FREQUENCY-DOMAIN SINGLE-CHANNEL INVERSE FILTERING FOR SPEECH DEREVERBERATION: THEORY AND PRACTICE

Ina Kodrasi, Timo Gerkmann, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics, Oldenburg, Germany

ABSTRACT

The objective of single-channel inverse filtering is to design an inverse filter that achieves dereverberation while being robust to an inaccurate room impulse response (RIR) measurement or estimate. Since a stable and causal inverse filter typically does not exist, approximate time-domain inverse filtering techniques such as single-channel least-squares (SCLS) have been proposed. However, besides being computationally expensive and often infeasible, SCLS generally leads to distortions in the output signal in the presence of RIR inaccuracies.

In this paper, a theoretical analysis is initially provided, showing that the direct inversion of the acoustic transfer function in the frequency-domain generally yields instability and acausality issues. In order to resolve these issues, a novel frequency-domain inverse filtering technique is proposed that incorporates regularization and uses a single-channel speech enhancement scheme. Experimental results demonstrate that the proposed technique yields a higher dereverberation performance and has a significantly lower computational complexity compared to the SCLS technique.

Index Terms— acoustic single-channel inversion, stability, causality

1. INTRODUCTION

Speech dereverberation techniques are crucial for improving speech intelligibility, perceptual speech quality, and the performance of automatic speech recognition systems in reverberant environments [1, 2]. When the room impulse response (RIR) between the source and the microphone can be measured or estimated, an inverse filter can be designed to invert the RIR, and hence, achieve dereverberation [3, 4, 5]. Although such an approach sounds attractive in theory, designing and applying an inverse filter results in several drawbacks in practice. Acoustic transfer functions are generally mixed-phase functions [3] and therefore a stable and causal inverse filter does not exist. Furthermore, the inverse filter is typically designed using an inaccurate RIR, where inaccuracies can arise from temperature variations [6], spatial mismatch [7], or the sensitivity of blind system identification methods to near-common zeros and interfering noise [8]. This will commonly lead to distortions in the output signal [7].

In order to overcome some of these issues, alternative techniques such as single-channel least-squares (SCLS) [4] and homomorphic inverse filtering [4, 9] have been investigated. In SCLS, the squared error between the output of the system and a desired response is minimized. In homomorphic inverse filtering, the RIR is first decomposed into a minimum-phase component and a maximum-phase

component. An exact time-domain inverse filter is designed for the minimum-phase component, whereas the maximum-phase component is only approximately inverted using a delay and truncation. In a comparative study between SCLS and homomorphic inverse filtering in [4], it has been concluded that SCLS yields a superior dereverberation performance. However, the SCLS inverse filtering technique still results in several drawbacks in practice. The RIR can be several thousand taps long, resulting in a computationally complex and often infeasible inverse filter design [10]. Furthermore, the output signal is often distorted when the SCLS filter is designed using an inaccurate RIR [10].

In this paper, first a theoretical analysis is provided, showing that the direct inversion of the acoustic transfer function in the frequency-domain typically yields instability and acausality issues, giving rise to undesirable tones and pre-echoes. A frequency-domain inverse filtering technique is then proposed, which incorporates regularization to reduce the tones and uses a single-channel speech enhancement scheme to reduce the pre-echoes.

2. PROBLEM FORMULATION

Consider a single-channel acoustic system, where the reverberant microphone signal $x(n)$ at time index n arises from the convolution of the clean speech signal $s(n)$ with the time-invariant RIR $h(n)$ of length L_h, i.e.,

$$x(n) = s(n) * h(n). \tag{1}$$

In order to recover the clean speech signal, an inverse filter $g(n)$ should be designed such that

$$h(n) * g(n) = d(n), \tag{2}$$

with

$$\begin{cases} d(n) & = & 1 \text{ at } n = 0, \\ d(n) & = & 0 \text{ elsewhere.} \end{cases} \tag{3}$$

