

3GPP2 C.S0012-0

Version 1.0



**3RD GENERATION
PARTNERSHIP
PROJECT 2
"3GPP2"**

Minimum Performance Standard for Speech S01

COPYRIGHT

3GPP2 and its Organizational Partners claim copyright in this document and individual Organizational Partners may copyright and issue documents or standards publications in individual Organizational Partner's name based on this document. Requests for reproduction of this document should be directed to the 3GPP2 Secretariat at shoyler@tia.eia.org. Requests to reproduce individual Organizational Partner's documents should be directed to that Organizational Partner. See www.3gpp2.org for more information.

CONTENTS

1		
2	1 INTRODUCTION.....	1-1
3	1.1 Scope.....	1-1
4	1.2 Definitions.....	1-2
5	1.3 Test Model for the Speech Codec.....	1-4
6	2 CODEC MINIMUM STANDARDS.....	2-1
7	2.1 Strict Objective Performance Requirements.....	2-1
8	2.1.1 Strict Encoder/Decoder Objective Test Definition.....	2-3
9	2.1.2 Method of Measurement.....	2-4
10	2.1.3 Minimum Performance Requirements.....	2-7
11	2.2 Combined Objective/Subjective Performance Requirements.....	2-7
12	2.2.1 Objective Performance Experiment I.....	2-10
13	2.2.1.1 Definition.....	2-10
14	2.2.1.2 Method of Measurement.....	2-10
15	2.2.1.3 Minimum Performance Requirements for Experiment I.....	2-10
16	2.2.1.3.1 Minimum Performance Requirements for Experiment I Level -9 dBm0.....	2-10
17	2.2.1.3.2 Minimum Performance Requirements for Experiment I Level -19 dBm0.....	2-10
18	2.2.1.3.3 Minimum Performance Requirements for Experiment I Level -29 dBm0.....	2-11
19	2.2.2 Objective Performance Experiment II.....	2-11
20	2.2.2.1 Definition.....	2-11
21	2.2.2.2 Method of Measurement.....	2-11
22	2.2.2.3 Minimum Performance Requirements for Experiment II.....	2-11
23	2.2.3 Objective Performance Experiment III.....	2-12
24	2.2.3.1 Definition.....	2-12
25	2.2.3.2 Method of Measurement.....	2-12
26	2.2.3.3 Minimum Performance Requirements.....	2-13
27	2.2.3.3.1 Minimum Performance Requirements for Level -9 dBm0.....	2-13
28	2.2.3.3.2 Minimum Performance Requirements for Level -19 dBm0.....	2-14
29	2.2.3.3.3 Minimum Performance Requirements for Level -29 dBm0.....	2-14
30	2.2.4 Speech Codec Subjective Performance.....	2-15
31	2.2.4.1 Definition.....	2-15
32	2.2.4.2 Method of Measurement.....	2-15

1	2.2.4.2.1 Test Conditions	2-15
2	CONTENTS (Cont.)	
3	2.2.4.2.1.1 Subjective Experiment I	2-15
4	2.2.4.2.1.2 Subjective Experiment II	2-16
5	2.2.4.2.1.3 Subjective Experiment III	2-17
6	2.2.4.2.2 Source Speech Material	2-17
7	2.2.4.2.3 Source Speech Material for Experiment I	2-17
8	2.2.4.2.4 Source Speech Material for Experiment II	2-18
9	2.2.4.2.5 Source Speech Material for Experiment III	2-18
10	2.2.4.2.6 Processing of Speech Material	2-19
11	2.2.4.2.7 Randomization	2-19
12	2.2.4.2.8 Presentation	2-20
13	2.2.4.2.9 Listeners	2-20
14	2.2.4.2.10 Listening Test Procedure	2-21
15	2.2.4.2.11 Analysis of Results	2-22
16	2.2.4.3 Minimum Standard	2-25
17	2.2.4.4 Expected Results for Reference Conditions	2-27
18	3 CODEC STANDARD TEST CONDITIONS	3-1
19	3.1 Standard Equipment	3-1
20	3.1.1 Basic Equipment	3-1
21	3.1.2 Subjective Test Equipment	3-2
22	3.1.2.1 Audio Path	3-3
23	3.1.2.2 Calibration	3-3
24	3.2 Standard Environmental Test Conditions	3-3
25	3.3 Standard Software Test Tools	3-3
26	3.3.1 Analysis of Subjective Performance Test Results	3-3
27	3.3.1.1 Analysis of Variance	3-3
28	3.3.1.2 Newman-Keuls Comparison of Means	3-6
29	3.3.2 Segmental Signal-to-Noise Ratio Program - snr_c	3-7
30	3.3.2.1 Algorithm	3-7
31	3.3.2.2 Usage	3-8
32	3.3.3 Cumulative Distribution Function - cdf_c	3-9

1	3.3.3.1 Algorithm.....	3-9
2	3.3.3.2 Usage	3-9
3	CONTENTS (Cont.)	
4	3.3.4 Randomization Routine for Service Option 1 Validation Test - random.c	3-9
5	3.3.4.1 Algorithm.....	3-10
6	3.3.4.2 Usage	3-10
7	3.3.5 Rate Determination Comparison Routine - rate_chk.c.....	3-10
8	3.3.6 Software Tools for Generating Speech Files - stat.c and scale.c	3-11
9	3.4 Master Codec - qcelp_a.c	3-11
10	3.4.1 Master Codec Program Files	3-11
11	3.4.2 Compiling the Master Codec Simulation	3-12
12	3.4.3 Running the Master Codec Simulation.....	3-12
13	3.4.4 File Formats	3-14
14	3.4.5 Verifying Proper Operation of the Master Codec.....	3-14
15	3.5 Standard Source Speech Files	3-16
16	3.5.1 Processing of Source Material.....	3-16
17	3.5.1.1 Experiment I - Level Conditions	3-16
18	3.5.1.2 Experiment II - Channel Conditions.....	3-16
19	3.5.1.3 Experiment III - Background Noise Conditions	3-16
20	3.5.2 Data Format of the Source Material.....	3-16
21	3.5.3 Calibration File	3-17
22	3.6 Contents of Software Distribution.....	3-17

1
2
3
4
5
6
7
8
9
10
11
12

FIGURES

Figure 1.3-1. Test Model.....2-4

Figure 2.1-1. Flow Chart for Strict Objective Performance Test.....2-2

Figure 2.2-1. Objective/Subjective Performance Test Flow.....2-9

Figure 2.2.4.2.10-1. Instructions for Subjects2-21

Figure 2.2.4.3-1. Flow Chart of Codec Testing Process.....2-26

Figure 2.2.4.4-1. MOS versus MNRU Q.....2-28

Figure 3.1.1-1. Basic Test Equipment3-1

Figure 3.1.2-1. Subjective Testing Equipment Configuration.....3-2

Figure 3.6-1. Directory Structure of Companion Software3-19

TABLES

1		
2	Table 2.1.2-1. Computing the Segmental Signal-to-Noise Ratio.....	2-6
3	Table 2.2.3.2-1. Joint Rate Histogram Example	2-13
4	Table 2.2.3.3.1-1. Joint Rate Histogram Requirements for -9 dBm0.....	2-14
5	Table 2.2.3.3.2-1. Joint Rate Histogram Requirements for -19 dBm0.....	2-14
6	Table 2.2.3.3.3-1. Joint Rate Histogram Requirements for -29 dBm0.....	2-15
7	Table 2.2.4.2.11-1. Means for Codecs x Levels (Encoder/Decoder).....	2-23
8	Table 2.2.4.2.11-2. Comparisons of Interest Among Codecs x Levels Means	2-23
9	Table 2.2.4.2.11-3. Critical Values for Codecs x Levels Means	2-24
10	Table 2.2.4.2.11-4. Analysis of Variance Table	2-25
11	Table 3.3.1.1-1. Analysis of Variance	3-5
12	Table 3.3.2.2-1. Output Text from <code>snr . c</code>	3-9
13	Table 3.4.4-1. Packet File Structure From Master Codec.....	3-14

14

REFERENCES

The following standards contain provisions, which, through reference in this text, constitute provisions of this Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. ANSI and TIA maintain registers of currently valid national standards published by them.

1. TIA/EIA/IS-96-C, *Speech Service Option Standard for Wideband Spread Spectrum Digital Cellular System*.
2. TIA/EIA/IS-98, *Recommended Minimum Performance Standards for Dual-Mode Wideband Spread Spectrum Mobile Stations*.
3. TIA/EIA/IS-97, *Recommended Minimum Performance Standards for Base Stations Supporting Dual-Mode Wideband Spread Spectrum Mobile Stations*.
4. TIA/EIA/IS-95, *Mobile Station — Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System*.
5. ANSI S1.4-1983, *Sound Level Meters, Specification for*
6. ANSI S1.4a-1985, *Sound Level Meters, Specifications for (Supplement to ANSI S4.1-1983)*
7. ITU-T, The International Telecommunication Union, Telecommunication Standardization Sector, *Blue Book*, Vol. III, Telephone Transmission Quality, IXth Plenary Assembly, Melbourne, 14-25 November, 1988, Recommendation G.711, *Pulse code modulation (PCM) of voice frequencies*.
8. ITU-T, The International Telecommunication Union, Telecommunication Standardization Sector, *Blue Book*, Vol. V, Telephone Transmission Quality, IXth Plenary Assembly, Melbourne, 14-25 November, 1988, Recommendation P.48, *Specification for an intermediate reference system*.
9. IEEE 269-1992, *Standard Method for Measuring Transmission Performance of Telephone Sets*.
10. ANSI/IEEE STD 661-1979, *IEEE Standard Method for Determining Objective Loudness Ratings of Telephone Connections*.
11. Gene V. Glass and Kenneth D. Hopkins, *Statistical Methods in Education and Psychology*, 2nd Edition, Allyn and Bacon 1984.
12. SAS[®] *User's Guide*, Version 5 Edition, published by SAS Institute, Inc., Cary, N.C, 1985.
13. Secretariat, Computer and Business Equipment Manufacturers Association, *American National Standard for Information Systems - Programming Language - C*, ANSI X3.159-1989, American National Standards Institute, 1989.
14. ITU-T, The International Telecommunication Union, Telecommunication Standardization Sector, *Blue Book*, Vol. III, Telephone Transmission Quality, IXth Plenary Assembly, Melbourne, 14-25 November, 1988, Recommendation G.165, *Echo Cancellers*.
15. ITU-T, The International Telecommunication Union, Telecommunication Standardization Sector, Recommendation P.50 (03/93), *Artificial voices*.

- 1 16. TIA/EIA-579, *Acoustic-to-Digital and Digital-to-Acoustic Transmission Requirements for ISDN*
2 *Terminals (ANSI/TIA/EIA-597-91).*

3 **REFERENCES (Cont.)**

- 4 17. J. Statist. Comput. Simul., Vol. 30, pp 1-15, Gordon and Breach Pub. 1988, *Computation of the*
5 *Distribution of the Maximum Studentized Range Statistic with Application to Multiple*
6 *Significance Testing of Simple Effects*, M. D. Copenhaver and B. Holland
- 7 18. ITU-T, The International Telecommunication Union, Telecommunication Standardization Sector,
8 Recommendation P.56 (02/93), *Objective Measurement of Active Speech Level.*
- 9 19. ITU-T, The International Telecommunication Union, Telecommunication Standardization Sector,
10 Revised Recommendation P.800 (08/96), *Methods for Subjective Determination of Transmission*
11 *Quality.*
- 12 20. ITU-T, The International Telecommunication Union, Telecommunication Standardization, Revised
13 Recommendation P.810 (02/96), *Modulated Noise Reference Unit (MNRU).*
- 14 21. ITU-T, The International Telecommunication Union, Telecommunication Standardization, Revised
15 Recommendation P.830 (02/96), *Subjective Performance Assessment of Telephone-band and Wideband*
16 *Digital Codecs.*
- 17 22. ITU-T, The International Telecommunication Union, Telecommunication Standardization,
18 Recommendation G.191 (11/96), *Software Tool Library 96.*
- 19 23. Bruning & Kintz, *Computational Handbook of Statistics*, 3rd Edition, Scott, Foresman & Co., 1987.
- 20 24. IEEE 297-1969, *IEEE Recommend Practice for Speech Quality Measuremet*, IEEE Transactions on
21 Audio and ElectroAcoustics, Vol. AU-17, No. 3, September 1969.

22
23 **TRADEMARKS**

24 Microsoft[®] is a registered trademark of Microsoft Corporation.

25 SAS[®] is a registered trademark of SAS Institute, Inc.

26 UNIX[™] is a registered trademark in the United States and other countries, licensed exclusively through X/Open
27 Company, Ltd.

1
2
3
4
5
6
7
8
9
10
11
12

CHANGE HISTORY

May 1995 Recommended Minimum Performance Standard for Digital Cellular Wideband Spread Spectrum Speech Service Option 1, IS-125

February 2000 Recommended Minimum Performance Standard for Digital Cellular Spread Spectrum Speech Service Option 1, IS-125-A

[This revision of the Minimum Performance Standard comprises the Alternative Partially-blocked ACR methodology for the testing of IS-96-C vocoders. IS-125-A also addresses substantive problems inherent in the performance of the original IS-125.]

1 Introduction

This standard provides definitions, methods of measurement, and minimum performance characteristics of IS-96-C [1]¹ variable-rate speech codecs for digital cellular wideband spread spectrum mobile stations and base stations. This standard shares the purpose of TIA/EIA-IS-98 [2] and TIA/EIA-IS-97 [3]. This is to ensure that a mobile station may obtain service in any cellular system that meets the compatibility requirements of TIA/EIA-IS-95-B [4].

This standard consists of this document and a software distribution on a CD-ROM. The CD-ROM contains:

- Speech source material
- Impaired channel packets produced from the master codec and degraded by a channel model simulation
- Calibration source material
- ANSI C language source files for the compilation of the master codec
- ANSI C language source files for a number of software data analysis tools
- Modulated Noise Reference Unit (MNRU) files
- A spreadsheet tool for data analysis

A detailed description of contents and formats of the software distribution is given in 3.6 of this document.

The variable-rate speech codec is intended to be used in both dual mode mobile stations and compatible base stations in the cellular service. This statement is not intended to preclude implementations in which codecs are placed at a Mobile Switching Center or elsewhere within the cellular system. Indeed, some mobile-to-mobile calls, however routed, may not require the use of a codec on the fixed side of the cellular system at all. This standard is meant to define the recommended minimum performance requirements of IS-96-C compatible variable-rate codecs, no matter where or how they are implemented in the cellular service.

Although the basic purpose of cellular telecommunications has been voice communication, evolving usages (for example, data) may allow the omission of some of the features specified herein, if that system compatibility is not compromised.

These standards concentrate specifically on the variable-rate speech codec and its analog or electro-acoustic interfaces, whether implemented in either the mobile station or the base station or elsewhere in the cellular system. They cover the operation of this component only to the extent that compatibility with IS-96-C is ensured.

1.1 Scope

This document specifies the procedures, which may be used to ensure that implementations of IS-96-C compatible variable-rate speech codecs meet recommended minimum performance requirements. This speech codec is the Service Option 1 described in IS-96-C. The Service Option 1 speech codec is used to digitally encode the speech signal for transmission at a variable data rate of 8550, 4000, 2000, or 800 bps.

¹Numbers in brackets refer to the reference document numbers.

1 Unlike some speech coding standards, IS-96-C does not specify a bit-exact description of the speech coding
2 system. The speech-coding algorithm is described in functional form, leaving exact implementation details of the
3 algorithm to the designer. It is, therefore, not possible to test compatibility with the standard by inputting
4 certain test vectors to the speech codec and examining the output for exact replication of a reference vector.
5 This document describes a series of tests that may be used to test conformance to the specification. These tests
6 do not ensure that the codec operates satisfactorily under all possible input signals. The manufacturer shall
7 ensure that its implementation operates in a consistent manner. These requirements test for minimum
8 performance levels. The manufacturer should provide the highest performance possible.

9 Testing the codec is based on two classes of procedures: objective and subjective tests. Objective tests are
10 based on actual measurements on the speech codec function. Subjective tests are based on listening tests to
11 judge overall speech quality. The purpose of the testing is not only to ensure adequate performance between
12 one manufacturer's encoder and decoder but also that this level of performance is maintained with operation
13 between any pairing of manufacturers' encoders and decoders. This interoperability issue is a serious one. Any
14 variation in implementing the exact standard shall be avoided if it cannot be ensured that minimum performance
15 levels are met when inter-operating with all other manufacturers' equipment meeting the standard. This standard
16 provides a means for measuring performance levels while trying to ensure proper interoperation with other
17 manufacturers' equipment.

18 The issue of interoperability may only be definitively answered by testing all combinations of encoder/decoder
19 pairings. With the number of equipment manufacturers expected to supply equipment, this becomes a
20 prohibitive task. The approach taken in this standard is to define an objective test on both the speech decoder
21 function and the encoder rate determination function to ensure that its implementation closely follows that of the
22 IS-96-C specification. This approach is designed to reduce performance variation exhibited by various
23 implementations of the decoder and rate determination functions. Because the complexity of the decoder is not
24 large, constraining the performance closely is not onerous. If all implementations of the decoder function
25 provide essentially similar results, then interoperation is more easily ensured with various manufacturers'
26 encoder implementations.

27 The objective and subjective tests rely upon the use of a master codec. This is a floating-point implementation
28 of IS-96-C written in the C programming language. The master codec is described more fully in 3.4. This
29 software is used as part of the interoperability testing.

30 By convention in this document, the `Courier` font is used to indicate C language and other software
31 constructs, such as file and variable names.

32 1.2 Definitions

33 **ANOVA** - Analysis of Variance. A statistical technique used to determine whether the differences among two
34 or more means are greater than would be expected from sampling error alone [11, 23].

35 **Base Station** - A station in the Domestic Public Cellular Radio Telecommunications Service, other than a mobile
36 station, used for radio communications with mobile stations.

37 **CELP** - Code Excited Linear Predictive Coding. These techniques use code-books to vector quantize the
38 excitation (residual) signal of a Linear Predictive Codec (LPC).

39 **Circum-aural Headphones** - Headphones that surround and cover the entire ear.

40 **Codec** - The combination of an encoder and decoder in series (encoder/decoder).

- 1 **dBA** - A-weighted sound pressure level expressed in decibels obtained by the use of a metering characteristic
2 and the weighting A specified in ANSI S1.4-1983 [5], and the addendum ANSI S1.4a-1985 [6].
- 3 **dBm0** - Power relative to 1 milli-watt. ITU-T Recommendation G.711 [7] specifies a theoretical load capacity with
4 a full-scale sine wave to be +3.17 dBm0 for μ -law PCM coding. In this standard, +3.17 dBm0 is also defined to be
5 the level of a full amplitude sine wave in the 16-bit 2's complement notation used for the data files.
- 6 **dBPa** - Sound level with respect to one Pascal, $20 \log_{10}$ (Pressure/1 Pa).
- 7 **dBQ** - The amount of noise expressed as a signal-to-noise ratio value in dB [20].
- 8 **dB SPL** - Sound Pressure Level in decibels with respect to 0.002 dynes/cm^2 , $20 \log_{10}$ (Pressure/0.002
9 dynes/cm²). dBPa is preferred.
- 10 **Decoder** - A device for the translation of a signal from a digital representation into an analog format. For the
11 purposes of this standard, a device compatible with IS-96-C.
- 12 **Encoder** - A device for the coding of a signal into a digital representation. For the purpose of this standard, a
13 device compatible with IS-96-C.
- 14 **FER** - Frame Error Rate equals the number of full rate frames received in error divided by the total number of
15 transmitted full rate frames.
- 16 **Full rate** - Encoded speech frames at the 8550 bps rate, 1/2 rate frames use the 4000 bps rate, 1/4 rate frames use
17 2000 bps, and 1/8 rate frames use 800 bps.
- 18 **Modified IRS** - Modified Intermediate Reference System [21, Annex D].
- 19 **MNRU** - Modulated Noise Reference Unit. A procedure to add speech correlated noise to a speech signal in
20 order to produce distortions that are subjectively similar to those produced by logarithmically companded PCM
21 systems. The amount of noise is expressed as a signal-to-noise ratio value in dB, and is usually referred to as
22 dBq [20].
- 23 **Mobile Station** - A station in the Domestic Public Cellular Radio Telecommunications Service. It is assumed that
24 mobile stations include portable units (for example, hand-held personal units) and units installed in vehicles.
- 25 **MOS** - Mean Opinion Score. The result of a subjective test based on an absolute category rating (ACR), where
26 listeners associate a quality adjective with the speech samples to which they are listening. These subjective
27 ratings are transferred to a numerical scale, and the arithmetic mean is the resulting MOS number.
- 28 **Newman-Keuls** - A statistical method for comparing means by using a multiple-stage statistical test, based on
29 Studentized Range Statistics [11, 23].
- 30 **ROLR** - Receive Objective Loudness Rating, a measure of receive audio sensitivity. ROLR is a frequency-
31 weighted ratio of the line voltage input signal to a reference encoder to the acoustic output of the receiver. IEEE
32 269 [9] defines the measurement of sensitivity and IEEE 661 [10] defines the calculation of objective loudness
33 rating.
- 34 **SAS** - Statistical Analysis Software.
- 35 **SNR_i** - The signal-to-noise ratio in decibels for a time segment, *i*. A segment size of 5 ms is used, which
36 corresponds to 40 samples at an 8 kHz sampling rate (see 2.1.1).
- 37 **SNRSEG** - Segmental signal-to-noise ratio. The weighted average over all time segments of SNR_i (see 2.1.1).

1 **T_{max}** - The maximum undistorted sinusoidal level that may be transmitted through the interfaces between the IS-
 2 96-C codec and PCM based network. This is taken to be a reference level of +3.17 dBm0.

3 **TOLR** - Transmit Objective Loudness Rating, a measure of transmit audio sensitivity. TOLR is a frequency-
 4 weighted ratio of the acoustic input signal at the transmitter to the line voltage output of the reference decoder
 5 expressed as a Transmit Objective Loudness Rating (TOLR). IEEE 269 defines the measurement of sensitivity
 6 and IEEE 661 defines the calculation of objective loudness rating.

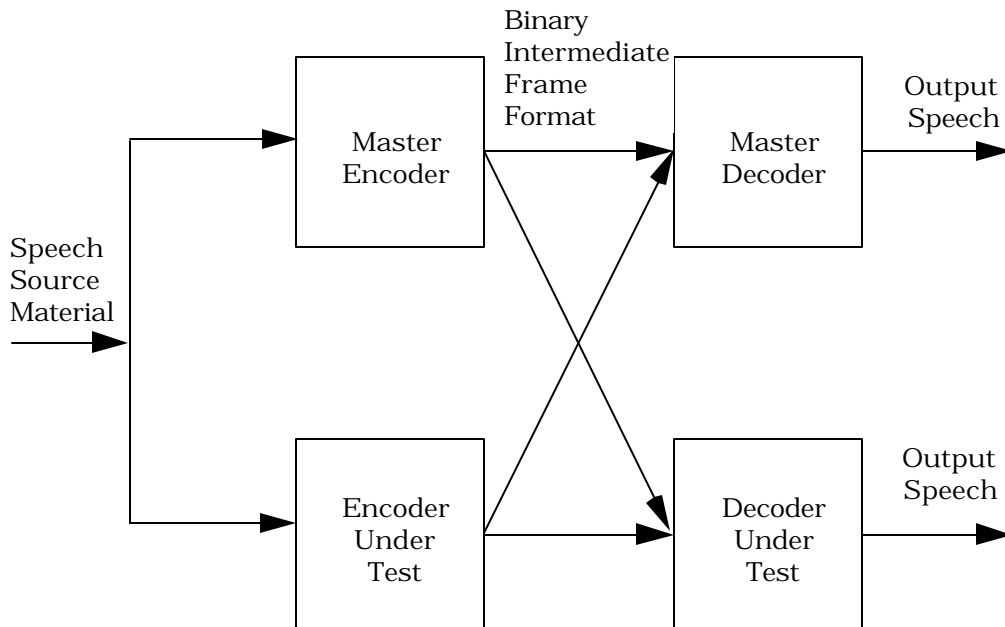
7 **1.3 Test Model for the Speech Codec**

8 For the purposes of this standard, a speech encoder is a process that transforms a stream of binary data samples
 9 of speech into an intermediate low bit-rate parameterized representation. As mentioned elsewhere in this
 10 document, the reference method for the performance of this process is given in IS-96-C. This process may be
 11 implemented in real-time, as a software program, or otherwise at the discretion of the manufacturer.

12 Likewise, a speech decoder is a process that transforms the intermediate low bit-rate parameterized
 13 representation of speech (given by IS-96-C) back into a stream of binary data samples suitable for input to a
 14 digital-to-analog converter followed by electro-acoustic transduction.

15 The test model compares the output streams of an encoder or decoder under test to those of a master encoder or
 16 decoder when driven by the same input stream (See Figure 1.3-1). The input stream for an encoder is a sequence
 17 of 16-bit linear binary 2's complement samples of speech source material. The output of an encoder shall
 18 conform to the packed frame format specified in 2.4.7.1-4 of IS-96-C corresponding to the rate selected for that
 19 frame. The representation of output speech is the same as that for input speech material.

20



21 **Figure 1.3-1. Test Model**

22

23 Various implementations of the encoder and decoder, especially those in hardware, may not be designed to
 24 deliver or accept a continuous data stream of the sort shown in Figure 1.3-1. It is the responsibility of the

1 manufacturer to implement a test platform that is capable of delivering and accepting this format in order to
2 complete the performance tests described below. This may involve parallel-to-serial data conversion hardware,
3 and vice versa; a fair implementation of the algorithm in software; or some other mechanism. A fair
4 implementation in software shall yield bit exact output with reference to any hardware implementation that it is
5 claimed to represent.

6 The input speech material has been precision-limited by an 8-bit μ -law quantization algorithm in which the
7 inverse quantized linear samples fill the entire 16-bit linear range. As specified in 2.4.2.1.1 of IS-96-C, the master
8 codec assumes a 14-bit integer input quantization. Thus the master encoder scales down the 16-bit linear
9 samples by 4.0 before processing, and the master decoder scales the output up by 4.0 after processing. In order
10 for the test codec to match the master codec, a similar pre-processing and post-processing scaling function shall
11 be performed. This may be done either internally by the test codec or externally by using a scaling function on
12 the input and output speech. (Note: one method of externally scaling the data is to use the program `scale.c`
13 supplied in directory `/tools` of the companion software, (see 3.3.6) with a 0.25 factor before encoder
14 processing and a 4.0 factor after decoder processing.)

1 No text.

2 CODEC MINIMUM STANDARDS

This section describes the procedures used to verify that the speech codec implementations meet minimum performance requirements. Minimum performance requirements may be met by either:

- Meeting strict objective performance requirements (see 2.1), or
- Meeting a combination of a less strict objective performance requirements and subjective performance requirements (see 2.2).

In the first case, objective performance requirements have to be met by both the encoder and decoder. In the latter case the decoder has to meet objective performance requirements for voice quality, while the encoder has to meet objective measures for encoding rate. Manufacturers are cautioned that the equivalent of a full floating-point implementation of the IS-96-C codec may be required to meet the performance requirements of the strict objective test. Minimum performance may be demonstrated, by meeting the requirements of either 2.1, or those of 2.2.

2.1 Strict Objective Performance Requirements

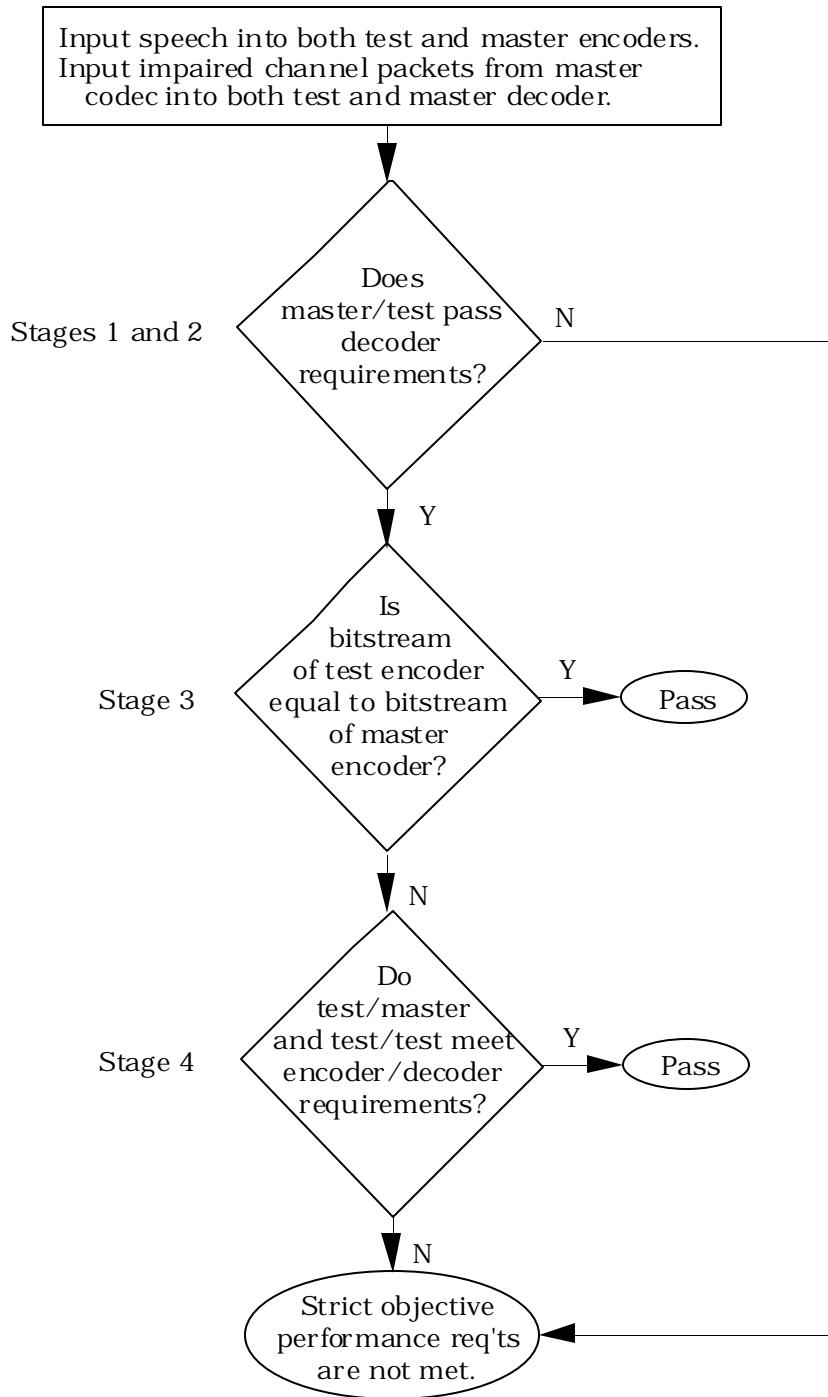
An objective measure of accuracy is obtained by computing the segmental signal-to-noise ratio (SNRSEG) between various encoder/decoder combinations. The specific measure of accuracy is based upon the distribution of SNR measurements collected over a large database.

The strict objective performance test may include up to four stages (see Figure 2.1-1).

In the first stage, the speech source material as described in 3.5 is input to both the master and test encoders. All speech source material supplied in the directory `/expt1/objective`, as described in Figure 3.6-1 of the companion software distribution, is to be used. The resulting output bit streams are saved. The output bit stream of the master encoder is input to both the test and master decoders. The output speech signals are compared using the procedure described in 2.1.1. If these signals agree according to the specific minimum performance requirements for the method defined in 2.1.1, they are considered to be in close agreement. If these signals are not in close agreement, then the strict objective performance requirements of this section are not met (see 2.2.1.3).

If there is close agreement, then the second stage of testing may proceed. The impaired bit stream from the master encoder, found in the file `/expt2/objective/expt2obj.pkt`, described in Figure 3.6-1, is input to both the test and master decoders. The output speech signals are compared using the procedure described in 2.1.1. If these signals are not in close agreement, then the strict objective performance requirements of this section are not met (see 2.2.1.3).

