ABSTRACT: This paper describes a system for the use of digital audio in foreign language instruction. The system consists of networked personal computers with digital audio adapters. The use of a digital audio adapter allows for the flexible manipulation of audio materials: natural sounding audio may be played, recorded, replayed, paused, and ordered under computer control with no distracting delays or distortion. In addition, audio can be coordinated with the display of text or graphics on the computer screen. Finally, the network allows for the central storage of the digital audio materials and computer programs. Only a single copy of any audio material needs to be produced and stored on a network file server. This material is then available for selection at any time for independent use by any number of students.

KEYWORDS: digital audio, tools, interactive audio, network, dictionary, dictation, reading comprehension.

In this paper we describe a media center featuring a network of computer workstations with integrated digital audio and interactive video capabilities. At present, the Center for Foreign Languages and Cultures (CFLAC) is nearing completion and, barring any unforeseen delays in the arrival of equipment, should be up and on-line by the time this paper appears. The information which follows represents a "work in progress" and is the product of the development and testing of individual software and hardware components and their interaction. Many of the decisions are of necessity arbitrary and will be modified as the facility is tested by the rigors of actual instructional applications.

Center for Foreign Languages and Cultures

Approximately four years ago Smith College began planning to replace its aging Language Lab. The existing lab had served the foreign languages well but was nearly 20 years old. At the time, a number of faculty members expressed a strong interest in the development of a "state of the art" facility which would provide audio, video, and computer capabilities without sacrificing any of the current lab's proven features. Despite its age, the old lab system did allow
students a degree of independence and convenience. Students could listen to
different passages, record on a separate audio track, and, (perhaps the most
important operational principle) students never had to handle an audio tape of
any kind.

The project quickly became CFLAC, the Center for Foreign Languages and
Cultures. The name reflects the fact that the Center supports teaching and
research in the field of language as part of a cultural context rather than
promoting the mechanistic view underlying much of the established
preconception of language. We began planning by looking at a number of
different commercially available "Media Lab" systems but soon concluded that
none of them provided the integration of audio, video, and computer that we
were seeking: the ability to combine audio from one source with video from
another; to combine audio with computer text and graphics displays; and so
forth. We ultimately decided to develop our own component system, built
around a personal computer, which would allow CFLAC to grow and change
with the technology. In this paper, we will emphasize the use of digital audio
and networks to integrate audio and the computer. Many of the ways in which
we have combined audio and computer applications, however, strongly suggest
our approach towards integrating video as well.

Pedagogical goals

Traditionally, the teaching of a foreign language has been divided into
four components: reading, writing, listening, and speaking. Although such a
division may at times prove useful, it is admittedly artificial: the teaching of one
component cannot (and should not) be done in complete isolation from the other
three. While we recognize that exercises often focus on one aspect or another, we
want CFLAC to support the integrated teaching of the language. This goal
reflects several concerns. First, there exists a great deal of literature indicating
that the use of multiple information sources (e.g. images and text, sound and
text, sound and images) improves the retention of information1.

Second, we sought to counter the impression that computer technology is
of little help in the acquisition of speaking skills2. By placing the teaching of
speaking back into the context of the other components, our new center could
demonstrate how valuable an aid technology could be in teaching speaking.
Third, we believed that an integrated approach was most similar to the
approaches faculty had found successful in the classroom setting, where, for
example, a listening exercise may also involve writing or a speaking exercise may
involve listening or reading. It was this combination of modalities that we
wanted our laboratory environment to support.

We also wanted to provide the benefits of individualization and
immediate feedback which computer workstations offer. Many of the necessary,
yet time-consuming drill and review exercises, as well as audio and video
previewing, could be completed outside of class time. In the lab setting, students
could receive more individualized support than is possible in a large group setting. More class time would then be available for human interaction and communication in the target language.

