ETSI TS 102 929 V1.1.1 (2011-04)

Technical Specification

Speech and multimedia Transmission Quality (STQ); Procedures for the identification and selection of common modes of de-jitter buffers and echo cancellers

Reference

DTS/STQ-00179

Keywords

QoS, jitter buffer

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Introduction

The present document describes the characteristics of a jitter buffer, including the requirement for in-band tone activating and other control mechanisms. Measurements for verifying the correct setting of jitter buffers and echo cancellers are currently not included in the present document, but they will be added in a next revision.

1 Scope

Jitter buffers and echo cancellers have a major effect on voice and data transmission quality in telecommunication networks. They affect the transmission of voiceband, data, fax, text telephones and transmission of unrestricted digital information (UDI).

Since the requirements for the settings of jitter buffers and echo cancellers differ for different services, the present document describes the activation and mode switching procedures of jitter buffers and echo cancellers, including the requirement for in-band tone activating and other control mechanisms.

The present document does not apply for fax transmissions with ITU-T Recommendation T.38 [i.8].

It is understood that the clock accuracy of all elements involved is sufficiently high for application of the present document.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

[1] ITU-T Recommendation V.8 bis: "Procedures for the identification and selection of common modes of operation between data circuit-terminating equipments (DCEs) and between data terminal equipments (DTEs) over the public switched telephone network and on leased point-to-point telephone-type circuits".

[2] ITU-T Recommendation G.168: "Digital network echo cancellers".

2.2 Informative references

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ITU-T Recommendation G.165: "Echo Cancellers".

[i.2] ITU-T Recommendation V.2: "Power levels for data transmission overtelephone lines".

[i.3] ITU-T Recommendation V.21: "300 bits per second duplex modemstandardized for use in the general switched telephone network".

[i.4] ITU-T Recommendation V.25: "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".

[i.5] ITU-T Recommendation G.131 (1996): "Control of talker echo".

[i.6] ITU-T Recommendation Q.115.1 (1999): "Logic for the control of echo control devices/functions".

[i.7] ITU-T Recommendation V.152:"Procedures for supporting Voice-Band Data over IP Networks".

[i.8] ITU-T Recommendation T.38 (2010): "Procedures for real-time Group 3 facsimile communication over IP networks".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

acoustic echo: acoustic echoes consist of reflected signals caused by acousticenvironments

NOTE: In these acoustic environments, an echo path is introduced by the acoustic path from the loudspeaker or earpiece to the microphone, e.g. echo created from hands-free speakerphones (see ITU-T Recommendation G.168 [2]).

echo canceller: 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 ITU-T Recommendation G.168 [2])

Non-Linear Processor (NLP): 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.

The present document assumes an echo canceller is equipped with an NLP function that can be enabled or disabled when performing the tests defined in the present document. An NLP function can be enabled or disabled by theuser (for the purpose of performing a particular test), or may also be disabled upon detection of an appropriate disabling tone (e.g. 2 100 Hz) (see ITU-T Recommendation G.168 [2]).

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ANM	Answer Message
ANS	V.25 Answer tone
CNG	CalliNG Tone
EC	Echo Canceller
IAD	Integrated Access Device
ISUP	ISDN User Part
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JB	Jitter Buffer
JBD	Jitter Buffer Delay
JBS	Jitter Buffer Size
MGW	Media Gateway
MSAN	Multi Service Access Nodes
NLP	Non-Linear Processor
PCM	Pulse Code Modulation
PSTN	Public Switched Telephone Network
RTP	Real Time Protocol
SIP	Session Initiation Protocol
UDI	Unrestricted Digital Information
VBD	Voice Band Data
VoIP	VoIP

4 Characteristics of jitter buffers

4.1 General

The present document describes the activation procedures of a jitter buffer, including the requirement for in-band tone activating and other control mechanisms. The jitter buffers are assumed to be dynamic jitter buffers and fixed jitter buffers. Fix jitter buffers shall be provided for fax and voice band data and 64 kbit/s bit sequence (UDI).

