INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.168

(04/97)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

International telephone connections and circuits – Apparatus associated with long-distance telephone circuits

Digital network echo cancellers

ITU-T Recommendation G.168

(Previously CCITT Recommendation)

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
General definitions	G.100–G.109
General Recommendations on the transmission quality for an entire international telephone connection	G.110–G.119
General characteristics of national systems forming part of international connections	G.120–G.129
General characteristics of the 4-wire chain formed by the international circuits and national extension circuits	G.130–G.139
General characteristics of the 4-wire chain of international circuits; international transit	G.140–G.149
General characteristics of international telephone circuits and national extension circuits	G.150–G.159
Apparatus associated with long-distance telephone circuits	**G.160–G.169**
Transmission plan aspects of special circuits and connections using the international telephone connection network	G.170–G.179
Protection and restoration of transmission systems	G.180–G.189
Software tools for transmission systems	G.190–G.199
INTERNATIONAL ANALOGUE CARRIER SYSTEM	
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA CHARACTERISTICS	
DIGITAL TRANSMISSION SYSTEMS	
TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999

For further details, please refer to ITU-T List of Recommendations.

ITU-T RECOMMENDATION G.168

DIGITAL NETWORK ECHO CANCELLERS

Source

ITU-T Recommendation G.168 was prepared by ITU-T Study Group 15 (1997-2000) and was approved under the WTSC Resolution No. 1 procedure on the 8th of April 1997.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, the ITU had/had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Page

Page

Recommendation G.168

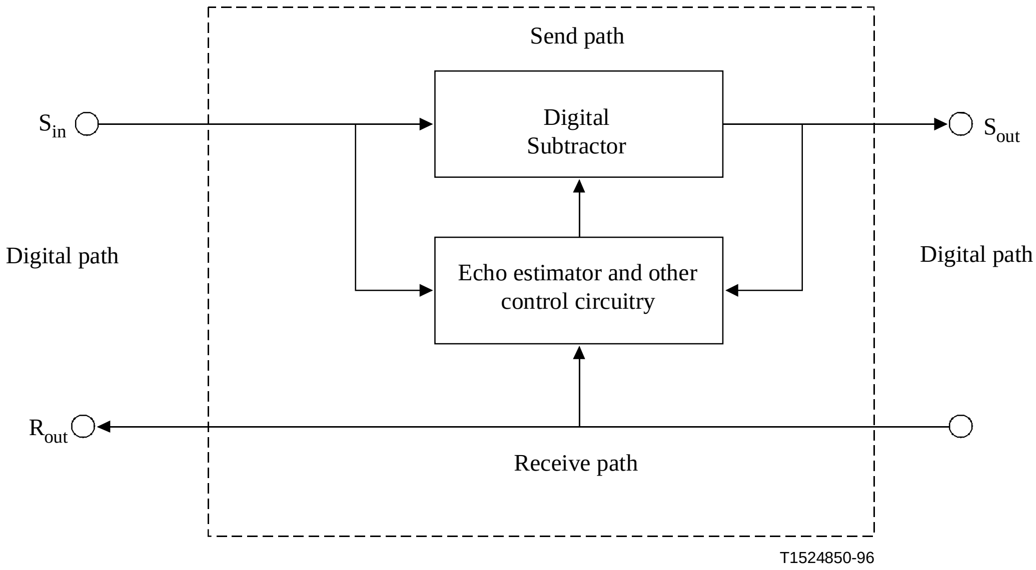
DIGITAL NETWORK ECHO CANCELLERS

(Geneva, 1997)

1 Introduction

1.1 Scope

Echo cancellers are voice operated devices placed in the 4-wire portion of a circuit (which may be an individual circuit path or a path carrying a multiplexed signal) and are used for reducing the echo by subtracting an estimated echo from the circuit echo. They may be characterized by whether the transmission path or the subtraction of the echo is by analogue or digital means (see Figure 1).



NOTE – Functionally, a Digital Echo Canceller (DEC) interfaces at 64 kbit/s. However, 24 or 30 digital echo cancellers for example may be combined corresponding to the primary digital hierarchy levels of 1544 kbit/s or 2048 kbit/s, respectively.

Figure 1/G.168 – Digital transmission path echo canceller using digital subtraction

This Recommendation is applicable to the design of echo cancellers using digital techniques, and intended for use in circuits where the delay exceeds the limits specified by Recommendations G.114 and G.131. Echo cancellers designed to this Recommendation will be compatible with each other, with echo cancellers designed in accordance with Recommendation G.165, and with echo suppressors designed in accordance with Recommendation G.164. Compatibility is defined in 1.4/G.164. Freedom is permitted in design details not covered by the requirements. This Recommendation is for the design of digital echo cancellers and defines tests that ensure echo canceller performance is adequate under wider network conditions than specified in Recommendation G.165, such as performance on voice, FAX, residual acoustic echo signals and mobile networks.

This Recommendation does not apply to echo cancellation through active 2-wire/4-wire hybrids or 2-wire repeaters. This Recommendation does not cover acoustic echo cancellation as per G.167.

This Recommendation defines objective tests that if passed will ensure (but will not guarantee) a minimum level of performance when installed in the network. An echo canceller which passes these tests should not harm equipment nor degrade transmission performance of voiceband signals and services below acceptable limits. These tests are lab-type tests and are not designed to be run in-service. Also, these tests are objective tests and do not replace or eliminate the need for subjective tests to measure the perceived quality of echo cancellers. Echo cancellers are complex devices with multiple parameters, and the correlation of these parameters and their interactions to the subjective quality of an echo canceller cannot be specified at this time. Thus, this Recommendation does not specify nor imply a selection criteria, however, guidelines are provided herein, and administrations have the freedom to specify criteria in their selection process. This set of criteria may include some or all of the thresholds and/or tests in this Recommendation.

1.2 References

The following ITU-T Recommendations, and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

– ITU-T Recommendation G.114 (1996), *One-way transmission time*.

– ITU-T Recommendation G.122 (1993), *Influence of national systems on stability and talker echo in international connections*.

– ITU-T Recommendation G.131 (1996), *Control of talker echo*.

– CCITT Recommendation G.164 (1988), *Echo suppressors*.

– ITU-T Recommendation G.165 (1993), *Echo cancellers.*

– ITU-T Recommendation G.167 (1993), *Acoustic echo controllers*.

– CCITT Recommendation G.223 (1984), *Assumptions for the calculation of noise on hypothetical reference circuits for telephony*.

– CCITT Recommendation G.229 (1984), *Unwanted modulation and phase jitter*.

– CCITT Recommendation G.711 (1988), *Pulse code modulation (PCM) of voice frequencies.*

– ITU-T Recommendation G.712 (1996), *Transmission performance characteristics of pulse code modulation channels*.

– CCITT Recommendation G.726 (1990), *40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).*

– CCITT Recommendation H.12 (1984), *Characteristics of telephone-type leased circuits.*

– CCITT Recommendation H.51 (1988), *Power levels for data transmission over telephone lines*.

– ITU-T Recommendation M.1050 (1993), *Lining up an international point-to-point leased circuit*.

– ITU-T Recommendation P.310 (1996), *Transmission characteristics for telephone-band (300-3400 Hz) digital telephones.*

– ITU-T Recommendation P.50 (1993), *Artificial voices*.

– ITU-T Recommendation P.56 (1993), *Objective measurement of active speech level.*

– ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.

– ITU-T Recommendation P.501 (1996), *Test signals for use in telephonometry*.

– ITU-T Recommendation Q.141 (1993), *Signal code for line signalling*.

– CCITT Recommendation Q.271 (1988), *General*.

– ITU-T Recommendation Q.552 (1996), *Transmission characteristics at 2-wire analogue interfaces of digital exchanges.*

– CCITT Recommendation Q.724 (1988), *Telephone user part signalling procedures*.

– ITU-T Recommendation T.30 (1996), *Procedures for document facsimile transmission in the general switched telephone network.*

– CCITT Recommendation V.2 (1988), *Power levels for data transmission over telephone lines*.

– ITU-T Recommendation V.8 (1994), *Procedures for starting sessions of data transmission over the general switched telephone network.*

– CCITT Recommendation V.21 (1984), *300 bits per second duplex modem standardized for use in the general switched telephone network.*

– CCITT Recommendation V.23 (1988), *600/1200-baud modem standardized for use in the general switched telephone network.*

– ITU-T Recommendation V.25 (1996), *Automatic answering equipment and general procedures for automatic calling equipment on the general switched telephone network including procedures for disabling of echo control devices for both manually and automatically established calls*.

– CCITT Recommendation V.26 *ter* (1988), *2400 bits per second duplex modem using the echo cancellation technique standardized for use on the general switched telephone network and on point-to-point 2-wire leased telephone-type circuits.*

– CCITT Recommendation V.27 *ter* (1984), *4800/2400 bits per second modem standardized for use in the general switched telephone network.*

– CCITT Recommendation V.29 (1988), *9600 bits per second modem standardized for use on point-to-point 4-wire leased telephone-type circuits.*

– ITU-T Recommendation V.32 (1993), *A family of 2-wire, duplex modems operating at data signalling rates of up to 9600 bit/s for use on the general switched telephone network and on leased telephone-type circuits.*

– ITU-T Recommendation V.34 (1996), *A modem operating at data signalling rates of up to 33 600 bit/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits.*

– IEC Publication 651 (1979), *Sound level metres.*

1.3 Terms and definitions

In the definition and text, *L* will refer to the relative power level of a signal, expressed in dBm0 (as defined by Recommendation G.711) and *A* will refer to the attenuation or loss of a signal path expressed in dB. These definitions assume that non-linearities are not present in the echo path and that the signal at S_{in} is purely echo. It is recognized that non-linearities may be present in a network.

1.3.1 acoustic echo

F: écho acoustique

S: eco acústico

Acoustic echoes consist of reflected signals caused by acoustic environments, e.g. analogue hands-free phones which are connected with a 2-wire circuit to a hybrid. An echo path is introduced by the acoustic path from earphone to microphone.

1.3.2 combined loss (A_{COM})

F: affaiblissement combiné (A_{COM})

S: atenuación combinada (A_{COM})

The sum of echo return loss, echo return loss enhancement and non-linear processing loss (if present). This loss relates L_{Rin} to L_{RET} by:

$$L_{RET} = L_{Rin} - A_{COM}$$

where:

$$A_{COM} = A_{ECHO} + A_{CANC} + A_{NLP}$$

1.3.3 comfort noise

F: bruit de confort

S: ruido confortativo

Insertion of pseudo-random noise during the silent interval when the NLP operates or allowance of some of the background or idle channel noise to pass through the NLP in order to prevent the annoyance of intervals of speech with background noise followed by intervals of silence.

1.3.4 composite echo

F: écho composite

S: eco compuesto

Composite echoes consist of the electric echoes and acoustic echoes caused by reflected signals at hybrids and acoustic environments, e.g. analogue hands-free telephones.

1.3.5 convergence

F: convergence

S: convergencia

The process of developing a model of the echo path which will be used in the echo estimator to produce the estimate of the circuit echo.

1.3.6 convergence time

F: temps de convergence

S: tiempo de convergencia

For a defined echo path, the interval between the instant a defined test signal is applied to the receive-in port of an echo canceller with the estimated echo path impulse response initially set to zero, and the instant the returned echo level at the send-out port reaches a defined level.

1.3.7 echo canceller

F: annuleur d'écho

S: compensador de eco; cancelador de eco

A voice-operated device placed in the 4-wire portion of a circuit and used for reducing near-end echo present on the send path by subtracting an estimation of that echo from the near-end echo (see Figure 2).

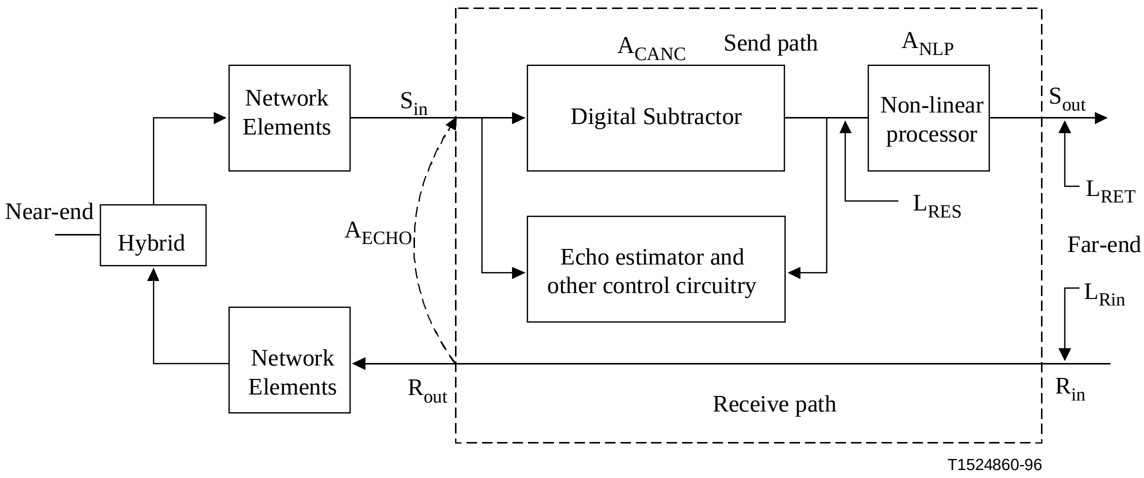


Figure 2/G.168 – Location of levels and loss of an echo canceller

1.3.8 echo path

F: trajet d'écho

S: trayecto del eco

The transmission path between R_{out} and S_{in} of an echo canceller. This term is intended to describe the signal path of the echo.

1.3.9 echo path capacity (Δ)

F: capacité en trajet d'écho (Δ)

S: capacidad del trayecto del eco (Δ)

The maximum echo path delay for which an echo canceller is designed to operate.

1.3.10 echo path delay (t_d)

F: retard de trajet d'écho (t_d)

S: retardo del trayecto del eco (t_d)

The delay from the R_{out} port to the S_{in} port due to the delays inherent in the echo path transmission facilities *including* dispersion time due to the network elements. In case of multiple echo paths, all delays and dispersions of any individual echo path are included. The dispersion time, which varies with different networks, is required to accommodate the band-limiting, and hybrid transit effects.

1.3.11 echo return loss (ERL) (A_{ECHO})

F: affaiblissement d'adaptation pour l'écho (ERL) (A_{ECHO})

S: atenuación del eco (ERL) (A_{ECHO})

The attenuation of a signal from the receive-out port (R_{out}) to the send-in port (S_{in}) of an echo canceller, due to transmission and hybrid loss, i.e. the loss in the (near-end) echo path.

NOTE – This definition does not strictly adhere to the echo loss definition given in 2.2/G.122, which applies to loss of the *a-t-b* path viewed from the virtual switching point of the international circuit. The echo canceller may be located closer to the echo reflection point.

1.3.12 echo return loss enhancement (ERLE) (A_{CANC})

F: renforcement de l'affaiblissement d'adaptation pour l'écho (ERLE)(A_{CANC})

S: atenuación reforzada del eco (ERLE) (A_{CANC})

The attenuation of the echo signal as it passes through the send path of an echo canceller. This definition specifically excludes any non-linear processing on the output of the canceller to provide for further attenuation.

1.3.13 electric echo

F: écho électrique

S: eco eléctrico

Electric echoes consist of reflected signals caused by the near-end impedance mismatch, e.g. at a 2-wire/4-wire conversion unit (hybrid).

1.3.14 far end

F: côté distant

S: extremo lejano

The side of an echo canceller which does not contain the echo path on which the echo canceller is intended to operate.

1.3.15 H register

F: registre H

S: registro H

The register within the echo canceller which stores the impulse response model of the echo path.

1.3.16 leak time

F: temps de fuite

S: tiempo de fuga

The interval between the instant a test signal is removed from the receive-in port of a fully-converged echo canceller and the instant the echo path model in the echo canceller changes such that, when a test signal is reapplied to R_{in} with the convergence circuitry inhibited, the returned echo is at a defined level.

This definition refers to echo cancellers employing, for example, leaky integrators in the convergence circuitry.

1.3.17 near-end

F: côté local

S: extremo cercano

The side of an echo canceller which contains the echo path on which the echo canceller is intended to operate. This includes all transmission facilities and equipment (including the hybrid and terminating telephone set) which is included in the echo path.

1.3.18 non-linear processor (NLP)

F: processeur non linéaire (NLP)

S: procesador no lineal (NLP)

A device having a defined suppression threshold level and in which:

a) signals having a level detected as being below the threshold are suppressed; and

b) signals having a level detected as being above the threshold are passed although the signal may be distorted (for example, see Annex B).

NOTE 1 – The precise operation of a NLP depends upon the detection and control algorithm used.

NOTE 2 – An example of a NLP is an analogue center clipper in which all signal levels below a defined threshold are forced to some minimum value.

1.3.19 non-linear processing loss (A_{NLP})

F: affaiblissement de traitement non linéaire (A_{NLP})

S: atenuación por procesamiento (o tratamiento) no lineal (A_{NLP})

Additional attenuation of residual echo level by a NLP placed in the send path of an echo canceller.

NOTE – Strictly, the attenuation of a non-linear process cannot be characterized by a loss in dB. However, for purposes of illustration and discussion of echo canceller operation, the careful use of A_{NLP} is helpful.

1.3.20 pure delay (t_r)

F: retard pur (t_r)

S: retardo puro (t_r)

The delay from the R_{out} port to the S_{in} port due to the delays inherent in the near-end echo path transmission facilities, not including dispersion time due to the network elements. In this case, the transit time directly across the hybrid is assumed to be zero (see Figure 3).

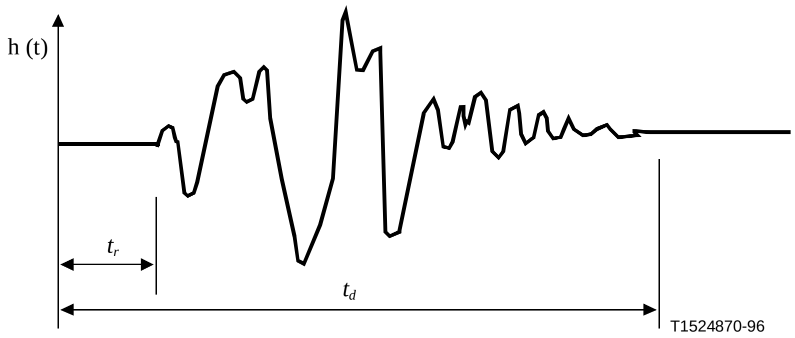


Figure 3/G.168 – Example of an impulse response of an echo path

1.3.21 residual echo level (L_{RES})

F: niveau d'écho résiduel (L_{RES})

S: nivel de eco residual (L_{RES})

The level of the echo signal which remains at the send-out port of an operating echo canceller after imperfect cancellation of the circuit echo. It is related to the receive-in signal L_{Rin} by:

$$L_{RES} = L_{Rin} - A_{ECHO} - A_{CANC}$$

Any non-linear processing is not included.

1.3.22 returned echo level (L_{RET})

F: niveau de retour d'écho (L_{RET})

S: nivel del eco devuelto (L_{RET})

The level of the signal at the send-out port of an operating echo canceller which will be returned to the talker. The attenuation of a NLP is included, if one is normally present. L_{RET} is related to L_{Rin} by:

$$L_{RET} = L_{Rin} - \left(A_{ECHO} + A_{CANC} + A_{NLP}\right)$$

If non-linear processing is not present, note that $L_{\mathrm{RES}} = L_{\mathrm{RET}}$.

1.4 Abbreviations

This Recommendation uses the following abbreviations.

ADPCM	Adaptive Differential Pulse Code Modulation
ATME	Automatic Test and Measurement Equipment
CED	Called Station Identification
CNG	Calling Tone
CPE	Customer Premises Equipment
CSI	Called Subscriber Identification
CSS	Composite Source Signal
DCME	Digital Circuit Multiplication Equipment
DCS	Digital Command Signal
DEC	Digital Echo Canceller

DIS	Digital Identification Signal
DTDT	Double Talk Detection Threshold
FAX	Facsimile
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
HDLC	High-level Data Link Control
IEC	International Electrotechnical Commission
NEST	Near-end Speech Threshold
NSF	Non-Standard Facilities
NSS	Non-standard Set-up
PCME	Packet Circuit Multiplication Equipment
PCM	Pulse Code Modulation
RMS	Root Mean Square
TBD	To Be Determined
TSI	Transmitting Subscriber Identification

2 Test signals

The tests in this Recommendation use special signals such as noise, tones, Group 3 facsimile signals, and a subset of the Composite Source Signals (CSSs) consisting of the bandlimited CSS with speech-like power density spectrum (pseudo noise signal generated using 8192 pt. FFT) and the bandlimited CSS for double talk (see Annex C/P.501). The CSS emulates the characteristics of speech, and its use as a test signal improves the ability to measure echo canceller performance for speech signals.

Furthermore, network echo cancellers should perform adequately on many non-speech signals, e.g. voiceband data, as well as under non-ideal network scenarios. Tests are included to test performance for Group 3 facsimile signals, residual acoustic echoes, and non-linearities in the echo path such as may arise with low bit rate encoding in the echo path.

3 Characteristics of echo cancellers

3.1 General

This Recommendation is applicable to the design of echo cancellers. The echo cancellers are assumed to be "half" echo cancellers, i.e. those in which cancellation takes place only in the send path due to signals present in the receive path.

3.2 Purpose, operation and environment

Echo cancellers have the following fundamental requirements:

1) rapid convergence;
2) low returned echo level during single talk;
3) low divergence during double talk;

4) assured double talk detection;

5) proper operation during facsimile and low speed (< 9.6 kbit/s) voiceband data transmissions.

Echo cancellers may remain active for several non-voice signals as well, in particular, Group 3 facsimile and low speed (< 9.6 kbit/s) voiceband data transmissions. Tests 10 and 14 address these issues.

It is increasingly common to have echo cancellers operate in tandem, especially in cellular applications. Tests for adequate performance are not defined. Test 11 is under study for this purpose.

When echo cancellers are located on the subscriber side of the international signalling equipment, signalling tones do not pass through the cancellers so no special action is necessary. When cancellers are on the international side of the signalling equipment, they are normally disabled by the switch during the active signalling exchange intervals in order to prevent distortion of the signalling tones by the echo canceller. When signalling tones simultaneously appear at the canceller receive and send ports (double talk), the receive signal will be processed through the echo path model contained in the canceller. The signal estimate produced by the canceller may sufficiently distort the send side signal so that it will not be properly recognized by the signalling receive unit (Note 1).

An echo canceller should be disabled during the transmission of the ITU-T No. 6 and No. 7 continuity check signal (Note 2). If an echo canceller conforming to this Recommendation is located on the international side of a circuit with ITU-T No. 6 or No. 7 signalling and is not externally disabled by the switch, it will not corrupt the return of the continuity check tone only if it is able to pass the optional Test No. 8 of this Recommendation. Similarly, if an echo canceller conforming to this Recommendation is located on the international side of ITU-T No. 5 signalling units and is not disabled by the switch, it will not corrupt the continuously compelled line signalling exchange only if it is able to pass the optional Test No. 8 of this Recommendation.

NOTE 1 – For some echo cancellers, this problem may not occur when the send and receive frequencies are different.

NOTE 2 – Recommendation Q.271 on ITU-T No. 6 and Recommendation Q.724 on ITU-T No. 7 both include the following statement: "As the presence of active echo suppressors in the circuit would interfere with the continuity check, it is necessary to disable the suppressors during the check and to re-enable them, if required, after the check has been completed." This consideration also applies to echo cancellers.

3.3 External enabling/disabling

Certain digital echo cancellers may be disabled directly by a digital signal. These echo cancellers should provide 64 kbit/s bit sequence integrity (i.e. if integrated, the A-law/μ-law conversion will also be disabled) in the externally disabled state. This is for further study.

3.4 Tests and requirements for performance with inputs signals applied to the send and receive paths

3.4.1 Transmission performance

The appropriate transmission performance requirements of Recommendations G.164 and G.165 also apply to echo cancellers except as noted below.

