2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

New Anthropomorphic Talking Robot having a Three-dimensional Articulation Mechanism and Improved Pitch Range

Kotaro Fukui, Yuma Ishikawa, Takashi Sawa, Eiji Shintaku, Masaaki Honda, Atsuo Takanishi,
Waseda University, Tokyo, Japan

Abstract— **We have developed a new three-dimensional talking robot Waseda Talker No. 6 (WT-6), which generates speech sounds by mechanically simulating articulatory motions as well as aero-acoustic phenomena in the vocal tract. WT-6 has 17-DOF vocal mechanism. It has three-dimensional lips, tongue, jaw and velum which form the 3D vocal tract structure. It also has an independent jaw opening/closing mechanism, which controls the relative tongue position in the vocal tract as well as the oral opening. The previous robot in the series had a 2D tongue and was not able to realize precise closure to produce human-like consonants such as /s/ or /r/. The new tongue, which can be controlled to form 3D shapes, is able to produce more realistic vocal tract shapes. The vocal cord model was also improved by adding a new pitch control mechanism that pushes from the side of the vocal cords. The pitch range is broader than that of the previous robot; it is sufficiently broad so as to be able to reproduce normal human speech. Preliminary experimental results showed improved synthesized speech quality for the vowels /a/, /u/ and /o/.**

I. Introduction

GIVEN the importance of speech in human communication, considerable research has been carried out to clarify the mechanisms of human speech. However, it is difficult to simulate human speech mechanisms using computer models since they involve complex aero-acoustics and the movement of speech organs. Since 1998, we have been developing mechanical models of the human speech organs to clarify the mechanisms of speech. We have developed talking robots that produce vowel and consonant sounds in a similar way to humans by making changes in the vocal tract area [1]. The voice production mechanism of an anthropomorphic talking robot is the same as that of a human. The airflow from the mechanical lungs causes the vocal cords to vibrate, producing a source sound that is articulated in the vocal tract and controlled by the tongue, the palate, the nasal cavity and the lips.

Other voice synthesis machines have been developed by Kempelen [2], Umeda [3], Kawamura [4], Sawada [5] and the other researchers.

In 2005, Waseda Talker No. 5 (WT-5) was developed with a two-dimensional tongue mechanism, which was able to produce a transition in the vocal tract area in the same manner as a human to produce the Japanese vowels (/a/, /i/, /u/, /e/ and /o/) and consonant sounds. WT-5 also had a new vocal cord mechanism that mimics the human biomechanical structure. And its voice was more human-like compared to those of previous robots [6].

However, a 2D tongue model was found to be inadequate for clarifying the mechanisms of human speech. And the voices of the previous robots were still not acceptable. The sound parameters, in particular the first and second formants (F1, F2), which are very important parameters in the production of vowels, were in the human range, however the robot did not sound like a human. It was considered that approaches that more closely reproduce the 3D tongue shape were required to further investigate human speech. In view of this, a novel approach must be taken if the human voice is to be reproduced accurately.

One major problem still remained in the WT-5's vocal cords, namely, the pitch control range was narrower than previous robots and was not wide enough to reproduce human speech.

We investigate a new mechanism to overcome these problems. The new anthropomorphic talking robot, Waseda Talker No. 6 (WT-6), had 1-DOF lungs, 5-DOF vocal cords, and articulators (5-DOF tongue, 1-DOF jaws, 1-DOF soft palate, 4-DOF lips and nasal cavity). Thus, WT-6 had a total of 17 DOF. The tongue had a 3D shape to improve the closure of the vocal tract and to reproduce the human vocal tract shape. The robot also had a new pitch control mechanism that pushes from the side of the vocal cords.

In this paper, we describe the details of the new anthropomorphic talking robot, WT-6, which is designed to reproduce human speech mechanisms in a more biological manner—the design is shown in Fig. 1.

Manuscript received September 15, 2006. This work was supported in part by the Grant-in-Aid for Scientific Research (A), 16200015 from MEXT, Japan.

