Adaptive Time-Frequency Analysis for Noise Reduction in an Audio Filter Bank With Low Delay

Kristian Timm Andersen, *Student Member, IEEE,* and Marc Moonen, *Fellow, IEEE,*

***Abstract*—In this paper an adaptive time-frequency analysis scheme is proposed along with a synthesis scheme using an asymmetric window. The proposed scheme is suitable for audio noise reduction with a low delay in the range of 0 to 4 ms. The adaptive analysis scheme can adapt to the incoming signal independently in both time and frequency by employing a complex filter on a DFT modulated filter bank. A number of adaptive time-frequency schemes are described that are suitable for low delay and low computational complexity implementations. The synthesis scheme uses an asymmetric window to achieve a good tradeoff between low delay and a sharp frequency response. Examples are given of the adaptive analysis and measurements of the synthesis scheme are given to show that the filter bank has a gain dependent non-linear phase response. Finally, a noise reduction task is performed that shows superior performance to a reference implementation in terms of segmental SNR and PESQ.**

***Index Terms*—Adaptive time-frequency analysis, noise reduction, speech enhancement, low delay.**

I. Introduction

ONE of the most important tasks for a real-time audio device is to present a clear and audible signal to the user at a low delay. For a hearing aid in particular, a low delay is critical since sound traveling through the vent into the ear canal should not get too out of sync with the sound coming from the hearing aid speaker, otherwise creating a comb-filter effect. Studies have shown that delays exceeding approximately 10ms can be objectionable while delays around 3-5ms can still be detected [1]. The quality of the signal can be improved using noise reduction, which in the single-channel case is done by applying a frequency-dependent gain to the signal. To achieve some tradeoff between time and frequency resolution in the noise reduction, the filter bank is often designed to have a nonuniform frequency resolution with more narrow bands in the low frequencies.

A popular framework for nonuniform frequency resolution is the wavelet transform [2], for instance the critically sampled tree-structured filter bank. For our application, however, the iterated use of the so-called "mother" wavelet results in a high group delay that makes it inappropriate for very low delay applications. Also, the need for noise reduction in the filter bank necessitates oversampling to avoid aliasing in the reconstruction [3], [4]. A frequency warped low delay filter bank that can approximate the Bark frequency scale has been proposed in [5] and [6].

A number of adaptive time-frequency (TF) resolution schemes have been suggested, see for instance [7], [8] and the references therein. Also, it has been shown that adapting the TF resolution can lead to improvements in noise reduction [9], [10]. Common for these approaches is that they are not suitable for low delay implementations since they require longer time windows to determine the TF resolution and/or that they have a high computational complexity. A window switching approach to adaptive TF resolution with a low delay of 10ms has been proposed in [11].

The approach in this paper differs from these methods in that an adaptive TF analysis that can adapt its resolution independently in both time and frequency is developed which is suitable for low delay implementation and has a low computational complexity. The adaptive TF analysis is incorporated into the framework of the DFT modulated filter bank and paired with a synthesis that is, again, suitable for low delay implementation.

The paper is structured as follows. In section II, the underlying basis function of the adaptive TF analysis is derived and it is shown that the adaptive TF analysis can be realized as a filtering on top of a DFT modulated filter bank. In section III, the estimation of the filter coefficients for the adaptive TF analysis is described. It is shown how the bandwidth of a frequency band can be calculated from the filter coefficients, which makes it possible to design a nonuniform analysis with any desired bandwidth for each individual frequency band. It is shown how the analysis filter bank can be time-varying and a number of different adaptive TF analysis schemes are described. An example is given to compare the different adaptive TF analysis schemes. In section IV, the synthesis using an asymmetric window is described. It is shown how the asymmetric window achieves a tradeoff between low delay and sufficient band attenuation. Further reduction in the delay is achieved by applying the frequency dependent gain for the noise reduction as a finite impulse response (FIR) filter, either by reusing the asymmetric window or by calculating the minimum phase response. The section ends with a discussion of the synthesis scheme with some experimental results. Section V contains the conclusion and future outlook.

II. Adaptive time-frequency analysis scheme

Assume an overcomplete set of L basis functions:

$$g_k[n] = h[n]W_L^{kn}, \quad 0 \leq k < L-1 \tag{1}$$

Manuscript submitted August 10, 2015.

This work was supported in part by The Danish Agency for Science, Technology, and Innovation and conducted in collaboration between KU Leuven and Widex A/S through the industrial Ph.D. program (case number 13-135472).

The authors are with the Department of Electrical Engineering (ESAT - STADIUS), KU Leuven, Leuven, Belgium (e-mail: kristian@esat.kuleuven.be; Marc.Moonen@esat.kuleuven.ne). K.T. Andersen is also with Widex A/S, Lynge, Denmark.

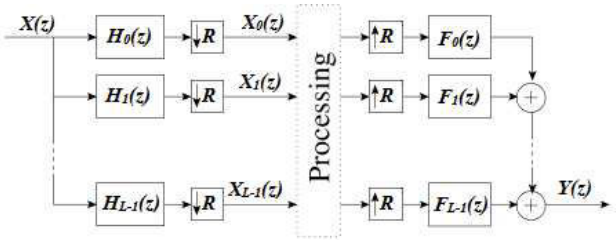

Fig. 1. Analysis/synthesis scheme.

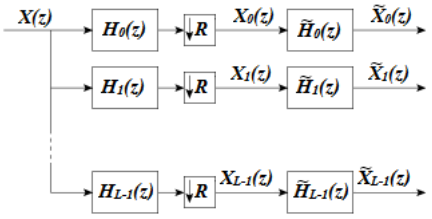

Fig. 2. Adaptive TF analysis scheme.

where $h[n]$ is a suitable real-valued window with a low-pass characteristic and with length N ($N < L$) that localizes the basis functions in time and $W_L = e^{2\pi j/L}$. A time-domain signal $x[n]$ is mapped to the transform domain by the sliding inner product decimated with a factor R ($R < N$):

$$x_k[i] = (g_k \star x)[iR] = \sum_{n=-\infty}^{\infty} g_k[n]^* x[n+iR] = \sum_{n=0}^{L-1} h[n]x[n+iR]e^{-2\pi jnk/L} \quad (2)$$

where $\star$ is the cross-correlation and * is the complex conjugate operator. Thus, the transformation is calculated with the discrete Fourier transform (DFT), $X_{k,i} = x_k[i]$ is the short-time Fourier transform (STFT) with hop-size R, N is the frame length and L is the number of frequency bands. Throughout this paper we will use the values $L = 512$, $N = 64$, $R = 16$ for examples, but it is possible to use other values depending on the application.

Equation (2) can be interpreted as the signal $x[n]$ being filtered with frequency modulated versions of the low-pass filter $h[-n]$, which has given rise to the term DFT modulated filter bank. This is efficiently implemented using the fast Fourier transform (FFT). A combined analysis/synthesis scheme representation is seen in Figure 1. The input signal $X(z)$ is first filtered by L band-pass filters $H_k(z)$ and then downsampled by R to give $X_k(z)$. Each subband signal $X_k(z)$ is then processed, which in a noise reduction application is where the gain is applied. The processed subband signals are then upsampled by R, filtered with a synthesis filter and finally summed. This can be written as:

$$y[m] = \sum_{i=-\infty}^{\infty} f[m-iR]\frac{1}{L}\sum_{k=0}^{L-1} Y_{k,i}e^{2\pi jk(m-iR)/L} \quad (3)$$

where $f[m]$ is a synthesis window and $Y_{k,i}$ is the signal to be reconstructed after processing. The analysis filters and synthesis filters correspond to frequency shifted versions of the low-pass analysis $h[n]$ and synthesis $f[n]$ windows respectively.

