REFINEMENT METHODS AS PREPROCESSING OR POSTPROCESSING FOR SHORT-TERM SPECTRA ANALYSIS

Mohammed Krini

Aschaffenburg University of Applied Sciences
Faculty of Engineering Sciences
Signal Processing
mohammed.krini@h-ab.de

Abstract:

In this contribution a generalized spectral refinement method of speech signals will be presented which can either be applied at a preprocessing stage before or at a postprocessing stage after a conventional frequency analysis. The refinement approach is based on a linear combination of weighted and shifted input speech signal segments. By applying the refinement method - either in time or in the subband domain - a refined version of the original short-term spectrum is obtained. Furthermore, the relationship between a polyphase filterbank and a spectral refinement technique will be considered. It will be shown, that the proposed refinement method applied as a preprocessor of conventional DFT-based analysis filterbanks is an extension of polyphase-based filterbank structures. The proposed spectral refinement technique is also very efficient and requires only few multiplications and additions. Experimental results based on the proposed refinement methods have shown that the robustness of pitch frequency estimation can be enhanced considerably.

1 Introduction

The speech signals are often exposed to noisy environments such as engine and wind noise in a car. Speech enhancement algorithms are utilized to attenuate undesired signal components while keeping the desired speech signal as natural as possible. For computational complexity reasons speech enhancement algorithms, e.g. noise suppression or echo cancellation for hands-free telephony, are often realized in the subband domain [4]. Therefore, the noisy speech signal is first divided into overlapping segments and an appropriate window function is applied to each segment. Afterwards each windowed segment is transformed into the frequency domain using a DFT. The resulting subband signals are utilized to estimate the power spectral density of the background noise or the fundamental frequency (pitch frequency) needed for speech enhancement algorithms.

Due to the windowing of input frames a signicant frequency overlap of adjacent DFT subband signals arises. This has the negative effect that trajectories of pitch frequencies cannot be easily separated which is often important for speech enhancement schemes that involve pitch frequency estimation. Computational complexity can be reduced considerably using filterbanks with large sub-sampling factors. But aliasing components appear which remarkably degrade the convergence behavior of echo cancellation schemes. In order to reduce the spectral overlap and/or the aliasing effects, the DFT size might be increased. However, the overall delay produced by the filterbank in the signal path will be higher. Sometimes the delay restrictions for different applications like hands-free telephony or in-car communication can not be fulfilled anymore [3]. To overcome these drawbacks, the so-called spectral renement method – applied as preprocessor before or as a postprocessor after a conventional frequency analysis – can be utilized.

2 DFT-Modulated Analysis Filterbank

Speech enhancement algorithms aim to reduce undesired noise components while preserving the undisturbed speech signal. The time domain microphone signal, $y(t)$, is represented by $y(t) = s(t) + b(t)$, where $s(t)$ and $b(t)$ are speech and noise respectively. In order to separate the desired and the undesired signal components, the microphone signal is usually first segmented into overlapping blocks of appropriate size:

$$\boldsymbol{y}(n) = \left[y(nR), ..., y(nR-N+1)\right]^{\mathrm{T}} = \left[y_0(n), ..., y_{N-1}(n)\right]^{\mathrm{T}}, \tag{1}$$

where R denotes the subsampling factor, n the frame index, and N the block length. Applying a window function

$$\boldsymbol{h} = \left[h_0, ..., h_{N-1}\right]^{\mathrm{T}}, \tag{2}$$

of length N to the input vector $\boldsymbol{y}(n)$ and computing a DFT to the weighted input vector results in N frequency supporting points of the short-term spectrum (STS) of the current frame:

$$Y\left(e^{j\Omega_\mu}, n\right) = \sum_{k=0}^{N-1} y(nR-k)\, h_k e^{-j\Omega_\mu k}. \tag{3}$$

The discrete frequencies Ω_μ are equidistantly distributed over the normalized frequency range: $\Omega_\mu = 2\pi\mu/N$, with $\mu \in [0, N-1]$.

