Published in final edited form as:

J Voice. 2016 November ; 30(6): 770.e9–770.e21. doi:10.1016/j.jvoice.2015.11.005.

A measure of the auditory-perceptual quality of strain from electroglottographic analysis of continuous dysphonic speech: Application to adductor spasmodic dysphonia

Keerthan Somanath, BSEE and Ted Mau, MD, PhD

Clinical Center for Voice Care, Department of Otolaryngology-Head and Neck Surgery, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A

Abstract

Objectives—(1) To develop an automated algorithm to analyze electroglottographic (EGG) signal in continuous, dysphonic speech, and (2) to identify EGG waveform parameters that correlate with the auditory-perceptual quality of strain in the speech of patients with adductor spasmodic dysphonia (ADSD).

Study Design—Software development with application in a prospective controlled study.

Methods—EGG was recorded from 12 normal speakers and 12 subjects with ADSD reading excerpts from the Rainbow Passage. Data were processed by a new algorithm developed with the specific goal of analyzing continuous dysphonic speech. The contact quotient (CQ), pulse width (EGGW), a new parameter peak skew, and various contact closing slope quotient (SC) and contact opening slope quotient (SO) measures were extracted. EGG parameters were compared between normal and ADSD speech. Within the ADSD group, intra-subject comparison was also made between perceptually strained syllables and unstrained syllables.

Results—The opening slope quotient SO7525 distinguished strained syllables from unstrained syllables in continuous speech within individual ADSD subjects. The standard deviations, but not the means, of CQ, EGGW50, peak skew, and SO7525 were different between normal and ADSD speakers. The strain-stress pattern in continuous speech can be visualized as color gradients based on the variation of EGG parameter values.

Conclusions—EGG parameters may provide a within-subject measure of vocal strain and serve as a marker for treatment response. The addition of EGG to multi-dimensional assessment may lead to improved characterization of the voice disturbance in ADSD.

Send correspondence to: Ted Mau, MD, PhD, Department of Otolaryngology-Head and Neck Surgery, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390, Phone: (214) 648-2042, Fax: (214) 648-9122, ted.mau@utsouthwestern.edu. This work was presented in part at the 10th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research, June 4, 2013, Cincinnati, Ohio.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

electroglottography; EGG; vocal strain; continuous speech; spasmodic dysphonia

INTRODUCTION

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. Auditory-perceptual measures have been the conventional benchmark for the assessment of vocal strain. 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 evaluation³, 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/pa/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 continuous 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 (e.g., Watts & Awan⁸; Lowell et al.⁷). On the other hand, while the aerodynamic parameters of flow, pressure, and flow resistance are directionally sensitive to dimensions such as strain or breathiness⁵, aerodynamic measurements are typically collected via vocal tasks that may not fully reflect the pathology of the voice disorder as it manifests in continuous speech.

A physiologic measurement that could in principle quantify some aspect of vocal strain and is readily obtained in continuous 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. As such, it provides a physiologically derived measure of vocal function connected to the oscillatory kinematics at the sound source.^{9–11} Two studies have demonstrated a relationship between the EGG output and the

vocal fold contact area (VFCA). Scherer et al.¹² employed a mechanically-driven contact-decontact model in excised canine hemilarynges to show that an approximately linear relationship between the normalized EGG value and VFCA was obtained in some cases. Hampala et al.¹³ more recently used an excised deer hemilarynx phonation model to show that the agreement between EGG and VFCA was better in the decontacting phase than in the contacting phase. Other studies suggested that EGG pulse width measures relate to the degree of vocal fold medial compression.^{14,15} While other elements of hyperfunctional phonatory behavior such as supraglottic squeeze, excessive subglottal pressure, and hyperfunction of extrinsic laryngeal muscles may also contribute to the auditory-perception of vocal strain, strain might be primarily associated with the degree of vocal fold compression.¹⁶ 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.^{17,18} 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.¹⁹

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.^{19,20}, 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 part of the information content inherent 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^{19,20} 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 almost all previous studies, EGG data were collected on a segment of sustained vowel in order to minimize signal baseline drift encountered in continuous speech. While the CQ from continuous speech has been reported (e.g., Zagolski & Carlson²¹; Vertigan et al.²²), it is unclear how the baseline drift was handled and how waveform cycles were selected for analysis.

