The Role of Laryngoscopy in the Diagnosis of Spasmodic Dysphonia

Pedram Daraei, BS; Craig R. Villari, MD; Adam D. Rubin, MD; Alexander T. Hillel, MD; Edie R. Hapner, PhD; Adam M. Klein, MD; Michael M. Johns III, MD

**IMPORTANCE** Spasmodic dysphonia (SD) can be difficult to diagnose, and patients often see multiple physicians for many years before diagnosis. Improving the speed of diagnosis for individuals with SD may decrease the time to treatment and improve patient quality of life more quickly.

**OBJECTIVE** To assess whether the diagnosis of SD can be accurately predicted through auditory cues alone without the assistance of visual cues offered by laryngoscopic examination.

**DESIGN, SETTING, AND PARTICIPANTS** Single-masked, case-control study at a specialized referral center that included patients who underwent laryngoscopic examination as part of a multidisciplinary workup for dysphonia. Twenty-two patients were selected in total: 10 with SD, 5 with vocal tremor, and 7 controls without SD or vocal tremor.

**INTERVENTIONS** The laryngoscopic examination was recorded, deidentified, and edited to make 3 media clips for each patient: video alone, audio alone, and combined video and audio. These clips were randomized and presented to 3 fellowship-trained laryngologist raters (A.D.R., A.T.H., and A.M.K.), who established the most probable diagnosis for each clip. Intrarater and interrater reliability were evaluated using repeat clips incorporated in the presentations.

**MAIN OUTCOMES AND MEASURES** We measured diagnostic accuracy for video-only, audio-only, and combined multimedia clips. These measures were established before data collection. Data analysis was accomplished with analysis of variance and Tukey honestly significant differences.

**RESULTS** Of patients with SD, diagnostic accuracy was 10%, 73%, and 73% for video-only, audio-only, and combined, respectively ($P < .001$, $df = 2$). Of patients with vocal tremor, diagnostic accuracy was 93%, 73%, and 100% for video-only, audio-only, and combined, respectively ($P = .05$, $df = 2$). Of the controls, diagnostic accuracy was 81%, 19%, and 62% for video-only, audio-only, and combined, respectively ($P < .001$, $df = 2$).

**CONCLUSIONS AND RELEVANCE** The diagnosis of SD during examination is based primarily on auditory cues. Viewing combined audio and video clips afforded no change in diagnostic accuracy compared with audio alone. Laryngoscopy serves an important role in the diagnosis of SD by excluding other pathologic causes and identifying vocal tremor.
Spasmodic dysphonia (SD), a focal neurologic dystonia, affects nearly 50,000 individuals in the United States. It is a central motor dysfunction resulting in a focal dystonia limited to the laryngeal musculature. The cause is unclear, and there do not appear to be environmental or hereditary patterns in SD. Regardless of the origin, SD leads to discoordination of vocal fold approximation during phonation and resultant characteristic auditory patterns. Seen more commonly in females, the mean age of onset is 45 years but has been reported to range from 13 to 71 years. It may occur in isolation of other neurologic conditions, but there are known associations with nonvocal tremor.

The 2 major categories of SD—adductor and abductor—are distinguished by vocal fold motion during the dystonia, resulting in the unique vocal characteristics of the given subtype (Table 1). In adductor SD, the patient has an unstable voice with glottal stops on vowel-laden phonation. Conversely, abductor SD is characterized by breathy, hypophonic vocal breaks most evident on voiceless to voiced phonemes.

The workup for a patient with dysphonia begins with a history and physical examination, often including laryngoscopy to visualize the vocal folds during phonation. Unfortunately, there is often a delay in diagnosis that spans multiple years. In 2008, an expert review determined that identifying methods of diagnosis and comparing SD with other focal dystonias are the highest priorities in research related to SD. This current study serves to investigate the role of laryngoscopy vs auditory cues in the diagnosis of SD. We also explore the importance of visual and auditory cues in diagnosing laryngeal tremor to illustrate key diagnostic aspects of SD.

