Ambiguity elimination in frequency–domain subspace identification

Koen Eneman[1,2] Marc Moonen[3]

[1] ExpORL – Dept. Neurosciences – Katholieke Universiteit Leuven
O & N 2, Herestraat 49 bus 721,
B–3000 Leuven – Belgium
koen.eneman@med.kuleuven.be

[2] GroepT Leuven Engineering School
Vesaliusstraat 13,
B–3000 Leuven – Belgium
koen.eneman@groept.be

[3] ESAT/SCD – Katholieke Universiteit Leuven
Kasteelpark Arenberg 10,
B–3001 Leuven (Heverlee) – Belgium
marc.moonen@esat.kuleuven.be

January 8, 2007

Abstract

During the last years, many blind subspace–based system identification techniques have been developed for channel equalization in digital communications. These techniques have been extended to speech enhancement applications, with limited success so far. In this paper both time–domain and subband/frequency–domain subspace–based system identification techniques are presented that are intended for speech enhancement and audio signal processing applications. Furthermore, an appropriate ambiguity elimination algorithm is proposed to remove the subband–dependent ambiguity factors that occur in frequency–domain subspace identification.

List of symbols:

Lower case bold–faced letters are used to denote vectors and upper case bold–faced letters to denote matrices. In addition the following notation is used throughout the paper:

$\star$	convolution operation
$(\cdot)_{D\downarrow}$	downsampling operator, retains each Dth sample
$\mathcal{F}$	Fourier transform
$x[k]$	discrete–time–domain signal x
$\underline{x}(f)$	frequency–domain representation of $x[k]$
$X(z)$	z–transform of $x[k]$
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of $\mathbf{A}$
$\mathbf{A}^\dagger$	pseudo–inverse of $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of $\mathbf{A}$
$\mathbf{1}$	vector with all elements equal to 1
$\mathbf{0}$	zero matrix
$\mathbf{0}_N$	$N \times N$ zero matrix
$\mathbf{0}_{M\times N}$	$M \times N$ zero matrix
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{J}$	exchange matrix : zero matrix with 1's along its main anti–diagonal
$\mathbf{F}$	DFT matrix
$[\mathbf{T}(z)]_{m,n}$	element (m, n) of polynomial matrix $\mathbf{T}(z)$

Other notation is explained in the text.

1 Introduction

During the last decades, many blind subspace–based system identification techniques have been successfully derived for channel equalization in digital communications [6] [7] [11] [12]. Later on, these techniques have been extended and applied to speech enhancement applications [5], with limited success so far. There are mainly three reasons to explain this. First of all, finding the correct system order is a first, yet essential step in subspace identification. Unfortunately, system order mismatches typically lead to an important decrease in the quality of the identified model, certainly for higher–order systems, which are regularly encountered in speech enhancement. Secondly, the quality of the model is compromised by the additive noise that is superimposed on the signals. Subspace techniques are known to be quite sensitive to small amounts of additive noise. Finally, subspace–based identification procedures tend to give rise to a high algorithmic cost for parameter settings that are typically used in speech and audio applications. Advanced matrix operations are required, which result in a computational cost of the order of $\mathcal{O}(N^3)$, where N is the length of the unknown transmission paths, and a memory storage capacity that is a multiple of N^2. This easily leads to computational and memory requirements that exceed the capabilities of many modern computer systems.
Dereverberation probably is the most important application of blind subspace identification techniques in speech enhancement. Various hands–free communication applications such as teleconferencing, hands–free telephony and voice–controlled systems suffer from degraded signal quality, which is most often attributed to the presence of so–called acoustic echo or background noise–like disturbances. A third type of disturbance, which is generally less often considered, is caused by the reverberation that is added to the signals as they propagate through the recording room and thereby reflect against walls, objects and people. Although signals (music e.g.) may sound more pleasant when reverberation is added, (especially for speech signals) the intelligibility is typically reduced. In order to cope with this kind of deformation, dereverberation or deconvolution techniques are called for. Whereas enhancement techniques for acoustic echo and noise reduction are well–known in the literature, high–quality, computationally efficient dereverberation algorithms are, to the best of our knowledge, not yet available. An overview and evaluation of speech dereverberation algorithms can be found in [2].

2 Blind subspace–based system identification

During the last decade, many blind system identification techniques have been developed for channel equalization in digital communications [7] [11]. These techniques can also be extended to other types of applications such as audio signal processing and signal enhancement.

2.1 Data model

Consider the $\mathcal{M}$–channel setup of figure 1. An unknown signal x is filtered by unknown impulse responses $h_1 \ \ldots \ h_{\mathcal{M}}$, resulting in $\mathcal{M}$ sensor signals $y_1 \ \ldots \ y_{\mathcal{M}}$. Sometimes, a compensator $\boldsymbol{\mathcal{C}}$ is added to reconstruct the unknown signal x, which is the case for the dereverberation application mentioned in section 1.
The following data model will be used throughout the text:

$$\mathbf{y}[k] = \mathbf{H} \cdot \mathbf{x}[k], \tag{1}$$

relating the system input to the sensor outputs, where

$$\mathbf{y}[k] = \left[\begin{array}{ccc|c|ccc} y_1[k] & \ldots & y_1[k-L+1] & \ldots & y_{\mathcal{M}}[k] & \ldots & y_{\mathcal{M}}[k-L+1] \end{array} \right]^T, \tag{2}$$

and

$$\mathbf{x}[k] = \left[\begin{array}{cccc} x[k] & x[k-1] & \ldots & x[k-L-N+2] \end{array} \right]^T. \tag{3}$$

