ENGLISH TRANSLATION

PERSONAL DIGITAL CELLULAR TELECOMMUNICATION SYSTEM

QUALITY RECOMMENDATION AND VALIDATION TEST FOR SPEECH CODEC

(STANDARD TECHNICAL CHARACTERISTICS AND VALIDATION TESTING METHODS RELATED TO SPEECH CODEC CONNECTIVITY AND SPEECH QUALITY)

ARIB TECHNICAL REPORT

ARIB TR-T1 Rev. 1.1

ISSUED ON DECEMBER 26, 1995
REVISION 1.1 J U L Y 25, 2000

Association of Radio Industries and Businesses

General Notes to the English translation of ARIB Standards and Technical Reports

1. The copyright of this document is ascribed to the Association of Radio Industries and Businesses (ARIB).

3. The ARIB Standards and ARIB Technical Reports are usually written in Japanese and approved by the ARIB Standard Assembly. This document is a translation into English of the approved document for the purpose of convenience of users. If there are any discrepancies in the content, expressions, etc., between the Japanese original and this translated document, the Japanese original shall prevail.

4. The establishment, revision and abolishment of ARIB Standards and Technical Reports are approved at the ARIB Standard Assembly, which meets several times a year. Approved ARIB Standards and Technical Reports, in their original language, are made publicly available in hard copy, CDs or through web posting, generally in about one month after the date of approval. The original document of this translation may have been further revised and therefore users are encouraged to check the latest version at an appropriate page under the following URL:

http://www.arib.or.jp/english/index.html

Preface

The Association of Radio Industries and Businesses (ARIB) has been investigating and summarizing the basic technical requirements for establishing standards for developing a digital mobile telephone system. These will appear in the form of standards or technical reports goveming the use of radio facilities and equipment for systems that transmit over radiowaves. Such standards are being developed based on the participation of and discussions with the various radio equipment manufactures, operators and users.

Technical reports such as this serve as guidelines for developing private standards for regulating measurement and testing methods for use of the pertinent radio equipment based on the publicly establish standards so as to ensure the necessary quality levels and compatibility of the radio equipment being developed.

This technical report specifies standard technical characteristics and validation testing methods related to speech codec connectivity and speech quality for the Personal Digital Cellular Telecommunication System. In order to ensure faimess, impartiality and openness among all parties involved, during the drafting states, we are inviting operators and users both domestically and overseas to participate openly in the activities of the Standard Assembly so as to develop standards based on the total agreement of all parties involved.

The scope of application of this technical report covers the basic items for ensuring the interconnectivity and speech quality of speech codecs used in components of the Personal Digital Cellular Telecommunication System.

We hope that this technical report will aid all parties involved, including radio equipment manufactures, telecommunications operators, equipment users.

Note: The original Japanese version of the technical report TR-T1 "Personal Digital Cellular Telecommunication System – Quality Recommendation and Validation Test for Speech Codec" was approved by the Third ARIB Standard Assembly on December 26, 1995. This document is the translation of TR-T1 into English, approved by the RCR STD-27 English Version Sub Working Group.

CONTENTS

CHAPTER 1 GENERAL

1.1 Overview

Speech codecs used in the Personal Digital Cellular Telecommunication System (hereinafter referred to as "speech codecs") shall be developed and designed according to the radio interface and speech encoding system specified in RCR STD-27 "Standard for Personal Digital Cellular Telecommunication System". The recommended quality validation tests for the speech codecs shall be limited to the basic functions and options specified in the Standard.

This technical report describes quality recommendations required for validation and testing methods.

1.2 Scope

Two different methods are used in the tests to validate the recommended quality of the speech codecs developed and designed in accordance with the Standard, one using objective evaluation (below referred to as the objective evaluation test including delay time evaluation) and the other using subjective evaluation (below referred to as the subjective evaluation test including delay time evaluation).

The objective evaluation tests, in which objective methods are used to test the degree of completion of the speech codec and to get a good idea of its characteristics, should preferably be carried out prior to the subjective evaluation. The subjective evaluation tests are used to make a pass/fail judgment on the speech codec characteristics.

As described in Chapter 6, two types of testing methods are available: "bit exactness verification test" (hereinafter referred to as "Bit Exact Verification Test) which includes delay time evaluation") and the "subjective evaluation test" (which not include delay time evaluation).

Purpose of bit exact verification test is to test the success of the implementation of speech codec by checking the output bitstream of the speech. This test shall be conducted prior to the subjective evaluation test.

Also, for ACELP speech codec, a subjective evaluation test, i.e., only the bit exactness validation test is available and the subjective evaluation test is not required.

1.3 Precautions

1) The tests specified in this technical report detail the minimum conditions which speech codecs must satisfy, including interconnectivity and speech quality under steady conditions with a master speech codec as a reference. The tests do not guarantee interconnectivity between speech codecs and their speech quality accepted in the tests under all operating conditions. Furthermore, the evaluation tests are conducted only with the speech codec part. The acoustic system is not tested. Therefore, evaluation of the audibility in a real operating system cannot be obtained.

2) In using this technical report, it is important to confirm accurately that equipment, tools, speech data characteristics, listening order, listening level and listening environment etc. have been set as planned. Users of this report should properly confirm the above items according to their test purposes.

3) The test equipment and tools (except the standard test tools) defined in Chapters 3 and 4 are

recommended. In case it is not possible to use the recommended products, e.g. when manufacturing has been discontinued, products of equal quality may be substituted.

1.4 Definitions

The main terms used in this technical report are defined below. In the Japanese version, the terms are ordered by the Japanese syllabary except for acronyms which are arranged alphabetically at the end. In the English version of the Report, all the terms have been arranged alphabetically.

Artificial voice:
Artificial voice mathematically described/generated to reproduce statistical speech characteristics such as short time spectrum and speech source characteristics according to Recommendation P.50 in ITU-T Blue Book Vol.V.

Cepstral Distance (CD):
Spectral distortion based on the difference between the Cepstral coefficients for reference speech and decoded speech (one of the evaluation indices in objective evaluation tests)

Delay time measurement test:
This is a test which measures the speech codec delay time. In this report, the test measures the real-time signal processing time only and does not include the algorithmic delay.

Encoded data, Decoded speech:
Data encoded/decoded by speech codecs. In the case of using the master speech codec, they are especially called encoded data for reference and decoded speech for reference, respectively.

Encoded data for reference, decoded speech for reference:
Data encoded and decoded by a master speech codec, used for comparative evaluation of test codecs.

Error data:
Error data simulating error characteristics of the transmission link in the host laboratory equipment. This is to be added to encoder output data.

Forward Error Correction (hard/soft decision decoding) :
A method for correcting bit errors caused by transmission on the receiver side. In PDC speech codecs, FEC interleaved encoding is applied. The decoding technique can be classified into hard-decision decoding using 2-digit quantization of the received signal and soft-decision decoding using more than 2 digits in the quantization.

Host laboratory equipment:
A speech quality evaluation system (hardware and software) used in the validation tests according to this technical report. The system has interfaces with a speech codec and inputs and outputs test speech, encoded data and decoded speech between the system and the speech codec.

Low segmental SNR (SNRfrq):
One of the evaluation indices in objective evaluation tests. The ratio (%) between the number of segmental SNRs which fail to reach the threshold and that of effective segments. The term effective segments is used to mean segments targeted to calculate the segmental SNR calculations.

Mean Opinion Score (MOS):
One type of successive category method by the psychological measurement method specified in Supplement 14 to the Series P Recommendations in ITU-T Blue Book Vol. V. Used as an evaluation

index in subjective evaluation tests. Normally, listeners are asked to evaluate quality judgment in the five levels of excellent, good, fair, poor and bad. Marks of 5, 4, 3, 2 and 1, respectively, are given to the evaluated levels and evaluation is made by the weighted mean method at the selected time interval.

Master speech codec:
A speech codec obtained by faithfully implementing the full-rate and half-rate speech encoding systems of the Standard for the Personal Digital Cellular Telecommunication System. This codec is used in comparative evaluation of test speech codecs in the validation tests specified in this technical report.

Modulated Noise Reference Unit (MNRU) Speech:
Speech obtained by adding white noise in proportion to the amplitude level of speech signal by MNRU specified by Recommendation P. 81 in ITU-T Blue Book Vol. V. The S/N ratio between the speech signal and white noise is called the Q value and the Q value, which equals the quality of encoded speech to be evaluated, is called an "opinion-equivalent Q value" of encoded speech. Nine Q values are used in subjective evaluation tests in the validation tests under this standard, namely 0, 5, 10, 15, 20, 25, 30, 35 and 40dB.

Objective evaluation test:
A test used to evaluate the technical characteristics of speech codecs implemented by the method specified in the standard, by numerically evaluating the physical speech measures such as segmental SNR, low segmental SNR and Cepstral distance, rather than by listening tests of encoded and decoded speech. Tests use a speech quality evaluation system, and evaluate encoding and decoding characteristics of the test codecs by comparing them with those of a master speech codec used as a reference.

Quality recommendation:
The technical characteristics (quality criteria) which speech codecs manufactured in accordance with the speech encoding system of the Standard for the Personal Digital Cellular Telecommunication System should meet for interconnection within a system.

Reference speech for subjective evaluation test:
Source speech and MNRU speech used for obtaining opinion-equivalent Q values in subjective evaluation test.

Segmental Signal to Noise Ratio (SNRseg):
One of the evaluation measures in objective evaluation tests. An S/N ratio calculated for each segment is averaged so that evaluation of low-level parts can be reflected.

Source speech:
Speech which has passed through a bandpass filter conforming to Recommendation G.714 in ITU-T Blue Book Vol.III.
In this evaluation test, this term construes as speech data sampled with a 8kHz sampling frequency, and a quantization accuracy of 14 bits. This includes artificial voice in this technical report.

Speech sample:
Decoded speech and reference speech used for objective evaluation test or subjective evaluation test

Subjective evaluation test:
A test used to evaluate the technical characteristics of speech codecs implemented by the method specified in the standard by a listening test of encoded and decoded speech. Listening test are conducted by randomly editing speech created by test speech codecs with reference speech (MNRU speech and source speech). Mean opinion score (MOS) is used as an index.

Test speech codec:
A speech codec implemented in hardware by a specification of the full-rate and half-rate speech encoding systems of the Standard for the Personal Digital Cellular Telecommunication System. This is an object of the validation tests specified in this technical report.

CHAPTER 2 OUTLINE OF VALIDATION TEST

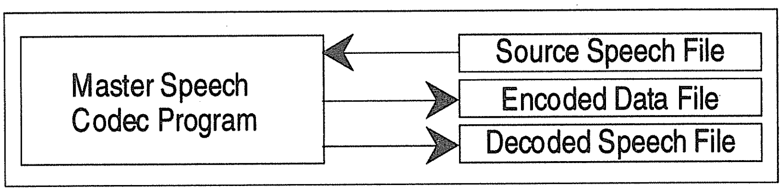
2.1 Test System

This test shall, in principle, be conducted using test speech codecs as test objects which contain the same speech codec LSI which will be incorporated in product codecs.

Tests can be conducted by using a substitute which obtains the same arithmetic results as those of a manufactured LSI and operates in real time, if the same speech codec LSI cannot be contained as a test speech codec. The differences between the substitute and a manufactured LSI should, however, be documented to clarify that the substitute is not the real product but just a satisfactory substitute.

In the tests, encoded data and decoded data speech suitable for the test conditions should first be obtained by using a master speech codec program, a source speech file, a computer for master speech codec, a test speech codec, host laboratory equipment and a host laboratory equipment interface etc. as shown in Fig.2.1-1 and Fig.2.1-2. Encoded data and decoded data are obtained for input signals of the decoder and for speech validation and the delay time measurement, respectively.

Two different methods, the objective and the subjective methods, are prepared for speech evaluation. Decoded speech is used for each method. It is also used for delay time measurement. The test outline, the quality recommendation and the testing methods are provided in the corresponding chapters or sections. Refer to Sections 3.3 and 4.3 for the details of the test equipment and tools. Refer to Section 3.3.2 for the details of the host laboratory equipment interface.



Master Speech Codec Computer
Fig.2.1-1 Master speech codec test connection configuration

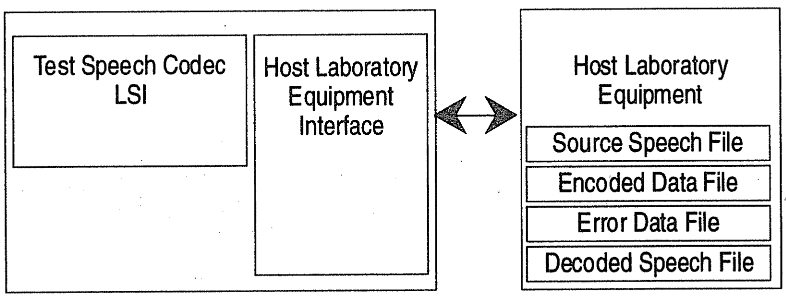


Fig. 2.1-2 Test speech codec test connection configuration

2.1.1 Testing methods

In order to judge whether or not test speech codecs comply with the quality recommendation, the following two evaluation tests can be used as testing methods:

1) Objective evaluation test

The test is conducted prior to the subjective evaluation tests. This test is not mandatory, but it is recommended that it be conducted in order to increase the likelihood of a successful subjective evaluation test. In the objective evaluation test, the test speech codec is evaluated comparing the coding and decoding characteristics with the master speech codec, using the speech quality evaluation method specified in this technical report and in which the test speech codec processing delay time is evaluated, and each test is conducted based on the corresponding test items and conditions of this report.

2) Subjective evaluation test

The test is conducted to make a pass/fail judgment on the test speech codec. In the subjective evaluation test, the decoded speech of the combination of the master speech codec and the test speech codec is evaluated by more than one listener, using the speech quality evaluation method specified in this technical report and in which the test speech codec processing delay time is evaluated, and each test is conducted based on the corresponding test items and conditions of this report.

2.1.2 Objective evaluation test

1) Test overview

The objective evaluation test involves evaluation of all functions including the optional items of the Standard (as shown in Fig. 2.1.2-1) : speech encoding (including noise canceller and low-volume suppression for half-rate speech codecs), channel encoding, channel decoding and speech decoding. The test is conducted by using the objective evaluation values segmental SNR (SNRseg), low segmental SNR frequency (SNRfrq) and Cepstral distance (CD) for respectively the encoder, the decoder (containing or not a postfilter) and the codec. After objective evaluation, the delay time is measured in order to confirm delay characteristics. Section 2.1.4 gives an overview of the delay time measurement test.

The objective evaluation value is the comparison value of the test speech with the reference speech which is obtained by passing source speech through a master speech codec and which is obtained by calculation. A method using source speech as reference speech may be also accepted as well as the above mentioned reference speech, according to the method defined in this report.

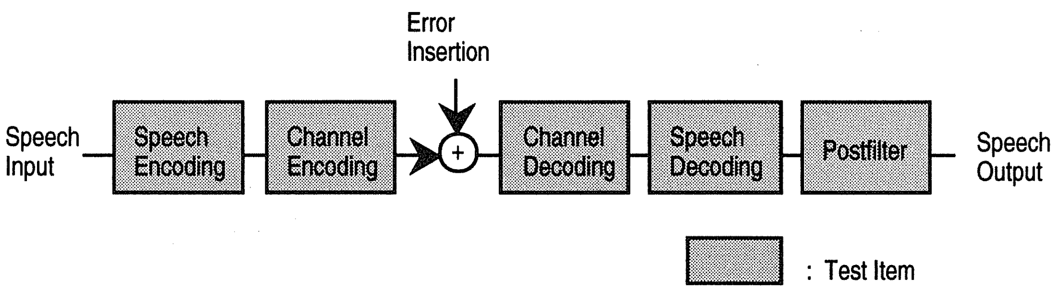


Fig. 2.1.2-1 Objective evaluation test

The reference system and the test system are defined as shown in Tables 3.1-1 and 4.1-1 in accordance with the combination of an encoder and a decoder and whether or not a postfilter and other devices are connected. Test Systems 1 to 6 are used for the tests.

2) Test procedure

Tests are conducted by the procedures shown in Figs. 3.1-1 and 4.1-1.

3) Target characteristics and judgment

The required characteristics of the objective evaluation test are not specified in this report. The person who is going to conduct an objective evaluation test should judge the test results according to the target characteristics value he/she defines. In the delay time measurement test, however, one should judge the test results according to the defined target value within the required characteristics of quality recommendation specified in Section 3.1.1.3 or 4.1.1.3.

2.1.3 Subjective evaluation test

1) Test overview

The test involves evaluation of all functions: speech encoding, channel encoding, channel decoding and speech decoding, including the optional items of the standard. Values are evaluated by a MOS evaluation conducted by preselected listeners. For listening, a listening tape is used, in which decoded speech and the reference speech (source speech, MNRU speech) are recorded.

Decoded speech includes reference speech obtained by passing source speech through a master speech codec. The MOS value of this speech is used as a reference in making the pass/fail judgment in the subjective evaluation. The reference system for MOS evaluation is System 1 of Table 3.1-1 or 4.1-1 and the test systems are Systems 2, 3 and 4 of the same tables.

The equipment configuration for producing listening tape is shown in Fig. 2.1.3-1 and that for the listening test is shown in Fig. 2.1.3-2. After subjective evaluation, the delay time is measured in order to confirm delay characteristics. Section 2.1.4 gives an overview of the delay time measurement test.

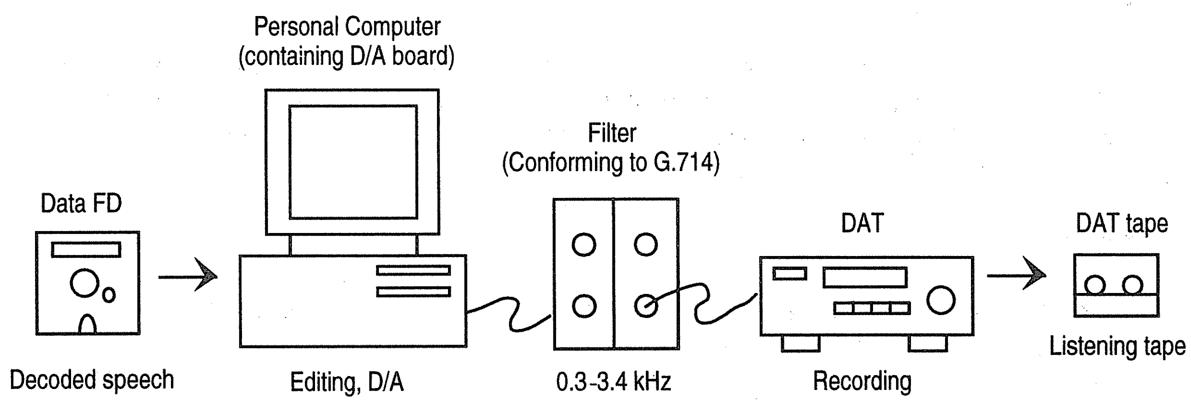


Fig. 2.1.3-1 Equipment configuration for producing the listening tape

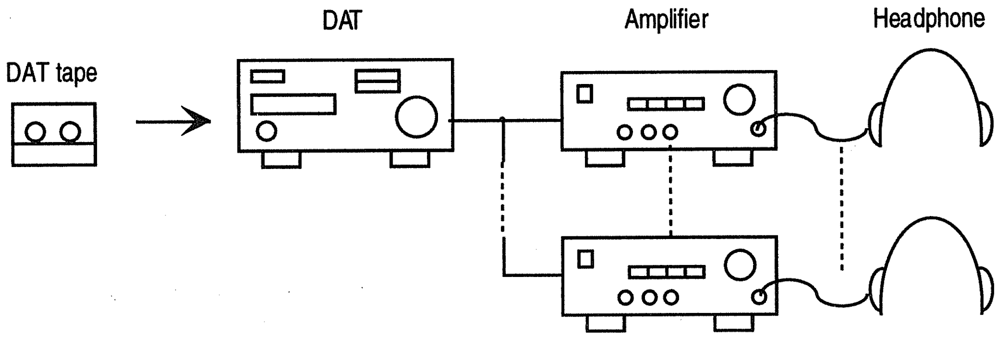


Fig. 2.1.3-2 Equipment configuration for listening test

2) Test procedure

Figs. 3.1-2 and 4.1-2 show the procedures for the tests.

Sections 3.2.2 and 4.2.2 describe the details of subjective evaluation.

3) Required characteristics and judgment

The codec is considered to have passed the subjective evaluation test, if it satisfies the requirements specified in Sections 3.1.1.2 and 3.1.1.3 and if is judged successful in all the judgment items specified in Section 3.2.2.10, or if it satisfies the requirements specified in Sections 4.1.1.2 and 4.1.1.3 and it is judged successful in all the judgment items specified in Section 4.2.2.10.

2.1.4 Delay time measurement test

1) Test overview

This test is conducted by measuring the delay time of the test speech codec using the master speech codec. The delay time to be measured should be the sum of the speech processing time in the speech codec and the delay time due to hardware, excluding the algorithmic delay time (signal reading time, interleaving time and signal outputting time).

2) Test procedure

Decoded speech samples of the master speech codec and the tested speech codec are prepared. Each decoded speech sample should use the same source speech, and there should be no artificial channel error. The time relationship between the two decoded speech samples is as follows with reference to the host laboratory decoded speech output starting time (Refer to Fig. 8 Signal Timing (1/2), Appendix 1):

a) : Decoded speech of master speech codec is set
1 : to start outputting 48.125 ms before the reference time in full-rate
2 : to start outputting 80 ms after the reference time in half-rate

b) : Decoded speech of the test speech codec starts outputting after an interval corresponding to the delay times of the self-interface and test speech codec, counted from the start of the host laboratory decoded speech output, in accordance with the rules of the host laboratory interface in both full-rate and half-rate.

Note that self-interface delay time means the delay time contribution of additional interface devices attached at the convenience of the tester to the test circuit when the host laboratory equipment is connected with the test speech codec.

c) : Consequently, the time relationship between decoded speech of the master speech codec and the test speech codec is fixed when the host laboratory decoded speech output starting time is common. This is illustrated in Figs. 2.1.4-1 and 2.1.4-2.

As shown in the figures, the delay time of the master speech codec (D_m), the delay time of the test speech codec system measured with decoded speech of the master speech codec as the reference (D_T), the self-interface delay time (D_{IF}) and the test speech codec processing time (excluding algorithmic delay time) (D) obey the following relation :

$$D = D_m + D_T - D_{IF}$$

Note that D_m is a negative value for full-rate and positive for half-rate.
Since parameters except processing time in speech codec configuring a delay time are known, the processing time in the speech codec, i. e. the required delay time, can be measured.

The algorithm-induced delay time is supposed to be equal in speech codecs using the same algorithm. Therefore, the parameter for this is not provided in the relative methods explained above.

