Digital Audio Standards

REPORT OF THE AES DIGITAL AUDIO MEETING

The AES held its second committee meeting on digital audio on 1978 February 1 and 2, in Atlanta, Georgia, at the time of the SMPTE Television Conference. Thirty-four members attended from the U.S.A., Japan, and Europe. The meeting was chaired by John G. McKnight of Magnetic Reference Laboratory.

A report of the previous meeting, held 1977 December 1 and 2 in Salt Lake City, was presented (see *J. Audio Eng. Soc.*, Jan/Feb. 1978, p. 52).

Then a report titled "A Review of Digital Audio Techniques," prepared by M. Willcocks was received *(J. Audio Eng. Soc.*, Jan/Feb. 1978, p. 56).

The discussion of sampling frequency (begun at the Salt Lake City meeting) was continued. Several reports were presented proposing the use of a 44.05594-kHz sampling frequency for all consumer and professional applications. They included: "Sampling Frequency Considerations" by M. Kosaka (p. 234); "On Several Standards of Forms for Converting PCM Signals into Video Signals" by T. Doi et al. (to be published in a future issue of the *Journal*); "Sampling Frequency Considerations" by K. Tanaka and Y. Ishida (p. 248); and "Sampling Frequency Considerations in Digital Audio Standards" by T. Muraoka et al. (p. 252). The committee felt that 44.05594 kHz would be an appropriate sampling frequency for use with low-bandwidth rotary-head type video recorders.

A. Heaslett's report titled "Some Criteria for the Selection for Sampling Rates in Digital Audio Systems," and R. Youngquist's report, "Sampling Frequency Considerations," *(J. Audio Eng. Soc.*, Jan/Feb. 1978, p. 66 and p. 54 respectively) were discussed. After consideration of the requirements for variable-speed recording, and for international interchange between all media, the committee felt that a second sampling frequency (in addition to 44.05594 kHz) such as 50 kHz or 54 kHz, would be necessary. From the point of view of digital signal processing, the group agreed that 50 kHz appeared to be more desirable, but further studies will be presented at the next meeting.

The discussion of source encoding (D/A conversion) begun at Salt Lake City was continued. There was a consensus of opinion on certain principles: For main-channel applications in professional equipment, the digital word should be a 16-bit, 2's compliment, format; pre- or de-emphasis should not be used; the aperture correction (sin x/x) should be at the D/A side; positive internal digital numbers should represent positive analog quantities; the polarity of analog input signals should be preserved throughout the digital system to the analog output; the equipment may be designed such that it does not set or detect least significant bits in the digital words; unused bits should be fixed; and the exact method for converting the words from and to analog signals should be at the discretion of the designer.

P. K. Burkowitz persented a report, "Users' Note on Digital Audio," on operational requirements for digital recorders (p. 242). This pointed out the need for measurement techniques for quantization noise, headroom, distortion of the signal, and distortion caused by out-of-band signals. It also led to a discussion of requirements for editing (particularly the "punch in" mode), and questions about whether the data needs to be in block form in order not to lose data on "punch in" and "punch out." Reports on these subjects are to be prepared for the meeting which will take place on 1978 April 29 and 30 in conjunction with the Los Angeles AES Convention.

A meeting was held on 1978 March 1 during the AES Convention in Hamburg. Han Tendeloo of Polygram was the chairman and a report will be published in a future issue of the *Journal*.

JOHN G. MCKNIGHT
Chairman

COMMENTS ON "TECHNICAL FACTORS TO BE STANDARDIZED"

The author would like to make the following comments on the technical factors to be standardized as proposed during the meeting on 1977 December 1 and 2. The numbers refer to the points as listed in the references.[1]

1) If more than one frequency is standardized, they should have a common denominator or common multiple, which can easily be sync locked to other frequencies generated or used in the system for signal processing purposes. Variable sampling frequencies, for example, as a function of other parameters, should be avoided.

2) In the given context "PCM" should be replaced by "quantization." Nonlinear quantization will generate additional quantization noise. Floating point should be avoided since it is not sufficiently precise.

