Mobile Voice Health Monitoring Using a Wearable Accelerometer Sensor and a Smartphone Platform

Daryush D. Mehta*, *Member, IEEE*, Matías Zañartu, *Member, IEEE*, Shengran W. Feng, *Student Member, IEEE*, Harold A. Cheyne II, and Robert E. Hillman

Abstract—Many common voice disorders are chronic or recurring conditions that are likely to result from faulty and/or abusive patterns of vocal behavior, referred to generically as vocal hyperfunction. An ongoing goal in clinical voice assessment is the development and use of noninvasively derived measures to quantify and track the daily status of vocal hyperfunction so that the diagnosis and treatment of such behaviorally based voice disorders can be improved. This paper reports on the development of a new, versatile, and cost-effective clinical tool for mobile voice monitoring that acquires the high-bandwidth signal from an accelerometer sensor placed on the neck skin above the collarbone. Using a smartphone as the data acquisition platform, the prototype device provides a user-friendly interface for voice use monitoring, daily sensor calibration, and periodic alert capabilities. Pilot data are reported from three vocally normal speakers and three subjects with voice disorders to demonstrate the potential of the device to yield standard measures of fundamental frequency and sound pressure level and model-based glottal airflow properties. The smartphone-based platform enables future clinical studies for the identification of the best set of measures for differentiating between normal and hyperfunctional patterns of voice use.

Index Terms—Accelerometer sensor, vocal hyperfunction, voice production model, voice use, wearable voice sensor.

*D. D. Mehta is with the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114 USA, with the Department of Surgery, Harvard Medical School, Boston, MA 02115 USA, with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA, and also with the Department of Otolaryngology, The University of Melbourne, Parkville, Victoria 3010, Australia (e-mail: daryush.mehta@alum.mit.edu).

M. Zañartu is with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso 1680, Chile (e-mail: matias.zanartu@usm.cl).

S. W. Feng is with the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114 USA, and also with the Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: willfeng@mit.edu).

H. A. Cheyne II is with the Bioacoustic Research Program, Laboratory of Ornithology, Cornell University, Ithaca, NY 14853 USA (e-mail: hac68@cornell.edu).

R. E. Hillman is with the Center for Laryngeal Surgery and Voice Rehabilitation and Institute of Health Professions, Massachusetts General Hospital, Boston, MA 02114 USA, and also with Surgery and Health Sciences and Technology, Harvard Medical School, Boston, MA 02115 USA (e-mail: hillman.robert@mgh.harvard.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2012.2207896

I. INTRODUCTION

VOICE disorders affect approximately 6.6% of the adult population in the United States at any given point in time [1]. Many common voice disorders are chronic or recurring conditions that are likely to result from faulty and/or abusive patterns of vocal behavior referred to generically as *vocal hyperfunction* [2]. These behaviorally based voice disorders can be especially difficult to assess accurately in the clinical setting and potentially could be much better characterized by long-term ambulatory voice monitoring as individuals engage in their typical daily activities.

Thus, an ongoing goal in clinical voice assessment is the development and use of measures derived from noninvasive procedures to quantify behaviorally based voice disorders. Common hyperfunction-related voice disorders [2], [3] are believed to arise from a variety of voice use–related patterns that include excessive loudness, inappropriate pitch, reduced vocal efficiency, and speaking for excessive durations of time [4]. Clinicians currently rely on a patient's self-assessment and self-monitoring to evaluate the prevalence and persistence of vocally abusive behaviors during medical diagnosis and management. Patients, however, tend to be highly subjective and prone to unreliable judgments of their own voice use. Low reliability of self-reported data is most likely due to patterns of voice use and misuse becoming habituated and therefore carried out below an individual's level of consciousness.

Wearable voice monitoring systems seek to provide more reliable and objective measures of voice use. Treatment of hyperfunctional disorders would be greatly enhanced by the ability to unobtrusively monitor and quantify detrimental behaviors and, ultimately, to provide real-time feedback that could facilitate healthier vocal function.

A. Current-Generation Technologies

The general concept of mobile monitoring of daily voice use has motivated several early attempts to develop speech and voice "accumulators," e.g., [5]–[7]. Our group [8], [9] and other investigators [10] have shown that a particular miniature accelerometer (model BU-27135; Knowles Corp., Itasca, IL) currently offers the best potential as a phonation sensor for long-term monitoring of vocal function because it 1) can be worn unobtrusively at the base of the neck above the collarbone; 2) has sensitivity and dynamic range characteristics that make it well suited to sensing vocal fold vibration; 3) is relatively immune to external environmental sounds; and 4) produces a voice-related signal that is not filtered by vocal tract resonances, which makes

subglottal neck–skin acceleration easier to process, more robust for sensing disordered phonation, and alleviates confidentiality concerns associated with the acoustic speech signal [11].

