Efficacy of a Behaviorally Based Voice Therapy Protocol for Vocal Nodules

*†Eva B. Holmberg, ‡Robert E. Hillman, *Britta Hammarberg, *Maria Södersten, and †Patricia Doyle

*Department of Logopedics and Phoniatrics, Karolinska Institute, Huddinge University Hospital, Stockholm, Sweden; †The Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts; ‡The Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary; Department of Otology and Laryngology, Harvard Medical School; Communication Sciences and Disorders, Massachusetts General Hospital Institute of Health Professions, Boston, Massachusetts

Summary: The aim of this study was to assess the effects on vocal function of voice therapy for vocal nodules. Perceptual and physiological progressive changes were examined during a strictly structured, behaviorally based voice therapy protocol in which 11 women with vocal nodules participated. Randomized audio recordings from pretherapy and from each of the therapy approaches (vocal hygiene, respiration, direct facilitation, carryover) were used for perceptual evaluations. Six speech-language pathologists rated ten voice quality parameters. Two evaluation procedures were performed and compared. Interlistener reliability was sufficiently high in both tests. Significant effects of therapy were found for decreased overall dysphonia, press, instability, gratings, roughness, vocal fry, and “scrape.” Nonsignificant group effects were found for breathiness, aphasis, and lack of sonority. No significant parameter changes occurred between baseline assessment and the completion of the initial (vocal hygiene) phase of therapy. Significant changes were found following the direct facilitation and respiration phases of therapy. Videostroboscopic evaluations made by two laryngologists showed that in no case were the nodules completely resolved. However, the nodules had decreased in size and edema was reduced after therapy for all clients, but one. Combined results suggest: (1) Alterations in vocal function were reflected in perceptual parameters, and (2) the voice therapy had a positive effect on voice quality, vocal status, and vocal function for the majority of the vocal nodule clients. Key Words: Efficacy—Voice therapy—Vocal nodules—Perceptual evaluations—Voice quality.
nodules, and their incidence appears to be related to occupation. It is widely believed that a primary etiological factor to vocal nodules is trauma to the vocal fold tissues. The nodules occur mostly at the midpoint of the membranous vocal folds where impact sources are the largest. It is generally held that an underlying component in vocal nodules is vocal hyperfunction. Vocal hyperfunction is assumed to be a result of increased and poorly regulated laryngeal muscle tension. Imbalances and increases in muscle tension produce abnormally stiff vocal folds and/or incomplete closure of the membranous glottis (typically seen as a posterior membranous chink). In order to achieve voicing, subglottal pressure is increased, which often results in abnormally high vocal fold collision forces. With the formation of nodules, complete closure of the membranous glottis may be further prevented causing an increase of turbulent air through the glottis. The result is a breathy and weak voice, which may in turn cause compensatory increased subglottal pressure and heightened vocal fold collision forces, thereby trigging a “vicious circle” and adding to the vocal trauma.

Clinicians use a variety of therapy approaches to treat vocal nodules. A main goal of the treatment techniques is to reduce muscle tension and hyperfunction and optimize vocal behavior to reduce the trauma to the vocal folds, so that aerodynamic forces and acoustic properties approach normal values. Schemes have been developed for evaluation of treatment. The design of the voice therapy protocol in this study was based on a review of literature most commonly referred to by clinicians. It is generally acknowledged that behaviorally based voice therapy has a positive effect on the vocal status of individuals with hyperfunctional voice disorders. The realm of the therapy can be divided into four basic approaches: (1) vocal hygiene, (2) direct facilitation, (3) relaxation, and (4) respiration. In addition, it is generally agreed that some time should be devoted to “carryover” of newly learned vocal behaviors to “real life” situations outside the therapy setting. However, a major clinical issue concerns the efficacy of voice therapy in treating hyperfunctional voice disorders, and clinical opinions vary concerning the relative effectiveness of therapy approaches. There is a lack of documentation of the efficacy of treatment techniques.

Perceptual judgments have long been the major tool for voice evaluation. The clients seek treatment in part because their voice does not sound normal. Speech clinicians use both formal and informal perceptual judgments in each session with the client, and perceptual judgments always serve as a complement to visual inspection of the larynx and vocal folds. In addition, perceptual parameters are used as standards against which objective measures of voice are evaluated. A problem has been that large intralistener and interlistener variations have been found in perceptual evaluations, which therefore have been criticized as being too subjective to have real clinical value. In contrast, good agreement among listeners as well as among different groups of listeners have been shown in other studies.

Much of recent research has focused on objective measurements of vocal function. Unfortunately, the techniques for obtaining such signals are not always easily accessible in clinics, and analyses and interpretations of the results are not straightforward. Identifying voice quality parameters that reflect underlying vocal function should have strong clinical value. An important goal is to develop reliable methods for perceptual evaluations of voice quality, and to establish reliable relationships between perceptual judgments and objective measures of underlying physiological conditions.

There is a paucity of data on relationships between changes in vocal behavior and the manifestations of such changes in voice quality parameters. The present perceptual and endoscopic study is part of a larger study of objective assessments of vocal hyperfunction, examining the efficacy of behaviorally based voice therapy for women with vocal nodules. The main goals of the study were to (1) document actual effects on vocal fold status and voice quality across the different approaches of the voice therapy, (2) provide critical information concerning the relative influence of each of these therapy approaches on voice quality, and (3) begin to define relationships between therapy approaches and desired alterations in vocal function with manifestations in voice quality parameters, and thus
METHODS

Procedures for the voice therapy and assessments

Subjects
Eleven American women with bilateral vocal nodules served as subjects. They were recruited from the client population at the Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary (MEEI) in Boston, Massachusetts. They all participated in a behaviorally based voice therapy protocol as part of a project with the aim to evaluate the effects of voice treatment. Shortly before initiation of the voice therapy a diagnosis of bilateral vocal nodules had been confirmed for each subject by an otolaryngologist. Nodule size and time since voice problem onset varied among the clients, but for all of them the nodules were considered chronic. The clients’ ages ranged from 19 to 35 years with an average of 23.3 years. They were native speakers of American English, nonsmokers, and had not received professional singing or speaking training. None of them had received previous voice therapy or undergone laryngeal treatment. All clients passed a hearing screening test at the initial assessment and were free from respiratory infections at all assessments. All clients worked outside their homes and continued to work during the voice treatment period.

