Development of a Talking Robot with Vocal Cords and Lips Having Human-like Biological Structures

Kotaro Fukui[*1], Kazufumi Nishikawa[*1, *6], Shunsuke Ikeo[*1], Eiji Shintaku [*1],
Kentaro Takada[*1], Hideaki Takanobu[*2, *4], Masaaki Honda[*5] and Atsuo Takanishi[*1, *2, *3]

[*1] Department of Mechanical Engineering, [*2] Humanoid Robotics Institute,
[*3] Advanced Research Institute for Science and Engineering,
Waseda University; 3-4-1 Ookubo, Shinjuku-ku, Tokyo 169-8555, Japan
[*4] Department of Mechanical Systems Engineering, Kogakuin University, Tokyo, Japan
[*5] School of Sport Sciences, Waseda University, Saitama, Japan
[*6] Research Fellow of the Japan Society for the Promotion of Science
Email: kotaro@toki.waseda.jp

Abstract - **We have developed a new talking robot, WT-5 (Waseda Talker No. 5), having novel vocal cords and lips based on human biological structures. The vocal cords are made from the thermoplastic rubber "Septon", available from Kuraray Co. Ltd, which has a similar elasticity to human tissue. The vocal cord model was constructed with a similar structure to the biological structure of the human vocal cords. The vocal cords vibrated like those of a human. The new lips were able to attain a large deformation in a similar manner to human lips, without the leakage of air. This allowed clear pronunciation of vowels. With these new mechanisms, the robot could reproduce human speech in a more biological manner and could produce voices closer to those of a human.**

1. INTRODUCTION

Given the importance of speech in human communication, much research has been carried out to clarify the mechanisms of human speech. However, because they involve the complexity of aero-acoustics and the movement of speech organs, it is difficult to simulate human speech mechanisms using computer models. Since 1998, we have been developing mechanical models of the human speech organs to clarify the mechanisms of speech. We developed talking robots to produce vowels and consonant sounds in a similar way to a human by considering changes in the area of the vocal tract [1].

Kempelen W.V. [2], Umeda [3], Kawamura [4], Sawada [5], and the other researchers have developed other voice

