Restoration of Click Degraded Speech and Music Based on High Order Sparse Linear Prediction

Bisrat Derebssa Dufera[1], Eneyew Adugna[2], Koen Eneman[3] and Toon van Waterschoot[4]
[1,3,4]*ESAT-ETC, KU Leuven - Group T Leuven Campus, Belgium*
[3,4]*ESAT-STADIUS, KU Leuven, Belgium*
[1,2]*Addis Ababa Institute of Technology, Addis Ababa University, Ethiopia*
[1]bisratderebssa.dufera@kuleuven.be, [2]eneyew_a@yahoo.com, [3]koen.eneman@kuleuven.be,
[4]toon.vanwaterschoot@esat.kuleuven.be

***Abstract*—Clicks are localized degradation that affect most archived audio media. Click degradation are objectionable to the listener and should be suppressed to make the audio acceptable. The use of linear prediction (LP) modeling for the restoration of audio signal that has been corrupted by click degradation has been extensively researched. However, it is hampered by the need of a pitch predictor and by its poor performance for voiced speech and music. High-order sparse linear prediction has been shown to offer better representation of voiced speech and music over conventional linear prediction. In this paper, the use of $\ell 1$-norm and $\ell 0$-norm regularized high-order sparse linear prediction is proposed for restoration of audio signal that is corrupted by click degradation that can work equally well for speech and music without a priori information of the type of signal. High-order sparse linear prediction is used to obtain a better model of the spectral envelope and harmonics in the presence of click degradation and background noise. Evaluation with clean speech and music shows that the proposed method achieves SNR improvement from 3dB to 5dB over conventional LP approach for a wide range of click durations. Tests with speech and music corrupted by background noise in addition to click degradation show that the proposed method achieves a better SNR than the restoration of click degraded speech and music that is not corrupted by background noise using conventional LP. Perceptual evaluation of audio quality (PEAQ), used to estimate the subjective quality audio, shows that the proposed method performs better than conventional LP methods in terms of perceived quality of the restored audio by a listener. A computational requirement analysis shows that even though the proposed method is not real-time, it only takes 2 to 3 times the duration of the frame being restored on a present day general-purpose processor.**

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

I. Introduction

The term 'click' according to [1] refers to "finite duration artifacts which occur at random positions in an audio signal". These are due to damages on the physical medium [2]. Clicks can be modeled as an additive or as a replacement degradation. An additive model, where the click degradation is assumed to be added to the underlying audio signal, has been shown to be acceptable for most surface defects in recording media, such as dust, dirt and small scratches [1]. A replacement model, where the degradation replaces the signal entirely for some short period of time, may be applicable for breakages and large surface scratches which may completely destroy the underlying signal information.

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. The research leading to these results has received funding from the KU Leuven Internal Funds C2-16-00449 and VES/19/004, and the European Research Council under the European Unions 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.

Generally, restoration of click degraded audio can be seen as missing sample estimation if the underlying signal during the occurrence of the click is assumed to be lost and the location of the click degradation is known. To avoid any undesirable distortion on the sample values that are not affected by click degradation a two step approach is usually followed [1]: identification of degraded signal samples and estimation of the underlying sample values for the click degraded samples. In this paper, only the estimation of the underlying click degraded samples is considered assuming the location of the click degraded samples is known a priori.

The Least Square (LS) Autoregressive (AR) estimation [1] uses the minimum square error criterion assuming that the excitation signal has a Gaussian distribution. The click degradation are assumed to be mutually independent zero-mean Gaussian process. The click degraded samples can then be obtained from a priori knowledge of the LP coefficients of the audio signal, the undegraded samples and the location of the click degraded samples.

The performance of the LS Autoregressive (AR) interpolator is limited by the fact that the LP coefficients of the undegraded audio signal are not known a priori. Janssen et al. [3] proposed to iteratively solve for both the LP coefficients and the missing samples through the minimization of the $\ell 2$-norm of the residual, i.e. the difference between the original undegraded signal and the predicted signal as a function of the unknown LP coefficients and the unknown samples. Even though such modeling works well for unvoiced speech [4], it is not a good model for music and voiced speech, where the excitation is quasi-periodic and spiky [3]. For voiced speech and music, the minimization of the $\ell 2$-norm of the residual puts more emphasis on the periodic peaks of the residual [5]. As a result, it trades off spectral envelope estimation accuracy against estimating the harmonics [5]. A better decoupling between the spectral envelope and pitch harmonics has been reported by using high-order sparse linear prediction (HOSpLP) [4], [6], [7].