4.1.1 Jitter Buffers

A jitter buffer is designed to remove the effects of jitter from the decoded voice stream, buffering each arriving packet for a short interval before playing it out synchronously. A **fixed jitter** buffer maintains a constant size whereas an **adaptive jitter buffer** has the capability of adjusting its size dynamically in order to optimize the delay/discard tradeoff. The disadvantage of a**daptive jitter buffer** is that a part of the jitter budget is transferred to the user. While the human perception of audio delay variation is low, modem and fax applications are extremly sensitive to delay variation in the audio path. For this reason adaptive jitter buffer are not aplicable for fax and modem transmission. Fixed jitter buffers try to maintain a constant End-to-End audio delay.

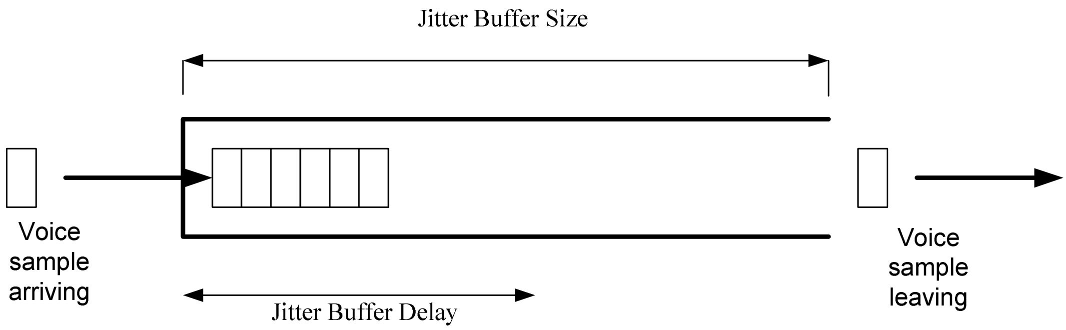

Figure 1: Jitter Buffer Size and Delay

- Jitter Buffer Size (JBS): The maximum amount of time packets can stay in the buffer.
- Jitter Buffer Delay (JBD): The jitter buffer delay is also called de-jitter delay, holding time or play-out delay. It corresponds to the time packets stay in the buffer, which is less than the jitter buffer size. The time of departure of each packet is determined by reading out the timestamp information provided by RTP.

4.2 Purpose, operation and environment

For proper operation for VBD-services, Jitter buffers have the following fundamental requirements:

1) fast switching between dynamic and fix jitter buffer mode;

2) proper operation during facsimile and data transmissions.

4.3 External enabling of fixed jitter buffers

The fixed jitter buffer for 64 kbit/s bit sequence (UDI) and V.152 [i.7] VBD shall be activated directly by signalization.

5 Characteristics of jitter buffer activator

5.1 General

The jitter buffer covered by the present document should be equipped with a tone detector that conforms to this clause. This tone detector shall activate the fixed Jitter Buffer only upon detection of a signal which consists of a 2100 Hz tone. The frequency characteristics of the tone detector are given in figure 2.

5.2 Detector characteristics

The tone detector shall detect a tone in the frequency range of 2 100 Hz ± 21 Hz (see ITU-T Recommendation V.21 [i.3]). The detection channel bandwidth should be chosen wide enough to encompass this tone (and possibly other 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 tone. The band characteristics shown in figure 2 will permit changing the jitter buffer behaviour by the 2 100 Hz tone as well as others used in North America. Figure 2 indicates that in the frequency band 2 079 Hz to 2 121 Hz detection **must** be possible whilst in the band 1 900 Hz to 2 350 Hz detection **may** be possible. Providing that only the recommended 2 100 Hz tone is used internationally, interference with signalling equipment will be avoided. The dynamic range of the detector should be consistent with the input levels as specified in ITU-T Recommendation V.2 [i.2] with allowances for variation introduced by the public switched telephone network.

