International Telecommunication Union

ITU-T   Technical Paper

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(30 July 2010)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA,
DIGITAL SYSTEMS AND NETWORKS
Digital sections and digital line system – Access networks

GSTP.CSS
The composite source signal as a measuring signal and a summary of various investigations on speech echo cancellers

Summary

New ITU-T Technical Paper GSTP.CSS on "The composite source signal as a measuring signal and a summary of various investigations on speech echo cancellers" has been approved by ITU-T Study Group 16 on 30 July 2010 based on the draft in TD 253/Plen. The purpose of this technical paper is to make publicly available the information early found in COM 15-27 (1993), which is of particular interest to digital network echo cancellers implemented according to ITU-T G.168.

Change Log

This document contains Version 1 of the ITU-T Technical Paper on "*The composite source signal as a measuring signal and a summary of various investigations on speech echo cancellers*" approved at the ITU-T Study Group 16 meeting held in Geneva, 19-30 July 2010.

Editor:	Harald Kullmann	Tel:	+49 6151 628 2296
	Deutsche Telekom	Fax:	+49 521 921 00 678
	Germany	E-mail:	harald.kullmann@telekom.de

Table of Contents

List of Tables

List of Figures

ITU-T Technical Paper GSTP.CSS

The composite source signal as a measuring signal and a summary of various investigations on speech echo cancellers

Summary

This Technical Paper describes how the composite source signal (CSS) was developed as a replacement to white Gaussian noise for testing echo cancellers. The use of CSS was one of the main differences between Recommendations ITU-T G.165 and ITU-T G.168 when the latter was first approved in 1997. The Technical Paper describes how the CSS was designed to reproduce the convergence characteristics of echo cancellers with real speech. For double-talk tests a special double-talk CSS was also developed.

1 Introduction

The test method used to determine the convergence characteristics of speech echo cancellers according to ITU-T G.165 is based on band-limited White noise signal. Comparative measurements were carried out using White noise signal, artificial voice and real speech and these investigations pointed out, that convergence characteristics under real conditions cannot be determined sufficiently with White noise test signal. Consequently, Composite Source Signal (CSS) was developed and proposed that is suitable to reproduce the average convergence characteristics of speech echo cancellers.

Additional comparative echo canceller performance measurements were conducted using CSS test signal and real speech test signals on different speech echo cancellers in laboratory tests. This Technical Paper provides a brief summary of the results obtained throughout various investigations and it shows the reason for adopting CSS test signal into ITU-T G.168.

2 Problems of the test procedure according to ITU-T G.165

The test signal according to Recommendation ITU-T G.165 [1] used for measuring speech echo cancellers is a band-limited White noise signal. However, this test signal does not reproduce the echo canceller convergence characteristics that occur when speech signals are processed. Comparative measurements of echo canceller convergence characteristics using real speech and artificial voice according to Recommendation P.50 [2] resulted in a completely different behaviour of the echo canceller in comparison to using White noise test signal.

Obvious differences concern the effect of over-modulation with high signal levels, the longer adaption time, and the deterioration of echo attenuation after double talk using speech signals. Hence, using White noise signal is inadequate to determine characteristics of echo canceller's performance, objectively. These problems have their roots in the totally different signal characteristics in the time and frequency domain of the noise signal compared to speech-similar signals. In fact, White noise test signal and speech test signals have completely different characteristics in time and frequency domains and are the root cause of these totally different echo canceller performance characteristics measurements.

Consequently, this led to the demand for a new test signal (CSS) with voice-like characteristics. Such a signal has to be suitable to determine the characteristic parameters of echo cancellers. On the other hand, it must be easy to apply in terms of measurement techniques, and precisely, so it can result in reproducible test results by various test instruments and test organizations.

3 Test arrangement and new measurement signals

3.1 Test arrangements

The test arrangement used in the investigations is given in Figure 1. All test sequences were generated and captured using a PC based test instrument that allowed appropriate calibration and use of any user specified test sequences.
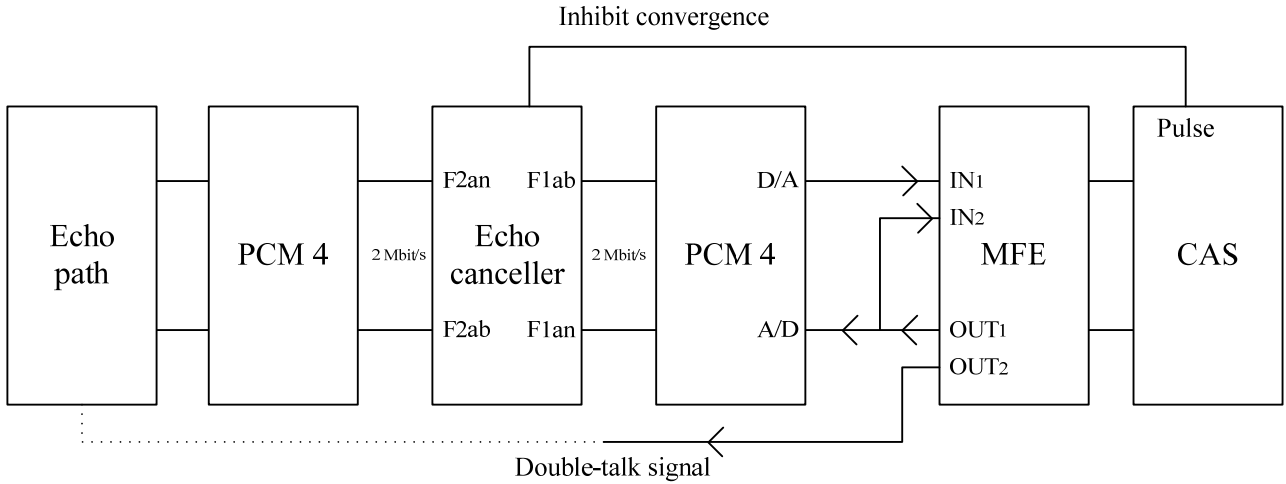

