Acoustic and Kinematic Characteristics of Vowel Production through a Virtual Vocal Tract in Dysarthria

Jeff Berry[1], *Andrew Kolb*[2], *Cassandra North*[2], *and Michael T. Johnson*[2]

[1] Marquette University Speech Pathology & Audiology, Milwaukee, WI, USA
[2] Marquette University Electrical and Computer Engineering, Milwaukee, WI, USA
jeffrey.berry@marquette.edu

Abstract

Broadening our understanding of the components and processes of speech sensorimotor learning is crucial to furthering methods of speech neurorehabilitation. Recent research in limb sensorimotor control has used virtual environments to study learning in novel sensorimotor working spaces. Comparable experimental paradigms have yet to be undertaken to study speech learning. We present acoustic and kinematic data obtained from participants producing vowels in unfamiliar articulatory-acoustic working spaces using a virtual vocal tract. Talkers with dysarthria and healthy controls were asked to produce vowels using an electromagnetic articulograph-driven speech synthesizer for participant-controlled auditory feedback. The aim of the work was to characterize performance within and between groups to generate hypotheses regarding experimental manipulations that may bolster our understanding of speech sensorimotor learning. Results indicate that dysarthric talkers displayed relatively reduced acoustic working spaces and somewhat more variable acoustic targets compared to controls. Kinematic measures of articulatory dynamics, particularly peak speed and movement jerk-cost, were idiosyncratic and did not dissociate talker groups. These findings suggest that individuals with dysarthria and healthy talkers may use idiosyncratic movement strategies in learning to control a virtual vocal tract, but that dysarthric talkers may nonetheless exhibit acoustic limitations that parallel deficits in speech intelligibility.

Index Terms: sensorimotor learning, auditory feedback, dysarthria, articulatory resynthesis.

1. Introduction

Broadening our understanding of the components and processes of speech sensorimotor learning is crucial to forwarding methods of speech neurorehabilitation. Research into the nature of articulatory adaptations elicited in normal talkers through conventional acoustic feedback is well-developed and clearly establishes that auditory feedback manipulations can affect articulation [1]. Moreover, these sensorimotor adaptations cannot be consciously controlled [2]. Formant frequency patterns can be used as measures to reflect articulator movements and provide sensory goals for articulatory movements of vowels [3, 4]. Perceived shifts in formant patterns can evoke real time compensatory changes in articulator movement patterns in typically-functioning talkers [5-8]. For example, lowering the perceived value of the first formant (F1) frequency evokes adaptive changes in articulation that compensate for the perceived sensory error by raising F1. In vowel contexts, a perceived decrease in F1 evokes compensation in the form of a relatively lower tongue position (higher F1) [5]. Currently, very little work has addressed the mechanisms of auditory feedback in disordered talkers [9, 10]. Moreover, the literature on speech-related influences of auditory feedback has focused on articulatory changes resulting from acoustic perturbations within a talker's familiar acoustic working space. The current work is novel in that the acoustic working space utilized by participants is wholly unfamiliar, reflecting a completely different vowel-acoustic space, presumably requiring the development of a novel sensorimotor calibration using novel auditory-acoustic targets for the production of speech sounds.

Haith and Krakauer [11] suggest that, while the bulk of sensorimotor learning research has utilized paradigms examining sensorimotor adaptations resulting from perturbations to familiar sensory targets, learning sensorimotor control within an unfamiliar sensory working space (with or without perturbation) is characteristic of much real human experience and a poignant research consideration. Several researchers have studied limb sensorimotor control using virtual environments to create novel sensorimotor working spaces [12-16]. However, the study of sensorimotor learning in unfamiliar sensory environments has yet to be undertaken for the auditory-motor transformations that are important to speech sensorimotor learning.

