

International Telecommunication Union

**ITU-T**

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

**G.161**

(06/2012)

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

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

---

**Interaction aspects of signal processing  
network equipment**

Recommendation ITU-T G.161



ITU-T G-SERIES RECOMMENDATIONS

**TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS**

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
Transmission planning and the E-model	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
<b>Apparatus associated with long-distance telephone circuits</b>	<b>G.160–G.169</b>
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
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

*For further details, please refer to the list of ITU-T Recommendations.*

# Recommendation ITU-T G.161

## Interaction aspects of signal processing network equipment

### Summary

The evolution towards interactive multimedia applications and implementation of new signal processing devices, both in terminals and communication networks, means that a number of sophisticated signal processing algorithms will be deployed for echo control, noise reduction, level control, voice activity detection, speech coding, speech recognition and voiceband traffic technologies. These technologies may interact with one another and with the imperfections of transmission systems, such as delay and error bursts, to adversely affect the quality of speech and voiceband signals.

The purpose of Recommendation ITU-T G.161 is to clarify interaction aspects and to provide guidance to minimize the undesirable interactions of signal processing network equipment (SPNE), between different SPNE, between SPNE and terminals and between SPNE and transmission systems in telecommunication networks.

### History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T G.161	2002-06-13	15
2.0	ITU-T G.161	2004-06-13	15
3.0	ITU-T G.161	2012-06-29	16

## FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). 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 Assembly (WTSA), 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 WTSA 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.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

## 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, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2013

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

## Table of Contents

	<b>Page</b>
1 Scope .....	1
2 References.....	1
3 Definitions .....	3
3.1 Terms defined elsewhere.....	3
3.2 Terms defined in this Recommendation.....	3
4 Abbreviations and acronyms .....	7
5 Interaction aspects of signal processing network equipment .....	8
5.1 Interaction aspects of echo suppressors.....	8
5.2 Interaction aspects of echo cancellers .....	9
5.3 Interaction aspects of voice enhancement devices (VEDs).....	18
5.4 Interaction aspects of automatic level control .....	20
5.5 Interaction aspects of circuit multiplication equipment .....	21
5.6 Interaction aspects of acoustic echo controllers .....	24
5.7 Interaction aspects of voice gateways and IP-based circuit multiplication equipment .....	26
5.8 Delay introduced by signal processing network functions.....	30
5.9 Some considerations on the preferred locations of voice enhancement SPNE .....	30
6 Conformance and interoperability .....	32
Appendix I – Tandem-free operation (TFO).....	33
I.1 Introduction .....	33
I.2 Essence of the current standards.....	34
Bibliography.....	35



# Recommendation ITU-T G.161

## Interaction aspects of signal processing network equipment

### 1 Scope

The evolution towards interactive multimedia applications and implementation of new signal processing devices, both in terminals and communication networks, means that a number of sophisticated signal processing algorithms will be deployed for echo control, noise reduction, level control, voice activity detection, speech coding, speech recognition and voiceband traffic technologies. These technologies may interact with one another and with the imperfections of transmission systems, such as delay and error bursts, to adversely affect the quality of speech and voiceband signals.

The purpose of this Recommendation is to clarify interaction aspects and to provide guidance to minimize the undesirable interactions of signal processing network equipment (SPNE), between different SPNE, between SPNE and terminals and between SPNE and transmission systems in telecommunication networks.

### 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; 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. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- [ITU-T G.108] Recommendation ITU-T G.108 (1999), *Application of the E-model: A planning guide*.
- [ITU-T G.108.2] Recommendation ITU-T G.108.2 (2003), *Transmission planning aspects of echo cancellers*.
- [ITU-T G.113] Recommendation ITU-T G.113 (2001), *Transmission impairments due to speech processing*.
- [ITU-T G.131] Recommendation ITU-T G.131 (2003), *Talker echo and its control*.
- [ITU-T G.164] Recommendation ITU-T G.164 (1988), *Echo suppressors*.
- [ITU-T G.165] Recommendation ITU-T G.165 (1993), *Echo cancellers*.
- [ITU-T G.168] Recommendation ITU-T G.168 (2004), *Digital network echo cancellers*.
- [ITU-T G.169] Recommendation ITU-T G.169 (1999), *Automatic level control devices*.
- [ITU-T G.177] Recommendation ITU-T G.177 (1999), *Transmission planning for voiceband services over hybrid Internet/PSTN connections*.
- [ITU-T G.711] Recommendation ITU-T G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.
- [ITU-T G.712] Recommendation ITU-T G.712 (2001), *Transmission performance characteristics of pulse code modulation channels*.
- [ITU-T G.722] Recommendation ITU-T G.722 (1988), *7 kHz audio-coding within 64 kbit/s*.

- [ITU-T G.722.2] Recommendation ITU-T G.722.2 (2003), *Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)*.
- [ITU-T G.723.1] Recommendation ITU-T G.723.1 (1996), *Speech coders: Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*.
- [ITU-T G.729] Recommendation ITU-T G.729 (1996), *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)*.
- [ITU-T G.763] Recommendation ITU-T G.763 (1998), *Digital circuit multiplication equipment using G.726 ADPCM and digital speech interpolation*.
- [ITU-T G.764] Recommendation ITU-T G.764 (1990), *Voice packetization – Packetized voice protocols*.
- [ITU-T G.765] Recommendation ITU-T G.765 (1992), *Packet circuit multiplication equipment*.
- [ITU-T G.766] Recommendation ITU-T G.766 (1996), *Facsimile demodulation/remodulation for digital circuit multiplication equipment*.
- [ITU-T G.767] Recommendation ITU-T G.767 (1998), *Digital circuit multiplication equipment using 16 kbit/s LD-CELP, digital speech interpolation and facsimile demodulation/remodulation*.
- [ITU-T G.768] Recommendation ITU-T G.768 (2001), *Digital circuit multiplication equipment using 8 kbit/s CS-ACELP*.
- [ITU-T G.769] Recommendation ITU-T G.769/Y.1242 (2004), *Circuit multiplication equipment optimized for IP-based networks*.
- [ITU-T G.799.1] Recommendation ITU-T G.799.1/Y.1451.1 (2004), *Functionality and interface specifications for GSTN transport network equipment for interconnecting GSTN and IP networks*.
- [ITU-T G.827] Recommendation ITU-T G.827 (2003), *Availability performance parameters and objectives for end-to-end international constant bit-rate digital paths*.
- [ITU-T G.828] Recommendation ITU-T G.828 (2000), *Error performance parameters and objectives for international, constant bit-rate synchronous digital paths*.
- [ITU-T P.340] Recommendation ITU-T P.340 (2000), *Transmission characteristics and speech quality parameters of hands-free terminals*.
- [ITU-T P.342] Recommendation ITU-T P.342 (2000), *Transmission characteristics for telephone band (300-3400 Hz) digital loudspeaking and hands-free telephony terminals*.
- [ITU-T P.835] Recommendation ITU-T P.835 (2003), *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithms*.
- [ITU-T Q.50.1] Recommendation ITU-T Q.50.1 (2001), *Signalling between International Switching Centres (ISC) and Digital Circuit Multiplication Equipment (DCME) including the control of compression/decompression*.
- [ITU-T Q.115.1] Recommendation ITU-T Q.115.1 (2002), *Logic for the control of echo control devices and functions*.
- [ITU-T Q.761] Recommendation ITU-T Q.761 (1999), *Signalling System No. 7 – ISDN User Part functional description, plus Amendment 1 (2001)*.



- [ITU-T Q.764] Recommendation ITU-T Q.764 (1999), *Signalling System No. 7 – ISDN user part signalling procedures*, plus Amendment 1 (2001).
- [ITU-T T.30] Recommendation ITU-T T.30 (2003), *Procedures for document facsimile transmission in the general switched telephone network*.
- [ITU-T T.38] Recommendation ITU-T T.38 (2004), *Procedures for real-time Group 3 facsimile communication over IP networks*.
- [ITU-T V.17] Recommendation ITU-T V.17 (1991), *A 2-wire modem for facsimile applications with rates up to 14 400 bit/s*.
- [ITU-T V.18] Recommendation ITU-T V.18 (2000), *Operational and interworking requirements for DCEs operating in the text telephone mode*.
- [ITU-T V.21] Recommendation ITU-T V.21 (1988), *300 bits per second duplex modem standardized for use in the general switched telephone network*.
- [ITU-T V.23] Recommendation ITU-T V.23 (1988), *600/1200-baud modem standardized for use in the general switched telephone network*.
- [ITU-T V.26] Recommendation ITU-T V.26 (1988), *2400 bits per second modem standardized for use on 4-wire leased telephone-type circuits*.
- [ITU-T V.27] Recommendation ITU-T V.27 *ter* (1988), *4800/2400 bits per second modem standardized for use in the general switched telephone network*.
- [ITU-T V.29] Recommendation ITU-T V.29 (1988), *9600 bits per second modem standardized for use on point-to-point 4-wire leased telephone-type circuits*.
- [ITU-T V.32] Recommendation ITU-T V.32 (1993), *A family of 2-wire, duplex modems operating at data signalling rates of up to 9600 bit/s for use on the general switched telephone network and on leased telephone-type circuits*.
- [ITU-T V.34] Recommendation ITU-T 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*.
- [ITU-T V.150.1] Recommendation ITU-T V.150.1 (2003), *Modem-over-IP networks: Procedures for the end-to-end connection of V-series DCEs*.
- [ITU-R S.1522] Recommendation ITU-R S.1522 (2001), *Impact of loss of synchronisation and timing recovery on path availability in hypothetical reference digital paths*.

### 3 Definitions

#### 3.1 Terms defined elsewhere

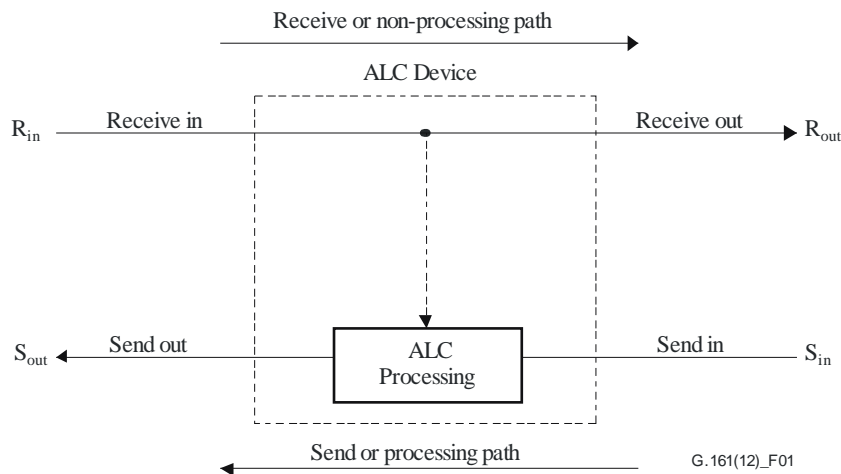
None.

#### 3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

**3.2.1 acoustic echo:** An acoustic echo is the reflected signal resulting from the acoustic path between the earphone/loudspeaker and microphone of a terminal, hand-held or hands-free mobile station.

