LEAST SQUARES DESIGN OF NONUNIFORM FILTER BANKS WITH EVALUATION IN SPEECH ENHANCEMENT

Jan Mark de Haan, Ingvar Claesson

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

Harald Gustafsson

Ericsson Mobile Platforms
221 83 Lund, Sweden

ABSTRACT

This paper presents a method for least squares design of nonuniform filter banks for application in subband signal processing. Design objectives aim to optimize the filter bank frequency response while minimizing subband and output aliasing. Aliasing is minimized although magnitude and phase changes affect the aliasing terms. Filter banks with increasing bandwidth are designed with the proposed method and evaluated in speech enhancement using a spectral subtraction algorithm. When using a nonuniform frequency resolution approximating that of the human auditory system it is shown that an increased noise reduction and SNR improvement is achieved while maintaining the speech quality for a fixed number of frequency-bands.

1. INTRODUCTION

Perfect reconstruction filter banks have been of great interest in subband coding [1]. The perfect reconstruction property with aliasing cancellation is not maintained when the subband signals are modified by individual adaptive subband filters with arbitrary magnitude and phase response. This implies that aliasing is present in the reconstructed output signal from the synthesis filter bank. Design methods for filter banks for subband processing dealing with this aliasing problem have been presented in [2, 3]. Nonuniform filter banks have been proposed for speech enhancement in [4]. Least squares approximation techniques for the design of filter banks has previously been presented in [5]. A two stage least squares design procedure can be found in [6].

This paper proposes a two stage frequency domain least squares method for the design of nonuniform analysis and synthesis filter banks with frequency domain criteria, similar to the approach presented in [2, 7]. The main goals with the design criteria are to minimize aliasing in the subbands and to minimize aliasing in the output signal with optimized amplitude and phase response. Parameters set prior to the design are number of subbands, group delays and subband dependent decimation factors.

With the introduction of the Wide-Band Adaptive Multi-Rate (WB-AMR) speech codec in GSM and UMTS networks the need of a noise reduction method handling a larger audio bandwidth of 50 Hz to 7 kHz is apparent [8]. Over this frequency range the human auditory system has a critical bandwidth that grows from approximately 0.1 kHz to 1.3 kHz [9]. In mobile terminals with limited computational resources it is beneficial to use a low frequency resolution when applying a noise reduction method.

In section 2, the filter bank structures are addressed. Section 3 deals with the design criteria and section 4 describes the speech enhancement algorithm, which is used in the evaluation simulations. In Section 5, the simulation results are presented and section 6 concludes the paper. The following notations are used in the paper: $(\cdot)^*$ denotes complex conjugate, $(\cdot)^T$ denotes transpose and $(\cdot)^H$ denotes conjugate transpose.

2. FILTER BANK STRUCTURES

2.1. Analysis Filter Bank

The analysis filter bank transforms a signal, $X(z)$, into subband signals, $X_m(z)$, using analysis filters, $H_m(z)$ and D_m-fold decimators, according to $X_m(z) = [X(z)H_m(z)]_{\downarrow D_m}$. In this paper a polyphase implementation is considered, see Fig. 1. The structure consists of a chain of M allpass functions $Q(z)$ and polyphase components E_l defined as

$$E_l(z) = \sum_{n=0}^{N-1} e_l(n) z^{-n}, \quad l = 0, \dots, M-1, \tag{1}$$

where the polyphase component coefficients are denoted by $e_l(n)$, $n = 0, \dots, N-1$. Filter bank structures with allpass functions are previously presented in [10]. The polyphase signals are input to a DFT matrix operation $\mathbf{W}_M^*$, with matrix entries $[\mathbf{W}_M]_{m,l} = W_M^{ml}$, where $W_M = \exp(-j2\pi/M)$. The undecimated subband signals $V_m(z)$ can be expressed in terms of the input signal $X(z)$ according to

$$V_m(z) = \sum_{l=0}^{M-1} E_l\left(Q^{-M}(z)\right) Q^l(z) W_M^{-ml} X(z) \tag{2}$$

