First Quarterly Report

January 1, 1996 through March 31, 1996

NIH Contract N01-DC-6-2100

Speech Processors for Auditory Prostheses

submitted by

Donald K. Eddington
Nico Garcia
Victor A. Noel
William M. Rabinowitz
Mario A. Svirsky
Joseph Tierney
Margaret E. Whearty

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge Massachusetts

1 Introduction

The work reported in this Quarterly Progress Report is a continuation of the research effort started under the previous Contract #N01-DC-2-2402 which ended December 31, 1995. Under the present contract we are continuing work directed at the design, development, and evaluation of speech processors for use with implanted auditory prostheses in deaf humans. The major research efforts are proceeding in four areas: (1) developing and maintaining a laboratory based, software controlled, real-time, speech processing facility where processor/stimulator algorithms for monaural and binaural eight-channel implants can be implemented/tested and a wide range of psychophysical measurements can be made, (2) using the laboratory facility to refine the sound processing algorithms used in the current commercial and laboratory processors, (3) using the laboratory facility to explore new sound processing algorithms in implanted subjects, and (4) designing and fabricating programmable, wearable speech processors/stimulators and using these systems to: (a) field test processor algorithms developed and tested in the laboratory, (b) evaluate the effects of learning using longitudinal evaluations of speech reception, and (c) compare asymptotic performance of different speech processors across subjects.

During this first quarter, we concentrated on the software and hardware preparations necessary to deliver new wearable sound processors to approximately twenty of our subjects using the Ineraid cochlear implant system. The majority of effort in this area was directed at the design and procurement of a new earhook assembly (including the associated pedestal connector and cable) and the four-cell, nickel-metal hydride (NimH) battery packs and rechargers. We spent a small amount of time cleaning up several small problems with the processors themselves.

We also explored procedures for adjusting parameters related to the nonlinear mapping of amplitudes used in CIS processors. The parameters associated with the mapping function provide a wide range of operating points that result in a comfortable listening level but not in equivalent transfer of information. During the past quarter, we settled on a procedure for setting these parameters that will be used in the initial fitting of the new wearable processor for the Ineraid subjects.

2 Wearable Processors

The new wearable sound processing system will replace all of the external components of the Ineraid hardware system in approximately twenty subjects. These components includes the sound processor, processor/earhook cable and connectors, earhook assembly and the earhook/pedestal cable and connector shown in Figure 1.

The specifications for the new programmable sound processor were developed jointly by us, the Geneva Group and RTI. The Microprocessor Laboratory of the Geneva Engineering School was responsible for the circuit design, printed circuit board layout, manufacturing and testing of the sound processor. This system was described in the Final Report for contract N01-DC-2-2402[3]. The processor/stimulators have been manufactured and delivered and are undergoing initial bench testing.

Precision Interconnect (PI) designed and manufactured the cable and associated connectors that provide the connections between the sound processor and the earhook assembly. This cable includes leads that: (1) carry signals from the microphone amplifier housed in the earhook assembly to the input circuitry of the sound processor and (2) connect the six stimulation signals (plus return) produced by the sound processor to the earhook assembly where they are routed to the electrode leads through the subject's pedestal.

We designed a custom earhook assembly that includes a commercial housing (Phonic Ear) with an electret microphone (Lectret) and a custom printed circuit (PC) board. The earhook assembly serves as a miniature junction box that distributes the microphone signals to the processor/earhook cable and the stimulation signals to the earhook/pedestal cable. In addition to providing the traces for the distribution of these signals, the earhook's PC board also includes a preamplifier for the microphone signal. The short, seven-conductor cable that connects the stimulating signals from the earhook assembly to the electrode leads at the pedestal is customized in length and orientation for each subject's pedestal location and orientation. We have procured all of the subunits required: the basic earhook shell, the PC boards, the pedestal connectors, and the cable sockets. Assembly of a unit for each of our subjects is underway.

We have also received delivery of the processor/earhook cables which include nine conductors, a shield, a polyurethane wrapping, and two connectors: one to mate with the earhook socket, the second to mate with the processor socket. We are hopeful that these new cables will not suffer from the same degree of stiffening with use that we have observed with the corresponding Ineraid cables.

The power sources for these processors consist of a four-cell battery pack that uses type 150AFH (roughly AA size) NimH rechargeable units. We have received approximately 100 of these units to be distributed to our subjects. The recharging units, designed by our colleagues at the Geneva Engineering School, have also arrived and are in the initial phase of testing. At the present time, we estimate that the CIS processor algorithm we designed for the wearable processor will operate in excess of 16 hours on a single, charged battery pack.