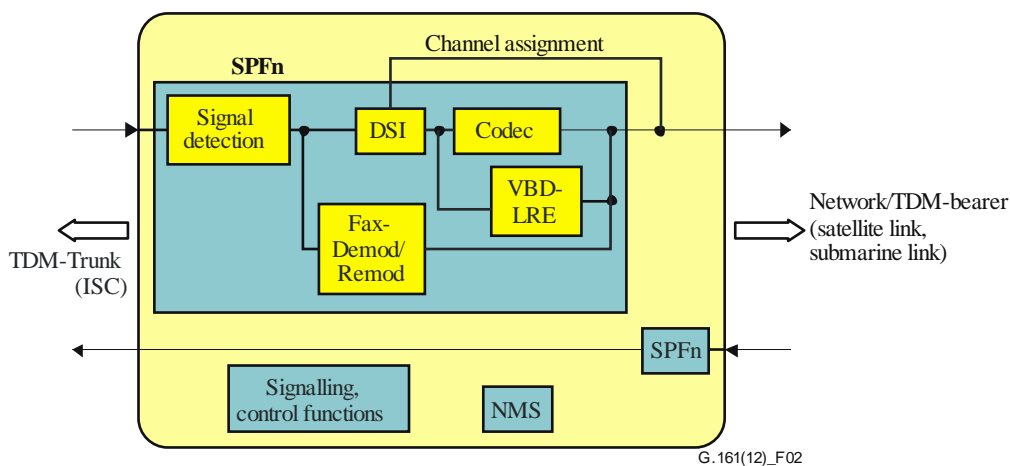
**3.2.2 automatic level control device:** An automatic level control (ALC) device is a signal processing function located in the digital transmission path which automatically adjusts the level of a signal towards a predetermined value. Devices which modify the frequency response or spectral content of the signal in such a way as to affect the overall level of the signal are also defined as ALC devices for the purpose of this Recommendation. An ALC device is designed to process signals in one direction of transmission (see Figure 1).



**Figure 1 – Block diagram of an automatic level control device [ITU-T G.169]**

**3.2.3 cancelled end:** 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.

**3.2.4 circuit multiplication equipment (CME):** Circuit multiplication equipment is a general class of equipment which permits concentration of a number of 64 kbit/s PCM encoded input trunk channels on a reduced number of transmission channels (see Figure 2).



**Figure 2 – Block diagram of digital circuit multiplication equipment (DCME) ([ITU-T G.763] and [ITU-T G.768])**

**3.2.5 IP-based circuit multiplication equipment (CME) (IP-CME):** IP-CME constitutes a general class of equipment that permits concentration of a number of IP ports on a reduced number of transmission channels over IP-based networks (see Figure 3).



NOTE 1 – Examples of managed objects are: receive/transmit port, power supply, plug-in cards, SPNFs. The communications function provides facilities for the transport of telecommunications management network (TMN) messages to and from the management application function, as well as facilities for the transit of messages. The message communications function does not originate or terminate messages. A management application function is the origin and termination for all TMN messages.

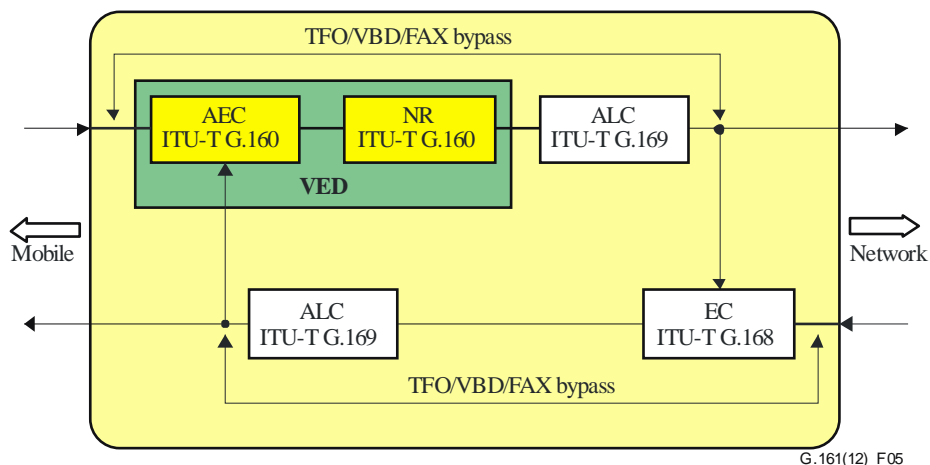
NOTE 2 – An SPNE in this context is a combination of hardware and software that performs SPNFs.

**3.2.11 signal processing network function (SPNF):** A function within a physical entity (e.g., SPNE, SPND) that performs signal processing to provide support or services to the transport network and/or to the users. Examples include electric or acoustic echo control, noise reduction, automatic level control, digital speech interpolation, low-rate encoding, transcoding.

**3.2.12 terminal acoustic echo controller:** Terminal acoustic echo controllers are voice operated devices installed in audio terminals on the customer premises, used for the purpose of eliminating acoustic echoes and protecting the communication from howling due to acoustic feedback from loudspeaker to microphone.

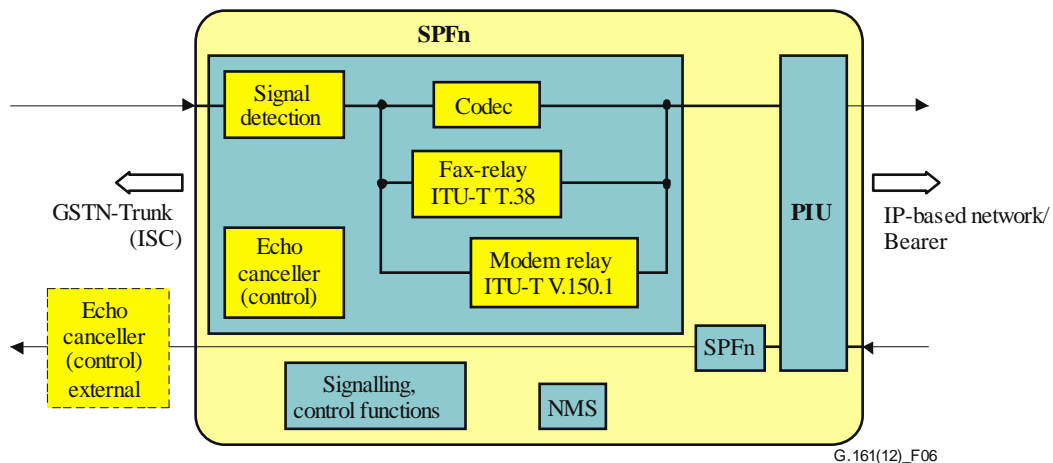
**3.2.13 voice enhancement device (VED):** A VED is defined as certain signal processing network functions such as noise reduction and acoustic echo control (see [ITU-T G.160], [ITU-T G.168] and [ITU-T G.169]) in the digital transmission path that perform voice enhancement functions on voiceband signals.

Voice enhancement functions in a mobile network environment (see Figure 5) include the control of acoustic echo generated by wireless handsets, noise reduction, and the recognition and accommodation of tandem-free operation (TFO) and interworking function (IWF) signals.



**Figure 5 – Block diagram of a voice enhancement device [ITU-T G.160]**

**3.2.14 voice gateways:** A voice gateway is a subset of a gateway that deals with voice and voiceband traffic only, and not data or video traffic (see Figure 6).



**Figure 6 – Block diagram of a voice gateway (TIGIN) [ITU-T G.799.1]**

#### 4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

A/D	Analogue-to-Digital converter
ADPCM	Adaptive Differential Pulse Code Modulation
AEC	Acoustic Echo Controller
ALC	Automatic Level Control
BSC	Base Station Controller
BTS	Base Transmitter Station
CED	Called station identification
CEST	Cancelled-End Speech Threshold
CME	Circuit Multiplication Equipment
CN	Comfort Noise
CNG	Comfort Noise Generation
CPE	Customer Premises Equipment
D/A	Digital-to-Analogue converter
DCME	Digital Circuit Multiplication Equipment
DSI	Digital Speech Interpolation
DTDT	Double Talk Detection Threshold
EC	Echo Canceller
ERL	Echo Return Loss
ERLE	Echo Return Loss Enhancement
GSM	Global System for Mobile Communications
GSTN	General Switched Telephone Network
HDLC	High-level Data Link Control
HRDP	Hypothetical Reference Digital Path
IP	Internet Protocol

IWF	InterWorking Function
LRET	Returned Echo Level
LSB	Least Significant Bit
MS	Mobile Station
MSB	Most Significant Bit
MSC	Mobile Switching Centre
NLP	Non-Linear Processor
NR	Noise Reduction
PBX	Private Branch exchange
PCM	Pulse Code Modulation
PIU	Packet Interface Unit
PLC	Packet Loss Concealment
PSTN	Public Switched Telephone Network
SPNE	Signal Processing Network Equipment
SPNF	Signal Processing Network Function
TASI	Time Assignment Speech Interpolation
TCE	Transcoder Equipment
TDM	Time Division Multiplex(ing)
TDMA	Time Division Multiple Access
TFO	Tandem-Free Operation
TIGIN	Transport Network Equipment for Interconnecting GSTN and IP Networks
UMTS	Universal Mobile Telecommunications System
VBD-LRE	VoiceBand Data-Low Rate Encoding/decoding
VED	Voice Enhancement Device
VoIP	Voice over IP

## **5 Interaction aspects of signal processing network equipment**

### **5.1 Interaction aspects of echo suppressors**

The principle of echo suppressors 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 in the send path, but the echo is much more attenuated than the direct speech. Other refinements are possible, as indicated in [ITU-T G.164]. The following text relates to ITU-T G.164 echo suppressors.

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 tandeming of echo suppressors is not recommended. In the case of voiceband data, a 2100 Hz tone is specified to permit disabling of the echo suppressor before the beginning of data transmission; this is for two reasons:

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

Facsimile is a special case. Even if an echo suppressor is disabled by 2100 Hz tone, it may be re-enabled during a facsimile transmission. The tone disabler hangover time of an echo suppressor is specified as  $250 \pm 150$  ms in clause 5.7 of [ITU-T 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. *Test results showed that the periods of silence during sequences of message exchange may extend to more than the maximum release time of 400 ms.* In addition, some facsimile manufacturers have chosen to exceed the signal separation intervals specified in [ITU-T T.30]; therefore, echo suppressors may be re-enabled.

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

[ITU-T T.30] specifies that the guard time between ITU-T V.21 and ITU-T V.29 transmission should be  $75 \pm 20$  ms. If a return signal from the local facsimile machine (within an ITU-T V.21 message-response sequence or an ITU-T V.21/ITU-T 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.

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.

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.

Enabled echo suppressors may cause errors in facsimile transmission and other low-speed modem transmission. 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.

## 5.2 Interaction aspects of 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 can be designed as "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. The appropriate transmission performance requirements of echo cancellers are shown in [ITU-T G.165] and [ITU-T G.168].

Echo cancellers are designed to:

- cancel linear echo path signals, which means a routing consisting of only linear transforming elements in conjunction with a standard coding/decoding (see [ITU-T G.711]);
- refrain from cancelling the echo when requested to do so by an in-band disabling signal;
- return to an operational mode after being disabled when the in-band signal power level drops below a specified level for a specified period of time. This design allows some networks to transport voiceband data on the same speech channels. It also allows the echo canceller to re-enable during a voice call after it has been turned off erroneously (talkoff).

