THE AUSTRALIAN JOURNAL OF
EDUCATIONAL TECHNOLOGY

Creating and using digital audio files under the Windows operating environment

Larry Nelson
Curtin University of Technology

Macintosh and Amiga users have had good sound playback facilities for years; recent models of the Mac have recording capabilities as well. While sound boards for MS DOS PCs have been around for quite some time, audio formatting standards have been nonexistent, and memory restrictions have placed severe limits on the amount of sound which can be recorded. The advent of Windows has changed things. This paper suggests that PCs with Windows are capable of serious audio work, and mentions factors to be aware of before setting forth on sound ventures with a PC.

This paper discusses experiences gained in the process of creating a series of language learning software modules for the Department of Employment, Education and Training.

The DEET-funded 'Talk Project' at Curtin University of Technology was instigated in order to determine the feasibility of developing cost-effective audio-based software for beginning students of Burmese, Indonesian, and Spanish[1]. The software was to be applied on personal computers capable of running the Microsoft Windows operating environment. It was to operate on what is now a garden variety PC, one which had a VGA colour monitor, sufficient random access memory (RAM) to run Windows, and a small hard disk. The only extra accessory permitted was to be a sound card compatible with Windows. In late 1992 such a machine could be purchased for a tax free price of around two thousand dollars (Australian).

Approximately a dozen interactive lesson modules had been produced at the time of writing. Most of them offered two screens for a user to interact with. One screen would have two lists of words or phrases, one list for English, the other in one of the three languages of concern to the project. A user could hide one of the lists, and test her or his vocabulary and pronunciation skills by clicking on individual words or phrases, hearing them pronounced by both male and female native speakers. On clicking a

'Record' button, the user could speak the word or phrase into a microphone, and then playback his or her voice for comparison with the voices of native speakers.

The second lesson screen provided basically the same skills practice, but used a set of game modes to challenge the user at time intervals of selectable duration. For example, the 'Faces' lesson randomly selects the name of a head part, such as ear, mouth, nose, etc., plays the voice of a native speaker pronouncing the name of the part, and gives the user so many seconds to point to the 'right answer'.

The 'Talk' lessons were highly similar to those Nelson (1992) developed on an Amiga computer, and parallel to language exercises which Rehn (1992) has experimented with on Macintosh hardware.

Both the Amiga and Macintosh families of personal computers have had in-built audio playback facilities since their inception, at least on most models. In recent years Apple has added built-in recording capabilities to Macs. The audio 'picture' on DOS-based PCs, however, did not have a clear focus until the release of Windows, which brought much needed format standards to such machines.

The appearance of a plethora of 'multimedia' articles and books has resulted in an outbreak of new Windows capable sound boards for the PC. We have tried several of these, and have found that not all were created equal, as it were, despite the standards promulgated by Windows and the so called 'multimedia PC'.

We detail some of our discoveries below. Before getting into results, it must be mentioned that the systems used were purchased at different times, and in different places: Sound Blaster Pro was obtained in January, 1992 in Western Australia; the Pro AudioSpectrum was purchased in July, 1992 in the United States; and the Windows Sound System was delivered from a Perth based supplier in January, 1993. It is likely that the latest versions of these systems differ to the ones used in the present study; some of the limitations discussed below may well have disappeared by the time readers take up this article. Nonetheless, it is felt that a sufficient number of generalities may come through in the following discussion to make readers aware of issues to take heed of in the audio digitising process.

Memory resources required by digital audio

As an example of matters which arise when creating digital voice clips, we relate here some of the procedures followed in setting up the Talk lesson known as 'alphabet'.

Input to the process: a tape recording of a native Indonesian speaker reciting the alphabet. The Indonesian alphabet is the same as that used by English, having 26 letters. It took the native speaker a total of 47 seconds to recite the letters at a predetermined pace. The native speaker's voice was captured using off the shelf tape recording components, such as those found in Radio Shack and Dick Smith retail outlets.

