INTERSPEECH 2014

Contribution of Tongue Lateral to Consonant Production

Jun Wang [1, 2], William F. Katz [2], Thomas F. Campbell [2]

[1] Department of Bioengineering
[2] Callier Center for Communication Disorders
University of Texas at Dallas, Dallas, Texas, USA

{wangjun, wkatz, thomas.f.campbell}@utdllas.edu

Abstract

Speaking requires the coordinated movements of individual articulators. Understanding each articulator's contribution to speech is fundamental not only for understanding how speech is produced, but also for optimizing speech assessment and treatment. Our recent work has studied the individual contributions of tongue tip, tongue blade, tongue body front, tongue body back, upper lip, and lower lip movement to speech sound production by tracking the motion of sensors attached on the midline of tongue and lips. An optimal set of articulators (tongue tip, tongue body back, upper lip, and lower lip) has been found. However, the tongue lateral (side)'s contribution to speech is still poorly understood. We therefore investigated the contribution of the tongue lateral region to consonant production by analyzing the motion of a sensor attached to the side of tongue. Repeated productions of 12 consonants (including the lateral approximant /l/) were collected from six native English speakers. Consonant classification accuracy based on articulatory movement data obtained using a support vector machine was used as an indication of contribution level. The results suggest that sagittal movement of the tongue lateral sensor did not significantly benefit consonant classification, over and above the optimal set. Implications of these findings are discussed.

Index Terms: speech production, consonant production, tongue lateral, support vector machine

10.21437/Interspeech.2014-47

1. Introduction

Although most people speak effortlessly under everyday conditions, the underlying coordination required to produce fluent speech is very complex, involving dozens of muscles. How exactly individual articulators contribute to speech production is poorly understood [1]. One major barrier to speech production research has been the logistic difficulty of tongue motion data collection [2]. Fortunately, recent advances in electromagnetic tracking devices have made speech production data collection more feasible. Tongue motion tracking using electromagnetic articulograph (EMA) technology is accomplished through the placement of small sensors on the surface of the tongue.

In prior work, the location and number of tongue sensors used in EMA studies have been justified partly upon long-standing assumptions about tongue movement patterns. For the majority of these studies, this involves three to four sensors attached along the midsagittal line of the tongue, plus lips and jaw [2-10]. Qin and colleagues showed that three to four sensors are able to predict the tongue contour with only 0.3-0.2 mm error per point on the tongue surface [11]. However, it is unclear how many lingual sensors (and which locations) are best for a particular study because an individual articulator's contribution to speech production has rarely been studied [12]. It is generally assumed that the sensors should be placed along the tongue midline in order to track vertical and anterior-posterior (front-back) movement, as the left-right movement is not significant in normal speech production [2, 3].

Understanding the contribution of individual articulators to speech has implications not only for speech science, but also for speech assessment and treatment, including (1) optimizing silent speech interfaces for assisting the oral communication of persons after laryngectomy (surgical removal of larynx due to treatment of cancer) or with severely impaired voice [13-16], (2) improving the accuracy of (acoustic) speech recognition with articulatory information [17, 18], and (3) providing a reference for the use of sensors in speech therapy using visual feedback of speech movements [19, 20]. In addition, the use of more sensors than is necessary comes at a cost for both investigators and participants for studies using EMA; the procedure for attaching sensors to the tongue is time intensive and can cause discomfort and, therefore, may limit the scope of research on persons with speech impairment [12].

Wang and colleagues recently examined the individual contribution of six articulation points ("articulators" for the rest of the paper), tongue tip, tongue blade, tongue body front, tongue body back, upper lip, and lower lip to the articulatory distinctiveness of vowels and consonants [12]. All tongue sensors were attached on the midline of the tongue [12]. An optimal set of articulators (i.e., tongue tip, tongue body back, upper lip, and lower lip) was found [12].

The contribution of tongue lateral sensor information to speech has not been explicitly described in EMA studies, although the movement of tongue lateral is likely important for consonant production, especially for lateral consonants, with evidence provided in studies using 3D MRI [21, 22], electropalatography [21], electromyography [23], and palatography [24]. The use of lateral sensors could conceivably provide information about the convex/concave shape of the tongue, critical for particular sounds. For example, American English /l/ involves lateral air currents [25-28] and /ɹ/ has been described as having retracted anterior lateral tongue features [21].

