Two-Stage Filter-Bank System for Improved Single-Channel Noise Reduction in Hearing Aids

Alexander Schasse, Timo Gerkmann, *Member, IEEE*, Rainer Martin, *Fellow, IEEE*, Wolfgang Sörgel, Thomas Pilgrim, and Henning Puder, *Member, IEEE*

***Abstract*—The filter-bank system implemented in hearing aids has to fulfill various constraints such as low latency and high stop-band attenuation, usually at the cost of low frequency resolution. In the context of frequency-domain noise-reduction algorithms, insufficient frequency resolution may lead to annoying residual noise artifacts since the spectral harmonics of the speech cannot properly be resolved. Especially in case of female speech signals, the noise between the spectral harmonics causes a distinct roughness of the processed signals. Therefore, this work proposes a two-stage filter-bank system, such that the frequency resolution can be improved for the purpose of noise reduction, while the original first-stage hearing-aid filter-bank system can still be used for compression and amplification. We also propose methods to implement the second filter-bank stage with little additional algorithmic delay. Furthermore, the computational complexity is an important design criterion. This finally leads to an application of the second filter-bank stage to lower frequency bands only, resulting in the ability to resolve the harmonics of speech. The paper presents a systematic description of the second filter-bank stage, discusses its influence on the processed signals in detail and further presents the results of a listening test which indicates the improved performance compared to the original single-stage filter-bank system.**

***Index Terms*—Cascaded filter-bank system, hearing aids, low delay processing, single-channel noise reduction.**

I. Introduction

THE purpose of hearing aids is to compensate auditory dysfunctions as experienced by hearing impaired people. The most common class is the *sensorineural* hearing loss [1], which results in several, and in general frequency dependent effects. First of all, the listening threshold is increased due to a dysfunction of the outer hair cells in the cochlea. However, the threshold of uncomfortable sound level is usually only slightly affected. Therefore, the first task of a hearing aid is to compress the captured signals to match the reduced dynamic range perceived by the impaired listener. This is done through an amplification of soft sounds, while loud sounds remain unchanged. Secondly, the temporal and spectral resolution of the cochlea is decreased, which affects the ability to separate target speech from ambient background noise or competing speakers. As a countermeasure, hearing aids apply speech enhancement to improve the signal-to-noise ratio (SNR) which has a positive effect, not only on the listening comfort but also on speech intelligibility [1]. Modern hearing aids usually make use of multiple microphones on each device to apply beamforming algorithms or even provide a wireless binaural link between the hearing aids on both ears. Since the microphone distances are usually very small on a single device, the largest gain can be achieved by utilizing differential beamformers [2], [3]. Such techniques can be used to reduce the sound coming from the back direction or to adaptively cancel a target noise source. However, after directional processing, single-channel noise reduction (NR) is still of great importance to further boost the signal quality and reduce listening effort.

The effects caused by a *sensorineural* hearing loss are in general frequency dependent. Therefore, it is reasonable to separate the microphone signal(s) into several frequency channels and apply, e.g., different compression rules and NR gains to each of these sub-bands. One efficient way to implement a frequency analysis, are uniformly-modulated filter-bank systems (FBS) [4], [5]. In this case, a fixed prototype low-pass filter is modulated to the center frequency of the respective subband, and finally used as a band-pass filter. This approach is especially efficient since it can be implemented using the fast Fourier transform (FFT) algorithm. Other filter-bank structures try to mimic the frequency decomposition of the human cochlea, e.g. the gammatone filter-bank system [6] or the constant Q transform [7], [8], but have usually a higher computational complexity and/or do not provide perfect signal reconstruction.

When developing a uniformly-modulated FBS for hearing aid purposes, [9]–[11] propose methods for the design of prototype low-pass filters for signal analysis and resynthesis. The design of these filters, as well as the proper choice of the DFT length K, the window length L and the decimation factor R are the design criteria which define the computational complexity as well as the resulting stop-band attenuation. Especially the latter is of great importance, since strong amplification of up to 60 dB [12] is applied to the frequency channels. The prototype filters or window functions have to provide a high stop-band attenuation to minimize the crosstalk between different sub-bands.

Manuscript received May 26, 2014; revised September 03, 2014; accepted October 15, 2014. Date of publication October 30, 2014; date of current version January 16, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Roberto Togneri.

A. Schasse and R. Martin are with the Institute of Communication Acoustics (IKA), Ruhr-Universität Bochum, 44780 Bochum, Germany (e-mail: alexander.schasse@rub.de; rainer.martin@rub.de).

H. Puder, W. Sörgel, and T. Pilgrim are with Siemens Audiologische Technik GmbH, 91058 Erlangen, Germany.

T. Gerkmann was with IKA, Ruhr-Universität Bochum, 44780 Bochum, Germany. He is now with the Speech Signal Processing Group, Cluster of Excellence "Hearing4all," Universität Oldenburg, 26111 Oldenburg, Germany (e-mail: timo.gerkmann@uni-oldenburg.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2014.2365992

This design process usually results in a FBS which is optimized in terms of computational complexity, processing delay (max. of 10 ms [13]) and stop-band attenuation, but it also does not allow a high frequency resolution.

In the context of NR, we assume the additive noise model in the short-time frequency domain

$$Y(k,m) = X(k,m) + V(k,m), \tag{1}$$

where $X(k,m)$ is the clean speech signal, $V(k,m)$ is the noise signal, and $Y(k,m)$ is the noisy mixture. The indexes m and k represent the time-frame and frequency-bin index, respectively. NR algorithms like the Wiener filter [14], [15], or the minimum mean-square error (MMSE) (log) amplitude estimators proposed in [16], [17] apply instantaneous and real-valued gains $G(k,m)$ which result in the estimated clean speech component

$$\hat{X}(k,m) = G(k,m)Y(k,m). \tag{2}$$

The respective time-domain signals are defined as $y(n)$ for the noisy input signal, $x(n)$ is the clean speech signal, $v(n)$ is the additive noise and $\hat{x}(n)$ is the enhanced output signal. Of course, the performance of the NR algorithm depends on the properties of the spectral analysis system. If the frequency resolution is too low, the noise between the harmonics cannot be removed, which leads to annoying residual noise. Therefore, and as an extension of [18], we further investigate the concept of a cascaded FBS with the purpose to improve the frequency resolution for NR, while the remaining tasks of the hearing aid, such as amplification and compression, can still be performed in the original FBS. We evaluate new concepts of an efficient implementation, discuss the influence when reducing the computational load and evaluate the different FBS implementations in the context of single-channel noise reduction in a listening test.

The paper is structured as follows. In Section II, we analyze the effect of poor frequency resolution to noise reduction algorithms. To do that, we define a filter-bank system appropriate for hearing aids and implement a frequency-domain Wiener filter. Section III then proposes the concept of the cascaded FBS, and discusses approaches to implement the second filter-bank stage with low additional algorithmic delay. Section IV then discusses procedures to reduce the computational load systematically by sharing the NR gains in spectrally overlapping sub-bands. Section V presents objective results, as well as the results of a listening test. Finally, in Section VI we summarize the main results of our contribution and give a short outlook on future research topics.

II. Single-Channel Noise Reduction in Hearing Aids

Assuming the single-channel, noisy signal model of (1), we first design a FBS using the approach proposed in [9]. As design parameters, we set the DFT length to $K = 64$, the length of the analysis and synthesis prototype filters to $L_{\mathrm{Ana}} = L_{\mathrm{Syn}} = 80$, the decimation factor to $R = 16$, and we process signals at a sampling frequency of $f_{\mathrm{s}} = 16$ kHz. This leads to a band distance of $\Delta f_1 = 250$ Hz and a frame rate of $\Delta T_1 = 1$ ms. The delay of this filter-bank system is 4 ms.

Since hearing aid manufacturers only provide little information about their NR systems, we implemented a frequency-domain Wiener filter as a well established reference algorithm. The Wiener filter with gain $G(k,m)$ is based on the *decision-directed* approach [16] to estimate the *a-priori* SNR $\hat{\eta}(k,m)$ in each frequency bin k at each time frame m as

$$\hat{\eta}(k,m) = \lambda \frac{\left|\hat{X}(k,m-1)\right|^2}{\hat{\phi}_V(k,m)} + (1-\lambda)\max[\hat{\gamma}(k,m)-1,0] \tag{3}$$

$$G(k,m) = \max\left[\frac{\hat{\eta}(k,m)}{1+\hat{\eta}(k,m)}, \mathrm{SF}\right]. \tag{4}$$

Here, $\hat{\gamma}(k,m) = |Y(k,m)|^2/\hat{\phi}_V(k,m)$ is an estimate of the *a-posteriori* SNR. Furthermore, we apply a spectral floor to the Wiener gains of $\mathrm{SF} = -15$ dB to ensure a natural sounding background noise in the processed signals. We use an idealized noise power estimator based on the usually unknown noise signal $V(k,m)$ defined as

$$\hat{\phi}_V(k,m) = \lambda_V \hat{\phi}_V(k,m-1) + (1-\lambda_V)|V(k,m)|^2. \tag{5}$$

This facilitates the comparison of different FBS settings in later sections, since we only need to adapt the smoothing parameter λ_V to the respective sampling frequency and decimation ratio to provide comparable noise estimates.

