Title	Reliability and Validity of Speech Evaluation in Adductor Spasmodic Dysphonia
Author(s)	Yanagida, Saori; Nishizawa, Noriko; Hashimoto, Ryusaku; Mizoguchi, Kenji; Hatakeyama, Hiromitsu; Homma, Akihiro; Fukuda, Satoshi
Citation	Journal of Voice, 32(5), 585-591 https://doi.org/10.1016/j.jvoice.2017.06.022
Issue Date	2018-09
Doc URL	http://hdl.handle.net/2115/75316
Rights	© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
Rights(URL)	http://creativecommons.org/licenses/by-nc-nd/4.0/
Type	article (author version)
File Information	JoV32_585.pdf

Reliability and Validity of Speech Evaluation in Adductor Spasmodic Dysphonia

Saori YANAGIDA, Noriko NISHIZAWA, Ryusaku HASHIMOTO, Kenji MIZOGUCHI, Hiromitsu HATAKEYAMA, Akihiro HOMMA, Satoshi FUKUDA

Abstract

Summary: Objectives. In order to establish a reliable diagnostic tool for adductor spasmodic dysphonia (ADSD), it is necessary to determine the proper terms representing its characteristic voice symptoms and to relate them to objective measures such as acoustic parameters or speech perturbation. The aim of this study was to evaluate speech in patients with ADSD by perceptual evaluations and acoustic measures, and to examine the reliability and validity of the measures by comparison with normal controls.

Methods. Twenty-four patients with ADSD and 24 healthy volunteers matched to the ADSD patients with regard to age and sex participated in the study. Speech materials, consisting of three short sentences, were constructed from serial voiced syllables to elicit abductor voice breaks. Three otolaryngologists specializing in phoniatrics rated the degree of voice symptoms using a visual analog scale (VAS). VAS sheets with five 100-mm horizontal lines were given to each rater. The ends of the lines were labeled normal versus severe, and the five lines were labeled as overall severity and each of the four voice symptoms; strangulation, interruption, tremor and strained speech. Nine words were selected from the speech materials for acoustic analysis, and abnormal acoustic events were classified into one of the three categories; percentage of frequency shifts, percentage of aperiodic segments, or percentage of phonation breaks. Acoustic measures were performed by a speech-language-hearing therapist specializing in voice disorders and five healthy university students. To evaluate the intra- and inter-rater/measurer reliability of the VAS scores or acoustic measures, Pearson r correlations were calculated. To examine the validity of perceptual evaluations and acoustic measures, the sensitivity, specificity, positive predictive value and negative predictive value were calculated.

Results. Pearson r correlation coefficients for overall severity showed the highest intra- and inter-rater reliability, and reliability coefficients for one of the four voice symptoms, strangulation (r = 0.816–0.937), were slightly higher than those for the other symptoms. For acoustic events, intra-measurer reliabilities were $r = 0.645$ (frequency shifts), $r = 0.969$ (aperiodic segments), and $r = 1.0$ (phonation breaks), and inter-measurer reliability ranged from $r = 0.102$ to $r = 1.0$ (average $r = 0.861$). The Pearson r correlation coefficient for phonation breaks was higher than those for the other acoustic events. Perceptual evaluation using VAS showed high sensitivity (91.7 %) and specificity (100 %),

and acoustic analysis showed low sensitivity (70.8 %) and high specificity (100 %). Eight of the 24 patients were judged to be within normal limits by one or both evaluation methods.

Conclusions. Both perceptual evaluation and acoustic measures alone were found to be likely to miss true ADSD patients. It is important to obtain a range of speech materials and to use a combination of perceptual evaluation and acoustic measures based on our understanding of the advantages and disadvantages of both methods.

Key Words: Adductor spasmodic dysphonia, speech evaluation, reliability, validity

INTRODUCION

Spasmodic Dysphonia (SD) is a disorder of speech and voice resulting from irregular and uncontrolled spastic movements of the laryngeal muscles. (1) It is characterized as a chronic neurological disorder that causes task-specific contractions of the laryngeal muscles during speech. (2) There are three types of SD: 1) Adductor SD (ADSD), which is the most common type and affects the adductor laryngeal muscles; 2) Abductor SD (ABSD), which accounts for about 30 percent of all diagnosed patients and affects the abductor laryngeal muscles; and 3) Mixed SD, which is quite rare and presents with signs observed in both ADSD and ABSD.

The voice symptoms in ADSD are characterized by an irregular and uncontrollable tight or strained voice, sudden changes in pitch or loudness and involuntary voice interruption. For experienced phoniatricians, a diagnosis of SD is not difficult because of its characteristic symptoms. However, the exact determination of this disease requires a combination of a number of special evaluation methods such as a detailed interview and speech-pathological evaluation along with neuro-phonological examination and laryngeal endoscopy. Therefore, general otolaryngologists who are not specialized in speech pathology may miss signs of SD resulting in misdiagnosis and delayed or inadequate therapeutic intervention.

In order to establish a reliable diagnostic tool for ADSD, it is necessary to determine the proper terms representing its characteristic voice symptoms and to relate them to objective measures such as acoustic parameters or speech perturbation.

