A novel iterative signal enhancement algorithm for noise reduction in speech

Simon Doclo, Ioannis Dologlou, Marc Moonen

Department of Electrical Engineering - ESAT
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94, 3001 Heverlee, Belgium
Tel. +32/16/321899, Fax +32/16/321970
{simon.doclo,ioannis.dologlou,marc.moonen}@esat.kuleuven.ac.be

Abstract

This paper presents an iterative signal enhancement algorithm for noise reduction in speech. The algorithm is based on a truncated singular value decomposition (SVD) procedure, which has already been used as a tool for signal enhancement [1][2]. Compared to the classical algorithms, the novel algorithm gives rise to comparable improvements in signal-to-noise ratio (SNR). Moreover the algorithm has an improved frequency selectivity for filtering out the noise and performs better with respect to the higher formants of the speech. It can also be extended easily to multiple channels.

1 INTRODUCTION

In many speech communication applications, like audio-conferencing and hands-free mobile telephony, the recorded and transmitted speech signals contain a considerable amount of acoustic background noise. This is mainly due to the fact that the speaker is located at a certain distance from the recording microphones. Background noise can stem from stationary noise sources, but most of the time the background noise is non-stationary and broadband, with a spectral density depending upon the environment. Background noise causes a signal degradation which can lead to total unintelligibility and which decreases the performance of speech coding and speech recognition systems.

Some approaches for noise reduction are based on the singular value decomposition (SVD) [1][2]. The idea is to consider the signal as a vector in an N-dimensional space and to separate the noisy signal into a clean signal and a noise signal, lying in orthogonal subspaces. This is done by constructing a Hankel matrix containing the given signal, reducing this matrix to a lower rank and restoring the Hankel structure (see section 2). A FIR filter interpretation of this approach has been given which provides some insight into the frequency domain properties [3]. The classical algorithm, which consists of repeating this approach a number of times, is covered in section 3.

In section 4 the iterative signal enhancement (ISE) algorithm is presented, which consists of two loops. In the inner loop an enhanced signal is computed based on the largest singular value (most energetic spectral region) and its residual signal. The outer loop consists of the repetition of the inner loop for a number of times, depending of the noise level. In section 5 some simulations and results are discussed, comparing the SNR improvement and the frequency behaviour of both algorithms.

2 TRUNCATED SVD PROCEDURE

2.1 Outline of procedure

Consider the clean speech signal $x[k]$ and the noise signal $n[k]$ (both unknown). If we assume the noise to be additive, we can write

$$y[k] = x[k] + n[k], \tag{1}$$

where $y[k]$ corresponds to the recorded noisy signal. From the vector $\mathbf{y} = [y[0], y[1], \ldots, y[N-1]]^T$ we can construct the Hankel matrix $\mathbf{Y} \in \mathbb{R}^{L \times M}$,

$$\mathbf{Y} = \begin{bmatrix} y[0] & y[1] & \ldots & y[M-1] \\ y[1] & y[2] & \ldots & y[M] \\ \vdots & \vdots & & \vdots \\ y[L-1] & y[L] & \ldots & y[N-1] \end{bmatrix}, \tag{2}$$

with $L \geq M$ and $M + L = N + 1$.

If we assume that the clean signal $x[k]$ consists of a sum of p complex exponentials, then the Hankel matrix containing the clean signal is rank-deficient and has rank $p \leq M$. This is a model that is often attributed to clean speech [4]. If $n[k]$ consists of broadband noise, the matrix $\mathbf{Y}$ will in general not be rank-deficient and will have rank M.

From the SVD of $\mathbf{Y}$ it is possible to construct a least-squares estimate of the Hankel matrix containing the clean signal. When we set the $M - p$ smallest singular values, corresponding to the noise, to zero and we only retain the p largest singular values, corresponding to the signal, we are able to construct the matrix $\mathbf{Y}_p$,

$$\mathbf{Y}_p = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^T, \tag{3}$$

which is the best rank-p approximation of the original matrix $\mathbf{Y}$.

In general, the matrix $\mathbf{Y}_p$ does not have a Hankel structure. A simple procedure for restoring the Hankel structure is to average along the anti-diagonals of the matrix

and to construct a Hankel matrix $\widehat{\mathbf{X}}$,

$$\widehat{\mathbf{X}} = \begin{bmatrix} \hat{x}[0] & \hat{x}[1] & \dots & \hat{x}[M-1] \\ \hat{x}[1] & \hat{x}[2] & \dots & \hat{x}[M] \\ \vdots & \vdots & & \vdots \\ \hat{x}[L-1] & \hat{x}[L] & \dots & \hat{x}[N-1] \end{bmatrix} \quad (4)$$

$$\hat{x}[k] = \frac{1}{\beta - \alpha + 1} \sum_{i=\alpha}^{\beta} \mathbf{Y}_p(k-i+2, i) \quad (5)$$

$$\alpha = \max(1, k-L+2), \beta = min(M, k+1) \quad (6)$$

Because of the averaging, the matrix $\widehat{\mathbf{X}}$ in general does not have rank p any more. Still, because $\widehat{\mathbf{X}}$ is closer to $\mathbf{Y}_p$ than the original matrix $\mathbf{Y}$, the signal $\hat{\mathbf{x}} = [\hat{x}[0], \hat{x}[1], \dots, \hat{x}[N-1]]^T$ will be more compatible with the p-th order model than the signal $\mathbf{y}$. It has been shown that for speech applications this simple procedure is indeed able to reduce additive noise [5].

2.2 FIR filter representation

In [3][6] a complete FIR filter representation of this algorithm in terms of the SVD of $\mathbf{Y}$ is described. The signal $\hat{x}[k]$, extracted from $\widehat{\mathbf{X}}$, essentially consists of the sum of p zero-phase filtered versions of the original signal $y[k]$. The zero-phase filters used are constructed from the right singular vectors $\mathbf{v}_i$ of the matrix $\mathbf{Y}$. The whole procedure can be considered a signal-dependent filtering operation of the signal $y[k]$ with a length-(2M-1) FIR filter. Figure 1 gives a schematic overview of this FIR filter representation. In this figure $\mathbf{J}$ is the reverse identity matrix. Multiplication with a length-N sequence $d[k]$ is necessary because of the different lengths of the anti-diagonals,

$$d[k] : \left[1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{M}, \frac{1}{M}, \dots, \frac{1}{M}, \dots, \frac{1}{3}, \frac{1}{2}, 1\right]. \quad (7)$$

Because the SVD can be considered a decomposition based on an energy criterion, the zero-phase filtered versions corresponding to the large singular values correspond to frequency components with large amplitudes. For speech this means that the zero-phase filters corresponding to the large singular values capture the formants of the speech, while the other zero-phase filtered versions mainly contain noise. Although this procedure provides some noise reduction, the obtained signal enhancement is generally not sufficient. However this procedure will be used as a tool in the algorithms described in sections 3 and 4.

3 CLASSICAL ALGORITHM

The classical algorithm for noise reduction repeats the previous procedure a number of times, using the output of one stage as the input to the following stage. This algorithm is described in table 1.

If we keep the order M and the rank p fixed, we obtain an 'enhanced' signal $\hat{x}[k]$ that is exactly represented by a p-th order model [7]. However, these iterations turn out not to be very good in terms of the resulting speech quality. If the order p is too low, the speech will sound low-pass filtered and the high frequency components of the speech will be lost. If the order p is higher, annoying 'musical tones' will be introduced. Therefore we propose an alternative procedure, referred to as the Iterative Signal Enhancement (ISE) algorithm.

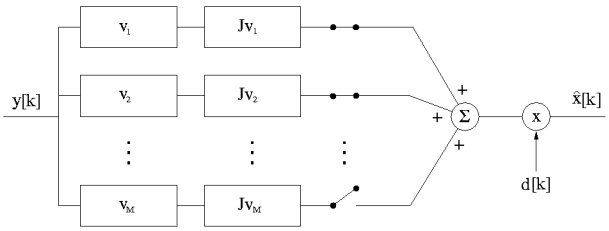

Figure 1: FIR filter representation

1.	Initialisation : $y^0[k] = y[k]$
2.	**for** $i = 0 \dots l-1$,
	• construct Hankel matrix $\mathbf{Y}^i$ from $y^i[k]$
	• compute SVD of $\mathbf{Y}^i$
	• truncate $\mathbf{Y}^i$ to rank p
	• average and extract output signal $\hat{x}^i[k]$
	• $y^{i+1}[k] = \hat{x}^i[k]$
	end
3.	Enhanced signal : $\hat{x}[k] = y^l[k]$

Table 1: Classical algorithm

4 ITERATIVE SIGNAL ENHANCEMENT ALGORITHM

4.1 Algorithm

The algorithm consists of two loops. The inner loop is an iterative procedure which computes an enhanced signal $\hat{x}[k]$ from a noisy input signal $y[k]$. Since the enhanced signal coming out of the inner loop still contains some noise, the outer loop consists of repeating the inner loop a number of times, depending on the noise level.

