Acoustic Nature of The Whisper

Masahiro Matsuda and Hideki Kasuya
Faculty of Engineering, Utsunomiya University
7-1-2 Yoto, Utsunomiya, 321-8585 Japan
Tel: +81 028 689 6122 E-mail: matsuda@klab.ishii.utsunomiya-u.ac.jp

10.21437/Eurospeech.1999-37

ABSTRACT

The lower formant frequencies of the whispery vowel are known to be slightly higher than those of the modal vowel. This paper attempts to interpret this phenomenon acoustically, based on an electrical circuit model of the vocal tract, taking into account acoustic coupling with the subglottal system. A three-dimensional vocal tract shape was measured from a magnetic resonance image (MRI). It was found that the narrowing of the tract in the false vocal fold regions and weak acoustic coupling with the subglottal system are primary causes of the rise of the lower formant frequencies.

Keywords: whisper, acoustic model, MRI measurement

1. INTRODUCTION

The acoustic nature of the whisper is not yet fully understood, although it is often used in everyday verbal communication. The whisper has been regarded occasionally as a simple modification of breathy voice [1]. Tsunoda et al. made a physiological measurement of the laryngeal shape during whispering by using magnetic resonance imaging and found that the supra-glottal structures were not only constricted but also shifted downward, attaching to the vocal fold to prevent vocal fold vibration [2]. In the meantime, the lower formant frequencies of the whispery vowel are known to be slightly higher than those of the modal vowel [3]. Acoustic interpretation, however, has not yet been carried out on this phenomenon. In this paper, we attempt to explain this phenomenon acoustically, based on an electrical circuit model of the vocal tract, taking into account acoustic coupling with the subglottal system. A three-dimensional vocal tract shape was measured from the magnetic resonance image (MRI). This paper will show that the narrowing of the tract in the false vocal fold regions and weak acoustic coupling with the subglottal system are primary causes of the rise of the lower formant frequencies.

2. ACOUSTIC ANALYSIS

2.1 Method

We made an acoustic analysis for distinguishing the acoustical difference between the modal voice and whisper. Five Japanese vowels, /i/, /e/, /a/, /o/, and /u/, of the modal voice and whisper were uttered alternately in a sound proof room by seven adult male subjects (labeled Subject A through G) who were trained to maintain the

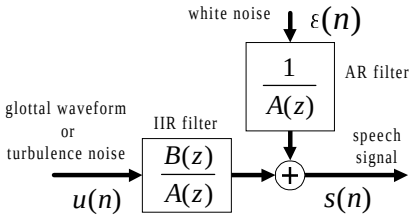

Figure 1. ARX model of speech production.

articulatory gesture for the two types of vowels as consistently as possible and were recorded with a digital tape recorder. The auto-regressive with exogenous input (ARX) analysis method was applied for the measurement of the formant frequencies of the vowel data [4]. Figure 1 shows the ARX model consisting of IIR and AR filters. The vocal tract transfer function is represented by *B(z)/A(z)*. In the analysis of the modal vowel, the input glottal waveform *u(n)* is the one produced by R+ model

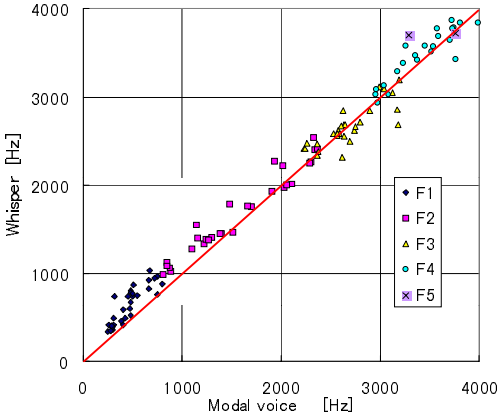

Figure 2. Formant frequency of the modal and corresponding whispery vowels.

Table 1. Measured formant frequencies of subject A.

		/a/	/i/	/u/	/e/	/o/
Modal vocie	F1	654	302	375	541	458
	F2	1060	2057	1208	1784	807
	F3	2375	3004	2165	2452	2379
	F4	3293	3486	3129	3341	3066
	F5	4363	4526	4146	4589	4064
Whisper	F1	813	336	617	695	559
	F2	1227	1957	1303	1763	906
	F3	2281	2954	2281	2315	2487
	F4	3220	4299	3151	3199	3162
	F5	4510	5342	4466	4606	4398
Ratio	F1	1.24	1.11	1.65	1.28	1.22
	F2	1.16	0.95	1.08	0.99	1.12
	F3	0.96	0.98	1.05	0.94	1.05
	F4	0.98	1.23	1.01	0.96	1.03
	F5	1.03	1.18	1.08	1.00	1.08

[5] and *B (z)*=1. For the whispery vowel, on the other hand, a white noise signal is used as the input to the vocal tract IIR filter, where *B(z)* is composed of two zeros. The two zeros have been introduced so that *B(z)* compensates for the frequency spectrum of the turbulence noise source which is assumed to be slightly inclined toward a higher frequency.

