# Improved Subglottal Pressure Estimation From Neck-Surface Vibration in Healthy Speakers Producing Non-Modal Phonation

Jon Z. Lin, Víctor M. Espinoza, Katherine L. Marks, Matías Zañartu, *Senior Member, IEEE*, and Daryush D. Mehta, *Member, IEEE*

***Abstract*—Subglottal air pressure plays a major role in voice production and is a primary factor in controlling voice onset, offset, sound pressure level, glottal airflow, vocal fold collision pressures, and variations in fundamental frequency. Previous work has shown promise for the estimation of subglottal pressure from an unobtrusive miniature accelerometer sensor attached to the anterior base of the neck during typical modal voice production across multiple pitch and vowel contexts. This study expands on that work to incorporate additional accelerometer-based measures of vocal function to compensate for non-modal phonation characteristics and achieve an improved estimation of subglottal pressure. Subjects with normal voices repeated /p/-vowel syllable strings from loud-to-soft levels in multiple vowel contexts (/a/, /i/, and /u/), pitch conditions (comfortable, lower than comfortable, higher than comfortable), and voice quality types (modal, breathy, strained, and rough). Subject-specific, stepwise regression models were constructed using root-mean-square (RMS) values of the accelerometer signal alone (baseline condition) and in combination with cepstral peak prominence, fundamental frequency, and glottal airflow measures derived using subglottal impedance-based inverse filtering. Five-fold cross-validation assessed the robustness of model performance using the root-mean-square error metric for each regression model. Each cross-validation fold exhibited up to a 25% decrease in prediction error when the model incorporated multi-dimensional aspects of the accelerometer signal compared with RMS-only models. Improved estimation of subglottal pressure for non-modal phonation was thus achievable, lending to future studies of subglottal pressure estimation in patients with voice disorders and in ambulatory voice recordings.**

***Index Terms*—Subglottal pressure, clinical voice assessment, neck-surface accelerometer, ambulatory voice monitoring.**

## I. Introduction

Voice disorders affect approximately 30% of the adult population in the United States at some point in their lives and up to 7.6% at any given point in time [1], [2], with far-reaching social, professional, and personal consequences [3]. Subglottal air pressure (Ps) plays a major role in voice production and is a primary factor in controlling voice onset, offset, and intensity, and contributes to volitional control of fundamental frequency. In terms of clinical voice assessment, Ps alone and ratios incorporating Ps and airflow (e.g., aerodynamic resistance and vocal efficiency measures), have been shown to differentiate between normal and disordered voice production and to provide insight into changes in vocal function associated with treating voice disorders [4]–[11]. Ps is a central component of vocal efficiency metrics [12]–[16] and is associated with aspects of perceived vocal effort [17]–[19]. Other aerodynamic measures showing discriminatory power include parameters of the glottal airflow waveform, such as peak-to-peak airflow and maximum flow declination rate (MFDR) [11], [20].

Measurements of Ps, however, are underutilized in clinical settings due to the invasive techniques or specialized/expensive equipment required. Direct Ps measurement includes rarely-used invasive methods such as tracheal puncturing [21], [22] and transglottal passage of miniature pressure transducers [23], [24]. Cumbersome indirect methods include full body plethysmography [25], [26] and esophageal balloon techniques [23], [27]. In specialized settings where clinical estimates of Ps are obtained, the typical approach involves well-controlled productions of sustained vowels (constant pitch and loudness at a set syllable rate) interrupted volitionally by bilabial closure (/p/ or /b/ consonants) to temporarily equilibrate Ps with intraoral pressure, which is measured via pressure sensor attached to a translabial catheter [28]. A related mechanical airflow interruption technique has been subsequently developed [29] but also is limited to estimating Ps during isolated vowel contexts. Even though such Ps estimates provide valuable information about vocal function,

Manuscript received May 22, 2019; revised October 9, 2019; accepted November 25, 2019. Date of publication December 12, 2019; date of current version April 8, 2020. This work was supported in part by the NIH National Institute on Deafness and Other Communication Disorders under Grants R21 DC015877 and P50 DC015446 and in part by CONICYT BASAL FB0008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Claudia Manfredi. *(Corresponding author: Jon Z. Lin.)*

J. Z. Lin is with the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114 USA (e-mail: jzlin@mgh.harvard.edu).

V. M. Espinoza is with the Department of Sound, Universidad de Chile, Santiago 8340380, Chile (e-mail: vespinoza@uchile.cl).

K. L. Marks is with the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114 USA, and also with the MGH Institute of Health Professions, Boston, MA 02129 USA (e-mail: kmarks@mghihp.edu).

M. Zañartu is with the Department of Electronic Engineering, Universidad Técnica Federico Santa Maria, Valparaíso 2390123, Chile (e-mail: matias.zanartu@usm.cl).

D. D. Mehta is with the Center for Laryngeal Surgery and Voice Rehabilitation, Department of Surgery, Massachusetts General Hospital–Harvard Medical School, Boston, MA 02114 USA, the MGH Institute of Health Professions, Boston, MA 02129 USA, and also with the Speech and Hearing Bioscience and Technology Program, Division of Medical Sciences, Harvard Medical School, Boston, MA 02115 USA (e-mail: mehta.daryush@mgh.harvard.edu).

Digital Object Identifier 10.1109/JSTSP.2019.2959267

the information is inherently limited by uncertainties about how well the vowel-based measures reflect glottal function during natural speech where pitch, loudness, and rate of speech vary rapidly.

An inexpensive, non-invasive accelerometer sensor has shown promise to unobtrusively estimate voice characteristics during natural voice production [30]–[33]. When positioned on the anterior neck during phonation, the accelerometer signal consists of components related to tissue-to-tissue transmission of vocal fold collision forces through the thyroid cartilage and air-to-tissue transmission of aerodynamic energy through the tracheal wall to the neck surface [30], [34]. The field of ambulatory voice monitoring has taken advantage of accelerometers and contact microphone sensors to estimate basic characteristics of fundamental frequency ($f_o$) and sound pressure level (SPL), with the primary objective of quantifying the accumulated impact of prolonged voice use [35]–[43]. Additional salient measures related to the glottal airflow waveform—peak-to-peak airflow, open quotient, and MFDR—have been extracted from the accelerometer signal using subglottal impedance-based inverse filtering [44].

Previous work has shown that average Ps is correlated with the root-mean-square (RMS) amplitude of the neck-surface accelerometer signal during normal modal voice production across multiple pitch and vowel contexts [45]. Those results suggest that a linear fit between accelerometer RMS amplitude and Ps can be used to calibrate the accelerometer signal level in terms of Ps estimates that can be performed in a continuous (frame-based) manner during natural speech production. Critically, the accelerometer-based estimation of Ps during normal modal voice production exhibited less uncertainty than traditional estimation of SPL from accelerometer RMS amplitude [46]. The coefficient of determination between accelerometer RMS amplitude and Ps within 10 adult participants was high ($r^2 = 0.68$–$0.93$). These relationships were stronger than between accelerometer RMS amplitude and SPL ($r^2 = 0.46$–$0.81$). Higher degrees of uncertainty are problematic, as SPL estimates obtained from accelerometer data are often used to derive voice-use parameters such as distance dose and energy dissipation dose [37], [38].

The current work follows up on these results to quantify the impact of non-modal phonation on amplitude-based accelerometer estimates of Ps and compensate for this impact by incorporating additional accelerometer-based measures of glottal function. Non-modal phonation refers to voicing that deviates from the most common type of voice qualities that are characterized by periodic vocal fold vibration [47]. Examples of non-modal phonation include categorical qualities such as vocal fry and diplophonia, as well as more continuously scaled qualities of breathiness, roughness, and strain. Auditory perceptions of breathy, rough, and strained/pressed voice qualities are often evaluated during clinical voice assessment due to their presence in the speaking voice of individuals with voice disorders [48].

Accelerometer-based estimates of Ps have recently been evaluated in studies with vocally healthy speakers who produced breathy, rough, and strained p-vowel syllable strings [49] or who were instructed to modulate their vocal effort [50]. The take-home message of these studies was that the baseline regression line between accelerometer RMS level and Ps for modal phonation was significantly affected when non-modal phonation or higher vocal effort was produced. In particular, the intercepts of the regression line generally increased for non-modal phonatory conditions without concomitant changes in the slope. Thus, the Ps required for speakers to initiate voicing and maintain phonation at given levels of neck-surface vibration tended to increase when their phonation was more breathy, strained, or rough. Similar results have also been reported for pressed voice quality, where lower MFDR values were produced for the same levels of Ps [51].

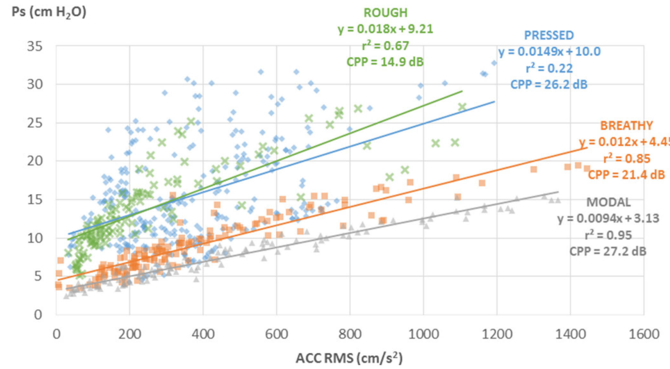


Fig. 1. Effect of non-modal phonation on the relationship between accelerometer root-mean-square (ACC RMS) and subglottal pressure (Ps), and the ability of mean cepstral peak prominence (CPP) to quantify non-modal characteristics.