3 Fitting Procedures for the Wearable Processor

Each channel of a Continuous Interleaved Sampling (CIS) processor algorithm includes a nonlinear mapping function of the type shown in Figure 2. This mapping function maps an input range of envelope amplitudes (60 dB in the figure) to an output range (DR dB in the figure) that is based on the dynamic range of the particular electrode connected to that channel. Fitting a wearable processor to each of our subjects means that a procedure to adjust the parameters of this mapping function had to be developed. In addition to the shape of the mapping function, the values for Gin and Gout must be specified.

We have discussed a number of issues related to the operation and specification of this style mapping function in previous QPRs[1, 2]. In the following sections, we describe the rationale and method for adjusting the mapping parameters we have adopted for the Ineraid subjects we will fit in the next Quarter.

3.1 Setting Mapping Function Parameters: Method I

Method I for setting the mapping parameters was evaluated by using our real-time laboratory system to implement a five-channel CIS processor to test 12 subjects. The magnitude of the frequency response for each of the bandpass filters and for the lowpass filter used for smoothing of the quadrature envelope are shown in Figure 3. The pulsatile stimulation for each electrode pair was a 2000 Hz, cathodic-first, bipolar pulse train with phase durations of 31.25 μs (total pulse width of 62.5 μs). The ordering of stimulation across the electrodes was 5, 2, 4, 1, 3, ending with no stimulation for 187.5 μs. Each channel's pulsatile output signal was amplitude modulated by the output signal of the respective channel's mapping operation.

The input to each mapper, e(t), was the smoothed quadrature envelope from its respective bandpass filter. The mapping transforms the envelope levels (peak value: 32767, range: 60 dB) into a range of output levels (peak value: 32767) determined by the dynamic range of the associated electrode[1]. The dynamic range of each electrode was defined by two psychophysical measures: most comfortable level (MCL) and threshold (THR). These behavioral measurements were made using a 2000 Hz, pulsatile stimulus with cathodic-first bipolar pulses with a phase duration of 25 μs.

Once the mapping function is defined as above, the parameters Gin and Gout must be specified. In the case of Method I, the Gout (see Figure 2) for each channel was adjusted such that for m(t)=32767, i(t)=MCL=Imax. In the case of Gin, a single value was used across all channels. While playing sentences from our TIMIT[4] database,[2] Gin was adjusted until a comfortable loudness was achieved.

Examination of the Gin established using this method for each of 12 subjects shows that they result in relatively low input amplitudes to the mapping function. This can be appreciated by examining Figure 4 together with Table 1 where Gin is given for each of the 12 subjects. Gin is referenced to the gain that causes 1% clipping at the mapper input in channel 2, the channel with the greatest energy when processing the TIMIT sentences.

Figure 4 plots the cumulative percentage of mapper input levels (Gin=0dB) for each processor channel. It shows that for channel#2 (the highest energy channel), 99% of its amplitude samples fall below the input clipping level (0 dB). Table 1 shows that all of the subjects set Gin to values 18 to 46 dB below this level. From Figure 4 we see that a Gin 18 dB below that resulting in 1% of the levels being clipped means that over 25% of channel#2's samples fall below the 60 dB input range (-44 dB shown as the 25% level). In the other channels, the percentage of levels falling below the input range will be even larger. S02 set Gin 46 dB below that producing 1% clipping. This results in more than 75% of channel#2's envelope levels falling below the mapper's input range and channel#5 loses almost 99%.

While the speech reception scores obtained for the Gin and Gout settings specified using Method I compare favorably to those obtained with the commercial CA system (see Table1) the resulting operating points seem undesirable for two reasons. First, a large percentage of the envelope levels fall below the mapper's input range resulting in a large percentage of envelope levels that produce inaudible stimuli. We were surprised that mapper operating points generating this degree of distortion did not result in consonant recognition scores well below those of the Ineraid. We have, therefore, begun a study comparing measures of consonant recognition with the range of envelope levels mapped. Preliminary results from this study will be reported next Quarter.

A second result of the operating points established by Method I is that a relatively large portion of the mapper's input and, therefore, output ranges are not used. This means that the range of input amplitudes being mapped could be represented with significantly higher resolution. A second method for setting the mapping function parameters that is discussed in the following

[1] The mapping relationship used is of the form $y = ax^p + b$ with $p = 0.001$ with a and b chosen to map the 60 dB input range onto a $20 * log_{10}(MCL/THR)$ output range. Input levels below this range were mapped to an output of 0.

