Estimation of the glottal flow from speech pressure signals: Evaluation of three variants of iterative adaptive inverse filtering using computational physical modelling of voice production

Parham Mokhtari a,⁎, Brad Story b, Paavo Alku c, Hiroshi Ando a

a Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan
b Speech Acoustics Laboratory, Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721, USA
c Department of Signal Processing and Acoustics, Aalto University, School of Electrical Engineering, P.O. Box 12200, Aalto FI 00076, Finland

ARTICLE INFO

Keywords:
Glottal inverse filtering
IAIF
Voice production
Physical modelling

ABSTRACT

The aim of this study is to comparatively review and evaluate three variants of the glottal inverse filtering algorithm based on iterative adaptive inverse filtering (IAIF): the Standard algorithm, and two recently proposed variants that use iterative optimal preemphasis (IOP) and a glottal flow model (GFM), respectively. To enable an objective evaluation, a computational physical model of voice production is used to generate time-domain signals pertaining to both the input glottal flow and the output speech pressure, for a wide range of vowels, fundamental frequencies, and voice qualities (including co-variation of phonation type and loudness). Furthermore, for a fair comparison, the three key parameters of IAIF are selected by an exhaustive search to minimize the root-mean-square error between the estimated and reference glottal flow derivative in each analyzed frame and performance is assessed with two time-domain and two frequency-domain error measures. A conventional evaluation is also carried out with fixed parameter values determined by cross-validation. Results indicate that IOP tends to yield the lowest errors for nonback vowels (reducing errors by 31% on average compared with Standard), especially for not too high fundamental frequencies and not too pressed voice qualities; GFM becomes competitive for normal phonations when fixed parameter values are used; and in other cases, Standard IAIF is still recommended. In addition, the results suggest that not only the overall spectral tilt (as controlled by IOP and GFM) but also the balance between the levels of different spectral regions, can be important for accurate estimation of the glottal flow.

1. Introduction

Glottal inverse filtering refers to the process of estimating the source of voiced speech sounds, the volume velocity airflow that passes through the orifice between the vibrating vocal folds. Most research efforts in this area have focused on the analysis of the speech pressure signal recorded with a microphone in the free-field in front of the lips, as this offers a convenient and non-invasive form of measurement, in contrast with, e.g., oral airflow measured with a pneumotachograph mask (Rothenberg, 1973). The most common approach is to parametrically model the acoustic filtering of the supralaryngeal vocal tract (mainly the vocal tract resonances and acoustic radiation at the lips), and then to remove these effects by filtering the speech signal through the inverse of the model, thus leaving an estimate of only the laryngeal signal: the glottal volume-velocity waveform, also known as the glottal flow. The difficulty is that, as neither the vocal tract response nor the glottal source can be measured directly and independently, this inverse problem amounts to a blind decomposition of the speech signal, and the consequent lack of ground-truth data makes evaluation of inverse filtering results problematic. Despite these issues, glottal flow estimation by inverse filtering has received considerable attention during the past six decades (Alku, 2011).

One of the most widely used algorithms, and one that is still regarded as a benchmark for comparison with newer methods (e.g., Drugman et al., 2012; Airaksinen et al., 2014a; Chien et al., 2017), is Iterative Adaptive Inverse Filtering (IAIF: Alku, 1992). It is founded on the classical model of speech production (at least for non-nasalized voiced sounds) as a linear cascade of three processes in the frequency domain (Fant, 1970): G(f), the spectrum of the volume-velocity excitation provided by the glottal source; V(f), the transfer function of the vocal-tract airway that imparts resonances which appear as formant peaks in the magnitude spectrum; and L(f), the lip-radiation effect which is a (leaky)

⁎ Corresponding author.
E-mail address: parham@nict.go.jp (P. Mokhtari).

https://doi.org/10.1016/j.specom.2018.09.005
Received 5 April 2018; Received in revised form 30 August 2018; Accepted 11 September 2018
Available online 12 September 2018
0167-6393/© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)
differentiator converting volume-velocity at the lips to acoustic pressure in the farfield. Thus, in the z-domain, the speech sound S is written as:

\[ S(z) = G(z) V(z) L(z), \]

where lip-radiation is of the form:

\[ L(z) = 1 - b z^{-1}, \quad 0 < b \leq 1. \]  

IAIF is an automatic algorithm that successively models using linear prediction (LP) analysis (Markel and Gray, 1976), then removes by inverse filtering, the effects of \( V(z) \), \( L(z) \), and \( G(z) \). After two such iterations, it ultimately removes \( V(z) \) and \( L(z) \) to leave an estimate of the glottal flow \( g(n) \), where \( n \) is the discrete-time index.

Recently, two potential improvements to IAIF appeared in the literature. First, motivated by the idea of completely flattening the overall spectral tilt assigned to the vocal tract transfer function, Mokhtari and Ando (2017) proposed iterative optimal preemphasis (IOP) to replace the first step of the standard algorithm, thus forming a modified algorithm named IOP-IAIF. Second, motivated by the idea of constraining the glottal flow model (GFM) to one derived by a 3rd-order LP analysis, Perrotin and McLoughlin (2017) proposed a modified algorithm named GFM-IAIF.

The aim of the current study is to comparatively evaluate the original and two recent variants of IAIF, henceforth referred to in short as Standard, IOP, and GFM, respectively. However as mentioned earlier, the ground-truth glottal flow is not available with recordings of natural speech. To circumvent this problem and enable an objective evaluation of glottal inverse filtering, a few studies have used instead numerical simulations of voice production by physical modelling (Alku et al., 2006; Airaksinen et al., 2014a; Chien et al., 2017), and one study proposed an experimental setup enabling simultaneous acoustic measurements of output pressure and input flow on a physical (hardware) model (Chu et al., 2013). Similar to the first group of studies, the current work uses computational physical modelling of voice production. This approach has two benefits. First, the physical modelling approach provides both the vocal production system’s input (i.e., the ground truth represented by the glottal flow) and output (i.e., the speech pressure signal at the lips) as time-domain signals. Second, this approach enables avoiding the utilization of the same modelling principle of voice production (i.e., a linear source-tract model) both in the inverse filtering algorithm to be tested and in the generation of the algorithm’s test material.

