The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Three Dimensional Tongue with Liquid Sealing Mechanism for Improving Resonance on an Anthropomorphic Talking Robot

Kotaro FUKUI, *Member, IEEE,* Yuma ISHIKAWA, Keisuke OHNO, Nana SAKAKIBARA, Masaaki HONDA and Atsuo TAKANISHI, *Member, IEEE*

Abstract— **We have developed a new, three dimensional vocal tract mechanical model for an anthropomorphic talking robot WT-7R (Waseda Talker No. 7 Refined), to improve the resonance of the vocal tract. The Waseda Talker robot series aims to reproduce the human speech mechanism with three-dimensional accuracy. The tongue of the previous model, WT-7 (Waseda Talker No. 7), was made of rigid links and covered with a thermoplastic rubber (Septon). This mechanism could deform the tongue shape and work as a part of the vocal tract, however, the cover thickness was not sufficient enough to prevent sound leakage, and did not have sufficient resonance. As a result, the produced sounds were unclear. To resolve this problem, the inner area of the tongue was filled with liquid. We experimented to select the filling liquid which is minimizes damage to the Septon cover and provides adequate resonance characteristics. The ethylene glycol was selected because it does little damage to the Septon and is relatively non-flammable. An oil seal and liquid gasket prevent leakage into the robot. WT-7 also has problems with its tongue link deformation range and open lip control range—these problems were also addressed. The improvements made the vowel production of WT-7R clearer than that of the previous robot, and the bandwidth of the formant peak in spectral analysis became sharper.**

I. Introduction

FROM 2005, we are developing anthropomorphic talking robots mimicking human biomechanical structure. The characteristic of these robots is three-dimensional tongue mechanism and fold based vocal cord mechanism, which is made of thermoplastic rubber, Septon [1]. The purpose is clarifying human speech mechanism. Much research has been done to clarify the human speech mechanism because of the importance of speech in communication. Speech production, however, includes aero-acoustic complexities, and the mechanism is still not understood. We use robotics to mimic the human vocal mechanism precisely and to confirm acoustic theory of voice production, which is difficult to experiment in human. The voice production mechanism of the anthropomorphic talking robot was designed exactly like that of a human voice production mechanism. Airflow from mechanical lungs vibrates the vocal cords, producing a source sound. The vocal tract resonance characteristics are controlled by articulating the tongue, the jaw, the lips, and the velum. Other voice synthesis machines have been developed, including those by Kempelen [2], Umeda [3], Kawamura [4], and Sawada [5].

This research is originally started 1999, and before WT-4 (Waseda Talker No. 4) in 2004, we focused on the reproduction of its functions, and produced the Japanese vowels (/a/, /i/, /u/, /e/, and /o/) and consonant sounds with a two-dimensional (2D) vocal cord and vocal tract. However, a 2D model is not adequate for the reproduction of the movement of human speech organs. We began development of a three-dimensional (3D) model of the speech organs and aimed to apply it as an intraoral mechanism simulator. The 3D model is focused on the biomechanical characteristics of each organ. For WT-5 (Waseda Talker No. 5), we developed a vocal cord model, which mimicked the human biomechanical structure, and reproduced human-like vibrations and source sounds. With WT-6 (Waseda Talker No. 6) [6], we started development of a 3D tongue model, which mimics human tongue deformation. The WT-6 tongue model consisted of a release mechanism, and it had accuracy problems. We developed WT-7 (Waseda Talker No. 7) in 2007[7] with a rigid link to improve accuracy. The sound produced by the WT-7 is analyzed from the first and second formants, which are important parameters for vowel recognition, and the sound has human-like parameters. However, the sounds it made were not perceived as human.

We looked for the reason and we made the hypothesis that the problem was a lack of clarity of the voice, and that the vocal tract resonance characteristics were the reason for the lack of clarity. This was confirmed by a liquid packing experiment, thus, we developed new articulation mechanism for WT-7R (Waseda Talker No. 7 Refined), as shown in Fig. 1. We also updated the tongue link and lip mechanisms to improve the produced sound. In this paper, we describe the development of WT-7R and its speech production.

II. Three Dimensional Tongue

In WT-5, the tongue was controlled by a seven control point

Manuscript received March 1, 2009. This work was supported in part by a Grant-in-Aid for Scientific Research (A), 16200015 from MEXT, Japan

K. Fukui, Y. Ishikawa, K. Ohno, N. Sakakibara, and A. Takanishi are with the Department of Modern Mechanical Engineering, School of Creative Science and Engineering, Waseda University, Tokyo, 162-8480, Japan (Corresponding author, Phone: +81-3-5369-7329; Fax: +81-3-5269-9061; e-mail: kotaro@toki.waseda.jp).

