

Real-time estimation of aerodynamic features for ambulatory voice biofeedback

Andrés F. Llico, Matías Zañartu,^{a)} and Agustín J. González

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

andres.llico@usm.cl, matias.zanartu@usm.cl, agustin.gonzalez@usm.cl

George R. Wodicka^{b)}

Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana 47907, USA

wodicka@purdue.edu

Daryush D. Mehta,^{c),d)} Jarrad H. Van Stan,^{d)} and Robert E. Hillman^{c),d)}

Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

mehta.daryush@mgh.harvard.edu, jvanstan@mghihp.edu, hillman.robert@mgh.harvard.edu

Abstract: The development of ambulatory voice monitoring devices has the potential to improve the diagnosis and treatment of voice disorders. In this proof-of-concept study, real-time biofeedback is incorporated into a smartphone-based platform that records and processes neck surface acceleration. The focus is on utilizing aerodynamic measures of vocal function as a basis for biofeedback. This is done using regressed Z-scores to compare recorded values to normative estimates based on sound pressure level and fundamental frequency. Initial results from the analysis of different voice qualities suggest that accelerometer-based estimates of aerodynamic parameters can be used for real-time ambulatory biofeedback.

© 2015 Acoustical Society of America

[AL]

Date Received: January 10, 2015 **Date Accepted:** May 27, 2015

1. Introduction

Our research group has recently reported the development of a user-friendly and flexible platform for ambulatory voice monitoring referred to as the Voice Health Monitor (VHM).¹ The system consists of a smartphone as the data acquisition platform and a miniature accelerometer (ACC) as the phonation sensor mounted at the base of the neck. The VHM allows for recording of the unprocessed neck skin acceleration signal, with storage space for over 18 h of recording per day for at least 7 days, and it also has interactive capabilities. Recording the neck skin acceleration signal allows the investigation of new voice use-related measures based on a vocal system model and facilitates collecting data to better understand vocal behaviors related to daily activities.²

In current-generation technologies, approaches based on acoustic parameters have been implemented for real-time biofeedback using a threshold for either fundamental frequency (f_0) or sound pressure level (SPL).³ However, the VHM provides an opportunity to incorporate novel, more sophisticated real-time biofeedback capabilities based on advanced measures of vocal function to potentially improve the identification and treatment of detrimental vocal behaviors (e.g., vocal hyperfunction). A recent approach provides a method to estimate aerodynamic measures of vocal function from the ACC signal through modeling of the vocal system and performing subglottal impedance-based inverse filtering (IBIF).⁴ This vocal system model is a physiologically based transmission line model that uses the ACC signal as an input to estimate the glottal volume velocity (GVV) airflow waveform as an output, and it has been shown to provide accurate estimation of maximum flow declination rate (MFD_R) and unsteady flow amplitude (AC Flow). Using statistical analysis based on an assessment of regressed Z-scores, it has been shown that these aerodynamic measures are

^{a)} Author to whom correspondence should be addressed.

^{b)} Also at School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA.

^{c)} Also at Harvard Medical School, Boston, MA 02114, USA.

^{d)} Also at MGH Institute of Health Professions, Boston, MA 02114, USA.

associated with the presence of adducted vocal hyperfunction that can cause vocal fold trauma and the formation of lesions (e.g., nodules).⁵

The aim of this study is to incorporate new, more advanced real-time biofeedback capabilities into the VHM. The new approach is based on using aerodynamic measures of vocal function that are extracted from neck surface acceleration as a basis for biofeedback. Regressed Z-scores are used to compare recorded values to normative estimates of aerodynamic parameters (predicted normal thresholds) on the basis of sound pressure level and fundamental frequency. The present investigation provides a proof-of-concept for these new features, which still require further development and testing within a framework that incorporates current principles of motor learning and biofeedback.⁶

2. Real-time biofeedback approaches

In this section, two approaches for triggering real-time biofeedback are presented and implemented: (1) a threshold-based method using fundamental frequency and SPL estimates as trigger parameters and (2) a Z-score threshold approach based on aerodynamic measures derived from IBIF processing.

