Implementation of a Multirate Speech Digitizer

BYUNG SUK LEE, CHONG KWAN UN, SENIOR MEMBER, IEEE, HYEONG HO LEE, STUDENT MEMBER, IEEE, BYOUNG CHEOL SHIN, STUDENT MEMBER, IEEE, AND HWANG SOO LEE, MEMBER, IEEE

Abstract—**In this paper, implementation of a compact and efficient multirate speech digitizer with variable transmission rates of 2.4, 4.8, 9.6, and 14.96 kbits/s is presented. The multirate algorithm has been made based on the residual-excited linear prediction (RELP) vocoder with a transmission rate of 9.6 kbits/s. The residual encoder employed in the RELP vocoder uses hybrid companding delta modulation (HCDM). This HCDM is also used as a 14.96 kbit/s coder. If the residual in the RELP system is down-sampled before encoding, a 4.8 kbit/s coder can be realized. If the residual encoder is not used, a 2.4 kbit/s linear predictive coder (LPC) can be realized by incorporating a pitch extractor. In the 4.8 and 9.6 kbit/s coders the pitch-implanted residual excitation method has been used to generate the excitation signal to the synthesis filter.**

The multirate speech digitizer algorithm has been implemented using 2900 series bit-slice microprocessors. The external memory is composed of 2K RAM's and 2K ROM's. The system design is a two-bus structure with a 204 ns cycle time. With efficient hardware and software design, the multirate speech digitizer requires almost the same hardware complexity as compared with the conventional 2.4 kbit/s LPC vocoder.

I. INTRODUCTION

In recent years, as the number of users of digital communication is increasing rapidly, many bandwidth-efficient digital speech coding methods have been developed. Speech coding methods can be classified largely into three categories: waveform coders, vocoders, and a mixed type of these two coders. The waveform coder realizes waveform matching. Typical examples of the waveform coders are pulse code modulation (PCM), differential PCM (DPCM), adaptive DPCM (ADPCM), and adaptive DM (ADM) [1], [2]. In vocoding, only the features characterizing speech signal, such as the vocal tract parameters, voiced/unvoiced decision, and pitch period information, are extracted and transmitted. At the receiver, speech is synthesized using this information. Typical examples of the vocoders are the linear predictive coder (LPC) [3], [4], channel vocoder [5], and formant vocoder [6]. In mixed- (or hybrid-) type coders, the vocal tract information is extracted and transmitted as in vocoders, but pitch extraction is not done. Instead, some form of residual signal is transmitted. At the receiver this signal is used after processing as the excitation signal to the synthesizer. Since pitch extraction is not done, this category of coders is fairly robust to environmental disturbances. Examples of the mixed-type coders are the residual-excited linear prediction (RELP) vocoder [7], voice-excited vocoder [8], and adaptive predictive coder (APC) [9].

In circuit-switched or packet-switched data networks the selection of an appropriate speech coding method should be considered seriously. To reduce the queueing delay and to increase the number of channels in a given bandwidth, a coding method with a low transmission rate should be used, but this low rate coder would result in degradation of speech quality. If a coding method with high transmission rate is selected to improve the speech quality, it increases the queueing delay and decreases the number of channels multiplexed. Also, the possibility of channel congestion would increase in this case. One method to compromise between those contradictory conditions is to use a multirate speech digitizer that can be switched between various transmission rates according to the speech quality required, channel condition, and network flow control. This has motivated the present study in which we have developed a very compact and efficient multirate speech digitizer with variable transmission rates of 2.4, 4.8, 9.6, and 14.96 kbits/s.

To implement the multirate coding system, coders that exhibit good performance at these transmission rates but yield minimum system complexity should be selected. We have selected as a 14.96 kbit/s coder the hybrid companding delta modulation (HCDM) system that yields the best performance among ADM's [2], [10], the RELP vocoder as 9.6 and 4.8 kbit/s coders, and the LPC vocoder as a 2.4 kbit/s coder. We have also used HCDM as the residual coder when the multirate system is operated at 4.8 or 9.6 kbits/s.

Throughout this paper emphasis will be placed on the development of an efficient speech digitizer algorithm, and on the implementation of compact hardware and software of the system. Following this Introduction, we describe the overall algorithm of the multirate speech digitizer in Section II. In Section III we consider hardware design of the speech digitizer, and then discuss development and implementation of system software. Finally, conclusions follow in Section IV.

II. ALGORITHM OF THE MULTIRATE SPEECH DIGITIZER

A. Design Approach

One important goal in designing the software and hardware of our multirate speech digitizer was compactness. To achieve this goal the multirate algorithm has been made based on the principle of RELP vocoding originally proposed by Un and Magill [7]. The RELP vocoder that is used as a 9.6 or 4.8 kbit/s coder in our system uses the well-known linear prediction method for feature extraction and HCDM for residual

Paper approved by the Editor for Data Communications Systems of the IEEE Communications Society for publication without oral presentation. Manuscript received April 6, 1982; revised October 18, 1982.

B. S. Lee is with the Institute of Technology, Gold Star Electric Company, Osan, Korea.

C. K. Un, H. H. Lee, B. C. Shin, and H. S. Lee are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Chongyangni, Seoul, Korea.

0090-6778/83/0600-0775$01.00 © 1983 IEEE

encoding. Therefore, we can have an LPC vocoder for low-rate (2.4 kbits/s) coding and HCDM for high-rate (14.96 kbits/s) waveform coding without additional software. Also, in 2.4 kbit/s LPC vocoding we use residual signal rather than input speech for accurate pitch extraction. Since this residual signal is used in the 9.6 or 4.8 kbit/s coder, it need not be generated separately for pitch extraction. The 9.6 kbit/s RELP coder includes the entire parameters that are used in the 2.4 kbit/s LPC vocoder.[1] Hence, there remains 7.2 kbits/s available for residual coding. Since HCDM, in which the input sampling rate is equal to the output bit rate, is used for residual coding in our system, we decided the digitizer input sampling rate to be 7.2 kHz.