In this paper, since $N < L$, the transform domain representation has redundancy, which in a filter bank interpretation can be understood as a high degree of overlap between neighboring frequency bands. We improve the frequency resolution in frequency band k at time i by using a window $\tilde{h}_{k,i}[n]$ that is less localized in time, obtained from $h[n]$ by filtering with a hop-size R:

$$\tilde{h}_{k,i}[n] = \sum_{q=0}^{Q} b_{k,i,q}h[n+qR] + \sum_{p=1}^{P} a_{k,i,p}\tilde{h}_{k,i-p}[n+pR] \quad (4)$$

where $b_{k,i,q}$ and $a_{k,i,p}$ are filter coefficients that are variable in both frequency band k and time i. The improved basis functions are then:

$$\tilde{g}_{k,i}[n] = \tilde{h}_{k,i}[n]W_L^{nk}, \quad 0 \le k < L-1 \quad (5)$$

Depending on the filter coefficients $a_{k,i,p}$ and $b_{k,i,q}$, the improved basis functions $\tilde{g}_{k,i}[n]$ can have infinite length and can be chosen independently for each frequency band k. $\tilde{g}_{k,i}[n]$ is sampled with the same time- and frequency steps as $g_k[n]$ and has essentially traded some redundancy in frequency for some redundancy in time. This redundant filtering scheme is what allows the analysis to adapt its resolution independently for each frequency band while maintaining a low delay.

The improved STFT is given by:

$$\begin{aligned}\tilde{x}_k[i] &= \sum_{n=-\infty}^{\infty} \tilde{g}_{k,i}[n]^* x[n+iR] \\ &= \sum_{q=0}^{Q} b_{k,i,q} \sum_{n=-\infty}^{\infty} h[n+qR]x[n+iR]W_L^{-nk} \\ &\quad + \sum_{p=1}^{P} a_{k,i,p} \sum_{n=-\infty}^{\infty} \tilde{h}_{k,i-p}[n+pR]x[n+iR]W_L^{-nk} \\ &= \sum_{q=0}^{Q} b_{k,i,q}x_k[i-q]W_L^{qRk} + \sum_{p=1}^{P} a_{k,i,p}\tilde{x}_k[i-p]W_L^{pRk} \quad (6)\end{aligned}$$

It is seen that $\tilde{x}_k[i]$ can be obtained as a filtering of $x_k[i]$ without the explicit calculation of the underlying basis functions. However, (4) and (5) are useful for properly setting the coefficients of the adaptive TF analysis, as explained in the next section. The proposed adaptive TF analysis scheme can be seen in Figure 2.

III. Setting the Filter Coefficients for Adaptive Time-Frequency Analysis

The analysis presented in section II allows for the design of an individual analysis window in each frequency band using any filter design method as long as all the windows share the same underlying basis function $h(n)$. However, as we are interested in a low delay, low complexity analysis, we consider

the estimation of the filter coefficients using only one auto-regressive coefficient, i.e. $P = 1$, $Q = 0$. This leaves two coefficients to be estimated for each frequency band. To further simplify this, in section III-A we use a maximum a posteriori (MAP) approach to reduce the estimation to one coefficient for each frequency band. In section III-B we consider the estimation of the filter coefficients for a constant filter and in section III-C we consider a number of adaptive TF analysis schemes for a time-varying filter. Section III-D contains some experimental results obtained using the discussed adaptive TF analysis schemes.

A. Setting the filter coefficients using MAP

The $X_{k,i} = x_k[i]$ from (2) has an associated TF resolution that is determined by $h[n]$ and must therefore be viewed as only an estimate of the true TF distribution $\bar{X}_{k,i}$:

$$X_{k,i} = \bar{X}_{k,i} + N_{k,i} \tag{7}$$

where $N_{k,i}$ is noise that is assumed additive. If $N_{k,i}$ has a complex Gaussian distribution with zero mean, then $X_{k,i}$ is a maximum-likelihood (ML) estimate of $\bar{X}_{k,i}$.

The estimate can be improved by including a (Gaussian) prior on $\bar{X}_{k,i}$ given $\bar{X}_{k,i-1}$:

$$p(\bar{X}_{k,i}|\bar{X}_{k,i-1}) = \frac{1}{\pi\sigma_R^2} e^{\frac{-|\bar{X}_{k,i} - \bar{X}_{k,i-1}W_L^{Rk}|^2}{\sigma_R^2}} \tag{8}$$

where $\sigma_R^2 = \mathrm{E}[|\bar{X}_{k,i} - \bar{X}_{k,i-1}W_L^{Rk}|^2]$. This states that given $\bar{X}_{k,i-1}$, the most likely value of $\bar{X}_{k,i}$ is $\bar{X}_{k,i-1}$ phase shifted by $2\pi Rk/L$, which is the phase difference between two points on a sinusoid with frequency k/L radians per sample separated by R samples in time. The a posteriori probability then becomes:

$$\begin{aligned} &p(\bar{X}_{k,i}|X_{k,i}, \bar{X}_{k,i-1}) \\ &\quad = \frac{p(X_{k,i}|\bar{X}_{k,i}, \bar{X}_{k,i-1})p(\bar{X}_{k,i}|\bar{X}_{k,i-1})}{p(X_{k,i}|\bar{X}_{k,i-1})} \\ &\quad \propto p(X_{k,i}|\bar{X}_{k,i}, \bar{X}_{k,i-1})p(\bar{X}_{k,i}|\bar{X}_{k,i-1}) \end{aligned} \tag{9}$$

As the logarithm is a monotonic function, a MAP estimate can be found by minimizing the following cost function:

$$\begin{aligned} &C_{k,i} = -\log\left[p(X_{k,i}|\bar{X}_{k,i}, \bar{X}_{k,i-1})p(\bar{X}_{k,i}|\bar{X}_{k,i-1})\right] \\ &= \frac{|\bar{X}_{k,i} - X_{k,i}|^2}{\sigma_N^2} + \frac{|\bar{X}_{k,i} - \bar{X}_{k,i-1}W_L^{Rk}|^2}{\sigma_R^2} + \log\left[\pi^2\sigma_R^2\sigma_N^2\right] \end{aligned} \tag{10}$$

where $\sigma_N^2 = \mathrm{E}[|N_{k,i}|^2]$. This cost function is a convex quadratic function of $\bar{X}_{k,i}$ and can be minimized by setting the derivative equal to zero:

$$\frac{\partial C_{k,i}}{\partial \bar{X}_{k,i}} = 0 \tag{11}$$

which results in the MAP estimate:

$$\tilde{X}_{k,i} = \alpha X_{k,i} + (1-\alpha)\tilde{X}_{k,i-1}W_L^{Rk} \tag{12}$$

where $\alpha = \frac{\sigma_R^2}{\sigma_R^2 + \sigma_N^2}$ is a real-valued parameter between 0 and 1. By comparing to (6) it is seen that the derived equation corresponds to a filter with coefficients: $b_{k,i,0} = \alpha, a_{k,i,1} = 1 - \alpha$.

To further analyze this, we evaluate the discrete-time Fourier transform (DTFT) of $\tilde{h}_{k,i}[n]$ at $\omega = 0$:

$$\begin{aligned} \tilde{H}_{k,i}(0) &= \alpha H(0) + (1-\alpha)\tilde{H}_{k,i-1}(0)e^{jR0} \\ &= \tilde{H}_{k,i-1}(0) + \alpha\big(H(0) - \tilde{H}_{k,i-1}(0)\big) \end{aligned} \tag{13}$$

where

$$\tilde{H}_{k,i}(\omega) = \sum_{n=-\infty}^{\infty} \tilde{h}_{k,i}[n]e^{-j\omega n} \tag{14}$$

is the DTFT of $\tilde{h}_{k,i}[n]$. It is seen that for $\omega = 0$, $\tilde{H}_{k,i}(\omega)$ is a low-pass filtered version of $H(\omega)$, and since $h[n]$ is assumed constant, we get $\tilde{H}_{k,i}(0) = H(0)$ for $i \to \infty$. Since $\tilde{g}_{k,i}[n]$ is a frequency shifted version of $\tilde{h}_{k,i}[n]$, where $H(0)$ is shifted to the center frequency of the kth frequency band, α does not change the energy at the center frequency of each frequency band compared to $g_k[n]$. As this also holds if α is replaced with $\alpha_{k,i}$, i.e. parameters changing in time and frequency, the underlying basis functions can be interpreted as a time-varying analysis filter bank, where each frequency band has constant energy at its center frequency and $\alpha_{k,i}$ controls the bandwidth of frequency band k at time i. Note that, in the time-varying case, the auto-regressive nature of the filter means that the bandwidth also depends on previous values of $\alpha_{k,i}$.