2.1 Smartphone platform

The algorithms to provide real-time biofeedback are implemented on the Samsung Nexus S smartphone (Samsung, Seoul, South Korea) and a software application called the VHM. The system records the unprocessed ACC signal at an 11 025 Hz sampling rate, 16-bit quantization, and 80-dB dynamic range to obtain frequency content of neck surface vibrations up to 5000 Hz. Detailed specifications of the system have been previously published.¹

2.2 Fundamental frequency and SPL biofeedback

The accelerometer is attached a few centimeters above the suprasternal notch using double-sided tape (model 2811, 3M Company, Maplewood, MN). Even though there are multiple ways to compute f_0 in the acoustic domain,⁷ a simple time-domain technique is sufficient to estimate f_0 from the ACC signal through a normalized autocorrelation function every 50 ms. SPL was estimated in the time domain by calibrating the level of the ACC signal to the level of a microphone signal when recorded 15 cm from the lips. The calibration factor was obtained from a linear regression between the synchronized microphone and accelerometer signals as the speaker produces the sustained vowel /a/ with increasing loudness.¹

During real-time data acquisition, the neck skin acceleration signal was processed in frames of 50 ms, which were divided into sub-frames of 25 ms, where SPL was computed for each sub-frame and compared with a voicing threshold settable on the device (set to 62 dB). If both sub-frames surpass the threshold, the frame is labeled as voiced and then SPL and f_0 are re-computed for the entire frame where f_0 is restricted to the range of 60 to 500 Hz. Otherwise, the frame is labeled as unvoiced, and both SPL and f_0 are set to zero. These thresholds and ranges for SPL and f_0 have been used on previous studies,^{1,2} where the focus was on normal to loud vocal efforts. However, the device provides nearly 80 dB of dynamic range,¹ so the SPL threshold could be adjusted to study lower or higher SPL levels.

With regard to biofeedback, values of SPL and f_0 are used to trigger a vibrotactile cue based on SPL and/or f_0 thresholds as well as a duration threshold (i.e., how long the threshold must be exceeded before triggering a vibrotactile cue). The vibrotactile alert has a modifiable duration from 50 to 1000 ms. These settings replicate the biofeedback capabilities of the Ambulatory Phonation Monitor (APM, model 3200, KayPENTAX, Montvale, NJ).

2.3 Subglottal impedance-based inverse filtering based biofeedback

The IBIF method was implemented for real-time processing in the time domain via a finite impulse response (FIR) filter obtained through a separate calibration procedure where the oral volume velocity (OVV) airflow waveform and electroglottography signals were recorded (and processed offline) prior to the subject wearing the VHM. This procedure only uses sustained vowels /a/ performed at normal loudness, thus lasting less than 5 min to complete. The recordings provide information that allows to obtain subject-specific parameters, which have been initially shown to vary to a certain degree in running speech, for which the IBIF parameters adjustment from sustained vowels is considered valid.⁸ The length of the FIR filter impulse response was reduced from 50 ms to 15 ms to balance the need for computational efficiency and performance for real-time processing and biofeedback. As a result of filtering the ACC signal with the IBIF filter, an estimate of the GVV signal (in mL/s) and the time derivative (dGVV)

are obtained on a frame-by-frame basis. Both, dGVV and GVV signals are used to derive MFDR and AC Flow features, respectively. While MFDR is obtained from the negative peak of the first derivative of the GVV waveform and is an indirect estimate of maximum vocal fold closing velocity, AC Flow is the unsteady component of the flow and is assumed to indirectly reflect the amplitude of vocal fold vibration. MFDR and AC Flow have been demonstrated to robustly represent the same features derived from a gold-standard (the OVV signal) within 10% accuracy⁴ and have been previously used in the clinical assessment of vocal function.⁵ Furthermore, MFDR and AC Flow have been used empirically to differentiate subjects with normal voices from a small sample of patients with vocal hyperfunction in the form of regressed Z-scores.⁵ These measures are SPL and f_0 dependent; however, in this study, they were only computed based on SPL (instead of using both SPL and f_0)⁵ for simplicity.

Regressed Z-scores for MFDR and AC Flow are calculated as

$$Z = \frac{x - \bar{x}}{\sigma_x}, \quad (1)$$

where x is the frame-based MFDR or AC Flow value, and \bar{x} and σ_x are the mean and standard deviation, respectively, for a vocally normal group's MFDR or AC Flow at specific SPL values. This group consists of 62 adult speakers (28 females and 34 males). The males ranged in age from 20 to 56 years, while the females ranged from 20 to 49 years.

