has been extensively tested for both speech analysis-synthesis and recognition applications over a wide range of recording conditions and has been found to provide satisfactory results.

The major limitation of the method is the necessity for training the algorithm on the specific set of measurements chosen, and for the particular recording conditions. For nonstationary speaking environments, it may be preferable to adapt the means and covariance matrices continuously.

REFERENCES

[1] B. Gold, "Note on buzz-hiss detection," *J. Acoust. Soc. Amer.*, vol. 36, pp. 1659-1661, 1964.

[2] A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Amer.*, vol. 41, pp. 293-309, Feb. 1967.

[3] M. M. Sondhi, "New methods of pitch extraction," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, pp. 262-266, June 1968.

[4] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, pt. 2, pp. 637-655, Aug. 1971.

[5] J. I. Makhoul and J. J. Wolf, "Linear prediction and the spectral analysis of speech," Bolt Beranek and Newman Inc., Cambridge, MA, BBN Rep. 2304, Aug. 1972.

[6] J. D. Markel, "The SIFT algorithm for fundamental frequency estimation," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 367-377, Dec. 1972.

[7] A. V. Oppenheim and R. W. Schafer, "Homomorphic analysis of speech," *IEEE Trans. Audio Electroacoust.*, vol. AU-16, pp. 221-226, June 1968.

[8] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *J. Acoust. Soc. Amer.*, vol. 55, pp. 1304-1312, June 1974.

[9] A. H. Gray, Jr., and J. D. Markel, "A spectral flatness measure for studying the autocorrelation method of linear prediction of speech analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 207-217, June 1974.

[10] E. A. Patrick, *Fundamentals of Pattern Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1972, pp. 193-194.

[11] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968, pp. 189-191.

[12] T. Marill and D. M. Green, "On the effectiveness of receptors in recognition systems," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 11-17, Jan. 1963.

[13] L. R. Rabiner and M. R. Sambur, "Some preliminary experiments in the recognition of connected digits," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 170-182, Apr. 1976.

Tone Detection for Automatic Control of Audio Tape Drives

JOHN J. DUBNOWSKI, JOSEPH C. FRENCH, AND LAWRENCE R. RABINER, FELLOW, IEEE

***Abstract*—This paper describes digital hardware for automatically stopping a cassette recorder upon detection of a prerecorded tone. This hardware is used in conjunction with experiments on computer assisted voice wiring experiments being performed at Western Electric locations [1]. For these experiments a sequence of instructions is automatically recorded on a cassette tape by a computer voice response system. At the end of each instruction, a tone is recorded. The hardware detects this tone and stops the cassette recorder. The operator, after performing the prescribed wiring instruction, manually restarts the cassette recorder for the next instruction. The technique used to detect the tone is a simple digital method comparing the axis crossings of the signal to a fixed threshold. This threshold is determined based on knowledge of the tone frequency, duration, and amplitude. When the signal axis crossings exceeds this threshold during two consecutive 40 ms nonoverlapping intervals the tone is detected and the tape recorder is stopped. The method described is a robust one which is rather insensitive to normal tape recorder problems, e.g., wow and loss of signal level due to battery drainage. The tone detection hardware requires nominal power and is portable.**

INTRODUCTION

ONE very promising application of computer voice response systems is in computer assisted wiring of electronic circuitry by voice. In essence, a complex printed list of wiring instructions is replaced by a spoken sequence of instructions on cassette tapes. During playback, all the information required to make the wiring connections is defined by a verbal instruction. The advantage of this procedure is that a craftsman is able to maintain continual visual contact with the wiring assembly. This avoids any disorientation usually caused by referring to printed wire-lists; thus, the operator's performance should be improved through increased efficiency and a reduction in errors. Furthermore, a training period is avoided in which the operator must learn how to read complex wiring diagrams.

A sequence of spoken instructions for wiring of electronic assemblies appears to have many advantages over the more conventional methods of wiring. To study the practical aspects of the feasibility of this method, the multiline computer voice response system [2] of the Acoustics Research Department has been used to generate a variety of spoken wiring lists for experimental evaluation at several Western Electric locations.

Along with the spoken wiring instructions, a tone cue is recorded to indicate the end of each instruction. By using special hardware to automatically detect this tone and subsequently halt the cassette drive, the operator is free to execute the latest wiring instruction without the added distriction of halting the drive.

