

# 음성파형 부호화기의 실시간 성능측정 시스템

## A Real-Time Performance Evaluation System for Speech Waveform Coders

\* 김 용 철 (Kim, Y. C.)  
 \*\* 은 종 관 (Un, C. K.)

### Abstract

In this paper realization of a real-time performance evaluation system for speech waveform coders is presented. This system has been designed to measure the performances of speech coders when the input signal is real speech. The hardware design is based on a bit-slice microprocessor structure. With the real-time system developed, the performances of three types of hardware codecs have been evaluated. The evaluation results are compared with those obtained by using a distortion analyzer. The real-time system should be a useful tool for evaluation of speech codecs. One can avoid the tedious and costly process of subjective listening tests by using the system developed.

### 요 약

본 논문에서는 음성파형 부호화기의 성능을 실시간에 측정하기 위한 시스템의 구현에 관하여 연구하였다. 본 장비는 "bit slice" 마이크로프로세서로 설계되었다. 개발된 시스템으로 세 개의 codec의 성능을 측정 하였으며 이 결과를 distortion analyzer로 측정한 결과와 비교하였다. 개발된 장비는 음성 부호화기의 성능측정에 유용히 쓰여질 수 있으며, 이러한 장비를 사용함으로써 음성 부호화기의 성능시험을 위한 주관적 청취시험 과정을 피할 수 있게 되었다.

### I. INTRODUCTION

Since the early 60's, pulse code modulation

(PCM) has been widely used in digital communication systems. It has advantages over other low-rate speech coding systems such

---

\* 금성전기 (주)  
 \*\* 한국과학기술원 교수

as vocoders in noise immunity and simplicity. However, since it requires relatively large bandwidth, many researchers have studied extensively on the bandwidth reduction in waveform coding. As a result, waveform coders, such as adaptive delta modulation (ADM) and adaptive differential pulse code modulation (ADPCM) systems, that take only about half of the bit rate of the conventional PCM but have almost the same speech quality are now practically available.

In waveform coding of speech the main source of distortions introduced in the coding process is the quantization noise. In addition to this nonlinear distortion, linear distortions such as dc level shift, short time delay and phase distortion are also added. In general, these distortions do not give the same influence on the perceptive quality of the reconstructed speech signal. The unpleasantness of distortions depends on such factors as the speech level, correlation of the distortion with the speech signal and the clarity of the original speech. For example, slight noise in silence portion can be very disturbing, while noise in high-level speech portion may be masked out. In this respect, minimization of distortion has a rather vague meaning. Accordingly, accurate evaluation of a coder performance must be based on a human-oriented distortion measure.

An objective measure is a convenient tool that can be replaced with the tedious and costly subjective test. A variety of objective measures have been developed. Signal-to-noise ratio (SNR) is one of the most commonly used objective measures. But, it is well known that SNR is a poor indicator of the speech quality [1]-[3].

The segmented SNR ( $SNR_{SEG}$ ) and the frequency-weighted  $SNR_{SEG}$  appear to correlate well with subjective performance [1].

So far, most performance studies in coder design have been based on computer simulation with a digitized speech data base of short duration. However, those studies based on simulation have shortcomings in that it is often difficult to reflect all hardware limitations in the simulation study. Consequently, the performance of a hardware-implemented coder may be different from that obtained in computer simulation of the coder, as is often the case in real world applications.

So far, the performance of a hardware codec has normally been measured by a distortion analyzer. The test input signal is a single frequency sinusoid of usually 800 Hz or 1000 Hz, which is known to be the spectral point where the power spectral density of speech signal is most concentrated [3]. The energy of distortion in the output sinusoidal signal is measured by the distortion analyzer. The method of measuring distortion energy by the analyzer is different from that used in computer simulation. The distortion analyzer cannot measure distortion in the exact sense of the quantization noise. Furthermore, it cannot be used in measuring the performance of a hardware codec when the input signal is real speech. Although the distortion analyzer yields some indication of the performance of a hardware codec, it cannot give a performance measure that is closely correlated with human perception.

To overcome the shortcomings of computer simulation and the distortion analyzer,

we have implemented a real-time performance evaluation system for speech waveform codecs. By using this system, one can avoid tedious and costly subjective listening tests in the performance evaluation of a new speech codec. The real-time equipment also facilitates the parameter optimization of a hardware codec in laboratory design and can be used as a useful test equipment. The microprocessor-based evaluation system has been constructed using the AMD system 29 development system [4]. The microprogrammed structure of the processor enables real-time processing and full control of the hardware by a 64-bit control word. The features of our real-time system are:

Flexibility of loading any objective measure program, such as the segmented signal-to-noise ratio or the frequency weighted objective measure that has strong correlation with a subjective measure.

- Variable sampling rate with the maximum rate of 32 kHz.
- Capability of measuring SNR in the range of 50 dB.
- Capability of compensating time delay resulting from the coding process.

Following this introduction, we describe hardware design and software implementation of the performance evaluation system in Section II. In Section III we present the results of performance evaluation of three commercially available speech codecs, and in Section IV we discuss the results. Finally, we make conclusions in Section V.

## II. IMPLEMENTATION OF THE EVALUATION SYSTEM

### A. Overall Hardware Architecture

A simplified block diagram of the real-time performance evaluation system is shown in Fig. 1. It can be divided largely into three parts; a central processing part, an I/O part and a microprogram control part. The central processing part is composed of a central processing unit (CPU) and a data memory. The CPU is interfaced with other subsystems via two bidirectional buses, an input data bus (IBUS) and an output data bus (OBUS). All instructions are executed in a cycle time of 220 ns by a 4.5 MHz system clock. The CPU is composed of four Am 2903 microprocessor chips, a status register, a shift multiplexer, and a carry input multiplexer. With these multiplexers, double length arithmetic is possible. The 16-bit CPU fetches from one of three data sources; a microinstruction register, a memory output register (MOR) and an A/D converter. The output of the CPU is fed to one of six destinations; a D/A converter, two memory address registers (MAR1, MAR2), a memory buffer register and an output register via the OBUS lines.

We have used the pipe lining technique in this system to save the fetch time of the next instruction. The pipe line register contains the microinstruction currently being executed by the machine. The Am 2911 contains an address multiplexer that provides four different inputs from which the address multiplexer that provides four different inputs from which the address of the next

microinstruction can be selected. These are the direct input, the register input, the program counter, and the file. Of the sixteen test inputs that can be applied to the condition code multiplexer, eleven test inputs are used. These inputs are the contents of the status register and the I/O interrupt request signals. An I/O interrupt is checked by software and the jump address from the microinstruction register is fed into the direct branch address input of the sequencer.

The data memory is composed of 1k x 16 bipolar RAMS. The data memory is fully buffered by a memory buffer register, a memory output register, and the memory address registers 1 and 2. The memory address registers are used with an address multiplexer to facilitate the alternating access of two block data sets in different memory locations without additional addressing from the CPU.

The I/O system consists of an I/O device part, an interrupt request part, and a control part. The control part generates the clock pulses to control the I/O devices and the codecs in test, and provides the interrupt request to the interrupt control unit of the central processor. The sampling rate can be varied according to the transmission rate of the waveform coder in test. The input speech signal is first low-pass filtered, and its level is adjusted by an automatic gain control (AGC) circuit to match the allowable input range of the coder. Two sample-and-holders sample the input and output signals of the coder at the same time. After A/D conversion, an I/O interrupt request signal is generated. Two kinds of interrupt request signals are generated.

Priority is given to the input interrupt request signal over the output interrupt request signal. Upon checking the presence of these interrupt requests, the CPU goes into the I/O service routine.

A microinstruction used in this system is made up of 64 bits. The instruction word is divided into CPU, MCU and other control parts. Each part is composed of some microfields. The CPU part contains ALU source field, ALU function field, ALU destination field, ALU A-RAM and B-RAM address field, shift field, and carry field. The MCU control part is divided into microfields as defined in the computer control unit manufactured by Advanced Micro Devices (AMD), Inc. The fields are a next address field, a condition code multiplexing field, a branch address field, a status field and an interrupt clear field. The IBUS and OBUS fields are used to select the source of the tri-state input bus and the destination of the output bus. The auto-increment bit controls whether to increment the two address registers (MAR1 and MAR2) at the same time or not. The direct field is used as a direct input address of the sequencer and an operand of the CPU. The memory address select bit is used to select the two address registers when reading and writing.

## B. Software Development

One distinct feature of the performance evaluation system developed is its flexibility in getting an objective quality measure. One or more objective measures that are correlated with a subjective measure can be programmed into the system. In the development of our system we have consider-

ed several objective measures. Perhaps, the simplest and most commonly used one is the SNR that is defined as the ratio of the signal energy to the error energy. It is well known that this measure does not predict satisfactorily the subjective speech quality.

On the other hand, the  $SNR_{SEG}$  is known to have good correlation with the subjective measure [2]. To compute  $SNR_{SEG}$ , one whole utterance is divided into  $M$  adjacent segments of  $J$  samples and SNR is measured for each segment. The  $SNR_{SEG}$  is the average of these measures over  $M$  segments. That is,

$$SNR_{SEG} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{\sum_{j=0}^{J-1} s^2(j+mJ)}{\sum_{j=0}^{J-1} [s(j+mJ) - r(j+mJ)]^2}$$

where  $s(n)$  and  $r(n)$  are the  $n_{th}$  input and decoded signal samples, respectively. In order to avoid the unduly heavy weighting of an idle channel noise in silence portion, those segments whose signal power is below -34 dBm is discarded in this measure.

Although one can implement any objective measure in the evaluation system, let us discuss here software realization of the  $SNR_{SEG}$  as a typical example. The software realization of the segmented SNR can be divided largely into five routines; the initialization routine, the main routine, the logarithm routine, the A/D-D/A interrupt service routine, and the display routine. The overall block diagram of the software is shown in Fig. 2. In the initialization routine, the sample number counter is set to zero and the delayed sample number counter is set to the two's complement

of the delay sample number. The storages of accumulation of input and output signals are also cleared. In the main routine, the interrupt request signal is checked every 30 cycles. After the I/O service routine, the routine that calculates the logarithm of the ratio of the signal-to-noise energy is carried out. Here, interrupt checking is done also every 30 cycles. In this way, the I/O service routine and the logarithm evaluation routine is alternatively carried out. When the logarithm evaluation routine is finished, the main routine takes over the control. After 128 samples are processed, the initialization routine starts and the main routine repeats. The number of segments that can be processed may be adjusted by a slight modification of the main program according to the duration of input speech. All digital data are in 16-bit form. Double precision (32 bits) arithmetic is used for division which operates on the 32-bit wide dividend and divisor.

The logarithm evaluation routine is essentially the power series expansion of natural logarithm up to 64 terms. It occupies a large part of processing time. For the case of a codec with a sampling rate of 32 kbits/s, the logarithm routine takes 47 percents of the whole processing time.

### III. EVALUATION RESULTS

To test the effectiveness of the real-time evaluation system, we have measured the performances of three commercially available speech codecs; Motorola 3417 CVSD, AMI S3507 PCM and INTEL 2910A PCM codecs.

### Motorola 3417 CVSD Codec

The performance of the 3417 CVSD codec was evaluated for a sine wave input and real speech. The transmission rates were 16 and 32 kbits/s. First, an 800 Hz sine wave input signal was applied to the codec. In Fig. 3 the performance result obtained by the evaluation system developed is plotted together with that of the distortion analyzer. A phase shifter was used to compensate the time delay resulting from low-pass filtering. The 3.4 kHz analog low-pass filter used in the coding path was a sixth-order Butterworth filter, which had a nonlinear phase characteristic. Second, the performance was measured for real input speech. The input speech had a duration of 7.3s. The peak segmented SNR's at the rates of 16 and 32 kbits/s were 15 and 21 dB, respectively.

### AMI S3507 PCM Codec

The S3507 codec has the north American  $\mu$ -law companding characteristic. The chip contains a low-pass filter and a receiver low-pass filter. To investigate the transfer characteristics of the low-pass filters in the chip, a sine wave input of 5 volts (peak-to-peak) was applied to the coder at various frequency of 800 Hz. At frequencies other than 800 Hz, a sampling time offset error introduced phase distortion. In this case, the SNR became lower than the value measured with the distortion analyzer. Next, the performance for an 800 Hz sine wave input at various signal levels was evaluated by our real-time system. The result is shown in Fig. 4. We also used the same speech data

as in the CVSD codec for the S3507 PCM codec. In this case, the internal low-pass filters introduced phase distortion, and the performance could not be evaluated properly. For the exact evaluation, the distortions due to these filters should be compensated by passing the input signal through the same filters [5]. The maximum segmented SNR for speech input with our real-time evaluation system was 24 dB.

### INTEL 2910A PCM Codec

The 2910A chip has no internal filters. The input signal and the decoded output signal were bandlimited to 3.4 kHz by the PCM line filter. The chip was operated in the direct control mode for single channel communication.

The output signal was compared with the input signal to measure the performance of the coder. In Fig. 5 the result is shown together with that from the distortion analyzer for an 800 Hz sine wave input. Also, we examined the performance variation for different input frequencies. The result is plotted in Fig. 6. In addition, the performance for real speech was measured by the system developed. It is plotted in Fig. 7.

It is seen from Fig. 6 that the performance deteriorates as the input frequency increases. This degradation is believed to be due to the sample-and-holder within the chip, which acts as a one-pole integrator. Besides this one-pole attenuation, the physical limitation of the companding device adds some distortion to the quantization noise. These distortions account for the lower performance of the hardware-implemented codec than that of computer simulation.

#### IV. DISCUSSION

In most objective measures, noise is defined as the difference between the input and the output signals. When we use this definition, overestimation of noise often results if the two signals are not in exact time synchronizatin. In computer simulation, exact synchronization is implicitly assumed because the coding path is completely known. But, in real codecs, subtraction of the input signal from the decoded signal may not yield an exact estimate of noise. Also, if a constant dc offset voltage is added to the input and output of the codec as in Motorola 3417 codec, the energy of the sampled signal appears larger than the actual energy of the original ac input signal. This dc offset should be removed in the exact evaluation of the signal energy. The internal low-pass filters (as in AMI S3507 codec) cause another difficulty. These filters introduce nonlinear phase delays. Even if we can compensate the time delay for a single frequency input signal, the evaluated performance would be degraded when the input signal occupies the full bandwidth. In addition to phase distortion, the overall codec gain different from unity makes a contribution to the noise energy. The gain slightly deviated from unity may cause a considerable error between the two signals, one of which is an amplified version of the other. For the measurement of the distortion introduced in the coding process only, these filters should be separated from the encoder/decoder part.

For a nonseparable hardware codec, the characteristics of the linear components of the codec other than the encoder and decoder should be estimated, and the input

signal should go through the estimated linear filter before being compared with the output signal. The architecture of such a measuring equipment was proposed by Billi and Scagliola [6].

On the other hand, the distortion analyzer does not take these additional noise terms into account in the performance measure for a single frequency sinusoidal input. The output signal of a coder, a distorted sinusoidal signal, is filtered by a band-rejection filter, and the energy of the filtered output becomes the distortion energy. One may note that this method of measuring distortion energy is not the same as in computer simulation.

The nonuniform time delay of the internal filters and the nonunity gain do not have any significance in the measurement by the distortion analyzer, while these factors should be compensated in the measurement using the same algorithms as in computer simulation. This difference in the meaning of the distortion energy might have caused the performance measure to disagree with that measured by the real-time evaluation system.

It is empirically known that the overall codec gain decreases slightly as the frequency of the input signal increases. For two separate input signals of different frequencies, the distortion analyzer indicates almost equal performances. However, the performance for an input signal composed of the two sinusoidal waves of distinct frequencies is lower than that measured separately by the distortion analyzer for a single frequency input. Thus, the performance for nonstationary speech signal of composite frequencies cannot be measured by the distortion analyzer. The result obtained by the dis-

tortion analyzer turns out to be better by 1 to 2 dB than that obtained by the developed system. This discrepancy appears to be due to the difference of the measuring algorithms of the two systems. If the band rejection filter used in the distortion analyzer does not have very sharp characteristic, the out-band noise that passes through the filter would have smaller energy than its actual value. In this case, SNR measured with the distortion analyzer may appear larger. As one can expect, the performance for real speech input is worse than for a stationary sine wave input. The abrupt peaks which cannot be controlled by the automatic gain control circuit are apt to exceed the admissible input range of the coder. Such peaks are clipped. This clipped portion degrades the quality of the reconstructed speech. Besides this clipping, the low SNR in a segment of the weak signal is an important source of degradation of the performance of the hardware codec.

## V. CONCLUSIONS

In this paper a real-time performance evaluation system for speech waveform coders has been presented. A detailed account on the hardware design based on microprogramed structure and on the software development has been given. Using the developed system, we evaluated the performances of two PCM coders and one CVSD codec, and compared the results with those measured by the distortion analyzer.

The evaluation system developed should be useful in measuring the performances

of speech coders such as ADPCM and ADM using real speech. We believe that the use of the evaluation system can replace the tedious and costly process of subjective listening tests required for evaluation of a new speech codec hardware.

## REFERENCES

1. R.E. Crochiere, L. R. Rabiner, N. S. Jayant, and J. M. Tribolet, "A study of objective measures for speech waveform coders," presented at the Int. Zürich Sem., Zürich, Switzerland, Mar. 1978.
2. B. McDermott, C. Scagliola, and D. Goodman, "Perceptual and objective evaluation of speech processed by ADPCM," *Bell Syst. Tech. J.*, vol. 57, pp. 1597-1618, May-June, 1978.
3. N.R. French and J.C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Amer.*, vol. 19, pp. 90-119, Jan. 1947.
4. The Am 2900 Family Data Book, AMD Inc., 1981.
5. R. Billi and C. Scagliola, "Artificial signals and identification methods to evaluate the quality of speech coders," *IEEE Trans. Commun.*, vol. COM-30, No. 2, pp. 325-335, Feb. 1982.

## LIST OF FIGURE CAPTIONS

- Fig. 1 Block diagram of the performance evaluation system of waveform coders  
 Fig. 2 Overall flow chart of software  
 Fig. 3 Performance of the MC3417 CVSD



codec for an 800 Hz sine wave input  
(Transmission rate: 32 kbits/s)

Fig. 4 Performance of the S3507 PCM  
codec for an 800 Hz sine wave input  
(Transmission rate: 64 kbits/s)

Fig. 5 Performance of the 2910A PCM  
codec for an 800 Hz sine wave input  
(Transmission rate: 64 kbits/s)

Fig. 6 Performance of the 2910A PCM codec  
for a sinusoidal input at various  
frequencies (Transmission rate: 64  
kbits/s)

Fig. 7 Performance of the 2910A PCM  
codec for real speech input (Trans-  
mission rate: 64 kbits/s)

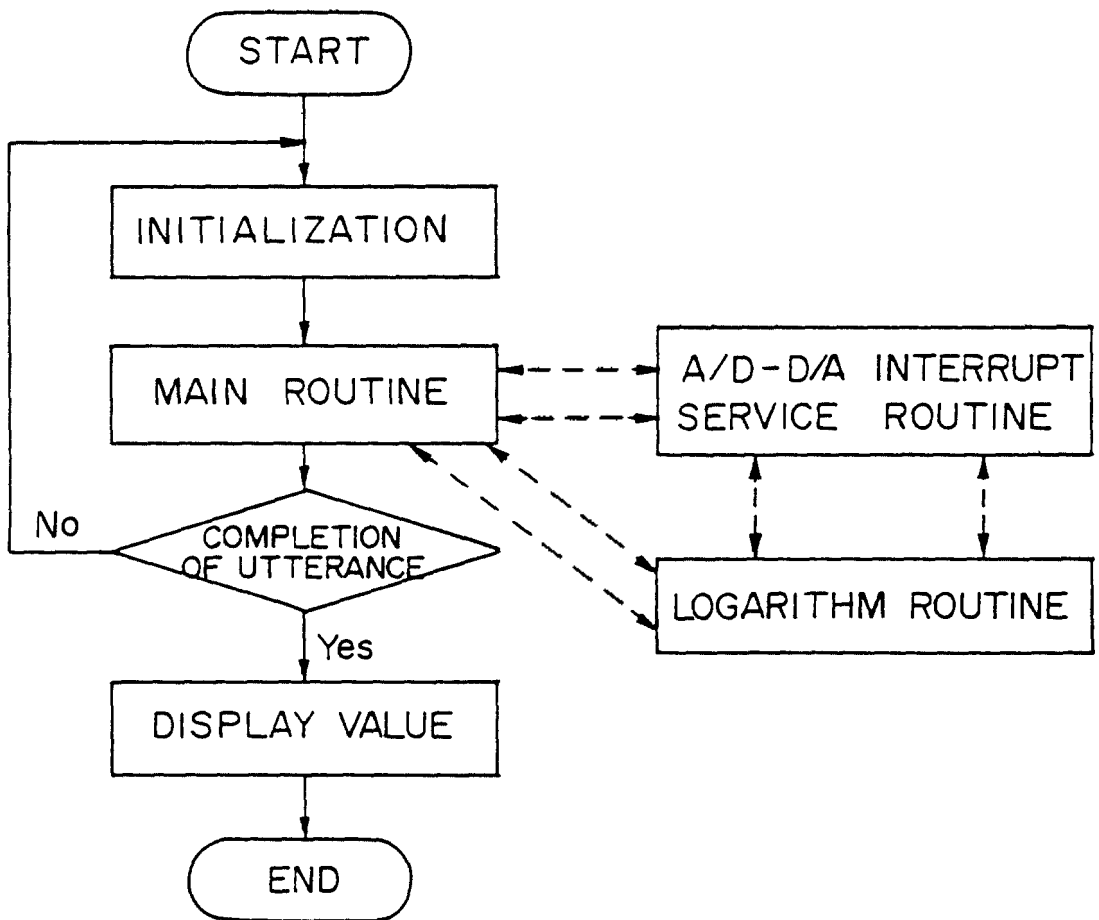


Fig. 1

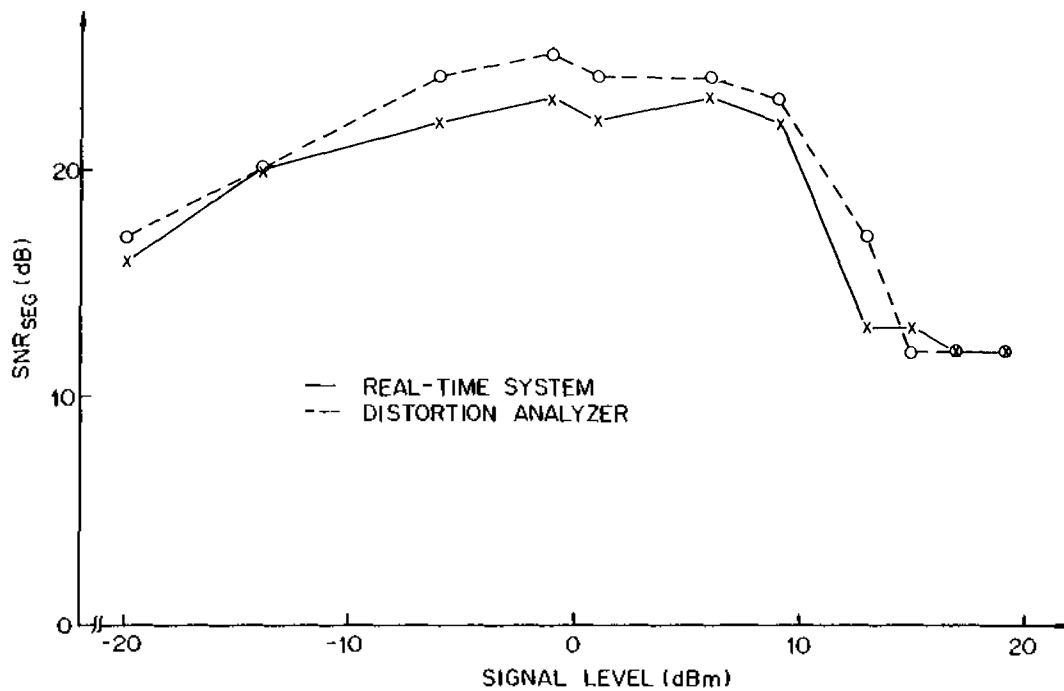


Fig. 2

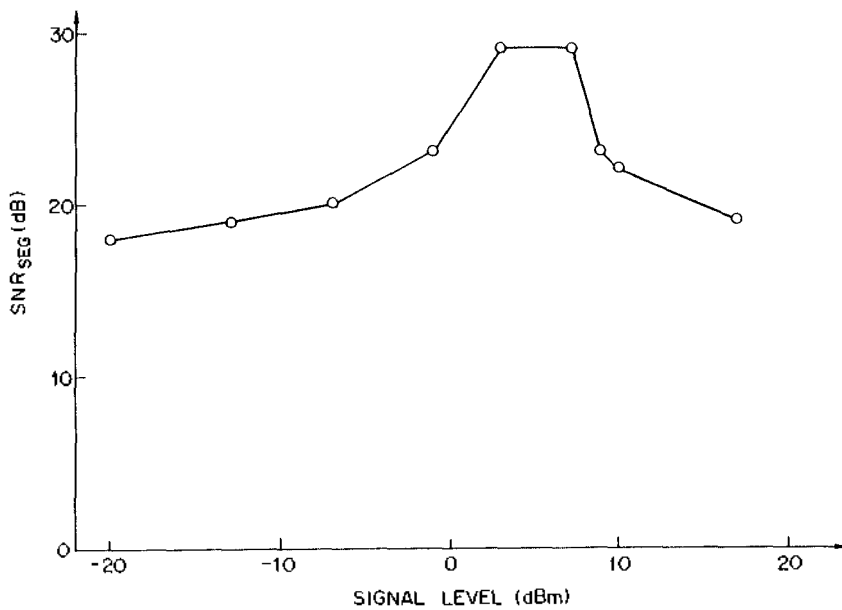


Fig. 3

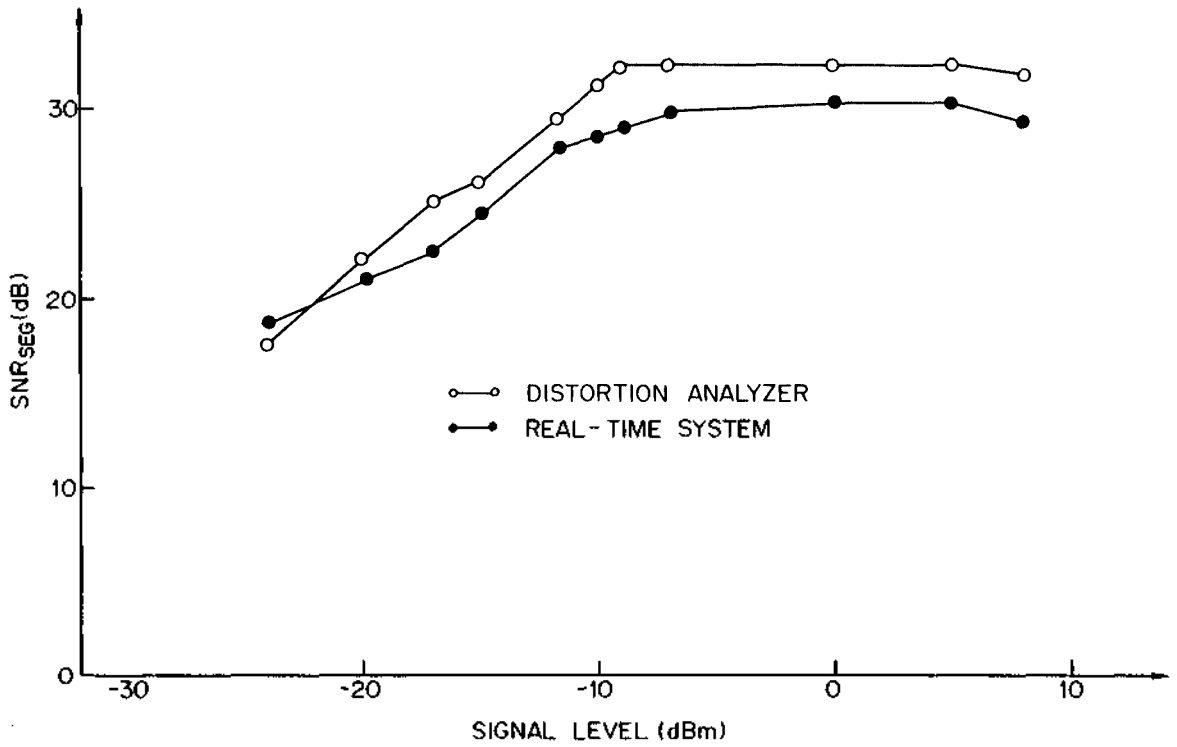


Fig. 4

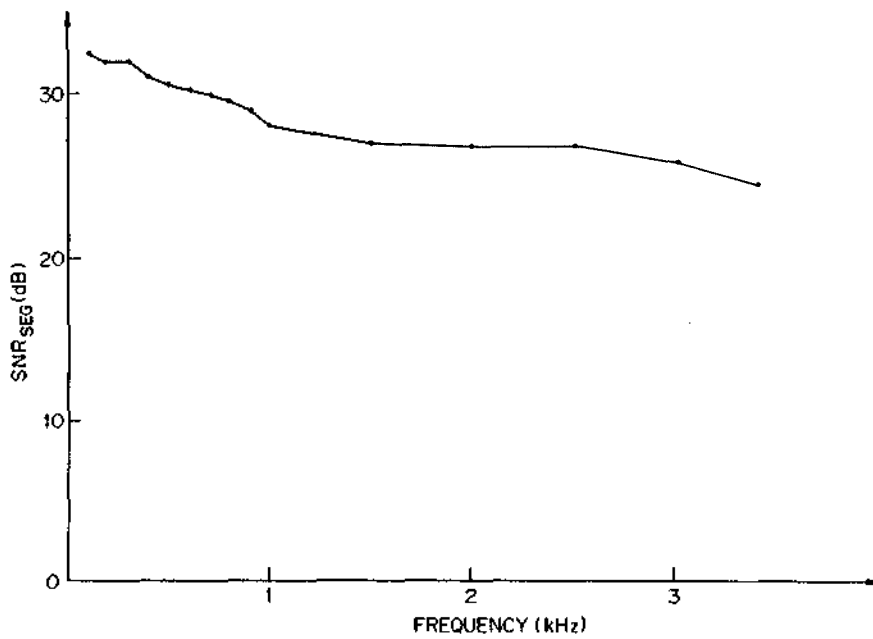


Fig. 5

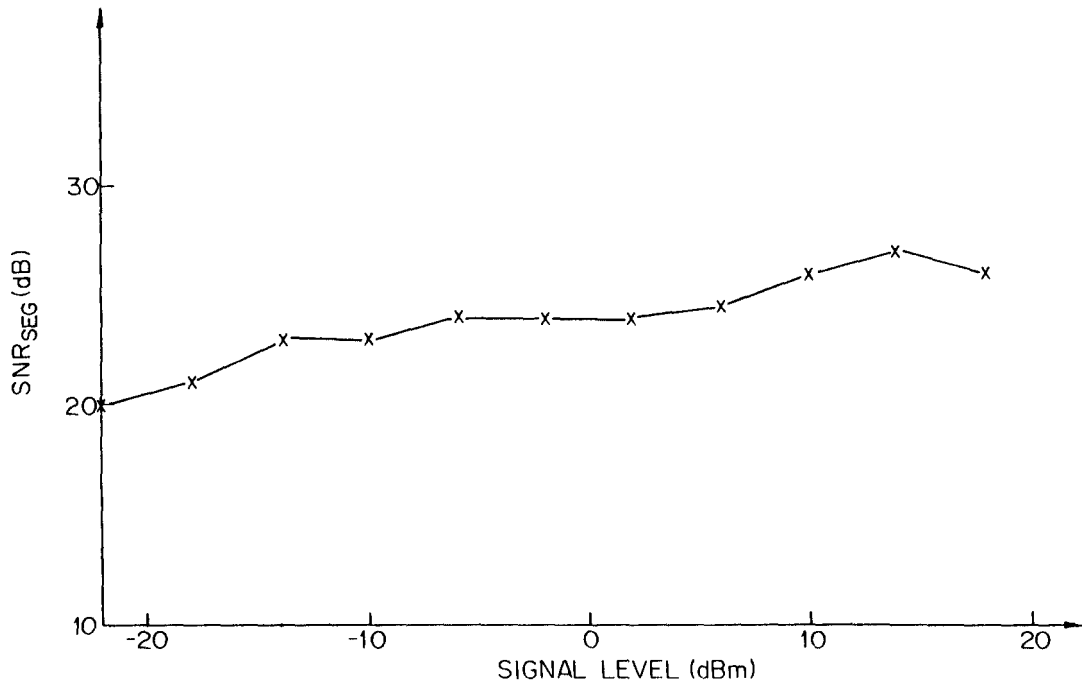


Fig. 6

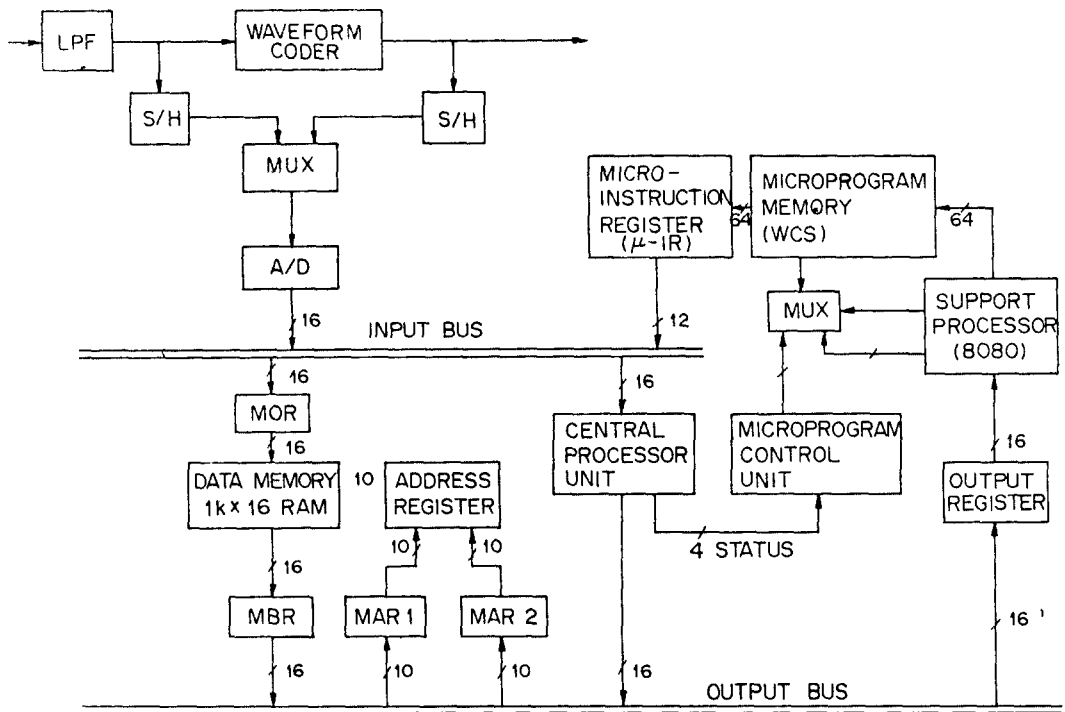


Fig. 7