Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Integrated magnetic resonance imaging methods for speech science and technology

Shinobu Masaki[a], Yukiko Nota[a], Sayoko Takano[a], Hironori Takemoto[b], Tatsuya Kitamura[c] and Kiyoshi Honda[d]

[a]ATR-Promotions, 2-2-2 Hikaridai, Seika-cho Soraku-gun, 619-0288 Kyoto, Japan
[b]ATR Cognitive Information Science Laboratories, 2-2-2 Hikaridai, Seika-cho Soraku-gun, 619-0288 Kyoto, Japan
[c]Konan University, Okamoto 8-9-1, Higashinada, 658-8501 Kobe, Japan
[d]Phonetics & Phonology Laboratories, UMR-7018-CNRS & University of Paris III, 19, rue des Bernardins, 75005 Paris, France
masaki@atr.jp

The objective of this paper is to introduce magnetic resonance imaging (MRI) data acquisition techniques for speech science and technology integrated at the Brain Activity Imaging Center (BAIC) of the Advanced Telecommunications Research Institute International (ATR). The history of the development of the MRI technology for speech production studies can be divided into three stages as follows: (1) applying MRI techniques to some restricted speech samples that can be produced without articulatory movements, (2) exploiting the MRI movie data acquisition technique for observation of articulatory movements for a variety of speech samples, and (3) refining data acquisition methods and visualization techniques for analyzing dynamic characteristics of speech organs. We will introduce the background of the development, the principles of the techniques, the results obtained from the experiments using the technique, and future MRI techniques applicable to speech science and technology.

1 Introduction

Magnetic resonance imaging (MRI) is a powerful non-invasive technique for capturing two-dimensional (2D) and three-dimensional (3D) images of the speech production system. The first application of the MRI technique to speech production studies was performed in 1986 [1]. Since then, fruitful results and insights have been acquired from many research projects. These results have been obtained with the advancement of MR imaging techniques over more than twenty years. The developmental history can be divided into three stages in terms of the expansion of the dimension, modality and quality of the MR images of speech organs. Table 1 shows representative topics and summaries for each stage. In this table, listed in bold characters are results obtained from a series of speech production research projects at ATR.

The objective of this paper is to introduce the crucial MRI techniques for speech production studies developed at the Brain Activity Imaging Center (BAIC) of the Advanced Telecommunications Research Institute International (ATR). We will present the background of the development, the principles of the techniques, and the results obtained from the experiments using the developed technique. In the final section, expected future MRI techniques will be discussed for speech science and technology.

Table 1. Chronological table of MRI technology for speech production studies. Listed in bold characters are the results obtained from a series of speech production research projects at ATR.

1st stage "Application of MR still imaging technique"
1986: Rokkaku et al. [1] Sagittal still images for Japanese vowel production. 1987: Baer et al. [2] Sagittal still images for English vowel production. **1992: Honda [3]** **Sagittal still images for Japanese vowel production.** **1996: Masaki et al. [5]** **Sagittal still images for /r/ and /l/ by native English speakers and by native Japanese speakers.** ***Summary:*** *Observations were restricted in the speech samples, which could be produced without articulatory movements. MRI motion imaging was expected to observe articulatory movements during speech production.*
2nd stage "Application of MRI cine sequence"
1990: Foldvik et al. [6] First application of cardiac cine (synchronized scan) sequence to observation of articulatory movements. **1999: Masaki et al. [7]** **Synchronized sampling method using timing synchronizer.** **2000: Kumada et al. [8]** **Tagged MRI movie for observation of tongue deformation during CV sequence.** 2001: Stone et al. [9] Tagged cine-MRI for modeling internal tongue motion for CV sequence. ***Summary:*** *Methods for acquiring tomographic movie of articulators during speech production were established based on the conventional cardiac cine MRI technique. 3D visualization of the articulatory movements was expected.*
3rd stage "Refinement of MRI data acquisition and visualization techniques for speech production studies"
2003: Takano et al. [15] **Application of synchronized scan method and special coil for high-resolution still images of larynx.** **2004: Takemoto et al. [12]** **A method of tooth superimposition of MRI data.** **2005: Nota et al. [17]** **Sound presentation using bone-conduction speaker for synchronized scan method.** **2006: Takemoto et al. [14]** **Visualization of 3D MRI movie and extraction of vocal tract area function for vowel sequence.** ***Summary:*** *3D MRI movie acquisition technique was established. Development of data acquisition method without repetition of speech production is expected.*