Manuscript received September 23, 1975; revised December 23, 1975.
The authors are with Bell Laboratories, Murray Hill, NJ 07974.

Both the tones and speech are recorded on the same tape track, thus minimizing the number of tape cassettes required for a long wirelist. Also, this allows the operator to use commercially available two-track cassette recorders with a single output channel. Finally, when one comes to the end of one side of the tape, one can continue by turning the tape over.

Since the tone and speech are combined into a composite signal, the detector must be capable of distinguishing the tone from speech. There are several ways in which this could be done in hardware. For example, a phase-locked loop (PLL) is capable of isolating and tracking a tone of any frequency. However, problems with many cassette recorders makes such a technique impractical. Most inexpensive cassette recorders do not run at uniform speed; thus, the tone frequency can vary. Widening the bandwidth sensitivity of the PLL only increases the systems susceptibility to being falsely activated by the speech signal. The additional circuitry required to compensate for this would increase the systems reliance on component selection; thus, restricting duplication.

In the technique proposed here, the tone is detected based on a simple axis-crossing algorithm. Using knowledge of the frequency, amplitude, and duration of the tone, a threshold can be determined for the required number of axis crossings of the signal over a given time interval. By setting the threshold conservatively, and by requiring that the threshold be exceeded in two consecutive, nonoverlapping time intervals, a robust method for detecting the tone is obtained.

In this paper we describe the implementation of this technique in unipolar digital hardware. The advantages of using a digital realization is that the tone detection hardware is more easily duplicated than for comparable analog implementations. In the remainder of this paper we describe the details of the implementation.

Tone Detection Algorithm

The characteristics of the tone to be detected are as follows.

Frequency	2.8 kHz
Duration	120 ms
Amplitude	approximately 0 VU

The speech which is on the same audio track is generated by an adaptive differential pulse code modulation (ADPCM) decoder and is low-pass filtered at 3 kHz. Thus, the frequency of the tone is in the speech frequency range. This poses an additional requirement on the tone detection hardware, namely that it not detect speech which has significant energy around 2.8 kHz.

Based on the above considerations, a tone detection algorithm was designed which was based solely on a measurement of the number of zero crossings in a given time interval. The zero-crossing measurement is a fairly robust indicator of the tone, because of its regular zero-crossing pattern. Furthermore, since speech is a complex signal whose spectral energy is generally concentrated in the region of the first formant (i.e., below 1 kHz), the zero-crossing count for a speech segment will generally be far below the threshold for a 2.8 kHz tone. Finally, the possibility that unvoiced sounds will be detected is fairly low because of the 3 kHz filter which eliminates most of the unvoiced energy.

Thus the basic algorithm for tone detection is to measure the zero-crossing count every 40 ms, and to compare the result with a fixed threshold based on the expected number of zero crossings for a 2.8-kHz tone. The 40-ms interval was chosen to be 1/3 of the tone duration. Since the 40-ms interval is clocked asynchronously with the signal, it is guaranteed that at least two such 40-ms intervals overlap the tone interval. If the threshold is exceeded in two consecutive intervals, the tape recorder is turned off. This technique of requiring the threshold to be exceeded twice provides added protection against the tone detector being triggered by speech whose zero-crossing rate exceeds the threshold by random occurrence.

Fig. 1 shows the basic tone interval and the gating intervals used in the zero-crossing measurement. In the case when the gating pulses line up with the beginning of the tone interval, the tone overlaps exactly three gating intervals. In the worst cases (i.e., when the gating intervals begin 20 ms into the tone) there are two full intervals in which the gating signal overlaps the tone. Thus the tone detector should always exceed the zero-crossing threshold in two consecutive intervals during the tone.

Additionally, it is seen that the operation of the tone detector is essentially independent of the amplitude of the tone. Thus as the battery on the cassette recorder drains and the signal level decreases, the tone detector should still perform correctly.

