INTERNATIONAL TELECOMMUNICATION UNION

# ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

# G.168

(04/2000)

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

(Formerly 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 |
| ***TESTING EQUIPMENTS*** | |
| ***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 the list of ITU-T Recommendations.*

**ITU-T Recommendation G.168**

# Digital network echo cancellers


## Summary

This revised version of ITU-T Recommendation G.168 covers the design and testing of digital network echo cancellers for use in circuits where the delay exceeds the limits specified by ITU-T Recommendation G.114 [1] and G.131 [3]. This revision improves the testing methodology and better defines the performance requirements. Many of the tests have been updated and improved. Annex D has been added on realistic echo-path models that are used in the testing of echo cancellers, and Appendix II has been added on the measurement methods for characteristics of echo paths.



## Source

ITU-T Recommendation G.168 was revised by ITU-T Study Group 15 (1997-2000) and approved under the WTSC Resolution 1 procedure on 4 April 2000.

## FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. 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 turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSC Resolution 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

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. 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, ITU 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

**Page**

**ITU-T Recommendation G.168**

# Digital network echo cancellers

## 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).

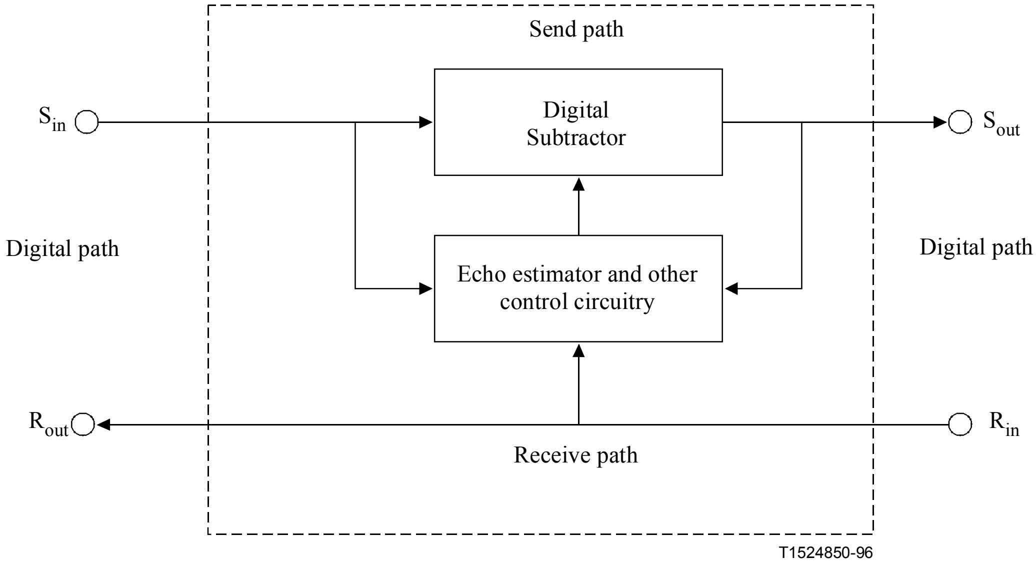



**Figure 1/G.168 – Digital transmission echo canceller using digital subtraction**

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.

This ITU-T 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 ITU-T Recommendations G.114 [1] and G.131 [3]. It is necessary for all echo control devices used on international connections to be compatible with each other. Echo cancellers designed to this ITU-T Recommendation will be compatible with each other, with echo cancellers designed in accordance with ITU-T Recommendation G.165 [5], and with echo suppressors designed in accordance with ITU-T Recommendation G.164 [4]. Compatibility is defined as follows:

Given:

1) that a particular type of echo control device (say Type I) has been designed so that satisfactory performance is achieved when any practical connections is equipped with a pair of such devices, and

2) that another particular type of echo control device (say Type II) has been likewise designed;

then Type II is said to be compatible with Type I if it is possible to replace an echo control device of one type with one of the other type, without degrading the performance of the connection to an unsatisfactory level. In this sense, compatibility does not imply that the same test apparatus or methods can necessarily be used to test both Type I and Type II echo control devices.

Freedom is permitted in design details not covered by the requirements. This ITU-T 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 ITU-T Recommendation G.165 [5], such as performance on voice, FAX, residual acoustic echo signals and mobile networks.

This ITU-T Recommendation does not apply to echo cancellation through active 2-wire/4-wire hybrids or 2-wire repeaters. This ITU-T Recommendation does not cover acoustic echo cancellation as per ITU-T Recommendation G.167 [7].

This ITU-T 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 voice-band 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 ITU-T 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 ITU-T Recommendation.

## 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.


[1] ITU-T Recommendation G.114 (2000), *One-way transmission time*.

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

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

[4] CCITT Recommendation G.164 (1988), *Echo suppressors*.

[5] ITU-T Recommendation G.165 (1993), *Echo cancellers*.

[6] ITU-T Recommendation G.167 (1993), *Acoustic echo controllers*.

[7] CCITT Recommendation G.223 (1988), *Assumptions for the calculation of noise on hypothetical reference circuits for telephony*.

[8] CCITT Recommendation G.229 (1988), *Unwanted modulation and phase jitter*.

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

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

[11] CCITT Recommendation G.726 (1990), *40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)*.

[12] ITU-T Recommendation M.1050 (1998), *Lining up an international point-to-point leased circuit with analogue presentation to the user*.

[13] ITU-T Recommendation P.310 (2000), *Transmission characteristics for telephone band (300-3400 Hz) digital telephones*.

[14] ITU-T Recommendation P.50 (1999), *Artificial voices*.

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

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

[17] ITU-T Recommendation P.831 (1998), *Subjective performance evaluation of network echo cancellers*.

[18] ITU-T Recommendation P.501 (2000), *Test signals for use in telephonometry*.

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

[20] CCITT Recommendation Q.271 (1988), *General*.

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

[22] CCITT Recommendation Q.724 (1988), *Telephone user part signalling procedures*.

[23] ITU-T Recommendation T.30 (1999), *Procedures for document facsimile transmission in the general switched telephone network.*

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

[25] ITU-T Recommendation V.8 (1999), *Procedures for starting sessions of data transmission over the public switched telephone network.*

[26] CCITT Recommendation V.21 (1988), *300 bits per second duplex modem standardised for use in the general switched telephone network.*

[27] CCITT Recommendation V.23 (1988), *600/1200-baud modem standardised for use on the general switched telephone network.*

[28] 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*.

[29] 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*.

[30] CCITT Recommendation V.27 *ter* (1988), *4800/2400 bits per second modem standardized for use in the general switched telephone network.*

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

[32] 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*.

[33] ITU-T Recommendation V.34 (1998), *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*.

[34] IEC 60651 (1979), *Sound level meters*.

## 3 Terms and definitions

This ITU-T Recommendation defines the following terms.

In the definition and text, $L$ will refer to the relative power level of a signal, expressed in dBm0 (as defined by ITU-T Recommendation G.711 [9]) 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.

### 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.

### 3.2 cancelled end:

*F: côté annulé*

*S: extremo compensado*

The side of an echo canceller which contains the echo path on which this echo canceller is intended to operate. This includes all transmission facilities and equipment (including the hybrid and terminating telephone set) which is included in the echo path. In previous versions of ITU-T Recommendation G.168 this was defined as the near end.

### 3.3 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 nonlinear processing loss (if present). This loss relates $L_{Rin}$ to $L_{RET}$ by:

$$L_{RET} = L_{Rin} - A_{COM}, \text{ where}$$

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

### 3.4 comfort noise:

*F: bruit de confort*

*S: ruido de confort, ruido nivelador*

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.

### 3.5 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.

### 3.6 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.

### 3.7 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.

### 3.8 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 the cancelled end echo present on the send path by subtracting an estimation of that echo from the cancelled end echo (see Figure 2).

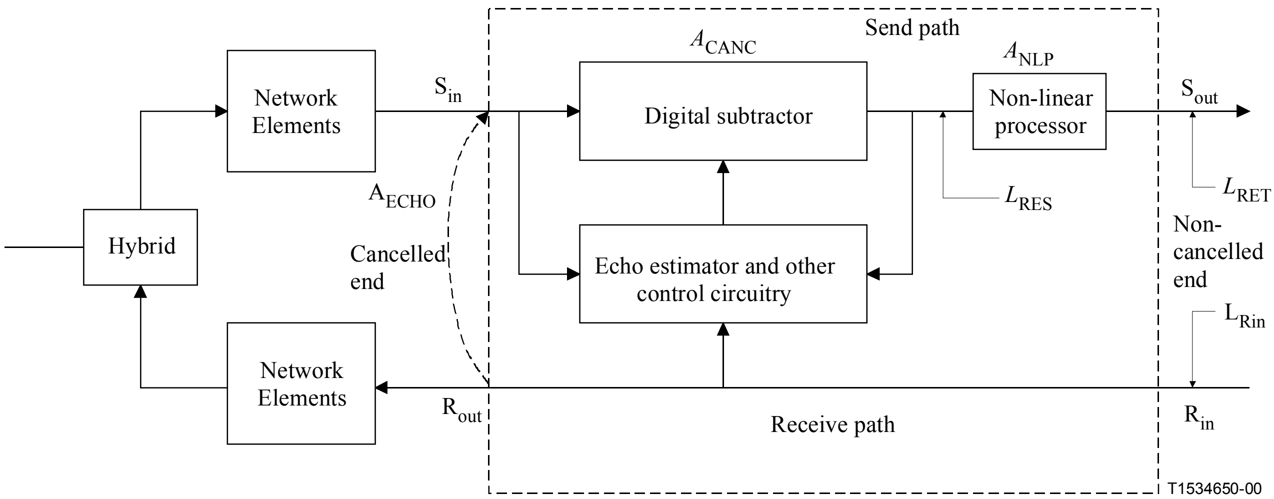



**Figure 2/G.168 – Location of levels and loss of an echo canceller**

### 3.9 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.

### 3.10 echo path capacity (Δ):

*F: capacité en trajet d'écho (Δ)*

*S: capacidad del trayecto del echo (Δ)*

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

### 3.11 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 (see Figure 3).

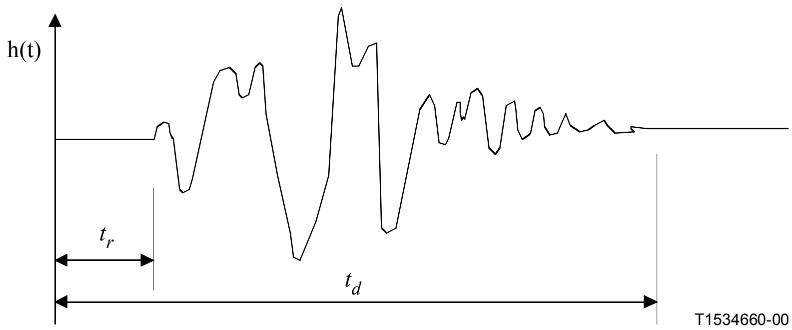



**Figure 3/G.168 – Example of an impulse response of an echo path**

### 3.12 echo return loss (ERL) ($A_{ECHO}$):

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

*S: atenuación del eco ($A_{ECHO}$)*

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 (cancelled-end) echo path.

NOTE – This definition does not strictly adhere to the echo loss definition given in 2.2/G.122 [2], 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.

### 3.13 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 nonlinear processing on the output of the canceller to provide for further attenuation.

### 3.14 electric echo:

*F: écho électrique*

*S: eco eléctrico*

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

### 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.

### 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 $R_{\text{in}}$ 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.

### 3.17 non-cancelled end:

*F: côté non annulé*

*S: extremo no compensado*

The side of an echo canceller which does not contain the echo path on which this echo canceller is intended to operate. In previous versions of ITU-T Recommendation G.168 this was defined as the far end.

### 3.18 nonlinear 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 centre clipper in which all signal levels below a defined threshold are forced to some minimum value.

### 3.19 nonlinear processing loss ($A_{NLP}$):

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

*S: atenuación por procesamiento 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 nonlinear 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.

### 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).

### 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 nonlinear processing is not included.

### 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

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

If nonlinear processing is not present, note that $L_{RES} = L_{RET}$.

### 3.23 open echo path:

*F: trajet d'écho ouvert*

*S: trayecto de eco abierto*

An echo path with infinite echo return loss.

## 4 Abbreviations

This ITU-T 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 |

## 5 Test signals

The tests in this ITU-T Recommendation use special signals such as noise, tones, Group 3 facsimile signals, and a subset of the composite source signals (CSS) 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 and ITU-T Recommendation P.501 [18]). 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. voice-band data, as well as under non-ideal network scenarios. Tests are included to test performance for Group 3 facsimile signals, residual acoustic echoes (optional), and non-linearities in the echo path such as may arise with low bit rate encoding in the echo path (optional).

## 6 Characteristics of echo cancellers

### 6.1 General

This ITU-T 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. A full echo canceller is possible consisting of two half echo cancellers, with each pointing in the opposite direction.

### 6.2 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) voice-band 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) voice-band 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 (see 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 ITU-T 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. Similarly, if an echo canceller conforming to this ITU-T 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.

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

NOTE 2 – ITU-T Recommendation Q.271 [20] on ITU-T No. 6 and ITU-T Recommendation Q.724 [22] 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.

### 6.3 External enabling/disabling

Certain digital echo cancellers may be disabled directly by a digital signal (e.g. see ITU-T Recommendation Q.55). These echo cancellers should provide 64 kbit/s bit sequence integrity (i.e. if integrated, the A-law to μ-law or μ-law to A-law conversion will also be disabled) in the externally disabled state.

### 6.4 Tests and requirements for performance with input signals applied to the send and receive paths

#### 6.4.1 Transmission performance

The appropriate transmission performance requirements of ITU-T Recommendations G.164 [4] and G.165 [5] also apply to echo cancellers except as noted below.

##### 6.4.1.1 Group delay

The group delay in the voiceband 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.

##### 6.4.1.2 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

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

$$S(k) = 3.17 + 20\log\left[\frac{\sqrt{\frac{2}{n}\sum_{i=k}^{k-n+1} e_i^2}}{8159}\right] \quad \text{(μ-law encoding)}$$

*S(k)* signal level in dBm0

$e_i$ linear equivalent of the PCM encoded signal at time i

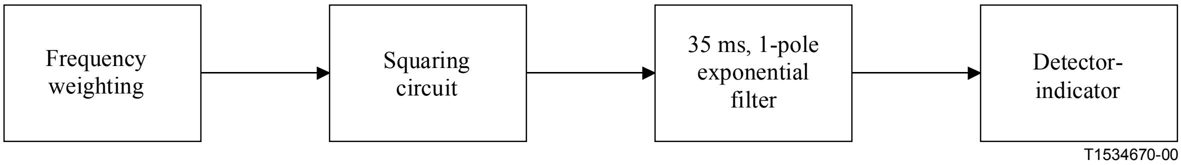
$k$ discrete time index

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

###### 6.4.1.2.1 Level Measurement Device

For some of the tests in this ITU-T Recommendation, e.g. Test No. 2, it is necessary to measure the short-term level of the signal. This is achieved using the following level measurement device.

The measurement device comprises a frequency-weighting network, a squaring circuit, an exponential filter, and a detector-indicator. This device may use either analog or digital methods. The impulse response of the frequency-weighting network is listed in the Table 1. The table is read in columns. The measurement device will have the characteristics that correspond to the following block diagram.




This frequency-weighting network is used to provide a greater attenuation of the frequencies outside the band of interest that can be achieved by the filters identified in IEC 60651 [34]. This filter is required because of D. C. effects due to the bias induced by A-law encoding. The filter is a 101-element finite impulse response bandpass filter with the impulse response shown in Figure 4 and the frequency response shown in Figure 5. The coefficients of the filter are:

**Table 1/G.168 – Coefficients of Bandpass Filter for Level Measurement Device**

| | | | |
|---|---|---|---|
| f0, f100 | 0.0000 | f26, f74 | 0.0044 |
| f1, f99 | 0.0006 | f27, f73 | 0.0095 |
| f2, f98 | 0.0005 | f28, f72 | 0.0017 |
| f3, f97 | 0.0004 | f29, f71 | 0.0188 |
| f4, f96 | 0.0011 | f30, f70 | 0.0000 |
| f5, f95 | –0.0000 | f31, f69 | 0.0225 |
| f6, f94 | 0.0015 | f32, f68 | 0.0024 |
| f7, f93 | –0.0003 | f33, f67 | 0.0163 |
| f8, f92 | 0.0012 | f34, f66 | 0.0092 |
| f9, f91 | –0.0002 | f35, f66 | 0.0000 |
| f10, f90 | 0.0000 | f36, f64 | 0.0164 |
| f11, f89 | 0.0002 | f37, f63 | –0.0210 |
| f12, f88 | –0.0020 | f38, f62 | 0.0161 |
| f13, f87 | 0.0005 | f39, f61 | –0.0375 |
| f14, f86 | –0.0040 | f40, f60 | 0.0000 |
| f15, f85 | 0.0000 | f41, f59 | –0.0406 |
| f16, f84 | –0.0047 | f42, f58 | –0.0357 |
| f17, f83 | –0.0019 | f43, f57 | –0.0267 |
| f18, f82 | –0.0033 | f44, f56 | –0.0871 |
| f19, f81 | –0.0047 | f45, f55 | –0.0000 |
| f20, f80 | –0.0000 | f46, f54 | –0.1420 |
| f21, f79 | –0.0068 | f47, f53 | 0.0289 |
| f22, f78 | 0.0036 | f48, f52 | –0.1843 |
| f23, f77 | –0.0057 | f49, f51 | 0.0475 |
| f24, f76 | 0.0054 | f50 | 0.8006 |
| f25, f75 | 0.0000 | | |

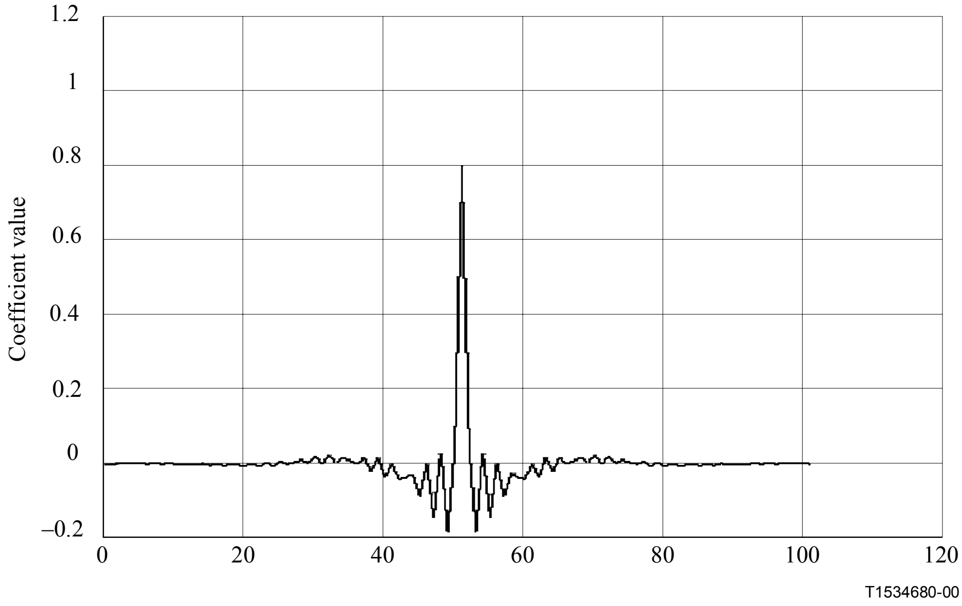



**Figure 4/G.168 – Impulse Response of Frequency Weighting Network**

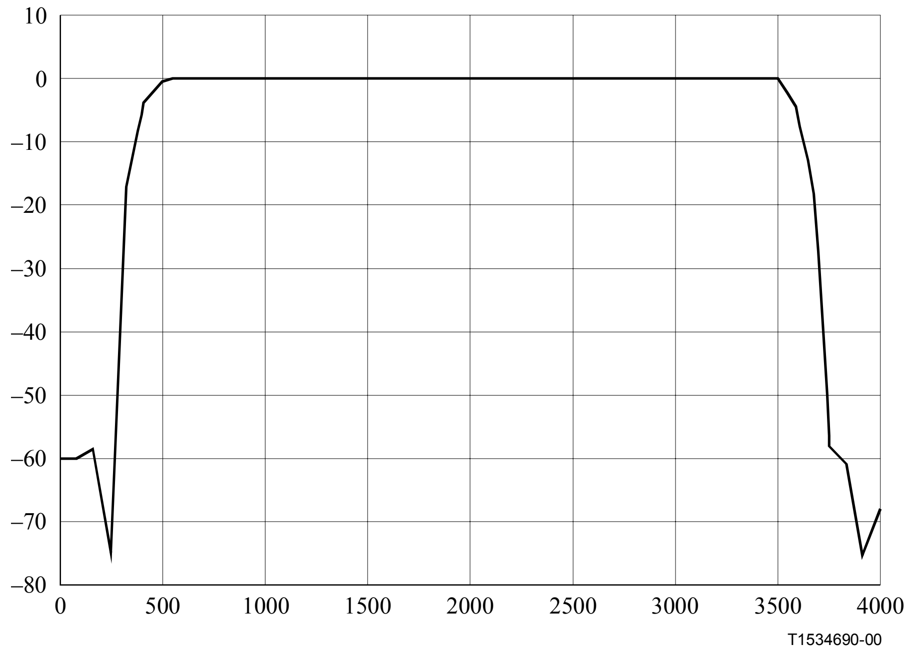



**Figure 5/G.168 – Frequency Response of Frequency Weighting Network**

#### 6.4.1.2.2 Level Measurement Device for Peaks

For tests that have requirements on the peaks at $S_{out}$, the measurement method used is a 35 ms. rectangular sliding window in place of the 35 ms one-pole exponential filter of 6.4.1.2.1.

### 6.4.2 Echo canceller performance

The performance requirements that follow are for echo cancellers that 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 reset (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 Figures 6 and 7. The ports are assumed to be at equal relative level points. For all values of $R_{in}$, and for all tests in this ITU-T 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 ITU-T Recommendation. The composite source signals, which consist of the receive-input test signal and send-input test signal (see Annex C and ITU-T Recommendation P.501 [18]) are used as the test signals, unless otherwise indicated. For multiple channel implementations, channel to channel independence is required, and any channels tested simultaneously should each meet the requirements of this ITU-T Recommendation. When performing the tests described in this ITU-T Recommendation, for enhanced repeatability, all channels not being tested should have idle code (e.g. 01111111 for 1544 kbit/s systems or 01010100 for 2048 kbit/s systems as outlined in ITU-T Recommendation Q.522) applied to the inputs. It is generally recognized that some operators may wish to apply simulated traffic loading to the untested channels. This type of channel loading is for further study.

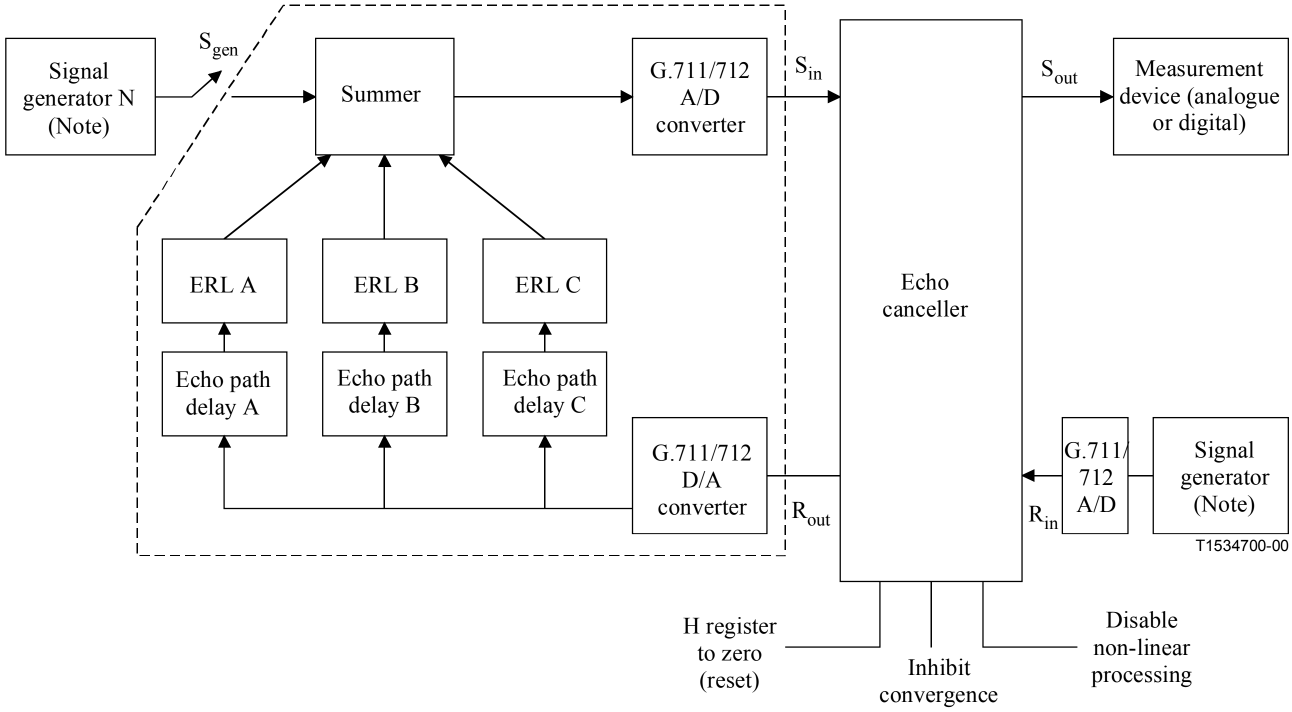
The ERL used in these tests have a minimum value of 6 dB. It should be noted that 6 dB is a typical worst case value encountered for most networks, and most current networks have typical ERL values better than this.

Also, it should be noted that the test configurations specified in this ITU-T Recommendation are artificial for purposes of test and result repeatability, and do not fully represent all conditions that would be expected in real networks.

The requirements in 6.4.2 are based on the use of the composite source signals, noise, tones, FAX signals, and voice-band data signals as the test signals.

Two echo path models can be used for the tests in this ITU-T Recommendation (as denoted in Figures 6 and 7).

Comfort noise should be disabled unless otherwise indicated.




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 ≤ Δ ms, echo path delay B ≤ Δ ms, and echo path delay C ≤ Δ ms.

**Figure 6/G.168 – Functional diagram for echo canceller performance measurements**

An echo path model which allows more realistic end-paths to be modelled can be realized by replacing the dotted box in Figure 6 with Figure 7. The characteristics of the end-path, which include the D/A and A/D converters, can be modelled as an impulse response g(k).

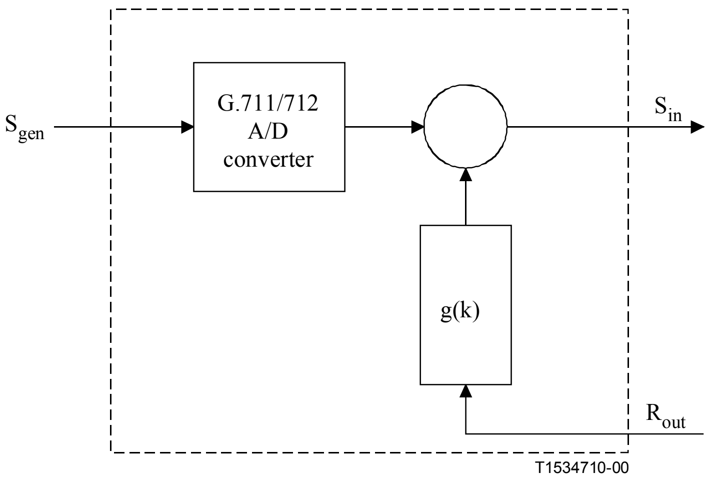



**Figure 7/G.168 – Echo Path Model using g(k)**

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 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 ITU-T 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 $\Delta$, whenever it appears in this ITU-T Recommendation, represents the maximum echo path delay for which the echo canceller is designed.

See I.9/Appendix I for some guidelines on the use of parameters for testing echo cancellers.

#### 6.4.2.1 Echo-Path Models for g(k)

See Annex D for the echo path models that may be used as the g(k) in Figure 7 for the tests in this ITU-T Recommendation. This does not represent an exhaustive set, and other models may be used provided that they meet the echo path requirements for each individual test. Note that the digital version of Figure 6 where three echo path reflection points are present may also be represented by a g(k). A specific model including this is not described in Annex D. A more complex model that includes realistic dispersion and other effects is for further study. See also Appendix III for additional information.

#### 6.4.2.2 Test No. 1 – Steady state residual and returned echo level test (deleted)

This test has been incorporated into test 2.

#### 6.4.2.3 Test No. 2 – Convergence and Steady state residual and returned echo level tests

This test is meant to ensure that the echo canceller converges rapidly for all combinations of input signal levels, echo paths, and certain echo path changes, and that the returned echo level is sufficiently low. This test is also meant to ensure that the steady state cancellation ($A_{\mathrm{CANC}}$) is sufficient to produce a residual echo level which is sufficiently low to permit the use of nonlinear processing without undue reliance on it. In general, given that all other variables are equal, a higher value of ERLE or lower values of $L_{\mathrm{RES}}$ will allow for less dependence on the NLP functionality.