Much of the basic technology for these current-generation systems was developed when limitations of digital memory technology forced the systems to only store frame-based estimates of vocal parameters such as fundamental frequency (f0) and sound pressure level (SPL) instead of the raw accelerometer signal. One system employs a personal digital assistant as the data acquisition platform [10] and has been used to produce valuable information about voice used by teachers [12], [13] and has provided a test bed for developing theoretical concepts about vocal dose measures [14], [15].

Our group developed a comparable system that culminated in the Ambulatory Phonation Monitor (APM Model 3200; KayPENTAX, Montvale, NJ), the first commercially available device for research and clinical use. Although accelerometer-based monitoring has produced valuable insight into daily patterns of voice use, and enthusiasm remains high for its potential to improve the diagnosis and treatment of voice disorders, its adoption into clinical practice has been limited due to 1) restrictions on measures and analysis algorithms, 2) the relatively high cost due to development costs and specialized circuitry, and 3) the lack of large-sample studies to determine the diagnostic capabilities of accelerometer-based voice measures.

B. Outline

The purpose of this paper is to report on our development of a versatile and cost-effective clinical tool for mobile voice monitoring that acquires the high-bandwidth accelerometer signal using a smartphone as the data acquisition platform. In Section II, we first describe the hardware and software components of the smartphone-based system. Section III follows with a description of two major signal processing approaches—traditional ambulatory voice measures (f0, SPL, phonation time) and novel model-based estimates of glottal airflow properties. In Section IV, we present pilot data from three vocally normal speakers and three individuals with voice disorders who each wore the device for 7 days. Finally, Section V concludes with a brief summary and discussion.

II. PLATFORM FOR A MOBILE VOICE MONITORING SYSTEM

The developed ambulatory monitoring system overcomes storage-related limitations of current-generation technologies and records high-bandwidth sensor data for over 18 h per day (with daily system charging) for at least 7 days before it is necessary to download data or swap out removable memory. This section describes the design specifications for hardware and software components of the device to be largely compatible with new generations of mobile device architecture. Fig. 1 highlights components of the proposed voice monitoring system.

A. Neck-Placed Miniature Accelerometer

The selected miniature accelerometer senses acceleration in 1-D and has a vibration sensitivity suitable for obtaining meaningful information about the voice. The sensor is calibrated using
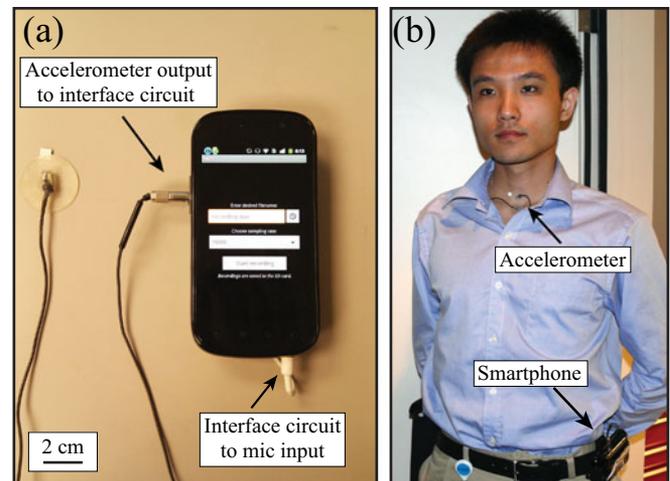

Fig. 1. Mobile voice monitoring system: (a) Smartphone and accelerometer input and (b) subject wearing the system with wiring underneath clothing and smartphone in belt holster.

a laser Doppler vibrometer system that outputs a chirp signal to a mechanical shaker. With no casing or mounting, the accelerometer sensitivity is approximately $88.5 \text{ dB} \pm 3 \text{ dB re } 1 \text{ cm/s}^2/\text{V}$ in the 100–3000 Hz frequency range. The accelerometer is wired to a three-conductor cable and mounted on a flexible silicone pad with a durable silicone sealant. A plastic cover around the accelerometer provides attenuation of undesirable mechanical disturbances such as wind noise and clothing contact.

The subject affixes the accelerometer assembly to his or her neck using hypoallergenic double-sided tape (Model 2181, 3M, Maplewood, MN). The tape's circular shape and small tab allow for easy placement and removal of the accelerometer on the neck skin a few centimeters above the suprasternal notch. The tape is strong enough to hold the silicone pad in place during a full day for the typical user.

B. Smartphone Platform

Acquiring a commercially available data acquisition device ensures professional-level support/upgrades for the device and a built-in comfort level with the technology by users. We selected the Nexus S smartphone manufactured by Samsung with Google's Android operating system installed. The Nexus S offers a commercially available device that can be used, in its most basic form, as a data logger, and, utilizing advanced features, as a comprehensive tool for long-term recording and user interactivity. The open source nature of the Java-derived Android operating system gives a high level of software control over device features. In addition to routines associated with controlling the data acquisition process, the programming environment enables the development of interactive applications for gathering subject self-report data, prompting daily calibration protocols, and verifying signal integrity.