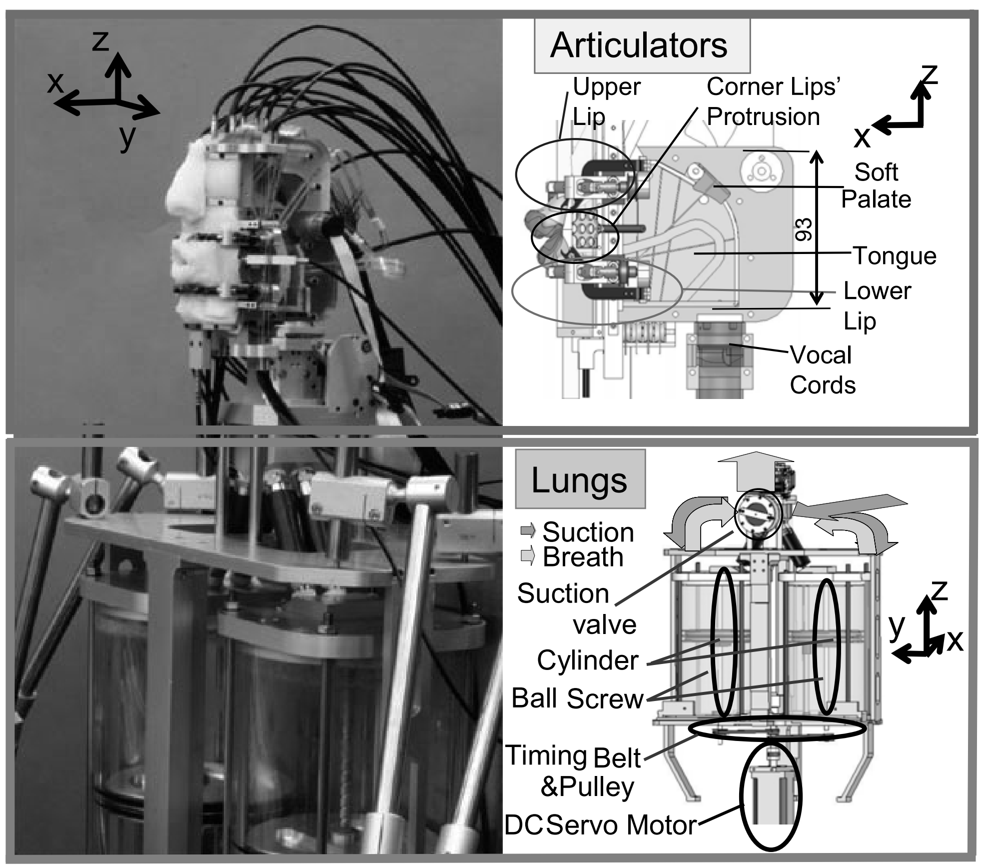
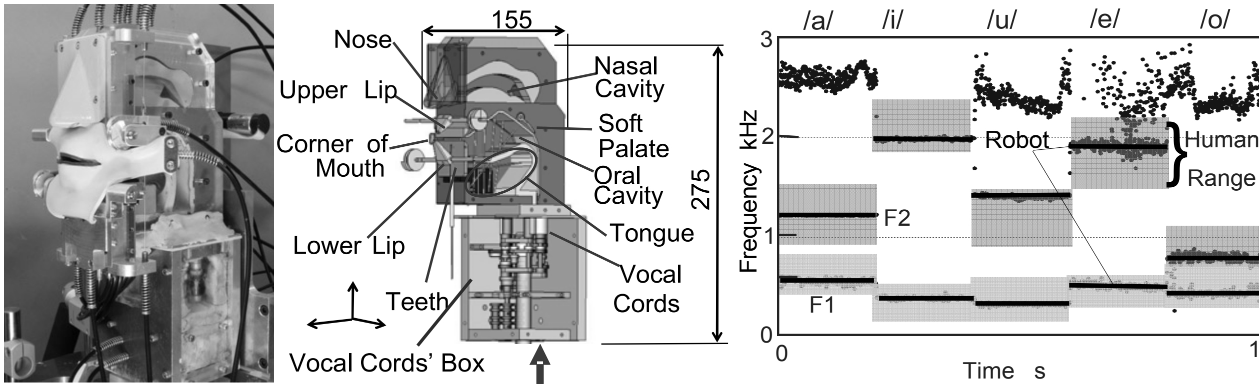


Fig. 1 New anthropomorphic talking robot WT-5



(a)Mechanism (b) Vowels

Fig. 2 Previous robot WT-4

synthesis machines.

In 2003, WT-3 (Waseda Talker No. 3) was developed with a 2D tongue mechanism, which was able to obtain a transition in the vocal tract area in the same manner as a human to produce the Japanese vowels (/a/, /i/, /u/, /e/, /o/) and consonant sounds. In 2004, we developed a talking robot WT-4 (Waseda Talker No.4), which improved on WT-3 and which had a more human-like connection between the vocal cords and the vocal tract. WT-4 could produce more human-like voices, and included an auditory feedback system to optimize the acoustic parameters of the voice [6].

However, this approach was found to be inadequate for clarifying the mechanisms of human speech. It was considered that approaches that more closely reproduced the biological mechanisms of the vocal organs were needed to further investigate human speech. Especially the measurement of the human vocal cords is very hard, and the mechanical model is very effective to clarify the mechanism.

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 recognition of vowels, were in the human range, however the robot did not sound like a human. In view of this, a new approach must be taken if the human voice is to be genuinely reproduced.

In this paper we describe the development a new anthropomorphic talking robot, WT-5(Waseda Talker No. 5), which is designed to reproduce human speech mechanisms in a more biological manner—the design is shown in Fig. 1.