K. Fukui, Y. Ishikawa, T. Sawa, E. Shintaku and A. Takanishi are with the Department of Mechanical Engineering, School of Science and Engineering, Waseda University, Tokyo, 196-8555, Japan (corresponding author, phone: +81-3-5286-3257; fax: +81-3-5273-2209; e-mail: kotaro@toki.waseda.jp).

K. Fukui is a JSPS research fellow, and A. Takanishi is a member of the Humanoid Research Institute and the Advanced Research Institute for Science and Engineering of Waseda University, Japan.

M. Honda is with the Department of Sport Medical Science, School of Sport Sciences, Waseda University, Saitama, Japan.

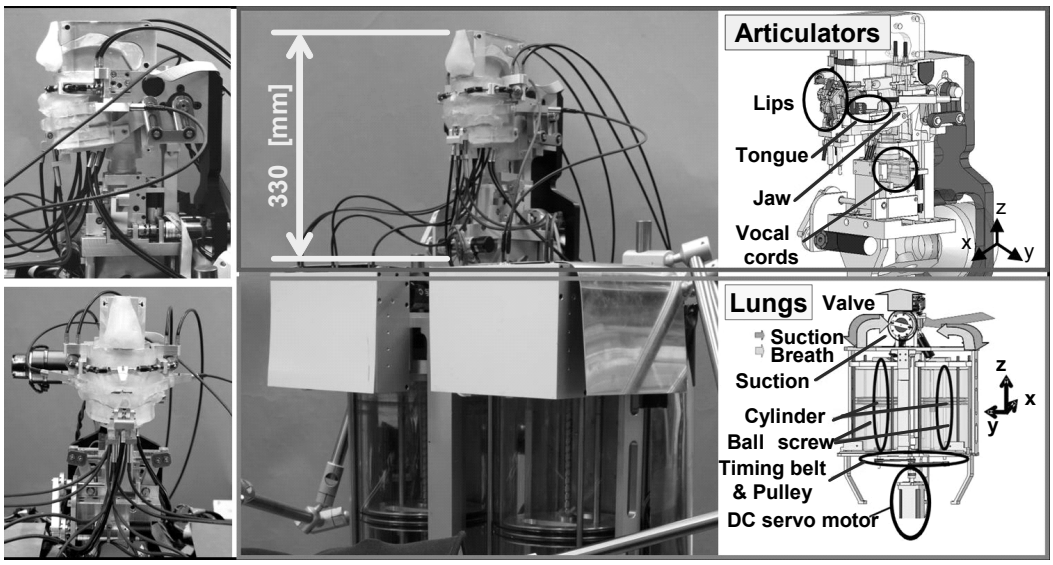


Fig. 1 WT-6's Overview

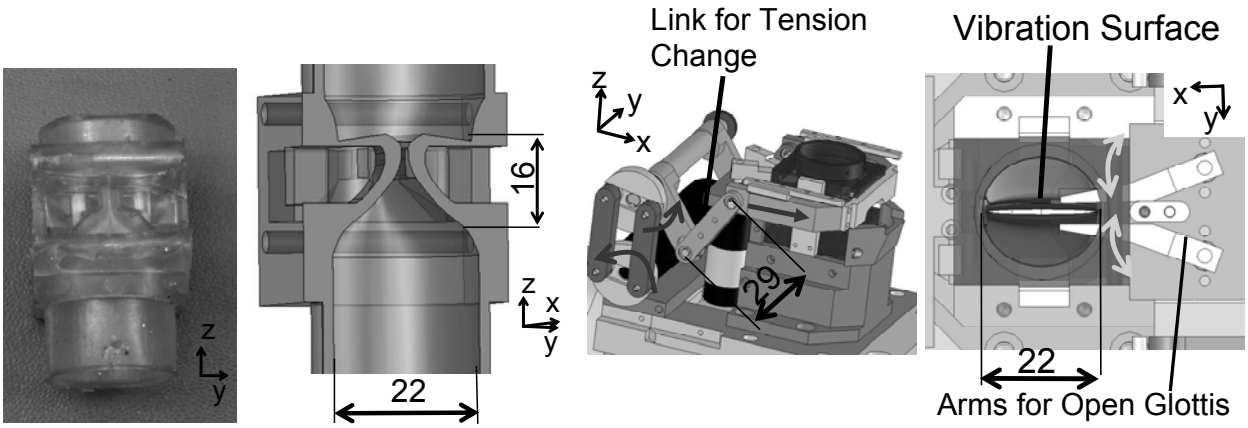


Fig.2 WT-5's vocal cord mechanism

II. Previous Talking Robot WT-5

WT-5 was developed in 2005 and has 1-DOF lungs, 3-DOF vocal cords, and articulators (7-DOF tongue, 1-DOF soft palate, 1-DOF teeth, 5-DOF lips and nasal cavity), so that it has a total of 18 DOF. The length of its vocal tract is 170 [mm], which is approximately equivalent to that of an adult male. Waseda Talker No. 5 can produce all 50 Japanese sounds, and its spectrum was similar to that of human speech. In this section we compare the speech production mechanism of WT-5 with that of WT-6.

A. Waseda Talker No. 5's Vocal Cords

The vocal cords we developed for WT-5 are based on the fold model which mimics the biomechanical structure of human vocal cords. The vocal cords of the previous robot, WT-4, were made from a thin soft rubber, EPDM [7]. In this mechanism, the pitch is controlled by rolling the rubber and the pitch control range is 140 [Hz]. This model is easy to make and to control pitch. However, the source sound from these vocal cords differed from that from humans. The spectrum of WT-4's vocal cords had many peaks and no attenuation at