Table 1: FREEDOM '93 Allophone Address Table (* = may be doubled)

PAUSES AND VOWELS			CONSONANTS					
Code	Allophone	Sample	Code	Allophone	Sample	Code	Allophone	Sample
11	PAUSE	10MS	35	B	BUSINESS	62	P	POW
12	PAUSE	30MS	36	B	BUSINESS			
13	PAUSE	50MS				63	R	REPAIR
14	PAUSE	100MS	37	CH	CHURCH	64	R	RURAL
15	PAUSE	200MS				65	R	BRAIN
			38	D	DO	66	R	ALARM
			41	D	COULD	67	R	CLEAR
						68	R	FIR
16	A	BEIGE	*42*	F	FOOD	71	R	FIR
17	AE	HAT						
18	AH	HOT	43	G	GOT	*72*	S	VEST
21	AW	AUGHT	44	G	GUEST	73	SH	SHIP
			45	G	DODGE			
22	E	SEE	46	G	WIG	74	T	TO
23	EH	END				75	T	PART
			47	H	HE	76	TH	THEY
24	I	SKY	48	H	HOE	77	TH	THEY
25	IH	SIT				*78*	TH	THIN
			51	K	COMB			
26	O	BEAU	52	K	CAN'T	81	V	VEST
27	OO	TO	53	K	SKY			
28	OO	FOOD				82	W	WOOL
31	OR	STORE	54	L	LAKE	83	WH	WHIG
32	OW	OUT	55	L	SADDLE			
33	OY	BOY				*84*	X	SUCCEED
			56	M	MILK			
34	UH	BOOK				85	Y	YES
			57	N	NO	86	Y	YES
			58	N	THIN			
			61	NG	ANCHOR	87	Z	ZOO
						88	Z	AZURE

a construct synthesis IC is again available, no practical improvements of the design will be possible.

Point of Contact

Address all correspondence to Professor James M. Conrad, Computer Systems Engineering Department, University of Arkansas, 313 Engineering Hall, Fayetteville, AR 72701-1201. Voice: (501) 575-6036, FAX: (501) 575-5339, email: jmc3@engr.engr.uark.edu. A more detailed description of this work can be obtained via anonymous ftp at `engr.engr.uark.edu`, directory `pub/cseg`, files `uarkcsegtr1993-2.*`.

References

[1] "Speech and Language Disorders", Fact Sheet, American Speech-Language-Hearing Association, Rockville, Maryland.

[2] Interview with Donna Edgmon, Speech Pathologist, 2700 American St., Springdale, AR, March 31, 1993.

[3] P.F. Adams, V. Benson, (1992). Current estimates from the National Health Interview Survey, 1991, Vital and Health Statistics, Series 10 (184), National Center for Health Statistics.

[4] G.C. Newman, A.R. Sparrow, F.E Hospod, "Two Augmentative Communication Systems for Speechless Disabled Patients", *The American Journal of Occupational Therapy*, August 1989, Volume 43, Number 8, page 529.

[5] D.L. Lee, F.H. Lochovsky, "Voice Response Systems", *Computing Surveys*, December 1983, Volume 15, Number 4, page 363.

[6] M.J. Walsh, "A Prototype Portable Electronic Speaking Aid for the Nonvocal Handicapped", *IEEE Engineering in Medicine and Biology Society*, October 1986, Volume BME-33, Number 10, page 986.

[7] D. Keating, D.C. Smith, A.L. Evans, D.J. Wyper, E. Cunningham, "The Pocket Speech Aid", *Medical and Biological Engineering and Computing*, May 1989, page 288, 289.

[8] R.J.H. Deliege, I.M.A.F.Speth-Lemmens, R.P. Waterham, "Development and preliminary evaluation of two speech communication aids", *Journal of Medical Engineering and Technology*, January/April 1989, Volume 13, Number 1/2, page 18, 21.

[9] M68HC11 Reference Manual, Motorola, Inc., 1988, pages 3-10.

[10] SP0256A-AL2 data sheet, General Instruments, 1984, page DS5005A-1.

