PERCEPTION ORIENTED, DELAY-CONTROLLED ECHO CANCELLATION IN IP BASED TELEPHONE NETWORKS

Wolfgang Brandstätter [1], Frank Kettler [2]

[1] Institute of Electrical Measurements and Circuit Design, Vienna University of Technology, Gusshausstrasse 25/354, 1040 Vienna, Austria, wolfgang.brandstaetter@gmx.at
[2] HEAD acoustics GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany, frank.kettler@head-acoustics.de

ABSTRACT

Echo cancellation is one of the most challenging tasks in the design of IP based telephone networks. The degradation of conversational speech quality caused by impairments introduced by echo cancellers can be considerably reduced when using an echo delay controlled approach: The non-linear processor provides a sufficient, but limited additional attenuation based on the talker echo tolerance curves of the corresponding ITU-T Recommendation. The important control parameter is the network delay. While restricting the echo level below the perception threshold, the near end signal of the echo canceller is transmitted with higher quality during double talk periods, especially when the double talk detection operates unreliably.

A third party listening test has been carried out to assess different speech quality parameters during a conversation. This test focused on the comparison of an echo canceller implementation with a standard non-linear processor and a delay-controlled residual echo attenuation as suggested in this contribution. The results show substantial improvements for one-way delays up to around 100 ms.

1. INTRODUCTION

Due to various technical reasons speech communication in IP scenarios is often significantly impaired by audible echo disturbances. As shown in figure 1 the echo disturbance is mainly influenced by the two parameters delay and echo level [1]. The talker echo loudness rating (TELR) denotes the level difference between the voice and echo signals. The “acceptable” curve represents the limit for acceptable talker echo performance for all-digital networks.

Typically delay is introduced 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

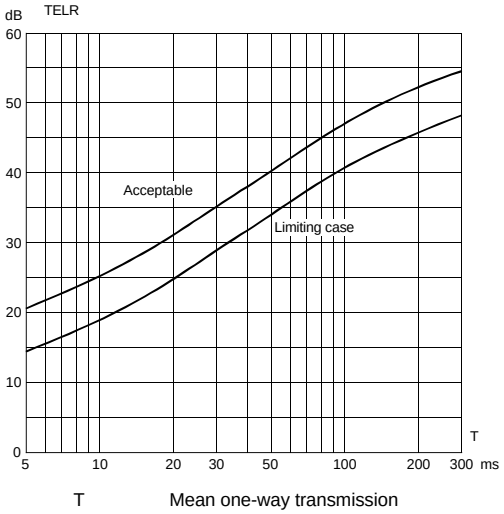

Figure 1: Talker echo tolerance curves

therefore high, time variant 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. Basically two aspects have to be considered both for network echo cancellers in IP scenarios and acoustic echo cancellers in IP terminals: On the one hand the echo attenuation has to be high enough to avoid echo disturbances even in connections introducing a high propagation delay, on the other hand the echo cancellers should not introduce undesired artifacts like syllable clipping especially under double talk conditions.

Digital network echo cancellers [2] are integrated in networks to eliminate echoes reflected at two- to four-wire hybrids. Adaptive filters are implemented 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 processor, NLP) is therefore neces-

sary, thus introducing an additional attenuation. Current specifications for network echo cancellers [2] recommend residual echo levels below –65 dBm0 independent of the network conditions like delay.

The design of the NLP implementation together with the robustness of the double talk detection (DTD) is extremely critical. Due to the imperfect function of the double talk detection the near end signal is disturbed during double talk, the background noise present at the near end subscribers location may be modulated by the NLP and conversational quality degradations occur. Corresponding results of subjective conversational tests for commercial echo cancellers are published in [3]. Objective measurement procedures determining the quality of NLP implementations are available, results – especially under double talk conditions – of echo cancellers implemented in Voice over IP (VoIP) gateways of different manufacturers are published in the Anonymous Test Report of the 2nd ETSI Speech Quality Test Event for VoIP equipment held in April 2002 [4].

2. PERCEPTION ORIENTED, DELAY-CONTROLLED APPROACH

The new concept on electric echo cancellation of an additional delay-controlled attenuation provided by non-linear processing is based on the assumption that echoes do not necessarily need to be completely suppressed (switched off), in order to avoid disturbances. Beside this application for network echo cancellers, the new approach may also be applied for acoustic echo cancellers. Due to typical lower filter attenuations provided by acoustic echo cancellers a more aggressive NLP implementation may be needed in order to sufficiently suppress residual echoes. This may lower the effectiveness of this approach compared to electric echo canceller but improvements can still be expected.

