DESIGN AND EVALUATION OF NONUNIFORM DFT FILTER BANKS IN SUBBAND MICROPHONE ARRAYS

J. M. de Haan, N. Grbić, I. Claesson

Blekinge Institute of Technology
Department of Telecommunications
and Signal Processing
372 25 Ronneby, Sweden

S. Nordholm

Australian Telecommunications
Research Institute
Curtin University
Perth, Australia

ABSTRACT

This paper presents a method for the design of nonuniform DFT filter banks for subband beamforming. Filter banks designed with the method are evaluated in subband beamforming in a real-world microphone array application. Different source positions in array applications give rise to different signal delays, which means that adaptive beamformers in the subbands alter the phase information of the subband signals in order to extract the source from a noisy background. Phase alterations in the subbands lead to signal degradations when perfect reconstruction filter banks are used for the subband decomposition and reconstruction. The objective of the proposed design is to minimize the magnitude of all aliasing components individually, such that aliasing distortion is minimized although phase alterations occur in the subbands. The proposed method is evaluated in a car hands-free mobile telephony environment with real speech signals and the results show that the performance can be increased by several decibel when using nonuniform filter banks instead of uniform filterbanks while maintaining the length of the subband filters.

1. INTRODUCTION

Perfect reconstruction filter banks have been of great interest in subband coding, [1]. However, these filter banks are less suitable for subband adaptive filtering, since the perfect reconstruction property with aliasing cancellation is not maintained when the subband signals are modified by individual subband filters with arbitrary magnitude and phase response. This implies that aliasing is present in the reconstructed output of subband adaptive filters. A design method for uniform DFT filter banks for subband adaptive filtering which deals with this problem has been presented in [2].

Nonuniform filter banks have been of interest in speech enhancement as a suitable model for the human auditory system, and they have been successfully applied to, e.g. speech recognition and speech coding. Nonuniform filter banks have also been proposed for subband adaptive filtering, e.g. in spectral subtraction for speech enhancement, [3]. A comprehensive theoretical study of nonuniform filter banks in subband beamforming is given in [4].

This paper proposes an extension to the method presented in [2] for the design of nonuniform analysis and synthesis DFT filter banks, where the main goal is that adaptive filtering in the subbands should cause minimal source signal degradations at the output. The design method aims at minimizing aliasing in the subband signals as well as minimizing magnitude, phase, and aliasing distortions in the reconstructed output signal.

The proposed method consists of two steps. In the first step the analysis filter bank is designed in such way that the aliasing terms in each subband are minimized individually, contributing to minimal aliasing at the synthesis filter bank output without allowing aliasing cancellation. In the second step the synthesis filter bank is designed to match the analysis filter bank where the analysis-synthesis response is optimized while all aliasing terms in the output signal are individually suppressed, rather than aiming at aliasing cancellation.

The filter banks are evaluated in subband beamforming for a real world microphone array arrangement in a handsfree voice communication situation in an automobile.

2. NONUNIFORM DFT FILTER BANKS

An M -channel nonuniform DFT filter bank can simply be obtained by replacing the delay operators in uniform DFT filter banks by first order Laguerre allpass functions, [5].

$$z^{-1} \rightarrow Q(z) = \frac{-\mu + z^{-1}}{1 - \mu z^{-1}} \quad (1)$$

The uniformity coefficient, $-1 < \mu < 1$, controls the bandwidth

Fig. 1. Polyphase implementation of a nonuniform analysis and synthesis filter banks with Laguerre allpass functions.

of the subbands. When $\mu = 0$, a uniform filter bank is obtained according to the method proposed in [2]. When μ is increased, the bandwidth of the subbands increases proportionally along the frequency axis. When μ is decreased, the bandwidth of the subbands decreases proportionally along the frequency axis.

