A VOICE-CONTROLLED ELECTRIC WHEELCHAIR

by

Cindy Furse

March 6, 1986

Electrical Engineering Department
University of Utah
Salt Lake City, Utah 84112

U/U-2-1

TABLE OF CONTENTS

ABSTRACT

Electric wheelchairs have greatly improved the self-mobility of many handicapped persons. The control systems for these chairs are still very limited, so a person without hand control is restricted to the use of physical mouth-held control devices. This paper describes a design for a voice-controlled system for directional control of an electric wheelchair. The design emphasizes use of integrated circuits, including the VRS1000 and SP1000 voice analysis chip set, with a minimum of external analog circuitry to produce a small, durable system. Although the system is still in the final design and testing stage, the results are encouraging that such a system for wheelchair control could be a functional reality.

I. INTRODUCTION

The 1970's was an era marked by an ever-increasing awareness to the needs of the handicapped. Programs to incorporate both mentally and physically handicapped persons into the general society have met with great success. Handicapped access laws have improved accessibility to buildings and facilities, and the design of an affordable electric wheelchair has improved the self-mobility of thousands who would previously have been completely dependent on the help of others.

The directional control systems for these electric wheelchairs are still somewhat limited. The most common method of control, the hand-held "joy" stick, is a very appropriate and effective method of directional control for people with the capacity for hand movement. However, many people confined to a wheelchair have little or no hand control and must therefore rely on one of two available mouth-held devices. The mouth-held "joy" stick is a large ball connected to directional control which is positioned by moving the head up and down and side to side. The Simpson Puff method employs a narrow plastic tube which is blown into to direct the chair. Both of these systems are uncomfortable for the user and somewhat demeaning, as they require a device actually held in the mouth.

The objective of building a voice-controlled wheelchair was to design a system for people with no hand control which would not require a device held in the mouth. A voice-controlled system needs only a small microphone mounted either on the wheelchair near the person's shoulder or on a lightweight headset similar to those worn by telephone operators.

II. BACKGROUND

Speech recognition systems can be categorized by three characteristics: speaker type, speech type, and vocabulary size. The simplest speaker type is a single speaker, or speaker-dependent system. This is much easier to implement than a multispeaker, or speaker-independent, system because of the great disparity in voice patterns from one individual to the next. The two varieties of speech types are isolated words and connected speech. Isolated words are commands, set apart by a space of silence which can be used to isolate an individual word. Connected speech is the way words are intertwined in normal conversation, marked only by sporadic intervals of silence. Algorithms to detect the end of a word in these systems are therefore very complicated. Increasing the vocabulary size of a system creates complications not only in required memory space, but also in increased word-search time and the possibility that two words may be phonetically similar and therefore difficult to distinguish. No standard definition of vocabulary size exists in the literature, but a small vocabulary would be on the order of 1-25 words, medium would be 25-100, and large vocabularies may accommodate up to 10,000 words.

The electronics industry has recently made great strides in the area of speech recognition. Improved VLSI design techniques have produced high-quality, low-cost lattice filters and microcomputer components capable of accurately analyzing voice with minimal addition of analog elements. This has brought the cost of a fundamental speech recognition system within the budget of an average consumer. Figure 1 shows the dramatic price reductions from 1980 to 1986.

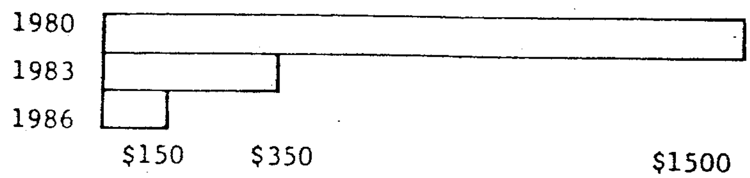

Fig. 1. Cost of a single speaker, isolated word small vocabulary voice recognition system from 1980 to 1986.

For a digital system such as a computer to analyze a variable analog signal such as voice requires a fairly extensive algorithm. First the signal must be filtered to eliminate power supply noise and limit the frequency range. The filtered signal is then sampled and digitized, converting it to data the computer can analyze. A single word is then isolated by an endpoint detection scheme. The system I used defines a word as a silence-to-rising, rising-to-plateau, and plateau-to-falling, as illustrated in Fig. 2. The endpoint is detected as a space of silence after a "word" is detected.