high frequencies. These were caused by the difference between the mechanisms of human vocal cords and WT-4's vocals cords. Human vocal cords have three layers, having a complex mechanism that involves the mucosa and the cartilages [8]. And the upper and lower sections of the fold vibrate with different phases. It is this

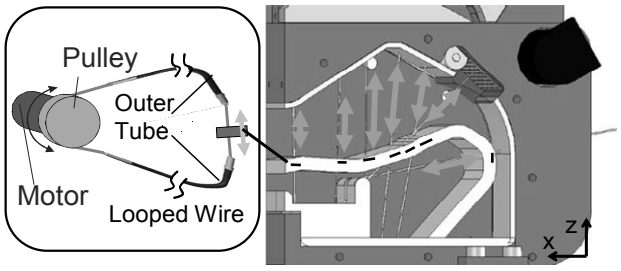


Fig.3 WT-5's Tongue Mechanism

which makes a human's voice sound soft.

To overcome these problems we developed WT-5's vocal cord's model having folds using thermoplastic rubber Septon by Kuralay Co. Ltd. [9], as shown in Fig. 2.Septon is a very elastic material and is easy to mold. In development, we found it difficult to reproduce the layer model using rubber, so instead we reproduced it by making hollows in the fold. This vocal cord model vibrated in the same way as human vocal cords, having different phases in the upper and the lower sections of the fold. In WT-5 voiced/unvoiced sounds were produced by using adduction/abduction of the glottal folds respectively, mimicking human arytenoid cartilages. The model also had a pitch control mechanism that involved varying the length of the glottis.

B. *WT-5's Tongue*

The tongue mechanism of WT-5 aimed to reproduce the area transition of the vocal tract as shown in Fig. 3. The tongue of WT-5 tongue is constructed from a 2D EPDM rubber plate, which has seven control points. The control points are made of little steel plates, controlled by looped wires connected to a DC servo motor. Using this mechanism the tongue could form various 2D tongue shapes, such as those necessary for producing the vowels and consonants sounds. WT-5 had a nasal cavity and a soft palate mechanism, and by opening the soft palate, the nasal cavity becomes connected to the oral cavity allowing nasal sounds, such as /n/ to be produced.