The subband signals $X_m(z)$ are described by

$$X_m(z) = \frac{1}{D} \sum_{d=0}^{D-1} \sum_{l=0}^{M-1} X(z^{\frac{1}{D}} W_D^d) W_M^l h_l \left(Q(z^{\frac{1}{D}} W_D^d) \right)^l, \quad (2)$$

where $W_M = e^{-j2\pi/M}$, the number of subbands is M , and the decimation rate is D . The polyphase implementation of the analysis filter bank described by Eq. (2) is given in Fig. 1. Eq. (2) can be compactly written as

$$X_m(z) = \frac{1}{D} \sum_{d=0}^{D-1} X(z^{\frac{1}{D}} W_D^d) \mathbf{h}^T \phi_{m,d}(z). \quad (3)$$

The polyphase coefficient vector, $\mathbf{h}$, and the complex basis function vector, $\phi_{m,d}(z)$, are defined as

$$\mathbf{h} = [h_0, \dots, h_{M-1}]^T, \quad (4)$$

$$\phi_{m,d}(z) = [\phi_{m,d}^0(z), \dots, \phi_{m,d}^{M-1}(z)]^T, \quad (5)$$

with $\phi_{m,d}^l(z) = (W_M^l Q(z^{\frac{1}{D}} W_D^d))^l$. Using these definitions, the analysis filters $H_m(z)$ are described by

$$H_m(z) = \mathbf{h}^T \phi_{m,0}(z^D). \quad (6)$$

The reconstructed output signal can be expressed in terms of the subband signals $Y_m(z)$ according to

$$Y(z) = \sum_{m=0}^{M-1} \sum_{k=0}^{M-1} Y_m(z^D) W_M^{-mk} g_k (Q(z))^k. \quad (7)$$

The polyphase implementation of the synthesis filter bank described by Eq. (7) is given in Fig. 1. Eq. (7) can be written as

$$Y(z) = \sum_{m=0}^{M-1} Y_m(z^D) \mathbf{g}^T \psi_m(z). \quad (8)$$

The polyphase coefficient vector, $\mathbf{g}$, and the complex basis function vector, $\psi_m(z)$, are defined as

$$\mathbf{g} = [g_0, \dots, g_{M-1}]^T, \quad (9)$$

$$\psi_m(z) = [\psi_m^0(z), \dots, \psi_m^{M-1}(z)]^T, \quad (10)$$

where $\psi_m^k(z) = (W_M^{-k} Q(z))^k$. Using these definitions, the synthesis filters $G_m(z)$ are described by

$$G_m(z) = \mathbf{g}^T \psi_m(z). \quad (11)$$

Inserting Eq. (3) into Eq. (8), and by setting $Y_m(z) = X_m(z)$, i.e. no subband filtering is applied, the analysis-synthesis relation is obtained

$$Y(z) = \sum_{m=0}^{M-1} \sum_{d=0}^{D-1} X(z^{\frac{1}{D}} W_D^d) \mathbf{h}^T \Phi_{m,d}(z) \mathbf{g}, \quad (12)$$

where

$$\Phi_{m,d}(z) = \frac{1}{D} \phi_{m,d}(z^D) \psi_m^T(z). \quad (13)$$

The total response of the analysis-synthesis filter bank can be written as

$$T(z) = \mathbf{h}^T \Psi(z) \mathbf{g}, \quad (14)$$

where

$$\Psi(z) = \sum_{m=0}^{M-1} \sum_{d=0}^{D-1} \Phi_{m,d}(z). \quad (15)$$

3. ANALYSIS FILTER BANK DESIGN

The polyphase coefficients of the nonuniform DFT analysis filter bank can be designed using a least squares passband objective function for the lowpass prototype filter, $H_0(z)$,

$$\alpha_h = \sum_{i=0}^{I-1} |H_0(e^{j\omega_i}) - H_d(e^{j\omega_i})|^2, \quad \omega_i \in \Omega_p, \quad (16)$$

where the desired frequency response is $H_d(z)$, defined in the passband region $\Omega_p = [-\omega_p, \omega_p]$

$$H_d(e^{j\omega}) = e^{-j\omega\tau_h}, \quad \omega \in \Omega_p. \quad (17)$$

The frequency range in Eq. (16) is discretized into I equidistant frequency points, $i = 0, \dots, I-1$. Eq. (16) can be written in the quadratic form

$$\alpha_h = \mathbf{h}^T \mathbf{A} \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + I, \quad (18)$$

where

$$\mathbf{A} = \sum_{i=0}^{I-1} \phi_{0,0}(e^{j\omega_i D}) \phi_{0,0}^H(e^{j\omega_i D}), \quad (19)$$

and

$$\mathbf{b} = \sum_{i=0}^{I-1} \text{Re} \left\{ e^{j\omega_i \tau_h} \phi_{0,0}(e^{j\omega_i D}) \right\}. \quad (20)$$

The inband aliasing is simultaneously minimized by complementing the passband objective function with the inband aliasing objective function

$$\beta_h = \sum_{i=0}^{I-1} \sum_{d \in \Delta} |\mathbf{h}^T \phi_{M,d}(e^{j\omega_i})|^2, \quad \omega_i \in [-\pi, \pi], \quad (21)$$

where $\Delta = [0, \dots, \frac{D}{2} - 1, \frac{D}{2} + 1, \dots, D-1]$. Only the aliasing terms in subband $m = \frac{M}{2}$ are addressed because this subband will have the lowest signal-to-aliasing ratio since it has the highest bandwidth. Note that aliasing term $d = \frac{D}{2}$ is omitted in the sum of Eq. (21) since it contains the desired spectral content of subband $m = \frac{M}{2}$. Eq. (21) can be rewritten to

$$\beta_h = \mathbf{h}^T \mathbf{C} \mathbf{h}, \quad (22)$$

where

$$\mathbf{C} = \sum_{i=0}^{I-1} \sum_{d \in \Delta} \phi_{\frac{M}{2},d}(e^{j\omega_i}) \phi_{\frac{M}{2},d}^H(e^{j\omega_i}), \quad \omega_i \in [-\pi, \pi], \quad (23)$$

The total analysis filter bank design objective function is then

$$\epsilon_h = \alpha_h + \beta_h = \mathbf{h}^T (\mathbf{A} + \mathbf{C}) \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + I, \quad (24)$$

which is minimized with respect to $\mathbf{h}$ by solving the set of linear equations

$$(\mathbf{A} + \mathbf{C}) \mathbf{h} = \mathbf{b} \quad (25)$$

4. SYNTHESIS FILTER BANK DESIGN

The synthesis filter bank is designed by minimizing the total response error and the residual aliasing distortion. The total response error is defined as

$$\gamma_{\mathbf{g}}(\mathbf{h}) = \sum_{i=0}^{I-1} \left| T(e^{j\omega_i}) - T_d(e^{j\omega_i}) \right|^2, \quad \omega_i \in [-\pi, \pi], \quad (26)$$

which can be written in the quadratic form

$$\gamma_{\mathbf{g}}(\mathbf{h}) = \mathbf{g}^T \mathbf{E} \mathbf{g} - 2\mathbf{g}^T \mathbf{f} + I, \quad (27)$$

where

$$\mathbf{E} = \sum_{i=0}^{I-1} \Psi^H(e^{j\omega_i}) \mathbf{h}^* \mathbf{h}^T \Psi(e^{j\omega_i}), \quad (28)$$

and

$$\mathbf{f} = \sum_{i=0}^{I-1} \text{Re} \left\{ e^{j\omega_i} \Psi^H(e^{j\omega_i}) \mathbf{h}^* \right\}. \quad (29)$$

The residual aliasing distortion is defined as

$$\delta_{\mathbf{g}}(\mathbf{h}) = \sum_{i=0}^{I-1} \sum_{m=0}^{M-1} \sum_{d=1}^{D-1} \left| \mathbf{h}^T \Phi_{m,d}(e^{j\omega_i}) \mathbf{g} \right|^2, \quad (30)$$

which can be written in the quadratic form

$$\delta_{\mathbf{g}}(\mathbf{h}) = \mathbf{g}^T \mathbf{P} \mathbf{g}, \quad (31)$$

where

$$\mathbf{P} = \sum_{i=0}^{I-1} \sum_{m=0}^{M-1} \sum_{d=1}^{D-1} \Phi_{m,d}^H(e^{j\omega_i}) \mathbf{h}^* \mathbf{h}^T \Phi_{m,d}(e^{j\omega_i}). \quad (32)$$

The total synthesis filter bank design objective function is then

$$\epsilon_{\mathbf{g}}(\mathbf{h}) = \gamma_{\mathbf{g}}(\mathbf{h}) + \delta_{\mathbf{g}}(\mathbf{h}) = \mathbf{g}^T (\mathbf{E} + \mathbf{P}) \mathbf{g} - 2\mathbf{g}^T \mathbf{f} + I, \quad (33)$$

which is minimized with respect to $\mathbf{g}$ by solving the set of linear equations

$$(\mathbf{E} + \mathbf{P}) \mathbf{g} = \mathbf{f} \quad (34)$$

5. EVALUATION IN A MICROPHONE ARRAY APPLICATION

The performance of the nonuniform filter banks is evaluated in the case of a subband LS optimal beamformer with real data recorded in a hands-free car situation. In this situation, there is a target signal, an interference signal causing echo at the far end of the communication link, and background noise, see Fig. 2.

A linear array of six, $S = 6$, microphones was mounted in a car on the visor at the passenger side. The distance between the speaker position and the microphone array was 350 mm and the position was perpendicular to the array axis at the center point. The spacing between adjacent elements in the array was 50 mm. During a calibration phase, a known white noise sequence was emitted from a human shaped doll in order to determine the LS beamformer weights. Recordings of real background noise taken in a car, running at 110 km/h on a paved road, a real speech target signal and a real speech interference signal serve as evaluation signals for the beamformer performance.

Fig. 2. Subband FIR Beamforming Structure.

The sensor signals are decomposed into subbands using the proposed method. The number of subbands is set to $M = 16$. Non-critical decimation is used with decimation factor $D = 2$ for all filter bank scenarios. The uniformity coefficient is varied $\mu = 0, 0.1, 0.2, 0.3, 0.4$, which means that 5 scenarios are compared. In Table 1, the filter bank design measures are given for the different scenarios. In Fig. 3 the analysis filter responses and total responses are given for $\mu = 0$ and $\mu = 0.4$.

Fig. 3. Analysis Filter Responses and Analysis-Synthesis Response for $\mu = 0$ and $\mu = 0.4$.

Finite impulse response beamformers are used in the subbands. The least squares solution of the beamforming weights are found by

$$\mathbf{f}_{LS,opt}^{(m)} = [\hat{\mathbf{R}}_{cc}^{(m)}]^{-1} \hat{\mathbf{r}}_c^{(m)}, \quad (35)$$

where $\mathbf{f}^{(m)}$ is a vector with beamforming weights in subband m according to

$$\mathbf{f}^{(m)} = [\mathbf{f}_1^{(m)}, \dots, \mathbf{f}_S^{(m)}]^T \quad (36)$$

and $\mathbf{f}_s^{(m)}$ is the s -th channel impulse response in subband m . Subsequently, $\hat{\mathbf{R}}_{cc}^{(m)}$ is an autocorrelation matrix estimate, and $\hat{\mathbf{r}}_c^{(m)}$ is a crosscorrelation vector estimate obtained from data recorded during the calibration phase. In the calibration phase, ordinary background noise and white noise calibration sequences, emitted from the passenger and hands-free loudspeaker positions, are recorded independently.

In order to measure the microphone array performance, the

L_f	$\mu=0$	$\mu=0.1$	$\mu=0.2$	$\mu=0.3$	$\mu=0.4$
α_h	-93.37	-63.08	-52.96	-46.53	-36.34
β_h	-80.27	-66.25	-60.71	-43.20	-29.47
$\gamma_g(h)$	-90.90	-20.40	-23.07	-16.06	-17.87
$\delta g(h)$	-88.45	-46.97	-51.23	-46.40	-45.99
	dB	dB	dB	dB	dB

Table 1. Filter Bank Design Measures.

normalized distortion quantity, D , is introduced as

$$D = \frac{1}{2\pi} \int_{-\pi}^{\pi} |C_d \hat{P}_{y_s}(w) - \hat{P}_{x_s}(w)| dw, \quad (37)$$

where C_d is defined as

$$C_d = \frac{\int_{-\pi}^{\pi} \hat{P}_{x_s}(w) dw}{\int_{-\pi}^{\pi} \hat{P}_{y_s}(w) dw}, \quad (38)$$

and where $\hat{P}_{x_s}(w)$ is a spectral power estimate of a single sensor observation and $\hat{P}_{y_s}(w)$ is the spectral power estimate of the beamformer output, when the source signal is active alone. The constant C_d normalizes the mean output spectral power to that of the single sensor spectral power. The measure of distortion in Eq. (37), is the mean output spectral power deviation from the observed single sensor spectral power. Ideally, the distortion is zero. The distortion quantities for the different scenarios are given in Table 2.

L_f	$\mu=0$	$\mu=0.1$	$\mu=0.2$	$\mu=0.3$	$\mu=0.4$
2	-28.74	-29.22	-28.24	-25.78	-25.31
4	-26.78	-29.13	-27.16	-25.21	-24.51
6	-26.88	-28.74	-27.12	-25.20	-24.48
8	-27.20	-27.58	-26.15	-25.02	-24.15
	dB	dB	dB	dB	dB

Table 2. Distortion quantities for the different scenarios. The uniformity coefficients is set to $\mu = 0, 0.1, 0.2, 0.3, 0.4$ and filter length of the subband filters is set to $L_f = 2, 4, 6, 8$.

To measure the suppression performance of the microphone array, the normalized noise suppression quantity, S_N , is introduced as

$$S_N = \frac{\int_{-\pi}^{\pi} \hat{P}_{y_N}(w) dw}{C_d \int_{-\pi}^{\pi} \hat{P}_{x_N}(w) dw}, \quad (39)$$

and the normalized interference suppression quantity, S_I , is introduced as

$$S_I = \frac{\int_{-\pi}^{\pi} \hat{P}_{y_I}(w) dw}{C_d \int_{-\pi}^{\pi} \hat{P}_{x_I}(w) dw}, \quad (40)$$

where $\hat{P}_{y_N}(w)$ and $\hat{P}_{x_N}(w)$ are spectral power estimates of the beamformer output and the reference sensor observation, respectively, when the surrounding noise is active alone. In the same way $\hat{P}_{y_I}(w)$ and $\hat{P}_{x_I}(w)$ are spectral power estimates when the interference signal is active alone. The noise suppression and interference suppression quantities are given in Table 3 and Table 4 respectively, for the different scenarios.

L_f	$\mu=0$	$\mu=0.1$	$\mu=0.2$	$\mu=0.3$	$\mu=0.4$
2	5.10	9.59	5.87	7.12	6.15
4	7.99	8.85	8.14	8.28	6.97
6	9.78	9.13	8.37	8.06	7.94
8	10.48	10.84	10.67	10.68	9.81
	dB	dB	dB	dB	dB

Table 3. Noise suppression quantities for the different scenarios. The uniformity coefficients is set to $\mu = 0, 0.1, 0.2, 0.3, 0.4$ and filter length of the subband filters is set to $L_f = 2, 4, 6, 8$.

L_f	$\mu=0$	$\mu=0.1$	$\mu=0.2$	$\mu=0.3$	$\mu=0.4$
2	3.76	5.73	5.54	8.06	9.86
4	8.57	9.13	9.27	9.87	10.82
6	9.19	9.23	8.98	9.75	11.03
8	8.27	8.83	8.40	8.91	10.35
	dB	dB	dB	dB	dB

Table 4. Interference suppression quantities for the different scenarios. The uniformity coefficients is set to $\mu = 0, 0.1, 0.2, 0.3, 0.4$ and filter length of the subband filters is set to $L_f = 2, 4, 6, 8$.

6. CONCLUSIONS

A design method for nonuniform DFT filter banks with reduced inband and residual aliasing has been proposed for the application of subband adaptive microphone arrays. The design aims at minimizing aliasing components individually, in order to bound the aliasing effects when phase alterations are applied by adaptive beamformers. The evaluation of subband beamformers in a microphone array in a automobile environment with designed filter banks show that hands-free suppression levels are improved by 2-6 dB compared to the uniform filter bank at the expense of minor signal degradations while maintaining or even slightly improving the level of noise suppression (1-2 dB improvement).

7. REFERENCES

- [1] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, 1993.
- [2] J. M. de Haan, N. Grbić, I. Claesson, and S. Nordholm, "Design of Oversampled Uniform DFT Filter Banks with Delay Specification using Quadratic Optimization," in *Proceedings of ICASSP*, May 2001, pp. 3633-3636.
- [3] T. Gulzow, A. Engelsberg, and U. Heute, "Comparison of a Discrete Wavelet Transformation and a Nonuniform Polyphase Filter Bank Applied to Spectral Subtraction Speech Enhancement," in *Elsevier Signal Processing*, 1998, pp. 5-19.
- [4] S. Nordholm, I. Claesson, and N. Grbić, "Structures and Performance Limits in Subband Beamforming," *Submitted to IEEE Transactions on Speech and Audio Processing*, Apr. 2001.
- [5] M. Kappelan, B. Strauss, and P. Vary, "Flexible Nonuniform Filter Banks using Allpass Transformation of Multiple Order," in *Proceedings of EUSPICO*, 1996, pp. 1745-1748.