Where $Q(z) = (-\mu + z^{-1})(1 - \mu z^{-1})^{-1}$ and μ is the uniformity coefficient. The analysis filters $H_m(z)$ are described by

$$H_m(z) = \sum_{l=0}^{M-1} E_l\left(Q^{-M}(z)\right) Q^l(z) W_M^{-ml} = \mathbf{e}^T \boldsymbol{\phi}_m(z). \tag{3}$$

In the vector notation for $H_m(z)$ in Eq. (3), a concatenated coefficient vector is used, which is defined as $\mathbf{e} = \left[\mathbf{e}_0^T, \dots, \mathbf{e}_{M-1}^T\right]^T$, where $\mathbf{e}_l = [e_l(0), \dots, e_l(N-1)]^T$ are the polyphase component coefficient vectors . The basis function vector $\boldsymbol{\phi}_m(z)$ in Eq.

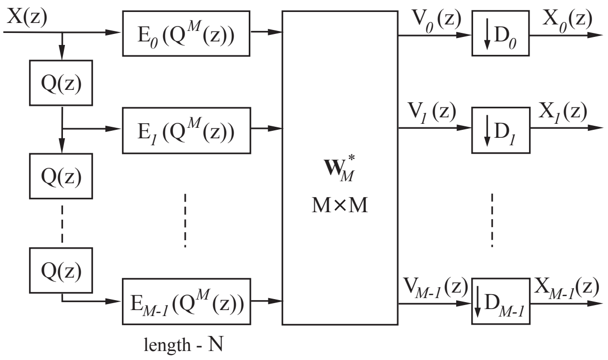

Fig. 1. *Polyphase analysis filter bank structure.*

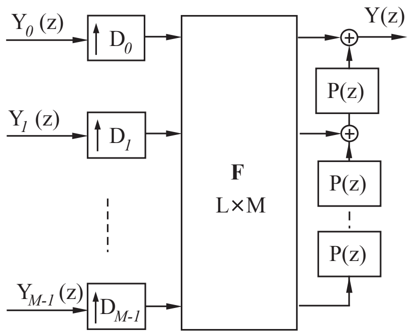

Fig. 2. *Direct-form synthesis filter bank structure.*

(3) is defined as

$$\boldsymbol{\phi}_m(z) = \left[\boldsymbol{\phi}_{m,0}^T(z), \ldots, \boldsymbol{\phi}_{m,M-1}^T(z)\right]^T, \tag{4}$$

where

$$\boldsymbol{\phi}_{m,l}(z) = Q^l(z) W_M^{-ml} \left[1, Q^M(z), \ldots, Q^{(N-1)M}(z)\right]^T. \tag{5}$$

The vector notation for $H_m(z)$ in Eq. (3) will be used for the problem formulation of the analysis filter bank design in Section 3.1.

2.2. Synthesis Filter Bank

The synthesis filter bank transforms the subband signals, $Y_m(z)$, into output signal, $Y(z)$, with synthesis filters, $G_m(z)$, according to $Y(z) = \sum_{\forall m} [Y_m(z)]_{\uparrow D_m} G_m(z)$. The synthesis filters $G_m(z)$ are implemented in direct-form according to

$$G_m(z) = \sum_{k=0}^{M-1} f_{m,k} P^k(z), \quad m = 0, \ldots, M-1. \tag{6}$$

This corresponds a matrix operation with $L \times M$ matrix $\mathbf{F}$, which has entries $[\mathbf{F}]_{k,m} = f_{m,k}$, and a chain of allpass functions $P(z) = z^{-1}$, see Fig. 2. The design of the synthesis filters intends to aim at the properties of the analysis and synthesis filter banks as a whole. These properties are described in the next section.

2.3. Analysis-Synthesis Transfer Functions

The synthesis filter bank output signal can be expressed in terms of the analysis filter bank input signal according to

$$Y(z) = T(z)X(z) + \sum_{m=0}^{M-1} \sum_{d=1}^{D_m-1} S_{m,d}(z) X(z W_{D_m}^d). \tag{7}$$

where, transfer function $T(z)$ affects the linear term in the output signal. The aliasing transfer functions $S_{m,d}(z)$ affect the nonlinear terms in the output signal, which are modulations of the input signal. The transfer function for the linear term is given by

$$T(z) = \sum_{m=0}^{M-1} \frac{1}{D_m} H_m(z) G_m(z) = \mathbf{f}^T \boldsymbol{\psi}(z). \tag{8}$$