2. The Previous Talking Robot —WT-4

We developed the anthropomorphic talking robot WT-4 shown in Fig. 2(a) in 2004, but this robot had a number of problems. WT-4 had 1-DOF lungs, 4-DOF vocal cords and articulators (7-DOF tongue, 5-DOF lips, 1-DOF teeth, nasal cavity and 1-DOF soft palate); the total DOF being 19. The average length of the vocal tract was approximately 175 [mm], the same as that of an adult male. WT-4 could produce 50 Japanese vowels and consonant sounds. The F1 and F2 of WT-4's vowels analyzed by Praat [7] are shown in Fig. 2(b) and are within the human range [8]. However, this robot did not prove sufficient for the purpose of clarifying the mechanism of human speech through the modeling of the human vocal organs. Human speech mechanisms comprise very complex movements involving both muscles and cartilage, and the modeling of these movement mechanisms was untouched in previous robots. In particular vocal cords and lips needed to more accurately reproduce the biological form. Additionally, there was still a need for improvement in the robot's voice.

A. *WT-4's Vocal Cords*

WT-4's vocal cords, as shown in Fig. 3(a), consisted of two thin super soft rubber (EPDM) [9], having a thickness of 2.0[mm]. The vocal cords had a 4-DOF mechanism with two flexible arms. An independent mechanism rolled around the rubber on each side equally to change pitch. The flexible arms were able to change the glottal length from 30[mm] to 35[mm], which was near to the width of the robot's vocal tract, to decrease the sound leakage. This mechanism could switch between voiced and voiceless sounds, keeping the tension in the vocal cords by opening the glottis with the arm. The sound source spectrum of the vocal cords is shown in Fig. 3(b), and is stable from the lowest to the highest frequencies. However the sound source spectrum was not the same as data estimated data for humans, which is attenuated at higher frequency.

This difference was caused by a difference in vibration. Human vocal cords have thin folds and the lower and upper edges of the vocal cords can vibrate in different phases. In contrast, the vibration of WT-4's vocal cords was very simple.

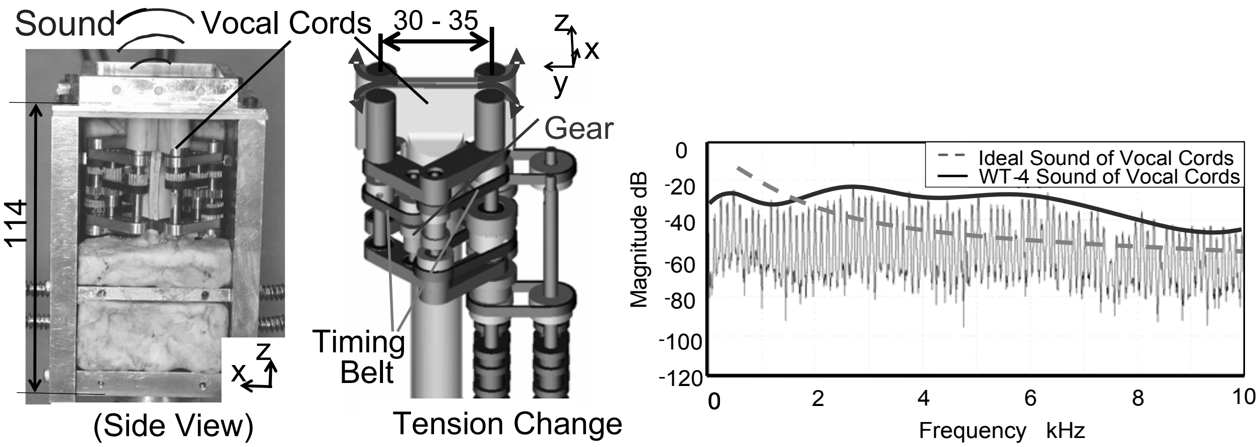
We needed to develop vocal cords which approximated the biological vocal cords of a human more in order to obtain a robot sound source closer to that of a human.