Considering the 4.8 kbit/s coder, one can think of two ways of getting its algorithm. One method is to improve the 2.4 kbit/s LPC coding by assigning more bits in parameter coding, and the other is to reduce the transmission rate in coding the residual signal of the 9.6 kbit/s RELP system. We have chosen the latter approach because by having additional 2.4 kbits/s in parameter coding we could attain little improvement of output speech quality over that of the 2.4 kbit/s coder. To have a 4.8 kbit/s coder we simply down-sampled the residual signal to 2.4 kHz.

Determination of the sampling rate for the high-rate (14.96 kbits/s) HCDM needs another consideration. In our HCDM system we transmit the gain estimated by a feedforward approach for better quality of speech. For the transmission rate of the HCDM system one could conceivably have 16 kbits/s, for which case the input sampling rate must be 15.52 kHz, since gain coding requires 6 bits/frame. Considering from the viewpoint of hardware implementation, the better sampling rate would be 14.4 kHz which is twice the sampling rate of the lower rate coders of the digitizer. In our simulation we could not detect any degradation of output speech quality by reducing the input sampling rate from 15.52 to 14.4 kHz. Accordingly, we have chosen the transmission rate of the high-rate HCDM coder to be 14.96 kbits/s.

B. The Multirate Algorithm

An algorithmic block diagram of the multirate speech digitizer is shown in Fig. 1. Since the individual algorithms have been described in detail in the literature, here we describe just the overall system with emphasis on the differences among the algorithms at different transmission rates.

Before the input speech is analyzed, it is first preprocessed. In preprocessing, the input speech is automatically gain-controlled (AGC), low-pass filtered with the cutoff frequency of 3 kHz, and preemphasized with the breaking point frequency of 400 Hz to reduce the spectral dynamic range of input speech [4]. The signal is then digitized using a 12-bit A/D converter at the sampling rate of 14.4 kHz if the transmission rate of 14.96 kbits/s is desired, or 7.2 kHz if other transmission rates are used. The sampling rate is selected using a panel switch.

For the 2.4 kbit/s coder, only the routines enclosed by the dotted lines are used. The digitized input speech is analyzed by the autocorrelation method to obtain ten prediction coefficients $\{a_i\}$ and the same number of reflection coefficients $\{k_i\}$ [4]. Also, speech samples are passed through an inverse filter to generate the prediction error or residual signal. This residual is used to estimate the gain (residual energy) and to make voiced/unvoiced (V/UV) decision. In our system the method of average magnitude difference function (AMDF) applied to the LPC residual signal has been used for pitch extraction [12]. Pitch period, reflection coefficients, and residual gain thus obtained are multiplexed after coding, and transmitted at the transmission rate of 2.4 kbits/s. In our system, the reflection coefficients have been coded by the piecewise linear quantization method [13], and the others have been coded by linear quantization. At the receiver, they are demultiplexed and decoded. Depending on whether the current speech frame is voiced or unvoiced, either the pitch generator or the random noise generator generates an excitation signal. The signal is gain-adjusted using the decoded gain and fed into the lattice-form synthesis filter.

For the 4.8 kbit/s coder all the routines shown in Fig. 1 are used. LPC analysis and pitch extraction are done in the same way as in the 2.4 kbit/s vocoder. In addition, coding of the residual signal is done. The residual signal obtained by inverse filtering in the LPC analyzer is first band limited to 600 Hz using a four-pole Butterworth low-pass filter to reduce the transmission rate [7]. The band-limited baseband residual is then encoded by HCDM [10], which will be described later.

To realize the transmission rate of 4.8 kbits/s, the baseband residual is down-sampled with the ratio of 3 to 1 so that the sampling rate is reduced to 2.4 kHz from the actual rate of 7.2 kHz. Before down-sampling, the dynamic range of the baseband residual is reduced by using a pitch prediction loop (see Fig. 1). This improves the signal-to-quantization noise ratio (SQNR) in HCDM coding of the baseband residual signal by about 2 dB over the case without the pitch prediction loop used. The bit stream at the output of the HCDM coder is multiplexed with coded pitch period, reflection coefficients, and residual gain, and then transmitted.

At the receiver, information related to these data is demultiplexed and decoded. The received bit sequence of the baseband residual signal is decoded by the HCDM decoder. The decoder output is interpolated and up-sampled to obtain the original sampling rate (i.e., 7.2 kHz). The up-sampled signal is smoothed by a four-pole Butterworth low-pass filter with the cutoff frequency of 600 Hz. The smoothed output is then fed to the inverse of the pitch prediction loop to obtain the decoded baseband residual signal.

To obtain the excitation signal to the synthesizer, high-frequency harmonics of the residual signal must be regenerated using the decoded baseband residual, and they must be spectrally flattened. For this purpose we have used the pitch implantation method [14] which is shown in Fig. 2. This method is known to be very effective in high-frequency generation and to yield excellent synthetic speech [15]. In this method high frequencies are generated by adding pitch pulses (for voiced speech) or random noise (for unvoiced speech) to the decoded baseband residual. When random noise is added, it is

[1] This scheme is essentially the same as the embedded vocoding concept [11].

