LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

LA-UR--89-443

DE89 007758

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under Contract W-7405-ENG-36

THE POWER OF DIGITAL AUDIO IN INTERACTIVE INSTRUCTION: AN UNEXPLOITED MEDIUM

Author(s): Jennifer Pratt
Mary Trainor

Submitted to: Seventh Conference of the Interactive Instruction Delivery

DISCLAIMER

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

The Power of Digital Audio in Interactive Instruction: An Unexploited Medium

Jennifer Pratt and Mary Trainor

Abstract

Widespread use of audio in computer-based training (CBT) occurred with the advent of the interactive videodisc technology. This paper discusses the alternative of digital audio, which, unlike videodisc audio, enables one to rapidly revise the audio used in the CBT and which may be used in nonvideo CBT applications as well. We also discuss techniques used in audio script writing, editing, and production. Results from evaluations indicate a high degree of user satisfaction.

Introduction

An abundance of literature on computer-based training (CBT) issues is available. However, one aspect, the area of interactive audio, is not very thoroughly addressed. If audio is even mentioned, most references make statements similar to the following:

"The most common type of auditory communication from machines to people is the tonal output of such devices as buzzers, bells, horns, or whistles."(1)

"The use of sound in computer-based instruction is not as advanced as the visual modes. Its most common use is for primitive sound effects such as beeps or explosions. The use of sound for music or human voice synthesis is more difficult and infrequently done."(2)

At its inception, CBT corresponded more or less exactly to the description given in the second quote above. Audio was used primarily as an alerting mechanism, useful to apprise the user of mistakes or readiness states. We are still exploiting this rather useful, if somewhat monotonous, feature of audio to make ourselves aware of inappropriate input to our keyboards.

In the early 1970s, the PLATO systems had a separate audio player interfacing with the program. It was not very successful, perhaps because it was not integrated with the program.

Audio was used for reinforcement as music made its way into the field. Music has been used to reinforce right answers as well as to fill in the dead space created by disk access times.

Digital audio technology has been in existence for some time but has experienced very limited usage in CBT applications. As videodisc technology began to appear in CBT, audio in the form of spoken language began to have an impact on the design. Audio has been used to deliver instructions, provide sound effects, create scenarios, explain concepts, and prescribe remediation.

The advent of digital audio created the circumstances necessary for effective application of audio in CBT. Now it was possible to add a new and useful dimension to instruction through the computer.

This paper begins with a rationale supporting the use of digital audio in interactive CBT. Audio is effective in learning situations, especially when audio and visual material enhance each other(2). Next, the methods and equipment available for implementation of digital audio are discussed. Finally, this paper addresses the potential created by the emerging technologies that will facilitate the use of digital audio, specifically CD-ROM.

Audio and Learning

The advantages of dual-modality instruction--for instance, the visual mode combined with the auditory mode--are well known and well documented(2). For a variety of reasons, not the least of which is learner preference, people learn better when information is presented both visually and auditorily. Because of hardware constraints, most CBT has concentrated on the visual mode to the exclusion of the auditory mode. In fact, many reference books that purport to teach the design of CBT fail to even mention the auditory channel and certainly do not elucidate the opportunities that it presents.

An important factor in evaluating the potential benefits of audio is the extra information that the spoken word can convey. While text on the screen can provide the same lexical data, speech has a number of parameters to carry connotations that are not available through simple written text. The meaning of any utterance can be modified by intonation, which is really a combination of rate, volume, and pitch. The intonational contour on an utterance can imply shades of meaning ranging from emphasis to contradiction.

Types of Interactive Audio

The most popular way to incorporate audio is to play recorded audio material on the **videodisc**. The obvious advantage is the high quality and reliability of the audio

produced in this way. If video is intended to be part of the CBT project, using audio in this way is convenient. However, video is not always appropriate for every CBT application. It is a very effective, but a very expensive element of CBT. In addition, the audio on videodiscs has some constraints. The specific utterance associated with a picture or series of pictures on a videodisc is irrevocably attached to that segment of video. Although the audio portion may be suppressed, it may not be moved to another segment. Additionally, the utterance itself is absolutely immutable. Once pressed, the videodisc is permanent, both an asset and a liability, for many applications require occasional updating or revision as techniques or equipment change.

Another way to incorporate audio into CBT is to use a **speech synthesizer**. While new advances continue to improve the quality of synthesized speech, it remains unacceptably mechanical and unintelligible to many people's ears. The best synthesizers are still expensive; and until the voice quality more nearly approaches that of natural human speech, they remain of less value to the CBT field than digitized recorded speech(3).