Using the notation of matrix-vector-multiplication the STS can also be described as follows:

$$\boldsymbol{Y}\left(e^{j\Omega}, n\right) = \boldsymbol{D}_N \boldsymbol{H} \boldsymbol{y}(n), \tag{4}$$

The quantity $\boldsymbol{D}_N$ of order N is a DFT-matrix and $\boldsymbol{H}$ is a diagonal matrix containing the window coefficients.

3 Polyphase-based Analysis Filterbank

DFT-modulated filterbanks can be extended to so-called *non-critically subsampled polyphase filterbanks* [2]. This means that the length N_{lp} of the window functions $\tilde{h}_k$ (also called prototype lowpass filters) is allowed to be larger than the amount of subbands (determined by the DFT size N). This leads to much lower aliasing components without increasing the computational complexity significantly. Depending on the length N_{lp} of the filters $\tilde{h}_m$, frameshifts of size R close to N can be selected. Typical design procedures (e.g. [8]) can achieve a frameshift $R \approx 3/4N$. However, this comes with filter orders of about $N_{\mathrm{lp}} = 6 \ldots 8N$.

Nearly all considerations made until now are still valid if polyphase filterbanks instead of DFT-modulated filterbanks are utilized except Eq. 3. The necessary polyphase extensions for the analysis part lead to:

$$\widetilde{Y}\left(e^{j\Omega_\mu}, n\right) = \sum_{m=0}^{N-1} e^{-j\frac{2\pi}{N}\mu m} \sum_{p=0}^{N_{\mathrm{lp}}/N-1} y(nR-pN-m)\, \tilde{h}_{m+pN}. \tag{5}$$

For the sake of simplicity and in order to show the relationship between polyphase analysis filterbank and a filterbak that incooperates specral refinement the short-term spectrum $\widetilde{Y}\left(e^{j\Omega_\mu}, n\right)$ will be rewritten in matrix-vector notation. For this purpose, a vector $\boldsymbol{y}_{\mathrm{Block}}(n)$ is defined which consists of the current input signal frame as well as of the $M-1$ delayed ones each of lower order N:

$$\boldsymbol{y}_{\mathrm{Block}}(n) = \left[\boldsymbol{y}^{\mathrm{T}}(n), ..., \boldsymbol{y}^{\mathrm{T}}(n-M+1)\right]^{\mathrm{T}}. \tag{6}$$

The input vectors are subsequently multiplied by a block-diagonal window matrix:

$$\boldsymbol{H}_{\mathrm{Block}} = \begin{bmatrix} \boldsymbol{H}^{(0)} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}^{(1)} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{H}^{(M-1)} \end{bmatrix}, \tag{7}$$

where $\boldsymbol{H}^{(m)}$, $m \in [0, M-1]$, denotes a diagonal weighting matrix which consists of window function coefficients for the m-th input signal vector as follows:

$$\boldsymbol{H}^{(m)} = \begin{bmatrix} h_0^{(m)} & 0 & \dots & 0 \\ 0 & h_1^{(m)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h_{N-1}^{(m)} \end{bmatrix} . \tag{8}$$

Given the prototype filter which is usually much larger than the number of subbands, the M window function vectors of lower order N can therefore be set as follows[1]:

$$\boldsymbol{h}^{(0)} = \left[\tilde{h}_0^{(0)}, ..., \tilde{h}_{N-1}^{(0)}\right]^{\mathrm{T}}, \tag{9}$$

$$\boldsymbol{h}^{(p)} = \left[0, ..., 0, \tilde{h}_{N+(p-1)R}^{(p)}, ... \tilde{h}_{N+pR-1}^{(p)}\right]^{\mathrm{T}}, \quad p \in [1, M-1] \tag{10}$$

