Improvement of five-degree-of-freedom sensors for Northern Digital Incorporated’s Wave speech research system

Tatsuya Kitamura1,*, Yukiko Noto2,3,4, Michiko Hashi5 and Hiroaki Hatano3,4

1Faculty of Intelligence and Informatics, Konan University, 8–9–1 Okamoto, Higashinada, Kobe, 658–8501 Japan
2Brain Activity Imaging Center, ATR-Promotions Inc., 2–2–2 Hikaridai, Keihanna Science City, 619–0288 Japan
3Research Center for Media and Culture, Graduate School of Intercultural Studies, Kobe University, 1–2–1 Tsurukabuto, Nada, Kobe, 657–8501 Japan
4ATR Intelligent Robotics and Communication Laboratories, 2–2–2 Hikaridai, Keihanna Science City, 619–0288 Japan
5Department of Communication Sciences and Disorders, Faculty of Health and Welfare, Prefectural University of Hiroshima, 1–1 Gaiken-cho, Mihara, Hiroshima, 723–0053 Japan

(Received 20 October 2014, Accepted for publication 27 November 2014)

Keywords: Electromagnetic articulograph, Wave speech research system of Northern Digital Inc., Articulatory movements

PACS number: 43.70.Jt [doi:10.1250/ast.36.347]

1. Introduction
Various techniques have been developed to measure articulatory movements that are difficult to observe from outside of the body. An electromagnetic articulograph (EMA) is one such technique, and it enables the tracking of the positions of small pellets or sensors placed inside the mouth or on the face. By this method, positions can be measured with high space and time resolutions, although the number of observation points is limited. The EMA is advantageous for speech studies because it does not generate noise during measurement, unlike magnetic resonance imaging.

An EMA that can be used to measure the position and orientation of transmitter coils in a 2D space was developed by Schöne et al. [1], and a 3D EMA was established in the late 1990s [2]. Articulatory movement data obtained using an EMA enjoy wide use in the fields of speech science and technologies, such as the analysis of coarticulation [3], speech therapy [4], estimation of articulatory movement from speech [5], and speech synthesis [6].

The Wave speech research system (hereafter, Wave system) of Northern Digital Inc. (NDI) is a type of EMA. It can track the positions of sensors, whose size is \( 3 \times 3 \times 2 \) mm\(^3\), in a magnetic field generated by a “field generator” at a maximum sampling rate of 400 Hz. The Wave system has a number of advantages over other EMAs: it is portable, allows subjects to adopt various body postures, and does not require calibration. These features are expected to expand the range of application of EMAs.

Berry [7] assessed the accuracy of the Wave system and provided detailed data on its performance. Yunusova et al. [8] proposed a method of acquiring the surface shape of the palate by calculation using tracing data of the palate obtained using the Wave system. The use of the system is expected to increase.

However, a few problems remain to be solved for the Wave system. First, the twisted pair of wires that connects a sensor and a sensor control unit (SCU) is not flexible and can hinder natural articulation. Second, it takes a long time to connect the sensor to the SCU because the wires must be connected to the SCU via a small terminal block. Third, the viewer software for the measured data installed in the system is not user-friendly and inadequate for the analysis. Recently, the third problem has been remedied somewhat; VisArtico [9], a viewer software, has been made compatible with data acquired by the Wave system. Thus, in the present study, we addressed the first and second problems. The twisted pair of wires was replaced with a thinner, more flexible one, and a connector was added to its end. We then evaluated the improved sensors by comparing it with the original.

2. Improvement of 5DOF sensors
Multiple five-degree-of-freedom (5DOF) sensors and a single six-degree-of-freedom (6DOF) reference sensor are generally used in the Wave system. Both types of sensor are covered with silicone resin and connected to a SCU via twisted pairs of wires. The 5DOF sensors are adhered to the surface of the tongue, lips, and alveolar ridge, for example, and their spatial locations and angles are tracked in real time. The 6DOF sensor is used for defining a local reference point in a measurement space and correcting for head movement.