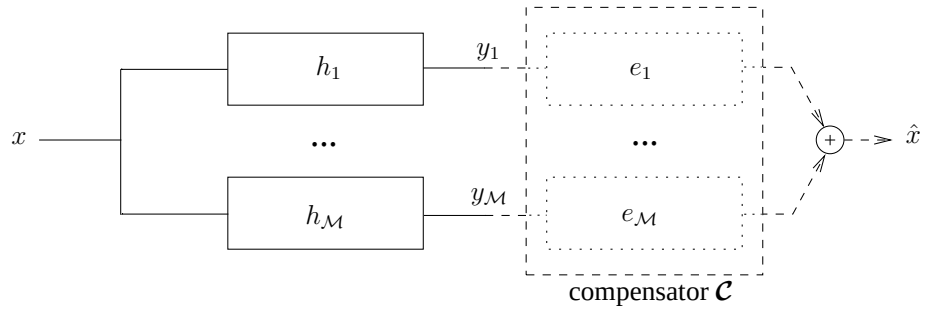

Figure 1: Multi–channel setup

Furthermore,

$$\mathbf{H} = \left[\begin{array}{ccc} \mathbf{H}_1^T & \dots & \mathbf{H}_{\mathcal{M}}^T \end{array} \right]^T, \tag{4}$$

with

$$\mathbf{H}_m \stackrel{\forall m}{=} \begin{bmatrix} \mathbf{h}_m^T & & \\ & \mathbf{h}_m^T & \\ & \ddots & \\ & & \mathbf{h}_m^T \end{bmatrix}, \qquad \mathbf{h}_m \stackrel{\forall m}{=} \begin{bmatrix} h_m[0] \\ \vdots \\ h_m[N-1] \end{bmatrix}, \tag{5}$$

where it is assumed that $h_1 \ \dots \ h_{\mathcal{M}}$ are FIR filters of length N.

2.2 Multi–channel subspace identification

System identifications aims to estimate the system impulse responses $h_1 \ \dots \ h_{\mathcal{M}}$, or in other words the system matrix $\mathbf{H}$. Thereto, the $L \times K$ Toeplitz matrices

$$\mathbf{Y}_m[k] \stackrel{\forall m}{=} \begin{bmatrix} y_m[k-K+1] & y_m[k-K+2] & \dots & y_m[k] \\ y_m[k-K] & y_m[k-K+1] & \dots & y_m[k-1] \\ \vdots & \ddots & \ddots & \vdots \\ y_m[k-K-L+2] & y_m[k-K-L+3] & \dots & y_m[k-L+1] \end{bmatrix}, \tag{6}$$

are defined. It follows from Eqs. 1–3 and 6 that

$$\mathbf{Y}[k] = \left[\begin{array}{ccc} \mathbf{Y}_1^T[k] & \dots & \mathbf{Y}_{\mathcal{M}}^T[k] \end{array} \right]^T = \mathbf{H} \underbrace{\left[\begin{array}{ccc} \mathbf{x}[k-K+1] & \dots & \mathbf{x}[k] \end{array} \right]}_{\mathbf{X}[k]}. \tag{7}$$

If $L \geq N$,

$$\mathbf{v}_{mn} \ = \ \left[\begin{array}{c|c|c|c|c} \mathbf{0}_{1\times(n-1)L} & \mathbf{h}_m^T \quad \mathbf{0}_{1\times(L-N)} & \mathbf{0}_{1\times(m-n-1)L} & -\mathbf{h}_n^T \quad \mathbf{0}_{1\times(L-N)} & \mathbf{0}_{1\times(\mathcal{M}-m)L} \end{array} \right]^T \tag{8}$$

can be defined. Then, for each pair (n, m) for which $1 \leq n < m \leq \mathcal{M}$, it is seen that

$$\mathbf{v}_{mn}^T \mathbf{H} \mathbf{X}[k] = \mathbf{v}_{mn}^T \mathbf{Y}[k] = \mathbf{0}, \tag{9}$$

as $\mathbf{v}_{mn}^T \mathbf{H} = \left[\begin{array}{cccccccc} w_{mn}[0] & \dots & w_{mn}[2N-2] & 0 & \dots & 0 \end{array} \right]$, where $w_{mn} = h_m \star h_n - h_n \star h_m$ is equal to zero. Hence, the transmission paths can be found in the left null space of $\mathbf{Y}[k]$, which has dimension

$$\nu = \mathcal{M}L - \underbrace{\text{rank}\{\mathbf{Y}[k]\}}_{r}. \tag{10}$$

By appropriately combining the ν basis vectors[1] $\mathbf{v}_\rho$, $\rho = r+1 \ldots \mathcal{M}L$, which span the left null space of $\mathbf{Y}[k]$, the filter h_m can be computed up to within a constant ambiguity factor α_m. This can for instance be done by solving the following set of equations :

$$\begin{bmatrix} \mathbf{v}_{r+1} & \ldots & \mathbf{v}_{\mathcal{M}L} \end{bmatrix} \begin{bmatrix} \beta_{r+1}^{(m)} \\ \vdots \\ \beta_{\mathcal{M}L-1}^{(m)} \\ 1 \end{bmatrix} = \left[\begin{array}{c} \alpha_m \mathbf{h}_m \\ \hline \mathbf{0}_{(L-N)\times 1} \\ \hline \mathbf{0}_{(m-2)L\times 1} \\ \hline -\alpha_m \mathbf{h}_1 \\ \hline \mathbf{0}_{(L-N)\times 1} \\ \hline \mathbf{0}_{(\mathcal{M}-m)L\times 1} \end{array}\right], \qquad \forall m : 1 < m \leq \mathcal{M}. \tag{11}$$

