HHS Public Access

Author manuscript

Speech Commun. Author manuscript; available in PMC 2021 October 01.

Published in final edited form as:

Speech Commun. 2020 October ; 123: 98–108. doi:10.1016/j.specom.2020.07.003.

Analysis of Glottal Inverse Filtering in the Presence of Source-Filter Interaction

Anil Palaparthi[1,2], **Ingo R. Titze**[1,2]

[1]National Center for Voice and Speech, The University of Utah, Salt Lake City, UT 84112, USA

[2]Department of Biomedical Engineering, The University of Utah, Salt Lake City, UT 84112, USA

Abstract

The validity of glottal inverse filtering (GIF) to obtain a glottal flow waveform from radiated pressure signal in the presence and absence of source-filter interaction was studied systematically. A driven vocal fold surface model of vocal fold vibration was used to generate source signals. A one-dimensional wave reflection algorithm was used to solve for acoustic pressures in the vocal tract. Several test signals were generated with and without source-filter interaction at various fundamental frequencies and vowels. Linear Predictive Coding (LPC), Quasi Closed Phase (QCP), and Quadratic Programming (QPR) based algorithms, along with supraglottal impulse response, were used to inverse filter the radiated pressure signals to obtain the glottal flow pulses. The accuracy of each algorithm was tested for its recovery of maximum flow declination rate (MFDR), peak glottal flow, open phase ripple factor, closed phase ripple factor, and mean squared error. The algorithms were also tested for their absolute relative errors of the Normalized Amplitude Quotient, the Quasi-Open Quotient, and the Harmonic Richness Factor. The results indicated that the mean squared error decreased with increase in source-filter interaction level suggesting that the inverse filtering algorithms perform better in the presence of source-filter interaction. All glottal inverse filtering algorithms predicted the open phase ripple factor better than the closed phase ripple factor of a glottal flow waveform, irrespective of the source-filter interaction level. Major prediction errors occurred in the estimation of the closed phase ripple factor, MFDR, peak glottal flow, normalized amplitude quotient, and Quasi-Open Quotient. Feedback-related nonlinearity (source-filter interaction) affected the recovered signal primarily when f_o was well below the first formant frequency of a vowel. The prediction error increased when f_o was close to the first formant frequency due to the difficulty of estimating the precise value of resonance frequencies, which was exacerbated by nonlinear kinetic losses in the vocal tract.

Corresponding Author: Anil Palaparthi, anil.palaparthi@utah.edu, 1901 S Campus Dr, Suite 2120, Salt Lake City, UT 84112, USA.

AUTHOR STATEMENT

Anil Palaparthi: Conceptualization, Methodology, Software, Formal analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing **Ingo R. Titze:** Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Funding acquisition.

CONFLICTS OF INTEREST: NONE

Declarations of Interest: None.

Keywords

Glottal Inverse Filtering; Linear Predictive Coding; Source-Filter Interaction; Speech Synthesis

I. INTRODUCTION

Voice production has traditionally been subdivided into three major components: the energy source (lungs), the sound source (vocal folds within the larynx), and the vocal tract resonator, also known as a filter (Flanagan, 1972, Zhang, 2016). In the classical linear source-filter theory, these components are considered independent in that downstream components do not affect upstream energy generation or signal composition. The vocal tract is assumed to be a linear time-invariant filter that filters the glottal flow waveform and then radiates the filtered acoustic signal into the surrounding air. Glottal inverse filtering (GIF) is described as a technique for obtaining an estimate of the glottal flow waveform during voiced speech from either the radiated acoustic pressure waveform or the acoustic flow at the mouth as input (Rothenberg, 1973). If the radiated pressure is used as the input signal, both the effect of vocal tract and lip radiation need to be cancelled from the input signal to obtain the glottal flow waveform (Alku, 2011). If the oral flow is used as the input, only the filtering effect of the vocal tract needs to be cancelled. In this paper, we address the more difficult challenge of inverse filtering the radiated pressure.

Given that usually less than 1% of the oral acoustic power is radiated from the mouth (Schutte, 1980, Titze and Palaparthi, 2018), there is much backward reflection of acoustic power that can influence the glottal airflow. Hence, the source is not independent of the filter, mainly due to energy feedback. In addition to the feedback nonlinearity, the vocal tract itself has nonlinear properties due to the presence of several energy losses, especially the nonlinear junction kinetic losses (Titze et al., 2014). It is also noted that the lip radiation can be nonlinear and its modeling as a parallel RL circuit is only an approximation (Flanagan, 1972; Titze and Palaparthi, 2018). As a result, the standard linear filtering and inverse filtering approaches can only result in an approximate to the glottal volume velocity waveform (Rothenberg and Zahorian, 1977).

Figure 1 shows a typical glottal flow (volume velocity) waveform. The waveform can be divided into three phases: the opening phase, the closing phase and the closed phase. Generally, the glottal flow waveform is skewed to the right (its peak is delayed relative to the glottal area) due to the acoustic inertance of the vocal tract (Rothenberg,1981). This phenomenon is part of what has been called Level 1 source-filter interaction (Titze, 2008; Titze and Worley, 2009; Maxfield et al., 2017). Level 2 interaction, not included in this study, is the disturbance of vocal fold tissue movement by the acoustic feedback. With Level 1 interaction, the duration of the rising portion of the flow is longer than the duration of the falling portion. Glottal flow can also have a DC offset, which arises when the glottis does not close completely. Often there is a ripple on the glottal flow waveform that reflects the standing waves in the vocal tract with their characteristic resonance frequencies.

Researchers have developed glottal inverse filtering techniques for over 50 years to estimate the vocal source waveform. For a thorough review of GIF techniques, the reader is referred

to Alku (2011), Walker and Murphy (2007), or Drugman et al. (2014). Starting with the early work of Miller (1959), who used analog elements to inverse filter the first resonance (formant) to the recent work of Airaksinen et al. (2017), who used quadratic programming to accurately estimate the closed phase of the glottal waveform, these algorithms have been able to estimate the glottal flow waveform with varied success. Some of these methods include linear prediction analysis (Markel and Gray, 1980), complex cepstral decomposition (CCD) (Drugman et al., 2011), zeros of the z-transform (ZZT) (Bozkurt et al., 2005), iterative adaptive inverse filtering (IAIF) (Alku, 1992), and Quasi Closed Phase (QCP) (Airaksinen et al., 2014) method.

More complex algorithms based on joint estimation of source and filter were also proposed over the years for better estimation of the glottal waveform. However, these algorithms also consider linear source-filter theory for the development of the techniques. The *joint estimation algorithm* was published several decades ago by Milenkovic (1986). Fröhlich et al. (2001) developed a “simultaneous inverse filtering and model matching” method based on a discrete all-pole modeling technique for inverse filtering. For synthesized signals, the accuracy obtained was higher compared to the conventional methods, but problems with robustness still persisted when tested on natural utterances. Fu and Murphy (2006) also used joint estimation of vocal source and filter by modeling the source using a linear filter (LF) model and the filter as a time-varying ARX speech production model. Vocal tract parameters were identified using the Kalman filtering process. Their approach yielded robust results when tested with synthetic as well as natural signals. However, they did not compare their method with other models to assess relative performance. Airaksinen et al. (2017) combined the vocal tract and lip radiation into a single filter and used quadratic programming to optimize the coefficients of such a filter. They achieved flatter closed phases with less formant ripple through this approach when tested on real speech signals compared to QCP, IAIF, and CCD methods. Alzamendi and Schlotthauer (2017) modeled the vocal source as a stochastic glottal model and the vocal tract as a time-varying autoregressive filter with exogenous input. State-space methods were then used to jointly estimate the glottal source and vocal tract filter (Sahoo and Routray, 2016). Here again, the glottal waveform signals obtained were smooth, with less formant ripple compared to IAIF and LPC methods when tested with synthetic and physical model-based signals. There are very few methods that treat glottal inverse filtering as a nonlinear filtering problem. Rothenberg and Zahorian (1977) used a linear time-varying model to represent the vocal tract and a nonlinear time-varying inverse filter with feedback to obtain the glottal source signal. It was found that the glottal waveform obtained using a nonlinear filter is more symmetric compared to the one obtained from linear filtering. Another study, Berezina et al., (2010) addressed the quasiperiodic nature of the glottal flow. To the best of our knowledge, no glottal inverse filtering method explicitly compensated for source-filter interaction nonlinearity.

At this stage, it is important to test the efficacy of the current glottal inverse filtering algorithms using voice simulators or synthesizers that incorporate the nonlinearities present in voice production, especially source-filter interaction. As far as we know, there are no studies that systematically assessed the role of source-filter interaction on glottal inverse filtering. Some studies used physical models that included source-filter interaction to test their algorithms (Airaksinen et al., 2014; Chien et al., (2017); Mokhtari et al., 2018; Alku et

al., (2019)), but did not assess the role of source-filter interaction on the performance of their glottal inverse filtering algorithms. Testing on natural signals alone cannot reveal the efficacy of the algorithms, as the true glottal flow waveform is not available for comparison. Hence, this study was designed to perform a thorough analysis of the efficacy of four linear glottal inverse filtering methods by using test signals generated by a mathematical model that included vocal tract nonlinearities and source-filter interaction.