3.4.1.1 Group delay

The group delay in the send path should be kept to a minimum and should not exceed 1 ms. No significant delay should occur in the receive path.

NOTE – The creation of frame slips in the echo path can lead to an occasional degradation of the echo cancellation. If a delay is necessary to synchronize the digital send and receive paths, the global admissible delay on the send path, including the group delay mentioned above, should not exceed 1 ms and on the receive path 250 μs.

3.4.1.2 Measuring input and output levels

For testing purposes, the method defined for measuring the input level of the composite source signals is a RMS method. Unless otherwise specified within a test, the RMS method should also be used for measuring the output levels at S_{out}. Other methods that would give equivalent results are possible (see Annex C). For the RMS method, specifically, CSS is measured using:

$$S(k) = 3.14 + 20\log\left[\frac{\sqrt{\frac{2}{n}\sum_{i=k}^{k-n+1} e_i^2}}{4096}\right] \quad \text{(A-law encoding)}$$

$$S(k) = 3.17 + 20\log\left[\frac{\sqrt{\frac{2}{n}\sum_{i=k}^{k-n+1} e_i^2}}{4096}\right] \quad \text{(μ-law encoding)}$$

S(k) = signal level in dBm.

e_i = linear equivalent of the PCM encoded signal at time i.

k = discrete time index.

n = number of samples over which the RMS measurement is made, and $n > \alpha\tau$ with $\alpha \geq 1$ (an integer) and τ = period of CSS (5600 for the single-talk portion and 6400 for the double talk portion of CSS).

3.4.2 Echo canceller performance

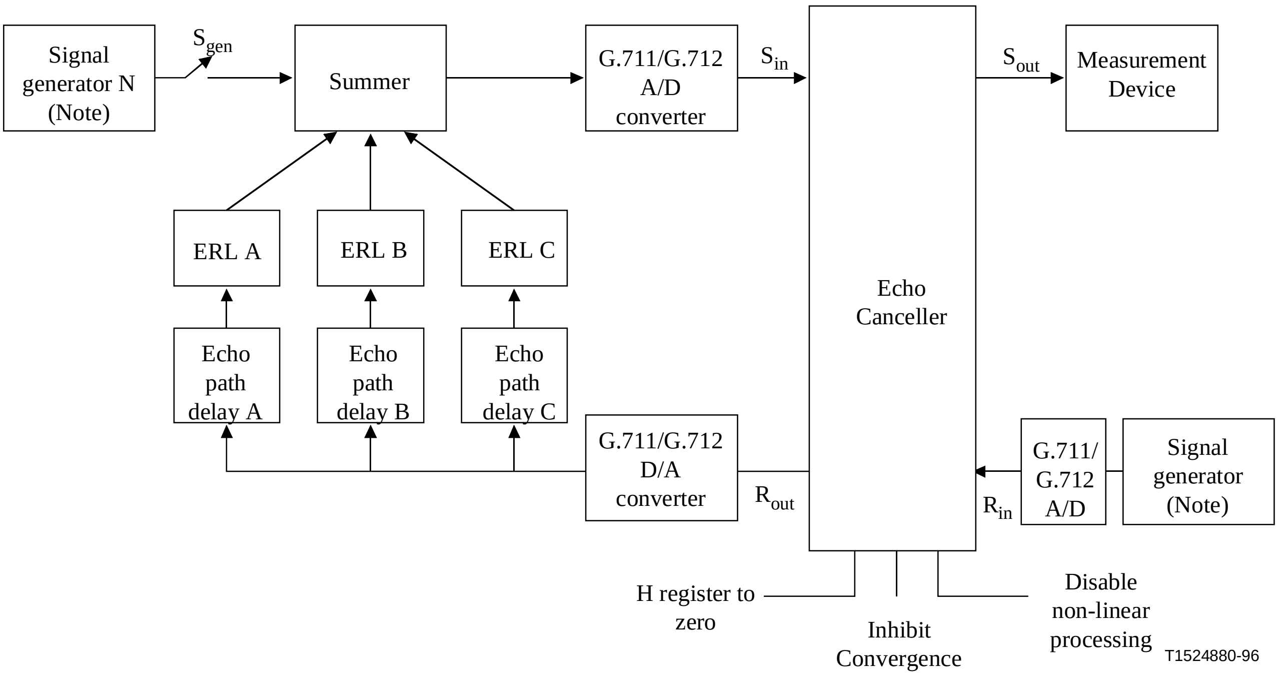
The performance requirements which follow are for echo cancellers which include NLPs.

For testing purposes, it is required that the NLP can be disabled, that the echo path impulse response store (H register) can be cleared (set to zero) and that adaptation can be inhibited.

The requirements are described in terms of tests made by applying signals to R_{in} and S_{in} of an echo canceller, and measuring the S_{out} signals. The test set-up is as shown in Figure 4. The ports are assumed to be at equal relative level points. For all values of R_{in}, and for all tests in this Recommendation, the level at R_{out} should be equal to the level at R_{in}. Any optional processing included in the echo canceller which may affect level transparency between R_{in} and R_{out} should be disabled during all tests in this Recommendation. The composite source signals, which consist of the receive-input test signal and send-input test signal (see Annex C/P.501) are used as the test signals, unless otherwise indicated. The ERL is independent of frequency. For multiple channel implementations, channel-to-channel independence is required, and any channels tested simultaneously should each meet the requirements of this Recommendation.

The ERL used in these tests have a minimum value of 6 dB. It should be noted that 6 dB is a typical worst case value encountered for most networks, and most current networks have typical ERL values better than this. Also, it should be noted that the test configurations specified in this Recommendation are artificial for purposes of test and result repeatability, and do not fully represent conditions that would be expected in real networks.

NOTE – The requirements in this subclause are based on the use of the composite source signals, noise, tones, FAX signals, and voiceband data signals as the test signals. A more complex echo path circuit including dispersions typically generated by hybrids, cable gauge changes and other sources of echo which occur in the real network is under study. This will include real or simulated hybrids, especially those providing wide dispersion and a relatively low ERL, which may provide a better characterization of the echo canceller under evaluation.



NOTE – The sum of the absolute values of the gains G_A, G_B, G_C that correspond to ERL A, B, C, respectively, taken in dB, should be less than or equal to –6 dB (i.e. $20 \log(|G_A| + |G_B| + |G_C|) \leq -6$ dB), and echo path delay A $\leq \Delta$ ms, echo path delay B $\leq \Delta$ ms, and echo path delay C $\leq \Delta$ ms.

Figure 4/G.168 – Functional diagram for echo canceller performance measurements

The primary purpose of an echo canceller is to control the echo of a speech signal. This is done by synthesizing a replica of the echo path impulse response and using it to generate an estimate of the echo which is subtracted from the actual circuit echo. The synthesis should be accomplished using a speech input signal. Because of the difficulty of defining a speech test signal, the following tests are type tests and rely upon the use of a composite source signal primarily for convenience and repeatability. These tests should be performed on an echo canceller only after the design has been shown to properly synthesize a replica of the echo path impulse response from a speech input signal and its corresponding echo. Speech signals are not used in the tests in this subclause. Additionally, the NLP in the echo canceller should be designed to minimize and potentially avoid undesirable effects such as double talk clipping, gaps in transmitted speech signals, and noise contrast (see Test 9 described later in this Recommendation for noise contrast, and see Appendix I for further discussion on double talk clipping). Tests to ensure proper operation are under study.

Different echo cancellers may be designed to work satisfactorily for different echo path delays depending on their application in various networks. Thus Δ, whenever it appears in this Recommendation, represents the maximum echo path delay for which the echo canceller is designed.

3.4.2.1 Test No. 1 – Steady state residual and returned echo level test

This test is meant to ensure that the steady state cancellation (A_{CANC}) is sufficient to produce a residual echo level which is sufficiently low to permit the use of non-linear processing without undue reliance on it. In general, given that all other variables are equal, a higher value of ERLE or lower values of L_{RES} will allow for less dependence on the NLP functionality.

The H register is initially cleared and a receive signal is applied for a sufficient time for the canceller to converge producing a steady state residual echo level (see Figure 5).

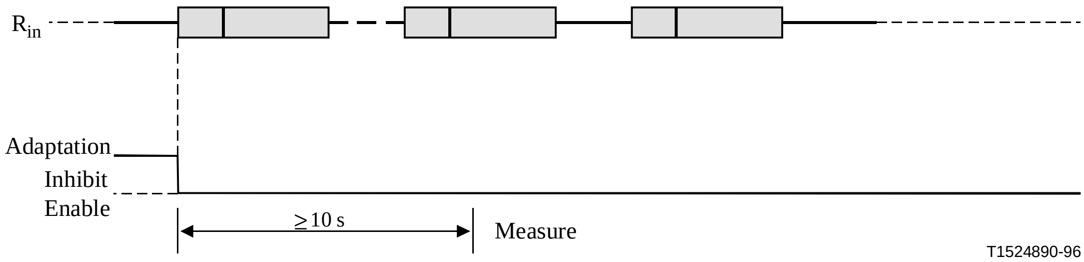


Figure 5/G.168 – Test No. 1 signal and time relationships

Requirement

With the H register initially set to zero, the NLP disabled for all values of receive input signal level such that $L_{Rin} \geq -30$ dBm0 and ≤ 0 dBm0 and for all values of ERL ≥ 6 dB and echo path delay, $t_d \leq \Delta$ ms, the residual echo level should be less than or equal to that shown in Figure 6. When the NLP is enabled, the returned echo level should be less than or equal to that shown in Figure 7. In addition, with the NLP either enabled or disabled, no peaks are allowed that exceed x dB (x under study) above the requirements in Figures 7 and 6 respectively. Peaks are either per-sample or RMS averaged (method under study).

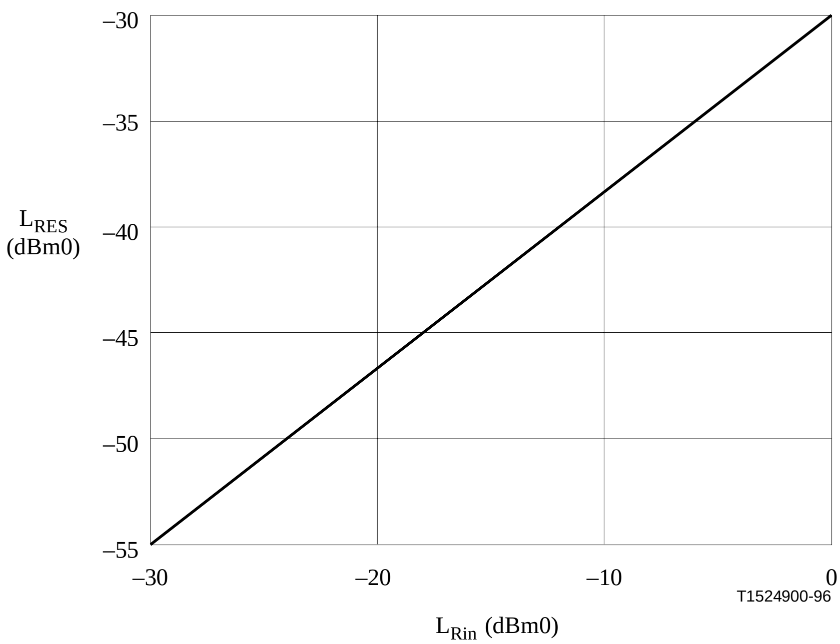


Figure 6/G.168 – Relationship between receive input level (L_{Rin}) and residual echo level (L_{RES}) with NLP disabled

The requirements in Figure 6 may not be met with echo cancellers containing a comfort noise feature, if enabled, and so, for the purposes of this test, comfort noise is disabled. For R_{in} signal levels exceeding –5 dBm0, CSS will be clipped. For this range, special care should be taken to ensure that the echo path is linear. Non-linearities in the real network may result in performance less than indicated in Figure 7.

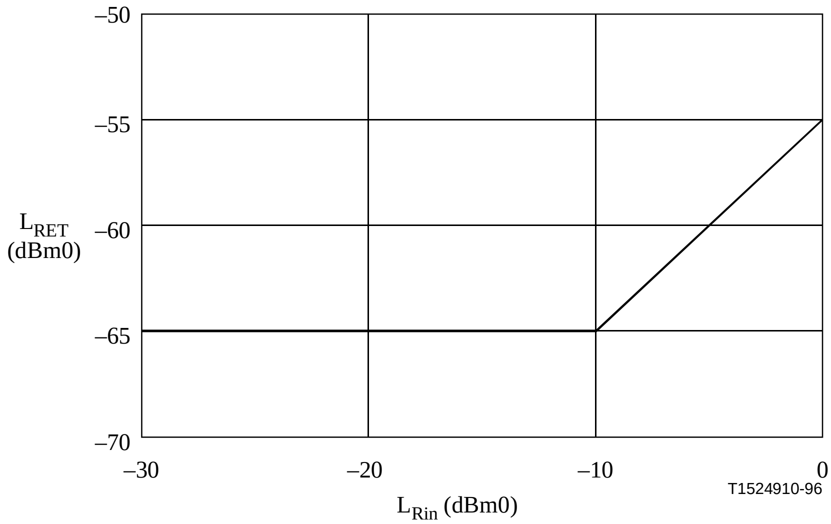


Figure 7/G.168 – Relationship between receive input level (L_{Rin}) and return echo level (L_{RET}) with NLP enabled

3.4.2.2 Test No. 2 – Convergence test

This test is meant to ensure that the echo canceller converges rapidly for all combinations of input signal levels and echo paths and that the returned echo level is sufficiently low. The H register is initially cleared and adaptation is inhibited. Adaptation is then enabled 50 ms before the start of a CSS burst (Figure 8). This 50 ms period is to allow for the latency time in the adaptation control of the canceller. The degree of adaptation, will depend on the convergence characteristics of the echo canceller.

At the beginning of a call the convergence should be fast enough to be subjectively unnoticeable. In general, the convergence should be fast enough to handle changes in the echo path in a subjectively transparent fashion. Faster convergence than required in Figures 8a and 8b is desirable, but only if no degradation is observed during single or double talk and the stability of the canceller can be maintained in all network conditions (e.g. various echo path conditions, including various hybrids) and for all voiceband signals.

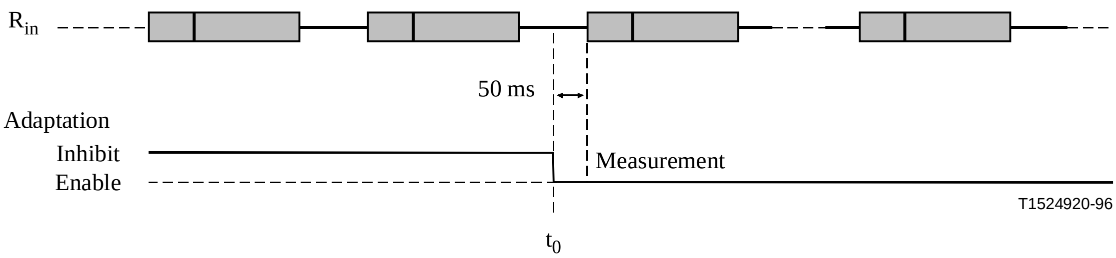


Figure 8/G.168 – Test No. 2A, 2B signal and time relationships

3.4.2.2.1 Test No. 2A – Convergence test with NLP enabled

Requirement

With the H register initially set to zero and the NLP enabled, for all values $L_{Rin} \geq -30$ dBm0 and ≤ 0 dBm0 and for all values of ERL ≥ 6 dB and echo path delay, $t_d \leq \Delta$ ms, the combined loss ($A_{COM} = A_{ECHO} + A_{CANC} + A_{NLP}$) should be greater than or equal to that shown in Figure 8a. The value of X_{CONV} should be greater than or equal to 16 dB, but the exact value is for further study. The level at S_{out} is measured using a level meter conforming to IEC 651 with impulse time constant (35 ms) modified to remove the peak detector and decay time constant block (see IEC 651, paragraph 7.3, Figure 2). The level at R_{in} is measured using the RMS method of 3.4.1.2, but modified to include only those samples of the CSS that are in the active portion of the CSS (i.e. excluding the gaps in the CSS signal). The modified IEC 651 method may also be used at R_{in}, but the input and output signals must also be synchronized.

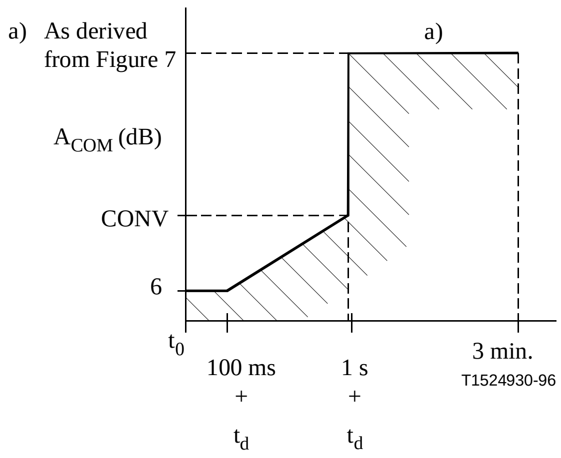


Figure 8a/G.168 – Convergence characteristics with NLP enabled

3.4.2.2.2 Test No. 2B – Convergence test with NLP disabled

Requirement

With the H register initially set to zero and the NLP disabled, for all values $L_{Rin} \geq -30$ dBm0 and ≤ 0 dBm0 and for all values of ERL ≥ 6 dB and echo path delay, $t_d \leq \Delta$ ms, the loss $A_{ECHO} + A_{CANC}$ should be greater than or equal to that shown in Figure 8b. The value of X_{CONV} should be greater than or equal to 16 dB, but the exact value is for further study. The level at S_{out} is measured using a level meter conforming to IEC 651 with impulse time constant (35 ms) modified to remove the peak detector and decay time constant block (see IEC 651, paragraph 7.3, Figure 2). The level at R_{in} is measured using the RMS method of 3.4.1.2, but modified to include only those samples of the CSS that are in the active portion of the CSS (i.e. excluding the gaps in the CSS signal). The modified IEC 651 method may also be used at R_{in}, but the input and output signals must also be synchronized.

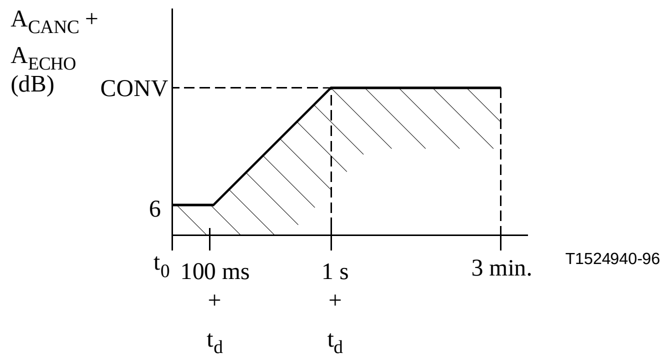


Figure 8b/G.168 – Convergence characteristics with NLP disabled

3.4.2.2.3 Test No. 2C – Convergence test in the presence of background noise

Test No. 2C is meant to ensure that the echo canceller converges rapidly for all combinations of input signal levels and echo paths in the presence of background noise.

The test procedure is to clear the H register and inhibit adaptation. A Hoth noise source (see Recommendation P.800) with level N is applied at S_{gen}. Adaptation is enabled coincident with the start of a CSS burst (see Figure 9). After the convergence time, inhibit adaptation and measure the residual echo level.

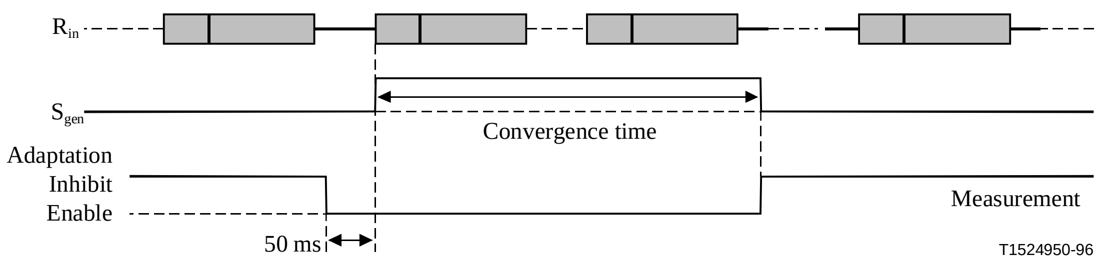


Figure 9/G.168 – Test No. 2C signal and time relationships

Requirement

With the H register initially set to zero and the NLP enabled, for all values of $L_{Rin} \geq -25$ dBm0 and ≤ 0 dBm0, $N = L_{Rin} -15$ dB but no higher than –30 dBm0, ERL ≥ 6 dB and echo path delay, $t_d \leq \Delta$ ms, convergence should occur within 1.0 s and L_{RET} should be $\leq N$.

With the H register initially set to zero and the NLP disabled, for all values of $L_{Rin} \geq -25$ dBm0 and ≤ 0 dBm0, $N = L_{Rin} -15$ dB but no higher than –30 dBm0, ERL ≥ 6 dB and echo path delay, $t_d \leq \Delta$ ms, convergence should occur within X_{h1} (under study) seconds and L_{RES} should be $\leq X_{h2}$ (under study).

3.4.2.3 Test No. 3 – Performance under conditions of double talk

The three parts of this test are meant to test the performance of the canceller under various conditions of double talk. During conditions of double talk the echo canceller can give rise to unwanted artefacts such as clipping, distortion, and noise contrast (see Appendix I). The tests make the assumption that, upon detection of double talk, measures are taken to prevent or slow adaptation in order to avoid excessive reduction in cancellation.

While CSS is proposed for this test it is recognized that it is only a statistical approximation of real speech. Double talk tests performed with actual speech samples may produce results somewhat different than those shown in this test. This test is intended to provide a guideline on how the double talk performance of an echo canceller should be measured. It is possible that this test and its requirements may change as the correlation between CSS and real speech is better understood. Use of different languages have been shown to provide considerable variation in the results for tests 3A and 3B [Reference: COM 15-27 (1993)].

Some concern has been expressed over the lack of specific numbers in this test. Those that have this concern should use Test 3 of Recommendation G.165 instead.

3.4.2.3.1 Test No. 3A – Double talk test with low near-end levels

Test No. 3A is meant to ensure that the double talk detection is not so sensitive that echo and low level near-end speech falsely cause operation of the double talk detector to the extent that adaptation does not occur. The test procedure is to clear the H register; then for some value of echo path delay and ERL, a signal is applied to R_{in}. Simultaneously (see Figure 10) an interfering signal, which is sufficiently low in level to not seriously hamper the ability of the echo canceller to converge, is applied at S_{gen}. This signal should allow adaptation and cancellation to occur. After the allowed convergence time the adaptation is inhibited and the residual echo measured. The NLP should be *disabled*.

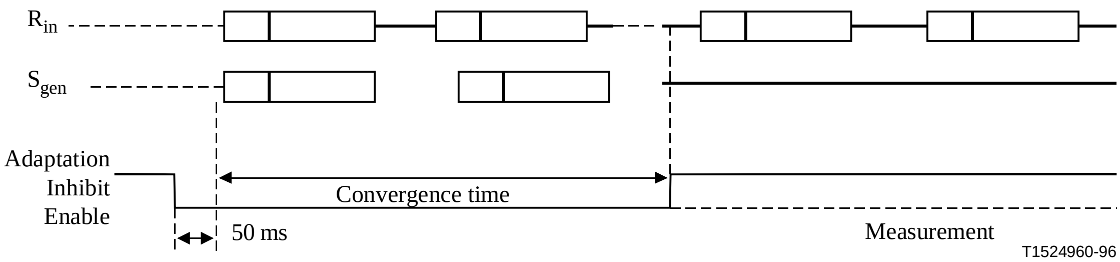


Figure 10/G.168 – Test No. 3A signal and time relationships

Requirement

With the H register initially set to zero for all values of $L_{Rin} \geq -25$ dBm0 and ≤ 0 dBm0, $N = L_{Rin} - 15$ dB, ERL ≥ 6 dB and echo path delay, $t_d \leq \Delta$ ms, convergence should occur within T_c seconds and L_{RES} should be $\leq N$. It is reasonable to expect that T_c should be less than 5 seconds, but the exact value is for further study.

