Adductor Spasmodic Dysphonia: Standard Evaluation of Symptoms and Severity

*†Celia F. Stewart, *‡Elizabeth L. Allen, *Phyllis Tureen, §Beverly E. Diamond, †Andrew Blitzer, and †Mitchell F. Brin

*Department of Speech-Language Pathology and Audiology, New York University, †Department of Neurology, Mount Sinai Medical Center, ‡Zilkha Center, International Center for the Disabled, §Irving Center for Clinical Research, Columbia-Presbyterian Medical Center, and †Department of Otolaryngology, Roosevelt Hospital, New York, New York, U.S.A.

Summary: Description and quantification of the symptoms of adductor spasmodic dysphonia often reflect the clinician's knowledge of the disorder, ideas about the cause of the disorder, and personal experience. No reliable instrument that identifies and quantifies the spectrum of perceptual symptoms has been available. Therefore, we developed a standardized measure called the Unified Spasmodic Dysphonia Rating Scale (USDRS) in cooperation with a team of 118 experienced voice judges. Consensual validations of content validity guided the incremental development of the scale. Using the USDRS allows more consistent and complete data collection, both clinically and in research clinical trials. Key Words: Spasmodic dysphonia—Evaluation—Rating scale—Voice—Interjudge agreement—Focal laryngeal dystonia.

ADDUCTOR SPASMODIC DYSPHONIA: A STANDARD EVALUATION

To adequately evaluate patients with spasmodic dysphonia, it is necessary to have a team of professionals consisting of a speech-language pathologist, an otolaryngologist, and a neurologist (1-3). Team evaluation rests on the identification of characteristic clinical phenomenology. Symptoms are identified based on history, perceptual evaluation of speech and associated symptoms, visualization of the anatomy and physiology of the vocal folds, and assessment of neurological control of general and specific body movements. Unfortunately, with the measures currently in use, professionals do not always agree on the diagnosis, the severity, the symptoms, or their importance.

An impressive body of literature describes the perceptual symptoms of spasmodic dysphonia. However, no standardized clinical instrument exists for evaluating the severity of the perceptual symptoms of adductor spasmodic dysphonia. Researchers have investigated general characteristics ranging from one overall dimension that described voice quality (4-6) to four voice symptoms: overpressure, harshness or hoarseness, tremor, and breathiness (7). In addition, researchers have used different numerical scales ranging from 4 points (8) to 100 points (9-10). They have rarely mentioned or assessed the validity of these measures. Highly sophisticated, pretrained judges who had extensive experience rating the severity of spasmodic dysphonia performed most of these studies. Some re-
searchers used only one highly sophisticated pretrained judge to rate their nonstandard measures (8,11) and others used three to five pretrained judges (5-7,12). Izdebski, in 1984, said that his research supported “the notion that the perceptual evaluation of voice qualities is valid if judges are pretrained” (12; p. 377). When those pretrained judges rated patients with spasmodic dysphonia, their interrater reliability was good.

Efficacy in treating any clinical disorder depends on having an accurate and practical means of quantifying the symptoms that can be compared with standard values. If no external standard exists, we can develop one through consensual validation. Feinstein (13) suggests developing a standard of reference by assembling a group of experts to test conformity. He stated: “For this process, a group of appropriate authorities is assembled. They deliberate over the particular issue at hand and agree on what will be regarded as the correct or standard result” (13; p. 192). Scales, such as the Glasgow Coma Scale (14) and the Katz Index of Activities of Daily Living (15), were developed through consensual validation.

The purpose of the present research was to develop a standard measure to rate the severity of the perceptual symptoms of adductor spasmodic dysphonia. We intend this instrument to help experienced voice judges rate the severity of perceptual symptoms of adductor spasmodic dysphonia quickly and uniformly.

MATERIALS AND METHODS

Subjects

We designed the Unified Spasmodic Dysphonia Rating Scale (USDRS) for experienced voice clinicians who specialize in voice and voice rehabilitation. Experienced voice clinicians are thought to be more reliable than inexperienced clinicians in their voice ratings (16-18). To identify judges with expertise in voice disorders, we compiled a list of speech pathologists who had published articles about spasmodic dysphonia, referred patients with spasmodic dysphonia to Columbia-Presbyterian Medical Center, or were affiliates of the Voice and Voice Disorders Special Interest Division (VVDSID) of the American Speech-Language-Hearing Association. We sent letters to those voice specialists asking them to participate in the study and included all 169 who agreed to participate in the study.

