EFFICIENT LOW DELAY FILTERING FOR RESIDUAL ECHO SUPPRESSION

Christelle Yemdji, Moctar Mossi Idrissa and Nicholas W. D. Evans

EURECOM Institute
06560 Sophia-Antipolis, France
{yemdji, mossi, evans}@eurecom.fr
web: www.eurecom.fr

Christophe Beaugeant

Infineon Technologies
06560 Sophia-Antipolis, France
christophe.beaugeant@infineon.com
web: www.infineon.com

ABSTRACT

In telecommunications terminals speech quality is often degraded by acoustic echo. Various approaches to echo cancellation have been proposed and generally involve two separate stages, namely those of adaptive echo cancellation and, as is the focus here, residual echo suppression. Whilst computationally efficient, residual echo suppression approaches based on gain loss control have poor double talk performance. Sub-band approaches give better performance but generally introduce a significant signal delay. This paper reports new experimental work which assesses the performance of three low delay approaches to suppress residual echo through time domain convolution. Results show that a new approach, based on the inverse discrete Fourier transform, performs as well as the existing approaches for both linear and non-linear echo whilst maintaining computational efficiency and low signal delay.

Index Terms — echo cancellation, echo postfiltering, linear echo, non-linear echo, sub-band filtering, FIR filter.

1. INTRODUCTION

An acceptable level of speech quality is an important requirement for any telecommunications terminal. With mobile devices, however, speech quality is often degraded by varying levels of ambient noise and acoustic echo. In noisy environments, the microphone is sensitive to near-end speech and ambient noise which are both transmitted to the far-end speaker. Acoustic echo results from the coupling between the loudspeaker and the microphone and, as a result, the far-end speaker can sometimes hear a delayed version of their own voice, where the delay is introduced by the communications link. Acoustic echo cancellation (AEC) and noise compensation are used to tackle these problems [1]. This paper is concerned solely with the problem of AEC.

Most approaches to AEC are based on adaptive filters [1]. As illustrated in Figure 1, an adaptive filter is used to generate an estimate of the echo signal which is then subtracted from the microphone signal. However, because of the limited filter order, changes in the acoustic path and non-linearities, the resulting signal generally contains some residual echo. Postfilters are commonly used to obtain further echo attenuation [1].

A simple and popular approach to residual echo suppression is that of gain loss control (GLC) [1, 2]. GLC algorithms simply consist in applying an attenuation to the uplink signal. Although this gain is generally calculated as a function of the loudspeaker power [2] it impacts on near-end speech during double talk periods because it is applied independently to the presence, or not, of near-end speech.

To overcome the poor double talk performance of GLC sub-band echo postfilters [3] are often used and are the focus of the work presented here. Sub-band postfilters are preferred to GLC because they consist of sub-band gains and can therefore specifically target frequencies where residual echo is audible. Such residual echo suppression filters can be applied to the uplink signal in the spectral domain. Even though sub-band filtering (SF) is advantageous because of its low computational complexity it, however, introduces significant delay in the output signal. This delay can be reduced by performing the filtering in the time domain through a convolution. In this case, the sub-band spectral gains are used to determine a broadband finite impulse response (FIR) filter. Popular approaches include the Filter Bank Equalizer (FBE) or the Low Delay Filter (LDF), presented in [4]. The LDF approach was used in [5] for joint noise reduction and residual echo suppression and in [6] for joint reverberation and noise suppression. Another alternative, reported by Hänsler in [7], involves the use of the inverse discrete Fourier transform (IDFT) of the spectral gains to suppress residual echo.

The side by side performance comparison of these approaches has, to our knowledge, not been previously reported. This paper presents the first comprehensive comparative assessment of the SF, FBE, LDF and IDFT approaches to sub-band residual echo processing.

The remainder of this paper is organized as follows. In the next section we present the algorithm used for the calculation of the spectral gains and different filtering scheme that can be used for processing residual echo. In Section 3, we compare the computational complexity of the different approaches. Section 4 presents our experimental setup used for simulations. Performances of the different approaches are presented in Section 5 and conclusions are presented in Section 6.

