

Design and Preliminary Evaluation of the Vocal Cords and Articulator of an Infant-like Vocal Robot “Lingua”

Nobutsuna Endo, Tomohiro Kojima, Hisashi Ishihara, Takato Horii, Minoru Asada, *Member, IEEE*

Abstract— Several observational studies have suggested that the vocal development of an infant towards language acquisition is by interaction with caregivers. However, owing to the difficulty of controlling infant vocalization, the underlying mechanisms of the process and how they are affected by the behavior of the caregiver are yet to be clearly established. Towards clarifying these outstanding issues, we developed an infant-like vocal robot, “Lingua,” as a controllable vocal platform that affords a model of real infant vocalization. Lingua has two features, namely, infant-like voice and high articulation capability. The shapes of its vocal cords and vocal tract are similar to those of a 6-month-old infant as determined from anatomical data. Seven degrees of freedom of tongue articulation were realized by a sophisticated design comprising linkage mechanisms inside a miniaturized vocal tract, and this enabled the achievement of high articulation performance. The relationship between the material hardness of the vocal fold and its acoustic performance was examined, and the results of preliminary experiments showed that the robot could vocalize almost the same ranges of fundamental frequencies and vowel-like utterances as an infant.

INTRODUCTION

Speech is one of the important means of human communication, and the experimental observations of developmental psychologists have suggested that an infant develops the capability through interaction with caregivers [1]–[4]. It has also been suggested that the non-verbal and emotive utterances of infants to caregivers affect the development of speech [5]. It is therefore necessary to consider the effects of infant emotive information on caregivers in order to understand the process of vocal development by interaction. However, owing to the difficulty of controlling infant vocalization, the underlying mechanisms of the process and how they are affected by the behavior of the caregiver are yet to be fully established. Nevertheless, infant robots with abilities similar to those of infant humans afford controllable platforms and constructive methods for understanding the mechanisms of vocal development [6].

Sasamoto et al. built a vocal robot as a platform for studying the process of vocal development [7]. Unlike speech simulators, robotic platforms have the advantages of real time responses, vocalization by means of flow acoustics, and interaction with humans. The anatomical shape of the articulator of the infant-like vocal robot of Sasamoto et al. [7] is similar to that of a human infant, and the vocal cords and tract could vocalize vowels within the same range of fundamental frequency as an infant human. However, owing to the much fewer degrees of freedom (DOFs) (four for the

tongue, one for the jaw, and one for the soft palate), the system could not be controlled to vocalize the same range of formant frequencies as an infant human.

However, vocal robots with many articulation DOFs, which enable vocalization similar to that of a human adult, have been developed [8, 9]. Particularly, Fukui et al. [8] developed the vocal robot WT-7RII, which is capable of vocalizing not only vowels, but also consonants by its many control DOFs (seven for the tongue, one for the jaw, one for the soft palate, and five for the lips). However, the design of the robot focused on the reproduction of adult utterance. The sizes of the vocal cords and articulators of adults and infants differ, and this causes their vocalizations to differ. Downsizing is therefore required to reproduce infant vocalization by means of many DOFs such as those of WT-7RII.

The authors have developed a new infant-like vocal robot name Lingua [10], aiming at coping with both infant size and high articulation capabilities. The infant-like vocal robot of Sasamoto et al. [7] is capable of vocalizing infant-like high-pitch sounds, but not a sufficient range of vowel sounds. WT-7RII [8] can vocalize sufficient ranges of vowel and consonant sounds, but is incapable of infant-like high-pitch vocalization. We therefore combined the two vocal robots in Lingua. The rest of this paper describes the design and preliminary evaluation of the vocal cords and articulator of Lingua.