Digitized speech, or **digital audio**, is recorded human speech that has been electronically sampled and digitized. The higher the sampling rate, the more natural sounding the speech. Of course, higher sampling rates require more storage as well. The large storage requirement is the main disadvantage of digital audio. For example, this sentence would require 82 kilobytes of storage. A typical one-hour CBT course would require anywhere from 9 to 12 megabytes of storage for the digital audio alone. The audio files can be stored on hard disk, but a particularly promising new avenue for storage is CD-ROM.

After the script has been completed, the utterances are recorded on audiotape. The tape is digitized and each utterance is indexed and stored in an audio file. The file is compressed to remove extraneous parts and to save space. Experience has demonstrated that smaller files are easier to create, manipulate, and modify than larger files. Whereas an audio editor will accommodate up to 128 indices in a file, it is difficult to work with more than about 40.

One of the great advantages of digital audio in CBT is its flexibility. Audio provided by a videodisc is static and unchanging. You can delete it, but for updating a program one cannot modify it without producing a new videodisc.

Digital audio, however, can be changed much more easily. New dialogue can be recorded, digitized, and inserted into the existing program relatively easily. This

capability provides each program greater longevity because it can be modified to meet changing needs. It is useful both in supplementing audio used on the videodisc or as the sole source of audio in graphics/text-based CBT. The Cognitive Systems Engineering (CSE) Group at Los Alamos has been using audio heavily in CBT development for three years.

A little forethought and the creative use of digital audio can allow a CBT developer to modify the CBT with relatively little effort. A few changes to the script and a few minutes with a tape recorder and an audio editor replace an old audio file with a new one. The program is suddenly more complete, reaping the advantages of the iterative design/development process.

Audio Script and Design Considerations

List 1 enumerates some design decisions that should be made when the design document is written.

List 1. Guidelines for Interactive Audio

- Consider the audience's (students') age, gender, education, and autonomy at the workplace when choosing voices.
- Designate each voice's purpose (e.g., male voice gives instructions, female voice gives feedback). More than one voice may improve the presentation.
- Balance clarity of pronunciation against formality of style.

Whenever speech is introduced into a CBT program, it carries with it the element of personality. This element has both assets and liabilities. While personality can enliven the presentation and offers opportunities for a variety of approaches, someone always dislikes any given personality.

It is critical to evaluate the prospective user audience accurately and thoroughly to minimize this kind of problem. Consider what kind of voice will be most acceptable to a given audience. A well-known story about the choice of voice for an application relates that the engineers chose a particular voice because they found it pleasant, but the target audience rejected it as sinister and evil(1). Although this incident occurred with a synthesized speech application, it illustrates that different groups attribute different "personalities" to voices.

A related concern is the balance between a natural sounding, conversational speaking style and a formal, clearly enunciated speaking style. It is normal to slow down and speak more clearly when one cannot see the listener's reaction, which is typical of lecture situations. This style, however, is slightly less interesting to listen to than conversational style. In conversation, where the level of listener comprehension may be inferred from the facial expression and verbal acknowledgment, speakers tend to use a faster rate and less ponderous enunciation.

Thus, one always compromises between formal, highly intelligible speech and less intelligible but more interesting speech. The requirement for good intelligibility in CBT applications suggests that this compromise be resolved more on the side of formality than would otherwise be desired.

The tone of the audio--public speaking style versus informal chat style and all the possibilities in between--depends also on the audience. For example, an audience that is accustomed to a large amount of autonomy in its workplace is likely to resent an authoritarian tone. Some audiences will react better to a voice of a particular gender, age, or dialect. In the absence of research on this subject, designers and developers would be wise to field-test several different voices and rely on user feedback to make this decision.

In some cases, it can be helpful to implement different functions within a program by using different voices. For instance, a male voice might be used to deliver narration and instructions, while a female voice might be used for corrective feedback. This system can help the student identify different instructional segments as well as provide some variety.

Several important considerations in making the actual audio recording should be contemplated. List 2 presents several suggestions for improving the intelligibility of the digital audio. Most of these can be implemented during the recording session itself, and they are largely self explanatory.

List 2. Technical Considerations in Recording

- Always rehearse well before the final recording session.
- Always take a trained listener when recording scripts; the listener can coach the speaker and identify which utterances need to be re recorded.

- Pay special attention to the first word or phrase; listeners tend to rely heavily on the initial phrase for comprehension.
- Consider the rate of speech; slow pacing is easier to understand but may make a listener impatient.
- Consider diction; more distinct diction is easier to understand but may sound stilted.
- Pronounce vowels distinctly. The vowels carry most of the sound energy.
- Use intonation to distinguish questions from statements.
- Make similar sounds that occur in sequence more distinctive by unusual intonations.
- Exaggerate mouth movements slightly and keep the mouth open to avoid the tendency to lock the jaw.

Experiences with Digital Audio

The CSE group has recently completed a CBT project for the Air Force, which uses audio both from a videodisc and in digital form. The program HEAT uses videodisc to set up a scenario involving a heating problem. The user, an Air Force heating specialist trainee, is asked to troubleshoot the problem through guided practice with instructive feedback displayed on the screen. The feedback is tailored to the specific errors made so that the user can be given hints that will help her him reason to the correct answer. A sample of the script is provided below.

HEAT Script Sample

Question: You have completed your boiler control inspection and found that all the components are working properly. Where do you want to start your burner inspection?

Combustion Chamber

Right Answer: Right.

First, second, and third hints if answer is wrong:

Reasoning 1: There's a safety check that needs to be made. Select another choice.

Reasoning 2: You need to make a safety check. Try again.

Reasoning 3: For safety, you need to inspect the combustion chamber. Choose combustion chamber

Primary Control

Right Answer: Right.

First, second, and third error messages for specific wrong answers:

Fuel Line

Message 1: You are not ready to inspect the fuel line yet.

Message 2: No, you need to think about safety before inspecting the fuel line.

Message 3: No, inspecting the fuel line now can present hazards.

Transformer

Message 1: No, the transformer is an electrical component.

Message 2: No, you need to think about safety before inspecting electrical components.

Message 3: No, inspecting the transformer now can present hazards.

Burner Assembly

Message 1: It's important to take precautions.

Message 2: No, you need to think about safety before inspecting the burner assembly.

Message 3: No, inspecting the burner assembly now can present hazards.

In the original conception, the project called for the audio to be provided by the videodisc. The unit portion of the program was limited to narration and the delivery of instructions. Feedback was provided through text on the screen. Although the level of interaction was

high and the program moved well, a demonstration of digital audio feedback was convincing enough that the client requested a modification of the program. The second version of the HEAT program included digital audio feedback for every response. Preliminary reactions indicate that the digital audio feedback increases the interest level of the program considerably. Using audio helped the trainee concentrate on the visual display of the equipment because it was not necessary to read the text feedback.

Summary

Learning is more effective when the instruction is delivered through both the auditory and visual channels. However, CBT applications have failed to exploit this fact. This paper has explained the relative merits of several different ways of incorporating audio into interactive CBT. Although each method has value, digital audio deserves closer attention from designers because of its distinct characteristics.

Digital audio is a versatile tool that can enhance other audio presentations or stand alone to provide interactive audio in video or graphics/text-based CBT. Digitized recorded speech is useful because of its flexibility, its relatively low cost, and its high voice quality and intelligibility.

References

1. Michaelis, P.R. and Wiggins, R.H. A Human Factors Engineer's Introduction to Speech Synthesizers. In Badre, A. and Shneiderman B. (Eds), Directions in Human/Computer Interaction. Norwood, New Jersey: Ablex Publishing Company, 1982.

2. Alessi, Stephen M. and Trollip, Stanley R. Computer-Based Instruction: Methods and Development. Englewood Cliffs, New Jersey:Prentice-Hall, 1985.

3. Marics, Monica A. and Williges, Beverly H. "The Intelligibility of Synthesized Speech in Data Inquiry Systems". Human Factors, 30 December, 1988: 719-732.

4. Schultejann, P.A., Spangenberg, L.M., Andrews, A.E., and Trainor, M.S. "The Application of Computer-Based Training Principles to On-the-Job Training for Air Force Heating Specialists." In Proceedings of the 1988 Conference on Technology in Training and Education ed. Cristy L. Vitale, pp.2-11. Biloxi, Mississippi, 1988.

About the Authors

JENNIFER PRATT
Cognitive Systems Engineering Group, A-6
Los Alamos National Laboratory
P.O. Box 1663, MS M997
Los Alamos, NM 87545

505 667-0990

Ms. Pratt earned an M.S. in Communicative Disorders and is employed at the Cognitive Systems Engineering Group, Los Alamos National Laboratory while pursuing a degree in computer science. She has worked in education for eight years, primarily in the area of communication and cognition. A background in audiology and an interest in computer science has led her to the area of interactive audio.

MARY S. TRAINOR
Cognitive Systems Engineering Group, A-6
Los Alamos National Laboratory
P.O. Box 1663, MS M997
Los Alamos, NM 87545

505 667-0990

Dr. Trainor holds a Ph.D. in science and mathematics education and is currently deputy group leader of the Cognitive Systems Engineering Group, Los Alamos National Laboratory. She has been an instructional psychologist in CBT projects for 14 years. She is currently the president-elect of the Association for the Development of Computer-based Instructional Systems (ADCIS).