Echo cancellers have the following main advantages over echo suppressors:

- Send path transparency is improved;
- Non-linear processor (NLP) hangover introduces fewer impairments;
- There is no receive insertion loss;
- Echo cancellation continues during double talk;
- Tandeming is possible (for well-designed echo cancellers).

Some echo cancellers are optioned to disable on the 2100 Hz tone as specified in [ITU-T G.164] for echo suppressors, and some are disabled with a 2100 Hz tone with periodic phase reversals of  $180^\circ \pm 25^\circ$ , as specified in [ITU-T G.165] and [ITU-T G.168] for echo cancellers. Use of the ITU-T G.165/ITU-T G.168 tone is intended to allow echo cancellers to be disabled independently of echo suppressors.

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.

### **5.2.1 Interaction of echo cancellers with facsimile transmission**

Although facsimile machines may transmit an ITU-T 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 echo canceller devices in the disabled state for the duration of the call. Echo cancellers conforming to [ITU-T G.165] and [ITU-T G.168] 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 cancellers conforming to [ITU-T G.165] and [ITU-T G.168] 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 clauses 5.2 and 5.5 of [ITU-T G.164] and clause 7.8 of [ITU-T G.168]). *Test results showed that the periods of silence during sequences of message exchange may extend to more than the maximum release time 400 ms.*

The ITU-T V.27 *ter* and ITU-T V.17 modulation scheme employed by [ITU-T T.30] are protected against the mutilation of the training sequence by echo suppressors (by using an unmodulated carrier prior to the training signal). In contrast, the ITU-T 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 ITU-T V.29 transmissions of the same format as that used during ITU-T V.27 *ter* and ITU-T V.17 transmissions). Unfortunately, these schemes are not universally recognizable by terminals produced by different modem manufacturers.



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

- The echo could be misinterpreted as an ITU-T 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:

- If echo cancellers are disabled according to the procedures of [ITU-T G.164] (2100 Hz tone without phase reversal), 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 ITU-T V.21 signal may confuse the facsimile machine at the other end and/or confuse the facsimile demodulator of the network circuit multiplication equipment (CME). The image quality may be affected as well.

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 ITU-T 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 public switched telephone network (PSTN) are not able to eliminate completely short echo bursts that could occur while the canceller is reconverging after transitions between the narrow-band signals, such as the CED tone or the ITU-T V.21 high-level data link control (HDLC) handshake, the wideband image signals (e.g., ITU-T V.29 or ITU-T V.27 *ter* signals), and again, narrow-band signals. As a minimum performance the requirements in Test 10 "Facsimile test" of [ITU-T G.168] should be fulfilled.

Current Recommendations imply that echo cancellers should be enabled during facsimile transmission. The main conclusion is that it is better to use echo cancellers that are controlled according to the ITU-T G.165/ITU-T G.168 procedures in preference to ITU-T G.164 procedures.

NOTE – This Recommendation does not discuss explicitly the case in which there is one echo canceller on one side of the connection and an echo suppressor on the other side.

### **5.2.2 Interaction of echo cancellers with modems**

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

- The echo canceller incorrectly identifies the cancelled-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 (electrical echo cancellers). These modems disable network-based echo cancellers using the disabling tone specified in [ITU-T G.165] and [ITU-T G.168]. 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.

Prior to [ITU-T V.32], most 2-wire modems used frequency division to provide duplex operation (i.e., different carrier frequencies were used for each direction of transmission). 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 V.21], [ITU-T V.23], [ITU-T V.26] (alternative B), [ITU-T V.27 *ter*] and [ITU-T V.29]. Therefore, it was accepted that these modems benefited from an active echo canceller and a disabled echo suppressor. Accordingly, in 1984, [ITU-T G.165] was modified to recommend that echo cancellers be disabled with a 2100 Hz tone with phase reversals.

Data have shown that interoperability problems occur between ITU-T V.34 high-speed fax and low-speed fax modulations ([ITU-T V.17], [ITU-T V.29] and [ITU-T V.27 *ter*]) if:

- a low speed fax machine calls a ITU-T V.34 fax machine that answers with ANSam with phase reversals; and
- there is insufficient bi-directional silence for the network echo cancellers to re-enable.

In this case, the call will be established using one of the low speed fax modulations (ITU-T V.17, ITU-T V.29 or ITU-T V.27 *ter*), but the network echo cancellers will remain disabled. The recommended mode of operation as defined in [ITU-T G.168] for this type of call is for the network echo cancellers to remain enabled. In order to solve these interoperability problems, the "75 ± 20 ms" silence interval in [ITU-T T.30] was extended to more than 400 ms (see clause 6.1.3 of [ITU-T T.30]) to allow echo cancellers to re-enable. Users should be aware that this silence period would also cause any echo suppressors conforming to [ITU-T G.164] to re-enable, which may degrade the call.

Data have also indicated that certain combinations of modems with different protocols and modulation schemes (ITU-T V.18], [ITU-T V.21] and [ITU-T V.23], DTMF and some ITU-T V.34 implementations) and some echo cancellers, in various simulated network configurations and in the network, exhibit degraded performance when the echo cancellers are enabled. Therefore it is strongly recommended to use the 2100 Hz answer tone with phase reversals as mandatory in any modem and protocol Recommendation where the presence of a network echo canceller is likely to affect the performance of the modem transmission.

In the case where echo cancellers are exposed to the initial portion of the call setup scenario, a speech signal is present at the receive input port, and a DTMF digit waveform is present at the Sgen port, the DTMF signal may be distorted.

Echo cancellers integrated in voiceband data modems, e.g., full duplex echo canceller modem ITU-T V.32, are not to be confused with the network echo cancellers that conform to [ITU-T G.164] and [ITU-T G.168], because the performance requirements for each are very different.

### 5.2.3 Interaction between echo cancellers (tandemed echo cancellers)

With the increasing use of:

- dynamic routing and special features such as call forwarding;
- the long delay introduced by low bit rate speech coders in cellular applications;
- deployment of CME functions in gateways; and
- the use of IP-/GSTN gateways equipped with echo cancellers in transmission links between IP-networks and the PSTN/ISDN,

it is very likely that some connections have more than one echo canceller.

Clause 6 of [ITU-T G.131], "Rules for connections with echo control devices" indicates that circuits with properly designed and thoroughly tested echo cancellers (meeting or exceeding the requirements of [ITU-T G.168]) can be connected in tandem without significant performance degradation. Specific guidance on planning aspects and tandeming of echo cancellers can be found in [ITU-T G.108.2].

Data from test results have shown that the following effects may arise and should be taken into consideration if echo cancellers are connected in tandem:

- The time to converge to steady-state can be increased or decreased when echo cancellers are placed in tandem.
- During initial convergence, the residual echo may be increased when echo cancellers are placed in tandem.
- The steady-state residual echo level may be increased when echo cancellers are placed in tandem.
- The echo canceller on the cancelled-end (the nearest echo canceller to the hybrid) may produce a non-linear echo path behaviour for the far-end echo canceller (tandemed echo canceller).
- The echo canceller on the far-end (tandemed echo canceller) may produce an additional echo signal if the echo path capacity of this echo canceller is exceeded.
- Tandemed echo canceller performance is mainly determined by the operation of the echo canceller on the cancelled-end (the nearest echo canceller to the hybrid).

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

To illustrate, assume that the NLP algorithm is designed to operate on the basis of the CEST/DTDT value (cancelled-end speech threshold (CEST)/double talk detection threshold (DTDT)). 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.

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.

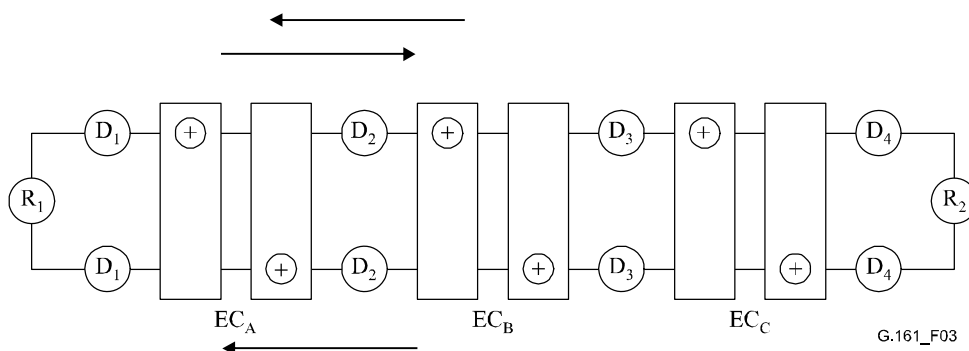
It has been observed that if an echo canceller converges too quickly, it can have annoying side effects if it is used in a situation where its echo path capacity is exceeded (such as sometimes occurs with tandem echo canceller operation). Therefore, the echo path capacity of an echo canceller should be 4 to 6 ms larger than the maximum expected network delay. This takes into account the

effect of dispersion. For example, to take into account a maximum pure delay of 58 ms, a 64-ms canceller could be selected.

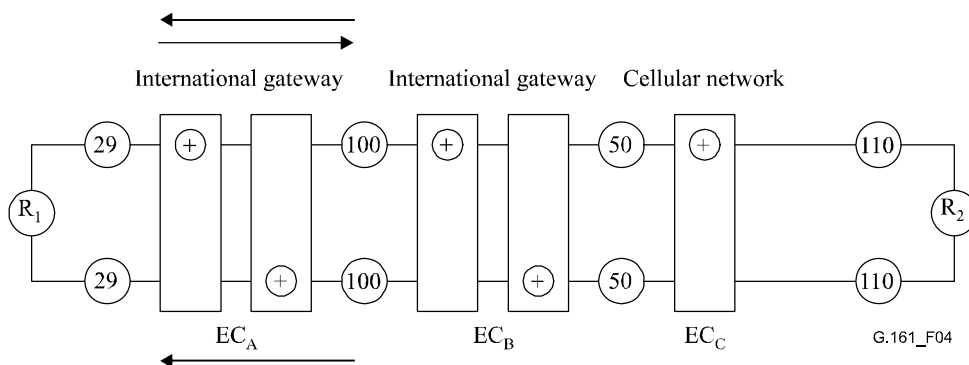
Figure 7 shows three pairs of back-to-back ECs ( $EC_A$ ,  $EC_B$ ,  $EC_C$ ), four delay generators ( $D_1$ ,  $D_2$ ,  $D_3$ ,  $D_4$ ), and two hybrids (designated by return loss  $R_1$  and  $R_2$ ) in a reference connection for tandem ECs. The values of  $R_1$  and  $R_2$  should be appropriate for proper operation of the nearest canceller (e.g., at least 6 dB). By selectively disabling ECs (either singly or in pairs), and varying the delays, it is possible to capture the relevant attributes of telephone connections with ECs. Guidance on how to control echo cancellers can be found in [ITU-T Q.115], and regarding tandem free operation in Appendix I.

As an example (see Figure 8), 29 ms delay at  $D_1$  50 ms at  $D_3$ , 100 ms at  $D_2$ , 80-110 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 5),  $EC_A$  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.