3.4.2.3.2 Test No. 3B – Double talk test with high near-end levels

Test No. 3B is meant to ensure that the double talk detector is sufficiently sensitive and operates fast enough to prevent large divergence during double talking.

The test procedure is to fully converge the echo canceller for a given echo path by applying CSS to R_{in}. After the canceller is fully converged (see Figure 11), a signal N is applied to S_{gen} which has a level at least that of R_{in}. This should cause the double talk detector to operate. After any arbitrary time, $\delta_t > 0$, the adaptation is inhibited, the S_{gen} signal is removed, and the residual echo measured. The NLP should be *disabled.*

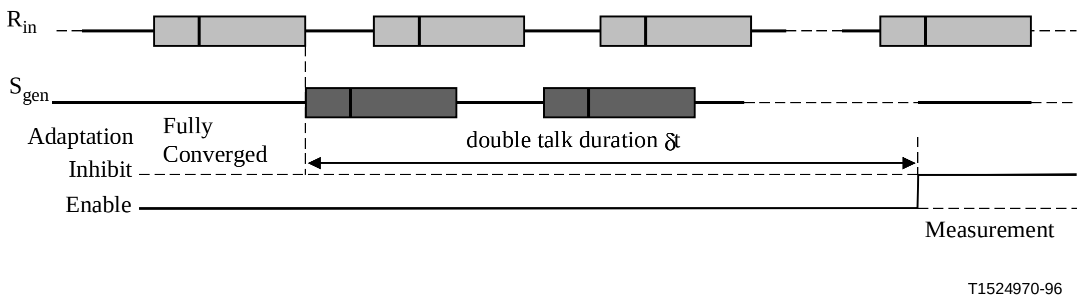


Figure 11/G.168 – Test No. 3B signal and time relationships

Note that Test No. 3B is even more sensitive to real speech variations and CSS may not provide adequate approximation of real speech for this test.

Requirement

With the echo canceller initially in the fully converged state for all values of $L_{Rin} \geq -30$ dBm0 and ≤ 0 dBm0, and for all values of $N \geq L_{Rin}$ and for all values of ERL ≥ 6 dB and echo path delay $t_d \leq \Delta$ ms, the residual echo level after the simultaneous application of L_{Rin} and N for any time period should not increase more than D dB over the steady state requirements of Test No. 1 (Figure 6). The value of D is for further study. In general, the lower the value of D, the better the performance.

3.4.2.3.3 Test 3C – Double talk test under simulated conversation (this test is under study)

Test No. 3C is meant to ensure that the echo canceller does not produce undesirable artefacts during and after periods of double talk.

The test procedure is to clear the H register. Then for some value of echo path delay and ERL a signal is applied to R_{in}. Simultaneously (see Figure 12), a signal N is applied to S_{gen} which has a level at least that of R_{in}. After a time t_1, S_{gen} is removed. After a time t_2, S_{out} is measured for a time t_3. After a time t_4, N is reapplied. The NLP should be enabled.

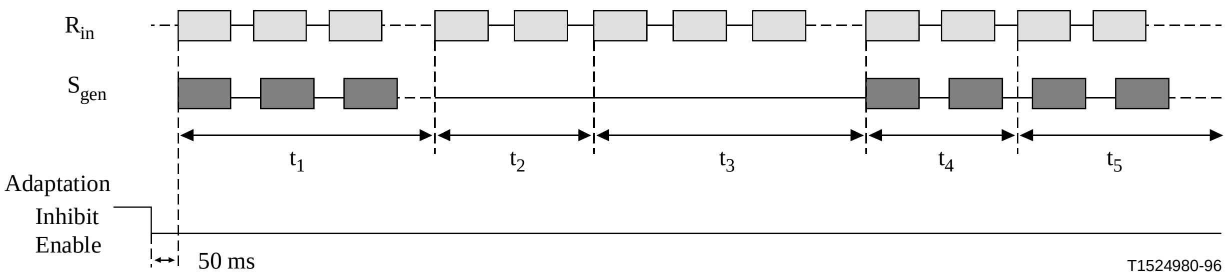


Figure 12/G.168 – Test No. 3C signal and time relationships

Requirement

With the H register initially set to zero, for all values of $L_{Rin} \geq -25$ dBm0 and ≤ 0 dBm0, and for all values of N ≥ or ≤ TBD and for all values of ERL ≥ 6 dB and echo path delay $t_d \leq \Delta$ ms, the residual echo level during time period t_3 should meet the requirements of Test No. 1 with NLP enabled (Figure 7).

(During period t_2, the goal of this test is to ensure that peaks are limited to a defined level. Specific methods for measuring these are for further study.)

(The intent of time period t_4, t_5 is to ensure that the echo canceller does not produce artefacts when double talk has resumed after a period of single talk. The requirements during this period are for further study.)

Level offsets between L_{Rin} and L_{Sgen} can cause inappropriate operation of the NLP and can cause speech degradation and is for further study. Variation of CSS may be useful for this purpose.

3.4.2.4 Test No. 4 – Leak rate test

This test is meant to ensure that the leak time is not too fast, i.e. that the contents of the H register do not go to zero too rapidly.

The test procedure is to fully converge the echo canceller using CSS for a given echo path and then to remove all signals from the echo canceller. After two minutes the contents of the H register are frozen, CSS is reapplied to R_{in} and the residual echo measured (see Figure 13). The NLP should be *disabled*.

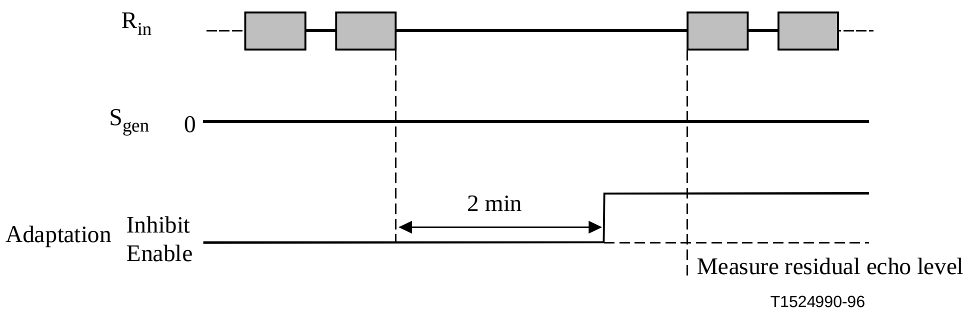


Figure 13/G.168 – Test No. 4 signal and time relationships

Requirement

With the echo canceller initially in the fully converged state for all values of $L_{Rin} \geq -30$ dBm0 and ≤ 0 dBm0, two minutes after the removal of the R_{in} signal, the residual echo level should not increase more than 10 dB over the steady state requirement of Test No. 1 (Figure 6).

3.4.2.5 Test No. 5 – Infinite return loss convergence test (this test is under study)

This test is meant to ensure that the echo canceller has some means to prevent the unwanted generation of echo. This may occur when the H register contains an echo path model, either from a previous connection or the current connection, and the echo path is opened (circuit echo vanishes) while a signal is present at R_{in}.

The test procedure is to fully converge the echo canceller using CSS for a given echo path. The echo path is then interrupted while a CSS is applied to R_{in}. 500 ms after interrupting the echo path the residual echo signal at S_{out} should be measured (see Figure 14). The NLP should be *disabled*.

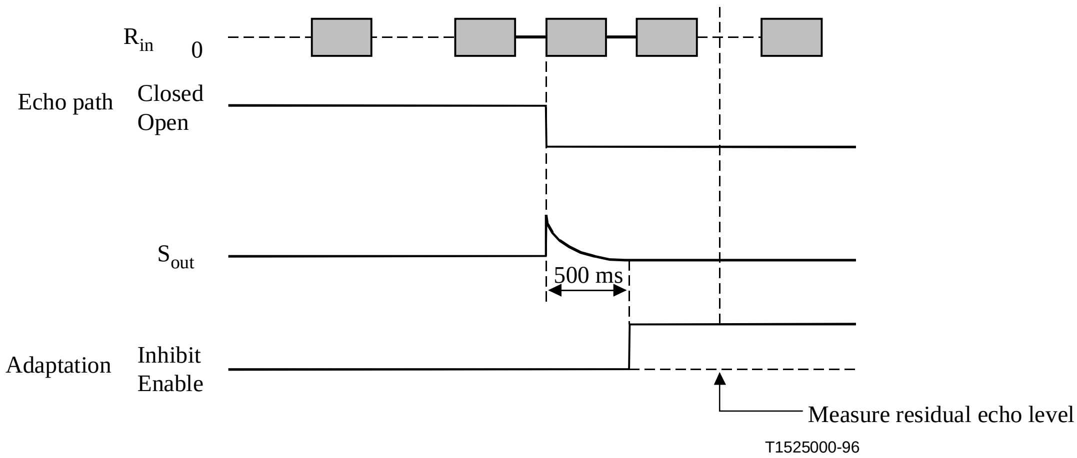


Figure 14/G.168 – Test No. 5 signal and time relationships

Requirement

With the echo canceller initially in the fully converged state for all values of ERL ≥ 6 dB, and for all values of L_{Rin} ≥ –30 dBm0 and ≤ 0 dBm0, the residual echo level at S_{out}, 500 ms after the echo path is interrupted, should be ≤ TBD dBm0.

3.4.2.6 Test No. 6 – Non-divergence on narrow-band signals (this test is under study)

This test has the object of verifying that the echo canceller will remain converged for subscriber-originated narrow-band signals after having converged on a wideband signal. The residual echo level is measured before and after the application of a sinusoidal wave or a wave composed of two frequencies.

The method consists of completely converging the echo canceller as in Test No. 1. One or more mono or bi-frequency signals from Table (TBD) are then applied in any sequence at R_{in}. After TBD minutes, the adaptation is inhibited and the residual echo is measured with the signal of Test No. 1. The NLP should be disabled.

Requirement

With the echo canceller first fully converged as in Test No. 1 and then after application at R_{in} of any sequence of mono or bi-frequency signals from Table (TBD) such that L_{Rin} ≥ TBD dBm0 and ≤ TBD dBm0 and for all values of ERL ≥ 6 dB and echo path delay $t_d \leq \Delta$ ms, and using the same test signal as in Test No. 1, and with adaptation inhibited, the residual echo levels should be less than or equal to TBD.

3.4.2.7 Test No. 7 – Stability test

The object of this test is to verify that the echo canceller will remain stable for narrow-band signals. The residual echo is measured before and after the application of a sinusoidal wave.

The test method is as follows: with the H register initially set to zero, and the NLP disabled, the echo canceller is converged on the sinusoidal signal. After two minutes, the residual echo is measured using the applied signal.

Requirement

With the echo canceller H register initially set to zero, and after the application at R_{in} of a mono-frequency signal, except for those identified in Table 1 of 3.4.2.8 for two minutes, for all values of L_{Rin} ≥ –30 dBm0 and ≤ +3 dBm0, and for all values of ERL ≥ 6 dB, with an echo path delay $t_d \leq \Delta$ ms, the residual echo levels should be less than or equal to that shown in Figure 6 during the application of the signal.

3.4.2.8 Test No. 8 – Non-convergence of echo cancellers on specific ITU-T No. 5, 6, and 7 in-band signalling and continuity check tones (optional)

Echo cancellers which are not externally disabled by the switch and which are located on the line side of Signalling System No. 5, 6 and 7 in international exchanges or are associated with national exchanges, should operate properly with specific in-band signalling and continuity check tones. This test is meant to ensure that echo cancellers will not remove or cancel a mono or bi-frequency signal transmitted in a handshaking protocol in the transmit direction either before or after receiving an identical signal (except for amplitude and phase) in the receive direction. This is intended to allow a correct transmission of specific signalling or continuity check tones without externally disabling the echo canceller. The NLP should be enabled.

Requirement

If the echo canceller is equipped with this optional capability, and with the echo canceller in any initially converged condition (for simplification, the fully converged state for an ERL of 6 dB may be chosen), the level at S_{out} should not vary more than 2 dB compared to any mono- or bi-frequency signal of Table 1 (according to the tolerance requirements from the appropriate signalling system) applied at S_{in} when the same signal (except for amplitude and phase) is applied to R_{in} within 90 ms (either before or after) of the signal applications at S_{in}. This requirement applies for all values of ERL $\geq$ 6 dB, with an echo path delay $t_d \leq \Delta$ ms. The signal level, N, of each frequency applied is such that the peak level of N is equivalent to the peak level M of a sinusoid with a level of –18 dBm0 $\leq$ M $\leq$ +3 dBm0. (Response time and Table 1 are for further study.)

Table 1/G.168 – Applicable signalling tones

System 5	System 6	System 7
2400 Hz	2010 Hz	2010 Hz
2600 Hz		
2400 Hz +2600 Hz		

3.4.2.9 Test No. 9 – Comfort noise test (this test is under study)

This test is meant to ensure that the echo canceller is able to provide a comfort noise signal on S_{out} which matches noise received on S_{in}. It also tests whether the canceller is able to adjust the level of this comfort noise signal to compensate for changes in the level of input noise. As this test is not intended as a test of echo cancellation capability, an ERL of 8 dB is used for the entire test. The steps of this test should be applied in sequence. This test covers a range of operation between –60 dBm0 and –40 dBm0. White noise is used for this test. The NLP and comfort noise feature should be enabled.

3.4.2.9.1 Part 1 (matching)

1) Set N to a level between –50 dBm0 and –40 dBm0.

2) Set L_{Rin} to silence (< –40 dBm0) and hold for 30 seconds.

3) Set L_{Rin} to –10 dBm0.

4) Measure L_{RET} after 2 seconds.

Requirement

For all values of N, L_{RET} should be within 2.0 dB of N. Also, this value should hold as long as noise level N remains constant.

3.4.2.9.2 Part 2 (adjustment down)

1) Lower N by 10 dB.
2) Set L_{Rin} to silence (< –40 dBm0) and hold for z1 seconds (z1 under study).
3) Set L_{Rin} to –10 dBm0.
4) Measure L_{RET} after 2 seconds.

Requirement

L_{RET} should be within 2.0 dB of N. Also, this value should hold as long noise level N remains constant.

3.4.2.9.3 Part 3 (adjustment up)

1) Raise N by 10 dB.
2) Set L_{Rin} to silence (< –40 dBm0) and hold for z2 seconds (z2 under study).
3) Set L_{Rin} to –10 dBm0.
4) Measure L_{RET} after 2 seconds.

Requirement

L_{RET} should be within 2.0 dB of N. Also, this value should hold as long noise level N remains constant.

3.4.2.10 Test No. 10 – Facsimile test during call establishment phase

These tests are meant to ensure that the echo canceller converges rapidly on initial FAX handshaking sequences and that the echo canceller has some means to prevent the unwanted generation of echo by these signals. The test should be performed with the G.165/G.168 tone disabler switched on.

For this purpose, the following signals should be applied (bits are transmitted left to right). The initial flag is repeated 37 times for each sequence.

FAX test sequences

Calling tone (CNG)

Conditions:

Signal 1100 Hz ± 38 Hz

Duration On for 0.5 s, Off for 3 seconds (±15%)

Called station identification (CED)

Conditions:

Signal 2100 Hz ± 15 Hz

Duration 2.6 s-4 s

Binary coded sequences

Sequence No. 1 (called station)

Non-Standard Facilities (NSF) frame

Flag	HDLC address field	HDLC Control field	Control field NSF	Information field, 8 octets (country, manufacturer, additional code number)	Frame check sequence	Flag
0111 1110	1111 1111	1100 1000	0000 0100	0101 0101, 0101 0101, 0101 0101, ...	1010 1010	0111 1110

Called Subscriber Identification (CSI) frame

Flag	HDLC address field	HDLC Control field	Control field CSI	Information field, 20 octets (receiver code number)	Frame check sequence	Flag
0111 1110	1111 1111	1100 1000	0000 0100	0101 0101, 0101 0101, 0101 0101, ...	1010 1010	0111 1110

Digital Identification Signal (DIS) frame

Flag	HDLC address field	HDLC Control field	Control field DIS	Information field, 3 octets	Frame check sequence	Flag
0111 1110	1111 1111	1100 1000	0000 0001	0101 0101, 0101 0101, 0101 0101	1010 1010	0111 1110

Sequence No. 2 (calling station)

Non-Standard Set-up (NSS) frame

Flag	HDLC address field	HDLC Control field	Control field NSS	Information field, 3 octets	Frame check sequence	Flag
0111 1110	1111 1111	1100 1000	1100 0100	0101 0101, 0101 0101, 0101 0101	1010 1010	0111 1110

Transmitting Subscriber Identification (TSI) frame

Flag	HDLC address field	HDLC Control field	Control field TSI	Information field, 20 octets (Transmitter code number)	Frame check sequence	Flag
0111 1110	1111 1111	1100 1000	1100 0010	0101 0101, 0101 0101, 0101 0101, ...	1010 1010	0111 1110

Digital Command Signal (DCS) frame

Flag	HDLC address field	HDLC Control field	Control field DCS	Information field, 20 octets (Transmitter code number)	Frame check sequence	Flag
0111 1110	1111 1111	1100 1000	1100 0001	0101 0101, 0101 0101, 0101 0101, ...	1010 1010	0111 1110

Data transmission conditions

The transmission of sequences No. 1 and No. 2 in the telephone channel is obtained by means of frequency shift (see Recommendation V.21).

Conditions:

Data signalling rate, synchronous	300 bit/s
Center frequency	1750 Hz

Frequency deviation	±100 Hz
Characteristic frequencies	1650/1850 Hz
Tolerances of the characteristic frequencies	±6 Hz

The higher characteristic frequency corresponds to a binary "0".

Tests

3.4.2.10.1 Test No. 10A – Canceller operation on the calling station side

The convergence test procedure is to clear the H register and to inhibit adaptation. Then adaptation is enabled for at least 7 s, while CED and sequence No. 1 are applied (see Figure 15). During the adaptation time, the residual/returned echo level is measured. This test should be performed with the NLP both enabled and disabled.

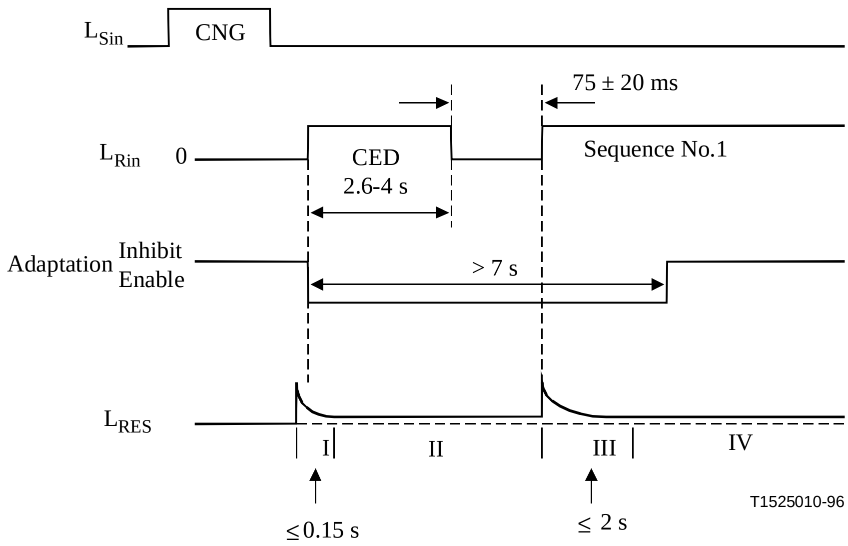


Figure 15/G.168 – Test No. 10A signal and time relationships

Requirement

With the H register initially set to zero and the value L_{Rin} = –13 dBm0 for the entire test, the following specific requirements apply for all values of ERL ≥ 6 dB and echo path delay $t_d \leq \Delta$ ms. In addition, there is an overall requirement that no unwanted echo bursts be produced. The test should run for 7 s as a minimum. Repeat sequence 1 as necessary.

Region I (converging on CED tone):

– the peaks of L_{RES} should be ≤ (–13 – A_{ECHO}) dBm0;

– the time to enter region II should be ≤ 0.15 s.

Region II (converged on CED tone):

– the peaks of L_{RES} should be ≤ –37 dBm0.

Region III (converging on sequence No. 1):

– the peaks of L_{RES} should be ≤ (–13 – A_{ECHO}) dBm0;

– the time to enter region IV should be ≤ 2 s.

Region IV (converged on sequence No. 1):

– the peaks of L_{RES} should be ≤ –24 dBm0.

If the NLP is enabled, L_{RET} should be ≤ X_f (under study) dBm0 in the regions II and IV. The value of X_f is for further study and should be no greater than the respective value with the NLP disabled. Note that according to Recommendation V.21, the sensitivity of the FAX receiver has a minimum value of –48 dBm.

3.4.2.10.2 Test No. 10B – Canceller operation on the called station side

The convergence test procedure is to clear the H register and to inhibit adaptation. Then adaptation is enabled for at least 10 s, while sequence No. 2 is applied (see Figure 16). During the adaptation time, the residual/returned echo level is measured. This test should be performed with the NLP both enabled and disabled.

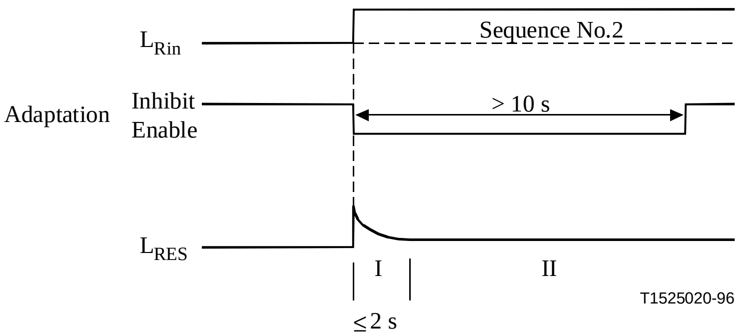


Figure 16/G.168 – Test No. 10B signal and time relationships

Requirement

With the H register initially set to zero and the value L_{Rin} = –13 dBm0 for the entire test, the following specific requirements apply for all values of ERL ≥ 6 dB and echo path delay t_d ≤ Δ ms. In addition, there is an overall requirement that no unwanted echo bursts be produced. The test should run for 10 s as a minimum. Repeat sequence 2 as necessary.

Region I (converging on sequence No. 2):

– the peaks of L_{RES} should be ≤ (–13 – A_{ECHO}) dBm0;

– the time to enter region II should be ≤ 2 s.

Region II (converged on sequence No. 2):

– the peaks of L_{RES} should be ≤ –24 dBm0.

If the NLP is enabled, L_{RET} should be ≤ X_f (under study) dBm0 in the regions II and IV. The value of X_f is for further study and should be no greater than the respective value with the NLP disabled. Note that according to Recommendation V.21, the sensitivity of the FAX receiver has a minimum value of –48 dBm.

3.4.2.11 Test No. 11 – Tandem echo canceller test

Under study. See Appendix I for further discussion on this issue.

3.4.2.12 Test No. 12 – Residual acoustic echo test

Under study. See Appendix I for further discussion on this issue.

3.4.2.13 Test No. 13 – Performance with ITU-T low bit rate coders in echo path

Under study. (The intention is to put in a table of performance goals for each coder/algorithm.)

3.4.2.14 Test No. 14 – Performance with V-Series low-speed data modems

This test is meant to ensure that echo cancellers will not impair the performance of V-Series low-speed (< 9.6 kbit/s) modems, including V.22 *bis* modems, which do not send a 2100 Hz disable tone with phase-reversals. The bit-error rate is measured while the echo cancellers operate in a simulated network with low-speed data modems.

The echo canceller is placed in the test configuration of Figure 17. The H register is cleared and NLP enabled and the modems allowed to train. The modems are then operated for a minimum of three minutes. The test should be repeated with the echo canceller both disabled and enabled, and the bit-error rate monitored.

