Maximum Correntropy Linear Prediction for Voice Inverse Filtering: Theoretical Framework and Practical Implementation

Iván A. Zalazar, Gabriel A. Alzamendi, Matías Zañartu, *Senior Member, IEEE*, and Gastón Schlotthauer

***Abstract*—Voice inverse filtering methods aim at noninvasively estimating the glottal source information from the voice signal. These inverse filtering strategies typically rely on parametric models and variants of linear prediction for tuning the vocal tract filter. Weighted linear prediction schemes have proved to be the best performing for inverse filtering applications. However, the linear prediction and its variants are sensitive to the impulse-like acoustic excitations triggered by the abrupt glottal closure during voiced phonation. The present study examines the maximum correntropy criterion-based linear prediction (MCLP) for voice inverse filtering. Correntropy is a nonlinear, localized similarity measure inherently insensitive to peak-like outliers. Here, a theoretical framework is established for studying the properties of correntropy relevant for voice inverse filtering and for developing an algorithm to estimate vocal tract filter coefficients. The proposed algorithm results in a robust weighted linear prediction, where a correntropy weighting function is adjusted iteratively by a data-driven optimization scheme. The effects of correntropy kernel parameters on the performance of the MCLP method are analyzed. Characterization of the MCLP method for voice inverse filtering is addressed based on synthetic and natural sustained vowel signals. Simulations show that MCLP naturally overweights samples in the glottal closed phase, where the phonation model is more accurate. MCLP does not require prior information about the glottal instants, nor applying a predefined weighting function. Results show that MCLP performs similarly or better than other well-established inverse filtering methods based on weighted linear prediction.**

***Index Terms*—Correntropy, weighted linear prediction, voice inverse filtering, glottal source estimation, closed phase analysis.**

Received 3 April 2024; revised 27 September 2024 and 29 November 2024; accepted 29 November 2024. Date of publication 5 December 2024; date of current version 24 January 2025. This work was supported in part by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) under Grant PIP-CONICET 633, in part by the Ministerio de Ciencia, Tecnología e innovación (MINCyT) under Grant PICT-ANPCYT 2020 Serie A-01865 and Grant PICT-2021-I-INVI-00122, in part by the Universidad Nacional de Entre Ríos (UNER) under Grant PID-UNER 6224 and Grant PID-UNER 6228, in part by the NIDCD of the National Institutes of Health (NIH) under Award P50DC015446, and in part by ANID under Grant BASAL AFB240002. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The associate editor coordinating the review of this article and approving it for publication was Prof. Jong Won Shin. *(Corresponding author: Iván A. Zalazar.)*

Iván A. Zalazar, Gabriel A. Alzamendi, and Gastón Schlotthauer are with the Institute for Research and Development on Bioengineering and Bioinformatics, CONICET-UNER, 3100 Oro Verde, Argentina (e-mail: izalazar@ingenieria.uner.edu.ar; galzamendi@ingenieria.uner.edu.ar; gschlotthauer@ingenieria.uner.edu.ar).

Matías Zañartu is with the Department of Electronic Engineering and the Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, 2390123 Valparaíso, Chile (e-mail: matias.zanartu@usm.cl).

Digital Object Identifier 10.1109/TASLP.2024.3512187

I. Introduction

According to the source-filter theory, human phonation results from the interplay of three simple decoupled components: the glottal airflow source as the acoustic excitation at the glottal level, the vocal tract as an acoustic filter spectrally modulating the different voice sounds, and the lip radiation producing the free-field acoustic pressure [1], [2]. Voice inverse filtering, in turn, addresses the inverse problem of phonation, aiming at noninvasively estimating the acoustic excitation underlying the voiced phonation through digitally processing the voice signal [3]. First, tunable filters are adjusted from the voice signal to match the main vocal tract resonance frequencies and the lip radiation effect [4], [5]. Then, inverse versions of these filters are applied to cancel out the contributions of the supraglottic structures in the voice signal, which results in an estimate of the glottal airflow [3], [6]. Therefore, airflow estimates from voice inverse filtering depend on suitably modeling the vocal tract resonances and radiation at the lips [7].

Inverse filtering typically uses a first-order causal differentiator filter to approximate the lip radiation effect [2], [4]. Vocal tract contribution, in turn, is modeled as an autoregressive filter with transfer function [8]:

$$V(z) = \frac{1}{1 - \sum_{k=1}^{P} a_k z^{-k}}, \tag{1}$$

where P is the filter order, and a_k are the filter coefficients [2]. The baseline standard for optimally computing the vocal tract filter from a voice signal is the linear prediction (LP) method [9]. In the traditional LP scheme, the filter coefficients in (1) are computed by minimizing the mean squared prediction error [2]; theoretically, this is optimal only for zero-mean, white Gaussian prediction errors [10]. However, in the voiced phonation context, LP scheme is suboptimal because successive glottal closure instants (GCIs) trigger high-energy acoustic bursts in the glottal excitation, which in turn give rise to prediction errors with exceptionally high local amplitudes [8]. Additionally, it is well known that the mean squared error function is highly susceptible to outlier data [11]. Therefore, the LP method is prone to provide vocal tract filters featuring biased formant frequencies and inaccurate bandwidths [8].

Variants of classic LP have been proposed and applied to improve the vocal tract filter estimation. Weighted LP strategies

are specifically addressed here since they have demonstrated the best performance in inverse filtering applications [12], [13], [14]. These schemes take advantage of a weighting function conceived upon empirical criteria or prior knowledge to regulate the relative importance of the prediction error samples in the least squares problem [15]. For example, closed phase covariance (CPC) [16] and quasi closed phase linear prediction (QCP) [17] are two strategies inspired in the time-derivative of the glottal airflow, the so-called glottal function. These methods use a weighting function to emphasize the voice samples in the closed phase (relative to those in the open phase), thereby mitigating the detrimental effects in the vocal tract estimation of the GCIs and acoustic coupling between the sub- and supraglottal systems. However, a drawback of these methods is their inherent reliance on the location of the glottal opening and closing instants for the weighting functions implementation. Consequently, any error in the instants' location affects QCP and CPC performance. Vocal tract filter estimates obtained by QCP may exhibit a residual spectral tilt that affects the estimated glottal signals [18]. To address this issue, a QCP-based method with spectral tilt compensation (QCP-ST) has been recently proposed to enhance the glottal signals estimated using QCP by transferring the estimated tilt from the vocal tract filter [13]. Other methods obtain the weighting functions in a data-driven manner, i.e., built upon information extracted from the voice signal under consideration. For example, the weighted linear prediction (WLP) [19] and its stabilized version (SWLP) [20] make use of the short-time energy computed from the voice signal as a weighting function. The energy function yields large amplitude values in the samples with a higher signal-to-noise ratio. Thus, this approach increases the robustness of LP analysis to additive noise. However, the resulting weighting function tends to center around the GCIs, yielding a suboptimal estimation of the vocal tract filter [8]. More recently, the probabilistic weighted linear prediction (PWLP) scheme [12] has been developed, which employs a data-driven adaptive function that automatically centers the linear prediction analysis on the closed phase. PWLP employs a probabilistic interpretation of the weighted LP scheme based on the likelihood of voice signal and different model parameters priors, and Gibbs sampling is applied to generate estimates of the LP coefficients from the posterior distribution. This method has the disadvantage of relying on different prior probability distributions for the model parameters, which makes the physiological interpretation challenging.

All weighted LP strategies seek to overcome the inherent problems of the mean squared prediction error function. Instead, LP schemes based on alternative cost functions with well-documented properties have not received equivalent attention. For example, as the mean absolute error function is less sensitive to large-amplitude errors, it becomes an appealing cost function for developing a sparse linear prediction model for speech processing [10]. In turn, the discrete all-pole modeling applies the Itakura-Saito distortion measure to better fit the vocal tract filter in (1) to a small set of spectral points [21]. Here, we study the correntropy with Gaussian kernel as a cost function to extend the LP scheme.

