Voice Onset Time for the Word-Initial Voiceless Consonant /t/ in Japanese Spasmodic Dysphonia—A Comparison With Normal Controls

*,†Saori Yanagida, *,†Noriko Nishizawa, †Kenji Mizoguchi, †Hiromitsu Hatakeyama, and †Satoshi Fukuda,
*†Sapporo, Hokkaido, Japan

Summary: Objectives. Voice onset time (VOT) for word-initial voiceless consonants in adductor spasmodic dysphonia (ADSD) and abductor spasmodic dysphonia (ABSD) patients were measured to determine (1) which acoustic measures differed from the controls and (2) whether acoustic measures were related to the pause or silence between the test word and the preceding word.

Methods. Forty-eight patients with ADSD and nine patients with ABSD, as well as 20 matched normal controls read a story in which the word "taiyo" (the sun) was repeated three times, each differentiated by the position of the word in the sentence. The target of measurement was the VOT for the word-initial voiceless consonant /t/.

Results. When the target syllable appeared in a sentence following a comma, or at the beginning of a sentence following a period, the ABSD patients' VOTs were significantly longer than those of the ADSD patients and controls. Abnormal prolongation of the VOTs was related to the pause or silence between the test word and the preceding word.

Conclusions. VOTs in spasmodic dysphonia (SD) may vary according to the SD subtype or speaking conditions. VOT measurement was suggested to be a useful method for quantifying voice symptoms in SD.

Key Words: Adductor spasmodic dysphonia–Abductor spasmodic dysphonia–Acoustic analysis–Voice onset time.

INTRODUCTION

Spasmodic dysphonia (SD) is defined as a focal dystonia of the laryngeal muscles.[1] Criteria for SD as an independent voice disorder have not yet been established, and the current standardized assessment method is insufficient for the objective evaluation of its severity or effects. SD subtypes include adductor spasmodic dysphonia (ADSD), abductor spasmodic dysphonia (ABSD), and mixed spasmodic dysphonia.[2] Hyperadduction of the true vocal folds occurs during speech in ADSD, and the voice is hoarse and strained. In contrast, endoscopic observation of the vocal folds in ABSD patients during running speech has revealed that, with each breathy air release, a synchronous and untimely abduction of the true and false vocal folds occurs. Various methods have been used to characterize the severity of dysphonia and to evaluate the effects of treatment in patients with SD. These have included the voice handicap index[3] or voice-related quality of life,[4,5] and the acoustic,[2,6–9] aerodynamic,[10–13] and perceptual assessments[10,14,15] such as the Unified Spasmodic Dysphonia Rating Scale[16] or the Consensus Auditory Perceptual Evaluation of Voice.[14] Acoustic methods have focused on the assessment of vocal spasm severity, eg, the number of voice breaks, fundamental frequency (F_0), maximum phonation time (MPT), and voice onset time (VOT).

Edgar et al[2] measured VOTs for three voiceless consonants /k, p, t/ in two different sentences. As a result, the VOTs for voiceless consonants in ABSD patients were significantly longer than those in the controls. The purpose of this study was to investigate how the voice control difficulties in SD patients influence the segmental structure of Japanese speech production by means of VOT measurement with special reference to pauses or silence preceding the test word.

METHODS

Participants

Forty-eight patients with ADSD (40 women and eight men, mean age of 29.2 ± 9.9 years) and nine patients with ABSD (eight women and a man, mean age of 27.1 ± 6.5 years), evaluated at the Health Sciences University of Hokkaido Hospital between April 2008 and July 2013, participated as the subjects in this study. The diagnosis of SD was made by an otolaryngologist in conjunction with a speech-language-hearing therapist (ST) based on the results of a detailed clinical history, endoscopic observation of the larynx combined with videostroboscopy, and the acoustic and aerodynamic analysis of each patient's voice. Twenty normal controls (NC) without voice disorders (10 women and 10 men, mean age of 27.0 ± 6.8 years) were also tested.

