Improved subglottal pressure estimation from neck-surface vibration in patients with voice disorders

Jon Z. Lin[1], Victor M. Espinoza[2,3], Katherine L. Marks[1,4], Matías Zañartu[2], Daryush D. Mehta[1,4,5]

[1]Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, U.S.A.
[2]Department of Electronic Engineering, Universidad Técnica Federico Santa Maria, Valparaíso, Chile
[3]Department of Sound, Universidad de Chile, Santiago, Chile
[4]MGH Institute of Health Professions, Massachusetts General Hospital, Boston, MA, U.S.A.
[5]Harvard Medical School, Boston, MA, U.S.A.

Keywords: Subglottal Air Pressure; Ambulatory Voice Monitoring; Accelerometer; Inverse Filtering

Introduction

Prior work has shown promise for the noninvasive estimation of subglottal air pressure (Ps) using a linear model of the magnitude of neck-surface vibration during modal and non-modal voice production in vocally healthy speakers [1]. Subsequent work further developed this methodology by incorporating additional measures of vocal function from the neck-surface accelerometer (ACC) signal to achieve improved prediction of Ps during phonation [2]. This study expands on those studies by integrating these additional cepstral and glottal airflow measures to improve the prediction of Ps in patients with voice disorders, with the goal of tracking Ps in naturalistic, ambulatory settings.

Methods

Data were obtained from participants with voice disorders representing a variety of glottal conditions, including phonotraumatic vocal hyperfunction (PVH; associated with nodules/polyps), non-phonotraumatic vocal hyperfunction (NPVH; diagnosed as muscle tension dysphonia), and unilateral vocal fold paralysis (UVFP). Each patient repeated /p/-vowel syllables from loud-to-soft levels in multiple vowel contexts (/pa/, /pi/, and /pu/) and pitch conditions (comfortable, lower than comfortable, higher than comfortable) in their typical voice. Ps estimates were obtained via intraoral pressure (IOP) recordings during occlusive plosives using an intraoral catheter. Simultaneously, oral airflow was captured using a circumferentially vented pneumotachograph mask. Ps for each vowel was estimated by taking the average of IOP peaks preceding and following the vowel.

Two subject-specific, linear regression models were constructed to predict Ps using the ACC signal: Model 1 used the traditional ACC root-mean-square (RMS) magnitude, and Model 2 included additional ACC-based measures of vocal function. The additional measures in Model 2 included fundamental frequency (f_o), cepstral peak prominence (CPP), and glottal airflow parameters from subglottal impedance-based inverse filtering (IBIF) of the ACC signal [3]. Five-fold cross-validation within each patient's data set assessed the robustness of Model 1 and Model 2 performance using the root-mean-square error (RMSE) metric for each of the two regression models.

Results and Discussion

Each fold of the five-fold cross-validation exhibited a baseline prediction performance when Model 1 (ACC RMS alone) was used to predict Ps within each patient group. Improvements to Ps prediction performance (decreases in RMSE) were found when Model 2 added CPP, f_o, and the following glottal airflow measures: open quotient, speed quotient, normalized amplitude quotient, maximum flow declination rate, harmonic richness factor, peak-to-peak amplitude, and the difference between first and second harmonic amplitudes. Of note, similar model performance was achieved when the same glottal airflow-based IBIF measures were derived from the ACC signal using traditional airflow-based inverse filtering, thus showing promise for ACC-only prediction of Ps in clinical populations.

Patient group	Model 1 RMSE	Model 2 RMSE	ΔRMSE (cm H_2O)	ΔRMSE (%)
PVH	2.31 (1.06)	1.78 (0.65)	−0.53 (0.50)	−20.92 (10.87)
NPVH	2.40 (1.02)	2.07 (1.09)	−0.33 (0.24)	−15.81 (12.38)
UVFP	2.36 (0.87)	2.07 (0.81)	−0.29 (0.31)	−11.82 (13.59)

Table 1: Improvements in Ps prediction performance in terms of mean (standard deviation) root-mean-square error (RMSE) within each patient group, comparing the accelerometer RMS-only linear regression model (Model 1) with a multiple linear regression model (Model 2) that incorporated accelerometer-based measures of CPP, f_o, and glottal airflow measures derived using subglottal impedance-based inverse filtering. Change in (Δ) RMSE also reported in cm H_2O and as a percentage with Model 1 as reference.

Acknowledgements

This work is supported by the NIH National Institute on Deafness and Other Communication Disorders (Grants R21 DC015877 and P50 DC015446). 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] Marks *et al*, J Speech Lang Hearing Res, 62:3339–3358, 2019.
[2] Lin *et al*, IEEE J Sel Topics Signal Process, In Press, 2019.
[3] Zañartu *et al*, IEEE Trans Audio Speech Lang Process, 21:1929–1939, 2013.

jzlin@mgh.harvard.edu