B. Setting the filter coefficients for a constant filter

In this section we consider the calculation of the filter coefficients α_k for a constant filter of the form $b_{k,i,0} = \alpha_k, a_{k,i,1} = 1 - \alpha_k$, i.e. where the filter is time-invariant (constant) for every frequency band. To define a measure of bandwidth we use the distance between the 3dB cutoff points, that is we specify a desired bandwidth $\omega_k = \frac{\pi B_k}{f_s}$ for each frequency band, where B_k is the bandwidth and f_s is the sampling rate in Hz.

Evaluating the squared magnitude of the DTFT of the filter $\tilde{H}_k(\omega_k)$ using (4) for the specific setting of $a_{k,i,1}$ and $b_{k,i,0}$, we can write:

$$|\tilde{H}_k(\omega_k)|^2 = \frac{\alpha_k^2|H(\omega_k)|^2}{1 - 2(1-\alpha_k)\cos(\omega_k R) + (1-\alpha_k)^2} \tag{15}$$

To find α_k for a given bandwidth ω_k, we set $|\tilde{H}_k(\omega_k)|^2 = \frac{|H(0)|^2}{2}$ and rewrite the equation as a quadratic function of α_k:

$$\alpha_k^2 c_1 + \alpha_k c_2 - c_2 = 0 \tag{16}$$

where

$$c_1 = \frac{|H(\omega_k)|^2}{|H(0)|^2} - \frac{1}{2} \tag{17}$$

$$c_2 = 1 - \cos(\omega_k R) \tag{18}$$

The solution is:

$$\alpha_k = \frac{c_2}{c_1}\left(\sqrt{\frac{c_1}{c_2} + \frac{1}{4}} - \frac{1}{2}\right) \tag{19}$$

where the solution for negative α_k has been discarded, as α_k must be positive. $\frac{|H(\omega_k)|^2}{|H(0)|^2}$ is calculated from the DTFT of the

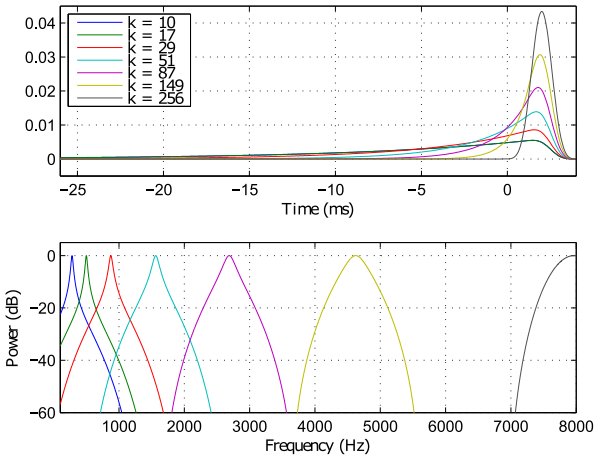

Fig. 3. 7 basis functions for a constant Q-transform with a Blackmann-Harris window ($N = 64$, $L = 512$, $R = 16$, $f_s = 16$kHz)). (Top) Improved analysis windows $\tilde{h}_k(n)$. (Bottom) Power spectrum of improved basis functions $\tilde{G}_k(\omega)$.

used window $h[n]$ and ω_k should be chosen small enough so that $\frac{|H(\omega_k)|^2}{|H(0)|^2} > \frac{1}{2}$. This is clear, as the new 3dB cutoff point must be smaller than that of $h[n]$.

Equation (19) can be used to design a filter bank with a desired bandwidth for each frequency band. As an example, a constant-Q filter bank can be designed by setting B_k to be proportional to k:

$$B_k = \max\left(\frac{2kB_H}{L}, \frac{f_s}{2L}\right) \tag{20}$$

where B_H is the 3dB bandwidth for $H(\omega)$. B_k is bounded downwards by $\frac{f_s}{2L}$ to ensure that the cutoff point of a frequency band is not smaller than half the distance to the neighbor band. An example of the basis functions for $h[n]$ equal to a Blackman-Harris window is seen in Figure 3. The frequency bands are linearly spaced on the frequency axis, but to avoid clutter in the plot only 7 logarithmically spaced basis functions are shown. It is seen that $\tilde{h}_k[n]$ is an asymmetric window with most of the energy concentrated at the most recent part of the signal and an exponential decay that depends on the parameter α_k. This asymmetry means that the analysis is mostly determined by the most recent part of the signal and makes the analysis more appropriate for low delay processing compared to symmetric basis functions that arise from for instance wavelets. It is seen that a smaller value of α_k in the lower frequencies sharpens the main lobe of $\tilde{G}_k(\omega)$. Further sharpening could be achieved by using a higher order filter, but then at the expense of the peak of $\tilde{h}_k[n]$ moving further backwards in time. The proposed TF analysis scheme is also significantly cheaper than processing the full filter in the time domain and the use of a single coefficient for each frequency band means that each frequency band only requires one complex and one real multiplication per time update to obtain a nonuniform TF resolution.

C. Setting the filter coefficients for a time-varying filter

In this section we consider the calculation of filter coefficients for a time-varying filter. More specifically, we develop three simple estimators for a time-varying TF analysis that is suitable for low delay and low complexity implementation. Many different adaptive TF analysis schemes have been suggested, but, to the best of our knowledge, no scheme exists that is suitable for both low delay and low complexity implementation and at the same time allows for full adaptability in both time and frequency.

As mentioned in section II, the derivation of section III-A can also be made for a time-varying filter, leading to:

$$\tilde{X}_{k,i} = \alpha_{k,i} X_{k,i} + (1 - \alpha_{k,i}) \tilde{X}_{k,i-1} W_L^{Rk} \tag{21}$$

Under the assumption that the signal can be decomposed into quasi-periodic signal components that can be separated in the combined TF domain, a simple optimal estimator can be derived.

In previous sections, it was shown that the energy of $\tilde{H}_{k,i}(0)$ is constant and equal to H(0) for $i \to \infty$ and any value of $\alpha_{k,i}$. Consequently, for a sinusoid with frequency $\frac{2\pi k}{L}$, the energy of $\tilde{X}_{k,i}$ is invariant to the choice of $\alpha_{k,i}$. Therefore any increase in energy, due to the change of $\alpha_{k,i}$, must come from signal components not centered on $\tilde{X}_{k,i}$. Consequently, an optimal estimate of $\alpha_{k,i}$ is the value $\hat{\alpha}_{k,i}$:

$$\hat{\alpha}_{k,i} = \arg\min_{\alpha_{k,i}} \left[|\tilde{X}_{k,i}|^2 \right] \tag{22}$$

subject to:

$$\alpha_{min} \leq \alpha_{k,i} \leq \alpha_{max} \tag{23}$$

where α_{min} and α_{max} are parameters that limits the TF resolution. A regularized version of this estimator is:

$$\hat{\alpha}_{k,i} = \arg\min_{\alpha_{k,i}} \left[|\tilde{X}_{k,i}|^2 + \lambda_k \alpha_{k,i} \right] \tag{24}$$

where λ_k ($\lambda_k > 0$) is a regularization parameter. The term $\lambda_k \alpha_{k,i}$ punishes large values of $\alpha_{k,i}$, which favors longer analysis windows for each basis function. This is justified since it improves the frequency resolution in the absence of a strongly defined minimum. Since $|\tilde{X}_{k,i}|^2$ is a quadratic function of $\alpha_{k,i}$ with positive curvature, the minimum can be found by setting the derivative to zero:

$$\frac{\partial \left[|\tilde{X}_{k,i}|^2 + \lambda_k \alpha_{k,i} \right]}{\partial \alpha_{k,i}} = 0 \tag{25}$$

which gives:

$$\hat{\alpha}_{k,i} = \min(\max(\bar{\alpha}_{k,i}, \alpha_{min}), \alpha_{max})$$
$$\bar{\alpha}_{k,i} = -\frac{\mathbf{R}\left\{ \tilde{X}_{k,i-1} W_L^{Rk} \left(X_{k,i} - \tilde{X}_{k,i-1} W_L^{Rk} \right)^* \right\} + \lambda_k}{|X_{k,i} - \tilde{X}_{k,i-1} W_L^{Rk}|^2} \tag{26}$$

where $\mathbf{R}\{x\}$ is the real part of x.