2. ECHO POSTFILTERING

Figure 1 illustrates the echo cancelling scheme used in our investigations: AEC followed by a postfilter to process the residual echo. The microphone signal $y(n)$ is composed of the near-end speech signal $s(n)$ and of the echo signal $d(n)$. The adaptive filter is used to generate an estimate of the echo signal $\hat{d}(n)$ which is subtracted from the microphone signal. The error signal $e(n)$ is composed of the residual echo $e_r(n)$ and of near-end speech $s(n)$. The postfilter aims to suppress the residual echo.

In the following, we describe the postfilter investigated. Section 2.1 details the common sub-band analysis used in our simulations. In Section 2.2, we present the algorithm used to calculate the spectral gains. Lastly, Section 2.3 presents the investigated filtering approaches.

2.1 Sub-band analysis

As shown in Figure 2, the error signal $e(n)$ and the loudspeaker signal $x(n)$ are split into sub-band signals $e_i(n)$ and $x_i(n)$ respectively, where i denotes the sub-band index and ranges from 0 to $M-1$. In our case, sub-band analysis and synthesis are performed through a discrete Fourier transform-modulated filter bank. One property of such filter banks is that each bandpass filter corresponds to a frequency shifted duplicate of a lowpass filter $h(n)$. In the literature $h(n)$ is referred to as a prototype filter [8].

Moreover, sub-band signals $e_i(n)$ and $x_i(n)$ have a reduced bandwidth compared to the original input signals $e(n)$ and $x(n)$. Therefore, the bit rate of the sub-band signals can be downsampled by a factor r, with the constraint that $r \leq M$ to avoid frequency do-

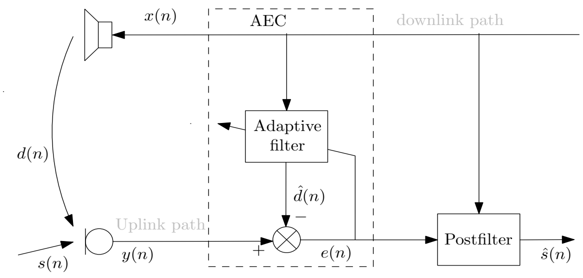


Figure 1: Echo cancelling scheme illustrating AEC followed by a sub-band echo postfilter with filtering in the sub-band domain.

main aliasing [8]. The benefit of downsampling is that the residual echo suppression filter update is performed at a sampling rate that is lower than that of the input sampling rate.

In our simulations, the above filter bank is implemented through a polyphase network (PPN) [7]. With the PPN implementation, the sub-band analysis stage requires one convolution and one Fourier transform instead of M convolutions.

2.2 Spectral gains

For each sub-band i, the gain of the echo postfilter is updated using a Wiener filtering rule [3]:

$$G_i(n) = \frac{\xi_i(n)}{1+\xi_i(n)}, \qquad (1)$$

where $\xi_i(n)$ is the signal (near-end speech) to echo ratio (SER). In practice, the SER is unknown and, consequently, it needs to be estimated. In our implementation, the SER is estimated through the Ephraim and Malah approach [9]:

$$\xi_i(n) = \beta \cdot \frac{\hat{s}_i^2(n-1)}{\hat{\gamma}_i^{e_r e_r}(n-1)} + (1-\beta)\cdot \max(\xi_i^{post}(n), 0) \qquad (2)$$

where the smoothing constant β lies in the interval $]0,1[$, $\hat{s}_i(n-1)$ is the i^{th} sub-band near-end speech signal estimate, $\hat{\gamma}_i^{e_r e_r}(n)$ is the residual echo spectral density and $\xi_i^{post}(n)$ is the a posteriori SER. The residual echo spectral density in Equation 2, $\hat{\gamma}_i^{e_r e_r}(n)$, is estimated according to [3]:

$$\hat{\gamma}_i^{e_r e_r}(n) = \frac{\gamma_i^{xe}(n)}{\gamma_i^{xx}(n)}, \qquad (3)$$

where $\gamma_i^{xe}(n)$ is the crosspower spectral density between $x(n)$ and $e(n)$ and $\gamma_i^{xx}(n)$ is the loudspeaker power spectral density. The a posteriori SER in Equation 2, $\xi_i^{post}(n)$, is calculated according to:

$$\xi_i^{post}(n) = \frac{e_i^2(n)}{\hat{\gamma}_i^{e_r e_r}(n)} - 1. \qquad (4)$$

In all cases, the spectral densities $\gamma_i^{xx}(n)$ and $\gamma_i^{xe}(n)$ are estimated through autoregressive smoothing as in [3].