An advantage of all three variants of IAIF tested here, compared with several other glottal inverse filtering methods (e.g., closed phase analysis: Wong et al., 1979; complex cepstrum decomposition: Drugman et al., 2012; zeros of z-transform: Bozkurt et al., 2005; quasi closed phase analysis: Airaksinen et al., 2014a), is that the glottal flow can be estimated from the speech signal using straightforward steps which involve only LP analysis and inverse filtering, without the need to locate instants of glottal opening or closure.

However, an often overlooked issue is that, for IAIF to be completely automated, the values of three key parameters must be somehow decided. As discussed later in connection with Fig. 1 (Section 2), the parameters are: the LP analysis order \( M_p \) used to model the vocal tract transfer function, the LP analysis order \( M_g \) used to model the glottal flow spectrum, and the lip radiation coefficient \( b \) (cf. Eq. (2)). In most studies involving IAIF, including those that evaluated IAIF using physical models (Alku et al., 2006; Airaksinen et al., 2014a; Chien et al., 2017), the values of the three parameters were simply fixed across all the data analyzed. In contrast, Chu et al. (2013) used four different values for \( M_p \) and selected the result that yielded the lowest mean-square error between estimated and reference glottal flows. Similarly, although using a glottal inverse filtering method other than IAIF, Airaksinen et al. (2014b) illustrated the sensitivity of estimated glottal flow to the lip radiation parameter, \( b \), and proposed one way of automatically tuning \( b \) to yield acceptable results. In the present study, to extract the best possible performance from each variant of IAIF and thus compare them on an equal footing, for every analyzed frame each algorithm is run across a wide range of parameter value combinations (as detailed in Section 2.4) and, through an exhaustive search, the glottal flow is selected whose derivative best matches (in a least squares sense) the derivative of the reference glottal flow. For comparison, a conventional evaluation is also performed with fixed parameter values, using cross-validation.

Section 2 describes the inverse filtering algorithms and parameter settings in more detail, focusing on the differences between the three variants. Section 3 summarizes the physical model of voice production used to generate the sound samples, which cover a range of vowels, fundamental frequencies, and phonation types. Section 4 details the time- and frequency-domain measures of error used in the evaluation. Section 5 presents the evaluation results, and Section 6 offers a concluding discussion.

2. Glottal inverse filtering methods

A flow diagram that unifies the three variants of IAIF is shown in Figs. 1 and 2. It is assumed that the input \( s(n) \) represents one frame of speech signal that has already been cleared of low-frequency ambient sounds by linear-phase high-pass filtering (with a cut-off around 70 Hz), and downsampled to a specified sampling rate. The next three sections summarize each algorithm in turn, then Section 2.4 describes the exhaustive search used with all three methods to find the best parameter values and Section 2.5 describes the evaluation with cross-validation and fixed parameters.

2.1. Standard IAIF

All three algorithms begin (step 1) by approximating modelling, then removing, the combined effects of glottal flow and lip radiation. The Standard method does this (cf. Fig. 2) by applying preemphasis just once, i.e., 1st-order LP analysis to define a single real pole that is optimal in the mean square sense (cf. Markel and Gray, 1976, p. 216) followed by inverse filtering to remove the effect of that pole. As the signal \( v_1(n) \) output from step 1 approximately contains the influence of only the vocal tract, step 2 involves modelling (step 2a), then removal from the speech signal by inverse filtering (step 2b), of an all-pole vocal tract transfer function of order \( M_p \). Having thus removed a first estimate of the vocal tract model from the speech signal, in step 3 the lip radiation effect is removed by integration (the inverse of Eq. (2), assuming a particular value for parameter \( b \)) to obtain the first, intermediate estimate \( g_{int}(n) \) of the glottal flow. While this concludes the 1st iteration of IAIF, a 2nd iteration is necessary to refine the estimates mainly because, as mentioned earlier, step 1 was only a rough approximation which can now be improved upon by using \( g_{int}(n) \).

Therefore, the 2nd iteration begins by obtaining a more accurate model of the glottal flow (by analysis of \( g_{int}(n) \) to find an all-pole model of order \( M_p \) in step 4a), and removing the effects of that model from the speech signal by inverse filtering (step 4b). The resulting signal is integrated to remove the lip radiation effect (step 5), leaving \( v_2(n) \) which is a more refined version of \( v_1(n) \) obtained earlier. Finally, analogous to steps 2 and 3 in the 1st iteration, step 6 models (step 6a) and removes from the speech signal by inverse filtering (step 6b) a more refined model of the vocal tract transfer function, and step 7 removes the lip radiation effect by integration, resulting in the final estimate \( g(n) \) of glottal flow.

Note that the final estimate of the vocal tract transfer function (LP model of order \( M_p \)) is given at the output of step 6a, and that in any given run of the algorithm the lip radiation effect is modelled as in Eq. (2) with a particular value for \( b \). Optimal selection of values for \( b \), \( M_p \), and \( M_g \), will be discussed in Section 2.4.
Fig. 1. A unified flow-diagram of the IAIF glottal inverse filtering algorithm, including the Standard, IOP, and GFM methods. For algorithm-specific details of step 1, see Fig. 2. Note that the capitalized versions of $v_1$, $s_{out}$, and $v_2$, refer to their LP models (and not merely to their z-transform).

Fig. 2. Algorithm-specific flow-diagrams for step 1 of the IAIF variants Standard, IOP, and GFM.
2.2. IOP-IAIF

The IOP variant of IAIF is the same as the Standard algorithm except for step 1, which is replaced with iterative optimal preemphasis (cf. Fig. 2). This entails repeatedly modelling (with 1st-order LP analysis in step 1a) then removing (by inverse filtering in step 1b) one real pole at a time, until the magnitude of the autoregressive coefficient \( a_1 \) becomes sufficiently small. Here a threshold of 0.01 is applied, which is small enough to ensure that any subsequent non-zero value will have negligible effect on the signal spectrum, while avoiding unnecessarily long sequences of \( a_1 \) that might arise from a more conservative threshold.

The main idea behind the IOP algorithm (Mokhtari and Ando, 2017) is to improve the approximate modelling in step 1, such that the input signal \( v_1(n) \) to the vocal tract modelling in step 2 has an optimally flat overall spectral tilt (as defined by zero autocorrelation at lag 1 sample).