Evaluation methods are classified as perceptual, acoustic or aerodynamic measurements. (3-11) Various words used to express voice symptoms in ADSD, such as staccato, feeble, intermittent, jerky, grunting, effortful, overpressured, aphonic, laryngealized, aspirated, aperiodic and dysprosody, have been reported. (12,13) However, there are few reports on correlations between perceptual evaluations using these terms and acoustic measures in ADSD. (14,15)

Erickson (16) and Cannito et al. (17) reported that voice symptoms in ADSD are more remarkable in voiced consonant sentences than in voiceless consonant sentences. It is speculated that abnormal laryngeal adductor/abductor motor coordination in SD can cause not only "voice" problems but also

"speech" problems in SD patients. We may, therefore, gain a better understanding of the symptoms in SD by considering syntactic complexity, articulation place or manner and voiced/voiceless distinction.

The aim of this study was to evaluate speech in patients with ADSD by perceptual evaluations and acoustic measures, and to examine the reliability and validity of the measures by comparison with normal controls.

METHODS

Participants

Twenty-four patients with ADSD (20 women and 4 men, mean age 39.7±14.5 years) participated in this study. They were all native Japanese speakers. All patients were evaluated at the Health Sciences University of Hokkaido Hospital between December 2013 and July 2016. 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 interview, laryngeal endoscopy, and speech-pathological evaluation. Twenty-four healthy volunteers matched to the ADSD patients with regard to age and sex also participated in the study as normal controls. Control subjects had no history of neurological disorders or vocal pathology. Their voices were perceptually screened by a ST and judged to fall within normal limits.

Speech materials

Speech materials were three short sentences: (1) Ame ga yandara umi ni mogurou. (Let's go diving in the ocean when the rain stops.); (2) Ano yama no ue niwa aoi yane no ie ga aru. (There is a blue-roofed house at the top of the mountain.); and (3) Daijobu no imi wa iroiro aru you desu. ("Okay" has various meanings.) The speech materials were constructed from serial voiced syllables to elicit abductor voice breaks (Table 1).

We performed acoustic recordings of the speech materials at the first examination for each patient. Recording was made in a quiet room using a digital audio recorder (Frontier TASCAM HD-P2 portable stereo audio recorder, TEAC Corp, Tokyo, Japan) at a sampling rate of 44.1 kHz, and a condenser microphone (AKG C1000 S CE Harman International, Stamford, CT) under similar conditions. The microphone was kept at a distance of 20 cm from the lips. Recordings of the third sentence were missing for three of the patients as they were evaluated using a different protocol. Another patient also misread part of the second sentence resulting in the deletion of the sentence from the voice record.

Assessment

1. Perceptual evaluation

Three otolaryngologists specializing in phoniatrics rated the degree of voice symptoms using a visual analog scale (VAS). VAS sheets with five 100-mm horizontal lines were given to each rater. The ends of the lines were labeled normal versus severe, the left end always corresponded to a score of 0 (i.e., normal) and the right end to 100 (i.e., severe). The five lines were labeled as overall severity and each of the four voice symptoms; strangulation, interruption, tremor and strained speech (Figure 1). Audio files for rating were created for each participant, totaling 48 files. Each file was made of a 1000 ms lead of sine wave (440Hz, amplitude 0.8) followed by 1000 ms of silence, and the speech materials (Figure 2). The raters were asked to listen to each file and place a single vertical line on the VAS to rate the overall severity and each of the four voice symptoms. In accordance with the evaluation method developed by Sanuki et al. (18), the following definitions were given to the raters.

1. Overall severity.

   A comprehensive evaluation through the entire speech sample.

2. Strangulation.

   Tightness of voice production.

3. Interruption.

   Abrupt voice initiation and arrest.

4. Tremor.

   Rapid fluctuations in pitch or loudness.

5. Strained Speech.

   An effortful, sustainable speech.

All files were played back with a digital audio recorder (LINEAR PCM recorder LS-7, OLYMPUS, Tokyo, Japan) and headphones (HD598 headphones, SENNHEISER ELECTRONIC GMBH & CO.KG, Germany).

All raters conducted these procedures twice at intervals one or more week(s).

2. Acoustic measures

Nine words constructed from serial voiced syllables were selected from the speech materials for acoustic analysis (Table 1). The acoustic waveforms of the target words were then analyzed using the *Praat,* free software program. (19)

Due to the absence of some speech materials for four patients, the total number of samples was 425. When the voice source was so aperiodic that automatic extraction of the fundamental frequency was not possible, the speech sample was excluded. At last, 405 of the original 425 samples were accepted as having adequate sound quality.

Following Cannito et al. (15), each abnormal acoustic event was classified into one of the three categories described below. When the result of the automatic analysis by *Praat* was irrelevant, the measurer verified the waveform visually, and classified the acoustic event.

Categories of abnormal acoustic events

The criteria of the abnormal acoustic events conformed to those of Cannito et al. (15) (Figure 3).

1. Percentage of Frequency Shifts.

A short-term (<50 ms) change in fundamental frequency greater than 50 Hz was defined as a frequency shift. For each observed frequency shift, the initial and final cycles were identified on the waveform and the duration was measured. The percentage of frequency shifts was then calculated by dividing the sum of all frequency shifts in milliseconds by the total duration of the 9 words.

2. Percentage of Aperiodic Segments.

Aperiodicity was defined as non-repetitive cycles on the waveform. The acoustic waveform of a voiced segment of each of the 9 words was inspected for deviation from periodicity. The percentage of aperiodic segments was calculated by adding the durations of such segments and then dividing it by the sum of durations of voicing in all 9 words.