The H register is initially set to zero, or converged to an open echo path and adaptation is inhibited. Adaptation is then enabled at least 200 ms before the start of a CSS burst (see Figure 8). This period is to allow for the latency time in the adaptation control of the canceller. The residual or returned echo level is then measured as a function of time to reveal the convergence and steady-state properties 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 10 and 12 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 voice-band signals.

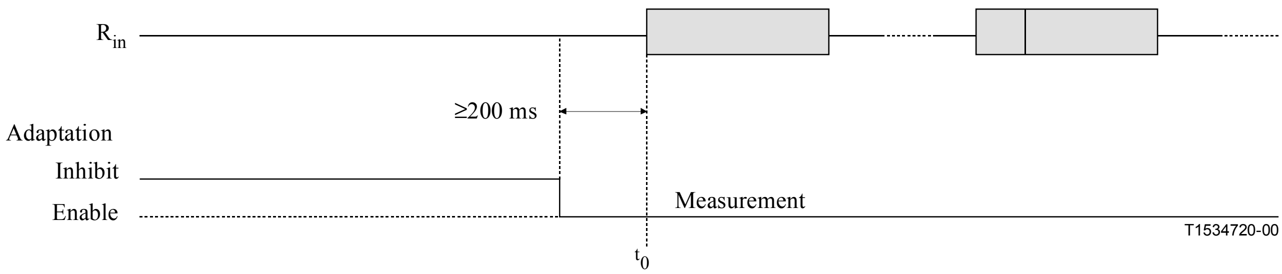



**Figure 8/G.168 – Test No. 2A, 2B signal and time relationships**

#### 6.4.2.3.1 Test 2A – Convergence test with NLP enabled

**Requirement**

With the H register initially set to zero, or alternatively, with the echo canceller initially fully converged to an open echo path, and the NLP enabled, for all values $L_{Rin} \geq -30$ dBm0 and $\leq 0$ dBm0 and for all values of ERL ≥ 6 dB and echo path delay, $t_d \leq \Delta$ ms, the combined loss ($L_{Rin} - L_{RET}$) should be greater than or equal to that shown in Figure 10. After $1+t_d$ s, the combined loss should be greater than or equal to that in Figure 9. The level at $S_{out}$ is measured using a meter conforming to the characteristics of 6.4.1.2.1. In addition, no peaks (see 6.4.1.2.2) are allowed that exceed 5 dB above the requirements in Figure 9. The level at $R_{in}$ is measured using the RMS method of 6.4.1.2, 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 method of 6.4.1.2.1 may also be used at $R_{in}$, but the input and output signals must also be synchronized.

With the H register in any initially converged state other than those in the paragraph above, the requirements of Figure 10 apply after time $1+t_d$ seconds.

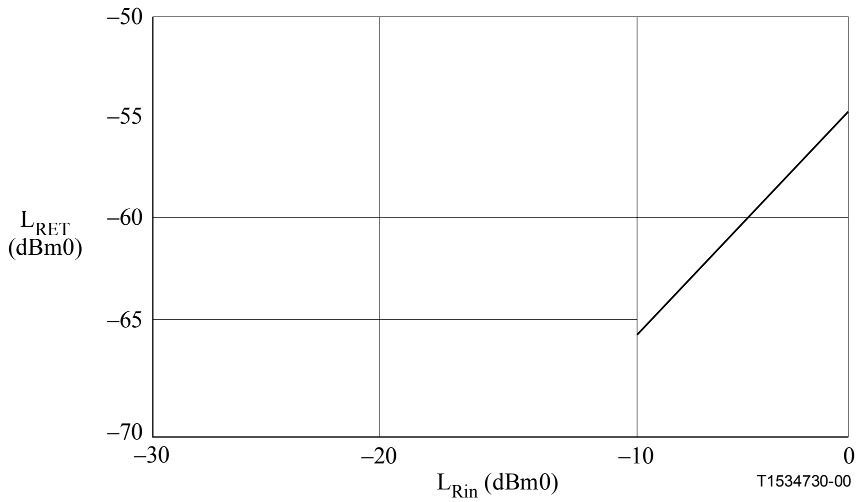
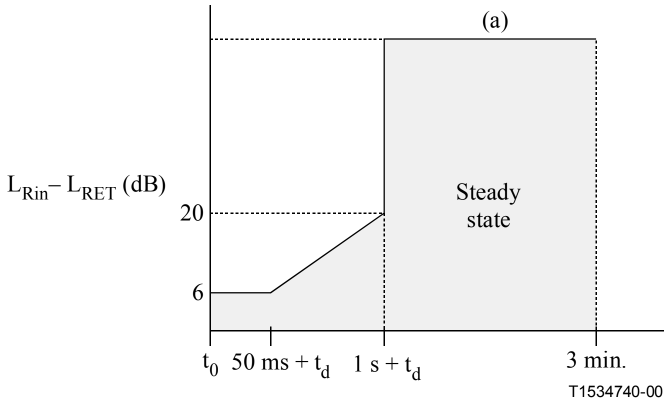



**Figure 9/G.168 – Relationship between receive input level ($L_{Rin}$) and return echo level ($L_{RET}$) with NLP enabled**

The requirements in Figure 9 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. This does not, however, imply that the requirements of Figures 9 and 10 need not be met at $R_{in}$ levels of –5 dBm0 or above. 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 the Figure.




(a) As derived from Figure 9.

**Figure 10/G.168 – Convergence characteristics with NLP enabled**

### 6.4.2.3.2 Test 2B – Convergence test with NLP disabled

**Requirement**

With the H register initially set to zero, or alternatively, with the echo canceller initially fully converged to an open echo path, and the NLP disabled, for all values $L_{Rin} \geq -30$ dBm0 and $\leq 0$ dBm0 and for all values of ERL $\geq$ 6 dB and echo path delay, $t_d \leq \Delta$ ms, the loss $L_{Rin} - L_{RES}$ should be greater than or equal to that shown in Figure 12. After 10 s, the loss $L_{Rin} - L_{RES}$ should be greater than or equal to that in Figure 11. The level at $S_{out}$ is measured using a meter conforming to the characteristics of 6.4.1.2.1. In addition, no peaks (see 6.4.1.2.2) are allowed that exceed 5 dB above the requirements in Figure 11. The level at $R_{in}$ is measured using the RMS method of 6.4.1.2, 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 method of 6.4.1.2.1 may also be used at $R_{in}$, but the input and output signals must also be synchronized. (Note: Some echo cancellers employ a supplementary NLP function which cannot be disabled. For information covering this case, see 8.2.6, Testing of NLP's.)

With the H register in any initially converged state other than those in the paragraph above, the requirements of Figure 12 apply after time $1+t_d$ seconds.

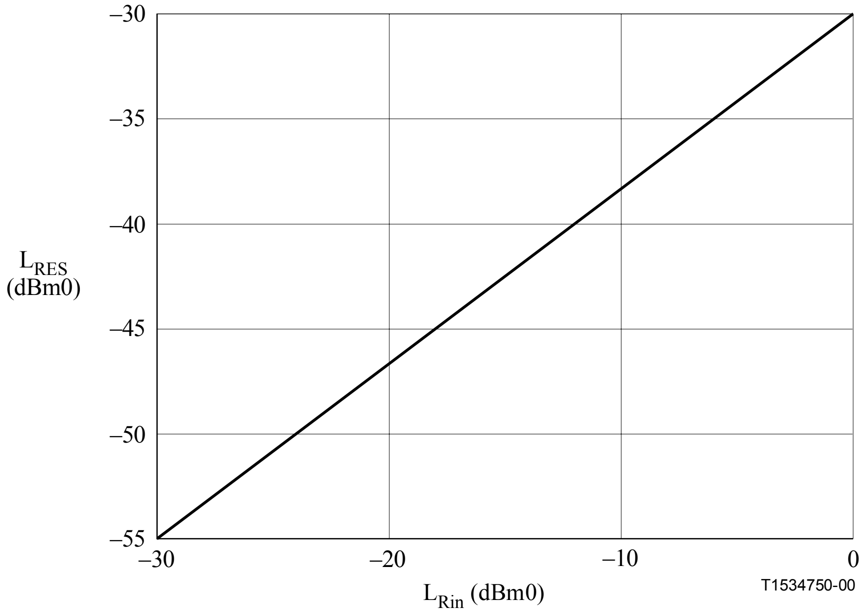



**Figure 11/G.168 – Relationship between receive input level ($L_{Rin}$) and residual echo level ($L_{RES}$) with NLP disabled**

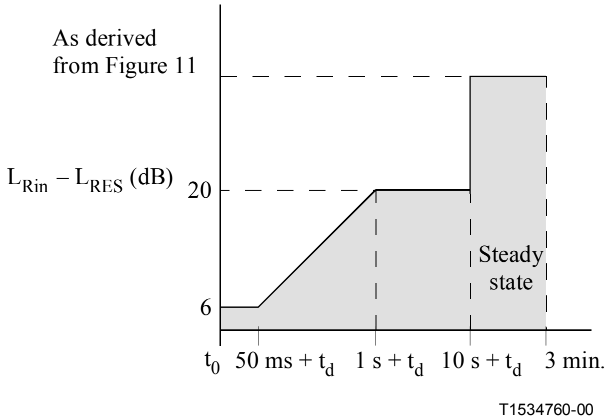



**Figure 12/G.168 – Convergence characteristics with NLP disabled**

### 6.4.2.3.3 Test 2C – Convergence test in the presence of background noise

Test No. 2C is meant to ensure that the steady state cancellation is sufficient to produce an echo level that is sufficiently low and 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 ITU-T Recommendation P.800 [16]) with level N is applied at $S_{gen}$. Adaptation is enabled at least 200 ms before the start of a CSS burst (see Figure 13). After the convergence time, inhibit adaptation, remove $S_{gen}$ and measure the residual echo level.

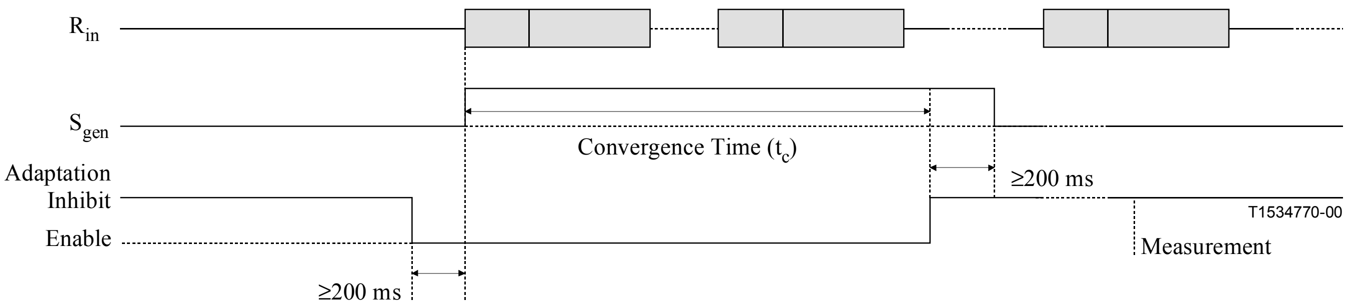



**Figure 13/G.168 – Test No. 2C signal and time relationships**

a) *Convergence Test with NLP Enabled*

**Requirement**

With the H register initially set to zero and the NLP enabled, for all values of $L_{\mathrm{Rin}} \geq -30$ dBm0 and $\leq 0$ dBm0, $N = L_{\mathrm{Rin}} - 15$ dB but no higher than $-30$ dBm0, ERL $\geq 6$ dB and echo path delay, $t_d \leq \Delta$ ms, convergence should occur within 1.0 s ($t_c$) and $L_{\mathrm{RET}}$ should be $\leq N$ (see Figure 14).

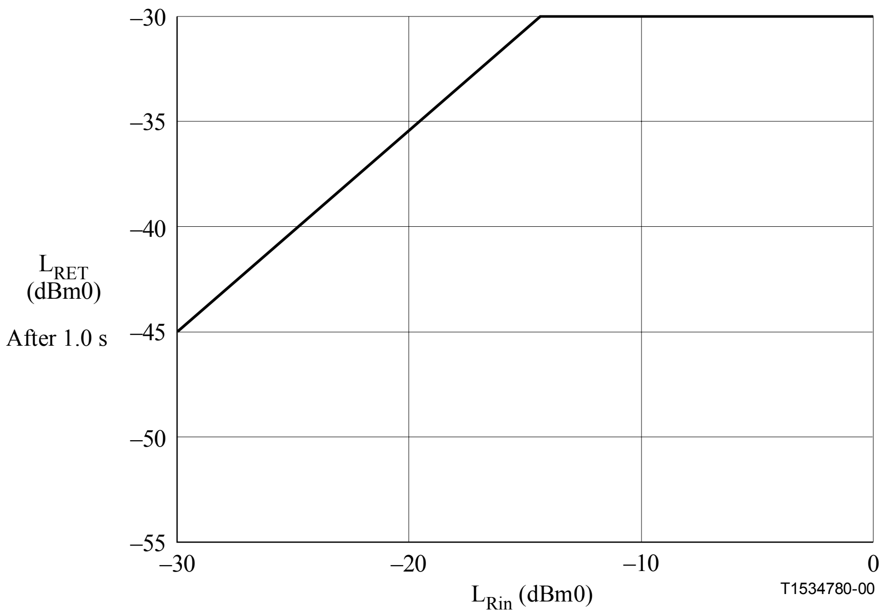



**Figure 14/G.168 – Test No. 2C Requirements NLP Enabled**

b) *Steady State Cancellation Test with NLP Disabled*

**Requirement**

With the H register initially set to zero and the NLP disabled, for all values of $L_{\mathrm{Rin}} \geq -30$ dBm0 and $\leq 0$ dBm0, N as given in Figure 15, ERL $\geq 6$ dB, echo path delay $t_d \leq \Delta$ ms, and convergence time $\geq 10$ s ($t_c$), $L_{\mathrm{RES}}$ should be less than that shown in Figure 15 for the corresponding value of N.

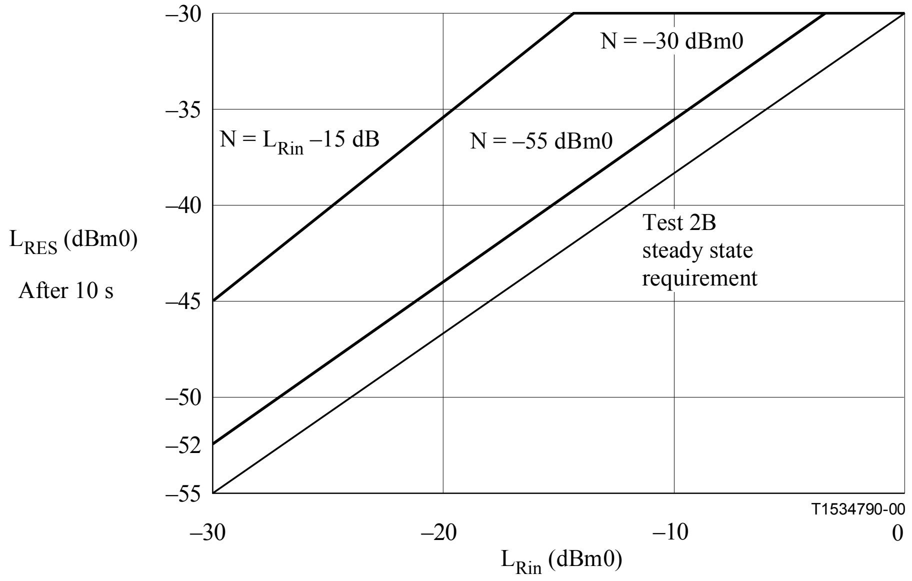



**Figure 15/G.168 – Test No. 2C Steady State Requirements NLP Disabled**

c) *Convergence Test with NLP Disabled (for further study)*

**Requirement**

With the H register initially set to zero and the NLP disabled, for all values of $L_{\mathrm{Rin}} \geq -30$ dBm0 and $\leq 0$ dBm0, $N = L_{\mathrm{Rin}} -15$ dB but no higher than –30 dBm0, ERL $\geq$ 6 dB and echo path delay $t_d \leq \Delta$ ms, the loss $L_{\mathrm{Rin}} - L_{\mathrm{RES}}$ should be greater than or equal to that shown in Figure 16. The value X is TBD.

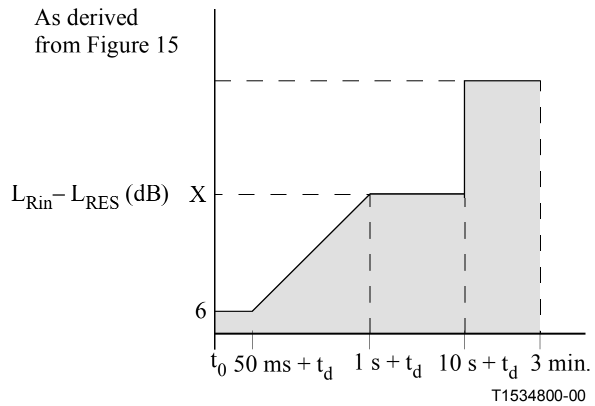



**Figure 16/G.168 – Test No. 2C Convergence Requirements NLP Disabled**

The level at $S_{\mathrm{out}}$ is measured using a meter conforming to the characteristics of 6.4.1.2.1. The level at $R_{\mathrm{in}}$ is measured using the RMS method of 6.4.1.2, 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 method of 6.4.1.2.1 may also be used at $R_{\mathrm{in}}$, but the input and output signals must also be synchronized.

### 6.4.2.4 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 artifacts 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.

For this test, the $R_{in}$ signal is CSS and the $S_{gen}$ signal is the double talk CSS. While CSS is used 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.

See I.8.4/Appendix I for guidelines on other double talk test methods for tests 3A and 3B.

#### 6.4.2.4.1 Test 3A – Double talk test with low cancelled-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 17) an interfering signal (double talk CSS) 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*.

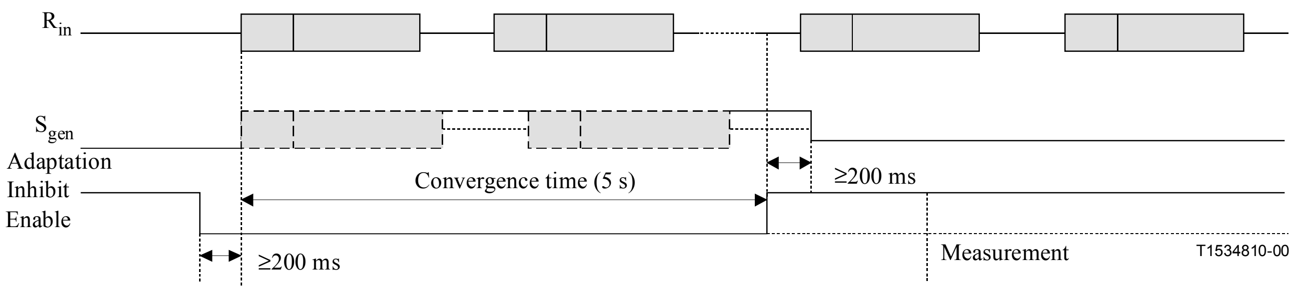



**Figure 17/G.168 – Test No. 3A signal and time relationships**

**Requirement**

With the H register initially set to zero for all values of $L_{Rin} \geq -25$ dBm0 and $\leq 0$ dBm0, $N = L_{Rin} - 15$ dB, ERL $\geq 6$ dB and echo path delay, $t_d \leq \Delta$ ms, convergence should occur within 5 s and $L_{RES}$ should be $\leq N$.

#### 6.4.2.4.2 Test 3B – Double talk test with high cancelled-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 18) 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.*

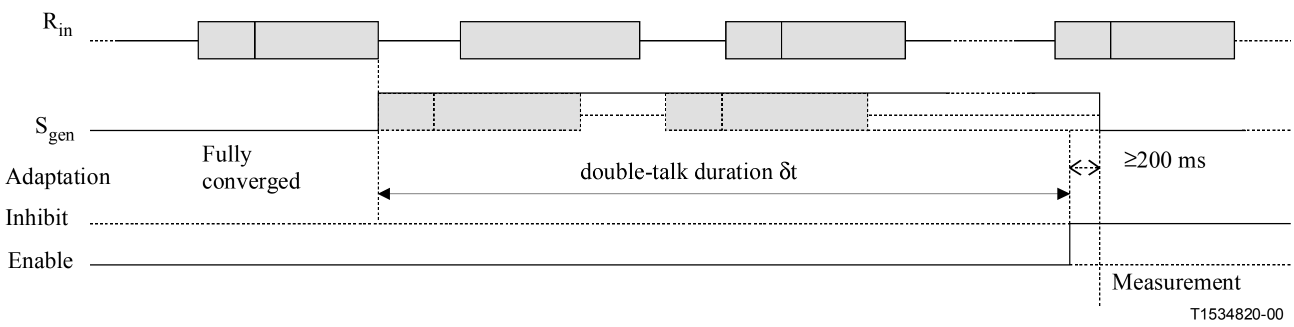



**Figure 18/G.168 – Test No. 3B signal and time relationships**

COM 15-27 (1993) shows a wide variation of performance for different languages for this test. Note that test 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 $L_{\mathrm{Rin}} \geq -30$ dBm0 and $\leq 0$ dBm0, and for all values of $N \geq L_{\mathrm{Rin}}$ 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 $L_{\mathrm{Rin}}$ and $N$ for any time period should not increase more than 10 dB over the steady state requirements of Figure 11.

### 6.4.2.4.3 Test 3C – Double talk test under simulated conversation

Test No. 3C is meant to ensure that the echo canceller does not produce undesirable artifacts during and after periods of double talk (see I.8).

The test procedure is to clear the H-register. Then for some value of echo path delay, $t_d \leq \Delta$ ms, and ERL ≥ 6 dB, a signal is applied to $R_{in}$. Simultaneously (see Figure 19), a signal $N$ is applied to $S_{gen}$ which has a level at least that of $R_{in}$. After a time $t_1$, $N$ is removed and $S_{out}$ is measured. During time $t_4$ and $t_5$, $N$ is reapplied and the output is evaluated for artifacts. The NLP should be enabled for this test.

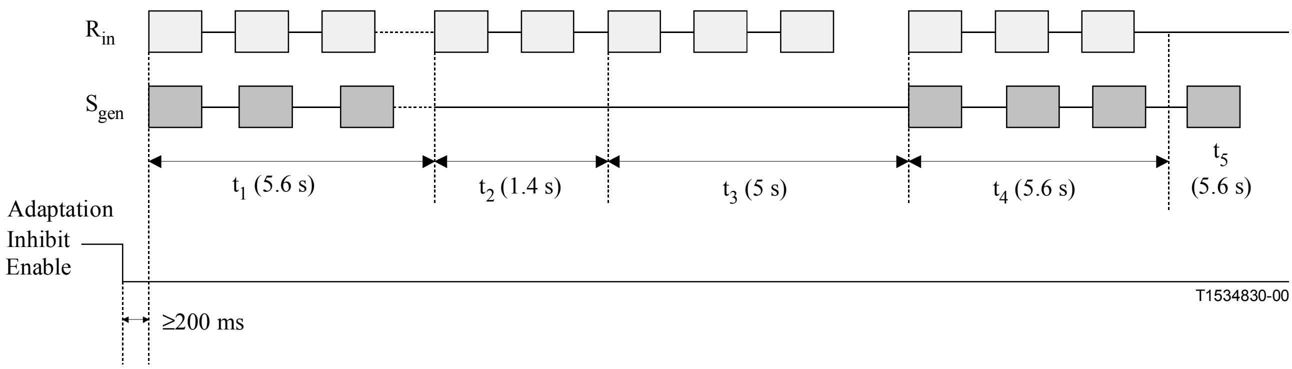



**Figure 19/G.168 – Test No. 3C signal and time relationships**

## Requirement

With the H-register initially set to zero, for all values of $L_{Rin} \geq -25$ dBm0 and $\leq 0$ dBm0, and for all values of N $\geq L_{Rin}$ and for all values of ERL $\geq 6$ dB and echo path delay $t_d \leq \Delta$ ms, any peaks (see 6.4.1.2.2) during period $t_2$ should not exceed the level of N during period $t_1$. The residual echo level during time period $t_3$ should meet the requirements of Figure 9 with NLP enabled. During $t_4$ and $t_5$, no peaks should exceed the level of N + 6 dB.

Level offsets between $L_{Rin}$ and $L_{S_{gen}}$ 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.

### 6.4.2.5 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 20). The NLP should be *disabled*.

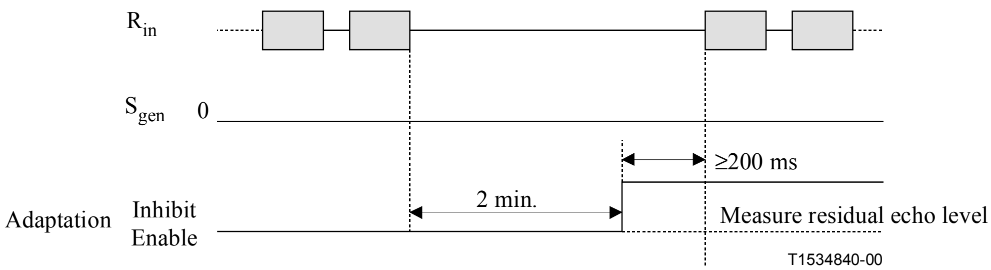



**Figure 20/G.168 – Test No. 4 signal and time relationships**

## Requirement

With the echo canceller initially in the fully converged state for all values of $L_{Rin} \geq -30$ dBm0 and $\leq 0$ 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. 2 (see Figure 11).

### 6.4.2.6 Test No. 5 – Infinite return loss convergence test

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 $R_{in}$, and the output at $S_{out}$ measured (see Figure 21). The NLP should be disabled.

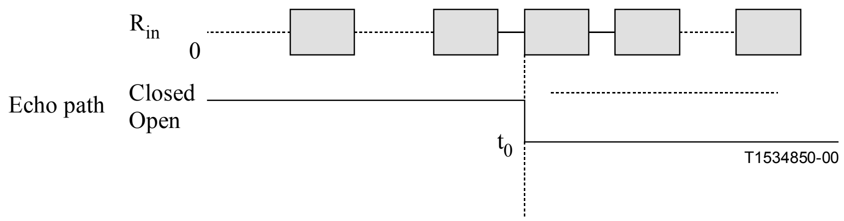



**Figure 21/G.168 – Test No. 5 signal and time relationships**

**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, and at time $t_0$ the echo path is interrupted with an open echo path, the combined loss $L_{Rin} - L_{RES}$ should meet the requirements of Figure 12, as measured using the method of 6.4.1.2.1.

### 6.4.2.7 Test No. 6 – Non-divergence on narrow-band signals

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 wide-band 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 2. The sequence of tones of Table 2 are then applied at $R_{in}$. After the sequence is completed, the adaptation is inhibited and the residual echo is measured with the signal of test 2. The NLP should be disabled.

**Requirement**

The echo canceller is fully converged as in test 2 for all values of ERL ≥ 6 dB, and echo path delay $t_d \leq \Delta$ ms. Once chosen, the echo path remains the same throughout the test. The sequence of tones of Table 2 (with the higher frequency 2 dB higher than the lower frequency for the dual tone combinations) is then applied for 5 s each to $R_{in}$ such that $L_{Rin}$ ≥ −30 dBm0 and ≤ 0 dBm0. After the application of the tone sequence, the adaptation is inhibited. Using the same signal as in Test No. 2 the measured residual echo should not degrade more than 10 dB from the requirements of Figure 11 of Test No. 2. The NLP is disabled for this test.

**Table 2/G.168**

| |
|---|
| 697 |
| 941 |
| 1336 |
| 1633 |
| 697 & 1209 |
| 770 & 1336 |
| 852 & 1477 |
| 941 & 1633 |

#### 6.4.2.8 Test No. 7 – Stability test

The object of this test is to verify that the echo canceller will remain stable for narrowband signals.

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 3 of test No. 8, for all values of $L_{Rin} \geq -30$ dBm0 and $\leq +3$ dBm0, and for all values of ERL $\geq$ 6 dB, with an echo path delay $t_d \leq \Delta$, the residual echo levels measured after two minutes should be less than or equal to that shown in Figure 22.

