Design and Preliminary Experiments of an Articulation Mechanism for an Infant-like Vocal Robot "Lingua" towards natural conversation with people@home

Nobutsuna ENDO, Tomohiro KOJIMA, Yuki SASAMOTO, Hisashi ISHIHARA, Takato HORII, Minoru ASADA

Dep. Adaptive Machine Systems, Graduate School of Engineering, Osaka University

endo@ams.eng.osaka-u.ac.jp

Abstract

Many teams @home utilize conventional speech recognition and generation systems of which sound is often heard unnatural due to a different sound generation from humans'. Further, conversation is based on existing AI strategy. Towards more natural conversation with humans, two big issues are considered; one is a sound generation mechanism like a human and the other is how to learn to vocalize like a human. As the first step towards these challenges, this article focuses on the mechanical design for infant-like vocal robot. Several observational studies suggest that the infant vocalization towards language acquisition develops through interactions with caregivers. However, what kind of underlying mechanisms works and how caregiver's behavior affects on this process have not been made clear since it is very difficult to control the infant vocalization. In order to attack this issue, we built an infant-like vocal robot "Lingua" as a controllable vocal platform. Lingua has two features; infant-like voice and high articulation capability. The shape of its vocal tract resembles that of a 6-month-old infant based on the anatomical data, and this may contribute to the former. 7-DOFs for articulation in the tongue is realized by sophisticated design of linkage mechanisms inside miniaturized vocal tract for high capability of its articulation. Preliminary experiments showed that the robot succeeded in vocalizing almost the same fundamental frequency vowel-like utterances similar to that of an infant.

1 Introduction

Spoken language is one of the important means for humans to communicate with others. Spoken language is one of the important means for humans to communicate with others. This is one of big issues dealt at RoboCup@home league where many teams utilize conventional speech recognition and generation systems [Chen 12]. The sound they generate is often heard unnatural due to a different sound generation from humans'. Further, conversation is realized mainly based on existing AI strategies. Towards more natural conversation with humans, there are two big issues to be considered:

1. How to generate vocal speech: due to properties of electromagnetic speakers, the sound properties generated by the conventional systems are different from humans' which may cause unnatural sounds.
2. How to learn to vocalize: if we adopt a human-like vocal system to attack the issue (1), the next issue is how the system learns to generate more natural vocal sounds.

As the first step towards these challenges, this article focuses on the mechanical design for infant-like vocal robot which is expected to learn to vocalize like a human infant.

In developmental psychology, it is suggested that an infant develops it through verbal interaction with caregivers by observation experiments [Bates 95]. However, what kind of underlying mechanism works for that and how caregiver's behavior affects on this process has not been fully investigated yet since it is very difficult to control the infant vocalization. On the other hand, there are several constructive approaches to understand the mechanisms by using infant robots with abilities equivalent to those of human infants, as a controllable platform [Asada 09].

Sasamoto et al. suggest a vocal robot as a platform for constructive investigation of the developmental process of vocalization [Sasamoto 13]. Unlike speech conversion and articulation simulators or speech synthesis, robotic platforms have advantages in terms of realtimeness, consonant vocalization by means of flow-acoustics, and interaction with humans. They actually built an infant-like vocal robot that mimics the anatomical shape of the articulator of human infant, and showed that its vocal cords and vocal tract could vocalize vowels in the same range of formant frequencies as that of human infant. However, the driving mechanism could not vocalize the same range because it did not comprise enough degrees of freedom (4-DOFs for tongue, 1-DOF for jaw, 1-DOF for soft palate).

On the other hand, vocal robots which have many degrees of freedom for articulation and can vocalize as well as human adults have been developed [Fukui 10, Sawada 08]. Particularly, Fukui et al. [Fukui 10] developed the vocalization robot WT-7RII which could vocalize not only vowels but also consonants by controlling many degrees-of-freedom (7-DOFs for tongue, 1-DOF for jaw, 1-DOF for soft palate, 5-DOFs for lip). However, this robot focused on reproduction of adults' utterance instead of infants'. Between adults and infants, the size of the articulator is different, which is closely related to the difference in their vocalization. Therefore, in order to reproduce infants' vocalization by means of many degrees of freedom (like WT-7RII), the problem of miniaturization has to be solved. Figure 1 shows the comparison of the sizes of the robots).

In this study, aiming at reproducing the infant vocalization, we miniaturized the articulation mechanism of WT-7RII, and developed a new infant-like vocal robot named "Lingua". This paper describes the design of its articulation mechanism and preliminary experiments.