The manuscript is organized as follows. Section II provides the details of the model, generation of test signals, details of the four glottal inverse filtering algorithms, and introduce the performance analysis metrics. Section III presents the results of the four glottal inverse filtering algorithms in the presence and absence of source-filter interaction. In section IV and V, the results and conclusions of this study are discussed.

II. Methods

A. Source-Filter Interactive Model

The model used for Level 1 source-filter interaction assumes a static (postural) configuration around which vibration takes place on the medial surface (Fig. 2, right vocal fold). Parameters for posturing are the glottal half-width at entry ξ_{01}, the glottal half-width at exit ξ_{02} (both at the arytenoid cartilage), the medial surface bulging ξ_b, the thickness of the vocal folds T, the depth of the vocal fold D, and the length of the vocal fold L. The glottal half-width $\xi_0(y, z)$ over the entire vocal fold surface is expressed as:

$$\xi_0(y,z) = (1 - y/L)[\xi_{02} + (\xi_{01} - \xi_{02} - 4\xi_b z/T)(1 - z/T)] \quad (1)$$

where y is the anterior-posterior dimension and z is the inferior-superior dimension. This equation has been used in multiple previous publications to describe the medial surface of the vocal folds (Titze, 1989; Titze, 2006, Eq. 4.194)

To include a vibrational displacement, a mucosal surface-wave approach was used (Titze, 1988). It accounts for differential movement of the upper and lower margins of the vocal folds on the medial surface. The mucosal surface-wave motion is defined by a wave velocity c and an inflection point z_{m} that changes the direction of motion by 180 degree for upper versus lower displacement. The vibrational displacement ξ at any vertical point z and any horizontal point y on the vocal fold surface is given by:

$$\xi(y,z,t) = \xi_m \sin(\pi y/L)[\sin\omega t - (\omega/c)(z - z_m)\cos\omega t] \quad (2)$$

where ξ_m is the vibrational amplitude, which has been empirically obtained as a function of lung pressure (Titze et al., 2003) and fundamental frequency f_o. The inflection point was also further defined empirically as $z_m = 0.6 - 2\xi_b T$.

Glottal airflow calculation included the effects of acoustic vocal tract pressures (subglottal and supraglottal) on the glottal flow. Fig. 3 shows a simplified coronal sketch of the vocal folds with incident and reflected partial pressures. An equation was previously derived analytically (Titze, 1984). If the flow-detachment area in the glottis is a_d, the glottal flow is

U_g, and the transglottal pressure loss coefficient is k_t, then continuity of pressure and flow at both inlet and outlet of the glottis results in the following equations

$$p_s = p_s^- + p_s^+ \tag{3}$$

$$p_e = p_e^- + p_e^+ \tag{4}$$

Assuming full reflection above and below the glottis,

$$p_s^- = -\frac{\rho c}{A_s} U_g + p_s^+ \tag{5}$$

$$p_e^+ = \frac{\rho c}{A_e} U_g + p_e^- \tag{6}$$

The transglottal pressure is

$$p_{tg} = p_s - p_e = \frac{1}{2} k_t \rho \, U_g^2 / a_d^2 \tag{7}$$

When Eqs. (3) and (4) are substituted into Eq. (7) and solved algebraically for the glottal flow U_g, we obtain

$$U_g = \left(\frac{a_d c}{k_t}\right)\left\{-\left(\frac{a_d}{A*}\right) \pm \left[\left(\frac{a_d}{A*}\right)^2 + \frac{4k_t}{\rho c^2}[p_s^+ - p_e^-]\right]^{1/2}\right\} \tag{8}$$

Here, ρ is the air density and c is the speed of sound in air. Further, if A_s is the subglottal area, and A_e is the supraglottal area, then A* in Eq. (8) is the effective interaction area for both subglottal and supraglottal airways, defined as:

$$A* = \frac{A_s A_e}{(A_s + A_e)} \tag{9}$$

and k_t as (Titze, 2006):

$$k_t = 1.37 - min[1, 2(a_d/A_e)(1 - a_d/A_e)], \tag{10}$$

Note that k_t is time-varying because the flow-detachment area a_d is time-varying. An impulse response of the vocal tract usually assumes constant boundary conditions.

In a wave-reflection algorithm (Liljencrants, 1985; Story, 1995), Eq. (8) uses both the forward travelling subglottal partial pressure (p_s^+) and the backward travelling supraglottal partial pressure (p_e^-) to produce the source-filter interaction (SFI). In this study, the vocal tract was modeled as a series of cylindrical sections connected to each other with boundary

conditions. Each cylindrical section included viscous and wall losses. The junction boundary conditions between adjacent sections contained the nonlinear kinetic losses (Titze et al. (2014)). The radiation losses were modeled as a parallel RL circuit. For a detailed description of vocal tract modeling, the readers are referred to (Story, (1995)).

B. Testing Paradigm

A brute force approach was used to generate test signals using the driven source-filter interactive model detailed above. Equation (8) was used to generate glottal flow waveform signals under full SFI test condition. The supraglottal-tract-only test cases were generated by replacing the subglottal time-varying partial pressure p_s^+ with a steady partial lung pressure ½ P_L in the Eq. (8). The test cases with no source-filter interaction were generated by setting $p_e^- = 0$ in the equation, in addition to the above-mentioned change. The prephonatory superior glottal half-width (ξ_{02}), prephonatory inferior glottal half-width (ξ_{01}), and the prephonatory medial surface bulging (ξ_b) were chosen as glottal parameters of interest. In addition, the cross-sectional area of epilarynx tube, a uniform vocal tract shape, and 11 vowels /ɛ/, /æ/, /ɑ/, /ɔ/, /ʌ/, /o/, /I/, /e/, /ʊ/, /i/, and /u/ were considered for vocal tract variations. The area functions for each vowel were taken from Story et al., (1996) which were further tuned so that the first and second formants of each area function are equal to the averages reported in Hillenbrand et al., (1995). The uniform tube vocal tract (symbol *ut*) was chosen to have an area of 3 cm^2. The area functions for the eleven vowels were given in Appendix. The prephonatory superior glottal half-width, ξ_{02} was varied from 0 to 0.1 cm in increments of 0.05 cm, the prephonatory inferior glottal half-width ξ_{01} was varied from 0 to (ξ_{02} + 0.2) cm in increments of 0.025 cm, and the medial surface bulging, ξ_b was varied from 0 to [(ξ_{01} + ξ_{02})/2] in increments of [(ξ_{01} + ξ_{02})/6] cm. The epilarynx tube was divided into three regions, the ventricle (one section), the false fold glottis (one section), and the vestibule (4 sections), as described in Titze and Palaparthi (2016). The area of each of the subdivisions was varied by doubling or dividing by 2 from the nominal values. The nominal values were chosen as 0.8, 0.4, and 0.5 cm^2 respectively. These variations resulted in 9408 test signals. Each signal was 0.2 s in duration.

These signals were generated under two f_o conditions: (1) a male speech f_o of 130 Hz for all test signals (referred to as fixed-f_o hereafter); (2) near-formant f_o for each vowel such that each f_o was 50 Hz below the first formant of that vowel. The f_o values for each vowel for near-formant f_o condition are given in Table I.

The near-formant f_o conditions were used to maximize the source-filter interaction (Maxfield et al., 2017). For each f_o condition, SFI was varied in 3 different ways: (1) with both subglottal and supraglottal tract interaction (fullSFI), (2) with supraglottal tract only interaction (supSFI), and (3) with no source-filter interaction (noSFI). These combinations resulted in 2 f_o × 3 SFI × 9408 = 56448 test signals. These test signals were used to assess the performance of the selected glottal inverse filtering algorithms.

C. Glottal Inverse Filtering Algorithms

Three glottal inverse filtering algorithms were chosen, including Linear Predictive Coding (LPC), the Quasi-Closed phase (QCP) method, and the Quadratic Programming (QPR)

approach. Along with these methods, the true supraglottal impulse response was also used for glottal inverse filtering as a benchmark. These methods are detailed in the following subsections to emphasize their assumptions.

1) True Supraglottal Impulse (TI) Response—The supraglottal impulse response, $h(n)$ for each test case was obtained by exciting the supraglottal vocal tract with a unit impulse and measuring the radiated pressure signal, p_o. The trachea and the glottal configuration were not considered while generating the impulse response. The sampling rate was set to 44100 Hz. The glottal flow, U_g was then obtained by convolving the inverse of the supraglottal impulse response with the corresponding radiated pressure p_o of the test signal.

$$U_g(n) = h^{-1}(n) * p_o(n) \tag{11}$$

2) Linear Prediction (LPC) based method—In the Linear Prediction based approach, the supraglottal vocal tract is modeled as an all-pole filter, where the current sample of the oral flow, $u_o(n)$, is modeled as a linear combination of past samples.

$$u_o(n) = \sum_{k=1}^{p} a_k u_o(n-k) \tag{12}$$

In the current study, the true supraglottal vocal tract resonance frequencies and their bandwidths were computed from the vocal tract impulse response for each test signal (Titze et al., 2014). The vocal tract filter coefficients, a_k were constructed using the resonance frequencies F_i and bandwidths B_i of the supraglottal vocal tract as reported by Markel and Gray (1980). The vocal tract poles, p_i, $i = 1, \ldots, N$ were computed using the equation

$$p_i = R_i e^{j\theta_i}, \tag{13}$$

Where the pole radii R_i were given by

$$R_i = e^{-\pi B_i / F_s}, \tag{14}$$

and pole angles θ_i were given by

$$\theta_i = 2\pi\ F_i / F_s\,. \tag{15}$$

Here, F_s is the sampling rate. The vocal tract filter coefficients a_k were then obtained by computing the coefficients of the polynomial whose roots were the above complex poles and their respective complex conjugates. The obtained poles were then inverted to obtain the zeros of the vocal tract inverse filter. Lip radiation was modeled as a parallel RL circuit (Flanagan, 1972, pp. 36, Titze et al., 2014). In this circuit, the radiation resistance $R = 128\ \rho c / 9\pi^2 A_m\ Pa\ s/m^3$, and the radiation inertance $L = 8\ \rho r / 3\pi A_m\ Pa\ s^2/m^3$, where r is the mouth radius and A_m is the mouth area.