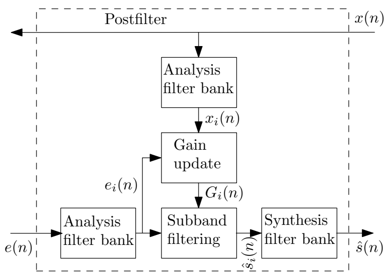
2.3 Postfiltering

Using the postfilter presented above echo can be processed either in the sub-band domain as described in Section 2.3.1 (see Figure 2(a)) or in the time domain as described in Section 2.3.2 (see Figure 2(b)).

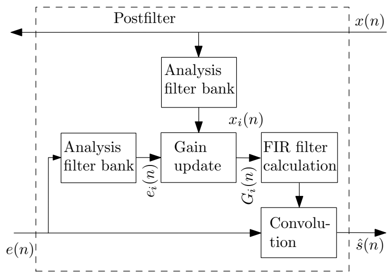
2.3.1 Sub-band filtering

Sub-band filtering (SF) is illustrated in Figure 2(a) and consists of applying the i^{th} sub-band gain $G_i(n)$ on the i^{th} sub-band microphone signal $e_i(n)$ as a multiplicative factor

$$\hat{s}_i(n) = G_i(n)\cdot e_i(n). \qquad (5)$$



(a) Echo filtering in the sub-band domain.



(b) Echo filtering with an FIR filter.

Figure 2: sub-band echo postfilter detailed scheme.

The full-band microphone signal $\hat{s}(n)$ is recovered by processing the sub-band signals $\hat{s}_i(n)$ through the appropriate synthesis filter bank.

Spectral gains are positive real numbers (zero phase), therefore the input signal phase is not modified. Moreover, for the analysis and synthesis stages, we restrict ourselves to linear phase prototype filters. The overall system (sub-band analysis and synthesis) introduces a signal delay of $L-1$ samples, where L is the length of the prototype filter.

2.3.2 Time domain filtering

Figure 2(b) shows the sub-band echo postfiltering scheme with a finite impulse response (FIR) filter. The sub-band signals $x_i(n)$ and $e_i(n)$ are used to calculate the sub-band gains $G_i(n)$ according to the equations in Section 2.2. The sub-band gains $G_i(n)$ are used to determine an FIR filter. This FIR filter is calculated according to the FBE, LDF or IDFT filter rule. The echo suppression is therefore done through convolution of the input signal $e(n)$ with any of the FIR filters. To avoid phase distortions and to ensure a constant signal delay, we ensure that these filters are linear phase filters [8]. We now present the three different FIR filters considered.

Filter bank equalizer (FBE): The FBE was introduced in [4] and is the mathematical time domain equivalent of the SF approach. The FBE is expressed as follows:

$$g_{fbe}(n) = h(n)\cdot \tilde{g}(n), \qquad (6)$$

where $h(n)$ is the prototype filter of the sub-band analysis stage and $\tilde{g}(n)$ is the IDFT of the spectral gains $G_i(n)$. As the spectral gains are positive (zero-phase), the linear phase property is assured if $h(n)$ has linear phase. This condition is fulfilled, for example, if $h(n)$ is symmetric.

The FBE process introduces a signal delay of $(L-1)/2$ samples, that is half the delay introduced by the SF method.

Low delay filter (LDF): Although the FBE has lower signal delay than the corresponding SF, smaller signal delays can be achieved by approximating the FBE by a lower degree filter [4, 6].

The LDF is obtained by truncating the FBE with a window of length P with $P < L$. The window can chosen arbitrarily or chosen so as to maintain linear phase. As the FBE has linear phase, one can use a window which is symmetric along $L/2$. We used the Hamming window in our simulations.

Experiments with different values of P showed that for an FBE of length L, an LDF of $L/2$ taps is a good match no matter what the number of sub-bands.

Inverse Discrete Fourier Transform (IDFT) filter: A more intuitive approach might be to obtain the FIR filter simply by applying the IDFT to the updated spectral gain factors [7]. The IDFT of the gains corresponds to a non-causal zero phase filter. A causal filter is obtained by applying a temporal shift of $(M-1)/2$. In the frequency domain, the temporal shift corresponds to a phase modification: the zero-phase filter then becomes a linear phase filter.

