IDIAP RESEARCH REPORT

Filter Bank Design based on Minimization of Individual Aliasing Terms for Minimum Mutual Information Subband Adaptive Beamforming

Kenichi Kumatani [a] John McDonough [b]
Stefan Schacht [b] Dietrich Klakow [b]
Philip N. Garner [a] Weifeng Li [a]

IDIAP–RR 07-77

January 2008

[a] IDIAP Research Institute, Martigny, Switzerland
[b] Spoken Language Systems at Saarland University in Saarbrücken, Germany

IDIAP Research Institute **www.idiap.ch**
Av. des Prés–Beudin 20 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 07-77

Filter Bank Design based on Minimization of Individual Aliasing Terms for Minimum Mutual Information Subband Adaptive Beamforming

Kenichi Kumatani John McDonough Stefan Schacht Dietrich Klakow
Philip N. Garner Weifeng Li

January 2008

Abstract.

This paper presents new filter bank design methods for subband adaptive beamforming. In this work, we design analysis and synthesis prototypes for modulated filter banks so as to minimize each aliasing term individually. We then drive the *total response error* to null by constraining these prototypes to be *Nyquist(M)* filters. Thereafter those modulated filter banks are applied to a speech separation system which extracts a target speech signal. In our system, speech signals are first transformed into the subband domain with our filter banks, and the subband components are then processed with a beamforming algorithm. Following beamforming, post-filtering and binary masking are further performed to remove residual noises.

We show that our filter banks can suppress the *residual aliasing distortion* more than conventional ones. Furthermore, we demonstrate the effectiveness of our design techniques through a set of automatic speech recognition experiments on the multi-channel speech data from the *PASCAL Speech Separation Challenge*. The experimental results prove that our beamforming system with the proposed filter banks achieves the best recognition performance, a 39.6 % word error rate (WER), with half the amount of computation of that of the conventional filter banks while the perfect reconstruction filter banks provided a 44.4 % WER.

1 Introduction

There has been great interest in subband adaptive processing applications. Subband adaptive filtering can reduce the computational complexity associated with time domain adaptive filters and improve the convergence property in estimating filter coefficients [1]. However, the filter bank design for adaptive filtering poses problems not encountered in more traditional applications such as speech coding. In [2], de Haan et al. noted that perfect reconstruction (PR) filter banks were not suitable for beamforming applications because PR is achieved through alias cancellation [3, §5], which can reconstruct an input signal correctly only if the outputs of the individual subbands are *not* subject to arbitrary magnitude scaling and phase shifts. They also proposed a method to design analysis and synthesis prototypes for modulated filter banks so as to minimize the weighted combination of the *response error* and *aliasing distortion*. The filter banks proposed in [2] are referred as de Haan filter banks here.

In this work, we drive the response error defined in [2] to null by constraining the analysis and synthesis prototypes to be *Nyquist(M)* filters [3, §4.6.1]. Thereafter, the minimization of the aliasing distortions is shown to reduce to the solution of an eigenvalue problem in the case of the analysis prototype, and to the solution of a set of linear equations in the case of the synthesis prototype. We also discuss the performance limitation of our filter banks due to numerical problems caused by singular matrices, and propose an alternate solution for the special case which can eliminate not only the total response error but also residual aliasing distortion completely. The filter banks proposed here are applied to *minimum mutual information* (MMI) beamforming where the *active weight vectors* are estimated so that mutual information of two beamforming outputs is minimized [4]. After that, the separated speech is further processed with Zelinski post-filtering and binary masking [5] in order to remove diffuse noises and a residual interference signal.

We show the effectiveness of our methods through speech recognition experiments on the far-field speech data from the *PASCAL Speech Separation Challenge*. The data were recorded in a reverberant room, not artificially convoluted with measured room impulse responses and the position of speaker's head varies as well as speaking volume.