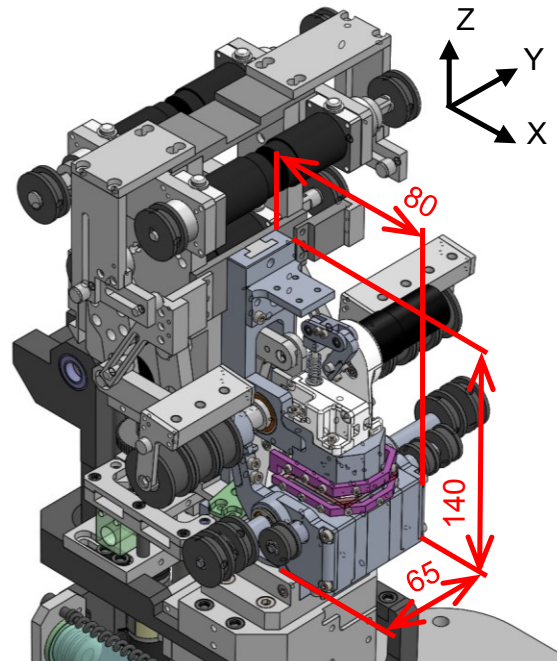


Fig. 1 Overview of Lingua

DESIGN OF THE INFANT-LIKE VOCAL ROBOT “LINGUA”

A. Overview

Fig. 1 and Table 1 present an overview and the DOF configuration of Lingua. The robot consists of a lung, vocal cords, and a vocal tract. The lung and driving mechanism of the vocal cords are those of WT-7RII. We used the same geometry of the vocal cords as that of the infant-like vocal robot of Sasamoto et al. [7]. As noted earlier, the sizes of the vocal cords and vocal tract affect the acoustic features of vocalization. In addition, the shapes of the vocal cords [11] and vocal tract [12] of infant humans gradually change, and it is necessary to put this into consideration in studying infant vocal development [13]. Following are descriptions of the designs of the vocal cords and tongue of Lingua.

B. Vocal Cords

It is known that the fundamental frequency varies with the size and shape of the vocal cords. As previously observed, the geometry of the vocal cords of Lingua is the same as that of the infant-like vocal robot of Sasamoto et al. [7], which imitates that of a 6-month-old infant. Based on anatomical

TABLE I. DOF CONFIGURATION OF LINGUA

Part	DOFs
Jaw	1
Tongue	7
Soft palate	1
Vocal cords	2
Lungs	1

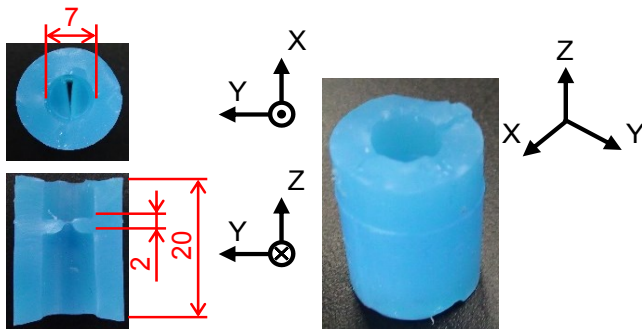


Fig. 2 Infant-like vocal cords

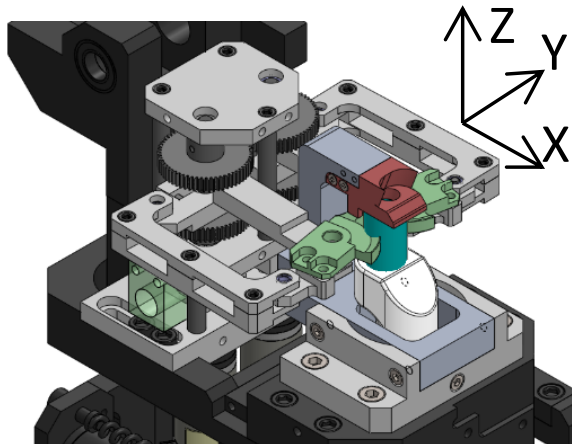


Fig. 3 Overview of the vocal cord mechanism

data [12], we set the length of the vocal cords of Lingua at 7 mm, and the thickness of its fold at 2 mm. The artificial vocal cords, which were made from a silicone resin, reproduced the deformation and vibration of the human vocal cords caused by expiration from the lung. Fig. 2 shows the artificial vocal cords.

