SESSION XII

SPEECH PROCESSING

Dominic Massaro, *Presider*
University of California, Santa Cruz

Voice reaction times with recognition for Commodore computers

DAVID A. WASHBURN and R. THOMPSON PUTNEY
Georgia State University, Atlanta, Georgia

Hardware and software modifications are presented that allow for collection and recognition by a Commodore computer of spoken responses. Responses are timed with millisecond accuracy and automatically analyzed and scored. Accuracy data for this device from several experiments are presented. Potential applications and suggestions for improving recognition accuracy are also discussed.

The Commodore 64 (C-64) and 128 have been popular computers with numerous strengths—documented in this journal (Hormann & Allen, 1987; Kallman, 1986; Pittenger & Hodge, 1986) and elsewhere (Hopkins & Morris, 1989; Rumbaugh, Hopkins, Washburn, & Savage-Rumbaugh, 1989)—as laboratory instruments. With the steady decline in cost of more powerful computer systems, many of the advantages offered by the inexpensive Commodore systems may seem to have diminished. However, the availability and relatively low cost of these computers and of their associated software and peripherals remain tenable reasons for selecting Commodore systems as dedicated data-collection and control units. A modification of one such peripheral, the inexpensive Voice Master (Covox, Inc.), to collect accurate voice response times (RT) is reported in the present paper.

Of course, voice-activated relays and voice keys have been used in psychological research since Donders measured vocal RT in 1868. The “smart” part of the present device is that, in addition to recording RT, the computer can analyze and “recognize” the spoken response and score the response as correct or incorrect automatically. Speech recognition by a computer remains a relatively young, imprecise, and expensive enterprise. Digitized

speech recognition has lagged behind digitized speech production in quality, reliability, and affordability. The Voice Master is an inexpensive speech and music processor that will: (1) accept and recognize spoken speech; (2) reproduce digitized speech; and (3) record and analyze musical input (humming or whistling). Although each of these would seem to apply to psychological research, only the first feature is our concern here.

The Voice Master plugs into one of the joystick ports on the Commodore 64. Software that accompanies the device enables a user to “train” the computer to recognize words—in fact, any sounds. The system will accept a vocabulary of up to 64 words, and additional vocabularies can be saved onto disk and utilized at any time. The machine-language wedge that drives the Voice Master also expands the native BASIC in the C-64 to include several new commands. The TRAIN command is used to generate a database of words that the computer can recognize. RECOG puts the C-64 into recognition mode, where it accepts speech and compares the digitized patterns against those which were trained. The computer can also reproduce the speech with the SPEAK command.

Hardware Modification

The Voice Master itself is fairly well documented and straightforward to use. It cannot be used as a voice key, however, without modification. The unit must be wired to alert the C-64 as soon as speech input is received, so that RT can be assayed. Only then should program control be switched to the speech-recognition routines.

The Voice Master is equipped with a light-emitting diode (LED), which becomes illuminated when a particular level of sound is received through the microphone.

This research was partially supported by Grant NAG2-438 from the National Aeronautics and Space Administration. Additional support was provided by a grant from the National Institutes of Health (NICHD 06016) and by the College of Arts and Sciences of Georgia State University. The authors gratefully acknowledge the invaluable contributions of Ray C. Taylor to the development of this device. Reprint requests should be sent to David A. Washburn, Department of Psychology, Georgia State University, Atlanta, GA 30303.

This level is adjustable through calibration of the Voice Master and adjustment of the Gain potentiometer. Thus, the most obvious operational definition of a vocal response is the production of a sound at a volume sufficient to illuminate the LED.

To modify the Voice Master, one must first remove the four screws in the base of the Voice Master and open the cartridge to expose the circuit board. One then locates the two resistors (R23 and R24) near the LED. One of these resistors (R23) will provide the signal (+5 V) for the voice key. Next, one carefully solders a wire to the high post of R23. Extreme caution must be exercised when soldering on the circuit board.

