NRSC REPORT

NATIONAL RADIO SYSTEMS COMMITTEE

NRSC-R58
Digital Audio Radio IBOC Laboratory Tests
Transmission Quality Failure Characterization and Analog Compatibility
August 11, 1995

Part IV – Appendices Q through W

NAB
NATIONAL ASSOCIATION OF BROADCASTERS

CEA
Consumer Electronics Association

NAB: 1771 N Street, N.W.
Washington, DC 20036
Tel: (202) 429-5356 Fax: (202) 775-4981

CEA: 1919 South Eads Street
Arlington, VA 22202
Tel: (703) 907-7660 Fax: (703) 907-8113

Co-sponsored by the Consumer Electronics Association and the National Association of Broadcasters
http://www.nrscstandards.org

NOTICE

NRSC Standards, Guidelines, Reports and other technical publications are designed to serve the public interest through eliminating misunderstandings between manufacturers and purchasers, facilitating interchangeability and improvement of products, and assisting the purchaser in selecting and obtaining with minimum delay the proper product for his particular need. Existence of such Standards, Guidelines, Reports and other technical publications shall not in any respect preclude any member or nonmember of the Consumer Electronics Association (CEA) or the National Association of Broadcasters (NAB) from manufacturing or selling products not conforming to such Standards, Guidelines, Reports and other technical publications, nor shall the existence of such Standards, Guidelines, Reports and other technical publications preclude their voluntary use by those other than CEA or NAB members, whether to be used either domestically or internationally.

Standards, Guidelines, Reports and other technical publications are adopted by the NRSC in accordance with the NRSC patent policy. By such action, CEA and NAB do not assume any liability to any patent owner, nor do they assume any obligation whatever to parties adopting the Standard, Guideline, Report or other technical publication.

This Report does not purport to address all safety problems associated with its use or all applicable regulatory requirements. It is the responsibility of the user of this Report to establish appropriate safety and health practices and to determine the applicability of regulatory limitations before its use.

Published by
CONSUMER ELECTRONICS ASSOCIATION
Technology & Standards Department
1919 S. Eads St.
Arlington, VA 22202

NATIONAL ASSOCIATION OF BROADCASTERS
Science and Technology Department
1771 N Street, NW
Washington, DC 20036

FOREWORD

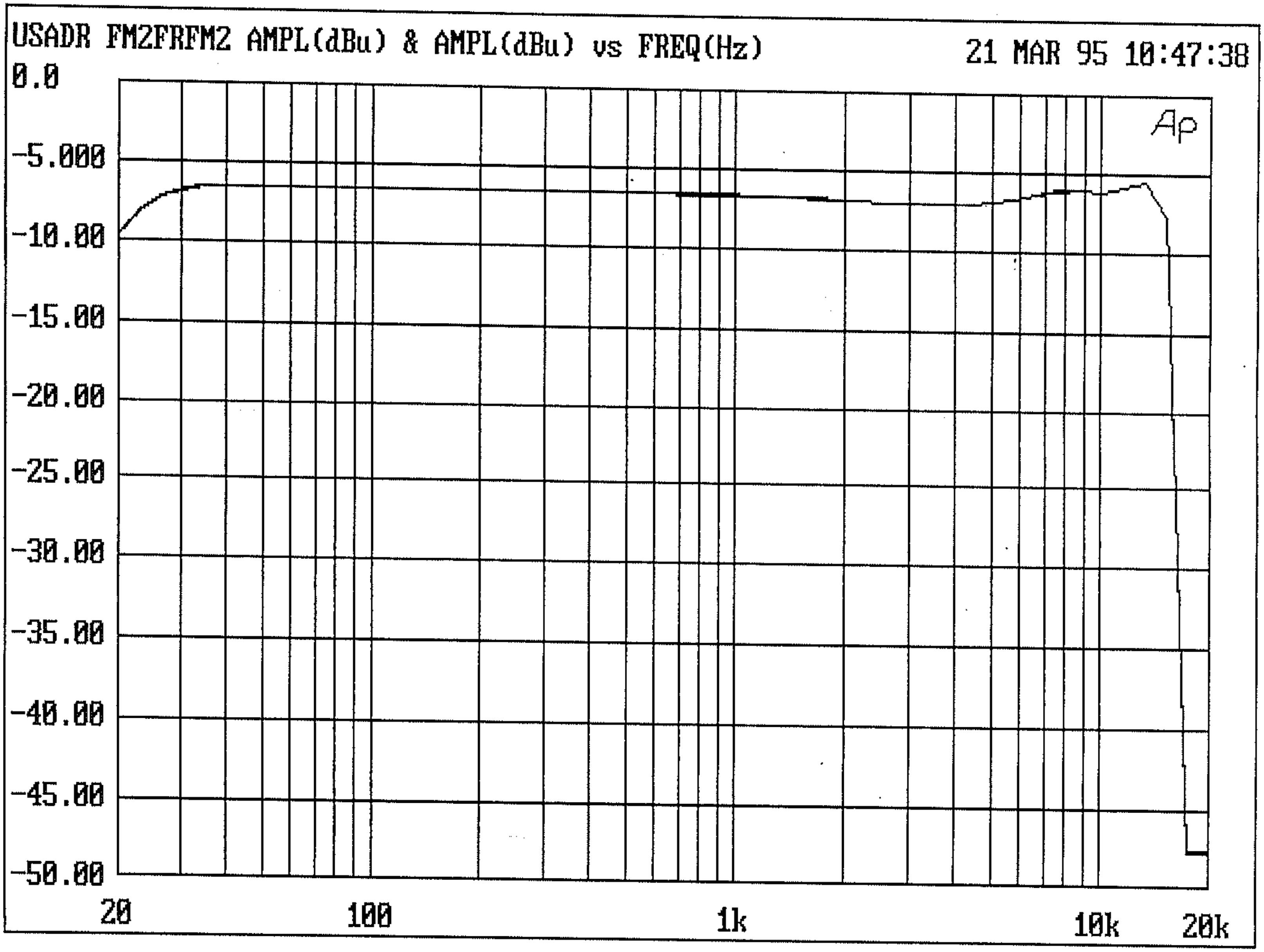
NRSC-R58, *Digital Audio Radio IBOC Laboratory Tests – Transmission Quality Failure Characterization and Analog Compatibility*, documents the first comprehensive testing of in-band/on-channel digital radio systems. This report was prepared for Working Group B and the Combined EIA DAR and NRSC DAB Subcommittees.

The NRSC is jointly sponsored by the Consumer Electronics Association and the National Association of Broadcasters. It serves as an industry-wide standards-setting body for technical aspects of terrestrial over-the-air radio broadcasting systems in the United States.

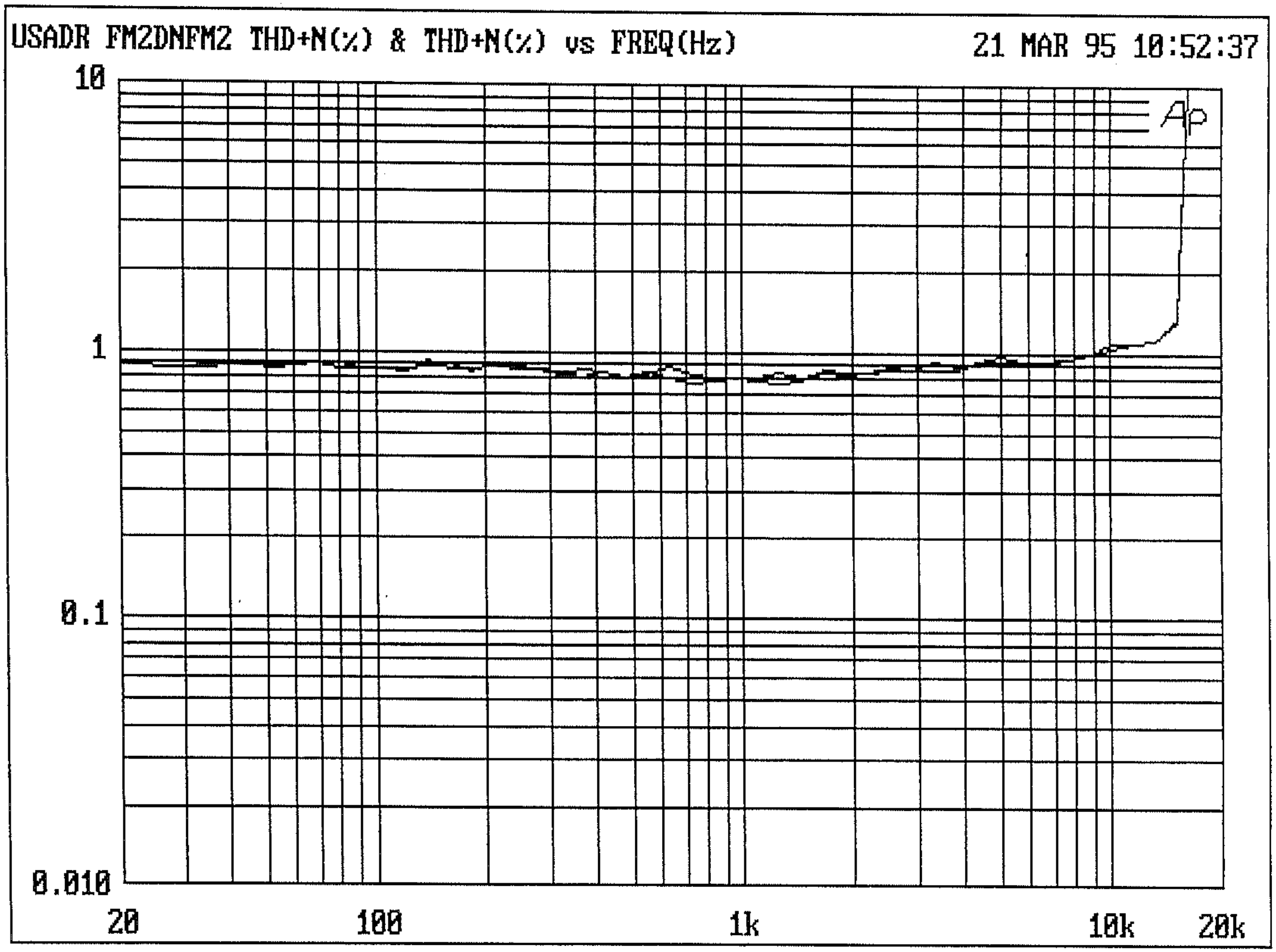
Contents

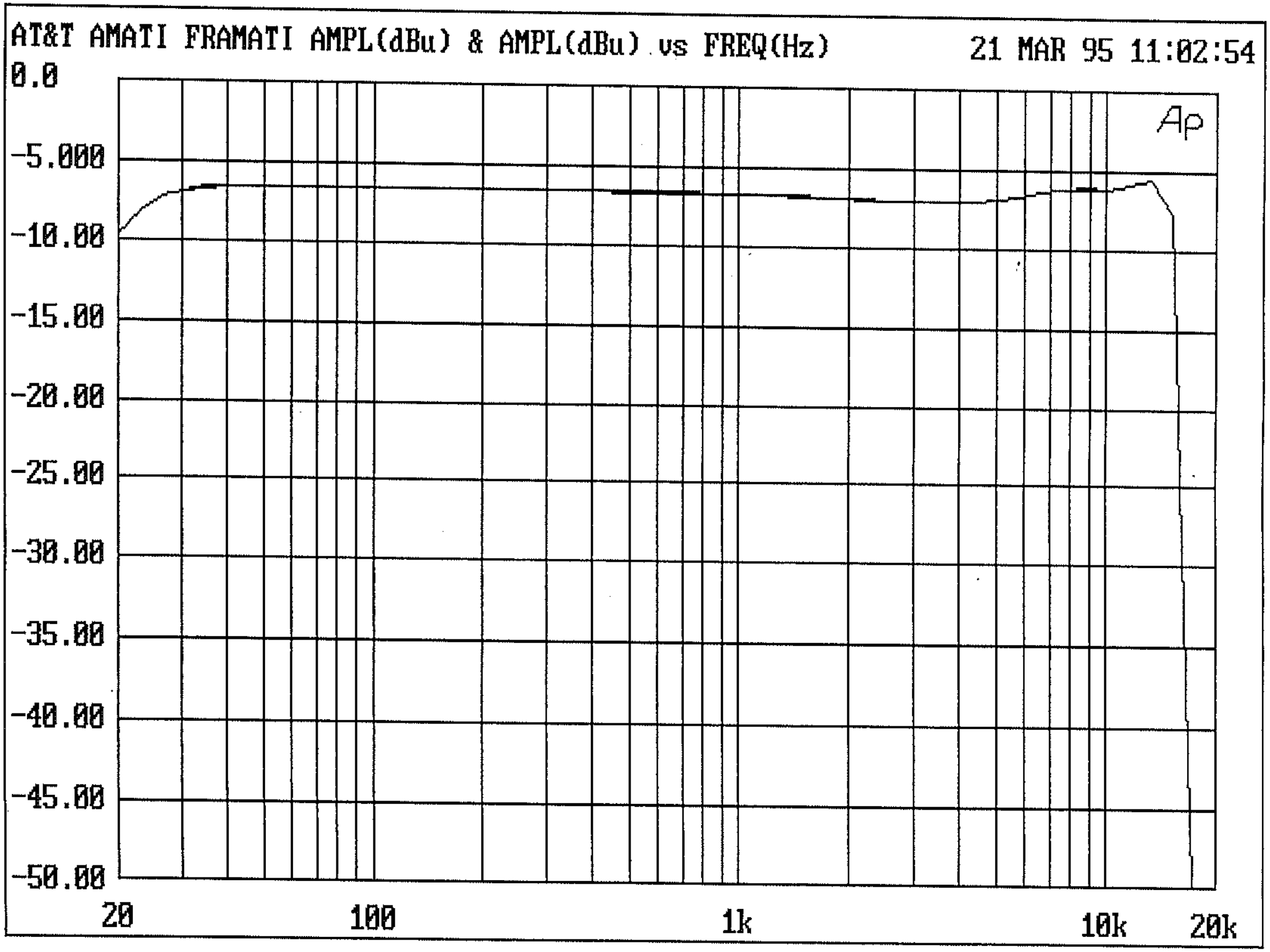
APPENDIX Q

FM Analog Transmitter Proofs

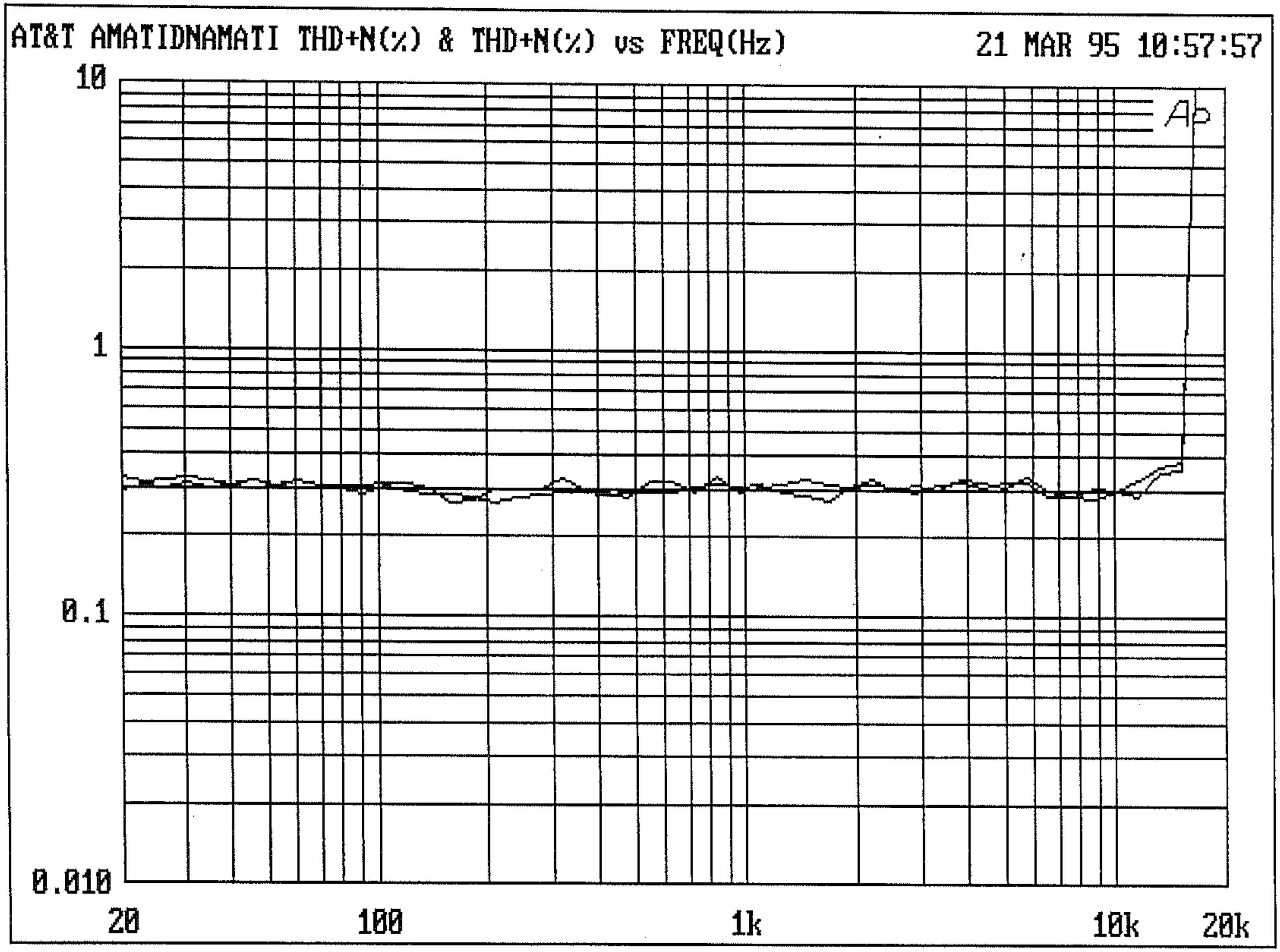


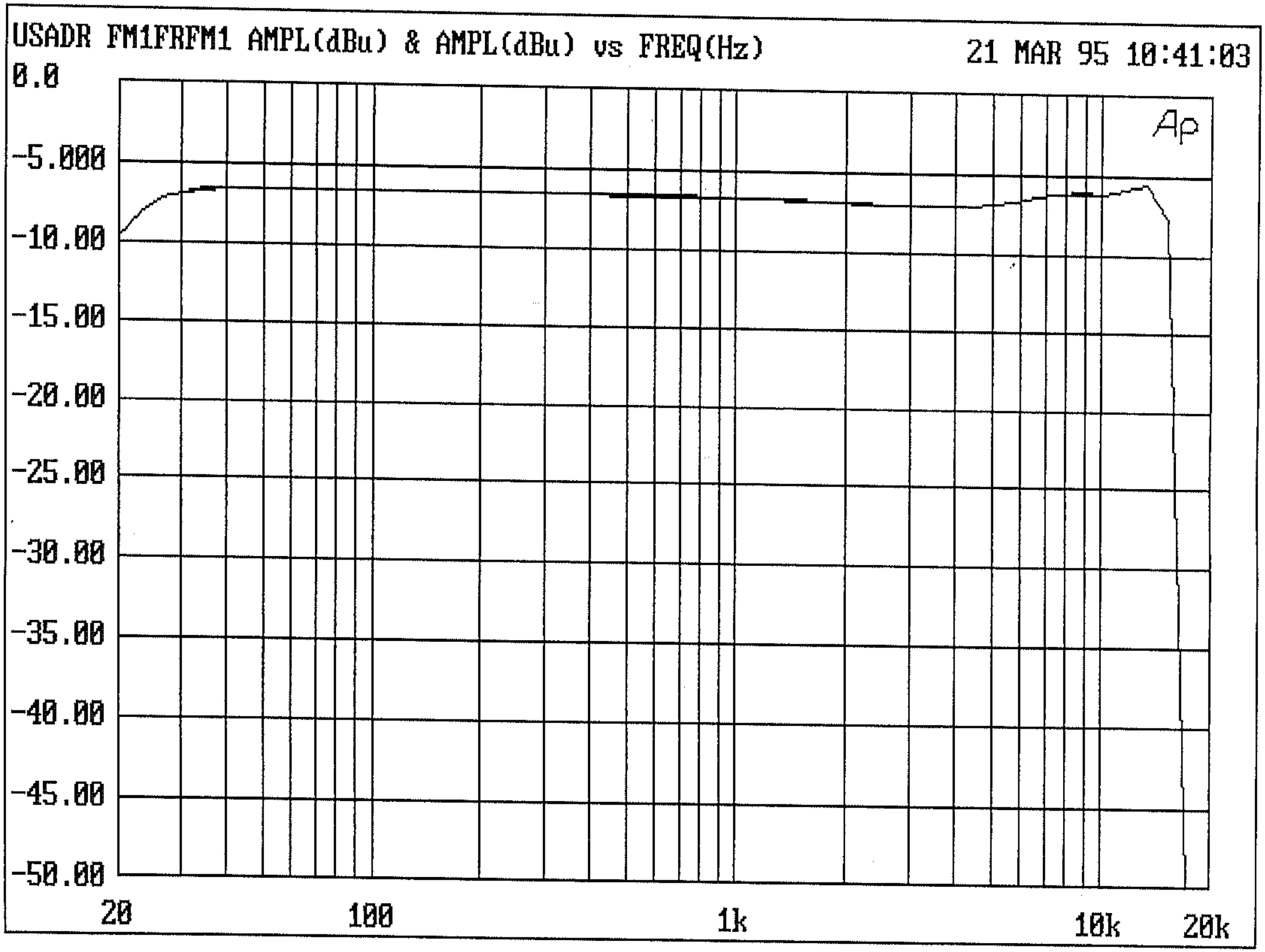
USADR FM2FRFM2 AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
21 MAR 95 10:47:38
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

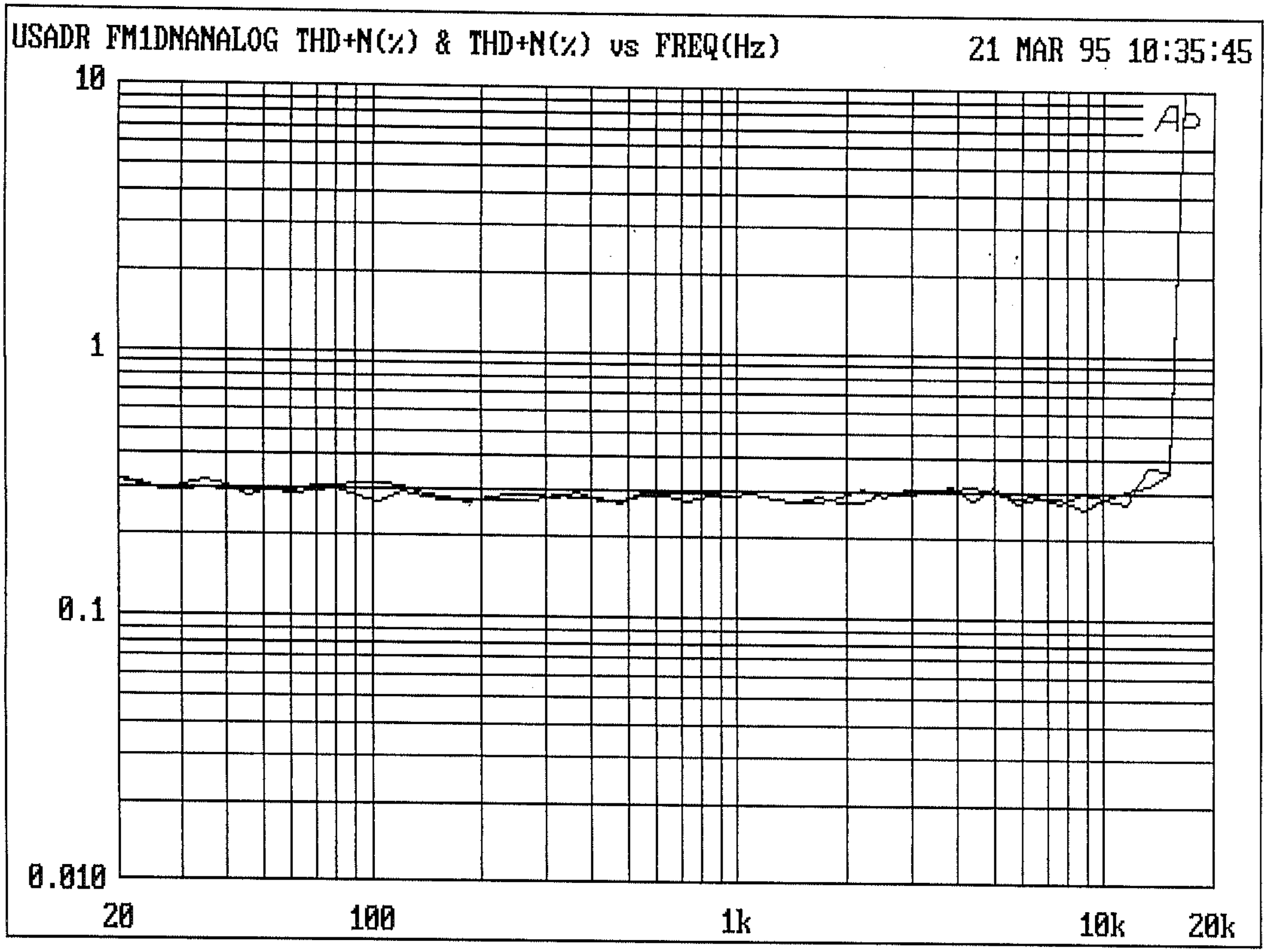

USADR FM2DNFM2 THD+N(%) & THD+N(%) vs FREQ(Hz)
21 MAR 95 10:52:37
10
1
0.1
0.010
20
100
1k
10k
20k
Ap

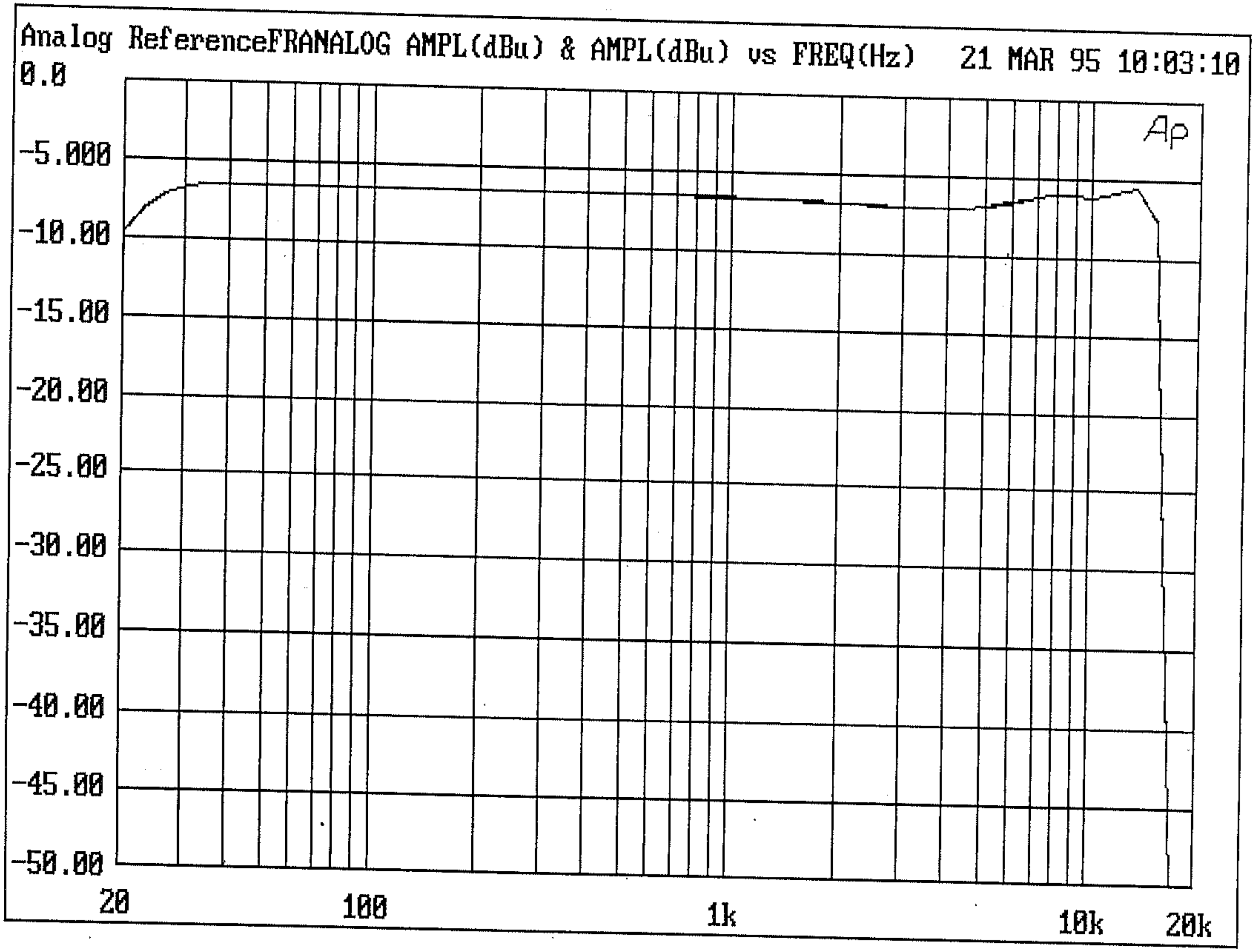


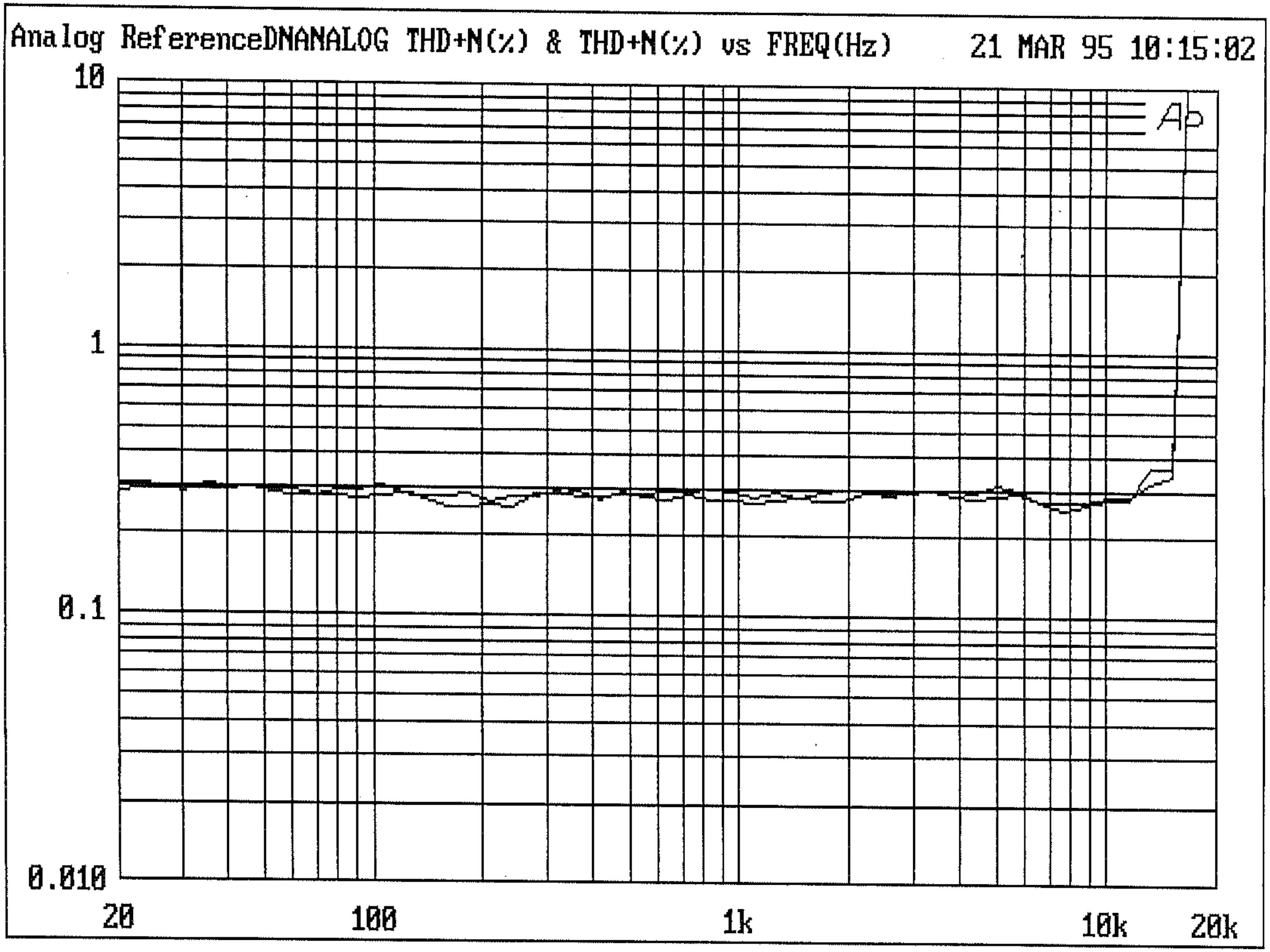
AT&T AMATI FRAMATI AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
21 MAR 95 11:02:54
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap


AT&T AMATIDNAMATI THD+N(%) & THD+N(%) vs FREQ(Hz)
21 MAR 95 10:57:57
10
1
0.1
0.010
20
100
1k
10k
20k
Ap


USADR FM1FRFM1 AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
21 MAR 95 10:41:03
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

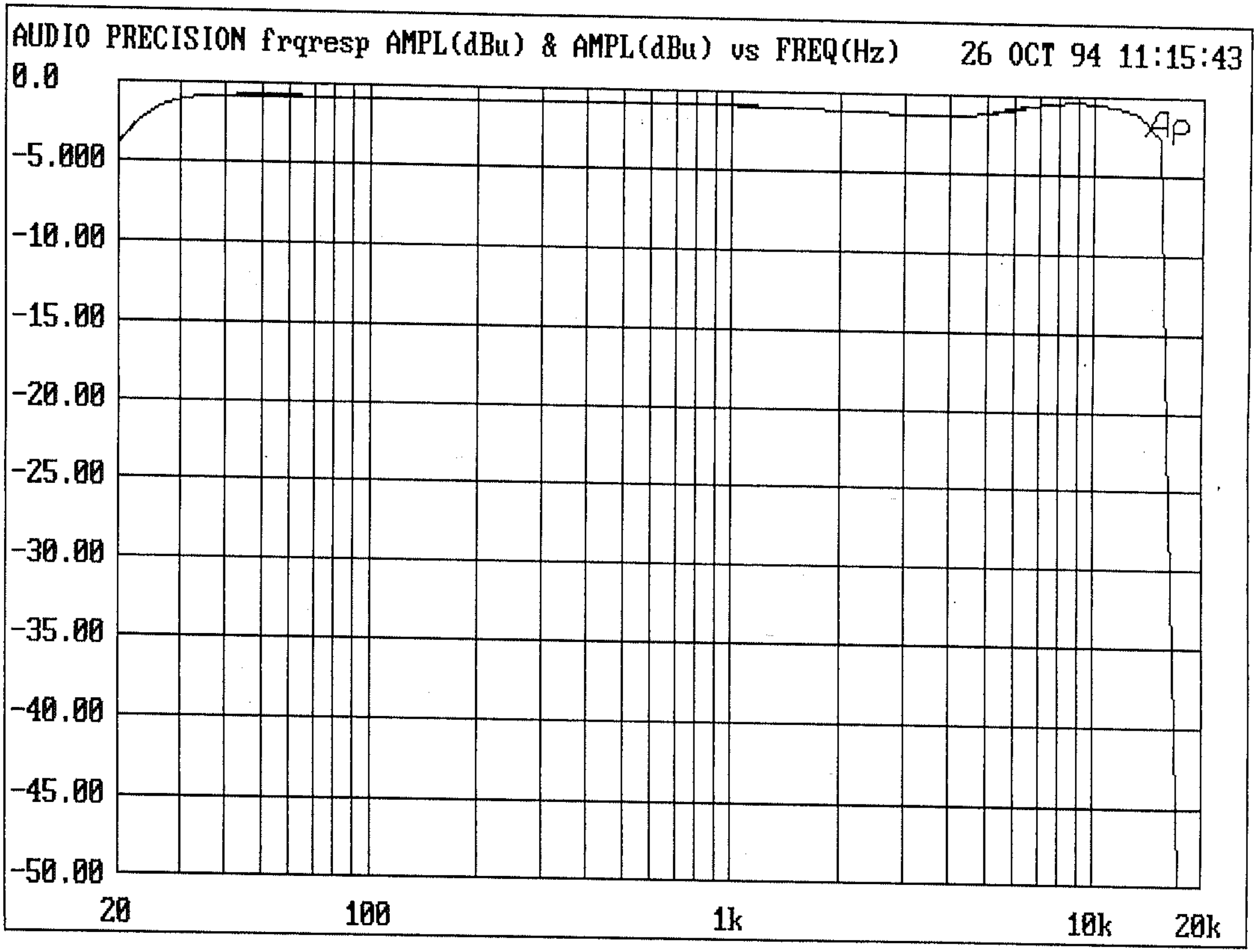

USADR FM1DNANALOG THD+N(%) & THD+N(%) vs FREQ(Hz)
21 MAR 95 10:35:45
10
1
0.1
0.010
20
100
1k
10k
20k
Ap

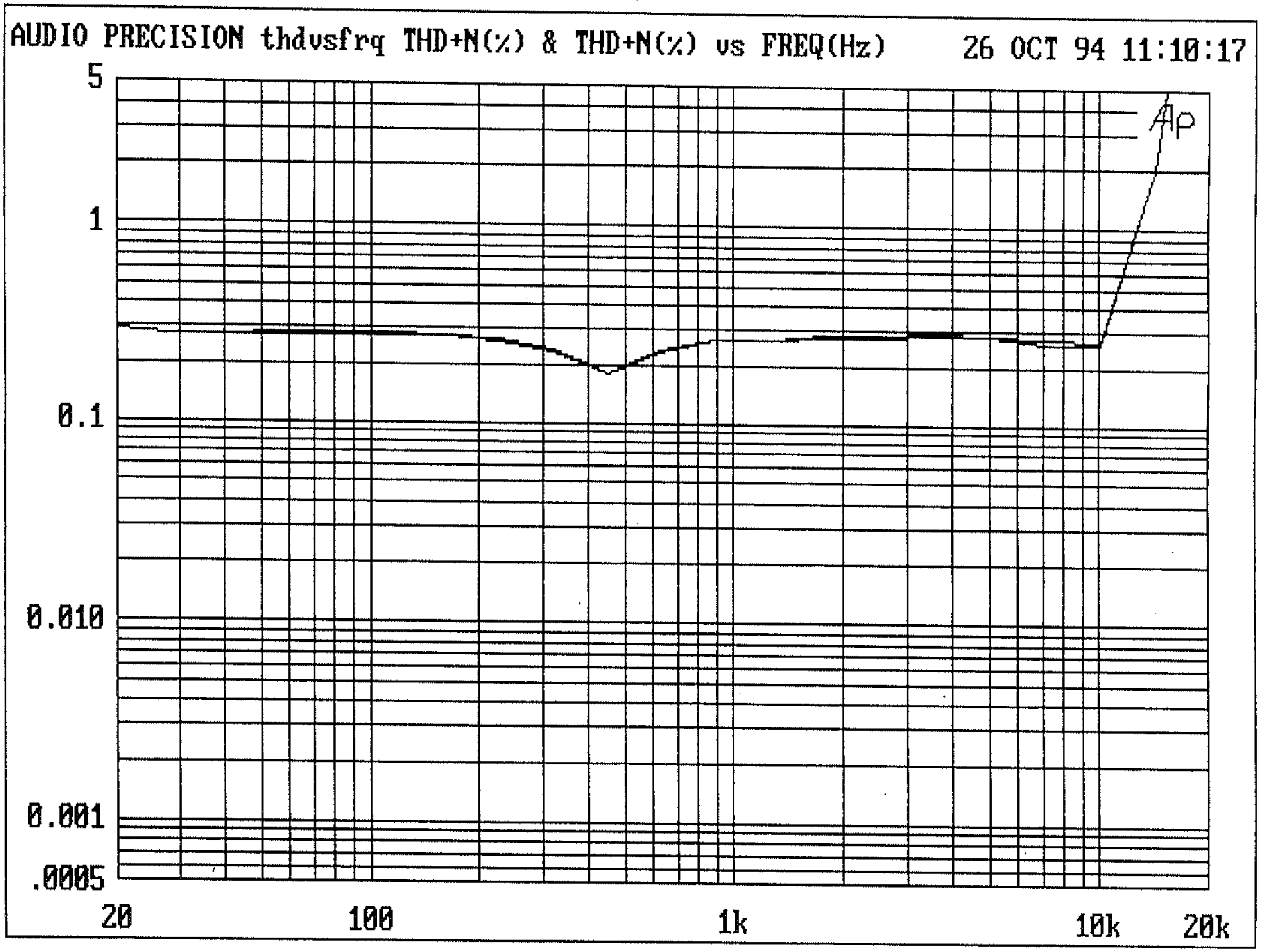

Analog ReferenceFRANALOG AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
21 MAR 95 10:03:10
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap



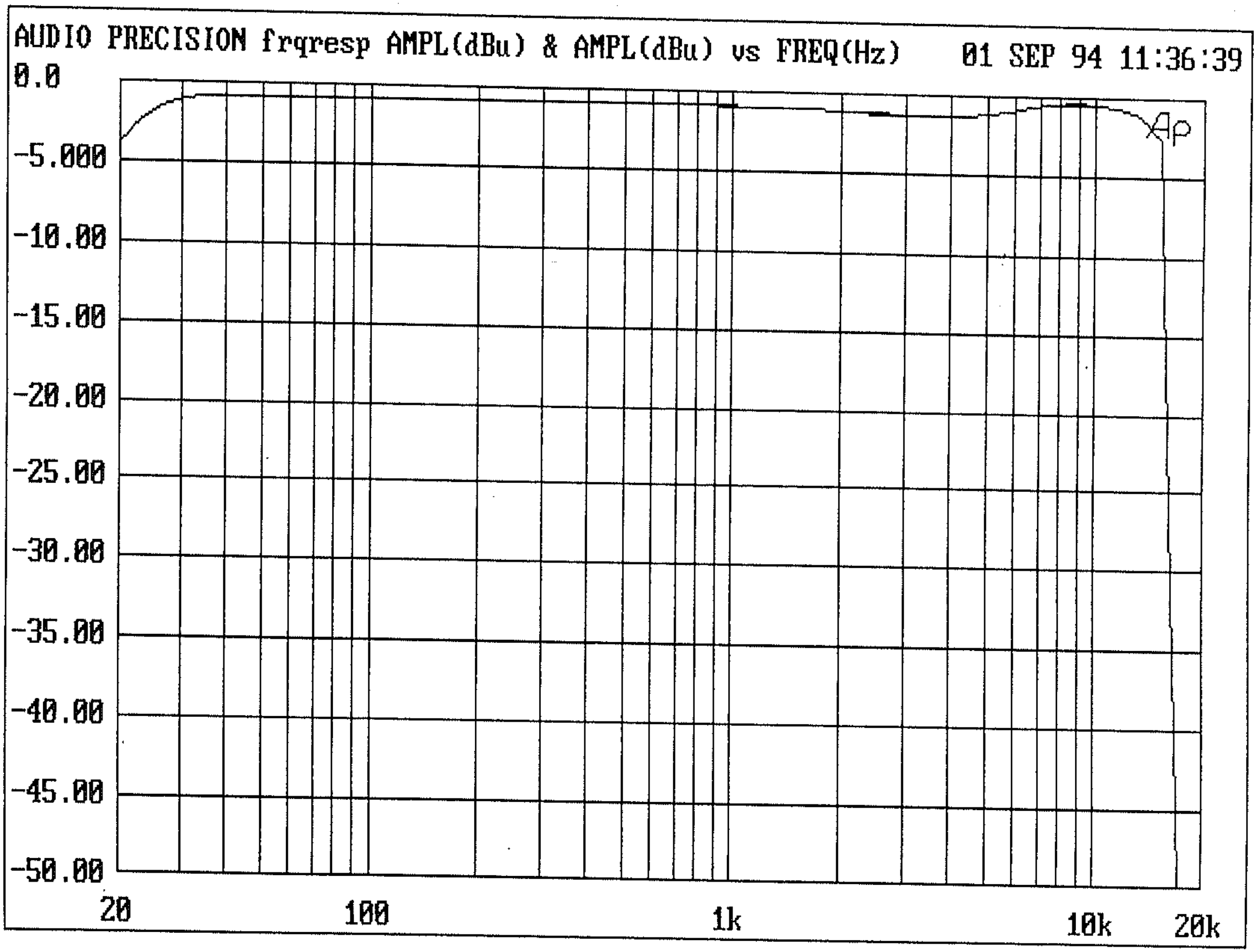
Analog ReferenceDNANALOG THD+N(%) & THD+N(%) vs FREQ(Hz)
21 MAR 95 10:15:02
10
1
0.1
0.010
20
100
1k
10k
20k
Ap

USADRFm1


AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
26 OCT 94 11:15:43
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

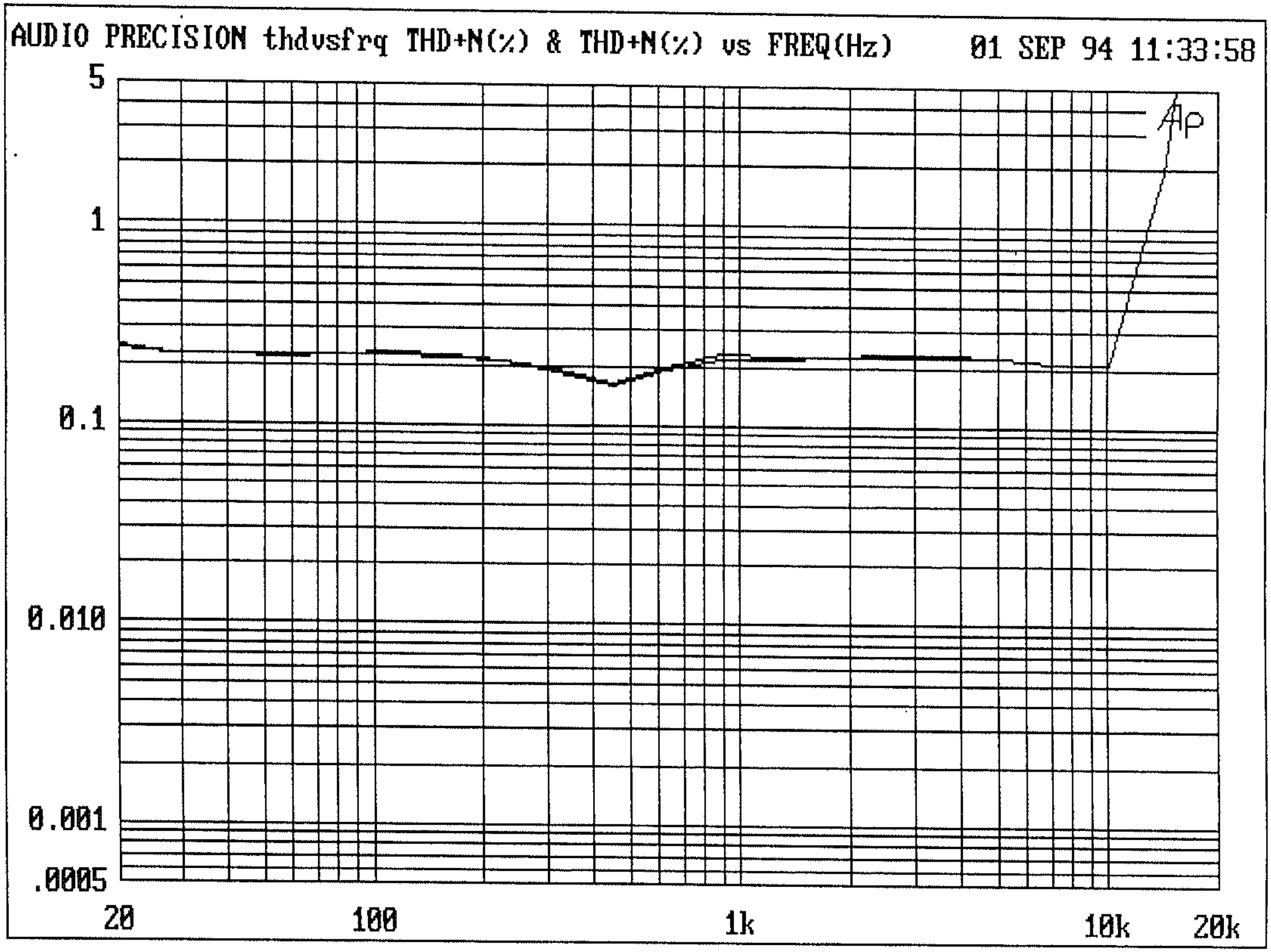


USADRPMI
WITH
DIGITAL

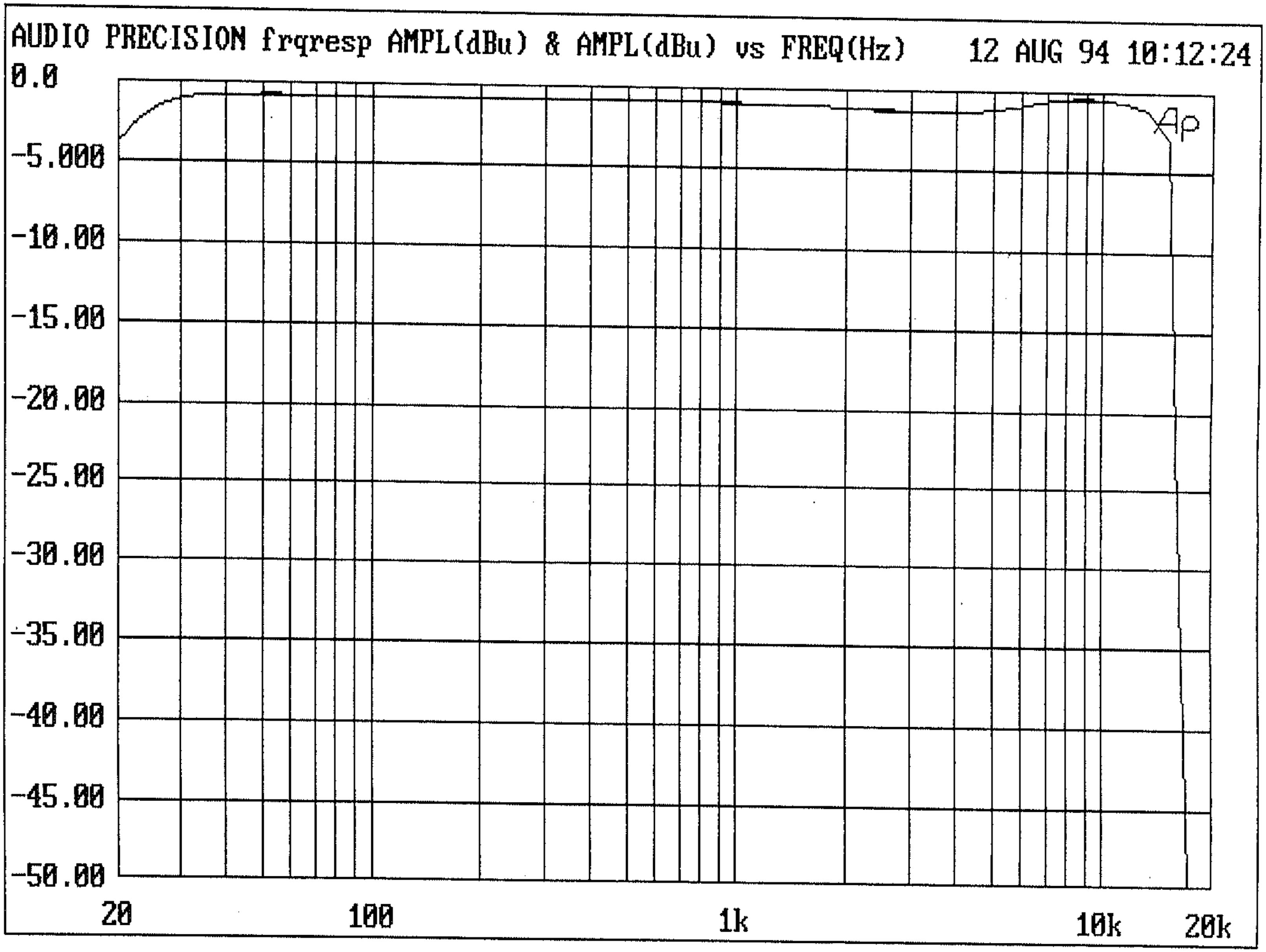

AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)   01 SEP 94 11:36:39
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

VSADRFM1

USADRFm1

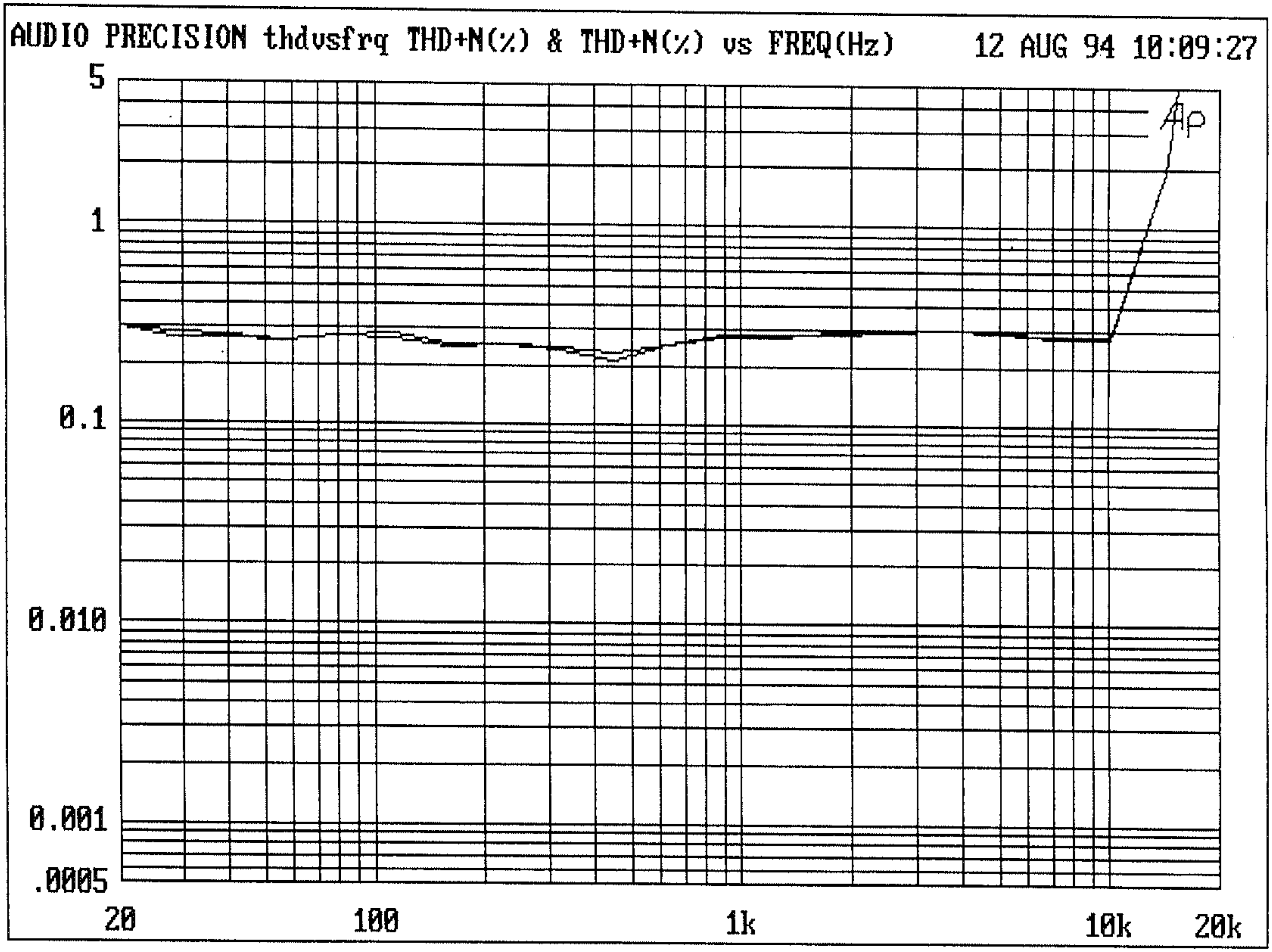

AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz) 01 SEP 94 11:33:58
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

VSADRFM1

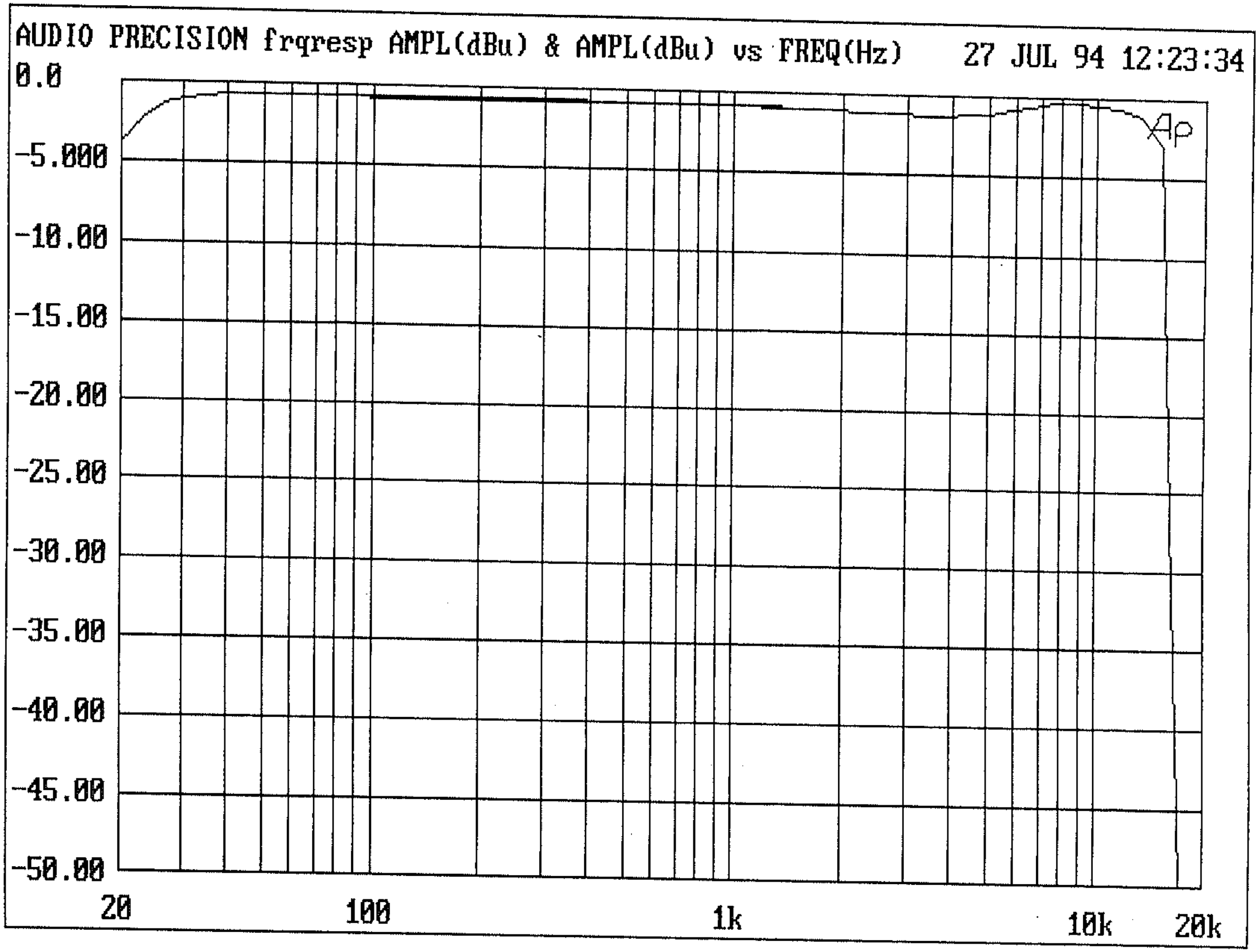


AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
12 AUG 94 10:12:24
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
AP

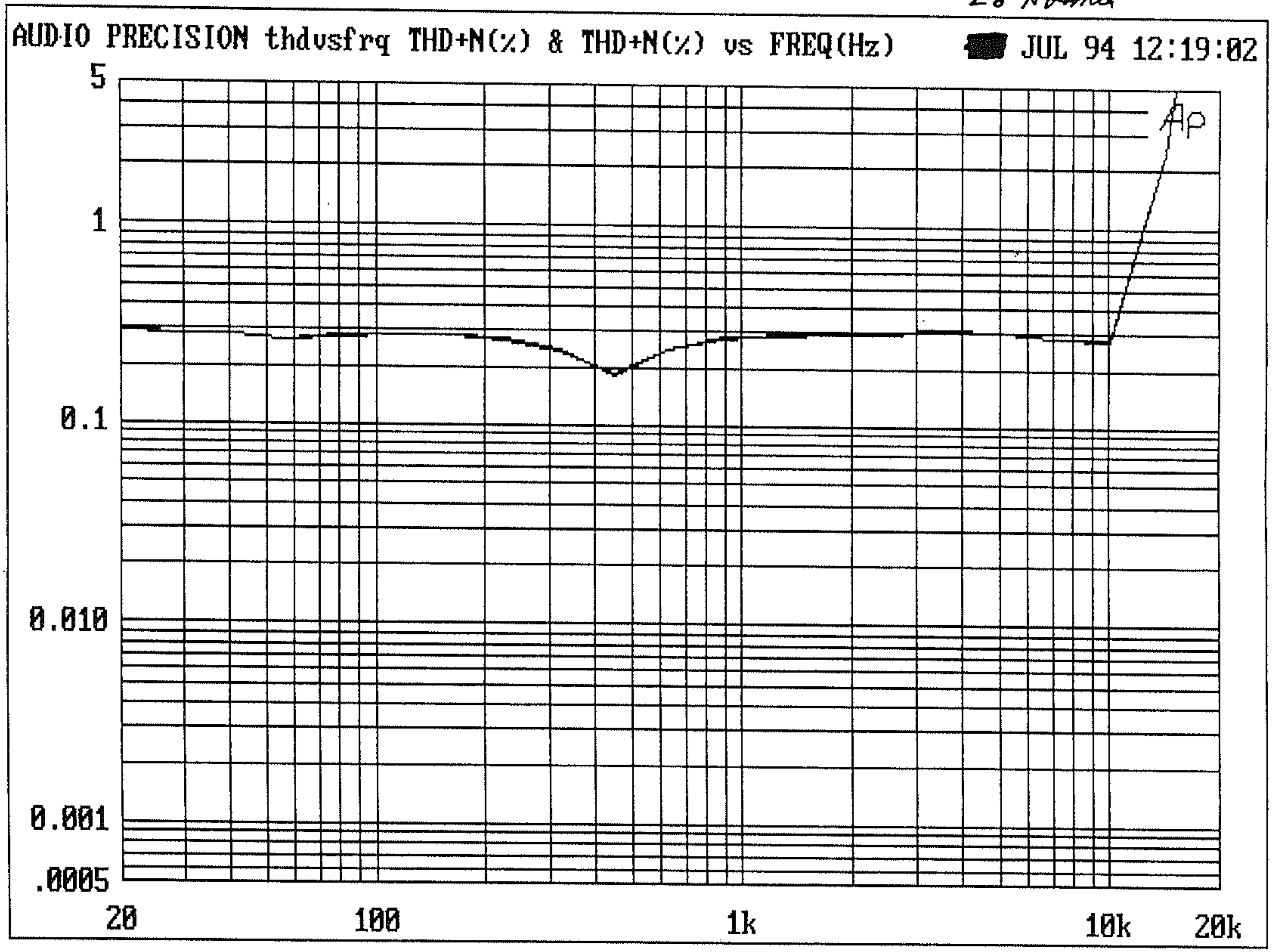
USADRFM1


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
12 AUG 94 10:09:27
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

VSA DRFml

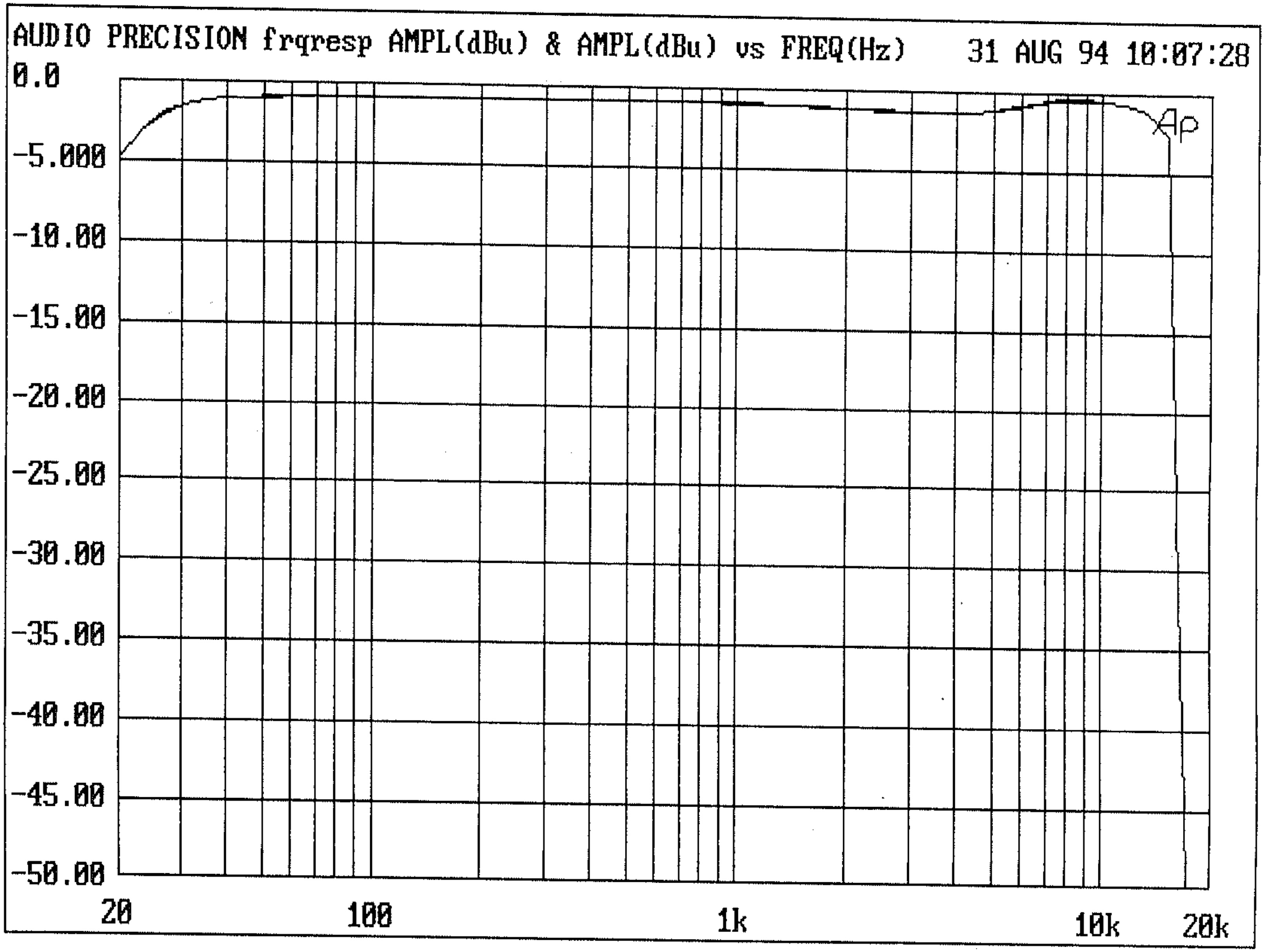

AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz) 27 JUL 94 12:23:34
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

VSA DRFM1

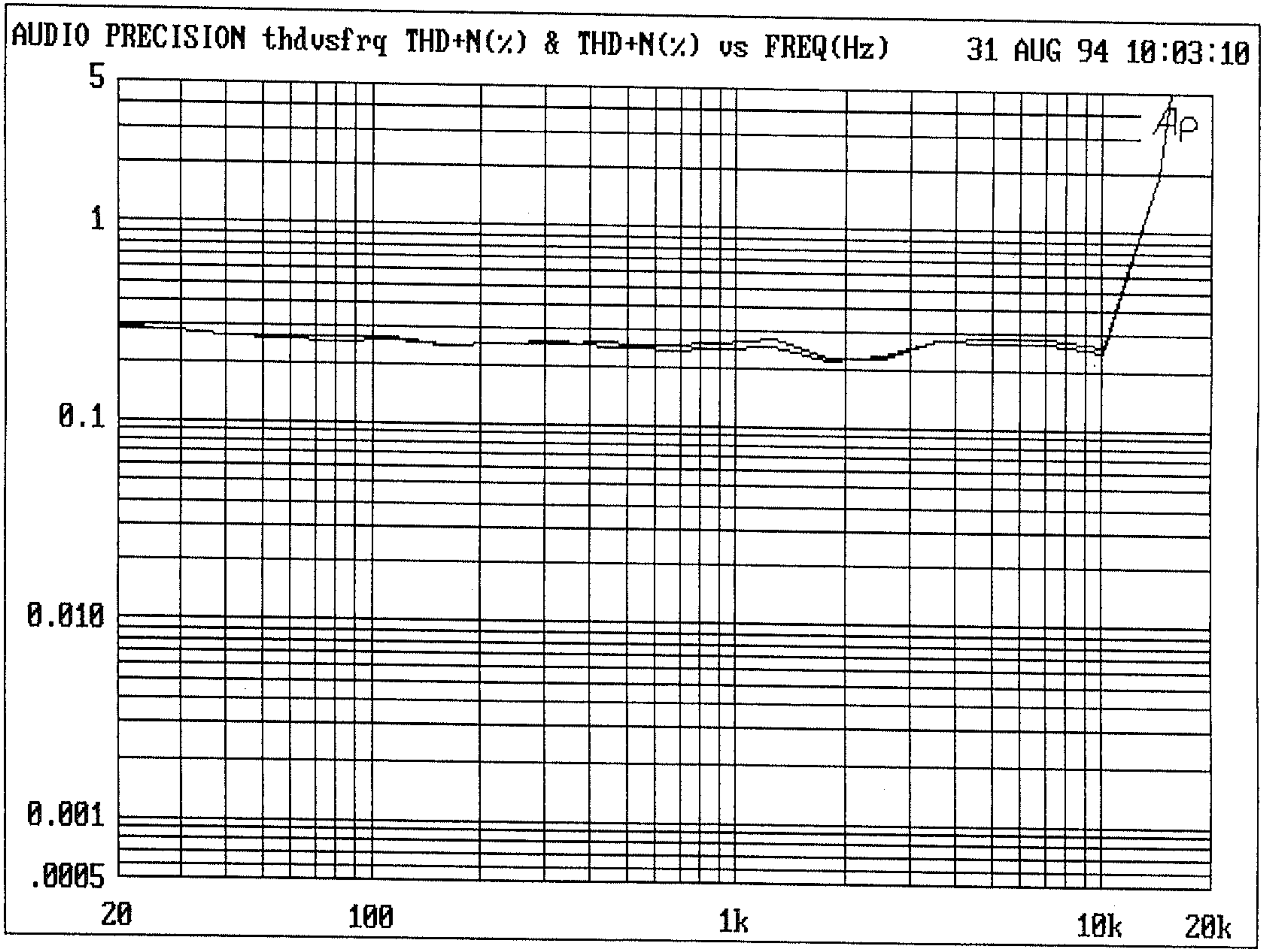


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
[redacted] JUL 94 12:19:02
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

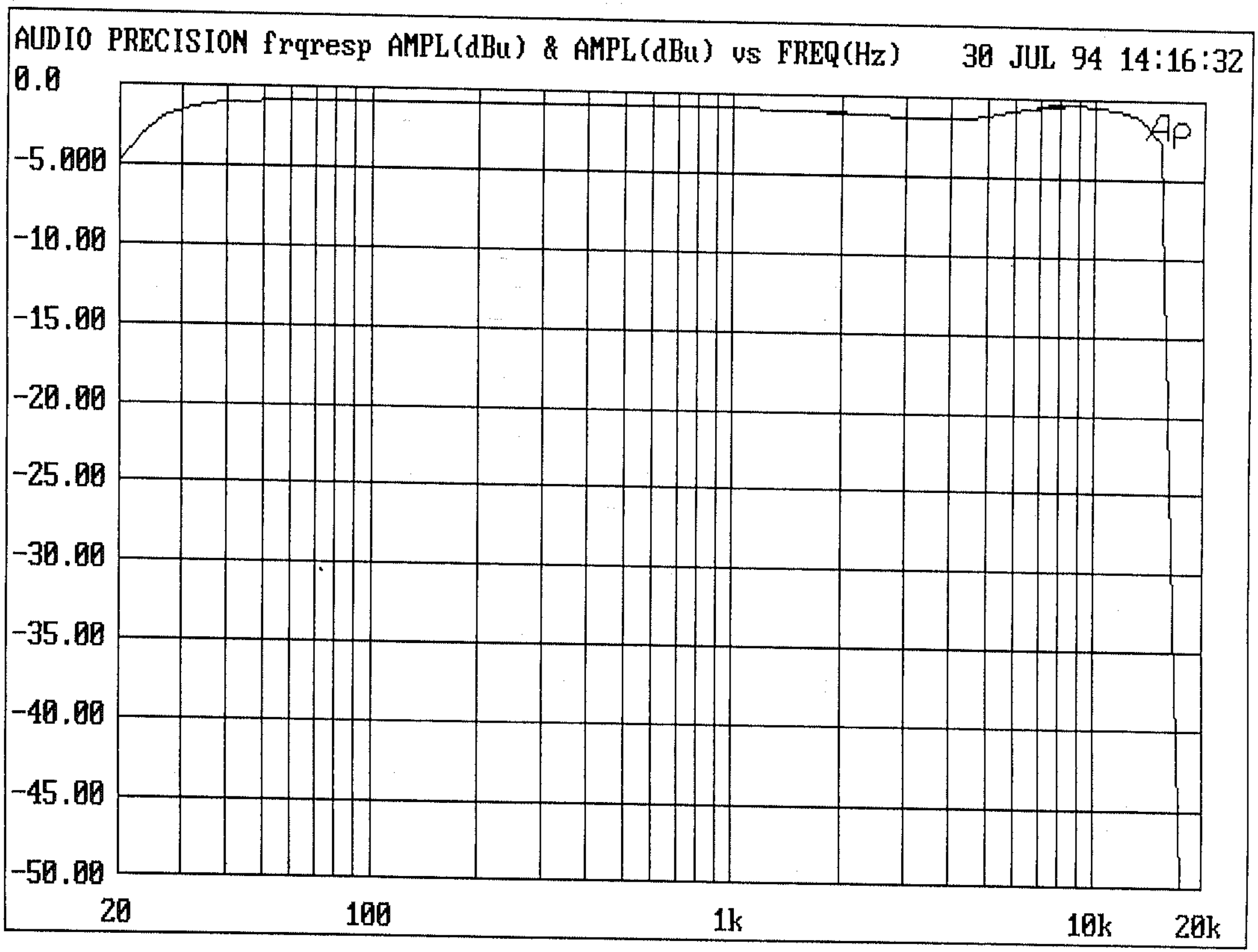
Amati LSB


AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
31 AUG 94 10:07:28
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

