

INVERSE FILTERING AND PRINCIPAL COMPONENT ANALYSIS TECHNIQUES FOR SPEECH DEREVERBERATION

Mohamed Anouar BEN MESSAOUD, Aicha BOUZID

Department of Electrical Engineering, National School of Engineers, University of Tunis El Manar,
BP 37 Belvedere, Tunis, Tunisia

anouar.benmessaoud@yahoo.fr, bouzidacha@yahoo.fr

DOI: 10.15598/aeec.v14i2.1533

Abstract. *In this work, we present a single channel approach for early and late reverberation suppression. This approach can be decomposed into two stages. The first stage employs the inverse filter to augment the signal-to-reverberant energy ratio. The second stage uses the kernel PCA algorithm to enhance the obtained dereverberant signal. It consists in extracting the main non-linear features from the speech signal after inverse filtering. Our approach appears to be efficient mainly in far field conditions and in highly reverberant environments.*

Keywords

Inverse filtering, kernel algorithm, principal components analysis, reverberant speech.

1. Introduction

It is well known that the intelligibility of speech in natural environments is influenced by room reverberation [1], [2]. The presence of reverberation alone can degrade speech perception. Several factors contribute to the influence of reverberation on intelligibility including: source distance, room reverberation time, and the characteristics of both late reverberation and early reflections. Speech dereverberation is an active area of research designed to mitigate the influence of reverberation on speech intelligibility under monaural presentation.

Several speech dereverberation approaches have been proposed in literature. As the linear prediction residual signal includes the effects of the reverberant channel, Brandstein et al. [3] propose to use the residual and the all-pole filter resulting from prediction analysis of the reverberant speech. Allen [4] aim to

suppress the effects of reverberation by identifying the Linear Predictive Coding (LPC) parameters.

Yegnanarayana et al. [5] apply the Hilbert envelopes to indicate the strength of the peaks in the Linear Prediction (LP) residuals. Gillespie et al. [6] propose to maximize the kurtosis of the residual by using an adaptive filter for multiple microphones. In [7], [8], the authors use a spatial averaging approach. Then the prediction residual is improved by applying a temporal averaging of larynx cycles. In [9], the authors identify the magnitude and phases of sinusoidal speech model by estimation of the pitch. Consequently, the speech dereverberation is applied to deduce an equivalent equalization filter. Habets et al. [10], Lebart et al. [11] have proposed to use the spectral subtraction technique to dereverberation. They used a statistical model of the room impulse response including Gaussian noise modulated by a decay rate of an exponential function, and then the power spectral density of the impulse response can be suppressed by spectral subtraction. A similar method proposed in [12] estimates the late reflection by using a multistep LP. In [13], the authors use the subspace method to determine the inverse filter independent of the source characteristics. Nakatani et al. [14] proposed to formulate the dereverberation problem as a maximum likelihood using a hill-climbing technique. In [15], the authors optimize the scaling factors by using the minimum mean square error criterion to eliminate the late reflection components.

Principal Component Analysis (PCA) is a subcategory of subspace approaches. It is an important tool for data analysis designed to reduce multidimensional signal. The essential object is to obtain a set of orthogonal factors that describe the variance of the observations and track the new factors considered to determine the necessary features. Speech processing by PCA [16] is extensively applied as a classical multi-

variate speech processing tool. PCA has been used in different fields of science to extract relevant information from complex matrix data [17]. For speech separation, a robust extension of classical PCA by generalizing eigenvalue decomposition of a pair of covariance matrices has been proposed by [18]. It may be used in speech identification [19], speech recognition [20], speech enhancement [21], [22], and [23] and speech dereverberation [24], [25], and [26].

In this work, we propose an approach of dereverberation in the field of single-channel speech enhancement which is appropriate for real-world applications. For suppressing the late and early reverberations, we apply a single-microphone two-step approach. In the first stage, the inverse filter is applied to enhance the speech to reverberation ratio and reduce the coloration caused by the early reflections and late reverberation. The second stage consists in using the kernel principal component analysis algorithm to enhance the obtained dereverberation signal.

Our approach is tested on a corpus of speech utterances from thirty two speakers. The evaluation shows that our approach is able to reduce early reflections and late reverberation in highly reverberant environments.

The rest of the paper is organized as follows. In Section 2. we present the first stage based on the inverse filter. Then, we integrate the kernel PCA into our approach. Next, in Section 3. we provide the experimental results of our approach.

2. Proposed Approach

We propose an approach of speech dereverberation employing monaural recordings. It can be decomposed into two essential stages, as shown in Fig. 1.

In the first stage of our approach, we apply a pre-processing procedure to reduce the short-term correlation of the speech, and identify the delayed late reverberations. Then, we determine the inverse filter by maximization of kurtosis. The second stage consists of using the kernel principal component analysis algorithm to reduce the effects of the long-term reverberation.

2.1. Inverse Filter Stage

1) Pre-Processing Procedure

The pre-processing procedure consists in adjusting the reverberant speech signal by the right choice of the short order linear prediction.