Instrumentation

Stimulus tasks

Stimulus tasks were selected to elicit a variety of voice productions. The symptoms of adductor spasmodic dysphonia are manifested in the habitual pitch range of the speaking voice during speech and oral reading (18-20). Therefore, the first group of stimulus tasks elicited connected speech. The first paragraph of “The Rainbow Passage” (21) was selected because it is well known and contains a wide range of phonemes. In addition, we included “The Cookie Theft Picture” (22) to elicit spontaneous continuous speech from all speakers including those with limited literacy or proficiency in English.

The perceptual symptoms of adductor spasmodic dysphonia often improve in some conditions. For example, the voice is usually better in the morning and when speakers laugh, whisper, sing, clear their throats, yawn, sigh, cough, or sustain vowels in conversational pitch or falsetto (23-28). They can often produce short segments of normal voice during brief greetings or one-word responses (18,29,30) or when speaking quietly (5,18,20,29,30). Alcoholic intoxication, anger, and joy can also trigger temporary improvement, a higher pitch, or shouting (18,24). Consequently, we developed a second group of stimuli to elicit improved voice production. These stimuli included samples of whispering, shouting, musical scales, and sustaining vowels at both modal and falsetto pitches—all of which have been shown to temporarily improve voice production in speakers with spasmodic dysphonia (18,23-28).

Videotapes

We developed two videotapes of speech samples to provide stimuli for consensual validation. We did all videotaping in a quiet office while speakers did the stimulus tasks as part of their initial voice evaluation. We recorded them on a Sony Hi-8 video cassette camera recorder (model EVO-9100) with a Shure headset-mounted unidirectional microphone (model SM10A) and Sony Hi-8 professional video cassettes (model P6-120 HMPX). The samples were dubbed onto the master videotape using a Sony Hi-8 video cassette recorder (model EVO-970). A commercial dubbing company produced multiple copies of the videotape on Maxell video cassettes (model FO42A) in VHS and PAL formats.

All speakers on the tapes exhibited the perceptual voice characteristics associated with adductor spasmodic dysphonia, hyperadduction of the vocal
ADDUCTOR SPASMODIC DYSPHONIA

folds, and no other neurological symptoms. We accepted all speakers for this study only on verification of the spasmodic dysphonia diagnosis and qualification for botulinum toxin injection. All speakers were diagnosed unanimously as having adductor spasmodic dysphonia by an otolaryngologist (AB), a neurologist (MB), and a speech-language pathologist (CS). If the three examiners were not in agreement, we did not include the speaking sample. Their evaluations consisted of a detailed history, a perceptual voice evaluation of severity and phonatory behavior that elicits changes in voice production, a fiber optic laryngeal examination, and a complete neurological assessment of those tasks that classically bring out movement disorders. The speakers on the videotape reflected the full range of symptom severity seen at the Columbia-Presbyterian Medical Center.

Two scales of severity were used on the rating form. A seven-point continuous scale of severity was used to rate symptoms observed in connected speech. The seven-point scoring system was used because increasing the number of scale values used to rate speaker behavior decreases rating errors (33–35). On the other hand, a two-point dichotomous scale of within normal limits or not within normal limits was used to rate the voice tasks that improve or worsen the perceptual symptoms of spasmodic dysphonia.

Definitions of symptoms
We developed definitions of the symptoms so that all voice clinicians would rate the symptoms in the same way. Definitions were first compiled from the literature. We modified the definitions through a series of consensual evaluations and then the judges re-evaluated the definitions [16,24,32,36–38; Allen EL (personal communication)].

Questionnaires
We developed three questionnaires to gather evidence of consensual validity. The first was an open-ended form on which the judges wrote criticism of the existing stimuli and symptoms. On the second and third, the judges made narrative comments and rated the importance of the stimulus tasks on the protocol, symptoms on the rating form, definitions, and videotape, using seven-point scales. For this study, they defined high content validity as a rating of 5, 6, or 7 on the seven-point questionnaires.

Procedures
We developed the USDRS through a series of consensual evaluations. To collect content validity data on the instructions, the rating form, and the definitions, we mailed three successive versions of the USDRS to experienced voice clinicians. Consensual validity was arrived at by recursively compiling content validations and resubmitting these syntheses to the judges for their comments. All comments helped to refine and focus subsequent versions of the scale that we then resubmitted for further evaluation.

First consensual evaluation
We sent the stimulus tasks, rating form, definitions, and first questionnaire to 26 voice clinicians. After they studied the stimulus task, rating form,
and definitions, they completed the first questionnaire. We carried out suggestions made by two or more clinicians.