Another limitation of Janssen's method is that it needs a pitch predictor to estimate long-term correlation. In [8], a sparse linear prediction approach was proposed that minimizes the $\ell 2$-norm to jointly model the long-term and short-term correlations. In [9] the joint optimization of the formant and pitch predictors has been proposed. It poses the estimation of the formant and pitch filter as a single LP problem given the a priori knowledge of the intermediate residual signal, i.e. the output of prediction error filter. It is an iterative optimization algorithm one where the intermediate residual signal of a previous iteration is used to jointly estimate the formant and pitch predictor filters. These methods however require a priori knowledge of the pitch period.

For musical sounds or tonal audio for which the signal contains a finite number of dominant frequency components, the LP model is

much less popular than in speech analysis as the generation of musical sounds is dependent on the instruments used [5]. This makes it hard to use a generic audio signal generation model [5]. In addition, each polyphonic audio signal should be modeled using multiple source-filter models, which seems to be rather impractical [5]. In the absence of noise, by using a model order which is twice the number of tonal components LP can be used to estimate the spectral peaks. In practice, noise is always present that may be due to imperfections in the tonal behavior, signal that is not tonal in nature, finite precision arithmetic, finite-length data windowing or noise in general. Therefore, such LP signal estimates are very often poor. In [5] extensive simulations were conducted to assess the performance of conventional and alternative LP models for tonal audio analysis in the presence of noise. It was reported that high-order all-pole models are better suited to the audio LP problem albeit being impractically complex in many applications. The high-order all pole method used in [5] minimize the $\ell 2$-norm of the residual to obtain the LP coefficients while the HOSpLP methods use sparsity of the residual and the coefficient vector in the optimization problem.

In [10] $\ell 1$-norm regularized HOSpLP was used for the restoration of click degraded audio. It was reported that the use of $\ell 1$-norm regularized HOSpLP coefficients in the Janssen algorithm provided a significant improvement over conventional LP and joint-optimization based LP. However, the HOSpLP coefficients were obtained by using the $\ell 1$-norm and the $\ell 0$-norm was not investigated. Furthermore, the noise robustness of the $\ell 1$-norm regularized HOSpLP coefficients was not investigated.

In this paper we move forward, proposing a novel method for the restoration of audio signal that is corrupted by click degradation that works for both speech and music without a priori information of the type of audio signal. It uses high-order sparse linear prediction models with different levels of sparsity ($\ell 0$-norm and $\ell 1$-norm of the coefficient vector) to estimate a high-order all-pole LP coefficients. The proposed method has several advantages, one of which is that no segmentation and annotation is needed when presenting the click degraded signal to the method. This will significantly decrease the need for manual annotation and segmentation needed for practical application.

The contribution of this paper is two fold. First, we extend the use of HOSpLP for the restoration of click degraded audio by using both $\ell 1$-norm and the $\ell 0$-norm of the coefficient vector into the optimization problem. Second, we investigate the noise robustness of these different LP models for the restoration of audio signal that is corrupted by click degradation by using the Janssen algorithm.

The organization of the paper is as follows. Section II formally describes conventional LP and HOSpLP with different levels of sparsity.Section III discusses the proposed method. Section IV describes the data used, the type of click degradation and the performance measure used. In section V the results are presented and discussed in comparison with conventional LP and pitch-prediction based joint optimization. Finally, section VI presents additional discussions and concludes the work.

II. Linear Prediction

The LP coefficient vector, $\mathbf{a}$, can be obtained from a set of observed samples $\mathbf{x} = \big[x(N_1), ..., x(N_2)\big]^T$ by the following optimization problem [4]:

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

where,

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

N_1 and N_2 are the start and end indexes of the observed frame $\mathbf{x}$,

M is the order of the prediction filter,

γ is the regularization parameter that determines by how much the sparsity of the LP coefficient vector and residual contribute to the optimization problem.

The ℓp-norm $||.||_p$ is defined as

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

In conventional LP, the $\ell 2$-norm is used, i.e. $p = 2$. In addition, no a priori information about the coefficient vector is assumed, i.e. $\gamma = 0$. Furthermore, the prediction order is usually set to a small value corresponding to twice of the number of formant frequencies to be modeled.

