Towards understanding the origin of infant directed speech: A vocal robot with infant-like articulation

Yuki Sasamoto, Naoto Nishijima and Asada Minoru
Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University
2-1 Yamadaoka, Suita, Osaka 565–0871
Email: {yuki.sasamoto, naoto.nishijima, asada}@ams.eng.osaka-u.ac.jp

***Abstract*—Infant-Directed Speech (IDS) is the non-standard form of caregivers' speech to their infants. Developmental studies indicate that IDS changes from infant-directed to adult-directed depending on infant's age and/or linguistic level. However, it is still unclear what features in infants cause IDS. This article introduces a vocal robot with an infant-like articulatory system to attack the issue by means of a constructive approach. A preliminary experiment implies that our robot can vocalize structurally similar to infant articulation although it is mechanically rather different.**

I. Introduction

Caregiver-Infant vocal interactions are important in many aspects, especially, for infant vocal development. It is well-known that caregivers modify their speech when they talk to their infants. This kind of speech is called Infant-Directed Speech (IDS), and has acoustically exaggerated characteristics: higher pitch [1], exaggerated intonation contours [2], hyperarticulated vowel [3], and articulatory exaggerations [4]. Developmental studies indicate that IDS changes from infant-directed to adult-directed depending on infant's age [5] and/or linguistic level [6]. However, it is unclear what features in infants provoke IDS.

Existing experimental paradigms in developmental psychology can hardly answer this question since infants' vocalization in the interaction cannot be controlled. Then, interdisciplinary approaches seem promising. Among them, the constructive approach [7] is aiming at understanding human's cognitive developmental processes by using controllable physical robots. Focusing on physical vocalization, few trials have been attempted [8], [9], [10]. However, no infant-like vocalization has yet been realized to the best of our knowledge.

We propose a vocal robot with an infant-like articulatory system to attack the issue by meant of a constructive approach. A preliminary experiment implies that our robot can vocalize structurally similar to infant articulation although it is mechanically rather different.

II. Design policy and specifications

The vocal cord and tract are essential parts for human vocalization since acoustic characteristics of vocal sounds, such as pitch and tone depend on them. Therefore, they are main targets for designing human-like, or especially in this article, infant-like articulation.

Anatomical studies [11] imply the following:

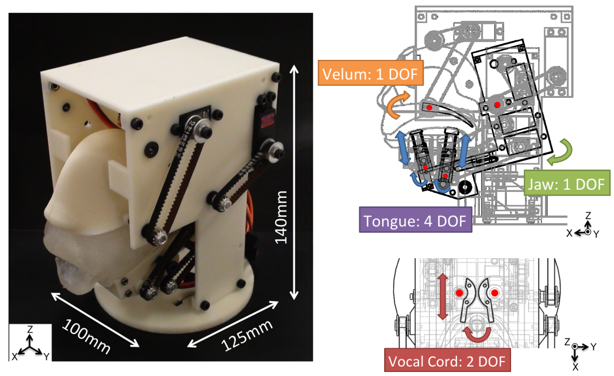


Fig. 1. Appearance and mechanical properties of the robot. Mainframes are made of ABS resin, and the vocal cord and tract are made of Septon by Kuraray Co, Ltd.

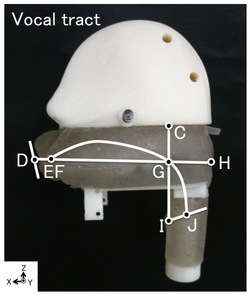
- Extension and rotation of the vocal cord regulate fundamental frequency and voiced/unvoiced switching.
- Tongue's upward/downward as well as forward/backward movements regulate formant frequencies.
- Opening and closing of the velum are involved in the production of nasal sounds.
- Moving the jaw up and down is involved in obstruent production.

Considering the above, the number of total degrees of freedom (DOF) is determined to be eight: two for the vocal cord, four for the tongue, one for the velum, and one for the jaw (See Fig. 1).

The morphologic development of the human larynx during the first year of life is investigated by measuring the sizes of plastinated whole organ serial sections of many infants' larynges [12]. Anatomical data of individual oral and pharyngeal portions of the vocal tract are also investigated in infants aged several months by using medical imaging techniques [13]. We designed the robot vocal tract and cord to meet these specifications. In the current version, we used anatomical data of infants at 6 months of age (Fig. 2), because infants at this age start babbling and therefore they are expected to have an articulatory system suitable for basic verbal communication.