Second consensual evaluation
The second assessment of consensual validity included evaluation of content validity and interrater reliability. The 26 voice clinicians received the revised stimulus tasks, rating form, definitions, and the first videotape of 15 speakers. The clinicians studied the rating form and definitions and then used the second rating form to rate videotaped samples of speakers reading “The Rainbow Passage.” After using the USDRS, the clinicians completed a seven-point questionnaire appraising its validity. We retained individual symptoms and voice stimuli when more than 50% of the judges rated items as 5, 6, or 7 on the questionnaire.

Third consensual evaluation
The final consensual assessment also included evaluation of both content validity and interrater reliability. The 169 clinicians received the modified stimulus tasks, rating form, definitions, and the second videotape of a training segment and 15 speakers. The clinicians studied the form and definitions, and then practiced rating patients on the training tape. They practiced with the training segment until their scores matched the suggested scores with 90% accuracy. After these training procedures, the judges scored the 15 additional samples of “The Rainbow Passage,” “The Cookie Theft Picture,” and the voice tasks. The clinicians then completed the third questionnaire.

We calculated intraclass correlations to assess interjudge reliability in continuous ratings where a random sample of judges (K) rates samples (N). The sources of the variance were partitioned into speaker mean square (PMS), judge mean square (JMS), and residual mean square (EMS). Fleiss’s formula (39; p. 26) for an estimator of the intraclass correlation coefficient for random raters is

\[
R = \frac{N \times (PMS - EMS)}{N \times PMS + K \times JMS + (NK - N - K) \times EMS}
\]

where reliability (R) is the “proportion of variance of an observation due to subject-to-subject [judge] variability in error-free scores” (39; p. 3). We analyzed reliability of dichotomous symptoms and stimuli rated on two-point categorical scales separately using a kappa (κ) statistic. The range of a reliability coefficient is 0.0 to +1.0, but as Fleiss (40) says, “no universally applicable standards are possible for what makes up poor, fair, or good reliability. In general, values of R below 0.4 or so may be taken to represent poor reliability, values above 0.75 or so may be taken to represent excellent reliability, and values between 0.4 and 0.75 may be taken to represent fair to good reliability” (40; p. 7). Therefore, an individual dimension or stimulus was kept on the scale only when fair to good or better interrater reliability was shown and when more than 60% of the judges rated it as a 5, 6, or 7 on the final questionnaire.

RESULTS

Many voice judges responded to each of the three consensual evaluations: 22 of 26 on the first; 25 of 26 on the second; and 118 of 169 on the third.

Validity
On the second consensual evaluation, 50 to 87% of the voice judges rated each stimulus task and 47 to 100% of the judges rated each symptom on the rating form as having high content validity. On the third consensual evaluation, 69 to 96% of the voice judges rated each stimulus task and 53 to 98% of the judges rated each symptom on the rating form and 83% rated the training segment as having high content validity (Table 1). In addition, narrative comments about the items were positive on all three questionnaires.

Reliability
On the second consensual evaluation, 25 judges’ ratings of “The Rainbow Passage” suggest good to excellent interrater reliability for only two symptoms, overall severity (R = 0.786) and strained-strangled voice quality (R = 0.768) and fair to good interrater reliability for seven: expiratory effort (R = 0.686), abrupt voice initiation (R = 0.537), pitch breaks (R = 0.525), voice loudness (R = 0.494), rough voice quality (R = 0.435), speech intelligibility reduced (R = 0.423), and breathy voice quality (R = 0.406). Poor reliability was found for the other eight symptoms.

For the third consensual evaluation, we calculated interrater reliability for the 118 voice judges’ ratings of “The Rainbow Passage” and “The Cookie Theft Picture” after viewing the training tape. Six symptoms exhibited good to excellent reliability: overall severity (R = 0.909), strained-
**TABLE 1.** Validity of tasks on the protocol and symptoms on the rating form from the third questionnaire