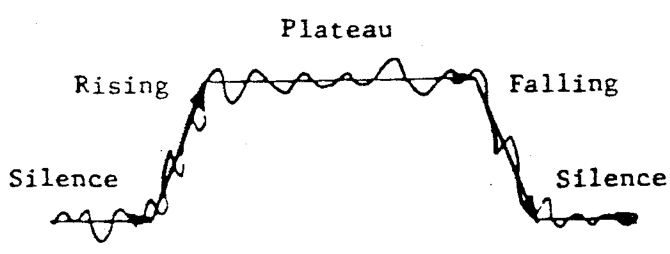

Fig. 2. Endpoint detection scheme.

Once a word has been isolated, its distinguishing features are extracted. Most discrete word systems employ linear predictive coefficient (LPC) analysis which defines the samples of the word as a linear approximation to the analog voice input. A less preferred method of feature extraction is dividing the word into phonemes, or elemental sounds, and defining a word as an ordered collection of phonemes. In the case of LPC analysis, the word is then time normalized to stretch or compress it so all words will appear of the same length, thus eliminating dependence on the speed a word is spoken. Time-weighted averaging is then performed to accentuate the relative length of sounds within the word. This is the final condition in which data describing a word is stored.

The final process in word recognition is dynamic time warping which maps the unknown word against word patterns, or templates, saved in the reference dictionary of the system. This algorithm looks for an optimal match, or if none of the reference templates is close enough, the system may reject the word as a nonmatch. Thus, a decision is produced. The complete word recognition flow diagram is illustrated in Fig. 3.

III. DESIGN

Six system specifications were defined for a voice-controlled electric wheelchair. First, it should respond only to the voice of the person in the wheelchair so that casual conversations nearby could not be mistaken for commands and cause the chair to respond unexpectedly.

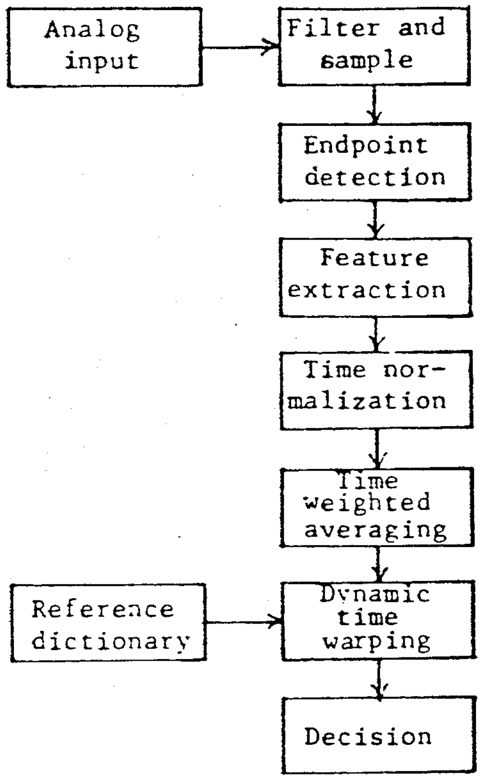

Fig. 3. Flow diagram of speech recognition algorithm.

Second, all directions would be given as commands, or isolated words. This is both a natural way of speaking to any device you wish to control and a preferred way in this case, as it would allow the person to carry on a normal conversation, perhaps even including words in the command vocabulary, without illiciting unexpected response from the chair. Third, only a small command vocabulary (go, stop, back, left, right) would be needed for complete directional control. These first three specifications define the simplest type of speech recognition system, a speaker-dependent, isolated-word, small-vocabulary system.

The microphone must be small and discrete so it will be comfortable to use and wear. It could be mounted either on the wheelchair near the person's shoulder or worn on a small, lightweight headset. The fifth specification is that the power supply must be self-contained so it can travel on the wheelchair. It will be derived from the battery used to power the wheelchair.

The final specification is that the directional response of the system must be both quick and accurate. Since the standard speed of a wheelchair is 5 mph, one-half second of delay from voice to wheelchair response would allow a chair going at top speed to travel nine inches before the chair would respond. This was chosen to be the maximum delay tolerance. Accuracy is of primary concern. It is essential that the chair respond correctly virtually 100 percent of the time. A safety mechanism will be implemented in future models to stop the chair in the event of an error.