3. Percentage of Phonation Breaks.

Any interruption of voicing lasting longer than 50 ms was defined as a phonation break. To obtain the percentage of phonation breaks, the duration of phonation breaks was summed and divided by the sum of the duration from voicing onset to offset in all 9 words.

Statistical design

1. Reliability

1.1 Perceptual evaluation

To evaluate the intra-rater reliability of the VAS scores, Pearson r correlations were calculated between the first and second listening sessions for each rater. Similarly, to evaluate the inter-rater reliability of the VAS scores, reliability coefficients of overall severity and each of the four voice symptoms were calculated for the first and second listening sessions.

1.2 Acoustic measures

One of the authors, a speech-language-hearing therapist specializing in voice disorders, performed the acoustic measures. To evaluate the intra-measurer reliability, approximately 20% of the total samples (n = 87) were randomly extracted. Additionally, to determine whether non-experts could make decisions automatically or not, the inter-measurer reliability was calculated. Approximately 30 % of the total samples (n = 130) were measured by five healthy university students who received lectures and practical skill training for four hours before the analysis.

2. Comparison between ADSD and normal controls

2.1 Perceptual evaluation

We first compared the overall severity of the speech samples using a t-test. Second, to determine which perceptual evaluations in ADSD differed from those in normal controls, we performed a two-way repeated measures analysis of variance (ANOVA) with the four voice symptoms (strangulation, interruption, tremor and strained speech) and subject types (ADSD and normal controls) as factors. A Greenhouse-Geisser correction for sphericity was used where necessary. When an interaction existed, one-way ANOVA for each level of the four voice symptoms in each group was applied. No adjustment for multiplicity was considered.

2.2 Acoustic measures

To determine which acoustic events differed from those in normal controls, we performed a two-way repeated measures ANOVA with three acoustic events (percentage of frequency shifts, percentage of aperiodic segments and percentage of phonation breaks) and subject type (ADSD and normal controls) as factors. A Greenhouse-Geisser correction for sphericity was used where necessary. When an interaction existed, one-way ANOVA for each level of the three acoustic events was applied to follow up this interaction. Fisher's exact test was performed when no acoustic events at all were measured in the normal controls. A t-test was conducted separately for each level of the other acoustic events.

Statistical software R was only used for the Fisher's exact test, with IBM SPSS Statistics ver.22 used for all other analyzes.

Validity

To examine the validity of perceptual evaluations and acoustic measures, the sensitivity, specificity, positive predictive value and negative predictive value were calculated. To calculate these items, overall severity on the VAS and the total percentage of the three acoustic events were used. Cut-off values were assumed as the average + 3 standard deviations of normal controls for each method. We examined whether the ADSD patients could be judged within normal limits or not, and what kind of perceptual and acoustic features they had.

RESULTS

1. Reliability

1.1 Perceptual evaluation

Pearson r correlation coefficients for perceptual evaluation ranged from $r = 0.802$ to $r = 0.978$, and showed high intra-rater reliability (Table 2). On the other hand, inter-rater reliability ranged from r

= 0.747 to r = 0.947 (Table 3). Pearson r correlation coefficients for overall severity showed the highest intra- and inter-rater reliability, and the reliability coefficients for one of the four voice symptoms, strangulation, was slightly higher than those for the other symptoms.

1.2 Acoustic measures

Intra-measurer reliability for each acoustic event were r = 0.645 (frequency shifts), r = 0.969 (aperiodic segments), and r = 1.000 (phonation breaks). Although some of the correlation coefficients were low, the average of the five measurers' inter-measurer reliability was r = 0.861 (Table 4). The Pearson r correlation coefficient for phonation breaks was higher than those for the other acoustic events.

2. Perceptual evaluation

To compare the overall severity VAS scores between the ADSD and normal controls, a t-test was conducted. As equal variance of the SD group and the control group could not be assumed by Levene's test, we have adjusted the degree of freedom. Overall severity in the ADSD patients was significantly higher than that of the normal controls (t (23.086) = 7.849, P < 0.001, Cohen's d = 2.27). (Table 5) Since the floor effect was observed in the control group, the overall severity was classified into 4 levels based on VAS scores and a nonparametric test was performed. For severity classification, 0: normal (0-5 mm) was assumed as the average + 3 standard deviations of normal controls, the remaining 95 mm is divided into the following 3 levels: 1: mild (6-37 mm), 2: moderate (38-69 mm), 3: (70-100 mm). The number of participants classified into each level is shown in Table 6. Fisher's exact test revealed that there were significant differences in the occurrence frequencies of each level in both groups (P < 0.001). Two-way ANOVA showed significant main effects for subject type (F (1, 46) = 53.183, P < 0.001) and each VAS element (F (1.947, 89.566) = 3.212, P = 0.046). The interaction between the two factors was shown to be marginally significant (F (1.947, 89.566) = 2.903, P = 0.061). To follow-up this interaction, one-way ANOVA for each subject type was applied. The one-way ANOVA showed no significant main effects for both ADSD patients (F (1.946, 44.747) = 3.064, P = 0.058) and normal controls (F (3, 69) = 1.040, P = 0.381) (Table 5).