In this paper, we investigated the contribution of tongue lateral sensor's movement to consonant production. To our knowledge, this is the first EMA study to analyze the contribution of tongue lateral region in speech production. Consonant production data were collected by attaching a sensor to the side of tongue. To compare our data with that of a previous study [12], this first series of analyses focused on movement in the sagittal plane (vertical and anterior-posterior movement) only. Similarly to [12], support vector machine (SVM, [29]) was used to classify those consonants from the data of individual or combined articulators. The resulting classification accuracies were used to analyze the contribution level of those individual articulators, including tongue lateral sensor. The hypothesis was that inclusion of the tongue lateral sensor information would benefit consonant classification,

14 – 18 September 2014, Singapore

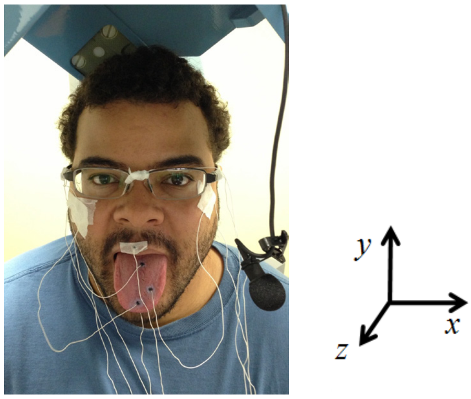

Figure 1. *Positions of sensors attached on the subject's head, tongue, and lips.*

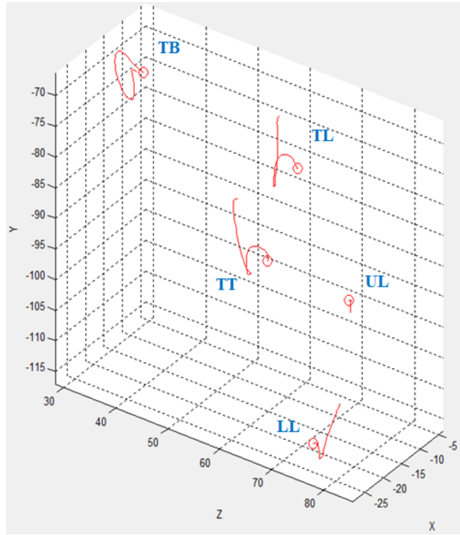

Figure 2. *Example of motion paths of the five articulators producing /ɹɑ/. The red circles indicate the end points of the motion paths. Sensor labels are described in text.*

over and above the optimal set consisting of sensors along the tongue midline (i.e., tongue tip, tongue body back, upper lip, and lower lip [12]).

2. Data collection

2.1. Participants and stimuli

Six native English speakers participated in the data collection. No history of speech, language, hearing, or any cognitive problems was reported from the participants.

Each speaker participated in one session in which he/she repeated a sequence of twelve consonants /s/, /z/, /ʃ/, /ʒ/, /dʒ/, /t/, /d/, /ɹ/, /l/, /j/, /k/, /g/, embedded in a short phrase "It's a /Cɑ/ game." These consonants were selected because they are representative of place and manner of articulation of consonants in English [2, 25].

2.2. Tongue motion tracking device and procedure

The electromagnetic articulograph (EMA) AG501 (Carstens Medizinelektronik GmbH, Bovenden, Germany) was used to collect 3D movement data of the head, tongue, and lips for all participants. EMA records tongue movements by establishing a calibrated electromagnetic field that induces electric current into small sensors attached to the surface of the articulators. AG501 is the latest model of EMA. The spatial precision of motion tracking using AG501 is approximately 0.3 mm [30]. The standard sampling rate of the motion data is 250 Hz [30].

Participants were seated with their head within the calibrated magnetic field. Sensors were then attached to the surface of each articulator using dental glue (PeriAcryl Oral Tissue Adhesive) or tape. The participants were then asked to produce the phrase sequences multiple times at their habitually comfortable speaking rate and loudness. Before the beginning of actual data recording, a five-minute training and practice session helped the participants adapt to the sensors. Previous studies have shown these sensors do not significantly affect speech output [31, 32].

