Duration of ambulatory monitoring needed to accurately estimate voice use

Daryush D. Mehta[1], *Rebecca Woodbury Listfield*[1], *Harold A. Cheyne II*[2], *James T. Heaton*[1], *Shengran W. Feng*[1], *Matías Zañartu*[3], *Robert E. Hillman*[1]

[1]Ctr. for Laryngeal Surgery & Voice Rehab., Mass. General Hospital, Boston, Massachusetts, USA
[2]Lab of Ornithology, Cornell University, Ithaca, New York, USA
[3]Dept. of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

```
daryush.mehta@alum.mit.edu, woodbury@alum.mit.edu, hac68@cornell.edu,
james.heaton@mgh.harvard.edu, willfeng@mit.edu, matias.zanartu@usm.cl,
hillman.robert@mgh.harvard.edu
```

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 into establishing 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 decreased after 26 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.

Index Terms: ambulatory voice monitoring, voice use, voice disorders, accelerometer

1. Introduction

The most common voice disorders are chronic or recurring conditions that are likely to result from faulty and/or abusive vocal behavior patterns [1, 2]. Since voice use is considered to play a major role in the etiology of many voice disorders, clinicians focus a great deal of attention on attempting to evaluate and modify how patients typically use their voices. Such efforts currently rely on patients self-reporting their perception of their own voice use. The subjective nature of self-reporting brings about uncertainty and reliability issues, particularly given that patterns of voice use (and misuse) become highly habituated and somewhat automatic and can conceivably be carried out below an individual's threshold of consciousness. This state of affairs underlies the fact that there is a lack of robust evidence that establishes the actual role of voice use in the etiology of voice disorders, as well as a paucity of objective information about what constitutes normal and/or healthy levels of daily voice use.

The long-standing need for computing objective measures of voice use has promoted several attempts to develop wearable devices for daylong ambulatory monitoring of voice use [3–8]. Additional research and development activity [9] produced the Ambulatory Phonation Monitor (APM) that is commercially available for clinical and research use (model 3200, KayPENTAX, Montvale, NJ). The APM device places a miniature accelerometer on the neck just above the sternal notch to sense phonation-induced vibrations of the skin. The accelerometer signal is processed in real time to estimate measures of fundamental frequency, sound pressure level, and phonation time for up to 14 hours per day [9–11].

Voice monitoring devices using neck-placed accelerometers have several advantages over systems that use an acoustic microphone as a monitoring device. Accelerometers are less susceptible to typical environmental noise or supraglottal vocal tract resonances, are unobtrusive to wear, and preserve the subject's and others' privacy because the acoustic speech signal is not recorded [12]. Guidelines, however, for the clinical use of the APM have yet to be established. In particular, the monitoring durations required to produce accurate estimates of voice use parameters are unknown.

This paper describes a pilot study that begins to investigate how long a subject must be monitored to adequately estimate long-term average values for phonatory parameters of interest. Shorter monitoring durations are associated with lower costs and less inconvenience to the subject and clinical researcher; however, shorter monitoring sessions are prone to larger errors due to the possibility of misrepresenting an individual's habitual vocal profile. This study provides initial benchmarks for quantifying the error in the estimation of several vocal parameters (fundamental frequency, sound pressure level, phonation time, cycle dose, and distance dose [13]) as a function of monitoring duration. Such data do not exist for various professions to provide evidence to investigate theories relating the impact of long-term vocal loading on voice use–related pathologies.

2. Methods

2.1. Subjects

The following three types of subjects were included in this study: patients with voice disorders ($n = 6$); control subjects matched to the patients in terms of age, gender, and occupation ($n = 6$); and low voice users ($n = 6$). All subjects underwent laryngeal endoscopic examination by an experienced clinician to confirm the presence (in patients) or absence (in controls and low voice users) of laryngeal pathology prior to subject participation and group assignment.

