Relation of second formant trajectories to tongue kinematics

Austin Thompson, and Yunjung Kim

ARTICLES YOU MAY BE INTERESTED IN

Identifying bilingual talkers after a language switch: Language experience matters
The Journal of the Acoustical Society of America **145**, EL303 (2019); https://doi.org/10.1121/1.5097735

Sentence recognition for native and non-native English listeners in quiet and babble: Effects of contextual cues
The Journal of the Acoustical Society of America **145**, EL297 (2019); https://doi.org/10.1121/1.5097734

Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging
The Journal of the Acoustical Society of America **145**, EL284 (2019); https://doi.org/10.1121/1.5097377

Orthography and second language word learning: Moving beyond "friend or foe?"
The Journal of the Acoustical Society of America **145**, EL265 (2019); https://doi.org/10.1121/1.5094923

Prediction of speech intelligibility using pseudo-binaural room impulse responses
The Journal of the Acoustical Society of America **145**, EL329 (2019); https://doi.org/10.1121/1.5099169
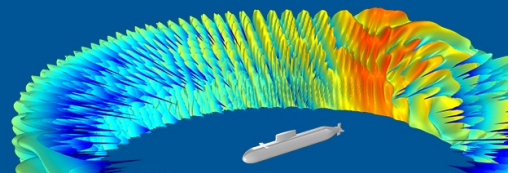
Discrimination of changes in spatial configuration for multiple, simultaneously presented sounds
The Journal of the Acoustical Society of America **145**, EL310 (2019); https://doi.org/10.1121/1.5098107

Relation of second formant trajectories to tongue kinematics

Austin Thompson and Yunjung Kim

Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, Louisiana 70803, USA
atho184@lsu.edu, ykim6@lsu.edu

Abstract: In this study, the relationship between the acoustic and articulatory kinematic domains of speech was examined among nine neurologically healthy female speakers using two derived relationships between tongue kinematics and F2 measurements: (1) second formant frequency (F2) extent to lingual displacement and (2) F2 slope to lingual speed. Additionally, the relationships between these paired parameters were examined within conversational, more clear, and less clear speaking modes. In general, the findings of the study support a strong correlation for both sets of paired parameters. In addition, the data showed significant changes in articulatory behaviors across speaking modes including the magnitude of tongue motion, but not in the speed-related measures.

1. Introduction

Although the complex relationship between speech acoustics and kinematics makes the mapping between acoustic and articulatory domains difficult, several paired parameters from each domain have been used to connect the articulatory and acoustic characteristics of speech. These parameters include (1) acoustic and articulatory vowel space area (VSA),[1,2] (2) formant frequencies and their corresponding lingual positions,[2–4] and (3) acoustic and articulatory Euclidean distance between two temporal points.[5]

Though few in number, studies investigating the relation between acoustics and articulatory kinematics have been examined in neurologically healthy speakers,[2,3,5] and even fewer studies have reported the relationship in disordered speech including apraxia of speech and dysarthria.[6–9] Particularly, initial research on dysarthria, a group of speech disorders secondary to various neurological conditions, has primarily pursued the following two goals: (1) to identify the overall nature of articulatory deficits in varying neuropathologies and (2) to examine articulatory modifications according to several speaking modes (e.g., slow, loud, or clear speech) in an effort to establish an empirical foundation for frequently used behavioral treatment approaches for dysarthria (e.g., LSVT Global[10]). The latter has been based partly on acoustic findings that some of those speaking modes yield positive changes in acoustic signals such as expanded vowel space, which subsequently enhances speech intelligibility.[11–14] However, one consistent challenge in this line of research is the complicated and inconsistent findings among kinematic-to-acoustic studies at least partly resulting from different methodologies including measurement points, target sounds and units (e.g., segment vs phrase), the varying neuropathologies under study, and interspeaker variability.[15] For example, Mefferd[7] found a strong relationship between acoustic vowel distance and tongue displacement among healthy speakers and speakers with Parkinson's disease (PD), but not among speakers with amyotrophic lateral sclerosis (ALS) when the speakers voluntarily modified their speaking rate. Conversely, Lee et al.[1] found a strong relationship during conversational speech between acoustic and kinematic VSA in both healthy speakers and speakers with ALS.