Fig. 1(a) shows the spectrogram of a short sequence of female speech (the fundamental frequency is in a range of 180 Hz - 220 Hz) mixed with traffic noise at a segmental input SNR of 0 dB, using the FBS as designed above for frequencies up to 1.5 kHz. The figure clearly shows the low frequency resolution which is not able to resolve the spectral harmonics of the speech signal. After applying the spectral gains of the Wiener filter, the noise power is clearly reduced as shown in Fig. 1(b), but a further analysis of the processed signal in Fig. 1(c) using a frequency resolution of 15.625 Hz reveals that noise between the harmonics is not affected. This leads to residual noise artifacts during speech activity, especially in case of high fundamental frequencies. As a comparison, Fig. 1(d) shows the spectrogram of an improved speech signal after processing in a FBS with a frequency resolution of 31.25 Hz. Finally, Fig. 1(e) compares the spectra of a single frame observed at 5.6 s in more detail. We can observe that the distance between the spectral harmonics of the speech signal and the residual noise is larger if the NR algorithm is applied at the higher frequency resolution. In case of a filter-bank system providing a band distance of 250 Hz in combination with single-channel NR, one experiences a distinct roughness in the processed signals due to the residual noise between the harmonics. If the fundamental frequency is larger, e.g., as in case of female or child speech, this effect is even increased. Therefore, we like to improve the frequency resolution for NR without affecting the stop-band attenuation or the low processing delay. In this context, we propose a cascaded filter-bank design as an extension of a well-tuned hearing-aid FBS.

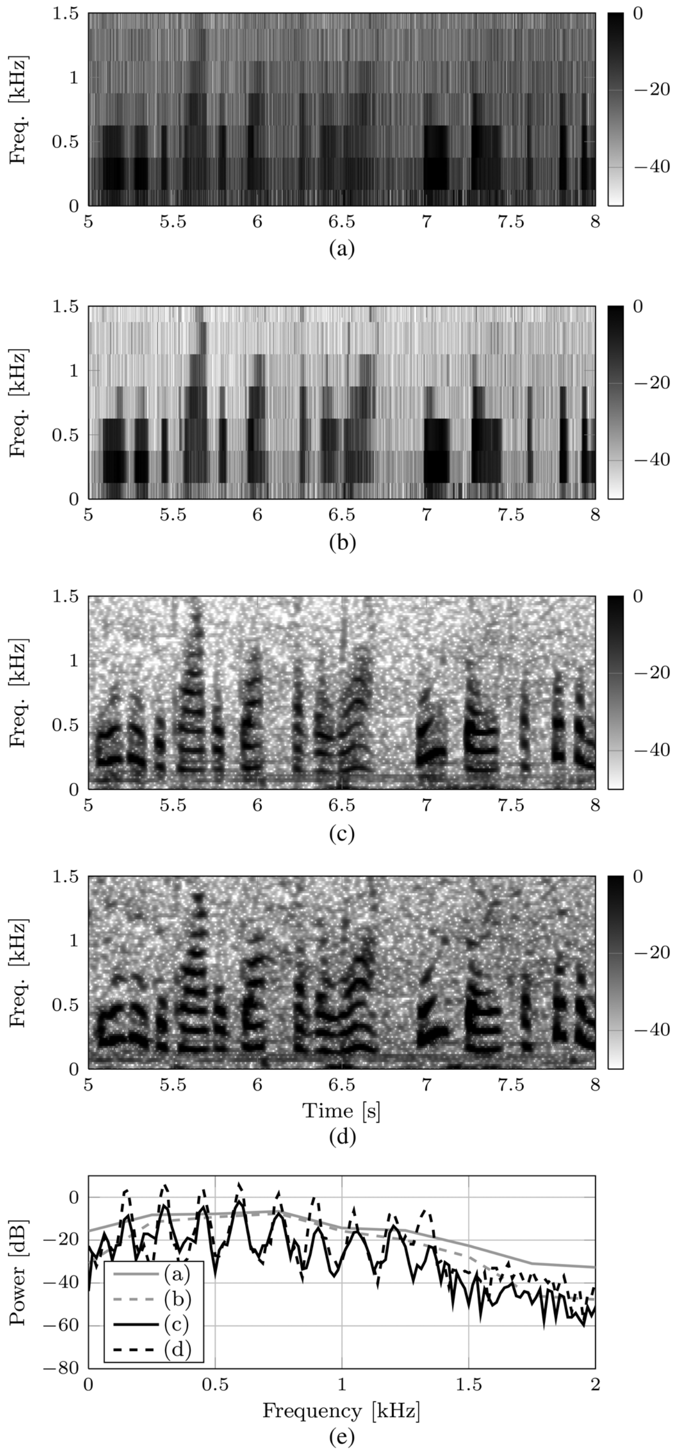

Fig. 1. Frequency resolution comparison for female speech mixed with traffic noise at 0 dB seg. input SNR: (a) noisy input speech, low frequency resolution; (b) processed speech, low frequency resolution; (c) same signal as in (b), but the spectrogram was computed at a high frequency resolution; (d) the NR was performed at a high frequency resolution. All spectrograms show a frequency range of $0-1.5$ kHz and the same dynamic range of 50 dB. The plot in (e) compares the spectra of a single voiced frame at 5.6 s.

III. Cascaded Filter-Bank System

The DFT coefficients computed in the FBS as discussed above can be interpreted as sub-band signals with time index m and sub-band index k. Therefore, we can easily implement a second FBS to perform a further frequency analysis and resynthesis of each sub-band signal. This leads to a short-time frequency-domain representation of the kth sub-band signal $Y(k,m)$ defined as $\mathcal{Y}_k(\kappa,\mu)$. The indexes μ and κ indicate the time frame and frequency bin within the second filter-bank stage. When applying no decimation in the second stage, the time-frame indexes of the first and second FBS are the same. Otherwise, we set $\mu = \mathcal{R}m$, where $\mathcal{R}$ is the decimation factor of the second FBS. The resulting frequency resolution is then given by the DFT length $\mathcal{K}$ and is defined as $\Delta f_2 = f_{\mathrm{s},2}/\mathcal{K}$, where $f_{\mathrm{s},2} = f_{\mathrm{s}}/R$ is the sampling frequency of the sub-band signals $Y(k,m)$. Fig. 2 shows an overview of the two-stage filter-bank system. For later purposes we have already introduced k', which defines the highest frequency channel of the first stage to which the second stage is applied. Since we are processing real-valued signals $y(n)$ and because of spectral symmetry, the figure considers the frequency channels $k = 0, 1, \ldots, K/2$ only.

A noise reduction algorithm such as the Wiener filter can now be implemented in the second stage likewise to Section II. All smoothing parameters, e.g. λ_V in (5), have to be adapted to the new filter-bank settings to achieve the same amount of variance reduction. The challenge of this two-stage design, however, is to achieve a low overall signal delay. Solutions to this challenge are reviewed next.

A. Filter-Bank Equalizer

The filter-bank equalizer (FBE) proposed in [19]–[21] is a technique to efficiently implement a signal modification by spectral weights as a time-domain filter. It is in principal also able to provide a non-uniform frequency resolution, however, we use a uniformly modulated filter-bank implementation. Furthermore, we especially focus on [20] which presents methods to reduce the algorithmic delay of the FBE by using short finite-impulse response (FIR) or infinite-impulse response time-domain filters (IIR). Fig. 3 shows a block diagram of the FBE in the context of the proposed two-stage FBS. Here, the noisy input samples of the kth sub-band signal $Y(k,m)$ are rearranged into segments of length $\mathcal{L}_{\mathrm{Ana}}$, multiplied with an analysis window and transformed to the frequency domain of the second filter-bank stage, using a fast discrete Fourier transform (DFT) implementation. Then, an NR algorithm computes the spectral gains $\mathcal{G}_k(\kappa,\mu)$. The idea of the FBE is now to compute linear-phase FIR or minimum-phase IIR filters, that correspond to the NR gains. This is done by transforming the spectral gains back to the sub-band time domain using the inverse generalized DFT (IGDFT), multiply the result with a prototype low-pass filter to ensure perfect reconstruction in case of $\mathcal{G}_k(\kappa,\mu) = 1$, and approximate this filter by (shorter) linear-phase FIR filters or minimum-phase IIR filters. These time-varying filters are then applied within each frequency channel of the first filter-bank stage instead of a conventional multiplication in the respective sub-band frequency domain.