C. *WT-5 Lips*

WT-5's lips are made of Septon, the same material used for its vocal cords, and they were formed into the shape of human lips pronouncing /u/. The upper and the lower lip each have two DOF associated with opening and protrusion actions. In addition the lip mechanism also had a protrusion action in the corner of lips. This mechanism could produce the various shapes, such as the protrusion of /u/, and the closure of /p/. The teeth are set just behind the lips, and the closure of the teeth could be varied. The teeth are used when producing fricative consonants.

III. Development of New Vocal Cords

A. *The problems associated with WT-5's vocal cords*

The vocal cords of WT-5 can produce human-like source sounds, however, the pitch range of the vocal cords is narrow. WT-5 can produce sounds in the range 100 [Hz] -110 [Hz], which is insufficient for reproducing the pitch range of a human (adult male), namely 100 [Hz] -190 [Hz]. Thus it was necessary to develop a new pitch control mechanism to broaden the pitch range.

B. *The mechanism of WT-6's vocal tracts*

To overcome this problem, we analyzed the vocal cords using FEM (Finite Element Method) software. From this analysis, we found that the WT-5 pitch control mechanism had little affected on the vibrating fold. We require a system in which tension can be applied tension to the vibrating part.

In the vocal cords of WT-6, we added a mechanism which pushes the sides of the vocal cords, based on the analysis results as shown in Fig. 4. This mechanism affected the stiffness of the fold. And by pushing from the side, the vibrations of the airway become smaller and their speed becomes faster. And the vibrating fold is more tensioned by new mechanism.

C. *Experiment with the new vocal cords*

We experimented with the new vocal cords mechanism. The experiment results shows the pitch could be varied in the range 110 [Hz] -160 [Hz], thus the pitch range was 50 [Hz], as shown in Table1. This is an improvement over WT-5. And the spectrum of the sound source is close to that a human, having nearly ideal attenuation and few peaks, as shown in Fig. 5.

However, a human has pitch range of about 90 [Hz], and a trained adult male singer (baritone) can produce sounds over a 300 [Hz] range (98 [Hz] - 392 [Hz]). Thus, the mechanism still requires further improvement to reproduce human speech.

IV. Development of New Vocal Tracts

A. *Three-dimensional vocal tracts*

The most important parameters of the vocal tract are

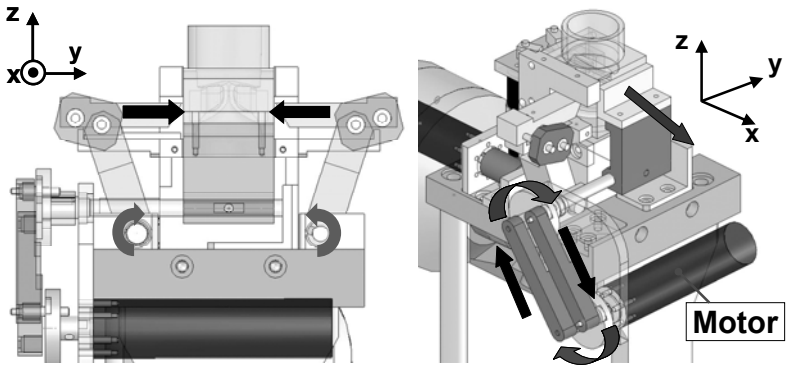


Fig. 4 WT-6'svocal cords mechanism

Table.1 Range of the pitch

	Frequency [Hz]	Pitch range [Hz]
New mechanism (WT-6)	117-152	35
WT-5	100-110	10
Adult male (Normal Speech)	80-180	100
Singing voice (Baritone)	98-392	294