3. Acoustic measures

None of the normal controls showed any phonation breaks in either of the two measures, whereas phonation breaks were observed in seven of the 24 patients. Fisher's exact test showed a significant difference between the ADSD and normal controls (P = 0.0094). Two-way ANOVA for the other two acoustic events (frequency shifts and aperiodic segments) showed significant main effects for subject type (F (1, 46) = 20.775, P < 0.001) and each acoustic event (F (1, 46) = 6.505, P < 0.05). Also the interaction between the two factors was shown to be significant (F (1, 46) = 6.726, P < 0.05). To follow-up this interaction, we applied one-way ANOVA for the ADSD patients. The one-way ANOVA showed significant main effects for the ADSD patients (F (1.461, 33.592) = 5.507, P = 0.015). We then

compared each pair, the percentage of aperiodic segments was found to be significantly higher than those of aperiodic segments ($P = 0.050$). On the other hand, other pairs showed no significant effects (aperiodic segments vs. phonation breaks; $P = 0.062$, frequency shifts vs phonation breaks; $P = 1.000$). As we did not observe any phonation breaks in the normal controls, we conducted a t-test for frequency shifts vs. aperiodic segments, the results of which were not significant (t (43.405) = 0.237, $P = 0.813$) (Table 7).

4. Sensitivity and specificity

Cut-off values for the VAS scores and acoustic events were 4.6 mm and 3.2 %, respectively. The sensitivity, specificity, positive predictive value and negative predictive value for perceptual evaluations were 91.7 %, 100 %, 100 % and 92.3 %, respectively. Similarly, those for acoustic measures were 70.8 %, 100 %, 100 % and 77.4 %, respectively.

Sixteen of the 24 patients were judged to be abnormal on the basis of both perceptual evaluation and acoustic measures. However, eight of the 24 patients were judged to be within normal limits by one or both evaluation methods in the present study.

The features of patients judged to be within normal limits were classified into the following three types. Type 1: by perceptual evaluations alone (n = 1). The VAS score for tremor in this patient was higher than that for overall severity. Type 2: by acoustic measures alone (n = 6). Although voice symptoms in these patients appeared to fluctuate markedly by task, this fact was not remarkable in terms of the speech materials used in this study. Perceptual evaluations assess voice at a sentence level and focus on not only the fundamental frequency but also suprasegmental features such as intonation. On the other hand, acoustic measures assess voice at a word level and focus on the fundamental frequency. These differences may have led to these misjudgments. Type 3: by both perceptual evaluation and acoustic measures (n = 1). Because of the absence of voice symptoms at the first clinical examination, we requested that this patient bring a recording of voice symptoms taken while she waited on customers at work. This patient had high task specificity, and no voice symptoms were observed during the medical interview, which perhaps explains why the patient was incorrectly assessed as a normal control by both methods.

In order to distinguish SD from motor dysphonia, which presents with similar symptoms, a comprehensive evaluation method, such as detailed interview, laryngeal endoscopy and speech-pathological evaluation, is required. It is important to obtain a range of speech materials and to use a combination of perceptual evaluation and acoustic measures based on our understanding of the advantages and disadvantages of both methods.

DISCUSSION

Reliability

High reliability was observed for both intra-rater and inter-rater perceptual evaluation. In the present study, three otolaryngologists with a good deal of clinical experience in SD rated the sound files. The high repeatability of perceptual evaluation might not, however, be assured when a clinician who is not an expert in voice disorders is required to assess voice symptoms. To establish a practical evaluation tool for general otolaryngologists and speech language pathologists, a large number of speech samples from SD patients need to be integrated into a database in association with rational severity scores to allow the development of an education program for perceptual evaluation.

Although the total percentage of acoustic events showed high intra-measurer reliability, frequency shifts demonstrated an only moderate correlation. This resulted from the low degree of accuracy of the visual verification required when automatic analysis by software was possible. Automatic acoustic measures showed high inter-measurer reliability for the five healthy non-expert university students. Acoustic analysis has advantages in that it is quantitative and not dependent on clinical experience. In the future, further development of acoustic analysis software can assist in limiting analytic error and increasing clinical convenience.

Perceptual evaluation

perceptual evaluation using VAS showed high sensitivity and specificity, and these results suggest that VAS is a valid tool for the evaluation of voice symptoms in SD. Perceptual evaluation is a useful method of real time evaluation without the need for any special equipment.

ADSD patients showed marked individual differences in the four voice symptoms, and no specific patterns in SD were observed. In the present study, only overall severity could be used to distinguish between patients and normal controls, and classification using the four voice symptoms was not significant. However, the four voice symptoms selected in this study might be useful as a subtest when comparing the severity of SD pre- and post-treatment. In fact, Nomoto et al. (20) reported that postoperative strangulation was significantly improved by thyroarytenoid muscle myectomy (TAM) in comparison to those treated by type II thyroplasty (TPII) regardless of a lack of any significant differences between in severity of the two treatment groups before surgery. Further, Nomoto et al. (21) reported that all patients were significantly improved after surgery with TAM, but the improvements in interruption and tremor after TPII were not significant.

Acoustic measures

Acoustic analysis showed low sensitivity and high specificity, in other words, acoustic measures were more likely to miss true ADSD patients than perceptual evaluation. Further, it is difficult to use acoustic methods when the voice source is too highly aperiodic to allow extraction of the fundamental frequency. On the other hand, the appearance of phonation breaks might allow an accurate diagnosis of ADSD as phonation breaks were only observed in ADSD patients in this study (7 of 24 patients). Acoustic analysis has advantages in terms of quantitation and objectivity, and it is suggested that perceptual and acoustic evaluation methods be combined for practical clinical diagnosis.

