Missing Sample Estimation Based on High-Order Sparse Linear Prediction for Audio Signals

Bisrat Derebssa Dufera
ESAT-ETC, Group T Leuven Campus
KU Leuven,
Leuven, Belgium
Addis Ababa Institute of Technology, AAU, Ethiopia
bisratderebssa.dufera@student.kuleuven.be

Koen Eneman
ESAT-ETC, Group T Leuven Campus
ESAT-STADIUS
KU Leuven
Leuven, Belgium
koen.eneman@kuleuven.be

Toon van Waterschoot
ESAT-ETC, Group T Leuven Campus
ESAT-STADIUS
KU Leuven
Leuven, Belgium
toon.vanwaterschoot@esat.kuleuven.be

***Abstract*—The restoration of click degraded audio signals is important to achieve acceptable audio quality in many old audio media. Restoration by missing sample estimation based on conventional linear prediction has been extensively researched and used; however, it is hampered by the limitations of the linear prediction model. Recently, it has been shown that high-order sparse linear prediction offers better representation of music and voiced speech over conventional linear prediction. In this paper, the use of high-order sparse linear prediction for missing sample estimation of click degraded audio signals is proposed. The paper also explores a possible computational time saving by combining the high-order sparse linear prediction coefficient determination and filtering operations. Evaluation with different types of speech and audio data show that the proposed method achieves an improvement in SNR over conventional linear prediction based filtering for all considered speech and audio data types.**

***Index Terms*—Missing sample estimation, Click degradation, Linear prediction, High-order sparse linear prediction**

I. Introduction

The term ‘click’ is used to refer to finite duration artifacts which occur at random positions in an audio signal [1]. Clicks are perceived by the listener as impulsive noise ranging from tiny ‘tick’ noises, ‘scratch’ to ‘crackle’ noise. These artifacts arise due to defects on the physical medium from microscopic surface irregularities to physical breakage of the medium [2]. Ideally, a missing sample estimation or interpolation technique used to restore click degraded audio signals should restore only those samples which are degraded, without modifying the undegraded signal. Two tasks are thus important for a successful click removal system: detection of degraded signal samples/segments and estimation of restored samples in degraded segments. In this paper, only the estimation problem is considered.

Several methods have been proposed for restoration of click degraded audio signals. Maximum a Posterior (MAP) Interpolator that uses Bayesian inference as a means of incorporating prior information about the restoration problem has been proposed in [3]. By assuming the clicks as a zero-mean multivariate Gaussian process, it combines the detection and estimation procedure. It does not make any assumption about the underlying signal generation process. It has the appeal of being the ‘most probable’ solution of all the possible solutions. However, the problem of click detection and estimation involves maximizing the probability by searching through all states. Searching for all possible states will be computationally impractical as there are a total of 2^N different click states for a signal of length N. Several approaches have been proposed to overcome this limitation of the MAP interpolator by incorporating additional a priori information in the interpolation. Linear prediction (LP) based methods that incorporate the source-filter model of speech production have been proposed and used extensively in the literature for restoration of click degraded speech and audio signals [1], [3], [4], [5].

The autoregressive interpolator [1] is based on the assumption that the underlying audio signal is generated by passing an excitation through an all-pole filter. The least square (LS) interpolator data is obtained by minimizing the squared prediction error under the assumption that the noise samples within click bursts are mutually independent and drawn from a Gaussian zero-mean process. It has been shown that the LS solution can be found from the LP coefficients, the range of known and range of missing samples [1].

