

Research Article

Acoustic Echo Cancellation Embedded in Smart Transcoding Algorithm between 3GPP AMR-NB Modes

Emmanuel Rossignol Thepie Fapi,¹ Dominique Pastor,² Christophe Beaugeant,³ and Hervé Taddei⁴

¹Media Application Server, Mas DEV, Genband Inc., 3500 Carling Avenue, Ottawa, ON, Canada K2H 8E9

²Institut Télécom, Télécom Bretagne, UMR CNRS 3192 Lab-STICC, Université Européenne de Bretagne, Technopôle Brest Iroise-CS 83818, 29238 Brest Cedex, France

³DSP Software, Business Group Communication Solutions, Infineon Technologies France-R&D Center, GAIA-2600, Route des Cretes-06560 Sophia-Antipolis, France

⁴Huawei Technologies Co., Ltd., EU R&D Center, Riesstrasse 25, Munich 80992, Germany

Correspondence should be addressed to Emmanuel Rossignol Thepie Fapi, emmanuel.thepie@telecom-bretagne.eu

Received 11 November 2009; Accepted 5 August 2010

Academic Editor: Andreas Spanias

Copyright © 2010 Emmanuel Rossignol Thepie Fapi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acoustic Echo Cancellation (AEC) is a necessary feature for mobile devices when the acoustic coupling between the microphone and the loudspeaker affects the communication quality and intelligibility. When implemented inside the network, decoding is required to access the corrupted signal. The AEC performance is strongly degraded by nonlinearity introduced by speech codecs. The Echo Return Loss Enhancement (ERLE) can be less than 10 dB for low bit rate speech codecs. We propose in this paper a coded domain AEC integrated in a smart transcoding strategy which directly modifies the Code Excited Linear Prediction (CELP) parameters. The proposed system addresses simultaneously problems due to network interoperability and network voice quality enhancement. The ERLE performance of this new approach during transcoding between Adaptive Multirate-NarrowBand (AMR-NB) modes is above 45 dB as required in Global System for Mobile Communications (GSM) specifications.

1. Introduction

Acoustic echo (AE) is due to the acoustic coupling between the mobile device transducers. It creates a feedback of the far-end speech through the whole communication path. The far-end user experiences the annoying effect of hearing his own voice with a delay introduced by the network of around 200–500 ms. AE does not only affect the user by the disturbance it adds to a telecommunication, it also leads to suboptimal behaviour of the speech codecs, meaning additive artifacts, and lower quality. AEC algorithms are recommended to improve speech quality. They can be implemented as preprocessing inside the mobile device. Unfortunately, the low complexity constraint restricts the choice of adaptive algorithms and limits the system performance.

AEC can be performed inside the network by decoding the bitstream, performing the AEC in the time or frequency

domain and re-encoding the enhanced speech signal [1]. However, the echo attenuation can be severely degraded by the non linearity and/or unpredictability of the effective acoustic echo path [2, 3]. An alternative approach [4] involves modifying the coded parameters. In [5] a complete noise reduction system was implemented by modifying the fixed codebook gain of a CELP coder. AEC system by filtering the fixed codebook gain of the CELP coder was proposed in [6] with performance similar to that of classical Normalized Least Mean Square (NLMS) approach. The advantage of the coded domain Voice Quality Enhancements (VQEs) is their low complexity. They do not require complete decoding and re-encoding.

For communications requiring interconnection between networks using incompatible codecs, transcoding from one codec format to another is necessary at the gateways. It is usually performed by decoding one bitstream and

re-encoding it into the target codec bitstream format. It induces computational load, delay increase, and speech quality degradation. To avoid these drawbacks smart transcoding solutions were proposed [7, 8], taking advantage that most speech codecs use CELP techniques with same kind of parameters. In this paper, we propose a centralized AEC unit embedded inside a smart transcoding. Parameters extracted during the transcoding are enhanced by a coded domain AEC. We compare our system with a classical NLMS.

This paper is organized as follows. Next section gives an overview of CELP techniques. Then smart transcoding principle is discussed. AEC by filtering the fixed gain of the microphone signal is described in Section 4. The new approach of integrating the coded domain AEC in smart transcoding strategy and experimental results are discussed in Sections 5 and 6, respectively. Finally, conclusion is drawn in Section 7.