Table 1. *Participant characteristics. Subjects appearing in the same row were matched for sex and approximate age.*

Sex	Patient ID, Age, Occupation	Control ID, Age, Occupation	Low Voice User ID, Age, Occupation
F	P1, 37 years, Teacher	C1, 49 years, Teacher	—
F	P2, 22 years, Salesperson	C2, 23 years, Salesperson	—
F	P3, 46 years, Clinician	C3, 52 years, Clinician	LVU3, 52 years, Researcher (Histologist)
F	P4, 54 years, Teacher	C4, 62 years, Teacher	LVU4, 59 years, Researcher (Histologist)
M	P5, 19 years, Singer	C5, 20 years, Singer	LVU5, 29 years, Researcher (Engineer)
F	P6, 30 years, Clinician	C6, 26 years, Clinician	LVU6, 29 years, Animal Care Technician
M	—	—	LVU1, 31 years, Researcher (Physiologist)
M	—	—	LVU2, 48 years, Researcher (Engineer)

Participant characteristics are listed in Table 1. The voice patients were recruited from the clinical population of a voice center. Patients had mild, moderate, or severe dysphonia and had occupations requiring substantial voice use. Each patient was asked to identify a colleague in his or her profession of the same sex and approximately the same age to serve as a corresponding control participant. Subjects in the low voice use group had occupations that did not require a substantial amount of talking on a daily basis and were recruited from the research staff in laboratory settings.

2.2. Data collection

Each subject was monitored for five work days, yielding 30–70 hours of monitoring data per subject. Subjects were instructed not to alter their normal voice use while being monitored. At the beginning of each day of data collection, the participant was fitted with the APM at either the voice center or at the participant's place of work.

In order to relate calibrate the APM's accelerometer level to sound pressure level (SPL) during the data analysis process, speech was recorded using a handheld microphone (model SM48, Shure Inc., Niles, IL) positioned 15 cm from the lips. One end of a spacing rod was fixed to the microphone and the other end was held against the skin just below the nose. The microphone signal was calibrated by an artificial larynx (Cooper-Rand, Luminaud, Inc., Mentor, OH) that produced a periodic tone complex at the participant's closed lips. SPL was measured at the microphone with a calibrated sound level meter (model NL-20, Rion, Japan). Skin acceleration was sensed using a miniature 1-axis accelerometer (model BU-7135, Knowles Corp., Itasca, IL) that was attached to the front of the neck just above the sternal notch using a medical grade adhesive (Skin-Bond, Smith & Nephew, London, UK; or model B-401, Factor II, Inc., Lakeside, AZ). The microphone was detached from the APM unit after calibration, and data collection proceeded throughout the subject's work day.

2.3. Data Analysis

The APM performed online processing of the accelerometer signal to yield and store a sound intensity and fundamental frequency estimate every 125 ms. To determine if a particular segment was voiced, each 125-ms frame (rectangular window) was divided into five 25-ms subintervals. The root-mean-square (RMS) value for each subinterval was calculated. If the RMS values of two contiguous subintervals exceeded a preset voicing threshold, the entire frame was designated as voiced.

If the frame was designated as voiced, then that frame's intensity was set by a linear relation between accelerometer signal magnitude and SPL using the microphone signals recorded at the start of each day's data collection. Then the frame's fundamental frequency (Hz) was set to the reciprocal of the largest peak in the frame's autocorrelation function [9].

Phonation duration, fundamental frequency, and intensity calculations were performed for each frame only if the following criteria were met: (a) the frame was voiced, (b) intensity was between 50 dB SPL and 130 dB SPL, and (c) fundamental frequency was between 70 Hz and 400 Hz. Otherwise, a zero was recorded in each unvoiced/non-speed frame. The parameter limits were chosen based on similar limits used in the commercial APM application and a widely used pitch-tracking software program (Computerized Speech Lab, KayPENTAX). The limits were imposed to eliminate spurious values resulting from two factors: (1) the linear acceleration-to-SPL calibration not being applicable at high SPL and (2) the prevalence of errors estimated in the microphone data.