Figure 1 shows the positions of eight sensors attached to a participant's head, tongue, and lips. Three of the sensors were attached to a pair of glasses. HC (Head Center) was on the bridge of the glasses; HL (Head Left) and HR (Head Right) were on the left and right outside edge of each lens, respectively. Three sensors - TT (Tongue Tip), TB (Tongue Body Back), and TL (Tongue Lateral) - were attached on the tongue. TT was about approximately 10 mm from the actual tongue tip. TB was as about 30 to 40mm from the tongue tip [2, 3, 4, 33]. Tongue lateral was about 10 to 15 mm back and right of TT. Lip sensors were attached to the vermilion borders of the upper (UL) and lower (LL) lips at midline. The data of TT, TB, TL, UL, and LL were used for analysis.

2.3. Data preprocessing

The sensor coordinate data recorded using EMA were preprocessed prior to analysis. First, head movements and orientations were subtracted from the tongue and lip data to provide head-independent measurements of the articulators. The orientation of the derived 3D Cartesian coordinate system is displayed in Figure 1. Second, a low pass filter (20 Hz) [33] was applied for removing noise.

As noted previously, only *y* (vertical) and *z* (front-back) coordinates (see Figure 5) of the tongue and lip sensors (i.e., TT, TB, TL, UL, LL) were used for this experiment because movement along the *x* axis (left-right) is not usually significant in speech production [2, 3]. The center of the magnetic field is the origin (zero point) of the EMA coordinate system.

Error samples (e.g., mispronunciation) were rare and were excluded from the analysis. In all, 816 consonant production samples were obtained and analyzed in this study.

3. Method

Support vector machine [29], a widely used machine learning classifier, was used to classify these consonants from the motion data of individual and combined articulators. SVM classifiers project training data into a higher dimensional space and then separate classes using a linear separator [29]. The linear separator maximizes the margin between groups of training data through an optimization procedure. Those training samples on the boundaries of the classes are called support vectors. A kernel function is used to describe the distance between two samples (i.e., u and v in Equation 1). The following radial basis function was used as the kernel

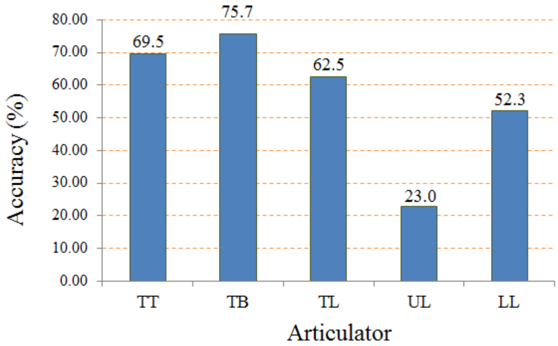

Figure 3: *Average consonant classification accuracy of single articulators across participants.*

function K_{RBF} in this study, where λ is an empirical parameter:

$$K_{RBF}(u, v) = \exp(1 - \lambda \| u - v \|) \quad (1)$$

For more details, please refer to [34], which describes the implementation of the SVM used in this study.

The same approach for constructing data samples in [12] was used in this study, where a sample (e.g., u or v in Equation 1) is a concatenation of discretized motion paths of multiple articulators. Initially, the data of each individual articulator for each consonant were time-normalized and sampled to a fixed length (i.e., 10 data points). The length was fixed, because SVM requires the input samples to be in a fixed-width array. The predominant frequency of tongue and lip movements is about 2 to 3 *Hz* for simple CV utterances; thus, 10 samples adequately preserve the motion patterns [2, 5]. Next, the arrays of y or z coordinates for those articulators were mean-normalized and concatenated into one sample (vector) representing the consonant. Overall, each sample contained 20 × p (10 data points × 2 dimensions × p articulators) attributes for p articulators ($1 \leq p \leq 5$). An integer (e.g., 1 for /sɑ/ and 2 for /zɑ/) was used for labeling the training data.