The variation of the fundamental frequency of the human vocal cords with the tension was reproduced in Lingua using the tensioner mechanism of WT-7RII [8]. The mechanism has 2-DOFs (one for rotating a clamp to pull the vocal cords, and

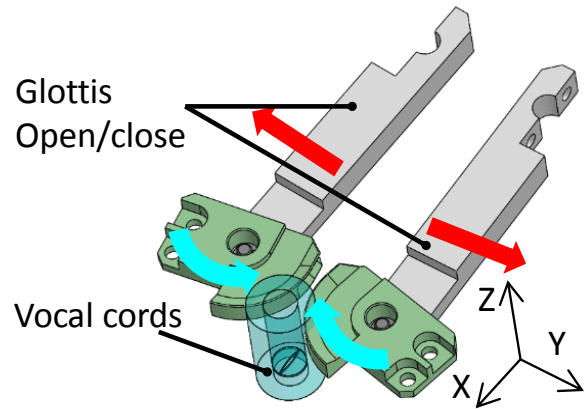


Fig. 4 Vocal cord mechanism

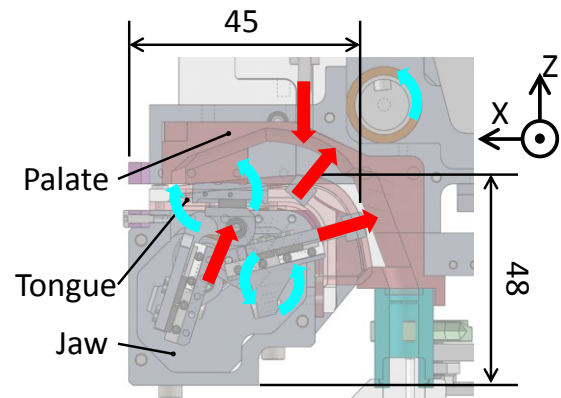


Fig. 5 Side view of the vocal tract

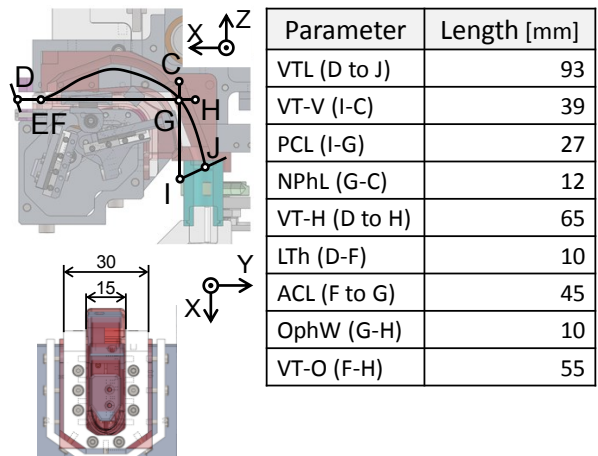


Fig. 6 Dimensions of the vocal tract

the other for adjusting the clamping force). Figs. 3 and 4 present an overview and the details of the vocal cord mechanism.

C. Articulators (Tongue and Jaw)

Fig. 5 shows the side view of the vocal tract of Lingua. The articulators contribute to the acoustic features of the sounds produced by the vocal cords by changing the cross-sectional area of the vocal tract and its length from the lips. We determined the mid-sagittal dimensions of the vocal tract and the widths of the oral cavity and the tongue based on anatomical data [12, 14] (Fig. 6). The DOFs and control points of the tongue were determined based on those of WT-7RII, which is capable of human-like vocalization.

To realize the tongue mechanism of WT-7RII on an infant scale, we downsized it by half. The Lingua tongue mechanism consisted of 7-DOFs of rotational and linear movements (Fig. 7) and utilized downsized linkages. The linkage of the tongue of WT-7RII comprised plural shafts connected by parallel and slider cranks (Fig. 8(a)). The number of parallel cranks was minimized by using a coaxial mechanism (Fig. 8(b)).