[2] Sentences spoken by male and female talkers from eight dialect regions in the U.S. The database subset we used consisted of one male and one female talker from each of the eight dialect regions, each speaking ten sentences.

section overcomes this problem.

3.2 Setting Mapping Function Parameters: Method II

Our present procedure for setting the parameters of the mapper uses the same mapping function described above and in Footnote 1. Gin is uniform across channels and is adjusted so that the highest 1% of the envelope levels in channel#2 (highest energy channel) reach the clipping level (0 dB) for our standard database of TIMIT sentences.

Once Gin is set, the output levels must be adjusted to provide a comfortable listening level. One method to accomplish this would be to adjust Gout. As shown in Figure 5a and 5b, this amounts to applying a scaling factor, k, that shifts both Imax and Imin by the same factor and maintains the DR for each channel.

If the selection of Imin is based on a threshold criterion, it may be desirable for its value to remained fixed. One way to accomplish this is to scale only Imax as represented in Figure 5c. When the dynamic ranges of a subject's electrode set are not uniform, this simple scaling method reduces the dynamic range unequally. This effect can be seen in Table 2 where the first three data rows show the MCL, THR and DR associated with each of subject 26's electrodes. The next two rows show the effect of equating Imin to THR and Imax to a scaled (k=0.6 or -4.4dB) MCL. Notice the distortions across the DRs that occur when they are reduced by a constant 4.4 dB using this method.

Method II's approach to scaling the mapper's output range and maintaining the relative size of the DRs across channels is to use the relationship Imax=k[MCL-THR]+THR to set Imax (see also Figure 5d). This results in the Imax and DR values shown in the last two rows of Table 2. Notice that the reductions (in dB) are roughly proportional to the DR measured for each electrode.

We plan to use Method II's linear scaling of the output range to set the mapping function's output characteristics as we fit subjects with new wearable processors. We have tested this method with twelve subjects using our laboratory system and found scaling values (k) that provide a comfortable listening level for a Gin that causes 1% clipping in channel#2 for our standard TIMIT sentences. These "k" values and the associated consonant recognition scores are shown in Table 3. When compared with the scores obtained with CIS mappers set by Method I, the preliminary scores for CIS mappers set by Method II show three subjects (A01, S23 and S27) that experience a significant reduction in performance. Considering our experience with one subject switching from CA to CIS processing who began with an initial decrease in performance and several weeks later showed superior performance with the CIS system and given the theoretical advantages of Method II, we decided to use Method II for the initial fitting of the new wearable processors.

4 Future Work

Next Quarter we will begin longitudinal trials of a CIS sound processing strategy using our new wearable and programmable sound processor. We will begin by fitting approximately twenty of our research subjects with this new system and conduct initial speech reception tests in noise and quiet. We will report on the results of these initial tests in the next QPR.

While the beginning of the longitudinal trials will consume most of our effort next Quarter, we also plan to use our laboratory system to conduct acute experiments in at least one subject that explore the relationship between speech reception and the range of envelope levels mapped in CIS sound processing.

References

[1] D.K. Eddington et al. (1995): "Speech Processors for Auditory Prostheses, Eleventh Quarterly Progress Report (April 1 through June 30, 1995)." *NIH Contract N01-DC-2-2402.*

[2] D.K. Eddington et al. (1995): "Speech Processors for Auditory Prostheses, Twelfth Quarterly Progress Report (July 1 through September 30, 1995)." *NIH Contract N01-DC-2-2402.*

[3] D.K. Eddington et al. (1996): "Speech Processors for Auditory Prostheses, Final Report (September 30, 1992 through December 31, 1995)." *NIH Contract N01-DC-2-2402.*

[4] W.M. Fisher et al. (1986): "The DARPA Speech Recognition Research Database: Specifications and Status. *Proc. DARPA Workshop on Speech Recognition, pp93-99, Palo Alto California.*

Table 1: Gin (dB, re the TIMIT reference) set by each subject to achieve a comfortable listening level when the Gout of each CIS channel was set to produce an output current range from THR to MCL. Lines three and four present recent Ineraid scores and scores obtained for the Five Channel CIS Processor using Method I fitting. The scores are obtained using the 16 consonant test (scores for S04 are for a 24 consonant test).

Subject	A01	S01	S02	S04	S11	S15	S18	S20	S23	S25	S26	S27
Gin (dB)	-27	-40	-46	-18	-28	-34	-27	-37	-27	-34	-18	-35
CA	47	56	44	85	39	64	55	81	66	64	76	58
CIS Method I	51	71	62	88	46	78	45	87	79	79	75	52

Table 2: Most Comfortable Loudness (MCL) levels, Threshold (THR) levels, and the resulting Dynamic Ranges (DR) associated with each electrode for subject S26. Additional entries demonstrate two approaches to output loudness reduction.

