Standardised Acoustic Echo Control Tests for Mobile Telecommunication Networks

Ville-Veikko Mattila [1), and Olli Kirla [2)

[1)Nokia Research Center, (+358 50 4835443, +358 7180 35899), ville-veikko.mattila@nokia.com

[2)Nokia Networks, (+358 40 5069140, +358 7180 65039), olli.kirla@nokia.com

Abstract

Two new acoustic echo control tests – Adaptation Test and Comfort Noise Test – have been included into the ITU-T[1] draft G.160 Recommendation to set the minimum performance requirements for acoustic echo controllers (AECs) implemented in mobile communication networks. Here, Adaptation Test was designed to verify that an AEC could rapidly adapt to all combinations of input signal levels and echo paths, and that the returned echo level was sufficiently low. Additionally, Comfort Noise Test was designed to verify that both the level and spectral shape of comfort noise matched the original noise process at cancelled end[2]. This is particularly important as mobile terminals are often used in noisy locations and as here the acoustic echo has passed through a mobile terminal. Both of these tests make use of an artificial voice as the source for the echo signal, which resembles the real-world processing environment of an AEC. The tests allow the measurements to be carried out without the need of any additional control signals. Depending on the type of the connection a voice enhancement device (VED) can provide towards the radio access network, two possible test configurations are described. The adopted echo path model simulates encoded echo by using appropriate low bit-rate (LBR) codecs, and consist of a single ideal reflection with an appropriate echo return loss (ERL) and a radio access delay simulation functionality.

Keywords

Acoustic echo control, AEC, audio quality

[1] International Telecommunication Union – Telecommunications Standardization Sector.

[2] “The side of an echo canceller which contains the echo path on which this echo canceller is intended to operate. This includes all transmission facilities and equipment which is included in the echo path.” [2].

1 Introduction

Acoustic echo has a major effect on speech quality and intelligibility in mobile communications. The objectionable effect of echo results from a partial cancellation of the acoustic echo in a mobile terminal, the echo leaking from the speaker back to the microphone. In these cases, the acoustic echo controller of the terminal has not managed to model the linear echo path properly, e.g., due to its rapid changes, leading to a residual echo signal that is sent back to the network. While hearing one’s own voice is perceptually annoying, especially with high echo levels and long delays, echo can also reduce speech intelligibility in double-talk situations.

Digital network acoustic echo controllers (AEC) are designed to eliminate acoustic echo for the user. Basically, the performance of an AEC may be measured in terms of the echo level experienced by a user and its behaviour during double-talk. In the worst case, an AEC may turn a phone conversation into a half-duplex mode by muting the cancelled end while non-cancelled end[3] is talking. In addition, if an AEC is making use of comfort noise, the artificial noise process should match the level and spectral shape of the original environmental noise

[3] “The side of an echo canceller which does not contain the echo path on which this echo canceller is intended to operate.” [2].

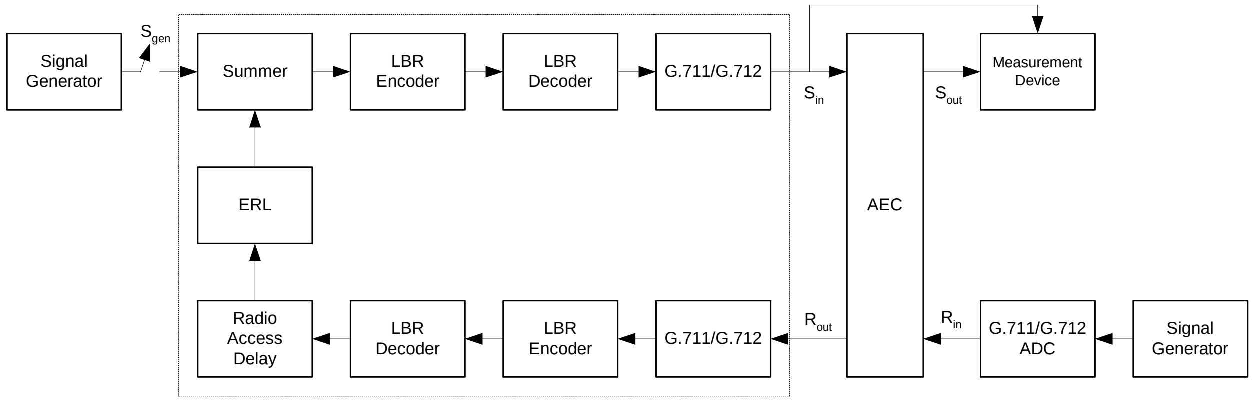

Figure 1. Test set-up in which AEC provides a G.711-connection to mobile direction.

process at cancelled end the addition of comfort noise to be indistinguishable. The transparency of environmental noise is an important issue in mobile communications because the likelihood for that a mobile terminal locates in a noisy condition is higher than that of a PSTN[4] terminal.