Fig. 1 illustrates the increased intercept effect and added variance of data points when adding non-modal phonatory qualities to the typical scatterplot mapping Ps to accelerometer RMS amplitude. This study hypothesizes that additional accelerometer-based measures of vocal function can compensate for non-modal phonation characteristics and achieve improved estimation of Ps. The mean accelerometer-based cepstral peak prominence within each phonatory condition in Fig. 1 is shown to illustrate a potential compensatory measure. Related work has shown that accelerometer-based measures of jitter, shimmer, spectral amplitudes, and spectral entropy can classify modal, breathy, and pressed, with accuracy reaching 82.5% [52]. Similarly, to systematically determine the impact of non-modal phonation on accelerometer-based estimates of Ps in a controlled manner, vocally normal individuals were taken from our prior study who produced voice samples in different voice qualities [49].

## II. Methods

### A. Subject Recruitment

Twenty-six vocally healthy adult speakers (18 women, 8 men) were recruited to participate in this study [49]. For women, the mean (SD) participant age was 26 (7.6) years, ranging from 19 to 47 years; for men, the mean (SD) participant age was 33 (9.9) years, ranging from 19 to 50 years. Sixteen of the 26 subjects had vocal training. Subjects had no history of voice disorders or current complaints related to their vocal status. They also underwent laryngeal videostroboscopy to verify that their vocal folds exhibited typical vibratory patterns with straight edges, as assessed by a licensed speech-language pathologist specializing in voice disorders.

### B. Subject Protocol

Subjects repeated /p/-vowel syllable strings from loud-to-soft levels in multiple vowel contexts (/pa/, /pi/, and /pu/), pitch conditions (comfortable, lower than comfortable, higher than comfortable), and voice quality type. A voice-specialized speech-language pathologist monitored the data collection and visually evaluated the flatness of the intraoral pressure plateaus. If plateaus were not visibly flat (see [53] for the various intraoral pressure waveshapes that speakers can exhibit), subjects were instructed to repeat that trial. To determine the impact of non-modal phonation on accelerometer-based estimates of Ps, participants were asked to produce four different voice conditions: *modal*, *breathy*, *strained*, and *rough*. The elicited voice qualities were chosen to mimic pathological glottal conditions and were drawn from the perceptually rated dimensions in the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) clinical form [48]. It should be noted that the intent of eliciting the non-modal phonatory conditions was not to obtain pure examples of breathy, strained, and rough qualities, but rather to elicit a variety of voice conditions that might influence the relationship between Ps and accelerometer signal measures.

The terms "modal" and "non-modal" were defined using an established nonmodal taxonomy [47], where "modal" referred to the usual or baseline type of phonation and "non-modal" referred to any phonation that differs from or contrasts with the usual voice quality. Since all the participants were speakers with healthy voices, modal phonation was used as the reference category when assessing the impact of non-modal phonatory conditions, consistent with prior studies [54].

For modal productions, participants were instructed to produce a string of p-vowel tokens in one breath starting from a loud vocal intensity and gradually decreasing in loudness to a soft vocal intensity. This method allowed for the acquisition of a wide range of loudness levels and large number of data points in a short period of time [13], [45], relative to the conventional method of eliciting one vocal intensity per syllable string. For breathy productions, participants were asked to produce the same task using a breathy or airy voice. For strained productions, participants were asked to perform the task using a voice as if they were lifting something heavy while speaking. For the rough productions, participants were asked to produce the task using a voice with a rough quality (e.g., mimicking "Cookie Monster" or "Batman" character voices). When necessary, the task was modeled by the investigators.

Participants produced two to three trials per pitch level for each modal/non-modal phonatory condition, yielding up to 36 trials (3 trials $\times$ 3 pitch levels $\times$ 4 phonatory conditions). It should be noted that for most participants, it was difficult to change pitch when producing the rough condition, so only comfortable pitch was included in the analysis. The entire recording session typically lasted approximately 20 minutes, and participants were encouraged to take breaks as needed to minimize any potential confounding effects of vocal fatigue.

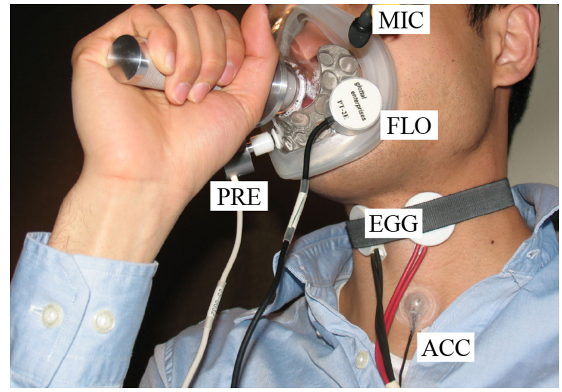


Fig. 2. In-laboratory data acquisition setup. (A) Synchronized recordings are made of signals from an acoustic microphone (MIC), electroglottography electrodes (EGG), accelerometer sensor (ACC), high-bandwidth oral airflow (FLO), and intraoral pressure (PRE). From [51].

### C. Data Collection

Fig. 2 shows the laboratory setup where synchronous recordings were made in a sound-treated booth using a pneumotachograph mask (Glottal Enterprises, Syracuse, NY) with oral airflow (PT-2E, Glottal Enterprises) and intraoral pressure (PT-75, Glottal Enterprises) sensors, electroglottograph (EG-2, Glottal Enterprises), and head-mounted condenser microphone positioned 15 cm from the lips (ME102, Sennheiser Electronic GmbH, Wennebostel, Germany). All signals were low-pass filtered at 8 kHz (CyberAmp Model 380, Axon Instruments, Union City, CA) prior to digital sampling at 20 kHz and 16-bit quantization (Digidata 1440A, Axon Instruments). FLO, IOP, and MIC signals were calibrated to physical units of mL/s, cm $H_2O$, and Pa, respectively.

A high-bandwidth accelerometer (ACC) sensor (BU-27135; Knowles Corp., Itasca, IL) was affixed halfway between the thyroid prominence and the sternal notch using hypoallergenic double-sided tape (Model 2181, 3M, Maplewood, MN) to measure neck-surface vibration in units of cm/s$^2$. Since data were collected as part of a larger study involving ambulatory voice monitoring, the accelerometer signal was recorded at an 11025 Hz sampling rate and 16-bit quantization onto a smartphone whose audio drivers and filters were modified for high-quality sampling instead of default telephone-optimized settings [55], [56].

### D. Signal Analysis

Fig. 3 displays an example of MIC, ACC, and IOP signals for one trial in a modal phonatory condition for one male participant. The voiceless /p/ plosives of the p-vowel gestures created a sequence of descending pulses in the IOP signal. Vowel segments can be seen in the MIC and ACC signals between IOP pulses.

Boundaries of the vowel segments were determined in the microphone signal using Praat version 6.0.30, which identified sounding/silent intervals [57]. The built-in algorithm was configured to detect a $-25$ dB change in signal amplitude from the maximum amplitude within 32-ms sliding windows

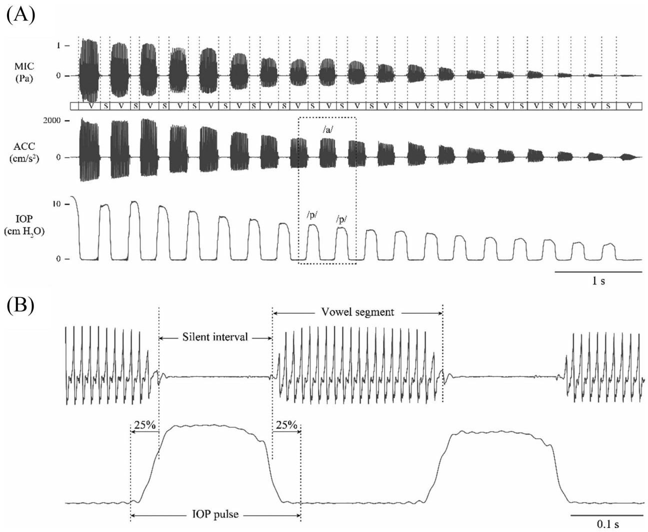


Fig. 3. An example of the repeated/pa/gesture with descending loudness for male participant M2. (A) Time-aligned signals from the acoustic microphone (MIC), neck-surface accelerometer (ACC), and intraoral pressure (IOP) sensor are displayed with voice activity label (S = silence, V = vowel). The boxed region is expanded in (B) to illustrate the boundary detection of each vowel segment and IOP pulse.