K. Fukui was 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.

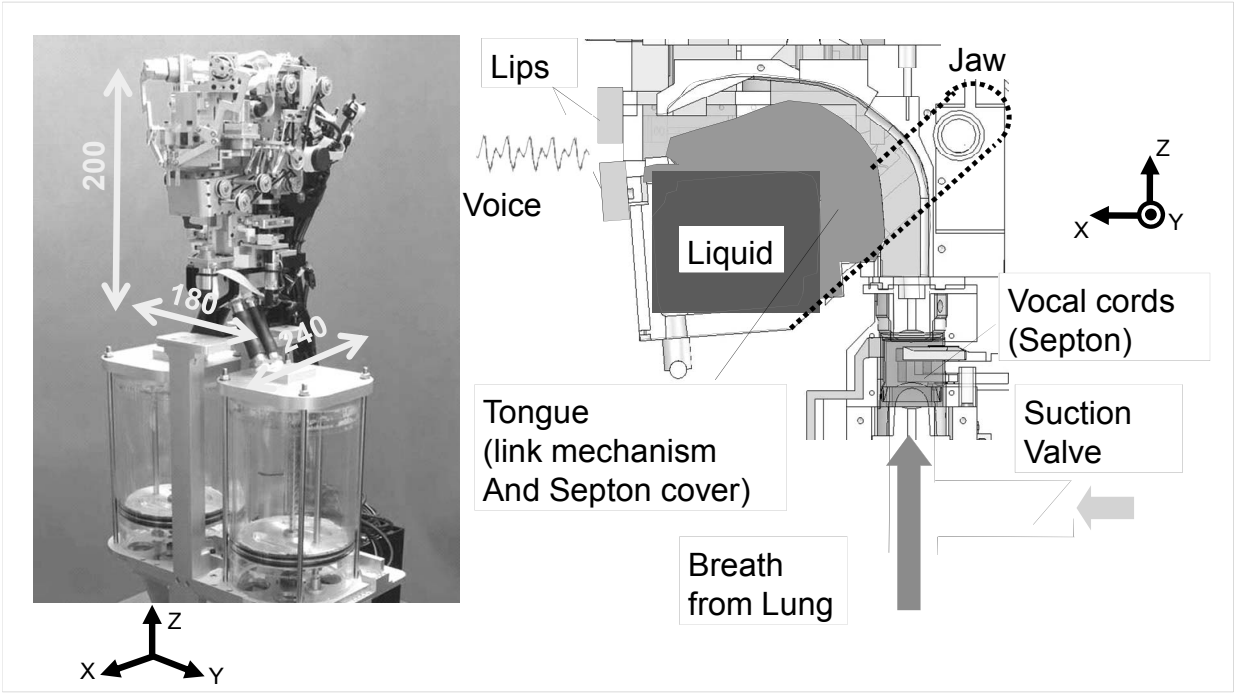


Fig. 1 Mechanical overview and control systems of the talking robot WT-7R

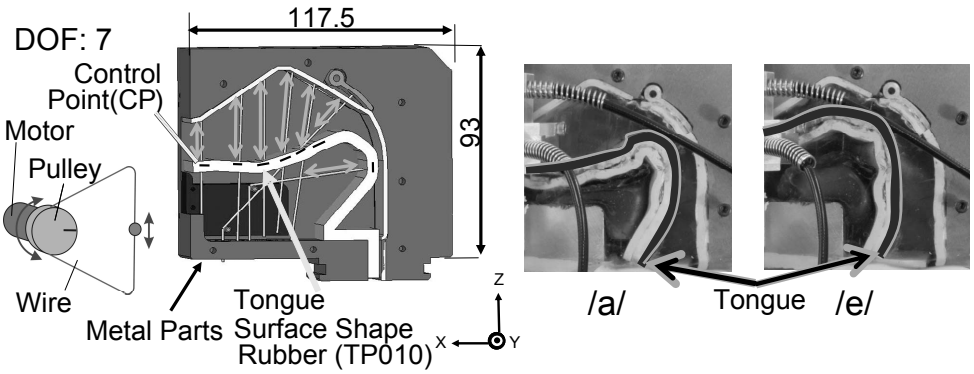


Fig. 2 Two-dimensional tongue model (WT-5)

