

1-1-2012

Tracking Articulator Movements Using Orientation Measurements

An Ji

Marquette University, an.ji@marquette.edu

Michael T. Johnson

Marquette University, michael.johnson@marquette.edu

Jeffrey J. Berry

Marquette University, jeffrey.berry@marquette.edu

Accepted version. © 2012 IEEE. Reprinted, with permission, from An Ji, Michael T. Johnson and Jeffrey Berry, "Tracking Articulator Movements Using Orientation Measurements," *Proceedings of the Third International Conference on Audio, Language, and Image Processing*, 2013.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of Marquette University's products or services. Internal or personal use of this material is permitted.

However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Tracking articulator movements using orientation measurements

An Ji

*Department of Electrical and Computer Engineering
Speech and Signal Processing Laboratory
Milwaukee, WI*

Michael T. Johnson

*Department of Electrical and Computer Engineering
Speech and Signal Processing Laboratory
Milwaukee, WI*

Jeffrey Berry

*Department of Electrical and Computer Engineering
Speech and Signal Processing Laboratory
Milwaukee, WI*

This paper introduces a new method to track articulator movements, specifically jaw position and angle, using 5 degree of freedom (5 DOF) orientation data. The approach uses a quaternion rotation method to accomplish this jaw tracking during speech using a single sensor on the mandibular incisor. Data were collected using the NDI Wave Speech Research System for one pilot subject with various speech tasks. The degree of jaw rotation from the proposed approach is compared with traditional geometric calculation. Results show that the quaternion based method is able to

describe jaw angle trajectory and gives more accurate and smooth estimation of jaw kinematics.

Introduction

The tracking of jaw movement tracking speech is of significant interest to a large number of disciplines, including occupational therapy, orthodontics, psychology, speech language pathology and acoustic speech processing. A number of technologies have been used for recording articulator movements. X-ray cinematography [1][2] is effective, but the radiation to the subject's head is a concern. Cine MRI can provide dynamical 3D measurement of the vocal tract but it is cumbersome and expensive [3][4]. In contrast, the ultrasound technique is able to capture the surface of the tongue [5][6] but noise, echo artifacts and refractions may affect the results. High speed optical tracking is an attractive option for jaw movements but this approach requires much more complex reconstruction algorithms, and the inaccessibility of mandible is a challenge. Recently, electromagnetic articulography (EMA) sensing systems have been developed to measure articulatory movements. This system is able to track both position and orientation of the sensors, which can include both 5 DOF (planar orientation) and 6 DOF (full orientation). This enables detailed study of the relationship between acoustic data and articulator movements. However, most of the published work in acoustic and articulator inversion [7][8][9] only utilizes the 3D position data to extract articulator movement features. The orientation information consists of quaternion data and has thus far not been incorporated in most kinematic EMA studies.

Jaw movement during speech has generally been described as a pure translation and pure rotation of a single point on the jaw. Westbure and Edwards develop a two-dimensional rigid-body model of jaw movement during speech [10][11]. This model decomposes speech-related jaw movements into three components: rotation degree around the transverse axis located approximately through the condyles, and horizontal and vertical translation of this axis in the mid-sagittal plane. The purpose for tracking jaw movement here is to find appropriate features to describe the degree of jaw openness for the acoustic to articulatory mapping, with a focus on the rotation angle of the jaw in the following experiments.

Proceedings of the Third International Conference on Audio, Language, and Image Processing, Shanghai, July 16-18, 2012, (2012): pg. 292-296. DOI. This article is © Institute of Electrical and Electronics Engineers (IEEE) and permission has been granted for this version to appear in e-Publications@Marquette. Institute of Electrical and Electronics Engineers (IEEE) does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Institute of Electrical and Electronics Engineers (IEEE).

This paper is a preliminary study of applying quaternion orientation data to the task of jaw movement tracking during speech. The kinematics data of jaw movements were recorded by the NDI Wave Speech Research System which is a motion capture system specifically designed for tracking speech related articulatory or ofacial movements. As noted, the kinematic data stores both position and orientation information for each sensor attached on articulators. Since the 6DOF sensors are significantly bulkier and can thus create articulation difficulty, it is beneficial to be able to use 5DOF sensors to capture the necessary orientation information.