TABLE I
ACCELEROMETER-BASED MEASURES AND GLOTTAL AIRFLOW MEASURES ESTIMATED USING INVERSE FILTERING AND SUBGLOTTAL IMPEDANCE-BASED INVERSE FILTERING

| Feature | Description | Units |
|---|---|---|
| RMS | RMS amplitude | $cm/s^2$ |
| $f_o$ | Fundamental frequency | Hz |
| CPP | Cepstral peak prominence | dB |
| ACFL | Peak-to-peak of the AC glottal airflow waveform | mL/s |
| MFDR | Maximum flow declination rate: Negative peak of the first derivative of the glottal waveform | $L/s^2$ |
| OQ | Open quotient: Ratio of the open time of the glottal vibratory cycle to the corresponding cycle period | % |
| SQ | Speed quotient: Ratio of the opening time of the glottis to the closing time | % |
| $L_1$–$L_2$ | Difference between the log-magnitude of the first two harmonics | dB |
| HRF | Harmonic richness factor: Ratio of the sum of magnitudes of second harmonic and above to the magnitude of the first harmonic | dB |
| NAQ | Normalized amplitude quotient: Ratio of ACFL to MFDR divided by the glottal period ($1/f_o$) | |

(minimum silent interval = 25 ms, minimum sounding interval = 50 ms). Fig. 2(A) displays the resulting TextGrid of labeled vowel segment and silent interval boundaries. Boundaries for the first and last plosive of each breath group were verified visually to create a trial label for each combination of pitch, vowel, and phonatory conditions.

*1) Reference Subglottal Pressure Estimation:* Boundaries of each intervocalic IOP pulse were detected automatically using a custom algorithm (Fig. 2(B)). The IOP signal was low-pass filtered with a fifth-order Butterworth filter (80 Hz 3-dB cutoff frequency) to remove harmonic information that might confound the boundary determination. Next, the silent interval boundaries were extended by 25% to the left and right, resulting in IOP pulse boundaries that compensated for the slight overlap between the preceding vowel segment and the rise of the subsequent IOP signal.

The IOP plateaus were defined as the peak amplitude of each IOP pulse. Estimates of Ps for each vowel segment were determined by computing the mean of the IOP pulse peak amplitudes preceding and following each vowel. Alignment of the smartphone-recorded ACC signal was achieved using a custom algorithm in MATLAB that shifted the accelerometer signal (up-sampled to the acoustic sampling rate of 20 kHz) such that the absolute value of the cross-correlation between the two signals was maximized.

*2) Inverse Filtering of the Oral Airflow Signal:* Fig. 4 illustrates an example airflow waveform and its inverse filtered counterpart. A common inverse filtering technique was applied to the oral airflow signal to cancel out the effects of the first formant and estimate the glottal airflow waveform from which measures were extracted to characterize the glottal volume velocity voicing source [20]. For each vowel segment, the oral airflow signal was lowpass filtered at 1100 Hz due to the bandwidth of the pneumotachograph mask, which exhibited an antiresonance at 1500 Hz. Then, a single-notch filter (a conjugate pair of zeros with unity gain at DC) was applied to reduce waveform ripple during the glottal closed phase without the need for closed-phase detection. The center frequency of the filter was swept from 200 Hz to 1000 Hz in 1 Hz steps (filter bandwidth was fixed at 70 Hz). Each single-notch filter was applied to each vowel waveform. The optimal center frequency was determined when the following expression was minimized: $\sum_{n=0}^{N-1} |\Delta^2 x_{IF}[n]|$, where $x_{IF}[n]$ is the inverse-filtered waveform at sample $n$, and $N$ is the number of samples in the vowel segment. The associated glottal airflow waveform was chosen for further parameterization.

*3) Subglottal Impedance-Based Inverse Filtering of the Accelerometer Signal:* Subglottal impedance-based inverse filtering (IBIF) was applied to the same vowel segments to estimate glottal airflow measures from the accelerometer signal [44]. This estimation was optimized on a per-segment basis, i.e., optimized within each vowel segment. Although computationally expensive, this was to provide the best possible *automated* IBIF, in lieu of applying a single IBIF inverse filter to all vowel segments per subject. The parameter space given by the skin model and tracheal geometry was adjusted to minimize the error between oral airflow and accelerometer-based glottal volume velocity waveforms [20], [58]. Five parameters were estimated for each subject—three parameters for a skin model (skin inertance, resistance, and stiffness) and two parameters for tracheal geometry (tracheal length and accelerometer position relative to the glottis). The waveforms were aligned, and model properties were obtained via particle swarm optimization, a constrained multivariate optimization procedure [44], [58].

*4) Accelerometer-Based Features:* Table I lists the accelerometer- and glottal airflow–based measures that were used in multiple linear regression models to better estimate average Ps. Fig. 4 illustrates the parameterization of the oral airflow signal

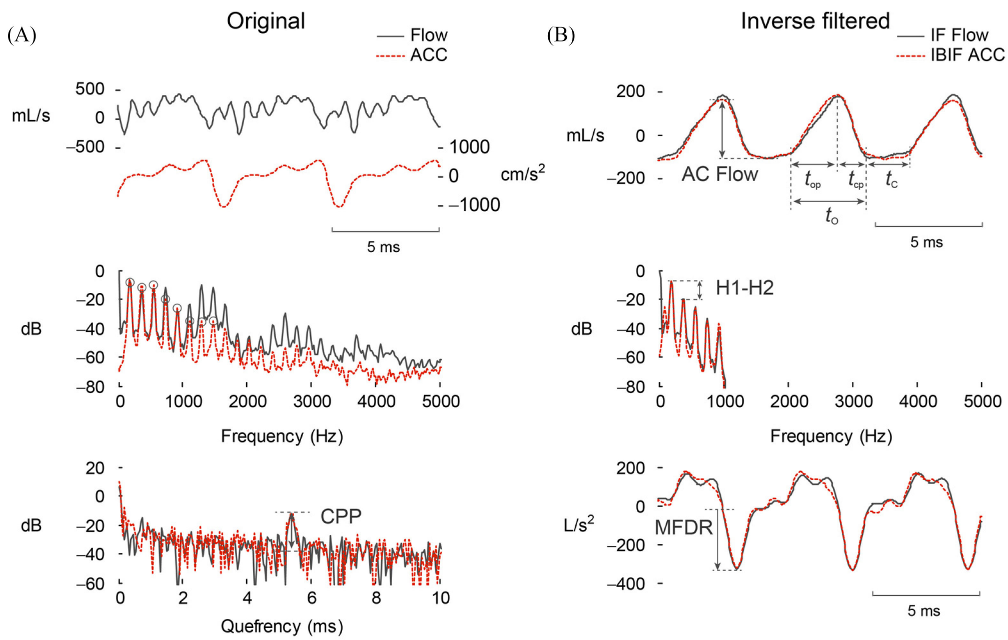


Fig. 4. Parameterization of the (A) original and (B) inverse-filtered waveforms from the oral airflow (black) and neck-surface acceleration (ACC, red-dashed) waveform processed with subglottal impedance-based inverse filtering (IBIF). Adapted from [42].

before and after inverse filtering and the accelerometer signal before and after IBIF.

The first set of measures quantifies accelerometer signal properties related to RMS amplitude, $f_o$ [55], and cepstral peak prominence (CPP) [43]. In particular, accelerometer-based CPP has been shown to correlate highly with acoustic-based CPP [31], which is often used as an indicator of breathiness [59] and overall dysphonia [60], [61]. Data in Fig. 1 illustrate the potential of CPP to categorize the modal (27.2 dB), breathy (21.4 dB), and rough (14.9 dB) voice qualities and thus to potentially act as a strong factor in the accelerometer-based Ps prediction equation.

The second set of measures was extracted from the glottal airflow waveform derived from the neck-surface accelerometer signal using IBIF [43], [44]: peak-to-peak flow (ACFL), maximum flow declination rate (MFDR), open quotient (OQ), speed quotient (SQ), spectral slope ($L_1$–$L_2$) [33], harmonic richness factor (HRF), and normalized amplitude quotient (NAQ). OQ is defined as $t_O/(t_O + t_C)$, and SQ is defined as $100(t_{op}/t_{cp})$. NAQ is a measure of the closing phase and is defined as the ratio of AC Flow to MFDR normalized by the period duration $(t_O + t_C)$ [62].

## *E. Stepwise Linear Regression Modeling*

Subject-specific, linear regression models were constructed using accelerometer signal RMS alone and in combination with the additional accelerometer-based measures to estimate Ps across vowel, pitch, and voice quality contexts. Glottal airflow measures from inverse filtering the oral airflow waveform (IF) were initially added to the regression models to assess whether IBIF exhibited any significant change in performance. Cross-validation assessed the robustness of model performance using the root-mean-square error (RMSE) metric for each regression model.

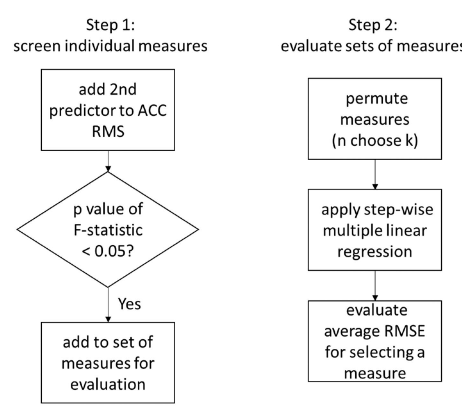


Fig. 5. Two-step process to select accelerometer-based measures for improving the prediction of Ps. Step 1 includes creating a two-predictor linear regression model and testing the statistical significance of the additional measure's impact on RMSE. Step 2 includes creating permuted subsets of the screened-in measures to determine the benefit of k additional measures.