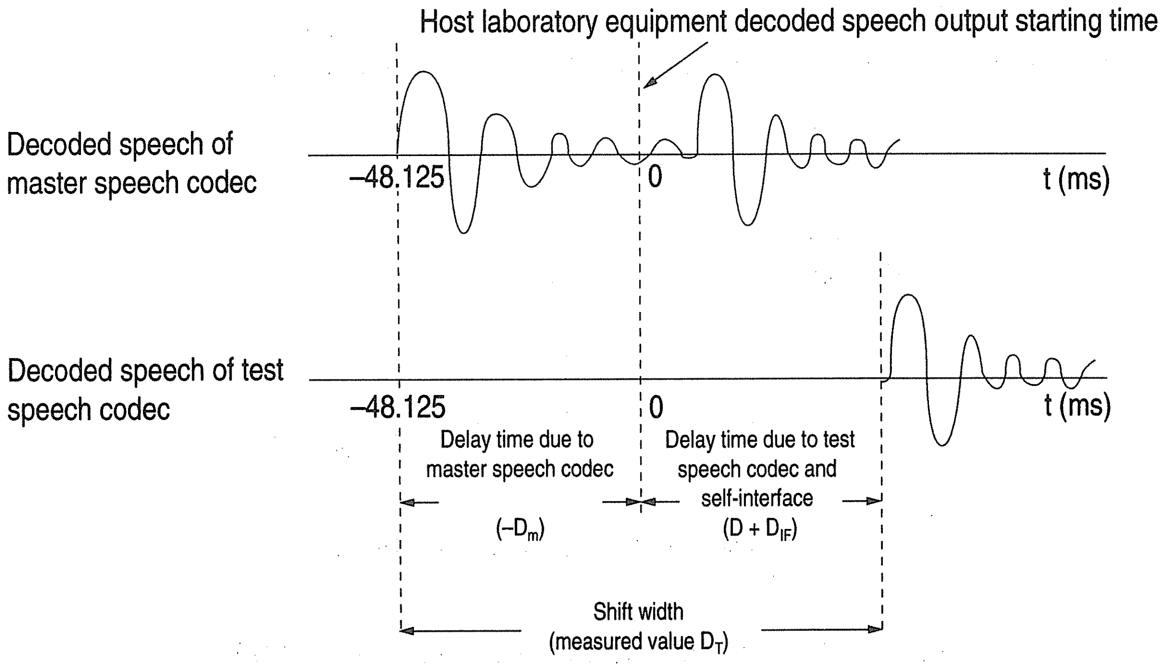


Fig.2.1.4-1 Full-rate speech codec delay time measurement principle

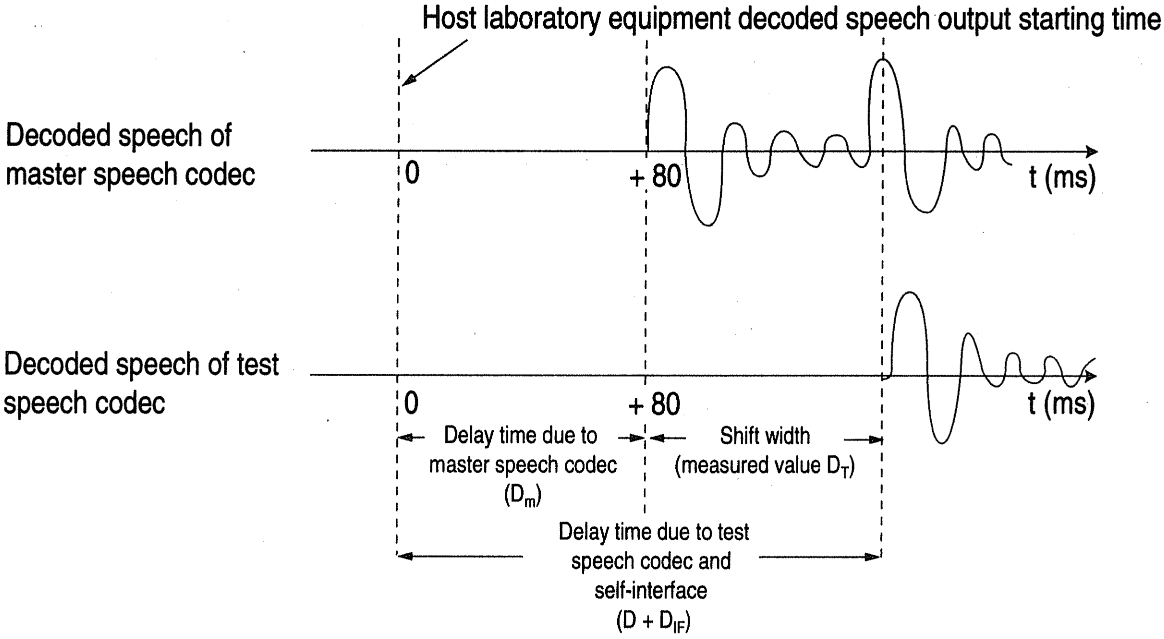
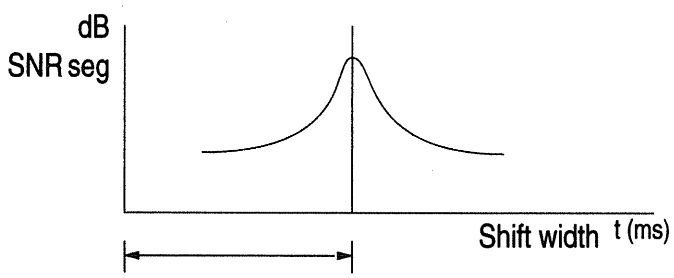


Fig.2.1.4-2 Half-rate speech codec delay time measurement principle

d) : The measured D_T value is obtained as the shift width with the maximum SNRseg value of the test speech codec decoded speech, using decoded speech of the master speech codec as the time reference. The delay time D (excluding algorithmic delays) is then calculated by the relational expression.



Shift width with maximum value of SNRseg
= Delay time (D_T)

Fig.2.1.4-3 The relation of decoded speech delay time and SNRseg

3) Required characteristics and judgment.
The tested items are assumed as having passed, if the delay time requirements specified in Sections 3.1.1.3 and 4.1.1.3 are fulfilled. (Refer to Sections 2.1.2 and 2.1.3)

CHAPTER 3 FULL-RATE SPEECH CODEC (VSELP SYSTEM)

3.1 Quality Recommendation

1) Test and procedure

This section describes requirements for speech quality and delay time in the objective evaluation test and the subjective evaluation test described in Sections 2.1.2 and 2.1.3, respectively.

The objective evaluation test is, based on target characteristics decided in advance by the tester, conducted in accordance with the procedures specified in Fig. 3.1-1 and concluded by judging whether or not the characteristics are satisfied.

The subjective evaluation test is, based on the subjective evaluation requirements, conducted in accordance with the procedures specified in Fig. 3.1-2 and concluded by a pass/fail judgment.

The delay time measurement test is based on the delay time requirement and is executed in accordance with the procedures specified in Figs. 3.1-1 and 3.1-2 after which the result is determined.

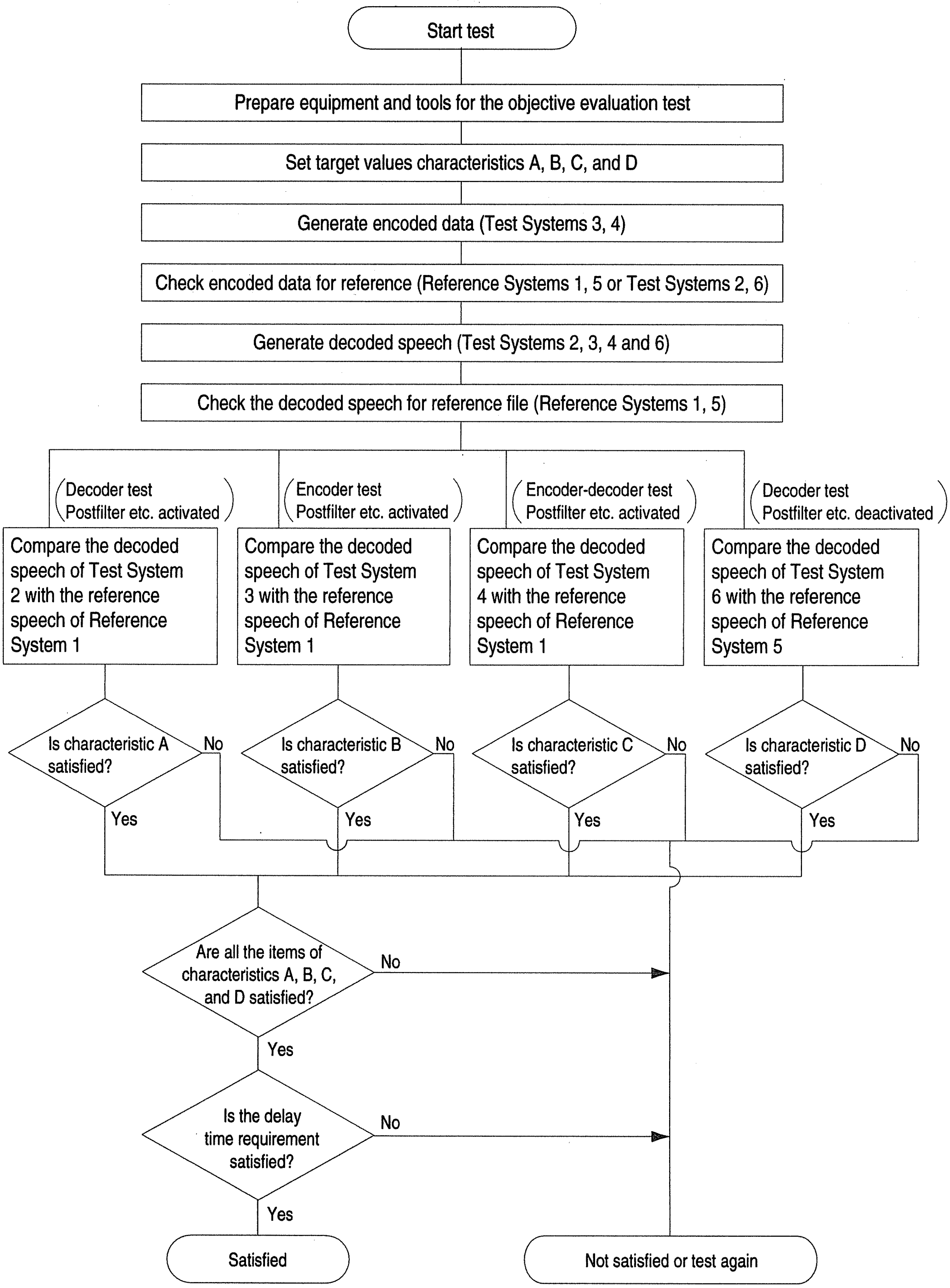


Figure 3.1-1 Procedure for Objective Evaluation Test of Full-Rate Speech Codec

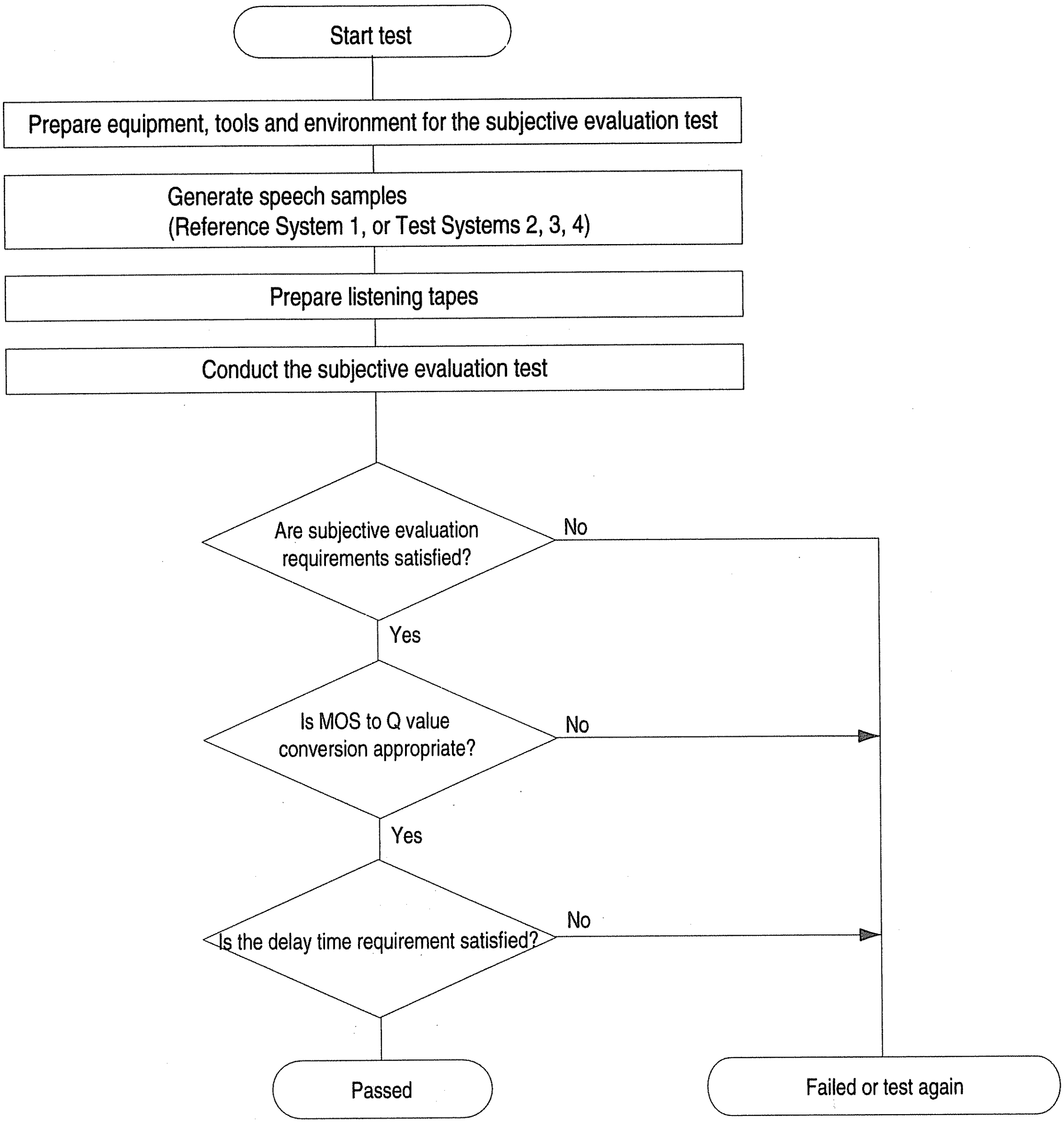


Figure 3.1-2 Procedure for Subjective Evaluation Test of Full-Rate Speech Codec

2) Speech codec reference and test systems

Reference or test systems : The reference and test systems are numbered as follows in accordance with combinations of master and test speech codecs.

Table 3.1-1 Reference and Test Systems of Full-Rate Speech Codec

Reference System (i) or Test System (j)	Encoder/Decoder	Postfilter, etc*
i=1	Master encoder/Master decoder	Activated
j=2	Master encoder/Test decoder	Activated
j=3	Test encoder/Master decoder	Activated
j=4	Test encoder/Test decoder	Activated
i=5	Master encoder/Master decoder	Deactivated
j=6	Master encoder/Test decoder	Deactivated

* Postfilter, etc. for full-rate speech codecs represents a high-pass filter (5.1.1.3 in RCR STD-27), adaptive pitch prefilter (5.1.4.4.8), and adaptive spectral postfilter (5.1.4.4.9).

3) Codec test items

Test items : Each item is numbered as follows.

Table 3.1-2 Test Items for Full-Rate Speech Codec

Test Item (k)	Condition
1	Basic Characteristics
2	Error Robustness
3	
4	Level Variation
5	
6	Background Noise
7	
8	Talker Dependency

4) Explanations of symbols used in this chapter

SNRseg : Segmental SNR (objective evaluation value calculated for each test item). $SNRseg_k$ (j/i) is the mean value of the segmental SNR of the output of Test System j to that of Reference System i in Test Item k.
Section 3.2.1.5.1 defines the calculation method.

SNRfrq : Low segmental SNR frequency (objective evaluation value calculated for each test item). $SNRfrq_k$ (j/i) is the fraction of effective segments in Test Item k with SNR, measured as the ratio of the output of Test System j to that of Reference System i, of less than 15 dB.
Section 3.2.1.5.2 defines the calculation method.

CD : Cepstral distance (objective evaluation value calculated for each test item). CD_k (j/i) is the mean value of the Cepstral distance from the output of Test System j to that of Reference System i in Tcst Item k.
Section 3.2.1.5.3 defines the calculation method.

MAv : MAv_{k1} is the mean value of subjective evaluation marks when Reference System 1 is applied to Test Item k (MOS value).
Section 3.2.2.10 defines the calculation method.

MVr : MVr_{k1} is the unbiased variance of subjective evaluation marks when Reference System 1 is applied to Test Item k.
Section 3.2.2.10 defines the calculation method.

TAv : TAv_{kj} is the mean value of subjective evaluation marks when Test System j (j=2,3,4) is applied to Test Item k (MOS value).
Section 3.2.2.10 defines the calculation method.

TVr : TVr_{kj} is the unbiased variance of subjective evaluation marks when Test System j (j=2,3,4) is applied to Test Item k.
Section 3.2.2.10 defines the calculation method.

t : t_{kj} is the index value in a significant difference test between MAv_{k1} and TAv_{kj} during Test Item k.
Section 3.2.2.10 defines the calculation method.

3.1.1 Speech quality targets and requirements

3.1.1.1 Objective evaluation target characteristics

As target characteristics, the tester sets thresholds for the objective evaluation values SNRseg, SNRfrq and CD. Thresholds are assigned to SNRseg as a lower limit, and to SNRfrq and CD as upper limits.

The objective evaluation target characteristics shown in Fig. 3.1-1 are defined as follows:

Characteristic A: The similarity of the test decoder with the master decoder is evaluated.

Test System j = 2 should be compared with Reference System i = 1 and the following must be satisfied in all of Test Items k = 1 to 8:

$SNRseg_k$ (2/1) $\geq$ Threshold (lower limit) of $SNRseg_k$(2/1)

$SNRfrq_k$ (2/1) $\leq$ Threshold (upper limit) of $SNRfrq_k$ (2/1)

CD_k (2/1) $\leq$ Threshold (upper limit) of CD_k (2/1)

The tester sets the thresholds (lower or upper limits) of $SNRseg_k$ (2/1), $SNRfrq_k$ (2/1), and CD_k (2/1) and enters the values in a free-format form. It is recommended that a column for evaluation values also be included in the form (this applies to B, C and D as well).

Characteristic B: The similarity of combinations of test encoder and master decoder with the master speech codec is evaluated.

Test System j = 3 should be compared with Reference System i = 1 and the following must be satisfied in all of Test Items k = 1 to 8:

$SNRseg_k$ (3/1) $\geq$ Threshold (lower limit) of $SNRseg_k$(3/1)

$SNRfrq_k$ (3/1) $\leq$ Threshold (upper limit) of $SNRfrq_k$ (3/1)

CD_k (3/1) $\leq$ Threshold (upper limit) of CD_k (3/1)

The threshold values (lower or upper limits) of $SNRseg_k$ (3/1), $SNRfrq_k$ (3/1) , and CD_k (3/1) are entered in a free-format form.

Characteristic C: The similarity of combinations of test encoder and test decoder with the master speech codec is evaluated.

Test System j = 4 should be compared with Reference System i = 1 and the following must be satisfied in all of Test Items k = 1 to 8:

$SNRseg_k$ (4/1) $\geq$ Threshold (lower limit) of $SNRseg_k$(4/1)

$SNRfrq_k$ (4/1) $\leq$ Threshold (upper limit) of $SNRfrq_k$ (4/1)

CD_k (4/1) $\leq$ Threshold (upper limit) of CD_k (4/1)

The threshold values (lower or upper limits) of $SNRseg_k$ (4/1), $SNRfrq_k$ (4/1), and CD_k (4/1) are entered in a free-format form.

Characteristic D: The similarity of the test decoder with the master decoder is evaluated.

Test System j = 6 should be compared with Reference System i = 5, and the following must be satisfied in all of Test Items k = 1, 4, 5 and 8:

$SNRseg_k$ (6/5) $\geq$ Threshold (lower limit) of $SNRseg_k$(6/5)

CD_k (6/5) $\leq$ Threshold (upper limit) of CD_k (6/5)

The thresholds (lower or upper limits) of $SNRseg_k$ (6/5) and CD_k (6/5) are entered in a free-format form.

Section 1 of Appendix 6 shows sample data of the objective evaluation test for the full-rate speech codec.

3.1.1.2 Subjective evaluation requirements

The subjective evaluation requirements shown in Fig. 3.1-2 are defined as follows:

After comparing with References System i = 1, all the Test Systems j = 2 to 4 for all the Test Items k = 1 to 8 should satisfy $t_{kj} \leq 1.645$, where t_{kj} is calculated with a five-per-cent, one-sided significant difference test.

Conversion of MOS values into opinion-equivalent Q values shall be verified to be appropriate.

3.1.1.3 Delay time requirement

The delay time of the test speech codec (D) measured by the method specified in Section 3.2.1.6 shall be 40 ms or less, excluding the algorithmic delay time.

3.2 Validation Test Methods

3.2.1 Objective evaluation method

3.2.1.1 Definitions

Objective evaluation tests are conducted to obtain objective evaluation data of test speech codecs to judge conformance of speech quality.
For the evaluation, the segmental SNR (SNRseg), low segmental SNR frequency (SNRfrq), Cepstral distance (CD), and delay time characteristic values are used.

3.2.1.2 Reference and test systems

Master and test speech codecs are combined in accordance with the reference and test systems shown in Table 3.1-1.

3.2.1.3 Test items

Table 3.2.1.3-1 lists eight test items covering basic characteristics (1), error robustness (2), level variation (2), background noise (2) and talker dependency (1).

Table 3.2.1.3-1 Test Items for Full-Rate Speech Codec

Test	Test Item	Error Condition	Input Level	Superposed Noise	Number of speech samples	
					Speech	Artificial Voice
Basic Characteristics	1	Error-free	Normal	None	16	2
Error Robustness	2	1% (4km/h)	Normal	None	4	2
		1% (20km/h)	Normal	None	4	2
		1% (60km/h)	Normal	None	4	2
		1% (60km/h)*	Normal	None	4	2
	3	3% (4km/h)	Normal	None	4	2
		3% (20km/h)	Normal	None	4	2
		3% (60km/h)	Normal	None	4	2
		3% (60km/h)*	Normal	None	4	2
Level Variation	4	Error-free	Normal - 10dB	None	4	2
	5	Error-free	Normal - 20dB	None	4	2
Background Noise	6	Error-free	Normal	Low noise SNR30 dB	8	2
	7	Error-free	Normal	High noise SNR15 dB	8	2
Talker Dependency	8	Error-free	Normal	None	8	—

NOTES 1: Error data marked with an asterisk are due to fading at 1.5 GHz. Error data without an asterisk are due to fading at 800 MHz.
2: "Normal" level is the level whose mean level is 21 dB lower than the sine-wave full scale of the A/D converter.
3: Superposed noise is defined as follows:
None: Noise which is unavoidably mixed during generation of source speech and quantization noise.
Low noise: SNR 30dB
High noise: SNR 15dB

3.2.1.4 Test preparations

Encoded data and decoded speech are generated using the speech samples defined in Table 3.2.1.4-1.

1) Source speech

Source speech is composed of speech signals and artificial voice signals. Eighty speech signals defined in Section 3.3.1 are used. Ten artificial voice signals defined in Section 3.3.1 are used. In the artificial voice column for the basic characteristics and error robustness tests in Table 3.2.1.4-1, artificial male and female voices are denoted PM1-9 and PF1-9, respectively, to distinguish the speech samples. Source speech from one male and one female artificial voice is used in common as source speech for these artificial voices. Artificial voice is not used in the talker dependency test.

The data format of individual speech signals and artificial voice signals is 16 bits (only the most significant 14 bits are valid) linear PCM with a time duration of 6s.

2) Generation of encoded data

For Systems 3 and 4 shown in Table 3.1-1, encoded data is generated from source speech with the test encoder. The number of encoded data samples is 80 for speech signals and 10 for artificial voice signals, totaling 90.

The encoded data for reference defined in Section 3.3.1 should be used for Systems 1, 2, 5, and 6.

3) Error data

From the 32 error data samples defined in Section 3.3.1, one sample each is used with each speech sample of Groups 1 to 4 in Test Items 2 and 3 of Table 3.2.1.4-1.

Error data from the same test items as those of Group 1 are used for Group 5 regarding error data for artificial voice. Error data from the same test items as those of Group 2 are used for Group 6.

4) Generation of decoded speech

Decoded speech is generated for Systems 2 to 4 and 6 shown in Table 3.1-1 by using the host laboratory equipment and master speech codec defined in Section 3.3. Level-variation speech samples in Test Items 4 and 5 are not returned to the normal level. (Level correction is not performed.) For artificial voice signals of Test Items 1-3 (basic characteristics and error robustness), 9 male and 9 female decoded speech samples are generated by applying different error data to each of the male and female speech encoded data samples. The number of decoded speech samples used for each of Systems 2 to 4 is 80 for speech signals and 26 for artificial voice signals, totaling 106. The number of decoded speech samples in System 6 is 32 for speech signals and 6 for artificial voice signals, totaling 38.

The decoded speech samples defined in Section 3.3.1 are used with Systems 1 and 5. Decoded speech samples should be regenerated if noticeable alternants are found in audition.

