

# A BINAURAL SYSTEM FOR THE SUPPRESSION OF LATE REVERBERATION

*K. Lebart\*, J.M. Boucher\* and P.N. Denbigh†*

\* ENST-Br, Dept. SC., Technopôle de Brest Iroise, BP 832, 29285 Brest Cedex, France.  
e-mail: Katia.Lebart@enst-bretagne.fr / J.M.Boucher@enst-bretagne.fr

† University of Sussex, Falmer, Biomedical Engineering, BN1 9QT, Brighton, England.  
e-mail: P.N.Denbigh@sussex.ac.uk

## ABSTRACT

The propriety of spatial decorrelation of late reverberation has often been used in binaural systems of dereverberation [1] [2]. However, the use of a small array of two microphones limits the performance of such systems, especially at low frequencies. We present in this paper a new algorithm, derived from the classical method proposed by Bloom et al. [2][3]. Both methods are assessed in terms of gain in signal to noise ratio (SNR), noise reduction, distortion and improvement in performances of automatic speech recognition. The proposed method leads to better noise reduction and consistent improvement in speech recognition scores.

Keywords : Speech dereverberation, front end for hearing aids, Short Time Fourier Transform.

## 1 INTRODUCTION

Reverberation is an acoustical phenomenon of which a normal listener, in most situations, is barely aware. However, for hearing impaired listeners, it can reduce speech intelligibility (eg. [4]). Moreover, automatic speech processing algorithms are often not very robust to reverberation. A dereverberation algorithm would be of benefit as a front end to a speech recognition system or a sophisticated hearing aid. Many signal processing approaches have been proposed in the past to suppress reverberation. We concentrate here on one type of algorithm, exploiting through binaural processing the spatial decorrelation of late reverberation. At low frequencies, where an important part of the speech information is concentrated, the efficiency of such algorithms is limited by the use of closely spaced microphones. We present here a new method, based on the theoretical study of the physical problem. We will then describe the assessment protocol used, and results of the comparative study of the performances of the different methods will be outlined.

## 2 DEREVERBERATION ALGORITHM

### 2.1 Principle

The principle of the algorithm is represented in Fig. 1.

Late reverberation can be modelled as a diffuse noise, that is spatially decorrelated. In the algorithm, the left and right input signals are decomposed into a Short Time Fourier Transform (STFT) filter bank after suitable zero padding. In each sub band  $k$ , a gain is estimated from the short time spectra as the absolute value of the coherence function [2] of the inputs. Since the speech signal and the reverberation are non stationary, the gain is estimated adaptively:

$$G_k(n) = \frac{|\Phi_{xy}(n)|}{\sqrt{\Phi_{xx}(n)\Phi_{yy}(n)}} \quad (1)$$

with:

$$\Phi_{xy}(n) = E[XY^*] = \alpha\Phi_{xy}(n-1) + (1-\alpha)X(n)Y^*(n) \quad (2)$$

where  $X$  and  $Y$  are the short time Fourier transforms of the inputs at each microphone.  $G_k$  is then applied as a multiplicative modification to  $k$ th coefficient of the STFT arising from one channel.

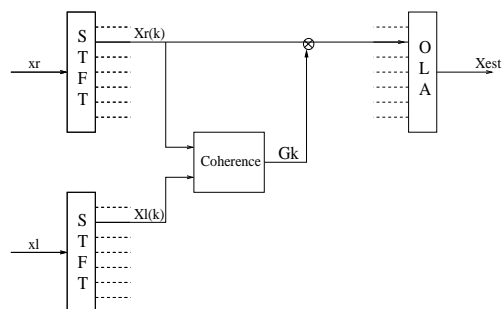


Figure 1: Dereverberation system [2]

Therefore, when the spatial coherence is low in one frequency band, reflecting a strong presence of reverberation, the signal is strongly attenuated. When it is high, the direct signal is dominant, and little or no attenuation is applied. An estimate of the dereverberated signal is then obtained by overlap-and-add reconstruction from this modified spectrum.