Besides the parameters of the second filter bank stage, the FBE is defined by the linear phase FIR filter length $\mathcal{L}_{\mathrm{FIR}}$ or the respective minimum-phase IIR filter length $\mathcal{L}_{\mathrm{IIR}}$. In case of the linear-phase FIR filter, the filter length defines the algorithmic delay of the processing which is given as $(\mathcal{L}_{\mathrm{FIR}}-1)/2$. In case of the minimum-phase IIR filter, the algorithmic delay is close

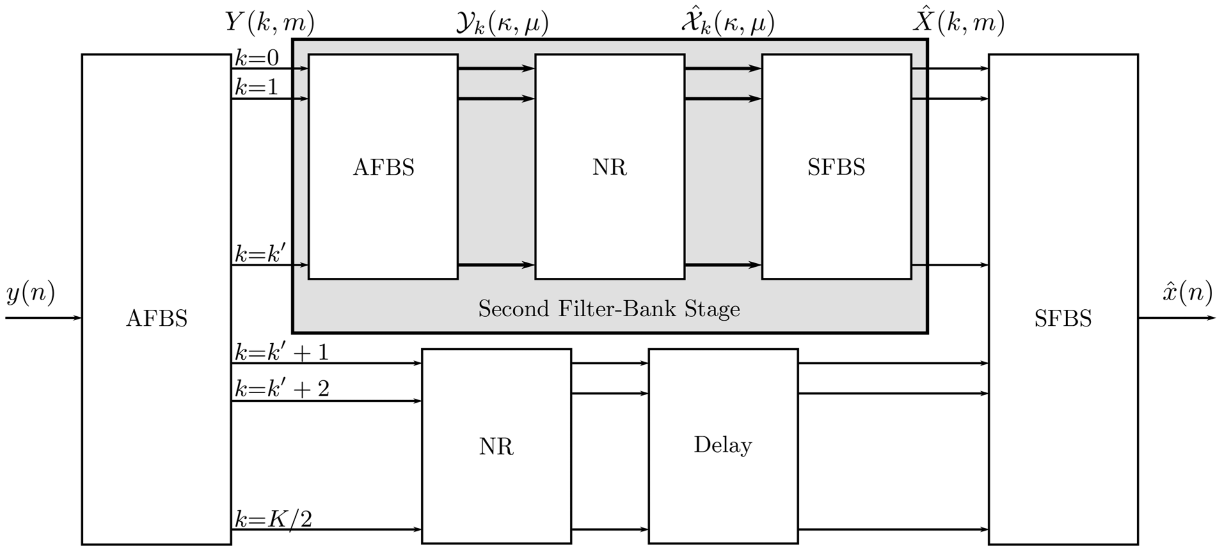

Fig. 2. Overview of the frequency decomposition and NR processing within the two-stage filter-bank system. The time-domain signal $y(n)$ is decomposed into K frequency channels by the first analysis filter-bank system (AFBS). The second filter-bank stage is only applied in lower frequency channels up to the sub-band index k', while the remaining channels are delayed w.r.t. the implementation of the second stage. Finally, the time-domain signal $\hat{x}(n)$ is the output of the last synthesis filter-bank system (SFBS).

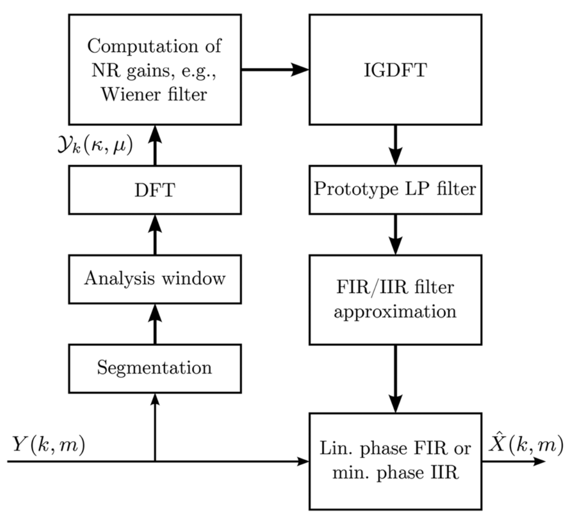

Fig. 3. Block diagram of the FBE [19]–[21]. The noisy input samples of the kth sub-band signal are arranged into segments and weighted by an analysis window. Applying the DFT to these segments generates the second frequency decomposition $\mathcal{Y}_k(\kappa,\mu)$. Then, the spectral gains of a NR algorithm like the Wiener filter are computed and transformed to the sub-band time domain using the IGDFT. The resulting coefficients are multiplied with a prototype low-pass filter and approximated by a linear-phase FIR or a minimum-phase IIR filter, which is then applied to the noisy signal $Y(k,m)$.

to 0. However, the computation of the IIR filter coefficients is demanding [21], which turns this approach almost unsuitable for an application in hearing aids. Therefore, we will mainly use the FIR implementation of the FBE in later evaluations.

B. Non-Symmetric Window Functions

In [22], the authors propose the design of non-symmetric analysis and synthesis window functions, originally to adapt the time and frequency resolution of a single-stage FBS to the different characteristics of transient and more stationary speech sounds. In the context of this contribution, we distinguish between the length of the synthesis filter $\mathcal{L}_{\mathrm{Syn}}$ which directly defines the algorithmic delay as $\mathcal{L}_{\mathrm{Syn}}-1$ sub-band samples, and the length of the analysis window, which also defines the DFT length, i.e., $\mathcal{L}_{\mathrm{Ana}}=\mathcal{K}$. Now, using the approach of [22], we can freely adjust the algorithmic delay and the frequency resolution of the second filter-bank stage, however, the choice of $\mathcal{L}_{\mathrm{Syn}}$ also defines the maximum frame advance $\mathcal{R}_{\max}=\mathcal{L}_{\mathrm{Syn}}/2$ in case of Hann-like windows and thus the computational complexity. The non-symmetric windows are designed such that their element-wise multiplication results in a Hann window function with the same length as the synthesis window. This leads to perfect reconstruction in case of no signal modification when using an appropriate overlap-add scheme for the synthesis of the sub-band signals as shown in Fig. 2.

Fig. 4 shows two examples of window function pairs. The first example on the left side shows two square-root Hann windows with $\mathcal{L}_{\mathrm{Ana}}=\mathcal{L}_{\mathrm{Syn}}$, while in the second example on the right hand, a shorter synthesis window length is used which affects the shape of the analysis window. For our filter-bank design we must require that at least the first element of the window functions shown in Fig. 4 has to be 0, which guarantees the perfect reconstruction property. The delay caused by the second filter-bank stage is now defined by the number of nonzero elements reduced by 1 in the synthesis window. Therefore, in the first example of Fig. 4, i.e. $\mathcal{L}_{\mathrm{Syn}}=15(+1)$, the delay is 14 samples, in the second example, the delay is reduced to 2 samples. Therefore, we use a unified notation of the window lengths as $\mathcal{L}_{\mathrm{Ana}}=15(+1)$ and $\mathcal{L}_{\mathrm{Syn}}=15(+1)$ or $\mathcal{L}_{\mathrm{Syn}}=3(+13)$, corresponding to the two examples in Fig. 4, where the number in parenthesis accounts for the number of elements equal to zero.

C. Low-Delay Spectral Analysis

When using the second filter-bank stage the frequency decomposition of the whole FBS is affected. The different effects of the two-stage approaches are shown in Fig. 5. The (a)–(d) show the frequency responses of the first-stage sub-band $k=2$ within the second filter-bank stage. It also shows the frequency responses of the first-stage sub-band k

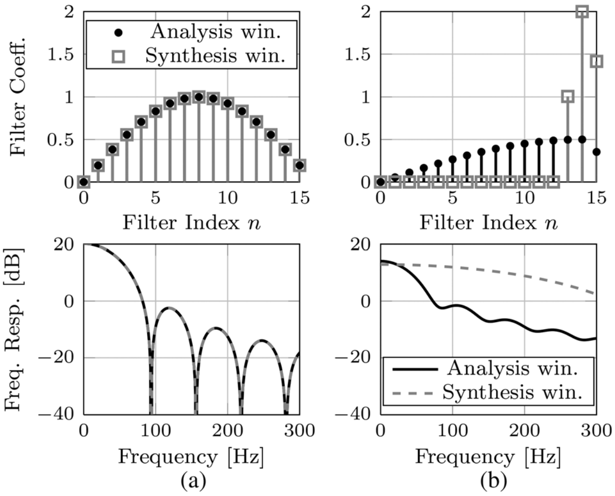

Fig. 4. The window functions defined in [22] allow different lengths of the analysis (black) and synthesis windows (gray). The figure shows two examples for $\mathcal{L}_{\mathrm{Ana}} = 15(+1)$, i.e., one element of the window is zero, the remaining 15 elements are nonzero. In case of $\mathcal{L}_{\mathrm{Syn}} = 15(+1)$, the resulting windows are both square-root Hann windows, in case of $\mathcal{L}_{\mathrm{Syn}} = 3(+13)$, the windows are non-symmetric. The plots at the bottom show the respective frequency responses. $\mathcal{L}_{\mathrm{Syn}} = 15(+1)$, $\mathcal{L}_{\mathrm{Syn}} = 3(+13)$.