Therefore the prototype filter can also be described by a total sum of appropriately chosen low-order window filter coefficients according to:

$$\widetilde{h}_{mR+i} = \sum_{k=0}^{M-1} h_{i+kR}^{(m-k)} . \tag{11}$$

Finally the higher resolution STS $\widetilde{Y}\left(e^{j\Omega_\mu}, n\right)$ with N subbands computed by the polyphase filterbank can be determined in matrix-vector notation by:

$$\widetilde{\boldsymbol{Y}}\left(e^{j\Omega}, n\right) = \boldsymbol{D}\,\boldsymbol{I}_{\text{Block}}\,\boldsymbol{H}_{\text{Block}}\,\boldsymbol{y}_{\text{Block}}(n) \tag{12}$$

The *block-unit matrix* $\boldsymbol{I}_{\text{Block}} = [\text{diag}\{\mathbf{1}\}, \text{diag}\{\mathbf{1}\}, ...]$ comprises M identity matrices each of size N. Fig. 3 shows the structure of a polyphase analysis filterbank.

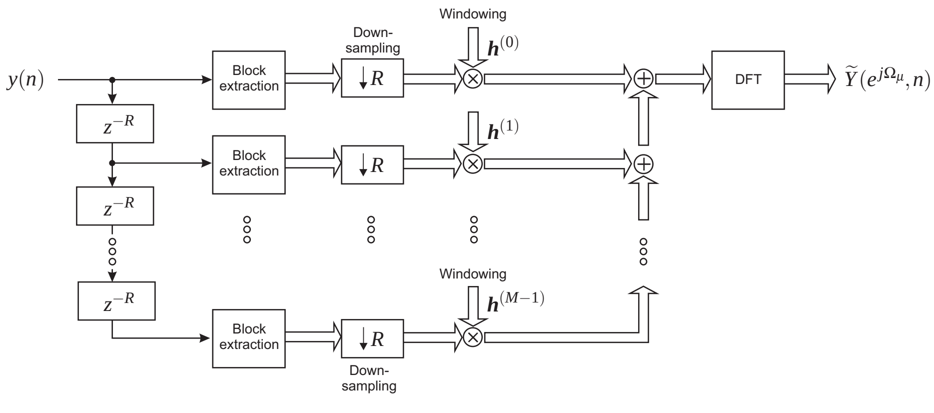

Figure 1 - Structure of a polyphase analysis filterbank.

4 Revisiting Spectral Refinement Method

The SR method determines a refined STS, $\widetilde{\boldsymbol{Y}}\left(e^{j\widetilde{\Omega}}, n\right) = [\widetilde{Y}(e^{j\widetilde{\Omega}_0}), ..., \widetilde{Y}(e^{j\widetilde{\Omega}_{\widetilde{N}-1}})]$ with $\widetilde{N} = k_0 N$ and $k_0 \in \{2, 3, ...\}$, by using the current spectrum and a number of time-delayed spectra $\boldsymbol{Y}\left(e^{j\Omega}, n-m\right)$ of lower

[1] It is assumed that the prototype filter length can also be described by $N_{\text{lp}} = N + (M-1)R$.

order N without the need for an additional DFT of higher order $\widetilde{N}$:

$$\boldsymbol{S}\begin{bmatrix} \boldsymbol{Y}\big(e^{j\Omega},n\big) \\ \vdots \\ \boldsymbol{Y}\big(e^{j\Omega},n-(M-1)\big) \end{bmatrix} = \widetilde{\boldsymbol{Y}}\big(e^{j\widetilde{\Omega}},n\big)\,. \tag{13}$$

The matrix $\boldsymbol{S}$ refers to the refinement matrix with a dimension of $\widetilde{N}\,\mathrm{x}\,NM$, where M is the amount of input spectra, each shifted by a frameshift of R samples. It is assumed that the lower order short-term spectra are already available. They might be used, e.g., to estimate the noise power for speech enhancement within a hands-free system. However, in some situations it is desired to determine a higher resolution STS in order to enhance feature extraction schemes such as pitch frequency estimation. For that purpose a linear combination of the lower order short-term spectra as stated in Eq. 13 can be applied. The derivation of the spectral refinement matrix $\boldsymbol{S}$ was firstly presented in [6] and will be shortly explained in the following.