Therapy protocol
The voice therapy protocol consisted of five basic behaviorally based approaches or phases: vocal hygiene, direct facilitation, respiration, relaxation, and carryover. The approaches are described below. Each therapy approach consisted of three sessions. In general, clients received one therapy session per week. The total therapy period per client varied between four and six months. (The clients sometimes postponed sessions in case of illness or for other reasons. Thus, the treatment period varied.) The therapy began with the vocal hygiene approach and ended with carryover for all clients. For the other approaches (relaxation, respiration, and direct facilitation) the scheduled order varied. (The clients had been divided into three groups in terms of order of the therapy approaches—respiration, direct facilitation, and relaxation—in the interest of studying potential order effects. However, in this perceptual study, the number of clients varied in the order groups and therapy order was not taken into account.) The voice therapy was provided at the Voice and Speech Clinic at MEEI by one ASHA certified speech and language pathologist. The design of the treatment protocol was based on current clinical practices in voice therapy. Standard protocols were used to increase the degree to which treatment could be closely replicated across the clients. As each new phase of therapy was started, work continued on the approaches already introduced. However, effort was made to not anticipate any aspects of therapy from approaches later on in the therapy sequence, prior to their formal introduction. At the end of each therapy session the client was given a short multiple-choice quiz to assess her understanding of the information covered during the session. The clients practiced techniques, worked on in the therapy session, at home in two 15-minute sessions per day. To improve compliance, the clients recorded their home practice sessions on cassette recorders, and the tapes were spot-checked by the clinician.

Voice therapy approaches
Vocal hygiene. Vocal hygiene included: (1) client education about how normal voice is produced; (2) identification of voice-use related (e.g., yelling) and non-voice-use related (e.g., smoky environment) abuses; (3) education about vocal nodules in terms of suspected etiology and effects on voice production; and (4) reduction/elimination of vocal abuse via the “Vocal Abuse Reduction Program,” VARP.43 The clients viewed a custom-made videotape that covered standard information about how voice is produced and described sources of vocal abuse. The speech pathologist presented information about vocal nodules. Together the clinician and client identified sources of vocal abuse that were specific to the client’s daily life and discussed how to limit or eliminate these sources.

Respiration. Work on respiration did not focus on developing a specific “type” of breathing (e.g., diaphragmatic) per se.44 Instead, respiration treatment was directed toward two related areas: (1) facilitation (reinforcement) of speech breathing that was free of extraneous effort, and (2) identification...
and amelioration of behaviors that reflect poor management of the air supply for voice/speech production. Facilitation of respiration for “effort-free” voice/speech involved a set of exercises that emphasized “easy” and relaxed breathing, starting with rest breathing and progressing through a hierarchy of voice and speech tasks up to the level of conversational speech. The emphasis was on good management of the air supply; timing between respiration and initiation of phonation, inhalation of a normal amount of air prior to initiation of phonation, and inhalation at appropriate points of juncture.\textsuperscript{41}

**Direct facilitation.** Two basic facilitation approaches were used: (1) reduction of loudness, and (2) yawn-sigh exercises.\textsuperscript{20,21}

**Reduction of loudness.** It is generally held that chronic use of an excessively loud voice can contribute to the formation and persistence of vocal nodules.\textsuperscript{1,2} Reduction of loudness involved auditory discrimination using tape-recorded samples of the client’s soft, normal, and loud voice. The intensity display of the Visipitch (Kay Elemetrics, Inc., Lincoln Park, NJ) was then used to help establish target (reduced) loudness levels to provide feedback as the client progressed through a hierarchy of speech tasks while maintaining target levels. Steps in the hierarchy included production of vowels in isolated phonation, syllables, multisyllabic words, phrases, sentences, reading, monologue, and dialogue.

**Yawn-sigh.** While potential pathologic mechanisms have been found to be primarily related to increased sound pressure level (SPL) in a group of nodule clients, evidence was found that similar vocal mechanisms may exist also in clients who do not display abnormal and/or proportionally increased SPL.\textsuperscript{10} For such clients, reducing loudness may not be sufficient to ameliorate pathologic mechanisms. The “yawn-sigh” approach is a frequently cited and frequently used facilitation approach for reducing muscular tension, decreasing effort, and reducing abruptness of vocal onset, thereby helping to reduce vocal hyperfunction.\textsuperscript{45} First the clinician explained the physiology of the yawn with an emphasis on widening the airway and relaxing the associated musculature. Then each step of the approach was demonstrated by the clinician and performed by the client.

**Relaxation.** A combination of progressive (differential) relaxation\textsuperscript{46,47} and imagery\textsuperscript{48,49} techniques were used in a program of generalized relaxation. In addition, each client also received basic instruction in stress management.\textsuperscript{50–52}

**Carryover.** This approach involved attempts to transfer “newly learned” vocal behaviors to “real life” situations outside the therapy setting. The process started in the therapy room with the client mimicking pertinent speaking situations (e.g., talking on the telephone), continued in next-door localities outside the therapy room, and ended in locations outside the clinic (e.g., a restaurant). The clinician gave feedback to the client concerning her voice production.