Hardware Structure of the Tone Detector

Fig. 2 shows a block diagram of the hardware structure used to implement the tone detector. The input signal is first converted to a rectangular waveform by a switching amplifier where positive logic transitions correspond to axis crossings. A switching threshold prevents low level signals from causing the amplifier to switch. The output of the switching amplifier is fed into an axis-crossing counter (which is periodically cleared by the gating pulse). Whenever the count exceeds the tone cue threshold (which is manually set) a comparator generates an event signal. A double event sequence detector determines when two consecutive events have occurred. Whenever this occurs, it automatically stops the tape recorder through a relay switch. A manual reset button restarts the tape recorder by clearing the sequence detector. A gating interval generator, controlled by a free running clock, provides the count intervals over which the threshold comparisons are made.

The operating sequence begins after the momentary reset switch has been closed. The axis-crossing count is periodically cleared during short periods when X_2 is negative as shown in Fig. 1. This gating sequence generates uniform counting intervals during which the threshold comparisons are made. The double event sequence detector is clocked by both X_2 and the threshold event X_1. Only a sequence in which an event occurs both preceding and immediately following an X_2 clear pulse will cause the tape drive to be halted. All other sequences have no effect on the remote switch.

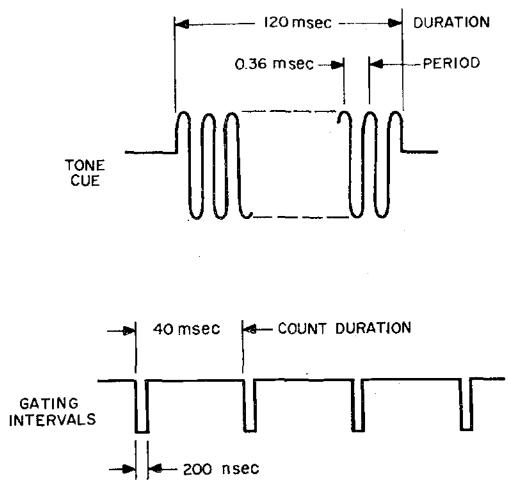

Fig. 1. Basic tone interval and the gating intervals used in the zero-crossing measurement.

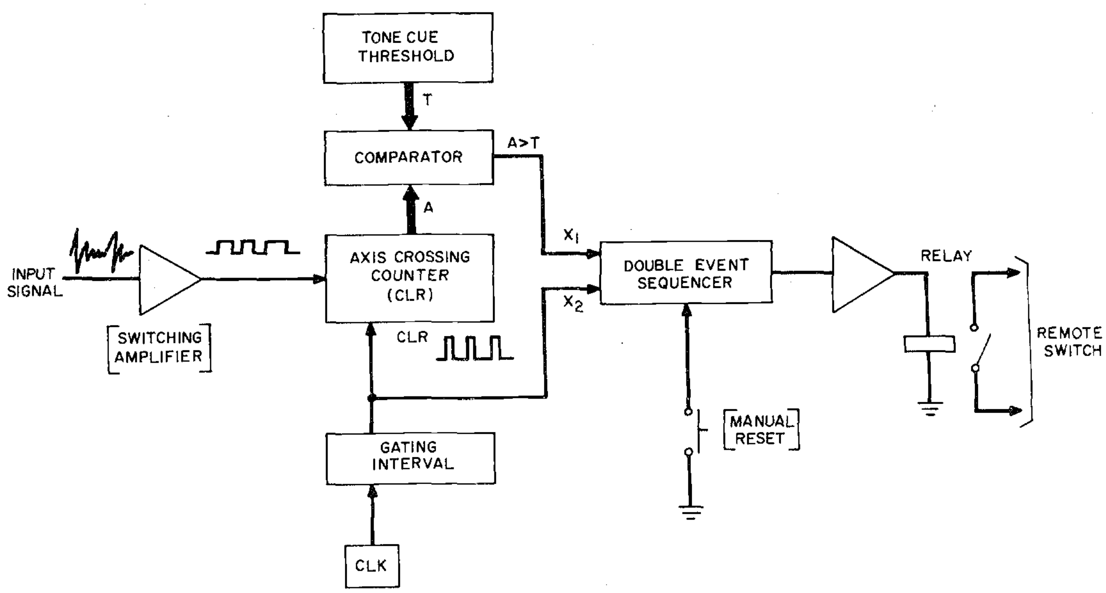

Fig. 2. Block diagram of the hardware structure used to implement the tone detector.