2.2 Result

Figure 2 illustrates the distribution of the formant frequencies of the modal and corresponding whispery vowels. The x-axis shows the modal and the y-axis the whispery. A slanted line passing through the origin

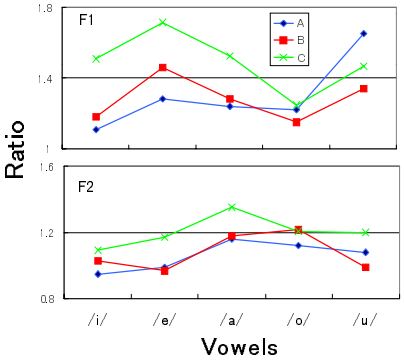

Figure 3. Ratios of the first two formant frequencies for subjects A, B and C.

corresponds to the identical frequency between the modal and the whispery vowels. From the figure, it is apparent that the formant frequencies of whisper below 1,800 Hz are slightly higher than those of the modal. Table 1 summarizes the measured formant frequencies of subject A, where the ratio of the whispery to the modal formant frequency is tabulated as well. The frequencies shown are averaged values over the utterance of about one second. The ratios of the first formant frequency are larger than one for all the five vowels but only the vowels /a/ and /o/ yield the ratio exceeding one in the second formant frequency. The first formant frequency of the

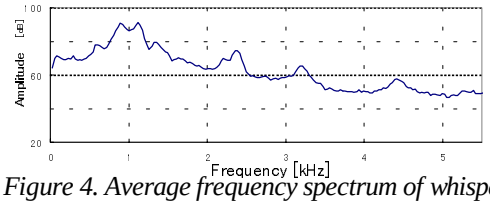

Figure 4. Average frequency spectrum of whispered /a/.

whispery vowel /u/, in particular, shows a larger rise. Ratios of the first two formant frequencies for Subjects A, B, and C are shown in Figure 3, which indicates that the tendency of the rise of the first two formant frequencies of the whispery vowels seems to be consistent among the three subjects.

Figure 4 shows an average frequency spectrum of the whispery vowel /a/ uttered by Subject A. No significant anti-formant exists, suggesting a weak coupling with the subglottal system. This is the case for all the other Japanese vowels.

3. OBSERVATION OF THE LARYNX BY LARYNGEAL ENDOSCOPE

In order to observe the laryngeal structure during the whispering of the five Japanese vowels, a laryngeal endoscope was inserted through the nasal tract of the subject. Figure 5 illustrates images of the laryngeal structure for

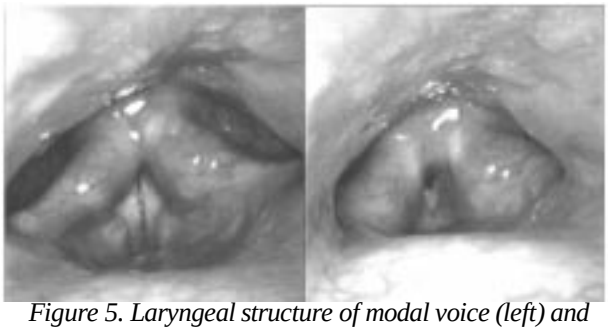

Figure 5. Laryngeal structure of modal voice (left) and whisper (right) of subject A.

the two different types of vowels, modal and whispery, taken with the endoscope. We see two major differences between the modal voice and whisper: In the whisper, 1)the supra-glottal structure is constricted in the false vocal fold regions and the vocal folds are covered by the false folds, and 2) the glottis is opened to a slight extent.

4. MEASUREMENT OF VOCAL TRACT SHAPE FROM MRI

4.1 Method

The MRI during whispering was taken to measure the three-dimensional vocal tract shape quantitatively. Subject A participated in the experiment. He was again instructed before the experiment to maintain the same articulatory gesture for the modal and whispery vowels. In order to obtain images of high resolution, the imaging

process consisted of five sessions, each of which took ten seconds. In each session, ten images were taken at every 5mm depth. From these MRI data, the three dimensional vocal tract shape was reconstructed, which was followed by the computation of the cross-sectional

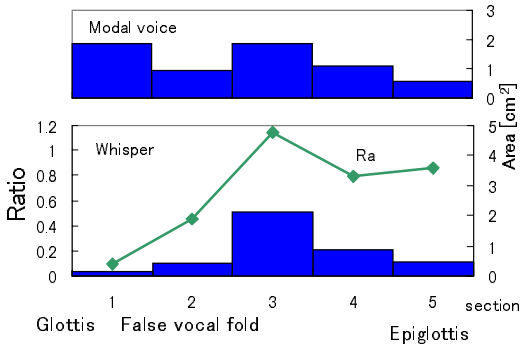

Figure 6. Vocal tract cross-sectional areas and the ratio for the modal and whispery vowels /a/.

area functions [6].

