VOICE-BASED LEARNING
William D. Wagers

ABSTRACT

A voice-based learning system is a computer system which uses voice input and voice output in an educational environment. Such systems promise two-way communication with the computer using natural language. In addition to reading and writing skills, the computer can now teach listening and speaking skills.

Foreign language educators around the country have been experimenting with voice-based learning to determine how best to utilize these new tools in the classroom and in the language lab.

The Babel fish is a fictional creature in the novel *Hitchhiker’s Guide to the Galaxy*. This fabulous creature has the ability to translate from any foreign tongue into one’s native language. Simply insert the fish into your ear and you have a universal translator. The Babel fish does not understand what is being said. It is a passive device like a hearing aid.

A voice-based learning system is built around this very capability. It can understand and speak virtually any language. However, instead of inserting the fish into your ear, the fish is in the computer’s ear. In this case, the fish translates from spoken language to computer language. While the computer does not understand what it hears in the human sense, the computer can associate sound patterns with written words and phrases. Embodying these associations in the form of a lesson is the function of a voice-based learning system.

A voice-based learning system includes a computer, a voice input terminal and a voice output device. It interacts with the student in oral language. While a computer does not possess the intelligence to interact fully as a human would, it can communicate.

Voice Input

Before describing a voice-based learning system in detail, a few words need to be said about voice input. A voice entry-terminal is simply a computer peripheral which accepts spoken input, compares that utterance against several known utterances, and when it finds a match, supplies the computer with the character string corresponding to that utterance. In other words, when you say, freund, the voice entry terminal selects the character string freund. The association between the spoken and written word is arbitrary from the computer’s point of view. The associations are established by the author of the lesson. The author could just as easily associate freund with friend.

Voice entry terminals fall into two categories. Speaker independent devices recognize a limited number of words when spoken by any user. The vocabulary, the utterances it recognizes, of such a device is fixed. This characteristic limits its usefulness to multiple choice or true/false questions. Also, speaker independent systems are quite expensive.

Speaker dependent devices must be trained to recognize the voice of a particular speaker. You train the device by repeating each utterance into the system microphone several times. The vocabulary is created by the user (author) and is kept on disk, so each lesson can have a separate vocabulary. Speaker dependent devices are available on many microcomputers for under $1,000.

Typically, when speaking to a voice entry terminal, you must speak clearly and consistently. You must also pause between the phrases you want recognized. The pauses tell the computer when an utterance begins and ends.
Voice Output

High quality audio is critical in voice-based learning, so speech synthesizers were never an option. While synthesizers are cheap and reliable, they simply are not good at reproducing the richness of expression of the human voice. Videotape was eliminated because of the reliability problems and videodisc was ruled out because of the expensive mastering process. And though high-quality tape and cassette players have been around for a long time, voice-based learning requires computer control of the player.

There are currently at least two devices on the market which I have found are suitable for voice-based learning: the Instavox RA-12A and the Tandberg TAL 822. Both devices offer high-quality audio and computer control. The Instavox is a giant floppy disk drive which plays analog recordings but allows random positioning. The Tandberg can position to any message on the disk in under half a second. This means the computer can prompt or respond to the student immediately throughout a lesson. The Instavox retails for under $5,000. The media costs $25 per diskette.

The Tandberg TAL 822 is a random access cassette player. It uses a highly accurate frame counter (0.5 second) to position at rewind speed to any place on the tape. The media is the standard audio cassette the positioning is not as accurate as that of the Instavox, and the response time is slower. However, the TAL 822 retails for under $1,500, so for a slight degradation in performance a significant savings can be achieved.

A Sound Basis for Learning

The VBLS Authoring System was announced by Scott Instruments Corporation in 1982. It was designed to allow classroom teachers to create courseware which a student would study by voice. The System consists of an Apple-compatible microcomputer, a VET voice entry terminal, and an optional voice output device.

This system gives the following advantages over traditional CALI:
1. **Visual Pronunciation Feedback.** The System has a mode in which the student's pronunciation is checked against that of a native speaker. The System displays a bar graph which shows how closely the student's utterance matches the standard. This is an extremely powerful tool for progressively refining the student pronunciation skills.