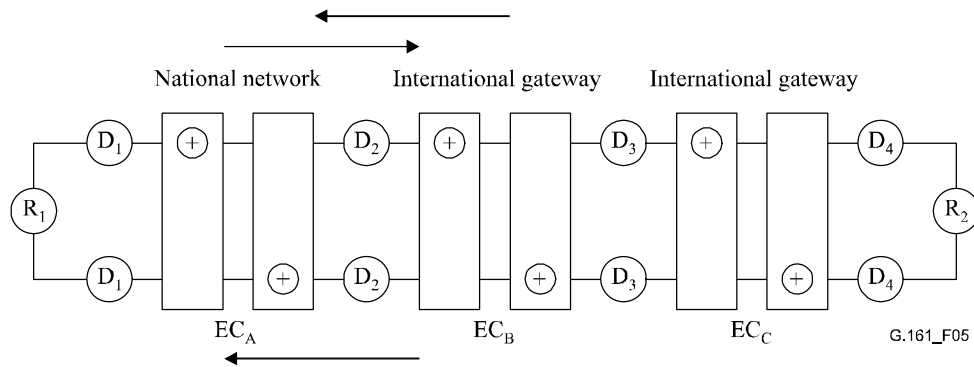
The sample configuration in Figure 9 can be extended easily if more pairs of ECs are required. In particular, inclusion of a fourth pair of ECs (and another delay generator) would capture the important features of an international connection with ECs in each national network as well as at the ends of the international facility.



**Figure 7 – Reference connection for tandem echo cancellers**



**Figure 8 – Example of international connection originating at analogue station and terminating in a digital cellular network**



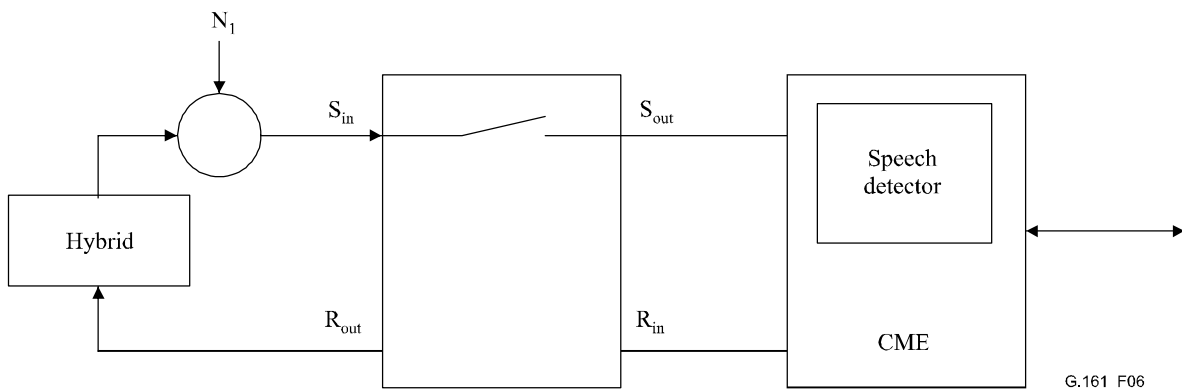
**Figure 9 – Example of international connection**

#### 5.2.4 Interaction of echo cancellers with CME

Echo cancellers may be needed for CME, even for a short terrestrial circuit, because of the additional buffering delay in a CME. The appropriate transmission performance requirements of CME are shown in [ITU-T G.763], [ITU-T G.766], [ITU-T G.767] and [ITU-T G.768].

One possible interaction relates to the potential loading effect of the comfort noise injected by the echo canceller on a CME (see Figure 10). 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.

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 CME to falsely classify this change in noise level as the presence of speech. In this case, the CME 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. It also has a negative effect on the subjective quality.



**Figure 10 – Speech detector/echo control device interaction**

##### 5.2.4.1 Detailed interaction

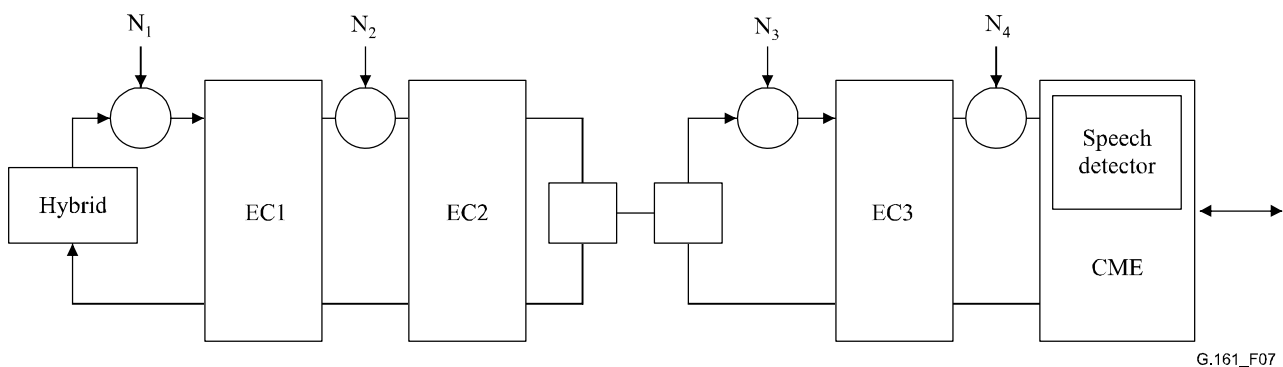
This interaction occurs as follows:

- 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 CME 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 CME 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 CME 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 11 shows a possible network configuration, with multiple echo cancellers interacting with a CME speech detector. In this configuration, the CME speech detector may respond to unit step increases in noise power, which result from echo canceller centre clipper activations in the send paths of echo cancellers 1 and 3. (The role of the NLP is to remove the residual echoes due to imperfect cancellation.) The CME 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 canceller 1 switch activation. The extent to which the CME 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 CME speech detector threshold adaptation algorithm. For example, the dual step increases in noise presented to the CME 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 11 – Multiple echo control devices in a CME network configuration**

#### 5.2.4.2 Possible solutions

There are several methods for dealing with the interactions between the echo cancellers and the CME speech detector. In one approach, the echo canceller 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 CME, 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 cancellers in use and the uniqueness of this application.

In a second approach, the speech detector adaptive threshold of the CME 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 cancellers in use and the uniqueness of the proposed application. Further, the large base of echo 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 G.165] and/or [ITU-T G.168] for a new generation of echo cancellers. The main point of this clause is that the solution depends on the speech detection procedures of both the CME and the echo canceller.

## **5.2.5 Interaction with non-linearities and time variant signal processing techniques**

### **5.2.5.1 Effects of a non-linear echo path**

The theory of echo cancellation assumes that the echo path is linear and time-invariant. Therefore, it is critical that clipping and non-linear distortion do not occur in the echo path between  $R_{out}$  and  $S_{in}$  of an echo canceller. 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.

Imperfect cancellation can occur, for example, in networks conforming to [ITU-T G.108] in which quantization distortion units (qdu) (see Table 1 of [ITU-T G.113]) are added in an echo path. The accumulated quantization distortion from these codecs may prevent an echo canceller from achieving the necessary returned echo level ( $L_{RET}$ ) by using linear cancellation techniques alone. It is, therefore, recommended that all echo cancellers capable only of modelling the linear components of echo paths but intended for general network use should incorporate suitable NLPs. In specific network environments with low delay or high echo return loss (ERL), it may be possible to disable the NLP in an echo canceller with a sufficiently high echo return loss enhancement (ERLE). This may result in higher overall speech quality, as NLPs sometimes cause speech degradation.

Some echo cancellers use the signal at  $R_{in}$  as its internal received signal  $R_{rev}$ , 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.

With the increased use of digital techniques in processing voiceband signals, digital insertion losses are being implemented increasingly in digital pads. Such digital padding typically occurs in PSTN end-offices when they act as a host to a digital remote line module as well as in customer premises equipment (CPE), such as private branch exchanges (PBXs). Improperly designed digital pads may add substantial non-linearities to the transmitted signal, including the returned echo signal, therefore degrading the canceller performance. The need to maintain linearity in digitally padded signals should be recognized.

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

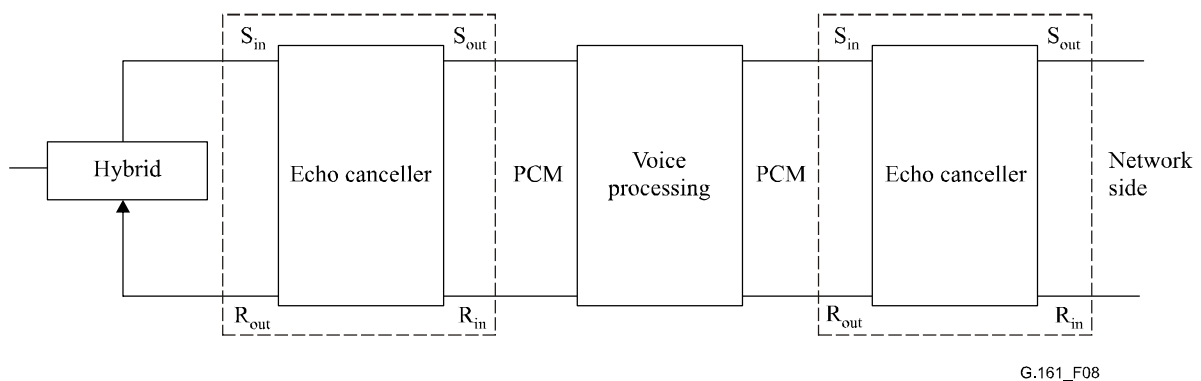
### **5.2.5.2 Voice compression in the echo path**

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 G.726]), the occurrence of a voice compression codec in the echo path becomes more likely. Measurements carried out with echo cancellers including an ADPCM circuit in the echo path have shown that the deterioration of the residual echo level may exceed 8 dB and the convergence speed decreases. If the NLP is

enabled it may cause a reduced speech quality performance, e.g., due to a disturbed echo signal caused by an unintentional switching of the NLP. An echo canceller may provide a certain amount of echo return loss enhancement.

### 5.2.5.3 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 12 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 experience non-linear or time variant echo paths. The performance of the tandem still may be acceptable if the canceller closer to the network remains stable and maintains a return loss enhancement. Theoretically, the canceller on the network side would not see an echo because the canceller on the distant end has removed it. However, it is recommended that the cancellers on the network side should be removed effectively from the connection or disabled.



G.161\_F08

Figure 12 – Voice compression between tandem cancellers

### 5.2.6 Pulse code modulation (PCM) offset

Distortion can occur in the presence of PCM offset. PCM offset is an unwanted, fixed DC signal level, caused by some analogue-to-digital converter or filter in the echo path, in the  $S_{in}$  signal relative to the  $R_{out}$  signal of an echo canceller. PCM offset can also happen in the  $R_{in}$  signal. This can result in degraded performance of an echo canceller and/or other SPNE ranging from reduced cancellation to unintentional speech clipping.

