QUANTIFYING THE PERCEPTUAL QUALITY OF STRAIN:

AN ELECTROGLOTTOGRAPHIC ANALYSIS OF CONTINUOUS DYSPHONIC SPEECH

by

KEERTHAN SOMANATH, BSEE

DISSERTATION

Presented to the Faculty of the Medical School

The University of Texas Southwestern Medical Center

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF MEDICINE WITH DISTINCTION IN RESEARCH

TABLE OF CONTENTS

ABSTRACT

Introduction: Excessive vocal effort perceived as strain may involve increased vocal fold compression. Electroglottography (EGG) in principle could provide a measure of strain directly connected to oscillatory kinematics at the sound source. The utility of EGG has been limited by considerable inter- and intra-subject variability.

Objective: This work aims to (1) develop an algorithm to analyze EGG signal in continuous, dysphonic speech and (2) identify parameters that correlate with strain.

Methods: EGG signal from 8 normal speakers and 8 subjects with adductor spasmodic dysphonia (ADSD) reading two-sentence excerpts from the Rainbow Passage was processed by the new software developed in MATLAB. The contact quotient (CQ), pulse width at the 50% amplitude level (EGGW50), and various closing slope and opening slope measures were extracted from selected speech segments. Intra-subject and inter-subject comparisons were then made.

Results: None of the EGG parameters differed between normal and ADSD speakers. Within-subject comparison among ADSD speakers showed that the opening slope measure SO7525 distinguished between the strained and unstrained syllables.

Conclusion: These results provide further insight into the utility and limitations of EGG. While EGG may have limited utility in inter-subject comparison, it may provide a useful objective measure of vocal strain in the same subject with variable degrees of strain or over time.

INTRODUCTION

The human voice can be evaluated by a variety of methods. Electroglottographic (EGG) signal is produced when vocal fold vibrations produce cyclic fluctuation in the electrical impedance across the larynx. The EGG signal thus reflects the degree of contact between the vocal folds during voice production and provides a measure of voice quality based on phonatory physiology. However, the utility of EGG has been limited because existing methods of EGG signal analysis focus on the evaluation of 2-3 parameters in a segment of sustained vowel production, which does not reflect pathologies more apparent in conversational speech. Patients' symptom severity may not be accurately reflected in a sustained vowel. The perceptual quality of a sustained vowel is one measure used in assessing the need for intervention, but this cannot be the only gauge. For example, a patient on maintenance Botox therapy for spasmodic dysphonia may be near-normal on the Rainbow passage when their voice starts to break. The potential for EGG analysis to clinically guide therapy was explored in this study.

Excessive vocal effort perceived as vocal strain is a cardinal feature in many types of voice disorders. The identification and assessment of strain is important because they may guide possible targets in the phonatory system for intervention, and the reduction of vocal strain is often one of the primary goals of voice therapy. The "gold standard"' for the assessment of vocal strain continues to be auditory-perceptual measures. Strain is one of the key attributes assessed in standardized clinical voice assessment protocols such as the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)[1] and the Grade, Roughness, Breathiness, Asthenia, Strain scale (GRBAS)[2]. To circumvent limitations of perceptual evaluation[3], researchers have sought objective correlates of vocal strain in aerodynamic and acoustic measurements. It was reported 30 years ago that the perception of strain was associated with low glottal flow and high subglottal pressure in a repeated /pα/ syllable task in a group of dysphonic patients[4]. A subsequent study of trained voice users showed that laryngeal resistance distinguished normal voice from pressed voice, which is qualitatively similar to strained voice[5]. Measures derived from the acoustic signal have been investigated as well. In a comparison of comfortable versus maximally projected acting voice in trained actors, a moderate positive correlation was found between the perceptual rating of strain and the average intensity difference between the lower and higher spectral regions in the long-term average spectrum (LTAS) [6]. More recently, the cepstral peak prominence, the spectral L/H ratio, and the Cepstral/Spectral Index of Dysphonia were all shown to have moderate to strong correlations with strain severity in the reading samples of dysphonic patients[7]. Compared to the aerodynamic measures, the acoustically-derived parameters generally show stronger correlation with the perceptual ratings and have the advantage of reliable measurements from connected speech. The stronger correlation is perhaps not surprising since auditory perception is derived from acoustic input and is further removed from the aerodynamics of voice production. Despite significant progress, however, all objective measures have limitations. For example, differentiation of vocal quality based on acoustic measures may not be specific for the vocal dimension of strain, as the same acoustic measures can be deviant in both strained voice and breathy voice[7,8]. On the other hand, while the aerodynamic parameters of flow, pressure, and resistance are directionally sensitive to dimensions such as strain or breathy[5], aerodynamic measurements are typically collected via vocal tasks that may not fully reflect the pathology of the voice disorder as manifested in continuous speech.

A physiologic measurement that could in theory quantify some aspect of vocal strain and is readily obtained in connected speech is the electroglottograph (EGG). The EGG measures the change in tissue impedance across the two sides of the larynx as the vocal folds contact and separate during phonatory vibration. The change in EGG waveform during the glottal cycle reflects the change in vocal fold contact area (VFCA) and therefore provides a physiologically derived measure of vocal function connected to the oscillatory kinematics at the sound source[9-11]. EGG measures are thought to indicate the degree of vocal fold medial compression[12,13]. While other elements of hyperfunctional phonatory behavior such as hyperadduction of the ventricular folds, excessive subglottal pressure, and hyperfunction of extrinsic laryngeal muscles may contribute to the auditory-perception of vocal strain, strain may also be associated with the degree of compression of the vocal folds [14]. It is reasonable to postulate that the perception of strain would correlate with increased vocal fold compression or VFCA in at least a subset of voice disorders in which excessive strain originates primarily from the glottic level. This rationale to employ EGG in the assessment of strain is supported by previous data. The EGG closed quotient was shown to distinguish pressed phonation from normal or resonant phonation on sustained /a/ and /i/ in trained voice users [15,16]. Further support came from a study of patients with adductor spasmodic dysphonia (ADSD) in which EGG measures of vocal fold adduction correlated with symptom severity in 2 of the 5 patients[17].

While the prior studies provided valuable, albeit at times inconsistent, data on how EGG measures relate to the perceptual quality of strain, there were also important limitations. First, much of the affirmative data came from vocal performers with no laryngeal pathology. The applicability of the findings to a more clinically relevant, dysphonic population is unknown. Second, with the exception of Fisher et al.[18] and Fisher et al. [17], most studies of dysphonic voice using EGG have only investigated a single measure, i.e. the closed quotient or the contact quotient (CQ), or the percent time the vocal folds are thought to be in contact during the glottal cycle. This single time-domain parameter captures only a small amount of the information content in the complex EGG waveform, which also provides information on the rate of contact and de-contact, for example. The use of the CQ to represent EGG's utility as an assessment tool likely underestimates the potential of EGG. The work by Fisher and colleagues[17,18] showed that the EGG pulse width and slope measures, parameters that are thought to reflect vocal fold adduction, also correlated with symptom severity in a subset of patients. These parameters, however, have not been further investigated by other groups. Finally, in all previous studies, EGG data were collected on a segment of sustained vowel in order to minimize signal distortion encountered in connected speech. While the CQ from connected speech has been reported[19], it is unclear how the EGG signal was processed to take into account such distortion.

There are several rationales to extend the capability of EGG analysis to connected speech. First, compared to sustained vowels or single syllables, connected speech is a more complete representation of speaking patterns that constitute dysphonic symptoms in daily activity. Second, the type and frequency of phonatory aberrations may differ between sustained vowel and connected speech, as has been shown in ADSD[20]. Finally, analysis of connected speech in principle avoids the significant inconsistencies introduced by the manual selection of a sustained vowel segment for analysis[21]. There is, however, a knowledge gap on how EGG signal in connected speech is best handled. The calculation of EGG waveform parameters beyond CQ is affected by the substantial DC drift in connected speech due to laryngeal movement relative to the skin-fixed electrodes. The effect of high-pass filtering, which is commonly applied to attenuate the DC drift, on waveform parameters such as the opening and closing slopes has not been systematically investigated. Signal processing is especially problematic in dysphonic speech, where distorted EGG waveforms are common and the signal-to-noise tends to be low. Given the potential utility of EGG to provide a physiologic correlate of strain, there is a need for improved processing of dysphonic EGG signal and critical evaluation of waveform parameters beyond the CQ.