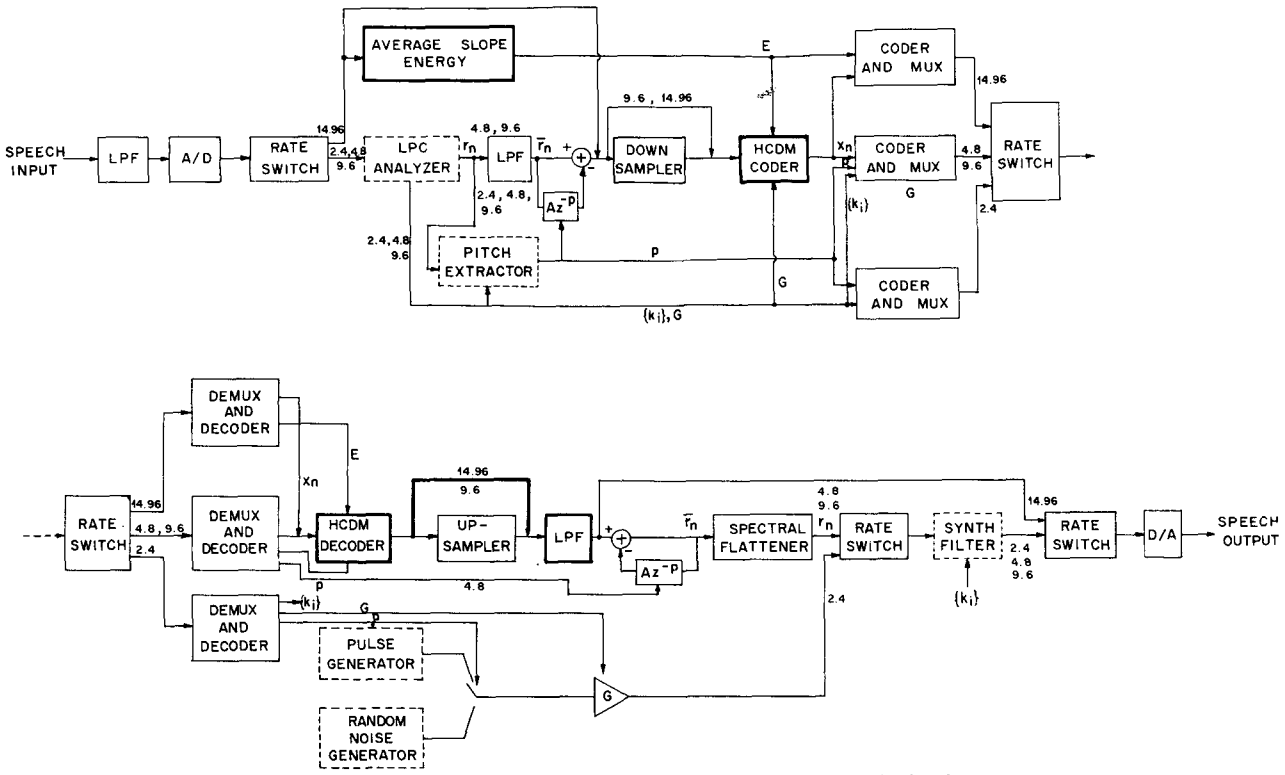


Fig. 1. Functional block diagram of the multirate speech digitizer system.

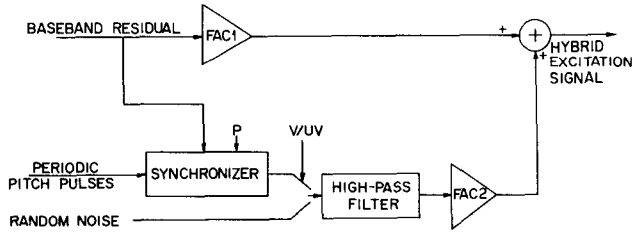


Fig. 2. Generation of RELP excitation signal by the pitch implantation method.

done asynchronously; but when pulses are added, it must be done synchronously with the baseband residual. To implant pitch pulses on the baseband residual, the "blanking-and-rundown" method [14] has been used. The pitch-implanted or random noise-added residual signal is fed through a four-pole Butterworth high-pass filter with the cutoff frequency of 600 Hz before they are summed with the decoded baseband residual. This output signal becomes the excitation signal of the synthesis filter. The synthesizer output signal is converted into an analog signal, and finally deemphasized with the breaking-point frequency of 400 Hz.

For the 9.6 kbit/s coder, all the routines used in the 4.8 kbit/s coder are used. But in this case the procedure of down-sampling and up-sampling is not necessary because 7.2 kbits/s is available for transmission of the coded residual signal. In the 4.8 kbit/s coder the original and decoded residual signals have been low-pass filtered with the cutoff frequency of 600 Hz. The same filtering operation is done in the 9.6 kbit/s coder, but the cutoff frequency is extended to 800 Hz for better speech quality.

For the 14.96 kbit/s coder, only the routines enclosed by the heavy lines are used. The digitized input speech is directly fed into the HCDM encoder. ADM is the most effective method in low-rate (below 24 kbits/s) waveform coding [16]. Particularly, HCDM is known to yield superior performance among various ADM systems [2], [10]. The HCDM system with feedforward estimation of the long-term step size is shown in Fig. 3. It uses both syllabic and instantaneous schemes in companding the quantizer step size. The former is used to update the long-term basic step size of the quantizer according to input signal energy variation, and the latter is used for changing the step size at every sampling instant based on the three consecutive sign bits. Details of the HCDM system including performance analysis may be found in [10].

One may note that since no reduction of spectral dynamic range is needed in a waveform coder, preemphasis and deemphasis are not done in the 14.96 kbit/s HCDM coder. At the receiver the HCDM decoded output is smoothed by a low-pass filter with the cutoff frequency of 3 kHz, and directly fed into the D/A converter.

C. Computer Simulation Results

The multirate algorithm described above has been simulated on a computer to determine its effectiveness and to optimize the parameter values. The optimum values of various parameters determined by simulation are shown in Table I. For determination of the optimal HCDM parameter values we have used the conventional SQNR measure; and for the gains (FAC1, FAC2) of the RELP excitation signal generator we have relied on informal subjective listening tests.