We present kinematic and acoustic data describing participant efforts to produce vowel sounds within an unfamiliar articulatory-acoustic working space using a virtual vocal tract. Typically-functioning participants and participants with dysarthria were asked to produce vowels using an electromagnetic articulograph (EMA) driven articulatory speech synthesizer to provide participant-controlled auditory feedback. The aim of this work is to characterize performance similarities and differences within and between groups to generate hypotheses regarding future experimental manipulations that may support neurorehabilitation. While typical methods for eliciting involuntary speech sensorimotor adaptations using LPC resynthesis are not viable for many individuals with dysarthria because they require robust a speech-acoustic signal from the participant, the current "virtual vocal tract" method provides a viable alternative for eliciting sensorimotor adaptation that circumvents this limitation [17].

2. Methods

The NDI Wave EMA system was used to register the real time movements of the tongue, lips, and jaw (see Figure 1). Five sensors were attached along the midsagittal plane (two on the dorsal surface of the tongue, one on each lip, and one at the

juncture of the central mandibular incisors near the gingival border). Reference sensors corrected for participant head movements. Sensor movements were transformed into control parameters for an articulatory speech synthesizer [18, 19]. The mathematical method for transforming articulator movements was speaker-independent, though speaker-specific calibrations based on differences in the available physical working space within the oral cavity were necessary. The "speech" heard by all participants was indistinguishable since the acoustic vowel space defining the novel auditory-sensory working space was the same for all participants except for subtle differences in vowel quality reflecting idiosyncrasies in articulatory movement and differences in proficiency between participants in reliably controlling the synthesizer.

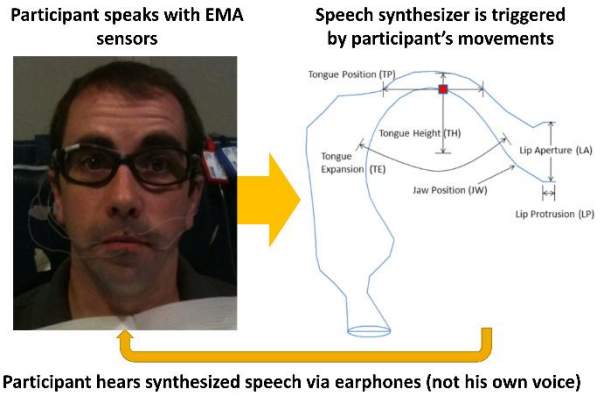

Figure 1: *Schematic of EMA-driven articulatory synthesis.*

Four typically-functioning speakers and four speakers with dysarthria completed the experimental protocol. All participants in the dysarthria group met the inclusionary criteria of being survivors of traumatic brain injury with functionally significant speech-intelligibility deficits [20], no indicated deficits in speech discrimination [21], and sufficient sustained attention and short-term auditory memory to participate in the experiment [22]. All participants passed a brief audiometric pure-tone screening to assure hearing was within functional limits. Participants read and signed informed consent documents. All procedures and documentation were approved by the Institutional Review Board of Marquette University. Each participant received a $50 cash incentive for participating in the approximately two hour experimental process.

EMA sensors were adhered to the articulators using Periacryl™ adhesive (Glustich Inc., Delta, BC, Canada). To improve the duration and reliability of sensor adhesion, lingual sensors were bonded with small squares of silk between the sensor and lingual surfaces. Similarly, labial sensors had 2 mm diameter circles of Super Poligrip Strips® denture wax (GlaxoSmithKline Consumer Healthcare L.P.) and the dental sensor had a 3 x 5 mm strip of Stomahesive® periostomal barrier (ConvaTec, Skillman, NJ, USA) used as intermediaries to support adhesion. Five articulatory sensors were used for each participant (central mandibular incisors, lower lip, upper lip, tongue dorsum, and tongue blade sensors were all placed approximately in the midsagittal plane).