The LS solution has several limitations, one of which is the fact that the AR coefficients of the undegraded underlying signal are unknown and need to be determined. Janssen *et. al.* [6] proposed a method that minimizes a sum of squared residual errors involving the unknown samples, the LP coefficients, and the known samples from a sufficiently large neighborhood as a function of the unknown samples and the unknown LP coefficients. It is an iterative method whereby in each iteration minimization with respect to the LP coefficients and, subsequently, minimization with respect to the unknown samples are performed. The LP coefficients of the filter are identified by minimizing the $\ell2$-norm of the residual, the difference between the actual and predicted signal. This works well for unvoiced speech; that is, when the samples of the excitation are Gaussian and independent identically distributed [7]. This is not the case for voiced speech, as the excitation is spiky and quasi-periodic, and music [6]. In this case, the estimation of the LP coefficients by minimizing the $\ell2$-norm is observed to give more emphasis to the periodic peaks of the residual [8]. Therefore, it trades off spectral envelope estimation accuracy to estimating the position of the poles due to the excitation [8]. Several methods have been proposed to mitigate these effects. One of the most recent is the use of sparse linear prediction (SLP), that tries to maximize the sparsity of the residual as well as the prediction coefficients. By using high order sparse linear prediction (HOSpLP) a more efficient decoupling between the pitch harmonics and the spectral envelope has been achieved in [7], [9], [10]. Even though it was originally used for speech processing purposes, it has found applications in many fields such as radar processing [11], geology and general signal representations [12].

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of the HGPP project (International University Partnership Services for the Establishment of Postgraduate Programmes in Ethiopia) funded through GIZ GmbH, KU Leuven internal funds C2-16-00449, VES/16/032, IMP/14/037, and VLAIO TETRA Project no. HBC.2016.0085. The research leading to these results has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors’ views and the Union is not liable for any use that may be made of the contained information.

ISBN 978-90-827970-1-5 © EURASIP 2018

Another drawback of LS interpolator is that it requires the use of a pitch predictor to predict long-term correlation. As such, the performance of the system is always limited by the accuracy of the pitch predictor. In [13], a sparse linear prediction approach that minimizes the $\ell 2$-norm was proposed to jointly model the short-term and long-term correlations. It showed improved SNR over conventional LP methods. Similar work was proposed in [14] that uses sparse AR modeling to construct a cascade of formant filter and pitch filter to eliminate impulsive disturbances from archive speech signals. However, in both works the limitation due to the use of the $\ell 2$-norm was not addressed.

In this paper, the use of HOSpLP coefficients in missing sample estimation of click degraded speech and music is proposed. The use of HOSpLP can jointly solve both problems (*emphasis on quasi-periodic excitation and pitch prediction*) of conventional LP based missing sample estimation. It should also be noted that HOSpLP-based signal restoration can be considered as an extension of sparsity-based audio inpainting [15] with the novel aspect of using signal-dependent dictionaries.

This paper is organized as follows. Section II formally describes SLP. Section III presents the proposed method and discusses different aspects of the algorithm. Section IV presents the data used, the type of artificial click degradation and the performance measure used. Finally, section V presents the results of the proposed method in comparison with conventional LP. Finally, section VI presents additional discussions and concludes the work.

II. Sparse Linear Prediction

The solution for finding the LP coefficient vector $\mathbf{a}$ from a set of observed samples $\mathbf{x}$ can be formulated as the following optimization problem [7]:

$$\mathbf{a} = \arg\min_{\mathbf{a}} \left|\left|\mathbf{x} - \mathbf{X}\mathbf{a}\right|\right|_p^p - \gamma\left|\left|\mathbf{a}\right|\right|_k^k \tag{1}$$

where,

$$\mathbf{X} = \begin{bmatrix} x(N_1-1) & \cdots & x(N_1-P) \\ \vdots & \ddots & \vdots \\ x(N_2-1) & \cdots & x(N_2-P) \end{bmatrix}$$

N_1 and N_2 are the start and end indexes of the frame,
P is the prediction order,
γ is the regularization parameter;

The ℓp-norm operator $||.||_p$ is defined as

$$\left|\left|\mathbf{x}\right|\right|_p = \left(\sum_{n=N_1}^{N_2} |x(n)^p|\right)^{\frac{1}{p}} \tag{2}$$

In conventional LP, the $\ell 2$-norm is used, therefore $p = 2$. In addition, no a priori information about the coefficient vector is assumed and therefore the regularization parameter $\gamma = 0$.