This work aimed to develop methodology to analyze EGG signal in connected dysphonic speech and to apply it to ADSD in a pilot study. The two objectives were to (1) establish the feasibility of automated waveform parameter extraction from EGG data that contain high aperiodicity, distorted waveforms, and high-frequency noise, and (2) to use the new algorithm to identify EGG parameters that correlate with the perceptual quality of strain in ADSD. We focused the initial application to one specific voice disorder in order to avoid the heterogeneity of disease-specific EGG findings across different disorders. We chose ADSD with the following considerations: (1) Strain is one of the hallmark perceptual features in ADSD speech, with a hyperadductive component as the dominant perceived characteristic in ADSD[22]; (2) the primary deficit in ADSD is involuntary spasms or over-adduction of the vocal folds. Localization of the source of vocal strain to the vocal folds makes ADSD a better disease entity to investigate than other hyperfunctional voice disorders, e.g. muscle tension dysphonia, in which other components of the phonatory system may also make significant contribution; (3) the vocal folds in ADSD are morphologically normal, which allows a more direct connection between the EGG signal and hyperadduction, versus disorders in which strain is present in the context of vocal fold lesions, e.g. nodules and polyps; (4) the diagnosis of ADSD rests on the identification of stress-strain patterns in speech that vary with phonetic content[23], yet this determination is dependent on listener training and experience. There is therefore a need for objective methods to identify stress-strain variation; and (5) since ADSD symptom is sensitive to phonetic content, the less-strained segments of speech can serve as internal control for each subject.

We hypothesized that (1) EGG parameters could differentiate between normal and ADSD speech, and (2) EGG parameters could distinguish ADSD speech segments that are perceptually different in their degree of strain. In particular, we were interested in the diagnostic utility of the waveform parameters that convey the speed of glottal contact and de-contact, in order to explore the potential utility of EGG beyond the CQ.

METHODS

Speakers and Data Collection

The study protocol was approved by the Institutional Review Board of UT Southwestern Medical Center. A group of ADSD speakers and a gender-matched group of normal speakers participated in this study with prospective data collection. The ADSD speakers were recruited from patients seen at the UT Southwestern Voice Center. All have the diagnosis of ADSD based on assessment by a speech-language pathologist with extensive experience in the treatment of voice disorders, assessment by a laryngologist, and laryngoscopic examination. Since the intent of the study was to determine if it was feasible to use EGG to detect vocal strain, we selected twelve ADSD speakers whose vocal symptoms were relatively severe at the time of data collection, out of a pool of 40 screened. Our rationale was that if the utility of EGG could not be demonstrated in this subset of speakers, then it probably did not exist. Of the 12 ADSD speakers, 1 had never received treatment for ADSD, 1 was seen 18 months following his previous BOTOX injection, and 1 was seen 6 months following her previous BOTOX injection. These 3 speakers were severely symptomatic at the time of data collection. The remaining 9 speakers were recorded just before their scheduled BOTOX injection, at 3-5 months following their previous injection, while they were becoming more symptomatic but before their voice quality had completely declined to a non-treatment level. At the time of data collection, they ranged from mildly to severely symptomatic compared to their own range of symptom fluctuation through cycles of BOTOX treatment.

Speakers were asked to read a two-sentence excerpt from the Rainbow Passage with comfortable loudness and pitch while EGG and acoustic signals were collected. The second and third sentences were used, from “The rainbow is…” to “…beyond the horizon”. EGG data were acquired with the EG2 electroglottograph (Glottal Enterprises, Syracuse, NY) using the vocal fold contact area (VFCA) output and a 10 Hz low frequency limit. The manufacturer recommends a low frequency limit of 20 Hz for routine clinical use and states that 40 Hz may also be acceptable for female voices. We chose the lower cutoff of 10 Hz to further minimize potential waveform distortion. Lower cutoffs, e.g. 5 Hz, resulted in signal drifts that consistently exceeded the dynamic range of the acquisition. The signal in each channel was digitized at 25 kHz by a DATAQ DI-720 A/D converter and recorded on a computer using WINDAQ (DATAQ Instruments, Akron, OH).

Software: Overview

Software was developed in MATLAB (versions R2011-R2013a, The Math-Works, Natick, MA) to display and analyze EGG data via a graphical user interface (GUI). The acoustic and EGG signals were displayed in a stacked manner to allow speech segment selection based on either signal. The GUI makes possible the inspection of EGG waveform, playback of the corresponding acoustic signal, and control of data processing flow.

Signal Flow

Software was developed in MATLAB to acquire and display the hardware-filtered EGG, which was in the form of a .csv file. Next the EGG baseline was to be corrected by either linear correction or Butterworth filtering. Following this a Savitsky-Golay smoothing filter was applied to the EGG, in order make further signal processing easier. Next the EGG signal was differentiated in order to acquire the DEGG signal. This DEGG signal was processed through a peak-detection algorithm we created in order to find DEGG maxima/minima. These maxima/minima were then used to identify EGG maxima/minima, which then allowed us to pursue further parameter calculations. Figure 1 below shows a diagram of the general signal flow described above.

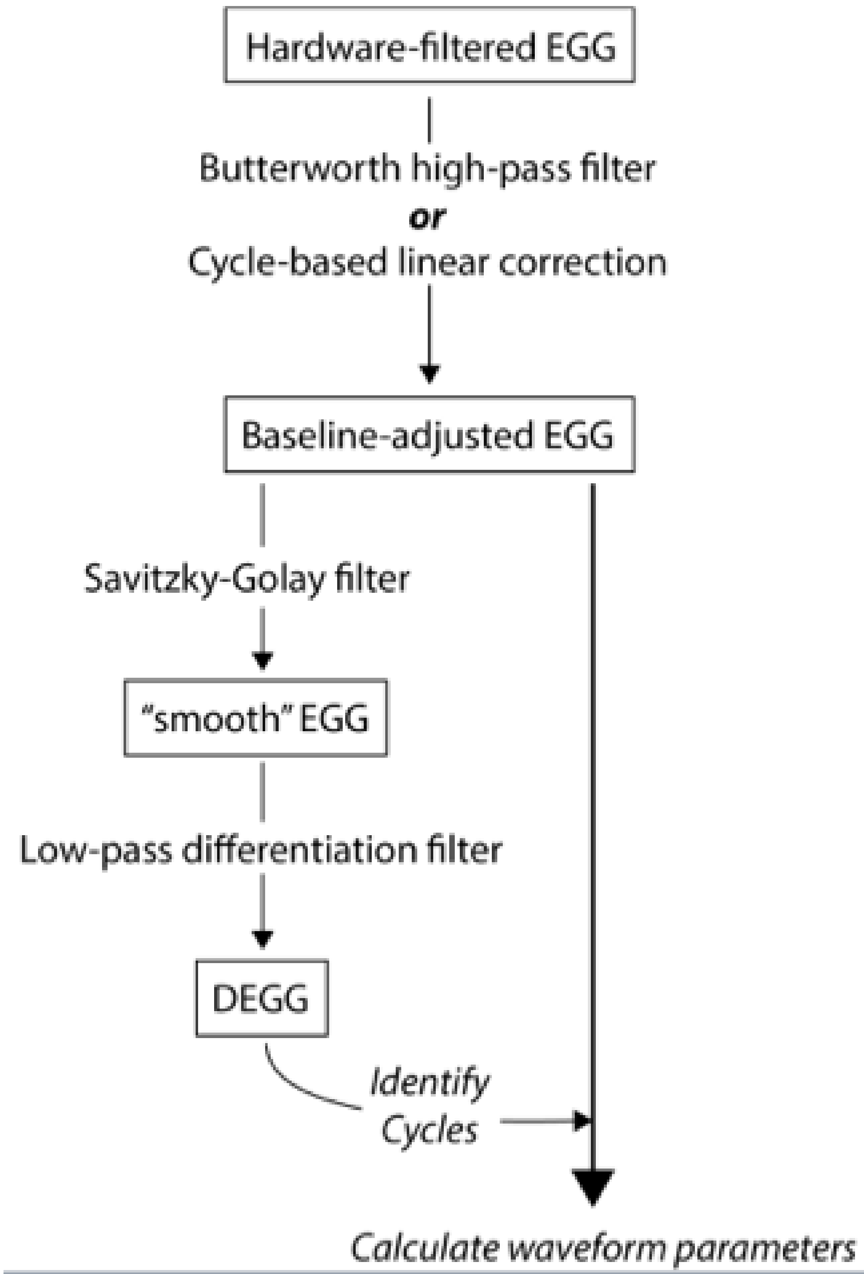

Figure 1. Diagram illustrating general signal flow from EGG signal acquisition to parameter calculations.

Signal Processing: Baseline Correction

Two alternative strategies were implemented to address the DC drift in the EGG signal. The first utilized a second-order Butterworth high-pass filter, and the second used cycle-based linear correction.

Butterworth high-pass filter

To determine the effect of the high-pass filter cutoff frequency on the morphology of the EGG waveform, select EGG parameters were calculated with Butterworth filter normalized cutoff frequencies of 0 (no cutoff), 1, 5, 10, 20, and 50 Hz in the sentence “The rainbow is a division of white light into many beautiful colors” in 5 test speech samples from patients with ADSD. The SC5075, SC2575, SO9050, SO7525, EGGW50, and NormalizedArea (defined below) were computed with each frequency cutoff and the percent deviation from values calculated without Butterworth filter was tabulated. This process yielded 30 values (6 parameters per each of the 5 frequency measurements) of percent-deviation for each cutoff frequency, and they were binned in 0.5% increments.