Participants were seated in a non-metallic chair (constructed primarily of polyvinyl chloride) with the field generator for the NDI Wave EMA system positioned approximately 5 cm from the head in left profile to assure that all sensors fell within the 300mm^3 field setting of the EMA system. Participants wore a pair of plastic glasses frames with a 6DOF reference sensor attached at midline to allow for use of the EMA system's automated head movement correction algorithm. A computer screen was positioned approximately 1 m away to display the stimuli "A" and "O". Participants wore insert earphones through which pink noise was mixed with synthesizer audio feedback. Participants were instructed to use their articulators to control the speech synthesizer in sustaining the indicated vowel sound when it appeared on the screen and for as long as they could hear it. Audio gating was achieved in synchrony with visual cues using Matlab scripts in conjunction with the TransShiftMex (Audapt) software [6]. For each trial, auditory feedback was available for 2.5 s windows with 5 s breaks between adjacent repetitions.

Following preparation, participants were given a period of accommodation to the EMA sensors, during which they spoke casually with the experimenter and lab staff and read a paragraph of text to become accustomed to the size and location of the text-prompted stimuli. Next a series of "calibration" maneuvers was recorded to define the physical limits of each participant's kinematic working space and obtain a viable mapping to the virtual vocal tract. Calibration maneuvers included: exaggerated CV reps, maximum jaw wags, maximum lip protrusions and retractions, and sustained corner vowels. The mapping was determined adequate if the participant was able to achieve a reasonable approximation of the target vowel qualities within a few attempts without repeated, complete constriction of the virtual articulators. This approach assured that participant articulatory movements would affect the synthesized sound without mapping directly onto their own speech articulation. Thus, the articulation required to produce a particular vowel was novel, requiring participants to learn to control the unfamiliar articulatory-acoustic working space. Once the calibration process was completed the baseline experiment was initiated. Participants were asked to produce 80 (total) repetitions of each /e/ and /o/ using only the synthesized speech for auditory feedback. The experiment required a total of 80 vowel productions and lasted approximately 30 minutes.

1st and 2nd formant frequency measures (F1 & F2) were obtained from the synthesized productions using the TF32 software [23]. Formant tracks were obtained via pitch-synchronous LPC (26 coefficients). All spectrograms were analyzed with a 300 Hz bandwidth. Given the very robust signal generated by the articulatory synthesizer, formant tracking errors were rarely apparent. When such errors occurred, F1 and F2 values were manually corrected within visualized windows of approximately 500 ms and a frequency range around 5 kHz. Indices were placed around an approximately 100 ms span characterized by a steady state F1 and F2 (the "off-glide" for /e/). Formant values were averaged within this window to obtain formant frequency values. Kinematic measures of performance included: 1) overall sensor working space (calculated per vowel based on the area of a convex hull circumscribing all sensor positions); 2) peak sensor speed (calculated per repetition using a 3-point central difference method); and 3) movement jerk-cost.

3. Results

Figure 2 illustrates all of the acoustic values in Hertz and positions of the participants' articulations. The values are distinguished by both vowel sound and by group. Vowels are represented by triangles and circles for /e/ and /o/, respectively. The control group is represented by green and

yellow, for /e/ and /o/, respectively and the dysarthria group is represented by red and black, for /e/ and /o/, respectively. The axes in Figure 1 are oriented to represent an approximation of the articulatory vowel space. The inverted F1 axis represents the tongue height, while F2 represents tongue advancement. As such, since /e/ is classified as a high front vowel, one would expect it to appear in the upper right quadrant (low F1, high F2) on this figure. Qualitative analysis of these formant values reveals distinct vowel positions for /e/ and /o/ for the typically-functional speakers. When looking at the values for the speakers with dysarthria, more overlap between the vowel positions is evident, which illustrates less discriminant and more centralized vowels.