2 1st stage: "Application of MR still imaging technique to speech production studies"

<u>First MRI application to speech production study</u>

In the early stage of MRI applications to speech production studies in the 1980s, speech materials were restricted to vowels and some consonants that can be produced without articulatory movements because of the long data acquisition time in MRI.

The earliest reports of the application of MRI to speech production study were reported by Rokkaku et al. [1] and Baer et al. [2]. In the report by Rokkaku et al., they took four images of sagittal slices of speech organs for the Japanese vowel /a/ produced by one male subject. According to the report, it took 25 seconds to acquire each slice of data with 1.2mm x 1.2mm resolution for 5mm thickness. Some edge lines of the MR images were ambiguous, probably due to the long acquisition time. They tried to extract vocal tract shape using the sagittal slice data.

MRI experiment at ATR

At ATR, we started speech production studies using MRI in 1991. We used a 1.0T MRI scanner (Shimadzu Corporation, SMT-100GUX) installed at Takanohara Chuo Hospital for vowel articulation studies. Honda took sagittal slice images for /a/, /i/ and /u/ out of five Japanese vowels [3]. Data acquisition time was more than 16 minutes for 16 slices with 4 times averaging for each vowel. The subject maintained the same posture of the articulators to produce each vowel during the entire data acquisition time. To do so, the subject breathed entirely through the oral cavity. Thanks to the performance by the subject and operators, the obtained images were clear enough to analyze the shapes of the articulators. As shown in Fig. 1, the deformation of the internal structure of the tongue can also be observed.

The static MRI was also applied to the observation of sustainable consonants, which are produced in a stable manner (fricatives [4], and liquids [5]). Masaki et al. applied the MRI to the observation of English /r/ and /l/ produced by native Japanese and native English speakers [5]. In this study, subjects were asked to keep the posture of the articulators stable. Although these consonants were sustainable enough, it was difficult to discuss the variation of tongue shapes in different phonetic environments due to the restriction of MRI data acquisition time.

Technical problems of MRI for speech production study

Results of these early MRI studies exposed some technical problems in the application of MRI to speech production studies as listed below: (1) insufficient temporal resolution to observe dynamic characteristics of articulatory movements, (2) insufficient spatial resolution to observe fine structures in the organs, such as those in the larynx, (3) invisibility of the teeth due to the poor contrast of signal intensity between the teeth and air, (4) the subject's unnatural supine position during MRI data acquisition, which might cause some effects of gravity on the position of the articulators. Especially, the acquisition of motion images was in high demand for the future application of MRI to speech production studies.

3 2nd stage: "Application of MRI cine sequence to movie data acquisition of articulators during speech production"

First MRI movie of articulatory movements

When the movie technique was thought to be a breakthrough for MRI-based speech production studies, especially for consonants, Foldvic et al. presented an "MRI film" showing a 1.5 s sequence of the midsagittal vocal tract composed of 55 frames of MR images [6]. They used the cardiac cine MRI technique to acquire the articulatory movements. The key technology in this method was to use a subject's heart beat signal obtained from a cardiograph system for synchronizing the repetitive utterances with MR scans in the raw data acquisition stage. In the following image reconstruction stage, the successive MR images are reconstructed from the raw data.