A. $\ell 1$-norm regularized HOSpLP

Sparse linear prediction can be used to decrease the emphasis on the quasi periodic peaks of the residual in LP [4]. To achieve this, the sparsity of the residual as well as the sparsity of the coefficient vector can be used. To measure sparsity the '$\ell 0$-norm' is a natural candidate, however it is NP hard as it leads to a combinatorial problem. The $\ell 1$-norm, $p = 1$, has been used as a convex relaxation of the '$\ell 0$-norm' [4] to alleviate this problem. That is by setting $k = 1$, $\gamma \neq 0$ and using a high-order LP, the short-term predictor and the long-term predictor can be jointly estimated [4]. The requirement of sparsity in the coefficient vector can be attributed to the fact that a cascade of the long-term and short-term predictor filters leads to a filter that has few non-zero coefficients [11].

It has been shown in [4] that for music and speech, the use of the $\ell 1$-norm and high-order linear prediction outperforms conventional LP in spectral envelope estimation, the sparsity of the prediction coefficients and the sparsity of prediction residual. It has also been shown in [10] that the use the $\ell 1$-norm HOSpLP for the restoration of click degraded audio outperforms conventional LP approaches.

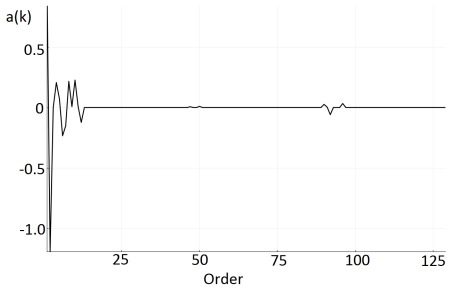
B. $\ell 0$-norm regularized HOSpLP

The a priori knowledge of the structure of the coefficient vector resulting from the cascading of the long-term and short-term predictor filters can also be incorporated as the following optimization problem:

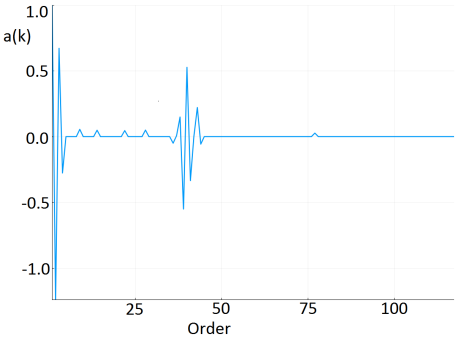
$$\mathbf{a} = \arg\min_{\mathbf{a}} \big|\big|\mathbf{x} - \mathbf{X}\mathbf{a}\big|\big|_2^2 \quad s.t. \quad \big|\big|\mathbf{a}\big|\big|_0 \leq Q \tag{3}$$

where Q is the sum of the order of the long-term and the short-term prediction filter.

In this formulation, no a priori structure is imposed on the coefficient vector except that the coefficient vector has a fixed maximum number of non-zero coefficients. As such, it can put emphasis on the formant filter coefficients if the frame is composed of speech and on the tonal components if the frame is composed of music. To illustrate this in Fig. 1 a plot is shown of the coefficient vector resulting from solving (3) with Q set to 16 for vowel and music. It is observed that the coefficient vector obtained for vowel by solving (3) has more of the non-zero coefficient values for the short term predictor while having very few non-zero coefficient values corresponding to the long-term predictor. For music on the other hand, the coefficient vector has few non-zero coefficients values for



(a) Vowel



(b) Music

Figure 1: Coefficient vector resulting from using $\ell 0 - norm$ regularized HOSpLP to obtain high-order all pole model.