<table>
<thead>
<tr>
<th>Training Tape</th>
<th>RATINGS 5 to 7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stimuli</strong></td>
<td>83%</td>
</tr>
<tr>
<td>Take a deep breath and say a sustained “ah”</td>
<td>96%</td>
</tr>
<tr>
<td>Loudly say: “Taxi”</td>
<td>95%</td>
</tr>
<tr>
<td>Make your voice go up and down like mine</td>
<td>91%</td>
</tr>
<tr>
<td>Whisper “Shh, the baby is sleeping”</td>
<td>90%</td>
</tr>
<tr>
<td>Take a deep breath and say a sustained “e”</td>
<td>89%</td>
</tr>
<tr>
<td>Say a sustained “e” in falsetto</td>
<td>88%</td>
</tr>
<tr>
<td>Count to 10</td>
<td>87%</td>
</tr>
<tr>
<td>Spontaneously laugh</td>
<td>86%</td>
</tr>
<tr>
<td>“Ambling along Rainey Island Avenue”</td>
<td>80%</td>
</tr>
<tr>
<td>Cough</td>
<td>79%</td>
</tr>
<tr>
<td>“He saw half a sea shell”</td>
<td>75%</td>
</tr>
<tr>
<td>Quietly say “Shh, the baby is sleeping”</td>
<td>73%</td>
</tr>
<tr>
<td>Say the alphabet from A to N</td>
<td>69%</td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
<td>98%</td>
</tr>
<tr>
<td>Strained-strangled voice quality</td>
<td>94%</td>
</tr>
<tr>
<td>Overall severity</td>
<td>93%</td>
</tr>
<tr>
<td>Expiratory effort</td>
<td>87%</td>
</tr>
<tr>
<td>Voice tremor</td>
<td>86%</td>
</tr>
<tr>
<td>Related movements and grimaces</td>
<td>85%</td>
</tr>
<tr>
<td>Speech intelligibility reduced</td>
<td>84%</td>
</tr>
<tr>
<td>Rough voice quality</td>
<td>84%</td>
</tr>
<tr>
<td>Voice arrest</td>
<td>84%</td>
</tr>
<tr>
<td>Aphonla</td>
<td>79%</td>
</tr>
<tr>
<td>Areas of movement</td>
<td>78%</td>
</tr>
<tr>
<td>Breathy voice quality</td>
<td>75%</td>
</tr>
<tr>
<td>Abrupt voice initiation</td>
<td>74%</td>
</tr>
<tr>
<td>Speech rate</td>
<td>71%</td>
</tr>
<tr>
<td>Voice loudness</td>
<td>67%</td>
</tr>
<tr>
<td>Deviated pitch</td>
<td>63%</td>
</tr>
<tr>
<td>Bursts of loudness</td>
<td>61%</td>
</tr>
<tr>
<td>Other</td>
<td>55%</td>
</tr>
<tr>
<td>Atypical intonation</td>
<td>53%</td>
</tr>
</tbody>
</table>

Note: Rating consisted of a seven-point continuous scale: 1) no importance through 7) essential. N = 116. Two judges did not complete the questionnaire. This table is excerpted from Stewart, 1994 (43).

Strangled voice quality (R = 0.886), voice arrest (R = 0.862), abrupt voice initiation (R = 0.799), aphonia (R = 0.790), and expiratory effort (R = 0.787). Eight continuous symptoms showed fair to good reliability: speech intelligibility reduced (R = 0.740), speech rate (R = 0.712), related movements and grimaces (R = 0.685), bursts of loudness (R = 0.670), breathy voice quality (R = 0.629), rough voice quality (R = 0.535), voice tremor (R = 0.500), areas of movement (R = 0.411), and voice loudness (R = 0.410). Two symptoms showed poor reliability. Confidence intervals were narrower for overall severity, strangled-strangled voice quality, and voice arrest than for the other symptoms (Table 2).

On the third consensual evaluation, judges also scored 13 voice and speech tasks on a two-point scale. According to Fleiss’s (40) guideline, these data suggest fair to good agreement beyond chance for six tasks on the protocol: loudly say “Taxi” (κ = 0.627), whisper “shh, the baby is sleeping” (κ = 0.570), say a sustained “e” in falsetto (κ = 0.557), make your voice go up and down like mine (κ = 0.514), take a deep breath and say a sustained “ah” (κ = 0.485), and take a deep breath and say a sustained “e” (κ = 0.425). Poor agreement was found on the other seven tasks (Table 3).

Following the third consensual review modifications were made to the USDRS. We included only those items that showed either excellent or fair to good interrater reliability and high content validity. Consequently, 16 symptoms and 6 voice tasks were retained on the scale (Tables 4 and 5).