The points to consider in speech evaluation

We should keep in mind that both perceptual and acoustic measures are only useful for diagnostic purposes when combined appropriately. In fact, eight of the 24 patients fell within normal limits on the basis of one or both evaluation methods in the present study.
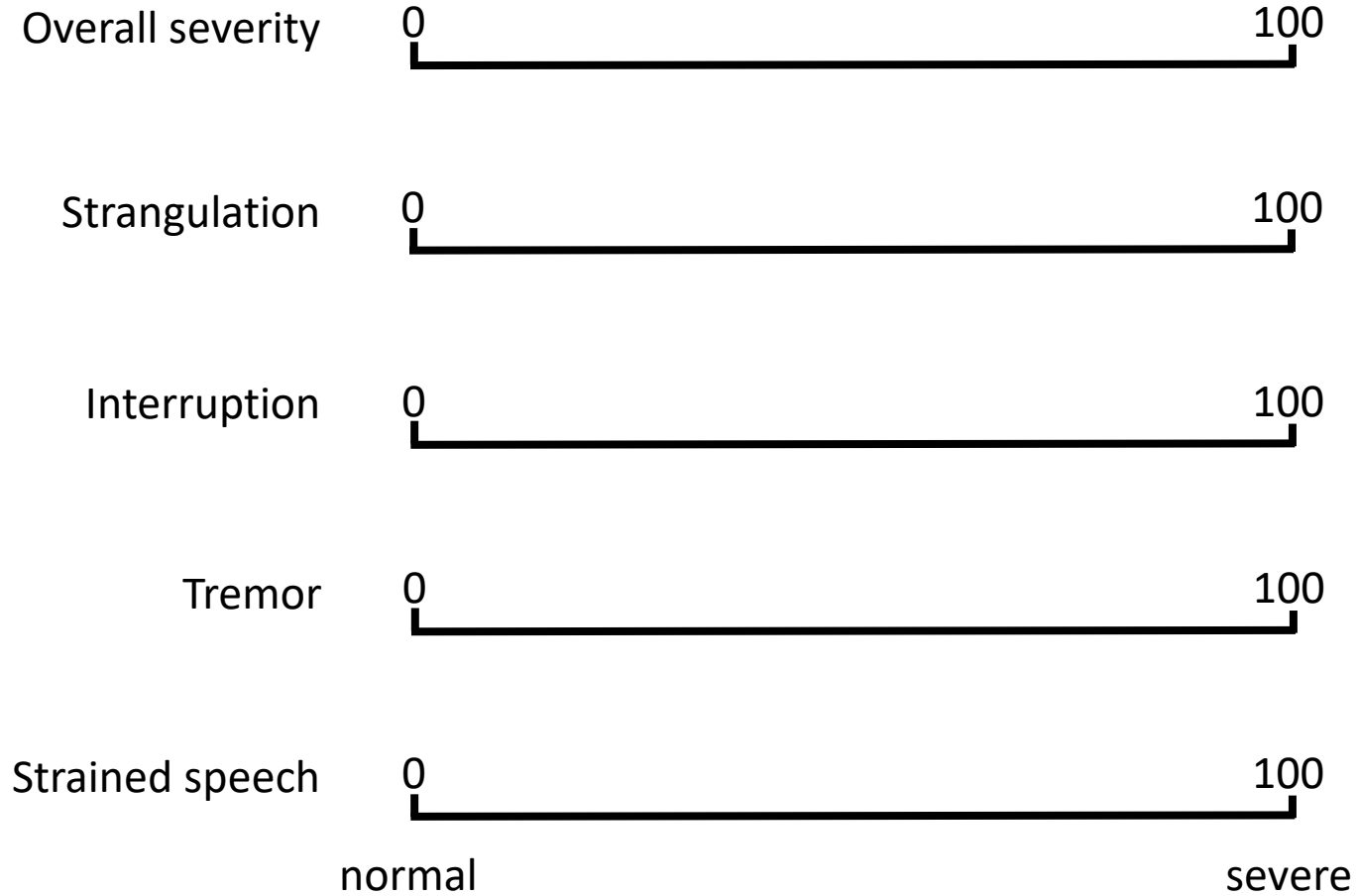
Acknowledgments

REFERENCES

1. Murry T. Spasmodic Dysphonia: let's look at that again. *J Voice*. 2014;28:694-699.
2. Hu A, Hillel A and Meyer T. Factors associated with patient-perceived hoarseness in spasmodic dysphonia patients. *J Voice*. 2016;30:769.e23-769.e26.
3. 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.
4. Cannito MP, Woodson GE, Murry T, Bender B. Perceptual analyses of spasmodic dysphonia before and after treatment. *Arch Otolaryngol Head Neck Surg*. 2004:130:1393-1399.
5. Edgar JD, Sapienza CM, Bidus K, Ludlow CL. Acoustic measures of symptoms in abductor spasmodic dysphonia. *J Voice*. 2001;15:362-372.
6. Sapienza CM, Murry T, Brown WS. Variations in adductor spasmodic dysphonia: acoustic evidence. *J Voice*. 1998;12:214-222.
7. Sapienza CM, Walton S, Murry T. Adductor spasmodic dysphonia and muscular tension dysphonia; acoustic analysis of sustained phonation and reading. *J Voice*. 2000;14:502-520.
8. Zwirner P, Murry T, Swenson M, Woodson GE. Acoustic changes in spasmodic dysphonia after botulinum toxin injection. *J Voice*. 1991;5:78-84.
9. Yanagida S, Nishizawa N, Mizoguchi K, Hatakeyama H, Fukuda S. Voice onset time for the word-initial voiceless consonant /t/ in Japanese spasmodic dysphonia- a comparison with normal controls. *J Voice*. 2015;29:450-454.
10. Woo P, Colton R, Casper J, Brewer D. Analysis of spasmodic dysphonia by aerodynamic and laryngostroboscopic measurements. *J Voice*. 1992;6:344-351.
11. 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.
12. Aronson AE, Brown JR, Litin EM, Pearson JS. Spastic dysphonia. I. Voice, neurologic, and psychiatric aspects. *J Speech Hear Disord*. 1968;33:203-218.
13. Izdebski K. Symptomatology of adductor spasmodic dysphonia: a physiologic model. *J Voice*. 1992;6:306-319.

