International Telecommunication Union

ITU-T **V.cvm**

TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU (October 2026)

SERIES V: DATA COMMUNICATION OVER THE TELEPHONE NETWORK

Interworking with other networks

Data transmission over the speech paths of mobile telephony networks

Draft new Recommendation ITU-T V.cvm

Proposal

Status of this document. This document is a *proposal* for a new Recommendation, written in the form and the conventions of an ITU-T Recommendation so that it may be contributed to the ITU-T study group responsible for the V series. It has not been approved by ITU-T, its designation V.cvm is a working name, and nothing in it binds ITU-T. The capability bits in the V.8 and V.8bis signals that it needs (Annex A) are allocations that only ITU-T can make; until they are made, the non-standard facilities mechanism of those Recommendations is used, as Annex A says.

Where its numbers come from. The behaviour of the speech codecs towards modem signals, on which the design rests, was measured with the D-Modem project's modems and test tools through the AMR and AMR-WB codecs (the OpenCORE and VisualOn implementations of 3GPP TS 26.090 and 26.190, checked bit-exact against the 3GPP reference code) and through the EVS codec (the 3GPP reference code of TS 26.442). Appendix I gives the measurements and how to repeat them; Appendix II those of EVS.

Summary

Modern mobile telephony carries speech as packets of a low bit rate speech codec: AMR in GSM and narrowband VoLTE, AMR-WB (ITU-T G.722.2) in wideband VoLTE, EVS in VoLTE and 5G voice (VoNR). These codecs model speech, not waveforms, and the voice-band modems of the V series fail through them: of the modulations of Recommendations ITU-T V.21 to V.34, only V.21 and V.22 come through the AMR codecs at all, and at 300 and 1200 bit/s. A circuit-switched data bearer is no longer offered. Where the only path between two parties is a telephone call – the general case between networks, operators and countries – data must go through the codec.

Recommendation ITU-T V.cvm specifies a data circuit-terminating equipment (DCE) for that path. It reuses the V series wherever the codec leaves it usable and replaces only what the codec breaks:

- the call starts as a V.8 session (ANSam, CI, CM, JM in V.21), with an indication of this Recommendation's capabilities, and falls back to V.22 or V.21 towards a DCE that does not have them;
- the line is probed, in the manner of V.34's line probing, for the codec's bandwidth and its noise against a multitone signal, and a mode is chosen from the measurements;
- the signal is a comb of tones with differential phase modulation, a symbol every codec frame or two, in the band the codec carries; the tones are loaded by the probing, and the mode is renegotiated as the codec's mode changes, in the manner of V.34's rate renegotiation;
- where no comb passes – AMR at 7.95 and 7.4 kbit/s, AMR-WB at 6.6, EVS at 9.6 – a parametric mode carries the data as the period of a pulse train, which the codec transmits as the lag of its long-term predictor: 200 bit/s, or 400 through codec modes that carry the shorter symbol, without error where no waveform passes; it also carries the procedures' messages;
- a Reed-Solomon code with interleaving (the code of ITU-T G.975) corrects the codec's errors and the erasures of lost frames, and V.42's LAPM, V.42bis and V.44 run above it unchanged;
- the DTE interface is V.250's, with the additions this Recommendation needs; facsimile is carried as T.38 packets.

The measured raw rate of the signal set is 1600 bit/s through AMR at 12.2 kbit/s, 800 through EVS at 13.2 kbit/s, 2200 to 2800 through AMR-WB and EVS at 24.4 kbit/s wideband, 4950 through EVS at 48 kbit/s and 10 350 through EVS at 128 kbit/s super wideband; the bound for any signal set is the codec's own bit rate. Through AMR at 7.95 and 7.4 kbit/s, AMR-WB at 6.6 and EVS at 9.6, the parametric mode carries 200 to 400 bit/s without error where no waveform does; through AMR at 4.75 kbit/s nothing passes.

Keywords

AMR, AMR-WB, data over voice, EVS, long-term predictor, modem, pitch lag, speech codec, V.22, V.8, VoLTE, VoNR.

Contents

1 Scope **7**

2 References **7**

3 Definitions **8**

3.1 Terms defined elsewhere . . . 8

3.2 Terms defined in this Recommendation . . . 8

4 Abbreviations and acronyms **9**

5 Conventions **9**

6 General description **10**

6.1 The speech path as a channel . . . 10

6.2 What this Recommendation reuses . . . 10

6.3 Overview of a call . . . 10

7 Reference configuration and interfaces **11**

7.1 Reference configuration . . . 11

7.2 The signal interface (reference point P) . . . 12

7.3 Interchange circuits . . . 12

8 Start-up procedures **12**

8.1 General . . . 12

8.2 Capability indication . . . 12

8.3 Outcome . . . 13

8.4 Timing . . . 13

9 Probing and mode selection **13**

9.1 Purpose . . . 13

9.2 Phase 2: the probing signals . . . 13

9.3 Modulation of the INFO and MP messages: the message mode . . . 14

9.4 The loading . . . 14

9.5 Phase 3: parameter exchange . . . 15

9.6 Renegotiation and retrain . . . 15

10 Modulation **16**

10.1 Signal set . . . 16

10.2 The comb . . . 16

10.3 Symbol timing and the cyclic prefix . . . 16

10.4 The codec frame and the symbol . . . 18

10.5 Differential phase modulation . . . 18

10.6 Loading and raw rate . . . 18
10.7 The parametric mode . . . 19
10.8 Control sequences during data . . . 20
10.9 Scrambling . . . 21
10.10 Quick start . . . 21
10.11 Loading with amplitude . . . 21

11 Error control and framing **22**
11.1 Structure . . . 22
11.2 The code . . . 22
11.3 Erased frames . . . 23
11.4 Mapping . . . 23

12 Link layer, compression and DTE interface **23**
12.1 Link layer . . . 23
12.2 DTE interface . . . 24
12.3 S-registers . . . 24

13 Interworking **24**
13.1 With V series DCEs on the PSTN . . . 24
13.2 With DCEs of this Recommendation across two codecs . . . 24
13.3 With facsimile . . . 25

14 Performance **25**

Annex A – Indication of this Recommendation's capabilities in V.8 and V.8bis **26**

A.1Encoding in CM and JM **26**

A.2The information field **26**

A.3V.8bis **27**

Annex B – The probing signals and the INFO messages **28**

B.1L1, L2 and L3 **28**

B.2What the receiver measures **28**

B.3INFO0 **28**

B.4INFO1 and MP **29**

Annex C – The inner code and its interleaver **30**

C.1The code **30**

C.2The interleaver 30

C.3Erasure marking 30

Annex D – Facsimile 31

D.1Principle 31

D.2Carriage of the IFP packets 31

D.3Rates 31

Annex E – A preset dictionary for V.44 32

E.1Use 32

E.2The text 32

Annex F – Zstandard compression 35

F.1 Negotiation 35

F.2 The frames 35

F.3 The dictionary 36

Appendix I – The speech codecs as a channel: measurements 37

I.1 Method 37

I.2 The modulations of the V series through the codecs 37

I.3 Steady tones 38

I.4 One carrier 39

I.5 The combs 39

I.6 Tandem codecs and frame erasures 40

I.7 The parametric mode 40

I.8 The two layers in one waveform 42

I.9 Cross-check of the codec implementations 43

Appendix II – The EVS codec 44

II.1The codec 44

II.2 Measurements **44**

II.3 What they show **44**

II.4 Consequences for the signal set **46**

Appendix III – Terminals and networks **47**

III.1 The terminal **47**

III.2 The network **47**

Appendix IV – Implementation notes **48**

Bibliography **49**

1 Scope

This Recommendation specifies a DCE for data transmission over the speech path of a telephone call carried by a mobile telephony network, that is, through one or more low bit rate speech codecs: the adaptive multi-rate codec (AMR, [3GPP TS 26.090]), the adaptive multi-rate wideband codec (AMR-WB, [ITU-T G.722.2], [3GPP TS 26.190]) and the enhanced voice services codec (EVS, [3GPP TS 26.445]), as the IP multimedia subsystem carries them in VoLTE and VoNR calls ([3GPP TS 26.114]), and through the circuit-switched speech bearers of earlier networks. It specifies:

- the reference configuration and the interface at which the DCE's signal enters and leaves the speech path (clause 7);
- the start-up procedures, built on [ITU-T V.8] (clause 8);
- the probing of the path and the selection of a mode of operation (clause 9);
- the signal set: the combs, their symbol timing and tone loading, and the parametric mode (clause 10);
- error control, framing and the use of [ITU-T V.42], [ITU-T V.42bis] and [ITU-T V.44] above it (clause 11);
- the DTE interface, as additions to [ITU-T V.250] (clause 12);
- interworking with the V series DCEs of the PSTN and with facsimile (clause 13, Annex D).

It does not specify the speech codecs, the mobile network, the terminal's acoustic processing or the IMS signalling, and it does not change the speech path in any way: it is a signal the path carries; what the terminal and the network should do for the DCE is in Appendix III (informative). Data transmission over the packet data bearers of the same networks is out of scope: this Recommendation is for the case where the telephone call is the path available.

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.711] Recommendation ITU-T G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.

[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.975] Recommendation ITU-T G.975 (2000), *Forward error correction for submarine systems*.

[ITU-T T.30] Recommendation ITU-T T.30 (2005), *Procedures for document facsimile transmission in the general switched telephone network*.

[ITU-T T.38] Recommendation ITU-T T.38 (2010), *Procedures for real-time Group 3 facsimile communication over IP networks*.

[ITU-T V.8] Recommendation ITU-T V.8 (2000), *Procedures for starting sessions of data transmission over the public switched telephone network*.

[ITU-T V.8bis] Recommendation ITU-T V.8bis (2000), *Procedures for the identification and selection of common modes of operation between data circuit-terminating equipments (DCEs) and between data terminal equipments (DTEs) over the public switched telephone network and on leased point-to-point telephone-type circuits*.

[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.22] Recommendation ITU-T V.22 (1988), *1200 bits per second duplex modem stan-*

dardized for use in the general switched telephone network and on point-to-point 2-wire 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.42] Recommendation ITU-T V.42 (2002), *Error-correcting procedures for DCEs using asynchronous-to-synchronous conversion*.

[ITU-T V.42bis] Recommendation ITU-T V.42bis (1990), *Data compression procedures for data circuit-terminating equipment (DCE) using error correction procedures*.

[ITU-T V.44] Recommendation ITU-T V.44 (2000), *Data compression procedures*.

[IETF RFC 8878] IETF RFC 8878 (2021), *Zstandard Compression and the 'application/zstd' Media Type*.

[ITU-T V.250] Recommendation ITU-T V.250 (2003), *Serial asynchronous automatic dialling and control*.

[ITU-T T.35] Recommendation ITU-T T.35 (2000), *Procedure for the allocation of ITU-T defined codes for non-standard facilities*.

[3GPP TS 26.090] 3GPP TS 26.090, *Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Transcoding functions*.

[3GPP TS 26.190] 3GPP TS 26.190, *Speech codec speech processing functions; Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; Transcoding functions*.

[3GPP TS 26.445] 3GPP TS 26.445, *Codec for Enhanced Voice Services (EVS); Detailed algorithmic description*.

[3GPP TS 26.114] 3GPP TS 26.114, *IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

answering DCE, calling DCE [ITU-T V.8].

call function, modulation mode [ITU-T V.8]: the categories of the CM and JM signals.

frame erasure [3GPP TS 26.090]: a codec frame that the decoder does not receive, or receives marked as bad, and conceals.

LAPM [ITU-T V.42]: the link access procedure for modems.

line probing [ITU-T V.34]: the transmission of a known multitone signal so that the receiver may measure the channel.

rate renegotiation [ITU-T V.34]: a change of the data signalling rate during data transmission, without retraining.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

speech path the path from the calling DCE's signal interface to the answering DCE's, through the terminals' speech processing, one or more speech codecs and the networks between them. It is the channel of this Recommendation.

signal interface the point at which the DCE's transmitted signal enters the speech path and the received signal leaves it, as linear PCM samples (reference point P, clause 7).

codec frame the block of samples a speech codec encodes together: 20 ms for the codecs of this Recommendation.

comb the set of tones of the signal set, at a common spacing.

tone symbol the information carried by one tone in one modulation symbol: a phase difference from the tone's previous symbol.

mode a signal set, a comb or a lag alphabet, a symbol period, a modulation order and a code rate, agreed between the DCEs by the probing procedure.

loading the set of tones in use and the modulation order of each.

lag alphabet the set of periods the parametric mode uses with a symbol period (clause 10.7).

lag symbol the information carried by one symbol of the parametric mode: the period of the pulse train in it, one of the lag alphabet's values.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACELP	Algebraic code excited linear prediction
AMR	Adaptive multi-rate (speech codec)
AMR-WB	Adaptive multi-rate wideband (speech codec)
ANSam	Modified answer tone ([ITU-T V.8])
CI	Call indicator signal ([ITU-T V.8])
CM, JM	Call menu, joint menu signals ([ITU-T V.8])
DCE	Data circuit-terminating equipment
DPSK	Differential phase shift keying
DTE	Data terminal equipment
DTX	Discontinuous transmission
EVS	Enhanced voice services (speech codec)
FEC	Forward error correction
IMS	IP multimedia subsystem
JBM	Jitter buffer management
LAPM	Link access procedure for modems
LPC	Linear predictive coding
MDCT	Modified discrete cosine transform
PCM	Pulse code modulation
PLC	Packet loss concealment
PSTN	Public switched telephone network
RS	Reed-Solomon (code)
SDP	Session description protocol
SER	Symbol error ratio
UE	User equipment
VAD	Voice activity detection
VoLTE, VoNR	Voice over LTE, voice over new radio (5G)

5 Conventions

In this Recommendation, "shall" indicates a requirement, "should" a recommendation and "may" a permission, as in ITU-T practice. Data signalling rates are in bit/s; the rate before the Reed-Solomon code is the *raw* rate and the rate after it the *coded* rate. Levels are relative to the full scale of the 16 bit linear PCM samples at the signal interface (dBFS); −12 dBFS is the level of the test signals. Frequencies of tones are in Hz. Octets are shown most significant bit first; bit 1 is the least significant. Figures and tables give the values of this Recommendation; where a value was measured, the measurement is in Appendix I.

6 General description

6.1 The speech path as a channel

A speech codec of the ACELP family (AMR, AMR-WB, the ACELP modes of EVS) describes each 20 ms frame of its input by a linear predictive spectral envelope, a long-term (pitch) predictor with a lag and a gain per 5 ms subframe, and a sparse algebraic excitation of a few pulses per subframe. The decoder synthesizes a signal from these parameters. The result matches the input's waveform only as far as the parameters can describe it and as far as the perceptually weighted error criterion of the encoder cares to. For a modem's signal the consequences, measured in Appendix I, are:

a) the signal-to-noise ratio of the output against the input is low, in the order of 10 to 15 dB for signals with the spectrum of speech, and it does not improve with the input level: the noise is the codec's, proportional to the signal;
b) a signal of many tones with a smooth spectral envelope is carried better than a single tone, whose line spectrum the predictive model cannot place exactly; single-carrier modulations therefore fare worse than their bandwidth suggests, and multitone signals better;
c) the phase of each tone is carried from one frame to the next well enough for differential phase modulation of four states per tone and symbol, when the symbol lasts a codec frame or two, and not for eight states or for amplitude modulation;
d) the band is the codec's: 200 to 3400 Hz for AMR, 50 to 6400 Hz for AMR-WB (its 6.4 to 7 kHz is a bandwidth extension that carries no waveform), and for EVS by its bandwidth mode and bit rate: at 24.4 kbit/s to about 8 kHz and at 48 kbit/s and above to 12 or 14 kHz (Appendix II);
e) a lost packet erases a frame; the decoder conceals it with an extrapolation, which to the modem is 20 ms of noise; the jitter buffer of the receiving terminal may also repeat or drop a frame, or stretch the signal, to adapt its delay;
f) the codec's mode (its bit rate) is chosen by the network and changes during the call; the noise changes with it;
g) the one-way delay is 100 to 300 ms, and differs between the two directions;
h) the lag of the long-term predictor is transmitted at every mode of every codec, to the sample, since nothing of voiced speech survives without it: the period of a periodic signal came through every codec mode measured but AMR at 4.75 kbit/s, where no waveform did (Appendix I, Table I.8).

Nothing in the path is a linear channel; the effects of a) to c) are those of a channel with a noise floor some 10 to 15 dB below the signal, and the design of this Recommendation is a modem for such a channel, with the measures of e) to g) around it; the parametric mode of clause 10.7 is a modem for h), for where that channel closes.

NOTE – The throughput of any signal set through a codec is bounded by the codec's bit rate: 12.2 kbit/s for AMR at its highest mode, 23.85 for AMR-WB, 13.2 to 128 kbit/s for EVS. The codec's parameters describe the signal, and no more information passes than they carry. Practical signal sets reach a fraction of the bound; the one of this Recommendation reaches in the order of 10 to 15 per cent of it through AMR and EVS.

6.2 What this Recommendation reuses

Table 1 lists the parts of the V series and of other Recommendations that this Recommendation uses as they are, and what it replaces.

6.3 Overview of a call

Figure 1 shows the phases of a call. The answering DCE sends ANSam; the calling DCE sends CI then CM with this Recommendation's capabilities; the answering DCE answers with JM. If both have the capability, the DCEs exchange probing signals and INFO messages (phase 2), each chooses a loading for the direction it receives and tells the other (phase 3, MP), and data transmission begins (phase 4). During data transmission either DCE may ask for a renegotiation of its loading or a retrain. If either DCE lacks

Table 1 – The V series and other Recommendations in this Recommendation

Function	Reused	New in this Recommendation
Start-up	[ITU-T V.8]: ANSam, CI, CM, JM, CJ, in V.21 (300 bit/s FSK, which the codecs carry)	capability indication (Annex A)
Fallback	[ITU-T V.22] (1200 bit/s, through AMR 12.2) and [ITU-T V.21] (300 bit/s, through every mode), as they are	–
Probing	the principle of [ITU-T V.34] clause 10: a multitone probing signal, INFO exchanges, MP parameter exchange, rate renegotiation and retrain	the probing signals and the parameters exchanged (clause 9)
Modulation	the differential phase modulation of [ITU-T V.22] per tone; the multitone structure of ADSL ([ITU-T G.992.1]) with a cyclic prefix	the comb, symbol timing and loading (clause 10); the parametric mode on the codec's pitch lag (clause 10.7)
Inner code	the Reed-Solomon code of [ITU-T G.975] over GF(2^8), shortened	its interleaving over symbols and frames (Annex C)
Link layer	[ITU-T V.42] LAPM with its detection phase, [ITU-T V.42bis] and [ITU-T V.44] compression; Zstandard ([IETF RFC 8878])	its negotiation in XID and its frames (Annex F)
DTE interface	[ITU-T V.250]	the +MS values, S-registers and result codes of clause 12
Facsimile	[ITU-T T.38] internet facsimile packets, [ITU-T T.30]	their carriage in the data link (Annex D)
Codec	[ITU-T G.722.2] (AMR-WB) as a channel model	–

the capability, the call continues as a [ITU-T V.8] session to V.22 or V.21 as the common modulation modes allow.