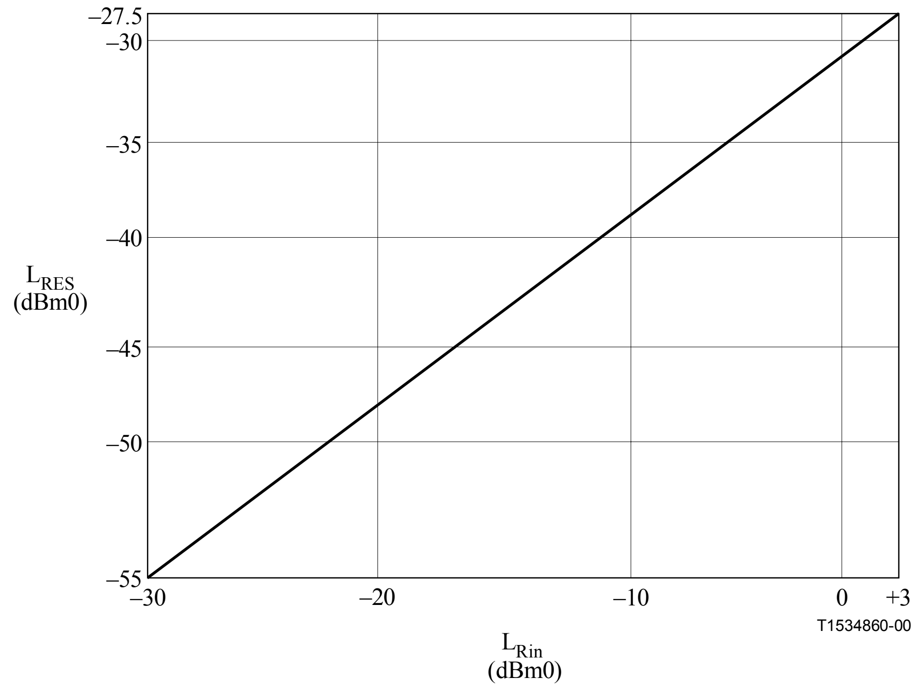



**Figure 22/G.168 – Performance requirements for test 7**

#### 6.4.2.9 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.

For an echo canceller equipped with this optional capability, the echo canceller is initially converged using CSS to any echo path with an ERL $\geq$ 6 dB, and an echo path delay $t_d \leq \Delta$. For simplification, the fully converged state for an ERL of 6 dB may be chosen. First the CSS is removed and then the echo path is opened. Then, any signal from Table 3 is applied at $S_{in}$. Within 90 ms (either before or after the application of the signal at $S_{in}$), the same signal is applied at $R_{in}$. After the detection time, the level at $S_{out}$ is measured. The peak level of each frequency signal applied is equivalent to the peak level of a sinusoid with a rms level, M, of $-18 \leq M \leq +3$ dBm0.

The echo canceller response should be sufficiently long to provide immunity from false operation due to voice signals. Also, the echo canceller response should not be so long as to needlessly extend the time for the signal to appear on the network.

**Requirement**

The level at $S_{out}$, as measured using the method of 6.4.1.2.1, should not vary more than 2 dB with respect to the level at $S_{in}$. The echo canceller should respond to the signals (detection time) within 1 s after application.

**Table 3/G.168 – Applicable Signalling Tones**

| System 5 | System 6 | System 7 |
|---|---|---|
| 2400 ± 6 Hz | 2000 ± 20Hz | 2000 ± 20 Hz |
| 2600 ± 6 Hz | | 1780 ± 20 Hz |
| 2400 ± 6 Hz & 2600 ± 6 Hz | | |

### 6.4.2.10 Test No. 9 – Comfort noise test

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 all input signals for this test. The NLP and comfort noise feature should be enabled.

#### 6.4.2.10.1 Part 1 (matching)

1) Set $N$ to a level between −50 dBm0 and −40 dBm0.
2) Set $L_{\mathrm{Rin}}$ to silence (< – 60 dBm0) and hold for 30 s.
3) Set $L_{\mathrm{Rin}}$ to −10 dBm0.
4) Measure $L_{\mathrm{RET}}$ after 2 s.

**Requirement**

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

#### 6.4.2.10.2 Part 2 (adjustment down)

1) Lower $N$ by 10 dB from the level in Part 1.
2) Set $L_{\mathrm{Rin}}$ to silence (< – 60 dBm0) and hold for 10 s.
3) Set $L_{\mathrm{Rin}}$ to −10 dBm0.
4) Measure $L_{\mathrm{RET}}$ after 2 s.

**Requirement**

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

#### 6.4.2.10.3 Part 3 (adjustment up)

1) Raise $N$ by 10 dB from the level in Part 2.

2) Set $L_{Rin}$ to silence (< – 60 dBm0) and hold for 10 s.

3) Set $L_{Rin}$ to −10 dBm0.

4) Measure $L_{RET}$ after 2 s.

**Requirement**

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

#### 6.4.2.11 Test No. 10 – Facsimile test during call establishment phase

This test is meant to ensure that the echo cancellers located at each end of a connection converge rapidly on the initial handshaking sequences of a facsimile call and have some means to prevent the unwanted generation of echo. The test is split into three parts. Test 10A looks at the performance of the echo canceller located on the calling station side, and Test 10B looks at the performance of the echo canceller on the called station side. Test 10C looks at the performance of the echo canceller on the calling station side during page transmission.

The test has been designed to run in a laboratory environment using an echo canceller and a fax simulator. The tests should be run separately.

The tests 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 s (± 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 Setup (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 ITU-T Recommendation V.21).

Conditions:

| | |
|---|---|
| Data signalling rate, synchronous | 300 bit/s |
| Centre 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".

**6.4.2.11.1 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 while CNG, CED and sequence No. 1 are applied (see Figure 23). During the adaptation time, the residual/returned echo level is measured. This test should be performed with the NLP both enabled and disabled.

**Requirement**

With the H register initially set to zero and the value $L_{\mathrm{Rin}} = -13$ 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. The test should run for at least 7 s while CED and sequence No. 1 are applied. Repeat sequence 1 as necessary.

*Region I (converging on CED tone)*

– the peaks (see 6.4.1.2.2) of $L_{\mathrm{RES}}$ should be ≤ ($-13 - A_{\mathrm{ECHO}}$) dBm0,

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

*Region II (converged on CED tone)*

– the peaks (see 6.4.1.2.2) of $L_{\mathrm{RES}}$ should be ≤ −37 dBm0.

*Region III (converging on sequence No. 1)*

– the peaks (see 6.4.1.2.2) of $L_{\mathrm{RES}}$ should be ≤ ($-13 - A_{\mathrm{ECHO}}$) dBm0,

– the time to enter region IV should be ≤ 1.1 s.

*Region IV (converged on sequence No. 1)*

– the peaks (see 6.4.1.2.2) of $L_{\mathrm{RES}}$ should be ≤ −24 dBm0.

If the NLP is provisioned on, the peaks (see 6.4.1.2.2) of $L_{\mathrm{RET}}$ should be ≤ −37 dBm0 in the regions II and IV.

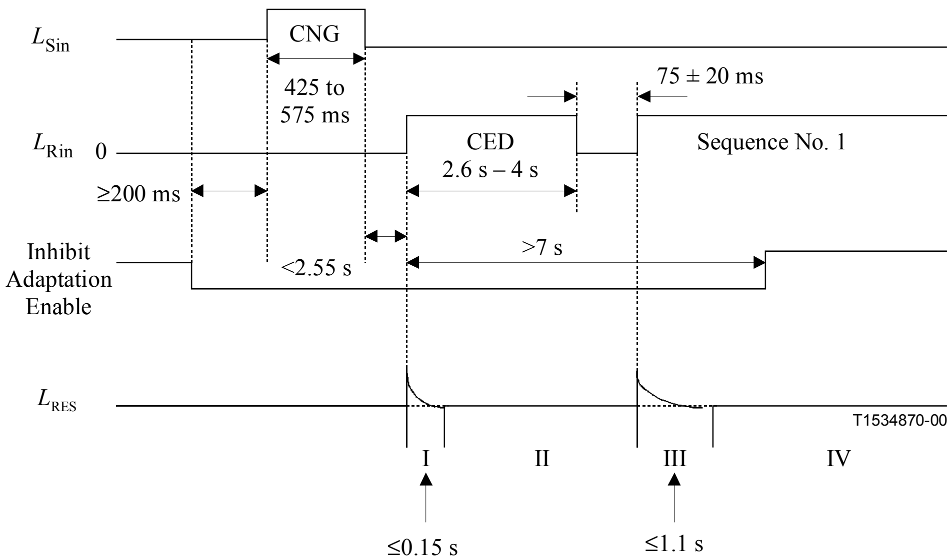



**Figure 23/G.168 – Test No. 10A signal and time relationships**

#### 6.4.2.11.2 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 24). During the adaptation time, the residual/returned echo level is measured. This test should be performed with the NLP both enabled and disabled.

**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. The test should run for 10 s as a minimum. Repeat sequence 2 as necessary.

*Region I (converging on sequence No. 2)*

– the peaks (see 6.4.1.2.2) of $L_{RES}$ should be ≤ (−13 – $A_{ECHO}$) dBm0,

– the time to enter region II should be ≤ 1.1 s.

*Region II (converged on sequence No. 2)*

– the peaks (see 6.4.1.2.2) of $L_{RES}$ should be ≤ −24 dBm0.

If the NLP is provisioned on, the peaks (see 6.4.1.2.2) of $L_{RET}$ should be ≤ −37 dBm0 in the region II.

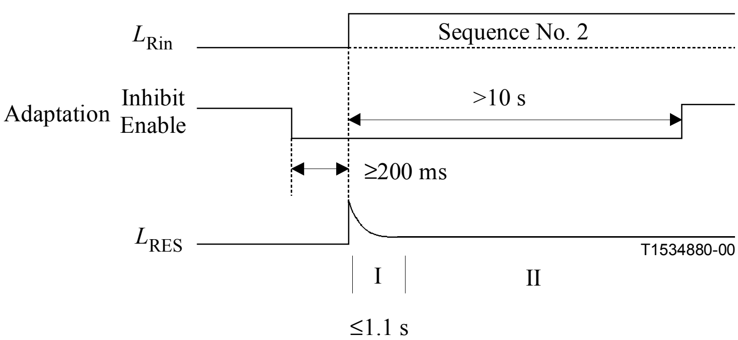



**Figure 24/G.168 – Test No. 10B signal and time relationships**

#### 6.4.2.11.3 Test No. 10C – Canceller operation on the calling station side during page transmission and page breaks (for further study)

Figure 25 shows the sequence of message exchange for a typical fax transmission consisting of two pages. The sequence begins with a V.21 [26] message handshake procedure. Operation and performance of the echo cancellers at each end of the link are tested during this period by tests 10A and 10B. This test is designed to check the operation and performance of the echo canceller at the calling station side during page transmission and page breaks as shown in Figure 25.

Test 10C uses data files A and B and software implementations of FSK and PSK modems. The test described uses V.29 modulation for the transmission of image data, but can be further extended to cover different types of modems such as V.17 and V.27 *ter* [30] as well as different timing relationships between handshake signals (represented by the V.21 [26] modem) and page transmission (represented by the V.29 [31] modem). The timing relationship shown in Figure 26 has been designed to mimic the real facsimile T.30 [23] protocol.

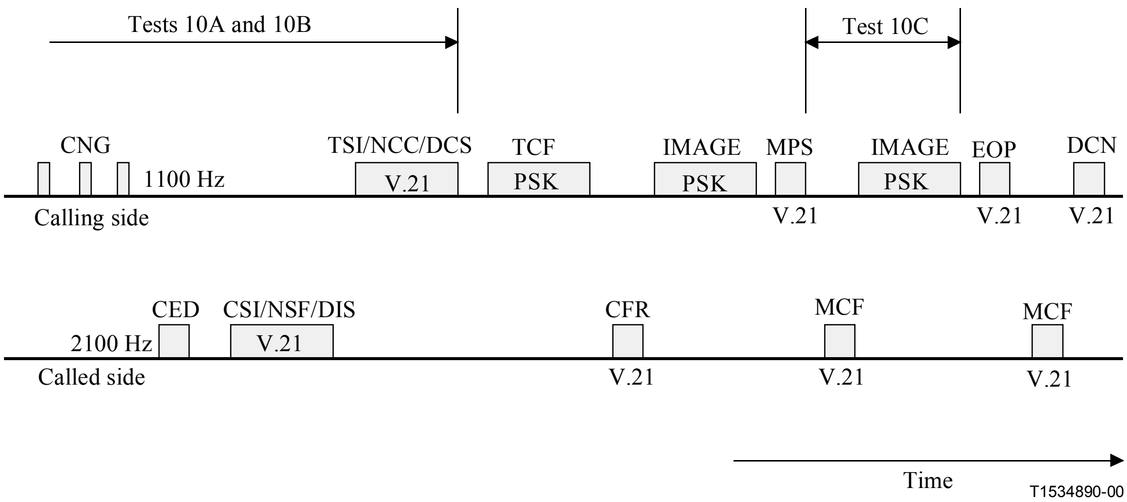



**Figure 25/G.168 – Sequence of message exchange for a typical 2-page fax transmission**

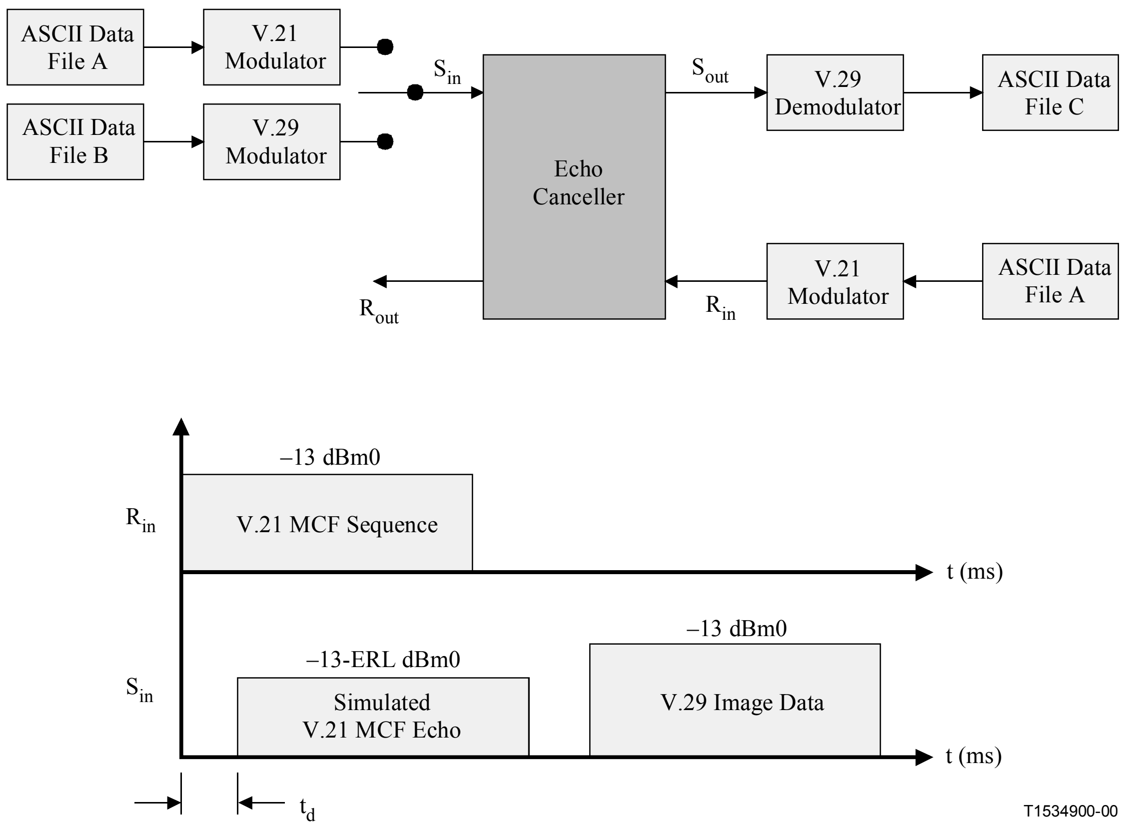



**Figure 26/G.168 – Canceller operation during page transmission and page breaks**

The test procedure is to clear the H register and inhibit adaptation. Adaptation is then enabled and the canceller is converged by sending data file A via the V.21 [26] modulator, into $R_{in}$ of the echo canceller (see Figure 26). This is meant to simulate the MCF signal transmitted by the called fax machine at the end of the first page. To simulate the echo signal, data file A is also transmitted (using the V.21 modulator) to the echo canceller's $S_{in}$ input, at a level of ≥6 dB below $R_{in}$ and with a timing offset of $t_d$. Following this, data file B is transmitted (using the V.29 [31] modulator) to the echo canceller's $S_{in}$ input.

The V.29 data from the echo canceller's $S_{out}$ output is first collected, stored and demodulated off-line. The recovered file C is then analysed. The data from the echo canceller's $R_{out}$ output is discarded (on the assumption that no impairments are introduced).

Note that in the latest version of T.30 [23], an Echo Protect Tone (EPT) is specified before all V.29 transmissions (TCF and image data). This EPT signal was originally intended to 'turn around' echo suppressors and prevent any front-end clipping of TCF or image data. In the case of echo cancellers, it helps to open the NLP before the transmission of TCF or image data to avoid corruption. However, many of the fax implementations currently in the field do not incorporate this EPT signal, so it has not been included in this test. In this respect, this test therefore represents a worst-case scenario.

The NLP should be provisioned enabled during this test. Note that some echo cancellers will automatically disable the NLP on detection of a fax call.

**Data File A**

In the following table, all data is given in hex. The initial flag is repeated 37 times.

| Flag | HDLC address field | HDLC control field | Control field MCF | Frame check sequence | Flag |
|---|---|---|---|---|---|
| 7E | FF | C8 | B1 | D4, 07 | 7E |

**Data File B**

Data File B should be taken from the following ITU-T CD ROM set:

ITU-T Recommendation T.24 (1998), *Standardized digitized image set*, Document No. 1, 'slerexe' letter.

**Requirement**

The intent is to compare the results with the echo canceller first disabled and then enabled. The actual requirements are for further study. The requirements should be based on a comparison of files B and C.

### 6.4.2.12 Test No. 11 – Tandem echo canceller test (for further study)

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

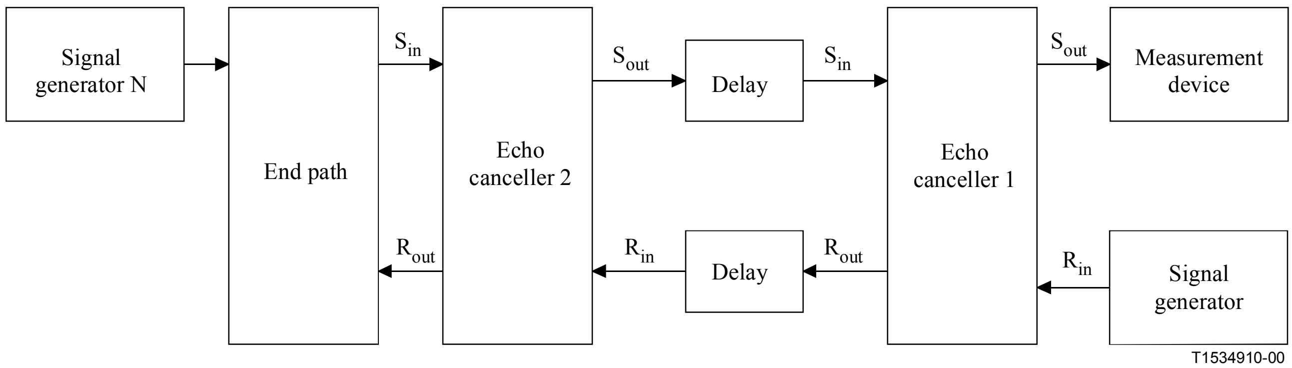



**Figure 27/G.168 – Test No. 11 configuration**

NOTE – For tests 12 (see 6.4.2.13) and 13 (see 6.4.3.1).

### 6.4.2.13 Test No. 12 – Residual acoustic echo test (for further study)

Under study. See I.6.6, I.6.6.1 and I.6.6.2/Appendix I for further discussion on this issue. This test is meant to check the performance of the echo canceller in the presence of residual acoustic echo and to ensure that its performance with electric echo is not overly degraded in the presence of acoustic echo.

### 6.4.2.14 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 28. 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 setup, 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 [21].)

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 ITU-T Recommendation G.122 [2], is obtained.

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

**Requirement**

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

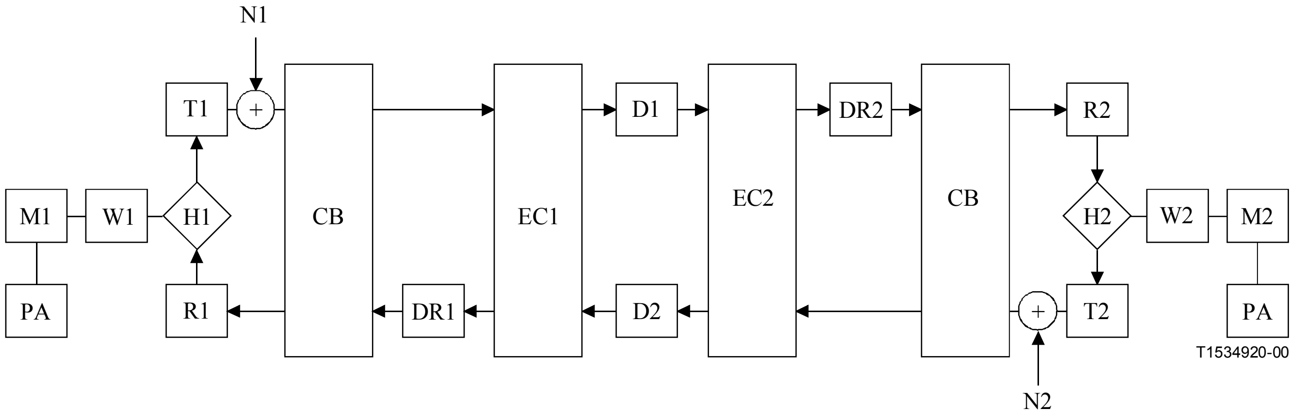
DR1, DR2 = echo path delay $\leq \Delta$ ms

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.




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

**Figure 28/G.168 – Test No. 14 Configuration**

### 6.4.3 Nonlinear echo paths

All tests related to nonlinear echo paths are optional unless otherwise indicated.

See I.5.3, I.5.4, I.6.2, I.6.3, I.6.6, and I.7/Appendix I for further discussion on this issue.

#### 6.4.3.1 Test No. 13 – Performance with ITU-T low bit rate coders in echo path (Optional – for further study)

Under study. (The intention is to put in a table of performance goals for each coder/algorithm.) A preliminary test configuration is given in Figure 29.

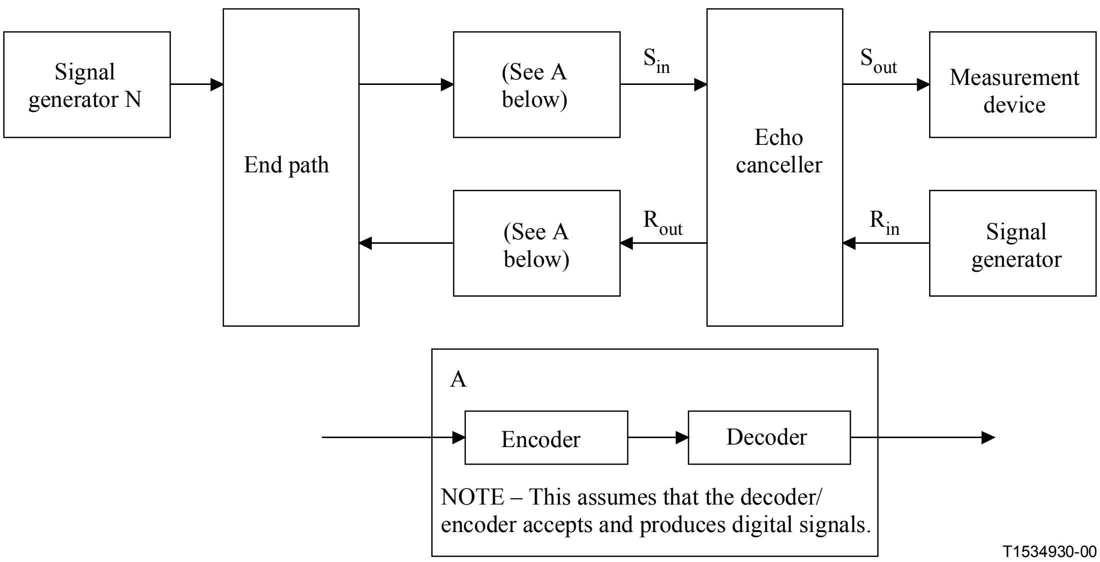



**Figure 29/G.168 – Test No. 13 configuration**

### 6.4.3.2 Test No. 15 – PCM Offset Test (Optional – for further study)

This test is meant to ensure that the echo canceller will operate properly in the presence of PCM offset in the speech signal applied to $S_{in}$.

PCM offset is an unwanted, fixed PCM signal level in the $S_{in}$ signal relative to the $R_{out}$ signal. This can be caused by some network equipment, such as PBX systems, and can result in degraded performance of echo canceller and other speech processing equipment. The method consists of applying a PCM offset error to the $S_{in}$ signal relative to the $R_{in}$ signal as indicated in Figure 30.

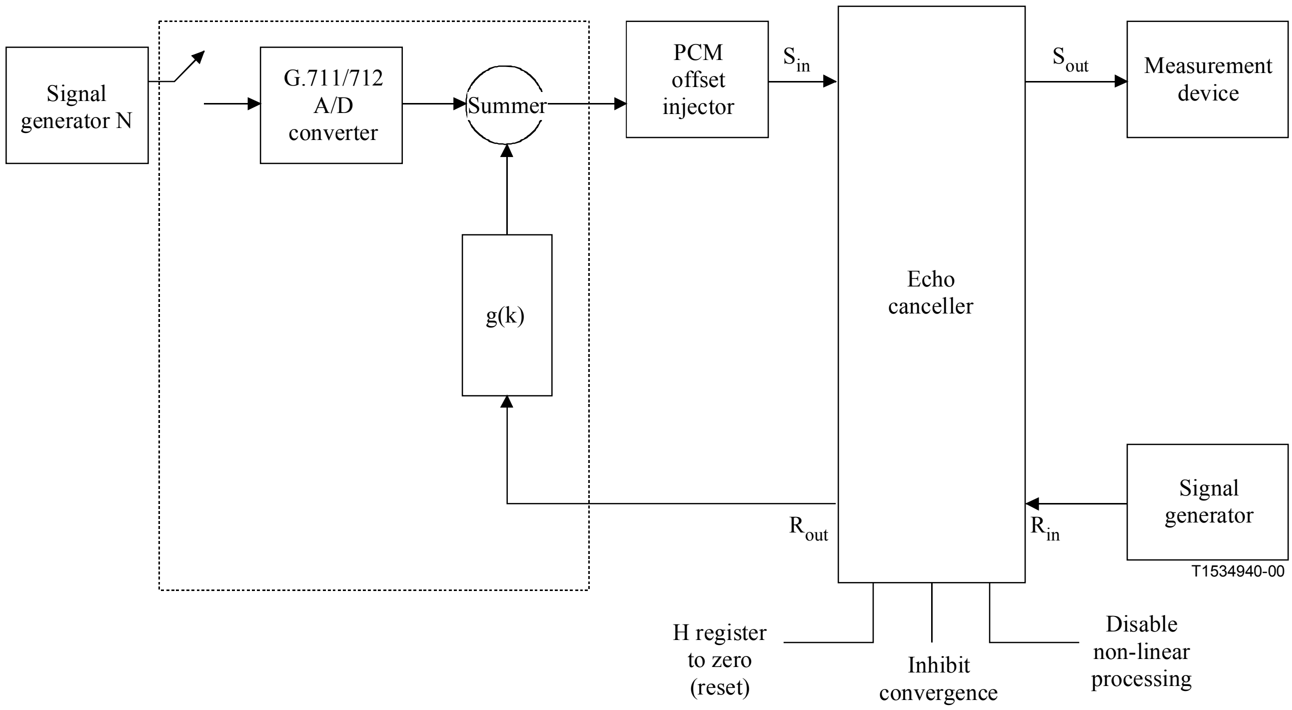



**Figure 30/G.168 – Test configuration for PCM offset test**

## 7 Characteristics of an echo canceller tone disabler

### 7.1 General

The echo cancellers covered by this ITU-T Recommendation should be equipped with a tone detector that conforms to this subclause. This tone detector should disable the echo canceller only upon detection of a signal which consists of a 2100 Hz tone with periodic phase reversals inserted in that tone, and 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.

To improve the operation of the echo canceller for fax signals and low-speed voiceband data, it may be beneficial for some echo cancellers to disable the NLP for such calls. In this case, the echo canceller may optionally detect any 2100 Hz tone without phase reversals. If 2100 Hz tone without phase reversal is detected, the echo canceller shall remain enabled, and the NLP may optionally be disabled. The frequency characteristics of the tone detector are given in Figure 31. The tone disabler characteristics as specified in 7.4 through 7.9 also apply for this NLP disabling detector. Note that if the 2100 Hz tone contains phase reversals, then the echo canceller shall be disabled as defined elsewhere in this subclause.

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 6.4.1. However, no relationship between the circuit conditions before and after disabling should be assumed. 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 the echo canceller should provide 64 kbit/s bit-sequence integrity when disabled.

A reference tone disabler is described in Annex A.

## 7.2 Detector characteristics

The tone detector shall detect a tone in the frequency range of 2100 Hz ± 21 Hz (see ITU-T Recommendation V.21 [26]). The detection channel bandwidth should be chosen wide enough to encompass this tone (and possibly other disabling tones used within national networks). At the same time, the detection channel bandwidth should be such that, in conjunction with guard action and timing, adequate protection is provided against false operation of the detector by speech signals. The detector channel sensitivity (threshold level) should be such that the detector will operate on the lowest expected power of the disabling tone. The band characteristics shown in Figure 31 will permit disabling by the 2100 Hz disabling tone as well as others used in North America. The figure indicates that in the frequency band 2079 Hz to 2121 Hz detection **must** be possible whilst in the band 1900 Hz to 2350 Hz detection **may** be possible.

