Consistency of Voice Frequency and Perturbation Measures in Children

Courtney A. Hill, MD1, Shilpa Ojha, MBChB2, Stephen Maturo, MD3, Rie Maurer, MA4,5, Glenn Bunting, MS, CCC-SLP6, and Christopher J. Hartnick, MD2

Abstract

Objective. Evaluate normal pediatric voice frequency and perturbation measures with Voice Evaluation Suite (VES) and Multi Dimensional Voice Program (MDVP), determine the consistency of these measures over time, and understand which measures might be most useful for evaluating children with voice disorders.

Study Design. Prospective, longitudinal study of normal voices of 50 children aged 4 to 17 years.

Setting. Pediatric otolaryngology clinic within tertiary hospital.

Subjects and Methods. Two tests of sustained utterances from each child were evaluated by 2 computerized voice analysis programs for frequency and perturbation. Intraclass correlation coefficient (ICC) was used to assess the reliability between the samples.

Results. Children (male/female, 1.08:1) with a mean age of 8.34 years were tested on an average of 54.2 minutes apart. Each test included 4 utterances; 1 was analyzed by MDVP, and 3 grouped utterances were averaged and evaluated by VES. Fundamental frequency had excellent reliability (ICC = 0.95) in both VES and MDVP. Jitter, shimmer, and noise to harmonic ratio were poorly reliable (ICC ≤0.4) in MDVP but had good to excellent reliability (ICC 0.66-0.8) in VES.

Conclusion. Single, sustained utterances in children provide consistent measures of frequency. Perturbation is not reliably measured by such testing, but averaging multiple samples yields improved consistency. Evaluating acoustic measure stability in spontaneous speech and in sustained utterances cued by a tuning frequency can provide further insight on pediatric voice consistency.

Keywords

pediatric, voice

Received October 22, 2012; revised January 9, 2013; accepted January 17, 2013.

Human voice has long been evaluated for acoustic measures such as frequency and perturbation. Historically, any variations in normal voice were designated as abnormalities; however, Baken hypothesized that abnormalities, particularly in perturbation measures, were in fact representative of the variability found within the normal adult voice. In a normative study of pediatric vocal acoustics, Maturo et al observed ranges of jitter (variation in frequency) and shimmer (variation in amplitude) that were much wider than those found for fundamental frequency (F0), with more variability found in boys than in girls. Although these variations may fall within the realm of normal, they have raised the question of whether acoustic measures such as F0, jitter, shimmer, and noise to harmonic ratio (NHR) are consistent, that is, whether they are reliably reproduced without significant difference between multiple measurements. In addition, it questions whether a single measurement can be considered a reliable representation of the subject’s voice at any given time.

Fundamental frequency has often been found to be reliable in repeated voice measurements in normal adults. Bough and colleagues observed high correlations among repeated F0 measures of the sustained vowel /a/ in the same day in normal adult voices. Fundamental frequency in 21- to 26-year-olds...
tested 7 to 10 days apart was not found to be significantly different in measurements derived from sustained vowels, reading, or spontaneous speech. Leong et al found that F0 was reliable when calculated from multiple samples of sustained /a/ over the course of a month.

On the other hand, studies of perturbation measures have revealed varied reliability. Bough et al found that jitter and shimmer were not correlative in intraday, sustained vowel testing, and in fact, correlation worsened when samples from different days were compared. Leong et al observed that jitter was highly reliable but only in male voices. Noise to harmonic ratio has been found to be highly correlative in repeated voicing of sustained /a/ in the same day but not reliable when measured multiple times over the course of a month.

In light of these variances, many investigators have studied factors that might explain the lack of consistency in acoustic measurements. Caffeine intake, mood, sleep deprivation, hydration status, hormonal effects of the menstrual cycle, time of day, and ambient noise are a few of these factors that have been hypothesized to affect vocal quality. Akhtar and colleagues observed no significant difference in irregularity of fundamental frequency before and after caffeine consumption. Importantly, they noted significant intersubject variability, meaning that individual differences in this study population were more significant than the effects of increased levels of caffeine in each individual. Energetic, stress-free, and positive mood was found to be associated with lower jitter. Amount of sleep and hydration had no correlation with fundamental frequency before and after caffeine consumption. Importantly, they noted significant intersubject variability, meaning that individual differences in this study population were more significant than the effects of increased levels of caffeine in each individual. Energetic, stress-free, and positive mood was found to be associated with lower jitter. Amount of sleep and hydration had no correlation with fundamental frequency before and after caffeine consumption. Importantly, they noted significant intersubject variability, meaning that individual differences in this study population were more significant than the effects of increased levels of caffeine in each individual.