The ground post of the gain potentiometer (the post nearest the microphone jack) is used for the ground. One must carefully solder one end of the second conductor of the wire to this post. The holes in the Voice Master cartridge may require enlarging to allow the line to exit the Voice Master cartridge near the microphone jack.

Although the Voice Master plugs into joystick port 2 on the C-64, it does not require the use of all of the pins on this port. Specifically, pin 1 (JOYBO) is available and is a perfect input for the voice key line. One solders the line from R23 to pin 1 of the D-connector. The line from the gain potentiometer, ground, will need to share pin 8 (JOYB7) of joystick port 2 (the easiest way to do this is to make a Y-adaptor).

Software Modifications

After these modifications have been made, bit 1 of DC00 (56320) will toggle low whenever the Voice Master receives sound input. This bit will be the signal to stop

the RT counter. The modified Voice Master can then be integrated into the software requirements of any particular experimental paradigm. For millisecond accuracy, a machine-language-based timer such as the one described by Hormann and Allen (1987) can be used. This particular program must be modified according to instructions provided by Hormann and Allen for use with a joystick. In addition, the program must be modified for use with the modified Voice Master. The code given in Listing 1 should be substituted for the corresponding lines in the timing program (see Hormann and Allen, 1987, p. 40). Note that the timing subroutine should return immediately back to a RECOG statement in the BASIC program if speech recognition is to occur.

Reliability

The modified Voice Master was used in several experiments in our lab to determine the reliability of the speech-recognition unit. In each of these experiments, audio tapes were made of subjects' responses. The tapes were used to score and to evaluate the accuracy of the Voice Master in recognizing responses. In each experiment, subjects were tested in a soundproofed room, using the headset microphone that accompanied the Voice Master. A machine-language millisecond timer, developed at our laboratory and similar to the one described by Hormann and Allen (1987), was used in the timing of all responses.

In two experiments, subjects were required to respond as quickly as possible by saying a number (0-9). In one of these experiments, 18 subjects were given 24 trials in an event-frequency estimation task. In the second experiment, 20 subjects were given 80 trials of a dot-

Listing 1
Code for Modifying Timing Program

```

10 REM ***** JOYSTICK MODIFICATIONS *****
:
30 IF T < > 13990 THEN PRINT "ENTRY ERROR: CHECK YOUR DATA VALUES: STOP
:
:
1007 DATA 17,208,162,17,142,14,221,173
1008 DATA 0,220,133,252,201,126,208,247
1009 DATA 234,234,234,234,169,0,141,14
1010 DATA 221,234,173,6,221,73,225
:
:
***** MACHINE LANGUAGE *****
037B LDA $DC00      ; Check joystick port #2
037E STA $FC        ; Store joystick value
0380 CMP #$7E       ; Is sound received?
0382 BNE 037B       ; NO, continue polling
0384 NOP            ; No operation
0385 NOP            ; No operation
0386 NOP            ; No operation
0387 NOP            ; No operation
0388 LDA #$00       ; Stop the clock
038D NOP            ; No operation
038E NOP            ; No operation
:
:

```

enumeration task. The modified Voice Master was also used for vocal RT in a lexical decision task, where subjects were asked to respond as quickly as possible by saying either "word" or "nonsense." Twenty subjects were given 30 trials, 15 with words and 15 with nonwords. Likewise, 20 subjects were given 30 trials of a lexical decision task in which responses of "yes" or "no" were required.

To test further the accuracy of this device, and to demonstrate its versatility of application, it was employed in an additional experiment that required general recognition but not millisecond accuracy in timing. A Commodore computer equipped with a modified Voice Master was placed in a room with two rhesus monkeys (*Macaca mulatta*). Each monkey worked on one of two video-formatted tasks, administered by an IBM computer and described in detail elsewhere (Washburn, Hopkins, & Rumbaugh, 1989). One task required the animals to compete against one another, whereas the other did not. Vocalizations by each animal were tape-recorded and played to the Commodore in TRAIN mode. These vocalizations were to be distinguished in real time from other noises (such as the bells and whistles that accompany both video-formatted tasks) and tabulated. To evaluate the accuracy of recognition, a continuous video/audio tape-recording was made during each of six 30-min test sessions and scored for the number of vocalizations made.