Correntropy with Gaussian kernel is a robust, nonlinear similarity measure that proved to be a suitable cost function for linear models because of its robustness to non-Gaussian impulse-like prediction errors [22]. Recently, in a conference paper [23], we took advantage of this alternative cost function to develop the maximum correntropy-based LP (MCLP), a data-driven weighted LP scheme that iteratively adjusts both the weighting function and the vocal tract filter coefficients. We showed that MCLP is capable of overweighting the closed phase samples without requiring prior knowledge of glottal instants. Correntropy has been used for processing voice signals (e.g., for fundamental frequency estimation [24], speech enhancement [25], and speech recognition [26]); however, only in [23] has been applied in the context of voice inverse filtering, to the best of our knowledge. The present work extends the ideas in [23] in the following aspects. First, the attention is paid to the theoretical properties of the correntropy measure relevant to the LP model and for voice inverse filtering. This sets up a framework to study the MCLP method in detail, particularly the effects of Gaussian kernel parameters on the local weighting imposed on LP prediction errors. Second, a computational algorithm is proposed for adjusting the vocal tract filter coefficients based on the step-like update of the correntropy kernel parameter. Third, performance of MCLP for voice inverse filtering is thoroughly evaluated in synthetic and natural phonation data using different evaluation metrics. Analysis of the estimated vocal tract filters and glottal waveforms for synthetic signals indicates that MCLP outperforms QCP and QCP-ST, and achieves a comparable performance to PWLP at a significantly reduced computational time. Results for natural signals show that MCLP yields glottal waveforms characterized by a flatter closed phase with fewer spurious oscillations associated with inverse filtering errors.

The organization of the paper is as follows. Section II analyzes the correntropy measure with Gaussian kernel, and addresses the MCLP method. Section III describes the voice inverse filtering data and the experimental setup. Section IV provides a detailed analysis of the proposed method for voice inverse filtering. Section V presents the results, whereas Section VI discusses the results and future works. Finally, we deliver the conclusions of this work in Section VII.

II. Maximum Correntropy Criterion for Linear Prediction

A. Correntropy With Gaussian Kernel

Correntropy with Gaussian kernel is a nonlinear similarity measure between two arbitrary scalar random variables X and Y, given by [27]:

$$V(X,Y) = \mathcal{E}\{G_\sigma(X-Y)\}, \tag{2}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator, and $G_\sigma(\cdot)$ is the Gaussian kernel with kernel size σ defined as:

$$G_\sigma(X-Y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(X-Y)^2}{2\sigma^2}\right). \tag{3}$$

Kernel G_σ is positive-definite and symmetric about the origin. An alternative expression of (2) can be obtained using the finite-sample estimator of the expectation operator [22] :

$$\hat{V}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(x_i - y_i). \tag{4}$$

where N is the number of samples available.

Correntropy assesses the probability of how similar two random variables are with respect to a neighborhood of the joint space determined by the kernel size σ [22]. This proves to be very useful in reducing the detrimental effects of outliers from gross differences between X and Y [28]. Correntropy is endowed with a strong theoretical framework; here, we present the most relevant properties in the context of linear prediction [27], [29]:

Property 1: Correntropy is positive-definite and bounded, i.e., $0 < V(X,Y) \leq 1/(\sqrt{2\pi}\sigma)$. Additionally, it reaches its maximum value if $X = Y$. This property guarantees the existence of an optimal solution for the LP model.

Property 2: Correntropy involves all the even moments of the difference between X and Y:

$$V(X,Y) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{n=0}^{\infty}\frac{(-1)^n}{2^n n!}\mathcal{E}\left\{\frac{(X-Y)^{2n}}{\sigma^{2n}}\right\}. \tag{5}$$

Compared to the mean squared error (second-order statistical moment) used in LP and its variants, correntropy includes statistical information on second and higher-order moments. Notice that, as σ increases, the high-order moments decay faster and the second-order moment dominates, thus approaching to the mean squared error.

Property 3: Assume an i.i.d. data sample $\{(x_i, y_i)\}_{i\in N}$ drawn from the joint probability density function $f_{X,Y}(x,y)$. The Parzen estimator with kernel size σ of the probability density function of the error samples $e_i = x_i - y_i$ is defined as $\hat{f}_\sigma(e)$. Then, the correntropy measure $\hat{V}(X,Y)$ is the value of $\hat{f}_\sigma(e)$ evaluated at the point $e = 0$, i.e., $\hat{V}(X,Y) = \hat{f}_\sigma(0)$ [22]. Since $f_\sigma(0) = p(X = Y)$ [30], maximizing the correntropy in the context of LP leads to an increase in the probability that the predicted and desired signals are equal.

Property 4: Express correntropy of X and Y as a cost function:

$$J = V(X,Y), \tag{6}$$

J is concave in the range of $[-\sigma, \sigma]$. The concavity guarantees the existence and uniqueness of the optimal solution in problems built around maximizing the correntropy cost function [29].

Property 5: Correntropy, as a sample estimator, induces a metric in the sample space with important geometric characteristics. As explained in [22], the induced metric changes progressively from L_2 norm-like in the Euclidean zone (where samples X and Y are close) to L_1 norm-like in the transition zone away from the Euclidean zone to eventually approaching L_0 norm and becoming insensitive to distance in the rectification zone (where two points are further apart). This explains the inherent robustness of the correntropy measure against outlier samples. Furthermore, the kernel bandwidth σ controls the size of each zone [22]. Then, increasing σ produces a large Euclidean zone and reduces the rectification zone, whereas decreasing σ has the opposite effect.

B. Linear Prediction Based on Maximum Correntropy Criterion

Correntropy has proved to be a suitable cost function for information theoretic learning, establishing the foundations for the maximum correntropy criterion in estimation-related problems [29], [31]. In the same vein, we proposed in [23] the maximum correntropy criterion-based LP (MCLP) method. In this section, we describe in detail the derivation of MCLP based on the correntropy properties discussed previously.

According to the LP model of phonation, the voice signal, $s[n]$, for index time $1 \leq n \leq N$, follows an autoregressive process [2]:

$$s[n] = \mathbf{a}^T\mathbf{s}[n] + e[n], \tag{7}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_P]^T$ are the LP coefficients, with P being the model order, $\mathbf{s}[n] = [s[n-1], s[n-2], \ldots, s[n-P]]^T$ is a vector gathering the P last past voice samples, and $e[n]$ is the prediction error. In classic LP, vector $\mathbf{a}$ is typically obtained by minimizing the mean squared prediction error in (7) [9]. Instead, as proposed in [23], we consider correntropy as a cost function of the prediction error, e, in the computation of the coefficients $\mathbf{a}$ of (7). Based on properties 1, 3, and 4 mentioned above, the optimal coefficients $\mathbf{a}$ can be determined by maximizing the correntropy with Gaussian kernel of the difference between the voice signal and the LP model in (7):

$$\begin{aligned} J &= \mathcal{E}\{G_\sigma(e[n])\} \\ &= \mathcal{E}\left\{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(s[n]-\mathbf{a}^T\mathbf{s}[n]\right)^2}{2\sigma^2}\right)\right\}. \end{aligned} \tag{8}$$

This method for computing the LP coefficients is called maximum correntropy criterion based-LP (MCLP). Maximum of (8) is obtained by setting $\frac{\partial J}{\partial \mathbf{a}} = 0$, which yields:

$$\mathcal{E}\left\{\exp\left(-\frac{\left(s[n]-\mathbf{a}^T\mathbf{s}[n]\right)^2}{2\sigma^2}\right)\left(s[n]-\mathbf{a}^T\mathbf{s}[n]\right)\mathbf{s}[n]\right\} = 0. \tag{9}$$

Considering property 3 and assuming that the prediction error is a discrete-time stationary stochastic process, then we can use in (9) the sample estimator of expectation operator [32], obtaining:

$$\frac{1}{N}\sum_{n=1}^{N} h_e[n]\left(s[n]-\mathbf{a}^T\mathbf{s}[n]\right)\mathbf{s}[n] = 0, \tag{10}$$

where, according to the weighted LP scheme, $h_e[n]$ takes the form of a positive weighting function:

$$h_e[n] = \exp\left(-\frac{\left(s[n]-\mathbf{a}^T\mathbf{s}[n]\right)^2}{2\sigma^2}\right). \tag{11}$$

After some manipulation in (10), coefficient vector $\mathbf{a}$ becomes:

$$\mathbf{a} = \left[\sum_{n=1}^{N} h_e[n]\mathbf{s}[n]\mathbf{s}[n]^T\right]^{-1} \left[\sum_{n=1}^{N} h_e[n]s[n]\mathbf{s}[n]\right] = \mathbf{R}_h^{-1}\mathbf{r}_h, \tag{12}$$

where $\mathbf{r}_h$ and $\mathbf{R}_h$ represent the weighted estimates of the correlation vector and the correlation matrix, respectively. As can be seen, equation (12) is in the form of the *Wiener-Hopf* solution; two key differences are worth noting, though [28]:

- A time-domain weighting function, $h_e[n]$, is applied from the Gaussian kernel with size σ. By $h_e[n]$, small errors with respect to σ are emphasized, whereas large amplitude errors, i.e., significantly higher than σ, are underweighted (as explained in property 5). Thus, the computation of coefficient vector $\mathbf{a}$ from (8) is robust against impulse-like prediction errors.
- Equation (12) is not a closed-form solution. Computation of the $\mathbf{a}$ vector involves the weighting function $h_e[n]$, which in turn depends on coefficients $\mathbf{a}$ according to (11).