1



2

3

Figure 2.1-1. Flow Chart for Strict Objective Performance Test

1 If there is close agreement, then the third stage of testing may proceed. The output bit streams of the master
 2 encoder and test encoder are compared. If the bit streams are identical, then the test encoder/decoder has met
 3 the minimum performance requirements and no further testing is required.

4 If the bit streams are not identical, then the fourth stage of the strict test may be attempted. In this stage, the
 5 speech output from the test encoder/master decoder and test encoder/test decoder combinations are both
 6 compared with the output of the master encoder/master decoder using the procedure described in 2.1.1. If both
 7 combinations pass the acceptance criteria of 2.1.1, then the test encoder/decoder has met the minimum
 8 performance requirements and no further testing is required. If both combinations do not pass the acceptance
 9 criteria, then the strict objective performance requirements of this section are not met.

10 2.1.1 Strict Encoder/Decoder Objective Test Definition

11 The strict encoder/decoder objective test is intended to validate the implementation of the speech codec under
 12 test. The strict objective test is based upon statistical comparisons between the output of the master codec (see
 13 2.1) and combinations of the test encoder and decoder as listed below:

- 14 • Master encoder/test decoder (Stage 1)
- 15 • Master encoder with impaired packets/test decoder (Stage 2)
- 16 • Test encoder/master decoder (Stage 4)
- 17 • Test encoder/test decoder (also Stage 4)

18 This test is a measure of the accuracy of the codec being tested in terms of its ability to generate a bit-stream
 19 identical (or nearly identical) to that of the master codec when supplied with the same source material. The test
 20 is based upon statistics of the signal-to-noise ratio per 5 ms time segment, SNR_i . Defining the output samples of
 21 the master codec as $x_i(n)$ and those of the combination under test as $y_i(n)$, the SNR_i for a segment i is defined as

$$22 \quad SNR_i = \begin{cases} 10 \log_{10} \left(\frac{P_{x_i}}{R_i} \right) & \text{for } P_{x_i} \text{ or } P_{y_i} = T \\ 0 & \text{for } P_{x_i} \text{ and } P_{y_i} < T \end{cases}$$

24 where

$$25 \quad P_{x_i} = \frac{1}{K} \sum_{n=1}^K x_i(n)^2,$$

$$26 \quad P_{y_i} = \frac{1}{K} \sum_{n=1}^K y_i(n)^2, \text{ and}$$

$$R_i = \frac{1}{K} \sum_{n=1}^K [x_i(n) - y_i(n)]^2$$

where the summations are taken over $K=40$ samples and T is a power threshold value. Silence intervals are eliminated by employing the threshold T . If powers P_{x_i} and P_{y_i} are less than T , then this segment will not be included in any further computations. The threshold is chosen conservatively to be 54 dB ($20\log_{10}[512]$) below the maximum absolute value (R_{\max}) of the signal samples. The power threshold T is given by

$$T = \left(\frac{R_{\max}}{512} \right)^2 = 4096$$

Since the samples are 16 bits each with values in the range $[-32768, 32767]$, $R_{\max} = 32768$; furthermore, to prevent any one segment from dominating the average, the value of SNR_i is limited to a range of -5 to +80 dB. In particular, if $x_i(n)$ and $y_i(n)$ are identical, the SNR_i value is clipped to 80 dB.

Two statistics of the SNR_i are used. These are the average segmental signal-to-noise ratio (SNRSEG) [See 3.3.2]:

$$\text{SNRSEG} = \frac{1}{M} \sum_{i=1}^N \text{SNR}_i$$

where M is the number of segments where $\text{SNR}_i \neq 0$

and the cumulative distribution function $f(x)$ [See 3.3.2.2 and 3.3.3]:

$f(x)$ expresses the probability that an SNR_i value takes on a value less than or equal to some particular value x :

$$f(x) = p(\text{SNR}_i = x).$$

The encoder/decoder objective test is met when SNRSEG and $f(x)$ satisfy the specified criteria of the appropriate minimum performance specification.

2.1.2 Method of Measurement

For the codec being tested, the source-speech material, supplied in the directory `/expt1/objective` of the companion software distribution (see Figure 3.6-1), is processed by the master codec and by the codec being tested. The post-filters shall be used in both the master codec and the codec being tested. The master codec is to be initialized with all internal states equal to 0. The resulting linear output samples (8 kHz sampling rate) are time-aligned. The time-aligned output of both the master codec and the codec being tested is processed with the `snr.c` program or the equivalent (see 3.3.2) to produce the SNR_i for each 5 ms frame. The program `snr.c` generates a sample value of SNRSEG and a sample histogram, $h[1:n]$, for each continuous data stream presented to it. To generate the sample statistics for the entire database, the input files shall be concatenated

1 and presented in one pass to the analysis program. The resulting measurements are used to compute the
2 segmental

1 signal-to-noise ratio, SNRSEG, and a sample histogram, $h[1:n]$, of the SNR_i . The histogram is computed
 2 using 64 bins, with boundaries being the values two through 64. SNR_i values below 2 dB are included in the first
 3 bin, while SNR_i values above 64 are included in the last bin. Segments corresponding to silence should not be
 4 included in the histogram; thus, given a non-zero SNR_i value, x , assign it to a bin according to Table 2.1.2-1.

6 **Table 2.1.2-1. Computing the Segmental Signal-to-Noise Ratio**

SNR Values	Bin Assignment
$x < 2$	bin 1
$2 \leq x < 3$	bin 2
...	...
$n \leq x < n+1$	bin n
...	...
$x \geq 64$	bin 64

7
 8 The procedure is outlined below, where the function $\text{int}(x)$ truncates to the nearest integer whose value is
 9 less than or equal to the value of x :

```

10
11     snr[1:n]    SNRi values, snr[i] = 0.0 indicates silence
12     h[1:64]    histogram values
13     npoint     number of data points used to compute histogram
14     SNRSEG     average value of SNRi values
15     SNRSEG = 0.0
16     npoint = 0
17     for (i = 1, N)
18         if(snr[i] not equal to 0.0) then
19             {
20                 npoint = npoint + 1
21                 SNRSEG = SNRSEG + snr[i]
22                 if(snr[i] < 2.0) then
23                     h[1] = h[1] + 1
24                 else if(snr[i] ≥ 64.0) then
25                     h[64] = h[64] + 1
26                 else
27                     h[int(snr[i])] = h[int(snr[i])] + 1
28             }
29     SNRSEG = SNRSEG / npoint
30

```

31 The histogram array is converted into a sample cumulative distribution function, $f[1:n]$, as follows:

```

32
33     h[1:64]    histogram values
34     npoint     number of data points used to compute histogram
35     f[1:64]    normalized cumulative distribution function

```

```

1      f[1] = h[1]/npoint
2      for (i = 2, 64)
3          f[i] = f[i-1] + h[i] / npoint

```

4 An ANSI C source language program, `cdf.c`, that implements this procedure is given in 3.3.3.

5 2.1.3 Minimum Performance Requirements

6 The codec being tested meets minimum performance requirements if all of the three following requirements are
7 met over the entire source speech material database of directory `/expt1/objective`:

- 8 • $f[14] = 0.01$, which means that the SNR_i can have a value lower than 15 dB no more than 1% of the time;
9 and
- 10 • $f[24] = 0.05$, which means that the SNR_i can have a value lower than 25 dB no more than 5% of the time;
11 and
- 12 • $SNRSEG = 40$ dB.

13 2.2 Combined Objective/Subjective Performance Requirements

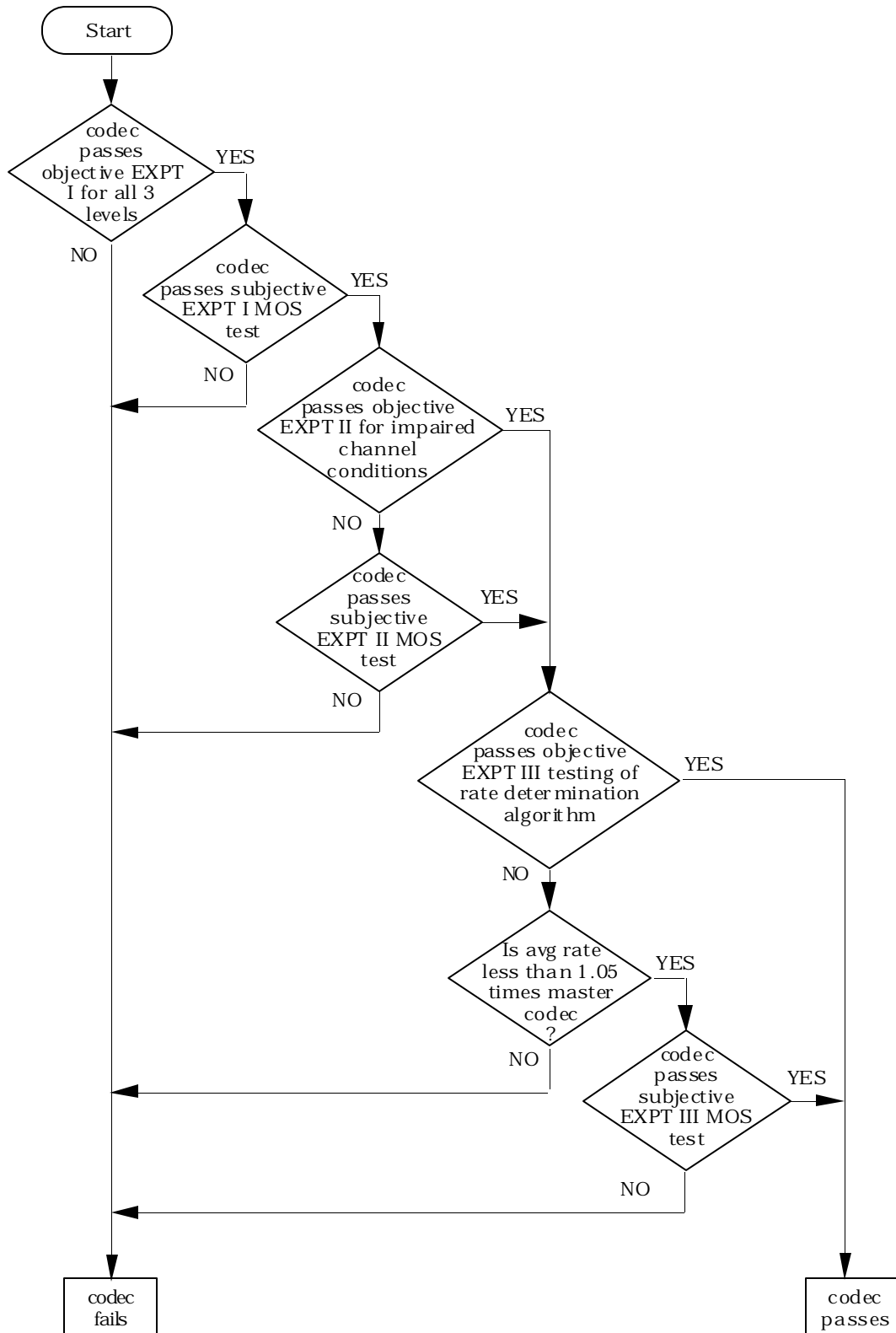
14 If the strict objective test has failed or is not attempted, then the codec being tested may be evaluated with the
15 combined objective and subjective tests described in this section. These objective and subjective tests are
16 divided into three experiments:

- 17 • Experiment I tests the codec under various speaker and input level conditions. Both objective and
18 subjective tests shall be met to satisfy minimum performance requirements.
- 19 • Experiment II tests the decoder under impaired channel conditions. If the objective performance criteria
20 of this test are met, it is not necessary to perform the subjective component of the test. If the objective
21 performance criteria of this test are not met, the subjective test for channel impairments shall be met to
22 satisfy minimum performance requirements.
- 23 • Experiment III of the test insures that the rate determination algorithm in the encoder performs correctly
24 under a variety of background noise conditions. If the objective performance criteria of this test are
25 satisfied, it is not necessary to perform the subjective component of the test. If the objective
26 performance criteria of this test are not met, and the average data rate of the codec being tested is less
27 than 1.05 times the master codec average data rate, the subjective test for background noise degradation
28 shall be met to satisfy minimum performance requirements. If the average data rate of the codec being
29 tested is greater than 1.05 times the master codec average data rate, then the codec being tested fails the
30 minimum performance requirements.

31 Figure 2.2-1 summarizes the steps necessary to satisfy the combined objective/subjective test requirements.

32 The objective performance requirement for Experiments I and II (see 2.2.1 and 2.2.2) evaluate the decoder
33 component of the codec being tested. The philosophy of the test is the same as the strict objective test (see
34 2.1). The objective performance requirement for Experiment III evaluates the encoder component of the codec
35 being tested, specifically the rate determination mechanism. The method of measurement for comparing the rate
36 output from the master encoder and the test encoder is described in 2.2.3.2.

37 The subjective performance requirements for Experiments I, II and III (see 2.2.4) evaluate the codec being tested
38 against the master codec performance through the use of a panel of listeners and a Mean Opinion Score criteria.



1
2

Figure 2.2-1. Objective/Subjective Performance Test Flow

1 2.2.1 Objective Performance Experiment I

2 2.2.1.1 Definition

3 The objective performance Experiment I is intended to validate the implementation of the decoder section of the
 4 speech codec being tested under various speaker and input level conditions. The objective performance
 5 Experiment I that tests the decoder is based upon the same philosophy as the strict objective test (see 2.1.1).
 6 Only one combination of encoder and decoder is evaluated against the master encoder/decoder, with the master
 7 encoder driving the test decoder. The performance requirements are based upon sample statistics SNRSEG and
 8 $f[1:n]$, which estimate segmental signal-to-noise ratio, SNRSEG, and the cumulative distribution function,
 9 $F(x)$, of the signal-to-noise ratios per segment, SNR_j . The codec being tested passes the objective performance
 10 requirement of Experiment I if the master encoder/test decoder combination satisfies the criteria of 2.2.1.3.

11 2.2.1.2 Method of Measurement

12 The method of measurement for decoder comparison is the same as for the strict objective test (see 2.1.2).

13 Note: the post-filter shall be used in these objective experiments.

14 2.2.1.3 Minimum Performance Requirements for Experiment I

15 The codec being tested meets the minimum requirements of the objective performance Experiment I, if all of the
 16 three criteria are met over the entire source speech material database in directory `/expt1/objective`.

17 The objective test shall be run three times, once for each of the three levels, -29, -19, and -9 dBm0. These levels
 18 are provided in the source files

19 `/expt1/objective/expt1obj.x29` ,
 20 `/expt1/objective/expt1obj.x19` and
 21 `/expt1/objective/expt1obj.x09`.

22 The numeric bounds given in Sections 2.2.1.3.1 through 2.2.1.3.3 are the performance requirements for Experiment
 23 I.

24 2.2.1.3.1 Minimum Performance Requirements for Experiment I Level -9 dBm0

- 25 • $f[1] = 0.01$, which means that the SNR can have a value lower than 2 dB no more than 1% of the time; and
- 26 • $f[15] = 0.06$, which means that the SNR can have a value lower than 16 dB no more than 6% of the time;
 27 and
- 28 • SNRSEG = 22 dB.

29 2.2.1.3.2 Minimum Performance Requirements for Experiment I Level -19 dBm0

- 30 • $f[1] = 0.01$, which means that the SNR can have a value lower than 2 dB no more than 1% of the time; and
- 31 • $f[15] = 0.14$, which means that the SNR can have a value lower than 16 dB no more than 14% of the time;
 32 and
- 33 • SNRSEG = 20 dB.

2.2.1.3.3 Minimum Performance Requirements for Experiment I Level -29 dBm0

- $f[1] = 0.03$, which means that the SNR can have a value lower than 2 dB no more than 3% of the time; and
- $f[15] = 0.37$, which means that the SNR can have a value lower than 16 dB no more than 37% of the time; and
- $SNRSEG = 16$ dB.

2.2.2 Objective Performance Experiment II

2.2.2.1 Definition

The objective performance Experiment II of the combined test, which validates the implementation of the decoder section of the speech codec under impaired channel conditions, is based on the same philosophy as the strict objective test (see 2.1.1). Only one combination of encoder and decoder is evaluated against the master encoder/decoder: the master encoder driving the test decoder.

Impaired channel packets from the master encoder used to drive the master decoder and test decoder for this test are found in the file `/expt2/objective/expt2obj.pkt` of the companion software. The performance requirements are based upon sample statistics $SNRSEG$ and $f[1:n]$, which estimate segmental signal-to-noise ratio, $SNRSEG$, and the cumulative distribution function, $F(x)$, of the signal-to-noise ratios per segment, SNR_j . The codec under test passes, if the master encoder/test decoder combination satisfies the criteria of 2.2.2.3.

If the codec being tested passes the objective test in Experiment II, impaired channel operation has been sufficiently tested and subjective Experiment II is not required for minimum performance validation. However, if the codec fails the objective test in Experiment II (presumably because the codec manufacturer has elected to implement a different error masking algorithm than specified in IS-96-C), then subjective Experiment II is required.

2.2.2.2 Method of Measurement

The method of measurement for decoder comparison is the same as for the strict objective test (see 2.1.2).

Note: the post-filter shall be used in these objective experiments.

2.2.2.3 Minimum Performance Requirements for Experiment II

The codec being tested meets the minimum requirements of the objective performance part if all of the three following criteria are met over the entire packet speech material database found in the file `expt2/objective/expt2obj.pkt`.

The following numeric bounds are the performance requirements for Experiment II:

- $f[1] = 0.01$, which means that the SNR can have a value lower than 2 dB no more than 1% of the time; and
- $f[15] = 0.05$, which means that the SNR can have a value lower than 16 dB no more than 5% of the time; and
- $SNRSEG = 22$ dB.

1 2.2.3 Objective Performance Experiment III

2 2.2.3.1 Definition

3 The objective performance Experiment III of the combined test is intended to validate the implementation of the
4 rate determination algorithm in the encoder section of the speech codec under various background noise
5 conditions. The background noise and level-impaired source material may be found in the files

```
6 /expt3/objective/expt3obj.x29,  
7 /expt3/objective/expt3obj.x19, and  
8 /expt3/objective/expt3obj.x09
```

9 of the companion software. These files vary the signal-to-noise ratio of the input speech from 15 dB to 35 dB
10 and back to 23 dB (in 2 dB steps) while keeping the signal energy constant at -29, -19, and -9 dBm0, respectively,
11 for each of the three files.

12 Each file is approximately 68 seconds in duration and contains seven sentence pairs. Each sentence pair is
13 separated by at least four seconds of silence. The noise source added digitally to each speech file varies the
14 noise level in 2 dB steps every four seconds to produce the signal-to-noise ratios described above. The codec
15 being tested passes the objective test in Experiment III, if the test encoder rate statistics match the master
16 encoder rate statistics within the criteria of 2.2.3.3 for all three input levels represented by the three input source
17 files.

18 If the codec being tested passes the objective test in Experiment III, the test-codec rate-determination algorithm
19 has been sufficiently tested and subjective Experiment III is not required for minimum performance validation.
20 However, if the codec fails the objective test in Experiment III (presumably because the codec manufacturer has
21 elected to implement a different rate determination algorithm than that which is specified in IS-96-C) then
22 subjective Experiment III shall be performed and shall be passed.

23 2.2.3.2 Method of Measurement

24 For the rate determination objective test, the method of measurement is performed by the program
25 `rate_chk.c`. The background noise corrupted source files

```
26 /expt3/objective/expt3obj.x29,  
27 /expt3/objective/expt3obj.x19, and  
28 /expt3/objective/expt3obj.x09
```

29 are encoded through both the master encoder and the test encoder representing the three input levels -29, -19,
30 and -9 dBm0, respectively.

31 The encoded packets are then input into the program `rate_chk.c`. This program strips off the rate
32 information from both the test and master packets and compares the rates on a frame by frame basis, compiling a
33 histogram of the joint rate decisions made by both the master and the test codecs. This histogram of the joint
34 rate decisions and the average rate from both the test and master codecs are used to judge the test-encoder rate-
35 determination algorithm. An example of the joint rate histogram statistics and average rate calculations output
36 by the program `rate_chk.c` is shown in Table 2.2.3.2-1.

37

1 **Table 2.2.3.2-1. Joint Rate Histogram Example**

Master Codec	Test Codec			
	1/8 Rate	1/4 Rate	1/2 Rate	Full Rate
1/8 rate	.4253	.0012	0.0	0.0
1/4 rate	.0187	.1559	.0062	0.0
1/2 rate	0.0	.0027	.0676	.0027
full rate	0.0	0.0	.0003	.3328

- 2
- Average rate of the master encoder: 3.710 kbps
 - Average rate of the test encoder: 3.696 kbps
- 3

4

5 The C source language program, `rate_chk.c`, that implements this procedure is described in 3.3.5.

6 **2.2.3.3 Minimum Performance Requirements**

7 The codec being tested meets the minimum requirements of Experiment III of the objective test, if the following
 8 rate statistic criteria are met over all three levels of the input speech source. If the average data rate of the test
 9 encoder is less than 1.05 times the average data rate of the master encoder, then the subjective part of this test
 10 (see subjective Experiment III in 2.2.4.2.1) shall be used to meet the minimum performance requirements.

11 If the test encoder average data rate is more than 1.05 times the average data rate of the master encoder, and the
 12 test codec does not meet the requirements of objective Experiment III, then the codec being tested fails the
 13 minimum performance requirements, and the subjective part of this test cannot be used to meet the minimum
 14 performance requirements.

15 **2.2.3.3.1 Minimum Performance Requirements for Level -9 dBm0**

16 In order for the minimum performance requirements for level -9 dBm0 to be met, the joint rate histogram
 17 probabilities output from the program `rate_chk.c` shall be no greater than the non-diagonal elements
 18 specified with boldface characters in Table 2.2.3.3.1-1. The overall average rate of the test codec shall be less
 19 than 1.05 times the average rate of the master codec. The numeric bounds given in 2.2.3.3.1 through 2.2.3.3.3 are
 20 the performance requirements for Experiment III.

1 **Table 2.2.3.3.1-1. Joint Rate Histogram Requirements for -9 dBm0**

Master Codec	Test Codec			
	1/8 Rate	1/4 Rate	1/2 Rate	Full Rate
1/8 rate	.4409	.0176	0.0	0.0
1/4 rate	.0394	.1577	.0063	0.0
1/2 rate	0.0	.0136	.0679	.0027
full rate	0.0	0.0	.0133	.3334

2
3 2.2.3.3.2 Minimum Performance Requirements for Level -19 dBm0

4 In order for the minimum performance requirements for level -19 dBm0 to be met, the joint rate histogram
5 probabilities output from the program `rate_chk.c` shall be no greater than the non-diagonal elements
6 specified with boldface characters in Table 2.2.3.3.2-1. The overall average rate of the codec being tested shall
7 be less than 1.05 times the average rate of the master codec.

8
9 **Table 2.2.3.3.2-1. Joint Rate Histogram Requirements for -19 dBm0**

Master Codec	Test Codec			
	1/8 Rate	1/4 Rate	1/2 Rate	Full Rate
1/8 rate	.4841	.0193	0.0	0.0
1/4 rate	.0352	.1409	.0056	0.0
1/2 rate	0.0	.0107	.0535	.0021
full rate	0.0	0.0	.0129	.3216

10
11 2.2.3.3.3 Minimum Performance Requirements for Level -29 dBm0

12 In order for the minimum performance requirements for level -29 dBm0 to be met, the joint rate histogram
13 probabilities output from the program `rate_chk.c` shall be no greater than the non-diagonal elements
14 specified with boldface characters in Table 2.2.3.3.3-1. The overall average rate of the test codec shall be less
15 than 1.05 times the average rate of the master codec.

1 **Table 2.2.3.3.3-1. Joint Rate Histogram Requirements for -29 dBm0**

Master Codec	Test Codec			
	1/8 Rate	1/4 Rate	1/2 Rate	Full Rate
1/8 rate	.5095	.0204	0.0	0.0
1/4 rate	.031	.1240	.0050	0.0
1/2 rate	0.0	.0108	.0538	.0022
full rate	0.0	0.0	.0125	.3128

2
3 2.2.4 Speech Codec Subjective Performance

4 This section outlines the subjective testing methodology of the combined objective/subjective performance test.

5 2.2.4.1 Definition

6 The codec subjective test is intended to validate the implementation of the speech codec being tested using the
7 master codec defined in 3.4 as a reference. The subjective test is based on the Absolute Category Rating or
8 Mean Opinion Score (MOS) test as described in ITU-T Recommendation P.800, Annex B [19].

9 2.2.4.2 Method of Measurement

10 The subjective test involves a listening-only assessment of the quality of the codec being tested using the
11 master codec as a reference. Subjects from the general population of telephone users will rate the various
12 conditions of the test. Material supplied with this standard for use with this test includes: source speech,
13 impaired packet files from the master codec encoder, and source speech processed by various Modulated Noise
14 Reference Unit (MNRU) conditions. The concept of the MNRU is fully described in ITU-T Recommendation
15 P.810 [20]. These MNRU samples were generated by the ITU-T software program mnrudemo. The basic
16 Absolute Category Rating test procedure involves rating all conditions using a five-point scale describing the
17 listener's opinion of the test condition. This procedure is fully described in ITU-T Recommendation P.800,
18 Annex B [19].

19 2.2.4.2.1 Test Conditions

20 The conditions for the test will be grouped into the following three experiments described in 2.2.4.2.1.1 through
21 2.2.4.2.1.3.

22 2.2.4.2.1.1 Subjective Experiment I

- 23 • Four codec combinations (master encoder/master decoder, master encoder/test decoder, test
24 encoder/master decoder, test encoder/test decoder).
- 25 • All codec combinations performed with three input levels (-29, -19, and -9 dBm0).
- 26 • Six Modulated Noise Reference Unit (MNRU reference conditions (40 dBq, 34 dBq, 28 dBq, 22 dBq, 16
27 dBq, and 10 dBq).

28 All 18 conditions listed in 2.2.4.2.1.1 shall be tested with each of 16 talkers.

1 For conditions with increased or reduced input level, the output levels shall be restored to the same nominal
2 average output level as the rest of the conditions.

3 The 18 combined conditions together with the use of 16 talkers per condition yield a total of 288 test trials. Each
4 listener will evaluate each of the 288 stimuli.

5 To control for possible differences in speech samples, the four encoder/decoder combinations shall be tested
6 with the same speech samples. This is done in the following way to avoid having individual listeners hear
7 samples repeatedly. Four samples for each talker shall be processed through each encoder/decoder
8 combination, at each level. These samples will be divided into four sets of samples called file-sets (one each of
9 the master/master, master/test, test/master, and test/test combinations at each level), with a unique sample per
10 talker for each combination within each file-set. Each file-set will be presented to one-quarter of the 48 listeners.
11 This design will contribute to control of any differences in quality among the source samples.

12 The 12 codec conditions together with the use of four sentence-pairs for each of 16 talkers per codec condition
13 yield a total of four sets of 192 stimuli. Each of these in combination with six reference conditions together with
14 the use of one sentence-pair for each of 16 talkers yields a total of four sets of 288 test trials per set. One-quarter
15 of the 48 listeners will each evaluate one of the four different file-sets of 288 stimuli.

16 2.2.4.2.1.2 Subjective Experiment II

17 Subjective Experiment II is designed to verify that the test decoder meets minimum performance requirements
18 under impaired channel conditions. If the test decoder doesn't meet the objective measure specified in 2.2.2.3,
19 then the subjective test in Experiment II shall be performed to pass minimum performance requirements. The
20 codecs, conditions and reference conditions used in the test are the following:

- 21 • Two codec combinations (master encoder/master decoder, master encoder/test decoder).
- 22 • All codec combinations performed under two channel impairments (1% and 3% FER).
- 23 • Three Modulated Noise Reference Unit (MNRU) reference conditions (30 dBq, 20 dBq, 10 dBq).

24 All seven conditions listed in 2.2.4.2.1.2 shall be tested with each of six talkers.

25 Four samples for each talker shall be processed through each encoder/decoder combination, at each channel
26 condition. These samples will be divided into two sets of samples (two each of the master/master, master/test
27 combinations at each channel condition) with two unique samples per talker for each combination within each
28 set. Each set will be presented to one-half of the listeners. This design will contribute to control of any
29 differences in quality among the source samples.

30 The four codec conditions together with the use of four sentence-pairs for each of 6 talkers per codec condition
31 yield a total of two sets of 48 stimuli. Each of these in combination with three reference conditions together with
32 the use of two sentence-pairs for each of 6 talkers yields a total of two sets of 84 test trials per set. One-half of
33 the listeners will each evaluate one of the two different sets of 84 stimuli.

34 The impaired channel packets for FERs of 1% and 3% to be processed by the master and test decoders for this
35 subjective test may be found under the directory /expt2/subjective/ed_pkts.

36

2.2.4.2.1.3 Subjective Experiment III

The purpose of Experiment III is to verify that the test encoder meets minimum performance requirements under various background noise conditions. If the test encoder does not produce rate statistics sufficiently close to that of the master encoder as specified by the objective measure defined in 2.2.3.3, and the test codec produces an average data rate of no more than 1.05 times the master codec average data rate, then the subjective test in Experiment III shall be performed to pass minimum performance requirements. The codecs, conditions and reference conditions used in this test are the following:

- Two codec combinations (master encoder/master decoder, test encoder/master decoder).
- All codec combinations shall be performed under three background noise impairments combined with three input noise levels. The three noise conditions are no added noise, street noise at 20 dB SNR, and road noise at 15 dB SNR. The three levels are -9, -19, and -29 dBm0.
- Six Modulated Noise Reference Unit (MNRU) reference conditions (35 dBq, 30 dBq, 25 dBq, 20 dBq, 15 dBq, 10 dBq).

All 24 conditions listed in 2.2.4.2.1.3 shall be tested with each of four talkers.

Two samples for each S/N condition for each talker shall be processed through each encoder/decoder combination, at each level. These samples will be divided into two sets of samples (one each of the master/master, test/master combinations at each S/N condition, at each level) with a unique sample per talker for each combination within each set. Each set will be presented to one-half of the listeners. This design will contribute to control of any differences in quality among the source samples.

The 18 codec conditions together with the use of two sentence-pairs for each of 4 talkers per codec condition yield a total of two sets of 72 stimuli. Each of these in combination with six reference conditions together with the use of one sentence-pair for each of 4 talkers yields a total of four sets of 96 test trials per set. One-half of the listeners will each evaluate one of the two different sets of 96 stimuli.

The MNRU conditions for all three experiments are used to provide a frame of reference for the MOS test. In addition, the MNRU conditions provide a means of comparing results between test laboratories. The concept of the MNRU is fully described in ITU-T Recommendation P.810 [20].

2.2.4.2.2 Source Speech Material

All source material will be derived from the Harvard Sentence Pair Database [24] and matched in overall level. While individual sentences will be repeated, every sample uses a distinct sentence pairing. Talkers were chosen to have distinct voice qualities and are native speakers of North American English.

2.2.4.2.3 Source Speech Material for Experiment I

The source speech material for subjective Experiment I consists of 288 sentence pairs. Each sentence will be full band and μ -law companded. The talkers in subjective Experiment I consist of seven adult males, seven adult females, and two children.

Directory `/expt1/subjective` contains the subdirectories of source material for Experiment I which consists of 18 sentence pairs from 16 different speakers for a total of 288 speech files. The speech database also includes source samples processed through the MNRU in directory `/expt1/subjective/mnru`. A total of 96 MNRU samples have been provided at the appropriate levels as defined by Experiment I. Practice samples for this experiment have been provided in the directory `/expt1/subjective/samples`. The subdirectories

1 under /expt1/subjective partition the speech database according to what processing is required on the
2 source material (see 3.6). For example, the speech files in the directory
3 /expt1/subjective/ed_srcs/level_29 are level -29 dBm0 files that shall be processed by all 4
4 encoder/decoder codec combinations for subjective Experiment I.