One of the motivations of using SLP is to decrease the emphasis on outliers in the residual so that the estimated spectral envelope is less affected by the quasi-periodic excitation [7]. This can be achieved by considering the sparsity of the residual, i.e. by minimizing the '$\ell 0$-norm', instead of the $\ell 2$-norm. Even though, the '$\ell 0$-norm' is an ideal candidate for measure of sparsity, it results in a combinatorial problem that is NP hard. To alleviate this problem the $\ell 1$-norm, $p = 1$, can be used as a convex relaxation of the '$\ell 0$-norm' [7]. In addition, by taking the sparsity of the coefficient vector into account by making $\gamma \neq 0$, setting $k = 1$ and using a high-order sparse linear predictor, the short-term predictor, associated with the formants, and the long-term predictor, associated with pitch excitation, can be jointly estimated [7]. This stems from the fact that a cascade of the short-term predictor filter and long-term predictor filter results in a filter that has few non-zero coefficients [16]. As such, the sparsity of the coefficient vector can be used as a regularization term in the cost function in addition to the sparsity of the residual.

The regularization parameter, γ, controls the trade-off between the sparsity of the residual and the sparsity of the predictor coefficients. In [17], the modified L-curve was used to find an optimum value of γ as the point of maximum curvature of the curve $(||\mathbf{x} - \mathbf{X}\mathbf{a}||_1, ||\mathbf{a}||_1)$ for different values of of γ. An adaptive update algorithm for estimating the regularization parameter that was based on the observation that the optimal γ is related to the pitch gain was proposed in [10].

It has been shown in [7] that for voiced speech and music, the use of the $\ell 1$-norm and HOSpLP outperforms conventional LP in spectral envelop estimation, sparsity of prediction residual and sparsity of the prediction coefficients.

III. Proposed Method of Restoration

In this paper a novel method of restoration of click degraded audio signals is proposed that uses HOSpLP coefficients instead of using the conventional LP coefficients in the iterative Janssen algorithm [6] for estimating the missing samples as shown in Algorithm 1. The click degraded segment location and duration are assumed to be known a priori. The HOSpLP coefficients are estimated by using the Alternating Direction Method of Multipliers (ADMM) algorithm [12] shown in algorithm 2.

Algorithm 1 HOSpLP Based missing sample estimation. Modified from [15] by incorporating HOSpLP model

1: **procedure** JANSSENINTERPOLATION_HOSPLP
2: **Input:** $\mathbf{s}, \mathbf{v}_m, \mathbf{v}_{Obs}, P, L, \gamma, K, \epsilon$
3: **Output:** $\hat{\mathbf{x}}$
4: $\mathbf{\Theta} = \left|\mathbf{v}_m \mathbf{1}_{1\times N} - \mathbf{1}_{M\times 1}[1, 2, \cdots N]\right|$;
5: $\hat{\mathbf{x}}_{\mathbf{v}_{Obs}} = \mathbf{s}_{\mathbf{v}_{Obs}}$; $\hat{\mathbf{x}}_{\mathbf{v}_m} = 0$; $\mathbf{\Phi} = \mathbf{0}_{M\times N}$; $l = 0$;
6: **for** $l \leq L$ **do**
7: $\mathbf{a}$ =L1L1_SLP_ADMM($\hat{\mathbf{x}}$,P,γ,K,ϵ);
8: $\mathbf{b} = \begin{bmatrix} 1 & -\mathbf{a}^T \end{bmatrix} \mathbf{A}$;
9: $\mathbf{\Phi}_{i,j} = \mathbf{b}_{\mathbf{\Theta}_{i,j}+1}$; $\forall i, j : \mathbf{\Theta}_{i,j} > P$
10: $\hat{\mathbf{x}} = -\mathbf{\Phi}^{-1}_{(1:M,\mathbf{v}_m)} \mathbf{\Phi}_{(1:M,\mathbf{v}_{Obs})} \mathbf{s}_{\mathbf{v}_{Obs}}$;
11: $l \leftarrow l + 1$;
12: Return

Where:

$\mathbf{s}$ is the click degraded signal to be restored;
M is the number of missing samples;
$N = N_2 - N_1$ is the number of samples in each frame;
$\mathbf{v}_{Obs}$ is a vector index of known samples;
$\mathbf{v}_m$ is a vector index of click degraded samples;

$$\mathbf{A} = \begin{bmatrix} 1 & -a_1 & -a_2 & \cdots & -a_P \\ -a_1 & -a_2 & \cdots & -a_P & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_P & 0 & 0 & \cdots & 0 \end{bmatrix};$$

L is the number of Janssen iterations; and
K is the maximum number of ADMM iterations.