Following the work by Singh and Principe [28], an iterative method is used based on assuming that the optimal solution $\mathbf{a}^*$ is a fixed point of (12). Thus, given an initial coefficient vector, $\mathbf{a}_0$, the solution is obtained through the following fixed-point iteration [28]:

$$\mathbf{a}_{k+1} = \left[\mathbf{R}_h(\mathbf{a}_k)\right]^{-1} \mathbf{r}_h(\mathbf{a}_k), \tag{13}$$

where $\mathbf{a}_k$ denotes the solution for iteration $k = 0, 1, 2, \ldots$. A condition guaranteeing the convergence of this fixed-point solution was provided by property 5.

Algorithm 1 describes the computation of fixed-point solution in (13). First, the coefficients $\mathbf{a}_0$ are initialized by classic LP. On the other hand, the kernel size in (11) is a free parameter that must be chosen by the user using concepts of density estimation [22]. In our case, σ is determined from the prediction error using Silverman's rule [33]:

$$\sigma_S = 1.06\sigma_e N^{-1/5}, \tag{14}$$

where σ_e stands for the minimum between standard deviation and interquartile range scaled by 1.34 extracted from prediction error computed by (7) using the coefficients $\mathbf{a}_0$. Thresholds, ϵ_1 and ϵ_2, are defined so that ϵ_1 is much greater than ϵ_2. Finally, a pre-emphasis filter is applied to the voice signal, s, and the resulting signal is then normalized to its maximum absolute value. In the second stage of Algorithm 1, the LP coefficients $\mathbf{a}_k$ are computed iteratively according to (13). If the difference between two consecutive solutions is less than the threshold ϵ_1, the kernel size σ is then updated using Silverman's rule in (14) from the last estimate of the prediction error. Steplike update of σ improves the vocal tract filter coefficients estimation, compared to updating σ at each iteration or keeping it constant. Finally, the computation stops when the difference between two consecutive solutions is less than ϵ_2.

Algorithm 1: Computation of LP Coefficient Vector $\mathbf{a}$ Based on MCLP.

Initialization: $\mathbf{a}_0$, σ, ϵ_1, ϵ_2
Pre-emphasis and normalization of $s[n]$
$k = 0$
do
  Compute: $h_e[n]$, $\mathbf{R}_h$, $\mathbf{r}_h$
  $\mathbf{a}_{k+1} = \mathbf{R}_h^{-1}\mathbf{r}_h$
  $e_{k+1}[n] = s[n] - \mathbf{a}_{k+1}^T\mathbf{s}[n]$
  if $\left(||\mathbf{a}_{k+1} - \mathbf{a}_k||_2^2 < \epsilon_1\right)$ **then**
    Compute: σ_S
  end
  $k = k + 1$
while $\left(||\mathbf{a}_{k+1} - \mathbf{a}_k||_2^2 > \epsilon_2\right)$;
return $(\mathbf{a}_{k+1})$

III. Materials and Methods

A. Phonation Data for Inverse Filtering Evaluation

The OPENGLOT database [34], a comprehensive data set on glottal excitation during phonation, was considered to benchmark the inverse filtering methods. In this work, only repositories II and IV were considered.

Repository II contains synthetic signals obtained from a physical model of phonation that includes vocal fold kinematics, aero-acoustic interactions at the glottal level, pressure wave propagation through the vocal tract, and voice sound radiation at the lips. This repository of synthetic or theoretical signals includes the voice signal (radiated pressure) s, the glottal airflow (volume velocity) u_g, and its time-derivative (volume acceleration) v_g. The latter will be subsequently referred to as the "glottal function". Therefore, a direct evaluation of voice inverse filtering methods is possible for these signals. The phonation data correspond to three vowels: [/a/, /i/, /u/], with four fundamental frequency cases: 82, 110, 156, and 220 Hz for male speakers, and 175, 196, 220, and 294 Hz for female speakers. The sampling frequency of the signals is $f_s = 44.1$ kHz.

In contrast, repository IV contains voice and electroglottogram signals from the natural production of vowel sounds at modal and breathy phonation qualities. Additionally, this repository contains high-speed video of the vocal folds. However, inverse filtering methods only consider voice signals, so this information is not taken into account in the present work. The signals correspond to five male and five female speakers that produce a vowel sound with three different pitch levels (low, medium, and high). The sampling frequency of the signals is $f_s = 44.1$ kHz. Glottal airflow and vocal tract information is not provided in this repository. So, a direct comparison with a ground truth signal is impossible.

B. Methodology and Reference Methods

Simulations were conducted to assess the performance of the MCLP method for voice inverse filtering based on different descriptors for the estimated glottal airflow and its time derivative.

Three voice inverse filtering methods were considered for comparison: PWLP [12], QCP [17], and QCP-ST [13]. QCP-based approaches have become benchmarks for voice inverse filtering methods due to its good performance; as recent studies [13], [14] have shown that other methods in the literature, such as LP, WLP, and SWLP, exhibit low performance compared to QCP methods, they are thus excluded from the present study. On the other hand, PWLP is a suitable inverse filtering method requiring no prior detection of glottal instants, since the weighting function is obtained in a fully data-driven manner similar to MCLP.

Following the hypothesis supporting the source-filter theory [1], all signals used in the simulations were resampled at 8 kHz. Inverse filtering analysis was performed for every 50 ms non-overlapping segment of the voice signal. A pre-emphasis filtering, with transfer function $R(z) = 1 - 0.99z^{-1}$, was applied to enhance the higher frequencies on the voice signal before computing the vocal tract filter with each voice inverse filtering method.

Vocal tract filters computed through MCLP, QCP, QCP-ST, and PWLP were processed to remove the poles with positive real parts, and magnitude less than 0.8 before applying it for voice inverse filtering. To correctly represent the vocal tract formant frequencies, the poles of the transfer function (see (1)) in the z-plane must be close to the unit circle [1]. Poles that do not satisfy this condition would provide poor inverse filtering estimates [16].

C. Performance Measures

Direct evaluation of voice inverse filter estimates is only possible for synthetic data from repository II. As a direct performance metric, a normalized waveform error of the glottal function is defined as:

$$E_{v_g} = \frac{m_e}{\text{RMS}(v_g)}, \tag{15}$$

where m_e denotes the median value of the absolute waveform error between the theoretical glottal function, v_g, and its voice inverse filtering estimate, $\hat{v}_g$, given by:

$$e_{v_g}[n] = \left|v_g[n] - \hat{v}_g[n]\right|, \text{ for } n = 1, 2, \ldots, N. \tag{16}$$

Normalization by the RMS value of the theoretical glottal function is applied. The signals in (16) are time-aligned to compensate for any phase delay due to the acoustic wave propagation along the vocal tract, or the applied inverse filtering method. Furthermore, the voice inverse filtering estimates $\hat{v}_g$ are normalized by the orthogonal projection proposed in [35].