**Assessments**

Each client’s pretherapy baseline performance was established prior to initiation of the voice therapy, after which the voice therapy was started within a week. Subsequent assessments were made directly after the last session of each therapy approach. Each assessment included recordings of a set of acoustic, aerodynamic, and physiological signals that will be reported on elsewhere, and ended with laryngoscopic examinations, including videostroboscopic recordings.

**Acoustic recordings**

The recordings were made in a soundproof booth. Among several tasks the protocol included free field acoustic recordings of the client reading a standard text (the Rainbow Passage)\textsuperscript{53} at a comfortable loudness and pitch. The speech signal was picked up by a small Sony electret microphone (Sony Corporation, Allandale, NJ), which was mounted on a headset to ensure a constant, reproducible distance (15 cm) from the mouth. The signal was audiotaped with a digital audiotape (DAT) recorder that was placed outside the recording booth. These recorded readings served as the listening material in the perceptual evaluations and for acoustic measurements of mean SPL and fundamental frequency ($F_0$).

**Laryngeal examinations and videostroboscopic recordings**

All clients underwent laryngeal examinations using videoendoscopy with stroboscopy at the end of
each assessment. Both flexible trans-nasal (Olympus, ENF-P3, Olympus Corporation, Melville, NY) and rigid trans-oral (Kay Elemetrics, 9105) were performed. When necessary, nose spray was used for surface anesthesia (3% lidocaine hydrochloride and 0.25% phenylephrine hydrochloride). Video recordings were made [CCD camera model GP-KS152, Panasonic Corporation, Secaucus, NJ, a Mitsubishi super VHS tape deck, model BV2000, Woburn, MA, and Kay Elemetrics Rhino-Laryngeal Stroboscope 1900].

Evaluations and ratings

Ratings of vocal fold status and voice quality took place at the Department of Logopedics and Phoniatrics, Huddinge University Hospital, Stockholm Sweden.

Endoscopic rating of vocal fold status before and after voice therapy

Two Swedish laryngologists evaluated in consensus recordings of ten of the clients. (For one subject the videoendoscopic recorded material was not complete and could not be used for evaluation). They were informed that the recordings were of American females with vocal nodules recorded twice over some time. Material from the videoendoscopic recordings was used for evaluation of vocal fold status before and after therapy. (The recordings with the highest quality were chosen for the evaluation; for 7 clients they had been made via a rigid laryngoscope, while for 3 clients flexible fiberoptic had been used). These recordings had been made for /i/ phonations at a comfortable loudness and pitch. The video recordings were presented without sound; thus the evaluations were done from visual impression only. All clients were evaluated in one session. The subjects were presented in random order, and the order of each client’s recordings were also randomized. Four subjects’ recordings were shown twice for reliability. Focus was on each client’s pretherapy and posttherapy changes, and the ratings were made of one client at a time. The questionnaire asked about differences, or no differences, between the client’s pretherapy and posttherapy recordings. The responses were given on paper forms, one for each client.

Ratings were made of: (1) nodule location (very front, front, mid, back of the vocal fold membranous portion); (2) nodule shape (narrow-based, broad-based); (3) nodule size (difference between the two recordings, larger/smaller/no difference); (4) surrounding edema (yes/no and difference between the two recordings: larger/smaller/no difference).

Auditory perceptual rating of voice quality

Although there were relatively large differences in voice quality among the clients, each client’s changes in voice quality across the assessments were subtle. Pilot work showed that the rating of these subtle differences was a difficult task. Thus, the experimental design for the rating of the perceptual changes became an important part of the study, and two perceptual tests (henceforth called Test 1 and Test 2) with different experimental designs and procedures were conducted and compared. The design of Test 2 was in part based on methodological experiences (not results) for Test 1.

Listeners. Six experienced Swedish female speech and language pathologists from three hospitals in the greater Stockholm area served as the listener panel. The listeners had an average time of 19 years, ranging from 4 to 34 years in the professional field. All listeners had extensive experience in perceptual evaluations and in using the Stockholm Voice Evaluation Approach. 

Listening material. The standard reading passage (the Rainbow Passage) read at a “comfortable” loudness and pitch was used as the listening material. Recordings of the whole reading paragraph (98 words; about 45 seconds of text) from one (the last of three) baseline assessment and from the therapy assessments for vocal hygiene, direct facilitation, respiration, and carryover were used for perceptual evaluations. (In the interest of keeping the listening material to a manageable size, recordings from the relaxation approach were omitted for the perceptual evaluations. For three clients there were no carryover assessments. For those clients direct facilitation was the last approach. For another client reading material was missing for direct facilitation.) The acoustic recordings of the material were digitized into a computer and edited into a file, using the Soundswell signal workstation program.
Voice quality parameters and their definitions. The voice quality parameters were selected after pilot listening to the whole material in a stepwise procedure described in the Stockholm Voice Evaluation Approach.\textsuperscript{42,54} Ratings were made of the following parameters:

- **Press/hyperfunction.** Strained, effortful voice, as produced with tension and compression of the vocal folds, larynx, and pharynx.
- **Breathiness.** Audible noise, as from air leakage through the glottis due to insufficient glottal closure.
- **Aphonic instances.** Intermittent and short devoiced portions.
- **Lack of sonority.** “Closed in” voice, voice without carrying sonority.
- **Instability.** Shifts in pitch and/or intensity and/or register, or diplophonia.
- **Gratings.** High-frequency noise during phonation as from irregular vocal fold vibrations.
- **Roughness.** Low-frequency noise during phonation as from irregular vocal fold vibrations.
- **Vocal fry/creak.** Low-frequency aperiodic/periodic vocal fold vibrations.
- **Scrape.** Summarizing term including gratings, roughness, and other perceptual impressions resulting from irregular vocal fold vibrations, as well as vocal fry/creak.
- **Overall dysphonia.** Severity of dysphonia from an esthetic as well as functional point of view in comparison with a healthy voice.