Before calculating the refinement matrix $\boldsymbol{S}$ a constraint for the higher resolution STS is introduced, where M window functions of lower order N are appropriately weighted and subsequently adjacent window functions are shifted by the chosen subsampling factor r. The so-obtained modified window functions were summed up to obtain a desired higher order window function $\widetilde{h}_l$, $l \in [0,\widetilde{N}-1]$. Therefore the higher order window function can be described by a weighted sum of appropriately chosen low-order window filter coefficients:

$$\widetilde{h}_{mR+i} = \sum_{k=0}^{M-1} a_{i+kR}^{(m-k)}\, h_{i+kR}^{(m-k)} \tag{14}$$

with $m \in [0,M-1]$ and $i \in [0,R-1]$. Consequently, the window function $\widetilde{\boldsymbol{h}}$ consists of a weighted sum of shifted window functions $\boldsymbol{h}^{(m)}$. The coefficients $a_k^{(m)}$, $k \in [0,N-1]$, can be designed in such a way that a set of low-order window functions $h_k^{(m)}$ are transformed into a desired window function of higher order $\tilde{h}_l$. The resulting order of the window function $\widetilde{\boldsymbol{h}}$ from Eq. 14 is given by:

$$\widetilde{N} \;=\; N + r\,(M-1)\,. \tag{15}$$

In the upper part of Fig. 2 an example of weighted and shifted Hann-windows each of lower order (dashed lines with $N = 256$, $M = 5$, $R = 64$) as well as the resulting window function of higher order (solid line with $\widetilde{N} = 512$) is shown. The coefficients used for weighting the window functions have been chosen as follows: $a_k^{(0)} = a_k^{(M-1)} = 0.28\,K_0$, $a_k^{(1)} = a_k^{(M-2)} = 0.72\,K_0$ and $a_k^{((M-1)/2)} = K_0$. As normalization constant $K_0 = 1/3$ has been applied. In the lower part of Fig. 2 the corresponding analyses of the short-term spectra are depicted. By comparing the results one can see that the main-lobe width as well as the side-lobe amplitudes are reduced when using the weighted sum of shifted window functions $\boldsymbol{h}$. Based on the expression from Eq. 13 and the constraint from Eq. 14 a solution for the SR matrix is given by:

$$S_{i,mN+l} = \frac{1}{N}\, e^{-j\frac{2\pi}{N} imR} \sum_{k=0}^{N-1} a_k^{(m)}\, e^{-j2\pi\left(\frac{i}{\widetilde{N}}-\frac{l}{N}\right)k}\,. \tag{16}$$

Once the general solution for the spectral refinement matrix is formulated, a simplification and approximation of the matrix can be performed [6].

5 Temporal Refinement of Short-term Spectra

As described in the previous section, a higher resolution STS is computed using a weighted sum of M low order short-term spectra. The SR is applied at the output of a DFT-modulated analysis filterbank as a postprocessor by means of FIR filters [6]. In this section it will be shown that the SR can also be applied as a preprocessor of a conventional analysis filterbank and the resulting structure looks similar