Our interest was focused on the relationship of formant *trajectories* with kinematic parameters especially given our long-term goal of expanding the current approach to include speakers with *movement* disorders (i.e., PD and ALS). Only limited data are available that directly examine articulatory gestures and formant trajectory measures such as transition extent and F2 slope. A study by Rong et al.[8] reported strong correlations between F2 slope and lingual and jaw movement in healthy speakers, but inconsistent correlations for speakers with cerebral palsy (CP). Yunusova et al.[9] found moderately strong associations for F2 slope and lingual speed and much weaker associations for F2 extent and lingual displacement in speakers with and

without ALS. Finally, relevant data exist that separately examine acoustic and articulatory measures without attempting to connect these measures. Tasko and Greilick[16] reported increases in duration and magnitude of acoustic and kinematic measures when speakers clearly produce a diphthong /ɑɪ/ of the word *combine*. However, no significant changes were found for speed-related measurements (i.e., F2 slope or lingual speed) between conversational and clear productions of the word.

The present study aimed to offer an account of variations in acoustic and articulatory measures associated with a voluntary change in speech clarity. In recognition that previous literature has focused on clear speech compared to casual speech, we added a unique mode, *less* clear speech, to investigate a full spectrum of speech clarity and the corresponding articulatory modifications when speakers voluntarily vary the degree of speech clarity.

Three acoustic parameters related to F2 transition were selected, and their associations with kinematic parameters were examined: transition duration, transition extent, and F2 slope. The F2 transition was the focus of this investigation because it is well-established that speech intelligibility is sensitive to F2 variables, such as extent and slope. Furthermore, F2 measures may be sensitive to the presence of dysarthria.[17–19] Two questions were explicitly posed. (1) What is the relationship between articulatory and acoustic transition measures (i.e., F2 slope vs articulatory speed and F2 extent vs articulatory displacement)? (2) What is the effect of speaking modes on acoustic and kinematic measures?

2. Methodology

2.1 Participants, speech tasks, and recording procedure

The participants included nine female speakers ranging in age from 19 to 23 ($M = 20.78$, SD = 1.20). All speakers were native speakers of Southern White English (SWE) and reported no history of communication problems. Speakers were asked to produce four repetitions of three sentences using three speaking modes: conversational, more clear, and less clear. The sentences were "*Buy* Bobby a puppy," "Tess told Dan to *stay* fit," and "*Carl* got a croaking frog." Three words including vocalic nuclei (*buy, stay, Carl*) requiring a relatively large change of vocal tract configuration were the analysis targets.[20] A different degree of clarity was elicited using a direct magnitude method with a modulus of 100.[21] For clear speech, speakers were instructed to speak with a clarity level of 200, as if they were speaking to someone with a hearing loss. For less clear speech, speakers were instructed to speak with a clarity level of 50, as if they were telling their friend a comment they did not want other people in the room to hear. The data for this study were collected as part of a larger study examining segment specific articulatory markers in speakers with dysarthria.

Acoustic and kinematic data were collected simultaneously in a sound-attenuating booth, with a sampling rate of 20 kHz and 16-bit resolution. An AKG C1000S microphone positioned approximately 30 cm from the speaker was used to record the speech stimuli. Kinematic data were collected using the Wave system (NDI, Canada) at a sampling rate of 100 Hz and were low-pass filtered at 10 Hz. Only lingual data from the tongue front (TF) (attached 2 cm from the tongue tip) and tongue blade (TB) (attached 3 cm from the tongue tip) were reported. Data from three reference sensors (one affixed to the bridge of a pair of glasses and two from a bite plate) were obtained from each participant to define the maxillary occlusal and midsagittal planes. Movement data from the jaw (adhered to the labial surface of the lower central incisors) were used to decouple the lingual sensors from the jaw using the estimated rotation method.[22]

2.2 Acoustic and kinematic analysis

The acoustic data were segmented using the spectrographic view in Time-Frequency Analysis Software Program for 32-bit Windows (TF32).[23] F2 slopes (Hz/ms) were calculated using the 20/20 rule.[24] That is, using the linear predictive coding algorithm of TF32, the F2 transition onset and offset are identified by a change of at least 20 Hz during a 20 ms increment. Kinematic data were extracted based on acoustic segmentation and were used to calculate x- (anteroposterior plane) and xy- (vertical and anteroposterior plane) displacement (mm) (i.e., straight line Euclidean distance between onset and offset of movement) and speed (mm/ms) (i.e., displacement/duration) for the two marker locations (i.e., TF, TB).