Cycle-based linear correction

In this method, suggested by Ron Scherer, the baseline correction was performed for each EGG cycle independent from other cycles. The y values over the cycle were corrected by an amount $\Delta y = B \times (\Delta x/W)$, where B was the difference between the starting and ending y values, and W was the cycle width. Figure 2 below illustrates this method.

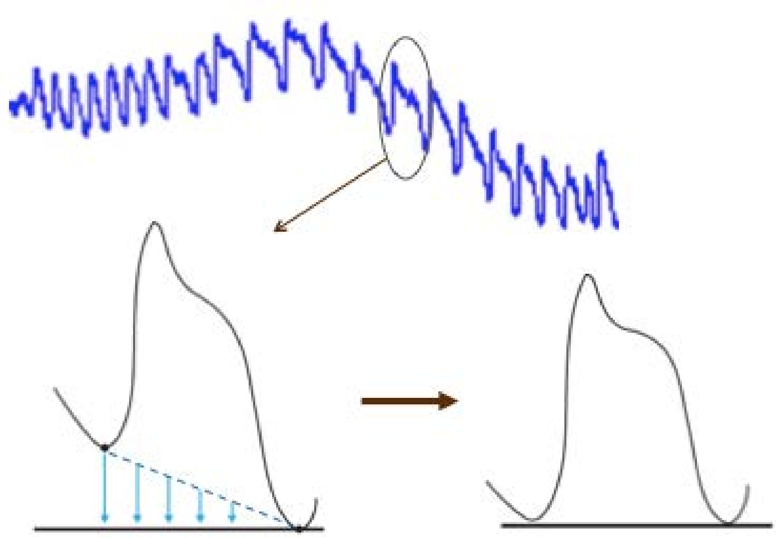

Figure 2. Schematic illustrating cycle-based linear correction.

Definition of EGG Waveform Parameters

The parameters are illustrated in Figure 3. ***Contact quotient (CQ)*** is defined by maxima and minima in the DEGG waveform[13]. ***EGGW50*** is the pulse width at the 50% amplitude level[13]. Contact opening and contact closing slopes are defined based on previous studies[9,18], and the normalized version is used[18]. In the current work, the closing height H_C and opening height H_O of each cycle are not assumed to be the same. The contact closing slope ***SC1090*** denotes the normalized rise (B/H_C) divided by the normalized run (A/T), where B is the EGG signal height corresponding to the segment between 10% and 90% of the closing height H_C, A is the time segment corresponding to B, and T is the period of that cycle (Figure 3a). ***SC2575*** is similarly defined to denote the slope between 25% and 75% of the closing height, and ***SC5075*** the slope between 50% and 75% of the closing height. The opening slopes ***SO7550, SO7525, and SO9050*** are defined analogously and normalized to the opening height H_O instead of H_C. A new parameter ***peak skew*** denotes the relative position of the EGG cycle peak within the cycle period (Figure 3b). On a scale of 0 to 1 (normalized to cycle period T), the earlier the peak occurs within the cycle, the smaller the peak skew value. While the peak skew parameter is motivated by a previous study[24], we use it solely as an empiric, geometric parameter and do not make inference about glottal shape based on its value.

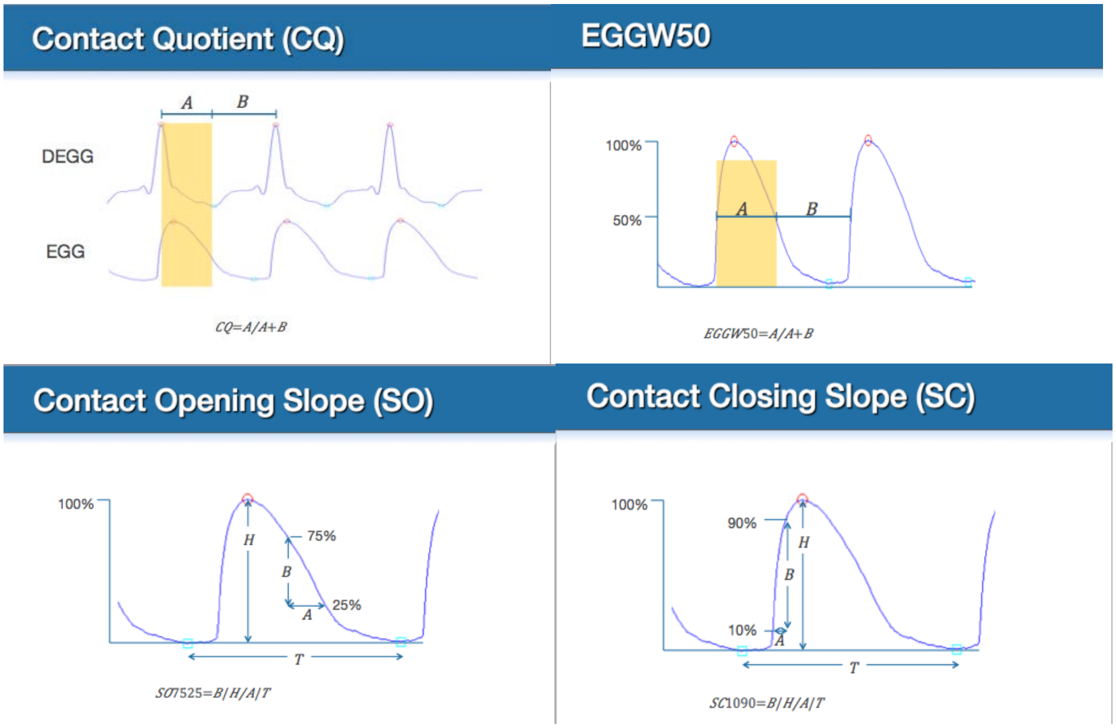

Figure 3a. EGG parameter definitions.

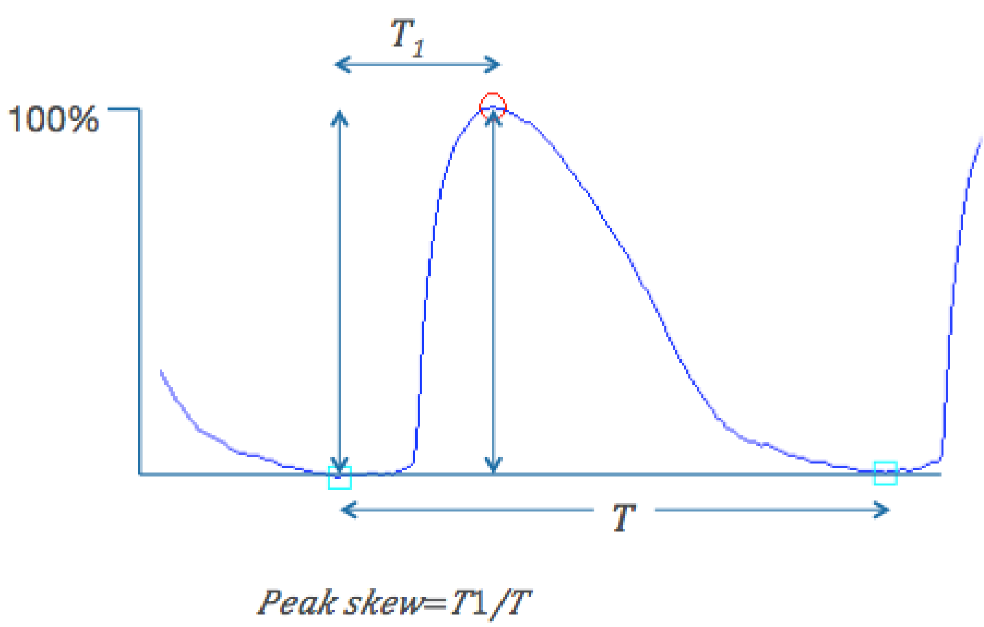

Figure 3b. Peak skew definition.

Signal Processing: Cycle Selection

Phonatory cycles in the EGG signal were identified indirectly using the derivative of the EGG (DEGG). The premise of this approach has been illustrated previously[25,26], but modifications were implemented in the present work to facilitate its application to dysphonic data. High-frequency noise in the EGG signal was removed with an Savitzky-Golay (polynomial) smoothing filter. The DEGG was then generated using a low-pass differentiation filter[27,28]. Positive peaks in the DEGG signal, which corresponded to rapid glottal closing, were identified using a predefined threshold that can be changed by the user. The time-locations of the DEGG peaks were then used as the starting points for a local maximum search in the original, baseline-corrected EGG signal (without application of the polynomial smoothing filter) to identify the EGG maxima. EGG minima were then identified by a local minimum search between successive maxima.

