Digital Audio Standards

MINUTES OF THE MEETING OF THE DIGITAL AUDIO STANDARDS COMMITTEE

Date: 1977 December 1 and 2

Time: 1830 hours

Place: Snowbird Resort, Salt Lake City, Utah

Present: Chairman, John G. McKnight (Magnetic Reference Laboratory); members, Stanley Becker (Scully/Dictaphone); Gregory Boganz (RCA Records); Vic Goh (Victor Company of Japan (JVC)); Thomas Hay (MCI, Inc.); Alastair Heaslett (Ampex Corporation); M. Carlos Kennedy (Ampex Corporation); William Kinghorn (Telex Communications); K. Kimihira (Akai America); Masahiro Kosaka (Wireless Research Lab, Matsushita Elect. Inc. Co.); Alfred H. Moris (3M Company); Thomas G. Stockham, Jr. (Soundstream, Inc.); Martin Willcocks (Willcocks Research Consultants); James V. White (CBS Technology Center); Yoshito Yamagudi (Melco Sales Inc. Mitsubishi Electric Corp.); Robert J. Youngquist (3M Company).

1. Because everyone agreed on the primary importance of the choice of the sampling frequency, the committee worked as a whole on this subject. (See list of subjects following minutes.)

1.1 Sampling frequencies now used were discussed. All agreed on the desirability in principle for *all* digital audio systems to use the same sampling frequency. If this is impossible or impractical, two standard frequencies might be necessary.

1.2 Mr. Heaslett summarized the criteria for frequencies compatible with 525- and 625-line TV systems with integral numbers of samples for 30, 25, and 24 frames per second, and editable more often than once-per-frame. Based on these criteria, he suggests possible sampling frequencies be 45, 54, or 60 kHz.

1.3 Mr. Youngquist showed the reasons that they preferred a 50-kHz frequency in developing the 3M Company recording studio system.

1.4 Mr. Kosaka gave the basis for using 44.05594 kHz in Japanese magnetic tape and optical disk systems now in development for consumer applications. He said that they would consider the possibility of using the 45-kHz frequency proposed by Heaslett.

1.5 Mr. Willcocks gave the available technical details of some 14 presently-used digital audio systems. He subsequently prepared a report containing this information for distribution to the committee (see page 56).

1.6 Several members expressed the urgency for sampling frequency standardization because of the number of digital audio recording systems — both studio and consumer types — now nearing completion and commercial availability.

1.7 The committee was unable to find an acceptable single frequency, given the conflicting requirements of the present TV-compatible proposal, the 3M studio recorder, and the Japanese consumer recorders. The committee asked Messrs. Heaslett, Youngquist and Kosaka each to prepare a report giving details explaining why they chose the frequency they did, and what penalties the other frequencies discussed would entail. The reports of Heaslett and Youngquist appear on pages 66 and 54 respectively. Mr. Kosaka's report will be published in a future issue. The Chairman mailed these reports to the Committee and "information" mailing lists, so that they may reach a better-informed opinion for standardizing one frequency (or possibly two frequencies). This subject will be continued at the next meeting.

2. Preliminary discussions were held on source encoding (A-D conversion).

2.1 Number of bits: a general preference was expressed for a record written with space for 16 bits.

2.2 Pre- and post-emphasis are not presently used in professional systems, which is considered desirable. Present consumer systems do use this processing and this should be further studied. Stockham pointed out the fact that simple digital equalizers do not have the same frequency response as simple analog equalizers and that a pre-emphasized signal must be de-emphasized in order to process it (mix, equalize, etc.) digitally.

2.3 The location of the compensator for the response loss of the D-A converter ($\sin x/x$ loss) was discussed. Most felt that the location should be standardized; some felt that the best location was at the output, but some systems now have it at the input. Further discussion is needed.

3. The next meeting was set for 1978-02-01 and 02, at the Sheraton Atlanta Hotel, in Atlanta, GA in conjunction with the SMPTE TV Conference.

4. The meeting was closed at 1200 on 1977-12-02.

JOHN G. MCKNIGHT
Chairman

TECHNICAL FACTORS TO BE STANDARDIZED (in priority order)

1. Sampling frequency (a standard value is necessary to avoid extra A-D/D-A conversions).

2. Source encoding (A-D conversion: what the bits mean — linear PCM, log PCM, pre/post emphasis, floating point, etc.)

3. Code formats themselves for each storage medium (for example, Miller, NRZ, sync, redundancy, error correction, space for other things such as time code, editing and mixdown information).

4. Format for storage (for example, for fixed-head tape, width, length, thickness, speed, track format, packing density. For disks: diameter, speed, pitch, etc.).

5. Digital interfacing of the bit-stream on wire(s) (for example, serial vs parallel; voltages).

6. System block diagram (what goes in console; in peripheral equipment; on storage medium).

7. Analog interface (for example, program level measurement; reference voltage; "lead").

8. Measuring and evaluating system performance (how meaningfully to measure signal minus noise level in digital audio, frequency response, how to define a "qualified" tape, how to align a machine for interchangeability, specifications (standard reporting of performance), electromagnetic interference generated and received; "headroom").