Table 1. Pearson's correlation (r) for each set of measures across speaking modes. N/S = no significance, TF = tongue front, TB = tongue back, F2 = second formant.

Paired parameters		Buy		Stay		Carl	
		r	Sig.	r	Sig.	r	Sig.
F2 extent x	TF x-displacement	0.37	<0.00625	0.51	<0.00625	0.31	<0.00625
	TF xy-displacement	0.75	<0.00625	0.44	<0.00625	0.43	<0.00625
	TB x-displacement	0.57	<0.00625	0.51	<0.00625	0.31	<0.00625
	TB xy-displacement	0.75	<0.00625	0.52	<0.00625	0.39	<0.00625
F2 Slope x	TF x-speed	0.41	<0.00625	0.24	N/S	−0.09	N/S
	TF xy-speed	0.50	<0.00625	0.10	N/S	−0.06	N/S
	TB x-speed	0.52	<0.00625	0.25	N/S	−0.10	N/S
	TB xy-speed	0.42	<0.00625	0.25	N/S	0.03	N/S

2.3 Statistical analysis

Prior to analysis, measures were normalized using z-scores. Relationships between the acoustic and kinematic data of four trials from each speaker (n = 36 for each word) for three speaking modes and three words were examined using Pearson's correlation (r). Multiple comparisons were accounted for using the Bonferroni method. The acoustic-to-articulatory relationships were investigated for the following four pairs of parameters for both TF and TB: F2 slope vs x-speed, F2 slope vs xy-speed, F2 extent vs x-displacement, and F2 extent vs xy-displacement (modified from Yunusova et al.[9]). Within the current study, x-movement was separately investigated due to the known primary effects of tongue advancement on F2. Additionally, a series of repeated measures analysis of variance (RM-ANOVA) tests were used to assess the effect of speaking mode on each target acoustic and kinematic measure. Measures with significant speaking mode effects were subjected to post hoc pairwise comparisons to reveal the significance among speaking modes. Post hoc comparisons were adjusted using the Bonferroni method to control for multiple comparisons.

3. Results

Table 1 summarizes the correlations between target acoustic and kinematic measures, among which were several significant relationships. For F2 extent and articulatory displacement, significant relationships were found for all words, with a relatively large variance (10%–56%) of tongue displacement associated with the variance in F2 extent. Meanwhile, F2 slope and articulatory speed relationships were significant only for buy, with 17%–28% of the variance in tongue speed associated with the variance of F2 slope. The correlations between target acoustic and kinematic measures were comparable for x- and xy-movement within stay and Carl. However, the relationship between target measures was stronger for xy-movement within buy.

Table 2 summarizes the results of the RM-ANOVA indicating the effect of speaking modes on each measure for each word. Speaking modes elicited significantly different measures with the exception of TF x-speed for buy, and F2 extent, slope, and TF and TB xy-speed for Carl.

4. Discussion

The findings of the study support a weak-to-moderate correlation, in general, between the sets of paired parameters from acoustic and articulatory domains, at least examined by the selected measures. In addition, speech style from less-to-more clear yielded significant changes in articulatory behaviors such as the magnitude of tongue motion, but not in the speed-related measures (i.e., F2 slope or lingual speed).

4.1 Acoustic and articulatory relations

The correlation of two sets of acoustic and articulatory measurements reached significance in every context (i.e., across words and measurement points) except for the relation between F2 slope and lingual speed in stay and Carl. A stronger correlation was found for F2 extent and lingual displacement compared to F2 slope and articulatory speed (Fig. 1). Interestingly, Yunusova et al.[9] observed the opposite trend (i.e., stronger relations were observed for F2 slope and lingual speed compared to F2 extent and lingual displacement). This inconsistent finding may be partly due to the different speaking modes examined (i.e., clarity-related speaking modes vs speaking rate[9]).

Table 2. Results of RM-ANOVA for each measure among the speaking modes of each word. N/S = no significance; TF = tongue front; TB = tongue back; df = 2.