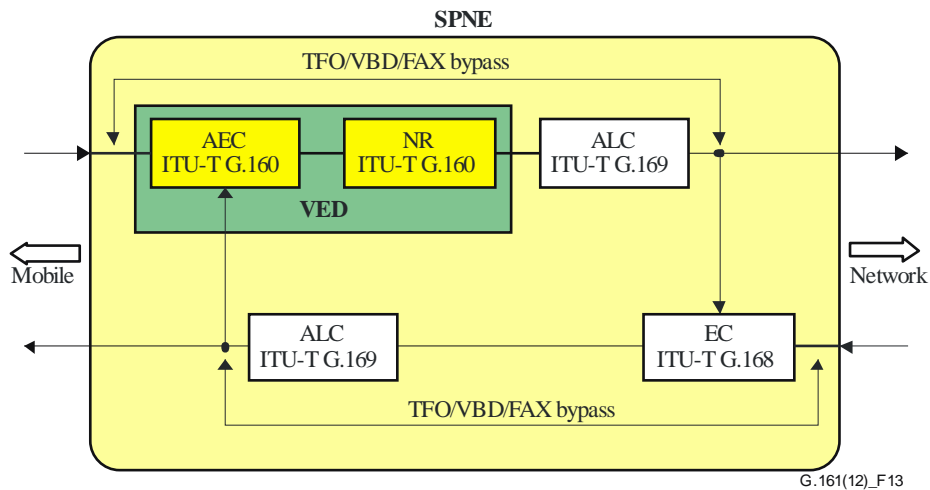
## 5.3 Interaction aspects of voice enhancement devices (VEDs)

A voice enhancement device (VED) is defined as certain signal processing network functions such as noise reduction and acoustic echo control (see [ITU-T G.160], [ITU-T G.168] and [ITU-T G.169]) in the digital transmission path that perform voice enhancement functions on voiceband signals.

### 5.3.1 Interaction aspects between voice enhancement devices, mobile applications and mobile networks

Voice enhancement functions, e.g., in a mobile network environment (see Figure 13), include the control of acoustic echo generated by wireless handsets, noise reduction, and the recognition and accommodation of tandem-free operation (TFO) and interworking function (IWF) signals.





G.161(12)\_F13

**Figure 13 – Network configuration of SPNEs, including VEDs, in a mobile network**

When including VEDs in mobile networks, the interworking between VEDs and the mobile network is mainly affected by the following points:

- the performance of the mobile terminal including:
  - the occurrence of acoustic echoes;
  - implemented echo cancellers in the terminal including additional level switching like NLPs or centre clippers;
  - comfort noise injection;
  - noise reduction algorithm;
- the speech transcoder part including the effects of speech processing and radio transmission between the mobile handsets and the network connection point.

Consequently, this may lead to tandemed signal processing units like echo cancellers, NLPs, different kinds of comfort noise injection or noise reduction algorithms.

**5.3.1.1 Interaction aspects between VEDs and acoustic echoes – Acoustic echo in digital mobile applications is typically time-variant**

In the digital mobile network acoustic echo can affect the system. The microphone of mobile terminal receives locally-generated signals like speech, multiple echoes caused by reflections and environmental noise.

The acoustic echo controller should take into account the reverberation time of the surroundings. Furthermore, the system is driven by signals such as periodic and aperiodic components with highly fluctuating magnitudes and pauses. This changes the characteristics of the acoustic echo continuously and the acoustic echo controller has to follow the changes in the impulse response.

The methodology for reducing acoustic echo in mobile networks correctly without appending side effects to the cancellation process is dependent on the acoustic echo controller's ability to characterize a signal as an acoustic echo. The correct determination is essential to a proper operation. Classification errors would be intensely harmful to the quality of the performance of VEDs.

**5.3.1.2 Acoustic echo in digital mobile applications is non-linear**

The speech from the far end (acoustic echo), injected into the mobile phone microphone, changes its characteristics by means of speech compression techniques. This causes non-linearities in these signals.

Loudspeakers and microphones in the terminals introduce non-linearities, especially if hands-free terminals are used. In this situation the typically small loudspeakers have to reproduce a sufficiently high signal level to playback the received speech signal. Typically, this leads to large non-linear distortions.

Another source of non-linearity might be radio interfaces in the echo path.

For optimal acoustic echo controller performance, it is important that the signal fed into the acoustic echo controller's  $R_{in}$  port should be linear related to the signal at the echo canceller's  $S_{in}$  port. Additionally, clipping or other non-linear distortion should not be "added" to the signal at the  $S_{in}$  port. This is most important when echo is present only at the  $S_{in}$  port, or both echo and cancelled-end speech are present and the double talk detector has not been triggered, since clipping (distorting) one affects the other.

### **5.3.1.3 Interaction aspects between echo cancellers/VEDs in tandem or noise reduction systems/algorithms in tandem**

Acoustic echo controllers and noise reduction algorithms are increasingly implemented in both mobile handsets and hands-free terminals. Special care should be taken to consider and evaluate the possible interaction between tandemed AECs implemented in VEDs and in mobile terminals. This may lead to increased speech clipping introduced by tandemed NLPs or centre clippers or the interaction between different kinds of comfort noise injection, resulting in an audible noise, switched between noise with different characteristics.

Investigations of tandeming of acoustic echo control functions in network-based VEDs and hands-free terminals in a digital mobile network showed that impaired double talk performance may occur.

In addition, the influence of noise reduction algorithms in tandem on speech quality has to be considered (guidance on the methodology for evaluating the subjective quality of speech in noise and the evaluation of noise suppression algorithms can be found in [ITU-T P.835]).

### **5.3.1.4 Interaction aspects of VEDs using tandem-free operation (TFO)**

VEDs may support tandem-free operation (TFO). Once in TFO mode, a VED may optionally choose to invoke speech enhancement functions, causing the termination of TFO. The decision should be based on whether TFO or speech enhancement techniques will provide better overall speech quality.

The following factors should also be taken into consideration:

- The decision to terminate TFO should take into account the apparent noise generated by the TFO signalling in the least significant bits (LSBs).
- AMR-Wideband speech [ITU-T G.722.2] is possible only with TFO in GSM or other 64 kbit/s TDM-based networks, and also in some universal mobile telecommunications system (UMTS) networks. Because the most significant bits (MSBs) of the PCM samples carry a down-sampled narrow-band signal, VEDs should not interrupt TFO, if TFO transports AMR-WB signals.
- VEDs need to detect TFO signalling messages in order to determine the codec type.
- VEDs should take care to avoid frequent switching between TFO and non-TFO. Frequent switching of this kind with changing contrast of noise and speech would affect the perceived speech quality negatively.

Information regarding tandem-free operation can be found in Appendix I.

## **5.4 Interaction aspects of automatic level control**

Under study.

## **5.5 Interaction aspects of circuit multiplication equipment**

Circuit multiplication equipment (CME) operation shall be tolerant to burst of errors. If SPNE amplifies error bursts more than is strictly necessary, then the duration of the disturbance to the traffic carried becomes more intrusive and the error burst may be amplified. Note that a period of unavailable time begins at the onset of 10 consecutive severely errored seconds (SES) events where changeover procedures will cause even greater disruption. Guidance on the nature of error bursts encountered on general switched telephone network (GSTN) transport systems may be found in [ITU-T G.827] and [ITU-T G.828] and [ITU-R S.1522].

### **5.5.1 Interaction aspects of circuit multiplication equipment related to speech quality effects**

Analyses from subjective tests show that the voice quality of DCME depends basically on three parameters:

- the type of generated comfort noise;
- the total number of hops;
- the bearer load.

In conversational tests and expert listening, subjects were asked to note the characteristics of the eventual impairment (noise, cut, etc.), which occurred during the conversation.

The voice degradations, which were noted, are of several types:

- cut during the communication;
- modification of voice timbre;
- variation of voice level;
- variation of comfort noise level.

The parameter which has the most determining influence on the speech quality degradation, is the number of hops, i.e., the number of successive coding/decoding. From two hops, the voice timbre seems different, less clear, and there are, in general, some variations of voice level. The quality decreases very quickly with the number of hops. Therefore, it is recommended that tandeming of CME should be prevented. Two potential approaches are discussed with tandem avoidance of CME. One is by the control from a switching equipment through signalling interfaces, and the other is internal control between the CME. Procedures for control of tandem CME avoidance under the control of a switching equipment are contained in [ITU-T Q.50.1], Amendment 1 of [ITU-T Q.761] and Amendment 1 of [ITU-T Q.764].

The other influential parameter is the bearer overload, which creates sometimes micro-cuts (freeze-outs). In this situation, the voice is often found distorted and metallic. Voice quality and overall quality are both degraded when the higher load level is applied on the system.

Concerning the comfort noise, it was observed, that the "adaptive" noise level is always preferred to the "fixed" noise: the overall quality of the communication is found better when the inserted noise level is adaptive. When set to "fixed", comfort noise is found disturbing (too high or too low a level).

### **5.5.2 Synchronous reset in CME**

In those DCME conforming to [ITU-T G.763], [ITU-T G.767] and [ITU-T G.768], limited numbers of encoders and decoders are pooled and assigned to the active speech channels utilizing external VAD/DTX/CNG schemes.

In such DCME, the speech encoders and corresponding decoders are not always connected to each other throughout the call, but only connected while the active speech signal exist on the channels. Especially in the case of DCME operating on multi-destination configuration, the codec pool is

shared for the communication to multiple destinations, and the encoder is not always used with a specific decoder.

Because of these circumstances, the internal parameters of an encoder and corresponding decoder are not always the same at the beginning of a new voice spurt after silence elimination, which may result in the extension of convergence time and potential degradation of speech quality.

The synchronous reset between the encoder and decoder is the procedure to prevent such potential degradation in decoded speech quality, and it is preferable that the reset of internal parameters to a predetermined state takes place whenever the coder pair gets the new assignment after the silence period (see [ITU-T G.763] and [ITU-T G.768]). The necessity of synchronous reset can be identified by the transition of internal assignment table in the DCME.

External synchronous reset is not necessary if silence suppression is supported according to the speech codec used, e.g., Annex B of [ITU-T G.729], or Annex A of [ITU-T G.723.1] and also if comfort noise generation is supported according to the speech codec used, e.g., Appendix II of [ITU-T G.711], Annex B of [ITU-T G.729], or Annex A of [ITU-T G.723.1].

### **5.5.3 Impact of synchronization loss**

Impact of synchronization loss at the trunk side of the DCME will not amplify its effect. The DCME will cease transporting the traffic data of relevant trunk channels, but not cease its operation.

On the other hand, the impact of synchronization loss at the bearer side of the DCME may result in amplifying its effect. If the synchronization loss occurs at the DCME bearer side, not only will the link synchronization, but also the internal DCME frame synchronization, be lost between the transmit unit of the DCME and the receive unit of a remote DCME; and the control channel messages notifying the information about the changes in the trunk channel (TC) to the bearer channel (BC) mapping table in the transmit unit to the receive unit of remote DCME might be lost as well.

From the moment this incident occurs to the moment the transmit unit of the DCME is notified of the DCME synchronization loss, the transmit unit keeps on assigning the TC to the BC, based on its trunk side channel activities, and sends out corresponding assignment messages, which are used to track the internal TC to the BC assignment mapping table in the receive unit of a remote DCME, to that of the transmit unit of DCME.

After recovering the link synchronization at the bearer side, it takes in the order of several tens of milliseconds to recover DCME synchronization loss, and both the transmit unit and the receive unit of a DCME will become available. But, if a discrepancy between the TC to the BC mapping table of the transmit unit and that of the receive unit in the remote DCME should occur, misconnections between incoming TCs of the transmit unit and corresponding outgoing TCs of a remote receive unit will be introduced. Such misconnections between the TCs may exist until the tracking of the TC to the BC assignment map to that of the transmit unit is recovered. Renewal of TC to BC mapping at the receiving side can be attained by receiving the control channel messages relevant to the bearer channel, either for mapping change or refreshment of the mapping information. The control channel message for the change of mapping will take place, based on the changes of the input trunk channels activities, and the refreshment message, to refresh the existing TC to BC mapping relations, will be exchanged only with the lowest priorities among several types of control channel messages.