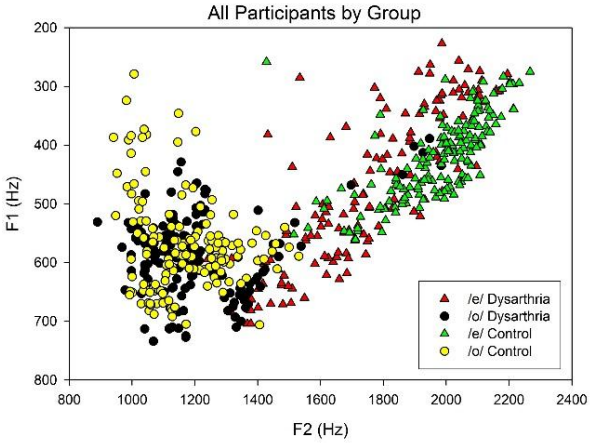

Figure 2: *Formant "target" values by vowel & group.*

Table 1: *Individual mean (SD) formant "target" values (Hz).*

	/e/		/o/	
	F1	**F2**	**F1**	**F2**
S4- D	461 (85)	1758 (176)	536 (45)	1176 (48)
S6- D	586 (82)	1612 (239)	612 (82)	1137 (212)
S8- D	325 (49)	1940 (170)	573 (57)	1206 (244)
S9- D	563 (61)	1588 (153)	648 (36)	1361 (63)
DYSARTHRIA GROUP MEANS	***476 (127)***	***1738 (235)***	***587 (71)***	***1207 (188)***
S2- C	445 (46)	1995 (105)	582 (40)	1284 (136)
S1- C	452 (62)	1930 (175)	642 (41)	1068 (45)
S7- C	392 (61)	2011 (114)	584 (34)	1238 (56)
S5- C	393 (79)	1956 (185)	489 (95)	1084 (89)
CONTROL GROUP MEANS	***420 (69)***	***1973 (151)***	***574 (80)***	***1168 (129)***
GROUP DIFFERENCES	**56* (58)***	**-235* (84)***	**13 (-9)**	**39* (59)**

Table 1 contains all of the individual mean and standard deviation (SD) values of F1 and F2 for both /e/ and /o/. It also contains the group mean and standard deviation values, as well as the group differences. The dysarthria group produced F1 for /e/ 56Hz higher and with more variability (58Hz greater SD). On average, the dysarthria group produced F1 of /o/ 13Hz higher than the controls. For F2 of /e/, the dysarthria group produced it 235Hz lower and again with more variability (84Hz greater SD) than the controls. For F2 of /o/, the dysarthria group was 39Hz higher. Levene's test of homogeneity of variance indicated significantly greater variability among the dysarthria group for /e/ measures ($p<0.001$), but no differences in variability between groups for /o/ measures. Independent samples t-tests were completed between groups within measure. F1 values for /e/ were significantly higher ($t(218.483)=4.748$, $p<0.001$) and F2 values for /e/were significantly lower ($t(243.616)=-10.310$, $p<0.001$) for the dysarthria group. F1 values for /o/ did not differ significantly between groups ($t(304)=1.435$, $p=0.150$), while F2 values for /o/ were significantly higher for the dysarthria group ($t(304)=2.107$, $p=0.039$). Comparison of results from individual participants reveals that some controls were more variable than some individuals with dysarthria.

Acoustic results revealed that both typically-functioning speakers and speakers with dysarthria demonstrated the ability to learn a novel articulator-acoustic mapping and produce discriminable speech sounds. Comparisons of acoustic measures between groups revealed that participants with dysarthria produced significantly more centralized (reduced range of motion) and variable acoustic targets compared to control.

Kinematic measures were analyzed to determine if there were group differences in the dynamics of articulation. Figure 3 shows examples of articulatory working space measures for 4 participants (2 controls, 2 dysarthric). These measures were obtained by determining the area of a convex hull circumscribing all tongue blade sensor movements during articulation of all replicates of each of the two target vowels. Data for /e/ are shown in black and data for /o/ are shown in purple. Differences in articulatory working space were idiosyncratic with no clear group differences in the size of the articulatory space used during the experiment. One qualitative difference that may be group related is the apparently increased overlap between vowels within the dysarthric group.

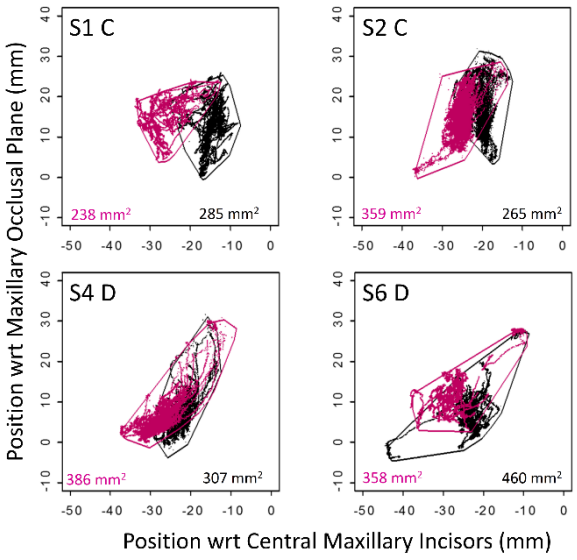

Figure 3: *Kinematic working space by vowel for 4 participants.*

Figure 4 shows box and whisker plots describing peak sensor speed across all vowel productions for each participant. The four control talkers are on the left side of the plot and the four dysarthric talkers are on the right side. Participants were clearly highly idiosyncratic with respect to the peak speed of tongue movement. No group differences are apparent for overall peak speed or peak speed variability.

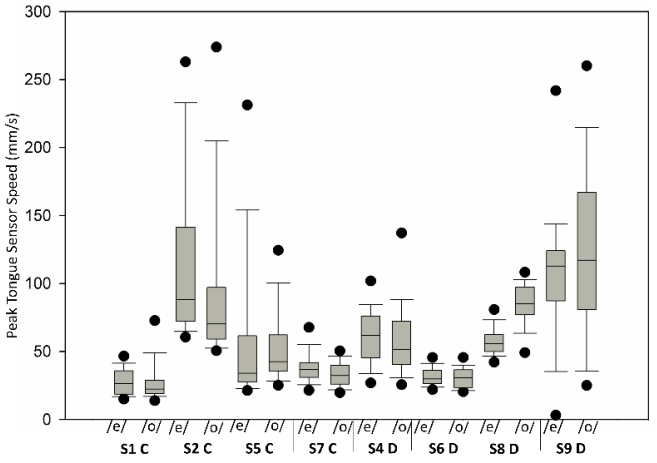

Figure 4: *Peak tongue sensor speed by vowel and participant.*

To assess movement irregularity (stability) during the vowel task, a measure of movement jerk was used. Such measures have long been applied to characterizing movement differences resulting from motor disorder. Hogan and Sternad [24] argue that dimensionless measures of jerk that are not contingent on movement duration and amplitude provide superior quantification of movement irregularity. We used a discretized version of the normalized jerk-cost calculation originally presented by Takada et al. [25] for characterizing jaw movement irregularity.

Figure 5 shows box and whisker plots describing normalized jerk-cost across all vowel productions for each participant. The four control talkers are on the left side of the plot and the four dysarthric talkers are on the right side. While the dysarthric group revealed overall larger mean normalized jerk cost values (suggesting greater movement irregularity) this was not a significant result. Moreover, participants were highly indiosyncratic with respect to normalized jerk-cost and normalized jerk-cost variability.