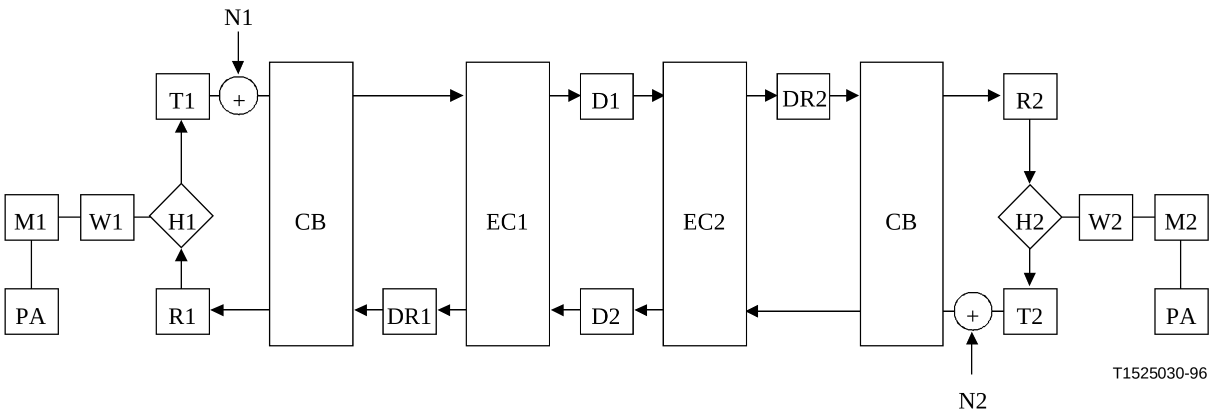
A specific selection of modem(s) to be tested should be done by the Administrations, depending on the most critical and prevalent types in the network. In the test set-up, the artificial 2-wire lines and the hybrids should simulate the actual range of echo paths that the echo canceller under test is intended to cope with.

For the hybrid this means a specification of the equivalent balance network.

NOTE – Examples of typical balance networks are given in Figure 11/Q.552.

For the artificial line this means a specification of the fundamental cable parameters, e.g. ohms/km and nF/km for unloaded cables. The length of the artificial lines should be variable. Test cases should include minimum and maximum lengths as well as that length for which the highest weighted echo loss, calculated according to Recommendation G.122, is obtained.

The hybrid and artificial line arrangements should be equal at each side of the test set-up.



CB	Channel Bank	D1	Long-Haul Delay	D2	Long-Haul Delay
DR1	Echo path delay	DR2	Echo path delay	EC1	Echo Canceller
EC2	Echo Canceller	H1	Hybrid	H2	Hybrid
M1	Modem under test	M2	Modem under test	N1	Noise Generator
N2	Noise generator	PA	Protocol Analyser	R1	Receive attenuator
R2	Receive attenuator	T1	Transmit attenuator	T2	Transmit attenuator
W1	Wire simulator	W2	Wire simulator		

Figure 17/G.168 – Test No. 14 configuration

Requirements

The values of the settings should be as follows:

R1, R2 = 6 dB to simulate access/egress loss.

T1	= 3 dB to 9 dB (3 dB is the nominal level, 9 dB simulates a 6 dB level offset).
T2	= 3 dB.
DR1, DR2	= echo path delay $\leq \Delta$ ms.
M1, M2	= modem data transmission levels should be between –8 dBm and –20 dBm.
N1, N2	= set to produce signal-to-noise ratios of not less than 25 dB, and, no noise.
D1, D2	= set to produce a round trip delay of up to 520 ms, with D1 = D2.

With the H register initially set to zero and the NLP enabled, for the conditions specified above, the percentage of data errors should not increase when the echo canceller is enabled, compared with when the echo canceller is disabled, when data is exchanged between the two terminals for a period of at least three minutes.

4 Characteristics of an echo canceller tone disabler

4.1 General

The echo cancellers covered by this Recommendation should be equipped with a tone detector that conforms to this subclause. This tone detector responds to a disabling signal which is different from that used to disable the echo suppressor as described in clause 5/G.164 and consists of a 2100 Hz tone with periodic phase reversals inserted in that tone. The tone disabler should disable the echo canceller only upon detection of this signal. It should not disable with any other in-band signal, e.g. speech, or a 2100 Hz tone without phase reversals. The tone disabler should detect and respond to a disabling signal which may be present in either the send or the receive path.

The requirements for echo canceller disabling to ensure proper operation with ATME No. 2 equipment that transmits the 2100 Hz tone with phase reversals could be met by using either the tone disabler specified in this subclause, or the echo suppressor tone disabler specified in clause 5/G.164. However, use of the 5/G.164 disabler does not assure proper operation with all currently specified V-Series modems.

The term disabled in this subclause refers to a condition in which the echo canceller is configured in such a way as to no longer modify the signals which pass through it in either direction. Under this condition, no echo estimate is subtracted from the send path, the non-linear processor is made transparent, and the delay through the echo canceller still meets the conditions specified in 3.4.1. However, no relationship between the circuit conditions before and after disabling should be assumed. For one thing, the operation of echo cancellers with tonal inputs (such as the disabling tone) is unspecified. Additionally, the impulse response stored in the echo canceller prior to convergence (and prior to the disabling tone being sent) is arbitrary. This can lead to apparent additional echo paths which, in some echo canceller implementations, remain unchanged until the disabling tone is recognized. Also note that echo suppressors could be on the same circuit and there is no specified relationship between their delay in the enabled and disabled states. In spite of the above, it is possible, for example, to measure the round trip delay of a circuit with the disabling tone but the trailing edge of the tone burst should be used and sufficient time for all devices to be disabled should be allotted before terminating the disabling tone and starting the timing.

It should be noted that this condition does not necessarily fulfill the requirements for 64 kbit/s bit sequence integrity, for which case other means of disabling in line with 3.3 will apply.

A reference tone disabler is described in Annex A.

4.2 Disabler characteristics

The echo canceller tone disabler requires the detection of a 2100 Hz tone with phase reversals of that tone. The characteristics of the transmitted signal are defined in Recommendations V.25 and V.8. Phase variations in the range of 180° ± 25° should be detected while those in the range of 0° ± 110° should not be detected.

The frequency characteristics of the tone detector are the same as the characteristics of the echo suppressor tone detector given in 5.2/G.164.

The dynamic range of this detector should be consistent with the input levels as specified in Recommendations V.2 and H.51 with allowances for variation introduced by the public switched telephone network.

4.3 Guardband characteristics

Similar to that defined in 5.3/G.164 consistent with the dynamic range given in 4.2 above with the following exception. The detector should operate perfectly with white noise less than or equal to 11 dB below the level of the 2100 Hz signal. No definitive guidelines can be given for the range between 5 and 11 dB because of the variations in the test equipment used. In particular, performance may vary with the peak-to-average ratio of the noise generator used. As a general guideline, however, the percentage of correct operation (detection of phase variations of 180° ± 25° and non-detection of phase variations of 0° ± 110°) should fall by no more than 1% for each dB reduction in signal-to-noise below 11 dB. It is noted that it is possible to design a detector capable of operating perfectly at 5 dB signal-to-noise ratio.

4.4 Holding-band characteristics

The tone disabler, after disabling, should hold in the disabled state for tones in a range of frequencies. The bandwidth of the holding mode should encompass all present or possible future data frequencies. The release sensitivity should be sufficient to maintain disabling for the lowest level data signals expected, but should be such that the disabler will release for the maximum idle or busy circuit noise. Thus, the requirement follows:

The tone disabler should hold in the disabled mode for any single-frequency sinusoid in the band from 390-700 Hz having a level of –27 dBm0 or greater, and from 700-3000 Hz having a level of –31 dBm0 or greater. The tone disabler should release for any signal in the band from 200-3400 Hz having a level of –36 dBm0 or less.

4.5 Operate time

The operate time should be sufficiently long to provide immunity from false operation due to voice signals, but not so long as to needlessly extend the time to disable. The tone disabler is required to operate within one second of the receipt of the disabling signal.

4.6 False operation due to speech currents

It is desirable that the tone disabler should rarely operate falsely on speech. To this end, a reasonable objective is that, for an echo canceller installed on a working circuit, usual speech currents should not on the average cause more than 10 false operations during 100 hours of speech. In addition to the talk-off protection supplied by the disabling channel bandwidth, by guardband operation and by the operate time, talk-off protection can be supplied by recycling. That is, if taken place the operate timing mechanism should reset. However, momentary absence or change of level in a true speech which simulates the disabling signal is interrupted because of inter-syllabic periods, before disabling has disabling signal should not reset the timing.

4.7 False operation due to data signals

It is desirable that the tone disabler should rarely operate falsely on data signals from data sets that would be adversely affected by disabling the echo canceller. To this end, a reasonable objective is that, for an echo canceller installed on a working circuit, usual data signals from such data sets should not, on the average, cause more than 10 false operations during 100 hours of data transmissions.

4.8 Release time

The disabler should not release for signal drop-outs less than the ITU-T recommended value of 100 ms. To cause a minimum of impairment upon accidental speech disabling, it should release within 250 ± 150 ms after a signal in the holding band falls at least 3 dB below the maximum holding sensitivity in both directions of signal transmission.

4.9 Other considerations

Both the echo of the disabling tone and the echo of the calling tone may disturb the detection of the echo canceller disabling tone. As such, it is not recommended to add the receive and transmit signal inputs together to form an input to a single detector.

Careful attention should be given to the number of phase reversals required for detection of the disabling tone. Some Administrations favour relying on 1 to improve the probability of detection even in the presence of slips, impulse noise, and low signal-to-noise ratio. Other Administrations favour relying on 2 to improve the probability of correctly distinguishing between non-phase-reversed and phase-reversed 2100 Hz tones.

5 NLPs for use in echo cancellers

5.1 Scope

For the purpose of this Recommendation the term "NLP" is intended to mean only those devices which fall within the definition given in 1.3 and which have been proven to be effective in echo cancellers. It is possible to implement such NLPs in a number of ways (center clippers being just one example), with fixed or adaptive operating features, but no recommendation is made for any particular implementation. General principles and guidelines are given in 5.2. More detailed and concrete information requires reference to specific implementations. This is done in Annex B for the particular case of a "reference NLP". The use of this term denotes an implementation given for guidance and illustration only. It does not exclude other implementations nor does it imply that the reference NLP is necessarily the most appropriate realization on any technical, operational or economic grounds.

5.2 General principles and guidelines

5.2.1 Function

5.2.1.1 General

The NLP is located in the send path between the output of the subtractor and the send-out port of the echo canceller. Conceptually, it is a device which blocks low level signals and passes high level signals. Its function is to further reduce the residual echo level (L_{RES} as defined in 1.3.21) which remains after imperfect cancellation of the circuit echo so that the necessary low returned echo level (L_{RET} as defined in 1.3.22) can be achieved.

5.2.1.2 Network performance

Imperfect cancellation can occur because echo cancellers which conform to this Recommendation may not be capable of adequately modeling echo paths which generate significant levels of non-linear distortion (see I.6.2). Such distortion can occur, for example, in networks conforming to Recommendation G.113 in which up to five pairs of PCM codecs (conforming to Recommendation G.712) are permitted in an echo path. The accumulated quantization distortion from these codecs may prevent an echo canceller from achieving the necessary L_{RET} by using linear cancellation techniques alone. It is therefore recommended that all echo cancellers capable only of modeling the linear components of echo paths but intended for general network use should incorporate suitable NLPs. In specific network environments with low delay or high ERL, it may be possible to disable the NLP in an echo canceller with a sufficiently high ERLE. This may result in higher overall speech quality, as NLPs sometimes cause speech degradation.

5.2.1.3 Limitations

This use of NLPs represents a compromise in the circuit transparency which would be possible by an echo canceller which could achieve the necessary L_{RET} by using only modeling and cancellation techniques. Ideally, the non-linear processor should not cause distortion of near-end speech. In practical devices it may not be possible to sufficiently approach this ideal. In this case, it is recommended that NLPs should not be active under double talk or near-end single-talk conditions. From this it follows that excessive dependence should not be placed on the NLP and that L_{RES} should be low enough to prevent objectionable echo under double talk conditions.

5.2.1.4 Data transmission

NLPs may affect the transmission of data through an enabled echo canceller. This is under study.

5.2.2 Suppression threshold

5.2.2.1 General

The suppression threshold level (T_{SUP}) of a NLP is expressed in dBm0 and is equal to the highest level of a sine wave signal at a given moment that is just suppressed. Either fixed or adaptive suppression threshold levels may be used.

5.2.2.2 Fixed suppression threshold

With a fixed suppression threshold level the appropriate level to use will depend upon the cancellation achieved and the statistics of speech levels and line conditions found in the particular network in which the echo canceller is to be used. Values of fixed suppression threshold levels to be used are under study – see Notes 1 and 2.

NOTE 1 – As an interim guide, it is suggested that the suppression threshold level should be set a few decibels above the level that would result in the *peaks* of L_{RES} for a "2σ-talker" and a "2σ-ERL" being suppressed.

NOTE 2 – Results of a field trial reported by one Administration indicated that a fixed suppression threshold level of –36 dBm0 gave a satisfactory performance. A theoretical study, by another Administration, of an echo path containing five pairs of PCM codecs showed that for an L_{RIN} of –10 dBm0, the quantization noise could result in an L_{RES} of –38 dBm0.

5.2.2.3 Adaptive suppression threshold

A good compromise can be made between using a high T_{SUP} to prevent it being exceeded by loud talker residual echo and using a low T_{SUP} to reduce speech distortion on break-in by making T_{SUP} adaptive to the actual circuit conditions and speech levels. This may be achieved in a number of ways and no recommendation is made for any particular implementation. General guidelines applicable to the control algorithm and suppression threshold levels are under study.

5.2.3 Control of NLP activation

5.2.3.1 General

To conform to the recommendation made in 5.2.1.3, it is necessary to control the activation of the NLP so that it is not active when near-end speech is likely to be present. When "active", the NLP should function as intended to reduce L_{RES}. When "inactive", it should not perform any non-linear processing on any signal passing through the echo canceller.

5.2.3.2 Control guidelines

It is recommended that the following two guidelines should govern control of the activation of a NLP. First, because they are intended to further reduce L_{RES}, they should be active when L_{RES} is at a significant level. Second, because they should not distort near-end speech, they should be inactive when near-end speech is present. Where these two guidelines conflict, the control function should favour the second.

5.2.3.3 Static characteristics

A conceptual diagram showing the two operational states of a NLP is shown in Figure 18. The L_{Sin} L_{Rin} plane is divided into two regions, W and Z by the threshold WZ (T_{WZ}). In the W region the NLP is inactive while in the Z region it is active. Proper control of the NLP to ensure operation in the appropriate region requires recognition of the double talk condition or the presence of near-end speech. Imperfect detection of double talk combined with a high suppression threshold level will result in distortion of near-end speech. The echo canceller then exhibits some of the characteristics of an echo suppressor. A low suppression level will permit easy double talking, even if a detection error is made because the near-end speech will suffer only a low level of non-linear distortion. If the suppression threshold level is too low, then peaks of residual echo may be heard.

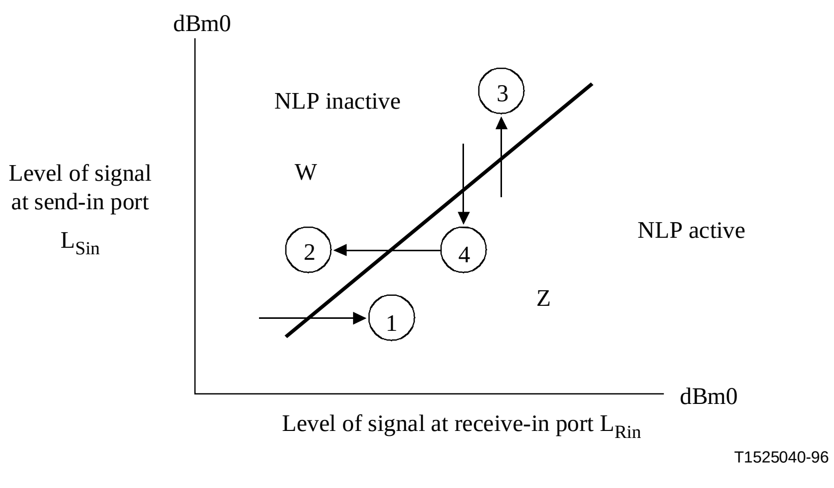


Figure 18/G.168 – NLP operating regions

5.2.3.4 Dynamic characteristics

The dynamic characteristics can be specified by stating the time that elapses when the signal conditions pass from a point in one area to a point in the other area before the state appropriate to the second area is established. Four such transitions are shown by arrows in Figure 18.

Transition No. 1 – W to Z, L_{Sin} constant, L_{Rin} increasing

In this case the L_{Sin} signal occurred first and the L_{Rin} is increasing to a sufficiently high level to override the L_{Sin} signal in the control path and cause the NLP to change from the inactive to the active state. Since this will cause distortion of the L_{Sin} signal (near talker speech in this case) the action should not be initiated too quickly.

Transition No. 2 – Z to W, L_{Sin} constant, L_{Rin} decreasing

In this case the L_{Rin} signal has overridden the L_{Sin} signal in the control path and the NLP is in the active state. The L_{Rin} signal is now decreasing. The NLP should remain in the active state sufficiently long to prevent echo, which is stored in the echo path, from being heard by the far talker.

Transition No. 3 – Z to W, L_{Rin} constant, L_{Sin} increasing

This transition is replicating the onset of double talk. As soon as possible after the L_{Sin} signal is detected, the NLP should be switched to the inactive state in order to minimize any distortion of the near talker speech.

Transition No. 4 – W to Z, L_{Rin} constant, L_{Sin} decreasing

In this case L_{Sin} has been recognized but is decreasing. Any action which is taken should favour continuing to permit the L_{Sin} signal to pass. This implies there should be some delay in switching the NLP back to the active state.

5.2.4 Frequency limits of control paths

Under study.

NOTE – Depending on the particular implementation of the NLP, the considerations and frequency response limits given in 3.2.4.2/G.164 for the suppression and break-in control paths of echo suppressors may also be applicable to similar control paths used in NLPs. These control paths may include the activation control and adaptive suppression threshold level control.

5.2.5 Signal attenuation below threshold level

The attenuation of signals having a level below that of the suppression threshold level of a NLP in the active state should be such that the requirements of 3.4.2.1 are met.

5.2.6 Testing of NLPs

The NLP may be considered as a special case of an echo suppressor which is limited to suppressing only low level signals. The types of test required to determine the NLP performance characteristics are very similar to the echo suppressor tests given in Recommendation G.164. However, depending on the specific implementation of a NLP, the transitions between areas W and Z of Figure 18 may not be as sharply defined as is the case for echo suppressors. Signals observed at the send-out port of the echo canceller may be distorted for short periods when transitions between the W and Z operating regions occur. Although Recommendation G.164 may be used as a guide to the testing of NLPs, it may be necessary to introduce unique test circuit modifications in order to make measurements on some specific NLP implementations. No recommendation can be given for a universal test circuit appropriate for all NLP implementations.

ANNEX A

Description of an echo canceller reference tone disabler

A.1 General

This Annex describes the characteristics of an echo canceller reference tone disabler. The use of the term reference denotes a disabling implementation given for guidance only. It does not exclude alternative implementations of a tone disabler which responds to the signals defined in Recommendations V.25 and V.8, and which also meets all of the criteria for reliability of operation and protection from false operation by speech signals.

A.2 Disabler characteristics

The echo canceller reference tone disabler described in this Annex detects a 2100 Hz tone with periodic phase reversals which occur every 450 ± 25 ms. The characteristics of the transmitted signal are defined in Recommendations V.25 and V.8.

A.2.1 Tone detection

The frequency characteristics of the tone detector used in this reference tone disabler are the same as the characteristics of 4.2, except that the upper limit of the dynamic range is –6 dBm0.

A.2.2 Phase reversal detection

The reference tone disabler responds to a signal which contains phase reversals of 180° ± 10° at its source (as specified in Recommendation V.25) when this signal has been modified by allowable degradation caused by the network, e.g. noise, phase jitter, etc. This disabler is insensitive to phase jitter of ±15° peak-to-peak in the frequency range of 0-120 Hz. This accommodates the phase jitter permitted by Recommendations H.12 and G.229. In order to minimize the probability of false disabling of the echo canceller due to speech currents and network-induced phase changes, this reference tone disabler does not respond to single phase changes of the 2100 Hz tone in the range 0° ± 110° occurring in a one second period. This number has been chosen since it represents the approximate phase shift caused by a single frame slips in a PCM system.

A.3 Guardband characteristics

Energy in the voiceband, excluding the disable band, must be used to oppose disabling so that speech will not falsely operate the tone disabler. The guardband should be wide enough and with a sensitivity such that the speech energy outside the disabling band is utilized. The sensitivity and shape of the guardband must not be such that the maximum idle or busy circuit noise will prevent disabling. In the requirement, white noise is used to simulate speech and circuit noise. Thus, the requirement follows:

Given that white noise (in a band of approximately 300-3400 Hz) is applied to the tone disabler simultaneously with a 2100 Hz signal, the 2100 Hz signal is applied at a level 3 dB above the midband disabler threshold level. The white noise energy level required to inhibit disabling should be no greater than the level of the 2100 Hz signal and no less than a level 5 dB below the level of the 2100 Hz signal. As the level of the 2100 Hz signal is increased over the range of levels to 30 dB above the midband disabler threshold level, the white noise energy level required to inhibit disabling should always be less than the 2100 Hz signal level.

NOTE – The possibility of interference during the phase reversal detection period has been taken into account. One potential source of interference is the presence of calling tone as specified in Recommendation V.25. If the calling tone interferes with the detection of the phase reversal, the entire disabling detection sequence is restarted, but only one time. Recommendation V.25 ensures at least one second of quiet time between calling tone burst.

A.4 Holding-band characteristics

The tone disabler, after disabling, should hold in the disabled state for tones in a range of frequencies. The bandwidth of the holding mode should encompass all present or possible future data frequencies. The release sensitivity should be sufficient to maintain disabling for the lowest level data signals expected, but should be such that the disabler will release for the maximum idle or busy circuit noise. Thus, the requirement follows:

The tone disabler should hold in the disabled mode for any single-frequency sinusoid in the band from 390-700 Hz having a level of –27 dBm0 or greater, and from 700-3000 Hz having a level of –31 dBm0 or greater. The tone disabler should release for any signal in the band from 200-3400 Hz having a level of –36 dBm0 or less.

A.5 Operate time

The reference tone disabler operates within one second of the receipt, without interference, of the sustained 2100 Hz tone with periodic phase reversals, having the level in the range –6 to –31 dBm0. The one second operate time permits the detection of the 2100 Hz tone and ensures that two phase reversals will occur (unless a slip or impulse noise masks one of the phase reversals).

A.6 False operation due to speech currents

It is desirable that the tone disabler should rarely operate falsely on speech. To this end, a reasonable objective is that, for an echo canceller installed on a working circuit, usual speech currents should not on the average cause more than 10 false operations during 100 hours of speech. In addition to the talk-off protection supplied by the disabling channel bandwidth, by guardband operation and by the operate time, talk-off protection can be supplied by recycling. That is, if speech which simulates the disabling signal is interrupted because of inter-syllabic periods, before disabling has taken place the operate timing mechanism should reset. However, momentary absence or change of level in a true disabling signal should not reset the timing.

A.7 False operation due to data signals

This meets the requirement in 4.7. To this end, the tone disabler circuitry becomes inoperative if one second of clear (i.e. no phase reversals or other interference) 2100 Hz tone is detected. The detector circuit remains inoperative during the data transmission and only becomes operative again 250 ± 150 ms after a signal in the holding band falls at least 3 dB below the maximum holding sensitivity. Thus the possibility of inadvertent disabling of the echo canceller during facsimile or low speed (< 9.6 kbit/s) voiceband data transmission is minimized.