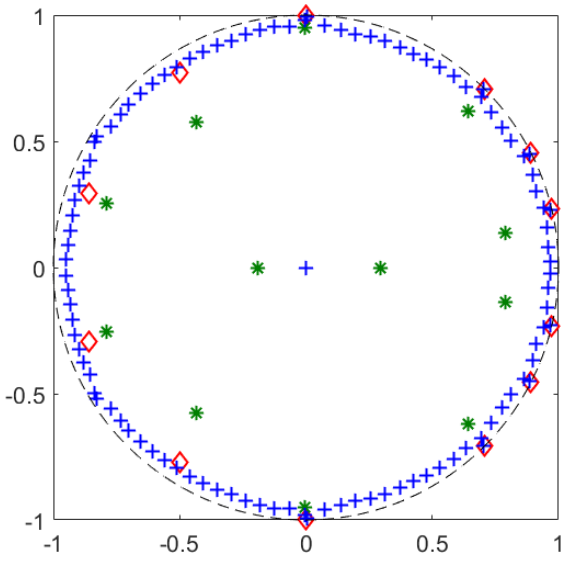


Figure 2: Pole plot of click degraded tonal audio. $\diamond$ represent actual pole locations for the original tonal audio. $*$ are the poles of a filter computed using conventional LP for the click degraded noisy tonal audio. $+$ are the poles of the $\ell 0$-norm regularized HOSpLP filter solution to (3) with 16 non-zero coefficients for $M = 128$ for the click degraded noisy tonal audio.

the short-term predictor while having more non-zero coefficients for the long-term predictor distributed over the whole coefficient vector length. As the location of these non-zero coefficients is neither incorporated into equation (3) nor dependent on a pitch predictor, a priori information regarding the type of signal is not needed.

To illustrate that the coefficients solved via (3) model the different tones in the tonal audio, an audio signal with tonal components at 300Hz, 600Hz, 1000Hz and 2000Hz was synthetically constructed, white noise was added so that the signal SNR is 5dB and a 5msec segment of the signal was artificially degraded with clicks. The conventional LP and $\ell 0$-norm regularized HOSpLP filter coefficients were obtained by the Levinson-Durbin algorithm and solution to (3) respectively. A pole plot of the conventional LP and $\ell 0$-norm regularized HOSpLP filter is shown in Fig. 2. It is seen from Fig. 2 that in the presence of noise and click degradation, the poles of the conventional LP filter drift away from the actual poles. On the other hand, the poles of the $\ell 0$-norm regularized HOSpLP that lie on the unit circle are accurate in the presence of noise and click degradation. As such, multiple tones in the music signal are represented more accurately by using the $\ell 0$-norm regularized HOSpLP. The other poles are due to the high-order of the $\ell 0$-norm regularized HOSpLP filter.

Given that problem (3) is non-convex, its relaxation (LASSO), i.e. (1) with $p = 2$ and $k = 1$, is typically solved instead [12]. Nevertheless, proximal gradient methods can solve (3) if a good initialization is given, e.g., the solution of LASSO [12]. In recent work, Antonello et al. [12] developed the StructuredOptimization package for Julia programing language that can solve (3) in a reasonable time.

III. Proposed Method of Restoration

In this paper a novel method is proposed for the restoration of click degraded audio signals that works for speech, tonal audio and music that uses high-order sparse linear prediction coefficients in the iterative Janssen algorithm for estimating the missing samples. The Janssen algorithm is used as a framework for the implementation of the different high-order sparse linear prediction based restoration approaches [10]. The LP coefficient vector is calculated with the proposed $\ell 1$-norm regularized HOSpLP and $\ell 0$-norm regularized HOSpLP coefficients as well as, for comparison, with conventional LP and with a Joint optimization of the LP coefficients.

- **Conventional LP**: obtained via the Levinson-Durbin algorithm.
- **Joint optimization of linear predictors**: the estimation of the short-term and long-term predictor filter is formulated as a single LP problem given the a priori knowledge of the intermediate residual signal after the inverse formant filter [9].
- **$\ell 1$-norm regularized HOSpLP**: the alternating direction method of multipliers (ADMM) algorithm for solving $\ell 1$-norm regularized linear regression problem [13] is used to obtain the HOSpLP coefficients [10].
- **$\ell 0$-norm regularized HOSpLP**: In this approach (3) is solved via the StructuredOptimization Julia package to obtain the $\ell 0$-norm regularized HOSpLP coefficients.

IV. Data used, Click noise model and Performance Measures

A. Data used

To fairly assess the restoration performance of the proposed methods the experiments were conducted using the following datasets:

- Speech: ten male and ten female speech of different speakers from the Voxforge dataset [14]; and
- Music: ten segments consisting of instrument audio, male and female singing voice signals from the Sparse Models, Algorithms and Learning for Large-scale data (SMALL) dataset [15].