An example of this estimator is seen in Figure 4. The signal consists of a number of sinusoids that have been frequency modulated to a varying degree. The top plot shows the analysis with $\alpha_k = 1$, which corresponds to the analysis with only the

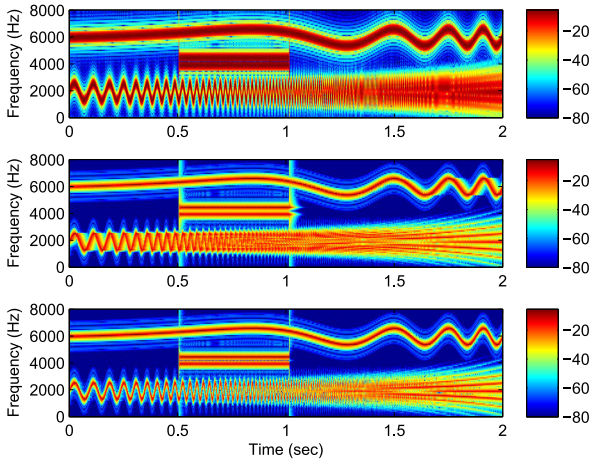

Fig. 4. Frequency modulated sinusoids. (Top) Short analysis window $h(n)$, $\alpha_k = 1$. (Middle) Long analysis window, $\alpha_k = 0.1$. (Bottom) Improved adaptive analysis window.

short window $h[n]$. The middle plot shows the analysis with $\alpha_k = 0.1$ and the bottom plot shows the analysis using the proposed estimator ($\lambda_k = 1e^{-5}$, $\alpha_{min} = 0.1$, $\alpha_{max} = 1$). It is seen that the top plot has poor frequency resolution due to the short window, while having a good time resolution. The middle plot has a good frequency resolution and can discriminate between two closely spaced sinusoids while having a poor time resolution, which results in smearing of the over time. It is also seen in the right part of the plot that when the modulations become so fast that they have several periods within the length of the window, the modulation is seen as a periodic function itself and is represented as harmonics of the underlying carrier. This illustrates the ambiguity that the same signal can be seen as fundamentally different depending on the window length. In this case, we prefer the longer window, since it is a more sparse representation, i.e. it represents the signal using fewer signal components. In the lower plot it is seen that the adaptive TF analysis gives a superior time- and frequency-resolution compared to the two fixed-resolution analyses and can adapt to each sinusoid independently. Unlike most adaptive TF analysis schemes, which can only adapt in either time or frequency, the proposed method can adapt to several signal components individually as long as they are separated in the combined TF domain.

An alternative estimator is given in [12] and [13] where it is shown that a minimax cross-entropy estimate of the squared magnitude TF distribution can be found as the minimum value of a set of M squared magnitude TF distributions at each (k, i) coordinate :

$$|\hat{X}_{k,i}|^2 = E \min_m S_{k,i,m} \tag{27}$$

where E is a normalization constant and $S_{k,i,m}$ is the set of M squared magnitude TF distributions $\{S_{k,i,m} : m \in (1, ..., M), M \geq 2\}$ calculated with equal-energy basis functions. In our case, the M different squared magnitude TF distributions are given by:

$$S_{k,i,m} = \frac{|\tilde{X}_{k,i,m}|^2}{E_{\tilde{h}_{k,i,m}}} \tag{28}$$

where

$$\tilde{X}_{k,i,m} = \alpha_m X_{k,i} + (1 - \alpha_m)\tilde{X}_{k,i-1,m} W_L^{Rk} \tag{29}$$

is the TF distribution for α_m and $E_{\tilde{h}_{k,i,m}}$ is the energy of the underlying window function that can be precomputed from (4) with $b_{k,i,0} = \alpha_m, a_{k,i,1} = 1 - \alpha_m$ and other coefficients set to zero. Thus, the minimax cross-entropy estimator is obtained by calculating M TF distributions and taking the minimum of the set of squared magnitude equal-energy windowed TF distributions in each (k, i) coordinate. This is in contrast to the estimator in (26) where there is no normalization of the energy before estimation.

The preceding estimators assume that the signal can be decomposed into signal components that are non-overlapping in the TF domain. For signals that are contaminated by high levels of noise, this is not a reasonable assumption and in this case we revert to a simpler estimator that uses a non-stationarity test to decide what resolution should be used. If a non-stationarity is detected, the TF distribution is calculated from the short window $h[n]$:

$$\tilde{X}_{k,i} = \alpha_k X_{k,i} \tag{30}$$

Otherwise, the TF distribution is updated using equation (21). Following a non-stationarity, the support of the underlying basis function $\tilde{g}_k[n]$ grows after each update, which results in a narrower bandwidth in the frequency domain. After a while it converges to the stationary basis function derived in section III-B depending on what value of α_k is used. It is noted that before the basis function has converged to the stationary case, the energy of the basis function will grow with time. However, if the value of α_k is known in advance, the energy and shape of the basis function can be precomputed and stored in a look-up table.

The non-stationarity test is based on the criteria from [14] which we adapted to our purpose in [15]:

$$\text{LR} = \sqrt{R} e^{-\frac{1}{2}(R-1)} \tag{31}$$

$$R = \frac{|X_{k,i}|^2/(\alpha_k^2 E_h)}{|\tilde{X}_{k,i-1}|^2/E_{\tilde{h}_{k,i-1}}} \tag{32}$$

where E_h is the energy of $h[n]$ and $E_{\tilde{h}_{k,i-1}}$ is the energy of the underlying window function at time $i-1$. LR is compared to a threshold value λ and if LR $< \lambda$ a non-stationarity is detected. To improve the robustness of the estimate, a non-stationarity is only detected if V adjacent frequency bands fail the test.

D. Discussion on adaptive time-frequency analysis

In section III-B and III-C different TF analysis schemes have been described. A constant analysis filter scheme that allows a specified bandwidth to be set for each frequency band as well as three adaptive schemes have been presented. A requirement for all the presented schemes is that they must

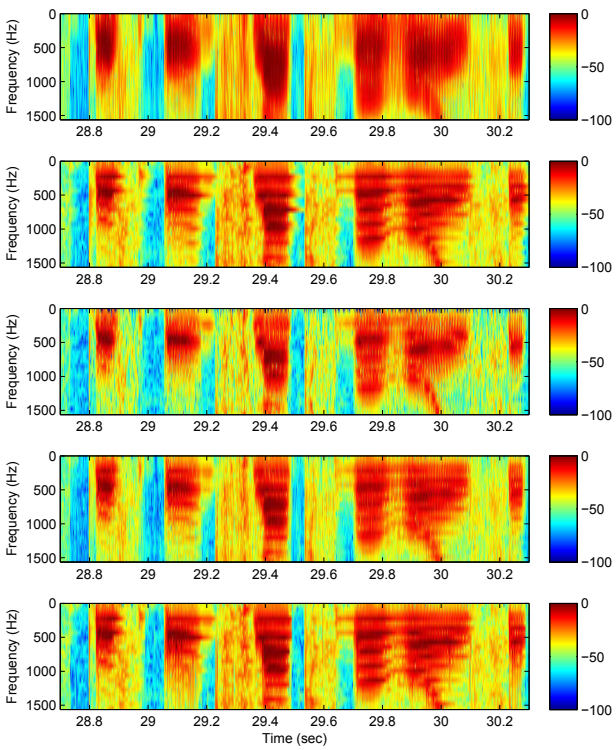

Fig. 5. Speech signal processed with 5 different TF analysis schemes. In descending order, (1) TF analysis using the short window $h[n]$, (2) Constant Q transform TF analysis, (3) Adaptive TF analysis that minimizes the energy, (4) Adaptive minimax cross-entropy TF analysis and (5) Adaptive TF analysis using a non-stationarity criteria.

be computationally simple and have a low delay to allow for a real-time implementation in an audio processing device such as a hearing aid. The low computational complexity is achieved by employing a first order filter on top of an oversampled DFT modulated filter bank and the low delay follows from the short analysis window $h[n]$ of length N and the exponential decay of the filter response that puts emphasis on the most recent part of the signal.

To compare the described TF distributions, a speech signal was processed through each of the different TF analysis schemes. The results are seen in Figure 5. The figure shows the low frequency area of the speech signal where the harmonics can be resolved by the TF analysis schemes, which illustrates the improved frequency resolution. The constant Q analysis scheme, however, has a poor time resolution due to the long analysis window. Especially it is seen that the speech energy is smeared in time at the end of the speech components due to the exponential decay of the analysis windows illustrating the well-known trade-off between time and frequency resolution. The three adaptive TF analysis schemes all have an improved frequency resolution compared to the short window $h[n]$ while also having a much better time resolution than the constant Q analysis scheme.

IV. Synthesis with low delay

The TF analysis presented in section II and III corresponds to an overcomplete transform of the signal, and therefore there is an infinite number of ways to reconstruct the signal from the analysis. There are numerous articles on perfect reconstruction modulated filter banks and other similar filter bank structures, see for instance [16], [17] and [4]. However, in many real-time audio processing applications it is not the reconstruction per se that is of interest, but the ability to apply a certain processing, for instance to apply a specified gain to each individual frequency band, and to do so with a low delay. This processing is time-varying and requires oversampling in the filter bank to avoid aliasing as in [18], [19] or [20]. The proposed adaptive TF analysis presented in this paper already has oversampling built into it. In this section, we focus on how to apply a desired gain to each individual frequency band with a low delay, by using an asymmetric synthesis window. The asymmetric window results in a non-linear phase response when a gain is applied to each individual frequency band, but since the ear is relatively insensitive to small phase distortions, there is no perceptual degradation of the signal. In the following, we disregard delays due to input/output buffering and processing of the data and measure the delay by the effective length of the combined analysis and synthesis window. Furthermore, we use the asymmetric window in a FIR filter design to further reduce the delay to N/2, similar to the filter bank equalizer in [6]. This is compared to a minimum phase FIR filter that represents the lowest possible delay for a given gain in each individual frequency band.

Since the adaptive TF analysis uses an individual analysis basis function in each frequency band, it would also require an individual synthesis basis function in each frequency band. However, since the adaptive TF analysis is calculated as a first order minimum-phase auto-regressive (AR) filtering of $h[n]$, this filtering itself is in theory invertible, although there could be numerical problems with trying to invert a time-varying AR filter. However, since the input to the AR filter is known a priori, it can be removed from the signal path and then only used to calculate a frequency dependent gain $\boldsymbol{G}$ vector applied to $X_k(z)$ as seen in Figure 6. This also greatly simplifies the design of the synthesis window and will be followed in the rest of this section.

A. Synthesis using an assymmetric window

Following Figure 1 and inserting the DFT-modulated synthesis formula (2) into (3), without applying a gain i.e. $Y_{k,i} = X_{k,i}$, the criterion for perfect reconstruction (PR) of the input signal is derived as:

$$\sum_{i=-\infty}^{\infty} h[m - iR]f[m - iR] = 1 \tag{33}$$

where it is used that $h[n]$ has length $N < L$. This condition is met for Hann windows for $R = 2^a, a \in \mathbf{N}$ and under similar conditions for many other windows such as the higher-order generalized cosine windows. It is noted that this requirement only ensures perfect reconstruction and does not imply that there is sufficient filtering being done by the windows. Several

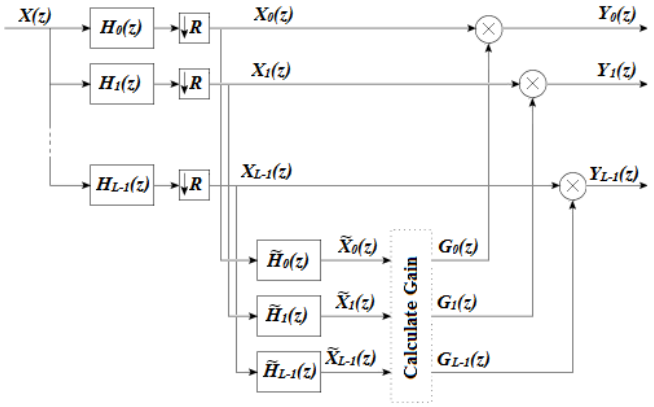

Fig. 6. Analysis scheme where the adaptive TF analysis is used to calculate a gain that is then applied to the primary signal path.

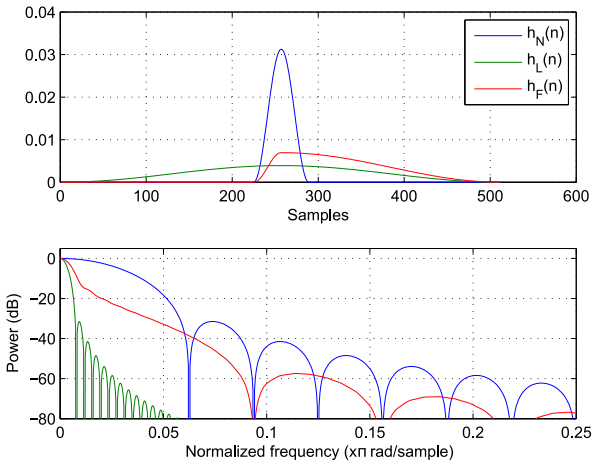

Fig. 7. Time domain (Top) and frequency domain (Bottom) representation of two symmetric Hann windows $h_N[n]$ and $h_L[n]$ and the asymmetric window $h_F[n]$

proposals have been made for optimizing these windows, while sometimes settling for near-perfect reconstruction, see for instance [21], [22] or [23].

An asymmetric window can be decomposed into the three following components:

$$\boldsymbol{h}_F[n] = [\mathbf{0}^T, \boldsymbol{h}_{lh}[n]^T, \boldsymbol{h}_{rh}[n]^T]^T \tag{34}$$

where $\mathbf{0}$ is a zero vector of size $(L-N)/2$, $\boldsymbol{h}_{lh}[n]$ is the left half of a window and $\boldsymbol{h}_{rh}[n]$ is the right half of a window. An asymmetric window of this type is seen in Figure 7 along with two symmetric Hann windows of length N and L respectively. The asymmetric window consists of the zero vector $\mathbf{0}$, the left half of a Hann window of length N and the right half of a Hann window of length L.

Setting the synthesis window equal to the analysis window $f[n] = h[n] = h_N[n]$ would achieve a delay of N samples and fulfills the PR criterion, but the resulting filtering would be poor. Setting $f[n] = h_L[n]$ achieves a higher degree of filtering in the synthesis stage as seen in the lower part of Figure 7. This comes at the price of an increased delay equal

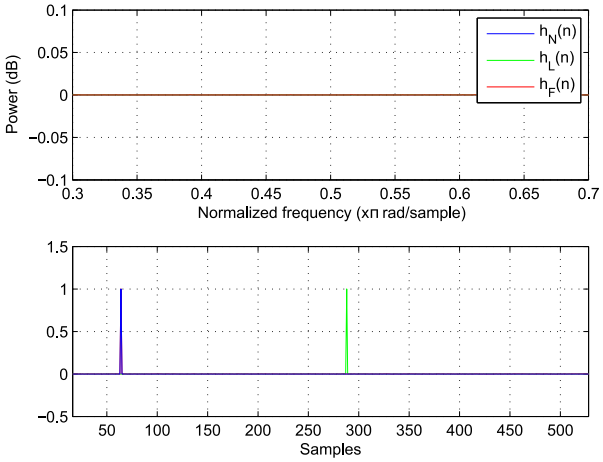

Fig. 8. Overall output of the analysis/synthesis scheme. (Top) Power spectrum. (Bottom) Impulse response. The blue impulse is on top of the red one.

to the length of the combined analysis and synthesis window $(L+N)/2$. Using the asymmetric window $f[n] = h_F[n]$ maintains the low delay of N samples, equivalent to the delay of $h_N[n]$, with a better filtering. The PR criteria is only fulfilled approximately, however as seen in the following paragraph, the overall frequency response only exhibits minor ripples.

To measure the response of the three different synthesis windows, white noise is sent through the analysis/synthesis scheme and the output is measured. The impulse response is measured by deconvolving the output with the input using division in the frequency domain. The overall impulse response is seen in Figure 8. It is seen that all three synthesis windows leads to a flat overall frequency response with less than 0.01dB ripple and that $f[n] = h_L[n]$ indeed leads to the predicted delay of $(L+N)/2 = 288$ samples while the two other synthesis windows leads to an equal delay of $N = 64$. If we assume a sample rate of 16kHz, this corresponds to a delay of 18ms and 4ms respectively. The impulse response of one frequency band is seen in Figure 9. It is seen that $f[n] = h_L[n]$ and $f[n] = h_N[n]$ leads to a symmetric impulse response which is due to the symmetric analysis and synthesis windows. The sharper frequency response with $f[n] = h_L[n]$ compared to $f[n] = h_N[n]$ comes at the price of the increased delay. $f[n] = h_F[n]$ leads to an asymmetric impulse response where the degree of asymmetry depends on how sharp the frequency response is. Thus, we characterize the analysis/synthesis system with synthesis windows $h_L[n]$ or $h_N[n]$ as linear phase filter banks and using synthesis window $h_F[n]$ as a mixed phase filter bank where the phase shift depends on the gain that is applied to the frequency bands (it is clearly not minimum phase due to the minimum delay of N samples).

B. Further reduction in delay using FIR filter and Minimum phase estimation

The asymmetric synthesis window used in the previous section reduces the delay to N samples by setting the first

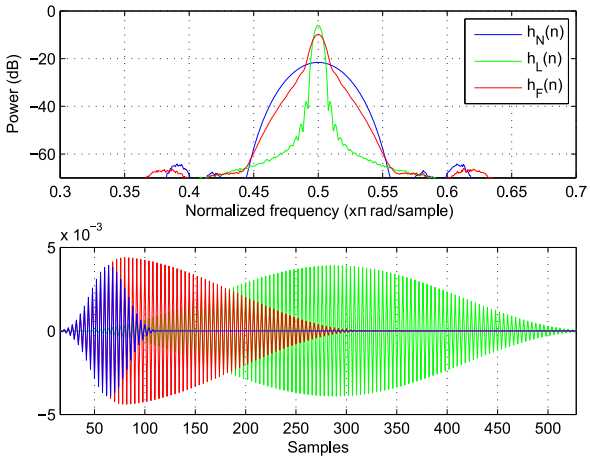

Fig. 9. Output of one frequency band in the analysis/synthesis scheme. (Top) Power spectrum. (Bottom) Impulse response.

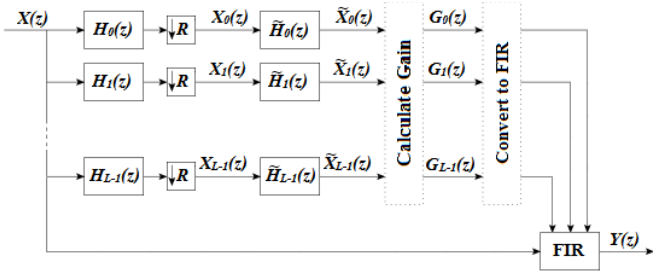

Fig. 10. Adaptive TF analysis where the gain is applied in a parallel signal path with a time-varying FIR filter.

$(L-N)/2$ samples of the synthesis window to zero. If a further reduction of the delay is needed, a parallel signal path without the R-fold decimation can be used, where the frequency dependent gain is applied to the signal as seen in Figure 10. The gain is applied to the signal by transforming it to the time-domain using a IDFT and filtering the signal with a corresponding FIR filter [24]. In this case, the gain can be interpreted as the magnitude response of the FIR filter. The delay for a time-domain filter, using a symmetric window of length N is $N/2$. Using an asymmetric window would in this case also provide a way to maintain the low delay of $N/2$ while giving a better filtering than a symmetric window. However, an even further reduction of the delay can be achieved by calculating the minimum phase filter for a given magnitude response. A minimum phase filter has the property of minimum group delay, which means that the energy is maximally concentrated in the low delay coefficients. Infinite impulse response (IIR) filters calculated by solving the set of Yule-Walker equations have been proposed in the context of speech enhancement [25], but minimum phase can equally well be realized using FIR filters, which is usually simpler to deal with when using time-varying filters, since stability is not an issue.

The gain vector $\boldsymbol{G}$ of length L is equal to the magnitude response of the desired minimum phase filter:

$$\boldsymbol{H}_{min} = |\boldsymbol{H}_{min}|e^{j\Theta} = \boldsymbol{G}e^{j\Theta} \tag{35}$$

where $\boldsymbol{H}_{min}$ is the minimum phase frequency response vector and Θ is the phase vector that must be estimated. The filter coefficients are found by taking the IDFT of $\boldsymbol{H}_{min}$ and then lowpass filtering each coefficient over time to smooth it. The phase Θ can be found by calculating the zeros of the corresponding linear phase response and then reflecting the zeros that are outside the unit circle inside the unit circle [26]. This method however is prone to numerical inaccuracies, especially for large filter orders. Instead, an approximate nonparametric method involving the cepstrum is used. It has been shown [27] that Θ can be calculated as:

$$\Theta = -jDFT[\boldsymbol{s} \cdot IDFT[\log \boldsymbol{G}]] \tag{36}$$

where

$$s(k) = \begin{cases} 0 & \text{if } k = 0, k = L/2, \\ 1 & \text{if } 0 < k < L/2, \\ -1 & \text{if } L/2 < k < L. \end{cases} \tag{37}$$

The accuracy of this method depends on the length of the (I)DFT and how steep the cuts in the magnitude response are. In audio noise reduction, the gain is usually limited to a range between Gmin and 0 dB, where Gmin is the minimum gain that is used. Setting Gmin to -20 dB is usually sufficient to give good noise reduction and avoid noisy artifacts, and also prevents significant artifacts from the minimum phase estimation of Θ.

C. Performance analysis

Table I shows a comparison of the synthesis schemes that have been discussed and lists the total delay for a 4ms analysis window at 16kHz ($N = 64$). y_N is the output signal when using a symmetric window of length N as synthesis window, y_L is the output when using a symmetric window of length L, y_A is the output when using an asymmetric window, FIR_A is the output when using a FIR filter designed with an asymmetric window, and FIR_{min} is the output when using a minimum phase FIR filter. The phase shift is due to the asymmetric window or the minimum phase filter calculation that is resulting in a non-linear phase response. The computational complexity of the different schemes can be seen in Table II. The DFT is considered to be computed with a Radix-2 FFT algorithm and the cost of calculating the FFT of a real sequence of length L is approximately half of the cost of a complex sequence of length L. All FFT's are applied to real sequences except the last inverse FFT to calculate the filter coefficients of the minimum phase filter FIR_{min}. The symbols $\mathcal{M}$ and $\mathcal{A}$ represent the multiplication and addition cost respectively of an exponential and a logarithm operation, whose values depend on the actual implementation. The coefficients of the FIR filters are smoothed over time using a first-order auto-regressive filter, which can be implemented using 1 multiplication and 2 additions per sample. Since FIR_{min} has most of the energy concentrated in the lowest order coefficients, its filter order is reduced to $2N$ to save

TABLE I
COMPARISON OF DELAY FOR DIFFERENT SYNTHESIS SCHEMES.

synthesis scheme	window length	delay
y_N	4ms	4ms
y_L	32ms	18ms
y_A	32ms	4ms+phase shift
FIR_A	32ms	2ms+phase shift
FIR_{min}	32ms	0ms+phase shift

TABLE II
COMPARISON OF COMPUTATIONAL COMPLEXITY FOR DIFFERENT SYNTHESIS SCHEMES.

synthesis scheme	multiplications	additions
y_N	$\frac{1}{R}(L\log_2 L + N)$	$\frac{1}{R}(\frac{3}{2}L\log_2 L + N)$
y_L	$\frac{1}{R}(L\log_2 L + L)$	$\frac{1}{R}(\frac{3}{2}L\log_2 L + L)$
y_A	$\frac{1}{R}(L\log_2 L + \frac{L+N}{2})$	$\frac{1}{R}(\frac{3}{2}L\log_2 L + \frac{L+N}{2})$
FIR_A	$\frac{L}{R}\log_2 L + \frac{L+N}{2}$	$\frac{3L}{2R}\log_2 L + \frac{L+N}{2}$
FIR_{min}	$\frac{4L}{R}\log_2 L + \frac{\mathcal{M}}{R} + 4N$	$\frac{6L}{R}\log_2 L + \frac{\mathcal{A}}{R} + 6N$

computations. This is found to have no significant impact on the output signal.

