Phase II Testing and Evaluation of Low Data Rate Voice CODEC Equipment

Thomas Dehel
Marcus Grable
Joseph Child
(CTA Incorporated)

August 1989

DOT/FAA/CT-TN89/49

U.S. Department of Transportation
Federal Aviation Administration

Technical Center
Atlantic City International Airport, N.J. 08405

Technical Report Documentation Page

1. Report No. DOT/FAA/CT-TN89/49	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle PHASE II TESTING AND EVALUATION OF LOW DATA RATE VOICE CODEC EQUIPMENT		5. Report Date August 1989
		6. Performing Organization Code ACD-330
7. Author(s) T. Dehel, M. Grable and J. Child (CTA Incorporated)		8. Performing Organization Report No. DOT/FAA/CT-TN89/49
9. Performing Organization Name and Address Department of Transportation Federal Aviation Administration Technical Center Atlantic City International Airport, N.J. 08405		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. T0704F
12. Sponsoring Agency Name and Address Department of Transportation Federal Aviation Administration Advanced System Design Service Washington, D.C. 20591		13. Type of Report and Period Covered Technical Note August 1989
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The Federal Aviation Administration (FAA) is currently evaluating low rate voice digitizing coder/decoder (CODEC's) for use with the Aeronautical Mobile Satellite Service (AMSS). Phase II of this evaluation consisted of air traffic control (ATC) personnel participating in an objective intelligibility test of several CODEC's under operational conditions. The results of the testing show that the intelligibility of the low rate 4.8 kilobits per second (kbps) CODEC's is essentially equivalent to the intelligibility of the 9.6 kbps CODEC. The results also show that the 4.8 kbps CODEC's can operate with high intelligibility under conditions of high bit error rates and operational background noise.

17. Key Words CODEC, Voice Coding, Intelligibility	18. Distribution Statement This document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	42	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

EXECUTIVE SUMMARY

The Federal Aviation Administration (FAA) is currently evaluating low rate voice digitizing coder/decoder (CODEC's) for use with the Aeronautical Mobile Satellite Service (AMSS). Phase II of this evaluation consisted of air traffic control (ATC) personnel participating in an objective intelligibility test of several CODEC's under operational conditions. The results of the testing show that the intelligibility of the low rate 4.8 kilobits per second (kbps) CODEC is essentially equivalent to the intelligibility of the 9.6 kbps CODEC. The results also show that the 4.8 kbps CODEC's can operate with high intelligibility under conditions of high bit error rates and operational background noise.

1. INTRODUCTION.

1.1 OBJECTIVE.

The primary objective of the Phase II coder/decoder (CODEC) testing is to evaluate the voice intelligibility and acceptability of low data rate voice digitizing CODEC's for use in air traffic control (ATC).

1.2 BACKGROUND.

The International Civil Aviation Organization (ICAO) has developed requirements for the Aeronautical Mobile Satellite Service (AMSS) for use by the civil aviation community; satellite systems are currently under development to meet these needs. In the 1990's, such systems will provide data and digitized voice communications with aircraft for ATC, airline operations and administration, and passenger use.

Testing was conducted by the Airline Electronics Engineering Committee (AEEC) to select a 9.6 kilobits per second (kbps) CODEC for passenger telephony over AMSS. The Federal Aviation Administration (FAA) is continuing to evaluate lower data rate CODEC's for suitability in the ATC environment.

The purpose of evaluating the intelligibility of low data rate voice CODEC's is to determine their suitability for use in the ATC environment. The 9.6 kbps CODEC's are generally accepted as practical for voice communications. However, the lower data rate CODEC's, if found to be suitable, offer system performance advantages due to their lower radio frequency (RF) transmission power and bandwidth requirements. The technology of lower rate CODEC's, especially those operating at transmission rates of 4.8 kbps, is rapidly improving. With improved processing algorithms, some CODEC's at these lower rates may be acceptable for ATC use.

The FAA completed a subjective CODEC evaluation in Phase I of its CODEC testing program in which ATC personnel from the New York Center listened to recordings of voice messages processed by the various CODEC's under test. The controllers gave their opinion of voice acceptability and intelligibility, with Mean Opinion Scores (MOS) ranging from 0 for "bad" to 5 for "excellent."

The Phase I tests were conducted under quiet background noise conditions. In preparing the audio test tapes, the voice messages were processed through CODEC's or CODEC software algorithms without inducing any transmission bit errors. The test results indicated that the intelligibility and acceptability of the best 4.8 kbps CODEC's were within 0.5 MOS points of the 9.6 kbps CODEC.

The results of Phase I testing demonstrated that the performance of 4.8 kbps CODEC's was close to that of 9.6 kbps CODEC's. However, Phase I included only subjective opinion testing; no background noise or bit errors were introduced to simulate actual field operational conditions. Phase I testing concentrated on opinion score testing of CODEC's to evaluate their intelligibility performance. Phase II of the test program, built on the results of Phase I, focused on more objective intelligibility tests, as opposed to subjective opinion scoring, and used only CODEC's available in hardware form.

The principal intent of Phase II was to develop objective intelligibility test procedures meaningful for the ATC environment and to measure the intelligibility and acceptability of several CODEC's using those procedures.

1.3 RELATED DOCUMENTATION.

a. International Civil Aviation Organization (ICAO), Report of the Fourth Meeting of the Special Committee on Future Air Navigation Systems (FANS/4), Montreal, 1988.

b. Aeronautical Radio, Inc. (ARINC), Aviation Satellite Communications System, Part 2, System Design and Equipment Functional Description, ARINC Characteristic 741, Annapolis, March 15, 1989.

c. Airlines Electronic Engineering Committee (AEEC), Report of Satellite Subcommittee's Voice Coding Working Group Meeting Held December 13, 1988, in Arlington, Virginia, AEEC Letter 89-006/SAT-96, February 13, 1989.

d. Child, J., Cleve, R., and Grable, M., Evaluation of Low Data Rate Voice CODECS for Air Traffic Control Applications, Technical Note, CTA Inc. for the FAA Technical Center, DOT/FAA/CT-TN89/13, January 1989.

2. TEST PROGRAM.

2.1 TEST OBJECTIVES

The test objectives of Phase II CODEC testing were:

a. To develop objective test procedures to evaluate the intelligibility of CODEC's in an ATC environment.

b. To use these procedures to evaluate the intelligibility of low bit rate CODEC's (4.8 and 2.4 kbps) available in hardware to determine if low data rate CODEC technology is mature enough for application in the ATC environment.

c. To quantify the effects of bit error rate and background noise on CODEC intelligibility.

d. To determine the acceptability of the CODEC's under test by collecting the opinions of the ATC personnel and to relate this subjective opinion to the more objective intelligibility results.

e. To conduct these evaluations over simulated telephone lines.

2.2 TEST METHODOLOGY.

Normally, an intelligibility test involves people listening to recordings of aural messages. The ability of the listener to properly understand the messages is a measure of intelligibility. There are many different forms of test message texts and methods of scoring intelligibility. Intelligibility test texts can be classified in terms of their "openness." A "closed" text is a set of messages that are well known to the listener or have a limited number of possible interpretations. For example, the listener may be asked to choose between two words such as "show" and "snow." An example of such a "closed" test is the industry standard Diagnostic Rhyme Test (DRT).

An "open" test consists of text with a large number of possible interpretations by the listener. An example of "open" text is a list of polysyllabic nonsense words. When using such text, there is a very large number of possible interpretations by the listener.

Other intelligibility tests fall between the "closedness" of the DRT and the "openness" of nonsense text. Specialized intelligibility tests are often desirable to help determine the suitability of a voice system for specific applications. In some circumstances, voice messages are normally spoken in a partially "closed" environment. For example, when talking to an ATC Center, aircraft pilots normally use a limited vocabulary. Almost all their words are spoken in a context that is well known to the listener. This enhances the intelligibility of words in the messages, resulting in higher intelligibility than would be experienced testing with nonsense syllables. Likewise, the pilot's messages will be less "closed" than a DRT text.

To evaluate the performance of a CODEC in the ATC environment, tests were run at three levels of "openness." Part 1 of the test is a "closed" pseudo-DRT test. It differs from the industry standard DRT test in that some of the word pairs of Part 1 are polysyllabic. Also, a number of words common to the ATC environment were included in the text without regard to the phonetic balance of the resulting word set. Although Part 1 differs from standard DRT tests, it is "closed" and allows the overall test results to be equated to generally accepted intelligibility metrics.

Part 2 consists of pilot-ATC Center messages wherein the test listener identifies spoken numerals within sentences. For example, the message may be "United flight 62 requesting weather ......" The listener is to identify the number 62. Part 2 is considered a partially "open" test.

The text for Part 3 consists of isolated, unrelated words. Some of the words, such as "pilot," are commonly used in the ATC environment. Therefore, Part 3 is not fully "open" since such words are well known and, to an extent, anticipated by the listener. By using such words, the test provides some understanding of the utility of each CODEC in the ATC environment.

Test sheets used for all three parts are shown in appendix A.