Once discrepancies between the TC to BC mapping tables occur, it takes a while to recover them, but it is difficult to estimate the necessary time for recovery.

### **5.5.4 Interaction aspects of digital speech interpolation used in CME**

Time assignment speech interpolation (TASI) was originally used for undersea analogue cable systems then transposed to digital satellite transmission systems. Later on, speech interpolation was

combined with low bit rate coding in CME (see [ITU-T G.763], [ITU-T G.764], [ITU-T G.767] and [ITU-T G.768]) to increase the efficient use of the digital transmission path. Today, an increasing amount of IP-CME is using this technique together with speech compression algorithms.

Speech signals occurring on telecommunication links are generally the product of two-way conversations. It is customary for one talker to pause while the other speaks; thus, an active speech signal is present on each direction of the trunk channel for only a fraction of the available time. In addition, even when only one talker is speaking, pauses occur between utterances, so there are times when the circuit is idle. Measurements show that speech is present on each direction of the trunk channel approximately 30 to 40% of the time, averaged over a large number of busy trunks.

Digital speech interpolation (DSI) (also called silence removal) takes advantage of inactive periods in a conversation to insert speech from other conversations, and reduces the transmission capacity needed to handle a multiplicity of telephone trunk channels by exploiting the low average channel activity by transmitting active speech.

It is recommended to consider the following aspects in order to guarantee a high quality of transmitted speech and voiceband data.

#### **5.5.4.1 Occurrence of clipping**

To recognize that speech is being transmitted, a highly sensitive speech detector is required. The quality of the speech detector at the originating end is one of the important factors that determines the overall quality of the speech. If the speech detector does not detect speech correctly, it could chop the beginning of words (front-end clipping), thereby causing severe degradation to the speech quality. In contrast, if the speech detector is too sensitive, then more silence intervals will be passed and the gain will be reduced. Usually, the speech detector extends the effective duration of the speech burst, an interval denoted as hang-over to avoid tail-end clipping of speech.

#### **5.5.4.2 Overload strategy and data discrimination in DCME**

An overload strategy consisting of variable bit rate encoding and dynamic load control techniques is utilized to limit speech clipping. Data detectors are necessary to discriminate between voiceband data and speech in order to assign the voiceband data signal to a bearer channel protected against the formation of overload channels which degrade the voiceband data performance. It is also necessary to discriminate between voiceband data and facsimile in order to introduce facsimile demodulation/remodulation processing (see [ITU-T G.766]).

#### **5.5.4.3 Disturbances caused by audible noise contrast**

At the terminating end, a comfort noise or noise fill may be played out instead of silence to minimize the discontinuities between the background noise for speech and silence. Careful selection of the noise power is necessary to avoid the problem of "noise pumping", an annoying contrast between the background noise during the silence period and the background noise during speech bursts.

#### **5.5.4.4 Additional delay, freeze-out in DCME**

Another problem, called freeze-out, occurs when a talker requires a channel while all channels are occupied so that the leading edge of the incoming speech is not transmitted. This is alleviated by adding a fixed delay in DCME.

#### **5.5.4.5 Need of echo cancellers**

Note that, in the public switched telephone network (PSTN), echo control is needed whenever digital speech interpolation is used because the speech detector may erroneously classify echoed signals as speech and reduce the compression gain. This is in addition to the need for echo cancellation to compensate for the delay due to packetization, or to avoid "freeze-out".

## **5.6 Interaction aspects of acoustic echo controllers**

### **5.6.1 Scope**

The following specifications are the requirements for acoustic echo controller(s) (AECs) which can be installed either in the terminal itself (terminal acoustic echo controller) or in remote places in the network (network acoustic echo controller) to interwork properly with other devices on the network side. They cover the applications for equipment with digital or analogue line interfaces for use in the following areas of telecommunications (denoted by applications):

- teleconferencing;
- loudspeaking (hands-free) telephones;
- videophone terminals;
- mobile and personal applications;
- interworking with speech codecs.

There are two cases of interest here: speech coding on the  $R_{in}$  path and speech coding on the  $S_{out}$  path.

### **5.6.2 $R_{in}$ path speech coding**

#### **5.6.2.1 Bit rate**

The obvious item is that the bit rate at the edge of the processing unit, including the AEC at  $R_{in}$ , should match the incoming rate.

#### **5.6.2.2 Bandwidth**

The other consideration is that the bandwidth of the AEC should, if at all possible, match the incoming bandwidth within a reasonable tolerance.

When activated, the processing in the AEC should maintain the bandwidth requirements which are given in [ITU-T G.712] for telephone band speech, and in [ITU-T G.722] for wideband speech.

If the AEC is not capable of equalling the incoming bandwidth, it should introduce compensating filters.

#### **5.6.2.3 Tandem speech coding effects**

If speech is re-encoded during processing within the AEC, the effects of tandeming with acceptable  $R_{in}$  speech codings should be considered. As a side note, the frequency shift method of howling control could be considered in this category. Also, any speech activity detection which results in processing of the  $R_{in}$  signal before being played out on  $R_{out}$  would be considered in this category. It is under study whether D.C. removal should also be in this category: current thinking is that it should not be. Overall, these tandem effects are considered to be less serious than those applying to the  $S_{in}$  to  $S_{out}$  path (clause 5.6.3.3) since any effect in this section would be localized to the user of the AEC.

### **5.6.3 $S_{out}$ path speech coding**

#### **5.6.3.1 Bit rate**

The obvious item is that the bit rate at the edge of the processing unit, including the AEC at  $S_{out}$ , should match the outgoing line rate.

#### **5.6.3.2 Bandwidth**

The other consideration is that bandwidth of the AEC should, if at all possible, match the outgoing bandwidth within a reasonable tolerance.

When activated, the processing in the AEC should maintain the bandwidth requirements which are given in [ITU-T G.712] for telephone band speech, and in [ITU-T G.722] for wideband speech.

If the AEC is not capable of equalling the outgoing bandwidth, it should introduce compensating filters.

### 5.6.3.3 Tandem speech coding effects

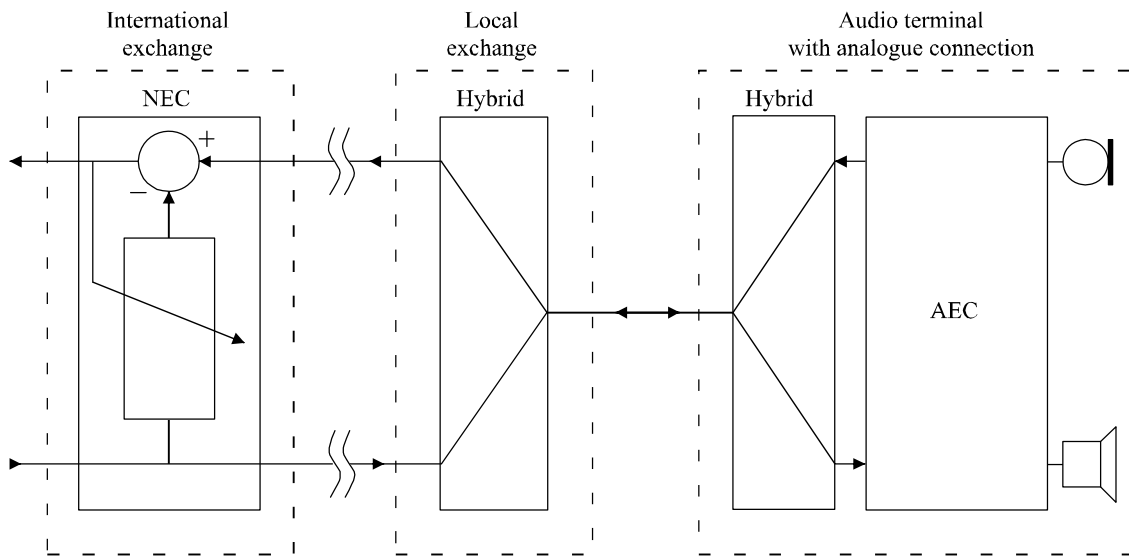
If speech is re-encoded during processing within the AEC, the effects of tandeming with acceptable  $S_{out}$  speech coding should be considered. The echo cancelling process itself is a form of speech coding in the sense of this clause and could have implications for the external speech coder on  $S_{out}$ . In general, processing which is representable as a linear filter would not pose a problem, but processing which is non-linear might pose a problem. As a side note, the frequency shift method of howling control could be considered in this category. Also, any speech activity detection which results in modulating the  $S_{in}$  signal before being played out on  $S_{out}$  would be considered in this category. It is under study whether d.c. removal should also be in this category: current thinking is that it should not be. These tandem effects are considered to be more serious than the ones applying to the  $R_{in}$  to  $R_{out}$  path (clause 5.6.2.3) since these effects are observable to the rest of the network and to the (possibly) non-AEC user. It is particularly important that the AEC does not negatively impact on the performance of cascaded network equipment.

### 5.6.4 Interworking of AECs with network echo cancellers

Acoustic echo control is becoming an important issue due to the increase in hands-free telephone sets. Although there is some commonality between issues encountered for acoustic echo control 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. Hands-free telephones are responsible for controlling their own acoustic echo. If the network echo canceller is to reduce the acoustic echo due to terminal equipment, it must be noted that the processing window should be long enough to cover the round trip end-to-end delay between the canceller and the room in addition to the acoustical echo paths in the room itself. See [ITU-T P.340] for more information on acoustical echo paths. In addition, it is important that an acoustic echo controller 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.

The most important item, from the point of view of network echo cancellers, is that there be no signal (or signal component) on  $S_{out}$  which appears to be correlated with  $R_{in}$  (within the constraints of the tail delay of the network echo canceller) unless it is during an instance of obvious double-talk. For purposes of discussion, obvious double-talk is defined as having speech power at  $S_{out}$  greater than the power at  $R_{in}$ . In fact, this constraint is perhaps a bit generous since one might (incorrectly) read it as permitting non-linear distortions in the echo path from  $R_{in}$  to  $S_{out}$ . A safer (but perhaps more strenuous than is required) condition would be to require that after initial convergence of the AEC, the level at  $S_{out}$  be greater than the least specified in [ITU-T P.342], including non-linear processing, from the level at  $R_{in}$  unless double-talk is occurring. This latter condition protects the network echo canceller since there is no acoustic echo on which to adapt (see Figure 14).