Figure 2 shows the row motion paths of all five articulators using the Smash software program [33]. The duration in the example of Figure 2 was 356 ms. Thus, there were about 89 data points (0.356 s × 250 Hz), for each articulator, which was then down-sampled to 10 data points.

Cross validation is a standard procedure for evaluating the performance of classification algorithms in machine learning, where training data and testing data are unique. In this study, a balanced leave-one-out cross validation strategy was used. In each execution, one sample for each consonant (totally N consonants) for each participant was selected for testing, and the rest were used for training. This strategy uses the same number of training samples per consonant in each execution. There were a total of m executions; where m is the number of samples per consonant. The average classification accuracy (number of correctly classified samples over the number of all samples in test) of all m executions was considered the overall classification accuracy [35].

4. Results and Discussion

4.1. Classification based on individual articulators

Figure 3 gives the average consonant classification accuracies across participants for each individual articulator. Accuracy obtained from any single tongue articulator (TT, TB, or TL) was significantly higher than that from LL or UL; accuracy for LL was significantly higher than that for UL (p < 0.001). Accuracy obtained from TB was significantly higher than that obtained from TL (p < 0.01). There were no significant differences between the other tongue articulators.

Table 1. *Consonant classification accuracies using data of individual or combined articulators across speakers.*

Articulator (Sensor) Combinations	Average (%)	Standard Deviation
{TT}	69.5	7.7
{TB}	75.7	9.6
{TL}	62.5	3.9
{UL}	23.0	7.9
{LL}	52.3	6.7
{TT, TB, UL, LL}	82.2	9.6
{TT, TL, UL, LL}	74.8	5.3
{TL, TB, UL, LL}	82.2	9.0
{TT, TB, TL, UL, LL}	82.5	6.9

Table 1 shows show the average classification accuracies (with standard deviations) obtained from each single articulator of all participants. These sensor-specific findings (excluding TL) are consistent with our previous results [12]. A key finding is that TL is not significantly different from TT, possibly because the sensor location of TL is close to TT, and movements of the two articulators are highly correlated, as previously illustrated in Figure 2.

4.2. Classification on articulator combinations

An optimal set of sensors (minimum set of sensors that can be used to accurately classify speech movements) has been determined (for sagittal movement) [12]. However, there were no tongue lateral (TL) data in [12]. In this study, we investigated if TL data can help distinguish those consonants. To address this question, we compared the classification accuracies of relevant combinations of articulators. For ease of explanation, we name the optimal combination/set (TT was named T1, and TB was named T4 in [12])

$$\boldsymbol{A} = \{\text{TT, TB, UL, LL}\} \quad (2)$$

First, the accuracy obtained from the whole set of all five articulators {TT, TB, TL, UL, LL} was compared to that of ***A*** (Table 1). A paired t-test showed there was no significant difference between the two articulator combinations. This finding suggests adding TL to ***A*** did *not* benefit general consonant classification. This could be also explained by the correlation of TT and TL.

Second, a comparison was conducted by replacing TT or TB with TL in ***A***, (i.e., {TT, TL, UL, LL}, {TB, TL, UL, LL}) (Table 1). Paired t-test shows there was no significant difference between ***A*** and {TT, TL, UL, LL}, nor between ***A*** and {TL, TB, UL, LL}. This finding suggests TL may be correlated with TT or TB.

Finally, {TT, TL, UL, LL} and {TB, TL, UL, LL} was compared to the whole set (i.e., {TT, TB, TL UL, LL}, see Table 1). There was a significant difference between the

Table 2. *Consonant classification matrix (in percentage) for all five articulators. Row numbers are actual number of samples; column numbers are the classified number of samples.*

	/s/	/z/	/ʃ/	/ʒ/	/dʒ/	/t/	/d/	/ɹ/	/l/	/j/	/k/	/g/
/s/	**.78**	.16		.03					.02			.01
/z/	.13	**.71**		.13		.01			.02			
/ʃ/	.01		**.87**	.05	.04					.02		
/ʒ/	.05	.11	.07	**.75**	.03							
/dʒ/			.04	.06	**.86**	.03	.01					
/t/					.03	**.76**	.21					
/d/					.03	.19	**.78**					
/ɹ/							.03	**.92**		.05		
/l/		.03							**.96**	.02		
/j/							.01	.02		**.91**	.01	.04
/k/											**.78**	.22
/g/										.04	.14	**.82**