14. Zwirner P, Murry T, Woodson GE. Perceptual-acoustic relationships in spasmodic dysphonia. *J Voice*. 1993;7:165-171.

15. Cannito MP, Doiuchi M, Murry T, Woodson GE. Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates. *J Voice*. 2012;26:818.e5-818.e13.

16. Erickson ML. Effects of voicing and syntactic complexity on sign expression in adductor spasmodic dysphonia. *Am J Speech Lang Pathol*. 2003;12:416-424.

17. Cannito MP, Chorna LB, Kahane JC, Dworkin JP. Influence of consonant voicing characteristics on sentence production in abductor versus adductor spasmodic dysphonia. *J Voice*. 2014;28:394.e13-394.e22.

18. Sanuki T, Isshiki N. Overall evaluation of effectiveness of type II thyroplasty for adductor spasmodic dysphonia. *Laryngoscope*. 2007;117:2255-2259.

19. Available at: http://www.fon.hum.uva.nl/praat/　Accessed February 2017.

20. Nomoto M, Tokashiki R, Hiramatsu H, et al. The comparison of thyroarytenoid muscle myectomy and type II thyroplasty for spasmodic dysphonia. *J Voice*. 2015;29:501-506.

21. Nomoto M, Tokashiki R, Hiramatsu H, Motohashi R, Sakurai E, Toyomura F, Suzuki M. Surgical treatments for spasmodic dysphonia. *J Jpn Bronchoesophagol Soc*. 2013;64:8-13.

［Procedures］

1. Listen to the speech sample.

2. Scoring the degree of overall severity and each voice symptoms (strangulation, interruption, tremor, strained speech).

