Reprinted from APPLICATION OF SIGNAL PROCESSING CONCEPTS IN HEARING AIDS
Edited by Paul Yanick, Jr., M.A. and Stephen Freifeld, M.D., F.A.C.S.

Chapter 11

A MICROPROCESSOR-BASED SPEECH PROCESSOR

H. Drucker, Ph.D.

Background

Persons with sensorineural hearing loss suffer typically from recruitment, an inadequate frequency response, and a reduced dynamic range. In Figure 1 is shown the frequency response for an idealized sensorineurally-impaired person. Compared to the normal hearing, the discomfort level for the hearing impaired has been reduced and the threshold has been increased. This causes a reduced dynamic range, most significantly at higher frequencies. Superimposed is the range of conversational speech (1). The dynamic range of conversational speech at high frequencies is thus greater than the hearing impaired's dynamic range. With amplification, we can move the dotted lines representing the extremes of conversational speech up. With filtering, we can emphasize the higher frequencies--but the dynamic range of the conversational speech cannot be changed by conventional amplification or filtering. Because the actual dynamic range of the speech is greater than the perceived dynamic range at high frequencies, the listener does not receive all information necessary to determine the identity of individual speech sounds. Ordinarily, because of the redundancy present in speech, a listener can fill in any perceptual gaps. However,

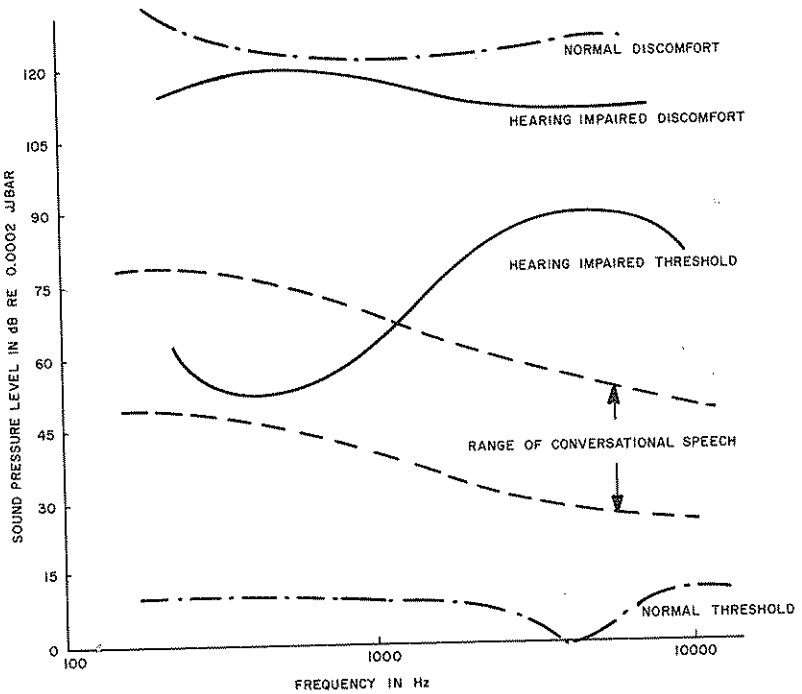


Fig. 1. Discomfort and thresholds for hypothetical hearing impaired and normal persons. Superimposed is the conversational level of speech. (From Drucker and Yanick, 1976).

the higher frequency sounds carry more of the speech intelligibility than the low frequency sounds (2) since the high frequency sounds are typically consonants, the low frequency sounds are typically vowels, and the consonants are the main information bearing components of speech.

Recruitment is a common symptom of the sensorineurally impaired. To the recruiting ear, incremental changes in input volume cause a larger perceived change. Thus loud sounds may sound unnaturally loud. Sometimes a recruitment threshold is present below which the perception of loudness is normal. Both the threshold and the degree of recruitment are functions of frequency. Recruitment impairs the perception of consonants following vowels. In general, the total

energy in consonants is less than in vowels so that the vowels, being perceived as abnormally loud mask the following consonants with a resultant loss in intelligibility. The recruiting ear does not function well in a competing noise environment where the noise is speech-like such as at cocktail parties, cafeterias, and in general where more than two people are trying to converse.