3) For regular data recorders preferably phase encoding (PE); for video recorders preferably direct pulse registration. Basic Miller and non-return-to-zero (NRZ) are outdated. Sync, depending upon storage medium; redundancy, not realistic for music; error correction, in any case, depending upon type of interference.

5) Both are recommendable, 5 through 12 volts.

7) Program level measurement should by all means be digital, and preferably converter capacity related.

8) Possibly independent of the system, related to generalized quality parameters.

9) Strongly recommendable. For some time to come no abbreviations should be stated without sufficient interpretation.

P. K. BURKOWITZ
Polygram B.V.
Baarn, The Netherlands

[1] *J. Audio Eng. Soc.*, vol. 26, p. 54 (Jan/Feb. 1978).

Sampling Frequency Considerations

MASAHIRO KOSAKA

Wireless Research Laboratory, Matsushita Electric Industry Company, Ltd., Osaka, Japan

1. INTRODUCTION: This report describes what should be considered in determining the sampling frequency to be employed for sampling audio signals in PCM tape recorders for consumer use. The considerations are based on the premise that standard video tape recorders (VTR) designed to record/reproduce usual TV signals or equipment equivalent to such standard VTRs are used for the PCM tape recorders. This report also proposes 44.05594406. . . kHz as an optimum sampling frequency.

2. TECHNICAL REQUIREMENTS

The following technical requirements are necessary to be considered in determining the sampling frequency for sampling audio signals for consumer PCM tape recorders.

2.1 The PCM tape recorders should be capable of satisfactorily sampling the whole range of dc to 20 kHz as a transmission band for audio signals.

2.2 The design and manufacture of antialiasing filters to prevent alias noises should be easy so as to reduce the cost of antialiasing filters to an appropriate cost for consumer use.

2.3 The conversion speed (data speed) of analog-to-digital (A/D) converters including sample-hold circuits should not be excessively high.

2.3.1 The conversion speed of A/D converters directly and greatly influences the cost of the A/D converters.

2.3.2 If the conversion speed of A/D converters is excessively high, data signals cannot be stably or easily recorded/reproduced by small consumer VTRs with rotary heads.

2.4 Usual VTRs for processing TV signals should be able to be used as PCM tape recorders without necessitating any modification.

2.5 The PCM tape recorders should be able to reproduce recorded signals with high time accuracy.

2.6 The circuit arrangement of the PCM tape recorders including various timing signal generators should be as simple and inexpensive as possible for consumer use.

3. DETERMINATION OF SAMPLING FREQUENCY

Assuming that usual small consumer VTRs with rotary heads for standard NTSC systems are used for the PCM tape recorders, the sampling frequency of sampling pulses to meet the above technical requirements can be determined by the following procedure.

In the NTSC system the horizontal and vertical scanning frequencies f_H and f_V are

$$f_H = 4.5 \text{ MHz}/(572/2) = 1.57342657343 \ldots \text{ MHz}$$

$$f_V = f_H \times 2/525 = 59.9400599402 \ldots \text{ Hz}.$$

Assuming that a sampled analog signal concerning one channel of two-channel signals (stereophonic signals) is converted to a digital signal with a resolution of 13 bits, a digital datum of 26 bits composed of two parallel digital signals constitutes one word with regard to both channels of the stereophonic signals:

1 word = 26 bits.

Upon converting the 26-bit digital datum by time sharing to a serial signal and inserting the integral number of such serial signals to a horizontal scanning period $T_H = (1/f_H)$, the bit rate (BR) of the thus inserted signals is

$$BR = N \times f_H \times 26 \times (1/K_{BL}) \times K_R$$

where N is the number (integer) of words in $1H$; K_{BL} is the effective time ratio of T_H, that is, the ratio of the T_H minimum nonrecordable period (the sum of the horizontal and vertical blanking periods, head switching time, etc.) to T_H (usually $K_{BL} \simeq 0.78$); and K_R is the redundancy representing bits (on the basis of bits of inserted serial signals) necessary to correct data error due to, for example, dropout (usually $K_R \simeq 1.5$ to 2.0).