Results

Numerical responses (0-9) from the first two experiments were correctly recognized on an average of 93% of the trials. A substantial increase in recognition accuracy was realized in the dot-enumeration task relative to the earlier frequency-estimation task (95% versus 89%, respectively). The improvement in reliability in the dot-enumeration task can be attributed to the fact that the vocabulary was trained in this experiment in conjunction with the RT subroutine, as described later (see the Discussion section, Suggestion 8). Recognition on the two lexical decision experiments, in which subjects responded "yes/no" or "word/nonsense," was 98.7% and 99.1%, respectively.

In the application with the rhesus monkeys, the computer correctly recognized all vocalizations as such ($N = 25$) and made no false positives in scoring. Although it is premature to interpret the meaning of the observed patterns of vocalizations, the implications for the reliability of the hardware are unequivocal.

Discussion

It is generally necessary to omit incorrect responses from the analyses of response times, and, typically, one would like to re-present trials in which incorrect responses are provided. This is not problematic with manual

responses or when a human experimenter is available to evaluate responses in real time. The device described in this article appears to offer a reasonably priced, efficient means of accomplishing these goals with vocal responses. Recognition accuracy of 95%-99% can be obtained with the device, and accurate timing in milliseconds is available. The following suggestions, gleaned from several attempts to employ this device in an experimental setting, are offered to improve effectiveness and efficiency.

1. Test subjects in an area free of extraneous noise.
2. Train each word twice for each subject, and clear the vocabulary between subjects with the BLANK command.
3. Carefully calibrate the Voice Master and adjust the Gain level before use. Instructions are included with the device.
4. Instruct subjects to provide "pure" responses without superfluous sounds (e.g., "uh ...") or paraphrasing (e.g., "twice" or "two times," rather than "Two").
5. Use responses as dissimilar as possible (e.g., "word/nonsense" versus "word/nonword").
6. Adjust the sensitivity (i.e., the ACCEPT value) of the recognition unit as instructed in the Voice Master documentation.
7. Always train the words to be recognized with a program that includes RT if RT is to be taken in the experiment. Because program control shifts to a RECOG statement only after the RT timer is stopped, the initial portion of many words may be lost. Train the vocabulary under these conditions as well to ensure accurate recognition.
8. Always use a good low- or medium-impedance microphone, and position the microphone in such a way as to maximize spoken volume without disruption from inadvertent sounds.

REFERENCES

- HOPKINS, W. D., & MORRIS, R. D. (1989). Laterality for visual-spatial processing in two language-trained chimpanzees (*Pan troglodytes*). *Behavioral Neuroscience*, *103*, 227-234.
- HORMANN, C. A., & ALLEN, J. D. (1987). An accurate millisecond timer for the Commodore 64 or 128. *Behavior Research Methods, Instruments, & Computers*, *19*, 36-41.
- KALLMAN, H. J. (1986). A Commodore 64-based experimental psychology laboratory. *Behavior Research Methods, Instruments, & Computers*, *18*, 222-227.
- PITTINGER, D. J., & HODGE, M. H. (1986). Statistical programs for Commodore computers. *Behavior Research Methods, Instruments, & Computers*, *18*, 50-52.
- RUMBAUGH, D. M., HOPKINS, W. D., WASHBURN, D. A., & SAVAGE-RUMBAUGH, E. S. (1989). Lana chimpanzee learns to count by "Numath": A summary of a video-taped experimental report. *Psychological Record*, *39*, 459-470.
- WASHBURN, D. A., HOPKINS, W. D., & RUMBAUGH, D. M. (1989). Video-task assessment of learning and memory in macaques (*Macaca mulatta*): Effects of stimulus movement upon performance. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*, 413-420.