Criteria for inclusion in this study for all participants were (1) normal resonance, and language ability as judged by a licensed, certified, ST; (2) no diagnosis of neurological disorder such as Parkinson disease, cerebrovascular attack (CVA), or essential tremor; and (3) no history of professional singing or voice training. Other specific criteria for the participants with SD included (1) no previous treatment with recurrent laryngeal nerve surgery, (2) no previous treatment with botulinum toxin injection, (3) evidence of repetitive hyperadduction/abduction during speech on videoendoscopic examination, (4) symptomatic at the time of testing, and (5) no voice change on stimulus with facilitation techniques such as muscle relaxation and resonance focus.

Accepted for publication September 25, 2014.
From the *Department of Communication Disorders, School of Psychological Sciences, Health Sciences University of Hokkaido, Sapporo, Hokkaido, Japan; and the †Department of Otolaryngology-Head and Neck Surgery, Hokkaido University Graduate School of Medicine, Sapporo, Hokkaido, Japan.
Address correspondence and reprint requests to Saori Yanagida, Department of Communication Disorders, School of Psychological Sciences, Health Science University of Hokkaido, 2-5, Ainosato, Sapporo, Hokkaido 0028072, Japan. E-mail: s.yanagi@hoku-iryo-u.ac.jp
Journal of Voice, Vol. 29, No. 4, pp. 450-454
0892-1997/$36.00
© 2015 The Voice Foundation
http://dx.doi.org/10.1016/j.jvoice.2014.09.028

Speech materials

Our diagnostic protocol consisted of several speech tasks in a natural speaking voice. The tasks included production of (1) sustained vowels/i, e, a, o, u/, (2) MPT for /a/, (3) counting from 0 to 50, (4) reading a story, and (5) repeating sentences which elicit voice symptoms associated with SD (cf., adductor or abductor sentences).

From our diagnostic protocol, we selected the reading task for analysis. Text for reading was "The North Wind and the Sun," ("Kitakaze to Taiyo" in Japanese). The word "Taiyo" (the Sun) was repeated three times, each time differentiated by its position in the sentence: (1) in a sentence without preceding punctuation, (2) in a sentence following a comma, and (3) at the beginning of a sentence following a period (Table 1). The target of the measurement was the VOT for the word-initial voiceless consonant /t/ in "Taiyo."

Acoustic measures

We performed acoustic recordings at the first medical examination. All patients were recorded in a quiet room using a digital audio recorder (Frontier TASCAM HD-P2 portable stereo audio recorder, TEAC Corp, Tokyo, Japan) and a condenser microphone (AKG C1000 S CE Harman International, Stamford, CT) under similar conditions. The microphone was maintained at a distance of 20 cm from the lips. The acoustic waveform was displayed using the *Praat* free software program,[17] and the pause or silence and VOTs were measured. The pause or silence was measured as the length of time between the spike of the test word and the final pulse of the preceding word (Figure 1). According to Edgar et al,[2] VOT for the test word was measured as the interval between the release of an oral constriction and the start of glottal pulsing.

Reliability

Intrajudge reliability was determined by the measurement of VOT. The same investigator rescored all measures without access to the original scores. Pearson r correlations were calculated for the two sets of measures. Correlation coefficients showed high intrameasurer reliability ranging from $r = 0.924$ to $r = 1.000$.

Statistical analysis

Using the statistical analysis tool SPSS Statistics 21, we performed a two-way repeated measures analysis of variance (ANOVA) with sentence type (in a sentence without preceding punctuation, in a sentence following a comma, and at the beginning of a sentence following a period) and subject type (ADSD, ABSD, and NC) as factors. A Greenhouse-Geisser correction for sphericity was used where necessary. To follow-up this interaction, a one-way repeated measures ANOVA was conducted separately for each level of sentence type and subject type using Bonferroni corrected P values.

