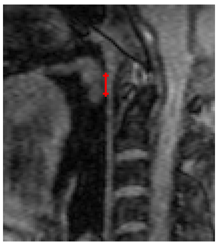

**Figure 2** Contact length between the velum and the vocal tract (red arrows).

For consonants, it is observed that the timing of the pronaunce is partially messed up, and its image blurs, since one movie consists of 128 times pronounce. When we measure "contact length," we make judgments as "do not contact" in the case that the image blurs compared with image when the velum and the back vocal tract wall were closely-attached.

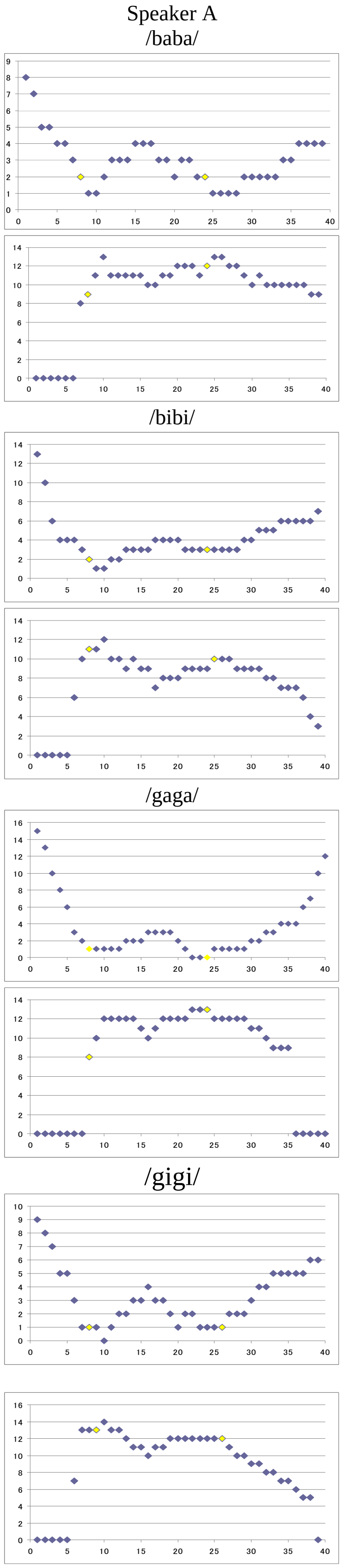


**Figure 3** Distance between the velum and the vocal tract (upper) and contact length between them (lower) uttered by speaker A. (vertical unit:mm, holizontal number: frame number, and yellow square: instance of plosion)

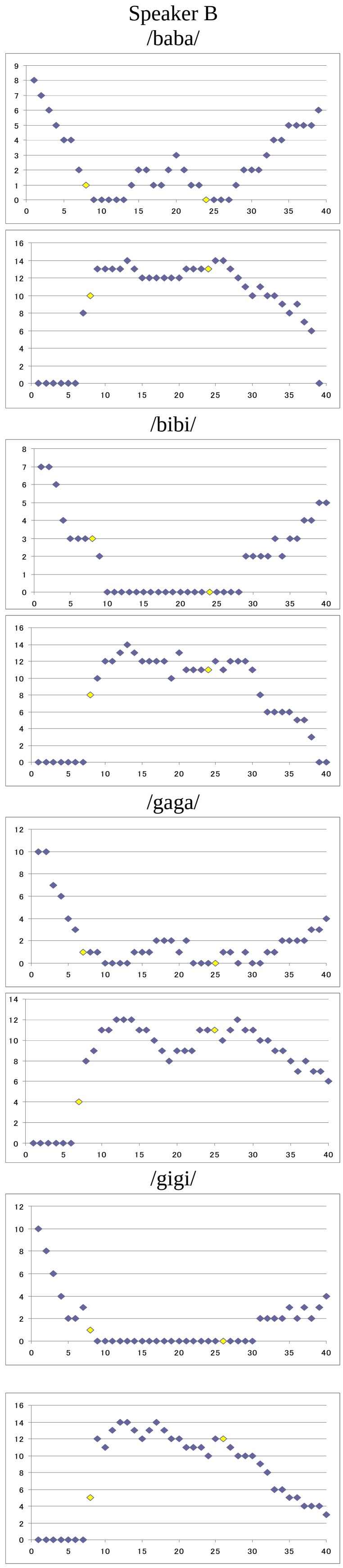


**Figure 4** Distance between the velum and the vocal tract (upper) and contact length between them (lower) uttered by speaker B. (vertical unit:mm, holizontal number: frame number, and yellow square: instance of plosion)