Although the estimated vocal tract and glottal flow models are subsequently refined during the 2nd iteration as in the Standard algorithm, step 1 is in fact highly influential in determining the proportions of the overall spectral tilt of the input speech signal that are allocated to the vocal tract response and the glottal spectrum. By design, in IOP most of the tilt is assigned to the glottal spectrum, and the vocal tract is regarded as a passive resonator with essentially no, or little, spectral tilt.

2.3. GFM-IAIF

The GFM variant of IAIF also differs from the Standard algorithm mainly in step 1. As shown in Fig. 2, the input speech signal is first integrated to remove the assumed lip radiation effect (step 1a), then optimal preemphasis is applied (i.e., 1st-order LP modelling in step 1b followed by inverse filtering in step 1c) exactly three times in succession.

The main idea behind the GFM algorithm (Perrotin and McLaughlin, 2017) is to constrain the modelling of the glottal flow spectrum to a strictly 3rd-order autoregressive model. Therefore, in GFM the LP model order \( M_p \) used subsequently in step 4a is fixed to a value of 3. Note, however, that in step 1, a 3rd-order model is not used because that would unintentionally model and remove a formant peak belonging to the vocal tract; instead, as a compromise, three consecutive 1st-order models are applied. In this respect, it is worth noting that even though the underlying motivations of GFM are quite different from that of IOP, in the end GFM is simply a more constrained version of IOP: optimal preemphasis is applied exactly three times following integration (i.e., essentially twice, as the first preemphasis basically reverses the initial integration step) and the glottal flow model is exactly of the 3rd order.

Furthermore, as opposed to the Standard and IOP algorithms, GFM does not include steps 6b and 7; while the final estimate of the vocal tract transfer function is still the output of step 6a, as shown in Fig. 1 the final GFM estimate of glottal flow is \( g_{est}(n) \). Although GFM’s limited use of the 2nd iteration may appear disadvantageous, our aim here is not to propose a modification to any of these algorithms, but rather to compare them in their original form.

2.4. IAIF parameter selection by exhaustive search

As mentioned earlier, most studies employing IAIF in fully automatic mode have assumed fixed values for the three parameters \( b, M_p, \) and \( M_g \). However, the inverse filtering results depend to a great extent on these parameters, especially on \( b \) (e.g., Airaksinen et al., 2014b) and \( M_p \) (e.g., Chu et al., 2013). Indeed, the publicly available Aalto Aparat software tool (Alku et al., 2017) was designed to allow the user to visually explore the resulting glottal flows yielded by IAIF, and thereby to choose the parameter value combination deemed to give the best result for each frame analyzed. Meanwhile, research efforts continue to improve quantification of the quality of estimated glottal flow (e.g., Moore and Torres, 2008) and automate experts’ subjective decisions (e.g., Kane and Gobl, 2013), with the aim of fully automating the entire process including choice of parameter values.

In the present study where the true (reference) glottal flow is known thanks to computational modelling of voice production, the three parameters were varied over a wide range and the best result was selected objectively by minimizing an error between the estimated flow derivative and its reference. In particular, for every analyzed frame:

- the lip radiation coefficient \( b \) was varied from 0.750 to 0.999 in steps of 0.001 (total 250 steps);
- the vocal tract LP model order \( M_p \) was varied according to the sampling frequency \( F_s \) (in Hz), from \( \text{round}(F_s/1000) - 2 \) to \( \text{round}(F_s/1000) + 6 \) in steps of 2 (total 5 steps), and
- the glottal flow LP model order \( M_g \) was set to 3 in GFM according to the explicit model constraint, and varied from 3 to 6 in steps of 1 (total 4 steps) in Standard and IOP.

As the results of LP modelling in general, and the overall spectral tilt in particular, depend somewhat on the available frequency bandwidth, all analyses were repeated at four different sampling rates \( F_s = \{4000, \ 8000, \ 12000, \ 16000\} \) Hz. Thus, for example, at the lowest sampling rate of 4 kHz, \( M_p \) was varied across even integers from 2 to 10, while at the highest rate of 16 kHz, \( M_g \) was varied across even integers from 14 to 22; the ranges for \( M_g \) were chosen to encompass the expected number of poles needed to model the formants across a given frequency range.

For every combination of parameters \( b, M_p, \) and \( M_g \) (total 1250 combinations for GFM, and 5000 combinations for Standard and IOP), an error measure between the estimated and reference glottal flow derivatives (obtained by simple 1st-difference of the flow) was calculated as follows. First, the estimated flow derivative was time-aligned with the reference flow derivative to eliminate the delay introduced by the propagation of acoustic waves along the vocal tract from the glottis to the measurement point at the lips; this was done by searching for the time-shift (discrete number of samples \( N_d \) maximizing the correlation between estimated and reference flow derivatives. Second, the estimated flow derivative was scaled in amplitude to make its total energy equal to that of the reference flow derivative; this step is necessary because the glottal flow estimated by inverse filtering of the acoustic pressure signal has an arbitrary amplitude scale, i.e., the absolute scale and any DC offset are not recoverable. Lastly, the error \( \Delta_{\text{RMS}} \) was calculated as the root mean square (RMS) of the difference between the reference and the estimated (time aligned and energy normalized) flow derivatives (in cm³/s²):

\[
\Delta_{\text{RMS}} = \sqrt{\frac{1}{N_d} \sum_{n=1}^{N_d} (d_{\text{ref}}(n) - d_{\text{est}}(n))^2},
\]

where \( N_d \) accounts for time alignment and the hat symbol denotes energy normalization. The best parameter value combination was then selected as the one yielding the smallest \( \Delta_{\text{RMS}} \) and from the selected glottal flow derivative, the corresponding glottal flow with an amplitude scale matching the reference (cm³/s) was obtained simply by integration followed by mean normalization. As described in Section 4, \( \Delta_{\text{RMS}} \) was also included as one of the error measures used to compare the performance of the inverse filtering algorithms.