Electrode #	1	2	3	4	5	6
MCL (μA p-p)	840	1050	1125	1500	1250	1580
THR (μA p-p)	287	289	368	403	540	1087
DR (dB)	9.4	11.2	9.7	11.4	7.2	3.2
0.6MCL (μA p-p)	504	630	675	900	750	948
DR (dB)	5.0	6.8	5.3	7.0	2.8	-1.2
0.6[M-T]+T (μA p-p)	619	746	822	1061	966	1383
DR (dB)	6.7	8.2	7.0	8.4	5.1	2.1

Table 3: The reduction factors used for loudness adjustment. TIMIT database speech is input so that channel#2 is clipping 1% of the time. Lines three and four compare the CIS processor scores for the two methods of fitting described in the text for 16 consonant tests (S04 was tested with 24 consonants).

Subject	A01	S01	S02	S04	S11	S15	S18	S20	S23	S25	S26	S27
k	0.5	0.6	0.5	0.5	0.6	0.5	0.6	0.4	0.6	0.65	0.7	0.55
CIS Method I	51	71	62	88	46	78	45	87	79	79	75	52
CIS Method II	43	76	57	99	52	83	46	91	65	76	75	39

Figure 1: Components of the wearable sound processing system. Starting at the bottom left: sound processor, processor/earhook cable and connectors, earhook assembly, and the earhook/pedestal cable and connector.

Figure 2: Diagram of the non-linear mapping function used in a CIS processing scheme. The total gain before the mapper is characterized by a single input gain ("Gin" that results in signal at the mapper input representing "e(t)," the band envelope. The mapper output, "m(t)," amplitude modulates the pulse train that is converted to a current stimulus, "i(t)," by the voltage-to-current source converter ("V/I" The amplitude of the current pulses is determined by "m(t)" and "Gout." From QPR12 figure #3 NIH Contract # NO1-DC-2-2402.

Figure 3: Frequency response magnitudes for the five bandpass filters (top) and the lowpass filters used for envelope smoothing (bottom) in the 5-channel, CIS processor.

Figure 4: Cumulative level distributions at the mapper input (Gin=0dB) for each processor channel for our standard TIMIT database. The inset tables for each channel show the percentage of levels that fall at and below specified input levels.

Figure 5: Three methods for manipulating output stimulating current ranges to achieve subjective loudness reduction.