It is important to determine if these variations exist in children’s voices and if one vocal sample is truly representative of voice over time. Children are more often afflicted with vocal disorders than adults yet pediatric voice is much less extensively studied. However, recently there has been an increase in pediatric vocal studies, including the establishment of normative acoustic measures in children, and this study aims to add consistency data to the growing literature on pediatric voice. We evaluate frequency and perturbation measures in normal pediatric voices with the goal of determining the consistency of these measures over time.

### Methods

Institutional review board approval was obtained from the Massachusetts Eye and Ear Infirmary. Children aged 4 to 18 years who were evaluated in the outpatient pediatric otolaryngology clinic between May and October 2012 were offered participation in this study. Exclusion criteria included reason for visit being hearing loss or posttonsillectomy check, smoking, history of voice disorders, developmental or cognitive delay, and subjective difficulty with speech production or hearing. Participation in the study was contingent upon the child’s agreement. In addition, informed consent was obtained from each child’s parent, and informed assent was obtained from children 12 years and older.

Children were seated in a quiet, soundproof room, and a headset-mounted, adjustable microphone (Shure Beta 53; Shure, Inc, Niles, Illinois) was placed 3 centimeters from the right oral commissure. Recordings were performed with a Dell (Dell, Inc, Round Rock, Texas) Optiplex 960 personal computer (Microsoft Windows XP Professional Version 2002; Microsoft Corp, Redwood, Washington) with an Intel Core Duo 2 CPU (3.1 GHz, 1.94 GB of RAM). Multi Dimensional Voice Program (MDVP) Model 5105 software option for the Computerized Speech Lab Model 4500 (KayPENTAX, Lincoln Park, New Jersey) and Voice Evaluation Suite (Estill Voice International, Pittsburgh, Pennsylvania) were used to evaluate F0, jitter, shimmer, and NHR of sustained vowel utterances. For MDVP evaluation, patients were asked to sustain the vowel /a/ at a comfortable pitch and volume using a normal speaking voice for 4 seconds. The middle 3.5 seconds of voicing was captured for analysis. The subjects were asked to voice a similar /a/ for analysis with Voice Evaluation Suite (VES). The program detected the start of voicing and recorded from the beginning for a total of 4 seconds. This was repeated for a total of 3 samples that were then averaged by the VES program to yield 1 set of data for each test session. These differences in methods of voice capture and analysis are those that have been established by the individual programs for clinical use. Each child underwent 1 testing session before and 1 session after their clinic visit. Subjects were not included if less than 30 minutes elapsed between the 2 sessions.

The data produced from each voice analysis program were evaluated separately. The mean of all voice samples for each variable was compared between the 2 tests. Intraclass correlation coefficients (ICCs) and 95% confidence intervals were calculated for the mean of each of the 4 variables. Fleiss described that ICC values above 0.75 represent excellent reliability, values between 0.4 and 0.75 represent fair to good reliability, and values below 0.4 represent poor reliability. Calculation prior to the start of the study revealed that a sample size of 50 would yield statistical significance (α < 0.05) with 99% power supporting an alternative hypothesis that these acoustic measures would have excellent reliability with an ICC of 0.75.

### Results

Fifty children aged 4 to 17 years with a male to female ratio of 1.08 to 1 were included in this study. Subjects were 8.34 years old on average with a median age of 8 years. Time...
between testing sessions ranged from 30 to 115 minutes, with a mean of 54.2 minutes.

Mean fundamental frequency was found to have an ICC of 0.95 when analyzed by both MDVP and VES, indicating excellent reliability (Tables 1 and 2). Mean jitter, shimmer, and NHR had good to excellent reliability in averaged, repeated samples in VES with ICCs of 0.66, 0.67, and 0.8, respectively. However, these variables were poorly reliable when measured with only a single sample in each testing session in MDVP; ICCs were 0.4, 0.3, and 0.02 for jitter, shimmer, and NHR, respectively.