A.8 Release time

The disabler should not release for signal drop-outs less than the ITU-T recommended value of 100 ms. To cause a minimum of impairment upon accidental speech disabling, it should release within 250 ± 150 ms after a signal in the holding band falls at least 3 dB below the maximum holding sensitivity.

ANNEX B

Description of a reference NLP

B.1 General

This Annex, which is for the purposes of illustration only and not intended as a detailed design (see 5.1), describes a reference NLP based upon concepts that are as simple as possible but having included in it a sufficient number of features to give guidance for a wide range of possible implementations. To this end two variants of the reference NLP are included. Both are based on a center clipper having either of the idealized transfer functions illustrated in Figure B.1. The suppression threshold level (determined, in this case, by the clipping level) in the first variant is adaptive, adaptation being by reference to L_{Rin}. Activation control is by reference to the difference between L_{Rin} and L_{Sin}. In the second variant the suppression threshold is fixed. It is assumed that the reference NLP is used in an echo canceller which can achieve a cancellation of the linear components of any returned echo of at least N dB. The value of N is under study.

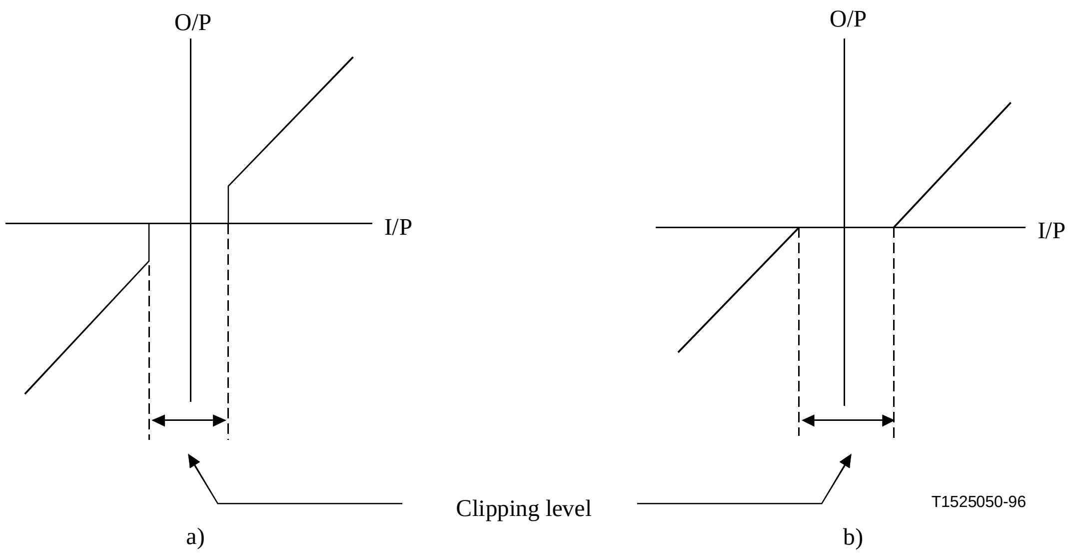


Figure B.1/G.168 – Two examples of idealized center clipper transfer function

B.2 Suppression threshold (T_{SUP})

Adaptive $T_{SUP} = (L_{Rin} - x \pm 3)$ dBm0 for $-30 \le L_{Rin} \le -10$ dBm0

Fixed $T_{SUP} = x'$ dBm0

NOTE – Values of x and x' are under study. Values of 18 for x and –36 for x' have been suggested but confirmation is required that these values are appropriate for use in all networks.

B.3 Static characteristics of activation control

$T_{WZ} = (L_{Rin} - y \pm 3)$ dBm0 for $-30 \le L_{Rin} \le -10$ dBm0

NOTE 1 – T_{WZ} is as defined in 5.2.3.3.

NOTE 2 – The value of y may be different for each variant, and this is under study. Values of x dB in the case of the adaptive T_{SUP} and ≥ 6 dB for y in the case of the fixed T_{SUP} seem reasonable.

B.4 Dynamic characteristics of activation control

Dynamic characteristics of the activation control are given in Tables B.1 and B.2. Also see Figure 18.

Table B.1/G.168 – NLP hangover times

Boundary		Initial signal (dBm0)		Final signal (dBm0)		Recom-mended value (ms)	Test No. (Rec. G.164)	Excursion (see Figure 18)	Test circuit, Figure:	Oscilloscope trace
		Send L_{Sin}	Receive L_{Rin}	Send L_{sin}	Receive L_{Rin}					
Z/W	Fixed	–25	–10	–25	–30	15-64	5	Transition 2	14/G.164	Trace 1 and trace 2 of Figure B.3 (β)
	Adaptive	–55 –40 –30	–20 –15 –5	–55 –40 –30	–40 –40 –30	Δ				
W/Z	Fixed	–15	–25	–40	–25	16-120	6	Transition 4	17/G.164	Trace 1 and trace 2 of Figure B.2 (β)
	Adaptive	–40 –40 –25	–50 –30 –15	–55 –55 –40	–50 –30 –15	30-50				

Table B.2/G.168 – NLP operate times

Boundary		Initial signal (dBm0)		Final signal (dBm0)		Recom-mended value (ms)	Test No. (Rec. G.164)	Excursion (see Figure 18)	Test circuit, Figure:	Oscilloscope trace
		Send L_{Sin}	Receive L_{Rin}	Send L_{Sin}	Receive L_{Rin}					
W/Z	Fixed	–25	–30	–25	–10	16-120	4	Transition 1	14/G.164	Trace 2 of Figure B.3 (α)
	Adaptive	–55 –40 –30	–40 –40 –30	–55 –40 –30	–20 –15 –5	15-75				
	Fixed	–40	–25	–15	–25	≤ 1	6	Transition 3	17/G.164	Trace 2 of Figure B.2 (α)
Z/W	Adaptive	–55 –55 –40	–50 –30 –15	–40 –40 –25	–50 –30 –15	≤ 5				

B.5 Frequency limits of control paths

See 5.2.4.

B.6 Testing

Tables B.1 and B.2 indicate, by reference to Recommendation G.164, how the dynamic performance of NLP activation control may be checked using sine wave signals. Figures B.2 and B.3 show the traces obtained on an oscilloscope for these tests.

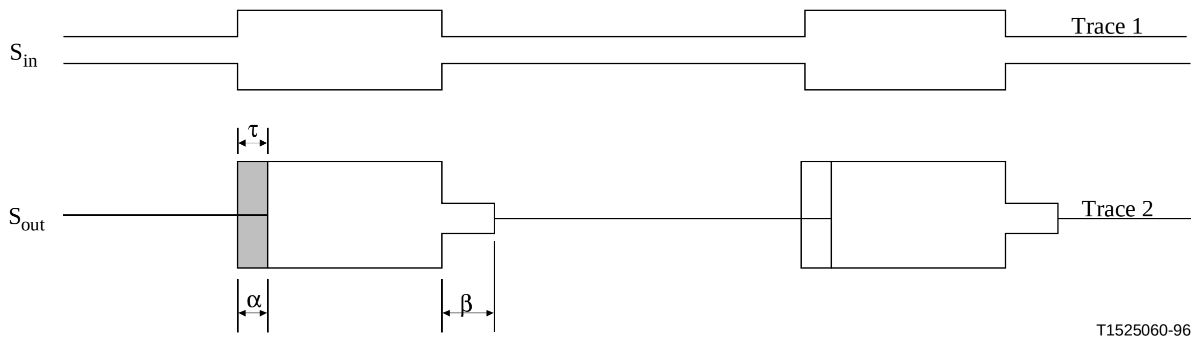
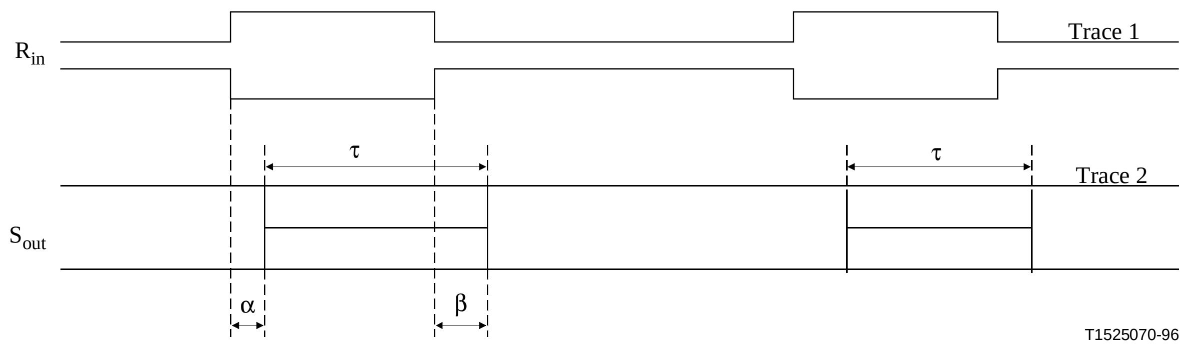


Figure B.2/G.168 – Traces for NLP operate and hangover times, L_{Rin} constant



α Operate time
β Hangover time
τ Time interval in which the distorted signal may be observed

Figure B.3/G.168 – Traces for NLP operate and hangover times, L_{Sin} constant

ANNEX C

Composite Source Signals for Testing of Speech Echo Cancellers – Signal, Description and Analysis

C.1 Introduction

This Annex describes the subset of Composite Source Signals that are used for testing speech echo cancellers in the network under single and double talk conditions. The exact definition of these signals is part of Recommendation P.501, Test Signals for Use in Telephonometry. First, a general description of Composite Source Signals is given. The following subclauses give the exact definition of both signals for testing speech echo cancellers under single and double talk conditions. Moreover, kinds of analysis are considered and described to test the specific parameters of echo cancellers especially for the tests of this Recommendation.

C.2 Composite Source Signal – General considerations

C.2.1 General description of the different sequences

Composite Source Signals, in general, consist of different sequences including voiced and unvoiced sounds as well as pauses.

Voiced signal produced from the "artificial voice" signal according to Recommendation P.50

The voiced signal part of CSS is the conditioning signal intended to activate possible speech detectors in voice-controlled systems and to reproduce voiced sounds of real speech in general. As the duration, beginning and end of the voiced signal are known exactly, this signal can also be used to measure the switching time for the direction of transmission under test. By means of the signal shape in the time domain, the switching time and delay time of the entire system can be determined. The duration of the signal amounts to 50 ms approximately.

Pseudo Noise Signal

The signal presented after the voiced artificial speech sound is the Pseudo Noise (PN) signal. This signal has certain noise-like features. The magnitude of its Fourier transform is initially constant with frequency while the phase is changing. For tests usually only the magnitude of the transfer function is of interest, the phase is not that important but can be determined as well.

The signal is produced as follows:

First a complex spectrum is produced in the frequency domain according to the following equation:

$$H(k) = W(k) \cdot \mathrm{e}^{j \cdot i_k \cdot \pi} \text{ where } \begin{cases} k = -M/2, \ldots\ldots, M/2, \text{ without } 0 \\ i_k \{+1, 0\},\ i_k = -i_{-k} \text{ random} \end{cases} \tag{C.2-1}$$

The index M is adjusted to the chosen FFT size (e.g. 2048, 4096 or 8192 points). The equation shows that the amount of the produced complex spectrum is constant for all frequencies if $W(k)$ is chosen equal to 1 for all frequencies, whereas the phase may be π or 0 for each frequency, corresponding to a random sequence. However, to produce a different weighting in the frequency domain, $W(k)$ can easily be adjusted in order to produce different spectra for the duration of the PN-sequence. Then, this spectrum will be transformed into the time domain by means of the inverse Fourier transform producing the following signal:

$$S(n) = \frac{1}{M} \sum_{k=-M/2,\, k \neq 0}^{M/2} H(k) \cdot \mathrm{e}^{j 2\pi \cdot n \cdot k/M} \qquad n = -M/2, \ldots\ldots, M/2 - 1 \tag{C.2-2}$$

NOTE 1 – Thus, a signal is produced which is limited in time (corresponding to the chosen length of the Fourier transform) and which is adjusted to the chosen FFT size correctly. If a longer time sequence is wanted, the signal can be cycled. This method permits time sequences of any length. The duration of this measurement signal amounts to about 200 ms by appropriate choice of M, the sampling rate and numbers of repetitions.

The Pseudo Noise sequence of the Composite Source Signal for measurements of speech echo cancellers is calculated in that way that $W(k)$ is chosen constant and the corresponding signal $S(n)$

(calculated by inverse Fourier transform) is filtered with a transfer function which is given below in C.3.1.

NOTE 2 – Typically the length of the FFT should be short for systems with highly time variant parameters such as companding techniques in order to get a good short time estimation of the time variant transfer function. For systems incorporating adaptive techniques such as echo cancellers or noise cancellers, a higher number of M (close to 200 ms signal duration) may be appropriate in order to have the autocorrelation function of the measurement signal not periodically within the processing window of the device under test.

Pause

The third part of the Composite Source Signal is a pause. Regarding the Composite Source Signal as a measurement signal that reproduces important characteristics of real running speech, the pause has the purpose to provide suitable amplitude modulation to the composite signal. Moreover it reproduces real speech pauses that occur in running speech signals as well. This also means a certain period without excitation signal, which gives the possibility to analyse noise or artefacts produced by the system under test. The length of the pause is chosen between 100 ms and 150 ms.

In order to achieve a long term offset free sequence the repeated CS-sequence should be inverted in amplitude (phase shift by 180º).

C.2.2 Calculation and analysis using a Composite Source Signal

When using CSS for measurements the sequence of voiced sound, pseudo noise signal and pause can be cycled. This means that after the pause the sequence starts again beginning with a voiced sound. Using this procedure, sequences of any length may be produced.

Having created a sequence as described above this signal can be handled like a standard measurement signal, e.g. a white noise signal or a switched pink noise. The level calibration (acoustical and electrical) is done using the whole sequence including voiced sounds, PN-sequences and pauses. In principle, a standard RMS meter with a bandwidth of 20 kHz operating with "fast" averaging can be used. Another method is to use a FFT analysis for level calculations. The parameters for the FFT based calculation are:

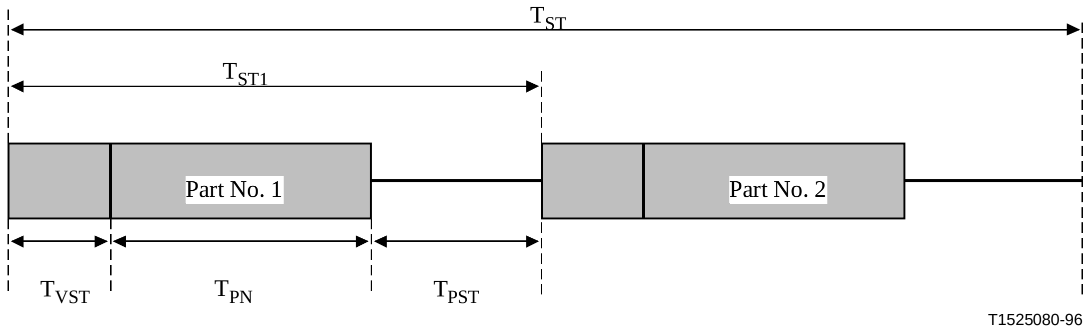
- sampling rate according to the one chosen for signal generation (preferred 44.1 kHz or 48 kHz);
- FFT length according to the one chosen for signal generation;
- rectangular windowing;
- no overlap;
- averaging over the **whole (cycled) sequence**, including voiced sounds, PN-sequences, pauses;
- calculation of the level from the power density spectrum derived by the FFT calculation (integration of the levels over all frequency components).

C.3 Bandlimited Composite Source Signal with speech-like power density spectrum – Practical realization for measurements of echo cancellers

Both Composite Source Signals described below in this Annex have a speech-like power density spectrum. This means that the noise sequences of both signals (the measurement signal and the signal to simulate double talk) are shaped with a decrease of 5 dB/octave towards higher frequency. The convergence characteristics of speech echo cancellers highly depends on the power density spectrum of the input signal. Therefore these Composite Source Signals were adapted in this way to reproduce the power density spectrum of real speech.

C.3.1 Composite Source Signal for single talk

Figure C.1 shows the principle construction of the Composite Source Signal for single talk.



Duration:	T_{VST} (voiced sound):	48.62 ms
	T_{PN} (pseudo noise):	200.00 ms
	T_{PST} (pause):	101.38 ms
	T_{ST1} (one period):	350.00 ms
	T_{ST} (whole period):	700.00 ms

Figure C.1/G.168 – Composite Source Signal for measuring echo cancellers (schematic)

Figure C.2 shows the power density spectrum of the bandlimited CSS and Figure C.3 the power density spectrum of the bandlimited voice signal.

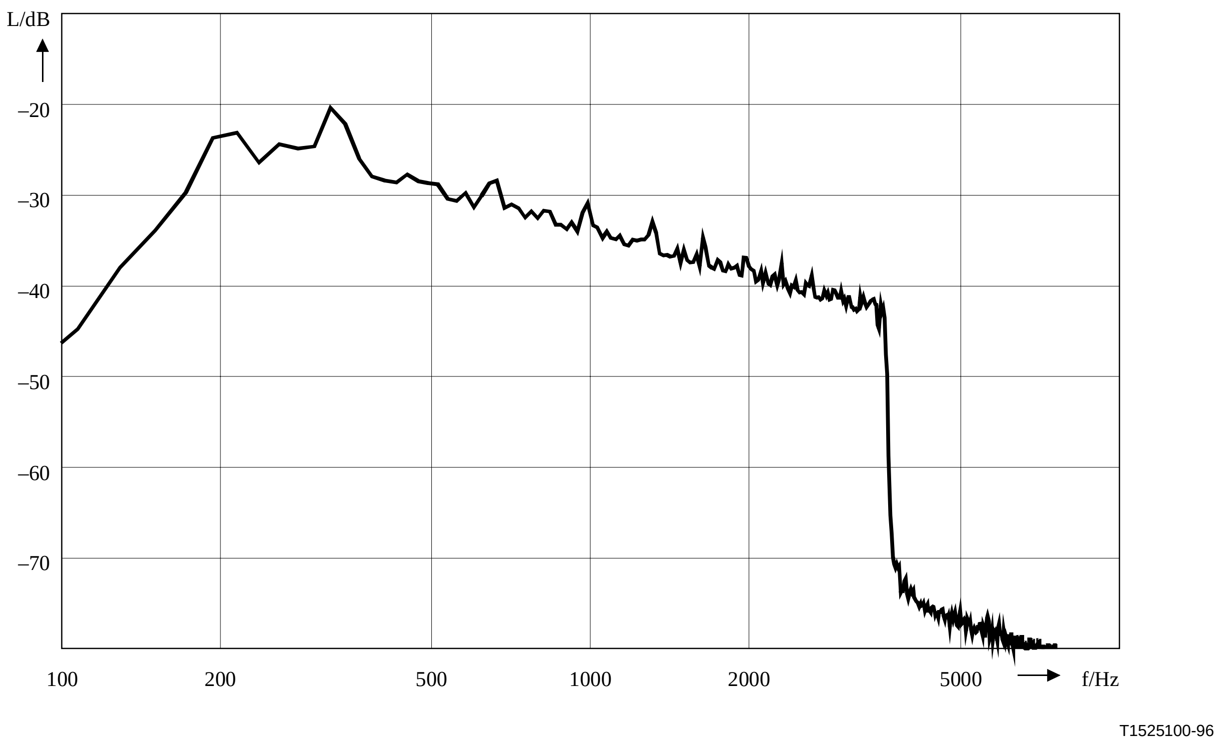


Figure C.2/G.168 – Power density spectrum of the bandlimited CSS (single talk signal, analysis window: Hanning)

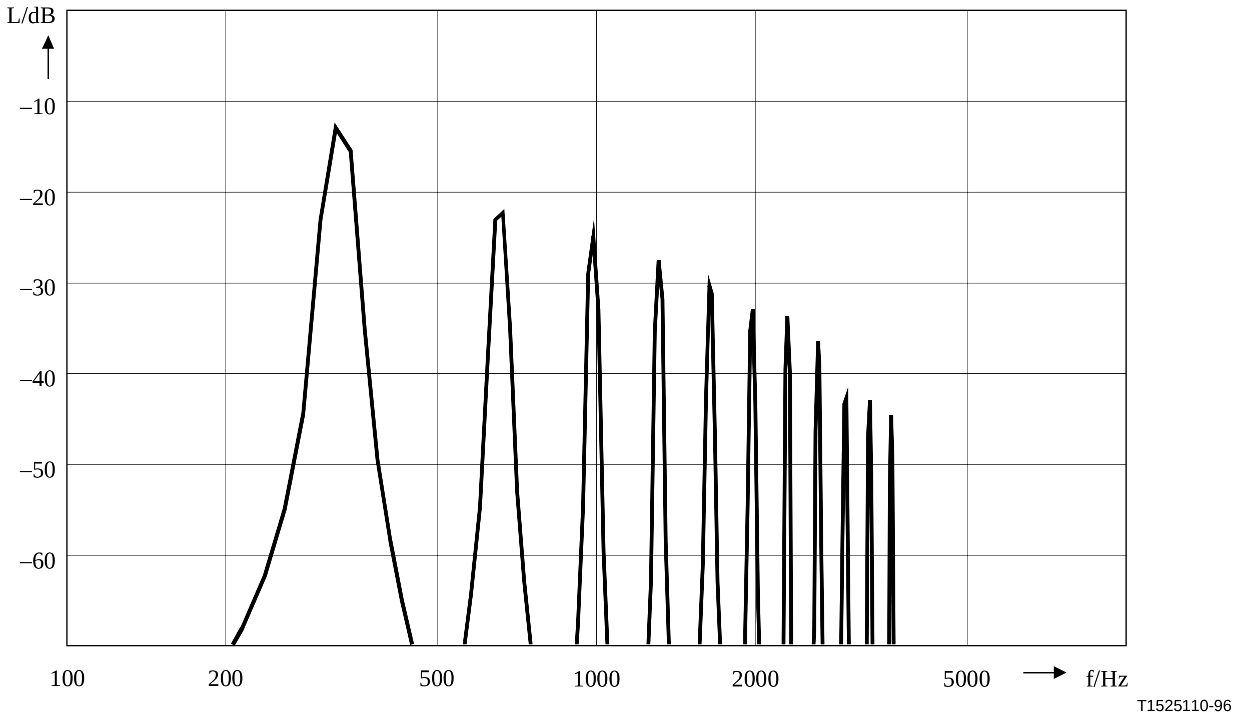


Figure C.3/G.168 – Power density spectrum of the bandlimited voiced signal (single talk signal, analysis window: Hanning)

Bandlimited voiced signal

In Table C.1, the 16-bit word values for the voiced signal, bandlimited between 200 Hz and 3.6 kHz can be found. According to a sampling rate of 44.1 kHz the 134 16-bit word values amount to 3.04 ms. The values are to be read in columns:

Table C.1/G.168 – 16-bit word values of the bandlimited voiced signal

–155	948	3224	4000	3129	1440	241	–888	–1853	–6137	–3474
276	1362	3370	4043	3043	1310	190	–957	–2121	–6560	–2508
517	1741	3500	4034	2914	1146	103	–1034	–2414	–6948	–1595
578	2043	3569	3974	2750	965	–9	–1103	–2707	–7301	–802
491	2276	3603	3862	2560	776	–138	–1146	–3017	–7568	
302	2422	3603	3724	2353	603	–267	–1181	–3319	–7732	
86	2500	3595	3577	2155	448	–388	–1190	–3612	–7758	
–103	2552	3586	3439	1991	345	–491	–1198	–3913	–7620	
–207	2595	3595	3336	1853	276	–569	–1215	–4224	–7310	
–198	2655	3638	3267	1750	250	–638	–1259	–4560	–6810	
–60	2758	3724	3224	1672	250	–698	–1327	–4922	–6155	
190	2896	3819	3198	1603	267	–759	–1457	–5301	–5344	
543	3060	3922	3172	1534	267	–813	–1629	–5715	–4439	

The values of the voiced signal in the frequency range 200 Hz-3.6 kHz again are calculated such that the RMS value of the voiced signal and the PN-sequence are equal. The sequence is repeated 16 times to achieve a length of 48.62 ms.

