OPTIMIZATION OF SWITCHABLE WINDOWS
FOR LOW-DELAY SPECTRAL ANALYSIS-SYNTHESIS

Dirk Mauler and Rainer Martin

Institute of Communication Acoustics, Ruhr-Universität Bochum, Germany
{Dirk.Mauler, Rainer.Martin}@rub.de

ABSTRACT

We present a novel iterative method for the optimization of switchable pairs of window functions. These windows may be used for block-based spectral analysis-synthesis (AS) in low-delay speech enhancement systems, where the energy compaction of speech sounds is improved by switching the spectral AS windows. Optimization objectives of the approach take the frequency response, quasi perfect reconstruction (PR) of each window pair and quasi-PR during window switching into account. An example of window pairs obtained with the proposed method clearly outperforms a reference design. The improved aliasing and imaging suppression is particularly important for hearing aids where high spectral gains may lead to audible reconstruction artifacts.

Index Terms— spectral analysis-synthesis, perfect reconstruction, switchable windows, low delay, hearing instruments.

1. INTRODUCTION

Speech enhancement is frequently performed in the spectral domain by means of a short-time spectral analysis and synthesis system. Besides filterbanks, systems based on discrete Fourier transform (DFT) are often used, since they provide good signal decorrelation properties and allow efficient implementations. In these systems the temporal and spectral properties of the analysis process is controlled by weighting the time domain data with a tapered analysis window. After the DFT the spectral coefficients might then be modified according to a speech enhancement algorithm. Then, the time domain signal is reconstructed via an overlap-add procedure by weighting the inverse transform of the modified spectrum with a synthesis window. To reduce the computational complexity the spectral data is frequently processed in a subsampled domain with a block advance of about 1/2 or 1/4 of the block length. Therefore, imaging artifacts that arise during upsampling to the original rate need to be attenuated by the synthesis window. This is an important factor in particular in the context of hearing aid applications where aliasing and imaging artifacts are amplified by very high gains, which need to be applied to spectral bands to compensate for the individual hearing loss.

The block length of the spectral analysis has to strike a balance between spectral and temporal resolution: A short block of data forbids a high spectral resolution but conserves the energy compaction of short non-stationary sounds. A long block lacks temporal resolution but allows for a high spectral resolution. Since many natural signals contain both, non-stationary and short-term stationary sections (e.g., speech: plosives and vowels), a signal adaptive spectral analysis is necessary to obtain improved temporal-spectral representations [1]. A signal adaptive switching of an analysis window has been proposed already in [2] and is extensively used in audio coding applications [3]. However, those windows are not suited for use in low-delay applications, e.g. hearing aids, where the overall delay of the spectral analysis-synthesis system must be small, e.g., below 10 ms, and must not vary when the analysis window is switched. While in the last years a couple of approaches for the optimal numerical design of analysis and synthesis filters have been proposed (e.g [4], [5], [6]), there has been less attempts in finding sets of optimal switchable windows for spectral analysis and synthesis. In [7] necessary conditions for perfect reconstruction are derived for a system with biorthogonal lapped transform using switchable spectral analysis and synthesis windows. However, this work does not discuss important constraints such as small group delay or the frequency response of the windows.

In [8] an example for a pair of switchable spectral analysis and synthesis windows has been introduced that maintains the same small delay irrespective of the pair of windows under use. The windows were designed in an intuitive manner using sections of square-root Hann windows. While PR was ensured, other important parameters of the window design, e.g. stopband frequency or stopband attenuation could not be varied. In this contribution we propose a systematic approach for the design of switchable window pairs for low-delay spectral analysis and synthesis. As opposed to [8] we relax the constraint of PR, requiring now only quasi-PR. From a perceptual point of view, quasi-PR is sufficient in many applications and the additional degrees of freedom enable improved overall optimization results. In the next section we introduce the notation and give an example for a pair of switchable windows. We then present our optimization approach in Section 3. An example of an improved set of windows generated with the new approach is given in Section 4.