6 2.2.4.2.4 Source Speech Material for Experiment II

7 The source speech material for subjective Experiment II consists of 84 sentence pairs. Each sentence will be IRS
8 filtered and μ -law companded. The talkers in subjective Experiment II consist of three adult males and three adult
9 females. Since this experiment is only a decoder test, impaired packets from the master codec have been
10 provided as input to both the master and test decoder at FER of 1% and 3% (see directories
11 /expt2/subjective/ed_pkt/fer_1%, and /expt2/subjective/ed_pkt/fer_3%. Some of
12 the impaired packets have been selectively impaired to check internal operation under error bursts, all ones rate
13 1/8 packets (see 2.4.8.6.5 of IS-96-C), and blanked packets (see 2.4.8.6.4 of IS-96-C). For testing purposes these
14 selectively impaired packets have been duplicated and renamed in the files

15 /expt2/objective/bursts.pkt,
16 /expt2/objective/null1_8.pkt, and
17 /expt2/objective/blanks.pkt.

18 The speech database also includes source samples processed through the MNRU in directory
19 /expt2/subjective/mnru. A total of 36 MNRU samples have been provided at the appropriate levels
20 as defined by Experiment II. Practice samples for this experiment have been provided in the directory
21 /expt2/subjective/samples. The subdirectories under /expt2/subjective partition the
22 speech database according to required processing of the source material (see 3.6). For example, the packet files
23 in the directory /expt2/subjective/ed_pkts/fer_3% are 3% FER packets that should be processed
24 by the master decoder and test decoder for subjective Experiment II.

26 2.2.4.2.5 Source Speech Material for Experiment III

27 The source speech material for subjective Experiment III consists of 96 sentence pairs. Each sentence will be full
28 band and μ -law companded. The talkers in subjective Experiment III consist of two adult males and two adult
29 females. The desired signal-to-noise ratio was realized by digitally mixing additive noise to the input speech files
30 at the appropriate levels.

31 The source material for this test may be found under the directory /expt3/subjective. The 24 MNRU
32 reference conditions required by this experiment may be found in the directory /expt3/subjective/mnru.
33 Practice samples for this experiment have been provided in the directory /expt3/subjective/samples.
34 The subdirectories under /expt3/subjective partition the speech database according to required
35 processing of the source material (see 3.6). For example, the speech files in the directory
36 /expt3/subjective/ed_srcs/20dB/level_29 are -29 dBm0 level, 20 dB SNR files that need to be
37 processed by the master encoder/master decoder and test encoder/master decoder codec combinations for
38 subjective Experiment III.

1 2.2.4.2.6 Processing of Speech Material

2 The source speech sentence-pairs shall be processed by the combinations of encoders and decoders listed in
 3 the descriptions of the three experiments given in 2.2.4.2.1. The speech database has been structured to facilitate
 4 the partitioning of the speech material for the respective master/test encoder/decoder codec combinations and
 5 conditions of each of the three experiments. For instance, the source material for Experiment I has the following
 6 subdirectories under `/expt1/subjective`:

```
7     ed_srcs/level_29 ,
8     ed_srcs/level_19 , and
9     ed_srcs/level_9 .
```

10 where `ed_srcs` represent the 4 sources for the encoder/decoder combinations, at each level. The level
 11 directories contain the source used for the various level conditions under test. Thus the source processed by all
 12 encoders and decoders for the level -9 dBm0 condition may be found in the speech database at
 13 `/expt1/subjective/ed_src/level_9`. The subdirectories have been organized in a similar manner for
 14 subjective Experiments II and III (see Figure 3.6-1).

15 For Experiment I, the source material for all four of the encoder/decoder combinations for each input level shall
 16 be processed through each of the four encoder/decoder combinations, at that input level. A similar
 17 consideration should be given to Experiments II and III as well.

18 The master codec software described in 3.4 shall be used in the processing involving the master codec. All
 19 processing shall be done digitally. The post-filter shall be enabled for both the master decoder and the test
 20 decoder.

21 The digital format of the speech files is described in 3.4.4.

22 2.2.4.2.7 Randomization

23 For each of the three subjective experiments, each presentation sample consists of one sentence pair processed
 24 under a condition of the test. The samples shall be presented to the subjects in a random order. The subjects
 25 shall be presented with the 16 practice trials for subjective Experiments I, II, and III. The randomization of the
 26 test samples shall be constrained in the following ways for all three experiments:

- 27 1. A test sample for each codec combination, talker and level, channel condition, or background noise
 28 level (Experiment I, II, or III) or MNRU value and talker is presented exactly once.
- 29 2. Randomization shall be done in blocks, such that one sample of each codec/level, codec/channel
 30 condition, or codec/background noise level (again depending on Experiment I, II, or III) or MNRU
 31 value will be presented once, with a randomly selected talker, in each block. This ensures that subjects
 32 rate each codec/condition being tested equally often in the initial, middle and final parts of the session
 33 controlling the effects of practice and fatigue. Such a session will consist of 16 blocks for Experiment I,
 34 6 blocks for Experiment II, and 4 blocks for Experiment III. A particular randomization shall not be
 35 presented to more than four listeners.
- 36 3. Talkers shall be chosen so that the same talker is never presented on two consecutive trials within the
 37 same block.

1 The 16 practice trials are already randomized for Experiments I, II, and III. Readme files have been provided in the
2 "samples" directories that give the randomization for the practice trials. Software to accomplish the required
3 randomization is described in 3.3.4.

4 2.2.4.2.8 Presentation

5 Presentation of speech material shall be made with one side of high fidelity circum-aural headphones. The
6 speech material delivery system shall meet the requirements of 3.1.2.1. The delivery system shall be calibrated to
7 deliver an average listening level of -16 dBPa (78 dB SPL). The equivalent acoustic noise level of the delivery
8 system should not exceed 30 dBA as measured on a standard A-weighted meter.

9 Listeners should be seated in a quiet room, with an ambient noise of 35 dBA or below.

10 2.2.4.2.9 Listeners

11 The subject sample is intended to represent the population of telephone users with normal hearing acuity.
12 Therefore, the listeners should be naïve with respect to telephony technology issues; that is, they should not be
13 experts in telephone design, digital voice encoding algorithms, and so on. They should not be trained listeners;
14 that is, they should not have been trained in these or previous listening studies using feedback trials. The
15 listeners should be adults of mixed sex and age. Listeners over the age of 50, or others at risk for hearing
16 impairment, should have been given a recent audio-gram to ensure that their acuity is still within the normal
17 sensitivity range.

18 Each listener shall provide data only once for a particular evaluation. A listener may participate in different
19 evaluations, but test sessions performed with the same listener should be at least one month apart to reduce the
20 effect of cumulative experience.

1 2.2.4.2.10 Listening Test Procedure

2 The subjects shall listen to each sample and rate the quality of the test sample using a five-point scale, with the
3 points labeled:

- 4 1. Bad
- 5 2. Poor
- 6 3. Fair
- 7 4. Good
- 8 5. Excellent

9 Data from 48 subjects shall be used for each of the three experiments. The experiment may be run with up to four
10 subjects in parallel; that is, the subjects shall hear the same random order of test conditions.

11 Before starting the test, the subjects should be given the instructions in Figure 2.2.4.2.10-1. The instructions
12 may be modified to allow for variations in laboratory data-gathering apparatus.

13

<p>This is an experiment to determine the perceived quality of speech over the telephone. You will be listening to a number of recorded speech samples, spoken by several different talkers, and you will be rating how good you think they sound.</p> <p>The sound will appear on one side of the headphones. Use the live side on the ear you normally use for the telephone.</p> <p>On each trial, a sample will be played. After you have listened to each passage, the five buttons on your response box will light up. Press the button corresponding to your rating for how good or bad that particular passage sounded.</p> <p>During the session you will hear samples varying in different aspects of quality. Please take into account your total impression of each sample, rather than concentrating on any particular aspect.</p> <p>The quality of the speech should be rated according to the scale below:</p>				
Bad	Poor	Fair	Good	Excellent
1	2	3	4	5
<p>Rate each passage by choosing the word from the scale which best describes the quality of speech you heard. There will be 288 trials, including 16 practice trials at the beginning.</p> <p>Thank you for participating in this research.</p>				

14

15

Figure 2.2.4.2.10-1. Instructions for Subjects

16

2.2.4.2.11 Analysis of Results

The response data from the practice sessions shall be discarded. Data sets with missing responses from subjects shall not be used. Responses from the different sets of encoder/decoder processed files shall be treated as equivalent in the analysis.

A mixed-model, repeated-measures Analysis of Variance of Observations (ANOVA) [11, 23] shall be performed on the data for the codec conditions. The α to be used for tests of significance is 0.01. That is, $p < 0.01$ will be taken as statistically reliable, or significant. The appropriate ANOVA table is shown in Table 2.2.4.2.11-4. A sample SAS program to perform this computation is given in Table 3.3.1.1-1 [12]. This program also computes the required cell means of the four terms: *codecs x levels*, *codecs x talkers*, *levels x talkers*, and *codecs x levels x talkers*. Any equivalent tool capable of repeated measures analysis may be used. The MNRU condition data is not used for analysis.

In subjective Experiment I, if the main effect of codec combinations (*codecs*) is significant, a Newman-Keuls comparison of means [11, 23] should be performed on the four means (see 3.3.1). In addition, should any of the *codecs x talkers*, *codecs x levels*, or *codecs x levels x talkers* interactions be significant, a Newman-Keuls comparison of means should be computed on the means for that interaction. Follow the procedure for Newman-Keuls comparison of means using the appropriate mean square for the error term and the number of degrees of freedom. A similar analysis should be performed with channel conditions substituted for levels in Experiment II and background noise conditions substituted for levels in Experiment III.

Computational procedure:

Specifically, the following discussion will work through the analysis for the *codecs x levels* interaction in subjective Experiment I. The other interaction means are tested in a similar fashion. A sample spreadsheet calculator (incomplete with respect to the analysis of the experiments in this test plan) is presented in the `/tools/spread` directory. Once configured to the design of the individual experiments, the spreadsheet may be used to automatically perform the needed Newman-Keuls analyses for the *codecs x talkers*, *codecs x levels*, or *codecs x levels x talkers* interactions. The following discussion will provide an outline of an appropriate manual analysis procedure:

1. Obtain the mean for each combination of codec pair and level. There will be twelve such means (see Table 2.2.4.2.11-1). These means will be available automatically if the SAS program provided (or an equivalent procedure) is used to analyze the data [12].

1

Table 2.2.4.2.11-1. Means for Codecs x Levels (Encoder/Decoder)

<i>Codecs x Levels</i>	Encoder/Decoder
-9 dBm0	master/master master/test test/master test/test
-19 dBm0	master/master master/test test/master test/test
-29 dBm0	master/master master/test test/master test/test

2

3

4

5

6

7

8

9

10

2. Tabulate the means in ascending order. Among the twelve *codecs x levels* means, there are nine comparisons of interest. Using Table 2.2.4.2.11-2 as a guide, determine which pairs need to be compared in order to complete the test. For each comparison, determine *r* by counting the number of means spanned by the two means to be compared, including the to-be-compared means. For example, if the mean for -9 dBm0 master/master is immediately followed in the ordered means by the mean for -9 dBm0 test/master codec combination, the *r* for this comparison is 2.

Table 2.2.4.2.11-2. Comparisons of Interest Among Codecs x Levels Means

<i>Codecs x Levels</i>	Comparisons of Interest
-9 dBm0	master/master/ master/test master/master/ test/master master/master/ test/test
-19 dBm0	master/master/ master/test master/master/ test/master master/master/ test/test
-29 dBm0	master/master/ master/test master/master/ test/master master/master/ test/test

11

12

13

14

3. Read the mean square for the error term used in the calculation of the *F*-ratio for the *codecs x levels* effect. This will be the mean square (MS) for the error term on the fourth row from the bottom in the ANOVA table given in Table 2.2.4.2.11-4.

- 1 4. Obtain an estimate of the standard error of the mean for the interaction cells by dividing the MS above
 2 by the number of scores on which it is based: in this case, each mean is based on 768 judgments.
 3 Taking the square root of this quotient yields the standard error of the mean.
- 4 5. Determine the *df* for the MS used. For 48 subjects, the *df* for the *codecs x levels* comparisons is 282
 5 (for *codecs* it is 141, for *codecs x talkers* it is 2115, and for *codecs x levels x talkers* it is 4230).
- 6 6. Refer to a table of significant Studentized ranges to obtain the coefficients for the critical differences
 7 [11, 23]. Since no value of *df*=282 appears in the table in the reference, other tables or methods [18]
 8 should be used where values for *df* greater than 120 are required. For the *codecs x levels*, there are 12
 9 means, thus critical differences for *r*=2 up to *r*=12 will be needed (for *codecs* it is 4, for *codecs x talkers*
 10 it is 64 and for *codecs x levels x talkers* it is 192). Choose the values corresponding to *a*=0.01 because
 11 this is the significance criterion chosen for this test.
- 12 7. Construct a table of critical values as shown in Table 2.2.4.2.11-3.

13
 14 **Table 2.2.4.2.11-3. Critical Values for Codecs x Levels Means**

<i>r</i>	<i>q_r</i>	<i>critical values</i>
2	3.70	(These values will be the value in <i>q_r</i> column times the standard error calculated in step 4.)
3	4.20	
4	4.50	
5	4.71	
6	4.87	
7	5.01	
8	5.12	
9	5.21	
10	5.30	
11	5.38	
12	5.44	

- 15
- 16 8. Compute the difference between each pair of means to be tested. Compare this to the critical value
 17 obtained in step 7 for the *r* corresponding to the span across that pair of means. If the actual
 18 difference between the two means exceeds the critical difference, the means are statistically different at
 19 the 0.01 level.

Table 2.2.4.2.11-4. Analysis of Variance Table

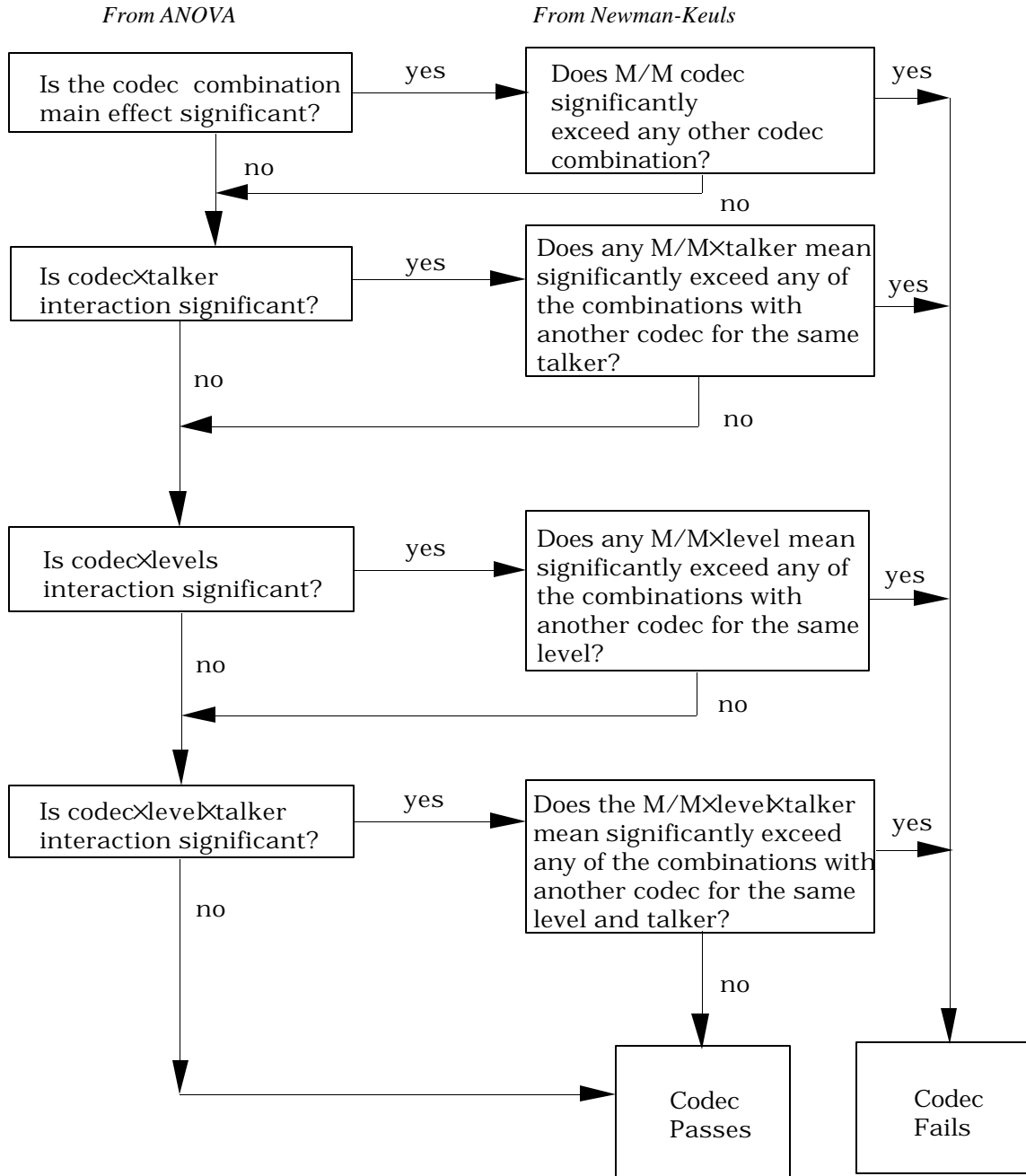
Effects	df	SS	MS	F	p
Total	9215				
Codec Combination	3				
Input Level	2				
Talker	15				
Codecs x Level	6				
Codecs x Talker	45				
Level x Talker	30				
Codecs x Level x Talker	90				
Subjects	47				
Error for Codec Combination: Codecs x Subjects	141				
Error for Level: Level x Subjects	94				
Error for Talker: Talker x Subjects	705				
Error for Codecs x Level: Codecs x Level x Subjects	282				
Error for Codecs x Talker: Codecs x Talker x Subjects	2115				
Error for Level x Talker: Level \times Talker x Subjects	1410				
Error for Codecs x Level x Talker: Codecs \times Level x Talker x Subjects	4230				

Degrees of freedom are those for 48 subjects.