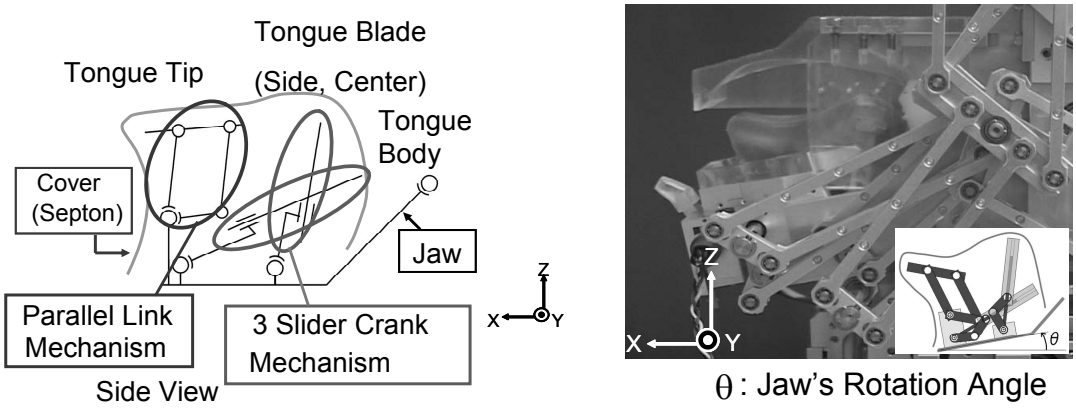


Fig. 3 The tongue mechanism of WT-7

(metal plate), driven by looped wires. The control point was set on a rubber plate in one line, and could reproduce 2D vocal tract area transitions, as shown in Fig. 2. With this mechanism, it could produce Japanese vowels and various

consonant sounds; however, it could only reproduce 2D shapes, and needed a more precise mechanism to reproduce the human biomechanical structure. Additionally, for voice production, a 3D mechanism was needed to produce consonant sounds accurately. Therefore, we developed WT-6, which had 3D tongue.

In developing WT-6, we aimed to reproduce the human tongue mechanism precisely. The human tongue has several closely associated muscles [8]. As extrinsic muscles, the styloglossus muscle primarily pulls the tongue body for upper, and the genioglossus muscle primarily pulls it for lower. As the intrinsic muscles, the superior and inferior longitudinal muscles control the shape of the tip of the tongue by antagonistic control [9][10]. However, the muscles of the tongue not only pull to one side, the volume of the muscles itself have important roll. This mechanism is complex, and has not been clearly clarified.

Several human vocal tract simulators have been proposed with two dimensional. And, analytical and computing advances have led to the development of a three-dimensional simulator based on human speech mechanisms [11].

We attempted to develop a more human-like tongue mechanism. Though the wire driven model is developed [12], very few actuators can move in a manner similar to human muscles, and artificial muscles cannot reproduce the control of the oral cavity. Therefore, we developed a 3D model using electric actuators, and analyzed the human tongue in an attempt to reproduce its movement using the actuators. Using magnetic resonance imaging (MRI) data of five Japanese vowels being pronounced, we determined that three parts are important in reproducing human tongue shapes: the tongue tip, the tongue blade, and the tongue body. In addition, each part has to provide deformation of 13 [mm], without the jaw mechanism. However, by adopting a jaw mechanism, this required movement can be reduced to 7 [mm], which is very useful with respect to design, because the mechanism must be packed into a small tongue. Therefore, we adopted a jaw mechanism, based on the human biomechanical structure, for the 3D tongue models.

The jaw mechanism of WT-6 varied the length from the jaw to the palate, but its looped wire mechanism was not adaptable. WT-6's 3D tongue is shaped by covering the release mechanism, which has five DOF (degrees of freedom), and setting it on the jaw with thermoplastic rubber Septon. Septon is a very elastic material that is easy to mold. The shape is designed as symmetric, because, in usual human vowel pronunciation, the tongue shape is symmetrical [13].The lip mechanism also used a release mechanism.

III. Development of WT-7

WT-6 could pronounce the vowel /a/, clearly and sounded similar to a human voice. However, the release mechanism could not be controlled accurately because of a bend in the inner wire, which made modeling difficult and caused low reproducibility. We developed the WT-7 anthropomorphic talking robot to solve this problem. WT-7 possesses 19 DOF, including the lips (5 DOF), the teeth (1 DOF), and the tongue and nasal cavity (7 DOF) as articulators; the soft palate (1 DOF), vocal cords (4 DOF), and lungs (1 DOF) act as vocal organs. The length of the vocal tract is 180 [mm], which is similar to that of an average adult male.