In [3] it has been experimentally validated that acoustic transfer functions are typically mixed-phase functions, such that a causal and stable inverse filter does not exist. Alternatively, the SCLS technique [4] has been proposed, which aims at designing an approximate inverse filter. In SCLS, a time-domain filter $\mathbf{g}$ of length L_g is designed, with $\mathbf{g} = [g(0)\ g(1)\ \dots\ g(L_g - 1)]^T$, by expressing (2) in terms of a matrix/vector multiplication as

$$\mathbf{Hg} = \mathbf{d}, \tag{4}$$

with $\mathbf{H}$ being the $(L_h + L_g - 1) \times L_g$-dimensional convolution matrix of $h(n)$ and $\mathbf{d}$ being the $(L_h + L_g - 1) \times 1$-dimensional desired response vector $\mathbf{d} = [1\ 0\ \dots\ 0]^T$. Minimizing the least-squares error $J_{\text{SCLS}} = \|\mathbf{Hg} - \mathbf{d}\|_2^2$ yields the SCLS filter

$$\mathbf{g}_{\text{SCLS}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{d}. \tag{5}$$

Designing and using $\mathbf{g}_{\text{SCLS}}$ however encounters several drawbacks in practice. The true RIR is generally not available and should either be measured beforehand or estimated [11]. Independently of

whether the RIR is measured or estimated, there is typically a mismatch between the true RIR $h(n)$ and the available RIR $\hat{h}(n)$ [6, 7, 8]. The SCLS filter designed using the inaccurate $\hat{h}(n)$ may fail to achieve dereverberation, leading to distortions in the output signal [10]. Furthermore, $\hat{h}(n)$ can be several thousand taps long, resulting in a computationally complex, numerically unstable, and often infeasible filter design [10], due to the multiplication and inversion of the high-dimensional convolution matrix in (5).

In order to avoid the inversion of the convolution matrix, a frequency-domain inverse filter can be designed instead. Consider the STFT-representation of the signal model in (1), i.e.,

$$X(k,l) = S(k,l)H(k), \qquad (6)$$

with $k = 0, \ldots, K-1$, denoting the frequency index and l denoting the frame index. We assume that $K \geq L_h$, such that the convolution in (1) can be accurately represented by multiplication in the STFT-domain. Given an estimate $\hat{H}(k)$ of the acoustic transfer function $H(k)$, an inverse filter $G(k)$ can be directly designed as

$$G(k) = \frac{1}{\hat{H}(k)}. \qquad (7)$$

In this paper, we will first analyze the instability and acausality issues associated with $G(k)$ and show that applying this inverse filter will typically give rise to undesirable tones and pre-echoes. A fast frequency-domain inverse filtering technique is then proposed, which aims at realizing a stable approximate inverse filter to avoid the tones as well as at reducing the pre-echoes arising due to acausality.

3. FREQUENCY-DOMAIN CHANNEL INVERSION

3.1. Instability and acausality of $G(k)$

In this section, the fundamental concepts of analyzing the stability and causality of discrete-time systems via the $\mathcal{Z}$-transform and the region of convergence (ROC) are briefly reviewed and used to show that using the frequency-domain inverse filter $G(k)$ in (7) encounters two main drawbacks: (i) the zeros of the RIR on the unit circle yield an unstable inverse filter, causing undesirable tones in the processed signal, and (ii) the zeros of the RIR outside of the unit circle result in an acausal inverse filter, and thus in pre-echoes in the processed signal.

The $\mathcal{Z}$-transform of a sequence $p(n)$ is defined as

$$P(z) = \sum_{n=-\infty}^{\infty} p(n)z^{-n}, \qquad (8)$$

while the ROC is defined as the set of complex numbers z for which $P(z)$ converges. For a filter to be stable, the ROC must include the unit circle, i.e., $|z| = 1$, whereas for a filter to be causal, the ROC must extend outward from the outermost pole. It should be noted that the $\mathcal{Z}$-transform is unique if and only if its ROC is also specified. If a $\mathcal{Z}$-transform without a ROC is provided, an inverse $\mathcal{Z}$-transform can be determined based on whether stability or causality is desired [12].

Since $G(k)$ in (7) is related to the $\mathcal{Z}$-transform $G(z)$ of the inverse filter $g(n)$ by

$$G(k) = G(z)\big|_{z=e^{j\frac{2\pi k}{K}}}, \qquad (9)$$

analysis of the ROC of $G(z)$ can be used to determine whether $G(k)$ realizes a stable or a causal filter.

Fig. 1(a) depicts the pole-zero representation of a measured RIR from the MARDY database [13], with a zoomed in portion presented

in Fig. 1(b). As illustrated in this figure, the zeros of a typical RIR tend to cluster near the unit circle [14], lying slightly inside, outside, and on the unit circle. Hence, the inverse $G(z)$ of a typical RIR will contain poles inside, outside, and on the unit circle. Given the relation in (9), clearly the poles of $G(z)$ on the unit circle will result in an unstable inverse filter $G(k)$, causing undesirable tones in the processed signal. In the following, a simple illustrative example is presented demonstrating how the remaining poles of $G(z)$ inside and outside the unit circle affect the causality of the inverse filter $G(k)$ in (7).

Example 1: Consider the impulse response with transfer function $\hat{H}(z) = (1 - 0.5z^{-1})(1 - 2z^{-1})$. The $\mathcal{Z}$-transform of its inverse is given by

$$G(z) = \frac{1}{(1 - 0.5z^{-1})(1 - 2z^{-1})}, \qquad (10)$$

with the pole at $z = 0.5$ and at $z = 2$ being inside and outside the unit circle, respectively. The ROC of rational $\mathcal{Z}$-transforms has to be bounded by poles or extend to infinity [12], hence, for the rational $G(z)$ in (10) only 3 different ROC definitions are possible, i.e., $|z| < 0.5$, $|z| > 2$, or $0.5 < |z| < 2$. When a frequency-domain inverse filter is designed as in (7), a finite response is obtained since $G(z) \neq \infty$ for $z = e^{j\frac{2\pi k}{K}}$. Therefore it can be said that $G(k)$ in (7) realizes the filter with the $\mathcal{Z}$-transform in (10) and the ROC including the unit circle, i.e., $0.5 < |z| < 2$. The inverse $\mathcal{Z}$-transform of (10) with the ROC being $0.5 < |z| < 2$ is given by [12]

$$g(n) = -\frac{1}{3}0.5^n u(n) - \frac{4}{3}2^n u(-n-1), \qquad (11)$$

with $u(n)$ denoting the unit step function. The sequence in (11) represents a stable bilateral filter, with a causal component due to the pole inside the unit circle and an acausal component due to the pole outside of the unit circle.

Based on this example, it can be concluded that for a typical RIR with zeros outside the unit circle, the inverse filter $G(k)$ in (7) results in acausal filtering of the microphone signal, hence yielding undesirable pre-echoes.

3.2. Frequency-Domain Single-Channel Dereverberation

In order to overcome the previously described issues of instability and acausality, the frequency-domain inverse filtering technique de-
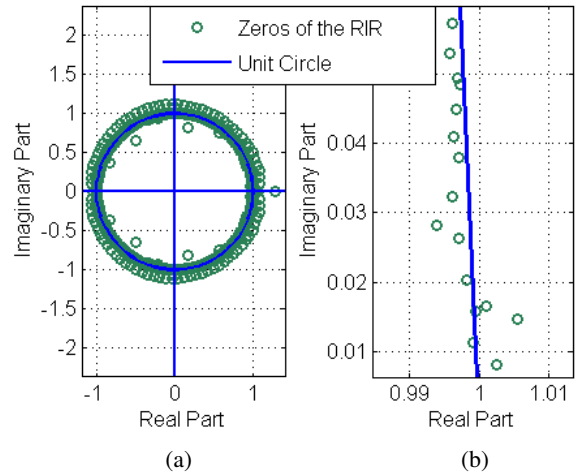

Fig. 1: Pole-zero representation of a typical RIR.

picted in Fig. 2 is proposed, where the undesirable tones and pre-echoes are reduced by incorporating regularization and a speech enhancement scheme.

3.2.1. Designing a stable approximate inverse filter

The advantage of a frequency-domain inverse filter design lies in the fact that the transfer function is being evaluated on the unit circle, hence the unstable poles can be directly manipulated. In order to obtain a stable approximate inverse filter, we propose designing a regularized filter $G_\delta(k)$, i.e.,

$$G_\delta(k) = \frac{\hat{H}^*(k)}{|\hat{H}(k)|^2 + \delta}, \tag{12}$$

where $\hat{H}^*(k)$ denotes the conjugate of $\hat{H}(k)$ and δ is a frequency-independent regularization parameter. The incorporation of regularization moves the unstable poles of $G(z)$ on the unit circle inside the unit circle, hence replacing the unstable filtering components by causal filtering components (c.f. Example 1). While regularization strongly reduces the tones in the processed microphone signal, the synthesized signal $s_\delta(n)$ exhibits pre-echoes due to the remaining acausality in $G_\delta(k)$. To remove the acausality in $G_\delta(k)$, clearly the same approach as in homomorphic inverse filtering can be applied, i.e., delaying and truncating the inverse discrete Fourier transform of $G_\delta(k)$. However, such an approach penalizes the achievable dereverberation performance and has been shown to be inferior to the SCLS technique.

3.2.2. Reducing the pre-echoes in the processed signal

In order to reduce the pre-echoes in $s_\delta(n)$, we propose to apply a single-channel speech enhancement scheme which estimates the pre-echo power spectral density (PSD) and employs this estimate to compute an enhancement gain function. In the STFT-domain, $s_\delta(n)$ can be written as

$$S_\delta(k', l') = S_\delta^d(k', l') + E(k', l'), \tag{13}$$

where $S_\delta^d(k', l')$ represents the desired speech spectral coefficients, $E(k', l')$ represents the spectral coefficients of the pre-echoes, $k' = 0, \ldots, K' - 1$, denotes the frequency index, and l' denotes the frame index. It should be noted that to account for the short-time stationarity of speech, the number of frequency bins K' in this stage is $K' \ll K$. Similarly to the commonly made assumption in spectral subtraction-based dereverberation techniques that the direct signal and late reflections are uncorrelated [15], the direct desired speech $S_\delta^d(k', l')$ and the pre-echoes spectral coefficients $E(k', l')$ can be assumed to be uncorrelated in each observation frame. Furthermore, given that the pre-echoes in $s_\delta(n)$ can be regarded as fairly non-stationary noise, we propose to estimate the PSD $\hat{\sigma}_E^2(k', l') = \mathcal{E}\{|E(k', l')|^2\}$ by employing the noise PSD estimator based on the speech presence probability with fixed priors [16], since it has been experimentally validated that this estimator exhibits a fast tracking performance for non-stationary noise. It should be noted that due to $K' \ll K$, $E(k', l')$ is significantly more stationary than $S_\delta^d(k', l')$,
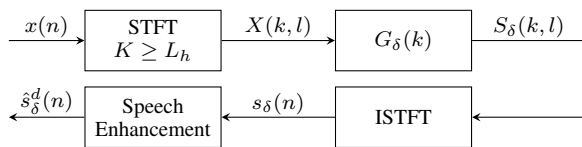

$$x(n) \rightarrow \boxed{\begin{array}{c} \text{STFT} \\ K \geq L_h \end{array}} \xrightarrow{X(k,l)} \boxed{G_\delta(k)} \xrightarrow{S_\delta(k,l)}$$

$$\hat{s}_\delta^d(n) \leftarrow \boxed{\begin{array}{c} \text{Speech} \\ \text{Enhancement} \end{array}} \xleftarrow{s_\delta(n)} \boxed{\text{ISTFT}} \leftarrow$$

Fig. 2: Schematic representation of the proposed technique.

making it plausible for the noise PSD estimator in [16] to discriminate between desired speech and pre-echoes.

Using $\hat{\sigma}_E^2(k', l')$, the a priori SNR $\hat{\xi}(k', l')$ is then estimated using the cepstral smoothing approach proposed in [17], with

$$\hat{\xi}(k', l') = \frac{\mathcal{E}\{|S_\delta^d(k', l')|^2\}}{\hat{\sigma}_E^2(k', l')}. \tag{14}$$

Finally, a Wiener gain function $G_W(k', l')$ is computed and applied to $S_\delta(k', l')$ to obtain an estimate of the desired speech spectral coefficients $\hat{S}_\delta^d(k', l')$, i.e., $\hat{S}_\delta^d(k', l') = G_W(k', l')S_\delta(k', l')$, with

$$G_W(k', l') = \frac{\hat{\xi}(k', l')}{1 + \hat{\xi}(k', l')}. \tag{15}$$

4. EXPERIMENTAL RESULTS

To investigate the performance of the proposed technique and compare it to the performance of SCLS, we have considered a measured RIR $\mathbf{h} = [h(0) \ \ldots \ h(L_h - 1)]^T$ with reverberation time $T_{60} \approx 450$ ms as the true RIR to be inverted. The sampling frequency is $f_s = 16$ kHz and the RIR length is $L_h = 7200$. The inaccurate RIR $\hat{\mathbf{h}} = [\hat{h}(0) \ \ldots \ \hat{h}(L_h - 1)]^T$ used for the inverse filter design is simulated as in [18], i.e., $\hat{h}(n) = h(n) [1 + e(n)]$, with $e(n)$ being an uncorrelated Gaussian noise sequence with zero mean and an appropriate variance, such that a normalized mismatch N_M, defined as

$$N_M = 10 \log_{10} \frac{\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2}{\|\mathbf{h}\|_2^2}, \tag{16}$$

is generated. The considered normalized mismatch values are $N_M \in \{-33 \text{ dB}, -30 \text{ dB}, \ldots, -9 \text{ dB}\}$.

For the proposed technique, the regularized inverse filter $G_\delta(k)$ is computed using $K = 16384$ and an experimentally determined regularization parameter $\delta = 10^{-2}$. The computed $G_\delta(k)$ is applied to the received microphone signal $X(k, l)$, where zero-padding of 1984 samples, a Hanning window of length K, and an overlap of 50% is used for the spectral analysis. The re-synthesized signal $s_\delta(n)$ is then processed by the proposed speech enhancement scheme, where the spectral analysis is done using $K' = 512$ and an overlap of 50%.

Since it has been shown that regularization is also effective in increasing the robustness of least-squares techniques to RIR inaccuracies [6, 19], we believe it is fair to compare the performance of the proposed technique to the performance of the regularized SCLS technique. The regularized SCLS filter is designed as $\mathbf{g}_{\text{SCLS}}^R = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I})^{-1}\hat{\mathbf{H}}^T\mathbf{d}$, with $\delta = 10^{-2}$ being an experimentally determined regularization parameter and $\mathbf{I}$ being the identity matrix. For a valid comparison between the techniques, the length of the regularized SCLS filter is also set to $L_g = K = 16384$.

For the sake of clarity, the experimental results are structured into two parts. In the first experiment, we validate the provided theoretical considerations as well as show that the proposed technique is effective in resolving the instability and acausality issues arising with the inverse filter design. These validations are done for $N_M = -33$ dB by analyzing the spectrograms of the processed signals, where undesirable tones and pre-echoes can be visualized. In the second experiment, the dereverberation performance and the perceptual speech quality of the proposed technique and the regularized SCLS technique is compared for all considered N_M. The dereverberation performance is evaluated using the speech to reverberation modulation energy ratio (SRMR) [20], whereas the perceptual

speech quality is evaluated using the objective speech quality measure PESQ [21]. The reference signal used in PESQ is the anechoic speech signal.

Experiment 1. Fig. 3(a) depicts the spectrogram of the clean speech signal, whereas Fig. 3(b) depicts the spectrogram of the received reverberant signal. To validate that the inversion of the acoustic transfer function in the frequency domain leads to instability and acausality, Fig. 3(c) depicts the spectrogram of the processed signal when directly applying the inverse filter $G(k)$ from (7). It can be seen that tones appear at several frequencies (e.g. at $f \approx 300$ Hz, $f \approx 3$ kHz), confirming that the inverse filter $G(k)$ results in instability. Furthermore, also the pre-echoes due to acausality are manifested, for example in the silence region before the speech signal starts. Applying the regularized inverse filter $G_\delta(k)$ significantly reduces the tones as shown in Fig. 3(d), however, the processed signal still contains pre-echoes. Applying the proposed speech enhancement scheme after $G_\delta(k)$ reduces the pre-echoes, as shown by the spectrogram in Fig. 3(e), where the pauses between words (e.g. at $t \approx 1.2$ s) are partly restored again. Therefore, the instability and acausality issues arising with the frequency-domain inverse filter design are resolved to a large extent by the proposed technique. Fur-
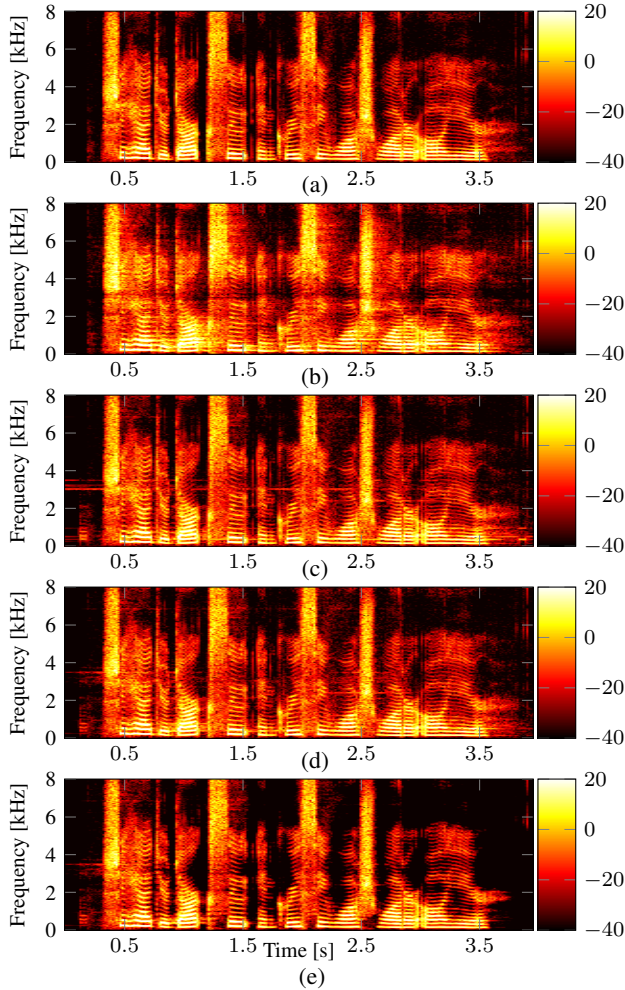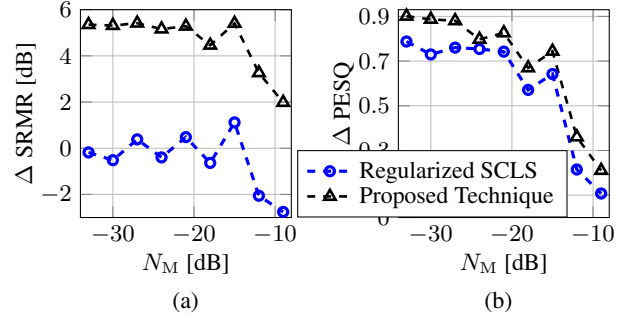

Fig. 4: Performance of the regularized SCLS technique and of the proposed technique for different N_M in terms of (a) SRMR improvement and (b) PESQ score improvement.

thermore, by comparing the spectrograms of the reverberant signal in Fig. 3(b) and of the signal processed by the proposed technique in Fig. 3(e), it can be seen that the proposed inverse filtering technique achieves dereverberation.

It should be noted that delaying and truncating the inverse discrete Fourier transform of $G_\delta(k)$ to remove the pre-echoes yields a lower performance than when using the proposed speech enhancement scheme. However, these results are omitted here due to space constraints.

Experiment 2. To analyze the dereverberation performance of the proposed technique and regularized SCLS, Fig. 4(a) presents the obtained SRMR improvement compared to the unprocessed microphone signal for all considered mismatches N_M. The presented results show that the proposed technique yields an SRMR improvement of approximately 5.5 dB for moderate mismatch values, whereas for high mismatch values such as $N_\mathrm{M} = -9$ dB the SRMR improvement is approximately 2 dB. On the other hand, the regularized SCLS technique often leads to a deterioration of the SRMR when compared to the unprocessed microphone signal, as indicated by the negative SRMR improvement values. In order to compare the perceptual speech quality, Fig. 4(b) depicts the PESQ score improvement for both techniques. It can be noticed that for all N_M, the proposed technique yields a higher perceptual speech quality than regularized SCLS. As expected, the higher amount of dereverberation achieved by the proposed technique leads to a higher perceptual speech quality.

As an indication of the computational complexity of the techniques, we also state the processing time of their MATLAB implementation, including the filter design and all the necessary processing steps to obtain a final time-domain output signal. To process a 4 s long input signal, the proposed technique requires only 0.5 s, whereas regularized SCLS requires approximately 4×10^3 s.

The presented experimental results illustrate that the proposed technique is not only significantly faster than SCLS, but it also results in a higher dereverberation performance and perceptual speech quality.

5. CONCLUSION

In this paper, it has been theoretically illustrated that the direct inversion of the acoustic transfer function in the frequency-domain yields undesirable tones and pre-echoes due to instability and acausality. A novel frequency-domain inverse filtering technique has been proposed, which incorporates regularization to reduce the tones and exploits a speech enhancement scheme to reduce the pre-echoes. Experimental results demonstrate that the proposed technique yields a higher performance and is significantly faster than the alternative SCLS inverse filtering technique.

Fig. 3: Spectrogram of (a) the clean signal $s(n)$, (b) the reverberant signal $x(n)$, (c) the signal processed with $G(k)$, (d) the signal processed with $G_\delta(k)$, and (e) the signal processed with the proposed technique.

6. REFERENCES

[1] M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1732–1745, Sept. 2010.

[2] R. Maas, E. A. P. Habets, A. Sehr, and W. Kellermann, "On the application of reverberation suppression to robust speech recognition," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 297–300.

[3] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *Journal of the Acoustical Society of America*, vol. 66, no. 1, pp. 165–169, July 1979.

[4] J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Paris, France, May 1982, vol. 7, pp. 1858–1861.

[5] E.-E. Jan and J. Flanagan, "Sound capture from spatial volumes: matched-filter processing of microphone arrays having randomly-distributed sensors," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Georgia, USA, May 1996, vol. 2, pp. 917–920.

[6] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[7] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.

[8] K. Hasan and P. A. Naylor, "Analyzing effect of noise on LMS-type approaches to blind estimation of SIMO channels: robustness issue," in *Proc. European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sept. 2006.

[9] B. D. Radlovic and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.

[10] P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005.

[11] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[12] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems (2nd ed.)*, Prentice Hall, 1996.

[13] J. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, Sept. 2006.

[14] C. P. Hughes and A. Nikeghbali, "The zeros of random polynomials cluster uniformly near the unit circle," *Compositio Mathematica*, vol. 144, pp. 734–746, May 2008.

[15] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of late reverberation effect on speech signal using long-term multiple-step linear prediction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 534–545, May 2009.

[16] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[17] C. Breithaupt, T. Gerkmann, and R. Martin, "A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, USA, Apr. 2008, pp. 4897–4900.

[18] J. Cho, D. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 718–724, Nov. 1999.

[19] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1879–1890, Sept. 2013.

[20] T. Falk and W.-Y. Chan, "A non-intrusive quality measure of dereverberated speech," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, Sept. 2008.

[21] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.