Table 3.2.1.4-1 Speech Samples of Full-Rate Speech Codec

Test		Test Item	Speech				Artificial Voice	
			Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Basic Characteristics		1	M1 F3 M5 F7	F1 M3 F5 M7	M2 F4 M6 F8	F2 M4 F6 M8	PM1	PF1
Error Robustness	1%	2	M9	F9	M10	F10	PF2	PM2
			F11	M11	F12	M12	PM3	PF3
			M13	F13	M14	F14	PF4	PM4
			F15	M15	F16	M16	PM5	PF5
	3%	3	M17	F17	M18	F18	PF6	PM6
			F19	M19	F20	M20	PM7	PF7
			M21	F21	M22	F22	PF8	PM8
			F23	M23	F24	M24	PM9	PF9
Level Variation		4	M25	F25	M26	F26	PF10	PM10
		5	F27	M27	F28	M28	PM11	PF11
Background Noise		6	M29 F31	F29 M31	M30 F32	F30 M32	PF12	PM12
		7	M33 F35	F33 M35	M34 F36	F34 M36	PM13	PF13
Talker Dependency		8	B1 G3	G1 B3	B2 G4	G2 B4	—	—

3.2.1.5 Speech quality measurement method

Using decoded speech for reference from System 1 or 5 shown in Table 3.1-1 as a reference, segmental SNRs (SNRseg), low segmental SNR frequencies (SNRfrq) and Cepstral distances (CD) of decoded speech for testing from one of Systems 2, 3, 4, or 6 are obtained for individual speech samples.

3.2.1.5.1 Calculation of segmental SNR (SNRseg)

SNRseg is the value obtained by averaging SNR for each segment length of 10ms (80 samples) by non-silent intervals. The calculation procedures are shown below:

1) The SNR of the jth segment is defined as shown in the following expression:

$$
\mathrm{SNRj} = \begin{cases} \mathrm{Vclip+}, & (\mathrm{Ps} > \mathrm{P} \text{ or } \mathrm{Pd} > \mathrm{P}) \text{ and } \mathrm{Rdif} = 0 \\ \mathrm{Vclip+}, & (\mathrm{Ps} > \mathrm{P} \text{ or } \mathrm{Pd} > \mathrm{P}) \text{ and } \mathrm{Rdif} \neq 0 \text{ and } \mathrm{Vsubf} > \mathrm{Vclip+} \\ \mathrm{Vsubf}, & (\mathrm{Ps} > \mathrm{P} \text{ or } \mathrm{Pd} > \mathrm{P}) \text{ and } \mathrm{Rdif} \neq 0 \text{ and } \mathrm{Vclip-} \leq \mathrm{Vsubf} \leq \mathrm{Vclip+} \\ \mathrm{Vclip-}, & (\mathrm{Ps} > \mathrm{P} \text{ or } \mathrm{Pd} > \mathrm{P}) \text{ and } \mathrm{Rdif} \neq 0 \text{ and } \mathrm{Vsubf} < \mathrm{Vclip-} \\ \text{invalid}, & \mathrm{Ps} \leq \mathrm{P} \text{ and } \mathrm{Pd} \leq \mathrm{P} \end{cases}
$$

$$Vsubf = 10\log_{10}\left\{\sum_{i=j*80}^{j*80+79}[s(i)]^2/Rdif\right\}$$

$$Ps = 10\log_{10}\sum_{i=j*80}^{j*80+79}[s(i)]^2$$

$$Pd = 10\log_{10}\sum_{i=j*80}^{j*80+79}[d(i+T)]^2$$

$$Rdif = \sum_{i=j*80}^{j*80+79}[s(i)-d(i+T)]^2$$

s (i) : ith sample of decoded speech for reference

d (i) : ith sample of decoded speech for testing

T : Speech codec delay time (expressed in sample units)
Delay time differs in accordance with test system.

P : Threshold level of non-silence/silence (P = -62dB)

Provided $10\log_{10}\sum_{i=0}^{79} Smax^2$ is 0dB.

Smax : Maximum input value (32768 for 16 bits)

Vclip+ : Maximum value of SNRj (Vclip+ = 80dB)

Vclip- : Minimum value of SNRj (Vclip - = -5dB)

2) Define Mj as shown in the following expression:

$$Mj = \begin{cases} +1, & SNRj \neq invalid \\ 0, & SNRj = invalid \end{cases}$$

3) Define total segmental SNR as follows:

$$SNRseg = \frac{1}{M}\sum_{j=0}^{N-1}[SNRj]$$

$$M = \sum_{j=0}^{N-1} Mj$$

N : Number of segments

3.2.1.5.2 Calculation of low segmental SNR frequency (SNR_{frq})

1) Define Hj as shown in the following expression:

$$Hj = \begin{cases} +1, & SNRj \neq \text{invalid and } SNRj < s \\ 0, & SNRj = \text{invalid or } SNRj \geq s \end{cases}$$

s : Threshold of SNR inside segment (15 dB).

2) Calculate SNR_{frq} as follows:

$$SNR_{frq} = \frac{1}{M} \sum_{j=0}^{N-1} [Hj] \times 100\ (\%),$$

where SNRj, M and N are those used in segmental SNR calculations.

3.2.1.5.3 Calculation of Cepstral distance (CD)

1) Define total CD as follows:

$$CD = \frac{1}{M} \sum_{j=0}^{N-1} CDj$$

CDj is calculated only for Segment j for which segmental SNR is calculated for the speech file consisting of N segments. Therefore, the number of effective segments M is the same as that used in segmental SNR.

2) Define CD of the jth segment as follows:

$$CDj = \frac{10}{\ln 10} \sqrt{2 \sum_{n=1}^{q} (c_n - \hat{c}_n)^2}$$

$c_n, \hat{c}_n$: nth order LPC Cepstral coefficients obtained by the pth order linear prediction of the decoded speech for reference and decoded speech for testing of the jth segment, respectively ($1 \leq n \leq q$, $p = 10$, $q = 30$).

3) The procedure to calculate LPC Cepstral coefficients is shown below:

$$c_n = \begin{cases} -\alpha_n, & n = 1 \\ -\sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \alpha_k c_{n-k} - \alpha_n, & 1 < n \leq p \\ -\sum_{k=1}^{p} \left(1 - \frac{k}{n}\right) \alpha_k c_{n-k} & p < n \leq q \end{cases}$$

α_k : Short-term prediction coefficient of the kth order in the linear prediction of the pth order. ($1 \leq k \leq p$)

Refer to 5.2.1.4 in the Personal Digital Cellular Telecommunication System RCR STD-27 for the method of calculating short-term prediction coefficients.

3.2.1.6 Delay time measurement method

Measure the delay time (excluding algorithmic delays) of the test codec according to the method described in Section 2.1.4. Alternatively, the delay time of the test codec can be calculated by measuring and adding the delay times of the test encoder and decoder (respectively Denc and Ddec).

1) Delay time of test codec

$D = D_m + D_T - D_{IF}$

D_m : Delay time of master speech codec (-48.125 ms)

D_T : Delay time at which SNRseg of decoded speech by System 4 becomes a maximum, using decoded speech by System 1 as the reference (in ms).

D_{IF} : Delay time caused by connecting with host laboratory equipment interface (by self-declaration).

2) Delay time of test encoder

$D_{enc} = D_{me} + D_{Te} - D_{IFe}$

D_{me} : Delay time of master speech encoder (-28.125 ms)

D_{Te} : Delay time at which SNRseg of decoded speech by System 3 becomes a maximum, using decoded speech by System 1 as the reference (in ms).

D_{IFe} : Delay time on the encoder side caused by connecting with host laboratory equipment interface (by self-declaration).

3) Delay time of test decoder

$D_{dec} = D_{md} + D_{Td} - D_{IFd}$

D_{md} : Delay time of master speech decoder (-20 ms).

D_{Td} : Delay time at which SNRseg of decoded speech by System 2 becomes a maximum, using decoded speech by System 1 as the reference (in ms).

D_{IFd} : Delay time on the decoder side caused by connecting with host laboratory equipment interface (by self-declaration).

3.2.1.7 Analysis, judgment and recording

Calculate the mean values of the segmental SNR, low segmental SNR frequency and Cepstral distance for each test item shown in Table 3.2.1.3-1 and record these values as well as the measured delay time in a free-format form. The values entered should be rounded off to the nearest two decimal places.

Judge whether or not these mean values satisfy the target values and the delay times satisfy the requirements specified in Section 3.1.1.3.

3.2.2 Subjective evaluation method

3.2.2.1 Definitions

Subjective evaluation tests are conducted to evaluate speech quality conformance of test speech codecs based on the quality of decoded speech of the master speech codec. The subjective evaluation test is performed with listening tests on a subject population with an age and sex distribution representative for telephone users. The MOS value is used as the evaluation measure.

After completion of the subjective evaluation test, a measurement of the delay time is carried out in accordance with the method specified in Section 3.2.1.6.

3.2.2.2 Test preparations

The following speech samples shall be prepared:

1) Decoded speech

Decoded speech is generated by combining Reference System 1 and Test Systems 2 to 4 shown in Table 3.1-1.

2) Reference voice

The following ten types are used as reference voices. MNRU speech is used to verify that the subjective evaluation test is correct and to compare test results from different test organizations by converting MOS values to opinion-equivalent Q values.

(a) Source speech

(b) MNRU speech (9 types: Q = 0, 5, 10, 15, 20, 25, 30, 35, and 40 dB)

3.2.2.3 Test items

Table 3.2.1.3-1 lists the test items.

3.2.2.4 Speech samples

Speech samples of Groups 1 to 4 in Table 3.2.1.4-1 are generated with a postfilter and other devices using the equipment and tools defined in Section 3.3 by the four decoded speech reference and test systems specified in Section 3.2.2.2. (Artificial voices are not generated.) Level-variation voices in Test Items 4 and 5 are returned to the normal level (refer to Section 3.2.1.3) by using the level-correction program described in Section 3.3.5 after they have been generated. Use a total of 40 voices as reference voices generated by using source speech files of two males and females from source speech files, which are used to generate decoded speech, in accordance with Section 3.2.2.2.

Consequently, the number of speech files M will be:

M = 80 x 4 (decoded speech) + 40 (reference voice) = 360

3.2.2.5 Randomization

As shown in Table 3.2.1.4-1, decoded speech can be divided into four groups. Taking the ratio with reference voice into consideration, the group of Groups 1 and 2 with reference voice added is named Listening Group 1. Similarly, the group of Groups 3 and 4 with reference voice added is named Listening Group 2. The number of speech samples N of one listening group will be:

N = (20 + 20) x 4 (decoded speech) + 40 (reference voice) = 200

(x4 is speech codec combinations)

The order of listening speech samples inside listening groups is randomized for each listening group using randomization tables (refer to Section 3.3.4). Different randomizing should be used for each listener group (refer to Section 3.2.2.7) and the listening sequence of the listening group should be changed for each listener group.

3.2.2.6 Listening environment

Voices are heard through only one ear, wearing a headphone. Listeners can freely decide by themselves whether to hear through the left or right ear. After D/A conversion (16bit accuracy), decoded speech from the speech codec and reference voices are passed through a filter to remove out-band noise within 0.3 kHz and 3.4 kHz and over, and are presented to the listeners through a headphone. Except for an amplifier or an attenuator to adjust the volume to a level more comfortable to the ear, any equipment which affects speech quality should not be connected. Speech may be presented by using DAT recording and replay. A quiet place without ambient noise should be selected as a listening place.

The recommended listening level is -24dBPa (70dB SPL) and noise level is 35dBA (standard A characteristic) and under.

Section 3.3.6 describes the recommended listening equipment.

3.2.2.7 Listeners

The number of listeners must be 40 or more. The listener sexes and age composition shall be as follows:

Sex ratio	Males	(60±10)%
	Females	(40±10)%
Age composition	15-29 years	(30±10)%
	30-49 years	(50±10)%
	50 years or older	(20±10)%

Acoustic experts must not be included among listeners.

Listeners are divided into at least five groups (the number of each group members should be averaged) and each group listens to different randomized speech samples.

3.2.2.8 Test procedure

Before starting the test, distribute the following sheet to the subjects (listeners) and explain the test purpose and guidelines for procedures and scoring.

About Listening Test
Thank you for cooperating with us today in this test. You will now hear several examples of telephone sounds. Please evaluate the individual sounds, based on how you would feel if you used them in a telephone conversation, according to the following five-level scale: 5 Excellent 4 Good 3 Fair 2 Poor 1 Bad You will soon hear examples of telephone sounds. Evaluate the sounds by selecting one of the five levels. We will now start the test.

The listening conditions will be:

- One listener will listen to all speech samples.
- Speech samples are presented at intervals of 1s. Therefore, if 6s speech samples are used, the listening time per listening group will be:

(6 + 1) x 200 = 1400s = about 24 min

provided the number of speech samples per listening group N is 200.

- One listening group is listened to continuously. Between each session of 70 speech samples (about 8 min), listeners take a break for 1-2 minutes without leaving the room. A break of more than 20 minutes should be provided between the listening groups.

When beginning the test, present about ten training samples which are suitably selected to include various conditions from the speech samples to let the listeners practice how to mark their evaluation, but they do not have to be informed about this.

3.2.2.9 Delay time measurement method

Use the method specified in Section 3.2.1.6.

If measurements from the objective evaluation test can be applied as they are, it is allowed to use them without performing an additional measurement.

3.2.2.10 Analysis, judgment and recording

1) Results

Conduct a significant difference test of the mean subjective evaluation values for test speech codecs and the master speech codec for each test item shown in Table 3.2.1.3-1 using the method specified in Item (2). Judge whether or not these mean values satisfy the subjective evaluation requirements specified in Section 3.1.1.2. Enter the subjective evaluation values (MAv_{k1}, TAv_{kj}), unbiased variances

(MVr_{k1}, TVr_{kj}), index values (t_{kj}) and judgments in Table 3.2.2.10-1. Enter "Passed" in the Judgment column if the index value (t_{kj}) satisfies the requirements specified in Section 3.1.1.2. If not, enter "Failed". If the conversion error of MOS opinion-equivalent Q value described below is larger than the designated value, the subjective evaluation test is considered not relevant and is reconducted. Judge whether the delay time satisfies the requirements specified in Section 3.1.1.3 and record the judgment in a free-format form.

Table 3.2.2.10-1 Full-Rate Subjective Evaluation Values and Judgments

Test Item (k)	Reference System i = 1		Test System j = 2				Test System j = 3				Test System j = 4			
	MAv_{ki}	MVr_{ki}	TAv_{k2}	TVr_{k2}	t_{k2}	Judgment	TAv_{k3}	TVr_{k3}	t_{k3}	Judgment	TAv_{k4}	TVr_{k4}	t_{k4}	Judgment
1														
2														
3														
4														
5														
6														
7														
8														

2) Significant difference test between mean values

MAv_{k1}, TAv_{kj}, MVr_{k1} and TVr_{kj} defined in Section 3.1 express MOS values or the unbiased variance of the subjective evaluation marks for each test item. They are calculated by the formulas 3.2.2.10-1 to 3.2.2.10-4, respectively.

t_{kj} is calculated by formula 3.2.2.10-5 in accordance with significant difference tests between mean values without the unique equivalent variance assumption. Based on the value of t_{kj}, the tests are executed as one-sided tests with a significance level of 5% and a degree of freedom of ∞ (infinity).

$$MAv_{k1} = \frac{1}{n} \sum_{m=1}^{n} X_{k1}(m) \qquad (3.2.2.10\text{-}1)$$

$$TAv_{kj} = \frac{1}{n} \sum_{m=1}^{n} X_{kj}(m) \qquad (3.2.2.10\text{-}2)$$

$$MVr_{k1} = \frac{1}{n-1} \sum_{m=1}^{n} (X_{k1}(m) - MAv_{k1})^2 \qquad (3.2.2.10\text{-}3)$$

$$TVr_{kj} = \frac{1}{n-1} \sum_{m=1}^{n} (X_{kj}(m) - TAv_{kj})^2 \qquad (3.2.2.10\text{-}4)$$

$$t_{kj} = \frac{MAv_{k1} - TAv_{kj}}{\sqrt{\dfrac{MVr_{k1} + TVr_{kj}}{n}}} \qquad (3.2.2.10\text{-}5)$$

where $X_{k1}(m)$ and $X_{kj}(m)$ are the subjective evaluation marks for Test Item k in respectively Reference

System 1 and Test System j (j=2, 3, 4), and n is the number of evaluation marks for each test item.

3) Conversion from MOS to opinion-equivalent Q values

As mentioned above, subjective evaluation results are judged by using MOS values. However, the MOS evaluation depends on the listeners and selected test sounds. By converting MOS values obtained using MNRU speech as reference sound to opinion-equivalent Q values, the validity of the subjective evaluation test can be verified and a comparison with test results and other data from different subject groups can be carried out. The method used to convert MOS values to opinion-equivalent Q values is described below.

Generally, the relationship between MOS values and opinion-equivalent Q values can be explained as shown in Fig. 3.2.2.10-1.

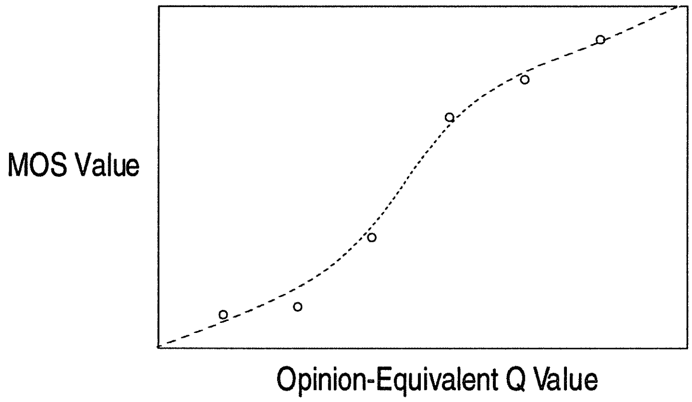


Figure 3.2.2.10-1 Diagram showing the general relationship between MOS values and opinion-equivalent Q values
Plots in the diagram show MOS evaluation values of MNRU speech. The dotted-line curve shows an MOS-opinion-equivalent Q value conversion curve.

The MOS-opinion-equivalent Q conversion model is expressed by Eq. 3.2.2.10-6.

$$Q = G \times \ln\left(\frac{MOS - 1}{MOS_{mx} - MOS}\right) + I \qquad (3.2.2.10\text{-}6)$$

where MOS is an MOS value (1 to 5) and Q is an opinion-equivalent Q value. MOS_{mx}, I, and G are parameters decided by the following procedures:

Step 1 : An initial value is given to MOS_{mx}. (Example: 3.50)

Step 2 : Using data from three points (MNRU 15dB, 20dB and 25dB) in the interval where the relationship between the MOS values and the opinion-equivalent Q values shown in Fig. 3.2.2.10-1 is comparatively linear, G and I are calculated by the linear regression method. The MOS-Q conversion mean square error MSE in regard to MOS values is then calculated by taking all the MNRU reference samples into account.

Step 3 : Steps 2 and 3 are repeated by applying $MOS_{mx} = MOS_{mx} + 0.01$ until $MOS_{mx} = 5.00$ is obtained.

Step 4 : The set of MOS_{mx}, G and I values with which MSE is minimized is used as parameter values in the MOS-to-Q conversion model.

If the conversion error MSE obtained above in the conversion from MOS values to opinion-equivalent Q values is greater than 0.0100, the subjective evaluation test is considered to be not relevant.

3.3 Test Equipment and Tools

This section describes the test equipment and tools for the full-rate speech codec system (VSELP) validation tests. The following equipment will be needed to obtain test data:

1) For both objective and subjective evaluation tests

- Test data
- Computer for host laboratory equipment
- Host laboratory equipment interface board
- Host laboratory equipment software
- Computer for master speech codec
- Master speech codec program

2) For objective evaluation tests

- SNRseg, SNRfrq and CD calculation programs

3) For subjective evaluation tests

- MNRU speech
- Level-correction program
- Listening tape editing computer
- Randomization table
- Randomization editing program
- Recording DAT
- Listening test environment described in Section 3.2.2.6
- Opinion-equivalent Q value calculation program
- MOS evaluation value aggregation and significant difference calculation program

3.3.1 Test data

Test data consist of files of stored source speech, encoded data for reference, decoded speech for reference, error data and MNRU speech. Data are supplied on a 3.5" magneto-optic disk. Refer to Appendix 5 for speech file naming rules.

1) Source speech file

Source speech is recorded by linear PCM and each sample value is recorded in a unit of words (16 bits each). Only the higher 14 bits are valid. The lower two bits are always 0. Each word is recorded in the order of lower and higher bytes. The source speech file provides 80 speech samples and 10 artificial voices. Technical specifications are listed in Appendix 4.

2) Encoded data for reference file

The following encoded data samples for reference are supplied:

With postfilter, etc.: 80 speech, 10 artificial voice

Without postfilter, etc.: 32 speech, 6 artificial voice

3) Decoded speech for reference file

The file stores decoded speech in the same format as that of the source speech. The following decoded speech samples are supplied:

With postfilter, etc.: 80 speech, 26 artificial voice

Without postfilter, etc.: 32 speech, 6 artificial voice

4) Error data file

Error data are recorded in bytes. Bit positions corresponding to errors become 1. MSBs are recorded first. The data format of error files is described in Appendix 2. The same error data as that used in Speech Groups 1 and 2 are used for artificial voices. A total of 32 error data samples are used.

5) MNRU speech file

MNRU speech is recorded in the same format as source speech. Nine types of speech are supplied. Four talkers talk at Q = 0, 5, 10, 15, 20, 25, 30, 35 and 40 dB. 36 speech samples are recorded in total.

3.3.2 Host laboratory equipment

The host laboratory equipment is composed of a computer for host laboratory equipment, a host laboratory equipment interface board and host laboratory equipment software. Refer to Ref. 1 in Appendix 7.

1) Computer for host laboratory equipment

The following computer and accessories are needed:

NEC personal computer PC 9801RA5

MS-DOS Ver. 5

Hard disk drive

3.5" magneto-optic disk unit

The equipment has only been ascertained to operate with this computer. The users are requested to check compatibility if any other type of computer is to be used. To exchange files with a master speech codec, an interface to allow data exchange with the computer for master speech codec by ftp, etc. will be needed.

2) Host laboratory equipment interface board

The host laboratory equipment interface board and test codec are interfaced by a 25-pin D-sub connector. Appendix 1 shows the signal pin layout, timing and electrical characteristics.

3) Host laboratory equipment software

The supplied software is implemented for operation in the host laboratory equipment computer. Data are input to and output from the test speech codec through the host laboratory equipment interface.

3.3.3 Master speech codec

The master speech codec is composed of a computer and a codec program.

1) Master speech codec computer

The following computer and software are needed:

Sun Microsystems SparcStation series

Sun OS 4.1.3

The equipment has only been ascertained to operate with this computer. The users are requested to check compatibility if any other type of computer is to be used.

2) Master speech codec program

The program is supplied in an executable file for floating point operations in the master speech codec computer.

3.3.4 Listening tape editing equipment

The following equipment will be needed to edit listening tapes for subjective evaluation tests:

1) Listening tape editing computer

The following equipment is recommended:

Compaq DeskPro 486/50

Data Translation DT-2823 (16bits)

MS-DOS Ver. 6

RAM Disk (16MBytes or more)

2) Randomization tables

Randomization tables generated for each listening group mentioned in Sections 3.2.2.5 and 4.2.2.5 are supplied in text files. A total of 20 randomization tables are generated for each listening group, one for each of the 20 listener groups. At least 5 tables are randomly selected to use. Each row in the randomization tables has the following format:

```
Number        Blank        Speech File Name
(4 columns)   (1 column)

  31   m31_f1.dec
 198   m02_00.mnr
 163   f02.pcm
  48 . m09_f2.deh
     .
     .
```

Refer to Section 3.3.1 for speech file naming rules.

The first unit with four columns shows numbers starting with 0. By sorting according to this number, items to be averaged are arranged next to each other. Tables before randomizing are supplied in text files in accordance with the above-mentioned format. For example, files for the Full-Rate Listening Group 1 before randomizing contain the following information:

File Content

0	m01_f1.dec	[Reference System 1]	[Test Item 1]
1	f01_f1.dec		⋮
2	f03_f1.dec		⋮
	⋮		⋮
8	m09_f1.dec		[Test Item 2]
	⋮		⋮
36	b01_f1.dec		[Test Item 8]
	⋮		⋮
40	m01_f2.dec	[Test System 2]	[Test Item 1]
	⋮	⋮	⋮
80	m01_f3.dec	[Test System 3]	[Test Item 1]
	⋮	⋮	⋮
120	m01_f4.dec	[Test System 4]	[Test Item 1]
	⋮	⋮	⋮
160	m01.pcm	[Source Speech]	
	⋮	⋮	
164	m01_40.mnr	[MNRU 40dB]	
	⋮	⋮	
168	m01_35.mnr	[MNRU 35dB]	
	⋮		
196	m01_00.mnr	[MNRU 0dB]	
197	f01_00.mnr		
198	m02_00.mnr		
199	f02_00.mnr		

3) Specification for creating the randomization editing program

The specification for generating listening tapes for subjective evaluation tests based on the randomization table is listed below. The speech sample names designated in each row of the randomization table are read sequentially beginning with the first sample name. The corresponding files are output to the DAT deck through the D/A converter. A one-second silence is inserted between speech samples and a short tone is inserted after every five samples. Silence for one or two minutes is inserted after every 70 samples, enabling listeners to pause.