Quaternion orientation representation

In computer visualization and animation, quaternion format is a commonly used method to represent rotation and orientation [12]. A quaternion is a 4-D unit vector $q = [q_0, q_x, q_y, q_z]$ satisfying the following equation:

$$q_0^2 + q_x^2 + q_y^2 + q_z^2 = 1 \quad (1)$$

A quaternion rotation thus lies on the 4-D unit hyper-sphere. The key application of quaternions to tracking jaw movement during speech in this paper lies in their use to represent rotations. A quaternion can represent a rotation by an angle around an unit axis v .

$$q = \left(\cos\left(\frac{\theta}{2}\right), \sin\left(\frac{\theta}{2}\right)v_x, \sin\left(\frac{\theta}{2}\right)v_y, \sin\left(\frac{\theta}{2}\right)v_z \right) \quad (2)$$

where the vector part $\sin\left(\frac{\theta}{2}\right)v = [q_x, q_y, q_z]$ defines the rotate axis direction, and the scalar part $\cos\left(\frac{\theta}{2}\right) = q_0$ defines the degree of rotation. To rotate a point whose position is represented by the vector \vec{p} by an angle θ along the axis v to a new position p_{final} we apply the quaternion multiply operation

$$p_{final} = PQQ^* \quad (3)$$

Where $Q = \left[\cos \frac{\theta}{2}, \sin \left(\frac{\theta}{2} \right) v \right]$, $P = [0, \bar{p}]$

In the EMA system, each sensor's orientation is represented at each data point by a single quaternion vector, representing the rotation that would be required to move the sensor from its original baseline orientation to its current orientation. In these experiments the quaternion data is used to calculate jaw movement by tracking the sensor's orientation change over time, to determine jaw angle. The global coordinate system is referenced to a fixed head orientation with origin at the front upper incisors. The mid-sagittal plane is the x-y plane in the system. The 5DOF sensor's base orientation is parallel to the x-y plane, so in this orientation the quaternion data shows no rotation. The initial sensor's plane can be defined as two orthogonal base vectors $[1 \ 0 \ 0]$ and $[0 \ 1 \ 0]$. The quaternion data can be applied to these base vectors to calculate the final orientation position of the sensor plane at any sampling time.

Method

System set up

In the **NDI** Wave Speech Research System, 5-DOF sensors allow tracking of x, y, and z spatial coordinates, as well as angular coordinates characterizing rotation about the transverse axis (pitch) and anterior-posterior axis (roll). 6-DOF sensors have the added capacity for tracking angular coordinates characterizing rotation about the inferior-superior axis (yaw). The standard **NDI** Wave (used for the current work) has eight input channels and records sensor movements with a 400-Hz sampling rate. Figure1 shows the system set up and the subject's head orientation relative to the field generator. A subject sits at the right side of the field generator within the range of the magnetic field.

Figure 1 Coordinate system direction and sensor placement



Figure 2 Bite-plate with sensor positions



Sensor placement and data collection

In order to locate the maxillary occlusal plane which will form the x-z plane of the local coordinate system, a calibration bite-plate recording is done initially.

One 6DOF sensor is attached at the center of forehead and two 5DOF sensors are placed on the bite-plate, one at the maxillary central incisors (OS) and one along the mid-sagittal plane at the bisection between the back molars (MS) (See figure below for more detail). To create the bite-plate, two pieces of wax were softened in warm water and molded onto a tongue depressor. This softened wax is then placed into the subject's mouth and an impression of the bite is taken. Immediately afterward (while the wax is still soft), experimenters measure the midpoint between the back molars and create an indentation for the placement of the MS sensor. The OS sensor is

Proceedings of the Third International Conference on Audio, Language, and Image Processing, Shanghai, July 16-18, 2012, (2012): pg. 292-296. DOI. This article is © Institute of Electrical and Electronics Engineers (IEEE) and permission has been granted for this version to appear in e-Publications@Marquette. Institute of Electrical and Electronics Engineers (IEEE) does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Institute of Electrical and Electronics Engineers (IEEE).

placed directly in front of the central maxillary incisors. These sensors are pressed into the wax until they are secured and the bite-plate is replaced in the subject's mouth for the bite-plate recording.

In normal recording, the bite-plate wax is taken out and 5DOF sensors are placed at the desired recording points. For this experiment a simple configuration was used with two 5DOF sensors placed on the mandibular incisor and back molar. The 6DOF sensor remains attached in the same forehead position which is used to as a reference to calibrate all the other sensors data to eliminate the error of heat movement. The two 5DOF sensors are used for tracking jaw movement.