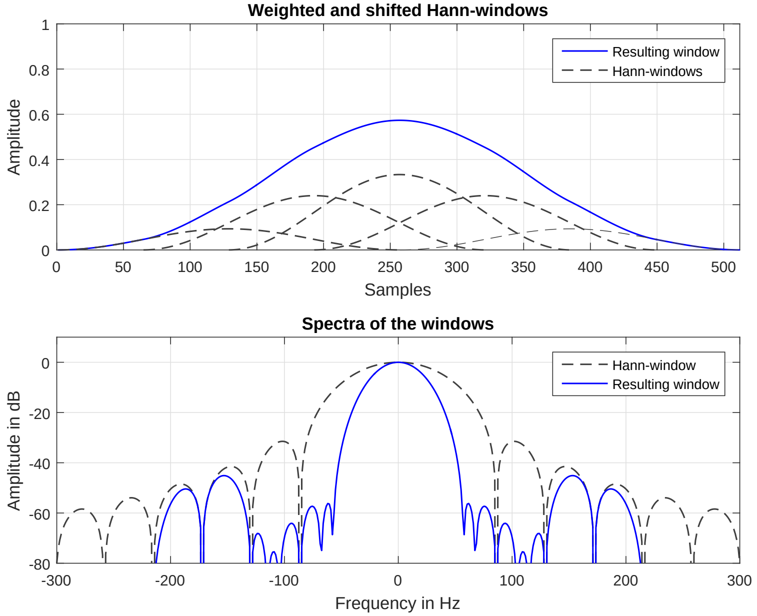

Figure 2 - Upper part shows the weighted and shifted Hann-windows and the resulting window function, lower part depicts the corresponding spectra.

to polyphase-based filterbanks. Considering Eqn. 13 and 16 the refined spectrum can be rewritten in the following way:

$$
\begin{aligned}
\widetilde{Y}\left(e^{j\Omega_i}, n\right) &= \sum_{m=0}^{M-1} \sum_{l=0}^{N-1} S_{i,mN+l}\, Y\left(e^{j\Omega_l}, n-m\right) && (17)\\
&= \sum_{m=0}^{M-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} a_k^{(m)}\, Y\left(e^{j\Omega_l}, n-m\right) e^{j\frac{2\pi}{N}kl}\, e^{-j\frac{2\pi}{N}(mR+k)i} && (18)\\
&= \sum_{m=0}^{M-1} \left(\sum_{k=0}^{N-1} a_k^{(m)} y\big((n-m)R-k\big)\, h_k\, e^{-j\frac{2\pi}{N}ki} \right) e^{-j\frac{2\pi}{N}mRi} && (19)\\
&= \sum_{m=0}^{M-1} \left(\sum_{k=0}^{N-1} a_k^{(m)} y_k(n-m)\, h_k^m\, e^{-j\frac{2\pi}{N}ki} \right) e^{-j\frac{2\pi}{N}mRi}\,. && (20)
\end{aligned}
$$

The inverse transformation of Eq. 20 into the time-domain leads to a sum of M circular shifted and weighted input frames:

$$
\widetilde{y}_k(n) = \sum_{m=0}^{M-1} a_l^{(m)}\, y_l(n-m)\, h_l^{(m)}\,, \tag{21}
$$

with $l = (k - mR) \bmod M$. The abbreviation "mod" describes a modulo operation. Eq. 21 can be rewritten in matrix-vector notation as follows:

$$
\widetilde{\boldsymbol{y}}(n) = \boldsymbol{A}_{\text{Block}}\, \boldsymbol{H}_{\text{Block}}\, \boldsymbol{y}_{\text{Block}}(n)\,. \tag{22}
$$

The block diagonal weighting matrix $\boldsymbol{A}_{\text{Block}}$ of size $NxMN$ consists of appropriate weights $a_l^{(m)}$ from Eq. 21 and it has the following structure:

$$
\boldsymbol{A}_{\text{Block}} = \left[\boldsymbol{A}^{(0)}, \boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(M-1)}\right]. \tag{23}
$$

Each element of $\boldsymbol{A}_{\text{Block}}$ characterizes a diagonal weighting matrix defined by:

$$\boldsymbol{A}^{(m)} = \begin{bmatrix} a_0^{(m)} & 0 & \dots & 0 \\ 0 & a_1^{(m)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{N-1}^{(m)} \end{bmatrix} \boldsymbol{I}^{(m)}, \tag{24}$$

whereas $\boldsymbol{I}^{(m)}$ denotes a shifted version of an identity matrix $\boldsymbol{I}^{(m)} = diag\{1\}$ of size N given by:

$$I_{i,j}^{(m)} = I_{i,\text{mod}(j+mR,N)}^{(0)}. \tag{25}$$

Fig. 5 shows the realization of the spectral refinement as a preprocessing stage before a standard frequency analysis. The resulting structure looks similar to the polyphase filterbank introduced in Fig. 3 except that the window fuctions $h_k^{(m)}$ are additionally combined with appropriately chosen weighting coefficients $a_k^{(m)}$. In order to find appropriate weights, first a vector $\boldsymbol{a}^{(m)}$ is defined which gathers all