SOLUTIONS

A partial solution to the reduced dynamic range problem is the use of a non-linear amplifier, termed a compressor, to reduce the dynamic range of the speech into the listener's dynamic range. The appropriate curves are shown in Figure 2. Figure 2a shows the input-output characteristics of a compressor with a slope of one (45°). When the slope is less than unity, input changes are reflected as smaller changes in the output. Chosen to be approximately the ratio of the person's perceived dynamic range to the dynamic range of conversational speech is k. Since the dynamic range is a function of frequency, so must the slope - we shall show how the variation in k is accounted for later in this paper. Figure 2b shows how the gain of the compressor varies with the input amplitude. This characteristic for a conventional amplifier will be a horizontal line but for the compressor, the slope is negative. Thus the gain is high for low amplitude sounds and low for high amplitude sounds.

Although the compressor is a step in the right direction, it needs further refinement because the compressor, in making loud sounds softer, makes soft sounds louder. This causes low level background noise to be raised to annoyance levels. This may be rectified by insertion of a threshold (Fig. 3) below which the amplifier is in expansion. The slope of the gain curve below threshold is positive reducing those inputs below the threshold to inconsequential levels. The input range over which the amplifier compresses is the range of conversational speech (30-36 dB). The compression slope is set as indicated before and will vary among individuals. The expansion slope is not critical

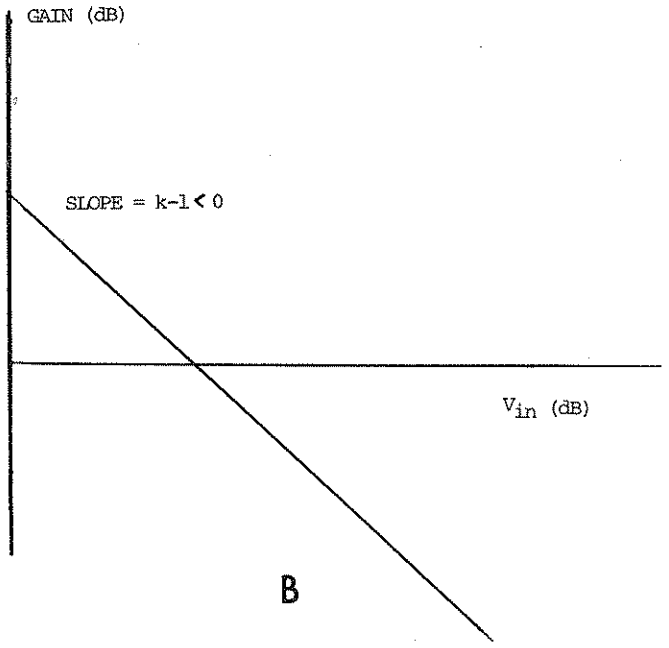
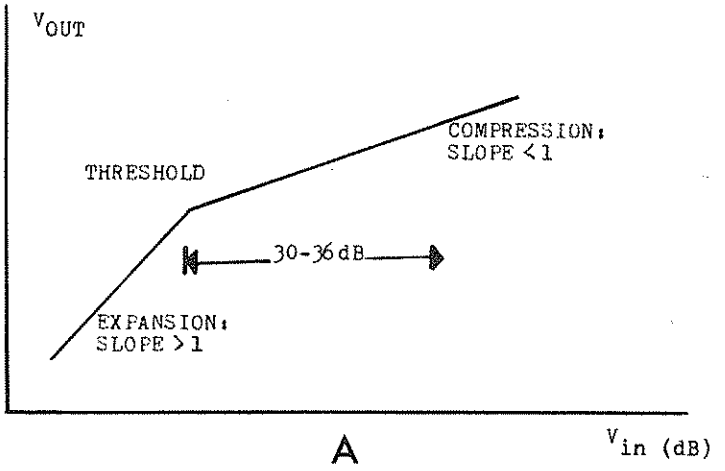


Figure 2. Input-Output (a) and Input-Gain (b) curves for compression amplifier. (From Drucker and Yanick, 1976).



