# MULTI-BAND PRE-ECHO CONTROL USING A FILTERBANK EQUALIZER

*Bernd Geiser, Martin Roggendorf* and Peter Vary*

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{geiser|roggendorf|vary}@ind.rwth-aachen.de

## ABSTRACT

We propose a new method to mitigate so called "pre-echo artifacts" that occur with transform-based audio codecs owing to their comparatively large block lengths. Our algorithm operates entirely in the time domain. In contrast to previous time domain approaches, a *frequency selective* pre-echo control (PEC) is achieved by applying the concept of a filterbank equalizer (FBE) with non-uniform frequency resolution.

As an example application, the new PEC approach has been used to enhance the ITU-T G.722.1C super-wideband codec. The performance of the algorithm has been objectively measured and optimized w.r.t. various FBE parameters. Considerable quality gains could be obtained which are confirmed by the subjective listening impression.

The proposed algorithm can be easily extended to act as an adaptive *spectro-temporal pre-* and *deemphasis* filter.

## 1. INTRODUCTION

Nowadays, virtually all successful standards for lossy audio compression are based on transform domain coding techniques and a perceptually relevant use of the available bit budget. An excellent overview is provided in [1].

A popular transform is the Modified Discrete Cosine Transform (MDCT) which is for instance used in the MPEG Advanced Audio Coding (AAC) standard [2, 3] or in the Ogg Vorbis coder [4]. It can be observed that the typical coding techniques that are used in these codecs are also gaining importance for high-quality real-time *conversational applications*. Several (partially) transform-based codecs with this focus have been standardized over recent years, in particular by ITU-T: G.711.1 [5, 6], G.718 [7, 8], G.719 [9, 10], G.722.1C [11, 12], and G.729.1 [13, 14]. Although all of these codecs employ a shorter transform length to reduce the algorithmic delay from several 100 ms (which is a typical value for non-conversational audio codecs) down to about 40 ms, the well-known problem of pre-echoes persists. Therefore, a number of different algorithms have been developed to overcome this issue.

In [15] we have described a time domain approach for pre-echo control (PEC), which has been applied in the context of a super-wideband extension of the ITU-T G.729.1 wideband codec [16]. A good quality gain could be shown for the high frequency range (8 – 14 kHz) of strongly transient signals. Moreover, the realization in the time domain provided the advantage of reusing the transmitted information for the purpose of frame erasure concealment (FEC).

* Martin Roggendorf has been working as a research assistant with the Institute of Communication Systems and Data Processing, RWTH Aachen University, Germany. He is now with RWE Power AG, Hürth, Germany.

However, time domain PEC approaches are usually limited to straightforward gain manipulation without any frequency selectivity. For obvious reasons, enforcing a gain contour for a full-band signal results in an unnatural and "snatchy" sound character. Therefore, such methods are only applicable with subband codecs that already provide an inherent subband decomposition. An alternative approach that can provide the desired frequency selectivity is linear prediction over frequency (transform coefficients). In the AAC codec, this method is called "Temporal Noise Shaping."

In the present paper, we propose a new approach for time domain PEC which achieves a *uniform or non-uniform frequency selectivity*. The algorithm is based on the concept of a filterbank equalizer (FBE) [17, 18]. Thereby, the applied spectral weights are adapted in a similar manner as for the single-band algorithm of [15].

**Organization of the Paper**

Sec. 2 reviews our previously proposed PEC method [15]. In Sec. 3, the concept of the FBE and its applications to speech processing and enhancement are summarized. The proposed modifications of the FBE and our new PEC method are detailed in Sec. 4. The new algorithm has been tested as an extension to the ITU-T G.722.1C super-wideband codec at codec rates of 24, 32, and 48 kbit/s. Various FBE parameter settings have been evaluated (Sec. 5). The paper is concluded in Sec. 6.

