Computer-Assisted Voice Analysis

Establishing a Pediatric Database

Paolo Campisi, MD, MSc; Ted L. Tewfik, MD, FRCSC; John J. Manoukian, MD, FRCSC; Melvin D. Schloss, MD, FRCSC; Elaine Pelland-Blais, MOA, SLP(C); Nader Sadeghi, MD, FRCSC

Objectives: To establish and characterize the first pediatric normative database for the Multi-Dimensional Voice Program, a computerized voice analysis system, and to compare the normative data with the vocal profiles of patients with vocal fold nodules.

Design: A cross-sectional, observational design was used to establish the normative database. The comparative study was completed using a case-control design.

Setting: University-based outpatient pediatric otolaryngology clinic.

Participants: One hundred control subjects (50 boys and 50 girls) aged 4 to 18 years contributed to the normative database. The voices of 26 patients (19 boys and 7 girls) with bilateral vocal fold nodules were also analyzed.

Main Outcome Measures: Demographic data, including sex, age, height, weight, body mass index, and cigarette smoke exposure, were obtained. The Multi-Dimensional Voice Program extracted up to 33 acoustic variables from each voice analysis.

Results: The mean (SEM) values of each of the acoustic variables are presented. At age 12 years, boys experience a dramatic decrease in fundamental frequency measurements. The voices of patients with vocal fold nodules had significantly elevated frequency perturbation measurements compared with control subjects (P<.001).

Conclusions: The vocal profile of children is uniform across all girls and prepubescent boys. Patients with vocal fold nodules demonstrated a consistent acoustic profile characterized by an elevation in frequency perturbation measurements. Normal acoustic reference ranges may be used to detect various vocal fold pathologic abnormalities and to monitor the effects of voice therapy.


Assessment of pediatric dysphonia has proven to be problematic for speech pathologists, pediatricians, and otolaryngologists. Clinical judgments of vocal quality have been commonly derived from subjective grading systems rather than from objective measures, which has resulted in the development of inconsistent descriptive terminology and severity classifications. Furthermore, standard adult diagnostic modalities have demonstrated limited usefulness in the assessment of pediatric voice disorders. Fiberoptic endoscopy, for example, is often difficult and rushed in the uncooperative child, and stroboscopic examination is technically challenging in any young patient.

Computer-assisted voice analysis represents an important diagnostic advancement because it provides objective acoustic measurements, and it is well tolerated by children. The Multi-Dimensional Voice Program (MDVP), in conjunction with the Computerized Speech Lab (Kay Elemetrics Corp, Lincoln Park, NJ), is a highly versatile voice-processing and spectrographic analysis software package ideally suited for use in the pediatric population. It provides an objective, reproducible, and noninvasive measure of vocal fold function. The MDVP extracts up to 33 acoustic variables from each voice analysis and compares them graphically or numerically with a built-in normative database. The normative data, however, were derived solely from adults. It is apparent that a pediatric database must be developed if acoustic measures are to be applied to the identification of pediatric vocal pathologic abnormalities.

The main objective of this study, therefore, was to establish and characterize the first pediatric normative database for the MDVP. To our knowledge, a pediatric normative database has not been

From the Departments of Otolaryngology (Drs Campisi, Tewfik, Manoukian, Schloss, and Sadeghi) and Speech-Language Pathology (Ms Pelland-Blais), The Montreal Children’s Hospital, McGill University Health Centre, Montreal, Quebec.
PARTICIPANTS AND METHODS

MDVP ANALYSIS

Apparatus

An IBM-compatible personal computer is used to operate MDVP model 4305. The MDVP is used in combination with Computerized Speech Lab model 4300. The Computerized Speech Lab consists of a hardware and software system that uses an MS-DOS–based computer as host. The Computerized Speech Lab includes signal conditioning capability, 16-bit analog/digital converters, and dual digital signal processors. The MDVP uses the signal conditioning and analog/digital hardware to sample speech at 50 kHz for sustained voicing.