4.2 Result

Figure 6 shows the vocal tract cross-sectional areas and the ratios of the whispery to the modal vowel /a/. The cross-sectional area of whisper reduces between the glottis and the top of the epiglottis, especially in the false vocal fold regions. This result agrees with the qualitative observations through the endoscope and with the report of Tsunoda et al. [2].

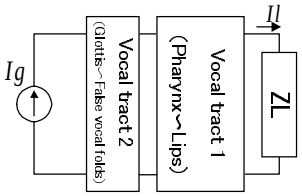

Figure 7. Electrical circuit model for the modal vowel.

5. ELECTRICAL CIRCUIT MODEL

5.1 Electrical circuit model

To make a comparison of the formant frequency configuration between the modal and whispery vowels, the vocal tract was divided into two parts, Vocal tracts 1 and

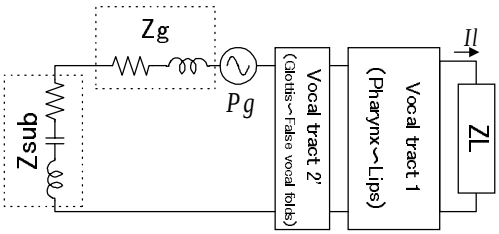

Figure 8. Model 1 of whispering.

2 (or 2') as shown in Figure 7, where an electrical circuit model is shown for the production of the modal vowel. Vocal tract 1 represents the pharynx to the lips and is identical in all the models described below. Vocal tract 2 (or 2') simulates the first two sections corresponding to the false vocal fold regions shown in Figure 6. Four electrical circuit models were taken into consideration for

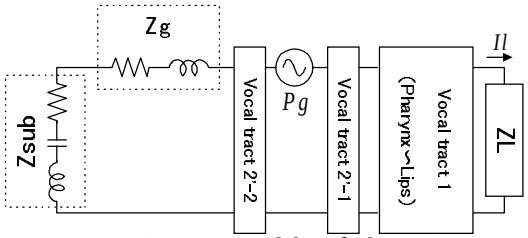

Figure 9. Model 4 of whispering.

the production of whispery vowel.

Model 1: The turbulence noise source is assumed to be located at the glottis and the vocal tract shape is based on the measurements from the MRI (Vocal tract 2'). The glottal and subglottal impedance elements, *Zg* and *Zsub*, are incorporated (Figure 8).

Model 2: The Vocal tract 2' of Model 1 is identical to the Vocal tract 2 of the modal vowel. This is to examine the effect of narrowing in the false vocal fold regions.

Model 3: The third vocal tract section in Figure 6 was removed. This is to examine the effect of shortening of the vocal tract length due to the rise of the glottis.

Model 4: The turbulence noise source is assumed to be located downstream the glottis by one section (Figure 9).

Norton's theorem was applied to the models of whis-

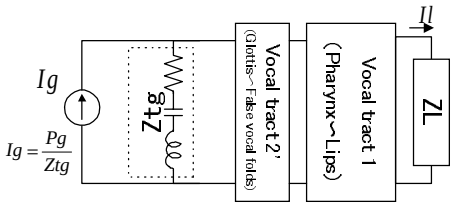

Figure 10. Modified Model 1 obtained by Norton's theorem.

pery vowels, resulting in Figure 10. The transfer function was calculated for each of the models using the method by Sondhi & Schroeter [7]. Impedance values of the subglottal system used were from Westbury [8].

5.2 Result

Figure11 shows an example of the calculated frequency transfer characteristics for the modal and the whispery vowels /a/. The lower formant frequencies of the whisper are slightly higher than those of the modal voice and the spectral tilt is very large for the whispery vowel. Table

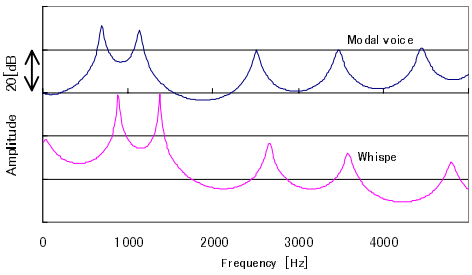

Figure 11. Calculated frequency transfer characteristics of the modal and whispery vowels /a/.

