Direct Comparison of Three Commercially Available Devices for Voice Ambulatory Monitoring and Biofeedback

Jarrad H. Van Stan
Center for Laryngeal Surgery & Voice Rehabilitation, Massachusetts General Hospital
Boston, MA
Center for Interprofessional Studies & Innovation, MGH Institute of Health Professions
Charlestown, MA

Joakim Gustafsson
Division of Speech & Language Pathology, Department of Clinical Sciences, Intervention & Technology, Karolinska Institute
Stockholm, Sweden

Ellika Schalling
Division of Speech and Language Pathology, Department of Clinical Sciences, Intervention & Technology, Karolinska Institute
Department of Speech and Language Pathology, Karolinska University Hospital
Stockholm, Sweden

Robert E. Hillman
Center for Laryngeal Surgery & Voice Rehabilitation, Massachusetts General Hospital
Boston, MA
Communication Sciences & Disorders, MGH Institute of Health Professions
Charlestown, MA
Department of Surgery, Harvard Medical School
Boston, MA

Disclosure: Financial: Robert E. Hillman has a financial interest in the Ambulatory Phonation Monitor based on a contractual agreement between Sensimetrics, Inc. (R&D for the initial version of the Ambulatory Phonation Monitor) and KayPENTAX, Inc. (manufacturer of the Ambulatory Phonation Monitor). Jarrad H. Van Stan, Joakim Gustafsson, and Ellika Schalling have no financial interests to disclose.
Nonfinancial: Jarrad H. Van Stan, Joakim Gustafsson, Ellika Schalling, and Robert E. Hillman have no nonfinancial interests to disclose.

Abstract

Purpose: To provide a direct comparison of three commercially available devices for voice ambulatory monitoring and biofeedback to assist voice clinicians and researchers in choosing the device that best meets their needs.

Methods: The Ambulatory Phonation Monitor (APM), VocaLog, and VoxLog were descriptively compared regarding cost, availability, physical characteristics, and operational features; and quantitatively compared regarding measures of loudness (dB sound pressure level [SPL]), pitch (fundamental frequency – F0), and phonation time. The quantitative comparison used simultaneous registrations acquired during a 90-minute lecture which also included a Smartphone-based system designed to capture the raw voice signal as a reference.

Results: Differences regarding cost, physical characteristics, and operational features could influence device choice. All three devices register SPL and phonation time, and also biofeedback based on SPL. Furthermore, the APM and VoxLog provide monitoring...
and biofeedback for F0 with the associated capability of providing additional measures related to vocal dose. All devices produced comparable results for common measures, except for the overestimation of phonation time by the VocaLog.

Conclusions: The cost, operational features, and performance characteristics of the three commercially available devices for voice ambulatory monitoring and biofeedback differ in ways that can significantly impact decisions about which one is best suited for a particular application.

Behaviorally based voice disorders (e.g., vocal nodules, polyps, and muscle tension dysphonia) can be difficult to accurately assess in the clinical setting since they are thought to result from faulty and/or abusive patterns of vocal behavior exhibited in daily life (Hillman, Heaton, Masaki, Zeitels, & Cheyne, 2006; Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989). The actual role, however, of vocal behaviors in the etiology of what are considered voice-use-related disorders is still not well understood. Speech-language pathologists must mostly rely on patients reporting about their own vocal behavior outside of the therapy session, which is prone to unreliability. Vocal behaviors and the impact of voice therapy on those behaviors can potentially be better characterized by long-term ambulatory monitoring of vocal function as individuals engage in their typical daily activities (Ghassemi et al., 2014). Furthermore, biofeedback delivered in the ambulatory setting has the potential to significantly improve one of the hardest aspects of voice therapy, which is carryover outside of the treatment session (KayPENTAX, 2009; Schalling et al., 2013).