To evaluate the performance of each synthesis scheme for a noise reduction task, 100 speech sentences from the TIMIT database recorded at 16kHz were mixed with noise at different signal-to-noise ratios (SNR) and used as input to each of the described schemes. Two different noise types were used; a traffic noise signal and a babble noise signal consisting of 12 individual speakers. For the analysis, the following values were used: $L = 512, N = 64, R = 16$, $h[n]$ is the Hann window and the adaptive TF analysis with non-stationarity detection was used with $\alpha_k = 0.05$. The gain is calculated as the Wiener gain with the recursive a-priori SNR estimator [28] with the smoothing parameter $a = 0.97$. The noise estimate is a lowpass filtered version of the noise signal. The gain coefficients are the same for all schemes, which only differ in the synthesis method. The average segmental SNR and PESQ value [29] is calculated and plotted in Figure 11 and 12. It should be noted that the use of SNR values for a noise reduction scheme that is not linear phase is not fully justified since it is not possible to align the input and output signals completely. This is clearly seen for FIR_{min} that performs more than 2dB worse than any other scheme, which is indeed due to the minimum phase property of the filter. For FIR_{min} and the asymmetric window schemes, the performance should be judged more on the perceptual quality which is measured by PESQ. It is seen that using y_L gives an improvement of about 1-2dB in segmental SNR compared to y_N, except at very low SNR values, at the cost of an added delay of 14ms. The asymmetric window y_A gives a good tradeoff between the two symmetric windows with a low delay of 4ms and about 0.5dB less than y_L. FIR_A leads to a slight deterioration in performance compared to y_A. Looking at the PESQ values, y_L performs about 0.1-0.3 better than y_N while y_A again has a good tradeoff between delay and performance. The FIR filters perform better in the perceptual test which is as expected, although FIR_{min} still performs worse. Further investigation has revealed that this is because the gain, which is calculated from the analysis window is delayed compared to the 0ms

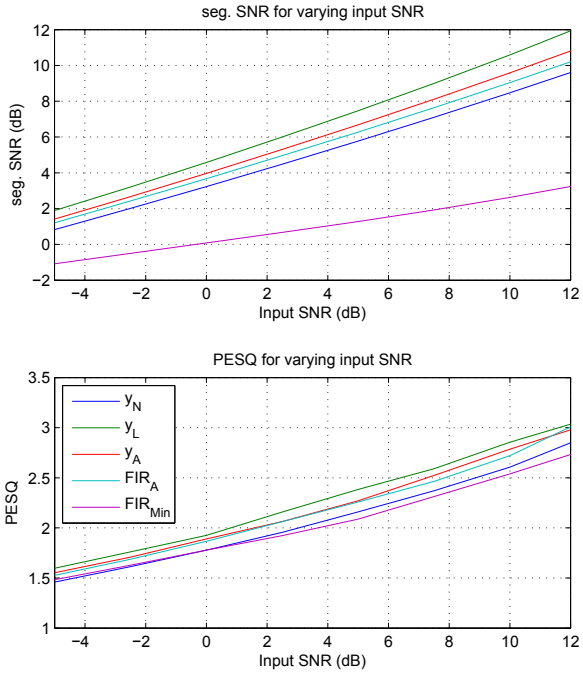

Fig. 11. Experimental results for different synthesis schemes for traffic noise.

delay in the signal path for FIR_{min}. Therefore, the gain does not react sufficiently fast enough to catch the onsets of the speech. If a delay of 4ms is inserted into the signal path, then FIR_{min} performs similarly to FIR_A. The measures for traffic noise are higher than for babble noise, which is explained by the fact that babble noise is a harder problem due to the highly modulated frequency content in the signal. Informal listening tests have been performed that are in agreement with the PESQ results. Perceptually y_N has more artifacts than the other output signals and creates more distortion of the speech.

The data sets have also been tested on a combination of different TF analysis schemes and synthesis schemes - namely the adaptive TF analysis scheme with the asymmetric synthesis window h_F used in the previous tests, the analysis using the short window $h[n]$ together with the short symmetric synthesis window h_N and the analysis using the short window $h[n]$ together the asymmetric synthesis window h_F. The gain is calculated in the same way as in the previous experiments, except that the noise is estimated using an unbiased MMSE-based noise estimation method [30], to investigate the effect of the adaptive analysis scheme. The PESQ results can be seen in Figure 13. It is seen that using the asymmetric synthesis window together with the short analysis window gives no improvement over the short symmetric synthesis window. Using the adaptive analysis together with the asymmetric window gives an improvement at higher SNR's of about 0.3 for traffic noise and 0.1 for babble noise compared to the two other analysis/synthesis schemes. This can be explained by the fact that the short analysis is matched with h_N as they have the same frequency resolution. The better frequency

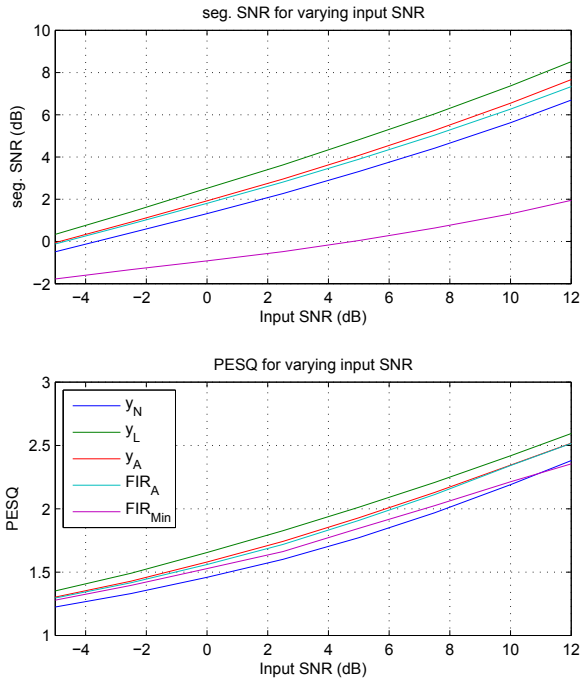

Fig. 12. Experimental results for different synthesis schemes for babble noise.

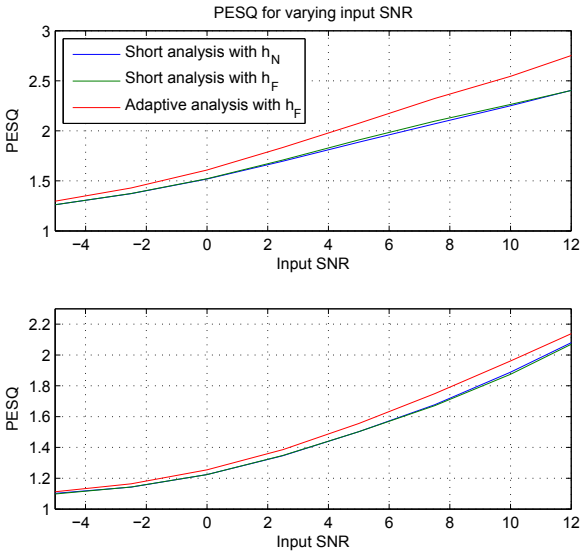

Fig. 13. Experimental results for different combinations of analysis and synthesis schemes. (Top) Traffic noise, (Bottom) Babble noise.

resolution of the synthesis window h_F compared to h_N can not be used by the short analysis window and therefore there is no improvement in performance. Only using the adaptive TF analysis scheme leads to an improvement.