At many colleges there has been a shift in foreign language pedagogy towards greater linguistic proficiency. At Smith, this shift has been coupled with the maxim of "proficiency in context" or cultural proficiency. To be effective, we believe that foreign language acquisition must include a cultural component as well. Teaching a foreign language course involves more than teaching the language per se. Verbal and non-verbal components are linked intrinsically in establishing true communication. This desire to develop proficiency in context has resulted in an increased demand for culturally authentic audio and video materials, and more importantly for our present concern, for new ways to assist students in gaining access to these materials earlier in their foreign language instruction. The use of authentic sources from the very beginning can lead to the more rapid development of communicative competence. They promote students’ acquisition of strategies to cope with "realistic" natural-speed speech patterns, rather than with a "textbook" language of controlled vocabulary and structures. Integrating audio and video with text and graphics will assist the student in training herself to understand both the foreign speech spoken naturally and the cultural context in which it was spoken.

Since many faculty had found the computer a valuable tool for their own research in language, culture, and literature, it was important that CFLAC support these tools so that students would come to appreciate their value for their own research and would learn to use them actively. This approach promotes not only a better understanding of the intricacies of a language and its culture, but also new techniques for exploring them.

Finally, to view CFLAC—despite the obvious advantages its technology offers as a drill tutor and research tool—as just a repository for the collection of standard learning tasks would be a mistake. The attraction of authentic audio and video materials should invite students to use the center freely to explore the foreign culture on their own. Faculty must be encouraged to engage in research projects involving technology and pedagogy to seek creative new ways to use the facility to help improve student learning.

Operational Goals

One of our primary operational goals was to minimize the procedural complexity of using and maintaining the lab facility. There should be no need for a large support staff scurrying around behind the scenes to make the center work. We wanted students to walk into CFLAC, choose a workstation, sit down, and be able to start work immediately. They should be able to select the course,
the lesson, and the task for study without requiring the intervention of a staff member. This approach affords the students more freedom to schedule their own activities and minimizes the required number of staff. Again, it should encourage the students to visit the lab voluntarily and frequently.

A second goal was to ease the creation and to preserve the quality of the audio materials. Since many faculty members prefer to create their own audio materials, or modify those provided by publishers, the creation of new materials and the repurposing of old ones should be straightforward. The procedure for doing so should not discourage the development of the most appropriate materials for a specific task.

Our operational goals are not goals of necessity but goals of convenience. If a facility is more convenient to use, students will choose to use it more. If a facility is easier to maintain, staff can work on more important and creative activities. If audio and textual materials are easy to modify, faculty will individualize them to meet specific needs.

**Hardware Selection**

With these goals in mind, we began to investigate the current level of development and anticipated improvements in technology. From the beginning, it was apparent that personal computers would be an integral part of any solution we found. CFLAC had to provide the standard personal computer-based productivity tools which have greatly influenced the creation, translation, transcription, and structural analysis of materials for teaching and research. Students should be introduced to their use in conjunction with the foreign languages they are learning. This technology also offered the possibility of using computer-based learning materials while taking advantage of the computer’s ability to individualize instruction. With personal computers, this flexible control could be extended from text-based materials to include text, graphic, audio, and video materials.

Developments in digital audio technology also seemed to offer valuable capabilities. For digital audio, a stream of speech becomes a sequence of numbers through the process of digitization. For playback, the sequence of numbers is transformed back into a stream of speech. With an appropriate digitizing process, speech can be coded to an arbitrary level of quality and it will not deteriorate because the numbers do not change. While digital, the speech does not pick up any hiss or static from power lines or other sources. In contrast to audio tape use, there is no need to worry about sound degradation due to dirty record or playback heads, tapes being wound too tightly, or tapes stretching with repeated usage. These are some of the reasons for the current popularity of audio Compact Discs (CDs).

For our purposes, once the sound is digital, it is easy to start and stop the speech at arbitrary points any number of times with a high degree of
precision—accuracy to the nearest hundredth or thousandth of a second is routine. Once digital, the speech can be analyzed (e.g., for pitch or tonal qualities), stored on computer storage media (e.g., floppy disks, hard disks, optical disks, or digital tape), and moved from one location to another via computer networks. This flexibility to control, analyze, store, and move digital audio, without degradation, made it a very appealing choice.