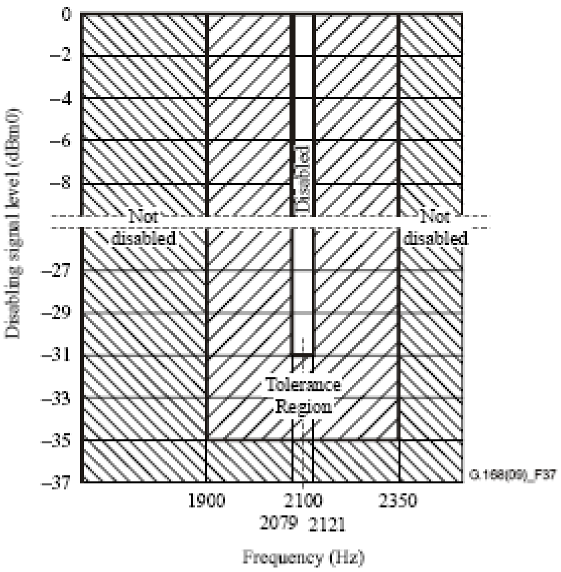

Figure 2: Required band characteristics

5.3 Noise tolerance

The detector should operate correctly with white noise less than or equal to 11 dB below the level of the 2 100 Hz signal. No definitive guidelines can be given for the range between 5 dB 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. It is noted that it is possible to design a detector capable of operating correctly at 5 dB signal-to-noise ratio.

5.4 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 jitter buffer activator is required to operate within one second of the receipt of the activating signal. The one second operate time permits the detection of the 2 100 Hz tone.

5.5 False operation due to speech signals

It is desirable that the jitter buffer activator should rarely operate falsely on speech signals. To this end, a reasonable objective is that, for an jitter buffer installed on a working circuit, usual speech signals 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 signal is interrupted because of inter-syllabic periods, before changing the jitter buffer behaviour has taken place, the operate timing mechanism should reset. However, momentary absence or change of level in a true signal should not reset the timing.

5.6 Release time

For further study.

5.7 Other considerations

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

6 Activation of jitter buffer for VBD

During telephony mode, the initiating station sends the caling tone (for fax called CNG, 1 100 Hz, a series of interrupted bursts of binary 1 signal or the 1 300 Hz signal (V.25 [i.4])) and while this takes place the user of the receiving station may be continuing to speak or send audio.The station on the left (figure 3) is the initiating station. The speech or audio signal from the station on the right have placed the jitter buffer in the dynamic state. The 2 100 Hz tone (ANS) shall drive both, jitter buffers JB2 and JB1 into the fixed mode.

NOTE: Whereas the operation of JB2 does not constitute a problem, the activation of JB1 may need special attention. This scenario is illustrated in figure 3.