C. Interface Circuit

The smartphone provides power to and acquires the accelerometer signal using the microphone channel. The accelerometer requires at least 1.5 V, which is satisfied by the 2-V

bias signal on the microphone channel. The interface circuit provides appropriate impedance matching so that the accelerometer acts as a typical handsfree microphone. Although signal bandwidth and power requirements warrant a wired setup, future device designs could consider wireless interfaces with smaller footprints.

The interface circuit is designed to be simple, consisting of an inverting operational amplifier configuration with passive components. The circuit couples the accelerometer pinout (power, ground, signal) to the smartphone microphone channel (signal+bias, ground) and is housed in extra space available in the plastic cover of an extended-life battery (Mugen Power, Hong Kong, China). The final device packaging insulates the interface circuit from the user and provides an input jack for the accelerometer assembly. A three-pole subminiature connector (nanoCON; Neutrik AG, Schaan, Liechtenstein) provides a lockable coupling between the accelerometer and the interface circuit. Fig. 1(a) displays the connector fitting into the side of the battery cover.

D. Data Acquisition Specifications

Minimum device recording specifications to record the raw accelerometer signal include an 11 025 Hz sampling rate, 16-bit quantization, 50-dB dynamic range, and 9.5 GB of nonvolatile memory. These specifications satisfy the requirements of obtaining voice-related neck skin vibrations from quiet-to-shouting voiced sounds at frequencies up to 4000 Hz. The Nexus S comes with a 1500-mAh battery that is 6.5 mm in width. Due to the requirements for extended recording durations, the extended-life battery extends the output current capacity to 3900 mAh and is 16.5 mm in depth.

The Nexus S smartphone contains a high-fidelity audio codec (WM8994; Wolfson Microelectronics, Edinburgh, Scotland, U.K.) that encodes the accelerometer signal using sigma-delta modulation (128× oversampling). Sampling rates available include standard frequencies from 8000 Hz to 48 kHz. The Nexus S smartphone enables one mono (microphone) input of the two stereo channels available on the codec. Operating system root access allows control over driver settings related to highpass filters (to remove dc offsets and low-frequency noise) and programmable gain arrays (to maximize dynamic range).

Fig. 2 shows the power spectrum during a silence region (noise floor) and during a loud voiced segment of a recording taken from female subject P1. In this in-field recording, the average noise floor across the spectrum is -86.0 dB relative to full scale (dBFS). The first harmonic at 522 Hz is 83.7 dB above the noise floor, illustrating the adequacy of the system's dynamic range for voice-related activity. Alert methods available on the Nexus S smartphone include audible, visual, and vibrotactile cues. In the current implementation, we disable audible outputs and employ the built-in vibration alert capability to prompt users. A combination of these three alert methods could be chosen based on subject preference.

The recording duration is limited primarily by memory and battery life. At a sampling rate of 11 025 Hz and 16-bit quantization, 18 h of data per day for seven days equates to 9.3 GB of
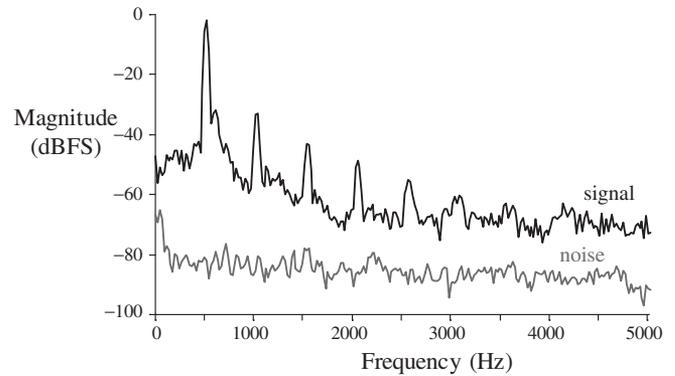

Fig. 2. Short-time power spectra of signal (black) and noise (gray) of the accelerometer data from female subject P1.

recorded data (GB $= 2^{30}$ bytes). The Nexus S smartphone has 16 GB of onboard flash memory, exceeding the required weekly allotment before needing to download data using the USB connection. In Section IV, a male subject (N2) was able to record over 18 h in one day.

E. User Experience

Fig. 1(b) illustrates the wearable nature of the mobile voice monitor. An Android application was coded to perform three basic system functions: 1) data acquisition, 2) sensor calibration, and 3) user prompting. The application creates a log file containing time-stamped records of when these basic system functions occur during the day.

Subjects enable acquisition of the accelerometer sensor signal at the start of their day. A level meter outputs the signal level on the screen. An option to pause the recording enables the subject to indicate that he or she is temporarily removing the sensor for various reasons that include taking a shower, swimming, or participating in high-impact sporting activities. Recording is stopped at the end of the day prior to sleeping.