**DISCUSSION AND CONCLUSIONS**

The findings suggest that the objectives of identifying the symptoms of adductor spasmodic dysphonias and rating their severity have been largely realized. We developed a standard measure that considered the perceptual symptoms observed in connected speech and definitions of those symp-

**TABLE 2.** Evaluation of reliability by intraclass correlation of continuous symptoms: 118 judges from the second videotape

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>R¹</th>
<th>LB²</th>
<th>UB³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall severity</td>
<td>0.909</td>
<td>0.833</td>
<td>0.958</td>
</tr>
<tr>
<td>Strained-strangled voice quality</td>
<td>0.886</td>
<td>0.800</td>
<td>0.947</td>
</tr>
<tr>
<td>Voice arrest</td>
<td>0.862</td>
<td>0.758</td>
<td>0.935</td>
</tr>
<tr>
<td>Abrupt voice initiation</td>
<td>0.799</td>
<td>0.658</td>
<td>0.868</td>
</tr>
<tr>
<td>Aphonla</td>
<td>0.790</td>
<td>0.654</td>
<td>0.896</td>
</tr>
<tr>
<td>Expiratory effort</td>
<td>0.787</td>
<td>0.644</td>
<td>0.895</td>
</tr>
<tr>
<td>Speech intelligibility reduced</td>
<td>0.740</td>
<td>0.580</td>
<td>0.868</td>
</tr>
<tr>
<td>Speech rate</td>
<td>0.712</td>
<td>0.555</td>
<td>0.830</td>
</tr>
<tr>
<td>Related movements and grimaces</td>
<td>0.685</td>
<td>0.516</td>
<td>0.834</td>
</tr>
<tr>
<td>Bursts of loudness</td>
<td>0.670</td>
<td>0.498</td>
<td>0.825</td>
</tr>
<tr>
<td>Breathy voice quality</td>
<td>0.629</td>
<td>0.455</td>
<td>0.797</td>
</tr>
<tr>
<td>Rough voice quality</td>
<td>0.535</td>
<td>0.364</td>
<td>0.726</td>
</tr>
<tr>
<td>Voice tremor</td>
<td>0.500</td>
<td>0.325</td>
<td>0.700</td>
</tr>
<tr>
<td>Areas of movement</td>
<td>0.411</td>
<td>0.259</td>
<td>0.616</td>
</tr>
<tr>
<td>Voice loudness</td>
<td>0.410</td>
<td>0.253</td>
<td>0.618</td>
</tr>
<tr>
<td>Atypical intonation</td>
<td>0.378</td>
<td>0.225</td>
<td>0.587</td>
</tr>
<tr>
<td>Deviated pitch</td>
<td>0.360</td>
<td>0.213</td>
<td>0.568</td>
</tr>
</tbody>
</table>

Note: R¹ = Intraclass correlation judges reliability for the symptoms on the fourth draft; LB² = the lower boundary of the confidence level; UB³ = the upper boundary of the confidence level. N = 118. This table is excerpted from Stewart, 1994 (43). Journal of Voice, Vol. II, No. 1, 1997
TABLE 3. Evaluation of reliability by kappa of categorical stimuli: 118 judges from the second videotape

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loudly say: &quot;Taxi&quot;</td>
<td>0.627</td>
</tr>
<tr>
<td>Say a sustained &quot;e&quot; in falsetto</td>
<td>0.557</td>
</tr>
<tr>
<td>Whisper &quot;Shh, the baby is sleeping&quot;</td>
<td>0.570</td>
</tr>
<tr>
<td>Make your voice go up and down like mine</td>
<td>0.514</td>
</tr>
<tr>
<td>Take a deep breath and say a sustained &quot;ah&quot;</td>
<td>0.485</td>
</tr>
<tr>
<td>Take a deep breath and say a sustained &quot;e&quot;</td>
<td>0.425</td>
</tr>
<tr>
<td>Count to 10</td>
<td>0.334</td>
</tr>
<tr>
<td>Quietly say &quot;Shh, the baby is sleeping&quot;</td>
<td>0.314</td>
</tr>
<tr>
<td>Spontaneously laugh</td>
<td>0.260</td>
</tr>
<tr>
<td>Say the alphabet from A to N</td>
<td>0.249</td>
</tr>
<tr>
<td>&quot;'He saw half a sea shell&quot;</td>
<td>0.213</td>
</tr>
<tr>
<td>&quot;’Ambling along Rainey Island Avenue&quot;</td>
<td>0.043</td>
</tr>
<tr>
<td>Cough</td>
<td>0.030</td>
</tr>
</tbody>
</table>

Note: Kappa evaluates the extent to which the judges were reliable. N = 118. This table is excerpted from Stewart, 1994 (43).