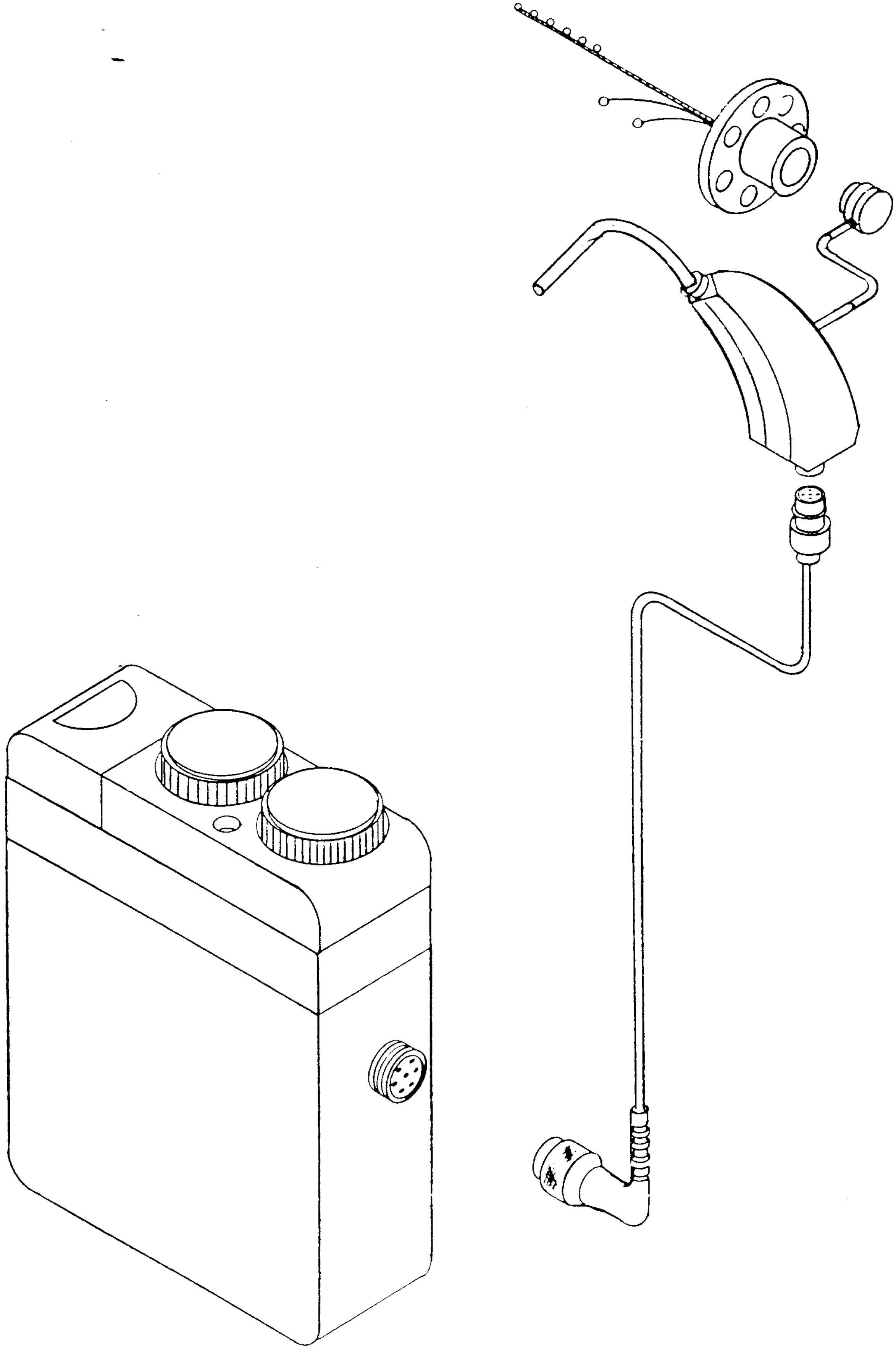
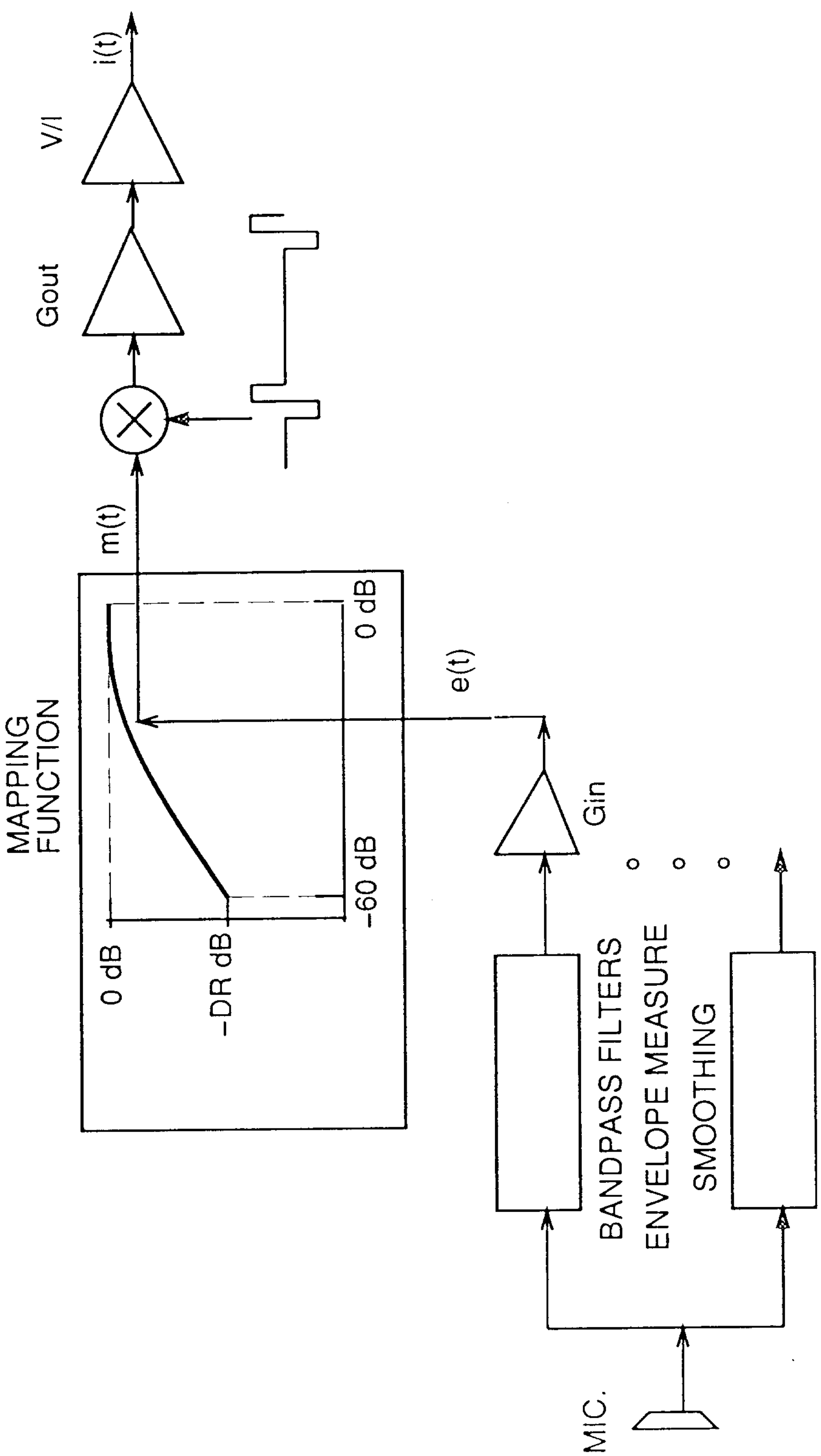