## 2. PRE-ECHO CONTROL (PEC) ALGORITHM

The PEC algorithm of [15] was designed for the 8 – 14 kHz frequency range of the input signal. The corresponding subband signal was assumed to be real-valued. Now, as a generalization, $S_i(k)$ may also be complex-valued. It represents the $i$-th subband of the input signal $s(n)$. As in [15], the PEC side information is computed based on the $\lambda$-th input frame of the subband signal $S_i(k)$, whereby the frame length is $L_{\mathrm{F}}$. More specifically, we use the overall gain

$$g^{(i)}(\lambda) = \frac{1}{2}\log_2 \frac{\sum_{k=0}^{L_{\mathrm{F}}-1} |S_i(\lambda L_{\mathrm{F}}+k)|^2}{L_{\mathrm{F}}}, \quad \lambda \in \mathbb{N}_0, \tag{1}$$

and $N_{\mathrm{SF}}$ subframe gains per frame with length $L_{\mathrm{SF}} = L_{\mathrm{F}}/N_{\mathrm{SF}}$

$$g_{\mathrm{SF}}^{(i)}(\lambda,\lambda_{\mathrm{SF}}) = \frac{1}{2}\log_2 \frac{\sum_{k=0}^{L_{\mathrm{SF}}-1} |S_i(\lambda L_{\mathrm{F}}+\lambda_{\mathrm{SF}}L_{\mathrm{SF}}+k)|^2}{L_{\mathrm{SF}}}, \tag{2}$$

where $\lambda_{\mathrm{SF}} \in \{0,\ldots,N_{\mathrm{SF}}-1\}$ and $i \in \{0,1,\ldots,M-1\}$ with $M$ denoting the number of subbands.

Each frame $\lambda$ is then classified as either "transient" or "stationary." This decision is communicated to the decoder using one bit. For *stationary* frames, the frame gain from (1) is quantized and transmitted. For *transient* frames the subframe gains according to (2) are encoded. Moreover, if this information is suitably distributed across neighboring bitstream frames, it can be efficiently reused for purposes of frame erasure concealment as shown in [15]. However, in the scope of the present paper, we will not consider this option.

In the original proposal, the quantized PEC information was used to construct a smooth "temporal gain function" (TGF) which could then be used to correct the temporal envelope of the decoded subband signal. Here, the filterbank equalizer described in the subsequent section will perform this task while, at the same time, providing a (non-uniform) frequency selective temporal signal shaping.

## 3. FILTERBANK EQUALIZER (FBE)

A block diagram of the employed filterbank equalizer (FBE) is shown in Fig. 1. In principle, the FBE applies spectral gain modifications to its input signal and, therefore, it can be used as an efficient alternative for a DFT analysis-synthesis filterbank. The FBE weight adaptation is, in contrast to the conventional approach, performed entirely in the (non-uniform) frequency domain while the actual signal processing is performed with a time domain filter. A typical use-case for the FBE is speech enhancement, in particular noise reduction [17] and near end listening enhancement [19].

### 3.1 Uniform FBE

The FBE input signal $\tilde{s}_{\mathbf{dec}}(n)$ (which denotes the decoded signal affected by pre-echoes in our case) is sampled with a rate of $f_s$. The FBE analysis splits this signal into $M$ subband signals $S_i(k)$ by means of a DFT analysis filterbank:

$$S_i(k) = \sum_{\nu=0}^{L} s(k-\nu) \cdot h_0(\nu) \cdot \mathrm{e}^{-\mathrm{j}\frac{2\pi}{M} i\nu}, \tag{3}$$

where $i \in \{0, 1, \ldots, M-1\}$ is the subband index and $k$ is the time index in the subsampled domain, i.e., $k = \lfloor n/r \rfloor \cdot r$ with downsampling ratio $r$. The prototype filter of length $L+1$ is denoted by $h_0(n)$. For $L > M$, (3) can be implemented with a polyphase network (PPN) and a FFT of length $M$.

The computation of the spectral weights $W_i$ depends on the desired application. For noise reduction, the weights $W_i$ are computed from the subband signals $S_i$ alone with the help of a noise tracking algorithm [17]. If the FBE is being used for near end listening enhancement, a noise estimate has to be supplied externally [19]. Here, we index the weights $W_i$ by the frame and subframes as used for PEC gain computation according to (1) and (2), i.e., $W_i(\lambda, \lambda_{\mathrm{SF}})$. The exact weight computation rule for the PEC case is detailed in Sec. 4.2.