The MDVP extracted up to 33 acoustic voice variables from each voice analysis. These variables were displayed numerically and graphically and were classified into 1 of 6 groups: (1) fundamental frequency information; (2) frequency perturbation; (3) amplitude perturbation; (4) noise and tremor evaluation; (5) voice break, subharmonic, and voice irregularity; or (6) miscellaneous. Definitions of the individual variables listed in Table 1 are available from the authors.

Technique

A consistent technique was used for each MDVP analysis. Seated in a quiet room, the subject held a microphone at a fixed distance (8 cm) and at a 45° off-axis position to reduce aerodynamic noise from the mouth. The subject was then instructed to vocalize and sustain the vowel a 3 times in a flat tone, at a comfortable pitch and a constant amplitude. To standardize the input amplitude, the input signal was adjusted to a predetermined level. This adjustment prevented signal loss and system overloading. Each subject’s third production of a was recorded. A 3-second voice sample was captured and incorporated into the MDVP using a microphone (Visi-Pitch; Kay Elemetrics Corp). The voice sample was not trimmed. The MDVP analysis was then performed, and the acoustic voice variables were displayed.

ESTABLISHING THE NORMATIVE DATABASE

Control Subjects

One hundred control subjects (50 boys and 50 girls) aged 4 to 18 years contributed to the normative database. Subjects were recruited from a pediatric otolaryngology outpatient clinic (Montreal Children’s Hospital, Montreal, Quebec). All subjects were healthy and had no history of sensorineural or voice pathologic abnormalities. Patients with moderate to severe conductive hearing loss or any degree of sensorineural hearing loss were excluded from the study.

Demographic data, including sex, age, height, weight, body mass index, and cigarette smoke exposure, were obtained for each subject. The absence of pathologic abnormalities of the vocal fold was verified in each subject using indirect mirror laryngoscopy or flexible nasolaryngoscopy. An MDVP analysis was then performed, as described in the “Technique” subsection. All laryngoscopies and voice analyses were performed by a single observer (P.C.) to eliminate interobserver variability.

Statistical Analysis

The normative data were analyzed using a statistical software program (SPSS/PC+; SPSS Inc, Chicago, Ill). Backward stepwise multiple linear regression was used to identify statistically significant associations between the acoustic voice variables and the independent variables of sex, age, height, weight, and body mass index. A 2-tailed $P<.05$ was considered statistically significant. The mean (SEM) of each acoustic variable was calculated.

CASE-CONTROL STUDY

Patients

Twenty-six patients (19 boys and 7 girls) with vocal fold nodules, diagnosed using flexible nasolaryngoscopy, were recruited into the study. All the patients had bilateral vocal fold nodules at the junction of the anterior one third and posterior two thirds of the vocal folds. The presence of hemorrhagic nodules or other laryngeal pathologic abnormalities resulted in exclusion from the study. Recruited patients were evaluated by a speech language pathologist (E.P.-B.) and underwent a perceptual evaluation of the voice. An MDVP analysis was then performed as described in the “Technique” subsection.

Statistical Analysis

Determination of a statistically significant difference in voice variable values between the control group and the vocal fold nodule group was achieved using 1-way analysis of variance. Again, a 2-tailed $P<.05$ was considered statistically significant.

If a statistically significant difference in a voice variable was detected, a threshold value was assigned as the upper limit of the 95% confidence interval (mean + 1.96 × SD) of the control group value. Based on the threshold value, data for the control and nodule groups were dichotomized, and a $2 \times 2$ table was constructed. A $\chi^2$ test was then used to assess the statistical significance of the distribution of the dichotomized data. An odds ratio was also calculated to quantify the association between the presence of vocal fold nodules and a voice variable value greater than the assigned threshold value.

RESULTS

NORMATIVE DATABASE

Voice samples from 100 control subjects were used to develop the normative database. Backward stepwise mul-
multiple linear regression revealed a statistically significant association between the fundamental frequency measurements and the independent variables of age and sex. This association was strongly affected by the peripubescent changes in the male voice pattern. All other variables were not affected by age and sex. When boys 12 years and older were excluded from the analyses, the association between the fundamental frequency measurements and age and sex was not significant (*P = .001*; Jitt, jitter; RAP, relative average perturbation; PPQ, pitch period perturbation quotient; and sPPQ, smoothed PPQ. Values are expressed as means ± 1.96 *SEM*).