9. Vocabulary (defining a track, a channel, etc., in both the analog and the digital parts of the system).

LIST OF APPLICATIONS FIELDS FOR DIGITAL AUDIO

1. Sound recording studios: multitrack mastering; 2-track mixdown.

2. Motion picture and TV sound production.

3. Signal processing (nonrecording): delay and reverberation; filtering and equalizing; mixing; time compression.

4. Distribution: satellite; radio and TV; long-lines; microwave.

5. Consumer recording: optical disk; mechanical disk; magnetic disk; flat-plate optical; magnetic tape with stationary heads; magnetic tape with moving heads.

6. Digitally synthesized sound and music, for research, entertainment, information storage and retrieval, etc.

NEXT STEP

Further discussion will continue at the next meeting, 1978 February 1 and 2 in Atlanta, GA, when the goal will be to arrive at one standard sampling frequency for digital audio systems and the rationale for this frequency. If this is impossible, an attempt will be made to arrive at a small set of frequencies and the rationale for these.

Sampling Frequency Considerations

Robert J. Youngquist

Research Manager, Mincom Laboratory, 3M Company

The minimum sampling frequency is a function of the audio frequency response required and what is economical to achieve in the anti-aliasing filter. If 20 kHz is chosen as the upper audio band edge and a filter with a sharp cutoff is used, it is possible to have a minimum sampling frequency of 45 kHz. If the recorder must have a ±10% speed change and the standard sampling frequency is 50 kHz, then the sampling range is 45 to 55 kHz which satisfies the previous 45 kHz requirement. On the higher sampling rate side, any increase would require an increase in recording density or an increase in tape speed. With the 3M system, if we were to go to a standard sampling frequency of 54 kHz, our density would go from 27.78 kilobits per inch to 30 kilobits per inch. The alternative of raising tape speed would take us from 45 in/s to 48.6 in/s. This would reduce our 12.5-inch reel playing time from 33.3 minutes down to close to 30 minutes, which we feel is undesirable.

The other requirement of sampling frequency is its relationship to NTSC and PAL television rates and to the 24 frame per second film rates. At a sampling rate of 50 kHz the relationship to PAL is that there are exactly 2000 digital audio words per PAL frame. In the NTSC system, we find there are 5005 digital audio words in three frames, and in film systems at 24 frames per second, we would have 6250 words in three frames. Therefore, it can be seen that a sampling rate of 50 kHz has an integer relationship to all three systems. Because of the above considerations, we feel 50 kHz is a very good compromise sampling rate.

A Review of Digital Audio Techniques

MARTIN WILLCOCKS

Willcocks Research Consultants, Santa Monica, CA

Techniques currently in use for digitally encoding, transmitting and recording audio signals are reviewed. A comparison is drawn on the bases of sampling rates, linear and nonlinear encoding of data samples, formating of the data for storage and transmission, error correction schemes and characteristics of recording and transmission media.

INTRODUCTION: The advent of digital recording and transmission techniques for high-fidelity audio signals in both professional and consumer applications has spawned a number of similar, but conflicting, systems. In the infancy of this new technology, the need for early establishment of standards for digital audio is recognized throughout the industry. Such standards will enable the transfer of data from one medium to another to be accomplished without intermediate digital-to-analog and analog-to-digital conversions. The data presented here are derived from manufacturers' literature, technical papers, articles and news reports, and underline both the disparity between the various sytems now available and the urgency of the task of standardization.

The principal parameter to be standardized is the sampling frequency, which determines the audio bandwidth of the system, but, more importantly, in practical situations determines whether digital-to-analog conversion is needed when transferring data from one medium to another. Although conceptually it is possible to transcode digital data at one sampling rate into digital form at a simply related sampling rate, the complexity of the algorithms for so doing may render such digital transcoding economically unattractive compared with conversion to analog form and re-encoding into the new digital code, with the inevitable degradation of signal quality incurred thereby.

Transcoding from linear to quasi-logarithmic nonlinear codes and vice versa presents less serious problems, provided that it is recognized that each such transcoding introduces error terms, and the cumulative effects may become unacceptable. Such effects could be exaggerated if any form of pre-emphasis or de-emphasis is used in conjunction with companding techniques, and any pre-emphasis scheme used should also be standardized. Indeed, it is preferable that no pre-emphasis be used at all, or that it should, if used, be capable of digital de-emphasis by a simple algorithm.

Transferability of digitally recorded tapes between different machines will require not only compatible sampling frequencies and codes, but also compatible formating of the individual data words into blocks, and compatible recording formats.

REVIEW OF EXISTING TECHNIQUES

Two digital audio systems which have been in use for some years are the BBC PCM transmission links in England [1, 2] and the Nippon Columbia disk mastering system using a VTR-based PCM audio recorder [1-6]. The BBC has also been developing fixed-head digital audio recording techniques for several years [1-10].

The BBC's PCM link between Broadcasting House and its Wrotham transmitter has been operational since September 14, 1972. It was extended to Sutton Coldfield in November 1972 and to Holme Moss in February 1973. By June 1975 it had been extended to Kirk O'Shotts in the North and also to the West. The sampling frequency is 32 kHz, using 13-bit linear encoding, with an extra parity

check bit, for each of the 13 channels provided. Eleven synchronizing bits and five auxiliary control bits are included in each frame of 198 bits, so that the bit rate is 6.336 Mb/s.

The Nippon Columbia PCM recording system uses a 4-head low-band VTR to record up to eight channels on 2-inch (50.8-mm) video tape. The sampling frequency is 47.25 kHz, using 13-bit linear encoding plus a parity bit and a check bit for phase shift detection. A clock frequency of 7.1825 MHz is used, so that three blocks of 120 bits can be packed to each horizontal period of a TV-type signal (no vertical synchronizing signal is incorporated). The tape speed is 38 cm/s, with a head-to-tape relative speed of 40 m/s. For four- or two-channel use, each sample can be recorded twice or four times, and error correction employs a three-level scheme; for single sample errors the parity bits are used to determine which sample is correct, or if both are incorrect the missing sample is computed by interpolation between the preceding and following samples. If a series of consecutive errors occurs, the last sample known to be correct is repeated. The recorder is played back at half speed, and the analog signal recovered from the digital-to-analog converter is passed through a tracing distortion simulator to a half-speed cutting machine. A considerable library of master tapes employing this technique exists [11].

Sony recently announced an accessory for the Betamax video cassette recorder which permits PCM audio recording and playback, [11–14] using a sampling rate of 44.1 kHz[1] and 13-bit non-linear encoding[1]. The bit rate is 1.762 Mb/s. Matsushita has unveiled a similar prototype [12], followed by Mitsubishi [15] and Japan Victor Co. [14], the latter being compatible with the VHS system.

Mitsubishi, TEAC and Tokyo Denka Co. demonstrated a PCM laser disk and player at the 58th Convention of the Audio Engineering Society in New York, November 4–7, 1977. The system was described in a paper [16] given at the Convention and in manufacturers' literature [17, 18]. Sony and Hitachi have similar developments under way [13, 14]. Encoding characteristics of the disks were not published at the Convention, but in reply to questions the sampling frequency was given as 47 kHz and the encoding format was described as 12-bit floating-point, using a 15-segment curve specified by an 8-bit mantissa and 4-bit exponent. An extra parity bit is added, and each sample is doubly recorded. Error correction employs a voting system in conjunction with interpolation if both samples are incorrect. These systems are based on video disc technology, the latest entry being a Matsushita mechanical system [19].

Some other machines recently announced are a fixed head PCM recorder [20] by Mitsubishi, with a sampling rate of 48 kHz and 14-bit linear encoding, and a PCM cassette recorder by the same company using a sampling

frequency of [21] 47.52 kHz and 13-bit non-linear encoding. TEAC announced a PCM 4-channel [22] encoder for use with a U-matic VTR employing a sampling frequency of 46.08 kHz and 13-bit non-linear encoding, also using a 15-segment A-law compression, with one bit used for a parity check. In the TEAC system, each set of four samples, making a 52-bit unit, is recorded twice, and separation of the two copies is achieved as follows. A block of 128 such units is assembled, the even-numbered units 0-126 preceding the odd-numbered units 1-127. Two units are recorded to each line of the false TV signal, that is 104 bits. Each block is immediately repeated, and three such repeated blocks are transmitted in each field, with 60 fields per second. This format accounts for the 46.08 kHz sampling frequency.

Also announced at the AES Convention was a new 4-channel PCM recorder by Soundstream, Inc. This uses 16-bit linear encoding, with a sampling frequency of either 37.5 kHz or 42.5 kHz [23, 24]. The Mincom Division of 3M announced a 32-channel fixed-head machine with a tape speed of 45 in/s (1143 mm/s), which uses a sampling frequency of 50 kHz $\pm$ 10%, and 16-bit linear encoding [25]. For this machine, developed in conjunction with the BBC, an elaborate error-correction scheme is used, described in a paper given at the convention. [26].

The data presented above are summarized in Table I.

Although the data in Table I represent a range of different and seemingly incompatible possibilities, there are several areas of convergence, and some of the systems are much firmer than others. The Japanese VTR manufacturers appear to be converging on a 44.05594-kHz sampling frequency and 13-bit nonlinear code because of constraints imposed by the VTR systems employed. These standards if adopted will probably be adequate for consumer VTR-based audio recorders, but not good enough for professional use. Therefore, it is likely that there will be a dichotomy between standards for these uses. However, the VTR-based standard, if any, does not necessarily dictate what is necessary for other uses of digital audio, with one exception: the transmission of digitally-encoded audio via TV channels. The only conceivable use of such transmissions would be to give the audiophile with a digital audio VTR the opportunity to copy onto his own tapes high quality musical performances to supplement those which would be available to him on prerecorded video cassettes and disks. It is unlikely that video cassettes will be sold as prerecorded audio software, due to the high cost, and it is equally unlikely that, were the copyright laws to be relaxed, the video cassette would be economically attractive relative to laser PCM disks, which can apparently be produced to sell at around $10.