III. Vocalization experiment

We conducted a vocalization experiment in order to evaluate the performance of the robot vocalization in which the first



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-E)	10
ACL (F-to-G)	45
OPhW (G-to-H)	10
VT-O (E-to-H)	55

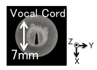
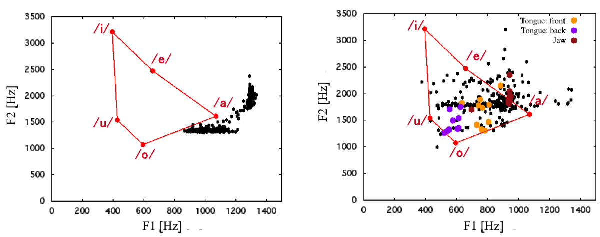


Fig. 2. Structural properties of the vocal cord and tract of the robot. Parameters of the vocal tract correspond to those of anatomical data [13]



(a) With actuators (b) Without actuators

Fig. 3. Acoustic features of the robot during vocalization: (a) with actuators, and (b) without actuators which means that its articulatory systems is manipulated by human-hand during vocalization

and second formant frequencies were measured as the main acoustic features. Fig. 3 (a) shows the formant frequencies of the robot vocalizing with its actuators. In this condition, two significant actuators for vowel vocalization, that is tongue and jaw, are focused and the rotation of each actuator is discretized into three levels. Black dots correspond to combinations of each actuator's level, and red line indicates the vowel space of Japanese children [14]. We can see that the vocal space of the robot is grossly disjoint to the infant's one. This might be because transformations of the vocal tract were not adequate. On the other hand, the vocal space of the robot is widely distributed inside the infant's one if it vocalizes without any actuators (Fig. 3 (b)) but the vocal tract of the robot is transformed by human-hand. Additional color dots correspond to each actuator's movement which is independently transformed. This implies structurally similar to infants' vocal tracts and has the potentioal to generate infant-like articulation.

IV. Summary and Discussions

In this paper, we introduce the first step of developing a vocal robot, which has an infant-like articulatory system to study the relationship between speech features in infant and resulting caregiver's utterances. The robot is developed by replicating the infant's articulatory system mechanically and structurally. As a result of the vocalization experiment, we found that the robot can vocalize structurally similar to infant articulation although it is mechanically rather different. Therefore, we need to redesign the actuation system so that robot can vocalize mechanically as an infant. We also need to investigate what kinds of features elicit IDS by measuring humans' behavior in response to the robot's vocalization in human-robot interactions.

Acknowledgment

This study has been supported by Grant-in-Aid for Specially promoted Research (No. 24000012) and Grant-in-Aid for JSPS Fellows (No. 24925). We also thanks to the Kuraray Co, Ltd for their provision of material "Septon".

References

[1] Gerald W. McRoberts and Catherine T. Best. Accommodation in mean f0 during mother-infant and father-infant vocal interactions: a longitudinal case study. *Journal of Child Language*, 24:719–736, 1997.

[2] Anne Fernald, Traute Taeschner, Judy Dunn, Mechthild Papousek, and Benedicte de Boysson-Bardies. A cross-language study of prosodic modifications in motehrs' and fathers' speech to preverbal infants. *Journal of Child Language*, 16:477–501, 1989.

[3] Patricia K. Kuhl, Jean E. Andruski, Inna A. Chistovish, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Viktoria L. Ryskina, Elvira I. Stolyarova, Ulla Sundberg, and Francisco Lacerda. Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277:684–686, 1997.

[4] Jordan R. Green, Lgnatius S. B. Nip, Erin M. Wilson, Antje S. Mefferd, and Yana Yunusova. Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, 53:1529–1542, 2010.

[5] Shigeaki Amano, Tomohiro Nakatani, and Tadahisa Kondo. Fundamental frequency of infants' and parents; utterancers in longitudinal recordings. *Journal of the Acoustical Society of America*, 119(3):1636–1647, 2005.

[6] T. G. Nienhuys, T. G. Cross, and K. M. Horsborough. Child variables influencing maternal speech style deaf and hearing children. *Journal of Communication Disorders*, 17:189–207, 1984.

[7] Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, 1(1):12–34, 2009.

[8] Kotaro Fukui, Yuma Ishikawa, Eiji Shintaku, Masaaki Honda, and Atsuo Takanishi. Anthropomorphic talking robot based on human biomechanical structure. In *Proceedings of the 3rd International Conference "SMART MATERIALS, STRUCTURES AND SYSTEMS"*, pages 153–158, 2008.

[9] Mitsuki Kitani, Tatsuya Hara, Hiroki Hanada, and Hideyuki Sawada. A taking robot for the vocal communication by the mimicry of human voice. In *Proceedings of the 3rd Conference on Human System Interactions*, pages 728–733, 2010.

[10] Yuichiro Yoshikawa, Junpei Koga, Minoru Asada, and Koh Hosoda. Primary vowel imitation between agents with different articulation parameters by parrot-like teaching. In *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 149–154, 2003.

[11] Joel C. Kahane. *Anatomy and Physiology of the Speech Mechanism*. Pro ed, 1986.

[12] Hans Edmund Eckel, Juergen Koebke, Christian Sittel, Georg Mathias Sprinzl, Claus Pototsching, and Eberhard Stennert. 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.

[13] Houri K. Vorperian, Shubing Wang, Moo K. Chung, E. Michael Schimek, Reid B. Durtschi, Ray D. Kent, Andrew J. Ziegert, and Lindell 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*, 125(3):1666–1678, 2009.

[14] Hideki Kasuya, Hisayoshi Suzuki, and Ken'iti Kido. Changes in pitch and first three formant frequencies of five japanese vowels with ahe and sex of speakers. *Journal of the Acoustical Society of Japan*, 24:355–364, 1968.