Another metric used to evaluate the performance is the l_1 norm, i.e., the sum of the absolute values of the estimated glottal airflow samples on the closed phase. This measure quantifies the waveform flatness in the closed phase, which is a looked-for feature in the estimated glottal airflow, $\hat{u}_g$ [7]; a large value of l_1 norm indicates the presence of spurious oscillations or an incomplete closed phase, resulting from inadequate inverse filtering [36].

In addition to the normalized waveform error E_{v_g} and the l_1 norm, the inverse filtering evaluation of the MCLP method was based on five aerodynamic parameters computed from the estimated glottal airflow and its time-derivative: Normalized Amplitude Quotient (NAQ), Quasi-Open Quotient (QOQ), Closing Quotient (ClQ), the difference between first two harmonic (H1H2), and the Spectral tilt (ST)[1] [4], [5], [13]. Estimation errors for NAQ, QOQ, and ClQ are reported as average absolute relative errors. For example, for the NAQ parameter:

$$\text{Error NAQ} = \mathcal{E}\left\{\frac{\left|\text{NAQ}_{\text{ref}} - \text{NAQ}_{\text{est}}\right|}{\text{NAQ}_{\text{ref}}}\right\}, \tag{17}$$

where NAQ_{ref} and NAQ_{est} are the NAQ values from the theoretical glottal airflow and the inverse filtered estimate, respectively. Errors for the parameters QOQ, and ClQ are computed similarly to (17). On the other hand, the error for the parameters H1H2 and ST are reported through the average absolute errors:

$$\text{Error H1H2} = \mathcal{E}\left\{\left|\text{H1H2}_{\text{ref}} - \text{H1H2}_{\text{est}}\right|\right\}, \tag{18}$$

$$\text{Error ST} = \mathcal{E}\left\{\left|\text{ST}_{\text{ref}} - \text{ST}_{\text{est}}\right|\right\}. \tag{19}$$

For the natural signals in repository IV, the MCLP method is assessed similarly as in [12], [36]. A first evaluation measures the l_1 norm on the closed phase of the estimated glottal airflow. Another measure to evaluate the MCLP method for the natural data is the NAQ parameter, where the results from QCP were considered the benchmark values. As QCP has proved to be one of the best-suited voice inverse filtering methods [37], we assume that any methods producing similar results as those obtained by QCP could be considered a befitting choice for analyzing natural voice signals.

D. Experimental Setup

Model order P and the threshold ϵ_1 in Algorithm 1 were explored in the simulations involving the repository II. The following sets were considered: $P \in (8, 9, \ldots, 12)$, and $\epsilon_1 \in (0.000001, 0.00001, \ldots, 0.1, 1)$. To fulfill the constraint $\epsilon_1 >> \epsilon_2$ we set $\epsilon_2 = 0.1\epsilon_1$. From P, ϵ_1, and ϵ_2, the vocal tract filter coefficients were computed using Algorithm 1, and the estimated glottal function $\hat{v}_g$ was obtained through inverse filtering. Finally, the parameter combination yielding the lowest average normalized waveform error, E_{v_g}, for each signal was selected.

For QCP and PWLP methods, its parameters are set up based on the values suggested in [12], [17]. Spectral tilt compensation for QCP-ST was performed as described in [13]. Additionally, QCP methods requires prior knowledge about the GCI locations; here, two different strategies are employed. For the synthetic signals in repository II, where the theoretical glottal data are available, the GCIs were determined from the location of the minimum value in the glottal function, v_g, for each glottal cycle. Before determining the GCIs, the voice signal and the glottal function were time-aligned to avoid any phase lag. On the other hand, for the natural voices, the GCIs and the closed phase were estimated from the electroglottogram signal using the SIGMA algorithm [38]. Time alignment between the voice and electroglottogram signals was performed to avoid any delay generated during the recording of the signals. For all cases, lag compensation was computed via cross-correlation measure [2].

[1] Computed following the code in [13].

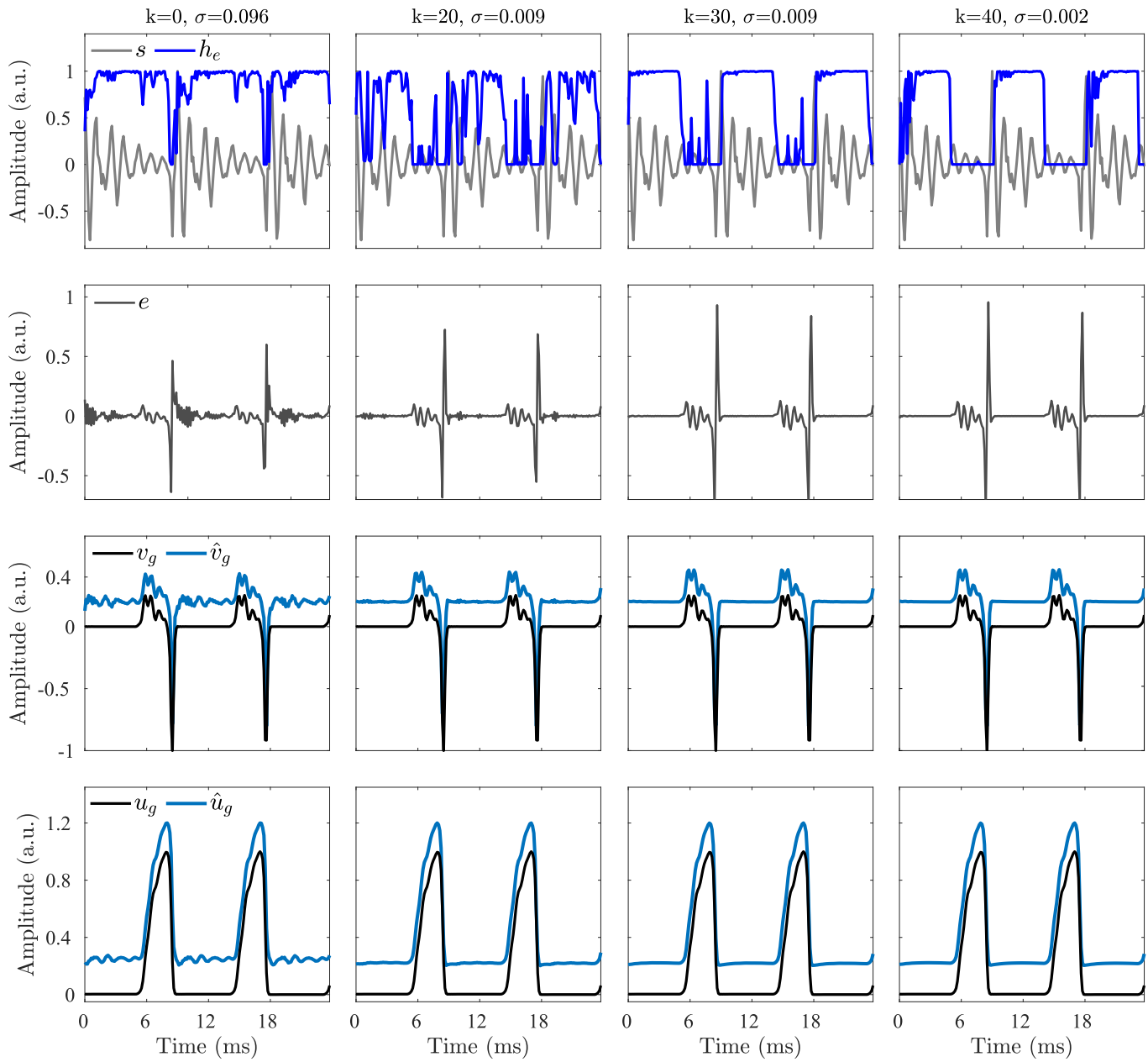

Fig. 1. Iterative adjustment of MCLP in a synthetic signal example for different iterations k of the Algorithm 1. First row: voice signal s and resulting weighting function h_e. Second row: prediction error e. Third row: estimated glottal function $\hat{v}_g$ together with the theoretical waveform v_g. Fourth row: theoretical glottal airflow u_g and its resulting estimation $\hat{u}_g$. For better visualization, the estimated $\hat{v}_g$ and $\hat{u}_g$ are shifted vertically in the third and fourth rows, respectively.

IV. MCLP Analysis

This section discusses some interesting features of the MCLP method based on the simulations for the synthetic signals from repository II.