Table 3. *Consonant classification matrix (in percentage) for articulators without TL (Tongue Lateral). Row numbers are actual number of samples; column numbers are the classified number of samples.*

	/s/	/z/	/ʃ/	/ʒ/	/dʒ/	/t/	/d/	/ɹ/	/l/	/j/	/k/	/g/
/s/	**.73**	.16		.08			.01		.02			
/z/	.13	**.71**		.11		.01			.03			
/ʃ/	.01		**.87**	.05	.04					.02		
/ʒ/	.05	.14	.07	**.72**	.03							
/dʒ/			.04	.04	**.86**	.04		.02				
/t/					.04	**.74**	.21					
/d/					.03	.21	**.76**					
/ɹ/						.01	.01	**.97**				
/l/		.04							**.94**	.02		
/j/							.01			**.94**	.01	.03
/k/											**.82**	.18
/g/										.04	.15	**.80**

accuracy obtained from {TT, TL, UL, LL} and the whole set ($p < 0.02$); There was no significant difference between the accuracy obtained from {TB, TL, UL, LL} and that obtained from the whole set. This finding suggests TT and TL may be functionally interchangeable. TB, however, contains unique information that TT or TL does not have.

4.3. Classification matrices with and without using Tongue Lateral data

To further understand the contribution of tongue lateral (TL) in consonant production, two classification matrices (or confusion matrices) were provided, one with the whole set {TT, TB, TL, UL, LL} and one without TL.

Table 2 and 3 give the classification matrices with or without using Tongue Lateral (TL) data, respectively. Row numbers are actual number of samples; column numbers are the classified number of samples. Ideally, the diagonal numbers (in bold) are ones; and others are zeros (not displayed). There are consistencies of correct classifications (diagonal numbers) and misclassifications (non-diagonal numbers) in Tables 2 and 3. For example, the highest accuracies were obtained from the consonants /ɹ/, /l/, /j/; the lowest accuracies were obtained from /s/, /z/, /ʒ/; middle accuracies were obtained from other consonants. Tables 2 and 3 further supported that adding TL did not benefit the consonant classification in addition to ***A***.

This finding that TL (tongue lateral) does not benefit consonant distinctiveness over the midsagittal sensors is consistent with previous single-subject MRI studies which indicate tongue midsagittal information is enough to account for the tongue shape characterization for a set of French vowels and consonants [36, 37].

The observation that TB played a predominant role in consonant classification for these /ɑCɑ/ stimuli replicates a previous EMA study of English talkers [12]. Although many studies suggest that four to five degrees of freedom are required to model the variety of tongue shapes found in larger English speech samples (e.g., all consonants across different vowel contexts), the present findings suggest that TB may be the best single articulator to describe smaller corpuses, and that the set {TT, TB, UL, LL} is optimal [12].

The current data suggest a midline sensor array {TT, TB, UL, LL} may be sufficient for many applications, for example, silent speech interfaces based on electromagnetic sensing [16, 38, 39], and 2D EMA visual feedback systems [20]. However, in emerging technologies that use 3D animated tongue shape for visual feedback [40], TL provides critical information in addition to midline tongue sensors.

Limitations. Although the experiment results are encouraging, the data set used in the experiment was collected from only a small number of subjects. Also, as noted previously, only sagittal movements were analyzed in this study; relevant analyses on x (left-right) movement of the lateral tongue sensor remain to be done.

5. Conclusions & Future Work

This paper investigated the contribution of Tongue Lateral information to consonant classification, in addition to an optimal set (Tongue Tip, Tongue Body Back, Upper Lip, and Lower Lip in the middle line) found previously. The results suggest that adding Tongue Lateral information to the optimal set did not increase the consonant classification accuracy.

Future work includes (1) computing correlation coefficients of the tongue sensor data, (2) verification of the present findings using a dataset collected from more participants, and (3) analysis of the contribution of x (left-right) movement of the tongue lateral sensor.