RESULTS

Pause or silence

The pause or silence between the test word and the preceding word tended to be prolonged successively under conditions 1 to 3. In ADSD patients, the mean pause or silence were 91.27 ± 90.75 ms under condition 1, 256.75 ± 162.45 ms under condition 2, and 894.73 ± 376.74 ms under condition 3. Similarly, in ABSD patients, values were 108.78 ± 69.18 ms, 239.11 ± 125.02 ms, and 826.00 ± 194.49 ms, respectively, whereas those in the controls were 76.05 ± 19.25 ms, 254.30 ± 207.67 ms, and 1014.55 ± 301.66 ms (Figure 2).

Voice onset time

In ADSD patients, the mean VOT was 26.38 ± 17.38 ms under condition 1, 34.85 ± 23.40 ms under condition 2, and 45.25 ± 29.31 ms under condition 3. Similarly, in ABSD patients, the values were 46.00 ± 48.82 ms, 98.33 ± 67.96 ms, and 134.33 ± 103.03 ms, respectively, whereas those in the controls were 17.95 ± 9.57 ms, 26.40 ± 12.79 ms, and 25.65 ± 10.35 ms.

The two-way ANOVA showed a significant main effect for subject type ($F(2, 74) = 28.24$, $P < 0.001$), but no main effect

TABLE 1.
Stimulus Sentences (From "The North Wind and the Sun")

Condition	Sentences
1. Without punctuation	Aruhi Kitakaze to **Taiyo** ga chikara-kurabe o shimashita. (The North Wind and the Sun had a quarrel about which of them was the stronger.)
2. With a comma	Kondo wa, **Taiyo** no ban ni narimashita. (Then the Sun began to shine.)
3. With a Period	Kondo wa, Taiyo no ban ni narimashita. **Taiyo** wa kumo no aida kara kao o dashite atatakana hizashi o okurimashita. Tabibito wa dandan yoi kokoromochi ni nari shimai niwa gaitou o nugisute mashita. (Then the Sun began to shine. At first his beams were gentle, and in the pleasant warmth after the bitter cold of the North Wind, the Traveler unfastened his cloak and let it hang loosely from his shoulders. The Sun's rays grew warmer and warmer. The man took off his cap and mopped his brow. At last he became so heated that he pulled off his cloak, and, to escape the blazing sunshine, threw himself down in the welcome shade of a tree by the roadside.)

Notes: The bold words (Taiyo) are targets to measure VOTs.

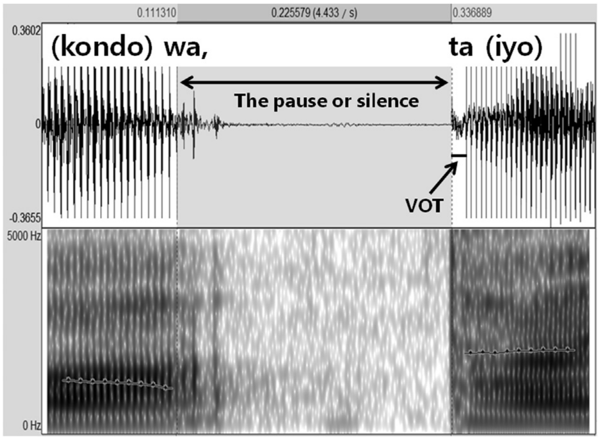

FIGURE 1. An example of the measures of the pause or silence and VOT under the condition 2 in controls.

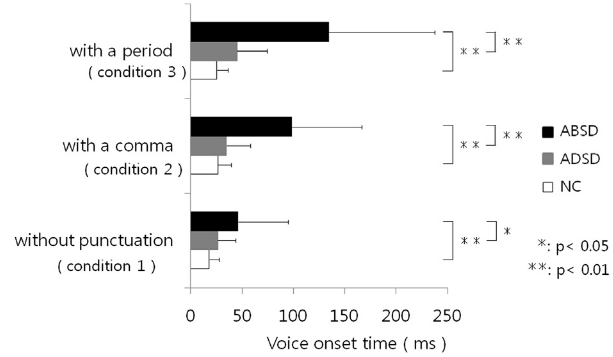

FIGURE 3. Comparison among subject groups. The mean voice onset time (VOT) (in milliseconds) in each patient with ABSD and ADSD and NC subjects. In all of three sentence types, the ABSD's VOTs were significantly longer than that of the ADSD and of NC.

for sentence type ($F(1.7, 126.3) = 26.23$, $P < 0.001$). The interaction between the two factors was shown to be significant ($F(3.4, 126.3) = 8.02$, $P < 0.001$). To follow-up this interaction, a one-way repeated measures ANOVA using Bonferroni correction was conducted separately for each sentence type and subject type.