A. Iterative Adjustment of MCLP

In MCLP, the linear prediction coefficients are iteratively computed based on the prediction error, e, and the resulting weighting function h_e (see (12)). Fig. 1 illustrates the iterative adjustment of MCLP variables for a synthesized signal from the repository II. Columns in Fig. 1 represent different iterations k of Algorithm 1 for a kernel size σ computed by Silverman's rule in (14). The first row of Fig. 1 displays the voice signal, s, and the weighting function, h_e. The second row shows the prediction error. Finally, the third and fourth rows of Fig. 1 show the theoretical waveform of the glottal airflow, u_g, and its time-derivative, v_g, together with the estimates, $\hat{u}_g$ and $\hat{v}_g$, obtained by using the vocal tract filter coefficients computed in each iteration k. Note that all estimated signals are shifted vertically for better visualization.

The proposed MCLP method becomes an iterative weighted LP scheme that performs a data-driven closed-phase voice inverse filtering analysis over multiple glottal cycles. As shown in Fig. 1, the prediction error becomes sparser during the iterations, i.e., evident local error diminutions are observed, especially

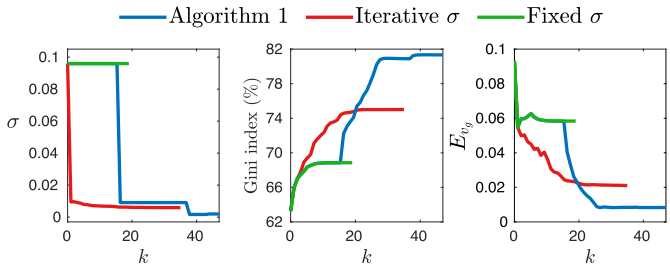

Fig. 2. Analysis of different σ update strategies in a synthetic signal from repository II. Left: σ values for each iteration. Middle: Gini index of the prediction error. Right: normalized waveform error of the glottal function.

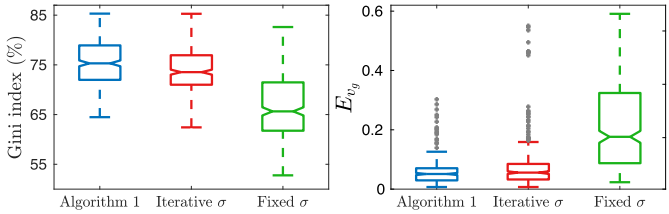

Fig. 3. Boxplots for the three alternatives for σ update obtained from the voice signals in the repository II. Left: Gini index for the prediction error. Right: normalized waveform errors E_{v_g}.

in the closed phase segments, while releasing the spike-like components characteristic of the maximum glottal excitations. The changes in the prediction error are the results of updating the kernel size σ, the weighting function h_e, and the LP coefficients $\mathbf{a}$. Algorithm 1 applies a steplike update of kernel size based on prediction errors, where the increases in the sparse level of the prediction error produce a diminution in the σ value computed by (14). In addition, h_e overweights the glottal closed phase, i.e., open-phase samples with high amplitude prediction error progressively receive significantly lesser weights than the closed-phase ones (see Fig. 1). So, the iterative adjustment of the vocal tract filter becomes increasingly meaningful, as it is based mostly on information from the closed phase, where the effect of the maximum glottal excitation is minimal [17]. The iterative adjustment of the vocal tract filter leads to more accurate estimates of the glottal airflow and its time-derivative obtained by voice inverse filtering. The third and fourth rows in Fig. 1 illustrate the progressive improvement in the estimates $\hat{v}_g$ and $\hat{u}_g$ across the successive iterations, compared to the theoretical waveforms.

B. Correntropy Kernel Size Update

The update of the kernel size σ is a worth analyzing factor affecting the performance of Algorithm 1. Fig. 2 describes three different strategies for updating σ from the voice signal of Fig. 1. The analyzed strategies are: initialize σ using Silverman's rule, perform no update and keep it fixed (in green color), update σ in each iteration by Silverman's rule (in red color), and steplike update σ as proposed in Algorithm 1 (in blue color). Fig. 2 shows the σ value (left panel), the Gini index of the prediction error (middle panel), and the normalized waveform error of the glottal function E_{v_g} (right panel) for successive iterations k. The Gini index allows quantifying the sparsity level of the prediction error [11], [39]. As explained above, a correct estimation of the coefficients $\mathbf{a}$ yields sparse prediction errors and, consequently, a high Gini index. On the contrary, a low Gini index indicates less sparse prediction errors resulting from an imprecise adjustment of the LP model.

As shown in Fig. 2, for fixed σ, the computation of the MCLP model converges faster (fewer iterations are required); however, it results in the lowest Gini index, indicating non-sparse prediction errors, and in the highest waveform error E_{v_g}, characteristic of a deficient voice inverse filtering. In the case of updating the kernel size in each iteration, a significant decrease in σ takes place, and a better fitting of the MCLP model results, as suggested by a higher Gini index and a lower waveform error, at the expense of requiring more iterations until convergence. However, the best results (showing the topmost Gini index and minimum waveform error) are obtained by the steplike update of σ, as in Algorithm 1. This strategy seeks to iteratively tune the LP model leveraged by a smooth-changing weighting function that gradually adjusts to the glottal closed phase (see Fig. 1), thus allowing a closed phase weighted linear prediction analysis that improves the voice inverse filtering estimations.

To further study the update of the kernel size, MCLP schemes considering the three update strategies described above were applied for processing all voice signals from repository II, and the performance of the resulting LP models was assessed. Fig. 3 depicts boxplots for the Gini index of the prediction error (left panel) and the normalized waveform error E_{v_g} (right panel). As expected, a deficient model adjustment (e.g., smallest Gini index and highest waveform error overall) results from keeping kernel size fixed. However, when updating σ the steplike strategy of Algorithm 1 yields a significant improvement (e.g., elevated Gini index and reduced estimation error) for LP model tuning. Statistical analysis based on the Friedman test with the Bonferroni correction shows that the Gini indices are different at a significance level greater than 95%. Similarly, normalized waveform errors are also significantly different for all three update strategies.

Based on our simulations for the synthetic signals from repository II, the best parameter combination for Algorithm 1 is $P = 12$, $\epsilon_1 = 0.0001$, and $\epsilon_2 = 0.00001$.

V. Results

This section studies the performance of the MCLP method for inverse filtering the voice signals in repositories II and IV. MCLP is compared against QCP, QCP-ST, and PWLP. All voice signals were processed considering $P = 12$ for the vocal tract filters. For the MCLP method, the thresholds for Algorithm 1 were set to $\epsilon_1 = 0.0001$ and $\epsilon_2 = 0.00001$.

A. Synthetic Voice Data

The performance for estimating the glottal function in repository II is addressed. The top row in Fig. 4 shows boxplots for the normalized waveform error E_{v_g} for different vowels and fundamental frequency ranges (Low: $f_0 < 200$ Hz and High: $f_0 \geq 200$ Hz). As can be seen, the data-driven methods, PWLP and MCLP, presented the lowest waveform errors, whereas

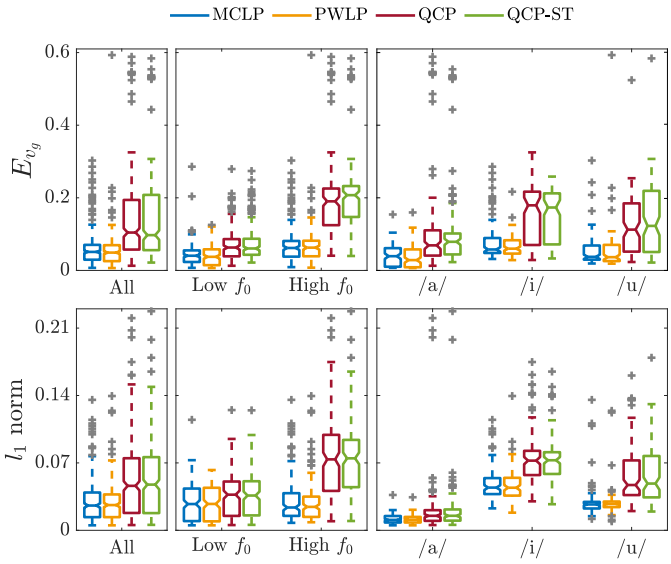

Fig. 4. Boxplots of the normalized waveform error E_{v_g} (top) and l_1 norm on closed phase of the estimated glottal airflow (bottom) for synthetic signals from repository II. Left: all signals. Middle: high and low f_0 ranges. Right: vowel sounds.

the QCP-based approaches showed the highest errors for all categories considered. The error values showed statistically significant differences greater than 95% between the MCLP and PWLP methods compared to the QCP-based approaches, while no significant differences were found between QCP and QCP-ST error values. However, for vowels /i/ and /u/, and for the low f_0 range, no statistically significant differences were obtained between the MCLP and PWLP errors. Likewise, the bottom row in Fig. 4 shows boxplots for the l_1 norm on closed phase of the estimated glottal airflow for the same categories. As can be seen, MCLP and PWLP exhibited low l_1 norm values for all categories. According to Friedman test with the Bonferroni correction, statistically significant differences were obtained between l_1 norms from QCP-based approaches and those from MCLP and PWLP for all categories.