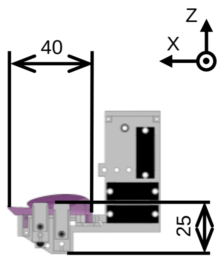
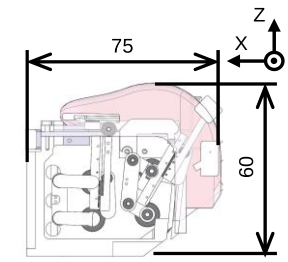
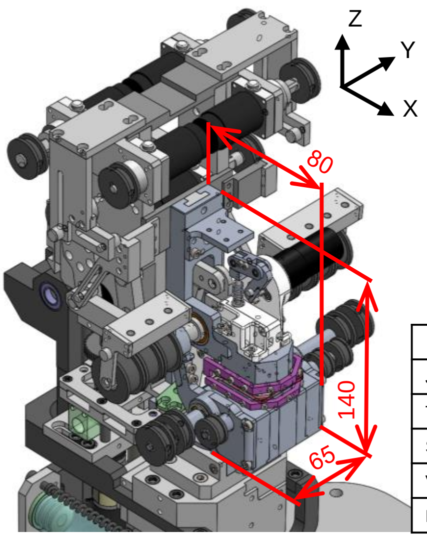
	Infant-like vocal robot[4]	WT-7RII[5]
Size		
DOF	4	7

Fig. 1 Size of articulation mechanism and degrees of freedom of the robots

2 Design of the articulation mechanism

2.1 Overview

Figure 2, 3, and 4 show Ligua's overview, DOF configuration, the mechanism of the vocal tract, and the mechanism of the vocal cords. This robot consists of a lung, vocal cords, and a vocal tract. The lung and the driving mechanism of the vocal cords are those of WT-7RII. We used the same vocal cords made from soft material as for the infant-like vocal robot by Sasamoto et al. [Sasamoto 13]. The following explains the design of the tongue which is compact as well human infant and has many degrees-of-freedom.



Part	DOF
Jaw	1
Tongue	7
Soft Plate	1
Vocal cords	2
Lungs	1

Fig. 2 Lingua's overview and DOF configuration

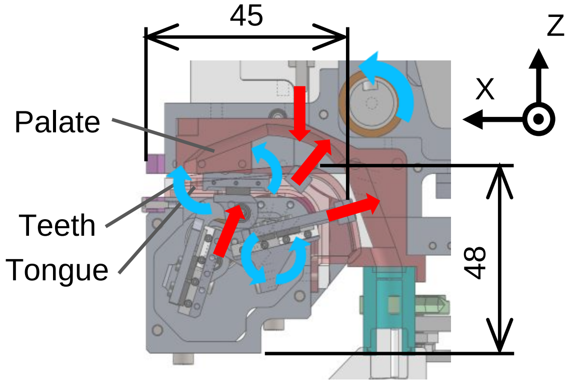


Fig. 3 Lingua's vocal tract

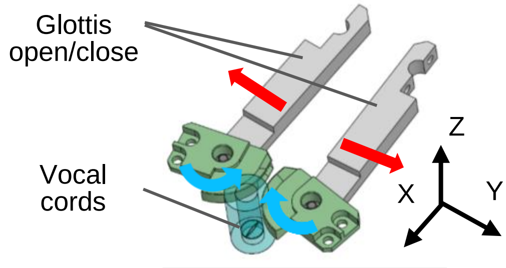
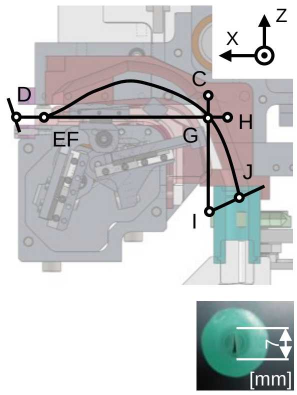


Fig. 4 Lingua's vocal cords mechanism



Parameter	Length [mm]
VTL (D-to-J)	93
VT-V (I-to-C)	39
PCL (I-to-G)	27
NPhL(G-to-C)	12
VT-H (D-to-H)	65
LTh(D-to-F)	10
ACL (F-to-G)	45
OPhW (G-to-H)	10
VT-O (F-to-H)	55

Fig. 5 Structural properties of the vocal cords and tract of Lingua. Parameters of the vocal tract correspond to those of anatomical data [Vorperian 09]

2.2 Requirements of the infant-like articulators

It is known that the sizes of articulators such as vocal cords and tracts influence acoustic features of its vocalization. The shapes of vocal cord [Eckel 99] and vocal tract [Vorperian 09] change with the growth. It is therefore necessary to consider the changes in order to understand infant's vocal development [Mugitani 12]. Therefore, we set these anatomical dimensions as a requirement specification (Figure 5). The widths of the tongue and oral cavity are decided as 15 mm and 30 mm based on a report in which changes in infant's teeth row were measured [Hayama 99].

