INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Appendix II

Rec. G.728

(11/95)

GENERAL ASPECTS OF DIGITAL TRANSMISSION SYSTEMS

SPEECH PERFORMANCE

Appendix II to

ITU-T Recommendation G.728

(Previously "CCITT Recommendation")

FOREWORD

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

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

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993).

Appendix II to ITU-T Recommendation G.728 was prepared by ITU-T Study Group 15 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 13th of November 1995.

NOTE

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

CONTENTS

Appendix II to Recommendation G.728

SPEECH PERFORMANCE

(Geneva, 1995)

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

II.1 Introduction

The purpose of this appendix is to give a broad outline of the speech performance of the 16 kbit/s LD-CELP algorithm when interacting with other parts of the network. Some general guidance is also offered on voice-like and non-voice signals.

In this appendix, the use of 16 kbit/s LD-CELP is meant to refer to the algorithm specified in the main body of Recommendation G.728; 32 kbit/s ADPCM is meant to refer to the 32 kbit/s algorithm specified in the main body of Recommendation G.726; and 64 kbit/s PCM is meant to refer to the algorithm specified in the main body of Recommendation G.711.

II.2 Speech performance

II.2.1 Single encoding

Under error-free transmission conditions the perceived quality of a 16 kbit/s LD-CELP codec is lower than that of a 64 kbit/s PCM codec, but equivalent to that of a codec conforming to 32 kbit/s ADPCM.

The performance of the 16 kbit/s LD-CELP codec has been found to be substantially unaffected by noise (Gaussian distribution) with Bit Error Ratio (BER) up to 10^{-3}, and has the equivalent performance to that of a codec conforming to 32 kbit/s ADPCM at a BER of 10^{-2}.

The planning value assigned to 16 kbit/s LD-CELP can be found in Recommendation G.113.

II.2.2 Speech performance when interconnected with coding systems on an analog basis

II.2.2.1 Multi-tandeming of 16 kbit/s LD-CELP codec

When a 16 kbit/s LD-CELP codec is tandemed with multiple speech coding devices, its performance appears to be equivalent to that of 32 kbit/s ADPCM for up to three devices in tandem. Precise rules for tandeming 16 kbit/s LD-CELP codecs are found in Recommendation G.113.

II.2.2.2 Performance with 32 kbit/s ADPCM

The speech performance when 16 kbit/s LD-CELP is connected in tandem with encoding(s) conforming to 32 kbit/s ADPCM is equivalent in the following two configurations:

G.728 + G.726 + G.728,
G.726 + G.728 + G.726,

where a "+" in the above expression indicates an interconnection.

II.2.3 Speech performance of the 16 kbit/s LD-CELP codec for synchronous tandeming

Subjective experiments have shown that the speech performance of synchronous tandeming (that is, interconnection of two or more LD-CELP codecs via 64 kbit/s PCM interfaces) is equivalent to the speech performance for asynchronous tandeming (see II.2.2.1). Therefore, LD-CELP does not have a synchronous tandeming property equivalent to that of 32 kbit/s ADPCM (see Recommendation G.726).

II.2.4 Performance with encoding other than 16 kbit/s LD-CELP and 32 kbit/s ADPCM

In general, interconnection of 16 kbit/s LD-CELP with up to two other devices appears to be equivalent with interconnection of 32 kbit/s ADPCM with the same devices. However, at the time of generating this appendix, only limited data were available on this subject. Consequently, great care must be exercised when interconnection is made to codecs with encoding different from that of 16 kbit/s LD-CELP and 32 kbit/s ADPCM.

II.3 Performance with non-speech signals

It should be noted that the 16 kbit/s LD-CELP codec is an adaptive system which has been optimized for speech. Great care must be taken when making measurements with non-speech signals because the normal assumptions of time invariance and linearity should not be made (e.g. when conducting network maintenance tests using test tones).

II.3.1 Performance with information tones

Experiments have shown that network originated information tones, conforming, to Recommendation Q.35, are easily recognizable when passed through the 16 kbit/s LD-CELP codec (one encoding).

II.3.2 Performance with music

Experiments have shown that no annoying effects were found over a wide variety of music (see also II.4).

II.3.3 Performance with Dual-Tone Multi-Frequency (DTMF) signalling

In general, the performance of the 16 kbit/s LD-CELP codec in single encoding has been found to be equivalent to that of 32 kbit/s ADPCM and 64 kbit/s PCM.

II.3.4 Performance with Signalling System 5 interregister signalling

In general, the performance of the 16 kbit/s LD-CELP codec in single encoding has been found to be equivalent to that of 32 kbit/s ADPCM and 64 kbit/s PCM.

II.3.5 Performance with voiceband data

In general, the performance of the 16 kbit/s LD-CELP codec (even for a single encoding) has been found to be significantly inferior to that of 32 kbit/s ADPCM and 64 kbit/s PCM. However, it should be noted that 16 kbit/s LD-CELP is able to accommodate most (but not all) voiceband data modems operating at 2400 bit/s or less, under realistic channel conditions, provided its perceptual weighting filter and postfilter are both disabled.

II.4 Artificial voice signals

In the course of testing of the 16 kbit/s LD-CELP codec, it was discovered that a certain type of input material can cause encoder and decoder stages to diverge, resulting in a highly distorted output signal. This behaviour was first observed for certain music signals and for artificial speech (see Recommendation P.50). It was subsequently observed for sustained vowel sounds for certain speakers. It has not been observed during actual fluent speech.

In order for this behaviour to happen, the encoder and decoder states must be slightly different at the time the input material begins. This difference could be caused by slight differences in operation, as will be normal with an encoder and a decoder from different manufacturers. Even in encoders and decoders with exactly identical operation, such differences could arise as the result of channel errors or due to different reset instants in the encoder and decoder. It has been observed that the amount of distortion on the output signal increases with reduced precision in the encoder or decoder implementations.

A second requirement for the divergence to happen is that the input signal possesses certain characteristics. First, the signal should contain a number of sharp spectral peaks. (The lowest number of peaks which have been observed to cause this problem is 14.) Second, the frequencies of these peaks should remain relatively stationary over several hundred milliseconds to allow the deviation to build up over time.

During the period of the divergence, the output signal of the decoder does not become unstable in the sense of becoming unbounded. It remains a bounded signal but its spectral content is significantly changed from that of the original signal. When the input signal changes, the encoder and decoder will again converge and give high quality output.

Bibliography

Speech Communication, Vol. 21, No. 2, June 1993 (Special Issue on CCITT Standard on 16 kbit/s Speech Coding).