The mean value of each of the acoustic variables is presented in Table 1. The corresponding SEM is also presented to provide an estimate of the variability in the population at large. To eliminate the effect of peripubescent voice changes on the fundamental frequency measure-
significant increase in frequency perturbation measurements corroborates the findings of the perceptual voice evaluation performed by the speech language pathologist. The perceptual evaluation revealed anomalies in the control of pitch rather than in the control of intensity.

When the control (n=94) and vocal fold nodule (n=23) group data were dichotomized relative to the defined normative threshold values (mean ± 1.96 × SD), a significant distribution of the data was observed for each of the frequency perturbation measurements (P<.001 by χ² test). The calculated odds ratios suggest a high risk of having vocal fold nodules with a variable value greater than the assigned threshold value. The calculated odds ratios range from 28.4 to 41.6, and the 95% confidence intervals do not approach unity (Table 3).

### Table 3. Two-by-Two Tables Generated by Dichotomizing Data for Frequency Perturbation Measurements in Individuals With and Without Vocal Fold Nodules

<table>
<thead>
<tr>
<th>Acoustic Variable</th>
<th>Threshold Value, Mean ± 1.96 × SD†</th>
<th>Control (n = 94)</th>
<th>Vocal Fold Nodule (n = 23)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jita</td>
<td>≤95.23</td>
<td>88</td>
<td>6</td>
<td>41.6 (12.0-144.4)</td>
</tr>
<tr>
<td></td>
<td>&gt;95.23</td>
<td>6</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Jitt</td>
<td>≤2.52</td>
<td>87</td>
<td>7</td>
<td>28.4 (8.8-92.1)</td>
</tr>
<tr>
<td></td>
<td>&gt;2.52</td>
<td>7</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>RAP</td>
<td>≤1.54</td>
<td>88</td>
<td>7</td>
<td>33.5 (10.0-112.9)</td>
</tr>
<tr>
<td></td>
<td>&gt;1.54</td>
<td>6</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>PPQ</td>
<td>≤1.44</td>
<td>88</td>
<td>7</td>
<td>33.5 (10.0-112.9)</td>
</tr>
<tr>
<td></td>
<td>&gt;1.44</td>
<td>6</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>sPPQ</td>
<td>≤1.54</td>
<td>89</td>
<td>8</td>
<td>33.4 (9.6-115.8)</td>
</tr>
<tr>
<td></td>
<td>&gt;1.54</td>
<td>5</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

*OR indicates odds ratio; CI, confidence interval; Jita, absolute jitter; Jitt, jitter percentage; RAP, relative average perturbation; PPQ, pitch period perturbation quotient; and sPPQ, smoothed PPQ.
†Threshold value equals the upper limit of the 95% CI of the control value. P<.001 by χ² test for all.

The functional assessment of pathologic voices is commonly achieved using perceptual and equipment-based clinical tools.1 Perceptual analyses such as the Wilson Voice Profile System and the GRBAS (grade, roughness, breathy, asthenic, strained) scale are based on the subjective interpretation of the individual speech language pathologist.1,2 The lack of consistency and standardization in the basic methods of perceptual assessment continues to be a major clinical problem. Instrumental diagnostic modalities such as video stroboscopy, electroglottography, and phonetography are indispensable components of a modern voice laboratory.1 These equipment-based tools, however, require costly and specialized instrumentation, an experienced operator, cooperative patients, and interpretation of complicated graphs and mathematical formulas.

The assessment of pediatric dysphonia presents unique challenges to the voice scientist. First, it is difficult for children to cooperate with lengthy, uncomfortable examinations. Fiberoptic endoscopy, for example, is often rushed in the uncooperative child. Second, pediatric normative data are unavailable. Therefore, the question arises as to whether a given voice measurement is normal or pathologic and, if pathologic, how severe. Computer-assisted voice analysis, such as the MDVP, represents a clinically important contribution to the assessment of pediatric dysphonia. This system provides objective and reproducible results. Assessments are non-invasive, completed in a short time, and well tolerated by patients as young as 4 years. However, the development and characterization of a normative pediatric database is a prerequisite to the application of this technology to the assessment of pediatric voice pathologic abnormalities.