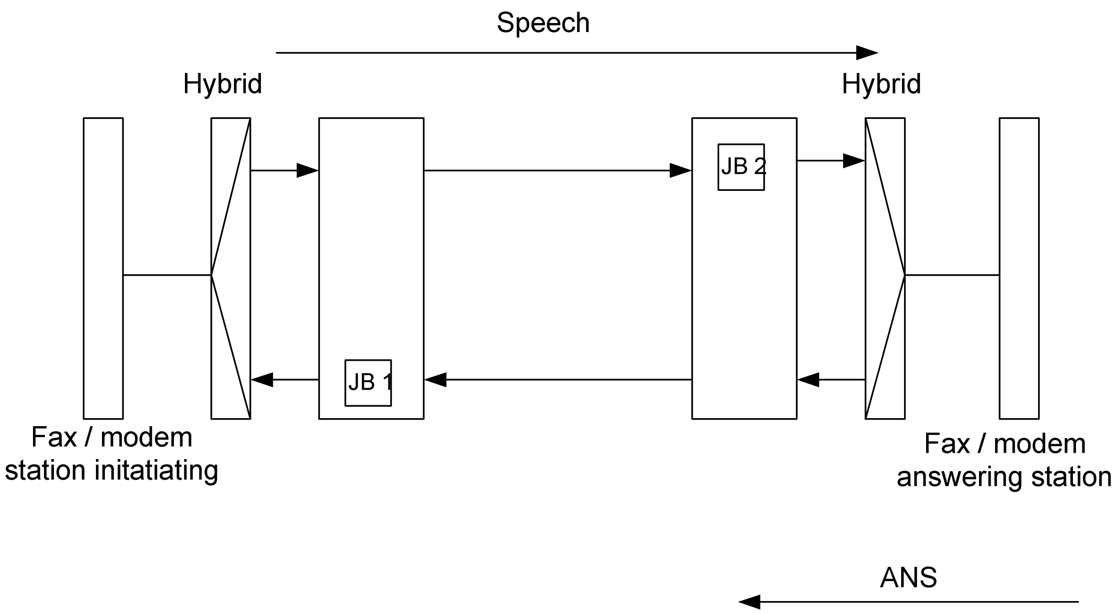

Figure 3: Activation of jitter buffer for VBD inband

7 Activation of jitter buffer for 64 kbit/s bit sequence (UDI) and V.152 VBD

The fixed jitter buffer from the calling and called side for 64 kbit/s bit sequence (UDI) shall be activated directly by a signalization. The activation takes place latest with the reception of Connect/ANM (ISUP) or 200 OK (SIP) message. See figure 4.

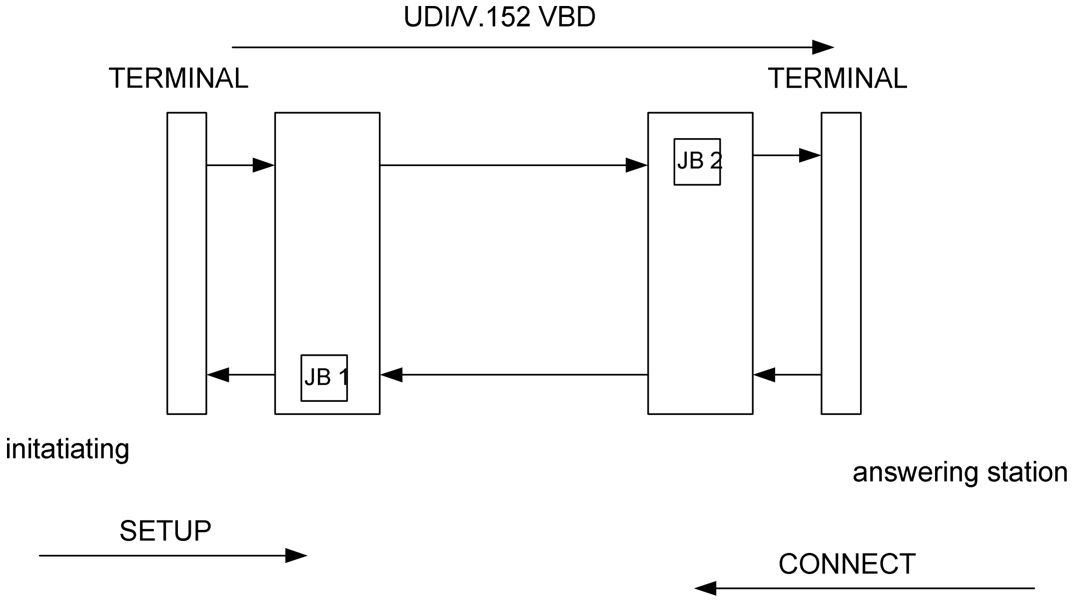

Figure 4: Activation of jitter buffer for VBD activated directly by signalization

8 Recommendation for values of jitter buffers

8.1 Fixed jitter buffers

In case VBD it is the goal to keep the audio end-to-end delay constant during the entire call. The jitter buffer has to be implemented in such a way that any jitter occuring during the entire call will not change the end to end delay.

The jitter buffer may adapt if there is a overflow or underrun.