At the time Local Area Networks (LANS) appeared to offer another valuable technology. Essentially, a LAN is a modality-independent information conduit. It allows systems to communicate by sending digital information from one system to another. The digital information can represent almost anything: computer programs, stock market quotes, digital audio, etc. At the time most LANs were beginning to evolve into more powerful information conduits, consisting of personal computers connected to one or more file servers. The personal computers offered distributed computing power while the file servers presented a means of sharing resources including a common repository for programs and data. This offered an economical way to provide expensive resources to many computer stations. With a file server, we would only need disk space for one copy of any program (with an appropriate license arrangement) or any digital speech we would use, not for a dozen copies. The possible convenience as well as economic and maintenance advantages of a LAN made it a natural choice for CFLAC.

The future with respect to video was not as easy to divine. Digital video was on the horizon, but exactly which version of digital video to select and when it would become affordable were difficult questions to answer. We therefore chose a middle course: prepare both for a digital future and meet our immediate needs. In preparation, we selected computer systems with some excess capacity and with expansion capability which we expect will allow us to accommodate digital video when it becomes practical. For the present, we selected videotape as well as videodisc players. Since the range of videodisc materials available in the foreign languages is still limited, we selected interfaces for the computer control of both videodisc and videotape players.

With these goals and these views of some emerging technologies, we selected the specific components for our system (see Appendix for details). The system consists of networked personal computers with digital audio adapters. The use of a digital audio adapter allows for the flexible manipulation of audio materials: natural sounding audio may be played, recorded, replayed, paused, ordered, and re-ordered under computer control with no distracting delays or distortion. In addition, audio can be coordinated with the display of text or graphics on the computer screen, allowing for many imaginative applications which would be virtually impossible without digital speech and a controlling computer. Finally, the network allows for the central storage of the digital audio
materials and computer programs. Only a single copy of any audio material needs to be produced and stored on a network file server. This material is then available for selection at any time for independent use by any number of students. Individual students at separate workstations can use the same or different material, in the same or different ways, simultaneously, or at their own pace, without having to leave their workstation or wait for another student to finish.

Software Examples

A variety of programs have been and are being developed at Smith College for use with this system. All have been designed with options allowing them to be adapted to the individual teaching style of the instructor and learning style of the student. All programs also have been developed in a modular fashion to allow the easy combination of features from different programs.

Reading Program

The first program is a graduated reading comprehension program which was designed as a first step toward a reading tutor. It also can be viewed as an expansion of a traditional language lab function to include integrated text. In an ideal world, each student would have a reading tutor who could speak the foreign language fluently, could highlight the correspondence between the written and spoken language, could point out the differences between the student's and native's enunciation, and would be infinitely patient. While the initial version of the reading program cannot point out the student's enunciation errors, it does provide each of the other capabilities of a reading tutor.

The reading program first displays a passage on the computer screen which students read and record. The computer digitizes the speech and stores it on a network file server. When the students finish speaking, the program speaks the passage, with each word highlighted as it is spoken. This spoken version of the passage can be created by the instructor or taken from a pre-recorded audio source. Next, students listen to their own production of the passage. At this point, they may repeat the comparison between the standard version and their own production or re-record the passage and practice those portions which present difficulties.

Beyond its applications in pronunciation exercises or preparation of material on a specific topic for a free in-class discussion, the reading program can be used in preparation for literature classes to help teach dramatic or poetic recitation. From the student perspective, lab preparation of this sort allows a privacy and an individual attention that is not always possible in the classroom.
Dictation (Listening) Program

The dictation program was conceived in response to two basic questions which arose while correcting a stack of traditional dictation exercises: First, do the students really read these corrections? Second, and more important, even if they do read them conscientiously, do these red marks really help them to understand that which they failed to hear or comprehend? Practically speaking, students seldom receive a corrected lab dictation back the same day. Even the most highly motivated students experience difficulty associating the written corrections with the sounds they had mistakenly thought they had heard the day before. If the instructor simply marks the errors and sends students back to the lab for a second listen, they may correct careless mistakes but would most likely still have the same difficulties associating sound and word when the final corrections were returned. With the potential for precise integration which digital audio provides, the pre-correction marking and final correction steps can be moved into the lab and provide students with meaningful feedback at the moment of learning.