Pseudo noise signal generated using 2048 pt. FFT

The parameters for the PN-sequence are:

Sampling rate 44.1 kHz, 16-bit word length, length of Fourier transform 2048 points.

$$H(k) = \begin{cases} W(k) \cdot e^{j - i_k \cdot \pi}; k = -928, \ldots, +928 \text{ except } 0, i_k \{+1,0\}, \text{ random, } i_k = -i_{-k} \\ 0 \text{ else} \end{cases} \quad \text{(C.3.1)}$$

According to the above described Formula (C.2-2) the time signal is calculated by inverse Fourier-Transformation. This sequence is repeated 4.307 times to achieve a length of 200 ms for the PN-sequence. The crest factor of the PN-sequence is 11 dB ± 1 dB.

According to the frequency resolution of 21.5 Hz (44.1 kHz/2048) there are 928 FFT-values in the frequency range between 0 and 20 kHz. Each value *W*(*k*) (before filtering) is 152 680. It is calculated such that levels within a bandwidth of 20 kHz are the same for the voiced signal and the PN-sequence.

Pseudo noise signal generated using 8192 pt. FFT

According to the above described Formula (C.2.2) the time signal is calculated by inverse Fourier-Transformation. This sequence is repeated 1077 times to achieve a length of 200 ms for the PN-sequence. The crest factor of the PN-sequence is 11 dB ± 1 dB.

According to the frequency resolution of 5.4 Hz (44.1 kHz/8192) there are 3715 FFT-values in the frequency range between 0 and 20 kHz. Each value *W*(*k*) before filtering is 305 360. It is calculated such that levels within a bandwidth of 20 kHz are the same for the voiced signal and the PN-sequence.

In order to achieve the same RMS value for the bandlimited PN-sequence, the filter function shown in Figure C.4 should be applied. The filter is chosen such, that the levels of the filtered and the unfiltered PN-sequence are equal. In Table C.2 the filter corner frequencies are shown.

NOTE – By appropriate up- or down-sampling other sampling rates for the described sequence can be achieved. The interpolation filter used for up- and down-sampling should be close to an ideal rectangular filter. The stopband attenuation should be > 60 dB, the passband ripple < ± 0.2 dB.

Figure C.4/G.168 – Transfer function of the filter for bandlimiting the PN-sequence

Table C.2/G.168 – Table of filter corner frequencies

50 Hz	100 Hz	200 Hz	215 Hz	500 Hz	1 kHz	2.85 kHz	3.6 kHz	3.66 kHz	3.68 kHz
–25.8 dB	–12.8 dB	17.4 dB	17.8 dB	12.2 dB	7.2 dB	0 dB	–2 dB	–20 dB	–30 dB

For adaptive systems such as echo cancellers, a longer PN-sequence may be preferable in order not to have correlated measurement signals within the adaptation window. For those systems the FFT-length should be extended to 8192 points when using 44.1 kHz sampling rate as described above.

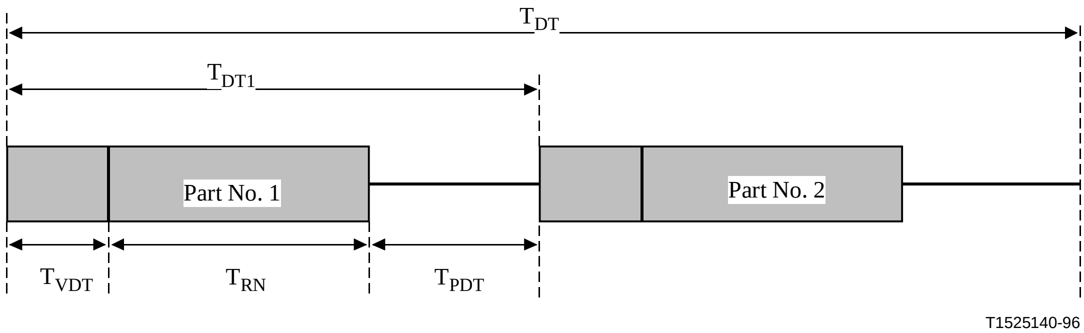
Pause

The length of the pause is chosen to 101.38 ms in order to achieve a complete length of 350 ms for the voiced sound, the Pseudo Noise sequence and the pause.

To achieve a long term offset free sequence this CS-sequence of 350 ms is repeated and inverted in amplitude (phase shift by 180º). The complete length amounts to 700 ms.

C.3.2 Bandlimited Composite Source Signal to simulate double talk

The double talk sequence is generated in the same way as the single talk signal. Figure C.5 shows the principle construction of the double talk signal. However the times of the voiced signal and the pause are slightly different in order to achieve a typical double talk condition with two signals applied the same time, signal present only in one channel, voiced signals present on both sides as well as voiced signals and unvoiced signals present the same time in the different channels. The correlation between single talk signal and double talk signal is low. This is achieved by choosing a different voiced signal with a different pitch frequency and a random noise signal instead of the PN-sequence. The duration of the voiced signal is 72.69 ms, the duration of the random noise signal is 200 ms and the duration of the pause amounts to 127.31 ms.



Duration:	T_{VDT} (voiced sound):	72.69 ms
	T_{RN} (random):	200.00 ms
	T_{PDT} (pause):	127.31 ms
	T_{DT1} (one period):	400.00 ms
	T_{DT} (whole period):	800.00 ms

Figure C.5/G.168 – Composite Source Signals to simulate double talk (schematic)

Figure C.6 shows the power density spectrum of the bandlimited double talk CSS and Figure C.7 the power density spectrum of the bandlimited double talk signal.

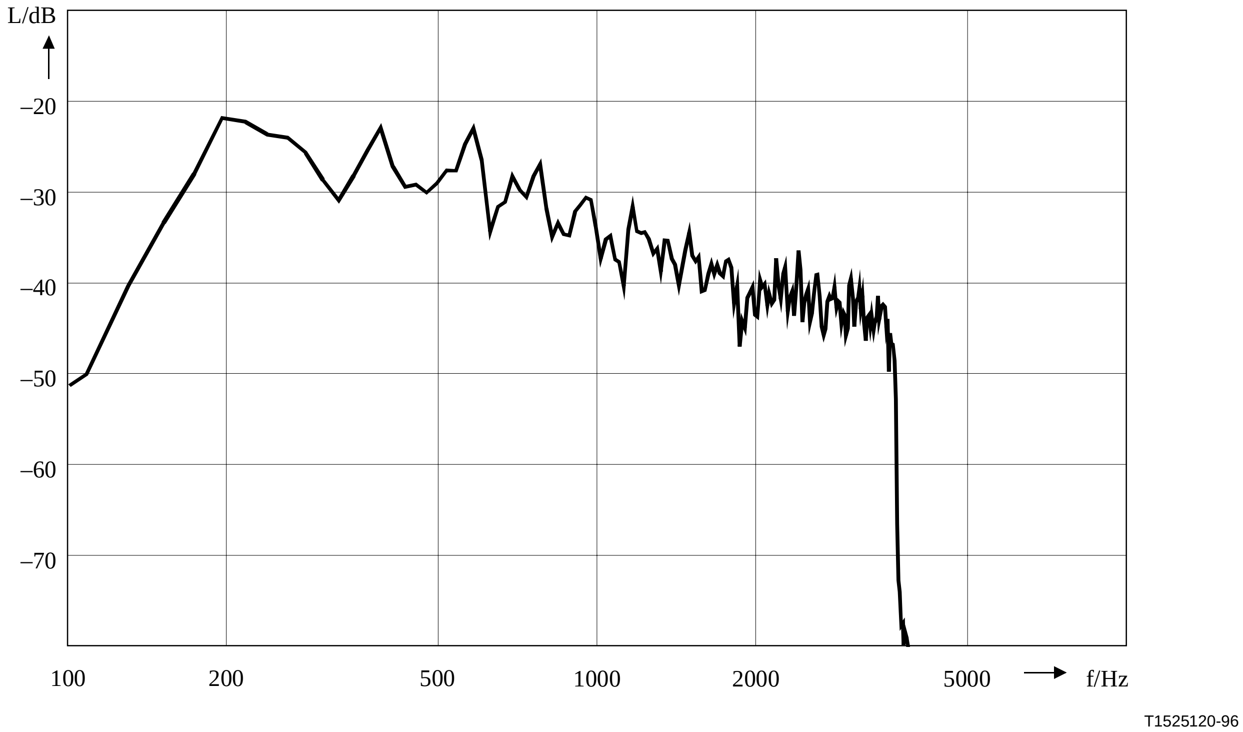


Figure C.6/G.168 – Power density spectrum of the bandlimited double talk CSS (analysis window: Hanning)

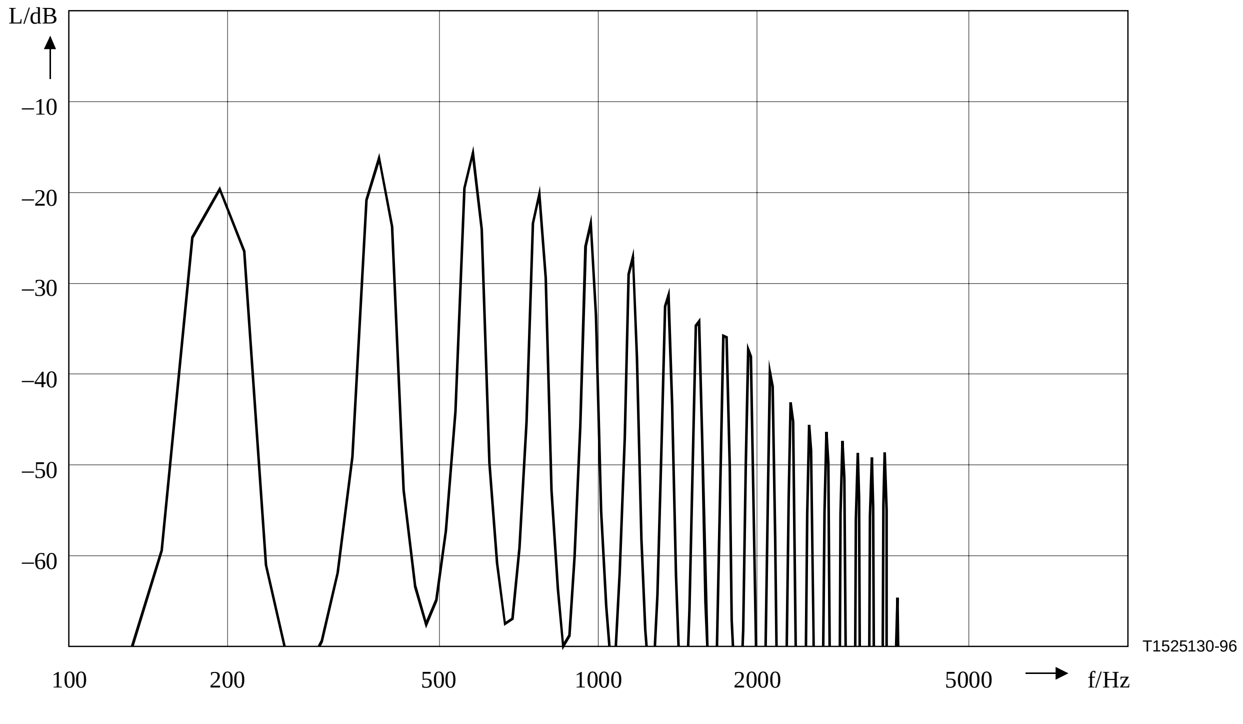


Figure C.7/G.168 – Power density spectrum of the bandlimited double talk voiced signal (analysis window: Hanning)

NOTE – By appropriate up- or down-sampling other sampling rates for the described sequence can be achieved. The interpolation filter used for up- and down-sampling should be close to an ideal rectangular filter. The stopband attenuation should be > 60 dB, the passband ripple < ± 0.2 dB.

Voiced signal

The voiced signal for double talk was chosen to have a different base frequency than the signal talk voiced signal. The values for the voiced signal for double talk can be found in Table C.3. The level of this sound again is the same as the one for single talk. Using a sampling rate of 44.1 kHz 229 16-bit word values represent 5.19 ms. The table is to be read in columns:

Table C.3/G.168 – 16-bit word values for the bandlimited double talk voiced signal

–198	1146	–8292	4827	5853	1422	–1293	–810	–690	–1052	–621
–112	871	–8715	5094	5715	1224	–1302	793	–724	–1043	–560
–9	560	–9077	5344	5560	1026	–1293	–767	–767	–1043	–509
103	233	–9370	5594	5387	819	–1267	–741	–793	–1052	–457
233	–121	–9542	5827	5215	603	–1250	–698	–819	–1060	–397
388	–491	–9542	6043	5043	388	–1233	–672	–845	–1060	–345
543	–871	–9361	6215	4879	181	–1224	–638	–853	–1060	–276
724	–1250	–8956	6344	4732	9	–1224	–603	–871	–1052	–207
896	–1638	–8327	6413	4586	–181	–1224	–595	–879	–1034	–112
1060	–2043	–7465	6422	4439	–328	–1224	–586	–888	–1017	
1233	–2465	–6396	6379	4276	–448	–1215	–595	–896	–991	
1388	–2896	–5163	6310	4086	–543	–1198	–603	–922	–957	
1517	–3345	–3827	6215	3870	–629	–1172	–621	–948	–931	
1638	–3819	–2448	6120	3629	–707	–1129	–629	–974	–905	
1747	–4310	–1103	6051	3370	–784	–1077	–938	–1009	–888	
1810	–4810	155	6000	3086	–871	–1026	–638	–1026	–862	
1845	–5319	1293	5991	2801	–948	–974	–638	–1052	–845	
1845	–5836	2241	5991	2534	–1026	–922	–638	–1069	–819	
1802	–6353	3034	6000	2267	–1112	–888	–638	–1077	–793	
1707	–6853	3655	6008	2034	–1181	–871	–638	–1069	–767	
1569	–7353	4138	5991	1819	–1241	–845	–647	–1060	–724	
1379	–7836	4517	5939	1612	–1276	–828	–664	–1060	–672	

In order to achieve the required length of 72.69 ms the values are to be repeated 14 times.

Random noise

The random noise is chosen as a white gaussian noise bandlimited at 20 kHz. The crest factor of the signal is 12 ± 1 dB. The RMS value of the bandlimited random noise is chosen to be the same as the one for the voiced signal.

In order to bandlimit the random noise between 200 Hz and 3.6 kHz, the filter function shown in Figure C.1 is used. This ensures the same RMS value for the bandlimited random noise.

Pause

The pause is chosen to 127.31 ms in order to achieve a length of 400 ms for the voiced sound, the random noise sequence and the pause.

Again, in order to achieve a long-term signal which is free of offset, this sequence of 400 ms is repeated and inverted in amplitude (phase shift by 180º). Thus the resulting length of the double talk signal is 800 ms.

Application

The application of the bandlimited Composite Source Signals for single talk as well as for double talk is for all testing where bandlimited systems need to be tested working non-linear and time variant and requiring the typical long-term power density spectrum of speech. The typical application is the testing of speech echo cancellers in the network. For all one directional tests the bandlimited CSS for single talk tests should be used. In case of tests in double talk conditions the double talk signal should be used in double talk direction (S_{gen}), whereas the single talk signal is fed in the far-end direction (R_{in}).

C.4 Appropriate analyses to determine convergence characteristics of speech echo cancellers using the composite source signal

The Composite Source Signal for testing speech echo cancellers and the second Composite Source Signal to simulate double talk are described above. If the echo signal level should be measured there are several possibilities of analysis technique. Calculations can be made in the time or frequency domain.

C.4.1 Calculation in the frequency domain

The signal level can be determined by calculations in the frequency domain, after the time sequence has been transformed by Fourier Transformation. This allows the level calculations in a certain frequency range, i. e. the telephone bandwidth of 300 Hz to 3.4 kHz. Another advantage is that the Fourier Transformation gives the possibility to analyse further characteristics of the echo signal in the frequency range, for example the echo attenuation versus frequency. For the Composite Source Signal a rectangle window should be used before calculating the Fourier Transformation. The Pseudo Noise sequence is generated with a 8192 points FFT. The sampling rate should be 44.1 kHz as described above in C.3.1 and C.3.2 for generating the Composite Source Signals. The sequence length used for transformation should be the complete length of 700 ms including the voiced sound, the Pseudo Noise sequence and the pause. Various measurements showed that due to signal delay or noise produced by the circuit under test, additional artefacts may appear during the pauses (e.g. switched residual echo signal or modulated background noise). Therefore it is suitable to analyse the echo signal over a sequence length of 700 ms, i. e. one whole period of the Composite Source Signal. A disadvantage of level calculations from the frequency range is due to the fact that this gives only a limited time resolution of one Fourier Transformation length. The level calculation in the frequency domain should be used to determine signal levels and residual echo levels after full convergence of after inhibiting adaptation.

C.4.2 Calculation in the time domain

The echo signal level calculation from the time domain is necessary for analysis of echo attenuation versus time because of its high resolution in the time domain. A suitable method is given through IEC 651, sound level meters. It describes the sound level measurement and recommends three different time constants, "Slow" (1000 ms), "Fast" (125 ms) and "Impulse" (35 ms). If measurement results of different laboratories should be compared, an agreement about the measurement procedure is necessary. A short time constant has advantages because of the highest possible resolution in the time domain, whereas longer time constants have the advantage that the results obtained with this kind of calculation demonstrate more the average level of the time sequence that is analysed. Especially if several measurements calculated for example as the level versus time are represented in the same picture, very short time constants may lead to confusing representations. This is due to the fact that using a very short time constant of, for example, 35 ms ("Impulse"), the calculation is more sensitive to even very small signal variations. For this reason, the use of the time constant "Fast" (125 ms) according to IEC 651 is more suitable for level calculations versus time.

This is a suitable method to analyse the convergence speed of speech echo cancellers at the beginning of adaptation. The echo signal level is calculated using the time constant "Fast" according to IEC 651. Level fluctuations due to input signal fluctuations can be eliminated if the echo signal level is referred to the input signal level. This represents the Echo Return Loss Enhancement (ERLE) versus time. A disadvantage is that no further analysis is possible in the frequency domain.

When using the meters of IEC 651, any peak detection or decay time constants referenced in IEC 651 should not be incorporated for measurements in this Recommendation.

C.4.3 Level calculations according to the active speech level P.56

Level calculations can also be done according to Recommendation P.56. This calculation is made from the time domain as well. It delivers one value and a percentage of speech activity. It may be suitable to calculate the residual echo level but there are more parameters that have to be defined to guarantee the same implementation of this algorithm. Difficulties may appear if echo signals with a very low level are analysed. It may fall below the recognition level for active speech. Another disadvantage is, although this is a calculation in the time domain it delivers only one value. It is not possible to achieve the level variation versus time, as it is important for convergence measurements. Therefore it is more suitable to analyse residual echo signal level using the Fourier Transformation as described in C.4.1 or the level calculation in the time domain for time varying echo signals (e.g. the convergence of echo cancellers) based on IEC 651 as described in C.4.2.

APPENDIX I

Guidance for application of echo cancellers

I.1 Scope

Echo cancellers are adaptive signal processors used to control echo; they are expected to replace echo suppressors in modern telecommunication networks. Echo cancellers are increasingly present on nearly every long distance connection and may be encountered singly or in tandem on a given connection. The purpose of this Appendix is to:

- explain the general principles of operation of echo cancellers;
- identify a limited set of application rules and the constraints under which echo cancellers operate;
- explain the relationship among the roles of the planners of a Public Switched Telephone Network (PSTN), modem manufacturers, private network planners, and end users regarding the control of echo (from sources inside and outside the PSTN) and the associated terminal design considerations;
- identify how echo cancellers may affect the perceived quality of speech, the quality of voiceband data, as well as the performance of various signal processing systems (such as digital and packetized circuit multiplication systems);
- identify both public and private network changes that may require additional study of echo cancellers, to fully understand how these changes may impact the functionality of present echo cancellers;
- explain how new services, if accepted for implementation, could have an evolutionary impact on echo canceller design.

I.2 Echo control in the PSTN

I.2.1 PSTN transmission planning

In the telephone network, the access line is typically a 2-wire facility between a customer premises and the switch, while the transmission facilities between the switches are typically 4-wire on long connections. At the 4-wire-to-2-wire conversion point, which typically occurs in a switch line card, a perfect impedance match cannot be achieved and thus a return signal, referred to as echo, results. Therefore, one of the major concerns of the PSTN planners is to ensure adequate echo control to provide satisfactory transmission performance.

For low-delay connections, echo is controlled by the insertion of appropriate transmission path losses, as defined in Recommendation G.131. Longer delay connections need echo control devices. It is the PSTN planners' role to design PSTNs so that the echo control devices installed provide adequate control of the echo from the 4-wire-to-2-wire conversions in the PSTN, and to ensure that the customer obtains satisfactory transmission performance.

In the past, echo suppressors were used to control echo in long distance networks. Today, however, the echo canceller is the device of choice. While PSTN planners and designers typically deploy the most current and modern technologies, it should be understood by modem designers, end users, and others, that for the foreseeable future the worldwide embedded plant may include some older echo control technologies on some connections. For example, connections through the PSTN may include some combinations of G.164 analogue or digital echo suppressors, G.165 analogue or digital echo cancellers equipped with G.164 tone disablers, and G.165 analogue or digital echo cancellers and G.168 digital echo cancellers equipped with G.165/G.168 tone disablers. The following two subclauses summarize the reasons for the use of echo cancellers instead of echo suppressors in modern telephone networks.

I.2.2 Echo suppressors

The principle of echo suppressors is well-known; it is summarized as follows: When speech is detected on the receive path, a very high attenuation is inserted in the send path. When double talk is detected, the send path is closed and a receive loss is inserted in the receive path. Thus, during double talk, there is no echo suppression, but the echo is much more attenuated than the direct speech. Other refinements are possible, as indicated in Recommendation G.164.

Many problems can occur in the operation of echo suppressors; this is because the decision as to which end is talking and which is listening is based essentially on the transmission levels. If the level of the echo is high and the level of direct speech is low, speech could be mutilated and/or it could be difficult to distinguish between single talk and double talk. This could also be the case at the beginning or at the end of a speech burst.

The problems are compounded on long-delay transmission paths because the pattern of conversation is usually changed. In addition, the cascading of echo suppressors is not recommended. In the case of voiceband data, a 2100 Hz tone is specified to permit disabling of the echo suppressor before the beginning of data transmission; this is for two reasons:

- to avoid insertion losses for modems with a secondary channel;
- to avoid delays due to hangover at turnarounds, thereby increasing the throughput.

Facsimile is a special case. Even if an echo suppressor is disabled by 2100 Hz tone, it may be re-enabled during a facsimile transmission. The tone disabler hangover time of an echo suppressor is specified as 250 ± 150 ms in 5.7/G.164. Therefore, periods of silence greater than 100 ms and smaller than 400 ms at the echo suppressor may cause the echo suppressor to be disabled, while periods greater than 400 ms cause it to be re-enabled. During a facsimile call, there are a number of silent periods that may be long enough to permit the re-enabling of the echo suppressor. In addition, some facsimile manufacturers have chosen to exceed the signal separation intervals specified in Recommendation T.30; therefore, echo suppressors may be re-enabled.

Enabled echo suppressors may distort the facsimile signals. One type of distortion is the truncation of fast turnaround signals. Typically, the echo suppressor operates in a single talk mode, so that when a signal arrives at the receive port, the suppression switch is activated and remains in that state until no signal arrives for a certain time. The recommended hangover time associated with each state transition is in the range of 24 to 36 ms[1], as specified in Table 4/G.164. The suppression hangover time guards against echo stored in the local echo path.