Comparison among subject groups

We compared subject types for each sentence type. For all of three sentence types, the VOTs in the ABSD patients were significantly longer than those in the ADSD patients and NC (Figure 3). Under condition 1, significant differences were observed between subject types ($F(2, 74) = 5.16$, $P < 0.01$). The VOTs in the ABSD patients were significantly longer than those in the ADSD patients ($P = 0.046$) and controls ($P = 0.006$). Significant differences were also observed between subject types under condition 2 ($F(2, 74) = 20.02$, $P < 0.001$). Again, the VOTs in the ABSD patients were significantly longer than those in the ADSD patients ($P < 0.001$) and controls ($P < 0.001$). Similarly, there were significant differences between subject types under condition 3 ($F(2, 74) = 22.35$, $P < 0.001$), with the VOTs in the ABSD patients were significantly longer than those in the ADSD patients ($P < 0.001$) and controls ($P < 0.001$).

Comparison within subject groups

Intragroup comparisons were also examined among sentence types. Significant differences among sentence type were observed for the ADSD patients ($F(2, 141) = 7.53$, $P = 0.001$). VOTs under condition 3 were significantly longer than those under condition 1. No significant differences under any conditions were observed for either the ABSD patients ($F(2, 24) = 3.02$, $P = 0.067$) or controls ($F(2, 57) = 3.62$, $P = 0.055$) (Figure 4).

DISCUSSION

In this study, one of the criteria for inclusion was the absence of any diagnosis of neurological disorders such as Parkinson disease, CVA, or essential tremor. We also excluded subjects

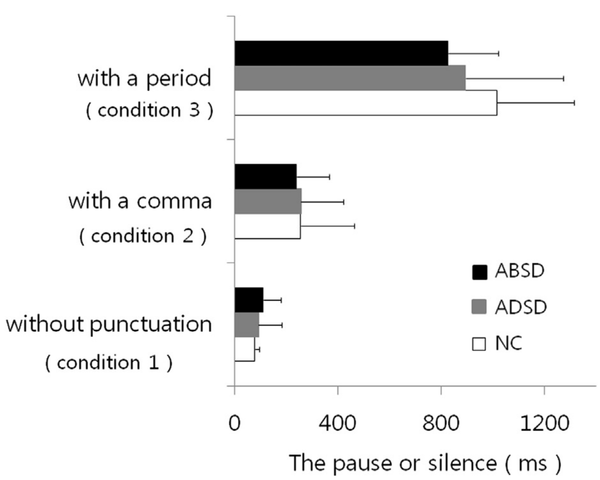

FIGURE 2. The mean pause or silence (in milliseconds) in each patient with abductor (ABSD) and adductor spasmodic dysphonia (ADSD) and normal control (NC) subjects. The pause or silence tended to be prolonged gradually under the condition 1 to 3.

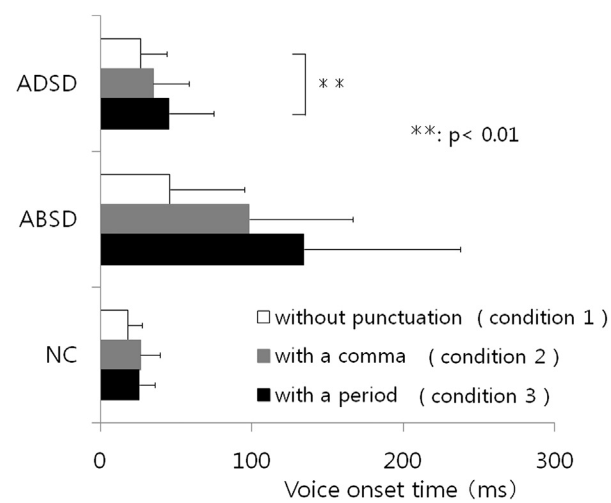

FIGURE 4. Comparison within subject groups. The mean VOT (in milliseconds) in each patient with ABSD and ADSD and NC. In ADSD, VOTs under the condition 3 were significantly longer than those of under condition 1.