Figure 1: Test arrangement for measuring speech echo cancellers

An additional communication channel with DUT allows the control of the adaptive filter, which inhibits the coefficients. The different echo cancellers were connected to two PCM channel banks (depicted as PCM-4 in Figure 1) and each 64 kbit/s channels could be accessed through the corresponding A/D and D/A codecs. Investigations were carried out using additional echo path delay simulations, and hybrid simulators. A digital echo path simulator, and an analog hybrid termination with an echo return loss of 6 dB or 7 dB were used in these tests, respectively.

3.2 Adaptation of the composite source signal for measuring speech echo cancellers

The convergence characteristics of speech echo cancellers depend on the time signal at the echo canceller's inputs. These signal-dependent differences occur not only regarding the residual echo attenuations, but they are evident when examining the temporal behaviours such as; convergence speed, the function of the nonlinear processor, or the echo canceller performance under double talk conditions.

Therefore, to determine the average convergence characteristics of an echo canceller, it is necessary to carry out measurements using different speech signals. First the average convergence characteristics of three different echo cancellers were determined using English, French, Japanese and German speech samples. Test sentences from the NTT-AT Multi-Lingual Speech Database [3] were used in these assessments.

Based on the outcome of these comparative EC performance analysis, the CSS was adapted to reproduce the echo canceller convergence characteristic measurements. The Composite Source Signal was developed to determine the transfer characteristics of voice-controlled devices with time-varying and nonlinear behaviour. In comparison to the original shape this signal was adapted during the measurements of echo cancellers with slight modifications as described in the following list. It consists of the following segments:

− A voiced segment derived from artificial voice, according to Recommendation P.50 [2], periodically repeated for a duration of 50 ms,

- A pseudo noise sequence as a measuring signal that consists of 200 ms duration. To approximate the average long term spectrum of speech this spectrum was shaped with a 5 dB/Octave attenuation characteristic.

- A pause of 100 ms to approximate gaps of real speech.

This sequence is shown in Figure 2. The voiced part and the pseudo noise sequence were chosen to have identical signal levels. To prevent any residual DC offset, this signal was inverted, and appended once to the original signal. The resulting test signal sequence length is 700 ms. This sequence is periodically repeated during the measurements.
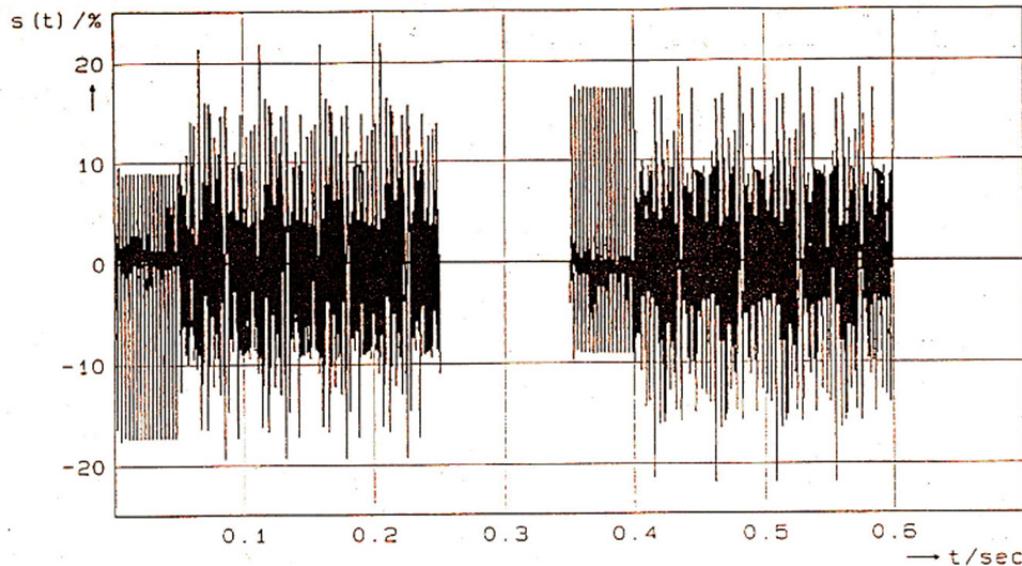

Figure 2: Composite Source Signal

The band-pass limitation was chosen from 200 Hz to 3600 Hz as given through the test arrangement. All measurements with the CSS were carried out during the pseudo noise signal over a sequence of approximately 186 ms, i.e. four 2048-point FFTs using a sampling rate of 44.1 kHz.