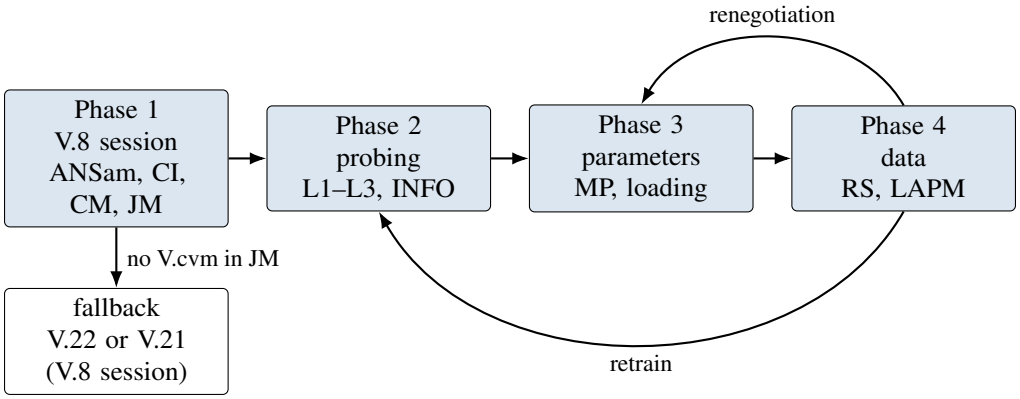


Figure 1 – The phases of a call

7 Reference configuration and interfaces

7.1 Reference configuration

Figure 2 gives the reference configuration. Reference point P is the signal interface: the DCE gives its transmitted signal to the terminal as linear PCM samples and takes the received signal from it in the same form. The terminal is whatever carries the speech path: a UE, an IMS client, a gateway between the IMS and another network, or a circuit-switched mobile terminal. Everything from one reference point P to the other is the speech path.

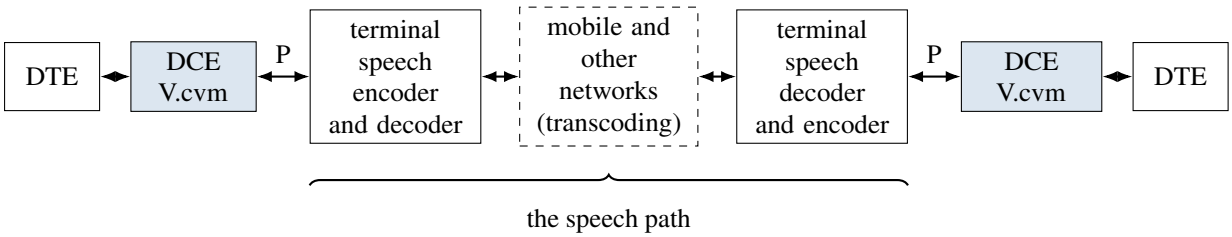


Figure 2 – Reference configuration

7.2 The signal interface (reference point P)

At reference point P:

a) samples shall be 16 bit linear PCM, two's complement, at 8000 Hz (narrowband operation) or 16000 Hz (wideband operation). A DCE shall support 8000 Hz and may support 16000 Hz. The rate is the terminal's; the DCE learns the band the path carries by probing, not from the rate.
b) the DCE's transmitted signal shall not exceed −9 dBFS rms in any 20 ms frame and should be at -12 ± 2 dBFS in the comb modes and -15 ± 2 dBFS in the parametric mode (whose pulse train has a peak factor of 13 dB, clause 10.7), so that no terminal gain stage limits it and the codec's gain quantizers are in their range. The received level is whatever the path delivers; the DCE shall accept −30 to −3 dBFS.
c) the terminal should pass the samples to and from the speech codec without acoustic processing (echo cancellation, noise suppression, automatic gain control, voice activity detection with comfort noise) between P and the codec. Where the terminal cannot disable them, the DCE shall still operate, with the modes the probing finds usable.
d) the sample clocks of the two terminals are independent. The DCE shall tolerate a difference of ±200 ppm between its transmit clock and the received signal's, and the steps of clause 10.3.

NOTE – A DCE in a UE must reach the speech codec's PCM interface. The acoustic path (loudspeaker to microphone) is not a signal interface of this Recommendation.

7.3 Interchange circuits

The DTE interface is the serial asynchronous interface of [ITU-T V.250] (clause 12), on any physical interface a V.250 DCE may have.

8 Start-up procedures

8.1 General

A DCE of this Recommendation shall perform the start-up procedures of [ITU-T V.8]: the answering DCE transmits ANSam; the calling DCE transmits CI and, on detecting ANSam, CM; the answering DCE answers with JM; the calling DCE acknowledges with CJ. The V.21 modulation of CI, CM, JM and CJ passes through every mode of the codecs (Appendix I, Table I.2), and ANSam's 2100 Hz tone, with its phase reversals, disables the network's echo cancellers (as it is meant to), which matters at a gateway to the PSTN.

8.2 Capability indication

The calling DCE shall indicate its support of this Recommendation in CM and the answering DCE in JM, as Annex A specifies, together with the modulation modes of its fallback (V.22 and V.21) and the protocol (LAPM). The indication carries:

- the version of this Recommendation (1);

- the signal interface rate: 8000, or 16000 Hz;
- the modes of clause 10 the DCE supports;
- whether the DCE can carry facsimile (Annex D).

What the link layer can do (its compression) is indicated later, in INFO0 (clause 9.2, Table B.1).

8.3 Outcome

When both CM and JM carry the indication, both DCEs proceed to the probing of clause 9 on the completion of the V.8 procedures (CJ sent, or JM ended), within the time [ITU-T V.8] allows for the start of the selected modulation. Otherwise the DCEs proceed to the common modulation mode of highest rate among those indicated, V.22 then V.21, as [ITU-T V.8] and that Recommendation specify, and this Recommendation has nothing more to say about the call.

An answering DCE that has agreed on this Recommendation and detects the end of CM (30 ms of the line 30 dB below CM's level) without having decoded CJ shall take the end of CM for CJ: a codec at a low mode can garble CJ's three zero octets, and the calling DCE's INFO0 follows CJ by only 175 ms.

NOTE – A legacy V series DCE on the PSTN, reached through a gateway, indicates its own modulation modes (V.34, V.32bis, ...) and none of this Recommendation's. Of its modulation modes, V.22 is the fastest the codec carries (Table I.2): a DCE of this Recommendation should not indicate modulation modes above V.22 in CM when it knows the speech path to be a codec's, so that the legacy DCE selects V.22 at once rather than fail at V.34 and renegotiate.

8.4 Timing

The DCEs' timers for the V.8 procedures are [ITU-T V.8]'s. The probing of clause 9 shall begin within 200 ms of the end of the V.8 procedures and shall be complete, with the first data symbols sent, within 21 s (each direction's INFO0, L1, L2, L3 and INFO1 take about 7 s with the messages at 200 bit/s, and the MP exchange 1 to 4 s; the start-up takes 15 to 21 s); a DCE that has not received its peer's INFO0 within 6 s of the end of the V.8 procedures shall fall back to V.22 by restarting the V.8 procedures, once.

9 Probing and mode selection

9.1 Purpose

Probing measures the speech path as the signal set will see it:

a) the band carried, from which the signal interface rate of the far end and the codec's bandwidth mode follow;
b) the noise per tone against a multitone signal of the signal set's own form, from which the loading follows;
c) the round-trip delay, for the timers of the link layer.

Each direction is probed and loaded separately: the two directions of a mobile call go through different codecs, radio links and jitter buffers.

9.2 Phase 2: the probing signals

After the V.8 procedures each DCE transmits, in turn as Figure 3 shows:

1. **INFO0**: its signal interface rate, the modes it supports and the comb it will probe with, modulated as clause 9.3 specifies;
2. **L1**: the probing combs of Annex B: each comb the DCE supports for the band of its signal interface, in turn (200, 300 and 400 Hz spacing, Table 3), all tones loaded, each tone with a differential phase sequence known to both, 0.6 s a comb;

3. **L2**: the 200 Hz comb with the 22.5 ms symbol, for 0.5 s, so that the receiver may compare the two symbol periods;
4. **L3**, where the DCE indicated the parametric mode: that mode's probing signal of Annex B, 40 symbols of P20 and 80 of P10 with lags known to both, each part with its preamble;
5. **INFO1**: what it measured of its peer's L1 to L3 as the loading it asks for (clause 9.4), and its own round-trip estimate.

Each part follows the last after 100 ms of silence. The calling DCE transmits its INFO0, L1, L2 and L3 first; the answering DCE begins its INFO0 50 ms after the end of the calling DCE's L3 and transmits INFO0, L1, L2, L3 and its INFO1; the calling DCE then transmits its INFO1 50 ms after the end of the answering DCE's. The calling DCE's INFO1 cannot carry what it measured before it has received anything, which is why it comes last. A DCE that misses its peer's INFO1 takes the loading from the peer's first MP (clause 9.5), which has INFO1's layout. The probing signal's level is −12 dBFS.

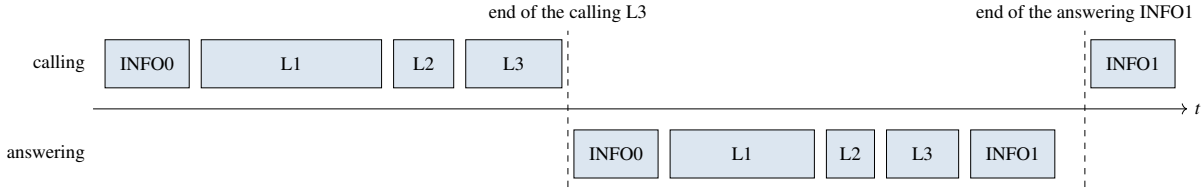


Figure 3 – Phase 2: the probing signals (durations not to scale)

9.3 Modulation of the INFO and MP messages: the message mode

INFO0, INFO1 and the MP messages of clause 9.5 shall be transmitted in the *message mode*, which is:

a) where both DCEs indicated the parametric mode in CM and JM (Annex A), the parametric mode with the 20 ms symbol (P20, clause 10.7): the preamble of that clause, then the message's bits four to a lag symbol as clause 11.4 maps them, unscrambled, padded with flags to whole symbols;
b) otherwise the *robust mode* of clause 10.1: the FSK of [ITU-T V.21] at 300 bit/s, channel 1 from the calling DCE and channel 2 from the answering DCE as [ITU-T V.8] uses them for CM and JM.

A message is an HDLC frame (flags, FCS) of [ITU-T V.42] with the message's octets as the information field. Each message is sent twice; the receiver takes the first copy whose FCS checks. V.21's bit error ratio was measured below 10^{-2} through every mode of AMR but 4.75 kbit/s, of AMR-WB but 6.6, and of EVS but 9.6 wideband (Appendix I, Table I.3); the parametric mode's lag symbol error ratio was zero through all of them but AMR 4.75 (Table I.8), which is why it carries the messages where both DCEs have it. Measured end to end, through a receiver of two filters matched over the bit with its own bit clock (Appendix IV), V.21's bit error ratio was 1 to $3 \cdot 10^{-2}$ at AMR 5.15 to 6.7 kbit/s: messages in V.21 are surest from AMR 7.4 kbit/s up. A receiver may retry a P20 message whose FCS fails with the second most likely lag on the few least reliable symbols.

NOTE – V.34 modulates INFO and MP at 600 bit/s with differential phase modulation on two tones. Through AMR at 7.4 kbit/s and below no differential phase modulation measured kept a symbol error ratio below 10^{-1}, while 300 bit/s FSK did (Table I.3); so this Recommendation sends its messages as V.8 sends CM and JM, or on the codec's pitch lag, which is surer still.

9.4 The loading

From its peer's L1 and L2 the receiving DCE shall estimate, for each tone of the comb and each of the two symbol periods, the root mean square phase error of the received tone symbols against the known sequence, and for each tone the mean received level and the rms change of its magnitude from one symbol to the next (in dB). It shall then choose:

a) the band: the tones whose level is within 20 dB of the strongest tone's; tones outside are not loaded (this finds the codec's band);

b) the comb and the symbol period: those giving the greatest number of bits per second over the tones loaded by c); a comb of fewer tones, more widely spaced, wins where the codec's noise against each tone of a denser comb is too high (Appendix I);
c) the modulation order of each tone, from its phase error σ: DQPSK where $\sigma \leq 20°$, 8-DPSK where $\sigma \leq 9°$ and the mode set allows it, none where $\sigma > 20°$; and on a loaded tone, from the rms change σ_A of its magnitude from one symbol of L1 to the next, two amplitude rings where $\sigma_A \leq 0.8$ dB and four where $\sigma_A \leq 0.4$ dB, where both DCEs have the amplitude loading of clause 10.11 and the mode set allows it;
d) the code rate of the Reed-Solomon code (Annex C) from the measured tone symbol error ratio over the loaded tones: rate 7/8 below 10^{-3}, 3/4 below 10^{-2}, 1/2 otherwise (in the parametric mode, from the lag symbol error ratio of L3 the same way);
e) the mode: the comb of b) where it loads four tones or more and carries more after the code of d) than the parametric mode would; otherwise the parametric mode (clause 10.7), with the 10 ms symbol where no lag symbol of L3's P10 part was in error, else the 20 ms symbol where at most one of its P20 part's 40 symbols was (at the code rate 1/2 where one was); else no data mode: the DCE asks in MP for no loading (Annex B) and both DCEs wait in the message mode, retraining as clause 9.6 says, until the codec's mode carries one (AMR at 4.75 kbit/s is the one codec mode measured that carries none; a network leaves a call there only at the worst coverage).

The thresholds in c) give a tone symbol error ratio of about 10^{-2} at the threshold for Gaussian phase error, which the code of Annex C corrects to below 10^{-6} at rate 3/4. The loading is sent to the peer in INFO1 and MP as a bitmap of the comb with two bits a tone (none, DQPSK, 8-DPSK), with a second such bitmap for the amplitude rings where any tone has them (Annex B), or as the parametric mode and its symbol period.

9.5 Phase 3: parameter exchange

Each DCE transmits an **MP** message with the loading it asks for in the direction it receives, the symbol period and the code rate, in the manner of [ITU-T V.34]'s MP sequence; it shall transmit MP until it has received its peer's MP with the acknowledgement bit set, then once more with its own acknowledgement bit set, and begin data transmission (phase 4) with the loading its peer asked for. Both directions enter phase 4 within 100 ms of each other.

9.6 Renegotiation and retrain

During phase 4, a DCE whose receiver measures a phase error or an error ratio over the thresholds of clause 9.4 on tones it loaded, or well under them on tones it did not load, should ask for a *renegotiation*: it sends the control sequence of clause 10.8 and the DCEs exchange MP messages in the message mode (clause 9.3) with the new loading, then resume data; the code's interleaver is flushed and restarted. A DCE that loses the symbol timing, or the signal, or decodes no codeword for 2 s (or 2.5 interleaver blocks where that is longer) shall ask for a *retrain*: the control sequence, then MP messages with the retrain bit set (Table B.2); after the exchange both DCEs return to phase 2, the calling DCE's INFO0 3 s after it (so that the answering DCE's last MP is through). The codec's mode changes with the radio conditions, and the loading shall follow it; a renegotiation shall take no more than 500 ms and a DCE shall not refuse one.

A DCE whose receive direction is in the parametric mode, or in no data mode (clause 9.4 e)), has nothing that measures the combs: it should ask for a retrain 30 s after entering that state and every 60 s while it lasts, so that a codec mode that has improved is found and used. In the parametric mode it shall keep $\bar{c}$, the mean over the last 2 s of the decided lags' normalized autocorrelation c (clause 10.7). While $\bar{c}$ is below 0.85 it need not ask for those retrains: no codec mode measured that gives $\bar{c}$ below 0.85 carries a comb. When $\bar{c}$ rises to 0.85 or more from below, it should ask for a retrain at once rather than at the next 60 s, though not in the first 30 s of phase 4. The threshold has 0.02 of hysteresis: $\bar{c}$ is below it once under 0.83 and rises through it at 0.87, so that $\bar{c}$ wavering about 0.85 does not retrain again and again. At or

above 0.85 the periodic retrains stay: $\bar{c}$ there does not say whether a comb would pass.

NOTE – $\bar{c}$ does not order the codec modes by what they carry, so it cannot replace the retrain (Appendix I, Table I.9). Every mode measured in which a comb loads four tones or more gave $\bar{c}$ of 0.875 or more (AMR 10.2 the least), but modes in which none does gave as much as 0.938 (EVS 9.6) and 0.920 (AMR 7.95 loads five): the long-term predictor of a codec at a low mode may code the pulse train's period better than one at a higher mode that spends its bits on the excitation. Below 0.85 are only the lowest modes (AMR 4.75 to 5.9, AMR-WB 6.6, EVS 7.2 and 8.0), none of which carries a comb, and $\bar{c}$ over 2 s varies by at most 0.075 there; so $\bar{c}$ saves the retrains where they cannot succeed, and its rise through 0.85, a codec mode leaving that range, is the earliest sign one may.

10 Modulation

10.1 Signal set

The signal set has four modes, Table 2. A DCE shall support the robust mode and the narrowband mode; it should support the parametric mode, which alone carries data through the lower codec modes and carries the messages of clause 9.3 between DCEs that have it; it may support the wideband mode (which needs the signal interface at 16000 Hz).