3. Draw a vertical line where you think the most suitable section.

Overall severity   0 _____ 100

Strangulation   0 _____ 100

Interruption   0 _____ 100

Tremor   0 _____ 100

Strained speech   0 _____ 100

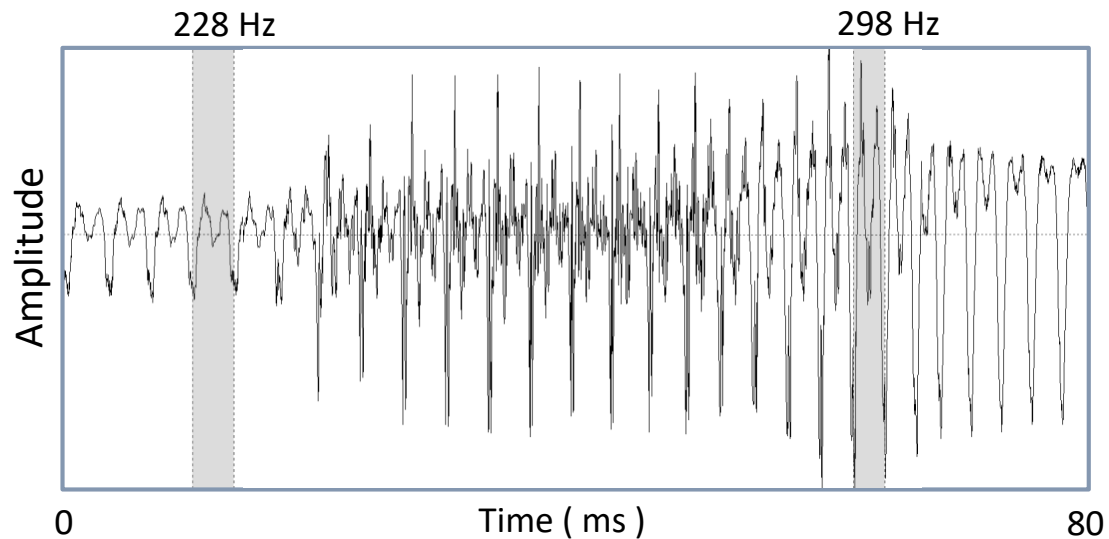   normal   severe

FIGURE 1. Perceptual evaluation form using visual analog scale (VAS). Each rater listened to the speech sample and asked to place a single vertical line on the 100 mm horizon VAS sheet.
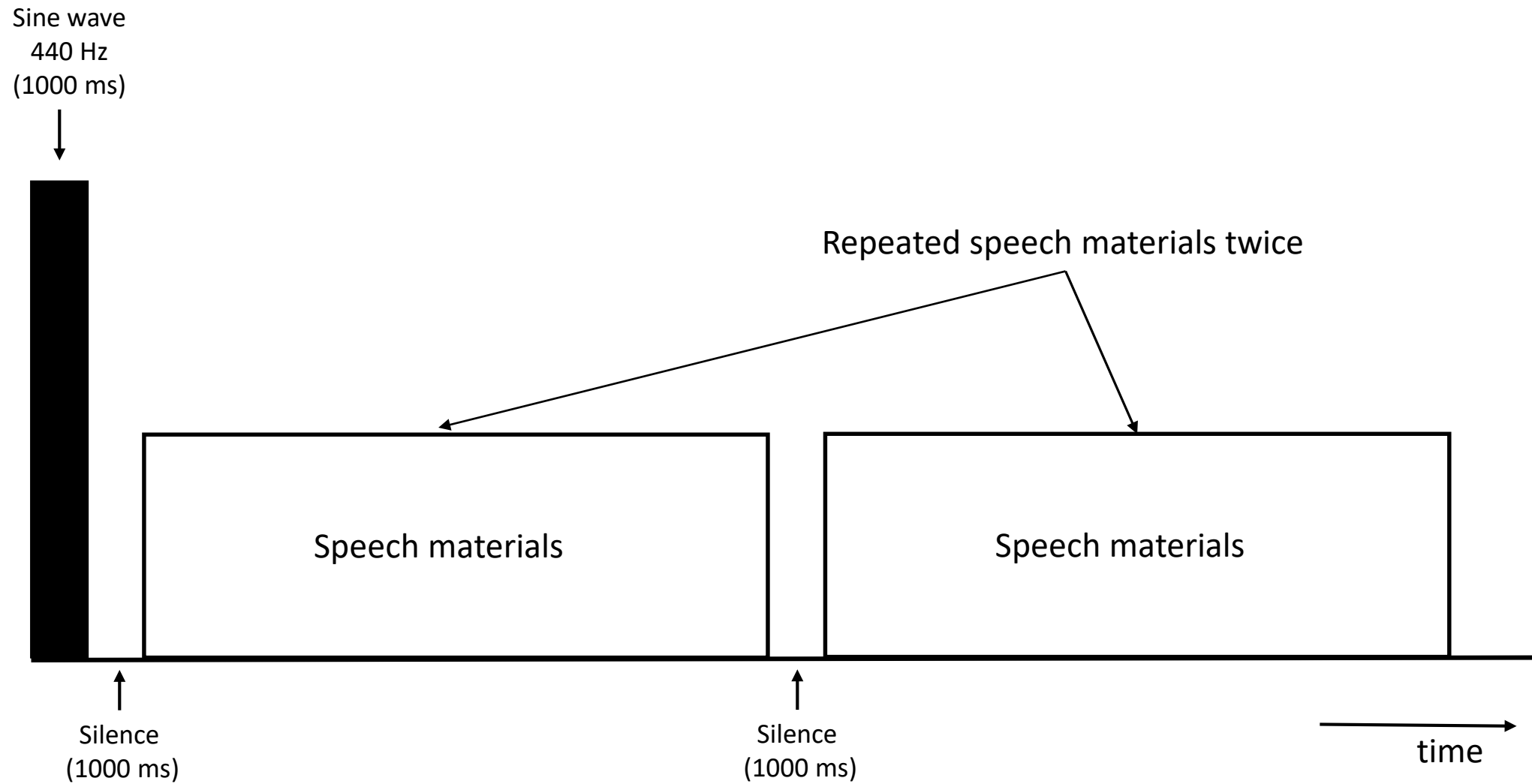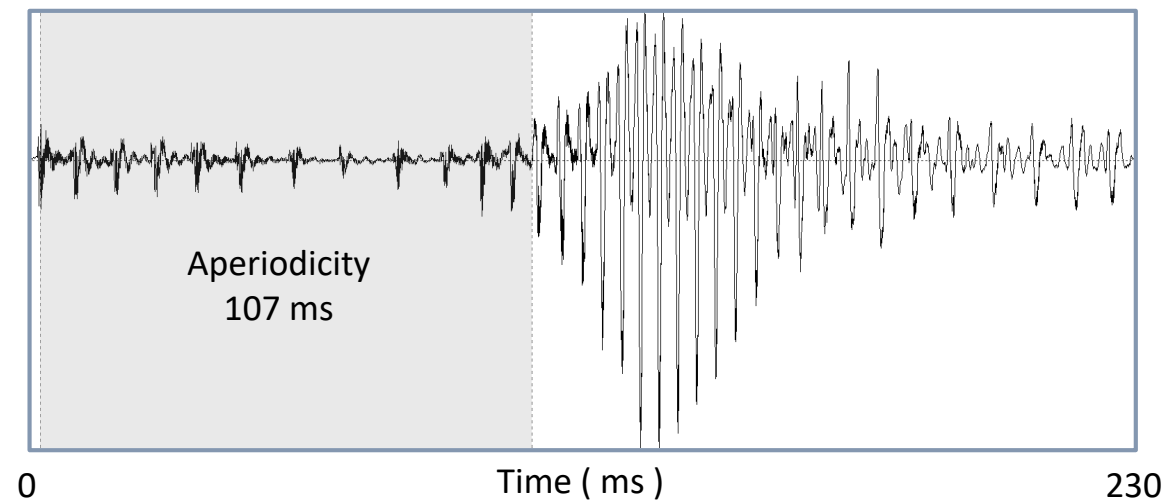
FIGURE 2. A speech sample for perceptual evaluation constructed with sine wave, some silences and repeated speech materials.
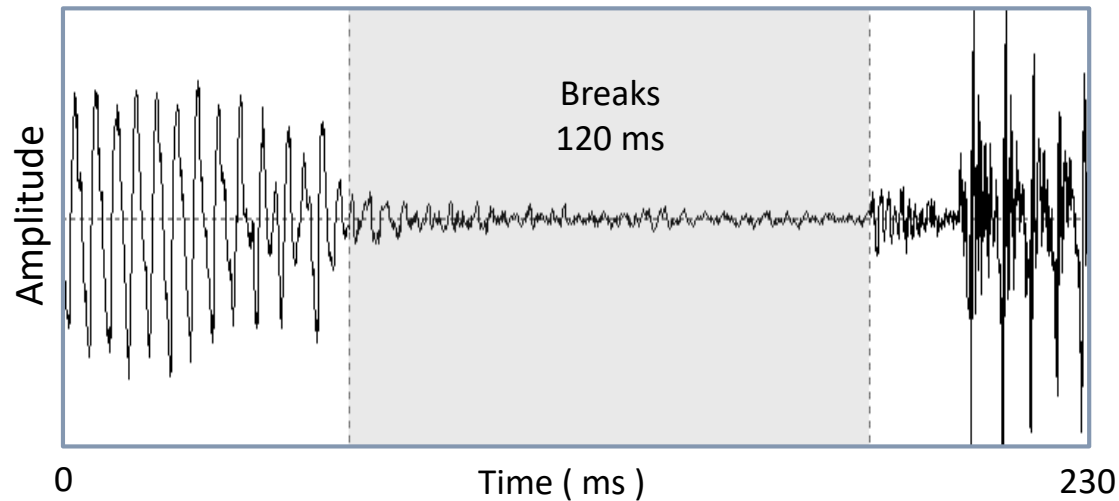
A : Frequency shifts