2. The AMR Speech Codec

To study our AEC system, we simulate transcoding between the Adaptive MultiRate (AMR) codec modes [9]. AMR is a multirate speech coder, ranging from 4.75 to 12.2 kbps and using algebraic CELP technique. Here we focus on transcoding between AMR 12.2 and AMR 7.4 modes. AMR uses a linear prediction analysis of order $M = 10$. The Linear Prediction Coefficients (LPCs) (a_i), $i = 1, \dots, 10$ are computed twice in AMR 12.2 and only once in other modes. The LPC coefficients are transmitted to the decoder through their Line Spectral Frequencies (LSFs) representation. The residual signal obtained after the LPC analysis is quantized in two steps. First an adaptive codebook search is performed every subframe leading to a pitch delay T and an adaptive gain value g_a . Using these parameters, a new residual signal is computed by subtracting the adaptive codebook contribution. This new target signal is used to process another codebook search, the so-called fixed codebook search [10]. The resulting parameters are the index j of that fixed codebook vector and its fixed gain value g_f . These gains are separately quantized in AMR 12.2 mode and jointly quantized in other modes. At the decoder side, the received adaptive gain, the pitch value and its fractional part are used to reconstruct the adaptive excitation $v(n)$. The received fixed codebook index and its associated gain are used to build the fixed codebook excitation $c(n)$. Both excitations are added and enter the LPC synthesis filter. Finally, a postprocessing algorithm is applied.

The AMR synthesis filter can be approximated [4] by

$$H_m(z) = \frac{g_f(m)}{(1 - g_a(m) \cdot z^{-T(m)}) \cdot (1 + \sum_{i=1}^M a_i(m) \cdot z^{-i})}. \quad (1)$$

In (1), g_f appears as a multiplication factor, hence a weighting of g_f modifies the signal amplitude. AEC applied to the fixed codebook gain is motivated by this remark and was experimented in [6].

3. Smart Transcoding

In standard transcoding approach, bitstream A is first decoded at decoder A (dec.A). The obtained signal $s_A(n)$ is then encoded by encoder B (enc.B), leading to bit-stream B . The smart transcoding principle is described in Figure 1, it exploits the similarity of the parameters transmitted inside bitstreams A and B and consists in avoiding the computation of some coded parameters at (enc.B). In this work, we restrict the smart transcoding strategy to the LPC coefficients, the fixed, and the adaptive codebook gains. At (dec.A), the LPC coefficients (A) $_{\text{dec.A}}$, the fixed gain (g_f) $_{\text{dec.A}}$, and the adaptive gain (g_a) $_{\text{dec.A}}$ are extracted. These parameters are mapped inside (enc.B) to directly compute the target parameters (A) $_{\text{enc.B}}$, (g_f) $_{\text{enc.B}}$, and (g_a) $_{\text{enc.B}}$. As a consequence, the LPC analysis is not performed anymore at enc.B . Smart transcoding is thus achieved by intelligently mapping the parameters extracted from bit-stream A inside those of bitstream B . The AMR codec modes transmit similar set of parameters: the LPC coefficients, the pitch delay, the fixed codebook vector index, the fixed and the adaptive codebook gains. The parameter estimation method, the resolutions of the estimation and the quantization technique used, differentiate one mode to another. In AMR 12.2, the fixed and adaptive gains are separately quantized, so that they are jointly quantized in AMR 7.4. Based on experiments, the following smart transcoding strategy is considered:

$$\begin{aligned} (g_a)_{\text{enc.B}}(m) &= (g_a)_{\text{dec.A}}(m), \\ (g_f)_{\text{enc.B}}(m) &= (g_f)_{\text{dec.A}}(m-1). \end{aligned} \quad (2)$$

Dealing with the AMR 12.2 or 7.4, four sets of LPC coefficients ($m = 1, 2, 3, 4$) are obtained each frame p at dec.A . At enc.B , two sets of LPC coefficients should be computed in AMR 12.2. The first set is computed at subframe 2 and the second at subframe 4. In AMR 7.4, one set only is computed, localized at subframe 4. It follows that the appropriated smart transcoding strategy is given during each frame by

$$(A)_{\text{enc.B}}(2m) = (A)_{\text{dec.A}}(2m). \quad (3)$$

In transcoding from AMR 12.2 to AMR 7.4, $m = 1, 2$ whereas $m = 2$ in transcoding from AMR 7.4 to AMR12.2. The LPC analysis and the computation of the gains are no more performed in enc.B .