In the present study, we made two improvements to the twisted pair of wires of the sensor, which was composed of two wires of 0.4 mm diameter. Because the wires are thick and stiff, articulatory movements are encumbered when multiple sensors are adhered to speech organs. We thus first replaced the sensor wires with thinner (0.16 mm diameter) and flexible enameled wires (Kyowa Harmonet Ltd., 2UEW 0.16 mm). The twisted pair of wires was cut at about 1 cm from the sensor, and the enameled wires were then soldered to the original wire. The joint was coated with a silicone tube for insulation.

*e-mail: t-kitamu@konan-u.ac.jp
Since the 5DOF sensors are disposable, users need to connect new sensors to the SCU prior to each experiment. This is, however, time-consuming because each sensor wire must be manually attached to a small terminal block using a screwdriver. Since each 5DOF sensor has a twisted pair of wires, users need to perform the task twice for each sensor. Also, the wires sometimes detach from the terminal block, disrupting the measurement. We thus added a male Molex connector to the end of the twisted pair of wires and replaced the terminal block with a female Molex connector. Figure 1 shows the ends of the sensor wires of the original and improved sensors.

When replacing the sensor wires, caution must be exercised to maintain the insulation property of the sensors and the joints. If not, the sensor often short-circuits in the mouth. It is also necessary to twist the pair of wires connected to the sensor to suppress electrical noise.

3. Evaluation

We compared the articulatory posture of a coronal consonant [c] in continuous speech measured using the original and improved sensors.

3.1. Method

3.1.1. Participants

Two male speakers (A and B) participated in recording sessions of articulatory movements and speech. Speakers A and B were in their 40s and 30s, respectively, and Japanese was their native language.

3.1.2. Adhering sensors

One 6DOF reference sensor was attached to the nasion, assuming that its relative position on the head was invariant during the experiment, and seven 5DOF sensors were adhered to positions on the midsagittal plane, as shown in Fig. 2.

To adhere the 5DOF sensors to the tongue and the incisors (T1, T2, T3, UI, and LI in Fig. 2), we used a medical superglue (Toagosei Co., Ltd., Aron Alpha). The position of T1 was roughly 10 mm posterior to the apex. T3 was glued to the position as far back as the participants could tolerate without eliciting the gag reflex. T2 was then placed about midway between T1 and T3. In our experience, the 5DOF sensors attached with medical superglue came off about one hour after gluing and do not damage the skin surface unless the participants forcibly peeled off the sensors.

The 6DOF reference sensor and the 5DOF sensors for the upper and lower lips (UL and LL) were adhered using cosmetic glue and surgical tape.

3.1.3. Data acquisition

The articulatory movements of the speakers were measured on different days using the original and improved 5DOF sensors. The relative positions of the 5DOF sensors to the 6DOF reference sensor were acquired at a sampling rate of 400 Hz. The volume of the electromagnetic field generated by the field generator was $300 \times 300 \times 300 \text{ mm}^3$. We asked the speakers to maintain a distance of less than 200 mm between their midsagittal plane and the field generator on the basis of evaluation results of the accuracy of the system [7]. Speech data were recorded in synchrony with the articulatory data using a condenser microphone (Sony ECM-77B) and a USB sound interface (M-Audio Fast Track) at a sampling rate of 22.05 kHz and 16-bit resolution.

The speakers uttered the beginning of a Japanese folk story “Momotaro” with the original and improved sensors, and the measurements were conducted every 10 min immediately after sensor attachment. The duration of speech data was roughly from 12 to 20 s. The articulatory and speech data were obtained three times: that is, 0, 10, and 20 min after sensor attachment. During intervals between the measurements, the participants were asked to converse with the experimenters to facilitate adaptation to the sensors.

The midsagittal shape of the palate of the participants was also measured using a pen-style 6DOF sensor or palate probe. The speakers held the palate probe and traced the midsagittal plane of their own palate.