The initial objective was to make a complete digital copy of the entire recording, and then use a waveform editor to pick out the individual letters of the alphabet, saving them as separate audio clips. Since the audio clips were to be used under the Windows operating environment, they needed to be saved in the 'WAV' format. This is a standard Microsoft Resource Interchange File Format (Microsoft Corporation, 1991; also see Sheldon, 1992); under Windows, the format supports 8 and 16 bit resolution at three sampling frequencies: 11.025 kHz, 22.05 kHz, and 44.1 kHz.

Many readers may be aware of the considerable memory resources needed to digitise audio using present technology. A monaural digitised version of the 47 second source tape would require 518 kilobytes of storage if recorded with a resolution of 8 bits at 11.025 kHz, 1.04 megabytes at 22.05 kHz, and 2.07 megabytes at 44.1 kHz. Increasing the resolution to 16 bits would double each of these figures; using stereo would double them again.

Two veteran and one new audio systems were used to digitise the source tape. The seasoned systems were Sound Blaster Pro from Creative Systems, and Pro AudioSpectrum from Media Vision. The new system was the Windows Sound System from Microsoft. The Windows Sound System is identical to the Business Audio resources which come as standard equipment in recent versions of Compaq DeskPro computers. (Certain Compaq computers are now audio-ready, coming equipped for immediate sound recording and playback without the need for any additional hardware or software.)

After the source tape was digitised, a variety of waveform editors were applied to the digital copy in order to extract the 26 distinct letters of the alphabet.

Digitising from tape or microphone

It is not necessary to use a tape as the audio source for digitising; one can use a microphone directly. In fact, some systems, such as the MicroKey AudioPort from Video Associates Labs, have only a microphone input jack. Level attenuators can be used with such systems to permit tape input, and are readily available at electronics hobby stores.

Direct use of a microphone for frequent audio digitising is, in the end, not yet as convenient and parsimonious as is using a tape recorder, at least not for what might be termed 'production' work.

The most significant problem encountered in using a microphone as input, instead of a tape, relates to computer memory requirements. Tapes come in standard, convenient lengths of 15, 30, and 45 minutes. However, corresponding pre-set lengths for digitised recordings do not exist. In fact, recordings which exceed 15 seconds (or thereabouts) do not always come through well on a personal computer. Depending on the resolution and sampling frequency selected, 15 seconds of recording can exceed a computer's RAM capacity, and require that the recording be buffered out to a hard disk.

Some digitising software, such as the professional Wave for Windows suite of programs from Turtle Beach Systems, use disk buffering as the standard procedure - no attempt is made to detect and use any extended memory which might be present on the computer. Hard disk access is always slower than RAM, and a system such as Wave for Windows can lose data when higher sampling frequencies are used.

A problem encountered with regularity during the project related to the fact that much of the software available was descended from the days when PCs had at most 640 kilobytes of working RAM. Some software, such as that which came with the Sound Blaster Pro system involved in the study, predeternines how much RAM is available for digitising, and automatically cuts out when this limit is exceeded. This would not be a bad approach at all, were it not for the fact that the software used was unable to see beyond the old 640 kilobyte DOS working RAM limit. Sound Blaster's settings allow the recording to be buffered out to disk if there is insufficient RAM, but once this option is selected one can run the risk of data loss if the hard disk is not fast enough to keep up with the sampling frequency and recording resolution selected.

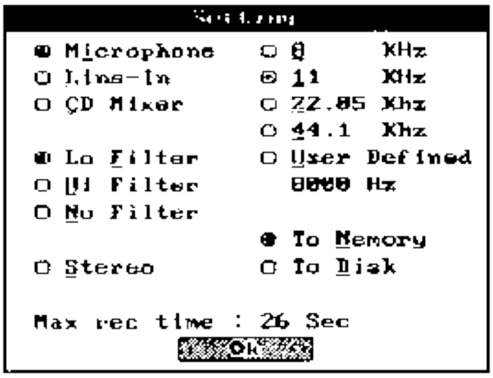
The end result of these limitations is that digitising directly from a microphone has a few unknowns attached to it. The recording can be abruptly terminated by the software, or data loss can occur when using a hard disk incapable of capturing all the data being sent it.

