A New Approach of Echo Cancellation Design in IP based Telephone Networks

Wolfgang Brandstätter: *Institute of Electrical Measurements and Circuit Design, Vienna University of Technology, wolfgang.brandstaetter@tuwien.ac.at*
Bernhard Handl: *Telekom Austria AG, bernhard.handl@telekom.at*
Harald Jammernegg: *AOSA Telekom, harald.jammernegg@aosa.at*
Waltraud Müllner: *Telekom Austria AG, waltraud.muellner@telekom.at*
Peter Tschulik: *Siemens AG, peter.tschulik@siemens.at*

Introduction

Due to various technical reasons like the interconnection to the traditional PSTN, a heterogeneous network design and a high and unpredictable signal delay speech communication in IP based networks (VoIP, Voice over IP) is often significantly impaired by audible echo disturbances.

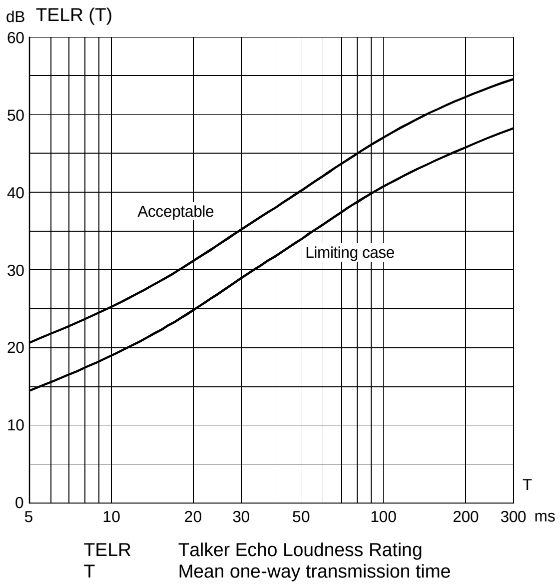

Figure 1: Talker echo tolerance curves

As shown in figure 1 the echo disturbance is mainly influenced by the two parameters echo round trip delay and echo level [1]. Typically the delay is influenced by the speech coders, the propagation delay in the network itself, delay jitter and the de-jitter buffers at the receiving side. The delay is therefore high and unpredictable in these networks, consequently echo disturbances and the need to implement high quality network echo cancellers (EC) in these connections may rise to a dominant factor determining speech quality.

Echo Canceller

Digital network ECs [2] are integrated in networks in order to eliminate reflected echoes. They consist of an adaptive filter calculating an echo model in order to subtract the estimated echo from the reflected echo. The maximum echo attenuation provided by these algorithms is not high enough to cancel the echo below the perception threshold. Additional signal processing (non-linear processing, NLP) is therefore necessary, thus introducing an additional attenuation. The design of the NLP implementation together with the robustness of the double talk detection is extremely critical. Due to the non-perfect function of the double talk detection (DTD) the near end signal is disturbed and conversational quality degradations occur. Corresponding results of subjective conversational tests for commercial echo cancellers are published in [3].

Delay-controlled Echo Canceller

The idea of this new implementation is to provide a sufficient, but limited additional attenuation by the NLP instead of a Center Clipper with adaptive clipping levels [2] as often implemented in practice. The minimum requirements on the overall echo attenuation along the echo path (TELR) as a function of the echo round trip delay are based on the "Acceptable" curve of figure 1. That guarantees that only 1% of the subscribers perceive objectionable echo.

The non-positive attenuation of the NLP (A_{NLP}) can be derived as follows:

$$A_{NLP} = ERL + ERLE + (SLR + RLR) - TELR\ (T) \qquad (1)$$

The current damping values of the hybrid and the adaptive filter – the echo return loss (ERL) and the echo return loss enhancement (ERLE), respectively – are monitored continuously in single talk situations. Furthermore the sending and receiving loudness ratings (SLR, RLR) of the terminal equipment are assumed to be fixed values.