FIG. 1



MIC.
BANDPASS FILTERS
ENVELOPE MEASURE
SMOOTHING
Gin
e(t)
MAPPING
FUNCTION
0 dB
–DR dB
–60 dB
0 dB
m(t)
Gout
V/I
i(t)

FIG. 2

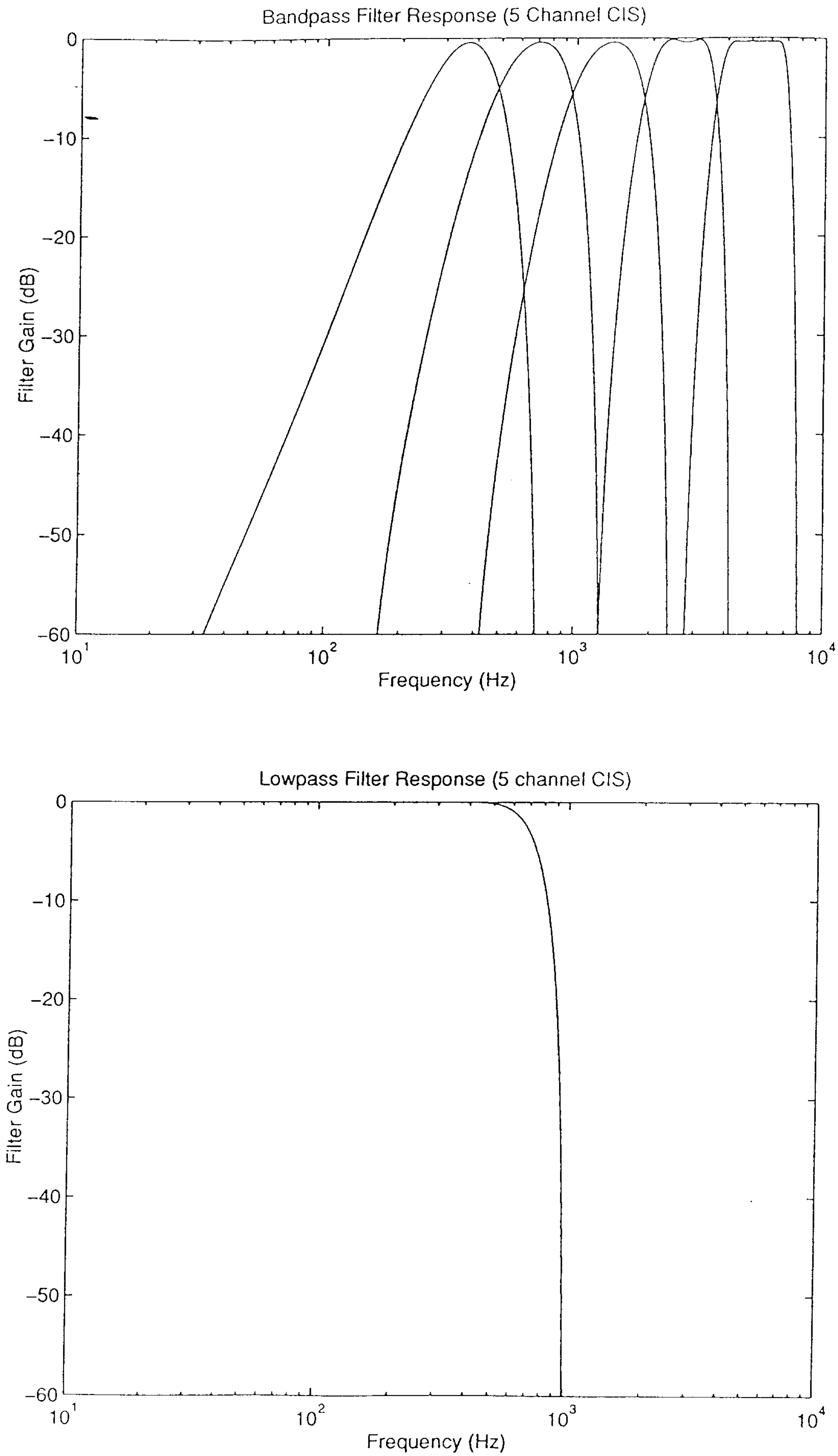
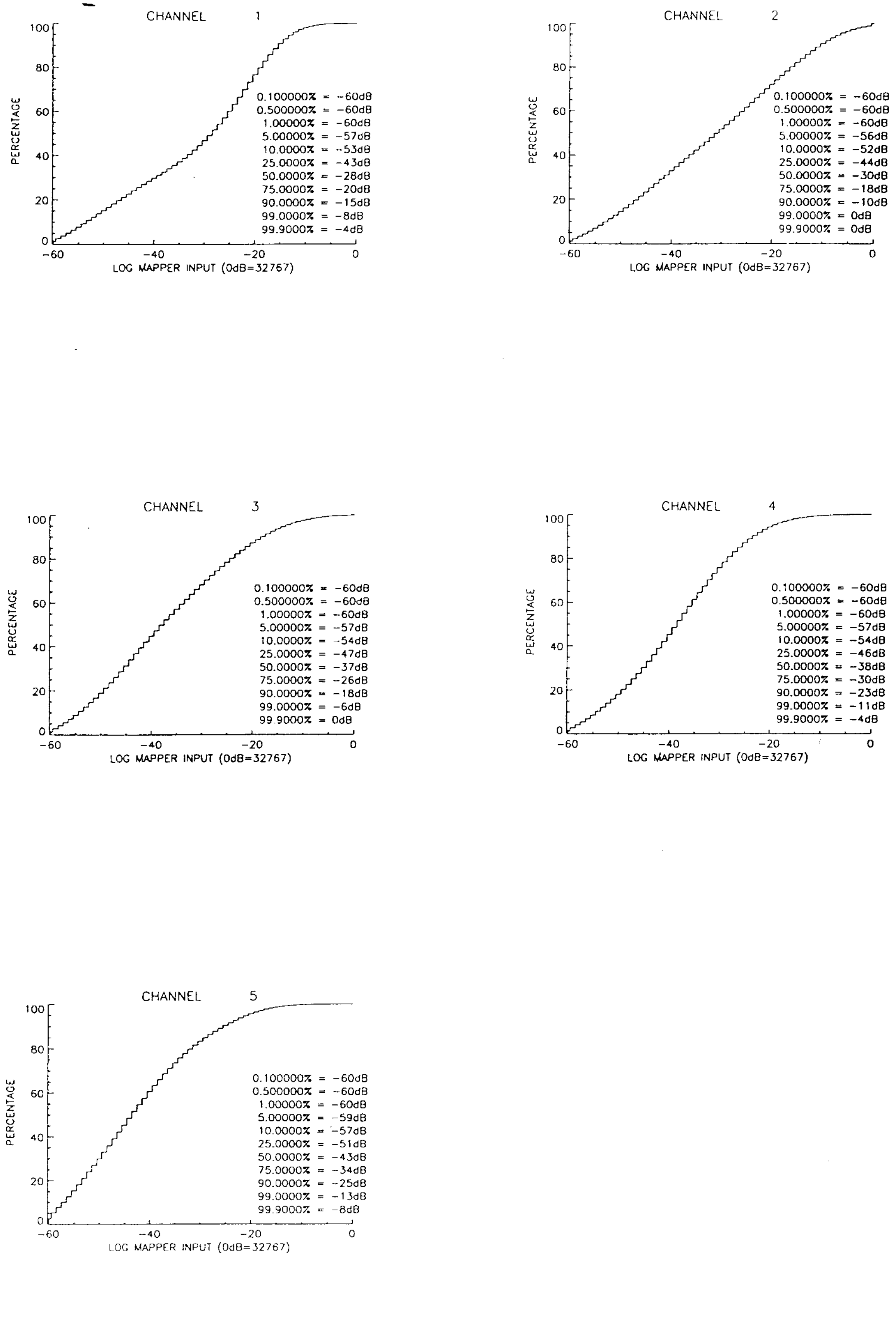