## 2.2 Problem Linked to the Use of Closely Spaced Microphones

Late reverberation can be modelled as a diffuse noise. For two closely spaced microphones (distance  $d$ ), the theoretical coherence function for diffuse noise and a propagative signal is given by  $\gamma(f)$  ([7]) :

$$|\gamma(f)| = \left| \frac{R(f) + \text{sinc}\left(\frac{2\pi f d}{c}\right)}{1 + R(f)} \right| \quad (3)$$

where  $R(f)$  is the propagative to diffuse energy ratio. In Fig. 2 the theoretical value of  $\gamma(f)$  is plotted for different values of the propagative to diffuse energy ratio,  $R : R = 0$  and  $R = 1$ .

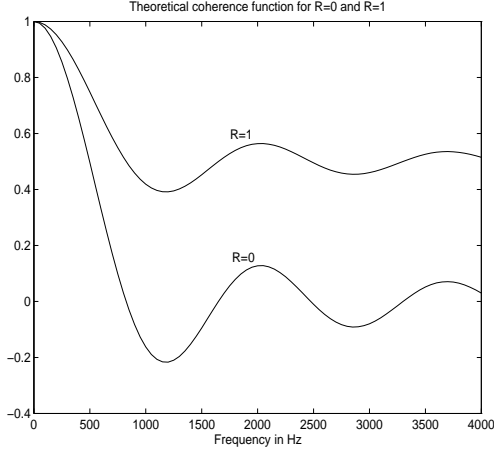


Figure 2: Theoretical coherence function  $\gamma$  for  $R = 0$  and  $R = 1$

At high frequencies the  $\text{sinc}\left(\frac{2\pi f d}{c}\right)$  term becomes negligible, and  $|\gamma(f)| \simeq R/(1 + R)$ . For low frequencies however, up to about 800Hz, the values of the coherence function remain fairly high, even if there is only reverberation. Therefore the attenuation applied to the signal will not suppress as much reverberation as expected in this frequency range, where an important part of the energy and information of a speech signal, corresponding to the first formant, is concentrated.

## 3 PROPOSED ALGORITHM

We propose the use of a new gain,  $\tilde{\gamma}(f)$  derived by transformation of the coherence function, in order to correct the behaviour of the cross-correlation gain at low frequencies (up to 800Hz) : if the measured coherence is less than its theoretical value  $\gamma_0(f)$  for reverberation only (when  $R = 0$ ), then the gain is set to 0. If it is greater, then the gain is set to an intermediary value so that the interval  $[\gamma_0(f) - 1]$  is mapped into the interval  $[0 - 1]$ . Different mapping functions have been tried, and a linear one has been selected that requires little extra computation. This is

$$\tilde{\gamma}(f) = a(f)\gamma(f) + b(f), \text{ for } f \leq 800 \text{ Hz} \quad (4)$$

with:

$$\begin{aligned} a(f) &= \frac{1}{1 - \text{sinc}\left(\frac{2\pi f d}{c}\right)} \\ b(f) &= \frac{-\text{sinc}\left(\frac{2\pi f d}{c}\right)}{1 - \text{sinc}\left(\frac{2\pi f d}{c}\right)} \end{aligned}$$

For the considered frequencies, the resulting theoretical gain is then equal to:

$$\tilde{\gamma}(f) = \frac{R(f)}{1 + R(f)} \quad (5)$$

This now provides an attenuation at all frequencies that reflects the proportion of diffuse and propagative energy in the signal.

## 4 ASSESSEMENT

### 4.1 Measurements

The reverberated speech signals used were obtained by convolution of anechoic phrases by real room impulses (RIRs), measured at two closely spaced omnidirectional microphones on a dummy head. Two different RIRs were used, of respective reverberation times 1.1 s (RIR1) and 1.7 s (RIR2). To assess the efficiency of the transformation we have used four types of objective measurement:

**Input to Output SNR gain [5] :** We used a time varying method proposed in [5]. The reverberated signal is decomposed into a sum of a direct signal  $s_{in}$  and a reverberant part  $r_{in}$ , obtained by convolving the anechoic signal with the first 5 ms of the RIR, and the RIR less its first 5 ms. While the complete reverberated signal is being processed, the time varying, signal-dependent gain is recorded. The recorded gain is then applied separately to the direct signal and reverberant part, giving respectively  $s_{out}$  and  $r_{out}$ . The SNR gain is then defined as:

$$G_{SNR} = 10 \log_{10} \left( \frac{\sum_{\text{Frame}} s_{out}^2 \sum_{\text{Frame}} r_{in}^2}{\sum_{\text{Frame}} s_{in}^2 \sum_{\text{Frame}} r_{out}^2} \right) \quad (6)$$

It is calculated either globally or in four equal sub-bands, over the periods of speech activity.

**Noise Reduction :** When no speech energy is present in a frame, the noise reduction is calculated in the same way by:

$$NR = 10 \log_{10} \left( \frac{\sum_{\text{Frame}} r_{in}^2}{\sum_{\text{Frame}} r_{out}^2} \right) \quad (7)$$

The separation between speech and silence zones has been made through manual segmentation.

**Distortion** : A cepstral distance [6] between the input and the output direct signal is used as a measure of distortion. Only the 8 first cepstral coefficients, which are linked to the first LPC coefficient, are taken into account. The distance used therefore reflects the dissimilarity in term of the formant structure of the two signals.

#### Speech Recognition Scores:

An isolated word recognition was used. The recognizer was first trained to 99% recognition on a set of 250 english anechoic words, pronounced by a male speaker. Recognition scores were then measured for the longest RIR, both on the reverberated words and the reverberated word processed respectively by the classical method and the proposed method.

## 4.2 Results

The results are displayed in Tables 1 and 2 for the two RIRs considered.

RIR1	Gain= $\gamma$	Modified Gain
<b>Noise Reduct.</b>	5dB	7.5dB
<b>SNR Gain</b>	0.6dB	0.4dB
<b>Distortion</b>	0.03	0.06

Table 1: Performances for RIR1

RIR2	Gain= $\gamma$	Modified Gain
<b>Noise Reduct.</b>	6.6dB	9.5dB
<b>SNR Gain</b>	0.8dB	0.8dB
<b>Distortion</b>	0.06	0.07

Table 2: Performances for RIR2

The proposed algorithm yields higher noise reduction than the classical one. For both RIRs, additional improvements of about 3dB are obtained. In periods of silence, only the reverberated energy reaches the two microphones. Because of the diffuse nature of reverberation, the spatial coherence function has high values at low frequencies, and therefore little attenuation is applied. Stronger attenuation is provided by the modified gain.

It was judged by informal listening that the residual reverberation presented a high pitched spectral coloration. This can be explained by the fact that the modified gain is applied only to low frequencies, whereas the unmodified coherence function is used as a multiplicative gain at high frequencies. In cases where there is only reverberation, the modified gain yields strong attenuation. However, even if the theoretical value of the coherence function at high frequencies is low, the estimated gain lies around 0.2. This is due to the bias and

variance of estimation, whose estimates by simulation are displayed on Fig 3. For low values of the coherence function (top figure), the bias can be as high as 0.2.

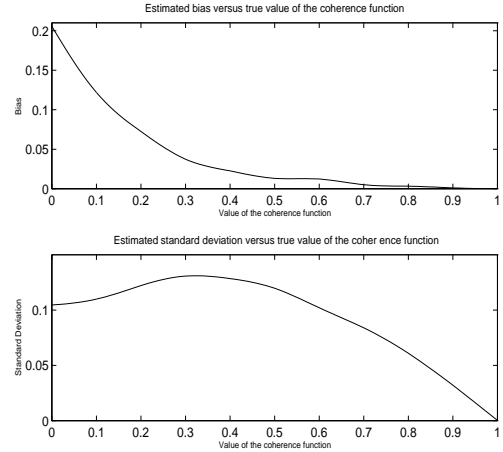


Figure 3: Bias and variance of estimation, plotted versus the true value of the coherence function

Therefore, the remaining energy after processing lies mainly above 800Hz, hence the coloration of the residual reverberation.

The modified gain yields slightly higher signal distortion, due to the unwanted attenuation of some low-energy speech at low frequencies. Over speech periods, little or no improvement in SNR gain is obtained. However, significant improvements were obtained on speech recognition scores : they are displayed, together with the 90% confidence interval, in Table 3 below.

Method	Recognition score
<b>Reverberated</b>	36.5% [31.6 - 41.7]%
<b>Gain=<math>\gamma</math></b>	37.3% [32.2 - 42.2]%
<b>Modified Gain</b>	42% [36.9 - 47.1]%

Table 3: Speech recognition scores using RIR2

## 4.3 Discussion

The method proposed relies heavily on the validity of the physical model of diffuse noise. Fig. 4 plots the measured coherence function estimated for a white noise process convolved by the late part of the second set of two room impulse responses used (RIR2), together with a dotted line showing the theoretical sinc function. It can be seen that in this case, there is good agreement between the theoretical coherence function and the model. However, this is no longer true for shorter RIRs, for which the method would be less efficient.

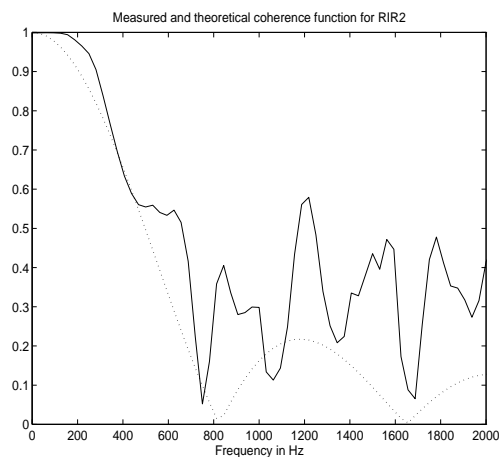


Figure 4: Theoretical and estimated coherence function for RIR2

## 5 CONCLUSION

Based on a theoretical study of the physics of the problem, we have proposed a new method derived from a classical binaural algorithm of late dereverberation. The performances of both the algorithms are assessed in terms of four different criteria. The new algorithm yields similar speech distortion but reduces noise and gives a significant improvement in automatic speech recognition performance at little computational cost.

## References

- [1] J.B. Allen, D.A. Berkley, J. Blauert, *Multi-microphone Signal-Processing Technique to Remove Room Reverberation from Speech Signals*, J. Acoust. Soc. Am., Vol. 62, no 4, pp. 912-915, Oct. 1977.
- [2] J.P. Bloom et al., *Evaluation of Two Input Speech Dereverberation Techniques*, ICASSP 1982, Paris, pp. 164-167.
- [3] J.P. Bloom, *Evaluation of a Dereverberation Process by Normal and Impaired Listeners*, ICASSP 1980, pp. 500-503.
- [4] *Acoustical Factors Affecting Hearing Aid Performance*, Ed. G.A. Studebaker, I. Hochberg, 2nd. Ed., 1993.
- [5] C. Marro, *Traitements de Déréverbération et de Débruitage pour le Signal de Parole dans des Contextes de Communication Interactive*, Thèse de l'Université de Rennes I, Déc. 1996.
- [6] R.F.Kubichek, *Standards and Technology Issues in Objective Voice Quality Assessment*, Digital Signal Processing 1, pp. 38-44, 1991.

- [7] A.G. Piersol, *Use of Coherence and Phase Data Between Two Receivers in Evaluation of Noise Environments* Journal of Sound and Vibration, 1978, Vol. 56 (2), pp 215-228.