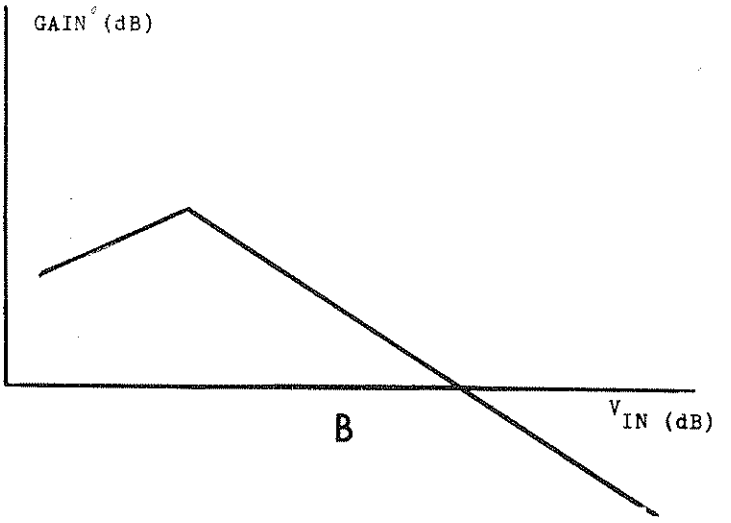


Figure 3. Input-Output (a) and Input-Gain (b) curves for amplifier with threshold.

and is set to 1.2 for all persons. The type of amplifier described here should not be confused with what is termed output compression amplifier. In output compression amplifiers, the amplifier is linear over most of its usable range and goes into hardlimiting only when the input is at such a level as to cause saturation of the hearing aid receiver or cause an output level greater than 120 dB SPL.

Surprisingly enough, hearing aids with the amplifiers described here do not function better than conventional aids, probably due to the vowel masking of consonants as described before since even with compression the vowels are still more intense than the consonants. The two-channel compression system in Figure 4 is much better however (3). Filtering is done at 1500 Hz to separate the speech band into (approximately) high frequency consonants and low frequency vowels. The slope for the compressor in each band is set depending on the perceived dynamic range in that band and the range of conversational speech in that band (36 dB for the high frequency band and 30 dB for the low band). The maximum output level for each compression amplifier is the same so that the consonants have the same intensity as the vowels. The low and high frequency bands are combined after compression and equalized to the listener's preference.

Another consideration in compression-amplifier design is the choice of attack and release time. In Figure 5, we show the output waveform when the input is a square wave. When the input goes from a low level to a high level, the amplifier goes from a high gain state to a low gain state. Because of the finite attack time in going to the low gain, the output amplitude overshoots before reaching a steady state low gain situation. The attack time is set at 2 msec. A similar situation exists when the input value is reduced thereby going from a low gain state to a high gain state causing an undershoot in the output waveform.

This release time is very critical - too short and the speech sounds unnatural - too long and low level sounds following loud sounds will not be amplified to perceivable levels. The release time is set at 20 msec for the low band and 10 msec for the high band. Explicitly these release times are defined as the time

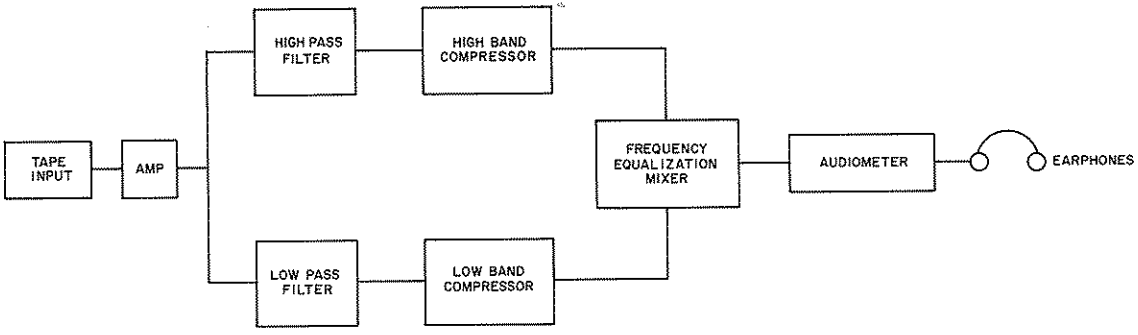


Fig. 4. A block diagram of a signal processing system.

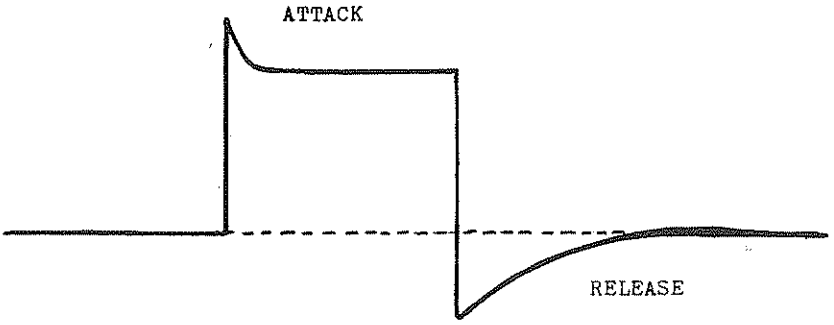


Figure 5. Showing attack and release times in the output of a compression amplifier when the input is a square wave.

it takes the output to be within 2 dB of the final value due to an input level change of 40 dB (square wave modulated sinusoid).