Bandpass Filter Response (5 Channel CIS)
Filter Gain (dB)
0
-10
-20
-30
-40
-50
-60
10^1
10^2
10^3
10^4
Frequency (Hz)
Lowpass Filter Response (5 channel CIS)
Filter Gain (dB)
0
-10
-20
-30
-40
-50
-60
10^1
10^2
10^3
10^4
Frequency (Hz)

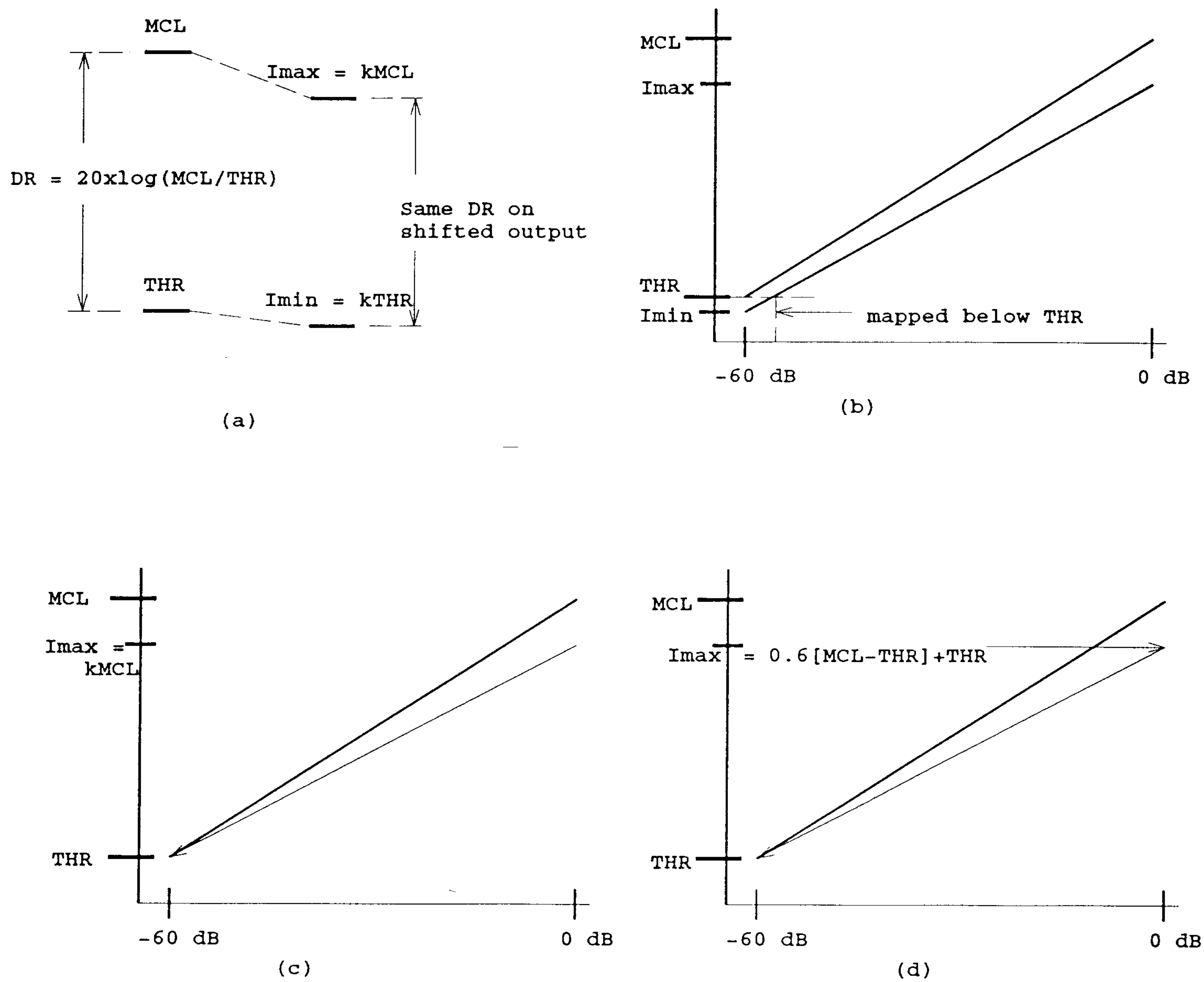
FIG. 3



dr1-8.hist_23_5c_1

binsize=1dB

FIG. 4



Three Methods for Output Loudness Reduction

FIG. 5