There are several rationales to extend the capability of EGG analysis to continuous speech. First, compared to sustained vowels or single syllables, continuous speech is a more complete representation of speaking patterns that constitute dysphonic symptoms in daily activity and may be a more relevant target of evaluation.²³ Second, the type and frequency of phonatory aberrations may differ between sustained vowel and continuous speech, as has been shown in ADSD.²⁴ Finally, analysis of continuous speech in principle avoids the significant inconsistencies introduced by the manual selection of a sustained vowel segment

for analysis.²⁵ There is, however, a knowledge gap on how the EGG signal in continuous speech is best handled. The calculation of EGG waveform parameters beyond CQ is affected by the substantial DC drift in continuous 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 a methodology to analyze EGG signals in the continuous speech of normal and dysphonic speakers and 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. This initial application focused on one specific voice disorder in order to avoid the heterogeneity of disease-specific EGG findings across different disorders. ADSD was chosen 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.²³ (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 content²⁶, yet this determination is dependent on listener training and experience. There is therefore a need for objective methods to identify stress-strain variation. (5) Since ADSD symptom is sensitive to phonetic content, the less-strained segments of speech can serve as internal control for each subject.

This study tests the following hypotheses: (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. Particular attention is given to 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 at all to use EGG to detect vocal strain in continuous speech, the twelve ADSD speakers were selected for relatively severe vocal symptoms at the time of data collection compared to their asymptomatic treated baseline, out of a pool of 40 screened. The 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 (age range: 40–86 years; $M = 60$, $SD = 12$), 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-treated 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. The 12 ADSD speakers included 8 women and 4 men. The 12 normal speakers included 8 women and 4 men with no vocal pathology (age range: 24–45 years; $M = 32$, $SD = 6$). At the time of data collection, their voices were perceptually normal.

Speakers were asked to read a two-sentence excerpt from the Rainbow Passage at 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. The lower cutoff of 10 Hz was chosen 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-R2015a, 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 Processing: Baseline Correction

Two alternate 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. The user can choose which one to use from the GUI, and the selected method is automatically applied.

Butterworth High-pass Filter—High-pass filtering is known to change EGG waveform morphology. To determine the extent to which high-pass filtering affected EGG waveform morphology, a preliminary analysis using a limited dataset was performed. Select EGG parameters (SC5075, SC2575, SO9050, SO7525, EGGW50, and CQ, defined below) in a test sentence in 6 ADSD subjects 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”. The mean values of the parameters in this test sentence were compared with their corresponding values calculated from uncorrected EGG, i.e., without high-pass filtering. The results are shown in Figure 1. With a cutoff frequency of 10 Hz, approximately 90% of the parameters remained within 1% of their values calculated from uncorrected EGG. However, when the cutoff was increased to 20 Hz, the percentage of parameters that remained within 1% of uncorrected values dropped to about 80%. Based on this preliminary analysis, a 10 Hz cutoff for the Butterworth filter was thought to provide a reasonable compromise between sufficient baseline correction and acceptably small changes in the EGG parameters. A 10 Hz cutoff was therefore utilized for subsequent parameter calculations.

Cycle-based Linear Correction—In this method, suggested by Ron Scherer^{14,15}, a linear baseline correction was applied to each EGG cycle independent from other cycles. This correction is based on the assumption that the DC shift of the EGG waveform over a single cycle most likely is due to a relative position change between the electrodes and the vocal folds, so that the baseline shift is an essentially additive linear distortion of the EGG waveform. Thus, the linear baseline correction is an approximate correction to this relative shift. It is assumed that the baseline shift is not a vocal fold contact shift per se. The EGG values over the cycle were corrected by an amount linearly proportional to the baseline offset over the entire cycle as shown in Figure 2.

Definition of EGG Waveform Parameters

The parameter definitions are illustrated in Figure 3. **Contact quotient (CQ)** is defined by maxima and minima in the derivative of the EGG (DEGG) waveform as in Scherer et al.¹⁵ **EGGW50** is the pulse width at the 50% amplitude level as in Scherer et al.¹⁵ Contact opening and contact closing slopes are defined after Orlikoff⁹ and Fisher et al.²⁰, and the normalized version (slope quotient) in Fisher et al.²⁰ was used. In the current work, the closing height (H_C) and opening height (H_O) of each cycle were not assumed to be the same and were determined separately. The contact closing slope quotient **SC2575** 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 25% and 75% of the closing height H_C , A is the time segment corresponding to B, and T is the period of that cycle. **SC1090** is similarly defined to denote the slope quotient between 10% and 90% of the closing height, and **SC5075** the slope quotient between 50% and 75% of the closing height. The opening slope quotients **SO7525**, **SO7550**, and **SO9050** are defined analogously and normalized to the opening height H_O instead of H_C . The parameter **peak skew** denotes the relative position of the EGG cycle peak within the cycle period. 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 Titze²⁷ and Childers and Lee²⁸, it is used here solely as an empiric parameter without inference to glottal shape or kinematics based on its value.

Signal Processing: Cycle Selection

Phonatory cycles in the EGG signal were identified indirectly using the DEGG. The premise of this approach has been illustrated previously (e.g., Childers et al.²⁹; Henrich et al.³⁰), 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 a Savitzky-Golay (polynomial) smoothing filter. The DEGG was then generated using a low-pass differentiation filter.^{31,32} 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. The search was performed on the un-smoothed EGG wave form to avoid the minor distortion caused by the smoothing filter. EGG minima were then identified by a local minimum search between successive maxima. A schematic of the overall algorithm is shown in Figure 4.