The movement range of each linkage tip was determined based on an articulation simulation of the three vowel utterances of an infant (/a/, /i/, /u/) using a Variable Linear Articulatory Model (VLAM) [15] (Fig. 9). The red lines in the figure represent the tips of the linkages of the tongue tip, tongue blade, and tongue body. The layout and dimensions of the linkages were determined based on the ranges determined by inverse kinematics. Fig. 10 shows the arrangement of the linkages.

EXPERIMENTAL EVALUATION

It was necessary to determine the vibration characteristics of the artificial vocal cords owing to their simplified structure compared to those of an infant human. However, nonlinear analytical simulation of the fluid-structure interaction of the vocal cords is very complicated, and the analytical method is not well-established. Although some studies have utilized approximations based on simple wave equations [16]–[18], it was difficult to apply such to our vocal cords owing to the oversimplification of the present model. In addition, although some studies have been conducted to measure the elastic modulus of the human vocal fold [19, 20], the results vary

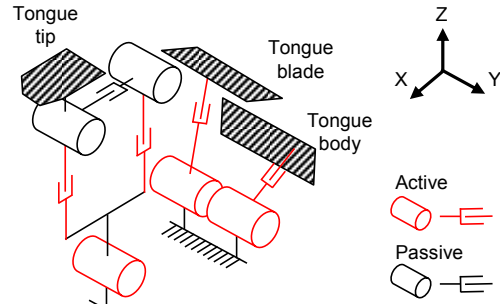


Fig. 7 DOF configuration of the tongue

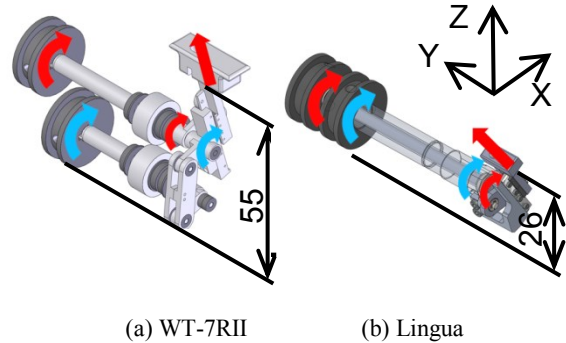


Fig. 8 Comparison of the linkage mechanisms of WT-7RII and Lingua

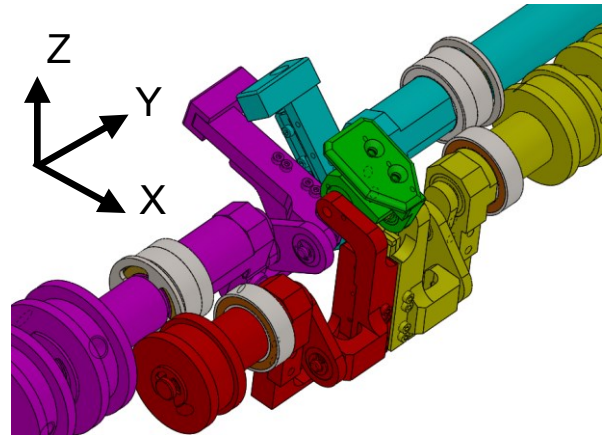


Fig. 10 Overview of the tongue mechanism

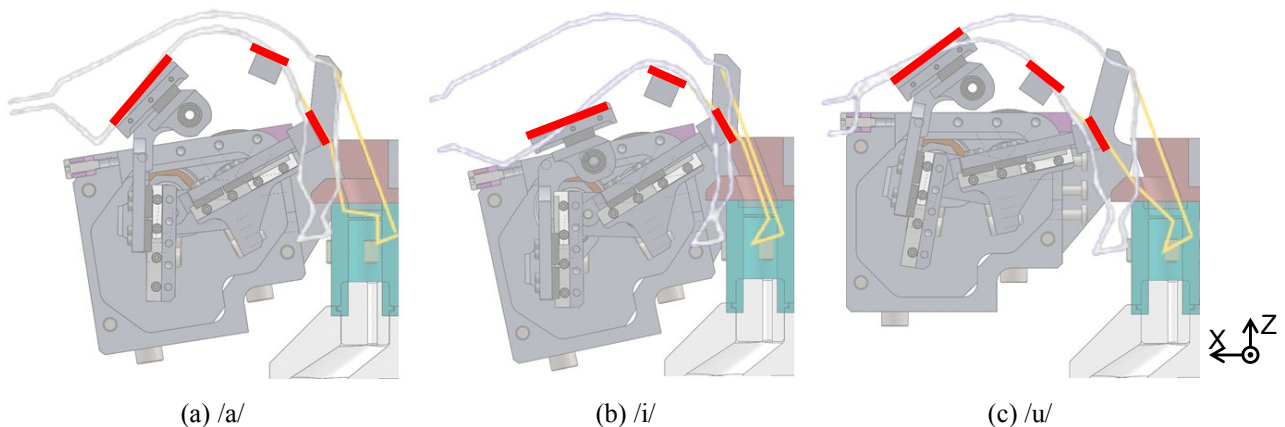


Fig. 9 VLAM simulation of the shapes of the vocal tract of Lingua while imitating a 7-month-old child uttering /a/, /i/, and /u/ (the yellow lines indicate the shapes corrected based on the angle of the neck)