The synthesis filter bank coefficient vector $\mathbf{f}$ in Eq. (8) is defined as $\mathbf{f} = \left[\mathbf{f}_0^T, \ldots, \mathbf{f}_{M-1}^T\right]^T$, with coefficient vectors $\mathbf{f}_m = [f_{m,0}, \ldots, f_{m,L-1}]^T$, so that $\mathbf{F} = [\mathbf{f}_0 \cdots \mathbf{f}_{M-1}]$. The basis function vector $\boldsymbol{\psi}(z)$, in Eq. (8) is given by

$$\boldsymbol{\psi}(z) = \left[\frac{1}{D_0} H_0(z) \boldsymbol{\psi}_0^T(z), \ldots, \frac{1}{D_{M-1}} H_{M-1}(z) \boldsymbol{\psi}_{M-1}^T(z)\right]^T, \tag{9}$$

where

$$\boldsymbol{\psi}_m(z) = \left[1, P^M(z), \ldots, P^{(L-1)M}(z)\right]^T. \tag{10}$$

The aliasing transfer functions $S_{m,d}(z)$ in Eq. (7) are defined by

$$S_{m,d}(z) = \frac{1}{D_m} H_m(z W_{D_m}^d) G_m(z) = \mathbf{f}_m^T \boldsymbol{\varphi}_{m,d}(z), \tag{11}$$

where

$$\boldsymbol{\varphi}_{m,d}(z) = \frac{1}{D_m} H_m(z W_{D_m}^d) \boldsymbol{\psi}_m(z). \tag{12}$$

3. DESIGN CRITERIA

3.1. Analysis Filter Bank Design Criterion

The proposed least squares frequency domain design criterion for the analysis filter bank is described in this section. The coefficients $\hat{\mathbf{e}}$ minimize the quadratic form

$$\sum_{m=0}^{M-1} \left\{ \sum_{\omega_i \in \Omega_{p,m}} \left| \mathbf{e}^T \boldsymbol{\phi}_m(e^{j\omega_i}) - \tilde{H}_m(e^{j\omega_i}) \right|^2 + \right. \\ \left. + \sum_{\omega_i \in \Omega_{s,m}} \left| \mathbf{e}^T \boldsymbol{\phi}_m(e^{j\omega_i}) \right|^2 \right\}, \tag{13}$$

where ω_i, $i = 1, \ldots, I$, is a set of frequency grid points. Passbands $\Omega_{p,m}$ are defined for analysis filters $H_m(z)$ in the vicinity of the subband center frequencies $\rho(2\pi m/M)$ according to

$$\Omega_{p,m} = \left[\rho\left(\frac{2\pi m - \delta\pi}{M}\right), \rho\left(\frac{2\pi m + \delta\pi}{M}\right)\right], \tag{14}$$

where $0 < \delta \leq 1$ is an additional passband width parameter, and $\rho(\omega)$ is a μ-dependent frequency warping function

$$\rho(\omega) = \omega - 2\arctan\left(\frac{\mu \sin(\omega)}{\mu \cos(\omega) + 1}\right). \tag{15}$$

Stopbands $\Omega_{s,m}$ are defined for analysis filters $H_m(z)$, which depend on the decimation rate D_m, according to

$$\Omega_{s,m} = \left[-\pi, \rho\left(\frac{2\pi m}{M} - \frac{\pi}{D_m}\right)\right] \cup \left[\rho\left(\frac{2\pi m}{M} + \frac{\pi}{D_m}\right), \pi\right]. \tag{16}$$