2. **Auditory Feedback.** The System plays the standard phrases as spoken by a native speaker. The student can listen and emulate the native.

3. **Helps Eliminate Unlearning.** A major difficulty in learning a foreign language is that a student frequently learns incorrect pronunciation due to inadequate visual and auditory feedback. It is far more difficult to unlearn poor pronunciation, than to learn correctly from the outset.

4. **Privacy.** Students are free to experiment, make and correct pronunciation errors without fear of embarrassment. Only the System is listening and it offers only encouragement.

5. **Success Is Reproducible.** Voice-based courseware allows successful language instruction techniques to proliferate. Each student is presented with the same learning opportunity, whereas the quality of classroom instruction varies from instructor to instructor and even from day to day.

6. **Active Learning.** Many traditional instructional technologies are passive, e.g., books, filmstrips, videotape, and cassettes. Voice-based learning permits the student to interact with otherwise passive aids in an interactive process.

7. **Ideal Computer Interface.** The human voice is the ideal means of communicating with a computer, especially in foreign language instruction. While the ideal technology is not yet available, current technology is a dramatic improvement over the QWERTY keyboard. Students are introduced to the computer in an unobtrusive manner. Virtually no typing skills are needed.

8. **Multisensory Learning.** The System stimulates multiple senses which reinforce learning.

Preparing Voice-Based Courseware

To create a lesson, the author first creates the text frames and associated responses which will appear on the CRT. The author is assumed to be a native speaker. The author then creates the audio sequences which correspond to the lesson text. Finally, the author creates the standard pronunciations, or templates, against which the student's utterances will be compared.

The menu-based authoring system makes the process very simple, even for the novice computer user.

The VBLS Grant Program

While the VBLS system was designed for general education, many of its early users were foreign language educators. At the 1982 FLINT Conference, Dr. Brian Scott announced the VBLS Grant Program. Its purpose was to stimulate ideas on improving the VBLS system for use in teaching foreign language. We offered a free VET voice entry terminal and VBLS software in exchange for the completion of a project defined by the grantee. Typical project proposals were for subjective studies or for units of courseware.

We awarded 15 systems to various applicants. A fair proportion of the awards went to FLINT attendees. In addition, we placed a System at the C.A.C.I. Language Center in Arlington, Virginia. C.A.C.I. supplies language training to governmental agencies and proved to be a fountain of ideas and approaches to voice-based learning. At the time of this report, some grantees have yet to fulfill the terms of their grant. However, the preliminary results were very useful in defining a foreign language version of VBLS.

The Grant Program is now closed but we hope to have an ongoing relationship with those grantees who demonstrated skill and imagination in executing their projects. For example, Dr. Robert Baker completed his Russian courseware despite the absence of a Russian character set in the original VBLS software. Dr. Harry Wohlerl worked with us to implement a version which used an ordinary cassette recorder. Nick Stancioff went so far as to hire a programmer...
to try various voice-based learning schemes.

The version of the VBLS system which was available at the time of the Grant Program did not have foreign character sets or voice output. The current version has both foreign characters and voice output via either the TAL 822 or the Instavox. As you might imagine, the two most frequent criticisms of the original version were that you could not hear the correct pronunciation and that you could not see the correct spelling. These features are now a part of the VBLS system.

When all the reports from the grantees have been filed, perhaps their studies can be made available in some form. Their work is valuable because the voice entry terminal's performance varies with the language in use. Each language has particular problems related to the sensitivities of the terminal.

**VBLS System Today**

VBLS Version 1 is in use in approximately 100 installations. About fifty percent of these systems were acquired for research in areas such as auditory rehabilitation, communication training, and many others. Approximately thirty percent were acquired for use in special education by slow learners, foreign students, hyperactive children, or physically handicapped students. Perhaps twenty percent were acquired for foreign language training.

VBLS Version 1.4 with foreign characters and voice output has just become available. It is our hope that this unique system will now begin to gain acceptance with foreign language educators. We have created only a tool. The quality and quantity of courseware for the System depends upon interest and commitment by the educational community. We work closely with prospective authors to introduce them to the new technology.