3. COMPUTATIONAL COMPLEXITY

Our study of time domain filters for sub-band adaptive filtering is mainly motivated by the reduction of signal delay. However, computational complexity is an aspect which is of crucial importance to real time implementations.

Table 1 shows the algorithmic complexity, memory requirements and signal delay characteristics of each different filtering approach (SF, FBE, LDF and IDFT). The number of multiplications and addition operations presented in columns 2 and 3 of Table 1 do not include the number of operations needed for the calculation of spectral gains nor the Fourier transform requirements since these are the same for all approaches.

The FBE and SF are mathematically equivalent filtering schemes. However, from Table 1, we can see that, although the FBE has lower signal delay than the correspondent SF, it is more computationally demanding (more multiplications). The LDF is a more computationnally efficient alternative to FBE, has a lower signal delay and is therefore a better trade-off between the FBE and the SF in terms of algorithmic complexity and memory requirements. If the IDFT and LDF filters have the same length ($P=M$), the delay and memory requirements are the same for each approach but the IDFT method requires even fewer operations than the LDF.

	Multiplications	Additions	Memory	Signal delay
SF	$\frac{2.L+M}{r}$	$\frac{L-M}{r}+(L-1)$	$2.L$	$L-1$
FBE	$\frac{2.L}{r}+L$	$\frac{L-M}{r}+(L-1)$	$2.L$	$\frac{L-1}{2}$
LDF	$\frac{2.L+P}{r}+P$	$\frac{L-M}{r}+(P-1)$	$L+P$	$\frac{P-1}{2}$
IDFT	$\frac{L}{r}+M$	$\frac{L-M}{r}+(M-1)$	$L+M$	$\frac{M-1}{2}$

Table 1: Number of operations required per sample for each approach to sub-band filtering. The analysis and synthesis stages are implemented through a polyphase network.

4. EXPERIMENTAL SETUP

The different filtering methods presented above are compared through simulations as described below. In Section 4.1, we describe our system setup. Section 4.2 presents the method used to generate our test signals. Our approach to assessment is described in Section 4.3.

4.1 System setup

In each case, the sub-band signals are extracted using a DFT-modulated filter bank where the prototype filter $h(n)$ of length L is defined as:

$$h(n)=\frac{1}{M}\cdot\mathrm{sinc}\left[\frac{2\pi}{M}\left(n-\frac{L}{2}\right)\right]\cdot w_L(n), \tag{7}$$

where w_L is a Hamming window also of length L. For all experiments reported here, L is equal to 128. The length of the LDF filter P is set to $L/2=64$. The number of sub-bands M is equal to 64 and was chosen to give a satisfactory compromise between acceptable frequency resolution and realistic computational complexity of the Fourier transform. When using fewer sub-bands, i.e. reduced frequency resolution, there is a noticeable degradation in performance whereas the use of more sub-bands leads to excessive computation. Finally, we use a downsampling factor of $r=M/2=32$.

4.2 Speech signals

All speech signals used in our simulations are sampled at 8 kHz. Microphone signals contain an echo-only period followed by a double talk period. The echo-only period is of interest to evaluate echo suppression whereas the double talk period is of interest to assess near-end speech quality.

The echo signal is obtained by convolving a linear or non-linear loudspeaker signal with an acoustic path response. The resulting echo signal is then added artificially to a near-end speech signal to synthesize the microphone signal. The SER of resulting signals ranges from -5dB to 10dB with the near-end speech active level being set at -26dB. The near speech and echo level are set using the ITU-T speech voltmeter [10]. The acoustic path responses used to generate the echo were measured using real mobile terminals in an office environment and are identical to those used in [11].

Loudspeaker non-linearities are simulated according to a Volterra model [12] as used in [11]:

$$x_{nl}(n)=x(n)+a\cdot x^2(n)+b\cdot x^3(n) \tag{8}$$

where $x_{nl}(n)$ is the non-linear loudspeaker signal, $x(n)$ is the linear loudspeaker signal and (a,b) are weighting components ranging between 0 and 1. For the experiments reported in this paper, we used $a=1$ and $b=1$. Other experimental work conducted by the authors of [11] shows that this configuration typifies the case when a mobile terminal is used in handsfree mode.