In ITU-T, minimum performance requirements for voice enhancement devices that implemented on network side are currently under study in the draft G.160 Recommendation [1]. In 2004, two new laboratory tests – Adaptation Test and Comfort Noise Test – were included into this Recommendation under section "Acoustic Echo Control Tests" to set the minimum performance requirements for network AECs. Here, Adaptation Test was designed to verify that an AEC could rapidly adapt to all combinations of input signal levels and echo paths, and that the returned echo level was sufficiently low. Additionally, Comfort Noise Test was designed to verify that both the level and spectral shape of comfort noise matched the original noise process at cancelled end.

This paper first describes the possible test configurations that can be applied to different telecommunication network architectures, depending what kind of connection they provide to mobile direction. Furthermore, the basic principles and structure of the tests are discussed. Finally, the performance requirements for network AECs are considered by taking into account the unique nature of the test signals used in these tests.

[4] Public-Switched Telephone Network

2 Acoustic Echo Control Tests

2.1 Test Configuration

The leading principle in the design of the test was that the required measurements could be performed directly at the selected network points without the need to use any control signals. This would make the conduction of measurements and, therefore the tests as straightforward as possible. Depending on the type of the connection a voice enhancement device (VED) can provide towards the radio access network, two possible test configurations can be considered; these are shown in Figure 1 and Figure 2 where echo path models are shown within dotted boxes. These figures closely resemble the functional diagram for echo canceller performance measurements in the ITU-T G.168 Recommendation [2]. Because returned acoustic echo is always encoded in real network implementations, an echo path model was adopted where encoded echo is simulated by using appropriate low bit-rate (LBR) codecs both in the receive and send directions of the echo path. Here, the codecs could be chosen based on what codecs are supported in the target mobile network.

In addition to LBR codecs, the echo path model was decided to consist of a single ideal reflection with an appropriate echo return loss (ERL) (see G.168 [2] for a definition) and a radio access delay simulation functionality. In this way, the total echo path delay would be caused by the radio access delay and twice the LBR codec delay.

If a VED can provide an A-law or a μ-law PCM-coded (G.711 [3]) connection towards a radio access network, the test set-up depicted in Figure 1 may be used. If a VED can only provide an LBR-encoded connection towards a radio access network, the test set-up depicted in Figure 2 may be used instead. In this case, the

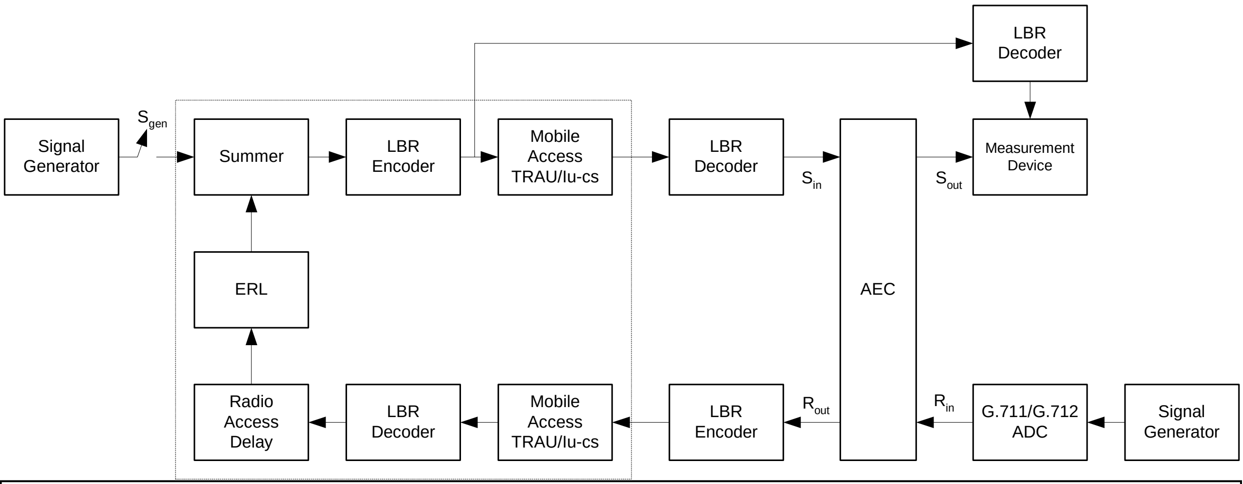

Figure 2. Test set-up in which AEC provides LBR-encoded connection to mobile direction.

measurement device should be able to decode the speech signal before any measurements can be made.