The inverse lip radiation filter coefficients were obtained as

$$b_r = \left[\frac{L + R\Delta T}{RL} \quad \frac{1}{R}\right],\ a_r = [1\ -1], \tag{16}$$

where ΔT is the sampling period, which was set to 1/12000 to limit the number of formants to six.

3) Quasi Closed Phase (QCP) method—The QCP method was detailed in Airaksinen et al., (2014). The method uses weighted linear prediction (WLP) with attenuated main excitation (AME) weight function. The AME function attenuates the contribution of the glottal source in the linear prediction model optimization in the vicinity of glottal closure instants. The AME function uses three parameters, Position Quotient (PQ), Duration Quotient (DQ), and length of linear ramp (N_{ramp}) to perform glottal inverse filtering. The parameter values were set as suggested in Airaksinen et al., (2014). For test cases where f_o is 130 Hz, the parameters PQ, DQ, and N_{ramp} were set to 0.05, 0.95, and 7 respectively. For vowel dependent f_o test cases, the three parameters were set to 0.05, 0.7, and 7 respectively. The sampling rate was set to 12000 Hz.

4) Quadratic Programming (QPR) Approach—The QPR method was detailed in Airaksinen et al., (2017). The method jointly models the effect of vocal tract and lip radiation using a single filter whose coefficients are optimized using quadratic programming. The optimization is based on the principles of closed phase analysis, where the flatness of the closed phase is targeted as the output of the optimization problem. The optimization uses three coefficients γ_1, γ_2, *and* γ_3 corresponding to the three criterion that are minimized. In our study, the three coefficients were set to 40, 5000, and 500 respectively as suggested in Airaksinen et al., (2017). The sampling rate was set to 12000 Hz.

These four methods were chosen to cover the entire gamut of information available about vocal tract to perform inverse filtering i.e. from most of the information to no information. The TI method has most of the information about the vocal tract *a priori* in terms of impulse response, LPC method has partial information about the vocal tract *a priori* in terms of vocal tract resonances and bandwidths, and QCP and QPR algorithms have no *a priori* information about the vocal tract.

D. Performance Analysis

The performance of the glottal inverse filtering methods was evaluated using several glottal flow parameters. The parameters included the standard metrics that are based on glottal closure instants (GCI) such as the Normalized Amplitude Quotient (NAQ), Quasi-Open Quotient (QOQ), and Harmonic Richness Factor (HRF) (Airaksinen et al., 2014) along with the Mean Squared Error (MSE). Most of these parameters do not directly quantify the glottal flow waveform shape. Hence, along with these standard parameters, other useful glottal flow parameters that quantify the glottal flow shape directly were used in this study for performance evaluation purposes. They are the Glottal Open Phase Ripple Factor (RF_o), Glottal Closed Phase Ripple Factor (RF_c), Maximum Flow Declination Rate (MFDR), and Glottal Flow Amplitude (UG_m) (Titze and Palaparthi, 2016).

NAQ measures the relative length of the glottal closing phase, QOQ measures the approximate length of the glottal open quotient, and HRF is the ratio of sum of the intensities of the harmonics to the intensity of the fundamental. The error for these three standard parameters for each test signal was reported as the average of absolute relative error in percentage computed as

$$E_p = E\left[\frac{|p_d - p_m|}{p_d}\right] \times 100. \tag{17}$$

Here, p is the parameter of interest among NAQ, QOQ, and HRF; p_d is the desired parameter value from the known glottal flow waveform; and p_m is the measured parameter value from the predicted glottal flow waveform. E is the expectation symbol that measures the mean value. The relative error is computed cycle-to-cycle between the parameters obtained from the known and the estimated glottal flow waveforms. The COVAREP voice analysis repository version 1.4.2 was used to compute the GCI based parameters (Degottex et al., 2014).

The parameters RF_o, RF_c, MFDR, and UG_m were computed based on the entire signal. The error for these parameters for each test signal was quantified in percentage using the equation:

$$E_p = \frac{|p_d - p_m|}{p_d} \times 100. \tag{18}$$

The approximate closed phase (UG_c) of the glottal flow waveform is separated from the approximate open phase (UG_o) with a 20% threshold from the maximum glottal flow in a glottal cycle (see Fig. 1). The ripple factor for the open phase and the closed phase were then computed as follows

$$RF_o = \frac{rms(UG_o)}{mean(UG_o)},\ RF_c = \frac{rms(UG_c)}{mean(UG_c)}. \tag{19}$$

The term *rms* is the root mean square value. MFDR is the absolute maximum negative peak of the glottal flow waveform derivative as shown below

$$MFDR = \left|min\left\{\frac{dUG}{dt}\right\}\right|. \tag{20}$$

It quantifies the closing phase of the glottal flow waveform. The MSE is computed using the following equation

$$MSE = mean((UG_d - UG_e)^2)$$

Here, UG_d is the known glottal flow waveform and UG_e is the predicted glottal flow waveform.

III. Results

A. True Supraglottal Impulse Response

The known and predicted glottal flow waveforms for the six combinations of f_o-SFI test scenarios for vowel /i/ were shown in Fig. 4 as an example. The known impulse response was used here for inverse filtering. The top row was from the fixed-f_o test case, and the bottom row was from the near-formant f_o test case. The complexity in the known glottal flow waveform increased with increase in SFI more in the near-formant f_o case compared to the fixed-f_o case. It can be observed from Fig. 4 that the glottal flow waveform was predicted rather well with inverse filtering, especially the strong open-phase formant ripple (bottom right). The prediction in the glottal closing region is much better for test cases with low fixed-f_o compared to the test cases with near-formant f_o. Under both f_o scenarios, the prediction is progressively better from no SFI cases to full SFI cases except for the glottal flow amplitude in the fixed-f_o scenario. However, the estimation of the glottal flow waveform was not entirely accurate even with the availability of the true supraglottal impulse response. The closing phase, closed phase, and the glottal flow amplitude estimation were slightly inaccurate. This is because glottal inverse filtering is based on the principle of linear time-invariant system (Rothenberg and Zahorian, 1977). However, the supraglottal vocal tract is non-linear due to the presence of various losses, especially the nonlinear junction kinetic losses (Titze et al., 2014), and the time-varying nature of the source-filter interaction.

B. Glottal Inverse Filtering Algorithms

Figure 5 shows the known and predicted glottal flow waveforms for an /i/ vowel from the three glottal inverse filtering algorithms for the fixed-f_o test cases. For each algorithm, the estimated glottal flow waveform is similar among the three different SFI cases (along the rows in Fig. 5). The LPC and QCP algorithms had similar predictions of the glottal flow waveform. They both estimated the open phase better than the closed phase. The main error was obtained in the prediction of the closed phase, the later part of the closing phase, and glottal flow amplitude. On the other hand, the QPR algorithm predicted the closed phase better but resulted in larger error in the prediction of open phase and glottal flow amplitude compared to the other two algorithms.

Figure 6 shows the known and predicted glottal flow waveforms of an /i/ vowel from the three glottal inverse filtering algorithms for the near-formant f_o test scenario. It is evident that the prediction error has increased compared to the fixed-f_o scenario. The prediction error is small for the LPC algorithm as true resonance frequencies and bandwidths were known a priori. That is not the case for the QCP and QPR algorithms, given that they performed inverse filtering without any knowledge of the vocal tract. Similar to the results from the fixed-f_o scenario, for the LPC algorithm, the error is higher in the prediction of closed phase, closing phase, and amplitude. For the QCP algorithm, it is higher in the prediction of opening phase, closed phase, and amplitude, and for the QPR algorithm in the prediction of open phase and amplitude. The open phase ripple was better estimated by the LPC algorithm and the closed phase by the QPR algorithm. These results indicate that across

all the test signals the error might be relatively high in the measures of RF_c, UG_m, NAQ, QOQ, and MFDR.

C. Performance Analysis

The error for all the performance analysis metrics was computed for each of the 9408 test signals in each f_o-SFI test cases. An example boxplot for % Error in MFDR for full SFI and near-formant f_o condition is shown in Fig. 7. An adjusted boxplot for skewed data was used for this purpose as the data is non-Gaussian (Hubert and Vandervieren, 2008). The prediction error had several outliers. Hence, the median of errors across all 9408 test signals was considered instead of the mean for comparison purposes. A one-sided Mann-Whitney U test was used to identify if the median of error from the no SFI test case was significantly lower than the median of error from supraglottal SFI and full SFI test cases.