Fig. 5 shows bar plots reporting the estimation errors for parameters NAQ, QOQ, ClQ, H1H2, and ST for all signals from repository II (top row). Additionally, separate bar plots are presented according to the f_0 ranges (middle row) and the vowel sounds (bottom row) to provide a more detailed description. As can be seen in the top row, MCLP and PWLP methods show superior performance for all parameters in contrast to QCP-based approaches. In particular, no statistically significant differences were obtained between MCLP and PWLP errors for the time-domain parameters, QOQ and ClQ, and the frequency-domain parameters, H1H2 and ST. Instead, PWLP yields lower NAQ errors than MCLP, QCP, and QCP-ST. Results for f_0 ranges indicate that MCLP shows similar performance to PWLP. However, MCLP showed inferior performance for the NAQ parameter in high f_0 cases. These results coincide with those reported in [12], where PWLP shows a lower NAQ error than QCP for physical model-based synthetic signals with high f_0. Finally, similar results were observed for vowel analysis, where

TABLE I
COMPUTATIONAL TIME DISPERSION FOR ADJUSTING VOCAL TRACT FILTER COEFFICIENTS FROM VOICE SIGNALS IN REPOSITORY II

	Q_1	Q_2	Q_3
QCP	$2.48\text{x}10^{-4}$	$2.71\text{x}10^{-4}$	$2.88\text{x}10^{-4}$
QCP-ST	$7.02\text{x}10^{-4}$	$7.28\text{x}10^{-4}$	$7.68\text{x}10^{-4}$
MCLP	$1.68\text{x}10^{-2}$	$2.27\text{x}10^{-2}$	$3.13\text{x}10^{-2}$
PWLP	$1.29\text{x}10^{0}$	$1.42\text{x}10^{0}$	$1.51\text{x}10^{0}$

Time is reported in seconds for the lower quartile (Q_1), median (Q_2), and upper quartile (Q_3).

the MCLP method obtained comparable results to PWLP, except for the NAQ parameter estimation.

The computational burden of each inverse filtering method was also analyzed based on the time taken to compute the vocal tract filter coefficients from voice signal segments from repository II. Table I reports the lower quartile (Q_1), median (Q_2), and upper quartile (Q_3) in seconds describing the dispersion of the computational times for QCP, QCP-ST, PWLP, and MCLP methods.[2] Simulations show that MCLP requires, on average, a 60 times less computational time than PWLP. This differences are due to PWLP being a computationally expensive method compared to MCLP. Table I also shows that the QCP method requires the lowest computational time, as expected, since it applies a simple, closed-form rule for computing vocal tract filter coefficients. Instead, QCP-ST demands more computations than QCP because of the processing of the spectral tilt in the estimated vocal tract filter.

B. Natural Voice Data

The results for natural voice signals from repository IV are described in this section. The first column in Fig. 6 shows bar plots for the average l_1 norm on the closed phase of the estimated glottal airflow for four voice inverse filtering methods. As can be seen, MCLP showed the minimum l_1 norm values, where statistically significant differences with l_1 norms from PWLP and QCP-based methods were found. On the other hand, the second column in Fig. 6 depicts bar plots for the average NAQ values obtained from the estimated glottal airflow waveform. No statistically significant differences were found in the estimated NAQ values for the four inverse filtering methods. These findings are similar to those reported in [12] for natural voices, where PWLP demonstrated superior performance over QCP for the l_1 norm while achieving similar NAQ values.

Fig. 7 shows examples of estimated waveforms for the glottal function (top rows) and the glottal airflow (bottom rows) obtained from a modal and a breathy natural voice of repository IV using four inverse filtering methods. All curves are shifted vertically for better visualization. We can observe that the estimates obtained by all methods present small spurious oscillations in the closed phase, which indicates suitably inverse filtering [17]. The results in Fig. 7 indicate that MCLP can provide glottal airflow and glottal function estimates comparable

[2]Computational times for all considered methods were measured in the same desktop computer with a Core i7-12700, and 32GB of RAM.

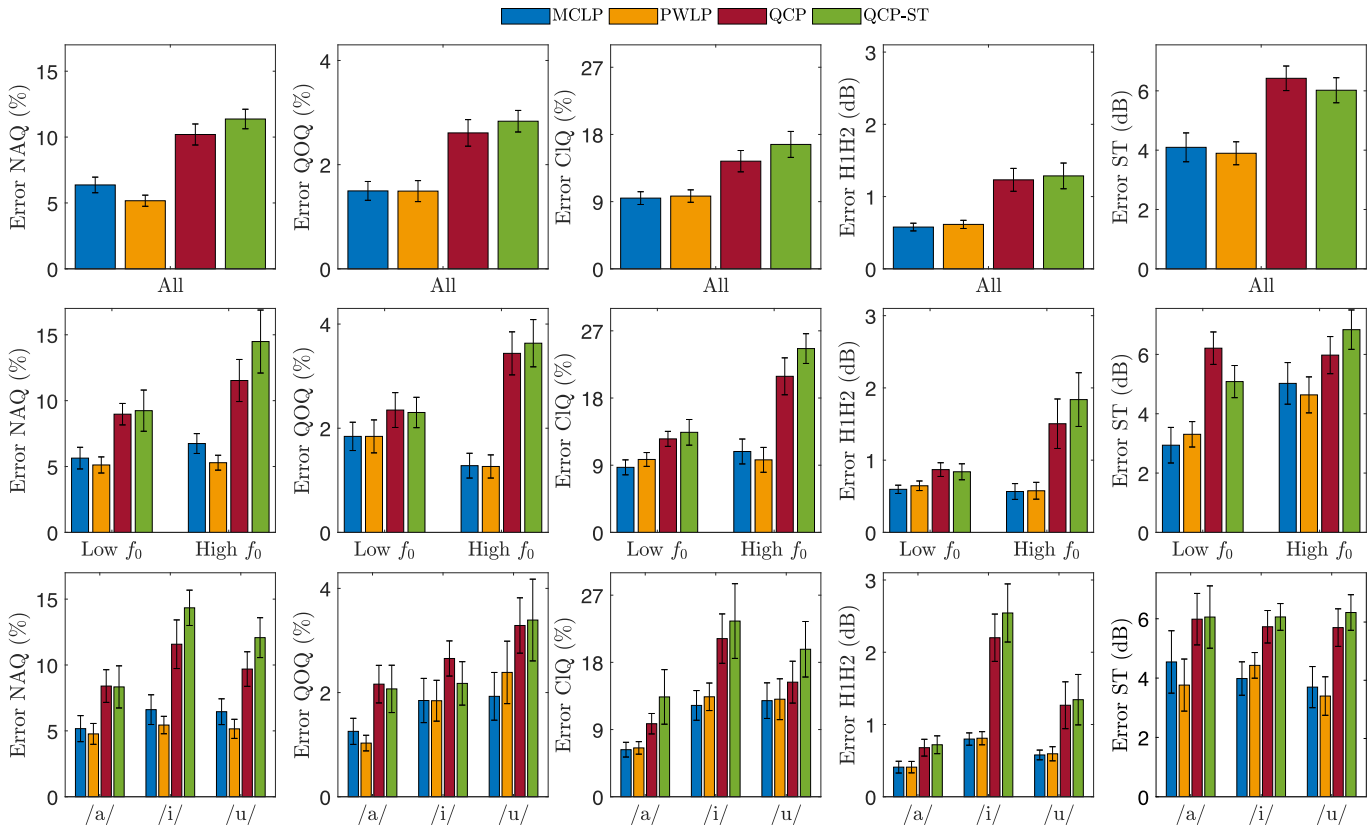

Fig. 5. Bar plots of the estimation error for parameters NAQ, QOQ, ClQ, H1H2, and ST for the synthetic signals from repository II (with the 95% confidence intervals) for different voice inverse filtering methods. The errors are shown according to: all signals (top row), f_0 ranges (middle row), and vowel sounds (bottom row).