Eq. 11 can be rearranged, resulting into the following set of $\mathcal{M}L$ equations in $\nu+2N-1$ unknowns

$$\left[\begin{array}{ccc} \mathbf{v}_{r+1} & \ldots & \mathbf{v}_{\mathcal{M}L-1} \end{array}\right. \left|\begin{array}{c|c} \mathbf{0}_N & \mathbf{I}_N \\ \hline \mathbf{0}_{(L-N)\times N} & \mathbf{0}_{(L-N)\times N} \\ \hline \mathbf{0}_{(m-2)L\times N} & \mathbf{0}_{(m-2)L\times N} \\ \hline -\mathbf{I}_N & \mathbf{0}_N \\ \hline \mathbf{0}_{(L-N)\times N} & \mathbf{0}_{(L-N)\times N} \\ \hline \mathbf{0}_{(\mathcal{M}-m)L\times N} & \mathbf{0}_{(\mathcal{M}-m)L\times N} \end{array}\right] \underbrace{\begin{bmatrix} \beta_{r+1}^{(m)} \\ \vdots \\ \beta_{\mathcal{M}L-1}^{(m)} \\ \alpha_m \mathbf{h}_1 \\ \alpha_m \mathbf{h}_m \end{bmatrix}}_{\boldsymbol{\epsilon}_m} = -\mathbf{v}_{\mathcal{M}L}. \tag{12}$$

The vector of unknowns $\boldsymbol{\epsilon}_m$ can be computed if Eq. 12 defines an (over)determined set of equations that has full column rank $\nu+2N-1$. Observe that in practice an (over)determined set of equations is obtained if the left-hand matrix in Eq. 12 does not have more columns than rows, i.e. if $\mathcal{M}L \geq L+N-1$, which by inserting Eq. 10 reduces to $\text{rank}\{\mathbf{Y}[k]\} \geq 2N-1$. An exact solution to Eq. 12 does exist in the noise–free case. If noise is present Eq. 12 has to be solved in a least–squares sense.
In order to eliminate the different ambiguity factors α_m, it is sufficient to compare the coefficients of e.g. $\alpha_2 \mathbf{h}_1$ with $\alpha_m \mathbf{h}_1$ for $m > 2$. In this way the different scaling factors α_m can be compensated for, such that only a single overall ambiguity factor α remains.

2.3 Channel order estimation

From Eq. 12 the transmission paths h_m can be computed [7], provided that the channel order N is known. Now, observe from Eq. 7 that

$$\text{rank}\{\mathbf{Y}[k]\} = \min\{\mathcal{M}L, L+N-1\} \tag{13}$$

if it is assumed that

1. $K \geq L+N-1$, i.e. $\mathbf{X}[k]$ has more columns than rows
2. $\mathbf{X}[k]$ has full rank $\min\{K, L+N-1\} = L+N-1$
3. $\mathbf{H}$ has full rank $\min\{\mathcal{M}L, L+N-1\}$.

These assumptions are fulfilled if signal x is randomly chosen or obtained from a generic setup and if the impulse responses h_m do not have common zeros. So for generic systems for which $K \geq L+N-1$, the rank of $\mathbf{Y}[k]$ equals $L+N-1$ if $L \geq \frac{N-1}{\mathcal{M}-1}$. Hence, the channel order can be found from

$$N = \text{rank}\{\mathbf{Y}[k]\} - L + 1, \tag{14}$$

provided that there is no noise added to the system. If there is noise in the system one typically attempts to identify a "gap" in the singular value spectrum to determine the rank of $\mathbf{Y}[k]$. This

[1] Assuming $\mathbf{Y}^T[k] \overset{\text{SVD}}{=} \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, $\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_r & \mathbf{v}_{r+1} & \ldots & \mathbf{v}_{\mathcal{M}L} \end{bmatrix}$ is the singular value decomposition of $\mathbf{Y}^T[k]$.

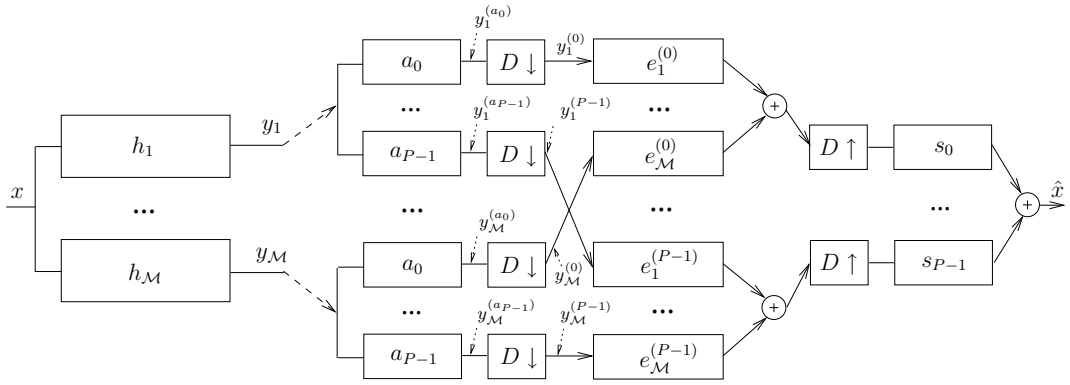

Figure 2: Multi–channel subband setup

gap is due to a difference in amplitude between the large singular values, which are supposed to correspond to the desired signal, and the smaller noise-related singular values.
Finally, following the same reasoning it appears that for multi-channel generic systems, Eq. 12 is an (over)determined set of equations with full column rank if $L \geq N$ and $K \geq L + N - 1$. Hence, once N is known, the transmission paths can be found based on Eq. 12.