What other choices of sampling frequency are possible? At the AES Digital Audio Standards Committee meeting on December 1-2, 1977, Mr. Alastair M. Heaslett of Ampex presented a rationale for a sampling frequency intimately related to the horizontal frequencies of both NTSC and PAL television, the frame rates of both TV systems, and the frame rate of the movie industry, arguing that these are areas in which professional recording of

---

[1] Reference [11] gives the sampling frequency as "about 44 kHz" while reference [14] quotes 13 bits per sample. The exact frequency has not been confirmed by Sony. These data are incompatible with the bit rate of 1.762 Mb/s quoted in the references.

Table I. Published data on PCM systems.

Manufacturer	Type of system	Sampling rate (kHz)	Encoding format	No. of channels
BBC	Microwave link	32	13-bit linear	13
Denon/Nippon Columbia	PCM on VTR	47.25	13-bit linear	2/4/8
Various Japanese	Betamax/VHS based PCM adaptors	44.05594†	13-bit non-linear	2
TEAC	PCM adaptor U-matic	46.08	13-bit non-linear	4
Mitsubishi	fixed-head tape	48	14-bit linear	2
Mitsubishi	cassette tape	47.52	13-bit non-linear	2
TEAC/Mitsubishi/ Tokyo Denka	laser PCM disk	47†	13-bit non-linear	2
Soundstream	computer tape	37.5/42.5	16-bit linear	4
3M Mincom	fixed head tape	45–55	16-bit linear	32
*Hitachi	tape	35.7	12-bit	†
*Toshiba	tape	50	14-bit non-linear	†
*Technics	tape	49.2	15-bit non-linear	†

* These systems have been reported to me verbally as experimental but have not been confirmed by the manufacturers concerned, and are included only to show other possible divergences from any standards established.
  † Information so marked has not been definitely confirmed as exact.

sound is important, and that it should be possible to edit such material in data blocks which are compatible with these frame rates. Frequencies which are suggested by this argument are 45 kHz, 54 kHz and 60 kHz.[2] Other suggestions are a rate closely related to the European landline and satellite networks, which operate at 8 kHz, so that a rate of 48 kHz would be convenient. Indeed, this rate is directly compatible with Heaslett's criteria. A rate of 50 kHz has been suggested by 3M, but would not fit these criteria directly.[3]

On encoding formats, there is apparent unanimity among U.S. manufacturers that 16-bit linear encoding, without any pre-emphasis, is both as good as the state of the art will permit and also good enough to permit all required mixing and filtering operations to be carried out without loss of quality. Most computers use 16-bit words as standard format, which makes elaborate digital processing of digitally recorded signals, such as the blind deconvolution performed by Soundstream on some early Caruso records, a great deal easier to perform.

It might be advantageous to employ the same sampling frequency of, say, 48 kHz for professional tape recording and for PCM laser disks. If so, there would also be advantages in allowing for up to 16 bits per sample to be recorded on the disc. Even if only 13-bit nonlinear coding were used on early software, with relatively inexpensive digital-to-analog converters, it would be possible, as the state of the art progressed, to increase the fidelity of later software by using, say, 12-bit mantissa with 3-bit ranging and a parity bit within the 16-bit format.

CONCLUSIONS

Current digital audio techniques employ a variety of sampling frequencies within the range 32–50 kHz, and several encoding formats. VTR-based PCM systems appear to be converging upon a sampling frequency of 44.05594 kHz, which is considered too low for professional use, and a 13-bit non-linear format. A 16-bit linear format is considered satisfactory for professional work,

with a sampling frequency in the neighborhood of 50 kHz. To allow for future development, it would be desirable to adopt the same sampling frequency and a 16-bit "slot" for laser PCM disk-based consumer products.

APPENDIX I: Heaslett's criteria (p. 66)

Heaslett's basic premises are:
1. Data should be organized into blocks for easier editing. An integral number of blocks should occur in each frame period of movie, PAL or NTSC TV sound tracks, so that block boundaries coincide with frame boundaries.
2. An integral number of samples should constitute a block.
3. The system clock frequency should be a multiple of the horizontal frequency in PAL or NTSC TV systems so that it can be locked to this frequency.
4. The system clock frequency should be a multiple of the sampling frequency.

From 1, the smallest possible block frequency is the lowest common multiple (l.c.m.) of 24 Hz, 25 Hz and 30 Hz, or 600 Hz. (This ignores the slightly different frame frequency in NTSC color TV.) From 3, the l.c.m. of the two line frequencies is 2.25 MHz = 143 × 15734.265 Hz or 144 × 15625 Hz. The clock frequency f_c can be an integer multiple of this frequency.

Thus,

$$f_b = 600 \text{ Hz}$$
$$f_c = k \times 2.25 \text{ MHz}$$

where k is an integer.
If there are n samples per block, and the sampling frequency is the mth subharmonic of the clock frequency,

$$f_s = nf_b$$
$$f_c = mf_s$$

where n and m are integers.

Therefore,

$$mnf_b = f_c$$
$$mn = 3750\,k$$

[2] See Appendix I.
[3] See Appendix II.

m, n, k being integrals.