A series of experiments on persons with sensorineural hearing impairments showed the increase in intelligibility using a prototype two-band system over conventional hearing aids (4). Intelligibility tests were run at 0 and 6 dB signal-to-noise levels where the noise was cafeteria noise. The subjects used their own hearing aids and the two-channel system. The compression ratios were experimentally determined by the patients' listening to an endless audio tape loop of speech and noise and adjusting parameters to their optimum perceivable levels. At a signal-to-noise level of 6 dB, the two-band system had an average intelligibility of 87% vs. 46% for conventional aids and at a 0 dB signal-to-noise ratio, the values were 69% vs. 21%, respectively.

MICROPROCESSOR-BASED SPEECH TRAINER

The techniques discussed in the last section can be used as hearing aids or speech trainers. The construction of a hearing aid, while not a simple matter, is a straightforward application of microcircuit techniques and is state-of-art to many electronic

industries although the hearing aid industry tends to lag. In this paper we concentrate on the applications of microprocessors, a new tool in the hearing field to speech trainers. In a speech trainer, one speaker (the instructor) is talking to one or many listeners. The microprocessor can be more cost-efficient than conventional analog techniques as we shall see.

A microprocessor plus memory and peripheral (Fig. 6) makes up what is termed a microcomputer. A microcomputer is nothing but a small computer that in this application can fit on a circuit board 4" by 4". Since many of you will be unfamiliar with microcomputer technology and since it will become an all invasive tool in the next few years, I am going to summarize some of the technical aspects. The microcomputer physically consists (Fig.6) of the microprocessor, memory, and peripherals. The microprocessor, which is usually called the CPU (Central Processing Unit) and the memory exist in the form of a silicon-based chip of material, sometimes only .1" by .1" in size upon which is deposited thousands of miniature electronic circuits. This is obtained in the manufacturing process by selectively adding "doping" impurities to the silicon chip. Because there are so many circuits on one small chip, it is an example of what we call LSI (Large Scale Integration). The design of these chips and the manufacturing process is very expensive but since the semiconductor industry sells millions of these chips, the cost per chip is relatively low. Microprocessors can be bought for as little as $10. In fact, in designing a microcomputer system, the cost of the physical devices, termed hardware is usually much less than the expense of writing the program, called the "software". The gap between hardware and software costs will further increase as the industry becomes more efficient.

The peripherals are used to communicate between the microcomputer and the outside world. Typical peripherals are teletypes, keyboards, TV's, card readers, line printers, and magnetic type storage devices such as tape or discs. The peripherals, also called I/O (Input/Output) devices are usually the most expensive of all hardware costs.

Memory is used to store the program that tells the computer what to do. A program consists of a

sequence of instructions that are decoded by the CPU and then executed. I will explain decoding in more detail. First of all, you must understand that the computer only works on certain types of numbers called binary. Binary numbers have only two values - 0 and 1. Humans are used to the decimal system which is the ten numbers, 0 to 9. Thus to count in binary we would count 0, 1, 10, 11, 100, 101, 110, 111, ... while in decimal we would count 0, 1, 2, 3, 4, 5, 6, 7, ... Although you see that it takes more binary digits (bits) to represent a number than using decimal notation, electronically it is easier to manufacture devices that hold two states - ON (for binary 1) and OFF (for binary 0) rather than ten states. Therefore, each instruction in the computer consists of a unique sequence of bits which when interpreted by the CPU, causes the CPU to take some action such as addition, subtraction, etc. Of course, the programmer must know what the instructions are in binary to write the program. Memory may be considered to be a collection of storage boxes. In the typical microcomputer, storage locations are specified by sequential 16 bit numbers starting from address 0000 0000 0000 0000 and terminating at 1111 1111 1111 1111 (this gives over 64,000 storage locations). In each box there are 8 bits which correspond to the instruction or part of an instruction. An analogy might be to a series of mailboxes. On each mailbox is the address and within the mailbox, the contents.