The command vocabulary and its effect on the wheelchair are illustrated in Fig. 4. The commands are simple and phonetically dissimilar. "Go" produces forward motion, "back" produces reverse. "Left" and "right" may be used either when the chair is in motion to alter direction or when the chair is stopped to pivot it to face in a desired direction. One statement of the command would produce 30 degrees of angular change, and it could be repeated to turn in increments of 30 degrees. "Stop", of course, stops the wheelchair as quickly as physically feasible. The commands "activate" and "deactivate" were inserted to ensure that the system would not accidentally respond to commands spoken in normal conversation. "Deactivate" inhibits the mechanical system's ability to move the wheelchair. Although the voice recognition system will still recognize words, the wheelchair will not move until "activate" is used to return the system to normal operating mode. It is important to note that these words are not the only possible command vocabulary. The system should respond equally well to words from other languages or to any vocal utterance. Some handicapped persons have limited vocal skills, but if they can make a set of seven distinguishable sounds, they should be able to operate the system.

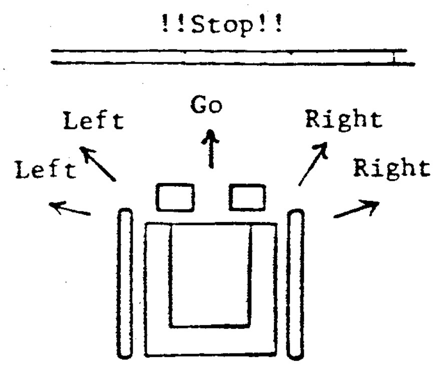

Fig. 4. Command vocabulary and chair response.

A simplified block diagram of the system is shown in Fig. 5. The analog input from the microphone goes through a digitizing system and into the speech recognition system where it is analyzed. The controller system converts the decision of the recognition system into data to control the wheelchair, and also initializes the recognition system with information supplied by the user through the initialization input.

The digitizer is made up of a bandpass filter, a variable gain control device, and an analog-to-digital converter. The bandpass filter is comprised of second-order low-pass and second-order high-pass

Butterworth filters. It cuts off at 100 Hz to eliminate 60 Hz power supply noise, and at 3.0 kHz to limit the frequency band of the signal. This is a legitimate filtering scheme for voice, as most of the voice information is stored in the band 100 Hz to 3.0 kHz.

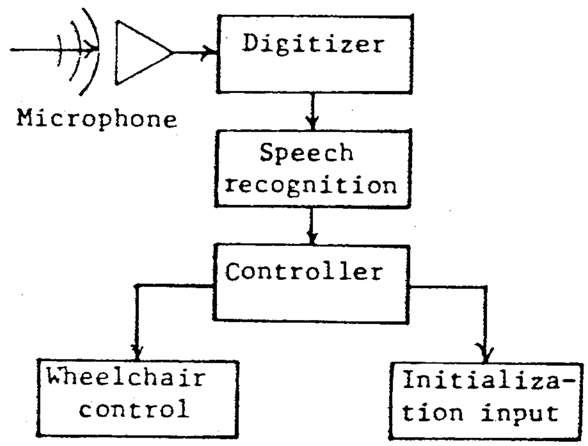

Fig. 5. Block diagram of complete system.

The LF13006 variable gain amplifier chip is used to control the gain of the analog signal to maintain maximum sensitivity to the low volume signals without saturating at high volume. This is controlled by the speech system. An ADC0831 analog-to-digital converter and CD4066 switch for sample-and-hold convert the analog signal to a digital signal. The information entering the speech system then contains one amplitude bit and eight data bits for every sample. Twelve samples are taken per word, requiring 108 bytes of storage for each word. The maximum allowable length of a word, as controlled by the recognition system, is two seconds, with a minimum of 200 milliseconds of silence between words.

The heart of the speech recognition system is the VRS1000 pre-programmed microprocessor manufactured by General Instrument Corporation. It includes four kilobytes of factory-programmed on-board ROM to perform both speech recognition and synthesis. Its sister chip is the SP1000, a specialized integrated circuit for speech recognition. It includes a set of programmable, reconfigurable lattice filters capable of being used for both speech synthesis and recognition. The VRS1000 microprocessor controls the SP1000 which, in turn, controls the variable gain amplifier and sample-and-hold circuit. The VRS1000 stores data retrieved from the SP1000 in a 64k static RAM. Data stored when the system is in "training" mode become templates, or patterns, in the reference dictionary. When the system is used in the "recognize" mode, data stored are mapped against these templates in search of an optimal match which will produce a word recognition decision.

