THE VOICE DATA COLLECTION PROGRAM
A GENERALIZED RESEARCH TOOL FOR STUDIES IN SPEECH RECOGNITION

ROBERT BREAUX
Naval Training Equipment Center

MICHAEL W. GRADY
Logicon, Inc.

INTRODUCTION

The technology of machine speech understanding, integrated with advanced instructional technology, has demonstrated the capability to automate training for tasks which are characterized by the use of restricted, stylized speech. Such a training system for the GCA Controller has been described at NTEC/Industry Conferences in 1974 and 1975.[1,2] This paper discusses the development since that time of a more generalized laboratory research system designed to support speech recognition studies for evaluation of essentially any vocabulary set. Moreover, the program provides the experimenter with sufficient flexibility to investigate training the student on the vocabulary itself; that is, the critical task of teaching proper phraseology to the student.

BACKGROUND

The advantages of automated adaptive instruction are well known to the training community. Tasks requiring verbal commands and instructions had, until recently, been unamenable to these automated training techniques. Performance in speech-based tasks was traditionally measured by subjective instructor ratings. With the advent of the speech recognition capability, however, an individualized, self-pacing, automated training system for these tasks (now including objective performance measurement) became feasible.

The Naval Air Systems Command has supported the Naval Training Equipment Center's (NAVTRAEQUIPCEN) Human Factors Laboratory in efforts to establish design guidelines for such speech understanding-based systems. Indeed, the basic feasibility of these systems was demonstrated in a NAVTRAEQUIPCEN laboratory version of a Ground Controlled Approach Controller Training System (GCA-CTS) developed under contract with Logicon, Inc.[3,4]

LABORATORY R&D, AND THE VOICE DATA COLLECTION PROGRAM

Tasks such as those of the Landing Signal Officer, Air Intercept Controller, and Ship Conning Officer would also appear amenable to a speech understanding-based automated training system. Future R&D should proceed to address these new areas. The GCA Controller has served as a test application; now other vocabularies (applications) must be studied.

Furthermore, the problem of training the student in correct pronunciation and use of the radio terminology, operational brevity code, "standard commands," etc. is itself a subject for study. At least a basic knowledge of the relevant vocabulary is required prior to training for the task which employs the phraseology. R&D efforts can be wisely directed along these lines as well.

Finally, experience with the laboratory GCA-CTS has highlighted certain risk areas associated with the recognition of some phrases. Further R&D analysis of these areas with other vocabularies can support the definition of future training programs, identify problem areas early in the development cycle, and in general support the establishment of design guidelines.

Aware of each of these requirements, NAVTRAEQUIPCEN has supported the development of a highly flexible laboratory research tool called the Voice Data Collection (VDC) program. The VDC program provides the framework around which the researcher can:

a. Define vocabulary phrases associated with essentially any application.

b. Preprogram the presentation of phrases or groups of phrases to the speaker via text, graphics, and/or computer synthesized speech, resulting in an effective environment in which to learn the vocabulary phrases.

c. Test the ability of existing hardware/software algorithms to recognize these phrases, and extract hardcopy data on recognition reliability, correlation scores, and potential system confusions.

As such, the VDC program specifically addresses the three areas previously identified as possible directions of future R&D. The remaining sections of this paper are a brief description of these aspects of the VDC program. A more complete explanation of VDC functions and user capabilities is available in the references.[4]

VOCABULARY DEFINITION

The experimenter's capability to easily define the vocabulary set for investigation is provided via an off-line program which builds a disk file via a dialog with the user on the system teletypewriter. The program will accept the specification of 63 words or phrases in each file. Each of these 63 items is defined by giving:

a. The expected duration of the phrase; e.g., 800 milliseconds for "on glidepath." This datum is later used to (optionally) determine that the student is grossly mispronouncing or garbling the phrase.

b. An alphanumeric text sequence to be associated with the phrase. These data may simply represent the English text of the phrase or a specially defined code useful to an online program.

c. A phonetic text sequence to be associated with the phrase. Again, this may be the English pronunciation of the phrase or some other aural cue or response useful in the online environment.

d. A syntax structure. Associated with the VDC program is a rudimentary syntactical processor which enables the concatenation of phrases within the vocabulary list. For example, "assigned heading" must be followed by three digits. This specification is defined by the syntax structure associated with "assigned heading." Table 1 briefly describes the syntax options available.

In addition to its file creation capability, the off-line vocabulary definition program enables the experimenter to edit the disk file and generate hardcopy reports.