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. The criteria were biased 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, cycles with questionable morphologies were rejected. 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 $CQ < 0.1$, $CQ > 0.95$, 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 $\text{EGGW50} < 0.1$, $\text{EGGW50} > 0.95$, peak skew < 0.1 , and peak skew > 0.8 were implemented, since cycles with any of these morphologic features were 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 Rating Tasks

Three speech-language pathologists with a primary practice focus on voice evaluation and treatment 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) with the quartiles marked, 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 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. 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 9th sample.

Statistical Analysis

Statistical calculations were performed using SAS 9.3 (SAS Institute Inc., Cary, NC). 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.³³ To determine the correlation between EGG parameters, Spearman’s rho was calculated for pairwise correlations between CQ, EGGW50, peak skew, SC5075, and SO7525 within the phrase “long round arch” for all 24 subjects individually, i.e., using the values from each phonatory cycle, not the mean value for the entire segment.

RESULTS

Automated EGG Cycle Selection/Rejection

Figure 5 shows an example of how the algorithm handles dysphonic EGG data. The algorithm was designed to identify cyclic changes in the EGG waveform that are likely to correspond to phonatory cycles rather than noise. It then rejects cycles with morphologic features that lie significantly outside the norm. The insets highlight the capability of the algorithm to identify EGG cycles that are not normal but still represent true phonatory cycles.

Evaluation of Slope Measures

Different contact closing slope measures (SC2575 in Orlikoff¹¹ and SC1090 in Fisher et al.²⁰) have been reported, as well as one opening slope measure (SO9050 in Fisher et al.²⁰). 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 variation (CV – defined as 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 12 normal and 12 ADSD speakers. This phrase involves continuous voicing and was also used in previous studies of ADSD.³⁴ A smaller CV indicates that the measure has less relative variation from the mean over the speech segment and can be considered more stable, but it does not mean greater validity. The results, shown in Figure 6, show that SC5075 had the lowest CV among both normal speakers and ADSD speakers regardless of which baseline correction method was applied. The previously reported SC1090 had the highest CV. These results are consistent with the observation that the closing segment below 25% height is susceptible to waveform artifacts.¹¹ Based on the CV measure, SC5075 was chosen to represent the closing slope quotient for subsequent analysis. While this measure only covers 25% of the height range, it is possible that the 50–75% segment is sufficiently sensitive as a measure of the opening gesture. Of the 3 opening slope measures tested, SO7525 had the lowest CV among normal speakers, whereas among ADSD speakers the CV for SO9050 was slightly lower than that of SO7525, by only 5%. Based on these comparisons, SO7525 was chosen to represent the contact opening slope quotient for subsequent analysis. Not surprisingly, ADSD speakers had greater CV for all slope measures than normal speakers.

Correlation Between EGG Parameters

Since the EGG parameters for each cycle represent different aspects of the same waveform, it is reasonable to expect some degree of correlation between parameters. For each speech sample, a correlation can be calculated between two parameters obtained for all cycles within that sample. Figure 7 shows the correlation for all pairwise comparisons for each subject. Only correlations computed with a *p*-value less than 0.01 are shown. CQ and EGGW50, as well as SO7525 and EGGW50, showed a moderately strong positive correlation. SO7525 and peak skew, as well as EGGW50 and peak skew, showed weak to

moderate negative correlation. The other pairwise correlations were less consistently observed and no definitive conclusions can be drawn.

EGG Parameters in Normal vs ADSD Speakers

To determine if EGG waveform features could 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 were statistically different between normal and ADSD speakers.

As a measure of the stability of the EGG parameters within a speech segment, their standard deviations were compared and shown in Table 2. CQ, EGGW50, peak skew, and SO7525 had larger standard deviations in ADSD speech than in normal speech, reflecting greater instability of these waveform parameters in ADSD.

Intra-subject Strained vs Unstrained Syllables

The hypothesis that EGG can differentiate between strained and unstrained syllables within the same speaker was tested. Listeners were asked to rate the “strain variability” within each sample (see Methods section), as well as to identify strained and unstrained syllables in each. The overall inter-rater agreement for the rating of “strain variability” was fair (Kendall’s coefficient of concordance = 0.67). Consensus strained syllables were identified in 9 of the 12 ADSD samples (Table 3). These syllables were used as the strained token for intra-subject comparison. The phrase “take the shape” was used as the unstrained token for all except one sample. EGG parameters were compared between the strained token and the unstrained token for each speaker, and the results are shown in Table 4. The opening slope quotient SO7525 differed between the strained and unstrained tokens ($t(8) = 2.58$, $p = 0.032$ for linearly-corrected data and $t(8) = 3.11$, $p = 0.015$ for high-pass filtered data), with a large effect size (Cohen’s $d = 0.86$ and 1.04 for linearly-corrected data and high-pass filtered data, respectively). The other parameters were not significantly different between the strained and unstrained tokens.

