AR SOURCE MODELING BASED ON SPATIOTEMPORALLY DIVERSE MULTICHANNEL OUTPUTS AND APPLICATION TO MULTIMICROPHONE DEREVERBERATION

Mahdi Triki, Dirk T.M. Slock[*]

Eurecom Institute
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Email: {triki,slock}@eurecom.fr

ABSTRACT

In this paper, we consider the blind multichannel dereverberation problem for a single source. The multichannel reverberation impulse response is assumed to be stationary enough to allow estimation of the correlations it induces from the received signals. It is well-known that a single-input multi-output (SIMO) filter can be equalized blindly by applying multichannel linear prediction (LP) to its output when the input is white. When the input is colored, the multichannel linear prediction will both equalize the reverberation filter and whiten the source. We exploit the channel's spatiotemporal diversity to estimate the source correlation structure, which can hence be used to determine a source whitening filter. Multichannel linear prediction is then applied to the sensor signals filtered by the source whitening filter, to obtain pure source dereverberation. A key parameter in this dereverberation scheme is the order of the source whitening filter. It determines the tradeoff between the modeling error (limited source whitening) and the estimation error (due to the blind estimation of the source correlations). In this paper we propose, using a statistical room reverberation model, a design to optimize the whitening order (function of the room characteristics, and the number of sub-channels).

1. INTRODUCTION

The quality of speech captured in real-world environments is invariably degraded by acoustic interference. This interference can be broadly classified into two distinct categories: additive and convolutive. The convolutive interference (commonly referred to as reverberation) is due to sound wave reflections from surrounding walls and objects. It leads to a modification of the speech signal characteristics. Therefore, it constitutes a major problem in speech recognition, speaker verification, and general auditive comfort in "hands-free" telephony applications. Blind dereverberation is the process of removing the effect of reverberation from an observed reverberant signal.

A simple multi-microphone speech dereverberation system is the delay-and-sum beamfomer [1, 2]. The dereverberation is performed by a simple averaging over the sensor outputs, delayed to focus in the direction of the desired speaker. Note that beamforming exploits only a partial spatial information (relative delays), and ignores the input signal characteristics.

Another way to address the problem is to consider the whole Acoustic Impulse Response (AIR). Matched Filter (MF) is proposed to equalize the room response [3]. In such a way, one increases the dereverberation SNR (compared to the Delay-and-Sum beamformer). However, MF equalization introduces a large equalization delay (of about the AIR length), and produces a pre-echo that is annoying in several applications (speech recognition...). On the other hand, SIMO channel can be perfectly equalized using multiple FIR filters (transverse filters) [4]. Let us consider a clean speech signal, $s(n)$, produced in a reverberant room. The reverberant speech signal observed on M distinct microphones can be written as:

$$\mathbf{y}(n) = H(q)s(n) \tag{1}$$

where $\mathbf{y}(n) = [y_1(n) \cdots y_M(n)]^T$ is the reverberant speech signal, $H(q) = [H_1(q) \cdots H_M(q)]^T = \sum_{i=0}^{L_h-1} \mathbf{h_i} q^{-i}$ is the SIMO channel transfer function, L_h is the channel length, and q^{-1} is the one sample time delay operator. According to the Bézout's identity, if the channels $H_1(q) \cdots H_M(q)$ does not have common zeros, then $\exists \mathbf{F}(q) = [F_1(q) \cdots F_M(q)]$ such that:

$$F(q)H(q) \;=\; \sum_{m=1}^{M} F_m(q)H_m(q) \;=\; 1 \tag{2}$$

If $H(q)$ is known (or can be estimated), the coefficients of $F_m(q)$ can be computed by the well-known rules of matrix algebra. The AIR blind estimation should face the channel/speech identifiability problem. In fact, for any scalar filter $\alpha(q)$, $(H(q)/\alpha(q),\ \alpha(q)s(k))$ is also an acceptable solution of (1). To deal with identification ambiguities, one can exploit a prior information on source spectrum. Indeed, if the source is white, the channel can be equalized using multichannels linear prediction [5]. For colored input, we take advantage of the spatial diversity to estimate the source color; and we propose a tree-stage speech dereverberation procedure exploiting spatial, temporal, and spectral diversities [6, 7]:

- First, the colored non-stationary speech signal is transformed into an iid-like signal. Exploiting spatial and temporal diversities, we estimate the AutoRegressive(AR) whitening filter based on received correlations (averaged over the subchannels).
- Then, a blind channel predictor is computed based on pre-processed reverberant speech.
- Finally, speech signal dereverberation is performed using a zero-forcing equalizer based on the predictor computed in the previous step.