The balance of this work is organized as follows. In Section 2, we review the definition of a modulated filter bank. Section 3 considers the design of suitable analysis and synthesis prototypes for the modulated filter banks. In particular, Sections 3.1 and 3.2 briefly present the design methods of [2] for prototypes, and then show how slight modifications of those techniques can produce prototypes with zero response error and minimal aliasing distortions. In Section 4, we first compare the residual aliasing distortion of our method with de Haan filter banks. We then describe the configurations for speech recognition experiments and compare our design technique with that originally proposed in [2] as well as the popular paraunitary PR design. Finally, in Section 5 we present our conclusions and plans for future work.

2 Modulated Filter Banks

Figure 1 shows a schematic of a *modulated filter bank* with M *subbands* and a *decimation factor* of D.

Following [2], we define the impulse responses $h[n]$ and $g[n]$ for analysis and synthesis prototypes respectively, and express those modulated versions according to

$$h_m[n] \;=\; h[n]\,W_M^{-mn} \leftrightarrow H_m(z) = H(zW_M^m) \tag{1}$$
$$g_m[n] \;=\; g[n]\,W_M^{-mn} \leftrightarrow G_m(z) = G(zW_M^m) \tag{2}$$

where $W_M = e^{-j2\pi/M}$ denotes the M-th root of unity.

As indicated in Figure 1, the input spectrum $X(z)$ is first processed with analysis filters $H_m(z)$. Then the decimators expand the filtered signals $V_m(z)$. The decimated signal $X_m(z)$ consists of the sum of a stretched output of the m-th filter bank and $D-1$ aliasing terms. At this point, the "fixed" subband weights F_m can be applied to the decimated signals $X_m(z)$. The expanders then compress the weighted signals $Y_m(z)$. In the last step, the compressed signals $U_m(z)$ are processed with the

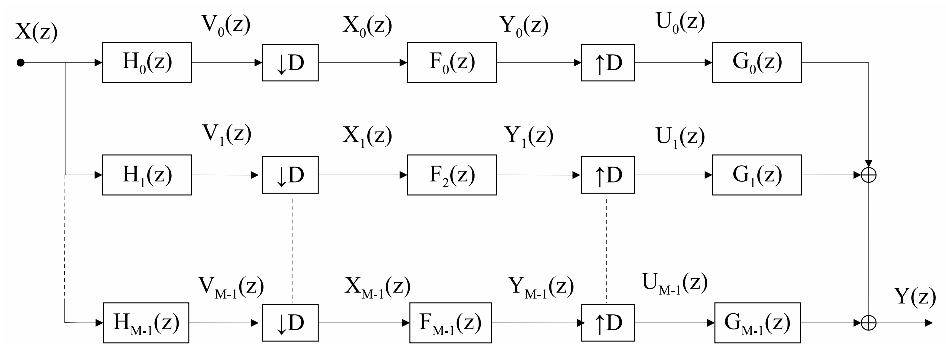

Figure 1: Schematic of a modulated filter bank.

synthesis filters $G_m(z)$ in order to suppress the spectral images created by expanders, and the outputs of the synthesis filters are summed together.

Upon defining

$$A_{m,d}(z) = \frac{1}{D}\, F_m\, H(zW_M^m W_D^d)\, G(zW_M^m), \tag{3}$$

the relationship between the input and output signals can be written as

$$Y(z) = \sum_{d=0}^{D-1} A_d(z)\, X(zW_D^d) \tag{4}$$

where

$$A_d(z) = \sum_{m=0}^{M-1} A_{m,d}(z). \tag{5}$$

The transfer function $A_0(z)$ produces the desired signal, while the remaining transfer functions $\{A_d(z)\}$ for $d = 1, \ldots, D-1$ give rise to the residual aliasing in the output signal.

3 Prototype Design

3.1 Analysis Prototype Design

In order to design the analysis prototype $h[n]$, de Haan *et al.* [2] define the objective function

$$\epsilon_{\mathbf{h}} = \alpha_{\mathbf{h}} + \beta_{\mathbf{h}} \tag{6}$$

where the *passband response error* is

$$\alpha_{\mathbf{h}} = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \left| H(e^{j\omega}) - e^{-j\omega\tau_H} \right|^2 d\omega, \tag{7}$$

and the *inband-aliasing distortion* is given by

$$\beta_{\mathbf{h}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{d=1}^{D-1} \left| H(e^{j\omega/D} W_D^d) \right|^2 d\omega. \tag{8}$$

In (7) the *desired filter bank response* corresponds to a pure delay of τ_H samples.

Defining $\mathbf{h} = \begin{bmatrix} h[0] & h[1] & \cdots h[L_{\mathbf{h}}-1] \end{bmatrix}^T$, de Haan *et al.* [2] then demonstrate that the passband response error can be expressed as

$$\alpha_{\mathbf{h}} = \mathbf{h}^T \mathbf{A} \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + 1 \tag{9}$$

where the components of $\mathbf{A}$ and $\mathbf{b}$ can be expressed as

$$A_{i,j} = \frac{\sin(\omega_p(j-i))}{\omega_p(j-i)} \text{ and } b_i = \frac{\sin(\omega_p(\tau_H - i))}{\omega_p(\tau_H - i)}.$$

The inband-aliasing term (8) can be expressed as

$$\beta_{\mathbf{h}} = \mathbf{h}^T \mathbf{C} \mathbf{h} \tag{10}$$

where the components of $\mathbf{C}$ can then be expressed as

$$C_{i,j} = \frac{\varphi[j-i] \, \sin\left(\frac{\pi(j-i)}{D}\right)}{\pi(j-i)}$$

and

$$\varphi[n] = D \sum_{k=-\infty}^{\infty} \delta[n-kD] - 1.$$

Combining all terms above, they then seek to minimize the objective function

$$\epsilon_{\mathbf{h}} = \alpha_{\mathbf{h}} + \beta_{\mathbf{h}} = \mathbf{h}^T(\mathbf{A}+\mathbf{C})\mathbf{h} - 2\mathbf{h}^T\mathbf{b} + 1 \tag{11}$$

Nyquist(M) Filters

The impulse response of a *Nyquist(M)* or *M-th band filter* [3, §4.6.1] satisfies

$$h[Mn] = \begin{cases} c, & n = m_d \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

If $H(z)$ is the Nyquist(M) filter, then the output of analysis filter bank would be equivalent to the input delayed by $m_d M$ samples; see McDonough et al. [6] for the proof.

Notice that (12) represents a much stronger condition than that aimed at by the minimization of (7), in that (12) implies the response error will vanish, not just for the pass band of a single filter, but for the entire working spectrum, including the transition bands between the passbands of adjacent filters. Hence, we replace the term $\alpha_{\mathbf{h}}$ in the optimization criterion (6) with a constraint of the form (12), then minimize the inband-aliasing distortion subject to this constraint. The inband-aliasing distortion reduces to (10), whose optimization clearly admits the trivial solution $\mathbf{h} = \mathbf{0}$. To exclude this solution, we impose the additional constraint $\mathbf{h}^T\mathbf{h} = 1$, which is readily achieved through the method of *undetermined Lagrange multipliers*. We posit the modified objective function

$$f(\mathbf{h}) = \mathbf{h}^T\mathbf{C}\mathbf{h} + \lambda(\mathbf{h}^T\mathbf{h} - 1) \tag{13}$$

where λ is a *Lagrange multiplier*. Then, by solving $\mathbf{C}\mathbf{h} = -\lambda\mathbf{h}$, we can find the optimal prototype $\mathbf{h}$. Clearly $\mathbf{h}$ is an eigenvector of $\mathbf{C}$. Moreover, in order to ensure $\mathbf{h}$ minimizes (10), it must be the eigenvector associated with the *smallest* eigenvalue of $\mathbf{C}$. Note that, in order to ensure that $\mathbf{h}$ satisfies (12), we must delete those rows and columns of $\mathbf{C}$ corresponding to the components of $\mathbf{h}$ that are identically zero. We then solve the eigenvalue problem (26) for the remaining components of $\mathbf{h}$, and finally reassemble the complete prototype by appropriately concatenating the zero and non-zero components. This is similar to the construction of the *eigenfilter* described in [3, §4.6.1].

3.2 Synthesis Prototype Design

In order to design the synthesis prototype, in [2], de Haan et al. take as an objective function

$$\epsilon_{\mathbf{g}}(\mathbf{h}) = \gamma_{\mathbf{g}}(\mathbf{h}) + \delta_{\mathbf{g}}(\mathbf{h}) \tag{14}$$

where the *total response error* is defined as

$$\gamma_{\mathbf{g}}(\mathbf{h}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|A_0(e^{j\omega}) - e^{-j\omega\tau_T}\right|^2 \, d\omega. \tag{15}$$

τ_T is the total analysis-synthesis filter bank delay and the *residual aliasing distortion* is

$$\delta_{\mathbf{g}}(\mathbf{h}) = \frac{1}{2\pi}\sum_{d=1}^{D-1}\sum_{m=0}^{M-1}\int_{-\pi}^{\pi} \left|A_{m,d}(e^{j\omega})\right|^2 \, d\omega. \tag{16}$$

Through manipulations similar to those used in deriving the quadratic objective criterion for the analysis filter bank, it can be shown that

$$\gamma_{\mathbf{g}}(\mathbf{h}) = \mathbf{g}^T \mathbf{E} \mathbf{g} - 2\mathbf{g}^T \mathbf{f} + 1. \tag{17}$$

The components of $\mathbf{E}$ and $\mathbf{f}$ are given by

$$E_{i,j} = \frac{M^2}{D^2} \sum_{k=-\infty}^{\infty} h^*[kM - i] h[kM - j] \text{ and } f_i = \frac{M}{\pi D} h[\tau_T - i].$$

The quadratic form for the residual aliasing distortion is

$$\delta_{\mathbf{g}}(\mathbf{h}) = \mathbf{g}^T \mathbf{P} \mathbf{g} \tag{18}$$

where the components of $\mathbf{P}$ are given by

$$P_{i,j} = \frac{M}{D^2} \sum_{l=-\infty}^{\infty} h^*[l + j]\, h[l + i]\, \varphi[i - j].$$

In [2], de Haan et al. introduce a weighting factor v to emphasize either the total response error ($0 < v < 1$) or residual aliasing distortion ($v > 1$):

$$\epsilon_{\mathbf{g}}(\mathbf{h}) = \gamma_{\mathbf{g}}(\mathbf{h}) + v\delta_{\mathbf{g}}(\mathbf{h}) = \mathbf{g}^T (\mathbf{E} + v\mathbf{P})\mathbf{g} - 2\mathbf{g}^T \mathbf{f} + 1 \tag{19}$$

Nyquist(M) Constraint

As with the analysis prototype, we impose the Nyquist(M) constraint on the *complete analysis-synthesis prototype* $(h * g)[n]$ such that

$$(h * g)[Mn] = \begin{cases} c, & n = m_d \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

in which case the total response error (15) must be identically zero. Subject to this constraint, we minimize the residual aliasing distortion (19). Satisfaction of (20) clearly reduces to a set of linear constraints of the form

$$\mathbf{H}^T \mathbf{g} = \mathbf{c} \tag{21}$$

where

$$\mathbf{H} = [\mathbf{h}_{-m+1}, \ldots, \mathbf{h}_0, \ldots, \mathbf{h}_{m-1}]\ , \tag{22}$$

$$\mathbf{c} = [0, \ldots, c, \ldots, 0]^T\ , \tag{23}$$

and $\mathbf{h}_k$ is obtained by shifting a time-reversed version of $\mathbf{h}$ by kM samples and padding with zeros as needed.

We can again resort to the method of undetermined Lagrange multipliers for this problem and obtain a solution of a synthesis prototype:

$$\mathbf{g} = \mathbf{P}^{-1} \mathbf{H} \left(\mathbf{H}^T \mathbf{P}^{-1} \mathbf{H} \right)^{-1} \mathbf{c}. \tag{24}$$

3.3 Alternate method for a special case

The optimal prototypes can be obtained by the methods mentioned above if matrices $\mathbf{C}$ and $\mathbf{P}$ are not singular. However, the matrices are often singular when decimation factor D is small.

If $\mathbf{C}$ is singular, we can consider its nullspace, $\mathbf{C}_{null}$, which consists of column vectors $\mathbf{q} \in \mathbf{R}^n$: $\mathbf{Cq} = \mathbf{0}$. Obviously, inband-aliasing distortion (10) can be driven to null by an analysis prototype which is represented as a linear combination of bases of the nullspace $\mathbf{C}_{null}\ \mathbf{x}$. We can then use the free parameters $\mathbf{x}$ for minimizing passband response error (9). Such a solution can be expressed as

$$\mathbf{h} = \mathbf{C}_{null} (\mathbf{C}_{null}^T \mathbf{A} \mathbf{C}_{null})^{-1} \mathbf{C}_{null}^T \mathbf{b} \tag{25}$$

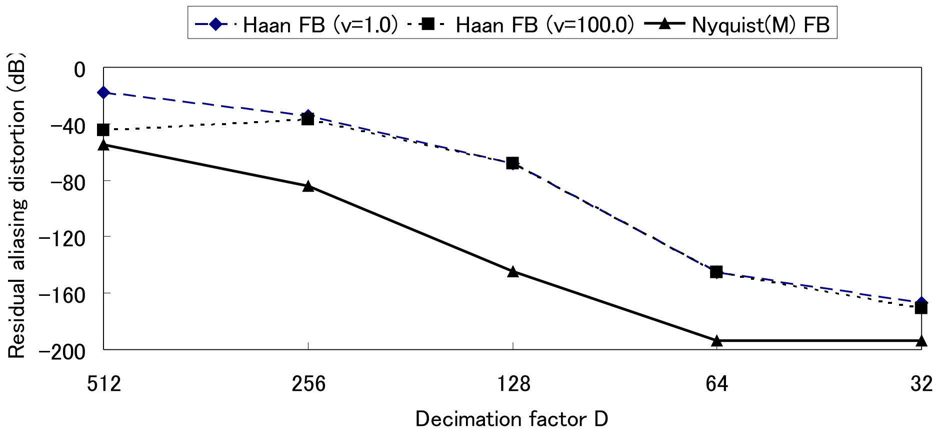

Figure 2: Residual aliasing distortion $\epsilon_{\mathbf{g}}(\mathbf{h})$ for decimation factor D, which was calculated with the number of subbands $M = 512$ and the filter length $L_{\mathbf{h}} = 1024$. The values for $D \leq 64$ were obtained with the alternate method.

where rows and columns of $\mathbf{C}_{null}$, $\mathbf{A}$ and $\mathbf{b}$ corresponding to the components of h that are identically zero are deleted, and $\mathbf{h}$ is reassembled so as to keep the Nyquist(M) constraint. For the synthesis prototype design, we can also erase residual aliasing distortion (18) in a similar manner. Defining the nullspace of $\mathbf{P}$ to be $\mathbf{P}^{null}$, we can express the synthesis prototype $\mathbf{g} = \mathbf{P}^{null}\mathbf{y}$. Then by substituting into (21), we have

$$\mathbf{y} = (\mathbf{H}^T \mathbf{P}^{null})^{+} \mathbf{c} \tag{26}$$

where $(\cdot)^{+}$ indicates the peseudoinverse of $(\cdot)$. If the number of column vectors of $\mathbf{P}^{null} \geq 2m - 1$, we can find a synthesis prototype $\mathbf{g} = \mathbf{P}^{null}\mathbf{y}$ with zero total response error and residual aliasing distortion. In practice, when the inband-aliasing distortion is very small, $\mathbf{P}$ becomes *computationally* singular.

4 Experiments

The residual aliasing distortion indicates how small the filter bank can keep the total response error even if the PR property is destroyed by arbitrary magnitude scaling and phase shifts. Figure 2 presents the residual aliasing distortions from (18), where de Haan filter banks are calculated with weighting factor $v = 1.0$ and 100.0, respectively. It is clear from Figure 2 that the proposed filter banks can provide better suppression performance for aliasing.

We performed far-field automatic speech recognition (ASR) experiments on development data from the *PASCAL Speech Separation Challenge* (SSC); see Lincoln *et al.* [7] for a description of the data collection apparatus. Prior to beamforming, we first estimated the speaker's position with the *Orion* source tracking system [8]. In addition to the speaker's position, Orion is also capable of determining when each speaker is active. This information is useful for speaker adaptation, given that utterances spoken by one speaker were often much longer than those spoken by the other. Based on the average speaker position estimated for each utterance, a beamformer was constructed. The active weights were estimated so as to achieve the minimum mutual information (MMI) of the outputs from the beamformers [4]. In this work, we assumed that subband snapshots were Gaussian-distributed. In addition to MMI beamforming, Zelinski post-filtering and binary masking [5] were performed.

We did four decoding passes on the waveforms obtained with the beamforming algorithms described above. Each pass of decoding used a different acoustic model or speaker adaptation scheme. Speaker adaptation parameters were estimated using the word lattices generated during the prior pass. The detail of the speech recognizer is presented in [9].

Table 1: WERs without post-filtering for every filter bank design algorithm after every decoding passes.

Filter bank	Pass (%WER)			
	1	2	3	4
FFT	88.5	71.1	58.8	55.5
PR	87.7	65.2	54.0	50.7
De Haan	88.7	68.2	56.1	53.5
Nyquist(M)	88.5	67.0	55.6	52.5

Table 2: WERs with post-filtering and binary masking for every filter bank design algorithm after every decoding passes. WERs of the Nyquist(M) FB with $M = 512$ & $D = 64$ were obtained with the alternate method.

Filter bank	Parameters		Pass (%WER)			
	M	D	1	2	3	4
PR	64	-	83.7	61.5	47.5	44.7
	512	-	84.6	60.5	47.6	44.4
De Haan	64	32	82.4	59.2	46.2	43.3
	512	256	83.9	59.1	43.2	41.3
	512	128	81.6	58.9	43.2	40.3
	512	64	82.7	57.7	42.7	39.6
Nyquist(M)	64	32	80.7	57.0	44.3	42.0
	512	256	84.1	58.6	43.4	40.6
	512	128	81.8	54.9	42.2	39.6
	512	64	81.4	56.5	42.6	40.3

We first conducted speech recognition experiments on speech separated with MMI beamforming only and investigated four methods : (1) normal frequency domain processing with a FFT [10], (2) cosine modulated filter bank [3, 6], which yields PR under optimal conditions, (3) de Haan filter bank, and (4) Nyquist(M) filter banks proposed here. Table 1 shows the word error rates (WERs) for every filter bank when we set parameters for each filter bank to obtain the best recognition performance. MMI beamforming with the PR filter banks provided the best recognition performance when post-filtering was not applied. Although it certainly scaled magnitudes and shifted phases of input subband components, we didn't observe strong aliasing noises. We consider that MMI beamforming with a Gaussian assumption can estimate active weight vectors while keeping aliasing cancellation. On the other hand, de Haan filter banks have the total response error which could deteriorate the recognition performance. FFT analysis achieved significantly worse performance than all the subband processing methods.