True, these limitations are also present when one uses a tape as input. But, in a production setting, tape recording a speaker is less likely to be disturbed than is direct digitising with a computer. Better to do the digitising off line, when the speakers have finished talking to a tape recorder. If the digitising from tape process is interrupted by the computer, the tape can be restarted with more ease than most speakers, and at less expense to the production budget.

A final factor to mention here is that of archiving. Digital audio disk files are easier to use than analogue tape files, a fact which derives from the sequential access nature of tape files, and the relative imprecision associated with indexing a tape file's start position. Digital files are eminently easier to access and archive, but, and it is a big but indeed, they can require massive magnetic disk resources. A 15 minute monaural audio tape would require 20 megabytes of disk storage if digitised with a sampling rate of 22.05 kHz, and 8 bit resolution, and even then one would end up with voice quality fidelity, not music quality.

Selecting options for digitising

The most professional of audio digitisers have settings options such as those in Sound Blaster Pro, as shown below:



```
● Microphone      ○ 8      KHz
○ Line-In         ● 11     KHz
○ CD Mixer        ○ 22.05  Khz
                  ○ 44.1   Khz
● Lo Filter       ○ User Defined
○ Hi Filter         8000 Hz
○ No Filter
                  ● To Memory
○ Stereo          ○ To Disk

Max rec time : 26 Sec
```

The Sound Blaster Pro card used in the study did not have a resolution option; those systems which permit audio capture at both 8 and 16 bit resolution will have one more option to set. In the present study 8 bit resolution was entirely adequate; 16 bit resolution provides for better audio fidelity, but for voice recordings. be they male or female, 8 bit resolution is entirely sufficient, and saves memory usage.

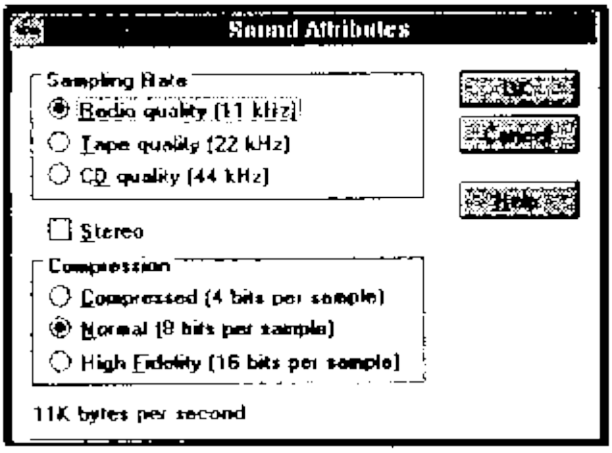
One thing to note in the settings box above is the presence of an 8 kHz sampling frequency. This option, if selected, will result in a file which Windows cannot use. Sound Blaster Pro will be happy with it, but not Windows, which insists on seeing 11.025, 22.05, or 44.1 kHz.

The Max rec time: 26 Sec shown above indicates that, with the settings as selected, a maximum of 26 seconds of digitising can take place before the computer runs out of RAM. Had the Sound Blaster software been able to peek beyond conventional DOS memory, it would have seen that a few megabytes of extended RAM were unoccupied, and the 26 second limit would have increased substantially.

We made some of our recordings at a sampling frequency of 11.025 kHz, and others at 22.05 kHz. Selecting the higher frequency in the settings above would knock the maximum record time down to a paltry 13 seconds. Under such conditions, digitising our 47 second source tape would take four steps, instead of the two required using the lower sampling frequency.