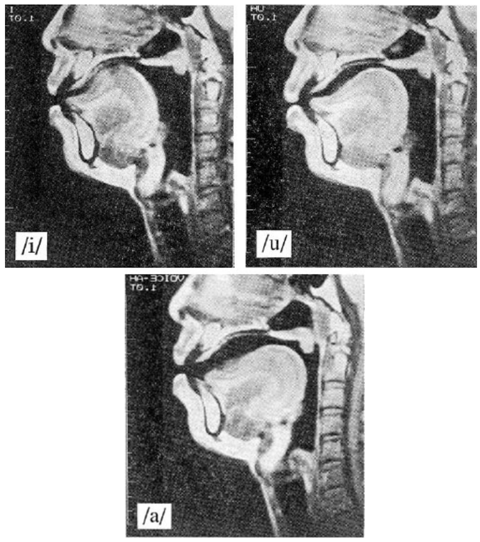


Fig. 1. Midsagittal T1-weighted MR images for Japanese vowels /i/, /u/ and /a/ obtained by a 1.0T MRI. Deformation of the internal structure of the tongue and difference in thickness of lips can be observed, while the teeth are invisible (from Honda [3]).

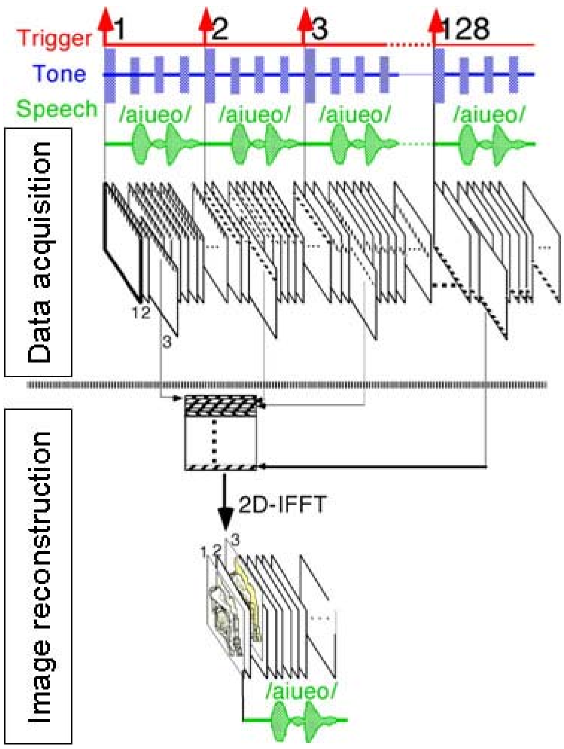


Fig. 2. Two-stage procedures for synchronized sampling method. The first stage is to acquire MR data during repetitive utterances (top), and the second stage is to transform the acquired data to sequential MR images (bottom).

MRI movie at ATR

Masaki et al. improved the synchronized sampling method by modifying the cardiac cine MRI [7]. They used a "timing synchronizer" instead of the cardiograph system to synchronize the utterances with MR scans. Figure 2 shows the schematic diagram of the synchronized sampling method. The procedure is composed of two parts: the "Data acquisition stage" and "Data reconstruction stage." At the first "Data acquisition stage," external trigger pulses were fed to the MRI scan controller 128 times to start each MR scan for acquiring raw data to reconstruct each frame. The trigger was generated by the "timing synchronizer," and the interval of the trigger was fixed to cover the period of the utterance to be analyzed. During this stage, the subject repeated the utterance 128 times so as to complete data acquisition. In order to synchronize the subject's utterance with the MR scans, tone bursts were presented through a loud speaker or earphones. The pattern of the tone bursts can be decided depending on the rhythm of the target word. In the "Data reconstruction stage," the entire set of raw data was used to reconstruct each frame of MR images. If the reconstructed MR images are displayed sequentially, articulatory movements can be observed as a movie.

Figure 3 shows a comparison of the tongue shape for /t/ production between (a) "dynamic condition" and (b) "sustained condition." Figure 3(a) shows the frame at the closure of the second /t/ in /tata/ obtained by the synchronized sampling method. Figure 3(b) shows the MR image of /t/ closure in /ata/ obtained by a "turbo-FLASH" sequence. In the turbo-FLASH sequence, the subject should keep the posture of /t/ closure for one second. The figure shows that the contact of the tongue against the palate in /t/ appears to be wider for the "sustained condition" than for the "dynamic condition," suggesting that articulation of the consonants is distorted in the "sustained condition" because the tongue contact had to be maintained at the place of articulation for one second in this condition. This result demonstrates an advantage of MRI movie imaging over the conventional static MR imaging for articulatory movement analysis.