Algorithm 1 is iterative and computationally expensive due to the nested ADMM iteration at step 7 of the algorithm. This nested iteration significantly increases the computational cost of the proposed algorithm. One way to alleviate this increased cost is to merge the two iterations into one. The merging can be achieved by re-estimating the

ISBN 978-90-827970-1-5 © EURASIP 2018

missing samples inside the ADMM iteration. The ADMM algorithm for solving L1-regularized linear regression problem [12] obtains SLP coefficients by starting from the conventional LP coefficients and iteratively minimizing the sum of $\ell 1$-norm of the estimation error and $\ell 1$-norm of the coefficients as shown in Algorithm 2.

Algorithm 2 ADMM [12]

procedure L1L1_SLP_ADMM
    Input: $\mathbf{x}$, P, γ, K, ϵ
    Output: $\mathbf{a}$
    $y, z = 0_{(N+2*P)\times 1}$; $Iter = 0$;
    $\mathbf{R_x} = autocorrelation(\mathbf{x})$;
    $\mathbf{a}_\gamma = levinsondurbin(\mathbf{R_x})$;
    $\mathbf{H} = \left(\begin{bmatrix} -\gamma \mathbf{I}_{P\times P}^T & \mathbf{X}^T \end{bmatrix}^T \right)^+$
    while do$\{E_n > \epsilon \quad \&\& \quad C_n > \epsilon \quad \&\& \quad Iter < K\}$
        $\mathbf{a} = \mathbf{a}_\gamma - \mathbf{H}(\mathbf{y} - \mathbf{u})$;
        $\mathbf{e} = \mathbf{Ax}$;
        $\mathbf{z} = \begin{bmatrix} \gamma\ \mathbf{a}^T & \mathbf{e}^T \end{bmatrix}^T$;
        $\mathbf{y} = Sm(\mathbf{z} + \mathbf{u}, \rho)$;
        $\mathbf{u} \leftarrow \mathbf{u} + \mathbf{z} - \mathbf{y}$;
        $E_n = ||\mathbf{e}||_1$; $C_n = ||\mathbf{a}||_1$;
        $Iter \leftarrow Iter + 1$;
    $\mathbf{a} = \mathbf{y}_{1:M}/\gamma$;
    Return

Where $Sm(x, y)$ is a soft thresholding operator and ρ is augmented Lagrangian parameter.

In this iterative procedure, the $\ell 1$-norm of the residual is decreased and the HOSpLP coefficients are made sparser by every iteration. As such, the HOSpLP coefficients at successive iterations are better representatives of the data. Therefore, it seems straightforward to do the restoration at each iteration by using the most recent HOSpLP coefficients. A novel restoration method that merges the two algorithms is shown in Algorithm 3. The number of Janssen iterations at each iteration of the ADMM algorithm is limited to one as the HOSpLP coefficients are fixed in each ADMM iteration.

A. Computational complexity

The computational complexity of Algorithms 1 and 3 are analyzed (only considering multiplications) as follows. Let:

$\Gamma =$ cost of each iteration of the ADMM algorithm.

- **Algorithm 1:** The most computationally expensive parts are the following.
  - Cholesky decomposition: efficient solution to matrix inversion at line 10;
  - Estimation of $\hat{\mathbf{x}}$: backward substitution after Cholesky decomposition at line 10; and
  - SLP coefficients calculation: at line 7.

  It can be shown that the number of arithmetic operations required by Algorithm 1 can be approximated by [18]:

$$L\frac{M^3}{3} + L\frac{N^2}{2} + LK\Gamma \tag{3}$$

- **Algorithm 3:** By comparing with algorithm 2, it is seen that the only additional computational cost is the estimation of the data in each iteration. Therefore, its computational cost can be approximated by:

$$K\frac{M^3}{3} + K\frac{N^2}{2} + K\Gamma \tag{4}$$

Algorithm 3 HOSpLP-based missing sample estimation with merging of ADMM iterations and Janssen iterations