A key parameter in our dereverberation scheme is the order of the whitening filter. In fact, if the correlation matrix of the pre-processed

[*]Eurecom Institute's research is partially supported by its industrial members: BMW, Bouygues Télécom, Systems, France Télécom , Hitachi Europe, SFR, Sharp, STMicroelectronics, Swisscom, Thales. The research reported herein was also partially supported by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space.

speech signal $\tilde{s}(k) = a(q)s(k)$ is spherical and if we take a long enough multichannel LP filter $\left(L_A \geq \frac{L_h - 1}{M-1}\right)$, Delay&Predict equalizes perfectly the channel. To investigate the choice of this parameter, we plot the curves of the output Direct to Reverberant Ratio (DRR) (function of the whitening filter order) using 2, 4, and 8 microphone array setups. The whitening order is plotted in logarithmic scale $20\log_{10}(l)$.

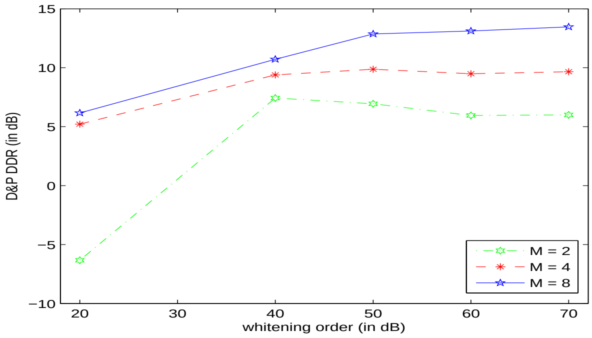

Fig. 1. The output DRR (function of the whitening filter order), using 2, 4, and 8 microphones

If we have only 2-microphones, the two-channel filter cannot be assumed to be all-pass (spatial diversity is not enough). Then, by increasing the order of the whitening filter ($l > 100$) we are capturing details belonging either to the clean speech, and/or the channel. The whitening in the first stage will also remove some channel correlation before the multichannel equalization. The fact that affects the overall dereverberation accuracy. However, for the 8 microphone array setup, the all-pass multi-channel assumption is better matched. Then by increasing the whitening LP order, we remove essentially more source correlation. And the whiteness assumption of $\tilde{s}$ is better fitted. Remark that this problem is quite different from the classic AR order selection problem, where the estimation of the source correlations is troubled by the finite number of the available observations [13]. In our problem, we assume having enough observations to have an accurate estimation of the received signals colors. The disturbance is due to the blind estimation of the source correlations: the channels are not flat and the number of microphones is not infinite. The whitening order should optimise the tradeoff between the modeling error (limited source whitening) and the estimation error (due to the blind estimation of the source correlations). In this paper we propose, using a statistical room reverberation model, a design to optimize the whitening order (function of the room characteristics, and the number of subchannels).

Notations: Upper- and lower-case boldface letters denote matrices and vectors, respectively; whereas upper-case normal letters denote spectral qualities. $E\{.\}$ and $\langle . \rangle$ represents respectively the statistical expectation and the spatial averaging. $(.)^H$ denotes the complex conjugate (Hermitian) transpose operator.

2. STATISTICAL ROOM REVERBERATION MODEL

In this section, we introduce the room reverberation model, which is built on some well-known results from statistical room acoustics.
Studying an empty rectangular room, the room impulse response $h(t)$ can be computed by solving a wave equation. At higher frequency, the complexity (in terms of the number of modes) of the deterministic wave equation modeling increases to a point where exact analysis is no longer feasible. To model $h(t)$, we will apply the theory of random (or diffuse) sound fields[9]. The crucial assumption of statistical room acoustics is that the distribution of amplitudes and phases of individual plane waves is so close to random that the sound field is uniformly distributed throughout the room volume. This theory closely describes the room acoustic behavior if the following conditions are met[8]:

A1) The dimension of the room are large relative to the wavelength of the source signal $s(t)$. For the frequencies of interest in speech processing, this condition is easily satisfied in almost all rooms.

A2) The average spacing of the resonance frequencies of the room must be smaller than on third of their bandwidth. In a room with volume V (in m^3), and reverberation time T_{60} (in seconds), this condition is fulfilled for frequencies that exceed the "Schroeder large room frequency":

$$f_s = 2000\sqrt{T_{60}/V} \tag{3}$$

A3) The source and the microphones are located in the interior of the room, at least a half-wavelength away from the walls.
Under the above conditions, the frequency response $H(f)$ (the Fourier transform of $h(t)$) can be treated as a random function of the source and microphone positions. These statistical properties are independent of the time-instant of the observation. They are determined by the room characteristics (volume, reverberation time, average wall absorption coefficient...).
We write the transfer function $H(f)$ as

$$H(f) = H^r(f) + jH^i(f) \tag{4}$$

where $H^r(f)$ and $H^i(f)$ are real and imaginary parts of $H(f)$, respectively; and $j = \sqrt{(-1)}$.
We next cite a couple of useful results derived using the Statistical Room Acoustics (SRA) theory[8, 9, 10, 11]. Assuming the assumption (A1-A3) to be fulfilled:
- $H^r(f)$, and $H^i(f)$ are independent, zero-mean, Gaussian process.
- $\langle |H(f)|^2 \rangle = \left\langle |H^r(f)|^2 + |H^i(f)|^2 \right\rangle = \dfrac{1-\beta}{\pi A \beta}$.
where $\langle . \rangle$ is the spatial expectation (estimated by averaging over all possible source and microphone positions), β is the average wall absorption coefficients, and A is the total wall surface area.
By denoting $r_h(t) = \sum_\tau h(\tau)h(\tau - t)d\tau = \text{IFFT}\left(|H(f)|^2\right)$ the autocorrelation of the room impulse response, one can show that:

$$\begin{cases} \langle r_h(t) \rangle = \dfrac{1-\beta}{\pi A \beta}\,\delta(t) \\ \langle r_h^2(t) \rangle = \text{cst}\ \exp\left(-|t|/\tau_0\right) \\ \langle r_h(t_1) r_h(t_2) \rangle = 0 \quad \forall t_1 \neq t_2 \end{cases} \tag{5}$$

where τ_0 is the time for which the sound energy in the room decays to $1/e$ of its initial value after impulsive excitation ($\tau_0 = T_{60}/13.8$).

We also assume that the room impulse response between a source and M microphones (and the corresponding autocorrelations) are i.i.d. Thus, by averaging the correlation of the different sub-channel ($r_M(t) = \frac{1}{M}\sum_{m=1}^{M} r_{h_m}(t)$), we have

$$\begin{cases} \langle r_M(t) \rangle = \langle r_{h_1}(t) \rangle = \dfrac{1-\beta}{\pi A \beta}\,\delta(t) \\ \text{var}\left(r_M(t)\right) = \dfrac{1}{M}\text{var}\left(r_{h_1}(t)\right) \end{cases} \tag{6}$$

where $\text{var}(x) = \langle x^2 \rangle - \langle x \rangle^2$ denotes the spatial variance. Note that the previous results can be also derived by modeling the room impulse response as one realization of a non-stationary stochastic process[12]:

$$h(t) = b(t)e^{-t/\delta} \tag{7}$$

where $b(t)$ is a centered stationary Gaussian noise.

3. SPEECH SOURCE WHITENING

As it is reported in the previous section, using the SRA theory one can show that for frequencies $f > f_{sch} = 2000\sqrt{T_{60}/V}$, the average reverberation spectrum is flat , i.e.,

$$\left\langle \left| H\left(\exp^{2j\pi f}\right)\right|^2 \right\rangle = \frac{1-\beta}{\pi A \beta} \tag{8}$$

Simulations shows that the superposition of the SIMO sub-channels spectrums tends to be flat as the number of microphones increases [6]. Then, the superposition of the spectra of the received signals can estimate (up to a multiplicative factor) the source spectrum. As this common part is due to the anechoic speech signal, it can be modeled as an AR process, i.e.,

$$s(n) = \frac{1}{A_s(q)} u_s(n) \tag{9}$$

where $u_s(n)$ is a zero-mean white process. The common AR coefficients can be estimated as those that minimize the sum of the squares of the prediction errors, averaged over the microphones:

$$e = \sum_{k=1}^{M}\sum_{n=0}^{\infty} e_k^2(n) = \sum_{k=1}^{M}\sum_{n=0}^{\infty}\left[y_k(n) - \sum_{j=1}^{l} a_j y_k(n-j)\right]^2 \tag{10}$$

The previous optimization problem leads to the normal equations:

$$\underbrace{\begin{bmatrix} r_{y,M}(0) & r_{y,M}(1) & \cdots & r_{y,M}(l-1) \\ r_{y,M}(1) & r_{y,M}(0) & \cdots & r_{y,M}(l-2) \\ \vdots & & \ddots & \vdots \\ r_{y,M}(l-1) & \cdots & r_{y,M}(1) & r_{y,M}(0) \end{bmatrix}}_{\mathbf{R}_{y,M}} \underbrace{\begin{bmatrix} a_M(1) \\ a_M(2) \\ \vdots \\ a_M(l) \end{bmatrix}}_{\mathbf{a}_M} = - \underbrace{\begin{bmatrix} r_{y,M}(1) \\ r_{y,M}(2) \\ \vdots \\ r_{y,M}(l) \end{bmatrix}}_{\mathbf{p}_{y,M}}$$

where - $r_{y,M}(t) = \frac{1}{M}\sum_{k=1}^{M} r_{y_k y_k}(t)$ is the correlation of the received signal at time-lag t (averaged over the M microphones).

- $r_{y_k y_k}(t)$ represents the correlation at the time-lag j of the received signal at the k^{th} microphone.

- $\{\widehat{a}_M(t)\}$ are the common AR parameters.

If the whitening filter is estimated using the source correlations, $\widetilde{s}(n) = \widehat{A}_M(q)s(n) = \frac{\widehat{A}_M(q)}{A_s(q)}u_s(n)$ will be perfectly white if the AR order goes to infinity. However, as source correlations $r_s(t)$ are unknown and only $r_{y,M}(t)$ are available to estimate the source color, infinite order is no-longer optimal. The optimal whitening order should be choosing as that minimizing the mean of the prediction error variance $\sigma_{\widetilde{s}}^2 = E\left\{\widetilde{s}(n)^2\right\}$, i.e.,

$$\hat{l} \quad = \quad \arg\min_{l} \sigma_{\widetilde{s}}^2 \tag{11}$$

On the other hand, the averaged received correlations $r_{y,M}(t)$ can be written as a function of the source correlations $r_s(t)$ and the averaged channel correlations, i.e.,

$$r_{y,M}(t) = r_s(t) * r_{h,M}(t) \tag{12}$$

By decomposing the averaged channel correlation into a deterministic and a zero-mean random processes, we have:

$$r_{y,M}(t) = c_0\left(r_s(t) + \frac{c_1}{\sqrt{M}}\underbrace{r_s * r_{\tilde{h},M}(t)}_{r_{e,M}(t)}\right) \tag{13}$$

where $c_0 = \frac{1-\beta}{\pi A \beta}$, and $r_{\tilde{h},M}(t)$ is a zero-mean random process.
If we assume that $\frac{c_1}{\sqrt{(M)}} \ll 1$, using second-order approximation one can show that

$$\begin{aligned} \widehat{\mathbf{a}}_M \quad &= \quad \mathbf{R}_s^{-1}\mathbf{p}_s + \frac{c_1}{\sqrt{M}}\left(\mathbf{R}_s^{-1}\mathbf{p}_{e,M} - \mathbf{R}_s^{-1}\mathbf{R}_{e,M}\mathbf{R}_s^{-1}\mathbf{p}_s\right) \\ &\quad + \frac{c_1^2}{M}\left(\mathbf{R}_s^{-1}\mathbf{R}_{e,M}\mathbf{R}_s^{-1}\mathbf{R}_{e,M}\mathbf{R}_s^{-1}\mathbf{p}_s - \mathbf{R}_s^{-1}\mathbf{R}_{e,M}\mathbf{R}_s^{-1}\mathbf{p}_{e,M}\right) \end{aligned} \tag{14}$$

where $\mathbf{R}_s$, and $\mathbf{R}_{e,M}$ (resp. $\mathbf{p}_s$, $\mathbf{p}_{e,M}$) have the same structure as $\mathbf{R}_{y,M}$ (resp. $\mathbf{p}_{y,M}$), where $r_{y,M}(t)$ is replaced by $r_s(t)$ and $r_{e,M}(t)$. Then, we use the predictor $\widehat{\mathbf{a}}_M$ (performed using the noisy source correlation $r_{y,M}(t)$) to whiten the speech source. The prediction error variance is given by:

$$\begin{aligned} \sigma_{\widetilde{s}}^2 \quad &= \quad \sigma_s^2 - \widehat{\mathbf{a}}_M^H\mathbf{p}_s - \mathbf{p}_s^H\widehat{\mathbf{a}}_M + \widehat{\mathbf{a}}_M^H\mathbf{R}_s\widehat{\mathbf{a}}_M \\ &= \quad \sigma_s^2 - \mathbf{p}_s^H\mathbf{R}_s^{-1}\mathbf{p}_s \\ &\quad + \frac{c_1^2}{M}\left(\mathbf{p}_{e,M} - \mathbf{R}_{e,M}\mathbf{R}_s^{-1}\mathbf{p}_s\right)^H\mathbf{R}_s^{-1}\left(\mathbf{p}_{e,M} - \mathbf{R}_{e,M}\mathbf{R}_s^{-1}\mathbf{p}_s\right) \end{aligned} \tag{15}$$

We observe that the prediction error variance can be decomposed into two terms:

- A deterministic term $\sigma_s^2 - \mathbf{p}_s^H\mathbf{R}_s^{-1}\mathbf{p}_s$ representing the error due to the use of finite order filter predictor.
- A stochastic term

$$\frac{c_1^2}{M}\left(\mathbf{R}_{e,M}^{-1}\mathbf{p}_{e,M} - \mathbf{R}_s^{-1}\mathbf{p}_s\right)^H\mathbf{R}_{e,M}\mathbf{R}_s^{-1}\mathbf{R}_{e,M}\left(\mathbf{R}_{e,M}^{-1}\mathbf{p}_{e,M} - \mathbf{R}_s^{-1}\mathbf{p}_s\right)$$

  representing the error due to the use of noisy correlations $r_{y,M}(t)$ to estimate to source color. Note that this term increases with the AR order, and is inversely proportional to the number of microphones.

The whitening order should be optimized to give the best tradeoff between these two terms.

4. WHITENING ORDER SELECTION

4.1. Stochastic Whitening Order Selection

As we can see from (15), $\sigma_{\widetilde{s}}^2$ depends on the channel realization (via $\mathbf{p}_{e,M}$ and $\mathbf{R}_{e,M}$). These information are not available (our goal is to perform blind equalization). Thus, we propose relaxing the cost function in (11), and computing the prediction order that minimize the spatially averaged prediction error variance, i.e.,

$$\hat{l} \quad = \quad \arg\min_{l}\left\langle \sigma_{\widetilde{s}}^2 \right\rangle \tag{16}$$

In such a way, we select a whitening order optimal in the average (over source and microphones positions), but not necessarily for the given channel realization. Note also that the (16) depends on the room statistics (function of the reverberation time, room volume...), but no-longer on the channel realization. Knowing the source correlations and the statistics of the room impulse response, one can have an analytical expression of $\left\langle \sigma_{\widetilde{s}}^2 \right\rangle$. However, this analytical expression is very complex to derive and to implement (even using second order approximations). So that, we propose computing the expectation using Monte-Carlo simulations:

1. We generate random channels $h(t)$ using (7) (having the same second order statistics as the room impulse responses)
2. We compute $\sigma_{\widetilde{s}}^2$ using the random channels $r_h(t)$.
3. We average over the random channels realizations.

Remark that $\sigma_{\tilde{s}}^2$ depends on the unknown source correlations (averaged over a given frame). However, the correlation details are not relevant, only the shape of the speech correlations is important. So that, we propose compute (16) using a priori speech correlation estimate $\overline{r}_s(t)$ (averaged over a long period of time, speakers ...). Fig. 2 subplots the curves of the averaged prediction error $\langle\sigma_{\tilde{s}}^2\rangle$ function of the whitening order for 2 and 4 microphones. As it was expected from (15), the optimal whitening order for 4 microphones is higher than the one for 2 microphones. We also remark that the optimization results are coherent with the dereverberation results (Fig. 1).

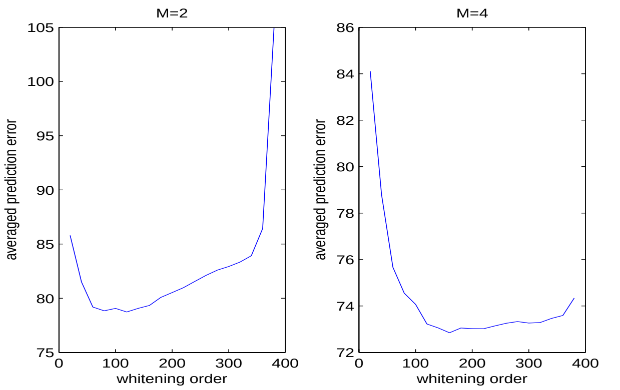

Fig. 2. The averaged prediction error $\langle\sigma_{\tilde{s}}^2\rangle$ function of the whitening order for 2 and 4 microphones.

4.2. Deterministic Whitening Order Selection

The order selected in the previous section is optimal in the average (over all possible channel realizations), but not necessarily for the given source and microphones position. In this section, we reconsider the it blind AR order selection for a given channel realization (solving (11)). To solve this problem, we propose looking to the AR modeling problem from a different point of view: we consider source correlations as the received-signal correlations corrupted by the inverse of the channel correlations. On the other hand, for large enough whitening order (such that the covariance matrices $\mathbf{R}_s$ and $\mathbf{R}_{\tilde{h},M}$ are almost band), we have:

$$\begin{aligned} \mathbf{R}_{y,M} &\approx \mathbf{R}_{\tilde{h},M}\mathbf{R}_s, \\ \mathbf{p}_{y,M} &\approx \mathbf{R}_{\tilde{h},M}\mathbf{p}_s \end{aligned} \qquad (17)$$

Using these approximations,the prediction error variance becomes

$$\sigma_{\tilde{s}}^2 = \left[\ 1\ \hat{\mathbf{a}}_M^H\ \right]\mathbf{R}_s\left[\begin{array}{c} 1 \\ \hat{\mathbf{a}}_M \end{array}\right] \approx \left[\ 1\ \hat{\mathbf{a}}_M^H\ \right]\mathbf{R}_{h,M}^{-1}\mathbf{R}_{y,M}\left[\begin{array}{c} 1 \\ \hat{\mathbf{a}}_M \end{array}\right]$$

where $\mathbf{R}_s$, $\mathbf{R}_{y,M}$, and $\mathbf{R}_{h,M}$ are $(l+1)\times(l+1)$ matrices defined as in previous. Finally, unknown matrix $\mathbf{R}_{h,M}^{-1}$ is replaced by its spatial average $\left\langle R_{h,M}^{-1}\right\rangle$:

$$\sigma_{\tilde{s}}^2 = \left[\ 1\ \hat{\mathbf{a}}_M^H\ \right]\langle\mathbf{R}_{h,M}^{-1}\rangle\,\mathbf{R}_{y,M}\left[\begin{array}{c} 1 \\ \hat{\mathbf{a}}_M \end{array}\right] \qquad (18)$$

The expectation is computed using Monte-Carlo simulations (as in the previous section). Fig. 3 subplots the curves of the prediction error $\sigma_{\tilde{s}}^2$ computed by (18) (using the source covariance matrix) or "blindly" by (18). We remark that the minima in the two curves match well; and that (18) can be used to select the whitening order. However, the approximation in (17) is valid only for large order $(l \geq 100)$. Thus, it can happen that we see some minima for $l < 100$. Those minima should be ignored. Another drawback of this approach is due to local minima. To alleviate this problem, we can use stochastic whitening order selection to situate approximately the optimal order. Then deterministic whitening order selection is derived to optimize the AR order for the given channel realization.

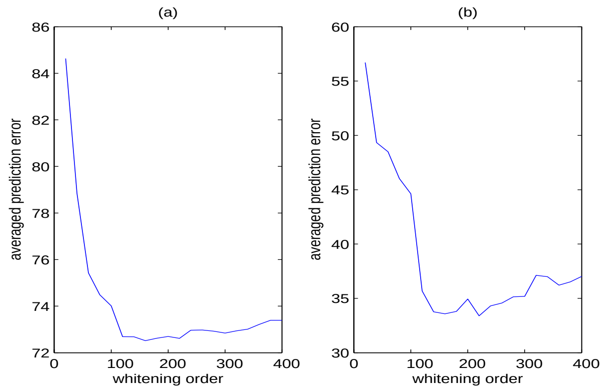

Fig. 3. The prediction error $\sigma_{\tilde{s}}^2$ computed using the source covariance matrix (a), or using (18) (b)

5. REFERENCES

[1] J. Flanagan, J. Johnston, R. Zahn, and G. Elko."Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Amer.*, Pages: 1508-1518, Nov. 1985.

[2] B.W. Gillespie, L.E. Atlas. "Acoustic Diversity for Improved Speech Recognition in Reverberant Environments," *In Proc. of the IEEE ICASSP*,Vol. 1, pp. 557-560, May 2002.

[3] J.L. Flanagan, A.C. Surendran, and E.E. Jan, "Spatially selective sound capture for speech and audio processing," *on Speech Communication*, Vol.13, pp.207-222, Oct. 1993.

[4] M. Miyoshi, and Y. Kaneda. "Inverse Filtering of Room Acoustics," *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol.36, Feb. 1988.

[5] C.B. Papadias, and D.T.M. Slock. "Fractionally Spaced Equalization of Linear Polyphase Channels and Related Blind Techniques Based on Multichannel Linear Prediction," *IEEE Trans. on Signal Processing*, Vol. 47, pp.641-654, Mar. 1999.

[6] M. Triki and D.T.M. Slock. "Blind Dereverberation of Quasi-periodic Sources Based on Multichannel Linear Prediction," *In Proc. of the IWAENC*, Sept. 2005.

[7] M. Triki and D.T.M. Slock. "Delay and Predict Equalization For Blind Speech Dereverberation," *In Proc. of the IEEE ICASSP*,Vol.5, pp.97-100, May 2006.

[8] N. D. Gaubitch and P. A. Naylor. "Analysis of the Dereverberation Performance of Microphone Arrays." *In Proc. of the IWAENC*, Sept. 2005.

[9] M.R. Schroeder. "Statistical Parameters of the Frequency Response Curves of Large Rooms," *J. of Audio Eng. Soc.*, Vol.35, pp.299-305, 1987.

[10] M.R. Schroeder, and H. Kuttruff. "On Frequency Response Curves in Rooms. comparison of experimental, theoretical and monte-carlo results for the average frequency spacing between maxima," *J. Acoust. Soc. Amer.*, Vol.34, pp.34-76, 1962.

[11] M.R. Schroeder. "Frequency Correlation Functions of Frequency Responses in Rooms," *J. Acoust. Soc. Amer.*, Vol.34, pp.1819-1823, 1963.

[12] J.D. Polack. "La transmission de l'ènergie sonore dans les salles," *Ph.D. Theses*, Maine University, Le Mans, 1988.

[13] P.M.T. Broersen. "Finite Sample Criteria for Autoregressive Order Selection," *IEEE Trans. on Signal Processing*, Vol.48, No.12, pp.3550-3558, Mec. 2000.