### Methods

#### Patient Selection

Institutional review board approval was obtained from Emory University School of Medicine with patient informed consent obtained in writing. This study includes patients seen at the Emory Voice Center in Atlanta, Georgia. Each patient underwent a multidisciplinary workup, including evaluation by a fellowship-trained laryngologist (M.M.J.) and a speech-language pathologist (E.R.H.) with more than 30 years’ experience in working with SD. Patients were selected from a cohort of patients seen during a 4-year period. Ten patients with typical features of SD, 5 with typical features of vocal tremor, and 7 controls were selected. Patients with SD were selected based on classic features previously described in the literature. All 10 patients had adductor-type SD. Control patients included 3 with unilateral vocal fold paralysis, 1 with muscle tension dysphonia, 1 with vocal fold granuloma, 1 with a vocal fold polyp, and 1 without disease.

#### Laryngoscopic Examination

The physical examination included flexible laryngoscopy. While undergoing laryngoscopic examination, patients were asked to phonate and sustain the letter E, read the first paragraph of the Rainbow Passage, and repeat the following phrases: “we eat eels every day,” “each time it oozes blue,” “she speaks pleasingly,” “Peter needs help at the peak,” “wait for me,” and “no, not now.” The laryngoscopic examinations and associated audio were recorded and stored on a secure server.

#### Multimedia Clip Formation and Editing

The initial laryngoscopic recordings for each patient were edited using iMovie (version 9.0.8; Apple, Inc.). First, the recordings were edited to remove extraneous footage, including, but not limited to, external visualization of the patient’s face before insertion into the nose, advancement through the nose to the hypopharynx, swallowing, and coughing. Next, the clips were edited to remove any information that could be considered identifying. This included name, age, sex, and dialogue between the clinician and the patient. The remaining footage was further edited to create a multimedia clip showing the patients’ vocal folds as they repeated the set of phonatory tasks described earlier. No clip exceeded 35 seconds. The multimedia clips were further edited to separate the audio from the video, creating 3 individual media clips for each patient—1 with video alone, 1 with audio alone, and 1 with combined audio and video (herein referred to as combined).

The clips were then randomized and inserted into a PowerPoint (version 14.3; Microsoft Corp) presentation. The presentation included a total of 79 slides—in introductory slide containing instructions for the raters, 26 consecutive slides containing video-only clips, 26 consecutive slides containing audio-only clips, and 26 consecutive slides with combined clips. The instructions slide asked the raters to view the presented media clips and determine a most probable diagnosis. The diagnosis form was open ended (ie, raters were not given choices to select from). A rating form was also created, asking the raters for the “most probable diagnosis” for each media clip.

The PowerPoint presentations were then sent to 3 experienced fellowship-trained laryngologists (A.D.R., A.T.H., and A.M.K.) masked to the purpose of the study. These raters were asked to open and complete the PowerPoint presentation with the corresponding form. No additional directions were given. Data were collected on the diagnosis forms and transferred to a secure database on a password-protected computer.

#### Reliability

Four video-only clips, 4 audio-only clips, and 4 combined clips were repeated in the presentation to assess for intrarater reliability, which was measured using the Cohen κ coefficient.

### Table 1. Classic Features of Spasmodic Dysphonia

<table>
<thead>
<tr>
<th>Feature</th>
<th>Adductor Spasmodic Dysphonia</th>
<th>Abductor Spasmodic Dysphonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of voice</td>
<td>Strain, strangled, harsh</td>
<td>Breathy, hypophonie</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Tremor, inappropriate pitch or pitch breaks, breathiness, glottal fry</td>
<td>Abrupt termination of voicing</td>
</tr>
</tbody>
</table>

Correlation of the Rainbow Passage, and repeat the following phrases: “we eat eels every day,” “each time it oozes blue,” “she speaks pleasingly,” “Peter needs help at the peak,” “wait for me,” and “no, not now.” The laryngoscopic examinations and associated audio were recorded and stored on a secure server.