This signal has several advantages:

- It is precisely defined in a short format and it is easily reproducible.

- Only a short average time of 186 ms is necessary to determine transfer characteristics.

- It reproduces characteristics of real speech, such as voiced and unvoiced parts, long term spectrum and pauses.

- It consists of well defined segments that allow detailed examination of the residual echo signal.

- It does not reproduce only a certain speech.

Note: In order not to have correlated measurement signals within the adaptation window a longer PN-sequence in ITU-T G.168 and P.501 was included. For adaptive systems such as echo cancellers the FFT length should be extended to 8192 points when using 44.1 kHz sampling rate as described in Table C.2/G.168 [4] and Table 3/P.501 [5] Table of filter corner frequencies.

3.3   Simulation of double talk conditions

For the examination of echo canceller double talk performance, a second signal should be presented to the near-end. Measurements based on the Composite Source Signal therefore require a secondary test signal to approximate the conditions that occur with speech under double talk operations.

Using speech signals the male and female voices were used to simulate double talk. Again, an average convergence behaviour was derived. The following may be observed when examining the time pattern of speech signals that occur under double talk conditions:

There are sections in the time sequences where the single talk speech signal is at a high level whereas the double talk speech signal is at a low level and vice versa. There are also periods where high levels as well as low levels for both speech signals occur at the same time.

In order to simulate Double Talk events also the second (double talk) CSS was developed that consists of the following segments:

- A voiced segment that consists of 75 ms duration.

- This segment is also derived from artificial voice, but uncorrelated to the one used in the single talk CSS.

- A White noise signal segment that consists of 200 ms duration.

- This segment is uncorrelated to the pseudo noise signal used in construction/definition of the single talk CSS. However, the spectrum was also shaped (or attenuated) at a rate of 5 dB/Octave.

- A pause segment consisting of 125 ms duration.

Figure 3 shows this signal. It is band-limited as well, inverted, appended to be free of offset and repeated periodically. The levels of the voiced and unvoiced parts were again chosen to have identical levels. Due to the different length of this sequence (800 ms) similar time pattern as described above for speech signals can be modelled with this combination, if both signals are repeated periodically.
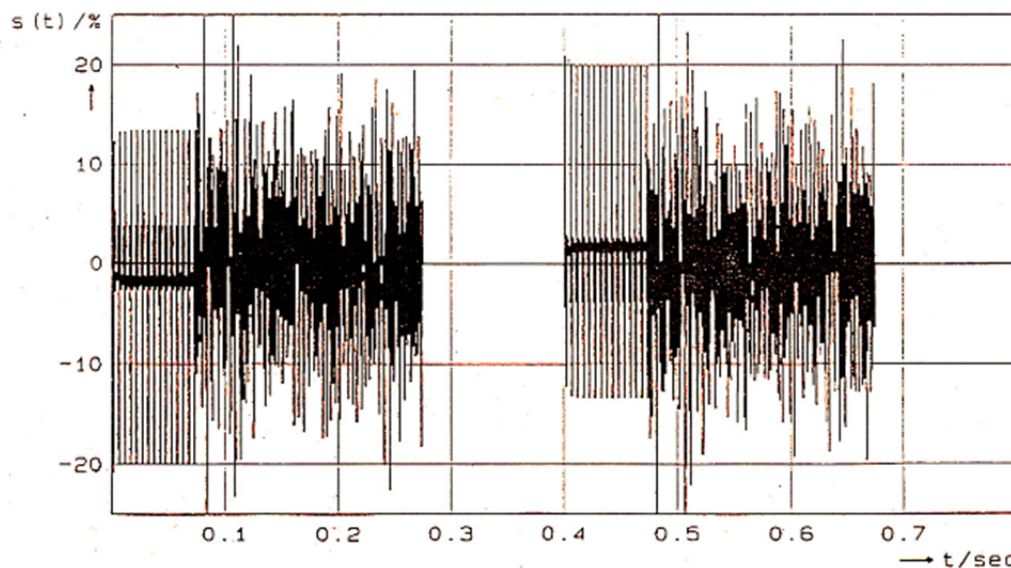

Figure 3: Composite Source Signal to simulate double talk

4   Results of various measurements

4.1   Comparative measurements with the composite source signal under single talk conditions

First the average convergence characteristics under single talk conditions were determined with different speech samples. The echo cancellers under test differ in their behaviour. Comparative measurements were carried out with the suggested Composite Source Signal. A digital echo path with an echo return loss of 6 dB and an additional echo delay of 48 ms was used.