B. *Lips*

The lips play a very important role in human speech. Positioned at the end of the vocal tract, they can not only change the area, but also the length of the vocal tract.

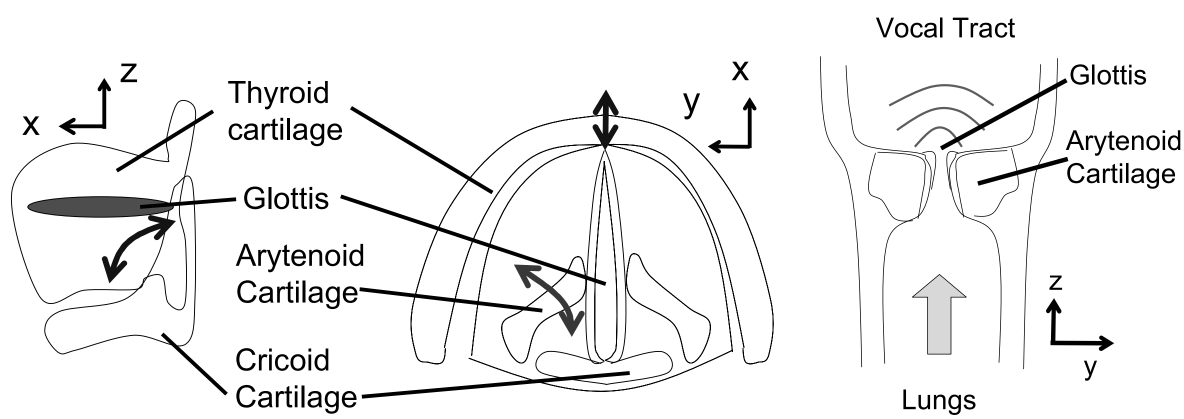
The mechanism of WT-4's lips had 5-DOF, with opening and protrusion mechanism on each lip and mechanism to change the width of the opening at the corners of the lips. The lip mechanisms were covered with the same super soft rubber as the vocal cords using quick drying glue, as shown in Fig. 6

However, this mechanism did not perform as expected. Although the change in the area of the opening seemed



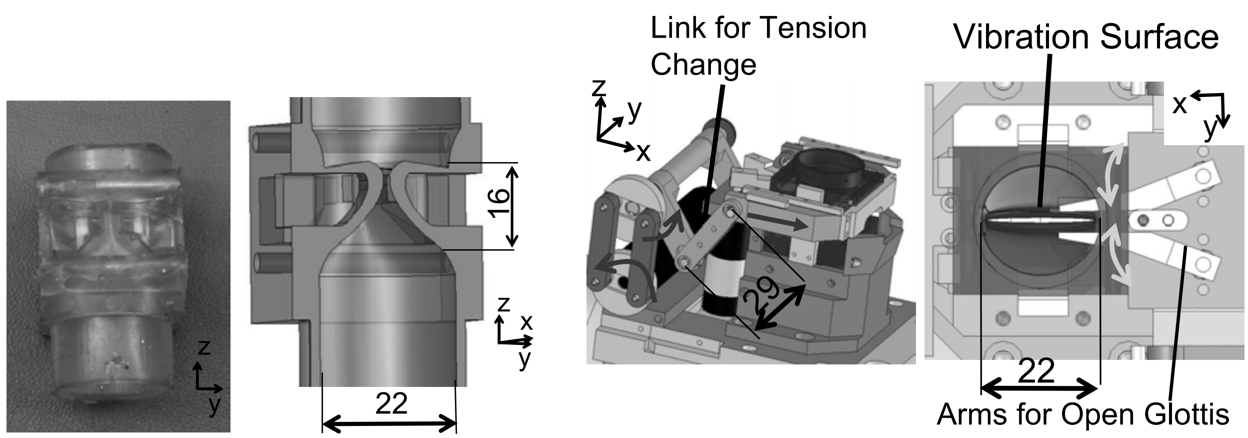
(a) Vocal cord mechanism (b) Vocal cords sound source