For small integer values of k there are several frequencies which lie within the useful range 40–60 kHz:

k	m	n	f_s	f_c
1	50	75	45 kHz	2.25 MHz
2	75	100	60 kHz	4.5 MHz
3	125	90	54 kHz	6.75 MHz

The clock frequency can be higher; a clock of 13.5 MHz corresponding to k = 6 can be used with any of these three sampling frequencies. The block length of 1/600 s may be too short for convenience; Heaslett suggests relaxing condition 1 with respect to movie sound blocks and allowing the block frequency to be 300 Hz. This also doubles n, so there is an even number of samples per block.

APPENDIX II: Alternative sampling frequencies

If the block frequency is reduced to 300 Hz or 200 Hz, so that block and frame boundaries coincide every two or three frames, and if the clock frequency can be higher, some additional frequencies emerge:

k	m	n	f_b	f_s	f_c
7	375	70	600 Hz	42 kHz	15.75 MHz
8	375	80	600 Hz	48 kHz	18 MHz
1	45	250	200 Hz	50 kHz	2.25 MHz
7	300	175	300 Hz	52.5 kHz	15.75 MHz

REFERENCES

[1] J. Dwyer, "Digital Techniques in Recording and Broadcasting," Wireless World, vol. 81, no. 1474, pp. 248–253, June 1975.

[2] "Sound Recorder Uses P.C.M.," Wireless World, vol. 79, no. 1457, pp. 548–549, November 1973.

[3] H. Iwamura, H. Hayashi, A. Miyashita and T. Anazawa, "Pulse-Code-Modulation Recording System," J. Audio Eng. Soc., vol. 21, no. 7, p. 535, September 1973.

[4] N. Sato, "PCM Recorder–A New Type of Audio Magnetic Tape Recorder," J. Audio Eng. Soc., vol. 21, no. 7, p. 542, September 1973.

[5] Sleeve notes of Denon PCM Recording Demonstration Record No. ST-6001, Nippon Columbia Co. Ltd.

[6] T. Anazawa, K. Yamamoto, S. Todoroki and A. Takasu, "Improved PCM (Pulse Code Modulation) Recording System," presented at the 56th AES Convention, Paris, March 1–4, 1977; AES Preprint No. 1206.

[7] A. H. Jones and F. A. Bellis, "Digital Stereo Sound Recorder," Wireless World, vol. 78, no. 1443, pp. 432–434, September 1972.

[8] M. A. Perry, "High-speed p.c.m. challenges conventional f.m. analogue recording," Electronic Engineering, vol. 45, no. 539, pp. 33–35, January 1973.

[9] W. E. Anderton, "Professional Sound Recording," Wireless World, vol. 80, no. 1462, pp. 211–214,

June 1974.

[10] F. A. Bellis and M. R. Brookhart, "An Error Correcting System for a Multichannel Digital Audio Recorder," presented at the 58th AES Convention, New York, November 4–7, 1977; AES Preprint No. 1298.

[11] "Japan Readying Ultrahigh-Fidelity Stereo System Built Around PCM Disk," Electronics, vol. 50, no. 20, pp. 42–43, September 29, 1977.

[12] "Sony PCM Unit Converts Betamax to Stereo Recorder," Electronics, vol. 50, no. 19, p. 60, September 15, 1977.

[13] Gerald M. Walker, "Digital Takeover Extends to TV and Auto Controls," Electronics, vol. 50, no. 22, pp. 140–146, October 27, 1977.

[14] Charles L. Cohen, "Audio Gets the Word: Digital," Electronics, vol. 50, no. 24, pp. 78–79, November 24, 1977.

[15] "A Third Firm Shows Audio Tape Systems Based on VTRs," Electronics, vol. 50, no. 20, p. 63, September 29, 1977.

[16] T. Iwasawa, A. Ogawa and T. Sato, "Development of the PCM Laser Sound Disc and Player," presented at the 58th AES Convention, New York, November 4–7, 1977; AES Preprint No. 1309.

[17] "PCM Technology Applied to a Novel Laser Disc Player," TEAC Corporation of America, P. O. Box 750, Montebello, CA 90640.

[18] "PCM Technology Applied to a Novel Laser Disc Player," Audio Topics AES 77-10, Mitsubishi Electric Corporation, Mitsubishi Denki Bldg., Marunouchi, Tokyo 100.

[19] "Matsushita Shows Its New Video Disc," Electronics, vol. 50, no. 25, pp. 38, 40, 42, December 8, 1977.

[20] "Making the Advantages of PCM Recording Available to the Professional Audio Engineer," Audio Topics AES 77-10, Mitsubishi Electric Corporation, Mitsubishi Denki Bldg., Marunouchi, Tokyo 100.

[21] "Stop-Press News of our New PCM Cassette Tape Deck," Audio Topics, AES Suppl. 77-10, Mitsubishi Electric Corporation, Mitsubishi Denki Bldg., Marunouchi, Tokyo 100.

[22] "Four Channel PCM Processor Developed for Use with U-matic Cassette VTR," TEAC Technical Information No. 33204, November 1977, TEAC Corporation of America, P.O. Box 750, Montebello, CA 90640.

[23] "Four Channel Digital Tape Recorder Features and Specifications," Soundstream, Inc., 375 Chipeta Way, Suite 2a, Salt Lake City, UT 84108, November 1977.