The weights $W_i(\lambda, \lambda_{\mathrm{SF}})$ are transformed into time domain filter coefficients using a generalized DFT (GDFT)

$$h(n, \lambda, \lambda_{\mathrm{SF}}) = h_0(n) \cdot \sum_{i=0}^{M-1} W_i(\lambda, \lambda_{\mathrm{SF}}) \cdot \mathrm{e}^{-\mathrm{j}\frac{2\pi}{M} i(n-n_0 \cdot L)} \tag{4}$$

with $n \in \{0, 1, \ldots, L\}$ and $n_0 = 1/2$, cf. [17]. Then, the time domain filter can be applied to the FBE input signal:

$$\tilde{s}(n) = \sum_{\nu=0}^{L} \tilde{s}_{\mathrm{dec}}(n-\nu) \cdot h(\nu, \lambda, \lambda_{\mathrm{SF}}). \tag{5}$$

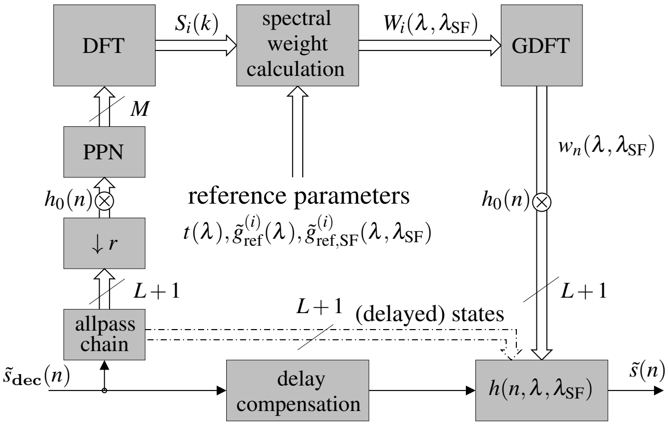


Figure 1: Filterbank Equalizer (decoder side)

### 3.2 Warped FBE

The FBE as described above can be easily extended to provide a non-uniform frequency resolution. This *frequency warping* is achieved by an allpass transformation of the delay chains in the filter as well as before the DFT [1], i.e.,

$$z^{-1} \to H_{\mathrm{A}}(z) = \frac{z^{-1} - a}{1 - az^{-1}} \tag{6}$$

with the allpass pole $-1 < a < 1$, also called "warping factor." The allpass element $H_{\mathrm{A}}(z)$ has a unity amplitude response and a frequency dependent phase response

$$\varphi_a(\Omega) = 2 \cdot \arctan\left(\frac{\sin(\Omega)}{\cos(\Omega) - a}\right) - \Omega \tag{7}$$

with the normalized frequency $\Omega = 2\pi f / f_s$. Effectively, the subband filters with frequency responses $H_i\left(z = \mathrm{e}^{\mathrm{j}\Omega}\right)$ are converted into warped filters with

$$H_i\left(z = \mathrm{e}^{\mathrm{j}\varphi_a(\Omega)}\right) = \sum_{\nu=0}^{L} h_0(n) \cdot \mathrm{e}^{-\mathrm{j}\frac{2\pi}{M} i \cdot \nu} \cdot \mathrm{e}^{-\mathrm{j}\varphi_a(\Omega)\nu}. \tag{8}$$

Examples for an allpass pole of $a = 0.5$ are shown in Fig. 2. In the graphs, "even stacking" denotes the regular DFT implementation according to (3) while "odd stacking" is a GDFT with a frequency shift of $n_0 = 1/2$ channel, cf. [20].

If $a \neq 0$, the then warped filter $h(n, \lambda, \lambda_{\mathrm{SF}})$ does not have a linear phase response and a tunable *phase equalizer* can be used to compensate objectionable phase distortions [17].

### 3.3 Filter Structure and Filter Shortening

Since $h(n, \lambda, \lambda_{\mathrm{SF}})$ is an adaptive filter, the actual implementation is important. A good choice is a crossfading of the output of two parallel filters, cf. [17]. This method effectively suppresses artifacts that stem from heavily varying filter coefficients. Note that a suitable delay compensation (as shown in Fig. 1) is required in this case so that the spectral weights take effect at the correct time instance.

Furthermore, in [17], a method to approximate the filter $h(n, \lambda, \lambda_{\mathrm{SF}})$ by a low-delay autoregressive (AR) filter is described. However, this option is currently not used here.

[1] In practice, a single allpass chain should be reused for both modules as indicated in Fig. 1 by the dash-dotted line (considering the delay).