Table 1 and 2 show the echo return loss enhancements measured for the two echo cancellers tested. Different speech signals (average results obtained from English and German, each male and female voice) and the CSS were used and fed with different input levels (Pe). The nonlinear processor (NLP) was enabled and disabled. The convergence was inhibited after 40 s to obtain the following results.

Table 1: Echo return loss enhancement (ERLE) after 40 s, echo canceller No.1

Pe	NLP disabled		NLP enabled	
	speech	CSS	speech	CSS
-15 dBm0	28.9 dB	26.8 dB	61.7 dB	54.2 dB
-20 dBm0	27.3 dB	24.8 dB	53.5 dB	46.7 dB
-30 dBm0	23.4 dB	23.3 dB	40.0 dB	37.4 dB

Table 2: Echo return loss enhancement (ERLE) after 40 s, echo canceller No.2

Pe	NLP disabled		NLP enabled	
	speech	CSS	speech	CSS
-10 dBm0	32.3 dB	31.3 dB	50.7 dB	52.8 dB
-20 dBm0	28.5 dB	35.0 dB	40.6 dB	42.8 dB
-30 dBm0	20.3 dB	26.9 dB	31.0 dB	32.6 dB

Comparative measurements with an input level of -10 dBm0 could not be carried out on echo canceller No.1 because this level leads to an over-modulation of the echo canceller when using English speech signals for measuring the ERLE. Therefore the first level was chosen to -15 dBm0. The results show a good correspondence for the CSS and speech.

To evaluate the echo canceller rate of convergence, the echo signals were recorded without inhibiting the coefficients of the adaptive filter. Examinations about operating time of the nonlinear processor in the time and frequency domain can be carried out based on spectral representations of the residual echo signals as shown in the lower parts of the following figures. Dark colours represent low levels, high residual levels are shown in lighter colours. The input signal is given in the upper parts. The convergence was not inhibited in these investigations.

Figures 4 and 5 show the convergence of echo canceller No.1 for the first 2 s approximately at an input level of -20 dBm0 for speech and the CSS. It is obvious that in the beginning the nonlinear processor is only enabled during pauses or time sequences with low signal levels. The black colour in the spectral representation during these sequences represents the residual low level.

That leads to an enabling and disabling of the nonlinear processor during the different time pattern until the echo attenuation is good enough to keep it enabled. This goes for speech and the CSS.

The same measurements were carried out with other speech echo cancellers. The following Figures 6 and 7 represent the same conditions for echo canceller No.2 at the beginning of the adaption. The convergence time is considerable longer, the NLP is finally enabled after 4 seconds, approximately.