The first time each day, the smartphone application prompts the user to perform a calibration routine in a quiet, minimally reverberant room. The calibration sequence seeks to match accelerometer signal levels to acoustic signal levels [16] that are simultaneously recorded by a handheld microphone (H1 Handy Recorder, Zoom Corporation, Tokyo, Japan). A custom metal standoff [see Fig. 3(a)] maintains a fixed distance of 15 cm between the subject's lips and the microphone. The noise floor of typical rooms is sufficient for such a calibration because the microphone is at a close distance and the accelerometer is largely insensitive to external acoustics. The application can be configured to periodically prompt subjects to respond to questions regarding their voice use during the day.

III. AMBULATORY MEASURES OF VOCAL FUNCTION

A. Traditional Measures of Voice Use

Traditional acoustic measures have been developed for quantifying the cumulative effects of voice use and rely on running estimates of phonation time, f0, and SPL from the raw signal
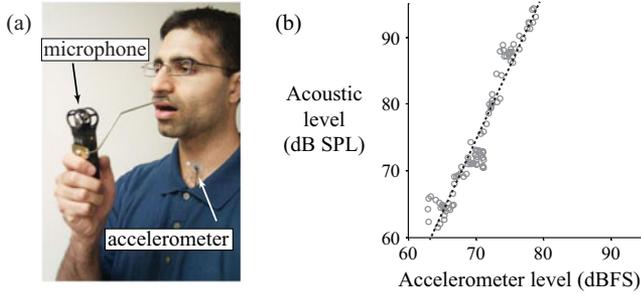
Fig. 3. Acoustic calibration procedure for a sustained vowel produced with increasing loudness. (a) Speaker holds the microphone at a set distance from the lips to derive a (b) linear regression (black line) between the accelerometer signal level (in dBFS; dB relative to full scale) and the acoustic level (in dB SPL). Gray circles indicate dB–dB levels from each 50-ms frame (rectangular window, no overlap).

obtained from the neck-mounted accelerometer sensor [8]. In this study, f0 and SPL are computed every 50 ms (nonoverlapping rectangular windows) for voiced frames. Each 50-ms frame is divided into two 25-ms subintervals, and the frame is considered voiced if the levels of both subintervals exceed 62 dB SPL. SPL is then re-computed over the entire frame duration. f0 for each voiced frame is equal to the reciprocal of the first peak location in the normalized autocorrelation function if the peak exceeds a threshold of 0.25. Otherwise, SPL and f0 values are set to zero for frames that are considered nonphonatory (containing either silence, unvoiced speech, or other nonspeech energy).

An estimate of the acoustic SPL is derived from the neck skin vibration level through the calibration routine that simultaneous records the accelerometer and acoustic speech signals [16]. Linear regression parameters are computed between frame-by-frame estimates of signal power during the production of sustained vowels of increasing loudness. Fig. 3(b) illustrates the regression performed in the dB–dB space to calibrate the accelerometer level in units of acoustic SPL.

Three vocal dose measures—phonation time, cycle dose, and distance dose [15]—are highlighted in the current study to quantify total daily voice use for each subject. Phonation time is the duration of voiced frames and is computed in terms of time (hours:minutes:seconds) and percentage of total time. The cycle dose is the number of vocal fold oscillations during a period of time. The distance dose estimates the total distance traveled by the vocal folds, combining cycle dose with estimates of vibratory amplitude based on SPL.

B. Model-Based Estimates of Time-Varying Airflow

Glottal airflow–based measures greatly enhance the potential of ambulatory monitoring to detect some of the features that have previously been shown to differentiate among hyperfunctionally related disorders and normal modes of vocal function [2], [3]. Robust automatic detection of these features, however, has yet to be achieved in an ambulatory setting. To derive such measures, a subglottal impedance-based inverse filtering (IBIF) technique is applied that expands upon previous work to produce glottal airflow measures from the neck surface acceleration signal [18].
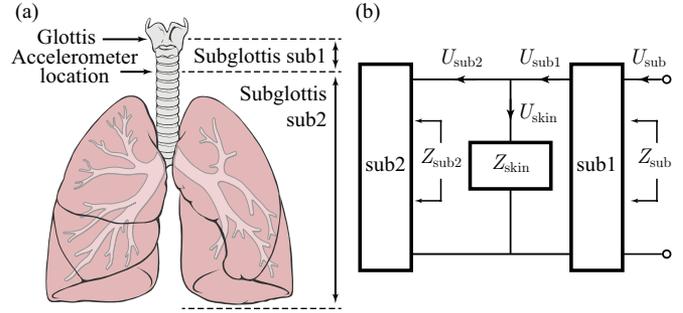

Fig. 4. Model of the subglottal system for impedance-based inverse filtering: (a) Anatomical diagram indicating the location of the accelerometer sensor on the skin surface approximately 5 cm below the glottis (figure adapted from [17]). (b) Mechanoacoustic one-port network indicating frequency-dependent velocities U and impedances Z of the subglottal system sub and subsections sub1 and sub2. Z_{skin} is the impedance due to the accelerometer sensor load and neck surface properties.