1) Entire Waveform based Parameters—The results for the five parameters RF_o, RF_c, MFDR, UG_m, and MSE which are measured on the entire waveform, are presented in Fig. 8. The top row (Fig. 8(a) to (e)) shows the results from the fixed-f_o test cases and the bottom row (Fig. 8(f) to (j)) shows the results from the near-formant f_o test cases for all the four methods. The error from no-SFI (noSFI), supraglottal SFI (supSFI), and full SFI (fullSFI) test cases are presented for all the vowels combined, as the error did not have a consistent trend for any vowel. The star (*) symbols on top of supSFI, and fullSFI bars indicate that their medians of error were significantly higher compared to that of the noSFI case according to the one-sided Mann-Whitney U test with 99% confidence level. The cases where such stars (*) were absent indicate that their medians were either lower or not significantly different.

The results in Fig. 8 suggest that all the algorithms estimated the open phase ripple factor (RF_o) better than the other three parameters (RF_c, MFDR, and UG_m). Across all four methods, the median of error in the prediction of RF_o was less than 4%, indicating excellent performance by all the algorithms. Compared to this, the median of error was significantly higher (10 times or more) in the prediction of RF_c, MFDR, and UG_m. Except in one instance for QCP method, the MSE decreased with increase in SFI level for both f_o scenarios and all the four methods. For the near-formant f_o test cases (Fig. 8(f) to (j)), the median of error was relatively higher than the fixed-f_o test cases (Fig. 8(a) to (e)). This finding suggests that the prediction error increases as f_o approaches F_1. This is probably due to the difficulty in measuring the resonances of the vocal tract when f_o is closer to F_1 (Alku, 2011) and the presence of nonlinearities in the vocal tract, but not due to source-filter interaction, as the error increased even for some noSFI cases.

The TI method, in which the true impulse response is known a priori, predicted the entire waveform-based parameters better than the other algorithms. The LPC algorithm, which estimated the vocal tract filter coefficients from known resonances and their bandwidths, had higher error compared to the TI method. This is because such an estimation of vocal tract coefficients results in only an approximation of the true vocal tract impulse response (Markel and Gray, 1980). The QCP algorithm performed rather well and the error in estimation was comparable to that of the LPC algorithm, even though the true characteristics

of the vocal tract were unknown a priori. Finally, the QPR algorithm had the highest error in the prediction of entire waveform-based parameters compared to the other three methods, except in the estimation of closed phase ripple factor, which the algorithm specifically targets during optimization. The error is similar to that of the LPC, and QCP algorithms.

Contrary to expectation, the prediction error did not consistently increase with increase in SFI level ('*' symbols in Fig. 8). There were many instances where the error decreased or remained similar. Not one particular method consistently resulted in higher prediction error with increase in SFI level across all five parameters and two f_o scenarios. It can also be observed that the fixed-f_o test cases had more instances with higher prediction error than the near-formant f_o test cases when SFI level was increased. There are 21 '*' symbols for fixed-f_o test cases compared to 10 for near-formant f_o test cases in Fig. 8.

2) Glottal Closure Instant Based Parameters—The results for Normalized Amplitude Quotient (NAQ), Quasi-open Quotient (QOQ), and Harmonic Richness Factor (HRF) are presented in Fig. 9. The top row (Fig. 9(a) to (c)) shows the results from fixed-f_o test cases and the bottom row (Fig. 9(d) to (f)) shows the results from near-formant f_o test cases for all four methods. The error in percentage for these parameters was measured cycle-to-cycle based on Glottal Closure Instants and then averaged for the entire waveform (Eq. 17). Here also, the error from no-SFI (noSFI), supraglottal only SFI (supSFI), and full SFI (fullSFI) test scenarios was presented for all the vowels combined. The star (*) symbols on top of supSFI, and fullSFI bars indicate that their medians of error were significantly higher compared to that of the noSFI case according to the one-sided Mann-Whitney U test with 99% confidence level.

The results in Fig. 9 suggested that all the algorithms estimated the Harmonic Richness Factor (HRF) better than the other two parameters (NAQ and QOQ). Across all the four methods, the median of error in the prediction of HRF was less than 8% which is comparable to that of RF_o. In contrast, the median of error for NAQ and QOQ is comparable to the error for RF_c, MFDR, and UG_m. For these parameters as well, the median of error is considerably higher for near-formant f_o cases compared to fixed-f_o test cases even for noSFI test cases. This agrees with the finding from whole waveform-based parameters, that difficulty in the estimation of vocal tract resonances and nonlinearities in the vocal tract might be a bigger factor than source-filter interaction at fundamental frequencies near first formant of a vowel.

Here also, the TI method resulted in the lowest prediction error across the three parameters followed by the LPC algorithm. The prediction errors from the QCP algorithm were slightly higher compared to that of the LPC algorithm, and the QPR algorithm had the highest error compared to the other methods even in the prediction of the cycle-based parameters.

The prediction error did not consistently increase with increase in SFI level ('*' symbols in Fig. 9) even for GCI based parameters. In the case of QOQ, the error significantly increased with increase in SFI level for both f_o scenarios and three methods. HRF followed next with the fixed-f_o test case having three methods and the near-formant f_o test case with one method where the prediction error increased with SFI level. The NAQ had no method with increase

in error when SFI level was increased across the two f_o scenarios. These results also show that fixed-f_o test cases had more instances with higher prediction error than near-formant f_o test cases when SFI level was increased. There are 11 '*' symbols for fixed-f_o test cases compared to 7 for near-formant f_o test cases in Fig. 9.

3) Vowel Based Prediction Error Dependence on Source-Filter Interaction— Table II summarizes the number of vowels that had significantly higher median of error statistically for fixed-f_o test cases when supSFI was compared with noSFI, and likewise when fullSFI was compared with noSFI. Table III presents similar results for near-formant f_o test cases. It can be observed that for all the parameters, the percentage of vowels with increase in error as SFI level increased was higher for fixed-f_o test cases than for near-formant f_o test cases. There were four out of eight parameters with percentages higher than 50% (RF_o with 62.5%, RF_c with 53.1%, QOQ with 53.1%, and HRF with 57.3%) in the case of fixed-f_o test cases. QCP algorithm is the major contributor to the percentages with both supSFI (74.0%) and fullSFI (66.7%) test cases resulting in percentages above 50%. The other methods contributed less than 50% to the number of vowels with a significant increase in median of error with higher SFI level. On the other hand, near-formant f_o test cases had only two out of eight with percentages above 50% (RF_c with 53.1%, and QOQ with 60.4%) as well as RF_o with 44.8%. All the other parameters had significantly lower percentages (<20%). In the case of near-formant f_o, MSE has no vowels that resulted in higher median of error with increase in SFI level. The QPR and LPC algorithms contributed the most to the percentages for the near-formant f_o test scenario (> 37%). The other two methods contributed in significantly lower percentages. Across both f_o scenarios, RF_c and QOQ are the only parameters with percentages higher than 50%, whereas MFDR, UG_m, MSE, and NAQ have percentages less than 40%.

IV. Discussion

In the present study, three algorithms were chosen to analyze the accuracy of glottal inverse filtering in the presence of source-filter interaction. For these algorithms, little to no *a priori* vocal tract information was provided. In a fourth procedure, a known impulse response of the vocal tract was utilized. Test signals with different levels of SFI were generated using *a flow-interactive model* by systematically varying the geometry of the glottis, the epilarynx tube above the glottis, and the vowels. The results indicated that all four methods were able to better estimate the open phase ripple factor than the closed phase ripple factor. An interesting finding from the current study is that inverse filtering is not entirely accurate even with the availability of the exact supraglottal impulse response. This is because inverse filtering requires that the vocal tract filter to be a linear system. However, due to the presence of constrictions and junction losses, the vocal tract does not have a linear transfer function. The amount of nonlinearity depends on the vowel produced and the varying glottal shape. Another reason for inaccurate estimation with a known impulse response is the time-varying nature of the source-filter interaction. Observation of the individual signals from the True Impulse (TI) method showed that the major prediction error occurred in the estimation of closing phase, closed phase, and glottal flow amplitude. These errors observed with the TI method were also evident for the three inverse filtering algorithms. As expected, the errors

were higher because less *a priori* information about the vocal tract was available to those algorithms. This error could also have been higher because of the generation of some unrealistic epilarynx tube variations with the brute force approach used in the study. One other possible cause of error might be the change in fundamental frequency between fixed-f_o and near-formant f_o while keeping the vowel area functions constant (Tom et al., 2001). These variations were generated following the method used in Titze and Palaparthi, (2016) to maximize source-filter interaction. Even though these variations provide useful information academically, more realistic variations will be used in future studies.