## OBSERVATION RESULTS

Figures 3 and 4 show measurement resuls of speakers A and B, respectively. From the top, /baba/, /bibi/, /gaga/ and /gigi/ are shown. Something that has a commonality, is to have differenct movement of the velum between buzz of the initial and the second consonants. For the buzz of the initial consonants, the velum is drooping and has some distance from the back vocal tract wall before pronounce as shown in figure 1. After pronounce, the velum moves up and to the back vocal tract wall, and touches the back vocal tract just before plosion, not more than two frames before. Therefor, for pronounce of buzz, the velum does not close up, has a 1 mm distance from the back vocal tract wall, or less tightly touch the back vocal tract wall than for vowels. For the the beginning of pronounce of buzz, the vocal cords can vibrat, since the velum do not close. For buzz of the second consonants, the velum closes and is not clear the difference from for the previous and following vowels.

## SIMULATION BY EQUIVALENT ELECTRICAL CIRCUT MODEL OF THE TUBE WITH LOSS

This time, for the losses of the tube, we consider that there are losses of vibration of the vocal tract wall, and of viscous friction and heat conduction of air, since the width of the tube is down to 1 mm.

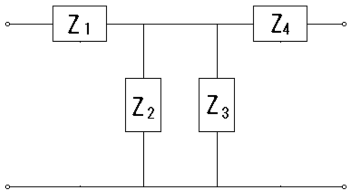


**Figure 5** One section of the equivalent electrical circuit model of the tube.

One section of the tube is shown in Fifure 5, where the acoustical impedance $Z_1$ and $Z_4$ considered the loss of the viscous friction of air [2],

$$Z_1(x,\Omega) = Z_4(x,\Omega) = \left[\frac{S(x)}{[A_0(x)]^2}\sqrt{\Omega\rho\mu/2} + j\Omega\frac{\rho}{A_0(x)}\right]\times\frac{1}{2},$$

the acoustical impedance $Z_2$ considered the loss of the heat conduction of air [2],

$$Z_2(x,\Omega) = 1\Big/\left[\frac{S(x)(\eta-1)}{\rho c^2}\sqrt{\frac{\lambda\Omega}{2C_p\rho}} + j\Omega\frac{A_0(x)}{\rho c^2}\right]\times l,$$

and the acoustical impedance $Z_3$ considered the loss of the vocal tract wall [2],

$$Z_3(x,\Omega) = \frac{b_\omega + j\Omega m_\omega}{S(x)l},$$

where

$S(x)$: circumference length of the cross sectional area $A_0(x)$

$\Omega$: angular frequency

$\rho$: density of air, 1.15kg/m$^3$, c: sound velocity in air, 353m/s

$\mu$: coefficient of viscosity of air, 0.1844x10$^{-4}$ Pa/s

$\eta$: ratio of specific heat at constant volume to that at constant pressure in air, 1.4

$\lambda$: coefficient of heat conduction in air, 23x10$^{-3}$ kg m/(s$^3$K)

$C_p$: coefficient of specific heat at constant presuure in air, 1005.6 m$^2$ / Ks$^2$

l: length of the section of the vocal tract

$b_w$: mechanical resistance of vocal tract wall, 16000 kg/(m$^2$K)

$m_w$: mass of vocal tract wall per unit area, 15 kg/ m$^2$.

The radiation impedance of the lips or the nostrils, Z, is approximated as the radiation from the infinite buffle:

$$Z_L(\Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r},$$

where

$$R_r = \frac{128\rho c}{9\pi^2 A},$$

$$L_r = \frac{8\rho}{3\pi\sqrt{\pi A}},$$

where *A* is radiation area of the lips or the nostril.