Speech materials

Initially the subject was asked to do wide range jaw motion. This large scale jaw motion repetition is used to evaluate the proposed rotation calculation method based on quaternions. In the reiterant speech task, the subject repeat the target three vowels [i], [a], [ea] 4 times each. The selection of vowels creates a wider range of movement amplitude. Of these, the vowel [ea] is associated with the largest amplitude jaw motion, whereas the vowel [i] is associated with smallest amplitude motion. The final speech task is a short word and sentence repetition using a normal speed pattern.

Experiment 1 (baseline): Calculate the jaw angle by using the position of two 5DOF sensors and the bite plate record

Calculating the jaw angle from the position data of the above two 5DOF sensors during normal recording requires straightforward geometric manipulation. In the bite-plate recording the position of OS and MS can be located by taking the average value of x, y, z coordinate of OS and MS sensors. The two points OS and MS forms vector $msos$. During normal recording, the mandibular incisor (MI) and back molar (BM) sensors' position also gives a vector $bmmi$. The jaw angle is calculated using the following equation

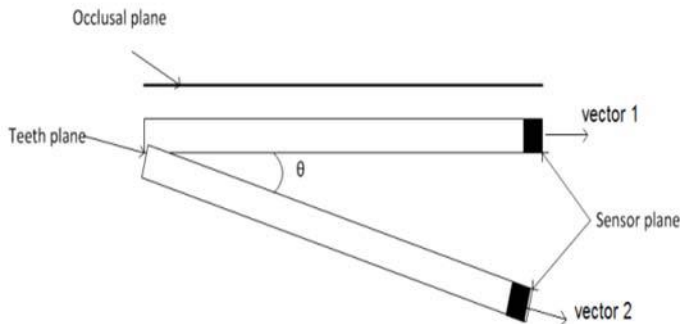
$$\theta = \frac{\langle \overline{moso}(x, y), \overline{bmmi}(x, y) \rangle}{|\overline{moso}(x, y)| |\overline{bmmi}(x, y)|} \quad (4)$$

where $\overline{moso}(x, y)$ represents the vector's projection in the $x-y$ plane, and represents $(\overline{moso}, \overline{bmmi})$ the dot product of these two vectors. The magnitude of lateral motion of the jaw in this task is not considered in these experiments.

Experiment 2: Calculate the jaw angle using a single sensor on the mandibular incisor, using quaternion orientation information

Since teeth are rigid, the sensor orientation changes during jaw movements, in direct correspondence to the angle of the jaw. This experiment uses the MI sensor to calculate the jaw angle by using the quaternion orientation data. Figure 3 illustrates the approach for the sensor attached on the front lower incisor. For the initial position, which is always the jaw closed condition, the sensor's quaternion data is used to calculate an original orientation position. The same method is used to obtain the jaw orientation over time, for each data point. *vector1* and *vector2* in Figure 3 represent the normal vectors perpendicular to the initial and final sensor's plane respectively. The angle between these two vectors is the jaw rotation angle.

Figure 3 x-y plane projection of the jaw angle calculation



The initial and final sensor plane can be defined by vector s_{i1}, s_{i2} and s_{f1}, s_{f2}

$$s_{i1} = q_i \overline{\text{base1}} q_i^* \quad (5)$$

$$s_{i2} = q_i \overline{\text{base2}} q_i^* \quad (6)$$

$$s_{f1} = q_f \overline{\text{base1}} q_f^* \quad (7)$$

$$s_{f2} = q_f \overline{\text{base2}} q_f^* \quad (8)$$

where q_i and q_f is the quaternion orientation of the initial and final position, respectively. $\text{base1} = [1], [0] [[0], [0] [[0]]$ and $\text{base2} = [0] [[0]], [1], [0] [[0]]$ are the vectors defining the plane, as previously introduced.

Following this, vector1 and vector2 , the normal vectors perpendicular to the initial and final sensor plane are determined:

$$v1 = s_{i1} \times s_{i2} \quad (9)$$

$$v2 = s_{f1} \times s_{f2} \quad (10)$$

where \times represents cross product of two vectors. The angle is then calculated by

$$\alpha = \frac{\langle v1(x, y), v2(x, y) \rangle}{|v1(x, y)| |v2(x, y)|} \quad (11)$$

where $v1(x,y)$ and $v2(x,y)$ represent the vector's projection in the $x-y$ plane.

Result and discussion

In this section, we present the mid-sagittal plane jaw motion paths and evaluate jaw rotation by both the baseline and new method. The goal is to assess the degree of jaw rotation in different speech behaviors and compare the quaternion method with the traditional geometrical approach. Figure 4 shows a typical jaw angle trajectory for the initial wide jaw movement pattern. The paths of the sensors attached to the mandibular incisor and back molar teeth display the jaw motion. From this plot, the maximum angle of jaw openness can be estimated as approximately 14 degrees.