4) Recording DAT

The following equipment is recommended:

Sony DTC-790

3.3.5 Measurement tools

The measurement tools needed to calculate objective and subjective evaluation values are shown below:

1) Level-correction program

The program will be supplied in a format enabling operation on a Sun workstation.

The program is used to correct decoded speech output signals to a normal level in subjective evaluation tests for level variation.

2) SNRseg, SNRfrq and CD calculation program

The program will be supplied in a format enabling operation on a Sun workstation. The program can also be used to measure the delay time by the method described in Section 3.2.1.6.

3) Opinion-equivalent Q value calculation program

A version running in Sun workstations will be proposed.

The program is used to calculate opinion-equivalent Q values from MOS.

3.3.6 Equipment for listening

The following equipment is recommended:

DA converter	:	Data Translation DT-2823 (16 bits)
Filter	:	Band-pass filters (0.3-3.4 kHz) conforming to ITU-T Recommendation G.714. For example, NF P82 (HPF, 0.2kHz cut-off) and NF P86 (LPF, 3.4 kHz cut-off).
Recording DAT	:	Sony DTC-790
Recording and replay amplifier	:	Panasonic SU-A900
Listening headphone	:	Sennheiser HD-250 Linear II or its equivalent

CHAPTER 4 HALF-RATE SPEECH CODEC (PSI-CELP SYSTEM)

4.1 Quality Recommendation

1) Test and procedure

This section describes requirements for speech quality and delay time in the objective evaluation test and the subjective evaluation test described in Sections 2.1.2 and 2.1.3, respectively.

The objective evaluation test is, based on target characteristics decided in advance by the tester, conducted in accordance with the procedures specified in Fig. 4.1-1 and concluded by judging whether or not the characteristics are satisfied.

The subjective evaluation test is, based on the subjective evaluation requirements, conducted in accordance with the procedures specified in Fig. 4.1-2 and concluded by a pass/fail judgment.

The delay time measurement test is conducted in accordance with the procedures specified in Fig. 4.1-1 or 4.1-2 based on the delay time requirement.

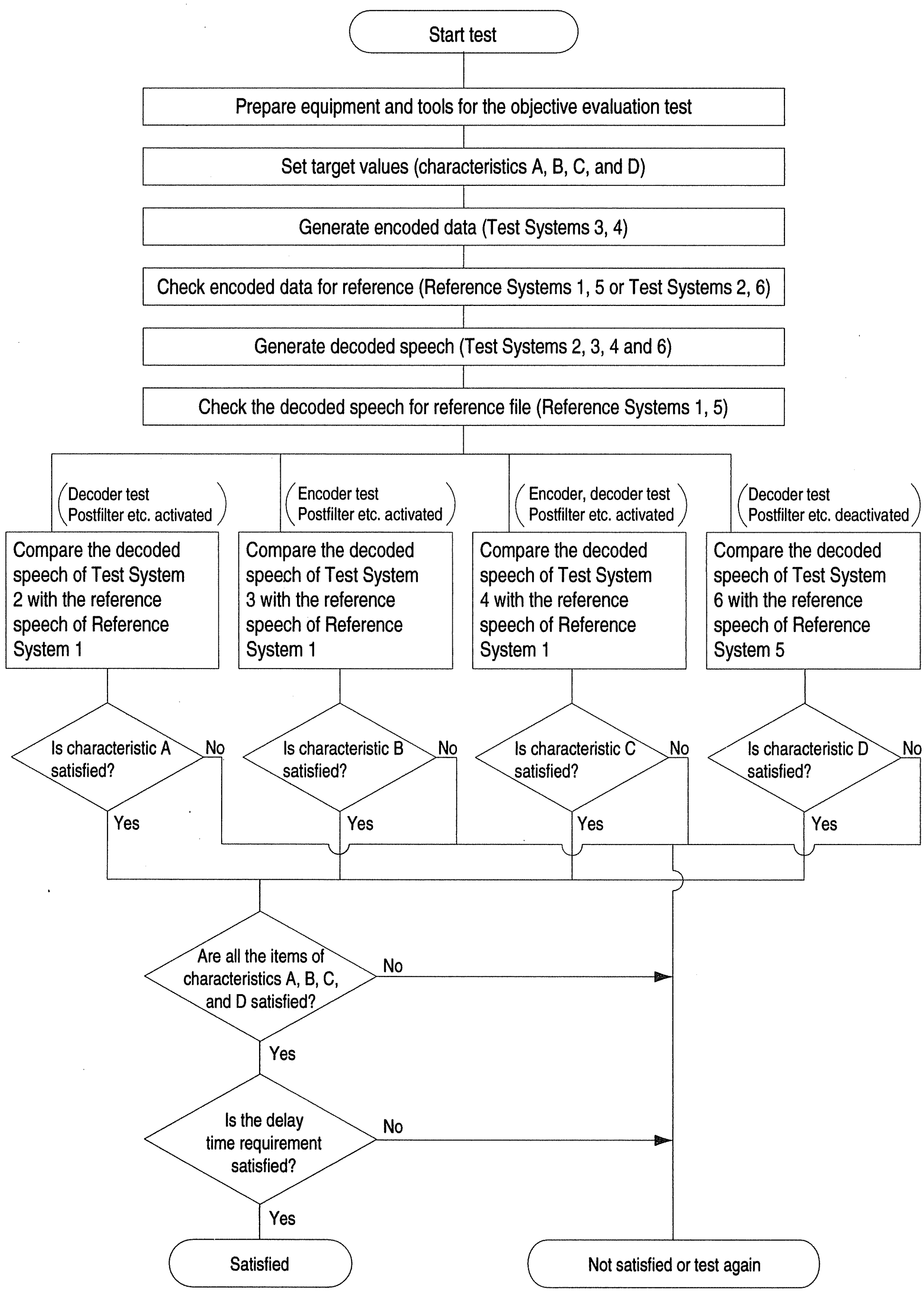


Figure 4.1-1 Procedure for Objective Evaluation Test of Half-Rate Speech Codec

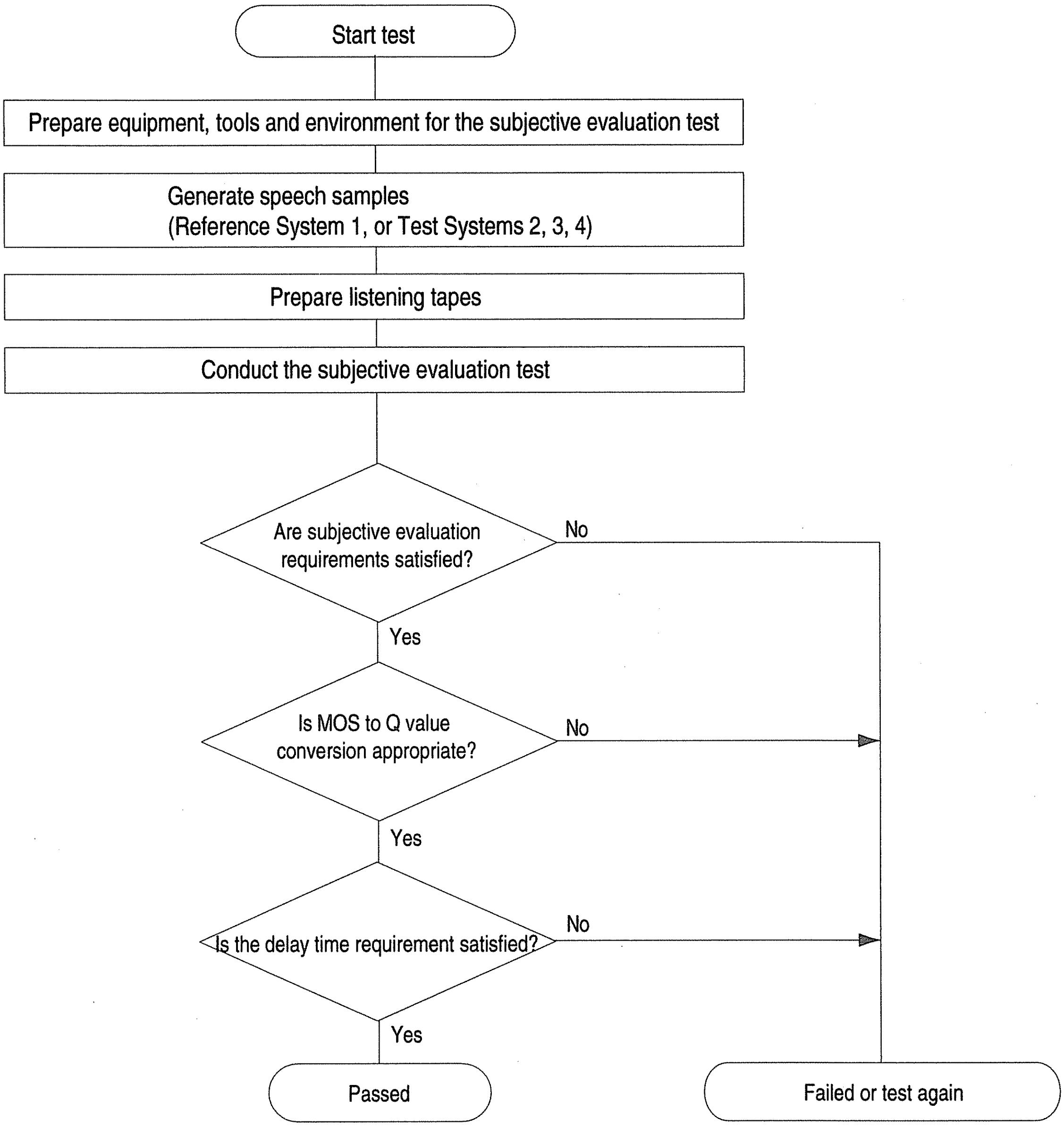


Figure 4.1-2 Procedure for Subjective Evaluation Test of Half-Rate Speech Codec

2) Reference and test systems of speech codec

Reference or test systems : The reference and test systems are numbered as follows in accordance with combinations of master and test speech codecs.

Table 4.1-1 Reference and Test Systems of Half-Rate Speech Codec

Reference System (i) or Test System (j)	Encoder/Decoder	Postfilter, etc*
i=1	Master encoder/Master decoder	Activated
j=2	Master encoder/Test decoder	Activated
j=3	Test encoder/Master decoder	Activated
j=4	Test encoder/Test decoder	Activated
i=5	Master encoder/Master decoder	Deactivated
j=6	Master encoder/Test decoder	Deactivated

* Postfilter, etc. for half-rate speech codecs represents noise canceller (Section 5.2.1.2 in RCR STD-27), low level signal suppression (Section 5.2.1.3), processing for decoder power increase (Section 5.2.4.3), and postfilter (Section 5.2.4.4.6). Processing for the decoder power increase can be inhibited by changing the equation spow_i=rspow_i · 1.44 to spow_i=rspow_i · 1.00 in Table 5.2.4.3-1 of RCR STD-27.

3) Test items for speech codec

Test items : Each item is numbered as follows. For hard decision codec, Test Items 1 to 8 are tested. For soft decision codec, Test Items 1 to 10 are tested.

Table 4.1-2 Test Items for Half-Rate Speech Codec

Test Item (k)	Condition	Hard Decision Speech Codec	Soft Decision Speech Codec
1	Basic Characteristics	√	√
2, 3	Error Robustness (Hard Decision)	√	√
4, 5	Level Variation	√	√
6, 7	Background Noise	√	√
8	Talker Dependency	√	√
9, 10	Error Robustness (Soft Decision)	—	√

4) Explanations of symbols used in this chapter

SNRseg : Segmental SNR (objective evaluation value calculated for each test item). $SNRseg_K$ (j/i) is the mean value of the segmental SNR of the output of Test System j to that of Reference System i in Test Item k.
Section 4.2.1.5 defines the calculation method.

SNRfrq : Low segmental SNR frequency (objective evaluation value calculated for each test item). $SNRfrq_k$ (j/i) is the fraction of effective segments in Test Item k with SNR, measured as the ratio of the output of Test System j to that of Reference System i, of less than 15 dB.
Section 4.2.1.5 defines the calculation method.

CD : Cepstral distance (objective evaluation value calculated for each test item). CD_k (j/i) is the mean value of the Cepstral distance from the output of Test System j to that of Reference System i in Test Item k.
Section 4.2.1.5 defines the calculation method.

MAv : MAv_{k1} is the mean value of subjective evaluation marks when Reference System 1 is applied to Test Item k (MOS value).
Section 4.2.2.10 defines the calculation method.

MVr : MVr_{k1} is the unbiased variance of subjective evaluation marks when Reference System 1 is applied to Test Item k.
Section 4.2.2.10 defines the calculation method.

TAv : TAv_{kj} is the mean value of subjective evaluation marks when Test System j (j = 2, 3, 4) is applied to Test Item k (MOS value).
Section 4.2.2.10 defines the calculation method.

TVr : TVr_{kj} is the unbiased variance of subjective evaluation marks when Test System j (j = 2, 3, 4) is applied to Test Item k.
Section 4.2.2.10 defines the calculation method.

t : t_{kj} is the index value in a significant difference test between MAv_{k1} and TAv_{kj} for Test Item k.
Section 4.2.2.10 defines the calculation method.

4.1.1 Speech quality targets and requirements

4.1.1.1 Objective evaluation target characteristics

As target characteristics, the tester sets thresholds for the objective evaluation parameters SNRseg, SNRfrq and CD. Thresholds are assigned to SNRseg as a lower limit, and to SNRfrq and CD as upper limits.

The objective evaluation target characteristics shown in Figure 4.1-1 are defined as follows:

Characteristic A: The similarity of the test decoder with the master decoder is evaluated.

Test System j = 2 should be compared with Reference System i = 1. The following conditions shall be satisfied for all of the Test Items k = 1 to 8 or k = 1 to 10:

$SNRseg_k$ (2/1) $\geq$ Threshold (lower limit) of $SNRseg_k$ (2/1)

$SNRfrq_k$ (2/1) $\leq$ Threshold (upper limit) of $SNRfrq_k$ (2/1)

CD_k (2/1) $\leq$ Threshold (upper limit) of CD_k (2/1)

The tester enters the thresholds (lower or upper limits) of $SNRseg_k$ (2/1), $SNRfrq_k$ (2/1) and CD_k (2/1) in a free-format form. It is recommended that a column for evaluation values also be included in the form (this applies to B, C and D as well).

Characteristic B: The similarity of combinations of test encoder and master decoder with the master speech codec is evaluated.

Test System j = 3 should be compared with Reference System i = 1. The following conditions shall be satisfied for all of the Test Items k = 1 to 8 or k = 1 to 10:

$SNRseg_k$ (3/1) $\geq$ Threshold (lower limit) of $SNRseg_k$ (3/1)

$SNRfrq_k$ (3/1) $\leq$ Threshold (upper limit) of $SNRfrq_k$ (3/1)

CD_k (3/1) $\leq$ Threshold (upper limit) of CD_k (3/1)

The tester enters the thresholds (lower or upper limits) of $SNRseg_k$ (3/1), $SNRfrq_k$ (3/1) and CD_k (3/1) in a free-format form.

Characteristic C: The similarity of combinations of test encoder and test decoder with the master speech codec is evaluated.

Test System j = 4 should be compared with Reference System i = 1. The following conditions shall be satisfied for all of the Test Items k = 1 to 8 or k = 1 to 10:

$SNRseg_k$ (4/1) $\geq$ Threshold (lower limit) of $SNRseg_k$ (4/1)

$SNRfrq_k$ (4/1) $\leq$ Threshold (upper limit) of $SNRfrq_k$ (4/1)

CD_k (4/1) $\leq$ Threshold (upper limit) of CD_k (4/1)

The tester enters the thresholds (lower or upper limits) of $SNRseg_k$ (4/1), $SNRfrq_k$ (4/1) and CD_k (4/1) in a free-format form.

Characteristic D: The similarity of the test decoder with the master decoder is evaluated.

Test System j = 6 should be compared with Reference System i = 5. The following conditions shall be satisfied for all of the Test Items k = 1, 4, 5 and 8:

$SNRseg_k$ (6/5) $\geq$ Threshold (lower limit) of $SNRseg_k$ (6/5)

CD_k (6/5) $\leq$ Threshold (upper limit) of CD_k (6/5)

The tester enters the thresholds (lower or upper limits) of $SNRseg_k$ (6/5) and CD_k (6/5) in a free-format form.
Section 2 of Appendix 6 shows sample data of the objective evaluation target characteristics test for half-rate speech codec.

4.1.1.2 Subjective evaluation requirements

The subjective evaluation requirements shown in Fig. 4.1-2 are defined as follows:

After comparing with Reference System i = 1, all the Test Systems j = 2 to 4 for all the Test Items k = 1 to 8 should satisfy $t_{kj} \leq 1.645$, where t_{kj} is calculated with a five-per-cent, one-sided significant difference test.

The same criterion shall be applied to the judgment for the Test Items k = 9 and 10 for soft decision speech codec. However, characteristic values of error robustness (hard decision) shall be used as reference values for error robustness (soft decision).

MAv_{k1}, MVr_{k1} in Test Items k = 9 and 10 shall be calculated by using MAv_{k1} and MVr_{k1} from Test Items k = 2 and 3. MOS to opinion-equivalent Q value conversion shall be verified to be appropriate.

4.1.1.3 Delay time requirement

The delay time (D) of test codecs measured by the method specified in Section 4.2.1.6 should be 80 ms or less, excluding the algorithmic delay time.

4.2 Validation Test Methods

4.2.1 Objective evaluation method

4.2.1.1 Definitions

Refer to Section 3.2.1.1.

4.2.1.2 Reference and test systems

Master and test speech codecs are combined in accordance with the reference and test systems shown in Table 4.1-1.

4.2.1.3 Test items

Table 4.2.1.3-1 lists ten test items covering basic characteristics, error robustness (4), level variation (2), background noise (2), and talker dependency. Error robustness shall be judged based on hard decision tests. The two test items of the soft decision test are not required for codec for hard decision only.

Table 4.2.1.3-1 Test Items for Half-Rate Speech Codec

Test	Test Item	Error Condition	Input Level	Superposed Noise	Number of speech samples: Speech	Number of speech samples: Artificial Voice
Basic Characteristics	1	Error-free	Normal	None	16	2
Error Robustness (Hard Decision)	2	1% (4km/h)	Normal	None	4	2
		1% (20km/h)	Normal	None	4	2
		1% (60km/h)	Normal	None	4	2
		1% (60km/h)*	Normal	None	4	2
	3	3% (4km/h)	Normal	None	4	2
		3% (20km/h)	Normal	None	4	2
		3% (60km/h)	Normal	None	4	2
		3% (60km/h)*	Normal	None	4	2
Level Variation	4	Error-free	Normal - 10dB	None	4	2
	5	Error-free	Normal - 20dB	None	4	2
Background Noise	6	Error-free	Normal	Low noise SNR30dB	8	2
	7	Error-free	Normal	High noise SNR15dB	8	2
Talker Dependency	8	Error-free	Normal	None	8	—
Error Robustness (Soft Decision)	9	1% (4km/h)	Normal	None	4	2
		1% (20km/h)	Normal	None	4	2
		1% (60km/h)	Normal	None	4	2
		1% (60km/h)*	Normal	None	4	2
	10	3% (4km/h)	Normal	None	4	2
		3% (20km/h)	Normal	None	4	2
		3% (60km/h)	Normal	None	4	2
		3% (60km/h)*	Normal	None	4	2

Notes 1: Error data marked with an asterisk are due to fading at 1.5 GHz. Those without an asterisk are due to fading at 800 MHz.

2: "Normal" level is the level whose mean level is 21dB lower than the sine-wave full scale of the A/D converter.

3: Superposed noise is defined as follows:
None: Noise which is unavoidably mixed during generation of source speech and quantization noise.
Low noise: SNR 30dB
High noise: SNR 15dB

4.2.1.4 Test preparations

Table 4.2.1.4-1 defines speech samples based on which generation of encoded data and decoded speech is performed.

1) Source speech

Source speech is composed of speech signals and artificial voice signals. Eighty speech signals defined in Section 4.3.1 are used. Ten artificial voice signals defined in Section 4.3.1 are used. In error robustness tests, use the same speech signals for hard and soft decisions.

In the artificial voice columns for the basic characteristics and error robustness tests in Table 4.2.1.4-1, artificial male and female voices are denoted PM 1-9 and PF 1-9, respectively, to distinguish the speech samples. Source speech from one male and one female artificial voice is used in common as source speech for these artificial voice. Artificial voice is not used in the talker dependency test.

The data format of individual speech signals and artificial voice signals is 16 bits (only the most significant 14 bits are valid) linear PCM with a time duration of 6s.

2) Generation of encoded data

For Systems 3 and 4 in Table 4.1-1, encoded data are generated from source speech with the test encoder. The number of encoded data samples is 80 for speech signals and 10 for artificial voice signals, totaling 90.

The encoded data for reference defined in Section 4.3.1 should be used for Systems 1, 2, 5 and 6.

3) Error data

From the 32 error data samples defined in Section 4.3.1, one sample each is used with each speech sample of Groups 1 to 4 in Test Items 2 and 3 in Table 4.2.1.4-1. Use error data for hard decision corresponding to each speech sample when making soft decisions.

Error data of the same test items as those of Group 1 are used with Group 5 regarding error data for artificial voice. Error data of the same test items as those of Group 2 are used with Group 6.

4) Generation of decoded speech

Decoded speech is generated for Systems 2 to 4 and 6 shown in Table 4.1-1 by using the host laboratory equipment and master speech codec defined in Section 4.3. Level-variation speech samples in Test Items 4 and 5 are not returned to the normal level. (Level correction is not performed.) For artificial voice signals of Test Items 1, 2, 3, 9 and 10 (basic characteristics and error robustness), 9 (17) each of male and female decoded speech samples are generated by applying different error data to each of the male and female encoded data samples. The total number of decoded speech samples in Systems 2 to 4 is 106 (154), 80 (112) for speech signals and 26 (42) for artificial voice signals. In System 6, the total is 38, 32 for speech signals and 6 for artificial voice signals. Numbers in parentheses ( ) refer to the case where both hard and soft decisions are used for generating decoded speech.

The corresponding decoded speech samples for reference defined in Section 4.3.1 are used for Systems 1 and 5.

Decoded speech should be regenerated if noticeable alternates are found in audition.

Table 4.2.1.4-1 Speech Samples for Half-Rate Speech Codec

Test		Test Item	Speech				Artificial Voice	
			Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Basic Characteristics		1	M1 F3 M5 F7	F1 M3 F5 M7	M2 F4 M6 F8	F2 M4 F6 M8	PM1	PF1
Error Ro-bust-ness (Hard Deci-sion)	1%	2	M9	F9	M10	F10	PF2	PM2
			F11	M11	F12	M12	PM3	PF3
			M13	F13	M14	F14	PF4	PM4
			F15	M15	F16	M16	PM5	PF5
	3%	3	M17	F17	M18	F18	PF6	PM6
			F19	M19	F20	M20	PM7	PF7
			M21	F21	M22	F22	PF8	PM8
			F23	M23	F24	M24	PM9	PF9
Level Variation		4	M25	F25	M26	F26	PF10	PM10
		5	F27	M27	F28	M28	PM11	PF11
Background Noise		6	M29 F31	F29 M31	M30 F32	F30 M32	PF12	PM12
		7	M33 F35	F33 M35	M34 F36	F34 M36	PM13	PF13
Talker Dependency		8	B1 G3	G1 B3	B2 G4	G2 B4	—	—
Error Ro-bust-ness (Soft Deci-sion)	1%	9	M9	F9	M10	F10	PF2	PM2
			F11	M11	F12	M12	PM3	PF3
			M13	F13	M14	F14	PF4	PM4
			F15	M15	F16	M16	PM5	PF5
	3%	10	M17	F17	M18	F18	PF6	PM6
			F19	M19	F20	M20	PM7	PF7
			M21	F21	M22	F22	PF8	PM8
			F23	M23	F24	M24	PM9	PF9

4.2.1.5 Speech quality measurement method

Using decoded speech for reference from System 1 or 5 shown in Table 4.1-1 as a reference, segmental SNRs (SNRseg), low segmental SNR frequencies (SNRfrq) and Cepstral distances (CD) of speech decoded by one of Systems 2 to 4, or System 6, are obtained for individual speech samples.