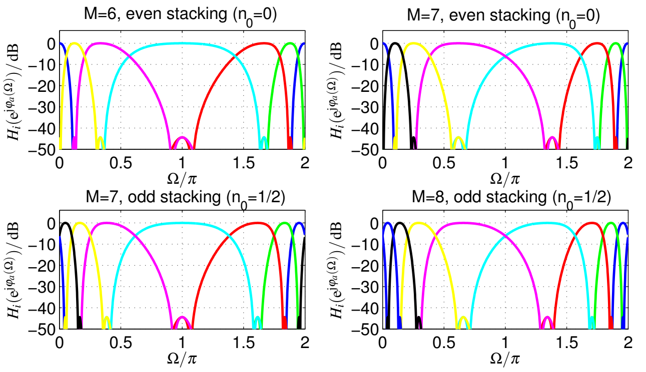


Figure 2: Frequency responses of the FBE analysis filters for $a = 0.5$. There are always four possibilities to obtain a certain number of *independent* DFT channels (here: 4).

## 4. FBE-BASED PEC IN THE TIME DOMAIN

The envisioned application of the FBE concept to the PEC problem requires, as opposed to its typical use in speech enhancement, to capture the detailed temporal evolution of audio signals within the FBE subbands. Therefore, different parameter settings and several slight algorithmic modifications are needed which are detailed below together with the specific weight computation rule.

### 4.1 Short Term Power Estimation

In speech enhancement applications, typically a strong recursive averaging is used to estimate the signal power in each subband. However, in our case we require an instantaneous and exact value for each subband $i$ and each subframe $\lambda_{\mathrm{SF}}$ of length $L_{\mathrm{SF}}$. Therefore, we set $r = 1$ so that the subband signals are *not* downsampled. Then, the short term power for each subband signal can be computed as described by (1) and (2). Though, aiming at further complexity reduction, a moderate downsampling (e.g., $r = 2$ or 4) could be useful.

### 4.2 Update of Spectral Weights and Filter Coefficients

To compute the desired spectral weights $W_i(\lambda, \lambda_{\mathrm{SF}})$, both the (sub)frame gains derived from the decoded signal $\tilde{s}_{\mathbf{dec}}(n)$ as well as a set of reference parameters is required. These reference parameters describe the temporal characteristics of the *original* audio signal $s(n)$. Specifically, the reference (sub)frame gains, $\tilde{g}^{(i)}_{\mathrm{ref}}(\lambda)$ and $\tilde{g}^{(i)}_{\mathrm{ref,SF}}(\lambda, \lambda_{\mathrm{SF}})$, and the transient flag[2] $t(\lambda)$ have to be determined at the encoder side. Therefore, an identical signal analysis chain (consisting of the allpass chain, the polyphase network, and the DFT) needs to be implemented at the encoder side. Now, the spectral weights for the FBE can be computed as follows:

$$W_i(\lambda, \lambda_{\mathrm{SF}}) = \begin{cases} 2^{\tilde{g}^{(i)}_{\mathrm{ref}}(\lambda) - g^{(i)}(\lambda)} & \text{if} \quad t(\lambda) = 0 \\ 2^{\tilde{g}^{(i)}_{\mathrm{ref,SF}}(\lambda, \lambda_{\mathrm{SF}}) - g^{(i)}_{\mathrm{SF}}(\lambda, \lambda_{\mathrm{SF}})} & \text{if} \quad t(\lambda) = 1. \end{cases} \tag{9}$$

Consequently, the update interval for the $L+1$ filter coefficients $h(n, \lambda, \lambda_{\mathrm{SF}})$ is $L_{\mathrm{SF}}$ samples within transient frames and $L_{\mathrm{F}}$ samples for stationary frames.

[2] For this paper we have restricted ourselves to a global transient flag $t(\lambda)$ instead of an individual flag per subband $i$. The transient detection is performed similar to the method of [15].