**Scales.** Ratings were made by markings on 100-mm visual analog scales (VAS) on paper forms, one form for each subject and assessment. The extremes of the scales were marked “nothing/never” and “very much/frequently.”

**Evaluation procedures**

Prior to the tests, the listeners were given written and spoken instructions with explanations of the procedures. They were informed that the subjects were American women with vocal nodules who had been recorded several times. No other information was given about the background or purpose of the study. The perceptual parameters were described and defined in consensus in the interest of minimizing listener variation due to different definitions of the terms. In addition, a standard for a healthy voice was discussed in terms of occurrence and degree of these parameters.\textsuperscript{32,56,57} Focus was on each client’s changes in voice quality from the pretherapy assessment and across the therapy approaches, and the perceptual evaluations were made of one client at a time. Since the clients’ changes in voice quality from one assessment to another were subtle, a reference, or anchor was required for the ratings.\textsuperscript{32,38}

**Test 1.** The evaluation sessions took place in an ordinary room, and the listening material was presented from a computer over loudspeakers. A total of 51 recordings were evaluated. Ratings were made of the parameters: press/hyperfunction, breathiness, aphonic instances, lack of sonority, gratings, roughness, and vocal fry/creak.\textsuperscript{6} (Perceived pitch was also included as a parameter. However, frequently pitch could not be given a single value because of unstable voice quality, which resulted in many blank responses. Thus, the pitch data were omitted.) A demonstration tape of the test design was played. Each subject’s baseline recording was used as a reference, to which her randomized therapy recordings were compared. (Recall that the listeners were unaware that the recordings were associated with different phases of voice therapy as well as of the nature of the reference recording).

The material was presented in the following way: For each client the baseline recording was evaluated first. The listeners were instructed to use a healthy voice as their inner standard for this first evaluation. Thereafter the client’s therapy recordings were presented and evaluated in randomized order, one at a time with breaks in between, during which the ratings were done. Prior to the presentation of each therapy recording, a short reminder of the reference (two sentences of the baseline recording) was played. The listeners rated the therapy recordings in comparison with the reference. If wanted, both the reference and therapy recordings were played back several times. The listeners made individual ratings and were not allowed to take part in each other’s ratings or discuss the voices during the session. A separate rating form was used for each recording. At any time the listeners were allowed to look at their own previous evaluation of the reference recording. When all of a client’s recordings had been rated, these recordings were presented again in one se-
sequence without breaks. The listeners reviewed their own ratings and were allowed to make adjustments if they wished, until they all felt satisfied with their ratings. Average evaluation time per client was 10 to 12 minutes. Because of the relatively large listening material, the evaluations were made in two sessions (with the same listeners) on separate days. Each session lasted for about 80 minutes.

**Test 2.** In Test 2 the subjects were the same clients as in Test 1. Six female speech-voice pathologists served as listeners. Five of them had also participated in Test 1. Ratings were made of the parameters: press/hyperfunction, breathiness, scrape, and overall dysphonia for each client’s baseline and last therapy recordings. One sentence (the first in the reading passage) from three repeated readings per assessment was used as the listening material, making six stimuli per client (2 recordings × 3 readings). (At each assessment, the clients had read the passage three times, with other speech tasks between readings.) An interactive computer program, VISOR, which allows custom-made design, was used for the ratings. Individual ratings were made of one client at a time in free order chosen by the rater. Rating and listening were done interchangeably, which allowed comparisons among all stimuli as well as parameters. Adjustments of ratings could be done as many times as wished. Average evaluation time per client was 5 minutes. Most of the listeners evaluated all clients in one session; two listeners divided the material up into two sessions.

**Statistic analyses**

Means and standard deviations were calculated across the six listeners for each client and assessment (individual data). Thereafter, means and standard deviations were calculated across all clients for each parameter and assessment (group data). Cronbach’s coefficient alpha was calculated to examine interlistener reliability. Spearman correlations (p < 0.05) were performed to examine pairwise relationships between the perceptual parameters within each assessment. Analysis of variance (p < 0.05) (with therapy approach as the independent variable and voice quality parameter as the dependent variable) was performed for each perceptual parameter to examine the effect of therapy. Post-hoc tests (Scheffé - p < 0.05) were performed for each parameter to examine differences between the assessments (Test 1).

**RESULTS**

**Vocal fold status before and after voice therapy**

There were no differences between evaluations of the 4 clients’ recordings that were repeated for reliability. Before therapy, all clients (n=10) were judged to display bilateral and broad-based nodules located at the mid portion of the membranous vocal folds. In no case were the nodules judged to have disappeared after the therapy. However, after therapy the nodules had decreased in size for 9 clients, either bilaterally (6 clients) or unilaterally (3 clients). One client displayed a decrease in nodule size on one side but an increase on the other side. Surrounding edema was found before therapy in 5 clients, either unilaterally (3 clients) or bilaterally (2 clients). After therapy all but one of these clients showed a reduction of edema. One client displayed a posttherapy unilateral increase of edema. This was the same client who displayed a unilateral posttherapy increase in nodule size.

**Changes in SPL and F₀ across therapy**

Group mean values of SPL did not change significantly between assessments or across therapy. Group mean value of F₀ was relatively low before therapy and increased significantly (p < 0.05) across therapy. Mean values and standard deviations of SPL and F₀ are presented in Table 1.
Breathiness had decreased for 3 clients, increased for 7 clients, and was unchanged for 1 client. Scrape decreased for all clients. Overall dysphonia decreased to various extent for all clients but one.