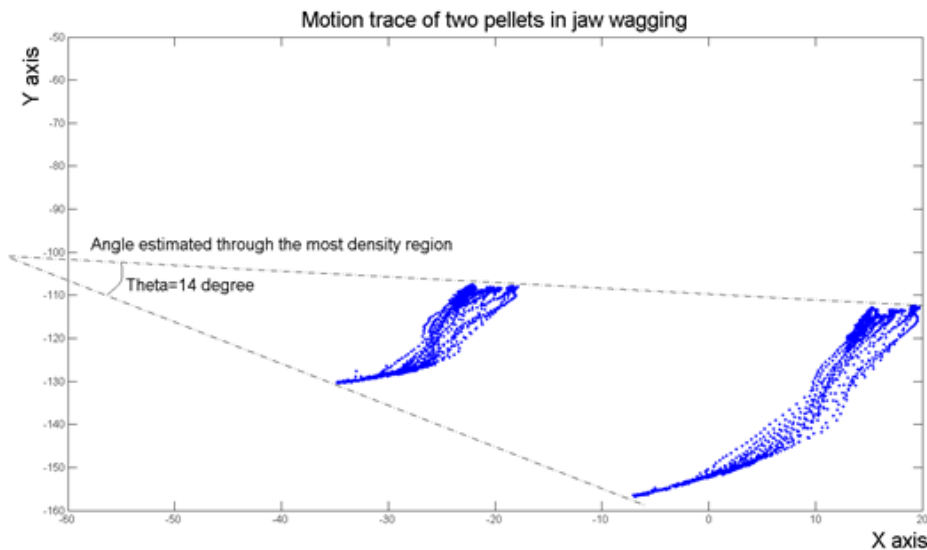


Figure 4 jaw trace for large range movement

Figure 5 shows the corresponding jaw angles. The top plot is calculated from one sensor using quaternion data, and the bottom plot is calculated from two sensors using position data only. These peaks are associated with five jaw movement repetitions. The smoothness and consistency of the quaternion-based plot indicate that the new approach is effective at tracking the motion of the jaw over time.

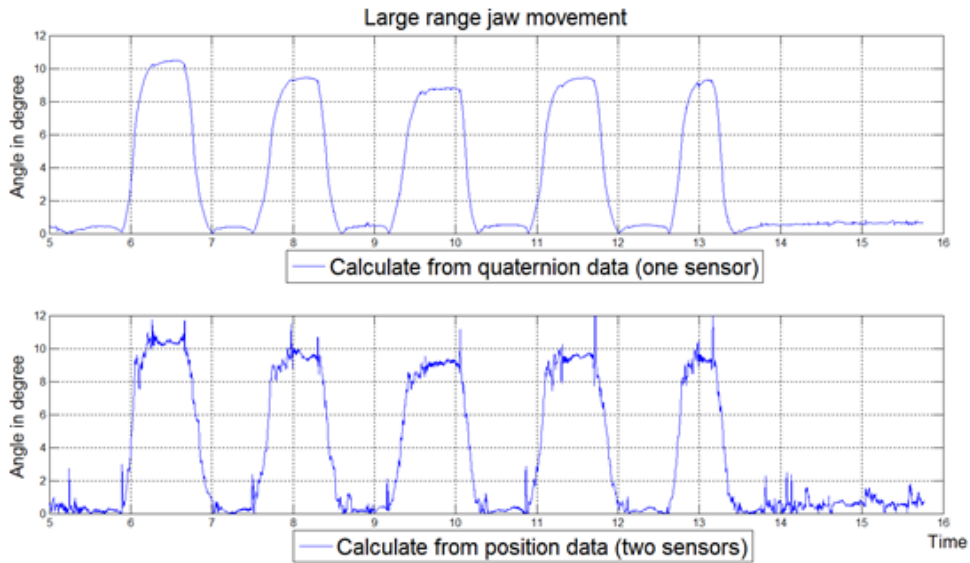


Figure 5 Rotation angle for large range jaw movement

In order to further evaluate the quaternion method, Figures 6, 7, and 8 show jaw movement for different speech tasks: vowels, words and sentences repetition.