After fixing the algorithm, we performed 16-bit integer

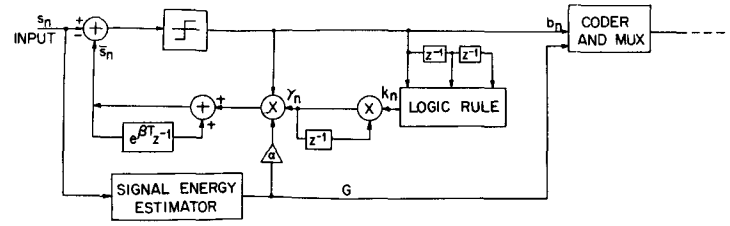


Fig. 3. Block diagram of HCDM encoder.

TABLE I
OPTIMUM PARAMETER VALUES

Transmission rate(kbits/s)	A	α	β	FAC1	FAC2
14.96	-	1.6	0.97	-	-
9.6	0.4	0.3	0.98	1	0.8
4.8	0.4	0.3	0.98	1	0.8

Note: A: Predictive loop constant (see Fig. 1).
α: Scale factor for basic step size estimation in HCDM (see Fig. 3).
β: Prediction filter leakage factor in HCDM (see Fig. 3).
FAC1, FAC2: Gains used in the RELP excitation signal (see Fig. 2).

simulation to predict the performance of the hardware to be built and to obtain the necessary scaling and normalization so that no underflow and overflow may occur. The results of the integer simulation are shown in Fig. 4. As one can expect, the waveform of the synthesized speech signal resembles more closely that of the input speech signal as the transmission rate increases. In addition to examining waveforms of the synthesized speech signals by eye, we have also performed informal listening tests. According to our subjective listening test, the speech quality was the best when the rate was 14.96 kbits/s and the worst when the rate was 2.4 kbits/s. At 9.6 kbits/s the synthesized speech sounded almost identical to the original speech. But at 4.8 kbits/s the synthesized speech had a little background echo. At 2.4 kbits/s the synthesized speech sound was not so natural as at 9.6 or 4.8 kbits/s, but its quality was satisfactory for the low rate.

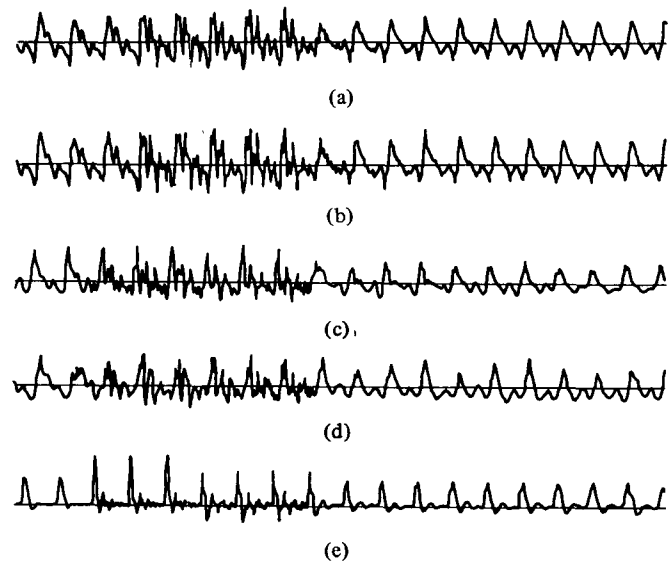


Fig. 4. Comparison of waveforms at different rates with the original speech. (a) Original. (b) 14.96 kbits/s. (c) 9.6 kbits/s. (d) 4.8 kbits/s. (e) 2.4 kbits/s.

III. FIRMWARE IMPLEMENTATION

A. Hardware Architecture

A simplified block diagram of the multirate speech digitizer hardware is shown in Fig. 5. The hardware architecture is very similar to that of the two-channel LPC vocoder developed by Shin *et al.* [17]. It can be divided largely into two parts: the processor part and the I/O part. The processor part is composed of four main sections: a central processing unit (CPU), a microprogram control unit (MCU), an external memory, and a multiplier. The CPU is interfaced with other subsystems (memory and I/O devices) for fast speed via two unidirectional buses [an input data bus (IBUS) and an output data bus (OBUS)] instead of the conventional bidirectional data bus. All instructions for this hardware are executed in a cycle time of 204 ns by a 4.896 MHz system clock.

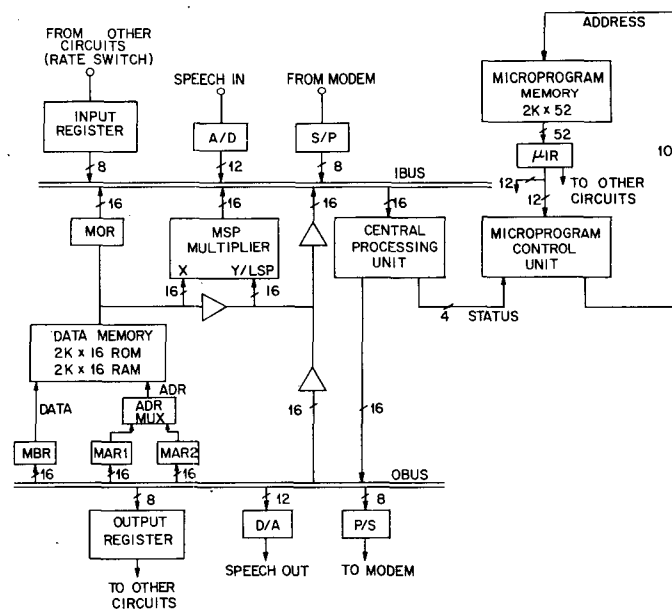


Fig. 5. Block diagram of the multirate speech digitizer hardware.

The center of this system is the CPU which is composed of a 16-bit arithmetic logic unit (ALU), a status register, a shift multiplexer, and a carry input multiplexer. The 16-bit ALU is composed of four Advanced Micro Devices (AMD) Am 2903 bipolar microprocessor slices and one Am 2902 carry-lookahead generator chip. The 16-bit CPU fetches data from one of the seven data sources: microinstruction register (μIR), memory output register (MOR), A/D converter, serial-to-parallel (S/P) converter, the upper and lower parts

CPU CONTROL (23)

2	4	4	4	4	3	2
ALU SOURCE	ALU FUNCTION	ALU DESTINATION	ALU A-ADDRESS	ALU B-ADDRESS	SHIFT	CARRY
$\bar{E}_A$, I_0	$I_1 - I_4$	$I_5 - I_8$	$A_0 - A_3$	$B_0 - B_3$	$S_0 - S_2$	$C_0 - C_1$

MCU CONTROL (10) | OTHER PART (20)

4	4	1	1	3	3	1	12	1
SEQUENCER	CONDITION MUX	STATUS REG.	INTERRUPT F/F CLEAR	IBUS	OBUS	AUTO-INCRE-MENT	DIRECT FIELD	HALT
$M_0 - M_3$	K_0-K_2, POL	TST	INTCLR	IB_0-IB_2	OB_0-OB_2	INC	$F_0 - F_{11}$	HLT

Fig. 6. Microinstruction format.

of the multiplier output, and the input register via the unidirectional tristate IBUS. The output of the CPU is transferred to one of the seven destinations: D/A converter, parallel-to-serial (P/S) converter, two memory address registers (MAR1 and MAR2), memory buffer register (MBR), Y input of the multiplier, and output register.

The microprogram memory is a 2K word by 52 bit PROM containing microinstructions. The format of a microinstruction is shown in Fig. 6, in which the halt bit is used during debugging operation only. Microinstructions stored in the PROM addressed by the MCU are transferred into the microinstruction register (μIR or "pipeline register") in synchronization with the system clock. The MCU is composed of an AM 2910 sequencer, a condition multiplexer, and an interrupt control unit (ICU).

The external memory is composed of 2K RAM's and ROM's. The RAM's are used for data buffering and for temporary memory, and the ROM's store various look-up tables needed to implement the multirate speech digitizer algorithm. To enhance the data processing capability, we use the pipeline technique [18] in data transfer between the CPU and the external memory. Also, we use two address registers with autoincrement function and an address multiplexer to facilitate the alternating access of two block data sets in different memory locations without additional addressing from the CPU. Accordingly, it is easy to copy or transfer a data set from one location to another as done in buffer copying, and also easy to calculate the sum of products of the two data sets $\{X_i\}$ and $\{Y_i\}$, as done in autocorrelation calculations.

A TRW MPY-16HJ multiplier accepts two 16-bit two's complement numbers (XIN and YIN) from the external memory or the OBUS and makes the full 32-bit product available to the CPU in two 16-bit pieces (LSP and MSP). The pipeline technique has been used for efficient multiplications as used for accessing the external memory, since the input and output of the multiplier are fully buffered and the multiplication time is less than one machine cycle.

The I/O system for the multirate speech digitizier hardware consists of three input devices (A/D, S/P, and an input register), three output devices (D/A, P/S, and an output register), and a device clock system. The device clock system generates various clocks to control the I/O devices and provides the interrupt request to the ICU.

In the multirate speech digitizer hardware, the sampling rates of A/D and D/A converters must be varied according to the transmission rate used. In our system the sampling rate is 14.4 kHz when speech is transmitted at 14.96 kbits/s, and 7.2 kHz when it is transmitted at 9.6, 4.8, or 2.4 kbits/s. This has been realized by using one D-flip-flop and one two-input multiplexer. The input selection control to the multiplexer is given from the panel switch.

A microinstruction is made up of 53 bits, the format of which is shown in Fig. 6. One instruction word is divided into CPU, MCU, and other control parts. In the CPU control part, source code, function code, destination code, and A-address and B-address control the 16-bit ALU. Other parts of the CPU control fields are used to control shift multiplexers and the carry input multiplexer. The MCU control part consists of sequencer and condition multiplexer control fields, a status register control bit, and an interrupt clear bit. The direct field is used as a direct input address to the 2910 sequencer and an operand to the CPU. The I/O control fields determine where the CPU gets its input and where its output is to go. The autoincrement bit controls whether or not to increment the two address registers (MAR1 and MAR2) at the same time. During debugging operation, the halt bit is used to stop the speech digitizer hardware.

Although it is not explicitly shown in Fig. 5, the analog part used in our system includes automatic gain control (AGC), preemphasis, and deemphasis circuits. As preemphasis and deemphasis circuits have no distinctive features compared with others, we discuss only the AGC circuit here. The voltage transfer characteristic of our AGC circuit is shown in Fig. 7. A feature of the AGC circuit is that it incorporates a silence detection circuit. When input speech is silent, the silence detector pulls down the AGC output to the ground level, thus preventing the input noise from being amplified. This has a good effect on our system, especially for the 2.4 kbit/s LPC coder that is very susceptible to environmental noise.

A summary of hardware design of the speech digitizer is shown in Table II.

B. Software Implementation

The real-time microprogrammed software of our multirate speech digitizer system can be divided largely into five routines: initialization routine, rate checking routine, main

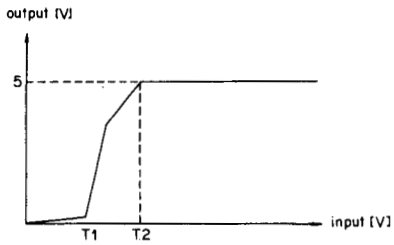


Fig. 7. Voltage transfer characteristics of automatic gain control circuit.