1: **procedure** ADMM_HOSPLP_ITER
2:     **Input:** $\mathbf{x}, \gamma, K, \epsilon, \mathbf{v}_m, \mathbf{v}_{Obs}$
3:     **Output:** $\hat{\mathbf{x}}$
4:     $\mathbf{\Theta} = \left| \mathbf{v}_m \mathbf{1}_{1\times N} - \mathbf{1}_{M\times 1}[1, 2, \cdots N] \right|$;
5:     $\mathbf{\Phi} = \mathbf{0}_{M\times N}$; $\hat{\mathbf{x}}_{\mathbf{v}_{Obs}} = \mathbf{x}_{\mathbf{v}_{Obs}}$; $\hat{\mathbf{x}}_{\mathbf{v}_m} = 0$; $Iter = 0$;
6:     $\mathbf{R_x} = autocorrelation(\mathbf{x})$;
7:     $\mathbf{a}_\gamma = levinsondurbin(\mathbf{R_x})$;
8:     $\mathbf{H} = \left(\begin{bmatrix} -\gamma \mathbf{I}_{P\times P}^T & \mathbf{X}^T \end{bmatrix}^T \right)^+$
9:     **while do**$\{E_n > \epsilon \quad \&\& \quad C_n > \epsilon \quad \&\& \quad Iter < K\}$
10:         $\mathbf{a} = \mathbf{a}_\gamma - \mathbf{H}(\mathbf{y} - \mathbf{u})$;
11:         $\mathbf{e} = \mathbf{Ax}$;
12:         $\mathbf{z} = \begin{bmatrix} \gamma \mathbf{a}^T & \mathbf{e}^T \end{bmatrix}^T$;
13:         $\mathbf{y} = Sm(\mathbf{z} + \mathbf{u}, \rho)$;
14:         $\mathbf{u} \leftarrow \mathbf{u} + \mathbf{z} - \mathbf{y}$;
15:         $E_n = ||\mathbf{e}||_1$; $C_n = ||\mathbf{a}||_1$;
16:         $\mathbf{b} = \begin{bmatrix} 1 & -\mathbf{a}^T \end{bmatrix} \mathbf{A}$;
17:         $\mathbf{\Phi}_{i,j} = \mathbf{b}_{\mathbf{\Theta}_{i,j}+1}$; $\forall i, j : \mathbf{\Theta}_{i,j} > P$
18:         $\hat{\mathbf{x}} = -\mathbf{\Phi}^{-1}_{(1:M,\mathbf{v}_m)} \mathbf{\Phi}_{(1:M,\mathbf{v}_{Obs})} \mathbf{s}_{\mathbf{v}_{Obs}}$;
19:         $Iter \leftarrow Iter + 1$;
20:     Return

It is not straightforward to compare (3) and (4) as the number of ADMM iterations, K, is not known a priori. The iteration can be stopped if the sparsity of the residual and HOSpLP coefficients is satisfied. Therefore, it is dependent on the norm of the residual and of the coefficient vectors, which in turn are dependent on the missing sample estimation done on line 18 of Algorithm 3. As such, the value of K in Algorithm 1 and 3 is not necessarily the same. However, the following observations can be made:

- $K \gg L$ as experiments show that no more than 5 Janssen iterations are necessary.
- The cost of the ADMM algorithm, the most computationally expensive part, is decreased by a factor of L in Algorithm 3 as compared to Algorithm 1.

IV. Data used, Click noise model and Performance Measure

A. Data

The experiments are conducted using the following four datasets so that the results can be representative for any kind of audio and speech data.

- Synthetic male and female vowels: ten vowels corresponding to /b*ea*t/, /b*i*t/, /b*e*t/, /b*a*t/, /p*ar*t/, /p*o*t/, /b*oo*t/, /b*oo*k/, /b*u*t/, /p*er*t/ synthesized using the Klatt speech synthesizer [19];
- Natural male and female vowels: same ten vowels from human speakers from Western University of Michigan dataset [20];
- Male and female speech: ten different male and female speech from Voxforge dataset [21]; and
- Music: ten excerpts consisting of male singing voice, female singing voice and instrument from Sparse Models, Algorithms and Learning for Large-scale data (SMALL) dataset [22].

In order to have comparable degradations among all signals, each signal is normalized so that the maximum amplitude is 1.