Tagged cine MRI

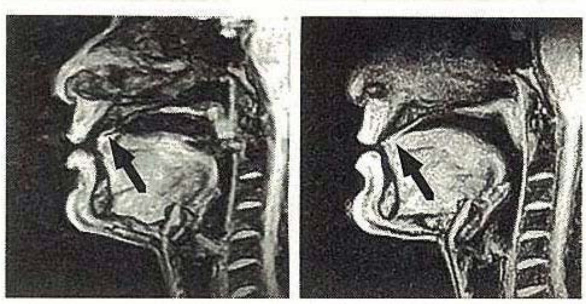
As an extension of this synchronized sampling method, the tagged cine MRI technique became popular for analyzing dynamic deformation of the internal structure of the tongue. In this technique, a lattice pattern is initially stamped in the MR images of the tongue, and the deformation of the lattice pattern can be used for the dynamic analysis of the internal structure of the organ [8][9][10]. Figure 4 shows an example of the tagged cine MRI for the Japanese vowel sequence /ei/ obtained with a Shimadzu-Marconi ECLIPSE 1.5T PowerDrive 250 installed at ATR.

4 3rd stage: "Refinement of MRI data acquisition and visualization techniques for speech production studies"

When the MRI movie acquisition technique was established, the next target was to obtain a 3D movie of the articulators with the tooth image in sufficient spatial resolution. What follow are our visualization techniques developed for speech production studies at ATR.

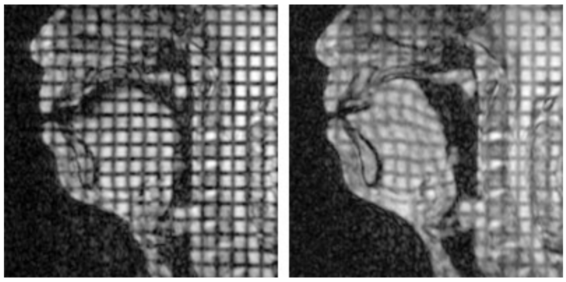
Superimposition of teech on the MR image

The first trial of tooth imaging by MRI was performed by Wakumoto et al. at ATR [11]. They used a pair of imaging plates containing contrast medium to visualize the dental crown. Although this method was successful at visualizing the teeth, the imaging plates were obstacles for articulation especially in some fricative consonants.



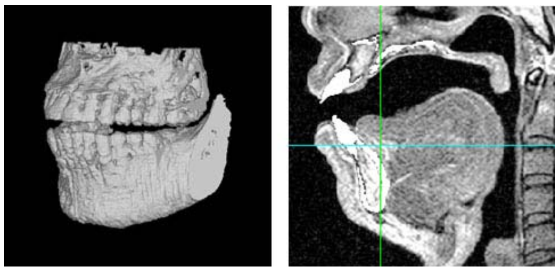
(a) "dynamic condition" (b) "sustained condition"

Fig. 3. Comparison of tongue shape for /t/ production between (a) "dynamic condition" (synchronized sampling method) and (b) "sustained condition" (turbo-FLASH). (from Masaki et al. [7])



(a) First frame for /ei/ (b) Last frame for /ei/

Fig. 4. Tagged cine MRI for Japanese vowel sequence /ei/. (a) The lattice pattern (1-cm interval) was stamped to the first frame of MR image. (b) Deformation of stamped lattice pattern can be used for the dynamic analysis of the internal structure of the tongue.



(a) Extracted teeth (b) Midsagittal MR image with the superimposed teeth

Fig. 5. Superimposition of the tooth image. (a) 3D tooth image extracted from MRI data obtained using contrast medium filled in the oral cavity. (b) Extracted tooth volume was superimposed based on position matching of the reference landmarks.

