Published in final edited form as:

IEEE Trans Biomed Eng. 2012 November ; 59(11): 3090–3096. doi:10.1109/TBME.2012.2207896.

Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform

Daryush D. Mehta [Member, IEEE],

Center for Laryngeal Surgery & Voice Rehabilitation, Massachusetts General Hospital, Boston MA 02114 USA, Department of Surgery, Harvard Medical School, Boston, MA 02115 USA, and the School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138 USA

Matías Zañartu [Member, IEEE],

Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

Shengran W. Feng [Student Member, IEEE],

Center for Laryngeal Surgery & Voice Rehabilitation, Massachusetts General Hospital, Boston MA 02114 USA and Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences & Technology, Massachusetts Institute of Technology, Cambridge, MA 02139

Harold A. Cheyne II, and

Bioacoustic Research Program, Lab of Ornithology, Cornell University, Ithaca, NY 14853 USA

Robert E. Hillman

Center for Laryngeal Surgery & Voice Rehabilitation and Institute of Health Professions, Massachusetts General Hospital, Boston MA 02114 USA and Surgery and Health Sciences & Technology, Harvard Medical School, Boston, MA 02115

Daryush D. Mehta: daryush.mehta@alum.mit.edu; Matías Zañartu: matias.zanartu@usm.cl; Shengran W. Feng: willfeng@mit.edu; Harold A. Cheyne: hac68@cornell.edu; Robert E. Hillman: hillman.robert@mgh.harvard.edu

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

voice use; vocal hyperfunction; voice production model; accelerometer sensor; wearable voice sensor

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 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], [6], [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 (f_0) 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 use 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 (f_0 , SPL, phonation duration) 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 seven days. Finally, Section V concludes with a summary and future directions.

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 hours 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. Figure 1 highlights components of the proposed voice monitoring system.

A. Neck-placed miniature accelerometer

The selected miniature accelerometer senses acceleration in one dimension and has a vibration sensitivity suitable for obtaining meaningful information about the voice. The sensor is calibrated using a laser Doppler vibrometer system that outputs a chirp signal to a mechanical shaker. With no casing or mounting, the accelerometer sensitivity is approximately $88.5 \text{ dB} \pm 3 \text{ dB re } 1 \text{ cm/s}^2/\text{volt}$ 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. Figure 1A 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 non-volatile 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 width.

The Nexus S smartphone contains a high-fidelity audio codec (WM8994; Wolfson Microelectronics, Edinburgh, Scotland, UK) that encodes the accelerometer signal using sigma-delta modulation (128x 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).

Figure 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 hours of data per day for seven days equates to 9.3 GB of recorded data ($\text{GB} = 2^{30}$ 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 hours in one day.

E. User experience

Figure 1B 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, non-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 (Figure 3A) 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, f_0 , and SPL from the raw signal obtained from the neck-mounted accelerometer sensor [8]. In the current study, f_0 and SPL are computed every 50 ms (non-overlapping 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. f_0 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 f_0 values are set to zero for frames that are considered non-phonatory (containing either silence, unvoiced speech, or other non-speech energy).

An estimate of the acoustic SPL is derived from the neck skin vibration level through the calibration routine that simultaneously 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. Figure 3B 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].

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.

Figure 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). $Z_{skin}(\omega) = Z_m(\omega) + Z_{rad}(\omega)$, where the mechanical impedance of the neck skin $Z_m(\omega)$ is in series with the radiation impedance due to the accelerometer sensor load $Z_{rad}(\omega)$. $Z_m(\omega)$ is defined as

$$Z_m(\omega) = R_m + j(\omega M_m - \frac{K_m}{\omega}), \quad (1)$$

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

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

where M_{acc} and A_{acc} are the per-unit-area mass and surface area, respectively, of the accelerometer sensor assembly.

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

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

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

The parameters of $Z_{skin}(\omega)$ are considered fixed for each speaker and are derived by optimizing measures of $U_g(\omega)$ derived from $\dot{U}_{skin}(\omega)$ 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 $\hat{U}_{skin}(\omega)$ without further calibration.

In Section IV, we demonstrate subglottal IBIF processing of the accelerometer signal to provide estimates of the following glottal airflow properties: 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 new 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.

Figure 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 f_0 , f_0 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 sound pressure level, e.g., throughout the day are highly averaged in Table II statistics, 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 ground-truth airflow reference and accelerometer signals, optimized for each utterance. Fundamental frequency and 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 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.

Acknowledgments

This work was supported by the NIH National Institute on Deafness and Other Communication Disorders (grants R21 DC011588 and T32 DC00038), the Institute of Laryngology and Voice Restoration, and a Chilean CONICYT grant (FONDECYT 11110147).