Providing that only the recommended 2100-Hz disabling tone is used internationally, interference with signalling equipment will be avoided.

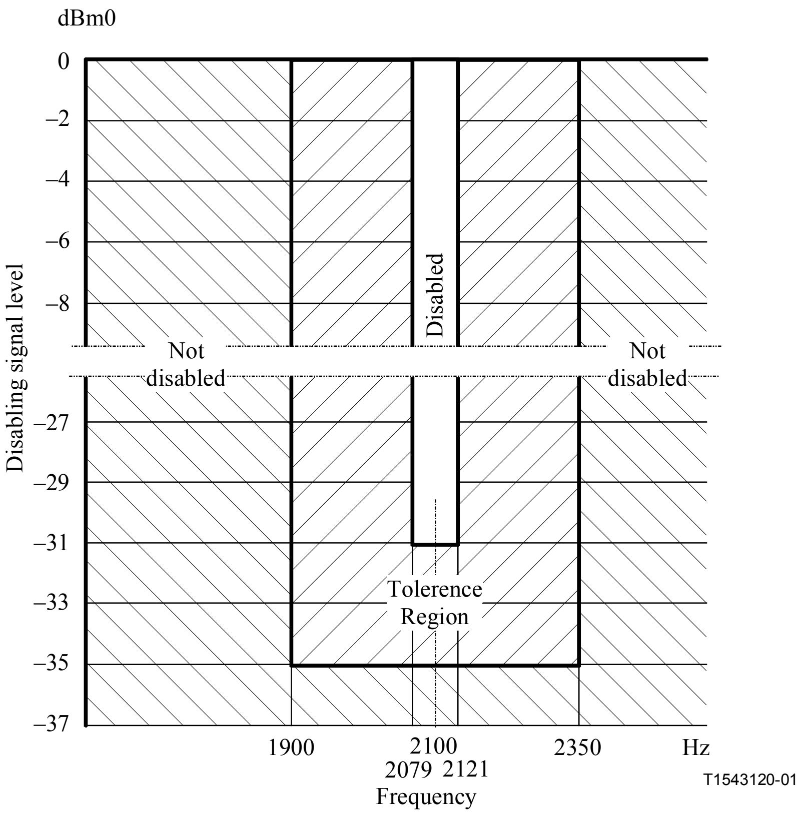



**Figure 31/G.168 – Required disabling band characteristics**

The echo canceller tone disabler requires the detection of a 2100 Hz tone with phase reversals. The characteristics of the transmitted signal are defined in Recommendation V.25 [28] and V.8 [25]. Phase variations in the range of 180º ± 25º should be detected while those in the range of 0º ± 110º should not be detected.

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

### 7.3 Guardband characteristics

The detector should operate correctly 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 180º ± 25º and non-detection of phase variations of 0º ± 110º) should fall by no more than 1% for each dB reduction in the signal-to-noise ratio below 11 dB. It is Noted that it is possible to design a detector capable of operating correctly at 5 dB signal-to-noise ratio.

### 7.4 Holding-band characteristics

The tone detector, after disabling either the NLP or the echo canceller, should hold the NLP or echo canceller in the disabled state for tones in a range of frequencies specified below. The release sensitivity should be sufficient to maintain disabling for the lowest level data signals expected, but should be such that the detector will release for the maximum idle or busy circuit noise. Thus the requirement follows:

The tone detector should hold the NLP or echo canceller in the disabled state 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.

### 7.5 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.

### 7.6 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 guard band 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.

### 7.7 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.

### 7.8 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.

### 7.9 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 favor 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 favor relying on 2 to improve the probability of correctly distinguishing between non-phase-reversed and phase-reversed 2100 Hz tones.

## 8 NLP's for use in echo cancellers

### 8.1 Scope

For the purpose of this ITU-T 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 NLP's in a number of ways (centre 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 8.2. 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.

## 8.2 General principles and guidelines

### 8.2.1 Function

#### 8.2.1.1 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 ($L_{\mathrm{RES}}$ as defined in 1.3.21) which remains after imperfect cancellation of the circuit echo so that the necessary low returned echo level ($L_{\mathrm{RET}}$ as defined in 1.3.22) can be achieved.

#### 8.2.1.2 Network performance

Imperfect cancellation can occur because echo cancellers which conform to this ITU-T Recommendation may not be capable of adequately modeling echo paths which generate significant levels of nonlinear distortion (see I.6.2). Such distortion can occur, for example, in networks conforming to ITU-T Recommendation G.113 in which up to five pairs of PCM codecs (conforming to ITU-T Recommendation G.712 [10]) are permitted in an echo path. The accumulated quantization distortion from these codecs may prevent an echo canceller from achieving the necessary $L_{\mathrm{RET}}$ 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 NLP's. 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 NLP's sometimes cause speech degradation.

#### 8.2.1.3 Limitations

This use of NLP's represents a compromise in the circuit transparency which would be possible by an echo canceller which could achieve the necessary $L_{\mathrm{RET}}$ 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 NLP's 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 $L_{\mathrm{RES}}$ should be low enough to prevent objectionable echo under double talk conditions.

#### 8.2.1.4 Data transmission

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

### 8.2.2 Suppression threshold

#### 8.2.2.1 General

The suppression threshold level ($T_{\mathrm{SUP}}$) 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.

### 8.2.2.2 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 $L_{\mathrm{RES}}$ for a "2s-talker" and a "2s-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 $L_{\mathrm{Rin}}$ of –10 dBm0, the quantization noise could result in an $L_{\mathrm{RES}}$ of –38 dBm0.

### 8.2.2.3 Adaptive suppression threshold

A good compromise can be made between using a high $T_{\mathrm{SUP}}$ to prevent it being exceeded by loud talker residual echo and using a low $T_{\mathrm{SUP}}$ to reduce speech distortion on break-in by making $T_{\mathrm{SUP}}$ 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.

### 8.2.3 Control of NLP activation

### 8.2.3.1 General

To conform to the recommendation made in 8.2.1.3, 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 $L_{\mathrm{RES}}$. When "inactive", it should not perform any nonlinear processing on any signal passing through the echo canceller.

### 8.2.3.2 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 $L_{\mathrm{RES}}$, they should be active when $L_{\mathrm{RES}}$ 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 favor the second.

### 8.2.3.3 Static characteristics

A conceptual diagram showing the two operational states of a NLP is shown in Figure 32. The $L_{\mathrm{Sin}}$ $L_{\mathrm{Rin}}$ plane is divided into two regions, W and Z by the threshold WZ ($T_{\mathrm{WZ}}$). 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.

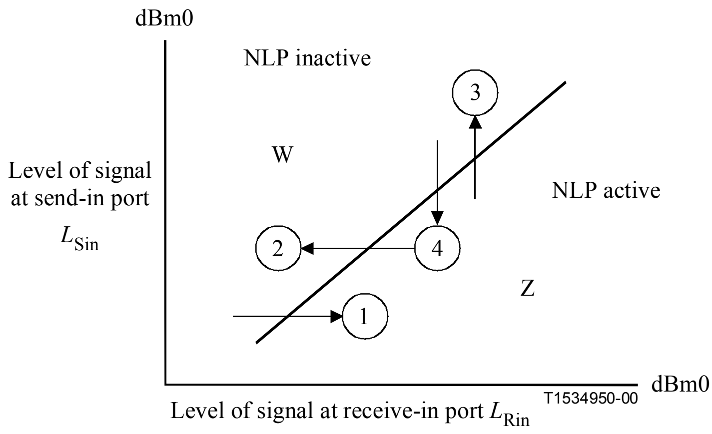



**Figure 32/G.168 – NLP operating regions**

### 8.2.3.4 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 32.

**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 favor 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.

### 8.2.4 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 NLP's. These control paths may include the activation control and adaptive suppression threshold level control.

### 8.2.5 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 6.4.2.3.1 are met.

### 8.2.6 Testing of NLP's

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 ITU-T Recommendation G.164. However, depending on the specific implementation of a NLP, the transitions between areas W and Z of Figure 31 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 ITU-T Recommendation G.164 may be used as a guide to the testing of NLP's it may be necessary to introduce unique test circuit modifications in order to make measurements on some specific NLP implementations. In particular, it is known that some echo cancellers employ a supplementary NLP technique independent of the NLP On/Off control. Those G.168 tests which call for NLP disabled do not strictly apply to such cancellers unless some further interpretation of results is made. The following paragraph suggests a technique for identifying the presence of such supplementary NLP functions.

#### 8.2.6.1 Testing for the Presence of an NLP

Set up G.168 Test 2B (Convergence without NLP). However, modify the test by mixing uncorrelated noise at various known levels below the returned CSS echo in the Send path. The noise level establishes a floor below which noise plus residual echo due to cancellation should not fall. Observing residual echo levels below the noise floor, then, is an indication of NLP action.

## 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 ITU-T Recommendations V.25 [28] and V.8 [25], 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 ITU-T 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 7.2, 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 ITU-T 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 ITU-T Recommendation G.229 [8]. 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 voice band, excluding the disable band, must be used to oppose disabling so that speech will not falsely operate the tone disabler. The guard band 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 guard band 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 ITU-T 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. ITU-T 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 guard band 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

The disabler meets the requirement in 7.7. 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) voice-band 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 8.1), 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 centre 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.

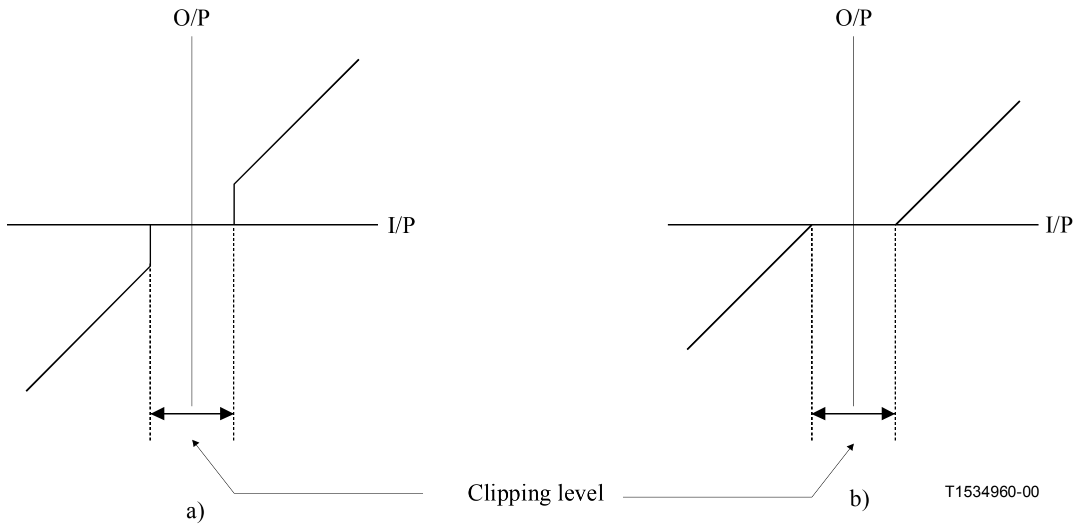



**Figure B.1/G.168 – Two examples of idealized centre clipper transfer function**

### B.2 Suppression threshold ($T_{SUP}$)

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

NOTE 1 – $T_{WZ}$ is as defined in 8.2.3.3.

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 $\geq 6$ 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 Table B.1 and B.2. Also see Figure 31.

### B.5 Frequency limits of control paths

See 8.2.4.

### B.6 Testing

Tables B.1 and B.2 indicate, by reference to ITU-T Recommendation G.164 [4], 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.

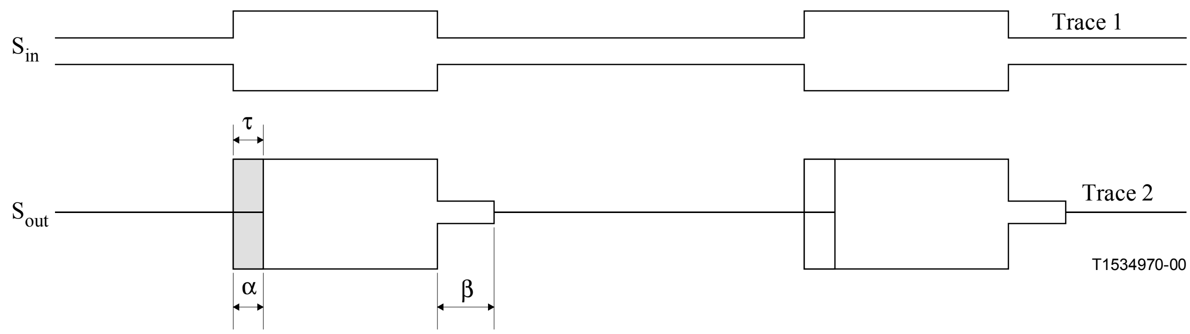
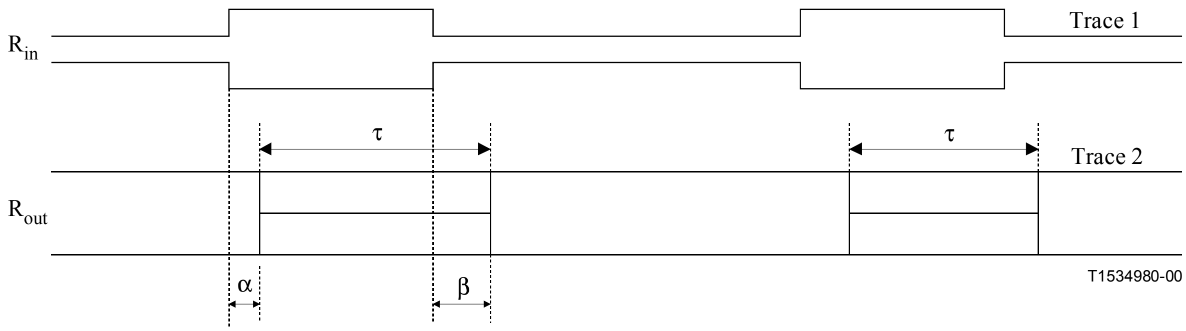



**Figure B.2/G.168 – Traces for NLP operate and hangover time, $L_{Rin}$ constant**




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**

**Table B.1/G.168 – NLP hangover times**

| Boundary | | Initial signal Send $L_{Sin}$ (dBm0) | Initial signal Receive $L_{Rin}$ (dBm0) | Final signal Send $L_{Sin}$ (dBm0) | Final signal Receive $L_{Rin}$ (dBm0) | Recommended value (ms) | Test No. (Rec. G.164) | Excursion (see Fig. 31) | Test circuit, Figure: | Oscilloscope Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 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<br>–40<br>–30 | –20<br>–15<br>–5 | –55<br>–40<br>–30 | –40<br>–40<br>–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<br>–40<br>–25 | –50<br>–30<br>–15 | –55<br>–55<br>–40 | –50<br>–30<br>–15 | 30-50 | | | | |

**Table B.2/G.168 – NLP operate times**

| Boundary | | Initial signal | | Final signal | | Recommended value (ms) | Test No. (Rec. G.164) | Excursion (see Fig. 31) | Test circuit, Figure: | Oscilloscope trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Send $L_{Sin}$ (dBm0) | Receive $L_{Rin}$ (dBm0) | Send $L_{Sin}$ (dBm0) | Receive $L_{Rin}$ (dBm0) | | | | | |
| W/Z | Fixed | −25 | −30 | −25 | −10 | 16-120 | 4 | Transition 1 | 14/G.164 | Trace 2 of Figure B.3 (α) |
| | Adaptive | −55<br>−40<br>−30 | −40<br>−40<br>−30 | −55<br>−40<br>−30 | −20<br>−15<br>−5 | 15-75 | | | | |
| Z/W | Fixed | −40 | −25 | −15 | −25 | ≤1 | 6 | Transition 3 | 17/G.164 | Trace 2 of Figure B.2 (α) |
| | Adaptive | −55<br>−55<br>−40 | −50<br>−30<br>−15 | −40<br>−40<br>−25 | −50<br>−30<br>−15 | ≤5 | | | | |

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 ITU-T Recommendation P.501 [10] on 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 ITU-T Recommendation.

## C.2 Composite Source Signal – General Considerations

### C.2.1 General Description of the Different Sequences

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

**Voiced signal produced from the "artificial voice" signal according to ITU-T Recommendation P.50 [14]**

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:

$$H(k)=W(k)\cdot e^{j\cdot i_k\cdot\pi};\ k=-M/2,\text{without }0;\ \mathrm{i_k}=-i_{-k}\text{ random} \qquad \text{(C.2-1)}$$

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:

$$S(n)=\frac{1}{M}\sum_{k=-M/2,k\neq 0}^{M/2} H(k)\cdot e^{j2\pi\cdot n\cdot k/M},\ n=-M/2,.....,M/2-1; \qquad \text{(C.2-2)}$$

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 analyze noise or artifacts 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 180 deg.).

### 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:

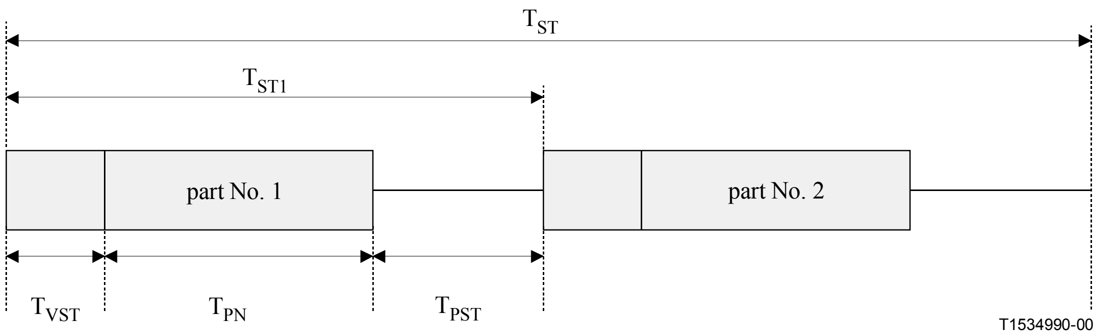
- sampling rate according to the one chosen for signal generation (preferred 44.1 kHz or 48 kHz);
- FFT length according to the one chosen for signal generation;
- rectangular windowing;
- no overlap;
- averaging over the **whole (cycled) sequence**, including voiced sounds, PN-sequences, pauses;
- calculation 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 largely depends on the power density spectrum of the input signal. Therefore these Composite Source Signal 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)**

**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.

$$H(k)=\begin{cases} W(k)\cdot e^{j\cdot i_k\cdot\pi}; k=-928,...,+928 \text{ except } 0, i_k\{+1,0\}, \text{ random}, i_k=-i_{-k} \\ O \text{ else} \end{cases} \qquad \text{(C.3-1)}$$

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 1.077 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.2 should be applied. The filter is chosen such, that the levels of the filtered and the unfiltered PN-sequence are equal. The filter owner frequencies are shown in Table C.2.

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.

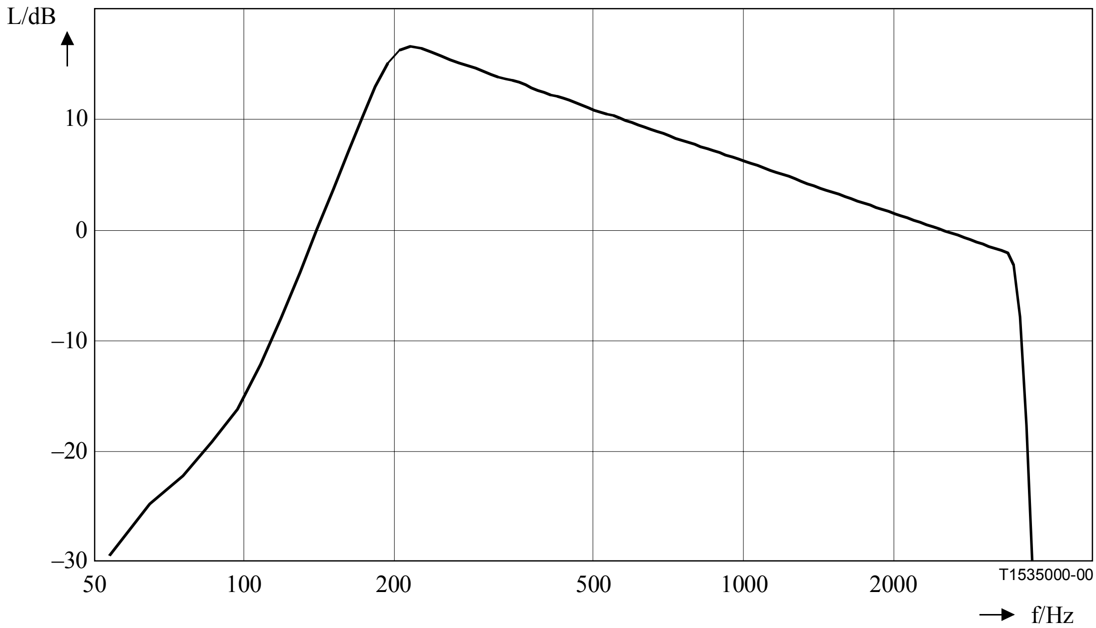



**Figure C.2/G.168 – Transfer function of the filter for band limiting the PN-sequence**

**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 180 deg.). 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.7 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.

**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 than 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.2 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 180 deg.) Thus the resulting length of the double talk signal is 800 ms.

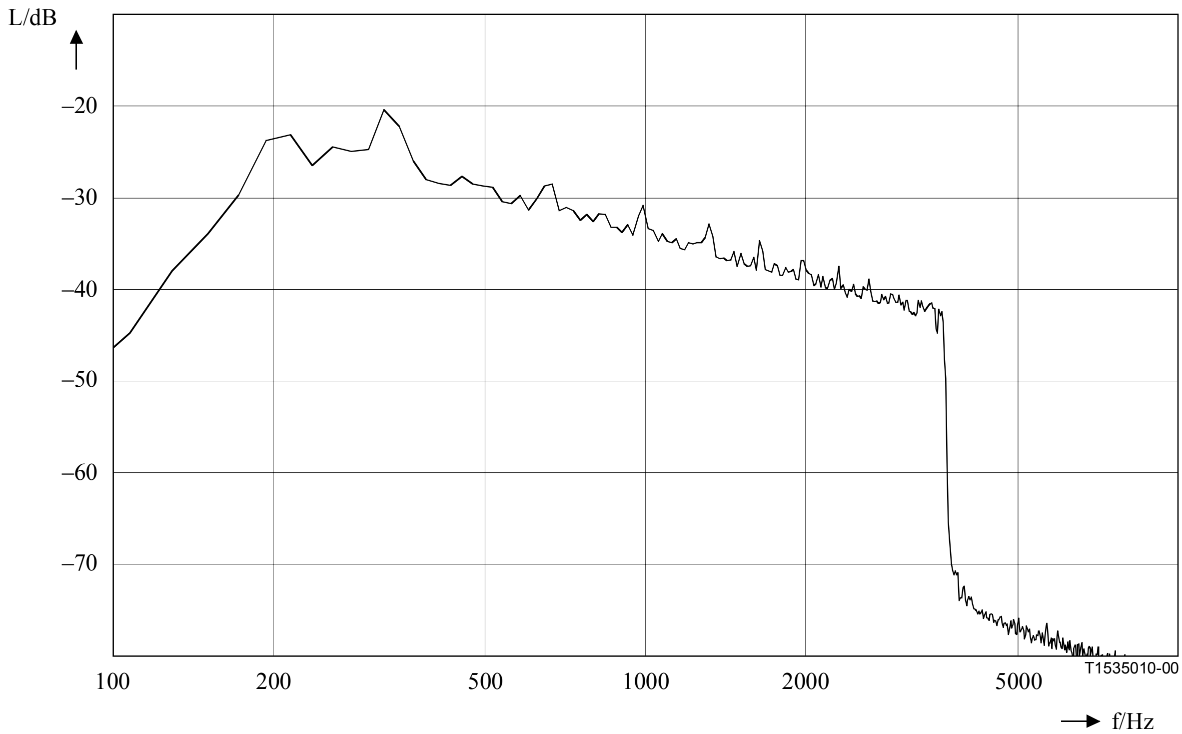



**Figure C.3/G.168 – Power density spectrum of the bandlimited CSS (single talk signal, analysis window: Hanning)**

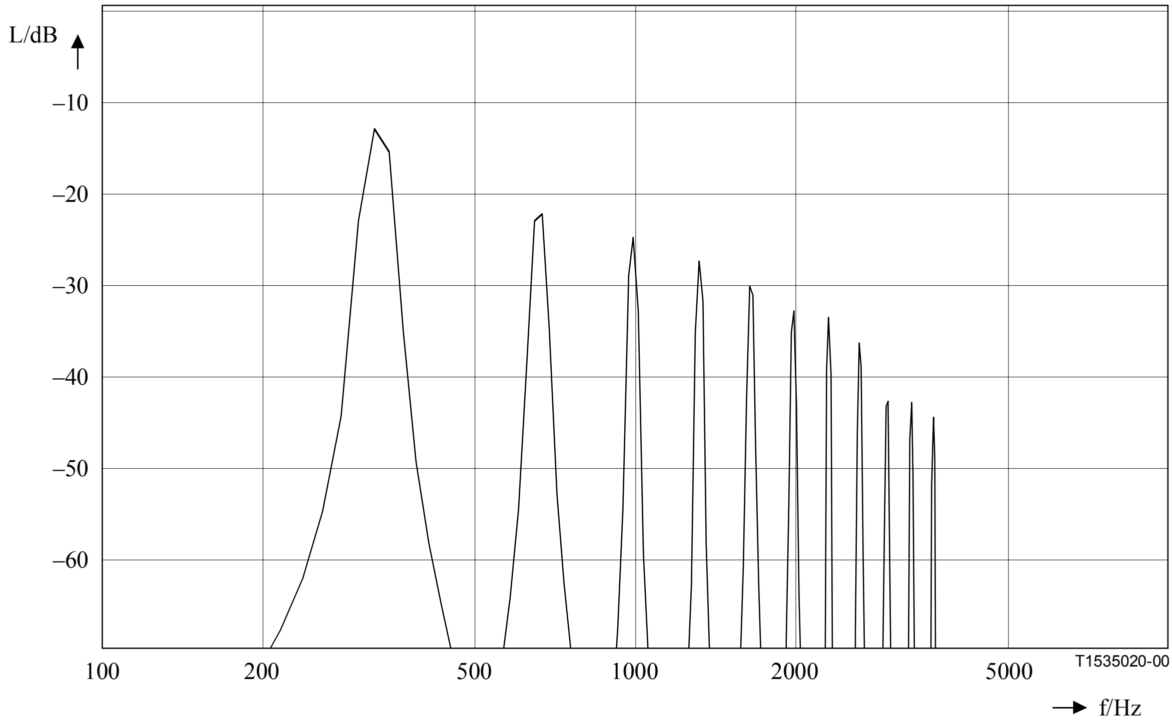



**Figure C.4/G.168 – Power density spectrum of the bandlimited voiced signal (single talk signal, analysis window: Hanning)**