As shown in the figures, no external control is required between a VED and a measurement device. While Adaptation Test only requires access to the signal at AEC's send-out port, S_{out}, Comfort Noise Test additionally requires a connection from AEC's send-in port, i.e., a connection representing the S_{in} signal, to a measurement device.

2.2 Adaptation Test

Mobile communications sets a lot of challenges for acoustic echo control. As mobile terminals are often used on the move, echo paths tend to change continuously. Moreover, as mobile calls often originate from noisy locations, speakers' volumes tend to be high, which together with a poor seal of the speaker on one's ear, can easily lead to a significant loop back to the terminal's microphone. On network side, the reduction of acoustic echo is even more challenging as many networks components, such as speech codecs, and discontinuous transmission (DTX), transfer it to a highly non-linear process. For the perceptual quality being acceptable for the user, a network AEC should be able to adapt rapidly to all combinations of input signal levels and echo paths to guarantee that the returned echo level was sufficiently low. Adaptation Test was designed to measure exactly these two characteristics by monitoring AEC's performance both during its adaptation stage and in its steady state.

Convergence tests in G.168 for network echo cancellers make use of a composite source signal (CSS) as the test signal, R_{in}, on the receiving side. On the other hand, 3GPP Technical Specification 26.132 [4] has specified artificial voice signals, defined in the ITU-T P.50 Recommendation [5], as the training sequence for echo level measurements in handset acoustic echo control, while a logarithmically spaced multi-sine or pseudo noise is used as the actual test signal. In G.160, artificial voices of P.50 were decided to be used as the test signal in Adaptation Test to guarantee a fair assessment of those network AECs that heavily rely on the input signal being speech-like. The artificial voices of P.50 were selected, as they have appeared to well mimic the changes in signal power of real speech unlike the artificial voice samples available in the ITU-T P.501 Recommendation [6]. As in 3GPP TS 26.132, a single period of the test signal consists of the concatenation of the female and male artificial voices, having duration of about 22 seconds. This signal period is then continuously repeated throughout the test. Finally before applying to the test, the signal should be filtered by an appropriate bandpass filter, such as GSM1[5] or IRS16[6] that are available in the ITU-T G.191 Recommendation [7], to become representative of a real cellular system frequency response.

At the beginning of a call the adaptation of an AEC should be fast enough to be subjectively non-disturbing. Particularly, an AEC should not amplify acoustic echo. However, as many speech-processing blocks, such as speech codecs, may have an effect on the echo signal along the echo path in mobile communications, a minimum requirement of not amplifying the echo by more than +3 dB at any time during adaptation was adopted. Moreover, small temporal variations may also be expected in AEC's performance due to the use of the time-varying test signal. In this requirement, it should

[5] GSM Mobile station input FIR

[6] Modified IRS weighting with factor 1:1

be noted that the echo level is measured using a root-mean-squared (RMS) method having an integration time over 35 ms. In Adaptation Test 1.5 s was selected as the time limit during which an AEC should have achieved its steady state. Here, it should be noted that the test signal has a preceding silence period of about 0.3 s in its beginning before the actual artificial voice starts. Faster adaptation than this is naturally desirable, but care should be taken that no degradation is observed during single and double talk. Moreover, it is important that the stability of the AEC was maintained in all network conditions and for all voice-band signals.

After the adaptation stage, the echo level is required to be below –50 dBm0, this level guaranteeing that practically no echo is perceived. Here, it should be noted that no requirement is actually set for how much acoustic echo should be suppressed by an AEC. Therefore, depending on the suppression of the echo from AEC's receive-out port (R_{out}) to its send-in port (S_{in}), due to the transmission and loudspeaker-to-microphone coupling losses, i.e. the loss in the (cancelled end) echo path (echo return loss, ERL), an AEC can provide different levels of additional suppression. This requirement set only for the send-out port, S_{out}, is perceptually valid as the corresponding signal is what users are listening to. In Adaptation Test ERL of 30 dB was considered as a suitable minimum value for the echo path model in mobile communications.

Figure 3 summarises the adaptation requirements. The measurement of the adaptation characteristics starts at t_0 simultaneously with the start of the test signal and t_d denotes echo path delay.

In mobile communications speech levels can be surprisingly high and more often higher than in land-line phone calls. Some reasons for this were already discussed in section 1. Many network operator studies show that speech signals may temporarily reach the 0-dBm0 level in some calls. Because of this, it has been proposed that Adaptation Test should apply test signal levels as high as up to 0 dB at the receive-in port (R_{in}). It is clear that a phone call of this quality would most probably be terminated quickly without a proper AEC. Adaptation Test should, therefore, also cover these cases where the need for a proper AEC is the highest.

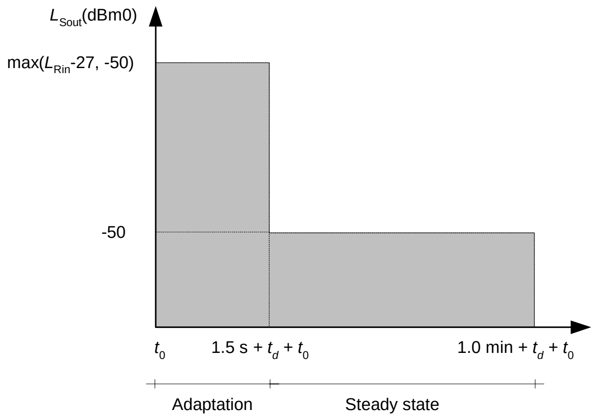

Figure 3. Adaptation characteristics.

2.3 Comfort Noise Test

As mobile calls often originate from noisy locations, the signal reaching AEC's send-in port may be noisy due to the cancelled end being under environment noise. In the echo path model in Figure 1 and Figure 2, this is simulated by introducing a noise signal generator, S_{gen}, at cancelled end. If an AEC applies echo suppression to a noisy signal, the underlying noise process can easily turn in to sounding very interrupted to a user, decreasing the perceptual quality of the phone call significantly. To give an illusion of a continuous noise process to the user, an AEC should generate comfort noise to fill in the dips caused by the suppression. Furthermore, this artificial noise at the AEC's send-out port should have the same level and spectral shape than the noise process at cancelled end, regardless of if S_{gen} was presenting a real environmental noise process or an artificial noise process due to the use of DTX.

Comfort Noise Test was added to G.160 to take into account the importance of a proper comfort noise functionality in mobile networks. This test was designed to ensure that an AEC is able to provide a comfort noise signal at S_{out} that matches noise received at S_{in}. It also tests whether the canceller is able to adjust the level and magnitude spectrum of this comfort noise signal to compensate for changes in the level and in the magnitude spectrum of input noise.

Due to similar reasons as in Adaptation Test, artificial voice of P.50 was adopted as the R_{in}-test signal in Comfort Noise Test too. The test was designed to take into account that the artificial voice signal may not keep AEC's suppressor active all the time due to the dynamic nature of the signal. Basically, the measurement periods were set to be long enough to include high-power signal periods, so that an inactive AEC that is simply passing cancelled end noise through, could not pass the test.

To test that the spectral shape of comfort noise was similar enough to the real noise present at cancelled end, the magnitude spectra of noise at AEC's send-in and send-out ports need to be measured. To keep the measurement procedure as simple as possible, but to guarantee a sufficient accuracy, the spectra were decided to be measured for three one-octave sub-bands, as defined in British Standard 61620 [8], covering the bandwidth from about 300 Hz to 4 kHz.

As in real life the noise process at cancelled end may change rapidly, e.g., as a user steps from a noisy street into a shop while talking into a phone, it was considered important that Comfort Noise Test was also testing the capability of an AEC to adapt its comfort noise to changing noise sources. To achieve this, the test was designed to introduce two different noise processes in turn at cancelled end. The level and spectral shape of the real noise process at the send-in port and of the corresponding comfort noise process at the send-out ports of an AEC are then measured and compared to test their similarity. Here, the measurement periods were not set to particularly recommend how quickly a comfort noise adaptation should happen, but to test that such a functionality would exist in an AEC. Furthermore, no strict similarity with respect to noise level and spectral shape between real noise and comfort noise was considered important as long as any mismatch would not be perceived by a user. Hence, a variation of ±4 dB is allowed for the overall noise level and ±6 dB for the level in each of the frequency subbands, when measured over a 1.4-s window. This window length was considered to provide averaging enough to adjust for any temporal level deviations. On the other hand, the window length is short enough to prevent an echo signal to have a significant effect on the averaged noise level. The overall level requirement is here 1 dB bigger than the allowed level fluctuation in the background noise transmission for hands-free terminals in the ITU-T P-340 Recommendation. This is partly due to taking into account the fluctuations possibly caused by speech codecs during the measurement periods.