EGG-based Strain-stress Pattern of Speech Segments

The preceding analyses solely depended on EGG parameter values averaged over all cycles within a speech segment and comparing the mean values between segments. Another way to depict EGG data is to graph the temporal variation of the parameter values throughout a speech segment, as illustrated in Figure 8. Although less quantitative than a comparison of means, the visual display of EGG parameter variation may constitute a closer parallel of auditory perceptual judgement of a speech segment since both share the same temporal axis. For example, it is straightforward to appreciate “long” and “round” as the strained syllables in the segment shown in Figure 8, where the “hotter” colors correspond with larger CQ, larger EGGW50, and less negative SO7525.

DISCUSSION

The overall goals of this work were to establish the feasibility of automated waveform parameter extraction from EGG data that contain high aperiodicity, distorted waveforms, and

high-frequency noise, and to identify objective or instrumental measures that correlated with the auditory perceptual quality of vocal strain. To the authors' knowledge, this represents the first effort to develop software with explicit consideration for automated handling of the large baseline drifts, significant high frequency noise, and low signal-to-noise inherent in EGG signals from dysphonic speech.

Signal Processing of EGG from Dysphonic Continuous Speech

The first step towards meaningful data extraction was to systematically evaluate the effect of baseline correction on EGG waveform parameters 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 continuous speech. However, the cutoff frequency used was often not reported in publications and/or not specified by the equipment manufacturer. The present data showed that, for dysphonic EGG signals, a software 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. It was concluded that a 10 Hz high-pass filter achieved a good balance between sufficient reduction of DC drift and minimizing signal distortion. It was further shown that a per-cycle linear correction scheme produced results comparable to a 10 Hz filter. The linear correction method has the advantage of avoiding the arbitrariness of setting a frequency cutoff for the high-pass filter. The disadvantage of linear correction is the increased complexity of data handling. To have the highest confidence in interpreting the EGG parameter values, both baseline correction methods were used (in alternate fashion) to make sure their results were consistent with each other.

The major challenge in automating EGG signal processing from dysphonic continuous speech is achieving the balance between selecting waveform cycles that represent true glottal cycles and rejecting waveform cycles that are spurious or highly dysmorphic. Several strategies were implemented in an attempt to achieve such a balance. Following baseline correction, high frequency noise in the EGG signal was removed with a Savitzky-Golay smoothing filter 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.³⁰ To minimize the effect of EGG waveform distortion produced by the smoothing filter, waveform parameters were calculated from the EGG waveform without the filter applied. Cycle rejection criteria were imposed based on several measures of waveform morphology to exclude cycles that had a high likelihood of being false positives. The underlying rationale for this 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 were true glottal cycles. The authors believe they have arrived at a reasonable balance of cycle acceptance/rejection (e.g., Figure 5) through an iterative process of optimizing the rejection criteria.

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.

Correlation Between EGG Parameters

The moderately strong positive correlation between CQ and EGGW50 was expected. The correlation between SO7525 and EGGW50 affirms a similar finding from Fisher et al.²⁰ that had not been further investigated since. Fisher et al.²⁰ suggested that this correlation was consistent with the concept that the EGG waveform width is driven by the speed of vocal fold separation. The negative correlation between SO7525 and peak skew implies that cycles in which the maximal contact is reached earlier in the cycle (smaller peak skew) have slightly more gradual separation (less negative opening slope). The negative correlation between EGGW50 and peak skew implies that as the extent of contact increases, the maximum contact tends to be reached earlier in the cycle. Although measures of vocal fold contact and separation suggested from the EGG signal are not equivalent to airflow measurements, this observation is in general agreement with the finding that patients with ASD have higher maximum flow declination rate (MFDR), which implies that vocal fold adduction and contact occur faster.³⁵

Mean EGG Measures Do Not Differentiate Normal from ASD Speakers

The hypothesis that EGG parameters could differentiate between normal and ASD speakers was tested. None of the EGG parameters investigated were significantly different between normal speakers and ASD speakers. Substantial within-group variability limited the utility of EGG measures as a means to differentiate the two groups. This is consistent with prior work that showed high EGG CQ variability among healthy subjects.¹⁸ High inter-subject variability has also been noted in certain acoustic and aerodynamic measures³⁶, including the mean phonatory air flow in ASD patients.³⁷ While the mean values of the EGG measures did not differentiate normal from ASD speech, the standard deviations of CQ, EGGW50, peak skew, and SO7525 were higher in ASD speakers. This is expected since dysphonic voice has higher short-term (cycle-to-cycle) and long-term (over an entire segment) variability.