If a Windows capable system were used for the digitising, it should be able to automatically use any free extended memory available. The Windows Sound System worked in this maimer; we used it on a Compaq 386 machine with four megabytes of RAM, and also on a Compaq 486 having an EISA bus, and eight megabytes of RAM. On both machines the 47 second source recording was digitised in a single step.

Recording options are set in this system by using the dialogue box shown below:



Sound Attributes

Sampling Rate
● Radio quality (11 kHz)
○ Tape quality (22 kHz)
○ CD quality (44 kHz)

□ Stereo

Compression
○ Compressed (4 bits per sample)
● Normal (8 bits per sample)
○ High Fidelity (16 bits per sample)

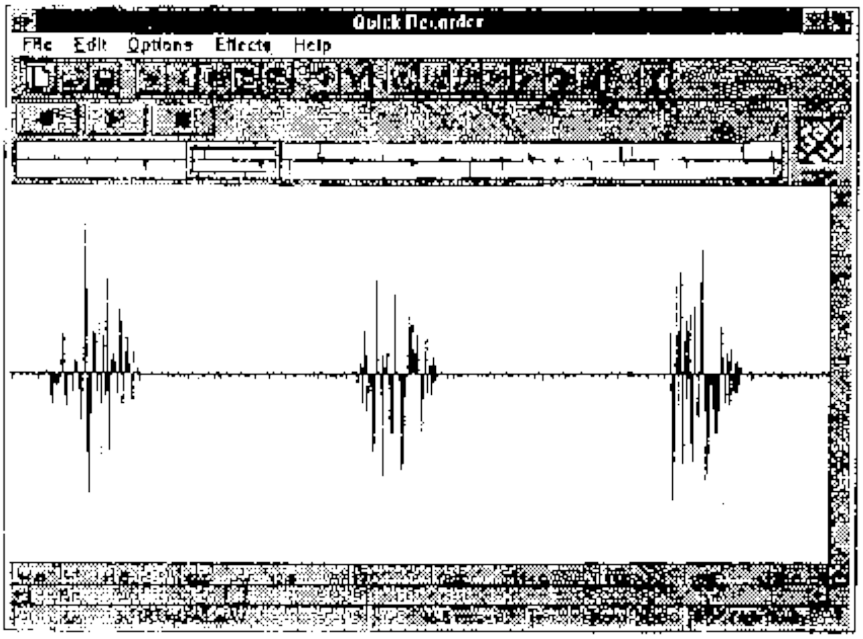
11K bytes per second

On both machines, Windows Sound System was content to accept whatever options we chose, and readily indicated the number of kilobytes of memory which would be required to digitise each second of audio. However, the system provided no indication of how much total time would be open for recording. Tests showed that the total time available depended not only on RAM, but also on the amount of free hard disk space at hand. As a practical operating guideline, a computer with four megabytes of RAM will permit about ten minutes of recording at 11.025 kHz, 8 bit resolution, and half that if sampling is increased to 22.05 kHz. These figures will be reduced if hard disk space is tight. We found that the small Pocket Recorder program from Media Vision could be used to provide a good index of maximum record time, when and if this figure was needed.

The Windows Sound System has an option for compressing audio clips, which, according to the manual, would be equivalent to recording with a 4 bit resolution. We found this compression scheme to be useable with voice recordings, but at an easily detectable degradation of fidelity.

It is the case that many manufacturers of audio boards provide some facilities for compression. All of these are meant to make it possible to squeeze more sound into limited computer memory and disk space. They do this, but audio fidelity is not the only sacrifice made if these schemes are applied one also loses portability. At the present time, a 4 bit recording made with Windows Sound System, for example, will only be playable on a computer equipped with the same sound hardware. It will not play back under Windows running a Sound Blaster or Media Vision sound driver.

If portability is a concern, users must stick to uncompressed audio files recorded at 11.025, 22.05, or 44.1 kHz. A resolution of eight bits is also best as the number of 16 bit sound cards is still limited.