2.5. Cross-validation with fixed parameters

The exhaustive-search optimization described in the preceding section allows evaluation of the best possible performance of each algorithm, by finding independently for each analyzed frame the values of parameters \( b, M_p, \) and \( M_g \) that yield the closest match between estimated and reference glottal flow derivative signals. However, the ground-truth glottal flow is not available when analyzing recordings of natural speech, and for this reason many studies employing IAIF have used fixed parameter values. To compare our optimization results with this more conventional method of evaluation while still taking advantage of the computational physical modelling approach, here we use cross-validation: i.e., a “test” subset of the data is analyzed using fixed parameter values obtained by optimization over the remaining “training” subset.
yielded the flows minimizing in group steps: including phonation to Middle physical Fig. 3. (a) Schematic representation of the computational physical model and associated signals. The vocal tract is shown configured as an /a/ vowel. (b) A segment of the 11 sets of signals produced with the configuration in (a) and $f_0 = 110$ Hz, where the thick black and grey lines in each panel denote the most breathy and pressed of the 11 parameter settings, whereas the thin black lines indicate intermediate parameter settings.

More precisely, cross-validation was carried out by the following steps: (i) the 660 samples in our dataset (detailed in Section 3) were randomly divided into 10 groups of 66 samples; (ii) the samples in each group were analyzed by the three algorithms using parameters fixed in value to their median across the remaining 9 groups (as obtained by minimizing $\Delta_{\text{RMS}}$, separately for each sampling rate); and (iii) the glottal flows estimated in this way were evaluated against the reference glottal flow using the four error measures described in Section 4. Although the errors thus obtained are expected to be generally greater than those yielded by exhaustive-search optimization, this 10-fold cross-validation serves to highlight the relative performance of the algorithms under more realistic conditions – e.g., as encountered with fully automated analysis of recorded speech, especially when the amount of data to be analyzed is so large as to preclude the use of human expert judgments.

3. Physical modelling of voice production

Section 3.1 describes the simulation model, and Section 3.2 details the ranges of parameter variation used to obtain the 660 pairs of voiced speech sounds and reference glottal flows analyzed in this study.
3.1. Computational physical model

A computational model of the speech production system, illustrated in Fig. 3(a), was used to simulate vowels with a wide range of variation in voice quality. The voice source component of the model is based on a kinematic representation of the vocal fold medial surfaces in which time-varying surface displacements are superimposed onto a postural configuration (Titze, 1984, 2006; Story, 2013) with a specified fundamental frequency. Structural characteristics of the vocal folds such as surface bulging, adduction, length, and thickness serve as control parameters, as well as fundamental frequency and respiratory pressure. As the vocal fold surfaces vibrate, the model produces a time-varying glottal area, $a_g(t)$, that is coupled to the acoustic pressures and airflows in the trachea and vocal tract through aerodynamic and acoustic considerations (Titze, 2002). The resulting glottal flow, $u_g(t)$, was determined by the interaction of the glottal area with the time-varying pressures present just inferior and superior to the glottis as calculated with the following equation (Titze, 1984):

$$
\frac{d}{dt}a_g(t) = \frac{a_g(t)c}{k_t} \left\{ \left( \frac{a_g(t)^2}{A^+} \right) - \left( \frac{a_g(t)}{A^+} \right)^2 + \frac{4k_t}{\rho c^2} (\rho_c^2(t) - \rho_i^2(t)) \right\}^{1/2},
$$

where $k_t$ is a transglottal pressure coefficient, $\rho$ is the air density, and $c$ is the speed of sound. Additionally, aspiration noise due to glottal turbulence is generated by adding a noise component to the glottal flow when the Reynolds number exceeds a threshold value; details can be found in Story (2013).

The configurations of the vocal tract and trachea were specified by area functions containing 44 and 34 cross-sectional areas, respectively, spaced equally along a length axis in increments of approximately 0.4 cm. Acoustic wave propagation in the trachea and vocal tract was computed with a wave-reflection model that included energy losses due to yielding walls, viscous, heat conduction, and radiation at the lips (Story, 1995).
The \( p^* \) and \( p^\prime \) in Eq. (4) are the partial pressures in the subglottal and supraglottal systems, respectively, that are incident upon the glottal area at each instant of time. \( A^* \) is an equivalent vocal tract area based on the entry areas of the trachea, \( A_t \), and the vocal tract, \( A_v \); it is calculated as \( A^* = (A_t + A_v) / A_t \). It can be noted that, according to Eq. (4), the glottal flow is nonlinearly-dependent on glottal area, vocal tract shape, and propagating acoustic pressures. This interaction of source and filter can cause the harmonic amplitudes in the glottal flow spectrum to be affected by the acoustic resonances (formants) of the vocal tract and tracheal configuration (Titze, 2008; Story and Bunton, 2013). Thus, the characteristics of the glottal flow signal will depend not only on the settings of the vocal fold parameters, but also on the particular shape of the vocal tract and trachea.

This model has been used to generate simulation samples of vowels, words, and phrases for experiments involving both signal analysis and perceptual testing (cf., Bunton and Story, 2010, 2012; Samlan and Story, 2011; Samlan et al., 2013; Lester and Story, 2015; Story and Bunton, 2016). Although a formal experiment to test the naturalness of the output has not been performed, listeners easily accept the stimuli as speech and respond accordingly.

### 3.2. Ranges of parameter variation

A total of 660 pairs of output speech pressure and input glottal flow signal, each of duration 501 ms and sampled with 44.1 kHz, were generated by all combinations of the following parameter variations:

- the shape of the vocal tract airway from glottis to lips was varied over 6 different vowels /a, æ, i, u, o, ò/ as measured by magnetic resonance imaging (MRI) of an adult male (Story et al., 1996; Story, 2009); the vocal tract length was maintained at 17.5 cm for all simulations,
- fundamental frequency of phonation was varied over 10 different values from F₂ (92.4986 Hz) to A₄ (440.0 Hz) in equal intervals of a minor third (equal frequency ratios): i.e., nominally \( f_o = \{92, 110, 131, 156, 185, 220, 262, 311, 370, 440\} \) Hz,
- voice quality was varied over 11 steps from a weak & breathy to a strong & pressed phonation, by linearly co-varying lung pressure \( P_L = \{4000, 5000, 6000, \ldots, 14,000\} \) dyn/cm² and the vocal-fold abduction parameter \( \zeta_o = \{0.0914, 0.0893, 0.0871, \ldots, 0.0700\} \) cm which specifies the prephonatory distance between the superior vocal process and the glottal midline.
For each of the 660 simulations, the parameter settings were held constant over the entire 501 ms duration, although there is variability associated with the generation of glottal turbulence. This was done to ensure that cycle-to-cycle temporal variations were not a factor in the initial comparisons of the inverse filtering techniques. In natural speech, however, there is typically variability in the acoustic characteristics. Thus, an additional set of 110 simulations was generated for the case of the neutral vowel where all parameters were assigned the same settings as before, but the fundamental frequency was perturbed with a random variation to impose a slight “jitter” on the $f_o$ contour. The perturbation was produced with a random number generator, low-pass filtered with a cutoff frequency of 150 Hz, and scaled with a value of 0.02. This had the effect of producing a fundamental frequency jitter ranging from about 0.4% at the lowest values of $f_o$ to 0.7% at the highest.