4. Acoustic Echo Cancellation by Filtering the Fixed Gain

Our proposed AEC is based on the attenuation of the fixed gain of the microphone ($g_{f,y}$)(m) by a gain factor $G(m)$. $G(m)$ depends on the echo signal fixed gain ($g_{f,z}$)(m) and the estimated useful signal fixed gain ($g_{f,s}$)(m) defined as follows:

$$g_{f,s}(m) = G(m) \cdot g_{f,y}(m). \quad (4)$$

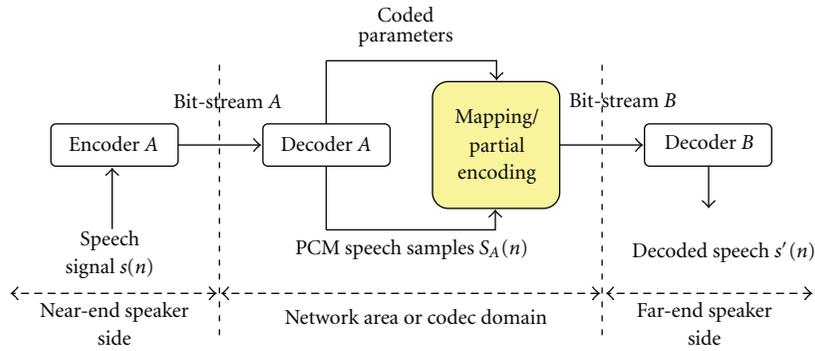


FIGURE 1: The smart transcoding principle.

Since $g_{f,z}$ and $g_{f,s}$ are unknown, $G(m)$ is computed based on a recursive estimation of the Signal-to-Echo Ratio (SER) [6]:

$$G(m) = \frac{\text{SER}(m)}{\zeta + \text{SER}(m)} \quad (5)$$

with ζ equal to 1 in single talk periods and 4/3 in double talk periods. Speech mode is detected using a normalized-cross correlation analysis of the microphone and loudspeaker fixed gains $g_{f,y}$ and $g_{f,x}$, respectively. These gains can be easily obtained from the far-end and near-end bitstream. The SER is estimated based on recursive estimation [5]:

$$\text{SER}(m) = \beta \frac{G(m-1)g_{f,y}(m-1)}{\hat{g}_{f,z}(m)} + (1-\beta) \frac{g_{f,y}(m)}{\hat{g}_{f,z}(m)} \quad (6)$$

with $\beta = 0.9$. The SER as defined in (6) requires only estimation of the echo fixed gain $\hat{g}_{f,z}$ [6].

The fixed gains are used as a sufficiently good representation of the energy. Acoustic echo presence is then detected by analyzing the relation between the smoothed gains and a threshold. An estimation of the fixed gain of the echo ($\hat{g}_{f,z}$)(m) is considered as a shifted and attenuated version of the fixed codebook gain of the loudspeaker:

$$\hat{g}_{f,z}(m) = k_{\text{opt}}(m) \cdot g_{f,x}(m - \tau_{\text{opt}}(m)), \quad (7)$$

where $k_{\text{opt}}(m)$ is the attenuation parameter and $\tau_{\text{opt}}(m)$ represents the shifted subframes. The shifted subframes $\tau_{\text{opt}}(m)$ are determined based on the normalized cross-correlation function between the loudspeaker and the microphone fixed gains. The attenuation factor derived from the ratio between the microphone gain and the shifted loudspeaker gain. This last parameter is also used to track down Double Talk periods. The microphone fixed gain $g_{f,y}$ is then replaced by the estimation of the fixed gain of the clean speech $\hat{g}_{f,s}$ obtained in (4).