Dissecting the digitised recording

The screen shot above was captured from Quick Recorder, a Windows Sound System waveform processor and editor.

The oscilloscope display in the main part of the screen has three obvious voice patterns, or breath groups. Each corresponds to a single letter of the alphabet, pronounced by an Indonesian speaker.

The smaller display above the main one shows the entire waveform, containing 26 small blips. Within this display bar, a frame indicates where the three letters in the main area are located with respect to the overall waveform. In this case the frame is around the letters G, H, and I.

We needed to make single letter voice clips given this waveform as input. The waveform above could be considered to be a long sentence with 26 words; we wanted to be able to highlight each 'word' and save it to disk as a separate file.

Not all waveform editors will do this. Portions of a wave can be highlighted, and options applied, but these options tend to perform various actions on the selection[2], and 'Save as ...' is not usually one of them. The *Creative Voice Editor* Ver. 2.08, which arrived with our Sound Blaster Pro package was the only editor we tested which had such a capability.

Some waveform editors, such as the Sound Recorder which comes as a standard utility with Windows, and the Pocket Recorder from Media Vision, can chop a waveform into two sections, and save just one of the sections, but this is not at all the same as being able to save what might be a small highlighted section in the middle of the waveform.

Other editors, such as that provided under the 'expanded view' option of the Quick Recorder program in the Windows Sound System, and that found with the Turtle Beach Wave for Windows software, allow any contiguous highlighted section of a waveform to be copied or cut to a clipboard, from where it can be pasted into another wave editing window. When in a new window it can be saved. This two step process is not as convenient as that found in the Sound Blaster Pro's voice editor, but it does have an important advantage: the properties of the highlighted section, such as the sampling frequency, can be changed before the section is pasted. We used this at times, taking 'words' from a 22 kHz waveform and halving the sampling rate before saving. This was done whenever we felt 11 kHz sampling provided playback quality adequate for our application's requirements, or when one of the lessons had more than twenty voice clips and disk space was at a premium. There were occasions when we trialed 11 kHz clips in the field, and later had to go back to the original 22 kHz format after it became apparent that fidelity was not satisfactory.

Factors to consider in selecting a sound card

There are indeed a variety of sound cards on the market now, and most of them claim compatibility with Microsoft's Windows operating environment.

Perhaps the two most important questions to consider when selecting a card would be: Is DOS compatibility important?, and, Will the card be used for recording as well as playback?

Two of the latest systems, the Microsoft Windows Sound System and Compaq's Business Audio, will generally not play audio from programs running under DOS.

If the card is to be used exclusively under Windows, a system specifically designed for Windows functioning, such as the two mentioned in the preceding paragraph, presents substantial advantages. These derive from two principal factors: systems not originally designed for Windows operation must often be run under DOS in order to gain access to all of their features, and, in the case of the Sound Blaster Pro and MicroKey AudioPort, produce files which must be converted to the WAV format by another program before Windows will use them.

The other factor results from the ability of most of the Windows specific systems to work with extended memory. This presents very real benefits, making it possible to record and edit audio files which might be several minutes long instead of just a few seconds[3]. But a caveat: we found one Windows system, Turtle Beach's Wave for Windows, which did not use extended memory efficiently.

Our tests of two veterans from the DOS era, the Sound Blaster Pro and Media Vision Pro AudioSpectrum 16, revealed some important limitations when applied in our Talk project. Both systems suffer from an inability to use extended memory. The Windows driver for the Sound Blaster Pro would not permit recording above 11 kHz; when the card runs under DOS this limitation disappears. The Media Vision system had mediocre documentation, and under Windows would not automatically mute its audio output when poked into record mode. This made the card unusable for Talk lessons; if a Talk user clicks on the 'Record' button on a machine running a Pro AudioSpectrum 16 card, the feedback can be deafening.

Some of these limitations may have vanished by the time this paper is read. The lack of automatic record muting has been fixed in Media Vision's new Audioport sound device, and one would think it will also soon be fixed on the rest of the Media Vision range, if it has not been already.

Using a microphone for voice recording produced quite variable results, depending on the microphone and sound card used. Of the systems we tested, the Microsoft Windows Sound System and Compaq's Business Audio produced the best quality audio. These systems come with their own microphone; the Microsoft system also comes with a pair of headphones.

Controlling volume levels with some cards can be inconvenient. The Sound Blaster Pro, the MicroKey AudioPort, and the Media Vision Audioport have knobs on them much like a small radio, and permit playback levels to be controlled easily. They use automatic gain circuitry when recording, which usually worked well enough.

Many other systems, however, use software to control playback and record levels. These are, generally, not convenient to use: we solved the playback part of the problem by using headphones and small amplifiers with volume knobs. It is also helpful to be able to control bass and treble levels on playback; the process of recording and playing back digital tracks is subject to interference from the very computer on which they are running - a treble control, in particular, can be invaluable.

The Microsoft Windows Sound System, and Compaq's Business Audio, include a utility for managing digital audio files, the 'Sound Finder'. We found it to be very useful as it permitted us to thoroughly label the hundreds of short voice clips used in our project, and to edit or play them back with ease.

Concluding comments

Our Windows audio work followed on from substantial prior experience with Amiga computers, as well as some experimentation with Macintosh audio. The author's original impression on creating, editing, and using audio under Windows was not exactly one of great enthusiasm. The software tools seemed limited, and the quality of the audio was initially felt to be inferior to that which could be achieved on other boxes.

Compaq's Business Audio system and, a bit later, the Microsoft Windows Sound System, did much to turn the picture around. These systems are carefully tuned to Windows, and use a microphone very closely coupled to the capabilities of their recording circuitry.

Digital audio on any current microcomputer platform requires considerable resources. Windows has added two crucial factors to the audio equation on Intel based machines: standard procedures for accessing extended RAM, and file format standards. Providing one is using a machine with sufficient RAM and hard disk space, our conclusion is that the current state of the art is adequate for producing reasonable quality digital audio files under Windows, at least for monaural voice recordings.

Notes

1. I gratefully acknowledge the assistance of several colleagues in completing the work described here, particularly that of Piet Herman Abik, Nur Hadi Amiyanto, and Brian Lawton.
2. These actions can be numerous, depending on the system selected. They commonly include volume and pitch controls, fading, mixing, echo effects, frequency filtering, muting, and automatic trimming of quiet spots.
3. A machine with a 486 processor is recommended when editing sound clips of more than half a minute duration; a 386 will handle them, but for frequent work the speed of the 486 is welcome.

References

Microsoft (1991). *Microsoft Windows Multimedia Programmer's Reference*. Redmond, Washington: Microsoft Press.

Nelson, L. R. (1992). Developing interactive digitised audio courseware on Amiga, Macintosh, and PC platforms: A comparison of common support facilities available. In *Proceedings of the International Interactive Multimedia Symposium*. Perth: Promaco Conventions Pty Ltd. http://www.ascilite.org.au/aset-archives/confs/iims/1992/nelson.html

Rehn, G. (1992). Two-way interactive sound on a stand-alone Macintosh platform. *Australian Journal of Educational Technology*, 8, 51-64. http://www.ascilite.org.au/ajet/ajet8/rehn.html

Sheldon, T. (1992). *Windows 3.1, the Complete Reference*. Sydney: Osborne McGraw-Hill.

Author: Larry Nelson is a senior lecturer at Curtin University's Faculty of Education, where he lectures in research procedures and computer applications in education. The work described here was undertaken as part of a DEET 'ILOTES' (Innovative Languages Other than English) Project.

Please cite as: Nelson, L. (1993). Creating and using digital audio files under the Windows operating environment. *Australian Journal of Educational Technology*, 9(1), 1-11. http://www.ascilite.org.au/ajet/ajet9/nelson.html