VOCABULARY PRESENTATION, TRAINING, AND VALIDATION

Readers familiar with the GCA-CTS efforts will recall that a necessary process is the extraction of voice characteristics of each potential user. In the early system, computer core restrictions necessitated a simple serial presentation of each individual vocabulary item. The phrase was typed onto the system teletypewriter, and the student was prompted to repeat the phrase the requisite number of times. The system collected data on the student's vocal characteristics, generated the necessary reference patterns, and moved on to the next vocabulary item.[2] Unfortunately, however, the accuracy of the recognition functions of the speech understanding subsystem (SUS) seemed to be directly related to the similarity between the student's voicing of a phrase when "configuring the recognition system" and during regular (GCA controlling) operation. It was difficult to achieve this vocal similarity using the scheme described above.

Also, during the early GCA-CTS efforts it was quickly realized that this "machine training" time — which was originally required only because of the necessity for collecting the speech reference data — could also be effectively utilized as "student training" time — to give exposure in the correct pronunciation and usage of the GCA vocabulary. The limitations of serial repetitions and individual phrases, however, severely hampered the exploitation of this dual mode.

In the VDC program, the experimenter is given complete flexibility in constructing (online or off-line) sequential presentations of multiphrase combinations. Additionally, the experimenter may associate a variety of cues with these phrases via the alphanumeric and phonetic texts in the vocabulary definition file. Furthermore, the presentations can be done independently, in combination with the reference data extraction process, or in combination with the recognition processes. Taken together, these capabilities provide the researcher with complete freedom in devising a uniquely tailored program which:

a. Allows the student to become acquainted with an unfamiliar vocabulary.

b. Transparently collects reference data on the vocabulary item spoken in a realistic sequence.

c. Transparently validates the accuracy of the reference data by attempting to recognize the phrase.

In short, a system can be quickly and easily developed which, from the student's perspective, is teaching him when and how to speak the vocabulary phrases, and, from the experimenter's perspective, is collecting data on the speech recognition and vocabulary training functions. A significant improvement in the quality (accuracy) of the extracted speech reference data is also achieved.

These presentation, training, and validation functions of the VDC program are facilitated by an easy and simple interface between the researcher and the software. Disk files and the presentation devices are named during a sign-on dialog with the system. The functional flow is directed via a command file created off-line, or by entering commands online via a named input device. Examples of these functional commands (regardless of their source) are given in Table 2.

If, during the validation mode of VDC program operation, the user does not specify

TABLE 1. VDC SYNTAX OPTIONS

Code	Definition	Examples*
1	Complete phrase	"execute missed approach"
2	Special word†	"zero," "one," "two," ... "point," "oh"
3	One of N phrases can precede this phrase	"above glidepath ... coming down"; "coming down"
4	One of N phrases can follow this phrase	"above glidepath ... coming down"; "above glidepath"
5.	One of N phrases must procede this phrase	"time to intercept ... track bravo"; "range and bearing ... track bravo"
6	One of N phrases must follow this phrase	"time to intercept ... track bravo"; "time to intercept ... track alpha"
7	N special words can precede this phrase	"7 ... aircraft"; "aircraft"
8	N special words can follow this phrase	"turn left ... 0 ... 1 ... 0"; "turn left"
9	N special words must precede this phrase	"6 ... miles from touchdown"; "4 ... miles from touchdown"
10	N special words must follow this phrase	"assigned heading ... 0 ... 1 ... 0"; "assigned heading ... 0 ... 1 ... 5"

†Special words are vocabulary items which:

1. may stand alone, and/or
2. may appear in a series, and/or
3. may precede a phrase, and/or
4. may follow a phrase.

*Phrases being defined are underscored. Other phrases provide context to explain the syntax.

TABLE 2. VDC FUNCTIONAL COMMANDS

Mode or File Type	Command	No. Arguments	Example	Meaning
"Presentation and Training"	P	1 – 12	"P 14"	Present phrase 14.
			"P 13, 14, 15"	Present multiphrase from concatenated phrases 13, 14, and 15.
	B	1	"B 0"	Begin collecting voice patterns for all phrases.
			"B 10"	Begin collecting voice patterns for phrase 10.
	S	1	"S 0"	Stop collecting voice patterns for all phrases.
			"S 10"	Stop collecting voice patterns for phrase 10.
	T	None	"T"	Terminate presentation function.
"Validation"	P	As above	As above	As above.
	A	None	"A"	Print all statistical data.
	S	None	"S"	Print summary statistical data.
	T	As above	As above	Terminate validation function.

either a command file or other input device, the system will (obviously) not present phrases to the student but rather will "echo" the phrases which it recognizes. The syntax structure defined earlier, together with some rather sophisticated logic, is used to concatenate phrases in this submode.