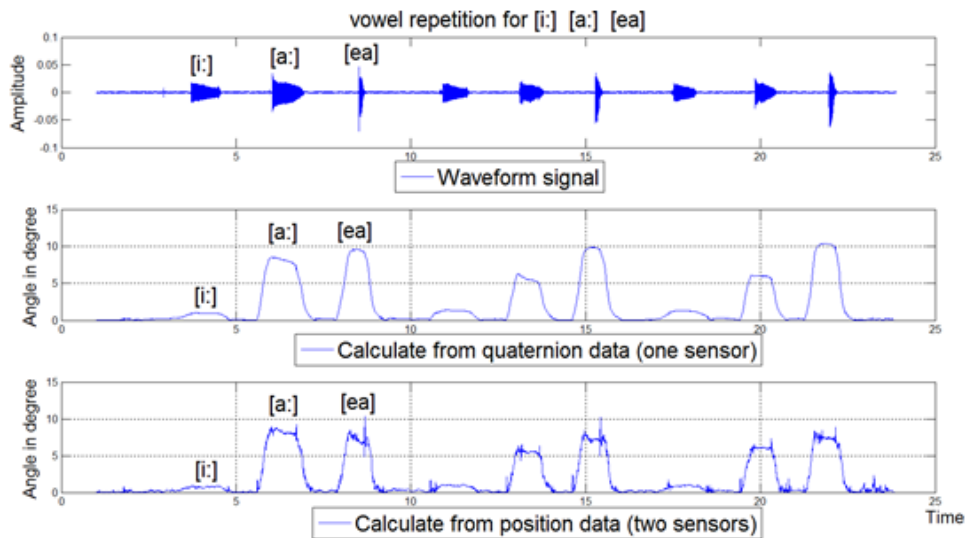


Figure 6 Rotation angle for vowel repetition

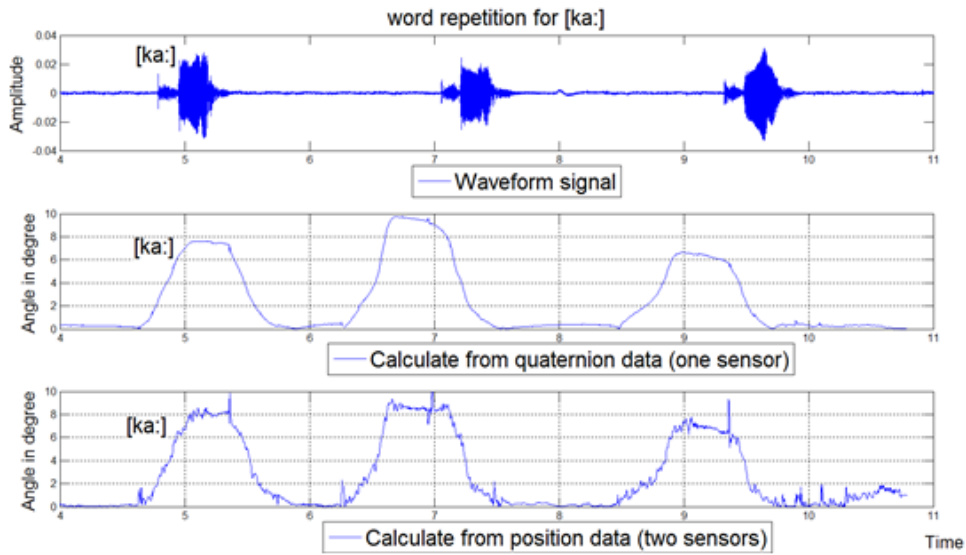


Figure 7 Rotation angle for word repetition

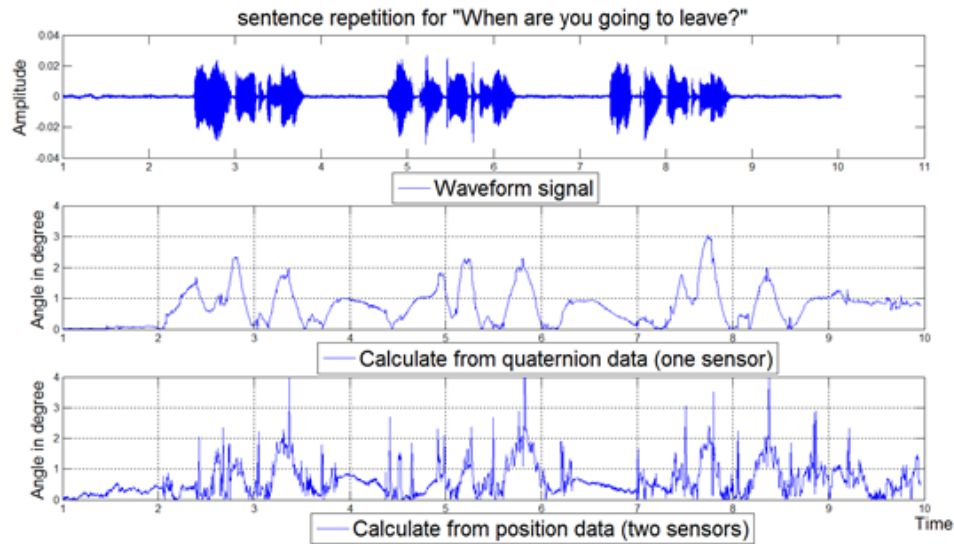


Figure 8 Rotation angle for sentence repetition

By carefully examining and comparing the angles from these two methods, the amplitude of the jaw openness does not match precisely.

From these plots, we can see that the two methods return similar results. There is a small deviation in peak value, as well as a significant difference in terms of the smoothness of the angle contours, with the quaternion approach yielding a much smoother

Proceedings of the Third International Conference on Audio, Language, and Image Processing, Shanghai, July 16-18, 2012, (2012): pg. 292-296. DOI. This article is © Institute of Electrical and Electronics Engineers (IEEE) and permission has been granted for this version to appear in e-Publications@Marquette. Institute of Electrical and Electronics Engineers (IEEE) does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Institute of Electrical and Electronics Engineers (IEEE).

trajectory than the position-based method. This may be due to the use of two sensor measurements rather than one, with two separate error sources, or alternatively due to differences in measurement error range for position versus orientation data from the sensors.

To summarize, we have shown that jaw movements can be accurately estimated using orientation data from a single sensor on the mandibular incisor, and that this approach gives a more accurate description of jaw openness during speech. This approach is suitable as a feature to characterize jaw motion in future applications, such as acoustic-to-articulatory inversion. Continued work in this area includes system calibration based on bite plate orientation and calculation of other articulatory measures from the position and orientation data.

This research is supported by grant IIS-1141826 from the National Science Foundation.

References

1. R. Houde. "A Study of Tongue Body Motion during Selected Consonant Sounds," PhD thesis, University Microfilms, Ann Arbor, Mich. 1967.
2. K.G. Munhall, E. Vatikiotis-Bateson, and Y. Tohkura. "X-ray film database for speech research," *Journal of the Acoustical Society of America*, pp. 1222-1224, 1998.
3. Masaki, S., Tiede, M.K., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., and Ninomiya, N., "MRI-based speech production study using a synchronized sampling method," *Journal of the Acoustical Society of America*, vol. 20, no. 5, pp. 375-379, 1999.
4. Narayanan, S., Nayak, K., Lee, S., Sethy, A., and Byrd, D., "An approach to real-time magnetic resonance imaging for speech production," *Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1771-1776, 2004.
5. Stone, M.L., Sonies, B.C., Shawker, T.H., Weiss, G., and Nadel, L., "Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system," *Journal of Phonetics*, vol. 11, no. 3, pp. 207-218, 1983.
6. Kaburagi, T. and Honda, M., "An ultrasonic method for monitoring tongue shape and the position of a fixed-point on the tongue surface," *Journal of the Acoustical Society of America*, vol. 95, no. 4, pp. 2268-2270, 1994.

- 7.** Richmond, K., "Estimating articulatory parameters from the acoustic speech signal". Edinburgh, Doctor of Philosophy, University of Edinburgh, 2001.
- 8.** Toda, T., A. Black and K. Tokuda, "Acoustic-articulatory inversion mapping with gaussian mixture model". International Conference on Spoken Language Processing, Jeju Island, Korea, 2004.
- 9.** S. Hiroya and M. Honda, "Estimation of articulatory movements from speech acoustics using an HMM-based speech production model," IEEE Transaction of Speech Audio Process. 12 (2) , pp. 175- 185, 2004.
- 10.** Westbury, J. R. "Mandible and hyoid bone movements during speech," Journal of Speech and Hearing Research, 31, pp. 405-416. 1988.
- 11.** Edwards J. "Mandibular rotation and translation during speech," Doctoral dissertation, City University of New York, 1985.
- 12.** John C. Hart, George K. Francis, Louis H. Kauffman, "Visualizing quaternion rotation," ACM Transactions on Graphics, v.13 n.3, pp. 256-276, 1994.

IEEE Keywords

Speech, Quaternions, Tracking, Vectors, Sensor phenomena and characterization, Acoustics