First, CPP, $f_o$, and IBIF measures were screened for potential contribution to improve Ps prediction. As shown in Fig. 5, each additional measure was added to the baseline regression model (accelerometer RMS as predictor of Ps) to create a two-predictor linear regression model. If a measure were sufficiently useful—i.e., selected by MATLAB R2018b's multilinear regression function *stepwisefit* (Statistics and Machine Learning Toolbox version 11.4) to be included in the two-predictor linear regression model—the measure was "screened-in" to a set of measures for follow-up evaluation. Second, the set of screened-in measures were permuted and added to the final multiple linear regression model to determine their collective utility for Ps prediction. Alternatives to this two-step selection process include all-possible-regressions, ridge regression, and lasso regression.

The relatively small number of measures considered here suggests that all-possible-regressions was at least computationally feasible. However, a first step of screening individual measures was performed before considering the performance of permuted subsets of measures to gain insight into the potential impact of individual measures.

*1) Step 1. Screening Individual Measures:* MATLAB's *stepwisefit* function was used to create and evaluate the baseline regression model with one additional measure, i.e., accelerometer RMS and one additional measure as co-predictors of Ps. The *stepwisefit* function was configured to consider the inclusion of a second predictor by adding the measure from a multilinear model based on its statistical significance in improving Ps prediction. The *p*-value of an F-statistic was computed to test models with and without the second predictor. The null hypothesis was that the second predictor would have a zero coefficient if added to the model. If there were statistically significant evidence to reject the null hypothesis ($p < 0.05$), the second measure was added to the model, i.e., screened in.

*2) Step 2. Evaluating n-Choose-k Sets of Measures:* After each of the additional measures was screened in according to the procedure in Step 1, the entire set of screened-in measures was permuted and evaluated based on an n-choose-k subset for Ps prediction accuracy. This step was designed to demonstrate the potential for Ps prediction improvement by modeling with an arbitrary subset of the additional measures. A multiple linear regression model was built with glottal flow measures derived from IBIF measures, and the model's Ps prediction performance was compared with one built with IF measures. The two non-IF measures—$f_o$ and CPP—were added to the models to demonstrate their utility as well.

Fig. 6 shows the order in which accelerometer CPP and $f_o$ were introduced into the *stepwisefit* function. Since the non-IF measures were computationally inexpensive compared with the inverse-filtered glottal flow measures and showed high inclusion frequency, CPP and $f_o$ were fixed to be always included before the inverse-filtered measures. The order in which the remaining glottal flow measures were introduced into the *stepwisefit* function was fully permuted because the order in which the measures presented to the *stepwisefit* function can affect the inclusion and exclusion of a subsequent parameter (for example, the second of two highly correlated measures was expected to be excluded if the first were included). Therefore, each permutation of the glottal flow measures, or ordered sequence thereof, was presented to the *stepwisefit* function. The resulting RMSE was averaged across all possible permutations to obtain an average change in ($\Delta$) RMSE to quantify the gain in incorporating accelerometer-based measures in addition to RMS amplitude.

A five-fold validation was performed for each of the permuted set of accelerometer-based measures in Table I selected as co-predictors along with RMS amplitude. Within each of the five folds, a training portion comprising 80% of the vowel tokens was used to construct a linear regression model of accelerometer RMS and the additional measures as predictors of Ps, and 20% of the remaining vowel tokens was used to calculate RMSE between the predicted and reference estimates of Ps.

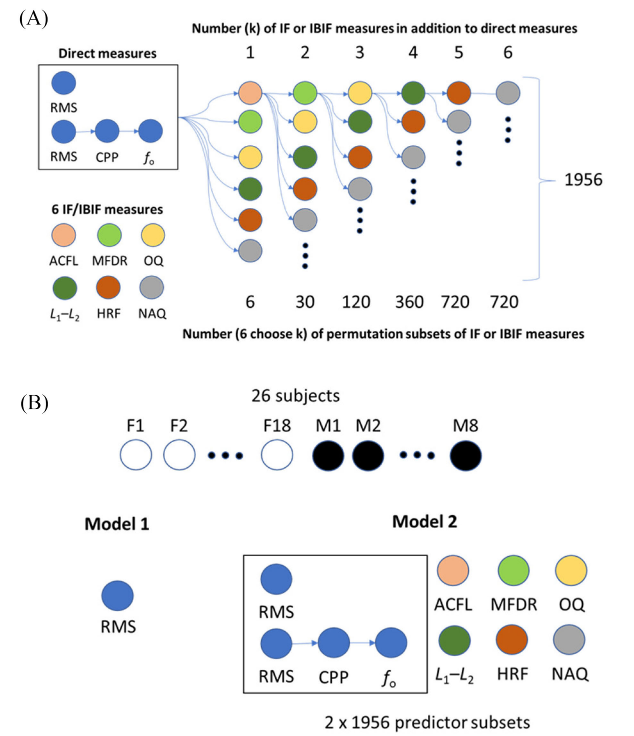


Fig. 6. Creation of n-choose-k permutation subsets of the screened-in IF/IBIF measures for evaluation. (A) Direct measures comprising either accelerometer RMS alone or accelerometer RMS, CPP, and $f_o$ with a permuted sequence of 1–6 screened-in IF/IBIF measures created predictor sets. (B) For each of the 26 subjects, prediction performance is evaluated for accelerometer RMS alone (Model 1) and each of $2 \times 1956$ predictor sets that each constructed a multiple linear regression model (Model 2).

## III. RESULTS

Across all subjects, baseline RMSE performance for predicting Ps for modal-only phonation using accelerometer RMS only was found to be 1.7 cm $H_2O$ on average. When non-modal phonation was added to the modal data points, each fold of the five-fold cross-validation exhibited an increase in RMSE when accelerometer RMS–alone models were used to predict Ps. Improvements to model performance (decreases in RMSE) were found when CPP, $f_o$, and glottal airflow measures of vocal function were included in the model. Critically, similar model performance was achieved when the same flow-based IF measures were derived from the accelerometer signal using IBIF, thus showing promise for accelerometer-only prediction of Ps for modal and non-modal phonation.

*1) Step 1. Screening Individual Measures:* Table II shows the frequency of how often each additional measure was selected for prediction of Ps along with the baseline predictor accelerometer RMS. CPP was included 72% of the time for female subjects and 88% of the time for male subjects (77% of the time across all subjects). $f_o$ was selected 67%, 38%, and 58% percent of the time for the female, male, and combined group, respectively. ACFL, MFDR, $L_1$–$L_2$, and HRF from both IF and IBIF signals occur with relatively high frequency across the 26 subjects. There appear to be sex-based differences for the selection of

TABLE II
INCLUSION FREQUENCY (%) WITHIN SUBJECT GROUPS OF ACCELEROMETER-BASED MEASURES INTO MULTIPLE REGRESSION MODEL FOR PREDICTION OF PS

| Group | Direct | | Oral airflow–based IF measure | | | | | | | Accelerometer–based IBIF measure | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CPP | $f_o$ | ACFL | MFDR | OQ | SQ | $L_1$–$L_2$ | HRF | NAQ | ACFL | MFDR | OQ | SQ | $L_1$–$L_2$ | HRF | NAQ |
| Female (n=18) | 72 | 67 | 83 | 78 | 50 | 28 | 78 | 67 | 56 | 78 | 78 | 61 | 6 | 72 | 78 | 78 |
| Male (n=8) | 88 | 38 | 63 | 75 | 88 | 50 | 75 | 88 | 75 | 63 | 75 | 63 | 13 | 75 | 75 | 50 |
| All (n=26) | 77 | 58 | 73 | 81 | 81 | 35 | 77 | 73 | 62 | 73 | 77 | 62 | 8 | 73 | 77 | 69 |

$f_o$ and NAQ (both higher in female subjects). The inclusion frequency of SQ was low for both IF and IBIF measures. Based on this screening step, CPP and $f_o$ were screened in. SQ was screened out and not included in any subsequent model.

*2) Step 2. Adding IF and IBIF Measures in Permuted Subsets:* For each of the five cross-validation folds, an RMSE value was calculated for the baseline regression model (accelerometer RMS only) and for multiple regression models using accelerometer RMS combined with a permuted sequence of additional measures (see Fig. 6). This cross-validation was performed for prediction of Ps during modal phonation and repeated for prediction of Ps during all phonatory conditions (modal, breathy, strained, and rough). An average RMSE per subject was computed across all folds and subsequently across all 26 subjects.

Fig. 7 illustrates the improvement in RMSE for one cross-validation fold of one of the subjects. The RMSE decreased from 4.1 cm $H_2O$ to 2.9 cm $H_2O$ (29.3% reduction) when additional IF measures were selected to create a multiple linear regression model to predict Ps (Fig. 7(A)). RMSE decreased similarly from 4.1 cm $H_2O$ to 3.1 cm $H_2O$ (24.4% reduction) when accelerometer-based measures derived using IBIF were selected (Fig. 7(B)).

Fig. 8 displays box-whisker plots of a grand-average RMSE as additional measures are added to the subject-specific regression models. For each permuted sequence of additional measures, the grand-average RMSE was calculated—first across the five folds of model building-testing runs per subject and then across the 26 subjects. Statistics of the grand-average RMSE were then accumulated across all permutations of a given sequence length. For example, the box-whisker plot for one additional IF/IBIF measure included six models that each included accelerometer RMS plus one of the six inverse-filtered measures from Table II (recall SQ never included). Permuting two IF/IBIF measures refers to prediction performance of 30 regression models (6-choose-2 permutations).