V. Conclusion and Future Outlook

An adaptive TF analysis/synthesis scheme for processing of audio with a low delay from 0 to 4ms and low computational complexity has been proposed in this paper. The scheme is suitable for real-time sound devices such as mobile communication devices or hearing aids. The scheme can obtain a nonuniform frequency resolution with any desired bandwidth for each frequency band. It can adapt the TF distribution independently in each frequency band according to the signal to obtain a time-varying TF analysis. A number of adaptive TF analysis schemes have been described that are suitable for a real-time implementation. Synthesis of the gain-modified signal with a low delay is achieved using an asymmetric synthesis window that results in a mixed phase filter bank that maintains the low delay of a reference filter bank but with an improved frequency response. Further reduction in the delay is achieved by applying the gain to the signal as a time-varying FIR filter. A noise reduction experiment was performed to show the proposed scheme's superior performance compared to a reference scheme with the same delay.

Future research includes investigating alternative ways to use the information from the adaptive TF analysis to further improve the processing of the audio signal. Many real-time audio processing applications, such as hearing loss compensation, frequency transposition and acoustic echo cancellation utilize a TF distribution and it should be investigated how such an adaptive analysis/synthesis scheme can improve their performance.

References

[1] J. Agnew and J. M. Thornton, "Just noticeable and objectionable group delays in digital hearing aids," *Journal of the American Academy of Audiology, vol. 11, no. 6, pp. 330336*, 2000.

[2] M. Vetterli and J. Kovačevic, *Wavelets and Subband Coding*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[3] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing*. Prentice-Hall, 1983.

[4] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Upper Saddle River, New Jersey: Prentice-Hall, 1993.

[5] J. M. Kates and K. H. Arehart, "Multichannel dynamic-range compression using digital frequency warping," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 3003–3014, Jan. 2005. [Online]. Available: http://dx.doi.org/10.1155/ASP.2005.3003

[6] H. W. Löllmann and P. Vary, "Low delay filter-banks for speech and audio processing," in *Speech and Audio Processing in Adverse Environments*, ser. Signals and Communication Technology, E. Hnsler and G. Schmidt, Eds. Springer Berlin Heidelberg, 2008, pp. 13–61. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-70602-1_2

[7] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, Jul 1989.

[8] M. Liuni, A. Robel, E. Matusiak, M. Romito, and X. Rodet, "Automatic adaptation of the time-frequency resolution for sound analysis and re-synthesis," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 5, pp. 959–970, May 2013.

[9] H. Sepehr, A. Y. Nooralahiyan, and P. V. Brennan, "Improving performance of a noise reduction algorithm by switching the analysis filter bank," in *Image and Signal Processing*, ser. Lecture Notes in Computer Science, A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass, and J. Meunier, Eds. Springer Berlin Heidelberg, 2010, vol. 6134, pp. 262–271. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13681-8_31

[10] D. Rudoy, P. Basu, and P. Wolfe, "Superposition frames for adaptive time-frequency analysis and fast reconstruction," *Signal Processing, IEEE Transactions on*, vol. 58, no. 5, pp. 2581–2596, May 2010.

[11] D. Mauler and R. Martin, "A low delay, variable resolution, perfect reconstruction spectral analysis-synthesis system for speech enhancement," *15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland*, 2007.

[12] P. Loughlin, J. Pitton, and B. Hannaford, "Approximating time-frequency density functions via optimal combinations of spectrograms," *Signal Processing Letters, IEEE*, vol. 1, no. 12, pp. 199–202, Dec 1994.

[13] ——, "Fast approximations to positive time-frequency distributions, with applications," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, vol. 2, May 1995, pp. 1009–1012 vol.2.

[14] D. Mauler and R. Martin, "Improved reproduction of stops in noise reduction systems with adaptive windows and nonstationarity detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, p. 469480, 2009. [Online]. Available: http://asp.eurasipjournals.com/content/2009/1/469480

[15] K. Andersen and M. Moonen, "An adaptive time-frequency analysis scheme for improved real-time speech enhancement," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 6265–6269.

[16] M. Vetterli and D. Le Gall, "Perfect reconstruction fir filter banks: some properties and factorizations," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 1057–1071, Jul 1989.

[17] X.-G. Xia and B. Suter, "Fir paraunitary filter banks given several analysis filters: Factorizations and constructions," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 720–723, 1996. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0030108619&partnerID=40&md5=514df8d1ec9cb473b922ab045f63ad31

[18] K. Eneman and M. Moonen, "Dft modulated filter bank design for oversampled subband systems," *Signal Processing*, vol. 81, no. 9, pp. 1947 – 1973, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168401000846

[19] H. Bolcskei and F. Hlawatsch, "Oversampled cosine modulated filter banks with perfect reconstruction," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 45, no. 8, pp. 1057–1071, Aug 1998.

[20] N. Grbic, J. de Haan, I. Claesson, and S. Nordholm, "Design of oversampled uniform dft filter banks with reduced inband aliasing and delay constraints," in *Signal Processing and its Applications, Sixth International, Symposium on. 2001*, vol. 1, 2001, pp. 104–107 vol.1.

[21] H. Dam, S. Nordholm, A. Cantoni, and J. de Haan, "Iterative method for the design of dft filter bank," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 51, no. 11, pp. 581–586, Nov 2004.

[22] K. F. C. Yiu, N. Grbic, S. Nordholm, and K. L. Teo, "Multicriteria design of oversampled uniform dft filter banks," *Signal Processing Letters, IEEE*, vol. 11, no. 6, pp. 541–544, June 2004.

[23] D. Hermann, E. Chau, R. Dony, and S. Areibi, "Window based prototype filter design for highly oversampled filter banks in audio applications," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2, April 2007, pp. II–405–II–408.

[24] H. W. Löllmann and P. Vary, "Generalized filter-bank equalizer for noise reduction with reduced signal delay," *Proc. of European Conf. on Speech Communication and Technology (Interspeech, Lisbon, Portugal*, Sep. 2005.

[25] ——, "Low delay filter for adaptive noise reduction," *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 205–208, Sep. 2004.

[26] O. Herrmann and W. Schuessler, "Design of nonrecursive digital filters with minimum phase," *Electronics Letters*, vol. 6, no. 11, pp. 329–330, May 1970.

[27] N. Damera-Venkata, B. Evans, and S. McCaslin, "Design of optimal minimum-phase digital fir filters using discrete hilbert transforms," *Signal Processing, IEEE Transactions on*, vol. 48, no. 5, pp. 1491–1495, May 2000.

[28] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 6, pp. 1109–1121, 1984.

[29] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2013.

[30] T. Gerkmann and R. Hendriks, "Unbiased mmse-based noise power estimation with low complexity and low tracking delay," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 4, pp. 1383–1393, May 2012.

Kristian Timm Andersen (S'13) was born in Herlev, Denmark, in 1983. He received the M.Sc. degree in electrical engineering from the Technical University of Denmark, Denmark, in 2008. He is currently pursuing the PhD degree at the Electrical Engineering Department of KU Leuven, and is a member of KU Leuvens STADIUS group. His research interests include audio signal processing, acoustical signal enhancement, and real-time audio applications.

From 2008 to 2013, he was a development DSP Engineer at Widex, A/S, Lynge, Denmark, where he developed novel DSP algorithms for hearings aids, including 8 patented (or patent-pending) inventions.

Marc Moonen (M'94-SM'06-F'07) received the electrical engineering degree and the PhD degree in applied sciences from KU Leuven, Belgium, in 1986 and 1990 respectively. Since 2004 he is a Full Professor at the Electrical Engineering Department of KU Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing.

He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received a journal best paper award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING (with Geert Leus) and from Elsevier Signal Processing (with Simon Doclo). He was chairman of the IEEE Benelux Signal Processing Chapter (19982002), and a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and is currently President of EURASIP (European Association for Signal Processing).

He has served as Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (20032005), and has been a member of the editorial board of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, IEEE Signal Processing Magazine, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking, and Signal Processing. He is currently a member of the editorial board of EURASIP Journal on Applied Signal Processing and Area Editor for Feature Articles in IEEE Signal Processing Magazine.