DIGITAL AUDIO PROCESSOR INTERFACE

B.J. Stone, D. Mead, and M. O'Kane

School of Information Sciences and Engineering
Canberra College of Advanced Education

ABSTRACT - The paper describes a computer interface for a digital audio processor, or pulse-code modulation adaptor, which records and plays back digital audio signals on videocassettes. The total cost of the prototype interface (in Multibus form), with digital audio processor and videocassette unit, was below $A2000. Adoption of this all-digital technique for speech database storage and shipping by the Australasian speech research community is advocated.

INTRODUCTION

Overview

This paper has two aims. The first is to demonstrate the simplicity of interfacing a digital audio consumer device to small computers, to carry out the analogue-digital (A-to-D, D-to-A) conversions of a computer-based speech laboratory. The second aim is to canvass the possibility of the Australasian speech research community's adopting the resultant format for data storage and interchange.

A prototype Multibus circuit board has been built and tested, together with control programs. Together with a Sony PCM-501ES digital audio processor (DAP), it provides the following main features:

- input gain control;
- 11, 22, or 44 kHz sampling of one or two channels;
- 16-bit samples delivered serially as twos complement numbers;
- continuous storage of samples by direct memory access (DMA) up to limit of available buffer space;
- encoding of samples as simulated video for videocassette recording and playback (the DAP's normal function);
- DMA transfers between buffer memory and video recorder;
- DMA playback from buffer; and
- total cost of interface, DAP, and videocassette recorder around $A2000.

Standardising Record-Transfer-Playback Methods

Speech databases are traditionally recorded and stored on high quality audio magnetic tape. This medium is subject to print-through and other long term deterioration, and to loss of quality through copying. For computer analysis its contents must be filtered for anti-aliasing and digitised. High resolution A-to-D systems have generally been a substantial cost in establishing a speech laboratory. Some years ago two groups, an IEEE Working Group (Pallett and Baker, 1983) and the International Working Group on Vowel Normalisation (Bladon, de Graaf and O'Kane, 1984), suggested that databases be stored on PCM encoded videocassettes. This technique's error correction ensures that loss of signal is rare and can be detected. Also, videocassette is a compact medium for storage and mailing. Another attractive feature is its error-corrected copying facility. A detailed discussion of the advantages can be found in Pallett (1986). Several large databases have been collected using this format. See, for example, Guyote, Lewis and Lijana (1986) and Itahashi (1986).

Although PCM-encoded videocassette is a digital recording, very few speech laboratories can use the data in this form directly. Rather, they use the reconstructed analogue output of the PCM unit and re-digitise this signal on whatever A-to-D system is available in the laboratory. In this paper we show how to transfer the PCM digital data directly to and from small computers.

Serial Audio Technology

Most consumers who take videocassette recorders and compact discs (CDs) for granted are probably unaware of a close connection between the systems. When digital recordings on vinyl discs were being placed on sale and the CD technology was being prepared for market, video recorders were the only devices able to record the required bandwidth, about 2MHz, economically. The digital data were encoded by pulse code modulation into a simulated video waveform, close enough to the genuine video signal to be acceptable to an unmodified video recorder. This recording method is now packaged with the titles "Digital Audio Processors" or "Pulse Code Modulation Adaptors" at a low price for home use with domestic videocassette units, while even recorders for studio preparation of CDs differ in details but are unaltered in principle.

The sampling rate for CDs is usually 44.1 kHz. Although this figure is set approximately by the Nyquist criterion and normal human hearing, it is in fact fixed precisely by the timing of both NTSC and PAL video. The well known differences between these standards in frame frequency and scan-lines/frame leave them with almost identical line frequencies. They both display slightly more than 14700 visible lines/second. In each of these normally visible lines, three audio samples are encoded for each stereo channel, giving 44100 samples/second/channel.

Video tape is subject to signal drop-outs lasting for several scan lines. (Dirt particles on a CD could obscure the signal data for similar periods). To minimise the alarming audible effects which this might produce, samples are buffered before recording on videotape. This allows the samples for blocks of about 35 lines to be recorded, together with numerous error-correction bits, in an interleaved format. When drop-outs occur during playback, a combination of parity, error-correction, and interpolation between channels gives useful protection even if the signal from tape is lost for more than 11 lines.

For transferring the data to and from a computer, this scrambled encoding of the audio samples is unnecessary. If unavoidable, it would be a major nuisance. Fortunately, the modular design of at least one DAP allows the stream of digital samples to be intercepted at points in both the recording and playback data paths where it is free of these complications.

The individual samples are produced serially as twos complement 16-bit numbers, most significant (sign) bit first in time, by the successive approximation A-to-D converter in the input section of the DAP. They retain this format in the playback section. For interfacing purposes, it is necessary to connect only three DAP signals - word clock, bit clock, and serial data - plus direction control signals between the DAP and computer.

Target Computers

The Digital Audio Processor (DAP) interface prorotype was built as a Multibus card. Only one channel was required for recording speech. The memory and bus bandwidth needed at the full 44.1 kHz sampling rate, with two bytes/sample, was therefore 88 kbytes/second. This fairly low rate did not necessitate 16-bit transfers on the data bus. Hence for testing purposes the host computers were initially an Intel SBC-80/10 supported by an Intel MDS (8080 based) and then an Eracom ERA50 (8085 based), both using 8-bit data transfers, 8-bit port addresses, and 16-bit memory addresses.

Experience with the prototype and the hope of wide use of the technique both now indicate that the IBM PC bus is a good choice for further development.

TECHNICAL DESCRIPTION

Interface Operating Principles

The operation of the DAP interface when recording speech is easily understood by considering in turn the units along its input data path. These are:

- the data direction multiplexer (a small card added to the PCM-501ES);
- the shift register serial-parallel converter; and
- the computer bus interface, with its memory management unit and DMA controller.

From the DAP input terminals, the stereo signals traverse a short analogue path for gain and frequency control, ending in sample-and-hold stages at the input of the analogue-to-digital converter (which is shared between the channels alternately). The two-way switch to select channels in turn, and the A-to-D converter, form one large scale integrated circuit. Another package of similar complexity, with external buffer memory, manages sample interleaving, addition of error control words, and video generation. The interface's data direction multiplexer takes its serial digital input from between these two major circuit blocks. At this point, samples are free of all added information and appear at a constant rate.

Entering the interface, the serial bit stream of samples is formed into parallel data bytes in a shift register, with the aid of the word and bit clock signals and counting circuits. The interface writes each byte into the memory of its host computer, via the Multibus. One channel of a single-chip DMA controller generates the writing control signals and the 16 least significant memory address bits. When it reaches its 64k addressing limit, it restarts without program intervention. The bit and word clock signals control a data byte counter whose higher order address bits, together with 16 bits from the DMA controller, cover all the available memory.

For prototype testing a 0.5 Mbyte dynamic memory card has been used. The 8-bit microcomputers used in the tests can only address the lowest 64k locations, but this constraint does not affect transfers to secondary (disc or tape) storage. For such transfers, two channels of the DMA controller are employed in its "memory to memory" mode to copy 2 kbyte blocks between the high memory locations used by the DAP interface and the locations visible by the microprocessor. In the prototype, diskette transfers and sampling do not proceed simultaneously. However, in any computer, transfers may be made to a fast enough disk or tape by a small task which keeps a queue of these 2 kbyte blocks, calling the operating system to write or read each block in turn.

For one input signal (not stereo), at the full sampling rate and with the first 64k locations reserved for program, the 0.5 Mbyte data memory limits recording to about six seconds in the absence of simultaneous transfers to disk. At 11.025 kHz the duration is increased to about 25 seconds.

Playback operation from memory to analogue output is precisely the reverse of what happens during recording in memory. The same shift register is used. The PCM-501ES DAP is as symmetrical in its recording and playback halves as the interface (e.g., the stereo channels share one D-to-A converter for output, as they share one A-to-D converter for input).

For off-line storage of data, or transfer of data between sites, on videocassette, the data direction multiplexer board added to the DAP can establish a data path either from computer memory (as input) to the video output, or from the video input to the computer memory. Also, for testing during recording, the audio output available from the DAP for monitoring can be reconstituted from samples which have passed through the complete interface, giving increased confidence that the digital logic is working correctly. All switching of the data paths is done under program control.

The original DAP functions of recording on videocassette from microphone inputs, and playback from cassette, remain unaffected after addition of the interface.

Use of Computer Disk (or Tape) during Processing

Utterances which will fit into available computer memory, as in the prototype, are long enough for only some applications. For longer utterances, there must be one program task using the DAP and its interface to put samples into an area of memory, while another task reads samples out from this memory to a disk or tape. The sample rate and volume set constraints upon the secondary storage device, its interface, and the program. The following discussion of these requirements is in terms of recording in memory and on disk. Similar arguments apply to playback and also to any type of computer digital tape.

Diskettes are too small and too slow for use in this application. The size objection does not apply to even the smallest of hard disk drives. However, problems may still lie in the drive's speed, its controller, other tasks in the computer, or the design of the computer. Minimum needs are:

- the multi-track average (not instantaneous) disk transfer rate must exceed the sampling data rate, viz. 22 up to 176 kbytes/sec;
- the disk controller, itself a DMA device, must be commanded to obtain samples from memory addresses as soon as possible after the DMA control chip in the DAP interface has finished with them; and
- if any other tasks (e.g. timer interrupts for refreshing dynamic memory) have higher priority than the DAP interface, then the latter must include hardware buffering for more samples than will be produced by the DAP during any high priority task.

Project Status

Prototype hardware development has been done in 8-bit Multibus microcomputers, as described earlier. Some control choices such as selection of sampling rate are made by altering jumpers on the DAP interface, but can easily be made accessible to the program. Driver software for the interface, coded in Intel 8080 assembler, is called by a simple user command interpreter in UCSD Pascal.

To put the interface into service in our own laboratory, its address decoding is being extended to work in a Sun-2 workstation (also a Multibus machine).

Design of an IBM-PC interface is proceeding. The points being investigated at present are the possible need for hardware on the interface to buffer samples (while the PC's DMA controller uses its top priority channel to refresh dynamic RAM), and whether to retain 8-bit data transfers or adopt the PC-AT bus and use only 16-bit transfers.

Alternative Components

The underlying standard in this technique is that of a country's broadcast television system. This can be recorded on U-Matic, VHS, or Betamax video-cassettes or reel-to-reel videotape.

To be confident of successfully reading a PCM speech database shipped on any of these media at a new location, the recipient should use the same DAP model as the source laboratory. The cost of the DAP used in this project is the same as that of a home videocassette recorder, below $A1000. Interface details would have to be slightly different for the previous Sony model, the PCM-F1, but the PCM formats of the two models are apparently the same.

This paper's suggestion is that the interface is simple enough to be built in laboratories, or as one or two small production runs (say, for Multibus and IBM-PC computers), in Australia.

It must also be acknowledged that a similar interface, for the PCM-F1, has recently been advertised by a French company (Oros, 1986).

CONCLUSION

Advantages and Limitations

The most obvious advantages of the proposal are the small physical size and low cost of the equipment, the high maximum sampling rate, and the universal, cheap availability of the videotape media.

The two known disadvantages are more subtle. Further information about them is given by Watkinson (1986). Firstly, when colour was added to NTSC broadcast television, the frame frequency had to be slightly decreased from 60 Hz (to eliminate beats between harmonics of the colour and sound subcarriers). However, the PAL frame frequency is precisely 50 Hz. The result is that DAPs for use in PAL countries sample at 44.100 kHz, but there is a second standard of 44.056 kHz for use with NTSC videocassette recorders. This limits the interchangeability of data on videotape. Secondly, consumer DAPs which share D-to-A and A-to-D converters between stereo channels, like the Sony PCM-501ES, have an 11 usec delay between the two samples of each stereo pair unless correction is built in. This does not affect the consumer, and is unlikely to affect speech research, but does depart from the Audio Engineering Society-European Broadcasting Union digital interconnect standard which governs studio recording systems.

Prospects for General Adoption

For speech laboratories the advantages of interfacing a DAP to their existing computer systems appear, as detailed in the previous section, to overwhelm the limitations.

Perhaps even more important than low cost, high resolution, and error correction is the ability to install this storage and shipment system without displacing a laboratory's existing analogue-digital conversion equipment. The two can be used independently as needs arise, and data can be copied from one to the other by either digital or analogue means. This results from the simplicity of the single-board interface: it is hard to imagine a computer system which would need to have other peripheral interfaces removed to accomodate the DAP interface.

The authors hope that conference delegates can decide informally how the Australasian speech research community should now develop this all-digital storage and shipping method for speech databases.

ACKNOWLEDGEMENTS

The adaptable Multibus interface circuit, used in its DMA variant on the prototype DAP interface, was developed by Winston Burhop during 1984. Craig Norris, of Sony (Australia) Pty. Ltd., readily produced complete technical documentation for the PCM-501ES DAP and for previous models, and offered useful advice. Mark Hill contributed in numerous ways to the project. He was involved in all planning during 1985, designed the memory management circuits, and wrote the low-level driver software. Recently, the prototype has been substantially rewired by John Andrew.

REFERENCES

BLADON, A., de GRAAF, Tj., and O'KANE, M. (1984), "Vowel Databases", Speech Communication 3, 169.

GUYOTE, M.F., LEWIS, K.A., and LIJANA, D. (1986), "A Speech Data Base at the United States Air Force Academy", Proc. IEEE Int'l. Conf. on Acoustics, Speech and Sig. Proc., Tokyo, 313-315.

ITAHASHI, S. (1986), "A Japanese Language Speech Database", ibid., 321-324.

PALLETT, D.S. (1986), "A PCM/VCR Speech Database Exchange Format", ibid., 317-320.

OROS SA (1986), "OROS.AI Adapter" (brochure), Meylan, France.

PALLETT, D.S. and BAKER, S.M. (1983), "Guidelines for Performance Assessment of Speech Recognisers", mimeo.

WATKINSON, J.R. (1986) "Compact Disc Mastering", Electronics & Wireless World 92, February, 47-50.