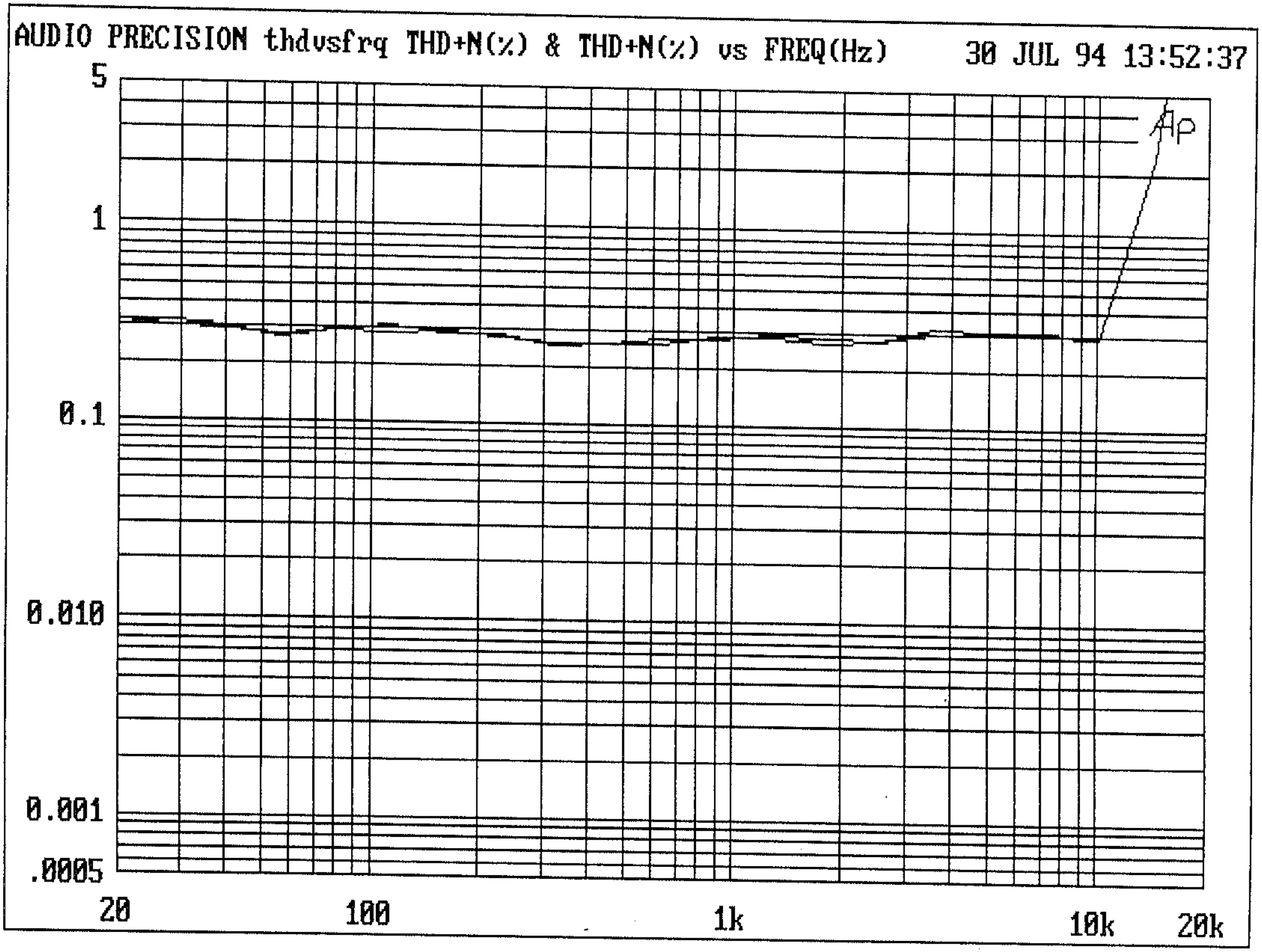
Amati LSB


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
31 AUG 94 10:03:10
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

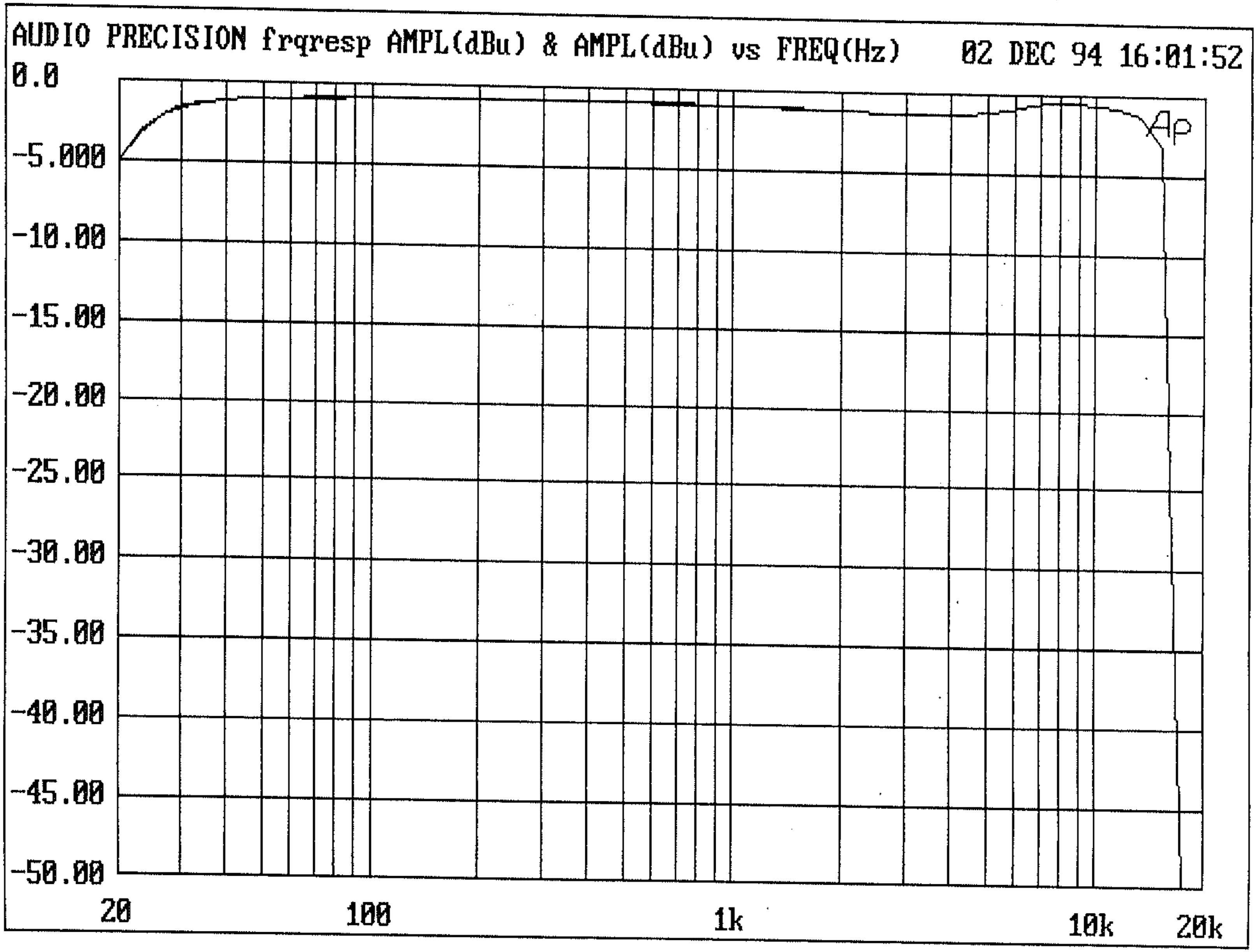
Amati LSB


AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz) 30 JUL 94 14:16:32
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

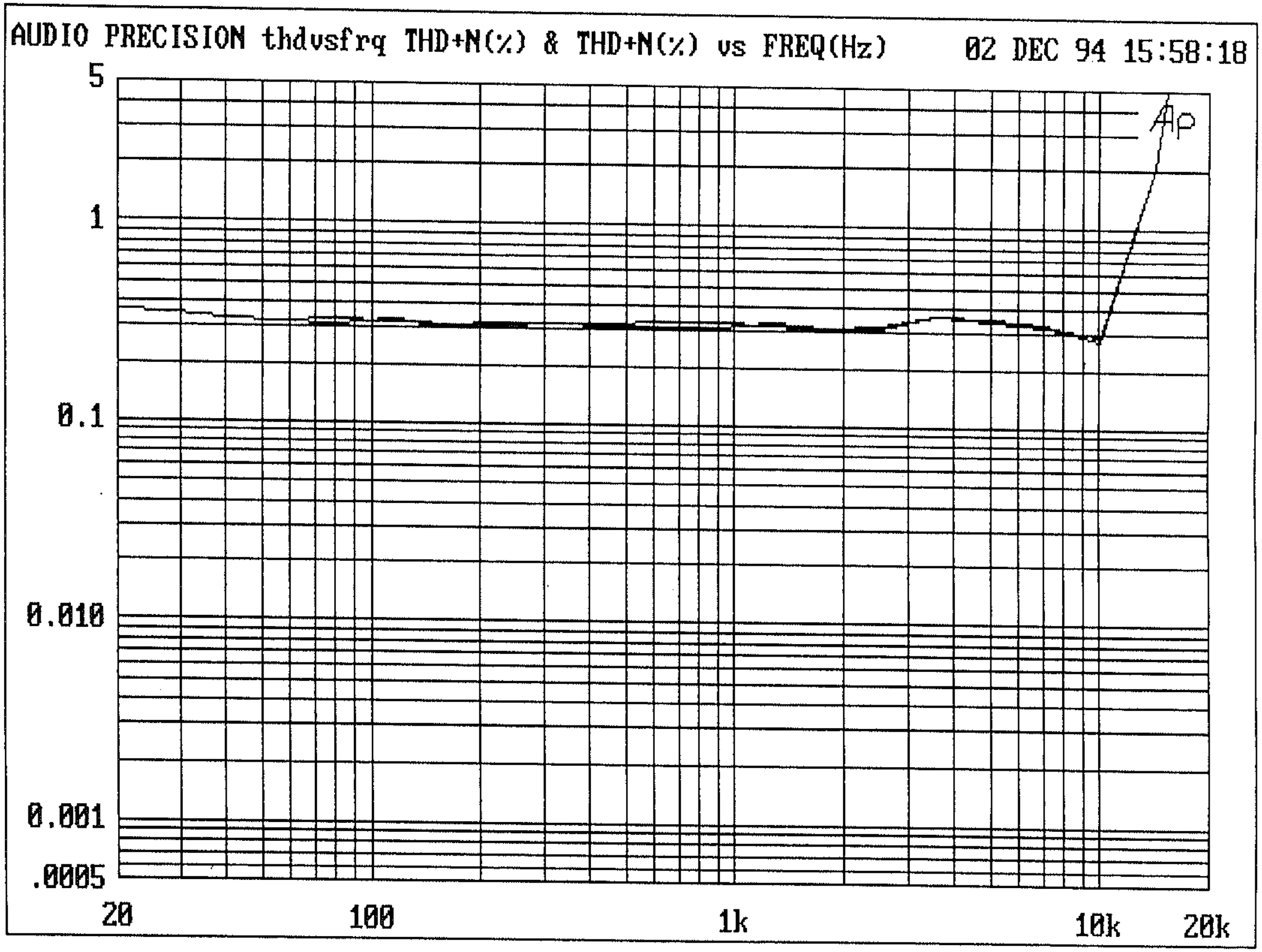
Amqti LSB


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
30 JUL 94 13:52:37
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

Amati BSB

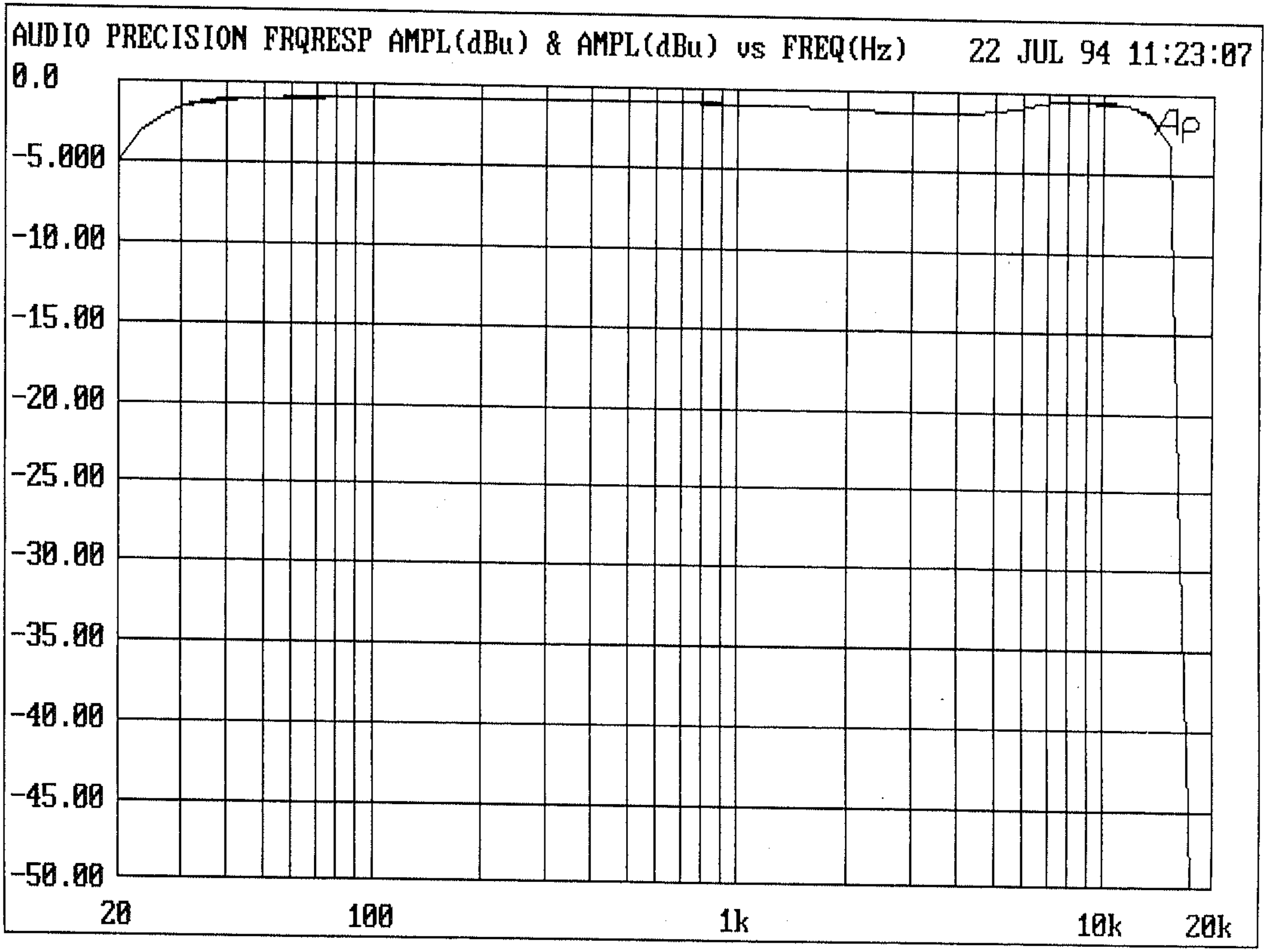

AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
02 DEC 94 16:01:52
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
AP

Amati DSB

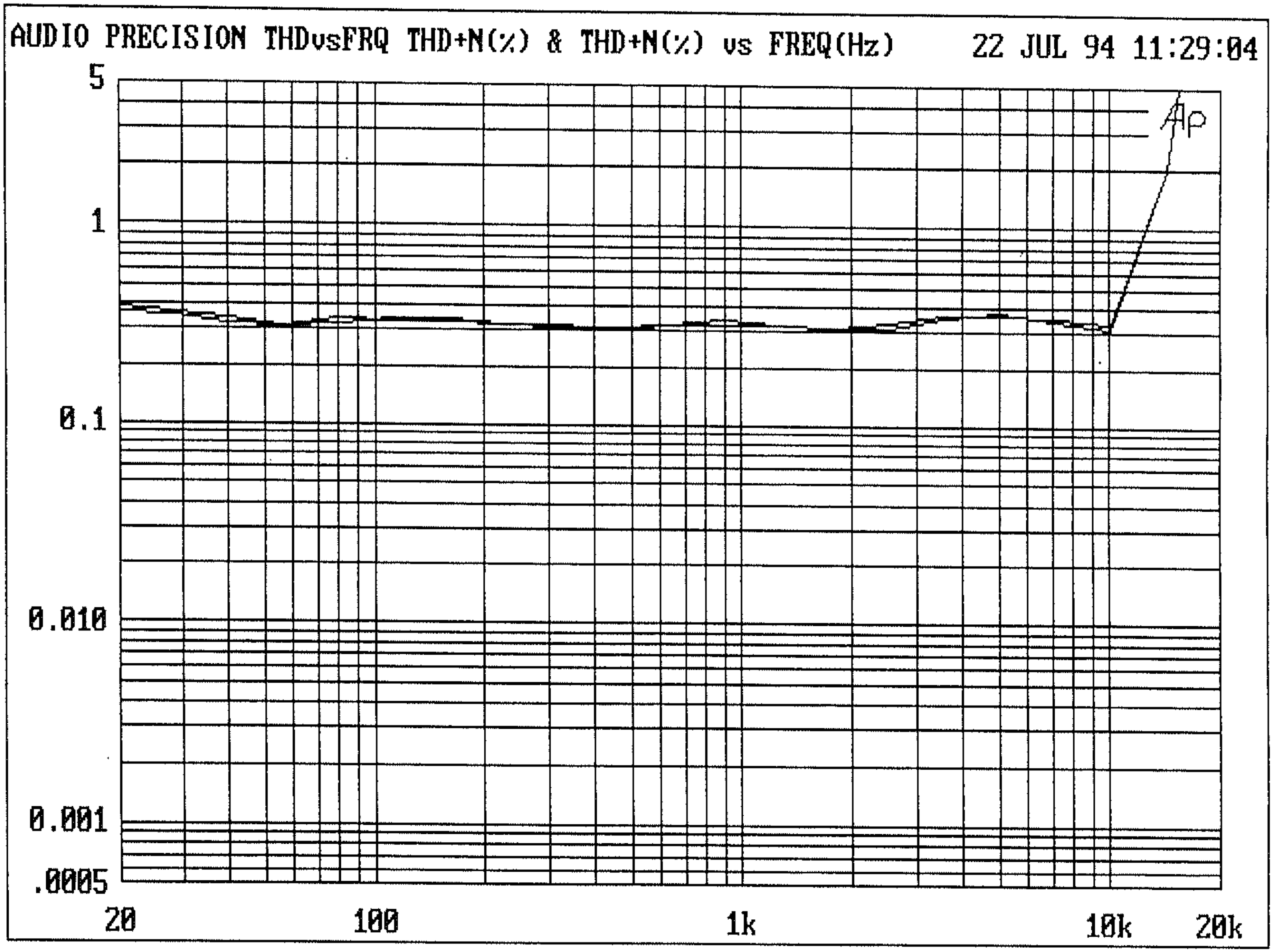


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
02 DEC 94 15:58:18
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

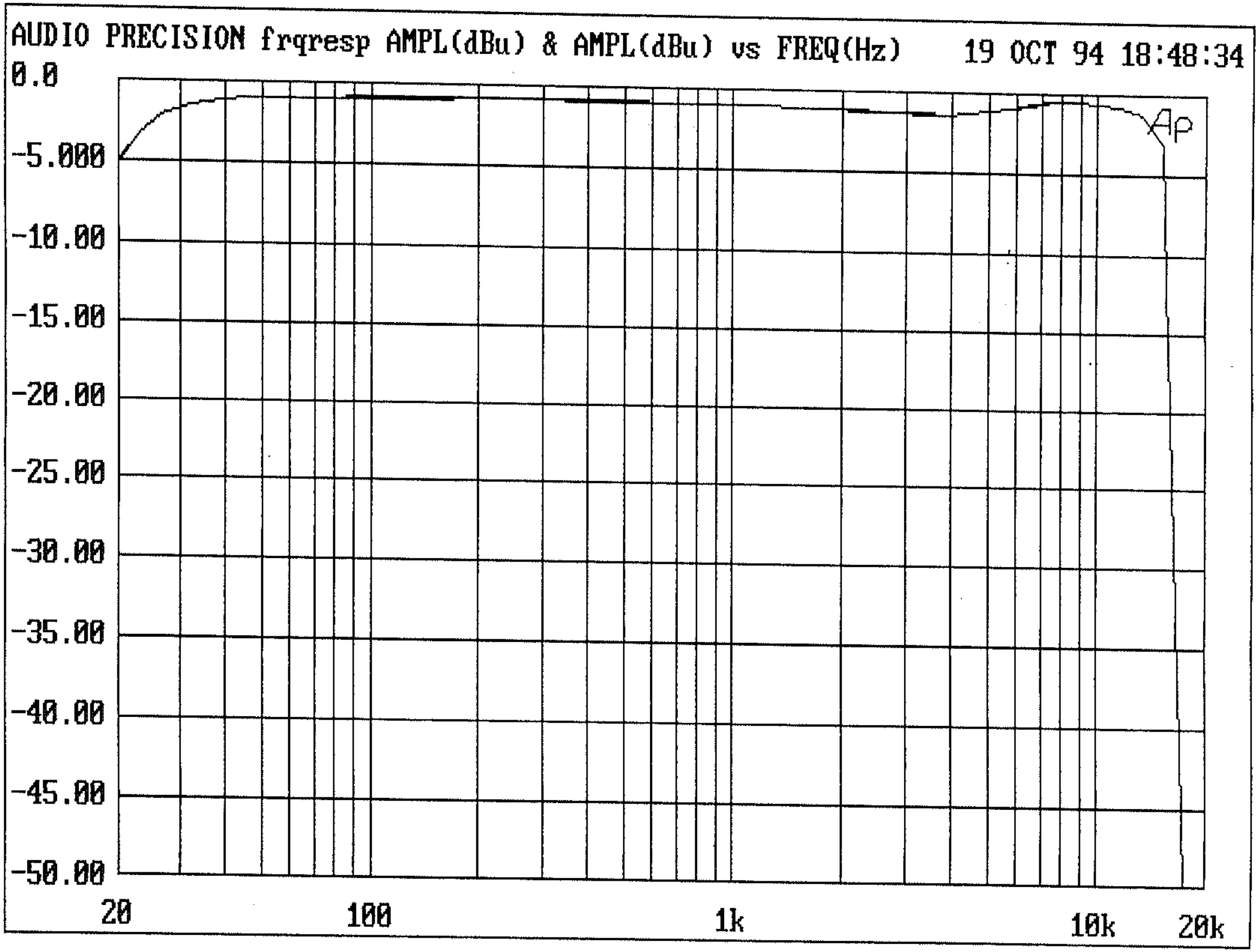
Amati DSB
AP TO DAT
DAT TO 8200
8200 TO PROPONENT



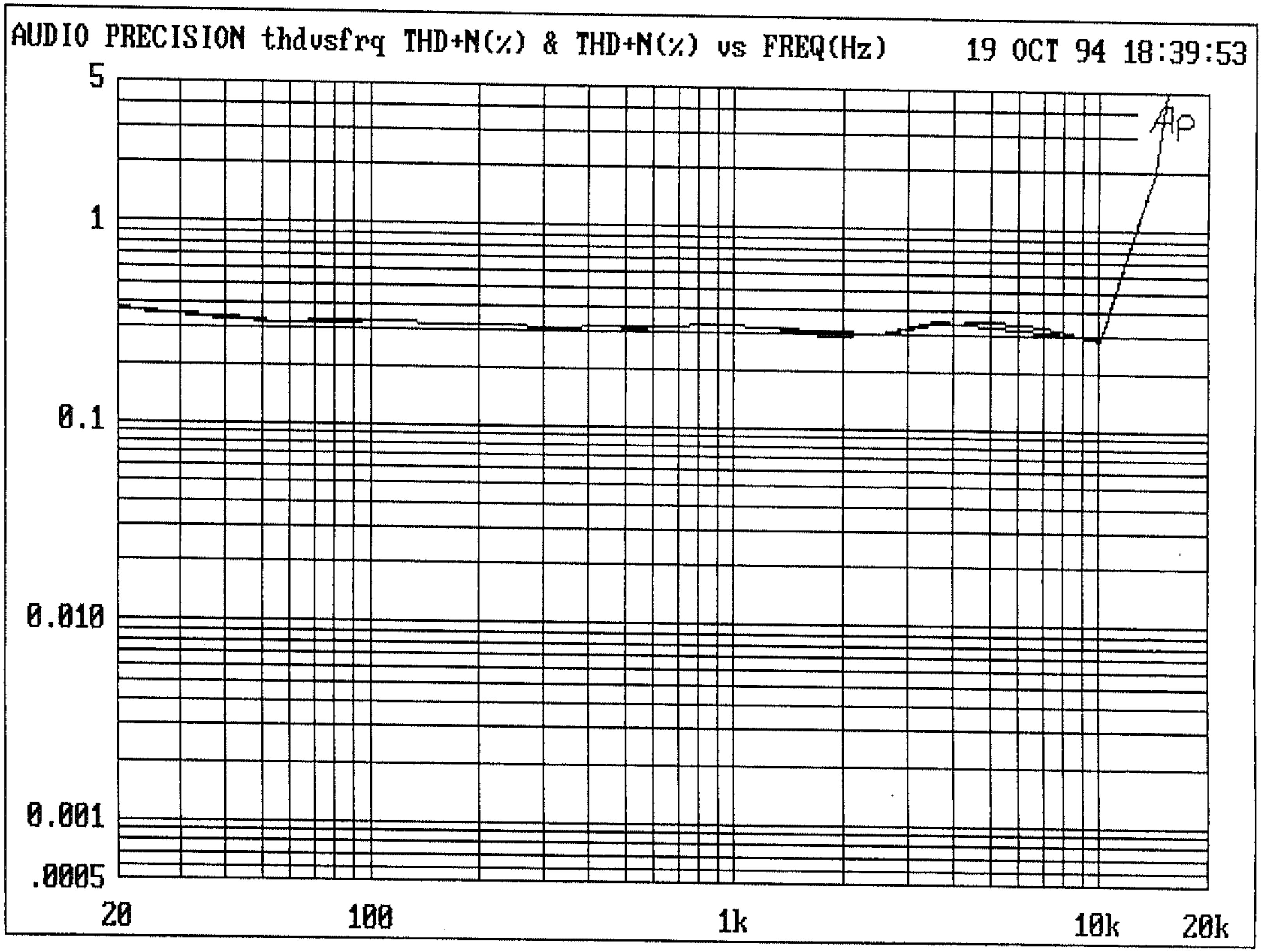
Amati DSB


AUDIO PRECISION THDvsFRQ THD+N(%) & THD+N(%) vs FREQ(Hz)
22 JUL 94 11:29:04
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

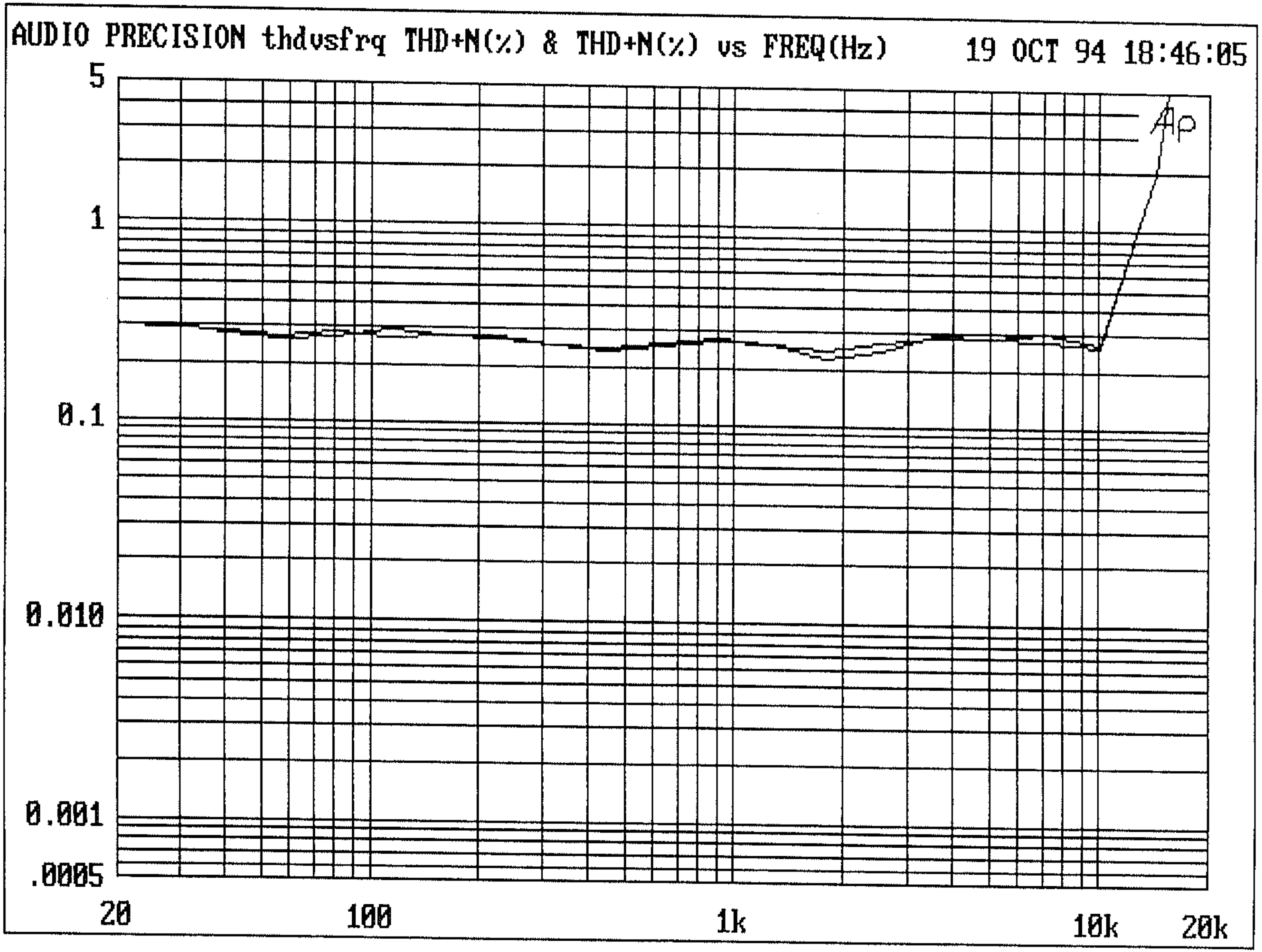
Amati DSB


AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
19 OCT 94 18:48:34
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

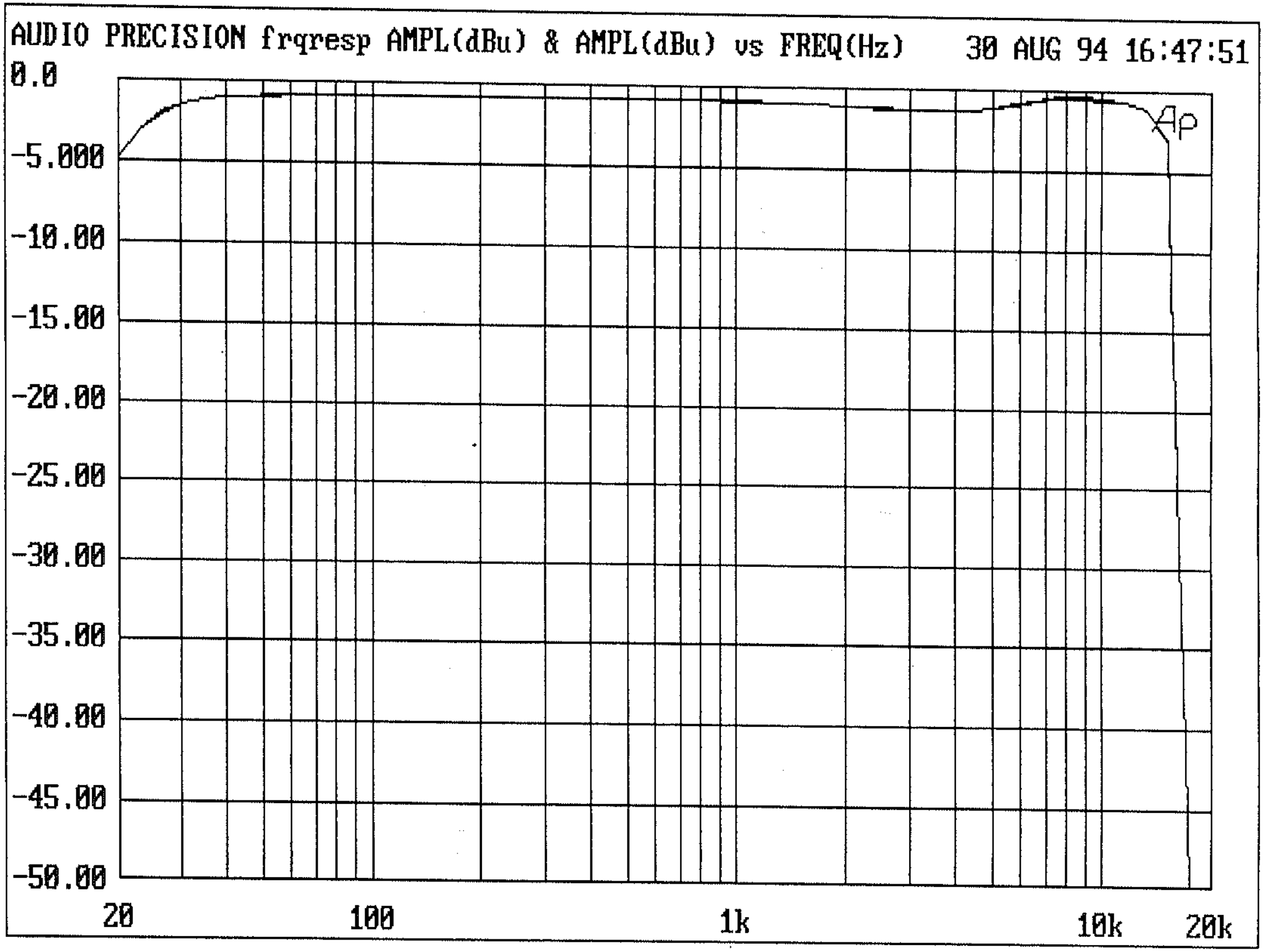
AmatibSB
WITH
DIGITAL



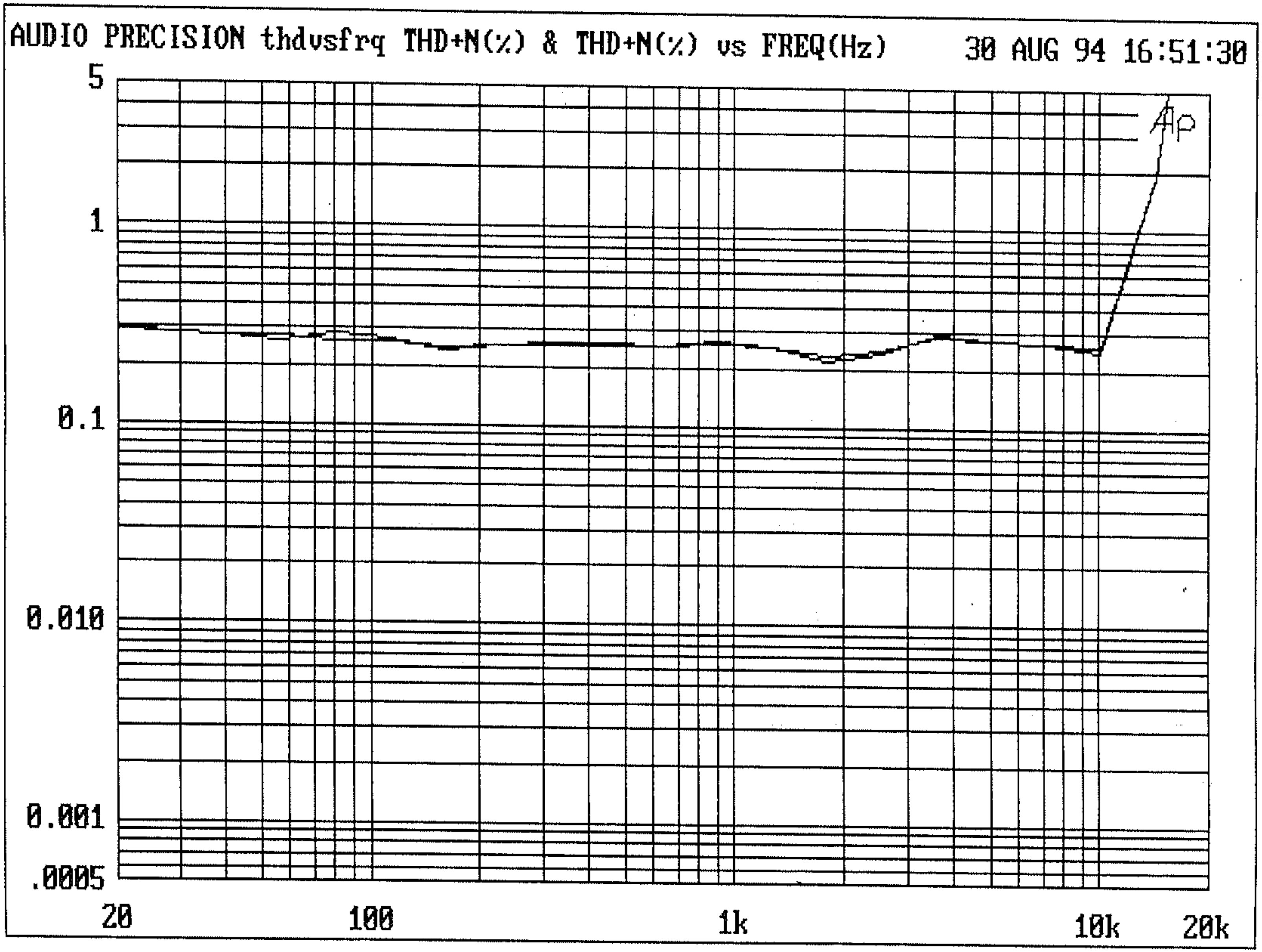
Amat: DSB
WITHOUT
DIGITAL



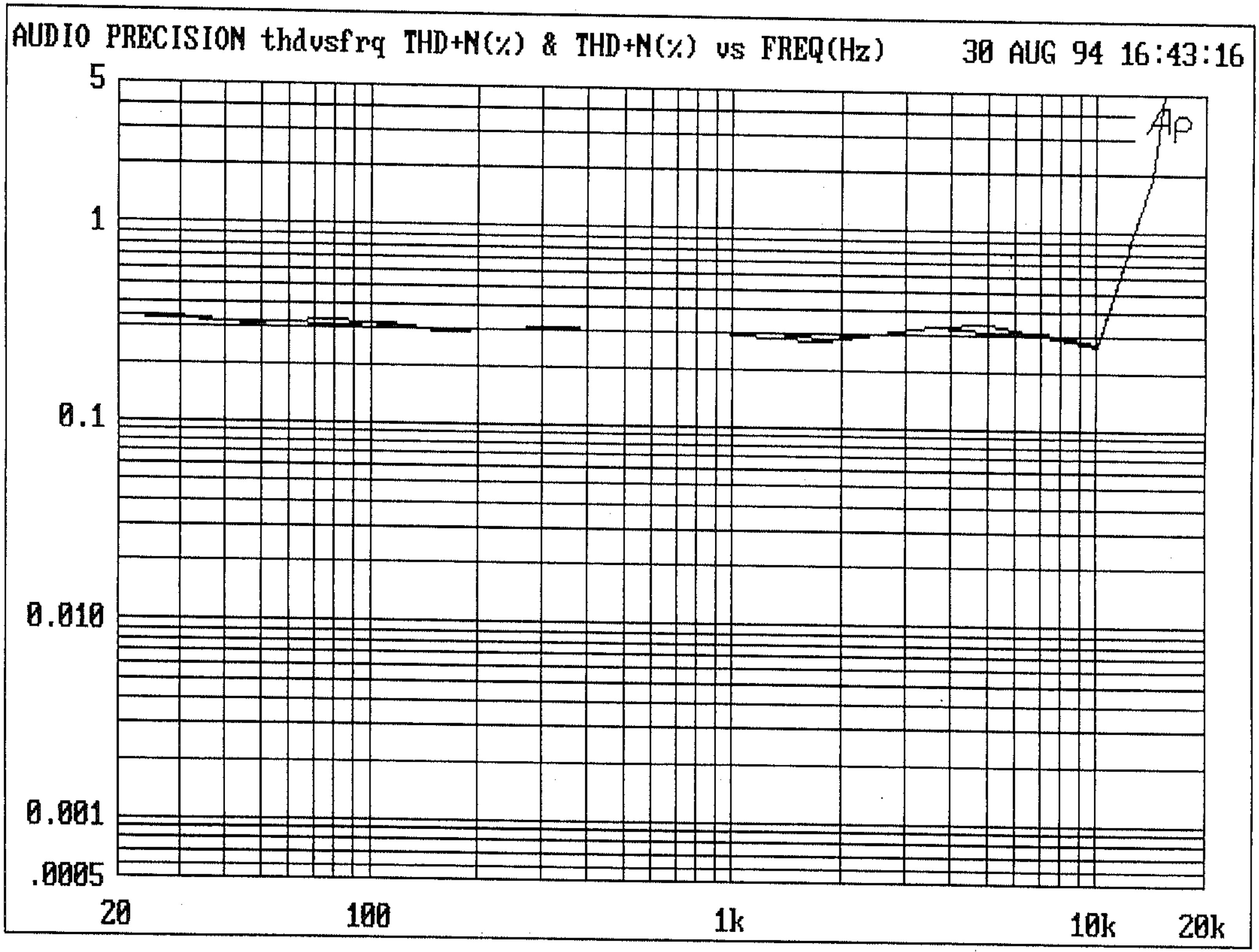
Amati DSB
DIGITAL ON



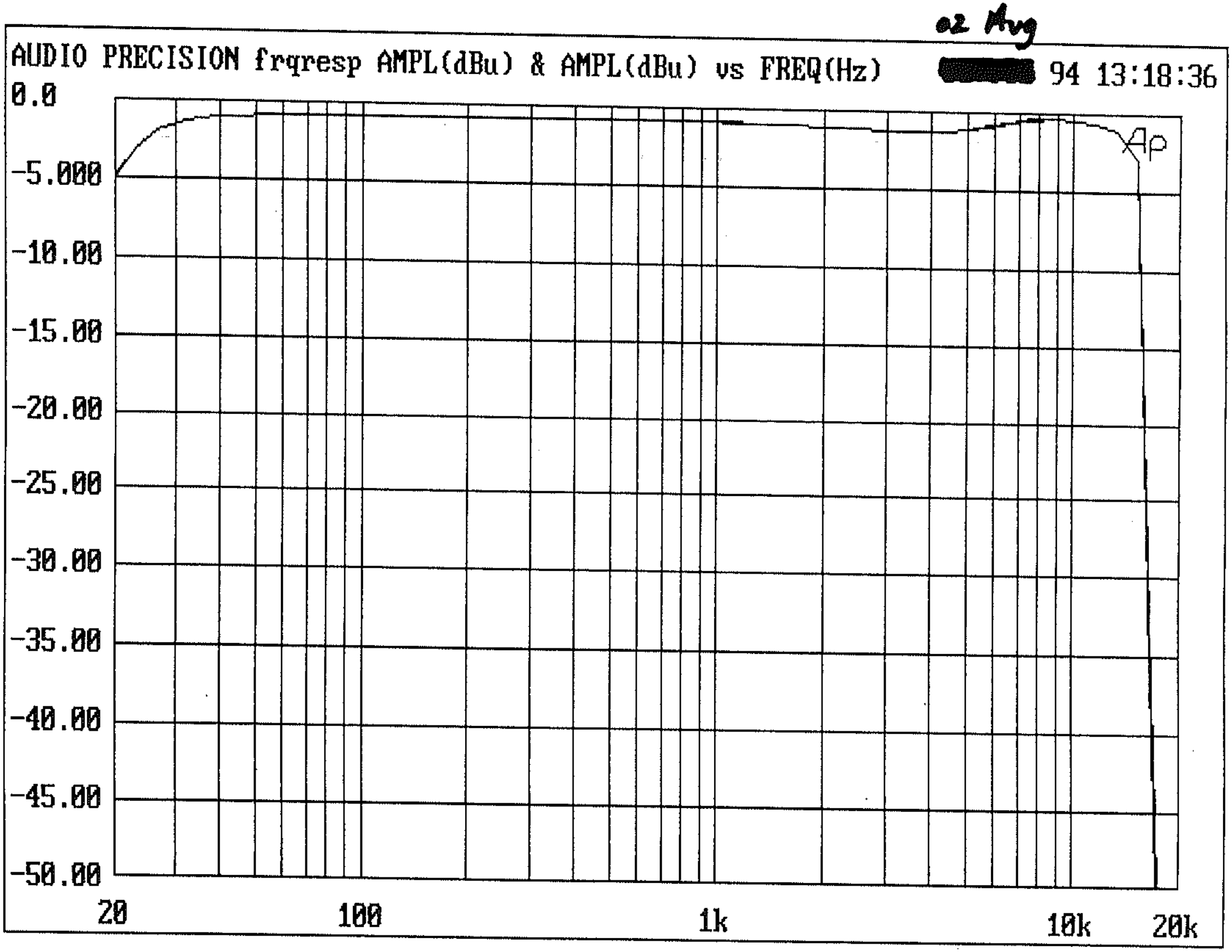
Amati DSB
DIGITAL OFF



Amati DSB
DIGITALON

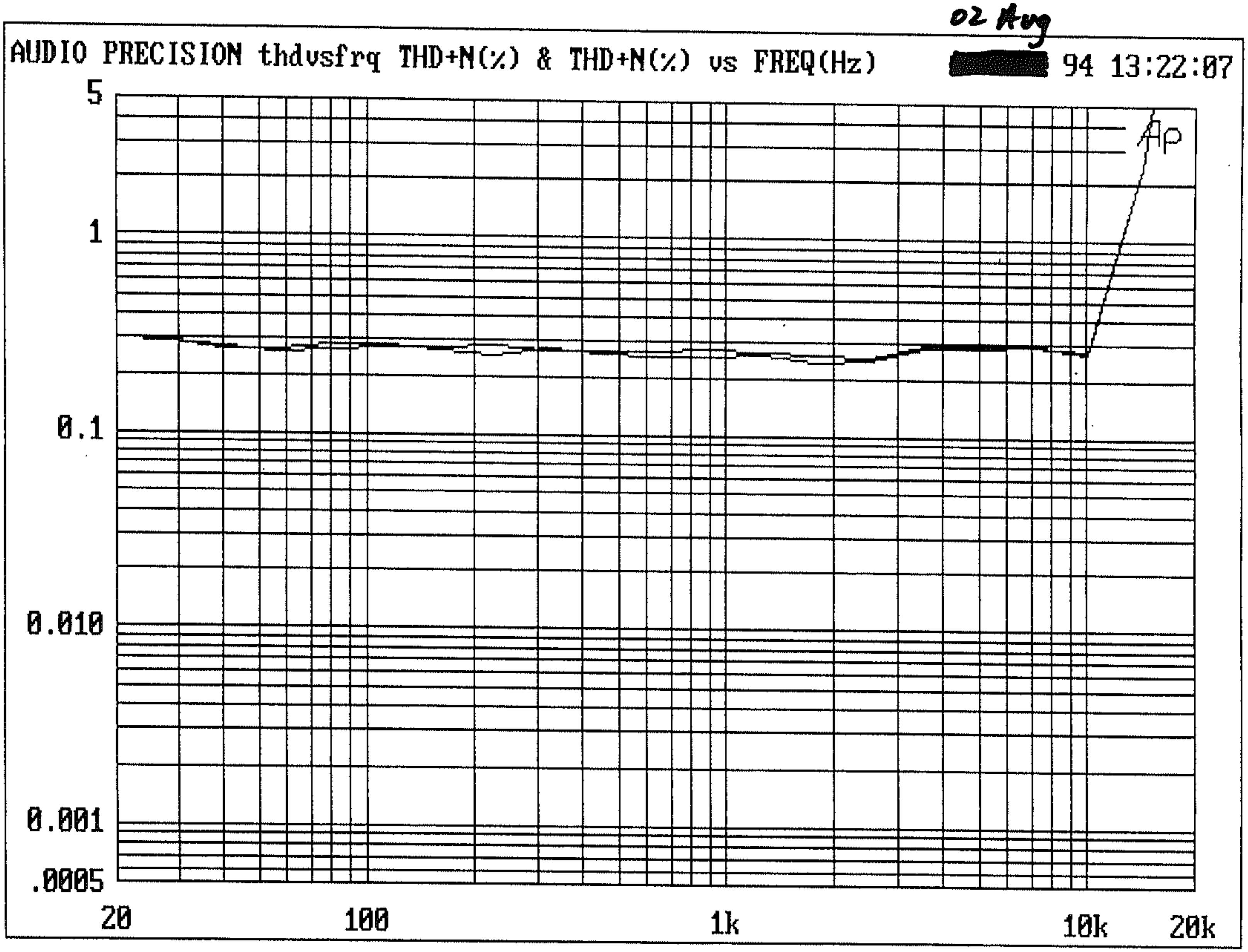


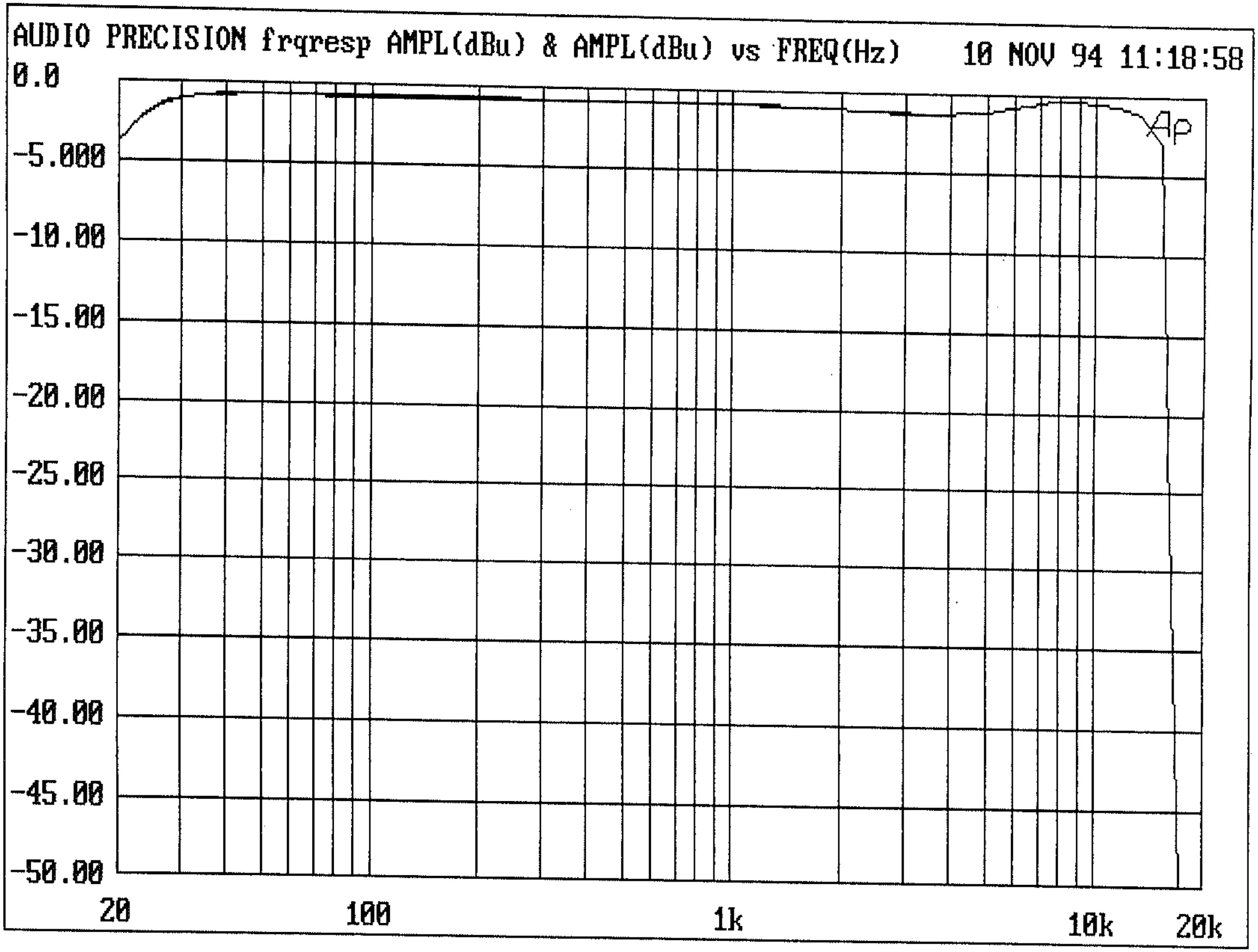
Amati DSB

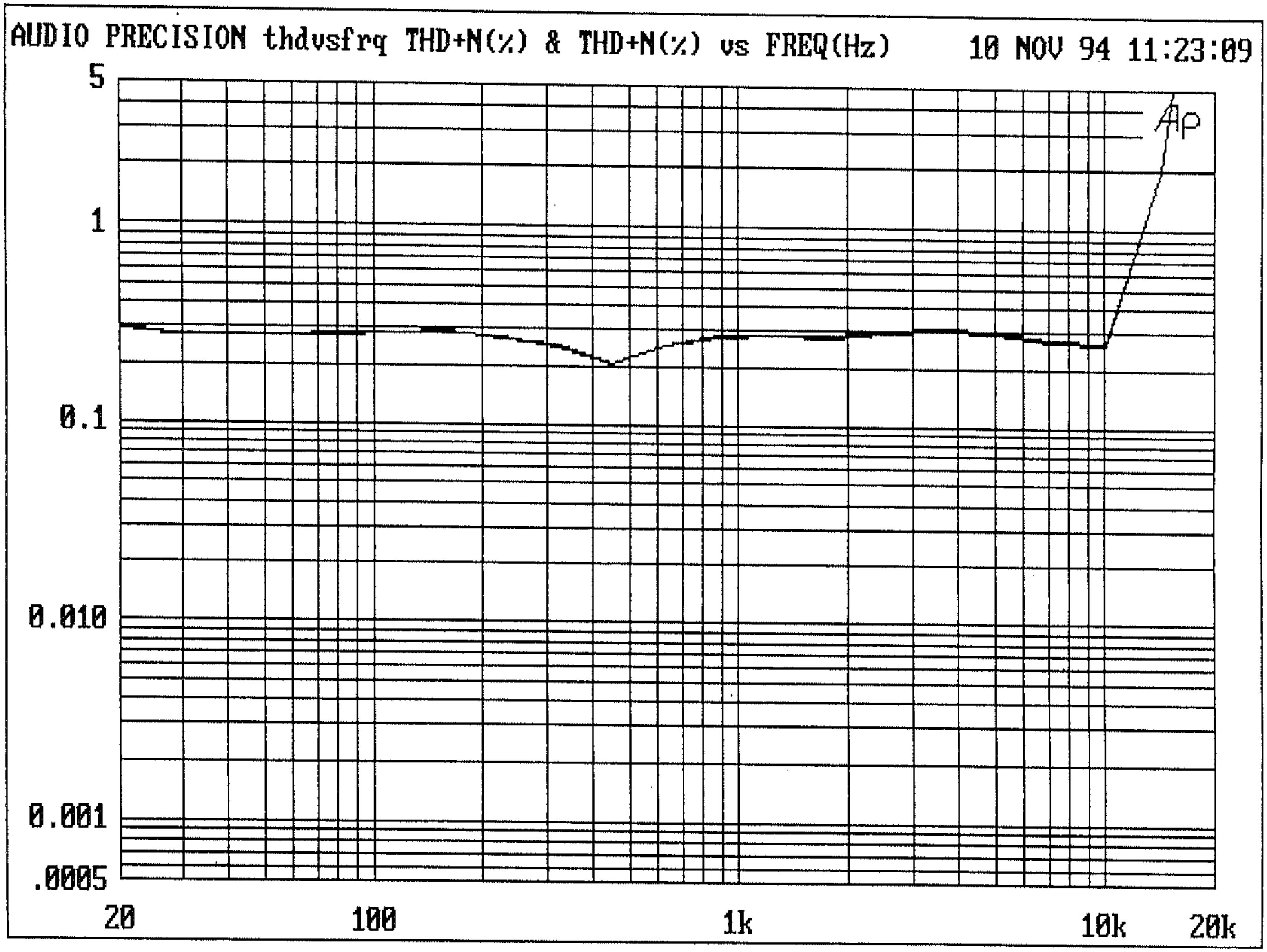


02 Aug
AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz) [redacted] 94 13:18:36
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

Amati DSB

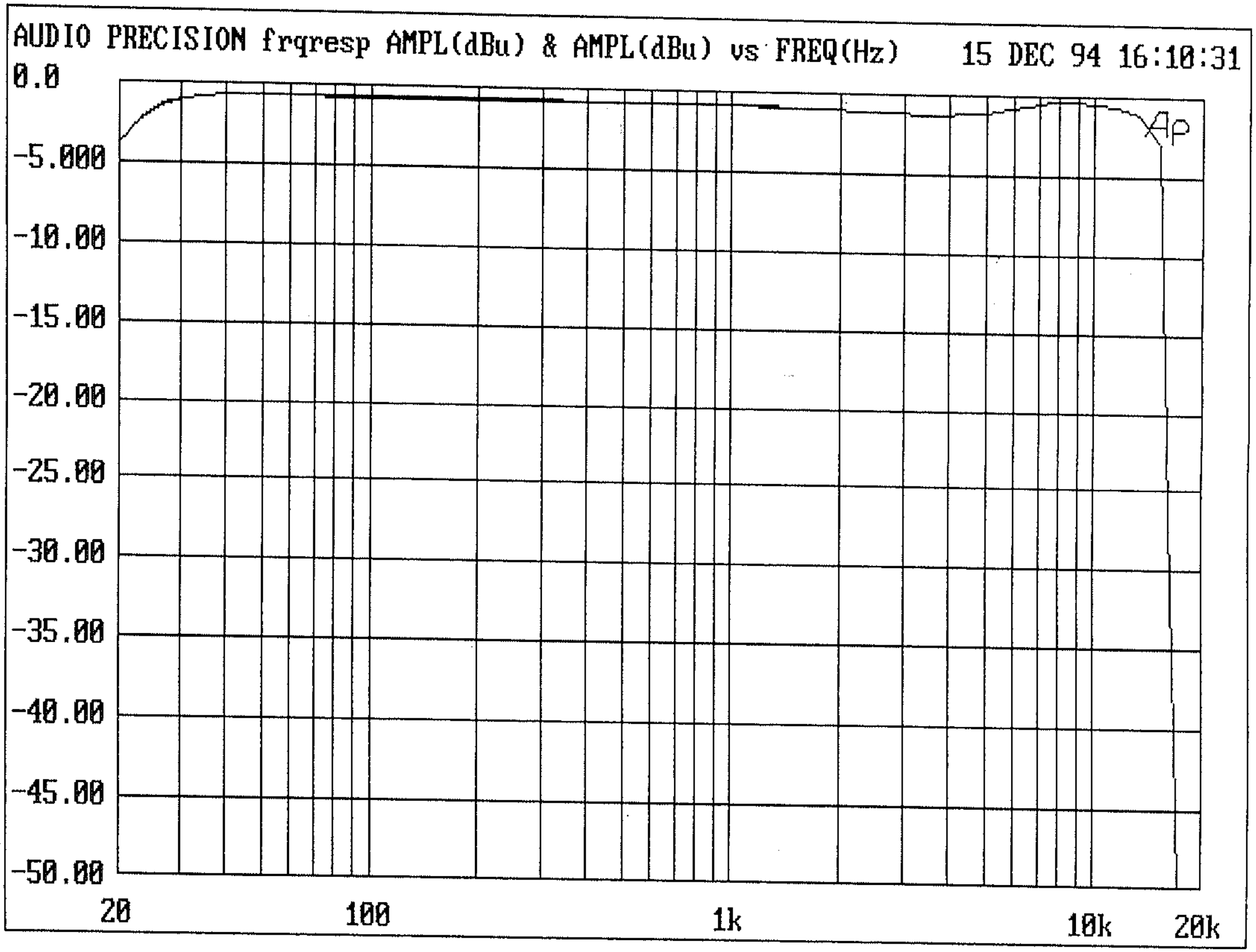

02 Aug
AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
[redacted] 94 13:22:07
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap


AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
10 NOV 94 11:18:58
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

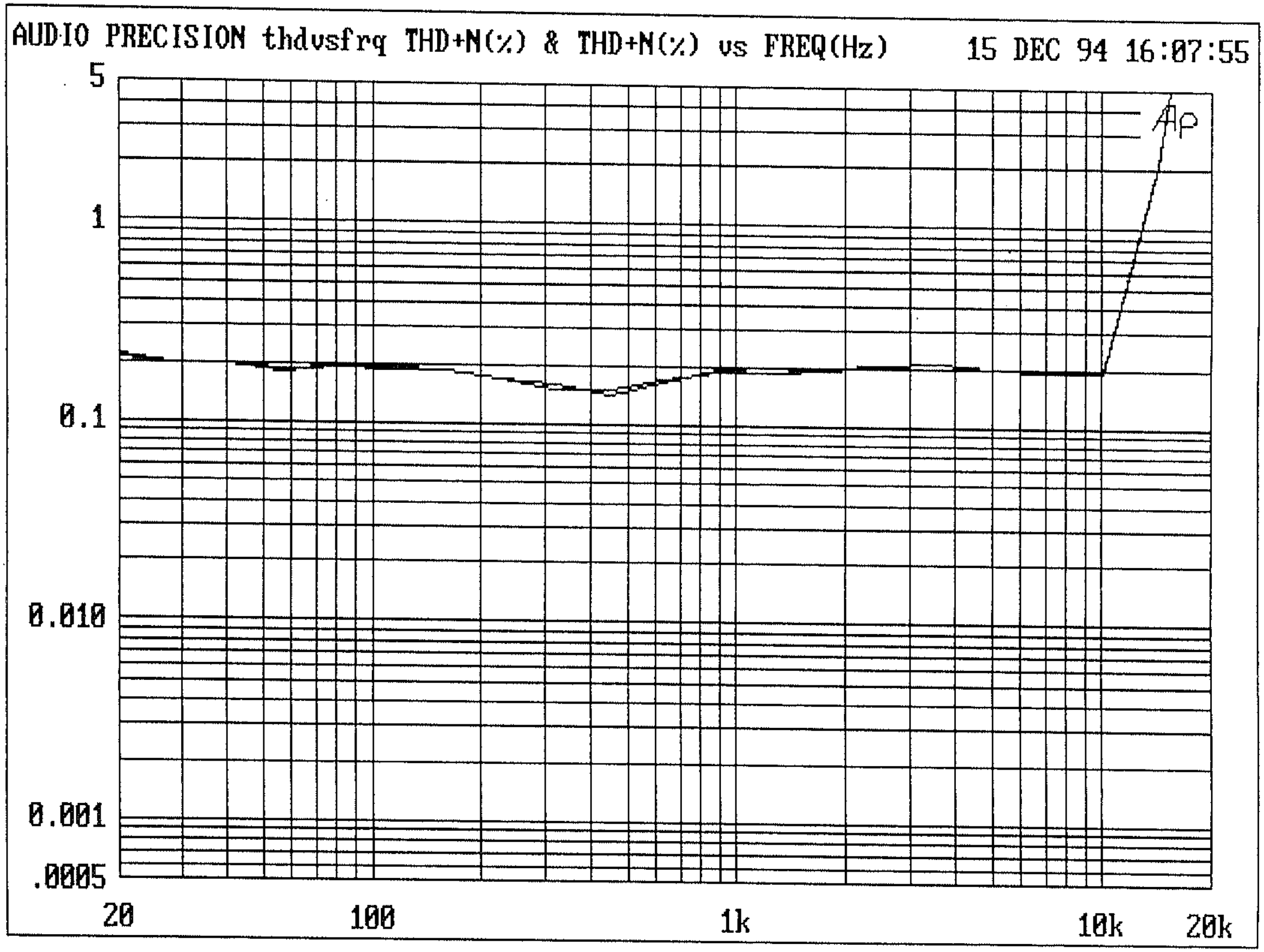


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
10 NOV 94 11:23:09
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
AP

VSA DR FM2

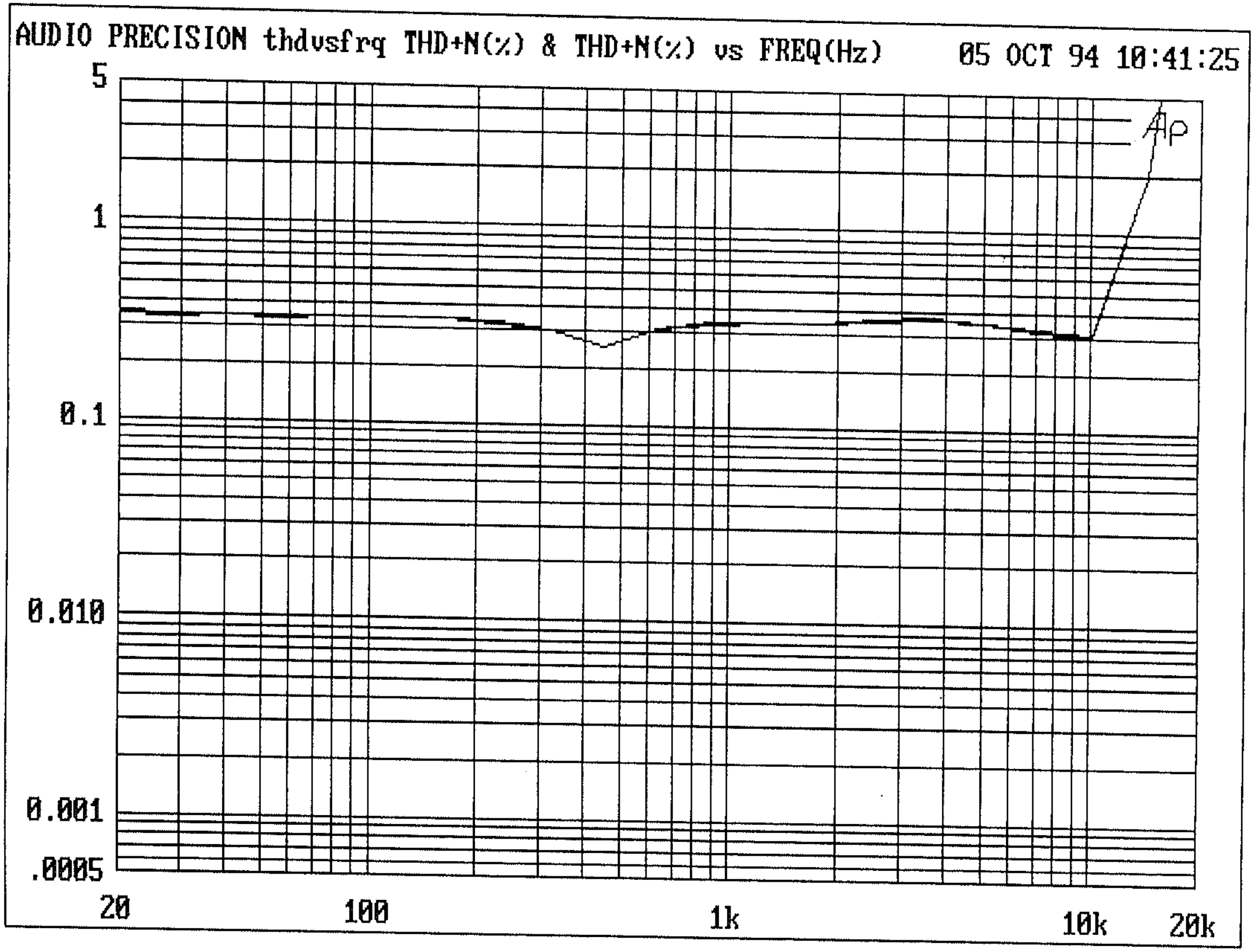

AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
15 DEC 94 16:10:31
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

USADRFM2

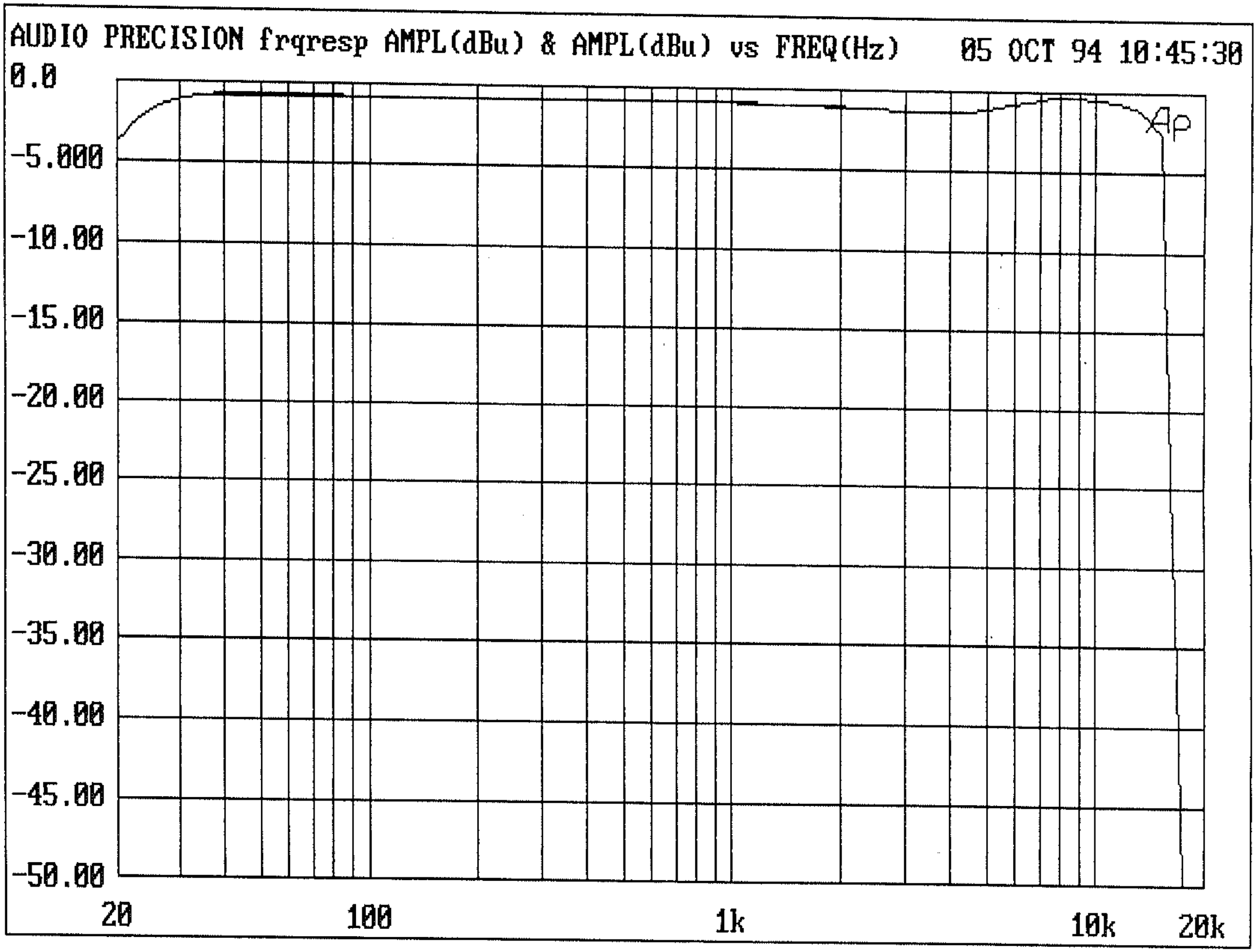


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
15 DEC 94 16:07:55
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

USADR FM2

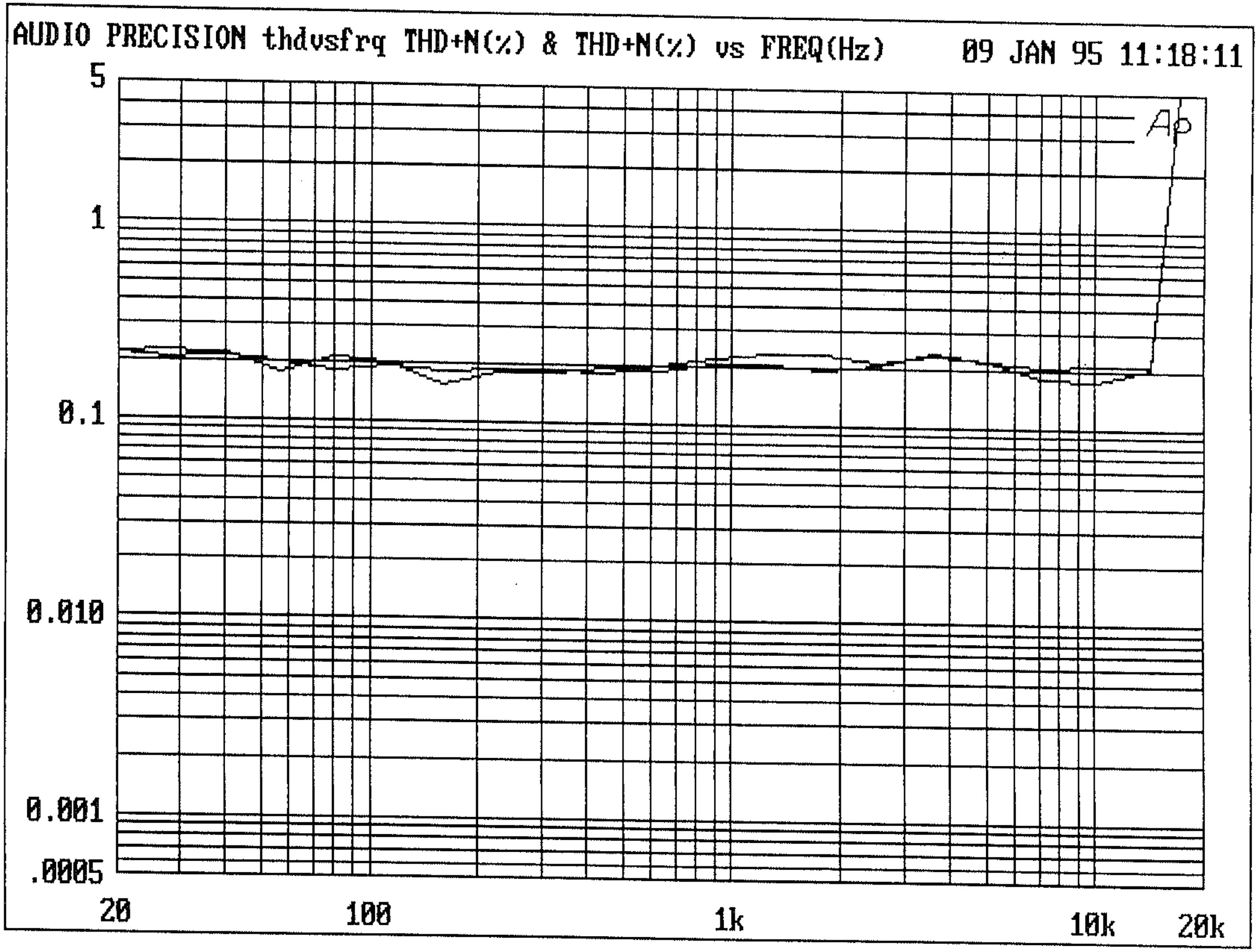

AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
05 OCT 94 10:41:25
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