TABLE 4. The USDRS—conversational speech and rating form

Patient Name: ____________________________ Date: ___________ Rater: ___________

Scale for rating severity: (circle 1)
1 no instance
2 mild
3 mild-to-moderate
4 moderate
5 moderate-to-severe
6 severe
7 profound

Overall Severity: 1 2 3 4 5 6 7
Rough Voice Quality: 1 2 3 4 5 6 7
Breathy Voice Quality: 1 2 3 4 5 6 7
Strained-Strangled Voice Quality: 1 2 3 4 5 6 7
Abrupt Voice Initiation: 1 2 3 4 5 6 7
Voice Arrest: 1 2 3 4 5 6 7
Aphonia: 1 2 3 4 5 6 7
Voice Loudness: 1 2 3 4 5 6 7
Bursts of Loudness: 1 2 3 4 5 6 7
Voice Tremor: 1 2 3 4 5 6 7
Expiratory Effort: 1 2 3 4 5 6 7
Speech Rate: 1 2 3 4 5 6 7
Speech Intelligibility Reduced: 1 2 3 4 5 6 7
Related Movements and Grimaces: 1 2 3 4 5 6 7
(circle all that apply then rate the most severe and mark it with an asterisk) upper face, lower face, neck, chest, shoulders

Note: Patients with adductor spasmodic dysphonia were chosen to evaluate the reliability of the USDRS. The hypothesis that we can quantify the voices of patients with adductor spasmodic dysphonia with high reliability were borne out for six of the perceptual symptoms, fair to good for eight additional symptoms, and six voice tasks that improve or worsen perceptual symptoms. The reliability of their ratings improved when the voice judges watched a training tape. It appears reasonable that other voice clinicians with appropriate training, could learn to use this scale and achieve highly comparable results. Training would include studying the definitions and using them as guidelines for rating the form and practicing scoring the training tape until 95% agreement is achieved with the scores on the tape.

Several other researchers have evaluated interrater reliability for ratings of highly sophisticated pretrained voice judges. Unfortunately, we cannot directly compare results among the studies (5-7) or with this study because the symptoms are different and different statistical procedures were used. Izdebski et al. calculated interrater reliability coefficient on breathiness, overpressure, aperiodicity, and tremor using analysis of variance statistics. The mean coefficient of interrater reliability was 0.88 and the range was 0.73 to 0.95. (7; p. 97). Sapir and Aronson found high interrater and intrarater reliability for three judges on their overall impressions of normality or abnormality ... based on such criteria as reduced intensity, or duration of
The following definitions were developed by blending definitions* (and response from the experienced voice judges).†

**Overall Severity** is an estimate of the global severity of the voice and speech disorder.

**Rough Voice Quality** is an abnormal voice quality that sounds harsh, hoarse, raspy, coarse, or uneven.

**Breathy Voice Quality** is an abnormal voice quality with a quavering sound similar to a vibrato.

**Rough Voice Quality** is an abnormal voice quality that sounds harsh, hoarse, raspy, coarse, or uneven.

**Abrupt Voice Initiation** is an abnormal initiation of phonation where sound is produced by first stiffening the vocal folds then adducting the vocal folds to the midline, building up pressure below the glottis, and finally initiating phonation; an explosive sound called the glottal attack.

**Breathy Voice Quality** is an abnormal voice quality with a quavering sound similar to a vibrato.

**Aphonia** is an abnormal involuntary whisper or loss of voice.

**Voice Arrest** is an abnormal unexpected interruption in voice production during a voiced sound, suggesting either the impedance of airflow by the vocal folds closing too tightly, or the reverse, a decrease in resistance to airflow by a sudden opening of the vocal folds.

**Aphonia** is an abnormal involuntary whisper or loss of voice.

**Stressed Strangled Voice Quality** is an abnormal voice quality that is squeezed and effortful.

**Abrupt Voice Initiation** is an abnormal initiation of phonation where sound is produced by first stiffening the vocal folds then adducting the vocal folds to the midline, building up pressure below the glottis, and finally initiating phonation; an explosive sound called the glottal attack.

**Stressed Strangled Voice Quality** is an abnormal voice quality that is squeezed and effortful.

**Voice Tremor** is an abnormal involuntary sudden increase or decrease in loudness.

**Abrupt Voice Initiation** is an abnormal initiation of phonation where sound is produced by first stiffening the vocal folds then adducting the vocal folds to the midline, building up pressure below the glottis, and finally initiating phonation; an explosive sound called the glottal attack.

**Strained Strangled Voice Quality** is an abnormal voice quality that is squeezed and effortful.

**Rough Voice Quality** is an abnormal voice quality that sounds harsh, hoarse, raspy, coarse, or uneven.

**Burst of Loudness** is an abnormal involuntary sudden increase or decrease in loudness.

**Strained Strangled Voice Quality** is an abnormal voice quality that is squeezed and effortful.

**Speech Rate** is the speed of producing utterances; reduced speed is related to lengthened consonants and vowels or to unexpected pauses between sounds, syllables, or words; increased speed is related to shortened consonants and vowels or decreased pauses between phrases and sentences.

**Speech Rate** is the speed of producing utterances; reduced speed is related to lengthened consonants and vowels or to unexpected pauses between sounds, syllables, or words; increased speed is related to shortened consonants and vowels or decreased pauses between phrases and sentences.

**Speech Intelligibility Reduced** is the quality of overall clarity with which an utterance can be understood or comprehended by the average listener.

**Speech Intelligibility Reduced** is the quality of overall clarity with which an utterance can be understood or comprehended by the average listener.

**Related Movements and Grimaces** are the involuntary visible changes in a patient’s body that accompany effortful speech including flushing of the face; facial tics such as lip movements, eye-blinking, and frowning; and movements in the neck, shoulder, arm, and leg.

* References 16, 24, 33, 36-38; Allen EL (personal communication).
† This table is excerpted from Stewart, 1994 (43).

---

strained hoarseness, smoother voice flow, reduced breathiness, increased tonality (as opposed to tonality or "noise" in the vocal tone) and fewer phonatory and respiratory indicators of effort or struggle" (5; p. 200). They calculated interrater reliability using the Pearson Product correlation coefficients as 0.83 between judge one and two, 0.87 between judges one and three, and 0.84 between judges two and three (5; p. 103). Sapi et al. found that interrater reliability coefficients were high and ranged from 0.830 to 0.877 for voice quality and from 0.802 to 0.882 on phonatory effort (6; p. 138).

These earlier studies have laid the groundwork for developing this rating scale that less experienced voice judges can use. The judges from previous studies were highly sophisticated and had extensive training and experience with patients with spasmodic dysphonia. When our judges used the training tape and definitions when scoring the USDRS, their responses showed good interrater reliability. Although our study results cannot be directly compared with these results from highly trained researchers, our interclass correlations for seven symptoms were high like those reported previously in the literature: overall (R = 0.91), strained-strangled voice quality (R = 0.89), voice arrest (R = 0.86), abrupt voice initiation (R = 0.80), aphonia (R = 0.79), expiratory effort (R = 0.79), and speech intelligibility reduced (R = 0.74).

The USDRS is a good tool for quantifying the voice symptoms of patients with adductor spasmodic dysphonia. Based on the results of our study, speech-language pathologists can use the USDRS to perceptually evaluate speakers with adductor spasmodic dysphonia (see Tables 4-6). Identifying the severity of the spectrum of the symptoms may help guide therapy. In addition, voice clinicians can evaluate the effectiveness of therapy more effectively when they use these 14-voice symptoms to evaluate the subtleties of a patient’s disorder.

Hirano describes five "major purposes of clinical examination of voice patients . . . (1) to diagnose the etiologic disease(s), (2) to determine the degree and the extent of the etiologic disease(s), (3) to evaluate the degree and the nature of dysphonia, (4) to determine the prognosis and (5) to monitor changes" (42; p. 89). Hirano says that no single measure can evaluate all these aspects. The USDRS can help voice specialists describe the third aspect by rating the severity and importance of perceptual symptoms.

Hirano also says "perceptual evaluation of the voice is a very important and useful procedure. . . . If we had an international standard for the protocol for perceptual evaluations of the voice, we would be able to compare the results across different periods all over the world" (42; p. 140). We believe that the USDRS can begin the process to satisfy Hirano’s concerns about allowing comparison of research across centers, researchers, and subjects (42). Through its use we may specify and compare the relative voice symptoms of one group with other
TABLE 7. Instructions

Before completing the scale, enter the patient identification information at the top of the rating form. Become familiar with the patient’s history including the development of the voice disorder, fluctuations in voice production, medical history, educational level, vocation, and awareness of spasmodic dysphonia.

If possible, videotape the evaluation to document the symptoms and their severity. If videotape equipment is not available, audiotape the evaluation. The tape can be replayed to aid in evaluation of the symptoms and to document progress. Check the microphone so that it is not blocking the view of the mouth and keep the camera and microphone distance constant. Have speakers remove bulky clothing and ask them to sit and look ahead while reading. Encourage speakers to talk normally and to refrain from tricks or compensatory techniques. The protocol takes approximately 10 minutes to do.

The rating form contains two parts: (1) Reading and Conversational Speech and (2) Voice and Speech Tasks. We rate both reading and conversational speech on the Conversational Speech and Reading section of the rating form. On the other hand, we score only the voice and speech tasks on the Voice and Speech Tasks section.

Conversational Speech and Reading

Before administering the scale, study the definitions and use them as guidelines when marking the rating form; then view the training tape and mark score sheets for the speakers on the tape. When 95% agreement is achieved with the scores on the training tape, you may begin to use the USDRS clinically.

Symptoms of spasmodic dysphonia usually occur during connected speech and not on isolated words or sounds. We designed the parameters to identify and quantify the symptoms observed most frequently in patients with adductor spasmodic dysphonia. Therefore, this section of the scale is used to evaluate the symptoms and their severity during monologue and reading aloud.

Ask the patient to read the first paragraph of “The Rainbow Passage” (21) aloud and to describe “The Cookie Theft Picture” (22). While the speaker is reading aloud, the judge marks the Conversational Speech and Reading section of the rating form by circling the number that most closely describes the severity of each parameter.

Scale for rating severity: 1 no instance (circle 1) 2 mild 3 mild-to-moderate 4 moderate 5 moderate-to-severe 6 severe 7 profound

If a speaker’s voice is not breathy and is moderately reduced in loudness, mark the Conversational Speech and Reading section of the rating form in the following way:

Breathy Voice Quality


Voice Loudness


If a patient speaks at a normal rate, mark the Conversational Speech and Reading section of the rating form as follows:

Speech Rate Normal = 4

If a patient speaks with a slow rate, mark the Conversational Speech and Reading section of the rating form as follows:

Speech Rate Normal = 4

If a patient exhibits moderate related movements in the lower face and mild movements in the upper face, neck, and chest, mark the Conversational Speech and Reading section of the rating form like this:

Related Movements and Grimaces

(circle all that apply then rate the most severe and mark it with an asterisk, as noted)

1 2 3 4 5 6 7

upper face, lower face,*

chest, neck, shoulders (continues, next column)

TABLE 7 (Continued)

Voice and Speech Tasks

Ask the speaker to turn to the Voice and Speech Task section of the protocol and to follow the instructions. Mark the rating form while the speaker is doing the tasks and make note of any fluctuation in voice production.

Use the second page of the rating form to score the Voice and Speech Tasks. The tasks are listed in an abbreviated form in the left column. If the task is produced with normal voice, mark the Voice and Speech Tasks in the within normal limits category. The speakers do not need to produce a specific type of disorder. If any deviation in voice production is observed, mark the not within normal limits category. If the patient’s voice improves with a task but is not normal, mark it in the not within normal limits column and make a note in the comments column.

If you wish to have patients repeat the voice and speech tasks, it is acceptable but not required. Repetition should not alter the patient’s performance and may allow more confidence about your ratings.

Say a sustained “Ah”

Ask the speaker to produce sustained vowels in the modal range. Because of the frequent improvement in speech production when speakers sustain vowels, we include two vowels in the protocol. Any variation between the two vowels should be identified on the rating form.

Say a sustained “E”

Some speakers produce one vowel more clearly than the other.

Make your voice go up and down.

Instruct the speaker to imitate your voice as you produce a slide from model register to loft and then to pulse register. Speakers may exhibit improvement in voice production in one or more register. We can evaluate and note vocal control during pitch changes on the rating form.

Say a sustained “E” in falsetto

Instruct the speaker to imitate your production of an /i/ in falsetto. Voice production in falsetto is usually appropriate except when speakers have severe spasmodic dysphonia. Thus, falsetto production can be helpful when quantifying the severity of the disorder.

Whisper: “Shh, the baby is sleeping.”

If the speaker does not vary the loudness of the voice, demonstrate appropriate production for the whisper and shout. Because a whisper is aphonie, speakers with spasmodic dysphonia frequently produce a whisper appropriately; however, some speakers have difficulty whispering.

Loudly say: “Taxi”

Ask the speaker to shout “Taxi.” Shouted speech is usually appropriate even when patients have very severe spasmodic dysphonia. However, shouting can be reduced in loudness and voice arrest can occur.

If a speaker produces a strained sustained “Ah” and a clear voice at the extremes but strained-stranded voice in speaking range, mark the Voice and Speech Task section of the rating form in the following way:

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Within Normal Limits</th>
<th>Not Within Normal Limits</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Say a sustained “Ah”</td>
<td>X Strained</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Make your voice go up and down</td>
<td>X Voice clear</td>
<td></td>
<td>at extremes*</td>
</tr>
</tbody>
</table>
groups, thus making research findings more meaningful.

Acknowledgment: We gratefully acknowledge the assistance of the 118 voice judges and 30 speakers with spasmodic dysphonia who participated in the standardization of the USDRS. Without their cooperation, assistance, and suggestions this scale could not have been completed. Special thanks to Dr. Arnold Aronson for his early and continued assistance with this enterprise.

REFERENCES