TABLE II
SUMMARY OF SPEECH DIGITIZER HARDWARE

Cycle time	204 ns
Basic logic family	TTL
Microprogram memory (ROM)	1K x 52 bits
	Thirteen 82S137's
External memory (ROM)	2K x 16 bits
	Eight 82S137's
External memory (RAM)	2K x 16 bits
	32 FCLD 93425's (1K x 1)
Hardware multiplier	1 TRW MPY 16-HJ
Microprocessor	4 Am 2903's (4-bit slice)
Microsequencer	1 Am 2910 (12-bit wide)
A/D Conversion	12-bit A/D, D/A conversion at
	7.2 kHz or 14.4 kHz sampling rate
Total power dissipation	5V x 10A = 50W max
Hardware size	19" x 15" x 5.25"

routine, idle routine, and A/D-D/A interrupt service routine.[2] The overall block diagram of the software is shown in Fig. 8. In the initialization routine all RAM data memories are cleared and the initial parameter values are set. In the rate checking routine the rate switch is checked via the input register. Whenever the rate switch is changed, the system is reinitialized. The idle routine is inserted to fill in the time intervals between the frames. In this routine, no operation is performed except interrupt checking and waiting until the speech input buffer is full and at the same time the speech output buffer is exhausted.

The main routine is the core of software. The overall flow chart of the main routine is shown in Fig. 9. At the start of the main routine, the synthetic speech data buffer of the previous frame is first copied to the output speech data buffer, and the input speech data buffer is copied to the processing buffer. The length of the input buffer is 180 samples and that of the processing buffer is 220 samples. Hence, the length of

[2] In addition, one needs to have an S/P and P/S interrupt routine for transmission and reception of coded data. Since this routine is of standard nature, it will not be discussed here.

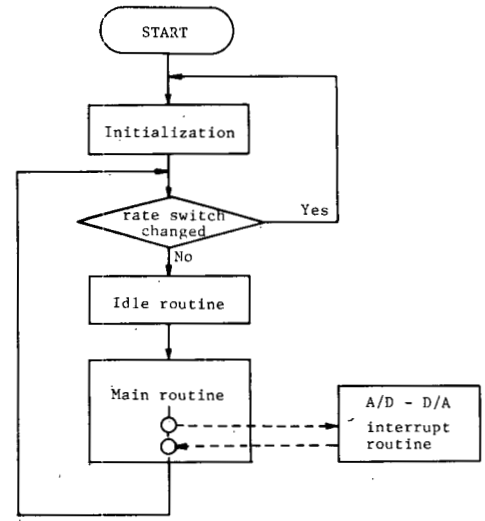


Fig. 8. Overall block diagram of multirate speech digitizer software.

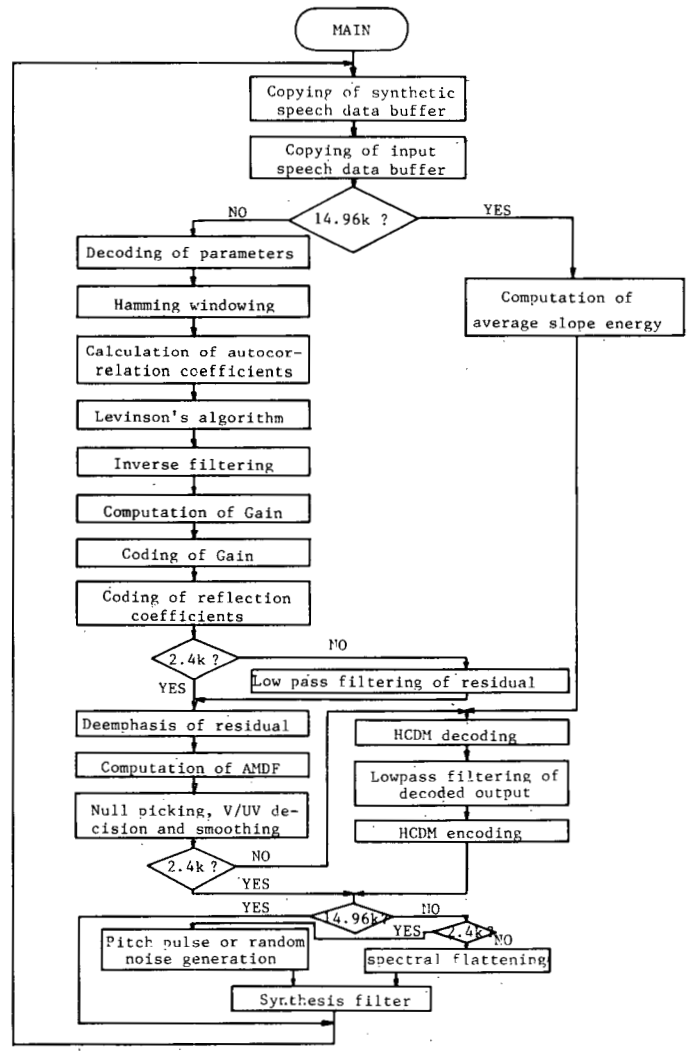


Fig. 9. Flow chart of main routine.

overlap between the frames is 40 samples. When the process of buffer copying is completed, speech analysis can start.

LPC analysis is performed for 2.4, 4.8, and 9.6 kbit/s coders. First, the parameters that have been coded three frames before (this delaying is required for smoothing pitch errors) are decoded. Here, the reflection coefficients $\{k_i\}$ are decoded using a look-up table to reduce the computation time. Second, the input speech data in the processing buffer are Hamming-windowed using the window values stored in ROM to generate windowed speech data, which are then used to calculate autocorrelation coefficients. To minimize the truncation error, autocorrelation coefficients are normalized with respect to signal power. By the modified Levinson's algorithm, the reflection coefficients and the prediction coefficients are calculated using these normalized autocorrelation coefficients. Third, using those prediction coefficients, the unwindowed speech data in the processing buffer are inverse filtered to generate the residual signal. The residual signal thus obtained is used to estimate the gain. Fourth, the reflection coefficients obtained by the modified Levinson's algorithm are coded by piecewise linear quantization using the number of quantization levels at each interval prepared in ROM. At this point, if the rate switch is set at 4.8 or 9.6 kbits/s, the residual is band limited using a four-pole Butterworth low-pass filter with the filter coefficients stored in ROM. For 2.4 kbit/s coding, the residual is not band limited.