5. Coded Domain AEC Embedded in Smart Transcoding

In Section 4, AEC is performed by directly modifying the fixed gain $g_{f,y}$ of the microphone signal. In this section, we

simultaneously perform coded domain AEC during smart transcoding. The fixed gain from decoder A is transmitted for each subframe to the AEC unit. After enhancement, the mapping strategy is performed as described in Section 3. The decoded signal at the far-end side is enhanced as the acoustic echo has been removed or attenuated.

5.1. Proposed Architecture. As depicted in Figure 2, during the decoding of the corrupted signal $y_A(n)$ at decoder A, the LPC coefficients A_y , fixed gain $g_{f,y}$, and adaptive gain $g_{a,y}$ are extracted. A_y and $g_{a,y}$ are directly mapped inside encoder B. At encoder B, the processing blocks (red boxes) needed are skipped, and $g_{f,y}$ is sent to the AEC unit (yellow box). The AEC unit also needs the microphone fixed gain of the loudspeaker gain $g_{f,x}$. The output of the AEC unit is the one which is mapped inside encoder B and decoder A. In our simulations, if encoder A is the AMR 12.2 then encoder B is the AMR 7.4 and vice versa.

6. Simulation Results and Comparison

Experiments were carried out by simulating a network environment. Corrupted files are divided in three groups of 5 test files each, containing both single and double talk periods. Each group of files was constructed using different car impulse responses h_i , $i = 1, 2, 3$. We measured the mean SER during double talk periods and obtained the following values: 11 dB for h_1 , 15 dB for h_2 , and 17 dB for h_3 . Our proposed system is compared to the standard NLMS [11]. With this classical approach, the corrupted speech is decoded and the NLMS is then applied to reduce the AE. Then the output if the NLMS is encoded with encoder B.

6.1. Computation Load and Delay Reduction. The LPC analysis and the preprocessing represent 20% and 7%, respectively, of the computational load during the encoding process [7]. With our approach, approximately 27% of the computational load is reduced during the decoding and encoding process, since the LPC analysis and the preprocessing are skipped at encoder B. The LPC coefficients are directly obtained from decoder A during the smart transcoding. In

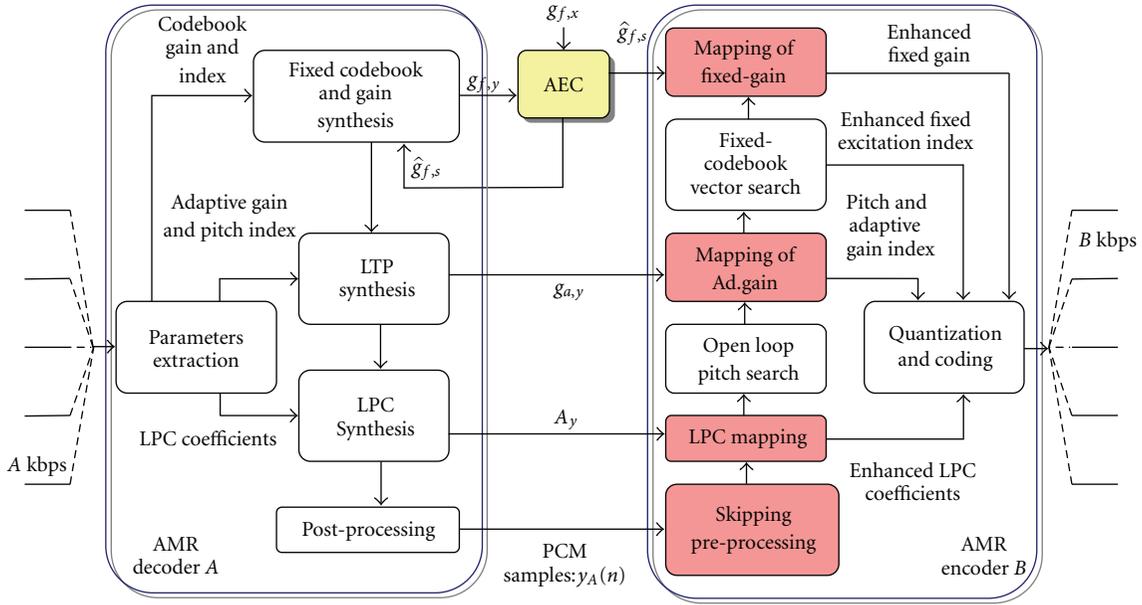


FIGURE 2: The proposed architecture.

transcoding from AMR 12.2 to 7.4, a delay reduction of 5 ms is achieved as the look-ahead required for LPC analysis is no more needed.