The initial laryngoscopic recordings for each patient were edited using iMovie (version 9.0.8; Apple, Inc.). First, the recordings were edited to remove extraneous footage, including, but not limited to, external visualization of the patient’s face before insertion into the nose, advancement through the nose to the hypopharynx, swallowing, and coughing. Next, the clips were edited to remove any information that could be considered identifying. This included name, age, sex, and dialogue between the clinician and the patient. The remaining footage was further edited to create a multimedia clip showing the patients’ vocal folds as they repeated the set of phonatory tasks described earlier. No clip exceeded 35 seconds. The multimedia clips were further edited to separate the audio from the video, creating 3 individual media clips for each patient—1 with video alone, 1 with audio alone, and 1 with combined audio and video (herein referred to as combined).

The clips were then randomized and inserted into a PowerPoint (version 14.3; Microsoft Corp) presentation. The presentation included a total of 79 slides—in introductory slide containing instructions for the raters, 26 consecutive slides containing video-only clips, 26 consecutive slides containing audio-only clips, and 26 consecutive slides with combined clips. The instructions slide asked the raters to view the presented media clips and determine a most probable diagnosis. The diagnosis form was open ended (ie, raters were not given choices to select from). A rating form was also created, asking the raters for the “most probable diagnosis” for each media clip.

The PowerPoint presentations were then sent to 3 experienced fellowship-trained laryngologists (A.D.R., A.T.H., and A.M.K.) masked to the purpose of the study. These raters were asked to open and complete the PowerPoint presentation with the corresponding form. No additional directions were given. Data were collected on the diagnosis forms and transferred to a secure database on a password-protected computer.

#### Reliability

Four video-only clips, 4 audio-only clips, and 4 combined clips were repeated in the presentation to assess for intrarater reliability, which was measured using the Cohen κ coefficient.
Interrater reliability was then measured using the Fleiss κ coefficient. Intrarater reliability was determined by comparing repeat media clips assessed by each rater. For example, if a rater identified the first of the 2 repeat clips as muscle tension dysphonia even if the correct diagnosis was SD and the second of the 2 repeat clips was again rated as muscle tension dysphonia, this was considered a successful reidentification. However, if an individual rater diagnosed one of the 2 repeat clips differently, this was considered a failure of reidentification.

Statistical Analysis

Data analysis was performed by first calculating accuracy for the video-only, audio-only, and combined groups. The 3 groups were compared using analysis of variance. Pairwise comparison was then performed using the Tukey honestly significant difference test. Statistical significance for analysis of variance was at the .001 level, given 2 degrees of freedom. For pairwise comparison, adjusted P values were considered significant at the .05 level.

Results

There were 22 patients in total. The mean age was 60.4 years, with a range of 35 to 78 years. There was no significant difference in age between the following groups of patients: SD and vocal tremor, SD and control, and vocal tremor and control (\( P = .27, P = .29, \) and \( P = .19, \) respectively). Other demographic characteristics are shown in Table 2. Statistical analysis results are outlined in Table 3.

Of the patients with SD, the raters reviewed 30 video-only clips, 30 audio-only clips, and 30 combined clips. Only 3 (10%) video-only clips were correctly identified as SD compared with 22 (73%) audio-only clips and 22 (73%) combined clips (\( P < .001, df = 2 \)) (Table 4). Raters were 7.3 times more likely to correctly identify SD in audio-only and combined clips compared with video-only clips (\( P < .001 \) and \( P < .001 \), respectively). There was no significant difference in diagnostic accuracy between audio-only and combined clips (\( P = .99 \)).

Of patients with vocal tremor, the raters reviewed 15 video-only clips, 15 audio-only clips, and 15 combined clips. Fourteen (93%) video-only clips were correctly identified compared with 11 (73%) audio-only clips and 15 (100%) combined clips. There was no statistical difference in raters’ ability to differentiate vocal tremor between video-only, audio-only, and combined clips (\( P = .05, df = 2 \)).