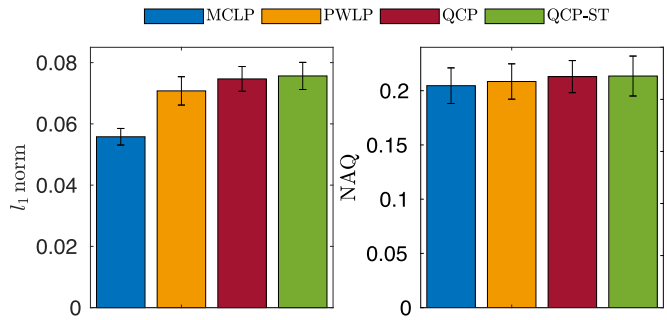

Fig. 6. Bar plots of the l_1 norm on the closed phase and NAQ values in the estimated glottal airflows from natural signals in repository IV by four voice inverse filtering methods. The bar plots show the average values with 95% confidence intervals.

to well-established state-of-the-art inverse filtering methods. However, MCLP estimates exhibit a flatter closed phase, which aligns with the lower l_1 norm values show in Fig. 6. This feature is highly desirable in estimated glottal waveforms [5], [34], [36].

VI. Discussion and Future Works

Results for signals from OPENGLOT indicate that MCLP and PWLP perform similarly, with both surpassing the QCP-based methods. However, our findings on synthetic signals indicate that PWLP outperforms MCLP regarding the NAQ parameter. The performance differences between MCLP and PWLP can be attributed in part to the strategies employed for computing the vocal tract filter coefficients. PWLP better explores the solution space by means of numerous estimations of the filter coefficients randomly generated using the Gibbs sampling method, increasing thus the likelihood of identifying an optimal set of coefficients at the expense of a significant increase in computational complexity. On the other hand, MCLP computes the vocal tract filter sequentially following the proposed algorithm, making the final estimation sensitive to the thresholds, ϵ_1 and ϵ_2, and the initial coefficients, $\mathbf{a}_0$. For initializing the vocal tract filter, classic LP method is applied in this work, which has been shown to perform poorly for voice inverse filtering, specifically for high fundamental frequency phonations. Moreover, the proposed algorithm does not guarantee obtaining the best global estimation of the vocal tract filter coefficients. Another possible reason could be the suggested PWLP hyper-parameters. Here, the parameters for the MCLP method are derived by minimizing the normalized waveform error, E_{v_g}. In contrast, the recommended hyper-parameters for PWLP detailed in [12] (the same considered in our simulations) were determined by minimizing the NAQ error; thus, they would provide the conditions for better performances in terms of NAQ error.

The results by MCLP are promising. However, in future work, we aim to validate the performance of the proposed method

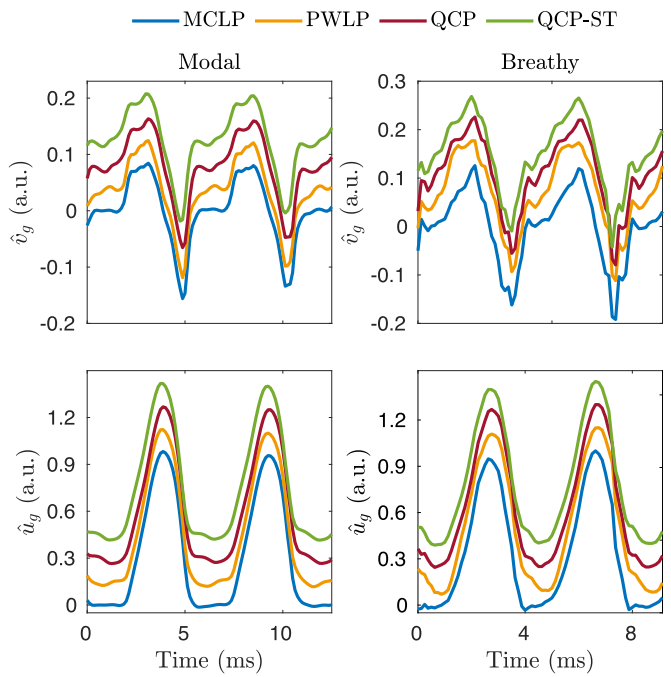

Fig. 7. Inverse filtering analysis for natural voice signals with different phonation qualities from repository IV. Estimations of the glottal function $\hat{v}_g$ (top row), and the glottal airflow $\hat{u}_g$ (bottom row) for four inverse filtering methods are shown. All estimated signals are shifted vertically for better visualization. Left column: Modal voice. Right column: Breathy voice.

on other phonation datasets including increasingly challenging scenarios involving different levels of source-filter interaction, continuous speech, and atypical phonations. [6], [35], [40].

Lastly, it is important to emphasize that the present effort was limited to the inverse filtering of voice signals; future avenues may study the use of MCLP for inverse filtering schemes based on alternative phonation signals, or for other signal processing applications. The correntropy cost function could also be applied in other inverse filtering methods. For example, an unconventional method was proposed in [41] for addressing the automatic glottal inverse filtering from the voice signal spectrogram, combining weighted linear prediction, time-frequency representations, and non-negative matrix factorization. Time-frequency analysis techniques, such as the spectrogram, provide a better description of the non-stationary dynamics in a signal (for more details, see [42], [43]). Moreover, the non-negative matrix factorization yields a part-based low-rank decomposition of the spectrogram in two components corresponding to the glottal acoustic excitation and the encompassing samples nearby, respectively. Maximum correntropy criterion could thus be applied to enhance the weighted linear prediction (as proposed here) and the non-negative matrix factorization [44], [45].

VII. Conclusion

The present study investigated weighted linear prediction based on the maximum correntropy criterion, referred to as MCLP in this study, as a novel method for voice signal analysis. The method involves maximizing the correntropy, a nonlinear similarity measure suitable for optimization problems and digital signal processing. In this work, we discussed the theoretical features of the correntropy with the Gaussian kernel that are relevant in the context of voice inverse filtering and glottal source estimation. Incorporating the maximum correntropy criterion in the linear prediction inverse filtering scheme provides a robust solution against the detrimental effects of impulse-like acoustic excitations during voiced phonation on vocal tract model adjustment.

Utilizing the maximum correntropy criterion, a solution for the MCLP (Maximum Correntropy Linear Prediction) coefficients can be obtained through an iterative fixed-point approach. In this work, we proposed an algorithm for this task based on the effect of the kernel size in the correntropy measure. The kernel size is a critical factor in MCLP model adjustment. We explored three strategies for managing the kernel size and found that using a steplike update of kernel size through Silverman's rule boosts the prediction error sparsity, especially on the closed phase, and inverse filtering performance. We also showed that the MCLP method implements an iterative weighted linear prediction analysis based on a data-driven weighting function emphasizing the closed phase region.

The performance of MCLP for voice inverse filtering was assessed based on the waveform error, the l_1 norm on the closed phase, and the estimation of aerodynamic parameters extracted from the estimated glottal waveforms. The results in a benchmark synthetic dataset showed that MCLP performs similarly or better than other inverse filtering methods well-established in the literature, the probabilistic weighted linear prediction, the quasi closed phase analysis and the quasi closed phase analysis with spectral tilt compensation, respectively. MCLP also showed a proper performance in simulations involving the inverse filtering of natural voices, resulting in smooth glottal airflow estimates with a flatter waveform on the closed phases.