Signal Processing: Cycle Rejection

The challenge in using an automated cycle-selection algorithm to process data that contain a high degree of noise and waveform shape irregularity is to determine which cycles to keep and which to reject. If the rejection criteria are too strict, some “true” cycles will be rejected; if the criteria are too forgiving, “false” cycles will be included. We chose to bias our criteria towards the rejection of false cycles based on the following rationale and assumptions: 1. Since the intent is to analyze speech segments, a relatively large number of cycles will be included in each selection; 2. in most selections, the number of true cycles will outnumber the number of false cycles; 3. the number of false cycles should be minimized to reduce their contribution to the averaged parameter values in a selection; 4. the unintentional rejection of a small number of true cycles is compensated by the larger true:false ratio in the remaining cycles following the application of rejection criteria; 5. the effect of any false cycles that remain is diluted by the larger true:false ratio in the remaining cycles; and 6. the averaged value of any particular EGG waveform parameter of the remaining cycles should therefore be representative of the speech segment selection.

With the above in mind, we sought to reject cycles with questionable morphologies. To determine waveform parameters that have a high likelihood of indicating a false cycle (a cycle identified by the automatic cycle-identification algorithm but in fact represents noise or a false peak), a preliminary analysis of 5 ADSD datasets was performed using the following rejection criteria to remove cycles with CQ < 0.1, CQ > 0.9, or period falling outside of two standard deviations of the cycles in the selected data segment. The EGG corresponding to the sentence “The rainbow is a division of white light into many beautiful colors” was analyzed and the results pooled. This yielded a total of 2087 automatically identified cycles, with 376 rejected based on the above criteria and 1711 accepted for further analysis. The distribution of accepted and rejected cycles across bins of values (0 to 1 in 0.1 increments) of CQ, EGGW50, and peak skew was examined. Based on this analysis, the additional rejection criteria of EGGW50 < 0.1, EGGW50 > 0.8, peak skew < 0.1, and peak skew > 0.8 were implemented, since cycles with any of these morphologic features were highly unlikely to represent true phonatory cycles.

Finally, the first two and last two cycles in each continuous segment of EGG cycles were also excluded from analysis to minimize waveform distortion from phonation onset and offset.

Listeners and Perceptual Tasks

Three speech-language pathologists with a primary practice focus on voice evaluation and therapy and a fellowship-trained laryngologist performed two rating tasks on each ADSD speech sample. For each sample, the listener was given a page printed with a 100-mm visual analog scale (VAS) marked in quartiles, as well as the text of the reading passage. First, the listeners were asked to place a single mark on the VAS to rate the “strain variability” or the dynamic range of the strain quality of the sample. A mark close to 0 signaled little difference in strain between the least-strained syllables and the most-strained syllables in that sample, whereas a mark close to 100 signaled a large difference in the degree of strain between the least-strained and most-strained syllables. For example, a sample that sounded very strained on some vowels but near-normal on others would receive a high rating. A sample that had little strain throughout would receive a low rating, and so would a sample that sounded evenly strained throughout with no variation of the strain with phonetic content. In the second rating task, the listeners circled the syllables that sounded more strained than the rest and underlined the syllables without any strain. Each sample was played 3 times. All 12 ADSD speech samples were rated. Three samples (25%) were repeated in a different session to measure intra-rater agreement. For each sample, if a syllable was circled by at least 3 listeners, it was considered a consensus strained syllable. Consensus strained syllables were identified in 9 of the 12 samples. The remaining 3 samples all had low strain variability and the listeners did not agree on which syllables were strained. These 9 samples were used for further analysis. The phrase "take the shape" was consistently unstrained across all 9 samples except one and was used as the unstrained token for 8 of the 9 samples. "Apparently" was used as the unstrained token for the 9^{th} sample.

Statistical Analysis

Kendall's coefficient of concordance as a measure of inter-rater agreement in the rating of "strain variability" among the 4 raters was calculated using the MAGREE macro in SAS 9.3. Cohen's *d* was calculated using a third-party macro in SAS[29].

RESULTS

Effect of Baseline Correction

To determine how the Butterworth filter affects EGG waveform morphology, a preliminary analysis using a limited dataset was performed. Select EGG parameters of a test sentence in 5 ADSD subjects were calculated after filters with various cutoff frequencies were applied. The results are shown in Figures 2 and 3. A cutoff frequency of up to 50 Hz did not appreciably change the percentage of parameters that remained within 2% of the values calculated without a filter. However, the percentage of parameters that remained within 0.5% declined with increasing frequency cutoff, showing that higher cutoffs do exert an effect on the parameter values. Figure 4 below illustrates this. Based on this preliminary analysis, a 10 Hz cutoff for the Butterworth filter was utilized for subsequent parameter calculations, although a 5 Hz cutoff was likely to produce comparable results.

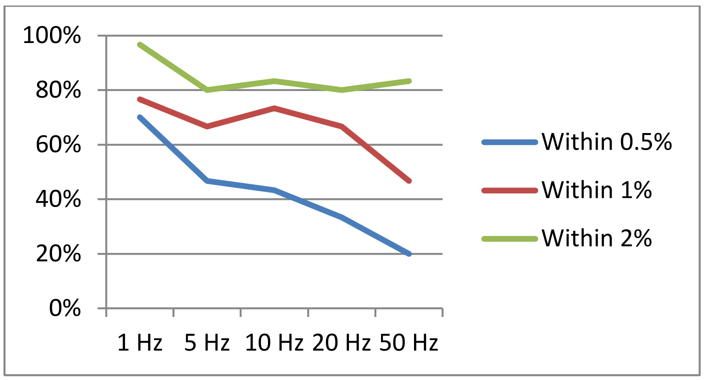

Figure 4. Effect of Butterworth filter cutoff frequency

Figure 5 shows the raw EGG signal. Figures 6-7 show the comparison between using cycle-based linear correction and a high-pass filter to remove the DC drift in an EGG signal from a moderately dysphonic patient.

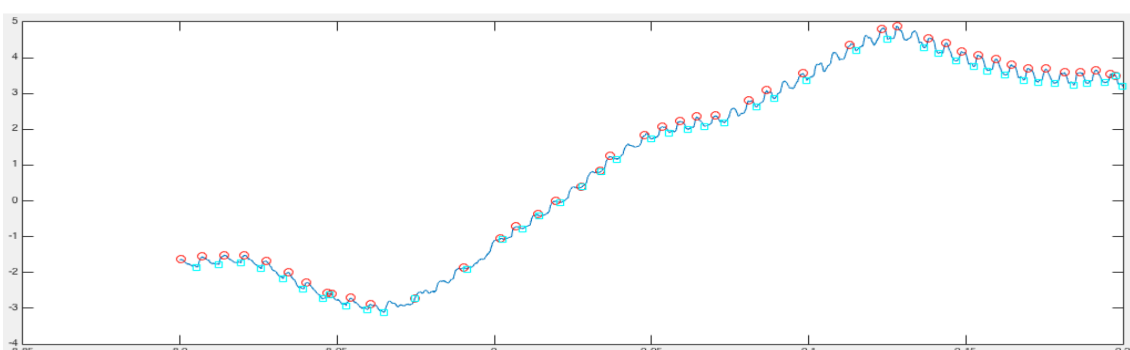

Figure 5. Raw EGG waveform [range ~9]