Table I: Experiment Parameters

No	Description	Value
1	Sampling frequency	8 kHz
2	Frame size	256 samples
3	Conventional LP order	12
4	HOSpLP order	128
5	Number of pitch taps	3
6	Click duration	0.25 msec - 10 msec

Each signal is normalized to have comparable degradation among all signals.

B. Click Degradation Model

Usually, the start, duration and amplitude of each click degradation is modeled probabilistically. Different probability distributions for the time between impulses and for their amplitudes can be used [1], [16]. In this work the location of click degradation was set randomly and the samples during the occurrence of click were replaced with zero-mean Gaussian noise to obtain a click degraded signal.

C. Performance Measures

To evaluate the restoration performance of the methods Signal-to-noise ratio (SNR) and perceptual evaluation of audio quality (PEAQ) are used. The SNR of the click restoration is computed for the click degraded samples only.

$$SNR(\mathbf{s}, \hat{\mathbf{s}}) = 10 * \log \frac{||\mathbf{s}||^2}{||\mathbf{s} - \hat{\mathbf{s}}||^2} \tag{4}$$

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

PEAQ is used to assess the subjective quality of the restored audio signal. It predicts the basic audio quality of a signal with respect to a reference signal by modeling the psychoacoustic properties of the human auditory system. It has a range of 0 to -4: 0 representing imperceptible distortion while -4 means very annoying distortion. PEAQ has been used for the assessment of click degraded signal restoration in [17].

V. Results

The Janssen algorithm [10] was used to restore the audio signal that is artificially click degraded using the four LP based restoration methods discussed in Section III. The SNR and PEAQ were averaged over all the audio data for each click duration. Table I lists some of the experiment parameters.

A. Comparison of Conventional LP, Joint optimization of LP coefficients and HOSpLP

A comparison of restoration of click degraded audio using the four methods listed in section III, is shown in Fig. 3 and Fig. 4 for speech and music respectively.

Restoration by using $\ell 0$-norm regularized HOSpLP offers the best (i.e. highest SNR) over most click degradation lengths except for very short clicks. The SNR obtained with joint optimization of LP coefficients is seen to be lower than $\ell 0$-norm regularized HOSpLP. This decrease in performance of the joint optimization of LP coefficients can be attributed to the fact that the iterative combined approach does not guarantee that the overall error decreases monotonically over the iterations. It only guarantees a mean-square error that is never worse than a conventional sequential solution which

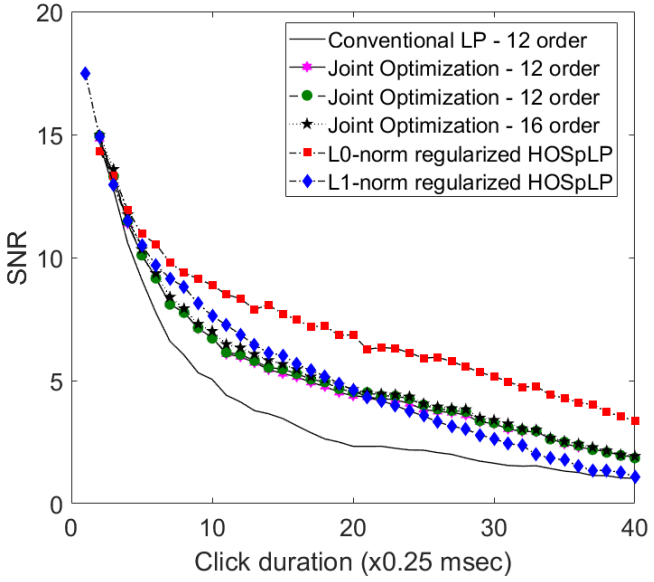


Figure 3: SNR of the restored signal for speech using conventional LP, Joint optimized LP, $\ell 1$-norm and $\ell 0$-norm regularized HOSpLP.