In Fig. 8(A) and Fig. 8(B), when one to six IF/IBIF measures were added to accelerometer RMS to build the multiple linear regression model for the prediction of Ps, the grand-average RMSE was lowered progressively from 2.9 cm $H_2O$ as the number of additional measures were included. For IF measures, the RMSE plateaued around 2.5 cm $H_2O$ when all six additional measures were included. For IBIF measures, the RMSE plateaued around 2.6 cm $H_2O$ when all six additional measures were included.

In Fig. 8(C) and Fig. 8(D), one to six IF/IBIF measures were added to a new baseline model that included accelerometer

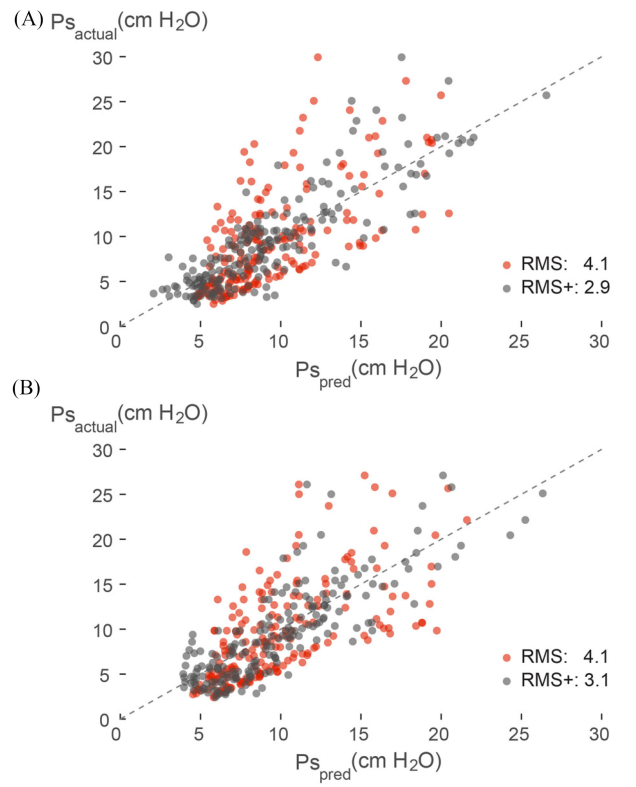


Fig. 7. Example of one cross-validation fold plotting reference Ps ($Ps_{actual}$) against predicted Ps ($Ps_{pred}$) for one subject. The root-mean-square error improves when adding glottal airflow measures from either (A) inverse-filtered oral airflow or (B) accelerometer IBIF signal.

RMS, CPP, and $f_o$. The grand-average RMSE decreased progressively from 2.9 cm $H_2O$ as the number of additional measures were included. For IF measures, the grand-average RMSE plateaued at approximately 2.4 cm $H_2O$. For IBIF measures, the average RMSE plateaued at approximately 2.5 cm $H_2O$ when all six additional measures were included.

Table III reports results of the multiple regression model performance from a subject-specific point of view. Shown is the improvement in Ps prediction performance in terms of RMSE for each subject comparing the accelerometer RMS-only model (Model 1) with multiple regression models (Model 2) incorporating CPP, $f_o$, and glottal airflow measures derived from the IF oral airflow waveform or from the IBIF accelerometer signal. This table indicates that the mean (standard deviation,

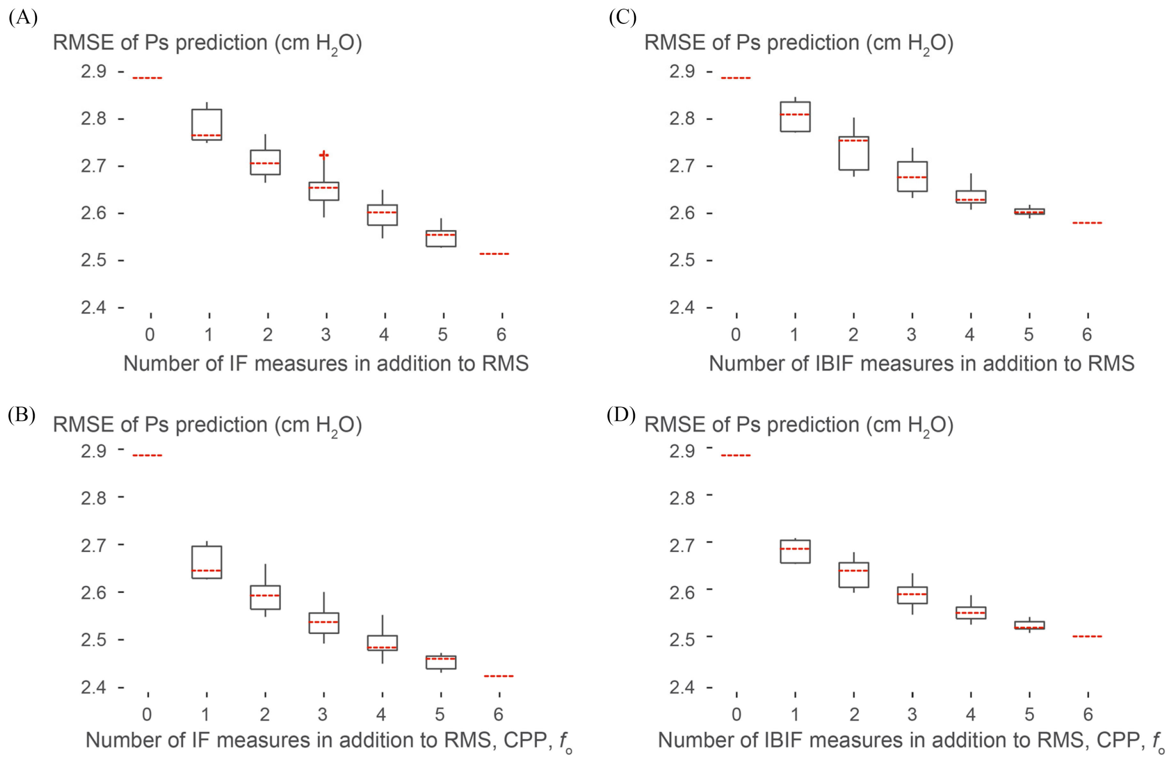


Fig. 8. Decrease in average root-mean-square error (RMSE) across all permutations of additional measures added to the subject-specific regression models. RMSE of Ps predictions are shown in A) by using accelerometer RMS and permuted subsets of 1–6 IF measures as co-predictors; in B) by using accelerometer RMS and permuted subsets of 1–6 IBIF measures as co-predictors; in C) by using accelerometer RMS, CPP, $f_o$ and permuted subsets of 1–6 IF measures as co-predictors; and in D) by using accelerometer RMS, CPP, $f_o$ and permuted subsets of 1–6 IBIF measures as co-predictors. In each plot, RMSE at 0 denotes the RMSE of using accelerometer RMS alone as the predictor of Ps.

SD) reduction in RMSE for the accelerometer-based multiple regression model is 12.5% (6.7 percentage points). This is compared with the mean (SD) reduction in RMSE when using oral airflow–based IF measures in the regression models of 15.0% (9.4 percentage points). There is variation in performance from subject to subject, with RMSE reduction as high as 25.1% (subject M1).

## IV. Discussion

The objective of this work was to develop a methodology for the improved prediction of Ps that incorporates accelerometer-based measures of vocal function to achieve improved prediction of Ps during non-modal phonation. The hypothesis was that the RMS amplitude of neck-surface vibration would not be enough to accurately predict Ps, especially in context of non-modal phonation such as that exhibited by speakers producing breathy, strained, and rough voice qualities. In this study, vocally healthy speakers were recruited to volitionally produce these different voice qualities. The advantage of this study design was to allow each subject to act as his or her own control to minimize across-subject variations in voice physiology and neck morphology.

A previous analysis of ten vocally healthy speakers producing modal phonation yielded an average 95% prediction interval of ±2.5 cm $H_2O$ when accelerometer signal RMS was the predictor variable for Ps estimation [45]. The corresponding accelerometer RMS–only Ps prediction performance in the current study of 26 vocally healthy speakers yielded an average RMSE of 0.7 cm $H_2O$. This error increased to 2.9 cm $H_2O$ when the non-modal phonatory conditions were included. To counteract the increase in error, a regression model with additional measures was proposed to improve Ps prediction performance. The measures were selected for their ability to be derived from the neck-surface accelerometer signal and to reflect changes in association glottal conditions.

The final subject-specific regression models incorporated nine measures, including three measures computed directly from the accelerometer signal (RMS amplitude, CPP, and $f_o$) and six measures parameterizing an estimate of the glottal airflow waveform (ACFL, MFDR, OQ, $L_1$–$L_2$, HRF, and NAQ). SQ did not contribute significantly to improved model performance. Estimates of the glottal airflow were computed using the IBIF algorithm that was optimized per vowel token [44]. RMSE decreased to 2.5 cm $H_2O$ with the final set of measures when averaged across all subjects. For certain subjects, RMSE decreased by up to 25% (e.g., from 4.1 cm $H_2O$ to 3.1 cm $H_2O$). In other subjects, the performance gain was not as dramatic (see Table III).