Table 2. Calculated formant frequencies of the modal vowels.

	/a/	/i/	/u/
Fm1	690	330	430
Fm2	1130	2100	1310
Fm3	2500	2590	2300
Fm4	3470	3450	3360
Fm5	4450	4160	4170

2 represents the calculated formant frequencies for the three modal vowels. Table 3 summarizes the calculated formant frequencies and the ratios of the whispery to the modal formant frequency for the four types of models to simulate the production of a whisper.

6. DISCUSSION

Tendency of the formant frequency ratios between the whispery and the modal obtained from Model 1 is consistent with the one observed in the natural utterances shown in Table 1. The second formant frequency rises only in the vowel /a/, while the first formant frequency of the vowel /u/ yields a larger value. The results from Model 2, which are in large disagreement with the natural utterance, suggest the importance of the constriction of the tract in the false vocal fold regions. An assumption that the turbulence noise source during whispering is located 0.85 cm downstream the glottis (Model 4) results in too much increase in the first formant frequency. The results from Model 3 are close to those of Model 1 but all the ratios seem to be slightly higher than the ones observed in the natural utterances. In summarizing the results shown in Table 3 and Figure 11 where no significant zero exists, weak acoustic coupling with the subglottal system and the constriction in the false vocal fold regions are primary causes of the rise of the lower formant frequencies.

7. CONCLUSION

Simulation experiments based on an electrical circuit model of the articulatory configuration of whispering indicate that narrowing in the false vocal fold regions and weak acoustic coupling with the subglottal system are primary causes of the rise of the lower formant frequencies in the whisper.

Further analysis with the simulation is now underway to gain more insights into the acoustic nature of the whisper, in particular the source spectrum characteristics and formant bandwidth change.

REFERENCES

[1] I. Titze, *Principles of Voice Production* (1994), Prentice Hall, New Jersey, p. 116.

[2] K. Tsunoda, Y. Ohta, Y. Soda, S. Niimi, and H. Hirose (1997), Laryngeal adjustment in whispering. *Annals of Otology, Rhinology & Laryngology*, Vol.106, pp. 41-43.

[3] K. J. Kallail and F. W. Emanuel (1984), An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *J. Phonetics*, Vol.12, pp. 175-186.

[4] W. Ding, H. Kasuya, and S. Adachi (1995), Simultaneous estimation of vocal tract and voice source parameters based on an ARX model. *IEICE Trans. Inf. & Syst.*, Vol.E78-D, pp. 738-743.

[5] D. Klatt and L. Klatt (1990), Analysis, synthesis and perception of voice quality variations among female and male talkers. *J.Acoust. Soc. Am.*, Vol.87, pp. 820-857.

[6] C.S.Yang and H.Kasuya (1996), Speaker individuality of vocal tract shapes of Japanese vowels measured by magnetic resonance images. *Proc. ICSLP96, Philadelphia*, Vol.2, pp. 949-952.

[7] M. M. Sondhi and J. Schroeter (1987), A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Trans. ASSP*, Vol.35, pp.955-967.

[8] J. R. Westbury (1983), Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *J. Acoust. Soc. Am.*, Vol.73, pp. 322-336.

Table 3. Calculated formant frequencies and ratios of the whispery to the modal formant frequency for the four types of models.

		Model_1			Model_2			Model_3			Model_4		
		/a/	/i/	/u/	/a/	/i/	/u/	/a/	/i/	/u/	/a/	/i/	/u/
Formant frequency	Fw1	880	410	830	1030	1930	1280	900	430	890	1060	2100	1310
	Fw2	1370	2170	1390	2380	2510	2160	1480	2300	1420	2580	2650	2470
	Fw3	2650	2680	2610	3380	3410	3240	2770	2990	2690	3580	3530	3720
	Fw4	3580	3520	3790	4260	4100	4110	3670	3670	3890	4830	4370	4530
	Fw5	4790	4350	4590			4990	4890	4720	4780			
Ratio (Fwi/Fmi)	R1	1.28	1.24	1.93	1.49	5.85	2.98	1.30	1.30	2.07	1.54	6.36	3.05
	R2	1.21	1.03	1.06	2.11	1.20	1.65	1.31	1.10	1.08	2.28	1.26	1.89
	R3	1.06	1.03	1.13	1.35	1.32	1.41	1.11	1.15	1.17	1.43	1.36	1.62
	R4	1.03	1.02	1.13	1.23	1.19	1.22	1.06	1.06	1.16	1.39	1.27	1.35
	R5	1.08	1.05	1.10			1.20	1.10	1.13	1.15			