2. SWITCHABLE WINDOWS FOR SPECTRAL ANALYSIS AND SYNTHESIS

While in principle an arbitrary number of switchable analysis-synthesis window pairs can be useful, we will restrict the description of our design method to a set of two pairs of analysis and synthesis windows, denoted as \( h_1 \), \( h_2 \) and \( h_3 \), \( h_4 \):

\[
\text{analysis win. I: } h_1 = [h_1(0), h_1(1), \ldots, h_1(M-1)]^T \quad (1)
\]

\[
\text{synthesis win. I: } h_2 = [h_2(0), h_2(1), \ldots, h_2(M-1)]^T \quad (2)
\]

\[
\text{analysis win. II: } h_3 = [h_3(0), h_3(1), \ldots, h_3(M-1)]^T \quad (3)
\]

\[
\text{synthesis win. II: } h_4 = [h_4(0), h_4(1), \ldots, h_4(M-1)]^T. \quad (4)
\]
Although the nominal length \( M \) is the same for each of the four windows, the effective lengths differ as filter coefficients at the end (or at the beginning) can be set to zero. Specifically, window pair II shall have an effective support of \( N < M \) and is thus suited to process sounds of short duration, e.g., vowels. The limited spectral resolution of the analysis window in pair II is not critical as the spectrum of these sounds normally lacks a detailed structure. The window pair I is required to have an analysis window with an effective support \( P \) close or equal to \( M \) in order to achieve a high spectral resolution. This is desirable for the analysis of short-term stationary sounds, e.g., vowels. The associated synthesis window, however, shall be of the same short (effective) length \( N \) as for the window pair II. This constraint is necessary to allow a low-delay overlap-add synthesis regardless which pair of windows is currently in use. Note, that the DFT length \( M \) is kept fixed for all window pairs. Thus, the seamless computation of spectral quantities for speech enhancement in between the analysis and the synthesis stage is greatly facilitated.

We further define the total effective window as the elementwise product of the coefficients of analysis and synthesis windows,

\[
\mathbf{g}_{1,2} = \left[ h_1(0)h_2(0), \ldots, h_1(M-1)h_2(M-1) \right]^T
\]

\[
\mathbf{g}_{3,4} = \left[ h_3(0)h_4(0), \ldots, h_3(M-1)h_4(M-1) \right]^T.
\]

When no spectral modifications are performed and PR is required the total effective windows must overlap and add to the constant value of one. The thin black lines in Figure 1 and the dashed black lines in Figure 2 show the example of switchable windows given in [8]. Here, we have \( M = 512, N = 128, P = 448 \) and a block shift of \( R = 32 \) samples. At the sampling rate of \( f_s = 16 \) kHz the delay of the analysis-synthesis system is 10 ms.

The frequency response \( H(e^{j\Omega}) \) of a filter is defined via the Fourier transform of the filter coefficients \( h \)

\[
H(e^{j\Omega}) = \mathbf{c}^T(e^{j\Omega})\mathbf{h},
\]

\[
\mathbf{c}(e^{j\Omega}) = [1, e^{-j\Omega}, e^{-2j\Omega}, \ldots, e^{-(M-1)j\Omega}]^T.
\]

### 3. Optimization of Switchable Windows

In this section we present a novel approach for the iterative design of switchable pairs of spectral analysis and synthesis windows. Some of the optimization criteria is based on [9] where a multi-objective optimization of a single FIR lowpass filter is presented.

#### 3.1. Optimization objectives

The optimization of the windows is performed subject to the objectives which are introduced in the following paragraphs. We use the subscript \( i \) to refer to any of the four windows, i.e. \( i \in \{1, 2, 3, 4\} \). The desired magnitude response of a window is defined as:

\[
H_{desi}(e^{j\Omega}) = \begin{cases} 
1, & 0 \leq \Omega \leq \Omega_{pi} \\
\alpha_i, & \Omega_{pi} < \Omega \leq \pi.
\end{cases}
\]

The parameter \( \alpha_i \) denotes the desired attenuation in the stopband, \( \Omega_{pi} \) the normalized passband frequency and \( \Omega_s \) the normalized stopband frequency for filter \( h_i \). Furthermore, for implementation the continuous normalized frequency \( \Omega \) is replaced by a sequence of \( K \) discrete frequency bins \( \Omega_k = 2\pi k/K, k = 0, 1, \ldots, K-1 \), such that \( K_{\Omega_i} \) represents the index of the discrete stopband frequency bin.

#### 3.1.1. Minimizing peak sidelobe in the stopband

Deviations of the desired and the actual filter response in the stopband are measured in the \( r \)-norm

\[
e_{pti}(\mathbf{h}_i) = \left[ \sum_{k=K_{\Omega_i}}^{K_{\Omega_i+1}} \left| |H(e^{j\Omega_k})| - |H(e^{j\Omega_k})| \right|^r \right]^{1/r}.
\]

An equiripple design is approximated by \( r = 20 \).