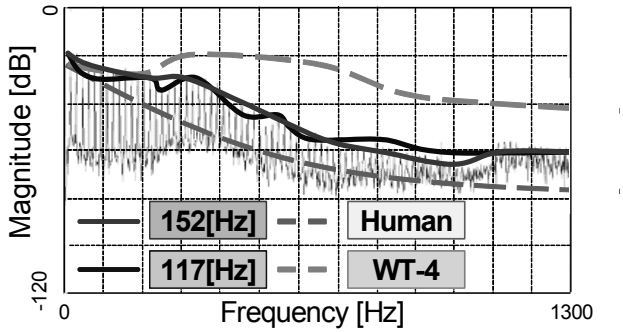


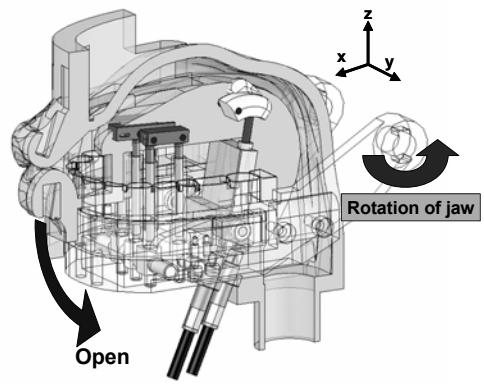
Fig. 5 Spectrum of glottal sound source

considered to be its length and its area transition [10]. However, these two parameters do not represent all the parameters required to produce a human-like voice, recent investigations have demonstrated the importance of shape parameters in addition to the area and the length of the vocal tract. However, the WT-5's tongue mechanism is based on a 2D model, making it impossible to reproduce the vocal tract shape precisely, especially in the case of liquid consonants such as /r/. The liquid consonants are produced by airflow through the side space produced by the palate and the tip of the tongue. It is therefore important to develop a 3D tongue shape that can reproduce human-like vocal tract shapes, and produce human-like vowels and consonant sounds.

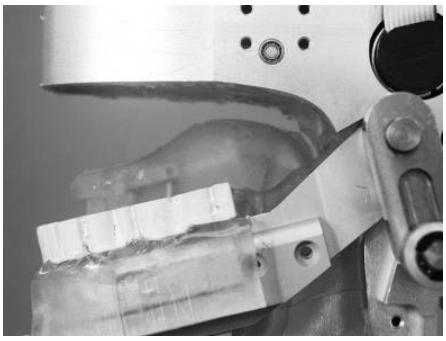
B. The mechanism of WT-6's vocal tracts

WT-6 had a jaw mechanism which mimics the ability of a human to reduce the tongue deformation, as shown in Fig. 6.We used a jaw mechanism which has a maximum tongue deformation of 7-8 [mm], which is smaller than the length of the fixed jaw which is 13 [mm].

However, the jaw mechanism varied the length from the jaw to the palate, and the looped wire mechanism is not adaptable. From analysis of the MRI (Magnetic Resonance Imaging) tongue data, at least 5 DOF are needed to control the tongue, and the tongue itself is too small to implement such a mechanism. To overcome these problems, we adopted a release mechanism. The efficiency of the power transition of the release mechanism depends on a pair of outer and inner tubes. In this mechanism we employed a coiled tube, the diameter of which is shown in Table 2.



(a) The mechanism



(b) Photo

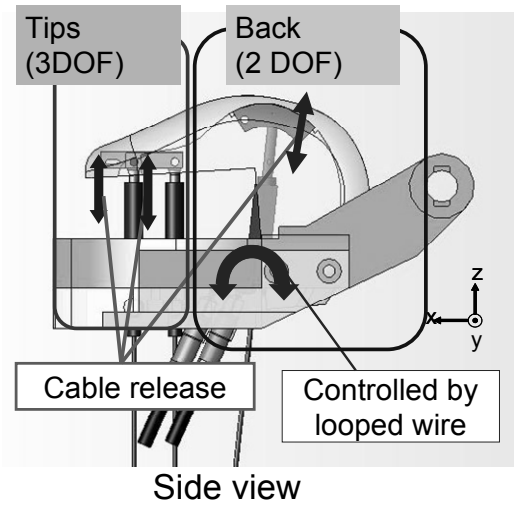
Fig. 6 WT-6's jaw mechanism

The tongue mechanism has seven release points (one in the middle of the first line, two at the side of the first line, two in the second line and two in the third line). The first line and the second line are connected to shape the parallel link mechanism, for varying the position and the inclination, and the mechanism could form a gutter in the center of the tip. This mechanism could change the shapes of the front part and make closure. The third line has a 1-DOF release mechanism and a 1-DOF rotation mechanism controlled by wire for changing the shape of the back position of the tongue as shown in Fig. 7.

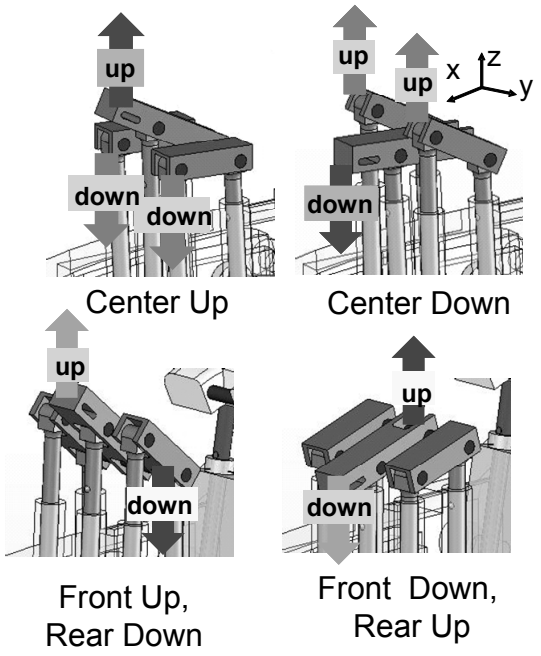
The lip mechanism also has a release mechanism to control the lower lip opening and the protrusion of the corner of the lips, as shown in Fig. 9.The lips are controlled by the jaws, the lower lip opening, the protrusion of the corner and protrusion of the upper and the lower lips. Each upper and lower protrusion mechanism is controlled by a looped wire. These mechanisms can produce various shapes of the lips. These tongue and lip mechanisms are covered by Septon rubber to avoid air leakage.

C. Experiments with the new vocal tracts

We experimented with the vocal tract mechanism by connecting it to the mechanical vocal cords. Fig. 10 shows the DFT (Discrete Fourier Transform) spectrum when pronouncing the vowel /a/. Compared to the DFT spectrum for WT-5, the WT-6 spectrum has peaks and troughs similar to those in the spectrum for humans, and the sound is more



(a) Jaw and tongue mechanism



(b) Parallel mechanism at tip of tongue

Fig.7 WT-6's tongue mechanism

human-like. Fig.11 shows the first and second formant mappings WT-6 compared with those for a human. It shows that the formants of the vowels for WT-6 lie within the human range, with the exception of the vowel /i/.

There are several reasons for this problem with the vowel /i/. The vocal tract for producing /i/ is narrow in the tips of the tongue. And the tongue mechanism has space at the side to allow it to move easily beside the original vocal tract. The second reason is the pressure. The vocal cords vibrate due to the difference in the pressure on the upper and the lower of the vocal cords. The narrow space causes the intraoral pressure to increase, making it difficult to vibrate the vocal cords. Thus WT-6 has limitations when constricted spaces are required.

Table 2 The property of release mechanism

	O.D. [mm]	I.D. [mm]
Outer Tube Stainless Tube with casing	2.3	1.2
Inner Tube Stainless Spring Tube	1.0	0.6

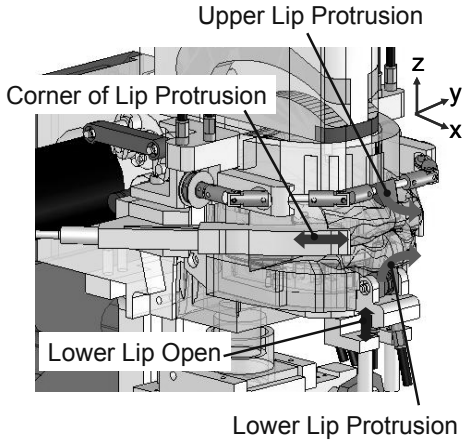
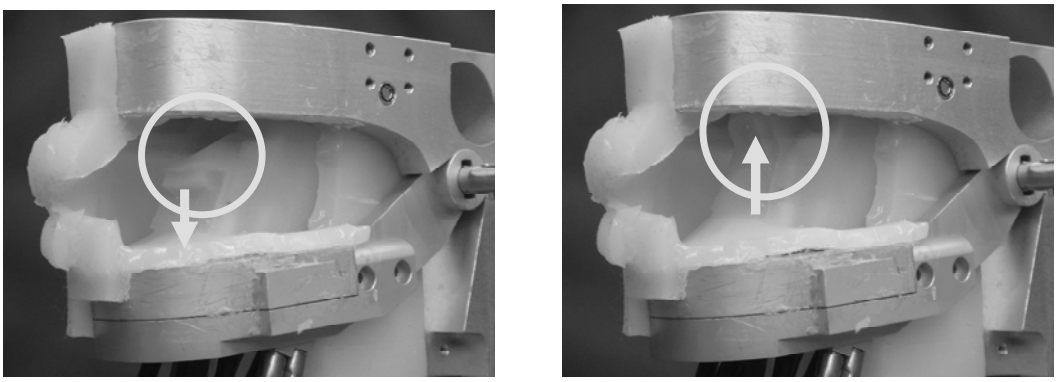
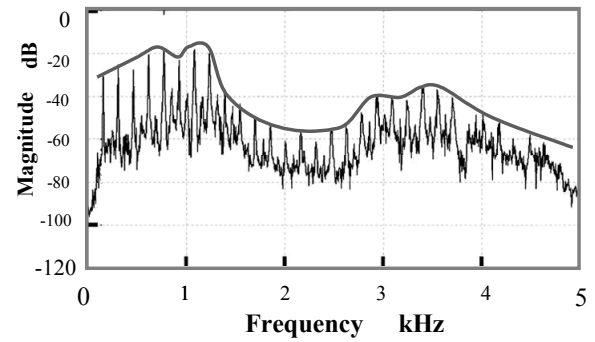


Fig. 9 WT-6's lip mechanism

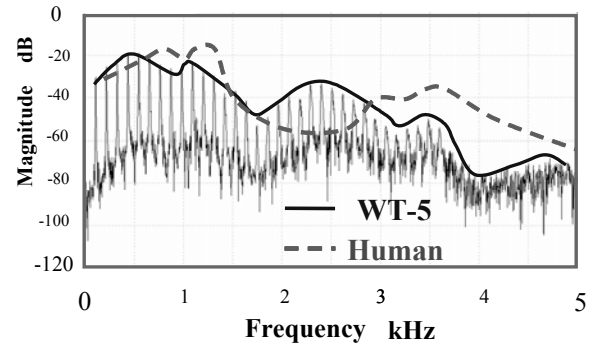


(a) Gutter by down of the center link (b)Peak by up of the center link

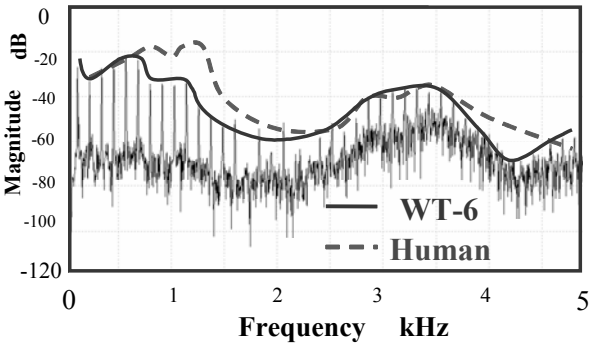
Fig. 8 WT-6's tongue shape



(a) Human's spectrum



(b) WT-5's spectrum



(c) WT-6's spectrum

Fig. 10 Spectrum of vowel /a/

These problems need to be solved in order to produce a human-like talking robot.