The desired analysis filter frequency response $\tilde{H}_m(e^{j\omega_i})$ is defined on the frequency grid ω_i as

$$\tilde{H}_m(e^{j\omega_i}) = \begin{cases} \exp(-j\omega_i\tau_H) & \text{for } \omega_i \in \Omega_{p,m} \\ 0 & \text{for } \omega_i \in \Omega_{s,m} \\ \text{undefined} & \text{for } \omega_i \notin \{\Omega_{p,m} \cup \Omega_{s,m}\} \end{cases} \tag{17}$$

where τ_H is the desired group-delay. The real-valued solution, which minimizes Eq. (13), is given by the normal equations

$$(\mathbf{A} + \mathbf{B})\,\hat{\mathbf{e}} = \mathbf{c} \tag{18}$$

where matrices $\mathbf{A}$ and $\mathbf{B}$ of size $MN \times MN$ are defined as

$$\mathbf{A} = \sum_{\omega_i \in \Omega_{p,m}} \sum_{m=0}^{M-1} \mathrm{Re}\left\{\boldsymbol{\phi}_m^*(e^{j\omega_i})\boldsymbol{\phi}_m^T(e^{j\omega_i})\right\}, \tag{19}$$

$$\mathbf{B} = \sum_{\omega_i \in \Omega_{s,m}} \sum_{m=0}^{M-1} \mathrm{Re}\left\{\boldsymbol{\phi}_m^*(e^{j\omega_i})\boldsymbol{\phi}_m^T(e^{j\omega_i})\right\}, \tag{20}$$

and vector $\mathbf{c}$ of size $MN \times 1$ is defined as

$$\mathbf{c} = \sum_{\omega_i \in \Omega_{p,m}} \sum_{m=0}^{M-1} \mathrm{Re}\left\{\boldsymbol{\phi}_m^*(e^{j\omega_i})\tilde{H}(e^{j\omega_i})\right\}. \tag{21}$$

3.2. Synthesis Filter Bank Design Criterion

The proposed least squares design criterion for the synthesis filter bank design is described in this section. The coefficients $\hat{\mathbf{f}}$ minimize the quadratic form

$$\sum_{\omega_i \in \Omega} \left\{\left|\mathbf{f}^T\boldsymbol{\psi}(e^{j\omega_i}) - \tilde{T}(e^{j\omega_i})\right|^2 + \right. \tag{22}$$

$$\left. + \sum_{m=0}^{M-1}\sum_{d=1}^{D_m-1}\left|\mathbf{f}_m^T\boldsymbol{\varphi}_{m,d}(e^{j\omega_i})\right|^2\right\}, \tag{23}$$

where the frequency grid points ω_i cover the total frequency range, i.e. $\Omega = [-\pi, \pi]$, and the desired response is a delay $\tilde{T}(e^{j\omega}) = \exp(-j\omega\tau_T)$. The solution is given by the normal equations

$$(\mathbf{E} + \mathbf{P})\,\hat{\mathbf{f}} = \mathbf{q} \tag{24}$$

where the matrix $\mathbf{E}$ and vector $\mathbf{f}$, of size $ML \times ML$ and $ML \times 1$ respectively, are defined as

$$\mathbf{E} = \sum_{\omega_i \in \Omega} \boldsymbol{\psi}^*(e^{j\omega_i})\boldsymbol{\psi}^T(e^{j\omega_i}), \tag{25}$$

$$\mathbf{q} = \sum_{\omega_i \in \Omega} \boldsymbol{\psi}^*(e^{j\omega_i})\tilde{T}(e^{j\omega_i}), \tag{26}$$

and matrix $\mathbf{P}$ of size $MN \times MN$ is defined as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_0 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{P}_{M-1} \end{pmatrix}, \tag{27}$$

where

$$\mathbf{P}_m = \sum_{d=1}^{D_m-1} \sum_{\omega_i \in \Omega} \boldsymbol{\varphi}_{m,d}^*(e^{j\omega_i})\boldsymbol{\varphi}_{m,d}^T(e^{j\omega_i}). \tag{28}$$

4. SUBBAND SPECTRAL SUBTRACTION

Spectral subtraction is a method to reduce background noise in noisy speech signals using an adaptive frequency gain function [11]. In subband spectral subtraction an adaptive gain is applied in each subband. The posed signal model which is used in subband spectral subtraction is given by

$$x_m(n) = s_m(n) + w_m(n), \quad m = 0, \ldots, M-1, \tag{29}$$

where $x_m(n)$ is the subband noisy signal, $s_m(n)$ is the subband clean speech signal, and $w_m(n)$ is the subband additive noise. Signals $s_m(n)$ and $w_m(n)$ are assumed uncorrelated. The background noise signal power is estimated during speech pauses by exponential averaging [11],

$$\overline{w}_m(n) = \alpha_m\overline{w}_m(n-1) + (1-\alpha_m)|x_m(n)|^a \tag{30}$$

where α_m is a decimation factor, D_m-dependent time constant and a is the norm. Subband dependent gains are calculated according to

$$u_m(n) = \left(1 - b\frac{\overline{w}_m(n)}{|x_m(n)|^a}\right)^{\frac{1}{a}}, \quad m = 0, \ldots, M-1, \tag{31}$$

where b is the subtraction factor [11]. The gains $u_m(n)$ are limited between the noise floor γ and 1 giving $u'_m(n)$.