Discussion

This study revealed that single, sustained utterances in children can provide consistent measures of frequency. This is similar to other reports of consistent fundamental frequency in which data from multiple samples were averaged.4,6 Bough and colleagues4 found that 3 averaged intraday samples of sustained /a/ and groups of 3 or 5 daily averages produced consistent F0. This study found that averaging multiple samples in a single session may not be necessary to obtain representative measurements of fundamental frequency.

Perturbation measures in children were found to be reliable only when 3 serial samples of sustained vowel were averaged. Single utterances did not yield the same reliability as in fundamental frequency. Historically, perturbation has been less reliable than frequency, and this study supports these findings. Bough and colleagues6 found that 3 averaged intraday samples of sustained /a/ produced consistent NHR measurements, but jitter and shimmer were not reliably reproduced in adults.6 In fact, inconsistency of shimmer increased as the number of intraday tests were evaluated.6 Jitter became less consistent when interday samples were compared in groups of 3 or 5 days.6 In this pediatric population, perturbation measures were less reliable than F0. This likely represents the variability within the normal pediatric population as is similarly proposed by Baken6 for adults. Alternatively, this could be a product of extralaryngeal factors that affect perturbation, such as anxiety and fatigue, that can be part of a child’s experience in a clinic setting. Furthermore, one must consider the statistical differences that can result from the 2 methodologies used in the 2 evaluation systems. It could be that the VES analysis yielded excellent reliability for jitter, shimmer, and NHR only because of the averaging of 3 samples.

There is a paucity of normative pediatric voice data to compare with our average acoustic measurements. There are, however, normative data for MDVP evaluation of fundamental frequency, jitter, shimmer, and NHR for children aged 4 to 18 years, grouped by sex and age.3 Given that the averages in this study were taken for all patients, both boys and girls together, it is not possible to compare directly with the normal values previously published by sex. It appears, however, that our F0, jitter, shimmer, and NHR data analyzed by MDVP fall within the normal range for either sex according to Maturo et al.3 Further studies are needed to elucidate the normative values of the remaining vocal parameters in children.

Strengths of this study include a moderately large sample size and minimization of extralaryngeal factors that might affect vocal acoustics in children. Our sample size of 50 is considerably larger than those of studies evaluating vocal consistency in adults that have evaluated 20 or fewer subjects.5-6,8 This sample size was chosen as it would yield statistical significance (α <0.05) with 99% power, supporting an alternative hypothesis that these measures would have excellent reliability with an ICC of 0.75.

Additional advantages of this study are intrinsic to its design. Sex differences in perturbation measures have been found in adult studies of consistency and in single samples.

### Table 1. Multi Dimensional Voice Program Variables

<table>
<thead>
<tr>
<th></th>
<th>Test 1, Mean ± SD</th>
<th>Test 2, Mean ± SD</th>
<th>ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0</td>
<td>244.57 ± 53.61</td>
<td>246.06 ± 52.92</td>
<td>0.95 (0.92 to 0.97)</td>
</tr>
<tr>
<td>Jitter</td>
<td>1.57 ± 0.80</td>
<td>1.52 ± 0.80</td>
<td>0.40 (0.13 to 0.61)</td>
</tr>
<tr>
<td>Shimmer</td>
<td>3.38 ± 1.43</td>
<td>3.32 ± 1.55</td>
<td>0.30 (0.02 to 0.53)</td>
</tr>
<tr>
<td>NHR</td>
<td>0.11 ± 0.02</td>
<td>0.12 ± 0.04</td>
<td>0.02 (−0.25 to 0.30)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; NHR, noise to harmonic ratio.

### Table 2. Voice Evaluation Suite Variables

<table>
<thead>
<tr>
<th></th>
<th>Test 1, Mean ± SD</th>
<th>Test 2, Mean ± SD</th>
<th>ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0</td>
<td>243.91 ± 53.90</td>
<td>242.23 ± 51.80</td>
<td>0.95 (0.91-0.97)</td>
</tr>
<tr>
<td>Jitter</td>
<td>1.11 ± 0.48</td>
<td>1.09 ± 0.53</td>
<td>0.66 (0.47-0.79)</td>
</tr>
<tr>
<td>Shimmer</td>
<td>1.79 ± 0.64</td>
<td>1.71 ± 0.61</td>
<td>0.67 (0.49-0.80)</td>
</tr>
<tr>
<td>NHR</td>
<td>8.87 ± 3.69</td>
<td>9.65 ± 3.37</td>
<td>0.80 (0.68-0.88)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; NHR, noise to harmonic ratio.
of pediatric voice. Matsuo et al found that girls aged 4 to 17 overall had significantly lower jitter and shimmer percent than boys when they were tested in a single session. Given that the goal of this study was not to determine acoustic differences between boys and girls, data from all children were grouped together to calculate means from each testing session, and thus the subjects functioned as their own controls. Also minimized was the variance of perturbation that has been seen secondary to time of day. The subjects were tested at approximately the same time of day with the longest interval between tests being 115 minutes. Last, although background noise level was never measured, ambient noise was minimized with a soundproof room. Background noise has been shown to affect the NHR primarily, and our measurements yielded excellent consistency of the NHR.