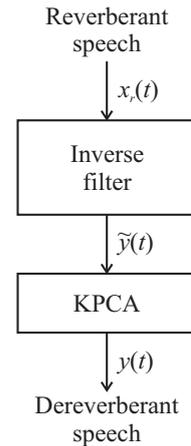


Fig. 1: Block diagram of our proposed approach.

We consider the reverberant speech signal $x_r(n)$ as the result of the convolution between the clean speech and the room impulse response.

$$\begin{aligned}
 x_r(n) &= \sum_{k=0}^N g(k)x(n-k), \\
 &= \sum_{i=0}^{Q-1} g(i)u(n-i),
 \end{aligned} \tag{1}$$

where g is the response impulse filter determined by combining the effects of the human speech production and the room impulse response, and u is the clean speech signal.

The finite impulse response filter is described as follows:

$$g(n) = \sum_{j=0}^M a(j)b(n-j). \tag{2}$$

Such a filter would produce the speech signal from the white noise $u(n)$. The reverberant speech signal can be written in vector form:

$$x_r(n) = Gu(n), \tag{3}$$

where:

$$\begin{bmatrix}
 g(0) & \dots & g(N-1) & 0 & \dots & 0 \\
 0 & g(0) & \dots & g(N-1) & \dots & 0 \\
 \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
 0 & \dots & 0 & g(0) & \dots & g(N-1)
 \end{bmatrix}$$

is a matrix G of size $(N-1) \times (Q+N)$ with Q and N are the length of the vector g and $x_r(n)$, respectively. With

$$\begin{aligned}
 g &= [g(0), \dots, g(N-1)], \\
 x_r(n) &= [x_r(n), \dots, x_r(n-Q)]^T.
 \end{aligned}$$

A bias provoked by the transfer function of the human speech production exists in estimated late elements of $G(z)$. So, we implement a linear prediction

with small-order. In [27], the authors have proposed to use an order of pre-whitening equal to 20 taps. However, in our approach, the order is arranged according to the length of the RIR. This pre-processing compensates for the bias provoked by the transfer function of the human speech production bearing in mind its convolution by the RIR. Therefore, the resulting pre-processing will be more adapted to the reverberant signal. In experiment, we have chosen for the room impulse response with reverberation time equal to 0.5 s and 0.7 s, a short linear prediction order equal to 6 and 20 taps, respectively.

After processing the short-order linear prediction, we reduce the effects of the long-term reverberation.

The long-term linear prediction is described by the following equation:

$$x_r(n) = \sum_{l=0}^L f(l)x_r(n-l-D) + er(n), \quad (4)$$

where $x_r(n)$ is the reverberant speech signal, $f(l)$ are the coefficients of filter, L is the number of the coefficients, D is the delay and $er(n)$ is the error of the speech signal.

In this step, we apply the Levinson-Durbin method to minimize the energy of the error $er(n)$.

Based on the method proposed by Kinoshita et al. [27], we apply the Wiener-Hopf equation specialized for delayed LP.

$$f = (E\{x_r(n-D)x_r^T(n-D)\})^{-1} \cdot E\{x_r(n-D)x_r(n)\}, \quad (5)$$

with

$$E\{x_r(n-D)x_r^T(n-D)\} = v_x^2 GG^T, \\ E\{x_r(n-D)x_r(n)\} = v_x^2 Gg_{late}^T,$$

where v_x^2 is the variance of white noise and

$$g_{late} = [g(D), \dots, g(L-1), 0, \dots, 0]^T.$$

As a result, we are having:

$$f = (GG^T)^{-1} Gg_{late}. \quad (6)$$

The estimated power of late reverberation is:

$$E\{(x^T f)^2\} = \|v_x g_{late}\|^2. \quad (7)$$

The linear prediction filter order, L , is a large number. Consequently, the residual speech signal each time is calculated based on samples $L+D$. Hence, the linear prediction residual speech signal is capable of providing the long-term reverberation of the signal.

2) Inverse Filtering

In order to suppress the long reverberations, we apply an inverse filtering by maximizing the kurtosis of linear prediction residual speech. As the reverberant speech signal has a lower kurtosis than that of linear prediction residual of clean speech signal, we suggest to estimate the inverse filter by using the kurtosis maximization.

Figure 2 shows the diagram for inverse filtering. We approximate the LP residual of the processed speech by the inverse-filtered LP residual of the reverberant speech.

The inverse-filtered speech signal is:

$$\tilde{y}(t) = h\hat{r}(t), \quad (8)$$

where $\hat{r}(t) = [r(t-K+1), \dots, r(t)]^T$ is the residual signal of the reverberant speech after the application of the pre-processing procedure, h is the inverse filter, and $\tilde{y}(t)$ is the inverse filtered speech signal.

The kurtosis of $\tilde{y}(t)$ is described as follow:

$$I(t) = \frac{E\{y^4(t)\}}{E^2\{y^2(t)\}} - 3. \quad (9)$$

According to the work of Gillespie et al. [6], the gradient of the kurtosis is described by the equation in below:

$$\frac{\partial I(t)}{\partial h(t)} = \frac{4(E\{\tilde{y}^2(t)\}E\{\tilde{y}^3(t)\hat{r}(t)\} - E\{\tilde{y}^4(t)\}E\{\tilde{y}(t)\hat{r}(t)\})}{E^3\{\tilde{y}^2(t)\}}. \quad (10)$$

Then, we approximate the gradient by the following equation:

$$\frac{\partial I(t)}{\partial h(t)} \approx \frac{4(E\{\tilde{y}^2(t)\}\tilde{y}^2(t) - E\{\tilde{y}^4(t)\}\tilde{y}(t))}{E^3\{\tilde{y}^2(t)\}}\hat{r}(t) = f(t)\hat{r}(t). \quad (11)$$

Then, we calculate the $E\{\tilde{y}^4(t)\}$ and $E\{\tilde{y}^2(t)\}$ to obtain the adaptive inverse filter.

In the time domain, the inverse adaptive filter can be described by the following equation:

$$h(t+1) = h(t) + \beta f(t)\hat{r}(t), \quad (12)$$

where β adapts the learning rate, and $f(t)$ is the function of the kurtosis.

However, on account of the several numbers of taps for RIR, signal is operated frame by frame into the frequency domain by using the Fast Fourier Transform (FFT) consistent with the work of Wu and Wang [28].

In the frequency domain, the Eq. (12) becomes:

$$H(n + 1) = H(n) + \frac{\beta}{P} \sum_{p=1}^P F(p)\hat{R}^*(p), \quad (13)$$

where $H(n)$ is the n^{th} iteration of h , $F(p)$, $\hat{R}^*(p)$ denote, respectively, the FFT of $f(t)$, $\hat{r}(t)$, with p is the p^{th} frame, and P is the number of frames.

Then, we obtain the inverse-filtered speech by convolving the inverse filter with the reverberant signal.

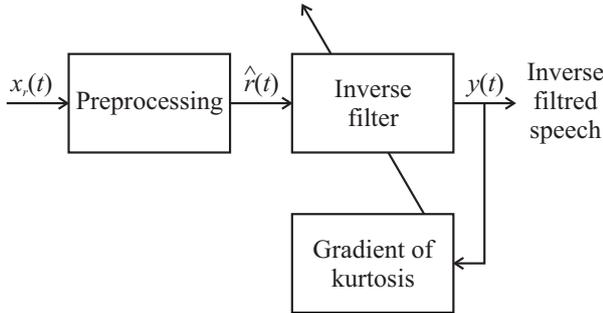


Fig. 2: Block diagram of the first stage.

In the next sub-section, we add the kernel PCA technique to enhance the processed signal.

2.2. Kernel PCA Stage

In Fig. 3, we illustrate the kernel PCA algorithm for speech dereverberation. In the KPCA technique, the speech components will be projected onto low-order

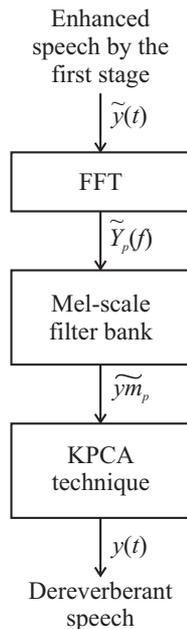


Fig. 3: Block diagram of the first stage.

while reverberant components will be projected onto high-order.

We compute the Fast Fourier Transform (FFT) of the obtained speech signal by the first stage, $\tilde{y}(t)$. We obtain the short spectrum for the clean speech and the transfer function of the frequency f at the p^{th} frame. Our goal is to determine the estimated clean speech $y(t)$.

The FFT of $\tilde{y}(t)$ can be expressed as:

$$\tilde{Y}_p(f) = Y_p(f) \cdot \tilde{H}_p(f). \quad (14)$$

Then, we calculate the log-spectrum of $\tilde{Y}_p(f)$. We obtain the following equation:

$$\tilde{Y}_{\log -p}(f) = \log(Y_p(f)) + \log(\tilde{H}_p(f)), \quad (15)$$

where $\tilde{Y}_{\log -p}(f)$, $\log(\tilde{H}_p(f))$, and $\log(X_i(f))$ are the logarithm spectra for the first stage obtained speech signal by inverse filtering, estimated clean speech signal, and transfer function, respectively.

So that, we extract the clean speech based on KPCA,

$$Y = VY_{\log}. \quad (16)$$

The eigen-vector matrix V , is derived by the eigen-value decomposition, in which the filter $V = [v^{(1)}, \dots, v^{(p)}]$ consists of the eigen-vectors corresponding to the Q dominant eigen-values. The component of the convolution reverberant signal $[v^{(p+1)}, \dots, v^{(M)}]$ is eliminated by the filtering operation in Eq. (16).

According to the Eq. (16), the $\tilde{y}(t)$ signal is represented under the assumption of non-correlation between the reverberant signal and the clean speech. Then, we apply a non-linear PCA called kernel principal component analysis (KPCA) to extract the clean speech.

Now, we detail the KPCA technique in the following subsection.

1) KPCA Technique

KPCA transforms the input logarithm spectra with nonlinear structure into a higher-dimensional feature space with linear structure, and then computes linear PCA on the mapped data. We can assume that reverberant signal will be eliminated in the high-dimensional space.

Consider the output of the mel-scale filter bank $\tilde{y}m_p$ at p^{th} frame, we define the covariance matrix as follow:

$$C = \frac{1}{P} \sum_{p=1}^P \bar{\Psi}(\tilde{y}m_p)\bar{\Psi}(\tilde{y}m_p)^T, \quad (17)$$

where Ψ is the nonlinear map, and P denotes the number of samples.

Now, we must solve the eigen-vectors v and eigen-values α .

We have:

$$\begin{aligned} \alpha V &= CV, \\ \alpha \bar{\Psi}(\widehat{y\tilde{m}}_k)^T V &= \bar{\Psi}(\widehat{y\tilde{m}}_k)^T CV \quad \text{for all } k, \end{aligned} \quad (18)$$

where $k = 1, \dots, P$.

Such as the eigen-vectors:

$$V = \sum_{j=1}^P \beta_j \bar{\Psi}(\widehat{y\tilde{m}}_j). \quad (19)$$

Replacing the Eq. (17) and Eq. (19) into Eq. (18) leads to:

$$\begin{aligned} \alpha \sum_{j=1}^P \beta_j \bar{\Psi}(\widehat{y\tilde{m}}_k)^T \bar{\Psi}(\widehat{y\tilde{m}}_j) &= \\ \frac{1}{P} \sum_{j=1}^P \beta_j \bar{\Psi}(\widehat{y\tilde{m}}_k)^T \sum_{p=1}^P \bar{\Psi}(\widehat{y\tilde{m}}_p) \bar{\Psi}(\widehat{y\tilde{m}}_p)^T \bar{\Psi}(\widehat{y\tilde{m}}_j), \end{aligned} \quad (20)$$

with $K_{jp} = \bar{\Psi}(\widehat{y\tilde{m}}_j)^T \bar{\Psi}(\widehat{y\tilde{m}}_p)$.

We formulate the equation above as:

$$\bar{\Psi}(\widehat{y\tilde{m}}_k) \cdot Cv = \frac{1}{P} \sum_j \beta_j \sum_p \bar{A}_{kp} \bar{A}_{pj}. \quad (21)$$

So we obtain

$$P\alpha K\beta = K^2\beta, \quad (22)$$

with K as the kernel matrix.

For dereverberation a test sample $\bar{\Psi}(\widehat{y\tilde{m}})$, it is projected onto the Eigen-vectors V eigenvalues. The projection is presented in terms of kernel functions based on the Eq. (19):

$$\gamma_k = (V^k)^T \bar{\Psi}(\widehat{y\tilde{m}}) = \sum_{j=1}^P \beta_j^k k(\widehat{y\tilde{m}}, \widehat{y\tilde{m}}_j). \quad (23)$$

Therefore, the projected sample in feature space is equal to:

$$O_n \bar{\Psi}(\widehat{y\tilde{m}}) = \sum_{k=1}^n \beta_k V^k, \quad (24)$$

where O_n presents the projection operator.

For dereverberation, we try to find a sample y that satisfies $\bar{\Psi}(y) = O_n \bar{\Psi}(\widehat{y\tilde{m}})$.

We consider the Gaussian kernel function:

$$K(\widehat{y\tilde{m}}_j, \widehat{y\tilde{m}}_p) = \exp\left(-\frac{\|\widehat{y\tilde{m}}_j - \widehat{y\tilde{m}}_p\|^2}{c}\right), \quad (25)$$

where c presents the variance.

The dereverberant speech is determined by an iterative update equation of y

$$y_{t+1} = \frac{\sum_{j=1}^P \eta_j k(y_t, \widehat{y\tilde{m}}_j) \widehat{y\tilde{m}}_j}{\sum_{j=1}^P \eta_j k(y_t, \widehat{y\tilde{m}}_j)}, \quad (26)$$

where $\eta_j = \sum_{k=1}^n \gamma_k \beta_j^k$.

We indicate that the pre-image y is a linear combination of the input speech $\widehat{y\tilde{m}}_j$.

The KPCA matrix was estimated only from the reverberant speech recordings. In order to perform the matrix, K , we choose $P = 1500$ frames from the training data, and we used the Gaussian kernel function. The KPCA uses 16 dimensions Mel-scale filter bank output.

3. Experiment Results

We evaluated the proposed dereverberation speech approach with the TIMIT database [29]. As we compare our approach to the state-of-the-art algorithms when the signals are sampled at 8 kHz [28] and [31] so we down-sample the speech signals from 16 kHz to 8 kHz. We use 32 English speakers of this database.

We apply the standard objective measures to determine the quality of noisy speech enhanced by our proposed approach. The reverberant speech signal is generated by convolving the room impulse response function and the clean speech with $T_{60} = 0.3, 0.5, \text{ and } 0.7$ s. In our approach, we use the mirror image model proposed in [30] to generate the RIR function. The RIR was created in a room with dimensions $6 \times 4 \times 3$ m, the microphone was installed at $4 \times 1 \times 2$ m, and the speaker was placed at $2 \times 3 \times 1.5$ m, see Fig. 4.

To evaluate the performance of our approach and to compare it with two existing methods named the Wu method [28], and MAP and variational deconvolution method (MA-VD) [31]. In [28], the authors proposed a two-step single-channel dereverberation method (Wu method) whose first step applied an adaptive inverse filtering scheme by kurtosis maximization. Then, they introduce the spectral subtraction technique in the second stage to improve the dereverberation performance for long reflections. In [31], the authors used a cost functions obtained from variational estimation and maximum posteriori to formulate a dereverberation in frequency bin.

We use the standard objective metrics namely respectively Global Signal Noise Ratio in dB (SNR),

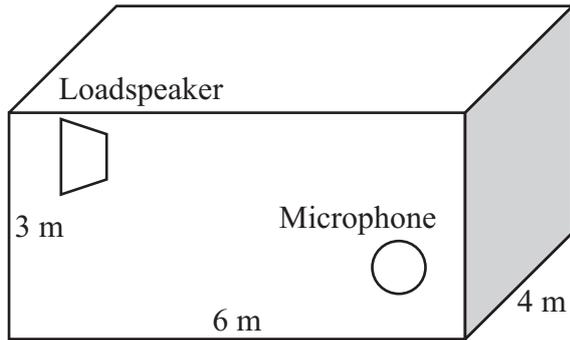


Fig. 4: Geometry of a varechoic room of the dimensions $6 \times 4 \times 3$ meters. Wall reflection coefficients are equal to 0.75. The microphone is at $4 \times 1 \times 2$ meters. The loudspeaker is at $2 \times 3 \times 1.5$ meters.

the Perceptual Evaluation of Speech Quality (PESQ) that measure the overall quality of the speech, and the Speech to Reverberation Ratio (SRR) with the three reverberation times.

Table 1 presents the results of global SNR in dB obtained by using the Wu, MA-VD, and proposed approach (Pro-App) for speech dereverberation using monaural speech recordings with the TIMIT database. We referred to the degraded reverberant speech by the word "Rev".

Tab. 1: Performance comparison of global SNR, in the presence of reverberant speech.

T_{60} (s)	Rev	Wu [28]	MA-VD [31]	Pro-App
Global SNR (dB)				
0.3	-2.87	1.84	4.77	6.08
0.5	-3.65	1.02	3.69	5.34
0.7	-4.39	0.61	2.86	3.15

It can be seen that our approach gives better SNR results than the others methods at SNR levels varied from -2.87 at -4.39 dB.

Table 2 gives the results in the case of our approach and that of Wu and Wang [28] and MA-VD [31].

Tab. 2: Performance comparison of PESQ, in the presence of reverberant speech.

T_{60} (s)	Rev	Wu [28]	MA-VD [31]	Pro-App
PESQ				
0.3	2.24	2.36	2.45	2.74
0.5	2.07	1.24	1.73	1.87
0.7	1.93	1.53	1.68	1.91

In Tab. 2, the PESQ scores show the intelligibility of our approach compared to that produced by the state-of-the-art methods. However, in the two reverberation times ($T_{60} = 0.5$ s and 0.7 s), the reverberant speech obtains the greater PESQ scores. This results show that the PESQ measure is not correlated with dereverberation.

Also, we compare the proposed approach with other methods in terms of SRR measures. The Speech to Reverberation Ratio (SRR) is a speech-based measure of reverberation that can be calculated even when the effect of the dereverberation method cannot be described in the impulse response. The SRR can be defined as:

$$SRR = 20 \log_{10} \left(\frac{\|x\|_2}{\|y - x\|_2} \right), \quad (27)$$

where x is the clean speech signal, and y is the estimated output of our dereverberation approach.

Tab. 3: Performance comparison of SRR, in the presence of reverberant speech.

T_{60} (s)	Rev	Wu [28]	MA-VD [31]	Pro-App
SSR (dB)				
0.3	-2.07	-1.46	0.68	1.42
0.5	-3.29	-2.95	-2.17	0.06
0.7	-5.08	-5.30	-5.13	-1.61

Results of the Tab. 3 indicate that our approach outperforms the two state of the art methods with SRR measures.

We observe that in the three reverberation times, the proposed approach demonstrate a greater SRR score compared to their corresponding methods. The SRR level of the reverberant speech is much lower than that of our approach. However, the two methods [28], and [31] fail to conserve its performance for $T_{60} = 0.7$ s. It can be seen that our approach outperforms the results of the compared methods when we added the KPCA technique to the first stage.

We can conclude that the proposed approach can suppress the long reflections in the first stage by applying a pre-processing procedure followed by a delayed long-term LP. The KPCA stage is used to enhance the obtained signal by the first stage. Furthermore, the proposed approach is more effective in suppressing the long reflections. However, the two state-of-the-art algorithms yield satisfactory results only when the T_{60} is short.

In order to show the enhanced speech obtained by using the proposed approach, spectrograms of the original speech, the reverberant speech, and the enhanced speech signals obtained by using the inverse filter, and the dereverberation by our approach after applying the kernel PCA are presented in Fig. 5.

The spectrogram of our proposed approach shows the inverse filter reconstructs the harmonic structure of the original speech. After the application of the kernel PCA algorithm, the spectrogram improves that the reverberation is greatly reduced in our proposed approach.

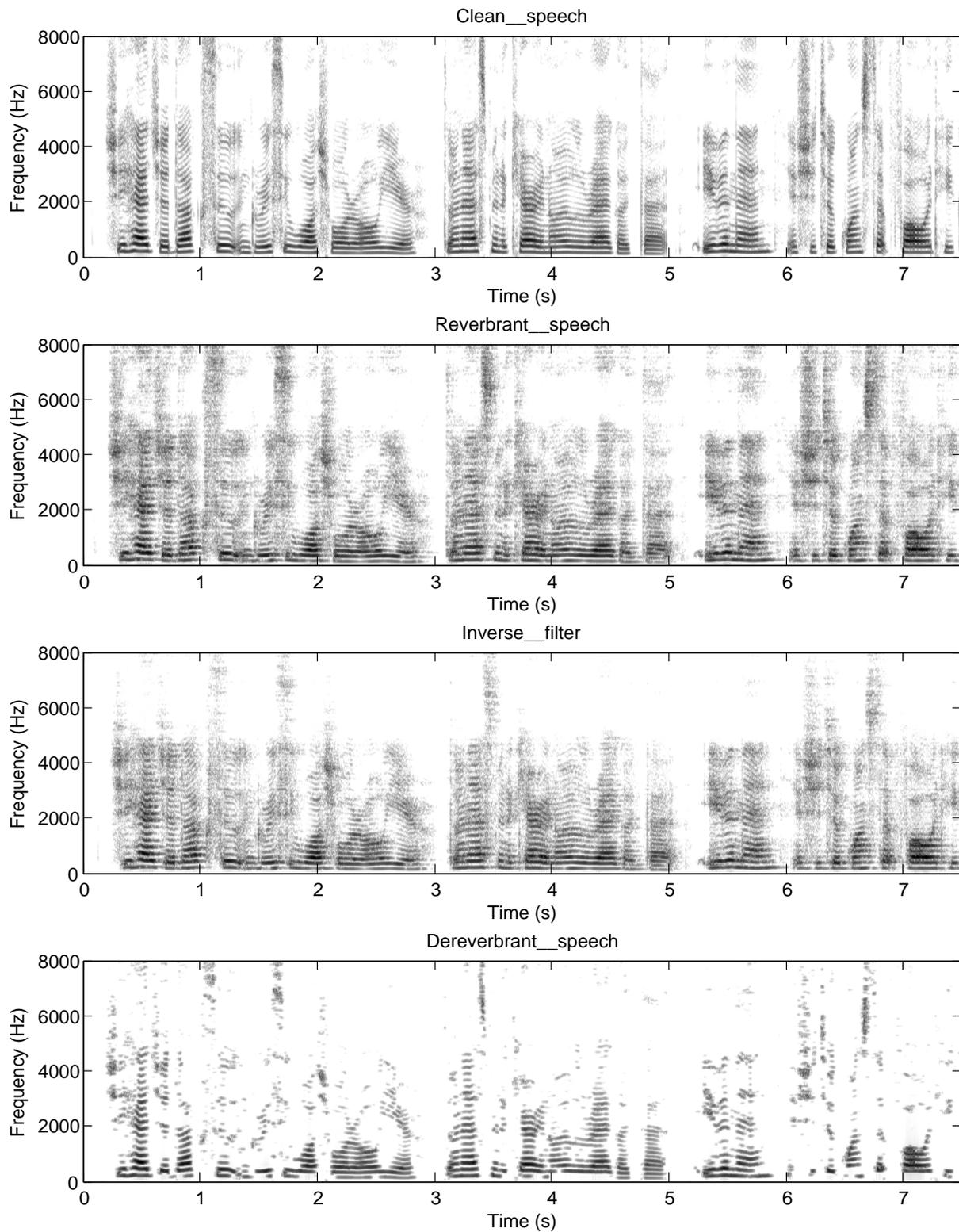


Fig. 5: Spectrograms of speech signal. (a) Spectrogram of clean speech signal. (b) Spectrogram of reverberant speech signal. (c) Spectrogram of dereverberant speech signal by inverse filter. (d) Spectrogram of enhanced dereverberant speech signal by our approach.

4. Conclusion

In this paper, we present a speech dereverberation approach of monaural recordings. The proposed approach can be summarized into two stages. First, we apply an inverse filter of reverberant speech with introducing a pre-processing procedure. Second, we propose to use the kernel principal component analysis algorithm to enhance the obtained speech by inverse filter.

The experimental results show that our approach achieves a significantly better speech quality than the state of the art algorithms and show a significant improvement in terms of these measures with minimal residual distortion. Also, the existing methods do not obtain better results for rooms with T_{60} of more than 0.5 s.

The future work consists in processing reverberant and noisy speech in real-time while keeping reasonable distortion levels.

References

- [1] HUNT, M. J. and C. LEFEBVRE. Speaker dependent and independent speech recognition experiments with an auditory model. In: *International Conference on Acoustics, Speech, and Signal Processing*. New York: IEEE, 1988, pp. 215–218. ISSN 0736-7791. DOI: 10.1109/ICASSP.1988.196552.
- [2] KLATT, D. Prediction of perceived phonetic distance from critical-band spectra: A first step. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Paris: IEEE, 1982, pp. 1278–1281. DOI: 10.1109/ICASSP.1982.1171512.
- [3] BRANDSTEIN, M. S. and S. M. GRIEBEL. *Non-linear, Model-Based Microphone Array Speech Enhancement*. New York: Springer, 2000. ISBN 978-1-4613-4656-2.
- [4] ALLEN, J. B. *Synthesis of pure speech from a reverberant signal*. US patent. US3786188.
- [5] YEGNANARAYANA, B. and P. S. MURTHY. Enhancement of reverberant speech using LP residual signal. *IEEE Transactions on Speech and Audio Processing*. 2000, vol. 8, iss. 3, pp. 267–281. ISSN 1063-6676. DOI: 10.1109/89.841209.
- [6] GILLESPIE, B. W., H. S. MALVAR and D. A. F. FLORENCIO. Speech dereverberation via maximum-kurtosis subband adaptive filtering. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Salt Lake City: IEEE, 2001, pp. 3701–3704. ISBN 0-7803-7041-4. DOI: 10.1109/ICASSP.2001.940646.
- [7] GAUBITCH, N. D. *Blind identification of acoustic systems and enhancement of reverberant speech*. London, 2006. Dissertation. Imperial College London.
- [8] GAUBITCH, N. D. and P. A. NAYLOR. Spatiotemporal Averaging Method for Enhancement of Reverberant Speech. In: *15th International Conference on Digital Signal Processing*. Cardiff: IEEE, 2007, pp. 607–610. ISBN 1-4244-0881-4.
- [9] NAKATANI, T., M. MIYOSHI and K. KINOSHITA. Single-Microphone Blind Dereverberation. *Speech Enhancement*. Berlin: Springer, 2005, pp. 247–270. ISBN 978-3-540-24039-6.
- [10] HABETS, E. A. P., S. GANNOT, I. COHEN, P. C. W. SOMMEN. Joint Dereverberation and Residual Echo Suppression of Speech Signals in Noisy Environments. *IEEE Transactions on Audio, Speech, and Language Processing*. 2008, vol. 16, iss. 8, pp. 1433–1451. ISSN 1558-7916. DOI: 10.1109/TASL.2008.2002071.
- [11] LEBART, K., J. M. BOUCHER and P. N. DENBIGH. A new method based on spectral subtraction for speech dereverberation. *Acta Acoust.* 2001, vol. 87, no. 3, pp. 359–366. ISSN 1610-1928.
- [12] KINOSHITA K., T. NAKATANI and M. MIYOSHI. Spectral subtraction steered by multi-step forward linear prediction for single channel speech dereverberation. In: *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*. Toulouse: IEEE, 2006, pp. 817–820. ISBN 1-4244-0469-X. DOI: 10.1109/ICASSP.2006.1660146.
- [13] HIKICHI, T., 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*. 2007, vol. 2007, iss. 1, pp. 1–4. ISSN 1110-8657. DOI: 10.1155/2007/34013.
- [14] NAKATANI, M., B.-H. JUANG, T. YOSHIOKA, K. KINOSHITA, M. DELCROIX and M. MIYOSHI. Speech dereverberation based on maximum-likelihood estimation with time-varying Gaussian source model. *IEEE Transactions on Audio, Speech, and Language Processing*. 2008, vol. 16, iss. 8, pp. 1512–1527. ISSN 1558-7916. DOI: 10.1109/TASL.2008.2004306.
- [15] GOMEZ, R., J. EVEN and H. SARUWATARI. Rapid unsupervised speaker adaptation robust in

- reverberant environment conditions. In: *9th Annual Conference of the International Speech Communication Association*. Brisbane: ISCA, 2008, pp. 187–190. ISBN 978-1615673780.
- [16] JOLIFFE, I. T. *Principal component analysis*. 2nd ed. New York: Springer Verlag, 2002. ISBN 978-0-387-95442-4.
- [17] MISRA, M., H. H. YUE, S. J. QIN and C. LING. Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Computer & Chemical Engineering*. 2002, vol. 26, iss. 9, pp. 1281–1293. ISSN 0098-1354. DOI: 10.1016/S0098-1354(02)00093-5.
- [18] BENABDERRAHMANE, Y., S.-A. SELOUANI and D. D. OSHAUGHNESSY. Blind speech separation for convolutive mixtures using an oriented principal components analysis method. In: *18th European Signal Processing Conference*. Aalborg: EURASIP, 2010, pp. 1553–1557. ISBN 978-3-642-03882-2.
- [19] ABOLHASSANI, A. H., S.-A. SELOUANI and D. OSHAUGHNESSY. Speech enhancement using PCA and variance of the reconstruction error model identification. In: *IEEE Workshop on Automatic Speech Recognition & Understanding*. Kyoto: IEEE, 2007, pp. 19–23. ISBN 978-1-4244-1746-9. DOI: 10.1109/ASRU.2007.4430077.
- [20] TAKIGUCHI, T. and Y. ARIKI. PCA-Based Speech enhancement for distorted speech recognition. *Journal of multimedia*. 2007, vol. 2, no. 5, pp. 13–18. ISSN 1796-2048. DOI: 10.4304/jmm.2.5.13-18.
- [21] SHINDE, V. D., C. G. PATIL and S. D. RUIKAR. Wavelet based multi-scale principal component analysis for speech enhancement. *International Journal of Engineering Trends and Technology*. 2012, vol. 3, iss. 3, pp. 397–400. ISSN 2231-5381.
- [22] VERTELETSKAYA, E. and B. SIMAK. Spectral subtractive type speech enhancement methods. *Advances in Electrical and Electronic Engineering*. 2010, vol. 8, no. 3, pp. 66–72. ISSN 1804-3119.
- [23] BEN MESSAOUD, M. A. and A. BOUZID. Speech Enhancement Based on Wavelet Transform and Improved Subspace Decomposition. *Journal Audio Engineering Society*. 2015, vol. 63, iss. 12, pp. 990–1000. ISSN 1549-4950.
- [24] GERKMANN, T. 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*. 2012, vol. 20, iss. 4, pp. 1383–1393. ISSN 1558-7916. DOI: 10.1109/TASL.2011.2180896.
- [25] KODRASI, I., S. GOETZE and S. DOCLO. Regularization for Partial Multichannel Equalization for Speech Dereverberation. *IEEE Transactions on Audio, Speech, and Language Processing*. 2013, vol. 21, iss. 9, pp. 1879–1890. ISSN 1558-7916. DOI: 10.1109/TASL.2013.2260743.
- [26] ZHENG, C., R. PENG, J. LI and X. LI. A Constrained MMSE LP Residual Estimator for Speech Dereverberation in Noisy Environments. *IEEE Signal Processing Letters*. 2014, vol. 21, iss. 12, pp. 1462–1466. ISSN 1070-9908. DOI: 10.1109/LSP.2014.2340396.
- [27] KINOSHITA, K., 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*. 2009, vol. 17, iss. 4, pp. 534–545. ISSN 1558-7916. DOI: 10.1109/TASL.2008.2009015.
- [28] WU, M. and D. WANG. A one-microphone algorithm for reverberant speech enhancement. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Hong Kong: IEEE, 2003, pp. 892–895. ISBN 0-7803-7663-3. DOI: 10.1109/ICASSP.2003.1198925.
- [29] GAROFOLO, J. *Getting started with the DARPA timit cd-rom: An acoustic phonetic continuous speech database*. Gaithersburgh: National Institute of Standards and Technology.
- [30] ALLEN, J. B. and D. A. BERKLEY. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*. 1979, vol. 65, iss. 4, pp. 943–950. ISSN 0001-4966. DOI: 10.1121/1.382599.
- [31] JUKIC, A., T. VAN WATERSCHOOT, T. GERKMANN and S. DOCLO. Speech dereverberation with convolutive transfer function approximation using MAP and variational deconvolution approaches. In: *14th International Workshop on Acoustic Signal Enhancement*. Juan-les-Pins: IEEE, 2014, pp. 50–54. ISBN 978-1-4799-6808-4. DOI: 10.1109/IWAENC.2014.6953336.

About Authors

Mohamed Anouar BEN MESSAOUD received the Ph.D. degree in Electrical Engineer-

ing from the National School of Engineers of Tunis in 2011. His research interests include topics in speech analysis and applications to engineering and computer science, particularly to pitch and multi-pitch estimation, voicing decision, speech separation, and also speech enhancement.

Aicha BOUZID received her Ph.D. in Electrical Engineering from the National School of Engineers of Tunis (ENIT) Tunisia in 2004. Currently, she is working as Professor in ENIT. Her current research interests include Speech Processing and Image Processing.