Performance comparisons were made between IBIF-derived glottal airflow measures and conventional glottal airflow measures derived from inverse filtering the oral airflow waveform. In general, although not exactly the same, Ps prediction performance was similar when computing the glottal airflow measures using IBIF. Thus, the IBIF algorithm did not introduce

TABLE III

IMPROVEMENT IN PS PREDICTION PERFORMANCE IN TERMS OF ROOT-MEAN-SQUARE ERROR (RMSE) FOR EACH SUBJECT COMPARING THE ACCELEROMETER RMS-ONLY MODEL (MODEL 1) WITH MULTIPLE REGRESSION MODELS (MODEL 2) INCORPORATING CPP, $f_o$, AND GLOTTAL AIRFLOW MEASURES DERIVED FROM THE IF ORAL AIRFLOW WAVEFORM OR FROM THE IBIF ACCELEROMETER SIGNAL. CHANGE IN ($\Delta$) RMSE ALSO REPORTED IN CM $H_2O$ AND AS A PERCENTAGE

| Subject ID | Oral airflow–based IF measures | | | | Accelerometer–based IBIF measures | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 RMSE | Model 2 RMSE | ΔRMSE (cm $H_2O$) | ΔRMSE (%) | Model 1 RMSE | Model 2 RMSE | ΔRMSE (cm $H_2O$) | ΔRMSE (%) |
| F1 | 1.75 | 1.71 | −0.04 | −2.40 | 1.75 | 1.68 | −0.07 | −4.09 |
| F2 | 1.43 | 1.28 | −0.15 | −10.71 | 1.43 | 1.23 | −0.20 | −14.13 |
| F3 | 1.68 | 1.36 | −0.33 | −19.40 | 1.68 | 1.35 | −0.34 | −19.90 |
| F4 | 2.38 | 2.18 | −0.20 | −8.31 | 2.37 | 2.23 | −0.14 | −6.01 |
| F5 | 1.97 | 1.82 | −0.15 | −7.55 | 1.97 | 1.66 | −0.31 | −15.51 |
| F6 | 2.96 | 2.42 | −0.54 | −18.28 | 2.97 | 2.62 | −0.35 | −11.72 |
| F7 | 2.94 | 2.62 | −0.32 | −10.97 | 2.94 | 2.77 | −0.17 | −5.70 |
| F8 | 1.91 | 1.63 | −0.28 | −14.80 | 1.91 | 1.78 | −0.13 | −6.89 |
| F9 | 1.49 | 1.39 | −0.09 | −6.30 | 1.49 | 1.44 | −0.05 | −3.36 |
| F10 | 2.57 | 2.34 | −0.23 | −8.88 | 2.57 | 2.40 | −0.17 | −6.60 |
| F11 | 2.98 | 2.55 | −0.43 | −14.43 | 2.98 | 2.20 | −0.77 | −25.90 |
| F12 | 2.17 | 1.87 | −0.30 | −13.93 | 2.17 | 1.99 | −0.18 | −8.10 |
| F13 | 3.09 | 2.70 | −0.39 | −12.52 | 3.09 | 2.79 | −0.30 | −9.72 |
| F14 | 2.07 | 2.00 | −0.07 | −3.25 | 2.07 | 2.02 | −0.05 | −2.39 |
| F15 | 2.61 | 2.27 | −0.34 | −13.13 | 2.61 | 2.32 | −0.29 | −11.18 |
| F16 | 4.68 | 3.72 | −0.95 | −20.41 | 4.67 | 4.09 | −0.59 | −12.53 |
| F17 | 7.38 | 6.33 | −1.05 | −14.26 | 7.38 | 6.43 | −0.94 | −12.77 |
| F18 | 2.01 | 1.77 | −0.24 | −11.92 | 2.01 | 1.74 | −0.27 | −13.59 |
| M1 | 4.15 | 2.91 | −1.23 | −29.75 | 4.15 | 3.11 | −1.04 | −25.09 |
| M2 | 3.34 | 2.98 | −0.36 | −10.77 | 3.34 | 3.01 | −0.33 | −9.76 |
| M3 | 2.35 | 1.71 | −0.64 | −27.34 | 2.35 | 1.90 | −0.46 | −19.35 |
| M4 | 4.36 | 3.24 | −1.12 | −25.74 | 4.36 | 3.41 | −0.95 | −21.80 |
| M5 | 2.21 | 1.33 | −0.88 | −39.74 | 2.21 | 1.71 | −0.49 | −22.39 |
| M6 | 4.21 | 3.73 | −0.48 | −11.30 | 4.19 | 3.51 | −0.68 | −16.31 |
| M7 | 2.68 | 2.62 | −0.07 | −2.43 | 2.68 | 2.49 | −0.19 | −7.20 |
| M8 | 3.72 | 2.51 | −1.21 | −32.44 | 3.72 | 3.24 | −0.48 | −12.82 |
| Mean | 2.89 | 2.42 | −0.47 | −15.04 | 2.89 | 2.51 | −0.38 | −12.49 |
| SD | 1.29 | 1.05 | 0.37 | 9.42 | 1.29 | 1.08 | 0.29 | 6.72 |
| Minimum | 1.43 | 1.28 | −1.23 | −39.74 | 1.43 | 1.23 | −1.04 | −25.90 |
| Maximum | 7.38 | 6.33 | −0.04 | −2.40 | 7.38 | 6.43 | −0.05 | −2.39 |

significant noise in the processing and, as expected, yielded measures that were good surrogates of IF-derived measures.

Ongoing work continues to demonstrate the need for novel clinically salient measures derived from the ambulatory accelerometer signal; e.g., average ambulatory estimates of sound pressure level and $f_o$ do not differentiate between patients with phonotraumatic lesions and matched healthy control subjects [63]. The results of the current study suggest that the error in estimating Ps in the laboratory setting is low enough such that Ps can be added to the suite of ambulatory voice measures that can be reliably derived from the neck-surface vibration signal. For example, the reduction in Ps prediction error becomes clinically meaningful when the error is low relative to differences in Ps ($\sim$4–5 cm $H_2O$) that have been found between patients with voice disorders and typical speakers, e.g., phonotraumatic vocal hyperfunction compared to vocally healthy speakers [20].

Airflow interruption techniques are limited to estimating Ps during isolated vowel contexts [28], [29]. Even though such Ps estimates provide valuable information about vocal function, the information is inherently limited by uncertainties about how well the measures reflect glottal function during natural speech where pitch, loudness, and rate of speech vary rapidly. The neck-surface accelerometer approach addresses these shortcomings by using an inexpensive, non-invasive sensor that can unobtrusively monitor Ps during natural speech. In practice, accelerometer-based Ps estimation requires an initial baseline model calibration with the oral airflow interruption technique, followed by the application of the subject-specific model to predict Ps during unconstrained, natural speech production. The unobtrusive accelerometer sensor can then be affixed to a speaker's neck for laboratory, clinical, and ambulatory assessment of vocal function, with the added potential of integration into smartphone applications for ease of use [43], [55]. Monitoring Ps as individuals go about their daily activities may provide clinicians with additional insight into a person's typical vocal functioning, including the potential to provide an objective measure of vocal effort in real-world environments [17], [19], [64].

Subglottal neck-surface vibration has been modeled as the output of the downward-traveling dipole voice source filtered by subglottal resonances and the transfer function between intra-tracheal acoustic pressure and the neck frequency response [44]. Thus, the glottal airflow waveform has been shown to be derivable from the neck-surface vibration signal, yielding AC voice signal properties such as MFDR, AC flow, OQ, etc. It may be surprising that information about a DC vocal function measure (mean Ps) can be derived reliably from an AC-only signal (neck-surface vibration). Strong associations have also been found between the DC signal property of mean Ps and the AC signal property of MFDR derived from the estimated glottal airflow waveform in vocally healthy speakers [65]. Consequently, for the same vowel, MFDR is known to correlate highly with acoustic SPL [66]. Following on that relationship, a first-order estimate of acoustic SPL (sound radiation from the oral opening) has traditionally been obtained using the magnitude of the neck-surface vibration signal measures by a contact microphone or accelerometer [46]. However, care must be taken to derive acoustic SPL from the subglottal neck-surface vibration signal when multiple vowel contexts are taken into account because of the impact of different vowel formant frequencies on radiated sound from the mouth. Estimating mean Ps from neck-surface vibration yields a lower uncertainty due to the subglottal placement of the accelerometer, which is minimally influenced by supraglottal vowel formants [45].

As phonation becomes non-standard, or non-modal, underlying assumptions and relationships among vocal function measures may be significantly affected. For example, pressed phonation has been shown to yield lower MFDR values for similar mean Ps values relative to the modal phonatory condition [51]. The relationship between neck-surface vibration magnitude and mean Ps is analogously affected by pressed/strained, breathy, and rough voice qualities; i.e., higher mean Ps values have been observed for the same accelerometer RMS values [49]. The current study follows on these past studies by incorporating MFDR and other source-related voice measures that can be estimated accurately from the neck-surface accelerometer signal to improve upon the baseline prediction of Ps. The RMSE improvement was expected to approach the average baseline RMSE (1.7 cm $H_2O$) exhibited when predicting Ps in the modal-only condition using accelerometer RMS signal amplitude. Instead, the average RMSE across subjects plateaued at 2.6 cm $H_2O$, indicating that alternative strategies for compensating for non-modal characteristics are areas of further investigation.