Figure 1 shows MFDR and AC Flow estimates for different SPL values. The subsequent normalization of MFDR and AC Flow estimates in real time corrects for the effect of loudness. The statistical parameters to compute the Z-scores are pre-loaded into smartphone memory before monitoring an individual, where the mean and standard deviation for a normal group were projected from Perkell *et al.*⁹ as a function of SPL. In order to account for the effects of variations in MFDR and AC Flow across SPL, a linear regression described the relationship using normative values (means and standard deviations from either MFDR or AC Flow) from the normal and loud conditions found in Perkell *et al.*⁹ Above the SPL for loud voice, the mean of MFDR and AC Flow is calculated according to the linear regression; the standard deviation is fixed at the value corresponding to a loud voice. Below the SPL for comfortable voice production, mean and standard deviation values are fixed at those given for the comfortable voice to avoid negative mean values. This is clearly a first approximation of real-time Z-score analysis, and thus requires further research to validate or enhance the method.

The approach described above is the first attempt to compute ACC-derived aerodynamic measures in real time. This Z-score scheme for biofeedback is hypothesized to be more clinically salient than biofeedback based on simple f_0 and SPL thresholds since (1) MFDR and AC Flow have been shown to be sensitive to the presence of hyperfunctional vocal behaviors⁵ and (2) the system provides a biofeedback measure that takes into account the subject's acoustic sound level.

3. Experimental setup for validation

In order to evaluate the functionality of the VHM's biofeedback capabilities, the APM was used as a reference for SPL based biofeedback. Since the APM does not compute

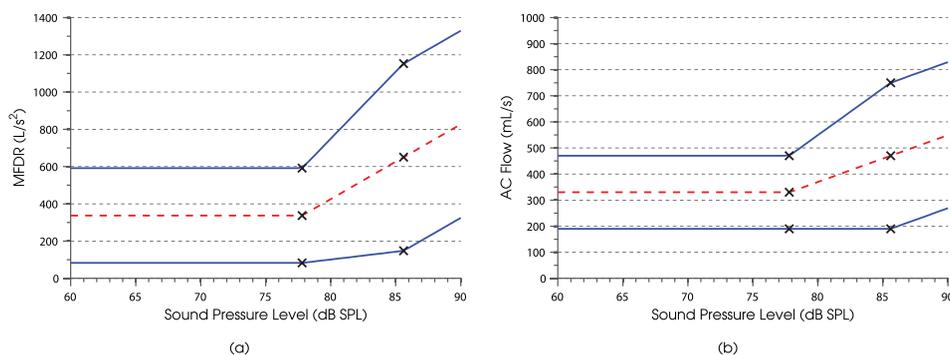


Fig. 1. (Color online) Method for obtaining Z-scores for (a) maximum flow declination rate (MFDR) and (b) AC Flow measures versus sound level. Plots show interpolated means (red dashed line) ± 2 standard deviation (blue solid line) for each measure, where x-marks (black) indicate empirical data points from speakers with normal voices (15 males).

Table 1. Summary statistics of ambulatory phonation measures for the Ambulatory Phonation Monitor (APM) and Voice Health Monitor (VHM) systems.

	Device	
	APM	VHM
Total time (hours:minutes:seconds)	02:08:47	02:08:52
Phonation time (hours:minutes:seconds)	00:38:03	00:32:45
Percent phonation (%)	29.61	25.42
Mean fundamental frequency (Hz)	145.9	148.1
Mean sound level (dB SPL)	82.7	81.6
Percent compliance (%)	93.5	96.3
Biofeedback time (hours:minutes:seconds)	00:00:46	00:00:19
Percent biofeedback (%)	2.03	1.01

MFDR and AC Flow, it does not estimate a parameter comparable to the VHM's aerodynamic Z-scores function.

3.1 Biofeedback based on SPL

The new real-time processing and biofeedback capabilities implemented on the VHM were validated through a comparison with an ambulatory voice monitor, the APM. Specifically, this comparison was carried out using a bioacoustics transducer tester (BATT)¹⁰ with an ambulatory recording of neck surface acceleration previously captured with the VHM from an adult male subject with a normal voice during a 90 min lecture. Both systems were calibrated with the same subject-specific parameters that related accelerometer level to acoustic SPL.¹ An ACC was mounted on a BATT that was set to have a flat, band-limited response between 70 Hz and 2 kHz.