We used a database of 16 speech signals and results presented in this paper correspond to typical observations.

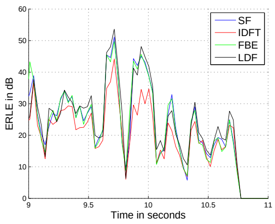
4.3 Assessment

Each filtering method is assessed stand-alone (i.e. without the AEC module) and in combination with AEC (i.e. for residual echo suppression). It is of interest to study the postfilter performance without AEC since it can be used alone in case of reduced computational load. Moreover such a configuration characterises performance when the postfilter is used for residual echo suppression while the AEC module has not yet converged. When the postfilter is used with AEC, we effectively evaluate its performance when used solely for residual echo suppression. In this case, we focus on periods where the AEC module has converged. The AEC algorithm used for all investigations reported here is the sub-band normalized least mean square method as described, in [2].

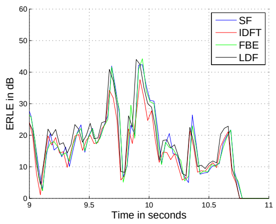
Performance is assessed in terms of echo return loss enhancement (ERLE) measurements and informal listening tests. The ERLE measures the amount of echo suppression and is defined as the energy ratio between the echo before the postfilter and that at the output. In our implementation, the ERLE is measured over windows of N samples:

$$ERLE(m)=10\cdot\log_{10}\left(\frac{\sum_N \phi^2(N)}{\sum_N \hat{s}^2(N)}\right) \tag{9}$$

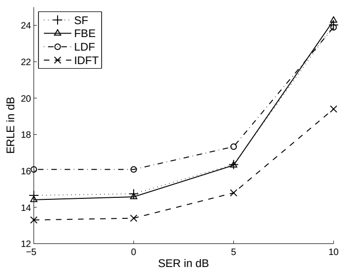
where N spans over 256 samples and $\phi(n)$ is the microphone signal $y(n)$ when the postfilter is used alone and is the residual echo signal $e(n)$ when it is used for residual echo suppression. Informal listening tests are necessary to complete the assessment of processed speech signal with subjective quality perception.



(a) ERLE against time for the different filtering methods where the postfilter is used for echo cancellation.

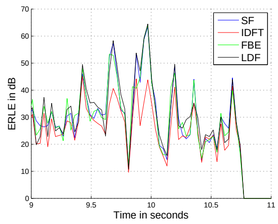


(b) ERLE against time for the different filtering methods where the postfilter is used for residual echo suppression.

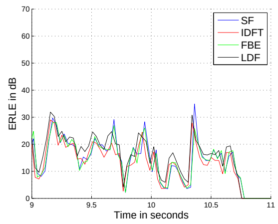


(c) Average ERLE against SER for the different methods when the postfilter is used for residual echo suppression. SERs are measured at the input of the AEC module.

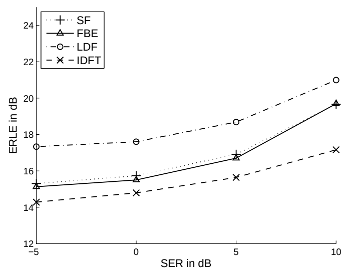
Figure 3: ERLE measurements in linear echo case.



(a) ERLE against time for the different filtering methods where the postfilter is used for echo cancelling.



(b) ERLE against time for the different filtering methods where the postfilter is used for residual echo suppression.



(c) Average ERLE against SER for the different methods when the postfilter is used for residual echo suppression. SERs are measured at the input of the AEC module.

Figure 4: ERLE measurements in non linear echo case.

5. RESULTS

In this section we present an assessment of the four different approaches to echo postfiltering presented in Section 2.3. Section 5.1 presents an assessment in the case of linear echo whereas results with loudspeaker non-linearities are presented in Section 5.2.

5.1 Simulation results with linear echo

Results where the postfilter is used stand alone and where the postfilter is used in combination with AEC are both reported here.

Postfilter without AEC: Figure 3(a) illustrates a typical ERLE profile against time during an echo-only period for each filtering method. Here, the postfilter is used without the AEC module. Figure 3(a) shows that all the filtering methods have approximately the same behavior; the curves are almost identical. Nevertheless, we can see that the IDFT filter sometimes achieves less echo suppression than do the other filtering methods. For example, at time $t = 10$s, the IDFT method achieves about 10dB less ERLE than the other methods. Periods where the IDFT method achieves less echo suppression correspond to intervals where the SER is locally low.