NOTE – The modes were chosen from measurements (Appendix I). Through AMR at 7.4 kbit/s and below, no differential phase modulation measured kept a symbol error ratio below 10^{-1}, on one carrier or many; only 300 bit/s FSK did. Through AMR at 12.2 and AMR-WB at 23.85 kbit/s, a comb of 16 tones at 200 Hz spacing with DQPSK and a symbol of 20 to 22.5 ms kept its tone symbol error ratio at or below 10^{-2}; combs of 32 tones at 100 Hz spacing did not (the codec's noise against each tone grows faster than the number of tones), and 8-DPSK held only on the better tones. Fewer tones, more widely spaced, keep working where more do not: through AMR at 10.2 kbit/s and through two AMR codecs in tandem the 11-tone comb at 300 Hz spacing kept $3 \cdot 10^{-3}$ while the 16-tone comb went to $3 \cdot 10^{-2}$, and through AMR-WB at 15.85 the 300 Hz comb loads 16 tones where the 200 Hz comb loads 10; so the signal set has three combs for each band and the probing chooses. A single carrier at 600 baud (V.22's) passes AMR 12.2 as well, at 1200 bit/s; the comb carries a third more at the same symbol error ratio and loads itself to the codec. The parametric mode came from the same measurements, which show that the one parameter every codec mode but AMR 4.75 transmits to the sample is the long-term predictor's lag (clause 10.7, Table I.8).

10.2 The comb

The comb's tones are at the frequencies

$$f_k = k \cdot \Delta f, \qquad k = k_{\min}, \ldots, k_{\max},$$

as Table 3 gives them. The narrowband combs N200, N300 and N400 are for the signal interface at 8000 Hz, the wideband combs W200, W300 and W400 for 16000 Hz, and the super wideband combs S200, S300 and S400 for 32000 Hz (for the EVS codec's super wideband modes, Appendix II). The probing (clause 9) decides which tones are loaded; the transmitter sends nothing on a tone that is not. The transmitted signal in the useful part of symbol n is

$$s_n(t) = A \sum_{k \in \mathscr{L}} \cos\bigl(2\pi f_k t + \varphi_{n,k}\bigr), \qquad 0 \le t < T_u,$$

where $\mathscr{L}$ is the set of loaded tones, $\varphi_{n,k}$ the tone's phase in this symbol and A the amplitude that gives the level of clause 7.2 b) over the loaded tones (with the amplitude rings of clause 10.11 the tone's amplitude is $A a_{n,k}$, of the same mean power). The useful period T_u is a whole number of cycles of Δf, so that the tones are orthogonal over it.

10.3 Symbol timing and the cyclic prefix

A symbol consists of a cyclic prefix of T_g followed by the useful part of T_u: the prefix is the last T_g of the useful part, transmitted first, so that the signal is continuous across each symbol except for the phase steps at the symbol's start. The symbol periods are

Table 2 – The modes of the signal set

	Robust (R)	**Narrowband (N)**	**Wideband (W)**	**Parametric (P)**
Signal	the FSK of [ITU-T V.21]: channel 1 (980/1180 Hz) from the calling DCE, channel 2 (1650/1850 Hz) from the answering DCE, 300 bit/s	a comb of tones at 200, 300 or 400 Hz spacing in the band 300 to 3400 Hz (16, 11 or 8 tones), DPSK	a comb of tones at 200, 300 or 400 Hz spacing in the band 300 to 6900 Hz (33, 23 or 17 tones), DPSK	a pulse train whose period is the symbol (clause 10.7): the codec transmits it as the lag of its long-term predictor
Symbol	3.33 ms (one bit)	20 ms (a codec frame): a useful part of whole cycles of the spacing and a cyclic prefix (clause 10.3); the 200 Hz comb also 22.5 ms	as N	20 ms (P20) or 10 ms (P10); the pulses restart at each symbol's start
Modulation	binary FSK	DQPSK, or 8-DPSK on tones with a low enough phase error; two or four amplitude rings on tones whose amplitude the codec keeps (clause 10.11)	as N	16 periods, four bits a symbol: 40 to 70 samples at 8000 Hz in steps of 2 (P20), 20 to 35 (P10)
Raw rate	300 bit/s	1600 bit/s (16 tones at 200 Hz, DQPSK, 20 ms), 1100 (11 at 300 Hz), 800 (8 at 400 Hz); half as much again with 8-DPSK	3300 bit/s (33 tones at 200 Hz, DQPSK, 20 ms), 2300 (23 at 300 Hz), 1700 (17 at 400 Hz); half as much again with 8-DPSK	200 bit/s (P20), 400 bit/s (P10)
Use	INFO and MP messages and control sequences (the message mode, clause 9.3) where a DCE lacks the parametric mode; data where no other mode passes	data	data	the message mode between DCEs that have it; data where no comb loads four tones
Passes (measured)	every AMR mode but 4.75, every AMR-WB mode but 6.6, every EVS mode but 9.6 wideband	AMR 12.2 (every comb), AMR 10.2 and two AMR 12.2 codecs in tandem (the 300 and 400 Hz combs), AMR-WB from 12.65, EVS 13.2 (N400) and 24.4 (every comb, 8-DPSK)	AMR-WB 23.85 (28 tones of W200), 15.85 (16 of W300), 12.65 (7 of W300); EVS 24.4 (W300), 48 and above (every comb, 8-DPSK); the super wideband combs through EVS at 48 kbit/s and above	P20 without a symbol in error through every AMR mode but 4.75, AMR-WB 23.85 to 6.6, EVS 9.6 to 24.4 narrowband and two codecs in tandem; P10 without error through AMR 12.2, AMR-WB 23.85 and 12.65, EVS 9.6 to 24.4 and the tandems, and with 1 to 5 per cent of its symbols in error through AMR 10.2 to 7.4 and AMR-WB 6.6

Table 3 – The combs

Code	Comb	Δf (Hz)	k	Tones	Band (Hz)	T_u (ms)	T_g (ms)
0	N200	200	2–17	16	400–3400	15 (or 20)	5 (or 2.5)
1	N300	300	1–11	11	300–3300	10	10
2	N400	400	1–8	8	400–3200	15	5
3	W200	200	2–34	33	400–6800	15	5
4	W300	300	1–23	23	300–6900	10	10
5	W400	400	1–17	17	400–6800	15	5
6	S200	200	2–69	68	400–13800	15	5
7	S300	300	1–46	46	300–13800	10	10
8	S400	400	1–35	35	400–14000	15	5
9–15	–	reserved					

- **20 ms**, the period of a codec frame, for every comb: T_u and T_g as Table 3 gives them (three cycles of the spacing for the 200 and 300 Hz combs, six for the 400 Hz combs), so that one symbol's phase steps fall in each frame and the prefix spans the codec's transient after them (a subframe of 5 ms, or two for the 300 Hz combs);
- **22.5 ms**, for the N200 comb only: $T_u = 20$ ms (four cycles), $T_g = 2.5$ ms, for codecs that gain from the longer integration (the probing compares the two).

The receiver shall integrate each tone over the useful part only. A receiver shall acquire the symbol timing from the cyclic prefix (the correlation of the signal with itself T_u earlier peaks over the prefix) and shall re-acquire it within two symbols after a step of the received timing of up to ± 20 ms, which a jitter buffer's adaptation causes (Appendix III). Differential modulation makes the symbol after such a step the only one lost.

10.4 The codec frame and the symbol

The symbols are not aligned to the codec's frames, and no alignment is signalled: the position of the symbols' phase steps within the frame was found to make no difference to the tone symbol error ratio (Appendix I, Table I.6: the 20 ms symbol at a frame boundary and 10 ms off it), which the codec's 5 ms subframes and its long-term predictor explain: the step is coded in the subframe it falls in, wherever that is. The 20 ms symbol is chosen for its rate (12 per cent more than the 22.5 ms symbol) and for its match with the frame erasures: one erased frame is one symbol.

10.5 Differential phase modulation

Each loaded tone carries, per symbol, the phase difference $\varphi_{n,k} - \varphi_{n-1,k}$, one of four values for DQPSK or eight for 8-DPSK, mapped from the bits as Table 6 shows. The first symbol of phase 4, and the first after a renegotiation, carries no data: it is the phase reference, and it is sent three times (the codec's first frames out of silence are its own, and the receiver takes the last of the three). The receiver decides each tone symbol from the difference of the tone's phases in two consecutive symbols; no carrier recovery and no equalization are required, and a frequency offset between the DCEs' clocks of Δf_c appears as a constant phase rotation of $2\pi\Delta f_c T$ per symbol, which the receiver should estimate from the decisions and remove.

10.6 Loading and raw rate

The raw rate is

$$R = \frac{1}{T_u + T_g} \sum_{k \in \mathcal{L}} b_k,$$

with $b_k = 2$ for a DQPSK tone and 3 for an 8-DPSK tone, and one or two more with the amplitude rings of clause 10.11: 1600 bit/s for the 16 DQPSK tones of N200 at the 20 ms symbol, 1100 for N300, 800

for N400, 3300 for W200, and half as much again with 8-DPSK on all tones. The loading of clause 9.4 chooses $\mathscr{L}$ and the b_k; a loading with fewer than 4 tones shall not be used (the DCE takes the parametric mode, clause 9.4 e)).

10.7 The parametric mode

10.7.1 Principle

The parametric mode carries the data not in a waveform but in a parameter the codec must transmit to describe the signal at all: the period of a periodic signal, which an ACELP codec codes as the lag of its long-term predictor in every subframe, at its lowest modes as at its highest (clause 6 h)). The transmitter sends a pulse train whose period is the symbol; the decoder reproduces a signal with that period; the receiver measures it. Neither DCE touches the codec or its bitstream, and the mode is a signal at the interface of clause 7.2 like the others.

10.7.2 Signal

In symbol n, of T samples, the transmitter sends pulses L_n samples apart, the first at the symbol's start:

$$e(m) = \sum_{j\geq 0,\ jL_n<T} \delta(m - nT - jL_n), \qquad x(m) = e(m) + 0.7\,x(m-1), \qquad s = A\,(x * h),$$

where L_n is the symbol's lag from the alphabet of Table 4; the one-pole filter gives the pulses the spectral tilt of a glottal pulse (about -6 dB per octave above 400 Hz, so that the codec's classifier takes the signal for voiced speech); and h is a band-pass pulse of 33 taps at 8000 Hz, the impulse response of an ideal band-pass from 300 Hz to the top of the codec's band (3400 Hz at the 8000 Hz interface) under a Hamming window. A sets the level: -15 ± 2 dBFS rms over a symbol; the signal's peak factor is 13 dB, and its peaks shall stay below -1 dBFS (clause 7.2 b)). At the 16000 Hz interface the lags and the pulse have twice the samples, so that the periods in time are the same.

10.7.3 Lag alphabets

Table 4 – The lag alphabets (samples at 8000 Hz) and the mapping of the bits

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bits (Gray code)	0000	0001	0011	0010	0110	0111	0101	0100	1100	1101	1111	1110	1010	1011	1001	1000
P20: lag L ($T = 160$)	40	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70
P10: lag L ($T = 80$)	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35

Each alphabet has 16 lags: four bits a symbol. No lag of an alphabet is twice another or more, so that a period cannot be read as its own multiple, and every lag is at most $T - 16$, so that the receiver sees more than one period in a symbol. The P20 lags are inside the range of every codec's long-term predictor (AMR: 18 to 143 samples); the shortest P10 lags are below AMR-WB's shortest (34 samples at 12.8 kHz, 2.66 ms) and pass it nonetheless (Table I.8): a signal of period L has period $2L$ as well. Lags above 70 samples were carried less surely by every codec and are not used (Appendix I).

10.7.4 Symbol timing

A transmission in the parametric mode – a message (clause 9.3), a part of the probing signal L3 (Annex B) or the data of phase 4 – shall begin with the *preamble*: eight symbols alternating the alphabet's lowest and highest lag (index 0 and index 15), the lowest first. The receiver shall acquire the symbol timing from the preamble: the signal is periodic within each symbol and not across the symbol boundaries, so the symbol phase at which the normalized autocorrelation at the decided lag, summed over the preamble's symbols, is greatest is the transmitter's. The receiver should track the timing on the data symbols in the same way, for the clock difference of clause 7.2 d), and shall re-acquire it after a step of the received timing (clause 10.3); in P20 a step of one codec frame is one symbol.

10.7.5 Receiver

The receiver shall decide each symbol's lag from the decoded samples y of the symbol: for each lag L of the alphabet the normalized autocorrelation

$$c(L) = \frac{\sum_{m=0}^{T-L-1} y(m)\, y(m+L)}{\sqrt{\sum_{m=0}^{T-L-1} y(m)^2 \; \sum_{m=0}^{T-L-1} y(m+L)^2}},$$

and the lag of the greatest $c(L)$; where a smaller lag's c is within 0.02 of the greatest, the smaller lag should be taken (as the measurements of Appendix I did). The receiver shall deliver, per symbol, the decided index's bits and a reliability (the margin of the decided c over the next), as clause 11.4 requires of a tone symbol. Nothing of the codec's own pitch estimate is used: the receiver has only the decoded samples.

10.7.6 Raw rate

Four bits a symbol: 200 bit/s with the 20 ms symbol, 400 with the 10 ms symbol. The probing chooses the symbol period, and chooses the parametric mode only where no comb loads four tones (clause 9.4 e)): a comb of four DQPSK tones carries what P10 does, and every comb that passes a codec mode carries more.

NOTE – The mode rests on one measured fact (Appendix I, Table I.8): with the P20 alphabet, the period of the pulse train came through every mode of AMR but 4.75 kbit/s, every mode of AMR-WB measured down to 6.6, EVS from 9.6 kbit/s and two codecs in tandem without a symbol in error in 198, while the combs carry nothing through AMR 7.4, AMR-WB 6.6 or EVS 9.6 and V.21's FSK has a bit error ratio of $4 \cdot 10^{-3}$ to $6 \cdot 10^{-3}$ through AMR 7.4 and up to $3 \cdot 10^{-2}$ through EVS 9.6. A codec at those modes spends its bits on the spectral envelope and the pitch and codes the lag to the sample (AMR: 8 or 9 bits in a frame's first subframe and 4 to 6 relative to it in the others, at every mode); what it drops is the excitation's detail, which is where a waveform's phases live. With frame erasures the mode's symbol error ratio is about twice the erasure ratio: the erased frame, whose concealment repeats the previous lag, and often the frame after it.

NOTE – The two layers – the comb's phases and the pulse train's period – were also measured as one waveform, with the comb's spacing as the pitch: a symbol of one period of L samples repeated, periodic at L so that the lag carries four bits, with a spectrum of harmonics of f_s/L whose phases carry the comb's bits as differences across frequency, so that L may change every symbol (Appendix I, Table I.10). It works and is not worth having. The lag's four bits cost the harmonics a reference and come to 1529 bit/s against the comb's 1600 through AMR 12.2, AMR-WB 23.85 and EVS 24.4; where a comb is marginal the stacked signal is worse (a tone symbol error ratio of 0.154 against 0.028 through AMR 10.2), because the codec's phase error is more correlated from one symbol to the next on a tone than from one tone to the next in a symbol, so that a difference across frequency loses to [ITU-T V.22]'s difference in time; and a harmonic, voiced-looking comb is not carried better than one with random phases (both 10^{-3} at 200 Hz spacing through AMR 12.2; the harmonic one worse through AMR 7.4). The modes are therefore separate, and the probing chooses.

10.8 Control sequences during data

A DCE that wishes to renegotiate (clause 9.6) shall send, in place of data, the control sequence, then its MP in the message mode (clause 9.3). In the comb modes the control sequence is eight consecutive symbols with the phase difference 180° on every loaded tone; a receiver that decodes four such symbols in a row shall stop delivering data to the inner decoder, flush, answer with the same sequence if it has not sent one, and proceed to the MP exchange. The sequence cannot occur in scrambled data for more than two symbols with a probability above 10^{-9} on 8 tones. In the parametric mode the control sequence is sixteen consecutive symbols at the alphabet's lowest lag (index 0), and a receiver that decodes ten in a row acts as above (ten equal symbols occur in scrambled data with a probability of 2^{-36} at any position).

A receiver shall count control symbols only once it has decoded a block of its peer's data in phase 4 (the peer's data is then *confirmed*), and only while the received signal is at the level of data: an MP's zero

octets are runs of lag index 0, and silence decodes as anything. The first MP of its peer's that a DCE receives in phase 4 is the peer's last of the exchange (clause 9.5), which may end after the DCE has begun phase 4: the peer's phase 4 begins at its end, and the receiver shall look for the data from there. A DCE that receives a further MP in phase 4 (the peer lost the last MP of the exchange and repeats its own) shall answer it with its own MP, its acknowledgement bit set, and begin phase 4 again.

10.9 Scrambling

The information octets, before the inner code (clause 11), shall be scrambled with the generating polynomials of [ITU-T V.34]: $1+x^{-18}+x^{-23}$ in the calling DCE's transmitter and $1+x^{-5}+x^{-23}$ in the answering DCE's (GPC and GPA), self-synchronizing as in that Recommendation. The scrambler is reset at the start of phase 4 and after each renegotiation.

NOTE – A self-synchronizing descrambler turns one line error into three (the error and its two taps). After the inner code that would cost a codeword of the parametric mode three or four octets for one lag symbol in error; before it, the decoder sees the line's errors as they are, and LAPM sees the descrambler's tripling only of what the code left.

10.10 Quick start

A DCE may keep, for the last number dialled or answered, the loading and the symbol period of the last call, and offer them in INFO0 as the mode to start with. If its peer accepts in its INFO0, phase 2 is reduced to L1 of 0.5 s for a check of the loading (or, where the loading kept is the parametric mode, to the part of L3 for its symbol period), and the DCEs proceed to phase 3; this is in the manner of [ITU-T V.92]'s quick connect. A loading so offered shall still be checked by the receiver against the thresholds of clause 9.4.