A. Tongue of WT-7

In WT-7, the release mechanism parts were replaced by

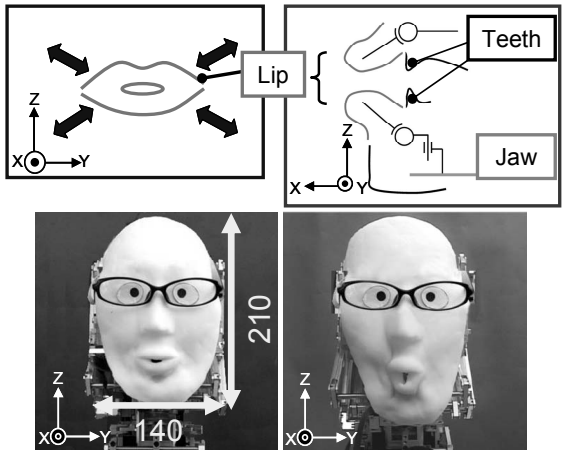
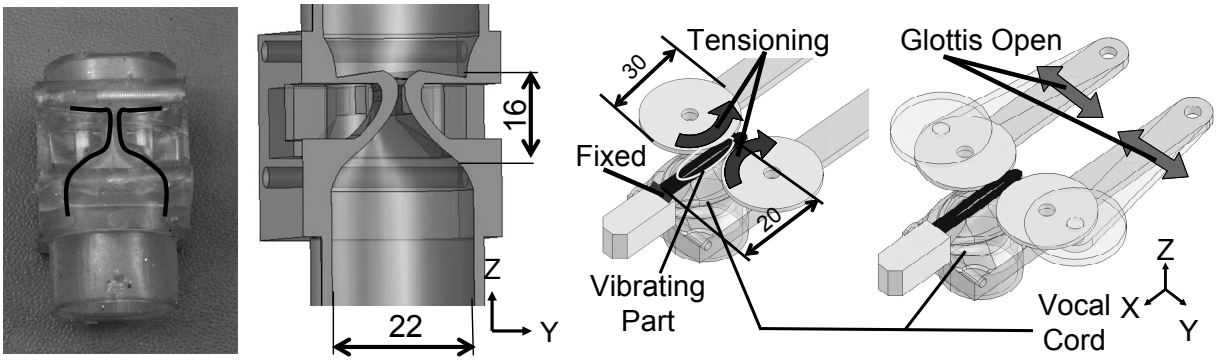


Fig. 4 WT-7's lip



(a)Vocal cords mimicking human structure (Same model as WT-5)

(b) Pitch control and voiced/unvoiced switch mechanisms (Developed in WT-7)

Fig. 5 WT-7's vocal cord mechanism

five DOF rigid link mechanisms. The mechanism for the tip of the tongue is a two DOF parallel link that controls the position and posture of the tip of the tongue. The tongue blade, a three DOF pair of slider crank links, controls the rotation of the tongue and the length of the sides and the center groove. The tongue body is reproduced by a set of two DOF slider crank links. Each slider crank mechanism controls the link length and the link gradient, and the blade link controls the length of the center groove, as shown in Fig. 3.

B. Lips of WT-7

The lips play two important roles in articulation. One is control of the vocal tract length by protrusion, as in the case of the /u/ vowel; the other function is the production of bilabial plosive consonants, such as /p/. We developed a mechanism to allow the lips to reproduce these functions. The lip mechanism of WT-7 had four DOF (in addition to the jaw mechanism)—upper lip protrusion, lower lip open, protrusion, and the width of the lip corners. The lip mechanism was constructed of a rigid link mechanism similar to the tongue. The control point was set inside the Septon lips. In combination with jaw movement, the lip mechanism produces a variety of lip shapes.

C. Vocal Cords of WT-7

Human vocal cords are vibrated by airflow from the lungs, generating the sound source of the voice. The vibration pattern differs in phase in the upper and lower parts of the vocal cord folds. This complex vibration is important for producing the various vibration patterns that create the human voice [7].

In WT-5, we developed a model of the vocal cord folds based on the human biomechanical structure, using Septon (Fig. 5(a)). The vibration pattern became more human-like, and the sound spectrum displayed human-like attenuation. However, the pitch control mechanism could only change the glottal length, and the pitch control range was only 15 [Hz]. In contrast, an adult male can control approximately 100 [Hz] pitch frequency range in normal speech.

To resolve this problem, in WT-7 we developed a new vocal cord control mechanism as shown in Fig. 5(b). This mechanism consisted of a pair of discs that are attached directly to the vibrating points. Using this mechanism, the vocal cords of WT-7 can produce pitches of 129–220 [Hz] and produced a human-like sound source spectrum. A separate mechanism controls the opening of the glottis.