The test signal level at the receive-in port (R_{in}) was set to –10 dBm0 during the noise measurements at the send-out to make sure that echo suppression would be applied, making the generation of comfort noise necessary. This together with the requirement to continue the 1.4-s measurements a total of 7 seconds in sequence should guarantee that echo suppression has to be applied and comfort noise generated. Here, it should be noted that the longest silence periods of the artificial voice test signal applied at R_{in} are only about 0.4 s.

As the noise spectrum is here measured only at three frequency bands and as some variation is allowed in these subbands, the noise sources at cancelled end were needed to represent spectra that are different enough. Otherwise, an AEC with a fixed comfort noise spectrum, representing an average of the noise spectra at cancelled end, could pass the test. As the spectra of the white noise source, defined in the P.501 Recommendation [6] and the internal moving vehicle noise source, defined in the ITU-T P.800 Recommendation [9], cannot be estimated with only one spectrum taking into account the required accuracy for the subbands, they were selected as the noise processes to be applied at cancelled end in the test. Moreover, the test was designed so that the levels of the noise sources should be different by 10 dB, which together with the required accuracy for the overall level makes it necessary for an AEC to adapt to changes in the cancelled end noise levels.

Hence, Comfort Noise Test can be summarised as a 7-step process as below:

1. Apply an internal vehicle noise source at S_{gen}.
2. Measure the level and average magnitude spectrum of the vehicle noise signal at S_{in} while keeping R_{in} silence.
3. Apply a White noise source at S_{gen} during the silence period. There should be an absolute difference of 10 dB compared to the vehicle noise level.
4. Measure the level and average magnitude spectrum of the White noise signal at S_{in} during the silence period.
5. Set the level of R_{in} to –10 dBm0. Measure the level and the average magnitude spectrum at S_{out}.
6. Apply the vehicle noise source used in point 1 at S_{gen} again after having silenced R_{in} first.
7. Set the level of R_{in} to –10 dBm0 again and measure the level and average magnitude spectrum at S_{out}.

3 Conclusion

Two acoustic echo control tests – Adaptation Test and Comfort Noise Test – have been included into the ITU-T draft G.160 Recommendation to set the minimum performance requirements for AECs operating in mobile communication networks. Here, Adaptation Test was designed to verify that an AEC could rapidly adapt to all combinations of input signal levels and echo paths, and that the returned echo level was sufficiently low. On the other hand, Comfort Noise Test was designed to verify that both the level and spectral shape of comfort noise matched the original noise process at cancelled end.

The leading principle of the design of the tests was that the measurements could be carried out without the need to use any additional control signals. This makes the arrangement of the test easier and quicker. Another principle in the design was that all the applied processes and test signals are taken from existing standards, which made the ITU-T's internal review process easier.

The acoustic echo control tests are unique in two ways; both Adaptation Test and Comfort Noise Test make use of artificial voice as the source for an echo signal. This resembles the real-world processing environment of an AEC. Secondly, Comfort Noise Test, for the first time, sets a requirement not only for the overall comfort noise level but also for its spectral shape. As here the acoustic echo that has passed through a mobile terminal is considered, this is an important requirement to guarantee an acceptable quality of experience for the user.

4 Acknowledgements

The authors would like to thank Mr. Jorma Mäkinen at Nokia Research Center for his valuable contribution to the design of the tests.

5 References

[1] ITU-T Recommendation G.160, "Voice enhancement devices", draft 12.
[2] ITU-T Recommendation G.168, "Digital network echo cancellers", 2002.
[3] ITU-T Recommendation G.711, "General aspects of digital transmission systems. Terminal equipments. Pulse code modulation (PCM) of voice frequencies", 1988.
[4] 3GPP TS 26.132, "Speech and video telephony terminal acoustic test specification", release 5, 2002.
[5] ITU-T Recommendation P.50, "Artificial voices", 1999.
[6] ITU-T Recommendation P.501, "Test signals for use in telephonometry", 2000.
[7] ITU-T Recommendation G.191, "Software tools for speech and audio coding standardization", 2000.
[8] British Standard, International Electrotechnical Commission (IEC) 61260, "Electroacoustics - Octave-band and fractional-octave-band filters", 1996.
[9] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", 1996.
[10] ITU-T Recommendation P.340, "Transmission characteristics and speech quality parameters of hands-free terminals", 2000.