### 4.3 Alternative System Architectures

The above described system is in principle applicable to all transform-based audio codecs since it can act as an "add-on." As a more bit rate efficient alternative, a *second* FBE could normalize the input signal $s(n)$ *before* it is supplied to the encoder. In this case, the FBE weights are computed as

$$W^{\mathrm{enc}}_i(\lambda, \lambda_{\mathrm{SF}}) = \begin{cases} 2^{-\tilde{g}^{(i)}_{\mathrm{ref}}(\lambda)} & \text{if} \quad t(\lambda) = 0 \\ 2^{-\tilde{g}^{(i)}_{\mathrm{ref,SF}}(\lambda, \lambda_{\mathrm{SF}})} & \text{if} \quad t(\lambda) = 1. \end{cases} \tag{10}$$

In effect, this FBE variant constitutes an adaptive *spectro-temporal preemphasis filter*; the audio codec only needs to encode a signal of strongly reduced dynamic range. The decoder side FBE, with weights adapted according to (9), can still be used to denormalize/deemphasize the decoded signal. A related (single subband) approach has been used in [16].

In the system as depicted in Fig. 1, the bit rate for the PEC side information will increase with the number of DFT channels $M$. Yet, with the availability of a *locally decoded signal* at the encoder side, it becomes possible to compute the required filter coefficients $h(n, \lambda, \lambda_{\mathrm{SF}})$ directly in the encoder and efficient quantization techniques can be employed, especially if AR approximation and filter shortening as mentioned in Sec. 3.3 is used. In this case, the additional PEC bit rate will be determined by the filter length.

## 5. PARAMETRIZATION AND EVALUATION

In this section, the influence of the FBE parameters on the obtained audio quality shall be investigated. Therefore, we have chosen the ITU-T G.722.1C super-wideband codec [11, 12] which is a classical transform codec but lacks a dedicated PEC mechanism. The G.722.1C codec operates at a sampling rate of $f_s = 32$ kHz and provides bit rate modes of 24, 32, and 48 kbit/s. For the codec *with* PEC, an instrumental quality assessment in terms of a PEAQ score improvement [21] over the codec *without* PEC has been used. The test material has been taken from the EBU SQAM corpus [22] (ABBA, Castanets, German Male, and German Female).

### 5.1 Parametrization

To investigate various FBE parametrizations, we have used *unquantized* versions of the reference PEC parameter set (i.e., $g_{\mathrm{ref}}$ and $g_{\mathrm{ref,SF}}$). Exemplary results for quantized parameters are given in Sec. 5.2.

#### *5.1.1 Sampling Rate & Framing*

The sampling rate is $f_s = 32$ kHz to match the G.722.1C input and output signals. For signal analysis with the FBE, we have used frames of 20 ms length, i.e., $L_{\mathrm{F}} = 640$. These frames have been subdivided into $N_{\mathrm{SF}} = 8$ subframes of length 2.5 ms each, i.e., $L_{\mathrm{SF}} = 80$. As reasoned in Sec. 4.1, no downsampling is used in the analysis part of the FBE ($r = 1$). Yet, the filter coefficient update is performed every $L_{\mathrm{SF}} = 80$ or $L_{\mathrm{F}} = 640$ samples, respectively.

#### *5.1.2 Filter Structure, Delay Compensation, Phase EQ*

The transition from subframe to subframe of the adaptive filter $h(n, \lambda, \lambda_{\mathrm{SF}})$ is implemented using a crossfading method with 80 samples overlap. To align the filter coefficients with the input signal, the delay compensation block in Fig. 1 has to delay the signal by 40 sample instants. For $a \neq 0$, an additional phase equalizer of length $2.5 \cdot L$ is applied to the FBE output signal.

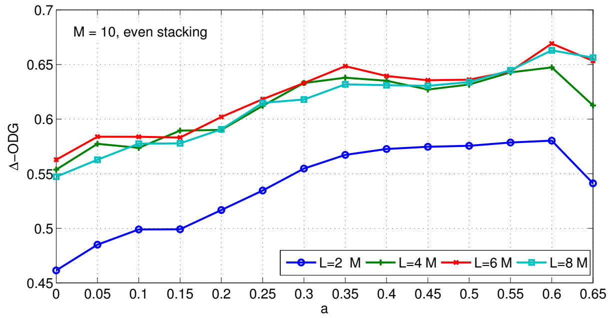


Figure 3: Improvement of PEAQ score (Δ-ODG): Influence of warping factor $a$ and choice of filter length $L$. The shown results are averaged over all codec bit rates and test items.