Characterization of Intra-subject Strain Variability

The hypothesis that EGG parameters could distinguish perceptually strained speech segments from unstrained segments in the same ASD speaker was also tested. This was motivated by the fact that the diagnosis of ASD relies on auditory-perceptual discrimination of stress-strain patterns in speech that vary according to the phonetic content.²⁶ In the perceptual rating part of this study, the raters found that while it was relatively easy to attribute a speech sample to an ASD speaker based on the overall “gestalt” of the presence of strain variation, it was relatively difficult to pinpoint which syllables were strained and which ones were not. In particular, the inter-rater agreement declined significantly when the degree of strain variation was low within a sample. There was therefore a motivation to objectively measure strain at the syllabic level within a speech sample. It was found that, with the exception of the opening slope quotient, the other EGG measures investigated in this study did not show statistically significant differences between the strained and unstrained syllables (Table 4). Despite the theoretical rationale of EGGW^{15,27} or CQ^{9,38} as a gauge of vocal fold adduction, neither measure sufficiently distinguished the strained syllables from unstrained ones on a statistical basis, although both

were greater in the strained syllables as expected. On one hand, this is consistent with the work of Hosokawa et al.³⁹, who showed that CQ did not change significantly following voice therapy in a group of patients with muscle tension dysphonia. On the other hand, since EGGW50 and CQ calculated in this study represented the means of the individual EGGW50 and CQ of each cycle included by the automated algorithm within the user-selected speech segment, it is possible that they may slightly underestimate the magnitude of EGGW50 and CQ in a strained voice due to the rejection criteria based on $EGGW50 > 0.95$ and $CQ > 0.95$. In other words, waveform cycles corresponding to true glottal cycles fitting these criteria would have been discarded by the algorithm. This was considered to be a reasonable tradeoff in exchange for the ability to reject waveform cycles that did not represent true glottal cycles and had unusual EGGW50 and CQ values.

The opening slope quotient (SO) was the only EGG measure investigated that was significantly different between the strained and unstrained syllables. This was consistent with the work of Fisher et al.²⁰, which showed that SO was among the parameters that showed the strongest correlation to the perceptual quality in the breathy-to-pressed dimension in one ADSD patient over a 10-week period following Botox injection. The follow-up work of Fisher et al.¹⁹ also showed that the SO became more negative in the breathy phase in the first weeks following Botox injection in 3 out of 5 subjects, and became less negative as the adductors recovered. While these previous studies monitored the change of SO over weeks as the voice went through relatively large changes in the breathy-to-pressed dimension, the present work aimed to characterize relatively small SO variations within a single setting. The direction of change in SO was consistent between the current work and the prior works, namely that the SO was more negative with less strained voice, and less negative with more strained voice within the same subject. It is not straightforward to explain the direction of change in SO in terms of the difference in vocal fold vibratory kinematics between a strained voice and an unstrained voice, as there may be simultaneous changes in the level of vocal fold adduction, convergence, medial surface bulging, and vertical phasing²⁵ that can affect the SO in opposite directions. Nevertheless, a plausible hypothesis can be formulated as follows. Less strain implies less compression or greater pliability of the medial vocal fold mucosa. Also, if the vocal folds are less adducted, there is more room for elastic recoil in the transverse direction. Both factors, i.e., greater pliability and more room for the contacting surface to “retreat”, would favor faster de-contact, or a steeper opening slope. This hypothesis remains to be tested.

Potential Utility of EGG Opening Slope Quotient

Results from this work suggest possible clinical applications of the SO. First, it may have utility as an objective measure to track symptom severity in ADSD as a function of behavioral, pharmacologic, or surgical intervention, particularly in the evaluation of possible new therapies which may require objective documentation of treatment effectiveness in addition to subjective ratings from patients and clinicians. The work of Fisher et al.^{19,20} has already established this potential utility of the SO in sustained vowel production in speakers with ADSD. The current work extends it to continuous speech in ADSD. Second, the temporal fluctuation in SO during continuous speech (e.g., Figure 8), as opposed to the mean SO in speech segments, may assist in the diagnosis of mild or early ADSD. As the

listeners in this study reported, making a reasonable judgment of the degree of strain variability in a speech sample required multiple playbacks, and even then it was difficult to rate the degree of strain. The SO may provide an objective measure for a vocal quality attribute that is difficult to quantify perceptually. The uncertainty of diagnosis based on auditory perception alone may be reduced by inspecting a pseudo-continuous plot of SO (e.g., Figure 8) to see if the SO trend mirrors the perception of the stress-strain pattern in the speech selection. Future work measuring the diagnostic accuracy of auditory perception with or without the aid of the SO plot will be required to determine if the SO plot indeed has utility.