B. Click Degradation Model

The onset, duration and amplitude of each click degradation is usually modeled probabilistically. Different distributions for the time

ISBN 978-90-827970-1-5 © EURASIP 2018

Table I
SIMULATION PARAMETERS

No	Description	Value
1	Sampling frequency	8 kHz
2	Frame size	256 samples
3	Conventional LP order	12
4	HOSpLP order	128
5	Dataset (counting male and female)	8
6	Cases for each data set	10
7	Number of file for each case	10
8	Number of simulations for each file	25
9	Click duration	0.25 msec - 10 msec

between impulses and for their amplitudes can be used [1], [4], [23]. As the main objective of this research was the study of the effect of using HOSpLP for missing sample estimation of click-degraded audio signals, the onset and duration of the click degradation was assumed to be known a priori. The click degraded samples were artificially replaced by zero-mean Gaussian noise.

C. Performance Measure

The performance of the proposed method can be assessed by computing the signal-to-noise ratio (SNR) computed on the click duration of the click-degraded fragment, defined by:

$$SNR(\mathbf{s}, \hat{\mathbf{s}}) = 10 * \log \frac{||\mathbf{s}||^2}{||\mathbf{s} - \hat{\mathbf{s}}||^2} \quad (5)$$

Where $\mathbf{s}$ is a vector of the original audio samples in the click duration and $\hat{\mathbf{s}}$ is a vector of the estimated audio samples in the click duration.

V. RESULTS

The artificially click-degraded audio and speech data were restored using iterative filtering by using HOSpLP and conventional LP coefficients. The SNR is computed in the missing sample range and averaged over all the audio data for each click duration. The parameters used during the simulations are shown in Table I.

A. Impact of Regularization Parameter

The regularization parameter plays an important role on the effectiveness of the HOSpLP representation. Even though different methods have been proposed to determine the optimum value, the methods are optimized with respect to sparsity of the residual and HOSpLP coefficients. They are not guaranteed to be optimal for the application of missing sample estimation. Therefore, an experiment was conducted by changing the regularization parameter between 0.01 and 10 to obtain the best γ value. Algorithm 1 was used to estimate the click-degrades samples. The SNR values of the estimated signal are averaged over all signal frames, and over all signals belonging to the same data type. Fig. 1 shows a SNR comparison of conventional LP and the proposed method for music data with different regularization parameter values.

From Fig. 1 it is seen that HOSpLP-based missing sample estimation with proper regularization parameter selection outperforms conventional LP-based missing sample estimation. A regularization parameter value between 0.1 and 1 is found to give consistently better SNR results for all types of data used. After similar experiments on all datasets, a regularization parameter of $\gamma = 0.1833$ was selected. Comparison of HOSpLP-based missing sample estimation with conventional LP-based missing sample estimation for all data types and $\gamma = 0.1833$ is shown in Fig. 2. Fig. 2 shows that for all

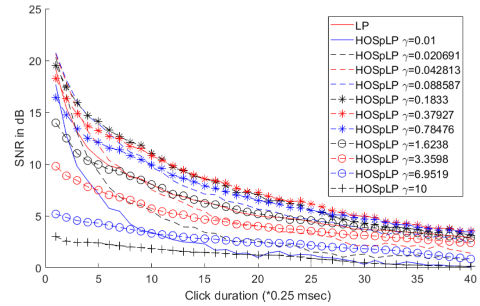

Figure 1. SNR for different regularization parameters for music.

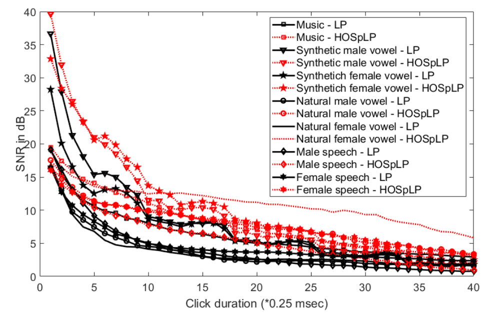

Figure 2. SNR for all data types with $\gamma = 0.1833$

the data types considered, the proposed method provides consistently higher SNR over conventional LP-based missing sample estimation except for very short or very long click duration. For very short click duration, it is seen that both HOSpLP and conventional LP achieve very high SNR. For click durations that are very long, in the range of 10msec and larger, the performance of the proposed method is seen to asymptotically approach conventional LP which itself approaches 0 dB. This is expected as the number of observed samples in a frame decreases with increasing click duration.

B. Effect of number of iterations on restoration performance

As the Janssen iterative algorithm does not have a stopping criterion apart from an initially set number of iterations, the selection of the number of iterations has significant impact on both the estimation accuracy and computational time. To see the effect of the number of Janssen iterations on the estimation performance of the proposed method, Algorithm 1 was executed for different number of iterations. A regularization parameter value of $\gamma = 0.1833$ was used in these experiments as this value led to the best estimation performance for the different datasets used. Fig. 3 shows the SNR of the proposed method for different iterations for music.

From Fig. 3 it is clear that after the fifth iteration, virtually no additional improvement in estimation performance is obtained. For the other data types, similar results are obtained where no additional SNR improvement is gained after the fifth iteration.

C. Combination of ADMM and Janssen iterations

The missing sample estimation performance of Algorithm 1 and Algorithm 3 as compared to the conventional LP-based iterative filtering for music, male speech and female speech with $\gamma = 0.1833$ is shown on Fig. 4. Algorithm 3 was implemented for different values of regularization parameter and it was observed that the regularization parameter γ has the same impact for both algorithms. In order to compare the computational time taken by the two algorithms,

ISBN 978-90-827970-1-5

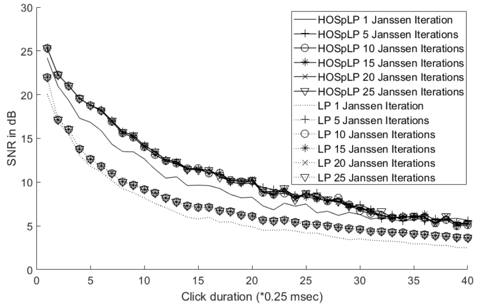

Figure 3. SNR for different Janssen algorithm iterations for music data.

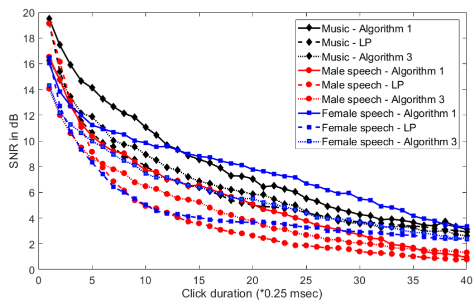

Figure 4. SNR of of algorithm 1 and 3 vs conventional LP for music data.

they were implemented in MATLAB and their execution time was measured. Although the obtained result from this execution may not be representative for all scenarios, it is an indicator of a comparison between the two algorithms.

- **Number of ADMM iterations:-** When using Algorithm 3 the number of ADMM iterations, K, is observed to be much lower than Algorithm 1. This can be traced to the re-estimation of the signal in each ADMM iteration cycle.
- **Overall computational time:-** Algorithm 3 is 3.95 times faster than Algorithm 1.

Fig. 4 shows that Algorithm 1 achieves the best result albeit with more computational cost. On the other hand, Algorithm 3 achieves a result inferior to Algorithm 1 but better than conventional LP for the three data types.

VI. Conclusion

This paper proposed the use of high-order sparse linear prediction for missing sample estimation of click-degraded audio signals. The proposed method achieved an improvement in SNR over conventional LP-based filtering for all considered speech and audio data types. It also investigated the effect of regularization on the performance of the proposed method and found a regularization parameter value that provided the best or very close to the best SNR for all data considered. The paper also explored a possible computational time saving by combining the iterative HOSpLP coefficient determination and the iterative filtering operation. Even though the obtained computational time improvement is not of the order of the HOSpLP number of coefficients or the data size, an improvement of a factor of 3.95 was obtained.

Even though the proposed method achieved better SNR over conventional LP, the perceptual quality of the restored audio signal should also be evaluated by scores from listening tests.

References

[1] S. J. Godsill and P. J. W. Rayner, *Digital audio restoration: a statistical model based approach*. Springer, 1998.

[2] M. K. Mathai and J. Deepa, "Design and implementation of restoration techniques for audio denoising applications," in *IEEE Recent Advances in Intelligent Computational Systems*, (Trivandrum, India), 2015.