It is acknowledged that the task of artificially producing non-modal voice characteristics does not mimic voice qualities produced during naturalistic speech contexts nor reflects the non-modal behavior exhibited by patients with voice disorders. The extreme behaviors elicited thus may have created a complex situation in which Ps prediction performance was overly challenging (and, perhaps, unrealistic). Also, although the oral airflow interruption method has been validated using direct measurements of Ps [67], [68], limited evidence exists for the validity of similar estimation of Ps in the context of non-modal voice production. Even in modal voice, over-estimation and under-estimation of the true mean Ps (directly measured via tracheal puncture) by the oral airflow interruption method has been reported in vocally healthy speakers [21], with larger Ps estimation errors observed during loud phonatory conditions [69]. Less information is available for individuals with voice disorders; individuals with spasmodic dysphonia have been studied, yielding inconsistent results for the indirect estimation of Ps [21]. Thus, as usual, caution is suggested when interpreting absolute values of mean Ps obtained using indirect methods.

Future work is needed to study patients with voice disorders to provide evidence that accelerometer-based estimation of Ps is feasible, valid, and accurate for clinical use. Indeed, with higher degrees of dysphonia (such as the rough phonatory condition), many measures of vocal function may become unreliable, including basic metrics such as fundamental frequency. The collection of data from patients with voice disorders is necessary to investigate the sources of error and applicability of the technique to monitoring treatment (before and after surgery, or longitudinal progress over the course of multiple therapy sessions).

## V. Conclusion

Improved estimation of subglottal pressure from neck-surface vibration during non-modal phonation is achievable by incorporating accelerometer-based measures of cepstral peak prominence, fundamental frequency, and of the subglottal impedance-based inverse filtered waveform. This non-invasive method for estimating Ps during natural speech should next be studied in the context of the clinical assessment of voice disorders, particularly for application to ambulatory monitoring and biofeedback as individuals go about their usual activities at home, work, and social settings.

## Acknowledgment

The authors would like to thank R. E. Hillman for his guidance and clinical perspective for this study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## 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, pp. 1988–1995, 2005.
[2] N. Bhattacharyya, "The prevalence of voice problems among adults in the United States," *Laryngoscope*, vol. 124, pp. 2359–2362, 2014.
[3] NIDCD, *2012-2016 Strategic Plan*. National Institute on Deafness and Other Communication Disorders (NIDCD), U.S. Department of Health and Human Services Bethesda, MD, USA, 2012.
[4] R. Speyer, "Effects of voice therapy: A systematic review," *J. Voice*, vol. 22, pp. 565–580, 2008.
[5] D. M. Hartl, S. Hans, J. Vaissière, M. Riquet, and D. F. Brasnu, "Objective voice quality analysis before and after onset of unilateral vocal fold paralysis," *J. Voice*, vol. 15, pp. 351–361, 2001.
[6] E. B. Holmberg, P. Doyle, J. S. Perkell, B. Hammarberg, and R. E. Hillman, "Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy," *J. Voice*, vol. 17, pp. 269–282, 2003.
[7] S. M. Zeitels, R. E. Hillman, R. A. Franco, and G. W. Bunting, "Voice and treatment outcome from phonosurgical management of early glottic cancer," *Ann. Otol., Rhinol., Laryngol.*, vol. 111, no. Suppl. 190, pp. 1–20, 2002.
[8] S. M. Zeitels, I. Hochman, and R. E. Hillman, "Adduction arytenopexy: A new procedure for paralytic dysphonia and the implications for implant medialization," *Ann. Otol., Rhinol. Laryngol.*, vol. 107, no. Suppl. 173, pp. 1–24, 1998.
[9] S. M. Zeitels, G. Lopez-Guerra, J. A. Burns, M. Lutch, A. M. Friedman, and R. E. Hillman, "Microlaryngoscopic and office-based injection of bevacizumab (Avastin) to enhance 532-nm pulsed KTP laser treatment of glottal papillomatosis," *Ann. Otol., Rhinol., Laryngol.*, vol. 118, no. Suppl. 201, pp. 1–13, 2009.

[10] J. Jiang and J. Stern, "Receiver operating characteristic analysis of aerodynamic parameters obtained by airflow interruption: A preliminary report," *Ann. Otol., Rhinol., Laryngol.*, vol. 113, pp. 961–966, 2004.

[11] 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 Hearing Res.*, vol. 32, pp. 373–392, 1989.

[12] R. H. Colton, J. K. Casper, and R. J. Leonard, *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. Baltimore, MD, USA: Lippincott Williams & Wilkins, 2006.

[13] S. Björklund and J. Sundberg, "Relationship between subglottal pressure and sound pressure level in untrained voices," *J. Voice*, vol. 30, pp. 15–20, 2016.

[14] I. Titze, "Quantifying vocal efficiency and economy - how can computation augment clinical assessment?" in *Proc. Meetings Acoust.*, 2013, vol. 19, Art. no. 060244.

[15] I. R. Titze, "Vocal efficiency," *J. Voice*, vol. 6, pp. 135–138, 1992.

[16] I. R. Titze, L. Maxfield, and A. Palaparthi, "An oral pressure conversion ratio as a predictor of vocal efficiency," *J. Voice*, vol. 30, pp. 398–406, 2016.

[17] A. L. Rosenthal, S. Y. Lowell, and R. H. Colton, "Aerodynamic and acoustic features of vocal effort," *J. Voice*, vol. 28, pp. 144–153, 2014.

[18] L. O. Ramig and C. Dromey, "Aerodynamic mechanisms underlying treatment-related changes in vocal intensity in patients with Parkinson disease," *J. Speech Hearing Res.*, vol. 39, pp. 798–807, 1996.

[19] V. S. McKenna and C. E. Stepp, "The relationship between acoustical and perceptual measures of vocal effort," *J. Acoust. Soc. Amer.*, vol. 144, pp. 1643–1658, 2018.

[20] V. M. Espinoza, M. Zañartu, J. H. Van Stan, D. D. Mehta, and R. E. Hillman, "Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction," *J. Speech, Lang., Hearing Res.*, vol. 60, pp. 2159–2169, 2017.

[21] R. L. Plant and A. D. Hillel, "Direct measurement of subglottic pressure and laryngeal resistance in normal subjects and in spasmodic dysphonia," *J. Voice*, vol. 12, pp. 300–314, 1998.

[22] J. Sundberg, R. Scherer, M. Hess, F. Müller, and S. Granqvist, "Subglottal pressure oscillations accompanying phonation," *J. Voice*, vol. 27, pp. 411–421, 2013.

[23] J. van den Berg, "Direct and indirect determination of the mean subglottic pressure: Sound level, mean subglottic pressure, mean air flow," subglottic power" and "efficiency" of a male voice for the vowel (a)," *Folia Phoniatrica*, vol. 8, pp. 1–24, 1956.

[24] B. Cranen and L. Boves, "Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production," *J. Acoust. Soc. Amer.*, vol. 77, pp. 1543–1551, 1985.

[25] S. Tanaka and W. J. Gould, "Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects," *J. Acoust. Soc. Amer.*, vol. 73, pp. 1316–1321, 1983.

[26] T. J. Hixon, "Some new techniques for measuring the biomechanical events of speech production: One laboratory's experiences," *Amer. Speech Hearing Assoc. Rep.*, vol. 7, pp. 68–103, 1972.

[27] H. K. Schutte, *The Efficiency of Voice Production*. Gröningen, The Netherlands: State University Hospital, 1980.

[28] M. Rothenberg, "A new inverse filtering technique for deriving glottal air flow waveform during voicing," *J. Acoust. Soc. Amer.*, vol. 53, pp. 1632–1645, 1973.

[29] J. Jiang, C. Leder, and A. Bichler, "Estimating subglottal pressure using incomplete airflow interruption," *Laryngoscope*, vol. 116, pp. 89–92, 2006.

[30] R. F. Coleman, "Comparison of microphone and neck-mounted accelerometer monitoring of the performing voice," *J. Voice*, vol. 2, pp. 200–205, 1988.

[31] D. Mehta, J. Van Stan, and R. Hillman, "Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 659–668, Apr. 2016.

[32] A. F. Llico *et al.*, "Real-time estimation of aerodynamic features for ambulatory voice biofeedback," *J. Acoust. Soc. Amer.*, vol. 138, pp. EL14–EL19, 2015.

[33] D. D. Mehta, V. M. Espinoza, J. H. V. Stan, M. Zañartu, and R. E. Hillman, "The difference between first and second harmonic amplitudes correlates between glottal airflow and neck-surface accelerometer signals during phonation," *J. Acoust. Soc. Amer.*, vol. 145, pp. EL386–EL392, 2019.

[34] H. E. Gunter, R. D. Howe, S. M. Zeitels, J. B. Kobler, and R. E. Hillman, "Measurement of vocal fold collision forces during phonation: Methods and preliminary data," *J. Speech, Lang., Hearing Res.*, vol. 48, pp. 567–576, 2005.

[35] R. Buekers, E. Bierens, H. Kingma, and E. Marres, "Vocal load as measured by the voice accumulator," *Folia Phoniatrica et Logopaedica*, vol. 47, pp. 252–261, 1995.

[36] A. Nacci *et al.*, "The use and role of the ambulatory phonation monitor (APM) in voice assessment," *Acta Otorhinolaryngologica Italica*, vol. 33, pp. 49–55, 2013.

[37] I. R. Titze, J. G. Švec, and P. S. Popolo, "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," *J. Speech, Lang., Hearing Res.*, vol. 46, pp. 919–932, 2003.

[38] I. R. Titze and E. J. Hunter, "Comparison of vocal vibration-dose measures for potential-damage risk criteria," *J. Speech, Lang., Hearing Res.*, vol. 58, pp. 1425–1439, 2015.

[39] P. Bottalico and A. Astolfi, "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms," *J. Acoust. Soc. Amer.*, vol. 131, pp. 2817–2827, 2012.

[40] F. Lindstrom, K. P. Waye, M. Södersten, A. McAllister, and S. Ternström, "Observations of the relationship between noise exposure and preschool teacher voice usage in day-care center environments," *J. Voice*, vol. 25, pp. 166–172, 2011.

[41] J. G. Švec, P. S. Popolo, and I. R. Titze, "Measurement of vocal doses in speech: Experimental procedure and signal processing," *Logopedics, Phoniatrics, Vocology*, vol. 28, pp. 181–192, 2003.

[42] 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, pp. 795–801, 2006.

[43] D. D. Mehta *et al.*, "Using ambulatory voice monitoring to investigate common voice disorders: Research update," *Frontiers Bioeng. Biotechnol.*, vol. 3, pp. 1–14, 2015.

[44] M. Zañartu, J. C. Ho, D. D. Mehta, R. E. Hillman, and G. R. Wodicka, "Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 9, pp. 1929–1939, Sep. 2013.

[45] A. S. Fryd, J. H. Van Stan, R. E. Hillman, and D. D. Mehta, "Estimating subglottal pressure from neck-surface acceleration during normal voice production," *J. Speech, Lang., Hearing Res.*, vol. 59, pp. 1335–1345, 2016.

[46] 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, pp. 1386–1394, 2005.

[47] B. R. Gerratt and J. Kreiman, "Toward a taxonomy of nonmodal phonation," *J. Phonetics*, vol. 29, pp. 365–381, 2001.

[48] G. B. Kempster, B. R. Gerratt, K. Verdolini Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol," *Amer. J. Speech-Lang. Pathol.*, vol. 18, pp. 124–132, 2009.

[49] K. L. Marks, J. Z. Lin, A. B. Fox, L. E. Toles, and D. D. Mehta, "Impact of nonmodal phonation on estimates of subglottal pressure from neck-surface acceleration in healthy speakers," *J. Speech, Lang., Hearing Res.*, vol. 62, pp. 3339–3358, 2019.

[50] V. S. McKenna, A. F. Llico, D. D. Mehta, J. S. Perkell, and C. E. Stepp, "Magnitude of neck-surface vibration as an estimate of subglottal pressure during modulations of vocal effort and intensity in healthy speakers," *J. Speech, Lang., Hearing Res.*, vol. 60, pp. 3404–3416, 2017.

[51] M. Millgård, T. Fors, and J. Sundberg, "Flow glottogram characteristics and perceived degree of phonatory pressedness," *J. Voice*, vol. 30, pp. 287–292, 2016.

[52] Z. Lei, E. Kennedy, L. Fasanella, N. Y.-K. Li-Jessen, and L. Mongeau, "Discrimination between modal, breathy and pressed voice for single vowels using neck-surface vibration signals," *Appl. Sci.*, vol. 9, 2019, Art. no. 1505.

[53] B. L. Perrine, R. C. Scherer, and J. A. Whitfield, "Signal interpretation considerations when estimating subglottal pressure from oral air pressure," *J. Speech, Lang., Hearing Res.*, vol. 62, pp. 1326–1337, 2019.

[54] E. U. Grillo and K. Verdolini, "Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects," *J. Voice*, vol. 22, pp. 546–552, 2008.

[55] D. 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.*, vol. 59, pp. 3090–3096, 2012.

[56] D. D. Mehta, M. Zañartu, J. H. Van Stan, S. W. Feng, H. A. Cheyne II, and R. E. Hillman, "Smartphone-based detection of voice disorders by long-term monitoring of neck acceleration features," in *Proc. IEEE Int. Conf. Body Sensor Netw.*, 2013, pp. 1–6.

[57] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer*, 5.3.39 ed. Amsterdam, The Netherlands, 2013. [Online]. Available: http://www.fon.hum.uva.nl/praat.

[58] J. P. Cortés *et al.*, "Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration," *PLoS One*, vol. 13, 2018, Art. no. e0209017.

[59] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *J. Speech Hearing Res.*, vol. 37, pp. 769–778, 1994.

[60] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," *J. Speech Hearing Res.*, vol. 39, pp. 311–321, 1996.

[61] S. N. Awan, N. Roy, M. E. Jetté, G. S. Meltzner, and R. E. Hillman, "Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V," *Clinical Linguistics Phonetics*, vol. 24, pp. 742–758, 2010.

[62] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," *J. Acoust. Soc. Amer.*, vol. 112, pp. 701–710, 2002.

[63] J. H. Van Stan, D. D. Mehta, S. M. Zeitels, J. A. Burns, A. M. Barbu, and R. E. Hillman, "Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects," *Ann. Otol., Rhinol., Laryngol.*, vol. 124, pp. 864–874, 2015.

[64] J. H. Van Stan, M. Maffei, M. L. V. Masson, D. D. Mehta, J. A. Burns, and R. E. Hillman, "Self-ratings of vocal status in daily life: Reliability and validity for patients with vocal hyperfunction and a normative group," *Amer. J. Speech-Lang. Pathol.*, vol. 26, pp. 1167–1177, 2017.

[65] J. Sundberg, "Flow glottogram and subglottal pressure relationship in singers and untrained voices," *J. Voice*, vol. 32, pp. 23–31, 2018.

[66] E. B. Holmberg, R. E. Hillman, J. S. Perkell, and C. Gress, "Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings," *J. Speech Hearing Res.*, vol. 37, pp. 484–495, 1994.

[67] A. Löfqvist, B. Carlborg, and P. Kitzing, "Initial validation of an indirect measure of subglottal pressure during vowels," *J. Acoust. Soc. Amer.*, vol. 72, pp. 633–635, 1982.

[68] S. Hertegård, J. Gauffin, and P.-A. Lindestad, "A comparison of subglottal and intraoral pressure measurements during phonation," *J. Voice*, vol. 9, pp. 149–155, 1995.

[69] M. A. McHenry, S. T. Kuna, J. T. Minton, and C. R. Vanoye, "Comparison of direct and indirect calculations of laryngeal airway resistance in connected speech," *J. Voice*, vol. 10, pp. 236–244, 1996.


**Jon Z. Lin** received the B.S. degree in engineering degree from Tufts University, Medford, MA, USA, in 2000, and the ALM degree in math and computation from Harvard University, Cambridge, MA, USA, in 2018.

He currently holds a Research Fellow position with the Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation, Boston, MA, USA.

**Víctor M. Espinoza** received the B.S. degree in engineering degree the Professional degree in acoustical engineering from Universidad Tecnológica Vicente Pérez Rosales, Santiago, Chile, in 1999 and 2001 and the Ph.D. degree in electronic engineering in 2018 from Universidad Técnica Federico Santa María (UTFSM), Valparaiso, Chile.

He currently holds an academic position as an Assistant Professor with the Department of Sound, Universidad de Chile, Santiago. His research interest focuses on the study of human voice production for clinical applications using signal processing, machine learning, and numerical models of voice production.

Dr. Espinoza received a CONICYT (Chilean government agency) scholarship for doctoral studies from 2012 to 2016 and a scholarship from UTFSM in 2015 as a Visiting Researcher with the Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation, Boston, MA, USA.

**Katherine L. Marks** received the M.S. degree in communication sciences and disorders in 2015 from the MGH Institute of Health Professions, Boston, MA, USA, where she is currently working toward the Ph.D. degree. She is currently a Speech-Language Pathologist who specializes in voice disorders and earned her certificate of clinical competence while working with the Lakeshore Professional Voice Center, St. Clair Shores, MI, USA, in the period of 2015 to 2017. Her research interests include vocal effort, acoustic and aerodynamic correlates of vocal hyperfunction, and clinical voice assessment and treatment.

Ms. Marks is currently a Doctoral Fellow with the Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation.

**Matías Zañartu** (S'08–M'11–SM'18) received the B.S. degree in acoustical engineering from Universidad Tecnológica Vicente Pérez Rosales, Santiago, Chile, in 1996, and the M.S. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2006 and 2010, respectively.

He is currently an Associate Professor with the Department of Electronic Engineering and the Director of the Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaiso, Chile. His research interests include the development of digital signal processing, system modeling, and biomedical engineering tools that involve speech, hearing, and acoustics. His research efforts have revolved around developing quantitative models of human voice production and applying these physiological descriptions for the development of clinical technologies.

Prof. Zañartu is a Fulbright Fellow.

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

He is currently Director of the Voice Science and Technology Laboratory at the Center for Laryngeal Surgery and Voice Rehabilitation, Boston, MA, USA, and holds appointments with the Massachusetts General Hospital (Assistant Investigator in the Department of Surgery), Harvard Medical School (Assistant Professor in Surgery), and the MGH Institute of Health Professions (Adjunct Assistant Professor in the Department of Communication Sciences and Disorders), Boston, MA, USA. He is also an Honorary Senior Fellow with the Department of Otolaryngology, University of Melbourne, Melbourne, VIC, Australia.

Dr. Mehta is a member of the Acoustical Society of America and the American Speech-Language-Hearing Association (ASHA) and received the Award for Early Career Contributions in Research from ASHA in 2015.