The fundamental frequency and SPL values for each frame were used to calculate cycle dose and distance dose, which estimate the total number of vocal fold oscillations and the total distance traveled by the vocal folds, respectively [13]. For proper comparison with SPL values used for these dose measures, which assumed a mouth-to-microphone distance of 50 cm, the SPL calibration factor, which was performed at 15 cm in the current study, was decreased by approximately 10.5 dB.

For each hour of monitoring, the total phonation time, cycle dose, and distance dose, as well as the average fundamental frequency and sound pressure level, were estimated from the accelerometer signal [13]. Data analysis was performed using a cumulative average technique, similar to that used previously to measure speech tokens [14]. From the hourly raw parameter data, cumulative parameter averages were calculated for each parameter. These cumulative averages were calculated such that the cumulative average for the first hour is equal to the parameter value for the first hour, the cumulative average for the second hour is equal to the average of the parameter values for the first and second hour, and so on.

We define a complete sample as 40 hours of monitoring (a full workweek), such that each parameter's cumulative average served as the true value of that parameter against which shorter-duration accumulations are compared. Only subjects who were monitored for a minimum of 40 hours (3 patients, 4 controls, and 3 low voice users) were included; the other subject data were shorter due to equipment issues and/or non-compliance with monitoring instructions. For each subject, the percent error in estimating a given parameter at each hour was calculated by taking the absolute difference between the cumulative average up to that hour (estimated value) and the average after 40 hours of monitoring (true value), divided by the true value.
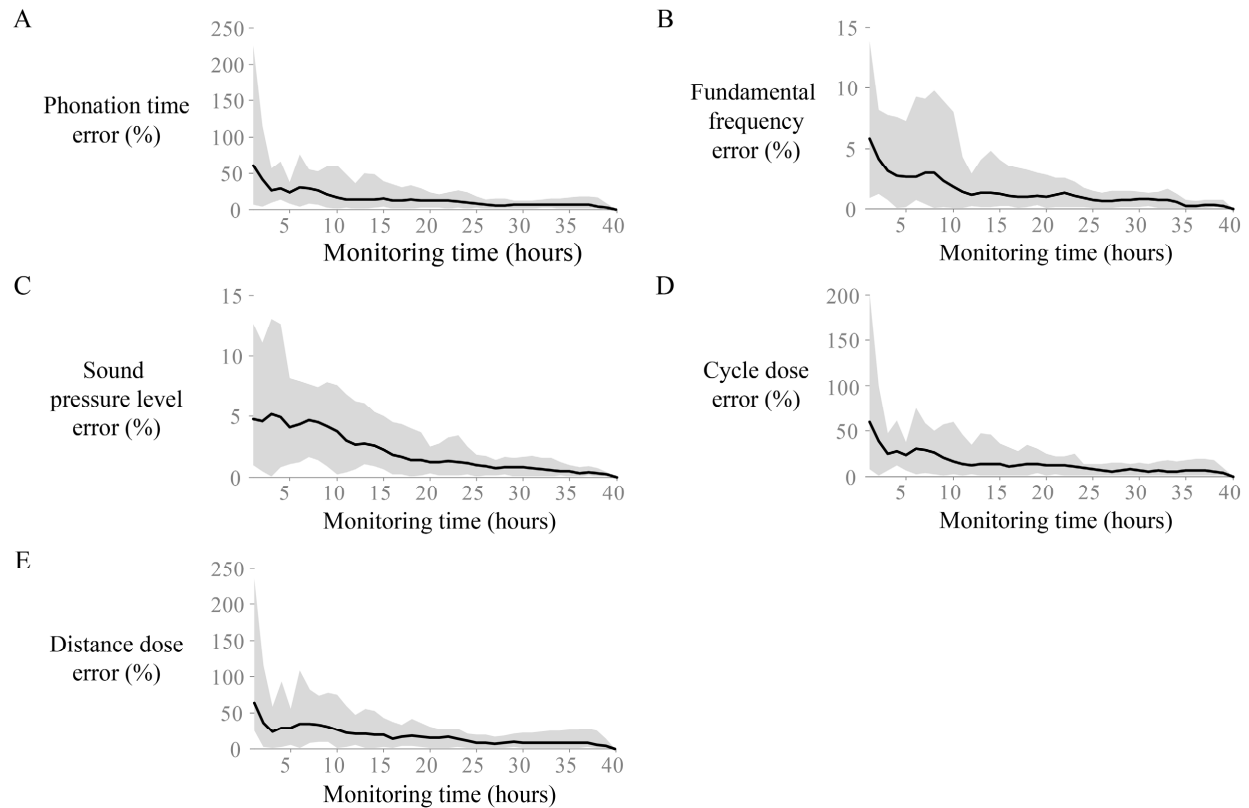
Figure 1: *Percent error across subjects monitored for at least 40 hours (n = 10) as a function of monitoring time in estimating (A) phonation time, (B) fundamental frequency, (C) sound pressure level, (D) cycle dose, and (E) distance dose. Dark lines represent the mean error, and gray shading indicates the range encompassing the minimum and maximum error.*