Several metrics were used in the current study to quantify the errors in the prediction of the open phase, closing phase, closed phase, and amplitude for all the test cases. These metrics included the standard ones used by researchers to evaluate glottal inverse filtering algorithms such as Normalized Amplitude Quotient (NAQ), Quasi-Open Quotient (QOQ), and Harmonic Richness Factor (HRF). These parameters are cycle-to-cycle based and rely on accurate measurement of glottal closure instants. Along with these standard metrics, other useful parameters that quantify error on the entire signal were also used, such as Open Phase Ripple Factor (RF_o), Closed Phase Ripple Factor (RF_c), Maximum Flow Declination Rate (MFDR), Glottal Flow Amplitude (UG_m), and Mean Squared Error (MSE). For measures computed on the entire signal, the error was minimal for all metrics when using the TI method. The LPC and QCP algorithms had similar performance and were better in the estimation of open phase parameters RF_o, MFDR, and UG_m compared to the QPR algorithm. The three algorithms performed similarly in the estimation of closed-phase ripple factor RF_c. The MSE consistently decreased with increase in SFI level suggesting that inverse filtering algorithm perform better in the presence of SFI. These results confirm the findings from the individual waveforms in Figs. 4 to 6. The results from cycle-to-cycle based methods also followed the whole waveform-based methods. The prediction error progressively increased from TI method to QPR method indicating that the error is directly proportional to the amount of vocal tract information available for the inverse filtering algorithms.

The error in predicting glottal flow showed no particular pattern across vowels. There was no vowel that resulted in consistently lower or higher error for all the parameters. There was, however, an error dependence on fundamental frequency. This result matches with the findings from earlier studies (Alku, 2011). For fixed-f_o test cases, SFI contributed significantly to prediction error. The prediction error for most of the parameters increased significantly for near-formant f_o test cases compared to fixed-f_o test cases, even when there was no source-filter interaction, suggesting that factors other than SFI led to the higher prediction error. For fundamental frequencies closer to the first formant, it appears that the difficulty with identifying vocal tract resonances and nonlinearities in the vocal tract might be the major source of error compared to SFI.

The results from this study compare well with the results from (Airaksinen et al., 2014; Airaksinen et al., 2017). For fixed-f_o test cases and the QCP algorithm, the error was less than 20% for NAQ, 10% for QOQ, and 3% for HRF. The average errors for NAQ and QOQ shown for physical model test in Fig. 5(b) of Airaksinen et al., (2014) match our median errors listed above. The average errors obtained for the QCP and QPR algorithms from

synthetic signals reported in Airaksinen et al., (2017) were very different from the numbers reported in Airaksinen et al., (2014), indicating that the test parameters might be different in those two studies. The average errors obtained for the QPR algorithm were higher than QCP in Airaksinen et al., (2017), comparing well with our findings. Overall, the error is less than 5% for RF_o, and less than 100% for RF_c, MFDR, and UG_m indicating that the glottal inverse filtering algorithms perform reasonably in the prediction of glottal flow even with no a priori knowledge of the vocal tract shape.

V. CONCLUSION

Source-filter interaction significantly influenced the glottal volume flow pulse obtained by inverse filtering the oral radiated pressure, as evident by the one-sided Mann-Whitney U test results. However, the error between the original volume flow and the inverse-filtered flow did not increase uniformly with increased source-filter interaction for all the parameters. In fact with an exception of one instance for QCP algorithm, the MSE decreased with increase in SFI level for all the four methods and two f_o scenarios. The prediction error is higher when f_o is closer to the first formant frequency of the vocal tract compared to when f_o is well below the first formant frequency. Our results indicated that glottal inverse filtering is less reliable due to source-filter interaction when f_o is below the first formant frequency of a vowel. In addition, when f_o was close to the first formant frequency, the difficulty in estimating the precise value of the formant frequencies with increased intrinsic vocal tract nonlinearities, raised the error. The glottal inverse filtering algorithms predicted the open phase ripple factor relatively well compared to the closed phase ripple factor, irrespective of the source-filter interaction level. Overall, the error was less than 5% for the open phase ripple factor, and less than 100% for close-phase ripple factor, MFDR, and the peak flow, indicating that the glottal inverse filtering algorithms can be useful without *a priori* knowledge of the vocal tract shape.

The current study included only Level 1 source-filter interaction, the effect of acoustic feedback on glottal airflow. Follow-up work will also include the role of Level 2 source-filter interaction, the effect of acoustic feedback on vocal fold movement (Palaparthi et al., 2019) in glottal inverse filtering. It must be recognized that near a resonance frequency, there is a greater likelihood that Level 2 interaction occurs (Titze, 2008). The amplitude of vibration can be modified by acoustic pressures, which will change the glottal flow. It should also be noted that the conclusions from the current study were based on one specific vocal fold model. Similar experiments on other computational models would strengthen the conclusions. The current study also did not assess the aspects of the linear glottal inverse filtering algorithms that make them perform better or worse in the presence of source-filter interaction. Whether better vocal tract estimation is enough should be explored in future studies.

Acknowledgment

This work was supported by the National Institute of Deafness and Communication Disorders under Grant 5R01DC012045. We thank Paavo Alku group for sharing the source code with us for the QCP and QPR algorithms.

Appendix

TABLE IV.

Area functions in cm^2 of eleven vowels used in the current study at distance × from the glottis.

X (cm)	/i/	/I/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/
0	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.8	0.397	0.372	0.353	0.328	0.318	0.274	0.279	0.301	0.336	0.359	0.303
1.2	0.341	0.318	0.273	0.255	0.213	0.194	0.235	0.294	0.335	0.381	0.267
1.6	0.35	0.329	0.256	0.242	0.169	0.164	0.242	0.35	0.405	0.483	0.289
2	0.516	0.484	0.362	0.34	0.222	0.215	0.344	0.528	0.619	0.753	0.423
2.4	0.939	0.879	0.675	0.63	0.433	0.402	0.608	0.909	1.08	1.3	0.748
2.8	1.55	1.44	1.14	1.05	0.77	0.67	0.945	1.38	1.66	1.99	1.17
3.2	2.05	1.88	1.51	1.36	1.02	0.811	1.1	1.61	2.02	2.44	1.41
3.6	2.27	2.03	1.63	1.41	1.08	0.732	0.966	1.49	2.02	2.49	1.34
4	2.45	2.14	1.71	1.41	1.11	0.618	0.781	1.28	1.92	2.43	1.2
4.4	2.82	2.41	1.96	1.55	1.27	0.588	0.683	1.17	1.94	2.5	1.17
4.8	3.32	2.79	2.31	1.78	1.52	0.606	0.632	1.1	2.02	2.64	1.19
5.2	3.84	3.18	2.69	2.01	1.81	0.627	0.581	1.03	2.08	2.75	1.2
5.6	4.3	3.52	3.03	2.22	2.07	0.637	0.518	0.939	2.09	2.79	1.18
6	4.61	3.73	3.27	2.34	2.28	0.623	0.44	0.811	2.01	2.71	1.11
6.4	4.8	3.85	3.44	2.42	2.46	0.611	0.37	0.682	1.89	2.55	1.04
6.8	4.95	3.96	3.62	2.53	2.67	0.649	0.343	0.603	1.8	2.42	1
7.2	5.08	4.07	3.83	2.69	2.96	0.751	0.365	0.575	1.75	2.3	1.03
7.6	5.12	4.13	4.01	2.86	3.26	0.911	0.425	0.58	1.71	2.18	1.08
8	4.98	4.07	4.08	2.98	3.52	1.11	0.511	0.596	1.64	2.01	1.15
8.4	4.61	3.81	3.98	2.98	3.65	1.3	0.602	0.597	1.5	1.74	1.18
8.8	4.08	3.45	3.75	2.9	3.7	1.52	0.719	0.61	1.35	1.46	1.2
9.2	3.52	3.06	3.49	2.82	3.72	1.81	0.904	0.668	1.23	1.22	1.27
9.6	2.92	2.63	3.17	2.7	3.66	2.14	1.15	0.761	1.12	1	1.36
10	2.27	2.16	2.75	2.5	3.48	2.45	1.42	0.872	1.01	0.802	1.44
10.4	1.63	1.66	2.27	2.23	3.16	2.73	1.72	1	0.907	0.625	1.51
10.8	1.08	1.22	1.79	1.95	2.78	2.97	2.06	1.18	0.833	0.497	1.59
11.2	0.708	0.919	1.44	1.76	2.47	3.31	2.56	1.5	0.865	0.468	1.79
11.6	0.5	0.763	1.23	1.7	2.28	3.8	3.28	2.05	1.04	0.557	2.16
12	0.374	0.676	1.08	1.69	2.1	4.3	4.14	2.77	1.32	0.738	2.63
12.4	0.291	0.621	0.956	1.67	1.91	4.72	5	3.59	1.67	0.995	3.13
12.8	0.252	0.608	0.871	1.67	1.72	5.03	5.83	4.48	2.1	1.35	3.65
13.2	0.27	0.666	0.859	1.73	1.6	5.31	6.62	5.43	2.63	1.83	4.21
13.6	0.349	0.798	0.922	1.86	1.57	5.52	7.31	6.35	3.24	2.43	4.76

X (cm)	/i/	/I/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ̂/
14	0.484	0.984	1.04	2.02	1.58	5.59	7.7	7.02	3.79	3.02	5.18
14.4	0.693	1.23	1.22	2.2	1.67	5.52	7.73	7.31	4.21	3.52	5.39
14.8	0.987	1.53	1.49	2.42	1.87	5.34	7.37	7.13	4.39	3.8	5.36
15.2	1.28	1.77	1.74	2.54	2.09	4.91	6.46	6.24	4.11	3.63	4.89
15.6	1.47	1.85	1.91	2.5	2.28	4.23	5.06	4.74	3.33	2.94	3.99
16	1.63	1.88	2.09	2.44	2.57	3.58	3.63	3.13	2.39	2.05	2.99
16.4	1.82	1.93	2.37	2.47	3.07	3.15	2.5	1.83	1.56	1.24	2.17
16.8	1.94	1.93	2.64	2.53	3.65	2.9	1.7	0.939	0.919	0.629	1.54
17.2	1.86	1.8	2.71	2.48	4.02	2.77	1.25	0.475	0.518	0.267	1.14
17.6	1.48	1.46	2.36	2.2	3.73	2.69	1.17	0.386	0.372	0.145	1

References

Airaksinen M, Raitio T, Story B, Alku P, 2014 Quasi closed phase glottal inverse filtering analysis with weighted linear prediction. IEEE/ACM Trans. Audio Speech Lang. Process. 22(3), 596–607.