2.2.4.3 Minimum Standard

Figure 2.2.4.3-1 displays a flowchart for the criteria of the subjective performance part of the combined objective/subjective test. The codec under test demonstrates minimum performance for this part if subjective testing establishes no statistically significant difference between it and the master codec. That is, if neither the main effect for codec combinations (*codecs*) and none of the *codecs x talkers*, *codecs x levels*, or *codecs x levels x talkers* interactions are significant, the codec under test has equivalent performance to the master codec within the resolution of the subjective test design.

1
2



3
4
5

Figure 2.2.4.3-1. Flow Chart of Codec Testing Process

1 If the main effect for codec combination is significant, then the differences in the *codecs* means shall be examined
2 for significance using the Newman-Keuls computational procedure of 3.3.1.2. There should be no case where the
3 mean rating on the master/master codec combination significantly exceeds the mean rating for any codec
4 combination, which includes the test codec. No significant differences between the master/master and the other
5 codec combinations indicate equivalent performance for the master codec and the test codec within the
6 resolution of the validation test design.

7
8 If the interaction between codec combination and talker is significant, the *codecs x talkers* means shall be
9 examined for significant differences using the Newman-Keuls computational procedure of 3.3.1.2. There should
10 be no case where the mean rating on the master/master codec combination significantly exceeds the mean rating
11 for any codec combination, which includes the test codec. No significant differences between the master/master
12 and the other codec combinations indicate equivalent performance for the master codec and the test codec
13 within the resolution of the validation test design.

14
15 If the interaction between codec combination and level is significant, the *codecs x levels* means shall be
16 examined for significant differences using the Newman-Keuls computational procedure of 3.3.1.2. There should
17 be no case where the mean rating on the master/master codec combination significantly exceeds the mean rating
18 for any codec combination, which includes the test codec. No significant differences between the master/master
19 and the other codec combinations indicate equivalent performance for the master codec and the test codec
20 within the resolution of the validation test design.

21
22 If the interaction between codec combination, level, and talker is significant, the *codecs x levels x talkers* means
23 shall be examined for significant differences using the Newman-Keuls computational procedure of 3.3.1.2. There
24 should be no case where the mean rating on the master/master codec combination significantly exceeds the mean
25 rating for any codec combination, which includes the test codec. No significant differences, between the
26 master/master and the other codec combinations, indicate equivalent performance for the master codec and the
27 test codec within the resolution of the validation test design.

28 If the above sequence of statistical tests indicates equivalent performance for the master codec and the test
29 codec, then the test codec meets the minimum performance standard of the subjective performance part of the
30 combined objective/subjective test.

31 The analysis of results for subjective Experiments II and III is analogous to the procedure described above for
32 Experiment I. Channel conditions in Experiment II and background noise conditions in Experiment III are
33 substituted for levels in the analysis of Experiment I. In Experiment II, there are two channel conditions and six
34 talkers with respective degrees of freedom of one and five for these factors. In Experiment III, there are nine
35 background noise conditions and four talkers with respective degrees of freedom of eight and three for these
36 factors.

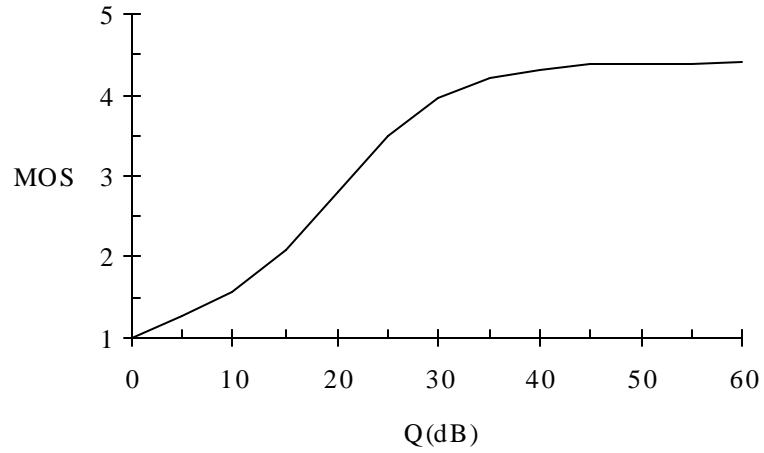
37 2.2.4.4 Expected Results for Reference Conditions

38 The MNRU conditions have been included to provide a frame of reference for the MOS test. Also, they provide
39 anchor conditions for comparing results between test laboratories. In listening evaluations where test
40 conditions span approximately the same range of quality, the MOS results for similar conditions should be

1 approximately the same. Generalizations may be made concerning the expected MOS results for the MNRU
2 reference conditions (see Figure 2.2.4.4-1).

3 MOS scores obtained for the MNRU conditions in any Service Option 17 validation test should be compared to
4 those shown in the graph below. Inconsistencies beyond a small shift in the means in either direction or a slight
5 stretching or compression of the scale near the extremes may imply a problem in the execution of the evaluation
6 test. In particular, MOS should be monotonic with MNRU, within the limits of statistical resolution; and the
7 contour of the relation should show a similar slope.

8



9

Figure 2.2.4.4-1. MOS versus MNRU Q

1 No text.

2

3

3 CODEC STANDARD TEST CONDITIONS

This section describes the conditions, equipment, and the software tools necessary for the performance of the tests of Section 2. The software tools and the speech database associated with 3.3, 3.4, and 3.5 are available on a CD-ROM distributed with this document.

3.1 Standard Equipment

The objective and subjective testing requires that speech data files be input to the speech encoder and that the output data stream be saved to a set of files. It is also necessary to input data stream files into the speech decoder and have the output speech data saved to a set of files. This process suggests the use of a computer based data acquisition system to interface to the codec under test. Since the hardware realizations of the speech codec may be quite varied, it is not desirable to precisely define a set of hardware interfaces between such a data acquisition system and the codec. Instead, only a functional description of these interfaces will be defined.

3.1.1 Basic Equipment

A host computer system is necessary to handle the data files that shall be input to the speech encoder and decoder, and to save the resulting output data to files. These data files will contain either sampled speech data or speech codec parameters. Hence, all the interfaces are digital. This is shown in Figure 3.1.1-1.

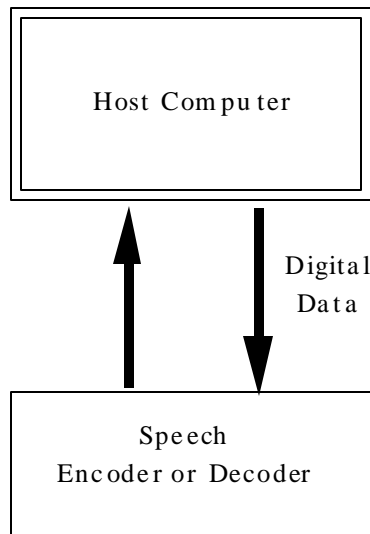


Figure 3.1.1-1. Basic Test Equipment

The host computer has access to the data files needed for testing. For encoder testing, the host computer has the source speech data files, which it outputs to the speech encoder. The host computer simultaneously saves the speech parameter output data from the encoder. Similarly, for decoder testing, the host computer outputs speech parameters from a disk file and saves the decoder output speech data to a file.

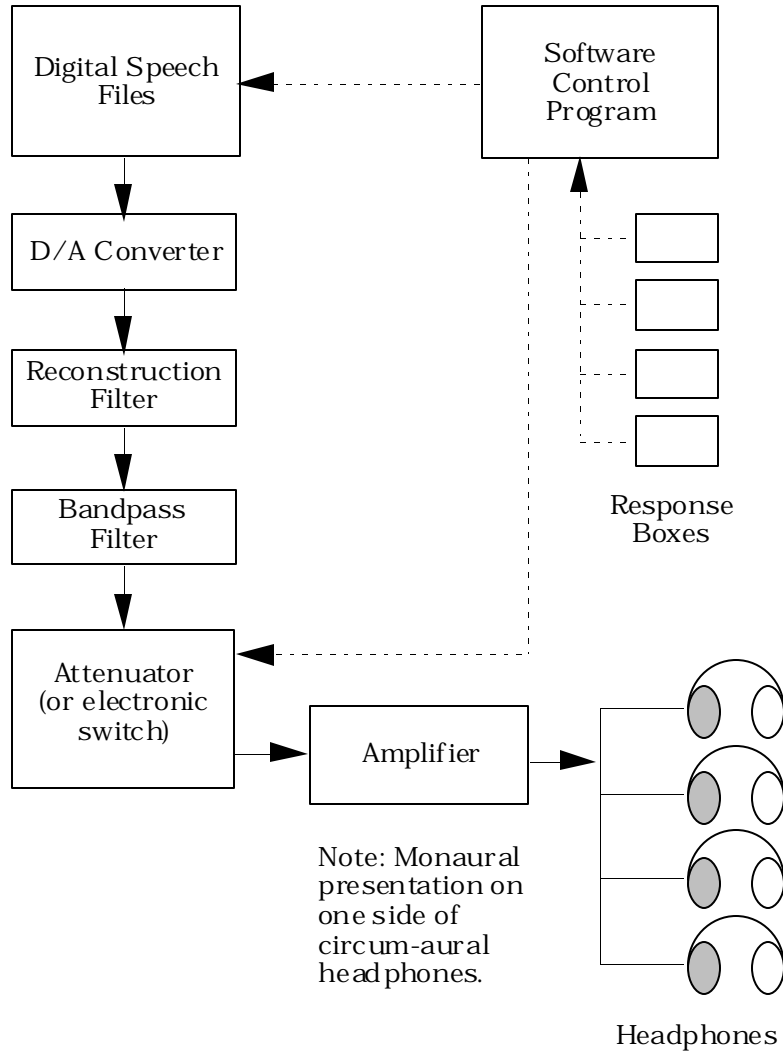
The choice of the host computer and the nature of the interfaces between the host computer and the speech codec are not subject to standardization. It is expected that the host computer would be some type of personal computer or workstation with suitable interfaces and adequate disk storage. The interfaces may be serial or

1 parallel and will be determined by the interfaces available on the particular hardware realization of the speech
 2 codec.

3 3.1.2 Subjective Test Equipment

4 Figure 3.1.2-1 shows the audio path for the subjective test using four listeners per session. The audio path is
 5 shown as a solid line; the data paths for experimental control are shown as broken lines. This figure is for
 6 explanatory purposes and does not prescribe a specific implementation.

7



8

9

Figure 3.1.2-1. Subjective Testing Equipment Configuration

3.1.2.1 Audio Path

The audio path shall meet the following requirements for electro-acoustic performance measured between the output of the D/A converter and the output of the headphone:

1. Frequency response shall be flat to within ± 2 dB between 200 Hz to 3400 Hz, and below 200 Hz the response shall roll off at a minimum of 12 dB per octave. Equalization may be used in the audio path to achieve this. A suitable reconstruction filter shall be used for playback.
2. Total harmonic distortion of less than 1% for signals between 100 Hz and 4,000 Hz.
3. Noise over the audio path of less than 30 dBA measured at the ear reference plane of the headphone.
4. Signal shall be delivered to the headphone on the listener's preferred telephone ear. No signal shall be delivered to the other headphone.

3.1.2.2 Calibration

The audio circuit shall deliver an average sound level of the stimuli to the subject at -16 dBPa (78 dB SPL) at the ear reference plane. This level was chosen because it is equivalent to the level delivered by a nominal ROLR handset driven by the average signal level on the PSTN network. This level may be calibrated using a suitable artificial ear with circum-aural headphone adapter and microphone. A test file with a reference signal is included with the source speech database for the purpose of calibration (see 3.5.3.) The audio circuit shall be calibrated so that the test signal has a level of -16 dBPa at the ear reference plane, while maintaining compliance with 3.1.2.1.

3.2 Standard Environmental Test Conditions

For the purposes of this standard, speech codecs under test are not required to provide performance across ranges of temperature, humidity or other typical physical environmental variables.

This standard provides testing procedures suitable for the assessment of speech codecs implemented in software. Where assumptions about software operating environments have been made in the construction of tools for testing, these have been based upon the UNIX™ environment.

3.3 Standard Software Test Tools

This section describes a set of software tools useful for performing the tests specified in Section 2. Where possible, code is written in ANSI C [13] and assumes the UNIX™ environment.

3.3.1 Analysis of Subjective Performance Test Results

3.3.1.1 Analysis of Variance

Section 2.2.4.2.11 of this IS-125-A test defines what is essentially a *mixed model*, which considers codecs, levels and talkers as *fixed* factors, while listeners are considered as a *random* factor. Therefore, the appropriate ANOVA will be a 4-way ANOVA with one random factor. This mixed model may be defined as follows:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + d_l + (\alpha d)_{il} + (\beta d)_{jl} + (\gamma d)_{kl} + (\alpha\beta d)_{ijl} + (\alpha\gamma d)_{ikl} + (\beta\gamma d)_{jkl} + (\alpha\beta\gamma d)_{ijkl} + e_{ijkl}$$

1 where μ is the main response, α_i is the contribution of the i -th codec-combination ($i=1..4$), β_j is the input level
2 effect ($j=1..3$), γ_k represents the contribution of the talker effect ($k=1..16$), and d_l represents the contribution of
3 the listener effect ($l=1..64$). The terms in parenthesis represent the interactions between the labeled effects. For
4 example, $(\alpha d)_{il}$ represents the interaction between codec combination i and listener l . Finally, the term e_{ijkl} is the
5 model (or experiment) error.

6 For the purpose of the IS-125-A Test, it is important to verify, at a 99% confidence level ($\alpha=1\%$), whether there is:

- 7 • a significant codec combination effect
- 8 • a significant interaction between codec and talker
- 9 • a significant interaction between codec and levels
- 10 • a significant interaction among codec, level, and talker

11 Should any of these items be significant at the 99% confidence level, this does not mean that the codec fails the
12 test, but only that different qualities were perceived. In other words, there is a chance that the test codec has
13 better performance than the reference codec, and in this case the test codec would pass the Conformance Test.

14 A sample SAS program to perform this computation is given in Table 3.3.1.1-1 [12]. This program also computes
15 the required cell means of the four terms: *codecs x levels*, *codecs x talkers*, *levels x talkers*, and *codecs x levels x*
16 *talkers*. The program will compute the Newman-Keuls comparison for the *codecs* main effect. The SAS program
17 does not compute the Newman-Keuls comparisons for the interaction effects.