There has been increased interest over the past several years in the development of devices for voice ambulatory monitoring and biofeedback. Currently there are three such devices available commercially for clinical and research use: the Ambulatory Phonation Monitor (APM; KayPENTAX, Montvale, NJ), the VoxLog (firmware 2.2.3, Sonvox AB, Umeå, Sweden), and the VocaLog (Griffin Laboratories, Temecula, CA). This brief report provides an initial direct comparison of the three commercially available devices for voice ambulatory monitoring and biofeedback in terms of operational features (including cost and availability) and performance characteristics. This is being done to assist voice clinicians and researchers in choosing the device that best meets their needs.

**Descriptive Comparison of Operational Features**

Currently, the APM is the only device that is available in the United States and Europe. The VocaLog is available only in the United States and the VoxLog is available only in Europe, but is expected to be available in the United States in the near future. The VocaLog is by far the least expensive of the three devices at $999 and its maker Griffin Laboratories is planning to release a VocaLog2 at an even lower price of $394. The APM and VoxLog are currently comparably priced at about $5,000.

The basic specifications for the three devices are shown in Table 1. In terms of hardware, all three devices employ transducers that are placed on the anterior neck below the larynx to sense phonation, and all have a wire running from the transducer to a system box where digital signal processing is performed to extract voice-related measures from the transducer signal. The extracted measures are stored in onboard memory and because none of these devices actually record the transducer signal, confidentiality is maintained (i.e., there is no capability to play back what subjects actually said). The VocaLog uses a contact microphone while the APM and VoxLog each employ a miniature accelerometer (ACC) encased in plastic or silicone. The VoxLog also has a calibrated air microphone built into the assembly that contains the ACC. Both the VocaLog and VoxLog systems use a collar-type device to hold the transducers against the neck (see Figure 1) and can be prone to changing position/contact (slippage) as subjects turn their necks, which could cause changes in the transducer signal. Without close adherence to the neck tissue, any throat sensor will become less sensitive to voicing and more sensitive to
environmental noise (Zañartu et al., 2009). The design of the VoxLog collar seems less prone to slippage than the one used by the VocaLog, but this issue has been addressed in the new VocaLog2, expected to be released sometime in mid-2014 (personal communication, Griffin Labs, January 2014). The APM employs a medical-grade adhesive to hold the ACC against the neck, which is generally effective in maintaining transducer position, but can be more cumbersome to place and remove than a neck collar, and in rare occasions can cause minor skin irritation.

Table 1. Important Differences Regarding Hardware/Software Specifications Between the Three Devices are Summarized Above.

<table>
<thead>
<tr>
<th>Basic Specifications</th>
<th>APM</th>
<th>VoxLog</th>
<th>VocaLog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (L × W × H - cm)</td>
<td>15.5 × 9 × 5</td>
<td>12 × 2.5 × 2.5</td>
<td>7 × 5.5 × 1.5</td>
</tr>
<tr>
<td>Cost (US Dollars)</td>
<td>~5,000</td>
<td>~5,000</td>
<td>999*</td>
</tr>
<tr>
<td>Analysis Window Size</td>
<td>50 ms</td>
<td>Adjustable** 100 ms − 5 s</td>
<td>1 s</td>
</tr>
<tr>
<td>Max Recording Duration</td>
<td>18 hours</td>
<td>7 days</td>
<td>3 weeks</td>
</tr>
<tr>
<td>Sensor</td>
<td>ACC</td>
<td>ACC and Air MIC</td>
<td>Throat MIC</td>
</tr>
<tr>
<td>Biofeedback</td>
<td>Vibrotactile</td>
<td>Vibrotactile</td>
<td>Vibrotactile &amp; Auditory</td>
</tr>
</tbody>
</table>

*The original VocaLog price is listed. A VocaLog2 will be released later in 2014 at the expected price of $394.00.

**The VoxLog analysis window size can be adjusted before starting monitoring. Also, phonation detection is completed every 4 ms.

In terms of ergonomics, all three devices are designed to be portable, but the APM has a much larger and heavier (cumbersome) system box that must be carried in a satchel or case attached to a belt. Both the VocaLog and VoxLog are much smaller than the APM and can be directly clipped to a belt or wristband or placed in a pocket. In addition, all three devices use rechargeable battery technology (the APM can also use non-rechargeable AA batteries), which provides maximum recording times between charges of approximately: APM=18 hours, VocaLog=3 weeks, VoxLog=7 days (see Table 1).