IV. Development of Anthropomorphic Talking Robot WT-7R for Resonance Characteristics

WT-7 could produce vowels having human-like formant parameters. However, the sound is not heard as a clear, human voice. From experiments, we hypothesized that this was caused by resonance characteristics that differed from those of a human, and the rigid vocal tract model. The difference of the tongue shape, caused by the limitations of link movement, is also part of the problem. We developed a new vocal tract mechanism for WT-7R to verify these hypotheses and improve the clarity of produced vowels. WT-7R's articulators consist of a seven DOF tongue, a one DOF jaw, four DOF lips, a one DOF velum, and a nasal cavity. The length of the vocal tract is 180 [mm]—the same as the previous robot. In this part, we describe the articulatory mechanism of WT-7R.

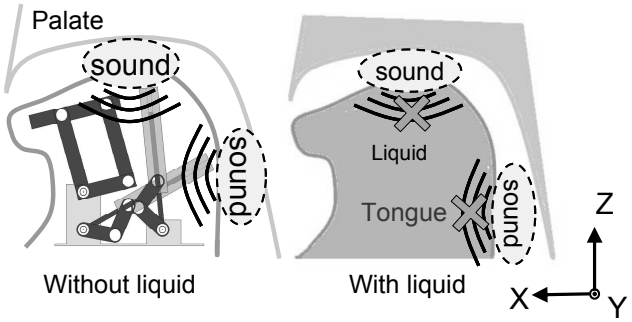
A. Improvement of Resonance Characteristics

i) Resonance Experiment

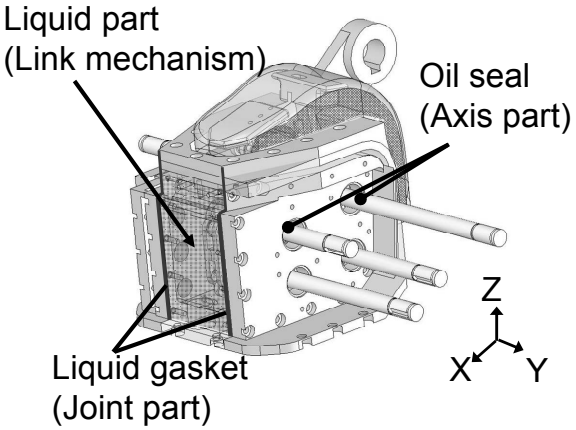
We experimented with many control parameters, to clarify the reason of the unclear sound. The pronunciation of vocal tract model with rigid resin was better than WT-7 and we hypothesized that the problems were caused by the resonance characteristic of the vocal tract wall, especially the tongue.

In the rigid vocal tract model sufficient wall thickness could prevent sound leakage, as in a human, which is needed to produce a clear voice. In 3D talking robot, however, vocal tract wall parts, such as the tongue, must deform into various shapes, so the part must be made of thin rubber.

We verified this hypothesis using liquid. We constructed a vocal tract model of thin rubber, and set it on a tank of the liquid. By comparing the produced sound, with and without liquid, we could calculate the liquid's effects. If the liquid is not viscous, the clarity of sound was improved by using it.



(a) Liquid packing in the tongue



(b) Sealing mechanism.

Fig. 6 Liquid mechanism

ii) Selection of Liquid

For the development of this robot, we required a highly resonant liquid; however, since the liquid is covered by Septon, damage to the rubber was less likely. In our experiments, the tongue was made with Septon, which includes paraffin oil, for elongation, the quality of the paraffin is extraordinary; however, the structure is unstable. In such a situation, if we include a liquid that can combine with the paraffin, it destroys the structure of the Septon. We experimented with many types of liquids. In the experiment, small piece of Septon sheet was put into the liquid with tensioned by clip and measured the time to break the sheet. The result is shown in Table 2 and ethanol, glycerol and ethylene glycol have very low invasiveness to Septon. And, the tongue mechanism is set on part of the robot; there are many cables near the packing area, so flammable liquid should be avoided. We selected ethylene glycol, because of its higher flash temperature than ethanol and it is enough fluid to prevent sound leakage.

Table 1 Effect to Septon of Liquids

Liquids	Time to Break
Acetone	Immediately
DMF	5min
1.3-Bis-(trifluoromethyl)benzene	7min
1.4-Bis-(trifluoromethyl)benzene	10min
DMSO	10min
Paraffin	90min
Oil	120min
Tap Water	24hour
Ethanol	Not Break
Glycerol	Not Break
Ethylene Glycol	Not Break

Table.2 Bandwidth of vocal cord models

Bandwidth	1st Formant [Hz]	2nd Formant [Hz]
Liquid	140	200
No liquid	150	270

iii) Sealing Mechanism

We used a liquid sealing mechanism on the packing area, as shown in Fig. 6. The shaft of the tongue mechanism was covered with an oil seal and the connection between the metal parts was sealed with a liquid gasket. The design of the jaw mechanism had an additional liquid pool to prevent leakage to the outer part of the robot.