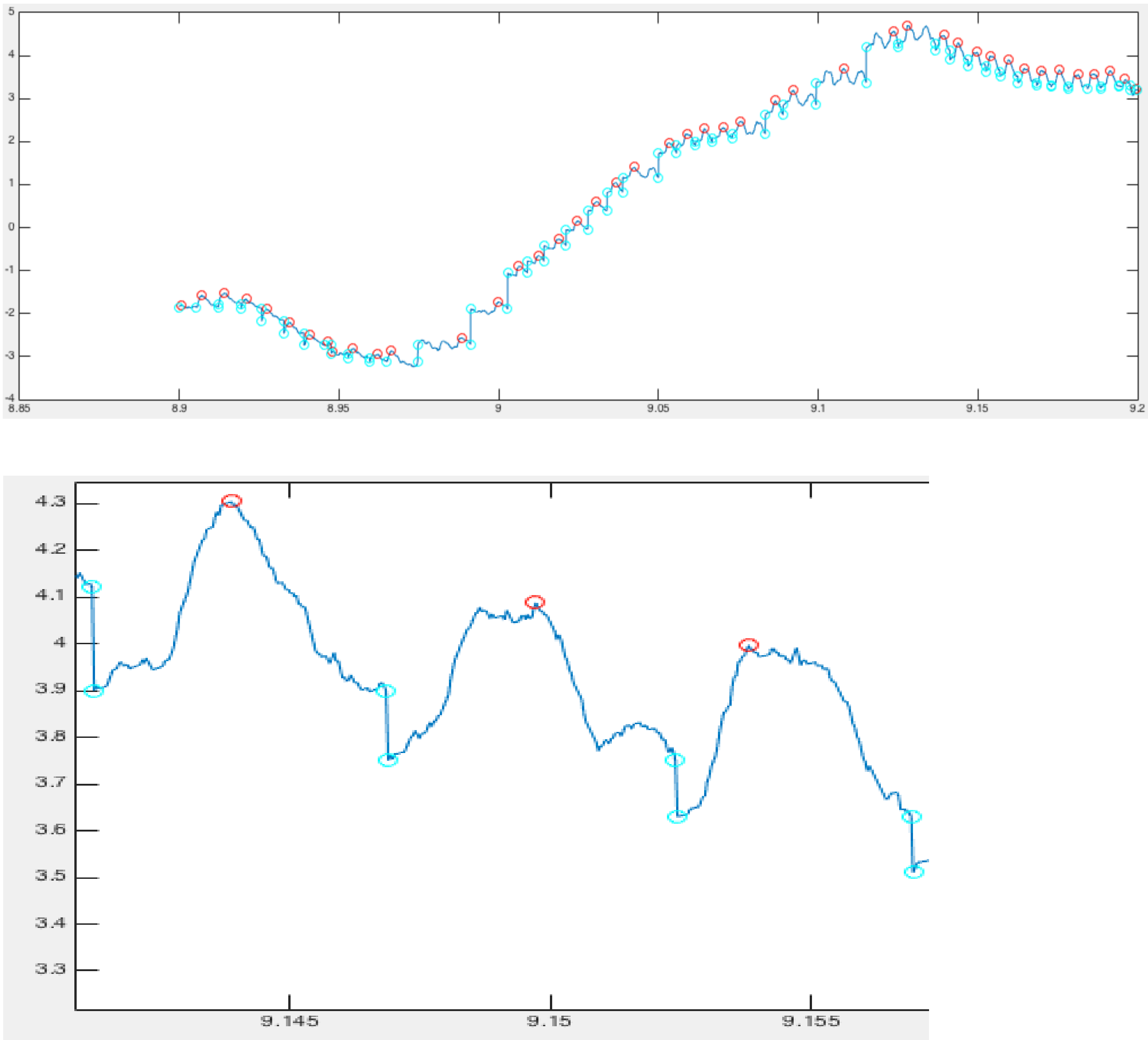

Figure 6: (top) EGG signal from moderately dysphonic patient after cycle-based linear correction [range ~9]. (bottom) Zoomed-in to show corresponding minimums on opening and closing side of cycle. Refer to Figure 2 above for schematic of linear correction.

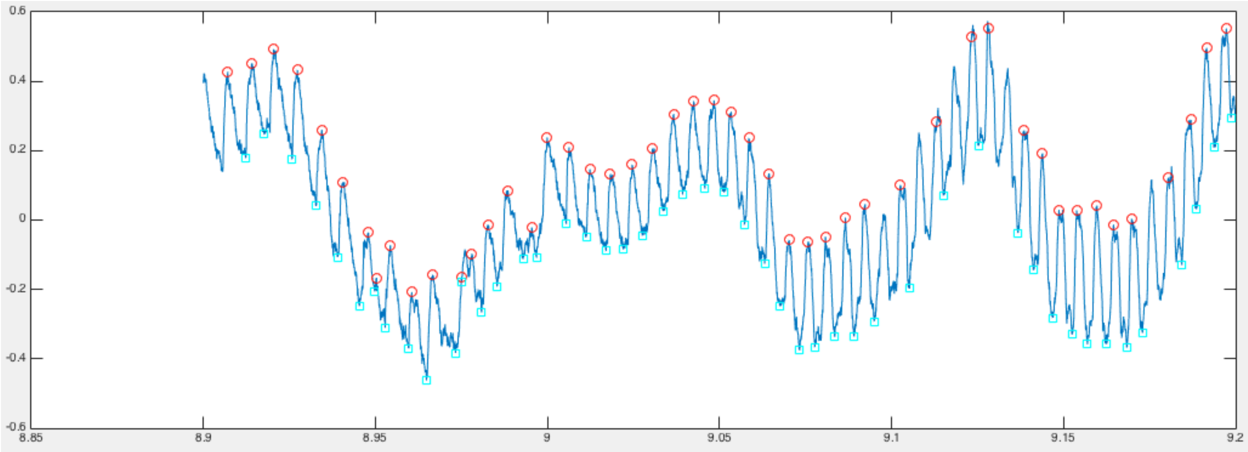

Figure 7. EGG signal from moderately dysphonic patient with 10 Hz Butterworth filter applied

(range ~1)

EGG Cycle Selection/Rejection

Figure 8 shows the automated selection of EGG maxima and minima from both a normal and ADSD speaker. Figure 9 shows both the selected and rejected EGG cycles from a dysphonic data set.

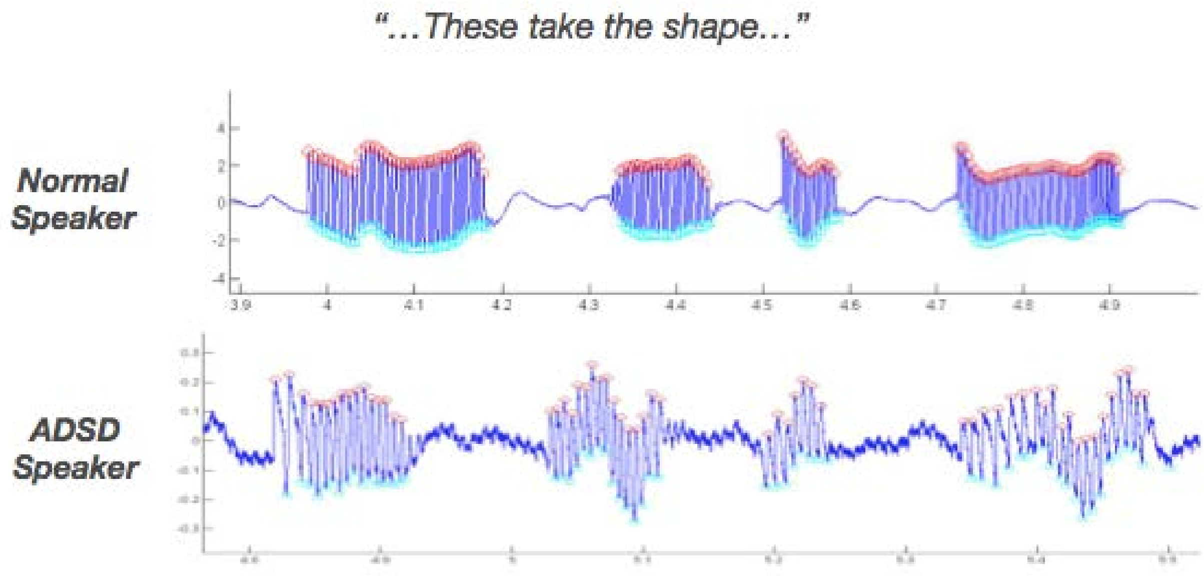

Figure 8. EGG data from a normal and ADSD speaker, on the phrase "These take the shape." Red circles denote cycle maximums and blue circles denote cycle minimums.