There are also differences between the devices regarding the voice-related measures that are provided and how some of these measures are obtained (see Tables 1 and 2). One fundamental difference is in the sizes of the analysis windows within which measures are averaged. The VocaLog only provides estimates of SPL and registers the presence of phonation once per second, whereas estimates of SPL, F0, and the presence of phonation are registered every 50 milliseconds (ms) by the APM (KayPENTAX, 2009). The VoxLog actually registers the presence of phonation every 4 ms and records measures of SPL and F0 across a range of 100 ms to 5 seconds (pre-set in software by the user; Lindström, 2011). The inclusion of an F0 measure in the APM and VoxLog gives these devices the capability of generating additional derived measures related to vocal dose (e.g., cycle dose, distance dose), which are hoped to have future potential in helping to differentiate between healthy versus damaging phonatory behaviors (Titze, Švec, & Popolo, 2003). In addition, the VoxLog’s simultaneous sensing and analysis of both the ACC and microphone signals gives it the unique capability of providing estimates of noise levels (e.g., competing sound from others talking, music, traffic noise) in the subject’s environment, which can have an impact on vocal behavior, particularly vocal loudness. The VoxLog does this by using the microphone signal to estimate background sound levels during periods when the ACC signal is below an amplitude threshold that indicates the presence of phonation. However, this arrangement can be problematic if the background sound levels exceed the vocal SPL (e.g., singing with a choir or musical accompaniment), in which case the registered dB values will not represent the actual phonation levels but the environmental noise.
levels. It can also be difficult to detect soft voicing over environmental noise, specifically below 65 dB SPL (Sonning, 2013). The devices also differ with respect to the procedures that are used to calibrate them for measuring SPL. The VoxLog does not require an SPL calibration because it has a calibrated microphone built into the system that is positioned approximately 10 cm from the lips depending on placement of the neck collar. The APM and VocaLog require a short calibration procedure while connected to a computer that is running the associated software, with the APM requiring daily calibration and the VocaLog only being calibrated once prior to the first recording for a given subject. When calibrating the APM, a handheld microphone is used to measure dB SPL at 15 cm from the lips while simultaneously recording the ACC signal from the neck as the subjects go from soft to loud phonation on a sustained /a/ vowel (Cheyne, Hanson, Genereux, Stevens, & Hillman, 2003). This allows the software to find calibration factors that map the ACC amplitude at the neck to dB SPL from the lips ($\text{vec, Titze, & Popolo, 2005}$). This daily calibration helps to adjust for conditions that could vary on a daily basis (e.g., vocal fold tissue/vibratory characteristics and sensor location on the neck skin). To calibrate the VocaLog, the amplitude of the signal from a contact microphone on the neck is compared to a calibrated dB SPL meter that is built into the device and positioned 50 cm from the lips. The VocaLog software requires that the patient phonate at his or her softest and loudest volumes, which creates a mapping for dB SPL from the neck sensor, a cutoff for the lowest detectable dB SPL, and settings for voicing onset/offset (personal communication, Griffin Labs, January 2014). It is important to point out that the relationship between the amplitude of the neck-surface ACC signal and vocal SPL measured with a microphone in front of the mouth can be influenced by a variety of factors (e.g., glottal conditions, mouth opening) that reduce the accuracy of this approach as compared to the VoxLog’s direct use of a microphone. For example, /a/ and /m/ can have similar ACC amplitude (RMS) values because the ACC is recording neck skin vibrations during voicing but the dB SPL recorded from a microphone at the lips will be different (6–10 dB SPL) due to an open versus closed mouth ($\text{vec et al., 2005}$).

Table 2. Important Differences Regarding Monitoring Parameters Between the Three Devices Are Summarized Above.

<table>
<thead>
<tr>
<th>Variables</th>
<th>APM</th>
<th>VoxLog</th>
<th>VocaLog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity (dB SPL)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Biofeedback</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Pitch (F0)</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Biofeedback</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Percent Phonation (%)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Environmental noise (dB SPL)</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Cycle Dose (cycles per day)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance Dose (meters per day)</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>dB SPL Calibration</td>
<td>Daily</td>
<td>None</td>
<td>Once</td>
</tr>
</tbody>
</table>

All three devices offer real-time ambulatory biofeedback for loudness (SPL) while the APM and VoxLog provide this for pitch (using F0) as well. Vibrotactile biofeedback is delivered by all devices when the parameter being targeted for feedback goes above or below an adjustable threshold. The vibrotactile feedback is delivered in the APM system by a separate belt-worn vibrator that is attached by a wire to the system box, whereas the system box itself vibrates to provide feedback for VoxLog and VocaLog devices. The VocaLog also provides the option to deliver auditory feedback by audible beeps from the system box. All devices have adjustable settings for
how long a subject must be above or below a threshold before they will receive biofeedback and the APM and VoxLog have adjustable settings for how long the vibrotactile cue will last. Additionally, the VoxLog and VocaLog have adjustable lengths of delay for a cue after the subject has triggered biofeedback.

According to the motor learning and control literature, being able to adjust the way biofeedback is delivered (e.g., frequency, timing) beyond simply having it triggered every time (100% feedback) a threshold is exceeded is necessary to enhance retention of a new motor behavior (c.f. Schmidt & Lee, 2011). For simple examples, frequency of feedback has to be considered, since cueing too often may result in subject annoyance. How fast feedback is triggered plays a role, since cueing too quickly may limit a subject’s ability to learn from the cueing (Swinnen, Schmidt, Nicholson, & Shapiro, 1990). In addition, how long the cue lasts may influence behavior change since a very short vibrotactile cue may not be noticed. As the capability to provide automated ambulatory biofeedback for voice is new, there is a need for research to determine the feedback frequency and timing parameters that are most effective. This will probably include the need to expand the feedback capabilities of ambulatory devices since they are currently quite limited.

Finally, the user interfaces for analyzing the data and displaying the results differ between devices. Because ambulatory monitoring and biofeedback is typically done for the entire day, it is not unusual to create files containing over 1 million data points. All three devices attach to a computer to download data files and generate summary statistics and graphs of the results using each system’s proprietary software. Also, the software interfaces for all 3 devices have the capability to export ambulatory monitoring data into other formats such as Microsoft Excel, Access, or simple text (.txt) and/or comma-separated value (.csv) files so that other statistical analysis approaches can be used. Summary statistics (means, modes, standard deviations) are provided for the primary measures generated by each of the three systems. In addition to overall summary statistics, all three devices provide time series plots of dB SPL data, allow the clinician to compare multiple monitoring days, and provide an overall “percent compliance” during dB SPL biofeedback days (i.e., percentage of voiced frames that the patient stayed below/above a dB SPL threshold). The APM and VoxLog software additionally permit the display of F0 data over time, dB SPL and F0 histograms, a scatterplot of dB SPL vs F0, and percent compliance during F0 biofeedback days. The APM’s software provides even more data processing options, such as the ability to highlight specific bandwidths in dB SPL histograms, F0 histograms, or dB SPL/F0 scatterplots. Furthermore, the APM software can offer more processing capabilities for biofeedback such as post-monitoring application of biofeedback thresholds for simulation and prediction purposes as well as modifications of biofeedback thresholds after the patient has completed a day of biofeedback to troubleshoot potential future consequences of threshold changes.

**Quantitative Comparison of Performance Characteristics**

Five ambulatory monitors (the APM, VHM, VocaLog, and two VoxLogs) were worn simultaneously while a speaker gave a 90-minute lecture (Figure 1). One VoxLog collar (inferior in Figure 1) was connected to a VoxLog system box and the other VoxLog collar (superior in Figure 1) was connected to a digital stereo voice recorder (Zoom H1) that was being used to collect data for a different study. A new Smartphone-based research system (not commercially available) named the Voice Health Monitor (VHM; Mehta, Zañartu, Feng, Cheyne, & Hillman, 2012) was used to record the raw ACC signal at 16 bits and 11025 samples per second so that it could be used as a reference (ground truth) for comparison purposes. The VocaLog (1 second analysis window) and APM (50 ms analysis window) were calibrated for SPL by having the subject perform a loudness glide (i.e., crescendo from soft to loud) during a sustained vowel. The VoxLog, which does not require calibration, was set to its shortest analysis window of 100 ms. A picture of how the sensors were worn during data collection is shown in Figure 1. The VoxLog sensor was taped
to the neck (not pictured) since the subject’s sternocleidomastoids routinely lifted the sensor off his neck when rotating his head left and right (this design issue has been addressed by Griffin Labs in the upcoming VocaLog2). Otherwise, all of the sensors were stable throughout the entire recording.

Figure 1. Subject Wearing Five Voice Monitors Labeled in the Photo.

The data files were aligned in post-processing using the first voiced frame at the start of the lecture. The raw ACC reference signal recorded with the VHM was post-processed using customized MATLAB scripts that mimicked the real-time processing performed by the APM so that its results could be derived and displayed using the APM software interface. Comparable upper and lower limits (filter cutoffs) were set for the data on all four devices to ensure consistency.

Table 3 shows a comparison of the overall values that were obtained for the four devices. There was good general agreement between devices for common measures of F0 (+2 Hz) and SPL (+3 dB). The only large discrepancy was for percent phonation where the VocaLog produced a value that was approximately twice as large as the values produced by the other three devices.

The largest discrepancy in SPL was displayed by the VoxLog, which can probably be attributed its shorter mouth-to-microphone distance (10 cm) as compared to the microphone distances used to calibrate the APM (15 cm) and VocaLog (50 cm). What is not readily apparent is why the SPL level reported by the VocaLog is not much lower than the values produced by the other devices, which would be expected based on the much longer mouth-to-microphone distance of 50 cm used to calibrate the VocaLog.

The large discrepancy (overestimation) in the percent phonation time reported by the VocaLog can be readily attributed to its use of a much longer analysis window as compared to

<table>
<thead>
<tr>
<th>Monitoring Device</th>
<th>Average dB SPL</th>
<th>Mode dB SPL</th>
<th>Average F0 (Hz)</th>
<th>Mode F0 (Hz)</th>
<th>Total Phonation Time (mm:ss:milliseconds)</th>
<th>Percent Phonation (% of total phonation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APM</td>
<td>67.45</td>
<td>68</td>
<td>111.55</td>
<td>95</td>
<td>37:06:200</td>
<td>41.34</td>
</tr>
<tr>
<td>VHM</td>
<td>65.56</td>
<td>68</td>
<td>109.79</td>
<td>95</td>
<td>39:24:000</td>
<td>43.14</td>
</tr>
<tr>
<td>VocaLog</td>
<td>66.49</td>
<td>67</td>
<td>N/A</td>
<td>N/A</td>
<td>83:27:000</td>
<td>91.34</td>
</tr>
<tr>
<td>VoxLog</td>
<td>69.87</td>
<td>71</td>
<td>109.23</td>
<td>94</td>
<td>36:28:640</td>
<td>39.93</td>
</tr>
</tbody>
</table>

The table lists each ambulatory voice monitor and its resulting summary statistics.
the other devices. In essence, the VocaLog registers an entire 1-second analysis window as phonation if any voicing occurs during that time period, no matter how brief. In a similar vein, the VoxLog’s reporting of the smallest percent phonation is most likely due to it having the shortest, and therefore the most accurate analysis window for phonation detection (4 ms).

**Conclusion**

The recent commercial availability of devices for voice ambulatory monitoring and biofeedback offers voice clinicians and researchers new options for potentially improving the assessment, treatment, and investigation of voice disorders. The cost, operational features, and performance characteristics of the three available devices differ in ways that can significantly impact decisions about which one is best suited for a particular application. This technology, as well as the associated knowledge about how best to utilize it, is expected to evolve and improve in the future. In the meantime, the information provided in this initial direct comparison of current devices is intended help voice clinicians and researchers in choosing the device that best meets their needs.

**References**