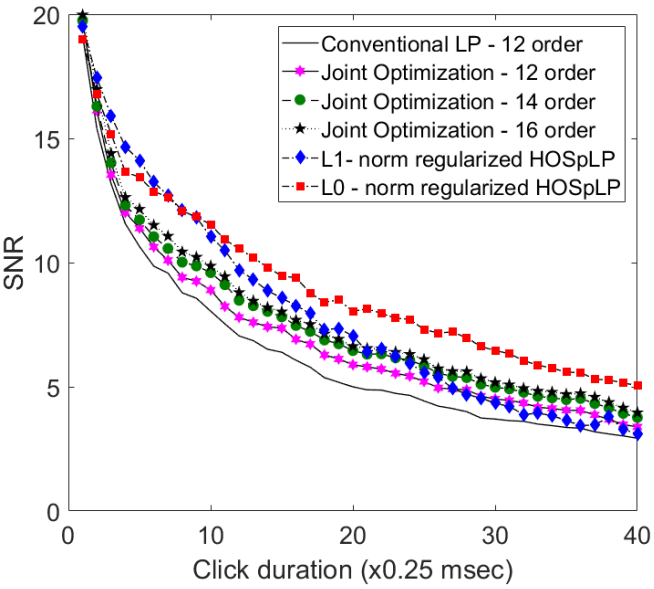


Figure 4: SNR of the restored signal for music using conventional LP, Joint optimized LP, $\ell 1$-norm and $\ell 0$-norm regularized HOSpLP.

is limited by the quasi-periodic nature of the intermediate residual signal [9]. The joint optimization of LP coefficients is observed to perform better than $\ell 1$-norm regularized HOSpLP for moderate to long click durations.

B. Noise Robustness

To assess the noise robustness of the proposed methods, additive white noise was added so that the SNR of the signal is 10dB, 20dB and 30dB. The four restoration methods were then used to remove the click degradation in the presence of background noise. Fig. 5 shows the SNR of the restored noisy signal for male speech.

It is seen that even though the performance of all the restoration approaches decrease with the addition of noise, the degradation in

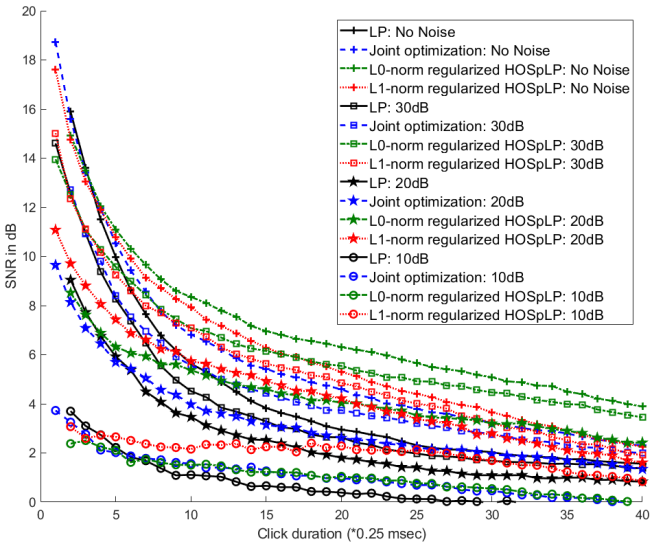


Figure 5: SNR of the restored signal for male speech in the presence of background noise.

Table II: Computational time needed to process a frame of length 32 msec

Method	**Time taken (in msec)**
Joint optimization LP	42.3
$\ell 1$-norm regularized HOSpLP	51.2
$\ell 0$-norm regularized HOSpLP	86.7

performance is graceful. It is also seen that the $\ell 0$-norm regularized HOSpLP performs better for high-SNR background noise cases, while $\ell 1$-norm regularized HOSpLP method seems to perform better for the low-SNR cases (10 dB and 20 dB).

C. Perceptual evaluation of audio quality

Fig. 6 and 7 show the PEAQ numbers obtained for speech and music for each of the four approaches without background noise. It is seen that both $\ell 1$-norm regularized and $\ell 0$-norm regularized HOSpLP based restoration achieve better (i.e. higher) PEAQ as compared to conventional LP and the joint optimization approach for both speech and music. While the $\ell 0$-norm regularized HOSpLP based restoration achieves the highest PEAQ for speech, the $\ell 1$-norm regularized HOSpLP based restoration achieves the highest PEAQ for music.

D. Computational complexity

The three methods proposed are iterative; therefore, it is difficult to mathematically derive their computational complexity. The computational complexity can nevertheless be estimated by measuring the time needed to process 32 msec of data. This value is averaged over all the datasets. Table II shows the processing time needed by a Core-i7-4510U dual core CPU running the Windows 10 Professional operating system and using Julia version 0.6.2.

It is seen that none of the methods are real-time. However, the $\ell 1$-norm regularized HOSpLP approach is only slightly slower than joint-optimization of LP coefficients approach while the $\ell 0$-norm regularized HOSpLP approach is only twice slower than joint-optimization of LP coefficients approach.

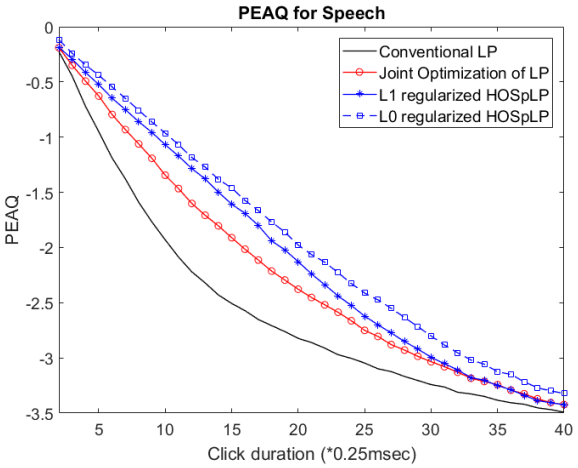


Figure 6: PEAQ for speech.

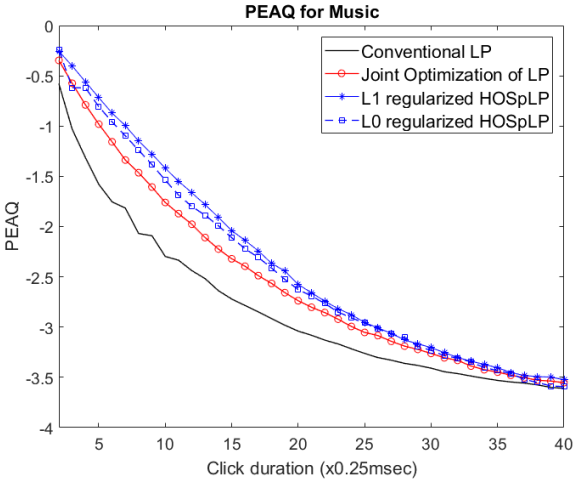


Figure 7: PEAQ for music.

VI. Conclusion

In this paper a high-order linear prediction based approach is proposed for the restoration of audio corrupted by click degradation working for both speech and tonal audio without a priori knowledge about the type of signal or pitch period. The proposed method achieves an improvement in SNR and PEAQ over conventional LP and joint optimization based LP coefficients for all considered speech and audio data types. Even though both the $\ell 1$-norm and $\ell 0$-norm regularized HOSpLP based restoration methods are not real time the processing takes only 2 to 3 times the duration of the frame in consideration on a present-day general-purpose processor. Considering the application at hand, which is the restoration of archived audio media, the processing time is not expected to be a significant limitation.

Only artificial click degradation was considered in this paper. Therefore, the performance of the methods should also be assessed under real click degradation. However, given that the samples during the occurrence of the click are discarded before restoration, it is expected that the results obtained in this paper will be valid also for real click degraded signals provided that the location of the clicks are known beforehand. In practice, the location of the click is not known a priori though, and therefore click detection methods are needed.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on January 11,2021 at 16:30:16 UTC from IEEE Xplore. Restrictions apply.

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] D. Giacobello, M. G. Christensen, J. Dahl, 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.

[8] 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.

[9] P. Kabal and R. P. Ramachandran, "Joint optimization of linear predictors in speech coders," *IEEE Trans. On Acoust., Speech and Signal Processing*, vol. 37(5), p. 642 650, 1989.

[10] B. D. Dufera, K. Eneman, and T. van Waterschoot, "Missing sample estimation based on high-order sparse linear prediction for audio signals," in *26th European Signal Processing Conference, EUSIPCO 2018, Roma, Italy, September 3-7, 2018*, pp. 2464–2468, 2018.

[11] 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.

[12] N. Antonello, L. Stella, P. Patrinos, and T. van Waterschoot, "Proximal gradient algorithms: Applications in signal processing," *arXiv:1803.01621*, March 2018.

[13] 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.

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

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

[16] 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.

[17] M. Niedwiecki, M. Cioek, and K. Cisowski, "Elimination of impulsive disturbances from stereo audio recordings using vector autoregressive modeling and variable-order kalman filtering," *IEEE Trans. Audio Speech Lang. Process.*, vol. 23(6), pp. 970–981, June 2015.