Fig. 3 WT-4's vocal cords



(a) Cartilages (b) Horizontal (C) Coronal

Fig. 4 Human vocal cords mechanical structure



(a)Vocal cords made by Septon (b) Coronal (c) Mechanisms

Fig. 5 New vocal cord mechanism of WT-5

sufficient, the effect of this on the formant was limited. This was due to insufficient sealing at the attachment points of the lips, such as between the lips and the palate.

A mechanism was required which allows the area and protrusion of the lips to be sufficiently changed without air leakage.

3. FORMING OF SOFT MATERIAL BASED ON HUMAN DATA

A. High Performance Thermoplastic Rubbers

"Septon", a thermoplastic rubber from Kuraray Co. Ltd [10], was adopted for the new vocal cords and lips. The material can be formed freely and has a greater flexibility than the other elastic materials investigated. The procedure for forming Septon is described below.

First, pellets of Septon compounds are added to paraffin oil (liquid). The hardness of Septon is determined by the quantity of paraffin oil. Next, the mixture is stirred at 200[°C]. After the pellets have sufficiently dissolved in the oil, air is removed from the liquid by means of a vacuum. The liquid is then poured into a mold. Once the mixture has cooled and set, it is removed from the mold.

B. Vocal Cords

It was considered that the robot's vocal cords needed to be closer to the biological vocal cords of humans. The mechanism of the human vocal cords is very complex as shown in Fig. 4. Human vocal cords consist of 3 layers. The thyroid cartilage can change the pitch of the voice by changing the length of the glottis. The arytenoid cartilage inside the vocal cords can change the opening of the glottis, allowing switching between voiced and voiceless sounds [11] [12].

The mechanism of the new vocal cords was based on that of the human vocal cords. The vocal cords had a 3-DOF mechanism: 1-DOF to change the tension and 2-DOF to open the glottis. The pitch was changed by the tension of the vocal cords, with the tension depending on the length of the vocal cords adjusted using the 1-DOF link mechanism. The 2-DOF arm mechanism was used to mimic the abduction and adduction of the human arytenoid cartilage. Switching between voiced and voiceless sounds was achieved by means of arms set on the fold. When the arms adducted, the glottis was closed and could vibrated with the passing air.

Robot vocal cords were produced based on human biological data, as shown in Fig. 5. Many types of mold having small variations in the shape parameters of the fold were produced in order to find a suitable sound source.

C. Mold of Lips

In the forming of the new lips, a mold was prepared from human lips. The mold was made using medical putty, with the