Refer to Section 3.2.1.5 for calculations of segmental SNR (SNRseg), low segmental SNR frequency (SNRfrq) and Cepstral distance (CD).

4.2.1.6 Delay time measurement method

Refer to Section 3.2.1.6. Use the following values for the delay time of a master speech codec. Note that the values are different from those for full-rate codecs.

D_m = 80 ms

D_{me} = 40 ms
D_{md} = 40 ms

4.2.1.7 Analysis, judgment and recording

Calculate the mean values of the segmental SNR, low segmental SNR frequency and Cepstral distance for each test item shown in Table 4.2.1.3-1 and record the values. Also, record measurements of the delay time. (In both cases, use a free-format form to record the values.) Evaluation values of objective evaluation characteristics are rounded off to the nearest two decimal places and then entered. Judge whether or not these mean values satisfy the target values and the delay times satisfy the requirements specified in Section 4.1.1.3.

4.2.2 Subjective evaluation method

4.2.2.1 Definitions

Refer to Section 3.2.2.1.
Follow Section 4.2.1.6 for the delay time measurement method.

4.2.2.2 Test preparations

Prepare the following:

1) Decoded speech

Decoded speech is generated by combining Reference System 1 and Test Systems 2 to 4 shown in Table 4.1-1.

2) Reference voice

The following ten types are used as reference voice. MNRU speech is used to verify that the subjective evaluation test is correct and to compare test results from different test organizations by converting MOS values to opinion-equivalent Q values.

(a) Source speech

(b) MNRU speech (9 types: Q = 0, 5, 10, 15, 20, 25, 30, 35 and 40 dB)

4.2.2.3 Test items

Table 4.2.1.3-1 lists the test items. The test items consist of basic characteristics, error robustness, level variation, background noise, and talker dependency, totaling ten. Error robustness shall be judged based on hard decision tests. The two test items of soft decision tests are not required for codec for hard decision only.

4.2.2.4 Speech samples

Speech samples of Groups 1 to 4 in Table 4.2.1.4-1 are generated with a postfilter and other devices using the equipment and tools defined in Section 4.3 by the four decoded speech reference and test systems specified in Section 4.2.2.2. (Artificial voices are not generated.) Test items only related to the correction of soft-decision errors are omitted for hard-decision-only codecs. Level-variation voices in Test Items 4 and 5 are returned to the normal level mentioned in Section 4.2.1.3 using the level correction program described in Section 4.3.5 after they have been generated. Use a total of 40 voices as reference voice generated using source speech files of two male and two female speakers from among the source speech files used to generate decoded speech in accordance with Section 4.2.2.2.

Therefore, the number of speech files M for codecs for hard decision only will be:

M = 80 x 4 (decoded speech) + 40 (reference voice) = 360

The number of speech files for codecs requiring test items for both hard and soft decisions will be:

M = 112 x 4 (decoded speech) + 40 (reference voice) = 488

4.2.2.5 Randomization

As shown in Table 4.2.1.4-1, decoded speech can be divided into four groups. Taking the ratio with reference voice into consideration, the group of Groups 1 and 2 with reference voice added is named Listening Group 1. Similarly, the group of Groups 3 and 4 with reference voice added is named Listening Group 2. The number of speech samples N of one listening group requiring only hard decision test items will be:

N = (20 + 20) x 4 (decoded speech) + 40 (reference voice) = 200

If test items for both hard and soft decisions are required:

N = (28 + 28) x 4 (decoded speech) + 40 (reference voice) = 264

(x4 is speech codec combinations)

The order of listening speech samples inside listening groups is randomized for each listening group using the randomization tables (refer to Section 4.3.4). Different randomization should be used for each listener group (refer to Section 4.2.2.7) and the listening sequence of the listening groups should be changed for each listener group.

4.2.2.6 Listening environment

Refer to Section 3.2.2.6.

4.2.2.7 Listeners

Refer to Section 3.2.2.7.

4.2.2.8 Test procedure

Before starting the test, distribute the following sheet to the subjects (listeners) and explain the test purpose and guidelines for procedures and scoring.

About Listening Test
Thank you for cooperating with us today in this test. You will now hear several examples of telephone sounds. Please evaluate the individual sounds, based on how you would feel if you used them in a telephone conversation, according to the following five-level scale: 5 Excellent 4 Good 3 Fair 2 Poor 1 Bad You will soon hear examples of telephone sounds. Evaluate the sounds by selecting one of the five levels. We will now start the test.

The listening conditions will be:

- One listener will listen to all speech samples.
- Speech samples are presented at intervals of 1s. Therefore, if 6s speech samples are used, the listening time per listening group will be:

(6 + 1) x 200 = 1400s = about 24 min.

provided the number of speech samples per listening group N is 200. If the number of speech samples N of one listening group is 264, the listening time per listening group will be:

(6 + 1) x 264 = 1848s = about 31 min.

- One listening group is listened to continuously. Between each session of 70 speech samples (about 8 min.), listeners take a break for 1-2 minutes without leaving the room. A break of more than 20 minutes should be provided between the listening groups.

When beginning the test, present about ten training samples which are suitably selected to include various conditions from the speech samples to let the listeners practice how to mark their evaluation, but they do not have to be informed about this.

4.2.2.9 Delay time measurement method

Use the method specified in Section 4.2.1.6.
If measurements of the objective evaluation test can be applied as they are, it is allowed to use them without performing an additional measurement.

4.2.2.10 Analysis, judgment and recording

Conduct a significant difference test of the mean subjective evaluation values for test codecs and the master codec for each test item shown in Table 4.2.1.3-1. Judge whether or not these mean values satisfy the subjective evaluation requirements specified in Section 4.1.1.2. Enter the subjective evaluation values (MAv_{k1} TAv_{kj}), unbiased variances (MVr_{k1}, TVr_{kj}), index values (t_{kj}) and judgments in Table 4.2.2.10-1. Values of MAv_{21} and MVr_{21} (k = 2) should be entered in the columns of MAv_{91} and MVr_{91} (k = 9), respectively. Values of MAv_{31} and MVr_{31} (k = 3) should be entered in the columns of MAv_{101} and MVr_{101} (k = 10), respectively. Enter "Passed" in the Judgment column if the index value (t_{kj}) satisfies the requirements specified in Section 4.1.1.2. If not, enter "Failed."

If the conversion error of MOS-opinion-equivalent Q value described in Section 3.2.2.10 is larger than the designated value, the subjective evaluation test is considered not relevant and is reconducted.

Refer to Section 3.2.2.10 for the significant difference test of mean values and the conversion from MOS values to opinion-equivalent Q values.

Judge whether or not the delay time measurements satisfy the requirements specified in 4.1.1.3 and record the judgment in a free-format form.

Table 4.2.2.10-1 Half-rate Subjective Evaluation Requirements and Judgments

Test Item	Reference System i = 1		Test System j = 2				Test System j = 3				Test System j = 4			
(k)	MAv_{k1}	MVr_{k1}	TAv_{k2}	TVr_{k2}	t_{k2}	Judgment	TAv_{k3}	TVr_{k3}	t_{k3}	Judgment	TAv_{k4}	TVr_{k4}	t_{k4}	Judgment
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														

4.3 Test Equipment and Tools

This section describes the test equipment and tools for the half-rate speech codec system (PSI-CELP system) validation tests. The equipment needed to obtain test data is as shown in section 3.3.

4.3.1 Test data

Test data consist of files with stored source speech, encoded data for reference, decoded speech for reference, error data and MNRU speech. Data are supplied on a 3.5" magneto-optic disk. Refer to Appendix 5 for speech file naming rules.

1) Source speech file

Refer to 1) in Section 3.3.1.

2) Encoded data for reference file

The following encoded data samples for reference are supplied:

With postfilter, etc. : 112 speech, 10 artificial voice

Without postfilter, etc. : 32 speech, 6 artificial voice

3) Decoded speech for reference file

The file stores decoded speech in the same format as that of the source speech. The following decoded speech samples are supplied:

With postfilter, etc. : 112 speech, 42 artificial voice

Without postfilter, etc. : 32 speech, 6 artificial voice

4) Error data file

The format of the error data files is shown in Appendix 3. A total of 32 samples of error data are used.

5) MNRU speech file

Refer to 5) in Section 3.3.1.

4.3.2 Host laboratory equipment

Refer to Section 3.3.2.

4.3.3 Master speech codec

The master codec is composed of a computer for a master codec and master codec program.

1) Computer for a master codec

The following computer and software are needed:

Sun Microsystems SparcStation series

Sun OS 4.1.3

The equipment has only been ascertained to operate with this computer. The users are requested to check compatibility if any other type of computer is to be used.

2) Master speech codec program

The program is supplied in an executable file for fixed point operations in the master codec computer.

4.3.4 Listening tape edit equipment

Refer to Section 3.3.4.

4.3.5 Measuring tools

Refer to Section 3.3.5.

4.3.6 Equipment for listening

Refer to Section 3.3.6.

(Intentionally blanked)

CHAPTER 5 STANDARD TEST TOOLS

5.1 Definitions

The recommended quality validation test requires constructing a test system by using the test equipment and tools described in Sections 3.3 and 3.4. In addition to commercially available products, standard tools are used as test equipment and tools in order to obtain strict evaluation results in the quality validation tests. The standard test tools are produced and supplied for standardization purposes by organizations involved in the preparation of the Standard. Copyrights and licensing rights belong to the owners (including original copyright owners) listed in Section 5.2. (The inclusion of original copyright owners also applies to the rest of Ch. 5.) Those who desire to conduct quality validation tests can obtain these standard test products from the owners (the original copyright owners) after completing the required formalities.

5.2 List and Owners

The standard test tools used in the quality validation tests consist of the host laboratory equipment interface board and its reference manual (the standard test hardware), and standard programs, standard data and their reference manuals (the standard test software).

Table 5.2-1 lists the standard test hardware and its owners. Tables 5.2-2 and 5.2-3 list the standard test software and its owners.

Table 5.2-1 Standard Test Hardware (Full-Rate and Half-Rate)

Standard Hardware	Owner (Copyright Owner and Licensor)
Full- and half-rate host laboratory equipment interface boards (See Section 3.3.2)	NTT Mobile Communications Network

Note: Parties interested in obtaining these tools should make an enquiry to ARIB.

Table 5.2-2 Standard Test Software (Full Rate)

Standard Software	Owner (Copyright Owner and Licensor)
Full-rate test data (See Section 3.3.1.)	
1) Source speech files	NTT Mobile Communications Network & NTT Corp. (in part)
2) Files of encoded data for reference 3) Files of decoded speech for reference	NTT Mobile Communications Network/ Nippon Motorola
4) Error data files 5) MNRU speech files	NTT Mobile Communications Network
Full-rate host laboratory equipment software (See Section 3.3.2.)	NTT Mobile Communications Network
Full-rate master codec program (See Section 3.3.3.)	Motorola Inc. & Nippon Motorola
Full-rate randomization table files (See Section 3.3.4.)	Mitsubishi Electric
Full-rate measuring tools (See Section 3.3.5.)	
1) Level change program	NTT Mobile Communications Network
2) SNRseg, SNRfrq and CD calculation program	NTT Corp.
3) Opinion-equivalent Q value calculation program	Nippon Motorola

Note: Parties interested in obtaining these tools should make an enquiry to ARIB.

Table 5.2-3 Standard Test Software (Half-Rate)

Standard Software	Owner (Copyright Owner and Licensor)
Half-rate test data (See Section 4.3.1.)	
1) Source speech files	NTT Mobile Communications Network & NTT Corp. (in part)
2) Files of encoded data for reference 3) Files of decoded speech for reference	NTT Mobile Communications Network
4) Error data files 5) MNRU speech files	NTT Mobile Communications Network
Half-rate host laboratory equipment software (See Section 4.3.2.)	NTT Mobile Communications Network
Half-rate master codec program (See Section 4.3.3.)	NTT Mobile Communications Network
Half-rate randomization table files (See Section 4.3.4.)	Mitsubishi Electric
Half-rate measuring tools (See Section 4.3.5.)	
1) Level change program	NTT Mobile Communications Network
2) SNRseg, SNRfrq and CD calculation program	NTT Corp.
3) Opinion-equivalent Q value calculation program	Nippon Motorola

Note: Parties interested in obtaining these tools should make an enquiry to ARIB.

5.3 Rights and Obligations for Use

The listed software among the standard test tools for use in the validation test of the standard tecnical characteristics shall be supplied only for the purpose of carrying out the validation test. Those who so desire can obtain the standard test software after agreeing on the conditions with the relevant software owners and completing the designated procedures such as entering into a software licensing agreement. Unauthorized utilization of the supplied materials for other purposes or by third parties is prohibited.

(Intentionally blanked)

CHAPTER 6 CS-ACELP SPEECH CODEC

6.1 Overview

This chapter describes the validation testing method for the CS-ACELP Speech codec defined in RCR STD-27, Section 5.3 "CS-ACELP Speech codec". Validation tests consist of bit exact verification test, including delay time measurement and subjective evaluation. In the tests defined in this section, the software attached with ITU-T Recommendation G.729 shall be used for CS-ACELP reference codec. Thus, host laboratory equipment interfaces shall not be used.

6.2 Technical validity test

(1) Detailed test items and procedures

In the validation tests, bit exact verification test shall be conducted first. Subjective evaluation test shall then be conducted only if the test codec is determined to be conformable by bit exact verification test.

Bit exact verification test shall be conducted by the procedure stipulated in Fig. 6.2-1 and the conformance of the test codec shall be determined based on the bit exactness requirements in conjunction with the result of delay time measurement.

Delay time measurement shall be conducted by the procedure stipulated in Fig. 6.2-1 and the result shall be evaluated based on the delay time requirement.

Subjective evaluation test shall be conducted by the procedure stipulated in Fig. 6.2-2 and result (pass or fail) shall be concluded based on subjective evaluation requirements.

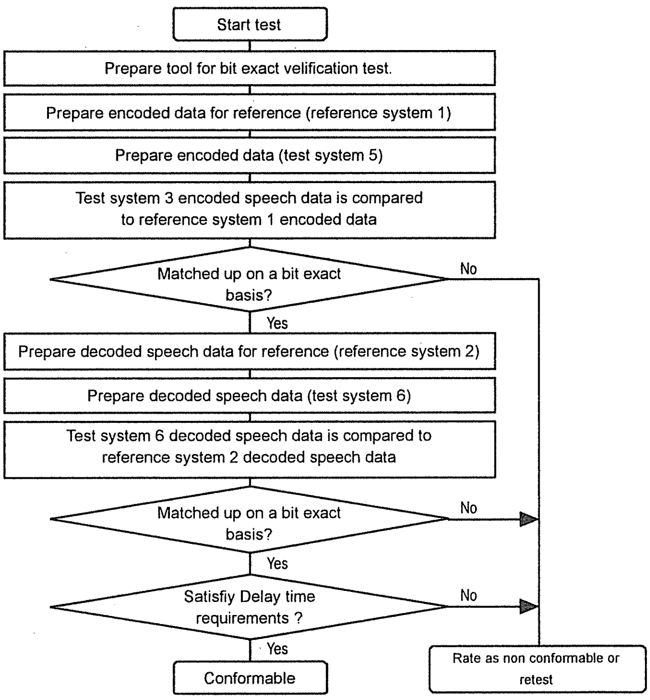


Fig. 6.2-1 Bit exact verification test procedure for CS-ACELP speech codec

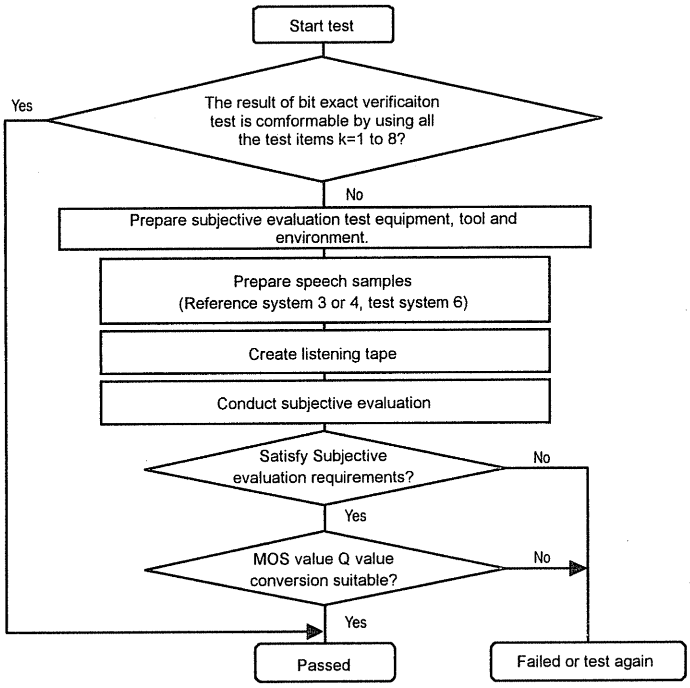


Fig. 6.2-2 Subjective evaluation test procedure for CS-ACELP speech codec.

(2) Speech codec reference systems and test systems

Reference system or test system: Reference systems and test systems shall be numbered as listed in Table 6.2-1 according to the combinations of reference speech codec and test speech codec.

Table 6.2-1 Reference/test systems for CS-ACELP speech codec

Reference system (i) or Test system(j)	Encoder/Decoder
i = 1	CS-ACELP reference encoder
i = 2	CS-ACELP reference encoder/CS-ACELP reference decoder.
i = 3	Full-rate reference encoder/full-rate reference decoder
i = 4	Half-rate reference encoder/half-rate reference decoder
j = 5	Test encoder
j = 6	Test encoder/test decoder

(Note) Channel encoding and decoding of the CS-ACELP reference codec shall be identical to that are applied to test codec.

(3) Speech codec test items

Test items: Each test item shall be numbered as listed in Table 6.2-2.

Table 6.2-2 Test items for CS-ACELP speech codec

Test item (k)	Requirements
1	Basic characteristics
2	Error robustness
3	
4	Level variation
5	
6	Background noise
7	
8	Talker dependency

6.2.1 Target speech quality and requirements

6.2.1.1 Bit exact requirements

Bit exactness means that the output bitstream of the test codec is identical to the output bitstream of the reference codec .

The bit exact requirements in Fig. 6.2-1 are defined below:

Encoded data: Verify that the output bitstream of the test encoder is identical to the output bitstream of the CS-ACELP reference encoder.
Bit exactness of encoded data shall be confirmed for all the test items, k=1 to 8, by using Test System j=5 and Reference System i=1 for comparison. In this test, the channel encoding part can be omitted and the output bitstream of the speech encoder can be used for comparison.

Decoded speech data: Verify that the output bitstream of the test decoder is identical to the output bitstream of the CS-ACELP reference decoder.
Bit exactness of decoded speech data shall be confirmed for all the test items, k=1 and 4 to 8, by using Test System j=6 and Reference System i=2 for comparison. In addition, bit exactness shall also be confirmed for test items k=2 and 3 if concealment of frame erasures (section 5.3.4.3 of STD-27) is identical to that defined in ITU-T Recommendation G.729.

6.2.1.2 Subjective evaluation requirements

The subjective evaluation requirements in Fig. 6.2-2 shall conform to section 5.1 of Appendix 8. Reference System i=3 or 4 and Test System j=6 shall be used.

6.2.1.3 Delay time requirement

If the test codec is implemented using an LSI, delay time excluding algorithmic delay (including interleaving delay) shall be less than or equal to 20ms. If the test codec is implemented by means of software, delay time requirement is not stipulated.

6.3 Validation test implementation method

6.3.1 Bit exact verification test procedure

6.3.1.1 Definition

In this test, speech encoding and decoding functions that are mandatory shall be evaluated by the verification of bit exactness of output bitstream of the encoder and decoder, respectively. Delay time of the test codec shall also be measured in this test.

6.3.1.2 Reference system and test system

Reference codec and test codec shall be set according to the reference and test systems listed in Table 6.2-1. Reference system i=1 and i=2 and Test system j=5 and j=6 shall be used in this test.

6.3.1.3 Test items

Test items is listed in Table 6.3.1.3-1. There are 8 test items in total, which consist of basic characteristics, error robustness, level variation, background noise, and talker dependency.

Table 6.3.1.3-1 Test items for CS-ACELP speech codec

Test	Test item	Error condition	Input level	Additional noise	Number of speech samples
Basic characteristics	1	Error-free	Normal	None	16
Error robustness	2	1% (4km/h)	Normal	None	4
		1% (20km/h)	Normal	None	4
		1% (60km/h)	Normal	None	4
		1% (60km/h)*	Normal	None	4
	3	3% (4km/h)	Normal	None	4
		3% (20km/h)	Normal	None	4
		3% (60km/h)	Normal	None	4
		3% (60km/h)*	Normal	None	4
Level variation	4	Error-free	Normal -10dB	None	4
	5	Error-free	Normal -20dB	None	4
Background noise	6	Error-free	Normal	Low noise SNR30dB	8
	7	Error-free	Normal	High noise SNR15dB	8
Talker dependency	8	Error-free	Normal	None	8

(Notes)

(1) Error data marked with an "*" results from fading at 1.5GHz; error data without an "*" results from fading at 800MHz.

(2) "Normal" is the level at which the level is 21dB lower than the sine-wave full scale of the A/D converter.

(3) Additional noise is defined as follows:

None: Noise that is unavoidably mixed during generation of source speech and quantization noise.

Low noise: SNR30dB

High noise: SNR15dB

6.3.1.4 Preparation

Encoded data and decoded speech data are generated using the speech samples defined in Table 6.3.1.4-1.

(1) Source materials

Source materials shall be consists of speech signal only. The data format for speech signal shall be binary 16-bit linear PCM with time duration of 6 seconds or longer. Japanese should be used in principle. When using other language, the used language shall be indicated specifically.

(2) Generation of encoded data

Encoded data for Reference system i=1 and Test system j=5 defined in Table 6.2-1 shall be generated from the source materials using the CS-ACELP reference encoder and test encoder, respectively. These encoded data shall be used for the decoder input of Reference system i=2 and Test system j=6. The total number of encoded data is 80 for each system.

(3) Error data

32 error data defined in Section 6.4.1 (2) shall be used for test item k=2 and 3 defined in Table 6.2.1.4-1. Unique error data shall be applied for each encoded data.

(4) Generation of decoded speech data

Decoded speech data for Reference system i=2 and Test system j=6 defined in Table 6.2-1 shall be generated from the encoded data using the CS-ACELP reference decoder and the test decoder, respectively. The level of the decoded speech data of the test item k=4 and 5 (level variation) shall not be compensated to the normal level. The total number of encoded data is 80 for each system.

Table 6.3.1.4-1 Speech samples for CS-ACELP speech codec

Name of test			Test item	Variety and number of speech samples
Basic characteristics			1	8 male 8 female
Error robustness	1%	4km/h	2	2 male 2 female
		20km/h		2 male 2 female
		60km/h		2 male 2 female
		60km/h*		2 male 2 female
	3%	4km/h	3	2 male 2 female
		20km/h		2 male 2 female
		60km/h		2 male 2 female
		60km/h*		2 male 2 female
Level variation			4	2 male 2 female
			5	2 male 2 female
Background noise			6	4 male 4 female
			7	4 male 4 female
Talker dependency			8	4 boy 4 girl

6.3.1.5 Bit exact verification method

The encoded data of Test system j=5 and the decoded speech data of Test system j=6 shall be compared to the encoded data of Reference system i=1 and the decoded speech data i=2, respectively.

6.3.1.6 Delay time measurement method

The delay time measurement method is left to the tester.