Logic and Timing

The timing for the system logic is shown in Fig. 3. An accumulated count of signal axis crossings is obtained every 38.4 ms (the basic timing cycle shown in Fig. 3). This assures that at least two such computations will be made during the random occurrance of the 120-ms tone. Based on a 6.2-kHz clock rate, the accumulation interval is generated by resetting a running counter with X_2 at the end of each 240 clock pulses. Signal B enables an output resulting from a comparison of the accumulated count and a fixed threshold, thereby generating a signal X_1, if the threshold is exceeded.

The logic implementation is shown in Fig. 4. A simple oscillator is constructed from two inverters [I.C. 6] feeding back through a simple RC filter. Control signals X_2 and B are generated by the gated counter [I.C. 7]. The input signal from the recorder is transformed into a rectangular pulse train through a transistor switch [I.C. 10] biased to avoid triggering on low-level fricative sounds. A 4-bit comparator [I.C. 4] continually compares the result of the axis-crossing counter [I.C. 3] against a threshold manually set by a switch block

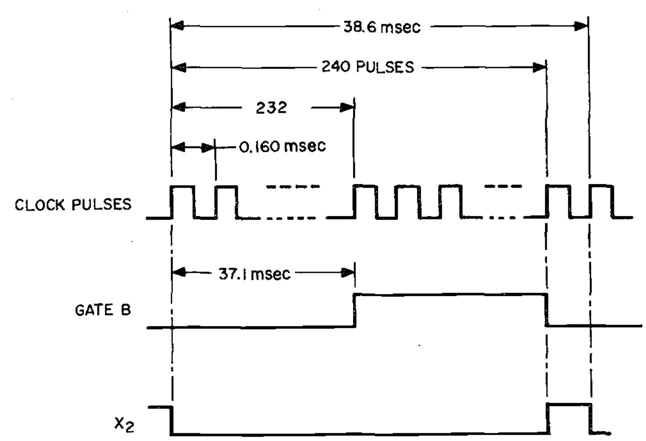

Fig. 3. Timing diagram for the tone detector.

[5]. This result is gated by signal B at [I.C. 2] to generate X_1 as previously described.

A double-event sequence detector having two memory elements (flip-flops) [I.C. 9] is triggered by both $\overline{X}_1$ and X_2.