with "essential voice tremor." Voice tremor is characterized by periodic fluctuations in the voice frequency or intensity at a rate between 3 and 12 Hz.[18] The condition is differentiated from SD, which is characterized by intermittent but nonperiodic voice stoppages or breaks. In SD, a tremorous voice may appear as abnormal phonatory behavior affecting the perception and/or differential diagnosis of ADSD from ABSD. However, the symptoms of the subjects in this study were clear cut and the differential diagnoses among ADSD, ABSD, and voice tremor could be clearly made.

Pauses are generally classified into two categories: those with breaths and those without breaths. The former are often called "physiological pauses," and latter "syntactic pauses." In Japanese speech production, breaths are a prerequisite condition at the beginning of a sentence following a period, but breaths are not necessary in sentences following a comma, or in sentences without preceding punctuation.[19] Taking these Japanese linguistic features into consideration, we applied three sets of conditions: segmentation without punctuation, segmentation with a comma, and segmentation with period. These types of segmentation reflect the semantic or grammatical distance between the words. Thus, breathing is assumed to depend on syntactic rather than physiological factors. In this study, each subject read a text at a natural speaking rate, without any order in the timing of pauses or silences. Judging perceptually, all subjects read without breaths under condition 1 (without punctuation), some of subjects read with breaths under condition 2 (with a comma), and all subjects read with breaths under condition 3 (with a period). As a result, the average duration of the pause in sentences with a period tended to be longer than that in sentences with a comma and without punctuation. These results corresponded to those of previous research.[19]

The purpose of this study was to demonstrate the characteristic voice symptoms in SD by measuring VOTs, and to investigate whether a pause or silence between the test word and the preceding word results in variations in the VOTs. The VOTs in the subjects with ABSD were significantly longer than those in NC and the subjects with ADSD under all three conditions. A comparison among subjects with ADSD revealed that the VOTs were significantly longer under condition 3, where a breath preceded the test word, than under condition 1, where a breath did not exist before the test word. Considering these results, prolongation of the VOTs for word-initial voiceless consonants seems to be related to the preceding pause or silence in ABSD subjects. In contrast, the VOTs in the ADSD patients were not significantly longer than those in the controls. Erickson[20] attempted to manipulate syntactic complexity and reported that sign expression for ADSD increased with syntactic complexity and in the voiced condition. Cannito et al[21] reported that for ADSD, voice quality in the voiceless condition was superior to that in the voiced condition. To test the unique influence of pause or silence between the test word and the preceding word, we would have needed to make identical phonetic environments (voiced vs voiceless, consonant vs vowels) in the words preceding the pause. Furthermore, we did not specifically control for syntactic complexity. Thus, strictly, the influence of pause time cannot be distinguished from that of phonetic environment or syntactic complexity although there is also a possibility that sample size affected our results. P values in the ABSD ($P = 0.067$) and in the controls ($P = 0.055$) compared with the ADSD results are very close to the level of "significance," and their nonsignificance can be explained by the fact that there were fewer subjects in both the ABSD ($n = 9$) and control ($n = 20$) groups versus the ADSD group ($n = 48$). Because of the small population in the ABSD group, in particularly, the nonsignificance of the findings for the ABSD and control groups could simply represent a type II error related to differences in sample size among the groups. By combining other analysis tools, these errors may be averted.

The observed difficulty in switching between the component movements of a complex task are consistent with the behavior seen in a wide range of dystonia, including focal, action-induced dystonia, and may be related to diminished cortical movement.[21] The larynx controls the distinction between voiced and voiceless consonants based on reciprocal activation of the posterior cricoarytenoid (PCA) and arytenoid muscles. Transient activation of the PCA muscle and restraint of the arytenoid muscle cause abduction of the vocal folds during speech in voiceless consonants.[22] That is, the PCA muscle plays a positive role in abduction activity not only during respiration but also during the production of voiceless consonants in speech.

In the present study, conditions 2 and 3 were both accompanied by a pause which might or might not be associated with breath. In such conditions, complex cooperation between the PCA and arytenoid muscles is required or both breathing and phonation. Supposing that a disturbance in the motor subroutine[23] in dystonia affects the voice symptoms of SD, abnormal prolongation of the VOTs in ABSD and ADSD might be explained by distortion of the central nervous system control of the coordination of the adductor and abductor laryngeal muscles.

The respiratory center is known to be located in the medulla oblongata, which is the lowest part of the brain stem. The localization of the central pattern generator for voicing is not yet identified in humans. Katada et al[24] reported that repetitive electrical stimulation to the midbrain periaqueductal gray and the ventrolateral pons evokes natural sounding vocalization in cat. These findings imply that the pattern generator for spontaneous phonation exists in the brain stem in mammals. Coordination between phonation and respiration for "speech" production must be differentiated from that for the spontaneous cries of a decorticated cat. However, we can at least assume that the failure of coordination between respiratory and phonatory motor controls in ABSD subjects lead to abnormal prolongation of VOTs, particularly after the pause with breath.

The present study demonstrated that, in ABSD patients, the VOTs were prolonged under conditions in which a breath did not necessarily precede the target. As for the conditions without a breath, we cannot attribute the prolongation of VOT to the failure of respiration/phonation control. Phonetically relevant properties of speech, including VOT, are influenced by individual differences in speakers, such as speaking rate and behavioral variations in articulation.[25,26] It is probable that the distinction between the semantic or grammatical distance between words might affect laryngeal adductor/abductor motor coordination in SD subjects although identical phonetic environments in the words preceding the pause or control of syntactic complexity is required to clarify this. These results suggest that VOT measurement helps the differential diagnosis of mild ABSD and ADSD.

CONCLUSIONS

When the target syllables appeared in a sentence following a comma, or at the beginning of a sentence following a period, the VOT for the word-initial voiceless consonant in ABSD patients were significantly longer than those in the controls. Abnormal prolongation of the VOTs was related to the pause or silence between the test word and the preceding word. VOTs in SD may vary according to the SD subtype or speaking conditions. VOT measurement of the word-initial voiceless consonant was suggested to be a useful method for quantifying voice symptoms in SD.

REFERENCES

1. Zwirner P, Murry T, Swenson M, Woodson GE. Acoustic changes in spasmodic dysphonia after botulinum toxin injection. *J Voice*. 1991;5:78–84.
2. Edgar JD, Spanienza CM, Bidus K, Ludlow CL. Acoustic measures of symptoms in abductor spasmodic dysphonia. *J Voice*. 2001;15:362–372.
3. Morzaria S, Damrose EJ. A comparison of the VHI, VHI-10, V-RQOL for measuring the effect of botox therapy in adductor spasmodic dysphonia. *J Voice*. 2012;26:378–380.
4. Hogikyan ND, Wodchis WP, Spak C, Kileny PR. Longitudinal effects of botulinum toxin injections on voice-related quality of life (V-RQOL) for patients with adductor spasmodic dysphonia. *J Voice*. 2001;15:576–586.
5. Rubin AD, Wodchis WP, Spak C, Kileny PR, Hogikyan ND. Longitudinal effects of botox injections on voice-related quality of life (V-RQOL) for patients with adductor spasmodic dysphonia. *Arch Otolaryngol Head Neck Surg*. 2004;130:415–420.
6. Murry T, Woodson GE. Combined-modality of adductor spasmodic dysphonia with botulinum toxin and voice therapy. *J Voice*. 1995;9:460–465.
7. Crevier-Buchman L, Laccourreye O, Papon JF, Nurit D, Brasnu N. Adductor spasmodic dysphonia: case reports with acoustic analysis following botulinum toxin injection and acupuncture. *J Voice*. 1997;11:232–237.
8. Sapiena CM, Murry T, Brown WS Jr. Variations in adductor spasmodic dysphonia: acoustic evidence. *J Voice*. 1998;12:214–222.
9. Sapoenza CM, Walton S, Murry T. Adductor spasmodic dysphonia and muscle tension dysphonia: acoustic analysis of sustained phonation and reading. *J Voice*. 2000;14:502–520.
10. Merson RM, Ginsberg AP. Spasmodic dysphonia: abductor type. A clinical report of acoustic aerodynamic and perceptual characteristics. *Laryngoscope*. 1979;89:129–139.
11. Woo P, Colton R, Casper J, Brewer D. Analysis of spasmodic dysphonia by aerodynamic and laryngostroboscopic measurements. *J Voice*. 1992;6:344–351.
12. Cantarella G, Berlusconi A, Maraschi B, Ghio A, Barbieri S. Botulinum toxin injection and airflow stability in spasmodic dysphonia. *Otolaryngol Head Neck Surg*. 2006;134:419–423.
13. Sanuki T, Yumoto E, Minoda R, Kodama N. Effects of type-II thyroplasty on adductor spasmodic dysphonia. *Otolaryngol Head Neck Surg*. 2010;142:540–546.
14. Braden M, Johns MM III, Klein AM, Delgaudio JM, Gliman M, Hapner ER. Assessing the effectiveness of botulinum toxin injections for adductor spasmodic dysphonia: clinician and patient perception. *J Voice*. 2010;24:242–249.
15. Cannito MP, Woodson GE, Murry T, Bender B. Perceptual analysis of spasmodic dysphonia before and after treatment. *Arch Otolaryngol Head Neck Surg*. 2004;130:1393–1399.
16. Stewart CF, Allen EL, Tureen P, Diamond BE, Blitzer A, Brin MF. Adductor spasmodic dysphonia: standard evaluation of symptoms and severity. *J Voice*. 1997;11:95–103.
17. Available at: http://www.fon.hum.uva.nl/praat/download_win.html. Accessed June 2014.
18. Gamboa J, Jimenez-Jimenez F, Nieto A, et al. Acoustic voice analysis in patients with essential tremor. *J Voice*. 1998;12:444–452.
19. Sugito M, Oyama G. A study of pauses and breathes in reading speech. *Stud Phon Speech Commun*. 1990;199–211.
20. Erickson ML. Effects of voicing and syntactic complexity on sign expression in adductor spasmodic dysphonia. *Am J Speech Lang Pathol*. 2003;12:416–424.
21. Cannito MP, Chorna LB, Kahane JC, Dworkin JM. Influence of consonant voicing characteristics on sentence production in abductor versus adductor spasmodic dysphonia. *J Voice*. 2014;28:394.e13–394.e22.
22. Hirose H, Ushijima T. Laryngeal control for voicing distinction in Japanese consonant production. *Phonetica*. 1978;35:1–10.
23. Hamano T, Kaji R, Katayama M, Kubori T, Ikeda A, Shibasaki H, Kimura J. Abnormal contingent negative variation in writer's cramp. *Clin Neurophysiol*. 1999;110:508–515.
24. Katada A, Nonaka S, Adachi M, et al. Functional electrical stimulation of laryngeal adductor muscle restores mobility of vocal fold and improves voice sounds in cats with unilateral laryngeal paralysis. *Neurosci Res*. 2004;50:153–159.
25. Klatt DH. Voice onset time, frication, and aspiration in word-initial consonant clusters. *J Speech Hear Res*. 1975;18:686–706.
26. Allen JS, Miller JL, DeSteno D. Individual talker difference in voice-onset-time. *J Acoust Soc Am*. 2003;113:544–552.