The experimental protocol was approved by the ethical and safety committees of Konan University.

3.2. Results

3.2.1. Effects of adhering sensors

The connectors made it possible to set up the Wave system quickly and easily. We could connect the improved 5DOF sensors to the SCU in a few seconds and they did not disconnect during the measurement.

3.2.2. Articulatory effects

Figures 3 and 4 depict the midsagittal positions of the 5DOF sensors for speakers A and B, respectively, at the
timing when T1 was closest to the palate during the production of [c]. The palate shape measured using the palate probe is superimposed on the plots as a black line. The anterior is towards the left in the figures.

In the third measurement using the original sensors, the distance between T1 and the palate was the smallest for both speakers. The distances calculated in 3D space were 5.1 mm, 4.9 mm, and 1.7 mm for the first, second, and third measurements of speaker A, and 5.1 mm, 2.5 mm, and 1.4 mm for the first, second, and third measurements of speaker B, respectively. These results indicate that the articulatory postures and movements approached the natural ones with time.

Conversely, the T1 positions measured using the improved sensors were almost constant and close to the palate, indicating that the speakers could speak naturally immediately after adhering the sensors. The speakers reported that the degree of difficulty they felt in speaking with the original wires was markedly decreased when using the improved sensors.

4. Conclusions

We improved the 5DOF sensors of the Wave system. The proposed sensors allowed speakers to speak naturally immediately after attaching the sensors. Also, the connectors added to the sensor wires enabled the experimenters to easily connect the sensors to the SCU. As a result, the improved sensors can shorten the setup time and lengthen the data collection time.

Acknowledgments

This study was supported by JSPS Kakenhi (Nos. 24652085, 25240026, 25280066, and 25884099). The authors wish to thank Dr. Shinobu Masaki of ATR-Promotions Inc., Mr. Minoru Nishitani of Advanced Systems Co. Ltd., and Mr. Okihiro Kobayashi of Kobatel Co. Ltd. for their kind assistance.

References

[1] P. W. Schönle, K. Graebe, P. Wenig, J. Höhne, J. Schrader and B. Conrad, “Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract,” Brain Lang., 31, 26–35 (1987).
[2] A. Zierdt, P. Hoole, M. Honda, T. Kaburagi and H. G. Tillmann, “Extracting tongues from moving heads,” Proc. 5th Semin. Speech Production, pp. 313–316 (2000).
[3] J. Dang, M. Honda and K. Honda, “Investigation of coarticulation in continuous speech of Japanese,” Acoust. Sci. & Tech., 25, 318–329 (2004).
[4] W. F. Katz and M. McNeil, “Studies of articulatory feedback treatment for apraxia of speech (AOS) based on electromagnetic articulography,” Perspect. Neurophysiol. Neurogenetic Speech Lang. Disord., 20, 73–80 (2000).
[5] S. Hiroya and M. Honda, “Estimation of articulatory movements from speech acoustics using an HMM-based speech production model,” IEEE Trans. Speech Audio Process., 12, 175–185 (2004).
[6] Z.-H. Ling, K. Richmond, J. Yamagishi and R.-H. Wang, “Integrating articulatory features into HMM-based parametric speech synthesis,” IEEE Trans. Audio Speech Lang. Process., 17, 1171–1185 (2009).
[7] J. J. Berry, “Accuracy of the NDI Wave speech research
system,” *J. Speech, Lang. Hear. Res.*, **54**, 1295–1301 (2011).

[8] Y. Yunusova, M. Baljko, G. Pintilie, K. Rudy, P. Faloutsos and J. Daskalogiannakis, “Acquisition of the 3D surface of the palate by in-vivo digitization with Wave,” *Speech Commun.*, **54**, 923–931 (2012).

[9] S. Ouni, L. Mangeonjean and I. Steiner, “VisArtico: A visualization tool for articulatory data,” *Proc. Interspeech 2012* (2012).