TABLE I
SUBJECT CHARACTERISTICS

Subject	Gender	Age	Occupation	Voice diagnosis
N1	male	24	researcher	normal
N2	male	30	researcher	normal
N3	male	59	researcher	normal
P1	female	18	student	hyperfunction
P2	female	22	student	nodules (post-surgery)
P3	female	63	retiree	laryngospasm

Subglottal IBIF is based on a biomechanical model linking mechanoacoustic analogies, transmission line principles, and physiological descriptions of voice production [19], [20]. The method follows a lumped-impedance parameter representation in the frequency domain that has been proven useful to model sound propagation in the subglottal tract [21], vocal tract [22], and boundaries experiencing source-filter interaction [23]. Time-domain techniques such as the wave reflection analog [24] and loop-equations [25] accomplish similar goals in numerical simulations, but are typically less attractive for the development of onboard signal processing schemes.

Fig. 4 illustrates the physiological model and one-port network topology of the frequency (ω)-domain subglottal IBIF model. The subglottal tract is divided into subglottal sections sub1 (between the glottis and accelerometer location) and sub2 (between the accelerometer location and the end of the bronchial branches). $Z_{\mathrm{skin}}(\omega) = Z_m(\omega) + Z_{\mathrm{rad}}(\omega)$, where the mechanical impedance of the neck skin $Z_m(\omega)$ is in series with the radiation impedance due to the accelerometer sensor load $Z_{\mathrm{rad}}(\omega)$. $Z_m(\omega)$ is defined as

$$Z_m(\omega) = R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) \tag{1}$$

where R_m, M_m, and K_m are the per-unit-area mechanical resistance, inertance, and stiffness of the neck skin, respectively. $Z_{\mathrm{rad}}(\omega)$ is defined as

$$Z_{\mathrm{rad}}(\omega) = \frac{j\omega M_{\mathrm{acc}}}{A_{\mathrm{acc}}} \tag{2}$$

where M_{acc} and A_{acc} are the per-unit-area mass and surface area, respectively, of the accelerometer sensor assembly.
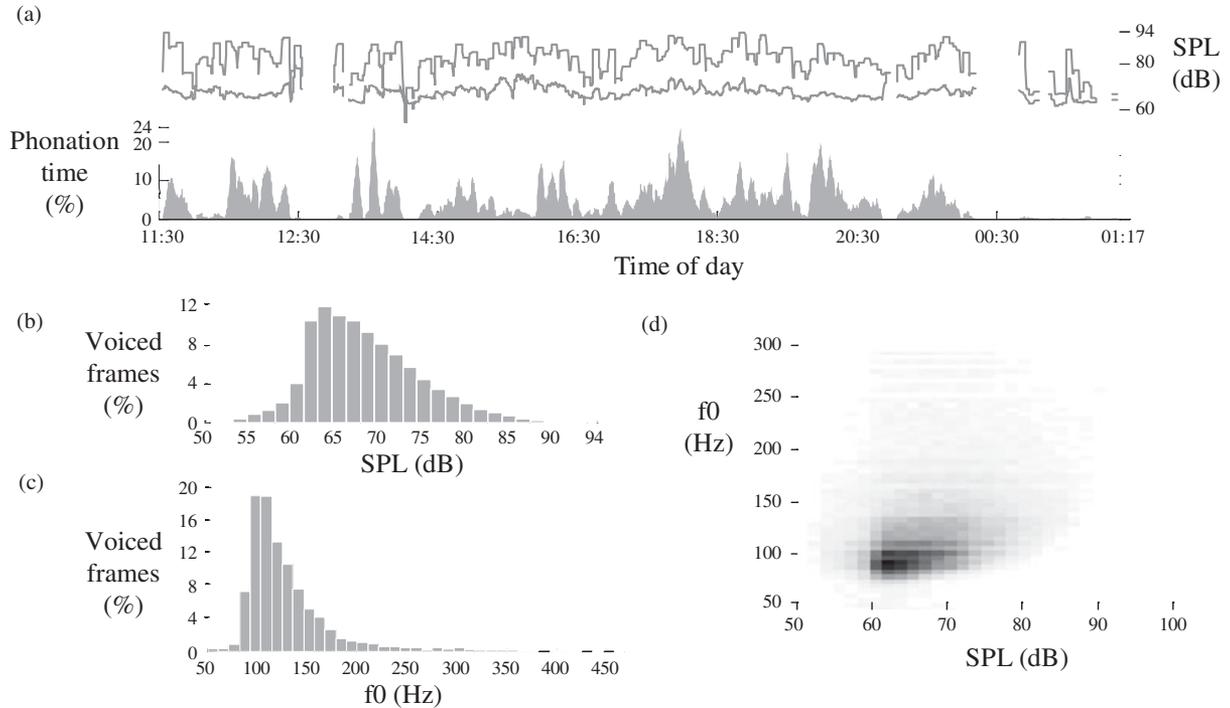
Fig. 5. Voice use profile of subject N2's first day (24-h time format). (a) 5-min moving average of phonation time (gray), average sound pressure level (SPL; lower line), and maximum SPL (upper line) for voiced frames; (b) histogram of fundamental frequency (f0); (c) histogram of SPL; and (d) phonation density showing the relative occurrence of particular combinations of SPL (horizontal axis) and f0 (vertical axis).