Airaksinen M, Backstrom T, Alku P, 2017 Quadratic programming approach to glottal inverse filtering by joint norm-1 and norm-2 optimization. IEEE/ACM Trans. Audio Speech Lang. Process. 25(5), 929–939.

Alku P, 1992 Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Commun. 11(23), 109–118.

Alku P, 2011 Glottal inverse filtering analysis of human voice production-a review of estimation and parameterization methods of the glottal excitation and their applications. Sadhana. 36, 623–650.

Alku P, Murtola T, Malinen J, Kuortti J, Story B, Airaksinen M, Salmi M, Vilkman E, Geneid A, 2019 OPENGLOT-An open environment for the evaluation of glottal inverse filtering. Speech Communication. 107, 38–47.

Alzamendi GA, Schlotthauer G, 2017 Modeling and joint estimation of glottal source and vocal tract filter by state-space methods. Biomedical Signal Processing and Control. 37, 5–15.

Berezina MA, Rudoy D, Wolfe PJ, 2010 Autoregressive modeling of voiced speech. IEEE International Conference on Acoustics, Speech, and Signal Processing 5042–5045.

Bozkurt B, Doval B, D' Alessandro C, Dutoit T, 2005 Zeros of z-transform representation with application to source-filter separation in speech. IEEE Signal Process. Lett. 12(4), 344–347.

Chien YR, Mehta DD, Guonason J, Zanartu M, Quatieri TF, 2017 Evaluation of glottal inverse filtering algorithms using a physiologically based articulatory speech synthesizer. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 25(8), 1718–1730.

Degottex G, Kane J, Drugman T, Raitio T, Scherer S, 2014 COVAREP-A collective voice analysis repository for speech technologies. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy.

Drugman T, Alku P, Yegnanarayana B, Alwan A, 2014 Glottal source processing: from analysis to applications. Computer Speech and Language, 28(5), 1117–1138.

Drugman T, Bozkurt B, Dutoit T, 2011 Causal-anticausal decomposition of speech using complex cepstrum for glottal source estimation. Speech Commun. 53, 855–866.

Flanagan JL, 1972 Speech Analysis: Synthesis and Perception. Springer-Verlag, Berlin pp. 36–38.

Fröhlich M, Michaelis D, Strube HW, 2001 SIM-simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals. J. Acoust. Soc. Am. 110(1), 479–488. [PubMed: 11508972]

Fu Q, Murphy P, 2006 Robust glottal source estimation based on joint source-filter model optimization. IEEE Trans. Audio, Speech, Lang. Process. 14(2), 492–501.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K, 1995 Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97(5), 3099–3111. [PubMed: 7759650]

Hubert M, Vandervieren E, 2008 An adjusted boxplot for skewed distributions. Computational Statistics & Data Analysis. 52(12), 5186–5201.

Liljencrants J, 1985 Speech synthesis with a reflection-type line analog. Ph.D. Dissertation, Royal Institute of Technology, Department of Speech Communication and Music Acoustics, Stockholm, Sweden.

Markel J, Gray AH Jr., 1980 Linear Prediction of Speech. Springer-Verlag, New York, New York, USA.

Maxfield L, Palaparthi A, Titze IR, 2017 New evidence that nonlinear source-filter coupling affects harmonic intensity and fo stability during instances of harmonics crossing formants. J. Voice. 31(2), 149–156. [PubMed: 27501922]

Milenkovic P, 1986 Glottal inverse filtering by joint estimation of an AR system with a linear input model. IEEE Trans. Acoust., Speech, Signal Process. 34(1), 28–42.

Miller R, 1959 Nature of the vocal cord wave. J. Acoust. Soc. Am. 31(6), 667–677.

Mokhtar P, Story B, Alku P, Ando H, 2018 Estimation of the glottal flow from speech pressure signals: Evaluation of three variants of iterative adaptive inverse filtering using computational physical modeling of voice production. Speech Commun. 104, 24–38.

Palaparthi A, Maxfield L, Titze IR, 2019 Estimation of source-filter interaction regions based on electroglottography. J. Voice. 33(3), 269–276. [PubMed: 29277351]

Rothenberg M, 1973 A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. J. Acoust. Soc. Am. 53(1), 1632–1645. [PubMed: 4719255]

Rothenberg M, Zahorian S, 1977 Nonlinear inverse filtering technique for estimating the glottal area waveform. J. Acoust. Soc. Am. 61(4), 1063–1071. [PubMed: 864095]

Rothenberg M, 1981 Acoustic interaction between the glottal source and the vocal tract in Vocal Fold Physiology, edited by Stevens KN, and Hirano M University of Tokyo Press, Tokyo, 305–328.

Sahoo S, Routray A, 2016 A novel method of glottal inverse filtering. IEEE/ACM Trans. Audio Speech Lang. Process. 24(7), 1230–1241.

Schutte H, 1980 The efficiency of voice production. Gronigen: State University Hospital.

Story BH, 1995 Physiologically based speech simulation using an enhanced wave-reflection model of the vocal tract. Ph.D. Dissertation, University of Iowa, Iowa City.

Story BH, Titze IR, Hoffman EA, 1996 Vocal tract area functions from magnetic resonance imaging. J. Acoust. Soc. Am. 100(1), 537–554. [PubMed: 8675847]

Suthers RA, Rothgerber JR, Jensen KK, 2016 Lingual articulation in songbirds. J. Experimental Biology. 219, 491–500.

Titze IR, 1984 Parameterization of the glottal area, glottal flow, and vocal fold contact area. J. Acoust. Soc. Am. 75(2), 572–580.

Titze IR, 1988 The physics of small-amplitude oscillation of the vocal folds. J. Acoust. Soc. Am. 83(4), 1536–1552. [PubMed: 3372869]

Titze IR, 1989 A four-parameter model of the glottis and vocal fold contact area. Speech Communication 8, 191–201.

Titze IR, Svec JG, Popolo PS, 2003 Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. J Speech Lang Hear Res. 46(4), 919–932. [PubMed: 12959470]

Titze IR 2006 The Myoelastic Aerodynamic Theory of Phonation. National Center for Voice and Speech, Salt Lake City, UT, pp. 197–200.

Titze IR, 2008 Nonlinear source-filter coupling in phonation: Theory. J. Acoust. Soc. Am. 123(5), 2733–2749. [PubMed: 18529191]

Titze IR, Palaparthi A, Smith SL, 2014 Benchmarks for time-domain simulation of sound propagation in soft-walled airways: Steady configuration. J. Acoust. Soc. Am. 136(6), 3249–3261. [PubMed: 25480071]

Titze IR, Palaparthi A, 2016 Sensitivity of source-filter interaction to specific vocal tract shapes. IEEE/ACM Trans. Audio, Speech, Lang. Process. 24(12), 2507–2515.

Titze IR, Palaparthi A, 2018 Radiation efficiency for long-range vocal communication in mammals and birds. J Acoust Soc Am. 143(5), 2813–2824. [PubMed: 29857705]

Titze IR, Worley AS, 2009 Modeling source-filter interaction in belting and high-pitched operatic male singing. J. Acoust. Soc. Am. 126(3), 1530–1540. [PubMed: 19739766]

Tom K, Titze IR, Hoffman EA, Story BH, 2001 Three-dimensional vocal tract imaging and formant structure: varying vocal register, pitch, and loudness. J Acoust Soc Am. 109(2), 742–747. [PubMed: 11248978]

Walker J, Murphy P, 2007 A review of glottal waveform analysis In: Stylianou Y, Faundez-Zanuy M, Esposito A, (eds), Progress in nonlinear speech processing. Lecture notes in computer science, Vol. 4391, Springer, Berlin, Heidelberg.

Zhang Z, 2016 Mechanics of human voice production and control. J. Acoust. Soc. Am. 140(4), 2614–2635. [PubMed: 27794319]

HIGHLIGHTS

- The study systematically assesses the role of source-filter interaction on glottal inverse filtering.
- Testing on natural signals alone cannot reveal the efficacy of the algorithms, as the true glottal flow waveform is not available for comparison.
- Hence, this study was designed to perform a thorough analysis of the efficacy of four linear glottal inverse filtering methods by using test signals generated by a mathematical model that included vocal tract nonlinearities and source-filter interaction.
- Feedback-related nonlinearity (source-filter interaction) affected the recovered signal primarily when fo was below the first formant frequency of a vowel.
- The prediction error increased when fo was close to the first formant frequency due to the difficulty of estimating the precise value of resonance frequencies, which was exacerbated by nonlinear kinetic losses in the vocal tract and time-varying glottal losses.