10.11 Loading with amplitude

10.11.1 Signal

A loaded tone may carry, besides its phase difference, a difference of its amplitude from one symbol to the next: differential amplitude and phase shift keying. The tone's amplitude in symbol n is one of R rings, $R = 2$ or 4:

$$a_{n,k} = g_R\, 10^{\rho_{n,k}\Delta_R/20}, \qquad \rho_{n,k} = (\rho_{n-1,k} + d_{n,k}) \bmod R,$$

with the rings $\Delta_2 = 6$ dB apart for two rings and $\Delta_4 = 3$ dB apart for four, the tone's amplitude digit $d_{n,k} \in \{0,\ldots,R-1\}$, and g_R the factor that gives the tone the mean power of a tone without rings, $g_R = \left(\frac{1}{R}\sum_{r=0}^{R-1} 10^{r\Delta_R/10}\right)^{-1/2}$. The transmitted signal of clause 10.2 is then

$$s_n(t) = A \sum_{k\in\mathscr{L}} a_{n,k} \cos\left(2\pi f_k t + \varphi_{n,k}\right),$$

with $a_{n,k} = 1$ on tones without rings. The phase reference symbol (clause 10) has $\rho = 0$ on every tone. A tone with rings carries its phase bits (Table 6) and then its amplitude bits: one for two rings (d the bit), two for four, Gray coded (00 is $d = 0$, 01 is 1, 11 is 2, 10 is 3).

10.11.2 Receiver

The receiver decides d from the ratio of the tone's magnitudes in two consecutive symbols, as it decides the phase from their phase difference, with no amplitude reference:

$$\hat{d}_{n,k} = \operatorname{round}\left(\frac{20\log_{10}\left(|r_{n,k}|/|r_{n-1,k}|\right)}{\Delta_R}\right) \bmod R,$$

where $r_{n,k}$ is the tone's integral over the useful part; the reliability of clause 11.4 is the distance of the ratio (in dB) from the nearest decision boundary.

10.11.3 Loading

The loading of clause 9.4 gives a tone two rings where the rms change σ_A of its magnitude from one symbol of L1 to the next (whose amplitude is constant) is at most 0.8 dB, and four rings where it is at most 0.4 dB, on the phase order that clause chose for the tone; it does so only where both DCEs indicated the capability in CM and JM (Annex A) and the mode set allows it. The raw rate's b_k (clause 10) is then the tone's phase bits plus one for two rings or two for four: 4 for 8-DPSK with two rings, 5 with four. A DCE that does not receive its peer's MP with the rings it asked for loads without them.

NOTE – The parameters come from the measurements of Appendix II (Table II.4). Of 725 tones measured through EVS (24.4 to 128 kbit/s), AMR and AMR-WB, two rings 6 dB apart held (a tone symbol error ratio below $2 \cdot 10^{-2}$ for the phase and for the rings) on 99 per cent of those whose σ_A was at most 0.8 dB and on 17 per cent of the others; four rings 3 dB apart held on all of those at or below 0.4 dB and on fewer than half of those up to 0.8. A tone's σ_A is therefore the loading's one criterion, as its phase error is for the phase order, and it finds the tones without a list of codecs: through EVS at 96 and 128 kbit/s the rings add 21 to 67 per cent to the comb's raw rate (15 100 bit/s on S200 at 128 kbit/s, against 10 350 with phase alone), through EVS at 48 kbit/s wideband and 24.4 kbit/s narrowband 19 to 51 per cent, through AMR-WB at 23.85 and AMR 12.2 a few tones of the 300 and 400 Hz combs, and through AMR 10.2 none. A coherent constellation (QAM with a trellis code) was not taken: it needs the reference and the equalization the comb's differential modulation does without, and the codec's phase error, correlated from one symbol to the next, is what differential detection cancels.

11 Error control and framing

11.1 Structure

The data of phase 4 pass, from the DTE towards the line, through:

1. [ITU-T V.42]: LAPM, with the zstd, V.44 or V.42bis compression the DCEs negotiate in XID (clause 12.1);
2. the inner code: the Reed-Solomon code of Annex C over the HDLC frames' octets, as a continuous octet stream (the LAPM frames with their flags are the stream; no further framing is added);
3. the interleaver of Annex C, which spreads each codeword over several modulation symbols so that an erased codec frame costs each codeword a fraction of its symbols (its depth is 1 in the parametric mode);
4. the mapping of the coded bits onto the tone symbols, or the lag symbols of the parametric mode, of clause 11.4.

11.2 The code

The inner code shall be the Reed-Solomon code over GF(2^8) of [ITU-T G.975] (the field of the primitive polynomial $x^8 + x^4 + x^3 + x^2 + 1$, the code generator of its clause on the RS(255,239) code), shortened to the codeword lengths of Annex C: a codeword of n octets with k information octets and $n - k = 2t$ check octets, correcting t octets in error or $2t$ octets erased. The code rates of Table 5 are offered (Annex C shortens the codewords to 32 octets, so that the interleaver's delay stays short). A receiver should mark the octets of an erased codec frame as erasures (clause 11.3).

Table 5 – Code rates of the inner code

Rate	Codeword n	Information k	Corrects t octets / $2t$ erasures
7/8	32	28	2 / 4
3/4	32	24	4 / 8
1/2	32	16	8 / 16

11.3 Erased frames

A speech decoder conceals a lost or bad frame by extrapolating the previous frame's parameters: the output over the erased frame keeps the spectral envelope and the pitch of the signal before it and loses its information. A receiver of this Recommendation should detect an erased frame – from the phase errors of all its tones at once, or from the tones' levels decaying together as concealment attenuates – and mark the tone symbols of the symbol or symbols it spans as erasures to the Reed-Solomon decoder, which corrects twice as many erasures as errors. With the interleaving depth of Annex C one erased 20 ms frame erases one symbol in each of D consecutive codewords' D symbols, and is corrected at any code rate of Table 5; frame erasure ratios of the mobile networks are below 2 per cent ([3GPP TS 26.114]'s design targets), and the measured effect is in Appendix I. In the parametric mode the concealment repeats the previous lag: the erased frame's symbol decodes as the symbol before it, and the frame after it is often wrong as well (the measured cost is two symbols an erasure, Table I.8); the receiver should detect the erasure from the level, which the concealment attenuates, since the lag gives no sign of it.

11.4 Mapping

The interleaver's output bits shall be taken in order and mapped onto the loaded tones of a symbol from the lowest tone to the highest, two bits to a DQPSK tone and three to an 8-DPSK tone, Gray coded onto the phase difference as Table 6 shows, and on a tone with amplitude rings one or two bits more, onto its amplitude digit as clause 10.11 says. Bits left over at the end of a symbol are carried into the next. The receiver shall deliver, per tone symbol, the hard decision and an indication of its reliability (the distance of the received phase difference from the decision's, and of the magnitude ratio from its decision boundary), so that the Reed-Solomon decoder may be given erasures. In the parametric mode the bits are mapped four to a lag symbol, the first bit the most significant, Gray coded onto the alphabet's index as Table 4 shows; the receiver delivers the decision and its reliability as above.

Table 6 – Gray mapping of the phase difference

Phase difference	0	45	90	135	180	225	270	315
DQPSK bits	00	–	01	–	11	–	10	–
8-DPSK bits	000	001	011	010	110	111	101	100

12 Link layer, compression and DTE interface

12.1 Link layer

Above the inner code the DCEs shall operate [ITU-T V.42] LAPM. The detection phase of [ITU-T V.42] (ODP and ADP) is not needed and shall not be used: both DCEs have indicated LAPM in the V.8 session; the originator shall start LAPM's XID exchange on entering phase 4. The DCEs should negotiate [ITU-T V.44] compression, or [ITU-T V.42bis] where V.44 is not available, in XID as those Recommendations specify. Where both DCEs indicate the preset dictionary of Annex E in INFO0 (Table B.1) and XID agrees on V.44, both shall start V.44 from it as that annex specifies. Where both indicate zstd in INFO0, both shall offer it in XID, and where XID agrees on it the call's compression is zstd, as Annex F specifies, in place of V.44 or V.42bis. LAPM's T401 shall be set from the round-trip delay measured in phase 2 (clause 9.2), with at least 1 s added for the interleaver's delay and the symbol period, and 3 s in the parametric mode (a codeword takes 1.3 s at 200 bit/s); N401 should not exceed 64 octets at raw rates below 2400 bit/s, so that a frame fits in an interleaver block.

NOTE – The code of Annex C leaves LAPM an error ratio below 10^{-6} in the loaded modes; LAPM is retained for the residual errors, for the erased frames the code cannot cover, for flow control and for compression.

12.2 DTE interface

The DCE's DTE interface shall conform to [ITU-T V.250], with:

a) in +MS (modulation selection), the value V_CVM for this Recommendation's modulation, with its automode fallback to V.22 and V.21 (+MS=V_CVM,1); a DCE should make it its default on a speech path known to be a codec's. +MS? reports the value, +MS=? includes it in the list.
b) the extended command +CVM (Table 7) to constrain the mode set and to read the loading and the measurements of the current connection.
c) in the CONNECT result code, with \V1 or X settings that give the rate, the coded data signalling rate (clause 5) and the text /CVM; the raw rate is read with +CVM?.
d) the result code +CVMR: <raw>,<coded> (unsolicited, if enabled by +CVM) on each renegotiation.

Table 7 – The +CVM command

Command	Meaning
+CVM=<modes>[,<rate>[,<report>]]	<modes>: the modes allowed, a sum of 1 (robust), 2 (narrowband), 4 (wideband), 8 (8-DPSK permitted), 16 (the amplitude rings of clause 10.11 permitted), 32 (parametric); default all. <rate>: the signal interface rate, 8000 or 16000. <report>: 1 enables +CVMR.
+CVM?	+CVM: <modes>,<rate>,<report> and, while connected, a second line +CVM: <raw tx>,<raw rx>,<coded tx>,<coded rx>,<symbol ms>,<tones tx>,<tones rx>,<code rate>,<renegotiations>,<erasures> (the tones are 0 in the parametric mode).
+CVM=?	the ranges.

12.3 S-registers

S-register 44 (as a DCE of this Recommendation uses it) holds the number of tones loaded in the receive direction of the last call, S-register 45 the number of renegotiations of the last call. Both are read-only; the loading kept for the quick start of clause 10.10 is the DCE's own.

13 Interworking

13.1 With V series DCEs on the PSTN

A DCE of this Recommendation falls back to [ITU-T V.22] and [ITU-T V.21] (clause 8). Through a gateway between the mobile network and the PSTN the speech path has one codec (the mobile side's); V.22 passes AMR at 12.2 kbit/s with its data intact and V.21 passes every mode measured (Appendix I). A gateway that detects the V.8 session and switches its IP leg to [ITU-T G.711] or to the modem relay of [ITU-T V.150.1] changes nothing for this Recommendation: the codec that limits the path is the mobile side's.

13.2 With DCEs of this Recommendation across two codecs

A mobile-to-mobile call, or a call through a transcoding gateway, has two or more codecs in tandem in each direction. The probing of clause 9 measures the tandem as one channel; the loading follows the measurement. The throughput through two codecs is lower than through one (Appendix I gives the measured figures for AMR after AMR), and the retrain timers allow for the longer delay.

13.3 With facsimile

Facsimile terminals are served through Annex D: the DCE terminates the [ITU-T T.30] session on its side and carries it as [ITU-T T.38] packets in the data link. The modulations of a facsimile terminal (V.27ter, V.29, V.17) do not pass the codecs (Table I.2).

14 Performance

A DCE of this Recommendation shall meet the following, measured through the reference codecs of Appendix I with no frame erasures and the signal interface at 8000 Hz, with the method of that appendix:

a) through AMR at 12.2 kbit/s it shall connect in a narrowband mode with a raw rate of at least 1200 bit/s in each direction, and with the Reed-Solomon code at rate 3/4 or better;
b) through AMR at 7.95 and 7.4 kbit/s and AMR-WB at 6.6 kbit/s it shall connect in the parametric mode with a raw rate of at least 200 bit/s in each direction, and through EVS at 9.6 kbit/s narrowband of at least 400 bit/s (no mode of this Recommendation passes AMR at 4.75 kbit/s, whose radio conditions a network leaves only at the worst coverage: the DCEs wait in the message mode for a better codec mode);
c) through AMR-WB at 23.85 kbit/s with the interface at 16000 Hz it shall connect in a wideband mode with a raw rate of at least 2400 bit/s in each direction;
d) through EVS at 13.2 kbit/s narrowband it shall connect in the narrowband mode (the N400 comb) with a raw rate of at least 800 bit/s; through EVS at 24.4 kbit/s narrowband with a raw rate of at least 2000 bit/s; through EVS at 48 kbit/s wideband, with the interface at 16000 Hz, of at least 4000 bit/s;
e) the error ratio delivered to LAPM shall be below 10^{-6} at the loading chosen by clause 9.4;
f) a frame erasure ratio of 1 per cent (independent erasures) shall not lower the coded rate in a comb mode by more than the code rate step of Table 5 (in the parametric mode an erasure costs two symbols, and the code rate 1/2 follows).

The rates achievable are in Appendices I and II.

Annex A

Indication of this Recommendation's capabilities in V.8 and V.8bis

(This annex forms an integral part of this Recommendation.)

A.1 Encoding in CM and JM

The CM and JM signals of [ITU-T V.8] are sequences of octets grouped in categories (call function, modulation modes, protocols, GSTN access, PCM modem availability, non-standard facilities). This Recommendation needs one bit of the modulation modes category to say "V.cvm available", and an information field of two octets to say what of it. The bit is an allocation that ITU-T makes in a revision of [ITU-T V.8]; until it is made:

a) the DCE shall send the information field of clause A.2 in the *non-standard facilities* category of CM and JM, as [ITU-T V.8] provides for it, with the country code and the provider code of [ITU-T T.35] that the implementer holds, followed by the octets of Table A.1 with the identifier octet `0x43` ('C') first, so that a DCE of this Recommendation recognizes the field among a provider's other uses of the category;
b) the non-standard facilities category is the category octet `0x0F` followed by the NS field of [ITU-T V.8] (its length octet $K+L+M+1$, the country code, the provider code's length L, the provider code, then the information: here the identifier and the two octets of the field), carried in extension octets of the form `xx010xxx`, five bits each, the field's bits from its first octet's least significant on, in bits 1 to 3 then 7 and 8 of each extension octet;
c) the DCE shall also set, in the modulation modes category, the bits of its fallback modulations (V.22 and V.21) and, in the protocols category, LAPM.

On the allocation of the bit, a DCE shall set it and shall still send the information field, in the category ITU-T assigns to it.

A.2 The information field

Table A.1 – The information field

Octet	Bits	Meaning
1	8–6	version of this Recommendation: 001
1	5	signal interface at 16000 Hz (1), or 8000 Hz (0)
1	4	facsimile of Annex D supported
1	3	the amplitude loading of clause 10.11 supported
1	2	reserved, 0
1	1	always 0 (see below)
2	8	robust mode (always 1)
2	7	narrowband mode (always 1)
2	6	wideband mode, 200 Hz comb
2	5	wideband mode, 400 Hz comb
2	4	8-DPSK
2	3	22.5 ms symbol (the 20 ms symbol is always supported)
2	2	quick start of clause 10.10
2	1	parametric mode (clause 10.7)

A DCE receiving an information field with a version it does not know shall treat the peer as having version 1 and the capabilities the field's bits it knows say. A DCE shall accept a field of more than two octets and use its first two: a later version may add octets.

Octet 1 bit 1 shall always be 0, in every version. K56flex clients (the PCM protocol before [ITU-T V.90]) take a non-standard block of [ITU-T V.8bis] as K56flex's from its country code (`0xB5`), its provider code's length (2) and two bits of its sixth and seventh octets after the length, without looking at the provider code; in this field's block those octets are the field's two, and octet 2 has the bits that they check. Octet 1 bit 1 clear is what keeps a K56flex client from taking this Recommendation's field, carried in [ITU-T V.8bis] (clause A.3), for a K56flex server's.

NOTE – The field is kept to two octets: CM and JM go through the codec's lowest modes in [ITU-T V.21], and every octet more is a chance more of a repetition lost. What only the link layer needs (its compression) is indicated in INFO0 (Table B.1), which is checked by its FCS and comes again until received.

A.3 V.8bis

A DCE that uses the procedures of [ITU-T V.8bis] shall carry the same information field in the capabilities list of its MS and CL messages, in the non-standard capability of that Recommendation, with the same identifier octet.

Annex B

The probing signals and the INFO messages

(This annex forms an integral part of this Recommendation.)

B.1 L1, L2 and L3

L1 is each comb the transmitting DCE supports for the band of its signal interface (Table 3: N200, N300 and N400 at 8000 Hz; W200, W300 and W400 at 16000 Hz; S200, S300 and S400 at 32000 Hz), in the order of their codes, all tones loaded, DQPSK, with the 20 ms symbol, for 0.6 s each (30 symbols); L2 is the 200 Hz comb with the 22.5 ms symbol for 0.5 s (22 symbols). The phase difference transmitted on tone k in symbol n is given by two bits taken from the scrambler of clause 10 fed with ones, started at the first symbol of L1 and run across the tones from the lowest and across the symbols in order, so that the receiver, knowing the sequence, measures the phase error of every tone symbol. L1's first symbol is the phase reference. The level is −12 dBFS over all tones.

L3 is transmitted by a DCE that indicated the parametric mode (clause 10.7): the P20 preamble and 40 symbols with the 20 ms symbol, then the P10 preamble and 80 symbols with the 10 ms symbol, 1.84 s in all, the P10 part beginning where the P20 part ends (the receiver takes the P10 part's timing from the P20 part's: the P10 preamble's lags are found poorly through AMR 10.2 kbit/s and below); each symbol's index is four bits from the same scrambler, continued from L2, so that the receiver, knowing the lags, counts the symbols in error. Its level is the parametric mode's, −15 dBFS.

B.2 What the receiver measures

For each tone k and each of L1 and L2 the receiver computes the root mean square phase error σ_k of the received phase differences against the transmitted ones, after removing the mean (the clock offset's rotation), the mean received level of the tone, the rms change $\sigma_{A,k}$ of its magnitude from one symbol to the next in dB (for the amplitude loading of clause 10.11; L1's amplitude is constant), and the fraction of tone symbols in error; for each part of L3, the number of lag symbols in error.

B.3 INFO0

Table B.1 – INFO0

Octet	Bits	Meaning
1	8–1	the information field's octet 2 (Annex A): the modes supported
2	8–1	the combs this DCE will probe with, a bit each for codes 0 to 7 of Table 3 (bit 8 for code 0)
3	8	the same for code 8
3	7	quick start offered (the loading of octets 5 onwards is the last call's)
3	6	quick start accepted (in the answering DCE's INFO0)
3	5	the amplitude rings of clause 10.11 allowed (the DCE's `+CVM` permits them)
3	4	the preset dictionary for [ITU-T V.44] of Annex E
3	3	the zstd compression of Annex F
3	2	reserved, 0
3	1	reserved, 0 (for an octet of further capabilities after octet 4)
4	8–1	the DCE's transmit level in the data phases, in 0.5 dB below −9 dBFS
5–		with quick start: the loading of Table B.2

B.4 INFO1 and MP

Table B.2 – INFO1 and MP: the loading asked for

Octet	Bits	Meaning
1	8	MP only: acknowledgement (the peer's MP was received)
1	7–6	symbol period: 00 20 ms, 01 22.5 ms, 10 10 ms (parametric mode only)
1	5–4	code rate: 00 1/2, 01 3/4, 10 7/8
1	3–1	reserved, 0
2	8	MP only: retrain (both DCEs return to phase 2 after the exchange, clause 9.6)
2	7–1	reserved, 0
3	8–1	round-trip delay this DCE measured, in 10 ms, 255 if unknown
4	8–5	the comb: its code of Table 3; 0 with the parametric mode
4	4	the parametric mode (clause 10.7) is asked for, with the symbol period of octet 1: *N* is then 0
4	3	amplitude rings: the second bitmap follows the first
4	2–1	reserved, 0
5	8–1	number of tones *N* in the comb, 8 to 68; 0 with the parametric mode, or when no data mode is asked for (clause 9.4 e))
6–		*N* two-bit fields, the lowest tone first, packed from bit 8: 00 not loaded, 01 DQPSK, 10 8-DPSK, 11 reserved; where octet 4 bit 3 is set, *N* more two-bit fields after them (from the next octet), the tones in the same order: 00 no rings, 01 two rings, 10 four rings (clause 10.11), 11 reserved

Each message is an HDLC frame of [ITU-T V.42] (flags, the FCS of ISO/IEC 13239, no address or control octet) carrying the octets above, transmitted twice in the message mode of clause 9.3; a receiver takes the first copy whose FCS checks. INFO0 is sent 100 ms after the end of the V.8 procedures by the calling DCE; the answering DCE's INFO0 50 ms after the end of the calling DCE's L3 (of L2 where there is no L3), and its INFO1 100 ms after the end of its own L3; the calling DCE's INFO1 50 ms after the end of the answering DCE's. Each DCE sends MP 100 ms after the last INFO1, and after it has received its peer's MP with the acknowledgement bit set, one more MP with its own set.

Annex C

The inner code and its interleaver

(This annex forms an integral part of this Recommendation.)

C.1 The code

The inner code is a Reed-Solomon code over GF(2^8). The field is generated by the primitive polynomial $p(x) = x^8 + x^4 + x^3 + x^2 + 1$, and the code generator polynomial is

$$g(x) = \prod_{i=0}^{2t-1} (x - \alpha^i),$$

with α a root of $p(x)$: the field and the generator of the RS(255,239) code of [ITU-T G.975] (where $t = 8$), with t as Table C.1 gives. Codewords are shortened to $n = 32$ octets: the encoder computes the check octets of the k information octets preceded by $255 - n$ zero octets, which are not transmitted, and the decoder restores them. The first octet of a codeword is the highest order coefficient.

Table C.1 – The code rates

Rate	n	k	t	**erasures corrected** ($2t$)
7/8	32	28	2	4
3/4	32	24	4	8
1/2	32	16	8	16

C.2 The interleaver

Codewords are interleaved in blocks of $D = 4$: the 4 codewords of a block are transmitted octet by octet across the codewords, that is, octet 0 of codeword 0, octet 0 of codeword 1, …, octet 0 of codeword 3, then octet 1 of each, and so on, 128 octets in all. Consecutive octets on the line thus belong to different codewords, and a symbol erasure of up to 4 octets (16 DQPSK tones) costs each codeword one octet. The interleaver's delay is one block: 128 octets, which is 0.64 s at 1600 bit/s and 0.31 s at 3300 bit/s. In the parametric mode $D = 1$: the codewords are sent one after the other (a block of 128 octets would take 5 s at 200 bit/s), and an erased frame costs the codeword it falls in one or two octets, which every code rate corrects. The receiver's de-interleaver is the inverse. The block boundary is marked by the first symbol of phase 4 and by the MP exchange of a renegotiation; the receiver recovers it, if lost, from the syndromes of the codewords (a block boundary error leaves every codeword uncorrectable).

C.3 Erasure marking

The receiver should give the decoder, per octet, an erasure flag set when the tone symbols that carried the octet's bits were marked unreliable (clause 11.4) or lay in a frame detected as erased (clause 11.3). A decoder given e erasures and v errors corrects the codeword when $2v + e \leq 2t$.

Annex D
Facsimile

(This annex forms an integral part of this Recommendation.)

D.1 Principle

A DCE that indicates facsimile in its information field (Annex A) terminates the [ITU-T T.30] session of the facsimile terminal on its side – as a Group 3 facsimile machine does on the PSTN, with the V.21, V.27ter, V.29 and V.17 modulations towards the terminal, or as a T.38 gateway does towards an IP network – and carries the session to its peer as the Internet facsimile protocol (IFP) packets of [ITU-T T.38] in the data link of this Recommendation. The peer DCE reproduces the session towards its own terminal. The procedure is that of a [ITU-T T.38] gateway pair with the data link of this Recommendation in place of the IP network, and of the facsimile relay of [ITU-T V.150.1] in its aim.

D.2 Carriage of the IFP packets

Each IFP packet shall be sent in the information field of one LAPM I frame, preceded by one octet of type (`0x01`: T.38 IFP packet, other values reserved) and two octets of length (most significant first). The UDPTL transport of [ITU-T T.38] and its redundancy are not used: LAPM delivers the packets in order and without loss. The T.38 options in use shall be version 0, no error correction and no fill bit removal, unless both DCEs indicated otherwise in a T.38 options exchange carried the same way with type `0x02`.

D.3 Rates

The slowest image modulation of [ITU-T T.30] is V.27ter at 2400 bit/s. Where the coded rate of the data link is below the modulation rate the terminals negotiate, the DCE shall accept only the error correction mode of [ITU-T T.30] Annex A from its terminal (it shall clear the non-ECM capability in the DIS it relays) and shall pace the sending terminal with the receive-not-ready procedure of that annex, as T.38 gateways do. The T.30 session's timers (T1, T2, T4) are met by the DCE towards its terminal; the DCE may spoof the terminal with preamble flags while the peer's answer is in transit, as [ITU-T T.38] allows.

Annex E

A preset dictionary for V.44

(This annex forms an integral part of this Recommendation.)

E.1 Use

Where both DCEs indicate it in their INFO0 (Table B.1, octet 3 bit 4) and the XID exchange of [ITU-T V.42] agrees on [ITU-T V.44] compression, each DCE shall, before the first data in either direction, prime its V.44 encoder and its V.44 decoder with the text of clause E.2, as follows:

a) with the parameters the XID exchange agreed (N2, N7 and N8 of [ITU-T V.44]), the text's characters are given, in order, to a V.44 encoder in compressed mode, from the start of the text, for as long as
   - the next free codeword, plus twice the number of characters taken but not yet encoded, plus 6, is below N2, and
   - the characters taken are fewer than $\lfloor N8/2 \rfloor - 1$;

   then the encoder's FLUSH ends them. When the parameters let the whole text through, all of it is taken (3505 characters);
b) the dictionary and the history that a V.44 decoder holds once it has decoded that encoder's output are the primed state, for the encoder and the decoder of both directions. A DCE need not run the encoder: any procedure that reaches the same state will do;
c) the data then follows, encoded and decoded from the primed state in compressed mode, the codeword size, the ordinal size, the next free codeword and the history position those of the primed state;
d) a REINIT control code, sent or received, returns the dictionary and the history to the primed state (and the characters the encoder has not encoded yet go into the history after the text), in place of the empty ones of [ITU-T V.44]; so does the decoder's ECM after transparent mode;
e) a new XID exchange that agrees on V.44 primes them anew.

Where either DCE does not indicate it, or the XID exchange agrees on [ITU-T V.42bis] or no compression, [ITU-T V.44] and [ITU-T V.42bis] run as those Recommendations specify.

NOTE – V.44 starts with an empty dictionary and learns from the data. At the rates of this Recommendation's parametric mode and narrowband combs (200 to 3500 bit/s) a call's data is often a few short messages, over before the dictionary has learnt much. Primed, V.44 halves such messages: six typical ones (an HTTP request and response, a mail, JSON, NMEA sentences, a short letter, 1264 octets) come to 958 octets with V.44 alone, 485 primed. The text costs nothing on the line and about 3.5 kB of history at each end.

E.2 The text

The text is the lines below, each followed by CR LF (`0x0D 0x0A`), in this order: 113 lines, 3505 octets, all of 7-bit characters ([ITU-T T.50]), so that the ordinals stay 7 bits. A visible space (␣) is a space (`0x20`); line 27 is empty.

```
GET␣/␣HTTP/1.1
Host:␣www.example.com
User-Agent:␣Mozilla/5.0␣(compatible)
Accept:␣text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language:␣en-US,en;q=0.5
Accept-Encoding:␣gzip,␣deflate
Connection:␣keep-alive
Cache-Control:␣no-cache
Content-Type:␣application/x-www-form-urlencoded
Content-Length:␣0
Authorization:␣Basic
Cookie:␣session=
POST␣/api/v1/␣HTTP/1.1
```

```
Content-Type:␣application/json;␣charset=utf-8
HTTP/1.1␣200␣OK
HTTP/1.1␣301␣Moved␣Permanently
HTTP/1.1␣304␣Not␣Modified
HTTP/1.1␣404␣Not␣Found
HTTP/1.1␣500␣Internal␣Server␣Error
Date:␣Mon,␣01␣Jan␣2024␣00:00:00␣GMT
Server:␣Apache
Last-Modified:␣Tue,␣02␣Feb␣2024␣12:00:00␣GMT
Location:␣https://www.example.com/index.html
Set-Cookie:␣id=;␣Path=/;␣HttpOnly
Transfer-Encoding:␣chunked
Content-Type:␣text/html;␣charset=UTF-8

<!DOCTYPE␣html>
<html␣lang="en">
<head>
<meta␣charset="utf-8">
<meta␣name="viewport"␣content="width=device-width,␣initial-scale=1">
<title></title>
<link␣rel="stylesheet"␣href="style.css">
<script␣type="text/javascript"␣src=""></script>
</head>
<body>
<div␣class=""><span></span></div>
<p><a␣href="https://www.">.com</a></p>
<ul><li></li></ul>
<table><tr><td></td></tr></table>
<img␣src=""␣alt=""><br>
<form␣action=""␣method="post"><input␣type="text"␣name=""␣value="">
</body>
</html>
<?xml␣version="1.0"␣encoding="UTF-8"?>
{"id":␣1,␣"name":␣"",␣"type":␣"",␣"value":␣0,␣"status":␣"ok",
␣"time":␣"2024-01-01T00:00:00Z",␣"data":␣[],␣"error":␣null,
␣"enabled":␣true,␣"count":␣0,␣"message":␣"",␣"result":␣false}
From:␣<user@example.com>
To:␣<user@example.com>
Cc:
Subject:␣Re:
Date:␣Wed,␣03␣Jan␣2024␣10:00:00␣+0000
Message-ID:␣<@mail.example.com>
Reply-To:
MIME-Version:␣1.0
Content-Type:␣text/plain;␣charset="us-ascii"
Content-Type:␣multipart/mixed;␣boundary="
Content-Transfer-Encoding:␣7bit
Content-Disposition:␣attachment;␣filename="
Received:␣from␣mail.example.com␣by␣mx.example.org␣with␣ESMTP␣id
EHLO␣example.com
250-SIZE␣10240000
MAIL␣FROM:<user@example.com>
RCPT␣TO:<user@example.com>
DATA
354␣Start␣mail␣input;␣end␣with␣<CRLF>.<CRLF>
250␣OK
220␣Service␣ready
221␣Bye
QUIT
+OK␣POP3␣server␣ready
USER
PASS
STAT
```

```
LIST
RETR␣1
DELE␣1
-ERR
A001␣LOGIN
A002␣SELECT␣INBOX
A003␣FETCH␣1␣(FLAGS␣BODY[HEADER])
*␣OK␣[CAPABILITY␣IMAP4rev1]␣ready
$GPGGA,000000.00,0000.0000,N,00000.0000,E,1,08,1.0,0.0,M,0.0,M,,*00
$GPRMC,000000.00,A,0000.0000,N,00000.0000,E,0.0,0.0,010124,,,A*00
$GPGSV,3,1,12,
$GPVTG,0.0,T,,M,0.0,N,0.0,K,A*00
login:
Password:
Last␣login:
Welcome␣to
Connection␣closed␣by␣foreign␣host.
ERROR
OK
0123456789␣10␣100␣1000␣2024␣00:00:00
https://www.example.com/index.html?id=&page=1&lang=en
http://www.example.org/␣.net/␣.gov/␣.edu/
The
the␣and␣of␣to␣in␣is␣that␣for␣it␣with␣as␣was␣on␣be␣at␣by␣this␣are␣from
␣have␣or␣not␣but␣which␣you␣an␣they␣his␣her␣were␣had␣has␣their␣been
␣one␣all␣there␣more␣if␣will␣can␣would␣what␣so␣out␣about␣up␣who␣when
␣also␣into␣its␣other␣new␣some␣time␣only␣these␣may␣than␣then␣them␣two
␣first␣any␣my␣like␣over␣such␣our␣well␣even␣most␣made␣after␣should
␣could␣because␣between␣through␣during␣before␣under␣while␣where␣your
␣please␣thank␣you␣regards␣information␣message␣number␣service␣report
␣system␣user␣name␣address␣phone␣email␣account␣date␣error␣status
␣data␣file␣page␣text␣total␣price␣order␣today␣tomorrow␣yesterday
ing␣tion␣ment␣ness␣able␣ally␣ed.␣es,␣ly␣s.
Dear␣Sir␣or␣Madam,
Thank␣you␣for␣your␣message.␣Best␣regards,
-----BEGIN-----
-----END-----
```

Annex F

Zstandard compression

(This annex forms an integral part of this Recommendation.)

F.1 Negotiation

Zstandard (zstd, [IETF RFC 8878]) is negotiated in the XID exchange of [ITU-T V.42], as a private parameter set in the private parameter negotiation group (GI `0xF0`), after the parameters of [ITU-T V.42bis] where those are present, with the parameters of Table F.1, each with its PI and PL as clause 12.2 of [ITU-T V.42] has them, values the most significant octet first.

Table F.1 – zstd's parameter set in XID

PI	PL	Name	Value
`0x60`	3	set identification	the characters `ZST` (`0x5A 0x53 0x54`)
`0x61`	1	P0	the directions, as P0 of [ITU-T V.42bis]: 1 initiator–responder, 2 responder–initiator, 3 both
`0x62`	1	window	the base-2 logarithm of the window, 10 to 23: the largest the DCE decodes (offer), the one used (response)
`0x63`	4	dictionary	the identifier of the dictionary, `0x56434D01` for that of clause F.3; 0 for none

a) A DCE that offers compression and supports this annex shall include the set in its XID command, besides [ITU-T V.42bis]'s parameters and [ITU-T V.44]'s, with a window of 15 (32 kB) and, where it indicated the preset in INFO0 (Table B.1, octet 3 bit 4), or outside this Recommendation's modulation, the dictionary `0x56434D01`. A V.cvm DCE shall offer it where both INFO0s indicate zstd (octet 3 bit 3), and not otherwise.
b) A responder that receives the set and accepts zstd shall answer with the set alone (neither [ITU-T V.42bis]'s parameters nor [ITU-T V.44]'s): P0 the directions it agrees to, the window the smaller of the two, the dictionary the one offered if it offers the same one, else 0. A responder that does not know the set passes over it, as [ITU-T V.42] requires, and answers for [ITU-T V.44] or [ITU-T V.42bis].
c) Where the response carries the set, both DCEs shall use zstd, with that response's directions, window and dictionary, in place of [ITU-T V.44] and [ITU-T V.42bis]; otherwise the compression is the one the response names, as before.

The set may be negotiated by any DCE operating [ITU-T V.42] LAPM, in any modulation; outside this Recommendation no indication precedes it.

NOTE – The set goes in the private parameter group, not in the user data subfield with [ITU-T V.44]'s: a DCE has been seen to drop the call when that subfield held a parameter set it does not know.

F.2 The frames

In each direction that compresses, the encoder's output is a single Zstandard frame of [IETF RFC 8878], from the first data until LAPM is released or a new XID exchange is completed (which starts a new frame in each direction):

a) the frame header: the magic number, a Frame_Header_Descriptor without content size, single segment or checksum, the Window_Descriptor of the window agreed (exponent the window's logarithm less 10, mantissa 0), and the Dictionary_ID, four octets, where a dictionary was agreed;
b) then blocks (raw, RLE or compressed) of no more than $\min(\text{window}, 128\ \text{kB})$ octets of content each, the encoder ending one whenever it has no further characters from the DTE to send, so that the data does not wait; Last_Block is not set;

c) no Content_Checksum (LAPM's FCS covers the data);
d) the decoder shall decode any block that [IETF RFC 8878] allows within the window agreed, and deliver its characters to the DTE once the block has arrived whole. A frame that cannot be decoded is a failure of the link, which shall be released.

With the dictionary, both directions' encoder and decoder start from it (its content as history, its entropy tables and repeat offsets) as [IETF RFC 8878] clause 5 specifies.

NOTE – zstd's blocks give an encoder the choice of its effort; the reference encoder of the DCEs (a hash chain and lazy matching, the entropy tables from each block's statistics) codes bulk text some three times smaller than [ITU-T V.44]. A block costs at least three octets of header, so that characters sent one at a time (typing) cost more than with [ITU-T V.44]; at such rates that is of no consequence. With the dictionary, the six messages of Annex E's note come to 474 octets each on its own (V.44 primed: 485), 414 in a row (449).