To demonstrate the output produced by the model, time-varying glottal area, glottal flow, and pressure radiated at the lips are shown in Fig. 3(b) for the 11 voice quality parameter settings when the vocal tract was configured as an /a/ vowel and $f_o = 110$ Hz. The thick black and grey lines in each panel denote the most breathy and pressed of the 11 parameter settings, whereas the thin black lines indicate intermediate parameter settings. It can be observed from both the glottal area and glottal flow signals that the breathiest (thick black line) case presents incomplete glottal closure and nearly sinusoidal vocal fold oscillation, whereas the most pressed case demonstrates rapid glottal closure and long closed phase.

To maintain a limited scope of model-based simulations, it is noted that the vocal tract length was always set to a constant 17.5 cm, and there was no attempt to simulate laryngeal raising or lowering for the different voice qualities. Thus, none of the simulations address differences between males and females, even though the fundamental frequency ranged from 92–440 Hz.

The simulated range of voice quality variations between two extremes, here referred to in short as “breathy” and “pressed”, includes co-variation in vocal effort and phonation type. In partial support of this method of co-varying lung pressure and vocal fold adduction, Fig. 4 compares three sets of output speech spectral harmonics for the vowel /a/ across a comparable range of voice qualities. The two sets of spectra measured on natural recordings (middle and right panels) show inter-speaker differences not only in the formants but also in the amplitude range and frequency dependence of voice quality variations. Nevertheless, overall they share qualitative similarities with the simulated spectra (left panel), especially in regard to the changes in overall spectral tilt and the level of the fundamental relative to the second and higher harmonics. In particular, the voice quality variations in the simulated spectra show an amplitude range that resembles recording 2 at lower frequencies (less than about 1.2 kHz), and recording 1 at higher
frequencies. This qualitative comparison helps to increase confidence in the use of computational physical modelling, and in the relatively simple co-variation of lung pressure and vocal-fold adduction to simulate a natural range of variation in phonation type and loudness (see also Gauffin and Sundberg, 1989, Fig. 4b).

Glottal inverse filtering with the Standard, IOP, and GFM algorithms was performed on a 200 ms Hann-windowed frame located at the centre of each output speech pressure waveform, after downsampling to each of the four sampling rates \( F_s = \{4000, 8000, 12000, 16000\} \) Hz. Thus a total of \( 660 \times 4 = 2640 \) frames were analyzed with each algorithm.

4. Glottal flow evaluation parameters

To evaluate the performance of the inverse filtering algorithms, the mismatch between estimated and reference glottal flow was quantified by four error measures. These four measures were selected because they are straightforward to calculate; two are time-domain and two are frequency-domain measures; and three are well known to quantify aspects of laryngeal voice quality.

The first error measure was \( \Delta_{\text{RMS}} \) as described earlier in Section 2.4: the RMS difference between estimated and reference glottal flow derivatives, after time alignment and energy normalization. This time-domain error is the same one that was used in automatic selection of the three IAIF parameters (\( b, M_s, \) and \( M_f \)).

The second error measure was based on the normalized amplitude quotient (NAQ), which is related to voice quality variation along the breathy-to-pressed continuum (Alku et al., 2002). While NAQ is robustly calculated as a ratio of two prominent amplitude values of the glottal flow and its derivative, divided by the length of the glottal cycle, it is nevertheless a dimensionless time-domain parameter that provides a measure of the relative duration of the glottal closing phase. The relative duration of the closing phase is in turn one of the most important time-domain parameters of the glottal flow which is known to affect, for example, the relative energy of higher harmonics and therefore the spectral tilt (e.g., Gauffin and Sundberg, 1989). For each analysis frame, NAQ was averaged over the values calculated for each glottal pulse within that frame. The error \( \Delta_{\text{NAQ}} \) was then defined as the absolute value of the deviation from the NAQ of the reference glottal flow:

\[
\Delta_{\text{NAQ}} = \left| \frac{NAQ_{\text{est}} - NAQ_{\text{ref}}}{NAQ_{\text{ref}}} \right|.
\]  

The third error measure was based on the harmonic richness factor (HRF), which is a frequency-domain parameter related to voice quality (Childers and Lee, 1991). HRF quantifies the balance of energy at the fundamental in relation to the sum of energies at higher harmonics of the glottal flow magnitude spectrum. The error \( \Delta_{\text{HRF}} \) was then defined as the absolute value of the difference (in dB) between estimated and
5. \[ \delta \Delta (\%) = \frac{\text{change in error} - \text{change in error}}{\text{error in Standard}} \times 100 \] 

The fourth error measure was based on H1 – H2 which is the difference between the amplitudes of the first and second harmonics of the glottal flow spectrum (e.g., Huffman, 1987). The error ΔH1–H2 was defined as the absolute value of the difference (in dB) between H1 and H2 calculated from estimated and reference glottal flows:

\[ \Delta_{H1-H2} = |H_{est} - H_{ref}| \]  

Finally, for any of the four error measures defined above, the relative improvement (reduction) or worsening (increase) of errors by IOP and GFM compared with Standard was calculated as the signed, percent change in error:

\[ \delta \Delta(\%) = 100 \left( \frac{\Delta_{\text{IOP or GFM}} - \Delta_{\text{Standard}}}{\Delta_{\text{Standard}}} \right) \]  

5. Results

Section 5.1 compares the aggregate error distributions broken down by simulation and analysis conditions (vowel, \( F_s, f_o \), and voice quality), Section 5.2 presents specific examples of glottal flow estimation, and Section 5.3 presents the cross-validation results.

5.1. Aggregate errors

The distributions of each of the four types of error are shown in Fig. 5 separately for each vowel. As the distributions were generally found to be asymmetric (skewed to low values and with a long upper tail), they were characterized and compared in terms of their median and quartiles. The box-plots in Fig. 5 indicate distinct results for the back vs nonback vowels: Standard IAIF retained the best performance for the two back vowels (especially /u/), but it was outperformed by at least one, if not both, of the two variant algorithms for the four nonback vowels (especially /a/ and /e/). In the nonback vowels, improved performance was indicated by a lower median error, and in most cases also by lower first and third quartiles. For IOP, this was achieved in all cases except a higher third quartile in \( \Delta_{\text{NAQ}} \) of /a/ and in \( \Delta_{\text{NAQ}} \) and \( \Delta_{\text{RMS}} \) of /a/. For GFM the results were more mixed, and in all cases worse than IOP except for the third quartile in \( \Delta_{\text{NAQ}} \) of /a/ and /e/. Due to the different results for back vs nonback vowels, the errors are broken down next in terms of sampling rate, fundamental frequency, and voice quality, separately for the two groups of vowels.

The results for the nonback vowels are broken down in Fig. 6, which shows the percent change in median error for IOP and GFM, relative to Standard IAIF, averaged over all four error measures; here, a negative bar indicates a reduction in error, i.e., an improvement, compared with Standard. In all conditions, IOP yielded a lower median error than GFM.
Fig. 10. Glottal inverse filtering results for vowel /a/ at \( F_0 = 8 \text{ kHz} \), \( f_o = 92 \text{ Hz} \) and a pressed voice quality (at the 10th step out of 11 along the breathy-to-pressed scale). The reference glottal flow and its energy-normalized harmonic spectrum are shown in solid light grey. Dotted: Standard \((b = 0.997, M_s = 8, M_g = 3)\). Dashed: GFM \((b = 0.999, M_s = 8, M_g = 3)\). Solid black: IOP \((b = 0.999, M_s = 10, M_g = 3)\).

Fig. 11. Glottal inverse filtering results for vowel /o/ at \( F_0 = 8 \text{ kHz} \), \( f_o = 92 \text{ Hz} \) and a breathy voice quality (at the 2nd step out of 11 along the breathy-to-pressed scale). The reference glottal flow and its energy-normalized harmonic spectrum are shown in solid light grey. Dotted: Standard \((b = 0.992, M_s = 8, M_g = 5)\). Dashed: GFM \((b = 0.998, M_s = 12, M_g = 3)\). Solid black: IOP \((b = 0.999, M_s = 12, M_g = 3)\).

(i.e., either a greater reduction, or a smaller increase, in error). The top two panels show that IOP reduced the median error by an amount that was nearly independent of the sampling rate, and for all four nonback vowels (by 31% on average). In contrast, GFM reduced the median error for only /ae/ and /a/ (by 13% on average), and only at the highest three sampling rates. The bottom two panels indicate that IOP's advantage over Standard waned for higher fundamental frequencies and more pressed voice qualities: Standard performed almost the same or better at \( f_o = (262, 311, 440) \text{ Hz} \) (as determined by lower median errors for \( \Delta_{\text{NAQ}} \) and \( \Delta_{\text{HRB}} \), although this breakdown is not shown here), and better at the two most pressed voice qualities (lower median errors for \( \Delta_{\text{NAQ}} \) and \( \Delta_{\text{H1-H2}} \), and for \( \Delta_{\text{HRB}} \) only at the most pressed voice quality). By comparison, GFM showed an advantage over Standard mainly up to \( f_o = 156 \text{ Hz} \), and mainly for voice qualities in the normal range.

For the back vowels, the results in Fig. 7 show that in all conditions the median error increased with IOP and GFM, except for a small reduction at \( f_o = 92 \text{ Hz} \) with IOP \((\Delta_{\text{RMS}} \text{ and } \Delta_{\text{HRF}})\), and at the breathiest voice quality with IOP \((\Delta_{\text{RMS}}, \Delta_{\text{HRF}}, \text{ and } \Delta_{\text{H1-H2}})\) and GFM \((\Delta_{\text{HRB}}, \Delta_{\text{NAQ}} \text{ and } \Delta_{\text{HRB}})\). With both algorithms the largest increase in median error was for /u/ (by 186% on average), for mid to high \( f_o \), and for normal to pressed voice qualities.

Another informative way to compare the three algorithms is to calculate the proportion of frames analyzed for which either Standard, IOP, or GFM gave the lowest errors (i.e., outright best performance). A breakdown of these results for the nonback vowels is shown in Fig. 8, which indicates that in every condition, IOP yielded a larger proportion of best performances than GFM. Standard yielded a larger proportion than IOP in only two conditions: at \( f_o = 262 \text{ Hz} \) and at the most pressed (11th) voice quality. IOP’s advantage over Standard and GFM was strongest for the vowels /a, æ, /, for the low to mid range of \( f_o \), and for the breathy to normal range of voice qualities.

For the back vowels, the results in Fig. 9 show that Standard had the largest proportion of best performances in all conditions, except at \( f_o = 92 \text{ Hz} \) and the breathiest (1st) voice quality where IOP most often performed best.

As mentioned in Section 3.2, an additional set of 110 samples for the neutral vowel were generated with \( f_o \) jitter. Glottal inverse filtering of these samples with all three algorithms resulted in error distributions and proportions of best-performing cases very similar to those obtained by analysis of the same samples without jitter. Therefore our conclusions regarding the relative performance of the three algorithms can be extended to include voice samples with at least a normal (non-pathological) amount of pulse-to-pulse variation in the fundamental period.

5.2. Specific examples

To gain a better sense of the relative performance of the three algorithms, Figs. 10-13 show examples of the reference and estimated
Fig. 12. Glottal inverse filtering results for vowel /i/ at $F_o = 16$ kHz, $f_0 = 311$ Hz and voice quality in the normal range (at the 7th step out of 11 along the breathy-to-pressed scale). The reference glottal flow and its energy-normalized harmonic spectrum are shown in solid light grey. Dashed: Standard ($b = 0.999$, $M_s = 16$, $M_f = 3$). Solid black: IOP ($b = 0.998$, $M_s = 20$, $M_f = 5$).

Fig. 13. Glottal inverse filtering results for vowel /u/ at $F_o = 12$ kHz, $f_0 = 220$ Hz and voice quality in the normal range (at the 7th step out of 11 along the breathy-to-pressed scale). The reference glottal flow and its energy-normalized harmonic spectrum are shown in solid light grey. Dashed: GFM ($b = 0.979$, $M_s = 18$, $M_f = 3$). Solid black: IOP ($b = 0.998$, $M_s = 18$, $M_f = 6$).

glottal flows, their harmonic spectra, and the corresponding vocal tract model transfer functions.