8.2 Adaptive jitter buffers

In case of voice the strategy of jitter buffer implementation is to keep the end to end audio delay as low as possible under all jitter conditions. Any jitter buffer implementation should mostly not impair the listening speech quality as perceived by the user.

For voice calls between MSAN, IAD, MGW adaptive jitter buffers are required. The minimum jitter buffer size should be smaller or equal to one packet size.

For adaptive jitter buffers the maximum aberration from the real jitter in the network should be one packetization time interval. It is recommended that the jitter measurement period for Jitter should be 2-3 packet intervals, not only on one packet interval. The adaptation interval towards higher values should be done immediately after the jitter measurement period. The adaptation towards lower values should be after at least several seconds or during silence periods.

9 Echo canceller

As a general rule, echo cancellers are required in VoIP systems due to the high transmission delay.

In accordance with ITU-T Recommendations G.131 [i.5] and Q.115.1 [i.6] echo canceller (EC) according to ITU-T Recommendation G.168 [2] shall be used if the mean one way delay of the "talker echo transmission path" exceeds the 25 ms limit.

9.1 Characteristics of an echo canceller tone disabler (ITU-T Recommendation G.168)

9.1.1 General

The echo cancellers implemented according ITU-T Recommendation G.168 [2] should be equipped with a tone detector that conforms to this clause. This tone detector should disable the echo canceller only upon detection of a signal which consists of a 2 100 Hz tone with periodic phase reversals inserted in that tone, and not disable with any other in-band signal, e.g. speech or a 2 100 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 2 100 Hz tone without phase reversals. If 2 100 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 5.

The tone disabler characteristics as specified in clauses 7.4 through 7.9 in ITU-T Recommendation G.168 [2] also apply for this NLP disabling detector.

Note that if the 2 100 Hz tone contains phase reversals, then the echo canceller shall be disabled as defined elsewhere in this clause.

The term disabled in this clause 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 clause 6.4.1.9 in ITU-T Recommendation G.168 [2]. 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.

9.1.2 Detector characteristics

The tone detector shall detect a tone in the frequency range of 2 100 Hz ± 21 Hz (see ITU-T Recommendation V.21 [i.3]).

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 5 will permit disabling by the 2 100 Hz disabling tone as well as others used in North America. Figure 5 indicates that in the frequency band 2 079 Hz to 2 121 Hz detection **must** be possible whilst in the band 1 900 Hz to 2 350 Hz detection **may** be possible.

Providing that only the recommended 2 100 Hz disabling tone is used internationally, interference with signalling equipment will be avoided. The dynamic range of the detector should be consistent with the input levels as specified in ITU-T Recommendation V.2 [i.2] with allowances for variation introduced by the public switched telephone network.