Of control patients, the raters reviewed 21 video-only clips, 21 audio-only clips, and 21 combined clips. Seventeen (81%) video-only clips and 13 (62%) combined clips were correctly identified compared with only 4 (19%) audio-only clips (\( P < .001, df = 2 \)). The difference in diagnostic accuracy was sig-

Table 2. Patient Demographic Data

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall</th>
<th>SD</th>
<th>Vocal Tremor</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5 (23)</td>
<td>1</td>
<td>0</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Female</td>
<td>17 (77)</td>
<td>9</td>
<td>5 (100)</td>
<td>3 (43)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>19 (86)</td>
<td>10</td>
<td>4 (80)</td>
<td>5 (71)</td>
</tr>
<tr>
<td>Black</td>
<td>2 (9)</td>
<td>0</td>
<td>1 (20)</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (4)</td>
<td>0</td>
<td>0</td>
<td>1 (14)</td>
</tr>
</tbody>
</table>

Abbreviation: SD, spasmodic dysphonia.

Table 3. Overall and Pairwise Statistical Analysis

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (( df ))&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Audio vs Video&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Audio vs Combined&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Video vs Combined&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>&lt;.001 (2)</td>
<td>&lt;.001</td>
<td>.99</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vocal tremor</td>
<td>.05 (2)</td>
<td>.18</td>
<td>.05</td>
<td>.82</td>
</tr>
<tr>
<td>Control</td>
<td>&lt;.001 (2)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.34</td>
</tr>
</tbody>
</table>

Abbreviation: SD, spasmodic dysphonia.

<sup>a</sup> Significant at the \( P < .001 \) level.

<sup>b</sup> Significant at the \( P < .05 \) level.

Table 4. Groups, Number of Participants, and Diagnostic Accuracy

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Video Only</th>
<th>Audio Only</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Correct</td>
<td>Accuracy, %</td>
</tr>
<tr>
<td>Cases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>30</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Vocal tremor</td>
<td>15</td>
<td>14</td>
<td>93</td>
</tr>
<tr>
<td>Control</td>
<td>21</td>
<td>17</td>
<td>81</td>
</tr>
</tbody>
</table>

Abbreviation: SD, spasmodic dysphonia.
nificant for video-only compared with audio-only \( (P < .001) \) and for combined compared with audio-only \( (P = .01) \). There was no significant difference in diagnostic accuracy when comparing video-only and combined clips \( (P = .34) \).

Intrarater reliability was evaluated using the repeated media clips. Raters correctly reidentified 32 of 36 (89%) repeat media clips \( (κ = 0.778) \) overall, 10 of 12 (83%) repeat video-only clips \( (κ = 0.667) \), 11 of 12 (92%) repeat audio-only clips \( (κ = 0.832) \), and 11 of 12 (92%) repeat combined clips \( (κ = 0.832) \).

Interrater reliability was also measured. Interrater agreement was 71% \( (κ = 0.414) \) for video-only, audio-only, and combined clips together; 79% \( (κ = 0.576) \) for video-only clips; 61% \( (κ = 0.212) \) for audio-only clips; and 73% \( (κ = 0.455) \) for combined clips.

**Discussion**

The workup for dysphonia involves evaluation by an otolaryngologist and ideally a speech-language pathologist. The examination often includes flexible laryngoscopy to allow for visualization of dynamic laryngeal function and videostroscopy to evaluate mucosal wave propagation and pliability of the vocal fold mucosa, whereby the clinician simultaneously listens to the patient’s voice.