F.3 The dictionary

The dictionary `0x56434D01` is a formatted dictionary ([IETF RFC 8878] clause 5) of 3660 octets: the 155 octets below (in hexadecimal, sixteen to a line, the decimal offset of its first at its left), then, as its content, the 3505 octets of the text of clause E.2. They are the magic number `0xEC30A437` and the identifier, least significant octet first, a Huffman table for all 256 octet values (its weights FSE coded), the FSE tables of the offset codes (accuracy 8), the match length codes (9) and the literal length codes (9), and the repeat offsets 1, 4 and 8.

```
000  37 A4 30 EC 01 4D 43 56 43 11 70 16 A3 B3 0D DD
016  3D 03 33 33 33 33 33 33 33 33 C1 46 EB 38 22
032  F5 8F D1 7F F4 90 94 11 17 CE 90 55 9D FA 5A 18
048  11 1E 30 13 0E 98 51 62 87 18 FC 82 D8 75 E1 36
064  70 C4 83 0C 4C D1 50 7B A2 D0 9A B9 A3 60 40 3C
080  3A A3 11 EA FB F0 24 1D 34 6C AC C6 63 A1 34 C9
096  B1 1C 84 18 31 C4 18 43 08 41 84 10 22 22 23 22
112  22 22 22 22 49 92 A4 03 44 A8 A9 07 06 B5 D1 4C
128  18 A5 31 4C 29 C3 98 52 33 23 23 49 92 A4 03 01
144  00 00 00 04 00 00 00 08 00 00 00
```

Appendix I

The speech codecs as a channel: measurements

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

I.1 Method

The measurements were made with the D-Modem project's tools [b-D-Modem], in October 2026, as follows.

a) **Codecs.** AMR: the OpenCORE AMR-NB library (encoder and decoder); AMR-WB: the VisualOn AMR-WB encoder and the OpenCORE AMR-WB decoder. Both are the implementations of the 3GPP reference code that Android carries; the AMR-WB pair was checked against the 3GPP reference code of [3GPP TS 26.173] on a comb signal (clause I.9). EVS: the 3GPP reference code of [3GPP TS 26.442] (fixed point, version 16.4.0), encoder and decoder, through files in the G.192 format. The codecs ran at the modes of Table I.1, without DTX unless stated, one encoding and decoding per direction.

b) **Signal interface.** 16 bit linear PCM at 8000 Hz (AMR, and AMR-WB through a 2:1 resampler with a 3400 Hz low pass, as a narrowband signal reaches a wideband codec), at 16000 Hz (AMR-WB and EVS as they are) or at 32000 Hz (EVS super wideband). The test signals were at −12 dBFS.

c) **Modulations of the V series.** Two D-Modem modems (`half_open8`) called each other through `netsim`, the project's network simulator, with the codec on its line (`NET_CODEC`), at each modulation of Table I.2 with the modems' default settings (V.42 and V.44 where the modulation carries them), and sent 10 s of data each way; a call passed when both modems reached CONNECT and the data arrived intact.

d) **Waveform probes.** Test signals were generated, passed through the codec (`codec_probe`, or the EVS programs) and analysed against the input, aligned by correlation: steady tones (the signal-to-noise ratio of the output against the input, with the best gain), binary FSK at 300 bit/s (bit errors of a non-coherent detector), DQPSK and 8-DPSK on one carrier (symbol errors and the root mean square phase error of the differential detector), and the combs of clause 10 with the phases changing at random every symbol (per tone: symbol errors, the rms phase error, and the rms change of the tone's amplitude from one symbol to the next). Each measurement is over 150 symbols or 800 bits; a tone counts as *good* when its symbol error ratio is below $2 \cdot 10^{-2}$, and the raw rate given is that of the good tones.

e) **Frame erasures.** Frames were erased at random, independently, at the ratio given, and the decoder told so (its concealment ran).

f) **Parametric probes.** A pulse train with a known sequence of periods from a lag alphabet, one period a symbol, as clause 10.7 defines the signal, at −15 dBFS rms, was passed through the codec and the period of each symbol decided by the normalized autocorrelation of that clause over the symbol, the receiver given the symbol timing (its acquisition from the preamble was not measured); 200 symbols a configuration with the 20 ms symbol and 400 with the 10 ms symbol, the first two priming the codec and not counted. The stacked signal of clause I.8 was measured the same way.

The scripts (`measure/probe.py`, `probe3.py`, `probe_pitch.py`, `probe_stack.py`, `matrix.sh`) and the generator of the tables below (`tables.py`) are in the D-Modem repository with this document, so that the measurements can be repeated and extended.

I.2 The modulations of the V series through the codecs

Table I.2 gives, for each modulation and codec, the rate of the CONNECT the modems reached (bold where the data then arrived intact in both directions; with an ^{e} where the data had errors, which means the modulation carried no error correction, or that V.42's detection phase failed and the modems ran without

Table I.1 – The codec modes measured

Codec	Modes (kbit/s)
AMR ([3GPP TS 26.090]), 8000 Hz	4.75, 7.4, 7.95, 10.2, 12.2
AMR-WB ([ITU-T G.722.2]), 16000 Hz	6.6, 12.65, 15.85, 23.85; also behind a 2:1 resampler at 8000 Hz
EVS ([3GPP TS 26.445]), primary modes	NB (8000 Hz): 7.2, 9.6, 13.2, 24.4; WB (16000 Hz): 9.6, 13.2, 16.4, 24.4, 32, 48, 64, 96, 128; SWB (32000 Hz): 13.2, 24.4, 32, 48, 64, 96, 128
EVS AMR-WB IO	23.85 (16000 Hz)

it) and a dash where the modems did not connect. V.21 (and Bell 103) connect through every mode and

Table I.2 – D-Modem's modulations through the codecs: CONNECT rate (bold: data intact; e: data in error; –: no connection)

Codec	V.21	Bell 103	V.23	V.22	Bell 212A	V.22bis	V.32	V.32bis	V.34
none (G.711)	**300**	**300**	**1200**	**1200**	**1200**	**2400**	**9600**	**14400**	**33600**
AMR 12.2	**300**	**300**	1200^{e}	**1200**	1200^{e}	2400^{e}	–	–	–
AMR 7.4	300^{e}	300^{e}	1200^{e}	–	300^{e}	2400^{e}	–	–	–
AMR 4.75	300^{e}	300^{e}	1200^{e}	–	300^{e}	–	–	–	–
AMR-WB 23.85	**300**	300^{e}	1200^{e}	–	300^{e}	2400^{e}	–	–	–
AMR-WB 12.65	300^{e}	300^{e}	1200^{e}	–	300^{e}	–	–	–	–
AMR-WB 6.6	–	300^{e}	–	–	300^{e}	–	–	–	–
AMR 12.2, 1% of the frames erased	300^{e}	300^{e}	1200^{e}	1200^{e}	300^{e}	2400^{e}	–	–	–
AMR 12.2, 3% of the frames erased	300^{e}	300^{e}	1200^{e}	–	300^{e}	2400^{e}	–	–	–
AMR-WB 23.85, 1% of the frames erased	300^{e}	300^{e}	1200^{e}	–	300^{e}	–	–	–	–

carry their 300 bit/s without error through AMR at 12.2 and AMR-WB at 23.85; at the lower modes their bits are in error (Table I.3). V.22 connects through AMR 12.2 and its LAPM comes up, so that its data arrives intact: 1200 bit/s. V.22bis connects through AMR 12.2 and AMR-WB 23.85 (its handshake is V.22's) but its 16-point constellation does not survive, nor does V.42's detection on it. V.32 and above do not complete their training; V.23's 1200 bit/s channel does, its 75 bit/s channel not.

Table I.3 – Bit error ratio of 300 bit/s FSK through the codecs (8000 Hz interface)

Codec	V.21 ch. 1	V.21 ch. 2	Bell 103
none (G.711)	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
AMR 12.2	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
AMR 10.2	0.0139	0.0139	0.0076
AMR 7.95	$< 10^{-4}$	0.0038	0.0038
AMR 7.4	0.0063	0.0038	$< 10^{-4}$
AMR-WB 23.85	$< 10^{-4}$	$< 10^{-4}$	0.0013
AMR-WB 15.85	$< 10^{-4}$	0.0013	0.0063
AMR-WB 12.65	0.0013	0.0025	0.0089
AMR 12.2 + AMR 12.2	0.0013	$< 10^{-4}$	$< 10^{-4}$
AMR-WB 23.85 + AMR 12.2	$< 10^{-4}$	$< 10^{-4}$	0.0013
AMR 12.2, 1% of the frames erased	0.0228	0.0241	0.0304
AMR 12.2, 3% of the frames erased	0.0253	0.0253	0.0342

I.3 Steady tones

Table I.4 gives the signal-to-noise ratio of a steady tone after the codec, which is high (25 to 37 dB through AMR 12.2) at most frequencies and low (10 to 18 dB) at a few: those whose period is a whole number of

samples at 8000 Hz (800, 1000, 1333, 1600, 2000, 2667 Hz) and at 1500 Hz. A steady tone is the easiest signal for the long-term predictor once it has locked, and the hardest for the linear predictor, whose line spectrum it cannot place; the values are no guide to a modulated signal, whose phases the predictor must follow, and the single-carrier and comb measurements below are.

Table I.4 – SNR (dB) of a steady tone at −12 dBFS after the codec, by frequency (Hz)

Codec	300	500	700	900	1100	1300	1500	1700	1900	2100	2300	2500	2700	2900	3100	3300
AMR 12.2	28	33	29	36	33	31	16	29	31	33	32	25	25	31	30	34
AMR 7.4	28	17	26	30	26	27	15	24	30	22	15	16	33	18	18	19
AMR 4.75	20	16	15	28	23	20	14	10	19	23	14	12	11	13	11	9
AMR-WB 23.85	28	24	30	27	31	22	16	26	27	25	26	15	27	33	31	32
AMR-WB 12.65	28	25	30	26	29	18	16	21	20	21	22	15	24	33	28	32
AMR-WB 6.6	27	23	21	23	22	18	19	14	5	4	6	14	6	10	8	21
AMR 12.2, 1% of the frames erased	15	14	15	-1	4	6	13	9	13	8	6	11	11	15	18	16
AMR 12.2, 3% of the frames erased	15	14	15	-1	4	6	13	9	13	8	6	11	11	15	18	16
AMR-WB 23.85, 1% of the frames erased	14	14	12	13	12	14	11	14	13	13	14	11	14	13	13	9

I.4 One carrier

Table I.5 gives the symbol error ratio and the rms phase error of DQPSK on one carrier at 1700 Hz, at 300 and 600 baud (600 baud is V.22's). Through AMR 12.2 and AMR-WB 23.85 the 600 baud signal has a symbol error ratio near 10^{-3} with 14° rms; through AMR 7.4 and below, and AMR-WB 12.65, it does not pass.

Table I.5 – DQPSK on one carrier at 1700 Hz: symbol error ratio / rms phase error

Codec	300 baud (600 bit/s)	600 baud (1200 bit/s)
none (G.711)	$< 10^{-3}$ / 2°	$< 10^{-3}$ / 3°
AMR 12.2	$< 10^{-3}$ / 10°	0.002 / 14°
AMR 10.2	0.010 / 15°	0.036 / 20°
AMR 7.95	0.040 / 21°	0.178 / 34°
AMR 7.4	0.064 / 24°	0.192 / 36°
AMR-WB 23.85	$< 10^{-3}$ / 10°	$< 10^{-3}$ / 12°
AMR-WB 15.85	0.002 / 13°	0.033 / 19°
AMR-WB 12.65	0.007 / 15°	0.047 / 23°
AMR 12.2 + AMR 12.2	0.004 / 13°	0.017 / 20°
AMR-WB 23.85 + AMR 12.2	0.002 / 13°	0.024 / 18°
AMR 12.2, 1% of the frames erased	0.010 / 17°	0.035 / 26°
AMR 12.2, 3% of the frames erased	0.012 / 18°	0.035 / 26°

I.5 The combs

Table I.6 gives, at the 8000 Hz interface, the tone symbol error ratio, the rms phase error and the raw rate of the good tones for the narrowband comb (200 Hz spacing, 400 to 3400 Hz, 16 tones) with the 20 ms and the 22.5 ms symbols, with 8-DPSK, with the shorter 12.5 ms symbol, and for the 400 Hz comb; Table I.7 the same at the 16000 Hz interface for the wideband combs (400 to 6800 Hz). The columns "+10 ms" are the 20 ms symbol with its boundaries 10 ms into the codec frame rather than at it. What the combs show, and clause 10 builds on:

- the 16-tone comb with DQPSK and the 20 ms symbol passes AMR 12.2 at a tone symbol error ratio of 10^{-3} (12° rms) and AMR-WB 23.85 better; it does not pass AMR 7.4 and below, nor AMR-WB 6.6;
- the position of the symbol in the codec frame makes no difference;

Table I.6 – Combs at the 8000 Hz interface: tone SER / rms phase error / raw rate of the good tones (bit/s)

Codec	200 Hz, 4-DPSK, 20 ms	200 Hz, 4-DPSK, 20 ms, +10 ms	200 Hz, 8-DPSK, 20 ms	200 Hz, 4-DPSK, 22.5 ms	200 Hz, 4-DPSK, 12.5 ms	400 Hz, 4-DPSK, 20 ms
none (G.711)	$< 10^{-3}$ / 0° / 1600	$< 10^{-3}$ / 0° / 1600	$< 10^{-3}$ / 0° / 2400	$< 10^{-3}$ / 0° / 1422	$< 10^{-3}$ / 0° / 2560	$< 10^{-3}$ / 0° / 800
AMR 12.2	0.001 / 12° / 1600	$< 10^{-3}$ / 12° / 1600	0.058 / 12° / 0	$< 10^{-3}$ / 13° / 1422	0.054 / 23° / 160	$< 10^{-3}$ / 7° / 800
AMR 10.2	0.027 / 20° / 700	0.023 / 19° / 700	0.249 / 19° / 0	0.033 / 21° / 355	0.163 / 34° / 0	0.020 / 15° / 600
AMR 7.95	0.146 / 32° / 0	0.139 / 32° / 0	0.457 / 33° / 0	0.172 / 35° / 0	0.368 / 54° / 0	0.055 / 23° / 200
AMR 7.4	0.219 / 39° / 0	0.225 / 40° / 0	0.541 / 40° / 0	0.248 / 41° / 0	0.409 / 58° / 0	0.151 / 31° / 0
AMR-WB 23.85	$< 10^{-3}$ / 8° / 1600	$< 10^{-3}$ / 8° / 1600	0.016 / 8° / 1650	$< 10^{-3}$ / 7° / 1422	0.007 / 14° / 2080	$< 10^{-3}$ / 6° / 800
AMR-WB 15.85	0.011 / 14° / 1100	0.008 / 14° / 1400	0.129 / 14° / 750	0.007 / 13° / 1155	0.087 / 25° / 1120	0.001 / 11° / 800
AMR-WB 12.65	0.060 / 21° / 700	0.059 / 21° / 700	0.265 / 21° / 300	0.048 / 20° / 622	0.169 / 35° / 320	0.017 / 17° / 500
AMR 12.2 + AMR 12.2	0.033 / 20° / 500	0.034 / 20° / 400	0.256 / 21° / 0	0.037 / 21° / 266	0.163 / 34° / 0	0.001 / 12° / 800
AMR-WB 23.85 + AMR 12.2	0.013 / 17° / 1200	0.018 / 17° / 900	0.178 / 17° / 0	0.016 / 17° / 888	0.127 / 31° / 0	0.002 / 11° / 800
AMR 12.2, 1% of the frames erased	0.026 / 20° / 300	0.035 / 24° / 0	0.094 / 23° / 0	0.022 / 20° / 444	0.070 / 28° / 0	0.030 / 21° / 0
AMR 12.2, 3% of the frames erased	0.044 / 26° / 0	0.057 / 29° / 0	0.117 / 27° / 0	0.047 / 27° / 0	0.070 / 28° / 0	0.048 / 26° / 0

Table I.7 – Combs at the 16000 Hz interface (AMR-WB): tone SER / rms phase error / raw rate of the good tones (bit/s)

Codec	200 Hz, 4-DPSK, 20 ms	200 Hz, 8-DPSK, 20 ms	400 Hz, 4-DPSK, 20 ms	400 Hz, 8-DPSK, 20 ms	300 Hz, 4-DPSK, 20 ms	100 Hz, 4-DPSK, 20 ms
none (G.711)	$< 10^{-3}$ / 0° / 3400	$< 10^{-3}$ / 0° / 5100	$< 10^{-3}$ / 0° / 1700	$< 10^{-3}$ / 0° / 2550	$< 10^{-3}$ / 0° / 2300	
AMR-WB 23.85	0.092 / 23° / 2800	0.198 / 26° / 900	0.087 / 20° / 1400	0.143 / 20° / 900	0.063 / 16° / 2000	
AMR-WB 15.85	0.163 / 33° / 1000	0.397 / 36° / 300	0.129 / 30° / 600	0.319 / 29° / 300	0.083 / 22° / 1600	
AMR-WB 12.65	0.279 / 46° / 600	0.502 / 47° / 0	0.233 / 40° / 300	0.476 / 41° / 150	0.148 / 31° / 700	
AMR-WB 23.85, 1% of the frames erased	0.111 / 29° / 1000	0.225 / 32° / 0	0.106 / 27° / 300	0.171 / 27° / 0	0.084 / 23° / 600	

- 8-DPSK on the same comb has a symbol error ratio of $6 \cdot 10^{-2}$ through AMR 12.2: too high for the code; on the 400 Hz comb (8 tones) it passes, at 1050 bit/s, less than DQPSK on 16 tones;
- the shorter 12.5 ms symbol raises the error ratio by an order of magnitude: the codec needs the integration time;
- at the 16000 Hz interface through AMR-WB 23.85, 28 of the 34 tones are good and the tones above 6400 Hz are not: AMR-WB codes 6.4 to 7 kHz by bandwidth extension, without the waveform;
- the amplitude of a tone changes from one symbol to the next by 1.5 to 3 dB rms through the ACELP codecs: amplitude carries no bits there.

I.6 Tandem codecs and frame erasures

Table I.6 also has AMR 12.2 followed by AMR 12.2 (a mobile-to-mobile call with a transcoding gateway, or two codecs in a call's path), and AMR-WB 23.85 followed by AMR 12.2 (a wideband terminal calling a narrowband one): the comb's error ratio roughly doubles, and the loading drops a few tones. With 1 and 3 per cent of the frames erased at random the comb's error ratio rises by about the erasure ratio itself (an erased frame is an erased symbol); the code of Annex C is dimensioned for this.

I.7 The parametric mode

Table I.8 gives the lag symbol error ratio of the pulse train of clause 10.7 through each codec, for the two alphabets of Table 4 (P20: 16 lags of 40 to 70 samples with the 20 ms symbol; P10: 16 lags of 20 to 35 with the 10 ms symbol) and for others measured: 8 lags of 40 to 96 in steps of 8, the same with the symbols 10 ms off the codec frame, and 32 lags of 32 to 63 at both symbol periods.

Table I.8 – The pulse train's period through the codecs: lag symbol error ratio by alphabet and symbol period (raw rate in bit/s in the heading); 0 is no symbol in error

Codec	P20: 16 lags 40–70, 20 ms (200)	P10: 16 lags 20–35, 10 ms (400)	8 lags 40–96, 20 ms (150)	8 lags 40–96, 20 ms, +10 ms (150)	32 lags 32–63, 20 ms (250)	32 lags 32–63, 10 ms (500)
none (G.711)	0	0	0	0	0	0
AMR 12.2	0	0	0	0	0	0.005
AMR 10.2	0	0.010	0	0	0	0.023
AMR 7.95	0	0.015	0	0	0	0.143
AMR 7.4	0	0.045	0.020	0.010	0.010	0.191
AMR 4.75	0.106	0.221	0.091	0.146	0.106	0.425
AMR-WB 23.85	0	0	0.025	0.056	0	0
AMR-WB 12.65	0	0	0.020	0.040	0	0.010
AMR-WB 6.6	0	0.038	0.187	0.197	0	0.108
EVS 9.6	0	0	0.010	0.015	0	0.083
EVS 13.2	0	0	0.005	0.010	0	0.008
EVS 24.4	0	0	0.025	0.010	0	0
AMR 12.2 + AMR 12.2	0	0	0	0.005	0	0.090
AMR-WB 23.85 + AMR 12.2	0	0	0.010	0.030	0	0.048
AMR 12.2, 1% of the frames erased	0.020	0.015	0.020	0.030	0.020	0.025
AMR 12.2, 3% of the frames erased	0.051	0.038	0.045	0.051	0.040	0.055

What it shows:

- with the P20 alphabet the period came through without a symbol in error through every codec and mode measured but AMR 4.75 (11 per cent) and the cases with frame erasures: through AMR 7.4, AMR-WB 6.6 and EVS 9.6 among them, where no comb loads a tone (Tables I.6 and II.1);
- the P10 alphabet, at twice the rate, came through without error from AMR 12.2, AMR-WB 12.65 and EVS 9.6 up and through both tandems, and with 1 to 5 per cent of its symbols in error through AMR 10.2 to 7.4 and AMR-WB 6.6: the codec codes a lag every 5 ms subframe, and two subframes a symbol are enough where it codes the lag well;
- the alphabet's range matters more than its size: 32 lags of 32 to 63 with the 20 ms symbol (250 bit/s) pass everything but AMR 7.4 (1 per cent) and 4.75, and with the 10 ms symbol (500 bit/s) pass AMR-WB 23.85 and EVS 24.4 without error but not the AMR modes; 8 lags of 40 to 96 (150 bit/s) are worse than 16 of 40 to 70 through every codec but AMR 12.2 to 7.95. The long lags are the less surely carried, and 40 to 70 samples (114 to 200 Hz, the pitch of a low voice) is the range every codec carried;
- the position of the symbol in the codec frame makes no difference (the column "+10 ms"), as for the comb;
- frame erasures cost about twice their ratio in symbols (2.0 per cent of the symbols at 1 per cent of the frames erased, 5 at 3): the erased frame and usually the one after it;
- a second codec in tandem costs nothing with either alphabet (AMR 12.2 after AMR 12.2, and after AMR-WB 23.85); the 32-lag alphabet with the 10 ms symbol shows the tandem's cost (9 and 5 per cent).

Table I.9 gives, for the P20 alphabet through each codec (300 symbols), the decided lag's normalized autocorrelation c of clause 10.7: its mean $\bar{c}$, its 10th percentile, the spread of $\bar{c}$ over windows of 2 s, and the mean margin over the next lag's c; and, from the same codec, the most tones a narrowband comb (N200, N300 or N400, DQPSK, 20 ms) loads at a tone symbol error ratio below $2 \cdot 10^{-2}$, which decides between the comb and the parametric mode (clause 9.4 e)). It is the measurement behind clause 9.6's use of $\bar{c}$.

Table I.9 – The parametric mode's autocorrelation by codec mode, against the tones a comb loads there

Codec	P20 SER	$\bar{c}$	c, 10th percentile	Spread of $\bar{c}$ over 2 s	Margin	Comb tones
AMR 4.75	0.148	0.617	0.328	0.015	0.377	0
AMR 5.15	0.101	0.652	0.388	0.025	0.415	0
AMR 5.9	0.003	0.843	0.684	0.008	0.667	0
AMR 6.7	0.003	0.851	0.692	0.009	0.664	0
AMR 7.4	0	0.883	0.745	0.008	0.683	0
AMR 7.95	0	0.920	0.833	0.002	0.758	5
AMR 10.2	0	0.875	0.729	0.015	0.689	10
AMR 12.2	0	0.951	0.908	0.004	0.767	16
AMR-WB 6.6	0	0.810	0.641	0.032	0.621	0
AMR-WB 8.85	0	0.915	0.833	0.004	0.703	2
AMR-WB 12.65	0	0.939	0.885	0.002	0.752	7
AMR-WB 15.85	0	0.949	0.897	0.002	0.750	11
AMR-WB 23.85	0	0.954	0.906	0.003	0.757	16
AMR 12.2 + AMR 12.2	0	0.918	0.854	0.012	0.744	11
AMR 7.4, 1% of the frames erased	0.020	0.862	0.679	0.020	0.660	0
AMR 12.2, 3% of the frames erased	0.050	0.915	0.861	0.067	0.728	0
EVS 7.2	0	0.757	0.618	0.075	0.525	3
EVS 8	0	0.788	0.639	0.053	0.573	3
EVS 9.6	0	0.938	0.875	0.009	0.775	2
EVS 13.2	0	0.884	0.813	0.027	0.707	10
EVS 16.4	0	0.961	0.926	0.003	0.810	16
EVS 24.4	0	0.968	0.941	0.004	0.828	16
EVS AMR-WB IO 6.6	0.003	0.890	0.712	0.022	0.683	0
EVS AMR-WB IO 12.65	0	0.954	0.909	0.006	0.776	6

What it shows:

- $\bar{c}$ falls with the codec's mode within a codec family at the low end (AMR 4.75: 0.62; 5.9: 0.84; AMR-WB 6.6: 0.81; EVS 7.2: 0.76) and is 0.94 to 0.97 at the high end;
- it does not order the modes by what a comb gets through them: EVS 9.6 (no comb) gives 0.938 and EVS 13.2 (ten tones) 0.884; AMR 7.95 (five tones) 0.920 and AMR 10.2 (ten) 0.875;
- every mode with a comb of four tones or more gave 0.875 or more, and every mode below 0.85 loads no comb: a threshold of 0.85 separates the modes where a retrain cannot find a comb from the others, and no threshold separates the modes where it would;
- frame erasures lower $\bar{c}$ (AMR 12.2: 0.951, with 3 per cent erased 0.915) and widen its spread over 2 s (0.004 to 0.067).

I.8 The two layers in one waveform

The pulse train's period and the comb's phases were also measured as one signal, to see whether the parametric mode could ride on a comb for nothing: a symbol of one period of L samples repeated over the 20 ms (periodic at L, so that the lag carries four bits), whose spectrum is the harmonics of $8000/L$ Hz between 300 and 3400 Hz, each harmonic's phase carrying two bits as its difference from the harmonic below (so that a symbol needs no reference to the symbol before it and L may change every symbol); a timing offset of a fraction of a sample is a constant in those differences, which the receiver estimates by their fourth power, and the lowest difference is a known reference against the fourfold ambiguity that leaves. Table I.10 gives the tone symbol error ratio of the comb of clause 10 (200 Hz spacing, random phases, differential in time) for comparison, of the harmonic signal at a fixed L (without the parametric layer, so the harmonic comb alone), and of the signal with L chosen per symbol from six lags of 36 to 46 (spacings of 174 to 222 Hz, whose harmonic combs carry most), with its lag symbol error ratio.

Table I.10 – The two layers in one waveform (8000 Hz interface): tone symbol error ratio and raw rate; the comb of clause 10 for comparison

Codec	Comb N200 (clause 10): SER / tones / bit/s	Harmonic $L = 40$ (200 Hz): SER / bit/s	Harmonic $L = 50$ (160 Hz): SER / bit/s	Harmonic $L = 70$ (114 Hz): SER / bit/s	Lag per symbol: tone SER / lag SER / bit/s
none (G.711)	0 / 16/16 / 1600	0 / 1400	0 / 1800	0 / 2500	0 / 0 / 1529
AMR 12.2	0.001 / 16/16 / 1600	0.001 / 1400	0.048 / 1800	0.150 / 2500	0.006 / 0 / 1529
AMR 10.2	0.028 / 8/16 / 800	0.065 / 1400	0.139 / 1800	0.292 / 2500	0.154 / 0 / 1529
AMR 7.4	0.222 / 0/16 / 0	0.363 / 1400	0.536 / 1800	0.606 / 2500	0.461 / 0 / 1529
AMR-WB 23.85	0 / 16/16 / 1600	0 / 1400	0.006 / 1800	0.014 / 2500	0.001 / 0 / 1529
AMR-WB 12.65	0.056 / 7/16 / 700	0.050 / 1400	0.155 / 1800	0.277 / 2500	0.114 / 0 / 1529
AMR-WB 6.6	0.392 / 0/16 / 0	0.484 / 1400	0.580 / 1800	0.633 / 2500	0.591 / 0.010 / 1529
EVS 13.2	0.036 / 10/16 / 1000	0.031 / 1400	0.108 / 1800	0.114 / 2500	0.106 / 0 / 1529
EVS 24.4	0 / 16/16 / 1600	0 / 1400	0.001 / 1800	0.003 / 2500	0 / 0 / 1529

It works – the lag came through without error while the harmonics carried data, through every codec but AMR-WB 6.6 (1 per cent) – and it is not worth having:

- the reference costs two harmonics, which is more than the six lags give back: 1529 bit/s against the comb's 1600 through AMR 12.2, AMR-WB 23.85 and EVS 24.4;
- where a comb is marginal the stacked signal is worse: through AMR 10.2 a tone symbol error ratio of 0.154 against 0.028, through AMR-WB 12.65 0.114 against 0.056, through EVS 13.2 0.106 against 0.036. The codec's phase error is more correlated from one symbol to the next on a tone than from one tone to the next in a symbol, so a difference across frequency loses to a difference in time: [ITU-T V.22]'s choice, and [ITU-T V.34]'s, holds here;
- a harmonic, voiced-looking comb is not carried better than one with random phases: at 200 Hz spacing ($L = 40$) both have a tone symbol error ratio of 10^{-3} through AMR 12.2, both fail below 160 Hz spacing, and through AMR 7.4 the harmonic one is worse (0.363 against 0.222). The limit on a comb's density is the codec's noise against each tone, not whether the signal looks like speech.

The modes of clause 10.1 are therefore separate, and the probing chooses between them: the parametric mode carries 200 bit/s where the comb carries nothing, and costs the comb nothing where it carries data.

I.9 Cross-check of the codec implementations

The AMR-WB encoder and decoder used (VisualOn, OpenCORE) were run on the wideband comb and compared with the 3GPP reference code of [3GPP TS 26.173] on the same input; the result is in the text of Table I.11.

Table I.11 – The AMR-WB implementations against the 3GPP reference code

Comparison	Result
OpenCORE AMR-NB encoder and decoder against the 3GPP TS 26.073 reference encoder and decoder, mode 12.2, 2 s of the N200 comb at 8000 Hz	decoded samples identical: 0 of 15 840 differ
VisualOn AMR-WB encoder and OpenCORE AMR-WB decoder against the 3GPP TS 26.173 reference coder and decoder, mode 23.85, 2 s of the W200 comb at 16000 Hz	decoded samples identical: 0 of 31 680 differ

Appendix II
The EVS codec

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

II.1 The codec

The enhanced voice services codec (EVS, [3GPP TS 26.445]) is the codec of VoLTE where both terminals have it, and the codec of 5G voice (VoNR, [b-GSMA NG.114]). It codes 20 ms frames at 5.9 to 128 kbit/s in four bandwidths (narrowband to 4 kHz, wideband to 8 kHz, super wideband to 16 kHz, full band to 20 kHz), choosing per frame between an ACELP core (speech), a transform core (MDCT: TCX and high-quality modes, for music and mixed content) and, at the low rates, a generic sound coding of the spectral envelope with a coarse residual; above the core's band it codes a bandwidth extension that carries no waveform. It also has an AMR-WB interoperable mode (the AMR-WB codec, bit-exact) and a channel-aware mode at 13.2 kbit/s with partial redundancy. Its signal classification decides which core codes a frame, so what it does to a modem's signal depends on what the signal looks like to it, not on the bit rate alone.

II.2 Measurements

The combs of clause 10 were passed through the 3GPP reference encoder and decoder of [3GPP TS 26.442] (fixed point, version 16.4.0, through G.192 files, DTX off) at the signal interface rates and modes of Table I.1, as Appendix I describes. Tables II.1 to II.3 give, per comb, the tone symbol error ratio, the rms phase error, the raw rate of the good tones, and the rms change of a tone's amplitude from one symbol to the next in dB (the last for the amplitude loading of clause 10.11, measured on its own in Table II.4).

Table II.1 – EVS at the 8000 Hz interface (narrowband): tone SER / rms phase error / raw rate of the good tones (bit/s) / rms amplitude change (dB)

Codec	200 Hz, 4-DPSK, 20 ms	200 Hz, 8-DPSK, 20 ms	300 Hz, 4-DPSK, 20 ms	300 Hz, 8-DPSK, 20 ms	400 Hz, 4-DPSK, 20 ms	400 Hz, 8-DPSK, 20 ms
EVS 7.2	0.457 / 70° / 300 / 8.8	0.664 / 73° / 0 / 9.1	0.504 / 74° / 300 / 8.0		0.418 / 64° / 200 / 8.6	0.622 / 70° / 150 / 8.8
EVS 13.2	0.036 / 18° / 1000 / 2.1	0.219 / 18° / 450 / 2.3	0.030 / 19° / 800 / 2.5	0.159 / 18° / 300 / 2.3	$< 10^{-3}$ / 8° / 800 / 1.4	0.021 / 9° / 750 / 1.4
EVS 24.4	$< 10^{-3}$ / 6° / 1600 / 0.9	0.002 / 6° / 2400 / 1.0	$< 10^{-3}$ / 4° / 1100 / 0.7	$< 10^{-3}$ / 5° / 1650 / 0.7	$< 10^{-3}$ / 5° / 800 / 0.8	$< 10^{-3}$ / 5° / 1200 / 0.9
EVS 9.6	0.259 / 45° / 100 / 5.4	0.527 / 46° / 0 / 5.2	0.178 / 36° / 0 / 4.7	0.448 / 34° / 0 / 4.4	0.138 / 32° / 200 / 4.4	0.339 / 32° / 0 / 4.1

Table II.2 – EVS at the 16000 Hz interface (wideband); EVS AMR-WB IO is the codec's AMR-WB mode

Codec	200 Hz, 4-DPSK, 20 ms	200 Hz, 8-DPSK, 20 ms	300 Hz, 4-DPSK, 20 ms	300 Hz, 8-DPSK, 20 ms	400 Hz, 4-DPSK, 20 ms	400 Hz, 8-DPSK, 20 ms
EVS 24.4	0.020 / 16° / 2200 / 2.9	0.177 / 17° / 600 / 2.7	0.001 / 10° / 2200 / 1.7	0.050 / 10° / 1350 / 1.7	0.001 / 9° / 1700 / 1.4	0.027 / 9° / 1650 / 1.4
EVS 13.2	0.444 / 67° / 500 / 7.0	0.631 / 67° / 300 / 7.0	0.323 / 55° / 600 / 7.2	0.545 / 57° / 150 / 7.4	0.196 / 42° / 300 / 6.2	0.379 / 35° / 300 / 5.5
EVS 48	$< 10^{-3}$ / 6° / 3400 / 0.9	0.003 / 6° / 4950 / 1.0	$< 10^{-3}$ / 5° / 2300 / 0.8	0.001 / 5° / 3450 / 0.7	$< 10^{-3}$ / 3° / 1700 / 0.5	$< 10^{-3}$ / 4° / 2550 / 0.5
EVS 96	$< 10^{-3}$ / 1° / 3400 / 0.3	$< 10^{-3}$ / 1° / 5100 / 0.3	$< 10^{-3}$ / 1° / 2300 / 0.3	$< 10^{-3}$ / 1° / 3450 / 0.3	$< 10^{-3}$ / 0° / 1700 / 0.3	$< 10^{-3}$ / 1° / 2550 / 0.3
EVS AMR-WB IO 23.85	0.118 / 27° / 2500 / 2.8	0.244 / 28° / 600 / 2.7	0.070 / 18° / 2000 / 2.0	0.113 / 17° / 1800 / 2.0	0.095 / 22° / 1400 / 2.2	0.178 / 22° / 600 / 2.4
EVS 24.4, 1% of the frames erased	0.041 / 24° / 400 / 3.4	0.201 / 24° / 0 / 3.3	0.021 / 19° / 900 / 2.9	0.075 / 18° / 450 / 2.4	0.022 / 21° / 700 / 2.0	0.055 / 19° / 150 / 2.1

Table II.3 – EVS at the 32000 Hz interface (super wideband): the S200, S300 and S400 combs to 14 kHz

Codec	200 Hz, 4-DPSK, 20 ms	200 Hz, 8-DPSK, 20 ms	300 Hz, 4-DPSK, 20 ms	300 Hz, 8-DPSK, 20 ms	400 Hz, 4-DPSK, 20 ms	400 Hz, 8-DPSK, 20 ms
EVS 24.4	0.390 / 61° / 1400 / 5.6	0.565 / 60° / 300 / 5.8	0.335 / 54° / 1500 / 5.0	0.473 / 54° / 900 / 5.0	0.348 / 55° / 1300 / 4.5	0.454 / 54° / 900 / 4.2
EVS 48	0.124 / 28° / 3300 / 3.1	0.287 / 27° / 2250 / 3.1	0.087 / 22° / 3000 / 2.4	0.220 / 23° / 1950 / 2.6	0.079 / 18° / 3100 / 1.4	0.133 / 19° / 2250 / 1.4
EVS 96	$< 10^{-3}$ / 7° / 6900 / 1.0	0.008 / 7° / 8400 / 1.1	$< 10^{-3}$ / 6° / 4600 / 0.9	0.003 / 6° / 6600 / 0.9	$< 10^{-3}$ / 3° / 3500 / 0.6	$< 10^{-3}$ / 3° / 5250 / 0.6
EVS 128	$< 10^{-3}$ / 3° / 6900 / 0.6	$< 10^{-3}$ / 4° / 10350 / 0.6	$< 10^{-3}$ / 3° / 4600 / 0.5	$< 10^{-3}$ / 3° / 6900 / 0.5	$< 10^{-3}$ / 2° / 3500 / 0.4	$< 10^{-3}$ / 2° / 5250 / 0.4
EVS 13.2	0.634 / 90° / 400 / 7.3	0.780 / 91° / 0 / 7.5	0.599 / 88° / 200 / 8.1	0.756 / 86° / 150 / 8.2	0.532 / 79° / 300 / 7.1	0.708 / 78° / 150 / 7.2

II.3 What they show

a) **At 7.2 and 9.6 kbit/s** (narrowband) and **at 9.6 and 13.2 kbit/s wideband and super wideband**, the combs do not pass: the tone symbol error ratio is 10^{-1} or worse on every comb, and the amplitude of

Table II.4 – The amplitude rings of clause 10.11: per comb, the raw rate (bit/s) of the best loading of each tone with phase alone / with the rings (DQPSK, 8-DPSK, and 8-DPSK or DQPSK with two rings 6 dB apart or four 3 dB apart; a tone loaded where its phase and ring symbol error ratios are both below $2 \cdot 10^{-2}$), 20 ms symbol

Codec	Interface	200 Hz comb	300 Hz comb	400 Hz comb
AMR 12.2	8000 Hz	N200: 1600 / 1600 (+0 %)	N300: 1650 / 1900 (+15 %)	N400: 1150 / 1250 (+9 %)
AMR 10.2	8000 Hz	N200: 1000 / 1000 (+0 %)	N300: 1200 / 1200 (+0 %)	N400: 600 / 600 (+0 %)
AMR 12.2 + AMR 12.2	8000 Hz	N200: 600 / 600 (+0 %)	N300: 1100 / 1100 (+0 %)	N400: 700 / 700 (+0 %)
EVS 13.2	8000 Hz	N200: 1150 / 1300 (+13 %)	N300: 1000 / 1050 (+5 %)	N400: 1000 / 1150 (+15 %)
EVS 24.4	8000 Hz	N200: 2400 / 2850 (+19 %)	N300: 1650 / 2250 (+36 %)	N400: 1200 / 1550 (+29 %)
AMR-WB 23.85	16000 Hz	W200: 3150 / 3250 (+3 %)	W300: 2800 / 3300 (+18 %)	W400: 1650 / 1700 (+3 %)
AMR-WB 15.85	16000 Hz	W200: 1100 / 1150 (+5 %)	W300: 1700 / 1800 (+6 %)	W400: 600 / 650 (+8 %)
EVS 24.4	16000 Hz	W200: 2350 / 2350 (+0 %)	W300: 2800 / 3050 (+9 %)	W400: 2150 / 2350 (+9 %)
EVS 48	16000 Hz	W200: 5000 / 6000 (+20 %)	W300: 3450 / 4450 (+29 %)	W400: 2550 / 3850 (+51 %)
EVS 48	32000 Hz	S200: 4150 / 4650 (+12 %)	S300: 3550 / 4100 (+15 %)	S400: 3850 / 4550 (+18 %)
EVS 64	32000 Hz	S200: 8950 / 9000 (+1 %)	S300: 6550 / 6600 (+1 %)	S400: 5250 / 6250 (+19 %)
EVS 96	32000 Hz	S200: 9800 / 11900 (+21 %)	S300: 6750 / 8650 (+28 %)	S400: 5250 / 7700 (+47 %)
EVS 128	32000 Hz	S200: 10350 / 15100 (+46 %)	S300: 6900 / 10200 (+48 %)	S400: 5250 / 8750 (+67 %)

a tone changes by 5 to 9 dB rms from one symbol to the next. The codec's low-rate generic coding keeps the spectral envelope and not the phases. The robust mode (300 bit/s FSK) passes: V.21's bit error ratio is below 10^{-2} at every mode but 9.6 wideband. A single carrier at 600 baud has a symbol error ratio of $3 \cdot 10^{-2}$ at 13.2 wideband: V.22 would connect and lose data.

b) **At 13.2 kbit/s narrowband**, the N400 comb passes with a tone symbol error ratio below 10^{-3} (8° rms, 800 bit/s raw), the N200 comb loads 10 of its 16 tones, and 8-DPSK does not hold. This is the mode a VoLTE call falls to in poor coverage; the probing's choice of the sparse comb is what keeps the DCE connected through it.

c) **At 24.4 kbit/s**, the transform core takes over for the comb and the picture changes: narrowband, every comb passes with 8-DPSK (N200: $2 \cdot 10^{-3}$, 6°, 2400 bit/s raw); wideband, W300 passes DQPSK at 10^{-3} (2200 bit/s) and W400 8-DPSK on 11 of 17 tones; super wideband, the band above about 8 kHz is the codec's bandwidth extension and carries nothing, and the probing loads 13 to 15 tones (1300 to 1500 bit/s) – less than the wideband interface gives at the same rate, because the signal's power is spread over tones the codec drops. A DCE at a 32000 Hz interface should therefore offer the wideband combs as well (Table 3: the probing finds the band).

d) **At 48 kbit/s and above**, the codec carries the comb nearly as a waveform: wideband, 8-DPSK on all 33 tones of W200 at $3 \cdot 10^{-3}$ (4950 bit/s raw) and on all tones of W400 without error; super wideband at 48 kbit/s, DQPSK on 31 of the 35 tones of S400 to 12.4 kHz (3100 bit/s); at 96 kbit/s, DQPSK on all 69 tones of S200 (6900 bit/s) and 8-DPSK on 56 of them (8400 bit/s); at 128 kbit/s, 8-DPSK on all 69 tones without error, 3.4° rms: 10 350 bit/s raw.

e) **Amplitude.** Through the ACELP codecs a tone's amplitude changes by 1.5 to 3 dB rms from one symbol to the next, and amplitude carries no bits but on a few tones. Through EVS at 24.4 kbit/s the change is 0.7 to 1.0 dB rms, at 48 it is 0.5 to 1.0 dB, at 96 and 128 it is 0.3 to 0.6 dB. The amplitude rings of clause 10.11 were measured on every comb (Table II.4): per comb, the raw rate of the best loading of each tone with phase alone and with two rings 6 dB apart or four 3 dB apart. Through EVS at 128 kbit/s two rings hold on all 46 tones of S300 and 68 of the 69 of S200 with 8-DPSK, and four rings on all 35 of S400: 15 100 bit/s on S200 against 10 350; at 96 kbit/s the rings add 21 to 47 per cent, at 64 kbit/s only on S400 (19 per cent), at 48 kbit/s 12 to 18 per cent super wideband and 20 to 51 per cent wideband, at 24.4 kbit/s narrowband 19 to 36 per cent. Through AMR 12.2 they add a few tones of N300 and N400 (9 to 15 per cent), through AMR 10.2 and the AMR tandem nothing, through AMR-WB 23.85 the few steadiest tones of W300 (18 per cent).

f) **AMR-WB IO** at 23.85 kbit/s behaves as the AMR-WB codec (which it is): W300 loads 20 of 23 tones (2000 bit/s), W200 25 of 33.

g) **Frame erasures** cost more through EVS than through AMR: with 1 per cent of the frames erased at

24.4 kbit/s wideband, the comb's tone symbol error ratio rises from 10^{-3} to $2 \cdot 10^{-2}$ (Table II.2), about twice the erasure ratio, because the transform core's overlap carries each erasure into the next frame as well: an erased frame costs two symbols. The receiver's erasure marking (clause 11.3) should mark both, and the probing should leave erased symbols out of its statistics, which are over a few tens of symbols.

h) **Mode changes.** A VoLTE call switches between 13.2 and 24.4 kbit/s with the radio conditions, and a VoNR call may hold higher modes. Between 13.2 and 24.4 the usable loading changes from one comb (N400, 800 bit/s) to another with 8-DPSK (N200, 2400 bit/s): the renegotiation of clause 9.6 is what follows the codec, and the DCE should expect it every few seconds in a moving terminal.

i) **The parametric mode** passes where the combs do not: through EVS at 9.6 kbit/s narrowband the pulse train's period came through without a symbol in error at both symbol periods (Table I.8): 400 bit/s where the combs carry nothing and V.21's bit error ratio reaches $3 \cdot 10^{-2}$ (item a)). At 13.2 and 24.4 kbit/s it passes as well, and the combs carry more; at 24.4 even the 32-lag alphabet with the 10 ms symbol passed without error (500 bit/s), which the signal set does not use.

II.4 Consequences for the signal set

The signal set of clause 10 needs no change for EVS: the combs, the two modulation orders and the probing's loading cover it from 13.2 kbit/s narrowband (N400, DQPSK) to 128 kbit/s super wideband (S200, 8-DPSK), a range of 800 to 10 350 bit/s raw, and the parametric mode covers 9.6 kbit/s, where no comb passes, at 400 bit/s. What EVS adds is the case for the super wideband combs (S200 to S400) and for the amplitude loading of clause 10.11, which this Recommendation defines from the measured amplitude stability: up to half as much again as phase alone at the codec's highest rates. The terminal's part is in Appendix III: a terminal that can hold the EVS mode at 48 kbit/s or more for a DCE's call multiplies its throughput.

Appendix III
Terminals and networks

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

This appendix says what a terminal and a network should do so that a DCE of this Recommendation gets the most of the speech path; none of it is a requirement of this Recommendation on them.

III.1 The terminal

a) **Signal interface.** The DCE needs the codec's PCM interface: in a UE, the point where the acoustic front end's output enters the encoder and the decoder's output leaves for the acoustic back end. The acoustic processing of [3GPP TS 26.131] (echo cancellation, noise suppression, automatic gain control, equalization) is for the acoustic path and should be bypassed for the DCE's samples; noise suppression in particular treats a comb as noise.
b) **Codec selection.** Where the terminal negotiates the codec (an IMS client, through SDP), it should offer, for a call the DCE will use, the highest modes available and the wideband codecs first: EVS at 24.4 kbit/s or more, AMR-WB at 23.85, AMR at 12.2; it should restrict the mode set (the `mode-set` parameter of [b-IETF RFC 4867]) to the highest modes, and set the EVS parameters `br` and `bw` of [3GPP TS 26.445] Annex A accordingly.
c) **DTX and VAD.** The terminal should turn discontinuous transmission off for the call (the `dtx` parameter of EVS; for AMR and AMR-WB it is the encoder's setting). A comb is not speech, and a voice activity detector may classify it as noise and send comfort noise in its place. The DCE's signals are continuous while a call is up, so that DTX, where it cannot be turned off, has nothing to detect as silence.
d) **Jitter buffer.** The jitter buffer management of [3GPP TS 26.114] adapts its delay by dropping or repeating frames, or by time-scaling the decoded signal ([3GPP TS 26.448] for EVS). Each adaptation is a step in the DCE's symbol timing of up to one frame, which the receiver re-acquires (clause 10.3); a terminal should prefer, for a DCE's call, a fixed delay large enough for the path's jitter.
e) **Level.** The terminal should pass the DCE's samples to the encoder at their level, without gain, and give the DCE the decoder's output at its level.

III.2 The network

a) **Transcoding.** Each codec in the path costs throughput (Appendix I); transcoder-free operation between terminals with the same codec is best. A gateway to the PSTN transcodes to [ITU-T G.711], which costs nothing further.
b) **Echo cancellers.** Network echo cancellers of [b-ITU-T G.168] disable on the 2100 Hz tone with phase reversals that [ITU-T V.8]'s ANSam is; a DCE sends it as a matter of course. Mobile-to-mobile IMS calls have no network echo canceller; the UE's acoustic echo canceller is bypassed with the rest of the acoustic processing.
c) **Codec mode control.** The network changes the codec mode with the radio conditions; the DCE follows with renegotiations (clause 9.6). A network that holds a call at the highest mode while a DCE is in it gives it the highest throughput; the DCE cannot ask for that, and this Recommendation does not define a way.
d) **In-band signalling.** DTMF and other tones sent out of band ([b-IETF RFC 4733]) do not reach the speech path; the DCE's signals are not DTMF and are unaffected. A network that detects the V.8 tones and switches a PSTN gateway to a voice-band data mode ([b-ITU-T V.152]) or to modem relay ([b-ITU-T V.150.1]) does so on its own leg; the mobile leg's codec remains, and remains the limit.

Appendix IV
Implementation notes

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

A DCE of this Recommendation can be built from the parts of a V series software modem, as the D-Modem project [b-D-Modem] has them:

- the V.8 procedures, with the information field of Annex A added to the CM and JM encoders and decoders;
- the V.21 modulator and demodulator for the robust mode and the INFO and MP messages, and the HDLC framing of V.42 for their frames;
- for the parametric mode, a pulse train generator (one 33-tap pulse and a one-pole filter) and the normalized autocorrelation at 16 lags over each symbol, with the timing search over the preamble: a few dozen lines (`measure/probe_pitch.py` has the generator and the detector);
- a multitone modulator (an inverse discrete Fourier transform of the tone phases, the cyclic prefix prepended) and demodulator (the correlation of each tone over the useful part, as the probe scripts of Appendix I do it), with the differential detection, the Gray mapping of Table 6 and the scrambler of V.34;
- a Reed-Solomon encoder and decoder over GF(2^8) for the code of Annex C, with erasure decoding; a block interleaver;
- the LAPM, V.42bis and V.44 of the V.42 implementation, unchanged, with the detection phase disabled;
- the V.250 command interpreter with the `+MS` value and the `+CVM` command of clause 12;
- for Annex D, the T.30 session machine and the T.38 packetizer of a T.38 gateway or an internet-aware fax implementation.

The computational load is small: the comb has at most 68 tones at 50 symbols/s, and the Reed-Solomon code at most 3300 bit/s. The probing's phase error estimates and the loading are the only parts not found in a V series modem, and they are a few hundred lines.

In D-Modem the DCE is a data pump (`dsp/dp_vcvm.cpp`, on `dsp/vcvm*.cpp` and `dsp/rs.cpp`) behind the existing pump interface, which the modem core (`core/`) drives through V.8 as it does for V.34 and V.90, with the link layer (`link/`) above it, and `netsim`'s `NET_CODEC` as the test line; the interoperability test harness runs the DCE against itself through G.711, AMR at 12.2 and 7.4 kbit/s and AMR-WB at 23.85 kbit/s. Its message mode's V.21 receiver is the detector of this appendix's measurements (two filters matched over the bit) with its own bit clock, quick to acquire over the first dozen transitions and then slow; its P20 messages whose FCS fails are retried with the second most likely lag on the five least reliable symbols, singly and in pairs; its Reed-Solomon decoder decodes the errors alone first, then with the 2, 4, … least reliable octets erased. `docs/dsp/vcvm.md` describes it.

Bibliography

[b-ITU-T G.168] Recommendation ITU-T G.168 (2015), *Digital network echo cancellers*.

[b-ITU-T G.992.1] Recommendation ITU-T G.992.1 (1999), *Asymmetric digital subscriber line (ADSL) transceivers*.

[b-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*.

[b-ITU-T V.152] Recommendation ITU-T V.152 (2010), *Procedures for supporting voice-band data over IP networks*.

[b-ITU-T V.22bis] Recommendation ITU-T V.22bis (1988), *2400 bits per second duplex modem using the frequency division technique standardized for use on the general switched telephone network and on point-to-point 2-wire leased telephone-type circuits*.

[b-ITU-T V.32bis] Recommendation ITU-T V.32bis (1991), *A duplex modem operating at data signalling rates of up to 14 400 bit/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits*.

[b-3GPP TS 26.071] 3GPP TS 26.071, *Mandatory speech CODEC speech processing functions; AMR speech CODEC; General description*.

[b-3GPP TS 26.093] 3GPP TS 26.093, *Mandatory speech codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Source controlled rate operation*.

[b-3GPP TS 26.441] 3GPP TS 26.441, *Codec for Enhanced Voice Services (EVS); General overview*.

[b-3GPP TS 26.447] 3GPP TS 26.447, *Codec for Enhanced Voice Services (EVS); Error concealment of lost packets*.

[b-3GPP TS 26.448] 3GPP TS 26.448, *Codec for Enhanced Voice Services (EVS); Jitter buffer management*.

[b-3GPP TS 26.131] 3GPP TS 26.131, *Terminal acoustic characteristics for telephony; Requirements*.

[b-IETF RFC 4867] IETF RFC 4867 (2007), *RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs*.

[b-IETF RFC 4733] IETF RFC 4733 (2006), *RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals*.

[b-GSMA IR.92] GSMA PRD IR.92, *IMS Profile for Voice and SMS*.

[b-GSMA NG.114] GSMA PRD NG.114, *IMS Profile for Voice, Video and Messaging over 5GS*.

[b-D-Modem] D-Modem, a software modem (Neo-Desktop), `https://github.com/Neo-Desktop/d-modem-re`; the modems, the codec simulation (`dsp/test/codec_sim.cpp`, `netsim`) and the measurement scripts of Appendix I are in it.