Figure 6 shows the equivalent electrical circuit model of the vocal tract.

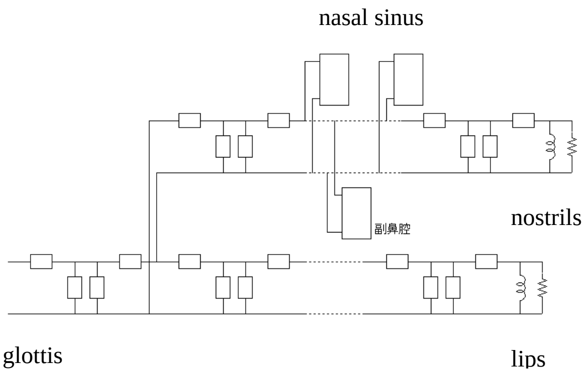


**Figure 6** The equivalent electrical circuit model of the vocal tract.

We simulate the output to input characteristics of the model as shown in figure 6. Simulated sounds are nasal consonant /m/, and voiced consonants /b/, /d/, and /g/ in the case of their following vowel /a/ and /i/. The lengths of the oral and nasal-cavities are 17.5 cm and 12.5 cm, respectively. The section numbers of the oral and nasal cavities are 10 and 9, respectively.

The cross-sectional area of each section of the oral cavity for /m/ is

{2.0, 5.0, 3.7, 3.8, 4.2, 2.6, 7.5, 9.0, 5.0, 5e-4} cm$^2$

from the glottis side. Its cross-sectional area for /b/ is

{2.0, 8.0, 8.0, 8.0, 8.0, 6.0, 2.0, 2.0, 2.0, 5e-4} cm$^2$

from the glottis side. The cross-sectional area each section of the nasal cavity is

{4.0, 4.0, 4.5, 6.0, 7.0, 5.5, 4.5, 3.0, 2.0} $cm^2$

from the velum side. The depth length of the first section of the nasal cavity of voiced consonants is 2cm after [3]. Their widths are changed from 0.3 mm to 5 mm. For their widths, sound does not become disturbed flow.

## SIMULATION RESULTS AND DISCUSSIONS

Figure 7 shows the ratio of output pressure to input pressure for nasal consonant /m/. It is seen that the peak gain less than 500Hz for /m/ is 15dB.Figure 8 shows the ratio of output pressure to input pressure for voiced consonant /b/ in the case that wide length of the velum is changed 0.5 mm to 4mm. Since the gain is changed at less than 500 Hz, Figure 9 shows the ratio of output pressure to input pressure at less than 500Hz. For widths 2 to 5 mm of the first section of the nasal cavity, peak frequencies are about 300 Hz and the peak gains are about 15 dB; these are the almost same as for /m/. For 1mm wide of the first section, the peak frequency is less than 300 Hz, the gain is about 10 dB.