The NLP needs to provide a sufficient, delay dependent but limited attenuation in order to reduce the echo level below values as indicated by the "acceptable" curve of figure 1. Center clippers with adaptive clipping levels [2] – as often implemented in practice – typically completely suppress the residual echo under single talk conditions. Consequently echo disturbances are prevented but in case of an erroneous detection of single talk in a double talk situation the near end signal is temporarily clipped. If the attenuation inserted by non-linear processing is limited using this new approach, speech quality degradations can be minimized even when the double talk detection wrongly turns on the NLP in a double talk situation.

As shown in figure 2 the necessary attenuation provided by the NLP (A_{NLP}) can be derived from monitoring the current hybrid and the adaptive filter performance – the echo return loss (ERL) and the echo return loss enhancement (ERLE), respectively:

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

Both values ERL and ERLE have to be continuously monitored in single talk situations. The sending and receiving loudness ratings (SLR, RLR) of the terminal equipment are assumed as constant values. These values characterize the sensitivity of the terminal used at the far end side. Assumptions like the use of digital phones with its SLR and RLR values e.g. specified in ETSI TBR 8 [5] or TIA-810-A [6] have to be made in equation 1. Assuming, for example, a digital connection with an overall loudness rating (OLR) of 10 dB (SLR + RLR = 10 dB), an echo attenuation of 25 dB (ERL + ERLE = 25 dB) and a one-way delay of 30 ms or 300 ms, the necessary attenuation according to equation 1 introduced by the NLP (A_{NLP}) is 0 dB for the 30 ms delay, respectively, 20 dB for the 300 ms delay. The higher the echo delays, the more attenuation has to be provided by the NLP. Both concepts – the traditional center clipper providing always an infinite attenuation and the delay-controlled version – converge for higher network delays. In other words, the improvement of transmission quality increases for low echo delays,

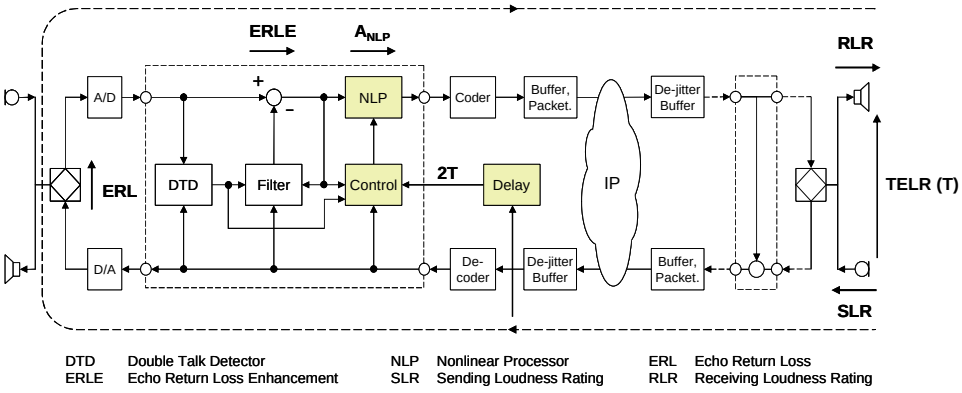

Figure 2: Delay-controlled Echo Canceller deployed in a VoIP network

as the disturbing influence of the NLP can be minimized, for low delays the NLP may even completely be disabled ($A_{NLP} = 0$) as shown in the example.

Moreover the new approach can easily be extended by considering the masking effect during double talk leading to a lower sensitivity of echo perception: Instead of disabling the NLP in double talk situations ($A_{NLP} = 0$ dB), the attenuation provided by the NLP can be reduced taking into account the additional masking caused by the double talk signal. Results from subjective tests for hands-free applications can be found in [7].

The delay-controlled approach can be applied independent of the implemented filter algorithm or the post processing component (attenuation, level switching device, Wiener Filter, ..). The additional complexity of the control mechanism can be neglected: It is limited to simple table look-up operations combined with multiplications.