228 Hz

298 Hz

Amplitude

0   Time (ms)   80

B : Aperiodic segments

Aperiodicity
107 ms

0   Time (ms)   230

C : Phonation breaks

Breaks
120 ms

Amplitude

0   Time (ms)   230

FIGURE 3. These figures showed three acoustic events (A: Frequency shifts, B: Aperiodic segments, C: Phonation breaks) . A: A short term (<50 ms) change in fundamental frequency greater than 50 Hz was defined as a frequency shifts. B: Aperiodic segments was defined as non-repetitive cycles on the waveform. C: Any disruption of voicing lasting longer than 50 ms was defined as a phonation break.

TABLE 1. Speech Materials

No.	Sentences
1	**Ame** ga yandara **umi** ni mogurou. (Let's go diving in the ocean when the rain stops.)
2	Ano **yama** no **ue** niwa **aoi yane** no **ie** ga aru. (There is a blue-roofed house at the top of the mountain.)
3	Daijobu no **imi** wa **iroiro** aru you desu. ("Okay" has various meanings.)

Notes: The bold nine words are targets to measure three acoustic events.
The bold faced font was not present during the reading aloud of each sentence by participants.

TABLE 2. Intra-rater Reliability for Perceptual Evaluation

	Overall severity	Strangulation	interruption	tremor	Strained speech
Rater 1	.943	.802	.918	.888	.874
Rater 2	.978	.934	.959	.921	.909
Rater 3	.957	.973	.906	.941	.919

TABLE 3. Inter-rater Reliability for Perceptual Evaluation

First session	Overall severity	Strangulation	interruption	tremor	Strained speech
Rater 1 vs 2	.938	.816	.840	.856	.798
Rater 2 vs 3	.933	.937	.851	.775	.879
Rater 1 vs 3	.898	.889	.907	.747	.868
Second session					
Rater 1 vs 2	.947	.891	.882	.817	.889
Rater 2 vs 3	.923	.904	.799	.902	.862
Rater 1 vs 3	.923	.859	.931	.804	.824

TABLE 4. Intra- and Inter-measurer Reliability for Acoustic Events

	Percentage (%)		
	Frequency shifts	Aperiodic segments	Phonation breaks
Intra-measurer	.645	.969	1.0
Inter-measurer			
Author SY vs St 1	.995	.934	1.0
vs St 2	.482	.102	.992
vs St 3	.699	.833	1.0
vs St 4	.938	.999	.968
vs St 5	.995	.977	1.0

TABLE 5. Mean VAS Scores by Perceptual Evaluation

Variable	Mean ± SD	
VAS scores (mm)	ADSD (n = 24)	NC (n = 24)
Overall Severity	48.0 ± 29.5	0.7 ± 1.3
Strangulation	41.4 ± 4.4	0.7 ± 4.4
Interruption	38.7 ± 3.9	0.5 ± 3.9
Tremor	29.7 ± 3.4	0.3 ± 3.4
Strained Speech	38.0 ± 4.1	0.4 ± 4.1

TABLE 6. A Cross-tabulation by Overall Severity using VAS

	VAS scores for overall severity			
	0 : Normal (0-5 mm)	1 : Mild (6-37 mm)	2 : Moderate (38-69 mm)	3 : Severe (70-100mm)
ADSD (n = 24)	3	5	10	6
NC (n = 24)	24	0	0	0

TABLE 7. Mean Percentages for Abnormal Acoustic Events

Variable	Mean ± SD	
Acoustic events (%)	ADSD (n = 24)	NC (n = 24)
Total of acoustic events	13.7 ± 13.2	0.5 ± 0.9
Frequency shifts	3.0 ± 4.1	0.3 ± 0.7
Aperiodic segments	8.3 ± 10.1	0.2 ± 0.6
Phonation breaks	2.4 ± 5.7	0