We note that echo remains slightly audible in processed signals, especially during double talk. Informal listening tests show that echo-only and double talk periods processed through the SF method are perceived as musical noise (random spectral peaks of short duration). In contrast, echo-only and double talk periods processed by FIR filters contain crackling noise; signals processed by FIR filters have a much smoother spectrogram than those processed by the SF method. This difference can be explained by the fact that the effective frequency response of the FIR filters is smoother compared to the original spectral gains. The crackling is slightly more perceptible in signals processed by the IDFT method: this can be explained by the fact that the effective frequency response of the IDFT filter has large variations between consecutive discrete frequency bins (here the discrete frequency bins are those where the spectral gains are calculated) where differences in gain are greater than 10dB. We can therefore assume that the use of a prototype filter in the definition of the FBE (and additional windowing for the LDF compared to the FBE) somehow results in a smoothing in the frequency domain because the effective frequency responses of the FBE and LDF do not have the large variations observed on the IDFT filter frequency responses. Nevertheless, for values of SER greater than 5dB, performance during echo-only and double talk periods is good. But, as the SER falls, the artifacts introduced during double talk (musical noise for SF method and crackling noise for FIR filters) become very annoying.

Postfilter with AEC: Figure 3(b) shows the ERLE profiles for each filtering method during an echo-only period for the sub-band echo postfilter when used for residual echo suppression (i.e. with combined AEC). Again, all the filtering methods have the same behavior, with a reduced gap between the IDFT method and the others. Improvements obtained with the IDFT method is due to the fact that the AEC improves the SER at the input of the postfilter.

Figure 3(c) shows mean ERLE against SER at the input of the AEC. The mean ERLE is calculated during echo-only periods after the AEC module has converged. The AEC achieves echo suppression of about 24dB. In Figure 3(c), we see that the SF and FBE methods are equivalent; they have the same ERLE. The LDF approach gives slightly better performances than FBE or SF methods in terms of ERLE whereas the IDFT method achieves an ERLE of between 1 and 3dB less than the SF and FBE methods. The overall system (AEC + postfilter) achieves an average ERLE of about 40dB when using any of the SF, FBE or LDF approaches.

Once the AEC module has converged, the differences between the four filtering methods is no longer perceptible during echo-only periods as echo is inaudible. During double talk periods, differences between the filtering methods are audible but not annoying. As is the case when the postfilter is used alone, we note the presence of small crackling between syllables (during the double talk) in signals processed by the FIR filter and the presence of musical noise in signals processed in the spectral domain.

5.2 Simulation results with non linear echo

Once again, we present results where the postfilter is used stand alone and where the postfilter is used for residual echo suppression.

Postfilter without AEC: Figure 4(a) shows the ERLE for each filtering method during an echo-only period. Here, the postfilter is used without the AEC module. Once again, all the filtering methods have the same behavior but in general, the IDFT filter achieves less echo suppression than do the others and, just as for the linear echo case, these intervals correspond to where the SER is locally low.

As for the linear case, we note the presence of crackling noise in FIR-processed signals. Signals processed by the FBE and LDF methods sound the same but differences between the IDFT method and the FBE or LDF methods are noticeable during echo-only and double talk periods. Signals processed by the IDFT method contain more echo and the crackling noise is also more audible which, in this case, is always perceived as annoying. The poor performance of the IDFT approach can once again be explained by the fact that the IDFT has difficulties to follow large gain variations. Crackling present in signals processed by LDF or FBE methods can be perceived as annoying for low SER (typically $SER \leq 0$dB). For speech signals processed in the sub-band domain, we note the presence of musical noise during both echo-only and double talk periods. The musical noise is perceived as annoying for SERs lower than 0dB.

Postfilter with AEC: Figure 4(b) shows the mean ERLE of each filtering method during an echo-only period for the sub-band echo postfilter when used for residual echo suppression. Here we take into account only the amount of echo suppressed by the postfilter. Once again, all the filtering methods have the same behavior and, as for the linear case, there are less disparities between the IDFT method and the others.