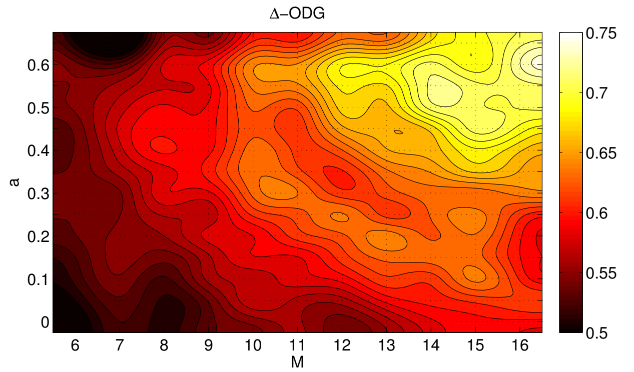


Figure 5: Improvement of PEAQ score (Δ-ODG): Dependency on $a$ and $M$ for even DFT stacking and $L = 4 \cdot M$. Results have been interpolated for better visualization. Exact values are shown at grid intersections. The shown results are averaged over all codec bit rates and test items.

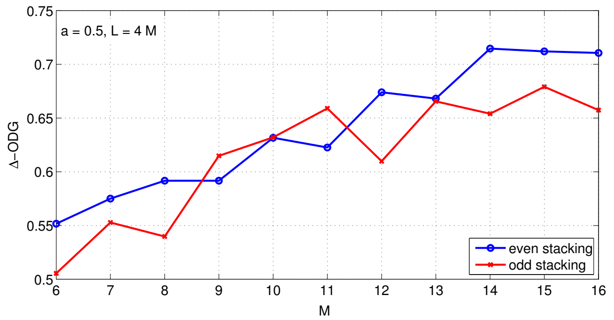


Figure 4: Improvement of PEAQ score (Δ-ODG): Influence of DFT stacking (even or odd). The shown results are averaged over all codec bit rates and test items.

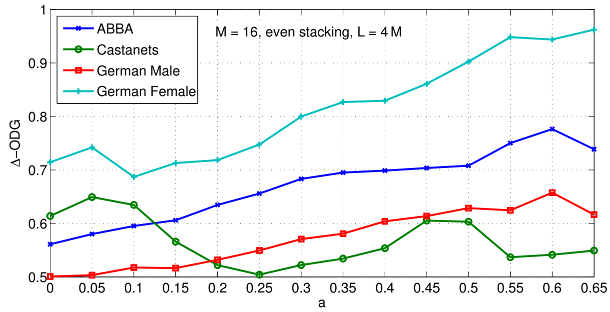


Figure 6: Improvement of PEAQ score (Δ-ODG): Test item variability concerning the warping factor $a$. The shown results are averaged over all codec bit rates.

#### *5.1.3 Filter Length*

For a DFT analysis filterbank with polyphase filtering, the prototype filter order $L$ is usually given as a multiple of the DFT size $M$. Experimental results are shown in Fig. 3. The average quality gain already saturates with $L = 4 \cdot M$. Note that this is not only true for the shown case ($M = 10$ and even DFT stacking) but for almost all other configurations. Therefore, we keep $L = 4 \cdot M$ in the following.

#### *5.1.4 DFT Stacking and Length*

An "evenly stacked DFT" is identical with the regular DFT algorithm while the "oddly stacked DFT" applies a frequency shift of $n_0 = 1/2$ channel [20]. The impact of DFT stacking on the subband division of the (warped) FBE has been shown in Fig. 2. Obviously, there is no direct relation between DFT length $M$ and the bit rate spent for PEC reference parameter transmission. In fact, the number of PEC parameter sets to be transmitted is identical with the number of *independent* DFT channels. In particular, evenly stacked DFTs of length $2M_0$ ($M_0 \in \mathbb{N}$), oddly *and* evenly stacked DFTs of length $2M_0 + 1$, as well as oddly stacked DFTs of length $2M_0 + 2$ all have the same number of independent channels ($M_0 + 1$). This has to be considered for the PEC parameter transmission.

Concerning the obtained average quality gain, there is a general tendency towards the DFT with even stacking. Yet, as illustrated Fig. 4, there are some cases where the odd stacking is better.