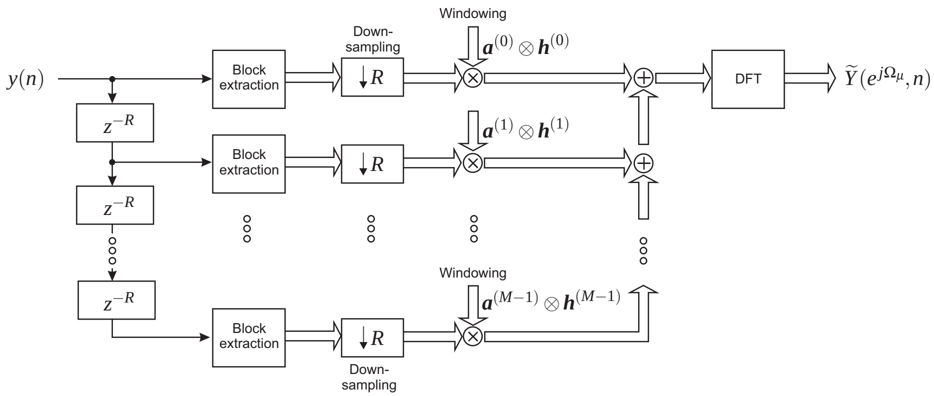

Figure 3 - Realization of spectral refinement as a preprocessing stage before a standard frequency analysis. Symbol $\otimes$ denotes element-wise multiplication.

weighting coefficients $a_k^{(m)}$ applied to the m-th window function $h_k^{(m)}$ of low order:

$$\boldsymbol{a}^{(m)} = \left[a_0^{(m)}, ..., a_{N-1}^{(m)}\right]^{\mathrm{T}}. \tag{26}$$

Furthermore an extended vector $\boldsymbol{a}_{\text{block}}$ of size $MN \times 1$ is defined which consists of all weighting coefficients for weighting M window functions of lower order N:

$$\boldsymbol{a}_{\text{block}} = \left[\left(\boldsymbol{a}^{(0)}\right)^{\mathrm{T}}, ..., \left(\boldsymbol{a}^{(M-1)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{27}$$

As already described in the previous section, a higher order window function can be described by a weighted sum of appropriately chosen low-order window filter coefficients. Therefore, the coefficients $a_l^{(m)}$ can be designed in such a way that a set of low-order window functions are transformed into a desired window function of higher order according to:

$$\min \left|\left|\boldsymbol{a}_{\text{block}}\right|\right|_2^2 \quad \text{s. t.} \quad \tilde{\boldsymbol{h}} = \widetilde{\boldsymbol{H}}_{\text{block}}\, \boldsymbol{a}_{\text{block}}, \tag{28}$$

whereas $||.||_2^2$ describes the $L2$-norm. The block diagonal matrix $\widetilde{\boldsymbol{H}}_{\text{block}}$ with a dimension of $N + (M - 1)R \times MN$ consists of M window submatrices according to:

$$\widetilde{\boldsymbol{H}}_{\text{block}} = \left[\widetilde{\boldsymbol{H}}^{(0)}, ..., \widetilde{\boldsymbol{H}}^{(M-1)}\right]. \tag{29}$$