We are continuing the research and development of the VBLS system. In future versions, we aim to reduce the cost of the hardware involved. We are also developing new learning modes for the System. And research is continuous on voice recognition, independent of its educational applications.

**Summary**

Voice-based learning is a new type of computer-assisted instruction. It holds the promise of increasing both the availability and the quality of language instruction. The fulfillment of this promise depends on the participation of foreign language educators, both in determining the proper role of voice-based learning and in embracing the technology by producing courseware.

**Definitions**

**Authoring System**—A systematic procedure for generating CAI courseware that does not require the author to have computer programming skills.

**Continuous Speech**—Everyday natural speech like humans use. Voice recognition products do not yet understand continuous speech without artificial pauses between words and with large vocabularies.

**Courseware**—A computerized course, typically including text, workbook, and software for a specific computer system.

**Frame**—One in a series of visual displays on the monitor of a computer system.

**Human Factors**—A voice recognition device must account for the variations inherent in human speech and human environments, e.g., inconsistent pronunciation, variations in tone, pitch, duration, and emphasis, and environmental noise variations. Humans find it impossible to say the same word twice in exactly the same way, especially if they are ill or under stress. And, of course, different speakers will pronounce the same word differently. A voice recognition device must, therefore, adjust itself to these variations.

**Isolated Word Recognition**—Current voice recognition products recognize only words or phrases which are in its vocabulary and which are spoken as short commands or answers. Affordable products cannot pick the desired word out of a stream of continuous speech.

**Keyboard Terminal**—An archaic typewriter-style keyboard for manually typing data into a computer.

**Prompt Square**—Symbol used to indicate whether the microphone is ON or OFF; located in the lower righthand corner of the monitor when the System is listening.

**Speaker Adaptive**—Refers to a system which overcomes the limitations of speaker dependence through software which adjusts itself to different speakers.

**Speaker Dependent**—Current products are speaker dependent, meaning they must be trained to recognize the voice of a particular speaker, and will respond reliably to no other speaker.

**Speaker Independent**—A product which could understand any speaker using any accent.

**Speech Synthesizer**—Electronic device capable of producing speech-like sounds.

**Template**—A pattern or guide; a computerized representation of a spoken word; speech patterns.

**Training**—The process of inputting several pronunciations of a word or phrase for future matching to one's own or another individual's voice; enrollment; creating a vocabulary.

**Vocabulary**—Group of words used by or known by a person or group of people; the words trained for a specific VBLS lesson, and therefore known to the System.

**Voice-Based Learning System(s)**—Computer systems that use vocal dialog between the user and the computer for instructional purposes.

**Voice Entry Terminal**—An input device enabling a human being to speak directly to the computer; VET; VET-2; ShadowVET.

**Voice Recognition**—The identification of spoken words by a device. A speech recognition device can respond to the human voice, instead of, or in addition to, a typewriter keyboard terminal.

**WORD DRLL**—A VBLS study session mode which allows the user to practice pronunciation and receive immediate visual feedback on the correctness of pronunciation.
VBLS SYSTEM MENU V 1.0

SELECTION FUNCTION
1 RUN LESSON
2 LIST LESSONS
3 DISK FUNCTIONS
4 CHANGE STUDY DISK PROFILE
5 CHANGE PROFILE RECORD
6 REVIEW PERFORMANCE HISTORY
7 AUTHOR NEW LESSON
8 CHANGE OLD LESSON
9 TRAIN OLD VOCABULARY

TYPE SELECTION NUMBER

1. España tiene cinco ríos
grandes y muchas _______.

SAY: montañas

Figure 2

...1...2...3...4...5

LAST SCORE: 3.2 PHRASE LENGTH OKAY
BEST SCORE: 3.2

SAY: montañas

(Press ESC for next word)

Figure 3

Bibliography: Books

Bibliography: Articles

Bibliography: Presentations
Horn, Carin E. and William D. Wagers, "Voice Controlled CAI, Presentation," Proceedings of the National Educational Computing Conference '82.

Bibliography: Unpublished Materials