The inside area of the tongue is connected to the tank set on the higher position. It keeps the pressure of the liquid same level in spite of the cavity change caused by the tongue deformation.

We compared the sound when packed with liquid to the sound with no liquid. In the experiment, the shape of the articulation mechanism has the same parameters—the pronunciation of the Japanese /o/. The spectrum of the experiment is shown in Fig. 7—the first and second formants approximated those of a human. An acoustic parameter of vowel clarity is bandwidth, which is calculated by the frequency range of 5[dB] below the top of the spectrum formant. The bandwidth comparison data is shown in Table. 2. This experiment showed that using the liquid sharpened the formant, and the bandwidth was 50 [Hz] narrower than without liquid.

B. Improvement of Tongue Link Mechanism

The WT-7's link mechanism also had problems. At the tongue tip, the link was a parallel link mechanism—vertical deformation was reproduced mainly by the jaw mechanism. This limited its ability to reproduce human deformation, therefore, in WT-7R we added one DOF on each side of the tip parallel link. In the tongue blade and tongue body the linear deformation range was not sufficient to reproduce tongue shape precisely, so we improved the range by integrating a more complex link mechanism, as shown in Fig. 8.

The link design of the new tongue mechanism had three DOF in the tongue tip, two DOF in the tongue blade and two

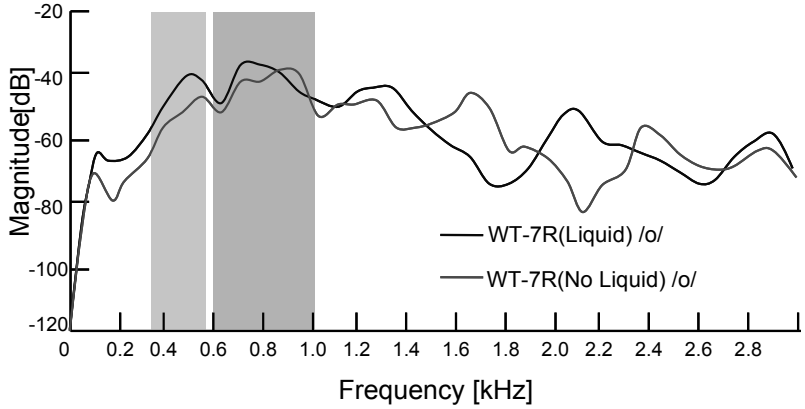
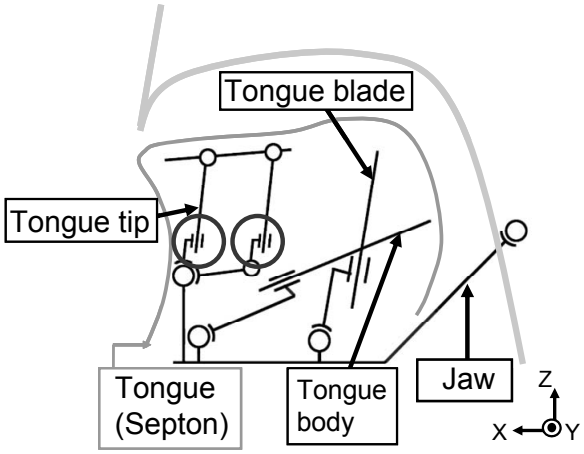
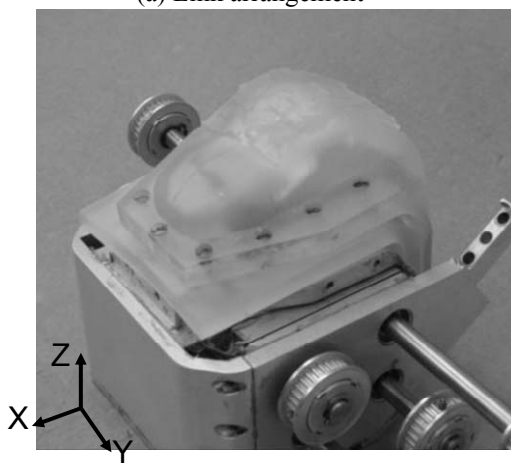


Fig. 7 Comparison of the spectra for two altered frequencies of vowel /o/, with liquid and without liquid

DOF in the tongue body. The tongue tip was a parallel link and controlled the front and back length, and its rotation. The blade and body mechanism was a slider crank mechanism, which could control the length and rotation.