The cancellers under test differ in their behaviour, but a good correspondence between the results measured with speech and the CSS can be noticed for each echo canceller.
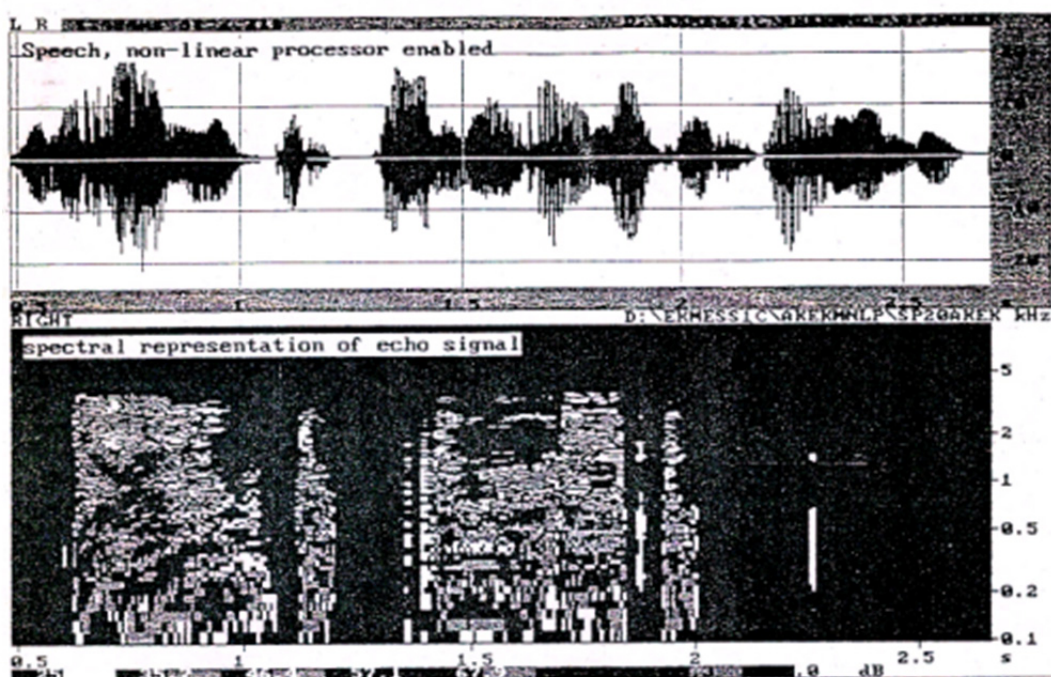
**Figure 4: Convergence with nonlinear processor for speech,
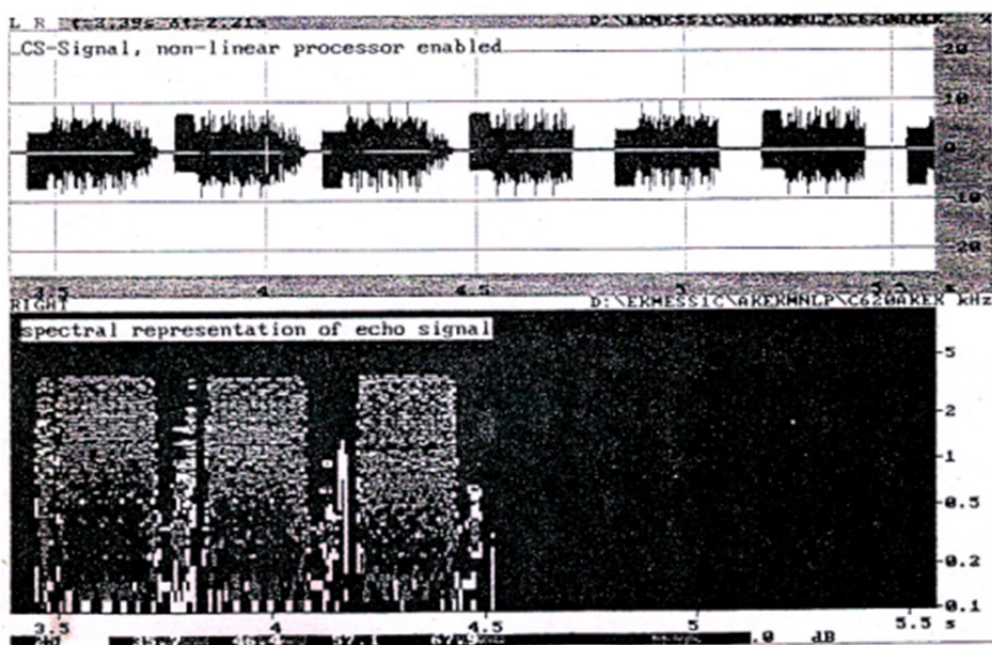echo canceller No.1, -20 dBm0**

**Figure 5: Convergence with nonlinear processor for CSS,
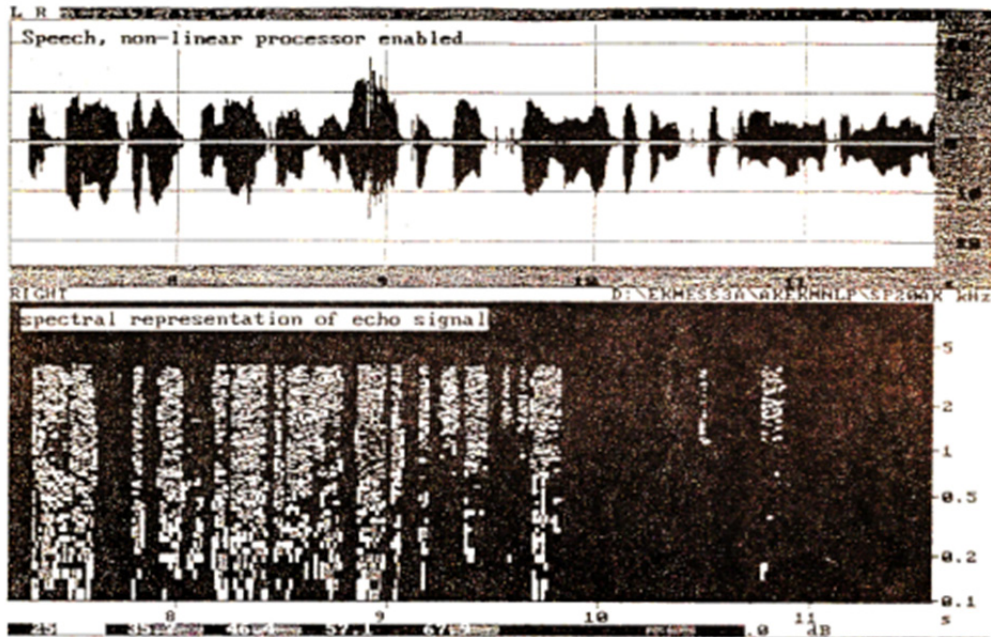echo canceller No.1, -20 dBm0**