The authors would like to acknowledge the contributions of Rob Petit for aid in designing and programming the smartphone app, Cedric Andrieu of Wolfson Microelectronics for audio codec information, and François Simond for Android operating system advice. The authors also thank James Heaton, PhD, and Jim Kobler, PhD, for their contributions to hardware design and building and John Rosowski, PhD, and Michael Ravicz for aid in mechanical calibration of accelerometers.

References

1. Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *Laryngoscope*. 2005; 115(11):1988–1995. [PubMed: 16319611]
2. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: An experimental framework and initial results. *J Speech Hear Res*. 1989; 32(2):373–392. [PubMed: 2739390]
3. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Phonatory function associated with hyperfunctionally related vocal fold lesions. *J Voice*. 1990; 4(1):52–63.
4. Verdolini, K.; Rosen, C.; Branski, RC. 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. Ryu S, Komiyama S, Kannae S, Watanabe H. A newly devised speech accumulator. *ORL J Otorhinolaryngol Relat Spec*. 1983; 45(2):108–114. [PubMed: 6341919]
6. Ohlsson AC, Brink O, Löfqvist A. A voice accumulation-validation and application. *J Speech Hear Res*. 1989; 32(2):451–457. [PubMed: 2739395]
7. Buekers R, Bierens E, Kingma H, Marres E. Vocal load as measured by the voice accumulator. *Folia Phoniatr Logop*. 1995; 47(5):252–261. [PubMed: 8563777]
8. Cheyne HA, Hanson HM, Genereux RP, Stevens KN, Hillman RE. Development and testing of a portable vocal accumulator. *J Speech Lang Hear Res*. 2003; 46(6):1457–1467. [PubMed: 14700368]
9. Hillman RE, Heaton JT, Masaki A, Zeitels SM, Cheyne HA. Ambulatory monitoring of disordered voices. *Ann Otol Rhinol Laryngol*. 2006; 115(11):795–801. [PubMed: 17165660]

10. Popolo PS, Švec JG, Titze IR. Adaptation of a Pocket PC for use as a wearable voice dosimeter. *J Speech Lang Hear Res.* 2005; 48(4):780–791. [PubMed: 16378473]
11. Zañartu M, Ho JC, Kraman SS, Pasterkamp H, Huber JE, Wodicka GR. Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface. *IEEE Trans Biomed Eng.* 2009; 56(2): 443–451. [PubMed: 19272887]
12. Nix J, Švec JG, Laukkanen A-M, Titze IR. Protocol challenges for on-the-job voice dosimetry of teachers in the United States and Finland. *J Voice.* 2007; 21(4):385–396. [PubMed: 16678386]
13. Titze IR, Hunter EJ, Švec JG. Voicing and silence periods in daily and weekly vocalizations of teachers. *J Acoust Soc Am.* 2007; 121(1):469–478. [PubMed: 17297801]
14. Švec JG, Popolo PS, Titze IR. Measurement of vocal doses in speech: Experimental procedure and signal processing. *Log Phon Voc.* 2003; 28(4):181–192.
15. Titze IR, Švec JG, Popolo PS. Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *J Speech Lang Hear Res.* 2003; 46(4):919–932. [PubMed: 12959470]
16. Švec JG, Titze IR, Popolo PS. Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *J Acoust Soc Am.* 2005; 117(3):1386–1394. [PubMed: 15807026]
17. Lynch PJ. Lungs-simple diagram of lungs and trachea. Creative Commons Attribution 2.5 License. 2006
18. Cheyne, HA. 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*; 2006. p. 118-121.
19. Zañartu, M. PhD dissertation. Purdue University; West Lafayette, IN: 2010. Acoustic coupling in phonation and its effect on inverse filtering of oral airflow and neck surface acceleration.
20. Zañartu, M.; Wodicka, GR.; Ho, JC.; Hillman, RE.; Mehta, DD. Estimation of glottal aerodynamics using an impedance-based inverse filtering of neck surface acceleration. US Patent, Application Number. 61444199. Filed on February 18, 2012
21. Wodicka G, Stevens K, Golub H, Cravalho E, Shannon D. A model of acoustic transmission in the respiratory system. *IEEE Trans Biomed Eng.* 1989; 36(9):925–934. [PubMed: 2777281]
22. Story B, Laukkanen A, Titze I. Acoustic impedance of an artificially lengthened and constricted vocal tract. *J Voice.* 2000; 14(4):455–469. [PubMed: 11130104]
23. Titze IR. Nonlinear source-filter coupling in phonation: Theory. *J Acoust Soc Am.* 2008; 123(5): 2733–2749. [PubMed: 18529191]
24. Story, BH. PhD dissertation. University of Iowa; 1995. Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract.
25. Ho JC, Zañartu M, Wodicka GR. An anatomically based, time-domain acoustic model of the subglottal system for speech production. *J Acoust Soc Am.* 2011; 129(3):1531–1547. [PubMed: 21428517]
26. Alku P, Magi C, Yrttiaho S, Backstrom T, Story B. Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering. *J Acoust Soc Am.* 2009; 125(5):3289–3305. [PubMed: 19425671]
27. Childers DG, Lee CK. Vocal quality factors: Analysis, synthesis, and perception. *J Acoust Soc Am.* 1991; 90(5):2394–2410. [PubMed: 1837797]
28. KayPENTAX. Ambulatory Phonation Monitor: Applications for Speech and Voice. Lincoln Park, NJ: KayPENTAX; 2009.