Delay Measurement

The determination of echo delay is one of the challenging tasks within this new approach. Basically two methods are suggested:

- The echo delay can be measured by reading out the timestamps of the received VoIP data [4]. For this purpose the clocks of transmitter and receiver are synchronized with adequate accuracy by dedicated time servers in the network [5]. Alternatively the deviation from the exact time can be minimized by satellite receiver cards, but this is a very costly method. The usage of timestamps guarantees latency values for every IP packet in the receiving direction.
- Each echo canceller should be equipped with a Tone Disabler [2]. The Tone Disabler disables the EC only upon detection of a signal which consists of a 2100 Hz tone with periodic phase reversals inserted in that tone. The disabling tone can be used to measure the round-trip delay along the echo path.

It is also possible to implement a combination of both methods: Before establishing the communication path between the two parties the offset between the two clocks can be determined using a disabling tone. Based on this offset the current delays can be determined by reading out the timestamps.

Results

In a third-party listening test 21 subjects assessed binaural speech samples of a female speaker situated at the far end of the echo canceller via free-field equalized headphones [6].

The recordings have been made at a head and torso simulator [7] using a standard ISDN terminal with artificially reduced loudness ratings: SLR + RLR = 8,7 dB. The RLR value of the terminal has

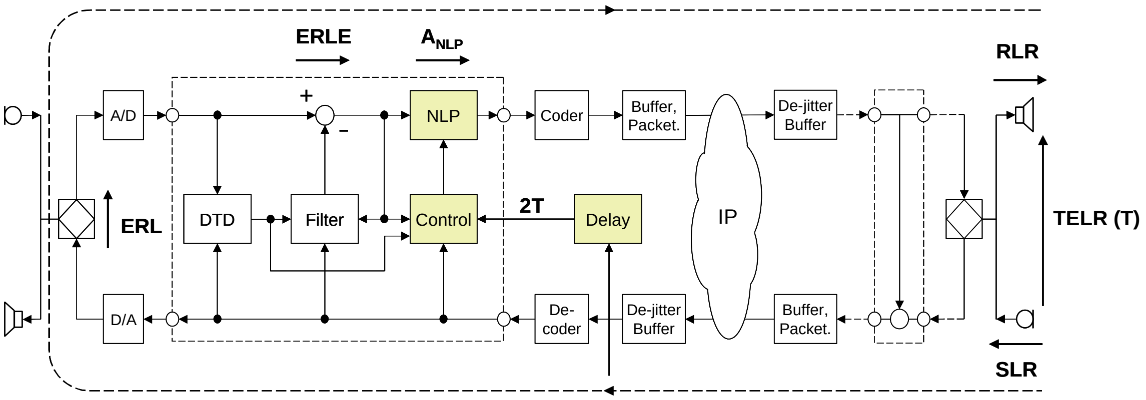

Figure 2: Delay-controlled Echo Canceller deployed in a VoIP network

been reduced by 3 dB. This value is still within the allowed range [8]. The test has been carried out according to the corresponding ITU-T Recommendation [9].

The assessment of the different aspects of the pre-recorded conversation has been made for a decreased level of the male voice at the ingress of the sending path (-6dB). That's why the MOS levels are at a quite low level compared to results obtained from undamped signal levels. The curves for the standard and the new approach in the diagrams of figure 3 converge – as expected – for higher latencies. The references have been voted under ideal conditions, i.e. the echo path has been open and the EC has been removed.

The overall quality is improved considerably for low one-way delays up to about 100 ms. The maximum difference between the standard and the new approach is 1,7 MOS at 15 ms. As the standard implementation doesn't depend on the delay in the echo path the MOS values are on a quite constant level of 2,0.

The degradation of speech quality caused by speech gaps has a similar appearance as the overall quality diagram. The standard EC performs slightly better than the delay-controlled version over the whole range of the mean one-way delays for echo disturbances caused by the female voice. The maximum deviation accounts for 0,7 MOS at 30 ms. The difference vanishes for lower latency values.