The input to the algorithm is the neck–skin acceleration $\dot{U}_{\text{skin}}(\omega)$ (the dot denotes the derivative operator), and the output is the inverse-filtered glottal airflow $U_g(\omega) = -U_{\text{sub}}(\omega)$. The equation relating input and output signals is expressed as

$$U_g(\omega) = -\dot{U}_{\text{skin}}(\omega) \frac{Z_{\text{sub2}}(\omega) + Z_{\text{skin}}(\omega)}{j\omega \cdot H_{\text{sub1}}(\omega) \cdot Z_{\text{sub2}}(\omega)} \qquad (3)$$

where $H_{\text{sub1}}(\omega) = U_{\text{sub1}}(\omega)/U_{\text{sub}}(\omega)$ is the transfer function of subglottal section sub1.

The parameters of $Z_{\text{skin}}(\omega)$ are considered fixed for each speaker and are derived by optimizing measures of $U_g(\omega)$ derived from $\dot{U}_{\text{skin}}(\omega)$ to a reference glottal airflow derived using a circumferentially vented pneumotachograph mask system (model MA-1L; Glottal Enterprises, Syracuse, NY). Optimization of the static model parameters (R_m, M_m, K_m, M_{acc}, A_{acc}, tracheal length, and accelerometer location) is performed by minimizing the mean-square error of measures derived from the accelerometer-based glottal airflow and the pneumotachograph-based glottal airflow during sustained vowel phonation. Closed-phase inverse filtering [26] yields the oral airflow-based reference estimates of glottal airflow, where glottal closure instants are identified using a synchronously acquired electroglottogram signal. After this one-time calibration that estimates these fixed speaker-specific parameters, the algorithm can be applied to an arbitrary in-field accelerometer recording spectrum $\dot{U}_{\text{skin}}(\omega)$ without further calibration.

In Section IV, we demonstrate subglottal IBIF processing of the accelerometer signal to provide estimates of the following glottal airflow properties: the amplitude of the modulated flow component (AC Flow), maximum flow declination rate

(MFDR), speed quotient (SQ), open quotient (OQ), spectral slope (H1–H2), and harmonic richness factor (HRF) [2], [27].

IV. PILOT DATA

The developed voice monitoring system recorded seven full days of the high-bandwidth acceleration signal from each of six subjects, three with no history of voice disorders and three undergoing voice therapy at the Massachusetts General Hospital Voice Center. Table I reports subject characteristics and diagnoses. The human studies protocol was approved by the hospital's institutional review board.

Fig. 5 displays an example voice use profile of Day 1 for subject N2. Such visualizations may ultimately enable clinicians to make informed decisions regarding the management or prevention of pathologies that could arise due to certain patterns of voice use. Table II reports weekly averages of seven traditional ambulatory measures for the six subjects in this pilot study: average f0, f0 mode, SPL, phonation time, percent phonation, cycle dose, and distance dose. The elevated cycle dose and distance dose values for the patient sample owe in large part to higher fundamental frequencies. Variations in SPL (e.g., throughout the day) are highly averaged in the type of statistics reported in Table II, suggesting that average measures might not be sensitive enough to reveal differences among subjects and that a detailed analysis of dosage measures and short-term patterns is warranted.

Table III shows the performance of the subglottal IBIF approach by comparing voice-related measures derived from the reference airflow signals and the accelerometer signals,

TABLE II
SUMMARY STATISTICS OF TRADITIONAL AMBULATORY PHONATION MEASURES FOR THREE VOCALLY NORMAL SPEAKERS (N1–N3)
AND THREE SUBJECTS WITH VOICE DISORDERS (P1–P3)

Measure	N1	N2	N3	P1	P2	P3
Total time (hh:mm:ss)	10:05:22	14:14:55	09:55:57	11:56:41	10:31:14	11:21:18
Phonation time (hh:mm:ss)	00:13:56	00:47:26	00:34:36	00:37:58	00:45:31	00:47:16
Percent phonation (%)	1.91	5.48	5.71	5.25	6.99	6.75
F0 mode (Hz)	159.0	110.7	117.6	324.6	215.1	168.8
F0 average (Hz)	137.4	130.5	116.4	305.4	227.7	186.1
Sound level (dB SPL)	71.1	72.9	71.6	71.9	72.1	73.1
Cycle dose (cycles)	113 309	368 440	245 542	697 692	631 959	528 356
Distance dose (m)	429.9	1524.0	920.9	1332.5	1375.4	1259.2

Values reported are averages of daily statistics over 7 days.