2.3 Miniaturization of the tongue mechanism

The tongue mechanism consists of 7-DOFs that combine rotational and linear movement (Figure 6). We downsized the linkage mechanism which connects them. The linkage of WT-7RII's tongue consisted of plural shafts by parallel and slider cranks, but we minimized the parallel crank by adopting a coaxial mechanism for it (Figure 7).

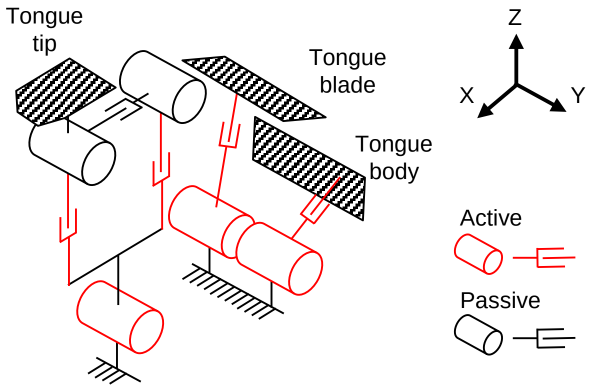
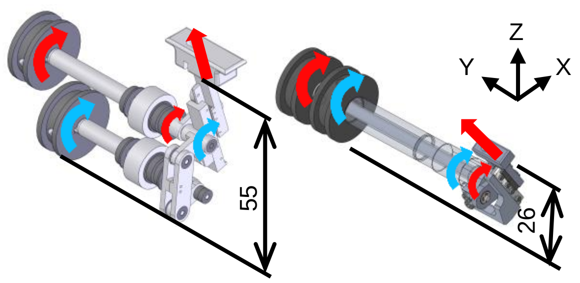


Fig. 6 DOF configuration of Lingua's tongue



(a) WT-7RII (b) Lingua

Fig. 7 Comparison of the link mechanisms

2.4 Miniaturization of the arrangement of the tongue mechanism

The movable range of each linkage tip was determined based on simulation results of infants' 3 vowels utterances (/a/, /i/, /u/) by a VLAM articulation simulation [Boë 13] (Figure 8). The red lines in the figure are the tips of the linkages of the tongue tip, tongue blade, and tongue body. The layout and dimensions of the linkage were determined based on the range calculated by inverse kinematics. Figure 9 shows the arrangement of the linkages.

2.5 Strength of the mechanism

The surface of the tongue was molded in silicone rubber. The linkages are put loads when they drives the surface. We had to save enough yield strength for reciprocating articulatory movements.

Figure 10 shows the deformation of the surface of the tongue blade. We regarded the surface as an elastic element and represented it by a model in which a spring is connected to a slider crank (Figure 11).

We calculated the elastic coefficient of the tongue based on measuring its stretching when exposed to external load. Then, we designed the mechanical parts such that the minimum yield safety ratio could be 5 by using FEM. Hence, the linkages have enough strength against reciprocating articulatory movements.