**Group mean values across therapy**

Figure 2 presents group mean values with 95% confidence intervals across therapy for each parameter and assessment. Tables 3 and 4 present analysis of variance (ANOVA) results for Test 1 and Test 2. Significant ($p<0.05$) effects of therapy were found for decreased values of press, instability, gratings, roughness, vocal fry, scrape, and overall dysphonia. Nonsignificant effects of therapy were found for breathiness, aphonic instances, and lack of sonority.

---

**TABLE 1. Group Mean Values and Standard Deviations of SPL and $F_0$ at Baseline and Across the Therapy Assessments Vocal Hygiene, Respiration, Direct Facilitation, and Carryover.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Baseline Mean (SD)</th>
<th>Vocal Hygiene Mean (SD)</th>
<th>Respiration Mean (SD)</th>
<th>Direct Facilitation Mean (SD)</th>
<th>Carryover Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPL (dB)</td>
<td>69 (4.5)</td>
<td>70 (4.6)</td>
<td>71 (4.4)</td>
<td>71 (3.5)</td>
<td>70 (2.3)</td>
</tr>
<tr>
<td>$F_0$ (Hz)</td>
<td>185 (13.6)</td>
<td>189 (14.3)</td>
<td>202 (15.2)</td>
<td>194 (18.6)</td>
<td>202 (11.6)</td>
</tr>
</tbody>
</table>

*Abbreviation: SD, standard deviation.*

**TABLE 2. Cronbach’s $\alpha$ Values Calculated for (Client) Group Mean Values of Each Parameter in Each Assessment (Baseline, Vocal Hygiene, Respiration, Direct Facilitation, Carryover) in Tests 1 and 2 (Pretherapy, Posttherapy).**

**TEST 1**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Base</th>
<th>Vocal Hygiene</th>
<th>Respiration</th>
<th>Direct Facilitation</th>
<th>Carryover</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Press</td>
<td>0.71</td>
<td>0.87</td>
<td>0.91</td>
<td>0.76</td>
<td>0.87</td>
<td>0.83</td>
</tr>
<tr>
<td>Breathiness</td>
<td>0.95</td>
<td>0.91</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Aphonic instances</td>
<td>0.76*</td>
<td>0.83*</td>
<td>0.95*</td>
<td>0.32*</td>
<td>0.49*</td>
<td>0.67</td>
</tr>
<tr>
<td>Lack of sonority</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
<td>0.90</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Instability</td>
<td>0.96</td>
<td>0.88*</td>
<td>0.87*</td>
<td>0.69*</td>
<td>0.65*</td>
<td>0.81</td>
</tr>
<tr>
<td>Gratings</td>
<td>0.86*</td>
<td>0.75*</td>
<td>0.67*</td>
<td>0.86*</td>
<td>0.75*</td>
<td>0.78</td>
</tr>
<tr>
<td>Roughness</td>
<td>0.81*</td>
<td>0.58*</td>
<td>0.66*</td>
<td>0.52*</td>
<td>−0.07*</td>
<td>0.50</td>
</tr>
<tr>
<td>Vocal fry</td>
<td>0.78</td>
<td>0.52</td>
<td>0.73*</td>
<td>0.84*</td>
<td>0.82*</td>
<td>0.74</td>
</tr>
<tr>
<td>Mean</td>
<td>0.84</td>
<td>0.79</td>
<td>0.84</td>
<td>0.73</td>
<td>0.67</td>
<td>0.78</td>
</tr>
</tbody>
</table>

**TEST 2**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Before</th>
<th>After</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Press</td>
<td>0.85</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td>Breathiness</td>
<td>0.91</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>Scrape</td>
<td>0.94</td>
<td>0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>Overall dysphonia</td>
<td>0.91</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Mean</td>
<td>0.90</td>
<td>0.85</td>
<td>0.88</td>
</tr>
</tbody>
</table>

An asterisk indicates skewed distribution due to reduction of range.
FIGURE 1. Pretherapy (unfilled circles) and posttherapy (filled circles) individual data for the 11 clients. The graphs show results for the parameters: press, breathiness, scrape, and overall dysphonia, respectively.
FIGURE 2. Changes in group mean VAS values with 95% confidence intervals across the five therapy assessments: baseline (B), vocal hygiene (VH), respiration (RS), direct facilitation (DF), and carryover (CO). The graphs show changes for the parameters: press, breathiness, instability, gratings, aphonics instances, lack of sonority, roughness, and vocal fry. Significant p values from ANOVA are shown for each parameter. Significant differences (Sheffé) between baseline and therapy assessment are indicated with an asterisk (*) after the therapy assessments.
Table 5 presents significant ($p < 0.05$) results from post-hoc tests (Scheffé) that were performed to examine between which assessments significant parameter changes occurred in Test 1. No significant (or close to significant) differences were found between pretherapy baseline and vocal hygiene, but significant differences were found between baseline and the therapy assessments direct facilitation, respiration and/or carryover.

Relationships between parameters

Spearman correlation coefficients ($p < 0.05$) were calculated for all parameter pairs within the assessments. Strong ($r > 0.70$) positive correlations were found in several assessments (Test 1) between press and gratings, roughness, and scrape; and between breathiness and aphonic instances, lack of sonority, instability. Vocal fry was not strongly correlated with any of the other parameters, but a few cases of significant negative correlations were found between vocal fry and breathiness and aphonics instances. Overall dysphonia (Test 2) before and after therapy was strongly correlated with press (before therapy $r = 0.96$, after therapy $r = 0.85$), breathiness (before therapy $r = 0.75$, after therapy $r = 0.92$), and scrape (before therapy $r = 0.74$). No cases of significant correlations were found between press and breathiness.