[3] S. J. Godsill and P. J. W. Rayner, "A bayesian approach to the restoration of degraded audio signals," *IEEE Trans. Speech Audio Processing*, vol. 3(4), pp. 267–278, July 1995.

[4] F. R. Avila and L. W. P. Biscainho, "Bayesian restoration of audio signals degraded by impulsive noise modeled as individual pulses," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20(9), pp. 2470–2480, November 2012.

[5] M. Lagrange and S. Marchand, "Long interpolation of audio signals using linear prediction in sinusoidal modeling," *J. Audio Eng. Soc.*, vol. 55(5), pp. 891–905, May 2005.

[6] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Trans., Acoust., Speech, Signal Process.*, vol. 34(2), pp. 317–330, Apr. 1986.

[7] 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(5), pp. 1644–1657, July 2012.

[8] T. van Waterschoot and M. Moonen, "Comparison of linear prediction models for audio signals," *EURASIP J. Audio, Speech, Music Process.*, vol. 20(5), pp. 1644–1657, July 2008.

[9] L. Shi, J. R. Jensen, and M. G. Christensen, "Least 1-norm pole-zero modeling with sparse deconvolution for speech analysis," in *Proc. 2009 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '09)*, vol. 20(3), (New Orleans, LA, USA), pp. 731–735, June 2009.

[10] D. Giacobello, M. G. Christensen, J. Dahl1, S. H. Jensen, and M. Moonen, "Joint estimation of short-term and long-term predictors in speech coders," in *Proc. 2009 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, (Taipei, Taiwan), pp. 409–412, Apr. 2009.

[11] I. Erer and K. B. Sarikaya, "Enhanced radar imaging via sparsity regularized 2d linear prediction," in *Proc. 22nd European Signal Process. Conf. (EUSIPCO '14)*, vol. 20(3), (Taipei, Taiwan), pp. 1751–1755, 2014.

[12] T. L. Jensen, D. Giacobello, T. van Waterschoot, and M. G. Christensen, "Fast algorithms for high-order sparse linear prediction with applications to speech processing," *Speech Communication*, vol. 76(5), pp. 143–156, July 2016.

[13] J. Koloda, A. M. Peinado, and V. Sánchez, "Speech reconstruction by sparse linear prediction," *in Advances in Speech and Language Technologies for Iberian Languages. (Torre Toledano D. et al., Eds.)*, vol. 328(5), pp. 247–256, 2012.

[14] M. Niedźwiecki and M. Cioek, "Elimination of clicks from archive speech signals using sparse autoregressive modeling," *J. Audio Eng. Soc.*, vol. 55(5), pp. 891–905, May 2005.

[15] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20(3), pp. 922–932, March 2012.

[16] D. Giacobello, T. van Waterschoot, M. G. Christensen, S. H. Jensen, and M. Moonen, "High-order sparse linear predictors for audio processing," in *Proc. 20th European Signal Process. Conf. (EUSIPCO '10)*, (Aalborg, Denmark), pp. 234–238, August 2010.

[17] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Review*, vol. 34(4), pp. 561–580, Dec. 1992.

[18] A. Krishnamoorthy and D. Menon, "Matrix inversion using cholesky decomposition," in *Signal processing algorithms, architectures, arrangements and applications*, vol. 20(3), pp. 170–72, Sept. 2013.

[19] D. H. Klatt, "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Amer.*, vol. 3(67), pp. 971–995, March 1980.

[20] J. M. Hillenbrand, "Acoustic characteristics of american english vowels." https://homepages.wmich.edu/ hillenbr/voweldata.html. accessed Dec. 14, 2017.

[21] Voxforge.org, "Free speech ... recognition (linux, windows and mac) - voxforge.org." http://www.voxforge.org/. accessed Dec. 14, 2017.

[22] SMALL, "Sparse models, algorithms and learning for large-scale data." http://www.small-project.eu/. accessed Oct. 13, 2017.

[23] M. Brandt, S. Doclo, T. Gerkmann, and J. Bitzer, "Impulsive disturbances in audio archives: signal classification for automatic restoration," *J. Audio Eng. Soc.*, vol. 65(10), pp. 826–840, Oct. 2017.

ISBN 978-90-827970-1-5 © EURASIP 2018