The EGG SO will likely have its greatest utility when used in combination with simultaneously collected acoustic and aerodynamic data in a multi-dimensional assessment of continuous speech. EGG captures an aspect of voice production complementary to those captured by acoustic and aerodynamic measures. A multi-dimensional or multi-parameter approach (e.g., Awan et al.⁴⁰; Wuyts et al.⁴¹; Dejonckere et al.⁴²) is likely to provide greater diagnostic value than individual measures. A future direction of research is to investigate the relative sensitivities of acoustic measures compared to the EGG SO in the differentiation of intra-subject strained vs. unstrained voice. There is some reason to believe EGG parameters may be advantageous in certain applications. For example, EGG perturbation parameters have been shown to have higher diagnostic accuracy than acoustic perturbation parameters for the detection of mild vocal roughness.⁴³ The eventual goal would be to derive a multi-parameter model that incorporates both EGG and acoustic information.

Advantages and Limitations of EGG

EGG readily lends itself to simultaneous acoustic and aerodynamic recording since the electrodes on the neck do not interfere with a face mask or a microphone. It is similarly non-invasive and the instrumentation relatively inexpensive. It is in fact less obtrusive than a face mask and does not interfere with facial motion in articulatory gestures in continuous speech. EGG also excludes voiceless consonants that are present in the acoustic signal but are extraneous to investigations focused on the sound source. Finally, EGG recording is not affected by ambient noise. On the other hand, the utility of EGG is limited by several factors. A large amount of subcutaneous adipose tissue will lead to low signal-to-noise. A low signal-to-noise is also encountered in a severely strained voice where the vocal folds do not fully separate between glottal cycles. There is also concern that contacts between the ventricular folds in patients with severe vocal strain¹⁹ or mucus bridges⁴⁴ can produce signal artifacts. This work represents an attempt to determine if, despite these limitations, EGG waveform parameters have diagnostic value in the differentiation of strained from unstrained voice.

Limitations of the Study

Several limitations to this study should be noted. First, the sample size of the study population in which the significant within-subject difference of SO was found was relatively small. Second, only ADSD subjects with sufficiently high strain variability in continuous speech were included. This was out of the necessity to generate a consensus set of strained syllables. ADSD subjects whose speech samples contained lower levels of strain variability

and could not yield a consensus set of syllables were not included. As a result, the sensitivity of SO to symptom severity was not assessed. A larger sample covering a more representative spectrum of symptom severity would be required to determine the sensitivity of SO as an objective measure of strain variability. Finally, this preliminary study was deliberately limited to a single hyperfunctional voice disorder, for the reasons detailed in the Introduction. Whether the conclusions can be applied to other types of hyperfunctional voice disorders in which a strained voice is a dominant perceptual feature will require further investigation.

CONCLUSION

This work presents an effort to critically analyze the EGG signal of dysphonic speech and to determine if any EGG waveform parameters correlated with the auditory perceptual quality of strain. An automated algorithm tailored for continuous speech with poor EGG waveforms was developed, with the capability to extract previously defined as well as new EGG parameters. While none of the parameters examined reliably differentiated normal speakers from ADSD speakers, the EGG opening slope quotient appeared to be sensitive to within-subject strain. This work is a step towards addressing the need for further knowledge regarding the role of EGG in voice assessment, an area of deficiency identified by an American Speech-Language-Hearing Association Ad Hoc Working Group in a systematic review of evidence-based clinical voice assessment.⁴⁵ EGG may be best used in a multi-dimensional voice assessment to complement acoustic and aerodynamic measures.

Acknowledgments

We thank 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. We also thank Professor Ron Scherer for critical review of and helpful comments on the manuscript. This work was supported by the NIH Loan Repayment Program (NIDCD grant L30 DC011942), the UT Southwestern Clinical Research Scholar program, and the UT Southwestern Summer Medical Student Research Program.

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. [PubMed: 18930908]
2. Hirano, M. Clinical examination of the voice. New York: Springer-Verlag; 1981. p. 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. [PubMed: 8450660]
4. Netsell R, Lotz W, Shaughnessy AL. Laryngeal aerodynamics associated with selected voice disorders. *Am J Otolaryngol*. 1984; 5:397–403. [PubMed: 6336354]
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. [PubMed: 17400426]
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. [PubMed: 16102670]
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. [PubMed: 22953661]