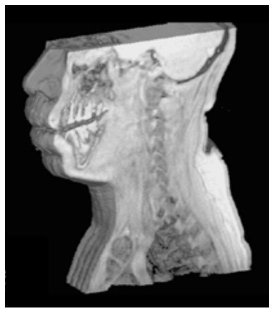


Fig. 6. 3D reconstruction of MRI movie. This is the first frame image of the MRI movie for Japanese vowel sequence /aiueo/, which was used for extraction of vocal tract area function shown in Fig. 7.

Takemoto et al. proposed another tooth imaging technique [12]. This method was divided into two steps. First, 3D tooth image was extracted from MRI data using a liquid contrast medium filled in the oral cavity. Second, extracted tooth volume was superimposed based on the position matching of the reference landmarks. Figure 5 shows an example of the tooth superimposition. Recently, Kitamura et al. proposed the use of a new flexible contrast medium for easier 3D tooth imaging [13].

3D MRI movie at ATR

The maximum temporal resolution of the reconstructed 2D MRI movie obtained by the MRI scanner installed at ATR (Shimadzu-Marconi ECLIPSE 1.5T PowerDrive 250) is 120 frames/s. If 30 frames/s is sufficient time resolution, we can obtain four slices of MRI movies. If 2D MRI movie data are accumulated for different slices, we can reconstruct and visualize a 3D MRI movie. Figure 6 shows an example of a 3D reconstructed MRI movie. It shows the first frame image of an MRI movie for Japanese vowel sequence /aiueo/, which is used for extraction of vocal tract area function shown in Fig. 7 [14].

Special coil for high spatial resolution MR imaging

Takano et al. proposed a high spatial resolution imaging technique [15][16]. They invented a special coil shown in the top panel of Fig. 8 (b). Together with this coil, they used a modified synchronized sampling method (called the "phonation-synchronized scan method") for sustained vowel production. In this method, the subject listened to the tone burst to pace the rhythm of phonation and respiration, while the MR scans were performed only in the phonation phases. This procedure could minimize the motion artifact due to respiration and/or swallowing, and provided clear MR images of the articulators for vowel production. Using these high-resolution MR images, they extracted the shape of the thyroid and cricoid cartilages of the larynx, and revealed rotation and translation characteristics of the joint for F0 control.

Figure 8 shows a comparison of performance between the conventional and special coils for detecting the fine structure of speech organs. The new coil resulted in the image with higher spatial resolution showing fine muscle structures of the lips and tongue.

Bone conduction transducer for MR imaging

Nota et al. proposed a sound presentation system using a bone conduction transducer together with high-insulation ear plugs for a synchronized sampling MRI movie and phonation-synchronized sampling static MRI [17][18]. This system improved the quality of MR images due to the higher intelligibility of the presented sounds and the availability of auditory feedback of the subject's own speech through enhanced bone conduction.

5 Expected MRI innovation

One of the crucial problems with our MRI movie technique for speech production studies is that a large number (around 100 or more) of repetitions of utterances is required to obtain an MRI movie with sufficient spatial resolution. It is a disadvantage when the MRI movie data acquisition is to be applied to a variety of subjects including patients with speech pathology.