To simulate real-world ATC conditions, some of the voice messages used for the tests were subjected to degradation such as background noise. Further, the processed voice signals, in their digital format, were exposed to bit errors. This process simulates corruption of the signal within a transmission system carrying the signal.

The background noise conditions tested were:

a. Quiet (no additional noise added).

b. ATC room noise (recorded at an ATC Center; average noise level was 73 decibels adjusted (dBa).

c. Cockpit noise (recorded in a Boeing 727 cruising at 14,500 feet; average noise level was 83 dBa).

The bit error rates (BER) tested were:

a. $1x10^{-5}$, which is considered achievable under normal circumstances.

b. $1x10^{-3}$, which is considered the normal operational BER.

c. $1x10^{-2}$, which is considered a worst case. Many CODEC's cannot maintain synchronization at error rates above this error rate.

The three noise and three BER conditions were randomly distributed over the test messages in all three parts of the Phase II tests. Therefore, all CODEC's were evaluated under nine different conditions of noise and error rate in each part.

Each of the three parts contained 36 test messages. They were divided among the nine possible conditions, providing four messages under each of the nine conditions.

A second three-part set of test messages was generated. It consisted of essentially the same text as the first set with slight variations, such as reversal of word pairs in the DRT test. These changes were intended to help prevent listeners from memorizing the messages during one CODEC test and transferring the knowledge to the next CODEC tested. The sets of messages are referred to as Script 1 and Script 2. The combinations of script, message, noise, and BER are shown in appendix B.

In addition to the intelligibility test, the ATC personnel were also asked to fill in a questionnaire regarding the acceptability of the CODEC under test and to compare the CODEC to the current voice communications. The questions are included at the end of the test sheets and are shown in appendix A.

2.3 TEST TAPE GENERATION.

To generate the message recordings for the CODEC testing, the messages were spoken by eight volunteer ATC controllers from the New York Air Route Traffic Control Center (ARTCC) and four pilot volunteers from the FAA Technical Center, Atlantic City International Airport, N.J. The volunteers first recorded test message text with a quiet background. This was followed with recording the messages with background noise (ATC room or cockpit noise) played at the appropriate level in the same room.

The recordings were edited to produce two master recordings corresponding to Script 1 and Script 2. The master recordings were then played through a phone line simulator, which was configured to model an analog telephone link, and then digitized by each CODEC. The digitized voice was then degraded by injecting bit errors using a satellite channel simulator. The digitized voice signal was fed to the receive port of the CODEC. The resulting reconstituted voice output was recorded on an audio cassette. The facility for generating these tapes is discussed in detail in appendix C. This process was repeated for each of the CODEC's under test. A total of 11 CODEC's were tested; one at 9.6 kbps, five at 4.8 kbps, and five at 2.4 kbps. Further, unknown to the listeners, the original, unprocessed voice scripts were evaluated as a means of control for the testing.

The tested CODEC's were obtained under loan from their manufacturers. As part of the loan arrangement, it was agreed that the test results would not be publicly associated with any specific CODEC. During this testing, the CODEC's are identified only by the letters A - M (CODEC F is actually clear voice, not processed through a CODEC). The CODEC algorithms and data rates are shown in table 1.

2.4 TEST ADMINISTRATION.

Both ATC personnel and non-ATC personnel took part in the testing. The non-ATC personnel were CTA Incorporated office staff, and the ATC personnel were controllers from the Washington ARTCC in Leesburg, Virginia, and the Oakland ARTCC in California. In order to minimize the time required for the ATC personnel, only 6 of the original 11 CODEC's were included in the ATC portion of the testing. The non-ATC personnel first evaluated the initial set of 11. Subsequent analysis led to elimination of the three worst performing CODEC's of the 11. Two additional CODEC's were eliminated from further consideration due to limited functional capability. These CODEC's were designated B (2.4 kbps) and D (4.8 kbps). Neither could operate at BER's worse than $1x10^{-5}$.

In the testing, test listeners scored entire script sets of test messages (either Script 1 or 2) as processed through a CODEC. The script-CODEC sets of messages were presented to groups of four listeners at a time. The groups of non-ATC listeners scored two sets of script-CODEC messages. The test facilities are described in appendix C. A total of 24 script-CODEC sets were evaluated using 11 CODEC's plus the clear voice. Each CODEC was evaluated with both Script 1 and 2. ATC listeners evaluated each script-CODEC set, each group of four ATC personnel evaluating three such sets.

At the completion of the testing the tests were examined, and the results entered into a data base. The use of a computerized data base permitted simplified calculation of each CODEC's score with presentations catagorized by background noise and BER conditions.

3. TEST RESULTS.

3.1 NON-ATC RESULTS.

The intelligibility test results for the non-ATC listeners are plotted in figure 1. This figure shows the percentage of incorrect responses for four listeners of Script 1 and four listeners of Script 2 averaged together. The result for the clear voice condition are also included in figure 1. These results average together all of the nine combinations of BER and background noise conditions.

The results for each CODEC under the nine conditions of background noise and BER are provided in table 2. It should be noted that clear voice was not subjected to BER conditions. From this data, it is apparent that CODEC's H, I, and C are the worst performers, especially at the higher BER and background noise conditions. Therefore, CODEC's H, I, and C were eliminated from further testing, leaving six CODEC's for testing by ATC personnel.

TABLE 1. CODEC DATA RATES AND ALGORITHMS

CODEC	DATA RATE (bps)	ALGORITHM
A	4800	STC
B	2400	LPC
C	2400	APC
D	4800	LPC
E	4800	SELP
F	- CLEAR VOICE -	
G	2400	LPC
H	2400	HDLPC
I	4800	APC
J*	9600	SELP
K	2400	STC
M	4800	STC

Algorithms:
- STC - Sinusoidal Transform Coding
- LPC - Linear Predictive Coding
- APC - Adaptive Predictive Coding
- SELP - Stochasticly Excited Linear Prediction
- HDLPC - High Definition Linear Predictive Coding

*Note: This CODEC was evaluated by the AEEC but was not selected for APC use.

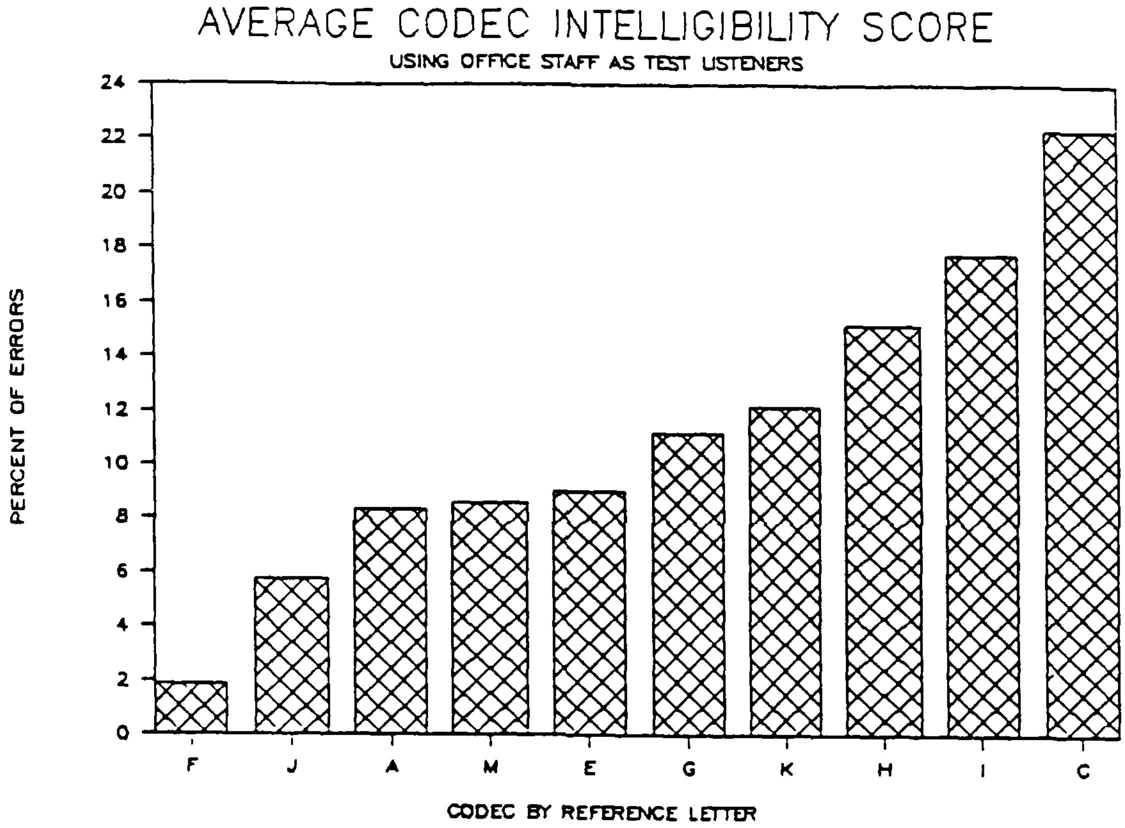


FIGURE 1. NON-ATC INTELLIGIBILITY RESULTS

TABLE 2. NON-ATC INTELLIGIBILITY RESULTS

Background ->	QUIET			COCKPIT			ATC		
BER ->	10^{-5}	10^{-3}	10^{-2}	10^{-5}	10^{-3}	10^{-2}	10^{-5}	10^{-3}	10^{-2}
CODEC ID bps	% of Errors *			% of Errors *			% of Errors *		
F (Clear)	2.5	0.8	0.8	1.7	5.8	1.7	0.0	3.3	0.0
J (9600)	0.8	3.3	4.2	5.8	6.7	12.5	3.2	9.2	6.1
A (4800)	1.7	5.0	6.7	11.7	7.5	10.8	5.6	11.7	14.8
M (4800)	7.3	5.2	11.5	9.4	10.4	10.4	5.0	12.5	5.4
E (4800)	6.9	4.2	5.6	10.2	11.1	11.6	9.8	11.1	11.1
G (2400)	7.5	5.8	8.3	15.8	9.2	19.2	8.8	11.7	14.8
K (2400)	7.5	9.2	6.7	10.8	18.3	14.2	5.6	18.3	19.1
H (2400)	10.8	1.7	5.8	12.5	20.8	16.7	13.6	32.5	22.6
I (4800)	5.0	2.5	12.5	15.8	23.3	29.2	14.4	20.8	37.4
C (2400)	5.8	6.7	15.0	22.5	19.2	45.8	18.4	35.0	33.0

*Note: There was no BER for the clear voice conditions. However, the same test items were used as in the CODEC tests where BERs were injected.

3.2 ATC RESULTS.

The test results for the ATC personnel are plotted in figure 2. This figure shows the average percentage of incorrect responses for all listeners for each CODEC. Sixty ATC controllers participated in the tests. Each CODEC was evaluated by 24 listeners. The results show that the 4.8 kbps CODEC's are within tenths of a percent of the 9.6 kbps CODEC. The results also show that the 2.4 kbps CODEC's performed worse than the 4.8 kbps CODEC's. However, the overall intelligibility of the 2.4 kbps CODEC's still exceeded 88 percent.

The test scores were broken out by BER and background noise for each CODEC. The scores are provided in table 3 and plotted in figure 3. The table and the figure show the expected trend that intelligibility declines as the CODEC's are subjected to combinations of high levels of operational background noise and BER. However, even under the worst conditions, the intelligibility of the 9.6 kbps CODEC's and the 4.8 kbps CODEC's remained high. The performance of the 2.4 kbps CODEC's worsened significantly at high noise and BER.

One apparent contradiction that can be seen in several of the plots is intelligibility scores at $1x10^{-3}$ are slightly worse than the same background noise conditions at $1x10^{-2}$. This may be due to different test messages being used with the different sets of conditions. This is suggested by the plot of results for clear voice messages. As in the non-ATC test result presentation, the plot has columns marked with error rates of $1x10^{-2}$, etc., and the clear voice was not actually subjected to bit errors. The markings are given only for referencing the results to other CODEC plots. As the clear voice plot indicates, the test messages used for CODEC evaluation at a BER of $1x10^{-3}$ were slightly less intelligible than those used at $1x10^{-2}$.

3.3 ATC OPINION SURVEY RESULTS.

At the conclusion of each scoring section, each listener was asked to complete a questionnaire. The questionnaire focused on acceptability and comparison of performance to existing communications systems. The results of the questionnaire are presented in figure 4 with the intelligibility scores. The figure shows that, in the opinion of the controllers, the 9.6 CODEC and the best 4.8 kbps CODEC were nearly equivalent to current communications. It must be pointed out that the controllers were listening to a variety of rapidly changing conditions, including conditions worse than normal operations. This may have affected their opinion scoring.

Most of the controllers used for the testing had a majority of their previous communications experience on very high frequency (VHF) systems. Some of the Oakland controllers also had HF experience. Those controllers who distinguished between VHF and HF voice communications rated the CODEC's considerably higher than HF, but equivalent to or worse than VHF.

The controllers also rated the CODEC's on acceptability (shown in figure 4). The figure shows the best 4.8 kbps CODEC is considered as acceptable as the 9.6 kbps CODEC. Again, it must be noted that the controllers based their opinions of acceptability on listening to the test scripts, which contained rapidly fluctuating conditions, some of which were worse than normal operations. It is also interesting to note that the best 2.4 kbps CODEC was rated nearly as highly as the worst 4.8 kbps CODEC. The other 2.4 kbps CODEC's were rated very poorly.

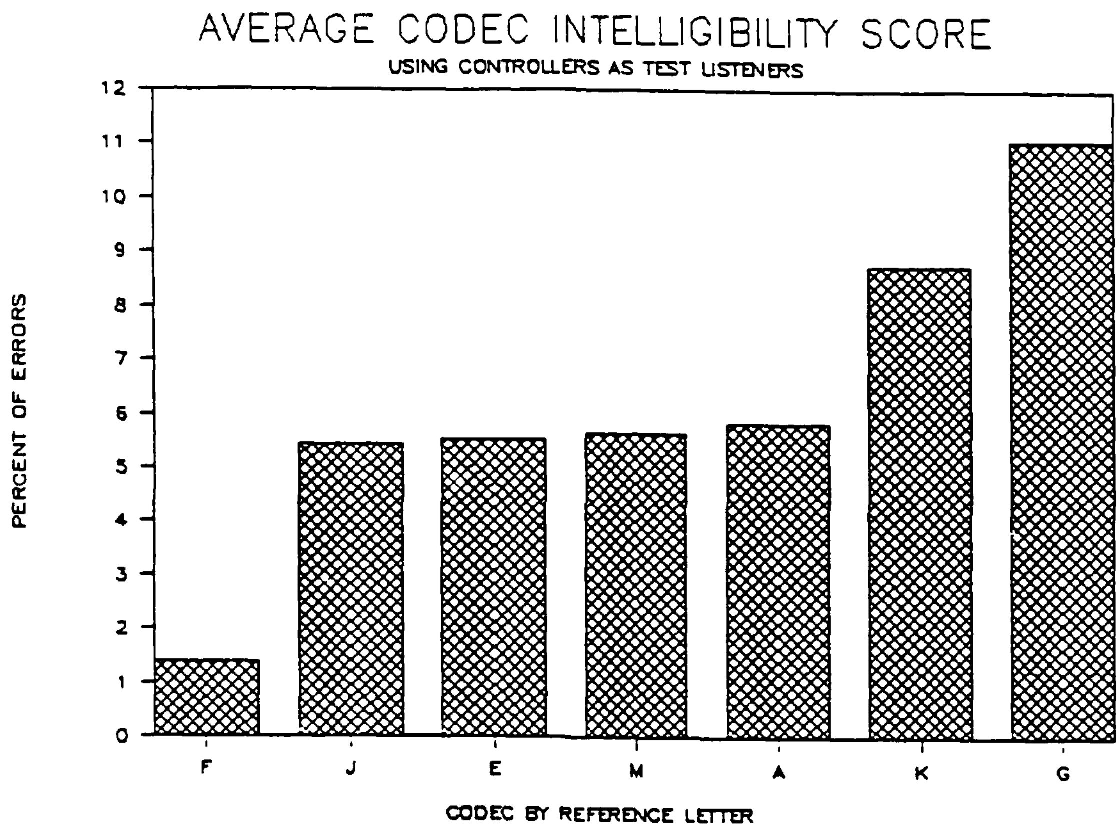


FIGURE 2. ATC INTELLIGIBILITY RESULTS

TABLE 3. ATC INTELLIGIBILITY RESULTS

Background ->	QUIET			COCKPIT			ATC		
BER ->	10^{-5}	10^{-3}	10^{-2}	10^{-5}	10^{-3}	10^{-2}	10^{-5}	10^{-3}	10^{-2}
CODEC ID bps	% of Errors			% of Errors			% of Errors		
F (Clear)	0.3	1.0	0.0	1.0	4.2	1.7	1.3	2.1	0.7
J (9600)	1.4	3.5	2.8	3.1	8.3	10.1	4.0	8.3	7.6
E (4800)	2.1	2.4	4.2	3.8	7.8	8.7	3.8	8.2	8.8
M (4800)	1.4	3.8	2.8	5.2	10.1	10.4	2.3	9.0	5.8
A (4800)	2.1	2.8	4.5	8.0	7.1	7.6	2.4	10.8	6.4
K (2400)	2.4	3.8	4.9	6.9	14.6	13.5	3.3	16.7	13.0
G (2400)	6.9	5.9	5.9	14.9	11.5	15.3	6.0	17.7	15.9