3. Results

Unexpectedly, there were no statistically significant differences among the average error curves for subjects within and across each subject group. Figure 1 thus shows the average, minimum, and maximum error at the end of each hour across the entire group of ten subjects.

Estimates of fundamental frequency and sound pressure level converged rapidly to the overall average value. The average fundamental frequency error decreased to approximately 1 % after 12 hours of monitoring time, with a worst-case scenario still exhibiting less than 5 % error. The average SPL curve similarly exhibits an error of about 5 % after only one hour and an average error of approximately 1 % after 20 hours of monitoring time.

For estimating phonation time, an average error of approximately 5 % was achieved after 26 hours with much of the error variation decreased after this duration of monitoring. Absolute errors for phonation time for one individual reached 18 % even after prolonged monitoring. Even after several hours of recording, the average error remained above 25 %, although certain subjects exhibited both extremes (accurate and inaccurate) estimation of phonation time across the monitoring time.

Average errors related to cumulative averages of cycle dose and distance dose were greater than 10 % until the 26-hour mark. As seen from the minimum and maximum error bounds of the graphs, though, individual subject behavior can yield varying levels of error over the entire 40-hour workweek. Whereas some individuals yielded measures that converged rapidly to the overall average in a matter of hours, other subjects exhibited errors up to 19 % for cycle dose and 26 % for distance dose after prolonged monitoring times.

4. Discussion

The error in estimating parameters as a function of APM monitoring time, as determined in this study, can be used as initial benchmarks to determine an optimal monitoring duration that yields the desired level of accuracy for a specific parameter of interest. The total monitoring duration can thus be adjusted depending on the desired level of accuracy. Reducing the monitoring time has the potential to reduce the cost and inconvenience associated with prolonged monitoring.

For estimating fundamental frequency and sound pressure level, the data provide initial evidence that satisfactory estimates can be obtained after only a few hours. Average errors decrease to about 1 % after 20 hours of ambulatory monitoring. In contrast, errors associated with phonation, cycle dose, and distance dose were higher over the first several hours, requiring at least 26 hours of data to yield errors below 10 %. It is acknowledged that the current data portray a small sample size and that speaker-to-speaker variation in phonatory behavior from day to day may confound notions of overall parameter averages.

If the hourly data were treated as a random process with a normal distribution, statistical theory dictates that the running average of the process approaches the true average depending on the variance of the data and the number of values averaged. The results of the current study reflect the expected trend of an increase in the estimation accuracy of a parameter over time, with the largest gains occurring over the first several hours.

The APM currently includes a biofeedback tool, which provides vibro-tactile feedback when certain vocal parameters are above or below a certain threshold. The approximate error in the parameter estimates can be incorporated into these targets for behavioral modification via biofeedback. Additional parameters and analysis methods, however, may prove to be more salient in tracking certain types of voice disorders.