TABLE III
GLOTTAL AIRFLOW MEASURES OBTAINED FROM THE REFERENCE ORAL AIRFLOW VOLUME VELOCITY (OVV) AND THE ACCELEROMETER SIGNAL (ACC).
RECORDED BY THE SMARTPHONE FROM SUBJECT N3 PRODUCING FIVE SUSTAINED VOWELS AT A COMFORTABLE PITCH AND LOUDNESS

Measure	Vowel /a/		Vowel /e/		Vowel /i/		Vowel /o/		Vowel /u/		Mean error
	OVV	ACC	OVV	ACC	OVV	ACC	OVV	ACC	OVV	ACC	
f0 (Hz)	100.6	100.6	94.1	94.1	97.7	97.7	94.7	94.7	93.1	93.1	0.0 %
AC Flow (mL/s)	609.9	637.6	736.2	796.5	631.6	640.7	626.0	621.6	691.2	698.8	3.2 %
MFDR (L/s^2)	644.7	607.9	804.6	720.6	511.1	485.9	533.8	526.0	638.5	617.8	5.2 %
SQ (%)	178.8	184.2	155.5	149.8	123.5	127.9	167.0	148.4	135.6	141.5	5.2 %
OQ (%)	55.0	58.7	61.4	65.5	58.6	61.0	57.2	55.9	55.9	55.7	4.0 %
H1–H2 (dB)	3.8	3.9	7.1	8.2	4.7	4.4	7.3	6.8	4.9	4.1	9.0 %
HRF (dB)	−3.5	−2.9	−6.6	−7.3	−4.4	−4.1	−6.8	−6.2	−4.7	−3.7	13.3 %

Mean error averages the absolute percent error over the vowels.

optimized for each utterance. Fundamental frequency and the six airflow-based measures are computed for five sustained vowels uttered by subject N3. The subglottal IBIF algorithm yields comparable estimates with respect to the reference and is capable of retrieving the signal structure during the open portion of the cycle. Similar results have been found for male and female speakers [19]. The match between the two signals depends to a minor extent on the vowel produced due to neck skin and laryngeal height changes.

The IBIF scheme yields comparable results for all measures, and higher accuracy is obtained for the temporal measures (AC Flow, MFDR, SQ, OQ) than for the spectral ones (H1–H2, HRF). These data are illustrative and do not represent a comprehensive error analysis of the IBIF algorithm. Model parameter estimation must be refined to operate under dynamic (continuous speech) scenarios, and a comprehensive set of recordings from large groups of subjects with and without voice disorders would provide much-needed data.

V. CONCLUSION AND DISCUSSION

In summary, we have developed and provided initial data from a new, versatile, and cost-effective system for ambulatory voice monitoring that uses a neck-placed miniature accelerometer as voice sensor and a smartphone as data acquisition platform. The main advantage of this device over current-generation systems is the collection of the unprocessed accelerometer signal that allows for the investigation of new voice use-related measures based on a vocal system model. In addition, the ease of use of the smartphone-based platform enables large-sample clinical studies that that can identify measures of high statistical power for differentiating between hyperfunctional and normal patterns of vocal behavior.

Ambulatory biofeedback has been shown in early case studies to have some potential to facilitate vocal behavioral changes being targeted in voice therapy [28]. Current devices provide simple feedback based on setting thresholds for f0 and SPL. Although apparently useful to some patients, thresholds based on these traditional measures do not provide the ability of facilitating and reinforcing changes in many other vocal behaviors that could be targeted in voice therapy. The potential for expanding the biofeedback capability of ambulatory voice monitoring systems could be increased dramatically once measures that differentiate pathological and normal patterns of vocal behavior are better delineated. Commercially available mobile devices have onboard signal processing capabilities that can support such expanded biofeedback applications while performing data acquisition. The incorporation of a remote connection with a central server could also provide additional online biofeedback capabilities and signal monitoring.

REFERENCES

[1] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, "Voice disorders in the general population: Prevalence, risk factors, and occupational impact," *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, 2005.

[2] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, "Objective assessment of vocal hyperfunction: An experimental framework and initial results," *J. Speech Hear. Res.*, vol. 32, no. 2, pp. 373–392, 1989.

[3] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, "Phonatory function associated with hyperfunctionally related vocal fold lesions," *J. Voice*, vol. 4, no. 1, pp. 52–63, 1990.

[4] K. Verdolini, C. Rosen, and R. C. Branski, *Classification Manual for Voice Disorders-I, Special Interest Division 3, Voice and Voice disorders, American Speech-Language Hearing Division*. Mahwah, NJ: Lawrence Erlbaum, 2005.

[5] S. Ryu, S. Komiyama, S. Kannae, and H. Watanabe, "A newly devised speech accumulator," *ORL J. Otorhinolaryngol. Relat. Spec.*, vol. 45, no. 2, pp. 108–114, 1983.