significantly with respect to age and sex, and this made it difficult to estimate the characteristics of an infant vocal fold. We therefore fabricated several vocal cords with differing hardness and measured their hardness and acoustic performance. Specifically, nine vocal cords were fabricated using different compounding ratios of the plasticizer of the silicone resin (silicone thinner). We then measured the fundamental frequencies of the vocal cords to evaluate the effects of their respective material properties on vocalization. We also evaluated the acoustic performance of the articulator mechanism.

A. Hardness of the Silicone Resin

Owing to the difficulty of directly measuring the hardness of the vocal fold, we prepared test pieces for the measurement. Nine test pieces were prepared using different compounding ratios of the plasticizer of the silicone resin, ranging between 60% and 140% (i.e., the ratio of the weight of the plasticizer to that of the resin). The size of the test pieces is shown in Fig. 11 based on the specifications of the durometers (Durometer HT6510F by Sato Shoji Co., Ltd.; ASKER Durometer Type FP by Kobunshi Keiki Co., Ltd.). The measurement method was basically in accordance with the JIS standard (JIS K 6253-3:2012, compatible with ISO 7619-1:2010). We allowed an interval of more than 60 s between measurements of the same test piece to compensate for deformation hysteresis. Each test piece was measured five times and the average reading of each durometer was obtained.

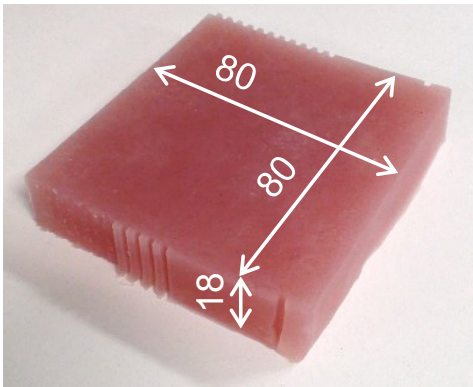


Fig. 11 Test piece made from silicone resin

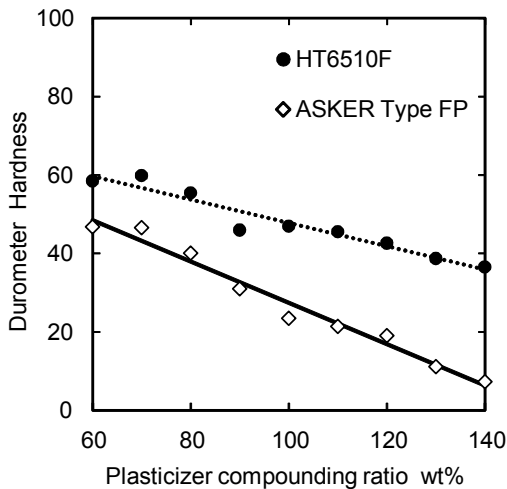


Fig. 12 Variation of the durometer hardness with the compounding ratio of the plasticizer

Fig. 12 shows the measurement results. The hardness measured by both durometers decreased linearly with increasing plasticizer compounding ratio. This suggests that the degree of deformation of the vocal cords by expiration increases with increasing compounding ratio.