Only dB-thresholds were tested in this study and to simulate biofeedback, both the APM and VHM were set with an upper dB-threshold of 95 dB SPL (above which biofeedback is triggered), duration threshold of 300 ms, and a vibrotactile alert duration of 300 ms.

3.2 Biofeedback based on aerodynamic measures

Given that no existing ambulatory device can be used to evaluate the aerodynamic features, the tests in this section focused on two objectives: (1) verifying the correct processing and discrimination of the Z-scores in real-time by recording and analyzing different voice modes and (2) assuring that the added processing does not interfere with the overall performance of the VHM. A vocally normal adult male sustained the vowel /a/ in modal and nonmodal (combination of breathy and rough) voice qualities at varying SPL while the Z-score for MFDR was computed online on the smartphone.

4. Results

4.1 Biofeedback based on SPL

The summary statistics for measures computed by the APM and VHM are shown in Table 1. The same conditions and calibration were provided to both systems, with phonation time, percent compliance, and biofeedback time qualitatively similar.

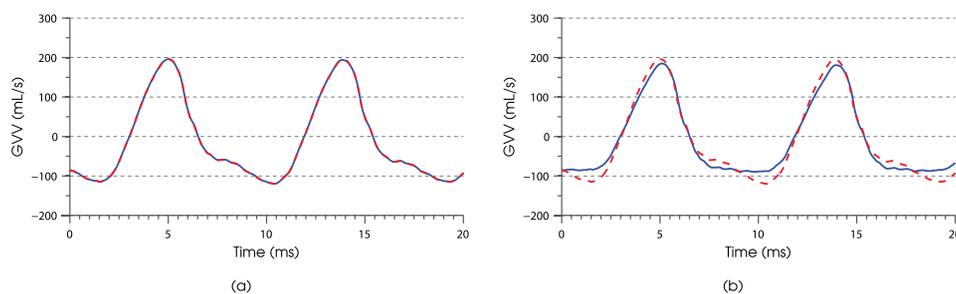


Fig. 2. (Color online) Comparison of glottal volume velocity (GVV) airflow estimation using online (red dashed line) and offline (blue solid line) subglottal impedance-based inverse filtering for the sustained vowel /a/ by an adult male. Also shown is the effect of (a) same-length finite impulse response (FIR) filter and (b) shorter FIR filter for online processing on the smartphone. Note that DC levels are not modeled by the AC accelerometer signal.

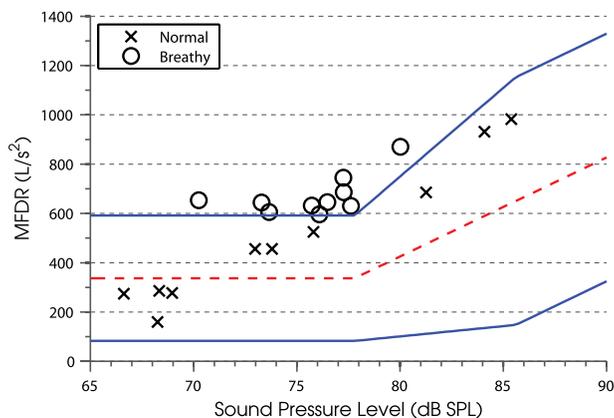


Fig. 3. (Color online) Estimation of maximum flow declination rate (MFDR) for a sustained vowel /a/ by an adult male. Measures derived from modal (black x-marks) and nonmodal (black rounded marks) voices indicate that vowel segments of the nonmodal vowel lie outside the normative bounds.

However, the two devices were not synchronized in time and operated with separate internal clocks, despite the fact the input stimuli was the same for both. Thus, onsets/offsets are captured in different ways, which can affect the phonation estimates. Differences in average f_0 and SPL estimates were approximately 2 Hz and 1 dB, respectively. Percent biofeedback showed the largest discrepancy. These results can be also addressed in terms of technological differences between the two devices since the VHM has a 16-bit instead of 7-bit signal quantization, which results in a wider dynamic range and a superior numerical precision than the APM.

4.2 Biofeedback based on aerodynamic measures

Figure 2 compares online and offline IBIF processing. Figure 2(a) illustrates that the GVV waveform estimated using the same IBIF filter length (15 ms) are similar for online and offline processing. In contrast, Fig. 2(b) notes slight differences when this shorter (15 ms) version of the IBIF filter implemented on the device is compared with the full (50 ms) version of the filter computed offline. However, both yield a similar waveform.

