Duration of ambulatory monitoring needed to accurately estimate voice use

Daryush Mehta1; Rebecca Woodbury Listfield1; Harold Cheyne II2; James Heaton1; Shengran Feng1; Matías Zañartu1; Robert Hillman1

1Ctr. for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital; 2Lab of Ornithology, Cornell University; 3Dept. of Electronic Engineering, Universidad Técnica Federico Santa María

Abstract

Voice use is considered to play a major role in the development of many voice disorders, and clinicians focus on evaluating and modifying how patients use their voices throughout the day. Some voice monitoring devices have used neck-mounted accelerometers to unobtrusively and confidentially track voice use–related measures, such as phonation time, fundamental frequency, and sound intensity. Guidelines for the clinical use of such monitoring devices have yet to be established. This is a preliminary investigation to establish initial benchmarks for obtaining robust estimates of long-term average voice use that may be used to begin examining basic relationships between vocal loading and voice use–related pathology. As expected, adequate monitoring durations depend on the inherent variability of the parameter of interest, with much of the error decreasing after 24 hours of monitoring. Investigations are currently under way to take advantage of a smartphone-based voice monitoring system that is designed to enhance device wearability and enable the derivation of new clinically relevant measures.

Motivation

Evaluating and modifying how voice–disordered patients typically use their voices currently relies on patients’ self-reporting. Self-reporting is associated with uncertainty and reliability issues, particularly given that patterns of voice use and misuse become highly habituated and somewhat automatic. Such issues have motivated the development of ambulatory voice monitors. This study provides initial benchmarks for quantifying the error in the estimation of five voice use–related measures as a function of monitoring duration. Such data do not exist for various professions and provide evidence to investigate theories relating the impact of long-term vocal loading on voice use–related pathologies.

Study Design

This pilot study includes 18 subjects in three groups matched for gender and approximate age, with additional matching for occupation between patients and controls. Subjects were monitored for five work days using the KayPENTAX Ambulatory Phonation Monitoring (APM), yielding 20–70 hours of data per subject, to address the following study question:

How long must a subject be monitored to adequately estimate long-term average values for phonation time, fundamental frequency, sound pressure level, cycle dose, and distance dose?

Example Voice Use Profile Using Neck Skin Acceleration

Voice Use Measures

Two measures were computed for each 125-ms frame (no overlap) that contained phonation:

1. Fundamental frequency (F0): Derived using a time-based autocorrelation, 70–400 Hz range
2. Sound pressure level (SPL): Calibrated using a linear mapping between neck skin acceleration amplitude and acoustic SPL at 15 cm; 50–130 dB SPL range

Three measures were then derived for a given monitoring duration:

1. Phonation time: Percentage of frames exhibiting phonation
2. Cycle dose: Integer number of glottal pulses
3. Distance dose: Theoretical distance the vocal folds travel

Results

Percent Error after Each Hour of Monitoring

Error Analysis

There were no statistically significant differences among the average error curves for subjects within and across each subject group.

Error compares the cumulative averages at each hour with the “ground truth average” at 40 hours for each measure:

\[
\text{Error} = \frac{\text{Average}_{40} - \text{Average}_{t}}{\text{Average}_{40}} \times 100
\]

Average errors in F0 and SPL decrease to about 1% after 20 hours of ambulatory monitoring. In contrast, errors associated with phonation time, cycle dose, and distance dose were high over the first several hours, requiring at least 26 hours of data to yield average errors below 10%.

Conclusions

We computed the error associated with estimating typical voice use parameters using varying monitoring durations. The average, minimum, and maximum errors associated with varying monitoring durations for five voice-use measures provide initial benchmarks to determine the monitoring time necessary to yield a desired level of accuracy for the parameters of interest.

Acknowledgments

This project is supported by grants from the NIH National Institute on Deafness and Other Communication Disorders (T32 DC0053 and R21 DC011588), the Institute of Laryngology and Voice Restoration, and by the Chilean CONICYT under Grant PONDECTIT-11110147. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

New Smartphone-Based Voice Monitor

A new smartphone-based ambulatory voice monitor is currently being developed by our group. The new system takes advantage of a smartphone as the data acquisition device, which is expected to increase subject comfort and compliance as compared with larger APM–like systems. A critical improvement is the recording of the raw neck skin acceleration waveform, which is in contrast to previous devices that did not save the waveform but instead used on–board processing to extract and store measures. Long–term recording of the accelerometer’s signal allows for post–processing to investigate alternative analysis algorithms, including the derivation of glottal airflow–based measures and pattern recognition techniques to differentiate healthy from potentially pathological vocal patterns.

<table>
<thead>
<tr>
<th>Patients</th>
<th>Controls</th>
<th>Low Voice Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>F 37 yrs Teacher</td>
<td>49 yrs Teacher</td>
<td>—</td>
</tr>
<tr>
<td>F 22 yrs Sales</td>
<td>23 yrs Sales</td>
<td>—</td>
</tr>
<tr>
<td>F 46 yrs Clinician</td>
<td>52 yrs Clinician</td>
<td>52 yrs Researcher, Histologist</td>
</tr>
<tr>
<td>F 54 yrs Teacher</td>
<td>62 yrs Teacher</td>
<td>59 yrs Researcher, Histologist</td>
</tr>
<tr>
<td>M 19 yrs Singer</td>
<td>20 yrs Singer</td>
<td>29 yrs Researcher, Engineer</td>
</tr>
<tr>
<td>F 30 yrs Clinician</td>
<td>26 yrs Clinician</td>
<td>29 yrs Animal Care Technician</td>
</tr>
<tr>
<td>M —</td>
<td>—</td>
<td>31 yrs Researcher, Physiologist</td>
</tr>
<tr>
<td>M —</td>
<td>—</td>
<td>48 yrs Researcher, Engineer</td>
</tr>
</tbody>
</table>

This is a preliminary investigation to establish initial benchmarks for obtaining robust estimates of long-term average voice use that may be used to begin examining basic relationships between vocal loading and voice use–related pathology. As expected, adequate monitoring durations depend on the inherent variability of the parameter of interest, with much of the error decreasing after 24 hours of monitoring.