6.2. AEC Simulation Results. During remote single talk, that is, $s(n) = 0$, the ERLE [12] is a suitable performance measure of the AEC. If the total number of frames is $C(N_p)$, the ERLE is computed as follows:

$$\text{ERLE} = \frac{1}{C(N_p)} \sum_{l=1}^{C(N_p)} \text{ERLE}(l), \quad (8)$$

where N_p is the frame length and

$$\text{ERLE}(l) = -10 \cdot \log_{10} \left(\frac{\sum_{n=1}^{N_p} \hat{\sigma} \tau (l \cdot N_p + n)^2}{\sum_{n=1}^{N_p} y(l \cdot N_p + n)^2} \right). \quad (9)$$

τ is the delay introduced by the process. Figure 3 represents the ERLE evolution and contains both echo only ($t = 0$ to $t = 3$ seconds) and the near-end single talk ($t = 3$ to $t = 6$ seconds) and double talk ($t = 6$ to $t = 10$ seconds). Table 1 indicated that during echo-only periods, our proposed system achieved the 45 dB as required in GSM [13]. The overall result reveals that our proposed algorithm increases the ERLE compared to the standard NLMS. During echo-only periods, the gain is, in average of about 33 dB during transcoding from AMR 12.2 to AMR 7.4. The gain average is more than 42 dB during transcoding from AMR 7.4 to AMR 12.2. The high ERLE is due to that strategy used during the mapping of the codebook gain at the encoder 12.2. In fact, the mapping of the gains in AMR 12.2 has higher impact to the decoded speech than in AMR 7.4 mode.

During single talk of the near-end speaker, the ERLE characterizes the distortion introduced by the AEC algorithm. In these periods, the average ERLE is 2 dB with our

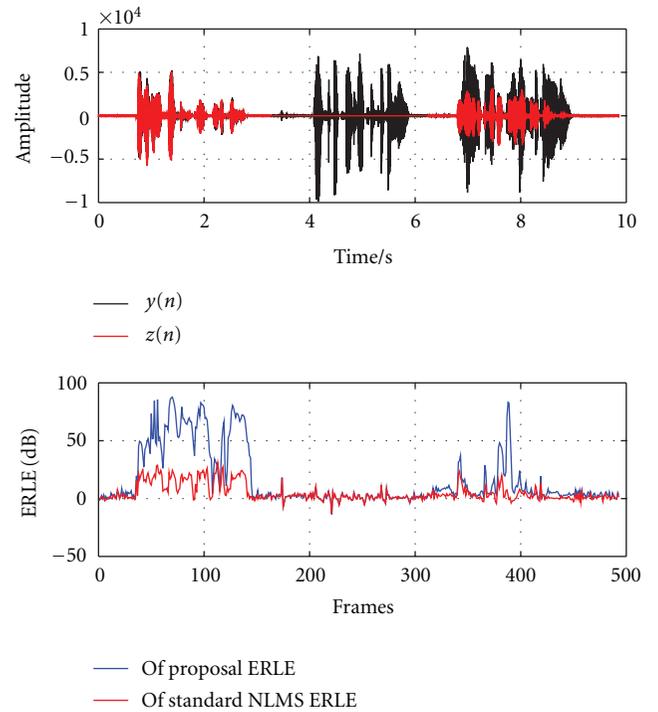


FIGURE 3: Typical ERLE evolution, AMR 12.2 kbps to 7.4 kbps.

proposal, and no audible distortion is noticed. In double talk periods, the effect of our method is noticeable. In average, the ERLE is 8 dB and 15 dB in transcoding from AMR 12.2 to 7.4 and AMR 7.4 to 12.2, respectively. The standard NLMS achieves an average of 3 dB and 1.3 dB in transcoding from AMR 12.2 to 7.4 and AMR 7.4 to 12.2, respectively. These measures reveal that our proposal can impact the acoustic echo during double talk periods.

TABLE 1: Overall Echo Return Loss Enhancement.