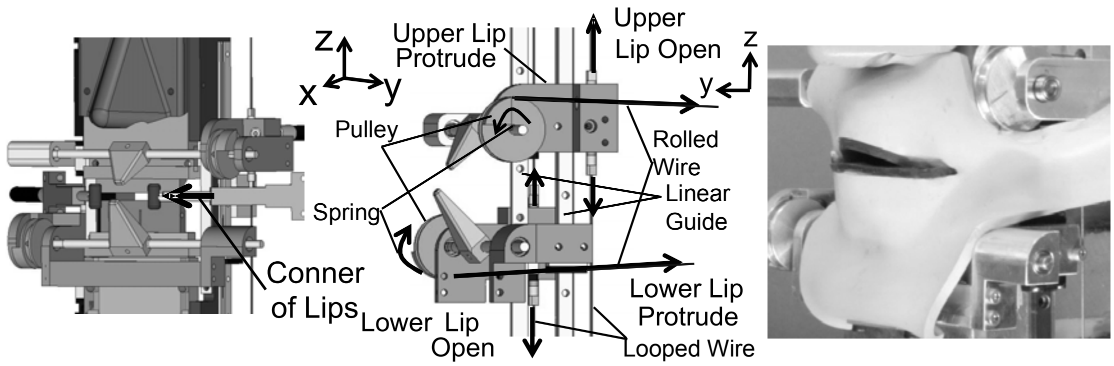


Fig. 6 WT-4's lip mechanism

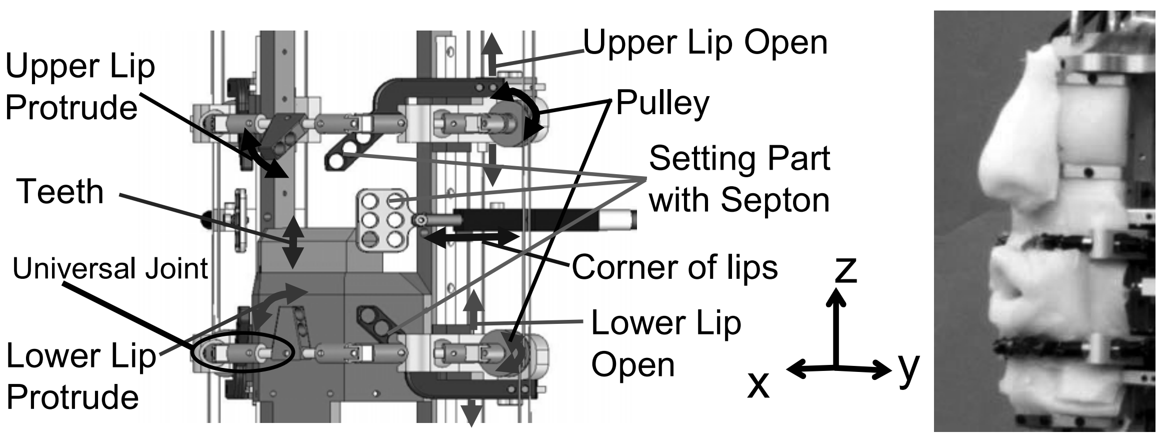


Fig. 7 WT-5's new lips mechanism

shape formed to the mouth shape of an adult male pronouncing the Japanese vowel /u/. The lips of the robot must be able to protrude to change the vocal tract length. Since elastic materials such as Septon are harder to compress than to stretch, the mold of the lips should be small with a protruded mouth shape.

The lip mechanism could control the opening area and protrusion of the lips with 5-DOF: the looped wires controlled the opening of the lips and the pulley with the rolled wires on each side controlled the protrusion of the lips. The corners of the mouth were changed using another 1-DOF linear mechanism as shown in Fig. 7. These lips could change the opening and protrusion of the lips without air leakage. The lips were thus able to function as part of the vocal tract.

4. EXPERIMENT

A. Evaluation of the New Vocal Cords

We observed the vibration of the new vocal cords using a high-speed-camera. The result was very similar to that of a human, as shown in Fig. 8. The vocal cords were vibrated by air flow from the lungs of the talking robot. The vibration took the form of a wave shape in the vocal cords, as occurs in the vocal cords of a human. The lower and upper edges of the vocal cords could vibrate in different phases.

In addition, the sound spectrum of the new vocal cords was attenuated at higher frequencies, which corresponded to the spectrum of estimated human vocal cords, as shown in Fig. 9.

B. Evaluation of the New Lips

The effect of the new lips was investigated by replacing WT-4's lips with the new lips. The sound spectra of the Japanese vowel /i/ produced by a human, WT-4 and WT-4 with new lips are shown in Fig. 10.

This demonstrates that the new lips produced a voice nearer to that of a human than the voice of WT-4 with the previous lips. As part of the vocal tract, the new lips were effective because of the decrease in air leakage.

C. Speech Production of WT-5

We analyzed WT-5's vowels with sound analysis software Praat [7]. WT-5 was built using the new vocal cords and lips, with its other parts being the same as in WT-4. The parameters of the WT-5 voices were optimized manually using FFT analysis. The first and the second formants of WT-5's Japanese vowels are shown in Fig. 11. WT-5's vowels had formant similar to those of a human. WT-5 also could produce Japanese consonant sounds as well as the vowels.