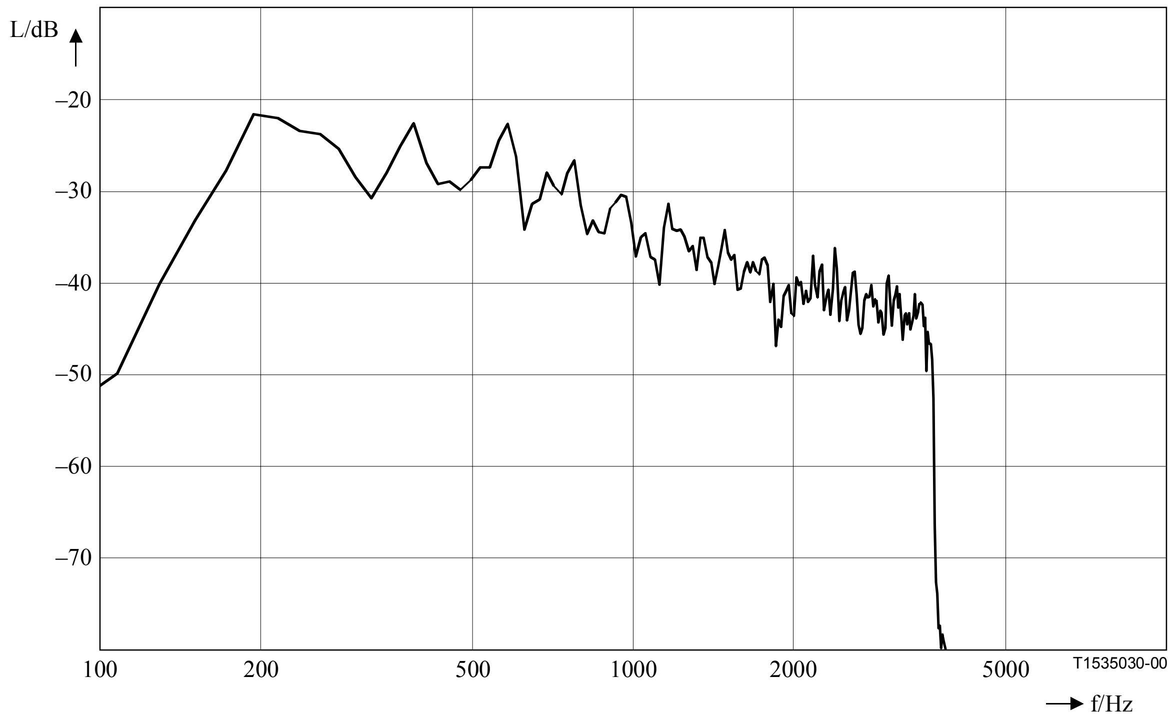



**Figure C.5/G.168 – Power density spectrum of the bandlimited double talk CSS (analysis window: Hanning)**

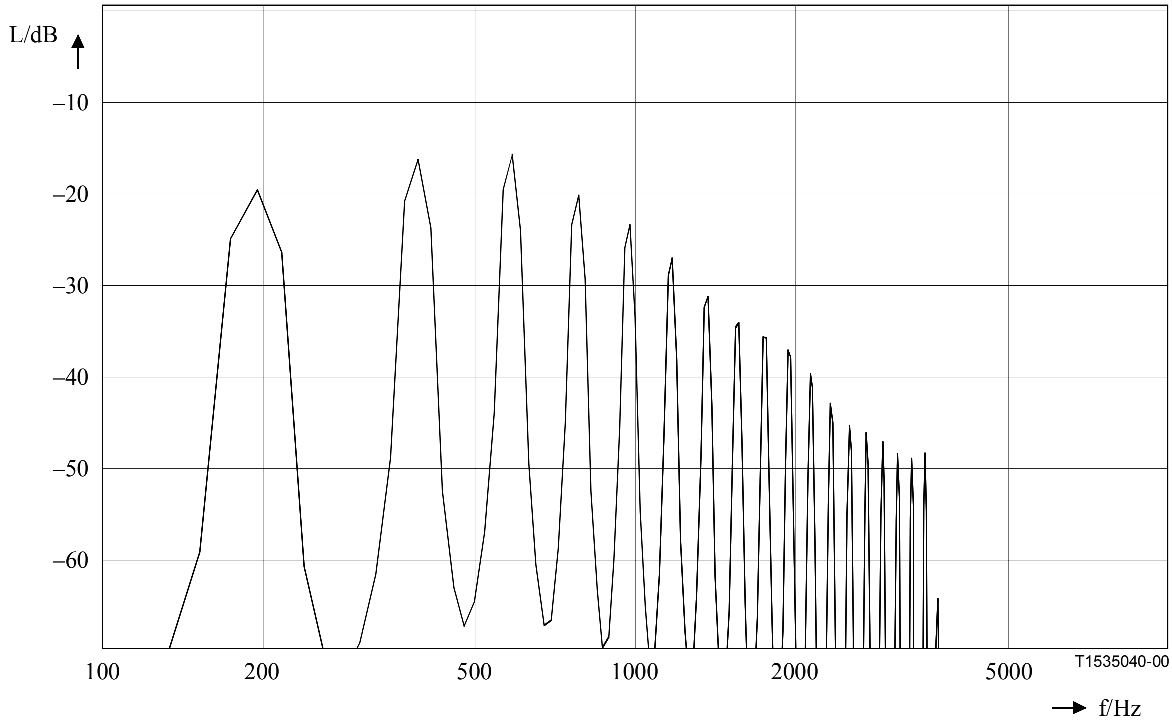
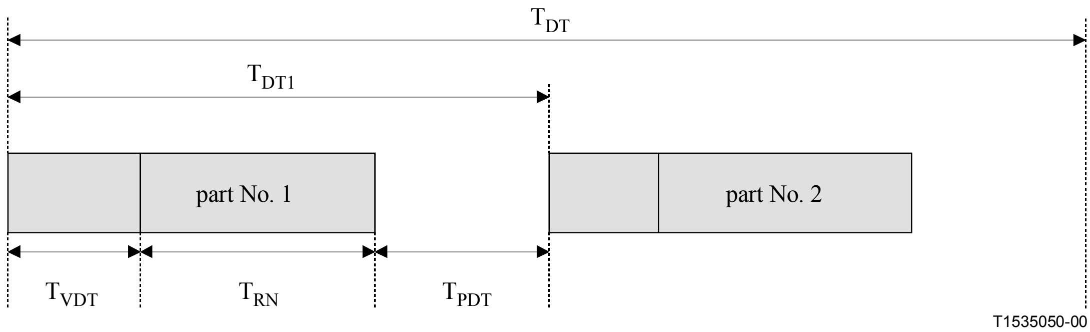



**Figure C.6/G.168 – Power density spectrum of the bandlimited double talk voiced signal (analysis window: Hanning)**

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.




| | | | |
|---|---|---|---|
| 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.7/G.168 – Composite Source Signals to simulate double talk (schematic)**

**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 analyze 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 artifacts may appear during the pauses (e.g. switched residual echo signal or modulated background noise). Therefore it is suitable to analyze 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 60651 [34], 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 analyzed. 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 60651 is more suitable for level calculations versus time.

This is a suitable method to analyze 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 60651 [34]. 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 60651, any peak detection or decay time constants referenced in IEC 60651 should not be incorporated for measurements in this ITU-T Recommendation.

### C.4.3 Level Calculations According to the Active Speech Level P.56

Level calculations can also be done according to the ITU-T Recommendation P.56 [15]. 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 analyzed. 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 analyze 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 60651 as described in C.4.2.

## ANNEX D

## Echo-Path Models for Testing of Speech Echo Cancellers

### D.1 Introduction

The following echo path models can be used for the tests in this ITU-T Recommendation. The echo path is simulated by a linear digital filter with the impulse response $g(k)$. To account for various delays, different ERLs and different dispersion characteristics and time widths, $g(k)$ is chosen as a delayed and attenuated version of any of the sequences $m_i(k)$, I = 1, 2, …, 7 that are given in tables in D.2 and D.3.

$$g(k) = \left(10^{-ERL/20} K_i\right) m_i(k - \delta) \tag{D.1-1}$$

The sequences $m_i(k)$ represent echo paths with various dispersion characteristics and different time widths. The delay $\delta$ should be chosen such that the non-zero values of $g(k)$ can be captured by the H-register of the echo canceller.

For the tests that use CSS or white noise as the input signals, the values of $K_i$ are given in the following Table D.1a for the seven sequences $m_i(k)$:

**Table D.1a/G.168 – Scaling factors $K_i$ for the seven digital echo path models**

| *i* | $K_i$ |
|---|---|
| 1 | $1.39 \times 10^{-5}$ |
| 2 | $1.35 \times 10^{-5}$ |
| 3 | $1.52 \times 10^{-5}$ |
| 4 | $1.77 \times 10^{-5}$ |
| 5 | $9.33 \times 10^{-6}$ |
| 6 | $1.51 \times 10^{-5}$ |
| 7 | $1.31 \times 10^{-5}$ |

NOTE – For echo-path models $m_2$ and $m_7$, the actual ERL measured from the echo produced by $g(k)$ are, respectively, 0.55 dB and 5.06 dB higher than the value of the ERL variable used in equation D.1-1. This is to ensure that the magnitude response of the scaled echo-path $g(k)$ does not exceed 0 dB over the appropriate frequency range.

For the tests that use tone(s) as input signals, the values of $K_i$ should be computed using the formula

$$K_i = \frac{1}{\max\limits_{f} \left| M_i(f) \right|} \tag{D.1-2}$$

where:

$$M_i(f) = \sum_{k=0}^{L-1} m_i(k) \exp\left( -\frac{j2\pi fk}{8192} \right) \tag{D.1-3}$$

with
- $i$ 1, 2, …, 7
- $f$ 0, 1, …, 4095
- $L$ length of $m_i(k)$

For the tests that use tones as the input signals, the values of $K_i$ are given in Table D.1b for the seven sequences $m_i(k)$ as follows:

**Table D.1b – Scaling factors $K_i$ for the seven digital echo path models for tones**

| $i$ | $K_i$ |
|---|---|
| 1 | $1.22 \times 10^{-5}$ |
| 2 | $6.78 \times 10^{-6}$ |
| 3 | $9.66 \times 10^{-6}$ |
| 4 | $1.07 \times 10^{-5}$ |
| 5 | $7.05 \times 10^{-6}$ |
| 6 | $8.60 \times 10^{-6}$ |
| 7 | $6.58 \times 10^{-6}$ |

## D.2 Echo-Path Models from Network Hybrid Simulator

This subclause contains four echo path models that are generated from a network hybrid simulator. Tables D.2, D.3, D.4, and D.5 show the values for $m_i(k)$ for i = 1,2,3,4. They are to be read in columns. The corresponding impulse responses are shown in Figures D.1, D.2, D.3 and D.4.

**Table D.2/G.168 – $m_1$(k); Impulse response of echo path model 1**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| −436 | 46150 | 390 | −3948 | −1098 | 745 | 1033 | 899 | 73 | −512 | −772 |
| −829 | 34480 | −8191 | −2557 | −618 | 716 | 1091 | 716 | −119 | −580 | −820 |
| −2797 | −10427 | −1751 | −3372 | −340 | 946 | 1053 | 390 | −109 | −704 | −839 |
| −4208 | 9049 | −6051 | −1808 | −61 | 880 | 1042 | 313 | −176 | −618 | −724 |
| −17968 | −1309 | −3796 | −2259 | 323 | 1014 | 794 | 304 | −359 | −685 | |
| −11215 | −6320 | −4055 | −1300 | 419 | 976 | 831 | 304 | −407 | −791 | |

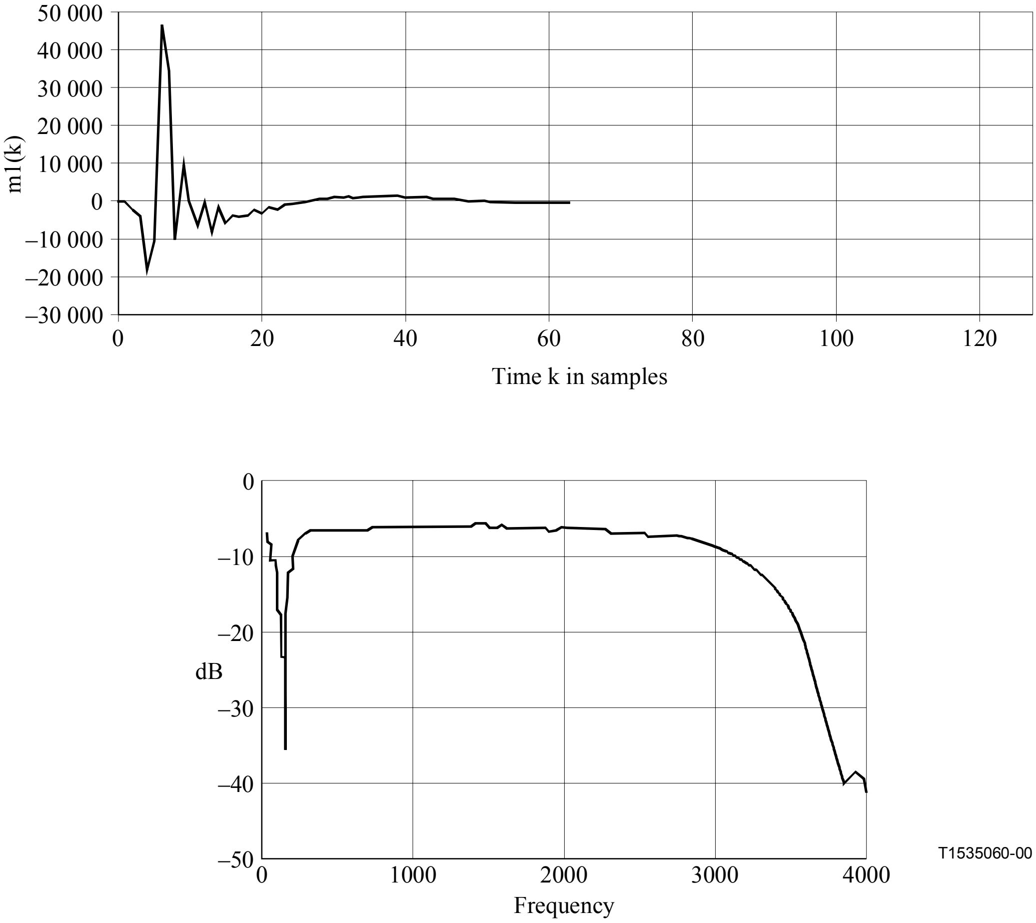



**Figure D.1/G.168 – Impulse response and magnitude response of echo-path model 1 (short dispersion, ERL = 7.6 dB)**

**Table D.3/G.168 – $m_2(k)$; Impulse response of echo-path model 2**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| −381 | −21370 | 13509 | −3858 | 1316 | −1468 | 789 | 658 | −331 | −479 | −249 |
| 658 | −5307 | 17115 | −1979 | −693 | −1221 | 954 | 476 | −347 | −479 | −216 |
| 1730 | 23064 | 13952 | 6029 | −759 | −842 | 756 | 377 | −430 | −512 | −249 |
| −51 | 24020 | 13952 | 5616 | −1517 | −463 | 839 | 377 | −314 | −479 | −265 |
| −3511 | 1020 | 97 | 7214 | −2176 | −298 | 872 | 262 | −430 | −397 | −166 |
| −1418 | −12374 | −9326 | 6820 | −2028 | −68 | 1020 | 97 | −463 | −430 | −232 |
| 7660 | −16296 | −9046 | 3935 | −2654 | 64 | 789 | −68 | −463 | −397 | |
| 8861 | −19524 | −15208 | 3919 | −1814 | 493 | 822 | −183 | −414 | −298 | |
| −8106 | −7480 | −9853 | 921 | −2077 | 723 | 558 | −232 | −381 | −265 | |

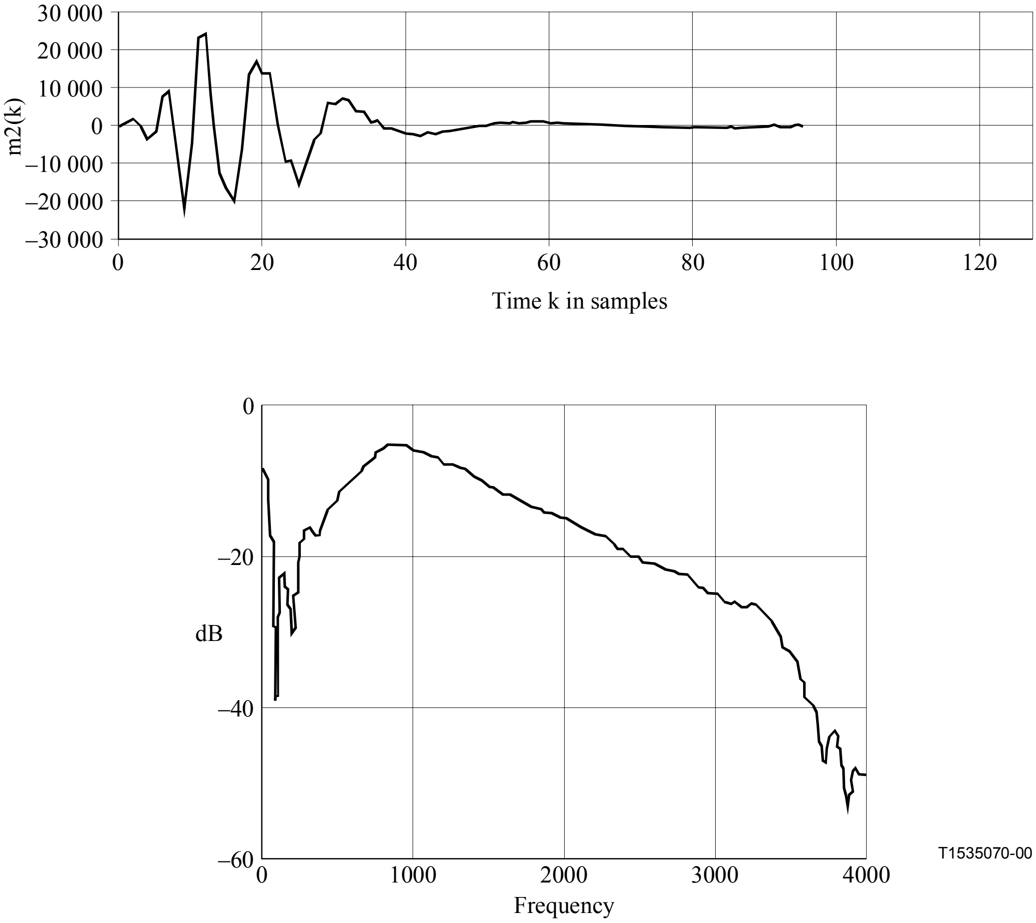



**Figure D.2/G.168 – Impulse response and magnitude response of echo-path model 2 (medium short dispersion, ERL = 12.2 dB)**

**Table D.4/G.168 – $m_3(k)$; Impulse response of echo-path model 3**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| −448 | −26261 | 14164 | 3271 | −3101 | 2172 | −139 | −1066 | −814 | −233 | −390 |
| −436 | −16249 | 13467 | 6566 | −9269 | 5387 | −573 | −1020 | −871 | −333 | −482 |
| 2230 | 21637 | 4438 | 4277 | −6146 | 4598 | −1100 | −1100 | −734 | −356 | −459 |
| 2448 | 25649 | 8627 | 11131 | −2553 | 3535 | −1157 | −1008 | −642 | −390 | −482 |
| −4178 | −2267 | 456 | 7562 | −6272 | 4004 | −1180 | −1077 | −562 | −310 | −551 |
| −7050 | −10311 | −11879 | 1475 | 811 | 2311 | −1455 | −1088 | −356 | −265 | −573 |
| 5846 | −4693 | −6352 | 3728 | 124 | 2150 | −1123 | −917 | −379 | −368 | |
| 18581 | −12690 | −5104 | −3525 | 788 | 1017 | −1386 | −917 | −345 | −310 | |
| 2322 | −7428 | −7496 | −7301 | 5147 | 330 | −1123 | −963 | −230 | −310 | |

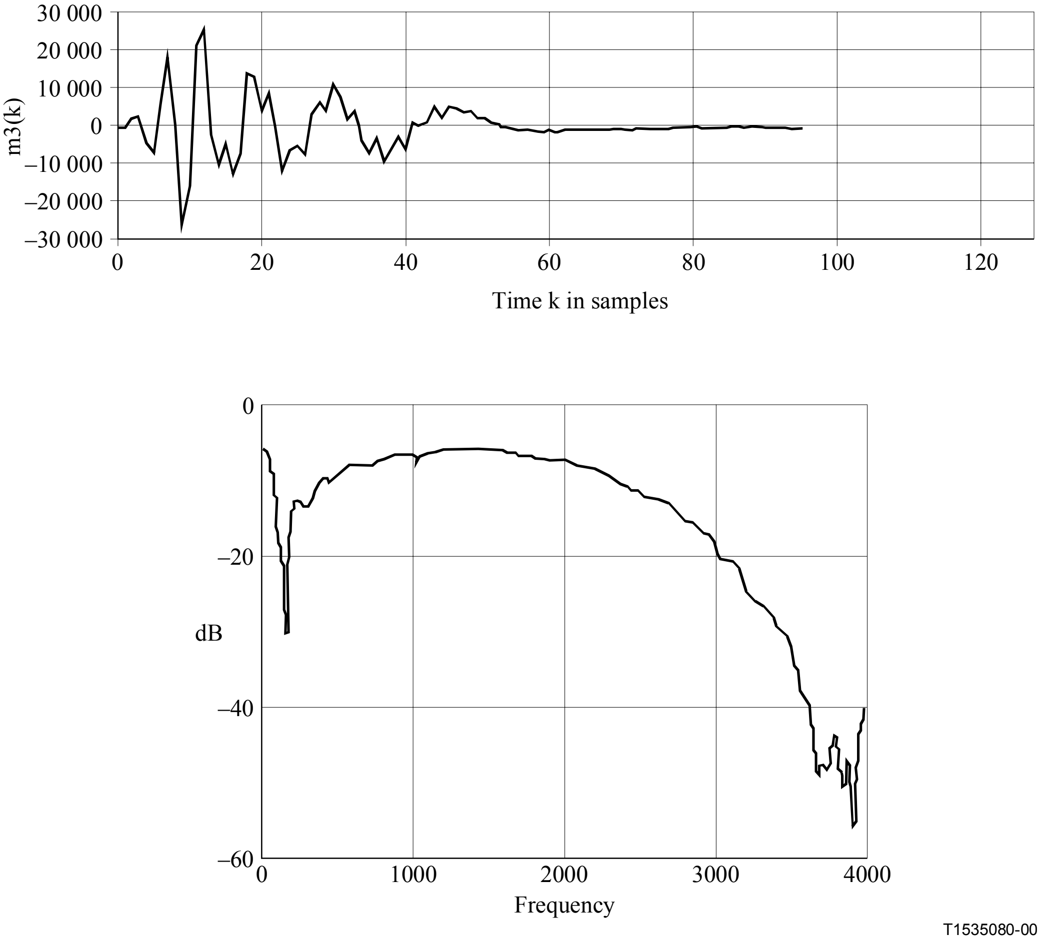



**Figure D.3/G.168 – Impulse response and magnitude response of echo-path model 3 (medium long dispersion, ERL = 9 dB)**

**Table D.5/G.168 – $m_4(k)$; Impulse response of echo-path model 4**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 160 | 4041 | 14624 | 6850 | 2393 | 1592 | 2654 | 4617 | –220 | 203 | –57 |
| 312 | 14484 | –6975 | 3944 | 2784 | –4752 | –881 | 3576 | –306 | –111 | –24 |
| –241 | –1477 | –17156 | 6969 | –892 | –3646 | –4113 | 2382 | 257 | 95 | 30 |
| –415 | –21739 | –187 | 8694 | –7366 | –5207 | –3244 | 2839 | 615 | –79 | –68 |
| 897 | –4470 | 149 | –4068 | –3376 | –5577 | –7289 | –404 | 225 | 30 | 84 |
| 908 | 25356 | 1515 | –3852 | –5847 | –501 | –3830 | 539 | 561 | 84 | –155 |
| –1326 | 11458 | 14907 | –5793 | –2399 | –1174 | –4600 | –1803 | 8 | –13 | –68 |
| –1499 | –19696 | 4345 | –9371 | 3011 | 4041 | –2508 | –1401 | 344 | –68 | 19 |
| 2405 | –11800 | –7128 | 453 | 1537 | 5647 | 431 | –1705 | 127 | –241 | |
| 3347 | 5766 | –2757 | 1060 | 6623 | 4628 | –144 | –2269 | –57 | –68 | |
| –3624 | 789 | –10185 | 3965 | 4205 | 7252 | 4184 | –783 | 182 | –24 | |
| –7733 | 6633 | –7083 | 9463 | 1602 | 2123 | 2372 | –1608 | 41 | 19 | |

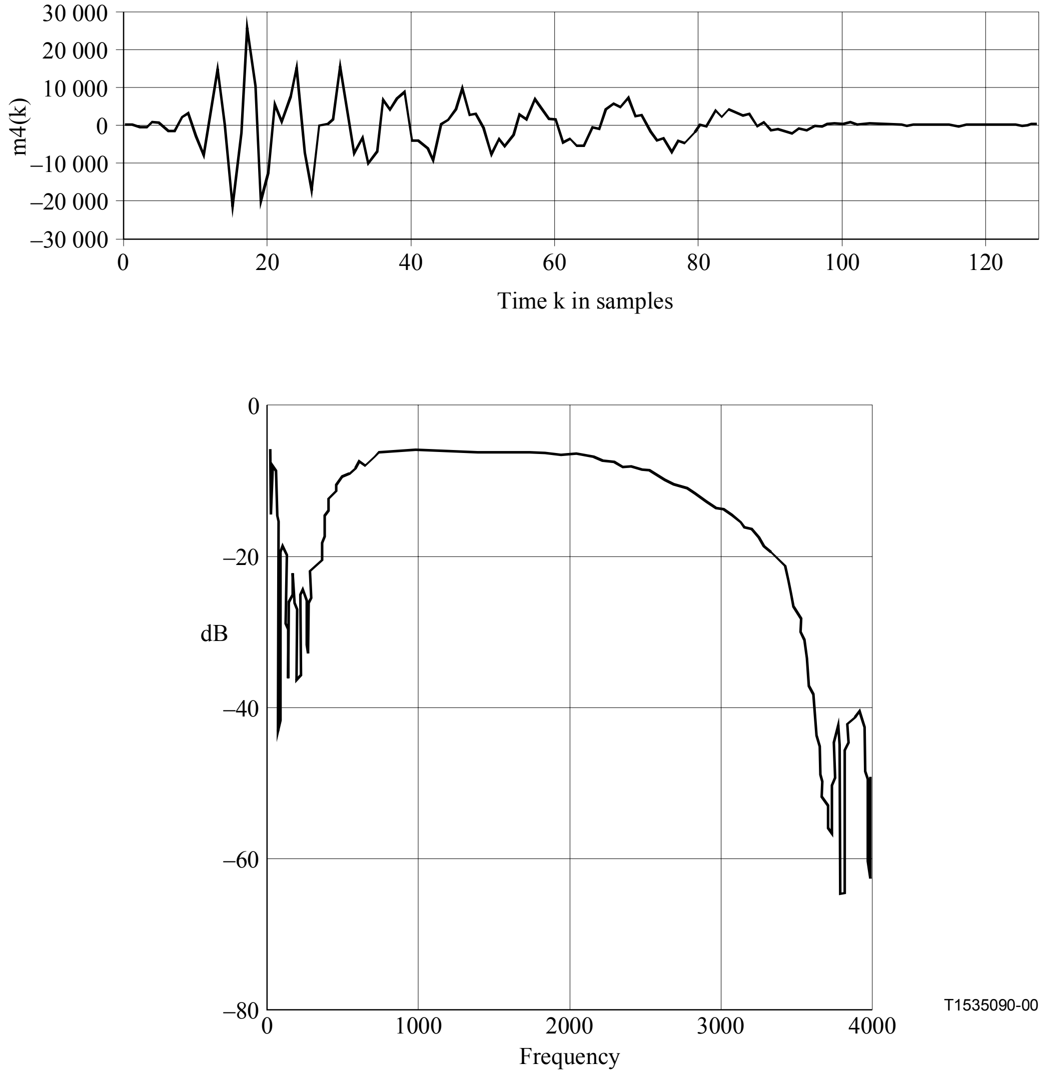



**Figure D.4/G.168 – Impulse response and magnitude response of echo-path model 4 (long dispersion, ERL = 8.6 dB)**

## D.3 Echo-path models measured from telephone networks in North America

This subclause presents three realistic digital echo-path models measured from telephone networks in North America. The measurement method for the echo-path models is described in Appendix II. Appendix II also includes the statistical characteristics (magnitude response, dispersion time width and the number of reflections) of echo paths measured in North America.

Tables D.6, D.7 and D.8 below tabulate the three echo-path models. The numbers are read in columns. The impulse responses as well as the magnitude responses of the models are shown in Figures D.5, D.6 and D.7 respectively.

The echo-path model shown in Table D.6 has a single reflection with a dispersion width of about 6 ms. This model has a frequency domain characteristic very close to the mean of the measured echo-path magnitude responses. This kind of echo path occurs most often in the measurements.

The echo-path model shown in Table D.7 has a single reflection. It has a longer dispersion width of about 9 ms. The long dispersion width is due to the spectral peak at about 250 Hz.

The echo path model shown in Table D.8 has double reflections. The impulse responses from the two reflectors are overlapped with each other. The dispersion width of this model is about 6 ms.

Note that due to the specific magnitude response of echo path model $m_7$, it may not be an effective model for tests 10A and 10B when the NLP is disabled. This is, coincidentally, due to notches in the magnitude response at the frequencies used in the tests.

**Table D.6/G.168 – $m_5(k)$; Impulse response of the echo-path model 5**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 293 | 896 | 20 | –22548 | 3889 | –5022 | –1608 | 1640 | 733 | –513 | –404 |
| 268 | 604 | –938 | –43424 | 7241 | –4039 | –645 | 1901 | 665 | –473 | –344 |
| 475 | 787 | –523 | 2743 | 925 | –4842 | –495 | 1687 | 323 | –588 | –290 |
| 460 | 561 | –1438 | 25897 | 2018 | –4104 | 279 | 1803 | 221 | –612 | –202 |
| 517 | 538 | –1134 | 7380 | –821 | –4089 | 471 | 1543 | –14 | –652 | –180 |
| 704 | 440 | –1887 | 21499 | –2068 | –3582 | 947 | 1566 | –107 | –616 | –123 |
| 581 | 97 | –1727 | 11983 | –2236 | –2978 | 1186 | 1342 | –279 | –566 | |
| 879 | 265 | –1698 | 10400 | –4283 | –2734 | 1438 | 1163 | –379 | –515 | |
| 573 | –385 | –4266 | 11667 | –3406 | –1805 | 1669 | 963 | –468 | –485 | |

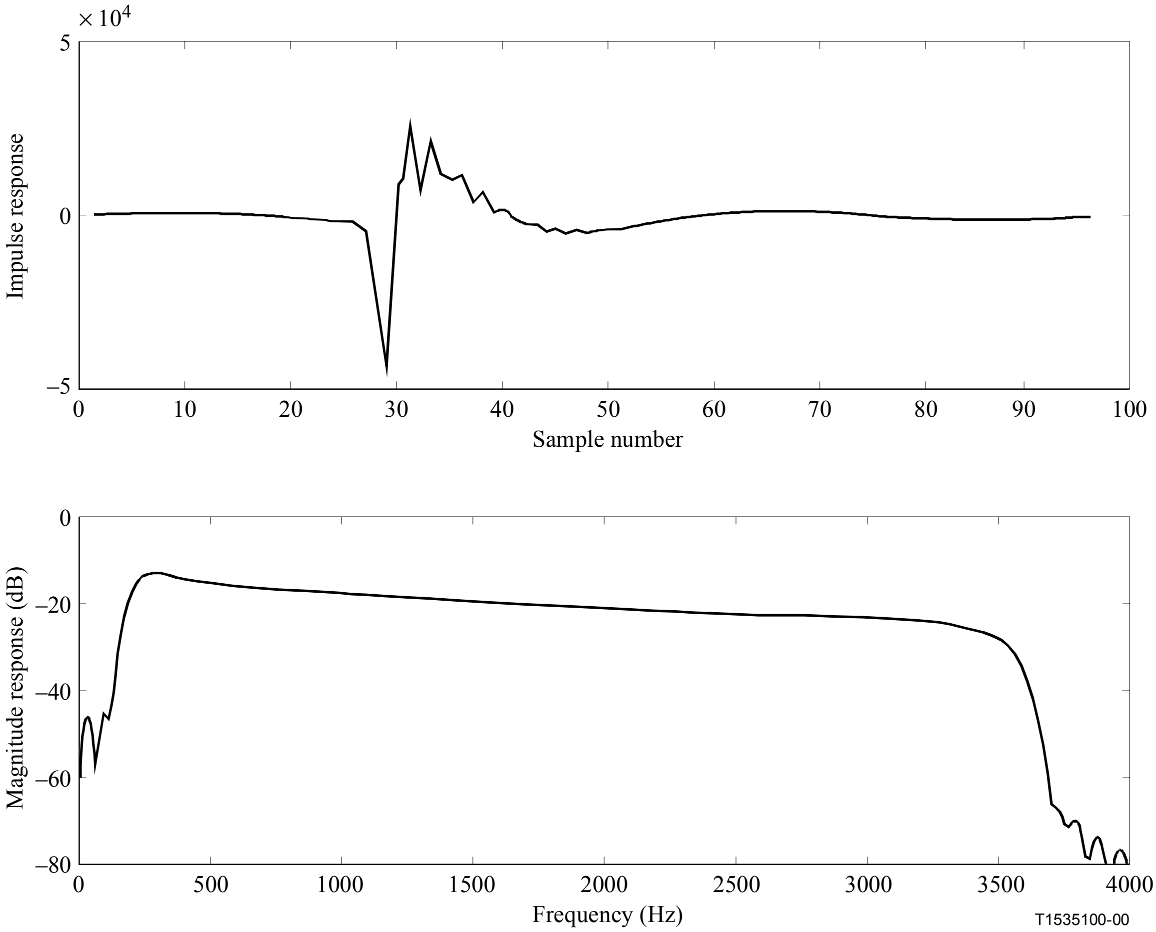



**Figure D.5/G.168 – Impulse Response and magnitude response of echo-path model 5**

**Table D-7/G.168 – $m_6(k)$; Impulse response of the echo-path model 6**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 29 | –450 | 158 | 19522 | 2469 | 5025 | –1117 | –1956 | 1724 | 482 | –838 |
| 109 | –105 | 1341 | 8421 | –7994 | 3946 | –2134 | –1539 | 1871 | 289 | –837 |
| –83 | –503 | 195 | –50953 | 490 | 4414 | –2547 | –1239 | 1767 | 54 | –834 |
| 198 | 145 | 1798 | –9043 | –3860 | 4026 | –2589 | –570 | 1802 | –137 | –740 |
| –294 | –490 | 344 | 18046 | –837 | 3005 | –3310 | –377 | 1630 | –321 | –673 |
| –135 | 267 | 1845 | –13553 | 490 | 3380 | –2778 | 251 | 1632 | –490 | –581 |
| –415 | –231 | 629 | 13336 | –636 | 1616 | –3427 | 331 | 1379 | –638 | –493 |
| –202 | 340 | 1604 | –3471 | 3682 | 2007 | –2779 | 964 | 1271 | –764 | –436 |
| –444 | 77 | 1182 | –107 | 1141 | 158 | –3116 | 1177 | 1063 | –836 | –327 |
| –337 | 343 | 940 | 1788 | 5019 | 388 | –2502 | 1449 | 856 | –800 | –201 |
| –313 | 783 | 5163 | –7409 | 2635 | –1198 | –2399 | 1564 | 711 | –859 | |

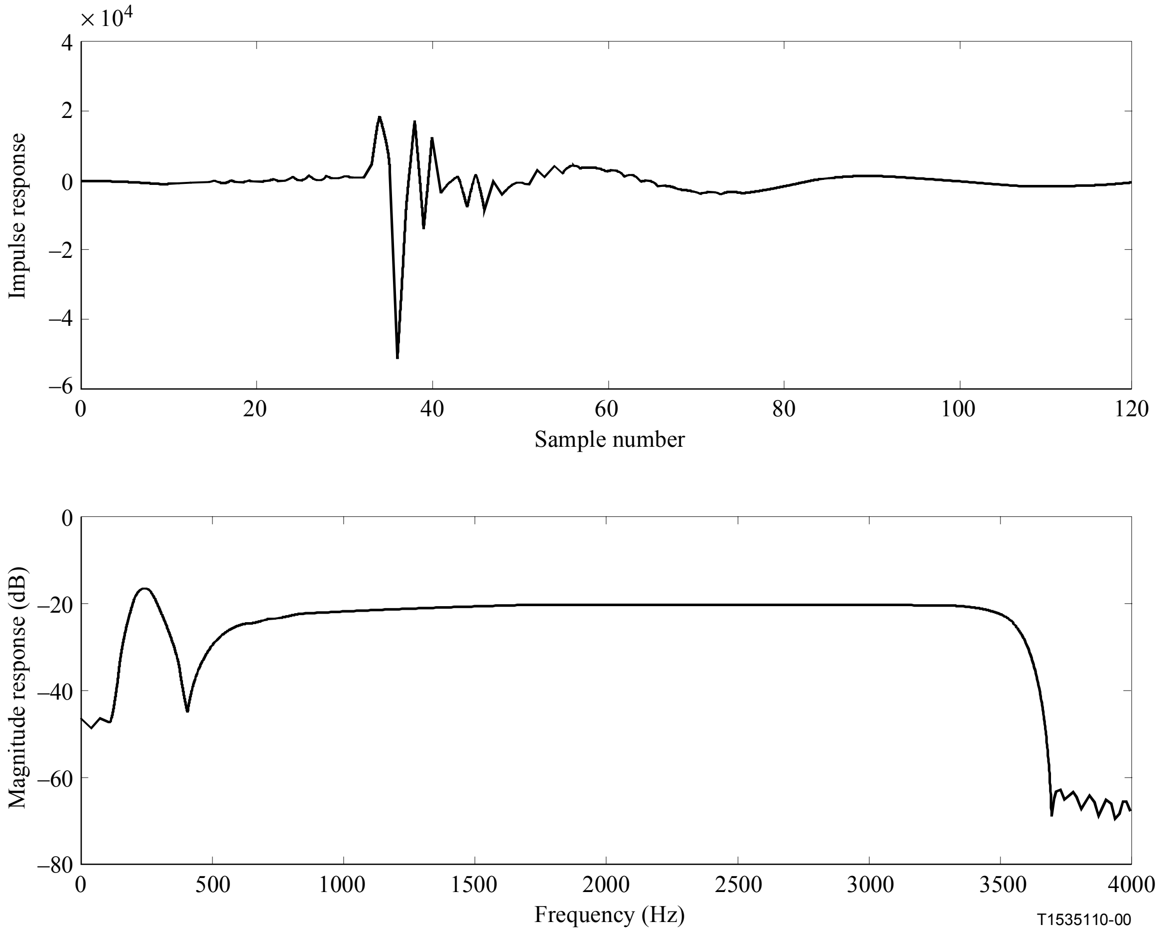



**Figure D.6/G.168 – Impulse response and magnitude response of echo-path model 6**

Note that due to the specific magnitude response of echo path model $m_7$, it may not be an effective model for tests 10A and 10B when the NLP is disabled. This is, coincidentally, due to notches in the magnitude response at the frequencies used in the tests.

**Table D.8/G.168 – $m_7(k)$; Impulse response of the echo-path model 7**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 258 | −343 | −1601 | 8950 | 18072 | −4342 | 6868 | 1239 | −415 | −325 | 365 |
| −111 | −596 | −1389 | −1574 | −14410 | −7415 | −2195 | 2 | −372 | −245 | 303 |
| 337 | −177 | −13620 | 758 | −7473 | 7929 | 3425 | −427 | −769 | −255 | 251 |
| −319 | −1187 | −720 | 3526 | 19836 | −10726 | 1969 | 596 | −183 | −60 | 230 |
| 347 | −52 | 33818 | −3118 | −16854 | 6239 | −109 | −1184 | −785 | 35 | 209 |
| −434 | −1781 | −10683 | 2421 | −3115 | −2526 | 3963 | 551 | −270 | 218 | 179 |
| 192 | −147 | −6742 | −8966 | 9483 | −1317 | −1275 | −1244 | −659 | 149 | |
| −450 | −1959 | 12489 | −4901 | −17799 | 5345 | 3087 | 141 | −377 | 340 | |
| −108 | −326 | −9862 | 11385 | 7399 | −4565 | −892 | −743 | −523 | 233 | |

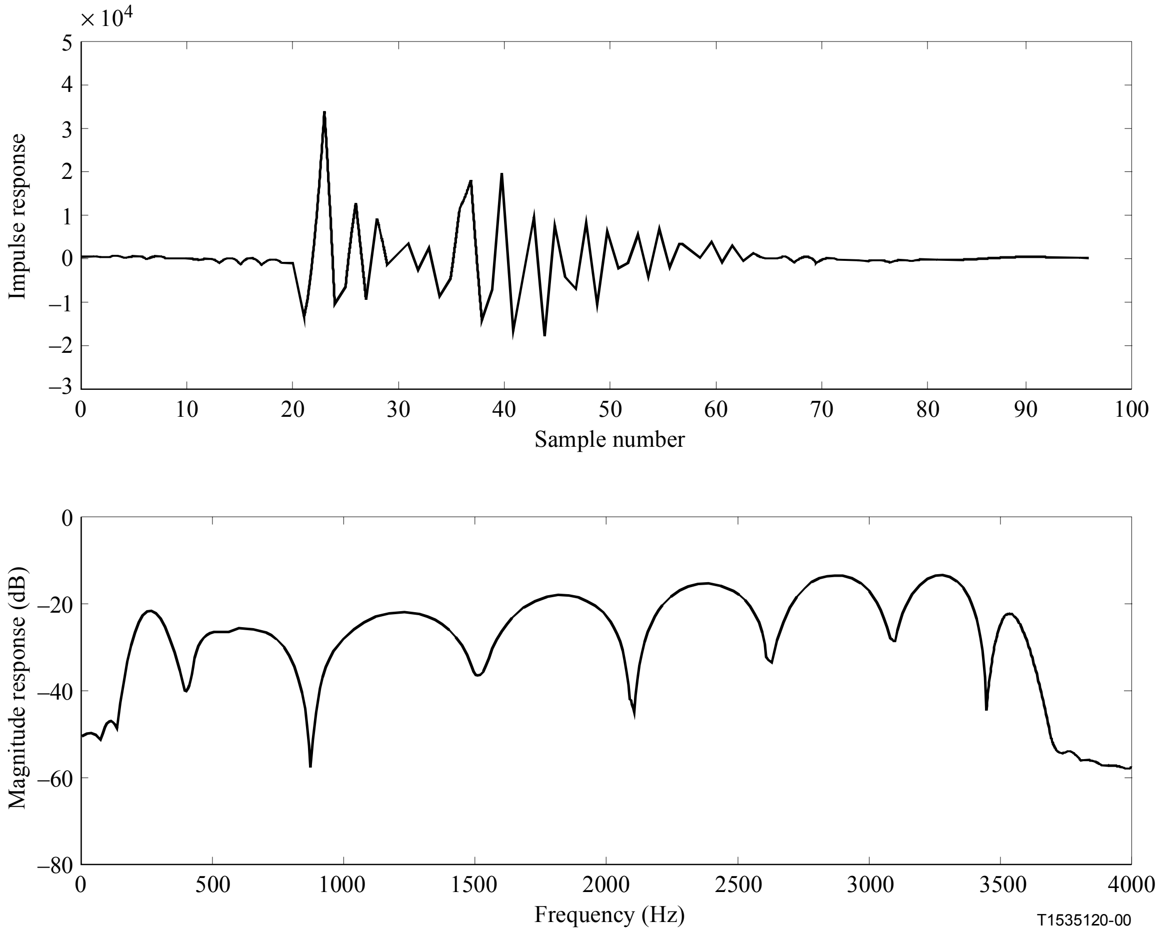



**Figure D.7/G.168 – Impulse response and magnitude response of echo-path model 7**

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 Telecommunications 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 voice-band 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 ITU-T Recommendation G.131 [3]. 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 [4] analog or digital echo suppressors, G.165 [5] analog or digital echo cancellers equipped with G.164 tone disablers, and, G.165 analog 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 ITU-T 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 voice-band 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 less than 400 ms at the echo suppressor may allow the echo suppressor to remain 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 ITU-T Recommendation T.30 [23]; 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, as specified in Table 4/G.164. The suppression hangover time guards against echo stored in the local echo path.

Now, ITU-T Recommendation T.30 specifies that the guard time between V.21 [26] and V.29 [31] 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 V.32 [32], most 2-wire modems used frequency division to provide duplex operation (i.e. different carrier frequencies were used for each direction of transmission). In the early 1980s, data showed that some echo cancellers did improve the operation (i.e. reduce or eliminate bit errors) for low-speed modems designed according to ITU-T Recommendations V.21 [26], V.23 [27], V.26 [29] (alternative B), V.27 *ter* [30] and V.29 [31]. Therefore, it was accepted that these modems benefited from an active echo canceller and a disabled echo suppressor. Accordingly, in 1984, Recommendation G.165 [5] was modified to recommend that echo cancellers be disabled with a 2100 Hz tone with phase reversals.

Data have indicated that certain combinations of modems and echo cancellers, in various simulated network configurations, exhibit degraded performance when the echo cancellers are enabled. This degradation has been reported by various Administrations. 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 voice-band data modem is not to be confused with the network echo cancellers that conform to ITU-T Recommendation G.165, because the performance requirements for each are very different.

### 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 voice-band 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 ITU-T Recommendation G.165).

Echo cancellers have the following main advantages over echo suppressors:

- Send path transparency is improved;
- NLP 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 ITU-T 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 ITU-T 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. V.32 [32], V.34 [33]), 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 ITU-T 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 ITU-T 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 ITU-T Recommendations G.122 [2], G.131 [3] and P.310 [13] (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 ITU-T Recommendations G. 223 [7], V.2 [24] and M.1050 [12]). 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 ITU-T Recommendations for PSTNs that utilize analogue accesses and for connections from digital cellular networks. Encoders should be consistent with ITU-T Recommendation G.711 [9]. For PSTNs with digital access, guidance for terminal design can be found in ITU-T Recommendation P.310 [13].

### 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 ITU-T Recommendations G.131 [3] and G.114 [1] 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. An example of this is given in Appendix III.

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 ITU-T 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 voice-band 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 Voice-band Data

Full-duplex data transmission in the voice-band 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 ITU-T 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 voice-band 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 ITU-T Recommendations G.164 (digital 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 ITU-T 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 and 5.5/G.164).

The V.27 *ter* modulation scheme employed by ITU-T 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 voice-band transmission). As explained earlier, enabled echo suppressors may cause errors in voice-band data transmission. However, it may be preferable to keep them enabled during facsimile transmission.

- If echo cancellers are disabled according to the procedures of ITU-T 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. To ensure that the echo control device re-enables, a period of at least 400 ms is required during which no energy could flow in either direction. 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 completely eliminate 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 explicitly discuss 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 ITU-T Recommendations imply that echo cancellers should be enabled during facsimile transmission. Generally, echo suppressors do not provide the same level of performance for speech, voice-band 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 in mind, ITU-T Recommendation G.165 was modified in 1992 to include an overload test (Test No. 8) at levels exceeding 0 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 does 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

ITU-T Recommendation G.165 was amended in 1992 to make it clear that a 2100 Hz disabling tone with phase reversals should cause the echo canceller to disable and provide an analogue clear-channel signal path. 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, 1992 revision). 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 ITU-T 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.
- With the increased use of digital techniques in processing voice-band 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 may still 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.

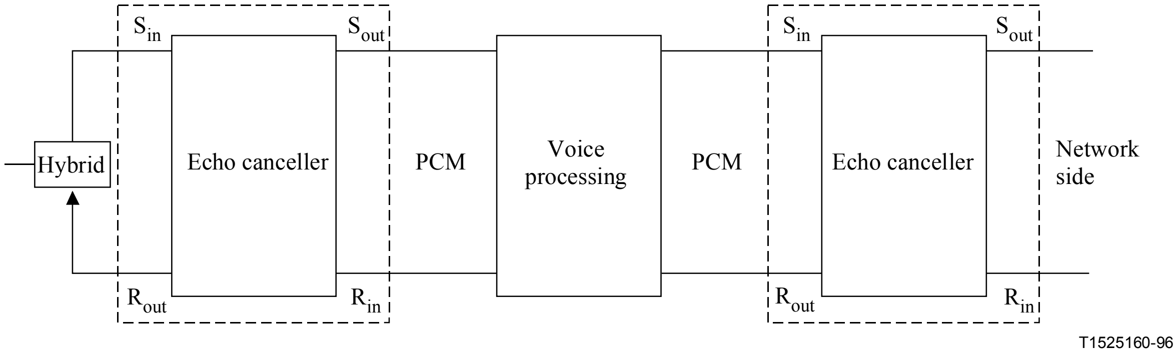



**Figure I.1/G.168 – Voice compression between tandem canceller**

### 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 ITU-T 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 NLP's, 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 [1]. 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.2a 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 possible to capture the relevant attributes of telephone connections with ECs.

As an example (see Figure I.2b), 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 $R_2$ is 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 $EC_C$ in the cellular network (in which case, the right-facing canceller of the pair might be inoperative or absent). Alternatively (see Figure I.2c), $EC_A$ 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.

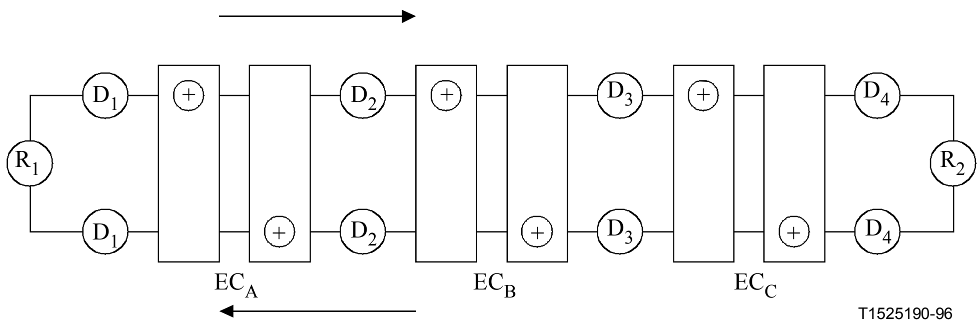



**Figure I.2a/G.168 – Reference connection for tandem ECs**

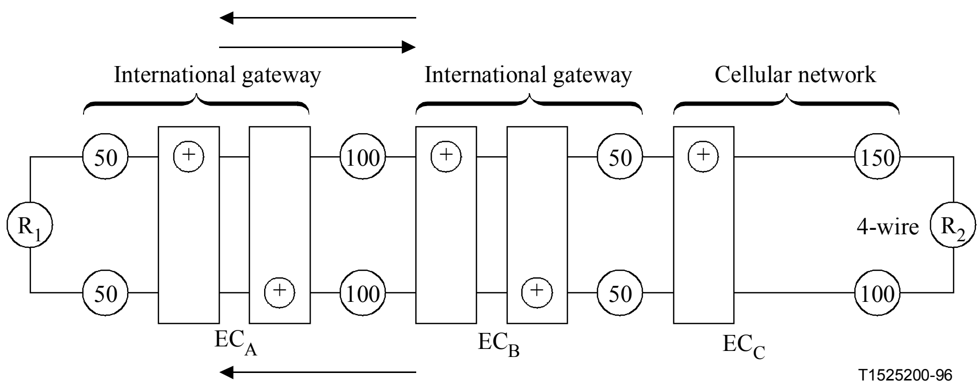



**Figure I.2b/G.168 – Example of international connection originating at analogue station and terminating in a digital cellular network**

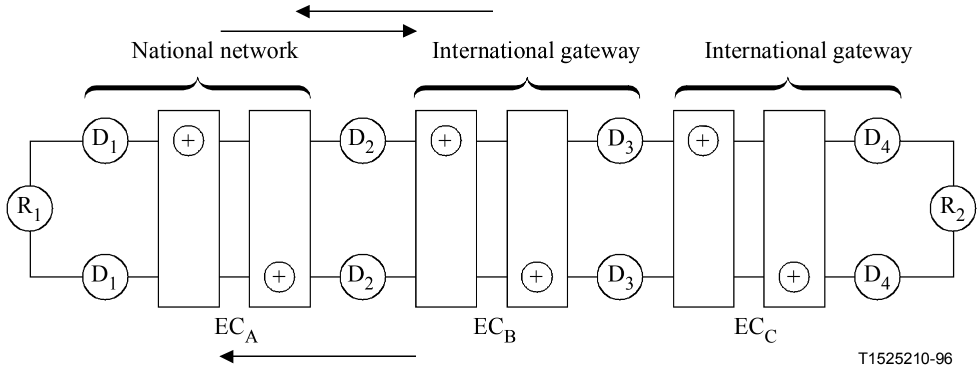



**Figure I.2c/G.168 – Example of international connection**

### 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, if the adaptation process is not 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 commonalty 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.6.6.1 Input from SG 12

Study Group 12 has published several ITU-T Recommendations on hands-free telephones:

ITU-T Recommendations P.30, Transmission performance of group audio terminals (GATS), P.340, Transmission characteristics of hands-free telephones, P.341, Transmission characteristics for wideband (150-7000 Hz) digital hands-free telephony terminals, P.342, Transmission characteristics for telephone band (300-3400 Hz) digital loudspeaking and hands-free telephony terminals.

Each ITU-T Recommendation was developed with the understanding that the terminal is responsible for controlling its own acoustic echo.

These ITU-T Recommendations present limits for different parameters like switching characteristics, Terminal Coupling Loss. These statements are reproduced in I.6.6.2.

However, a number of terminals can be found on the market which do not comply with these limits, and different terminals may have significantly different results. As the terminal technology matures we expect the number of terminals that comply with the ITU-T Recommendations to increase.

We may also be reminded of the Contribution COM 15-174 of September 1995: Test signals and other influences on the convergence behaviour of echo cancellers (GER).

Considering the possibility given by the network echo canceller to reduce the acoustic echo due to terminal equipment, it must be noted that the processing window should be at least 500 ms in order to take care of the typical room impulse response.

#### I.6.6.2 Extracts from ITU-T Recommendations related to acoustic echo control

**From ITU-T Recommendation P.30**

**3.2 Echo performance**

**3.2.1 Acoustic echo control**

To get satisfactory suppression of acoustic echoes it is necessary to provide the audio processor with either an echo canceller or an echo suppressor. The echo cancellation technology is recommended if highest possible speech quality performance is aimed at. However, it is recommended always to complement echo cancellation with a mild echo suppression, in order to prevent the undue transmission of room background noises when no talkers are active in the room. This condition should particularly be met in multi-conference environments.

**3.2.2 Echo return loss**

The echo return loss of the audio system shall be measured at reference point X of Figure 2/P.30, with the volume control in maximum position. When the electric test signal, as specified in clause 1, is applied to the input port (receive in), the level measured at the output port (send out) shall not be higher than −62 dBV.

An acoustic echo loss of 40 dB includes a margin of 5 dB in order to provide an echo return loss of 35 dB when several GATs are used in a conference situation. This value of 35 dB should be understood as a minimum value. The long-term target value for the acoustic echo loss must be considered as being 45 dB (especially, to take into account the case where a handset is connected to a hands-free terminal). This value is known to prevent any subjective degradations due to delayed acoustic echo [1] and [2]. The level measured at reference point X will then be −72 dBV.

**From ITU-T Recommendation P.341**

**6.1 Weighted terminal coupling loss (TCLw)**

The TCLw measured from the digital input to digital output shall be at least 35 dB when corrected to the nominal values of SLR and RLR as specified in 4.1 and 5.1, respectively. If a volume control is provided, the requirement applies at a setting as close as possible to the nominal value of RLR as specified in 5.1.

**From ITU-T Recommendation P.342**

**6.1 Terminal coupling loss**

**6.1.1 Hands-free function**

The weighted Terminal Coupling Loss (TCLw), in single talk operation, shall be greater than 40 dB with SLR + RLR normalized to OLR = +15 dB.

NOTE – This normalization is referred to the nominal setting of the receiving volume control.

It is assumed that this requirement is met if TCL and TCLw, respectively, meet the values of Table I.6.6.2-1/G.168, identical to Table 5/P.342 with the receive volume control in its maximum setting.

**Table I.6.6.2-1/G.168**

| TCL (Third-octave band) | TCLw |
|---|---|
| >25 dB | >35 dB |
| NOTE – These values assume no other echo control in the connection. | |

If information is available in the terminal about the one-way transmission time of the connection, and if the terminal operates in double talk, then the limits defined in Table I.6.6.2-2/G.168 may apply. X is under study.

The values in Table I.6.6.2-2/G.168, identical to Table 6/P.342, are derived from those defined in Table I.6.6.2-1/G.168. Values in brackets are under study.

**Table I.6.6.2-2/G.168**

| | One-way transmission time | TCL (Third-octave band) | TCLw |
|---|---|---|---|
| Single talk | ≤25 ms | > (18) dB | > (24) dB |
| Double talk | >25 ms | > (25 – X) dB | > (35 – X) dB |
| | ≤25 ms | > (12 – X) dB | > (18 – X) dB |

However, in order to meet G.131 [14] talker echo objective requirements, a weighted terminal coupling loss greater than 45 dB is desirable and should be striven for.

### 6.1.2 Loudspeaking function

TCL shall be greater than 25 dB and TCLw shall be greater than 35 dB.

When one-way transmission time is less than (25) ms, the TCL shall be greater than (18) dB and TCLw shall be greater than (24) dB.

If there is a voice switching device for the enhancement of the TCLw, it shall be assumed that, in double talk, the sending path of the associated handset gets priority over the loudspeaking path.

**From ITU-T Recommendation P.340**

Switching time $T_S$ – Time from one transmission direction to the other. The switching time $T_S$ should be approximately 100 ms. Limit: $T_S < 150$ ms

### I.6.7 New Circuit Switched Service

It has been suggested that there may be merit in modifying the disabling mode of G.165/G.168 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. This study is to be done in cooperation with SG 1.

### 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 pseudo-random noise during the silent interval. The second solution is to allow some of the background or idle channel noise to pass through the NLP.

NOTE – Input from SG 12 recommends the following:

- artifacts due to comfort noise insertions should be minimized;
- the noise used should match the background noise, both in frequency content and level (discussions indicate that this may not be a good idea for high and low noise levels);
- level of the inserted noise should match that of the background noise; appropriate level measurements and adjustments should be done using dBm0p;
- the time course of changes in the level of the inserted noise should match, as closely as possible, the level changes that are occurring in the background noise.

## 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.

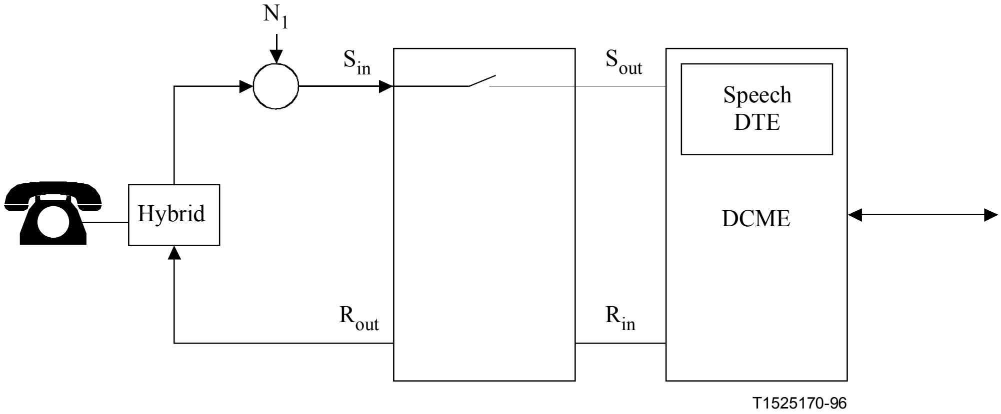



**Figure I.3/G.168 – Speech detector/echo control device interaction**

### 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 centre clipper activations in the send paths of echo control devices 1 and 3. (The role of the centre 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 centre 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.

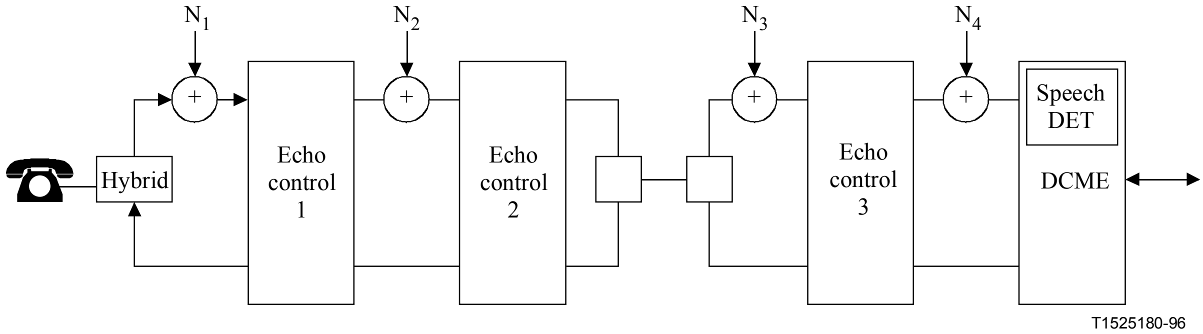



**Figure I.4/G.168 – Multiple echo control devices in a DCME/PCME network configuration**

### 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 prevents consideration of a fast phasing in of new echo cancellers.

This subject requires further study and may result in changes to ITU-T Recommendation G.165 and/or G.168 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.

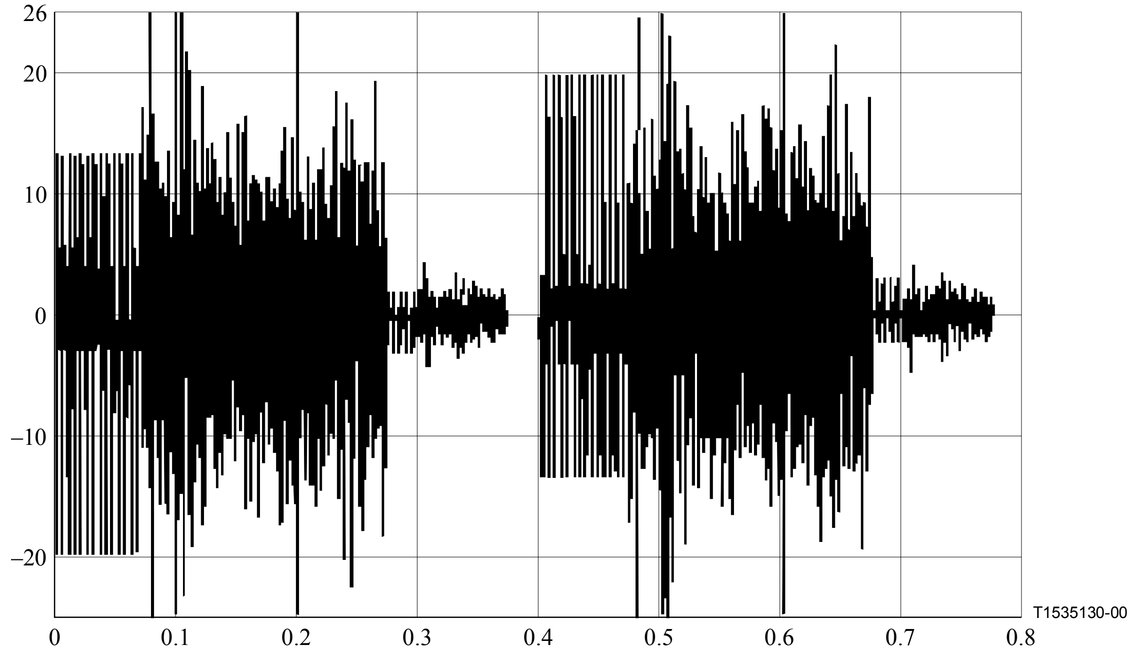



**Figure I.5/G.168 – Modified double talk signal**

### 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.
- Based on input from SG 12: Temporal clipping (i.e. syllable clipping or mutilation) introduced by the NLP for durations greater than 64 ms should always be avoided, and for durations less than 64 ms should be less than 0.1% of active speech.

### I.8.3 Analysis of Technical Parameters that Influences 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 ITU-T 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 Conducting the Double Talk Tests 3A and 3B Without Inhibiting the Adaptation

#### I.8.4.1 Introduction

In tests 3A and 3B the echo canceller is exposed to double talk for some time δt and the convergence C/divergence D is determined thereafter by removing the double talk and freezing the adaptation while the CSS remains active on the receive-in port. Thus, each experiment yields only one single point of the graph C = C(δt) or D = D(δt). Instead, one could measure the whole graph C = C(δt) or D = D(δt) in one experiment by subtracting the near-end double talk component sgen(k) from the signal e(k) at the send-out port. The difference $e_r(k) = e(k) - \mathrm{sgen}(k)$ would be the residual echo, which directly leads to C = C(δt) or D = D(δt).

#### I.8.4.2 Test Procedure

The test is performed using the test configuration of Figure I.6. For high levels of sgen(k), the magnitude of sum of sgen(k) and the echo, g(k)∗c(k), may exceed the linear range of the A/μ -law coder. As a result, the echo canceller sees the saturated value of sgen(k)+g(k)∗c(k). The double-talk component at the send-in port becomes: **sgen_sat(k) = codec[sgen(k)+g(k)*c(k))-g(k)*c(k)] – g(k)*c(k)**. Thereby the function codec(·) is defined as a linear to A/μ-law conversion followed by a A/μ-law to linear conversion. The signal sgen_sat(k) is computed by the far left blocks of Figure I.6.

Some echo cancellers contain a high- or bandpass filter in the send-path. If that is the case, sgen_sat(k) must be passed through such a filter before it can be subtracted from the send-out signal. This can be achieved in various ways. If the filter is known, it appears simplest to pass sgen_sat(k) directly through the filter. If it is unknown, one can pass sgen_sat(k) through the echo canceller while there is silence on the receive-in port (see Figure I.6). The obtained signal sgen_sat_filt(k) represents the double talk component of the send-out signal, and can be subtracted to compute the residual echo $e_r(k)$.

The requirements on the power level of $e_r(k)$ should be the same as on $L_{RES}$. This procedure provides a continuous (rather than discrete) view of the echo canceller behavior throughout the double-talk course. Hence, it does not require freezing the echo canceller adaptation, and it eliminates the possibility of a potential distortion of the echo canceller measured performance. Note that the $S_{out}$ signal is measured during the application of the double talk signal, $S_{gen}$.

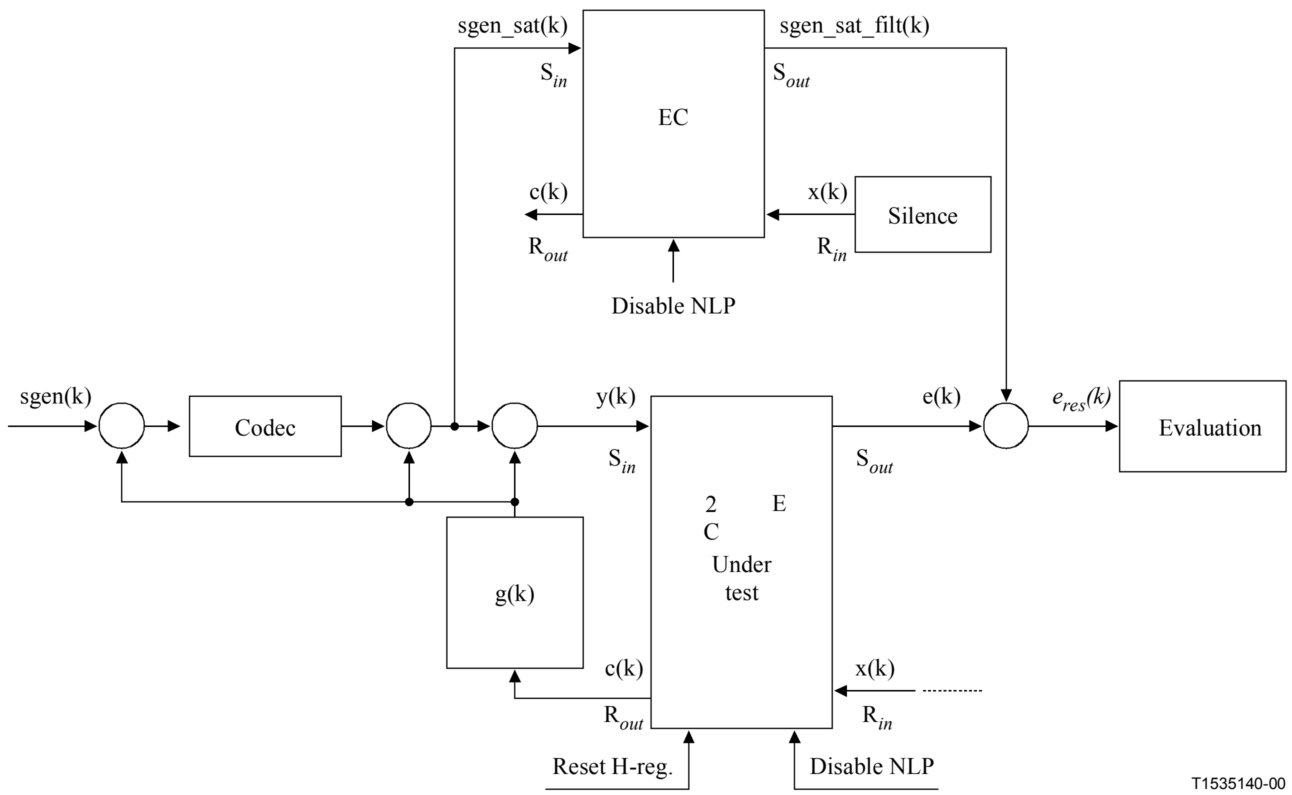



**Figure I.6/G.168 – Test to subtract the double talk component from the send-out signal**

### I.8.5 Subjective and Objective Echo Canceller Testing

#### I.8.5.1 Subjective and Objective Echo Canceller Testing – Fundamentals

One of the most important points in studying the performance of echo cancellers is to investigate how objective tests correlate with subjective tests. The subjective test procedures (conversational test, talking and listening tests, listening only tests) were suggested for standardization in ITU-T Recommendation P.831. Based on the data from subjective tests parameters determining the transmission quality for speech echo cancellers can be identified and in a second step conclusions can be drawn about important tests.

The following paragraphs summarize how subjective tests results were used in order to develop objective tests.

**Auditory Test Procedures**

Figure I.7 demonstrates the structure and relation of the different subjective test procedures like *conversational tests, talking and listening tests and listening only tests* together with parameters which can be assessed by these tests. The different procedures were developed as parts of the whole test setup. The procedures were not performed in isolation but each test for a certain purpose.

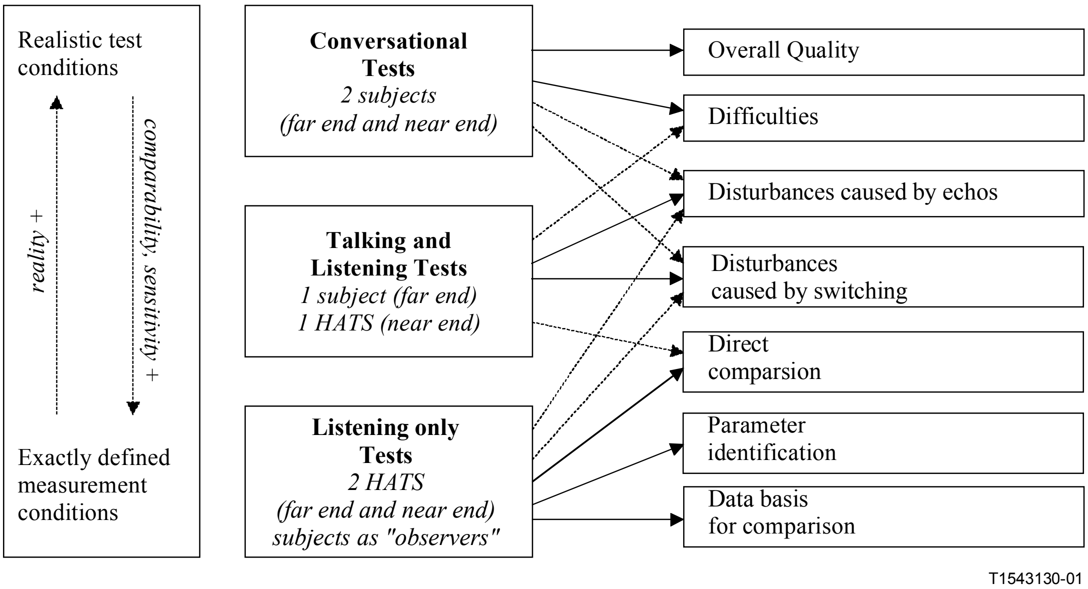



**Figure I.7/G.168 – Structure of subjective tests procedures for speech echo cancellers and a selection of parameters**

*Conversational tests* play an important role in the evaluation of speech echo cancellers, since the performance of the echo canceller influences the perceived quality of the overall telephone connection. During a conversational test, subjects give their rating of the quality of the *overall* connection after a *complete* conversation, thus taking into account all of the aforementioned performance factors. Conversational tests are therefore probably the most important tests designed to evaluate the influence of speech echo cancellers on overall transmission performance.

Furthermore the conversational tests are necessary to identify those parameters, which play an important role concerning the complex parameter "*overall quality*". But on the other hand, such tests are not sensitive enough to evaluate specific transmission performance characteristics. If laboratory tests shall be developed and requirements for specific transmission parameters for echo cancellers shall be fixed (residual echo level, initial convergence, NLP behavior, a.m.) other test procedures are necessary. For this purpose *talking and listening* tests and *listening-only* tests have been developed. If these tests are properly designed, they provide a powerful and effective method to investigate the effect which specific echo canceller parameters have on overall transmission performance. The results of these tests may therefore be directly correlated to the results of objective measurements.

**Parameter Identification by Conversational Tests**

In 1996, conversational tests were carried out with 4 commercially available echo cancellers being at least G.165 compliant. After the test was finished, the operators who were present in the test rooms interviewed subjects about the *overall quality rating* and the *percentage difficulty in talking or listening over the connection (% D)*. The echo path characteristics and the corresponding test results are given in Figures I.8 and I.9.

In addition to these recommended questions, all subjects who answered the question about difficulties with "Yes" where additionally asked about *the kind of difficulties* they had. At the end of the interview each subject was asked about *the most annoying aspect* during the call they just had finished. The answers are of special interests for the operators not only for the subjective tests itself but also for the evaluation of those technical parameters which cause the problems for the subscribers. This information gives an important idea about the relevance of instrumental parameters.

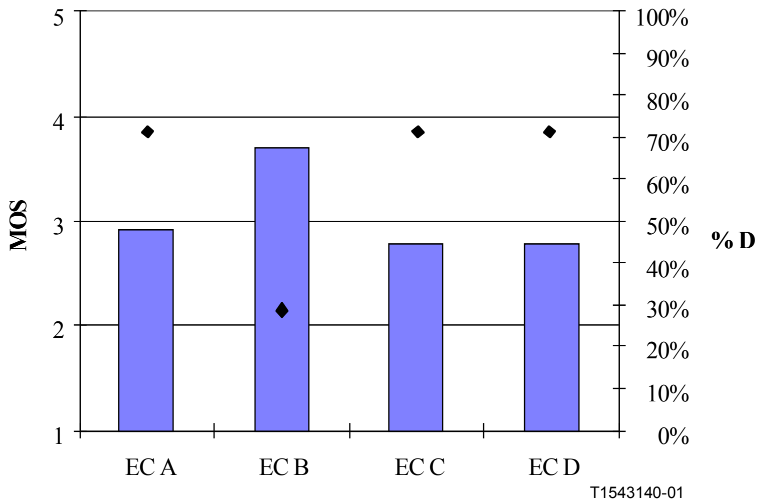



**Figure I.8/G.168 – Results from *conversational tests*, overall quality MOS and % D (rhombus) for the 4 echo cancellers, ERL 7 dB, room noise level 40 dB(A), corresponding level –61 $dB_{m0}$**

A significant difference between echo canceller B and the other cancellers A, C and D could be analyzed in Figure I.8. The main focus of these tests was not to compare different echo canceller but to identify the parameters that determine the subjectively perceived quality. If the comments about the difficulties during the conversation are analyzed, most of these quotations concern the following points:

- *Audible speech clipping during double talk (implementation of the NLP)*

  Instrumental measurements based on the Composite Source Signals demonstrate that echo canceller B in Figure I.8 shows a very good double talk performance. In connection with a high echo attenuation this leads to the best rating given through Figure I.8.

- *Disturbances caused by echoes (initial convergence and residual echo)*

  The echo signal of EC D itself was typically characterized as 'distorted' or '...like whispering'. This leads to significantly worse MOS values compared to echo canceller B although the double talk performance was good.

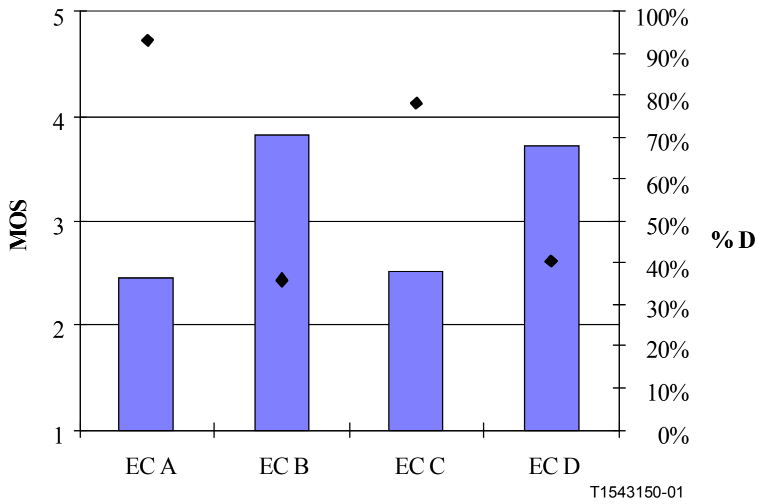



**Figure I.9/G.168 – Results from *conversational tests*, overall quality MOS and % D (rhombus) for the 4 echo cancellers, 4 wire echo path, room noise level 50 dB(A), corresponding level –55 dB$_{m0}$**

The echo cancellers can be divided into two groups in Figure I.9. Both cancellers A and C are judged significantly worse than the cancellers B and D. Analyzing the comments of subjects after the tests expressing their rating of percentage difficulty, it clearly points two important aspects:

- *Clipping during double talk as the most annoying impairment*

  Both echo cancellers B and D with good double talk performances in connection with this kind of echo path realization are judged significantly better.

- *Quality of background noise transmission*

  The modulation of background noise is audible and annoying for subjects specially if the noise disappears when they start talking and the echo canceller attenuates the send path.

Conversational tests conducted by another Administration under different network conditions also pointed out the importance of the NLP performance. Significant differences between two of the echo cancellers under test appeared in these evaluations, when the signal levels were either above or below nominal. In one condition, it was concluded that the difference in performance was due to the clipping of speech in time, caused by operation of the non-linear processor and its associated control.

The results of conversational tests published by two Administrations, and corresponding results of specific listening only tests under double talk conditions, indicate that the occurrence of clipping and gaps is very important to subjective performance. Objective measurements, concerning the switching characteristics of the NLP, give correlated parameters. The results of the different subjective test procedures clearly indicate that the control of the NLP and the associated switching characteristic under single and double talk conditions is one of the most important parameters. This influences the quality of background noise transmission and double talk performance. In addition, the echo attenuation during initial convergence and steady state conditions (under single and double talk conditions) is important.

**Specific Talking and Listening Tests**

In the conversational tests, complaints were made concerning the echo disturbances for some of the echo cancellers under test. This indicates the need for a specific subjective test and corresponding objective test results to fix requirements for laboratory tests. For this purpose, specific *talking and* *listening tests* were designed to evaluate talking related disturbances. Four echo cancellers were tested.

One part of this test considered the initial convergence of the echo cancellers, both with enabled and disabled NLP. The results of the subjective tests and correlating objective measurements are given in the following Figures.

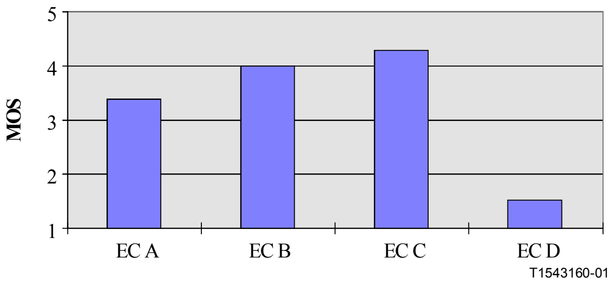



**Figure I.10/G.168 – Results from *talking and listening tests*, MOS, disturbance caused by echoes, digital echo path, ERL 6 dB, no near end background noise, NLP disabled**

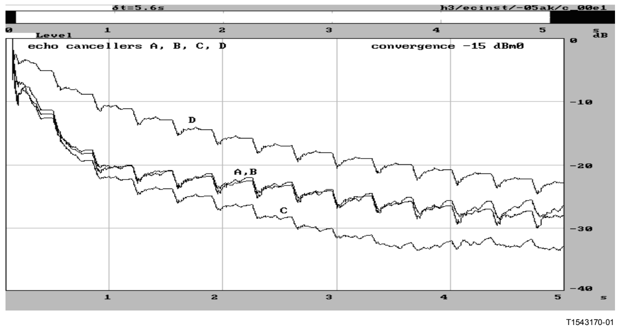



**Figure I.11/G.168 – Convergence test versus time, digital echo path, ERL 6 dB, no near end background noise, receive level –15 dB$_{m0}$, NLP disabled**

If the NLP is disabled, the results are given in Figure I.10. This test characterizes the convergence of the adaptive filters. The *talking and listening tests* are sensitive enough to point out differences between the tested echo cancellers. These differences could be expected from the test persons comments during the conversational tests but the conversational tests were not sensitive enough for a further comparison between the tested echo cancellers. The echo signal, which is produced by EC D causes the highest annoyance. These MOS values correlate to objective test results measured for the same echo cancellers as given in Figure I.11.

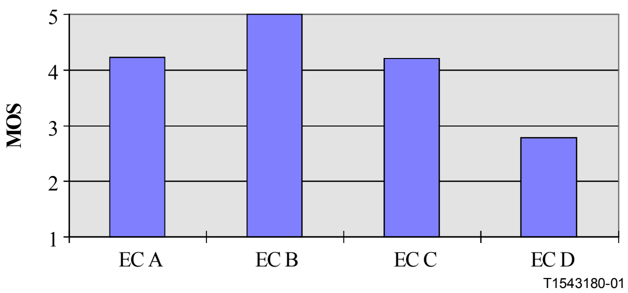



**Figure I.12/G.168 – Results from *talking and listening tests*, MOS, disturbance caused by echoes, digital echo path, ERL 6 dB, background noise –55 dB$_{m0}$, NLP enabled**

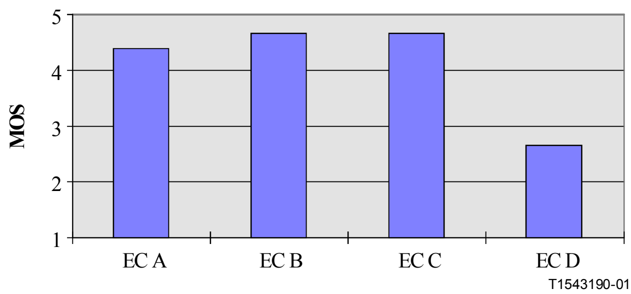



**Figure I.13/G.168 – Results from talking and listening tests, MOS, disturbance caused by audible switching, digital echo path, ERL 6 dB, background noise –55 dB$_{m0}$, NLP enabled**

Figures I.12 and I.13 give the results, if the NLP is enabled. Figure I.12 shows the ratings for the echo disturbances, if the NLP is disabled and Figure I.13 the corresponding results for audible clipping. EC D obtains the worse ratings with less than 3 points for both parameters. Echoes can be heard for a longer time until they are suppressed (Figure I.12). Audible switching causes a higher annoyance compared to the other three echo cancellers EC A, EC B and EC C (Figure I.13). The switching characteristic of EC D is more annoying compared to the echo cancellers.

Basically, several parameters cause the annoyance at the beginning of convergence: The initial convergence speed, determined as the echo attenuation vs. time is one important aspect as given through the correlated subjective and objective test results in Figures I.10 and I.11. The switching characteristics of the NLP and the echo attenuation vs. frequency influence the annoyance too. Correlated objective measurement results are given in the following Figures I.14 to I.17. The pictures show a spectral analysis of the echo signal versus time (x-axis) and frequency (y-axis) during initial convergence for the echo cancellers which were tested subjectively (Figures I.12 and I.13). High echo signal peaks are given in light colors, dark colors represent a better echo attenuation. These measurement results are again a very good example how subjectively obtained test results can be correlated to objective analyses. The tests clearly point out that the echo cancellers A, B and C which were judged significantly better in the talking and listening tests (Figures I.12 and I.13), attenuate and suppress the residual echo faster than echo canceller D.

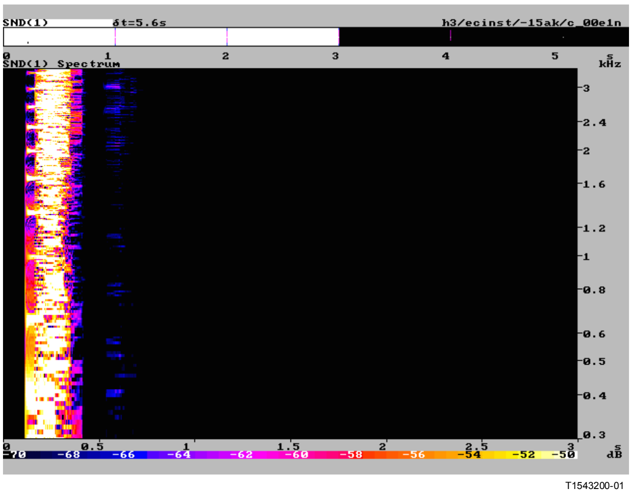



**Figure I.14/G.168 – Spectral analysis of residual echo during initial convergence: Echo cancellers A**

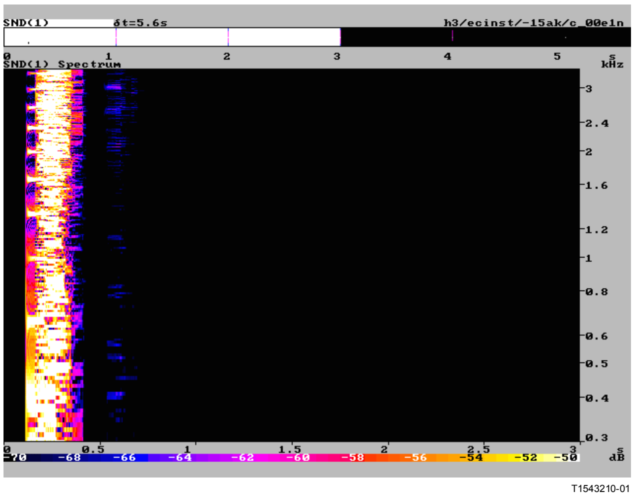



**Figure I.15/G.168 – Spectral analysis of residual echo during initial convergence: Echo cancellers B**

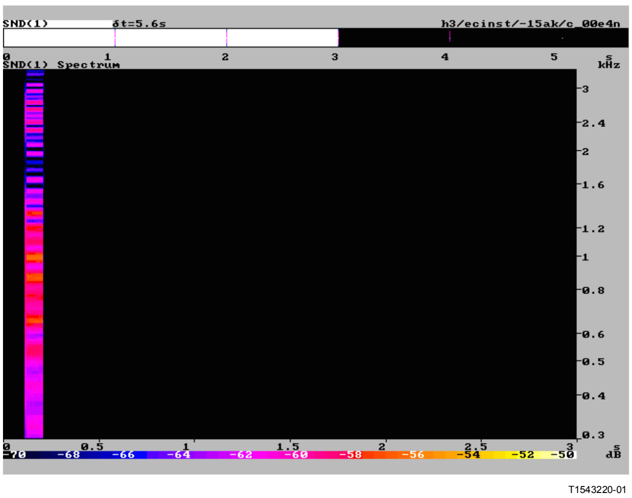



**Figure I.16/G.168 – Spectral analysis of residual echo during initial convergence: Echo cancellers C**

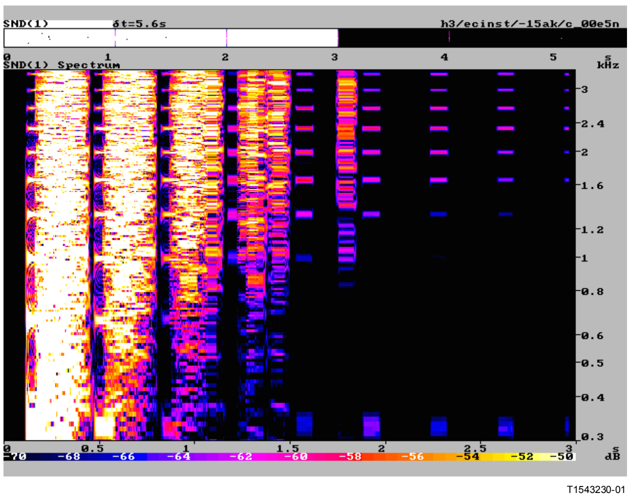



**Figure I.17/G.168 – Spectral analysis of residual echo during initial convergence: Echo cancellers D**

### Listening-only Tests

A listening-only test is probably the most sensitive test method for the direct comparison of echo cancellers and for the evaluation of single transmission parameters. In addition to conversational tests and talking and listening tests, this test method was used for the evaluation of specific transmission parameters like residual echo levels, initial convergence or double talk performance.

The following Figures I.18 and I.19 show one example how the subjective test results were used to find values for laboratory tests. The double talk performance of six echo cancellers was accessed by a group of untrained subjects (Figure I.18) and experts (Figure I.19) for an ERL of 24 dB, a receive level of –15 $dB_{m0}$ and a double talk level of –30 $dB_{m0}$.

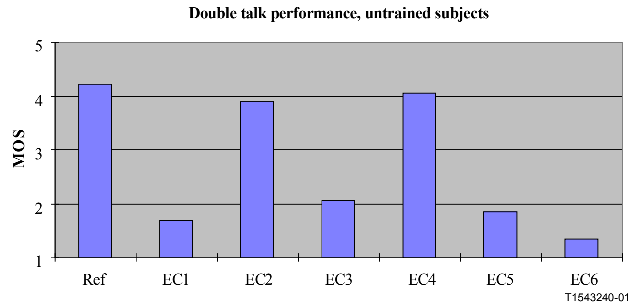



**Figure I.18/G.168 – Results from *listening only tests*, untrained subjects, double talk performance MOS, ERL 24 dB, receive level –15 $dB_{m0}$, double talk level –30 $dB_{m0}$**

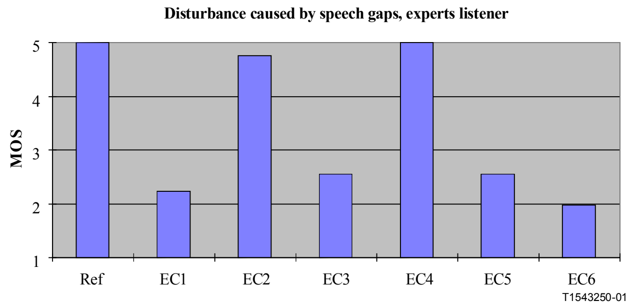



**Figure I.19/G.168 – Results from *listening only tests*, experts, disturbances during double talk caused by speech gaps MOS, ERL 24 dB, receive level –15 $dB_{m0}$, double talk level –30 $dB_{m0}$**

Although the double talk signal level is very weak, the two echo cancellers EC2 and EC4 reach comparable MOS values like the undisturbed reference in Figure I.18. The untrained subjects assessed the parameter 'double talk performance'. This demonstrates the sensitivity of listening-only tests and indicates that, even under these conditions, a good transmission quality can be expected for echo cancellers. The other four speech echo cancellers are judged significantly worse. The ratings of the experts group for the disturbances caused by speech gaps (given in Figure I.19) show the same order of preference of the echo cancellers. This result again correlates to the conversational test result and the corresponding comments. It demonstrates, that the annoyance caused by speech gaps seems to be the most important parameter.

Objective tests and test signals could be adapted to measure the correlated parameters. The tests pointed out that the degradation of a transmitted double talk signal is mainly determined by the insertion of the NLP during continuous speech. The correlated parameter is the switching characteristic of the NLP and its control during double talk.

### I.8.5.2 Subjective tests with the purpose of qualifying those effects of the echo cancellation process that cannot be captured by objective measurements

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.

### I.9 Guidelines on the use of parameters for testing echo cancellers

The tests in this ITU-T Recommendation have many parameters which are specified as ranges. To provide some guidelines on a set of test conditions which should be considered at a minimum, the following is suggested. It is strongly cautioned that these values do not imply a sufficient set for compliance to this ITU-T Recommendation, which is left to the discretion of the telecommunications providers. Also, it should be noted that all possible combinations of these parameters listed below would be time consuming.

1) *Echo path delay:*
   a) Δ - dispersion of g(k) – 4
   b) Δ/2 – 4
   c) Δ/10
2) *Echo return loss:*
   a) 6 dB
   b) 15 dB
   c) 30 dB
3) *Receive-in level* ($L_{Rin}$)*:*
   a) Maximum specified input level for the particular test
   b) −10 dBm0
   c) −20 dBm0
   d) Minimum specified input level for the particular test
4) *Initial Condition of H Register:*
   a) H register reset
   b) H register converged to an open echo path
   c) H register converged to some g(k).

## APPENDIX II

## Measurement Methods for Characteristics of Echo Paths and an example for the North American Network

### II.1 Introduction

Knowledge of an echo path is necessary in order to control voice echo efficiently in telecommunication systems. Two of the main characteristics of an echo path are the dispersion time and frequency response. This appendix summarizes the measurement techniques used to obtain the echo-path impulse responses in telephone networks. It also provides analysis results for the echo paths measured in North American networks.

The appendix is organized as follows. Subclause II.2 gives details on the measurement procedure. Subclause II.3 describes the test signal. The computation of impulse response from the signal measurements is illustrated in subclause II.4. The generation of echo-path characteristics is in subclause II.5. Subclause II.5 also contains the characteristics of echo paths measured in North America. Finally, subclause II.6 is the conclusion.

## II.2 Measurement Procedure

Figure II.1 shows a typical call configuration where $x(k)$ and $y(k)$ are, respectively, the signals to and from the cancelled end. Because of the impedance mismatch in the hybrid that connects the 2-wire and 4-wire circuits, part of the signal from the non-cancelled end will leak through the hybrid and propagate back to the talker to form an echo. If the user in the cancelled end is not talking, $y(k)$ will be the echo signal. We therefore measure the signals $x(k)$ and $y(k)$ to obtain the echo-path characteristics.

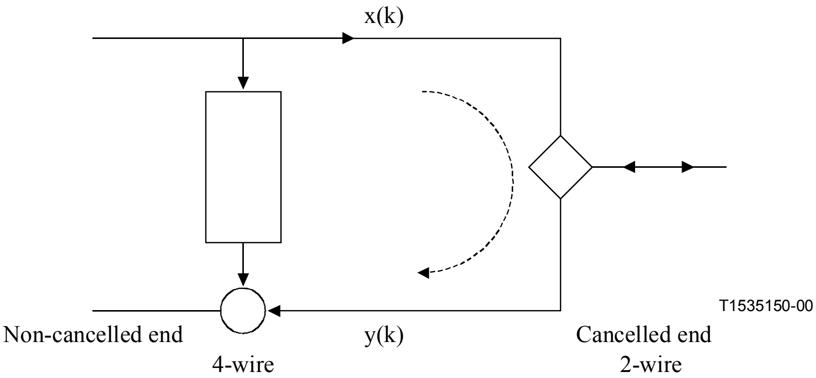



**Figure II.1/G.168 – A typical call connection**

Figure II.2 is the block diagram of the measurement setup. The equipment on the left-hand side generates the test signal and performs the signal recording. The equipment uses a 4-wire connection for sending and receiving signals. Through a T1 interface, the 4-wire port in the equipment is directly connected to a central office (CO). Apart from the distant section between a CO and the called user, the signal transmission path is all 4-wire and no 2-wire circuitry is expected to be involved.

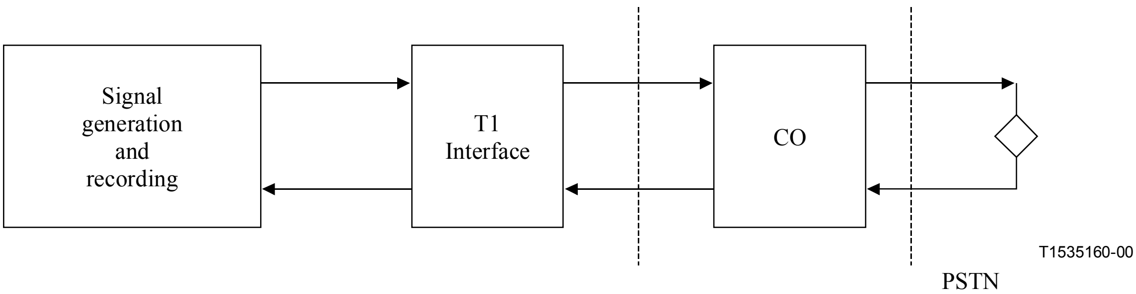



**Figure II.2/G.168 – Block diagram of measurement setup**

When a T1 interface is not available, the test signal could be generated at a 2-wire point. In any case, the signals (transmit and receive) should be measured at the 4-wire digital point where the signals are well defined (in terms of level).

During measurement, the equipment first dials a telephone number. It then sends out the test signal after the call is established. The test signal goes through the T1 interface, the CO and the PSTN to a distant user. Part of the transmitted signal is reflected back in the form of an echo. Both the transmitted signal and the returned echo are recorded in the equipment using a synchronous recorder. The two recorded signals are the $R_{in}$ and $S_{in}$ signals for an echo canceller.

### II.3 Test Signal

The test signal used in this exercise is shown in Figure II.3. It consists of three segments. The first segment is a 2100 Hz tone with phase reversal. The purpose of the tone is to disable all the echo cancellers and echo suppressors in the link during measurement. The duration of the tone is set to $T_1 = 1.35$ s. The amplitude of the tone is –12 dBm0. The second segment is a pause. The purpose of it is to obtain the background noise characteristics in the returned echo signal. Since a tone-disabled echo canceller, as specified in clause 7 and ITU-T Recommendations G.164, and G.165, will enable itself within 250 ms ± 150 ms when the signal level is below –39 dBm0, the duration of the pause is set to $T_2 = 80$ ms. The third segment is a Gaussian white noise signal which is used to identify the echo-path impulse response. The power level of the white noise signal is –18 dBm0 and its duration is $T_3 = 5$ s.

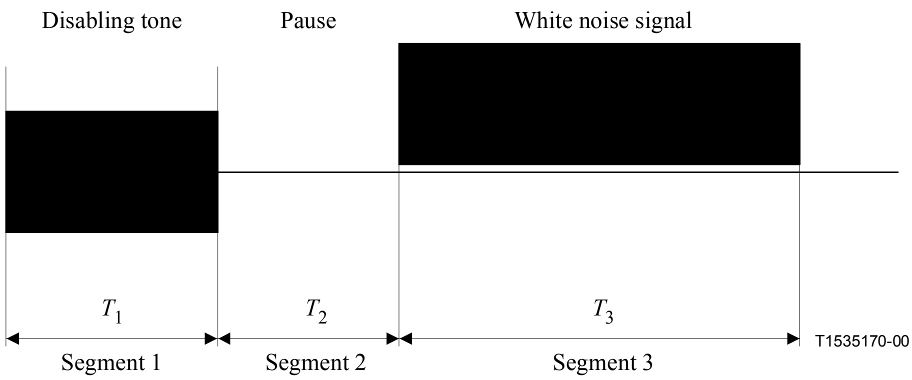



**Figure II.3/G.168 – Test signal for echo-path measurement**

### II.4 Computation of Impulse Response

$$y(n) = x(n) * h^\circ(n) + v(n) = \sum_{i=0}^{N-1} h^\circ(i) x(n-i) + v(n), n = 0,1,2,\ldots, L-1$$

where $N$ is the length of the echo-path impulse response and $L$ is the total number of data samples available. Define vectors

$$\boldsymbol{h}^\circ = \left[h^\circ(0), h^\circ(1), \ldots, h^\circ(N-1)\right]^T$$

and:

$$\boldsymbol{x}(n) = [x(n), x(n-1), \ldots, x(n-N+1)]^T$$

In vector notation,

$$y(n) = x(n)^T \boldsymbol{h}^\circ + v(n), n = 0,1,2,\ldots, L-1$$

The estimation problem can be stated as follows: Given $L$ samples of $x(n)$ and $y(n)$, compute the echo-path impulse response $\mathbf{h}^o$. This is a standard system identification problem which can be solved by two common approaches. One is the least-squares (LS) method and the other is the iterative method using the normalized least-mean square (NLMS) algorithm. The first method yields a good solution with short data record. However, it is more computationally intensive than the second method. In most cases, provided that the data length $L$ is large enough, the final solutions from both methods are very close and the difference is insignificant.

**LS Method**

Define error signal

$$e(n) = y(n) - \boldsymbol{h}^T \boldsymbol{x}(n)$$

where $\mathbf{h}$ is the echo-path impulse response estimate. The LS method minimizes

$$J = \sum_{n=0}^{L-1} e(n)^2$$

to determine $\mathbf{h}$. The corresponding solution can be shown to be

$$\boldsymbol{h} = \boldsymbol{R}^{-1}\boldsymbol{p}$$

where:

$$\boldsymbol{R} = \sum_{n=0}^{L-1} \boldsymbol{x}(n)\boldsymbol{x}(n)^T$$

$$\boldsymbol{p} = \sum_{n=0}^{L-1} y(n)\boldsymbol{x}(n)$$

**NLMS Method**

The NLMS method finds $\mathbf{h}$ interactively using the following equations:

$$e(n) = y(n) - \boldsymbol{h}(n)^T \boldsymbol{x}(n)$$

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \frac{\mu}{\delta + x(n)^T \boldsymbol{x}(n)} e(n)\boldsymbol{x}(n)$$

where $\mathbf{h}(n)$ denotes the echo-path estimate at time $n$, $\delta$ is a small positive number and $\mu$ is a positive constant called step-size. The step-size must be less than 2 to allow convergence.

## II.5 Analysis of Echo-Path Characteristics

This subclause describes methods of echo-path impulse response analysis. The echo-path characteristics considered are dispersion width and magnitude response of echo paths. Also included in this subclause are the results for the echo-path measurements made in North America.

Before the analysis of echo-path characteristics, the measured echo-path impulse responses should be post-processed by a bandpass filter which covers the same frequency region as a speech signal. The passband of the filter is from 200 Hz to 3600 Hz, the same frequency range as that of the CS signal specified in C.3.1/Annex C.

### II.5.1 Dispersion Time

The echo-path impulse response is not a single impulse. It has a finite duration which we shall refer to as dispersion time. The dispersion time is a factor to determine the length of the H register in an echo canceller. A region where the echo reflection occurs is first located from an impulse response measurement. The duration for this region should be long enough to cover the entire echo reflection. In our measurement, it was set to be 30 ms. This corresponds to 240 samples at 8 kHz sampling rate. The impulse response in this range is denoted as h(n), n = 0, 1,…, M, where M = 239. If we truncate h(n) by keeping the impulse response values from n = $N_1$ to $N_2$, then the relative square-error due to truncation is:

$$ERR = 1 - \frac{\sum_{n=N_1}^{N_2} h(n)^2}{\sum_{n=0}^{M} h(n)^2}$$

Figure II.4 shows such a possible truncation. According to Figure II.6, we need the largest amount of echo reduction at 0 dBm0 $L_{R_{in}}$ . With 6 dB ERL, this translates to 30 – 6 = 24 dB ERLE provided by an echo canceller. Hence the dispersion time is considered to be the shortest length $N_2 - N_1 + 1$ such that the *ERR* is less than –24 dB.

Note that the method suggested here to compute the dispersion time is for the worst case scenario. This is because a 6 dB ERL is assumed to compute the *ERR* threshold. In practice, the ERL is much higher. The dispersion time is shorter when ERL increases.

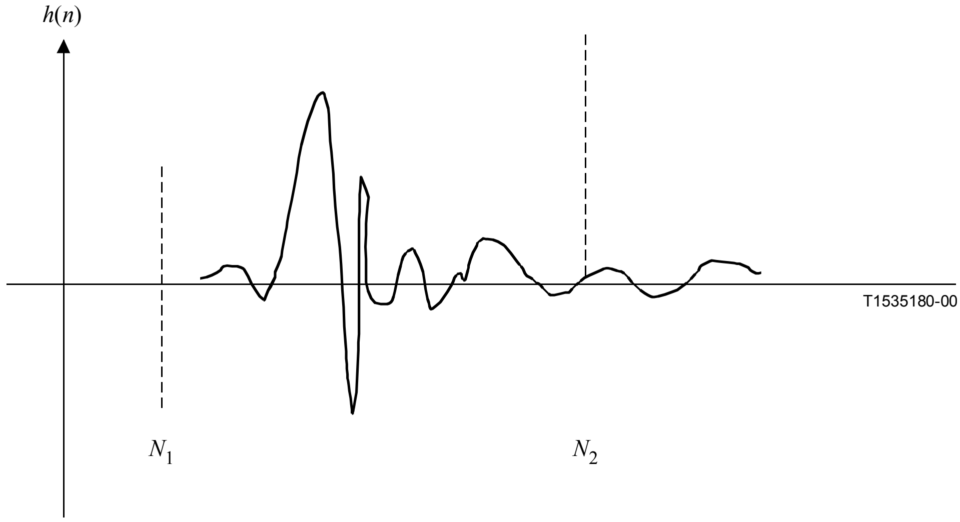



**Figure II.4/G.168 – Echo-path truncation for dispersion time estimate**

### II.5.2 Magnitude Response of Echo Path

While the echo-path impulse responses can be quite different in shape, the frequency domain characteristics of the echo paths are expected to have some similarity. The magnitude response characteristics of echo path can be generated by taking the average of the magnitude of the Fourier transform of each echo-path measurement.

### II.5.3 Echo-Path Characteristics from Measurements in North America

During the period from June 1998 to April 1999, 101 long-distance calls were made from Montreal to the following Provinces and States across North America: Arizona, British Columbia, California, Louisiana, Manitoba, Massachusetts, Michigan, Minnesota, Missouri, Nevada, New York, North Carolina, Ontario, Quebec, Saskatchewan, Texas and Wisconsin. The send-out and returned signals were recorded in each call and the echo-path impulse responses were computed using the method described in II.4. This subclause reports the echo-path characteristics generated from the above echo-path measurements.

### II.5.4.1 Dispersion Time

Figure II.5 is the histogram of the echo-path dispersion time. The largest percentage of dispersion time was between 5-7 ms. Only two calls had dispersion time between 11-12 ms. There was no call with dispersion time over 12 ms.

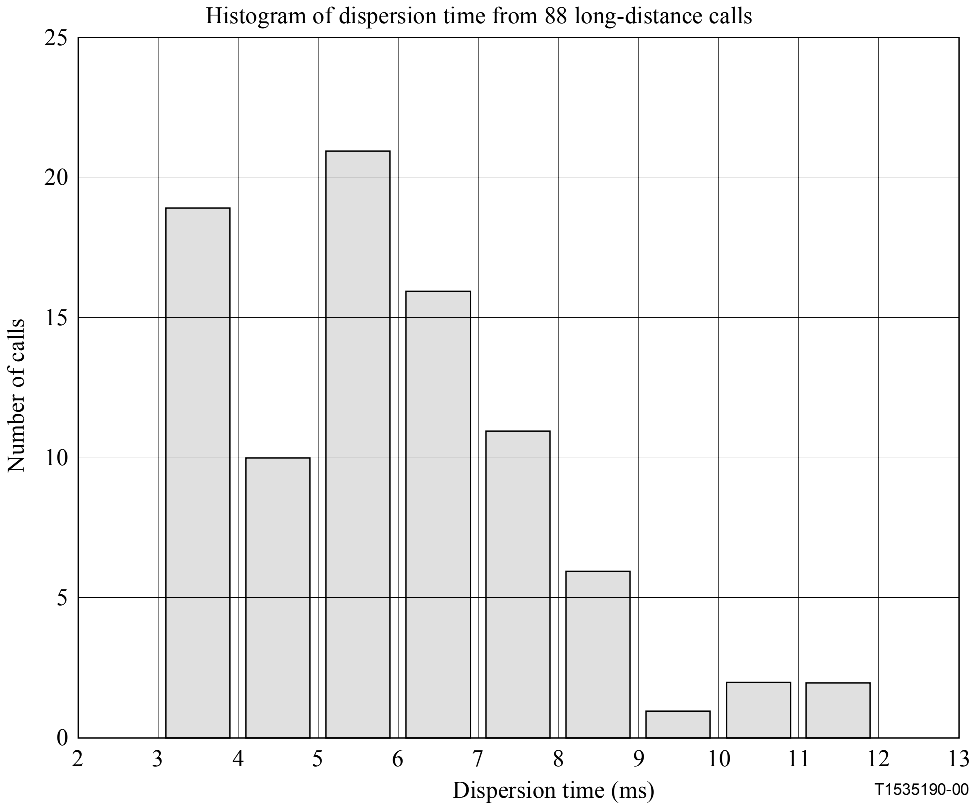



**Figure II.5/G.168 – Histogram of dispersion time for long-distance calls.
Mean = 6.02 ms, StD = 2.26 ms**

### II.5.4.2 Magnitude Response of Echo Path

Figure II.6 is the average of the magnitude spectra of the measured echo-path impulse responses. The impulse responses were normalized to have unit energy before the spectra were computed. The solid line is the mean and the two dotted lines represent the one standard deviation region. It appears that the variance of magnitude spectra is not large. This indicates that there is a high consistency among the magnitude responses in the echo-paths. The averaged magnitude spectrum has a small peak around 250 Hz. In addition, the magnitude responses are relatively flat.

The results are consistent with the study in the article "Echo Performance of Toll Telephone Connections in the United States" by F.P. Duffy, G.K. McNees, I. Nasell, and T.W. Thatcher, Jr., in Bell System Technical Journal, 1974.

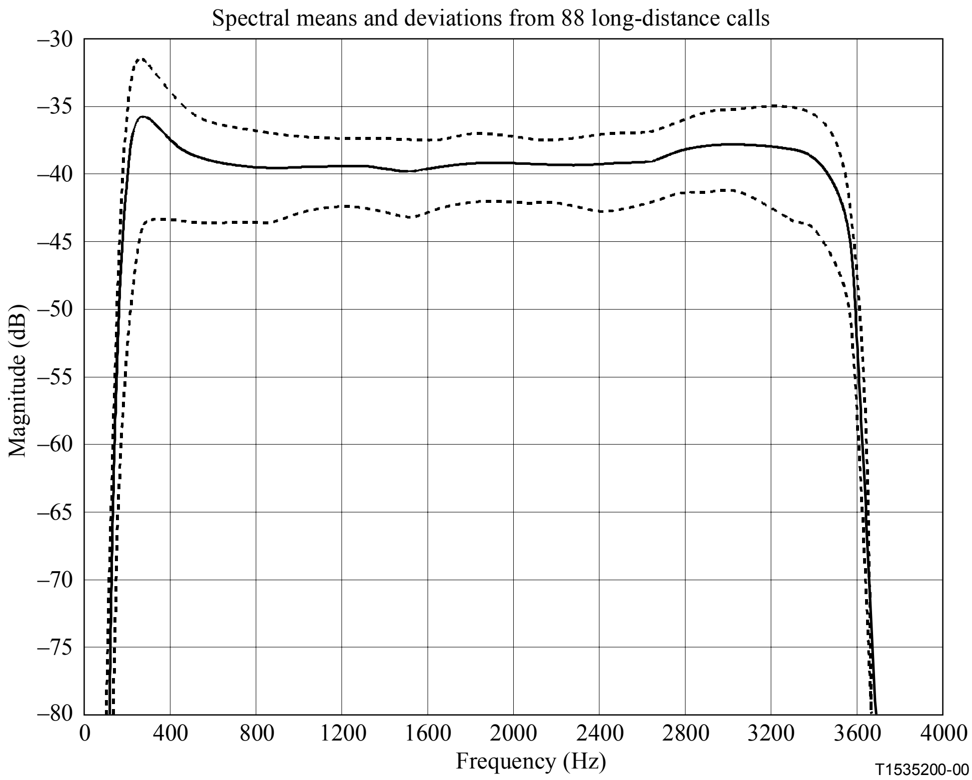



**Figure II.6/G.168 – Magnitude spectra of echo paths for long-distance calls, echo paths normalized to have unity energy**

### II.5.4.3 Multiple Reflections

We have observed 6 double reflections from the 101 long-distance calls. However, the total dispersion times of these double reflections were all limited to 10 ms. In other words, the two reflectors in each of these cases were closely located.

The number of reflections did not exceed two in the measurements.

## II.6 Conclusions

This appendix summarized the measurement procedures together with the computation of echo-path impulse response and its characteristics. Also included are the results from a number of echo paths measured in North America. The results can serve as references in designing a digital echo path for the testing of echo cancellers in this ITU-T Recommendation. Specifically, the following important properties of echo paths are observed:

– The dispersion time of an echo path is within 12 ms.

– The magnitude response of echo path is relatively flat, with a small peak around 250 Hz.

– On some occasions, double reflections may occur. Three or more reflections, however, rarely occur.

## APPENDIX III

## Multiple Tail Circuits

In modern networks, it may be the case that a two-party call is modified after the call is initially set up, so that one or more additional parties can participate in the conversation, as illustrated in Figure III.1.

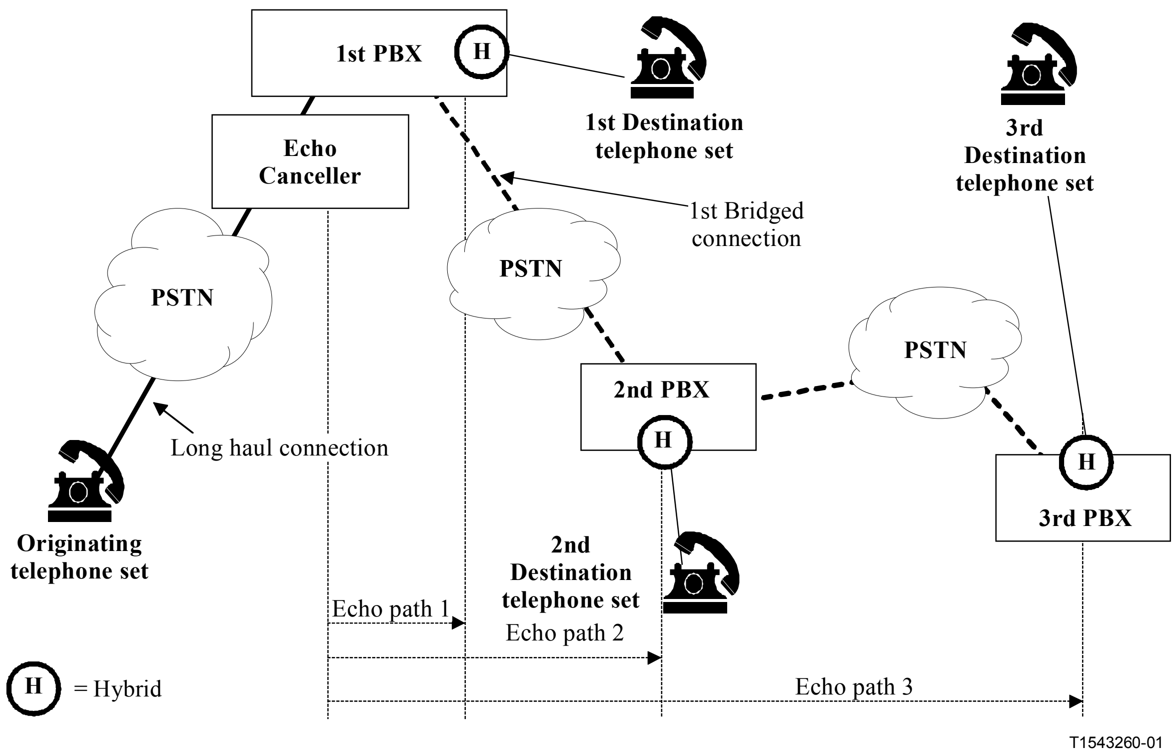



**Figure III.1/G.168 – Multiple tails in a typical bridged telephone call**

In Figure III.1, the originating telephone set initiates a call to the 1st destination set. The echo canceller sees the hybrid associated with the 1st PBX, illustrated as echo path 1. The recipient at the 1st PBX then bridges in a 2nd destination set by using the bridging function found on all modern PBXs. The echo canceller now sees the second hybrid, appended to the first, and delayed in time by the network delay between PBX 1 and 2. This is illustrated as echo path 2. The second destination may well bridge in a third destination, adding another hybrid tail associated with the hybrid in the 3rd PBX, and delayed by the sum of the network delay between PBX 1 and 2, and PBX 2 and 3. This is illustrated as echo path 3.

PBXs may not have any echo cancellation built into them, even though they perform this bridging function. This requires that the network echo canceller be able to support multiple tails up to the echo tail capacity of the canceller.

Network operators and administrators should take this into account when testing echo cancellers for use in their network.

## SERIES OF ITU-T RECOMMENDATIONS

| | |
|---|---|
| Series A | Organization of the work of 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 communications |
| Series Y | Global information infrastructure and Internet protocol aspects |
| Series Z | Languages and general software aspects for telecommunication systems |

Geneva, 2001