The operation of a typical cycle is coded via the numbers in Figure 6. (1) The PC (Program Counter) points at the address of the instruction to be executed. This instruction takes three storage locations. (2) The instruction is located in the decoder to be interpreted. In this example, the instruction is interpreted to mean to move the contents of 0000 0000 0000 0101 to another location. (3) The instruction is executed. The next step would be to change the PC to 0000 0000 0000 0100 and repeat the cycle.

What is the difference between the microcomputer and what people typically think of as being computers; the IBM 360/370, the UNIVACs, or the Control Data machines? The first difference is in their speed - while the micro may perform 10,000 operations per second - the larger computer may perform over a million

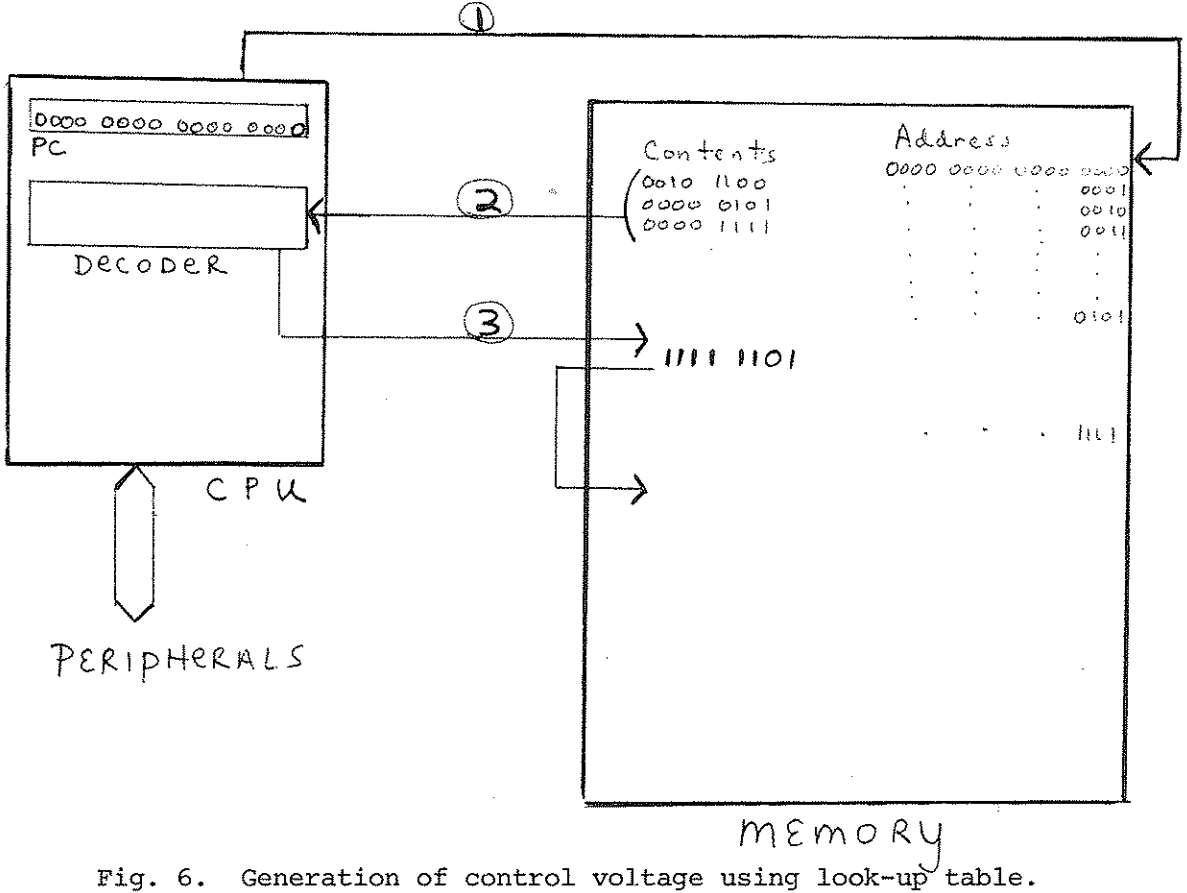


Fig. 6. Generation of control voltage using look-up table.

operations per second. Secondly, the numbers they handle will be much larger - while the microcomputer may only handle 8 bit numbers (the contents of each storage location), the large computer may handle 32 or 64 bits. Thirdly, the memory address may be specified by 32 bits rather than 16 bits. This means more than 64,000 times as many available memory locations and thus much larger programs may be written. Finally, the cost $10 versus hundreds of thousands of dollars. These are not all the differences, but it should be enough for you to see why the larger computers are ideal for "number crunching" applications or handling extensive data files while the micro is perfect for specialized, control applications as in the control of the ignition and fuel system in an automobile, the control of a home heating system, the control of automatic machinery, or in the application we are discussing here.