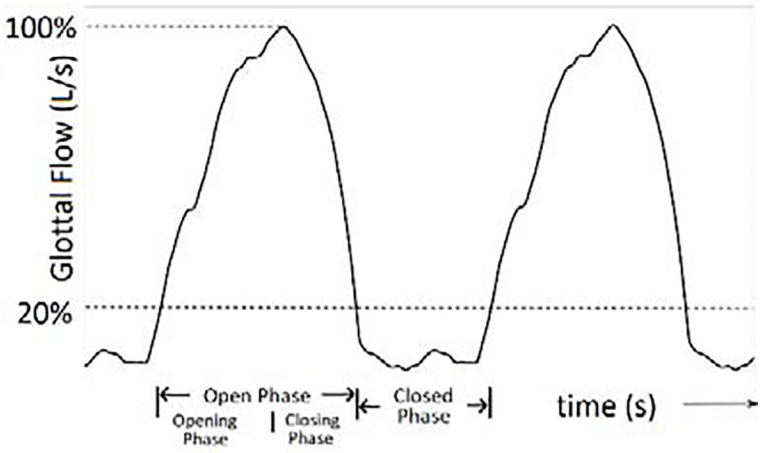

Fig. 1.

A typical glottal flow waveform depicting opening, closing, open, and closed phases. Horizontal line depicts the threshold at 20% from the maximum glottal flow that separates open phase from closed phase.

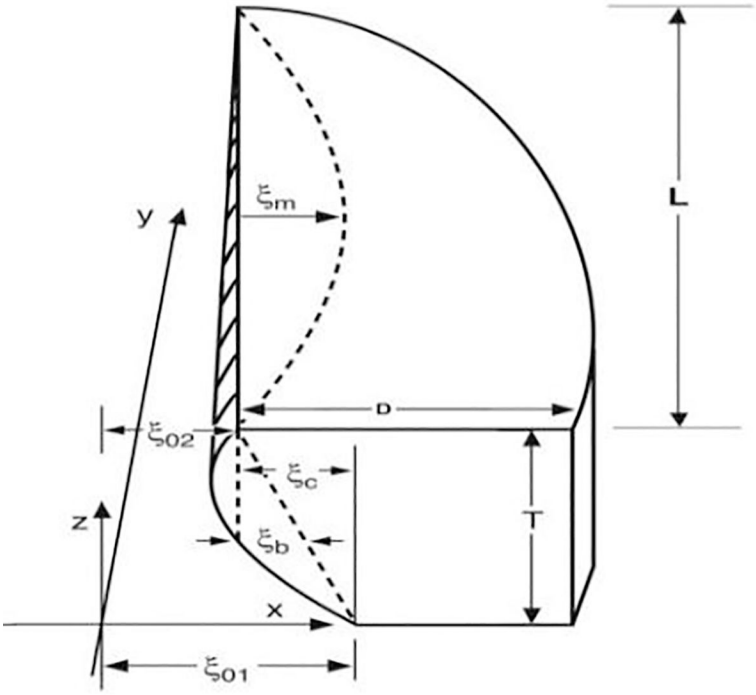

Fig. 2.

Parameters for kinematic movement and posturing. ξ_{01} is the entry glottal half-width, ξ_{02} is the exit glottal half-width, ξ_b is the medial surface bulging, ξ_m is the vocal fold vibrational amplitude.

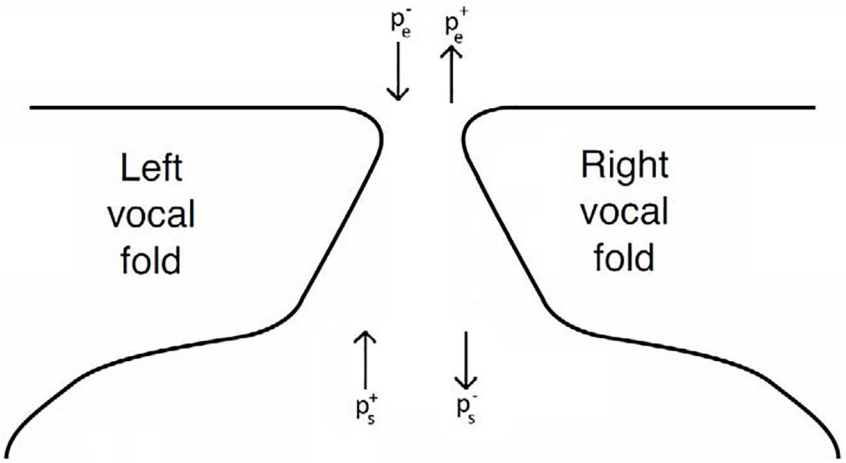

Fig. 3.

Coronal view of the vocal folds depicting the direction of supraglottal partial pressures (p_e^-, p_e^+) and subglottal partial pressures (p_s^+, p_s^-)

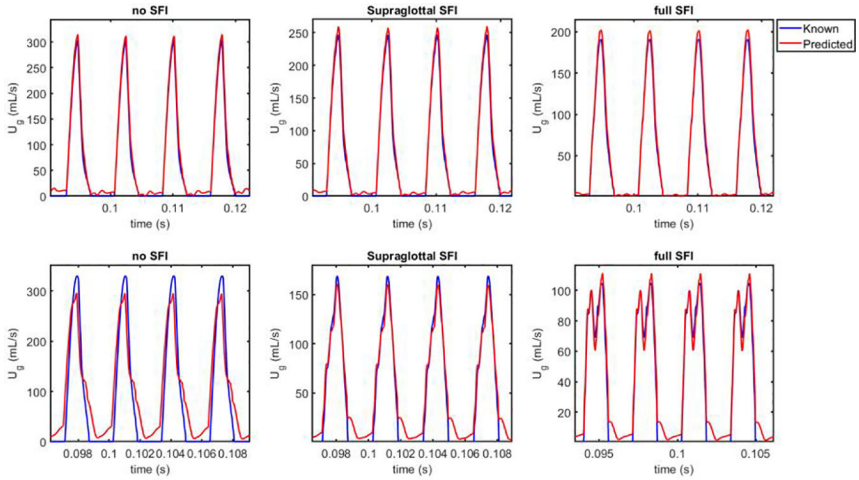

Fig. 4.

Known and Predicted glottal flow waveforms of /i/ vowel from True Impulse response inverse filtering. Top row is from fixed-f_o test cases and the bottom row is from the near-formant f_o test cases.

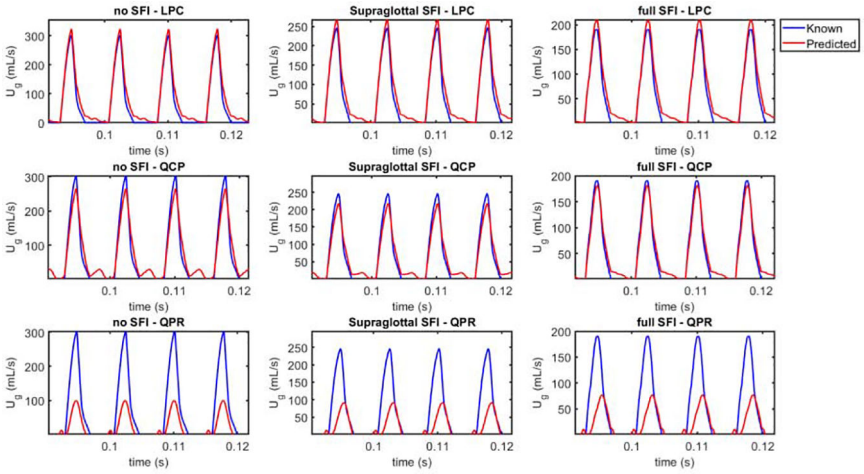

Fig. 5.

Known and Predicted glottal flow waveforms of /i/ vowel for fixed-f_o test cases from the three glottal inverse filtering algorithms.

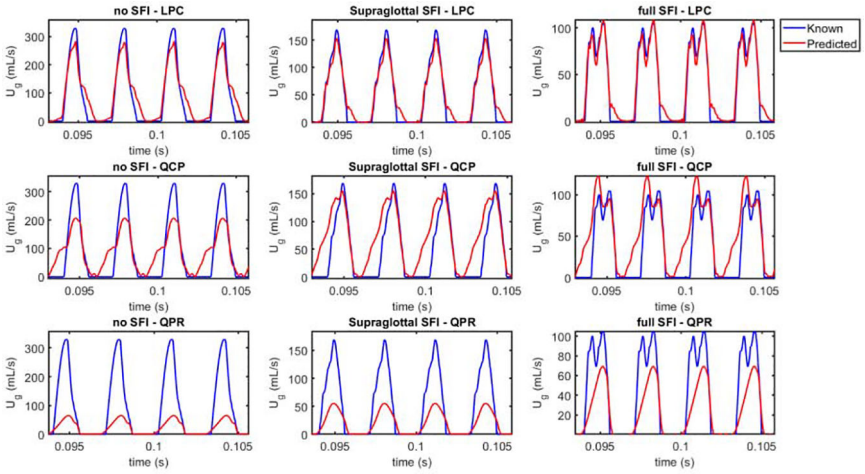

Fig. 6.