#### 3.1.2. Minimizing the stopband energy

While the preceding objective aims at a minimization of the peak sidelobe amplitude one can also aim at minimizing the total sidelobe energy [10]. The sidelobe energy is measured by

\[
e_{en}(\mathbf{h}_i) = \sum_{k=K_{\Omega_i}}^{K_{\Omega_i+1}} \left| |H(e^{j\Omega_k})| \right|^r.
\]

#### 3.1.3. Quasi-PR within a window pair

Let us temporarily assume that only window pair I is under use. In the overlap-add method [11] the output signal

\[
\mathbf{g}_{ola_{1,2}} = [g_{ola_{1,2}}(0), g_{ola_{1,2}}(1), \ldots, g_{ola_{1,2}}(R-1)]^T
\]

is generated after every block shift by summation of \( M/R \) shifted sections of \( g_{1,2} \), each of length \( R \) samples:

\[
\mathbf{g}_{ola_{1,2}} = \begin{bmatrix} I_R | I_R | \cdots | I_R \end{bmatrix} \mathbf{g}_{1,2},
\]

where \( I_R \) is the unity matrix of dimension \( R \times R \). If the vector \( g_{ola_{1,2}} \) contains only entries of the same constant value, the analysis-synthesis system is transparent (apart from a delay and possibly an overall amplification or attenuation which can be easily compensated). Any deviation from this constant is noticed as a ripple on the output signal and should be minimized. As a third objective we therefore demand quasi-PR. Based on the above considerations the maximum absolute ripple of window set I is monitored by

\[
e_{pr_{1,2}}(\mathbf{h}_1, \mathbf{h}_2) = \max_{n=1, \ldots, R-1} \left( \frac{g_{ola_{1,2}}(n)}{1/R \sum_{\eta=0}^{R-1} g_{ola_{1,2}}(\eta)} - 1 \right).
\]

Assuming now that only window pair II is used the corresponding error \( e_{pr_{3,4}}(\mathbf{h}_1, \mathbf{h}_4) \) can be defined in the same fashion.

#### 3.1.4. Quasi-PR when window pairs are switched

The measures defined in the preceding section quantify the deviations from PR when either one of the two analysis-synthesis window pairs is used. When the window pair is switched (\( I \rightarrow II \) or \( II \rightarrow I \)), however, PR is not guaranteed. If PR is still to be achieved, the analysis-synthesis products of each pair have to be equal, i.e. \( g_{1,2} = g_{3,4} \). Our fourth objective therefore is to establish quasi-PR during switching of the window pairs by minimizing the maximum distance between the total effective windows:

\[
e_{su_{1,2}}(\mathbf{h}_1) = \| g_{1,2} - g_{3,4} \|_{\infty}.
\]

The normalization in (14) is necessary to make the error independent from the window amplitude. By exchanging the indices \( \{1, 2\} \leftrightarrow \{3, 4\} \) the corresponding error \( e_{su_{3,4}}(\mathbf{h}_1) \) can be defined.
3.1.5. Minimizing the group delay

Given the group delay response of \( H(e^{j\Omega}) \)

\[
\tau_g(e^{j\Omega}, h_i) = \frac{-\partial \text{arg}(e^{j\Omega}h_i)}{\partial \Omega} = \text{Re} \left\{ \frac{e^T(e^{j\Omega})Dh_i}{e^T(e^{j\Omega})h_i} \right\}
\]

(15)

with the \( M \times M \) diagonal matrix \( D = \text{diag}\{0, 1, \ldots, M - 1\} \) and \( \text{Re}\{\} \) the real part of a complex-valued number, its maximum absolute value shall be minimized in the band \( \Omega \in [0, \Omega_{sp}] \):

\[
e_r(h_i) = \max_{0 \leq \Omega_h < \Omega_{sp}} \left| \tau_g(e^{j\Omega}, h_i) \right|
\]

(16)

3.2. Single objective formulation

For each window function \( h_i \), the error measures defined in the preceding paragraphs are aggregated in a vector

\[
e(h_i) = [e_{sl}(h_i), e_{es}(h_i), e_{wsl}(h_i), e_{wes}(h_i)]^T
\]

(17)

with the index \( \{j,k\} = \{1, 2\} \) for \( i = 1 \) or \( i = 2 \) and \( \{j,k\} = \{3, 4\} \) for \( i = 3 \) or \( i = 4 \). Using a weighting vector \( w_i \) of dimension \( 5 \times 1 \) a weighted scalar objective function is computed

\[
e(h_i) = w_i^T e(h_i) .
\]

(18)

This function is subject to minimization which can be performed e.g. by a gradient based procedure as implemented in function \texttt{fminunc} of the MATLAB optimization toolbox.