6.3.1.7 Judgment and recording

The result of bit exact verification for each speech samples listed in Table 6.3.1.4-1 shall be recorded. The result of delay time measurement shall also be recorded. Finally, it shall be evaluated if the encoded data and decoded speech data of the test codec fulfill the requirements defined in section 6.2.1.1 and if delay of the test codec fulfills the requirement defined in section 6.2.1.3.

6.3.2 Subjective evaluation method

The subjective evaluation method shall comply with the method defined in section 5 of Annex 8. "Guidelines for Standardization of Speech Codec for Personal Digital Cellular Telecommunication System". Reference system i=3 or 4 and Test system j=6 shall be used for this test. If the test codec is determined to be conformable by the bit exact verification test, the test codec can be automatically categorized as "passed". Thus, subjective evaluation test can be omitted.

6.4 Test equipment and tools

This section describes the test equipment and tools for the CS-ACELP speech codec validation test.

6.4.1 Test data

(1) Source speech file

The source speech file in section 3.3.1 (1) can be used.

(2) Error data file

The format of error data file is defined in Annex 3. Note that 112 words correspond to 20ms encoded data. The total number of error data is 32. If the test codec adopts hard decision only, error data in section 3.3.1 (4) can be used in stead of above.

(3) MNRU speech file

The MNRU speech file in section 3.3.1 (5) can be used.

6.4.2 Reference speech codec

The software attached with ITU-T Recommendation G.729 shall be used for reference CS-ACELP codec.

6.4.3 Listening tape edit equipment

Refer to section 3.3.4.

6.4.4 Measurement tools

The measurement tools described in section 3.3.5 can be used.

6.4.5 Listening equipments

Refer to section 3.3.6.

(Intentionally blanked)

CHAPTER 7 ENHANCED FULL-RATE SPEECH CODEC (ACELP SYSTEM)

7.1 Standard technical characteristics

(1) Detailed tests and procedure

This section describes the requirements for speech quality and delay time for the objective evaluation test for the ACELP codec defined in RCR STD-27. The ACELP codec is a standard which stipulates bit-exactness, thus, the objective evaluation test is performed only for verifying bit-exactness and the subjective evaluation test is not required.

7.1.1 Objective evaluation test

The objective evaluation test is performed by a method using the test vector defined in RCR STD-27, section 5.4.8. The output sequence which corresponds to the input test sequence for all encoders and decoders shall guarantee bit exactness with the corresponding reference output sequence.

The following tables list the correspondence between the input test sequence and reference output sequence for the objective evaluation test for encoders and decoders. Refer to RCR STD-27, section 5.4.8 for details on the respective test vectors. Each test vector is provided in the form of CDROM as attachment to RCR STD-27.

Table 7.1.1-1 Correspondence of test sequence for speech encoders

Input test sequence	Reference output sequence	Remarks
Tstseq1.inp	Tstseq1.cod	For LPC quantized codebook test
Tstseq2.inp	Tstseq2.cod	For LTP (adaptive) codebook test
Tstseq3.inp	Tstseq3.cod	Female speech signal
Tstseq4.inp	Tstseq4.cod	Male speech signal
Tstseq5.inp	Tstseq5.cod	DTMF and single tone
Tstseq6.inp	Tstseq6.cod	For VOX parameter test

Table 7.1.1-2 Correspondence of test sequence for speech decoders

Input test sequence	Reference output sequence	Remarks
Tstseq1.cod	Tstseq1.out	For LPC quantized codebook test
Tstseq2.cod	Tstseq2.out	For LTP (adaptive) codebook test
Tstseq3.cod	Tstseq3.out	Female speech signal
Tstseq4.cod	Tstseq4.out	Male speech signal
Tstseq5.cod	Tstseq5.out	DTMF and single tone

7.1.2 Delay time requirements

If the test codec is incorporated in the form of LSI, delay time, excluding the algorithm delay (which includes interleave delay), shall be 20ms or less. If the test codec is implemented in the form of software, the delay time requirement shall not be specifically specified.

(Intentionally blanked)

Appendix 1 Specification for Host Laboratory Equipment Interface

1. Scope

This specification applies to both full- and half-rate speech codecs.

2. Electrical Level

The electrical level of each signal is RS-422.

3. Connector

25-pin D-SUB connectors shall be used. The cable and the interface board sides shall be male and female, respectively.

The pin numbers of the signals are allocated as follows:

	+	−
• Clock sync	2	14
• Word sync	3	15
• Frame sync	4	16
• Source speech data	6	18
• Encoded data	11	23
• Encoded data with error	7	19
• Decoded speech data	10	22
• GND	1, 20, 21	

4. Other Notes

- Encoder and decoder delays must together not exceed 12 frames.
- The encoder and decoder should be able to detect reset frames and initialize their internal states.

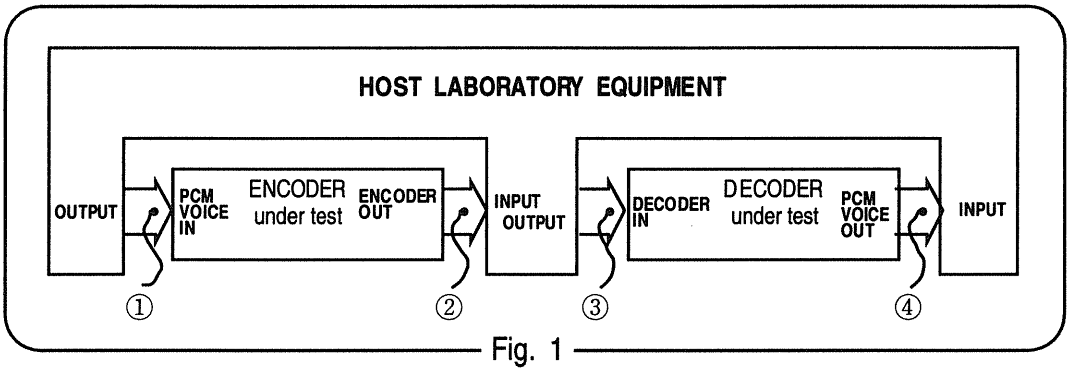
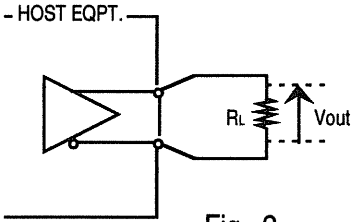


Fig. 1

1. GENERAL (Common to ① – ④ in Fig.1)

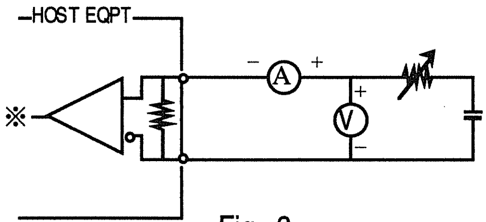
i) RS-422 OUTPUT LEVEL of HOST LABORATORY EQUIPMENT



Vout is guaranteed to be more than +2 Volts,
or less than -2 Volts,
when RL = 100 Ω ±10%.

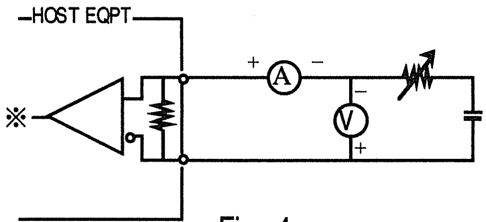
Fig. 2

ii) RS-422 INPUT LEVEL of HOST LABORATORY EQUIPMENT



When V is 200 mV,
{ A should be less than 2 mA.
output (※) should be logical "1".

Fig. 3



When V is 200 mV,
{ A should be less than 2 mA.
output (※) should be logical "0".

Fig. 4

iii) OUTPUT TIMING of HOST LABORATORY EQUIPMENT (Guaranteed)

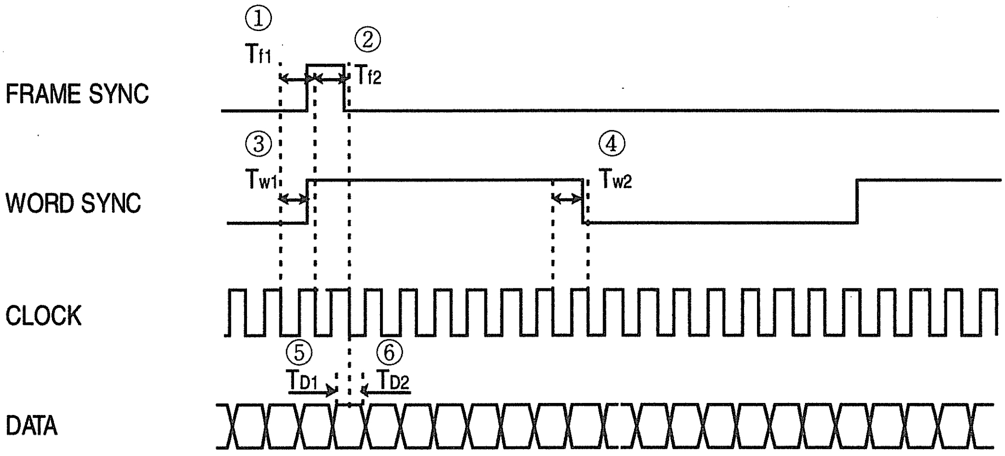


Fig. 5

① Rising edge of FRAME SYNC .... After the previous falling edge of the CLOCK, before the next falling edge of the CLOCK

② Falling edge of FRAME SYNC .... ditto.

③ Rising edge of WORD SYNC................. ditto

④ Falling edge of WORD SYNC................ ditto

⑤ Minimum stable period of DATA
before the falling edge of CLOCK .... More than 1.0 μs.
$(3.9\mu s = \frac{1}{2} \times \frac{1}{128{,}000} s)$

⑥ Minimum stable period of DATA
after the falling edge of CLOCK .... More than 1.0 μs.

iV) INPUT TIMING to HOST LABORATORY EQUIPMENT (Requirement for CODEC)

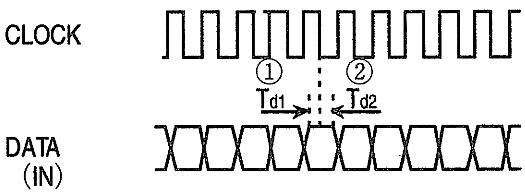


Fig. 6

① Minimum stable period of DATA
before the falling edge of CLOCK .... More than 1.0 μs.

② Minimum stable period of DATA
after the falling edge of CLOCK .... More than 1.0 μs.

2. SIGNAL FORMAT

Refer to Figs. 7 and 8 (following pages)

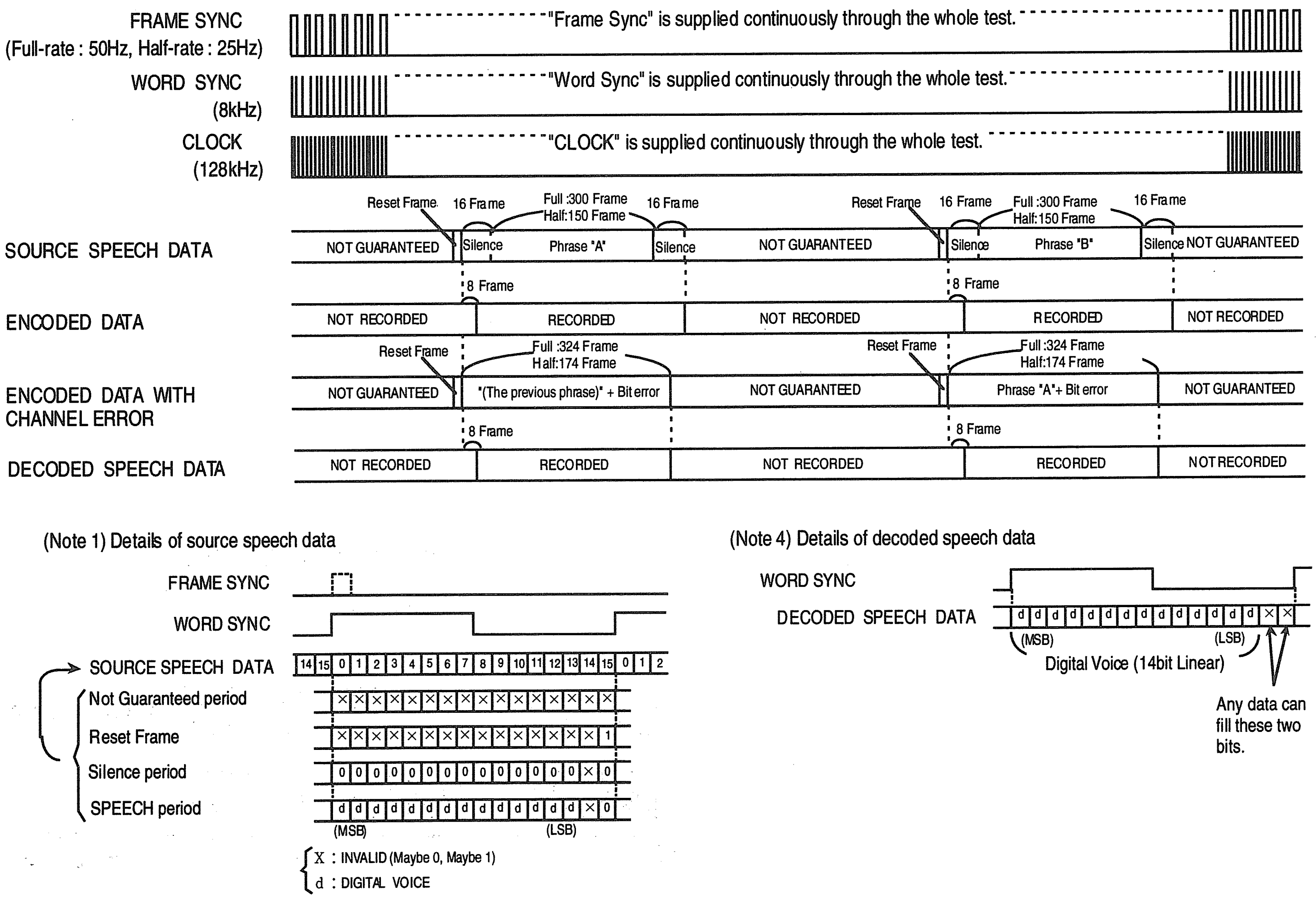
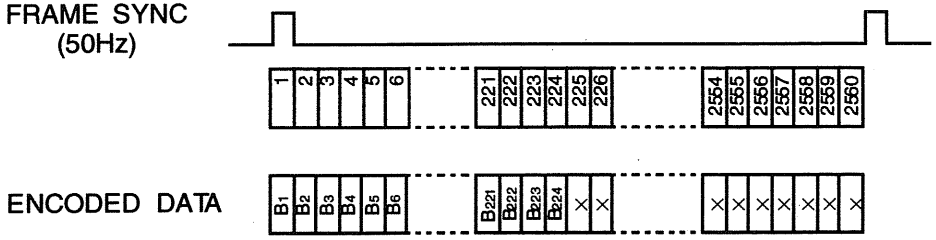


Fig. 7 Signal Format (1/5)

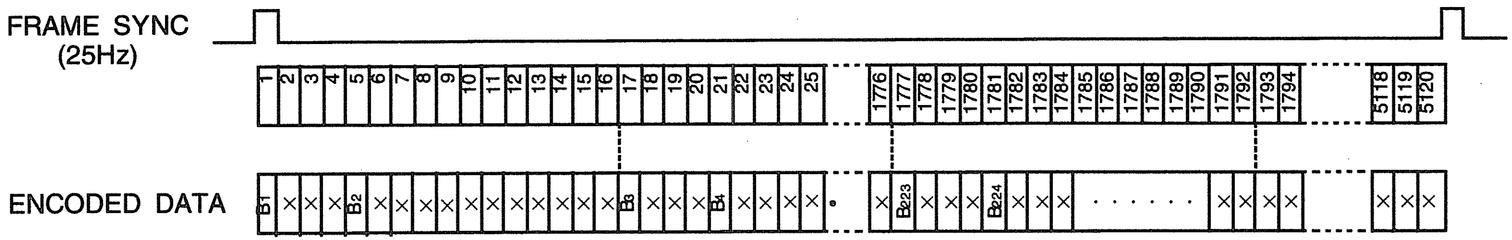
(Note 2-1) Details of encoded data



× : Don't Care

Fig. 7 Signal Format (2/5)

(Note 2-2) Details of encoded data (Half-rate)



B1 - B224 : ENCODED BITS IN ONE FRAME × : Don't Care

Fig. 7 Signal Format (3/5)

(Note 3-1) Details of encoded data with channel error (Full-rate)

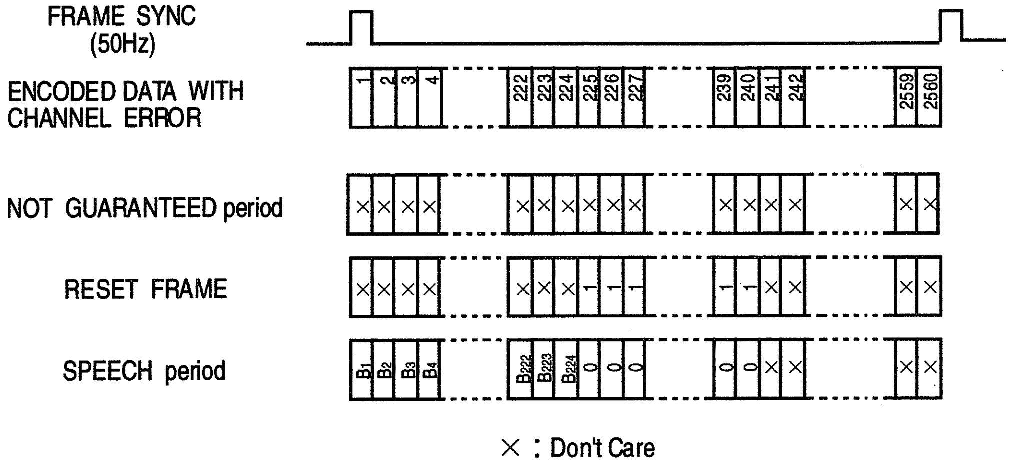
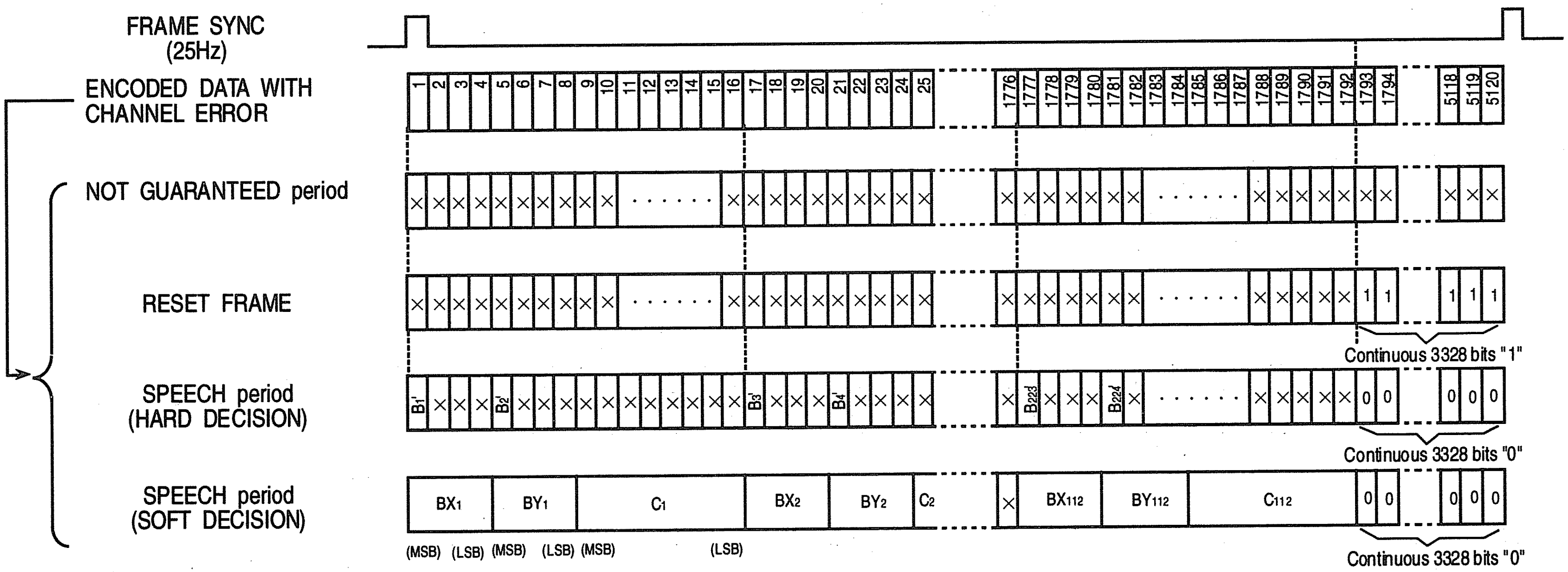


Fig. 7 Signal Format (4/5)

(Note 3-2) Details of encoded data with channel error (Half-rate)



X : INVALID (May be 0, May be 1)

B'_{2n-1} : ENCODED BIT (B_{2n-1}) $\oplus$ bX_n (MSB_1)

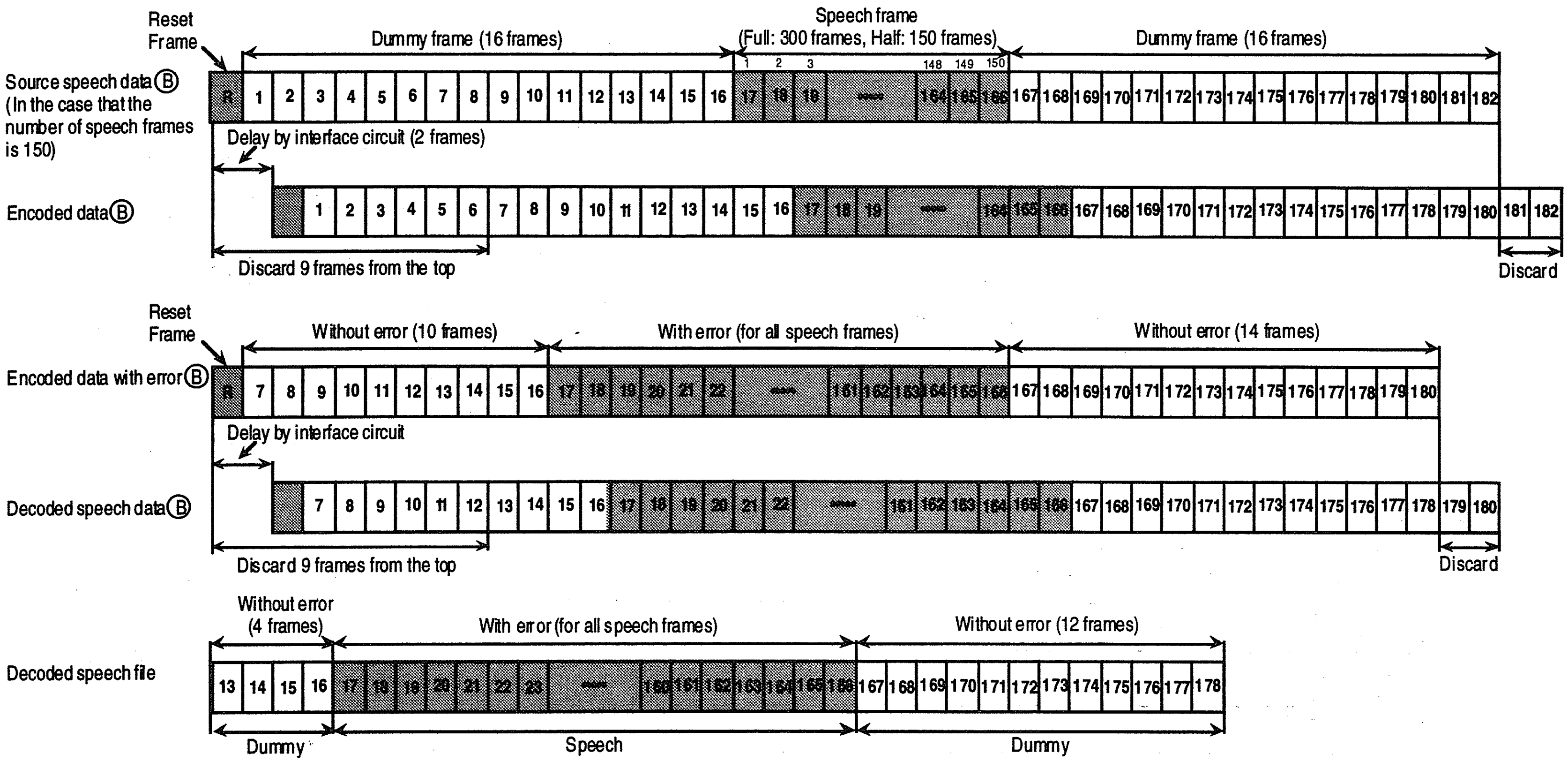
B'_{2n} : ENCODED BIT (B_{2n}) $\oplus$ bY_n (MSB_1)