Our study showed that SD is primarily diagnosed through interpretation of auditory cues during diagnostic evaluation. In addition, our results indicate that in the evaluation of a patient with dysphonia, laryngoscopy is used to rule out pathologic causes, such as laryngeal cancer. These findings support the American Academy of Otolaryngology-Head and Neck Surgery guideline on dysphonia, in which the expert committee suggested that laryngoscopy be used when the patient’s dysphonia is thought to be caused by a serious underlying disease. Fellowship-trained laryngologists were 7.3 times more likely to correctly diagnose SD when presented with a media clip that included audio cues (ie, audio-only and combined) compared with video-only cues \( (P < .001) \). When comparing audio-only with combined, there was no difference in diagnostic accuracy, suggesting that audio alone is sufficient for the diagnosis of SD and that visual cues offered by laryngoscopy in the workup of SD are necessary for diagnostic confirmation and to rule out other pathologic causes.

This study also reinforces anecdotal evidence that vocal tremor can be diagnosed effectively with either auditory or visual cues. Interestingly, it appears relatively easier for clinicians to diagnose vocal tremor using clips that include video cues (ie, video-only and combined) compared with auditory cues alone, although no statistically significant relationship was identified. While both SD and vocal tremor are complex neural disorders, vocal tremor appears easier to conceptualize for clinicians given its low-frequency, repetitive, almost rhythmic presentation. The lack of a statistically significant difference for strength of diagnosis between auditory and visual cues in vocal tremor may result from vocal tremor being a more straightforward process in which the physical manifestations more closely match the auditory outcome of the dystonia.

This study highlights the importance of auditory cues in the diagnosis of SD. As shown in the analysis of control data, raters were more likely to diagnose pathologic causes other than SD or vocal tremor using media with video cues (ie, video-only and combined) compared with audio cues alone \( (P < .001) \). Pairwise analysis further strengthens this finding since raters were 4.3 times more likely to correctly identify controls using video-only compared with audio-only and 3.3 times more likely to correctly identify controls using combined compared with audio-only \( (P < .001 \) and \( P < .01 \), respectively). Although diagnostic accuracy was higher for video-only compared with combined clips, this difference was not statistically significant \( (P = .34) \). This supports the role of visual cues in the workup of SD as a means to rule out other laryngeal pathologic causes.

Despite these findings, we recognize that laryngoscopy should be included for patients in whom SD is suspected by history and auditory evaluation given the possibility for coincidental pathologic causes that would not be identified without visualization of the larynx. In addition, laryngoscopy can be used to rule in vocal tremor that is not easily identified through auditory cues alone. Ludlow et al\(^6\) described the stepwise process to diagnose SD by using screening questions, followed by speech examination and nasolaryngoscopy for definitive diagnosis. Our results show that significant emphasis should be placed on a speech examination to identify and diagnose SD. Laryngoscopy should be used with the intent of ruling out other pathologic causes. This is important due to the diagnostic delay that patients experience. Given that SD can be accurately identified on audio cues alone, there is an opportunity for educating nonspecialist clinicians to improve the time to diagnosis and treatment for these patients.

A limitation of the study is the small number of patients who participated. Another limitation is that the multimedia clips were edited from their natural length to fit a 35-second timeframe to decrease rater fatigue during the consecutive evaluation of 52 media clips. As a result, raters were limited in the data they could evaluate to make their diagnosis. In the clinical setting, otolaryngologists are able to manipulate the laryngoscope and ask the patient to repeat phonation activities to strengthen their diagnosis. Although not studied with these data, we surmise that the accuracy of diagnosing SD increases with time spent examining the patient.

**Conclusions**

Spasmodic dysphonia is diagnosed primarily by recognizing specific patterns in auditory cues alone. The role of laryngoscopy in the evaluation of a patient with SD is to give the clinician the opportunity to visualize the vocal folds to rule out other pathologic causes and to rule in vocal tremor. Early recognition of SD portends a greatly improved voice-related quality of life and avoids delay in diagnosis. Educating the medical community on the typical auditory features that are present in SD and vocal tremor may lead to faster time to diagnosis and treatment.

---
Conflict of Interest Disclosures: None reported.
Previous Presentation: This study was presented as a poster at the Fall Voice Conference; October 17-19, 2013; Atlanta, Georgia.

REFERENCES