The controller system is made up of a 6809 microprocessor, a 1k EPROM, and a series of six data latches for input and output. The latches are independently addressed by the 6809 and control the initialization input (three latches), wheelchair relays, recognition system input and recognition system output.

The initialization input is made up of a 12-key hexidecimal key pad for user input and two seven-segment displays and eight labelled LED's for system prompts to the user. When the system is first powered up, the user is sequentially prompted to initialize a set of reference templates of the command words. This initialization process requires the assistance of someone with hand control to input data through the key pad as prompted, but it needs to be initialized only once.

The templates made are stored in the recognition system RAM and are protected by a battery backup which is recharged nightly with the wheelchair battery. One option of the system is that if a word is found not to be recognized reliably, or if the speaker's vocal characteristics undergo physical change, the system can be retrained at any time.

The wheelchair directional servos are activated by eight relay switches which are controlled by the 6809 controller system. The chair comes from the factory with a hand-held "joy" stick, and this is the only part of the chair which needs to be modified. The electrical contacts which were originally physically connected are now electrically connected and controlled using the relays.

Schematic diagrams for the complete system are available in the appendix.

IV. RESULTS

The system compared very well with the desired specifications. The speech recognition system recognized only the person whose voice it was trained to. With three training passes, it was above 90 percent accurate. The VRS1000/SP1000 chip set is rated at 97 percent accuracy. The system recognizes words within less than one second (chips are rated for 45 milliseconds). The microphone is very lightweight (0.14 oz.). Still under development are the power supply and a complete working model. Consideration needs to be made to reduce the power requirements of the system by using CMOS chips where possible to reduce drain on the wheelchair battery.

The system is still in the testing stage. The recognition system needs to be tested in a wide range of noise conditions for a wide variety of speakers. To verify the directional control abilities, it was connected to a radio-controlled toy car to verify the feasibility of the system and improve the design and programming before connecting it to a wheelchair.

The cost of the system was roughly $200 when built using wirewrap connections. This includes the retail cost of all chips, sockets and miscellaneous parts. Ultimately much of the system could be further integrated into a single chip and produced on a printed circuit board for a substantially smaller cost, perhaps $50 to $75.

V. RECOMMENDATIONS FOR IMPROVEMENT

An essential addition to this system is a safety switch to stop the wheelchair in the event of a malfunction. This is complicated by the basic assumption that a person using this system will be unable to use a physical switch. A volume-sensitive switch might be a good solution to the problem. In case of trouble, the person would yell, a normal reaction, and the chair would stop.

User friendliness of the system can still be further improved. One option is to use the synthesis capability of the speech system to prompt the user. In the initialization routine, it might say, "Say go!," for instance, instead of turning on the "Say word" LED and printing the word number on the seven-segment displays as it does now.

Another possible improvement would be to install proportional control. This would allow such options as "slow," "medium," and "fast," and "one left" for 10 degrees left, "two left" for 20 degrees, etc. Most available wheelchairs are not designed for more than one speed, so the chair itself would need to be modified for this improvement. Another possibility would be a "park" mode where all commands would be followed by an automatic "stop" for maneuvering in tight places.

VI. OTHER APPLICATIONS

The applications for this device are tremendous. In the area of handicapped devices, word recognition systems are already available to input to computer terminals, and this can be further expanded to include automatic stenographer machines and a range of other devices. A voice recognition system with a radio-controlled transmitter-receiver pair mounted on the wheelchair could be used to turn on electrical items around the house such as fans, television, lights, and more. It could be adapted to be a phone-answering aid for the handicapped.

This voice recognition system could also be used in security systems, even accommodating more than one person if each voice were trained as an individual word.

Voice recognition systems can also be used as a "spare hand." They are being used by the Navy in some prototype jet fighters for control of such nonessentials as temperature and lighting controls. They might also be used to assist in assembly line construction or handling of dangerous substances.

The possibilities for this system and other voice recognition systems are limited only by the ingenuity and imagination. The applications are endless!