I do not think any of you need to worry about actually programming microcomputers unless you want to. Programming cannot be learned overnight but it is interesting and opens up different research lines. If you do use microcomputers, it will probably be on a "turnkey" system where you push a button on your spectrum analyzer or turn the key on your car and the microcomputer automatically starts its program which has already been inserted in memory by the manufacturer. As micros get faster and cheaper, they will take over many of the applications where analog circuitry is now used and most importantly, create new applications.

In our application, (Fig. 7) the microcomputer is used to read the input level to voltage controlled amplifier (VCA) and then control the gain of the amplifier. The computer functions internally using the type of numbering system termed digital (or two level--0's and 1's). Because the environment external to the computer is analog (or multi-valued) there must be a conversion between the internal and external environment. The ADC (Analog to Digital Converter) is a peripheral that takes the multilevel analog input and converts to digital representation. The ADC converter is termed 8 bit because it breaks the input range 0-10 volts into $2^8 = 256$ levels and each level is represented by a unique 8 binary digit code. For example 0 volts may be represented as digital 0000 0000 while 10 volts will be

represented internally as 1111 1111. Any other of the values between 0 and 10 volts will be represented by some other 8 binary digit sequence of 0's and 1's. A DAC (Digital-to-Analog Converter) does just the opposite of the ADC, converting an 8 bit internal representation to one of 256 analog output values, 0 to 10 volts.

Figure 7 shows a compressor constructed of a VCA (voltage controlled amplifier), the microcomputer, a full-wave rectifier, DAC, and ADC. The full-wave rectifier (Fig. 8) takes the input waveform, inverts all negative signals and smooths it with a 2 msec time constant (the attack time) so that any rapid ripples in the original waveforms are smoothed. The ADC then samples the input waveform once every .1 msec, converting the sampled analog value to an 8 bit digital representation. The computer then goes to a look-up table in memory, different for each user (Fig. 9). For each of the 256 input levels, there exists a value in the look-up table that is directed to the DAC after proceeding through a decrement algorithm in the computer. The

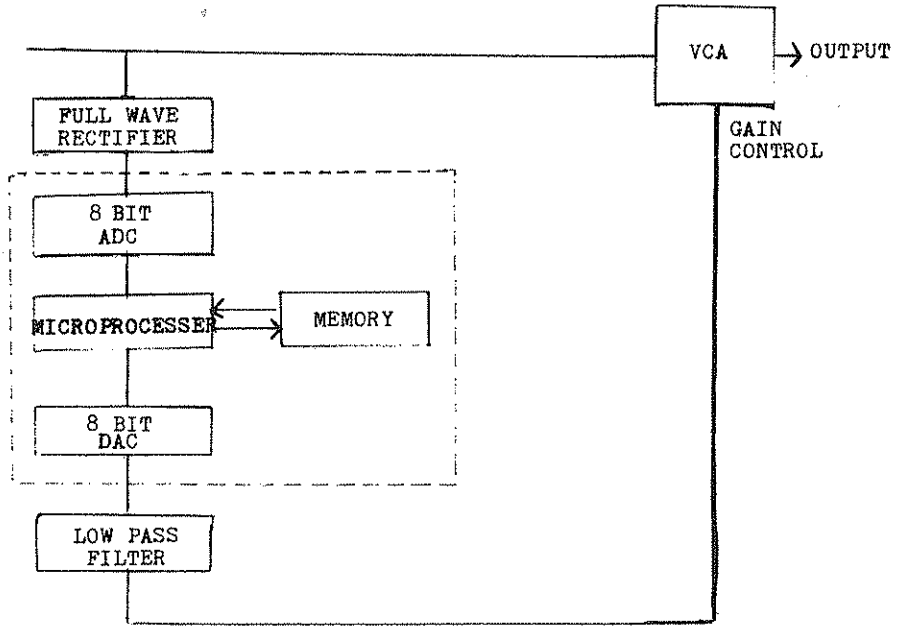


Figure 7. The microprocessor controlled compression amplifier.