The first element matrix $\widetilde{\boldsymbol{H}}^{(0)}$ adds $(M-1)R\text{x}N$ zero values below the diagonal window matrix $\boldsymbol{H}^{(0)}$, whereas the remaining matrices $\widetilde{\boldsymbol{H}}^{(1)}$,$\widetilde{\boldsymbol{H}}^{(2)}$, etc., represent cyclic shifts of $\widetilde{\boldsymbol{H}}^{(0)}$. This means that equal column indices of adjacent submatrices are rotated by R elements. Thus, the first and the last element matrices are defined according to:

$$\widetilde{\boldsymbol{H}}^{(0)} = \begin{bmatrix} \boldsymbol{H}^{(0)} \\ \boldsymbol{0}^{((M-1))R\text{x}N} \end{bmatrix} = \begin{bmatrix} h_0^{(0)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & h_{N-1}^{(0)} \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}, \qquad \widetilde{\boldsymbol{H}}^{(M-1)} = \begin{bmatrix} \boldsymbol{0}^{(R\text{x}N)} \\ \boldsymbol{0}^{(R\text{x}N)} \\ \vdots \\ \boldsymbol{H}^{(M-1)} \end{bmatrix}. \tag{30}$$

A weighting vector $\boldsymbol{a}_{\text{block}}$ has to be found that satisfies Eq. 28, where the $L2$-norm of the weighting coefficients is minimized. The solution of the problem is given by:

$$\boldsymbol{a}_{\text{block}} = \widetilde{\boldsymbol{H}}^{\dagger}\widetilde{\boldsymbol{h}}, \tag{31}$$

where $\widetilde{\boldsymbol{H}}^{\dagger}$ characterises the Moore-Penrose inverse which is defined as:

$$\widetilde{\boldsymbol{H}}^{\dagger} = \widetilde{\boldsymbol{H}}_{\text{block}}^{\text{T}} \left(\widetilde{\boldsymbol{H}}_{\text{block}} \widetilde{\boldsymbol{H}}_{\text{block}}^{\text{T}} \right)^{-1}. \tag{32}$$

6 Simulation Results

The refinement method can be applied to a wide variety of audio and speech signal processing applications. It is assumed that a lower order DFT is already available and is used for spectrum analysis and synthesis within communication systems. However, in some situations it is desired to determine a higher resolution STS in order to enhance feature extraction schemes such as pitch frequency or background noise estimation. For that purpose the refinement method can either be applied at a preprocessing stage before or at a postprocessing stage after a standard DFT of lower order. In the following the application of spectral refinement at a preprocessing stage before frequency analysis for enhanced fundamental frequency estimation will be considered. In the literature different algorithms for estimating the fundamental frequency of speech signals exist [1], [7]. For estimating the pitch frequency a method based on the autocorrelation has been implemented. The noisy speech signal $y(n)$ is first divided into overlapping blocks of appropriate size and a window function is applied. The proposed spectral refinement method applied to the input signal before applying the FFT according to Fig. 5 is utilized. Afterwards the short-term power spectral density (PSD) is estimated. In order to compute the autocorrelation function (ACF) an IFFT is applied to a normalized version of the PSD. A maximum search over the ACF in a certain

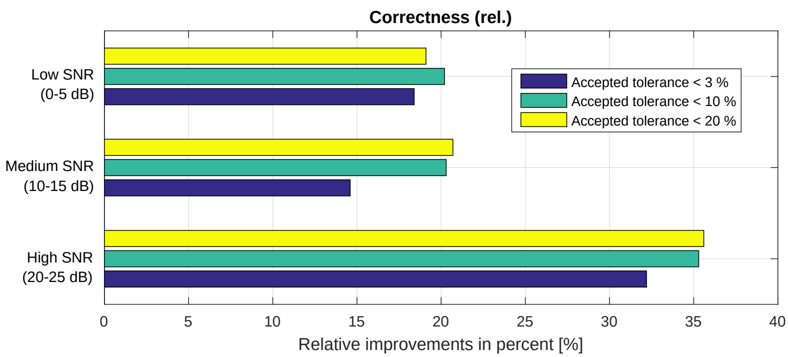

Figure 4 - Relative improvements of the pitch frequency correctness for different tolerance ranges.

range of indices is performed and the argument of the maximum represents the normalized pitch period.

By taking the inverse of the pitch period an estimate for the pitch frequency results. For more details, the reader is referred to [5]. For the refinement a basic DFT of order $N = 256$, a frameshift $R = 64$, and $M = 5$ hann-windows of order N have been used. The weighting coefficients were set by: $a_k^{(0)} = a_k^{(M-1)} = 1/10$, $a_k^{(1)} = a_k^{(M-2)} = 7/30$ and $a_k^{((M-1)/2)} = 1/3$. The performance and the accuracy of the pitch estimation method without and with spectral refinement has been analysed for low SNRs (0-5 dB), medium SNRs (10-15 dB), and high SNRs (20-25 dB). As reference a clean speech laryngograph database has been used consisting of a high number of pitch frequencies out of the interval [60 Hz, 400 Hz]. For performance analyses the correctness and the false detection of the estimated pitch frequencies have been considered. Three value ranges have been examined for the correctness: pitch frequency estimation falls within a tolerance range of 3 %, 10 %, and 20 %. A false detection occurs when the algorithm estimates a pitch frequency while no reference pitch frequency exists at that instance. Fig. 4 shows the relative improvements of the correctness of the pitch estimation method with spectral refinement for high, medium, and low SNR. From the results one can see that a significant increase of correctness by about $30 - 35$ % (rel.) at high SNR, approx. $15 - 20$ % (rel.) at medium SNR, and about $18 - 20$ % (rel.) at low SNR are achievable. Analyses have further shown that the miss detection rate is nearly kept unchanged for all considered SNRs while the correctness is considerably increased at the same time.

7 Summary and Outlook

In this contribution a refinement method of short-term spectra applied either as a preprocessing stage or as postprocessing stage of an analysis filterbank for speech signals was presented. The refinement method is particularly suitable for speech processing systems with already integrated analysis filterbanks or DFTs. Therefore, by applying refinement methods as a preprocessing stage or a postprocessing stage of a lower order DFT specific feature estimation schemes such as pitch frequency or noise power estimation can be further improved. In this paper the refinement method was realized as a preprocessor of an analysis filterbank and was applied for pitch frequency estimation. Results show that the performance of pitch frequency estimation can be enhanced considerably while the miss detection rate is kept unchanged for all considered SNRs. Furthermore, the relationship between polyphase filterbanks and spectral refinement methods was considered. It has been figured out that temporal refinement of short-term spectra applied as a preprocessor of DFT-modulated analysis filterbanks can be considered as an extension of polyphase filterbanks. Future work should be concentrated on how polyphase filterbanks could be extended appropriately to achieve improved performance. Also the refinement method applied for background noise estimation and for echo cancellation to reduce aliasing effects should be further investigated.

References

[1] A. de Cheveigne, H. Kawahara: *Yin, a Fundamental Frequency Estimator for Speech and Music,* J. Acoust. Soc. Am., vol. 111, no. 4, pp. 1917-1930, 2002.

[2] R. E. Crochiere, L. R. Rabiner: *Multirate Digital Signal Processing,* USA: Prentice-Hall, 1983.

[3] ETS 300 903 (GSM 03.50): *Transmission Planning Aspects of the Speech Service in the GSM Public Land Mobile Network (PLMS) System,* ETSI, France, 1999.

[4] E. Hänsler, G. Schmidt: *Acoustic Echo and Noise Control – A Practical Approach,* John Wiley & Sons, Hoboken, NJ, USA, 2004.

[5] M. Krini, G. Schmidt: *Model-based Speech Enhancement,* in E. Hänsler, G. Schmidt (eds.), Speech and Audio Processing in Adverse Environments, Springer, pp. 89–134, Berlin, Germany, 2008.

[6] M. Krini, G. Schmidt: *Spectral Refinement and its Application to Fundamental Frequency Estimation,* WASPAA, New York, USA, 2007.

[7] M. R. Schroeder: *Period Histogram and Product Spectrum: New Methods for Fundamental Frequency Measurements,* J. Acoust. Soc. Am., vol. 43, no. 4, pp. 829–834, 1968.

[8] G. Wackersreuther: *On the Design of Filters for Ideal QMF and Polyphase Filter Banks,* AEU, vol. 39, no. 2, pp. 123–13, 1985.