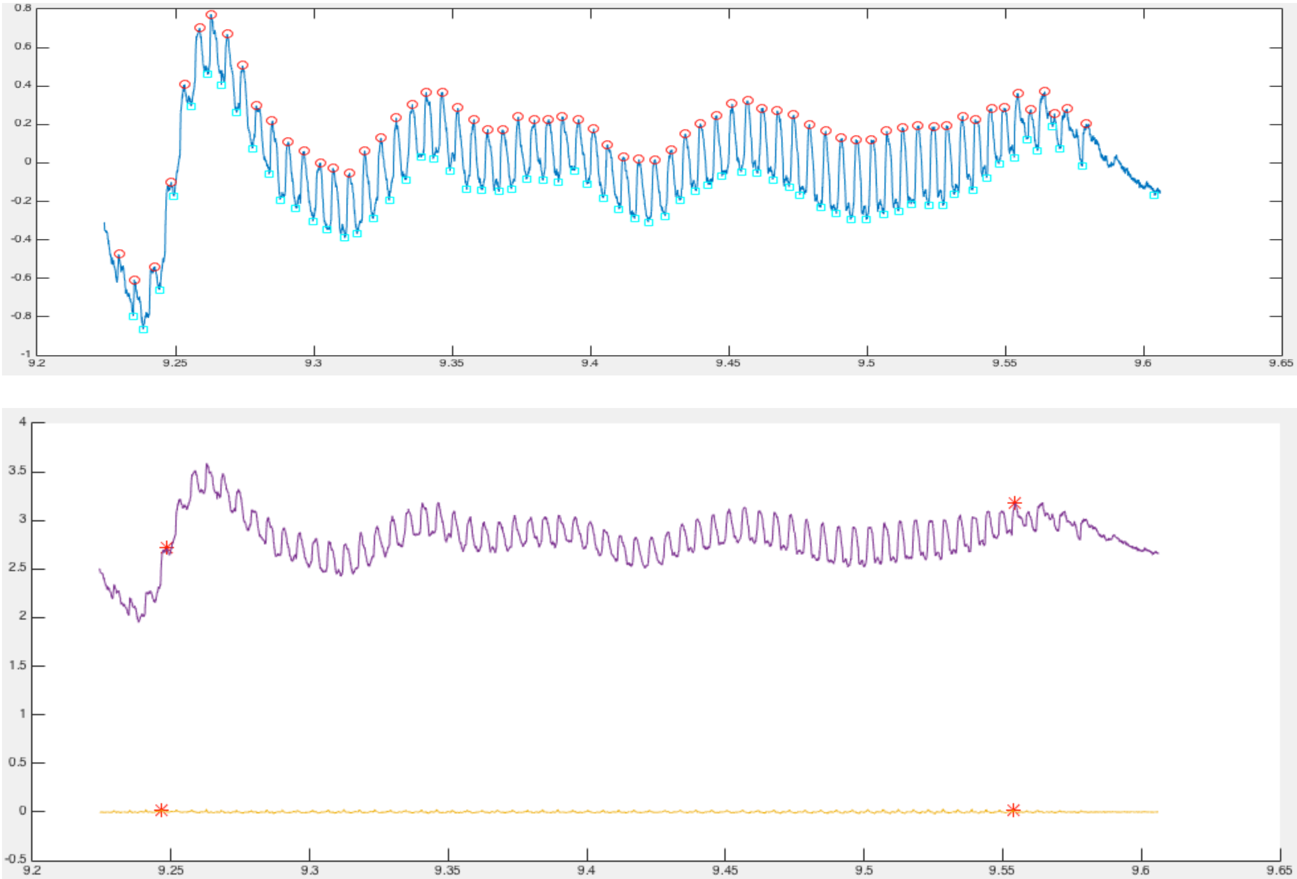

Figure 9. (top) Peak-picking in moderately dysphonic data set. (bottom) Rejected cycles marked by asterisks, shown on EGG (purple waveform) and DEGG (yellow waveform) and the yellow waveform being the DEGG.

Evaluation of Slope Measures

Different contact closing slope measures, SC2575[9] and SC1090[18], have been reported, as well as one opening slope measure, SO9050[18]. To determine which contact closing and contact opening slope formulations should be used for reporting in this study, these previously described measures as well as several new measures (SC5075, SO7525, SO7550) were compared in a preliminary analysis to determine which formulations had greater stability, i.e. less fluctuation across a speech segment. The coefficient of variance (COV-standard deviation divided by the mean, multiplied by 100), which serves as an index of irregularity, was calculated for each slope measure in the speech segment “long round arch” in the 8 normal and 8 ADSD speakers. This phrase involves continuous voicing and was also used in previous studies of ADSD.[30]

The results, shown in Figure 10, show that SC5075 had the lowest COV among both normal speakers and ADSD speakers regardless of which baseline correction method was applied. The previously reported SC1090 had the greatest COV. These results are consistent with the observation that the closing segment below 25% height is susceptible to waveform artifacts.[9] Of the 3 opening slope measures tested, SO7525 had the lowest COV among normal speakers, whereas among ADSD speakers the COV for SO9050 was slightly lower than that of SO7525, by only 5%. Based on these comparisons, **SC5075 and SO7525** were chosen to represent the contact closing and opening slopes, respectively, for subsequent analysis. Not surprisingly, ADSD speakers had greater COV for all slope measures than normal speakers.

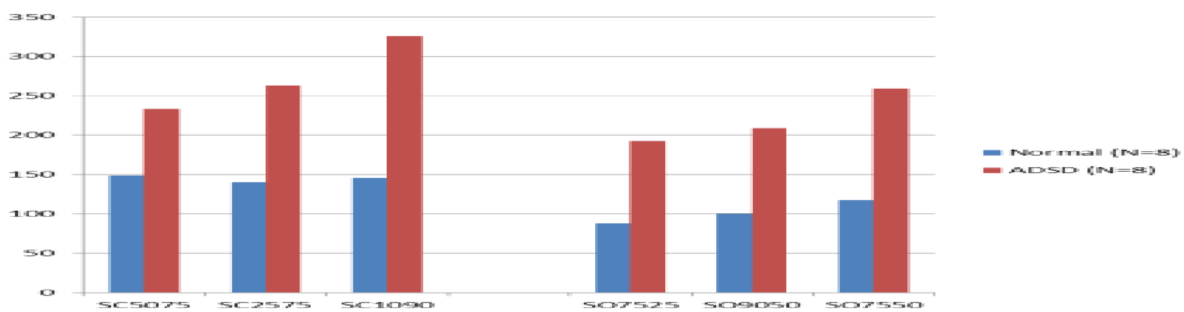

Figure 10. Comparison of slope measures. Importantly, SC5075 and SO7525 seen to have the least COV.

EGG Parameters in Normal vs ADSD Speakers

To determine if EGG waveform features can distinguish between normal and ADSD speech, the CQ, EGGW50, peak skew, SC5075, and SO7525 were calculated for the phrase "long round arch". The results are shown in Table 1. None of the parameters are statistically different between normal and ADSD speakers.

	High-Pass Filter			Linear Correction		
	Normal	**ADSD**	***p***	**Normal**	**ADSD**	***p***
CQ	0.486 (0.067)	0.491 (0.068)	0.85	0.486 (0.068)	0.490 (0.069)	0.88

EGGW50	0.454 (0.039)	0.480 (0.062)	0.24	0.454 (0.040)	0.476 (0.065)	0.32
Peak Skew	0.299 (0.058)	0.311(0.045)	0.57	0.299 (0.058)	0.319 (0.052)	0.39
SC5075	11.5 (2.9)	10.6 (3.0)	0.46	11.4 (2.9)	10.4 (3.1)	0.40
SO7525	-2.25 (0.36)	-2.15 (0.64)	0.65	-2.24 (0.36)	-2.20 (0.66)	0.86

Table 1. EGG waveform parameters in normal vs. ADSD speech. Values are expressed as mean (standard deviation) (N=12).

Strained vs Unstrained

We tested the hypothesis that EGG can differentiate between strained and unstrained voicing within the same speaker. Consensus strained syllables were identified in 9 of the 12 ADSD samples. These samples were rated to have high strain variability, i.e. at least 25% on the VAS. The syllables differed between samples (Table 2). These syllables were used as the strained token for within-subject comparison. The phrase “take the shape” was used as the unstrained token for all except one sample.

ADSD Subject ID	Strained syllables
1	long round
2	path
3	light, colors, arch
6	many, arch, above
7	rainbow, many
8	rain, long
9	light, these, arch, path
10	bow, long, high, ends
11	above, horizon

Table 2. Consensus strained syllables by speaker

EGG parameters were compared between the strained token and the unstrained token for each speaker, and the results are shown in Table 3. The opening slope SO7525 was significantly different between the strained and unstrained tokens ($t(8) = 3.11$, $P = 0.015$ and $t(8) = 2.78$, $P = 0.024$ for high-pass filtered and linearly-corrected data, respectively), with a large effect size (Cohen’s $d = 1.04$ and 0.93 for high-pass filtered and linearly-corrected data, respectively). The other parameters were not significantly different between the strained and unstrained tokens.

	High-Pass Filter	Linear Correction

	Strained	Unstrained	*p*	Strained	Unstrained	*p*
CQ	0.513 (0.076)	0.482 (0.073)	0.36	0.513 (0.076)	0.480 (0.073)	0.34
EGGW50	0.501 (0.065)	0.459 (0.074)	0.11	0.503 (0.060)	0.458 (0.072)	0.072
Peak Skew	0.303 (0.039)	0.327 (0.087)	0.46	0.331 (0.060)	0.307 (0.095)	0.53
SC5075	10.3 (3.7)	12.7 (4.8)	0.11	10.0 (3.7)	12.5 (4.8)	0.095
SO7525	-2.13 (0.72)	-2.53 (0.69)	0.015	-2.18 (0.72)	-2.54 (0.73)	0.024

Table 3. EGG waveform parameters in strained vs. unstrained syllables in ADSD speech. Values are expressed as mean (standard deviation) (N=9).

EGG-Based Strain-stress Pattern of Entire Passage

Since SO7525 appeared to correlate with listener's perception of strain within a speech segment, this parameter was plotted for the entire speech passage to obtain a visual representation of parameter variation in continuous speech. An example is shown in Figure 11.