3 Subband–domain subspace–based identification

In order to overcome the high computational and memory requirements of the time–domain subspace approach subband processing can be put forward as an alternative.

3.1 Subband implementation scheme

In a subband implementation all sensor signals $y_m[k]$ are fed into identical analysis filter banks $\{a_0, ..., a_{P-1}\}$, as shown in figure 2, and are in this way divided in P subbands. All subband signals are subsequently D–fold subsampled. Unlike in the standard time–domain approach of section 2, the system model is estimated based on the downsampled signals $y_m^{(p)}[k]$. The processed subband signals can be equalized by $e_m^{(p)}[k]$ and be upsampled, and are then finally recombined in the synthesis filter bank $\{s_0, ..., s_{P-1}\}$, leading to the system output $\hat{x}$. As the channel estimation procedure is performed in the subband domain at a downsampled sampling rate, a substantial cost reduction is expected.
If there are P subbands that are D–fold subsampled, one might expect that the transmission path length reduces to $\frac{N}{D}$, lowering the memory storage requirements from $\mathcal{O}(N^2)$ to $\mathcal{O}(P\frac{N^2}{D^2})$. As typically $P \approx D$, it follows that $\mathcal{O}(P\frac{N^2}{D^2}) \approx \mathcal{O}(\frac{N^2}{D})$. As far as the computational cost is concerned not only the matrix dimensions are reduced, also the updating frequency has been lowered by a factor D, leading to a huge cost reduction from $\mathcal{O}(N^3)$ to $\mathcal{O}(P\frac{N^3}{D^4}) \approx \mathcal{O}(\frac{N^3}{D^3})$. In practice however, the cost reduction is less spectacular, as the transmission path length will have to be larger than $\frac{N}{D}$ to appropriately model the acoustics (see also section 3.2). Secondly, we so far neglected the filter bank cost, which will furthermore lower the complexity gain that can be reached with the subband approach. Nevertheless, a significant overall cost reduction can be obtained, given the $\mathcal{O}(N^3)$ dependency of the algorithm.
In order to reduce the amount of overall signal distortion that is introduced by the filter banks and the subsampling, perfect or nearly perfect reconstruction filter banks are employed [3] [10]. Oversampled filter banks, for which $P > D$, are applied to minimize furthermore the amount of aliasing distortion that is added to the subband signals during the downsampling. DFT modulated filter bank schemes are then typically preferred. In many applications very simple so–called DFT filter banks are used (see also section 3.3).

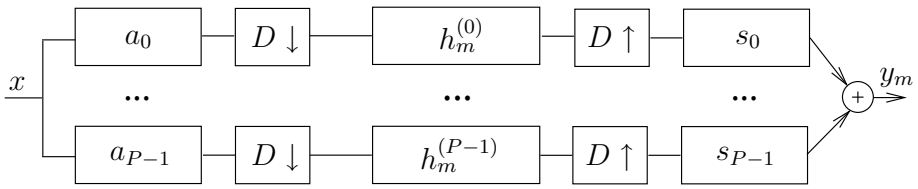

Figure 3: Subband–domain representation of $h_m[k]$

Summarizing, the advantages of a subband implementation are the substantial cost reduction and the subband dependent processing, which is expected to give rise to improved performance. The disadvantages are extra processing delay, possible signal distortion and aliasing effects caused by the subsampling [4].

3.2 Subband–dependent data model

If we leave out the subsampling for the time being, the following subband–dependent data model, similar to Eq. 7, can be put forward:

$$\mathbf{Y}^{(a_p)}[k] = \mathbf{H} \cdot \mathbf{X}^{(a_p)}[k], \qquad \forall p = 0 : P-1. \tag{15}$$

$\mathbf{H}$ is defined by Eqs. 3 and 5, and $\mathbf{Y}^{(a_p)}[k]$ and $\mathbf{X}^{(a_p)}[k]$ are constructed similar to Eq. 7, based on the subband signals $x^{(a_p)}[k] = a_p \star x[k]$ and $y_m^{(a_p)}[k] = a_p \star y_m[k] = a_p \star h_m \star x[k] = h_m \star x^{(a_p)}[k]$. If the analysis filter bank filters a_p are frequency selective and filter out contingent frequency bands with a total bandwidth smaller than $1/D$, the subband signals can be D–fold subsampled without introducing aliasing. The subsampling operation however affects the data model. It was proven in [4] that for a P–band, D–fold downsampled ideally frequency selective DFT modulated filter bank

$$y_m^{(p)}[n] = h_m^{(p)} \star x^{(p)}[n], \tag{16}$$

where $y_m^{(p)}[n] = \left(y_m^{(a_p)}[k]\right)_{D\downarrow}$ and $x^{(p)}[n] = \left(x^{(a_p)}[k]\right)_{D\downarrow}$ are D–fold subsampled equivalents of the filtered fullband signals $y_m^{(a_p)}[k]$ and $x_m^{(a_p)}[k]$, and $h_m^{(p)}[n]$ is an infinitely long, anti–causal interpolation–like filter, with decaying amplitude for increasing values of $|n|$:

$$h_m^{(p)}[n] \overset{\forall m}{=} \left(h_m[l] \star \left(e^{-j\frac{2\pi lp}{P}} \operatorname{sinc}\left(\frac{\pi l}{D}\right)\right)\right)_{D\downarrow}. \tag{17}$$