With this rationale as a guide, the dictation program was designed along the following steps, most of which were chosen to follow the standard pattern of pre-recorded lab dictations. First, the program speaks the passage at a natural rate and with natural intonation while students listen for global comprehension. Next, the passage is spoken with pronounced pauses at phrase boundaries, the duration of which can be controlled by the faculty member creating the exercise and varied from student to student, if so desired. When the second reading begins, students enter their draft transcription of the exercise in a word processor. Finally, the passage is spoken one more time at natural speed while the students correct and complete their transcription. When they are finished and have checked for spelling or grammar mistakes, they submit their text and the computer provides feedback about its accuracy.

In correction mode, the program immediately speaks and simultaneously marks those sources of error in the student responses on one line while providing the correct response—again simultaneously with the spoken word—on a line below. For the instructor who wants students to see where their errors were and attempt to correct them themselves, the feedback can be programmed to identify errors only and afford them a further chance to hear and correct those passages with mistakes before beginning the final correction sequence. Finally, to reestablish both the context of the words or phrases within the sentence and the sentence, within a context of several, the passage can be spoken one last time with natural intonation.

Mark Program

The purpose of this program is to allow authors to describe audio materials. To coordinate the presentation of text and speech, the program enables
the author to mark points in a stream of speech. This allows the attachment of
text to the speech or the insertion of silence and pauses at appropriate sentence
or phrase (or word or syllable) boundaries at the author's discretion. Note that by
marking the speech stream, we do not require that the text be presented or that
pauses be inserted. The marking simply indicates those points along the speech
stream where such action might be appropriate.

To attach text, the author marks the beginning and end of the speech
segment and then types in the associated text. For the author's convenience, the
speech may be played back at a slower rate to allow more precise placement of
the marks. To allow fine tuning of the marks, the mark can be moved forward or
backward in time by 50 milliseconds with a single key press and replayed at will.
Further, the position of the mark is displayed numerically so that the author can
make direct adjustments to the nearest millisecond if desired. By default, the end
of one segment is assumed to be the beginning of the next segment although the
author may vary this setting as described above. Boundary marks for possible
silence insertion are marked in a similar way.

This program is all that is required to create materials for the dictation and
reading programs. Once marked, the materials can be used with either program,
without change or augmentation. This is very important; for it allows a standard
set of materials to be created and then reused with a variety of programs.
Essentially, this provides reusable data to go with reusable programs. Just as the
reading program can be used with different (marked) audio materials, the same
(marked) audio materials can be used with different programs.

Dictionary Program

One of the greatest advantages of the principles we have used in
developing our software is the ability it affords us to "mix and match" programs.
Considering the time invested in creating software, the reusability of the
structure, as well as the content of such programs, is essential. It multiplies the
value of time spent developing computer programs and materials. Like the mark
program, the dictionary program provides a case in point. It can be linked as a
"help" utility to a basic foreign language word processor, to the dictation
program, to the reading program, and to the control programs used with
interactive videodisc and tape. The time spent developing a single vocabulary
can be used with multiple instructional programs.

Our dictionary program owes its inception in part to Système-D, a
commercially available software writing tool for French. Having seen and used
Système-D, we wanted a system which provided similar functions but, in
addition had the following features: an open architecture which could be
modified by the instructor to incorporate additional vocabulary, a shell which
could be used with other languages, and a module which could be integrated
with the other programs we were developing.
The dictionary program is a shell designed to support many different foreign languages on different levels. Currently, we have created a reasonably complete first-year German vocabulary as well as modest French and Spanish entries. Once the basic rules for declension and conjugation have been set, it is relatively easy to expand the vocabulary to support more advanced materials, such as the transcript for a video clip of native German speakers we have prepared.