MCLP presents a series of advantages that make it attractive for voice inverse filtering: 1) it is robust to the impulse-like acoustic excitations that impacts negatively on the estimated vocal tract filter; 2) it employs a weighting function that automatically adjusts in a data-driven manner; 3) it does not require prior information about the glottal instants; 4) the computation of vocal tract filter coefficients requires a low to moderate computational burden. However, MCLP has the disadvantage of not guaranteeing the estimation of the optimal vocal tract coefficients, since the proposed algorithm is sensitive to the initial conditions.

References

[1] G. Fant, *Acoustic Theory of Speech Production*. Mounton, The Hague, The Netherlands, 1970. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/9783110873429/html

[2] J. R. Deller, J. G. Proakis, and J. H. Hansen, *Discrete-Time Processing of Speech Signals*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2000.

[3] S. R. Kadiri, P. Alku, and B. Yegnanarayana, "Extraction and utilization of excitation information of speech: A review," *IEEE Proc.*, vol. 109, no. 12, pp. 1920–1941, Dec. 2021.

[4] P. Alku, "Glottal inverse filtering analysis of human voice production—A review of estimation and parameterization methods of the glottal excitation and their applications," *Sadhana*, vol. 36, no. 5, pp. 623–650, 2011.

[5] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana, "Glottal source processing: From analysis to applications," *Comput. Speech Lang.*, vol. 28, no. 5, pp. 1117–1138, 2014.

[6] A. Palaparthi and I. R. Titze, "Analysis of glottal inverse filtering in the presence of source-filter interaction," *Speech Commun.*, vol. 123, pp. 98–108, 2020.

[7] M. Airaksinen, T. Bäckström, and P. Alku, "Automatic estimation of the lip radiation effect in glottal inverse filtering," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, 2014, pp. 398–402.

[8] P. Alku, J. Pohjalainen, M. Vainio, A.-M. Laukkanen, and B. H. Story, "Formant frequency estimation of high-pitched vowels using weighted linear prediction," *J. Acoust. Soc. Amer.*, vol. 134, no. 2, pp. 1295–1313, 2013.

[9] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[10] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 5, pp. 1644–1657, Jul. 2012.

[11] T. Drugman, "Maximum phase modeling for sparse linear prediction of speech," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 185–189, Feb. 2014.

[12] A. Rao and P. K. Ghosh, "Glottal inverse filtering using probabilistic weighted linear prediction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 1, pp. 114–124, Jan. 2019.

[13] M. Freixes, L. Joglar-Ongay, J. C. Socoró, and F. Alías-Pujol, "Evaluation of glottal inverse filtering techniques on openglot synthetic male and female vowels," *Appl. Sci.*, vol. 13, no. 15, 2023, Art. no. 8775.

[14] I. A. Zalazar, G. A. Alzamendi, and G. Schlotthauer, "Symmetric and asymmetric gaussian weighted linear prediction for voice inverse filtering," *Speech Commun.*, vol. 159, 2024, Art. no. 103057.

[15] Y. Miyoshi, K. Yamato, R. Mizoguchi, M. Yanagida, and O. Kakusho, "Analysis of speech signals of short pitch period by a sample-selective linear prediction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 9, pp. 1233–1240, Sep. 1987.

[16] P. Alku, C. Magi, S. Yrttiaho, T. Bäckström, and B. Story, "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering," *J. Acoust. Soc. Amer.*, vol. 125, no. 5, pp. 3289–3305, 2009.

[17] M. Airaksinen, T. Raitio, B. Story, and P. Alku, "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 3, pp. 596–607, Mar. 2014.

[18] S. Seshadri, L. Juvela, O. Räsänen, and P. Alku, "Vocal effort based speaking style conversion using vocoder features and parallel learning," *IEEE Access*, vol. 7, pp. 17230–17246, 2019.

[19] C. Ma, Y. Kamp, and L. F. Willems, "Robust signal selection for linear prediction analysis of voiced speech," *Speech Commun.*, vol. 12, no. 1, pp. 69–81, 1993.

[20] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilised weighted linear prediction," *Speech Commun.*, vol. 51, no. 5, pp. 401–411, 2009.

[21] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. signal Process.*, vol. 39, no. 2, pp. 411–423, Feb. 1991.

[22] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[23] I. A. Zalazar, G. A. Alzamendi, M. Zañartu, and G. Schlotthauer, "Correntropy-based linear prediction for voice inverse filtering," *Proc. SPIE*, vol. 12567, pp. 356–365, 2023.

[24] J.-W. Xu and J. C. Principe, "A pitch detector based on a generalized correlation function," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1420–1432, Nov. 2008.

[25] X. Cui, Z. Chen, F. Yin, and X. Xu, "Correntropy-based multi-objective multi-channel speech enhancement," *Circuits, Syst., Signal Process.*, vol. 41, no. 9, pp. 4998–5025, 2022.

[26] R. Singh and J. C. Principe, "Correntropy based hierarchical linear dynamical system for speech recognition," in *Proc. 2018 IEEE Int. Joint Conf. Neural Netw.*, 2018, pp. 1–7.

[27] I. Santamaria, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2187–2197, Jun. 2006.

[28] A. Singh and J. C. Principe, "A closed form recursive solution for maximum correntropy training," in *Proc. 2010 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2010, pp. 2070–2073.

[29] S. Zhao, B. Chen, and J. C. Principe, "Kernel adaptive filtering with maximum correntropy criterion," in *Proc. 2011 Int. Joint Conf. Neural Netw.*, IEEE, 2011, pp. 2012–2017.

[30] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. Berlin, Germany: Springer Science & Business Media, 2010.

[31] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *Proc. 2009 IEEE Int. Joint Conf. Neural Netw.*, 2009, pp. 2950–2955.

[32] D. G. Manolakis, V. K. Ingle, and S. Kogan, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. New York, NY, USA: McGraw-Hill, 2000.

[33] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Evanston, IL, USA: Routledge, 2018.

[34] P. Alku et al., "Openglot–An open environment for the evaluation of glottal inverse filtering," *Speech Commun.*, vol. 107, pp. 38–47, 2019.

[35] Y.-R. Chien, D. D. Mehta, J. Gunason, M. Zañartu, and T. F. Quatieri, "Evaluation of glottal inverse filtering algorithms using a physiologically based articulatory speech synthesizer," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 8, pp. 1718–1730, Aug. 2017.

[36] M. Airaksinen, T. Bäckström, and P. Alku, "Quadratic programming approach to glottal inverse filtering by joint norm-1 and norm-2 optimization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 929–939, May 2017.

[37] S. R. Kadiri and P. Alku, "Analysis and detection of pathological voice using glottal source features," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 367–379, Feb. 2020.

[38] M. R. Thomas and P. A. Naylor, "The SIGMA algorithm: A glottal activity detector for electroglottographic signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 8, pp. 1557–1566, Nov. 2009.

[39] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.

[40] I. Langheinrich, S. Stone, X. Zhang, and P. Birkholz, "Glottal inverse filtering based on articulatory synthesis and deep learning," in *Proc. Interspeech*, 2022, pp. 1327–1331.

[41] M. Airaksinen, L. Juvela, T. Bäckström, and P. Alku, "Automatic glottal inverse filtering with non-negative matrix factorization," in *Proc. Interspeech*, 2016, pp. 1039–1043.

[42] S. Mallat, *A Wavelet Tour of Signal Processing*. (Wavelet Tour of Signal Processing), Cambridge, MA, USA: Academic Press, 1999.

[43] R. B. Pachori, *Time-Frequency Analysis Techniques and Their Applications*. Boca Raton, FL, USA: CRC Press, 2023.

[44] J. J.-Y. Wang, X. Wang, and X. Gao, "Non-negative matrix factorization by maximizing correntropy for cancer clustering," *BMC Bioinf.*, vol. 14, pp. 1–11, 2013.

[45] S. Peng, W. Ser, Z. Lin, and B. Chen, "Robust sparse nonnegative matrix factorization based on maximum correntropy criterion," in *Proc. 2018 IEEE Int. Symp. Circuits Syst.*, 2018, pp. 1–5.