4. CONCLUSION AND FUTURE WORK

We have developed a new anthropomorphic talking robot, WT-5. The robot has vocal cords and lips based on the biological structures of the human vocal cords and lips. A method of reproducing human vocal cords and lips using

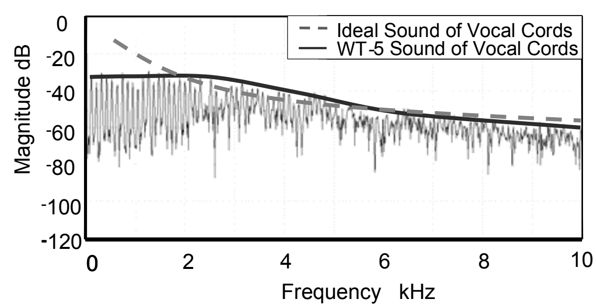


Fig. 9 WT-5's vocal cords characteristics in frequency domain

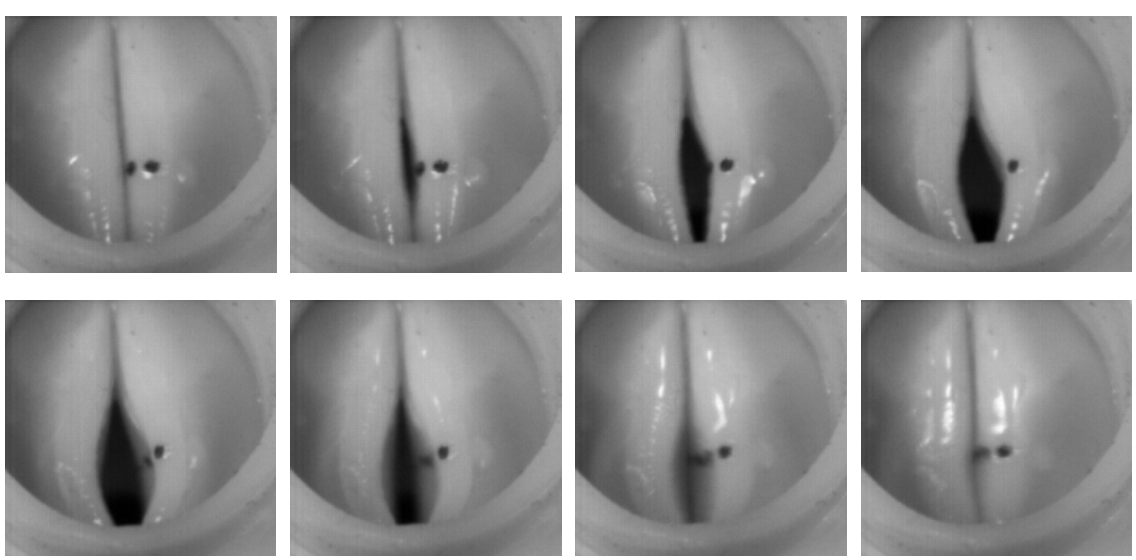


Fig. 8 WT-5's Vocal cords vibration (250[fps])

elastic materials was developed using the rubber Septon. The new vocal cords were made using a mold based on data obtained from human vocal cords, and vibrated in a manner similar to human vocal cords. By shaping the lips with Septon, the voices were more clearly pronounced than in previous robots.

In future work, we intend to develop other vocal organs, besides vocal cords and lips, to have more human-like biological forms. As the speech control of the robot becomes closer to those of a human it will be possible to simulate human speech more comprehensively. Through these developments, we aim to clarify the mechanisms of human speech, as well as to develop instruments that will help people with speech difficulties to communicate.