APPENDIX A

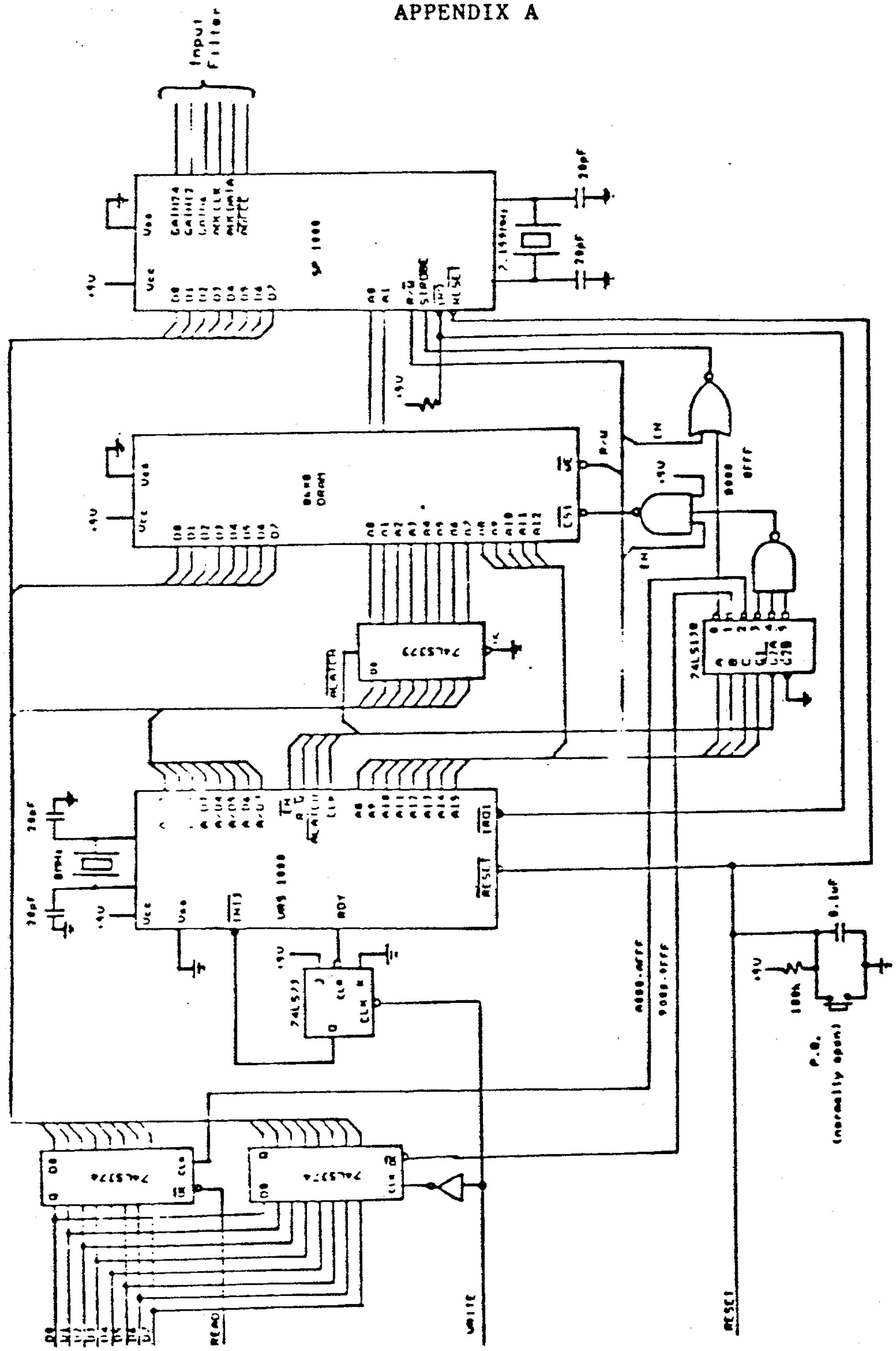

Fig. A1. Schematic diagram of speech recognition system.