Figure 2 displays a picture of an ambulatory voice monitor 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 the larger APM-like systems [15]. A critical improvement to the APM is the collection of the raw waveform from the accelerometer that allows for the investigation of alternative analysis algorithms to, e.g., derive glottal airflow-based measures [16] and employ pattern recognition techniques to reveal potentially abusive vocal patterns [17]. Large storage space on the smartphone enables the collection of the raw accelerometer signal for more than just a 40-hour workweek.

5. 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 parameters were estimated and reported in this study, highlighting the different error curves for each parameter. These results provide initial benchmarks to determine the monitoring time necessary to yield the desired accuracy of parameter measurements. Future work calls for gathering data on a larger subject sample and implementing a next-generation ambulatory voice monitor to improve the user experience and characterize individual patterns of vocal behavior (including vocal recovery times, etc.).

6. References

[1] Aronson, A., Clinical voice disorders, 2nd ed. New York: Thieme Medical Publishers, 1985.

[2] Boone, D. R. and McFarlane, S. C., The voice and voice disorders. Englewood Cliffs: Prentice Hall, 1988.

[3] Airo, E., Olkinuora, P., and Sala, E., "A method to measure speaking time and speech sound pressure level," Folia Phoniatr. Logop., 52(6):275–288, 2000.

[4] Buekers, R., Bierens, E., Kingma, H., and Marres, E., "Vocal load as measured by the voice accumulator," Folia Phoniatr. Logop., 47(5):252–261, 1995.

[5] Ohlsson, A. C., Brink, O., and Löfqvist, A., "A voice accumulation—validation and application," J. Speech Hear. Res., 32(2):451–457, 1989.

[6] Ryu, S., Komiyama, S., Kannae, S., and Watanabe, H., "A newly devised speech accumulator," ORL. J. Otorhinolaryngol. Relat. Spec., 45(2):108–114, 1983.

[7] Watanabe, H., Komiyama, S., Ryu, S., and Kannae, S., "A newly designed automatic phonometer," Agressologie, 25(9):1035–1037, 1984.

Figure 2: *Ambulatory voice monitor currently under development that utilizes a smartphone as a data acquisition device, a similar accelerometer assembly for sensing neck-skin vibration, and an interface circuit to provide power to the accelerometer and impedance matching with the smartphone.*

[8] Popolo, P. S., Švec, J. G., and Titze, I. R., "Adaptation of a pocket pc for use as a wearable voice dosimeter," J. Speech. Lang. Hear. Res., 48(4):780–791, 2005.

[9] Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., and Hillman, R. E., "Development and testing of a portable vocal accumulator," J. Speech. Lang. Hear. Res., 46(6):1457–1467, 2003.

[10] Cheyne, H. A., "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Proceedings of the 3rd IEEE-EMBS International Summer School and Symposium on Medical Devices and Biosensors:118–121, 2006.

[11] Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., and Cheyne, H. A., "Ambulatory monitoring of disordered voices," Ann. Otol. Rhinol. Laryngol., 115(11):795–801, 2006.

[12] Zañartu, M., Ho, J. C., Kraman, S. S., Pasterkamp, H., Huber, J. E., and Wodicka, G. R., "Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface," IEEE Trans. Biomed. Eng., 56(2):443–451, 2009.

[13] Titze, I. R., Švec, J. G., and Popolo, P. S., "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," J. Speech. Lang. Hear. Res., 46(4):919–932, 2003.

[14] Karnell, M. P., "Laryngeal perturbation analysis: Minimum length of analysis window," J. Speech Hear. Res., 34(3):544–548, 1991.

[15] Hunter, E. J., "Teacher response to ambulatory monitoring of voice," Logoped. Phoniatr. Vocol., http://informahealthcare.com/doi/abs/10.3109/14015439.2012.6646 57, 2012.

[16] Zañartu, M., "Acoustic coupling in phonation and its effect on inverse filtering of oral airflow and neck surface acceleration," PhD thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 2010.

[17] Syed, Z., Guttag, J., and Stultz, C., "Clustering and symbolic analysis of cardiovascular signals: Discovery and visualization of medically relevant patterns in long-term data using limited prior knowledge," EURASIP J. Adv. Sig. Process.:1–16, 2007.