Figure 4(c) shows mean ERLE against SER at the input of the AEC. The mean ERLE is calculated on an echo-only period where the AEC module has converged to its optimal response. Here, we can see that the LDF method achieves the best results. The AEC module improves the SER at the postfilter input, which is still higher than in the linear case since the AEC achieves less echo reduction. As for the linear case (Figure 3(c)), the LDF achieves more echo reduction than the other methods at low SERs. Figure 4(c) also shows that the SF and FBE methods are still equivalent whereas the IDFT method achieves about 1 to 2dB less ERLE than does the SF method.

In the presence of non-linearities, the AEC achieves about 13dB of echo suppression which is approximately 10dB less than in the linear case. The overall system (AEC + postfilter) achieves an average ERLE of about 32dB when using the SF, FBE or LDF approaches: that is about 6dB less than for the linear case. Informal listening tests reveal that the AEC output contains mostly non-linear echo. Therefore, the difficulty when analyzing signals processed by the postfilter is to distinguish artifacts due to non-linearities from those introduced by the postfilter.

Although most of this residual echo is suppressed by the postfilter, echo is still slightly audible during echo-only periods. As for the linear case, we note the presence of crackling in signals processed by FIR filters and the presence of musical noise in signals processed in the sub-band domain. The crackling is slightly more audible for IDFT-processed signals and is annoying for SERs lower than 0dB. Compared to the case where the postfilter is used stand alone, near-end speech quality during double talk is improved. This is explained by the fact the AEC module improves the SER at the postfilter input.

6. CONCLUSION

This paper presents the first comparison of four different filtering approaches that can be used within a sub-band echo postfilter. Our postfilter is assessed as a stand alone solution to echo cancellation and in combination with an adaptive approach to acoustic echo cancellation. Both linear and non-linear echo is considered.

Results show that, for high SERs, all the filtering methods lead to efficient echo cancellation and, when used for residual echo suppression with linear echo, they all produce speech of equivalent quality. For non-linear echo the IDFT method gives equivalent performance to the other approaches at high SERs, whereas the SF, FBE and LDF approaches give better performance at lower SERs. In the presence of high non-linearities, the LDF method gives a good compromise between effective echo suppression, computional complexity and signal delay.

It is shown that the IDFT method leads to small signal delays and is the most computationally efficient. For linear echo and moderate level of non-linearities, the IDFT method gives good results in terms of ERLE and speech quality. The experimental work presented in this paper thus shows that the IDFT method is an appealing alternative to SF, FBE or LDF methods.

REFERENCES

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Wiley-Interscience, 2004.

[2] P. Degry and C. Beaugeant, "Solution to speech quality improvement in telecommunication terminals," in *ITG Fachtagung Sprachkommunikation*, October 2008.

[3] C. Beaugeant, V. Turbin, P. Scalart, and A. Gilloire, "New optimal filtering approaches for hands-free telecommunication terminals," *Signal Processing*, vol. 64, no. 1, pp. 33–47, 1998.

[4] H. W. Löllmann and P. Vary, "Uniform and warped low delay filter-banks for speech enhancement," *Speech Communications*, vol. 49, no. 7–8, pp. 574–587, 2007.

[5] K. Steinert, M. Schönle, C. Beaugeant, and T. Fingscheidt, "Hands-free system with low-delay subband acoustic echo control and noise reduction," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2008, pp. 1521–1524.

[6] H. W. Löllmann and P. Vary, "A blind speech enhancement algorithm for the suppression of late reverberation and noise," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009, pp. 3989–3992.

[7] E. Hänsler and G. U. Schmidt, "Hands-free telephones - joint control of echo cancellation and postfiltering," *Signal Processing*, vol. 80, no. 11, pp. 2295–2305, 2000.

[8] R. E. Crochiere and L. R. Rabiner, *Multirate digital signal processing*. Prentice-Hall, 1983.

[9] Y. Ephraim and D. Malah, "Speech enhancement using optimal non-linear spectral amplitude estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April 1983, pp. 1118–1121.

[10] ITU-T, "ITU-T recommendation P.56: objective measurement of active speech level," 1993.

[11] M. Mossi Idrissa, N. W. D. Evans, and C. Beaugeant, "An assessment of linear adaptive filter performance with nonlinear distorsions," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[12] A. N. Birkett and R. A. Goubran, "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects," in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1995, pp. 103 – 106.