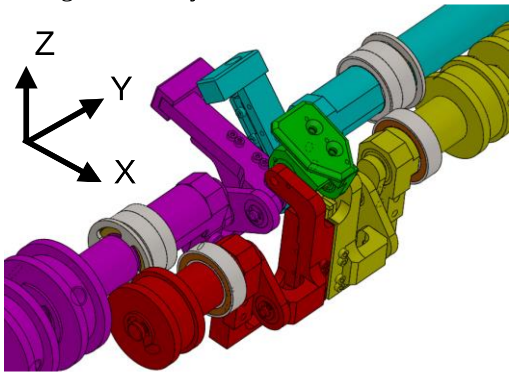
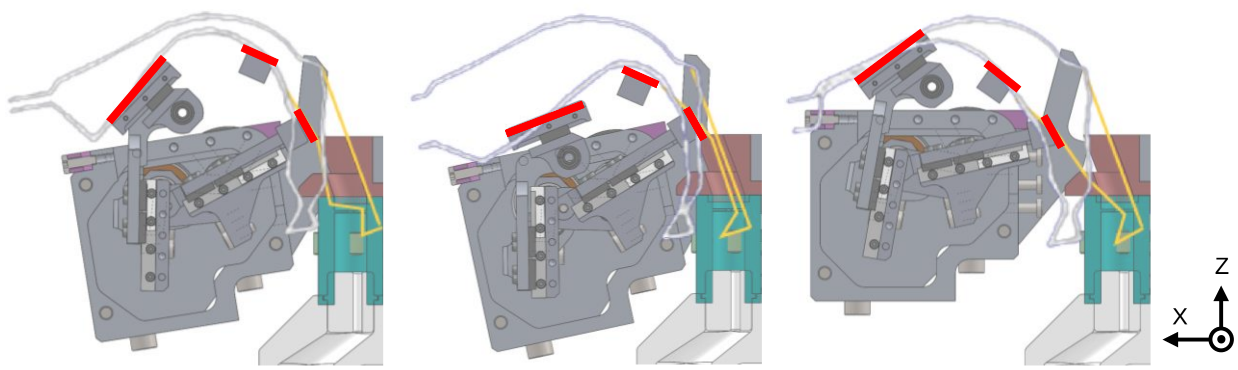


Fig. 9 Overview of the tongue mechanism



(a) /a/ (b) /i/ (c) /u/

Fig. 8 Lingua's articulation imitating VLAM simulated vocal tract shapes for a 7-month-old child uttering /a/, /i/ and /u/ (yellow lines indicate shapes corrected the angle of the neck)

Fig. 10 Deformation of the tongue made of soft material by the tongue blade linkage

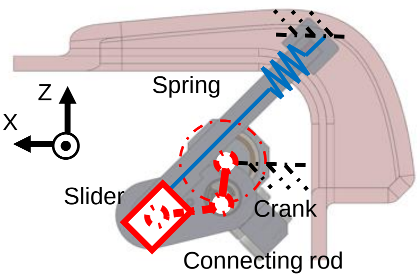


Fig. 11 Spring and linkage model of the tongue blade mechanism.

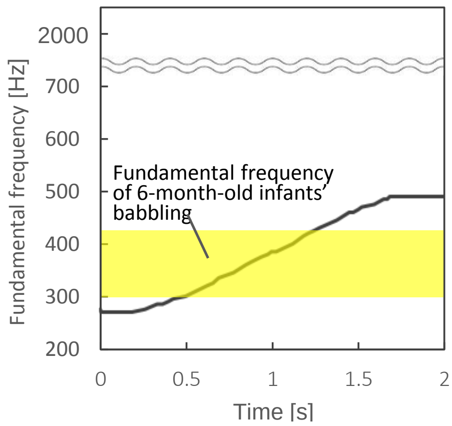


Fig. 12 Result of the fundamental frequency measurement

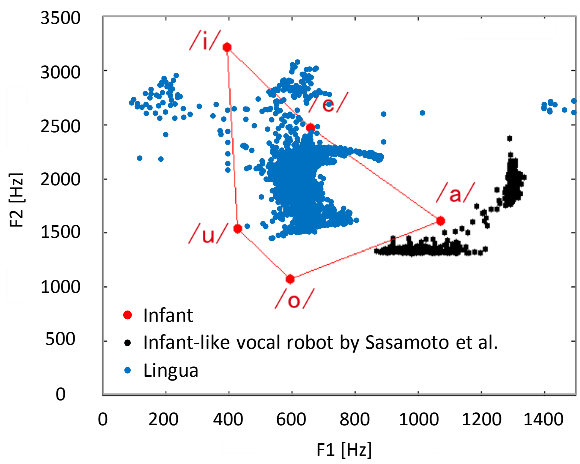


Fig. 13 F1-F2 space of the robot during vocalization

3 Preliminary evaluation

3.1 Fundamental frequency

We conducted preliminary experiments to validate that the vocal cords mechanism and tongue mechanism developed could vocalize as well as human infants. We measured the fundamental frequency and the formant frequencies of the robot and compared them with those of infant.

The fundamental frequency of human infants is generally regarded as from 200 Hz to 2000Hz. Figure 12 shows the temporal change in the fundamental frequency when the vocal cords mechanism changes the tension of the vocal cords. This indicates the fundamental frequency of the robot lies within the range of human infant. Particularly, it fills a fundamental frequency of babbling [Whalen 93] which plays a important role in vocal acquisition.