RECOGNITION ACCURACY AND DATA COLLECTION

An important aspect of the VDC program — especially when considered as a laboratory investigation tool — is its ability to easily and quickly assess the recognition reliability of the hardware/software system and identify misrecognized or confused phrases. Very often, once these risk areas are delineated, the eventual training system can be designed and implemented to support the recognition processes,* and/or minimize the impact of the recognition mistakes.

The results of the laboratory evaluations of the GCA-CTS, for example, have indicated that recognition confusions tend to occur between similar messages for the computer just as they do for humans, as, for example, the phrases "slightly above glidepath" and "slightly below glidepath." For some talkers, the plosive "buh" sound in both phrases tends to obliterate the other differences, particularly if the above/below portion carries unnatural emphasis, and thus the two can be confused. But consider the typical glidepath position error made by a student controller. He is not likely to say "slightly above" when he means slightly below. The more common mistake is failure to distinguish between adjoining zones; for example, by saying "slightly below" when the aircraft is actually below. In the latest GCA-CTS programs, delivered in May, 1976, the Performance Measurement Subsystem, in communion with the SUS, is able to resolve these recognition errors, resulting in understanding (as opposed to recognition) accuracies approaching 100 percent.

The real point to be made here is that, when typical recognition confusions are known, the system can often be designed to tolerate them. The VDC program enables the researcher to identify these problem areas very early in the analysis and definition phase, resulting in greater visibility going into the design phase.

Figure 1 demonstrates a portion of the hardcopy data available from the VDC program. The printout includes a header which provides identifying information followed by each prompted (presented) phrase and by the recognized phrase. Phrases which were misrecognized are flagged with an "*." Phrases that are recognized with low confidence are noted with an "@." Critical recognition parameters are saved for reference, and two overall phrase accuracy figures are calculated: for all phrases and excluding low confidence phrases. If requested by the user, the program will provide more detailed "internal" data collected during the recognition process. This information is particularly helpful to determine specifically which phrases are being confused, the degree of the confusion, and the effects of the various recognition parameters.

(In the "echo" mode described earlier, the comparative data shown in the figure cannot be calculated since the program has no *a priori* information on which phrases the student is speaking; i.e., no phrases are prompted. The VDC program will, nevertheless, save recognized phrases as well as the associated internal data.)

USE OF VDC IN THE TOTAL TRAINING PROGRAM

Though primarily a research tool, the VDC program is capable of supporting the laboratory task training programs as well. Indeed, the VDC concept was originally intended as a replacement for specific elements of the earlier GCA training system software, albeit with increased emphasis on vocabulary independence. By coupling the VDC and the latest GCA-CTS, together with the DEC display group and PDP-9 computer, the following experimental training scenario has been achieved for the GCA application.

A new student is signed into the system. He receives an audio/visual presentation of the GCA PAR vocabulary via the voice generation unit and graphics display. Via "show-and-tell," the student is instructed in the proper phrases for each radar pip position. The student participates in a follow-after-me mode in which he repeats GCA commands after an appropriate audio/visual cue. As he becomes familiar with the phraseology, the instructor quietly informs the system to start collecting voice data for specified phrases. The student continues this process, always being prompted as to the correct phrase for each situation. He garbles a phrase, but the system forces a repeat. The instructor, finally satisfied with the student's progress, commands the system to generate reference data based upon the information collected thus far.

The student returns later to practice more approaches. At first, he is still prompted by the system. The instructor commands a report on recognition accuracy. "Slightly above glidepath" is being confused

*Indeed, this is one of the distinctions between a speech understanding system and speech recognition. Understanding utilizes contextual and other cues to support the recognition algorithms.