$= 2$ and its close-by frequency channels. In Subfigure (a), we set $\mathcal{L}_{\mathrm{Ana}} = \mathcal{L}_{\mathrm{Syn}} = 15(+1)$ and $\mathcal{R} = 8$, which corresponds to a conventional, i.e., not a low-delay implementation. The figure also defines the indexing of the frequency channels within the second-stage, i.e., in case of $\mathcal{K} = 16$, $\kappa = -7, -6, \ldots, 7, 0^+$. The sub-bands $\kappa = 0$ and 0^+ have the same center frequency as the respective first-stage frequency channel $k = 2$, however, the sub-band $\kappa = 0^+$ has no distinct maximum and contains the least signal power. The remaining channels are arranged symmetrically around the centering sub-band $\kappa = 0$. Due to the spectral overlap of the first-stage frequency channels, the second filter-bank stage contains quite some redundancy, which can be used to reduce the computational complexity, as discussed in Section IV. When implementing the second filter-bank stage with less delay, the shape of the frequency responses changes. In case of the shorter synthesis window length $\mathcal{L}_{\mathrm{Syn}} = 3(+13)$ in Subfigure (b), the second filter-bank stage becomes less frequency selective. When using the FBE with an FIR filter length of $\mathcal{L}_{\mathrm{FIR}} = 9$, the frequency responses are even more distorted, which can be observed for $\kappa = -2$ and 4. In case of the IIR implementation of the FBE, the frequency responses clearly change their shape and are more selective than the FIR implementation.

IV. Computational Complexity Reduction

A. Application to a Reduced Subset of Frequency Channels

In general, speech signals consist of voiced and unvoiced sounds. Voiced sounds are defined by the fundamental frequency f_0, respective harmonics, and most of their energy is found at lower frequencies. Unvoiced sounds usually appear in a wider frequency range, and show fast temporal changes or noise-like characteristics. Therefore, similar to the idea proposed in [22], it seems to be reasonable to use the high temporal resolution of the first filter-bank stage for unvoiced sounds in higher frequency channels, while the second filter-bank stage is

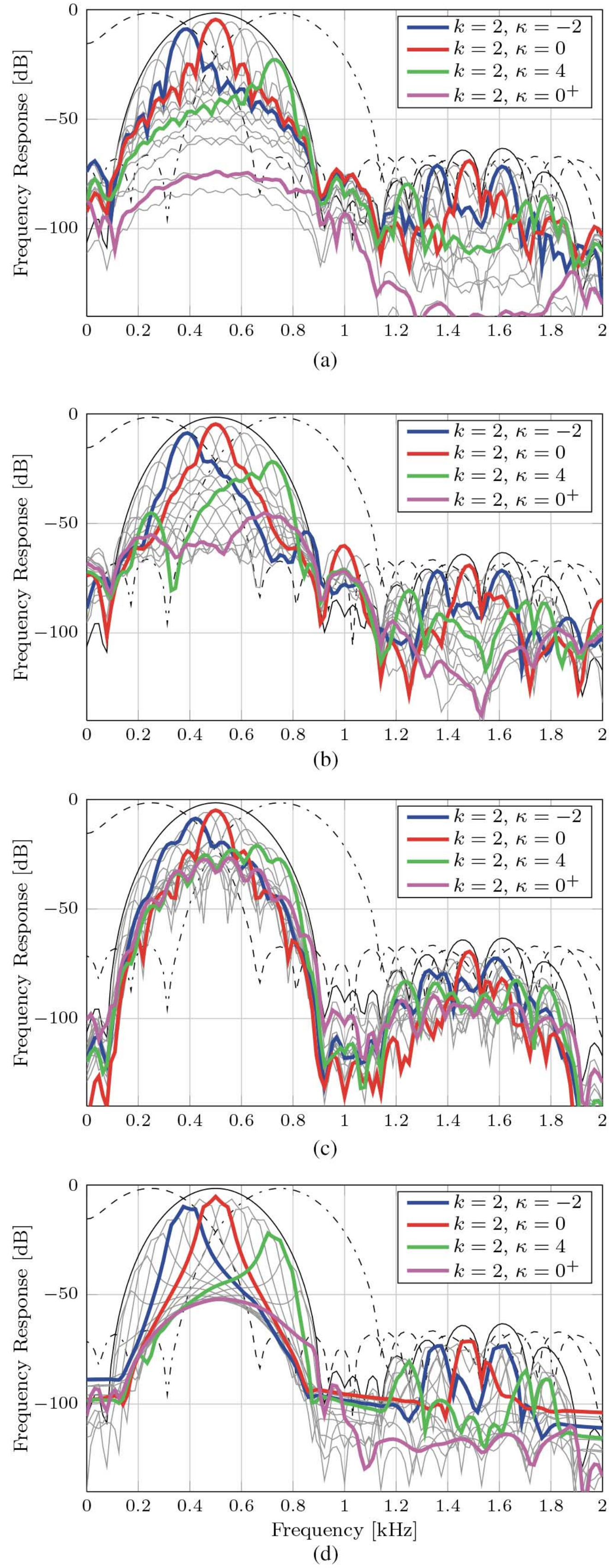

Fig. 5. Frequency responses of the two-stage filter-bank system in different implementations as defined in Table I. (a) 2Stage: Cascaded FBS, no low-delay implementation, (b) MM Ls4: Non-symmetric window functions, $\mathcal{L}_{\mathrm{Syn}} = 3(+13)$, (c) LV FIR8: FBE with $\mathcal{L}_{\mathrm{FIR}} = 9$, (d) LV IIR8: FBE with $\mathcal{L}_{\mathrm{IIR}} = 8$.

TABLE I

FILTER-BANK SETTINGS USED IN THE EVALUATION. IN ALL TWO-STAGE FBSS, k' IS SET TO 12 FOR LATER EVALUATIONS WHICH CORRESPONDS TO AN APPLICATION OF THE SECOND STAGE UP TO 3 kHz. TO INDICATE THE USE OF THE COPY PATTERN IN FIG. 7 WHICH REDUCES THE NUMBER OF COMPUTED SPECTRAL GAINS, WE EXTEND THE RESPECTIVE LABELS BY _RC, E.G. LV_FIR8_RC.

LowRes	FBS with low frequency resolution of 250 Hz as defined in Section II.
HighRes	Single-stage FBS with a higher frequency resolution of 62.5 Hz.
2Stage	Two-stage FBS as an extension of LowRes with $\mathcal{L}_{\text{Ana}} = \mathcal{L}_{\text{Syn}} = 15(+1)$ and $\mathcal{R} = 8$. The resulting frequency resolution is 62.5 Hz. The additional delay of the second stage is 14 ms.
MM_LsX	Two-stage FBS using the non-symmetric window functions of [22]. While the length of the analysis window is kept fixed at $\mathcal{L}_{\text{Ana}} = 15(+1)$, X defines the length of the synthesis window function $\mathcal{L}_{\text{Syn}} = \text{X} - 1(+16 - \text{X} + 1)$ and the decimation ratio $\mathcal{R} = \text{X}/2$. This leads to an additional delay of $(\text{X} - 2)$ ms. Example MM_Lg8: $\mathcal{L}_{\text{Syn}} = 7(+9)$, $\mathcal{R} = 4$ and 6 ms additional delay.
LV_FIRX	Two-stage FBS using the filter-bank equalizer proposed in [21]. The length of the analysis window is again set to $\mathcal{L}_{\text{Ana}} = 15(+1)$, X defines the length of the FIR filter $\mathcal{L}_{\text{FIR}} = \text{X} + 1$. The additional delay is then given as $\text{X}/2$ ms.
LV_IIRX	Two-stage FBS using the filter-bank equalizer proposed in [21]. The length of the analysis window is again set to $\mathcal{L}_{\text{Ana}} = 15(+1)$, X defines the length of the IIR filter $\mathcal{L}_{\text{IIR}} = \text{X}$. The additional delay is approximately 0.

implemented in lower frequency channels. We propose to apply the second filter-bank stage only up to a fixed frequency-bin index k' as shown in Fig. 2, while a time varying decision $k'(m)$ or even a voiced/unvoiced detection might further improve the performance. However, these latter algorithmic variants do not lie in the scope of this paper. As another benefit of fitting the temporal and spectral resolutions to the different needs of voiced and unvoiced sounds, we also reduce the computational complexity since the second filter-bank stage is implemented in a reduced number of first-stage frequency channels. In order to find an optimal value for k', we analyze its influence on the output SNR for female speech mixed with different noise signals at a seg. input SNR of 0 dB in Fig. 6. For all noise signals, we can observe that the higher frequency resolution is most efficient at lower frequencies. In case of white Gaussian noise, the best performance is achieved when setting $k' = 12$ which corresponds to a frequency of 3 kHz. After this boundary, the output SNR decreases, due to the reduced temporal resolution which affects the representation of unvoiced sounds. For stationary speech-shaped noise and traffic noise, the output SNR saturates between 1 kHz and 2 kHz, while there is no decline at higher values for k' due to less noise power at higher frequencies. Therefore, we set k' to a fixed value corresponding to 3 kHz for later evaluations. In comparison to the application of the second filter-bank stage in each first-stage frequency channel, the computational load is reduced significantly.