Assuming that $K_{BL} = 0.78$ and $K_R = 1.5$, BR values for various N values can be obtained by calculation:

N	BR
1	0.786713×10^6
2	1.57343×10^6
3	2.36014×10^6
4	3.14685×10^6
5	3.93357×10^6
.	.
.	.
.	.

From the preceding table it is apparent that when N is 4 or more, BR is higher than 3 MHz. If BR is higher than 3 MHz, data signals cannot be stably or easily recorded/reproduced by usual small-size VTRs. On the other hand, when N is 1 or 2, BR is lower than 2 MHz, which is too low to obtain a satisfactorily high sampling frequency, although recording/reproduction can be easily performed.

Therefore N should be 3, namely, the optimum number of words to be inserted in $1H$ is 3, that is, the sampling frequency f_s should be

$$f_s = 3 \times (525/2) \times f_V.$$

However, as set forth above, there is a nonrecordable period in each frame. Putting D to be the nonrecordable period in one frame per T_H, the above equation should be modified to

$$\begin{aligned} f_s &= 3 \times [(525 - D)/2] \times f_V \\ &= 3f_H\,(525 - D)/525. \end{aligned}$$

Each of f_s, f_H, and f_V is required to be stable, and the relation among f_s, f_H, and f_V is required to always satisfy the above equation. Therefore it is necessary to generate these f_s, f_H, and f_V from a stable single master signal by frequency division. Thus such a single master signal should have a frequency of a common multiple of f_s and f_H. The common multiple is preferred to be as small as possible, of course.

However, a mere least common multiple (LCM) of f_s and f_H is not satisfactory. In the standard NTSC system there is a timing difference of $0.5H$ between odd- and even-number fields. Therefore the single master signal is required to have a cycle period of one-half of that of the horizontal synchronizing pulses, that is, a frequency of two times f_H ($2f_H$).

Similarly, the sampling frequency is required to have a frequency of two times f_s ($2f_s$) in order to alternately sample signals of both channels of the stereophonic signals.

Therefore the single master signal should have a frequency of the LCM of $2f_s$ and $2f_H$. Among the LCMs of $2f_s$ and $2f_H$ with various D values it has been found by calculation that the most easy frequency to produce as master signal is obtained when $D = 35$ (see Table 1). The sampling frequency at $D = 35$ is 44.05594406 . . . kHz. Accordingly, 44.05594406 . . . kHz has been found to be an optimum sampling frequency.

Table 1. Calculation of LCM of $2f_s$ and $2f_H$.

```
Fh=
 1.573426573e 04
Fv=
 5.994005994e 01

D=            20

Fs=
 4.540459540e 04
Fm=
 3.178321678e 06

     M=      101
     N=       35
```

```
D=            21

Fs=
 4.531468531e 04
Fm=
 4.531468531e 06

     M=      144
     N=       50

D=            22

Fs=
 4.522477522e 04
Fm=
 1.582867133e 07

     M=      503
     N=      175
```

```
D=            23

Fs=
 4.513486514e 04
Fm=
 3.159440559e 07

     M=     1004
     N=      350

D=            24

Fs=
 4.504495505e 04
Fm=
 1.576573427e 07

     M=      501
     N=      175
```

Table 1 con't.

D=	Fs=	Fm=	M=	N=
25	4.495504496e 04	6.293706294e 05	20	7
26	4.486513487e 04	3.140559441e 07	998	350
27	4.477522478e 04	1.567132867e 07	498	175
28	4.468531469e 04	2.234265734e 06	71	25
29	4.459540460e 04	3.121678322e 07	992	350
30	4.450549451e 04	3.115384615e 06	99	35
31	4.441558442e 04	1.554545455e 07	494	175
32	4.432567433e 04	4.654195804e 07	1479	525
33	4.423576424e 04	1.548251748e 07	492	175
34	4.414585415e 04	1.545104895e 07	491	175
35	4.405594406e 04	4.405594406e 05	14	5
36	4.396603397e 04	4.616433566e 07	1467	525
37	4.387612388e 04	1.535664336e 07	488	175
38	4.378621379e 04	1.532517483e 07	487	175
39	4.369630370e 04	4.588111888e 07	1458	525
40	4.360639361e 04	3.052447552e 06	97	35
41	4.351648352e 04	1.523076923e 07	484	175
42	4.342657343e 04	4.342657343e 06	138	50

Table 1 con't.