Hence, for perfect system identification infinitely long subband filters need to be estimated. In practice however, we hope to obtain a quasi–perfect identification by taking into account only the most dominant N_s^D filter taps of $h_m^{(p)}[n]$, where typically $N_s^D = \mathcal{O}(\frac{N}{D})$.

3.3 Ambiguity elimination

With blind system identification techniques the transmission paths can only be estimated up to within a constant factor. Contrary to the fullband approach where a global uncertainty factor α is encountered (see section 2.2), in a subband implementation there is an ambiguity factor $\alpha^{(p)}$ in each subband. This leads to significant signal distortion if the ambiguity factors $\alpha^{(p)}$ are not compensated for.

Assume that a set of filters $\hat{h}_m^{(p)}[n]$, $m = 1 : \mathcal{M}$, $p = 0 : P-1$, are computed with a blind subspace channel estimation procedure in the subband domain. If there is no noise in the system and the filters $\hat{h}_m^{(p)}[n]$ are sufficiently long, then $\hat{h}_m^{(p)}[n] = \alpha^{(p)} h_m^{(p)}[n]$, where $h_m^{(p)}[n]$, $p = 0 : P-1$, are a subband representation (cf. Eq. 17) of the true m–th time–domain transmission path filter $h_m[k]$, as shown in figure 3.

Remark that in general, due to the ambiguity, the set of filters $\hat{h}_m^{(p)}[n]$ does not describe a valid

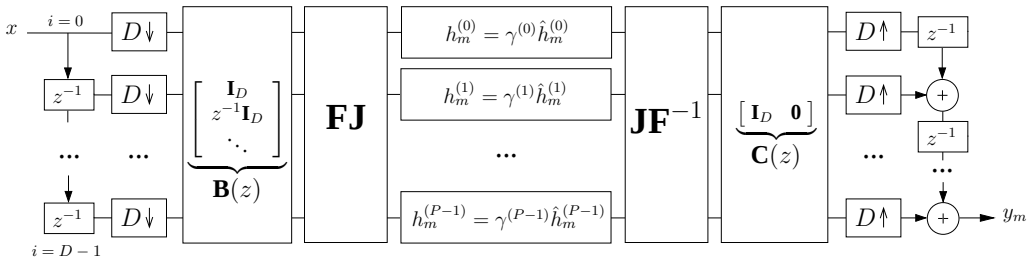

Figure 4: Partitioned block frequency–domain representation of $h_m[k]$. Matrix $\mathbf{J}$ is the $P \times P$ exchange matrix and $\mathbf{F}$ is the $P \times P$ DFT matrix. The $P \times D$ polynomial matrix $\mathbf{B}(z)$ and the $D \times P$ matrix $\mathbf{C}(z)$ are defined in the figure.

linear convolution operation. In order to remove the ambiguity, scaling factors $\gamma^{(p)} = 1/\alpha^{(p)}$ need to be found such that a valid convolution is obtained. In that case, $\gamma^{(p)}\hat{h}_m^{(p)}[n]$ will approach the true transmission paths $h_m^{(p)}[n]$. In this way the subband filters $\hat{h}_m^{(p)}[n]$ are transformed into a valid linear convolution operation.

In this section we will derive an ambiguity elimination procedure for subband schemes that make use of simple DFT filter banks. As mentioned in section 3.1, DFT filter banks are often used in practical applications for their low complexity and ease of implementation. If the analysis filter bank is defined as

$$a_p[k] = e^{-j\frac{2\pi(P-1-k)p}{P}}, \qquad k = 0 : P-1, \quad p = 0 : P-1, \tag{18}$$

for instance, and is combined with the following synthesis filter bank

$$s_p[k] = \frac{1}{P} e^{j\frac{2\pi(P-D+k)p}{P}}, \qquad k = 0 : D-1, \quad p = 0 : P-1, \tag{19}$$

an elegant perfect reconstruction subband system is obtained (in the absence of intermediate subband operations, i.e. if all $h_m^{(p)} = 1$). Despite the rather poor frequency characteristic of the filter banks an efficient solution is obtained: observe from Eqs. 18 and 19 that the filter bank operations come down to the computation of an (I)FFT. If the filter banks are as defined in Eqs. 18 and 19, figure 3 can be redrawn, resulting in figure 4. It can be shown that this subband system is related to overlap–save fast convolution techniques [4] [8].