**Figure 6: Convergence with nonlinear processor for speech,
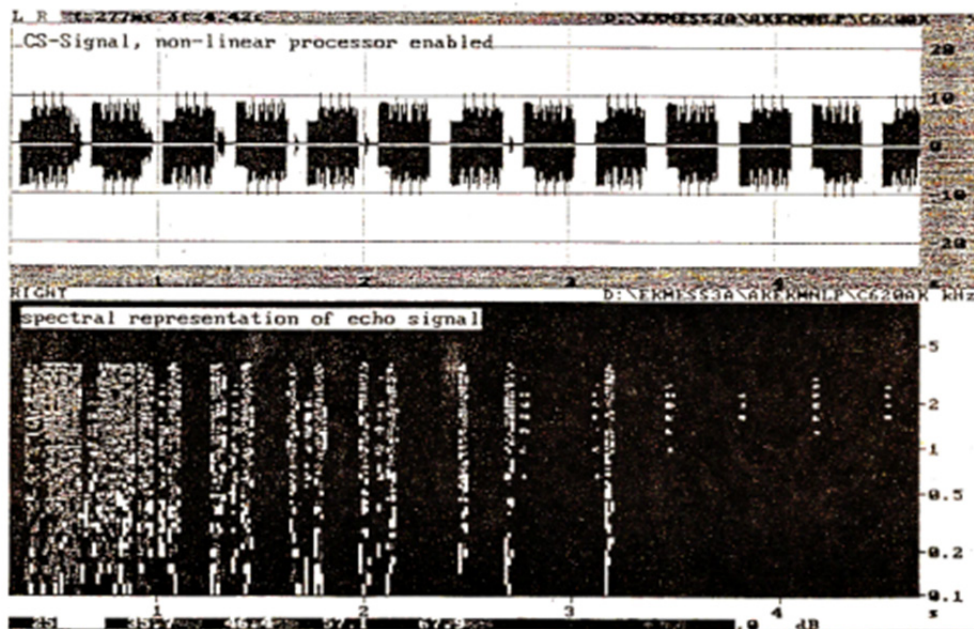echo canceller No.2, -20 dBm0**

**Figure 7: Convergence with nonlinear processor for CSS,
echo canceller No.2, -20 dBm0**

A closer examination of the nonlinear processor performance can be seen in Figures 8 and 9 using the CSS. The figures are taken from measurements on echo canceller No.1 with an additional echo delay of 48 ms and an input level of -20 dBm0. The input signal is shown in Figure 8 and the residual echo signal measured at the echo canceller's output in the send path is presented separately in Figure 9. The following can be noticed:
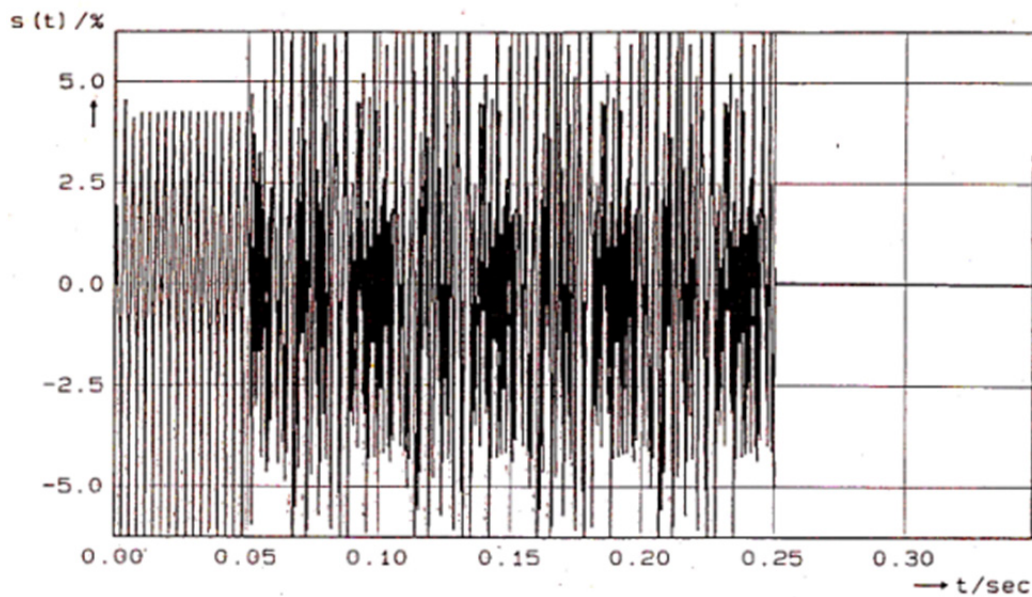