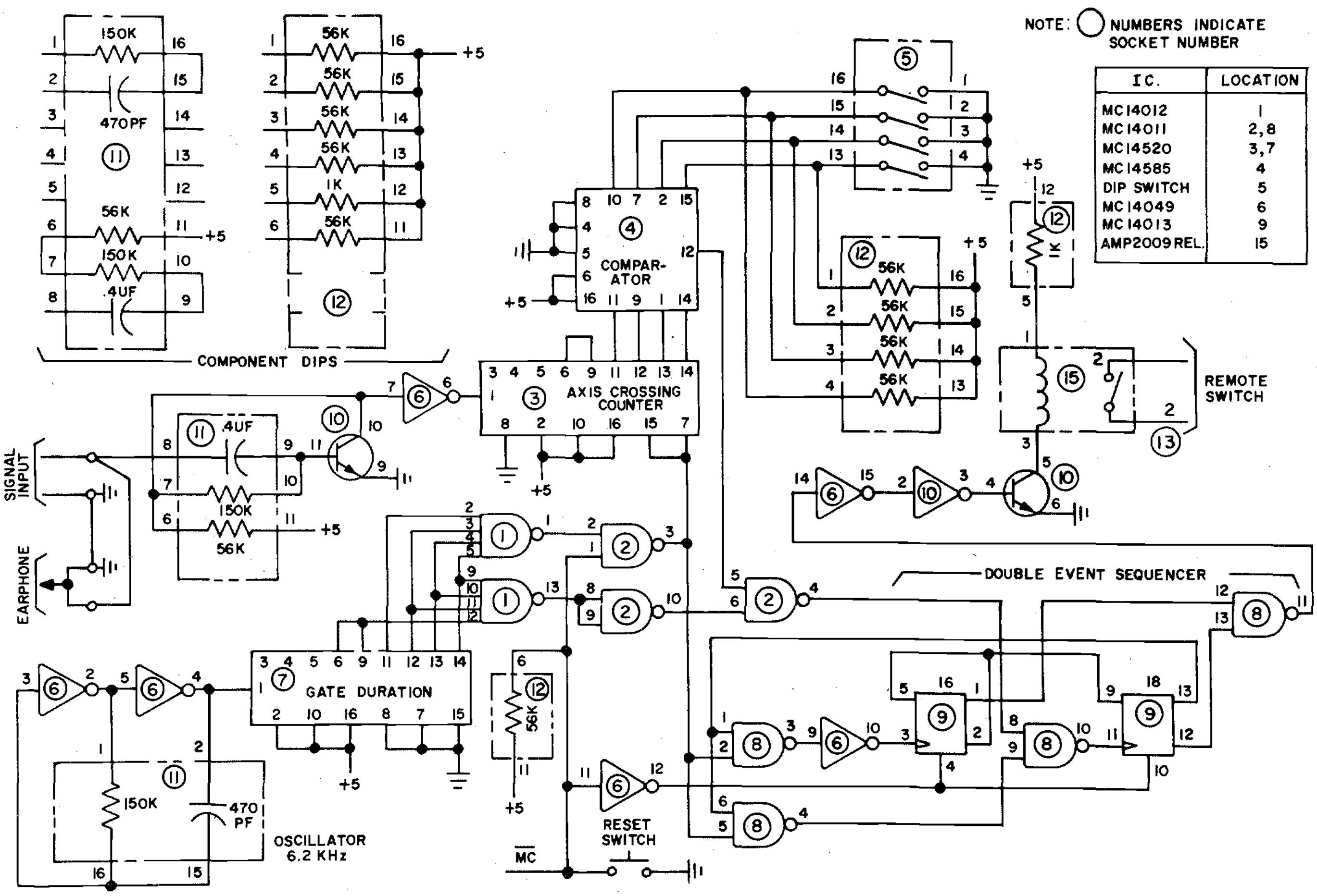

Fig. 4. Logic implementation of the tone detector.

The resulting detector output at NAND gate [I.C. 8] controls the remote control relay through driver [I.C. 10] to halt the tape drive. This only occurs when a sequence is detected where $\overline{X}_1$ is true both preceding and following a reset signal X_2. This indicates that the axis crossings have exceeded the threshold for two successive intervals. A manual reset switch restarts the tape drive by clearing both the detector and the axis-crossing counter.

Standard CMOS integrated circuits are used for the logic to minimize power consumption for a portable system. CMOS is also compatible with the dual bipolar drivers [I.C. 10] used to both detect the signal and to energize the remote relay.

Experimental Evaluation of the Tone Detector

The circuit of Fig. 4 was built and subjected to several tests to measure its performance as a function of tape recorder speed, operating voltage (battery drain), and tape recorder speed fluctuation (wow). It was found that the tone detector operated properly over a voltage range of 3.6 to 6 V. Below 3.6 V the tape recorder itself did not function properly—thus the tone detector worked over the full dynamic range of the tape recorder. In terms of the effects of tape recorder speed fluctuations, the use of a conservative tone cue threshold made possible proper operation of the tone detector for speed fluctuations of on the order of 20 percent. Finally, the power consumption of the circuitry was on the order of 350 mW.

After the construction of the circuit shown in Fig. 4, a second version of the tone detector was built with three modifications, including elimination of the relay used to remotely switch the tape recorder by replacing it with a transistor switch; utilization of the power supply of the tape recorder itself, thereby eliminating the battery pack and the on-off switch inside the tone detector; and simplification of the logic used in the circuit. These modifications reduced the number of hardware components, simplified construction and increased (somewhat) the performance of the tone detection unit. The final unit fit within a 3 × 5 in minibox and was powered entirely off the cassette recorder battery.

Summary

A portable digital hardware tone detector has been designed and implemented to automatically stop a cassette tape drive upon detection of a prescribed tone. The device uses a simple axis-crossing detection algorithm which has been implemented digitally, resulting in a reliable and stable system, with minimal power requirements.

References

[1] J. L. Flanagan, L. R. Rabiner, R. W. Schafer, and J. D. Denman, "Wiring telephone apparatus from computer-generated speech," *Bell Syst. Tech. J.*, vol. 51, pp. 391–397, Feb. 1972.

[2] L. H. Rosenthal, L. R. Rabiner, R. W. Schafer, P. Cummiskey, and J. L. Flanagan, "A multiline computer voice response system utilizing ADPCM coded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 339–352, Oct. 1974.