(a) Link arrangement



(b) Isometric view

Fig. 8 WT-7R's Tongue mechanism

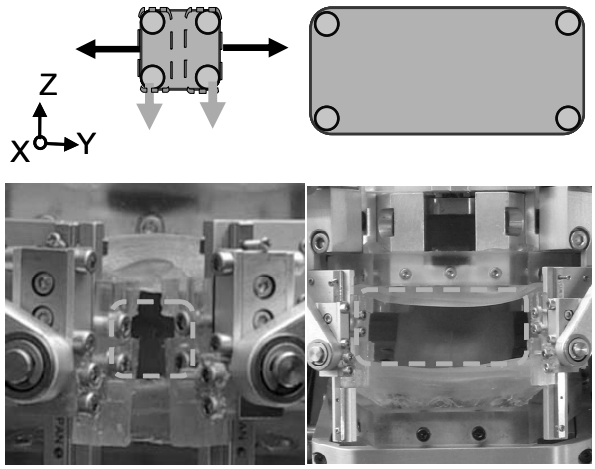
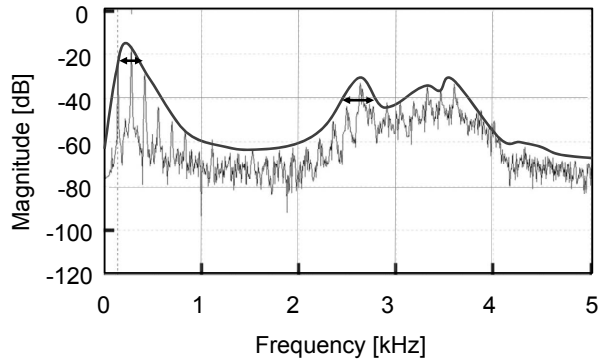


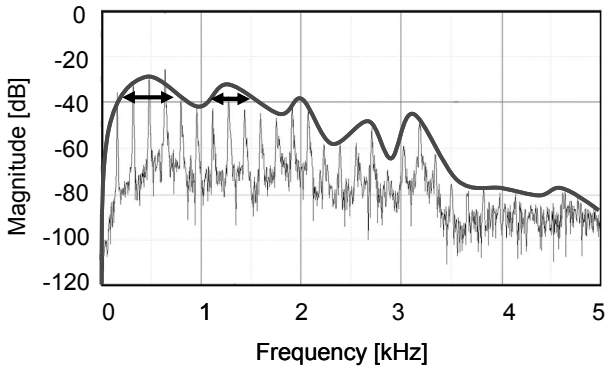
Fig. 9 WT-7R's lip

C. *Improvement of Lip Mechanism*

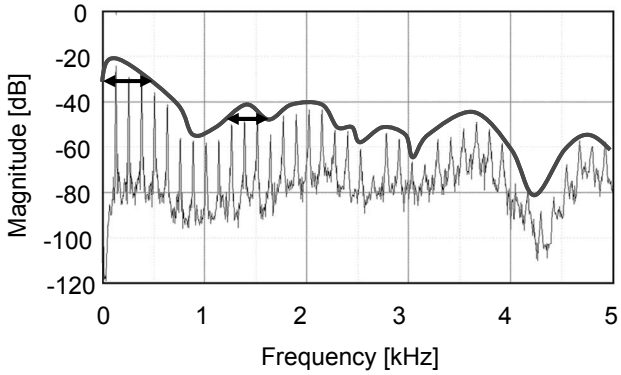
The WT-7's lip mechanism had a deformation range problem and caused WT-7's lack of clarity. The mouth corner mechanism was controlled by a link attached to the inside of the human-like lips; the Septon part between the control point and the edge of the open area did not stretch. As a result, the open deformation area was much smaller than the movable range of the control point. In the new lip mechanism, we formed the lip as /o/, which requires the smallest open area and largest protrusion, and the control point was placed inside the open area. The control point was set as a rectangle and with two DOF controlling the horizontal length by moving each control point pair. The upper two control points were controlled by this vertical movement alone, and the lower two control points were controlled vertically by an additional two DOF, as shown in Fig. 9. Though the deformation range was



(a) Human



(b) WT-7



(c) WT-7R

Fig. 10 Spectrum of vowel /a/

limited by the Septon, the open lip control range increased to 55–1200 [mm^2]. This was enough to reproduce the area for vowel /a/ (840 [mm^2]) and vowel /o/ (140 [mm^2]). The lips were made thicker to prevent sound leakage.