B. Fundamental Frequency

We installed the vocal cords in Lingua and conducted the vocalization experiment (Fig. 13). The objective of the experiment was to determine the range of fundamental frequencies that the vocal cords could produce. Hence, we did not install the vocal tract, and manually moved the tensioner mechanism of the vocal cords. We removed the non-modal sounds that were unlike human utterance from the recorded sounds, and obtained the maximum/minimum fundamental frequency by Praat, a speech analysis software [21].

The results are shown in Fig. 14. The graph excludes the data for 140% plasticizer compounding ratio because the vocal cords produced using this combination could not produce the modal voice. There was no significant change in the minimum fundamental frequency; the difference being approximately 100 Hz for most of the compounding ratios. There was gradual increase up to a compounding ratio of

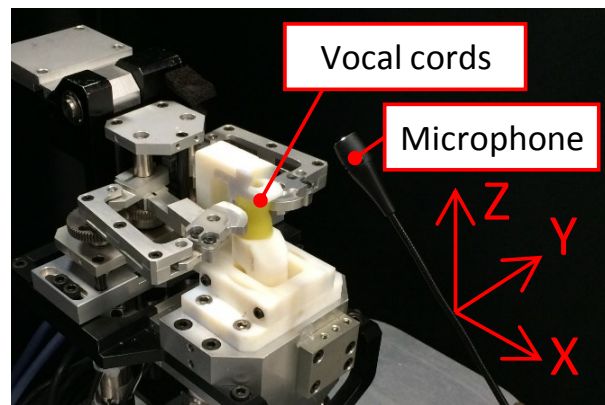


Fig. 13 Experimental setup for measuring fundamental frequency

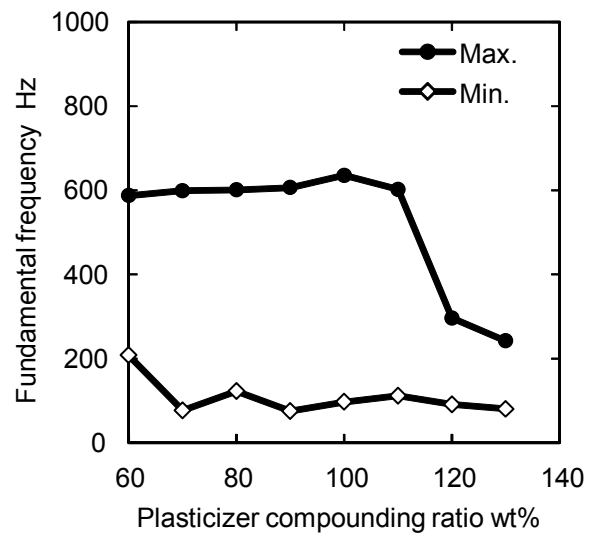


Fig. 14 Measured maximum/minimum fundamental frequencies

110% and a sharp decrease thereafter. It is suggested that the range of fundamental frequencies of vowels uttered by infants is 300–430 Hz [22]. Vocal cords made using no more than 110% plasticizer satisfy this range. The highest frequency was obtained for 100% plasticizer. A possible cause of the gradual increase in the maximum fundamental frequency up to 110% plasticizer compounding ratio is the inability of the vocal cords to vibrate adequately under tension owing to the excessive hardness of the folds. Conversely, the sharp decrease in the fundamental frequency above 110% compounding ratio may be attributed to the folds being unable to acquire sufficient tension and undergo significant deformation.