Known and Predicted glottal flow waveforms of /i/ vowel for near-formant f_o test cases from the three glottal inverse filtering algorithms.

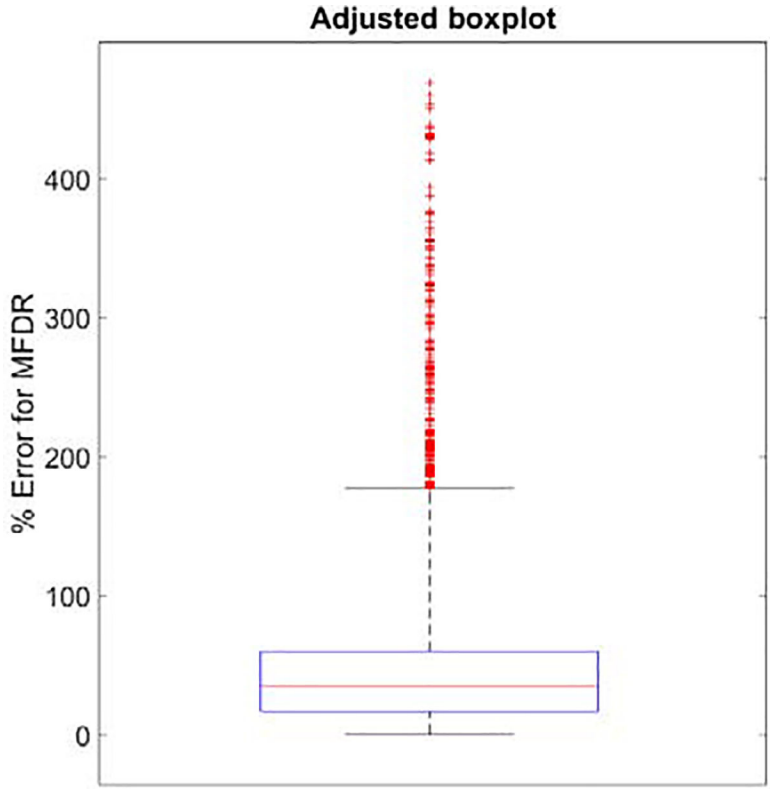

Fig. 7.

Boxplot for MFDR data under full SFI and near-formant f_o condition. Red '+' symbols in the boxplot denote outliers.

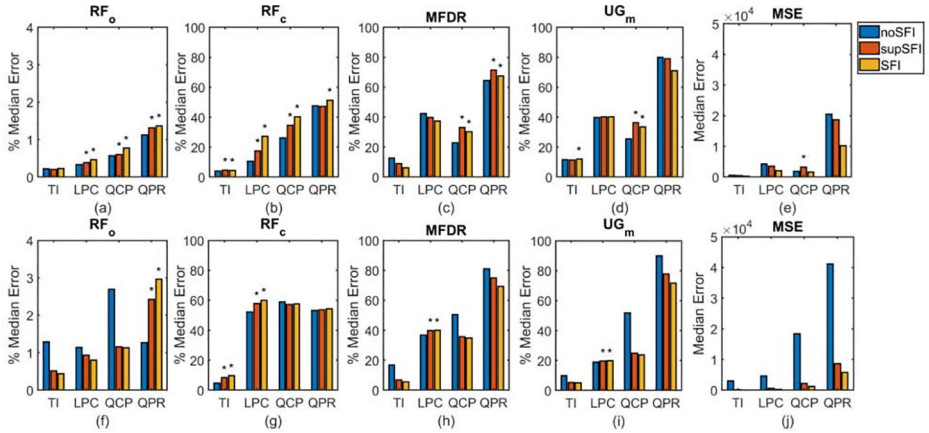

Fig. 8.

Median of Error in Percentage for Open Phase Ripple Factor (RF_o), Closed Phase Ripple Factor (RF_c), Maximum Flow Declination Rate (MFDR), and Glottal Flow Amplitude (UG_m), and Median of Error for Mean Squared Error (MSE) for (a to e) Fixed-f_o test cases and (f to j) Near-formant f_o test cases. TI – True Impulse method, LPC – Linear Predictive Coding based method, QCP – Quasi-closed phase method, QPR – Quadratic programming approach.

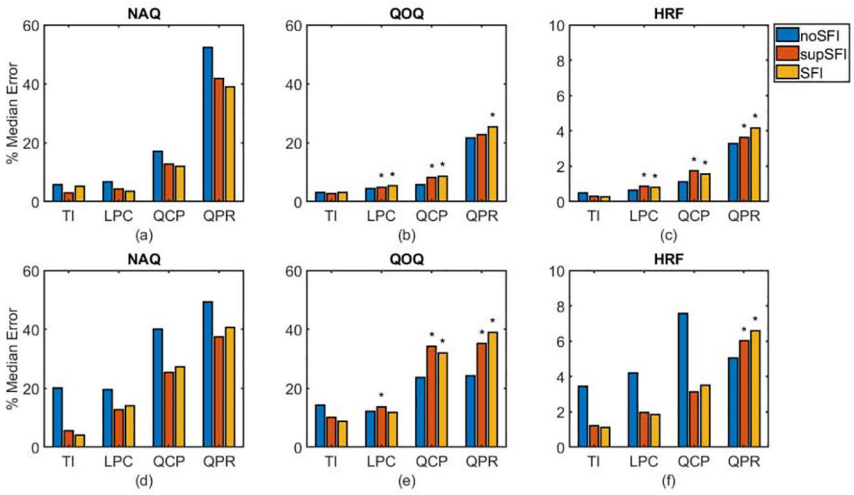

Fig. 9.

Median of Error in Percentage for Normalized Amplitude Quotient (NAQ), Quasi-open Quotient (QOQ), and Harmonic Richness Factor (HRF) for (a to c) Fixed-f_o test cases and (d to f) Near-formant f_o test cases. TI – True Impulse method, LPC – Linear Predictive Coding based method, QCP – Quasi-closed phase method, QPR – Quadratic programming approach.

TABLE I

F_o VALUES USED FOR EACH VOWEL UNDER NEAR-FORMANT F_O TEST CONDITION

Vowel	**/i/**	**/I/**	**/e/**	**/ɛ/**	**/æ/**	**/ɑ/**	**/ɔ/**	**/o/**	**/ʊ/**	**/u/**	**/ʌ/**	**ut**
f_o (Hz)	320	413	464	570	575	740	630	495	470	363	611	561

ut – uniform tube vocal tract

TABLE II

NUMBER OF VOWELS OUT OF 12 FOR FIXED-F_O TEST CASES WITH SIGNIFICANTLY HIGHER ERROR STATISTICALLY WHEN SUPSFI IS COMPARED TO NOSFI AND WHEN SFI IS COMPARED TO NOSFI

Method	SFI Level	RF_o	RF_c	MFDR	UG_m	MSE	NAQ	QOQ	HRF	Total
TI	supSFI	2	2	0	2	0	0	2	0	8 (8.3%)
	fullSFI	5	7	0	4	0	4	5	0	25 (26.0%)
LPC	supSFI	7	0	0	0	0	0	8	8	23 (24.0%)
	fullSFI	11	11	0	3	0	0	10	8	43 (44.8%)
QCP	supSFI	7	12	12	11	6	2	9	12	71 (74.0%)
	fullSFI	11	12	9	9	3	1	10	9	64 (66.7%)
QPR	supSFI	7	0	10	0	0	0	2	8	27 (28.1%)
	fullSFI	10	7	7	0	0	1	5	10	40 (41.7%)
Total		60 (62.5%)	51 (53.1%)	38 (39.6%)	29 (30.2%)	9 (9.4%)	8 (8.3%)	51 (53.1%)	55 (57.3%)	

TABLE III

NUMBER OF VOWELS OUT OF 12 FOR NEAR-FORMANT F_O TEST CASES WITH SIGNIFICANTLY HIGHER ERROR STATISTICALLY WHEN SUPSFI IS COMPARED TO NOSFI AND WHEN SFI IS COMPARED TO NOSFI

Method	SFI Level	RF_o	RF_c	MFDR	UG_m	MSE	NAQ	QOQ	HRF	Total
TI	supSFI	3	5	1	2	0	0	4	0	15 (15.6%)
	fullSFI	3	7	0	0	0	0	3	0	13 (13.5%)
LPC	supSFI	6	11	5	8	0	1	6	0	37 (38.5%)
	fullSFI	6	11	6	8	0	1	4	0	36 (37.5%)
QCP	supSFI	0	3	0	0	0	0	10	0	13 (13.5%)
	fullSFI	2	4	1	0	0	3	11	1	22 (22.9%)
QPR	supSFI	11	4	1	0	0	3	10	8	37 (38.5%)
	fullSFI	12	6	1	0	0	4	10	8	41 (42.7%)
Total		43 (44.8%)	51 (53.1%)	15 (15.6%)	18 (18.7%)	0 (0%)	12 (12.5%)	58 (60.4%)	17 (17.7%)	