References

[1] ITU-T Recommendation G.131, Control of Talker Echo

[2] ITU-T Recommendation G.168, Digital Network EC

[3] Conversational Tests with Speech ECs – Description of Test Procedures and Results, ITU-T Contribution, Deutsche Telekom, ITU-T Rapporteurs meeting, October 1996, Jerusalem

[4] RFC 1889: A Transport Protocol for Real-Time Applications

[5] RFC 2030: Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI

[6] Auditory Judgement of Echo: Talking and Listening Tests in Comparison to Third Party Listening Test, ITU-T Contribution COM 12-16-E, February 2001, Geneva

[7] ITU-T Recommendation P.58, Head and torso simulator for telephonometry

[8] ETSI TBR 8 Ed. 2, Integrated Services Digital Network (ISDN); Telephony 3,1 kHz teleservice, October 1998

[9] ITU-T Recommendation P.831, Subjective Performance Evaluation of Network Echo Cancellers

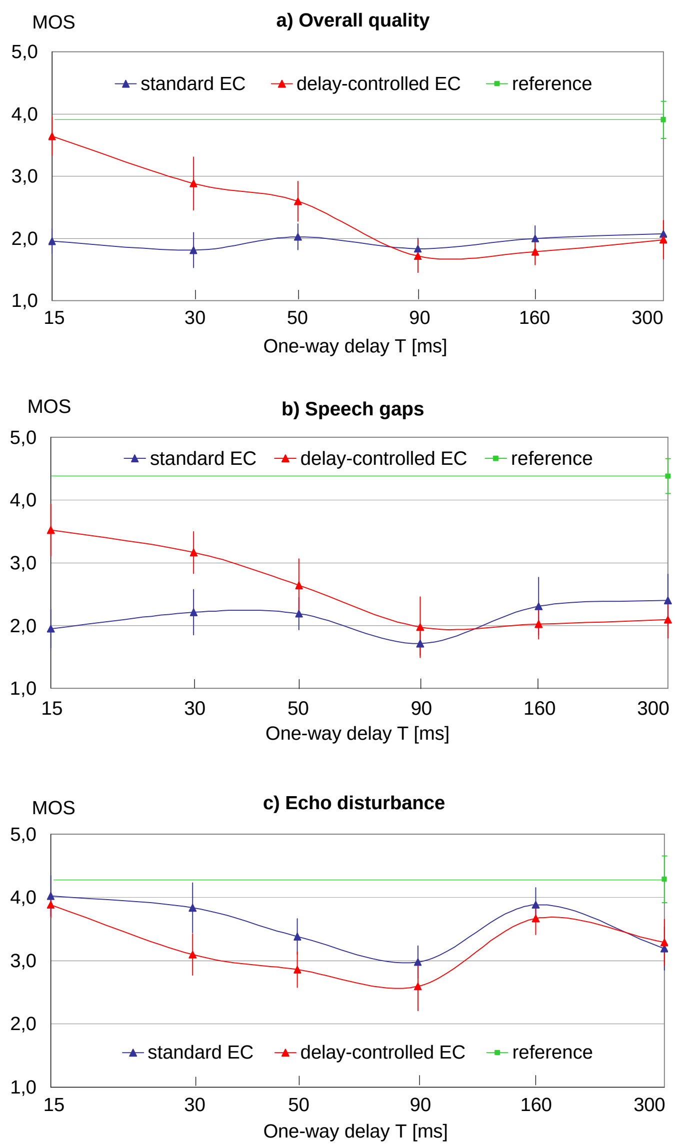

Figure 3: Interpolated MOS results of 21 subjects in the Listening Only Test for reduced level of the male voice (-6dB):
a) Overall perceived speech quality of the male voice,
b) Drop-outs and gaps in the male voice,
c) Echo disturbance caused by female voice.