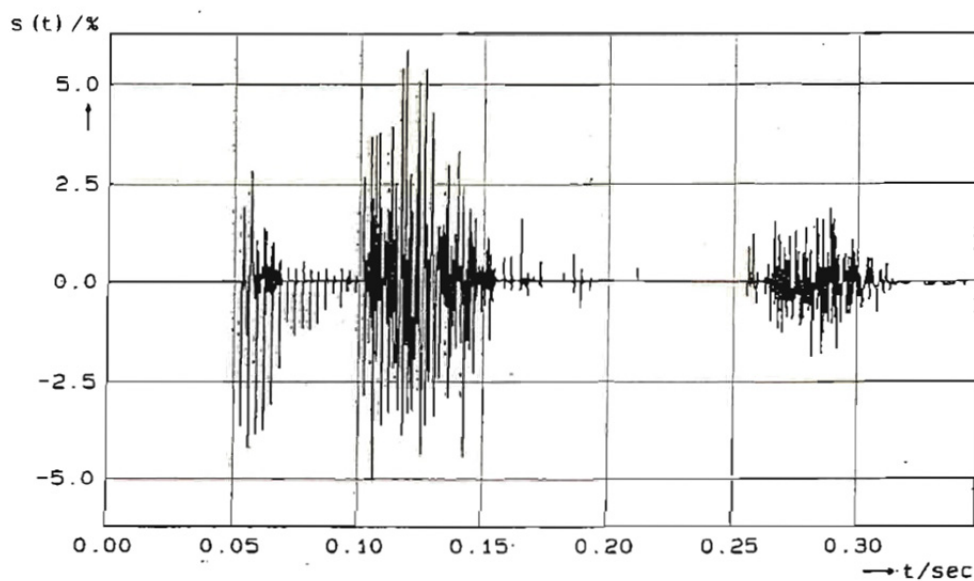**Figure 8: Input sequence (CSS)**

Figure 9: Residual echo signal of the CSS, convergence with nonlinear processor

After the delay time of 48 ms the voiced part of the input CSS appears at the output, the echo canceller starts to converge, which leads to a reduction of the echo signal. With the pseudo noise sequence at the echo cancellers input, the excitation of the adaptive filter changes. This leads to a higher residual echo signal and the echo canceller converges again.

After a short period, the residual echo level is below a certain threshold level to enable the nonlinear processor. Hence, it suppresses the echo signal. At the beginning of the pause of the CSS there is no excitation at the echo cancellers input in the receive path. Due to the time delay in the echo path, the pause is not yet present at the input in the send path. The echo of the pseudo noise sequence is still

present here and cannot be compensated because of the missing excitation. The level of this echo signal exceeds the threshold of the nonlinear processor. The processor is disabled and this part of the echo signal appears in the send path. This mechanism is the same when real speech is processed but it can better be examined using the CSS with its exactly defined segments.

4.2   Comparative measurements with the composite source signal under double talk conditions

To investigate the performance degradation of the residual echo after a double talk operation, the echo cancellers were fully converged prior to introducing 2 seconds of a double talk signal at the near end with the same level, adaption was inhibited, and the residual echo degradation level was measured. Again comparative measurements were carried out with different speech signals. The echo return loss of the analog hybrid termination was set to 7 dB.

Table 3 shows the differences in echo attenuation before and after the double talk operation for echo canceller No.1.

Measurements could now be carried out on echo canceller No.1 with an input level of -10 dBm0 without over-modulation. The reason for this may be found in the analog hybrid termination with the echo return loss of 7 dB. The deterioration values after double talk scatter over a large range especially at the input level of -10 dBm0. Therefore the average values for all speech signals are also mentioned in the table. The results for the CSS represent an average over six measurements. The outcome of test using speech samples and the CSS demonstrates a good correlation.

Table 3: Deterioration of echo level, echo canceller No.1

Pe	German	English	French	Japanese	Average speech	CSS
-10 dBm0	21.3 dB	9.2 dB	9.3 dB	1.6 dB	10.3 dB	10.6 dB
-20 dBm0	5.6 dB	5.6 dB	6.8 dB	0.5 dB	4.5 dB	5.6 dB
-30 dBm0	9.5 dB	2.5 dB	1.2 dB	0.4 dB	3.4 dB	3.5 dB

Table 4: Performance degradation of echo level, echo canceller No.2

Pe	German	English	French	Japanese	Average speech	CSS
-10 dBm0	16.3 dB	11.5 dB	8.4 dB	14.3 dB	12.6 dB	7.6 dB
-20 dBm0	17.6 dB	11.0 dB	9.9 dB	8.2 dB	11.7 dB	12.1 dB
-30 dBm0	4.0 dB	5.1 dB	5.5 dB	3.4 dB	4.5 dB	2.2 dB

The examinations made on echo canceller No.2 are represented in Table 4.

Again the deterioration values for speech (in average) and the CSS are close together. Only for the input level of -10 dBm0 the averaged deterioration of the residual echo level using the CSS is less than the average obtained with speech signals.