VDC SUMMARY DATA

```
NAME:  GRADY, MIKE                    DATE:  5/31/76
VOCAB LIST FILE:  VLG:GCACTS.VL       VOCAB LIST NAME:  GCA-CTS VOCABULARY REV 01
VOICE DATA FILE:  DP1:GRADY.VD        VOICE DATA DATE:  5/31/76

****************************************************************************

PHRASE #:   1
  P: SLIGHTLY ABOVE GLIDEPATH
 @R: SLIGHTLY ABOVE GLIDEPATH

PHRASE #:   2
  P: SLIGHTLY BELOW GLIDEPATH
  R: SLIGHTLY BELOW GLIDEPATH

PHRASE #:   3
  P: GOING FURTHER ABOVE GLIDEPATH
  R: GOING FURTHER ABOVE GLIDEPATH

PHRASE #:   4
  P: GOING FURTHER BELOW GLIDEPATH
  R: GOING FURTHER BELOW GLIDEPATH

PHRASE #:   5
 *P: FOUR MILES FROM TOUCHDOWN
 @R: TWO MILES FROM TOUCHDOWN

PHRASE #:   6
  P: TURN LEFT HEADING ZERO ONE FOUR
  R: TURN LEFT HEADING ZERO ONE FOUR

PHRASE #:   7
  P: WIND 1 1 0 A1 1 0
  R: WIND 1 1 0 A1 1 0

PHRASE #:   8
  P: WIND 1 1 0 A1 2 0
  R: WIND 1 1 0 A1 2 0

M = 150   T = 60   VBCLS = 6   VBTV = 30   VCONF = 50

PERCENT CORRECT RECOGNITION:  87.5

WHEN LOW CONFIDENCE ITEMS CAN BE RESOLVED,
PERCENT CORRECT RECOGNITION:  100.0
```

Figure 1. VDC Summary Data — With Prompting

with "slightly below glidepath," so the instructor tells the system to again gather voice reference data for these particular phrases.

Finally, when the instructor is confident in both the basic knowledge of the student and the proper configuration of the speech system*, the student is introduced to the GCA-CTS. In an effort to describe the basic system concept, the student is encouraged to speak any of the GCA phrases he has been practicing. The student listens and finds that the computer is repeating back to him whatever he says! He is amused by a strange "d u u u h h h" when he speaks illegal phrases into the live mike.

The student is anxious to have his performance objectively measured by the computer training system. In order to facilitate his explanation of errors, the instructor commands an automatic freeze whenever the system detects an error. The student catches on very quickly, so the instructor places the GCA-CTS in a fully automated mode and leaves the student to himself. Later the instructor returns, reviews the student's performance via hardcopy printouts, and decides to advance the student past several levels into a more difficult syllabus sequence. The student finally completes the entire syllabus and is ready for live GCA exercises.

As new areas for the application of speech recognition are identified and studied, a variety of similar scenarios will also be brought to reality.

REFERENCES

1. Goldstein, I., Norman, D., et. al. Ears for Automated Instruction Systems: Why Try? Proceedings of the Seventh NAVTRAEQUIPCEN/Industry Conference, 1974.

2. Breaux, R., Goldstein, I. Developments of Machine Speech Understanding for Automated Instructional Systems. Proceedings of the Eighth NAVTRAEQUIPCEN/Industry Conference, 1975.

3. Breaux, R. Training Characteristics of an Automated Adaptive Ground Controlled Approach Radar Controller Training System. Technical Note: NAVTRAEQUIPCEN TN-52, Orlando, Florida. To be published.

4. Grady, M.W., Hicklin, M. The GCA-CTS — A Demonstration Training System for the Ground Controlled Approach Controller. Final Technical Report: NAVTRAEQUIPCEN 74-C-0048-1, Orlando, Florida. To be published.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the guidance of Ira Goldstein (NAVTRAEQUIPCEN) and Larry Egan (Logicon) during the course of this effort. The VDC programs were designed and implemented by M. Grady, M. Hicklin, R. Watson, and R. Barnhart. The work was conducted under NAVTRAEQUIPCEN Contract N61339-74-C-0048.

*Ideally, the fully automated system would continually update the speech reference data until some criterion of recognition accuracy was met. At that point, the system would signal the instructor to shift (or automatically shift) to the performance scoring capability.

ABOUT THE AUTHORS

DR. ROBERT BREAUX is a Research Psychologist in the Human Factors Laboratory at the Naval Training Equipment Center. He has an interest in application of theoretical advances from the psychological laboratory to the classroom situation. Papers include computer application for statistics, basic learning research, concept learning math models, and learning strategies. He is an instrument rated commercial pilot.

MR. MICHAEL W. GRADY is the Project Leader at Logicon's Tactical and Training Systems Division in San Diego, California, for programs utilizing the advanced speech technologies: synthesized speech and speech recognition. He brings to this position several years of experience in real-time systems design and development, particularly in both large and small scale training programs.