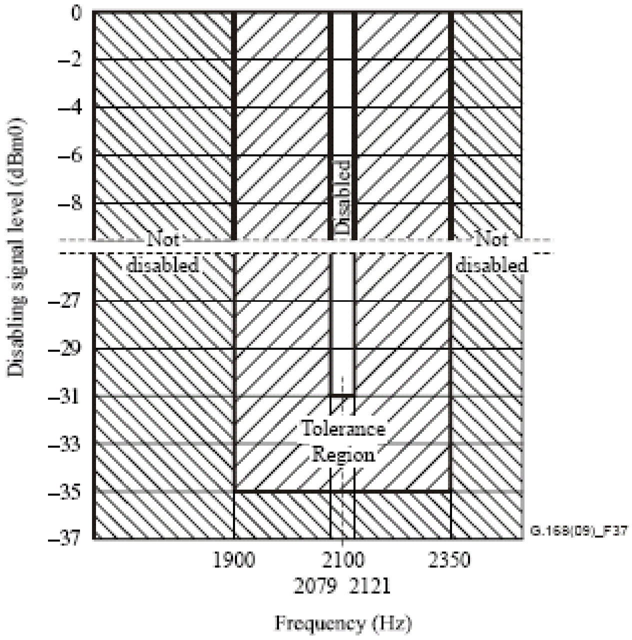

Figure 5: Required disabling band characteristics

9.1.3 Phase reversal detection

The echo canceller tone disabler requires the detection of a 2 100 Hz tone with periodic phase reversals which occur every 450 ms ± 25 ms. The characteristics of the transmitted signal are defined in ITU-T Recommendation V.25 [i.4] and ITU-T Recommendation V.8 bis [1]. Phase variations in the range of 180° ± 25° should be detected while those in the range of 0° ± 110° should not be detected. This restriction is to minimize the probability of false disabling of the echo canceller due to speech currents and network-induced phase changes. The ±110° range represents the approximate phase shift caused by a single frame slip in a PCM system.

9.1.4 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 Hz to 3 400 Hz) is applied to the tone disabler simultaneously with a 2 100 Hz signal, the 2 100 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 2 100 Hz signal and no less than a level 5 dB below the level of the 2 100 Hz signal. As the level of the 2 100 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 2 100 Hz signal level. These requirements, together with the noise tolerance requirements given in clause 9.1.5 are illustrated in figure 6.

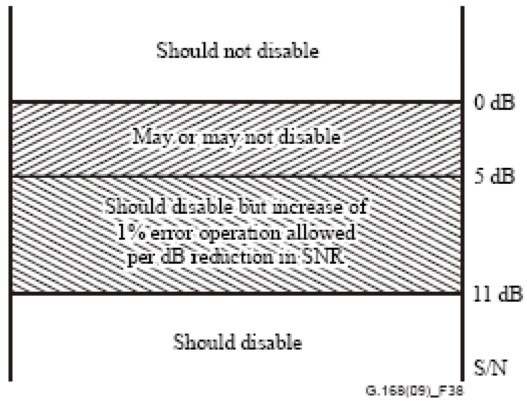

Figure 6: Guardband and noise tolerance requirements

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 [i.4]. 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 [i.4] ensures at least one second of quiet time between calling tone bursts.

9.1.5 Noise tolerance

The detector should operate correctly with white noise less than or equal to 11 dB below the level of the 2 100 Hz signal. No definitive guidelines can be given for the range between 5 dB and 11 dB because of the variations in the test equipment used. In particular, performance may vary with the peak-to-average ratio of the noise generator used. As a general guideline, however, the percentage of correct operation (detection of phase variations of 180°± 25° and non-detection of phase variations of 0°± 110°) should fall by no more than 1 % for each dB reduction in 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.

9.1.6 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 Hz to 700 Hz having a level of -27 dBm0 or greater, and from 700 Hz to 3 000 Hz having a level of -31 dBm0 or greater. The tone disabler should release for any signal in the band from 200 Hz to 3 400 Hz having a level of -36 dBm0 or less.

9.1.7 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. The one second operate time permits the detection of the 2 100 Hz tone and ensures that two-phase reversals will occur.

9.1.8 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 bandoperation 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.

9.1.9 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.

To this end, in the reference tone disabler described in annex B of ITU-T Recommendation G.165 [i.1], which meets the above requirements, the tone disabler circuitry becomes inoperative if one second of clear (i.e. no phase reversals or other interference) 2 100 Hz tone is detected. The detector circuit remains inoperative during the data transmission and only becomes operative again 250 ms ± 150 ms after a signal in the holding band falls at least 3 dB below the maximum holding sensitivity. Thus the possibility of inadvertent disabling of the echo canceller during facsimile or low speed (< 9,6 kbit/s) voiceband data transmission is minimized.

9.1.10 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 ms ± 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.

9.1.11 Other considerations

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

Careful attention should be given to the number of phase reversals required for detection of the disabling tone. Some Administrations favour relying on 1 to improve the probability of detection even in the presence of slips, impulse noise, and low signal-to-noise ratio. Other Administrations favour relying on 2 to improve the probability of correctly distinguishing between non-phase-reversed and phase-reversed 2 100 Hz tones, and to reduce the likelihood of false triggering of the tone disabler by speech or data signals.

History

Document history		
V1.1.1	April 2011	Publication