Comparisons between Tests 1 and 2

Five listeners had participated in both Tests 1 and 2. Qualitative comparisons between Tests 1 and 2 were made for each of these listener’s judgments of press and breathiness in terms of direction (increase or decrease) of change. A difference of 10 mm or more on the VAS was considered a change.
The agreement between the tests was higher for press (75%) than for breathiness (58%). The listeners had reported a higher number of pre-post therapy differences in Test 2 than in Test 1.

DISCUSSION

Vocal Status Before and After Therapy

Endoscopic ratings of vocal status in this study showed that none of the vocal nodules had disappeared after completion of the therapy. This finding is in agreement with histologic studies that suggest that more permanent types of tissue damage may be associated with the formation of nodules, e.g., fibroblastic response involving increased fibronectin decomposition.\(^{59}\) Thus, the positive impact of therapy is not necessarily associated with a complete amelioration of the nodular lesions. However, the fact that the nodules had decreased in size after therapy for all clients but one, as well as findings of reduced edema, strongly suggest that the voice therapy had a trauma-reducing effect for the majority of the clients. The hyperfunctional vocal behavior appears to have decreased. Thus, there seemed to be a good potential that the compensatory “vicious circle” of further increased muscle tension and increased subglottal pressure leading to escalating trauma could be stopped by voice therapy. Positive effects of the therapy were also experienced by the clients. Client subjective reports are not included in this paper, but only one client—the client for whom the nodules were judged to not have decreased in size—chose to have the nodules surgically removed after therapy. Long-time follow-up assessments would be of interest to show whether or not the improved vocal behavior and vocal status lasted.

Methods for the perceptual evaluations

Finding perceptual correlates to observed physiological changes across therapy is an important research goal for clinical work. One of the problems with perceptual ratings of pathological voices has been large interlister variation. Such variation may in part be due to variation in standards for the healthy voice, which is used as an inner reference for rating pathological voices.\(^{32}\) Since clinicians must continue to rely on their perceptual judgments in working with voice patients, it seems important to work on developing more reliable procedures for perceptual evaluations rather than dismissing such evaluations as simply not useful.

\[\text{TABLE 5. Significant (p<0.05) Results from Post Hoc Scheffé Analyses, Examining Parameter Differences Between the Assessments}\]

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>ASSESSMENT</th>
<th>Mean difficulty</th>
<th>Critical difficulty</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Press</td>
<td>Base - Direct Facilitation</td>
<td>24.16</td>
<td>19.53</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>Base - Respiration</td>
<td>23.60</td>
<td>19.53</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>Base - Direct Carryover</td>
<td>26.44</td>
<td>19.53</td>
<td>0.004</td>
</tr>
<tr>
<td>Instability</td>
<td>Base - Carryover</td>
<td>14.96</td>
<td>15.19</td>
<td>0.054</td>
</tr>
<tr>
<td>Gratings</td>
<td>Base - Direct Facilitation</td>
<td>15.23</td>
<td>14.54</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Base - Respiration</td>
<td>14.29</td>
<td>14.54</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>Base - Carryover</td>
<td>14.76</td>
<td>14.54</td>
<td>0.045</td>
</tr>
<tr>
<td>Roughness</td>
<td>Base - Direct Facilitation</td>
<td>6.93</td>
<td>7.06</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>Base - Carryover</td>
<td>8.87</td>
<td>7.06</td>
<td>0.008</td>
</tr>
<tr>
<td>Vocal fry</td>
<td>Base - Respiration</td>
<td>18.51</td>
<td>15.45</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Base - Carryover</td>
<td>19.04</td>
<td>15.45</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>Vocal Hygiene - Respiration</td>
<td>16.09</td>
<td>15.45</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>Vocal Hygiene - Carryover</td>
<td>16.61</td>
<td>15.45</td>
<td>0.030</td>
</tr>
</tbody>
</table>
The experimental designs used for the perceptual evaluations were an important part of this study. Efforts were made to eliminate irrelevant interlistener variation as much as possible. An inner reference of a healthy voice was used for rating the clients’ baseline recordings. Prior to the evaluations, the experimenter and listeners defined the perceptual parameters in consensus and discussed their potential occurrence and degree in healthy voices. However, in this discussion it became clear that the listeners agreed about the standards for healthy voice quality, and there seemed to be no problem with the use of a healthy voice reference for rating the baseline recordings. In addition, the fact that the ratings were always comparative between the baseline and therapy recordings should have served to decrease the sensitivity for where along the VAS the changes occurred.

An extensive parameter inventory, as well as description of features that were typically found in the material, was made prior to the listening tests, from which the final parameter set was chosen in a stepwise procedure. The parameter set should well characterize the typical voice quality of the client group as well as reflect changes in underlying vocal behavior that appear relevant to the goals of the therapy approaches. At the same time a goal was to keep the parameter number to a manageable size. In Test 2, the set was reduced to include only four parameters. Therefore, two of the parameters (overall dysphonia and scrape) were intentionally expanded to each include several perceptual features.

Comparisons between results from Tests 1 and 2 showed that the listeners had detected pretherapy and posttherapy parameter changes more frequently in Test 2 than in Test 1. This observation was somewhat surprising, since it could be more difficult to detect a quality change when listening to short listening material (one sentence) than to long material (the whole paragraph). The finding may reflect the fact that the interactive computer research design in Test 2, which allowed quick switching between stimuli for comparisons, facilitated detection of quality changes.

Feedback from the listeners on the test procedures

After completion of both tests informal feedback from the listeners on the test designs was encouraged. In both tests, the listeners felt that rating one client at a time mimicked a clinical situation. They appreciated the opportunity to adjust ratings until they felt satisfied. In Test 1, the repeated reference voice greatly facilitated the ratings. The five listeners who participated in both Tests 1 and 2 thought it was easier to concentrate, focus on one parameter at a time, and detect details with the use of the computer setup (Test 2) than with traditional paper-form methods (Test 1). The ability to quickly switch between stimuli, choose the order of stimuli for comparisons, and check and adjust ratings facilitated the evaluation. Furthermore, the fact that the listeners could independently determine the scheduling and length of the sessions may also have contributed to concentration and focus, in addition to being of practical importance for busy professionals.

Loudness and pitch

Excessive loudness is a common feature in vocal nodule clients. The finding that SPL was not decreased during the therapy shows that the exercises to reduce loudness were not quite successful. The clients decreased loudness during the actual sessions, but evidently the use of the lower loudness did not carry over to the assessment sessions outside the therapy room. It is possible that, at least in part, this was due to the experimental situation and to the fact that the recordings took place in a sound-proof booth. However, it is noteworthy that perceived press decreased in spite of the unchanged SPL, a result that may indicate a more relaxed flow phonation and thereby a physiologically more efficient voice.

SPL and $F_0$ did not co-vary in a normal fashion. Pitch was not directly manipulated in the therapy. Mean $F_0$ before therapy was in the low end of the normal range and increased during therapy, which to some extent may result from the fact that the nodules decreased and edema was reduced, as well as from the decreased vocal fry.

Perceptual results

Parameter occurrence at baseline

Group mean values for all parameters fell below 50 mm on the VAS. This finding indicates that the voice quality of this client group was moderately...
dysphonic. Before therapy, all clients were perceived as having a pressed voice quality. A majority of the clients were heard as also being breathy, lacking sonority, and displaying instability and vocal fry. Gratings and roughness varied in terms of occurrence and degree. Aphonics were heard in only a few clients and to a relatively low degree. Thus, in terms of parameter occurrence, the typical pretherapy quality for this group was a highly pressed, breathy, nonsonorous and unstable voice with vocal fry.

Based on previous studies, a pressed voice quality in this study was assumed to reflect the main underlying problem for this client group, and relatively high values of pressed quality were also found in all clients before therapy. Decreased pressed quality was considered to be an important indicator of successful voice therapy. Individual data showed that press decreased across therapy for most of the clients (Figure 1), and for the group the decrease in press was statistically significant. Values of press stayed low over the carryover assessment (Figure 2), which suggests that the newly learned behavior, i.e., handling vocal tension, was transferred to situations outside the clinic. Group mean value of press decreased from baseline to the vocal hygiene approach, but the change was not statistically significant. Individual data showed that press actually had increased for some clients from baseline to vocal hygiene. It appears that the vocal hygiene approach reduced press for some clients, but for others and for a significant group result, further therapy was needed.

Group values for press were lowest after the direct facilitation and respiration assessments in spite of a somewhat increased SPL (2 dB). The results of decreased pressed quality may indicate positive effects of the yawn-sigh exercises, which aimed at decreased muscle tension and facilitated “easier” voicing. Decreased pressed quality may also result from improved speech respiration behavior with better management of air supply and a more efficient relationship between subglottal pressure and glottal function.

A majority of the voices were perceived as (hyperfunctionally) breathy, which is to be expected for nodule clients. With vocal fold lesions such as nodules, complete closure of the membranous glottis can be prevented causing an increase of airflow through the glottis leading to a breathy voice quality. Decreased breathiness in therapy may reflect a situation in which the nodules have decreased in size, thus making possible more complete glottal closures. However, decreased breathiness alone may reflect increased muscle tension and increased hyperfunction in trying to overcome the breathiness. Therefore, breathiness alone cannot be a reliable indicator of efficacy of the voice therapy, but must be considered in combination with press. For this client group, increased breathiness in combination with decreased press may, in fact, reflect decreased hyperfunction in trying to overcome the breathiness.

In this study, the effect of therapy on breathiness was not significant for the group (Figure 2). Individual data showed that breathiness increased across therapy for the majority of the clients, but decreased or did not change for a few (Figure 1). With decreased press, both increased and decreased breathiness may indicate improved voice function. Data for client 1, from the bottom of Figure 1, show a posttherapy decrease in press in combination with a considerable increase in breathiness, while for client 4, from the bottom of the figure, decreased press was combined with unchanged breathiness. For both clients, scrape and overall dysphonia decreased, which suggests an improved voice function.

Test 2 showed significant effects of replication for both press and breathiness (Table 4). Data for the separate readings showed that press was significantly lower and breathiness significantly higher in the third reading than in the two first readings. These perceptual effects did not appear to depend on differences in SPL or F0, which did not vary across readings in a consistent way among the clients. The reason for the perceptual difference between the first two readings and the third is not clear. It may possibly reflect a fatigue effect due to the length of the recording protocol. Even though the total recording time of all tasks generally did not exceed 30 to 40 minutes per client, the clients may have fatigued somewhat during that time, which may have caused a change in voice production. These results suggest that voice production...
and voice quality may vary over a relatively short time for nodule clients. Thus, a strict recording scheme as well as listening order is important for both intersubject and intrasubject comparisons of voice quality.

All clients were rated with moderate to high levels of lack of sonority before therapy. This parameter was included in the set because it was observed frequently in this client group. However, the results that lack of sonority changed across therapy in an almost identical pattern as press (Figure 2), and was highly correlated with breathiness in all therapy assessments, suggest that this parameter was redundant. Since it did not appear to add any information on its own, it could be excluded in the interest of downsizing the evaluation procedures.

Aperiodic turbulent noise caused by irregular vibration pattern is commonly heard in the voice quality of individuals with vocal nodules. Cycle-to-cycle variation of the vibration pattern can result from increased muscle tension per se but can also be an effect of the vocal nodules. In Test 1 high-pitched noise (gratings) was separated from low-pitched noise (roughness). Before therapy, gratings were somewhat more frequently occurring and rated with higher values than roughness. Both gratings and roughness decreased significantly in the therapy (Figure 2). Observations of the individual listeners’ ratings of each client showed that gratings and roughness were often used interchangeably. It appeared as if the listeners agreed in that they heard a “scrape-noise,” but disagreed in what to call it. For evaluation of voice quality of nodule clients, it may be of less importance whether the noise is high or low pitched, since it is not possible to separate the effects of underlying vocal function from the effects of the nodules. Therefore these two parameters were merged into one parameter called “scrape” in Test 2. Scrape also included all other features that could be due to irregular vocal fold vibrations, as well as vocal fry. As seen in Figure 1, scrape decreased considerably after therapy for all clients but one. Scrape decreased significantly for the group from baseline, which agrees with the results for gratings, roughness, and vocal fry in Test 1.

Moderate to high levels of vocal fry occurred in most clients before therapy. Vocal fry decreased significantly during therapy (Figure 2). However, results of no significant correlations between gratings/roughness on the one hand and vocal fry on the other, in addition to always negative correlations between vocal fry and other parameters show that vocal fry should not be merged together with gratings and roughness or with any other parameter. The results strongly suggest that the listeners separated between the different glottal closure patterns underlying gratings/roughness and vocal fry. These results agree with results from a perceptual study by Hammarberg et al., in which breathy and vocal fry fell out as bipolar contrasts in factor analysis, whereas gratings/roughness were components of another factor.

Relatively high or moderate levels of instability were found before therapy in a majority of the clients and instability decreased across therapy. Instability should reflect unstable vocal vibration patterns and cycle-to-cycle variation. Significant correlations were found between instability and lack of sonority in all assessments, as well as between instability and gratings, roughness, press, breathiness, and aphonic instances in several assessments. These multiple high correlations reflect the intuitive impression that several components of voice production and associated perceptual features contribute to the perceptual impression of instability. The results of decreased instability (i.e., a more stable voice production) in combination with decreased values of press, gratings, roughness, and aphonic instances suggest that the clients gained better control over their voice production during the therapy in spite of the fact that the nodules were still present.

Aphonic instances occurred with low values in a few clients and assessments, and there was no significant effect of therapy for the group. However, occurrence of aphonic instances is a strong indicator of vocal pathology, and reflects high stiffness of the vocal folds, which demands high subglottal pressure for phonation. For the clients with aphonic instances before therapy, the aphonic instances approached zero or disappeared after therapy, which suggests improved vocal behavior. Thus, the selection of voice quality parameters in perceptual evaluations should be based on severity as well as frequency of occurrence.

Overall dysphonia (Test 2) was included in the parameter set to examine which of the other parameters contributed the most to the overall impression of the voice quality. The listeners reported that their judgments were made from knowledge about vocal function more than from an esthetic point of view. The parameter was deliberately not specified in detail for the listeners, and could include any or all of the other parameters, as well as not listed voice quality features. This is reflected in the correlation results, which showed that overall dysphonia was most strongly correlated to the highest rated parameter, that is, to press before therapy and to breathiness after therapy. Overall dysphonia decreased significantly after therapy.

Education and instruction in vocal hygiene has been suggested to perhaps be sufficient for learning better vocal behavior for clients with hyperfunctional voice disorders. However, in the present study, there were no significant parameter changes between baseline and vocal hygiene for the group, and individual data showed cases of increased values for some parameters (e.g., press) at the end of vocal hygiene. The results suggest that the whole therapy protocol was needed for significant changes and improvement of voice quality for the group.

Studies of respiration have supported the view that the respiratory system is involved in hyperfunctionally related voice disorders. There is evidence that some individuals with vocal nodules tend to initiate and/or terminate speech at lower than normal volumes. In addition, reports of higher than normal average airflow rates and transglottal air pressures for nodule patients during voice production lend indirect support to the notion of respiratory involvement in the form of increased respiratory effort. Therefore, a portion of the respiration phase of therapy in this study was directed toward ensuring that clients inhaled to normal lung volume levels before initiation of phonation to improve the management of air supply.

The perceptual results of generally high levels of press, breathiness, and instability before therapy suggest that some of the previously reported observations in respiratory and aerodynamic studies for individuals with nodules were reflected in the perceptual judgments of voice quality. The results showed that significant changes in perceptual parameters often occurred after completion of the direct facilitation and/or respiration therapy approaches. Studies of the relationships between these perceptual results and respiratory, aerodynamic, and acoustic spectral data will be the goal for future studies.

**CONCLUSIONS**

The high agreement among the listeners, as well as the significant parameter changes across therapy in this study, suggests that perceptual evaluations of voice quality can be made reliable and should have clinical value. The combined results from endoscopic and perceptual evaluations suggest that alterations in vocal status and voice function were reflected in parameters of voice quality. The results of decreased nodules and improved voice quality suggest that the voice therapy had a positive effect for a majority of the nodule clients.

**Acknowledgments:** This study was supported by grants from The National Institute for Deafness and Other Communication Disorders (NIDCD): R01 DC00266-15, and from the Karolinska Institute. We are grateful to Eva Borell, Eva Ihre, Jenny Iwarsson, Per-Åke Lindestad, Elisabeth Lundström, Anita McAllister, Annika Sääf-Rothoff, and Riitta Ylitalo for offering skill and time in the perceptual and endoscopic evaluations; to Joakim Westerlund for statistical advice and analyses; to Svante Granqvist for computer instructions; and to Kajsa Andersson, Hans Larsson, and Claudio Milstein for technical support.

**REFERENCES**

6. Hirano M, Kurita S, Matsuo K, Nagata K. Laryngeal tissue reaction to stress. Transcripts of the 9th symposium: Care
38. McAllister A. Acoustic Perceptual and Physiological Studies of Ten-Year-Old Children’s Voices. [doctoral the-