(a) Transcoding from AMR-NB 12.2 kbps mode to 7.4 kbps mode			
SER	11 dB (h_1)	15 dB (h_1)	17 dB (h_1)
Standard NLMS (ERLE)	11.7 dB	9.1 dB	9 dB
Proposal AEC (ERLE)	49.3 dB	40.3 dB	40.7 dB
(b) Transcoding from AMR-NB 7.4 kbps mode to 12.2 kbps mode			
SER	11 dB (h_1)	15 dB (h_1)	17 dB (h_1)
Standard NLMS (ERLE)	8.3 dB	4.6 dB	5.4 dB
Proposal AEC (ERLE)	55.7 dB	39.1 dB	49.7 dB

Our solution tends to be more robust against non linearity introduced by the codec. This observation can be explained by the fact that our algorithm only acts on the signal amplitude. During informal listening test of the processed files, and in echo only period, it follows that our approach totally cancels the echo while, with the standard NLMS, there is a noticeable amount of the remaining residual echo.

7. Conclusion

An integrated GSM network low cost AEC and a smart transcoding scheme have been presented. On the basis of our architecture, we have shown that it is suitable and beneficial to directly process the coded parameters instead of the decoded speech samples. Operations such as coding and quantization of the CELP parameters are nonlinear. These operations introduce nonlinearity to the decoded signal that may degrade the estimation of the acoustic echo path. As no acoustic echo path estimation is required in the coded domain AEC, the problems due to nonlinearity introduced by speech coders are greatly reduced. Simulation results and objective tests have confirmed that the proposed method is capable of delivering promising results at low cost. Problems due to network interoperability and network voice quality enhancement can be suitably improved.

Acknowledgment

Hervé Taddei was with Nokia Siemens Network (Munich, Germany) when this work was done.

References

- [1] X. Lu and B. Champagne, "A centralized acoustic echo canceller exploiting masking properties of the human ear," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 377–380, April 2003.
- [2] M. Rages and K. C. Ho, "Limits on echo return loss enhancement on a voice coded speech signal," in *Proceedings of the 45th Midwest Symposium on Circuits and Systems*, pp. 152–155, August 2002.
- [3] G. Enzner, H. Kruger, and P. Vary, "On the problem of acoustic echo control in cellular networks," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '05)*, pp. 213–215, 2005.
- [4] R. Chandran and D. J. Marchok, "Compressed domain noise reduction and echo suppression for network speech enhancement," in *Proceedings of the 43rd Midwest Circuits and Systems Conference (MWSCAS '00)*, pp. 10–13, August 2000.
- [5] H. Taddei, C. Beaugeant, and M. De Meuleneire, "Noise reduction on speech codec parameters," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 497–500, May 2004.
- [6] E. Thepie, C. Beaugeant, H. Taddei, N. Dutsch, and D. Pastor, "Echo reduction based on speech codec parameters," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '06)*, 2006.
- [7] C. Beaugeant and H. Taddei, "Quality computation load reduction achieved by applying smart transcoding between CELP speech coders," in *Proceedings of the European Signal Processing Conference (EUSIPCO '07)*, pp. 1372–1376, 2007.
- [8] S. W. Yoon, H. G. Kang, Y. C. Park, and D. H. Youn, "Transcoding algorithm for G.723.1 and AMR speech coders: for interoperability between VoIP and mobile networks," in *Proceedings of 8th European Conference on Speech Communication and Technology (Eurospeech '03)*, pp. 1101–1104, 2003.
- [9] 3GPP, "Mandatory speech codec speech processing functions; AMR speech codec; general description," Tech. Rep. TS 26.090, 3GPP, 1999.
- [10] W. C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*, chapter 16, Wiley Interscience, New York, NY, USA, 2003.
- [11] C. Beaugeant, M. Schoenle, H. Loellman, B. Sauert, P. Steinert, and P. Vary, "Hand-free audio and its application to telecommunication terminals," in *Proceedings of the 29th International Conference on Audio for Mobile and Handheld Devices*, September 2006.
- [12] E. Hansler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley Interscience, New York, NY, USA, 2004.
- [13] ITU-T, "ITU-T recommendation G.168, digital network echo cancellers," Tech. Rep., ITU, 2004.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