3.2 Formant frequencies

We drove each linkage of the tongue and the jaw manually, and measured formant frequencies of Lingua's vowel vocalization. Figure 13 draws the result, that of the infant-like vocal robot made by Sasamoto et al. [Sasamoto 13], and infants' vowel utterance [Kasuya 68] in F1-F2 space. While the robot by Sasamoto et al. could vocalize in the space far from infants, Lingua could vocalize in the similar space to it.

In this experiment, all of the configuration space was not fully examined. Near future, we will examine all of the configuration space.

4 Conclusion

In this article, we described the design of the articulation mechanism of the infant-like vocal robot "Lingua" towards natural conversation with people@home. Preliminary evaluation shows Lingua's ability to vocalize the vowels as well as human infant. In the future, we will develop the lip mechanism and examine the vocalization performance of the overall mechanism for vowels and consonants. We also aim to reproduce crying and babbling. Moreover, we will conduct interaction experiments between the robot and a caregiver in order to investigate how infants' vocalization develops.

Acknowledgments

This work was partially supported by the MEXT Grant-in-Aid for Specially Promoted Research (24000012), and the MEXT project "Creating Hybrid Organs of the future" at Osaka University.

We appreciate the provision of WT-7RII by Atsuo Takanishi Laboratory and Masaaki Honda Laboratory at Waseda University.

References

[Asada 09] Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Yoshida, C.: Cognitive developmental robotics: A survey, IEEE Transactions on Autonomous Mental Development, 1(1), 12--34 (2009).

[Bates 95] Bates, E., Dale, P. S., Thal, D. : Individual differences and their implications for theories of language development. The handbook of child language, 96--151 (1995).

[Boë 13] Boë, L. J., Badin, P., Ménard, L., Captier, G., Davis, B., Macneilage, P., Schwartz, J. L. : Anatomy and control of the developing human vocal tract: A response to Lieberman. Journal of Phonetics, 41(5), 379--392 (2013).

[Eckel 99] Eckel, H.E., Koebke, J., Sittel, C., Sprinzl, G.M., Pototsching, C., Stennert, E., : Morphology of the human larynx during the first five years of life studied on whole organ serial sections. The Annuals of Otology, Rhinology and Laryngology, 108(3), 232--238 (1999).

[Fukui 10] Fukui, K., Kusano, T., Mukaeda, Y., Suzuki, Y., Takanishi, A., Honda, M.: Speech Robot Mimicking Human Articulatory Motion. Proceedings of INTERSPEECH 2010, 1021--1024 (2010).

[Hayama 99] Hayama, Y., Sugimura, K., Yang, J., Shinji, H., Uchimura N. : Study on the Development of the Dental arch from the Pre-dental phase until the Deciduous Dentition. The Japanese Journal of Pediatric Dentistry, 37(3), 559--572 (1999).

[Kasuya 68] Kasuya, H., Suzuki, H., Kido, K. : Changes in pitch and first three formant frequencies of five Japanese vowels with age and sex of speakers. Journal of the Acoustical Society of Japan, 24(6), 355--364 (1968).

[Mugitani 12] Mugitani, R., Hiroya, S. : Development of vocal tract and acoustic features in children. Acoustical Science and Technology, 33(4), 215--220(2012).

[Chen 12] Chen, X., Stone, P., Sucar L. E., van der Zang, T. (Eds.) : RoboCup 2012: Robot Soccer World Cup XVI. Springer (2013).

[Sasamoto 13] Sasamoto,Y., Nishijima N., and Asada M.: Towards understanding the origin of infant directed speech: A vocal robot with infant-like articulation. Proceedings of the IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (2013).

[Sawada 08] Sawada, H., Kitani, M., Hayashi, Y.: A robotic voice simulator and the interactive training for hearing-impaired people. Journal of biomedicine ¥& biotechnology, 2008, 768232 (2008).

[Vorperian 09] Vorperian, H.K., Wang, S., Chung, M.K., Schimek, E.M., Durtschi, R.B., Kent, R.D., Ziegert, A.J., Gentry, L.R.,: Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. The Journal of the Acoustical Society of America, 125, 1666 (2009).

[Whalen 93] Whalen, D. H., Levitt, A. G., Smorodinsky, I., Hsiao, P. L. : Intrinsic F0 in the babbling of infants at ages 6, 9, and 12 months. The Journal of the Acoustical Society of America, 94(3), 1882--1883 (1993).