3. DELAY MEASUREMENT

The control mechanism for the NLP attenuation relies on the accurate determination of echo delay. Basically this round-trip delay through the network, the signal processing components, and along the echo path can be measured in two different ways:

- An explicit measurement signal, the so-called disabling tone, provides the control mechanism with round-trip delay values.
- The timestamps of the IP packets [8] enable the calcula-tion of the one-way delay experienced through the network.

A theoretical solution would be to use the tone disabler of the echo canceller [2] e.g. before the call is set up. It disables the echo canceller 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. This can only be seen as a theoretical idea because it would significantly change the network behavior during this time period.

When using the timestamps of the received VoIP data packets, synchronized receiver and transmitter clocks are required. On the one hand both parties can gather time information from dedicated time servers in the network [9]. On the other side the use of satellite receiver cards guarantees time information with very high accuracy. Timestamps provide the control mechanism with one-way network delay values for every IP packet in the receiving direction. In addition to that the delay introduced by the signal processing components (de-jitter buffer, coding schemes, ..) and the echo-path delay at the near end have to be taken into account. Due to unsymmetrical network conditions or different de-jitter buffer sizes at both ends of the connection the system also has to monitor the propagation delay in the send direction of the echo canceller: As this concept may be implemented in corporate networks or in-house communication systems special IP packets containing information about the sending delay can be introduced. This information can be transmitted from the far end side to the near end echo canceller. Information about the absolute propagation delay in sending direction could be transmitted or – after this information has been transmitted once – it could be limited to delay changes in the network in order to guarantee the exact attenuation control.

4. RESULTS

In a third-party listening test 21 subjects assessed the quality of recorded conversations between a male and female voice transmitted over a network simulation including the different echo canceller implementations. The recordings were carried out using two artificial head measurement systems (head and torso simulators according to [10]) equipped with artificial mouth, two type 3.4 artificial ears [11] and mounted handsets, thus reproducing important characteristics like the acoustical leakage between handset and the human ear, sidetone and self-masking.

The male voice was played back by the artificial head measurement system simulating the near end subscriber. The female voice was used at the far end side. The signals at the far end side were recorded consisting of

- the female voice coupled from the artificial mouth to the open ear microphone (not covered by the handset),
- the female voice coupled from the artificial mouth to the ear microphone covered by the handset via the handset sidetone and the acoustical leakage,
- echo signals from the female voice transmitted via the receiving direction of the terminal,
- the double talk signal (male voice) transmitted via the receiving direction of the terminal.

The binaural recordings were assessed via free-field equalized headphones [12, 13]. Typical ISDN handsets according to [5] were used.

The interpolated MOS results of the listening only test are given in figure 3 to 5 for the parameters overall quality (figure 3), disturbances caused by speech gaps (figure 4) and echo disturbances (figure 5). Each diagram shows a comparison of the delay-controlled approach with a standard implementation consisting of a center clipper with adaptive clipping levels. The modification of the NLP levels within this standard approach is based on the residual echo level in single talk situations. The level of the male voice was reduced by 6 dB in order to clearly point out the room for improvement of the new approach. The references were created under ideal conditions, i.e. with

infinite ERL (open echo-path) and the echo canceller disabled.

The curves for the standard and the new approach in the diagrams of figure 3 converge – as expected – for higher network delays. The overall quality is improved considerably for low one-way delays up to 80 ms. The maximum difference between the standard implementation and the new approach amounts to 1.7 MOS at 15 ms. As the standard implementation is independent of the delay the MOS values are quite constant around 2.0 MOS. The degradation of speech quality caused by speech gaps in figure 4 leads to similar ratings as shown for the parameter overall quality in figure 3. Echo disturbances in figure 5 are slightly more annoying in case of the delay-controlled NLP attenuation. Arround 90 ms the MOS decreases due to a lower masking effect in the specific speech material used in this test. Such effects can be minimized when various speech materials are used. These results are preliminary and need to be further analyzed. The new approach could certainly be tuned leading to an echo attenuation comparable to a standard implementation without losing the advantage of minimized double talk impairments.

5. SUMMARY

The suggested implementation of a delay-controlled nonlinear processor for echo cancellers limits the echo below the perception threshold as recommended in current ITU-T specifications. Audible distortions of the near end speech can be minimized if the double talk detection operates unreliably. The network delay can be monitored from the timestamps of the speech data packets in synchronized networks e.g. equipped with simple satellite receiver cards.