By the new mechanism the tongue could shape the peak or the gutter on the center of the vocal tract, as shown in Fig.8. It is necessary to reproduce human-like vocal tract.

V. Conclusion and Future Work

In order to clarify human speech mechanism, we constructed a new talking robot WT-6 that has vocal cords and a vocal tract based on human biomechanical structure. We improved the mechanism of the vocal cords to broaden the pitch control range; overcoming problem we encountered with WT-5's vocal cords. We constructed a 3D tongue shape in the vocal tract to solve the various problems of the 2D mechanism. The new mechanism improves the neutrality of the vowel /a/. However, the pitch control range is not wide enough to reproduce human speech, and the vocal tract could not reproduce the /i/ sound. We must solve these problems. Through the development of a talking robot, we aim to clarify the human speech mechanism, and we also aim to develop a robot which sounds like a human

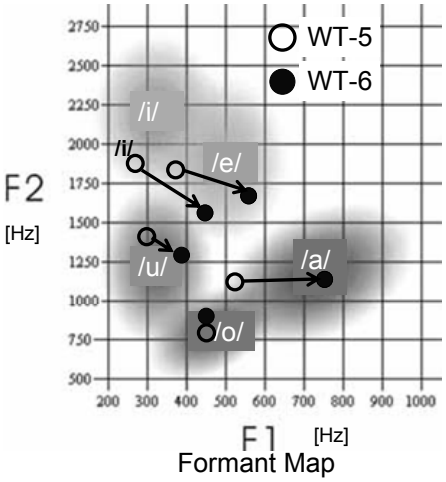


Fig. 11 Formant mapping of WT-5 and WT-6 compared with Japanese adult male [11]

Acknowledgment

The authors would like to thank to the following companies: Solid Works KK for provision of "SolidWorks" 3D CAD software and "COSMOS Works" FEM analysis software; Kuraray Co. for the provision and advice about Septon, and the members of ATR Bio Physical Imaging Project for advice about the biology of human speech mechanism.

References

[1] K. Fukui, K. Nishikawa, T. Kuwae, H. Takanobu, T. Mochida, M. Honda and A. Takanishi: "Development of an Human-like Talking Robot for Human Vocal Mimicry", *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp1449-1454, 2005

[2] J. L. Flanagan: *Speech Analysis Synthesis and Perception* 2nd ed., Springer, pp205-206, 1972

[3] N. Umeda, and R. Teranishi: "Phonemic Feature and Vocal Feature -Synthesis of Speech Sound, using an Acoustic Model of Vocal Tract-", *Journal of Acoustical Society Japan*, Vol.22, No.4, pp195-203, 1965

[4] A. Izawa, K. Hattori, Y. Matsuoka and S. Kawamura: "Speech Synthesis by Mechanical System Control", *Journal of Robotics Society of Japan*, pp. 273-278, 1993

[5] H. Sawada, M. Nakamura, T. Higashimoto: "Mechanical Voice System and Its Singing Performance", *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp1920-1925, 2004

[6] K. Fukui, K.Nishikawa, S. Ikeo, E. Shintaku, K. Takada, H. Takanobu, T. M. Honda and A. Takanishi: "Development of a Talking Robot with Vocal Cords and Lips Having Human-like Biological Structures", *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp272-278, 2005

[7] EPDM (Ethylene Propylene Diene Monomer) made by TOKYO RUBBER INDUSTRIAL CO., Ltd

[8] I. R. Titze: *Principles of Voice Production*, Prentice Hall, 1994

[9] http://www.septon.info/

[10] T. Chiba and M. Kajiyama: *The Vowel -Its Nature and Structure*, Tokyo-Kaiseikan, 1942

[11] R.A. Yamada, BLUEBACKS Scientific Inquiry: How to Improve English Speaking Skills, Kodansha, 2005(in Japanese)