C. Formant Frequencies

We drove each linkage of the tongue and jaw manually and measured the formant frequencies of the vowel vocalization of Lingua to assess the effectiveness of the articulator mechanism. Vocal cords produced using 70% plasticizer were employed for the experiment. The results are shown in Fig. 15, where they are compared with those for the infant-like vocal robot of Sasamoto et al. [7] and actual infant vowel utterance [23] in an F1-F2 space. Whereas the robot of Sasamoto et al. [7] could vocalize in a space far from that of real infants, the vocalization space of Lingua was comparable to that of real infants.

The configuration spaces of the motors were not fully examined in this experiment. This will be done in a future study.

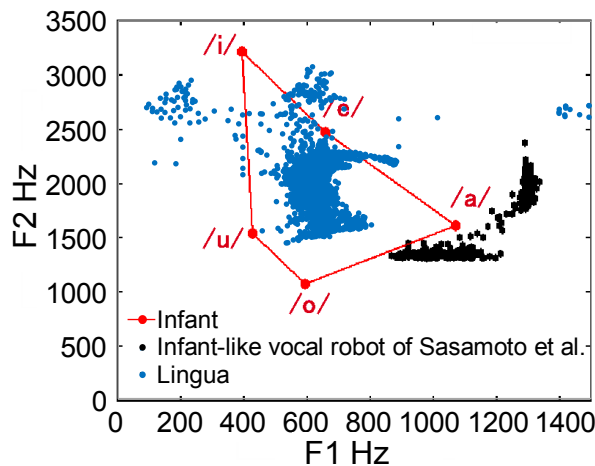


Fig. 15 Measured formant frequencies

CONCLUSION AND FUTURE WORK

In this paper, we described the design of the vocal cords and articulation mechanism of Lingua, an infant-like vocal robot that combines features of a previous infant-like vocal robot and the adult-like vocal robot WT-7RII. In particular, we used the vocal cords of the previous infant-like robot and examined how its hardness affected the acoustic performance. We also downsized the articulation mechanism of WT-7RII and evaluated the vowel vocalization. The results of these preliminary investigations showed that the vocalization ability of Lingua was comparable to that of an infant human.

In a future work, we will develop the lip mechanism of Lingua and examine the performance of the overall mechanism with regard to the vocalization of vowels and consonants. We also aim to reproduce crying and babbling, and plan to conduct experiments on interaction between the robot and a caregiver to investigate the development of infant vocalization.

ACKNOWLEDGMENT

This work was partially supported by the JSPS 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