BX_n, BY_n :

BX_n (MSB1)=ENCODED BIT (B_{2n-1}) $\oplus$ bX_n (MSB_1)
BX_n (MSB2)=bX_n (MSB_2)
BX_n (MSB3)=bX_n (MSB_3)
BX_n (LSB) =bX_n (LSB)
BY_n (MSB1)=ENCODED BIT (B_{2n}) $\oplus$ bY_n (MSB_1)
BY_n (MSB2)=bY_n (MSB_2)
BY_n (MSB3)=bY_n (MSB_3)
BY_n (LSB) =bY_n (LSB)
C_n =c_n

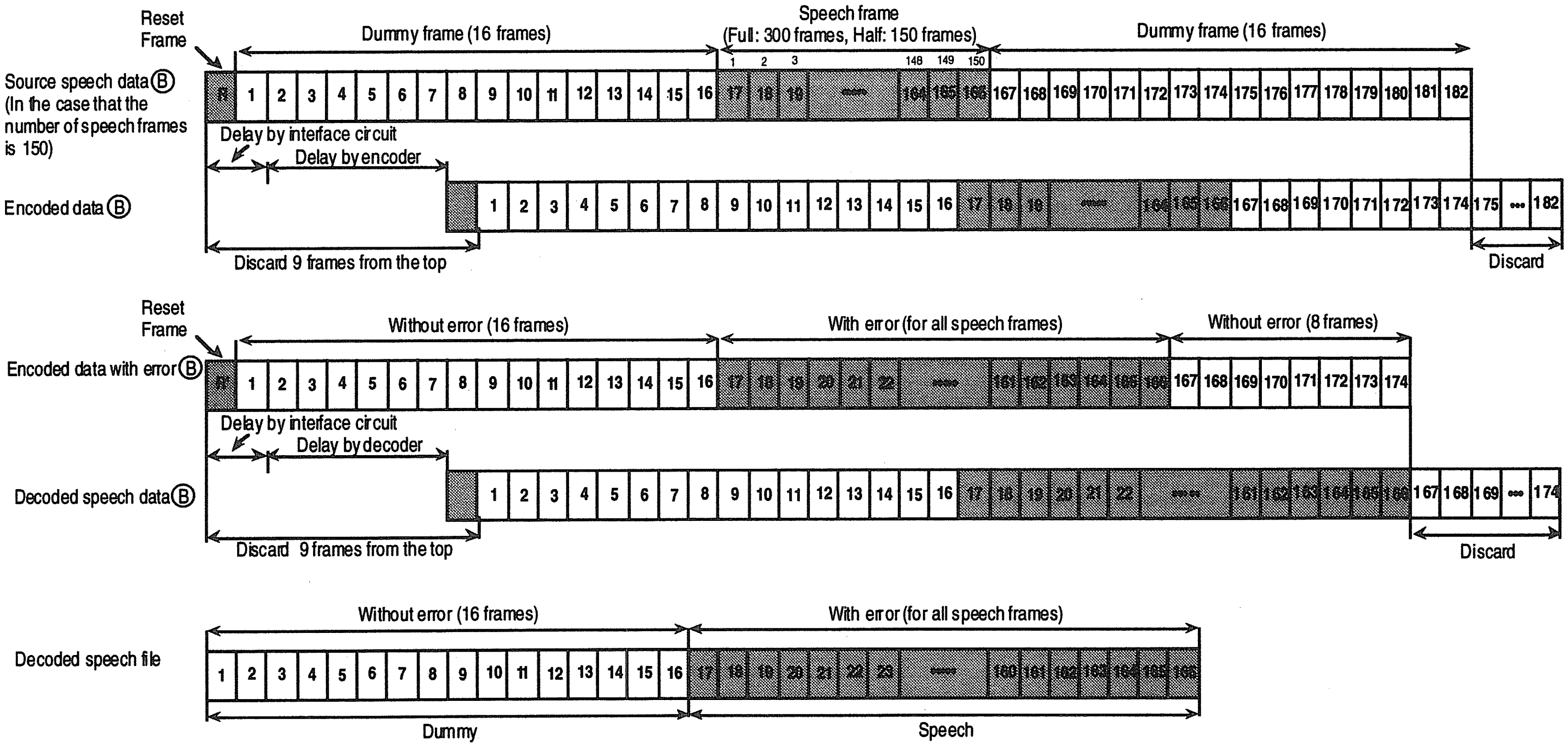
n= Syimbol number (1 to 112)
bXn, bYn, cn:
Outputs of the channel simulator which correspond with bx, by,c in Appendix 3 "Data Format of Error Files for Half-Rate Speech Codecs".

Fig. 7 Signal Format (5/5)



(Note)
- The regulation point of signal timing is Ⓑ in Fig. 9.
- Dummy frames correspond to silence (all "0") data.
- Frame numbers are for half-rate.
- The number of dummy frames before and after speech frames is the same in full-rate and half-rate regardless of the number of speech frames.

Fig. 8 Signal Timing (Frame order) (1/2)
(In case of 0-delay by CODEC)



(Note)

- The regulation point of signal timing is(B) in Fig. 9.
- Dummy frames correspond to silence (all "0") data.
- Frame numbers are for half-rate.
- The number of dummy frames before and after speech frames is the same for full-rate and half-rate regardless of the number of speech frames.

Fig. 8 Signal Timing (Frame order) (2/2)
(In case of 6-frame delay by encoder and 6-frame delay by decoder)

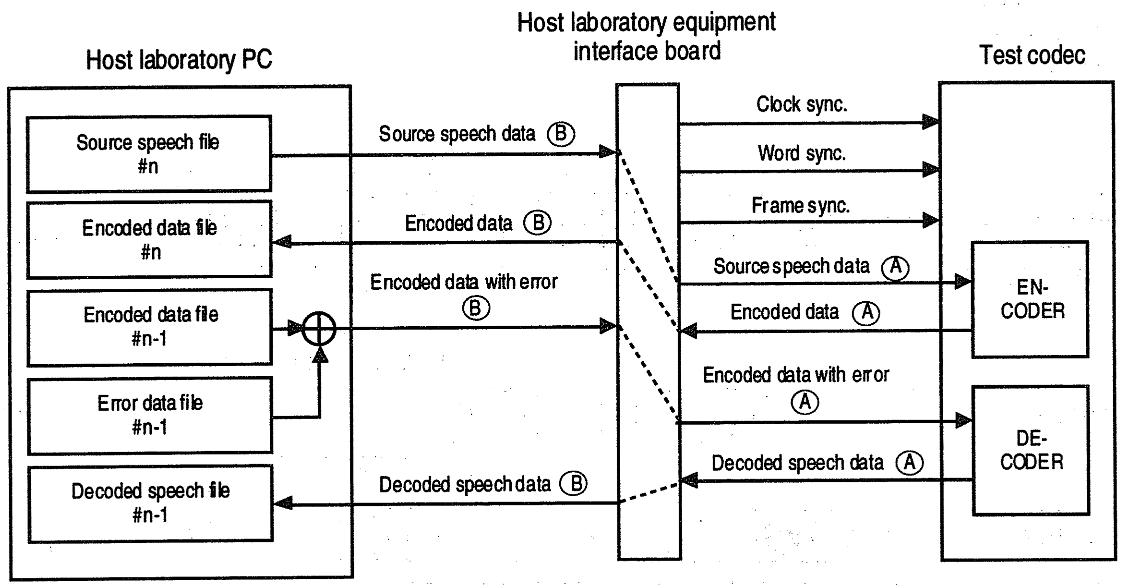


Fig. 9 Structure of Interface

Appendix 2 Data Format of Error Files for Full-Rate Speech Codecs

1. Principle of Quality Evaluation Channel Model for Full-Rate Speech Codecs

Fig.1 shows the block diagram of a channel model.

The modulation method is $\pi/4$ shift QPSK and the demodulation method is limiter delay detection. Signals are received after complex envelope fluctuations by Rayleigh fading and Gaussian noise are added to them.

The reception system has two systems for dual-branch diversity and detection output from the branch with a larger envelope level is selected for each symbol. (Selective diversity after detection) Envelope detection errors are taken into consideration as characteristic deterioration factors with actual receivers.

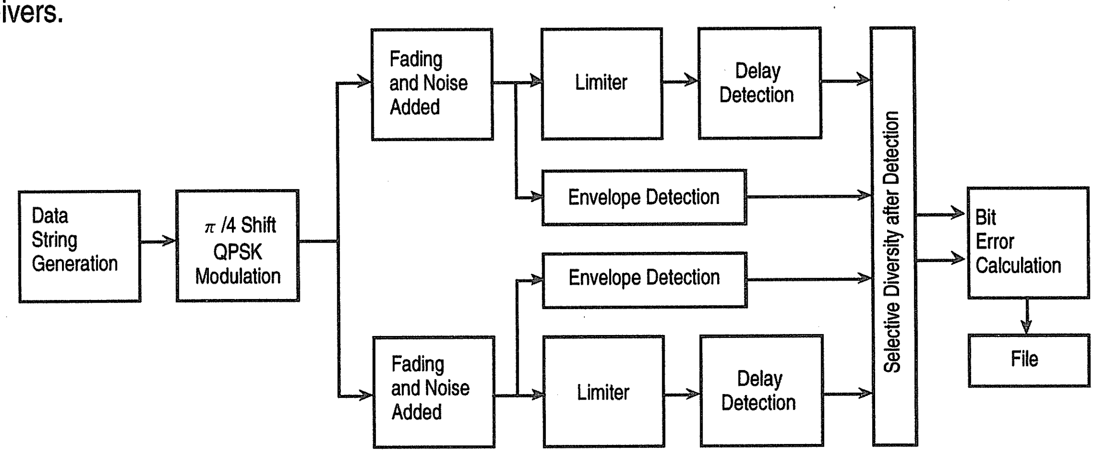


Fig.1 Block Diagram of Channel Model

2. Hard Decision Information

The full-rate hard decision error information for transmitted data is indicated by a bit string of "1" and "0" which are defined as "error" and "no error", respectively.

3. File Data Format

Fig.2 shows the file data format. The data is recorded in units of 8 bits (1 byte) with the MSB coming first and the LSB last. 28 bytes make up 1 frame (224 bits). A file is created by repeating this operation for the required speech length.

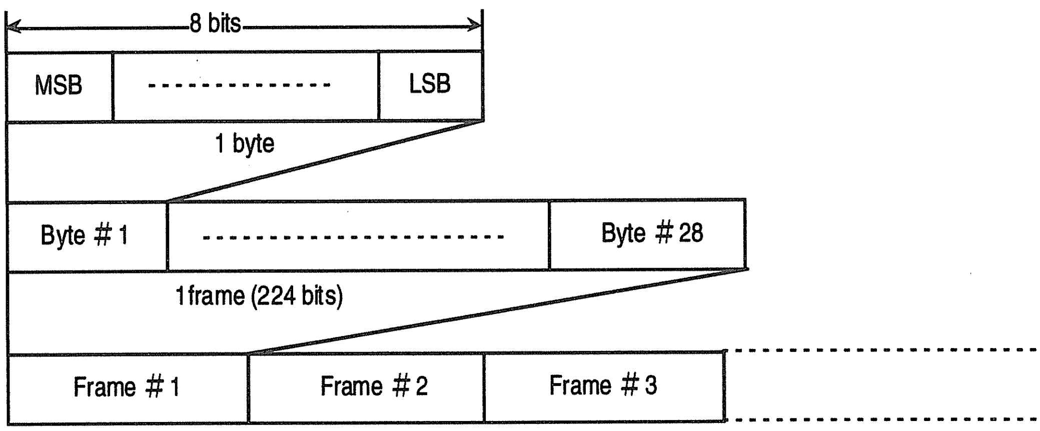


Fig.2 File Data Format

Appendix 3 Data Format of Error Files for Half-Rate Speech Codecs

1. Principle of Quality Evaluation Channel Model for Half-Rate Speech Codecs

Fig. 1 shows the block diagram of a channel model.

The modulation method is $\pi/4$ shift QPSK and the demodulation method is limiter delay detection. Signals are received after complex envelope fluctuations by Rayleigh fading and Gaussian noise are added to them.

The reception system has two systems for dual-branch diversity and detection output from the branch with a larger envelope level is selected for each symbol (selective diversity after detection). Envelope detection errors are taken into consideration as characteristic deterioration factors with actual receivers.

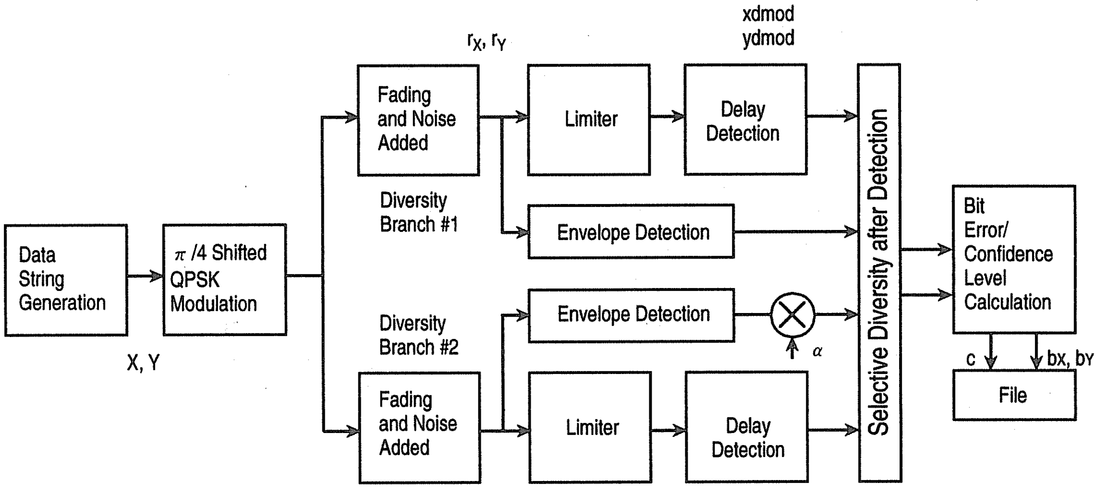


Fig. 1 Block Diagram of Channel Model

2. Hard Decision Information

Hard decision error patterns corresponding to the data X and Y can be obtained as patterns expressed by sign bits b_X and b_Y. Negative sign bits of b_X and b_Y are defined as "1" (error) and positive sign bits, "0" (no error).

3. Soft Decision Information

Confidence level information b_X and b_Y (each expressed by four bits) and envelope information c (expressed by eight bits) are generated for each transmitted symbol (2 bits: X, Y). b_X and b_Y have the following relationships with orthogonal detection output xdmod and ydmod and with X and Y:

$$b_X = -\,\mathrm{ydmod} \times (2X - 1),\ b_Y = -\,\mathrm{xdmod} \times (2Y - 1) \qquad (1)$$

Where X and Y are transmitted bit strings that can be expressed by "0" and "1". In the expressions (2X – 1) and (2Y – 1) in Eq. (1), "0" corresponds to negative and "1" to positive. For details, refer to the "Modulation method" section (3.3.1) in RCR STD-27.

b_X and b_Y express confidence levels. The confidence level is higher, the larger the positive number is.

Conversely, the confidence level is lower, the larger the negative number is. For example, in making a hard decision for each bit, as mentioned above, negative b_X and b_Y sign bits can be regarded as "1" (error) and positive sign bits, "0" (no error).
Envelope information c is defined by the following expression:

$$c = \mathrm{sqrt}\,(r_X^2 + r_Y^2)$$

Higher c values indicate higher reception levels.
Envelope detectors in actual receivers have characteristic dispersion. This effect should be taken into consideration. Diversity operation is carried out by multiplying envelope information of one branch by a constant (α<1). In this model, α = 1 is assumed.

4. File Data Format

Fig. 2 shows the file data format.
Using b_X, b_Y and c (4 + 4 + 8 = 16 bits) as one word, 112 words (corresponding to 40 ms of coded speech) are written in a file and 208 dummy words are written thereafter. A file is created by repeating this operation for the required speech length.

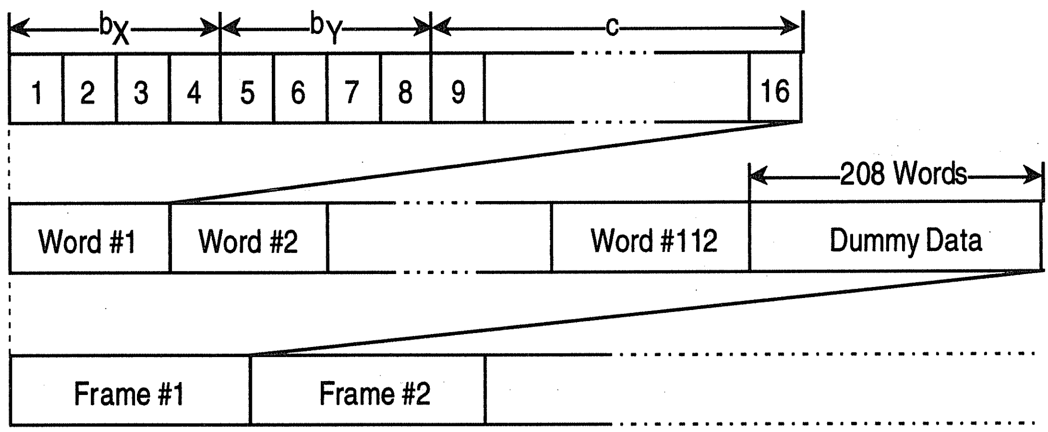


Fig. 2 File Data Format

Appendix 4 Technical Specifications for Source Speech File

Technical specifications for source speech file used in this technical report are listed below:

• Recording microphone characteristic:	Flat characteristic
• Band limitation:	0.3 kHz to 3.4 kHz
• A/D conversion:	8 kHz sampling, 16 bit linear PCM (Upper 14 bits are valid)
• Sample text:	Japanese text (6 seconds)
• Number of talkers:	Male: 36, Female: 36 Boy: 4, Girl: 4
• Source speech samples:	80 types of speech samples (source speech corresponding to speech types listed in Tables 3.2.1.4-1 and 4.2.1.4-1.) 10 types of artificial voice samples (source speech corresponding to the artificial voice samples listed in Tables 3.2.1.4-1 and 4.2.1.4-1. Note that source speech for PM1 and PF1 is used in common for PM1 to PM9 and PF1 to PF9, respectively.)

Appendix 5 Speech File Naming Rules

The naming rules for source speech files, encoded data for reference files, decoded speech for reference files, error data files, MNRU speech files, decoded speech files generated by the tester are listed below:

- Source speech files

    [Talker type] [Talker number] .pcm

[Talker type]	m: male, f: female, b: boy, g: girl pm: artificial male, pf: artificial female
[Talker number]	01 to 36: m, f; 01 to 04: b, g; 01,10 to 13: pm, pf

- Encoded data for reference files

    [Talker type] [Talker number] _ [Codec type] [Reference system] .air

[Talker type]	m: male, f: female, b: boy, g: girl pm: artificial male, pf: artificial female
[Talker number]	01 to 36: m, f; 01 to 04: b, g; 01,10 to 13: pm, pf
[Codec type]	f: full-rate, h: half-rate
[Reference system]	1

- Decoded speech for reference files

    [Talker type] [Talker number] _ [Codec type] [Reference system] .[Extension]

[Talker type]	m: male, f: female, b: boy, g: girl
[Talker number]	01 to 36: m, f; 01 to 04: b, g; 01 to 13: pm, pf
[Codec type]	f: full-rate, h: half-rate
[Reference system]	1, 5
[Extension]	dec: error-free, deh: hard decision, des: soft decision

- Error data files

    [Talker type] [Talker number] .err

[Talker type]	m: male, f: female
[Talker number]	09 to 24: m, f

    * Error-free: 0.err

- MNRU speech files

    [Talker type] [Talker number] _ [Q value] .mnr

[Talker type]	m: male, f: female
[Talker number]	01 to 02: m, f
[Q value]	00, 05, 10, 15, 20, 25, 30, 35, 40

- Decoded speech files

[Talker type] [Talker number] _ [Codec type] [Reference or test system] .[Extension]

[Talker type]	m: male, f: female, b: boy, g: girl
[Talker number]	01 to 04: m, f, b, g, 05 to 36: m, f
[Codec type]	f: full-rate, h: half-rate
[Reference or test system]	Reference systems 1 and 5: test systems 2, 3, 4, and 6
[Extension]	dec: error-free, deh: hard decision, des: soft decision

Appendix 6 Example Objective Evaluation Tests

1. Full-rate speech codec

This section shows the minimum, maximum and median values (the average of the 2 middle values in the case of an even number of data) for the objective evaluation requirements A through D based on the data obtained from the full-rate speech codec objective evaluation tests conducted by the 6 organizations which participated in the test run.

The evaluation equipment used by the six organizations participating in the test run can be categorized as follows:

(1) Types : Hardware (3 sets) and software (3 sets)

(2) Operation precision : 16-bit fixed point (5 sets) and 24-bit fixed point (1 set)

Note: The term "test run" refers to the tests conducted in 1995 to verify the contents of TR-T1.