Now, Recommendation T.30 specifies that the guard time between V.21 and V.29 transmission should be 75 ± 20 ms. If a return signal from the local facsimile machine [within a V.21 message-response sequence or a V.21/V.29 sequence such as a confirmation to receive (CFR) followed by training] reaches the echo suppressor transmit port within 24 to 36 ms of the termination of the signal at the receive port, the persistence of echo suppression insertion losses or open-circuit condition may introduce an attenuation. As a result, the echo suppressor mutilates the initial portion of that fast turnaround signal. When this signal is part of the training-/training check signal, training might be disrupted and rate fallback ensues, or in a worse case, the call is terminated.

Similarly, an enabled echo suppressor may block a low-level secondary channel signal. If the level of that signal is high enough, the suppressor may enter the double talk mode, in which a receive loss is inserted. The result is a reduction in the levels of both the transmit and the receive signals, if echo suppressors are at both ends of the connection and are both in the double talk mode.

Finally, for certain combinations of propagation times and insertion losses, listener echo may cause the 2100 Hz tone to persist long enough to disable the echo suppressors. This echo may then contribute to the degradation of the image quality by reducing the signal-to-noise ratio during page transmission.

Prior to Recommendation V.32, most 2-wire modems used frequency division to provide duplex operation (i.e. different carrier frequencies were used for each direction of transmission). Data showed that some echo cancellers did improve the operation (i.e. reduce or eliminate bit errors) for low-speed modems designed according to Recommendations V.21, V.23, V.26 (alternative B), V.27 *ter* and V.29. Therefore, it was accepted that these modems benefited from an active echo canceller and a disabled echo suppressor. Recommendation G.165 recommends that echo cancellers be disabled with a 2100 Hz tone with phase reversals.

Recently, preliminary data have indicated that certain combinations of modems/echo cancellers, in various simulated network configurations, exhibit degraded performance when the echo cancellers are enabled. However modem manufacturers committee have not experienced any problems using low speed modems over circuits equipped with echo cancellers.

V.32 modems, in contrast, use the same band of frequencies in both directions and achieve duplex operation through the use of an integrated echo canceller. The echo canceller integrated in this voiceband data modem is not to be confused with the network echo cancellers that conform to Recommendation G.165, because the performance requirements for each are very different.

---

[1] Analogue echo suppressors are still in service; they have a suppression hangover time of 40 to 75 ms. Accordingly, a signal may be mutilated if it reaches the transmit port before 40 to 75 ms.

I.2.3 Echo cancellers

Echo cancellers are devices that use adaptive signal processing to reduce or eliminate echoes. Echo cancellers are placed in the 4-wire portion of a circuit, and reduce (or cancel) the echo by subtracting an estimate of the echo from the returned echo signal. Echo cancellers may operate on a single circuit or on a multiplexed facility, e.g. echo cancellers operate on a 64 kbit/s speech facility that is multiplexed into a primary rate link.

Echo cancellers are designed to:

- cancel linear echo path signals;
- refrain from cancelling the echo when requested to do so by an in-band disabling signal;
- return to an operational mode after being disabled when the in-band signal power level drops below a specified level for a specified period of time. This design allows some networks to transport voiceband data on the same speech channels. It also allows the echo canceller to re-enable during a voice call after it has been turned off erroneously (talkoff).

Echo cancellers are characterized by whether the interface path is analogue or digital, and/or whether the subtraction of the echo is by analogue or digital means. This Appendix is limited to echo cancellers that have a digital input and digital subtractors (Type C echo canceller as defined in Recommendation G.165).

Echo cancellers have the following main advantages over echo suppressors:

- send path transparency is improved;
- hangover introduces fewer impairments;
- there is no receive insertion loss;
- echo cancellation continues during double talk;
- cascading is possible (for well-designed echo cancellers).

Some echo cancellers are optioned to disable on the 2100 Hz tone specified in Recommendation G.164 for echo suppressors, and some are disabled with a 2100 Hz tone with periodic phase reversals of 180º ± 25º, as specified in Recommendations G.165 and G.168 for echo cancellers. Use of the G.165/G.168 tone is intended to allow echo cancellers to be disabled independently of echo suppressors.

Most modem manufacturers feel that network echo cancellers should be disabled for modems with integrated echo cancellers (e.g. Recommendations V.32, V.34), because an active network echo canceller operating in conjunction with the integral echo canceller in the modem may cause undesirable phenomena under specific but unlikely circumstances. Some of these cases are:

- The echo canceller incorrectly identifies the near-end signal as an echo and attempts to cancel it.
- When there is frequency offset in the echo path, the echo canceller injects bursts of reinforced echo interspersed with quiet periods.

Although neither case is likely, it was decided that the onus for making the decision to disable the network canceller should rest with the end users. Modem manufacturers had to rely on a unique technique to disable echo suppressors and echo cancellers.

Historically, manufacturers of modems with integrated echo cancellers have designed their modems to disable network-based echo cancellers. These modems disable network-based echo cancellers using the disabling tone specified in Recommendation G.165. Modem-based echo cancellers should accommodate three types of echoes simultaneously:

1) near-end echo;

2) far-end echo; and

3) any echo generated between the near-end and the far-end.

Because the range of echo path capacities needed for each case varies widely, three echo cancellers may be needed.

I.2.4 Responsibilities of modem manufacturers and end users

It is the responsibility of the modem manufacturers and end users to understand the characteristics of the network-based echo canceller fully and decide whether the echo cancellers should be enabled or disabled. If the modem manufacturers and end users decide that the network-based echo canceller functionality should be disabled, they should ensure that the terminal uses the appropriate approved methods, defined in Recommendations, to disable cancellers. Additionally, it is the end user's responsibility to ensure that terminals and private networks are designed to operate in a fashion compatible with the PSTN network-based echo cancellers. For example:

- Digital telephone sets are expected to control their own echoes, see Recommendations G.122, G.131 and P.310 (the PSTN network is not responsible for cancelling acoustic echoes).
- Terminals and private networks should be designed to provide circuit extensions compatible with the design intent of the PSTN, e.g. echo paths outside the PSTN-network should be linear and time-invariant or the terminal should control its own echo.
- Either the delay of the terminal or private network should be within the operational limits of the network-based echo canceller, or the terminal/private network should control its own echo.

I.3 Application rules and operational constraints

I.3.1 Public network transmission planning

The evolving digital PSTN requires a loss plan to ensure that appropriate transmission levels exist at the various A/D conversion points (see Recommendations G. 223, V.2 and M.1050). With such a plan, Pulse Code Modulation (PCM) overload distortion is avoided and signal levels allow the echo canceller to operate as per its design intent.

Guidance for transmission levels can be found in the G.100-Series of Recommendations for PSTNs that utilize analogue accesses and for connections from digital cellular networks. Encoders should be consistent with Recommendation G.711. For PSTNs with digital access, guidance for terminal design can be found in Recommendation P.310.

I.3.2 Delay considerations

As previously mentioned, conversion from the 4-wire toll network transmission facilities to 2-wire loop plant facilities should be made on all long connections. On these connections, it is the impedance mismatch at the hybrid that causes reflections of the incident signal at the 4-wire interface to occur (see Figure 2 as the reference model of the echo canceller). Because loops vary in composition, e.g. their length varies and they may be loaded or unloaded, a perfect balance cannot be obtained. Based on empirical data, it is commonly accepted that the average ERL should be considered to be approximately 11 dB. For those loops in which a poor impedance match is obtained, the reflections (talker echo) can become noticeable and objectionable when the delay between two telephones is greater than about 16 ms (32 ms round trip). See Recommendations G.131 and G.114 for guidance in this regard. It is the network planners' responsibility to determine at what point, i.e. for what delay threshold, a network echo control device will be implemented. This is a business decision that requires a balance between performance and cost.

NOTE – If an appropriate transmission plan is not implemented, echo may still occur in a circuit equipped with echo cancellers.

I.3.2.1 Echo Return Loss

The Near-End Speech Threshold (NEST), or Double Talk Detection Threshold (DTDT), is the level at which the echo canceller declares the presence of near-end speech, i.e. the occurrence of double talk, and stops its adaptation process. In other words, double talk is declared when:

$$LR_{out} - LS_{in} \leq NEST / DTDT$$

For example, when the NEST/DTDT of an echo canceller is provisioned for 6 dB, the echo canceller declares near-end speech and stops its adaptation process if $LR_{out} - LS_{in} \leq 6$ dB.

It is important that the NEST/DTDT value be provisioned such that the ERL > NEST/DTDT. For example, when the echo canceller is provisioned for NEST/DTDT = 6 dB, the echo canceller works properly with a 4-wire circuit path whose ERL > 7 dB. However, if the hybrid has ERL ≤ 6 dB, the echo canceller assumes that the echo at the S_{in} is a near-end speech. Because there is no adaptation during double talk, the end result is the presence of echo on the S_{out} path.

When the ERL is less than a provisionable threshold, the ERL of the circuit should be increased through level adjustments. It is the network planners' responsibility to ensure ERL is greater than the NEST/DTDT for which the circuit is provisioned.

I.3.3 Provisioning of the echo path capacity and echo path characteristics

The link from the canceller to the hybrid is often referred to as the "echo path of the circuit". The delay of the echo to be cancelled is determined by specifying the "echo path capacity" of the canceller. To specify this echo path capacity correctly, it should be remembered that some of the received power at port R_{out} is reflected by the hybrid and multiple reflections respectively resulting in echo at port S_{in}. The time it takes the signal at R_{out} to travel from the echo canceller to the hybrid and return to the echo canceller at port S_{in} should not exceed the provisioned echo path capacity; otherwise the echo cancellation process will not work properly. This time should include round trip propagation time delay over the transmission media, all intermediate equipment, and the dispersion due to the transmission characteristics of the circuit. This dispersion increases the effective duration of the impulse response of the circuit that should be taken care of by echo canceller. Note that the echo path can still include more than one source of echo, e.g. additional hybrids, cable gauge changes or other sources of echoes; many network configurations exist in which multiple 2-wire to 4-wire conversions exist in the echo path of an echo canceller.

It is the network planners' responsibility to ensure that echo cancellers are implemented in such a way that their echo path capacity is not exceeded on normal network connections, so that echo cancellation occurs. Cooperation among the interexchange carriers and the exchange carriers is required.

An echo canceller should be able to synthesize a replica of the echo path impulse response. Many echo cancellers model the echo path using a sampled data representation, the sampling being at the Nyquist rate (8000 Hz). Such an echo canceller, to function properly, should have sufficient storage capacity for the required number of samples (the maximum echo path delay in the network in which the canceller will be used will determine the required storage capacity). Typically, too few storage locations will prevent adequate synthesis of all echo path: too many storage locations will create undesirable additional noise due to the unused locations which, because of estimation noise, are generally not zero. It should be recognized that an echo canceller introduces an additional parallel echo path. If the impulse response of the echo path model is sufficiently different from the echo path impulse response, the total returned echo may be larger than that due to the echo path only.

I.3.4 End user/manufacturer/private network transmission planning

For convenience, the term "end user/manufacturer/private network planner" is used synonymously with "private network planner".

I.3.4.1 Transmission levels

The private network planner is expected to implement equipment that is consistent with the network transmission loss plan. Guidance is available in the form of Recommendations (see I.3.1). Further, the private network planner is expected to meet the relevant available requirements.

I.3.4.2 Delay considerations

The private network planner, like the public network planner, needs to make a conscious decision about how to control talker echo, and at what delay threshold to implement an echo control device in the private network. Note that if the private network connects to the PSTN on a 4-wire basis, the echo generated by the 4-wire-to-2-wire conversion may be cancelled by the network-based echo canceller. However, if the private network connects to the PSTN on a 2-wire basis and then converts to 4-wire for carriage, the private network planner should consider how to handle the echoes generated at the 4-wire-to-2-wire conversion points in the private network.

I.3.4.3 Echo return loss

It is the private network planner's responsibility to ensure that the ERL is greater than the NEST/DTDT for which the circuit is provisioned.

I.3.4.4 Provisioning of the echo path capacity and echo path characteristics

It is the private network planner's responsibility to ensure that any delay added in the private network does not exceed the delay specified by the PSTN service provider thus causing echo in the PSTN. Accordingly, the private network planner should ensure that the amount of delay added does not exceed the PSTN service providers allowable delay specification for network connection. If this specification is exceeded, the private network planner should take appropriate action to control echo.

I.4 Effect of cancellers on voice and data services

Network-based echo cancellers are present on connections that experience long delays. They should be designed to allow a speech channel to support voiceband data, including facsimile. This means that they should retain the capability of being disabled upon an appropriate request from customer terminal equipment. However, the modem manufacturer is responsible for determining if network-based echo cancellers should be enabled or disabled.

I.4.1 Interaction with voiceband data

Full-duplex data transmission in the voiceband can occur, depending on the modem modulation scheme. New modulation schemes are being introduced, and manufacturers should determine the optimal state in which the echo canceller should be when the modem is operating, i.e. if the canceller should be enabled or disabled, or whether the call should be routed on a connection that never has an echo canceller functionality present.

I.4.2 Interaction of echo control with facsimile transmission

The designers of facsimile terminals generated these terminals with the understanding that network providers were installing network-based echo control devices as per Recommendations G.164 and G.165. Thus, PSTN network planners were expected to continue to evolve the network in such a way that it would not knowingly prevent the continued carriage of a permissive voiceband data/facsimile service.

Although facsimile machines may transmit a G.164 disabling tone at the beginning of a call, there is no requirement to guarantee that the power of in-band signals will continue to hold the echo control devices in the disabled state for the duration of the call. Echo control devices conforming to G.164 (echo suppressors), G.165 (echo cancellers) and G.168 (digital network echo cancellers) are designed to re-enable when the signal level drops below a predefined threshold for a predefined period of time, once the call is in progress. The reason for this is that echo control devices conforming to Recommendations G.164 and G.165 are designed to become re-enabled if no signal energy is present in both directions of signal transmission for a period greater than 100 ms (minimum) to 400 ms (maximum) (see 5.2/G.164 and 5.5/G.164).

The V.27 *ter* modulation scheme employed by Recommendation T.30 is protected against the mutilation of the training sequence by echo suppressors (by using an unmodulated carrier prior to the training signal). In contrast, the V.29 modulation scheme is not protected. Some implementations are based on proprietary solutions to this problem (most notably the addition of an unmodulated carrier prior to V.29 transmissions of the same format as that used during V.27 *ter* transmissions). Unfortunately, these schemes are not universally recognizable by terminals produced by different modem manufacturers. Thus, if the guard time between V.21 and V.29 transmission from the facsimile machine exceeds the T.30 time limit of 75 ± 20 ms, it is possible that an echo suppressor will be re-enabled. In this case, the initial portion of the training check sequence could be mutilated, preventing the connection establishment.

The presence of echo can interfere with facsimile transmission in two ways:

- The echo could be misinterpreted as a T.30 protocol message and then interrupt the handshake between the two ends machines. This is particularly important if the facsimile machines are not protected against echo.
- The echo can reduce the S/N ratio necessary for the good transmission of images data.

Echo could be present for the following reasons:

- Echo suppressors are disabled (to avoid errors in voiceband transmission). As explained earlier, enabled echo suppressors may cause errors in voiceband data transmission. However, it may be preferable to keep them enabled during facsimile transmission.
- If echo cancellers are disabled according to the procedures of Recommendation G.164 (2100 Hz tone), then, depending on the propagation delay and the response time of the facsimile machines, echo could be present during the initial handshake. This could disrupt the establishment of the call. This imposes a limit of 400 ms on the time during which no energy could flow in either direction for the echo control device to re-enable. If these echo cancellers remain disabled, the echo of the V.21 signal may confuse the facsimile machine at the other end and/or confuse the facsimile demodulator of the network Packetized Circuit Multiplication Equipment (PCME)/digital circuit multiplication equipment (DCME). The image quality may be affected as well.
- Echo cancellers that respond to the G.165/G.168 disabling tone are not disabled by the 2100 Hz tone without phase reversal.

Other vulnerable instances during the connection are when handshakes are exchanged between pages. Disabled echo cancellers could allow echo at these instances; enabled echo cancellers, in contrast, control echo, including listener's echo.

Under some conditions, echo cancellers disabled using the G.164 procedures (2100 Hz) may affect the connection establishment or the quality of facsimile transmission because they may be disabled inadvertently by the called station identification (CED) tone; hence, echo control does not function as expected.

It should be noted that a number of echo cancellers already deployed in the PSTN are not able to eliminate completely short echo bursts that could occur while the canceller is reconverging after transitions between the narrow-band signals, such as the CED tone or the V.21 High-level Data Link Control (HDLC) handshake, the wide band image signals (e.g. V.29 or V.27 *ter* signals), and again, narrow-band signals. In the future, it still will not be possible to guarantee that all echo cancellers will be able to avoid this problem.

NOTE – This Appendix does not discuss explicitly the case in which there is one echo canceller on one side of the connection and an echo suppressor on the other side; this "mixed case" can be deduced from I.2.2 and I.2.3.

Current Recommendations imply that echo cancellers should be enabled during facsimile transmission. Generally, echo suppressors do not provide the same level of performance for speech, voiceband data, or facsimile. Enabled echo suppressors could cause failures due to clipping and/or mutilation of the training check sequence, thereby preventing the establishment of the facsimile connection. However, it may be better to enable echo suppressors during facsimile transmission to protect against both talker and listener echoes and avoid their interference with facsimile at connection establishment and/or during image transmission.

The main conclusion is that it is better to use echo cancellers that are disabled according to the G.165/G.168 procedures.

I.5 High-level speech

I.5.1 Introduction

A number of sources could produce high speech levels in the network. In hands-free telephones, for example, the microphone may allow high speech levels to be generated. With this perspective, Recommendation G.165 includes an overload test (Test No. 8) at levels exceeding 0 dBm0 (the provisional values for the test are +3.05 dBm0 and +3.25 dBm0) and to increase the maximum test levels from –10 dBm0 to 0 dBm0.

The presence of high speech levels may cause increased non-linearities that would degrade the performance of some echo cancellers, especially echo cancellers that have not been implemented in a fully digital manner. Another area in which high signal levels may cause difficulty is in the double talk detection and non-linear processor control circuits. These are discussed in the following two subclauses.

I.5.2 Double talk detection and activity detection

The performance of echo cancellers is very dependent on the activity detection and double talk detection algorithms used. For example, if double talk is not recognized quickly, the near-end speech masks the residual echo that is used to update the impulse response model of the echo canceller.

The following items are for further study:

- The effect of activity detection algorithms for low bit rate coders.
- The effect of double talk detection in the presence of high signal levels.

New echo canceller requirements for echo canceller design may result.

I.5.3 Effects of low bit rate coders

This topic is for further study.

I.5.4 Effects of a non-linear echo path

The theory of echo cancellation assumes that the echo path is linear and time-invariant. Therefore, it is critical that clipping and non-linear distortion do not occur in the echo path between R_{out} and S_{in}. If any clipping does occur, it is important that it be slight, infrequent, and that it occurs only during double talk conditions. Otherwise, the environment needs to be corrected, e.g. frequency offset removed or implementation of an acceptable transmission plan ensured.

One potential source of problems with high-level speech stems from the resultant non-linearities in the echo path. For optimal echo canceller performance, it is essential that the signal fed into the echo canceller's R_{in} port be linearly related to the signal at the echo canceller's S_{in} port. If any non-linear distortion of high-level speech occurs, the distortion should occur before it is used by the echo canceller so that the same clipped signal is sent to the R_{out} port. However, echo canceller performance may still degrade if the echo path is not linear.

Some echo cancellers use the signal at R_{in} as its internal received signal R_{rcv}, and also pass R_{in} to the R_{out} port. This is acceptable provided that there is no clipping or other non-linear distortion of one signal leg that does not occur with the other. Otherwise, the echo path does not appear to be linear to the echo canceller and, consequently, performance suffers.

Additionally, clipping or other non-linear distortion should not be "added" to the signal at the S_{in} port. This is most important when:

1) echo is present only at the S_{in} port; or

2) both echo and near-end speech are present and the double talk detector has not been triggered, since clipping (distorting) one affects the other.

I.5.5 Guidelines for R_{out} usage in echo cancellers

The configuration in which the same signal feeds both R_{in} and the echo path may result in degraded performance if R_{out} is not digitally equivalent (bit for bit) to R_{in} under all signal conditions. The signal R_{rcv} used internally by the echo canceller after passing through the R_{in} port can be used as the source signal for the echo path. Therefore, it is recommended that R_{out} (which is used to drive the echo path) should be digitally equivalent to R_{rcv}.

I.6 Network and service evolutionary considerations

I.6.1 Bit transparency of echo cancellers

A 2100 Hz disabling tone with phase reversals should cause the echo canceller to disable and provide an analogue clear-channel signal path (see Recommendation G.165). In other words, a tone between 300 Hz and 3400 Hz should pass with its power level and frequency unaltered through the echo canceller, but 64 kbit/s bit-transparency is not guaranteed (see 3.3/G.165). It is noted that 64 kbit/s transparency is achievable and is implemented in some echo cancellers, but to remain in that state, the in-band power level should remain above a predefined power level.

If cancellers are to be applied to trunks and disabled by use of a "switch to echo canceller signalling channel", the canceller should support a 64 kbit/s clear channel capability, if such capability is to be provided.

I.6.2 Non-linearities and time variant effects in the echo path

Two issues are related to the introduction of non-linear and time variant signal processing techniques in the PSTN:

1) the occurrence of voice compression in the echo path; and

2) the occurrence of digital insertion losses.

- With the increasing use of voice compression in the public and private voice networks, specifically 32 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM, see Recommendation G.726), the occurrence of a voice compression codec in the echo path becomes more likely. Measurements carried out with echo cancellers including an ADPCM circuit in the echo path have shown that the deterioration of the residual echo level may exceed 8 dB and more.
- With the increased use of digital techniques in processing voiceband signals, digital insertion losses are being implemented increasingly in digital pads. Such digital padding typically occurs in PSTN End-Offices when they act as a host to a digital remote line module as well as in Customer Premises Equipment (CPE), such as Private Branch Exchanges (PBXs). Improperly designed digital pads may add substantial non-linearities to the transmitted signal, including the returned echo signal, therefore degrading the canceller performance. The need to maintain linearity in digitally padded signals should be recognized.

The effect of further voice compression techniques as regards non-linearity affecting canceller performance is for further study.

I.6.3 Voice compression between tandem cancellers

The use of voice compression as part of the voice transmission path could also affect connections that use tandem cancellers. Figure I.1 shows a circuit in which tandem cancellers are in place, and voice compression is used only between the two cancellers. Although the canceller closer to the hybrid would not be affected, the canceller on the network side would see a non-linear or a time variant echo paths as described in I.5.4 and I.6.2. The performance of the tandem still may be acceptable if the canceller closer to the network remains stable and maintains a return loss enhancement. Theoretically, the canceller on the network side would not see an echo because the canceller on the distant end has removed it. However it is recommended that the cancellers on the network side should be removed effectively from the connection.

The conditions under which the performance is not degraded is a candidate for further study.

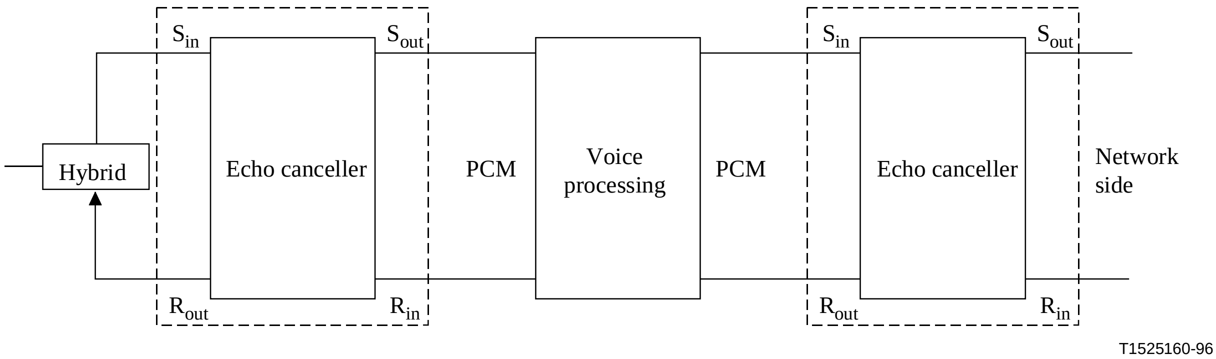


Figure I.1/G.168 – Voice compression between tandem cancellers

I.6.4 Tandeming of echo cancellers

It is generally accepted that properly designed echo cancellers can be tandemed with little or no penalty in performance. In Recommendation G.131, Rule B indicates that G.165 echo cancellers can be connected in tandem without echo performance degradation (see 2.3.2.1.1/G.131). With the increasing use of dynamic routing and special features such as call forwarding, and because of the long delay introduced by low bit rate speech coders in cellular applications, it is very likely that some connections have more than one echo canceller.

Subjective tests on some echo cancellers verify that tandeming poses no problems under most conditions. However, reports have suggested that other echo cancellers cannot be tandemed without problems. In these cases, it is imperative that the PSTN and/or private network planners ensure that echo cancellers that cause undue performance degradation when tandemed are not allowed to operate in a tandem mode.

Test results showed that improper design of some of the auxiliary circuits, such as NLPs, could cause problems when the echo path delay for one of the echo cancellers in tandem exceeds its echo path capacity. For example, in some echo cancellers, the NLP may operate at inappropriate times during double talk. This occurs when the hangover time in the NLP circuit does not match the echo path delay characteristics.

To illustrate, assume that the NLP algorithm is designed to operate on the basis of the NEST/DTDT value. In the case where the echo path delay capacity of an echo canceller is exceeded, the echo arrives later than the "expected" time. As a result, the comparison is in effect between power levels of a later far-end speech burst and an unrelated near-end speech burst. Based on this scenario, clipping can occur. However, it is reasons like these that make it important that PSTN and private network planners ensure that the echo path capacity of the echo canceller echo paths are never exceeded, unless additional echo control measures are taken inside the private network.

This problem is mitigated since it only occurs during double talk, and most situations involving tandeming of echo cancellers do not include many cases in which the echo path capacity is greatly exceeded. Finally, with some adjustments to the time constants of the NLP, partial improvements can be made.

It has been observed that if an echo canceller converges too quickly, it can have annoying side effects if it is used in a situation where its echo path capacity is exceeded (such as sometimes occurs with tandem echo canceller operation). Therefore, the echo path capacity of an echo canceller should be 4 to 6 ms larger than the maximum expected network delay, as estimated from Table 1/G.114. This takes into account the effect of dispersion. For example, to take into account a maximum pure delay of 44 ms, a 48 ms canceller could be selected.

Figure I.2 shows three pairs of back-to-back ECs (EC_A, EC_B, EC_C), four delay generators (D_1, D_2, D_3, D_4), and two hybrids (designated by return loss R_1 and R_2). The values of R_1 and R_2 should be appropriate for proper operation of the nearest canceller (e.g. at least 6 dB). By selectively disabling ECs (either singly or in pairs), and varying the delays, it is possible to capture the relevant attributes of telephone connections with ECs.

As an example (see Figure I.2a), 50 ms delay at D_1 and D_3, 100 ms at D_2, 150 ms at D_4, and 4-wire termination in place of R_2 are a reasonable representation of an international call originating at an analogue station and terminating in a digital cellular network. In this case, EC_A and EC_B might be at opposite ends of the international facility, with EC_C is in the cellular network (in which case, the right-facing canceller of the pair might be inoperative or absent). Alternatively (see Figure I.2b), EC_A might be in a national (land-based) network while EC_B and EC_C are at the ends of an international facility. In this case, D_1, D_2, and D_4 would be fairly short and D_3 would provide delay consistent with an international connection.

The sample configuration in Figure I.2 can be extended easily if more pairs of ECs are required. In particular, inclusion of a fourth pair of ECs (and another delay generator) would capture the important features of an international connection with ECs in each national network as well as at the ends of the international facility.

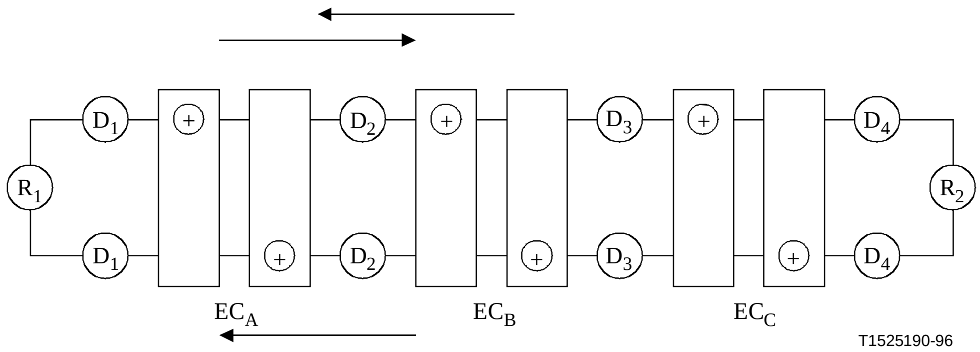


Figure I.2/G.168 – Reference connection for tandem ECs

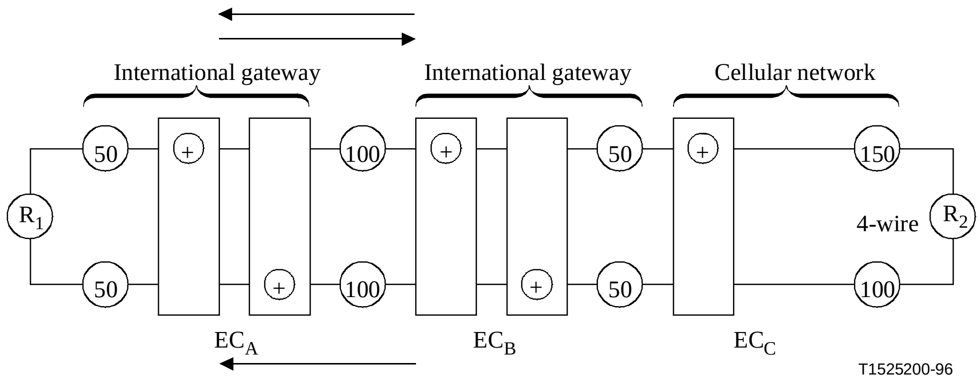


Figure I.2a/G.168 – Example of international connection originating at analogue station and terminating in a digital cellular network

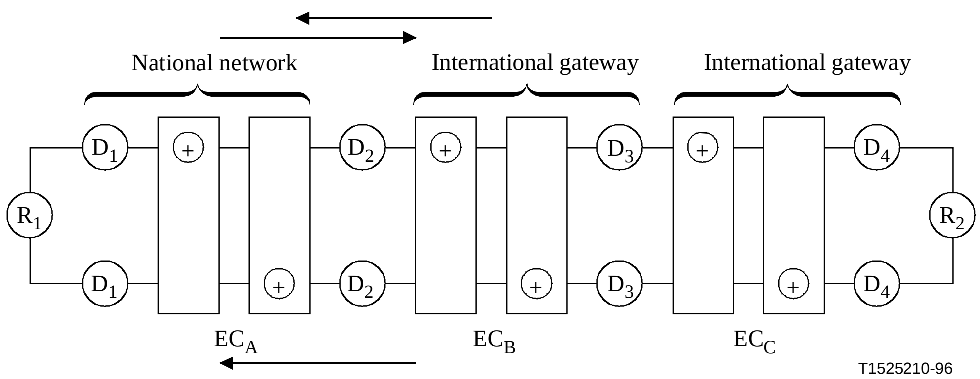


Figure I.2b/G.168 – Example of international connection

I.6.5 Convergence speed

High speed of convergence is desirable to reduce echo during initial acquisition, and to minimize echo when the echo path is changing.

Some echo cancellers generate noise in trying to continuously adapt to the echo path. This may be related to adaptation speed. The effect is very noticeable and annoying, especially during double talk, when the adaptation process is suspended. For some echo canceller implementations, as the speed of adaptation is increased beyond the optimum speed, the accuracy of the transfer function after adaptation becomes poorer. High speed of convergence is desirable for initial acquisition, while lower convergence may be needed for subsequent tracking, since the echo transfer function changes very slowly. The need of high convergence speed when time varying components are in the echo path is still under study.

I.6.6 Acoustic echo control and environments

Acoustic echo control is becoming an important issue due to the increase in hands-free telephone sets. Although there is some commonality between issues encountered for acoustic echo cancellation and network echo cancellation, there are also many differences. The issues of level points, natural echo path loss (or gain), degree of loss-switching, as well as level and/or type of singing (howling) protection are all important to a study of acoustic echo cancellers. In addition, it is important that an acoustic echo canceller is capable of working in harmony with a network-based electric echo canceller.

Analogue hands-free telephones which allow real double talk may produce an acoustic echo signal. This echo signal is added to the electrical echo signal coming from the 4-wire/2-wire connection of the hybrid termination and cannot be reduced sufficiently if it is decorrelated. Analogue hands-free telephones including dynamic compression devices may amplify the ambient room noise during speech pauses and transfer it to the echo canceller input in the send path. Due to the signal dependent switching of hands-free telephones, the level of a double talk signal may be reduced at the echo canceller input in the send path. This may lead to increased clipping by the non-linear processor because the level of this double talk signal may fall below the threshold level.

I.6.7 New circuit switched service

It has been suggested that there may be merit in modifying the disabling mode of G.165 cancellers so that upon the receipt of the disabling tone, the canceller disables until the connection is released.

It has been suggested that a customary procedure in some networks for initiating a digital transmission through a PCM-only digital voice network is to precede the digital transmission with a 2100 Hz tone to disable any echo cancellers/suppressors in the circuit. However, the cancellers remain disabled only as long as the transmitted digital data, when interpreted as PCM samples, contain sufficient energy to maintain the cancellers in the disabled state. The success of this non-standard approach depends upon the content of the digital data stream, and, as the maintenance of a sufficient power level cannot be guaranteed, proprietary means are usually used to ensure that the cancellers remain disabled. When the disabling signal is digitally generated, additional complexity is required for terminals that use a bit-level protocol and a serial interface, due to the inability of the terminal to establish octet alignment with the octets used in the transmission channel.

In this context, the need for an in-band, non-octet aligned echo canceller disable signal is for further study.

I.6.8 Comfort noise

As the telephone network migrates to more digital connections, it becomes more likely that the echo path will be analogue while the long distance connections path will be digital. One consequence is that the long distance path has a low idle channel noise while the echo path has a higher idle channel noise. This in turn leads to a situation called "noise modulation". When the NLP operates, the talker "hears" the idle channel noise of the digital long distance path, but when the NLP releases, the talker "hears" the idle channel noise of the echo path and the far-end environmental noise. Thus, the talker hears intervals of speech with background noise followed by intervals of silence, which can be very annoying in some instances.

There are two known approaches for comfort noise. The first solution is to insert pseudorandom noise during the silent interval. The second solution is to allow some of the background or idle channel noise to pass through the NLP.

I.7 Special DCME/PCME networking considerations

It is well known that echo control is needed in long-delay circuits, such as on satellite links. In addition, echo control may be needed, even for a short terrestrial circuit, because of the additional buffering delay in a DCME or a PCME. If echo is present, it may be classified as speech and reduce the compression gain.

One possible interaction relates to the potential loading effect of the comfort noise injected by the echo canceller on a DCME/PCME (see Figure I.3). The operation of the echo canceller may modulate the near-end analogue noise injected into the S_{in} port of the echo canceller. This could cause the adaptive speech detector of the DCME/PCME to falsely classify this change in noise level as the presence of speech. In this case, the DCME/PCME transmits the noise spurt as if it were speech and thus increases the activity factor of the circuit. The consequence is a decrease in the compression gain, and in some systems, an increase in the occurrence of freeze-out.

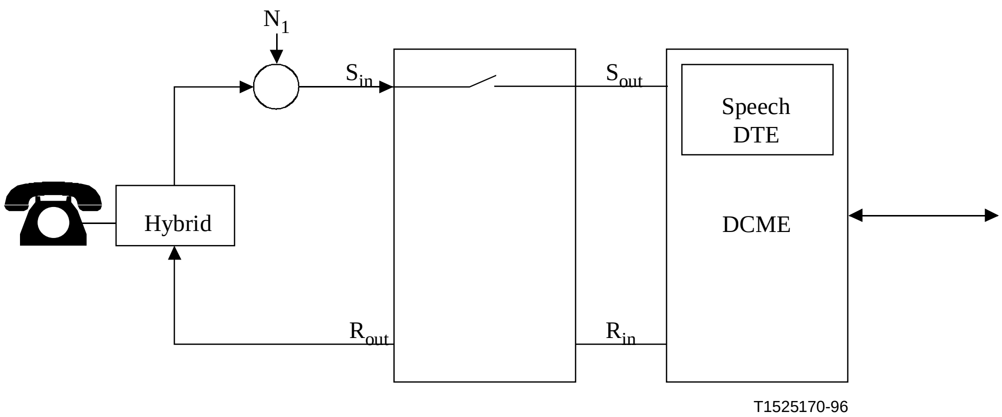


Figure I.3/G.168 – Speech detector/echo control device interaction

I.7.1 Detailed interaction

This interaction occurs as follows:

1) Receive speech arrives at the receive input (R_{in}) port of the echo control unit.

2) The echo suppression switch or canceller NLP activates, stopping the echo or residual echo and attenuating the near-end-generated analogue terrestrial noise (N_1) present at the send input (S_{in}) port.

3) If very little noise is generated between the echo control send output (S_{out}) port and the DCME speech detector input, the speech detector threshold adapts to its minimum level (typically, –50 dBm0).

4) When the receive speech stops, after a suitable echo control unit hangover time, the echo suppression switch or canceller NLP closes and the near-end-generated terrestrial noise (N_1), as seen by the DCME speech detector, reappears as a step change in noise level.

5) The step change in noise level may exceed the speech detector threshold, causing the DCME to transmit a noise spurt as if it were speech. The noise spurt duration is a function of the adaptation speed of the speech detector and the near-end-generated terrestrial noise level.

This sequence is repeated for every speech spurt and produces a very annoying speech-correlated noise spurt heard by the far-end talkers every time they stop speaking.

This interaction is not limited to single echo control device network configurations. Figure I.4 shows a typical network configuration, with multiple echo control devices interacting with a DCME/PCME speech detector. In this configuration, the DCME/PCME speech detector may respond to unit step increases in noise power, which result from echo suppressor switch or echo canceller center clipper activations in the send paths of echo control devices 1 and 3. (The role of the center clipper is to remove the residual echoes due to imperfect cancellation.) The DCME/PCME speech detector first experiences a unit step increase in noise power from echo control device 3 switch activation, followed by a second step increase from echo control device 1 switch activation. The extent to which the DCME/PCME speech detector incorrectly responds to these step increases in noise power is a function of the noise power levels N_1, N_2, N_3, and N_4 and the specific DCME speech detector threshold adaptation algorithm. For example, the dual step increases in noise presented to the DCME/PCME speech detector, which result from switch or center clipper activation at locations 1 and 3, are masked if the power level N_4 is excessively high. Likewise, high noise power levels at N_2 or N_3 may mask step increases in noise power caused by echo control unit 1.

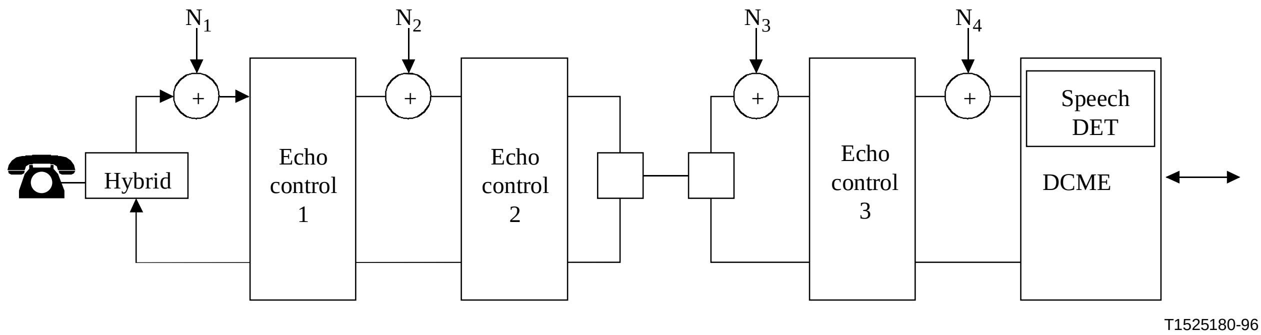


Figure I.4/G.168 – Multiple echo control devices in a DCME/PCME network configuration

I.7.2 Possible solutions

There are several methods for dealing with the interactions between the echo control devices and the DCME speech detector. In one approach, the echo control device could be modified to monitor the terrestrial-generated noise at the send-input port. When the send transmission path is broken, noise at the proper level is injected into the send-output toward the DCME, keeping the noise seen by the speech detector at a constant level (comfort noise) and avoiding speech detector activation. Not all echo cancellers may implement this approach, due to the number of different echo control devices in use and the uniqueness of this application.

In a second approach, the speech detector adaptive threshold of the DCME/PCME is frozen in the presence of speech on the corresponding receive channel.

A third approach is to specify an adaptive speech detector with a fast adaptation feature, which would track step changes in noise level and minimize the noise spurts.

The approaches described above may be unacceptable due to the number of different echo control devices in use and the uniqueness of the proposed application. Further, the large base of cancellers prevent consideration of a fast phasing in of new echo cancellers.

This subject requires further study and may result in recommendations to modify Recommendation G.165 for new generation echo cancellers. The main point of this subclause is that the solution depends on the speech detection procedures of both the DCME/PCME and the echo canceller.

I.8 Considerations regarding echo canceller performance during double talk

I.8.1 Introduction

A double talk situation (as the name suggests) could occur when both signals present at the input of an echo canceller have characteristics of active speech.

The CSS, which simulates double talk, consists of a burst (of constant energy) and a real pause. However, it was shown that a better double talk signal could be achieved with a signal in which the two bursts with high signal energy are identical to the original one, while the pause is filled up with a shortened CSS consisting of a voiced sound, a noise sequence and a real pause. Figure I.5 shows the modified double talk signal with the sequence length of 800 ms.

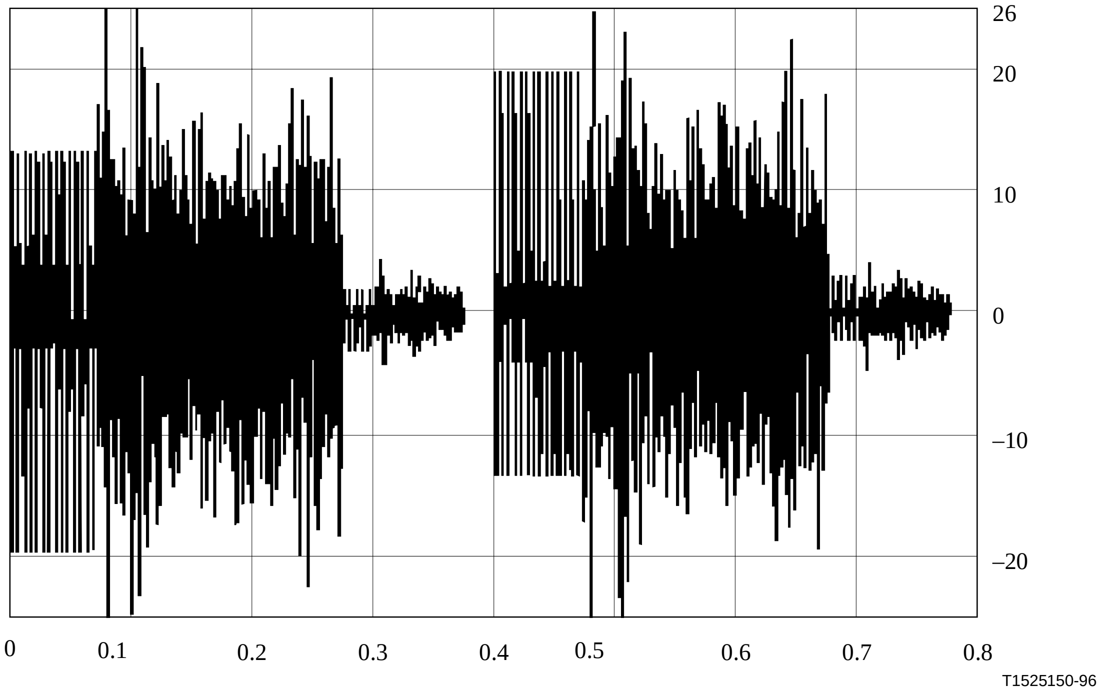


Figure I.5/G.168 – Modified double talk signal

I.8.2 Double talk parameters

The performance of echo cancellers under conditions of double talk is determined by many parameters. From recordings and listening tests, the following aspects can be derived:

- One of the most noticeable degradations when listening to the double talk signal is caused by the erroneous insertion of the NLP during continuous speech.
- Speech gaps caused by the NLP as mentioned above during continuous speech seem to be more annoying than clipping at the beginning of a double talk sequence (first word of the first sound).
- The detectability and annoyance of echo signals during double talk depend on the echo signal level and echo signal sound

- Echo cancellers behave in a different way if double talk occurs at the beginning of adaptation or after full convergence.

I.8.3 Analysis of technical parameters that influence performance under double talk conditions

The following parameters need to be taken into account when defining a test signal and the measurement procedure:

- signal levels at the R_{in} and S_{gen} port (receive signal and double talk signal);
- level ratio and time pattern of both signals at the R_{in} and the S_{gen} port;
- time of double talk (convergence status of echo canceller);
- duration of double talk.

The performance of the echo canceller itself is determined by technical parameters such as:

1) sensitivity of double talk detection;
2) threshold level of double talk detection (insertion of NLP, possible adaptive control);
3) reliability of double talk detection;
4) switching time of NLP;
5) double talk detection hangover time;
6) frequency characteristics of the residual echo signal loss measured between the R_{in} and S_{out} port (ERL versus frequency, "sound" of echo signal);
7) divergence during double talk.

Again these influencing parameters can be separated into different groups:

- the points 1 to 3 are determined by the performance of double talk detection (sensitivity, reliability);
- the switching characteristics of the NLP determine points 4 and 5;
- the points 6 and 7 (frequency characteristics, i.e. ERL vs. frequency and divergence) depend on the filter algorithm.

A suitable measurement procedure to evaluate double talk performance requires a suitable measurement sequence. A combination of two Composite Source Signals was derived to reproduce typical speech double talk sequences. Both signals are described in Recommendation P.501. The length of the measurement CSS is 700 ms, the second CSS, which simulates the double talk fed into the echo path, has a duration of 800 ms. Due to their different sequence length the level relationships on both echo cancellers inputs R_{in} and S_{gen} (or) S_{in} change, if both signals are periodically repeated. The same relationships can be observed if real speech signals are used. Various measurements on different echo cancellers demonstrate that this signal combination reproduces results under double talk conditions compared to speech.

I.8.4 Subjective testing

Subjective tests were performed with the purpose of qualifying those effects of the echo cancellation process that cannot be captured by objective measurements.

The results of the tests, judged by untrained and trained listeners, pointed out that one degradation of a transmitted double talk signal is mainly determined by the insertion of the NLP during continuous speech if the echo cancellers are fully converged. If CSS is used for the objective measurements, the switching characteristics can easily be determined after a burst of the double talk signal, because the time duration of all components is exactly defined for CSS. Subjective tests pointed out that a good double talk performance can be achieved even with double talk signal levels 15 dB lower than the receive input signal levels. If the bursts of the double talk CSS are not completely transmitted, the probability is high that longer speech gaps occur. The methodology of doing subjective tests for echo cancellers and their correlation with objective tests are for further study.

ITU-T RECOMMENDATIONS SERIES

Series A	Organization of the work of the ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	**Transmission systems and media, digital systems and networks**
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communication
Series Z	Programming languages