ACKNOWLEDGMENT

This research is supported by a Grant-in-aid for Scientific Research (A), 16200015, 2004 by MEXT, Japan. We would like to express our thanks to the following people: Dr. K. Honda, Dr. E. Murano and Dr. T. Kitamura of Biophysical Imaging Lab at ATR for the advice on the biology of vocal cords, Kuraray Medical Inc. for the supply of Septon and advice on the molding of Septon, SolidWorks Japan K.K. for the provision of 3DCAD software, and Chukoh Chemical Industries for the provision of Teflon coated wire.

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. Nishikawa, T. Kuwae, H. Takanobu, T. Mochida, M. Honda and A. Takanishi: "Mimicry of Human Speech Sounds using an Anthropomorphic Talking Robot by Auditory Feedback", *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp272-278, 2004

[7] http://www.praat.org/

[8] M. Mohri: *Otorhinolaryngology for Dentistry*, Ishiyaku Publishers, Inc., pp.95-99, 1996

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

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

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

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

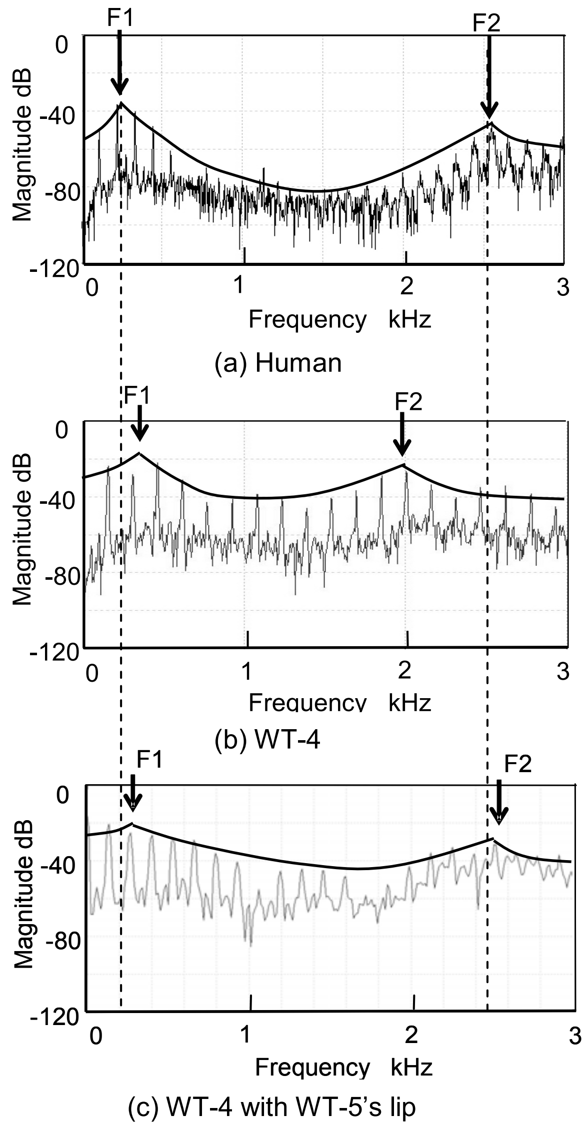


Fig. 10 Spectrums of vowel /i/

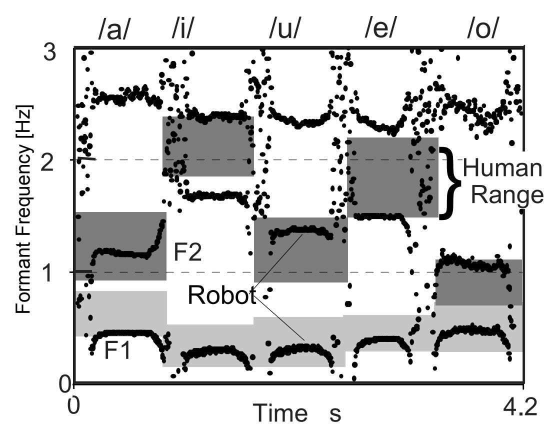


Fig. 11 Formants of WT-5's five vowels