One may note that pitch extraction is done for 2.4, 4.8, and 9.6 kbit/s coders. First, the non-band-limited residual is deemphasized with the breaking-point frequency of 100 Hz. Using the deemphasized residual, AMDF is computed, for which we have used every fourth sample of the residual to reduce the computation time. That is,

$$\text{AMDF}(p) = \sum_{n=0}^{30} (r_{4n} - r_{4n+p}),$$

$$p = 24, 25, \cdots, 39, 40, 42, 44, \cdots, 130, 132. \tag{1}$$

Since we have obtained AMDF for only 64 nonlinearly quantized levels of p, the value of p can be regarded as the coded pitch period. Using the AMDF values, the pitch period is determined by null-picking. In addition, V/UV decision is made based on AMDF and input energy.

Next, for 4.8 or 9.6 kbit/s coding, the band-limited residual is encoded by HCDM and stored in a buffer with the format of 15 bits of coded bit sequence per one buffer location. Also, the coded residual of the frame that precedes three frames before the current frame is decoded by HCDM and smoothed by a four-pole Butterworth low-pass filter.

In the 14.96 kbit/s coder, LPC analysis is skipped and the input speech data in the processing buffer are directly applied to the HCDM coding routine. Since average slope energy is used to estimate the basic step size, it must be estimated before the HCDM coding routine.

Finally, output speech is synthesized using the decoded parameters. In the 14.96 kbit/s coder, no operation is performed in the synthesis routine. But, for other rates, excitation signal should be generated. For the 2.4 kbit/s coder, pitch pulses or random noise are generated. To generate random noise, 256 samples of random numbers are prepared in ROM. For 4.8 and 9.6 kbit/s coders, the decoded baseband residual signal is spectrally flattened by the pitch implantation method. The excitation signal thus obtained is fed into the synthesis filter of lattice form.

In the main routine, the A/D-D/A interrupt request should be checked properly at the interval of about 680 instructions when the sampling rate is 7.2 kHz, and at the interval of about 340 instructions when the sampling rate is 14.4 kHz. Hence, in the routines of the 14.96 kbit/s coder, interrupt should be checked two times more frequently than the other routines. Whenever A/D-D/A interrupt occurs, the A/D converted speech input is stored in the input buffer and the output speech data in the output buffer are D/A converted.

C. Implementation Results

During the development of the real-time microprogrammed software, we have performed emulation under the control of a NOVA-4 minicomputer. By controlling the multirate system to execute instructions in real time to the system clock, and also in single step to the clock given by NOVA when the halt bit (see Fig. 6) is set,[3] we could test the multirate system successfully.

The quality of speech synthesized by the real-time speech digitizer was identical with that of integer simulation at each rate. Also, a good agreement was found between the synthetic speech waveform obtained by integer simulation and that collected by NOVA during real-time emulation. Therefore, our system proved to be working well.

Finally, let us consider timing of the system developed. As stated before, one frame of speech is composed of 180 samples. Therefore, the frame length of 2.4, 4.8, and 9.6 kbit/s coders (the sampling rate is 7.2 kHz) is 25 ms and that of a 14.96 kbit/s coder (the sampling rate in this case is 14.4 kHz) is 12.5 ms. Since the system clock is 4.896 MHz, the number of instructions executable during one frame is 122 400 for the 2.4, 4.8, and 9.6 kbit/s coders and 61 200 for the 14.96 kbit/s coder. In Table III the number of instruction cycles and time required for execution of each routine is shown. It is seen that in the worst case only 77.66 percent of one frame time is used to execute all the routines except the idle routine.

IV. CONCLUSIONS

Implementation of a multirate speech digitizer having transmission rates of 2.4, 4.8, 9.6, and 14.96 kbits/s has been

[3] During a single step execution, the values loaded on the registers, memories, and input and output bus could be read or modified.

TABLE III
NUMBER OF INSTRUCTIONS PER FRAME FOR CODERS AT DIFFERENT TRANSMISSION RATES

Functions	2.4 kbits/s	4.8 kbits/s	9.6 kbits/s	14.96 kbits/s
MAIN ROUTINE				
Copying I/O buffer	910	910	910	910
Decoding of parameters	232	232	232	-
LPC analysis	17,052	17,052	17,052	-
Low-pass filtering of residual	-	8,103	8,103	-
Pitch extraction	12,576	12,576	12,576	-
Average slope energy computation	-	-	-	1,888
HCDM decoding[a]	-	14,374	15,082	14,722
HCDM encoding	-	6,597	6,429	6,428
Speech synthesis	14,760	23,220	23,040	-
INTERRUPT SUBROUTINE				
A/D-D/A subroutine	1,800	1,800	1,800	1,800
P/S subroutine	2,268	2,916	4,212	2,106
S/P subroutine	3,024	3,888	5,616	2,808
Total	52,622 (10.735 ms)	91,668 (18.700 ms)	95,052 (19.391 ms)	30,662 (6.255 ms)

(1 instruction = 204 ns)

[a] The execution time of low-pass filtering of HCDM decoded output is included.

presented. The hardware complexity is almost the same as that required for the conventional 2.4 kbit/s LPC vocoder except an additional 50 percent increase of memory size. The speech quality of the digitizer at one transmission rate is just as good as that of any other digitizer designed for that transmission rate only. As one can expect, the quality and robustness to environmental noise become better as the transmission rate increases. We believe that the multirate speech digitizer developed is advantageous over other similar systems in system compactness and design efficiency.

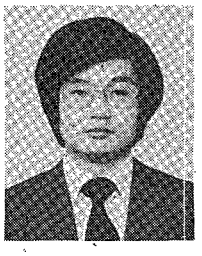
ACKNOWLEDGMENT

The authors wish to thank the reviewers for their helpful comments and suggestions in improving the earlier version of the paper.

