

Dual-Microphone Speech Dereverberation in a Noisy Environment

Emanuël A. P. Habets
Dept. of Electrical Engineering
Technische Universiteit Eindhoven
Eindhoven, The Netherlands
Email: e.a.p.habets@tue.nl

Sharon Gannot
School of Engineering
Bar-Ilan University
Ramat-Gan, Israel
Email: gannot@eng.biu.ac.il

Israel Cohen
Dept. of Electrical Engineering
Technion - Israel Institute of Technology
Haifa, Israel
Email: icohen@ee.technion.ac.il

Abstract—Speech signals recorded with a distant microphone usually contain reverberation and noise, which degrade the fidelity and intelligibility of speech, and the recognition performance of automatic speech recognition systems. In [1] Habets presented a multi-microphone speech dereverberation algorithm to suppress late reverberation in a noise-free environment. In this paper we show how an estimate of the late reverberant energy can be obtained from noisy observations. A more sophisticated speech enhancement technique based on the Optimally-Modified Log Spectral Amplitude (OM-LSA) estimator is used to suppress the undesired late reverberant signal and noise. The speech presence probability used in the OM-LSA is extended to improve the decision between speech, late reverberation and noise. Experiments using simulated and real acoustic impulse responses are presented and show significant reverberation reduction with little speech distortion.

I. INTRODUCTION

In general, acoustic signals radiated within a room are linearly distorted by reflections from walls and other objects. Early room echoes mainly contribute to coloration, or spectral distortion, while late echoes, or late reverberation, contribute noise-like perceptions or tails to speech signals. These distortions degrade the fidelity and intelligibility of speech, and the recognition performance of automatic speech recognition systems. Late reverberation and spectral coloration cause users of hearing aids to complain of being unable to distinguish voices in a crowded room. We have investigated the application of signal processing techniques to improve the quality of speech distorted in an acoustic environment.

Even after three decades of continuous research, speech dereverberation remains a challenging problem. Dereverberation algorithms can be divided into two classes. The classification depends on whether the Room Impulse Responses (RIRs) need to be known or estimated beforehand. Until now blind estimation of the RIRs, in a practical scenario, remains an unsolved but challenging problem [2]. Even if the RIRs could be estimated, the inversion and tracking would be very difficult. While these techniques try to recover the anechoic speech signal we like to suppress the tail of the RIR by means of spectral enhancement.

One of the reasons that reverberation degrades speech intelligibility is the effect of overlap-masking, in which segments of an acoustic signal are affected by reverberation components of previous segments. In [1] Habets introduced a multi-microphone speech dereverberation method based on spectral subtraction to reduce this effect. The described method estimates the Power Spectrum Density (PSD) of late reverberation directly from the reverberant, but noise-free, microphone signals.

In this paper we show how an estimate of the late reverberant energy can be obtained from two noisy observations. A more sophisticated speech enhancement technique based on the Optimally-Modified Log Spectral Amplitude (OM-LSA) estimator [3] is used to

suppress undesired late reverberation and noise. The speech presence probability used in the OM-LSA is modified to improve the decision between speech, late reverberation and noise. Experiments using simulated and real acoustic impulse responses are presented and show significant reverberation reduction with little speech distortion.

The outline of this paper is as follows. In Section II, we explain the problem in more detail. Section III describes the estimation procedure of the late reverberant energy. The dual microphone speech dereverberation algorithm based on the OM-LSA estimator is presented in Section IV. A modification of the speech presence probability estimator is presented in Section V. Experimental results are presented and discussed in Section VI, and finally we discuss our conclusions in the last section.

II. PROBLEM STATEMENT

The m^{th} microphone signal is denoted by $z_m(n)$, and consists of a reverberant speech component $b_m(n)$, and a noise component $d_m(n)$. The anechoic speech signal is denoted by $s(n)$. The Room Impulse Response from the source to the m^{th} microphone is modelled as a Finite Impulse Response (FIR) of length L , and is denoted by $\mathbf{a}_m(n) = [a_{m,0}(n), \dots, a_{m,L-1}(n)]^T$. The RIR is divided into two parts such that

$$a_{m,j}(n) = \begin{cases} a_{m,j}^{\text{d}}(n) & 0 \leq j < t_r, \\ a_{m,j}^{\text{r}}(n) & t_r \leq j \leq L-1, \end{cases}$$

where j is the coefficient index, t_r is chosen such that $\mathbf{a}_m^{\text{d}}(n)$ consists of the direct path and a few early echoes, and $\mathbf{a}_m^{\text{r}}(n)$ consists of all later echoes, i.e., late reverberation. The value t_r/f_s , where f_s denotes the sample frequency, usually ranges from 40 to 80 ms. In the sequel we assume that the array is positioned such that the arrival times of the direct speech signal are aligned. The observed signals are given by

$$\begin{aligned} z_m(n) &= b_m(n) + d_m(n), \\ &= \left(\mathbf{a}_m^{\text{d}}(n)\right)^T \mathbf{s}(n) + \left(\mathbf{a}_m^{\text{r}}(n)\right)^T \mathbf{s}(n) + d_m(n), \\ &= x_m(n) + r_m(n) + d_m(n), \end{aligned}$$

where $\mathbf{s}(n) = [s(n), \dots, s(n-L+1)]^T$, $x_m(n)$ is the desired speech component, and $r_m(n)$ denotes the late reverberant component.

Using the Short-Time Fourier Transform (STFT), we have in the time-frequency domain

$$\begin{aligned} Z_m(k, l) &= B_m(k, l) + D_m(k, l), \\ &= X_m(k, l) + R_m(k, l) + D_m(k, l), \end{aligned}$$

where k represents the frequency bin index, and l the frame index.

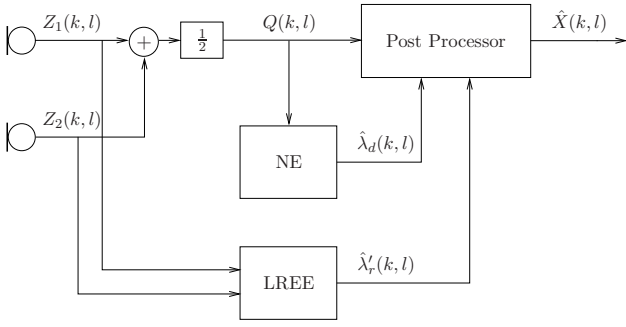


Fig. 1. Dual Microphone Speech Dereverberation System (NE: Noise Estimator, LREE: Late Reverberant Energy Estimator).

Figure 1 shows the proposed dual microphone speech dereverberation system. The time-frequency signal $Q(k, l)$ is the output of a Delay and Sum beamformer (in this case with zero delay), i.e.,

$$\begin{aligned} Q(k, l) &= \frac{1}{2} (Z_1(k, l) + Z_2(k, l)) \\ &= B(k, l) + D(k, l) \\ &= X(k, l) + R(k, l) + D(k, l). \end{aligned}$$

The Noise Estimator (NE) provides an estimate of the Power Spectral Density (PSD) of the noise in $Q(k, l)$, and is denoted by $\hat{\lambda}_d(k, l)$. We used the Improved Minima Controlled Recursive Averaging (IMCRA) approach [4] for noise estimation. The Late Reverberant Energy Estimator (LREE), see Section III, is used to obtain an estimate of the PSD of the late reverberant spectral component $R(k, l)$. It should be noted that the energy of the late reverberant spectral component $R(k, l)$ is reduced due to the Delay and Sum beamformer. The spectral speech component $\hat{X}(k, l)$ is then obtained by applying a spectral gain function $G_{\text{OM-LSA}}$, see Section IV, to each noisy spectral component, i.e.,

$$\hat{X}(k, l) = G_{\text{OM-LSA}}(k, l) Q(k, l).$$

The dereverberated speech signal $\hat{x}(n)$ can be obtained using the inverse STFT and the weighted overlap-add method.

III. LATE REVERBERANT ENERGY ESTIMATION

In this Section we explain how the late reverberant energy is estimated. There are two main issues that have to be dealt with. First, an estimate of the PSD of the reverberant signal $B_m(k, l) \forall m \in \{1, 2\}$ is needed for the estimation of the late reverberant energy (Section III-A). Second, we need to compensate for the energy contribution of the direct path, as will be explained in Section III-B.

A. Estimate Reverberant Energy

The PSD of the reverberant spectral component $B_m(k, l)$ is estimated by minimizing

$$\mathbb{E} \left\{ \left(|B_m(k, l)|^2 - |\hat{B}_m(k, l)|^2 \right)^2 \right\}$$

with $m \in \{1, 2\}$. As shown in [5] this leads to the following spectral gain function

$$G_m^{\text{SP}}(k, l) = \sqrt{\frac{\xi_m(k, l)}{1 + \xi_m(k, l)} \left(\frac{1}{\gamma_m(k, l)} + \frac{\xi_m(k, l)}{1 + \xi_m(k, l)} \right)}$$

where

$$\xi_m(k, l) = \frac{\lambda_{b_m}(k, l)}{\lambda_{d_m}(k, l)}, \text{ and } \gamma_m(k, l) = \frac{|Z_m(k, l)|^2}{\lambda_{d_m}(k, l)},$$

respectively, denote the *a priori* and *a posteriori* Signal to Noise Ratios (SNRs). The *a priori* SNRs are estimated using the Decision-Directed method proposed by Ephraim and Malah [6]. Estimates of PSD of the noise in the m^{th} microphone, i.e., $\lambda_{d_m}(k, l)$, are obtained using the IMCRA approach [4]. A noise-free estimate of the PSD of the reverberant signal is then obtained by:

$$\hat{\lambda}_{b_m}(k, l) = \left(G_m^{\text{SP}}(k, l) \right)^2 |Z_m(k, l)|^2.$$

B. Direct Path Compensation

In [1] Habets showed that, using Polack's statistical RIR model [7], the late reverberant energy can be estimated directly from the PSD of the reverberant signal using

$$\hat{\lambda}_{r_m}(k, l) = \alpha^{\frac{t_r}{R}}(k) \hat{\lambda}_{b_m} \left(k, l - \frac{t_r}{R} \right), \quad (1)$$

where $m \in \{1, 2\}$, R denotes the frame rate of the STFT, and

$$\alpha(k) = e^{-2\delta(k)\frac{R}{f_s}}.$$

The value t_r should be chosen such that $\frac{t_r}{R}$ is a positive integer value. Note that the PSD $\hat{\lambda}_{b_m}(k, l)$ in (1) was first smoothed over time using a first-order low-pass IIR filter, with filtering constant $\alpha(k)$. The exponential decay is related to the frequency dependent reverberation time $T_{60}(k)$ through

$$\delta(k) \triangleq \frac{3 \ln(10)}{T_{60}(k)}.$$

In case the spatial ergodicity requirement is fulfilled it was shown that the estimate of the late reverberant energy can be improved by spatial averaging, i.e.,

$$\hat{\lambda}_r(k, l) = \frac{1}{2} \sum_{m=1}^2 \alpha^{\frac{t_r}{R}}(k) \hat{\lambda}_{b_m} \left(k, l - \frac{t_r}{R} \right). \quad (2)$$

To incorporate the frequency dependent reverberation time we apply Polack's statistical RIR model to each sub-band. The energy envelope of the RIR in the k^{th} sub-band can be modelled as

$$\begin{aligned} h_k(z) &= \sum_{n=0}^{\infty} \alpha^n(k) z^{-n}, \\ &= \frac{1}{1 - \alpha(k)z^{-1}}. \end{aligned} \quad (3)$$

In [1] it was implicitly assumed that the energy of the direct path is small compared to the reverberant energy. However, in many practical situations the contribution of the energy related to the direct signal may cause a severe problem, since the model in (3) may not be valid. To eliminate the contribution of the energy of the direct path in $\lambda_{b_m}(k, l)$, we propose to apply the following filter to $\lambda_{b_m}(k, l)$,

$$f_{m,k}(z) = \frac{h_k(z)}{\kappa_m(k) + h_k(z)},$$

where $\kappa_m(k)$ is related to the direct and reverberant energy at the m^{th} microphone, and in the k^{th} sub-band. Using the energy envelope $h_k(z)$ we obtain

$$f_{m,k}(z) = \frac{1}{1 - \frac{\kappa_m(k)}{1 + \kappa_m(k)} \alpha(k) z^{-1}}. \quad (4)$$

Using the difference equation related to the filter in (4) we obtain an estimate of the reverberant energy with compensation of the direct path energy, i.e.,

$$\hat{\lambda}'_{b_m}(k, l) = \frac{\kappa_m(k)}{1 + \kappa_m(k)} \alpha(k) \hat{\lambda}'_{b_m}(k, l - 1) + \frac{1}{1 + \kappa_m(k)} \hat{\lambda}_{b_m}(k, l). \quad (5)$$

We now replace $\hat{\lambda}_{b_m}(k, l)$ in (2) by the PSD with compensation, i.e., $\hat{\lambda}'_{b_m}(k, l)$, to obtain the late reverberant energy $\hat{\lambda}'_r(k, l)$.

In case $\kappa_m(k) = 0$ (5) reduces to $\lambda'_{b_m}(k, l) = \lambda_{b_m}(k, l)$. The estimated late reverberant energy is then given directly by (2) as proposed in [1].

IV. DUAL-MICROPHONE DEREVERBERATION

We use a modified version of the Optimally Modified Log Spectral Amplitude estimator (OM-LSA) to obtain an estimate of the desired spectral component $X(k, l)$. The Log Spectral Amplitude (LSA) estimator proposed by Ephraim and Malah [8] minimizes

$$\mathbb{E} \left\{ \left(\log(A(k, l)) - \log(\hat{A}(k, l)) \right)^2 \right\},$$

where $A(k, l) = |X(k, l)|$ denotes the spectral speech amplitude, and $\hat{A}(k, l)$ its optimal estimator. Assuming statistical independent spectral components, the LSA estimator is defined as

$$\hat{A}(k, l) = \exp(\mathbb{E}\{\log(A(k, l)) | Q(k, l)\}).$$

The LSA gain function is given by

$$G_{\text{LSA}}(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \exp\left(\frac{1}{2} \int_{\nu(k, l)}^{\infty} \frac{e^{-t}}{t} dt\right),$$

where

$$\nu(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \gamma(k, l),$$

$$\xi(k, l) = \frac{\lambda_x(k, l)}{\lambda'_r(k, l) + \lambda_d(k, l)},$$

and

$$\gamma(k, l) = \frac{|Q(k, l)|^2}{\lambda'_r(k, l) + \lambda_d(k, l)}.$$

The OM-LSA spectral gain function, which minimizes the mean-square error of the log-spectra, is obtained as a weighted geometric mean of the hypothetical gains associated with the speech presence uncertainty [9]. Given two hypotheses, $H_0(k, l)$ and $H_1(k, l)$, which indicate, respectively, speech absence and speech presence, we have

$$H_0(k, l) : Q(k, l) = R(k, l) + D(k, l),$$

$$H_1(k, l) : Q(k, l) = X(k, l) + R(k, l) + D(k, l).$$

Based on a Gaussian statistical model, the speech presence probability is given by

$$p(k, l) = \left\{ 1 + \frac{q(k, l)}{1 - q(k, l)} (1 + \xi(k, l)) \exp(-\nu(k, l)) \right\}^{-1},$$

where $q(k, l)$ is the *a priori* signal absence probability [9]. Details w.r.t. this probability are presented in Section V.

The OM-LSA gain function is given by,

$$G_{\text{OM-LSA}}(k, l) = \{G_{H_1}(k, l)\}^{p(k, l)} \{G_{H_0}(k, l)\}^{1-p(k, l)},$$

with $G_{H_1}(k, l) = G_{\text{LSA}}(k, l)$ and $G_{H_0}(k, l) = G_{\text{min}}$. The lower-bound constraint for the gain when the signal is absent is denoted by G_{min} , and specifies the maximum amount of reduction in those frames.

In our case the lower-bound constraint does not result in the desired result since the late reverberant signal can still be audible. Our goal is to suppress the late reverberant signal down to the noise floor, given by $G_{\text{min}} D(k, l)$. We apply $G_{H_0}(k, l)$ to those time-frequency frames where the desired signal is assumed to be absent, i.e., the hypothesis $H_0(k, l)$ is assumed to be true, such that

$$\hat{X}(k, l) = G_{H_0}(k, l) (R(k, l) + D(k, l)).$$

The desired solution for $\hat{X}(k, l)$ is

$$\hat{X}(k, l) = G_{\text{min}}(k, l) D(k, l).$$

Minimizing

$$\mathbb{E} \left\{ |G_{H_0}(k, l) (R(k, l) + D(k, l)) - G_{\text{min}}(k, l) D(k, l)|^2 \right\}$$

results in,

$$G_{H_0}(k, l) = G_{\text{min}} \frac{\hat{\lambda}_d(k, l)}{\hat{\lambda}_d(k, l) + \hat{\lambda}'_r(k, l)}.$$

V. SIGNAL ABSENCE PROBABILITY

In this section we propose an efficient estimator for the *a priori* signal absence probability $q(k, l)$ which exploits spatial information. This estimator uses a soft-decision approach to compute four parameters. Three parameters, i.e., $P_{\text{local}}(k, l)$, $P_{\text{global}}(k, l)$, and $P_{\text{frame}}(l)$, are proposed by Cohen in [9], and are based on the time-frequency distribution of the estimated *a priori* SNR, $\xi(k, l)$. These parameters exploit the strong correlation of speech presence in neighbouring frequency bins of consecutive frames. We propose to use a fourth parameter to exploit spatial information. Since a strong coherency between the two microphone signals will indicate the presence of a direct signal, we propose to relate our fourth parameter to the Mean Squared Coherence (MSC) of the two microphone signals. The MSC is defined as

$$\Phi_{\text{MSC}}(k, l) \triangleq \frac{\mathcal{S}Z_{21}(k, l)}{\mathcal{S}Z_1(k, l)\mathcal{S}Z_2(k, l)}, \quad (6)$$

where $Z_{21}(k, l) = Z_2(k, l)Z_1^*(k, l)$, and the operator \mathcal{S} denotes smoothing in time, i.e.,

$$\mathcal{S}X(k, l) = \beta \mathcal{S}X(k, l - 1) + (1 - \beta) |X(k, l)|^2,$$

where β ($0 \leq \beta \leq 1$) is the smoothing parameter. The MSC is further smoothed over frequency using

$$\tilde{\Phi}_{\text{MSC}}(k, l) = \sum_{i=-w}^w b_i \Phi_{\text{MSC}}(k + i, l)$$

where b is a normalized window function ($\sum_{i=-w}^w b_i = 1$) that determines the frequency smoothing. The spatial speech presence probability $P_{\text{spatial}}(k, l)$ is related to (6) by

$$P_{\text{spatial}}(k, l) = \begin{cases} 0 & \tilde{\Phi}_{\text{MSC}}(k, l) \leq \Phi_{\text{min}}, \\ 1 & \tilde{\Phi}_{\text{MSC}}(k, l) \geq \Phi_{\text{max}}, \\ \frac{\tilde{\Phi}_{\text{MSC}}(k, l) - \Phi_{\text{min}}}{\Phi_{\text{max}} - \Phi_{\text{min}}} & \Phi_{\text{min}} \leq \tilde{\Phi}_{\text{MSC}}(k, l) \leq \Phi_{\text{max}}, \end{cases}$$

where Φ_{min} and Φ_{max} are, respectively, the minimum and maximum threshold values for $\tilde{\Phi}_{\text{MSC}}(k, l)$. The proposed *a priori* speech absence probability is given by

$$\hat{q}(k, l) = 1 - P_{\text{local}}(k, l) P_{\text{global}}(k, l) P_{\text{spatial}}(k, l) P_{\text{frame}}(l).$$

TABLE I
EXPERIMENTAL RESULTS IN TERMS OF SEGMENTAL SIGNAL TO INTERFERENCE RATIO AND BARK SPECTRAL DISTORTION.

Method	Room A @ 1 m		Room A @ 2 m		Room B @ 1 m		Room B @ 2 m	
	SegSIR	BSD	SegSIR	BSD	SegSIR	BSD	SegSIR	BSD
Unprocessed	-0.193 dB	0.314 dB	-0.198 dB	0.361 dB	2.192 dB	0.087 dB	-1.785 dB	0.265 dB
Delay & Sum Beamformer	-0.147 dB	0.268 dB	-0.141 dB	0.301 dB	2.405 dB	0.079 dB	-1.480 dB	0.242 dB
Proposed (without DPC)	-0.033 dB	0.313 dB	-0.028 dB	0.334 dB	4.292 dB	0.085 dB	0.680 dB	0.272 dB
Proposed (with DPC)	-0.053 dB	0.261 dB	-0.037 dB	0.313 dB	3.836 dB	0.078 dB	0.152 dB	0.264 dB
Parameter: $\kappa_m(k) \forall k, \forall m$	27.1		8.6		19.9		5.3	

VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we present experimental results that were obtained using synthetic and real Room Impulse Responses. A male voice of 20 seconds, sampled at 8 kHz, was used in all experiments. A moderate level of White Gaussian Noise was added to each of the microphone signals (segmental SNR \approx 20 dB). Note that too much noise will mask the late reverberation. The real RIRs were measured using a Maximum Length Sequence (MLS) technique in an office room (Room A). The (full-band) reverberation time was measured using Schroeders method, the parameter $T_{60} = 0.54$ seconds. The synthetic RIRs were generated using the image method (Room B), and the reflection coefficients were set such that the reverberation time was equal to the real acoustic room. Experiments were conducted using different distances between the source and the center of the array, ranging from 1 m to 3 m. The distance between the two microphones was set to 15 cm.

The parameters related to the OM-LSA where equal to those used in [9]. Parameters that were altered or added in Section IV and V are presented in Table II. The parameter t_r/f_s was set to 48 ms, $\kappa_m(k)$ was fixed for all k and m , and was determined experientially for each situation, its value can be found in Table I.

We used the Segmental Signal to Interference Ratio (SegSIR) and the Bark Spectral Distortion (BSD) to evaluate the proposed algorithm. As a reference for these speech quality measures we used the (properly delayed) anechoic speech signal. From the results presented in Table I we can see that the Direct Path Compensation (DPC) has a positive outcome in case the source receiver distance is (relatively) small and the energy related direct path energy is large.

In Figure 2 the spectrogram of the proposed method, using Room B @ 1 m, with and without DPC are depicted. One can clearly see that the DPC prevents over-subtraction of late reverberation, which is also indicated by the BSD measure.

In Figure 3 the microphone signal $z_1(n)$ and the output of the proposed algorithm (with DPC), using Room B @ 2 m, are depicted. Note that the noise, and smearing caused by late reverberation, are clearly reduced.

The results are available for listening on the following web page: <http://www.sps.ele.tue.nl/members/e.a.p.habets/isspit06>.

TABLE II
PARAMETERS RELATED TO THE OM-LSA IN SECTION IV AND V.

$\Phi_{\min} = 0.2$	$\beta = 0.46$	$G_{\min}^{\text{dB}} = 15$ dB
$\Phi_{\max} = 0.6$	$w = 9$	

VII. CONCLUSIONS

In this paper we have presented an algorithm for speech dereverberation in a noisy environment using two microphones. We showed

how the PSD of the late reverberant component can be estimated in a noisy environment, using little *a priori* information about the RIRs. A novel method is proposed to effectively compensate for the direct path energy. We used the OM-LSA estimator to suppress late reverberation and noise. The OM-LSA estimator is a well known speech enhancement technique that introduces considerably less musical tones compared to the spectral subtraction technique used in [1]. Additionally, we proposed two modifications for the OM-LSA, which resulted in a larger amount of interference suppression

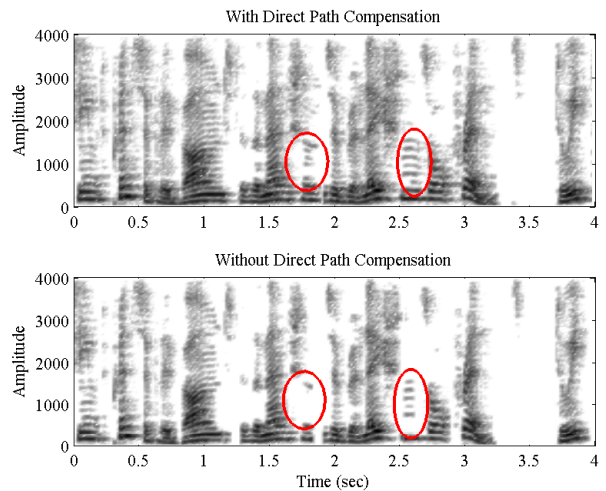


Fig. 2. Spectrogram of the proposed solution with, and without DPC, taken from experiment Room B @ 1 m.

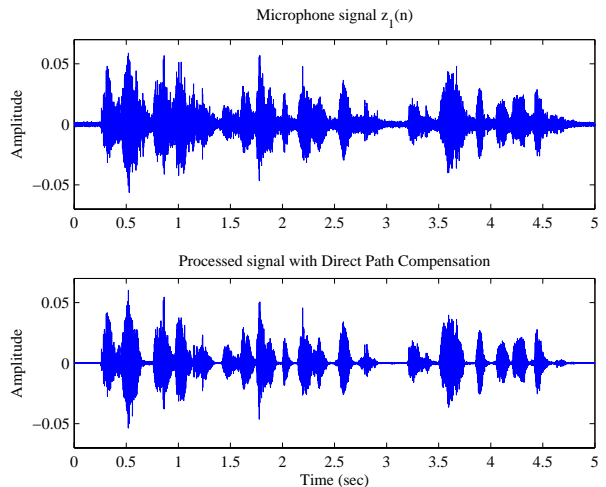


Fig. 3. Microphone signal $z_1(n)$ and the proposed algorithm with DPC, taken from experiment Room B @ 2 m.

and an improvement of the *a priori* speech absence probability.

ACKNOWLEDGMENT

This research is/was partially supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs. The authors express their gratitude to STW for funding.

REFERENCES

- [1] E. Habets, "Multi-Channel Speech Dereverberation based on a Statistical Model of Late Reverberation," in *Proc. of the 30th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005)*, Philadelphia, USA, March 2005, pp. 173–176.
- [2] Y. Huang, J. Benesty, and J. Chen, "Identification of acoustic MIMO systems: Challenges and opportunities," *Signal Processing*, no. 86, pp. 1278–1295, 2006.
- [3] I. Cohen, "Relaxed Statistical Model for Speech Enhancement and A Priori SNR Estimation," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, pp. 870–881, September 2005.
- [4] —, "Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 5, pp. 466–475, Sep 2003.
- [5] P. J. Wolfe and S. J. Godsill, "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," *EURASIP J. Appl. Signal Process., Special Issue on Digital Audio for Multimedia Communications*, vol. 2003, no. 10, pp. 1043–1051, Sep 2003.
- [6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," in *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, April 1985, pp. 443–445.
- [7] J. Polack, "La transmission de l'énergie sonore dans les salles," Thèse de Doctorat d'Etat, Université du Maine, La mans, 1988.
- [8] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," in *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, December 1984, pp. 1109–1121.
- [9] I. Cohen, "Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator," *IEEE Signal Processing Lett.*, vol. 9, no. 4, pp. 113–116, April 2002.