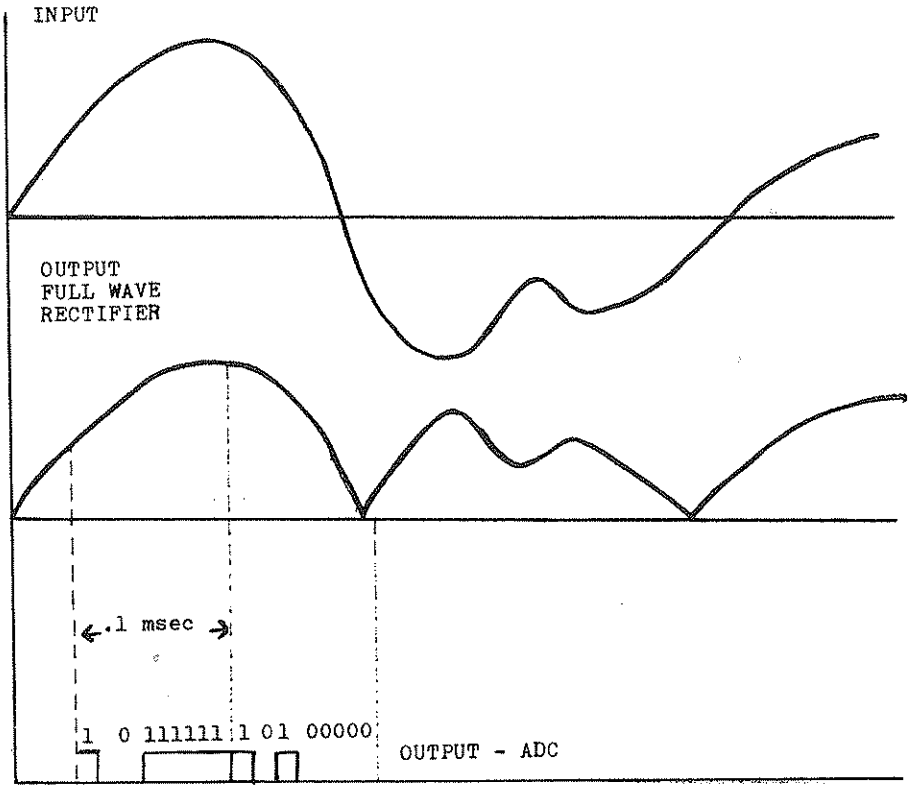


Figure 8. Showing the outputs of the full-wave rectifier and ADC for a typical input.

decrement algorithm is used to generate the release time constant (Fig. 10). What happens is that the value in the look-up table is compared with the look-up table value generated .1 msec ago which was decremented according to the required time constant (20 msec for low frequency band and 10 msec for the high frequency band). If the decremented old value is greater than the present value, the decremented old value is output through the ADC. On the other hand, if the decremented value is less than the present value, the present value is output. The decrement algorithm thus insures that when the input waveform decreases, the gain changes by no faster than a 40 dB/20 msec rate. A low pass filter

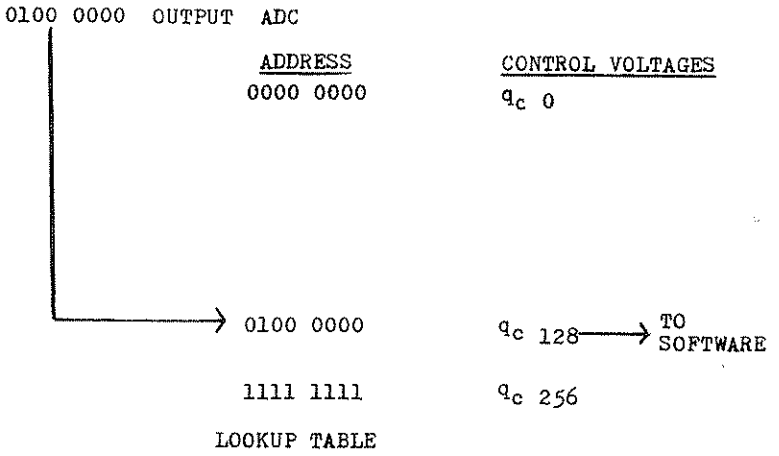


Figure 9. How the look-up table is used to generate the control voltages. The control voltages are in binary and are different for each individual.