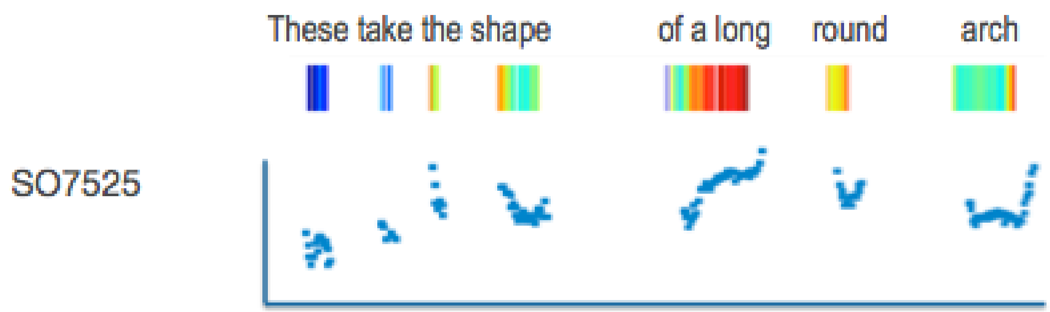

Figure 11. SO7525 visual representation of parameter variation in continuous speech.

DISCUSSION

The overall aim of this work was to identify objective or instrumental measures that correlated with the auditory perceptual quality of vocal strain. The two specific goals were to develop the methodology to process dysphonic EGG signal from connected speech, and to use it to examine the clinical utility of EGG measures as they pertain to vocal strain.

Objective vs Subjective

EGG analysis has the promise of objectifying a task that is difficult to quantify perceptually. It is difficult for listeners to differentiate a small amount of strain in running speech, so the goal of this study is to detect differences that we can hear but otherwise could not measure. When the EGG signal is carefully processed and the rating tasks are correctly defined, certain EGG waveform parameters correlate with perceived strain in ADSD speech. The clinical utility of an objective method of quantifying vocal strain is to measure symptom severity and monitor progress with behavioral, pharmaceutical, or surgical intervention.

Signal Processing of EGG from Dysphonic Connected Speech

The first goal of the study was to develop a robust algorithm to automate the extraction of EGG waveform measures from dysphonic speech. The first step towards meaningful data extraction was to systematically evaluate the effect of baseline correction on those measures in order to determine an optimal approach to baseline correction. Historically, a high-pass filter has been applied to the raw EGG signal to remove the DC drift from laryngeal movement in connected speech. However, the cutoff frequency used was often not reported in publications and/or not specified by the equipment manufacturer. Our data showed that, for dysphonic EGG signals, a high-pass filter with cutoff greater than 10 Hz resulted in increased deviation of parameter values from those obtained from the raw signal collected with a 10 Hz low-frequency limit in the hardware. We concluded that a 10 Hz high-pass filter achieved a good balance between sufficient reduction of DC drift and minimizing signal distortion. We further showed that a per-cycle linear correction scheme produced results comparable to a 10 Hz filter. By omitting the high-pass filter, the linear correction method has the advantage of avoiding the arbitrariness of choosing a frequency cutoff for the filter. The disadvantage of linear correction is the increased complexity of data handling. To have the highest confidence in interpreting the EGG parameter values, we recommend both baseline correction methods to be employed.

The major challenge in automating EGG signal processing from dysphonic connected speech is achieving the balance between selecting waveform cycles that represent true glottal cycles and rejecting waveform cycles that are spurious or highly dysmorphic. We implemented several strategies in an attempt to achieve such a balance. Following baseline correction, high frequency noise in the EGG signal was removed to improve the signal-to-noise in the DEGG signal. Automated cycle selection was carried out in the DEGG domain, following the approach of Henrich[26]. We imposed cycle rejection criteria based on several measures of waveform morphology to exclude cycles that have a high likelihood of being false positives. The underlying rationale for our approach was that the mean value of an EGG parameter will be representative of a selected speech segment if most cycles included in the calculation are true glottal cycles. We believe we have arrived at a reasonable balance of cycle acceptance/rejection (e.g. Figure 12) through an iterative process of optimizing the rejection criteria.

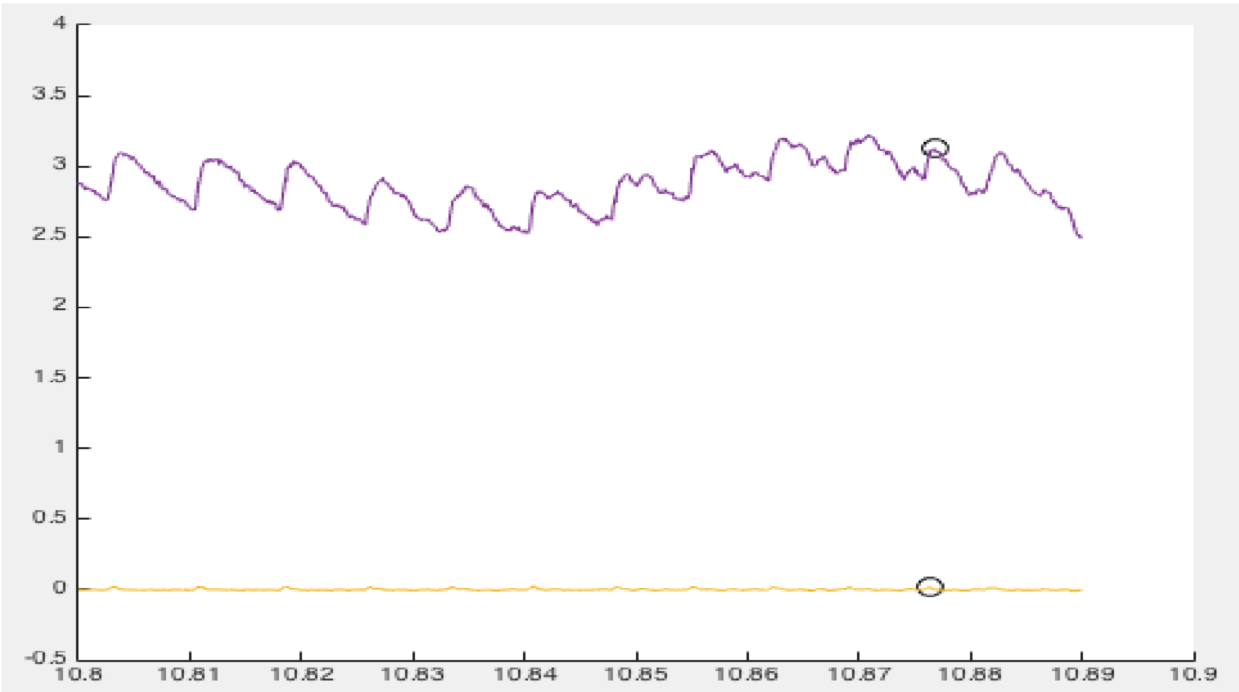

Figure 12. EGG (purple) and DEGG (yellow) in moderately dysphonic data with EGG/DEGG outlier combo shown (black circles).

A major advantage of the automated cycle-selection algorithm developed in this work is it allows arbitrary speech segment selection. The analysis does not depend on manual selection of the most stable mid-segment of a sustained syllable. The user can select the speech segment to encompass an entire word or phrase, with the starting and ending points in non-phonated breaks or in the phonated transition between two syllables.

EGG Measures Do Not Differentiate Normal from Dysphonic Voice

We found that none of the EGG parameters investigated were significantly different between normal speakers and ADSD speakers. Substantial within-group inter-subject variability limited the utility of EGG measures as a means to differentiate the two groups. High inter-subject variability has been noted in acoustic and aerodynamic measures[31]. High EGG CQ variability among healthy subjects is noted in Verdolini[16].

Baseline Correction

As this is the first work looking at EGG in continuous speech, careful attention must be given to how the baseline is treated. The baseline is subject to drift because of laryngeal movement during continuous speech. In order to properly analyze the cycles within the EGG signal, this baseline must be corrected for first. The two methods described above, using a Butterworth 10 Hz filter and linear correction, both seem to produce comparable results. Choosing a Butterworth cutoff frequency is somewhat arbitrary, which linear correction avoids. Additionally, linear correction does not depend on neighboring cycles or the length of the EGG selection. However, the downside of linear correction is that data handling is more complex whereas the Butterworth filter is straightforward, as it is a built-in MATLAB function. The cutoff frequency must be chosen to where the signal distortion that occurs at higher frequencies is balanced against the less robust baseline flattening at lower cutoff frequencies. The 10 Hz Butterworth filter provided the best balance in our study.

Whole versus Partial Segment Selection

Selection of a waveform segment in the EGG sample is equally as important as correcting the baseline. A previous study looked for an accurate and reliable method of selecting appropriate syllable segments, specifically aiming at identifying the steadiest segment of the voice sample[21]. It was found that selecting certain parts of a vowel segment as opposed to the entire segment introduces significant inconsistencies. For this reason, our analysis was performed on the entirety of a syllable.

Parameter Selection and Outcomes

Parameter selection was novel in our study, in that we explored measures that have not been previously used. Out of the three opening slopes measured, SO7525 was found to have the least coefficient of variance and was used for inter-subject and intra-subject comparisons. SC1090 is not a reliable measure for dysphonic data because the closing phase could have a large lag that accounts for more than 10% of the height[17]. The need for a higher cutoff, of at least 25%, has been previously mentioned[9]. Contact quotient (CQ) was unable to differentiate between controls and ADSD patients. In a previous study, it was found that "a specific range of CQs cannot be identified for resonant voice across all subjects[16]." SO7525 was able to show a significant difference within intra-subject comparisons, indicated by Table 3 in the Results section.

Advantages and Limitations

Advantages of using the EGG signal is that it does not need a quiet environment for collection, which lends well to simultaneous acoustic recording. Additionally, it only includes phonated information without consonant or aspirated sounds.

Though our study has used the EGG signal in a way it has not been before, there are limitations that must be discussed. Inherent to the EGG signal, signal-to-noise ratio is low in a patient with significant fat around their neck, making it difficult for DEGG peaks to be detected by our algorithm.

We also acknowledge that the numerical results include artifacts from other things, like frequency-dependent changes in laryngeal configuration, and supraglottic contact, even mucous bridges that can change the impedance. Specifically, constant pitches have been used in the past to avoid the effect of laryngeal drift on the EGG data[16]. However, this study is an attempt to see if, despite the possible presence of the aforementioned artifacts, we can still obtain a correlation with strain.

Absolute values of parameters from one syllable to another cannot be equated, as the effect from laryngeal drift during normal speech cannot be accounted for. As with many physiologic measures, inter-subject variability is high. High inter-subject variability is a common feature of acoustic and aerodynamic measures[31]. High EGG CQ variability among healthy subjects is noted in Verdolini[16].

Future Directions

Our intent is to identify waveform morphologic features that can be clinically useful. We are less interested in the individual parameters rather than groups of parameters, which is the direction we would like to take our EGG analysis in the future. EGG should be included as part of multidimensional voice assessment, since it may provide information complementary to acoustic measures[12]. Additionally, treatment response for spasmodic dysphonia needs to be multidimensional[32]. The potential contribution of EGG to clinical voice assessment should be a focus for future studies[33].

Specifically, for intra-subject data, EGG can be used as a baseline against which future comparisons can be made. Additionally, aside from spasmodic dysphonia, it can be applied to other voice disorders in which strain is prominent. It can also be used to differentiate between types of dysphonia (e.g. muscle tension dysphonia and spasmodic dysphonia).

CONCLUSION

These results provide further insight into the utility and limitations of EGG. The data suggest that meaningful information can be obtained from EGG of continuous dysphonic speech. While EGG may have limited utility in inter-subject comparison because of overlapping parameter values between normal and strained phonation, it may provide a useful objective measure of vocal strain in the same subject with variable degrees of strain or over time. Our findings support the conclusion from previous work using opening slope measures[17]. More extensive analysis of other speech tokens in the reading passage is required to confirm our preliminary finding and to determine its sensitivity.

ACKNOWLEDGMENTS

We thank our colleagues Janis Deane, MEd, CCC-SLP, Amy Hamilton, MA, CCC-SLP, Laura Toles, MS, CCC-SLP, and Lesley Childs, MD for participating in the perceptual voice assessments.

REFERENCES

1. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. Am J Speech Lang Pathol 2009; 18:124-132.
2. Hirano M. *Clinical examination of the voice*. New York: Springer-Verlag, 1981:83-84.
3. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. J Speech Hear Res 1993; 36:21-40.
4. Netsell R, Lotz W, Shaughnessy AL. Laryngeal aerodynamics associated with selected voice disorders. Am J Otolaryngol 1984; 5:397-403.
5. Grillo EU, Verdolini K. Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects. J Voice 2008; 22:546-552.
6. Pinczower R, Oates J. Vocal projection in actors: the long-term average spectral features that distinguish comfortable acting voice from voicing with maximal projection in male actors. J Voice 2005; 19:440-453.
7. Lowell SY, Kelley RT, Awan SN, Colton RH, Chan NH. Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. Ann Otol Rhinol Laryngol 2012; 121:539-548.
8. Watts CR, Awan SN. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. J Speech Lang Hear Res 2011; 54:1525-1537.
9. Orlikoff RF. Assessment of the dynamics of vocal fold contact from the electroglottogram: data from normal male subjects. J Speech Hear Res 1991; 34:1066-1072.

10. Rothenberg M, Mahshie JJ. Monitoring vocal fold abduction through vocal fold contact area. J Speech Hear Res 1988; 31:338-351.

11. Childers DG, Hicks DM, Moore GP, Alsaka YA. A model for vocal fold vibratory motion, contact area, and the electroglottogram. J Acoust Soc Am 1986; 80:1309-1320.

12. Scherer RC, Vail V. Measures of laryngeal adduction. J Acoust Soc Am 1988; 84(suppl 1).

13. Scherer RC, Vail VJ, Rockwell B. Examination of the laryngeal adduction measure EGGW. In: Bell-Berti F, Rapheal LJ, eds. *Producing Speech: Contemporary Issues*. Woodbury, NY: American Institute of Physics, 1995:269-290.

14. Askenfelt AG, Hammarberg B. Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. J Speech Hear Res 1986; 29:50-64.

15. Peterson KL, Verdolini-Marston K, Barkmeier JM, Hoffman HT. Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. Ann Otol Rhinol Laryngol 1994; 103:335-346.

16. Verdolini K, Druker DG, Palmer PM, Samawi H. Laryngeal adduction in resonant voice. J Voice 1998; 12:315-327.

17. Fisher KV, Scherer RC, Swank PR, Giddens C, Patten D. Electroglottographic tracking of phonatory response to Botox. J Voice 1999; 13:203-218.

18. Fisher KV, Scherer RC, Guo CG, Owen AS. Longitudinal phonatory characteristics after botulinum toxin type A injection. J Speech Hear Res 1996; 39:968-980.

19. Vertigan AE, Theodoros DG, Winkworth AL, Gibson PG. Acoustic and electroglottographic voice characteristics in chronic cough and paradoxical vocal fold movement. Folia Phoniatr Logop 2008; 60:210-216.

20. Sapienza CM, Walton S, Murry T. Acoustic variations in adductor spasmodic dysphonia as a function of speech task. J Speech Lang Hear Res 1999; 42:127-140.

21. Choi SH, Lee J, Sprecher AJ, Jiang JJ. The Effect of Segment Selection on Acoustic Analysis. J Voice 2011.

22. Cannito MP, Doiuchi M, Murry T, Woodson GE. Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates. J Voice 2012; 26:818 e815-813.

23. Ludlow CL, Adler CH, Berke GSet al. Research priorities in spasmodic dysphonia. Otolaryngol Head Neck Surg 2008; 139:495-505.

24. Titze IR. Interpretation of the Electroglottographic Signal. Journal of Voice 1990; 4:1-9.

25. Childers DG, Hicks DM, Moore GP, Eskenazi L, Lalwani AL. Electroglottography and vocal fold physiology. J Speech Hear Res 1990; 33:245-254.

26. Henrich N, d'Alessandro C, Doval B, Castellengo M. On the use of the derivative of e lectroglottographic signals for characterization of nonpathological phonation. J Acoust Soc Am 2004; 115:1321-1332.

27. Luo JW, Bai J, He P, Ying K. Axial strain calculation using a low-pass digital differentiator in ultrasound elastography. Ieee T Ultrason Ferr 2004; 51:1119-1127.

28. Usui S, Amidror I. Digital Low-Pass Differentiation for Biological Signal-Processing. Ieee T Bio-Med Eng 1982; 29:686-693.

29. Kadel RP, Kip KE. A SAS macro to compute effect size and its confidence interval from raw survey data *Annual Southeastern SAS Users Group Conference*, 2012.

30. Sapienza CM, Cannito MP, Murry T, Branski R, Woodson G. Acoustic variations in reading produced by speakers with spasmodic dysphonia pre-botox injection and within early stages of post-botox injection. J Speech Lang Hear Res 2002; 45:830-843.

31. Behrman A. Common practices of voice therapists in the evaluation of patients. J Voice 2005; 19:454-469.

32. Dejonckere PH, Neumann KJ, Moerman MB, Martens JP, Giordano A, Manfredi C.

Tridimensional assessment of adductor spasmodic dysphonia pre- and post-treatment with Botulinum toxin. Eur Arch Otorhinolaryngology. 2012 Apr;269(4):1195-203. doi: 10.1007/s00405-011-1890-6. Epub 2011 Dec 31.

33. Roy N, Barkmeier-Kraemer J, Eadie T, Sivasankar MP, Mehta D, Paul D, Hillman R. Evidence-based clinical **voice** assessment: a systematic review. Am J Speech Lang Pathol. 2013 May;22(2):212-26. doi: 10.1044/1058-0360(2012/12-0014). Epub 2012 Nov 26. Review.