G.161\_F10

**Figure 14 – Example of network configuration where disturbing interaction can occur between acoustic echo controllers and network echo cancellers**

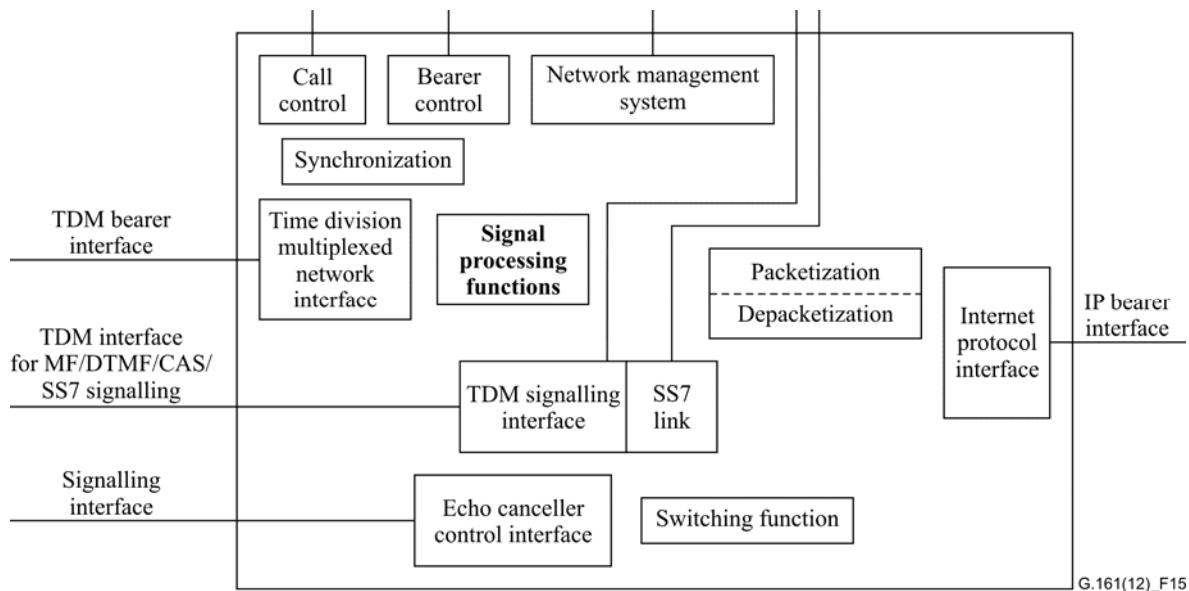
### 5.6.5 Interworking of AECs with CME

The constraints mentioned above for speech coding and network echo cancellers should largely satisfy the needs of CME equipment. One additional factor that is of interest is background noise effects. The AEC must keep the background noise level on  $S_{out}$  as constant and as low as possible. If the AEC includes a non-linear device like a centre clipper, or if it introduces a large amount of loss on the send path when there is speech on the receive side only, then appropriate comfort noise should be inserted by the AEC to avoid false activation of the CME speech detectors on modulated background noise which can result from non-linear processing or loss insertion.

### 5.7 Interaction aspects of voice gateways and IP-based circuit multiplication equipment

Gateways and IP-CME that cover voice, voiceband data, facsimile and narrowband digital data are equipment that interconnect GSTN and IP networks such that they will provide the degree of voice quality and interoperability required. In practice, such a gateway or IP-CME may be composed of multiple pieces of functions, each with specialized functions. Figure 15 illustrates some of the functions performed, e.g., in a trunking voice gateway according to [ITU-T G.799.1].





**Figure 15 – Illustration of functions performed by a trunking voice gateway [ITU-T G.799.1]**

Voice gateways and IP-CME supporting voice-related features may include the following:

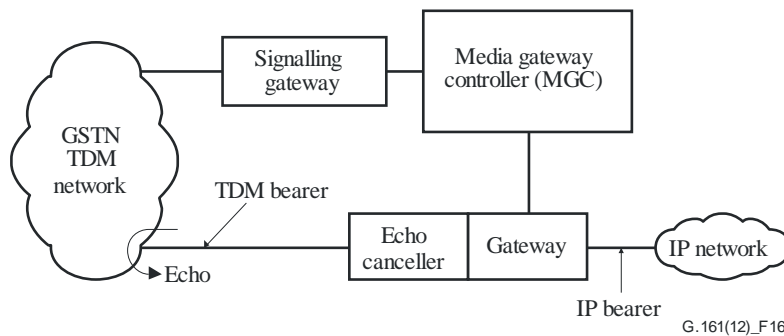
- Echo canceller functions including comfort noise generation;
- Silence suppression;
- Voice activity detection (VAD);
- Comfort noise generation (in conjunction with VAD on the sending side);
- Automatic level control;
- Selection of and/or negotiation of codec (e.g., [ITU-T G.711], [ITU-T G.723.1], and [ITU-T G.729]);
- Transcoding function for different codecs on the GSTN and IP (e.g., [ITU-T G.711] to [ITU-T G.729]);
- Packet loss concealment (PLC);
- Packet size selection;
- Continuity check tone generation and detection;
- Continuity check loop-back;
- Jitter buffer/adaptive jitter buffering control.

Some or all of the voice feature functions listed may be used during a call. The selection of these mechanisms may affect voice quality, e.g., problems due to transcoding, voice activity detection, and comfort noise generation. Echo cancellation has to cope with cancelling echo in different talking conditions and background noise types in order not to add impairments.

The following clauses give more information on this topic related to signal processing functions and speech quality.

### 5.7.1 Echo control

In compliance with [ITU-T G.177], a voice-over-IP connection traversing a hybrid is required to include an ITU-T G.168-compliant echo canceller. Since the echo canceller needs to have a constant delay for its echo path, it shall be placed so that the echo path is on the TDM side of the interface that includes the hybrid (see Figure 16). Network operators and service providers should ensure that echo cancellation is applied in the most appropriate location for the specific configuration. For more information of interaction aspects of echo cancellers, see clause 5.2.



**Figure 16 – Echo canceller integration in a trunking voice gateway**

### 5.7.2 Voice activity detection, silence suppression and comfort noise generation

In order to more efficiently use bandwidth, voice activity detection (VAD), silence suppression, and comfort noise generation (CNG) functions are used in gateways and IP-CME. As an example, silence suppression and comfort noise generation may be requested on a call that uses ITU-T G.711 coding. The network element in that case would take the GSTN data, interpret it as ITU-T G.711 coding and generate packets to the IP network using ITU-T G.711 coding only if it detects speech activity. In the reverse direction, it would receive packets from the IP network, interpret them as ITU-T G.711 encoded PCM, convert the packets into a synchronous (TDM) stream and send it to the GSTN network. During the period where no packets are received or so called silence packets are sent, it would generate comfort noise encoded as ITU-T G.711 PCM and send it to the GSTN. Silence suppression and comfort noise generation can be realized either internally or externally from the codec.

There should be a method of ensuring synchronous reset capability between the encoder and decoder of ITU-T G.726, ITU-T G.727, ITU-T G.728 and ITU-T G.729 codecs when external comfort noise generation is used with them. In these cases, an external synchronous reset will be beneficial to voice quality, particularly in the transient period from silence to active speech when the codec is used in the DTX condition. If generic SID comfort noise insertion is used with a low bit rate codec, synchronous reset between encoder and decoder should be employed (see also clause 5.5.2).

Voice activity detection may influence speech quality, e.g., noise generation during the transition from speech transmission to idle mode and when operating in noisy environments. VADs can cause too high or too low comfort noise. To mitigate these effects, voice activity detectors and the corresponding comfort noise generation should match the actual background noise in level and in spectral characteristics.

Also, an overly aggressive VAD can cause the leading or trailing edges of an active signal burst to be clipped. Clipping is a loss of speech at any time.

VAD and CNG algorithms are generally implemented in situations where transmission bandwidth resources need to be optimised. CNG is also used in echo canceller functionality to provide a comfort noise during NLP activation.

NOTE – Clipping may be due not only to an aggressive VAD as stated above, but it can also be caused by packet loss with no error concealment. The subjective impact of clipping depends on the duration and frequency of the clipping, percentage of speech clipped and overall speech activity.

### 5.7.3 Automatic level control

Interaction aspects of automatic level control are for further study.

#### **5.7.4 Transcoding function for different codecs on the GSTN and IP**

Transcoding might occur at the boundary between two networks that do not have the same codec (e.g., [ITU-T G.711] to [ITU-T G.729]). Transcoding is harmful to call quality and shall be avoided if possible.

Transcoding of speech signals between different coding schemes may affect the relative levels of the interfaced digital streams, i.e., a 0 dBm0 signal received in one stream may not result into a 0 dBm0 signal transmitted in the other stream.

#### **5.7.5 Effect of packet loss**

Packet loss causes more notable impairments than other network errors. Codecs will show different amounts of degradation depending on the effectiveness of the associated packet loss concealment (PLC) algorithm.

The choice of the PLC implementation can have an influence on the speech quality. The distortion impairment associated with a codec increases as the packet loss ratio increases. PLC can be an independent function, or implemented in a codec.

#### **5.7.6 Effects of jitter**

The presence of jitter in the network shall be considered by appropriate buffers at the receiving side. These jitter buffers are typically dynamically adjusted in order to minimize the additional delay. Care should be taken to avoid jitter buffer re-adjustments during speech activity so as not to cause additional disturbances.

#### **5.7.7 Overall speech quality of signal processing in gateways and IP-CME**

The speech quality of signal processing in gateways and IP-CME is influenced by parameters like:

- Quality of echo cancellation functions;
- Switching characteristics of NLP;
- Switching characteristic, minimum activation level, sensitivity of double talk detection;
- Behaviour of VAD, silence suppression and CN generation;
- Implemented ALC;
- Background noise transmission (background noise signal characteristics);
- Type of speech coder;
- Transcoding;
- Packet loss and jitter; and
- Quality of PLC implementation, quality of jitter buffer control and adjustment.

Basically, the overall quality can be separated into the following aspects:

- one-way transmission quality in the sending and receiving directions (listening speech quality related to codec, PLC and jitter buffer implementation);
- delay effects (conversational impact);
- echo performance (talking related);
- double talk performance (interactive conversational aspects); and
- quality of background noise transmission.

NOTE – The transmitted background noise carries important information about the environmental conditions for the other subscriber and therefore plays an important role for conversational quality determination.

### **5.7.8 Test considerations**

All the listed aspects are important. They have to be taken into consideration when developing voice quality measurement procedures, and when using voice quality measurement algorithms, in order to account for the potential functional interactions that may ultimately affect user perception of voice quality. All these considerations also have to be taken into account when comparing results from objective measurement algorithms with results from subjective tests.

A methodology to objectively predict and evaluate end-to-end listening speech quality is described in [b-ITU-T P.862] and [b-ITU-T P.862.1]. Additionally, [b-ITU-T P.502] provides objective test methods for evaluation of speech transmission quality of terminals and speech transmission systems and utilizes test signals defined in [b-ITU-T P.501].

### **5.7.9 Level offset**

Under study.

### **5.7.10 Background noise**

Under study.

### **5.7.11 Tandem signal coding**

Under study.

### **5.7.12 Tandeming of voice gateways**

Under study.

### **5.7.13 Low bit rate coding and other non-linear effects**

Under study.

### **5.7.14 Voiceband data signals, signalling tones, facsimile traffic, call-processing tones, speech**

Under study.

### **5.7.15 Interaction aspects of 3rd generation wireless networks related to speech quality effects**

Under study.

## **5.8 Delay introduced by signal processing network functions**

It is important for network administrations to understand and manage the delay contribution of an SPNE. An SPNE may have one or more SPNFs, and their individual contributions to delay must be properly understood for effective management. The processing delay of an active SPNF is of obvious importance in planning. During a normal voice call, the processing delay of an active SPNF should remain constant once the call is established and the answering side goes off-hook.