Table 6.1.1-1 Full-rate objective evaluation characteristics A (speech)

Test item (k)	Test system j = 2								
	$SNRseg_k(2/1)$			$SNRfrq_k(2/1)$			$CD_k(2/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	53.24	19.85	28.92	17.12	0.07	5.00	1.51	0.10	0.32
2	47.42	14.47	23.48	57.24	6.23	13.56	1.63	0.26	0.53
3	34.91	4.38	18.23	91.93	23.33	32.87	3.28	0.77	1.19
4	46.01	18.22	22.66	26.86	1.00	11.75	1.83	0.24	0.89
5	36.35	13.36	20.26	53.57	1.48	11.28	1.99	0.37	1.27
6	42.81	21.15	28.16	10.78	0.12	1.60	1.18	0.12	0.23
7	51.15	22.23	38.13	6.18	0.20	1.09	1.03	0.03	0.07
8	53.75	19.52	27.85	20.25	0.00	8.02	1.77	0.09	0.34

Table 6.1.1-2 Full-rate objective evaluation characteristics A (artificial voice)

Test item (k)	Test system j = 2								
	$SNRseg_k(2/1)$			$SNRfrq_k(2/1)$			$CD_k(2/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	52.60	20.79	30.11	22.64	0.00	1.66	1.35	0.06	0.30
2	39.26	10.80	19.46	68.87	16.69	22.23	1.95	0.52	0.91
3	25.69	1.76	15.38	97.99	32.91	39.89	3.56	0.97	1.48
4	47.24	20.28	24.39	22.39	0.00	3.47	1.49	0.12	0.52
5	40.54	12.77	18.62	64.04	0.14	24.99	1.76	0.27	1.31
6	49.10	21.15	31.74	9.13	0.00	2.40	1.18	0.08	0.18
7	53.95	22.14	38.48	5.28	0.10	0.94	1.22	0.04	0.07
8	–	–	–	–	–	–	–	–	–

Table 6.1.2-1 Full-rate objective evaluation characteristics B (speech)

Test item (k)	Test system j = 3								
	$SNRseg_k(3/1)$			$SNRfrq_k(3/1)$			$CD_k(3/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	9.68	7.21	9.20	93.44	81.10	85.16	2.99	2.40	2.49
2	7.73	5.13	7.60	96.74	89.41	90.75	3.11	2.46	2.51
3	4.83	2.34	4.73	99.44	96.61	97.73	3.53	2.73	2.88
4	11.08	8.19	10.39	90.53	75.08	82.75	3.04	2.32	2.65
5	10.99	7.90	9.90	96.96	79.10	84.81	3.11	2.33	2.68
6	6.10	2.43	5.50	99.90	92.53	95.28	2.97	2.52	2.65
7	4.85	2.15	4.28	99.96	97.09	98.08	2.96	2.59	2.78
8	10.62	8.23	10.21	92.33	78.05	81.51	2.98	2.48	2.57

Table 6.1.2-2 Full-rate objective evaluation characteristics B (artificial voice)

Test item (k)	Test system j = 3								
	$SNRseg_k(3/1)$			$SNRfrq_k(3/1)$			$CD_k(3/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	9.85	3.14	9.27	100.00	78.51	83.33	2.95	2.40	2.61
2	6.32	1.83	5.41	100.00	92.55	95.50	3.20	2.68	2.89
3	3.04	0.33	2.12	100.00	98.80	99.17	3.64	3.05	3.35
4	11.21	3.81	10.14	100.00	73.81	82.89	3.24	2.40	2.63
5	12.36	4.52	10.41	100.00	69.95	85.18	3.49	2.36	2.64
6	8.67	2.76	7.88	100.00	80.79	87.45	2.85	2.42	2.60
7	7.64	2.91	6.97	100.00	88.39	93.59	3.00	2.47	2.70
8	–	–	–	–	–	–	–	–	–

Table 6.1.3-1 Full-rate objective evaluation characteristics C (speech)

Test item (k)	Test system j = 4								
	$SNRseg_k(4/1)$			$SNRfrq_k(4/1)$			$CD_k(4/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	9.69	7.21	9.19	93.44	81.15	86.53	3.00	2.39	2.81
2	7.71	5.11	7.37	97.01	89.20	92.75	3.14	2.54	2.87
3	4.55	1.94	4.36	99.26	96.85	98.26	3.62	2.89	3.26
4	11.14	8.19	10.48	90.40	74.71	82.22	3.05	2.35	2.76
5	10.98	8.45	10.62	92.97	79.31	84.34	2.98	2.50	2.74
6	6.09	2.44	5.54	99.96	92.44	95.81	2.97	2.52	2.76
7	4.89	2.23	4.49	100.00	96.76	98.27	2.96	2.57	2.72
8	10.62	8.24	10.15	92.10	78.09	83.49	2.99	2.48	2.86

Table 6.1.3-2 Full-rate objective evaluation characteristics C (artificial voice)

Test item (k)	Test system j = 4								
	$SNRseg_k(4/1)$			$SNRfrq_k(4/1)$			$CD_k(4/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	9.84	3.25	9.20	100.00	79.56	85.91	3.01	2.41	2.73
2	6.17	1.77	5.21	100.00	93.14	96.35	3.26	2.72	3.05
3	2.36	0.29	1.81	100.00	98.66	99.32	3.69	3.20	3.56
4	11.20	3.83	10.31	99.88	73.81	81.52	3.25	2.41	2.71
5	12.36	4.55	10.63	100.00	69.95	81.54	3.51	2.40	2.71
6	8.44	2.90	7.83	100.00	82.67	87.92	2.85	2.42	2.71
7	7.61	3.00	7.00	100.00	90.43	93.50	2.93	2.54	2.75
8	–	–	–	–	–	–	–	–	–

Table 6.1.4-1 Full-rate objective evaluation characteristics D (speech)

Test item (k)	Test system j = 6					
	$SNRseg_k(6/5)$			$CD_k(6/5)$		
	Max.	Min.	Med.	Max.	Min.	Med.
1	41.02	30.97	33.22	0.27	0.11	0.21
4	39.63	26.01	29.70	0.48	0.21	0.39
5	37.91	19.41	25.40	0.82	0.24	0.61
8	41.47	28.79	32.02	0.32	0.11	0.22

Table 6.1.4-2 Full-rate objective evaluation characteristics D (artificial voice)

Test item (k)	Test system j = 6					
	$SNRseg_k(6/5)$			$CD_k(6/5)$		
	Max.	Min.	Med.	Max.	Min.	Med.
1	40.34	30.81	34.71	0.20	0.06	0.15
4	38.95	23.77	29.80	0.47	0.13	0.29
5	39.11	16.19	24.02	1.09	0.28	0.70
8	–	–	–	–	–	–

2. Half-rate speech codec

This section shows the minimum, maximum and median values (the average of the 2 middle values in the case of an even number of data) for the objective evaluation requirements A through D based on the data obtained from the half-rate speech codec objective evaluation test conducted by the 10 organizations which participated in the test run.

The evaluation equipment used by the ten organizations participating in the test run can be categorized as follows:

(1) Types :	Hardware (8 sets) and software (2 sets)
(2) Operation precision :	16-bit fixed point (6 sets), 24-bit fixed point (3 sets), and 32-bit floating point (1 set)

Note: The "test run" refers to the tests conducted in 1995 to verify the contents of TR-T1.

Table 6.2.1-1 Half-rate objective evaluation characteristics A (speech)

Test item (k)	Test system j = 2								
	$SNRseg_k(2/1)$			$SNRfrq_k(2/1)$			$CD_k(2/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	53.13	30.14	35.77	5.99	0.00	0.76	0.36	0.06	0.25
2	53.20	30.06	36.02	8.34	0.14	2.04	0.36	0.06	0.25
3	53.21	28.67	33.59	16.28	0.03	7.96	0.57	0.07	0.29
4	49.45	28.70	34.96	6.68	0.13	0.65	0.51	0.15	0.29
5	47.08	26.11	32.51	10.16	0.11	1.88	0.76	0.24	0.47
6	55.03	29.45	35.45	8.14	0.00	1.47	0.41	0.04	0.23
7	55.76	31.93	36.85	4.83	0.00	1.05	0.31	0.03	0.16
8	53.41	24.61	32.58	23.36	0.09	4.81	0.65	0.07	0.40
9	53.23	29.91	34.42	8.53	0.14	2.51	0.39	0.06	0.29
10	53.38	27.68	33.00	13.58	0.03	6.01	0.66	0.06	0.35

Table 6.2.1-2 Half-rate objective evaluation characteristics A (artificial voice)

Test item (k)	Test system j = 2								
	$SNRseg_k(2/1)$			$SNRfrq_k(2/1)$			$CD_k(2/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	58.93	29.21	32.99	14.76	0.00	2.27	0.38	0.03	0.26
2	58.58	29.24	32.70	15.60	0.03	3.95	0.39	0.03	0.26
3	58.08	27.21	31.96	20.97	0.21	7.67	0.61	0.04	0.29
4	53.57	27.46	32.78	21.89	0.36	2.05	0.52	0.09	0.27
5	50.85	25.01	32.34	18.73	0.15	1.78	0.83	0.17	0.41
6	57.95	26.93	32.78	23.04	0.10	1.72	0.53	0.03	0.25
7	55.21	29.30	34.59	13.43	0.10	1.24	0.68	0.03	0.22
8	–	–	–	–	–	–	–	–	–
9	58.70	28.58	32.13	16.17	0.03	5.78	0.40	0.03	0.30
10	58.34	26.19	31.14	19.71	0.21	7.79	0.69	0.04	0.31

Table 6.2.2-1 Half-rate objective evaluation characteristics B (speech)

Test item (k)	Test system j = 3								
	$SNRseg_k(3/1)$			$SNRfrq_k(3/1)$			$CD_k(3/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	30.69	16.69	20.27	61.41	41.84	56.07	1.72	1.16	1.57
2	31.12	16.09	18.91	68.81	44.35	62.50	1.79	1.15	1.66
3	26.89	13.60	16.44	76.61	54.61	70.18	2.08	1.41	1.91
4	30.35	15.60	18.58	63.48	38.59	56.40	1.90	1.12	1.65
5	24.72	13.40	16.19	70.52	45.74	60.22	1.92	1.29	1.71
6	38.98	17.92	23.02	61.82	32.05	52.89	1.64	0.93	1.46
7	50.60	11.04	25.26	75.39	24.88	51.68	2.23	0.72	1.40
8	25.53	15.54	17.19	63.75	43.78	57.01	2.03	1.45	1.85
9	31.26	15.73	18.78	68.69	44.23	62.12	1.79	1.15	1.65
10	27.10	13.96	16.86	76.18	54.30	70.18	2.02	1.38	1.91

Table 6.2.2-2 Half-rate objective evaluation characteristics B (artificial voice)

Test item (k)	Test system j = 3								
	$SNRseg_k(3/1)$			$SNRfrq_k(3/1)$			$CD_k(3/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	27.56	15.04	20.44	70.07	51.76	62.48	1.96	1.30	1.65
2	25.31	12.98	19.10	80.56	61.13	70.95	2.16	1.51	1.84
3	23.03	10.41	15.20	86.78	67.91	77.70	2.45	1.78	2.09
4	22.34	12.58	17.60	75.54	53.31	63.84	2.01	1.45	1.83
5	16.82	10.26	12.97	83.36	66.35	73.80	2.63	1.96	2.24
6	28.15	15.60	18.45	69.13	52.94	65.22	1.89	1.35	1.73
7	35.36	19.20	22.10	65.47	46.62	59.83	1.72	1.22	1.65
8	–	–	–	–	–	–	–	–	–
9	25.80	13.26	19.22	80.54	60.18	70.23	2.14	1.46	1.82
10	23.44	10.69	16.88	86.57	67.48	77.20	2.44	1.73	2.08

Table 6.2.3-1 Half-rate objective evaluation characteristics C (speech)

Test item (k)	Test system j = 4								
	$SNRseg_k(4/1)$			$SNRfrq_k(4/1)$			$CD_k(4/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	30.69	14.36	16.16	61.78	41.81	56.40	1.73	1.16	1.62
2	31.12	14.02	15.22	68.65	44.35	62.83	1.80	1.15	1.69
3	26.89	11.72	12.70	76.60	54.61	70.30	2.08	1.41	1.93
4	30.35	14.22	15.49	63.04	38.59	56.92	1.90	1.12	1.69
5	24.72	13.38	14.72	71.52	45.74	59.54	1.95	1.29	1.76
6	38.98	14.83	16.63	61.78	32.05	54.15	1.68	0.93	1.51
7	50.60	7.09	17.73	75.38	24.88	51.83	2.24	0.72	1.42
8	25.53	13.61	14.91	65.88	43.83	59.92	2.08	1.46	1.89
9	31.26	13.08	14.87	68.53	44.23	63.46	1.82	1.15	1.71
10	27.10	11.33	12.85	76.11	54.30	70.98	2.13	1.38	2.00

Table 6.2.3-2 Half-rate objective evaluation characteristics C (artificial voice)

Test item (k)	Test system j = 4								
	$SNRseg_k(4/1)$			$SNRfrq_k(4/1)$			$CD_k(4/1)$		
	Max.	Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.
1	27.56	12.75	14.43	72.03	51.76	65.62	2.35	1.30	1.77
2	25.17	10.75	13.22	80.54	61.13	71.56	2.20	1.51	1.92
3	23.03	8.21	10.45	86.75	67.91	78.27	2.51	1.78	2.23
4	22.34	12.08	13.54	75.90	53.19	68.61	2.52	1.45	1.90
5	14.62	9.91	11.82	83.50	68.15	74.78	2.71	1.97	2.39
6	28.15	12.16	14.50	69.58	52.94	65.31	1.96	1.35	1.76
7	35.36	13.93	15.89	65.37	46.62	60.39	1.78	1.22	1.65
8	–	–	–	–	–	–	–	–	–
9	25.80	10.93	13.36	80.48	60.18	70.91	2.18	1.46	1.87
10	23.44	8.30	10.89	86.77	67.48	77.69	2.49	1.73	2.21

Table 6.2.4-1 Half-rate objective evaluation characteristics D (speech)

Test item (k)	Test system j = 6					
	$SNRseg_k(6/5)$			$CD_k(6/5)$		
	Max.	Min.	Med.	Max.	Min.	Med.
1	52.95	32.72	39.74	0.20	0.06	0.11
4	49.47	31.43	38.89	0.31	0.13	0.21
5	47.69	30.73	36.09	0.56	0.22	0.37
8	53.52	32.42	40.04	0.31	0.07	0.11

Table 6.2.4-2 Half-rate objective evaluation characteristics D (artificial voice)

Test item (k)	Test system j = 6					
	$SNRseg_k(6/5)$			$CD_k(6/5)$		
	Max.	Min.	Med.	Max.	Min.	Med.
1	58.03	33.75	41.96	0.29	0.03	0.06
4	53.54	33.53	39.37	0.38	0.07	0.14
5	50.91	28.44	35.12	0.48	0.14	0.30
8	–	–	–	–	–	–

Appendix 7 Reference Material

1. "Voice Quality Measurement System for Digital Cellular Telephone Codec" by Miki Suda
   -- "The Institute of Electronics, Information and Communication Engineers (IEICE)" RCS 90-8, July 19, 1990

Appendix 8 PDC Speech Codec Standardization Guideline

1. Overview

PDC speech codec standardization guideline defines the procedure to standardize new PDC speech codecs in addition to full-rate speech codec defined in Section 5.1 of RCR STD-27 (hereinafter called full-rate speech codec) and half-rate speech codec defined in Section 5.2 of RCR STD-27(hereinafter called half-rate speech codec).

2. Requirements

Requirements for the new speech codec are defined below.

(1) Total bitrate including speech and channel coding shall be 5.6kbit/s or 11.2kbit/s.

(2) The speech quality shall be equal to or better than that of full-rate speech codec or half-rate speech codec.

(3) The frame structure shall comply with that of full-rate or half-rate channel defined in RCR STD-27.

(4) The algorithmic delay shall not exceed that of half-rate speech codec.

3. Standardization procedure

Speech codec, which fulfills all the requirements given in this guideline can be adopted as the new PDC speech codec standard.

Details of the standardization procedure is defined below.

(1) Subjective evalulation test

The proponent shall perform subjective evaluation test for its candidate codec according to the procedure described in this guideline.

(2) Submission of codec specifications and subjective test results.

The proponent shall submits algorithm description, and subjective evaluation test results including values of MA, TA, MV, TV and t defined in Section 3.2.2.10 of TR-T1 for significant difference testing.

(3) Working Group evaluation

Working Group shall evaluate the specifications and subjective evaluation test results of the candidate codec submitted by the proponent.

(4) Submission of the documents necessary for RCR STD-27 revision

If Working Group confirms that the candidate codec fulfills all the requirements defined in this guideline. then Working Group shall request the proponent to provide a document necessary for the revision of

RCR STD-27 and other necessary documents such as validation test method. Description of the document provided for the revision should be equivalent to that of Section 5 of RCR STD-27.

(5) Revision of the standard

Working group shall revise the standard based on the documents submitted by the proponent.

4. Documents to be provided

(1) Codec specifications

The proponent shall provide the specifications (See attachment 1; PDC Speech Codec Specification List) of the candidate codec. In case that the candidate codec has already been standardized by the other standardization body, the proponent shall provide the name of the standard, changes for complying with the PDC frame structure and other changes made by the proponent if any.

(2) Speech evaluation test results

The proponent shall provide the subjective evaluation test results and the values necessary for a significance test with t-test, i.e., MA, TA, MV, TV and t defined in Section 3.2.2.10 of TR-T1. The proponent shall also provide a demonstration tape including speech of reference codec(s).

5. Subjective evaluation test method

The subjective evaluation test shall be performed based on subjective requirements. The ACR test defined by ITU shall be used for evaluation.

5.1 Subjective requirements

The candidate codec satisfying the subjective requirements shall be considered as a codec to be standardized. Then the standard will be revised based on the documents provided by the proponent.

The subjective requirements are defined below.

Subjective evaluation test results of the candidate codec shall be less inferior than that of the full-rate speech codec or the half-rate speech codec (reference codecs defined in Section 3.3.3 and 4.3.3 of TR-T1), for test items 1 to 8 defined in Table 1. This requirement shall be evaluated by using one-sided inspection with the standard for determining the significant difference being 5%. Note that it is not mandatory to evaluate the test item 3 since it depends on the design of the radio access area. Hard decision or soft decision selection and noise suppresser on/off condition for the reference codec shall be reported. The reference codec shall be tested under the same noise suppresser on/off condition with the candidate codec. In addition, the soft/hard decision conditions of the reference codec preferably includes the same conditions as the candidate codec, if possible.

Also, that MOS value to equivalent Q-value conversion is adequate shall be evaluated according to the method described in Section 3.2.2.10 of TR-T1.

5.2 Test items

Test items are defined in Table 1. There are 8 test items in total, i.e. basic characteristics, error robustness, level variation, background noise, and talker dependency. The proponent shall report the test conditions of the candidate codec such as hard-decision or soft-decision and with or without noise suppresser.

Table 1. Voice subjective standard test item of speech CODEC

Test	Test item	Error condition	Input level	Additional noise	No. of speech samples
Basic characteristics	1	Error-free	Normal	None	16
Error robustness	2	1% (4km/h)	Normal	None	4
		1% (20km/h)	Normal	None	4
		1% (60km/h)	Normal	None	4
		1% (60km/h)*	Normal	None	4
	3	3% (4km/h)	Normal	None	4
		3% (20km/h)	Normal	None	4
		3% (60km/h)	Normal	None	4
		3% (60km/h)*	Normal	None	4
Level variation	4	Error-free	Normal -10dB	None	4
	5	Error-free	Normal -20dB	None	4
Background noise	6	Error-free	Normal	Low noise level SNR30dB	8
	7	Error-free	Normal	High noise level SNR15dB	8
Talker dependency	8	Error-free	Normal	None	8

(Notes)

(4) Error data marked with an "*" results from fading at 1.5GHz band. Error data without an "*" results from fading at 800MHz.

(5) "Normal" is the level at which the mean level is 21dB lower than the sine-wave full scale of the A/D converter.

(6) Superposed noise is defined as follows:

None: Noise, which is unavoidably mixed during recording of source speech and quantization noise.
Low level noise: SNR30dB
High level noise: SNR15dB

5.3 Speech samples

The speech samples used in subjective evaluation test are defined in Table 2. Creation of encoding and decoding data shall follow this table.

Table 2 Speech samples used for subjective evaluation test

<table>
<tr><th colspan="3">Name of test</th><th>Test item</th><th>Variety and number of speech samples</th></tr>
<tr><td colspan="3">Basic characteristics</td><td>1</td><td>8 male
8 female</td></tr>
<tr><td rowspan="8">Error robustness</td><td rowspan="4">1%</td><td>4km/h</td><td rowspan="4">2</td><td>2 male
2 female</td></tr>
<tr><td>20km/h</td><td>2 male
2 female</td></tr>
<tr><td>60km/h</td><td>2 male
2 female</td></tr>
<tr><td>60km/h*</td><td>2 male
2 female</td></tr>
<tr><td rowspan="4">3%</td><td>4km/h</td><td rowspan="4">3</td><td>2 male
2 female</td></tr>
<tr><td>20km/h</td><td>2 male
2 female</td></tr>
<tr><td>60km/h</td><td>2 male
2 female</td></tr>
<tr><td>60km/h*</td><td>2 male
2 female</td></tr>
<tr><td colspan="3" rowspan="2">Level variation</td><td>4</td><td>2 male
2 female</td></tr>
<tr><td>5</td><td>2 male
2 female</td></tr>
<tr><td colspan="3" rowspan="2">Background noise</td><td>6</td><td>4 male
4 female</td></tr>
<tr><td>7</td><td>4 male
4 female</td></tr>
<tr><td colspan="3">Talker dependency</td><td>8</td><td>4 boy
4 girl</td></tr>
</table>

5.4 Error data

The error data for the full-rate speech codec and the half-rate speech codec defined in Section 5 of TR-T1 will be provided by their owners through proper procedure.

5.5 Listening test

Listening test shall conform to the procedure defined in Section 3.2.2.6 and 4.2.2.6 of TR-T1.

One ear-piece in the headphone shall be used for listening and listener can select the side. Decoded speech and reference speech signals shall be converted from digital to analog with 16-bit resolution. Then they shall be presented to the listener by headphone through band pass filter in order to eliminate the out of band noise blow 0.3KHz and above 3.4KHz. Any equipment affects speech quality shall not be used except amplifier or attenuator for adjusting listening level. Recording and playing by DAT may be used for presentation of the speech. Listening room shall be adequately quiet and with less environmental noise.

Recommended listening level is -24 dBPa (i.e. 70 dBSPL). Environmental noise level shall be less than 35 dBA (standard A characteristic).

Recommended equipment for the listening test is described in Section 3.3.6 of TR-T1.

5.6 Listeners

At least 40 listeners shall participate. The gender and age of the listeners is defined below.

(Refer to Section 3.2.2.7 of TR-T1)

Male:Female Ratio	Male	(60±10)%
	Female	(40±10)%
Age Composition	Late teens to 20's	(30±10)%
	30's and 40's	(50±10)%
	50 and over	(20±10)%

Speech and sound experts shall not be included in the listeners. Listener's native language shall be noted in the document of the test results.

The listeners shall be divided into at least 5 groups (the population of each group should be averaged), and different randomization shall be used for each groups.

5.7 Test conditions

The test conditions are defined below:

- Listener shall listen all of the speech samples.
- The interval between two speech samples shall be 1 second.
- Each sample shall be 6 seconds or longer. (Refer to Appendix 4 of TR-T1. Note that the default language is Japanese. If a language other than Japanese is used for listening test, the language used shall be reported in the document provided.)
- One group shall listen whole samples sequentially.

5.8 Analysis and judgment

The test results shall be judged by the methods defined in Section 3.2.2.10 of TR-T1.

Significance difference between MOS score of the candidate codec and MOS score of the reference codec (defined in Section 3.3.3 and/or 4.3.3 of TR-T1) shall be tested for each test item. Error of MOS to Opinion-equivalent Q value conversion shall also be calcurated.

(Attachment 1)

PDC Speech Codec Specification List

Company name:

No.	Item			Data	Remarks (Special notations)
1	Algorithm				
2	Bit-rate				
3	Frame length				
4	Delay	Waveform input			
		Interleaving delay			
		Transmission delay			
		Interpolation wait time			
5	Sampling frequency				
6	Memory capacity	In ROM			
		Data ROM			
		Total ROM			
		RAM			
		Bit/W			
7	Processing volume	Encoding	MIPS		
			MOPS		
		Decoding	MIPS		
			MOPS		
8	Short term prediction	Prediction order			
		Prediction analysis method			
		quantization parameters			
		Number of bits			
		Quantization method			
9	Long Term Prediction	Prediction order			
		Prediction analysis method			
		Pitch analysis method			
		Pitch searching range			
		Number of bits			
		Time/frequency domain			
		Quantization method			

No.	Item			Data	Remarks (Special notations)
10	Residual Quanti-zation	Quantization method			
		Number of bits			
		Number of vector dimensions			
		Number of vector candidates			
		Vector search method			
		Vector expression method			
		Vector table	Size		
			Characteristics		
			Training method		
		Other characteristics			
11	Post filter		Yes/No		
			Characteristics		
12	Error protection	Error correction/detection codes			
		Assignment			
		Interleave	Method		
			Size		
		Interpolation, squelch, etc.			
		Other characteristics			
13	Others	Detailed block diagram			
		Frame structure			
		Processor			
		Floating/fixed point implementation			

(Attachment 2)

Contents and Result of Subjective Evaluation Test

Codec Operating Conditions during Subjective Evaluation Test

	Reference Codec	Candidate Codec
Hard decision/soft decision		
noise canceller		

Language for Speech Samples and Listeners

Language for speech sample		
Number of listeners (40 or more) and composition	Male Female	
	10s, 20s, 30s, 40s, 50s or older	
	Native language of listeners	

One-tailed t-test with 95% confidence interval shall be used for judgement.

The definition of each variable described in TR-T1.

Test item (k)	Reference system i=1		Test system j=4			
	MAv1	MVrk1	TAvk4	TVrk4	tk4	Result of judgement
1						
2						
3						
4						
5						
6						
7						
8						

MOS to equivalent Q value conversion

MOSmx	
G	
I	
mean square error (MSE) of MOS－Q value conversion	

List of samples in the demo tape to be submitted:

(Any form is acceptable.)

Attachment Amendment History of ARIB TR-T1

Version 1.1

Addition of CS-ACELP SPEECH CODEC and ACELP SPEECH CODEC
Revision of Section 1.2 and Addition of Chapter 6, 7 and Appendix 8.

PERSONAL DIGITAL CELLULAR TELECOMMUNICATION SYSTEM

TECHNICAL REPORT

ARIB TR-T1

Issued	DECEMBER	1995
REVISION 1.1	JULY	2000

Published by

Association of Radio Industries and Businesses

Nittochi Bldg. 11F
1-4-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-0013, Japan

TEL +81-3-5510-8590
FAX +81-3-3592-1103

Printed in Japan