In the case of DFT filter banks as defined in Eqs. 18 and 19, the filters $h_m^{(p)}[n]$ are a frequency–domain representation of the true time–domain transmission paths $h_m[k]$. For a typical parameter setting [4]:

$$\begin{bmatrix} h_m^{(0)}[n] \\ h_m^{(1)}[n] \\ \vdots \\ h_m^{(P-1)}[n] \end{bmatrix} = \mathbf{F} \begin{bmatrix} h_m[nD] \\ \vdots \\ h_m[nD+D-1] \\ \mathbf{0}_{(P-D)\times 1} \end{bmatrix}, \qquad \forall n = 0 : \frac{N}{D} - 1, \tag{20}$$

where it is assumed that the transmission path length N is a multiple of D. Otherwise, $h_m[k]$ must be padded with zeros. Moreover, one typically chooses $P = 2D$, where P is a power of 2 such that fast Fourier transforms can be used.

If now a set of frequency–domain filters $\hat{h}_m^{(p)}[n]$ are computed with the scheme of figure 4 using a blind channel estimation procedure as presented in section 2.2, a set of scaling factors $\gamma^{(p)} = 1/\alpha^{(p)}$ has to be found such that a valid linear convolution is obtained. As we want the subband system of figure 4 to be equivalent to the time–domain filter $h_m[k]$ (up to a delay), the chain of subband operations should represent a time–invariant filter operation. It can be proven that a time–

invariant filter operation is obtained[2] if

$$\mathbf{T}_m(z) = \mathbf{C}(z) \underbrace{\mathbf{J}\mathbf{F}^{-1}\mathrm{diag}\{\gamma^{(p)}\hat{H}_m^{(p)}(z)\}\mathbf{F}\mathbf{J}}_{\mathbf{U}_m(z)} \mathbf{B}(z) \tag{21}$$

is a pseudo–circulant matrix [10], i.e.

$$\left.\begin{array}{rcl} [\mathbf{T}_m(z)]_{q+1,n+1} &=& [\mathbf{T}_m(z)]_{q,n} \\ [\mathbf{T}_m(z)]_{q+1,0} &=& z^{-1}\,[\mathbf{T}_m(z)]_{q,D-1} \end{array}\right\} \quad \begin{array}{l} \forall q = 0 : D-2 \\ \forall n = 0 : D-2. \end{array} \tag{22}$$

As $\mathbf{F}\mathbf{J}\mathbf{U}_m(z)\mathbf{J}\mathbf{F}^{-1}$ is diagonal, $\mathbf{U}_m(z)$ must be a right–circulant matrix :

$$\mathbf{T}_m(z) = \underbrace{\left[\begin{array}{cc} \mathbf{I}_D & \mathbf{0} \end{array}\right]}_{\mathbf{C}(z)} \underbrace{\left[\begin{array}{cccc} U_{m_0}(z) & U_{m_1}(z) & \ldots & U_{m_{P-1}}(z) \\ U_{m_{P-1}}(z) & U_{m_0}(z) & \ldots & U_{m_{P-2}}(z) \\ \vdots & \ddots & \vdots & \\ U_{m_1}(z) & U_{m_2}(z) & \ldots & U_{m_0}(z) \end{array}\right]}_{\mathbf{U}_m(z)} \underbrace{\left[\begin{array}{c} \mathbf{I}_D \\ z^{-1}\mathbf{I}_D \\ \ddots \end{array}\right]}_{\mathbf{B}(z)}. \tag{23}$$

It can be easily verified based on Eqs. 22 and 23 that

$$U_{m_p}(z) = 0, \qquad \forall p : P - D < p < P. \tag{24}$$

must be true for $\mathbf{T}_m(z)$ to be a pseudo–circulant matrix.
As $\mathbf{U}_m(z)$ is a circulant matrix, the diagonal elements of $\mathbf{F}\mathbf{J}\mathbf{U}_m(z)\mathbf{J}\mathbf{F}^{-1}$ are the DFT coefficients of the first row of $\mathbf{U}_m(z)$. Hence, from Eqs. 21, 23 and 24 it follows that

$$\left[\begin{array}{c} \gamma^{(0)}\hat{H}_m^{(0)}(z) \\ \vdots \\ \gamma^{(P-1)}\hat{H}_m^{(P-1)}(z) \end{array}\right] = \mathbf{F} \left[\begin{array}{c} U_{m_0}(z) \\ \vdots \\ U_{m_{P-D}}(z) \\ \mathbf{0}_{(D-1)\times 1} \end{array}\right]. \tag{25}$$

By comparing Eq. 25 with 20, and by requiring that $\gamma^{(p)}\hat{h}_m^{(p)}[n] = h_m^{(p)}[n]$, it is observed that the filters $U_{m_p}(z)$ contain the coefficients of $h_m[k]$. As equation 20 holds for a typical parameter setting that is used for the scheme of figure 4, only the first D filters $U_{m_p}(z)$ are different from 0 :

$$U_{m_p}(z) = 0, \qquad \forall p : D \leq p < P. \tag{26}$$

Hence, it follows that $U_{m_p}(z)$, $p = 0 : D-1$ are the Dth order polyphase components of $h_m[k]$. Observe from Eq. 25 that for each filter tap n and each channel m an equation

$$\boldsymbol{\mathcal{H}}_m^{[n]} \underbrace{\left[\begin{array}{c} \gamma^{(0)} \\ \vdots \\ \gamma^{(P-1)} \end{array}\right]}_{\boldsymbol{\gamma}} = \left[\begin{array}{c} \mathbf{u}_m^{[n]} \\ \mathbf{0}_{(P-Q)\times 1} \end{array}\right] \tag{27}$$

can be set up with

$$\boldsymbol{\mathcal{H}}_m^{[n]} = \left[\begin{array}{c} \boldsymbol{\mathcal{H}}_{m1}^{[n]} \\ \boldsymbol{\mathcal{H}}_{m2}^{[n]} \end{array}\right] = \mathbf{F}^{-1} \left[\begin{array}{ccc} \hat{h}_m^{(0)}[n] & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \hat{h}_m^{(P-1)}[n] \end{array}\right], \tag{28}$$

[2] $\hat{H}_m^{(p)}(z)$ is the z–transform of $\hat{h}_m^{(p)}[n]$

$\boldsymbol{\mathcal{H}}_{m1}^{[n]}$ a $Q \times P$ matrix, $\boldsymbol{\mathcal{H}}_{m2}^{[n]}$ a $(P-Q) \times P$ matrix and

$$\mathbf{u}_m^{[n]} = \begin{bmatrix} u_{m_0}[n] \\ \vdots \\ u_{m_{Q-1}}[n] \end{bmatrix}, \tag{29}$$

with $U_{m_p}(z) = \sum_{n=0}^{L_n-1} u_{m_p}[n] z^{-n}$ and $Q \leq P - D + 1$. Remark that for the scheme of figure 4, $Q = D$ and $L_n = N/P$. Now, define

$$\boldsymbol{\mathcal{H}}_m = \begin{bmatrix} \boldsymbol{\mathcal{H}}_m^{[0]} \\ \vdots \\ \boldsymbol{\mathcal{H}}_m^{[L_n-1]} \end{bmatrix}, \quad \mathbf{u}_m = \left[\begin{array}{c} \mathbf{u}_m^{[0]} \\ \hline \mathbf{0}_{(P-Q)\times 1} \\ \hline \vdots \\ \hline \mathbf{u}_m^{[L_n-1]} \\ \hline \mathbf{0}_{(P-Q)\times 1} \end{array}\right] \quad \text{and} \quad \bar{\mathbf{u}}_m = \begin{bmatrix} \mathbf{u}_m^{[0]} \\ \mathbf{u}_m^{[1]} \\ \vdots \\ \mathbf{u}_m^{[L_n-1]} \end{bmatrix}. \tag{30}$$

Then, $\boldsymbol{\gamma}$ can be computed by combining Eq. 27 for each filter tap n and each channel m

$$\begin{bmatrix} \boldsymbol{\mathcal{H}}_1 \\ \vdots \\ \boldsymbol{\mathcal{H}}_{\mathcal{M}} \end{bmatrix} \underbrace{\begin{bmatrix} \gamma^{(0)} \\ \vdots \\ \gamma^{(P-1)} \end{bmatrix}}_{\boldsymbol{\gamma}} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{\mathcal{M}} \end{bmatrix}, \tag{31}$$

which can be rewritten as

$$\left[\begin{array}{c|ccc} \boldsymbol{\mathcal{H}}_{11}^{(0)} & \mathbf{I}_Q & \mathbf{0}_{Q\times(L_n\mathcal{M}-2)Q} & \mathbf{0}_Q \\ \boldsymbol{\mathcal{H}}_{12}^{(0)} & \mathbf{0} & \mathbf{0}_{(P-Q)\times(L_n\mathcal{M}-2)Q} & \mathbf{0} \\ \hline \vdots & \ddots & \ddots & \vdots \\ \hline \boldsymbol{\mathcal{H}}_{\mathcal{M}1}^{(L_n-1)} & \mathbf{0}_Q & \mathbf{0}_{Q\times(L_n\mathcal{M}-2)Q} & \mathbf{I}_Q \\ \boldsymbol{\mathcal{H}}_{\mathcal{M}2}^{(L_n-1)} & \mathbf{0} & \mathbf{0}_{(P-Q)\times(L_n\mathcal{M}-2)Q} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\gamma} \\ \hline -\bar{\mathbf{u}}_1 \\ \vdots \\ -\bar{\mathbf{u}}_{\mathcal{M}} \end{array}\right] = \mathbf{0}. \tag{32}$$

Hence, vector $\boldsymbol{\gamma}$ can be found in the right null space of the $(P-Q)L_n\mathcal{M} \times P$ matrix $\boldsymbol{\mathcal{H}}$:

$$\underbrace{\begin{bmatrix} \boldsymbol{\mathcal{H}}_{12}^{(0)} \\ \vdots \\ \boldsymbol{\mathcal{H}}_{\mathcal{M}2}^{(L_n-1)} \end{bmatrix}}_{\boldsymbol{\mathcal{H}}} \boldsymbol{\gamma} = \mathbf{0}. \tag{33}$$

If $(P-Q)L_n\mathcal{M} < P$ the null space becomes multi–dimensional, which should be avoided as in this case $\boldsymbol{\gamma}$ cannot be retrieved based on Eq. 33. The vectors $\mathbf{u}_m^{[n]}$ (and hence $h_m[k]$) can be obtained from

$$\begin{bmatrix} \mathbf{u}_1^{[0]} \\ \vdots \\ \mathbf{u}_{\mathcal{M}}^{[L_n-1]} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{H}}_{11}^{(0)} \\ \vdots \\ \boldsymbol{\mathcal{H}}_{\mathcal{M}1}^{(L_n-1)} \end{bmatrix} \boldsymbol{\gamma}. \tag{34}$$

In the presence of noise, $\boldsymbol{\gamma}$ can be found as the right singular vector that corresponds to the smallest singular value. Remark furthermore that $\boldsymbol{\gamma}$ and the set of filters $h_m[k]$ can only be estimated up to within an unknown global constant.

Rahbar, Manton et al. [9] proposed a similar, probably more noise robust method, to compensate for the frequency–dependent ambiguity in the case of 1–tap filters $\hat{h}_m^{(p)}[n]$. They maximize the energy in the first part of the impulse response (first N coefficients of $\mathbf{F}^{-1}[\hat{h}_m^{(0)} \ldots \hat{h}_m^{(P-1)}]^T$) under the constraint that the total energy of $\mathbf{F}^{-1}[\hat{h}_m^{(0)} \ldots \hat{h}_m^{(P-1)}]^T$ stays constant. This then leads to a

generalized eigenvalue decomposition.
Both ambiguity elimination algorithms are quite computationally involving, as the eigenvalue or singular value decomposition has to be computed of a large matrix. This certainly is a drawback of the subband–domain subspace identification approach. It further appeared that both ambiguity elimination methods are sensitive to system order mismatches. Finally, in our simulations with the algorithm of Rahbar and Manton we sometimes encountered numerical problems during the computation of the generalized eigenvalue decomposition.
In [1] a promising dereverberation algorithm was presented that relies on 1–dimensional frequency–domain subspace tracking. An LMS type updating scheme was proposed that offers a low cost alternative to the matrix based algorithms of section 2. A major drawback of this approach is that the norm of the transfer function matrix $\boldsymbol{\beta}(f) = \begin{bmatrix}\underline{h}_1(f) & \dots & \underline{h}_{\mathcal{M}}(f)\end{bmatrix}$ (with $\underline{h}_m(f)$ the frequency-domain representation of $h_m[k]$, see figure 1) needs to be known in advance. We can get around this by measuring parameter $\boldsymbol{\beta}(f)$ beforehand. This is however unpractical, hence an alternative is to fix $\boldsymbol{\beta}(f)$ to an environment–independent constant, e.g. $\boldsymbol{\beta}(f) = \mathbf{1}$. Now, if the frequency–domain subspace estimation algorithm is combined with the ambiguity elimination algorithm presented in Eqs. 27–34 the transmission paths $\underline{h}_m(f)$ can be determined up to within a global scaling factor. Hence, $\boldsymbol{\beta}(f)$ can be computed and does not need to be known in advance. Uncertainties on $\boldsymbol{\beta}(f)$, however, which are due to the limited precision of the channel estimation procedure and the lagging behind of the algorithm during tracking of time–varying transmission paths, affect the performance of the subspace tracking algorithm.

4 Acknowledgements

This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian State Interuniversity Poles of Attraction Programmes P5/22 and P5/11, the Concerted Research Action GOA-MEFISTO-666 and IWT project 000401 (MUSETTE-II) of the Flemish Government, and was partially sponsored by Philips–PDSL. The scientific responsibility is assumed by its authors.

5 Conclusions

In this paper subspace identification techniques are discussed for speech enhancement and audio signal processing. Both time–domain and subband/frequency–domain algorithms are presented. It is shown that the unknown transmission paths can be estimated only up to within a global constant, provided that appropriate ambiguity elimination algorithms are employed.

References

[1] S. Affes and Y. Grenier. A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech. *IEEE Transactions on Speech and Audio Processing*, 5(5):425–437, September 1997.

[2] K. Eneman, J. Duchateau, M. Moonen, D. Van Compernolle, and H. Van hamme. Assessment of Dereverberation Algorithms for Large Vocabulary Speech Recognition Systems. In *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003)*, pages 2689–2692, Geneva, Switzerland, September 2003.

[3] K. Eneman and M. Moonen. DFT Modulated Filter Bank Design for Oversampled Subband Systems. *Signal Processing*, 81(9):1947–1973, September 2001.

[4] K. Eneman and M. Moonen. Hybrid Subband/Frequency–Domain Adaptive Systems. *Signal Processing*, 81(1):117–136, January 2001.

[5] S. Gannot and M. Moonen. Subspace Methods for Multi–Microphone Speech Dereverberation. In *Proceedings of the 2001 IEEE/EURASIP International Workshop on Acoustic Echo and Noise Control (IWAENC01)*, pages 47–50, Darmstadt, Germany, September 2001.

[6] S. Haykin. *Blind Deconvolution*. Prentice Hall, Englewood Cliffs, New Jersey, 1994.

[7] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue. Subspace Methods for the Blind Identification of Multichannel FIR Filters. *IEEE Transactions on Signal Processing*, 43(2):516–525, February 1995.

[8] A. Oppenheim and R. Schafer. *Digital Signal Processing*. Prentice Hall, Englewood Cliffs, New Jersey, 1975.

[9] K. Rahbar, J. Reilly, and J. Manton. A Frequency Domain Approach to Blind Identification of MIMO FIR Systems driven by Quasi–Stationary Signals. In *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP02)*, Orlando, Florida, May 2002.

[10] P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[11] A.-J. van der Veen, S. Talwar, and A. Paulraj. Blind Identification of FIR Channels carrying Multiple Finite Alphabet Signals. In *Proceedings of the 1995 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP95)*, pages 1213–1216, Detroit, Michigan, May 1995.

[12] P. Vandaele. Space–Time Processing Algorithms for Smart Antennas in Wireless Communication Networks. PhD thesis, Katholieke Universiteit Leuven, Heverlee, Belgium, November 1999.