$$u'_m(n) = \begin{cases} \gamma & \text{for} \quad u_m(n) < \gamma \\ u_m(n) & \text{for} \quad \gamma \le u_m(n) \le 1 \\ 1 & \text{for} \quad u_m(n) > 1 \end{cases} \tag{32}$$

The variability of the gains is further reduced by exponential averaging,

$$\overline{u}_m(n) = \beta_m\overline{u}_m(n-1) + (1-\beta_m)u'_m(n), \tag{33}$$

in order to avoid the musical tones phenomenon. Parameter β_m is decimation factor, D_m-dependent. When the gains function is applied to the noisy input signal, an estimate of the clean speech signal is obtained [11],

$$y_m(n) = \overline{u}_m(n) \cdot x_m(n), \quad m = 0, \ldots, M-1. \tag{34}$$

5. SIMULATION RESULTS

Filter banks with $M = 32$ subbands are designed for uniformity coefficients, μ, ranging $0 \le \mu \le \frac{1}{2}$. Other filter bank parameter settings are $N = 2$, $L = 32$, $\tau_H = (MN-1)/2$, and $\tau_T = MN + L - 1$. The decimation factors are ranging $2 \le D_m \le \frac{M}{4}$ depending on the bandwidth according to μ, where $D_m = \frac{M}{4}, \forall m$ in the uniform case for $\mu = 0$.

Noisy speech has been simulated from a database of 10 male speakers, 10 female speakers, and 3 noise sources, with appropriate SNR levels. The noise sources are compartment noise from a car, babble noise from a restaurant and engine noise from a factory. The sampling frequency is 16 kHz.

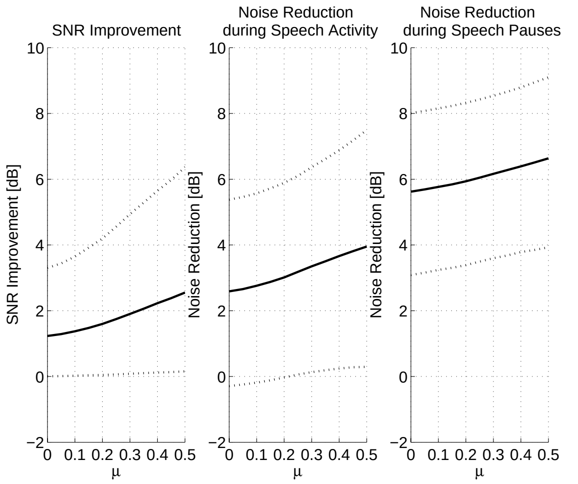

Fig. 3. *Distribution of the SNR improvement and noise reduction as a function of the uniformity coefficient. The whole line represents the mean and the dotted lines represent the 10%-boundaries of the distribution.*

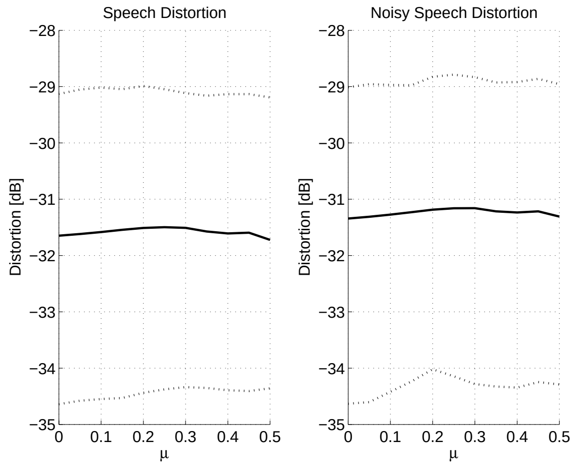

Fig. 4. *Distribution of the speech and noisy speech spectral distortion as a function of the uniformity coefficient. The whole line represents the mean and the dotted lines represent the 10%-boundaries of the distribution.*