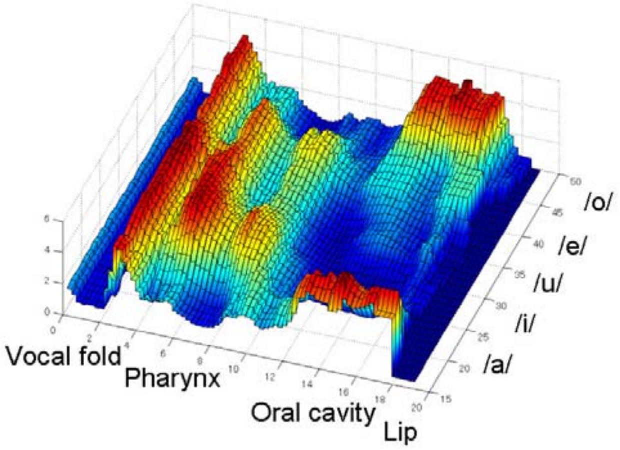
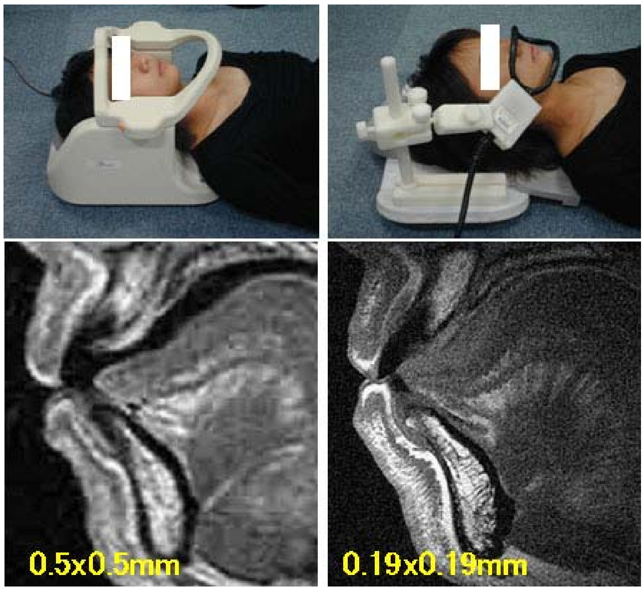
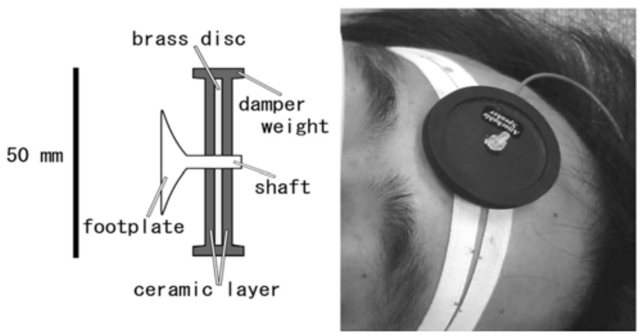


Fig. 7. Temporal changes in vocal tract area function for Japanese vowel sequence /aiueo/ extracted from 3D MRI movie data.



(a) Normal coil (b) Special coil

Fig. 8. Comparison of MR images between conventional and special coils. Top panels show the coils used to obtain the MR images shown at the bottom. The special coil provides MR image with higher spatial resolution.



(a) Piezoelectric transducer (b) Set-up of transducer

Fig. 9. Bone conduction sound transducer for high quality MR image acquisition with synchronized scans.

Some recently reported MRI movie techniques require a smaller number of repetitions of utterance or a single utterance. If they have sufficient spatial and/or temporal resolutions, they will contribute to the development of speech production study. A technical breakthrough might be established by recent higher field MRI scanners. At ATR, using a 3T MRI scanner with a 16-channel data acquisition system (Siemens, MAGNETOME Trio A Tim System) installed last September, we are trying to obtain MRI movies of more detailed structures of speech organ and laryngeal function for voicing and F0 controls.

6 Conclusion

MR imaging application techniques for speech production studies have been presented. There are many practical techniques developed by ATR and applied to speech production studies including a synchronized sampling method for MRI movies and some related techniques. Further MRI technical innovations are expected to contribute to speech science and technology using a higher field MRI scanner.

Acknowledgments

We thank Mr. Yasuhiro Shimada and Mr. Ichiro Fujimoto at BAIC for their technical support in MR imaging. This work was supported partly by a Grant-in-Aid for Science Research No. 18300069, Japan Society for the Promotion of Science.

References

[1] M. Rokkaku, K. Hashimoto, S. Imaizumi, S. Niimi and S. Kiritani, "Measurements of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique," *Ann. Bull. RILP*, 20, 47-54 (1986).

[2] T. Baer, J.C. Gore, S. Boyce, and P.W. Nye, "Application of MRI to the analysis of speech production," *Magn. Reson Imaging,* 5, 1-7 (1987).

[3] K. Honda, "Physiological background of speech production," *J. Acoust Soc. Jpn.*, 48, 9-14 (1992). (in Japanese)

[4] S.S. Narayanan, A.A. Alwan, and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am., 98, 1325-1347 (1995).

[5] S. Masaki, R. Akahane-Yamada, M.K. Tiede, Y. Shimada, and I. Fujimoto, "An MRI-based analysis of the English /r/ and /l/ articulations," *Proc. Int. Conf. Spoken Language Processing*, 1581-1584 (1996).

[6] A.K. Foldvic, O. Husby, J. Kvaerness, I.C. Nordlpi, and P.A. Rinck, "MRI (Magnetic Resonance Imaging) film of articulatory movements," *Proc. Int. Conf. Spoken Language Processing,* Vol. 1, 421-422 (1990).

[7] S. Masaki, M.K. Tiede, K. Honda, Y. Shimada, I. Fujimoto, Y. Nakamura, and N. Ninomiya, "MRI-based speech production study using a synchronized sampling method," *J. Acoust. Soc. Jpn. (E),* 20, 375-379 (1999).

[8] M. Kumada, S. Masaki, K. Honda, Y. Shimada, K. Mori, S. Niimi, "Function of Tongue-Related Muscles during Speech –A Tagging MRI Movie Study–," *Jpn. J. Logop. Phoniatr.*, 41, 170-178 (2000). (in Japanese)

[9] M. Stone, E.P. Davis, A.S. Douglas, M. NessAiver, R. Gullapalli, W.S., Levine, and A. Lundberg, "Modeling the motion of the internal tongue from tagged cine-MRI images," *J. Acoust. Soc. Am.,* 109, 2974-82 (2001).

[10] S. Takano and K Honda, "Measurement of internal deformation of the tongue during vowel sequence /ei/ using tagged-cine MRI," *Jpn. J. Logop. Phoniatr.*, 47, 283-290 (2006). (in Japanese)

[11] M. Wakumoto, S. Masaki, J. Dang, K. Honda, I. Fujimoto, Y. Nakamura, and Y. Shimada, "Dental crown imaging technique by MRI with imaging plate and its influence on articulation," *Proc. 1997 Autumn Meeting of Acoust. Soc. Jpn,* 397-398 (1997). (in Japanese)

[12] H. Takemoto, T. Kitamura, H. Nishimoto, K. Honda, "A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions," *Acoust. Sci. & Tech*. 25, 468-474 (2004).

[13] T. Kitamura, H. Hirata, K. Honda, I. Fujimoto, Y, Shimada, S. Masaki, T. Nishikawa, K. Fukui, and A. Takanishi, "A Method for Measuring Tooth Shape by Magnetic Resonance Imaging using a Thermoplastic Elastomer Dental Mouthpiece," *IEICE Technical Report*, MI2007-28 (2007). (in Japanese)

[14] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto, "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *J. Acoust. Soc. Am*, 119, 1037-1049 (2006).

[15] S. Takano, K. Honda, S. Masaki, Y. Shimada, I. Fujimoto, "High-resolution imaging of vocal gesture using a laryngeal MRI coil and a synchronized imaging method with external triggering," *Proc. 2003 Spring Meeting of Acoust. Soc. Jpn.*, Vol. I, 291-292 (2003) (in Japanese)

[16] S. Takano, K. Honda, and K. Kinoshita, "Measurement of cricothyroid articulation using high-resolution MRI and 3D pattern matching," *Acta Acustica united with Acustica,* 92, 725-730 (2006).

[17] Y. Nota, T. Kitamura, H. Hirata, K. Honda, "A bone-conduction system for auditory stimulation in MRI," *Proc. 2005 Spring Meeting of Acoust. Soc. Jpn.*, 385-286 (2005) (in Japanese)

[18] Y. Nota, T. Kitamura, K. Honda, H. Takemoto, H. Hirata, Y. Shimada, I. Fujimoto, Y. Shakudo, and S. Masaki, "A bone-conduction system for auditory stimulation in MRI," *Acoust. Sci. & Tech.,* 28, 33-38 (2007).