REFERENCES

[1] N. S. Jayant, Ed., *Waveform Quantization and Coding*. New York: IEEE Press, 1976.
[2] C. K. Un and H. S. Lee, "A study of the comparative performance of adaptive delta modulation systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 96–101, Jan. 1980.
[3] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, pp. 131–133, Apr. 1975.
[4] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. New York: Springer-Verlag, 1976.
[5] B. Gold *et al.*, "New applications of channel vocoder," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 13–23, Feb. 1981.
[6] C. K. Un, "A low-rate digital formant vocoder," *IEEE Trans. Commun.*, vol. COM-26, pp. 344–355, Mar. 1978.
[7] C. K. Un and D. T. Magill, "The residual-excited linear prediction vocoder with transmission rate below 9.6 kbits/s," *IEEE Trans. Commun.*, vol. COM-23, pp. 1466–1474, Dec. 1975.
[8] E. E. David, Jr. *et al.*, "Voice-excited vocoders for practical speech bandwidth reduction," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 101–105, Sept. 1962.
[9] B. S. Atal and M. R. Schroeder, "Adaptive predictive coding of speech signals," *Bell Syst. Tech. J.*, pp. 229–242, Oct. 1970.
[10] C. K. Un, H. S. Lee, and J. S. Song, "Hybrid companding delta modulation," *IEEE Trans. Commun.*, vol. COM-29, pp. 1337–1344, Sept. 1981.
[11] D. J. Goodman, "Embedded DPCM for variable bit rate transmission," *IEEE Trans. Commun.*, vol. COM-28, pp. 1040–1046, July 1980.
[12] C. K. Un and S. C. Yang, "A pitch extraction algorithm based on LPC inverse filtering and AMDF," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 565–572, Dec. 1977.
[13] ——, "Piecewise linear quantization of LPC reflection coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 1977.
[14] C. K. Un and W. Y. Sung, "A 4800 bps vocoder with improved excitation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Apr. 1980, pp. 142–145.
[15] C. K. Un and J. R. Lee, "On spectral flattening techniques in residual excited linear prediction (RELP) vocoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 1982.
[16] C. K. Un and M. H. Cynn, "A performance comparison of ADPCM and ADM coders," in *Proc. IEEE Nat. Telecommun. Conf.*, 1980.
[17] B. C. Shin, H. H. Lee, C. K. Un, S. Z. Lee, and K. H. Song, "Implementation of a two-channel LPC vocoder," *IEEE Trans. Commun.*, vol. COM-31, July 1983, to be published.
[18] "The AM 2900 family data book AM-BUB 003," Advanced Micro Devices, Inc., Sunnyvale, CA, 1979.

★



Byung Suk Lee was born in Seoul, Korea, in 1957. He received the B.S. degree in electronics engineering from Seoul National University, Seoul, in 1980, and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (K.A.I.S.T.), Seoul, in 1982.

He is currently employed at Gold Star Electric Company, Osan, Korea.

★

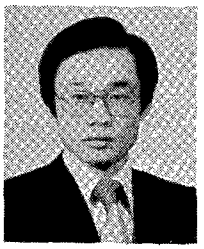
Chong Kwan Un (S'63–M'64–SM'81) was born in Seoul, Korea, in 1940. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Delaware, Newark, in 1964, 1966, and 1969, respectively.

From 1964 to 1969 he was a Research Fellow at the University of Delaware. From 1969 to 1973 he was Assistant Professor of Electrical Engineering at the University of Maine, Portland, where he taught communications and did research on synchronization problems. In May 1973, he joined the

staff of the Telecommunication Sciences Center, SRI International, Menlo Park, CA, where he did research on voice digitization and bandwidth compression systems. Since June 1977 he has been with the Korea Advanced Institute of Science and Technology (K.A.I.S.T.), Seoul, Korea, where he is Professor and Director of the Communication Research Laboratory, teaching graduate courses and doing research in the area of digital communication systems and digital signal processing. At present he is also Dean of the Faculty of Engineering at K.A.I.S.T.

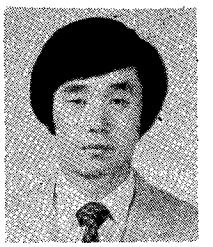
Dr. Un received a number of awards including the 1976 Leonard G. Abraham Prize Paper Award from the IEEE Communications Society and the National Order of Merits from the Government of Korea. He is a member of Tau Beta Pi and Eta Kappa Nu.

★



Hyeong Ho Lee (S'77) was born in Muju, Korea, in 1955. He received the B.S.E.E. degree from Seoul National University, Seoul, Korea, in 1977, and the M.S.E.E. degree from the Korea Advanced Institute of Science and Technology (K.A.I.S.T.), Seoul, in 1979.

He is currently working towards the Ph.D. degree at K.A.I.S.T. His research interests include voice digitization, integrated voice/data communication, and the microcomputer applications in digital communication systems.



Byoung Cheol Shin (S'80) was born in Choongju, Korea, in 1952. He received the B.S.E.E. degree from Seoul National University, Seoul, Korea, in 1975, and the M.S.E.E. degree from the Korea Advanced Institute of Science and Technology (K.A.I.S.T.), Seoul, in 1977.

From 1977 to 1979, he worked on the development of microcomputer systems at the Korea Institute of Electronics Technology. He is presently working towards the Ph.D. degree at K.A.I.S.T. His current interests include digital processing of speech signal, integrated voice/data communication systems, and industrial applications of microprocessors.

★

Hwang Soo Lee (S'79–M'83) was born in Seoul, Korea, in 1952. He received the B.S. degree in electrical engineering from Seoul National University, Seoul, in 1975, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (K.A.I.S.T.), Seoul, in 1978 and 1983, respectively.

In 1975 he was with the Hyundai Heavy Industries Company, where he was involved in designing marine instrumentation and automatic navigation systems. In March 1983, he joined K.A.I.S.T. as a Research Scientist and Assistant Professor of Electrical Engineering. His current research interests include digital coding and bandwidth compression of speech and image signals, and the design and implementation of a computer communication network.