Figures 10 and 11 demonstrate the deterioration of the residual echo levels for speech (examples shown for English) and the suggested CSS in the frequency domain. The results are derived from echo canceller No.2 for the input level of -20 dBm0 (echo canceller No.1 shows a similar behaviour). It can be seen that the use of the CSS (see Figure 11) leads to an identical deterioration of the residual echo level as compared to natural speech (Figure 10). Changes in the spectra occur mainly below 1 kHz.
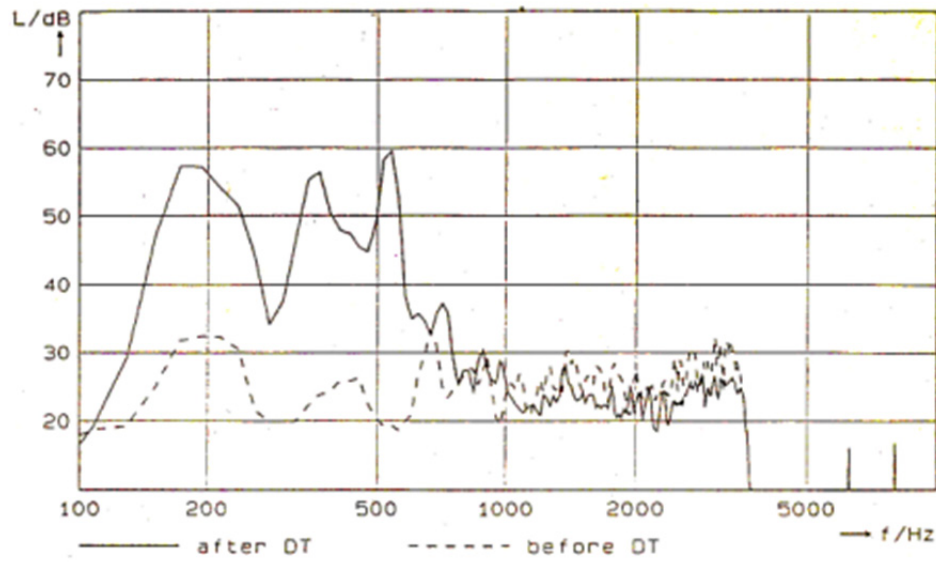
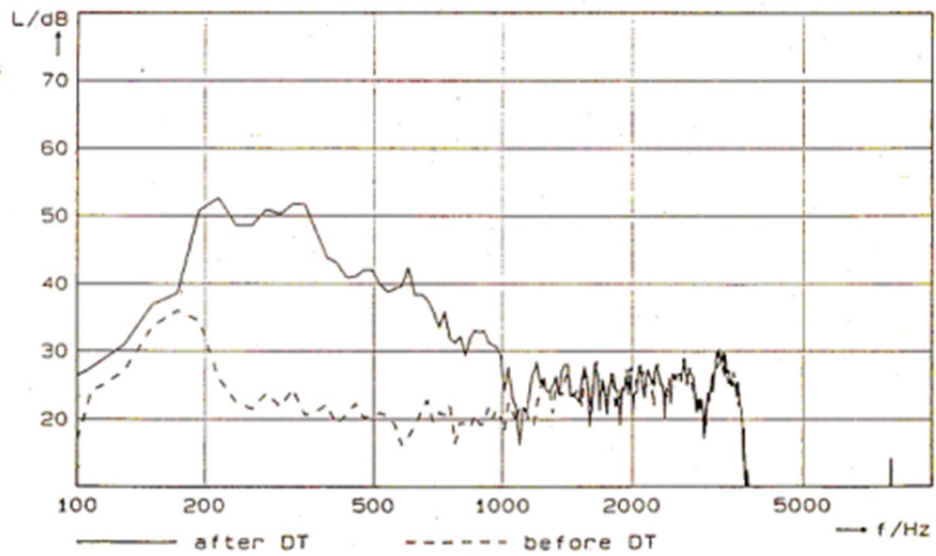**Figure 10: Deterioration of echo level after double talk, speech, -20 dBm0**

Figure 11: Deterioration of echo level after double talk, CSS, -20 dBm0

5   Conclusions

The band-limited White noise signal that is used in ITU-T G.165 to measure echo cancellers in the telephone network does not reproduce actual real speech characteristics. The problem arising from the fact that, the convergence parameters of echo cancellers occurring under real conditions processing speech samples can not be determined using this test procedure.

Measurements on different echo cancellers were carried out with several speech samples to derive average convergence behaviour. Based on these data, the Composite Source Signal was adapted to reproduce these characteristics. To evaluate the performance under double talk a combination of two Composite Source Signals simulates double talk conditions. The comparative measurements with the Composite Source Signal and real speech show a good correspondence for different echo cancellers. The convergence characteristics that occur with real speech can be reproduced very well. This test signal consists of exactly defined parts and allows detailed examinations of the residual echo signal in the time and frequency domain and the convergence parameters of echo cancellers.

References

[1]   Recommendation ITU-T G.165 (1993) "Echo cancellers"

[2]   Recommendation ITU-T P.50 (1999) "Artificial voices"

[3]   Multi-lingual speech database for telephonometry (1994), NTT Advanced Technology Corporation (NTT AT), Tokyo, Japan

[4]   Recommendation ITU-T G.168 (2009) "Digital network echo cancellers"

[5]   Recommendation ITU-T P.501 (2009) "Test signals for use in telephonometry"

_____