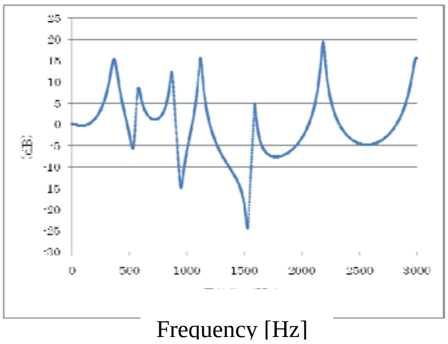


**Figure 7** Frequency characteristics of consonant /m/.

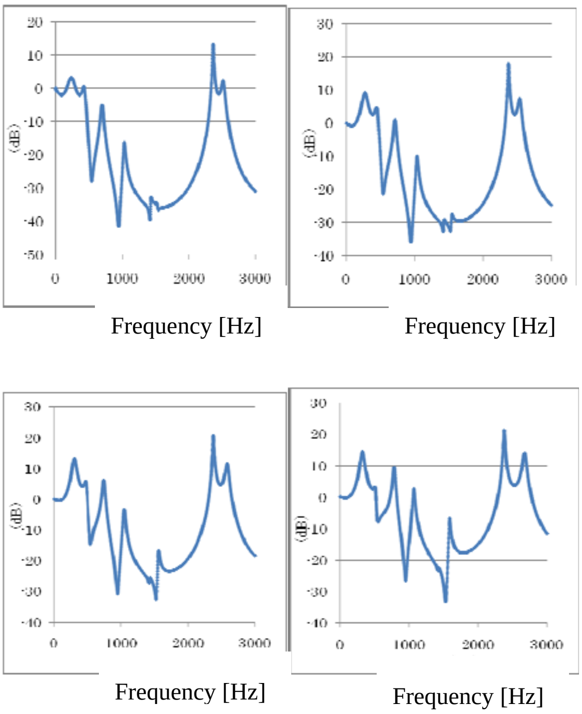


Figure 8 Frequency characteristics of buzz of voiced consonant /b/ in the case that widths of the velum are changed 0.5 mm(upper, left), 1 mm(upper, right), 2 mm(lower, left), and 4 mm(lower, right).

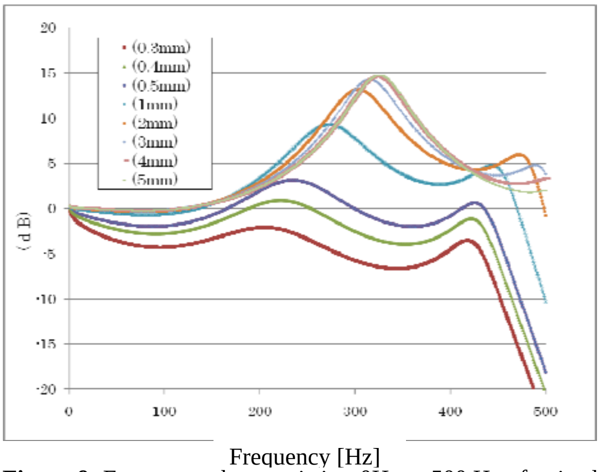


**Figure 9** Frequency characteristics 0Hz to 500 Hz of voiced consonant /b/ in the case that wide length of the velum is changed 0.3 mm to 5 mm.

From our acoustical measurements, for buzz of initial voiced consonants, the radiation from the nostrils is the same or 5dB less than for /m/ [4]. Therefore, for the beginning of buzz of initial voiced consonants, it is highly possible that the velum is open.

For buzz of second voice consonants, the velum is close. Therefor, the other model is needed as leakage from vibrating velum [1].

## CONCLUSIONS

The movies on mid-sagittal plane of the head were made using f-MRI. Speech materials are /cVcV/, where /c/ is one of three voiced consonants, and V is one of 5 vowels. The length whose velum touches the back vocal tract wall, and the distance between the upper part of the velum and the back vocal tract wall were measured on every frame. It is shown that at the beginning of the initial buzz of a word, the velum does not touch the back vocal tract and has a 1mm distance from the back vocal tract, however, for medial buzz of a word, the velum completely contacts the back vocal tract. We simulated frequency characteristics of the vocal tract at the beginning of the initial buzz of a word by equivalent electrical cuircut model of the tube. It is shown that it is highly possible that the velum is open.

## REFERENCES

1 Y. Piao, J. Dang, T. Nakai, and M. Akagi, "Modeling of Velum Function in Speech Production," TECHNICAL REPORT OF IEICE, SP2006-33, pp.37-42 (2006.7) in Japanese

2 L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech signals* (Prentice-Hall, Englewood Cliffs, New Jersey, 1978) pp. 66–69

3 A.Serrurier and P.Badin, "Three-dimensional articulatory modeling of velum," J.Acoust.Society of America, 123(4), pp2335-2355 (2008.4)

4 T. Nakai and M. Yamamoto, "Sound radiations from nostrils for viced speech and their individuality," 4th joint meeting of ASA/ASJ, 5pSc25, p.3375 (2006.12)