D. Vowel Production with New Mechanism

We experimented with Japanese vowel production for total validation. The variations in languages are not large in vowels, and hardware requirement for vowels are not so differed. A comparison of the spectrums of WT-7R, WT-7, and a human pronouncing the Japanese vowel /a/ is shown in Fig. 10. In the experiment, the mechanical vocal cords were vibrated by the mechanical lungs, and the vocal tract shape was controlled to articulate the source sound and produce vowels. The spectrum shows that the peak of WT-7R sharpened and the formant frequency become closer to that of the human. In formant mapping of five vowels, comparing WT-7R, WT-7 and a human, the WT-7R came near to the human. The production of /i/ (hard vowel) is relatively un clear. It caused by the source sound of vocal cords, and we need further improvement in the vocal cord.

V. Conclusion And Future Work

We have constructed a 3D tongue mechanism to reproduce various human-like voices and to better understand the human speech mechanism. Although a previous model (WT-7) could reproduce human-like tongue shapes, the produced vowels were not clear. We developed the articulation mechanism of WT-7R to improve the clarity of the vowels produced. During this development work, we hypothesized that the lack of clarity was caused by sound leakage in the tongue mechanism, and confirmed this by experiments with a thin rubber vocal tract model sitting on a tank of the liquid. From this result, we developed a new tongue mechanism, which we could pack with liquid. We selected ethylene glycol as the packing liquid because it does little damage to the Septon and is relatively non-flammable. We also improved the movement range of the tongue link mechanism and the open lip mechanism.

With these mechanisms, the sound produced by the new robot was clearer, and nearer to that of a human. This was confirmed by the bandwidth of the spectrum.

This robot showed improved vowel production. However, we still have not reproduced various consonant sounds with the 3D tongue. We also should investigate production of other language, because some unique consonant phonemes, such as trilled “r”, are used for certain language. For this reproduction, the tip of the tongue should be more elastic. Many other improvements, such as asymmetric tongue movement, are needed to realize the production of various consonant precisely. In the future, we want to develop the muscle based 3D tongue model by adopting elastic actuators. Our ultimate goal is to clarify the human speech mechanism, and we would like to examine the various speech control models using the 3D talking robot. Therefore, we intend to develop more human-like robot hardware that can reproduce the human biomechanical structure.

Acknowledgment

The authors would like to thank the following companies: Solid Works KK for providing CAD and FEM software; Kuraray Co. for the providing and advising us on Septon; Prof. Shimizu at Department of Applied Chemistry, Waseda University for advice of the liquid selection; and, the members of the ATR BioPhysical Imaging Project for advice on the biology of the human speech mechanism.

References

[1] Homepage of Septon(Kuraray corp.) http://www.septon.info/
[2] J. L. Flanagan: Speech Analysis Synthesis and Perception 2nd ed., Springer, pp. 205-206, 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, pp. 195-203, 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. pp. 1920-1925, 2004
[6] K. Fukui, Y. Ishikawa, T. Sawa, E. Shintaku, M. Honda and A. Takanishi: New Anthropomorphic Talking Robot having a Three-dimensional Articulation Mechanism and Improved Pitch Range, 2007 IEEE International Conference on Robots and Automations, pp. 2922-2927, 2007.
[7] K. Fukui, E. Shintaku, A. Shimomura, N. Sakakibara, Y. Ishikawa, M. Honda and A. Takanishi: Control Methods Based on Neural Network Forward and Inverse Models for a Biomechanical Structured Vocal Cord Model on an Anthropomorphic Talking Robot, 2008IEEE International Conference on Robotics and Automation, pp. 3648-3653, 2008
[8] I. R. Titze: Principles of Voice Production, Prentice Hall, 1994
[9] W. R. Zemlin: “Speech and Hearing Science –Anatomy and Physiology”, 4th Edition, Allyn and Bacon, pp.253-258, 1998
[10] H. Takemoto and K. Honda: “Measurement of temporal changes in vocal tract area function from 3D cine-MRI data”, Journal of Acoustic Society of America, Vol. 119, No. 2, pp. 1037-1049, 2006
[11] F. Vogt, J.E. Lloyd, S. Buchaillard, P. Perrier, M. Chabanas, Y. Payan, and S.S. Fels: “An Efficient Biomechanical Tongue Model for Speech Research”, Proceedings of ISSP 06. pp.51-58. 2006
[12] R. Hofe; R. K. Moore:v “Towards an investigation of speech energetics using ‘AnTon’: an animatronic model of a human tongue and vocal tract”, Connection Science, Vol. 20, No. 4, pp. 319–336, 2008
[13] S. Hiki and S. Imaizumi: “Observation of Symmetry of Tongue Movement by Use of Dynamic Palatography”, Annual Bulletin of Research Institute of Logopedics and Phoniatrics in University of Tokyo,No.8, pp. 69-74, 1974