Biographies

Daryush D. Mehta (S'01–M'11) received the B.S. degree in electrical engineering (summa cum laude) from University of Florida, Gainesville, in 2003, the S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 2006, and the Ph.D. degree from MIT in speech and hearing bioscience and technology in the Harvard–MIT Division of Health Sciences and Technology, Cambridge, in 2010.

He currently holds appointments at Harvard University (Research Associate in the School of Engineering and Applied Sciences), Massachusetts General Hospital (Assistant Biomedical Engineer in the Department of Surgery), and Harvard Medical School (Instructor in Surgery), Boston. He is also an Honorary Senior Fellow in the Department of Otolaryngology, University of Melbourne, in Australia.

Matías Zañartu (S'08–M'11) received the Ph.D. and M.S. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, in 2010 and 2006, respectively, and his C.E. professional title and B.S. degree in acoustical engineering from Universidad Tecnológica Vicente Pérez Rosales, Santiago, Chile, in 1999 and 1996, respectively.

Since 2011, he is an Academic Research Associate at the Department of Electronic Engineering from Universidad Técnica Federico Santa María, Valparaíso, Chile. His current interests include digital signal processing, nonlinear dynamic systems, acoustic modeling, speech/audio/biomedical signal processing, speech recognition, and acoustic biosensors.

Shengran W. Feng (S'09) received B.S. degrees in biomedical engineering and applied mathematics from Johns Hopkins University, Baltimore, MD, in 2009. He is currently a Ph.D. candidate in the Speech and Hearing Bioscience and Technology Program in the Harvard–MIT Division of Health Sciences and Technology, MIT, Cambridge, MA.

He currently conducts research on long-term voice monitoring at the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston. His research interest lies in identifying abusive patterns in voice use that may lead to vocal hyperfunction.

Harold A. Cheyne II received the B.S. degree in electrical engineering (summa cum laude) from Tufts University in 1993, and the Ph.D. degree from MIT in speech and hearing bioscience and technology in the Harvard–MIT Division of Health Sciences and Technology, Cambridge, in 2002.

He is currently the Director of Technology for the Bioacoustics Research Program at the Cornell Lab of Ornithology, and consults in audio electronics design and audio forensics through his firm L.A.S.E.R., LLC.

Robert E. Hillman received the B.S. and M.S. degrees in speech pathology from Pennsylvania State University, University Park, in 1974 and 1975, respectively, and the Ph.D. degree in speech science from Purdue University, West Lafayette, IN, in 1980.

He is currently Co-Director/Research Director of the MGH Center for Laryngeal Surgery and Voice Rehabilitation, Associate Professor of Surgery & Health Sciences and Technology at Harvard Medical School, and Professor and Associate Provost for Research at the MGH Institute of Health Professions, Boston, MA. His research has been funded by both governmental and private agencies since 1981, and he has over 100 publications on normal and disordered voice.

He is a Fellow of the American Speech-Language-Hearing Association (also receiving Honors of the Association, ASHA's highest honor) and the American Laryngological Association.

Fig. 1.
(Color online) Mobile voice monitoring system: (A) Smartphone and accelerometer input,
(B) subject wearing the system with wiring underneath clothing and smartphone in belt
holster.

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

Fig. 3. (Color online) Acoustic calibration procedure for a sustained vowel produced with increasing loudness. (A) The 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).

Fig. 4.

(Color online) 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.

Fig. 5.

Voice use profile of subject N2's first day (24-hour time format). (A) Five-minute 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 (f_0); (C) histogram of SPL; and (D) phonation density showing the relative occurrence of particular combinations of SPL (horizontal axis) and f_0 (vertical axis).

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

\$watermark-text

\$watermark-text

\$watermark-text

TABLE II

Summary statistics of traditional ambulatory phonation measures for three vocally normal speakers (N1–N3) and three subjects with voice disorders (P1–P3). Values reported are averages of daily statistics over seven days.

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

\$watermark-text

\$watermark-text

\$watermark-text

TABLE III

Glottal airflow measures obtained from the reference oral airflow volume velocity (OVV) and the accelerometer signal recorded by the smartphone (ACC) from subject N3 producing five sustained vowels at a comfortable pitch and loudness. Mean error averages the absolute percent error over the vowels.

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 %