Measure		Buy SS	F	p	ηp^2	Stay SS	F	p	ηp^2	Carl SS	F	P	ηp^2
	Duration (ms)	34.85	31.59	<0.01	0.474	41.59	35.76	<0.01	<0.01	14.05	11.22	<0.01	<0.01
	F2 extent (Hz)	48.60	38.60	<0.01	0.524	5.80	9.61	<0.01	<0.01	2.08	3.27	N/S	0.052
	F2 slope (Hz/ms)	3.86	6.09	<0.01	0.148	28.15	33.43	<0.01	<0.01	1.25	1.06	N/S	0.344
TF	x-displacement (mm)	13.37	8.20	<0.01	0.190	6.74	10.99	<0.01	<0.01	18.39	10.31[a]	<0.01	<0.01
	xy-displacement (mm)	46.62	34.74	<0.01	0.498	2.24	4.41	<0.05	0.017	11.27	7.60	<0.01	<0.01
	x-speed (mm/ms)	0.03	0.02	N/S	0.001	2.78	4.07	<0.05	0.026	8.77	3.47	<0.05	0.039
	xy-speed (mm/ms)	6.39	4.60	<0.05	0.116	2.38	3.54	<0.05	0.038	6.61	2.80	N/S	0.071
TB	x-Displacement (mm)	31.55	30.25	<0.01	0.464	2.02	3.75	<0.05	0.035	22.11	13.28	<0.01	<0.01
	xy-displacement (mm)	75.99	46.08	<0.01	0.568	3.71	10.04	<0.01	<0.01	7.99	8.55	<0.01	<0.01
	x-speed (mm/ms)	4.70	5.42	<0.05	0.134	4.22	5.87[a]	<0.05	0.01	9.60	3.97	<0.05	0.026
	xy-speed (mm/ms)	21.09	16.96	<0.01	0.326	5.15	6.50	<0.01	<0.01	6.04	2.69	N/S	0.076

[a]df = 1.

Additionally, the inconsistent finding may be due to the different phonetic contexts examined (i.e., /aI/, /eI/, and /aɹl/ vs /oI/, /jæ/, and /dɔ/[9]).

We were also interested in the relationship between acoustic and articulatory domains within one- and two-dimensional movement planes (x- and xy-planes, respectively). The rationale was to compare the well-established theoretical and empirical relationship between F2 and anteroposterior movement (e.g., perturbation theory) to the relatively recent reports of more complicated relations. Our data showed a stronger relationship between domains with xy-measurements compared to x-measurements, consistent with Lee et al.[2]

Although a direct comparison was not conducted, there appears to be a word effect, consistent with previous articulatory[3,9] and acoustic[20,25] studies. That is, buy and Carl demonstrated the strongest and weakest correlations, respectively. This may align with the speculation that the diphthongs requiring a greater degree of F2 change and slope show greater correlation to movement measures and may be more sensitive to the presence and severity of dysarthria.[20] In our data, the word buy demonstrated the greatest acoustic and articulatory magnitude and the highest correlation between the paired acoustic and articulatory parameters. However, within the current study, variations in /ɹ/ production (e.g., bunched vs retroflexed) were not controlled for, which may have confounded the results relating to Carl.

4.2 Effects of speaking mode

In general, across speaking modes, we observed the expected scaling effect on the measures. That is, while clear speech elicited greater measure values consistent with Tasko and Greilick,[16] less clear speech elicited shorter and smaller measures relative to conversational speech. The RM-ANOVA revealed lingual displacement measures were
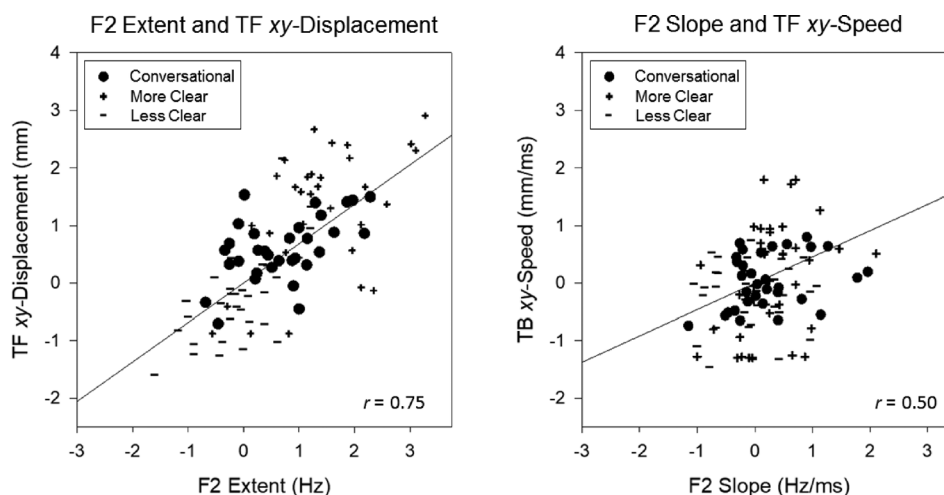

Fig. 1. Scatter plot of acoustic and articulatory measures across speaking modes for buy.
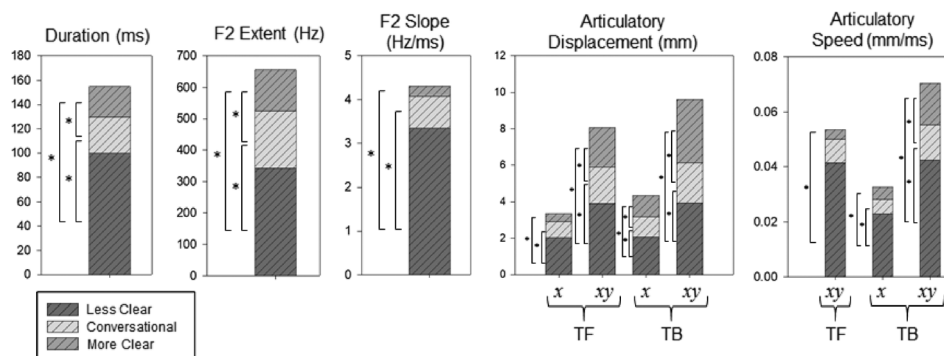
Fig. 2. Cumulative bar graphs displaying the *post hoc* pairwise comparison results following the RM-ANOVA results for buy (TF: tongue front, TB: tongue back, ms: milliseconds, Hz: Hertz, *$p < 0.05$). As seen in the left panel, the duration for less clear, conversational, and more clear speaking modes are 102, 129, and 151 ms, respectively.

significantly affected by speaking modes. However, in some cases, measures of rate (i.e., F2 slope and articulatory speed) were not significantly affected by speaking modes although speakers successfully modified their duration and lingual displacement (or extent in the case of F2 slope). This is likely due to duration and displacement measures being scaled to the same degree. Tasko and Greilick[16] also observed statistically insignificant changes in F2 slope between conversational and more clear speaking modes. As previously mentioned, F2 slope is a well-established measure reflecting the degree of speech intelligibility.[20] However, within the current study, F2 slope did not always statistically capture the observed change in speech clarity. This finding may suggest a possible ceiling effect for F2 slope's ability to capture speech intelligibility.

Figure 2 displays the *post hoc* pairwise comparison results for each measure for the word *buy*. The measures for *buy* that did not demonstrate a significant speaking mode effects (as seen in Table 2) were excluded from Fig. 2. Across words, *buy* demonstrated the strongest significant articulatory and acoustic relations (as seen in Table 1) and the greatest number of significant contrasts within each measure as indicated by *post hoc* pairwise comparisons.

In many cases, acoustic and articulatory measures were significantly different among the three speaking modes. Interestingly, in some instances, there was no significant change from conversational to more clear speech, despite significant changes from conversational to *less* clear speech (see F2 slope, TF *x*-displacement, and TB *x*-speed in Fig. 2). Mefferd and Green[5] describe a hypothetical speech clarity continuum within the context of phonetic specification and variability. On this continuum, typical (or conversational) speech is in the center with ideal (or clear) speech at one end, and dysarthric (or less clear) speech at the other end. Applying this model to the current findings reveals that conversational speech is located not in the center, but rather closer to the more clear end of the continuum. For example, F2 slopes for conversational speech are more like clear speech than less clear speech (as seen in Fig. 2). For speakers with dysarthria, it is worthy to investigate where conversational speech is located along the speech clarity continuum in relation to their more clear and less clear speech.

In conclusion, the current study found significant linear articulatory and acoustic relations. These findings serve as foundational work for future research investigating articulatory and acoustic relations for clarity-related speaking modifications within speakers with dysarthria.

References and links

[1]J. Lee, M. A. Littlejohn, and Z. Simmons, "Acoustic and tongue kinematic vowel space in speakers with and without dysarthria," Int. J. Speech-Lang. Pathol. **19**(3), 195–204 (2017).

[2]J. Lee, S. Shaiman, and G. Weismer, "Relationship between tongue positions and formant frequencies in female speakers," J. Acoust. Soc. Am. **139**(1), 426–440 (2016).

[3]C. Dromey, G.-O. Jang, and K. Hollis, "Assessing correlations between lingual movements and formants," Speech Commun. **55**(2), 315–328 (2013).

[4]G.-O. Jang, "The relationship between acoustic and kinematic measures of diphthong production," Master's thesis, Brigham Young University, Provo, Utah (2010).

[5]A. S. Mefferd and J. R. Green, "Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations," J. Speech Lang. Hear. Res. **53**, 1206–1219 (2010).

[6]W. F. Katz, S. V. Bharadwaj, and B. Carstens, "Electromagnetic articulography treatment for an adult with Broca's aphasia and apraxia of speech," J. Speech Lang. Hear. Res. **42**, 1355–1366 (1999).

[7]A. Mefferd, "Articulatory-to-acoustic relations in talkers with dysarthria: A first analysis," J. Speech Lang. Hear. Res. **58**(3), 576–589 (2015).

[8] P. Rong, T. Loucks, H. Kim, and M. Hasegawa-Johnson, "Relationship between kinematics, F2 slope and speech intelligibility in dysarthria due to cerebral palsy," Clin. Ling. Phon. **26**(9), 806–822 (2012).

[9] Y. Yunusova, J. R. Green, L. Greenwood, J. Wang, G. L. Pattee, and L. Zinman, "Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis," Folia Phon. Log. **64**(2), 94–102 (2012).

[10] L. Ramig, O. C. Fox, and S. Sapir, "Parkinson's disease: Speech and voice disorders and their treatment with the Lee Silverman Voice Treatment," Sem. Speech Lang. **25**(2), 169–180 (2004).

[11] P. McRae, A. K. Tjaden, and B. Schoonings, "Acoustic and perceptual consequences of articulatory rate change in Parkinson disease," J. Speech Lang. Hear. Res. **45**(1), 35–50 (2002).

[12] K. Tjaden and V. Martel-Sauvageau, "Consonant acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear and loud speaking conditions," Am. J. Speech-Lang. Pathol. **26**, 569–582 (2017).

[13] K. Tjaden and G. Wilding, "Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings," J. Speech Lang. Hear. Res. **47**, 766–783 (2004).

[14] J. A. Whitfield and A. M. Goberman, "Articulatory-acoustic vowel space: Application to clear speech in individuals with Parkinson's disease," J. Commun. Disord. **51**, 19–28 (2014).

[15] Y.-C. Tsao and G. Weismer, "Interspeaker variation in habitual speaking rate," J. Speech Lang. Hear. Res. **40**, 858–866 (1997).

[16] S. M. Tasko and K. Greilick, "Acoustic and articulatory features of diphthong production: A speech clarity study," J. Speech Lang. Hear. Res. **53**(1), 84–99 (2010).

[17] J. F. Kent, R. D. Kent, J. C. Rosenbeck, G. Weismer, R. Martin, R. Sufit, and B. Brooks, "Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis," J. Speech Hear. Res. **35**, 723–733 (1992).

[18] Y. Kim, R. D. Kent, and G. Weismer, "An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria," J. Speech Lang. Hear. Res. **54**(2), 417–429 (2011).

[19] K. L. Lansford and J. M. Liss, "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification," J. Speech Lang. Hear. Res. **57**, 57–67 (2014).

[20] Y. Kim, G. Weismer, R. Kent, and J. Duffy, "Statistical models of F2 slope in relation to severity of dysarthria," Folia Phon. Log. **61**, 329–335 (2009).

[21] G. Weismer and J. Laures, "Direct magnitude estimates of speech intelligibility in dysarthria," J. Speech Lang. Hear. Res. **45**, 421–433 (2002).

[22] J. R. Westbury, M. J. Lindstrom, and M. D. McClean, "Tongues and lips without jaws: A comparison of methods for decoupling speech movements," J. Speech Lang. Hear. Res. **45**, 651–662 (2002).

[23] P. Milenkovic, "TF 32" [computer program], University of Wisconsin–Madison, http://userpages.chorus.net/cspeech/ (Last viewed December 2, 2015).

[24] G. Weismer, R. D. Kent, M. Hodge, and R. Martin, "The acoustic signature for intelligibility test words," J. Acoust. Soc. Am. **84**(4), 1281–1291 (1988).

[25] G. Weismer, Y. Yunusova, and K. Bunton, "Measures to evaluate the effects of DBS on speech production," J. Neuroling. **25**, 74–94 (2012).