To access the shell, students press a help key which opens the dictionary windows. These windows appear below the window containing the material the students are working on so as not to obscure their view of the lesson at hand. Within the dictionary windows, they can either enter the word they need or go to an alphabetized scroll of the vocabulary. By pressing any key they can advance the scroll to that letter or symbol and can press additional keys to advance the scroll further and zero in on the word of interest. Upon selecting a word, a window will open to identify the part of speech of the entry and, in the case of nouns, to give the article and plural form. By pressing "D," students can find the definition, which can be programmed either as an English or foreign language entry. With verbs, the program provides the principal parts (infinitive, imperfect, and perfect forms) and with an additional keystroke, the full conjugation of the present or past tense. While the release of several commercial on-line foreign language dictionaries may meet the needs of advanced language students, not all these include the quick-check features for nouns and verbs described above. We regard the dictionary program as a useful utility through the intermediate level and as an easy-to-program key-word dictionary for use with interactive video Clips.

Summary

From a purely pedagogical point of view, what we have described here may not appear very novel. However, we have purposely chosen to take an incremental approach to the incorporation of technology into instruction. We are building a system capable of performing traditional tasks in innovative ways, while at the same time providing a creative environment where the technology supports rather than determines the direction of the pedagogy. No one would suggest that the reading program or the dictation program could replace an individual tutor. However, few would argue that the current way we actually do, and must, teach reading and listening skills to large groups of students, represents an optimal solution either. The improvements which the programs offer come about by letting the computer do what it does best: to allow students a degree of independence, to control the flow of information in multiple modalities, to be infinitely patient, and to provide prompt and thorough feedback.
Our incremental approach has two related advantages. First, it allows us to use standard (some might say out-dated, others might say proven) methods in new ways. Familiar methods encourage teachers to use the new technologies without expecting to have to change completely the way they teach. They can examine how well the technologies work for them, for their students, at their institution, and consider how else they might make use of them. Knowing the value of dictation exercises, the step up to the dictation program is small and easy to justify. Having discovered the value of the dictation program but wishing for more cultural context, the step up to a video dictation program also is small. Can task-based and pedagogically guided interactive video be far behind? Second, the incremental approach allows the people actually using the new technologies to determine their most appropriate use. The technologies, programs, and materials are developed in the directions indicated by the pedagogy and are proven by extensive use in courses under the close scrutiny of the teacher. If the new technologies lead a revolution in teaching, as with most revolutions, it will come one small step at a time—by a large number of individuals.