[24] "Digital Recording System," Soundstream, Inc., 375 Chipeta Way, Suite 2a, Salt Lake City, UT 84108

[25] John A. McCracken, "A High Performance Digital Audio Recorder," presented at the 58th AES Convention, New York, November 4–7, 1977, AES Preprint No. 1268.

BIBLIOGRAPHY

[26] L. Klein and R. Hodges, "Audiovideo," Stereo Review, vol. 39, no. 6, p. 41, December 1977.

[27] T. G. Stockham, Jr., "Records of the Future," J. Audio Eng. Soc., vol. 25, no. 10/11, pp. 892–895, October/November 1977.

[28] John P. Myers and Abe Feinberg, "High Quality Professional Recording Using New Digital Techniques,"

J. Audio Eng. Soc., vol. 20, no. 8, p. 622, August 1972.

[29] R. B. Ingebretsen, "A Strategy for Automated Editing of Digital Recordings," presented at the 58th AES Convention, New York, November 4–7, 1977, AES Preprint No. 1303.

[30] F. Granum, "Design Criteria for Digital Audio Tape," presented at the 58th AES Convention, New York, November 4–7, 1977, AES Preprint No. 1279.

[31] SP5 P. G. Hutson, Jr., "New Technology: The Impact of Digital Tape," presented at the 58th AES Convention, New York, November 4–7, 1977; AES Preprint No. 1269.

[32] R. Steele, "Chip Delta Modulators Revive Designers' Interest," *Electronics*, vol. 50, no. 1, pp. 86–93, October 13, 1977.

[33] C. L. Cohen, "Two Formats Go Head to Head in Consumer VTR Confrontation," *Electronics*, vol. 50, no. 24, pp. 106–113, November 24, 1977.

[34] Y. Tsunoda, S. Sawano, H. Nakamura, K. Saito, T. Tsukada and Y. Takeda, "Semiconductor Laser Pickup for Optical Video Disk Player," *IEEE Trans. Consumer Electronics*, vol. CE-23, no. 4, pp. 479–493, November 1977.

[35] R. Adler, "Video Disc System Alternatives," *IEEE Trans*. Consumer Electronics, vol. CE-22, no. 4, p. 302, November 1976.

[36] J. S. Winslow, "Mastering and Replication of Reflective Videodiscs," *IEEE Trans. Consumer Electronics*, vol. CE-22, no. 4, p. 318, November 1976.

[37] G. C. Kenney, "Special Purpose Applications of the Optical Videodisc System," *IEEE Trans. Consumer Electronics*, vol. CE-22, no. 4, p. 327, November 1976.

[38] J. C. Mallinson and J. W. Miller, "Optimal Codes for Digital Magnetic Recording," *The Radio and Electronic Engineer*, vol. 47, no. 4, p. 172, April 1977.

[39] F. F. Lee and D. Lipschutz, "Floating-Point Encoding for Transcription of High-Fidelity Audio Signals," *J. Audio Eng. Soc.*, vol. 25, no. 5, p. 266, May 1977.

[40] D. M. Freeman, "Slewing Distortion in Digital-to Analog Conversion," *J. Audio Eng. Soc.*, vol. 25, no. 4, April 1977.

AUTHOR'S NOTE

Some of the data presented here were reported to me verbally, not necessarily by the proprietors of particular systems, and I have indicated this where applicable. I would appreciate any comments or corrections to such data by the manufacturers concerned.

Some Criteria for the Selection for Sampling Rates in Digital Audio Systems

ALASTAIR HEASLETT

Ampex Corporation, Redwood City, CA 94063

1. It has been reasonably well established that the sampling rate for a program-originating professional digital audio system should lie in the range of 45 kHz to 60 kHz. This range derives immediately from the stipulation of a minimum bandwidth of 20 kHz. Sampling must take place at at least twice this frequency, or 40 kHz. For practicality a 45-kHz sampling rate implies that the input anti-aliasing filter should have the required attenuation at the Nyquist rate of 22.5 kHz, to prevent aliasing components from appearing in the audio passband. Typically, the filter will be a minimum of 50 dB to 60 dB down at the Nyquist rate.

The two extremes of sampling rate given here represent a reasonable range of compromise between the cost and/or implementation of high (attenuation) rate filters on the one hand, and the cost and/or implementation of wide dynamic range/high signal to noise ratio analog to digital converters on the other.

Variable sampling rate systems are not considered in detail in this paper in the choice of a standard sampling rate. However, it is pertinent to note that any special effects which may be obtained by grossly varying the sampling rate, may also be obtained by purely digital methods. This paper only looks at primary criteria for selection of a *standard* sampling frequency.

2. In virtually all mastering recording applications, extensive use is made of overdubbing, punch in, or insert editing procedures. In an analog recording the merge between old and new recorded material is aided by finite bias and erase current ramp up/down times. In the case of timed record/erase delay systems the actual transition between original and inserted material is extremely difficult to detect unless pure tones are used.

However, in the case of a digital audio recorder, random entry into, or exit from, a recording mode in the middle of previously recorded material will cause substantial disturbance to the data clock extractor, invariably resulting in a momentary loss of sync and a consequent loss of data. The area of tape immediately adjacent to record entry and exit points will thus have, in effect, a permanent dropout. Any reasonable error-correcting scheme may clearly cope with this, and the results will be inaudible.

In many instances, it is found that an insert may be attempted several times. This process would mean that some area around the record entry point and around the record exit point will have a considerable number of deliberately introduced permanent drop outs. If such drop-outs become too numerous, they may cause a failure of the error-correcting system, requiring a concealment technique to avoid audible effects.

In addition, it may be necessary at times to make very precise edits, on a sample by sample basis. Such edits would most conveniently be performed "off line", that is, in an outboard editing system. Hence, there would arise a need to be able to create a splice between two digital data streams, without damaging any data on the tape.

The obvious solution, which is employed in the computer data recording industry, is to arrange the data on the tape in block form, with gaps which contain no data between the blocks. Record entry or exit in these gaps would thus cause absolutely no loss or damage to existing data before record entry or after record exit. With this approach it is then possible to retrieve even a single block of data from the tape, perform complex digital operations on the data in an outboard processing system, and return the modified block to the same space in the tape which it originally occupied, without disturbing or modifying adjacent data blocks in any manner whatsoever.

3. Based upon the notion of arranging the data on the tape in blocks, the next requirement is to attempt to define some boundary range of block size.

The gaps between blocks or inter block gaps (IBG) will have their size largely determined by: 1) The worst case time required to resynchronize bit synchronizers at the beginning of each block; and 2) The uncertainty of mechanically locating the IBG prior to punching into record (assuming a separate record and reproduce head is used).

In addition, the IBG length represents tape unused for data, and hence adds to the overhead for error correction and synchronization. Thus, its length relative to the useful data in a block becomes important.

It should be noted that the finite size of the block will in effect cause a latency time between any record command and the actual entry of that channel into record mode. This latency time will have a maximum value of one block length. In a typical analog system there is also a latency time caused by the required operation of relays, or electronic switches, followed by the ramp time of the bias. In the case of a timed record erase delay system, the (latency) time is extended by the time required for the tape to travel between the erase and record heads. The overall time might be as much as 120 ms, for a 1.5-inch (38-mm) erase-to-record-head spacing at 15 in/s (381 mm/s) tape

speed, or as little as 30 ms for normal nondelayed bias. However, in the latter case, the familiar overwrite on punch in, and gap on punch out, will occur, leaving a gap or overlap time of about 100 ms. From this, it would appear that a latency time of 1 to 10 ms would be very attractive and essentially unnoticed in a digital system.

If the lower limit of a 1-ms block size is considered and we assume that 10% of the block period is allocated to the IBG, then at a 30-in/s (7620-mm/s) tape speed, the block is 30 mil long and the IBG only 3 mil long. Consideration of mechanical tolerances of record head/reproduce head placement, tape stretch, and machine-to-machine interchange suggest that this IBG length is probably as small as can be reasonably supported. Finally it seems quite reasonable that each block have an integer number of samples contained in it and that each block be basically self dependent in terms of error correction capability.

It should also be noted that the concept of blocked data is one which is internal to the particular system. Externally, the only evidence of a blocked format is perhaps the finite latency time on random record entry commands. External digital inputs or outputs would be seen on a sample-by-sample basis only.

4. The final major criterion for choice of a sampling rate is undoubtedly the desire to be able to synchronize the digital audio recorder to the major video standards and to film-only systems. This synchronization should ideally involve *no* compromise in the absolute timing and pitch of the digital audio being reproduced or recorded, limited only by the absolute accuracy of the reference frequencies of the different standards.

It should be recognized that for digital audio recorders to enjoy the most widespread use, commonality of operation, and interchangeability of recordings on a worldwide basis are extremely important.

There are perhaps two levels at which such synchronous operation is desirable:

1) The sampling rate should be such that *digital* audio information can be synchronously received from and transmitted to future digital video recorders of either major video standard, that is, NTSC 525 line, or the PAL/SECAM 25 frame per second (625) line standards; and

2) The block rate should be such that editing decisions based on either nominally 30-Hz NTSC frame boundaries, 25-Hz PAL/SECAM frame boundaries, or 24-Hz film frame boundaries, in a composite audio-video system, should create edits on the synchronized digital audio system which correspond very closely to the internal IBG areas of the audio system, that is, the audio latency time of such an edit should be repeatable for any one edit point.

This leads to the conclusion that the block rate chosen should provide integer numbers of blocks for each frame rate in common use, or, nominally 30 Hz, 25 Hz and 24 Hz. The block rate should certainly also be derivable from the sampling rate and thus also from the system master clock.

5. The actual NTSC frame rate is not exactly 30 Hz. To derive the actual frequency and also the actual horizontal

rate for NTSC video systems requires a slight diversion.

The color subcarrier frequency f_{sc} was desired to be an odd submultiple of the new horizontal scan rate, H, that is

$$f_{sc} = A \times H/2 \tag{1}$$

where A is odd.
In addition, the beat frequency between the existing (and to be retained) audio subcarrier of 4.500 MHz and the color subcarrier, was also desired to be an odd submultiple of the horizontal rate, H, that is

$$f_{sa} - f_{sc} = B \times H/2 \tag{2}$$

where B is odd.
The values chosen for A and B were $A = 455$ and $B = 117$. Combining Eqs. (1) and (2),

$$H_{NTSC} = \frac{4.5 \times 10^6 \times 2}{A + B}$$

$$= \frac{4.5 \times 10^6}{286}$$

$$= 15734.26573 \text{ Hz to 5 decimal place accuracy.}$$

This immediately yields a frame rate of $H_{NTSC} \div 525 = 29.97002996$ Hz to 8 decimal places. In addition it yields the fact that an exact integer division from the audio subcarrier frequency defines the NTSC horizontal rate, that is 2.25 MHz $\div$ 143.

Casual inspection of the exact NTSC frame rate, the 25-Hz frame rate for 625-line PAL and SECAM systems and the 24-Hz film system frame rate, shows that there is no possibility of integer numbers of blocks for each frame standard. If, therefore, it is considered that the same latency time for editing as defined in Section 3 is acceptable, it will suffice to assume a frame rate of 30 Hz.

The lowest block rate yielding integer blocks per frame in each standard is 600 Hz. This gives:

Frame Rate	Blocks/Frame
30	20
25	24
24	25

Thus, in a synchronous system, the latency for 25 f/s and 24 f/s would be fixed and the actual latency in an NTSC system would change frame by frame, but would always be less than 1 block. This block rate must be relatable to the sampling rate chosen, since we require an integer number of samples per block. It must also possess an integer relationship with some master clock frequency, which itself has an integer relationship with both NTSC standards, 625-line PAL/SECAM standards and 24-Hz film frame rates.

It happens that the lowest common multiple of the NTSC horizontal rate, and the 625-line horizontal rate, is the 2.25 MHz derived previously. Therefore,

$$H_{NTSC} \times 143 = 2.25 \text{ MHz}$$

$$H_{625} \times 144 = 2.25 \text{ MHz}$$

where $H_{625} = 15625$ Hz.

Thus if f_B is the block rate and f_S the sample rate, then

$$f_S = Nf_B$$

where N is an integer, and then

$$f_S = \frac{K \times 2.25}{M} \times 10^6 \, \text{Hz}$$

from which

$$\frac{MN}{K} = \frac{2.25 \times 10^6}{f_B}$$

where M, N, and K are integers.

For example, if $f_B = 600$ Hz, $MN/K = 3750$.

The inclusion of K simply permits the master clock frequency to possess an integer relationship with the LCM of the video horizontal rates. If any other block rate is used, it may not possess a simple integer relationship to the frame rates, unless it were a multiple of 600 Hz. However, since the product M and N must be an integer, then:

$$\frac{K \times 2.25 \times 10^6}{f_B} = MN$$

where M and N are integers.

Thus, for any value of K there will be a finite number of values for f_B which lie between $f_B = 100$ Hz and $f_B = 1500$ Hz.

For simplicity, if only integer block rates between these limits are examined, which will ensure that only relatively simple relationships exist between the frame rates of nominally 30 Hz, 25 Hz and 24 Hz, then it can be seen that since the number of samples per block N, must be an integer, then the sampling rate f_S, will also be an integer.

Table I shows all possible integer values of f_S between 45 kHz and 60 kHz and indicates the number of samples per block N, and the block rate f_B associated with the particular value of N. Those block rates giving integer numbers of blocks for one, two or three frames for the frame rates of nominally 30 Hz, 25 Hz and 24 Hz are marked with an asterisk. Of the nine sampling frequencies, four give ten or fewer block rates to choose between, three have 21 choices, one has 25 choices and one has 33 choices. To obtain maximum flexibility in the choice of block size it would seem that 54 kHz as a sampling rate would be the best choice, with 48 kHz running second and 60 kHz, 52.5 kHz and 45 kHz running an equal third.

Table I. Sample rates between 45 kHz and 60 kHz for integer block rates and N.

Values of Integer N for Block Rates f_B (Hz)	Sampling Rate f_S (kHz)									
	45.000	46.875	48.000	50.000	50.625	52.500	54.000	56.250	60.000	
100	450		480	500		525	540		600	
105						500				
108							500			
120	375		400				450		500	
125	360	375	384	400	405	420	432	450	480	
128			375							
135					375		400			
140						375				
144							375			
150	300		320			350	360	375	400	
160			300						375	
175						300				
180	250						300			
192			250							
200	225		240	250			270		300	
210						250				
216							250			
225	200				225		240	250		
240			200				225		250	
250	180		192	200		210	216	225	240	
270							200			
300	150		160			175	180		200	*
320			150							
350						150				
360	125						150			
375	120	125	128		135	140	144	150	160	
384			125							
400			120	125			135		150	*
405					125					
420						125				
432							125			
450	100						120	125		
480			100						125	
500	90		96	100		105	108		120	
525						100				
540							100			
600	75		80				90		100	*
625	72	75		80	81	84		90	96	
640			75							
675					75		80			
700						75				
720							75			
750	60		64			70	72	75	80	
800			60						75	
875						60				
900	50						60			
960			50							
1000	45		48	50			54		60	*
1050						50				
1080							50			
1125	40				45		48	50		
1200			40				45		50	*
1250	36			40		42		45	48	