D=	Fs=	Fm=	M=	N=
43	4.333666334e 04	1.516783217e 07	482	175
44	4.324675325e 04	1.513636364e 07	481	175
45	4.315684316e 04	3.020979021e 06	96	35
46	4.306693307e 04	1.507342657e 07	479	175
47	4.297702298e 04	1.504195804e 07	478	175
48	4.288711289e 04	3.002097902e 07	954	350
49	4.279720280e 04	2.139860140e 06	68	25
50	4.270729271e 04	5.970029979e 05	19	7
51	4.261738262e 04	4.474825175e 07	1422	525
52	4.252747253e 04	2.976923077e 07	946	350
53	4.243756244e 04	1.485314685e 07	472	175
54	4.234765235e 04	1.482167832e 07	471	175
55	4.225774226e 04	5.916083916e 06	188	70
56	4.216783217e 04	2.108391608e 06	67	25
57	4.207792208e 04	1.472727273e 07	468	175
58	4.198801199e 04	4.408741259e 07	1401	525
59	4.189810190e 04	1.466433566e 07	466	175
60	4.180819181e 04	5.853146853e 06	186	70

STEPHEN M. HOCKSTEIN

Music Masters, New York, NY 10036

I am writing to call your attention to an error which exists in the article entitled "Sampling Frequency Considerations."[1]

On page 234 in section 3 "Determination of Sampling Frequency," the first equation $f_H = 4.5$ MHz/(572/2) = 1.57342657343 . . . MHz. On checking the mathematics you will find that the correct expression is $f_H = 4.5$ MHz/(572/2) = 15.7342657343 . . . kHz.

[1] M. Kosaka, "Sampling Frequency Considerations," *J. Audio Eng. Soc.*, vol. 26, pp. 234–240 (April 1978).

User's Note on Digital Audio Standard Code and Procedures, Operational Requirements

P. K. Burkowitz

Polygram B.V., Baarn, The Netherlands

1. INTRODUCTION: As an industrial combine that works worldwide in the technical recording field, we feel that the upcoming era of digital audio urgently deserves an early attempt at the standardization of codes and formats and an international agreement on working modes.

This will be more important the stronger all participants expect to continue, also in a digitalized business, to exchange masters and to work on their recordings wherever they are, as they are currently accustomed to doing.

However, with the advent of the first lab sample machines, a number of noncompatible solutions has already appeared, greatly endangering the good auspices by an unnecessary and detrimental variety of codes and formats.

As we see it, it seems that this unfortunate trend has again been provoked (as was the case with quadraphony) by the lack of consumer specifications and requirements which have not early enough been compiled, harmonized internationally, and made know to everybody who intended to conceive hardware.

This summary of questions, remarks, and requirements is meant to serve the latter purpose, and, thus, hopefully will contribute to prevent the bright digital possibilities from being hampered for some time to come. It is hoped that this action may result in coordinated specifications that will fulfill the requirements and expectations of professional users of digital audio equipment.

2. SAMPLING FREQUENCY AND AUDIO BANDWIDTH

Sampling frequency and audio bandwidth most urgently require standardization. Without standard sampling no interchange of carriers will be possible, except by expensive (and unnecessary) code converters.

Pilot equipment has become known using 40, 44.056, 47.25, and 48 kHz as sampling frequencies. Other systems, such as for broadcasting lines, have recently been reported using around 14 kHz (speech communication) and 32 kHz (music transmission).

What (costly) procedure do the advocates of the various sampling proposals envisage when receiving tapes from place A to be processed at place B?

For quality recording the record industry may agree to only *one* standard sampling frequency which ensures a safe minimum audio bandwidth beyond the risks imposed by wideband audio signal processing. These risks are in the transients range above 15 kHz where today's technology carries the full content of unattenuated harmonics and overtones, not to speak of the distortion products caused by the usual overloading. As long as analog low-pass