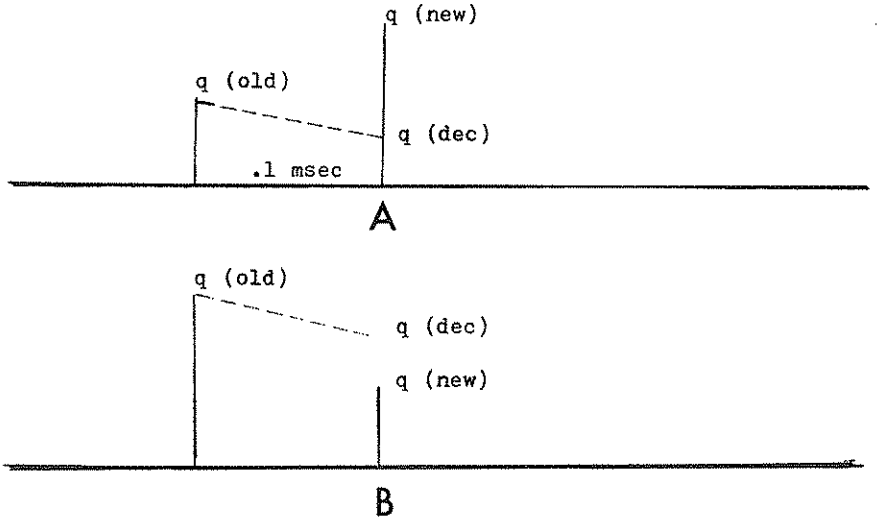


Figure 10. The generation of the release time.

is included to smooth the output of the DAC.

The system I have described to you is a hybrid system, that is - part digital and part analog. The digital computer is used to control the gain but the VCA is analog, that is, the speech waveform remains in its analog form and is not converted to a digital (or binary) format. The advantage of the hybrid system as a speech trainer over its analog version is in a multi-user environment where there may be one instructor for twenty or thirty students. In this situation, the one microprocessor can be "time shared" among the thirty students. You recall that we sampled the output of the full-wave rectifier every .1 msec. The micro is fast enough for the program to sample the output of the full wave rectifier, go to thirty look-up tables (a different one for each user), output the control voltage through the DAC and return for the next sample. Thus the hardware is shared among listeners as opposed to analog circuitry where each individual must have a duplicate unit. Listening tests among our students at Monmouth College indicate that the speech sounds as natural as that processed by analog units. There is one weakness to the microprocessor-based system. Because we have only 256 values in the look-up table, this limits the dynamic range to 20 log 256 = 48 dB. Thus the high band compressor has a 36 dB compression range, an additional range of 12 dB which is the expansion range, and below this, the slope is linear (k = 1). This has no practical effect since 48 dB below the maximum is far below the level of conversational speech under our assumptions.

It would be extremely advantageous if the hardware could be completely digital. What this would involve is that the speech would be converted to digital and then filtering, compression, and equalization be done by the computer. The processed speech would then be reconverted to analog form using the DAC. To do this we would have to sample the speech at least 15,000 times a sec (one sample every .06 msec) and use a 12 bit ADC because 8 bit ADC converted speech does not sound completely natural. In the hybrid scheme, we ADC converted the output of the envelope detector, not the speech itself. In addition to the sampling, the micro would also have to perform the filtering, equalization

and compression - this takes time. Micros are not fast enough at this time to perform these operations in the time required.

As to our future plans, we are in the process of designing a turnkey system which hopefully we will be able to demonstrate to you the next time this seminar is held.

REFERENCES

1. Dunn HK, White SD: Statistical measurements on conversational speech. J Acoust Soc Am 11:278-288, 1938
2. French NR, STeinberg JC: Factors governing the intelligibility of speech. J Acoust Soc Am 19: 90-119, 1949
3. Yanick P: Effects of signal processing on intelligibility of speech in noise for persons with sensorineural hearing loss. J Amer Aud Soc 1: No 5, 229-238, 1976
4. Yanick P, Drucker H: Speech processing to improve speech intelligibility in the presence of noise for persons with a ski-slope hearing impairment. IEEE Trans ASSP, Vol ASSP-24, No 6, 507-512, 1976