#### *5.1.5 Warping Factor vs. DFT Length*

Fig. 3 already indicates that frequency warping is beneficial for FBE-based PEC. However, the actual impact strongly depends on the chosen DFT length $M$. Therefore, the interdependency of warping factor $a$ and DFT length $M$ is analyzed in Fig. 5. Without warping, the achievable average quality gain saturates at Δ-ODG $= 0.6$ (for $M = 15$), a value which is already achieved for $M = 8$ when applying a warping factor of $a = 0.4$. With higher warping factors, average quality gains of up to Δ-ODG $= 0.75$ are possible.

The variability of the results over the test items has also been investigated. Fig. 6 reveals that the quality gain through warping is consistent over all items except for the castanets. This can be explained by the extreme transient nature of that signal. The castanets exhibit nearly identical temporal behavior over the entire frequency range which is, however, not a typical case. Yet, an average quality gain for the castanet signal of $0.5 - 0.6$ is still acceptable and not necessarily worse than for an unwarped FBE. It has also been verified that this improvement is clearly audible.

For an actual implementation on DSP hardware, a value of $a = 0.5$ appears to be a good compromise since the multiplications in the allpass chain (6) can be replaced by simple binary shift operations in this case. The choice of the DFT length is more a matter of allowable bit rate, at least within the investigated range of values, because the quality scores increase nearly monotonically with $M$.

Table 1: Improvement of PEAQ score ($\Delta$-ODG) for quantized reference parameters. $M = 7$, even stacking, $L = 4 \cdot M$, $a = 0.5$.

| Codec $\rightarrow$ <br> Test item $\downarrow$ | G.722.1C@24 kbit/s <br> unquant. | / quant. PEC | G.722.1C@32 kbit/s <br> unquant. | / quant. PEC | G.722.1C@48 kbit/s <br> unquant. | / quant. PEC | average added <br> bit rate (kbit/s) |
|---|---|---|---|---|---|---|---|
| Castanets | +0.20 | +0.20 | +0.38 | +0.39 | +0.84 | +0.80 | 1.15 |
| German Male | +0.37 | +0.34 | +0.43 | +0.41 | +0.42 | +0.38 | 3.64 |
| German Female | +0.59 | +0.58 | +0.79 | +0.77 | +1.14 | +1.10 | 2.97 |
| Pop Music (ABBA) | +0.46 | +0.44 | +0.59 | +0.55 | +0.68 | +0.60 | 3.31 |
| ∅ | +0.41 | +0.39 | +0.55 | +0.53 | +0.77 | +0.72 | 2.77 |

### 5.2 Parameter Quantization

Finally, our concept needs to be verified for quantized PEC reference parameters. We have used scalar quantization of the (sub)frame gains with a step size of $1/8$ (in the $\log_2$ domain, cf. (1), (2)) followed by Huffman coding of the (differential) quantizer indices. As a reasonable example, we have chosen a DFT length of $M = 7$ with even stacking, i.e., *four* independent DFT subbands. As argued above, a warping factor of $a = 0.5$ and a filter order of $L = 4 \cdot M$ has been selected.

In Tab. 1, the results for quantized and unquantized PEC parameters are tabulated together with the average added bit rate. For our test items, we have found that the G.722.1C coder at 24 kbit/s *with* PEC (quantized) outperforms G.722.1C at 32 kbit/s *without* PEC by an average of $\Delta$-ODG $= +0.23$ while saving bit rate. The same is true for G.722.1C at 32 kbit/s *with* PEC when compared to G.722.1C at 48 kbit/s *without* PEC ($\Delta$-ODG $= +0.12$).

## 6. DISCUSSION AND CONCLUSIONS

In this paper we have proposed a new method for pre-echo control in transform audio codecs which is based on a (warped) filterbank equalizer. The major advantage over previous time domain PEC approaches is the inherent frequency selectivity of the FBE concept. Therefore, a better representation of the relevant signal characteristics can be obtained which was verified with the instrumental PEAQ measure. The subjective listening impression confirms substantial quality gains over the G.722.1C coder without PEC.

The *frequency warped* version of the FBE is particularly beneficial since it elegantly matches the properties of the human auditory system. At lower frequencies, more spectral resolution is required than at higher frequencies. On the other hand, higher frequencies are more susceptible to temporal artifacts. In fact, the temporal resolution of the FBE subband filters increases with the subband bandwidth, i.e., higher (and therefore broader) subbands are more accurately represented in the time domain.