- [1] E. Bates, P. S. Dale, and D. Thal, “Individual differences and their implications for theories of language development,” *The Handbook of Child Language*, pp. 96–151, 1995.
- [2] M. M. Chouinard and E. V. Clark, “Adult reformulations of child errors as negative evidence,” *Journal of Child Language*, vol. 30, no. 3, pp. 637–669, 2003.
- [3] S. J. Pawlby, “Imitative interaction,” in *Studies in Mother-Infant Interaction*, H. R. Schaffer, Ed. London: Academic Press, 1977, pp. 203–223.
- [4] M. Pelaez, J. Virues-Ortega, and J. L. Gewirtz, “Reinforcement of vocalizations through contingent vocal imitation,” *Journal of Applied Behavior Analysis*, vol. 44, no. 1, pp. 33–40, 2011.
- [5] Y. Shimura and S. Imaizumi, “Infant’s expression and perception of emotion through vocalizations,” *Proceedings of the International Conference on Spoken Language Processing (ICSLP’94)*, vol. 4, pp. 1703–1706, 1994.
- [6] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, and C. Yoshida, “Cognitive developmental robotics: A survey,” *IEEE Transactions on Autonomous Mental Development*, vol. 1, no. 1, pp. 12–34, 2009.
- [7] Y. Sasamoto, N. Nishijima, and M. Asada, “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.
- [8] K. Fukui, T. Kusano, Y. Mukaeda, Y. Suzuki, A. Takanishi, and M. Honda, “Speech robot mimicking human articulatory motion,” *Proceedings of INTERSPEECH 2010*, pp. 1021–1024, 2010.
- [9] H. Sawada, M. Kitani, Y. Hayashi, “A robotic voice simulator and the interactive training for hearing-impaired people,” *Journal of Biomedicine & Biotechnology*, pp. 768232, 2008.
- [10] Nobutsuna Endo, Tomohiro Kojima, Yuki Sasamoto, Hisashi Ishihara, Takato Horii, and Minoru Asada, “Design of an articulation mechanism for an infant-like vocal robot ‘Lingua’,” *Proceedings of the 3rd Conference on Biomimetic and Biohybrid Systems (Living Machines 2014)*, July, 30, 2014.
- [11] H. E. Eckel, J. Koebke, C. Sittel, G. M. Sprinzel, C. Pototsching, and E. Stennert, “Morphology of the human larynx during the first five years of life studied on whole organ serial sections,” *The Annals of Otolaryngology and Laryngology*, vol. 108, no. 3, pp. 232–238, 1999.
- [12] H. K. Vorperian, S. Wang, M. K. Chung, E. M. Schimek, R. B. Durtschi, R. D. Kent, A. J. Ziegert, and L. R. Gentry, “Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study,” *The Journal of the Acoustical Society of America*, vol. 125, pp. 1666, 2009.
- [13] R. Mugitani and S. Hiroya, “Development of vocal tract and acoustic features in children,” *Acoustical Science and Technology*, vol. 33, no. 4, pp. 215–220, 2012.
- [14] Y. Hayama, K. Sugimura, J. Yang, H. Shinji, and N. Uchimura, “Study on the development of the dental arch from the pre-dental phase until

- the deciduous dentition,” *The Japanese Journal of Pediatric Dentistry*, vol. 37, no. 3, pp. 559–572, 1999.
- [15] L. J. Boë, P. Badin, L. Ménard, G. Captier, B. Davis, P. Macneilage, and J. L. Schwartz, “Anatomy and control of the developing human vocal tract: A response to Lieberman,” *Journal of Phonetics*, vol. 41, no. 5, pp. 379–392, 2013.
- [16] R. W. Chan, T. Siegmund, and K. Zhang, “Biomechanics of fundamental frequency regulation: Constitutive modeling of the vocal fold lamina propria,” *Logopedics Phoniatrics Vocology*, vol. 34, no. 4, pp. 181–189, 2009.
- [17] T. Riede, A. York, S. Furst, R. Müller, and S. Seelecke, “Elasticity and stress relaxation of a very small vocal fold,” *Journal of Biomechanics*, vol. 44, no. 10, pp. 1936–1940, 2011.
- [18] J. E. Kelleher, T. Siegmund, R. W. Chan, and E. A. Henslee, “Optical measurements of vocal fold tensile properties: Implications for phonatory mechanics,” *Journal of Biomechanics*, vol. 44, no. 9, pp. 1729–1734, 2011.
- [19] R. W. Chan, M. Fu, L. Young, and N. Tirunagari, “Relative contributions of collagen and elastin to elasticity of the vocal fold under tension,” *Annals of Biomedical Engineering*, vol. 35, no. 8, pp. 1471–1483, 2007.
- [20] D. K. Chhetri, Z. Zhang, and J. Neubauer, “Measurement of Young’s modulus of vocal folds by indentation,” *Journal of Voice*, vol. 25, no. 1, pp. 1–7, 2011.
- [21] “Praat: Doing Phonetics by Computer,” <http://www.fon.hum.uva.nl/praat/>
- [22] D. H. Whalen, A. G. Levitt, I. Smorodinsky, and P. L. Hsiao, “Intrinsic F0 in the babbling of infants at ages 6, 9, and 12 months,” *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1882–1883, 1993.
- [23] H. Kasuya, H. Suzuki, and K. Kido, “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*, vol. 24, no. 6, pp. 355–364, 1968.