USADRFMZ

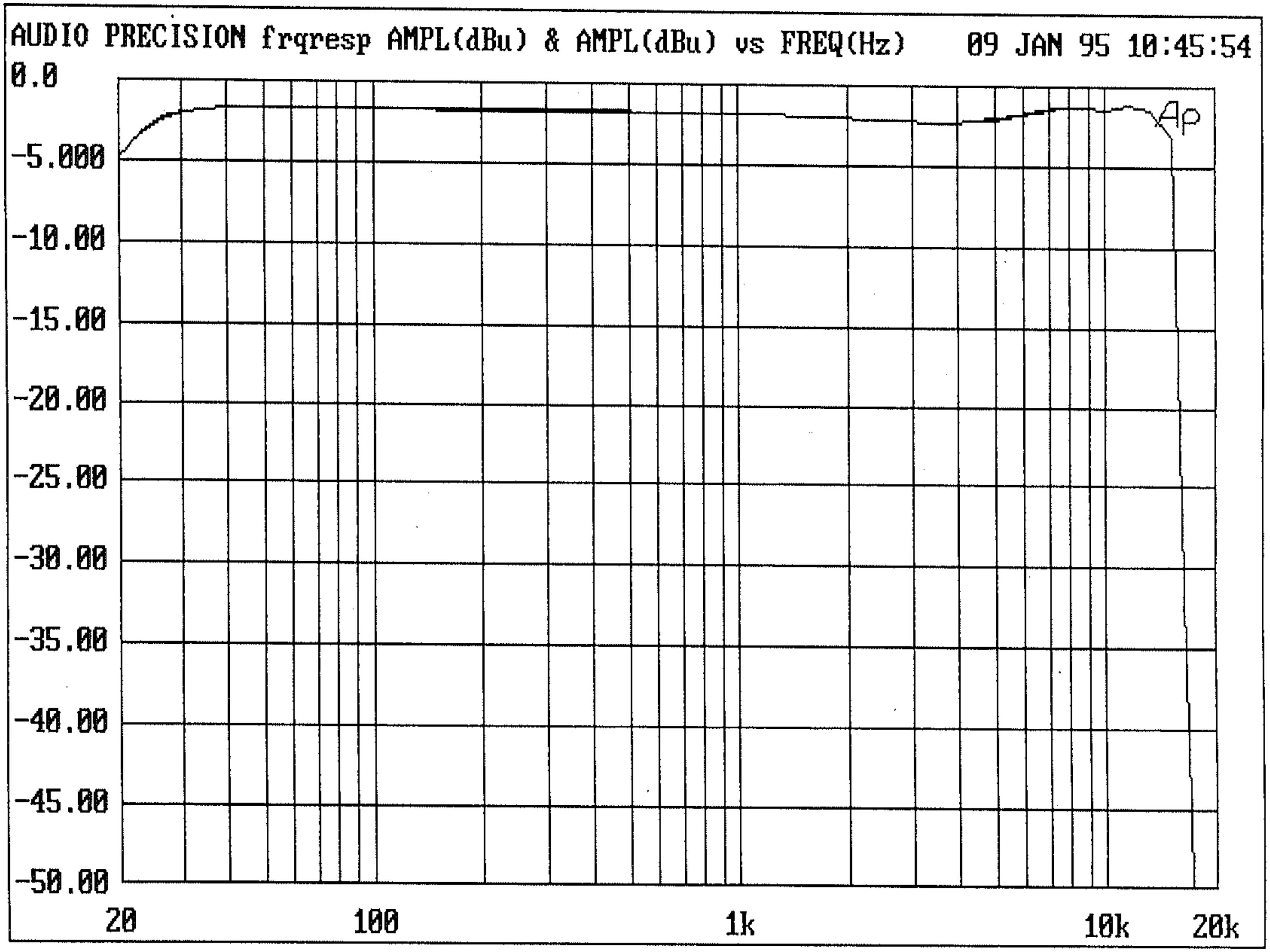


AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
05 OCT 94 10:45:30
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

USA DR FM2


AUDIO PRECISION thdvsfrq THD+N(%) & THD+N(%) vs FREQ(Hz)
09 JAN 95 11:18:11
5
1
0.1
0.010
0.001
.0005
20
100
1k
10k
20k
Ap

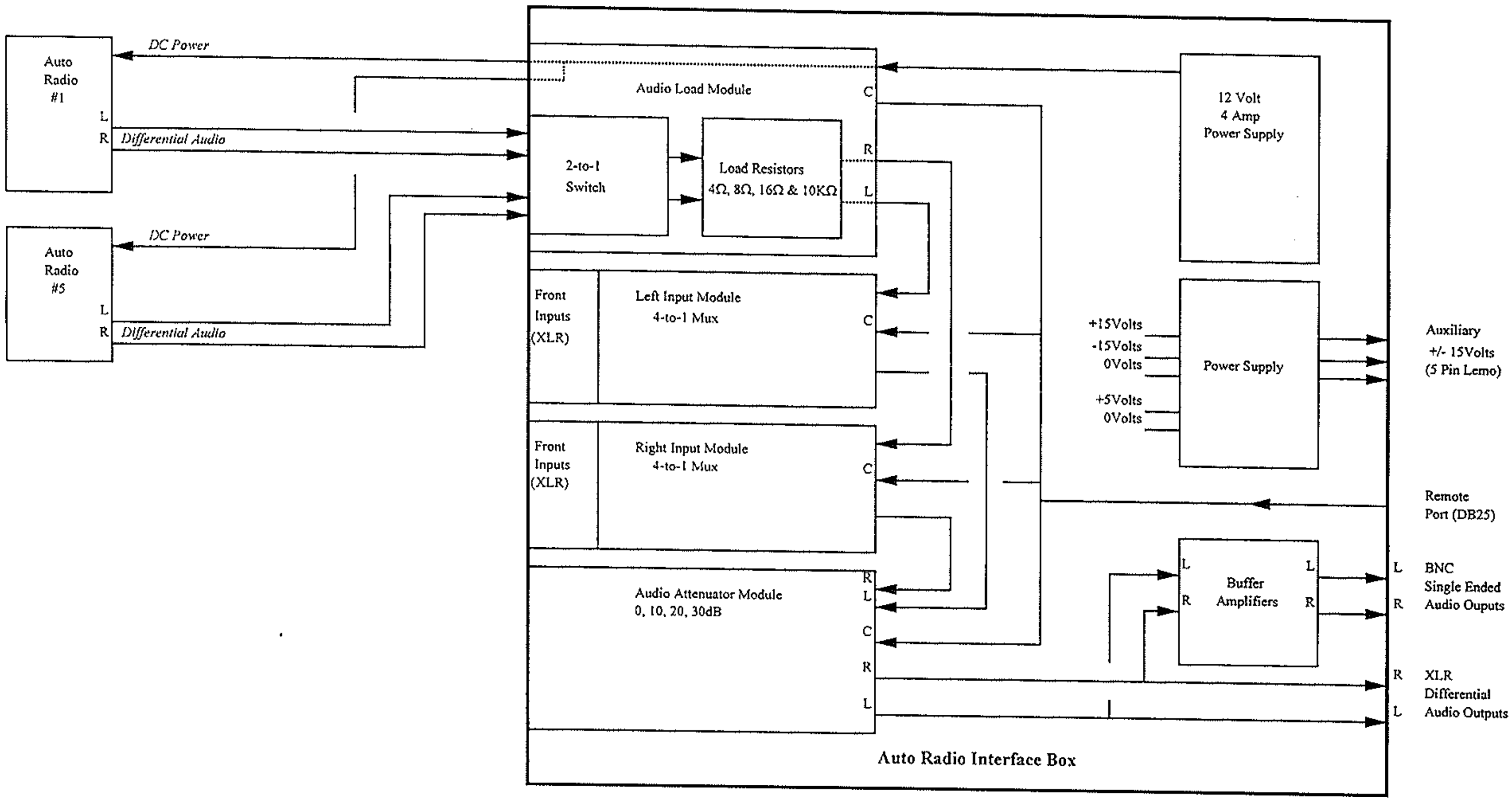
VSADR Fm2

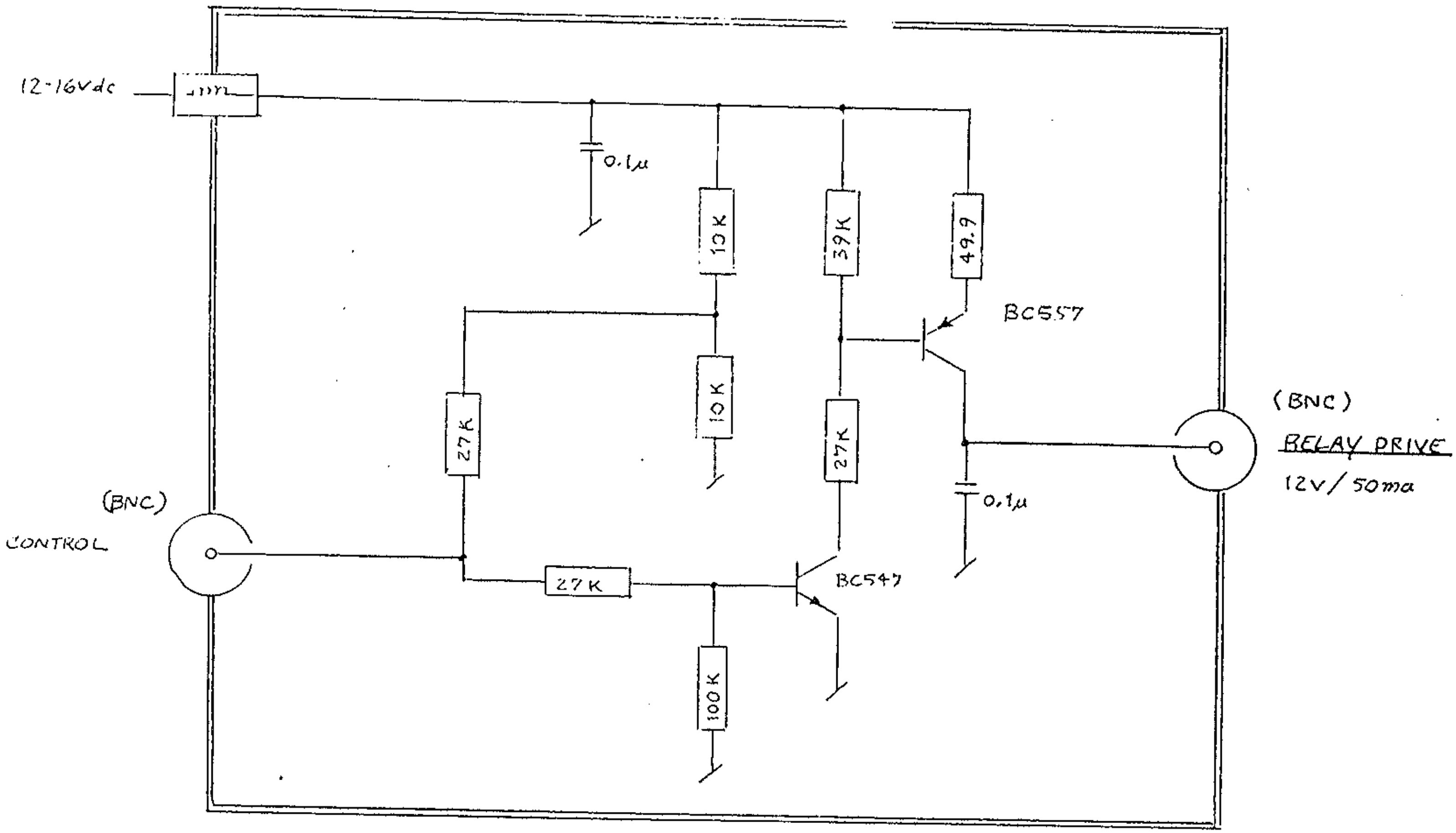

AUDIO PRECISION frqresp AMPL(dBu) & AMPL(dBu) vs FREQ(Hz)
09 JAN 95 10:45:54
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
20
100
1k
10k
20k
Ap

APPENDIX R

Custom Equipment

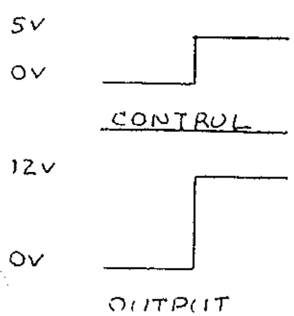
EIA Digital Audio Radio Test Laboratory

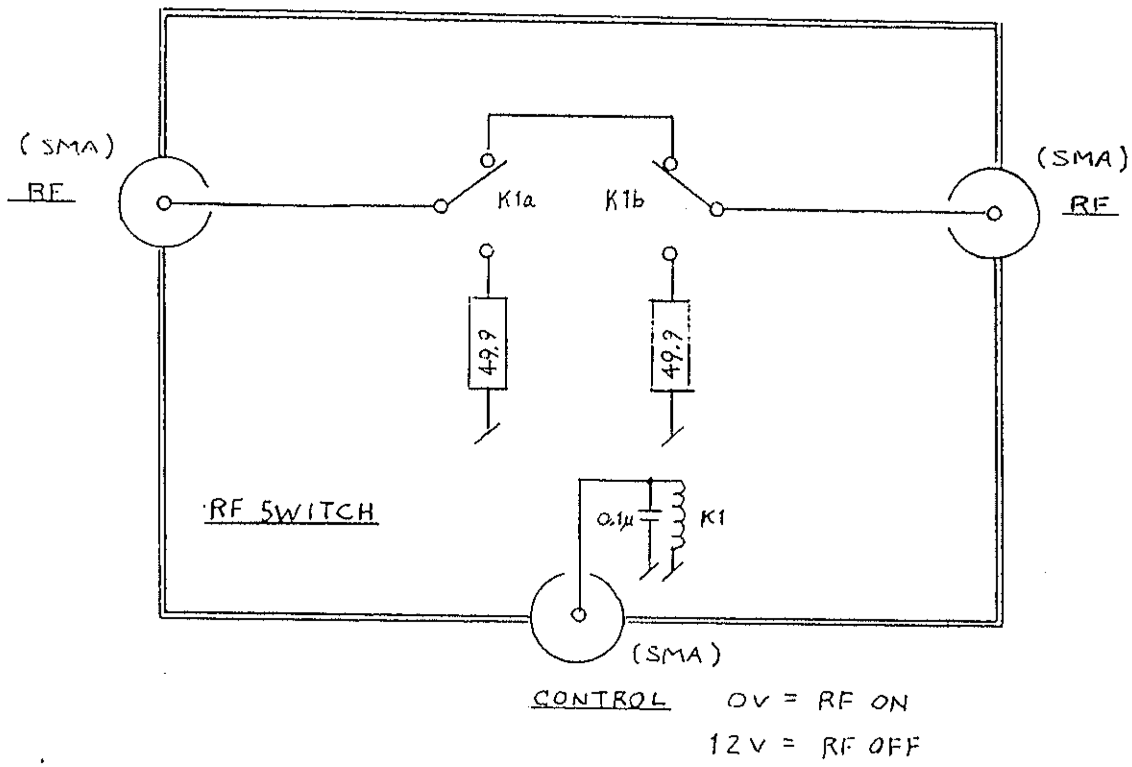




TTL RELAY DRIVER

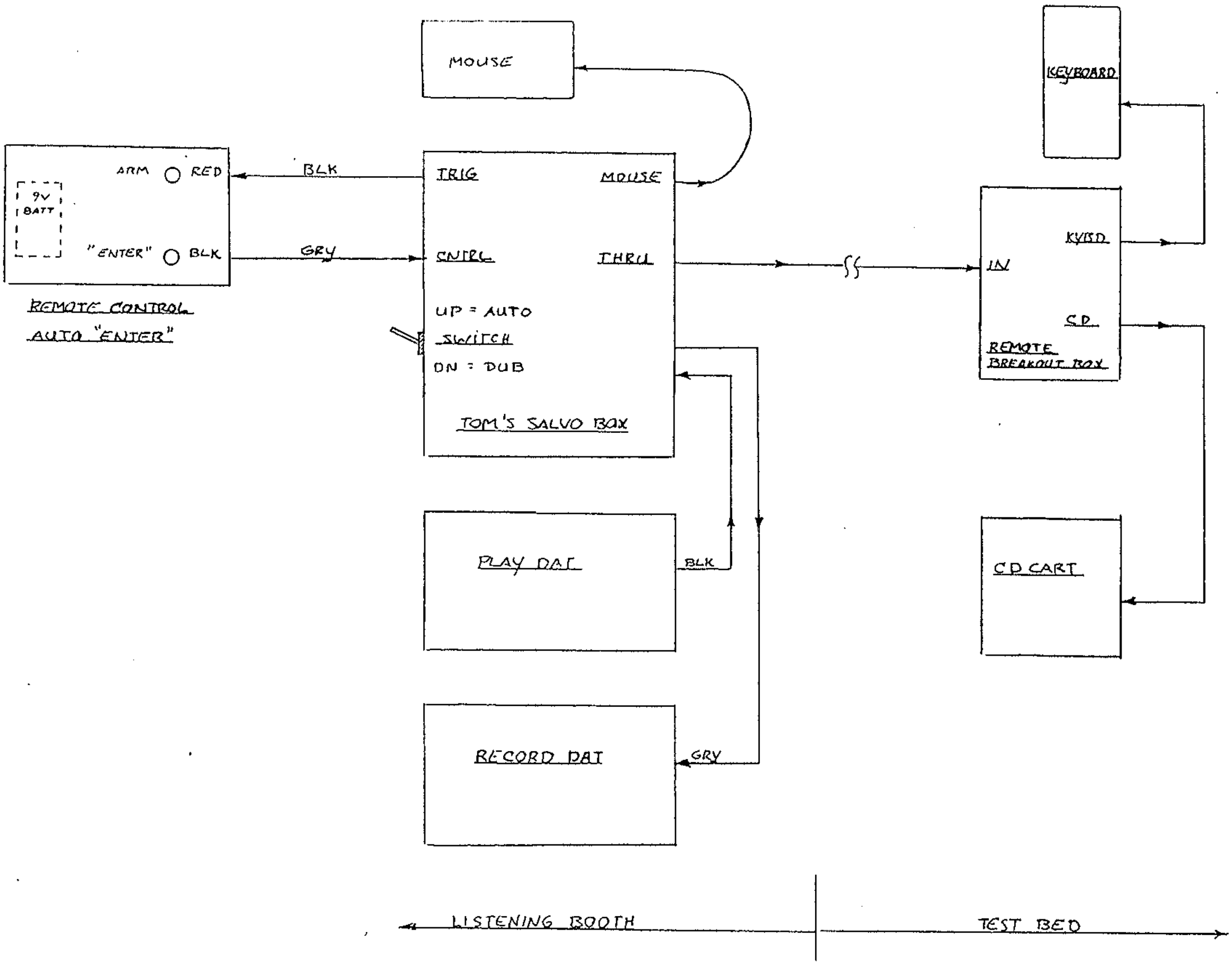
1/95 RMS



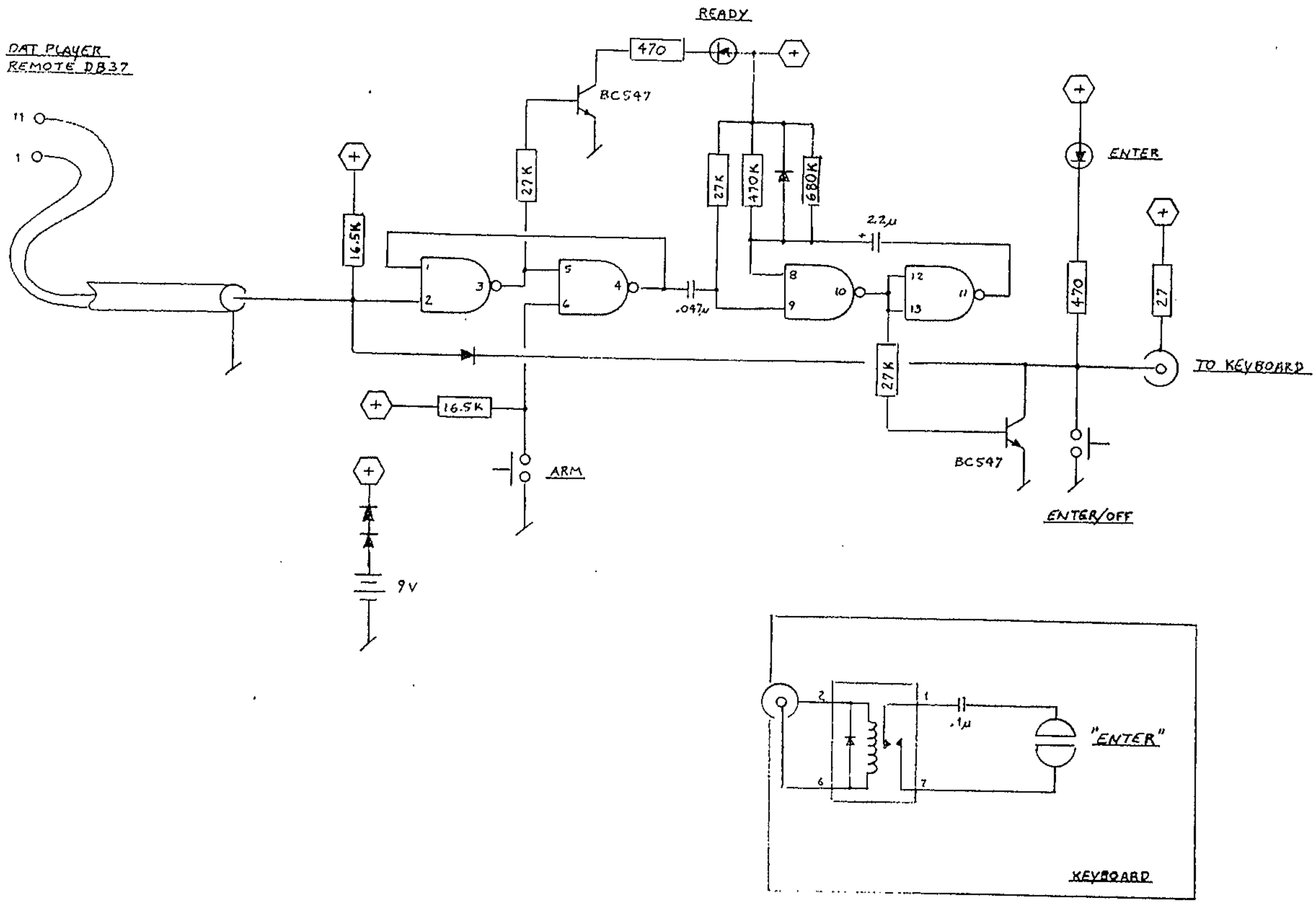


RF SWITCH

1/95 RMS



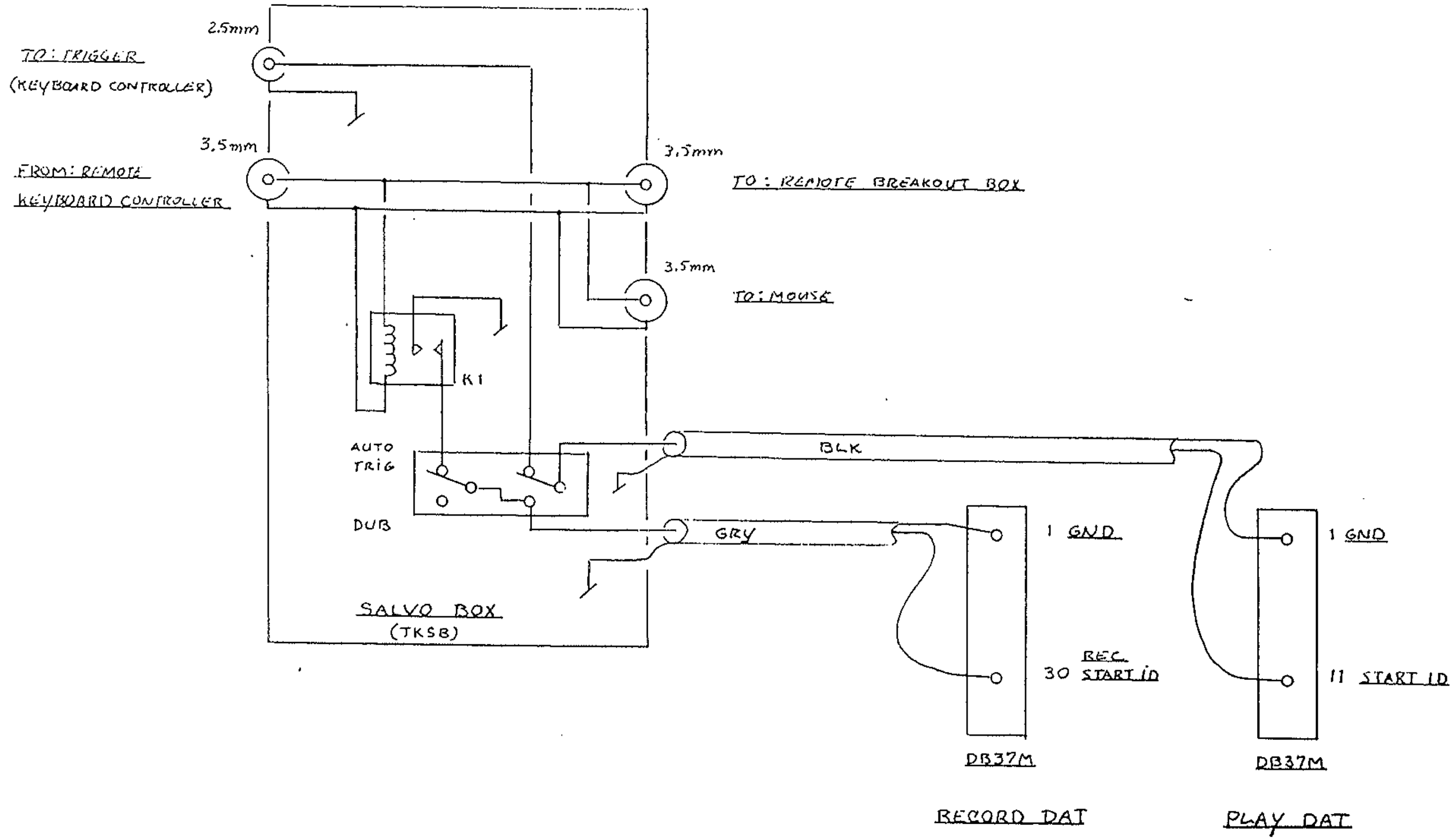
2/95 RMS
REMOTE CONTROL
CONNECTION DIA.



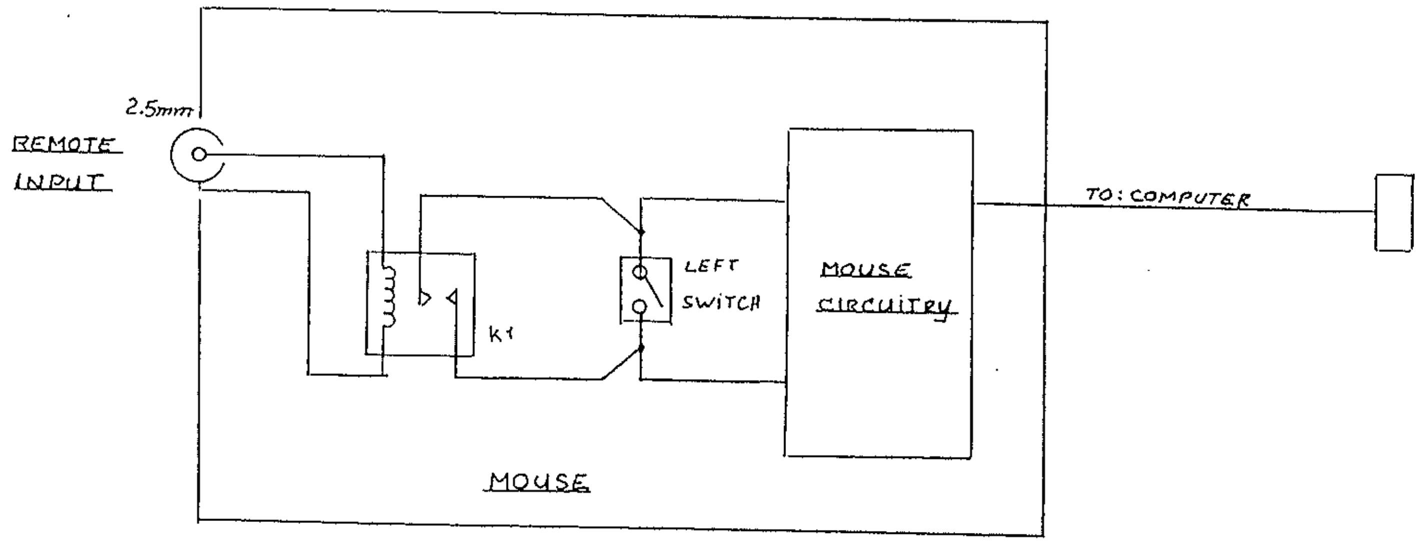
11/94 RMC

EIA

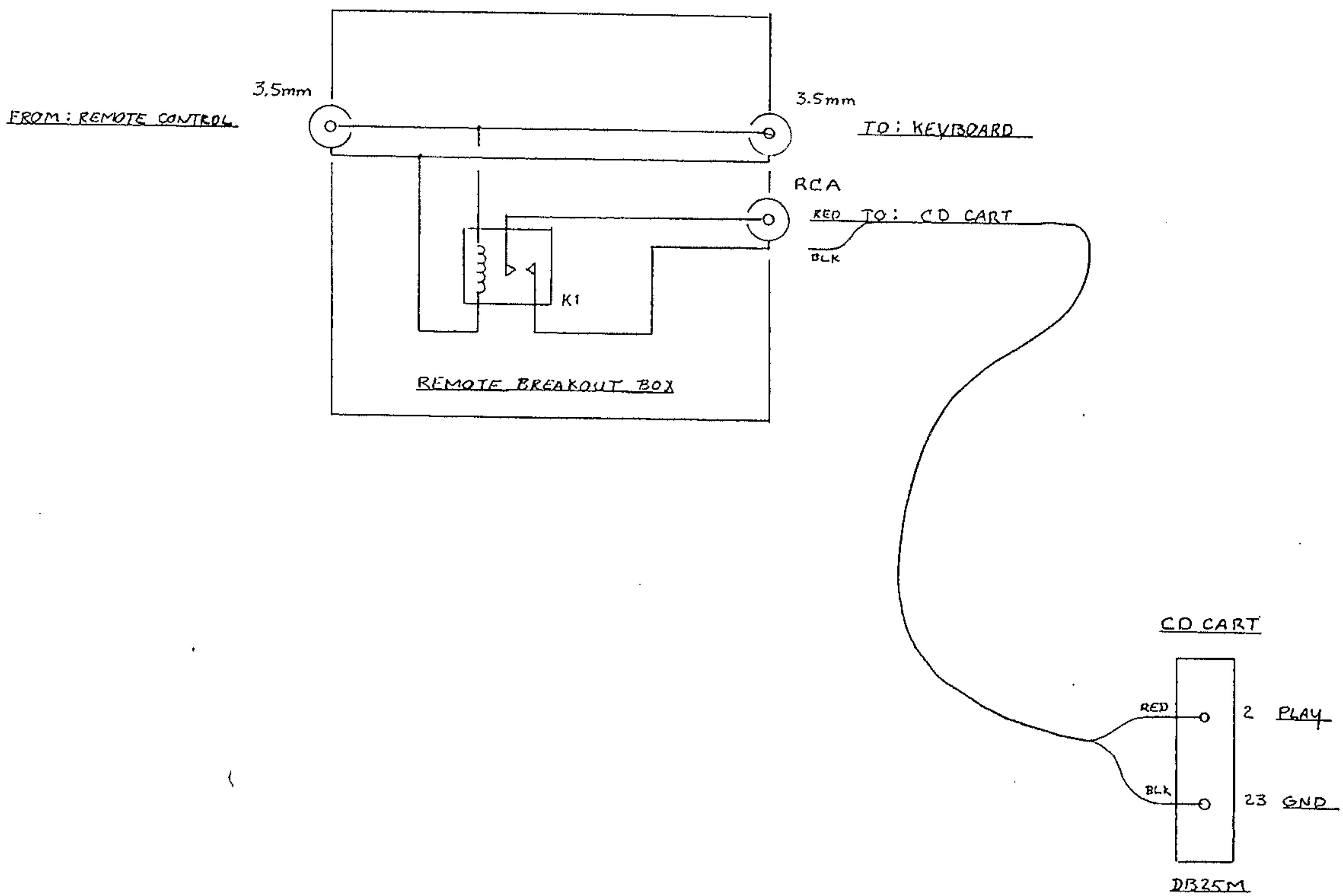
REMOTE KEYBOARD CONTROLLER



2/95
TOM'S SALVO CONTROL BOX



2/95 RMS
MOUSE MODIFICATION



2/95 RMS

REMOTE BRE[illegible]UT BOX

EIA/DAR

PROJECT: HIGH SPEED RF SWITCH
DATE: Nov/10/94
PREPARED BY: Robert McCutcheon

SPECIFICATIONS

FREQUENCY RANGE:
500KHZ - 110MHZ 1.2dB

INPUT/OUPUT IMPEADANCE:
50 OHMS

INSERTION LOSS VS FREQUENCY:
(RF INPUT LEVEL REFERENCE: 0dBm)

FREQ.	LOSS (dB)
200KHZ	25.86
500KHZ	24.50
1000KHZ	24.37
1660KHZ	24.21
5.0MHZ	24.35
10.0MHZ	24.31
50.0MHZ	24.18
94.1MHZ	23.68
100MHZ	23.53
110MHZ	23.44

MAXIMUM INPUT POWER:
+ 10dBm

SWITCH SETTLING TIME:
< 5us

ISOLATION: (Off Characteristics)
(RF OUTPUT LEVEL REFERENCE: -24dBm)

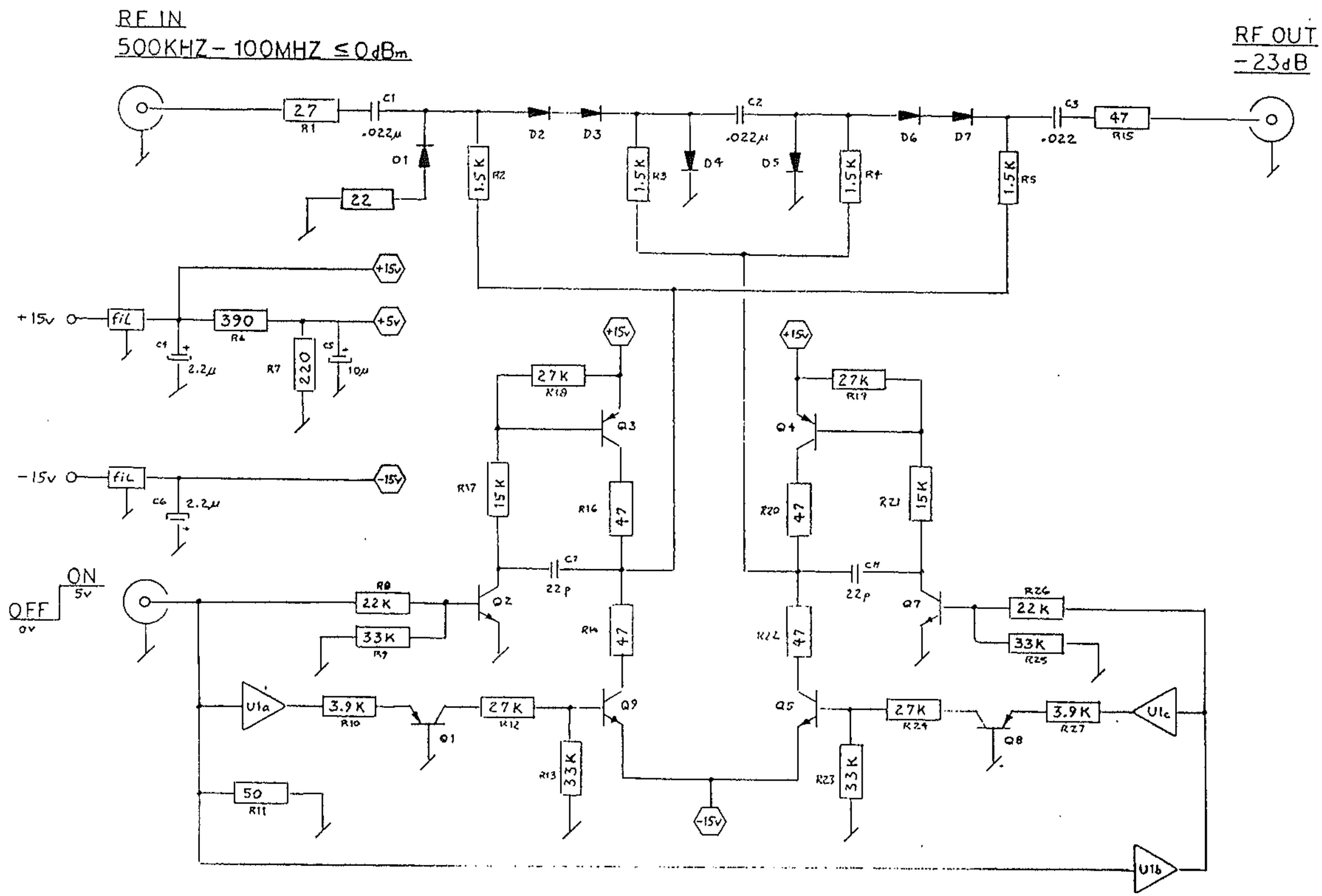
500KHZ - 1700KHZ	> 60dB
4MHZ	> 55dB
50MHZ	> 35dB
100MHZ	> 29dB

CONTROL:
* 5 Volt TTL compatable
* 0V = RF SIGNAL OFF
* 5V = RF SIGNAL ON

LINEARITY:

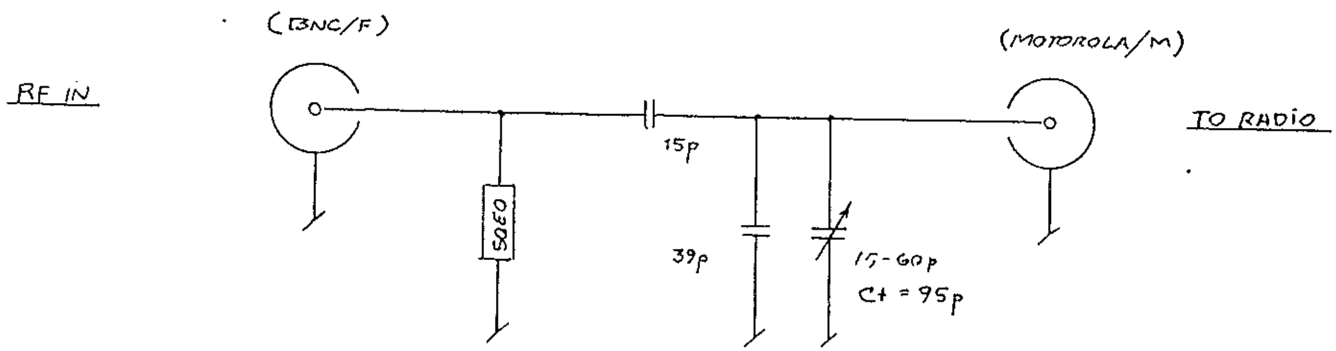
INPUT POWER (dBm)		OUTPUT POWER (dBm)	
14.18		-10.30	
13.00		-10.45	
12.00		-10.65	
11.07		-10.88	
10.06	*COMPRESSION*	-11.20	
9.09		-12.04	
8.11		-13.04	
0.05dBm	--------	-21.75	*- - RECOMMENDED OPERATING LEVEL - -*
-9.95		-31.70	
-20.08		-41.50	

POWER REQUIREMENTS:
+ 15Volts dc @ 50ma
- 15Volts dc @ 20ma



HI SPEED RF SWITCH

11/94 RMc



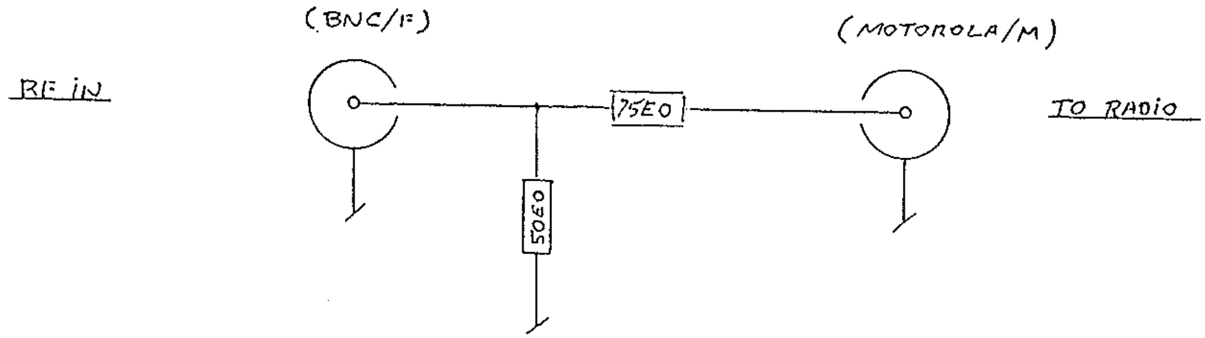
NOTES: * RF TIGHT ENCLOSURE

* INSERTION LOSS: VARIABLE W/FREQ.

EIA/DAR LAB

FORD AM DUMMY ANTENNA

1/95 RMS



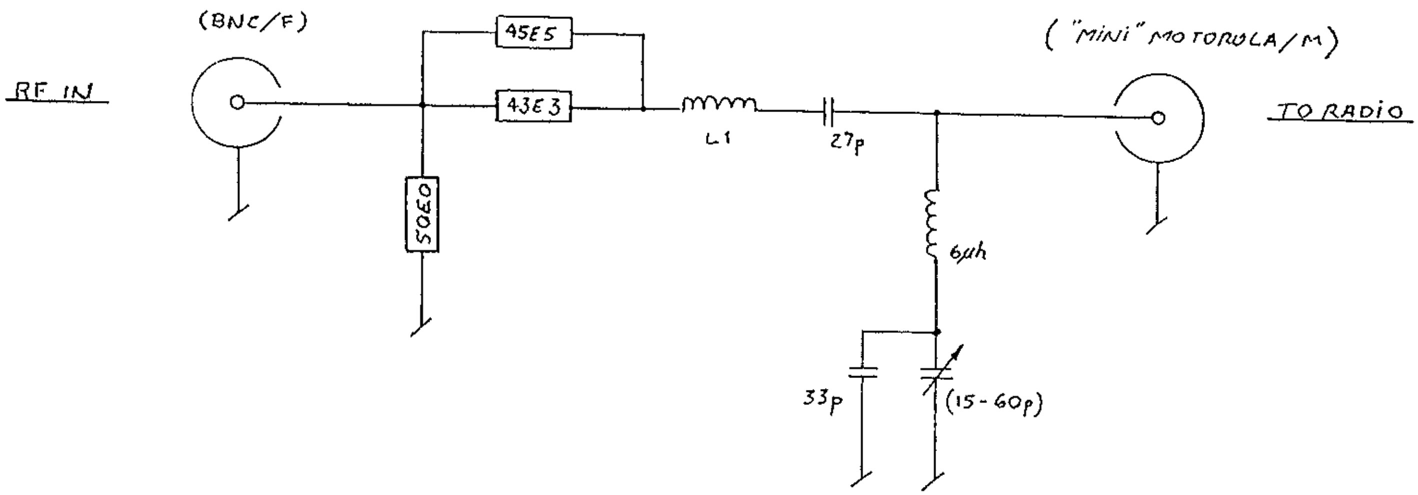
NOTES: * RF "TIGHT" ENCLOSURE

* INSERTION LOSS: 4.88dB W/ PAC/MIDLINE RADIOS (100Ω IMP)

EIA/DAR LAB

FORD FM DUMMY ANTENNA

1/95 RMS



NOTES: * L1 HAND WOUND - 14 T/24GA/0.125 DIA./0.75 L

* RF "TIGHT" ENCLOSURE

* "MINI" MOTOROLA: MALE "MOTOROLA" STYLE, SMALLER DIA.

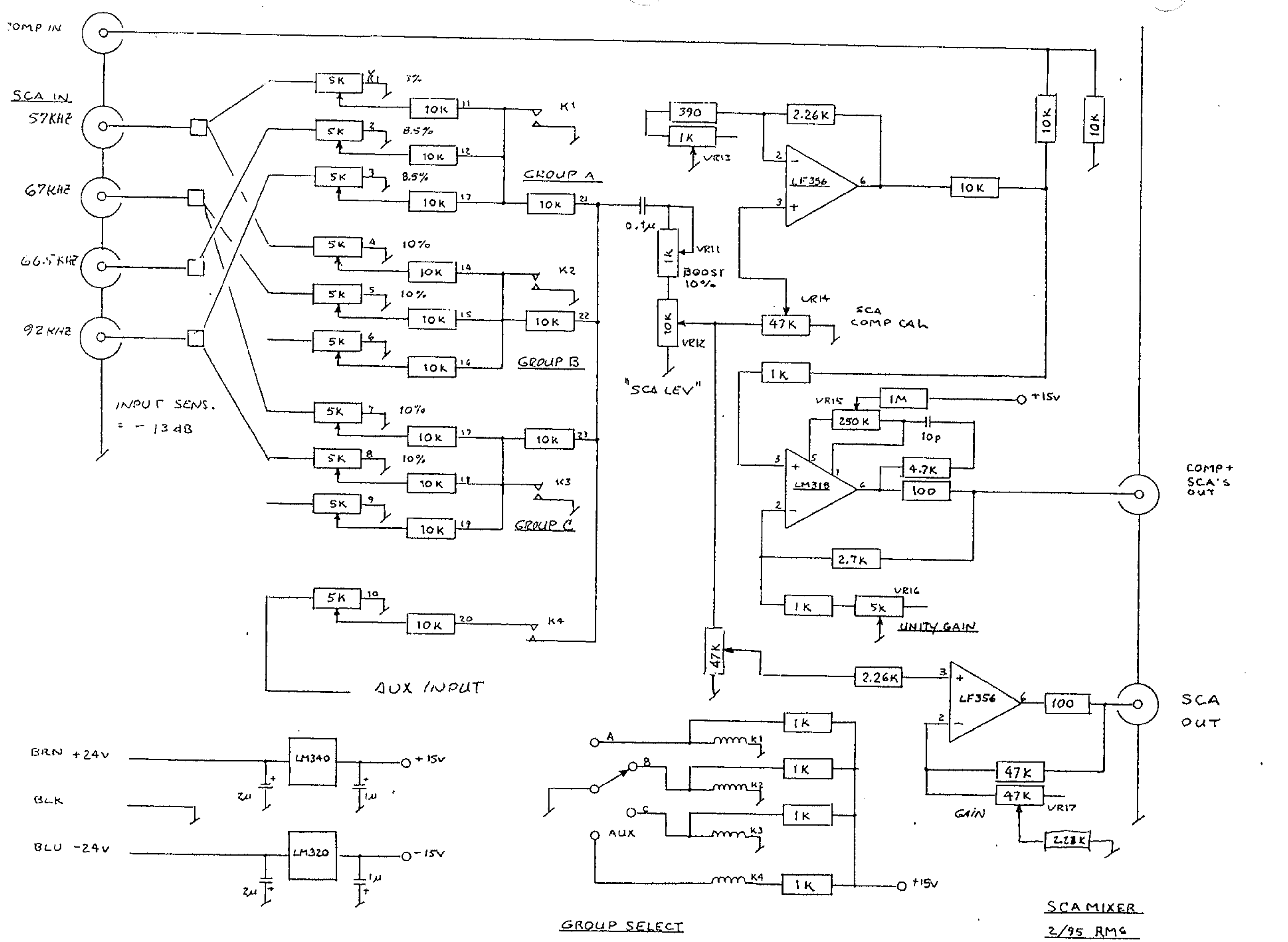
* INSERTION LOSS: FM/6dB
AM/ VARIABLE W/FREQ.

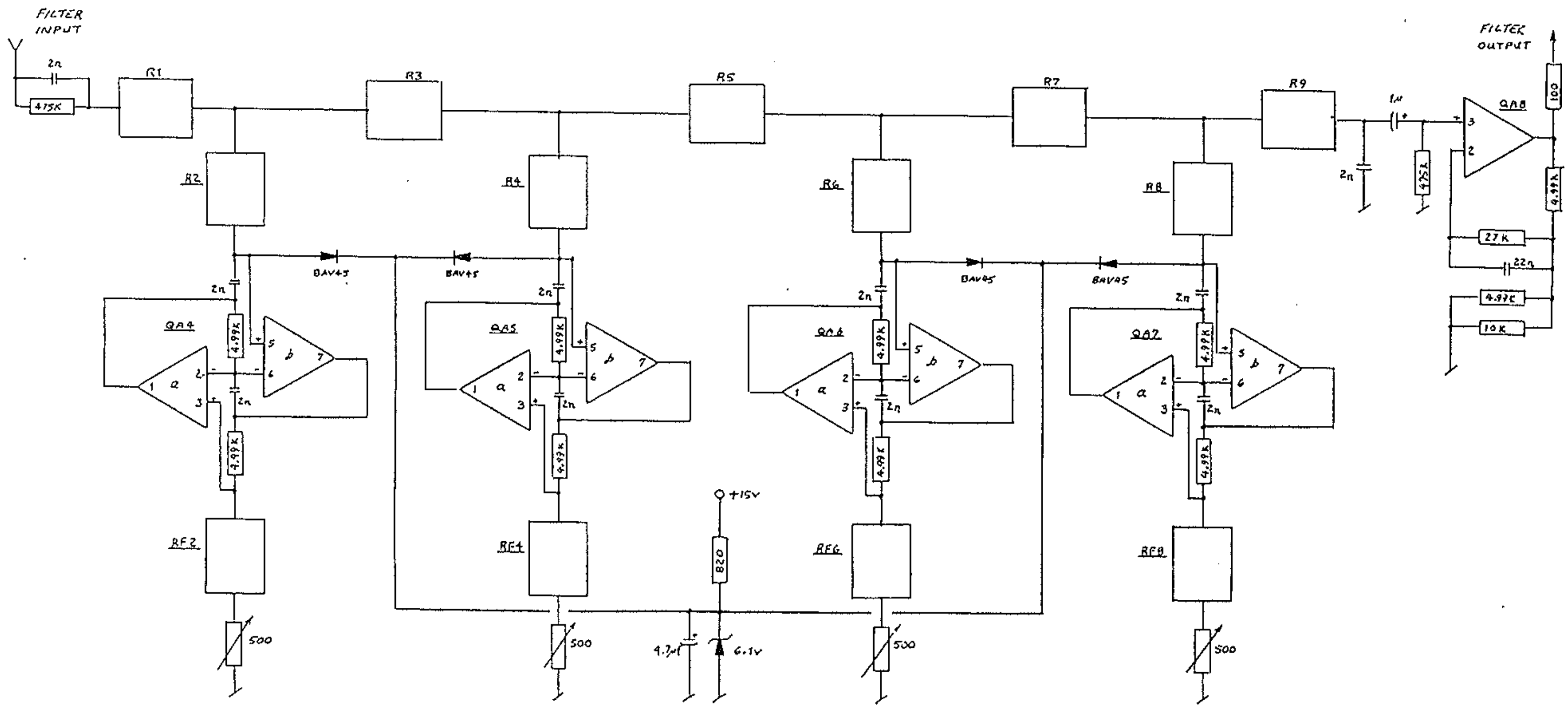
* MFG: JFW

EIA/DAR LAB

DELCO AM/FM DUMMY ANTENNA

1/95 RMS


COMP IN
SCA IN
57KHZ
67KHZ
66.5KHZ
92KHZ
INPUT SENS.
= -13dB
5K
10K
3%
8.5%
8.5%
10%
10%
10%
10%
K1
K2
K3
K4
GROUP A
GROUP B
GROUP C
AUX INPUT
0.1µ
1K
VR11
BOOST 10%
10K
VR12
"SCA LEV"
390
1K
VR13
2.26K
LF356
47K
VR14
SCA COMP CAL
1K
1M
VR15
+15V
250K
10p
LM318
4.7K
100
2.7K
1K
5K
VR16
UNITY GAIN
COMP + SCA'S OUT
47K
2.26K
LF356
100
SCA OUT
47K
47K
VR17
GAIN
2.23K
BRN +24V
BLK
BLU -24V
LM340
LM320
+15V
-15V
2µ
1µ
A
B
C
AUX
+15V
GROUP SELECT
SCA MIXER
2/95 RMG

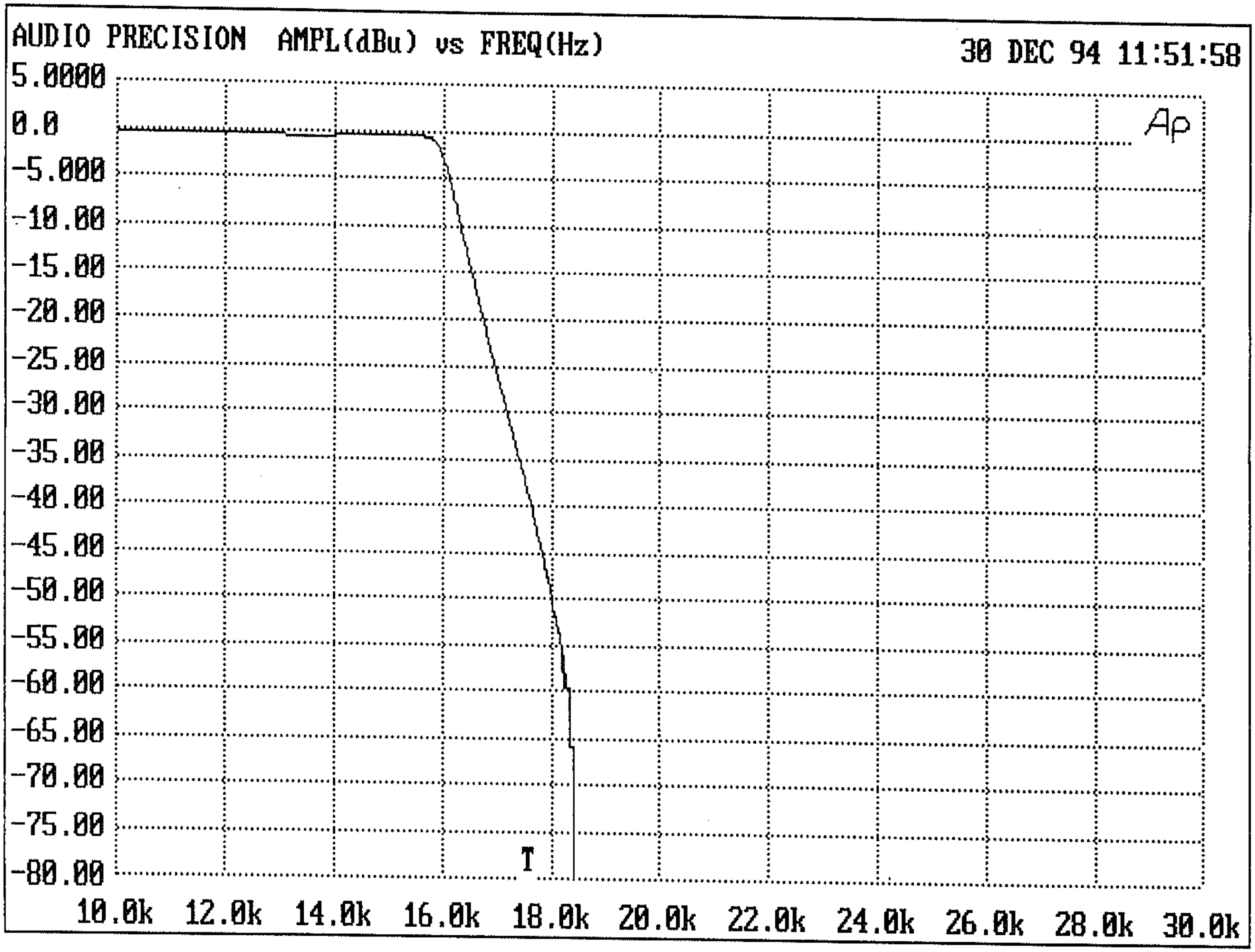


QA4 - 7 NE5532

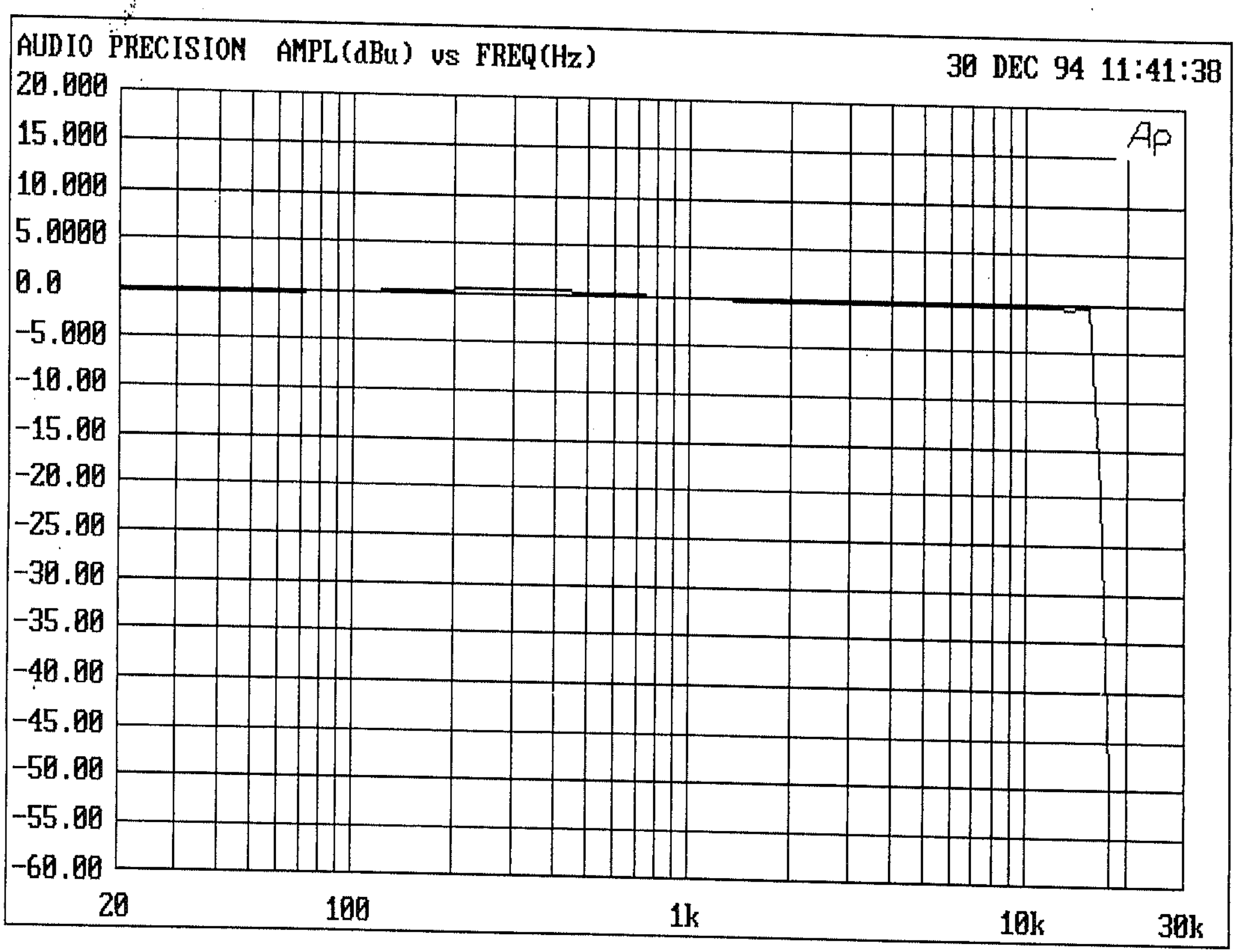
QA8 NE5534

ALL IC'S DECOUPLED WITH 0.1uf CER CAPACITORS

9 POLE ELLIPTICAL FILTER



AUDIO PRECISION AMPL(dBu) vs FREQ(Hz)
30 DEC 94 11:51:58
5.0000
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
-55.00
-60.00
-65.00
-70.00
-75.00
-80.00
Ap
T
10.0k
12.0k
14.0k
16.0k
18.0k
20.0k
22.0k
24.0k
26.0k
28.0k
30.0k


AUDIO PRECISION AMPL(dBu) vs FREQ(Hz)
30 DEC 94 11:41:38
Ap
20.000
15.000
10.000
5.0000
0.0
-5.000
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
-40.00
-45.00
-50.00
-55.00
-60.00
20
100
1k
10k
30k

APPENDIX S

Equivalent Noise Bandwidth of Noise Filters

Noise Equivalent Bandwidth Calculation by the Method of Least Squares Approximation

Introduction

If white noise with a two-sided power spectral density of $N_0/2$ is passed through a band-pass filter with a transfer function H(f), the average noise power at the filter output is given by

$$P_{N0} = \int_{-\infty}^{\infty} |H(f)|^2 (N_0/2)\,df = N_0 \int_{0}^{\infty} |H(f)|^2 df \qquad (1)$$

If the filter were ideal (i.e. a "brick-wall" magnitude response), with a bandwidth B_N and a mid-band gain (or point of minimum insertion loss) H_0, the output average noise power would be

$$P_{N0} = N_0 B_N H_0^2 \qquad (2)$$

By equating the expressions in equations (1) and (2), it is possible to determine the noise equivalent bandwidth of a real filter as

$$B_N = (1/H_0)^2 \int_{-\infty}^{\infty} |H(f)|^2 df \qquad (3)$$

where the mid-band, or maximum gain of the real filter has been adjusted to an amplitude of H_0. Note that only the filter's magnitude response is involved here. By means of equation (3) a band-pass filter's noise equivalent bandwidth may be determined if the band-pass filter's transfer function is known analytically, or can be determined experimentally.

While the analytic expressions for band-pass filters are well known, they do not take into account the de-tuning effects of lossy components or the parasitic coupling between various components and between components and the enclosure. Therefore, while it would be easier to integrate an analytical expression for H(f), it would not properly include the effects of the real filter that is being used. Consequently it was decided to obtain the filter transfer function experimentally and evaluate equation (3) using numerical methods.

A network analyzer is a natural choice for obtaining these type of measurements since it evaluates the frequency response (offering both magnitude and phase information) of devices under test. The analysis band can be easily adjusted and proper calibration can be done to place the analyzer's reference planes at the input and output of the filter to be characterized. Thus, only the filter's response is obtained, including the parasitic and package effects. Data may be obtained in linear magnitude format, which is the natural choice for use with equation (3). An HP 8753D automatic vector network analyzer was used for data collection.

There are several limitations involved in obtaining an accurate representation of H(f) which must be considered and compensated for before it is certain that the B_N value obtained in equation (3) is correct. These limitations are-

(1) the expression for H(f) thus obtained is truncated in frequency, that is, H(f) information is not obtained for all frequencies, $0 \le f \le \infty$, but rather over a finite range $f_L \le f \le f_U$.

(2) The analyzer provides a discrete representation of H(f) in the frequency domain, i.e. it is not a continuous response. The analyzer used could provide up to 1601 calibrated data points across any desired frequency band.

(3) The amplitude information thus obtained is subject to errors due to noise and numerical round-off.

(4) As a result of (2) and (3) above, the filter's actual shape, specifically it maximum transmission gain (minimum insertion loss) location, could be missed if proper care is not taken in obtaining the data.

To ensure that the filters are properly characterized given the above limitations, the following steps are taken-

(1) Each filter was characterized on the 8753D to find its minimum insertion loss point (H_0 in equations (2) and (3)), which was chosen as the filter's center frequency. The upper and lower frequencies, f_u and f_l respectively, over which the filter

characterization would occur were chosen as the frequency(s) at which the magnitude response was 5% of the filter's minimum insertion loss; i.e., $0.05H_0$. Since the integral in equation (3) involves $|H(f)|^2$ this choice ensures that 99.5% of the area under the filter's response curve is accounted for. The validity of this choice was ascertained by checking the $0.025H_0$ frequency band as well; the resulting B_N values were practically unchanged (<~0.1% difference) from the $0.05H_0$ interval.

Each filter's transmission maximum was normalized to 1.0, and the rest of the data points scaled accordingly. The magnitude response of each filter as obtained from the network analyzer appears at the end of this appendix. Note that the scale used is linear, not logarithmic.

It is the area under this curve which must be evaluated so that the equivalent "brick-wall" filter bandwidth may be obtained. Note that by normalizing the filter's insertion loss to 1.0, the $(1/H_0)^2$ term appearing in equation (3) is absorbed into the expression for $|H(f)|^2$.

(2) To verify that the transmission peak H_0, was the proper one, the frequency band, defined as $f_u - f_l$ was shifted by an amount $f_\Delta /2$, where

$$f_\Delta = (f_u - f_l)/1600 \tag{4}$$

where 1600 is the maximum number of intervals between the 1601 data points that the 8753D can display. This frequency shift also verified that the filter response in question had no spurious responses in the region of interest.

(3) and (4) To account for the HP 8753D's finite resolution and frequency step size, it was decided that rather than performing a Simpson's rule integration on the raw data, a curve fitting procedure would be employed using polynomial functions, as follows

$$|H(f_i)|^2 \approx \sum_{n=0}^{N} a_n \phi_i(n) \tag{5}$$

for $i=0,1,2,\ldots Imax$, where Imax is the number of data points obtained from the 8753D and the $\phi_i(n)$ represent the polynomial functions that will be used to fit the filter transfer function over the interval from f_l to f_u. The maximum polynomial order, N, is chosen based on a convergence criterion to be discussed later. The $H(f_i)$ are the filter transfer function data points as obtained from the 8753D using the linear scale at the discrete frequencies f_i. The $H(f_i)$ are normalized so that the filter's minimum insertion loss is 1.0. The coefficients a_n will be determined using the method of least squares so as to minimize the error in a mean square sense. Thus, the polynomial expansion on the right hand side of equation (5) will represent the best approximation to the actual filter transfer function.

To guarantee an accurate representation of H(f) over the interval from f_l to f_u, three sets of data points were taken, the first set consisting of 1601 points, the second consisting of 3201 points, and the third set consisting of 6401 data points taken over the interval from f_l to f_u. Since the 8753D can only generate 1601 calibrated data points over any one interval, the interval from f_l to f_u was sub-divided into two and then four sections to obtain 3201 and 6401 data points respectively. The sub-sections were combined to re-create the original interval of f_l to f_u in the computer program which performed the curve fit technique. Thus, in equation (5), Imax1=1601, Imax2=3201 and Imax3=6401 points respectively.

The choice of polynomial functions on the right hand side of equation (5) is appropriate since they form a linearly independent set of functions on an interval [a,b], and because the analytical expressions for the actual band-pass filter transfer functions, in both the pass-band and in the skirt regions can ultimately be expressed in terms of polynomial functions. Thus the use of polynomial functions to fit the data represent a "natural" choice.

To avoid numeric overflow that would inevitably occur if the curve fit was attempted in the region of the filter's center frequency, the filter's response was shifted from f_0, its original center frequency, down to zero frequency. Thus, f_l and f_u are shifted to f_l-f_0 and f_u-f_0, respectively. Finally the band of interest, namely f_u-f_l was re-defined to be the interval [-1,1], and this interval was subdivided into the appropriate number of sections corresponding to the appropriate f_Δ in equation (4). Once the unknown coefficients, the a_n, have been obtained, the noise equivalent bandwidth, B_n, is re-normalized to the band f_u-f_l.

It should be noted that there is no loss of accuracy introduced by re-defining the bandwidth of the frequency shifted filter response to the interval [-1,1], even though this process will shift the filter's minimum insertion loss point (i.e. which was defined as f_0 and then shifted to zero frequency) away from zero (unless of course it just so happens that f_0 also corresponds to the midpoint defined as $(f_u+f_l)/2$). This is not a problem since the method of least squares, which is used here, will fit the best curve to the data regardless whether the general shape of the data to be fit is symmetric about the ordinate or not. Since our purpose is to obtain the equivalent "brick-wall" filter response whose area is the same as that of the actual filter being characterized, maintaining the filter shape with respect to its position along an abscissa is not critical, maintaining the area under that filter shape is.

With the filter's band-pass so defined, there is no loss of accuracy if the abscissa is re-defined from a frequency scale to the x-axis, and the ordinate is now referred to as the y-axis. We now have a set of data points described in the x-y coordinate system. The expression in equation (5) may now be expressed as

$$f(x_i) \approx \sum_{n=0}^{N} a_n \phi_n(x_i) \qquad (6)$$

where i=0,1,2,...Imax as before. the coefficients a_n are obtained by applying the method of least squares (a brief description of which appears at the end of this section) to equation (6) followed by Gauss elimination to invert the resulting square matrix. The coefficients resulting from the application of this technique provide the best fit of the expansion function $\phi_n(x_i)$ to the data points. Once coefficients of the polynomial function have been obtained, the area under the curve may be obtained by an analytical integration of this function. Thus, the expression for B_n, the equivalent noise bandwidth of the filter given in equation (3) has been obtained.

The calculated noise equivalent bandwidth for the filters are presented below, along with the frequency band over which the filter data was obtained (the frequencies between which the filter response is $\geq 0.05H_0$) the resulting step size, f_Δ given in equation (4) and the filter's minimum insertion loss, H_0 as obtained from the network analyzer.

Filter Type	f_u-f_l (MHz)	f_Δ (kHz)	H_0 (dB)	B_n (MHz)
AM #1	2.72	1.700	-0.29	1.04400
FM #1	20.00	12.500	-2.72	6.44868
FM #2	20.00	12.500	-2.80	6.41530
L-Band	24.00	15.000	-1.03	11.58017
S-Band	36.00	22.500	-0.90	17.82766

Table 1. Band-pass filter data.

Remarks on the Accuracy of the Expression for B_n

There are two main questions regarding the accuracy of the technique used that must be properly addressed so as to have a high confidence level in the numerical results obtained by this method. They are-

(1) Have sufficient data points in frequency been obtained for the filter, and

(2) is the polynomial order of the $\phi_n(x)$ in equation (6) sufficiently high to ensure the expression for B_n thus obtained has converged to its limiting value.

To ensure that sufficient data points (1601) were obtained for the filter response, three things were done, they are-

1a) The band f_u-f_l was divided into two, then four sections so that the overall band was represented by 3201 and 6401 points, respectively. This was done by taking 1601 data points per band and properly joining the data together in the program to fill the entire band f_u-f_l with data.

1b) The start and stop frequencies, f_l and f_u respectively, where shifted by one half of the frequency step, f_Δ, for the case of 6401 data points for each filter. This was done to verify that there were no spurious transmission peaks between data points, i.e., the data was smooth. It should be noted that the frequency steps given in Table 1 are so fine relative to each filter's center frequency (frequency of minimum insertion loss) that any transmission anomalies, resulting from spurious resonances, would naturally appear in the data as measured. This procedure also improved the possibility that the point of minimum insertion loss was obtained for each filter.

1c) A Simpson's rule integration of the data was performed for each of the three sets of data (1601, 3201, and 6401 points). The resulting values of B_n thus obtained were checked for convergence using the criterion that $\Delta B_n \leq 1.0\%$, where ΔB_n = B_n(3201 points)-B_n(1601 points) and B_n(6401 points)-B_n(3201 points). By doing this, it was ascertained that 1601 data points would be sufficient to characterize the filters listed in Table 1.

For the filter designated FM #1, 801 data points were taken as well, as a check to see if using fewer points increased ΔB_n. In this case, B_n(1601)-B_n(801) produced a larger difference (~1.0%) than B_n(3201)-B_n(1601) did, however, it still met our criterion of $\Delta B_n \leq 1.0\%$ and so 801 data points could have been used. We decided to stay with 1601 data points because it gave better accuracy and cost very little extra run time in the program.

It should be noted here that in terms of the B_n obtained by this method, a 1% error in B_n represents a worst case error of <0.05 dB. This value, 0.05 dB, is well within the accuracy level of the rest of the components and test equipment used in the test bed in the DAR lab.

Performing a Simpson's rule integration on the raw data also provided a check on the values obtained for B_n obtained by the least squares curve fit approach. This ensured us that no gross error was made in obtaining the final values of B_n.

A final check on the B_n values was to perform the rather crude but reliable technique of graphically computing the area under the filter's response curve on the spectrum analyzer. This technique is fully described in a Hewlett Packard application note AN-150-4, (April 1974) and the description will not be repeated here. In each case where this technique was used as a check, the results were within 10% to 15% of those obtained by the method of least squares.

To ensure that an adequate order was used for the approximating polynomial, the B_n calculation was performed for each filter using several different values of N, the highest order of polynomial used, along with a criterion that $\Delta B_n \leq 0.01\%$ is achieved before the iteration in N is terminated. In this case

ΔB_n is defined as $B_n(Ncur.)-B_n(Ncur.-2)$ where Ncur. is the current order of the polynomial and only an even polynomial order was used in the procedure. In practice it was found that N=28 easily met this criterion. The difference between N=28 and N=30 was never more than 0.007% for any filter listed in Table 1. Using N>30 produced no significant improvement or change in the B_n values obtained for any filter beyond one part in 10^4.

Appendix: The Method of Least Squares

Let us consider a linear space of real functions f, g, h, ... which are defined on a set of points, x, (which can be continuous or discrete) on a closed interval [a,b] on the real axis. A scalar product can be defined in this space if, to any pair of elements defined in this space, there corresponds a real number, designated (f,g), which is called the scalar product of f and g. The scalar product must also satisfy the following properties-

(a) $(f,g)=(g,f)$ (A1a)

(b) $(f,f)\geq 0;\ (f,f)=0$ iff $f\equiv 0$ (A1b)

(c) $(\alpha f,g)=\alpha(f,g)$; were α is real (A1c)

(d) $(f_1+f_2,g)=(f_1,g)+(f_2,g)$ (A1d)

For continuous functions on the interval [a,b] the scalar product may be defined as

$$(f,g)=(1/(b-a))\int_a^b f(x)g(x)dx \tag{A2}$$

While for functions defined on a finite set of discrete points on the interval [a,b], the scalar product my be defined as

$$(f,g)=(1/(N+1))\sum_{n=0}^{N} f(x_n)g(x_n) \tag{A3}$$

In this case, these functions are defined to exist on a N+1 dimensional linear space, and the function f may be considered as a (N+1) dimensional vector $f=(f_0, f_1, f_2, \ldots f_N)$ on this space. Thus the scalar product as defined on equation (A3) may be considered as a projection of the vector f onto a basis function, or along a principal co-ordinate, of the space.

The norm of a scalar product may be defined as

$$|f|=(f,f)^{1/2} \tag{A4}$$

and a distance in the linear space as follows

$$d(f,g)=|f-g|=(f-g,f-g)^{1/2} \tag{A5}$$

Using the above terminology and definitions, we may now define a function consisting of elements of a linear space as follows

$$\Phi_n(x)=a_0\phi_0(x)+a_1\phi_1(x)+ \ldots +a_N\phi_N(x)=\sum_{n=0}^{N} a_n\phi_n(x) \tag{A6}$$

where the a_n are real numerical coefficients, as a generalized polynomial with respect to the system of functions $\phi_n(x)$. If an arbitrary function f(x) belongs to a linear space E on which the $\phi_n(x)$ are defined, the problem of choosing the coefficients a_n in (A6) such that the mean square error is minimized can be stated as a problem in which the distance $d(f,\Phi_n)$ is minimized. Using the concept of distance in the linear space as given in equation (A4), we can define $d(f,\Phi_n)$ as

$$d(f,\Phi_n)=|f-\Phi_n|=(f-\Phi_n, f-\Phi_n)^{1/2} \tag{A7}$$

The polynomial that satisfies the property of minimizing $d(f,\Phi_n)$ is the polynomial which minimizes the mean square error between f(x) and $\Phi_n(x)$ over the interval [a,b]. In this way, the generalized polynomial defined in (A6) represents the "best fit" to the data points.