The simulated noisy speech has been enhanced using subband spectral subtraction, with parameter settings $a = 1$, $b = 0.8$, $\gamma = 0.1$, and $\alpha_m = \beta_m = \exp(\log(0.999)D_m)$. Segmental SNR [dB] is measured by

$$\Delta\text{SNR} = 10\log_{10}\frac{\int R_{y_s}(\omega)d\omega}{\int R_{y_w}(\omega)d\omega} - 10\log_{10}\frac{\int R_{x_s}(\omega)d\omega}{\int R_{x_w}(\omega)d\omega}, \quad (35)$$

where $R_{(\cdot)}$ denotes short-time spectrum and x_s, x_w, y_s and y_w denote input speech, input noise, output speech and output noise components, respectively. Segmental noise reduction [dB] is measured by

$$\text{NR} = 10\log_{10}\frac{\int R_{x_w}(\omega)d\omega}{\int R_{y_w}(\omega)d\omega}. \quad (36)$$

Fig. 3 shows the distribution of the segmental SNR improvement and segmental noise reduction for a number of approx. 13000 noisy speech and 27000 noise-only segments. The segment duration was set to 20 ms.

Spectral distortion per utterance, of speech and noisy speech, is measured by

$$\begin{aligned} \text{DIST}_{\text{Speech}} &= \int|P_{x_s}(\omega) - P_{y_s}(\omega)|d\omega, \\ \text{DIST}_{\text{Noisy Speech}} &= \int|P_{x_s+k\cdot x_w}(\omega) - P_y(\omega)|d\omega \quad (37) \end{aligned}$$

where $P_{(\cdot)}$ are periodogram average estimates over one utterance and $k = \int P_{y_w}(\omega)d\omega / \int P_{x_w}(\omega)d\omega$ is the average noise reduction. Fig. 4 shows the distortion distributions for speech and noisy speech.

6. CONCLUSIONS

In this paper a method is proposed for the design of nonuniform filter banks, using frequency domain least squares criteria, for subband processing applications, e.g. subband adaptive filtering. The design objective of the proposed method is to minimize the magnitude of all aliasing components individually, such that aliasing distortion is minimized although phase alterations occur in the subbands. The filter banks have been applied and evaluated in speech enhancement using a subband spectral subtraction technique. The evaluation results show that filter banks designed using the proposed method can give higher noise reduction and SNR improvement while maintaining the speech quality.

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] J. M. de Haan, N. Grbić, I. Claesson, and S. Nordholm, “Filter Bank Design for Subband Adaptive Microphone Arrays,” *Accepted for publication in IEEE Transactions on Speech and Audio Processing*, Jul. 2002.

[4] 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,” pp. 5–19, 1998.

[5] M. S. Spurbeck and C. T. Mullis, “Least Squares Approximation of Perfect Reconstruction Filter Banks,” *IEEE Transactions on Signal Processing*, vol. 46, pp. 968–978, 1998.

[6] D. B. H. Tay, “Two Stage, Least Squares Design of Biorthogonal Filter Banks,” in *Proceedings of ISCAS*, 2000, vol. I, pp. 591–594.

[7] J. M. de Haan, N. Grbić, I. Claesson, and S. Nordholm, “Design and Evaluation of Nonuniform DFT Filter Banks in Subband Microphone Arrays,” in *Proceedings of ICASSP*, May 2002, pp. 1173 –1176.

[8] “TSG-SA; speech codec speech procesing functions; AMR wideband speech codec; transcoding functions, 3G TS 26.190 ver 5.1.0,” Tech. Rep., 3GPP, 2001.

[9] E. Zwicker and H. Fastl, *Psycho-acoustics, facts and models*, Springer-Verlag, 1999.

[10] 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.

[11] H. Gustafsson, S. Nordholm, and I. Claesson, “Spectral Subtraction using Reduced Delay Convolution and Adaptive Averaging,” *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 799–807, Nov. 2001.