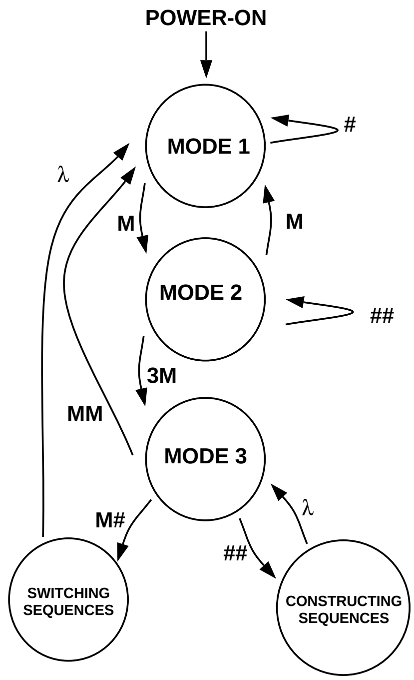

Figure 4: The Freedom '93 State Transition Diagram

further logic is involved. If, however, the M key is the second key, action is dependent on the first key. If the first key is any number other than 3, the assumption is that the user thought they were in MODE 1.

Therefore, the first key is ignored and the FREEDOM '93 switches to MODE 1.

3M is the gateway to MODE 3. Note that MODE 3 can only be accessed from MODE 2. Also, the only escape from MODE 3 is to MODE 1. This is accomplished by depressing the M key twice (MM) when MODE 3 is first entered or upon completion of either of the MODE 3 functions, "switching" or "construction".

"Switching" is the process by which the user can interchange the numerical assignments of an idea without having to reprogram each idea from scratch. Any two ideas may be interchanged, regardless of previous assignment (i.e. single-button or two-button).

"Construction" is the other function available in MODE 3, and is intended to be used for two purposes; programming any of the 72 single- or two-button ideas, or for impromptu conversation. Naturally, use for conversation is dependent on the user's familiarity with the allophone set and its use.

The speech data table in the FREEDOM '93 is limited to 2560 bytes in the NV RAM version and 4096 bytes in the EPROM version. This corresponds to an average of 45 or 73 allophones per idea for each of the 72 assignments. While speech data storage will probably never exceed these limits, the FREEDOM '93 does check to insure that any new assignment will fit into the available structure.

Allophone Usage

The Allophone Table supplied in Table 1 shows each of the 64 allophones for the FREEDOM '93. The allophones are listed pseudo-alphabetically in two tables, Pauses and Vowels, and Consonants. The FREEDOM '93 code is shown for each allophone, along with a sample word which demonstrates the positional sound for each.

Pauses are provided in five different lengths, 10ms through 200ms. The shorter pauses (10ms, 30ms) are normally used within the construct of a word to take the place of a hyphen, apostrophe, when syllables need special emphasis, or between syllables which have dramatically different sounds in order to prevent the FREEDOM '93 from slurring the two sounds. Examples include, "ninety-three", "o'clock", "about", and "freedom". Medium length pauses (50ms, 100ms) are useful to separate words, whereas the long 200ms pause is useful as a comma or as a period between sentences. Pauses may also be strung together to create custom lengths.

There are 10 allophones, denoted by asterisks, which may be doubled. These 10 allophones contain no tonal variations, which allow them to be used twice in succession in order to stretch the sound. More often than not, those allophones which may be doubled improve the construct, but final determination is, of course, left to the user.

The user should also take notice of the fact that many of the consonant allophones have different variations. Not only are the sounds associated with each variation distinguishable, but their positioning within the sample word is also key to their usage.

Conclusions and Future Research

The Freedom '93 speech aid design project has met the goals of the proposal. A prototype of a small programmable handheld device, with a voice output controlled by a keypad has been successfully built and tested. As an indication of both the interest and need for a device such as the Freedom '93, we have been informed that the Director of the Office for Student Access at the University of Arkansas wishes to make the prototype unit available for evaluation by vocally handicapped students.