6. Acknowledgements

This work was in part supported by Callier Excellence in Education Fund, University of Texas at Dallas, and grants awarded by the National Institutes of Health (R01 DC013547 and R43 DC013467). We like to thank Dr. Jordan R. Green, Dr. Ashok Samal, Sonya Mehta, Marcus Jones, Kristin J. Teplansky, Vedad "Kelly" Fazel, Loren Montgomery, Kameron Johnson, Amy Berglund, and Mona Dirbashi for their support or contribution in data collection and processing.

7. References

[1] Kent, R. D., Adams, S. G., and Tuner, G. S., "*Models of speech production*," in Principles of Experimental Phonetics, N. J. Lass, Ed., St Louis, MO: Mosby, 1996.

[2] Wang, J., Green, J. R., Samal, A., and Yunusova, Y., "Articulatory distinctiveness of vowels and consonants: A data-driven approach," *Journal of Speech, Language, and Hearing Research*, 56:1539-51, 2013.

[3] Westbury, J., *X-ray microbeam speech production database user's handbook*, University of Wisconsin, 1994.

[4] Green J. R. and Wang, Y., "Tongue-surface movement patterns during speech and swallowing," *Journal of Acoustical Society of America*, 113:2820-33, 2003.

[5] Wang, J., Green, J. R., Samal, A., and Marx, D. B., "Quantifying articulatory distinctiveness of vowels", *Proc. Interspeech*, 277-280, Florence, Italy, 2011.

[6] Yunusova, Y., Weismer, G., Westbury, J. R., and Lindstrom, M. J., "Articulatory movements during vowels in speakers with dysarthria and healthy controls," *Journal of Speech, Language, and Hearing Research*, 51(3):596-611, 2008.

[7] Wrench, A. A., "A multi-channel/multi-speaker articulatory database for continuous speech recognition research," *Phonus*, 5:1-13, 2000.

[8] Rudzicz, F., Namasivayam, A. K., and Wolff, T., "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," *Language Resources and Evaluation*, 46(4):523-541, 2011.

[9] Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., and Perkell, J., "Articulatory tradeoffs reduce acoustic variability during American English /r/ production," *Journal of Acoustical Society of America*, 105:2854-65, 1999.

[10] Jackson, P. J. B. and Singampalli, V. D., "Statistical identification of articulation constraints in the production of speech, *Speech Communication*, 51, 695-710, 2009.

[11] Qin, C., Carreira-Perpiñán, M. A., Richmond, K., Wrench, A., and Renals, S., "Predicting tongue shapes from a few landmark locations," *Proc. Interspeech*, 2306-2309, 2008.

[12] Wang, J., Green, J. R., and Samal, A., "Individual articulator's contribution to phoneme production", *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, 7795-89, 2013.

[13] Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., and Brumberg, J., "Silent speech interfaces," *Speech Communication*, 52:270-287, 2010.

[14] Wang, J., Samal, A., Green, J. R., and Rudzicz, F., "Sentence recognition from articulatory movements for silent speech interfaces," *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, 4985-88, 2012.

[15] Wang, J., Samal, A., Green, J. R., and Rudzicz, F., "Whole-word recognition from articulatory movements for silent speech interfaces," *Proc. Interspeech*, 2012.

[16] Fagan, M. J., Ell, S. R., Gilbert, J. M., Sarrazin, E., and Chapman, P. M., "Development of a (silent) speech recognition system for patients following laryngectomy," *Medical Engineering & Physics*, 30(4):419-425, 2008.

[17] King, S., Frankel, J., Livescu, K., McDermott, E., Richmond, K., and Wester, M., "Speech production knowledge in automatic speech recognition," *Journal of Acoustical Society of America*, 121(2):723-742, 2007.

[18] Rudzicz, F., Hirst, G., and Van Lieshout, P. "Vocal tract representation in the recognition of cerebral palsied speech," *Journal of Speech, Language, and Hearing Research*, 55(4): 1190-1207, 2012.

[19] Levitt, J. and Katz, W. F., "The effects of EMA-based augmented visual feedback on English speakers' acquisition of the Japanese flap: A perceptual study," *Proc. Interspeech*, 1862-1865, Makuhari, Japan, 2011.

[20] Katz, W. F. and McNeil, M., "Studies of articulatory feedback treatment for apraxia of speech (AOS) based on electromagnetic articulography," *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, 20(3):73-80, 2010.

[21] Narayanan, S.S., Alwan, A.A. and Haker, K., "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals," *The Journal of the Acoustical Society of America*, 101: 1064-1077, 1997.

[22] Martins, P., Oliviera, C., Ferreira, C., Silva, A. and Teixeira, A., "3D MRI and semi-automatic segmentation techniques applied to the study of European Portuguese lateral sounds," *The 9th International Seminar on Speech Production*, 235-242, 2011.

[23] Stavness, I., Lloyd, J. E. and Fels, S., "Automatic prediction of tongue muscle activations using a finite element model," *Journal of Biomechanics*, 45(16): 2841-2848, 2012.

[24] Hamlet, S. L. and Bunnell, T., and Struntz, B., "Articulatory asymmetries," *Journal of the Acoustical Society of America*, 79(4): 1164-69, 1986.

[25] Ladefoged P. and Johnson, K., *A course in phonetics* (6th Ed.). Independence, KY: Cengage Learning, 2011.

[26] Ladefoged, P., *Elements of Acoustic Phonetics*, London: University of Chicago Press, 1962.

[27] Fant, G., *Speech Sounds and Features*, Cambridge, Mass.: MIT Press, 1973.

[28] Stevens, K. N., *Acoustic Phonetics*, Cambridge, Mass.: MIT Press, 1998.

[29] Cortes C. and Vapnik, V., "Support-vector network," *Machine Learning*, 20:273-297, 1995.

[30] Carstens Medizinelektronik GmbH, "Articulograph AG501 Flyer", available online: http://www.articulograph.de/wp-content/uploads/2013/01/Articulograph-AG501-brochure-2012.pdf, retrieved on Mar 08, 2013.

[31] Katz, W. F., Bharadwaj, S., Rush, M., and Stettler, M., "Influences of EMA receiver coils on speech production by normal and aphasic/apraxic talkers", *Journal of Speech, Language, and Hearing Research*, 49:645-659, 2006.

[32] Weismer, G. and Bunton, K., "Influences of pellet markers on speech production behavior: Acoustical and perceptual measures," *Journal of the Acoustical Society of America*, 105: 2882-2891, 1999.

[33] Green, J. R., Wang, J., and Wilson, D. L., "SMASH: A tool for articulatory data processing and analysis", *Proc. Interspeech*, Lyon, France, 1331-35, 2013.

[34] Chang C. C. and Lin, C. J. "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, 2(3):1-27, 2011.

[35] Wang, J., *Silent speech recognition from articulatory motion*, Doctoral dissertation, University of Nebraska-Lincoln, 2011.

[36] Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C. and Savariaux, C., "Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images," *Journal of Phonetics*, 30(3): 533-553, 2002.

[37] Badin, P. and Serrurier, A., "Three-dimensional linear modeling of tongue: Articulatory data and models," *The 7th International Seminar on Speech Production*, Ubatuba, Brazil, 395-402, 2006.

[38] Wang, J., Balasubramanian, A., Mojica de La Vega, L., Green, J. R., Samal, A., and Prabhakaran, B., "Word recognition from continuous articulatory movement time-series data using symbolic representations," *ACL/ISCA Workshop on Speech and Language Processing for Assistive Technologies*, 119-127, 2013.

[39] Wang, J., Samal, A., and Green, J. R. (2014). Preliminary test of a real-time, interactive silent speech interface based on electromagnetic articulograph, *ACL/ISCA Workshop on Speech and Language Processing for Assistive Technologies* (In press).

[40] Katz, W. F., Campbell, T. F., Wang, J., Farrar, E., Eubanks, J. C., Balasubramanian, A., Prabhakaran, B., and Rennaker, R., "Opti-speech: A real-time, 3D visual feedback system for speech training", *Proc. Interspeech*, Singapore, 2014 (In press).