The iterative procedure (inner loop) proceeds as follows : from the noisy input signal $y[k]$ the rank-1 signal decomposition $s[k]$ is calculated by truncating the Hankel matrix to rank 1 and averaging along its anti-diagonals. In the frequency domain this signal decomposition will cover the most energetic spectral band of the input signal. Then the residual signal $r[k]$ is calculated by subtracting the signal decomposition $s[k]$ from the input signal $y[k]$ and the procedure is started over again using the residual signal as input signal for the next iteration. The enhanced signal $\hat{x}[k]$ is obtained by summing the signal decompositions $s[k]$ over all iterations. We stop iterating when the residual signal $r[k]$ contains 'only noise'-components.

The inner loop is represented in figure 2, keeping in mind that the superscript i represents the number of times the inner loop has been repeated. The complete ISE algorithm is described in table 2. The ISE algorithm can be viewed as a filtering operation of the signal $y[k]$ with a length-(2p(M-1)+1) FIR filter.

The advantage of the ISE algorithm is that it has a better frequency selectivity for filtering out the noise than the classical algorithm. The ISE algorithm performs systematically better with respect to the higher formants of the speech (see section 5). This algorithm can also be implemented efficiently because it requires only the computation of the largest singular value and its corresponding singular vector [8].

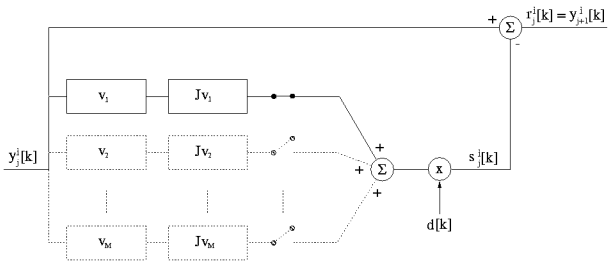

Figure 2: ISE algorithm (inner loop)

1. Initialisation : $y_0^0[k] = y[k]$
2. **for** $i = 0 \ldots l-1$, *(outer loop)*
3. - **for** $j = 0 \ldots p-1$, *(inner loop)*
   - construct Hankel matrix $\mathbf{Y}_j^i$ from $y_j^i[k]$
   - compute rank-1 decomposition $s_j^i[k]$
   - compute residual $r_j^i[k] = y_j^i[k] - s_j^i[k]$
   - $y_{j+1}^i[k] = r_j^i[k]$

   end
   - enhanced signal : $\hat{x}^i[k] = \sum_{n=0}^{p-1} s_n^i[k]$
   - $y_0^{i+1}[k] = \hat{x}^i[k]$

   end
4. Enhanced signal : $\hat{x}[k] = y_0^l[k]$

Table 2: ISE algorithm

4.2 Extension to multiple channels

The SVD-based algorithm can be extended to multiple channels, by considering block-Hankel matrices instead of Hankel matrices [9]. Consider the K-dimensional vector $\mathbf{m}[k] = [m_1[k], m_2[k], \ldots, m_K[k]]^T$, consisting of the K microphone signals at time k.

The ISE algorithm can be extended to multiple channels by replacing the Hankel matrix $\mathbf{Y}$ by the following block-Hankel matrix $\mathcal{H} \in \mathbb{R}^{L \times KM}$

$$\mathcal{H} = [\mathbf{H}_1 \mathbf{H}_2 \ldots \mathbf{H}_K] \tag{8}$$

$$\mathbf{H}_i = \begin{bmatrix} m_i[0] & m_i[1] & \ldots & m_i[M-1] \\ m_i[1] & m_i[2] & \ldots & m_i[M] \\ \vdots & \vdots & & \vdots \\ m_i[L-1] & m_i[L] & \ldots & m_i[N-1] \end{bmatrix}. \tag{9}$$

This way the correlation between the different channels is exploited, assuming that the noise is less correlated than the speech.

5 SIMULATIONS AND RESULTS

Several experiments have been carried out, where we processed a speech signal, corrupted by noise, with both algorithms. Figure 3 shows the maximum attainable SNR-improvement for the inner loop of both algorithms, when the speech signal is corrupted by white noise (SNR ranging from 0 to 30 dB). The SNR improvement is defined as

$$SNR(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left(\frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \right), \tag{10}$$

where $\mathbf{x} = [x[0], x[1], \ldots, x[N-1]]^T$ is the clean speech signal vector and $\hat{\mathbf{x}} = [\hat{x}[0], \hat{x}[1], \ldots, \hat{x}[N-1]]^T$ is the enhanced signal vector. As shown in the figure, the SNR improvements for both algorithms are comparable.

However, the frequency behaviour of the ISE algorithm is better than the classical algorithm. We have processed a speech signal (8 kHz, frames of 1000 samples), corrupted with white noise (SNR = 10 dB). Figure 4 shows the frequency spectrum of one frame of the clean and the noisy speech signal.

We have processed the noisy signal with both algorithms. Figure 5 shows the frequency spectrum of the enhanced signal for both algorithms. As can be seen, the ISE algorithm has a better frequency selectivity for filtering out the noise than the classical algorithm. The ISE algorithm performs systematically better with respect to the higher formants of the speech, which are clearly preserved.

Figure 6 shows the frequency spectrum of the residual signal (noisy signal minus enhanced signal) for both algorithms. Ideally, this spectrum should be flat, since we have added white noise to the clean speech signal. As can be seen, the spectrum obtained with the ISE algorithm resembles much more a white noise spectrum than the spectrum obtained with the classical algorithm. The residual signal for the classical algorithm mainly contains

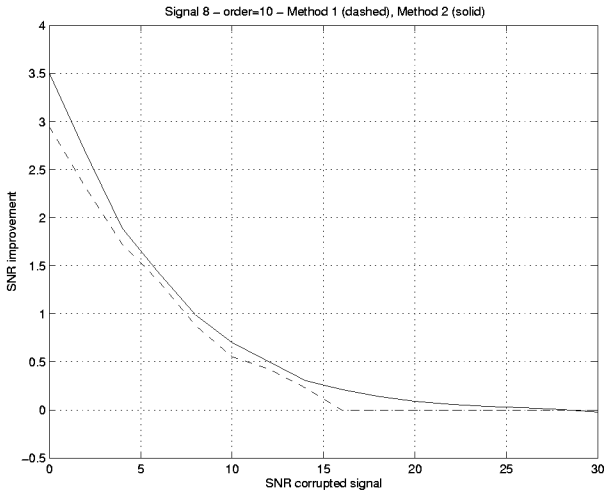

Figure 3: SNR improvement for both algorithms (*dashed*: classical algorithm, *solid*: ISE algorithm)

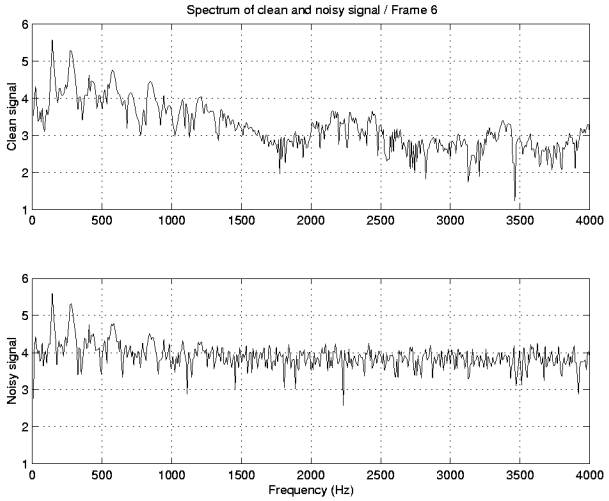

Figure 4: Frequency spectrum of the clean and the noisy speech signal (*top*: clean signal, *bottom*: noisy signal)

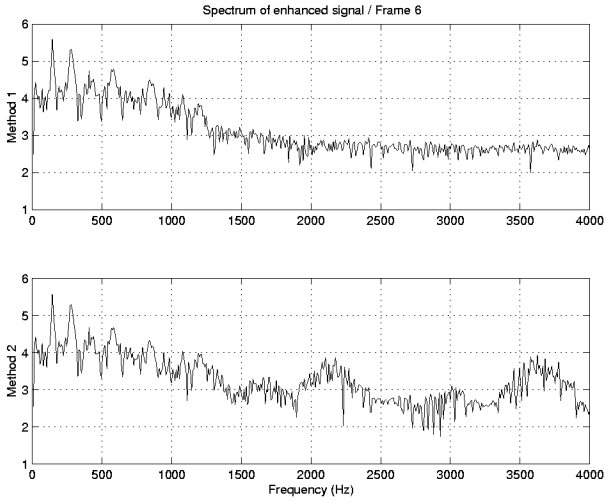

Figure 5: Frequency spectrum of the enhanced signal (*top*: classical algorithm, *bottom*: ISE algorithm)

high frequency components, so that executing the classical algorithm can be considered merely a low-pass filtering of the noisy speech signal, whereas the residual signal for the ISE algorithm also contains components in the low frequency region.

Acknowledgements

This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven. It was partially supported by the Concerted Research Action *Model-based Information Processing Systems* of the Flemish Government (GOA/MIPS/95/99/3) and the F.W.O. Research Project nr. G.0295.97, *Design and implementation of adaptive digital signal processing algorithms for broadband applications.* Simon Doclo is a Research Assistant supported by the I.W.T. Marc Moonen is a Research Associate with the F.W.O.-Vlaanderen.

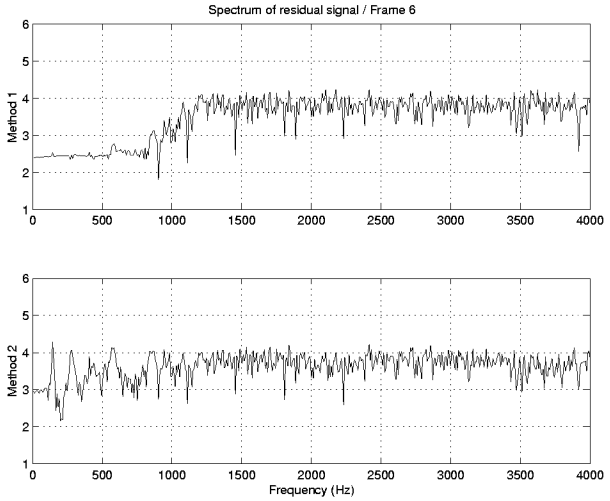

Figure 6: Frequency spectrum of the residual signal (*top*: classical algorithm, *bottom*: ISE algorithm)

References

1. S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of Broad-Band Noise in Speech by Truncated QSVD", *IEEE Trans. Speech, Audio Processing*, vol. 3, no. 6, pp. 439–448, Nov. 1995.
2. P. S. K. Hansen, *Signal Subspace Methods for Speech Enhancement*, PhD thesis, Technical University of Denmark, Lyngby, Denmark, Sept. 1997.
3. P. C. Hansen and S. H. Jensen, "FIR Filter Representations of Reduced-Rank Noise Reduction", Accepted for publication in *IEEE Trans. Signal Processing*, Sept. 1996.
4. R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation", *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, Aug. 1986.
5. M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise : A regenerative approach", *Speech Communication*, vol. 10, no. 2, pp. 45–57, Feb. 1991.
6. I. Dologlou and G. Carayannis, "Physical Representation of Signal Reconstruction from Reduced Rank Matrices", *IEEE Trans. Signal Processing*, vol. 39, no. 7, pp. 1682–1684, July 1991.
7. J. A. Cadzow, "Signal Enhancement – A Composite Property Mapping Algorithm", *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, Jan. 1988.
8. G. H. Golub and C. F. Van Loan, *Matrix Computations*, MD : John Hopkins University Press, Baltimore, 3rd edition, 1996.
9. I. Dologlou, J.-C. Pesquet, and J. Skowronski, "Projection-based rank reduction algorithms for multichannel modelling and image compression", *Signal Processing*, vol. 48, pp. 97–109, Jan. 1996.