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. [PubMed: 22180020]
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. [PubMed: 1749236]
10. Rothenberg M, Mahshie JJ. Monitoring vocal fold abduction through vocal fold contact area. *J Speech Hear Res.* 1988; 31:338–351. [PubMed: 3172751]
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. [PubMed: 3782607]
12. Scherer, RC.; Druker, DG.; Titze, IR. Electroglottography and direct measurement of vocal fold contact area. In: Fujimura, O., editor. *Vocal Physiology: Voice Production, Mechanisms, and Function.* New York: Raven Press; 1988. p. 279-291.
13. Hampala V, Garcia M, Svec JG, Scherer R, Herbst CT. Relationship between the electroglottographic signal and vocal fold contact area. *J Voice.* 2015 Aug 6. pii: S0892–1997(15)00060-0. Epub ahead of print. doi: 10.1016/j.jvoice.2015.03.018
14. Scherer RC, Vail V. Measures of laryngeal adduction. *J Acoust Soc Am.* 1988; 84(suppl 1)
15. Scherer, RC.; Vail, VJ.; Rockwell, B. Examination of the laryngeal adduction measure EGGW. In: Bell-Berti, F.; Rapheal, L.J., editors. *Producing Speech: Contemporary Issues.* Woodbury, NY: American Institute of Physics; 1995. p. 269-290.
16. Askenfelt AG, Hammarberg B. Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. *J Speech Hear Res.* 1986; 29:50–64. [PubMed: 3702379]
17. 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. [PubMed: 8179248]
18. Verdolini K, Druker DG, Palmer PM, Samawi H. Laryngeal adduction in resonant voice. *J Voice.* 1998; 12:315–327. [PubMed: 9763181]
19. Fisher KV, Scherer RC, Swank PR, Giddens C, Patten D. Electroglottographic tracking of phonatory response to Botox. *J Voice.* 1999; 13:203–218. [PubMed: 10442750]
20. 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. [PubMed: 8898251]
21. Zagolski O, Carlson E. Electroglottographic measurements of glottal function in vocal fold paralysis in women. *Clin Otolaryngol Allied Sci.* 2002; 27:246–253. [PubMed: 12169125]
22. Vertigan AE, Theodoros DG, Winkworth AL, Gibson PG. A comparison of two approaches to the treatment of chronic cough: perceptual, acoustic, and electroglottographic outcomes. *J Voice.* 2008; 22:581–589. [PubMed: 17485195]
23. 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. [PubMed: 23177751]
24. 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. [PubMed: 10025549]
25. Choi SH, Lee J, Sprecher AJ, Jiang JJ. The Effect of Segment Selection on Acoustic Analysis. *J Voice.* 2011; 26:1–7. [PubMed: 21889300]
26. Ludlow CL, Adler CH, Berke GS, et al. Research priorities in spasmodic dysphonia. *Otolaryngol Head Neck Surg.* 2008; 139:495–505. [PubMed: 18922334]
27. Titze IR. Interpretation of the Electroglottographic Signal. *Journal of Voice.* 1990; 4:1–9.
28. Childers DG, Lee CK. Vocal quality factors: analysis, synthesis, and perception. *J Acoust Soc Am.* 1991; 90:2394–2410. [PubMed: 1837797]
29. Childers DG, Hicks DM, Moore GP, Eskenazi L, Lalwani AL. Electroglottography and vocal fold physiology. *J Speech Hear Res.* 1990; 33:245–254. [PubMed: 2359265]
30. Henrich N, d'Alessandro C, Doval B, Castellengo M. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *J Acoust Soc Am.* 2004; 115:1321–1332. [PubMed: 15058354]

31. Luo JW, Bai J, He P, Ying K. Axial strain calculation using a low-pass digital differentiator in ultrasound elastography. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*. 2004; 51:1119–1127.
32. Usui S, Amidror I. Digital Low-Pass Differentiation for Biological Signal-Processing. *IEEE Transactions on Biomedical Engineering*. 1982; 29:686–693. [PubMed: 6897393]
33. 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;
34. 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. [PubMed: 12381042]
35. Sapienza CM, Crary MA, Gorham M. Laryngeal aerodynamic aspects of women with adductor spasmodic dysphonia. *Arch Otolaryngol Head Neck Surg*. 1996; 122:385–388. [PubMed: 8600922]
36. Behrman A. Common practices of voice therapists in the evaluation of patients. *J Voice*. 2005; 19:454–469. [PubMed: 16102671]
37. Higgins MB, Chait DH, Schulte L. Phonatory air flow characteristics of adductor spasmodic dysphonia and muscle tension dysphonia. *J Speech Lang Hear Res*. 1999; 42:101–111. [PubMed: 10025547]
38. Orlikoff RF. Scrambled EGG: The uses and abuses of electroglottography. *Phonoscope*. 1998; 1:37–53.
39. Hosokawa K, Yoshida M, Yoshii T, et al. Effectiveness of the computed analysis of electroglottographic signals in muscle tension dysphonia. *Folia Phoniatr Logop*. 2012; 64:145–150. [PubMed: 22965009]
40. Awan SN, Roy N, Dromey C. Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model. *Clin Linguist Phon*. 2009; 23:825–841. [PubMed: 19891523]
41. Wuyts FL, De Bodt MS, Molenberghs G, et al. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res*. 2000; 43:796–809. [PubMed: 10877446]
42. 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 Otorhinolaryngol*. 2012; 269:1195–1203. [PubMed: 22210475]
43. Hosokawa K, Ogawa M, Hashimoto M, Inohara H. Statistical analysis of the reliability of acoustic and electroglottographic perturbation parameters for the detection of vocal roughness. *J Voice*. 2014; 28:263e269–263 e216. [PubMed: 24216270]
44. Colton RH, Conture EG. Problems and pitfalls of electroglottography. *J Voice*. 1990; 4:10–24.
45. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol*. 2013; 22:212–226. [PubMed: 23184134]

Figure 1.