Finally we ran recognition experiments on speech enhanced with post-filtering and binary masking following MMI beamforming. In that case, the PR property was not kept because of the rapid change of filter weights. We observed the aliasing distortions when the PR filter banks were used. In contrast, de Haan and the proposed filter banks can suppress such aliasing noises because those filter banks are designed so as to minimize aliasing terms individually. Table 2 shows the WERs for each filter bank with different numbers of subbands M and decimation factors D. From Table 2, we can see that the systems equipped with de Haan and Nyquist(M) filter banks can reduce the absolute WER by about 5% compared to those with the PR filter banks. This proves that the PR filter bank is not suitable for adaptive processing. It is also clear from Table 2 that the proposed method achieved a bigger WER reduction than de Haan's algorithm. In particular, the improvements of the recognition performance are significant with $M = 64$ since differences of the residual aliasing and response errors between the Nyquist(M) and de Haan filter banks are larger than those with $M = 512$. The proposed filter banks achieved the best recognition performance, WER 39.6 % with the number of subbands $M = 512$ and decimation factor $D = 128$. On the other hand, de Haan filter banks provided the same number with $M = 512$ and $D = 64$. Therefore, our method can be thought of as halving the computational cost of that of de Haan.

5 Conclusions

In this work, we have proposed a new design method for filter banks that is suitable for adaptive processing. We have demonstrated the effectiveness of our design techniques through a set of automatic speech recognition experiments on the multi-channel speech data from the *PASCAL Speech Separation Challenge*. The proposed method achieved the smallest WER (39.6 %) with half as much computational costs as de Haan filter banks, while the PR filter provided a 44.4 % WER.

References

[1] John J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, pp. 14–37, 1992.

[2] Jan Mark de Haan, Nedelko Grbic, Ingvar Claesson, and Sven Erik Nordholm, "Filter bank design for subband adaptive microphone arrays," *IEEE Trans. Speech Audio Proc.*, vol. 11, no. 1, pp. 14–23, Jan. 2003.

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

[4] Kenichi Kumatani, Tobias Gehrig, Uwe Mayer, Emilian Stoimenov, John McDonough, and Matthias Wölfel, "Adaptive beamforming with a minimum mutual information criterion," *IEEE Transactions on Audio, Speech and Language Processing*, vol. in the printing stage.

[5] Iain McCowan, Ivan Himawan, and Mike Lincoln, "A microphone array beamforming approach to blind speech separation," in *Proc. MLMI*, 2007.

[6] John McDonough and Kenichi Kumatani, "Filter bank design for beamforming," Tech. Rep. 110, Spoken Language Systems, Saarland University, August 2007.

[7] M. Lincoln, I. McCowan, I. Vepa, and H. K. Maganti, "The multi-channel Wall Street Journal audio visual corpus (mc-wsj-av): Specification and initial experiments," in *Proc. ASRU*, 2005, pp. 357–362.

[8] Tobias Gehrig, Ulrich Klee, John McDonough, Shajith Ikbal, Matthias Wölfel, and Christian Fügen, "Tracking and beamforming for multiple simultaneous speakers with probabilistic data association filters," in *Proc. Interspeech*, 2006, pp. 2594–2597.

[9] John McDonough, Kenichi Kumatani, Tobias Gehrig, Emilian Stoimenov, Uwe Mayer, Stefan Schacht, Matthias Woelfel, and Dietrich Klakow, "To separate speech! a system for recognizing simultaneous speech," in *Proc. MLMI*, 2007.

[10] Futoshi Asano, Shiro Ikeda, Michiaki Ogawa, Hideki Aso, and Nobuhiko Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signals," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 3, pp. 204–215, May 2003.