It is common that redundant SPNFs are in a network path. For the purpose of maintaining voice quality and to minimize processing path delays added by SPNFs, it is desirable that redundant SPNFs be provisioned OFF.

Each SPNF of an SPNE should be capable of being provisioned OFF. Ideally, when the SPNF is provisioned OFF, it should add no delay or only a minimal delay (see [ITU-T G.160], [ITU-T G.168] and [ITU-T G.169]).

## **5.9 Some considerations on the preferred locations of voice enhancement SPNE**

Network connections are becoming more complex nowadays due to the need for increasing flexibility of transmission, and to provide different types of services. The introduction of gateways establishes an interface between GSTN and digital networks such as packetized and IP networks. A

voice signal could, therefore, traverse different network types in a single connection. Each of these networks is likely to support its own range of voice enhancement functions. It is, therefore, expected that several voice enhancement SPNE, of similar or different types, may happen to operate simultaneously in the same communication link. This can result in undesirable interactions between SPNE, as well as tandem operations of similar voice enhancement functions. In such scenarios it may be beneficial to disable selected SPNE in a communication link in order to prevent potentially detrimental effects on overall voice quality. The SPNE to be disabled should be those whose unfavourable location may contribute to unnecessary voice quality degradation.

In voice transmission over a network, voice quality can be affected by the relative positions of the SPNE in a transmission link. The text below presents a number of factors to be considered when determining preferred locations, from a voice quality viewpoint, of voice enhancement SPNE:

#### **Acoustic echo control (AEC):**

Acoustic echo control (AEC) is a signal processing function that improves voice quality by reducing the amount of perceived echo caused by acoustic coupling between the speaker and the microphone of a terminal. AEC may also feature a comfort noise generator. It is preferred that an AEC is located in a transmission link that:

- is as close as possible to the source of acoustic echo to minimize modifications in acoustic echo characteristics by other network equipment;
- will not cause the AEC to exceed its tail capacity;
- avoids the impact of comfort noise from SPNE in the upstream of the processing thread on noise detection and noise modelling.

It is also desirable to avoid the tandeming of AEC.

#### **Automatic level control (ALC):**

Automatic level control (ALC) is a signal processing function in the digital transmission path which automatically adjusts the level of a signal towards a pre-determined value. It is preferred that an ALC device is located in a position that:

- is close to the edge of the network and not in transit nodes;
- is before the  $R_{in}$  port and after the  $S_{out}$  port of an echo canceller to avoid the creation of non-stationary echo-path.

#### **Echo canceller (EC):**

Network echo cancellers (ECs) are voice operated devices placed in the 4-wire portion of a circuit and are used for reducing the echo generated by hybrid circuits. An EC may contain a comfort noise generator. It is preferred to locate EC in a position that:

- is as close as possible to the source of echo.
- avoids non-linearities such as speech coding and decoding in the echo-path.
- avoids any SPNE in the echo-path that can create non-stationary echo-tail. NR and ALC are SPNE examples that can create non-stationary echo-path.
- avoids the impact of comfort noise from SPNE in the upstream of the processing thread on noise detection and noise modelling.

### **Noise reduction (NR):**

Noise reduction (NR) is a signal processing function in the digital transmission path intended to reduce the perceived noise level, thereby aiming to improve subjective speech quality. An NR device may contain a comfort noise generator. It is desirable to locate an NR device within the network in a position that:

- is as close as possible to the source of background noise.
- is before the  $R_{in}$  port and after the  $S_{out}$  port of an echo canceller to avoid the creation of non-stationary and non-linear echo-path.
- avoids the impact of comfort noise from SPNE in the upstream of the processing thread on noise detection and noise modelling.

It is also preferred to avoid the tandeming of NR devices.

### **Speech coding:**

Speech coding facilitates efficient speech signal transmission with reduced bandwidth requirement. However, speech encoding and decoding processes inherently introduce noise into the reconstructed speech signal. In order to ensure optimal end-to-end voice quality in a communication, it is preferred that:

- a speech codec is placed near communication end points;
- there is only a single encoding and decoding process operating on speech in the transmission path, for the purpose of avoiding speech codecs in tandem;
- speech codecs being used are amenable to tandem-free operation;
- speech coding is after NR, such as in the application of voice transmission over packetized networks.

### **Tandem-free operation (TFO):**

The main advantage of the tandem-free operation (TFO) of voice codecs is to avoid the degradation of voice quality resulting from unnecessary speech decompression and compression processes, as typically encountered in wireless mobile-to-mobile calls. Voice quality could potentially be degraded by frequent TFO disruptions causing TFO and tandem operation transitions. It is preferred that:

TFO interruptions are minimized.

## **6 Conformance and interoperability**

This Recommendation does not contain any conformance and interoperability requirements to be tested.

# Appendix I

## Tandem-free operation (TFO)

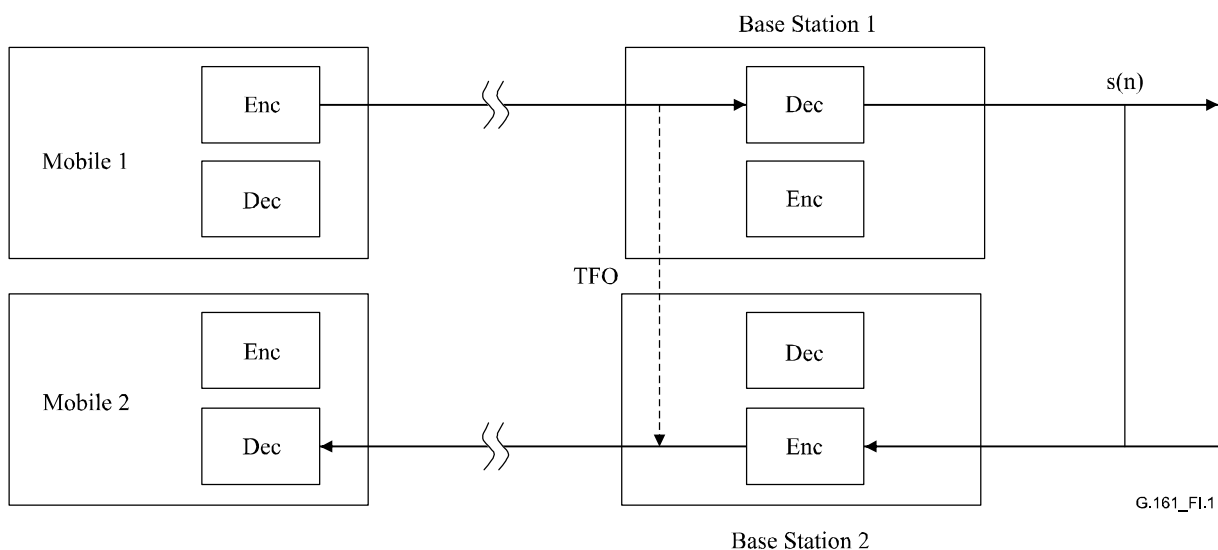
(This appendix does not form an integral part of this Recommendation.)

### I.1 Introduction

Speech coding is used extensively in telecommunication networks to reduce transmission bandwidth and increase capacity. High-quality speech codecs operating at low to medium rates (4 to 16 kbit/s) have been a primary factor in making packet voice networks and, in particular, digital wireless systems, feasible. Some popular low bit rate speech codecs developed by ITU-T are [ITU-T G.729] and [ITU-T G.723.1].

Despite the high quality of such codecs, they are lossy in the sense that the reconstruction of the speech signal is not perfect and the synthesized signal does not completely match the original speech in quality. Consequently, tandeming of codecs leads to further distortion of the processed signal. Depending on the quality/bit-rate of the codecs, as well as the characteristics of the processed signal, the added degradation due to tandem codecs may be significant. Music and noisy speech signals are examples of cases where the additional degradation due to back-to-back codecs, can be significant. In general, the lower the bit rate of a speech codec, the higher will be the degradation from tandeming. Tandeming codecs has been one of the technological hurdles in the way of the standardization of low bit-rate speech codecs.

The problem of codec tandems is often experienced in digital wireless telephony, where mobile-to-mobile calls involve two back-to-back codecs processing the signal in each direction. Figure I.1 shows such a situation, where the speech signal is assumed flowing from Mobile 1 to Mobile 2 for ease of illustration. Mobile 1 encodes the source speech signal and transmits the encoded speech data to Base Station 1. Base Station 1 decodes the speech data from Mobile 1 and produces  $s(n)$ . Because the call connection is mobile-to-mobile, Base Station 2 will encode the decoded speech signal  $s(n)$  again and transmit the encoded data to Mobile 2. As a result, the signal from Mobile 1 to Mobile 2 is encoded twice and the speech quality will be degraded more than in a single stage coding. Since the objective of digital wireless telephony is to deliver service with quality comparable to that of landline telephony, it has been necessary to develop a remedy.



G.161\_FI.1

Enc Speech coder  
Dec Speech decoder

**Figure I.1 – Tandem of speech codecs in a simple mobile-to-mobile-connection**

## **I.2 Essence of the current standards**

Generally speaking, TFO is implemented through an in-band signalling protocol that is designed to detect connections that flow through tandem codecs. Once such connections are identified, and after it is established that the codecs are interoperable, the processing flow is altered to remove the extra coding/decoding stages in the base stations. In addition, in the case of terminal devices that support multiple codecs, provisions are put in place to allow negotiation towards the highest-quality codec supported by the two terminals.

The recent development of TFO standards has attempted to ensure tandem-free operation even across systems of different access technologies, e.g., GSM/TDMA, as long as the speech codecs on the two ends are interoperable.

Different TFO standards have been established over the years for various wireless access techniques. All of these TFO standards were based on the same approach, but adapted for use with the different speech codecs in different wireless access technologies. Existing TFO standards are [b-ANSI/TIA 895A] and [b-ETSI TS 128.062].

While the TFO standards were originally designed for wireless services, they were also intended to provide interoperability with wireline packet voice networks such as VoIP. The TFO standards can be easily applied to any packetized voice transmission. The TFO protocol is designed to work without requiring out-of-band signalling or intervention of higher layers of protocol allowing for simple implementation. An ultimate objective in the development of TFO is to have a universal standard that is interoperable with the existing standards, in order to provide for seamless operation across various networks and services.



## Bibliography

- [b-ITU-T P.501] Recommendation ITU-T P.501 (2012), *Test signals for use in telephony*.
- [b-ITU-T P.502] Recommendation ITU-T P.502 (2000), *Objective test methods for speech communication systems using complex test signals*.
- [b-ITU-T P.862] Recommendation ITU-T P.862 (2001), *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*.
- [b-ITU-T P.862.1] Recommendation ITU-T P.862.1 (2003), *Mapping function for transforming P.862 raw result scores to MOS-LQO*.
- [b-ANSI/TIA 895A] ANSI/TIA-895A (10/2002), *CDMA Tandem Free Operation*.
- [b-ETSI TS 128.062] ETSI TS 128.062 (12/2005), *Digital cellular telecommunications system (Phase 2+) Universal Mobile Telecommunications System (UMTS) Inband Tandem Free Operation (TFO) of speech codecs; Service description; Stage 3, V6.2.0*.





## SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
<b>Series G</b>	<b>Transmission systems and media, digital systems and networks</b>
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and 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	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Terminals and subjective and objective assessment methods
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, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems