TOWARD AN OBJECTIVE AERODYNAMIC ASSESSMENT OF VOCAL HYPERFUNCTION USING A VOICE HEALTH MONITOR

Matías Zañartu[1*], Víctor Espinoza[1], Daryush D. Mehta[2], Jarrad H. Van Stan[2], Harold A. Cheyne II[3], Marzyeh Ghassemi[4], John V. Guttag[4], and Robert E. Hillman[2]

[1]Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[2]Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[3]Bioacoustics Research Program, Laboratory of Ornithology, Cornell University, Ithaca, NY, USA
[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

*Corresponding author. Email: matias.zanartu@usm.cl

Abstract: **Vocal hyperfunction is a description of abnormal patterns of vocal behavior that may lead to many common voice disorders. Previous studies demonstrated that disorders associated with hyperfunction could be detected in patients by measuring aerodynamic and acoustic parameters from recordings of a single sustained vowel using a Rothenberg mask setup. Although ambulatory systems have shown the best potential for unobtrusive long-term monitoring of vocal function, their ability to differentiate hyperfunctional from normal patterns of vocal behavior has not been assessed. This study provides an initial quantitative evaluation of the capabilities of a neck surface acceleration signal to objectively detect abnormal vocal behaviors associated with hyperfunctionally-related disorders. The goal is to verify if such detection is possible using a neck accelerometer signal rather than an airflow mask and incorporate vocal gestures from multiple vowels and running speech. An impedance-based inverse filtering algorithm is used to estimate aerodynamic parameters from the neck-surface acceleration signal. The results obtained when contrasting five patients with vocal nodules to five paired normal subjects indicate that the accelerometer-based assessment offers comparable discrimination capabilities as those from the aerodynamic recordings. The results also provide a first indication that this discrimination is possible with an expanded sample that includes other sustained vowels and running speech.**

Keywords: Voice use, ambulatory voice monitoring, neck accelerometer, vocal hyperfunction, inverse filtering.

I. INTRODUCTION

Many common voice disorders are likely to result from faulty and/or abusive patterns of vocal behavior, referred to as *vocal hyperfunction* [1]. These patterns can be difficult to assess accurately in the clinical setting and may be better characterized when individuals wear an ambulatory voice monitor while engaging in their typical daily activities. Current methods for ambulatory assessment of vocal function are based on measurements of neck surface acceleration and constitute a non-invasive, unobtrusive, noise-robust approach that maintains confidentiality [2]. However, there is a lack of statistically robust studies that demonstrate the true diagnostic utility of such systems. Our group strives to advance accelerometer-based ambulatory monitoring of vocal function by validating it as a reliable and cost-effective clinical tool that can be used to accurately identify and differentiate patterns of voice use that are associated with hyperfunctional voice disorders.

Our recently developed system, the Voice Health Monitor (VHM) shown in Fig. 1, is an accelerometer-based system that uses a smartphone platform that takes advantage of technological advances to allow for recording and storing raw acceleration data for at least 7 full days [3]. The acceleration data is subjected to an inverse filtering technique known as sub-glottal impedance-based inverse filtering (IBIF) to provide a

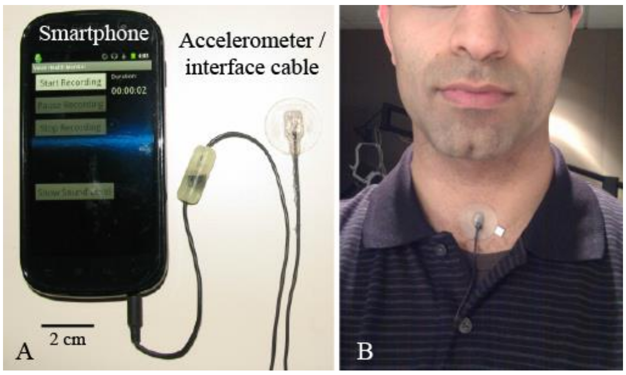

Fig. 1: Voice Health Monitor: (A) Smartphone and accelerometer assembly with (B) illustration of neck-surface sensor position.

non-invasive estimation of the glottal airflow based on the neck-surface acceleration signal [4]. To achieve that task, the IBIF algorithm uses a transmission line model of the subglottal system, a lumped representation of the mechanical properties of the neck, and subject-specific parameters estimated during a calibration session.

In this study, we provide an initial quantitative evaluation of the capabilities of the accelerometer-based aerodynamic parameters extracted using subglottal IBIF to discriminate disorders associated with vocal hyperfunction. Our goal is to replicate the analysis performed by Hillman *et al.* [1] for selected parameters and conditions of interest. To accomplish this goal, we aim to address the following research questions:

1. Can we discriminate between patients with hyperfunctionally-related disorders and subjects with normal voices using a neck accelerometer signal, comparable with previous results obtained with airflow mask–derived aerodynamic measures?
2. What are the best aerodynamic measures to extract from the acceleration signal?
3. How is discrimination affected by changing the articulatory gesture from a sustained vowel /a/ to other vowels and running speech?

II. METHODS

We recruited five adult female subjects with bilateral vocal fold nodules and five female subjects matched for age and occupation. We recorded the neck-surface acceleration signal using the VHM simultaneously with three physiological signals of interest: oral airflow, electroglottograph (EGG), and sound pressure level (SPL). These signals were used to compute both subject-specific parameters for the IBIF scheme (see [4] for details) and compensate for loudness variation when performing the statistical analyses. The subglottal IBIF algorithm estimated four aerodynamic measures from the neck-surface accelerometer signal windowed into 100 ms non-overlapping frames: maximum flow declination rate (MFDR), amplitude of the modulated flow (AC Flow), open quotient (OQ), and speed quotient (SQ). These measures were selected so that we could compare the results of this study with those from [1].

The aerodynamic measures were computed for each of four vocal gestures performed at a comfortable loudness level (no specific target SPL): sustained vowels /a/, /i/, /u/, and the Rainbow Passage. Normal and regressed Z-scores were obtained for each gesture following the procedure described in [1], but using the matched-normal subject as the reference (rather than using a normative data set). Thus, mean measure values within each gesture for each patient were normalized by the means and standard deviations of the same vocal gesture from the matched subject. The normal (Z_N) and regressed (Z_R) Z-scores are computed, respectively, as

$$z_N = \frac{\bar{x}_P - \bar{x}_N}{\sigma_N} \text{ and} \qquad (1)$$

$$z_R = \frac{\tilde{x}_P - \bar{x}_N}{\sigma_N}, \qquad (2)$$

where $\bar{x}_P$ and $\tilde{x}_P$ are the normal and regressed observations for the patient, respectively, and $\bar{x}_N$ and σ_N are the mean and standard deviation of the matched subject. Given that the comparison is performed against one gesture on a single paired subject, the standard deviation refers in this case to the stability of the signal every 100ms, rather than the variation of a population.

Fig. 2 displays a graphical representation of the method used to compute Z_R. A robust least-square linear regression [5] is calculated for the normal subject data and used to extrapolate each aerodynamic measure to correct for loudness differences. The Z_R score for each patient's measures is given by the distance to the regression line, normalized by the standard deviation measured for the same gesture by the matched subject. Z_R scores are reported when a high Pearson's correlation ($|r| \geq 0.7$) exists between a given measure and SPL.

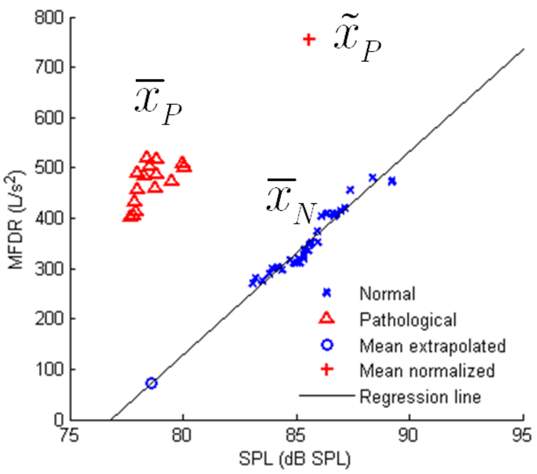

Fig. 2: Computation of regressed Z-scores for MFDR from the sustained vowel /a/ of subject pair 1

We present both Z_N and Z_R scores to evaluate the potential of the accelerometer signal to discriminate normal and pathological cases with and without the correction for the effect of SPL. This information may be used in the future to determine how to implement a real-time ambulatory application of this method for use in biofeedback.

Table I: Discrimination power between each subject pair. Regressed Z-scores scores for sustained vowel /a/ are comparable with those in [1]. Nomenclature: Z-scores ≥ 2 (+), Z-scores ≤ 2 (−), |Z-score| < 2 (no symbol).

Gesture	**Subject pair**	**Normal Z-score**				**Regressed Z-score**			
		MFDR	**AC Flow**	**OQ**	**SQ**	**MFDR**	**AC Flow**	**OQ**	**SQ**
Sustained vowel /a/	1		+	−		+	+		
	2	+	+				+		
	3	+				+			
	4		+				+		
	5	−	−				+		
Sustained vowel /i/	1	+	+			+	+		
	2		+				−		
	3	+	+			+	+		
	4	+	+			+	+		
	5				−				
Sustained vowel /u/	1	+	+			+	+		
	2	+	+			−	−		
	3	+	+			+	+		
	4	+	+			+	+		
	5	−	−			+	+		
Rainbow Passage	1		+						
	2								
	3								
	4		+						
	5								

III. RESULTS AND DISCUSSION

The summary of the Z-scores from the five subject pairs is shown in Table I. There are fewer Z_R scores Z_R reported because only those with $|r| \geq 0.7$ and $|Z_N| \geq 2$ were considered. For both scores, AC Flow is the most salient measure, followed by MFDR. The features OQ and SQ did not exhibit discriminating power. When comparing the results from vowel /a/ (the only gesture evaluated in previous studies) with the reference framework described in [1] for the nodules patients, the same trend in terms of salient measures is observed.

There were some negative values of Z_N for the /a/ vowel but no negative Z_R scores following adjustments/corrections for SPL (when correlations with SPL were high), which also matched previous results [1]. Our results for different sustained vowels yielded comparable discrimination power using Z_N but a slightly less robust behavior using Z_R; i.e., negative scores were observed for one subject pair (#2). It is possible that the SPL effect was not completely adjusted for due to the fact that only one matched subject was used in the linear regression, rather than a normative dataset. It is also possible that these lower vowels provide a different loading where reduced Z-scores are in fact correct. Given that these vowels have not been tested in previous studies, further investigations are needed.

The mean parameter values extracted from the Rainbow Passage did not correlate highly enough with SPL to allow for the calculation of Z_R scores. This may have been due to the averaging of measurements across many complex/variable phonemic environments. However, normal Z-scores using the unadjusted mean parameter values did provide salient scores in two of the patients for AC Flow. These findings suggest that smaller temporal windows may provide better discrimination and support the potential for computing real-time Z-scores in an ambulatory device.

V. CONCLUSION

This initial evaluation indicates that the accelerometer-based estimates of aerodynamic parameters obtained via subglottal IBIF provide comparable discrimination capabilities between normal and pathological subjects as was observed in previous studies using actual aerodynamic recordings. The results also provide a first indication that this discrimination is possible with other sustained vowels and even with running speech, thus motivating the continued development of these approaches for applications in ambulatory voice monitoring systems.

ACKNOWLEDGMENT

This work was supported by NIH-NIDCD grant R33 DC011588, CONICYT grant FONDECYT 11110147, and MIT MISTI grant MIT-Chile 2745333. V.E. acknowledges scholarships from CONICYT and Universidad de Chile.

REFERENCES

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

[2] R. E. Hillman and D. D. Mehta, "Ambulatory monitoring of disordered voices," *Perspectives on Voice and Voice Disorders*, vol. 21, no. 2, pp. 56–61, 2011.

[3] 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, no. 11, pp. 3090–3096, 2012.

[4] M. 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. Proc*., vol. 21, no. 9, pp. 1929–1939, 2013.

[5] R. A. Marona, R. Douglas Martin, V. J. Yohai, "Robust Statistics," John Wiley & Sons, 2006.