The Freedom '93 speech aid design can be expanded in several ways:

- Utilize the RS-232 port of the EVBU to provide a PC based interface for Mode 3 programming.
- Move the Freedom '93 code into the EPROM of a MC68HC711E9, providing additional space for the storage of word and phrase data.
- Remap the internal RAM of the HC11 (or HC711) and add additional decoding to allow for the increase of data memory, perhaps replacing the 2K X 8 NVRAMS with 8K X 8 devices.
- Add circuitry to detect a low battery condition.
- Add circuitry to provide software control of the volume, eliminating the mechanical adjustment.
- Utilize the HC11 STOP instruction, which halts the oscillator and MCU clocks, and provide start up of the unit by keypad presses, thereby eliminating the external power switch.
- Replace the 9 volt alkaline battery with a lithium equivalent, for longer operation.
- Replace the 9 volt battery with 6 nickel-cadmium cells, which are rechargeable and available in capacities of up to 4400 mAH. The use of 6 cells will reduce the power loss in the 5 volt regulator, resulting in better efficiency.

Lastly, and not insignificantly, additional research should be made concerning the SPO256A-AL2 speech processor. Project research indicated that this IC is no longer in production and further, no constructive synthesis speech device is apparently being made. The Freedom '93 speech aid is a flexible and versatile design. However, until such time as battery as the data retention power source. Because of this switch process, there is a required 2 to 125 ms delay before the NV SRAM can be safely accessed. Because the program code of the prototype unit resides in NV SRAM, the delay code is placed in the MCU's internal EEPROM.

The HC11 on the EVBU is forced, via the BUFFALO ROM vector to execute the 125 ms EEPROM code, which is followed by a jump instruction to the beginnning of the program code in NV SRAM.

Control of the Freedom '93 is achieved by keypad input. The nine keys are labeled "1" to "8" and "M." To simplify the program code for single port reads, a 10 to 4 line encoder (74LS147) is used.

To be portable, the speech aid design requires a battery as the power source. The power supply circuit consists of a single 9 volt transistor battery, driving a 5 volt regulator. The current drain of the prototype unit is approximately 120 mA, resulting in 30 minutes of operation from an alkaline 9 volt battery. Suggestions to extend the operating time of the Freedom '93 are listed in Section .

Software

The software for the FREEDOM '93 is completely written in M68HC11 assembly language for ease of bit manipulation and hardware interfacing. There are currently two versions of the FREEDOM '93 code: one for the NV RAM version and one for the EPROM version. Both programs use a 512 byte scratch pad for program variables and flags as well as queues and buffers. This scratch pad memory is completely allocated; there are no 'spares'. The speech data table for the NV RAM version is 2560 bytes and the EPROM version is 4096 bytes. The addresses for these memory sections, as well as the actual program storage, are:

	NV RAM	EEPROM
SCRATCH PAD	\$C000 - \$C1FF	\$0000 - \$01FF
PROGRAM	\$C200 - \$C5AD	\$D000 - \$D370
SPEECH DATA	\$C600 - \$CFFF	\$C000 - \$CFFF

The two versions operate in same manner with a few exceptions which will be pointed out below.

The FREEDOM '93 must monitor two input devices during operation; the keypad and the feedback from the speech processor (which informs the program when its input buffer is clear to accept a new allophone.) Since keypad operations are a slow function in relation to real-time, inputs are polled at each point in the program where a key is expected. The speech processor, on the other hand, is totally independent of actual program operation. Once a message is queued for transmission to the speech processor, the message may only be sent one allophone at a time, and only when the speech processor has completed processing the preceeding allophone. Also, the allophone string to the speech processor needs to be fed in a steady manner. One of the properties of the speech processor is that it will continue to emit its last allophone until instructed otherwise, via a new allophone or a pause. Due to these considerations, the "ready" signal from the speech processor is configured to cause a maskable interrupt in the MC68HC11.

A speech subroutine does the actual queueing of the allophone buffers to be spoken by the speech processor. The maximum capacity of the queue is 20 bytes, or 10 addresses. The queue is circular and the design assumption is that this will never overflow. If overflow did occur, it will merely result in some of the speech being lost. No harm to the system can occur.

Actual completion of a message requires that a final pause be sent to the speech processor to 'end' the previous allophone. The 50 ms pause is used as this has been determined to be the usual pause between words and allows the user to string messages together, achieving an even flow of speech. A software block diagram is shown in Figure 3.