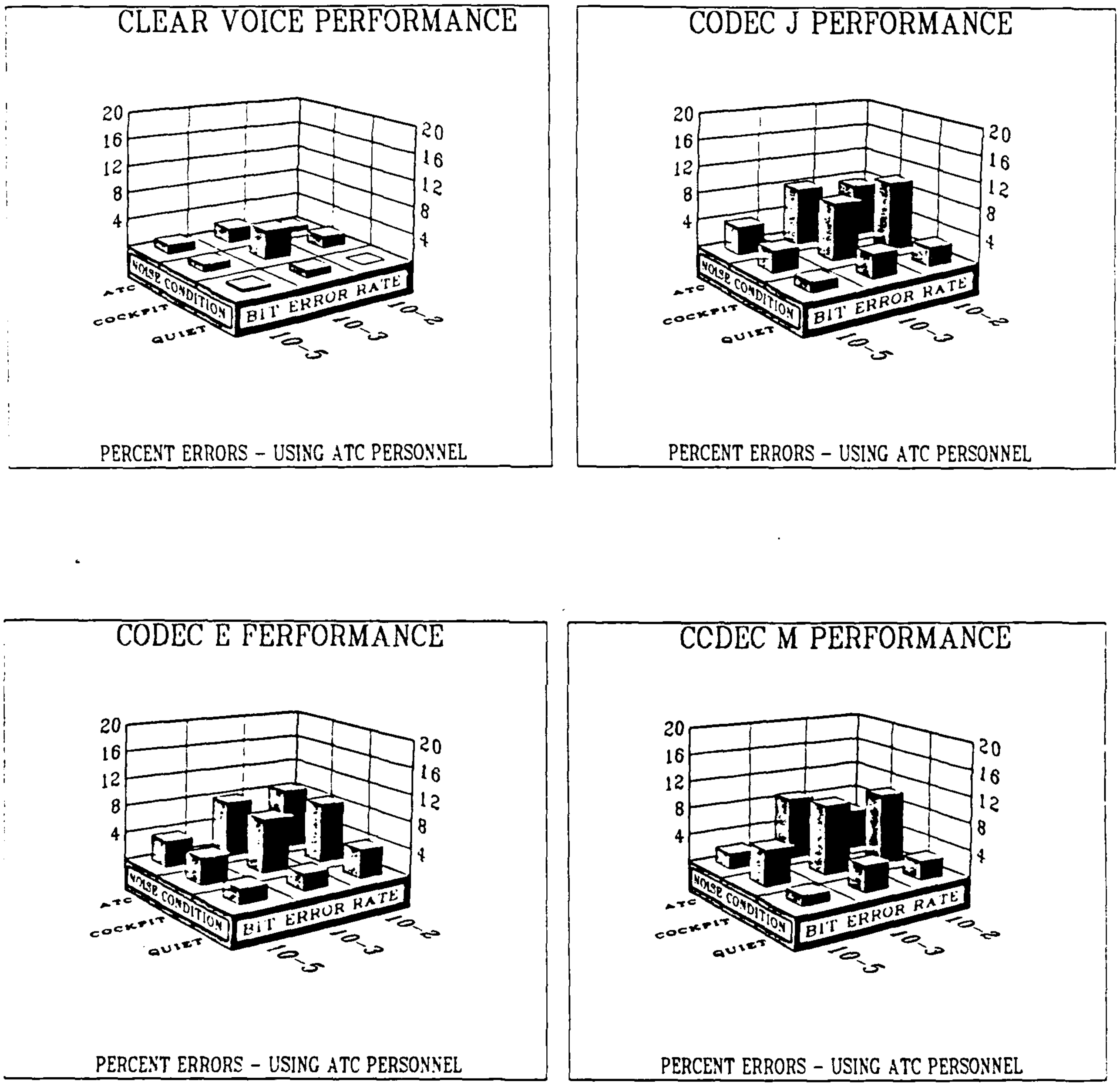
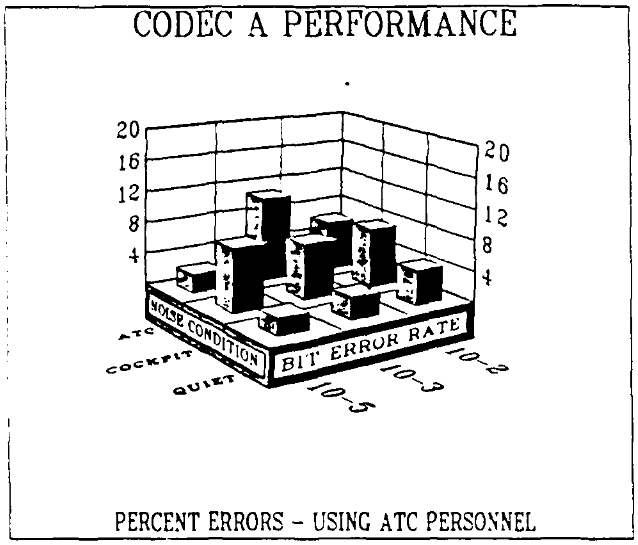
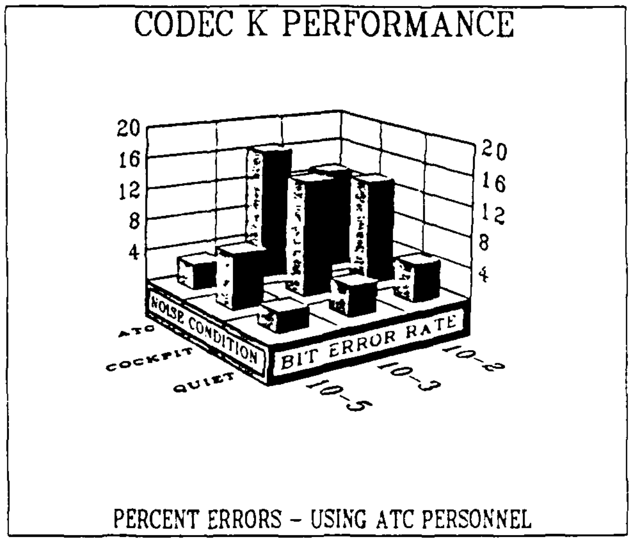


FIGURE 3. ATC INTELLIGIBILITY SCORES CATEGORIZED BY BACKGROUND NOISE AND BER (SHEET 1 OF 2)



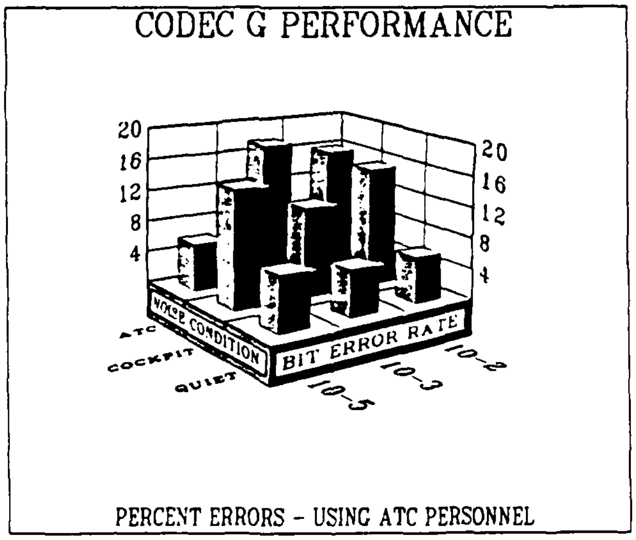
CODEC A PERFORMANCE

PERCENT ERRORS - USING ATC PERSONNEL



CODEC K PERFORMANCE

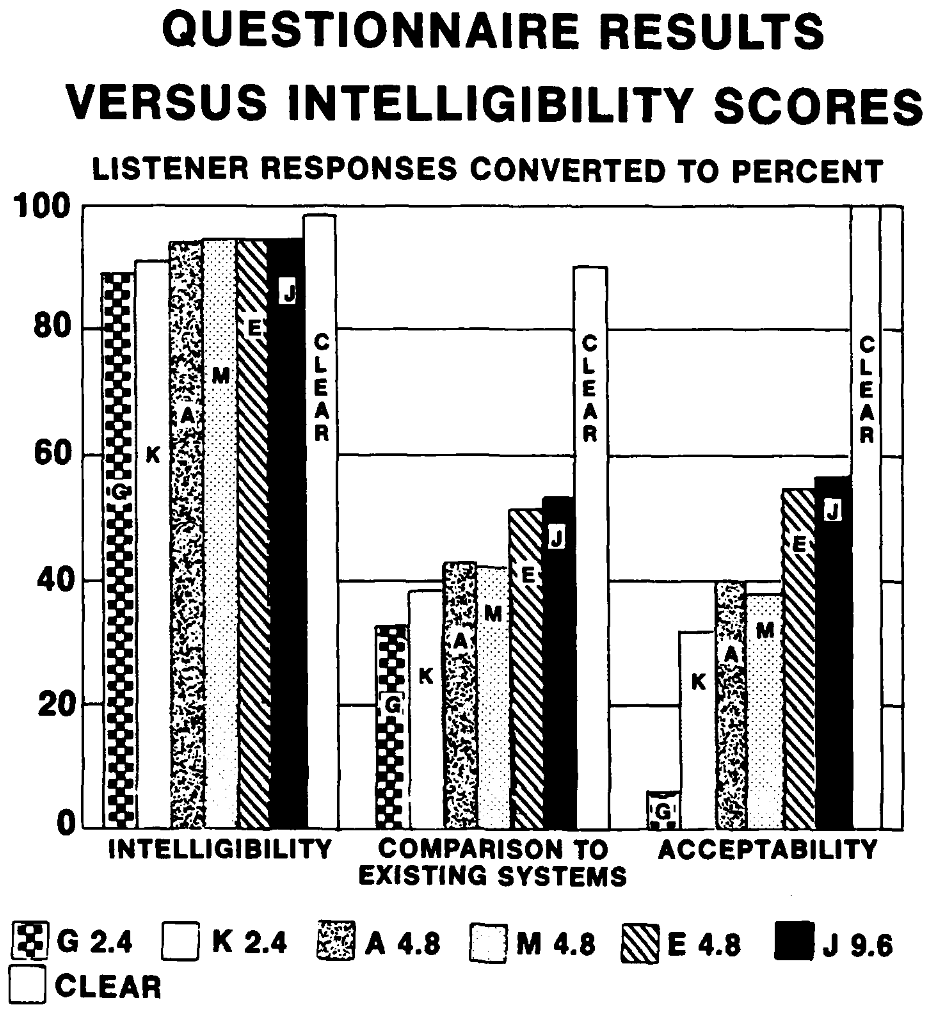
PERCENT ERRORS - USING ATC PERSONNEL



CODEC G PERFORMANCE

PERCENT ERRORS - USING ATC PERSONNEL

FIGURE 3. ATC INTELLIGIBILITY SCORES CATEGORIZED BY BACKGROUND NOISE AND BER (SHEET 2 OF 2)



Notes: 1. Intelligibility scores are shown as percent correct.

2. Code for "comparison to existing" category:

100% - Much better than current communications
75% - Better than current communications
50% - Equivalent to current communications
25% - Worse than current communications
0% - Much worse than current communications

3. Acceptability results show percentage of ATC personnel who considered the CODEC under test to be acceptable for ATC.

FIGURE 4. ATC OPINION RESULTS

3.4 COMPARISON OF ATC RESULTS WITH NON-ATC RESULTS.

Figure 5 compares the test results of the ATC and non-ATC participants. The figure shows the controllers generally had higher intelligibility scores than the non-ATC group; however, the two groups scored the CODEC's similarly in terms of relative performance. This agreement supports the use of the non-ATC scores to eliminate the three worst performing CODEC's as noted in paragraph 3.1.

3.5 COMPARISON OF TEST SECTION RESULTS.

Intelligibility testing was divided into three parts, representing three levels of "openness." Part 1 was a pseudo-DRT test, Part 2 contained numerals in context, and Part 3 contained isolated words. Figure 6 shows the percentage of total ATC test errors broken out by part. The figure shows the "closed" DRT-like Part 1 received approximately 22 percent of the errors, the partially "open" Part 2 received 30 percent of the errors, and the more "open" Part 3 received over 47 percent of the errors. These results agree with industry experience of words being more easily misinterpreted in an "open" test than a "closed" test.

4. CONCLUSIONS.

The following conclusions can be drawn from the results of the Phase II coder/decoder (CODEC) test program:

a. That the intelligibility of the 4.8 kilobits per second (kbps) CODEC's is nearly as high as the 9.6 CODEC in an air traffic control (ATC) environment. The 9.6 kbps CODEC was an improved version of one of the CODEC's evaluated in the Airline Electronics Engineering Committee (AEEC) selection of a CODEC for Aeronautical Passenger Communications (APC) use.

b. That the acceptability of the 4.8 kbps CODEC's range from approximately equivalent to slightly worse than the 9.6 kbps CODEC.

c. That the intelligibility of the 4.8 kbps CODEC's and the 9.6 kbps CODEC declined under conditions combining bit error rates (BER's) and operational background noise; however, even under these conditions the intelligibility remained high (better than 89 percent).

d. That the opinion of the ATC personnel was split over the acceptability of the 9.6 kbps and the 4.8 kbps CODEC's, with roughly half of the controllers stating that the CODEC's were acceptable for ATC use.

e. That the intelligibility of the 2.4 kbps CODEC's was low, especially under conditions of BER and background noise; the acceptability was also low.

5. RECOMMENDATIONS.

It is recommended that a Phase III coder/decoder (CODEC) test program be initiated to:

a. Evaluate the acceptability of available 4.8 kilobits per second (kbps) CODEC's compared to standard air traffic control (ATC) voice communications. This includes very high frequency (VHF), high frequency (HF), and the 9.6 kbps

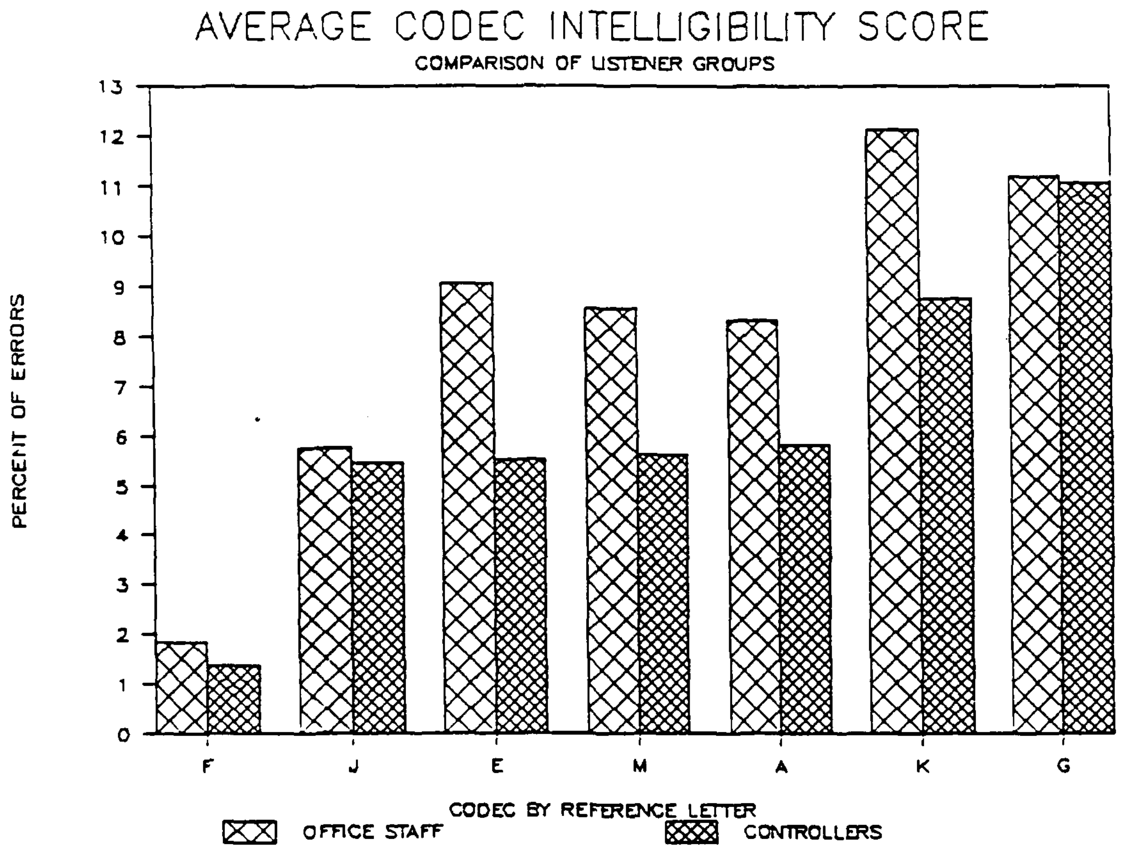


FIGURE 5. COMPARISON OF NON-ATC AND ATC RESULTS

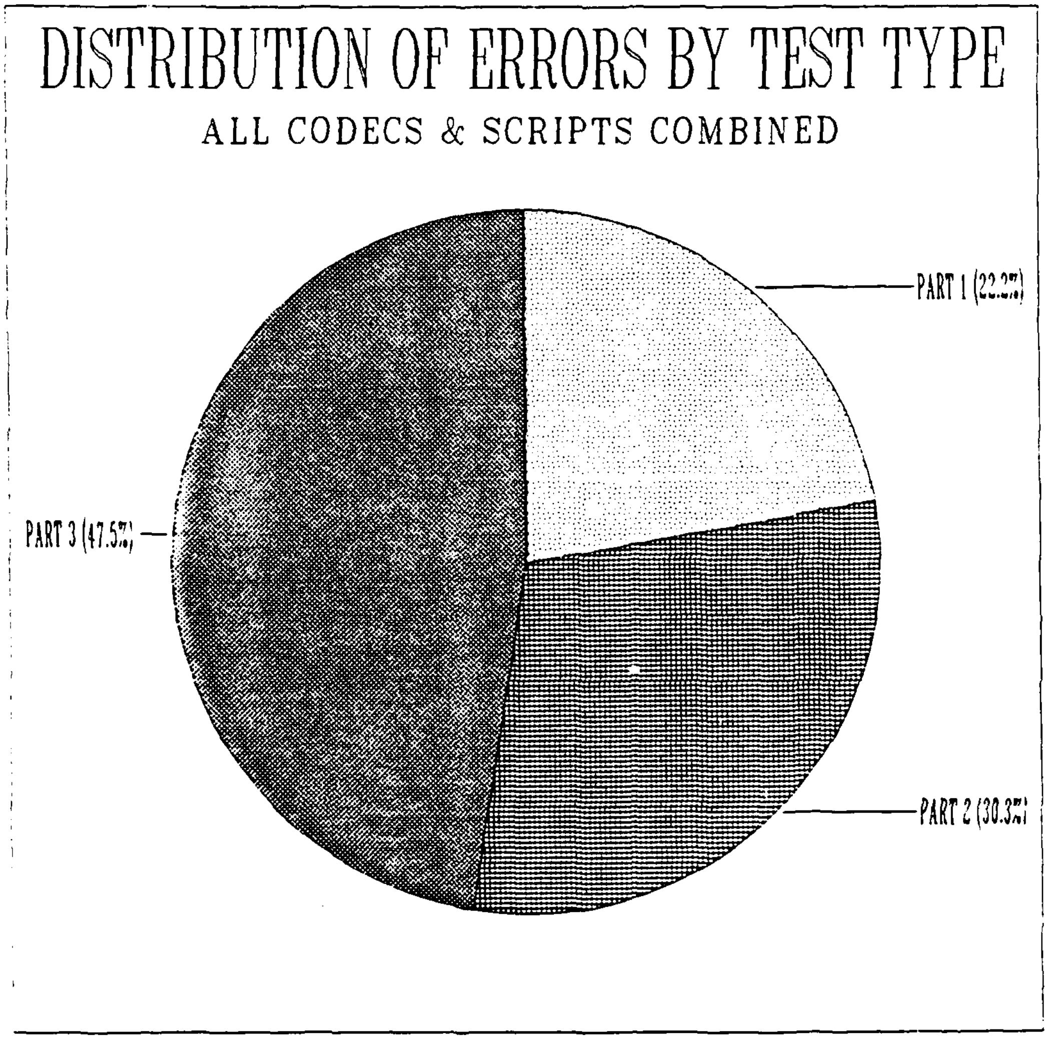


FIGURE 6. DISTRIBUTION OF ATC RESULTS OVER SECTION

CODEC which was selected by Airline Electronics Engineering Committee (AEEC) for Aeronautical Passenger Communications (APC) use.

b. Develop a test and criteria for selecting a 4.8 kbps CODEC for ATC use. This selection test would be based, in part, on an objective intelligibility test (similar to Phase II testing) and a subjective acceptability test.

Additionally, it is recommended:

a. That the test developed in the Phase III program be used to select a 4.8 kbps CODEC for ATC use.

b. That the algorithm of the selected 4.8 kbps CODEC be proposed for worldwide standardization for ATC use.

APPENDIX A - TEST SHEETS

This appendix contains the test sheets and the correct answers used for Script 1 and Script 2 of the Phase II intelligibility test.

FAA PHASE 2 SUBJECTIVE LISTENING EVALUATION

NAME : ______________________________

GROUP/SECTION : ______________________________

DATE : ______________________________

TIME : ______________________________

YEARS EXPERIENCE : ______________________________

TYPE EXPERIENCE : ______________________________

____

____

AGE (circle one) :

20 - 24
25 - 29
30 - 34
35 - 39
40 - 44
45 - 49
50 - 54
> 55

PART 1 : WORD PAIR TEST

1) SECTOR ____ VECTOR ____

2) BRAIN ____ RAIN ____

3) SIX ____ TRICKS ____

4) SEVEN ____ HEAVEN ____

5) ROCK ____ MACH ____

6) REPORT ____ RESORT ____

7) FLIGHT ____ HEIGHT ____

8) FOUR ____ FIVE ____

9) ONE ____ FUN ____

10) RIGHT ____ LIGHT ____

11) TWO ____ THREE ____

12) WEST ____ TEST ____

13) NINE ____ WINE ____

14) ALTITUDE ____ ATTITUDE ____

15) SOW ____ SOUTH ____

16) ZERO ____ HERO ____

17) TURN ____ BURN ____

18) CENTER ____ ENTER ____

19) TREE ____ THREE ____

20) PEACH ____ REACH ____

21) RUNWAY ____ SUNDAY ____

22) LEAR ____ CLEAR ____

23) SMILES ____ MILES ____

24) END ____ SEND ____

25) RHYME ____ CLIMB ____

26) DIVE ____ FIVE ____

27) DATE ____ EIGHT ____

28) HELL ____ HAIL ____

29) DANGER ____ RANGER ____

30) STOP ____ DROP ____

31) FIRE ____ HIGHER ____

32) DAMPER ____ DANDER ____

33) SLOW ____ FLOW ____

34) FRONT ____ RUNT ____

35) CLOUDS ____ CROWDS ____

36) INTERCEPT ____ INCORRECT ____

PART 2: FILL IN THE BLANK WITH #'S TEST

01. Continental __________, climb and maintain FL 290.
02. Interstate __________, Roger.
03. This is Falcon __________ Charlie Delta at flight level 270.
04. Power __________, say again.
05. Tango Aztec __________ Alpha, how do you read?
06. Intercept the Deer Park __________ radial.
07. Leaving flight level __________ for 290.
08. Bonanza __________ Golf, Roger.
09. Engine failure on Northwest __________ .
10. Eastern 731 contact Atlanta Center on __________ .
11. Hold southwest on the Hampton __________ radial.
12. Expect further clearance at __________ .
13. Climb and maintain flight level __________ .
14. Talan 3 reaching flight level __________ .
15. Any ride reports at flight level __________ southbound?
16. Cleared to land runway __________ left.
17. Bearcat __________ Yankee Romeo, Roger.
18. We are VFR at __________ .
19. Piedmont __________, cleared direct Boston.
20. Delta __________ at flight level 330.
21. Follow the American __________ .
22. Contact New York Center on __________ .
23. Beechcraft __________ Romeo Lima to flight level 310.
24. Intercept Jay __________ on your present heading.
25. Reaching and maintaining flight level __________ .
26. This is Northwest __________, out of FL 270 for FL 280.
27. Baron __________ Hotel, requesting traffic advisories.
28. United __________ have a good day.
29. Request flight level __________ .
30. United 60, climb and maintain flight level __________ .
31. Contact San Juan on __________ .
32. Contact New York Center __________ .
33. Cherokee __________ Alpha over Kansas City.
34. This is November __________ Romeo.
35. Increase to mach point __________ .
36. Traffic 3 o'clock __________ miles southbound.

PART 3 : WORD RECOGNITION TEST

01. ______________________
02. ______________________
03. ______________________
04. ______________________
05. ______________________
06. ______________________
07. ______________________
08. ______________________
09. ______________________
10. ______________________
11. ______________________
12. ______________________
13. ______________________
14. ______________________
15. ______________________
16. ______________________
17. ______________________
18. ______________________
19. ______________________
20. ______________________
21. ______________________
22. ______________________
23. ______________________
24. ______________________
25. ______________________
26. ______________________
27. ______________________
28. ______________________
29. ______________________
30. ______________________
31. ______________________
32. ______________________
33. ______________________
34. ______________________
35. ______________________
36. ______________________

PART 4: QUESTIONNAIRE

1. Overall, would you say that what you have just listened to is (circle one)

   a. much better than
   b. slightly better than
   c. about the same as
   d. slightly worse than
   e. much worse than

   your present form of communications.

2. Overall, would this form of communications be acceptable to you in your present environment (circle one)?

   a. yes
   b. no

3. Did the presence of background noise affect your answer in question 2 (circle one)?

   a. yes
   b. no

4. Please comment on the similarities and/or differences between what you have just heard and your present form of communications.

5. Any further comments on this evaluation?

SOLUTIONS TO SECTION 1
PHASE 2 CODEC TEST

Solutions to Script 1

01.	2/1
02.	1/2
03.	1/2
04.	2/1
05.	2/1
06.	2/1
07.	2/1
08.	1/2
09.	2/1
10.	1/2
11.	2/1
12.	1/2
13.	1/2
14.	1/2
15.	2/1
16.	2/1
17.	1/2
18.	2/1
19.	1/2
20.	1/2
21.	1/2
22.	2/1
23.	1/2
24.	2/1
25.	1/2
26.	1/2
27.	2/1
28.	1/2
29.	2/1
30.	2/1
31.	2/1
32.	2/1
33.	2/1
34.	1/2
35.	1/2
36.	1/2

Solutions to Script 2

01.	2/1
02.	1/2
03.	2/1
04.	1/2
05.	2/1
06.	2/1
07.	1/2
08.	1/2
09.	2/1
10.	1/2
11.	2/1
12.	2/1
13.	1/2
14.	1/2
15.	2/1
16.	2/1
17.	1/2
18.	2/1
19.	2/1
20.	2/1
21.	1/2
22.	1/2
23.	2/1
24.	1/2
25.	1/2
26.	2/1
27.	1/2
28.	2/1
29.	2/1
30.	1/2
31.	1/2
32.	2/1
33.	1/2
34.	2/1
35.	1/2
36.	1/2

SOLUTIONS TO SECTION 2
PHASE 2 CODEC TEST

Solutions to Script 1

01.	125
02.	953
03.	41
04.	98
05.	4723
06.	221
07.	390
08.	8231
09.	34
10.	133.45
11.	264
12.	2040
13.	260
14.	380
15.	250
16.	24
17.	5
18.	17500
19.	953
20.	845
21.	747
22.	134.5
23.	759
24.	121
25.	330
26.	323
27.	4365
28.	8
29.	280
30.	350
31.	125.8
32.	128.3
33.	351
34.	6731
35.	78
36.	20

Solutions to Script 2

01.	152
02.	.955
03.	51
04.	89
05.	4723
06.	212
07.	360
08.	8123
09.	43
10.	123.45
11.	462
12.	2050
13.	270
14.	320
15.	240
16.	42
17.	6
18.	16500
19.	552
20.	435
21.	727
22.	134.6
23.	781
24.	212
25.	220
26.	233
27.	4365
28.	2
29.	270
30.	340
31.	124.8
32.	132.8
33.	315
34.	6771
35.	75
36.	20

SOLUTIONS TO SECTION 3
PHASE 2 CODEC TEST

Solutions to Script 1

01. Chart
02. Boston
03. Position
04. Nine
05. Copilot
06. Contact
07. Seven
08. Caution
09. Clear
10. Three
11. One
12. Point
13. Mayday
14. Plenty
15. Five
16. West
17. North
18. Oakland
19. Flight
20. Light
21. Feather
22. Forty
23. Satellite
24. Zero
25. Airspace
26. Radar
27. Degree
28. Navigation
29. Depart
30. Continue
31. Cruise
32. Eleven
33. Atlantic
34. Separation
35. Station
36. Traffic

Solutions to Script 2

01. Two
02. Surface
03. Downwind
04. Tower
05. Failure
06. Ocean
07. Pilot
08. Four
09. Clock
10. Area
11. Level
12. Eight
13. Twenty
14. Wind Shear
15. Heavy
16. Control
17. Zone
18. Thirty
19. Ground
20. Turn
21. Takeoff
22. Pacific
23. Request
24. Alert
25. Visual
26. Good Morning
27. Discontinue
28. Weather
29. New York
30. Baltimore
31. Roger
32. Hijack
33. London
34. Northeast
35. Target
36. Update

APPENDIX B - TEST ITEM CONDITIONS

This appendix contains the list of conditions for each test item on each script. For example, the first entry on the list for Script 1 is for Scenario 22. The scenario number corresponds to a test item for each section of the test. By referring to the test in appendix A, the test item for Part 1, Number 22, is the word pair "lear" and "clear." Referring again to the list of conditions, it is indicated that this word pair was spoken by a female with quiet background condition, and that a BER of $1x10^{-3}$ was introduced. The BER was adjusted for each section of the test, as is shown in the last column.

TABLE B-1.1. TEST CONDITIONS FOR EACH ITEM ON SCRIPT 1

SCENARIO	VOICE GENDER	BACKGROUND NOISE	VOICE TYPE	ERROR		
				section 1	section 2	section 3
22	FEMALE 1	QUIET	----	1E-3	1E-2	1E-3
04	FEMALE 2	QUIET	----	1E-5	1E-3	1E-2
11	FEMALE 3	QUIET	----	1E-5	1E-3	1E-2
16	FEMALE 4	QUIET	----	1E-3	1E-2	1E-3
13	MALE 1	QUIET	----	1E-3	1E-2	1E-3
28	MALE 2	QUIET	----	1E-2	1E-5	1E-5
33	MALE 3	QUIET	----	1E-2	1E-5	1E-5
07	MALE 4	QUIET	----	1E-5	1E-3	1E-2
25	MALE 5	QUIET	----	1E-2	1E-5	1E-5
21	MALE 6	QUIET	----	1E-3	1E-2	1E-3
02	MALE 7	QUIET	----	1E-5	1E-3	1E-2
35	MALE 8	QUIET	----	1E-2	1E-5	1E-5
01	FEMALE 1	ATC ROOM	ATC	1E-5	1E-3	1E-2
12	FEMALE 2	ATC ROOM	ATC	1E-5	1E-3	1E-2
06	FEMALE 3	ATC ROOM	ATC	1E-5	1E-3	1E-2
24	FEMALE 4	ATC ROOM	ATC	1E-3	1E-2	1E-3
10	MALE 1	ATC ROOM	ATC	1E-5	1E-3	1E-2
31	MALE 2	ATC ROOM	ATC	1E-2	1E-5	1E-5
30	MALE 3	ATC ROOM	ATC	1E-2	1E-5	1E-5
08	MALE 4	ATC ROOM	ATC	1E-5	1E-3	1E-2
19	MALE 5	ATC ROOM	ATC	1E-3	1E-2	1E-3
23	MALE 6	ATC ROOM	ATC	1E-3	1E-2	1E-3
36	MALE 7	ATC ROOM	ATC	1E-2	1E-5	1E-5
32	MALE 8	ATC ROOM	ATC	1E-2	1E-5	1E-5
26	FEMALE 1	COCKPIT	PILOT	1E-2	1E-5	1E-5
29	FEMALE 2	COCKPIT	PILOT	1E-2	1E-5	1E-5
09	FEMALE 3	COCKPIT	PILOT	1E-5	1E-3	1E-2
18	FEMALE 4	COCKPIT	PILOT	1E-3	1E-2	1E-3
17	MALE 1	COCKPIT	PILOT	1E-3	1E-2	1E-3
15	MALE 2	COCKPIT	PILOT	1E-3	1E-2	1E-3
03	MALE 3	COCKPIT	PILOT	1E-5	1E-3	1E-2
34	MALE 4	COCKPIT	PILOT	1E-2	1E-5	1E-5
14	MALE 5	COCKPIT	PILOT	1E-3	1E-2	1E-3
20	MALE 6	COCKPIT	PILOT	1E-3	1E-2	1E-3
27	MALE 7	COCKPIT	PILOT	1E-2	1E-5	1E-5
05	MALE 8	COCKPIT	PILOT	1E-5	1E-3	1E-2

TABLE B-1.2. TEST CONDITIONS FOR EACH ITEM ON SCRIPT 2

SCENARIO	VOICE GENDER	BACKGROUND NOISE	VOICE TYPE	ERROR		
				section 1	section 2	section 3
22	FEMALE 1	QUIET	—	1E-2	1E-5	1E-5
04	FEMALE 2	QUIET	—	1E-5	1E-3	1E-2
11	FEMALE 3	QUIET	—	1E-5	1E-3	1E-2
16	FEMALE 4	QUIET	—	1E-2	1E-5	1E-5
13	MALE 1	QUIET	—	1E-2	1E-5	1E-5
28	MALE 2	QUIET	—	1E-3	1E-2	1E-3
33	MALE 3	QUIET	—	1E 3	1E-2	1E-3
07	MALE 4	QUIET	—	1E-5	1E-3	1E-2
25	MALE 5	QUIET	—	1E-3	1E-2	1E-3
21	MALE 6	QUIET	—	1E-2	1E-5	1E-5
02	MALE 7	QUIET	—	1E-5	1E-3	1E-2
35	MALE 8	QUIET	—	1E-3	1E-2	1E-3
01	FEMALE 1	ATC ROOM	ATC	1E-5	1E-3	1E-2
12	FEMALE 2	ATC ROOM	ATC	1E-5	1E-3	1E-2
06	FEMALE 3	ATC ROOM	ATC	1E-5	1E-3	1E-2
24	FEMALE 4	ATC ROOM	ATC	1E-2	1E-5	1E-5
10	MALE 1	ATC ROOM	ATC	1E-5	1E-3	1E-2
31	MALE 2	ATC ROOM	ATC	1E-3	1E-2	1E-3
30	MALE 3	ATC ROOM	ATC	1E-3	1E-2	1E-3
08	MALE 4	ATC ROOM	ATC	1E-5	1E-3	1E-2
19	MALE 5	ATC ROOM	ATC	1E-2	1E-5	1E-5
23	MALE 6	ATC ROOM	ATC	1E-2	1E-5	1E-5
36	MALE 7	ATC ROOM	ATC	1E-3	1E-2	1E-3
32	MALE 8	ATC ROOM	ATC	1E-3	1E-2	1E-3
26	FEMALE 1	COCKPIT	PILOT	1E-3	1E-2	1E-3
29	FEMALE 2	COCKPIT	PILOT	1E-3	1E-2	1E-3
09	FEMALE 3	COCKPIT	PILOT	1E-5	1E-3	1E-2
18	FEMALE 4	COCKPIT	PILOT	1E-2	1E-5	1E-5
17	MALE 1	COCKPIT	PILOT	1E-2	1E-5	1E-5
15	MALE 2	COCKPIT	PILOT	1E-2	1E-5	1E-5
03	MALE 3	COCKPIT	PILOT	1E-5	1E-3	1E-2
34	MALE 4	COCKPIT	PILOT	1E-3	1E-2	1E-3
14	MALE 5	COCKPIT	PILOT	1E-2	1E-5	1E-5
20	MALE 6	COCKPIT	PILOT	1E-2	1E-5	1E-5
27	MALE 7	COCKPIT	PILOT	1E-3	1E-2	1E-3
05	MALE 8	COCKPIT	PILOT	1E-5	1E-3	1E-2

APPENDIX C - TEST FACILITIES FOR RECORDING AND LISTENING

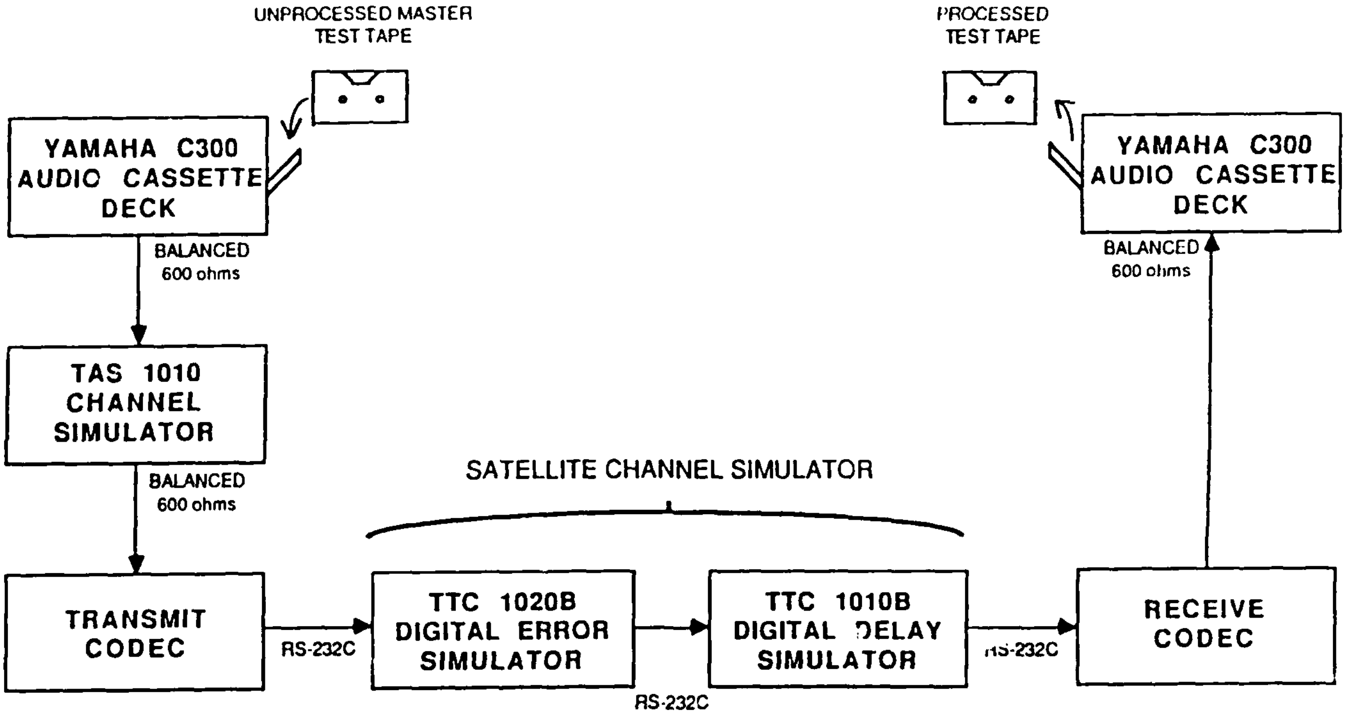
This appendix provides a description of the test setups used in Phase II for generating the test tapes and for conducting the listening tests.

Figure C-1.1 shows the equipment setup for the test tape generation. The unprocessed master tape contains recordings of the test scripts in the appropriate background environments. The recording is first played through the TAS 1010 phone line simulator, which was configured to simulate an "Average Bell" 2-wire circuit.

The analog output of the phone line simulator was then converted by the transmit CODEC into a digital bit stream, which was input into the TTC 1020B Digital Error and Delay simulators. These simulators injected the appropriate BER and introduced the 270 ms delay representing the satellite link. The digital bit stream was converted to an analog signal by the receive CODEC, which was recorded.

One tape was recorded for each CODEC on each script. These tapes were then used to conduct the Phase II tests.

Figure C-1.2 shows the setup for the listening tests. Each CODEC tape for each script was played for four listeners simultaneously; however, each listener could select the listening level they preferred.



NOTE: CODECs A, K, & M were configured without the delay simulator. CODEC H was configured without the error and delay simulator.

FIGURE C-1.1. PHASE II TAPE GENERATION EQUIPMENT SETUP

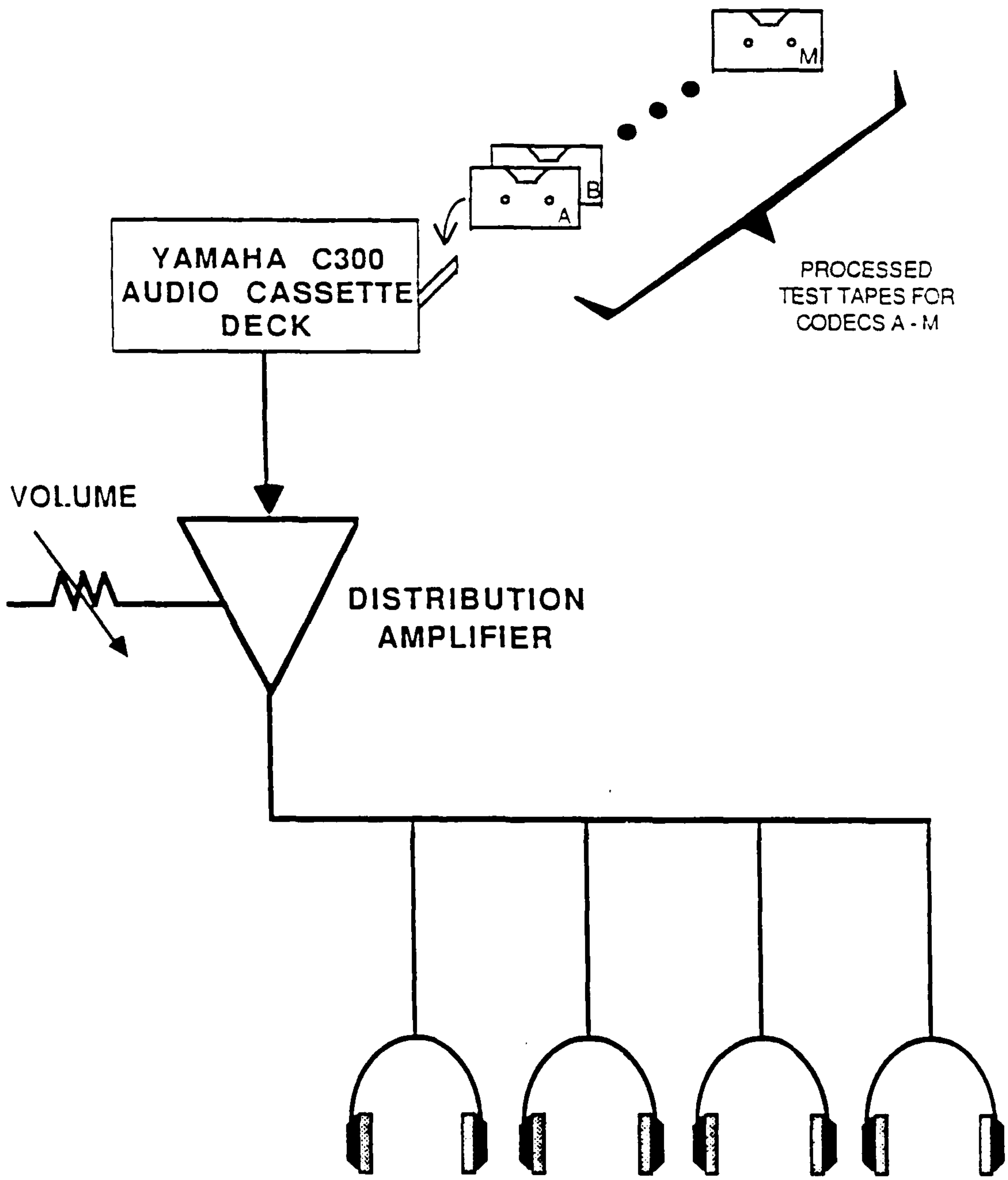


FIGURE C-1.2. PHASE II LISTENING TEST SETUP