To see that an appropriate choice of the coefficients a_j can minimize the distance function $d(f,\Phi_n)$ in equation (A7) we may proceed as follows. Inserting the generalized polynomial defined in equation (A6) into (A7) and squaring the resulting expression we can write

$$d^2(f,\Phi_n)=|f-\Phi_n|=(f-\Phi_n,f-\Phi_n)$$

$$=(f,f)+\sum_{j,k=0}^{N} a_j a_k(\phi_j,\phi_k)-2\sum_{j=0}^{N} a_j\ (f,\phi_j) \qquad \text{(A8)}$$

The quantity $d^2(f,\Phi_n)$ in equation (A8) is a quadratic form relative to the coefficients a_j, and consequently for any a_j, $d^2(f,\Phi_n)\geq 0$. The quantity $d^2(f,\Phi_n)$ will reach its non-negative minimum when the distance, $d(f,\Phi_n)=(d^2(f,\Phi_n))^{1/2}$, reaches its minimum.

To obtain the coefficients a_j for which this minimum occurs, we can take the partial derivatives of equation (A8) with respect to the a_j's and set them equal to zero. This will produce the following set of linear equations

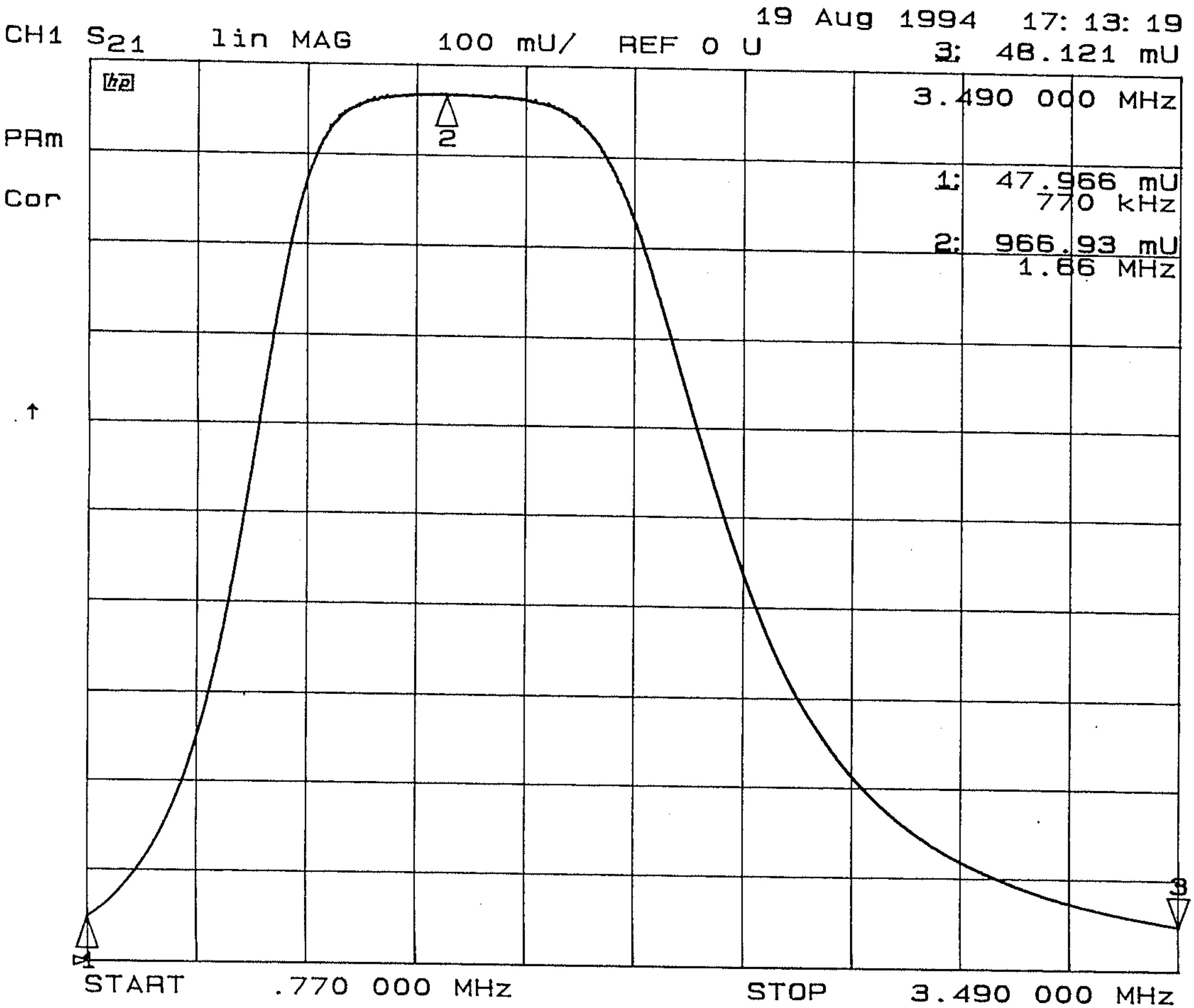
$$\begin{aligned}
a_0(\phi_0,\phi_0) + a_1(\phi_1,\phi_0) + a_2(\phi_2,\phi_0) + \ldots + a_n(\phi_n,\phi_0) &= (f,\phi_0)\\
a_0(\phi_0,\phi_1) + a_1(\phi_1,\phi_1) + a_2(\phi_2,\phi_1) + \ldots + a_n(\phi_n,\phi_0) &= (f,\phi_1)\\
\ldots \qquad\qquad \ldots & \qquad\qquad \text{(A9)}\\
\ldots \qquad\qquad \ldots & \\
a_0(\phi_0,\phi_n) + a_1(\phi_1,\phi_n) + a_2(\phi_2,\phi_n) + \ldots + a_n(\phi_n,\phi_n) &= (f,\phi_n)
\end{aligned}$$

The scalar products (ϕ_i,ϕ_j) in equation (A9) form an (N by N) array. Since the functions ϕ_n are linearly independent, this array will have a non-vanishing determinant and consequently it may be inverted. Gauss elimination is used to invert the array, and so equation (A9) can be solved for the a_n which will provide a minimum value for the $d(f,\Phi_n)$. This is the technique used to find the proper coefficients a_i in equation (5). The functions $\phi_n(x)$ are chosen to be polynomial functions, i.e. $\phi_n(x)=x^n$, for the reasons given in the main body of this paper. Note that with 1601 data points and N=30, the system of equations described in equation (4) is over specified, and that the method of least squares described above reduces this system of equations to one of N equations with N unknowns.

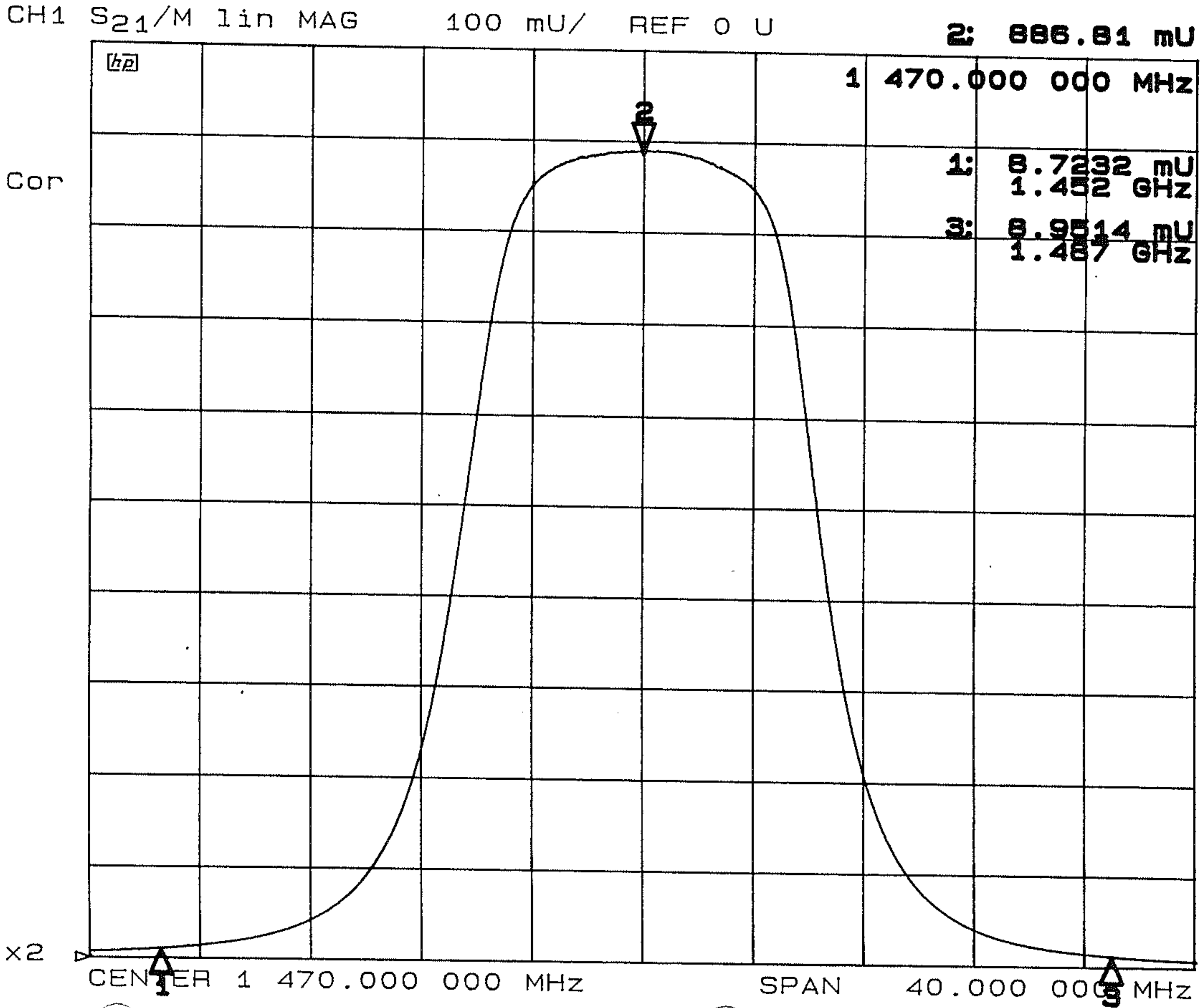
The program to evaluate the noise equivalent bandwidth of the filters was written in HP BASIC. Data on each filter was obtained from the HP 8753D using the linear scale. The program found the maximum amplitude of the data , normalized the data to this value, then squared it to create $|H(f_i)/H_0|^2$. The program then evaluated the appropriate scalar products as defined in equation (A9). Gauss elimination was used to invert the matrix and thus solve for the unknown a_n's. The corresponding value for B_n was obtained by integration of the resulting polynomial. A Simpson's rule integration was also performed on the raw data and the results displayed for B_n obtained in this fashion as a check on the least squares technique. The user may choose the order of the polynomial function used in the curve fit, as well as the number of data points to be used in the analysis.

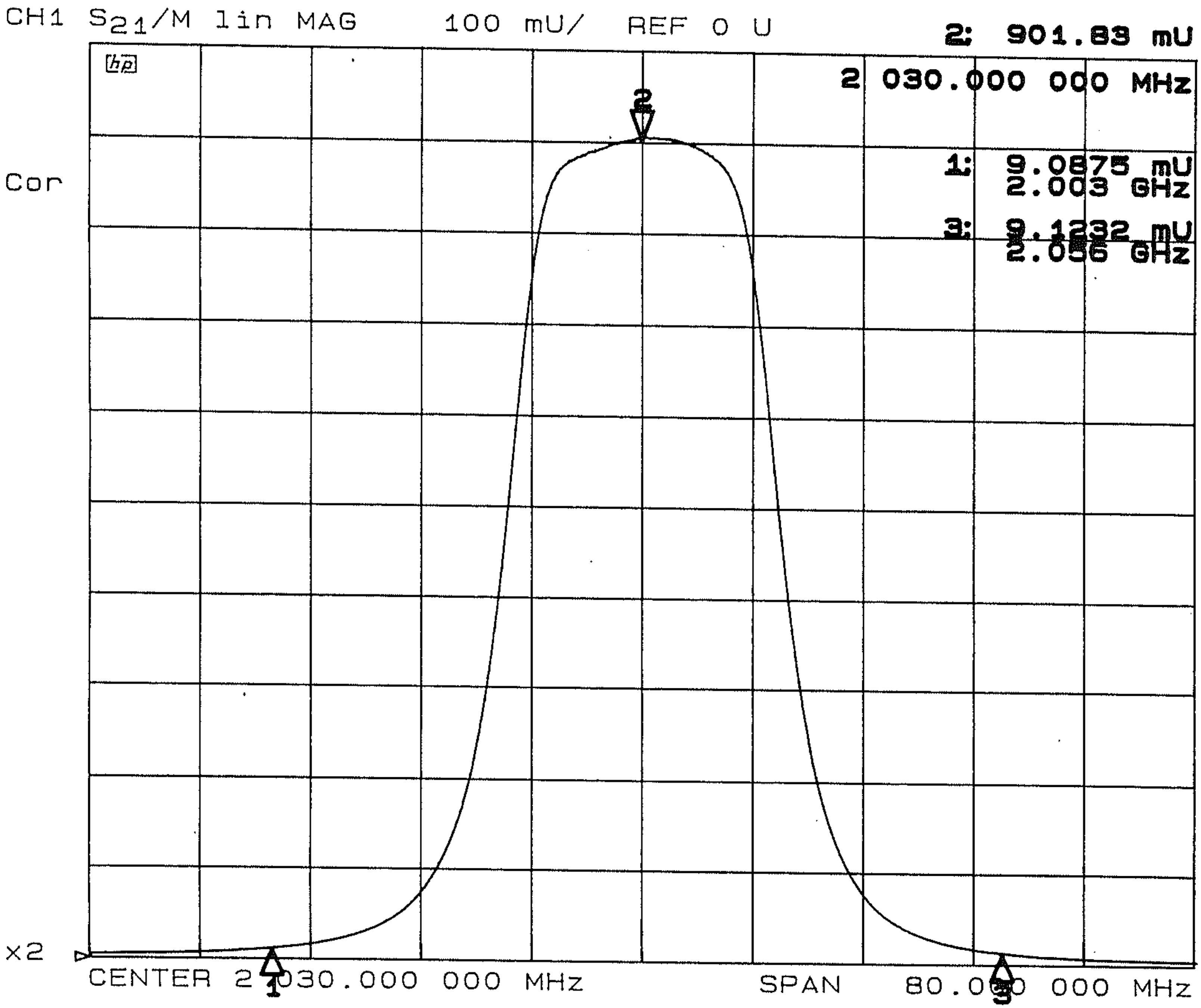
TTE BPF AM

Data over this band used to calc. BWN



8/8/94
S.S.T.
TTE L-band BPF
K2821-1470m-4m
-50/50
100 pts



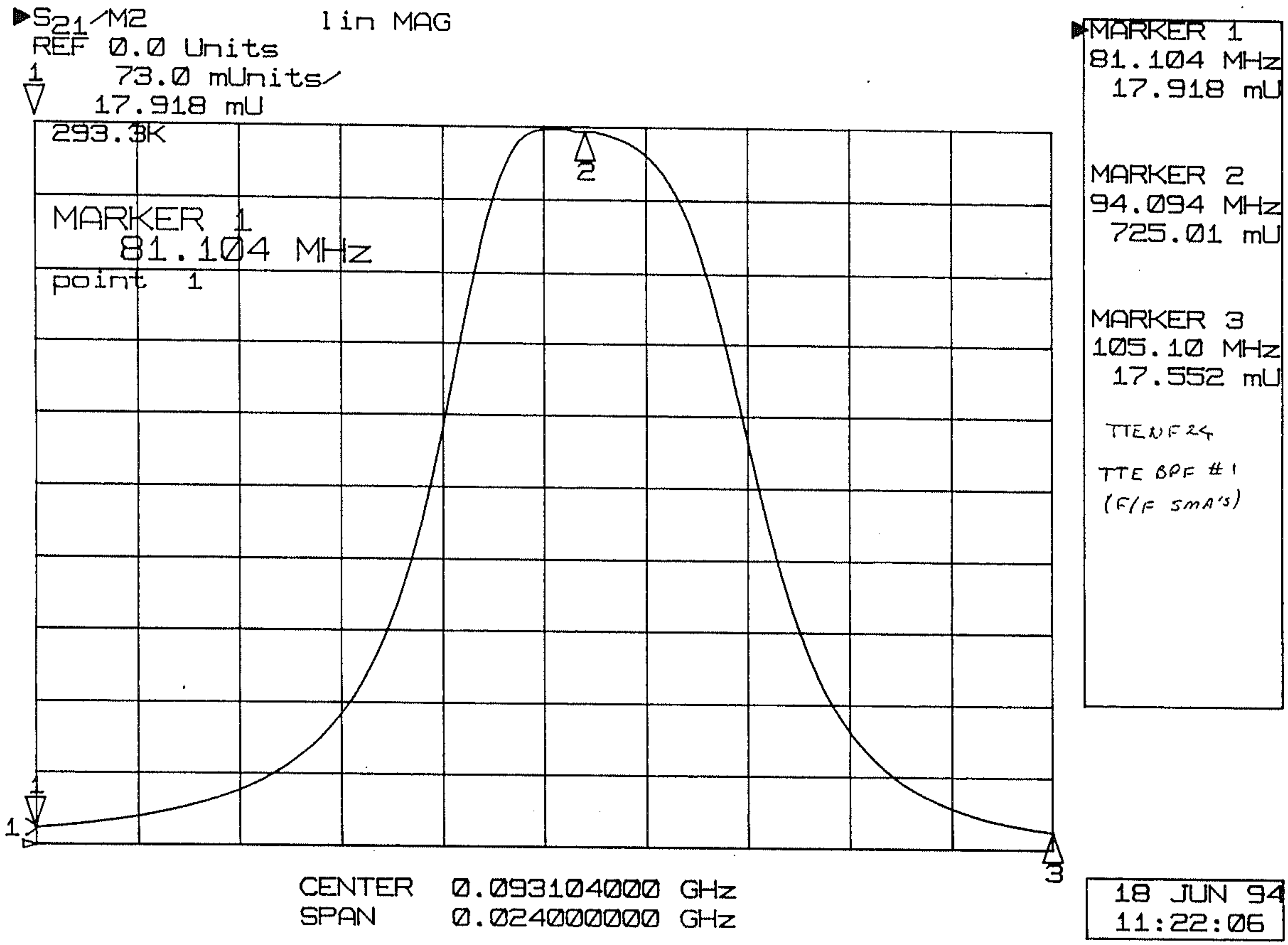


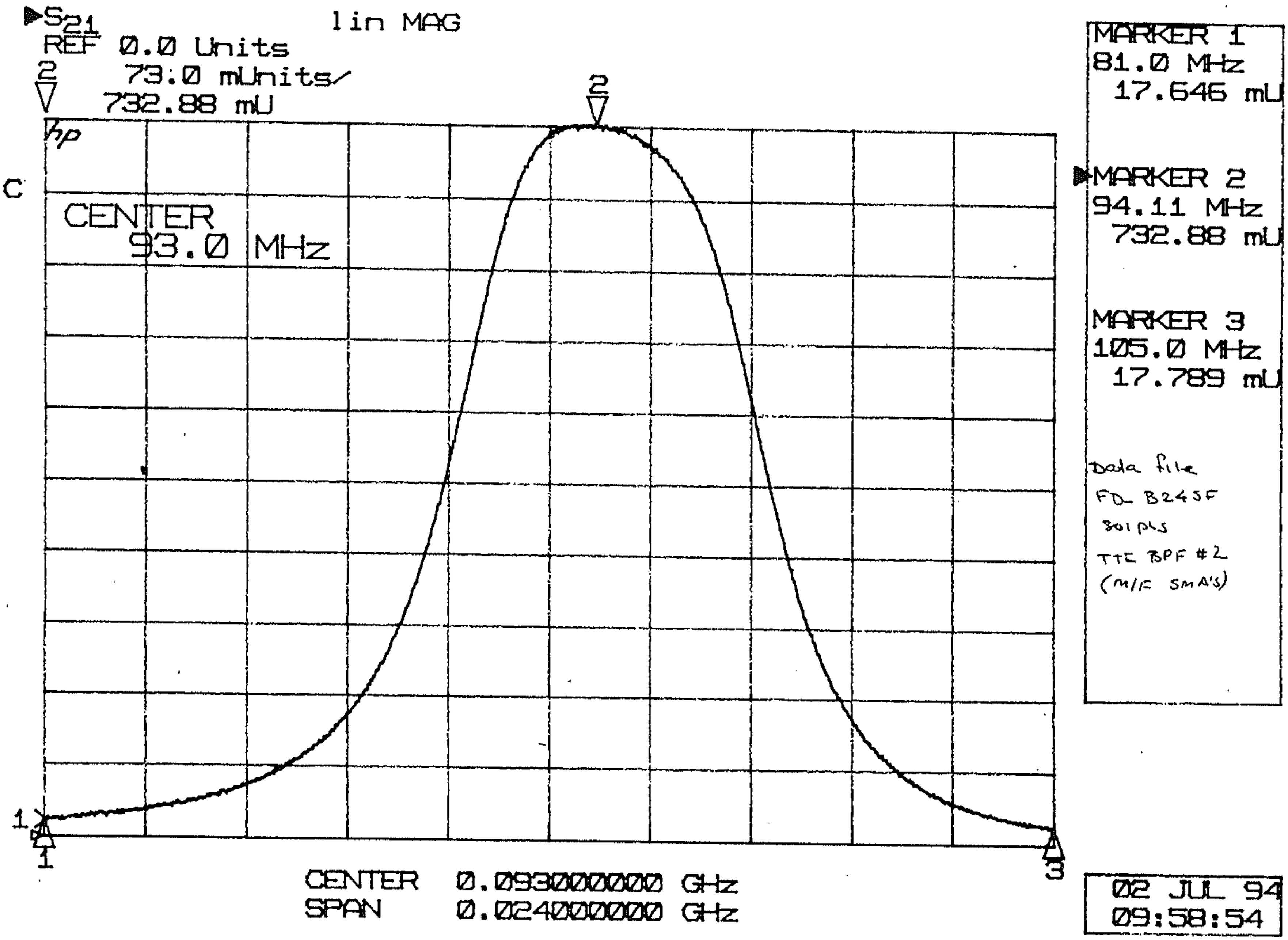
8/8/94
S.S.T.

TTE S-band BPF
K2830-2050m-4m-50/50

1601 pts.

Thru cal only, not a full two port.


S21/M2 lin MAG
REF 0.0 Units
73.0 mUnits/
17.918 mU
293.3K
MARKER 1
81.104 MHz
point 1
MARKER 1
81.104 MHz
17.918 mU
MARKER 2
94.094 MHz
725.01 mU
MARKER 3
105.10 MHz
17.552 mU
TTENF24
TTE BPF #1
(F/F SMA's)
CENTER 0.093104000 GHz
SPAN 0.024000000 GHz
18 JUN 94
11:22:06


S21
lin MAG
REF 0.0 Units
73.0 mUnits/
732.88 mU
hp
C
CENTER
93.0 MHz
MARKER 1
81.0 MHz
17.646 mU
MARKER 2
94.11 MHz
732.88 mU
MARKER 3
105.0 MHz
17.789 mU
Data file
FD_B245F
801 pts
TTE BPF #2
(M/F SMA's)
CENTER 0.093000000 GHz
SPAN 0.024000000 GHz
02 JUL 94
09:58:54

APPENDIX T

Selection of Critical Material

FROM: Emil Torick

TO: Ralph Justus (EIA)

DATE: June 30, 1994

RE: Report of the Selection Panel for DAR Subjective Test Material

The Selection Panel, consisting of Emil Torick, Jody Daub and Douglas McKinnie was convened at the Communications Research Centre (CRC) in Ottawa during the period May 25 - June 17, 1994. Torick and Daub are consultants to the EIA, and McKinnie is from McGill University on special assignment at the CRC. All are skilled listeners experienced in assessing the subjective performance of digital audio codecs.

The goal of the Panel, as communicated by the EIA DAR Subcommittee, was to choose nine selections of critical subjective test material for the evaluation of the proponent DAR systems. The guidelines followed for the selection process were those of ITU-R[1] which state in part:

> *"Critical material is that which stresses the systems under test. . . .Unless truly critical material is found for each system [to be tested], experiments will fail to reveal differences among systems and will be inconclusive. . . .Synthetic signals deliberately designed to break a specific system should not be included. The artistic or intellectual content of a programme sequence should be neither so attractive nor so disagreeable or wearisome that the subject is distracted from focusing on the detection of impairments."*

The Panel evaluated more than 70 selections of potential test material. These had been contributed, at the invitation of the DAR Subcommittee, by AT&T, CRC, R. Culver, A. Laird, Roland USA and the University of Miami. Additionally, a number of selections from the EBU subjective test disc (SQAM) developed specifically for the assessment of audio systems also were considered. All contributed material was from CD or DAT sources. At the EIA DAR Test Laboratory at the NASA-Lewis Research Center in Cleveland, each selection was re-recorded directly and through each proponent DAR system and then delivered to the CRC on DAT. At the CRC these recordings were transferred to hard disc files for playback in the CRC audio subjective test facility.[2] Neither the

---

[1] ITU-R (formerly CCIR) Recommendations 500, 562, 710 and 811 previously have established methods for assessing subjective quality of audio and video systems. The guidelines quoted here are from a TG 10/3 Draft New Recommendation: "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," dated 1 November 1993.

[2] For a description of the CRC facility including the computer-operated presentation equipment see T. Grusec and L. Thibault (CRC) and G. Dimino (RAI), "CCIR Listening Tests: Network Verification Tests Without Commentary Codecs," a document submitted to CCIR Task Group 10/2 (doc. 43), Geneva, Switzerland, 27 October 1993.

panelists nor CRC staff personnel knew the proponent system identity of any recording; systems were labelled only by letter designations A through I.

The Panel operated on the basis of consensus, observing the above ITU-R guidelines and those of the DAR Subcommittee which additionally specify that each proponent system be stressed critically by at least two test selections.[3] Further balance was achieved by selecting test materials which caused a broad range of artifacts to appear. During initial evaluations by the Panel more than half of the selections contributed were deemed "non-critical" and removed from further consideration. With the test material finally selected, most systems displayed only rather subtle degradation of the source material. However, the artifacts heard by the panelists included pre-echo, background noise, distortion, high-frequency attenuation, transient attenuation, chirps etc. and alteration of the stereo image location and depth. The final choices, which range in duration from 10 to 30 seconds each, are as follows:

TITLE (Artist)	SOURCE	CONTRIBUTOR
"Harpsichord"	SQAM, Track 40, Index 1	EBU
"Glockenspiel"	SQAM, Track 35, Index 1	EBU
"Ride Across the River" (Dire Straits)	Warner Bros. 7599-25264-2, Track 6	CRC
"Music and Rain"	Original production	AT&T
"Vega and Glass"	Original production	AT&T
"Pictures at an Exhibition"	Live recording of a muted trumpet	U. Miami
"Wet Environment"	"Roland Sound Space" Demonstration CD	Roland USA
"Daughter" (Pearl Jam)	Epic recording processed through: Aphex "Compellor" Model 300, Dolby "Spectral Processor" Model 740, Aphex "Dominator II" Model 720	A. Laird
"Bass Clarinet"	SQAM, Track 17, Index 1, processed through broadcast-equipment chain as described above.	A. Laird

[3] See "Laboratory RF Transmission Test Procedure," Revision 9, November, 1993, Test K1.

SUBJECTIVE TEST SELECTIONS

TITLE (Artist)	SOURCE	CONTRIBUTOR
"Harpsichord"	SQAM, Track 40, Index 1	EBU
"Glockenspiel"	SQAM, Track 35, Index 1	EBU
"Ride Across the River" (Dire Straits)	Warner Bros. 7599-25264-2	CRC
"Music and Rain"	Original production	AT&T
"Vega and Glass"	Original production	AT&T
"Pictures at an Exhibition"	Live recording of a muted trumpet	U. Miami
"Wet Environment"	"Roland Sound Space" Demonstration CD	Roland USA
"Daughter" (Pearl Jam)	Processed* Epic recording	A. Laird
"Bass Clarinet"	SQAM, Track 17, Index 1, processed*	A. Laird

* Representative broadcast processing chain incorporated Aphex "Compellor" Model 300, Dolby "Spectral Processor" Model 740 and Aphex "Dominator II" Model 720.

CRITICAL TEST MATERIAL

"Critical material is that which stresses the systems under test. . . . Unless truly critical material is found for each system [to be tested], experiments will fail to reveal differences among systems and will be inconclusive. . . .Synthetic signals deliberately designed to break a specific system should not be included. The artistic or intellectual content of a programme sequence should be neither so attractive nor so disagreeable or wearisome that the subject is distracted from focusing on the detection of impairments."

From ITU TG 10/13 Draft New Recommendation, dated 1 November 1993:
"Methods for the Subjective Assessment of Small Impairments in Audio Systems"

THE SELECTION PROCESS

Contributions of potentially critical test material received:	70
DAR systems to be evaluated (identified only as System A - System I):	9
Total number of recordings available for audition:	70 x 9 = 630
Target goal of the number of critical selections for subjective tests:	9

Artifacts observed: pre-echo, background noise, distortion, high-frequency attenuation, transient attenuation, chirps, etc. and alteration of stereo image, location and depth.

A further goal achieved: Each proponent system should be stressed critically by at least two test selections.

APPENDIX U

Quality Test Report

Communications
Research Centre
Centre de recherches
sur les communications

EIA/NRSC DAR Systems

Subjective Tests of Audio Quality and Transmission Impairments

Final Report

21 July 1995

TED GRUSEC, GILBERT SOULODRE and LOUIS THIBAULT
Signal Processing and Psychoacoustics
Radio Broadcast Technologies Research
Communications Research Centre
3701 Carling Avenue
Ottawa, Ontario, Canada K2H 8S2

References

[1] *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems,* ITU-R Recommendation BS.1116, Radiocommunications Study Group, International Telecommunication Union, Geneva, Switzerland

[2] Thibault L. and Grusec T., "*EIA-DAR Listening Tests - Quality and Impairment Tests Procedures:*, Signal Processing and Psychoacoustics, Radio Broadcast Technologies Research, CRC, Ottawa, Ont., Document submitted to EIA Working Group B, Updated version, 1 December 1993

[3] Londa, Dave, et al. "*EIA-DAR/NRSC DAB objective tests of transmission impairments*".

[4] Beaton R.J. and Wong P., *A Disk-based System for the Subjective Assessment of High Quality Audio*, Preprint 3497, AES 94th Convention, Berlin, Germany, 16-19 March 1993

Table of Contents

List of Figures

List of Tables

1. Introduction

This final report describes the procedures as well as the results of three series of subjective tests conducted at the Communications Research Centre (CRC), Ottawa, Ontario, Canada from June 1994 to March 1995. These tests were performed to assess the audio quality of Digital Audio Radio (DAR) systems submitted to the DAR subcommittee of the Electronics Industries Association (EIA) and the DAB subcommittee of the National Radio System Committee (NRSC).

A total of nine DAR systems were submitted for testing and these are labeled as *a* to *i* in this report. In the first series of tests, the subjective audio quality of the DAR systems was assessed in the absence of any transmission error. Essentially, this test evaluates the quality of the audio source coding component of each DAR system. In the second series of tests, the subjective audio quality of the DAR systems was assessed in the presence of transmission errors. Both the onset of detectability of transmission errors (namely the threshold of audibility) as well as the failure characteristic of the DAR systems (from threshold of audibility down to point of failure) were determined for the following seven types of impairment:

1. Gaussian noise
2. Co-channel interference
3. Urban slow multipath
4. Urban fast multipath
5. Rural fast multipath
6. Obstructed fast multipath
7. DAR lower first adjacent to DAR channel interference

All nine systems were tested in the presence of additive white gaussian noise and co-channel interference. A subset of the systems were tested for the other impairments. In the third and last series of tests, two modified DAR systems were retested. The audio quality (in absence of transmission error) of one of these systems was subjectively retested while the threshold of audibility and the point of failure of both DAR systems were re-evaluated for the gaussian noise and co-channel interference.

The procedures and results of the subjective tests of audio quality in the absence of transmission errors are described in chapter 2 of this report while those in the presence of transmission impairments are contained in chapter 3. Chapter 4 contains a description of the procedures and results of the subjective retests of audio quality while those of the impairment retests are contained in chapter 5. In chapter 6, the facilities used for the tests are described. Finally, acknowledgments are given in chapter 7.

2. Subjective Tests of Audio Quality

2.1 Selection of critical materials

Various organizations submitted potentially critical audio test materials from a wide variety of sources. Processing of these materials through the 9 DAR systems to be assessed for subjective quality took place at the NASA Lewis Research Center (LeRC) in Cleveland, USA. A panel of three expert listeners convened at the CRC for the selection of final test materials from the initial pool received from Cleveland. Two of these experts were US citizens and one was a Canadian.

The task of the panel was to find at least 2 stressful materials for each system. The panel uncovered a total of nine materials that met this requirement. The experimental results suggest that the panel's choices were excellent ones. Some evidence that supports this is presented in section 2.2.1.

One of the 9 systems was tested with two different comparison references because the sampling rate for that system was lower than for the other 8 systems. Accordingly, our report refers to "10" systems rather than 9. The identity of the systems was unknown to the CRC and, at the time of writing, still remains unrevealed. In the present report, the "10" systems are designated "*a*" to "*j*".

2.2 Test procedures

2.2.1 Listener expertise

A total of 21 listeners went through the test process for two days to complete the 90 rating trials (10 systems x 9 materials). Thirteen of these subjects were from the US. The other listeners included one from the UK, one from France and 6 from Canada.

The equipment, listening environment and procedures were the standard ones used in subjective tests at the CRC as described in ITU-R Rec. BS.1116 [1] and in the December 1st, 1993 revision of the *"Quality and impairment tests procedures"* document submitted to the EIA-DAR Working Group B [2]. As in all similar experiments at the CRC, we applied a rigorous statistical criterion to the data to assess each individual's listening expertise.

This criterion makes use of the fact that on each of the trials in a rating experiment (90 trials in the present case), the listener supplies two ratings; one for what the subject concluded was the "hidden reference", and one for the presumed system version of the audio material heard on that trial. (section 2.2.2 below provides more details about the subject's task and the rating procedure.) Over all the 90 trials, then, we have two distributions of grades for each subject - one distribution for the true hidden references; the other for the DAR system versions. If the subject was discriminating between these two classes of events in an objectively correct way, then these two distributions will have averages which are reliably different from each other. On the other hand, if the subject's discriminations between system and hidden reference versions was faulty, then the grades in the two distributions will resemble a "chance guessing" pattern, and the two distributions will be statistically identical.

Evaluation of these two possibilities is by way of a *t*-test, a standard statistical test applicable to comparing such distributions, as outlined in most textbooks of statistical analysis. For 90 data points, as in the present experiment, the magnitude of t must equal or exceed a value of approximately 2.00 in order that one may conclude ($p < 0.05$) that the subject was truly discriminating beyond chance between hidden reference and system versions. If the value of t for a listener is smaller than 2.00, one must conclude that the subject was, overall, guessing.

As will become apparent in the presentation of results later in this report, there were two DAR systems in the experiment where detecting degradations appeared to be obviously "easy". Listeners gave those systems (*i* and *j*, Fig. 2.2a) consistently low grades and all subjects made very few errors in correctly discriminating the system from the hidden reference versions. Including these easy trials in the calculation of the *t*-scores would artificially inflate the merit scores of each subject. To avoid this, the *t*-scores which we present in the following table conservatively omit those 18 very easy trials and are calculated on only the other 72 trials. The scores for all subjects in the quality experiment are shown below in descending order of merit.

Subject	t-score
1	11.62
2	8.51
3	7.35
4	6.50
5	6.34
6	7.56
7	6.49
8	6.38
9	6.19
10	5.68
11	5.61
12	5.16
13	5.00
14	4.57
15	4.30
16	3.48
17	3.35
18	2.84
19	2.57
20	2.46
21	2.43

Table 2.1 Listeners' *t*-scores

As is clear in the above Table, no listener who took part in the experiment scored below 2.00. Therefore, they all showed that they were able to discriminate correctly between hidden reference and system versions across all the trials in the experiment. No-one's data had to be rejected for insufficient listener merit. And so, the data of all the 21 participants are included in the analyses. As mentioned, this *t*-score was very conservative since the easiest two systems, *i* and *j*, were omitted in the calculation.

We must emphasize that these are high expertise scores. This fine listener performance was due, in addition to other things, to very well-chosen materials which were stressful to the DAR systems under test, while, at the same time, relatively easy for listeners to grade. This reveals that the three experts who chose the materials performed their task extremely well.

We should point out that listeners can all be, as in the present case, sufficiently expert, yet disagree with each other in the relative ratings they assign. If this were the case, then the results would be inconsistent, hence statistically unreliable. As the results presented below will show, the opposite was found. In other words, the listeners, all sufficiently expert, were highly consistent with each other. Hence the experiment provided clear, interpretable outcomes as will be seen in sections 2.3 and 2.4.

2.2.2 Grading scale

The test procedures used complied with those described in ITU-R Recommendation BS.1116 [1]. The 90 trials of the experiment were rated in two consecutive days of 45 trials each by the listeners. The procedure included a training phase followed by a blind rating phase. The morning of each day was devoted to the training session using those audio materials to be rated in the subsequent blind rating. Subjects worked together in groups of 2 or 3 and were allowed as much time as they required for training. Subjects trained using the same hard disk playback system used in the blind rating tests.

Blind rating took place in the afternoon of each day and was performed individually by each listener. The subjects alternated with each other in the listening room, so while one was working, the other(s) rested. Listeners used the disk-based playback system described in section 5 of this report which allowed seamless switching between the stimuli to be compared. Listeners were able to take as much time as they needed on each trial, switching as often as they liked, until satisfied with the numerical ratings they were asked to assign. They were also free to use either the loudspeakers or headphones to make a judgment, whichever they felt was the most critical transducer on any given trial.

The actual scale used by the subjects is shown in Fig. 2.1. It is a 5 grade rating scale (1.0 to 5.0) where listeners were instructed to use a single decimal point. In effect, this is a 41 point scale. The subjects were instructed to treat this as a continuous scale but, to facilitate the subjects' orientation, category labels were associated with the scale. Thus, 1.0 to 1.9 is a "very annoying" range; 2.0 to 2.9 is "annoying"; 3.0 to 3.9 is "slightly annoying"; 4.0 to 4.9 is "perceptible but not annoying". Finally, 5.0 is "imperceptible".

The listener's task on a trial is to compare each of two alternative versions, labeled as "B" and "C", of an audio material with a <u>known</u> Reference version, labeled "A", of the same material. The subject knows that one of the alternatives ("B" or "C") is a "hidden reference", identical to

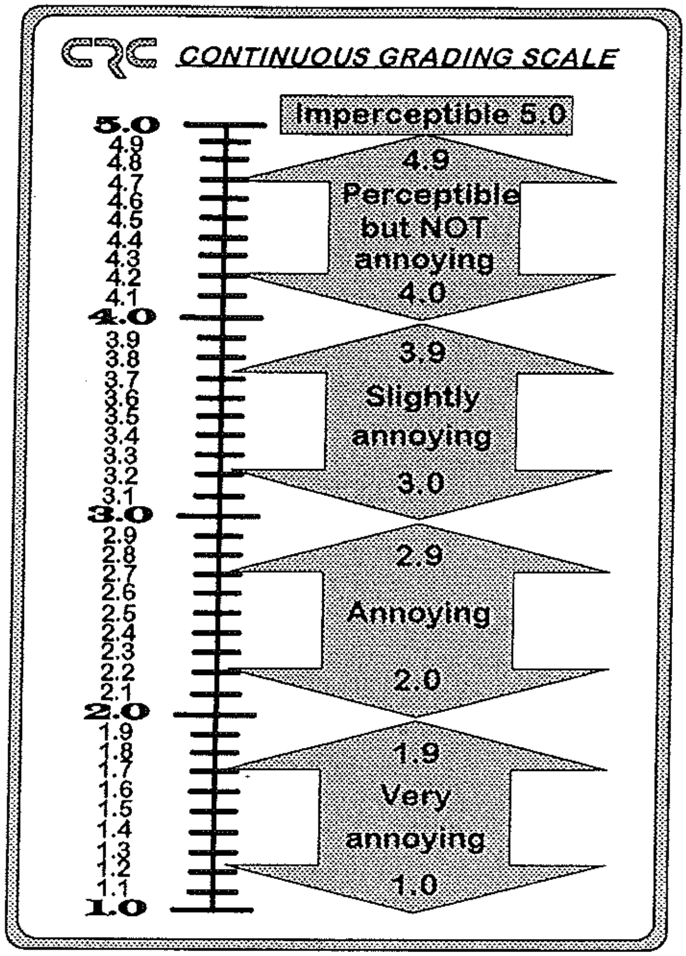


Fig. 2.1 ITU-R continuous 5-grade impairment scale

the Reference, and that the other alternative is one that has been processed through a DAR system. The subject does not know which is which, but must decide this through listening. He or she then assigns a grade to both "B" and "C" alternatives, as compared to the known Reference "A", using the 1.0 to 5.0 scale. A constraint is that the alternative the subject has decided is the "hidden reference" must be graded 5.0. And so, *at least one* of the two grades on each trial must be a 5.0.

Thus two totally interdependent scores from the listener are recorded on each trial. This deliberate interdependence is handled by subtracting the score given to the true hidden reference from the score given the true processed version (i.e., DAR System minus Reference). The reason for subtracting in this direction rather than the opposite one (Reference minus DAR System) is only so that in a graphical plot of outcomes, the data will fall in the same geometric quadrant as they would if the actual 1.0 to 5.0 scores used by the subjects were plotted.

This subtraction means, however, that the scores are transformed so that the 1.0 to 5.0 range of the original scale becomes, instead, -4.0 to 0.0 in the analysis and presentation of results. These difference grades or "diffgrades" represent the relative differences between the grades given to the hidden reference and the ones given to the DAR system under test. For practical purposes, it can be assumed that a direct analogy holds between the original 1.0 to 5.0 scale and the -4.0 to 0.0 diffgrades in the report. Accordingly, to facilitate interpretation for the reader who is familiar with the 1.0 to 5.0 scale used by subjects, the "annoyance" category labels are shown between the Y-axis numbers in Fig. 2.2a and 2.2b. Thus, in these figures, the words "perceptible but not annoying" are shown between 0.0 and -1.0, "slightly annoying" appears between -1.0 and -2.0, and so on.

The 90 materials to be rated were divided into 6 sessions of 15 trials each. Within each session, items were ordered in a quasi-random, unpredictable way. Moreover, the order of presentation of the sessions was changed from listener to listener to ensure that time-correlated factors (such as fatigue) would not differentially affect any level of any of the factors under test.

2.3 Tests results

2.3.1 Graphical and tabular presentations

For visual clarity, the average quality diffgrades of the experiment are divided between Fig. 2.2a and 2.2b rather than being shown within a single graph. The two highest ranking systems overall are shown in Fig. 2.2a along with the 7th through 10th ranking systems. The 3rd to 6th ones are found in Fig. 2.2b. The ranking referred to here is an ordering by the overall average diffgrade (average across all listeners and all audio materials for each system). These overall averages are plotted in the "Overall Averages" column at the right-hand side of these Figures. Also shown are the rank order of the systems in the legends in these figures.

Table 2.4a shows the overall average diffgrade for each system in the right-hand column. That Table, as well as Table 2.4b on the same page, shows all the numbers that are plotted in Fig. 2.2a and 2.2b. In Table 2.4a, the average diffgrades across all listeners for each audio material occupy a separate row for each DAR system. In Table 2.4b, the average diffgrades in each column are ordered by the magnitude of those diffgrades for each audio material. The system represented in each cell is clearly indicated in that table.

The ordering chosen for the audio materials along the X-axis in Fig. 2.2a and 2.2b reflects the average diffgrades given to each material across all systems. The averages for audio materials across systems are shown in the two bottom rows of Table 2.4a. It is the ones "without *i* and *j*" that were used to determine the order used in Fig. 2.2a and 2.2b since gradings for those two systems were erratic. This ordering means that the most transparent materials (those receiving the highest grades) are shown at the left and, as we move toward the right, the materials are less and less transparent, i.e., more and more critical for revealing system differences.

The five-letter abbreviations used in the text and the sources for the audio materials are shown in the table below:

Code	Description	Duration	Source
Dires	Dire Straits cut	30 s	Warner Bros. CD 7599-25264-2 (track 6)
Prljm	Pearl Jam cut	30 s	Sony/Epic CD ZK53136 (track 3) with processing[1]
Water	Sounds of water	30 s	Roland Dimensional Space Processor Demo. CD
Glock	Glockenspiel	16 s	EBU SQAM CD (track 35/Index 1)
Bascl	Bass Clarinet arpeggio	30 s	EBU SQAM CD (track 17/Index 1) with processing[1]
Mrain	Music and rain	11 s	AT&T mix
Vegla	Susan Vega with glass	11 s	AT&T mix
Trmpt	Muted trumpet	9 s	Original DAT recording, University of Miami
Hpscd	Harpsichord arpeggio	12 s	EBU SQAM CD (track 40/Index 1)

[1] Processing chain used: Aphex Compellor Model 300 (set for leveling only)
Dolby Spectral Processor Model 740
Aphex Dominator II Model 720

Table 2.2 List of audio test materials used in the quality tests

2.3.2 Overall systems comparison

The first criterion by which the DAR systems under test can be compared and ranked is by their overall average diffgrade. Statistical analysis (Analysis of Variance, or ANOVA) reveals that the overall experimental differences among systems have a very fine resolution of 0.17 of a grade in the transformed diffgrade scale. In other words, any two systems that are numerically different by 0.17 or more in overall average diffgrades are reliably different ($p < 0.05$) from each other. If they differ by less than 0.17, this difference is not considered statistically significant since it could too easily be due to chance (i.e., with $p > 0.05$, the generally accepted cut-off).

By this rigorous criterion, systems *a* and *h* (Fig. 2.2a, Table 2.4a) are not reliably different from each other overall, although system *a* is rated 0.10 of a grade higher than *h*. System *a* is rated significantly higher than the group consisting of systems *g*, *f*, *c* and *e* (Fig. 2.2b); however, system *h* does not differ reliably from these four systems. These four (Fig. 2.2b) form a cohesive group with little to differentiate among them in overall grades. There is only a 0.05 difference in overall average diffgrade between the highest of the four (*g*) and the lowest (*e*).

Another cohesive group of systems consists of systems *b* and *d* which differ from each other by only 0.09. These two (Fig. 2.2a) are reliably different from the four systems (*g*, *f*, *c* and *e*) of Fig 2.2b and from both *a* and *h*. Finally, systems *i* and *j* are virtually identical with only a 0.01 difference magnitude between their overall average diffgrade. By this measure, the different audio sampling rates (32 kHz for system *i* and 48 kHz for system *j*) of the two references that were used with this system made no significant difference (statistically speaking) in its overall diffgrades; nor, as will be seen in more detail in the next section, in its pattern among the 9 audio test materials. Overall, these "two" (*i* and *j*) rank lower by more than a full grade, in the 5-grade scale, from the systems that are nearest to them; thus they comprise a significantly different "group" from all the others.

To summarize, there are four distinct groups of systems, reliably different from each other on statistical grounds. Only the grouping of *h* is questionable, since it is not reliably different from either *a* or from the systems in the second ranking group :

Highest	1.	*a*, (and possibly *h*)
	2.	*g*, *f*, *c* and *e* (and possibly *h*)
	3.	*b* and *d*
Lowest	4.	*i* and *j*

It should be noted that, in overall average diffgrades, all the systems, except for the low ranking *i* and *j*, fall within the "perceptible but not annoying" category at the top of the grading scale. In fact, both *a* and *h* are in the very top half of that range, with overall diffgrades higher than -0.5 (equivalent, as explained before, to 4.5 in the 1.0 to 5.0 scale actually used by the subjects). The second ranking systems (*g*, *f*, *c*, *e*) are just slightly below -0.5; and the third group (*b*, *d*) is lower than that, but is still above -1.0 in the "perceptible but not annoying" range. Systems *i* and *j* stand apart quite clearly, falling into the "annoying" range, far below all the other systems.

2.3.3 Systems comparison by audio material

When comparing the audio quality of DAR systems, the kinds of overall averages discussed above need qualification by examining the pattern of interactions that occur between combinations of specific audio materials on the one hand, and specific systems on the other. These interactions may sometimes affect the interpretation of the results made from the overall picture as presented above. In the present experiment, these interactions do not greatly disturb the interpretations presented in the previous section. Rather they confirm them and provide some further illumination.

First, statistical analysis (ANOVA) reveals that the resolution for the interaction of audio materials and systems in this experiment is 0.45 of a grade. This too is a very fine degree of resolution for interactions of this type. It means that when comparing any two average diffgrades obtained for any given audio material and for any given system in Fig. 2.2a and 2.2b or Table 2.4a and 2.4b, a numerical difference of 0.45 or greater is required before it can be concluded that those two diffgrades are statistically different from each other rather than being due to chance ($p < 0.05$).

In this experiment, there is little interest in comparing average diffgrades between different audio materials. However, comparing the average diffgrade of each system within a given material is of importance. In particular, it is pertinent to see for any given material, whether any particular obtained difference between systems is meaningful.

To facilitate such comparisons, a chart of 9 figures, one for each audio material, was prepared. These are shown on a single page as Fig. 2.3. It should be noted that, in order to fit all 9 figures in one page, the full range of diffgrades is not shown in Fig. 2.3, but only the range between -2.0 and 0.0. This space saving was possible because only systems *i* and *j* have any average diffgrades below -2.0, and in all those cases, *i* and *j* were very reliably different from all the other systems (see the footnote under Table 2.4b).

Fig 2.3 shows the average diffgrades per system for each of the 9 audio test materials (corresponding to the abscissa or "columns" of Fig. 2.2a and 2.2b), arranged in descending order of magnitude from left to right. Table 2.4b contains the actual data shown in Fig. 2.3. The reader can make any comparisons desired using the 0.45 critical value mentioned above for interactions between systems and audio materials. To facilitate this comparison, "error bars" are shown in Fig. 2.3 for each system average in each audio material chart. Each error bar deviates from the average for a system by + and - one half of 0.45 (the critical magnitude). Thus, one can evaluate visually whether any two systems within each audio material should be considered to be statistically different by noting whether or not there is any overlap (along the Y-axis) between the error bars for the two systems being compared. If there is overlap, those two systems do not statistically differ on that material. If there is no overlap, it indicates that the difference between the average diffgrades of those two systems on that material is statistically reliable ($p < 0.05$).

The actual numerical differences underlying the charts of Fig. 2.3 are presented in the two pages of Table 2.5. There, in a manner that parallels the charts of Fig. 2.3, differences are seen numerically rather than visually as in the Figure. Those that are statistically reliable are shown as **bold** numbers. There is redundancy in Table 2.5 in that the numbers along the right hand side above the blank diagonal are mirror images of the left hand ones below that diagonal. However, retaining this redundancy means that a reader can get all the information needed more easily. For completeness, we should also point out that if a reader is interested in evaluating overall differences among audio materials independent of systems (as are shown in the averages in the bottom row of Table 2.4a), the critical value shown by the ANOVA is 0.23. This applies to the "without *i* and *j*" averages. Thus, any two of the 9 audio material averages ("without *i* and *j*") across systems must differ by at least 0.23 before they can be considered significantly different on statistical grounds.

2.4 Discussion and conclusions

The quality assessment results show that system *a* has the highest overall rank and the most consistent ratings across the audio materials tested (Fig. 2.2a). For all audio materials, *a* never falls below the "perceptible but not annoying" range. Even on the most revealing materials (Vegla, Trmpt, Hpscd) it only dips to the lower half of the "perceptible but not annoying" range. Taking the statistical error into consideration, *a* remains within the "perceptible but not annoying" range even under the most pessimistic view (i.e., by looking at the lower bound of error bars for Vegla, Trmpt, Hpscd in Fig 2.3).

System *h* is less consistent. Though it ranks significantly higher statistically than *a* in two materials, Vegla and Dires, the opposite is true for three (Glock, Bascl and Trmpt) where *a* is rated reliably higher than *h*. System *h* looks transparent for four stimuli (Dires, Water, Mrain and Vegla), but is rated in the "slightly annoying" range on Bascl and Trmpt. System *a*, like *h*, also appears transparent for four materials (Prljm, Glock, Bascl, and Mrain), but, as discussed previously, *a* is never rated in the "slightly annoying" range. Other systems (i.e., *c*, *e* and *f*) perform significantly better than *a* on two materials, namely Dires and Trmpt. But all those three systems that are better than *a* at any point, fall below *a* on other materials.

The inconsistent behavior across audio materials of the second group (*g*, *f*, *c*, *e*) produces the lower overall ranking obtained by those systems. Their descent into the "slightly annoying" range is quite clear for Mrain and Trmpt (Fig. 2.2b).

Turning to the third group, generally lower performance in the "perceptible but not annoying" range, coupled with more frequent and more severe dips into the "slightly annoying" range, accounts for the lower overall ranking of systems *b* and *d*.

Finally, the performance of systems *i* and *j* across audio materials is entirely consistent with its low "annoying" overall rating. The average diffgrades of *i* are virtually identical to those of *j* on every material, showing that they are perceived as being really the same system. This is true even though the reference used for system *i* was sampled at 32 kHz and that for system *j* was a CD quality signal sampled at 48 kHz. The sampling rate variation of the reference was therefore found to be irrelevant. The highest rated materials for systems *i* and *j* (Prljm and Mrain) are close to -1.0. Three materials (Bascl, Trmpt and Hpscd) are rated well within the "very annoying" range of the scale. The remaining four materials (Dires, Water, Glock and Vegla) were rated in the "slightly annoying" or "annoying" ranges.

One way to summarize the audio materials by systems interaction is to see how many times each system fell below -1.0 for the 9 materials. A reasonable way to do this that takes the statistical error into account is to count as "below -1.0" any system whose lower error bar in Fig. 2.3 falls into that range. The following list shows those counts:

Count	Systems
0	*a*
1	no system
2	*c,f,g,h*
3	*e*
4	*b*
5	*d*
9	*i* and *j*

Comparing this to the previous list of "Highest" to "Lowest" ranking which was based on overall averages (section 2.3.2) shows an almost identical pattern. And so these interactions support the previous conclusions completely <u>but place system *h* with the second best group rather than with system *a*</u>.

The tabulation below summarizes the entire discussion and presents the systems in descending order of overall merit, with groupings to reflect the patterns of significant differences.

	System	Overall Average Diffgrade	Number of transparent materials	Number of materials below -1.0
Group 1	*a*	-0.33	4	0
Group 2	*h*	-0.43	4	2
	g	-0.50	4	2
	f	-0.51	2	2
	c	-0.52	1	2
	e	-0.55	2	3
Group 3	*b*	-0.79	3	4
	d	-0.88	0	5
Group 4	*j*	-2.31	0	9
	i	-2.32	0	9

Table 2.3 Summary of the audio quality test results

This experiment is one of the least ambiguous in outcomes among related experiments undertaken at the CRC. There are very few interaction artifacts that obscure the meaning of the overall system averages. The very fine-grained data resolution of 0.17 of a grade for overall system effects and 0.45 of a grade for the interaction of audio materials and systems are among the smallest magnitudes seen in any CRC quality assessment.

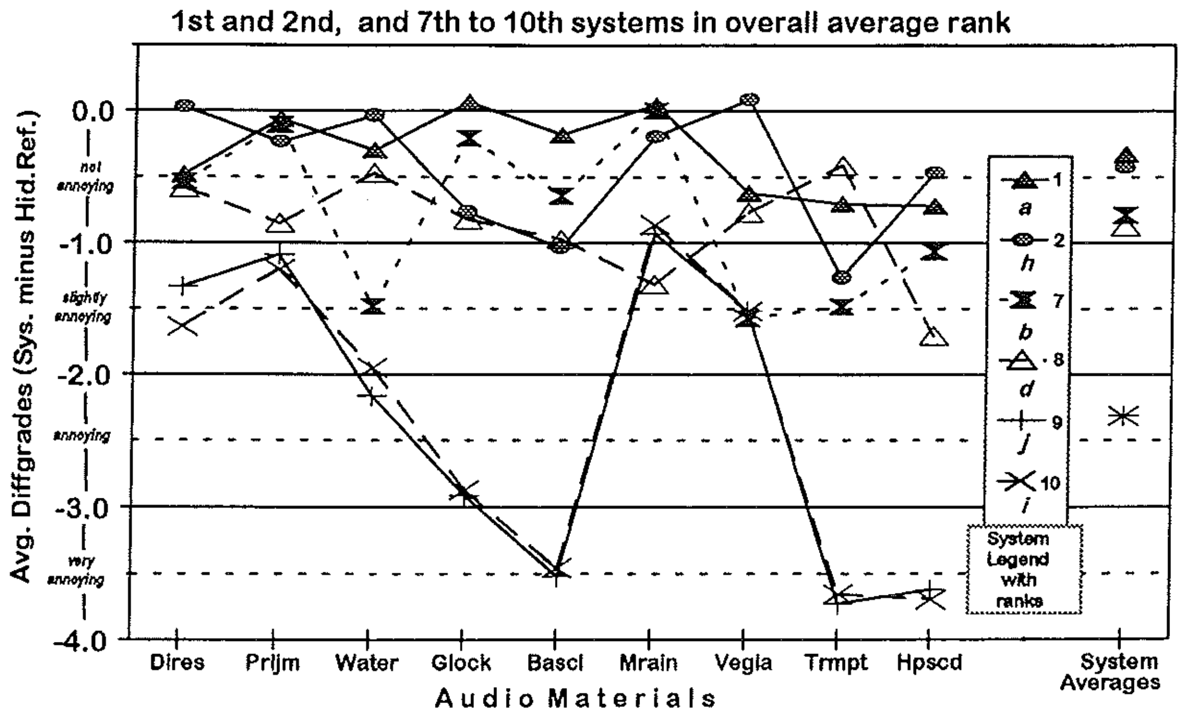


Fig. 2.2a. Quality test results - systems *a*, *b*, *d*, *h*, *i* and *j*

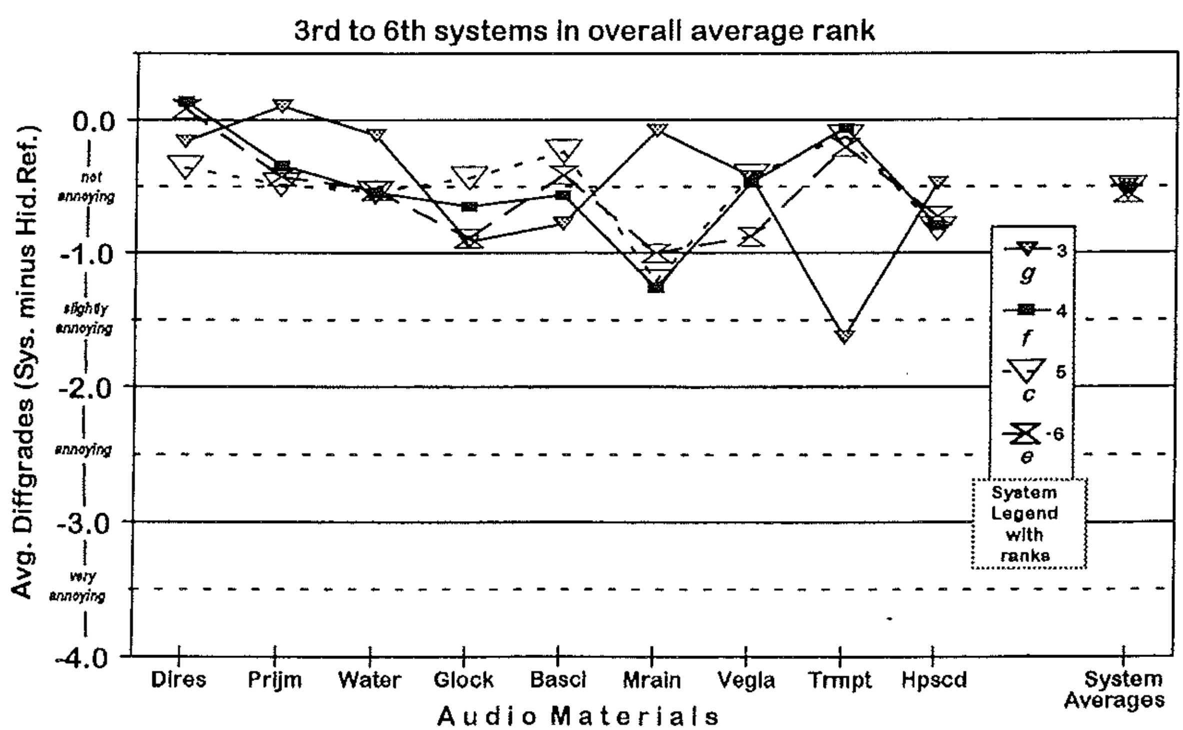


Fig. 2.2b. Quality test results - systems *c*, *e*, *f*, and *g*

Table 2.4a Average Difference Grades for each of the 9 Audio Materials (columns) by each of the 10 Systems

The data for a single system are shown throughout each row. Systems are arranged in descending average merit, with the highest rated system at the top, as shown in the far right-hand column (see Fig. 2.2a and 2.2b).

System	Dires	Prljm	Water	Glock	Bascl	Mrain	Vegla	Trmpt	Hpscd		Overall Averages
a	-0.49	-0.06	-0.30	0.07	-0.18	0.04	-0.62	-0.70	-0.72	*a*	-0.33
h	0.02	-0.24	-0.04	-0.77	-1.04	-0.20	0.08	-1.27	-0.47	*h*	-0.43
g	-0.16	0.10	-0.11	-0.92	-0.78	-0.08	-0.43	-1.63	-0.48	*g*	-0.50
f	0.14	-0.34	-0.55	-0.65	-0.57	-1.26	-0.47	-0.06	-0.80	*f*	-0.51
c	-0.36	-0.49	-0.54	-0.44	-0.24	-1.21	-0.42	-0.12	-0.82	*c*	-0.52
e	0.09	-0.43	-0.53	-0.89	-0.41	-1.00	-0.88	-0.20	-0.72	*e*	-0.55
b	-0.54	-0.10	-1.49	-0.21	-0.64	0.00	-1.58	-1.49	-1.07	*b*	-0.79
d	-0.59	-0.85	-0.47	-0.82	-0.97	-1.31	-0.77	-0.41	-1.70	*d*	-0.88
j	-1.34	-1.09	-2.16	-2.91	-3.52	-0.93	-1.51	-3.73	-3.62	*j*	-2.31
i	-1.64	-1.20	-1.95[1]	-2.87	-3.46	-0.86	-1.52	-3.66	-3.70	*i*	-2.32
Audio Material Averages	-0.49	-0.47	-0.81	-1.04	-1.18	-0.68	-0.81	-1.33	-1.41		-0.91
Averages Without *i* and *j*	-0.24	-0.30	-0.50	-0.58	-0.60	-0.63	-0.64	-0.74	-0.85		-0.56

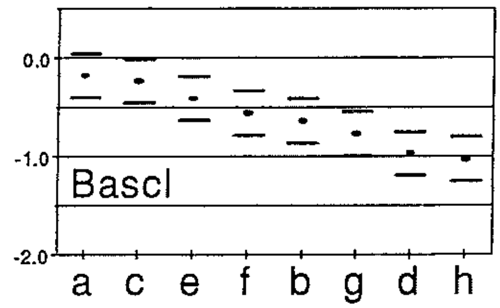
Table 2.4b Average Difference Grades for each of the 9 Audio Materials (columns) by each of the 10 Systems

The data in the cells under each audio material are arranged in descending order of system merit, with the highest ranking system for that material at the top (see Fig. 2.2a and 2.2b).

Dires	Prljm	Water	Glock	Bascl	Mrain	Vegla	Trmpt	Hpscd
f 0.14	*g* 0.10	*h* -0.04	*a* 0.07	*a* -0.18	*a* 0.04	*h* 0.08	*f* -0.06	*h* -0.47
e 0.09	*a* -0.06	*g* -0.11	*b* -0.21	*c* -0.24	*b* 0.00	*c* -0.42	*c* -0.12	*g* -0.48
h 0.02	*b* -0.10	*a* -0.30	*c* -0.44	*e* -0.41	*g* -0.08	*g* -0.43	*e* -0.20	*a* -0.72
g -0.16	*h* -0.24	*d* -0.47	*f* -0.65	*f* -0.57	*h* -0.20	*f* -0.47	*d* -0.41	*e* -0.72
c -0.36	*f* -0.34	*e* -0.53	*h* -0.77	*b* -0.64	*i* -0.86	*a* -0.62	*a* -0.70	*f* -0.80
a -0.49	*e* -0.43	*c* -0.54	*d* -0.82	*g* -0.78	*j* -0.93	*d* -0.77	*h* -1.27	*c* -0.82
b -0.54	*c* -0.49	*f* -0.55	*e* -0.89	*d* -0.97	*e* -1.00	*e* -0.88	*b* -1.49	*b* -1.07
d -0.59	*d* -0.85	*b* -1.49	*g* -0.92	*h* -1.04	*c* -1.21	*j* -1.51	*g* -1.63	*d* -1.70
j -1.34	*j* -1.09	*i* -1.95*	*i* -2.87	*i* -3.46	*f* -1.26	*i* -1.52	*i* -3.66	*j* -3.62
i -1.64	*i* -1.20	*j* -2.16	*j* -2.91	*j* -3.52	*d* -1.31	*b* -1:58	*j* -3.73	*i* -3.70

*System *i* received a grade of -1.95 for Water. In view of the statistical error, and also because *i* was not significantly different from the next higher rated system *b* on this audio material, *i* was omitted from Water in Fig. 2.3 on the next page, along with other instances of *i* and *j* in materials where either of these two systems obtained a diffgrade lower than -2.00. (No systems other than *i* and *j* received any diffgrades below -2.00.)

FIG. 2.3 System Differences Within Audio Materials
Upper and lower statistical boundaries are shown for the average of each system within each audio material. *Only systems with no horizontal overlaps among their boundaries are statistically different.* Within each chart, systems are ordered along the X-axis by the magnitude of their averages.
The vertical axes start at -2.0 rather than, as in Figs. 1a and b, at -4.0. Systems *i* and *j* are omitted from those charts where their averages fall below -2.0. At those low values, *i* and *j* are significantly different from all the other 8 systems in those audio materials without ambiguity.



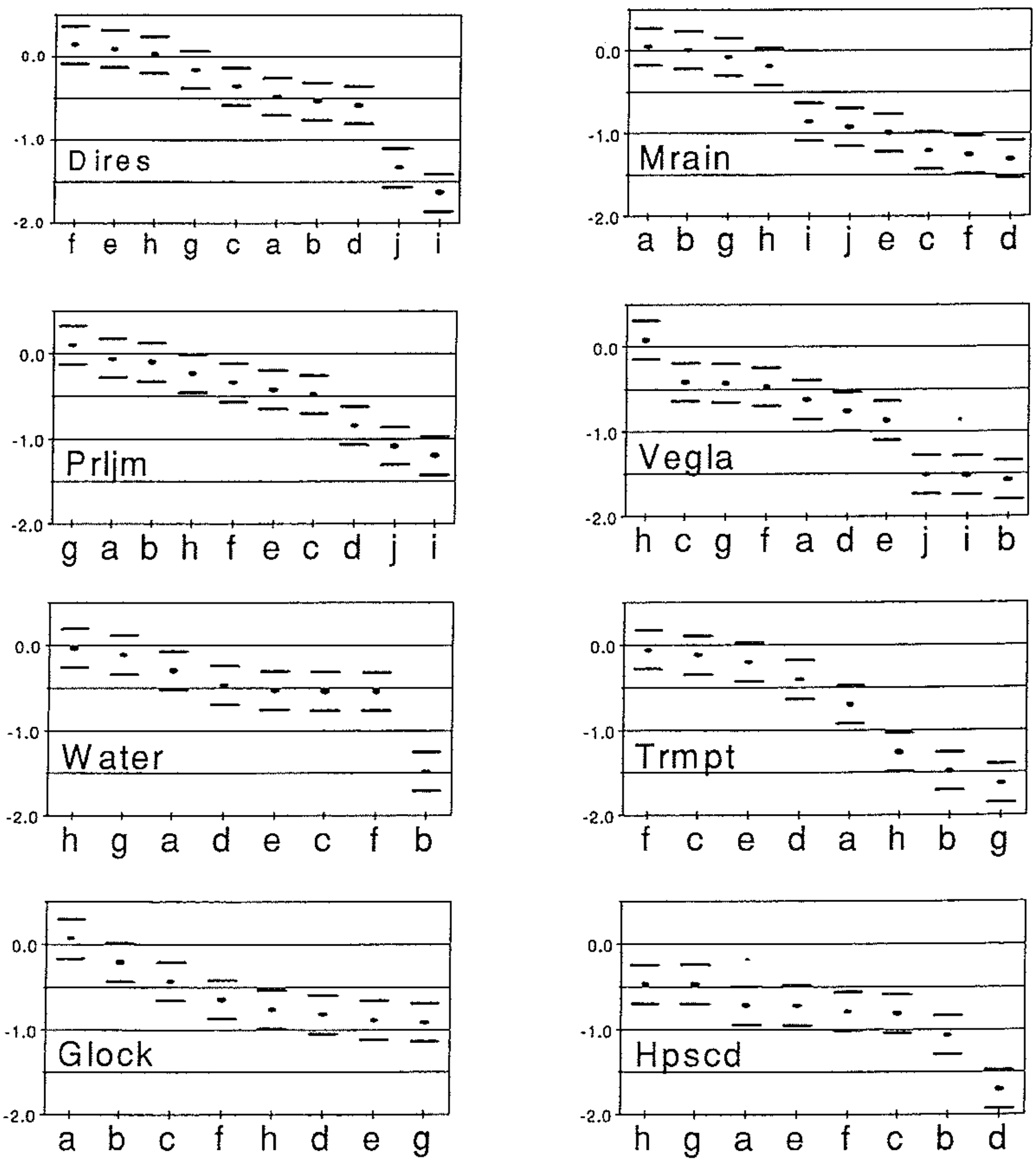


Table 2.5 Magnitudes of the differences between each pair of the 10 systems (a to j) for each of the 9 audio materials

This table parallels the presentation in Fig. 2. In both rows and columns, the systems are ordered by their average difference grades (see Fig. 2.2a and 2.2b). Significant differences (absolute magnitudes of 0.45 or greater) are shown in bold type

	Dires-f	Dires-e	Dires-h	Dires-g	Dires-c	Dires-a	Dires-b	Dires-d	Dires-j	Dires-i
f		0.05	0.11	0.30	**0.50**	**0.62**	**0.68**	**0.72**	**1.48**	**1.78**
e	0.05		0.06	0.25	**0.45**	**0.57**	**0.62**	**0.67**	**1.42**	**1.72**
h	0.11	0.06		0.19	0.39	**0.51**	**0.56**	**0.61**	**1.36**	**1.66**
g	0.30	0.25	0.19		0.20	0.32	0.38	0.42	**1.18**	**1.48**
c	**0.50**	**0.45**	0.39	0.20		0.12	0.18	0.22	**0.98**	**1.28**
a	**0.62**	**0.57**	**0.51**	0.32	0.12		0.05	0.10	**0.85**	**1.15**
b	**0.68**	**0.62**	**0.56**	0.38	0.18	0.05		0.05	**0.80**	**1.10**
d	**0.72**	**0.67**	**0.61**	0.42	0.22	0.10	0.05		**0.75**	**1.05**
j	**1.48**	**1.42**	**1.36**	**1.18**	**0.98**	**0.85**	**0.80**	**0.75**		0.30
i	**1.78**	**1.72**	**1.66**	**1.48**	**1.28**	**1.15**	**1.10**	**1.05**	0.30	

	Prljm-g	Prljm-a	Prljm-b	Prljm-h	Prljm-f	Prljm-e	Prljm-c	Prljm-d	Prljm-j	Prljm-i
g		0.16	0.20	0.34	0.44	**0.53**	**0.59**	**0.95**	**1.19**	**1.30**
a	0.16		0.04	0.18	0.28	0.37	0.42	**0.79**	**1.03**	**1.14**
b	0.20	0.04		0.13	0.24	0.32	0.38	**0.75**	**0.99**	**1.10**
h	0.34	0.18	0.13		0.10	0.19	0.25	**0.61**	**0.85**	**0.96**
f	0.44	0.28	0.24	0.10		0.09	0.14	**0.51**	**0.75**	**0.86**
e	**0.53**	0.37	0.32	0.19	0.09		0.06	0.42	**0.66**	**0.77**
c	**0.59**	0.42	0.38	0.25	0.14	0.06		0.37	**0.60**	**0.71**
d	**0.95**	**0.79**	**0.75**	**0.61**	**0.51**	0.42	0.37		0.24	0.35
j	**1.19**	**1.03**	**0.99**	**0.85**	**0.75**	**0.66**	**0.60**	0.24		0.11
i	**1.30**	**1.14**	**1.10**	**0.96**	**0.86**	**0.77**	**0.71**	0.35	0.11	

	Water-h	Water-g	Water-a	Water-d	Water-e	Water-c	Water-f	Water-b	Water-i	Water-j
h		0.08	0.26	0.43	**0.50**	**0.50**	**0.51**	**1.45**	**1.91**	**2.12**
g	0.08		0.19	0.36	0.42	0.43	0.43	**1.37**	**1.83**	**2.05**
a	0.26	0.19		0.17	0.23	0.24	0.25	**1.19**	**1.65**	**1.86**
d	0.43	0.36	0.17		0.06	0.07	0.08	**1.01**	**1.48**	**1.69**
e	**0.50**	0.42	0.23	0.06		0.01	0.01	**0.95**	**1.41**	**1.63**
c	**0.50**	0.43	0.24	0.07	0.01		0.00	**0.94**	**1.40**	**1.62**
f	**0.51**	0.43	0.25	0.08	0.01	0.00		**0.94**	**1.40**	**1.61**
b	**1.45**	**1.37**	**1.19**	**1.01**	**0.95**	**0.94**	**0.94**		**0.46**	**0.68**
i	**1.91**	**1.83**	**1.65**	**1.48**	**1.41**	**1.40**	**1.40**	**0.46**		0.21
j	**2.12**	**2.05**	**1.86**	**1.69**	**1.63**	**1.62**	**1.61**	**0.68**	0.21	

	Glock-a	Glock-b	Glock-c	Glock-f	Glock-h	Glock-d	Glock-e	Glock-g	Glock-i	Glock-j
a		0.28	**0.50**	**0.72**	**0.83**	**0.89**	**0.96**	**0.99**	**2.94**	**2.98**
b	0.28		0.23	0.44	**0.56**	**0.61**	**0.68**	**0.71**	**2.66**	**2.70**
c	**0.50**	0.23		0.21	0.33	0.39	**0.45**	**0.48**	**2.43**	**2.47**
f	**0.72**	0.44	0.21		0.11	0.17	0.24	0.27	**2.22**	**2.26**
h	**0.83**	**0.56**	0.33	0.11		0.06	0.12	0.15	**2.10**	**2.14**
d	**0.89**	**0.61**	0.39	0.17	0.06		0.07	0.10	**2.05**	**2.09**
e	**0.96**	**0.68**	**0.45**	0.24	0.12	0.07		0.03	**1.98**	**2.02**
g	**0.99**	**0.71**	**0.48**	0.27	0.15	0.10	0.03		**1.95**	**1.99**
i	**2.94**	**2.66**	**2.43**	**2.22**	**2.10**	**2.05**	**1.98**	**1.95**		0.04
j	**2.98**	**2.70**	**2.47**	**2.26**	**2.14**	**2.09**	**2.02**	**1.99**	0.04	

	Bascl-a	Bascl-c	Bascl-e	Bascl-f	Bascl-b	Bascl-g	Bascl-d	Bascl-h	Bascl-i	Bas[illegible]
a		0.06	0.23	0.39	**0.46**	**0.60**	**0.79**	**0.86**	**3.28**	3[illegible]
c	0.06		0.18	0.33	0.40	**0.54**	**0.73**	**0.80**	**3.22**	3.2[illegible]
e	0.23	0.18		0.15	0.23	0.37	**0.56**	**0.62**	**3.04**	**3.11**
f	0.39	0.33	0.15		0.08	0.21	0.40	**0.47**	**2.89**	**2.96**
b	**0.46**	0.40	0.23	0.08		0.14	0.33	0.40	**2.81**	**2.88**
g	**0.60**	**0.54**	0.37	0.21	0.14		0.19	0.26	**2.68**	**2.74**
d	**0.79**	**0.73**	**0.56**	0.40	0.33	0.19		0.07	**2.49**	**2.55**
h	**0.86**	**0.80**	**0.62**	0.47	0.40	0.26	0.07		**2.42**	**2.49**
i	**3.28**	**3.22**	**3.04**	**2.89**	**2.81**	**2.68**	**2.49**	**2.42**		0.07
j	**3.34**	**3.29**	**3.11**	**2.96**	**2.88**	**2.74**	**2.55**	**2.49**	0.07	

	Mrain-a	Mrain-b	Mrain-g	Mrain-h	Mrain-i	Mrain-j	Mrain-e	Mrain-c	Mrain-f	Mrain-d
a		0.04	0.12	0.24	**0.90**	**0.97**	**1.04**	**1.26**	**1.30**	**1.36**
b	0.04		0.08	0.20	**0.86**	**0.93**	**1.00**	**1.21**	**1.26**	**1.31**
g	0.12	0.08		0.11	**0.78**	**0.85**	**0.92**	**1.13**	**1.18**	**1.23**
h	0.24	0.20	0.11		**0.67**	**0.73**	**0.80**	**1.02**	**1.07**	**1.12**
i	**0.90**	**0.86**	**0.78**	**0.67**		0.07	0.14	0.35	0.40	**0.45**
j	**0.97**	**0.93**	**0.85**	**0.73**	0.07		0.07	0.29	0.33	0.39
e	**1.04**	**1.00**	**0.92**	**0.80**	0.14	0.07		0.21	0.26	0.31
c	**1.26**	**1.21**	**1.13**	**1.02**	0.35	0.29	0.21		0.05	0.10
f	**1.30**	**1.26**	**1.18**	**1.07**	0.40	0.33	0.26	0.05		0.05
d	**1.36**	**1.31**	**1.23**	**1.12**	**0.45**	0.39	0.31	0.10	0.05	

	Vegla-h	Vegla-c	Vegla-g	Vegla-f	Vegla-a	Vegla-d	Vegla-e	Vegla-j	Vegla-i	Vegla-b
h		**0.50**	**0.51**	**0.55**	**0.70**	**0.85**	**0.96**	**1.60**	**1.60**	**1.66**
c	**0.50**		0.01	0.05	0.20	0.35	**0.46**	**1.10**	**1.10**	**1.16**
g	**0.51**	0.01		0.04	0.20	0.34	**0.45**	**1.09**	**1.09**	**1.15**
f	**0.55**	0.05	0.04		0.15	0.30	0.40	**1.04**	**1.05**	**1.10**
a	**0.70**	0.20	0.20	0.15		0.14	0.25	**0.89**	**0.90**	0.[illegible]
d	**0.85**	0.35	0.34	0.30	0.14		0.11	**0.75**	**0.75**	0.[illegible]
e	**0.96**	**0.46**	**0.45**	0.40	0.25	0.11		**0.64**	**0.64**	**0.70**
j	**1.60**	**1.10**	**1.09**	**1.04**	**0.89**	**0.75**	**0.64**		0.00	0.06
i	**1.60**	**1.10**	**1.09**	**1.05**	**0.90**	**0.75**	**0.64**	0.00		0.06
b	**1.66**	**1.16**	**1.15**	**1.10**	**0.95**	**0.81**	**0.70**	0.06	0.06	

Trmpt-f Trmpt-c Trmpt-e Trmpt-d Trmpt-a Trmpt-h

3. Subjective Tests of Transmission Impairments

3.1 Background

This chapter describes the results obtained in a series of subjective tests performed to assess the effects of transmission errors on the audio quality of the DAR systems. Processing of the audio materials through the DAR systems took place at the NASA LeRC in Cleveland, USA. Reference and processed versions of the test materials were delivered to the CRC on DAT audio tapes and were transferred digitally, via the AES/EBU interface, onto a custom disk-based playback system which was used for the subjective tests. Details on this playback system as well as other facilities used during the tests are described in chapter 6 of this report.

A total of seven different types of impairment were assessed. These are:

1. Gaussian noise
2. Co-channel interference
3. Urban slow multipath
4. Urban fast multipath
5. Rural fast multipath
6. Obstructed fast multipath
7. DAR lower 1st adjacent to DAR channel interference

As will be shown in the section discussing the results, a total of nine DAR systems, labeled from *a* to *i*, were tested for additive gaussian noise and co-channel interference. A subset of four systems were tested in the urban fast multipath, three in both the urban slow and the obstructed fast multipath and two systems in the rural fast multipath. Only one system was tested for DAR lower 1st adjacent to DAR channel interference. Details on the characteristics of the simulated multipath mobile channels can be found in [3].

As described in [2], two series of subjective tests were performed, namely the threshold of audibility test, to accurately determine the point at which transmission impairment are just perceptible, and the failure characteristic test to determine the way in which the subjective quality degrades with increasing levels of transmission impairment.

3.2 Test procedures

3.2.1 Test materials

The three critical audio materials described in Table 3.1 below were used in all the impairment tests reported in this chapter. These materials were carefully selected by the staff at the NASA LeRC laboratory because they were particularly sensitive to revealing artifacts resulting from transmission impairments. Only one of these materials, namely Glockenspiel, was also used in the audio quality subjective tests described in chapter 2.

Code	Description	Duration	Source
Glock	Glockenspiel	16 s	EBU SQAM CD (track 35/Index 1)
Clarn	Clarinet solo	20 s	EBU SQAM CD (track 16/Index 2)
Sopra	Soprano solo	31 s	EBU SQAM CD (track 62 /Index 1)

Table 3.1 List of audio materials used in impairments tests

3.2.2 Impairment levels

As outlined in [2], each of the three audio test materials was recorded, for each DAR system and each type of impairment, at the following levels of impairment: CC, TOA_1, TOA_2, ..., TOA_N, S_1, S_2, ... , S_M, POF where:

CC = coded audio in a Clear Channel
TOA_1, TOA_2, ..., TOA_N = N stimuli (three or more) in the close neighborhood of the approximate TOA (threshold of audibility)
S_1, S_2, ... , S_M = M intermediate levels of impairments
POF = point of failure

For each of the above impairment levels, the C_o/N_o (or D/U for co-channel interference) ratio was noted. The CC level was recorded at a high value of C_o/N_o (or D/U) so that transmission errors could be considered as negligible, hence the label "clear channel" given to that level. The increments in the C_o/N_o (or D/U) ratios at which the other impairment levels were recorded varied from 0.25 dB, for those DAR systems which failed abruptly, up to 1.0 dB for those systems which failed more gracefully, with increments of 0.5 dB being the typical value used for most systems and most of the impairments.

3.2.3 Subjective test procedures

As described in [2], two separate experiments were performed for each of the seven types of impairment considered:

<u>Experiment 1</u>: Threshold of Audibility

The purpose of this experiment was to provide a sensitive and reliable measurement of the threshold of audibility (TOA) for the various types of channel impairments. The threshold of audibility is defined as the highest Co/No (or D/U) level where an audible artifact due to a transmission error can be detected. Determining TOA consisted of a two part process and was performed using CC (the coded audio in clear channel) as the reference signal. Simply stated, the

TOA values were determined by agreement between the in-house expert listeners at the NASA LeRC and the CRC.

a) Determining TOA_{Lab}:

The first step in this experiment was to determine TOA_{Lab}. TOA_{Lab} is the value for TOA determined initially by the LeRC team. To determine TOA_{Lab}, the LeRC team processed the critical source materials through a given system and listened to the output in real-time. The level of the impairment was adjusted in 0.25dB increments until the expert listeners determined TOA_{Lab} for that system/critical material/impairment combination. The determination of TOA_{Lab} was based on listening to a given critical material at least 3 times at each level of impairment or until the impairment was detected. This procedure was repeated for each combination of system, critical material, and impairment. During the process of determining TOA_{Lab}, the LeRC team also established POF_{Lab}.

Having determined TOA_{Lab} and POF_{Lab}, the LeRC team processed the source materials through the systems and recorded the output onto DAT tapes. Since these recordings were to be used in the failure characteristic tests, recordings were also made for several impairment levels on either side of TOA_{Lab} and POF_{Lab} as well as for intermediate levels between TOA_{Lab} and POF_{Lab} for those systems that failed gracefully. In particular, for the conditions near the TOA_{Lab} impairment level (i.e. TOA_{Lab} +/- 0.25dB and +/- 0.5dB), additional recordings were made. This was done to account for the statistical nature of the occurrence of impairment conditions near TOA. The DAT tapes containing the recorded materials were then sent to the CRC team while the LeRC team kept a duplicate set.

b) Determining the True TOA:

Upon receiving the DAT tapes the CRC team transferred the audio recordings to the CRC's hard disk playback system. The CRC expert listeners then auditioned the recordings in order to evaluate the values of TOA_{Lab} established by the LeRC team. Three possible scenarios resulted from these listening tests and are discussed below.

Scenario 1: The CRC team agrees with the LeRC team that TOA_{Lab} is in fact the true TOA.

In this case, the two teams agreed that TOA_{Lab} should be designated as the true TOA. This was by far the most common scenario.

Scenario 2: The CRC team cannot hear any audible artifact at TOA_{Lab}.

Due to the statistical nature of the occurrence of impairments, it was quite possible that no audible artifacts existed (at TOA_{Lab}) in some of the recordings sent to CRC. To minimize this possibility, the LeRC team recorded several samples at TOA_{Lab}. If the CRC team could not hear any audible artifacts in any of the samples (at TOA_{Lab}) then they notified the LeRC team. The LeRC team then listened to their duplicate tapes to determine whether or not they could detect any artifacts in the samples at TOA_{Lab}. If the LeRC team was not able to detect any artifacts at TOA_{Lab} then the next level of impairment (typically TOA_{Lab} +0.5dB) was designated as the true TOA.

If the LeRC team was able to identify audible artifacts at TOA_{Lab} then they provided the CRC team with a detailed description of each artifact and where it occurred. The CRC team then listened to the recordings to confirm the LeRC team's observations. Once the two teams agreed that an audible artifact existed, TOA_{Lab} was designated as the true TOA.

Scenario 3: The CRC team can hear artifacts at impairment levels lower than TOA_{Lab}.

During the course of evaluating TOA_{Lab}, the CRC team also listened to recordings with less impairment than TOA_{Lab} (i.e. TOA_{Lab} -0.25 or -0.5dB) to determine if any audible artifacts could be detected. Due to the statistical nature of the occurrence of the impairments, it was possible for audible artifacts to exist at these impairment levels in some of the recordings sent to CRC.

In those instances where the CRC team was able to detect audible artifacts at lower impairment levels, a detailed description of each artifact and where it occurred was sent to the LeRC team. The LeRC team then listened to their duplicate recordings to confirm the CRC team's evaluation. Once the two teams agreed that an audible artifact existed, that impairment level was designated as the true TOA.

Experiment 2: Failure Characteristic

The purpose of the second experiment was to determine how the subjective quality of each DAR system degraded with increasing levels of impairment. The uncoded CD original was used as the reference against which the following stimuli were compared and rated: CC, TOA (as determined in experiment 1 above), S_1, S_2, ... , S_M, POF.

The test procedure used was essentially the same as the one used in the audio quality tests (see section 2.2.2). The informed reader may skip the rest of this section up to the paragraph preceding Table 3.2. The procedure included a training phase followed by a blind rating phase. At least half a day (the morning of the first day) was devoted to the training session. Subjects worked together in groups of 2 or 3 and were allowed as much time as they required for training. Subjects were trained using the same hard disk playback system used in the blind rating tests. For each experiment, a subset of the stimuli to be rated in the blind rating test was used for the training session. This subset consisted of stimuli which were considered to be representative of those to be rated in the subsequent blind rating test.

For the blind rating phase, the triple-stimulus A-B-C presentation was used, where "A" was the known reference (unprocessed CD signals). One of "B" or "C" was the stimulus to be rated (i.e. one of the impairment levels) and the other one was the hidden reference (i.e. a perfect replica of "A"). The assignment of stimulus and hidden reference to "B" and "C" was not known to the subjects and was arranged to be unpredictable to the listeners from trial to trial. For each trial, subjects were asked to rate the *difference* between the known reference "A" and version "B" as well as the difference between "A" and "C" using the same ITU-R continuous 5-grade impairment scale as in the audio quality tests and shown in Fig. 2.1 of chapter 2.

Listeners were instructed to use scores with a single decimal point. In effect, this is a 41 point scale. The subjects were instructed to treat this as a continuous scale but, to facilitate the subjects' orientation, category labels were associated with the scale. Thus, 1.0 to 1.9 is a "very annoying" range; 2.0 to 2.9 is "annoying"; 3.0 to 3.9 is "slightly annoying"; 4.0 to 4.9 is "perceptible but not annoying". Finally, 5.0 is "imperceptible". A constraint was that the alternative the subject has decided is the "hidden reference" must be graded 5.0. And so, *at least one* of the two grades on each trial must be a 5.0.

Thus two totally interdependent scores from the listener were recorded on each trial. This deliberate interdependence was handled by subtracting the score given to the true hidden reference from the score given the true impaired version (i.e., DAR System minus Reference). The reason for subtracting in this direction rather than the opposite one (Reference minus DAR System) is only so that in a graphical plot of outcomes, the data will fall in the same geometric quadrant as they would if the actual 1.0 to 5.0 scores used by the subjects were plotted.

This subtraction means, however, that the scores are transformed so that the 1.0 to 5.0 range of the original scale becomes, instead, -4.0 to 0.0 in the analysis and presentation of results. These difference grades or "diffgrades" represent the relative differences between the grades given to the hidden reference and the ones given to the DAR system under test. For practical purposes, it can be assumed that a direct analogy holds between the original 1.0 to 5.0 scale and the -4.0 to 0.0 diffgrades. Accordingly, to facilitate comparison with the 1.0 to 5.0 scale used by subjects, the "annoyance" category labels are shown between the Y-axis numbers in the various figures shown in this chapter.

Blind rating was performed individually by each listener. Listeners used the disk-based playback system described in chapter 5 of this report which allowed seamless switching between the stimuli to be compared. Listeners were able to take as much time as they needed on each trial, switching as often as they liked, until satisfied with the numerical ratings they were asked to assign. They were also free to use either the loudspeakers or headphones to make a judgment, whichever they felt was the most critical transducer on any trial.

The total number of materials to be rated was divided into sessions of 10-15 trials each and, within each session, items were ordered in a way which was unpredictable by the listeners. Moreover, the order of presentation of the individual sessions was changed from listener to listener over the course of the test to ensure that time-correlated factors (such as fatigue) would not differentially affect any level of any of the factors under test.

The seven impairments tested were grouped into five separate experiments as shown in Table 3.2. The table also shows the number of DAR systems tested and the number of trials for each impairment. The number of days spent by each listener for each experiment is also shown in this table. The listening panel included a total of 6 expert-listeners for each of the experiments. Two subjects took part in all of the experiments. The remaining four panel members were in general different from experiment to experiment.

Experiment	Impairment	Number of systems tested	Number of trials	Number of days per listener
1	AWGN	9	150	3
2	Co-channel interference	9	150	3
3	Urban Fast Multipath	4	72	2
4	Urban Slow Multipath Rural Fast Multipath	3 2	54 36	2
5	Obstructed Fast Multipath DAR lower 1st Adj. to DAR channel interference	3 1	54 21	2

Table 3.2 Grouping of the seven impairments into five experiments

3.3 Test results

3.3.1 Data presentation and statistical analysis

The impairment test results are presented in three different forms. First, a table shows, for each impairment, each DAR system and each test material, the values of C_o/N_o (or D/U) at TOA, POF and the difference between the two values, labelled in this chapter as the *Failure Margin (FM)*. The failure margin gives an indication of how quickly the subjective quality degrades with a reduction in C_o/N_o (or D/U). **The definition of POF used in this report is any score which falls in the "very annoying" range of the rating scale, that is below -3.0.** When more than one impairment level have been rated below -3.0, the level with the **highest C_o/N_o (or D/U)** has been taken as the POF. In the results presented throughout this chapter, **the term TOA will be used to mean the true TOA level** as determined in the TOA test. Two different types of graphical representations of the results are provided. The first shows, for each impairment and each test material, the average diffgrade given to each presentation plotted against the corresponding C_o/N_o (or D/U) ratio. The resulting graph is called the failure characteristic curve. The second type of graph shows the failure characteristic curve of all three test materials for a single system and a given impairment.

In impairment studies, expert observation and commentary (EO&C) methodology is often used. In compliance with [2], a grading scale was used (the same one that was used in the audio quality assessments) in the present tests because this approach has advantages over EO&C. One advantage is that each person provides their ratings in individual blind sessions rather than in a group situation. Thus, the judgments of each individual are less influenced by those of the other judges. Also, the audio item presentation sequences are controlled by the experimenter and are varied from subject to subject to ensure that events correlated with time (fatigue, learning, etc.)

do not distort the average grades associated with each experimental condition. By contrast, with EO&C, all listeners hear the materials at the same time in a group situation. And so the effects of time-correlated events cannot be factored out by experimental design.

With EO&C, you cannot have a statistical analysis of results, since that methodology provides only a group consensus rather than sets of independent judgments, one set from each listener. However, even though we used a grading scale here, we cannot present a statistical analysis of results such as is usually associated with this methodology. The major reason for this is that, as required by the nature of the tests, gross impairments (e.g. POF) were included among the sets of stimuli, mixed in with small (e.g. clear channel) and intermediate impairments. The nature of the grading scale that was used, however, did not permit comparable discriminations along the whole range of impairments that were presented. The scale problem was at the "very annoying" end of the grading scale. Thus, the scale suffered a "floor effect" where listeners could not assign a grade worse than the minimal one allowed of 1.0, even though there might have been considerable subjective variation among those stimuli that were judged to be grossly impaired. Thus, the scale permitted rather uniform variance to occur as at the top and middle areas of the impairment scale used, but not so among the gross impairments at the bottom of the scale.

The resulting set of numerical subjective judgments, then, are "distorted". And this kind of distortion is one which precludes performing a meaningful statistical analysis since this requires that similar variance among judgments is obtained across the entire range of the scale that is used by subjects. In addition, the number of subjects was small (6 per impairment experiment) hence statistical variance was large.

Out of curiosity, ANOVA's were actually performed for each experiment, and it was found that, if these analyses had been legitimate, then the critical difference between two scores required for statistical significance was found to be half a grade, for the DAR lower 1st adjacent to DAR channel interference experiment and between 0.8 to a full grade for all the other experiments. However, because the violations of essential statistical requirements discussed above were serious ones, we cannot place any confidence in these outcomes and do not present or quote these analyses. All in all, then, we feel the results are quantitatively superior to those that would have been obtained using EO&C, but because of the inherent nature of the range of impairments, they are not, and could not be, fully equivalent to those in the quality tests.

All the conditions in the system audio quality tests which took place previous to the series of impairment tests discussed here, were "clear-channel" conditions. It might be suggested that the clear-channel baselines established in each test in the impairment series should show similar grades to those obtained by the comparable systems in the previous quality tests. However, in the quality tests, there were 9 audio materials, while there were only three in the impairment tests, and only one material, namely Glockenspiel, was in common between the two studies. Furthermore, as discussed above, the total range of impairments was quite different between the quality and the impairment tests, and introduced an unavoidable distortion in the grading scale for the impairment tests. This means that the total context of the impairment experiments was substantially different from that of the quality tests. And also, as mentioned previously, the number of subjects was considerably smaller in the impairment experiments. Moreover, the subjects in the quality tests were in general different than those who took part in the various impairment tests.

For all of these reasons, no comparisons can legitimately be made between the quality experiments on the one hand, and the clear-channel conditions in the impairment tests. The same statement can be made for comparisons between the clear-channel conditions of the seven impairment tests.

3.3.2 Gaussian noise

A total of nine DAR systems were tested with the gaussian noise impairment. The complete failure characteristic curves for systems *a* through *h* are shown in Fig. 3.1, 3.2 and 3.3 for *Glockenspiel*, *Soprano* and *Clarinet* respectively. Each curve shows the mean opinion score (average diffgrade across all 6 listeners) given to each impairment level vs the C_o/N_o value at each of these levels. The failure characteristic curves for system *i* are given in Fig. 3.12. The score given to the coded audio in a clear channel is plotted, for each system, as a separate point identified on the X-axis by the *Clear* label. For visual clarity, the *Clear* point has not been linked to the rest of the curve. The first point to the right of the *Clear* point on each curve is the *TOA* level as determined in the TOA test.

It can be seen in Fig. 3.1, 3.2 and 3.3 that a few curves are not completely monotonic, that is the scores do not always decrease with a reduction of the C_o/N_o value. These non-monotonic fluctuations are caused by the random nature of the gaussian noise impairment. Statistically, it is possible for a momentary high level noise transient to occur at a given impairment level and generate more audible degradation than the next higher impairment level which would be free of such transients. In some instances, it is found that the score at TOA is worse than the next higher level of impairment. This may be explained by the fact that, for TOA, more time was probably taken to detect and record an impaired audio segments that contained a noise transient than for the next higher level of impairment. It is also well known to the authors of this report that listeners base their judgements on the worst impairment they heard over the entire duration of a given test material. Therefore, a single isolated and momentary artifact ("click", mute, etc...) may result in a rather poor or marginal score.

These observations also apply to the failure characteristic curves shown in section 3.3.3 for co-channel interference. The co-channel interfering signal is generally independent of the desired signal and, like the gaussian noise, exhibits random level fluctuations. In the case of multipath impairments (section 3.3.4 to 3.3.7), the impairments are the results of the combination of two random processes: the mobile channel and the gaussian noise. The mobile channel goes through a series of fades in an unpredictable and random fashion. It is consequently possible for a particularly bad fade pattern to happen at a given impairment (i.e. noise) level and for a less severe one to occur during the recording of the next higher impairment level. It must be remembered that injected noise was the parameter varied between each impairment level in the various multipath tests. The interaction of the random fades with the random noise is probably the reason why the failure characteristic curves for the multipath impairments exhibit, in general, more non-monotonicity than the curves obtained in gaussian noise alone and co-channel interference.

As shown in Table 3.3, the C_o/N_o values at TOA cover a range of approximately 23 dB from system *f* to system *g* for all three test materials. In this table, the systems have been arranged in ascending order of their C_o/N_o value at TOA. System *f* is at one extreme of the range with a C_o/N_o value at TOA of 3.26 dB for all three test materials. At the other extreme lies system *g* which required a C_o/N_o value at TOA of around 25-26 dB for the three materials tested. Systems *a* and *b* are practically identical while the values for systems *e* and *h* are very close to each other, being separated by 1 dB or less for all three test materials.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
f	3.26	2.76	0.5	3.26	2.76	0.5	3.26	3.01	0.25
b	8.46	7.46	1.0	8.71	6.71	2.0	8.96	6.46	2.5
a	8.48	6.48	2.0	8.23	6.73	1.5	8.98	7.98	1.0
h	10.51	9.01	1.5	9.51	8.51	1.0	10.01	9.51	0.5
e	10.76	9.76	1.0	10.51	9.51	1.0	10.76	9.76	1.0
c	11.36	10.86	0.5	11.11	10.61	0.5	11.11	10.61	0.5
d	18.85	17.85	1.0	18.1	16.6	1.5	18.6	17.1	1.5
g	25.1	23.1	2.0	25.1	22.1	3.0	26.35	22.85	3.5
i	16.82			16.82			17.07		

Table 3.3 C_o/N_o ratios at TOA and POF and
failure margin for the gaussian noise impairment

Systems *f* and *c* showed an average (across all three test materials) failure margin around 0.5 dB while this value was 1.0 dB for both systems *e* and *h*. Systems *d* and *a* are next with 1.33 dB and 1.5 dB respectively. Systems *b* and *g* follow with average failure margin values of 1.8 and 2.8 dB respectively.

The failure characteristic curves for each individual system and all three test materials are plotted in Fig. 3.4 to 3.12 for systems *a* to *i* respectively. In general, the results obtained with the three test materials are within 1 dB or so of each other, with *Soprano* being the (slightly) less critical of the three materials.

3.3.3 Co-channel interference

As for the gaussian noise, a total of nine DAR systems were tested with the co-channel interference. The co-channel failure characteristic curves for systems *a* through *h* are shown in Fig. 3.13, 3.14 and 3.15 respectively for *Glockenspiel, Soprano* and *Clarinet*. The failure characteristic curves for system *i* are given in Fig. 3.24.

As shown in Table 3.4, the D/U values at TOA cover a range of approximately 36 to 38 dB from system *f* to system *g* for all three test materials. In this table, the systems have been arranged in ascending order of their D/U values at TOA. System *f* is at one extreme of the range with a D/U value at TOA of 5.5 dB for all three test materials. At the other extreme lies system *g* which required a D/U value at TOA of around 42-43 dB for the three materials tested. As for the gaussian noise impairment, systems *a* and *b* are practically identical while the D/U values at TOA for systems *e* and *h* are very close to each other, being separated by 1 dB or less for all three test materials. Systems *e* and *c* are also separated, in general, by less than 1 dB from TOA to POF.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
f	5.5	4.5	1.0	5.25	4.75	0.5	5.5	5.0	0.5
b	8.5	6.0	2.5	8.25	6.25	2.0	8.5	7.0	1.5
a	8.6	5.85	2.75	7.85	5.85	2.0	8.35	6.85	1.5
h	11.37	8.87	2.5	10.87	7.87	3.0	10.87	9.87	1.0
e	11.12	9.62	1.5	10.87	9.87	1.0	10.87	9.87	1.0
c	11.64	10.64	1.0	11.14	10.64	0.5	11.64	10.64	1.0
d	17.4	15.4	2.0	17.15	15.65	1.5	17.4	15.9	1.5
g	42.6	39.6	3.0	41.6	39.1	2.5	43.1	40.1	3.0
i	26.98	23.23	3.75	26.23	24.23	2.0	26.73	25.73	1.0

Table 3.4 D/U ratios at TOA and POF and failure margin
for the co-channel interference impairment

Systems *f* and *c* showed an average (across all three test materials) failure margin of around 0.7-0.8 dB while this value was 1.2 dB for system *e* and 1.7 for system *d*. Systems *b*, *a* and *h* are next with an average failure margin of around 2.0 dB. Finally, systems *g* has the slowest degradation rate with an average margin of 2.8 dB.

The failure characteristic curves for each individual system and all three test materials are plotted in Fig. 3.16 to 3.24 for systems *a* to *i* respectively. In general, the results for the three test materials were within 1 dB or so of each other, with *Clarinet* being the (slightly) more critical of the three materials if one considers the overall failure curves. By considering the TOA point only (Table 3.4), all three materials are within 0.5 dB of each other for most of the cases.

3.3.4 Urban slow multipath

Only systems *a, b* and *c* were tested subjectively with the urban slow multipath impairment. The failure characteristic curves of these systems are shown in Fig. 3.25, 3.26 and 3.27 respectively for *Glockenspiel, Soprano* and *Clarinet.* These figures (and Table 3.5) show that the performance of systems *a* and *b* are very close to each other, with system *b* requiring a C_o/N_o ratio 0.5 to 1 dB higher than that of system *a* at TOA across the three test materials. The C_o/N_o ratios at TOA for system *c* are around 32 dB for all three test materials, roughly 10 dB higher than for systems *a* and *b*. The failure margins for systems *a* and *b* are between 6 and 7 dB (except for system *b* on *Clarinet* where a failure margin of 5 dB was obtained (by applying the strict definition of POF of section 3.3.1) while those of system *c* are between 7 and 8 dB.

Examination of Fig. 3.28, 3.29 and 3.30, which show respectively the failure characteristics of systems *a, b* and *c* for all three test materials, reveals that, overall, *Clarinet* is slightly more critical than *Glockenspiel* and *Soprano*. This is particularly true for systems *a* and *b*.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
a	21.79	15.04	6.75	21.29	15.29	6.0	22.04	15.04	7.0
b	22.74	16.24	6.5	21.74	15.74	6.0	22.74	17.74	5.0
c	32.14	24.14[1]	8.0	31.64	24.64	7.0	32.14	24.64	8.0

Note 1: This value has been extrapolated from the lowest measured score of 2.9 at a C_o/N_o value of 24.64 dB (see Figures 3.25, 3.30). It has been assumed that if the C_o/N_o ratio had been reduced to 24.14 (an additional increment of 0.5 dB), the subjective quality would have been rated below -3.0.

Table 3.5 C_o/N_o ratios at TOA and POF and failure margin for the urban slow multipath impairment

3.3.5 Urban fast multipath

Only the four systems, *a*, *b*, *c*, and *e*, were tested subjectively under the urban fast multipath impairment condition. The failure characteristic curves for these systems are given in Fig. 3.31, 3.32, and 3.33 respectively for *Glockenspiel*, *Soprano* and *Clarinet.* Table 3.6 provides the C_o/N_o ratios at TOA and POF, as well as the failure margins (FM) for the four systems with the three test materials.

The failure characteristic curves for each system with all three test materials are plotted in Fig. 3.34, 3.35, 3.36, and 3.37 for systems *a*, *b*, *c*, and *e* respectively.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
a	17.79	14.79	3.0	17.54	14.54	3.0	19.29	14.79	4.5
b	20.24	15.49	4.75	20.99	12.79	8.2	19.24	16.99	2.25
e	26.12	22.62	3.5	24.87	19.12	5.75	25.62	21.12	4.5
c	26.64	22.64	4.0	27.14	24.64	2.5	26.64	20.64	6.0

Table 3.6 C_o/N_o ratios at TOA and POF and failure margin for the urban fast multipath impairment

3.3.6 Rural fast multipath

Only the two systems, *e* and *f*, were tested subjectively under the rural fast multipath impairment condition. The failure characteristic curves for these systems are given in Fig. 3.38, 3.39, and 3.40 respectively for *Glockenspiel*, *Soprano* and *Clarinet*. Table 3.7 provides the C_o/N_o ratios at TOA and POF, as well as the failure margins (FM) for the two systems with the three test materials.

The failure characteristic curves for systems *e* and *f* for all three test materials are plotted in Fig. 3.41 and 3.42 respectively.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
f	5.96	2.46	3.5	6.46	2.46	4.0	6.96	2.46	4.5
e	31.62	25.12	6.5	30.62	24.62	6.0	31.62	24.62	7.0

Table 3.7 C_o/N_o ratios at TOA and POF and failure margin for the rural fast multipath impairment

3.3.7 Obstructed fast multipath

Three systems, *a*, *b*, and *c*, were tested subjectively under the obstructed fast multipath impairment condition. System *c* was only tested with the *Clarinet* test material. The failure characteristic curves for these systems are given in Fig. 3.43, 3.44, and 3.45 respectively for *Glockenspiel*, *Soprano* and *Clarinet*. Table 3.8 provides the C_o/N_o ratios at TOA and POF, as well as the failure margins (FM) for the three systems with the three test materials. Note that the entries under *Glockenspiel* and *Soprano* are intentionally left blank for system *c* since this system was not tested subjectively with these materials.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
a	19.54	15.54	4.0	19.04	15.54	3.5	19.54	15.54	4.0
b	20.49	15.99	4.5	20.49	15.99	4.5	20.49	16.49	4.0
c							25.14	22.64	2.5

Table 3.8 C_o/N_o ratios at TOA and POF and failure margin for the obstructed fast multipath impairment

The failure characteristic curves for each system with all three test materials are plotted in Fig. 3.46, 3.47, and 3.48 for systems *a*, *b*, and *c* respectively. Again, note that for system *c* (i.e. Fig. 3.48) a failure characteristic curve is only available for the *Clarinet* test material.

3.3.8 DAR lower 1st adjacent to DAR channel interference

Only system *i* was tested subjectively under the DAR lower 1st adjacent to DAR channel interference condition. The failure characteristic curves for this system are given in Fig. 3.49. All three source materials, *Glockenspiel*, *Soprano* and *Clarinet*, are plotted together in this one figure. The C_o/N_o ratios at TOA and POF, as well as the failure margins (FM) are given below in Table 3.9 for the three test materials.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
i	31.63	30.13	1.5	31.13	30.63	0.5	31.13	31.13	0.0

Table 3.9 D/U values at TOA and POF and failure margin for the DAR lower 1st adjacent to DAR channel interference

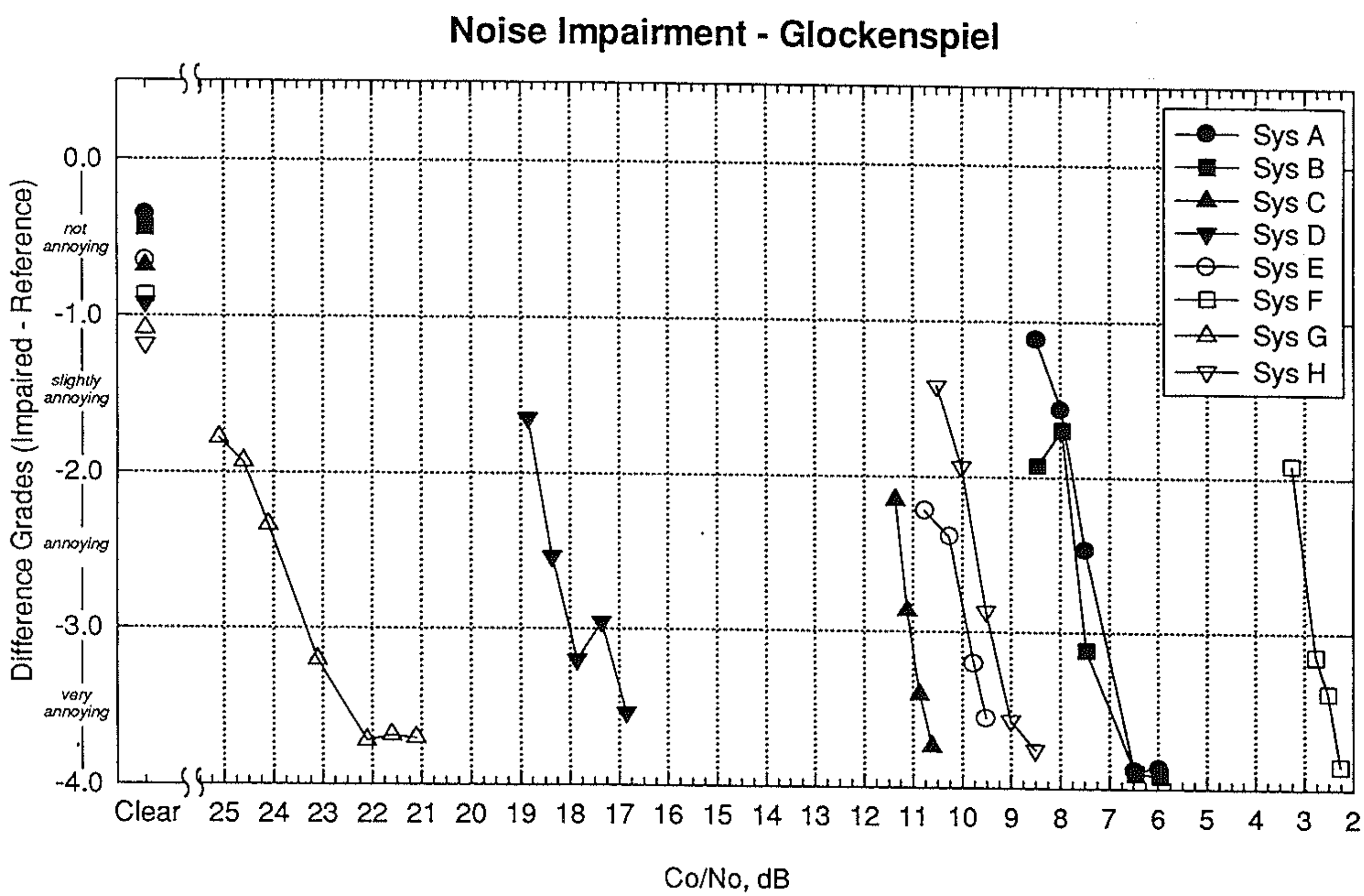


Figure 3.1. Failure characteristic curves of DAR systems *a* through *h* tested with the gaussian noise impairment and the Glockenspiel test material. (DAR laboratory was unable to certify digital power levels for system *g*)

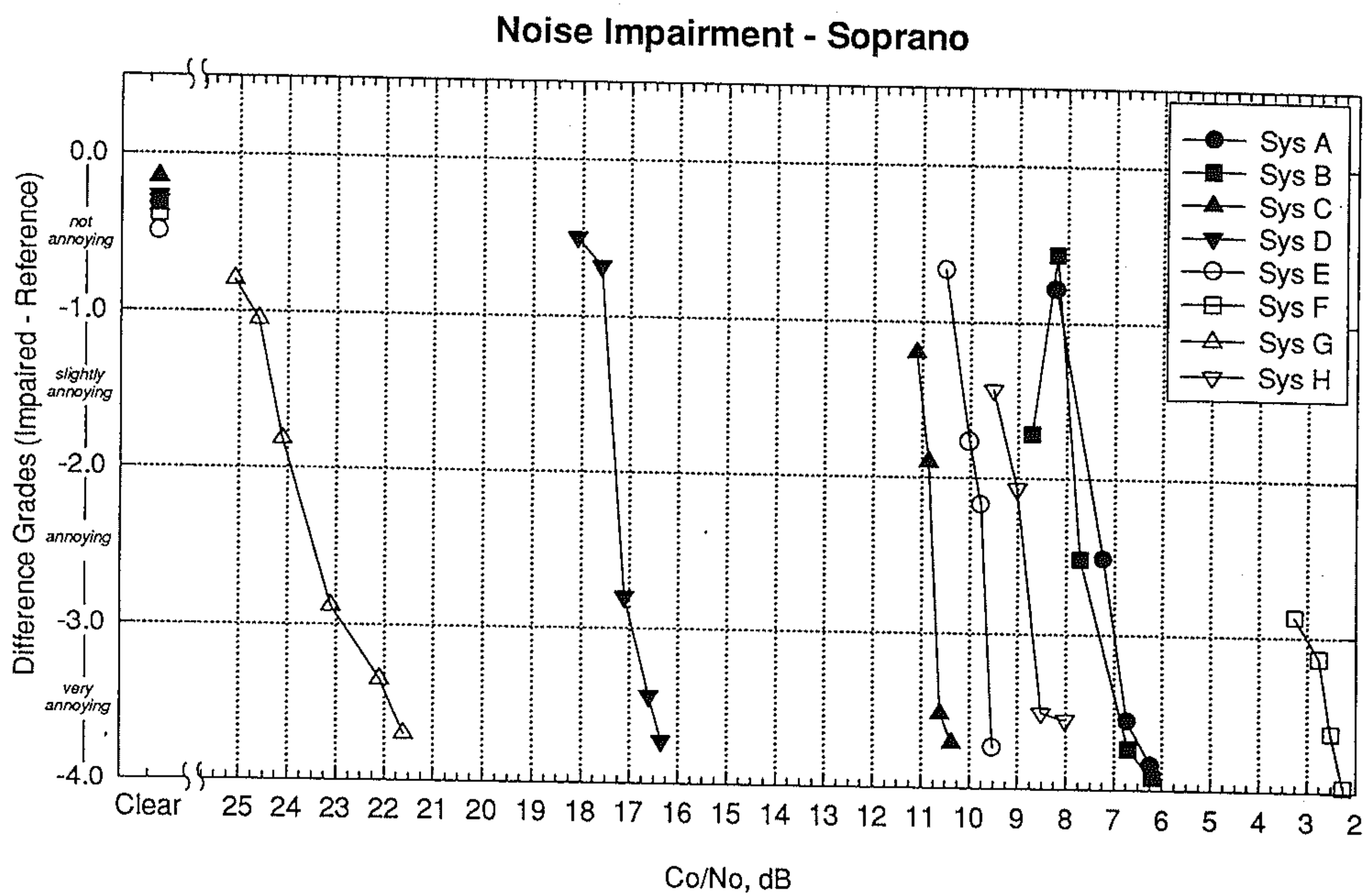


Figure 3.2. Failure characteristic curves of DAR systems *a* through *h* tested with the gaussian noise impairment and the Soprano test material. (DAR laboratory was unable to certify digital power levels for system *g*)

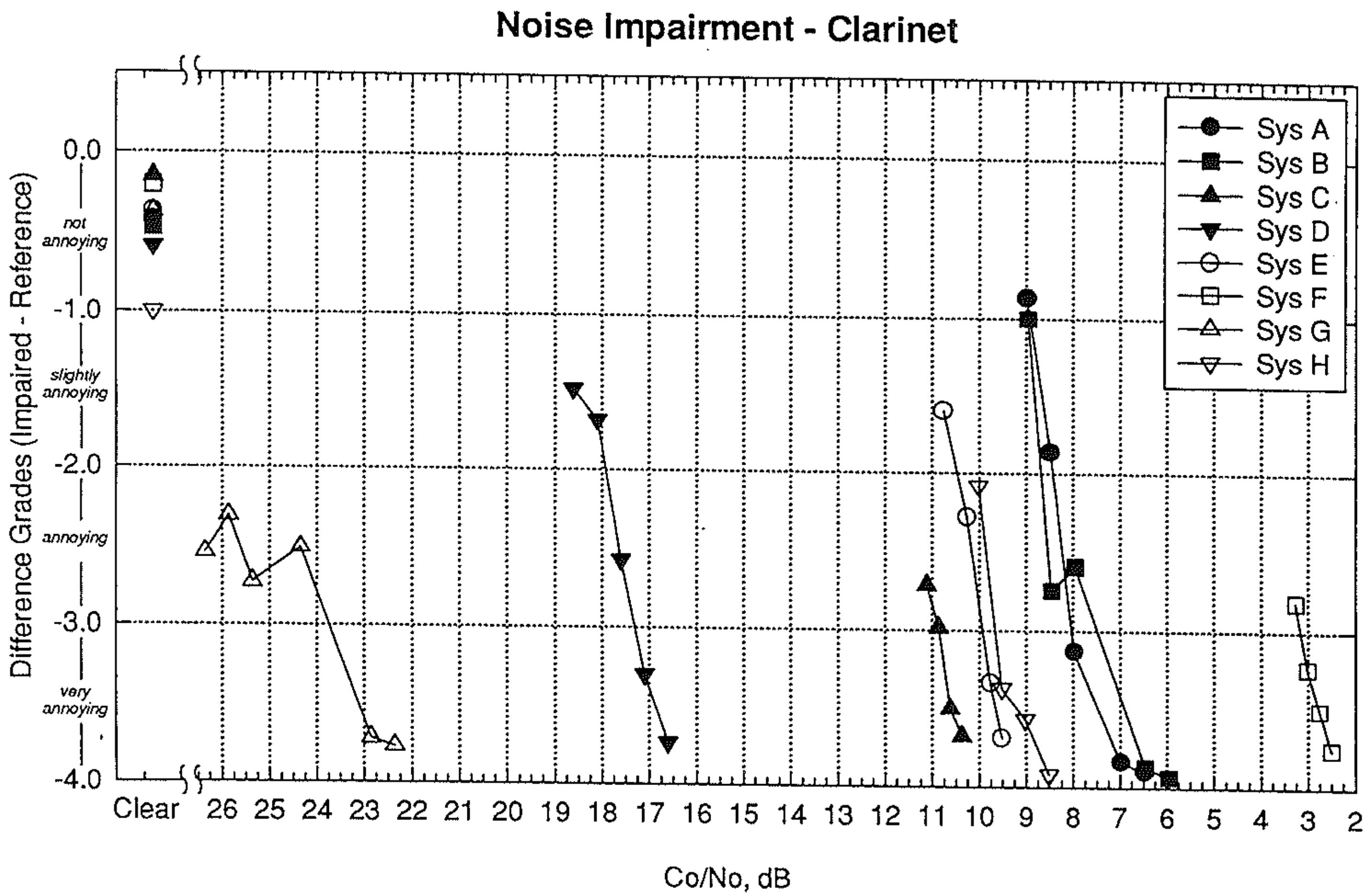


Figure 3.3. Failure characteristic curves of DAR systems *a* through *h* tested with the gaussian noise impairment and the Clarinet test material. (DAR laboratory was unable to certify digital power levels for system *g*)

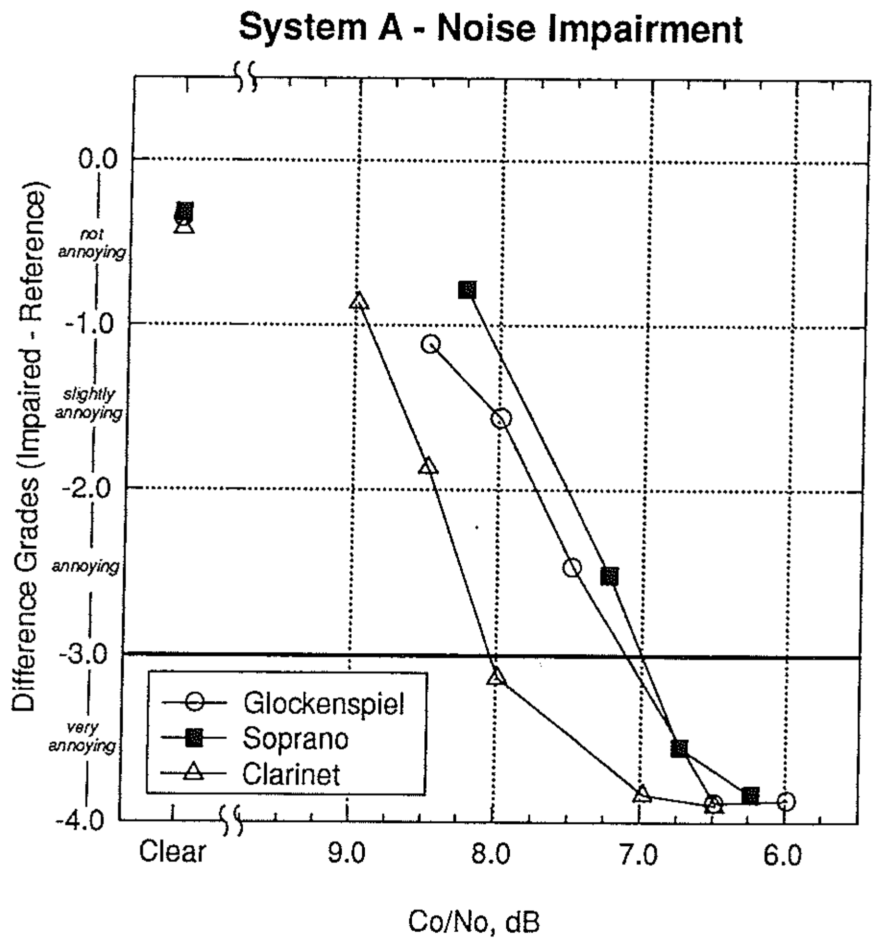


Figure 3.4. Failure characteristic curves for system *a* for all three test materials with the gaussian noise impairment.

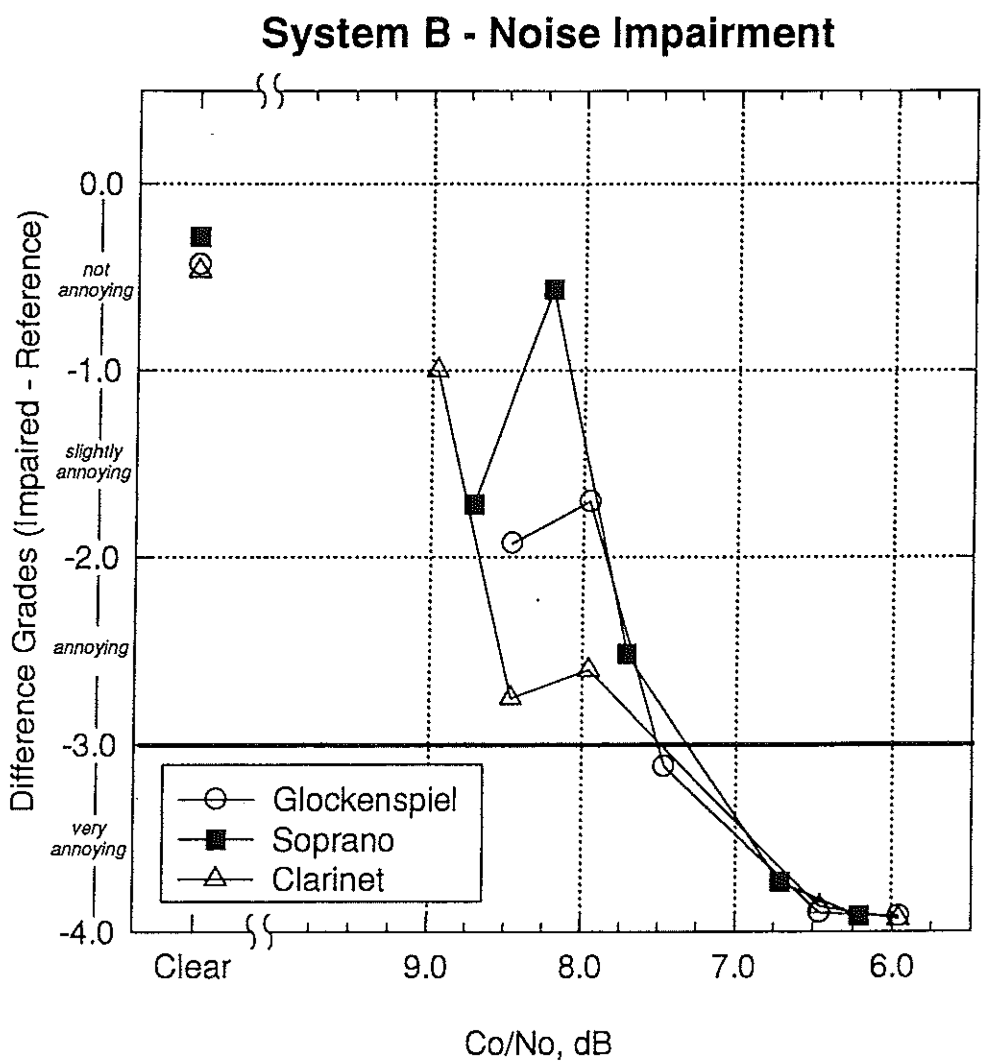


Figure 3.5. Failure characteristic curves for system *b* for all three test materials with the gaussian noise impairment.

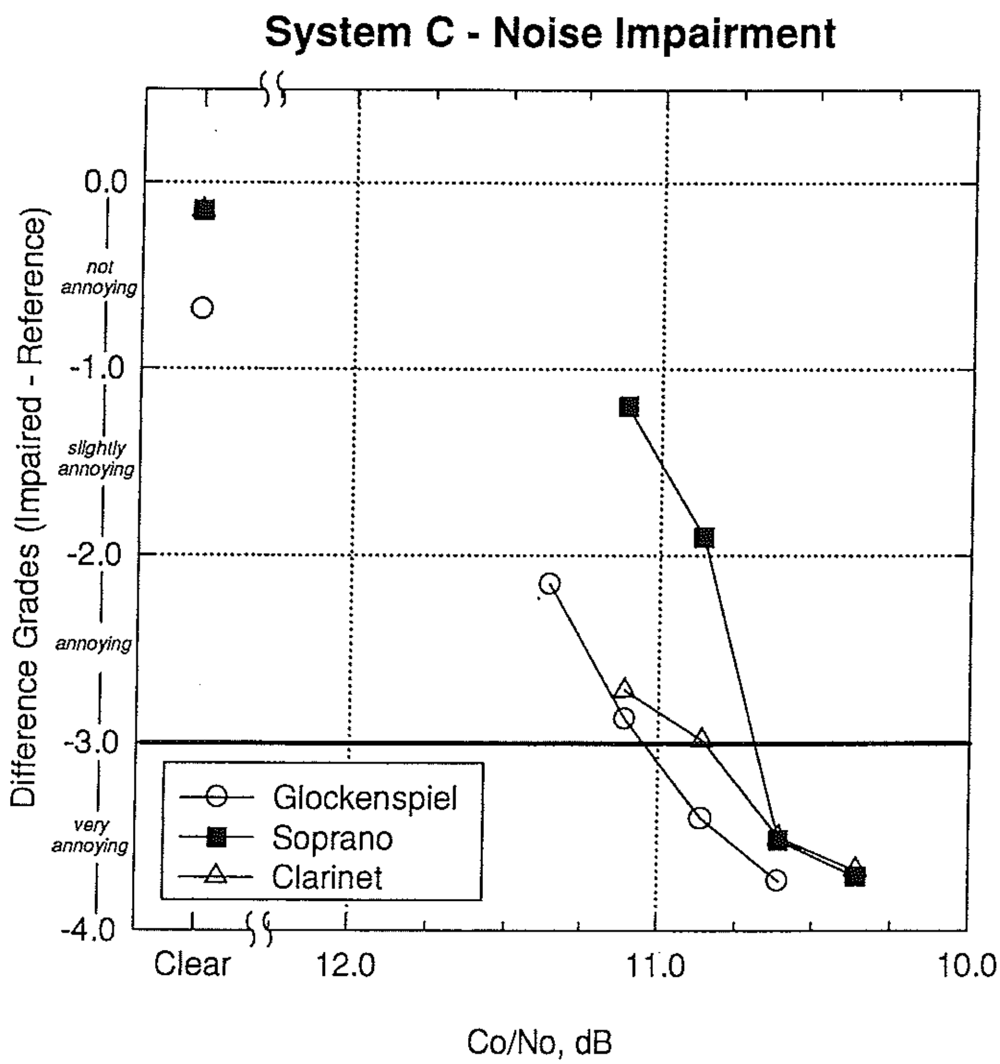


Figure 3.6. Failure characteristic curves for system *c* for all three test materials with the gaussian noise impairment.

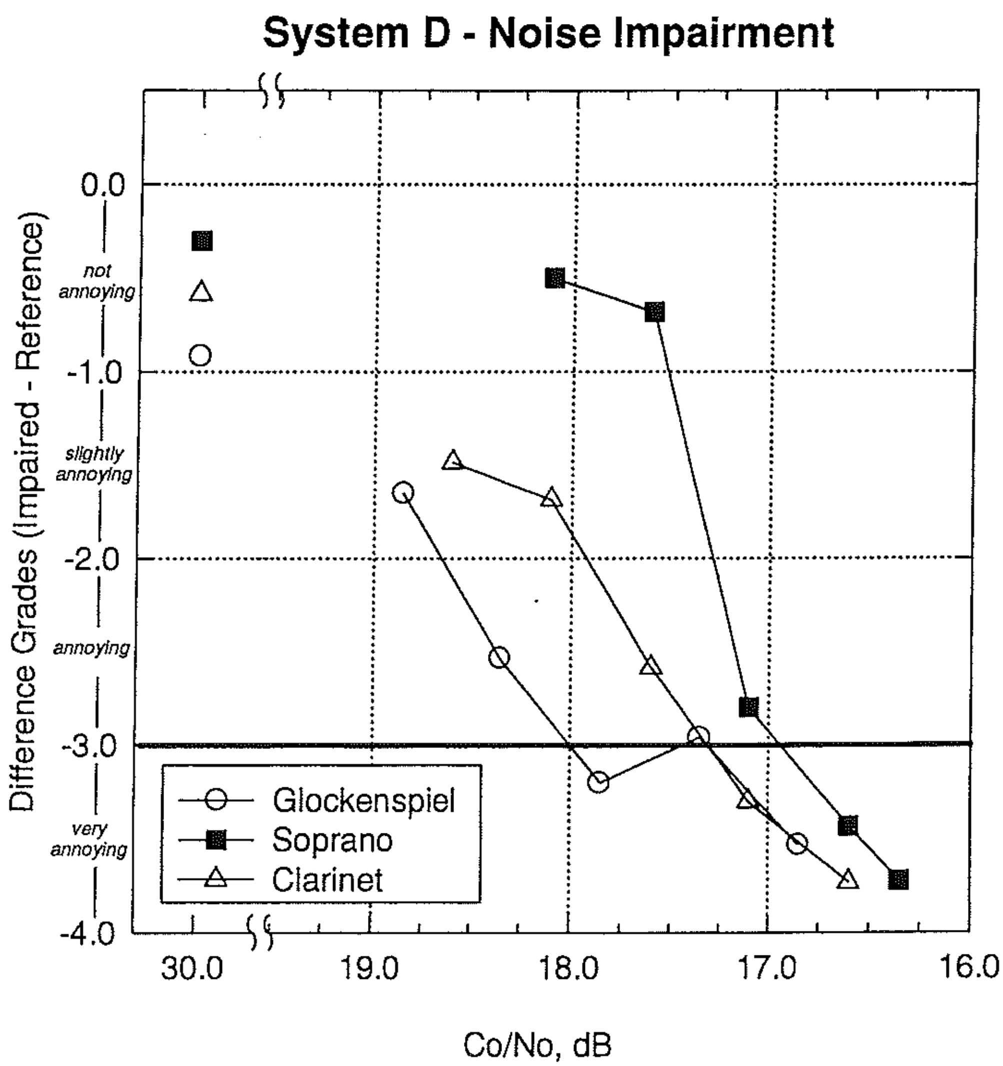


Figure 3.7. Failure characteristic curves for system *d* for all three test materials with the gaussian noise impairment.

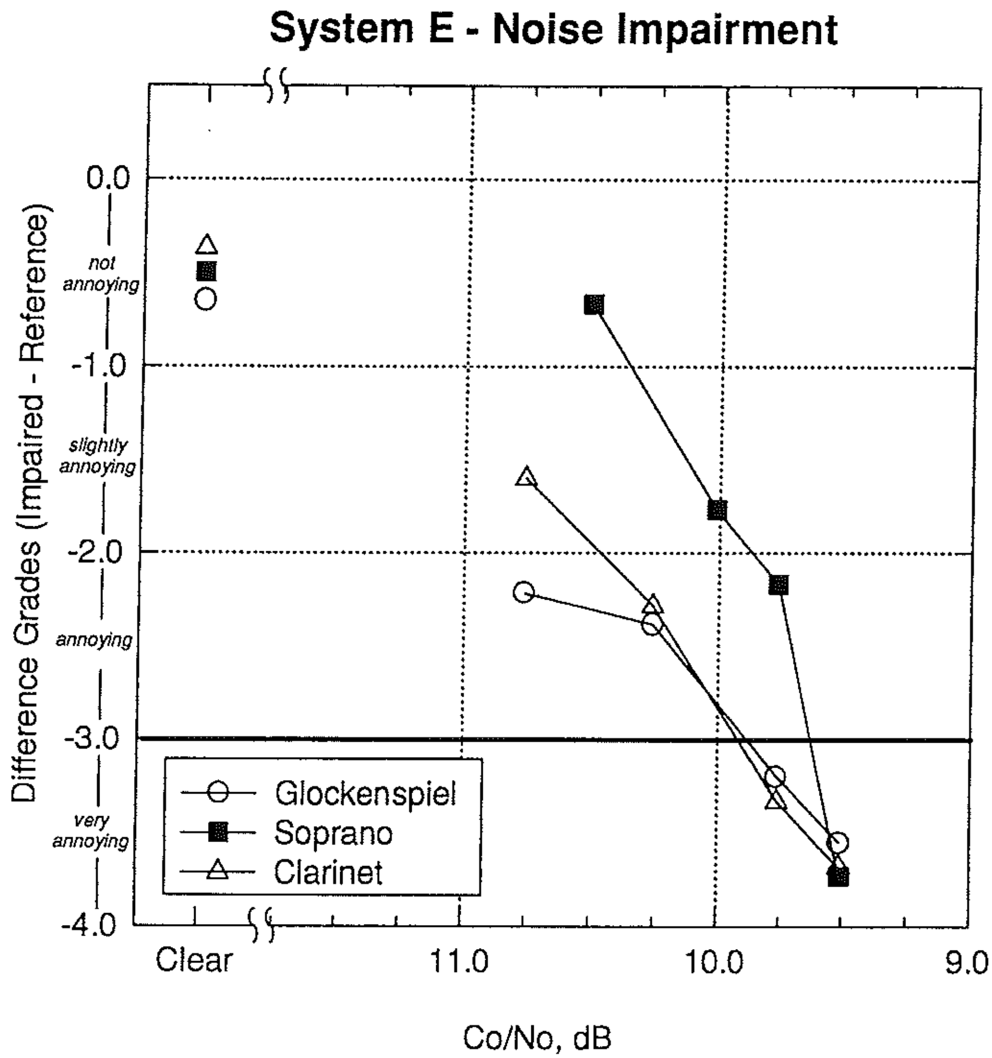


Figure 3.8. Failure characteristic curves for system *e* for all three test materials with the gaussian noise impairment.

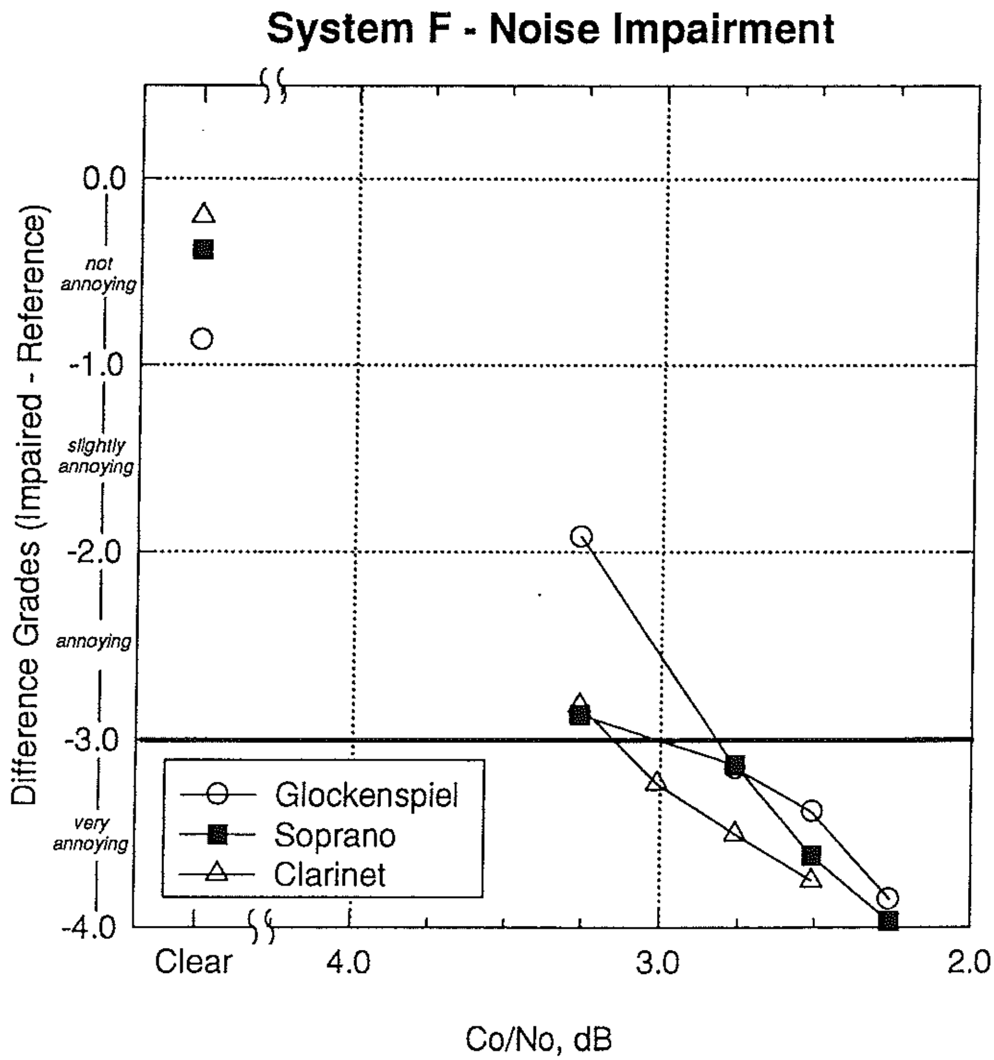


Figure 3.9. Failure characteristic curves for system *f* for all three test materials with the gaussian noise impairment.

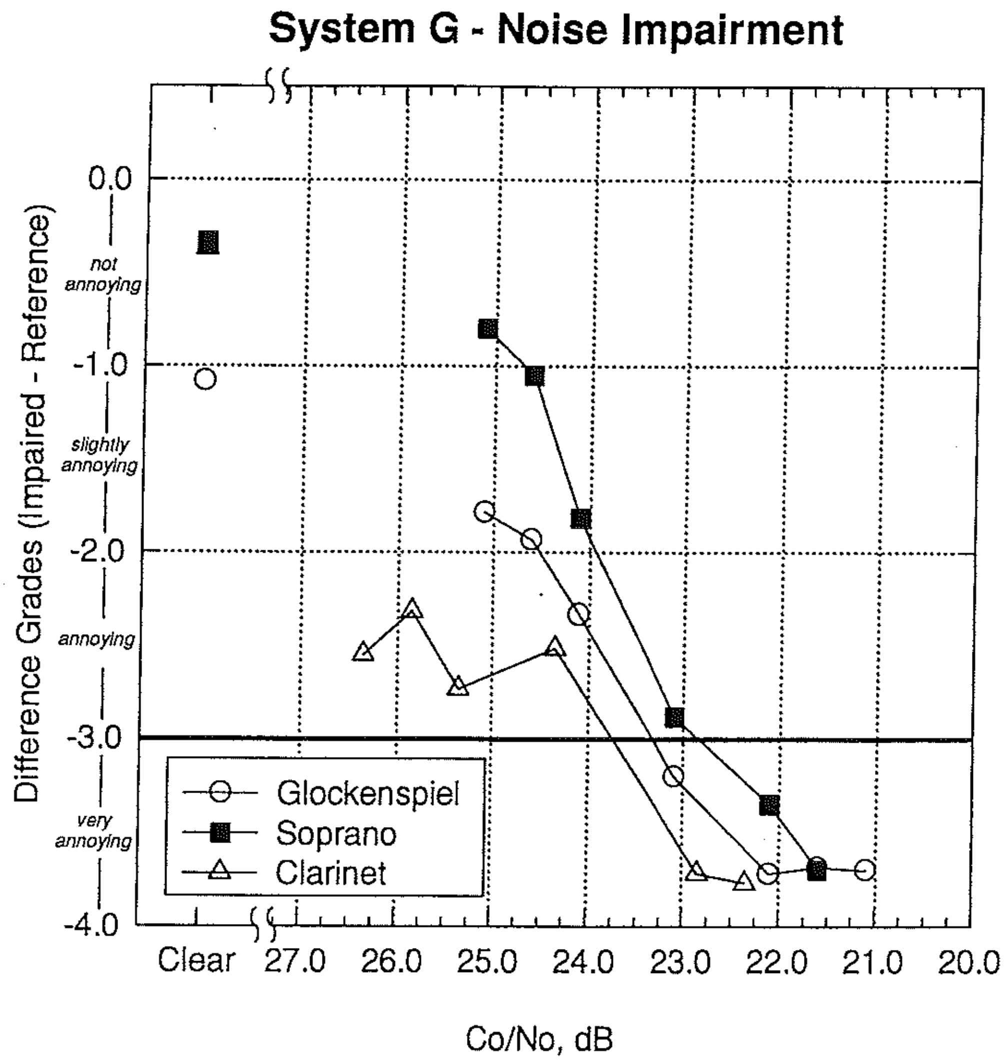


Figure 3.10. Failure characteristic curves for system *g* for all three test materials with the gaussian noise impairment. (DAR laboratory was unable to certify digital power levels for this system)

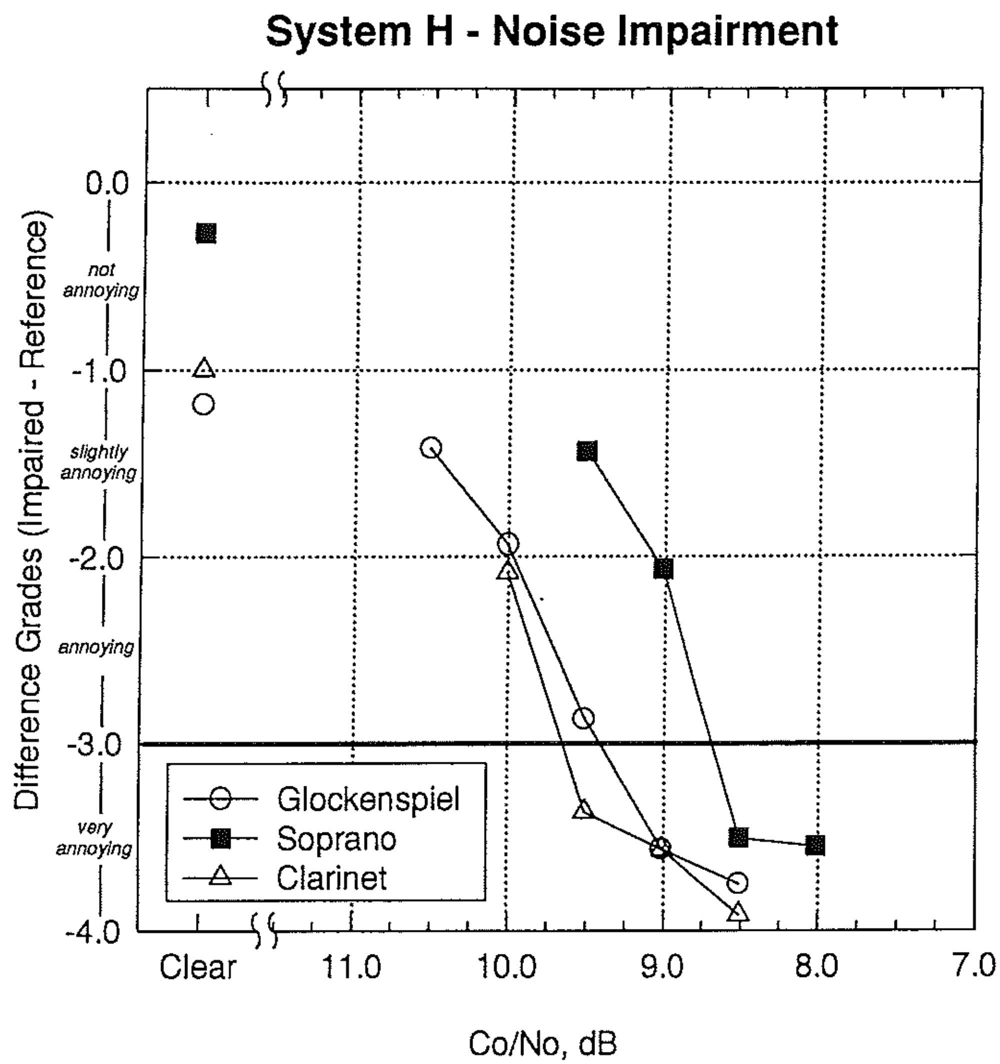


Figure 3.11. Failure characteristic curves for system *h* for all three test materials with the gaussian noise impairment.

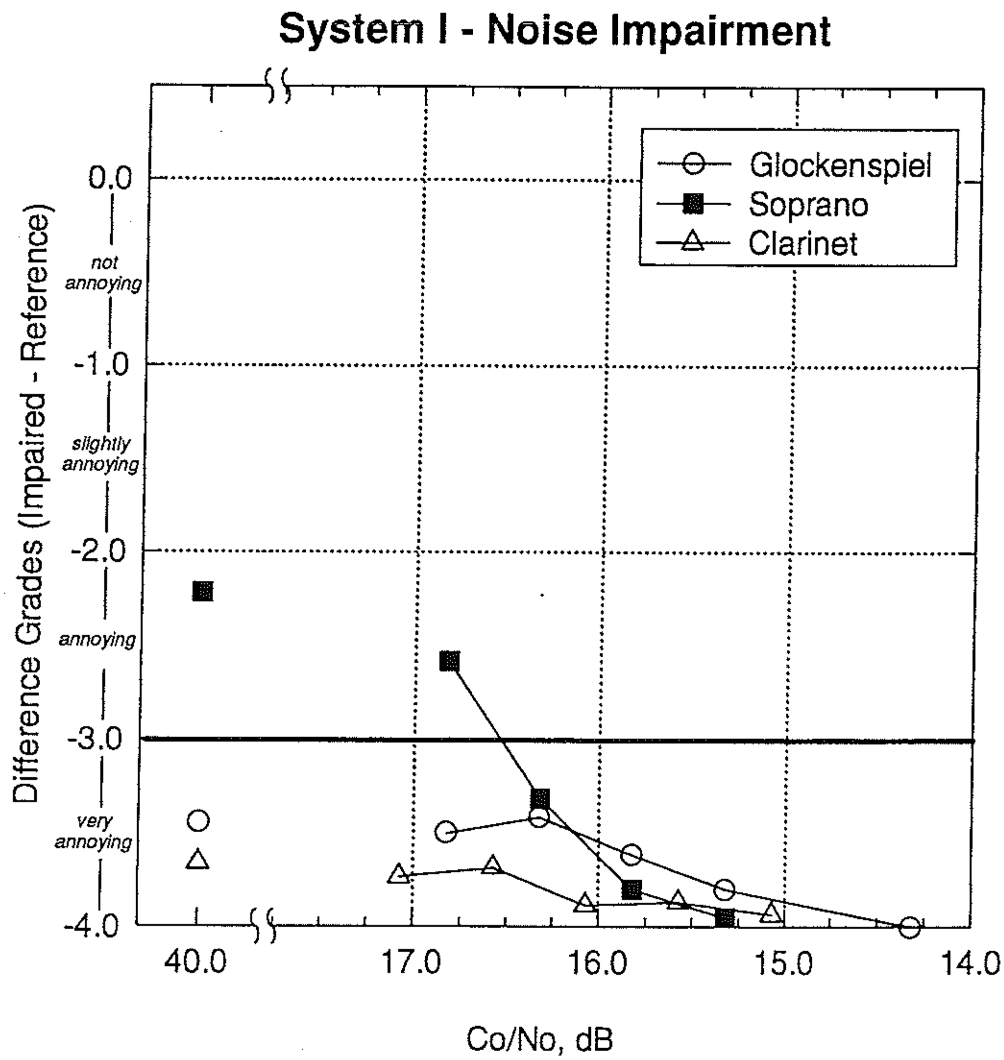


Figure 3.12. Failure characteristic curves for system *i* for all three test materials with the gaussian noise impairment. (DAR laboratory was unable to certify digital power levels for this system)

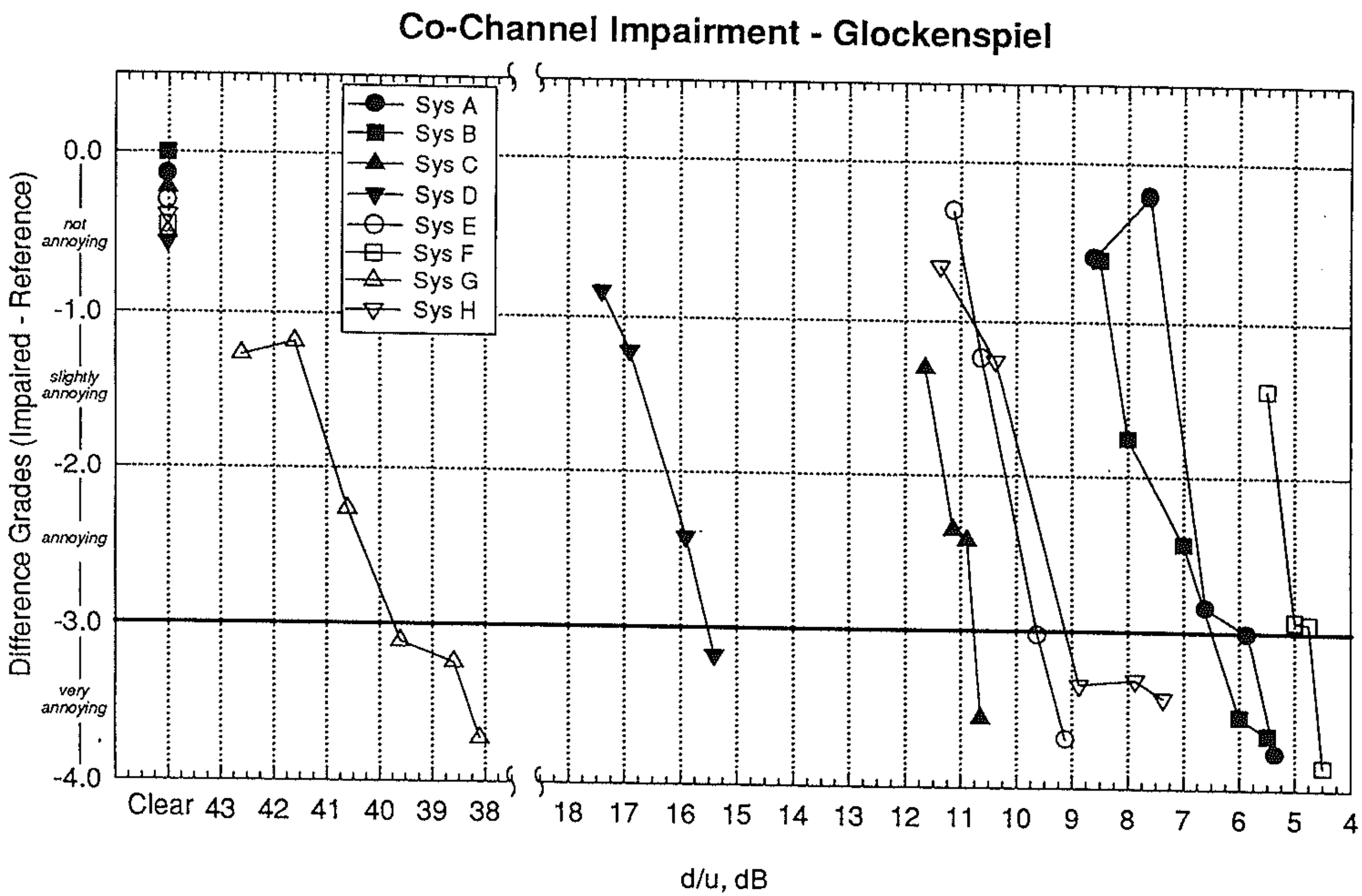


Figure 3.13. Failure characteristic curves of DAR systems *a* through *h* tested with co-channel interference and the Glockenspiel test material. (DAR laboratory was unable to certify digital power levels for system *g*)

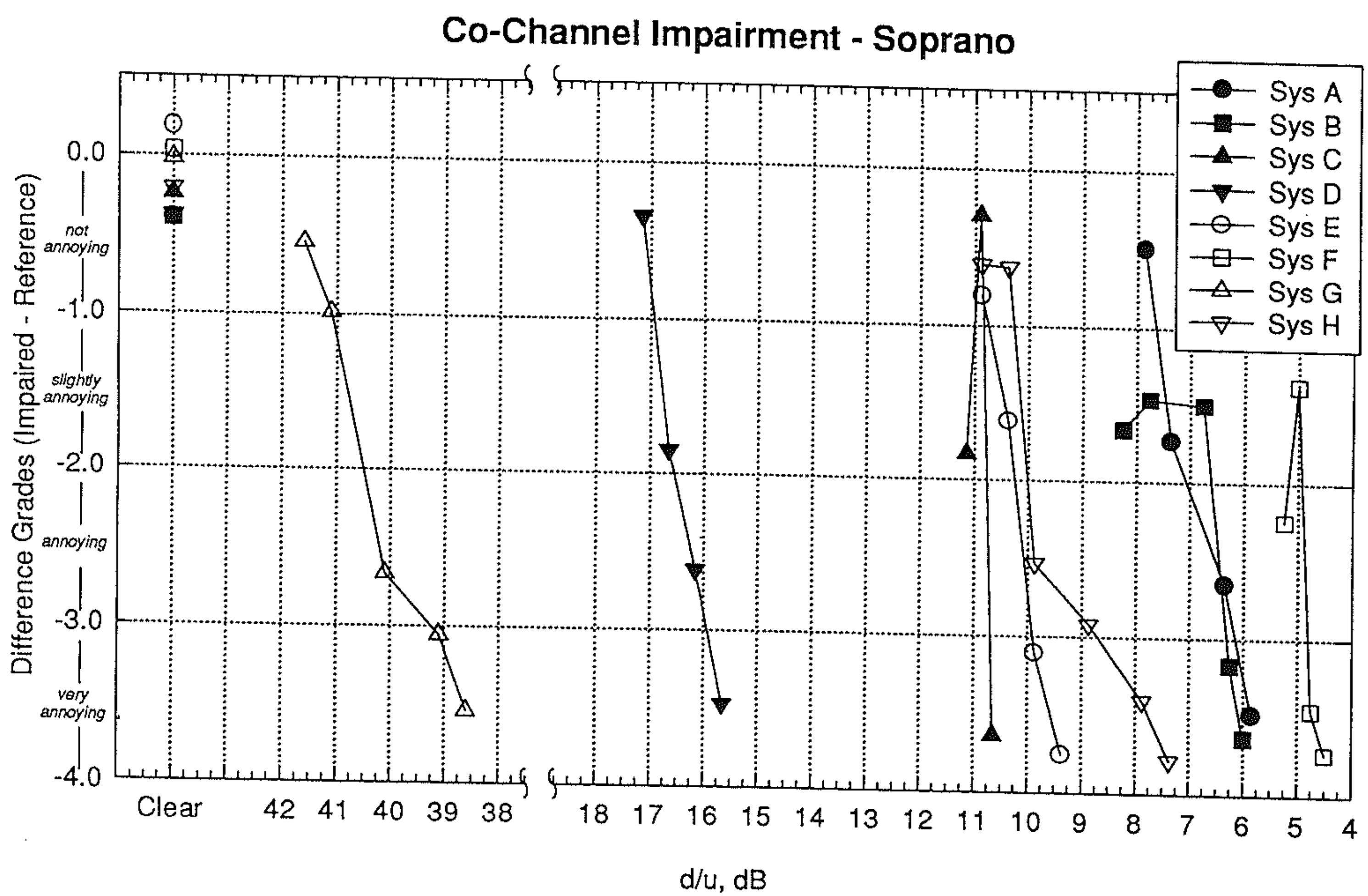


Figure 3.14. Failure characteristic curves of DAR systems *a* through *h* tested with co-channel interference and the Soprano test material. (DAR laboratory was unable to certify digital power levels for system *g*)

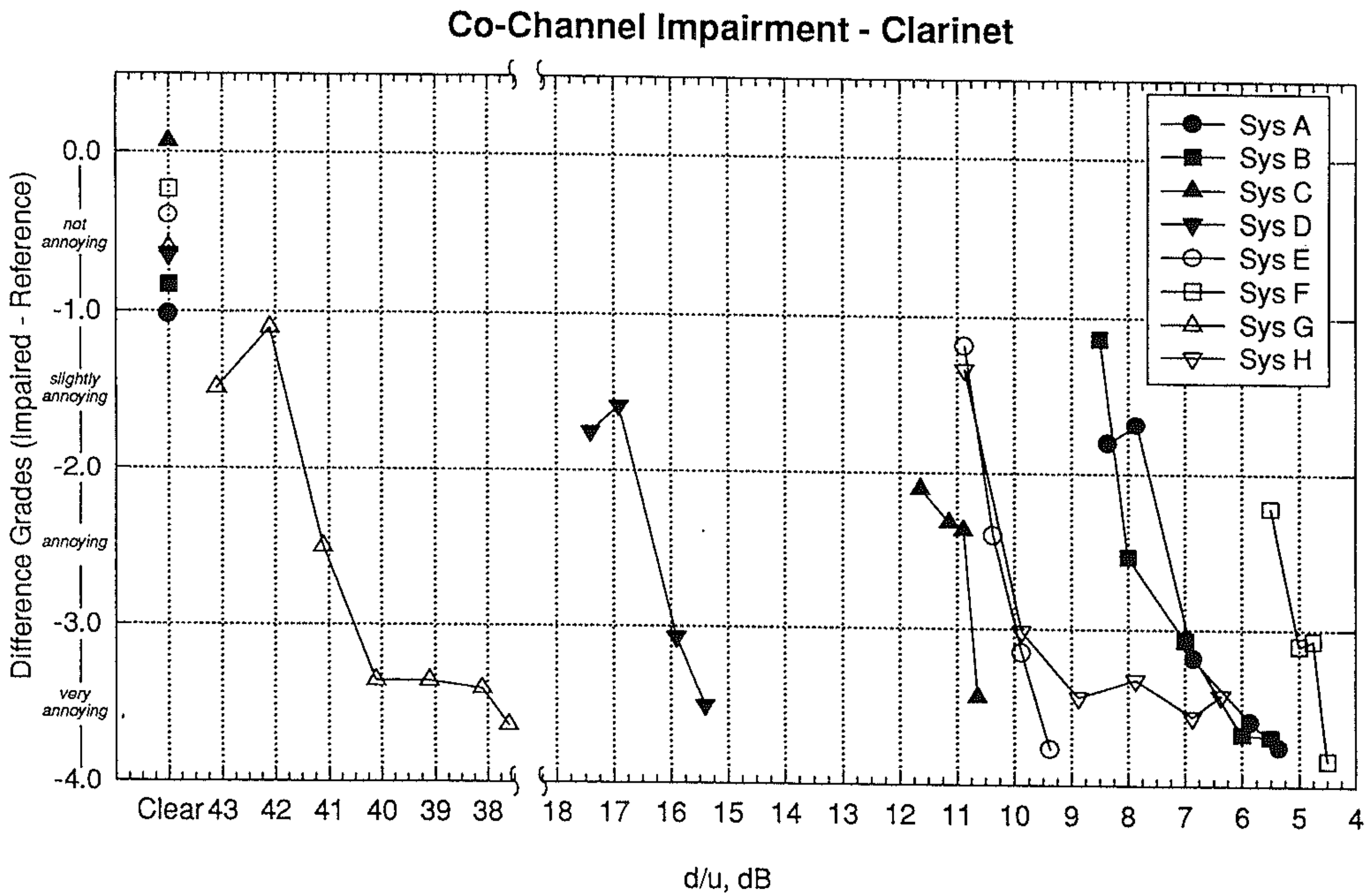


Figure 3.15. Failure characteristic curves of DAR systems *a* through *h* tested with co-channel interference and the Clarinet test material. (DAR laboratory was unable to certify digital power levels for system *g*)

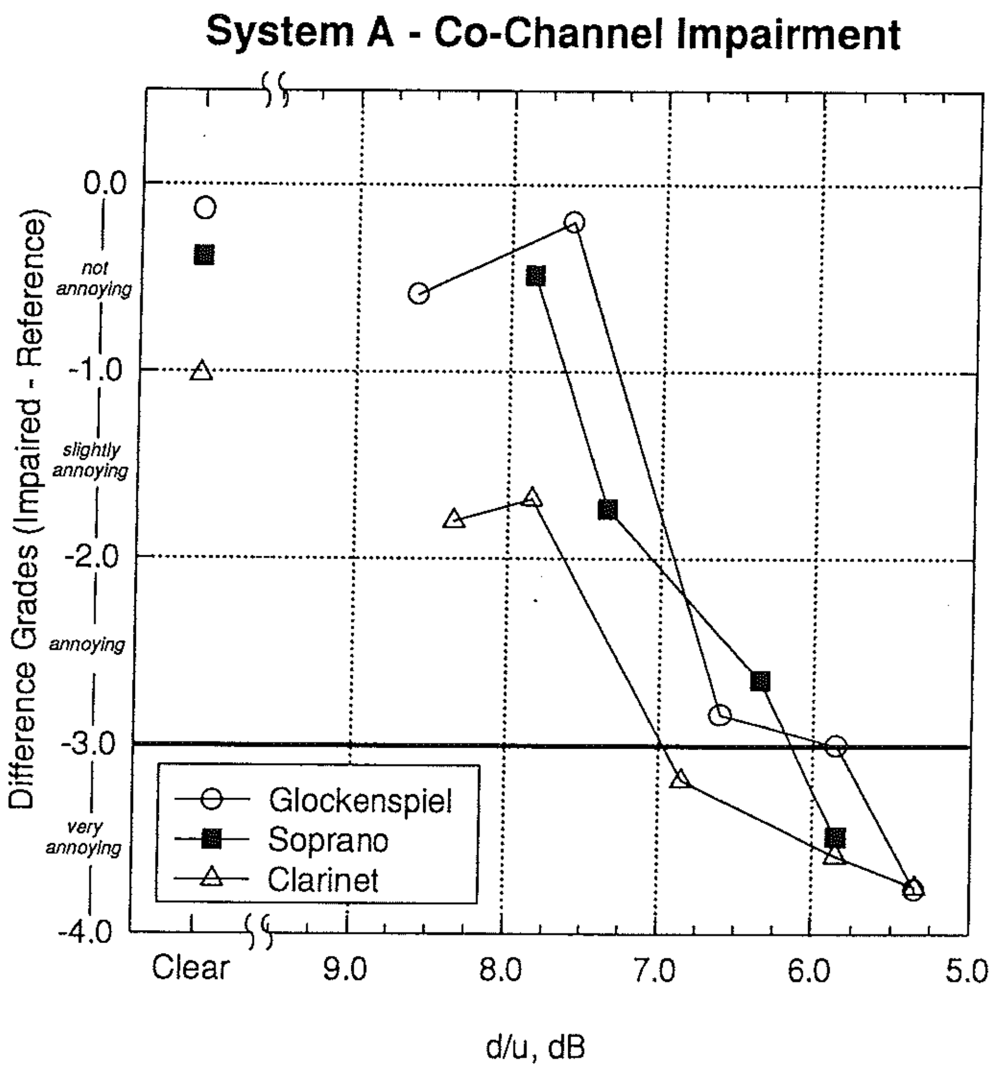


Figure 3.16. Failure characteristic curves for system *a* for all three test materials with the co-channel interference.

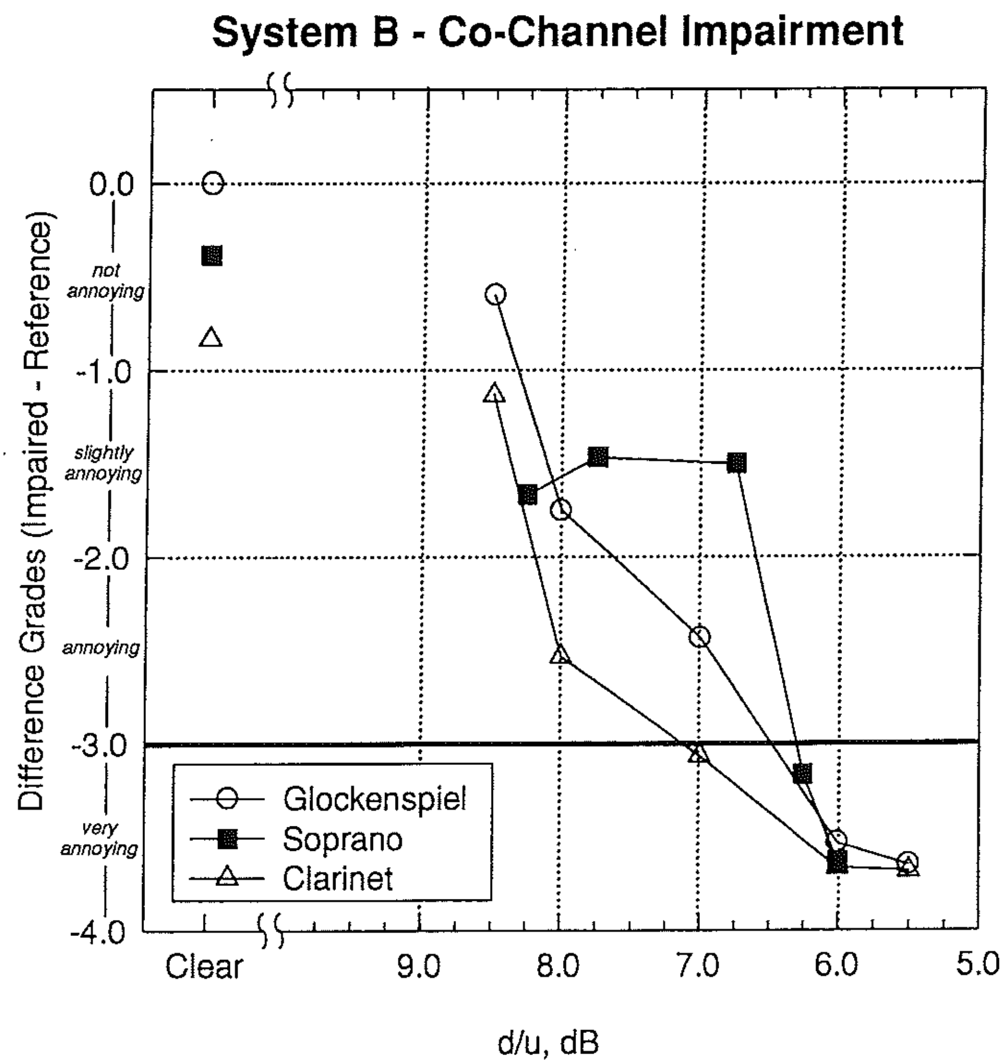


Figure 3.17. Failure characteristic curves for system *b* for all three test materials with the co-channel interference.

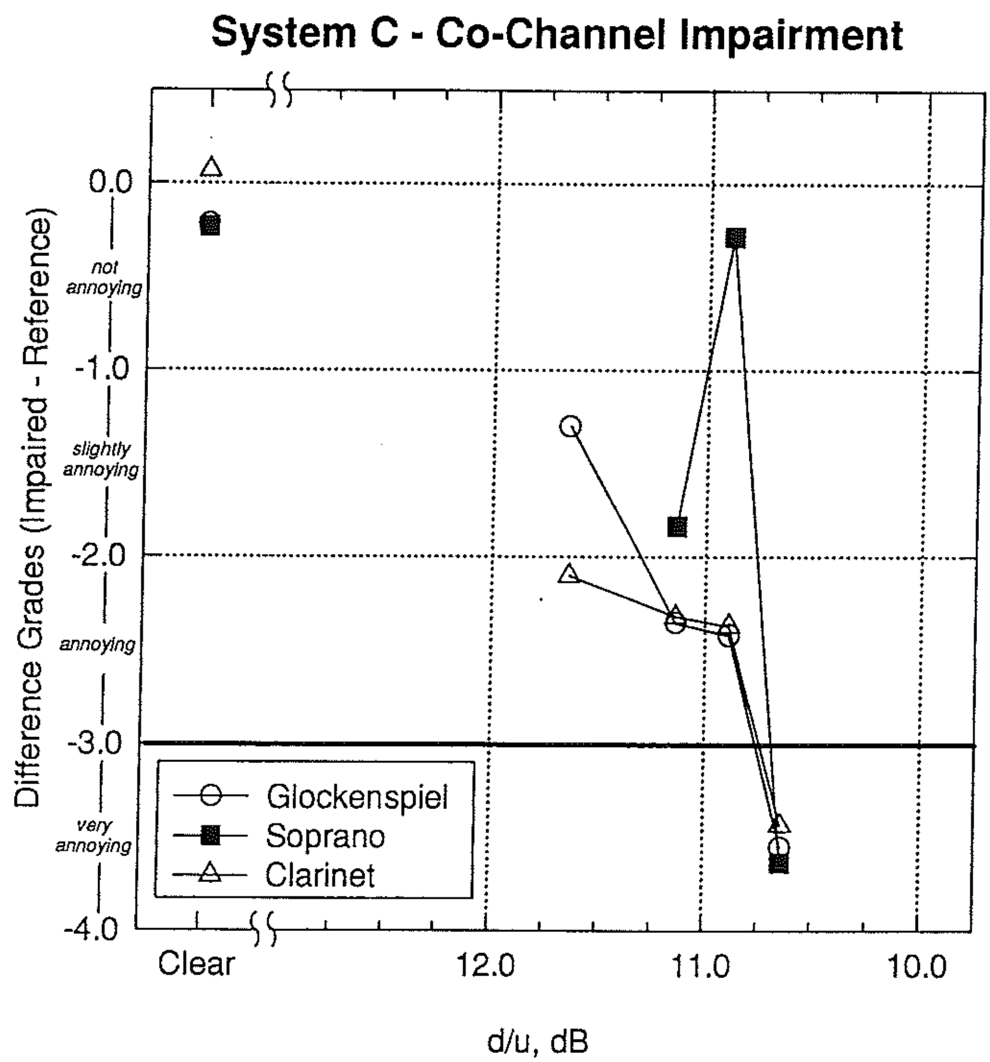


Figure 3.18. Failure characteristic curves for system *c* for all three test materials with the co-channel interference.

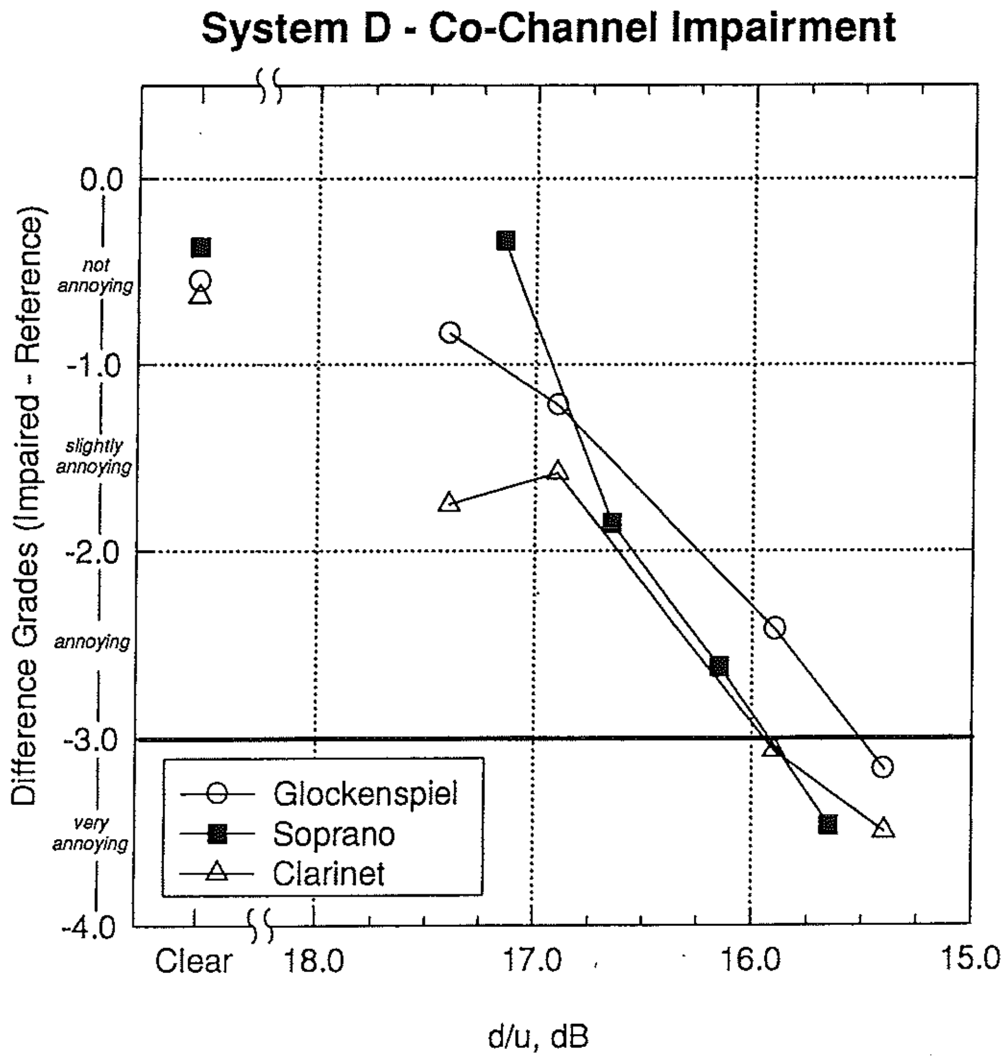


Figure 3.19. Failure characteristic curves for system *d* for all three test materials with the co-channel interference.

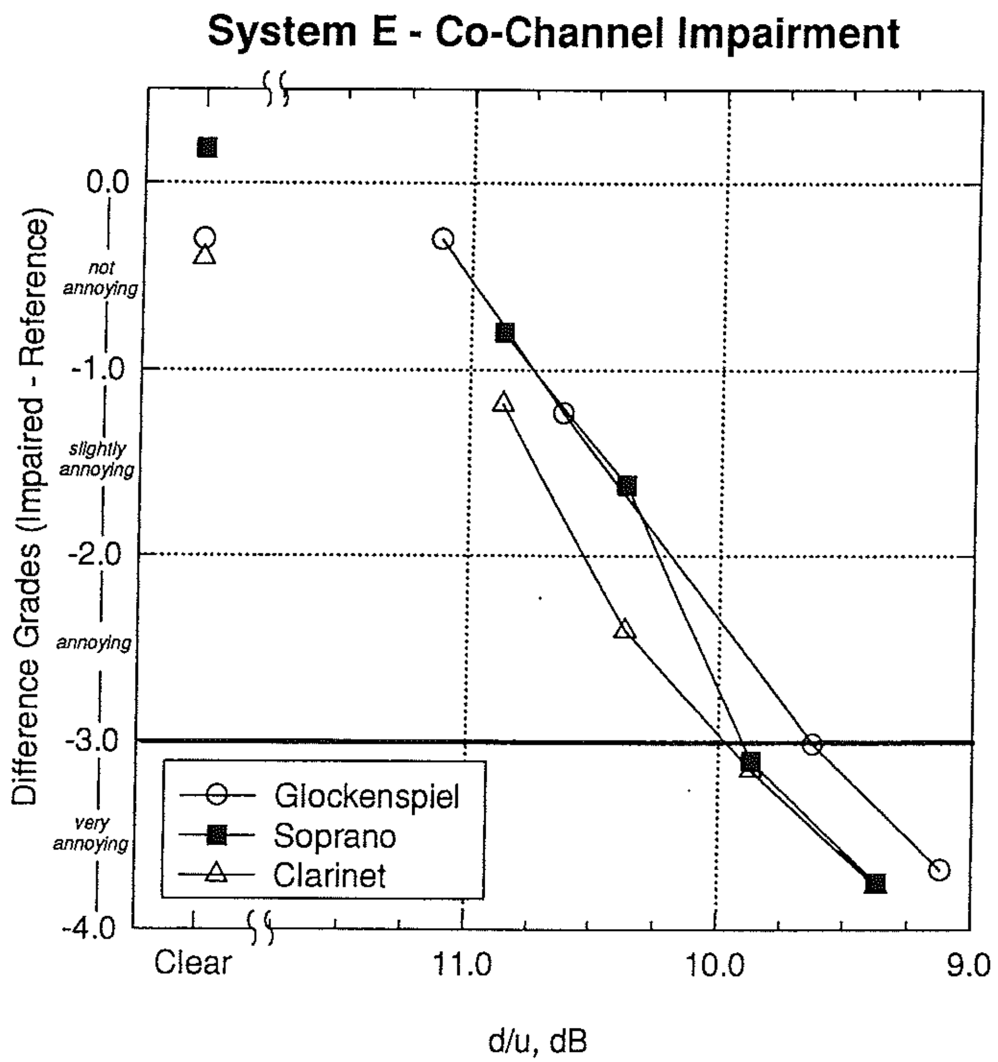


Figure 3.20. Failure characteristic curves for system *e* for all three test materials with the co-channel interference.

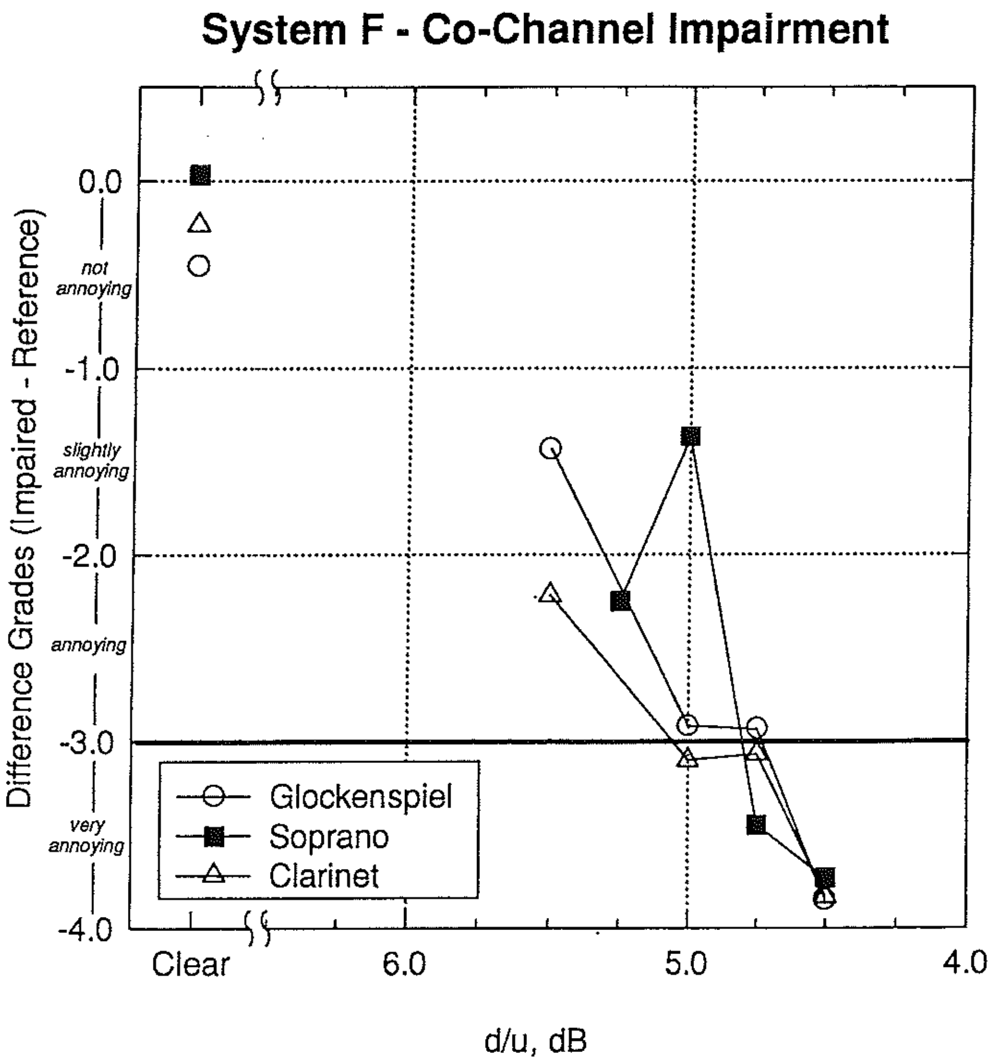


Figure 3.21. Failure characteristic curves for system *f* for all three test materials with the co-channel interference.

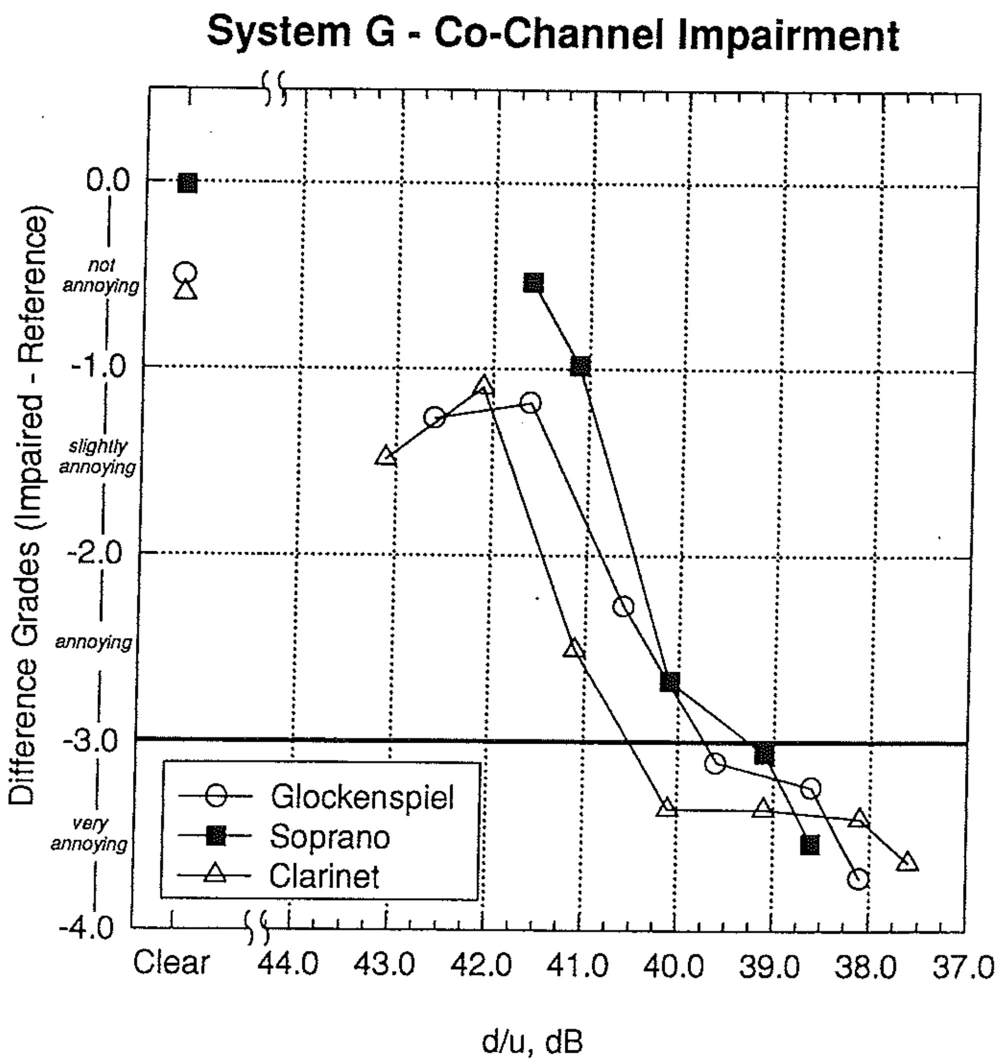


Figure 3.22. Failure characteristic curves for system *g* for all three test materials with the co-channel interference. (DAR laboratory was unable to certify digital power levels for this system)

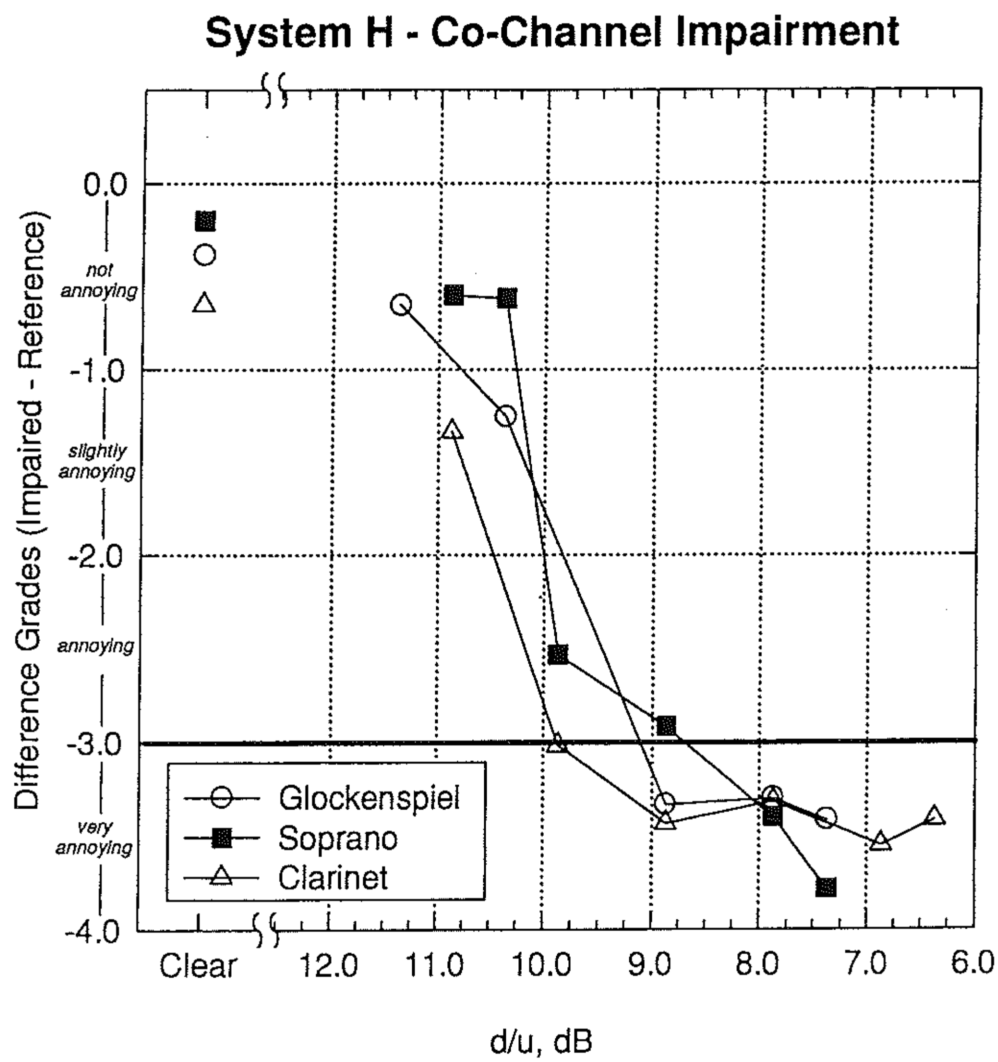


Figure 3.23. Failure characteristic curves for system *h* for all three test materials with the co-channel interference.

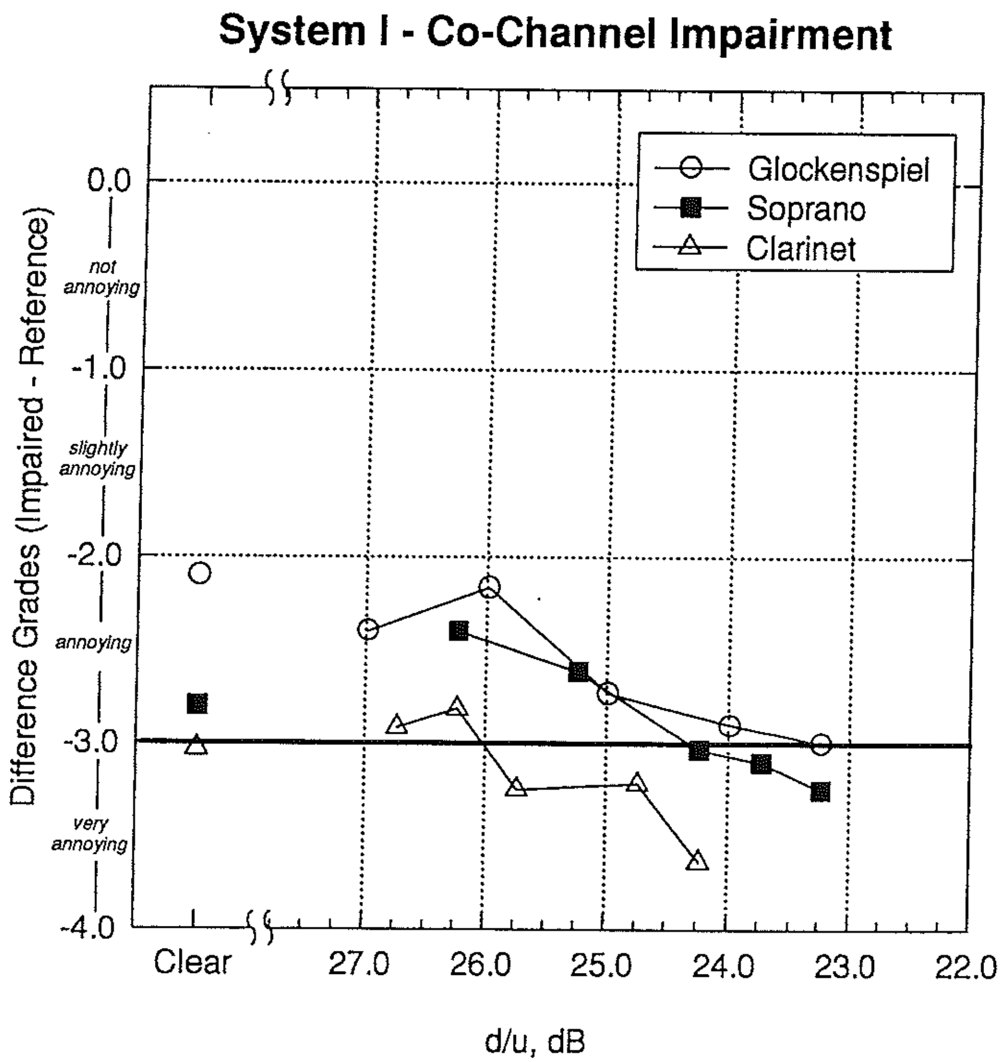


Figure 3.24. Failure characteristic curves for system *I* for all three test materials with the co-channel interference. (DAR laboratory was unable to certify digital power levels for this system)

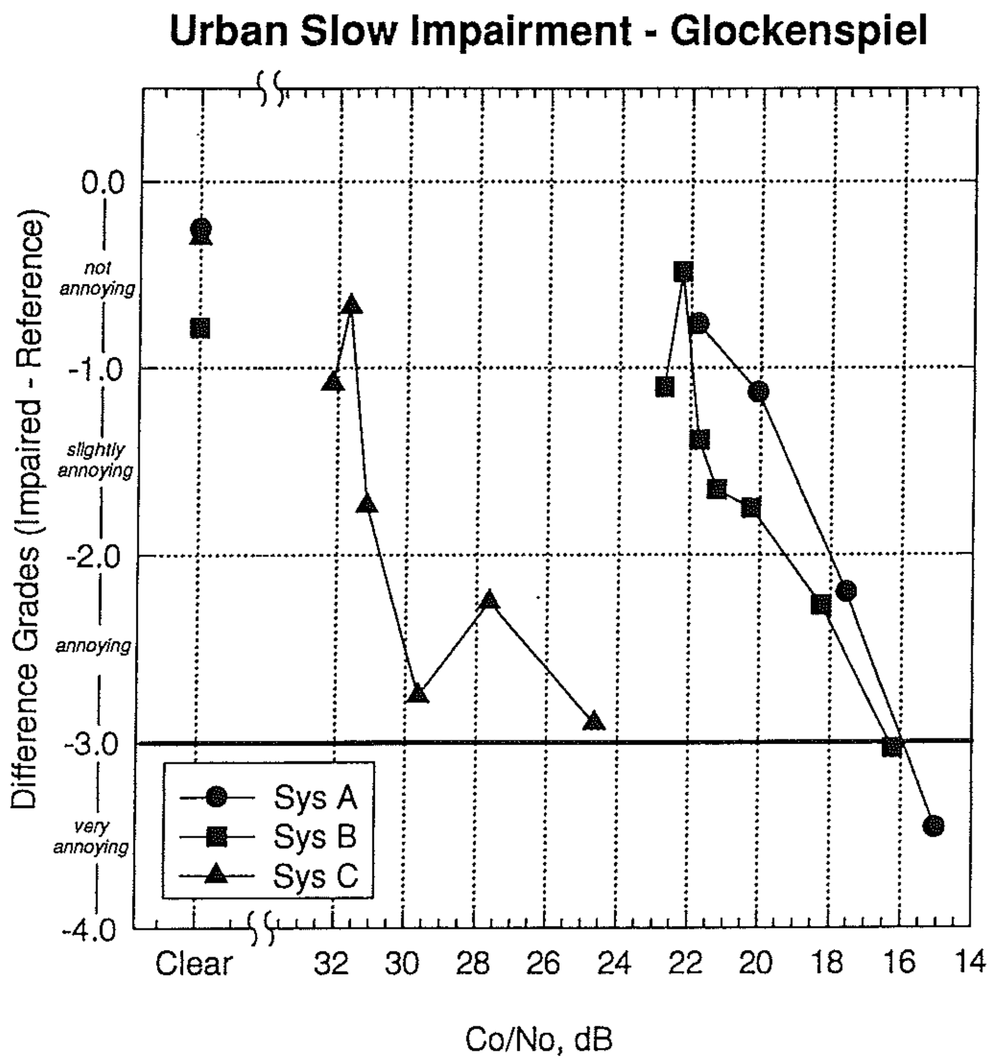


Figure 3.25. Failure characteristic curves of the three DAR systems tested with the Urban Slow multipath impairment and the Glockenspiel test material.

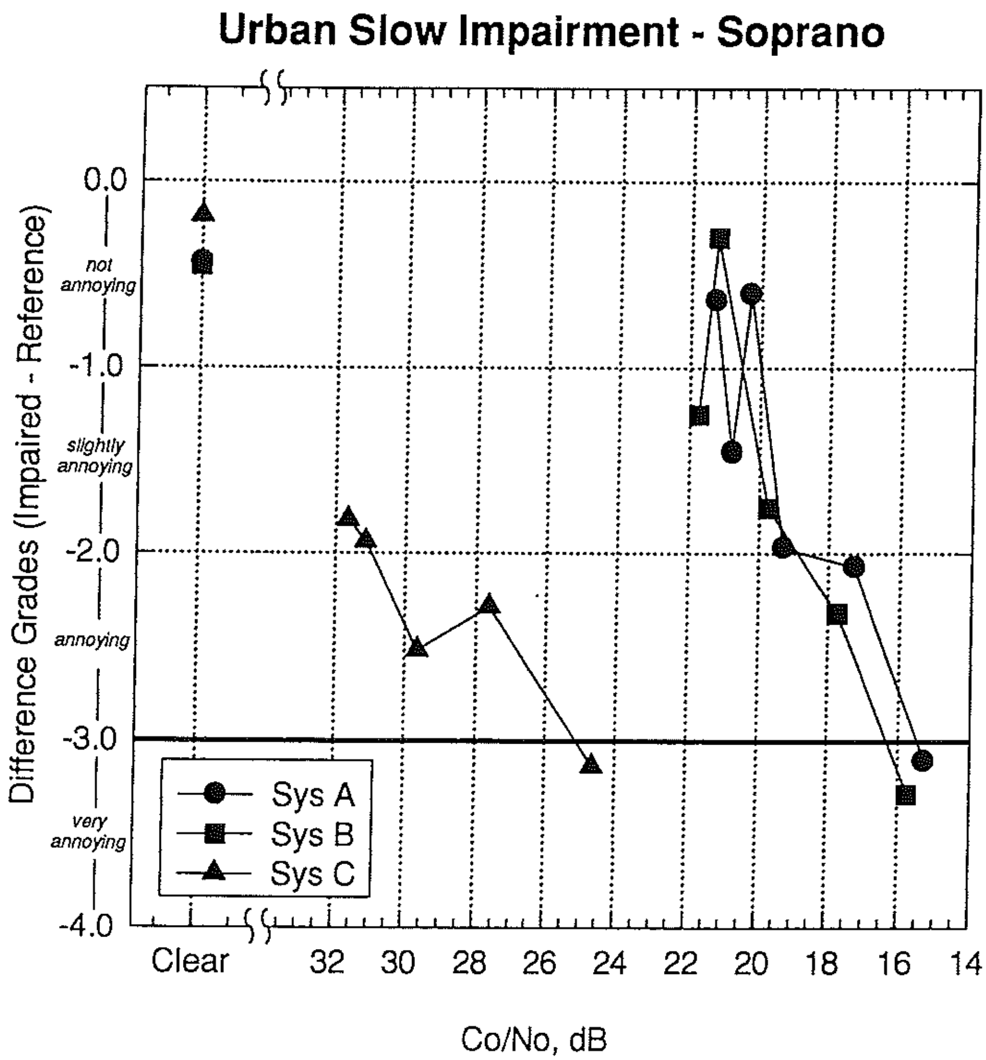


Figure 3.26. Failure characteristic curves of the three DAR systems tested with the Urban Slow multipath impairment and the Soprano test material.

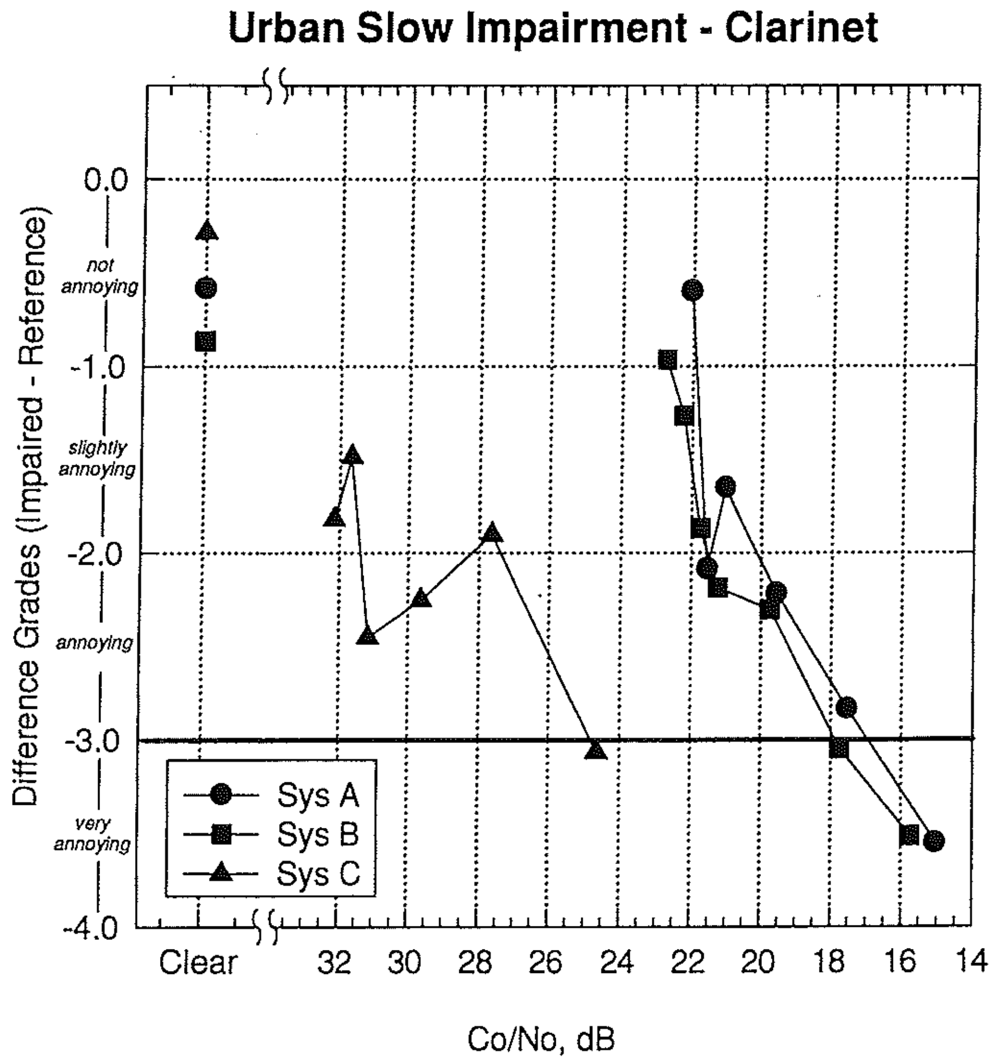


Figure 3.27. Failure characteristic curves of the three DAR systems tested with the Urban Slow multipath impairment and the Clarinet test material.

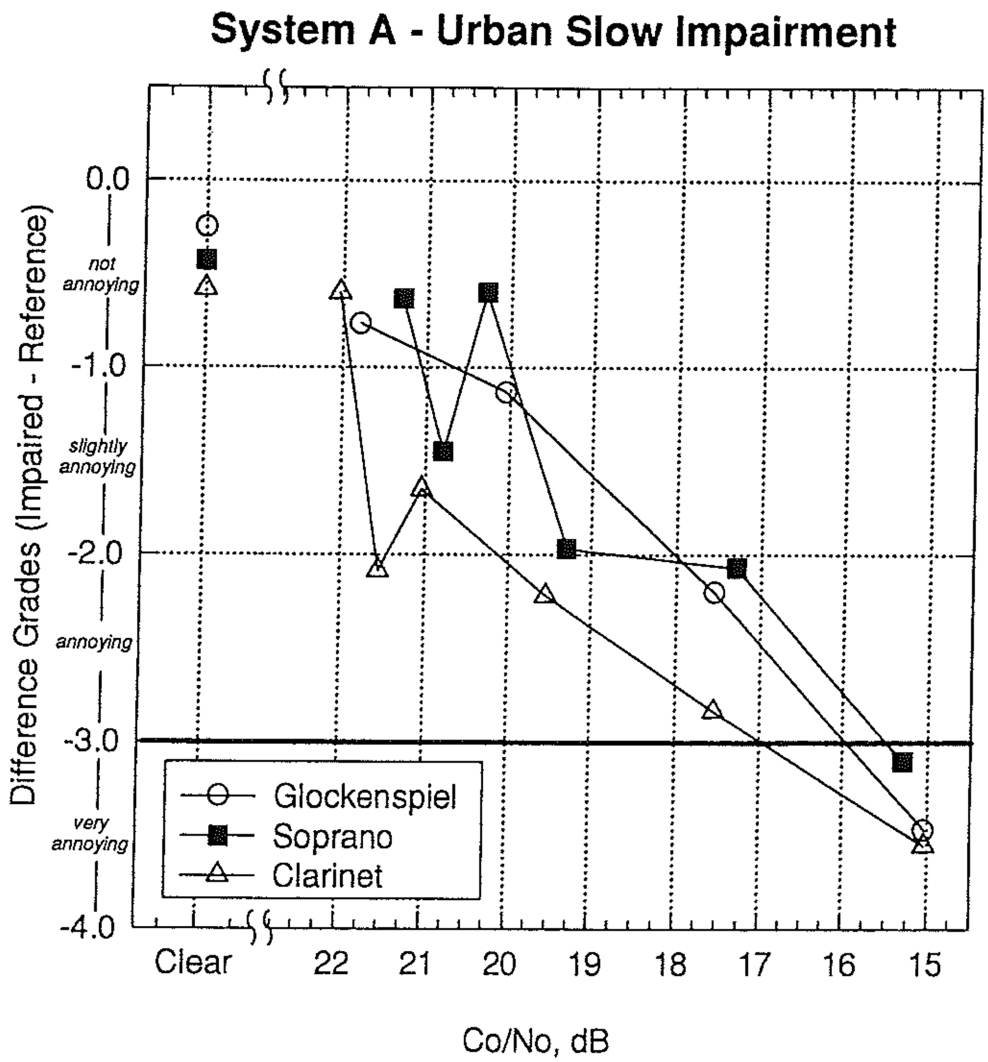


Figure 3.28. Failure characteristic curves for system *a* for all three test materials with the Urban Slow Multipath impairment.

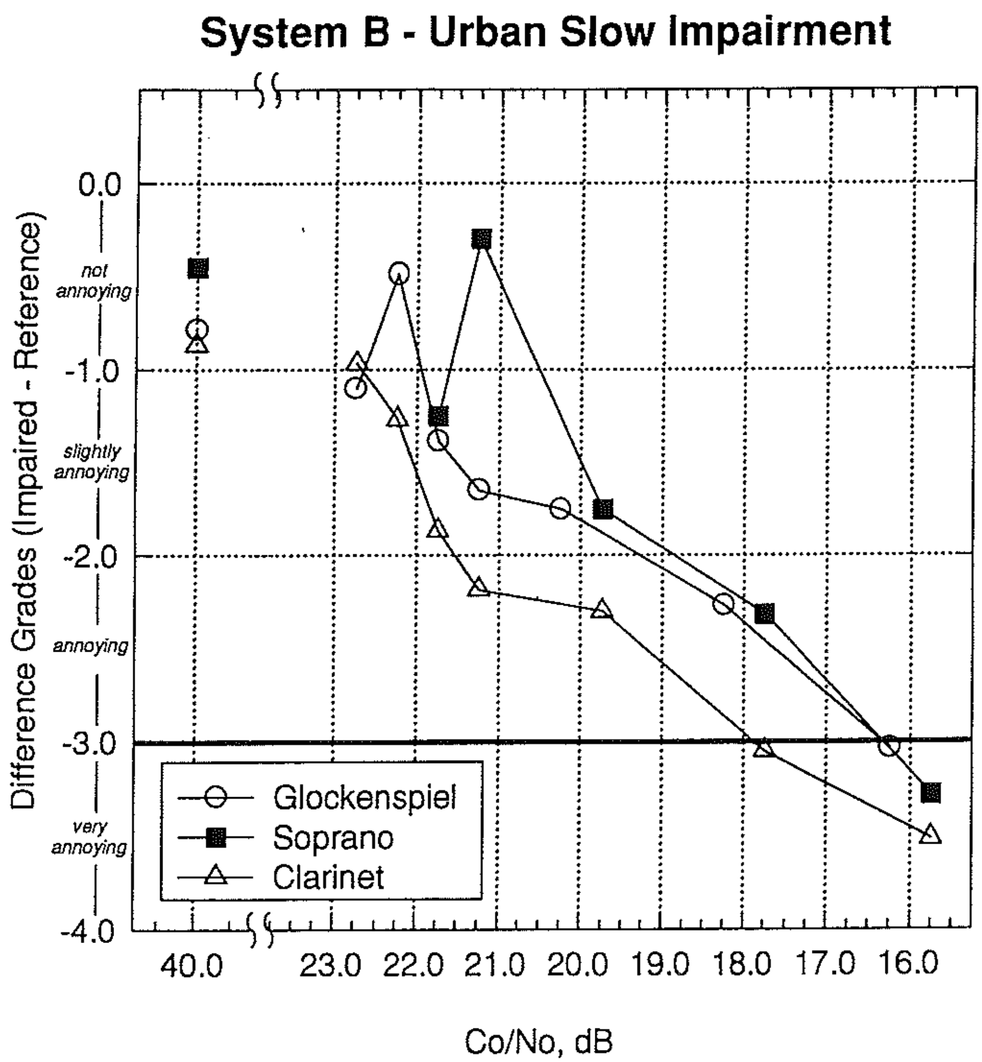


Figure 3.29. Failure characteristic curves for system *b* for all three test materials with the Urban Slow Multipath impairment.

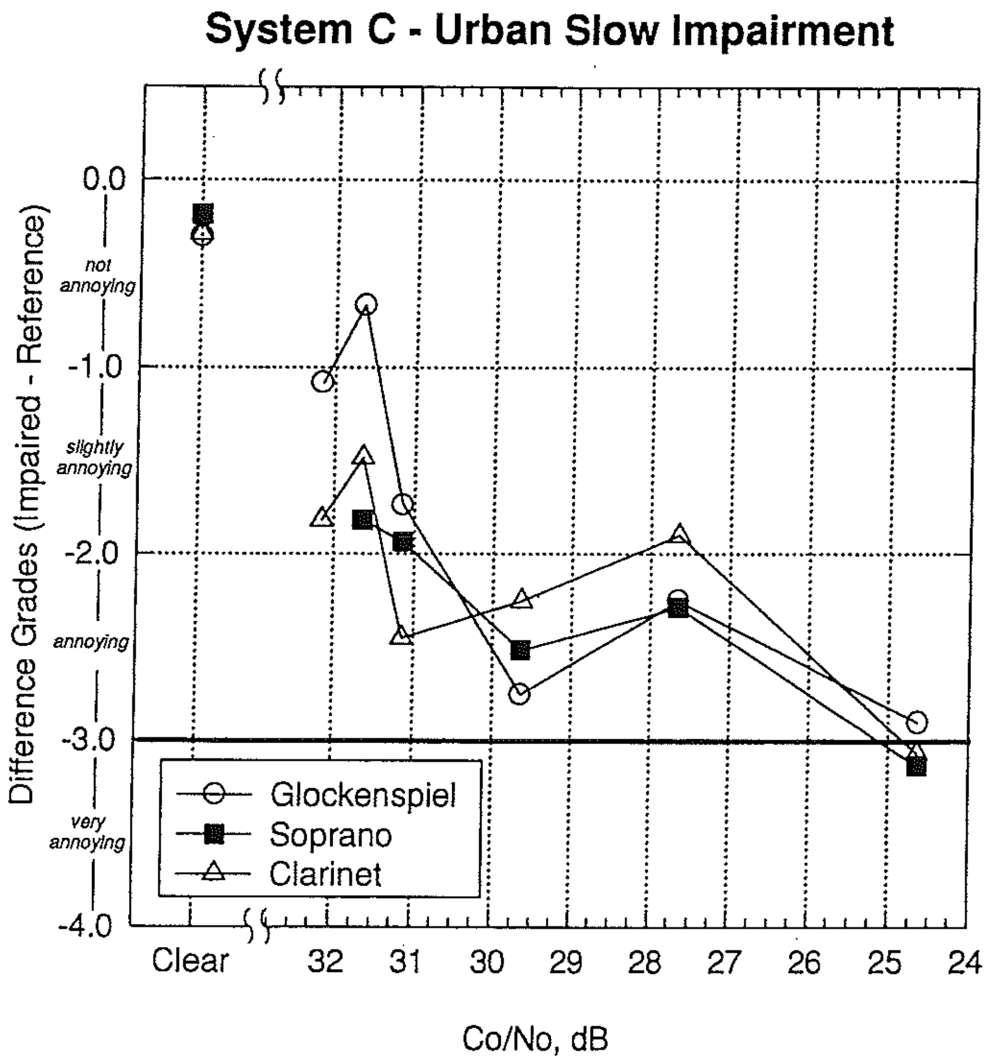


Figure 3.30. Failure characteristic curves for system *c* for all three test materials with the Urban Slow Multipath impairment.

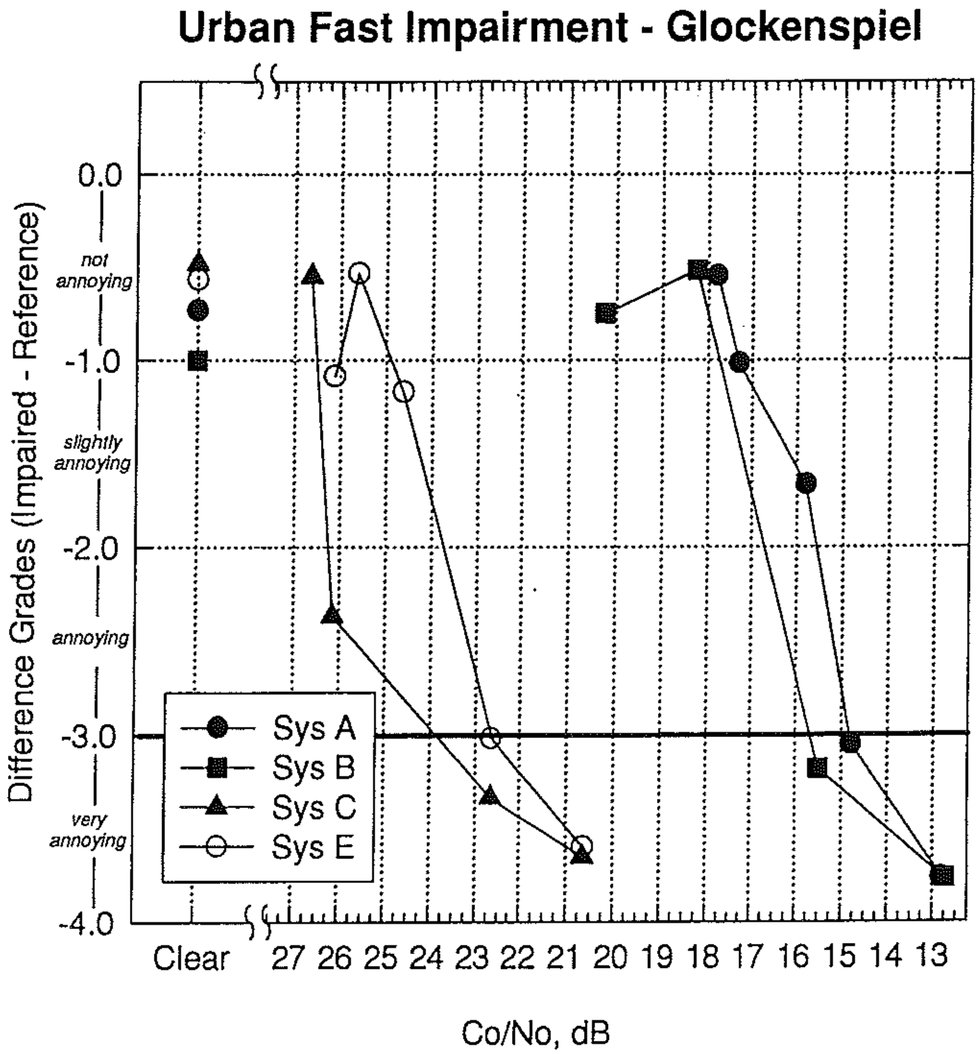


Figure 3.31. Failure characteristic curves of the four DAR systems tested with urban fast multipath impairment and the Glockenspiel test material.

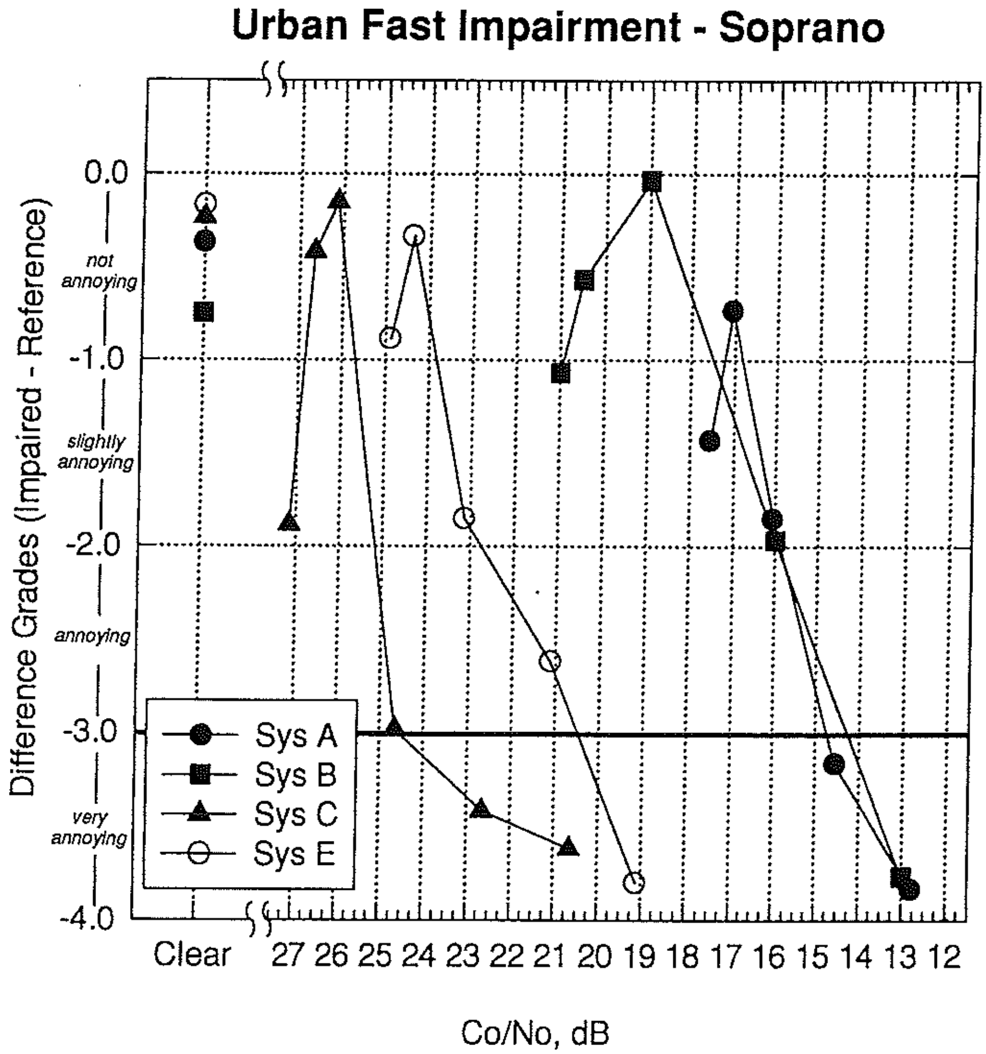


Figure 3.32. Failure characteristic curves of the four DAR systems tested with urban fast multipath impairment and the Soprano test material.

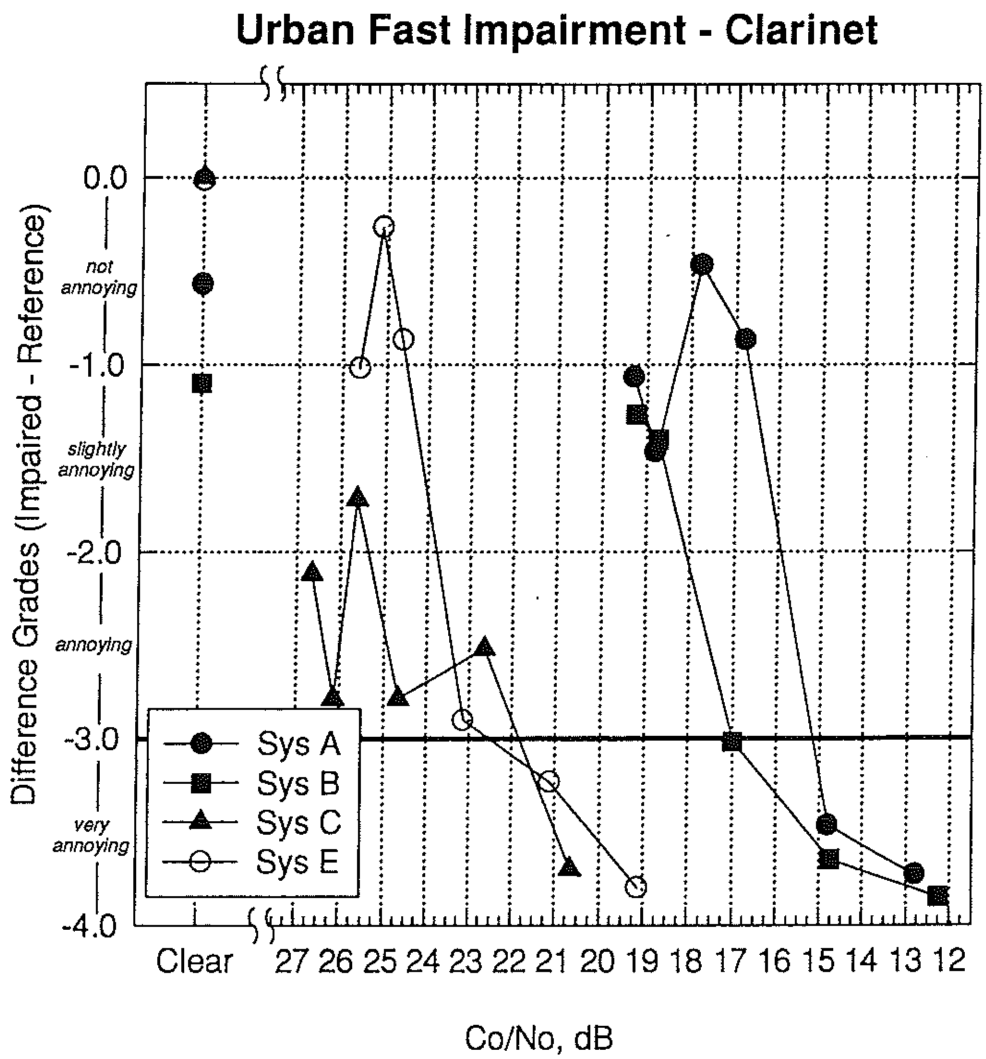


Figure 3.33. Failure characteristic curves of the four DAR systems tested with urban fast multipath impairment and the Clarinet test material.

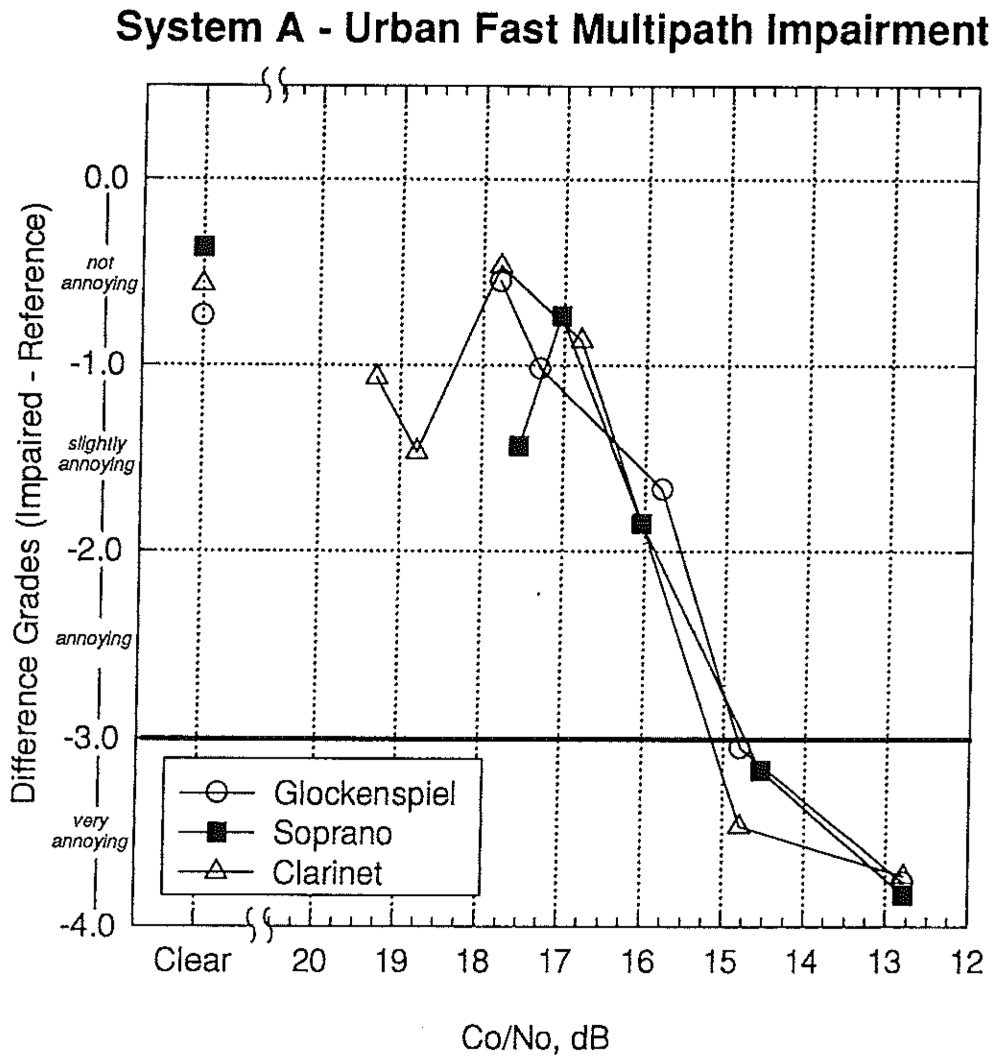


Figure 3.34. Failure characteristic curves for system *a* for all three test materials with the urban fast multipath impairment.

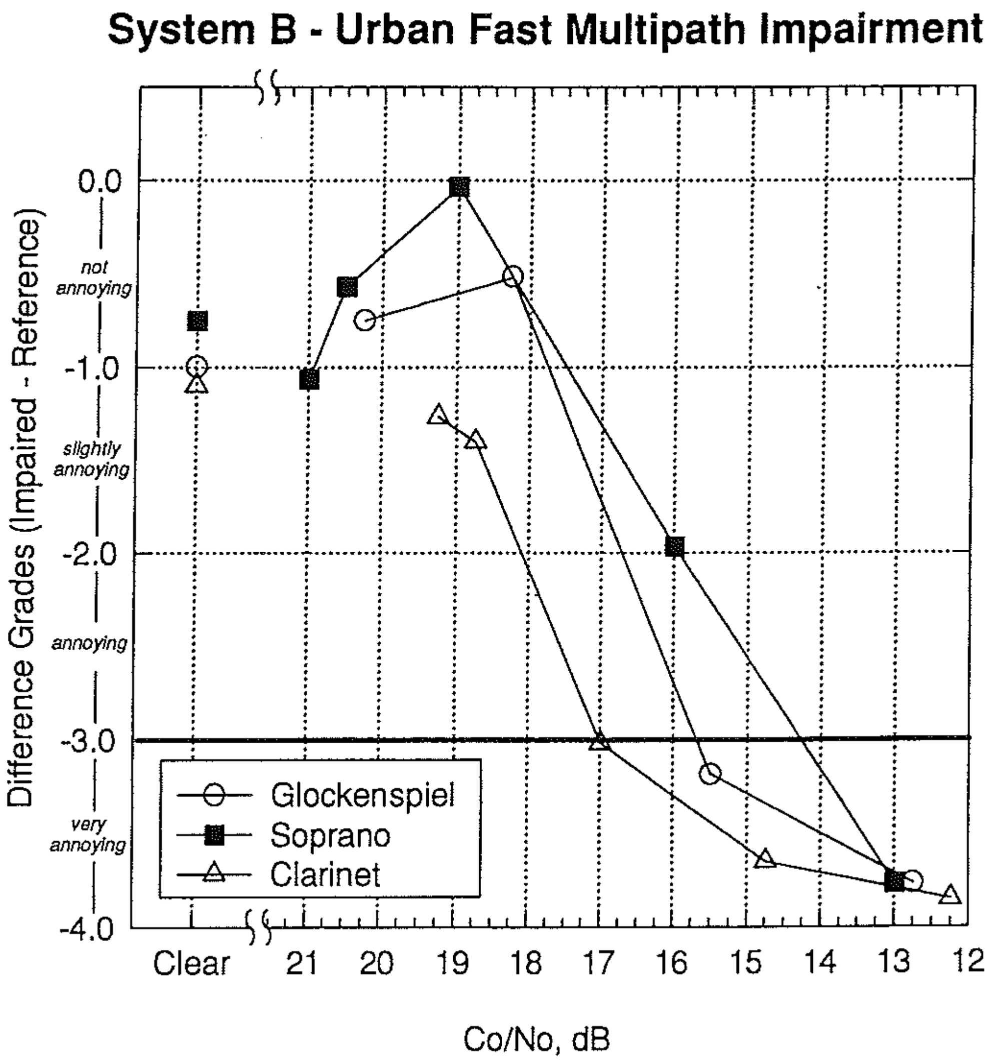


Figure 3.35. Failure characteristic curves for system *b* for all three test materials with the urban fast multipath impairment.

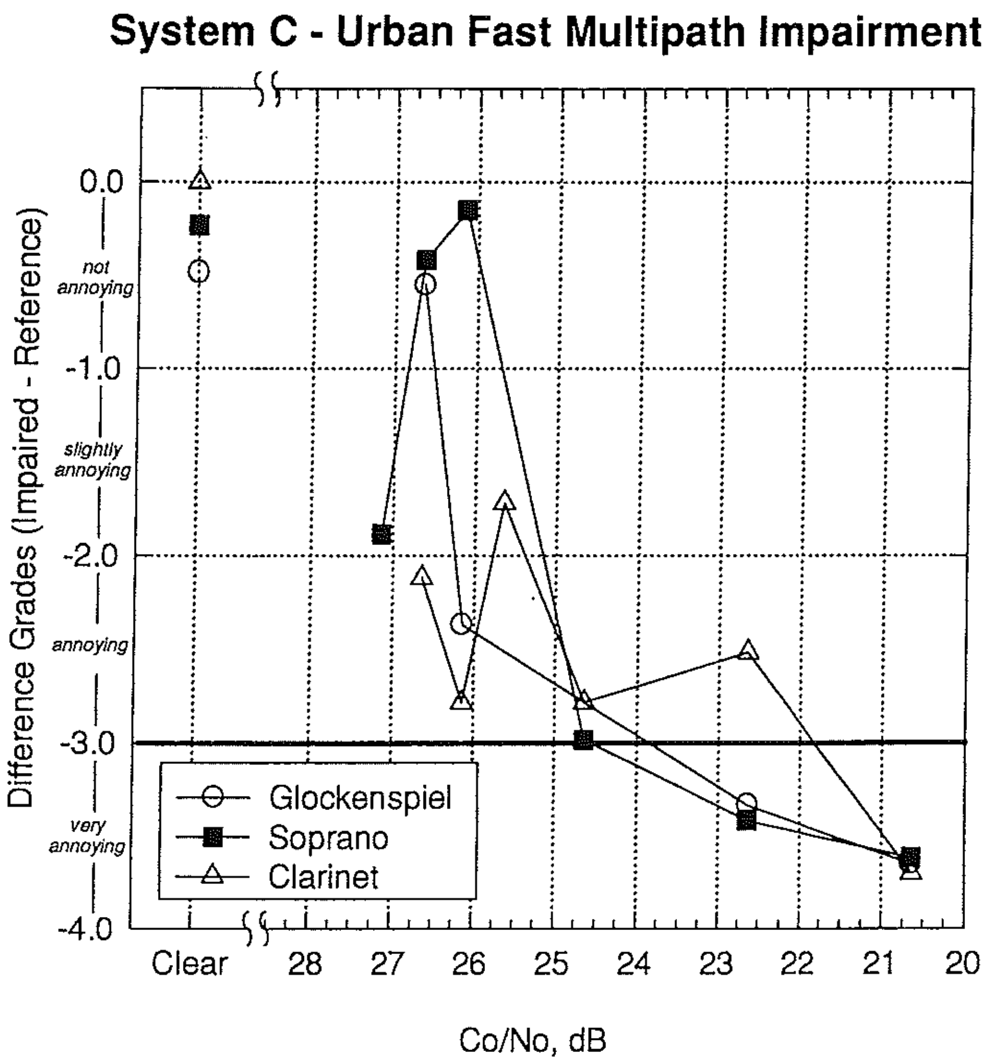


Figure 3.36. Failure characteristic curves for system *c* for all three test materials with the urban fast multipath impairment.

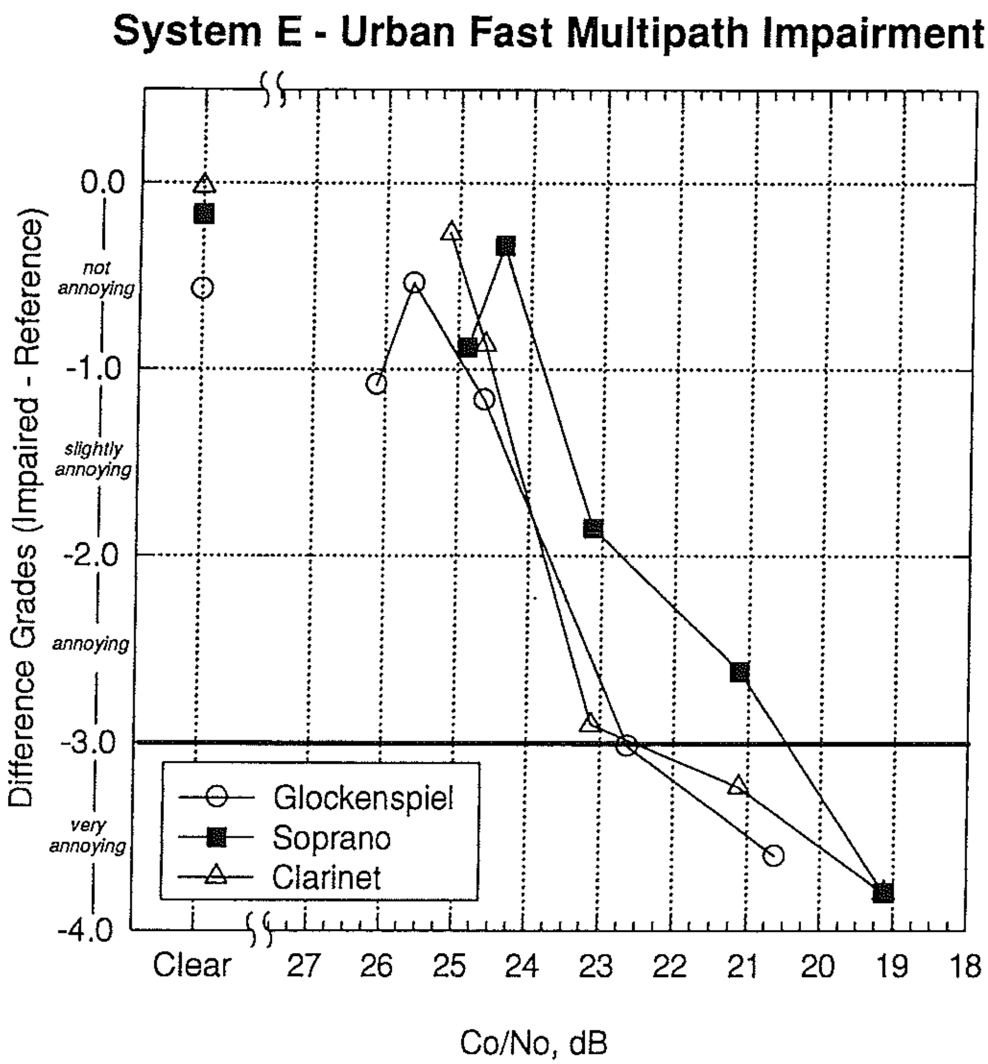


Figure 3.37. Failure characteristic curves for system *e* for all three test materials with the urban fast multipath impairment.

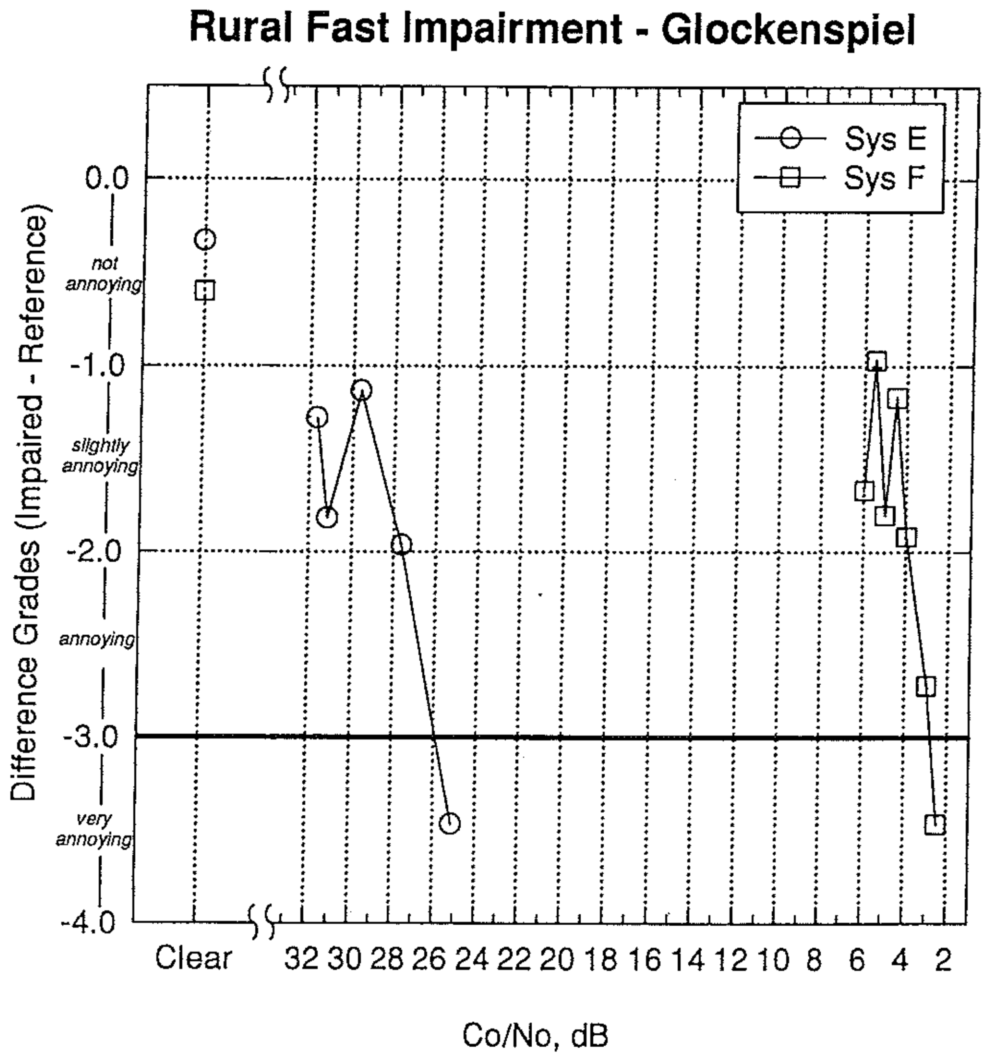


Figure 3.38. Failure characteristic curves for the two DAR systems tested with the rural fast multipath impairment and the Glockenspiel test material.

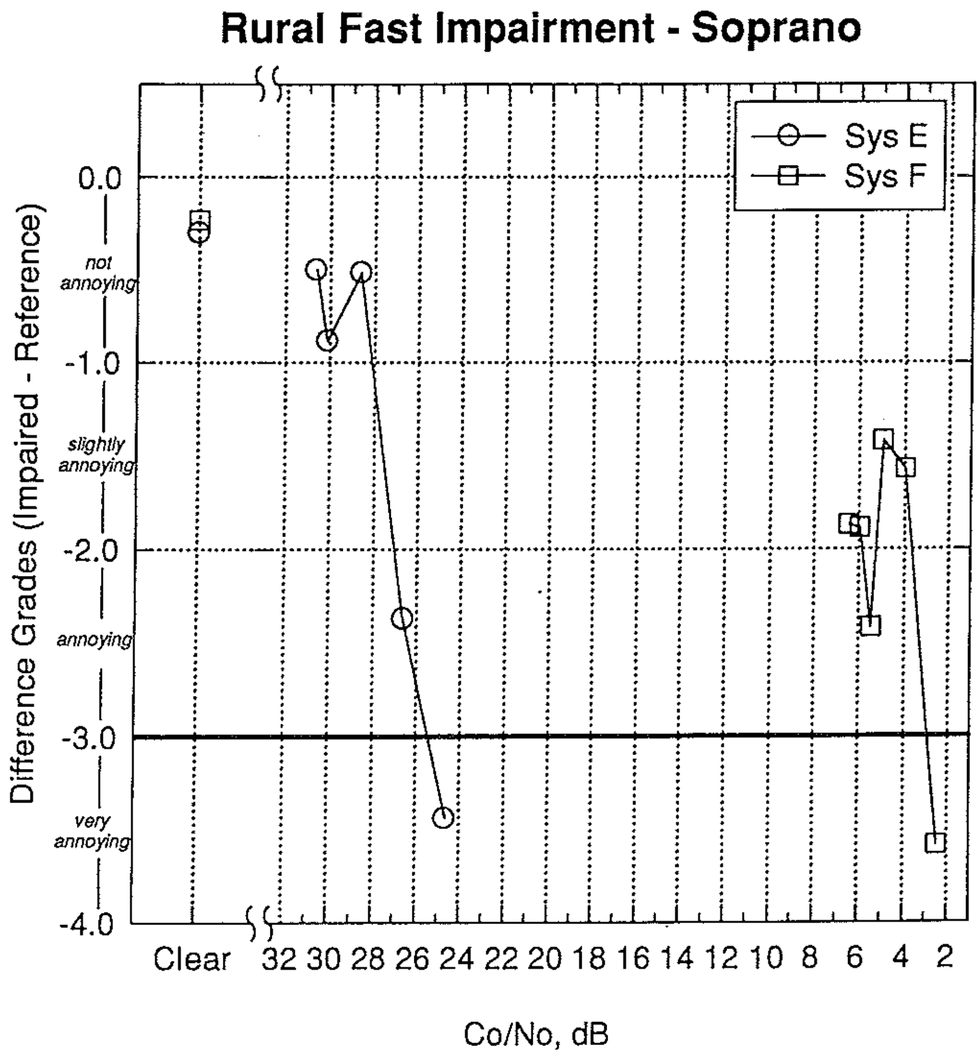


Figure 3.39. Failure characteristic curves for the two DAR systems tested with the rural fast multipath impairment and the Soprano test material.

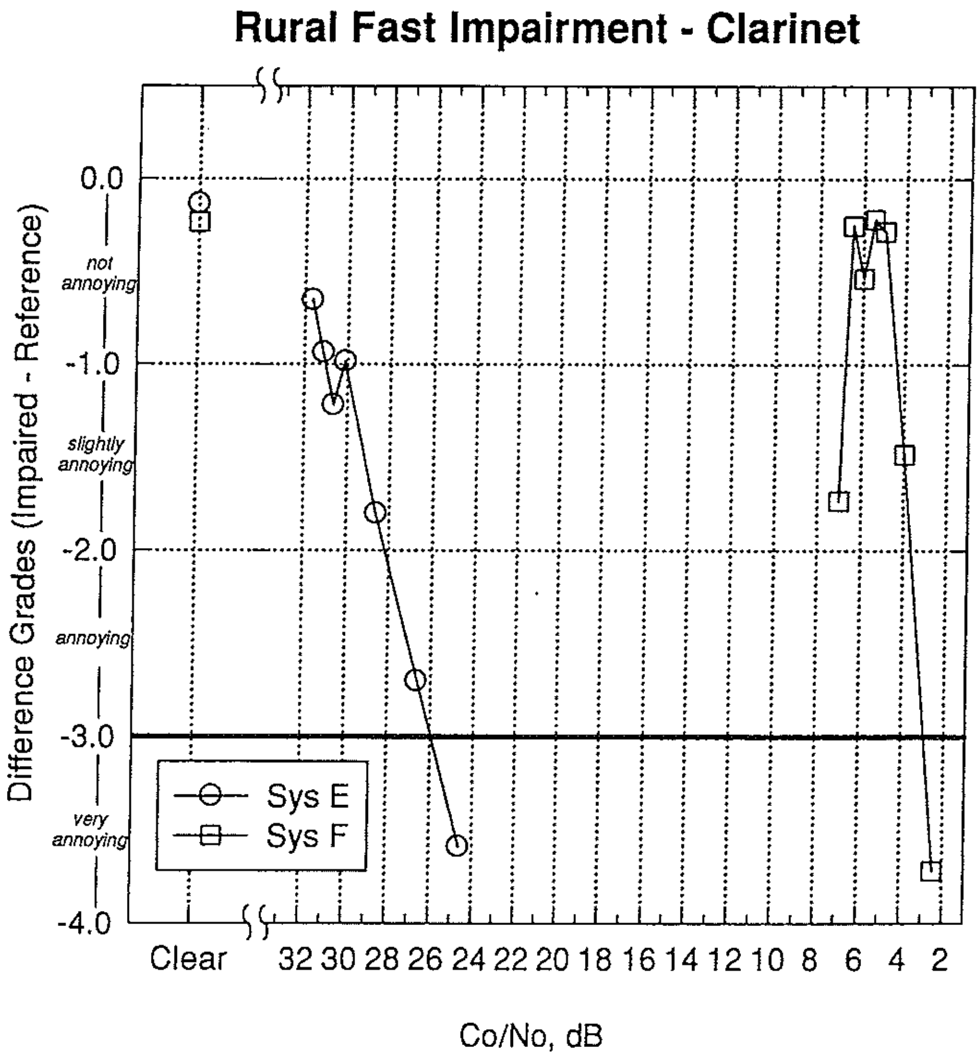


Figure 3.40. Failure characteristic curves for the two DAR systems tested with the rural fast multipath impairment and the Clarinet test material.

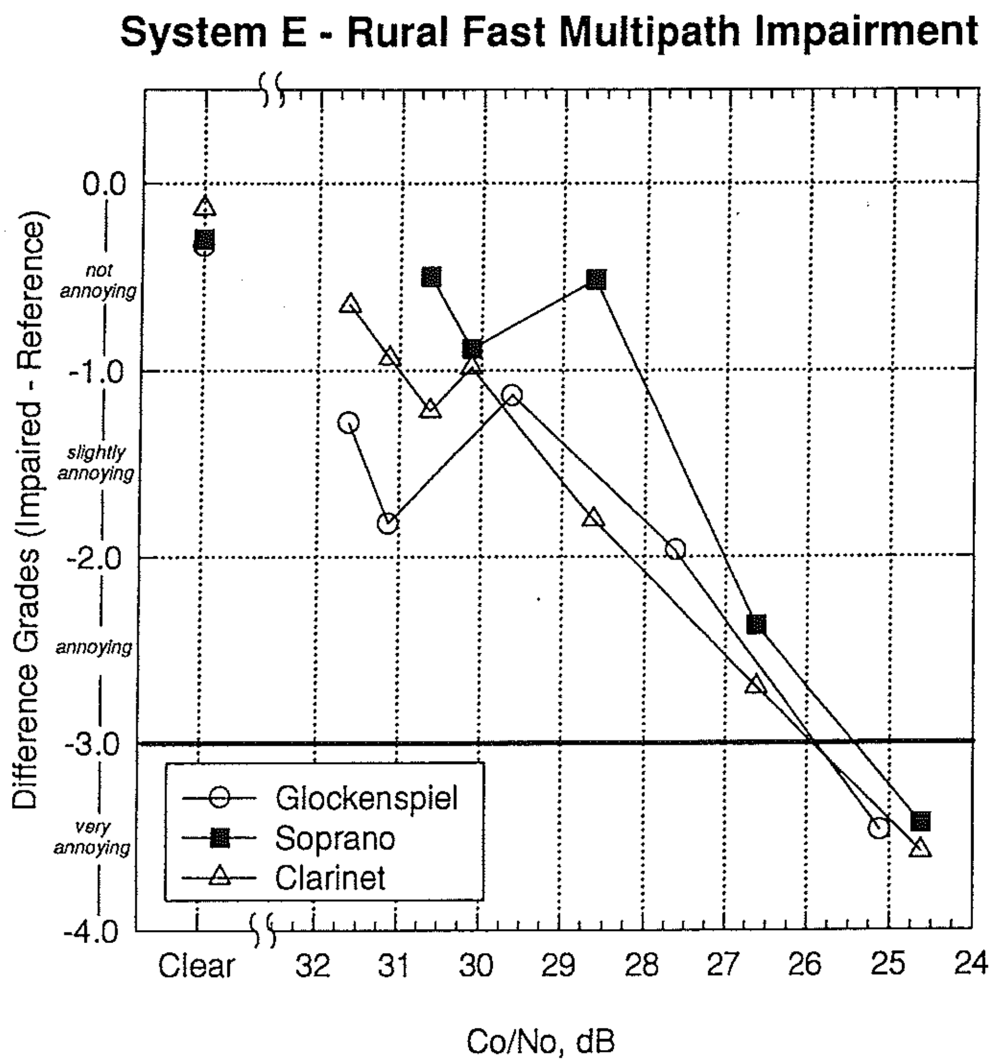


Figure 3.41. Failure characteristic curves for system *e* for all three test materials with the rural fast multipath impairment.

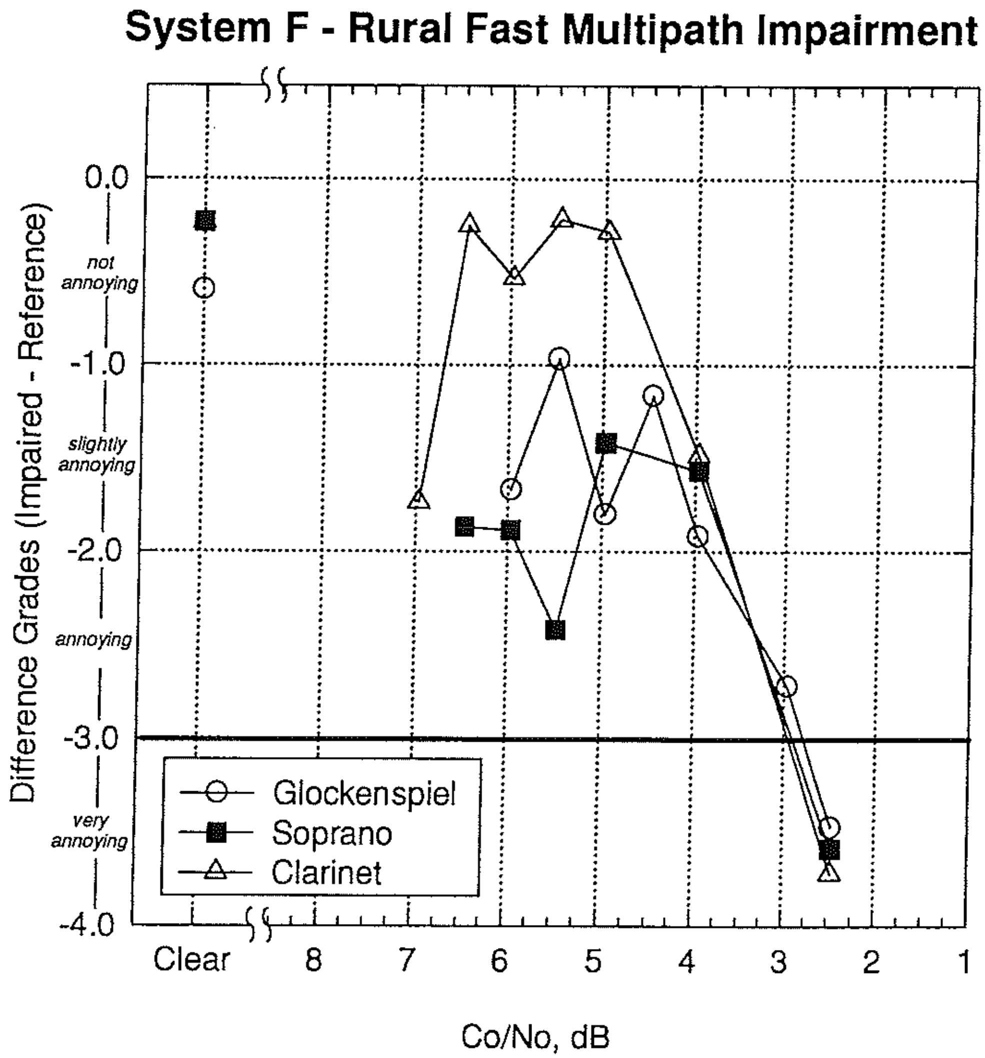


Figure 3.42. Failure characteristic curves for system *f* for all three test materials with the rural fast multipath impairment.

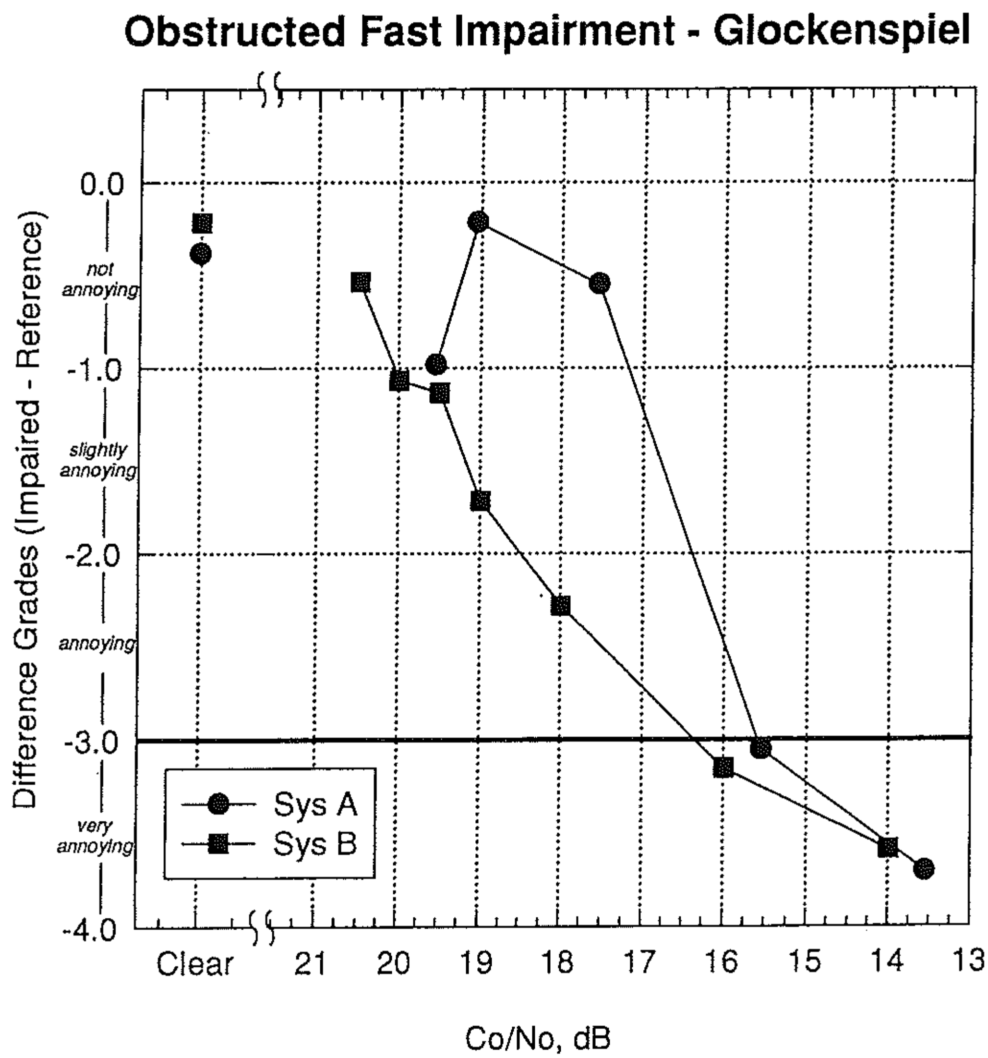


Figure 3.43. Failure characteristic curves of the two DAR systems tested under obstructed fast multipath impairment and the Glockenspiel test material.

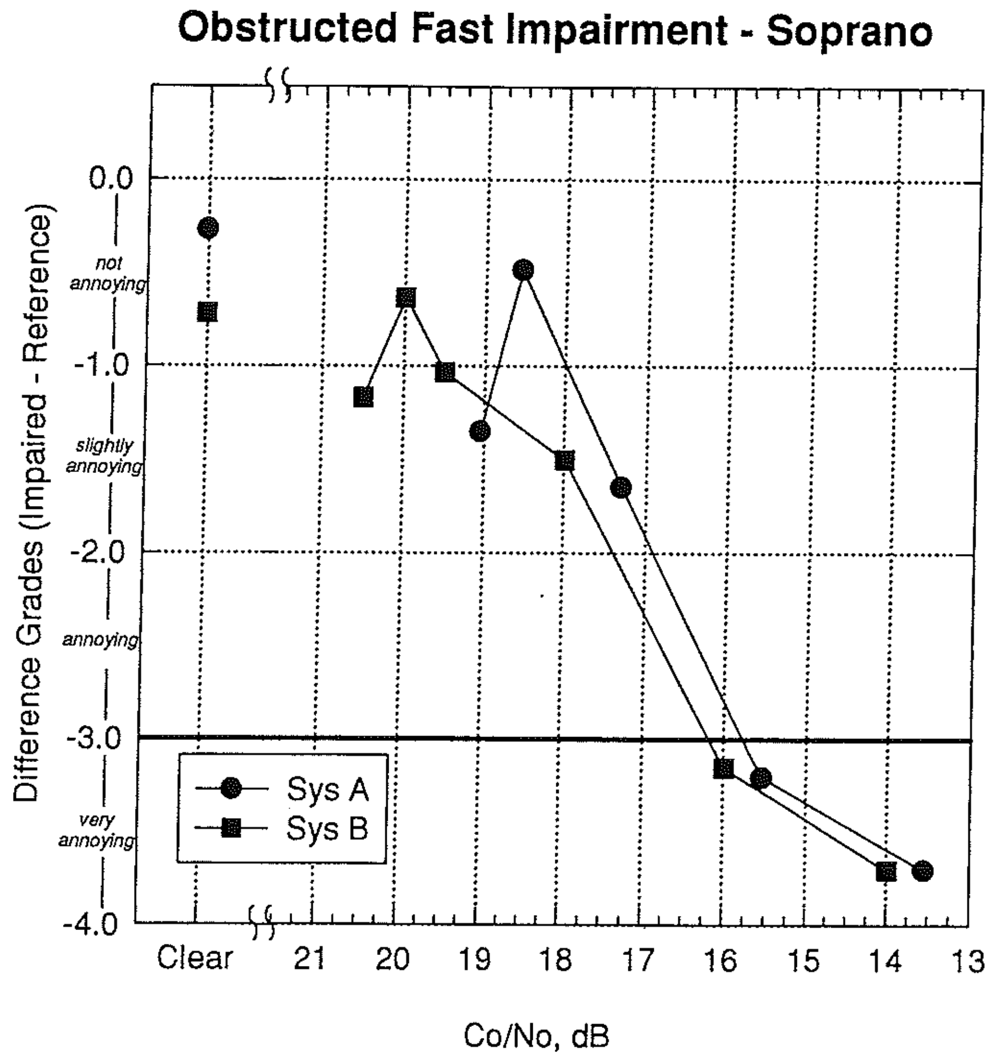


Figure 3.44. Failure characteristic curves of the two DAR systems tested under obstructed fast multipath impairment and the Soprano test material.

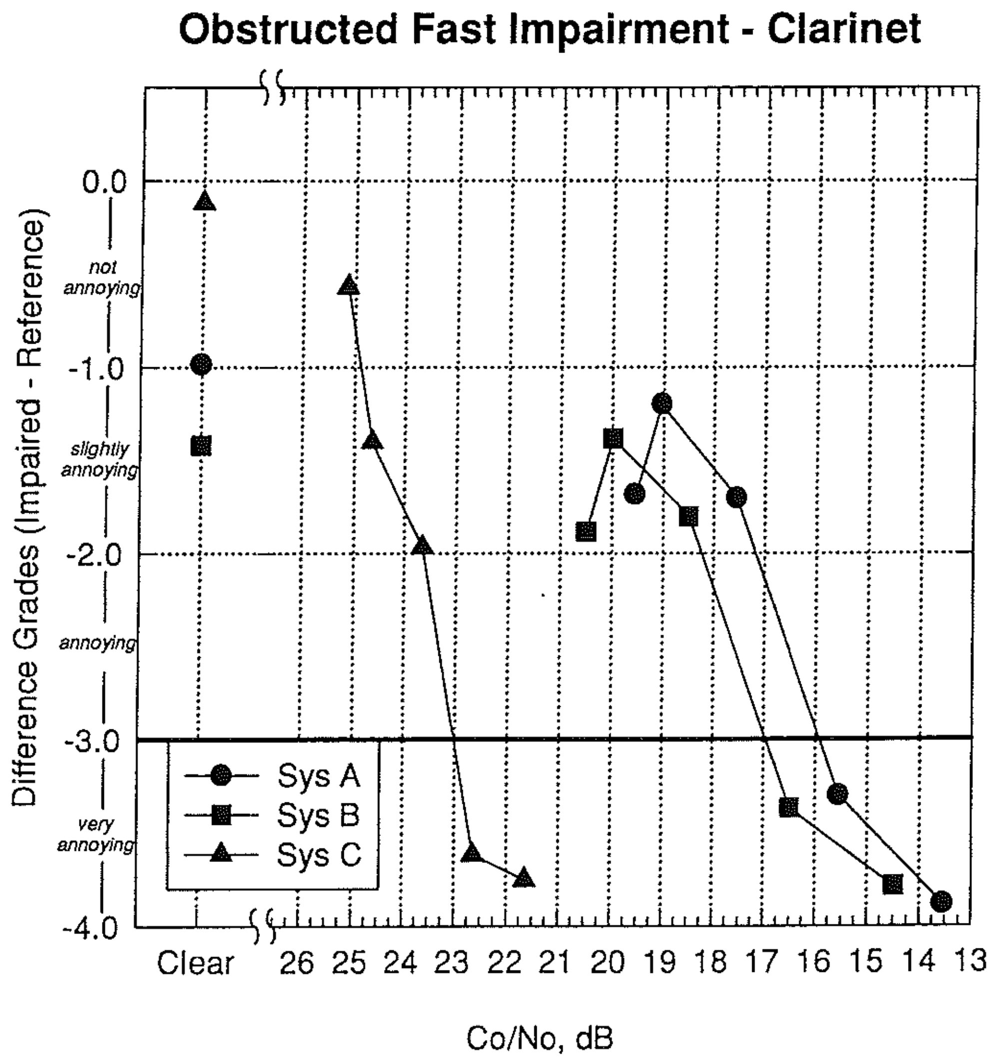


Figure 3.45. Failure characteristic curves of the three DAR systems tested under obstructed fast multipath impairment and the Clarinet test material.

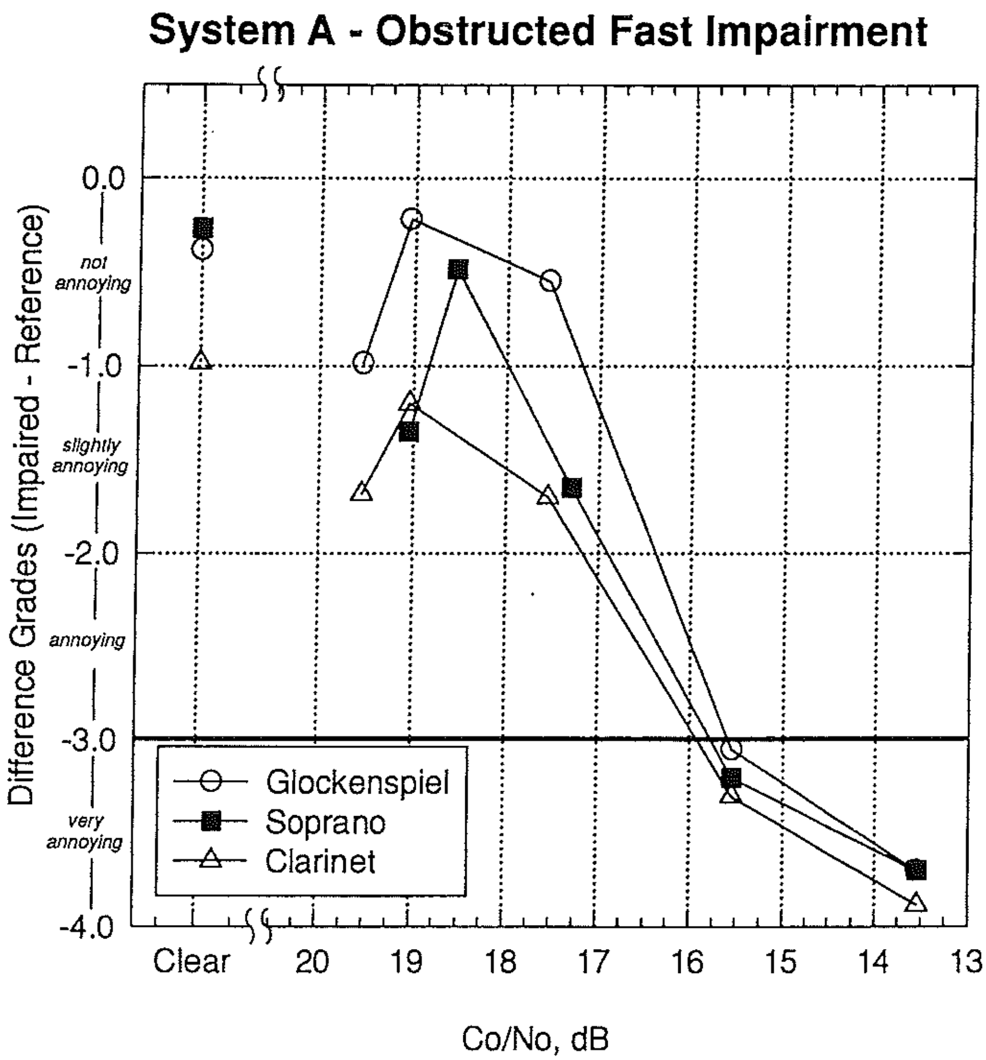


Figure 3.46. Failure characteristic curves for system *a* for all three test materials with the obstructed fast multipath impairment.

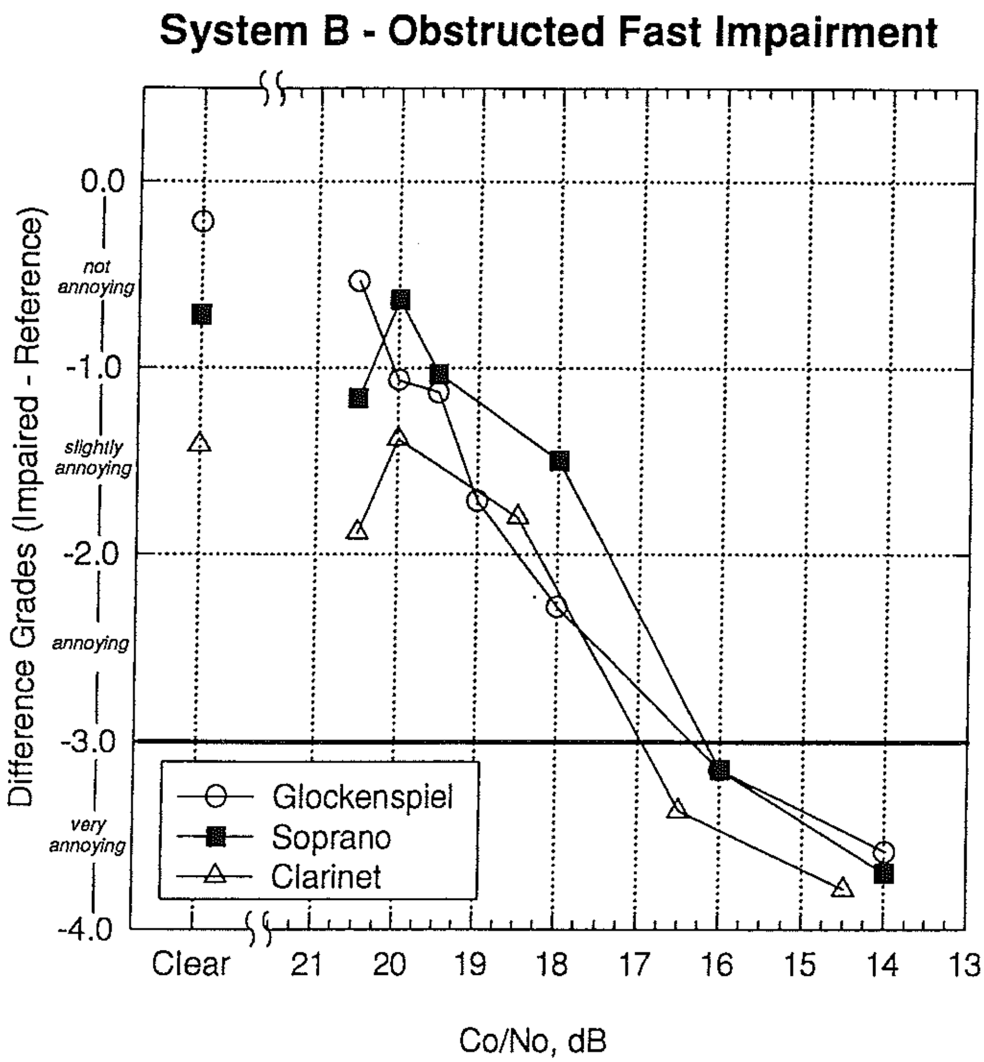


Figure 3.47. Failure characteristic curves for system *b* for all three test materials with the obstructed fast multipath impairment.

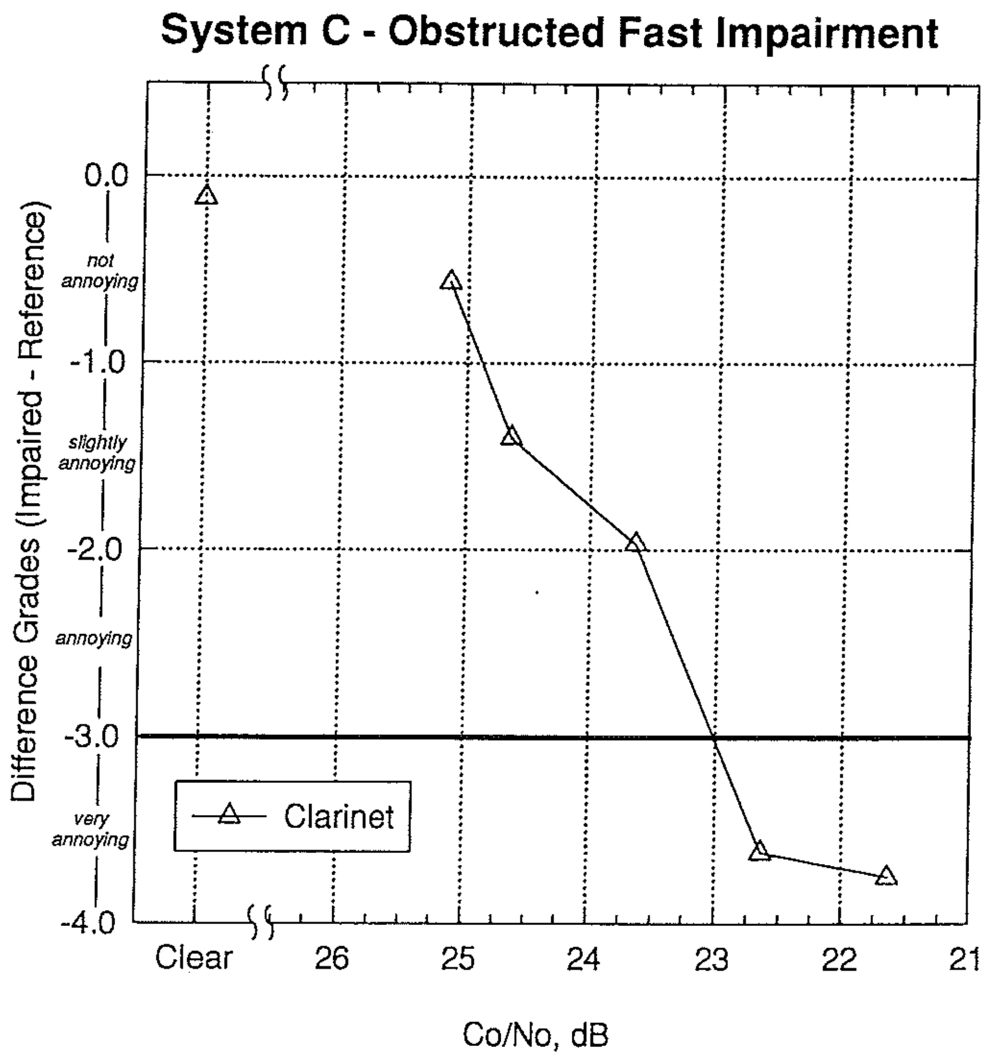


Figure 3.48. Failure characteristic curve for system *c* for the Clarinet test material with the obstructed fast multipath impairment.

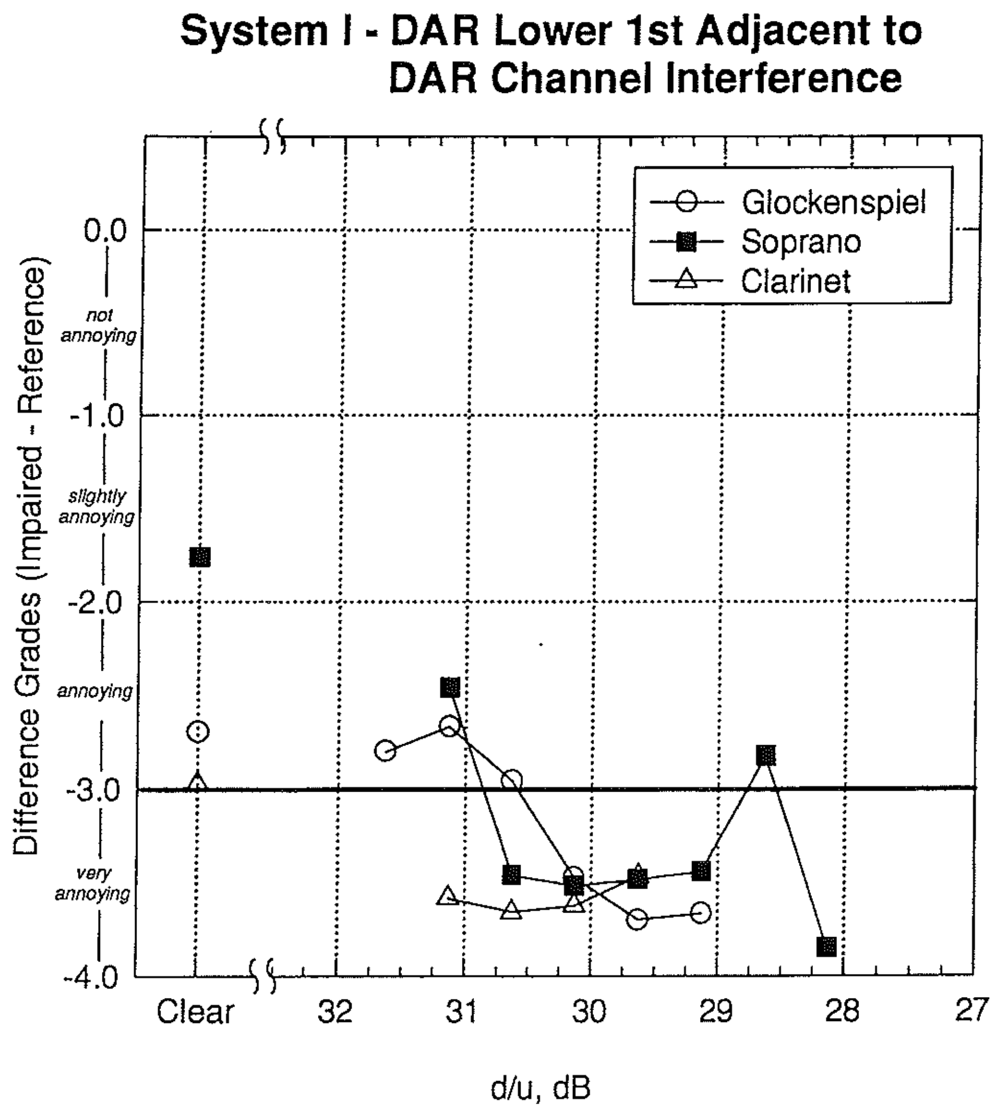


Figure 3.49. Failure characteristic curves for system *i* for all three test materials with the DAR lower 1st adjacent to DAR channel interference. (DAR laboratory was unable to certify digital power levels for this system)

4. Subjective Retests of Audio Quality

4.1 Background

After the original quality test had been completed at the CRC on 10 proposed DAR systems, system *h*, which was one of those 10 systems, was modified. The present retest was to determine whether or not the modification had altered the perceptual quality of that system.

There are many different ways to approach the question of identity in quality between the modified version (now called system *l*) and the original one (*h*). The fundamental question is whether or not *l* would have performed in the same way in the original quality test as *h* had done. The most direct way to answer the question would be to replicate the original quality experiment, but substituting system *l* for *h*. Such an approach, however, would bring many difficulties. For example, recruiting essentially equivalent subjects, might prove to be time-consuming. And the time needed for full replication of these tests would further delay the testing process and would escalate costs.

At the other extreme of cost and complexity is running an experiment in which subjects would compare *h* and *l*, exclusively and directly. The biggest problem with this approach is that one could never be sure, no matter what comparison methods were used, whether the results could be directly extrapolated to the outcomes of the original experiment. The many necessary differences in methodologies between the original experiment with 10 systems, and one with only 2 systems, would make such a projection questionable, regardless of how *h* and *l* were found to compare.

To minimize both problems - those of cost, and of veracity - a compromise approach was taken. The experiment reported here did follow the basic design of the original experiment, but version *l* was compared with only 4 of the other systems (including system *h*) rather than with all 10 systems. Other departures from the original quality experiment are outlined and discussed in the following "Test Procedures" section.

4.2 Test Procedures

No consequential changes in the room and equipment had occurred during the time elapsing between the original quality experiment and this retest. Accordingly, in using the same situation for the retest, it can be assumed that the basic acoustic, and other conditions were the same for the two experiments.

For reasons of efficiency and cost, 8 subjects were used rather than 21 as in the original experiment. To achieve stable results with this smaller number of listeners, exceptionally sensitive listeners were used exclusively, chosen for this characteristic from a pool of subjects who had established their expertise in prior research at the CRC labs and who were available for the present test. Of these 8 subjects, 5 had been in the original EIA quality test.

The same 9 audio materials were used as in the original experiment. A total of 5 systems were examined in the present test, rather than 10 as in the previous test. The 5 systems used were *a*, *d* and *j* as well as systems *h* and *l*. The primary systems of interest were, of course, *h* and *l*. The other 3 were chosen to represent the range of system qualities present in the original study. Thus *a* was from the highest rated Group 1 of the original test, *d* was from Group 3, and *j* came from Group 4, the lowest rated one. System *h*, itself, was from Group 2. The presence of the other systems besides *h* and *l* was an attempt to have a total evaluation context as similar to that of the original experiment as possible, short of a full replication. No matter what the outcome regarding systems *h* and *l*, the validity of that result required that the ratings of the other systems (*a*, *d* and *j*) were commensurate here with those in the original study.

Since the major question was establishing whether or not systems *h* and *l* were equivalent in subjective quality while using only a small group of subjects, an additional means of reducing variance and enhancing sensitivity (besides using exceptionally expert subjects) was implemented. This additional means was to include both of systems *h* and *l* three times, rather than only once, in all conditions. In effect, this was like having 9 systems in the experiment, with 9 observations per system, one per audio material. But of these 81 observations, systems *h* and *l* had 27 each, as though each of them were 3 separate systems, while the other 3 systems (*a*, *d*, and *j*) each had 9.

The subjects, of course, knew nothing about how many systems were in the test. They were simply evaluating items of audio material, with 9 observations for each of the 9 materials. Training was carried out in the same way as for the original test, and took up most of the morning of each subject's first day of the two day experiment.

One important departure from the original experiment was that not all 81 items that were to be rated in the later blind sessions were made available to the subjects during training. Specifically, the *h* and *l* versions of each material were included only once in the training materials, rather than three times as in the grading sessions. And so, there were 45 items in training, rather than the full 81 of the blind grading sessions. This was done so that subjects would not discover the fact of repeated presentations of identical items (the 3 inclusions of both *h* and *l* for each audio material) during training. Such discovery might have led a subject to adopt a grading strategy during the blind trials which anticipated repetitions. A strategy of this sort might detract a subject from listening to, and grading, each audio item as a stand-alone entity. This in turn, would defeat the purpose of the duplications for systems *h* and *l* which was specifically for providing <u>independent</u> repeated ratings.

In all other essential respects, the training and blind testing procedures were the same as in the original experiment. After a morning training session on the first day (done in a group when there was more than one subject scheduled for a two-day experimental time slot), each subject had three blind grading sessions in the afternoon of the first day. The second day consisted of six grading sessions during the morning and afternoon. Each grading session consisted of 9 trials. The 81 presentations were distributed over the sessions so that each of the 9 audio materials always occurred once per session. Within that limitation, the "9" systems (5 systems with 3 repetitions for both *h* and *l*) were distributed equally among the nine sessions to the greatest extent possible. Within sessions the systems were presented as unpredictably ("randomly") as possible as to their trial-by-trial presentation order. To control for any interaction between events in time (such as fatigue, learning) and the content of sessions, the actual temporal order of the nine programmed sessions was different for each subject.

4.3 Results and Discussion

Subjects had been told before training that only a subset of the materials to be graded in blind sessions were to be presented in training. One comment made by almost all of the subjects after the experiment was completed was that the grading sessions were much more difficult than the training sessions had led them to expect. It is reasonable to assume that they found this to be true because *h* (as seen in the original experiment) and *l* (the modified version of *h*) were very high quality codecs. The three repetitions of each of these codecs in grading sessions would, then, have provided many more materials that were difficult to distinguish from transparency than was the case in the original quality experiment. Since these repetitions were not presented in training (for reasons noted previously), that part of the experiment would seem relatively easy compared to the grading sessions.

Five of the 8 subjects had been in the original quality test, and these 5 listeners also commented that this test was more difficult in an absolute sense than the previous one had been. Some objective evidence for this is that the expertise scores of these subjects (explained in the report of the original quality test) was consistently lower in this retest than in the original study. On average, their expertise scores in the retest were 78% of the magnitudes of those scores in the previous test. From their comments, none of these subjects seemed to suspect that this was a retest which included some of the same systems as before.

Harking back to the relative ease during training compared to blind rating sessions, we might note that the subjects did not themselves volunteer reasons for the unexpected difficulties during blind grading sessions, even though this did seem to surprise them. This might be interpreted to indicate that the subjects did not discover the fact of repetitions of items during the grading sessions. If this is true (and it is likely that if not true, subjects would have mentioned finding repetitions), then, indeed, as intended, the ratings obtained were truly independent from item to item despite repetitions.

The average diffgrades (coded minus hidden reference grades) for each of the 5 systems at each of the nine audio materials are presented in Fig. 4.1. The actual numerical diffgrades plotted in the figure are shown in Table 4.1 below the figure. Since the differences between the averages for the various audio materials are of little interest here, the bottom row of this table shows the difference between *h* and *l* at each material instead of (as would be more usual in a table of this type) the average diffgrade per material.

It can be seen (in the figure and table) that *h* and *l* both received highly similar grades across all audio materials. Statistical analysis of the total data set (ANOVA, followed by *t*-tests, all *p*'s <

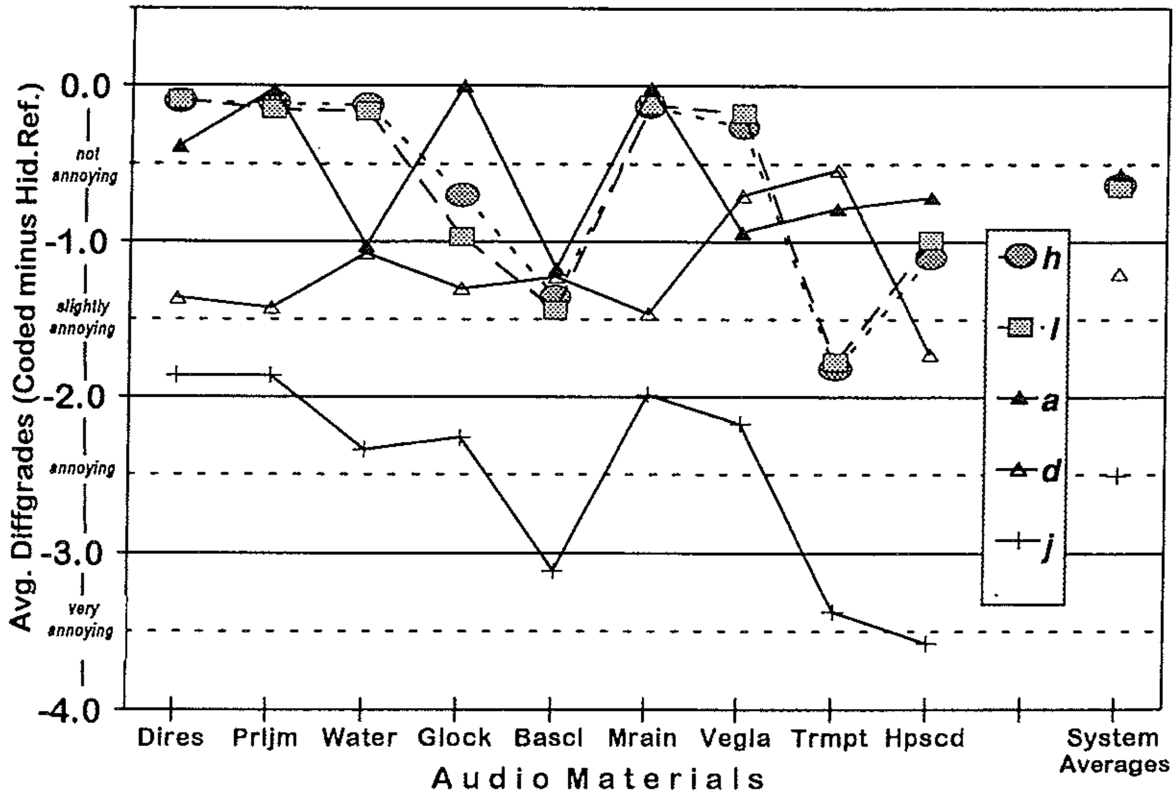


Fig. 4.1 Average diffgrades of the 8 subjects in the quality retest for all the systems in that experiment. Each audio material was presented 3 times for systems *h* and *l*, and once for each of the other 3 systems (*a*, *d*, and *j*) in the 81 trials of the retest.

	Dires	Prljm	Water	Glock	Bascl	Mrain	Vegla	Trmpt	Hpscd		SysAvg
h	-0.10	-0.11	-0.13	-0.70	-1.36	-0.13	-0.26	-1.82	-1.10		-0.64
l	-0.09	-0.15	-0.16	-0.97	-1.44	-0.12	-0.18	-1.78	-0.99		-0.65
a	-0.39	-0.03	-1.03	0.00	-1.18	-0.01	-0.94	-0.79	-0.71		-0.56
d	-1.36	-1.43	-1.08	-1.30	-1.23	-1.46	-0.70	-0.54	-1.73		-1.20
j	-1.86	-1.86	-2.34	-2.26	-3.11	-1.99	-2.18	-3.38	-3.58		-2.51
h - l	-0.01	0.04	0.04	0.27	0.08	-0.01	-0.09	-0.04	-0.11		0.01

Table 4.1 Average diffgrades for each system at each audio material and for overall system averages. The bottom row shows the algebraic difference between the diffgrades for systems *h* and *l* in the top two data rows.

.05) shows that, within any material, differences between any two systems must exceed 0.68 of a grade in absolute magnitude in order for those systems to be considered statistically different for that material. The bottom row of the table shows that the differences between *h* and *l* are less than 1/10th of a grade except for the Glockenspiel (0.27) and for the Harpsichord (0.11). All of these differences are considerably smaller than the 0.68 needed for the verdict of a reliable difference.

The relatively large 0.68 needed for significance (compared to 0.45) in the original study) is due to the smaller number of subjects used in this retest (8 versus 21 previously). The exceedingly small magnitudes of differences between *h* and *l* across materials here (mostly less than 1/10th of a grade) must be attributed to the exceptional expertise of these subjects as well as to the three repeated observations for those two systems.

It must be concluded that *h* and *l* cannot be considered to be different in quality across all 9 audio materials under the conditions of the experiment.

The results for the other systems (*a*, *d*, and *j*) are important too for several reasons. First, the results must show that the experiment was not insensitive to revealing differences. If it were, then no valid final judgment could be made about *h* and *l* since their apparent lack of difference might be explained as the outcome of a poorly executed experiment.

Examining the table and figure shows that many significant differences did emerge in the experiment. For example, looking at the "water" and "trmpt" materials, the differences between *h* (and *l*) on the one hand, and both *a* and *d* on the other, obviously exceed the criterion magnitude of 0.68. And, of course, almost all comparisons between system *j* and any of the other systems (with only 3 exceptions involving system *d*) are larger than 0.68. Hence it cannot be said that the experiment was too crude for reliable differences to emerge. And so the lack of difference between *h* and *l* is not due to experimental insensitivity.

Second, and quite important, the results for these other three systems must relate in some sensible way to the outcomes for those same systems in the original quality experiment. If such a relationship was obscure, then it might be argued that despite the apparent identity of *h* and *l* here, the conditions of this study were so different from the original one that one cannot assume that *l* would have performed identically to *h* if it had been in the original study instead of *h*.

A visual comparison of Fig. 4.1 here, with Fig. 2.2a (of chapter 2) of the original quality test results shows that the general pattern of results for each of the comparable systems are strikingly similar. This is most immediately obvious with system *j* because it occupies a distinct region towards the bottom of both figures. But it is also true of the other systems. A point-by-point comparison shows that the up-and-down pattern across the audio materials is generally the same for each system in both figures.

An even stronger case can be made for the identity of outcomes between the two experiments. Recalling that 5 of the subjects in the present retest were also in the original quality assessment, the results for those subjects alone can be isolated and compared. These comparison will not be presented in detail here. The reader can be assured, however, that they show virtually identical numerical results with no statistical differences between them.

Since the present results were obtained with exceptionally expert subjects, as compared with sufficiently expert ones in the original test, the implication is that whatever small differences are seen between the two experiments is due to the larger range of individual subject expertise in the first test. In fact, consistent with findings about the grades of subjects with exceptional expertise in many previous studies at the CRC, the outcomes here show somewhat harsher judgments (lower diffgrades) than in the first quality test. In other words, we are confident that had listeners with fully comparable expertise been used in this retest as in the first test (and if the number of subjects had been as large), the results would have been quite identical. This retest, then, provides excellent confirmation of the original study.

4.4 Conclusions

We conclude that *h* and *l* can be considered to be identical in quality. This finding is not due to insensitivity of the retest to finding differences. The outcomes of this retest strongly confirm the quality outcomes of the original experiment.

We conclude that if *l* had been in the original quality experiment instead of *h*, the outcomes of that first experiment would have been no different than they actually were.

5. Subjective Retests of Transmission Impairments

5.1 Introduction

After the original transmission impairment tests had been completed at the CRC on the proposed DAR systems, two systems, *e* and *h* were modified. The modified version of system *e* was named system *k* while the modified version of system *h* was named system *l*. As such, the present retests were conducted to determine the failure characteristics for the modified systems. These retests were only conducted for the gaussian noise and co-channel transmission impairments.

In order to minimize costs and time requirements it was decided that the retests would be limited to determining the threshold of audibility and the point of failure only. Furthermore, these points (TOA and POF) would be determined by means of expert observations and commentary, EO&C. The procedures for determining TOA and POF are described in greater detail in the following section.

5.2 Test Procedures

5.2.1 Test Materials

The three critical audio materials used in the impairment retests were the same as the ones used in the original impairment tests and were carefully selected by the staff at the NASA Lewis Research Laboratory because they were particularly sensitive to revealing artifacts resulting from transmission impairments. These materials are described in Table 3.1 in section 3.2.1 of this document.

5.2.2 Impairment Levels

As mentioned earlier, only the TOA and POF points were to be determined in the transmission impairment retests. Nonetheless, a complete set of recordings were provided by the staff at the NASA Lewis Research Laboratory for the two systems *k* and *l* as described below.

Each of the three audio test materials was recorded, for the two DAR systems (*k* and *l*) and the two types of impairment (gaussian noise and co-channel interference), at the following levels of impairment: CC, TOA_1, TOA_2, ..., TOA_N, S_1, S_2, ... , S_M, POF where:

CC = coded audio in a Clear Channel
TOA_1, TOA_2, ..., TOA_N = N stimuli (three or more) in the close neighborhood of the approximate TOA (threshold of audibility)
S_1, S_2, ... , S_M = M intermediate levels of impairments
POF = point of failure

For each of the above impairment levels, the C_o/N_o ratio for the gaussian noise impairment, or D/U for co-channel interference, was noted. The CC level was recorded at a high value of C_o/N_o (or D/U) so that transmission errors could be considered as negligible, hence the label "clear channel" given to that level. The increments in the C_o/N_o (or D/U) ratios at which the other impairment levels were recorded varied from 0.25 dB, for those situations where the DAR system failed abruptly, up to 1.0 dB for those situations where the system failed more gracefully.

5.2.3 Subjective test procedures

Two separate experiments were performed for each of the two systems. The purpose of the first test was to determine the TOA point for each system/impairment-type/audio-material combination. The procedure for determining the TOA points was the same as the one used in the original impairment tests. A detailed description of this procedure is given in section 3.2.3 of this document.

The second test was designed to provide an estimate of the POF points for the two systems. The procedure used for this test is described below.

Point of Failure

The purpose of the second experiment was to derive an estimate of the POF points for the two systems *k* and *l* as opposed to determining full failure characteristic curves for each system under various impairment conditions as was done in chapter 3. However, it was desirable to try to find some way of relating the results of the retests to the results of the original tests. To do this, a series of "matching experiments" were conducted wherein expert listeners identified the impairment levels required for the modified systems *k* and *l* to be perceptually equivalent to the POF's found for the original systems *e* and *h* in the original tests. The expert listeners consisted of three CRC staff members and were the same listeners who took part in the TOA EO&C tests.

In the figures of chapter 3, it can be seen that for many of the failure characteristic curves there are several points which fall in the POF range. (note that a diffgrade of -3 or less is considered to be POF). Therefore, in the retests, a perceptually equivalent impairment level for the modified systems was found for each of the points (for the old systems) which fell in the POF range.

As stated above, to determine the POF points for the two modified systems, a series of matching experiments were conducted at the CRC. In these tests the CRC staff members listened to the POF points for the original systems (*e* and *h*) and found the impairment levels for the modified systems (*k* and *l*) which were perceptually equivalent. This process was repeated for the two systems, the three test materials, and the two types of transmission impairments (gaussian noise and co-channel interference). Furthermore, this process was done for all points for the original systems which fell in the POF range.

In some instances, it was found that the perceptually equivalent impairment level fell between the impairment levels recorded at the NASA Lewis Research Center and provided to the CRC. In these instances, a value of C_o/N_o (or D/U) which was half-way between the levels provided was used even though this impairment level was not actually auditioned by the CRC listening team.

5.3 Test Results

5.3.1 Gaussian noise

The results of the retest for the gaussian noise impairment are given in Table 5.1 for systems *e* and *k*. The table provides the C_o/N_o ratios at TOA and POF, as well as the failure margin (FM). This is done for the three critical audio materials, Glockenspiel, Soprano, and Clarinet. It should be noted that the values for system *e* are taken from the original transmission impairment tests described in chapter 3 of this document. The bottom row of the table shows the *differences* between the original system and the modified system (i.e. *e-k*).

The results given in Table 5.1 show that the differences between the original system and the modified system are very small. None of the differences are greater than 0.72 dB. The largest differences tend to occur at TOA while the differences in the failure margins are extremely small (no more than 0.25 dB).

As stated earlier, the matching experiments were conducted for all points which fell in the POF range in the original transmission impairment tests. The POF points given in the table below represent the highest C_o/N_o ratio which fell into the POF range. The other POF points are omitted for clarity purposes since they do not alter the basic conclusions.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
*e**	10.76	9.76	1.0	10.51	9.51	1.0	10.76	9.76	1.0
k	10.29	9.54	0.75	10.04	9.04	1.0	10.04	9.29	0.75
difference	**0.47**	**0.22**	**0.25**	**0.47**	**0.47**	**0.0**	**0.72**	**0.47**	**0.25**

* These values were determined in the original transmission impairment tests described in chapter 3.

Table 5.1 Comparison of the C_o/N_o ratios at TOA and POF and failure margin (FM) for the gaussian noise impairment for systems *e* and *k*.

The results of the retest for the gaussian noise impairment are given in Table 5.2 for systems *h* and *l*. The form of the table is identical to that of Table 5.1. It should be noted that the values for system *h* are taken from the original transmission impairment tests described in chapter 3 of this document.

The results given in Table 5.2 show that the differences between system *h* and system *l* are small. They are, however, slightly larger than the differences shown in Table 5.1. None of the differences is greater than 1.57 dB. Again, the largest differences tend to occur at TOA. The difference in the failure margin is constant at -0.5 dB. The negative value for the failure margins indicate that they are 0.5 dB larger for the original system than for the modified system.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
*h**	10.51	9.01	1.5	9.51	8.51	1.0	10.01	9.51	0.5
l	11.33	9.33	2.0	10.83	9.33	1.5	11.58	10.58	1.0
difference	**-0.82**	**-0.32**	**-0.5**	**-1.32**	**-0.82**	**-0.5**	**-1.57**	**-1.07**	**-0.5**

* These values were determined in the original transmission impairment tests described in chapter 3.

Table 5.2 Comparison of the C_o/N_o ratios at TOA and POF and failure margin (FM) for the gaussian noise impairment for systems *h* and *l*.

As stated earlier, the matching experiments were conducted for all points which fell in the POF range in the original transmission impairment tests. The POF points given in the table represent the highest C_o/N_o ratio which fell into the POF range. The other POF points are omitted for clarity purposes since they do not alter the conclusions.

5.3.2 Co-channel interference

The results of the retest for the co-channel interference are given in Table 5.3 for systems *e* and *k*. The table provides the D/U ratios at TOA and POF, as well as the failure margin (FM). This is done for the three critical audio materials; Glockenspiel, Soprano, and Clarinet. It should be noted that the values for system *e* are taken from the original transmission impairment tests described in chapter 3 of this document. The bottom row of the table shows the *differences* between the origianl system and the modified system (i.e. *e-k*).

The results given in Table 5.3 show that the differences between the original system and the modified system are quite small. None of the differences are greater than 1.11 dB. As was the case for the gaussian noise impairment condition, the largest differences tend to occur at TOA. The differences in the failure margins are no greater than 0.5 dB for the three audio source materials.

As stated earlier, the matching experiments were conducted for all points which fell in the POF range in the original transmission impairment tests. The POF points given in the table below represent the highest D/U ratio which fell into the POF range. The other POF points are omitted for clarity purposes since they do not alter the basic conclusions.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
*e**	11.12	9.62	1.5	10.87	9.87	1.0	10.87	9.87	1.0
k	10.26	9.26	1.0	9.76	9.14	0.62	10.01	9.01	1.0
difference	**0.86**	**0.36**	**0.5**	**1.11**	**0.73**	**0.38**	**0.86**	**0.86**	**0.0**

* These values were determined in the original transmission impairment tests described in chapter 3.

Table 5.3 Comparison of the D/U ratios at TOA and POF and failure margin (FM) for the co-channel interference for systems *e* and *k*.

The results of the retest for the co-channel interference are given in Table 5.4 for systems *h* and *l*. The format of the table is identical to Table 5.3. It should again be noted that the values for system *h* are taken from the original transmission impairment tests described in chapter 3 of this document.

The results given in Table 5.4 show that the differences between the original system and the modified system are very small. None of the differences is greater than 0.83 dB. The largest difference in the failure margin is -0.5 dB which occurs for the Glockenspiel audio source material.

As stated earlier, the matching experiments were conducted for all points which fell in the POF range in the original transmission impairment tests. The POF points given in the table represent the highest D/U ratio which fell into the POF range. The other POF points are omitted for clarity purposes since they do not alter the conclusions.

	Glockenspiel			Soprano			Clarinet		
DAR System	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)	TOA (dB)	POF (dB)	FM (dB)
*h**	11.37	8.87	2.5	10.87	7.87	3.0	10.87	9.87	1.0
l	11.04	8.04	3.0	10.54	7.54	3.0	11.04	10.04	1.0
difference	**0.33**	**0.83**	**-0.5**	**0.33**	**0.33**	**0.0**	**-0.17**	**-0.17**	**0.0**

* These values were determined in the original transmission impairment tests described in chapter 3.

Table 5.4 Comparison of the D/U ratios at TOA and POF and failure margin (FM) for the co-channel interference for systems *h* and *l*.

6. Facilities

6.1 Playback system

The playback system used during the subjective tests is shown in Fig. 6.1. The particular brand name and model of equipment used in this set-up is described in Table 6.1 below.

The heart of the playback system is a custom audio workstation. It is implemented by means of a 486 personal computer (PC) equipped with a dual DSP56001 processor board and custom software. On two daughter cards, the DSP board also hosts an AES/EBU interface and a SCSI bus interface. Both the DSP board and software are developed by MPR Teltech Ltd, Burnaby, BC, Canada [4]. The workstation is also equipped with a large SCSI disk drive (1.75 Gbyte), a color VGA monitor and serial mouse. It operates in the Windows 3.1 environment.