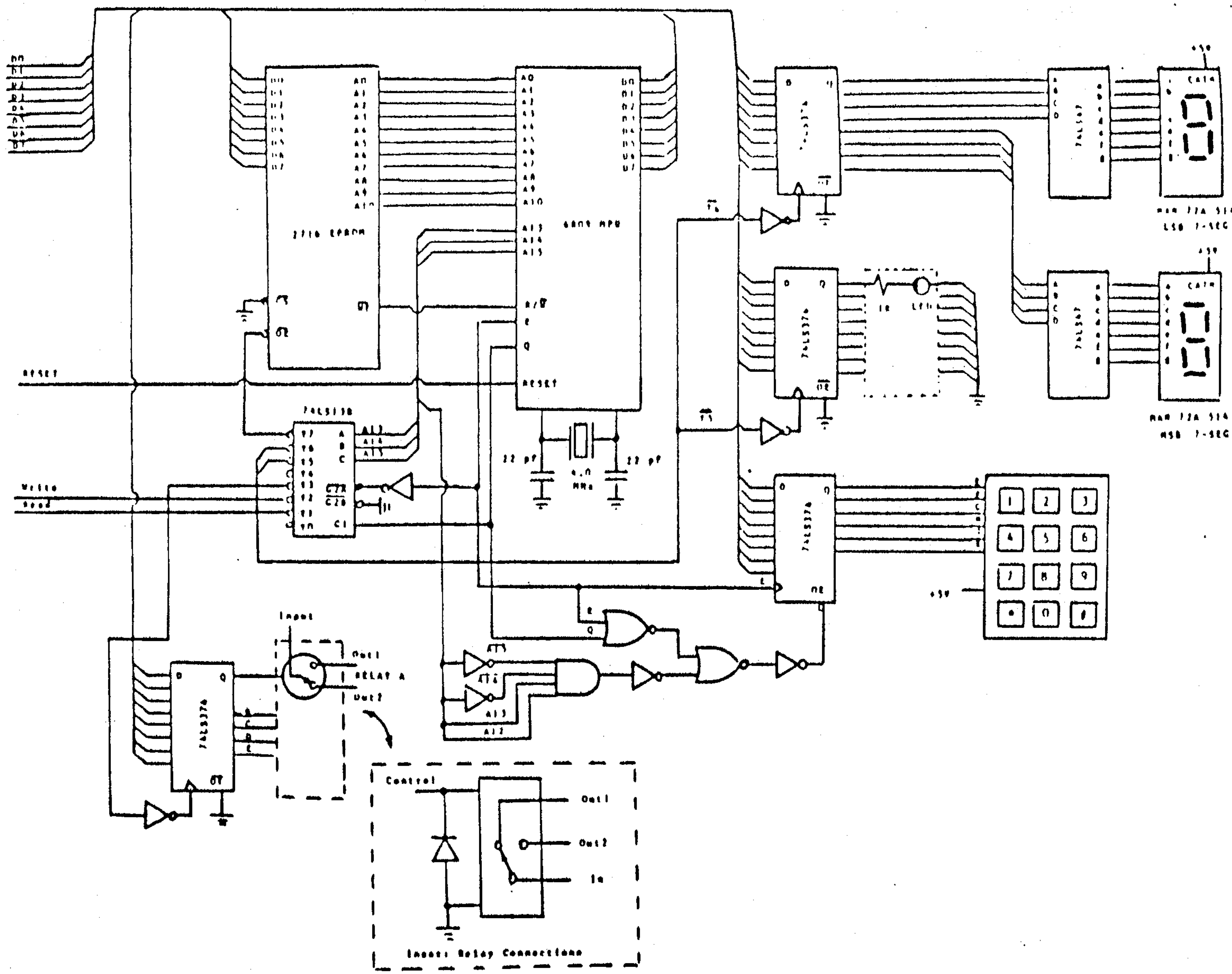

Fig. A2. Schematic diagram of 6809 controller system and other devices.

REFERENCES

1. C. L. Berney, "The Capabilities and Applications of Voice Recognition and Response Systems", Proc. Nat'l. Elec. Conf., 1980, U.34, pp. 530-533.

2. J. M. and J. K. Baker, "Digital Systems/Verbex Corporation Making Speech Recognition Practical Today and Tomorrow", Proc. Nat'l. Elec. Conf., 1980, pp. 534-539.

3. A. S. Poulton, Microcomputer Speech Synthesis and Recognition, Proc. Nat'l. Elec. Conf., 1980, pp. 534-539.

4. H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. Acoustics, Speech and Signal Processing, Vol. ASSP-20, No. 1, Feb. 1978, pp. 43-39.

ACKNOWLEDGEMENTS

I would like to acknowledge the support and contributions of the following companies and individuals:

Western Regional Marketing

General Instrument Corporation

Bell Industries

21st Century Scientific

Lindberg Corporation

Dr. James E. Dalley