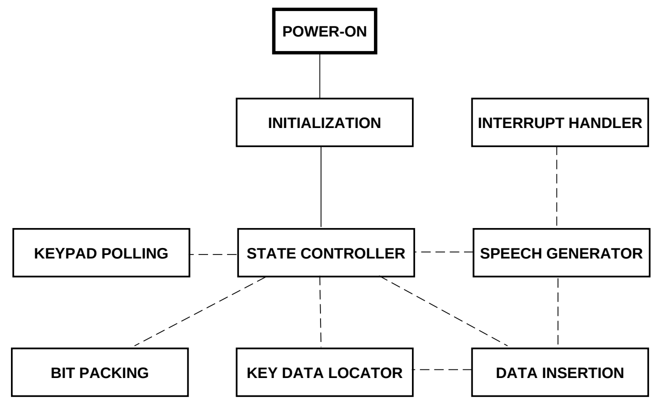

Figure 3: The Freedom '93 Software Block Diagram

Operation

The FREEDOM '93 has three modes of operation:

1. - SINGLE-BUTTON MODE
2. - TWO-BUTTON MODE
3. - EDIT MODE

These three modes are herein referred to as MODE 1, MODE 2, and MODE 3, respectively. The tri-color LED mounted on the panel indicates the current operating mode to the user: GREEN, RED, and YELLOW, respectively. The transition between modes (Figure 4) depends on keypad input of a number (#), the MODE key (M), or any keystroke (λ).

The FREEDOM '93 powers-up in MODE 1. MODE 1 enables the user to access any of 8 preprogrammed words, phrases, sentences, etc., by depressing any one of the eight number keys on the keypad. Throughout this report, these words, phrases, sentences, etc. will be referred to as "ideas".

Accessing an idea causes the FREEDOM '93 to speak that idea. Ideas may be accessed in a string to form larger ideas just as the words on this page are printed in a string to form larger ideas, i.e. sentences, paragraphs, etc. The FREEDOM '93 employs a queueing technique that enables the user to string together up to 10 ideas (including the one being currently being spoken) in rapid succession.

MODE 2 enables the user to access an additional 64 preprogrammed ideas by depressing any two-digit combination on the keypad. The stringing of ideas in MODE 2 can also be accomplished.

Further, ideas from both modes may be strung in an intermixed fashion simply by switching modes during the process.

Mode switching is accomplished by using the M key on the keypad. When operating in MODE 1, depressing the M key will cause the FREEDOM '93 to switch to MODE 2. Similarly, while operating in MODE 2, the FREEDOM '93 will return to MODE 1 by depressing the M key. If the M key is the first key in an expected two-digit combination, no

- Canon Communicator & Vortax Type 'n Talk [6]
- Pocket Speech Aid [7]
- Pocketstem and Tiepstem [8]
- Sharp El-7100, Taalzakboak, and Bliss-board [8]

These systems offer various combinations of features, including portability, control (keyboard vs. keypad), programmability, and technique. The programmable types all make use of a full keyboard approach, making them non-portable.

Since all speech aids need to be reprogrammed in some manner, those referred to as non-programmable are dependent on other devices for updates. They are connected by cables to a compatible host or PC. They are portable, but not very flexible as they require knowledge of the words, phrases, and sentences required for all foreseeable encounters until reprogramming is possible.

Two portable commercial speech aids manufactured in the United States are the Vocaid, produced by Texas Instruments, and the DynaVox by Sentient Systems Technology, Inc. The Vocaid produces a voice output by utilizing a text-to-speech process. While the Vocaid is inexpensive (approximately $50.00), it is not user programmable. Thus it is not designed to be custom adjusted to its user. The DynaVox is a more sophisticated device. It utilizes the DECtalk synthesized speech technology, an advanced text-to-speech system produced by Digital Equipment Corporation. While using what is arguably the highest quality speech synthesis available, the DynaVox comes at a high price - $4,305 for a minimal system.

Freedom '93 Design

The Freedom '93 addresses the shortcomings of the current market by providing a speech communication aid that is both portable and easily user programmable by employing constructive synthesis methods. The prototype can be constructed at an approximate cost of $100.00.

Hardware

As shown in Figure 2, the Freedom '93 speech aid is composed of five sections: system controller, speech generator, word/phrase memory, switch inputs, and system supply.

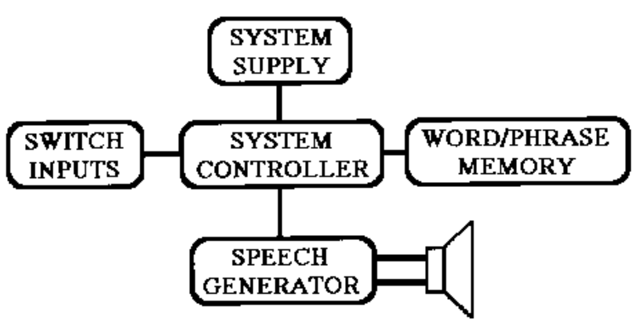

Figure 2: Functional Block Diagram of the Freedom '93 Speech Aid

The system controller is a Motorola M68HC11EVBU universal evaluation board (EVBU). These boards include an 8-bit, M68HC11E9 microcontroller unit (MCU), RS-232 terminal interface, and wire wrap access to all of the MCU connections. The MCU operates with a bus speed of 2 Mhz. The M68HC11E9 (HC11) has 12K of ROM (containing the BUFFALO monitor program), and 512 bytes of both RAM and EEPROM. The development of the Freedom '93 code is done with the HC11, using a PC based program from Motorola, PCbug11. Communication between the EVBU and the PC is established by means of a "talker" program, stored in the HC11's EEPROM, and a serial cable connected between the two units.

The HC11 on the EVBU can be operated in either a single chip or expanded mode. The latter allows external memory and peripheral devices to be accessed by a time-multiplexed address/data bus [9]. The Freedom '93 speech aid is operated in that mode to access the word/phrase memory. However, the expanded mode of operation changes the function of two 8-bit, bidirectional ports on the HC11 to make the address/data lines externally available. Because insufficient port lines remained to directly control the speech generator, a Rockwell R6522 Versatile Interface Adapter (VIA) is added to the memory map. The VIA also controls the output of the two MODE LEDs (contained in a single housing) located on the front panel.

Vocal output is generated by a General Instruments SPO256A-AL2 speech processor. The four basic functions incorporated by this device are:

1. A software programmable digital filter that can be made to model a VOCAL TRACT.
2. A 16K ROM which stores the allophones.
3. A MICROCONTROLLER which controls the data flow from the ROM to the digital filter, the assembly of the "word strings" necessary for linking speech elements together, and the amplitude and pitch information to excite the digital filter.
4. A PULSE WIDTH MODULATOR that creates a digital output which is converted to an analog signal when filtered by an external low pass filter [10].

The speech processor has eight lines (A1-A8) that are used as addresses to access the desired allophone in the internal 16K ROM. Only A1-A6 are used in this application (64 allophones), and the other lines are tied to ground. An address load line (ALD) is then used to initiate the speech sequence. An interface signal (SBY: Stand By) is also provided, to inform the microprocessor when an allophone output has been completed. In the Freedom '93, the SBY signal is used to generate a maskable interrupt on the HC11.

By driving the VIA's CA1 line - configured as a handshake input - with the SBY signal, and driving the HC11's IRQ line with the VIA's IRQ output, the correct interface signal is generated. All speech sequences must end with one of the pause allophones to silence the output.

To store the allophone data used by the speech processor to produce the desired word or phrases, two Dallas Semiconductor 2K X 8 nonvolatile static RAMs (NV SRAM) are used. The NV SRAM devices were chosen for three reasons. First, SRAM requires no special timing delays for proper write operations, as opposed to EEPROM.

Second, the NV SRAMs incorporate an internal lithium battery to maintain the memory contents. The specified minimum data retention time (t_{DR}) by the manufacturer is 10 years - well within the useful life of the Freedom '93 speech aid. Third, if the NV SRAM needs to be replaced with EEPROM at a later time, the device is pin compatable with the 2816 EEPROM, and only the addition of a delay routine to the program code would be required.

The Dallas Semiconductor NV SRAMs contain circuitry that monitors the rising voltage level of the Vcc supply pin during start up, determining when to switch off the internal lithium

FREEDOM '93: A PORTABLE SPEECH DEVICE

Kenneth L. Rusnok, Martin S. Lavine, James M. Conrad
University of Arkansas

Keywords: Constructive Speech Synthesis, Vocal Disability, Microcontroller, MC68HC11.

Abstract

The ability to communicate through speech is an important part of the structure of human society. For many people however, normal speech is not possible because they are vocally disabled. Thus, they are limited in being able to communicate their thoughts. The result is often a sense of isolation and frustration. To provide an effective form of communication for people so disabled, an easy-to-use device is needed. The device requirements include: hand-held size, operate from battery power, be adjustable to the user, and utilize a simple control method. The Freedom '93 speech aid generates artifical speech in a small programmable hand-held device, with the voice output controlled by a keypad.

Introduction

In the United States, an estimated 2.5 to 10 million people have a speech and/or language disorder [1]. The causes include neurological disorders from strokes, and diseases such as ALS (Lou Gehrig's disease) or cerebal palsy [2]. Another estimate, by the National Center for Health Statisitcs, reports the speech impaired population in the United States to be approximately 2.8 million, 39 percent of whom are children under 18 years of age [3]. Regardless of the cause of the speech disorder, if the ability to think and reason has not been affected, technology can often be used to provide a communication substitute. That technology is commonly referred to as augmetative communication [4].

Among the possible forms of augmentative communication for the vocally handicapped are writing tablets or sign language. Both methods, while simple, have several limitations. The use of a tablet is slow and further, depends on legible handwriting. In an emergency situation, there might not be time to write out a message, nor may legible handwriting be possible. The use of sign language as a communication method is dependant upon its widespread proficiency by the people with whom a vocally disabled person interacts. To achieve a versitile communication system for the vocally disabled, an electronic speech aid, capable of custom utterances, solves the limitations just addressed.

The Freedom '93 is a microcontroller based device utilizing construction synthesis to create a voice output and a keypad to determine the specific output. Figure 1 is a photograph of the prototype unit.

Proceedings of the 1994 ACM Symposium on Applied Computing, March 6-8, Phoenix, AZ.

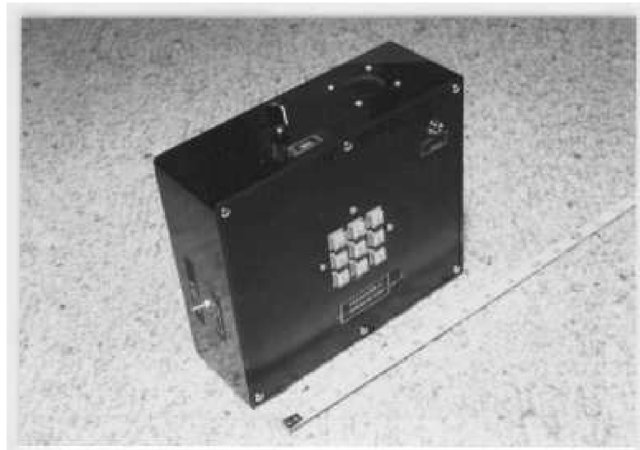

Figure 1: The Freedom '93 Speech Aid

Constructive systhesis is based on the use of phonemes and allophones. A phoneme is "the smallest segment of sound that distinguishes words of different meaning in a language [5]." Every language has its own set of phonemes, each slightly different. Further, phonemes are slightly different depending on their position within the word being spoken. An allophone is a positional variant of a phoneme. By combining allophones in the correct sequential order, virtually any word or phrase can be produced.

The Freedom '93 uses an integrated circuit capable of outputting 64 unique allophones (59 sound and 5 variable length pauses). The microcontroller monitors keypad input, provides the allophone data to the speech IC, and controls storage of the allophone data into system memory. The keypad controls all operations, including speech output for single and double button presses and allophone programming of speech to key assignment.

State of Current Technology

The three main methods of speech synthesis are waveform coding, synthesis by analysis, and constructive synthesis [7]. Waveform coding results in compaction of data by digitizing the original speech. An example of synthesis by analysis, linear predictive coding (LPC), digitizes a human voice, and reconstructs the original signal using stored vocal variables, such as pitch. As mentioned previously, constructive synthesis is achieved by concatenating allophones in the correct sequence to produce the desired speech sound.

Research into available speech aids yielded reports of work done in Europe and the United States. In the European marketplace, a number of speech communication aid products have already been introduced: