Digital Audio

by John Watkinson

Once in the digital domain, audio can be stored in a wide variety of ways. This article illustrates some of the more important digital recording techniques and formats, and compares the advantages and limitations of each in terms of the functions a user might expect from a given technology.

fundamentals

The advantages of digital audio recording are now well known, and so it is not necessary to repeat them all here. If the most significant feature of digital audio recording has to be stated, it is that the audio quality is made independent of the storage medium. In this case the quality conscious will immediately ask what constrains the quality. Assuming a sensible sampling rate, the number of bits in the sample determines the signal-to-noise ratio, a good rule of thumb being 6 dB/bit. This gives a theoretical performance that designers can never exceed. Converters and filters can and have fallen short of the theoretical performance. 16 bits is perfectly adequate, indeed in some cases excessive for consumer use, but professional users like to have a performance margin over the consumer to allow for some degradation in processing. This is a nice idea, and some recorders use formats that allow up to 20-bit samples.

Unfortunately technology does not currently permit affordable converters exceeding 16 bits in wordlength. This causes the unprecedented situation, which arose with the introduction of the Compact Disc (CD), where the consumer can hear everything that the professional can hear. It became bizarre when oversampling consumer converters started to outperform those fitted in professional machines. Digital technology is very good at standing situations on their heads.

Once an audio waveform is expressed as a series of binary numbers, it can be stored on any medium that offers a suitable bit rate and an acceptable error rate. The necessary data rate can be obtained by multiplying the sampling rate by the sample length. This results typically in 1 Mbit/s for a high-quality audio channel and means that a high recording density is necessary if a reasonable playing time is to be achieved. The error rate for quality audio must be better than about 1 bit in 10^{10}. Media with this kind of raw error rate are almost unknown, especially if the recording density is high, and so an error-correction system of some kind will usually be necessary. Error-correction systems work by adding redundant symbols to the original data, and so the overall storage requirement increases. In practice the use of error correction allows a higher raw error rate to be tolerated; thus a given medium can be recorded at higher density. The gain in density eclipses the added storage requirement. One fundamental characteristic of digital machines incorporating error correction is that the only way in which their audible performance can be degraded is if the uncorrectable error specification is not met. In other words, if the series of numbers which expresses the audio waveform is repeated exactly on replay, the recording process has been perfect. This has some interesting consequences.

First, if the uncorrectable error rate specification is met, the best transport to use is the cheapest, since no amount of overengineering will improve the audible quality. A certain amount of entertainment can be had currently by reading the popular hi-fi press where it is suggested that the use of exotic materials in the transports of CD players improves the sound quality.

Second, the designer has a degree of flexibility, because equal performance can be obtained by combining a powerful error-correcting format with an indifferent transport or by combining an indifferent format with a good transport.

Third, the user has generally no idea how close the machine is to breakdown. Wear and dirt buildup increase the raw error rate without audibly affecting the uncorrected error rate, until suddenly the raw error rate exceeds the capacity of the error-correction system and the quality nosedives. This is in contrast with analog machines, which go downhill gracefully between maintenance. The solution is to monitor the raw error rate, since this reveals the performance of the transport. Early machines buried the raw error indicators, in case the uninitiated thought that a

Recorders

machine that was making errors could not be any good. More recently the raw error rate has become a useful maintenance indicator. The provision of a "goodness" indicator on the meter bridge of Studer digital machines is an example of this concept. This can be seen in Fig. 1.

Although it has been possible to record digital audio for many years, the price penalty acted against widespread adoption until relatively recently. Digital audio was simply waiting for the appropriate technology. A combination of high-density recording and LSI chips suddenly reduced the capital and running costs of digital recorders, and they became viable as commercial products.

In professional use, a recorder has to be able to do much more than just record and play back. One has only to consider the uses (and abuses) to which an analog recorder is put in the real world to realize that sound quality is only the beginning of the specification of an audio recorder. Even an unpretentious reel-to-reel analog recorder can be used for tape cut editing. Many will run at variable speed for pitch correction or to squeeze an item into a broadcast time slot. Most offer a range of tape speeds so that the best compromise can be reached between frequency response, tape consumption, and ease of edit point location.

Simple comparison of analog and digital is missing the point, however. Once in the digital domain, the range of storage technologies available is much wider, opening up opportunities denied to analog technology. Samples can be stored in RAM or bubble memory, on converted analog video recorders, on digital magnetic tape with stationary or rotary heads, or on optical or magnetic disks of various types. One should not expect digital recorders necessarily to mimic analog machines in appearance, features, or method of operation. In the early days of digital audio, the only devices that could support the required bandwidth were hard disk drives and analog video recorders. At that time the storage capacity of disks limited the playing time, and so the video recorder became the basis for the first practical digital stereo machines. Analog video recorders

Fig. 1. In this Studer DASH format recorder the signal-quality indicator driven by the error-correction system can be seen to the right of the level meters.

could not be used for multitrack work, and the stationary-head digital multitrack was next to evolve. Recent advances in recording density, particularly Winchester technology, have allowed the magnetic hard disk to compete again for some applications. The optical disk has only had significant impact in the read-only CD, but recordable and erasable optical disks may increase in importance.

PCM adapters

Most digital recorders based on video recorders use a so-called PCM adapter. Incoming analog audio is sampled, quantized, and converted to a binary bit stream. This bit stream is then used to generate a waveform called pseudo video, which is sufficiently like a video signal that a video recorder will reproduce it. The active line is made to store bits as black and white levels. The synchronizing period and the vertical interval cannot be used for recording samples, so all of the audio data must be compressed into the active portion of unblanked lines. The replay section of the PCM adapter can easily combine the necessary expansion with the time-base correction process. In order to decode the pseudo video correctly, the bits and their weightings must have a stable relationship with the synchronizing pulses. This is only possible if the audio sampling rate used is related to the video standard. A problem immediately arises in that there are two line standards in common use, 525/60 and 625/50. The sampling rate of 44.1 kHz came about because it is high enough for adequate audio bandwidth, and it can simply relate to both line standards. In fact the limitations of using converted video recorders have determined the sampling rate of the CD. Even though it has no video structure, the CD simply has to have the same sampling rate as the PCM adapters used to make the master tapes.

PCM adapters have been made to complement Betamax (PCM F-1/701), VHS (JVC), and U-matic (PCM 1610/30) recorders. All of these work in basically the same way, although there are differences in the interleaving and error-correction strategies.

The F-1/701 is a consumer product and as such was designed to work with VCRs on both line standards. The EIAJ format for consumer PCM adapters actually uses 14-bit samples, but offers a powerful error-correction system using B-adjacent coding. Sony modified the EIAJ format to offer 16-bit working by putting the extra 2 bits from several samples into the position previously occupied by part of the error-correction redundancy. A characteristic of consumer video cassette recorders, with which these units are designed to work, is that they are seldom fitted with an input to allow playback locked to a reference. The player will run at its own speed, and the PCM adapter has to figure out what the sampling rate then is. As a result, speed instability in the video cassette recorder will show in the digital audio. Since these units were designed with only analog inputs and outputs, one converter was used at each end and time multiplexed between channels. Converter offset was of no consequence. These devices transcended their original market at the high end of consumer hi-fi recording, and became the introduction to digital recording for many professionals because of the low cost and the good audio quality. There were a number of drawbacks for professional use. Several companies modified units to offer digital outputs when the time multiplexing and offsets became apparent. The slow linear tape speed of consumer video cassette recorders meant that time code in the analog audio tracks was not very reliable. Digital filters were developed to time align the samples and remove offsets, and some adapters could convert the signal into PCM-1610 format for editing so that CDs could be made without reversion to the analog domain.

The PCM-1610 was designed primarily to make master tapes for CD production and so is a true professional device. To facilitate international interchange of master tapes, there is only one PCM-1610 format, using 525/60 monochrome U-matic video cassette recorders. For CD mastering at 44.1 kHz the field rate is exactly 60 Hz, but an alternative mode of operation is supported where the machine can be locked to the 59.94 Hz of NTSC broadcast video. In this case the sampling rate becomes 44.0559 kHz and drop frame time code is used. CD cutters will reject a tape with drop frame time code. In the PCM-1610 the sampling rate is determined by genlock or an internal crystal, and this produces a video reference to which the video cassette recorder locks. The wow and flutter will then always be unmeasurable. The U-matic format has a video bandwidth and signal-to-noise ratio far in excess of the needs of two digital audio channels, and so no great sophistication should be expected in the digital format. Error handling is by CRCC detection and parity correction in a simple crossword structure with 100% overhead. Extensive block interleaving allows more than 11 video lines to drop out before uncorrectable errors are caused. One of the biggest drawbacks of the PCM adapter is the lack of confidence replay during record. The DMR 4000 U-matic transport shown in Fig. 2 was developed by Sony to overcome the problem. It has additional heads on the drum, which allow simultaneous replay while recording. Using additional circuitry in the PCM 1630, this transport can replay with the record heads and the replay heads at the same time. The samples containing the least errors are used. This allows one head to clog during replay with no ill effect, which is useful when cutting CDs.

When producing a CD master, the source material for the various bands on the disk may be several different cassettes that vary in level, and each may contain errors in the performance that need to be replaced by material from retakes. Unlike vinyl disk cutting, where the cutter operator has control over the audio level, CD cutting is completely automated. Every sample on the master tape appears on the final disk. This means that the production of that tape is the last creative step, after which no correction is possible. All relative levels

Fig. 2. Sony's DMR-4000 is a U-matic transport specially designed for digital audio. It has additional heads on the drum to allow confidence replay in record or read after read in replay.

and pauses or cross fades between bands must be exactly as required. To perform digital audio edits, the PCM-1610 and two or three video cassette recorders are connected to a digital audio editor. Editing is by assembly, a process where the new piece is edited onto the end of existing pieces. There are three stages in the process, the location of the out point on the existing recording, the location of the in point on the new recording, and the edit process itself, controlled by the in and out points, the level of the new material, and the cross-fade period.

Using video cassette recorders for audio editing is inevitably complicated. The major problems to be overcome are as follows:

1) Variable-speed operation is not possible; edit points must be located in some other way.

2) Video cassette recorders can only edit to frame boundaries, whereas audio needs to be edited to greater accuracy.

3) The action of cross fading implies that two audio sources must be available simultaneously, yet only one PCM adapter is used for economic reasons.

4) Industrial video cassette recorders do not always accelerate at the same speed. Two machines parked on the same time code may be a frame apart when they achieve lock, even if they are started together.

To locate edit points, the editor contains a RAM. The area of the edit is played into the RAM, and this can then be heard at variable speed or, in reverse, by generating the RAM address from a hand-operated rotor. As the rotor turns, the time code of the edit point is updated and can be determined to a fraction of a frame. Both edit points need to be located in this way.

The edit process is completely automatic and uses the in and out points as control parameters. Since only one PCM adapter is used, several frames in the area of the out point must be played from the recorder into RAM. This is done in advance of the edit proper, and occurs automatically under control of the system microprocessor. The sight of the video cassette recorder busily transporting tape with no sound emerging is a little disconcerting if the reason is not appreciated. Once the preload is complete, the edit proper can commence. Both video cassette recorders back up to suitable time codes and are rolled together. The recorder is playing at this stage and is connected to the PCM adapter so that the end of the current recording can be heard. At this time the system also compares the time codes of the two transports to see whether they have the correct frame relationship, or whether there is an offset due to differing acceleration rates. When the time code of the beginning of the preloaded memory is reached, the samples from the memory become the source of monitor audio, freeing the PCM adpater. This is then connected to the player video cassette recorder. Samples from the new material are fed to a second area of memory. Here they experience delay, which is controlled in frame steps to compensate for small time-code slips between player and recorder. A further delay is then imposed, calculated from the positions of the in and out points relative to the frame structure. The effect of this delay is that the sample at the in point will be fetched from the preload memory at the same time as the sample at the output emerges from the delay. As this point is approached, both sample streams are fed to a cross fader, which is set to the first set of samples and will ignore samples from the player. At a frame boundary, the recorder begins to record the cross-fader output and thus rerecords what was already on the tape. However, the cross fader will then operate, and the rerecording will change to the new material. In this way the position of the cross fade can be made independent of the frame timing. The cross fade has to be completed before the preloaded memory runs out of samples. After this, samples from the player are dubbed to the recorder, and this continues until the operator intervenes. The only difference between an edit and a preview is that the recorder fails to record in preview.

While this mechanism is complicated, it works well because all of the various timing signals come from the control system automatically. The edit process does, however, take a long time, because of the need to preload the memory and then rewind again to perform the edit. U-matic recorders do not wind particularly quickly, and this exaggerates the problem. JVC also manufactures a VCR-based CD mastering system, which differs from the Sony unit primarily in the use of a more complex error-correction system. The JVC format also encodes frame numbers into the pseudo video, which replaces the time code on the audio tracks in the Sony. This allows the use of VHS decks instead of U-matic if desired. Fig. 3 shows the JVC AE-900 editor. The rotor used to locate the edit points can be seen. PCM adapters have also been made, which allow the recording of four digital audio channels. These must, however, all be recorded at once, since all four bit streams are combined into a single pseudo video signal. This is not at all the same as having a true multitrack machine. Fig. 4 shows the Denon DN-039R four-track PCM adapter.

direct digital rotary recorders

Some designers went beyond the use of an analog video recorder. In the mastering recorders built by Decca, only the transport and servos of a video recorder were used. The signal fed to the head was true digital, using a binary channel code and synchronizing patterns for timing purposes. The performance of the open-reel 1-inch video transport and the error-correction system is such that monitoring is generally carried out with error concealment disabled. In this way any uncorrectable errors can be heard, and a buildup of concealments from one generation to the next can be prevented. The Decca machines were actually the forerunners of R-DAT, the only conceptual difference being size.

R-DAT began life as an experimental Sony machine. An extensive standardization procedure involving some 80 companies led to the specifications of R-DAT and S-DAT. R-DAT is interesting because it achieves the highest recording density of any magnetic recorder. The cubic volume of tape in an R-DAT cassette is actually less than the volume of a CD. An advantage of rotary heads is that the head-to-tape speed is high, raising the replay signal above head noise. This is combined with azimuth recording, as used in Betamax and VHS video cassette recorders, which needs no gap between adjacent tracks. R-DAT tape is of the same width as compact cassette tape, but is coated with high-coercivity metal powder, allowing a useful output with incredibly narrow tracks which can only be traced with an active track-following system. The tracks on R-DAT are 13 μm wide, only six times the track pitch of a CD. A consequence of the very narrow tracks is that the linear tape speed is extremely slow, only 8 mm/s. This allows the transport to shuttle at up to 200 times play speed, making access to material rapid for a tape format. The track-following system of R-DAT uses a dedicated part of each head sweep. The alignment patterns form natural breaks between the audio samples and the subcode. Audio and subcode can be recorded and edited independently. R-DAT operates at three sampling rates, 32, 44.1, and 48 kHz. Consumer machines will not record at 44.1 kHz in order to prevent digital dubbing from CDs. Dubbing of CDs is still possible in the analog domain, which is no different from the situation with the Compact Cassette. Since an R-DAT machine is effectively a cross between a video cassette recorder and the electronics of a CD player, it will never be as cheap to produce as a CD player or a Compact Cassette recorder. Fears that the introduction of R-DAT would trigger an unprecedented wave of copyright piracy are groundless. Had this been the case, one might have expected a similar outburst over the PCM-F1, since this was a consumer PCM recorder that records at the same sampling rate as CD. The now-discredited copycode episode stands as a triumph of misinformed politics over honest technology. If R-DAT is not succeeding as a consumer format, it is finding friends in professional circles, since the sound quality is a match for digital machines costing many times more, and the 48-kHz sampling rate of the consumer players is, of course, the professional standard rate. Sony has introduced a professional version of an R-DAT machine, as shown in Fig. 5, which has AES/EBU digital inputs and outputs in addition to the consumer digital connections and the analog converters. This machine will record from the AES digital input at 44.1 kHz and so can be used to make CDs without rate conversion.

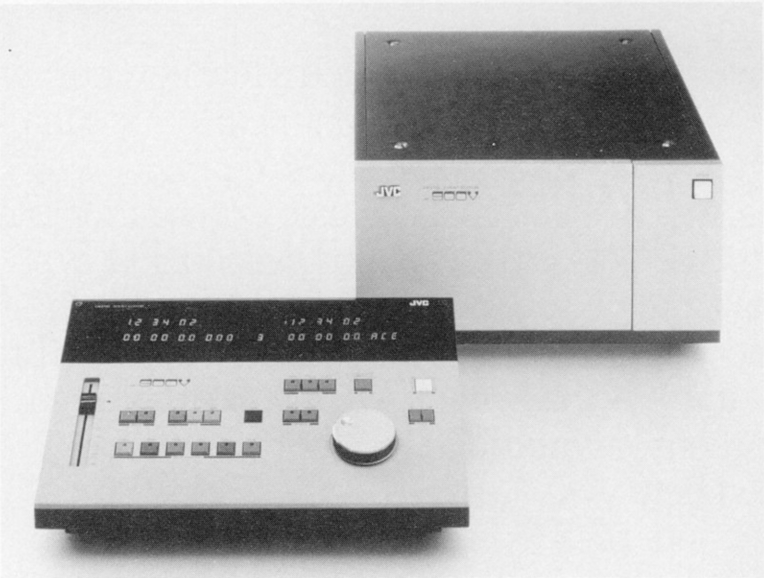


Fig. 3. Compact Disc mastering editor from JVC clearly shows rotor for edit point location.

stationary-head recorders

The stationary-head digital recorder has followed a different evolutionary route than the rotary-head machine to satisfy different requirements. When the analog multitrack

Fig. 4. The Denon DN-039R PCM adapter allows simultaneous recording of four digital audio channels on a video cassette recorder.

studio of a few years ago was analyzed, it was soon apparent that the quality was largely being determined by the tape machine, because the noise and distortion performance of the desk and microphones was significantly better. The situation was exaggerated by the tendency to work with multiple generations, at which analog recording does not excel.

The stationary-head digital recorder came into being as an exact replacement for an analog multitrack, so that it could be plugged into an otherwise analog studio to offer an improvement in overall quality.

The operational requirements of a multitrack recorder are quite specific. Individual control of tracks is essential, so that some will record while others play. It must be possible to bounce one track to another, and synchronous recording, where a performer listens to one track playing in order to record a second track, must be supported. Variable speed is needed for pitch correction and for tape editing by either cutting or using punch in.

Despite this formidable list of requirements, digital machines were developed to fulfill them. The requirement for splicing meant that the machines would be open reel, and this went well with the general attempt to make them appear the same as analog machines, both physically and in their use.

The analog video recorder could not be adapted for this purpose, since it lacked the bandwidth for more than about four tracks, and splicing was nearly impossible. (It was possible to edit transverse-scan Quadruplex video tape by splicing, but all modern video tape recorders use helical scan.) Digital multitracks needed to have separate tape tracks for each audio track, as analog machines always have had, so that the operation of each track would be independent of the others.

There are many contradictions in stationary-head digital machines since an exact copy of an analog machine is being attempted. The freedom of the digital domain is largely denied to a multitrack for this reason. This is not particularly healthy from a technological point of view and may be an evolutionary dead end. The contradictions can be listed here.

First, the bit rate of a digital audio channel demands high recording density with short wavelengths. This implies thin tape, which allows intimate contact with the heads. The roughness of the back coat has to be restricted to prevent embossing of the adjacent magnetic surface when the tape is on the reel, which, combined with the thinness of the tape, restricts the spooling speed achievable by digital multitracks. The short-wavelength requirement is directly contradicted by exposing the medium to contamination on an open-reel transport. Packing density has to be sacrificed with open reel to achieve a format that works when the tape has been handled.

Second, the need for operation in an environment where fingerprints and debris are the norm means that the error-correction system must use extensive interleaving. Interleaving spreads the effect of a large physical defect on tape to make it appear to be many small defects, which are more easily corrected. Unfortunately a large interleave is the last thing to have in a tape that is to be spliced,

Fig. 5. The PCM-2500 from Sony is a professional R-DAT machine which has comprehensive digital input–output facilities, including AES/EBU interface. It also holds world record for tape packing density.

since the damage caused by a splice is proportional to the interleave length. It is necessary to shuffle the samples before recording so that odd and even samples are displaced by a distance larger than the interleave length. A splice then causes two error bursts at different times, one with only odd samples corrupt and one with only even samples corrupt. Interpolation can be used to conceal the corruption.

The interleave processes cause delay. It takes about one tenth of a second for the input to a stationary-head machine to emerge from the confidence replay output. This delay is unacceptable in synchronous recording and prevents playback using the record head. An extra head is necessary, physically shifted along the tape path, so that it can make a synchronous recording about one tenth of a second after the playback process began. Fig. 6 shows the head stack of the PCM-3324, where the additional head can be seen.

A further complication arises where punch in is to be used with an interleaved recording. It is just not possible to start recording at the punch-in point since this would destroy the interleave. A punch in requires the use of two heads physically displaced along the track. The first replays the existing recording, which is then deinterleaved to the original real-time sample sequence. This sample sequence is fed to the record channel, which begins to rerecord the existing recording. The punch in is performed by cross fading new material into the real-time sample sequence. The punch out is performed by cross fading back to the replay signal. After a short time the rerecording can then stop.

Currently the stationary-head market is dominated by two incompatible formats, DASH from Sony, Studer, and Teac, and ProDigi from Mitsubishi, Telefunken, and Otari. The only thing the two formats have in common is that they both support the above mechanisms.

DASH as implemented in the PCM-3324 has 24 audio channels on half-inch tape and uses one tape track per audio channel. All data and redundancy are carried in that track with an extensive interleave. The error-correction system works by CRC error flags generated on tape blocks, which are passed to a simple parity correction process on a convolutional cross interleave through the data. Recording any one audio track on a new tape results in a control track being laid down in the center of the tape. This track is synchronous with the sample rate. It contains a contiguous sector address, which can be used for an absolute autolocator or to lock two machines together to sample accuracy. Subsequent tracks can then be recorded synchronously, so that the control track is played back to lock all new tracks to the existing ones. The only drawback of synchronous recording is that confidence replay is not available due to the recording being made by the delayed head. Splices are detected by looking for discontinuities in the sector addresses in the control track, and so the tape should be butted together snugly when it is spliced. The 3324 has no internal time-code generator, but has a dedicated track which will record any standard code fed in. This may be from a generator that is locked to the sampling rate or not, depending on the user's installation. The ProDigi format offers 32 audio channels on 1-inch tape, but uses many more tape tracks than audio channels since most of the redundancy is carried in separate tracks. Effectively the tape format forms a crossword, or product, code, where CRC checks along the tracks act as pointers for Reed–Solomon code words working across the tracks. The cross-track redundancy means that there is no way that a single track can be recorded. At least eight audio tracks must be recorded on a blank tape in order to produce one of the four cross-track patterns. If a single track is to be rerecorded later, it is necessary to play back all eight tracks in the group in order to calculate the different cross-track redundancy due to the changed track. In practice users tend to preempt operational problems by formatting new tapes. All subsequent recordings are then synchronous to the original format process. The Mitsubishi machine has an SMPTE 60-Hz time-code generator built in, which locks to the sample rate, and the output of this is usually recorded during the format process. As the time code is always synchronous in ProDigi, the control track is absent and there is no parallel with the sector addresses of DASH. For this reason it is necessary to leave a small gap in tape splices because the loss of signal in all tracks is the

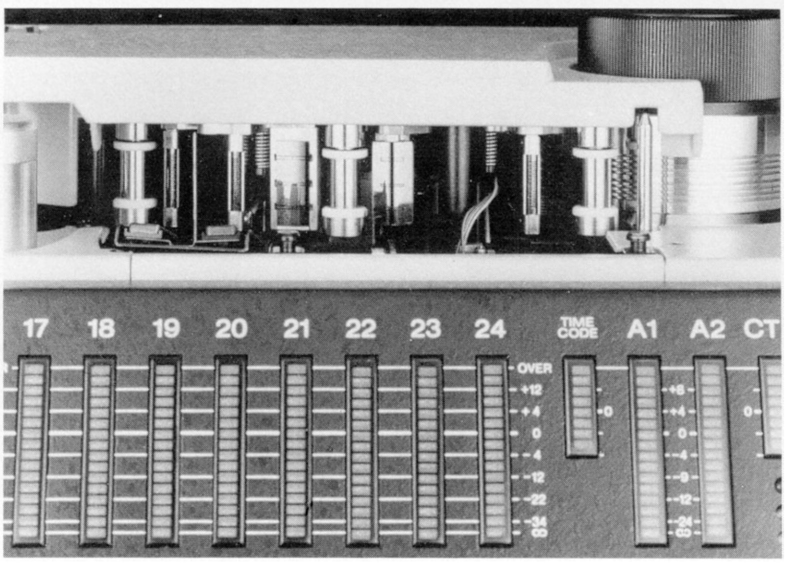


Fig. 6. Head stack of the popular PCM-3324 DASH multitrack showing the additional head (last head before capstan) needed for synchronous recording and punch in. 24 audio channels on half-inch tape, with one digital track per audio channel.

only way that the splice can be detected. Fig. 7 shows the Mitsubishi X-850 digital multitrack.

In practice the incompatibility of these two formats is less of a problem than it might appear. It is quite easy to perform digital dubs between DASH and ProDigi machines, in either direction, using commercially available adapters or a simple homemade interface. It is also possible to synchronize DASH and ProDigi machines using time code. If the DASH time code is not synchronized to the sampling rate when the rerecordings are made, the sampling rates of the DASH and PD machines will not be the same when the synchronizer forces the two time-code rates to be the same. This may not matter if playback is used in the analog domain, but a digital track bounce between the two machines could not be achieved unless both had sampling-rate synchronous time code.

Recently stereo versions of DASH and ProDigi have become available. These use one-quarter-inch tape and have compatible physical track positions across the tape, even if, as before, there is no other compatibility. It is thus feasible to construct a dual-standard machine, but the political problems involved would dwarf the engineering issues. In one version of the stereo DASH format, the tape speed is doubled, and every sample is recorded twice, once as normal and once with a reverse odd–even shuffle. In the presence of a splice, the reverse interleave ensures that all samples are always available, and no interpolation is necessary. The higher tape speed also makes the location of the edit point easier. The PCM-3402 shown in Fig. 8 is a two-speed DASH machine which supports normal DASH at 7.5 in/s or TWIN DASH at 15 in/s.

S-DAT is the stationary-head brother of R-DAT, using the same tape, but in a different cassette. S-DAT obtains a high recording density in a different way. Conventional magnetic heads respond to the rate of change of flux, so they work better at high speeds. S-DAT uses magneto-resistive heads which measure the tape flux instead of differentiating it. These heads have a noise advantage at low tape speeds. There are 20 tape tracks in each running direction in S-DAT, and these work in parallel to carry a digital stereo recording. Unlike R-DAT, the subcode data are combined with the audio samples and cannot be edited independently.

Fig. 7. Mitsubishi X-850 ProDigi format recorder offers 32 audio channels on one-inch tape. Due to cross tape error correction, 8 audio channels require 10 tape tracks.

disk drive recorders

The disk drive was originally developed as a rapid-access storage device for computers. Data are stored in concentric tracks on the disk surface, and the track can be selected by moving the head radially with a positioner. This positioning process, known as seeking, can be completed in a few tens of milliseconds from the beginning of the disk to the end, and is vastly superior to the access speed of tape, which requires lengthy winding. Disk drives were used for early experiments with digital audio, but lost out to PCM adapters because the capacity of disks was at that time uncompetitive and the medium cost per bit rather high. Progress in recording technology has increased the density of disk drives. High-coercivity media and improved track-following systems have allowed the use of narrower tracks, and improved channel coding has increased the longitudinal density. To reduce access time, hard disks turn at high speed. The head-to-track speed is around 100 km/h, which makes the instantaneous data transfer rate extremely high by audio standards. Unlike magnetic tape, where there is contact with the head, disks use an air bearing between the head and the medium to eliminate wear at the high relative speeds employed. This air film causes a severe spacing loss, and clearly the recording density

Fig. 8. Sony PCM-3402 is a two-speed DASH machine. At 7.5 in/s it records two channels in normal DASH for tape economy and long play time. At 15 in/s it records two-channel TWIN DASH, which has better performance for tape-cut splicing.

achievable is a function of the height at which the heads fly. The flying height cannot be reduced indefinitely, since it has to be high enough to allow contaminant particles to pass underneath. If a particle becomes attached to the head, it will collect other particles and spoil the aerodynamics that keep the head in flight. The head may then touch the disk, burning itself and destroying the magnetic layer, an event known as a head crash. Removable disks have to restrict their storage capacity to allow for the contamination experienced when the disk is out of the drive.

The Winchester disk was developed to overcome these problems. In Winchester disks the heads and disk platters are sealed inside an enclosure, which is purged of contaminants and allows the flying height to be reduced. The longitudinal density can be increased considerably. The boom in desktop computers has fueled demand for compact Winchester disk drives at low cost, and this has allowed disk drives to become viable again for digital audio use.

Since an audio channel sampled at 48 kHz requires about 1 Mbit/s, 1 Mbyte—the usual measurement unit for disks—allows 8 s of full-bandwidth audio. A typical Winchester disk can store 200 Mbytes, allowing about 25 track minutes of audio.

Like all digital recorders, a disk-based system must use a certain amount of memory to buffer the drive from the converters. Since the transfer rate of the drive is several times the sampling rate, the drive can fill the memory in a short burst, and the memory then supplies samples to the output at the correct speed, allowing the drive time to move its heads to another cylinder. By adding further converters, it is possible to make a multitrack machine since the speed of the drive allows samples for several channels to be fetched to the memory faster than the memory sends them to the converters. This is the reason for measuring disk-based systems in track minutes, since the capacity can be used for one audio track for maximum time, or for several audio tracks for a shorter time. The playing time in real life will be greater than the time obtained by dividing the capacity by the number of tracks. In a typical multitrack recording, all tracks are not necessarily recorded all of the time. A multitrack tape recorder will waste tape in these circumstances, whereas the disk-based system only stores wanted samples on the disk.

The disk drive splits up the data tracks into blocks, and each block can be uniquely specified by supplying three parameters, the cylinder, or positioner address, the head, or surface address, and the sector, or rotational angle address. Clearly it is vital that the correct block be supplied at the right time. One way in which this can be done is to use the block structure of the AES/EBU digital interface. If each disk block is made the same size as an AES/EBU channel status block, it is only necessary to construct a table that relates the time code of each block to the disk address. To play back the recording it is then only necessary to provide the appropriate time code. A system of this kind essentially has a built-in synchronizer and could provide a digital sound track capable of chasing video tape or film. Fig. 9 shows the Digital Audio Research Soundstation II, which is a disk-based recorder–editor.

A great strength of disk-based systems is that the speed of editing is increased dramatically by the random access of the disks. Editing follows the same steps as with any medium in that the in and out points must be located first of all. To provide an ergonomic interface to the operator, it must be possible to hear the recordings at various speeds. Some increase in speed can be had by raising the system clock rate, so that the sampling rate and the time code advance faster than real time. However, to obtain very high speeds, another technique is needed. When the recording is made, the disk blocks are filled as normal, but the samples are also fed to a sampling-rate reducer, which compresses several regular blocks into so-called spooling files. These are also recorded on the disk. If a spooling file is played back as if it were a regular audio file, the speed will rise by the compression factor.

Fig. 9. Digital Audio Research offers the Soundstation II hard-disk-based editor. It uses touch screen controls to enhance rapid access of disk drives.

The coarse edit point can be located using spooling files, and then the area of the edit point is loaded into memory for fine access by a hand rotor, as in the PCM adapter editing described earlier. Once two edit points are specified, an edit list is created, which contains the time codes of the points and a cross-fade time. To hear the effect of the edit, the time code is set prior to the out point and advances. Samples are fetched from the disk and can be heard. As the edit point approaches, samples are also fetched from the incoming recording to a different part of the memory. By careful control of memory addresses it is possible to obtain two sample streams from the memory in real time, such that the in point in one sample stream appears at the same instant as the out point in the other one. A digital cross fade takes place across this point. An advantage of disk-based systems is that the two sample streams can be supplied for a considerable time. Extremely long cross fades can be obtained. It should be appreciated that the original audio files have not been changed in any way. The edit was performed by executing the edit list using the audio files as read-only input. At any time the edit can be altered by changing the edit list.

The only drawback of a Winchester-disk-based system is that the disks cannot be removed. The samples held on the disks have to be transferred to another medium for long-term storage or shipment in order to free up the system for a new recording. If it is certain that the edits are perfect, it is only necessary to connect a suitable digital tape recorder and execute the edit list. The edited material will then be dubbed to tape. If the editing is not considered definitive, the entire data base can be backed up to tape so that all of the audio files can be re-created at a later date. Once the backup to tape is complete, the disks can be overwritten with new material. The rapid random access of disk-based systems can also be used for jingle and commercial broadcasting. In that environment, the bandwidth required will be less, and a sampling rate of 32 kHz can be used to increase playing time. For AM or speech applications it can be even lower. A commercial break can be broadcast simply by creating a suitable edit list. The system simply executes the edit list to air, fetching the appropriate audio files to memory and cross fading between them. This allows the commercial broadcast to be changed right up to the last moment since only the edit list need be changed. Fig. 10 shows the For-A Sirius, which is designed for broadcast applications. Disk-based systems of this kind compete directly with cartridge machines.

RAM-based recorders

The falling cost of semiconductor memory has led to the development of RAM-based audio machines. In these devices the memory needed by a disk-based recorder is extended until the disk can be dispensed with

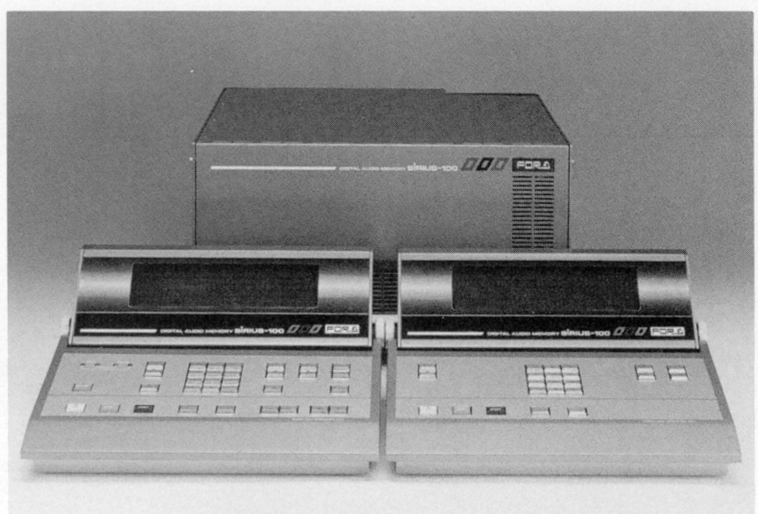


Fig. 10. For-A Sirius hard disk unit intended for broadcast applications uses lower sampling rate to suit broadcast bandwidth. It competes with cart machines for jingle and commercial break applications.

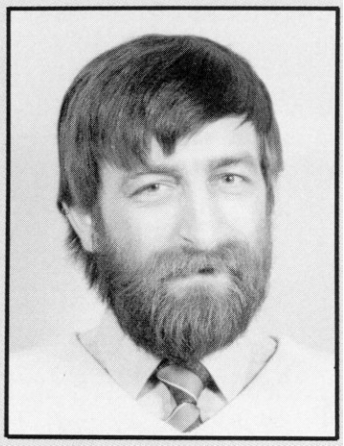
for some purposes. Clearly the access time will be negligible, with RAM storage giving an advantage for editing, but the recording time will be restricted by the 8-s/Mbyte rule. A disadvantage of RAM-based machines is that the recording is volatile and will be lost if the RAM loses power. The edited work needs to be backed up to a magnetic medium both for permanence and to free up the RAM for further work. Fig. 11 shows the Soundcraft Digitor, a RAM-based audio recorder.

A relative of the RAM-based recorder is the sampler, which records a short sound effect and allows its instant recall at various pitches. The effects achieved are much in evidence in current pop records.

Fig. 11 Soundcraft Digitor is a solid-state recorder–editor using large RAM. RAM capacity limits recording time, but access to any part of recording is immediate, speeding editing process. RAM is volatile; work must be dumped to storage medium afterward.

conclusion

This article has illustrated the tremendous variety of hardware available for digital audio recording. The technology is now changing so fast that this article must have omitted something. The future may hold recordable CDs for the consumer and larger capacity optical devices for professional use, which will have advantages over the current write-once optical disks. Despite this, tape recording will not go away. The development of R-DAT shows the kind of magnetic recording density possible today. The process of development will never stop.



the author

John Watkinson was born in Yorkshire, United Kingdom, and graduated from Southampton University in 1971 with an honors degree in electronic engineering. He earned an M.Sc. degree in sound and vibration from the Institute of Sound and Vibration Research in 1974. Following a period as an academic he spent six years with Digital Equipment Corporation specializing in mass storage. He then returned to the audio/video industry, initially with Sony Broadcast, then with Ampex Great Britain Ltd.

Mr. Watkinson is a committee member of the British Section of the AES and a member of the Royal Television Society and the British Computer Society. He has published numerous articles and papers on digital audio and video, coding and error correction, and his first book, *The Art of Digital Audio*, has recently become available.