The main objective of this study was to establish and characterize the first pediatric normative database for the MDVP. Multiple linear regression was used to assess the association between the derived acoustic variables and the independent variables of age, sex, height, weight, and body mass index. In general, we found that girls of all ages and boys younger than 12 years had the same vocal profiles. However, age and sex were significantly associated with fundamental frequency measurements when boys aged 12 years and older were included in the statistical analyses. Fundamental frequency measurements in boys sharply decreased and approached adult values beginning at age 12 years. No association was detected between the acoustic variables and the independent variables of height, weight, and body mass index. Our findings are consistent with previously published study results.9,10 Sussman and Sapienza4 examined the developmental and sex trends in fundamental frequency in 17 boys and 14 girls aged 6.1 to 9.2 years. They found that the fundamental frequency for vowel production of boys and girls (aged <12 years) was not significantly different but were markedly different from men. Harries and colleagues10 reported abrupt changes in male speaking and singing fundamental frequencies during puberty between Tanner stages G3 and G4, corresponding to an average age of 13 years.

Vocal fold nodules are a common cause of pediatric dysphonia. In fact, 38% to 78% of children evaluated for chronic hoarseness have vocal fold nodules.11 Nodules are the result of voice misuse or abuse, and they oc-
Nodules result from injury to the superficial layer of the lamina propria and the basement membrane zone caused by extensive vibration that destroys the tissue. An aberrant healing response occurs with excessive deposition of collagen type IV and fibronectin.

The normative acoustic data were compared with the vocal profiles of patients with vocal fold nodules using a retrospective, case-control study design. Patients with vocal fold nodules demonstrated a consistent acoustic profile characterized by markedly elevated frequency perturbation measurements. This finding was exemplified by abnormal jitter values. Jitter is defined as the cycle-to-cycle variation of frequency. In other words, it is a measure of unintended frequency unsteadiness during prolonged phonation. This abnormality was corroborated by the perceptual analysis, which demonstrated an inability of these patients to control pitch. Furthermore, several of our patients who were treated with voice therapy demonstrated normalization of their frequency perturbation measurements as the size of their nodules decreased.

Abnormally elevated frequency perturbation measurements in children with vocal fold nodules have been reported in the literature. In a small series of 10 patients, aged 6 to 8 years, Kane and Wellen demonstrated a correlation between perceptual severity ratings and elevated PPQ measurements. In 1997, Boltezar and colleagues failed to reveal an elevation in PPQ in 11 adolescents (aged 10-17 years) with vocal fold nodules compared with a control group (jitter and relative average perturbation were not measured). However, PPQ measurements were abnormally elevated compared with adult values published in the literature.

To evaluate the ability of the MDVP to detect vocal pathologic abnormalities, the frequency perturbation measurement data were dichotomized in controls and patients according to an assigned threshold value. The calculated odds ratios strongly suggest that an elevated frequency perturbation measurement is highly associated with the presence of vocal nodules. Indeed, the strong statistical association suggests that the MDVP may be useful in detecting early, subclinical pathologic abnormalities and in monitoring treatment. The role of the MDVP in detecting other vocal pathologic abnormalities, such as respiratory papillomatosis, polypoid degeneration, and congenital sulcus vocalis, needs to be further investigated.

In conclusion, the MDVP is an objective, noninvasive diagnostic tool that complements existing functional voice analysis modalities. It represents an important advancement in the assessment of pediatric dysphonia. The normative database may be used to reliably assess the voices of girls and prepubescent boys. However, boys older than 12 years should be compared with age-matched controls.

Accepted for publication September 10, 2001.

Presented in part at the 2000 Annual Meeting of The American Academy of Pediatrics, Section on Otolaryngology and Bronchoesophagology, Chicago, Ill, October 29, 2000.

We thank the Gustav Levinschi Foundation, Montreal, Quebec, for donation of the Voice and Speech Laboratory.

Corresponding author and reprints: Ted L. Tewfik, MD, FRCS, Director of Otolaryngology, The Montreal Children’s Hospital, 2300 Tupper St, Suite B240, Montreal, Quebec, Canada H3H 1P3.

REFERENCES