Effect of Butterworth high-pass filter cutoff frequency on parameter values. Shown are the percentages of parameters that stayed within 1% of their corresponding values when no filter was applied. Based on pooled data for 6 parameters (SC5075, SC2575, SO9050, SO7525, EGGW50, and CQ) in 6 ADSD samples. For example, when a 10 Hz cutoff was used, 89% of the 36 parameters are within 1% of their values calculated from EGG without the filter applied.

Figure 2.

Schematic of baseline linear correction. For a positive slope (top), an amount equal to the magnitude of the down arrow is subtracted from the y value according to the equation $y = S * (T/T)$, where S is the absolute value of the difference between the starting and ending y values, T is the time since the beginning of the cycle, and T is the cycle width (period). For a negative slope (bottom), $y = S * (1 - (T/T))$.

Figure 3.

EGG waveform parameter definitions. Small square boxes denote wave maxima, and small circles denote minima.

Figure 4.
Schematic of algorithm.

Figure 5.

Representative EGG waveform of a short phrase in normal speech (top) and in ADSD speech (bottom). Open circles and squares mark the maxima and minima detected by the algorithm, respectively. The Xs above the cycles mark the cycles rejected by the algorithm.

Figure 6.

Comparison of coefficient of variation (CV) for slope measures of EGG processed with (A) 10 Hz Butterworth high-pass filter or (B) cycle-based linear baseline correction. Slope measures for the phrase “long round arch” were computed for 12 normals and 12 subjects with ASD. The mean CV in each group is plotted for each slope measure. The CV is defined as the ratio of the standard deviation to the mean.

Figure 7.

Heatmap showing correlation between EGG parameters. Each cell shows the Spearman's rank correlation coefficient between two parameters (along top row) for a subject (along left column). Only correlations with a p value < 0.01 are shown. Correlations with $p > 0.01$ are left blank. The color of the cell reflects the value of the correlation coefficient, with dark green corresponding to 1, and dark red corresponding to -1 , as shown in the color scale bar to the right.

Figure 8.

EGG parameter variation in continuous speech. A color gradient plot of several EGG parameters is shown for an ADSD speech segment. The color gradient reflects the normalized value of the parameters within this segment. Perceptually, “long” and “round” were the strained syllables in this sample.

EGG waveform parameters in normal vs. ADSD speech. Values are expressed as mean (standard deviation) (N=12). The standard deviation shown is the standard deviation of the 12 values of that parameter, one from each subject.

Table 1

	High-Pass Filter			Linear Correction		
	Normal	ADSD	<i>p</i>	Normal	ADSD	<i>p</i>
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
S07525	-2.25 (0.36)	-2.15 (0.64)	0.65	-2.24 (0.36)	-2.20 (0.66)	0.86

Table 2

Standard deviation (SD) of EGG waveform parameters in normal vs. ADSD speech. Since each parameter was averaged over all cycles within the selected speech segment for each subject, a SD was obtained for each subject. The mean SD of the 12 subjects is shown.

	High-Pass Filter			Linear Correction		
	Normal	ADSD	<i>p</i>	Normal	ADSD	<i>p</i>
CQ	0.070	0.111	0.0024	0.070	0.110	0.0032
EGGW50	0.044	0.078	0.0009	0.045	0.078	0.0010
Peak Skew	0.058	0.090	0.025	0.061	0.104	0.0061
SC5075	2.243	2.875	0.13	2.316	2.873	0.21
S07525	0.274	0.494	0.0074	0.295	0.545	0.0225

Table 3

Consensus strained syllables.

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

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

Table 4

	High-Pass Filter			Linear Correction		
	Strained	Unstrained	<i>p</i>	Strained	Unstrained	<i>p</i>
CQ	0.513 (0.076)	0.482 (0.073)	0.36	0.522 (0.074)	0.500 (0.106)	0.59
EGGW50	0.501 (0.065)	0.459 (0.074)	0.11	0.510 (0.067)	0.467 (0.076)	0.11
Peak Skew	0.303 (0.039)	0.327 (0.087)	0.46	0.328 (0.054)	0.321 (0.088)	0.84
SC5075	10.3 (3.7)	12.7 (4.8)	0.11	10.1 (3.7)	12.3 (4.7)	0.12
S07525	-2.13 (0.72)	-2.53 (0.69)	0.015	-2.15 (0.70)	-2.51 (0.70)	0.032