This study was limited by study population, poor control of intrinsic factors that might affect voice, no control of frequency or intensity, evaluation of sustained utterances alone without evaluation of spontaneous speech, and a study period of a single day. In the outpatient pediatric otolaryngology clinic volunteered for participation in this study, but a sample more representative of the pediatric population could be recruited in other settings such as a school. Barriers to research in schools, however, would likely prohibit carrying out this design. Although our study population might not reflect all children, it should be representative of children without vocal abnormalities.

We did not record histories of, but also had no control over, caffeine intake, sleep deprivation, hydration status, or menstrual status of subjects. The average age of female participants was 8.42 years (median 9 years), with only 3 subjects older than 10 years; it was unlikely that any effects from menstrual cycle status of postpubescent females significantly altered the data. Furthermore, there are conflicting reports of the extent of effect that these intrinsic factors have on vocal consistency. We observed consistent frequency and perturbation without control of these factors.

Acoustic measures can be obtained from various vocal samples, including spontaneous speech, reading speech, and sustained vowels. This study focused on sustained vowels; however, there may be advantages to measuring acoustics in spontaneous speech. Fitch found that fundamental frequency was very consistent when tested with reading and spontaneous speech and that sustained vowels produced the greatest difference between 2 tests measuring F0. Although our fundamental frequencies tended to be consistent, it might be helpful in the future to know which type of vocal sampling produces the most reliable measurements in children.

Although our study design of sampling twice in one day may have eliminated the variance in voice throughout the day, this limited our ability to assess vocal consistency over longer periods of time. Studies in adults have shown reliability of various acoustic measures over several days to approximately 1 month. A longer time period between voice samples might be more representative of individual variations in voice and thus give a more true representation of a child’s acoustics at any given time.

Fundamental frequency and intensity were not controlled when evaluating perturbation measures in this study. Lee and colleagues found that controlling fundamental frequency produced very reliable jitter percent in vocal samples recorded over a period of 28 days, although there were no significant differences in perturbation when subjects found their own “comfortable” pitch. Control of intensity produces significantly different jitter percent than when fundamental frequency alone is controlled or when neither intensity nor frequency is controlled. The most variability resulted when subjects were allowed to choose their vocal intensity. Despite these previous findings, this study observed good to excellent consistency of perturbation measures in averaged, serial recordings when children were asked to produce a sustained vowel in a comfortable pitch and volume. Perhaps controlling the intensity would yield reliable jitter, shimmer, and NHR measures in single vocal samples and should be investigated further in studies of pediatric voice.

In conclusion, this is the first study to the authors’ knowledge that addresses consistency of voice acoustics in children. Single, sustained vowel utterances in children provide consistent measures of frequency. Perturbation is not reliably measured by such testing, but averaging multiple samples yields improved consistency. These findings should be considered when evaluating abnormal vocal acoustics in children. Future studies evaluating acoustic measure stability in spontaneous speech and in sustained utterances cued by a tuning frequency can provide further insight on pediatric voice consistency.

Author Contributions
Courtney A. Hill, data acquisition, data analysis, drafting and revision of paper, final approval of manuscript; Shilpa Ojha, data acquisition, revision of paper, final approval of manuscript; Stephen Matsuo, data analysis, paper revision, final approval of manuscript; Rie Maurer, study design (statistics), revision of paper, final approval of manuscript; Glenn Bunting, data interpretation, revision of paper, final approval of manuscript; Christopher J. Hartnick, study design, data analysis, revision of paper, final approval of manuscript.

Disclosures
Competing interests: None.
Sponsorships: None.
Funding source: None.

References

Downloaded from otol.sagepub.com at Harvard University on July 3, 2013