Figure 3 presents MFDR estimates using the short version of the online IBIF filter for the modal and nonmodal vowels. Whereas MFDR values for modal vowels were observed within the normative range, those derived from the nonmodal vowels were outside of the normative range. These initial results suggest that normalizing MFDR for SPL may aid in indicating the presence of nonmodal voice quality, which is frequently associated with vocal hyperfunction.

5. Conclusion

Two approaches to real-time biofeedback were incorporated into the VHM. First, a vibrotactile alarm based on SPL and f_0 thresholds was implemented and compared with the APM. The VHM provided comparable performance to the APM and differences noted are likely due to the increased quantization, dynamic range, synchronization, and computational precision of the VHM. In addition, a real-time biofeedback scheme was introduced that was based on the estimation of aerodynamic parameters using subglottal impedance-based inverse filtering and a Z-score assessment. The aerodynamic assessment discriminated between modal and nonmodal vowels in a proof-of-concept test. Also, the initial performance assessment indicated that the added computational load is not a limitation for the smartphone platform.

Future work will include the testing of innovative ambulatory biofeedback approaches based on motor control and learning theories to improve retention of desired vocal motor behaviors using the framework proposed herein.

Acknowledgments

This work was supported by NIH-NIDCD Grants Nos. R33 DC011588 and F31 DC014412, CONICYT grants FONDECYT 11110147 and Basal FB0008, MIT MISTI Grant No. MIT-Chile 2745333, and a grant from the Voice Health Institute. A.F.L. acknowledges support provided by CONICYT and UTFSM. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References and links

- ¹D. D. Mehta, M. Zañartu, S. W. Feng, H. A. Cheyne II, and R. E. Hillman, “Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform,” *IEEE Trans. Biomed. Eng.* **59**(11), 3090–3096 (2012).
- ²M. Ghassemi, J. H. Van Stan, D. D. Mehta, M. Zañartu, H. A. Cheyne II, R. E. Hillman, and J. V. Guttag, “Learning to detect vocal hyperfunction from ambulatory neck-surface accelerometer features: Initial results for vocal fold nodules,” *IEEE Trans. Biomed. Eng.* **61**(6), 1668–1675 (2014).
- ³J. H. Van Stan, J. Gustafsson, E. Schalling, and R. E. Hillman, “Direct comparison of commercially available ambulatory voice monitors: A clinical perspective,” *Perspect. Voice Voice Disorders* **24**(2), 80–86 (2014).
- ⁴M. Zañartu, J. C. Ho, D. D. Mehta, R. E. Hillman, and G. R. Wodicka, “Subglottal impedance-based inverse filtering of speech sounds using neck surface acceleration,” *IEEE Trans. Audio Speech Lang. Process.* **21**(9), 1929–1939 (2013).
- ⁵R. 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.* **32**(2), 373–392 (1989).
- ⁶J. H. Van Stan, D. D. Mehta, and R. E. Hillman, “The effect of voice ambulatory biofeedback on the daily performance and retention of a modified vocal motor behavior in participants with normal voices,” *J. Speech Lang. Hear. Res.*, in press (2015).
- ⁷A. Tsanas, M. Zañartu, M. Little, C. Fox, L. Ramig, and G. Clifford, “Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering,” *J. Acoust. Soc. Am.* **135**(5), 2885–2901 (2014).
- ⁸V. Espinoza and M. Zañartu, “Estudio dinámico de parámetros de filtrado inverso para el seguimiento ambulatorio de la función vocal” (“Dynamic study of inverse filtering parameters for ambulatory monitoring of vocal function”), in *IX Congreso Iberoamericano de Acústica, FIA2014*, Valdivia, Chile, December 2014.
- ⁹J. S. Perkell, R. E. Hillman, and E. B. Holmberg, “Group differences in measures of voice production and revised values of maximum airflow declination rate,” *J. Acoust. Soc. Am.* **96**(2), 695–698 (1994).
- ¹⁰S. S. Kraman, G. A. Pressler, H. Pasterkamp, and G. R. Wodicka, “Design, construction, and evaluation of a bioacoustic transducer testing (BATT) system for respiratory sounds,” *IEEE Trans. Biomed. Eng.* **53**(8), 1711–1715 (2006).