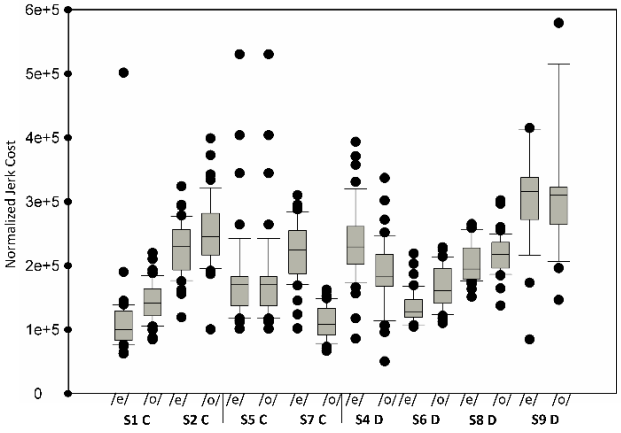

Figure 5: *Normalized jerk-cost by vowel and participant.*

4. Discussion

Taken together, the results of the current study suggest that, while talkers with dysarthria may display limits in the acoustic working space when producing vowels through a virtual vocal tract, kinematic measures reflecting more dynamic aspects of articulation do not appear to clearly dissociate groups during this task. These findings may reflect the possibility that typically-functioning talkers are not necessarily better able than talkers with dysarthria in learning to articulate within a novel articulatory-acoustic working space. Nonetheless, dysarthric talkers appear to maintain acoustic characteristics of the dysarthria that could reflect changes in acceptable auditory-acoustic targets for vowel production, particularly for talkers who are non-functionally unintelligible. It is hypothesized that experiment protocols that perturb the articulatory-acoustic mapping by reducing the acoustic sensitivity of articulatory movement may provide novel insights into sensorimotor learning in dysarthria and may prove useful in bolstering novel approaches to the treatment of dysarthria. A limiting consideration is that auditory-acoustic targets do not completely define the sensory goals of speech and adaptation can be driven by somatosensory perturbations [26, 27]. Moreover, auditory-perceptual processing can also be affected by somatosensory perturbation [28]. At this time, much remains to be elucidated about the integration of different sensory feedback channels in sensorimotor learning [28-30].

5. Conclusions

Results indicate that dysarthric talkers displayed relatively reduced acoustic working spaces and somewhat more variable acoustic targets compared to controls when producing vowels through a virtual vocal tract. Kinematic measures of articulatory dynamics, particularly peak speed and movement jerk-cost, were idiosyncratic and did not dissociate talker groups. These findings suggest that individuals with dysarthria and healthy talkers may use idiosyncratic movement strategies in learning to control a virtual vocal tract, but that dysarthric talkers may nonetheless exhibit acoustic limitations that parallel deficits in speech intelligibility.

6. Acknowledgements

This paper is based upon work supported by American Speech-Language-Hearing Foundation New Century Scholar Grant and the National Science Foundation under Grant No. IIS-1142826 and IIS-1320892.

7. References

[1] Houde, J. F., & Jordan, M. I. "Sensorimotor Adaptation in Speech Production," *Science*, 279, 1213-1216, 1998.

[2] Munhall, K. E., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. "Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate," *J. Acoust. Soc. Am.,* 125, 384-390, 2009.

[3] Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," *J. Phonetics*, 28, 233-272, 2000.

[4] Perkell, J. S. "Movement goals and feedback and feedforward control mechanisms in speech production," *J. Neurolinguistics*, 25, 382-407, 2012.

[5] Purcell, D. W. & Munhall, K. G. "Compensation following real-time manipulation of formants in isolated vowels," *J. Acoust. Soc. Am.*, 119, 2288-2297, 2006.

[6] Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong/iau/and its pattern of generalization." *J. Acoust. Soc. Am.*, 128, 2033-2048, 2010.

[7] Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. "Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech training." *J. Neuroscience*, 31, 16483-16490, 2011.

[8] Villacorta, V. M., Perkell, J. S., & Guenther, F. H. "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," *J. Acoust. Soc. Am.*, 122, 2306-2319, 2007.

[9] Mollaei, F., Shiller, D. M., & Gracco, V. L. "Sensorimotor adaptation of speech in Parkinson's disease," *Movement Disorders*, 28, 1668-1674, 2013.

[10] Tourville, J. A., Cai, S., & Guenther, F. H. "Exploring auditory-motor interactions in normal and disordered speech," *Proceedings of Meetings on Acoustics*, 060180, 2013.

[11] Haith & Krakauer. "Model-based and model-free mechanisms of human motor learning." *Advances in Experimental Medicine and Biology*, 782, 1-21, 2013.

[12] Mussa-Ivaldi, F. A., Casadio, M., Danziger, Z. C., Mosier, K. M., & Scheidt, R. A. "Sensory motor remapping of space in human–machine interfaces." *Progress in Brain Research*, 191, 45, 2011.

[13] Mosier, K. M., Scheidt, R. A., Acosta, S., & Mussa-Ivaldi, F. A. "Remapping hand movements in a novel geometrical environment." *J. Neurophysiology*, 94(6), 4362-4372, 2005.

[14] Liu, X., Mosier, K. M., Mussa-Ivaldi, F. A., Casadio, M., & Scheidt, R. A. "Reorganization of finger coordination patterns during adaptation to rotation and scaling of a newly learned sensorimotor transformation." *J. Neurophysiology*, 105(1), 454-473, 2011.

[15] Nagengast, A. J., Braun, D. A., & Wolpert, D. M. "Optimal control predicts human performance on objects with internal degrees of freedom." *PLoS Computational Biology*, 5(6), e1000419, 2009.

[16] Sternad, D., Abe, M. O., Hu, X., & Müller, H. "Neuromotor noise, error tolerance and velocity-dependent costs in skilled performance." *PLoS Computational Biology*, 7(9), e1002159, 2011.

[17] Berry, J. J., North, C., Meyers, B., & Johnson, M. T. "Speech sensorimotor learning through a virtual vocal tract." *Proceedings of Meetings on Acoustics*, 19, 060099, 2013.

[18] Maeda, S. "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model." In *Speech production and speech modelling* (pp. 131-149). Springer Netherlands, 1990.

[19] Huckvale, M. *VTDemo – Vocal Tract Acoustics Demonstrator* [computer software]. University College London, 2009.

[20] Yorkston, K.M., & Beukelman, D.R. *Assessment of Intelligibility of Dysarthric Speech.* Tigard, Oregon: C.C. Publications, Inc., 1984.

[21] Ross, M., Lerman, J., & Cienkowski, K.M. *Word Intelligibility by Picture Identification – WIPI*, 2nd Edition. St. Louis, MO: Auditec, 2004.

[22] Woodcock, R.W., McGrew, K.S., & Mather, N. *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside, 2001.

[23] Milenkovic, P. *TF32* [Computer software]. Madison, WI: University of Wisconsin–Madison, 2004.

[24] Hogan, N., & Sternad, D. "Sensitivity of smoothness measures to movement duration, amplitude, and arrests." *J. Motor Behavior*, 41, 529-534, 2009.

[25] Takada, K., Yashiro, K., and Takagi, M. "Reliability and sensitivity of jerk-cost measurement for evaluating irregularity of chewing jaw movements." *Physiol. Meas.*, 27, 609-622, 2006.

[26] Tremblay, S., Douglas, M. S., & Ostry, D. J. "Somatosensory basis of speech production," *Nature*, 423, 866, 2003.

[27] Nasir, S. M., & Ostry, D. J. "Somatosensory Precision in Speech Production," *Current Biology*, 16, 1918-1923, 2006.

[28] Nasir, S. M., & Ostry, D. J., "Auditory plasticity and speech motor learning," *Proc. Natl. Acad. Sci. U. S. A.*, 106, 20470-20475, 2009.

[29] Lametti, D. R., Nasir, S. M., & Ostry, D. J. "Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback," *J. Neuroscience*, 32, 9351-9358, 2012.

[30] Nasir, S. M., Darainy, M., & Ostry, D. J. "Sensorimotor adaptation changes the neural coding of somatosensory stimuli," *J. Neurophysiology*, 109, 2077-2085, 2013.