Appendix: Hardware

The Center for Foreign Languages and Cultures consists of two distinct facilities equipped with different hardware. This indicates some of the ways in which our software efforts are independent of the hardware decisions. The CFLAC Electronic Classroom consists of 17 personal computers networked to a central file server. The personal computers are IBM PS/2 Model 30’s with dual floppy disks (720KB Model 30’s each) and 640KB of memory. Because the storage will be provided by the file server and most of the work will be done by the digital speech board, a low end 8 MHz 8086-based personal computer was considered adequate. The systems are connected via a 10 Megabit Ethernet using Western Digital EtherCard Plus adapters. This provides the high data throughput required for digital speech, while imposing relatively low overhead on the personal computers. The digital speech adapter we finally selected was IBM’s Audio Capture and Playback Adapter. This adapter provides two channels of 16 bit sampling and playback at up to 88.2 KHz, on-board selection of line-level or microphone input, and line-level or headphone output, and contains a 10 million instruction per second TI 32025 coprocessor. This is substantially more computing power than provided by the host personal computer and allows a quality equivalent to music CDs, with adequate power for future expansion. While this represents more capability than required, it does allow us a great deal of flexibility, and the board is very reasonably priced. Finally, each system has a Mitsubishi 1381A "multi-scanning" color monitor. Because of the multi-scanning feature it can be used with almost any computer. More important, it also allows the display of standard NTSC video. While it does not allow for the simultaneous display of video and computer output (without special hardware),
The other half of CFLAC is the lab facility proper. This Fall it will initially house 23 student workstations but it is designed to expand to 61. These stations are IBM PS/2 Model 55s with one floppy drive (1.44MB), one hard disk (30MB), and 2 MB of memory. This is a medium to high-end personal computer based on the Intel 80386SX chip. We chose a more powerful system for this installation in order to meet our future video, graphic, and artificial intelligence needs. These stations are connected by a 16 Megabit Token Ring, using IBM’s Token Ring adapters. While the Ethernet is more than adequate for the classroom facility, we had some concerns about its capacity for handling the larger lab facility. As Ethernet approaches full utilization, its performance can begin to deteriorate. Because Ethernet (unlike Token Ring) does not guarantee delivery within any specific time frame, this could result in long delays in the reception of information for some workstations. Students would notice this as pauses randomly placed in the speech stream. While there are possible ways to remedy these unwanted pauses, the Token Ring allowed a much simpler solution. The workstations also have the Microchannel version of the IBM Audio Capture and Playback adapter and the Mitsubishi monitor. In addition, some of the workstations are equipped with Pioneer LDV-4200 videodisc players, while others house Panasonic AG-6100 videotape players. The videodisc players can be controlled by the built-in RS-232 port of the computers while the videotape players can be controlled by BCD’s Videolink 232 interface.

A single file server is shared by the two facilities although we anticipate expanding to two shared file servers at a future date. The file server is an IBM PS/2 Model 80 with over 2.5 Gigabytes of hard disk storage and 8 Megabytes of memory. We selected the amount of storage after estimating the amount of speech that would need to be available on line at any point in time. First, we estimated that, after removing all of the long periods of silence and ignoring copies of sentences in different orders, we had approximately 300 hours of actual, actively used, foreign language speech on tapes. Second, we estimated that we needed about four week’s worth of speech on line at any time. This reduces the amount of on-line speech to 50 hours or 180,000 seconds. Finally, we estimated that, given the power of the digital speech adapter, we would be able to compress one second of speech to about 12 kilobytes. This results in an estimate of 2.21 gigabytes required within our total of over 2.5 gigabytes of storage space.

Currently, the file server is running Novell NetWare 2.11, which we anticipate upgrading to release 3.1 when it becomes available next year. The server contains an Ethernet adapter for connection with the classroom facility, a Token Ring adapter for connection with the lab facility, and a third network adapter for connection with our campus backbone network. All speech is stored...
in files saved on the server, as are all computer programs. Backup is provided via the campus backbone to a 2.2 gigabyte 8mm video cassette device.

Notes


3 The markup program is designed to provide descriptive and referential markup in the spirit of the Standard Generalized Markup Language (SGML) which provides a standard notation and syntax for marking text. While our program intends to use the principles of SGML, it hides the specifics of the markup notation from the author. The author does not need to know anything about SGML in order to create marked audio materials. See James H. Coombs, A. Renear, and S. DeRose, “Markup systems and the future of scholarly text processing,” *Communications of the ACM*, 30 (11), 933-947, for a discussion of the benefits of SGML.


5 When there is an SGML format for dictionary entries, we will modify the dictionary program so that it can read them. This should make it compatible with many commercial on-line dictionaries.

Authors' Biodata

Richard Elias Parker is Director of Software Acquisition and Development for the Center for Academic Computing at Smith College. He holds a Ph.D. in Cognitive Psychology from the University of California, San Diego.

Robert Chapin Davis is Director of the Center for Foreign Languages and Cultures and member of the German Department at Smith College. He received a Ph.D. from Stanford University in German Studies and English Literature.

Authors' Addresses

Richard E. Parker  
Center for Academic Computing  
Smith College  
Northampton, MA 01063  
(413) 585-3079

Robert C. Davis  
Center for Foreign Languages and Cultures  
Smith College  
Northampton, MA 01063  
(413) 585-3470