Still, since this work is so far regarded as a concept study, a real-world application of FBE-based PEC requires further investigations. The open questions concern, e.g., a more efficient quantization scheme for the PEC parameters, an *individual* subframe division and transient detection for each DFT subband to achieve further bit rate savings, filter shortening (Sec. 3.3), and an optimized phase equalizer degree. Moreover, we believe that the use of our FBE proposal as an adaptive spectro-temporal preemphasis filter (Sec. 4.3) is an interesting option that deserves further attention.

### Acknowledgments

The authors would like to thank *Heinrich W. Löllmann* and *Bastian Sauert* for many helpful discussions and ideas.

## REFERENCES

[1] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–513, Apr. 2000.

[2] ISO/IEC JTC1/SC29/WG11 (MPEG), "International Standard ISO/IEC 13818-7 'Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding'."

[3] M. Bosi *et al.*, "ISO/IEC MPEG-2 Advanced Audio Coding," *Journal of the Audio Engineering Society*, vol. 45, no. 10, pp. 789–814, Oct. 1997.

[4] C. Montgomery, "Vorbis I specification," 2004, online: http://www.xiph.org/vorbis/doc/Vorbis_I_spec.html.

[5] ITU-T Rec. G.711.1, "Wideband embedded extension for G.711 pulse code modulation," 2008.

[6] Y. Hiwasaki *et al.*, "G.711.1: A Wideband Extension to ITU-T G.711," in *Proc. of EUSIPCO*, Lausanne, Switzerland, Aug. 2008.

[7] ITU-T Rec. G.718, "Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s," 2008.

[8] T. Vaillancourt *et al.*, "ITU-T EV-VBR: A robust 8-32 kbit/s scalable coder for error prone telecommunications channels," in *Proc. of EUSIPCO*, Lausanne, Switzerland, Aug. 2008.

[9] ITU-T Rec. G.719, "Low-complexity, full-band audio coding for high-quality, conversational applications," 2008.

[10] M. Xie *et al.*, "ITU-T G.719: A new low-complexity full-band (20 kHz) audio coding standard for high-quality conversational applications," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, Oct. 2009, pp. 265–268.

[11] ITU-T Rec. G.722.1, "Low complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," 2005.

[12] M. Xie, D. Lindbergh, and P. Chu, "ITU-T G.722.1 Annex C: A new low-complexity 14 kHz audio coding standard," in *Proc. of IEEE ICASSP*, Toulouse, France, May 2006.

[13] ITU-T Rec. G.729.1, "G.729 based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729," 2006.

[14] S. Ragot *et al.*, "ITU-T G.729.1: An 8-32 kbit/s scalable coder interoperable with G.729 for wideband telephony and Voice over IP," in *Proc. of IEEE ICASSP*, Honolulu, Hawai'i, USA, Apr. 2007.

[15] B. Geiser and P. Vary, "Joint pre-echo control and frame erasure concealment for VoIP audio codecs," in *Proc. of EUSIPCO*, Glasgow, Scotland, Aug. 2009.

[16] B. Geiser *et al.*, "Candidate proposal for ITU-T super-wideband speech and audio coding," in *Proc. of IEEE ICASSP*, Taipei, Taiwan, Apr. 2009.

[17] H. W. Löllmann and P. Vary, "Uniform and warped low delay filter-banks for speech enhancement," *Speech Communication*, vol. 49, no. 7-8, pp. 574–587, July 2007, Special Issue on Speech Enhancement.

[18] P. Vary, "An adaptive filterbank equalizer for speech enhancement," *Signal Processing*, vol. 86, no. 6, pp. 1206–1214, June 2006, special Issue on Applied Speech and Audio Proc. (dedicated to Prof. Hänsler).

[19] B. Sauert, H. W. Löllmann, and P. Vary, "Near end listening enhancement by means of warped low delay filter-banks," in *ITG Speech Comm.* Aachen, Germany: VDE Verlag GmbH, Oct. 2008.

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

[21] ITU-R Rec. BS.1387, "Method for objective measurements of perceived audio quality," 1998.

[22] European Broadcast Union, EBU Tech 3253, "Sound quality assessment material — recordings for subjective tests," 1988.