[6] A. C. Ohlsson, O. Brink, and A. Löfqvist, "A voice accumulation–validation and application," *J. Speech Hear. Res.*, vol. 32, no. 2, pp. 451–457, 1989.

[7] R. Buekers, E. Bierens, H. Kingma, and E. Marres, "Vocal load as measured by the voice accumulator," *Folia Phoniatr. Logop.*, vol. 47, no. 5, pp. 252–261, 1995.

[8] H. A. Cheyne, H. M. Hanson, R. P. Genereux, K. N. Stevens, and R. E. Hillman, "Development and testing of a portable vocal accumulator," *J. Speech. Lang. Hear. Res.*, vol. 46, no. 6, pp. 1457–1467, 2003.

[9] R. E. Hillman, J. T. Heaton, A. Masaki, S. M. Zeitels, and H. A. Cheyne, "Ambulatory monitoring of disordered voices," *Ann. Otol. Rhinol. Laryngol.*, vol. 115, no. 11, pp. 795–801, 2006.

[10] P. S. Popolo, J. G. Švec, and I. R. Titze, "Adaptation of a Pocket PC for use as a wearable voice dosimeter," *J. Speech. Lang. Hear. Res.*, vol. 48, no. 4, pp. 780–791, 2005.

[11] M. Zañartu, J. C. Ho, S. S. Kraman, H. Pasterkamp, J. E. Huber, and G. R. Wodicka, "Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 443–451, Feb. 2009.

[12] J. Nix, J. G. Švec, A.-M. Laukkanen, and I. R. Titze, "Protocol challenges for on-the-job voice dosimetry of teachers in the United States and Finland," *J. Voice*, vol. 21, no. 4, pp. 385–396, 2007.

[13] I. R. Titze, E. J. Hunter, and J. G. Švec, "Voicing and silence periods in daily and weekly vocalizations of teachers," *J. Acoust. Soc. Amer.*, vol. 121, no. 1, pp. 469–478, 2007.

[14] J. G. Švec, P. S. Popolo, and I. R. Titze, "Measurement of vocal doses in speech: Experimental procedure and signal processing," *Log. Phon. Voc.*, vol. 28, no. 4, pp. 181–192, 2003.

[15] I. R. Titze, J. G. Švec, and P. S. Popolo, "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," *J. Speech. Lang. Hear. Res.*, vol. 46, no. 4, pp. 919–932, 2003.

[16] J. G. Švec, I. R. Titze, and P. S. Popolo, "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," *J. Acoust. Soc. Amer.*, vol. 117, no. 3, pp. 1386–1394, 2005.

[17] P. J. Lynch, "Lungs-simple diagram of lungs and trachea," Creative Commons Attribution 2.5 License, 2006.

[18] H. A. Cheyne, "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," in *Proc. 3rd IEEE-EMBS Int. Summer School Symp. Med. Devices Biosens.*, 2006, pp. 118–121.

[19] M. Zañartu, "Acoustic coupling in phonation and its effect on inverse filtering of oral airflow and neck surface acceleration," Ph.D. dissertation, Dept. Electr. Comput., Eng., Purdue Univ., West Lafayette, IN, 2010.

[20] M. Zañartu, G. R. Wodicka, J. C. Ho, R. E. Hillman, and D. D. Mehta, "Estimation of glottal aerodynamics using an impedance-based inverse filtering of neck surface acceleration," U.S. Patent, Application Number 61 444 199, Filed on Feb. 18, 2012.

[21] G. R. Wodicka, K. N. Stevens, H. L. Golub, E. G. Cravalho, and D. C. Shannon, "A model of acoustic transmission in the respiratory system," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 9, pp. 925–934, Sep. 1989.

[22] B. H. Story, A.-M. Laukkanen, and I. R. Titze, "Acoustic impedance of an artificially lengthened and constricted vocal tract," *J. Voice*, vol. 14, no. 4, pp. 455–469, 2000.

[23] I. R. Titze, "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Amer.*, vol. 123, no. 5, pp. 2733–2749, 2008.

[24] B. H. Story, "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. dissertation, Dept. Speech Pathol. Audiol., Univ. Iowa, 1995.

[25] J. C. Ho, M. Zañartu, and G. R. Wodicka, "An anatomically based, time-domain acoustic model of the subglottal system for speech production," *J. Acoust. Soc. Amer.*, vol. 129, no. 3, pp. 1531–1547, 2011.

[26] P. Alku, C. Magi, S. Yrttiaho, T. Backstrom, and B. Story, "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering," *J. Acoust. Soc. Amer.*, vol. 125, no. 5, pp. 3289–3305, 2009.

[27] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *J. Acoust. Soc. Amer.*, vol. 90, no. 5, pp. 2394–2410, 1991.

[28] KayPENTAX, *Ambulatory Phonation Monitor: Applications for Speech and Voice*. Lincoln Park, NJ: KayPENTAX, 2009.