Figs. 10 and 11 show two examples where all three algorithms appear to have captured the shape of the glottal flow waveform in the time domain to a reasonably high accuracy, but where a frequency-domain comparison reveals clearer differences. Fig. 10 shows the results for /a/ with a pressed voice quality, and Fig. 11 the results for /o/ with a breathy voice quality. For the pressed /a/, time-domain errors revealed that Standard ($\Delta_{\text{RMSE}} = 5.19$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 24.7\%$) was outperformed by GFM ($\Delta_{\text{RMSE}} = 1.74$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 7.9\%$) and IOP ($\Delta_{\text{RMSE}} = 1.70$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 10.1\%$); similarly, for the breathy /o/, the already low errors yielded by Standard ($\Delta_{\text{RMSE}} = 1.04$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 4.1\%$) were further improved by GFM ($\Delta_{\text{RMSE}} = 0.28$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 0.3\%$) and IOP ($\Delta_{\text{RMSE}} = 0.20$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 0.4\%$). In the frequency domain, the vocal tract transfer functions in both examples show that the greater amount of total preemphasism applied by GFM and especially IOP, resulted in a better modelling of the second formant and consequently a clearer separation between the first two formant peaks. The frequency-domain errors for the pressed /a/ revealed that Standard ($\Delta_{\text{H1-H2}} = 0.37$ dB, $\Delta_{\text{HRF}} = 1.34$ dB) was outperformed by GFM ($\Delta_{\text{H1-H2}} = 0.37$ dB, $\Delta_{\text{HRF}} = 0.67$ dB) and even more so by IOP ($\Delta_{\text{H1-H2}} = 0.29$ dB, $\Delta_{\text{HRF}} = 0.02$ dB); for the breathy /o/, Standard ($\Delta_{\text{H1-H2}} = 0.37$ dB, $\Delta_{\text{HRF}} = 2.75$ dB) was almost equally outperformed by GFM ($\Delta_{\text{H1-H2}} = 0.02$ dB, $\Delta_{\text{HRF}} = 0.30$ dB) and IOP ($\Delta_{\text{H1-H2}} = 0.03$ dB, $\Delta_{\text{HRF}} = 0.15$ dB).

Fig. 12 shows an example of clear improvements afforded by the two recent variants of IAIIF in both the time and frequency domains. This example was chosen because of all the samples analyzed, with both IOP and GFM the vowel /i/ at $f_0 = 311$ Hz and $F_o = 16$ kHz showed the largest percent improvement over Standard (by 82% for IOP and 78% for GFM, when averaged over the 11 voice qualities and 4 error measures). This is particularly noteworthy, as the aggregate results reported earlier suggested only a modest shift in the median error for nonback vowels at this high fundamental frequency; nevertheless, the evaluation by Ali et al. (2006) had concluded that Standard IAIIF was not recommended for vowels such as /i/ where the first formant is low enough to overlap with the fundamental. The results in Fig. 12 confirm that Standard indeed did not perform well in this case, introducing a major ripple in the time domain ($\Delta_{\text{RMSE}} = 6.78$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 34.7\%$) and underestimating the balance of energy between the fundamental and all higher harmonics ($\Delta_{\text{H1-H2}} = 11.9$ dB, $\Delta_{\text{HRF}} = 13.8$ dB). In contrast, GFM ($\Delta_{\text{RMSE}} = 2.50$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 6.4\%$, $\Delta_{\text{H1-H2}} = 2.9$ dB, $\Delta_{\text{HRF}} = 1.6$ dB) and especially IOP ($\Delta_{\text{RMSE}} = 1.70$ cm$^3$/s$^2$, $\Delta_{\text{NAQ}} = 15.0\%$, $\Delta_{\text{H1-H2}} = 0.2$ dB, $\Delta_{\text{HRF}} = 0.9$ dB) yielded a better match to the reference, with IOP outperforming GFM on all measures except $\Delta_{\text{NAQ}}$. The vocal tract model spectra in the right panel suggest that the improvements were a result of
more accurate modelling of the formants including their overall balance (e.g., a less sharp first formant peak), and the spectral tilt. This example is also interesting in that the noticeable ripple in the open phase of the reference glottal flow, probably caused by interactions with subglottal resonances, was captured to varying degrees by all three algorithms, with IOP producing the closest match.

As an example of one situation where Standard retained best performance, Fig. 13 shows the results for back rounded vowel /u/ at \( f_o \) = 220 Hz and a normal voice quality, analyzed with \( F_s \) = 12 kHz. In the time domain, both IOP and GFM yielded much larger errors especially in the closed phase, and both failed to capture the existing ripple in the open phase. In the frequency domain, Standard yielded reasonably accurate results up to about the 9th harmonic (up to nearly 2 kHz); in contrast, while IOP and GFM gave a more accurate estimate of the overall spectral tilt up to the Nyquist frequency, they both underestimated the amplitudes of harmonics, from the 3rd harmonic through the mid-frequency range. The estimated vocal tract models in the right panel suggest that this error was likely due to overestimation of the vocal tract magnitude response across that frequency range, especially the amplitudes of the second and third formants (at about 1 and 2.5 kHz respectively). This example highlights a limitation of IOP and GFM, both of which attempt to improve on Standard by adjusting mainly the overall spectral tilt prior to vocal tract modelling: in general and among other factors, for accurate glottal inverse filtering not only the overall spectral tilt but also the balance between the levels of different spectral regions, can be of importance.

5.3. Cross-validation

The results presented thus far reflect the best possible performance by each method, due to optimization of the parameters \( b \), \( M_o \), and \( M_f \) for every analyzed frame. On the other hand, the cross-validation procedure described in Section 2.5 allows a slightly more realistic evaluation by assuming that the reference glottal flow is not available, thus replacing optimization with fixed parameter values derived from separate “training” data. Interestingly, the fixed values for each randomized 10% of data – calculated as the median across the remaining 90% of data analyzed by all three methods – turned out to be the same for all 10 folds: as listed in Table 1, the values for \( b \) and \( M_o \) increased with sampling rate, while \( M_f \) was constant.