B. *Efficient Share of Spectral Gains*

As proposed in [18], we can also make use of the inherent correlation of the second filter-bank stage due to spectral overlap.

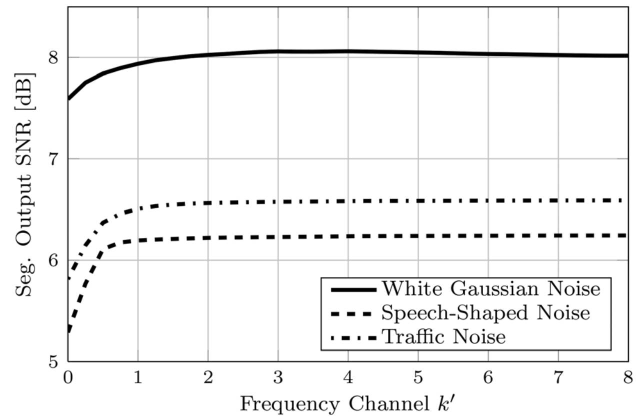

Fig. 6. When applying the second filter-bank stage to a reduced set of first-stage frequency channels, the seg. output SNR indicates an optimal value of 3 kHz which corresponds to $k' = 12$ for white Gaussian noise. In case of speech-shaped stationary noise and traffic noise, the output SNR saturates at lower frequencies. The results shown in this figure were averaged for 4 female speech signals for a fixed segmental input SNR of 0 dB (see Section V).

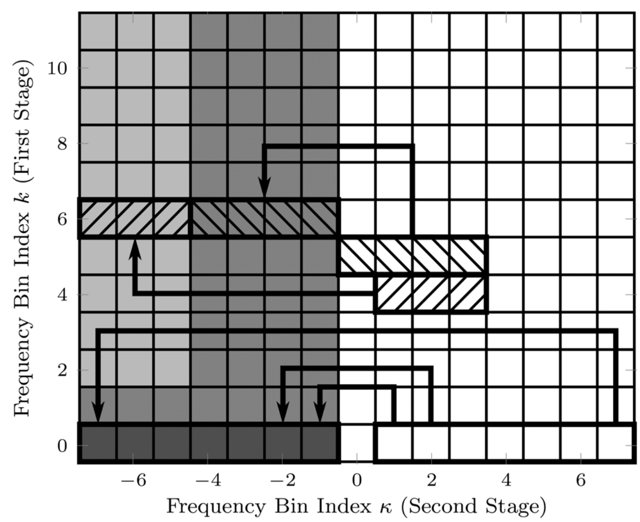

Fig. 7. The figure shows a copy pattern, designed w.r.t. the inherent correlation of the second filter-bank stage due to overlapping frequency channels. The gains within the white region are computed, and used in the remaining sub-bands on the left side (gray). The different shadings of the gray region describe the amount of correlation to those bands from which the respective gains are copied, indicated by arrows. The dark shading (■) indicates a correlation of 1, (■) defines a high correlation, and (■) describes less correlation. By using this pattern, we halve the amount of computed spectral gains in the second filter-bank stage.

Depending on the choice of all FBS parameters, it is even possible to find multiple frequency channels in the second stage that share a common center frequency, defined as

$$f_{\text{c},2}(k, \kappa) = \Delta f_1 k + \Delta f_2 \kappa, \tag{6}$$

with $\kappa = -K_2/2 + 1, \ldots, K_2/2 - 1, 0^+$. As shown in [18], the data in those channels sharing the same center frequency is highly correlated. Therefore, it is sufficient to compute the spectral NR gains only for a smaller subset of frequency channels. By exploiting the inherent redundancy of the two-stage FBS, we are therefore able to define reasonable strategies to reduce the amount of computed spectral gains. This leads to copy patterns such as shown in Fig. 7 for $\mathcal{K} = 16$, which will be used in later studies to reduce the amount of computed NR gains within the second filter-bank stage. In Fig. 7, the spectral gains of the white regions of the (k, κ) plane are computed. The remaining gains are then copied from these sub-bands, while the gains of the sub-band $\kappa = 0^+$ are set to zero. The copy pattern distinguishes between bins with maximum, higher and less correlation. The data in the first frequency channel $k = 0$ is real-valued, therefore the spectral gains of the second stage are symmetrical, i.e., $\mathcal{G}_0(\kappa, \mu) = \mathcal{G}_0(-\kappa, \mu)$. This copy process is described as one with maximum correlation, indicated by the dark gray shading. The copy processes with higher correlation are done between two adjacent frequency channels of the first stage in the pattern $\mathcal{G}_k(\kappa, \mu) \approx \mathcal{G}_{k-1}(\kappa+4, \mu)$. Finally, the copy processes with less correlation are applied across a wider spectral gap in the first stage, which leads to the pattern $\mathcal{G}_k(\kappa, \mu) \approx \mathcal{G}_{k-2}(\kappa + 8, \mu)$, indicated by the light gray shading. Even if the respective frequency responses show very different shapes, these bands share the same center frequency such that the sub-band signals are still correlated to each other.

C. Computational Complexity Assessment

To compare the computational complexity of the different approaches, we count the number of real-valued operations. Regarding a straightforward implementation of the frequency domain Wiener filter in combination with the *decision-directed* approach and a noise power estimator as defined in (5), we count 25 real-valued operations for each improved DFT coefficient. To evaluate the additional computational load of the cascaded filter-bank design, we count the real-valued operations to obtain $\mathcal{R}$ improved first-stage DFT coefficients $\hat{X}(k, m)$. When counting these operations, we do not distinguish between additions, multiplications and divisions. To process $\mathcal{R}$ DFT coefficients using the first filter-bank stage only, we need

$$\mathcal{O}_{1\ \text{Stage}} = 25\mathcal{R} \tag{7}$$

operations per frequency channel. To obtain $\mathcal{R}$ DFT coefficients from the second filter-bank stage, we need

$$\mathcal{O}_{2\text{Stage}} = \underbrace{4\mathcal{K}\log_2\mathcal{K}}_{\text{FFT and IFFT}} + \underbrace{25\mathcal{K}}_{\text{NR}} + \underbrace{2\mathcal{K}}_{\text{Windowing}} + \underbrace{2\mathcal{K}}_{\text{OLA}} \tag{8}$$

operations per first-stage frequency channel k. Again, we assume 25 real-valued operations for each improved DFT coefficient, but now we also have to include the number of computations for the FFT and IFFT (in case of $\mathcal{K} = 2^i$, $i \in \mathbb{N}^{>0}$), the windowing and the overlap-add process (OLA). When using the copy pattern in Figure 7 (or a similarly designed pattern for other DFT lengths $\mathcal{K}$), the number of operations per frequency channel is reduced to

$$\mathcal{O}_{2\text{Stage,red}} = \underbrace{4\mathcal{K}\log_2\mathcal{K}}_{\text{FFT and IFFT}} + \underbrace{25\mathcal{K}/2 + \mathcal{K}}_{\text{NR}} + \underbrace{2\mathcal{K}}_{\text{Windowing}} + \underbrace{2\mathcal{K}}_{\text{OLA}}. \tag{9}$$

Therefore, we can compute the additional computational load as

$$\mathcal{C}_{\text{full}} = \frac{\mathcal{O}_{2\text{Stage}}}{\mathcal{O}_{1\text{Stage}}} - 1 \tag{10}$$

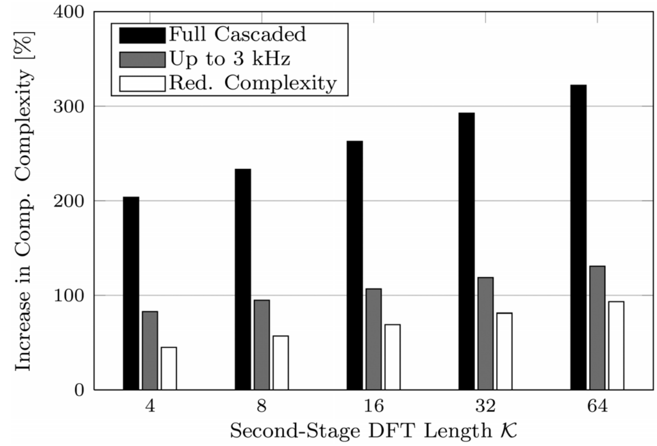

Fig. 8. Increase in computational complexity for different cascaded filter-bank implementations as a function of the DFT length $\mathcal{K}$ compared to the conventional single-stage FBS. *Full Cascaded* describes the application of the second filter-bank stage in all first-stage frequency channels, *Up to 3 kHz* is the application to a reduced set of first-stage frequency channels which is further optimized in *Red. Complexity* by using copy patterns similar to Fig. 7.

$$\mathcal{C}_{k'} = \frac{(k'+1)\mathcal{O}_{2\text{Stage}} + (K/2 - k' - 1)\mathcal{O}_{1\text{Stage}}}{\mathcal{O}_{1\text{Stage}}K/2} - 1 \tag{11}$$

$$\mathcal{C}_{k',\text{red}} = \frac{k'\mathcal{O}_{2\text{Stage,red}} + (K/2 - k')\mathcal{O}_{1\text{Stage}}}{\mathcal{O}_{1\text{Stage}}K/2} - 1. \tag{12}$$

Here, $\mathcal{C}_{\text{full}}$ describes the additional computational load when implementing the second filter-bank stage with DFT length $\mathcal{K}$ in each first-stage frequency channel, $\mathcal{C}_{k'}$ for an implementation of the second stage only in k' first-stage frequency channels, and $\mathcal{C}_{k',\text{red}}$ describes the case when we also apply efficient copy patterns. Fig. 8 shows the results for different second-stage DFT lengths $\mathcal{K}$ and critically oversampled FBSs, i.e., $\mathcal{R} = \mathcal{K}/2$. We observe that the implementation of the second stage in each first-stage sub-band (Full Cascaded, $\mathcal{C}_{\text{full}}$) increases the computational load by 200 - 300%, depending on the choice of $\mathcal{K}$. By applying the second stage only up to 3 kHz (Up to 3 kHz, $\mathcal{C}_{k'}$ with $k' = 12$), the computational extra cost is reduced to 75 - 125%, while the application of the copy patterns (Red. Complexity, $\mathcal{C}_{k',\text{red}}$) further reduce the extra cost to 40 - 90%.

In the context of the low-delay implementations, the use of the non-symmetric window functions by [22] is comparable to the results shown in Fig. 8, however, the decimation ratio $\mathcal{R}$ is usually set to smaller values, which increases the computational load. The implementation of the FBE will not be discussed in detail, however, especially the use of IIR filters comes with a high computational cost. The use of FIR filters results in very similar computational extra costs as the costs discussed in this section and shown in Fig. 8.

V. Evaluation Results

A. Objective Evaluation

To evaluate the performance of the two-stage filter-bank system in the context of single-channel noise reduction, we use parts of the TSP database [23] as speech corpus, resampled at

16 kHz. Each file has a length of 15 s and contains 6 sentences. For the objective evaluations, we distinguish between female and male speech by averaging the results for 4 speech files each. We add white Gaussian noise or speech-shaped stationary noise to these speech signals at 0 dB segmental input SNR to generate noisy speech files. For the listening experiment in Section V-B, we also mixed the same speech signals with babble and traffic noise taken from the Sound Ideas 6000 database [24]. The input and output SNR values are computed via

$$\mathrm{SNR} = \frac{10}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} \log_{10} \frac{\sum_{n=0}^{L-1} x^2(n+mR)}{\sum_{n=0}^{L-1} \left(x(n+mR) - \hat{x}(n+mR)\right)^2}, \tag{13}$$

where $\mathbb{M}$ is the set of time frames of length L that contain speech, identified in the clean speech signal $x(n)$. To quantify the amount of noise reduction and speech distortion, we also use the seg. NR and the seg. speech SNR, defined as

$$\text{Speech SNR} = \frac{10}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} \log_{10} \times \frac{\sum_{n=0}^{L-1} x^2(n+mR)}{\sum_{n=0}^{L-1} \left(x(n+mR) - \tilde{x}(n+mR)\right)^2}, \tag{14}$$

$$\mathrm{NR} = \frac{10}{|\mathbb{M}|} \sum_{m \in \mathbb{M}} \log_{10} \frac{\sum_{n=0}^{L-1} v^2(n+mR)}{\sum_{n=0}^{L-1} \tilde{v}^2(n+mR)}. \tag{15}$$

In these definitions, $\hat{x}(n) = \tilde{x}(n) + \tilde{v}(n)$ is the improved output signal, while the signals $\tilde{x}(n)$ and $\tilde{v}(n)$ represent the residual speech and noise components in the time domain after processing, respectively. Thus, the speech SNR measures the distance of the processed *clean* speech signal $\tilde{x}(n)$ to the unprocessed speech signal $x(n)$ and is therefore a measure of speech distortion. All NR algorithms are tuned to achieve the same amount of seg. NR of 11 dB. In case of the two-stage FBS, this was done when applying the second stage to all first-stage frequency channels and without using a low-delay implementation. Therefore, the seg. NR of the cascaded approaches might vary from the initial tuning. To compare the performance of the Wiener filter implemented in the different filter-bank settings, we define test cases as summarized in Table I.

Fig. 9 shows the NR results of the different implementations defined in Table I. When comparing the settings LowRes and High Res, the figure shows the effect of proper frequency resolution to remove the noise between the harmonics. At the same seg. NR of 11 dB, the HighRes FBS achieves less speech distortion (higher speech SNR) and better seg. SNR values. Furthermore, we can observe a slightly improved performance when applying the second filter-bank stage up to 3 kHz in case of 2Stage. When shortening the synthesis window length L_{Syn} in case of the non-symmetric window functions, the seg. SNR and Speech SNR are gradually decreased, while the seg. NR is slightly increased. A similar effect, but however more distinctive, can be observed for the filter-bank equalizer. Here, the shortening of the FIR filters has a stronger impact on the NR performance, while the use of the IIR filter degrades the signal quality even more in case of white Gaussian noise. For this

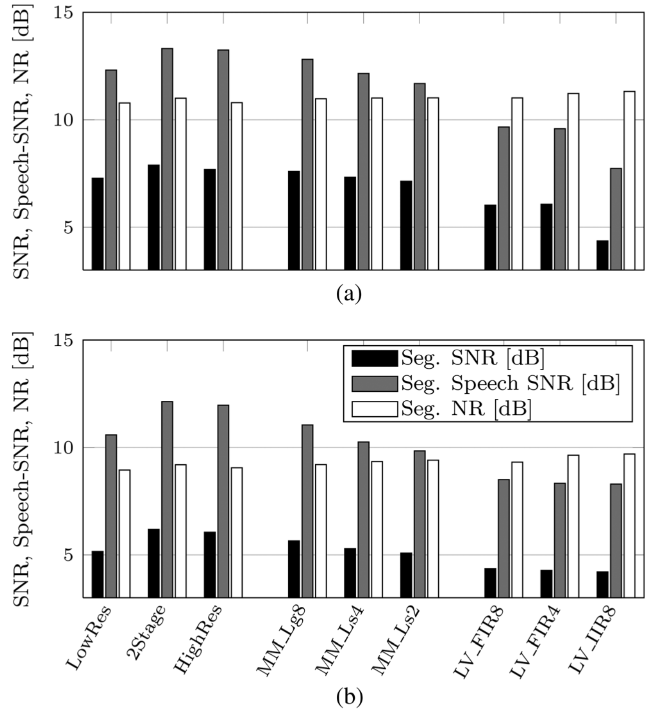

Fig. 9. The seg. SNR, seg. speech SNR and the seg. NR for female speech and two different, stationary noise signals at a seg. input SNR of 0 dB. (a) Female speech and white Gaussian noise at 0 dB seg. input SNR, (b) Female speech and stationary, speech-shaped noise at 0 dB seg. input SNR.

reason, we do not consider this implementation in our evaluation. The figure also shows the results for female speech and speech-shaped, stationary noise. In this case, the results are very similar to those observed for white Gaussian noise. However, the NR algorithms achieve less SNR gain, due to the worse SNR in lower frequency bands. We also observe that the difference between the HighRes and LowRes cases are more distinctive, since there is more noise between the harmonics of the speech signal which can be removed when using a high frequency resolution. Listening to the signals reveals that even if the objective measures are in some cases worse than the reference single-stage FBS, each of the two-stage approaches benefits of a higher frequency resolution. Therefore, the roughness of the processed signals is reduced, however, depending on the implementation of the second stage, the Wiener filter induces speech distortions which affects the objective measures.

Fig. 10 shows the results when reducing the computational complexity of the two-stage NR system by applying the copy pattern shown in Fig. 7 to reduce the amount of computed NR gains. We can observe that the exploitation of the inherent correlation of the second filter-bank stage only slightly affects the objective measures.

B. Subjective Evaluation

Since typical single-channel NR algorithms do not achieve a gain in speech intelligibility [25], we performed a listening test to evaluate signal quality improvements. We invited experienced, normally-hearing participants to the test. Since we only consider the effects of noise reduction, we can assume very similar results for hearing-impaired participants [26]. For a subjective evaluation of the proposed two-stage filter-bank

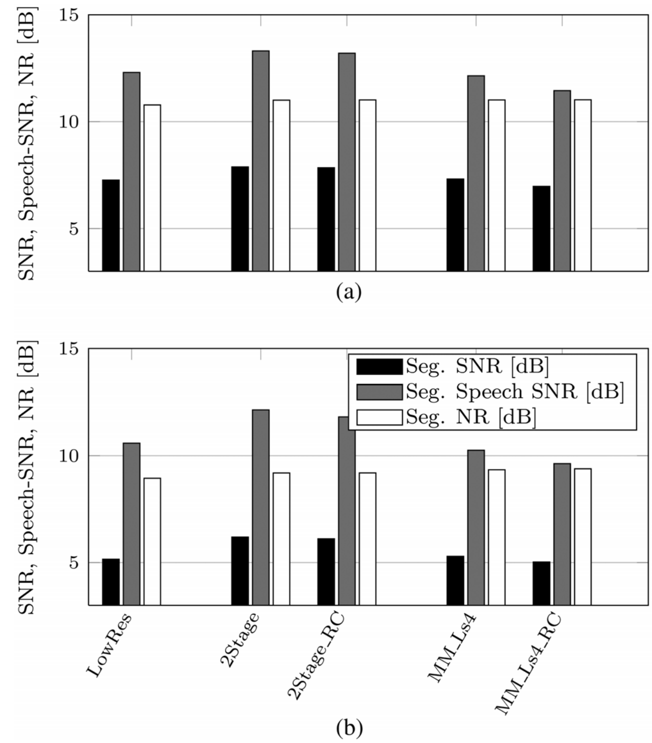

Fig. 10. The seg. SNR, seg. speech SNR and the seg. NR for female speech and two different stationary noise signals at a seg. input SNR of 0 dB. In case of the two-stage approaches, the Wiener gains were copied to reduce the computational complexity following Fig. 7. (a) Female speech and white Gaussian noise at 0 dB seg. input SNR, (b) Female speech and stationary, speech shaped noise at 0 dB seg. input SNR.

TABLE II
FILTER-BANK SYSTEMS ANALYZED IN THE LISTENING TEST

LowRes	NR performed in a single-stage FBS with low frequency resolution.
2Stage16	Two-stage FBS with $\mathcal{L}_{\text{Ana}} = 15(+1)$, $\mathcal{L}_{\text{Syn}} = 3(+13)$, and $\mathcal{R} = 2$.
2Stage8	Two-stage FBS with $\mathcal{L}_{\text{Ana}} = 7(+1)$, $\mathcal{L}_{\text{Syn}} = 1(+7)$, and $\mathcal{R} = 1$.
2Stage8_RC	Two-stage FBS with $\mathcal{L}_{\text{Ana}} = 7(+1)$, $\mathcal{L}_{\text{Syn}} = 1(+7)$, and $\mathcal{R} = 1$. Furthermore, we applied a copy pattern similar to Fig. 7 to reduce the amount of spectral NR gains.

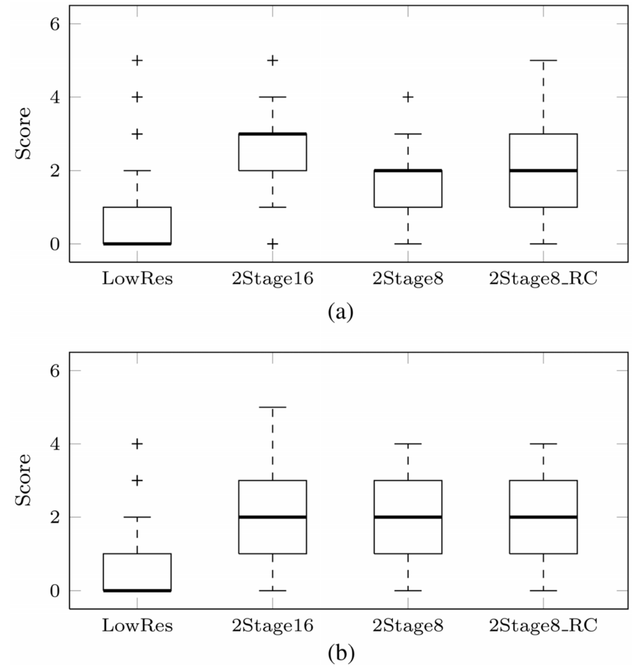

Fig. 11. Box plots of the data, measured in the listening test for noise signals at 0 dB (a) and -5 dB seg. input SNR (b) The median values are indicated by bold bars while outliers are denoted by (+).

TABLE III
SUBJECTIVE EVALUATION RESULTS AND p–VALUES, INDICATING STATISTICAL SIGNIFICANT DIFFERENCES BETWEEN THE NR APPLIED IN THE FILTER-BANK SYSTEMS (** IF $p < 0.01$, * IF $p < 0.1$)

	Score	p-Values LowRes	2Stage16	2Stage8	2Stage8_RC
LowRes	0,976		1,3 E-25 **	3,9 E-18 **	1,8 E-21 **
2Stage16	2,672			1,9 E-08 **	1,4 E-02 *
2Stage8	1,993				8,9 E-04 **
2Stage8_RC	2,375				

(a)

	Score	p-Values LowRes	2Stage16	2Stage8	2Stage8_RC
LowRes	1,023		1,2 E-21 **	7,0 E-19 **	1,7 E-18 **
2Stage16	2,508			2,4 E-02 *	3,2 E-02 *
2Stage8	2,244				9,7 E-01
2Stage8_RC	2,242				

(b)

system, we use the NR algorithm used in the previous sections with minor modifications, and extended the filter-bank system with a second stage. In order to perform a fair comparison, each of the authors was involved in tuning the NR algorithm for each filter-bank system to optimize the signal quality for the subjective evaluation. The second filter-bank stage is implemented using the non-symmetric window functions of [22]. In contrast to the implementation of the second stage based on the filter-bank equalizer, this approach not only shows better objective measures, it is also easier to be implemented in hearing aids. We compare two different settings of the second filter-bank stage, resulting in a different frequency resolution, algorithmic delay and computational complexity. Furthermore, we also evaluate the influence of the complexity reduction within this two-stage filter-bank system. The approaches compared in the listening test are summarized in Table II. The test was performed in a forced-choice paired comparison, i.e., for each noisy input signal, we asked the participants to rate the ease of listening in 6 paired comparisons. For each single comparison, the participants had the possibility to rate one signal better or much better than the other, resulting in a score of 1 or 2 points for the respective FBS. For each full comparison, i.e., 6 single comparisons for each noisy input signal, we collected the score points as the measured data. We performed the Kruskal-Wallis test as a non-parametric analysis to evaluate statistical significance [27] since the data was not normally distributed.

Table III and Fig. 11 show the results of the listening test, performed at the Institute of Communication Acoustics, Bochum. We provided a database with three speech (male, female, child) and 4 noise signals (white Gaussian noise, stationary speech-shaped noise, traffic noise and babble noise) mixed at seg. input SNR values of -5 and 0 dB. 20 normally-hearing participants in the age of 25-35 took part in the experiments, 10 for each SNR condition. The signals were played back via headphones in an acoustically and electromagnetically isolated chamber and the participants were asked to set the loudness of the playback to a convenient level. In the experiment, each participant was asked to rate the ease of listening in 72 single paired comparisons ($3 \times 4 = 12$ noisy signals, $12 \times \binom{4}{2}$ single comparisons) after a short learning phase. The results show that all FBSs with higher frequency resolution than LowRes are rated better, while the best performance is obtained by 2Stage16, due to the highest frequency resolution. The comparison of 2Stage8 and 2Stage8_RC reveals that in case of -5 dB seg. input SNR, there is no significant difference. In case of 0 dB seg. input SNR, the approach with reduced computational complexity 2Stage8_RC was even rated better than 2Stage8 with statistical significance. Due to the copy processes in 2Stage8_RC, we reduce the variance within the spectral gains which results in a smoothing effect. In terms of objective gains, we already showed in [18] that the Itakura-Saito distance also indicates an improved performance which supports the results of our listening experiment. Therefore, reducing the computational load of the two-stage filter-bank system improves the performance, at least in moderate SNR conditions. Nevertheless, it is difficult to hear distinct differences between signals processed in the cascaded FBS with and without copied NR gains.

VI. Conclusions and Outlook

We studied the implementation of a cascaded filter-bank system to improve the frequency resolution of an already existing, and well-tuned signal analysis and resynthesis scheme. We also evaluated different approaches to implement the second filter-bank stage with reduced or even with almost no algorithmic delay. Furthermore, we also analyzed means to reduce the computational complexity of a two-stage noise-reduction system which is, when designed properly, even able to improve the performance, supported by objective measures and the results of a listening experiment. In future works, we hope to further improve the noise reduction performance by adapting the number of first-stage frequency channels to the signal in terms of a voiced/unvoiced sound detection.

References

[1] H. Dillon, *Hearing Aids*, 2nd Ed. ed. Stuttgart, Germany: Thieme, 2012.

[2] H. Puder, E. Fischer, and J. Hain, "Optimized directional processing in hearing aids with integrated spatial noise reduction," in *Proc. Int. Workshop Acoust. Signal Enhance. (IWAENC)*, Aachen, Germany, Sep. 2012.

[3] V. Hamacher, U. Kornagel, T. Lotter, and H. Puder, "Binaural signal processing in hearing aids," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweilter, Eds. New York, NY, USA: Wiley, 2008, pp. 401–429.

[4] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, ser. Prentice-Hall signal processing series. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.

[5] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[6] V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," *Acta Acust. united with Acust.*, vol. 88, no. 3, pp. 433–442, May/Jun. 2002.

[7] J. C. Brown, "Calculation of a constant Q spectral transform," *J. Acoust. Soc. Amer.*, vol. 89, no. 1, pp. 425–434, Jan. 1991.

[8] A. Nagathil and R. Martin, "Optimal signal reconstruction from a constant-Q spectrum," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 349–352.

[9] T. Kurbiel, H. Göckler, and D. Alfsmann, "A novel approach to the design of oversampling low-delay complex-modulated filter bank pairs," *EURASIP J. Adv. Signal Process.*, vol. 2009, 2009, article ID 692861.

[10] D. Alfsmann, H. Göckler, and T. Kurbiel, "Filter banks for hearing aids applying subband amplification: A comparison of different specification and design approaches," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Glasgow, U.K., Aug. 2009.

[11] W. Sörgel and R. Bäuml, "Uniform polyphase filter banks for use in hearing aids: Design and constraints," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.

[12] M. Gray and R. Meredith, "Stable gain limits of instant-fit hearing aid couplings," *Hear. Rev.*, vol. 16, no. 12, pp. 16–19, 2009.

[13] J. Agnew and J. M. Thornton, "Just noticeable and objectionable group delays in digital hearing aids," *J. Amer. Acad. Audiol.*, vol. 11, no. 6, pp. 330–336, Jun. 2000.

[14] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. New York, NY, USA: Wiley, 2006.

[15] C. Breithaupt and R. Martin, "Noise reduction - statistical analysis and control of musical noise," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. New York, NY, USA: Wiley, 2008, pp. 107–133.

[16] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[17] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.

[18] A. Schasse, R. Martin, W. Sörgel, T. Pilgrim, and H. Puder, "Efficient implementation of single-channel noise reduction for hearing aids using a cascaded filter-bank," in *Proc. ITG Symp. Speech Commun.*, Braunschweig, Germany, Sep. 2012, pp. 287–290.

[19] P. Vary, "An adaptive filter bank equalizer for speech enhancement," *Signal Process.*, vol. 86, no. 6, pp. 1206–1214, Jun. 2006.

[20] H. W. Löllmann and P. Vary, "Generalized filter-bank equalizer for noise reduction with reduced signal delay," in *Proc. Eur. Conf. Speech Commun. Technol. (Interspeech)*, Lisbon, Portugal, Sep. 2005, pp. 2105–2108.

[21] H. W. Löllmann and P. Vary, "Efficient non-uniform filter-bank equalizer," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Antalya, Turkey, Sep. 2005.

[22] D. Mauler and R. Martin, "Improved reproduction of stops in noise reduction systems with adaptive windows and nonstationarity detection," *EURASIP J. Adv. Signal Process.*, vol. 2009, 2009, article ID 469480.

[23] Telecommunications & Signal Processing Laboratory, McGill Univ.. Montreal, Canada [Online]. Available: http://www-mmsp.ece.mcgill.ca/Documents/Data/

[24] "Sound ideas," [Online]. Available: http://www.sound-ideas.com/sound-effects/series-6000-sound-effects-library.html

[25] P. C. Loizou and G. Kim, "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 1, pp. 47–56, Jan. 2011.

[26] K. H. Arehart, J. M. Kates, M. C. Anderson, and L. O. Harvey, Jr., "Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Amer.*, vol. 122, no. 2, pp. 1150–1164, 2007.

[27] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*, 2nd Ed. ed. New York, NY, USA: Wiley-Interscience, Jan. 1999.

Alexander Schasse studied electrical engineering and information technology at Ruhr-Universität Bochum, Bochum, Germany, where he received the Dipl.-Ing. degree in 2008. Currently, he is pursuing the Dr.-Ing. degree from the Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany. His main research interests are single-channel speech enhancement and noise reduction. In this context, he also works on low-delay implementations for hearing instruments. Other research interests are multichannel speech processing, such as beamforming and acoustic sound source localization via Doppler processing.

Rainer Martin (S'86–M'90–SM'01–F'11) received the M.S.E.E. degree from the Georgia Institute of Technology, Atlanta, in 1989 and the Dipl.-Ing. and Dr.-Ing. degrees from RWTH Aachen University, Aachen, Germany, in 1988 and 1996, respectively. From 1996 to 2002, he was a Senior Research Engineer with the Institute of Communication Systems and Data Processing, RWTH Aachen University. From April 1998 to March 1999, he was a Technology Consultant at the AT&T Speech and Image Processing Services Research Lab (Shannon Labs), Florham Park, NJ. From April 2002 until October 2003, he was a Professor of Digital Signal Processing at the Technische Universität Braunschweig, Braunschweig, Germany.

Since October 2003, he has been a Professor of Information Technology and Communication Acoustics at Ruhr-Universität Bochum, Bochum, Germany, and from October 2007 to September 2009 Dean of the Electrical Engineering and Information Sciences Department. His research interests are signal processing for voice communication systems, hearing instruments, and human–machine interfaces. He is coauthor with P. Vary of *Digital Speech Transmission–Enhancement, Coding and Error Concealment* (Wiley, 2006) and coeditor with U. Heute and C. Antweiler of *Advances in Digital Speech Transmission* (Wiley, 2008). He served as an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and is a member of the Speech and Language Processing Technical Committee of the IEEE Signal Processing Society.

Timo Gerkmann studied electrical engineering at the universities of Bremen and Bochum, Germany. He received his Dipl.-Ing. degree in 2004 and his Dr.-Ing. degree in 2010 both at the Institute of Communication Acoustics (IKA) at the Ruhr-Universität Bochum, Bochum, Germany. In 2005, he spent six months with Siemens Corporate Research in Princeton, NJ, USA. During 2010 to 2011 he was a Postdoctoral Researcher at the Sound and Image Processing Lab at the Royal Institute of Technology (KTH), Stockholm, Sweden. Since 2011 he has been a Professor for Speech Signal Processing at the Universität Oldenburg, Oldenburg, Germany. His main research interests are digital speech and audio processing, including speech enhancement, modeling of speech signals, and hearing devices.

Wolfgang Sörgel received the M.S. degree from Clemson University, Clemson, SC, in 1995 and the Dipl.-Ing. from Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany in 1996, both in electrical engineering. From 1996–1998, he worked on medical image processing and classification at the clinic of oral and maxillofacial surgery of Friedrich-Alexander-Universität Erlangen-Nürnberg, from 1998–2004 on video communication at the chair of multimedia communications and signal processing of Friedrich-Alexander-Universität Erlangen-Nürnberg. Since 2004, he has been with Siemens Audiologische Technik (SAT), Erlangen, Germany, where his main work responsibilities within the audio signal processing group include single channel noise reduction, filter banks and ensuring applicability of algorithms for actual implementation in hearing aids.

Thomas Pilgrim received the diploma degrees in electrical engineering from the University of Applied Sciences, Jena, Germany, in 2007. In September 2007 he joined the Signal Processing Group at Siemens Audiologische Technik (SAT), Erlangen, Germany. Since then he worked on several audio signal processing topics for hearing instruments, such as single-channel noise reduction, frequency compression, and recovery of spatial perception.

Henning Puder (M'99–SM'11) received the diploma degrees in electrical engineering from Darmstadt University of Technology, Darmstadt, Germany, and Grande École Supérieure d'électricité (Supélec), Paris, France, in 1997 and the Dr.-Ing. degree (with distinction) in electrical engineering from Darmstadt University of Technology in 2003. Since 2002, he has been with Siemens Audiologische Technik (SAT), Erlangen, Germany, where his main research area is within audio signal processing applications for hearing aids, such as noise reduction, beamforming, and feedback cancellation. In 2006, he was appointed head of the Signal Processing Group within SAT's R&D Department. Since June 2011, he is additionally a part-time Professor ("industrial professorship") at Darmstadt University of Technology where he is responsible for the research topic "adaptive systems for processing of speech and audio signals." Dr. Puder was honored with the German ITG Johann Philipp Reis award in 2003 for his thesis dealing with the topic of Freisprecheinrichtungen für Kraftfahrzeuge (hands-free equipment for cars).