The new approach leads to the same conversational speech quality results as standard implementations for higher network delays. It may significantly improve conversational speech quality in VoIP scenarios with network delays up to 100 ms. A typical application could therefore be a VoIP scenario in a corporate network or in-house communication.

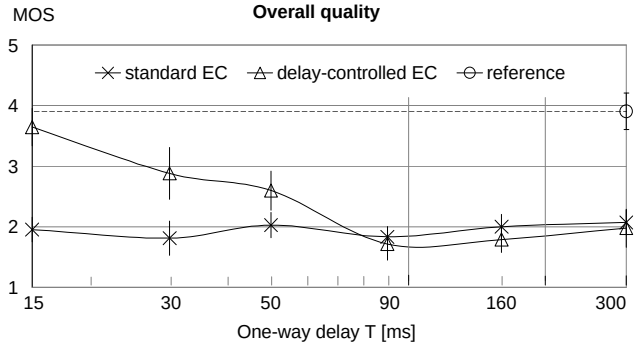

Figure 3: Overall speech quality of the male voice

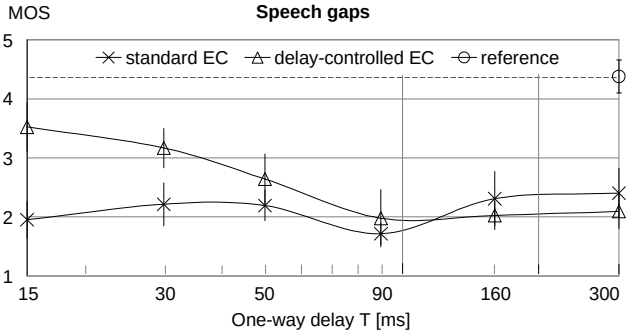

Figure 4: Drop-outs and gaps in the male voice

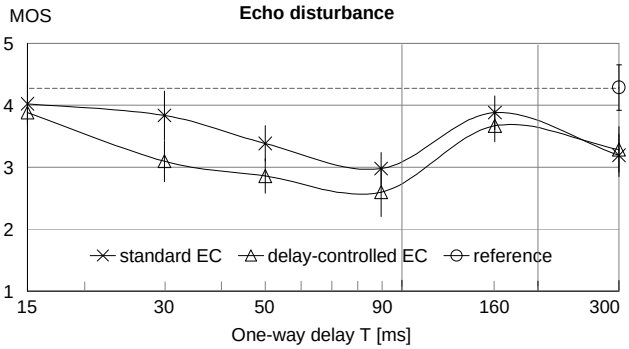

Figure 5: Echo disturbance caused by female voice

6. REFERENCES

[1] ITU-T Rec. G.131, “Control of talker echo”, Aug. 1996.

[2] ITU-T Rec. G.168, “Digital network echo canceller”, June 2002.

[3] ITU-T Contribution, “Conversational tests with speech echo cancellers – description of test procedures and results”, ITU-T Rapporteurs meeting, Jerusalem, Oct. 1996.

[4] F. Kettler et al., “Anonymized Test Report, 2nd ETSI Speech Quality Test Event for Voice over IP”, April 2001.

[5] ETSI TBR 8 Edition 2, “Integrated Services Digital Network (ISDN), Telephony 3,1 kHz teleservice”, Oct. 1998.

[6] TIA-810-A, “Transmission requirements for narrowband voice over IP and voice over PCM digital wireline telephones”, Dec. 2000.

[7] F. Kettler, H.-W. Gierlich, E. Diedrich, “Echo and level variations during double talk influencing hands-free telephone transmission quality”, IWAENC, Sept. 1999.

[8] H. Schulzrinne et al., “RTP: A Transport Protocol for Real-Time Applications”, IETF RFC 1889, Jan. 1996.

[9] D. Mills, “Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI”, IETF RFC 2030, Oct. 1996.

[10] ITU-T Rec. P.58, “Head and torso simulator for telephonometry”, Aug. 1996.

[11] ITU-T Rec. P.57, “Artificial Ears”, July 2002.

[12] ITU-T Contribution COM 12-16-E, “Auditory judgement of echo: Talking and listening tests in comparison to third party listening test”, Feb. 2001.

[13] ITU-T Rec. P.831, “Subjective performance evaluation of network echo cancellers”, Dec. 1998.