With all three methods using the fixed (only \( F_s \)-dependent) parameter values listed in Table 1, Fig. 14 shows the breakdown of results for the entire dataset. Compared with the earlier results in Figs. 8 and 9, it is clear that cross-validation with fixed parameter values leads to improvements in the relative performance of GFM in particular. While IOP still performs the best for (vowels /a, æ/), low to mid \( f_o \), breathy phonations)
and Standard still performs the best for {vowels /a/, /o/, some higher  \( f_0 \), pressed phonations}, GFM now performs best for {vowels /i/, /a/, some mid to high  \( f_0 \) normal phonations}.

These results are corroborated by the observation that, compared with optimization, the use of fixed parameters increased the overall median error by 66% for Standard, by 69% for IOP, and by only 21% for GFM. Thus with fixed parameter values, all three methods are competitive and partly complement each other’s performance across different conditions. However, our results also confirm that fixed parameter values do not allow the best possible results (especially for Standard and IOP); therefore, even with natural speech where the reference glottal flow is not available, whenever possible it is recommended to use human expert judgment in order to more optimally select the parameters that yield the “best” glottal flow (e.g., with the least ripple in the closed-phase).

6. Concluding discussion

This study compared the performance of three variants of glottal inverse filtering based on the IAIF method, using output speech signals and input glottal flows provided by computational physical modelling of the vocal tract for six different vowels and a wide range of fundamental frequencies and voice qualities. The results with exhaustive-search optimization to achieve best possible performance indicated that while the Standard algorithm performed generally the best for the back vowels /a/, /o/ and for very high fundamental frequencies and the most pressed voice qualities, IOP gave the best improvement in performance for the nonback vowels /a/, /e/, /i/, /a/, reducing the median error by 31% on average. The results obtained with cross-validation indicated that fixed parameter values can improve the relative performance of GFM, making it the best performer for vowels /i/, /a/ and for the normal range of phonations; nevertheless, these conclusions would likely depend on the exact choice for the fixed parameter values, and may therefore not easily extend to a different set of values applied to natural speech.

The performance improvement afforded by IOP (and to a lesser extent GFM) for breathy to normal voice qualities, is in line with the main effect of these algorithms on overall spectral tilt. As IOP, and to a lesser extent GFM, potentially applies greater preemphasis on the speech signal, it is precisely the voice qualities with a greater spectral tilt (in the breathy to normal range) that would be expected to benefit the most, and this was borne out by the results. Conversely for the most pressed voice qualities with relatively less spectral tilt owing to the stronger presence of higher harmonics, while in some cases IOP and GFM yielded improvements (as in Fig. 12), more often Standard retained better performance owing to details of spectral balance not easily corrected for by tilt alone.

The recent results of Perrotin and McLoughlin (2017) which suggested the superiority of GFM over both Standard and IOP, may appear to be at odds with the results presented here. However, the data used in that study were recordings of natural speech at a high sampling rate  \( F_s = 22,050 \) Hz (necessitating a high-order vocal tract model), analyzed with fixed parameters  \( M_s = 26 \) and  \( b = 0.99 \) across all conditions and algorithms, and with no reference glottal flow available for comparison. In contrast, thanks to computational physical modelling the present study enabled an objective comparison directly with reference glottal flows, and furthermore provided a breakdown of results across a wide range of vowels,  \( f_0 \) values, and phonation types and loudnesses. Nevertheless, it is noteworthy that GFM showed a relative performance gain with cross-validation and fixed parameters. It is therefore possible that Perrotin and McLoughlin’s (2017) conclusions regarding the superiority of GFM may have been influenced by their use of fixed parameter values which, according to our results, may have disproportionately penalized both Standard and IOP.

The use of fixed parameters in IAIF is often necessitated by the large amounts of speech recordings to be analyzed, and by a lack of a definitive and automated measure of the quality of estimated glottal flows, despite some progress in this direction (e.g., Moore and Torres, 2008). Alternatively, if the amount of data to be analyzed is not prohibitive and a certain amount of subjectivity can be tolerated, human expert judgments are sometimes used to select the “best” glottal flow – as enabled for example by the interface of the Aalto Aparat software for IAIF (Alku et al., 2017) or for other glottal inverse filtering methods such as with Glööri (Dalton et al., 2014).

In sum, the results here suggest that glottal inverse filtering with the IAIF method would benefit from a judicious choice of algorithm variant, depending primarily on the type of vowel being analyzed, and secondarily on  \( f_0 \) and voice quality. In general, IOP tends to produce superior results for nonback vowels, not too high fundamental frequencies and not too pressed voice qualities, although with fixed parameter values GFM may be competitive for normal phonations. In other cases, Standard IAIF is still recommended.

An additional but often neglected factor for accurate glottal inverse filtering in general, concerns the phase spectrum (e.g., Degottex et al., 2011). In this regard, while IAIF does not explicitly model nor control the phase spectrum during inverse filtering, our evaluations implicitly included phase information via the time-domain error measures \( \Delta_{\text{UAS}} \) and \( \Delta_{\text{SAO}} \). It may be useful in the future to develop explicit control of the phase of harmonics in IAIF-based methods, since phase modifications introduced by LP modelling and inverse filtering have an impact on the resulting time-domain signals.

It is worth noting that the present study used exclusively conventional LP analysis in all three variants of IAIF. Extension of the present results to IAIF variants using instead discrete all-pole (DAP) modelling (El-Jaroudi and Makhboul, 1991) or any alternative method which may enhance the accuracy of the vocal tract spectral models and consequently the glottal flow yielded by inverse filtering, is left for future work. It would also be of great interest to compare the performance of the IAIF variants with other state-of-the-art glottal inverse filtering algorithms. An implementation of IOP-IAIF will soon be provided in the freely available Aalto Aparat (Alku et al., 2017) software tool.

Declaration of interest

Authors PM and HA have filed an application for a patent on the IOP-IAIF algorithm.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jspcom.2018.09.005.

References


Table 1  
\( F_s \)-dependent, IAIF parameter values used for cross-validation.  

<table>
<thead>
<tr>
<th>( F_s ) (kHz)</th>
<th>( M_s )</th>
<th>( M_o )</th>
<th>( b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>0.991</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>10</td>
<td>0.996</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>14</td>
<td>0.997</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>18</td>
<td>0.998</td>
</tr>
</tbody>
</table>