3.3. Iterative optimization approach

Our goal is the joint optimization of the four windows, respecting the dependencies between the windows that have been defined in 3.1.3 and 3.1.4. To satisfy these dependencies we propose a round-robin optimization approach where the errors \( e(h_i) \) of each window \( i \) are minimized while the other three windows are kept fixed: After the initialization of all windows we minimize the weighted error \( e(h_i) = w_i^T e(h_i) \) of the \( i \)th window, \( i \in \{1, 2, 3, 4\} \), by performing \( V \) iterations of the gradient based solver. We then switch to the optimization of the next window until the optimization has achieved the desired window properties.

Depending on the initial window sets and the weights \( w_i \), the multidimensional optimization procedure may converge to a local minimum of the error function which might not satisfy the needs. Good results, however, are obtained when initializing the algorithm with the window set [8]. We also observed that after convergence, setting the weights \( w_i(1) = 0, i \in \{1, 2, 3, 4\} \), and continuing with optimization results in much better stopband attenuation at the cost of only a slight degradation of the desired stopband frequencies.

4. RESULTS

Figure 1 shows a design example for \( M = 512, N = 128, P = 448, R = 32, V = 10, r = 20, a_i = 0, \Omega_{sp} = 1/R \) for \( i \in \{1, 2, 3, 4\} \) and for the settings given in Table 1. The delay of the analysis-synthesis system is 10 ms. The weight \( w_1(5) = 0.01 \) emphasizes group delay minimization of the long analysis window \( h_1 \) and ensures the asymmetry of the window. Since the purpose of the synthesis windows is to attenuate imaging, the desired stopband frequency is set to the inverse of the blockshift (1/R). The same setting is used for the analysis window \( h_3 \). The window does not need to provide a high spectral resolution as this window is to be used for spectral analysis of burst-like sounds that do not convey many spectral details.

The time domain ripple is less than 0.017 dB, thus quasi-PR is achieved. Compared to the reference design [8] the frequency response of the optimized long analysis window \( h_1 \) has been slightly improved for frequencies above \( \Omega/\pi = 1/R \) while the performance for lower frequencies did not deteriorate (Fig. 1 (a)). The stopband attenuation of the corresponding synthesis window has been considerably improved (up to 20 dB, Fig. 1 (b)). The frequency responses of the analysis and synthesis windows of window pair II are virtually the same (Fig. 1 (c)). Compared to the reference they show a wider passband. The stopband attenuation has been improved by up to 12 dB.

In particular when applying large spectral gains, as necessary in hearing aid applications, aliasing and imaging artifacts are much better attenuated with the new design. This is also confirmed by informal listening tests for both window pairs (I,II). In conjunction with a Wiener filter noise reduction system a 40 dB amplification of a band of 1 kHz width led to a ringing noise when the reference windows [8] were used while the artifact did not appear when the optimized windows were used.

The results obtained with the current approach depend significantly on the parameter settings (e.g. weights \( w_i \)). Suboptimal results can be observed. This property can be circumvented by running several optimizations with different initial values and selecting the best solution [9].

5. CONCLUSIONS

We presented a novel iterative method for the optimization of a set of windows for low-delay block-based spectral analysis and synthesis. An important objective of the optimization is to make the resulting pairs of analysis-synthesis windows switchable without harming the quasi-PR property of the system during window switching. Compared to an existing solution of low-delay switchable pairs of analysis-synthesis windows we showed that the proposed method achieves windows with significantly improved spectral properties. The stopband attenuation of the synthesis windows in the example has been improved by 12 - 20 dB.

6. ACKNOWLEDGMENT

We thank T. Kurbiel, D. Alfsmann, and H. Göckler for inspiring discussions on the subject.

7. REFERENCES

Fig. 1. Design example - frequency responses. The thin black lines refer to the reference windows in [8] that were used here as initialization for the iterative optimization procedure. The plot in the upper right corner is a zoom into the respective frequency responses.

Fig. 2. Design example - time domain windows. The dashed black lines refer to the reference windows in [8] that were used here as initialization for the iterative optimization procedure. The windows are plotted normalized to their maximum amplitude.