Qty	Description
1	486-66 PC compatible with 1.75 Gbytes SCSI hard-disk and MPT Teltech Dual DSP56001 Processor Card (DSPC)
1	SVGA color video monitor
1	Spectral model ADDA 2218 D/A converter
1	Klark Teknik model DN410 parametric equalizer
1	Bryston model BP-5 professional stereo preamplifier
1	Bryston model 3B PRO professional stereo power amplifier
1	Bryston model 4B PRO professional stereo power amplifier
1	State-of-the-Art Elektronik model AAX2-2-750 active crossover system
2	State-of-the-Art Elektronik model CF 750 monitor loudspeakers
1	Stax model SRM-1/MK-2 professional headphone driver
1	Stax model ED-1 diffused field equalizer
3	Stax model Lambda Pro headphones
2	Sony model PCM-2500 DAT recorder/player
1	Panasonic SV-3700 DAT recorder/player

Table 6.1 List of equipment

The software driving the workstation consists in four different windows applications. Each of them corresponds to one of the four operations required for the preparation and the presentation of the audio materials for the listening tests. These operations are:

1. Recording of the audio materials on the audio workstation (*Record* application)
2. Synchronization of the audio files (*TimeSync* application)
3. Building of the session files (*Notepad* application)
4. Presentation of the audio materials (*ABC* application)

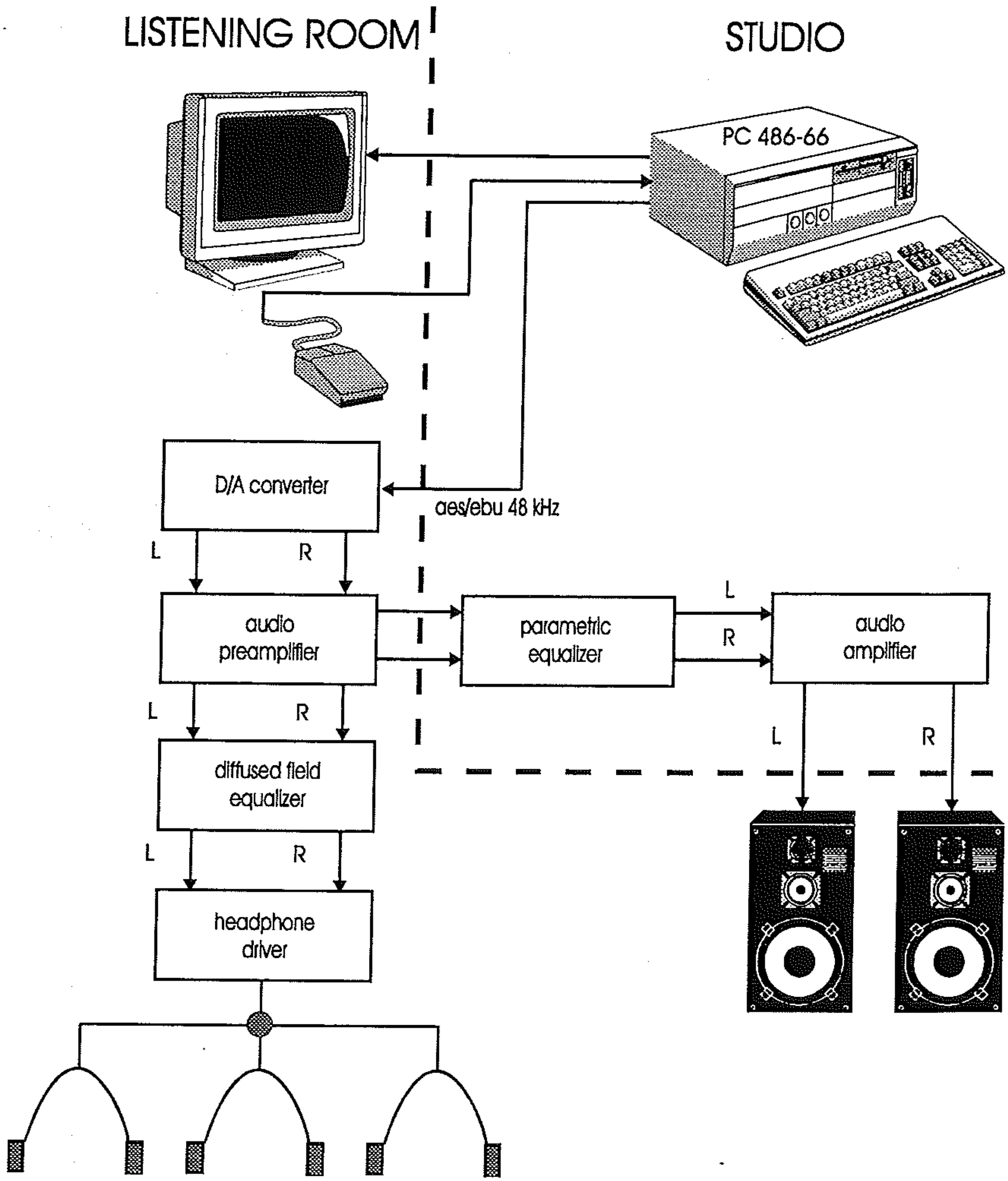


Figure 6.1 Equipment set-up

- ***Record* application**

Audio test materials were processed through the DAR systems under test at the NASA Lewis Research Centre, Cleveland, Ohio, USA and the output was recorded on DAT tapes. The DAT tapes were then shipped to the CRC where each of the audio materials required for the subjective tests (audio quality and transmission impairments) was individually transferred digitally to hard disk via the AES/EBU serial interface and stored as separate audio files with individual filenames.

- ***TimeSync* application**

Once on hard disk, the audio files containing the reference and all processed versions of each audio material are then precisely time aligned with *TimeSync*, a special software system developed at the CRC. This time synchronization process is required in order to ensure seamless switching between reference and processed versions of audio materials during the presentation.

- ***Notepad* application**

The next step is to build "session files". A trial consists of an A-B-C presentation of a given audio material and a session consists of one or more consecutive trials. A session file contains the list of audio files to be assigned to A, B and C (in the triple-stimulus A-B-C presentation) for each trial. The session file is built with any text editor that is capable of generating ASCII files, such as *Notepad* in the Windows environment.

The experimenter can subsequently re-order the audio files within a trial, or trials within a session, simply by editing a session file. Modifications are done in a few minutes compared to the many hours that an equivalent re-ordering would take to do on a DAT based playback system. This is a powerful feature of this disk-based playback system.

- ***ABC* application**

The screen used by the listener during the blind testing phase is shown in Fig. 6.2. During the training phase, there are additional buttons not seen in this figure, namely a *Session* button that is used to select a particular session file, and a *Trials* button that allows the listener to select a particular trial within the selected training session. These buttons are removed during the blind rating phase. This is to prevent the subject from accidentally changing the programmed session with the *Session* button, since this must remain under control of the experimental design. Removal of the *Trial* button prevents the listener from moving backwards through the trials during blind rating since trial order, as is true for session, must also remain under experimental design control. During blind rating, the listener can only move forwards (after spending as much time as he or she wants on a given trial) and this is done by using the *Next* button seen in Fig. 6.2. A pause button is provided, and by using the loopback button seen in Fig. 6.2, the listener can control whether playback of the trial stops at the end of each material, or whether looping is continuous. A small comment identifying the audio material being presented is also shown. This comment is added, as explained previously, in the session file for each trial.

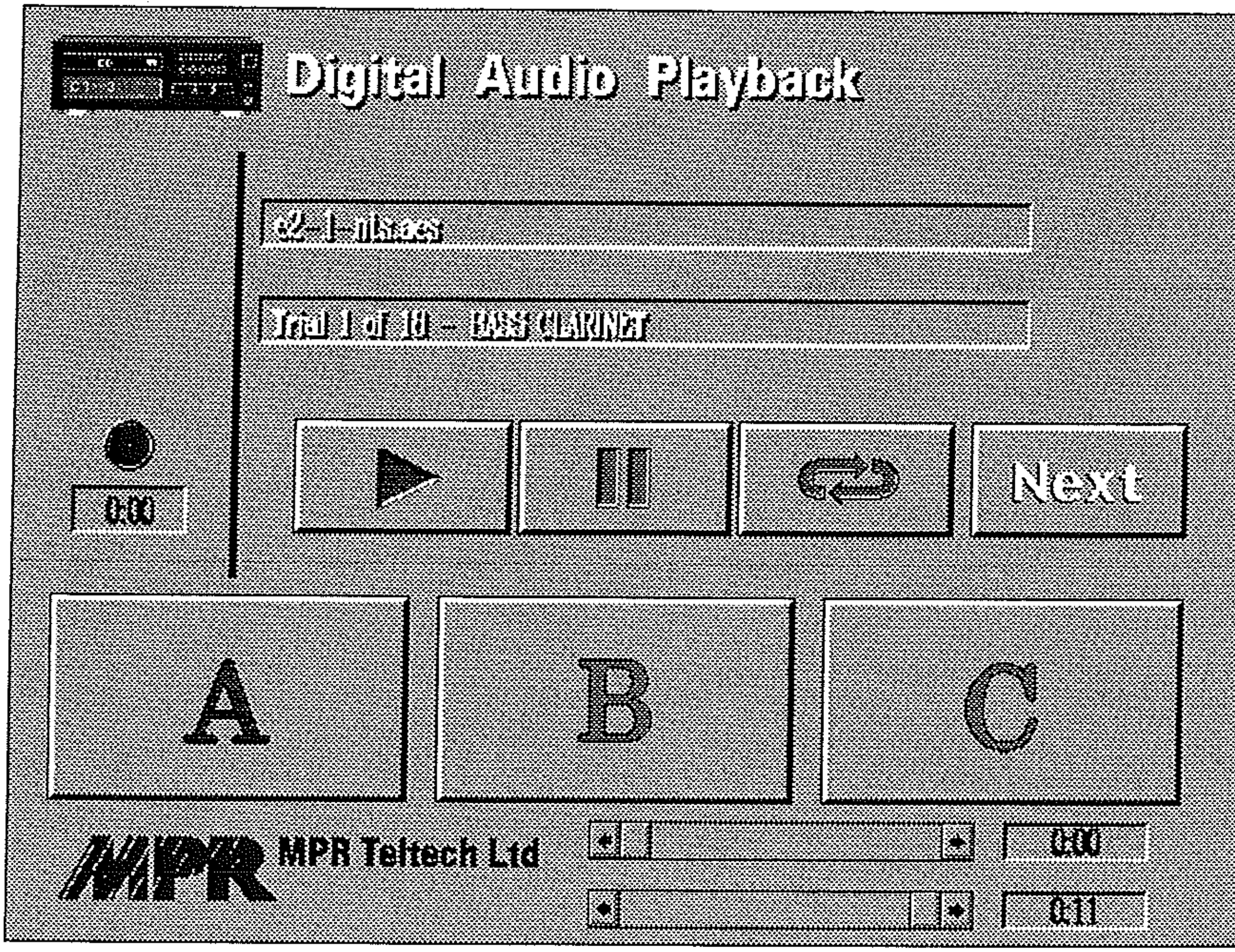


Figure 6.2 Video screen used by the listeners during blind rating sessions

Switching among A, B or C is done using anyone of the following two methods: (a) by clicking with the mouse pointer on the "A", "B" or "C" button displayed on the video screen of Fig. 6.2 or, (b) by pressing respectively the left, centre or right button of the three-button mouse (this method requires the mouse pointer to be first dragged outside any button area)

A "zoom" tool allows the listeners to listen to a smaller subsegment of the audio material they are asked to compare and assess in a given trial. The start and end time of the subsegment can be set anywhere within the audio material using the two horizontal scroll bars shown at the bottom right of Fig. 6.2. The top bar changes the start time of the subsegment while the bottom bar adjusts the end time. The continuous looping, if activated, is performed on the subsegment of the material. With this feature, listeners can focus more closely on a specific section of any material that appears to reveal suspected distortions.

Since the files are precisely time-aligned and since cross-fading is used, the switching is truly seamless in that there are no audible cues generated, and very exact continuity between the audio materials triggered by the buttons is maintained. When ready for rating, the listener clicks on the *Next* button. This action invokes a new scoring screen (not shown). Scores for B and C are entered by sliding a cursor to the desired impairment level in the corresponding scroll bar. After the *OK* button is clicked, another small window (not shown) appears requesting the listener to confirm his/her selected ratings.

One additional distinctive feature of the disk-based system is that any selected segment (A, B or C) is output digitally via the same single AES/EBU interface, and is fed through the same D/A convertor as any other selected segment. And so, all versions of the audio materials to be compared are presented through the very same pieces of hardware. This eliminates any possible contribution by hardware to differences detected by listeners among the different versions.

6.2 Listening Room

The specifications of the CRC listening room are summarized in the following three figures which show respectively the room layout (Fig. 6.3), the background noise levels (Fig. 6.4) and the reverberation time curve (Fig. 6.5). Both the background noise levels and the reverberation time curve of the room comply with the requirements of ITU-R Recommendation BS.1116 [1].

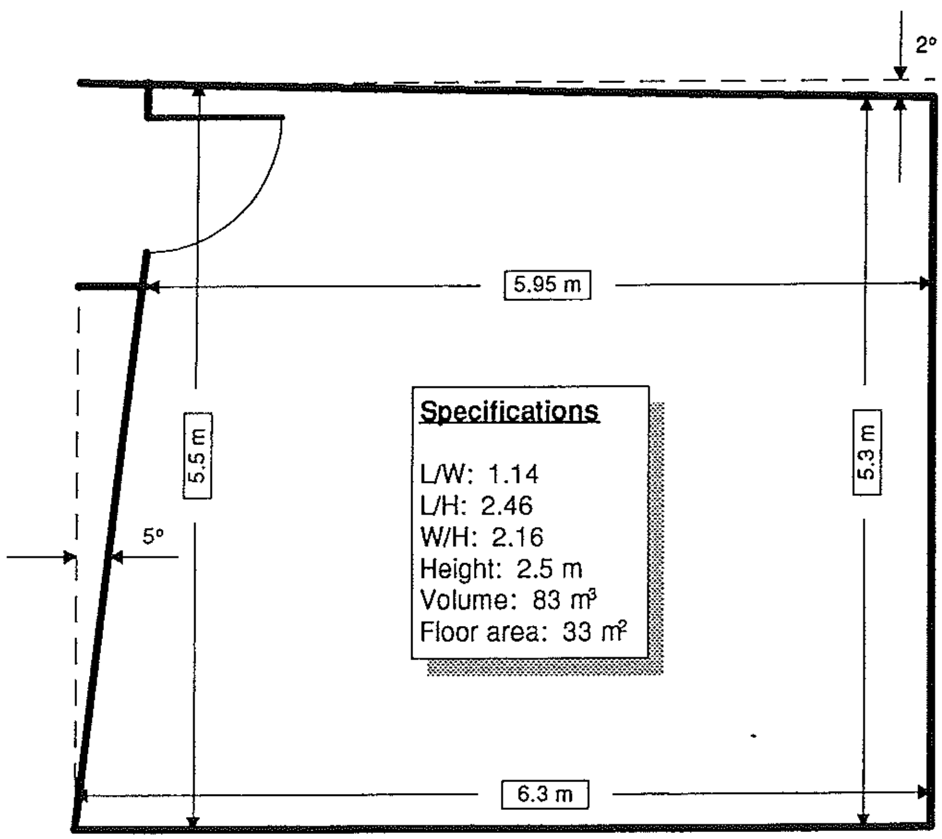


Fig. 6.3 CRC listening room layout

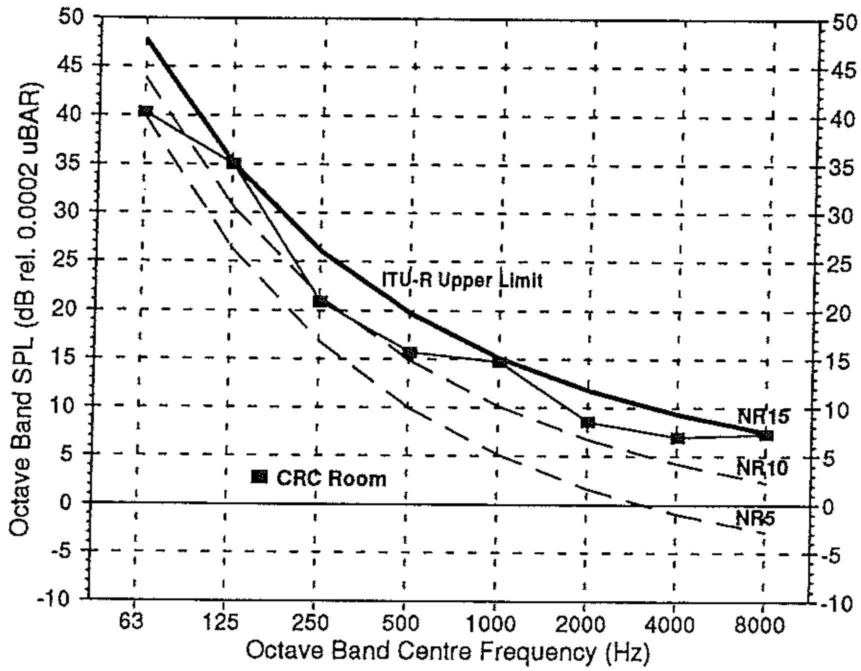


Figure 6.4 CRC listening room - Background noise levels

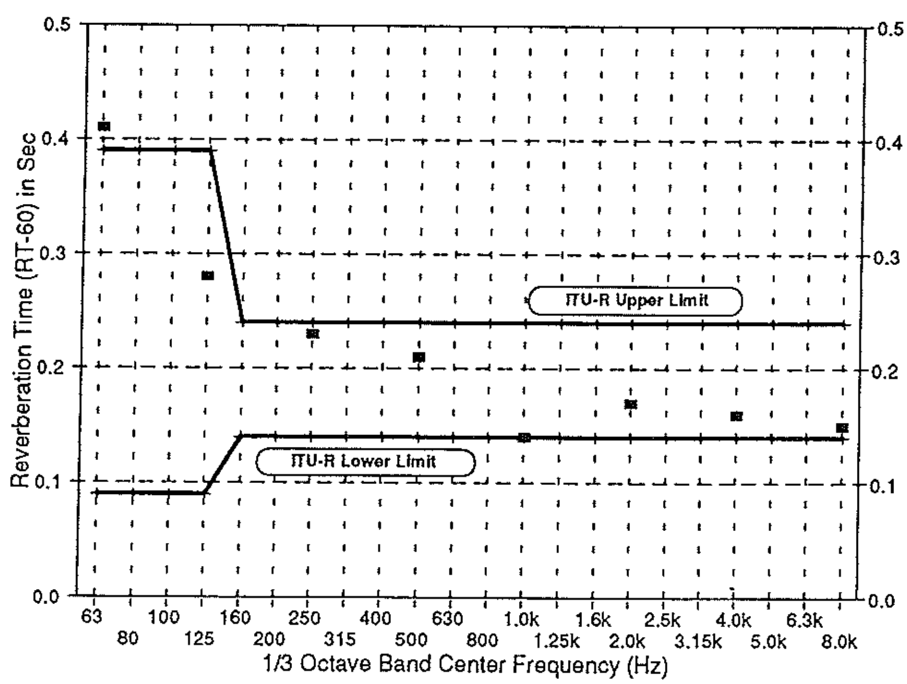


Figure 6.5 CRC listening room - Reverberation time (RT-60)

7. Acknowledgements

The authors of this report would like to gratefully acknowledge the work, support and collaboration of the following persons throughout the execution of the test described in this report:

- Mr. Ralph Justus of the EIA in Washington, USA

- Messrs. Tom Keller and Dave Londa at the NASA LeRC in Cleveland, USA

- Mr. Emil L. Torick for his participation in the selection of critical materials

The authors also wish to acknowledge the dedicated and invaluable work of Messrs. Gerald Hatton and Douglas McKinnie of the CRC who played key roles in the preparation and execution of the tests. Special thanks to Mr. Lewis Scott of the CRC for his assistance in the preparation of the various legal documents for this contract.

References

[1] *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems,* ITU-R Recommendation BS.1116, Radiocommunications Study Group, International Telecommunication Union, Geneva, Switzerland

[2] Thibault L. and Grusec T., "*EIA-DAR Listening Tests - Quality and Impairment Tests Procedures:,* Signal Processing and Psychoacoustics, Radio Broadcast Technologies Research, CRC, Ottawa, Ont., Document submitted to EIA Working Group B, Updated version, 1 December 1993

[3] Londa, Dave, et al. "*EIA-DAR/NRSC DAB objective tests of transmission impairments*".

[4] Beaton R.J. and Wong P., *A Disk-based System for the Subjective Assessment of High Quality Audio*, Preprint 3497, AES 94th Convention, Berlin, Germany, 16-19 March 1993

APPENDIX V

Subjective Assessment Report and Procedures

APPENDIX V
IBOC

Analog Compatibility Subjective Assessment Report

DAR -> Analog
F4 Co-Channel
F5 First Adjacent
F6 Second Adjacent
G1 Co-Channel with Multipath
G2 First Adjacent with Multipath
G3 Second Adjacent with Multipath

DAR -> Host Analog
L-3 Digital -> Host Analog
L-4 Digital -> Host Analog with Multipath

Enclosure #1
Instructions for the Reporting of Subjective Evaluation of Inband Compatibility Recordings

Enclosure #2
Summarized Results of the Subjective Tests

Instructions for the Reporting of Subjective Evaluation of Inband Compatibility Recordings

June 27, 1995

The tests are recorded on eight compact disks and are divided into 96 test groups. Each test group consists of one to five tests. Tests are designated on each CD by the track number. The total number of test tracks is 391. Each test has an audio reference.

Each test is identified by compact disk number and track number. The test form is further divided into sub groups that are designated alphabetically. Each test (track) within each test group represents a proponent system. Each subgroup represents one of five test receivers with a common impairment. Five DAR systems (IBOC & IBAC) were tested for the co-channel and adjacent channel interference and are recorded on disk 1 through 4. Four DAR systems were tested for the digital interference to host FM tests (IBOC) and are recorded on disk 5 through 8.

The test segments will be assessed by rating **CHANGES** between the reference and the test segment. The subjective effects of the interference will be rated using this seven point numerical rating scale.

3 Much Better

2 Better

1 Slightly Better

0 The Same

-1 Slightly Worse

-2 Worse

-3 Much Worse

Tests will be reported on the enclosed evaluation forms. The ratings are to be recorded by placing an indelible circle around one of the seven numbers for each test (track) on the test reporting evaluation form.

Four formats are used for the test recordings. No more than one test format is used for each reporting form page. The four formats are described in detail in this document.

The seven point scale is the primary assessment form. The expert observation and commentary column should only be used for your optional observations.

RECOMMENDED SUBJECTIVE ASSESSMENT PROCEDURE

1. Before starting the assessor should be familiar with the four test formats, seven point test assessment scale, and the test reporting form. It is advisable to have the description of the test format that is being used and the seven point assessment scale visible during the trials.

2. The test form designates the CD to be used, test format, and cut to be assessed. There are 22 pages of test forms for the eight CDs.

3. The tests should be conducted in a room with low noise.

4. Only the furnished Grado Model SR 225 reference headphones will be used.

5. It is advisable to have a CD player with programmable track selection. The number of tracks for each of the eight CDs will range from 32 to 64.

6. The expert listener should sign the cover of the test report and indicate the dates of the trials.

SAMPLE TEST

1. Place CD #1 in machine and go to track #1.

2. Select the reporting form that has the corresponding CD number and track number (page #1).

3. At the top of this reporting form, format #1 is indicated. The format should be thoroughly understood prior to the trials.

4. Play track #1 and listen for the second harp (cut #2 reference). Note any changes in the third harp (cut #3 test) on the 7 point scale by placing a circle around the appropriate number. See the description of the numeric scale on the previous page.

5. If needed, comments may be made in the Expert Observation and Commentary Column.

6. Continue to CD track #2.

NOTE: Test format is not the same on all evaluation form!

Format #1

Test Format #1: Each track is approximately 60 seconds long and consists of four cuts (Cut #4 will not be used).

> Cut #1 (15 sec.) is a clear FM channel and may be used to recognize FM receiver audio artifacts. The use of this cut is optional.
>
> Cut #2 (15 sec.) **REFERENCE** is FM -> FM interference set for an audio S/N of 45 dB.
>
> Cut #3 **TEST** (15 sec.) uses the same D/U ratios used in cut #2 with the DAR signal replacing the undesired FM signal.
>
> Cut #4 is **NOT TO BE USED.**

FORMAT

Test audio - harp

Track Number

Cut #1 (0–15)	Cut #2 (15–30)	Cut #3 (30–45)	Cut #4 (45–60)
Test receiver without interference	FM -> FM reference	DAR -> FM segment to be assessed	DAR -> FM **Disregard**

Format #1 Tracks and Locations

CD Number	Track Number	Test Group Number
1	1 through 25	A through E
2	1 through 25	A through E
3	1 through 20	A through D
4	1 through 14 & 32 through 47	A through E & M through P

Format #2

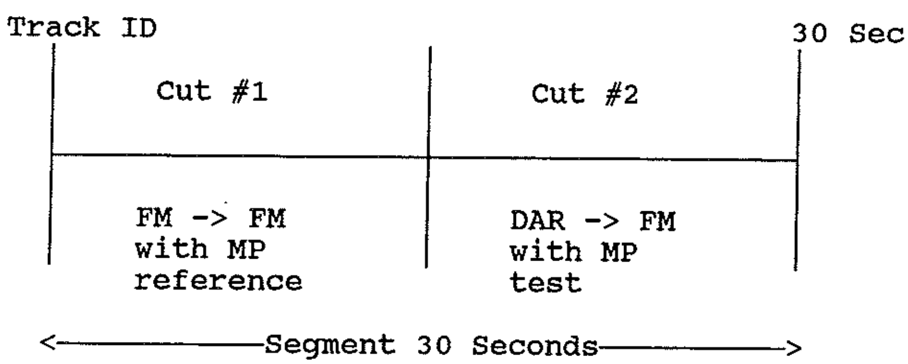
Test Format #2: Each track is approximately 30 seconds long and consists of two cuts with multipath:

Cut #1 **REFERENCE** (15 sec.) is FM -> FM with multipath and no interference. This is the reference cut.

Cut #2 **TEST** is for the subjective evaluation of DAR -> FM interference with multipath and the D/U set at the RF levels found in the corresponding FM -> FM objective test.

FORMAT

Test audio - harp



Format #2 Tracks and Locations

CD Number	Track Number	Test Group Number
1	26 through 60	F through L
2	26 through 60	F through L
3	21 through 45	E through I
4	15 through 31 & 48 through 64	F through L & Q through U

Format #3

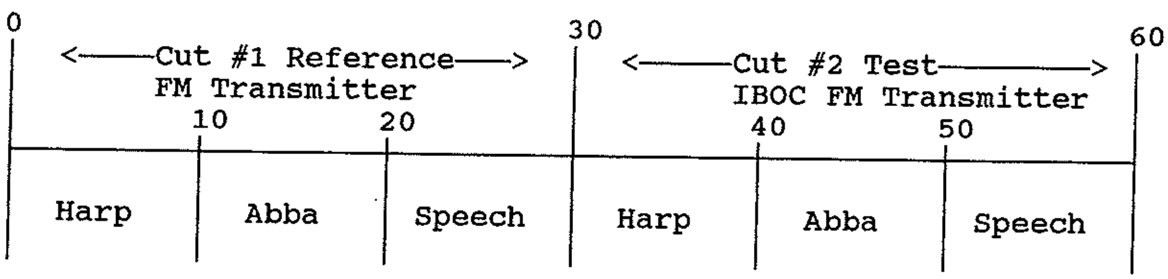
Test Format #3: Each track is approximately 60 seconds long and consists of two cuts.

Cut #1 **REFERENCE** (30 sec.) The reference is all three 10 second audio materials transmitted through the reference FM transmitter.

Cut #2 **TEST** (30 sec.) The test uses the three 10 second audio materials used in the reference with the addition of the IBOC digital signal. The rating of cut #2 should be done for all three audio segments as a single assessment.

FORMAT

Test audio - harp, Abba, and speech



Format #3 Tracks and Locations

CD Number	Track Number	Test Group Number
5	1 through 50	A through N

Note: Due to an artifact in the multipath simulator the first few seconds of the reference harp is lost in a receiver noise burst or a mute. The last part of harp, Abba, and speech are not effected. The test cut is not effected.

Format #4

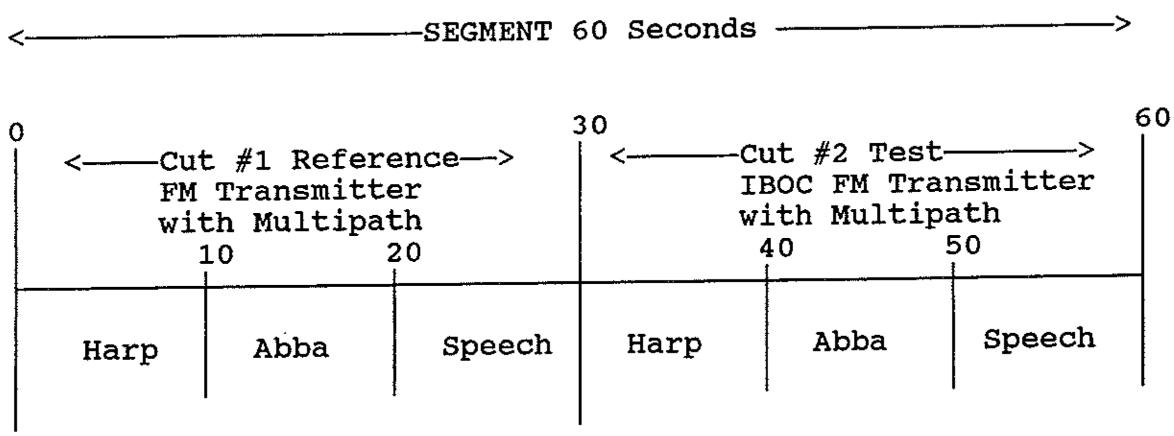
Test Format #4: Each track is approximately 60 seconds long and consists of two cuts with multipath.

Cut #1 **REFERENCE** is the FM transmitter with multipath.

Cut #2 **TEST** is the IBOC FM transmitter with multipath.

FORMAT

Test audio - harp, Abba, and speech



Format #4 Tracks and Locations

CD Number	Track Number	Test Group Number
6	1 through 40	A through J
7	1 through 40	A through J
8	1 through 32	A through H

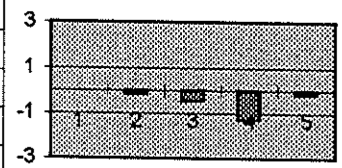
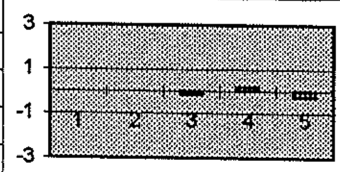
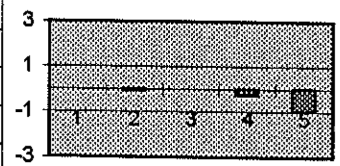
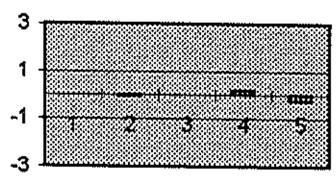
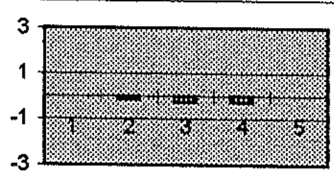
Note: Due to an artifact in the multipath simulator the first few seconds of the reference harp is lost in a receiver noise burst or a mute. The last part of harp, Abba, and speech are not effected. The test cut is not effected.

Subjective Rating Scale	
3	Much Better
2	Better
1	Slightly Better
0	The Same
-1	Slightly Worse
-2	Worse
-3	Much Worse

Subjective Tests Data

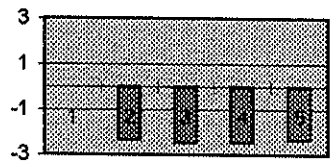
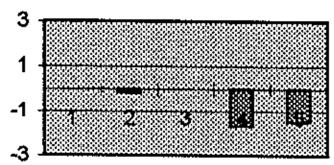
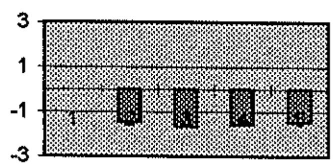
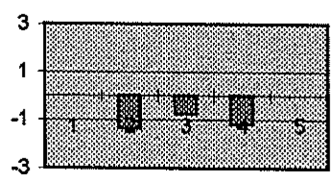
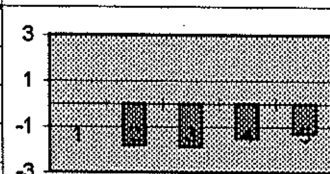
CO-CHANNEL DIGITAL TO ANALOG SUBJECTIVE

TEST F4

RECEIVER	CD #1 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#4	DSB	-0.2	
REFERENCE	#5	LSB	-0.5	
F1 45 dB S/N	#2	FM-1	-1.3	
D/U 36.17 dB	#3	FM-2	-0.2	
#2 DENON				
	#8	DSB	0	
REFERENCE	#7	LSB	-0.2	
F1 45 dB S/N	#9	FM-1	0.2	
D/U 43.39 dB	#10	FM-2	-0.3	
#3 PANASONIC				
	#12	DSB	-0.1	
REFERENCE	#15	LSB	0	
F1 45 dB S/N	#13	FM-1	-0.3	
D/U 40.94 dB	#14	FM-2	-1	
#4 PIONEER				
	#17	DSB	-0.1	
REFERENCE	#16	LSB	0	
F1 45 dB S/N	#19	FM-1	0.2	
D/U 44.18 dB	#20	FM-2	-0.3	
#5 FORD				
	#23	DSB	-0.2	
REFERENCE	#22	LSB	-0.3	
F1 45 dB S/N	#24	FM-1	-0.3	
D/U 35.22 dB	#25	FM-2	0	

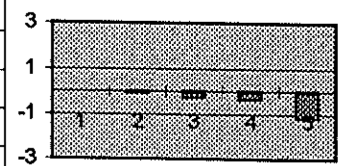
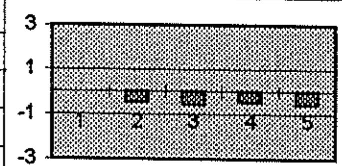
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

CO-CHANNEL DIGITAL TO ANALOG SUBJECTIVE
URBAN SLOW MULTIPATH
TEST G1

RECEIVER	CD #1 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#26	DSB	-2.4	
REFERENCE	#27	LSB	-2.5	
F1 45 dB S/N	#29	FM-1	-2.5	
D/U 36.17 dB	#30	FM-2	-2.4	
#2 DENON				
	#33	DSB	-0.2	
REFERENCE	#34	LSB	0	
F1 45 dB S/N	#32	FM-1	-1.7	
D/U 43.39 dB	#31	FM-2	-1.5	
#3 PANASONIC				
	#37	DSB	-1.5	
REFERENCE	#36	LSB	-1.7	
F1 45 dB S/N	#39	FM-1	-1.6	
D/U 40.94 dB	#40	FM-2	-1.5	
#4 PIONEER				
	#43	DSB	-1.4	
REFERENCE	#44	LSB	-0.8	
F1 45 dB S/N	#42	FM-1	-1.2	
D/U 44.18 dB	#41	FM-2	0	
#5 FORD				
	#48	DSB	-1.8	
REFERENCE	#49	LSB	-1.9	
F1 45 dB S/N	#47	FM-1	-1.5	
D/U 35.22 dB	#46	FM-2	-1.3	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

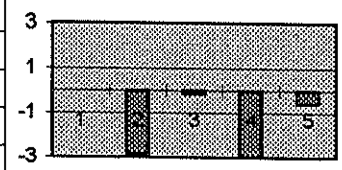
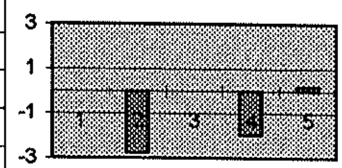
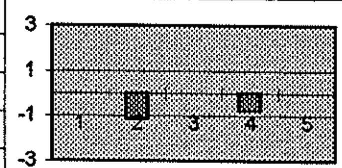
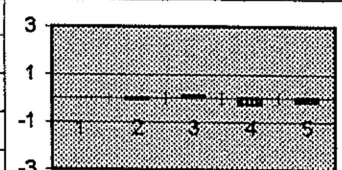
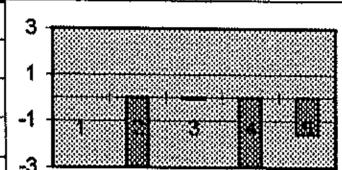
CO-CHANNEL DIGITAL TO ANALOG SUBJECTIVE
URBAN FAST MULTIPATH
TEST G1

RECEIVER	CD #1 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#54	DSB	-0.1	
REFERENCE	#53	LSB	-0.3	
F1 45 dB S/N	#52	FM-1	-0.4	
D/U 36.17 dB	#51	FM-2	-1.2	
#5 FORD				
	#58	DSB	-0.5	
REFERENCE	#57	LSB	-0.6	
F1 45 dB S/N	#59	FM-1	-0.5	
D/U 35.22 dB	#60	FM-2	-0.6	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

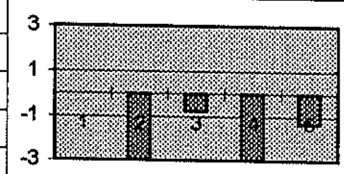
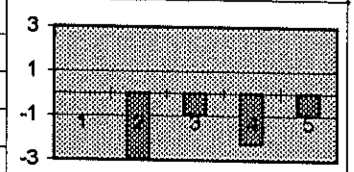
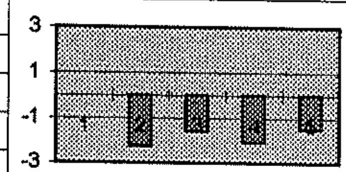
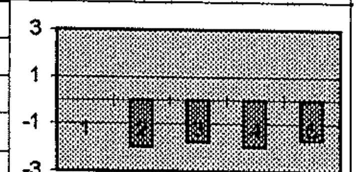
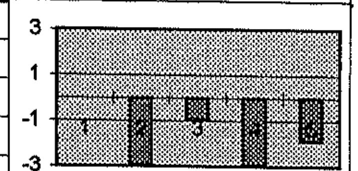
LOWER 1ST ADJACENT DIGITAL TO ANALOG SUBJECTIVE

TEST F5

RECEIVER	CD #2 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#2	DSB	-2.9	
REFERENCE	#3	LSB	-0.2	
F2L 45 dB S/N	#4	FM-1	-3	
D/U 4.09 dB	#5	FM-2	-0.6	
#2 DENON				
	#6	DSB	-2.8	
REFERENCE	#8	LSB	0	
F2L 45 dB S/N	#9	FM-1	-2	
D/U 23.61 dB	#10	FM-2	0.2	
#3 PANASONIC				
	#11	DSB	-1.2	
REFERENCE	#13	LSB	0	
F2L 45 dB S/N	#14	FM-1	-0.8	
D/U 27.33 dB	#15	FM-2	0	
#4 PIONEER				
	#16	DSB	-0.1	
REFERENCE	#18	LSB	0.1	
F2L 45 dB S/N	#19	FM-1	-0.3	
D/U 31.87 dB	#20	FM-2	-0.2	
#5 FORD				
	#21	DSB	-3	
REFERENCE	#23	LSB	-0.1	
F2L 45 dB S/N	#24	FM-1	-3	
D/U -6.18 dB	#25	FM-2	-1.6	

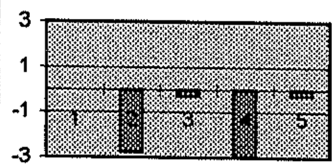
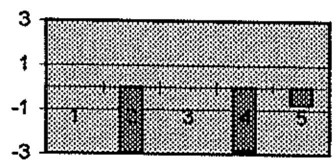
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

LOWER 1ST ADJACENT DIGITAL TO ANALOG SUBJECTIVE
URBAN SLOW MULTIPATH
TEST G2

RECEIVER	CD #2 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#30	DSB	-3	
REFERENCE	#28	LSB	-0.8	
F2L 45 dB S/N	#27	FM-1	-3	
D/U 4.09 dB	#26	FM-2	-1.4	
#2 DENON				
	#35	DSB	-3	
REFERENCE	#33	LSB	-1	
F2L 45 dB S/N	#32	FM-1	-2.3	
D/U 23.61 dB	#31	FM-2	-1	
#3 PANASONIC				
	#40	DSB	-2.3	
REFERENCE	#38	LSB	-1.6	
F2L 45 dB S/N	#37	FM-1	-2.1	
D/U 27.33 dB	#36	FM-2	-1.5	
#4 PIONEER				
	#43	DSB	-2	
REFERENCE	#41	LSB	-1.8	
F2L 45 dB S/N	#44	FM-1	-2	
D/U 31.87 dB	#45	FM-2	-1.7	
#5 FORD				
	#50	DSB	-3	
REFERENCE	#48	LSB	-1	
F2L 45 dB S/N	#47	FM-1	-3	
D/U -6.18 dB	#46	FM-2	-1.9	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

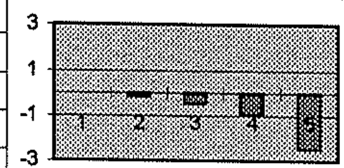
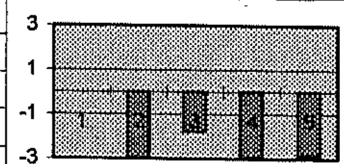
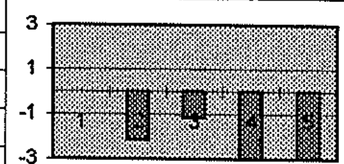
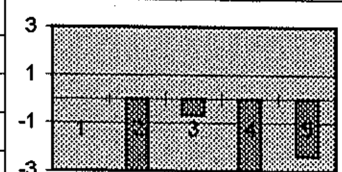
LOWER 1ST ADJACENT DIGITAL TO ANALOG SUBJECTIVE
URBAN FAST MULTIPATH
TEST G2

RECEIVER	CD #2 TRACK	PROPONENT	TEST SCOR	SUBJECTIVE SCORE
#1 DELCO				
	#51	DSB	-2.8	
REFERENCE	#55	LSB	-0.3	
F2L 45 dB S/N	#53	FM-1	-3	
D/U 4.09 dB	#54	FM-2	-0.3	
#5 FORD				
	#56	DSB	-3	
REFERENCE	#58	LSB	0	
F2L 45 dB S/N	#59	FM-1	-2.9	
D/U -6.18 dB	#60	FM-2	-0.8	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

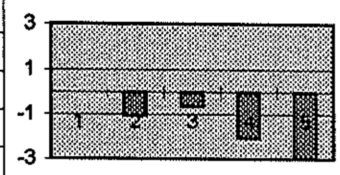
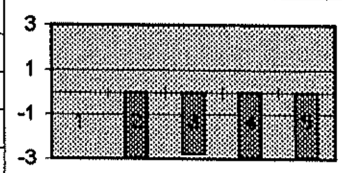
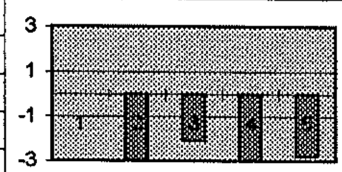
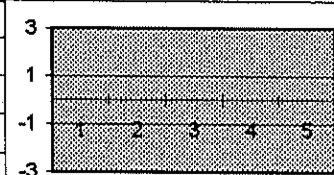
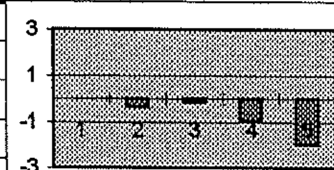
LOWER 2ND ADJACENT DIGITAL TO ANALOG SUBJECTIVE

TEST F6

RECEIVER	CD #3 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#1	DSB	-0.2	
REFERENCE	#3	LSB	-0.5	
F3L 45 dB S/N	#4	FM-1	-1	
D/U -24.17 dB	#5	FM-2	-2.5	
#2 DENON				
	#6	DSB	-3	
REFERENCE	#8	LSB	-1.8	
F3L 45 dB S/N	#9	FM-1	-3	
D/U -24.67 dB	#10	FM-2	-2.9	
#3 PANASONIC				
	#13	DSB	-2.2	
REFERENCE	#11	LSB	-1.2	
F3L 45 dB S/N	#14	FM-1	-3	
D/U -22.41 d	#15	FM-2	-3	
#4 PIONEER				
	#18	DSB	-3	
REFERENCE	#16	LSB	-0.7	
F3L 45 dB S/N	#19	FM-1	-3	
D/U -15.16 dB	#20	FM-2	-2.4	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

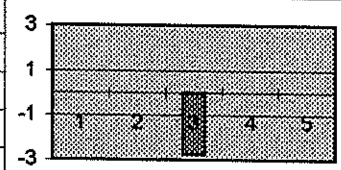
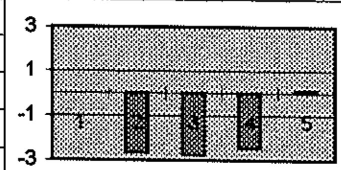
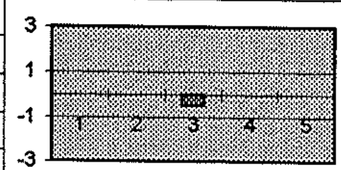
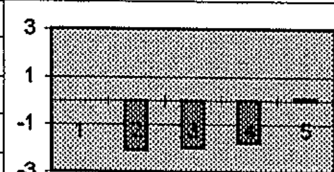
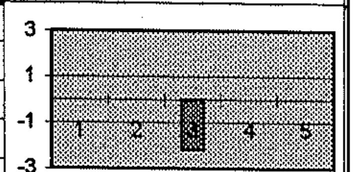
LOWER 2ND ADJACENT DIGITAL TO ANALOG SUBJECTIVE
URBAN SLOW MULTIPATH
TEST G3

RECEIVER	CD #3 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#23	DSB	-1.1	
REFERENCE	#21	LSB	-0.7	
F3L 45 dB S/N	#24	FM-1	-2.1	
D/U -24.17 dB	#25	FM-2	-3	
#2 DENON				
	#30	DSB	-3	
REFERENCE	#28	LSB	-2.8	
F3L 45 dB S/N	#27	FM-1	-3	
D/U -24.67 dB	#26	FM-2	-3	
#3 PANASONIC				
	#33	DSB	-3	
REFERENCE	#35	LSB	-2.1	
F3L 45 dB S/N	#32	FM-1	-3	
D/U -22.41 dB	#31	FM-2	-2.8	
#4 PIONEER				
	#38	DSB	0	
REFERENCE	#40	LSB	0	
F3L 45 dB S/N	#37	FM-1	0	
D/U -15.16 dB	#36	FM-2	0	
#1 DELCO				
UPPER 1ST U.F	#41	DSB	-0.4	
REFERENCE	#43	LSB	-0.2	
F1U 45 dB S/	#44	FM-1	-1	
D/U 5.41 d	#45	FM-2	-2	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

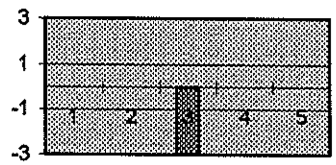
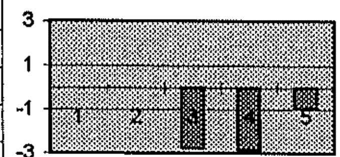
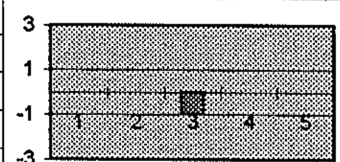
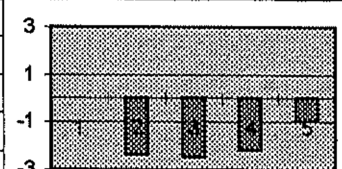
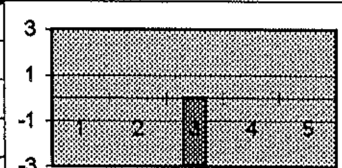
UPPER 1ST ADJACENT DIGITAL TO ANALOG SUBJECTIVE

TEST F4

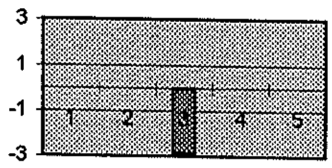
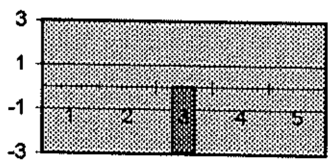
RECEIVER	CD #4 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
REFERENCE	#1	LSB	-2.8	
F2U 45 dB S/N				
D/U 5.41 dB				
#2 DENON				
	#2	DSB	-2.7	
REFERENCE	#4	LSB	-2.8	
F2U 45 dB S/N	#5	FM-1	-2.5	
D/U 12.46 dB	#6	FM-2	0.1	
#3 PANASONIC				
REFERENCE	#7	LSB	-0.5	
F2U 45 dB S/N				
D/U 27.19 dB				
#4 PIONEER				
	#10	DSB	-2.1	
REFERENCE	#8	LSB	-2	
F2U 45 dB S/N	#11	FM-1	-1.8	
D/U 21.33 dB	#12	FM-2	0.1	
#5 FORD				
REFERENCE	#14	LSB	-2.2	
F2U 45 dB S/N				
D/U -6.12 dB				

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

UPPER 1ST ADJACENT DIGITAL TO ANALOG SUBJECTIVE
URBAN SLOW MULTIPATH
TEST G2

RECEIVER	CD #4 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
REFERENCE	#15	LSB	-3	
F3U 45 dB S/N				
D/U 5.41 dB				
#2 DENON				
	#20	DSB	0	
REFERENCE	#18	LSB	-2.8	
F3U 45 dB S/N	#17	FM-1	-2.9	
D/U 12.46 dB	#16	FM-2	-1	
#3 PANASONIC				
REFERENCE	#21	LSB	-1	
F3U 45 dB S/N				
D/U 27.19 dB				
#4 PIONEER				
	#24	DSB	-2.4	
REFERENCE	#26	LSB	-2.5	
F3U 45 dB S/N	#23	FM-1	-2.2	
D/U 21.33 dB	#22	FM-2	-1	
#5 FORD				
REFERENCE	#27	LSB	-3	
F3U 45 dB S/N				
D/U -6.12 dB				

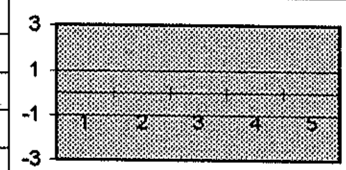
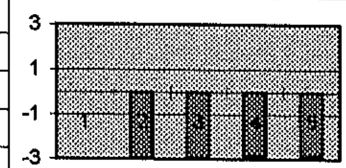
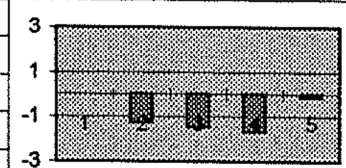
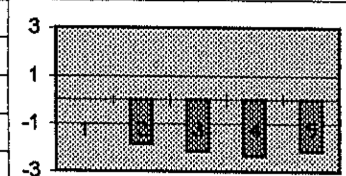
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

UPPER 1ST ADJACENT DIGITAL TO ANALOG SUBJECTIVE URBAN FAST MULTIPATH TEST G2				
RECEIVER	CD #4 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO REFERENCE F2U 45 dB S/N D/U 5.41 dB	#29	LSB	-2.9	
#5 FORD REFERENCE F2U 45 dB S/N D/U -6.12 dB	#31	LSB	-3	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

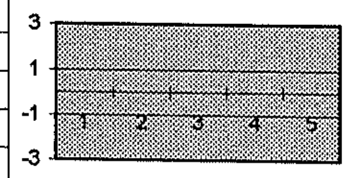
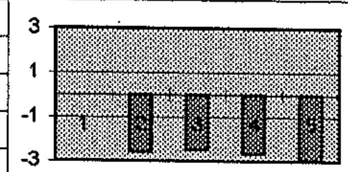
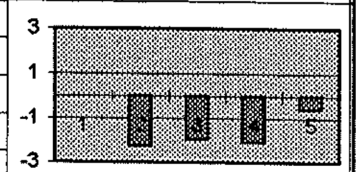
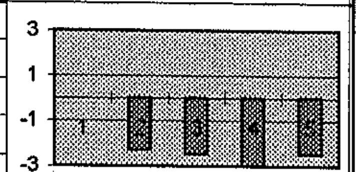
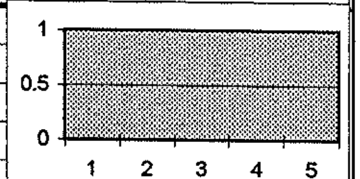
UPPER 2ND ADJACENT DIGITAL TO ANALOG SUBJECTIVE

TEST F6

RECEIVER	CD #4 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
REFERENCE				
F3U 45 dB S/N				
D/U -24.17 dB				
#2 DENON				
	#33	DSB	-3	
REFERENCE	#35	LSB	-3	
F3U 45 dB S/N	#36	FM-1	-2.9	
D/U -33.18 dB	#37	FM-2	-2.9	
#3 PANASONIC				
	#38	DSB	-1.3	
REFERENCE	#40	LSB	-1.5	
F3U 45 dB S/	#41	FM-1	-1.7	
D/U 2.16 d	#42	FM-2	-0.2	
#4 PIONEER				
	#45	DSB	-1.9	
REFERENCE	#43	LSB	-2.2	
F3U 45 dB S/N	#46	FM-1	-2.4	
D/U -14.92 dB	#47	FM-2	-2.2	

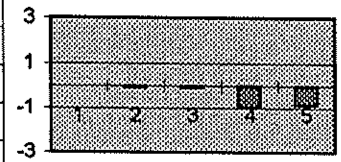
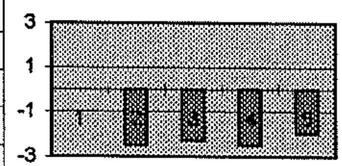
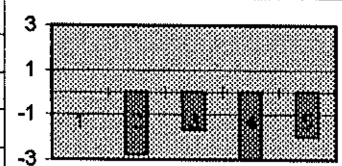
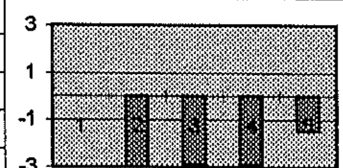
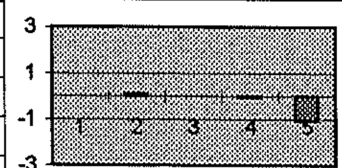
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

UPPER 2ND ADJACENT DIGITAL TO ANALOG SUBJECTIVE
URBAN SLOW MULTIPATH
TEST G3

RECEIVER	CD #4 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
REFERENCE				
F3U 45 dB S/N				
D/U -24.17 dB				
#2 DENON				
	#53	DSB	-2.6	
REFERENCE	#51	LSB	-2.5	
F3U 45 dB S/N	#50	FM-1	-2.7	
D/U -33.18 dB	#49	FM-2	-3	
#3 PANASONIC				
	#58	DSB	-2.3	
REFERENCE	#56	LSB	-2	
F3U 45 dB S/	#55	FM-1	-2.1	
D/U 2.16 d	#54	FM-2	-0.6	
#4 PIONEER				
	#61	DSB	-2.3	
REFERENCE	#63	LSB	-2.5	
F3U 45 dB S/N	#60	FM-1	-3	
D/U -14.92 dB	#59	FM-2	-2.5	
#1 DELCO				
URBAN FAST				
REFERENCE				
F3U 45 dB S/N				
D/U -46.18 dB				

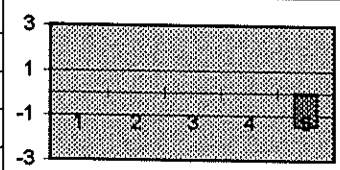
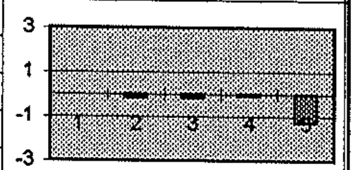
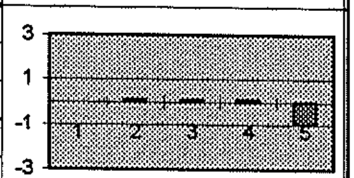
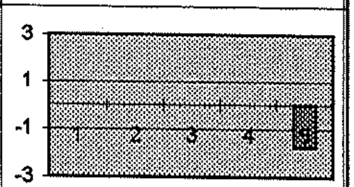
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
STRONG SIGNAL
TEST L3

RECEIVER	CD #5 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#4	DSB	-0.1	
	#1	LSB	-0.1	
	#2	FM-1	-1	
	#3	FM-2	-0.9	
#2 DENON				
	#8	DSB	-2.5	
	#5	LSB	-2.3	
	#6	FM-1	-2.5	
	#7	FM-2	-2	
#3 PANASONIC				
	#9	DSB	-2.8	
	#12	LSB	-1.7	
	#10	FM-1	-3	
	#11	FM-2	-2	
#4 PIONEER				
	#13	DSB	-3	
	#16	LSB	-2.9	
	#14	FM-1	-2.9	
	#15	FM-2	-1.5	
#5 FORD				
	#17	DSB	0.1	
	#20	LSB	0	
	#18	FM-1	-0.1	
	#19	FM-2	-1.1	

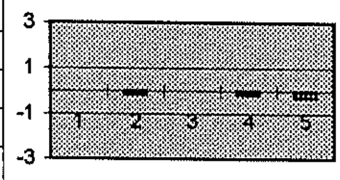
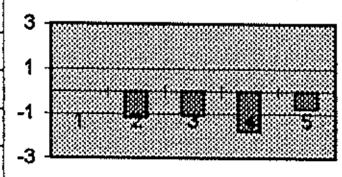
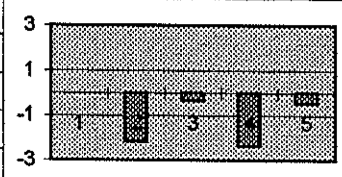
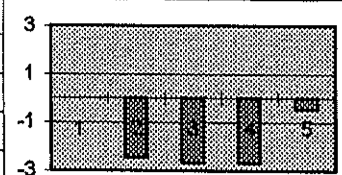
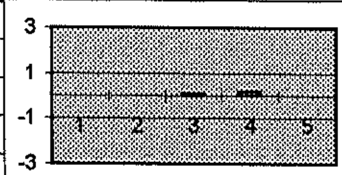
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
STRONG SIGNAL
TEST L3

RECEIVER	CD #5 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#4 PIONEER SCA GROUP B				
	#21	FM-2	-1.5	
#5 FORD SCA GROUP B				
	#22	DSB	-0.2	
	#25	LSB	-0.2	
	#23	FM-1	-0.1	
	#24	FM-2	-1.3	
#5 FORD SCA GROUP A				
	#26	DSB	0.1	
	#29	LSB	0.1	
	#27	FM-1	0.1	
	#28	FM-2	-1	
#4 PIONEER SCA GROUP A				
	#30	FM-2	-1.8	

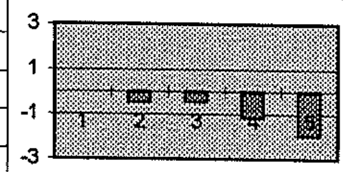
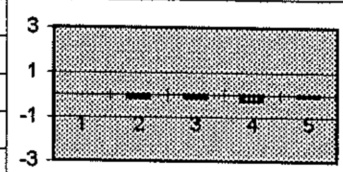
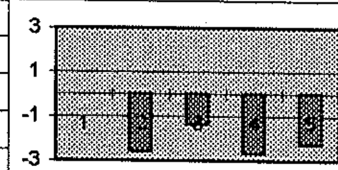
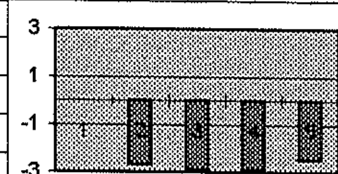
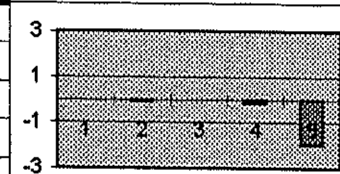
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
WEAK SIGNAL
TEST L3

RECEIVER	CD #5 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#34	DSB	-0.2	
	#31	LSB	0	
	#32	FM-1	-0.2	
	#33	FM-2	-0.3	
#2 DENON				
	#38	DSB	-1.2	
	#35	LSB	-1.1	
	#36	FM-1	-1.8	
	#37	FM-2	-0.8	
#3 PANASONIC				
	#39	DSB	-2.2	
	#42	LSB	-0.4	
	#40	FM-1	-2.4	
	#41	FM-2	-0.5	
#4 PIONEER				
	#43	DSB	-2.5	
	#46	LSB	-2.7	
	#44	FM-1	-2.7	
	#45	FM-2	-0.5	
#5 FORD				
	#47	DSB	0	
	#50	LSB	0.1	
	#48	FM-1	0.2	
	#49	FM-2	0	

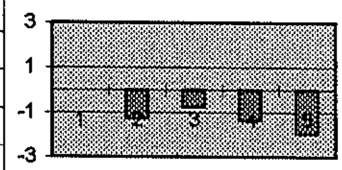
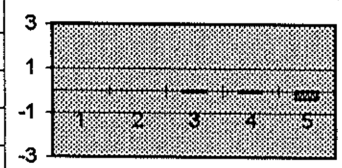
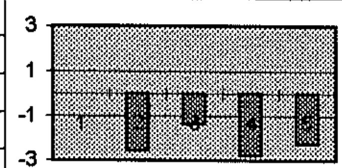
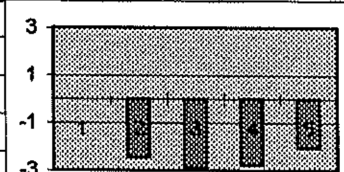
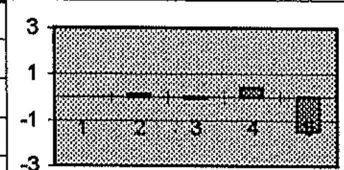
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
STRONG SIGNAL URBAN SLOW M.P.
TEST L4

RECEIVER	CD #6 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#1	DSB	-0.5	
	#4	LSB	-0.5	
	#2	FM-1	-1.2	
	#3	FM-2	-2	
#2 DENON				
	#5	DSB	-0.2	
	#8	LSB	-0.2	
	#6	FM-1	-0.3	
	#7	FM-2	-0.1	
#3 PANASONIC				
	#9	DSB	-2.6	
	#12	LSB	-1.4	
	#10	FM-1	-2.7	
	#11	FM-2	-2.3	
#4 PIONEER				
	#13	DSB	-2.7	
	#16	LSB	-2.9	
	#14	FM-1	-2.9	
	#15	FM-2	-2.5	
#5 FORD				
	#17	DSB	-0.1	
	#20	LSB	0	
	#18	FM-1	-0.2	
	#19	FM-2	-2	

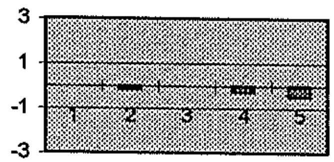
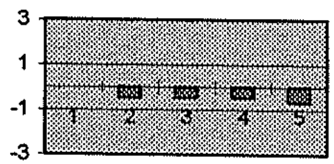
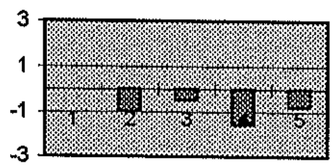
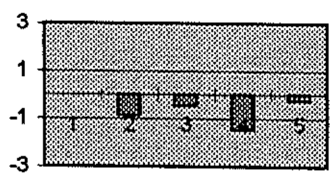
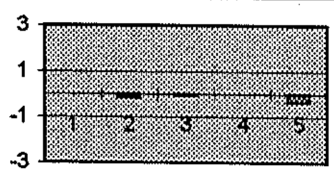
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
STRONG SIGNAL URBAN SLOW SCA GROUP B
TEST L4

RECEIVER	CD #6 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#21	DSB	-1.3	
	#24	LSB	-0.8	
	#22	FM-1	-1.4	
	#23	FM-2	-2	
#2 DENON				
	#25	DSB	0	
	#28	LSB	-0.1	
	#26	FM-1	-0.1	
	#27	FM-2	-0.4	
#3 PANASONIC				
	#29	DSB	-2.6	
	#32	LSB	-1.4	
	#30	FM-1	-2.8	
	#31	FM-2	-2.3	
#4 PIONEER				
	#33	DSB	-2.5	
	#36	LSB	-2.9	
	#34	FM-1	-2.8	
	#35	FM-2	-2.1	
#5 FORD				
	#37	DSB	0.1	
	#40	LSB	-0.1	
	#38	FM-1	0.4	
	#39	FM-2	-1.5	

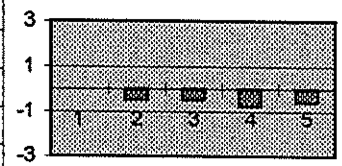
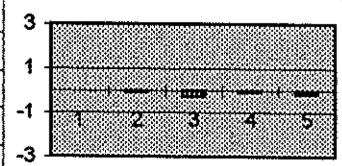
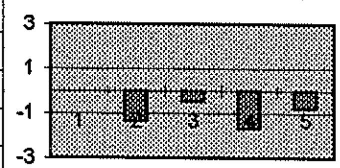
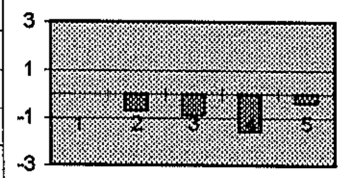
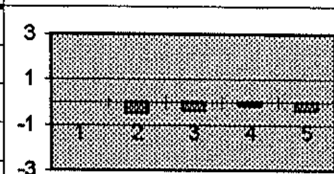
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
WEAK SIGNAL URBAN SLOW M.P.
TEST L4

RECEIVER	CD #7 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#4	DSB	-0.2	
	#1	LSB	0	
	#2	FM-1	-0.3	
	#3	FM-2	-0.5	
#2 DENON				
	#8	DSB	-0.5	
	#5	LSB	-0.5	
	#6	FM-1	-0.5	
	#7	FM-2	-0.7	
#3 PANASONIC				
	#12	DSB	-1	
	#9	LSB	-0.5	
	#10	FM-1	-1.6	
	#11	FM-2	-0.8	
#4 PIONEER				
	#16	DSB	-0.9	
	#13	LSB	-0.5	
	#14	FM-1	-1.5	
	#15	FM-2	-0.3	
#5 FORD				
	#20	DSB	-0.2	
	#17	LSB	-0.1	
	#18	FM-1	0	
	#19	FM-2	-0.4	

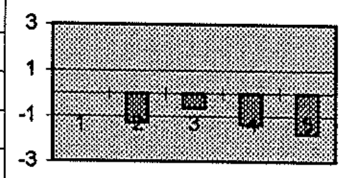
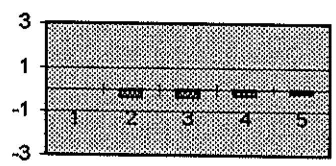
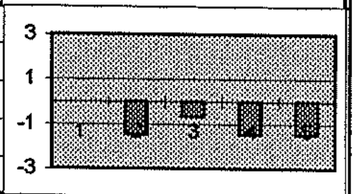
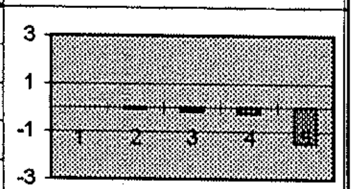
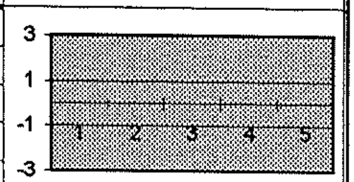
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
WEAK SIGNAL URBAN SLOW SCA GROUP B
TEST L4

RECEIVER	CD #7 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#24	DSB	-0.5	
	#21	LSB	-0.5	
	#22	FM-1	-0.8	
	#23	FM-2	-0.6	
#2 DENON				
	#28	DSB	-0.1	
	#25	LSB	-0.3	
	#26	FM-1	-0.1	
	#27	FM-2	-0.2	
#3 PANASONIC				
	#32	DSB	-1.36	
	#29	LSB	-0.5	
	#30	FM-1	-1.7	
	#31	FM-2	-0.8	
#4 PIONEER				
	#36	DSB	-0.7	
	#33	LSB	-0.9	
	#34	FM-1	-1.6	
	#35	FM-2	-0.4	
#5 FORD				
	#40	DSB	-0.5	
	#37	LSB	-0.4	
	#38	FM-1	-0.2	
	#39	FM-2	-0.4	

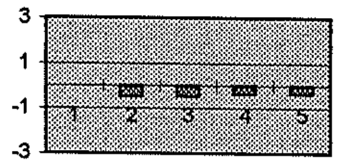
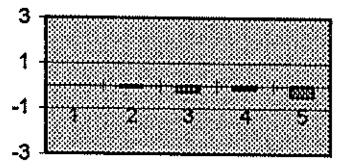
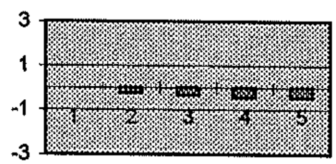
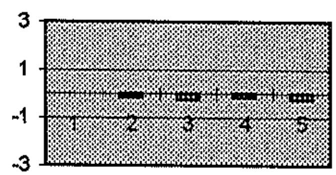
LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
STRONG SIGNAL URBAN FAST M.P.
TEST L4

RECEIVER	CD #8 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#1	DSB	-1.3	
	#4	LSB	-0.7	
	#2	FM-1	-1.4	
	#3	FM-2	-1.8	
#5 FORD				
	#5	DSB	-0.4	
	#8	LSB	-0.4	
	#6	FM-1	-0.3	
	#7	FM-2	-0.2	
#1 DELCO				
SCA GROUP B	#9	DSB	-1.5	
	#12	LSB	-0.7	
	#10	FM-1	-1.5	
	#11	FM-2	-1.5	
#5 FORD				
SCA GROUP B	#13	DSB	-0.1	
	#16	LSB	-0.2	
	#14	FM-1	-0.3	
	#15	FM-2	-1.5	
				

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

DAR TO HOST IBOC ANALOG SUBJECTIVE
WEAK SIGNAL URBAN FAST M.P.
TEST L4

RECEIVER	CD #8 TRACK	PROPONENT	TEST SCORE	SUBJECTIVE SCORE
#1 DELCO				
	#20	DSB	-0.5	
	#17	LSB	-0.5	
	#18	FM-1	-0.4	
	#19	FM-2	-0.4	
#5 FORD				
	#24	DSB	-0.1	
	#21	LSB	-0.3	
	#22	FM-1	-0.2	
	#23	FM-2	-0.5	
#1 DELCO SCA GROUP B				
	#28	DSB	-0.3	
	#25	LSB	-0.4	
	#26	FM-1	-0.5	
	#27	FM-2	-0.5	
#5 FORD SCA GROUP B				
	#32	DSB	-0.2	
	#29	LSB	-0.3	
	#30	FM-1	-0.2	
	#31	FM-2	-0.3	

LEGEND	
2	DSB
3	LSB
4	FM-1
5	FM-2

APPENDIX W

Impulse Documentation

ELECTRONIC INDUSTRIES ASSOCIATION

DIGITAL AUDIO RADIO LABORATORY

Engineers: RMc/DL
DATE: 5/9/95

TEST: C1

Pulse signal for Impulse Noise Test
Requirements call for pulse of 10 nanoseconds wide with a rise and decay time of 4 nanoseconds.

TEST SET-UP

Pulse Generator: Interststate P25
Oscilloscope: Tektronix 7104
Plug-in Modules: 7A19 Vert. Amp 7B85 Time Base

PHOTO No. 1

HORIZ.: sweep speed 10nsec/div. VERT.: 1.0v/DIV.

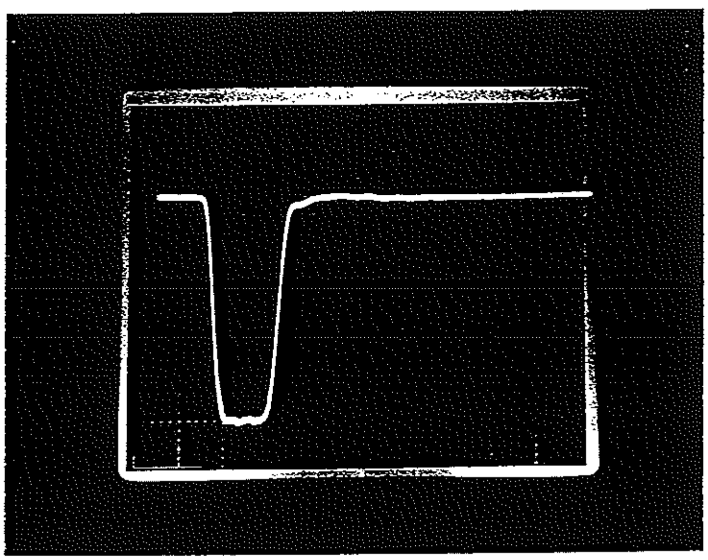


PHOTO No. 2

HORIZ.: sweep speed 10nsec/div. VERT.: 0.5v/DIV.

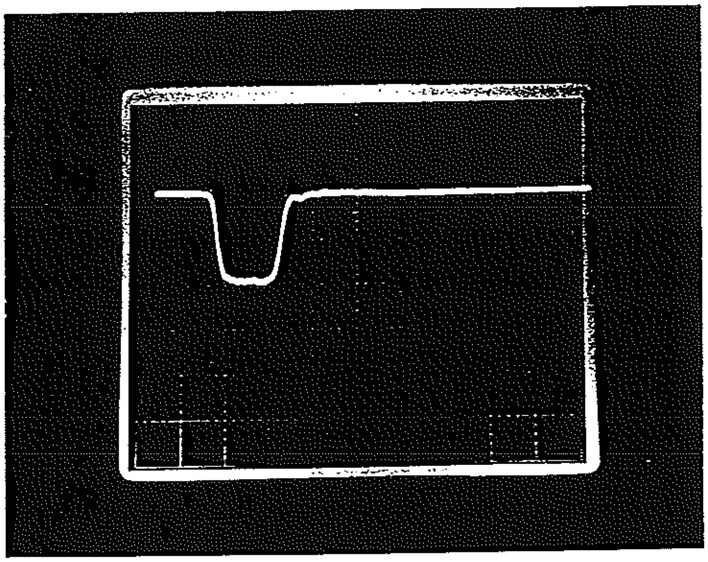
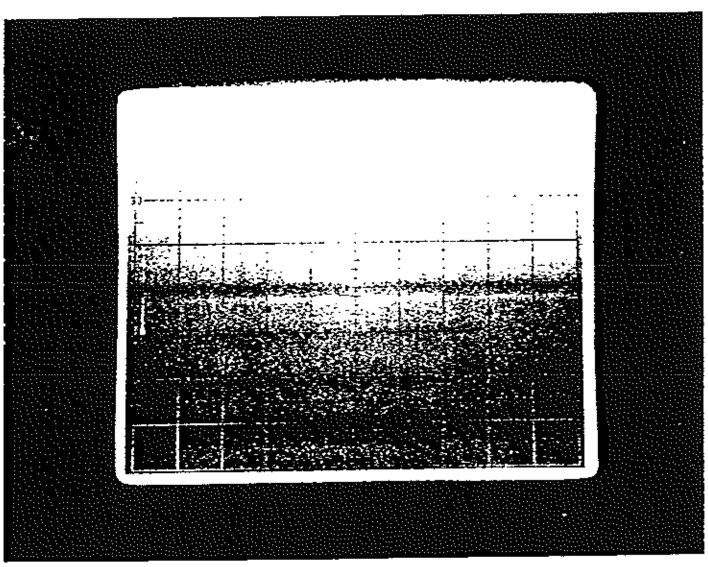


PHOTO No.3

HORIZ.: sweep speed 0.2msec/div. VERT.: 1.0v/DIV.

REPETITION RATE: 1KHZ



NRSC-R58

NRSC Document Improvement Proposal

If in the review or use of this document a potential change appears needed for safety, health or technical reasons, please fill in the appropriate information below and email, mail or fax to:

National Radio Systems Committee

c/o Consumer Electronics Association

Technology & Standards Department

1919 S. Eads St.

Arlington, VA 22202

FAX: 703-907-4190

Email: standards@ce.org

DOCUMENT NO.	DOCUMENT TITLE:
SUBMITTER'S NAME: COMPANY:	TEL: FAX: EMAIL:
ADDRESS:	
URGENCY OF CHANGE: _____ Immediate	_____ At next revision
PROBLEM AREA (ATTACH ADDITIONAL SHEETS IF NECESSARY):	

a. Clause Number and/or Drawing: b. Recommended Changes: c. Reason/Rationale for Recommendation:

ADDITIONAL REMARKS:	
SIGNATURE:	DATE:

FOR NRSC USE ONLY

Date forwarded to NAB Technology: ______________________

Responsible Committee: ______________________

Co-chairmen: ______________________

Date forwarded to co-chairmen: ______________________

CEA®

Consumer Electronics Association

NAB™

NATIONAL ASSOCIATION OF BROADCASTERS