18

Table 3.3.1.1-1. Analysis of Variance

```

1
2  data codectst;      (defines input data file)
3
4      infile logical;
5      drop i t;
6      input / group 17-18;
7      input ///// ///// ///// //;          (skips practice trials)
8      do t = 1 to 224;                        (reads data. This statement
9 shall                                         be modified to
10 read data
11 according to the input record
12         format)
13 input trial block cond level talker;
14 do i = 1 to 3;                              (assumes three ratings per line
15                                         in file)
16     subj = 3*(group-1) + i;
17     input rating @;
18     output;
19 end;
20 end;
21
22 proc print;
23
24     title "Data for validation test of TEST codec";
25
26 proc sort;
27
28     by cond level talker;
29
30 proc means mean std maxdec=2; (computes a table of means and
31                               standard deviations)
32     by cond level talker;
33     var rating;
34
35 data anova;          (selects data for codec combinations,
36                       leaving out MNRU conditions)
37     set codectst;
38     drop cond;
39     if cond<5;
40     codecs=cond;
41
42 proc anova;
43
44     class codecs level talker subj;
45     model rating=codecs|level|talker|subj;
46     test h=codecs e=codecs*subj;
47     test h=level e=level*subj;
48     test h=talker e=talker*subj;
49     test h=codecs*level e=codecs*level*subj;
50     test.h=codecs*talker e=codecs*talker*subj;
51     test h=level*talker e=level*talker*subj;
52     test h=codecs*level*talker e=codecs*level*talker*subj;

```

```

1      means codecs /snk e=codecsxsubj;
2      means level codecsxlevel levelxtalker codecsxlevelxtalker;
3      means talker /snk e=talkerxsubj;
4      means codecsxtalker;
5      title "ANOVA to test for main effects and comparisons of means";
6

```

7 3.3.1.2 Newman-Keuls Comparison of Means

8 Following the completion of the ANOVA, comparisons of cell means for certain interactions may be necessary.
9 A sample MicroSoft® Excel spreadsheet calculator (incomplete with respect to the analysis of the experiments in
10 this test plan) is presented in the `/tools/spread` directory (with instructions for use of this spreadsheet
11 supplied in the `Read_me.spd` file of the same directory) which, once configured to the design of the individual
12 experiments, may be used to automatically perform the needed Newman-Keuls analyses for the *codecs x talkers*,
13 *codecs x levels*, or *codecs x levels x talkers* interactions. The following discussion will provide an outline of an
14 appropriate manual analysis procedure:

15 In order to verify whether the significant effects identified by the ANOVA represent a passage or failure of the
16 IS-125-A test, multiple comparison methods are applied. In the IS-125-A test, the performances of codec
17 combinations involving the test encoder and/or the test decoder are compared to the performance of the
18 reference encoder/reference decoder combination. In the case of the IS-125-A test, however, the post-hoc
19 Newman-Keuls multiple comparison method is to be used, with a confidence level of 99%. The Newman-Keuls
20 method is considered to have an adequate Type-1 error rate, given the confidence level, and is based on the
21 Studentized range defined by:

$$q = \frac{\mu_m - \mu_t}{S_{\bar{\mu}}}$$

22
23 where μ_m represents the average score for the reference encoder and decoder combination, μ_t represents the
24 average score for a given combination which involves either the test encoder or test decoder ($\mu_m > \mu_t$) and $S_{\bar{\mu}}$ is
25 the pooled standard error defined by

$$S_{\bar{\mu}} = \sqrt{\frac{MS_{error}}{n}}$$

26
27 where n is the number of samples used to calculate μ_m and μ_t , and MS_{error} is the denominator of the F-test in
28 the ANOVA for the particular factor or interaction under consideration.

29 Once the q value above is computed, it is compared to the critical Studentized range value $q_{\alpha, v, r}$ for the
30 confidence level α (99%), v degrees of freedom, and a range of r means. For the IS-125-A test, $\alpha=99\%$, while v
31 and r will depend on the factor or interaction under consideration. Alternatively, a critical confidence interval
32 may be computed by multiplying $q_{\alpha, v, r}$ by the pooled standard error, $S_{\bar{\mu}}$. Since the test failure criterion is that
33 none of the test encoder/decoder combinations should be inferior in performance to the reference
34 encoder/decoder combination, the failure criterion may easily be expressed as:

$$\mu_m - \mu_t > q_{\alpha, v, r} \times S_{\bar{\mu}}$$

35

1 3.3.2 Segmental Signal-to-Noise Ratio Program - snr . c

2 This section describes a procedure for computing overall signal-to-noise ratios and signal-to-noise ratios per
 3 segment, SNR_i . It also describes the use of the accompanying C program, `snr . c`, that implements this
 4 procedure. The program needs as input, two binary data files containing 16-bit sampled data. The first data file
 5 contains the reference signal, the second data file contains the test signal. The output may be written to the
 6 terminal or to a file.

7 3.3.2.1 Algorithm

8 Let $x_i(n)$ and $y_i(n)$ be the samples of a reference signal and test signal, respectively, for the i^{th} 5 ms time segment
 9 with $1 = n = 40$ samples. Then, we can define the reference signal power as

$$10 \quad Px_i = \frac{1}{K} \sum_{n=1}^K x_i(n)^2 = \frac{Ex_i}{K}$$

11 and the power of the test signal as

$$12 \quad Py_i = \frac{1}{K} \sum_{n=1}^K y_i(n)^2 = \frac{Ey_i}{K}$$

13 and the power of the error signal between $x(n)$ and $y(n)$ as

$$14 \quad R_i = \frac{1}{K} \sum_{n=1}^K [x_i(n) - y_i(n)]^2 = \frac{N_i}{K}$$

15 where the summations are taken over $K=40$ samples, Ex_i is the energy in the i^{th} reference signal segment, Ey_i is
 16 the energy in the i^{th} test signal segment, and N_i is the noise energy in the i^{th} segment. The signal-to-noise ratio
 17 in the i^{th} segment (SNR_i) is then defined as

$$18 \quad SNR_i = \begin{cases} 10 \log_{10} \left(\frac{Px_i}{R_i} \right) & \text{for } Px_i \text{ or } Py_i = T \\ 0 & \text{for } Px_i \text{ and } Py_i < T \end{cases}$$

19 It is possible to compute an average signal-to-noise ratio, $\langle SNR \rangle$, as

$$20 \quad \langle SNR \rangle = 10 \log_{10} \left(\frac{\sum_i Ex_i}{\sum_i N_i} \right)$$

1 where the sums run over all segments. If SNR_i corresponds to the signal-to-noise ratio in decibels for a segment
 2 i , SNRSEG is then defined as the average over all segments in which

$$3 \quad P_{x_i} \text{ or } P_{y_i} = T$$

4 For these segments, SNRSEG is defined as

$$5 \quad SNRSEG = \frac{1}{M} \sum_{i=1}^N SNR_i$$

6 where it is assumed that there are N segments in the file, and M is the number of segments in which P_{x_i} or $P_{y_i} \geq$
 7 T . A segment size of 5 ms is used, which corresponds to 40 samples at an 8 kHz sampling rate.

8 Silence intervals are eliminated by employing a threshold T . If either power, P_{x_i} or P_{y_i} , for segment i exceed a
 9 threshold T , then this segment will be included in the computation of SNRSEG. The threshold T is chosen
 10 conservatively to be a factor of 512^2 below the maximum value.

11 The samples are 16-bit with a dynamic range of $[-32768, 32767]$, and T thus equals $(32768/512)^2 = 4096$. The
 12 energy threshold will be $K=40$ times larger than this. Furthermore, to prevent any one segment from dominating
 13 the average, the value of SNR_i is limited to a range of -5 to +80 dB.

14 3.3.2.2 Usage

15 The accompanying software distribution contains the program `snr.c`. The program is written in ANSI standard
 16 C and assumes the UNIX™ environment.

17 The input consists of the reference file containing the samples $x(n)$ and the test file containing the samples $y(n)$.
 18 The relative offset between the samples x and y may also be specified in samples. If the samples y are delayed,
 19 the offset is considered to be positive. Summary output is written to the standard output. A binary output file,
 20 which contains the SNR_i value for each segment, is written for use by the `cdf.c` program. Silent segments
 21 will have an SNR_i value of 0.0.

22 Table 3.3.2.2-1 is an example of the output of the program.

23

1

Table 3.3.2.2-1. Output Text from snr.c

Reference file	r.dat
Test file	t0.dat
Relative offset	0
Average SNR	14.94 dB
SNRSEG	13.10 dB
Maximum value	22.12 dB
Minimum value	-5.00 dB
Frame size	40 samples
Silence threshold	64.00
Total number of frames	512
Number of silence frames	46

2

3 3.3.3 Cumulative Distribution Function - `cdf.c`

4 The computation of a cumulative distribution function of signal-to-noise ratio values per segment, SNR_i , is
 5 required for objective testing. The program `cdf.c` performs this computation. The program is written in ANSI
 6 standard C and assumes the UNIX™ environment.

7 3.3.3.1 Algorithm

8 The program implements the pseudo-code given in 2.1.2.

9 3.3.3.2 Usage

10 The input consists of a binary file of sample SNR_i output, such as created by `snr.c` (see 3.3.3). The samples
 11 are read as floating point `double` values. The output is an ASCII formatted file of cumulative distribution
 12 function data for the range of values from less than 2 dB to 64 dB or greater. The first value, `f [1]`, includes all
 13 occurrences less than 2 dB; the last value, `f [64]`, includes all occurrences for 64 dB or greater. The last line of
 14 output gives the SNRSEG value. The program prompts for the names of the input (binary) file and the output
 15 (text) file.

16 3.3.4 Randomization Routine for Service Option 1 Validation Test - `random.c`

17 This routine creates a randomized array of integers corresponding to 18 test conditions and 16 talkers (see
 18 2.2.4.2.1). The randomization conforms to the required randomized block design in 2.2.4.2.7. Source code for the
 19 randomization is `random.c` in the directory `/tools` of the companion software. Note that the constants
 20 `BLOCKSIZE` and `NBLOCKS` have been defined in the source code to be 18 and 16, respectively, for use with
 21 subjective Experiment I. In order to use this randomization routine for subjective Experiments II and III, these
 22 constants need to be changed appropriately. For Experiment II, `BLOCKSIZE` equals 14 and `NBLOCKS` equals 6.
 23 For Experiment III, `BLOCKSIZE` equals 24 and `NBLOCKS` equals 4.

1 3.3.4.1 Algorithm

2 The program accomplishes the randomization in the following steps:

- 3 1. Creating an array of integers, `talker[i][j]`, corresponding to the talker used for the j^{th}
4 occurrence of test condition i .
- 5 2. Randomizing the test conditions in the j^{th} block.
- 6 3. Retrieving the talker number for that test condition and block from `talker[i][j]`, and storing the
7 test condition and talker together in array `trials[i][j][1]` (test condition) and
8 `trials[i][j][2]` (talker). The content of `trials[i][j][1]` gives the test condition for the
9 i^{th} trial of the j^{th} block of the session.
- 10 4. Checking for consecutive occurrences of the same talker. If two consecutive trials have the same
11 talker, the second is swapped with the trial following. If three consecutive trials have the same talker,
12 or the trials in question occur at the end of a block (in which case swapping would violate the
13 constraints of the randomized block design), the order of test conditions for that block is recalculated.

14 Blocks are flagged with `pass[i] = 1` if the randomization for that block meets the constraints. The routine
15 continues until `pass[i] = 1` for $i = 0$ to `NBLOCKS - 1`.

16 Any arbitrary assignment of test condition numbers and talker numbers to the integers output by the
17 randomization program will suffice. One simple correspondence is for the test condition numbers given in the
18 standard to equal the test condition numbers output in the randomization, and the talker numbers to correspond
19 in the following way:

20 Male talkers 1 - 8 ==> talkers 1 - 8.

21 Female talkers 1 - 8 ==> talkers 9 - 16.

22 The output of this routine or a similar custom routine should be used to determine the order of trials in the
23 validation test.

24 3.3.4.2 Usage

25 The program iterates until the array `trials[i][j][1]` meets the randomization conditions. It is output in
26 the file `rand_out.rpt`.

27 3.3.5 Rate Determination Comparison Routine - `rate_chk.c`

28 This utility program is used to check the rate decision algorithm of the test codec to ensure that the test codec
29 complies with the minimum performance specification. This program operates on two input packet files, whose
30 format is specified by the format of the master codec output packet file (see 3.4.4). The first input file is the
31 packet file produced by the master codec, and the second input file is the packet file produced by the test codec.
32 Both input files are parsed for the rate byte of each packet, and then they are compared to produce joint rate
33 histogram statistics of the test and master codec. These statistics are recorded in a log file of the same name as
34 the first input file with an extension of “.log”. The log file contains the individual rate statistics of both input

1 files, the joint rate statistics, and the packets of each input file where the frame rates differed. This utility
 2 program may be found in the `/tools` directory of the companion software.

3 3.3.6 Software Tools for Generating Speech Files - `stat.c` and `scale.c`

4 The first program, `stat.c`, is used to obtain the maximum, root mean square (rms), and average (dc) values of a
 5 binary speech file. Given that these values are known, the second program, `scale.c`, may be used to generate
 6 an output file whose values differ from an input binary file by a scale factor. Both programs ignore any trailing
 7 samples that comprise less than a 256 byte block. Source code for both files is located in the directory `/tools`
 8 of the companion software.

9 3.4 Master Codec - `qcelp_a.c`

10 This section describes the C simulation of the speech codec specified by the IS-96-C standard [1]. The C
 11 simulation is used to produce all of the master codec speech files used in the objective and subjective testing for
 12 the three experiments.

13 3.4.1 Master Codec Program Files

14 This section describes the C program files which are provided in the directory `/master` in the companion
 15 software.

16	<code>qcelp_a.c</code>	This file contains the <code>main</code> program function and subroutines which parse the command line and describe the proper usage of the master codec.
17		
18	<code>cb.c</code>	This file contains the search subroutine used in the encoder to determine the code-book parameters, <code>i</code> and <code>G</code> , for each code-book sub-frame.
19		
20	<code>decode.c</code>	This file contains the main decoder subroutine and the <code>run_decoder</code> subroutine which is used by both the encoder and decoder to update the filter memories after every sub-frame.
21		
22		
23	<code>encode.c</code>	This file contains the main encoder subroutine.
24	<code>filter.c</code>	This file contains various digital filtering subroutines called throughout the encoding and decoding procedures.
25		
26	<code>init.c</code>	This file contains the subroutine which initializes the codec parameters at the start of execution.
27		
28	<code>io.c</code>	This file contains the input and output subroutines for speech files, packet files, and signaling files.
29		
30	<code>lpc.c</code>	This file contains the subroutines which determine the linear predictive coefficients from the input speech.
31		
32	<code>lsp.c</code>	This file contains the subroutines which convert the linear predictive coefficients to line spectral pair frequencies and vice versa.
33		
34	<code>pack.c</code>	This file contains the packing and unpacking subroutines and the error detection and correction subroutines.
35		

1	<code>pitch.c</code>	This file contains the search subroutines used in the encoder to determine the pitch parameters, b and L, for each pitch sub-frame.
2		
3	<code>prepro.c</code>	This file contains the subroutine which removes the DC component from the input speech.
4		
5	<code>quantize.c</code>	This file contains the various quantization and inverse-quantization routines used by many subroutines throughout the encoding and decoding procedures.
6		
7	<code>ratedec.c</code>	This file contains the subroutine which determines the data rate to be used in each frame of speech using the adaptive rate decision algorithm described in IS-96-C.
8		
9	<code>target.c</code>	This file computes and updates the “target signal” used in the pitch and code-book parameter search procedures.
10		
11	<code>coder.h</code>	This file contains the constant parameters used by the IS-96-C speech codec.
12	<code>defines.h</code>	This file contains various define statements used by subroutines throughout the master codec simulation.
13		
14	<code>filter.h</code>	This file contains the structures for the digital filters used by subroutines throughout the master codec simulation.
15		
16	<code>struct.h</code>	This file contains other structures used by subroutines throughout the master codec simulation.
17		

18

19 3.4.2 Compiling the Master Codec Simulation

20 The code is maintained using the standard UNIX™ `make` utility. Manual pages on `make` are available in most
 21 UNIX™ environments. Typing `make` in the directory containing the source code will compile and link the code
 22 and create the executable file called `qcelp_a`.

23 3.4.3 Running the Master Codec Simulation

24 The `qcelp_a` executable file uses command line arguments to receive all information regarding input and
 25 output files and various parameters used during execution. Executing `qcelp_a` with no command line
 26 arguments will display a brief description of the required and optional command line arguments. These options
 27 are described below:

28	<code>-i infn</code> (required)	Specifies the name of the input speech file, or the name of the input packet file if only decoding is being performed (see the <code>-d</code> option below).
29		
30	<code>-o outf</code> (required)	Specifies the name of the output speech file, or the name of the output packet file if only encoding is being performed (see the <code>-e</code> option below).
31		
32	<code>-d</code>	Instructs the simulation to perform only the decoding function. The input file shall contain packets of compressed data.
33		
34	<code>-e</code>	Instructs the simulation to perform only the encoding function. The output file will contain packets of compressed data.
35		

1 If neither the `-d` nor the `-e` option is invoked, the coder performs both the
2 encoding and decoding functions by default.

3 `-h max` Sets the maximum allowable data rate to `max`, using the codes specified in the
4 first column of Table 3.4.4-1.

5 `-l min` Sets the minimum allowable data rate to `min`, using the codes specified in the
6 first column of Table 3.4.4-1.

7 If neither the `-h` nor `-l` option is invoked, the coder allows the data rate to
8 vary between Rate 1 and Rate 1/8.

9 In addition, if `max != min`, the data rate varies between `max` and `min` using
10 the same rate decision algorithm, where the data rate is set to `max` if the selected
11 data rate is \geq `max`, and the data rate is set to `min` if the selected data rate is
12 \leq `min`. See the `select_rate()` routine in the file `ratedec.c` for more
13 information.

14 `-f N` Stop processing the speech after `N` frames of speech. The frame-size is equal to
15 160.

16 If the `-f` option is not invoked, the coder processes all of the speech until the
17 end of the input file is reached.

18 `-p flag` If `flag` is set to 0, the post-filter is disabled. If the `-p` option is not invoked,
19 the post-filter is enabled during decoding.

20 `-s sigfn` Read signaling data from the file `sigfn`. This option allows rates to be set by
21 the user in the file `sigfn` on a frame by frame basis. The file `sigfn` takes the
22 form

23 `frame_num_1 max_rate_1 min_rate_1`
24 `frame_num_N max_rate_N min_rate_N`
25 `...`

26 This causes frame number “`frame_num_1`” to have a maximum allowable rate of
27 “`max_rate_1`” and a minimum allowable rate of “`min_rate_1`,” where the first frame
28 is frame number 0, not frame number 1. The `frame_num` parameters shall be in
29 increasing order. For example, to process the input file `input.spch`, put the
30 output speech in `output.spch`, force the 10th frame to “blank” and the 15th
31 frame between Rate 1/2 and Rate 1/8, the following could be run.

32 `qcelp_a -i input.spch -o output.spch -s signal.data`
33 where the file `signal.data` contains

34 `9 0 0`
35 `14 3 1`

36 In this case, every frame except for frame numbers 9 and 14 will have the rate
37 selected using the normal IS-96-C rate decision algorithm.

If the `-s` option is not invoked, the rate for every frame will be selected using the normal IS-96-C rate decision algorithm.

3.4.4 File Formats

Files of speech contain 2's complement 16-bit samples with the least significant byte first. The `qcelp_a` executable processes these files as follows: The samples are first divided by 4.0 to create overall positive maximum values of $2^{13}-1$. After decoding, the samples are multiplied by 4.0 to restore the original level to the output files. The $2^{13}-1$ value is slightly above the maximum μ -law value assumed in the IS-96-C standard [1], but the simulation performs acceptably with this slightly larger dynamic range. The subroutines in the file `io.c` perform the necessary byte swapping and scaling functions.

The packet file contains twelve 16-bit words with the low byte ordered first followed by the high byte. The first word in the packet contains the data rate while the remaining 11 words contain the encoded speech data packed in accordance with the tables specified in 2.4.7 of IS-96-C. The packet file value for each data rate is shown in Table 3.4.4-1.

Table 3.4.4-1. Packet File Structure From Master Codec

Value in Packet File	Rate	Data Bits per Frame
4 = 0x0004	1	171
3 = 0x0003	1/2	80
2 = 0x0002	1/4	40
1 = 0x0001	1/8	16
0 = 0x0000	Blank	0
15 = 0x000f	Full Rate Probable	171
14 = 0x000e	Erasure	0

Unused bits are set to 0. For example, in a Rate 1/8 frame, the packet file will contain the word 0x0100 (byte-swapped 0x0001) followed by one 16-bit word containing the 16 data bits for the frame (in byte-swapped form), followed by ten 16-bit words containing all zero bits.

3.4.5 Verifying Proper Operation of the Master Codec

Three files, `m01v0000.raw`, `m01v0000.pkt`, and `m01v0000.mstr.raw`, are included with the master codec software to provide a means for verifying proper operation of the software. The file `m01v0000.raw` is an unprocessed speech file. The file `m01v0000.pkt` is a packet file that was obtained by running

```
qcelp_a -i m01v0000.raw -o m01v0000.pkt -e
```

The file `m01v0000.mstr.raw` is a decoded speech file that was obtained by running

1 `qcelp_a -i m01v0000.pkt -o m01v0000.mstr.raw -d`

2 Once `qcelp_a.c` is compiled, verification files should be processed as follows:

3 `qcelp_a -i m01v0000.raw -o verify.pkt -e`

4 `qcelp_a -i m01v0000.pkt -o verify.mstr.raw -d`

5 The output files `m01v0000.pkt` and `m01v0000.mstr.raw` should exactly match `verify.pkt` and

6 `verify.mstr.raw`, respectively.

3.5 Standard Source Speech Files

The database used for testing consists of Harvard sentence pairs [24]. There were 16 talkers used: seven adult males, seven adult females, one male child, and one female child. The talkers were native speakers of North American English. The child talkers were less than 10 years of age.

3.5.1 Processing of Source Material

Recordings were made with a studio linear microphone with a flat frequency response. The analog signal was sampled at 16 kHz. The individual sentences of each sentence pair were separated by an approximate 750 ms gap. Recording was done in a sound-treated room, with an ambient noise level of not more than 35 dBA. Care was exercised to preclude clipping.

The source material used in this standard has been provided by Nortel Networks. Source material was prepared for the three experiments according to the following procedures. The decimation and Intermediate Reference System Specification filters are specified below. All measurement and level normalization procedures were done in accordance with ITU-T Recommendation P.56 [18]. The software program `sv56demo` that is included in ITU-T Software Tools G.191 [22] was used for this purpose.

3.5.1.1 Experiment I - Level Conditions

1. The sampling was decimated from 16 kHz to 8 kHz.
2. The level was normalized to -19 dBm0 (3.17 dBm0 is a full-scale sine wave).
3. Each speech sample is represented in 16-bit linear format.

3.5.1.2 Experiment II - Channel Conditions

1. The sampling was decimated from 16 kHz to 8 kHz.
2. Frequency shaping was performed in accordance with the IRS filter mask defined in [8].
3. The level was normalized to -19 dBm0 (3.17 dBm0 is a full-scale sine wave).
4. Each speech sample is represented in 16-bit linear format.

3.5.1.3 Experiment III - Background Noise Conditions

1. The sampling was decimated from 16 kHz to 8 kHz.
2. The level was normalized to -9 dBm0, -19 dBm0, and -29 dBm0.
3. Each type of noise was added to files with each input level as follows: car noise was added at SNR of 15 dB; street noise was added at SNR of 20 dB.
4. Each speech sample is represented in 16-bit linear format.

3.5.2 Data Format of the Source Material

The database for the first objective test consists of 3 files, each composed of 112 single sentences, which are derived from 16 talkers with seven sentences each. The total volume is approximately 16 Mbytes. Subjective Experiment I contains 288 sentence pairs, 16 talkers of 18 sentence pairs each, with a volume of approximately

1 28.8 Mbytes. The second objective test consists of 24 sentence pair packets from the master codec, six talkers of
 2 four sentence pairs each, with a volume of approximately 160 kbytes. Subjective Experiment II contains 36
 3 sentence pairs and 24-sentence-pair packets, six talkers of four sentence pairs each, with a volume of
 4 approximately 3.76 Mbytes. The third objective test requires three 84-second-long background noise impaired
 5 files with a volume of approximately 3.3 Mbytes. Subjective Experiment III contains 96 sentence pairs, four
 6 talkers of 24 sentence pairs each, with a volume of approximately 10 Mbytes. The total volume of the speech
 7 database for the three objective tests and subjective experiments is approximately 63 Mbytes.

8 The file names use the following conventions:

- 9 1. Each file name contains six characters with the extension .X19, .X09, .X29, .PKT, .Q10, .Q20, or
 10 .Q30. The extensions .X19, .X09, and .X29 indicate a -19, -9, or -29 dBm0 rms level, respectively.
 11 The extensions .Q10, .Q20, or .Q30 indicate an MNRU condition of 10, 20, or 30 dBQ. The
 12 extension .PKT indicates that the file is a packet file produced by the master codec. The filenames
 13 begin with a letter m or f, indicating the speech is from a male or female, respectively.
- 14 2. The second character of the filename is a digit that identifies the talker. The eight male and eight
 15 female talkers are simply identified by the digits from one through eight.
- 16 3. The third and fourth characters identify one of the nineteen utterances made by the identified talker.
- 17 4. The fifth character identifies the condition being imposed on the source or packet file. The value zero
 18 indicates no condition is imposed on the source, the value one indicates 1% FER channel impairments
 19 on the packet file, the value two indicates 3% FER channel impairments on the packet file, the value
 20 three indicates 15 dB SNR car noise impairing the source, and the value four indicates 20 dB SNR street
 21 noise impairing the source.
- 22 5. The sixth character identifies the source as being from Experiment I, II, or III of the objective and
 23 subjective tests using the values 1, 2, and 3, respectively.

24 For example, the file `m20301.x19` contains the third sentence from the second male talker for Experiment I at
 25 the -19 dBm0 level. The file `f11801.x09` contains the eighteenth sentence from female talker one for
 26 Experiment I at the -9 dBm0 level.

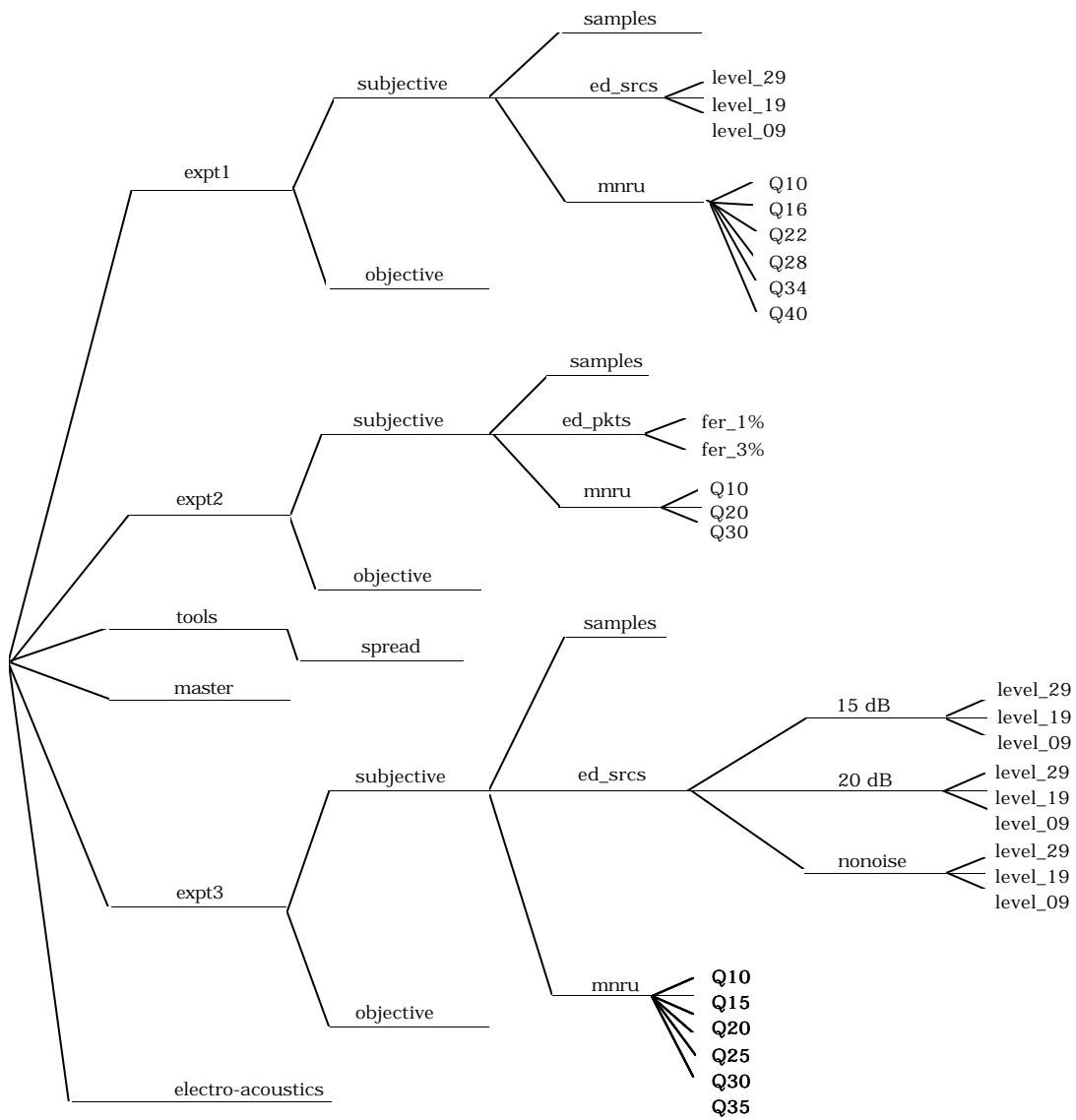
27 3.5.3 Calibration File

28 To achieve the calibration described in 3.1.2.2, a standard calibration file is provided with the source speech
 29 material. The file `1k_norm.src` is located in the directory `/tools` of the companion software. The calibration
 30 file comprises a -19.0 dBm0 1004 Hz reference signal. In addition, a 0.0 dBm0 1 kHz reference signal and a full
 31 scale 1 kHz reference signal are included in this directory and are named `1k_0dBm0.src` and `1k_full.src`
 32 respectively. All measurement and level normalization procedures done with respect to these calibration
 33 references should be in accordance with CCITT ITU-T Recommendation P.56 [18]. The software program
 34 `sv56demo` that is included in ITU-T Software Tools G.191 [22] should be used for this purpose. Acoustic output
 35 shall be adjusted to -16 dBPa (78 dB SPL) upon playback of the calibration file, as per 3.1.2.2.

36 3.6 Contents of Software Distribution

37 The contents of the software distribution are tree structured in order to facilitate the partitioning of the database
 38 required for the three objective and subjective minimum performance tests. The software distribution is provided
 39 on a CD-ROM. This software should be readable on any UNIX™ machine or other machine with suitable

1 translation software. The directories of the software distribution are represented in Figure 3.6-1. At the first
2 node of the tree the database is divided into the source material needed for the three experiments (/expt1,
3 /expt2, and /expt3), the software tools needed to analyze test results and prepare source material
4 (/tools), and the software implementation of the master codec (/master). The three experiments are further
5 divided into subjective and objective tests. The /objective directory contains the source material or packet
6 material (in the case of Experiment II) required to perform the objective component of the three experiments. The
7 /subjective directory is further broken down into the codecs and conditions of the test, the MNRU
8 samples, and the test samples required for each MOS test. The final leaf of the tree under the subjective
9 component of each of the three experiments contains the source speech files necessary for that part of the MOS
10 test. For instance, all the female and male sentence pairs necessary for testing the encoder/decoder codec
11 combinations at an input level of -29 dBm0 in subjective Experiment I may be found at
12 /expt1/subjective/ed_srcs/level_29.



1
2
3
4
5

Figure 3.6-1. Directory Structure of Companion Software