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 filters are used in front of A/D converters, all these signal components will heterodyne along the steep skirt of this low-pass filter and will then generate downbeats in the audible signal. (Conventional analog and former tube techniques did not have that problem since they could allow the transients to go as far up as available, and the "end" of the available channel normally smoothed down pretty slowly with little danger for heterodyning.)

The alternatives are either cutting off the audio response safely before entering the A/D circuit (this appears not to be acceptable in light of the fact that a novelty will pay only if it is definitely better than anything previously available) or to allow for an upper band limit of, say, 50 kHz (this is "common practice" today in any good professional amplifier), which means a sampling frequency of at least 100 kHz.

Since this may be beyond realism, we invite a clear explanation by hardware specialists and, in consequence, an agreement among potential users on that particular point.

3. DYNAMIC CAPACITY AND QUANTIZATION APPROACH

General agreement appears to exist that a new generation of technology should at least give us the long waited for overall 80-dB dynamic capacity without any tricky conditioning. Taking multitrack into account, a minimum figure of 90 dB per channel would thus be preferable (the geometrical addition of 32 times −90 dB channel noise would result in a noise figure of −75 dB; the addition of 32 times random-phase 0 dB musical signals results in a signal level between +5 and +15 dB, making up for a total dynamic capacity of 80 to 90 dB).

Along with less than 0.05% distortion (which should go without saying) this would for the first time bring our mixing console audio signal quality direct to the cutting equipment. (Decibel figures as stated above would become less critical upon introduction of digital mixing techniques.) However, it is already to be heard that tests are going on to find out how the ear can possibly be deceived in order to save electronics. Conditioning tricks are thus under discussion from logarithmic through polygonal quantization.

We feel that this sort of conditioning must definitely stop with a technology that offers all the possibilities to avoid it. As to this we plead for a linear quantization approach.

4. HEADROOM

It is worth remembering that no spectacular disaster happened when, in a vacuum-tube circuit, peak amplitudes exceeded the channel capacities. Transistor circuits are much more critical in this respect since their headroom ends abruptly. Much more headroom must therefore be provided in transistor circuits. In A/D converters, now, the clipping effect (with all audible annoyance) is just "perfect." Consequently, in digital audio safe headroom is vital not only in terms of frequency response but also with regard to the operating level. Instantaneous true peak level reading thus becomes as vital as avoiding any overload. Overload simply destroys the program, while conventional equipment at least still provides some compliance. The same goes for the lower end where a kind of switch-off effect occurs which can well be audible even if the remaining noise floor is of comparable magnitude.

New hardware must give a clear answer to that point and also stipulate how much of the extended dynamic capacity will be consumed by all these considerations before the dynamic range is specified that really remains for music.

5. SPURIOUS FREQUENCIES

Frequencies generated in the equipment, or fed to it for signal processing purposes, may not cause any spurious interference within the audible range. No frequencies may be used which require particular authorization unless such authorization can be granted worldwide.

6. ERROR RECOGNITION AND SIGNAL RESTORATION

Error recognition and signal restoration shall be provided and made possible without repeated run of the carrier.

7. A/D – D/A CONVERTER CODE

Codes should be computer arithmetic compatible, such as the two's complement code.

8. TIME CODE

Digital audio and time code registration go hand in hand (see also Section 9). However, there are several time code proposals. One should be standardized for use in the recording industry.

We would prefer an advanced SMPTE/EBU code (resolution capability, time plus date and user bits) which has a chance to be used worldwide and poses no interface problems when coworking with TV organizations.

9. EDITING AND TAPE FORMAT

Most probably editing will only be possible electronically. In any case, editing must at least be possible in the same flexible and sometimes "artistic" manner as was manual editing, preferably with greater operational ease and also externally programmable. This requires that all takes which have to be considered must either be real-time addressable from an appropriate number of machines or sequentially stored and assembled via appropriate storage means.

These capabilities are a prerequisite for the operator; but what about sync locking of time code, sampling frequency, and clock frequency generators, particularly if a multi-recorder or -channel setup is working where probably no phase lag or heterodyning can (or should) be tolerated at any point?

The carrier format will certainly be determined by the combination of all these requirements. In our view, if it is tape, 2-inch (50-mm) wide tapes should by all means be avoided since they impose too many physical handling risks, except if handling could be mechanized almost completely such as, for example, with computer technology.

Tape speed as such is not a prime design parameter as long as the playing time per reel or cassette is sufficient, at least 40 minutes, and the reel diameter does not exceed usual standards. Naturally for a particular format the speed should be standarized also.

Generally, parameters dictated by editing requirements, tape format, and track utilization require strict standardization.

10. VOCABULARY

The continuous mixup of the words digitizing and PCM should be eliminated.

11. POLYGRAM PREFERRED REQUIREMENTS FOR DIGITAL RECORDING

The requirements for digital recording preferred by Polygram are listed in Table 1.

Table 1. Preferred requirements for digital recording.

Audio bandwidth	10 Hz–20 kHz ±0.2 dB
Audio level headroom	
Upper	20 dB above peak program reference level, or no audible compression, clipping, or distortion on up to 20 dB overshoots
Lower	20 dB below analog system noise floor, or no audible switch-off effects
Actual dynamic capacity	>56 dB per channel, available between analog system peak noise floor and peak program reference level
Total harmonic distortion (within specified ambient temperature range)	0.05% at any audio level; no audible sidebands
High-frequency headroom	No audible distortion products when amplitudes occur in nonaudible ranges
Amplitude stability (within specified ambient temperature range)	Reference to absolute accuracy of converters, equal to: short time (minute) <0.05 dB long time (hour) <0.1 dB
RF proof	No audible products from any high frequencies produced for any purpose in the equipment or combination of equipment; certain frequencies may require licensing by national authorities
Error proof	Codes must provide for defined (such as click perception related) processing safety on widely available tape brands or other carriers, comparable to the quality obtainable from top analog technology
Wow and flutter	< 0.005%
Tape width	If audio tape, 1 inch (25 mm); if computer/video tape, standard width
Tape speed	To be standardized; if recorder is based on computer or video recorder, standard speed
Continuous recording time	> 40 min
Editing	From take selection and correction of tone fragments through programmable assembling, independent from session time and date
Sync modes	Full sync capabilities to any other machine that meets specifications (such as by standardized unassigned bits within data stream)
Time code	Upgraded SMPTE/EBU should be included (see sync modes)
Mechanical design	Portable
Interface	Interface prepared for later direct data input from digital mixing desks

Sampling Frequency Consideration

KUNIMARO TANAKA AND YOSHINOBU ISHIDA

Mitsubishi Electric Corporation; Products Development Laboratory, Hyogo, Japan

1. INTRODUCTION: It is the opinion of these authors that the same sampling rate should be used in all digital audio systems. (If this is practically impossible, two standard frequencies might be chosen.) The reasons that we recommend 44.05594 kHz in the 525-line area and 44.1 kHz in the 625-line area as standard sampling rates are given in this report.

2. REQUIREMENTS

2.1 The sampling rate of digital audio systems should be selected so as to satisfy the requirement of a minimum bandwidth of 20 kHz, and it should also be higher than 40 kHz from the economical point of view in the design of the antialiasing filter.

2.2 The choice of the sampling rate is furthermore restricted to meet the specific requirements of each component in the digital audio system. The possible components in the digital audio system are the video tape recorder used as a PCM tape recorder, the film, PCM disk, and mixing console. Some requirements are of great influence while others have less or none.

2.3 The possible components in the digital audio system should be in synchronism with the video tape recorders (NTSC 525-line 30-Hz, PAL/SECAM 625-line 24-Hz) and film (24 Hz).

2.4 Video tape recorders (helical/transversal) can be used as PCM recorders for professional applications (for mastering and as portable versions) as well as for home use. Consequently the sampling rate should be chosen such that the PCM signal will be able to be converted into the pseudocomposite video signal keeping the following requirements.

1) The line frequency and the frame frequency should be identical with the standard composite video signal rate.

2) It seems quite reasonable that each line has an integer number of left and right samples contained in it.

3) Some blank space should be provided for helically scanning video tape recorders to incorporate vertical sync pulses and to avoid any damage of the PCM signal by head switching noise. The length L between the switching position and the leading edge of the vertical sync pulse (Fig. 1) is standardized for each type of video tape recorder (Table 1). The blank space should be greater than the maximum L in Table 1 plus $6H$.

Requirements 1) through 3) formulate the following two equations:

$$f_s = N \frac{(525 - x)}{525} \times 15.73426 \quad \text{[kHz] for 525-line}$$

$$f_s = N \frac{(625 - x)}{625} \times 15.625 \quad \text{[kHz] for 625-line}$$

where N is the total number of samples in a horizontal line; x is the sum of the blank spaces of the even and odd fields; and f_s is the sampling frequency.

Since the frequency of the master clock is the least common multiple (LCM) of Nf_H, f_s, and the bit clock, the numerator and denominator of f_s/Nf_H should be small integer numbers in order to facilitate the manufacturing of the crystal oscillator of the master clock. Table 2 shows some possible standard sampling rates applicable to both 525-line NTSC and 625-line PAL/SECAM with slight differences of 0.1% in between, which, we believe is imperceptible. (*Note*: Theoretically 525-line color BTSC never shares exactly the same sampling rate with 625-line PAS/SECAM.)

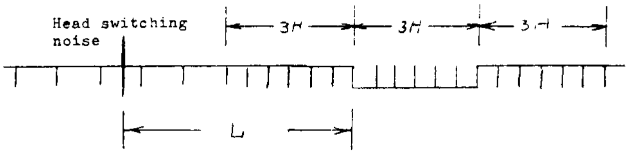


Fig. 1. Head switching point and vertical sync pulse.

2.5 In a few years the PCM disk and its player will be commercially available for home use at a very reasonable price. Since this system should obtain the PCM music source from a mastering PCM recorder, both should share exactly the same sampling frequency. For future development into long-play version as well as out of necessity the PCM disk should employ as low a sampling rate as possible within the guarantee of 0–20-kHz audio reproduction capability so that the bit density can be reduced. As noted above, the PCM disk is for home use. We suggest the use of low-price converters, under 12 bit, with logical companding schemes. The total bit number per sample should be chosen to be below about 13 or 14.

2.6 The film has 24 frames per second. As all digital audio systems suggest only one sampling rate, the chosen sampling rate should be in synchronism or at least in a simple integer relationship with 24 Hz.

Table 1. Length between switching position and leading edge of the vertical sync pulse.

Type of VTR	L
EIAJ type 1	$5 \pm 5H$
U matic	$5 \cdots 8H$
Vcord	$7 \pm 3H$
βmax	$7 \pm 2H$
VHS	$5 \cdots 8H$

3. SUGGESTED RATE

Based on the preceding discussions, we found 44.05594 kHz for the 525-line area and 44.1 kHz for the 625-line area as the best sampling frequencies. As noted, the difference between the two versions is only an imperceptible 0.1% which practically generates no problem in interfacing.

4. INTERFACING

Digital interfacing is desired, although analog interfacing is not really bad. Therefore the most important item to be standardized in the order of priorities is the sampling rate. It should be noticed that there is a variety of coding schemes, such as linear coding, nonlinear (companding/floating-point), and number of bits employed. With the use of only one standard sampling rate, the data conversion among them is, theoretically and practically, quite easy. We think basically that the interfacing word format should be linear because the word formats can be linearized for nonlinear coding. Furthermore, the total time slots per sample should be determined leaving ample margin for future technical advancement.

Table 2. Possible standard sampling rate.

	NTSC			PAL/SECAM			Difference (%)
	x	f_s/Nf_H	f_s(kHz)	x	f_s/Nf_H	f_s(kHz)	
$N = 3$	32	493/525	44.32565	34	591/625	44.325	−0.0016
	33	164/175	44.235755	35	118/125	44.25	+0.023
	35	14/15	44.055938	37	588/625	44.1	+0.101
	36	163/175	43.966026	39	586/625	43.95	−0.036
	37	488/525	43.876115	40	117/125	43.875	−0.0025
	38	487/525	43.786203	41	584/625	43.8	+0.0315
	39	162/175	43.606295	43	582/625	43.65	+0.1
	45	32/35	43.156834	49	576/625	43.2	+0.1
$N = 4$	32	493/525	59.100887	34	591/625	59.1	−0.0015
	33	164/175	58.981008	35	118/125	59.0	+0.0322
	35	14/15	58.741248	37	588/625	58.8	+1.0
	37	488/525	58.501487	40	117/125	58.5	−0.003
	40	20/21	58.141846	43	582/625	58.2	+0.1
	42	23/25	57.902087	46	579/625	57.9	−0.0036
	45	39/42	57.542446	49	576/625	57.6	+0.1
	47	478/525	57.302686	52	573/625	57.3	−0.0047
	50	19/21	56.943046	55	114/125	57.0	+0.1
	55	37/42	56.343645	61	564/625	56.4	+0.1
	60	18/21	55.744245	67	558/625	55.8	+0.1
	65	92/105	55.144843	73	552/625	55.2	+0.1
	70	13/15	54.545444	79	546/625	54.6	+0.1
	75	6/7	53.946044	85	108/125	54.0	+0.1

f_H—line frequency; f_s—sampling rate.

Sampling-Frequency Considerations in Digital Audio

TERUO MURAOKA, YOSHIHIKO YAMADA, AND MASAMI YAMAZAKI

Victor Company of Japan, Ltd. (JVC), *Yamato-shi, Kanagawa-ken, Japan*

The sampling frequency should be set at 44.05594 kHz for the following reasons.

1) Under the limited total bit capacity of the recording material, a lower sampling frequency is preferable to increase the recording time and channel and the reliability for protecting the drop-out noise with more sensitive parity codes.

2) According to the results of audibility tests, the 20-kHz bandwidth is considered to be quite sufficient to prevent deterioration of sound quality (including the effect of the phase characteristic of the low-pass filter; see Appendix). The sampling frequency of 44.056 kHz can cover this bandwidth.

3) From the viewpoint of introduction or use of digital-audio systems in broadcasting studios, it is advantageous that the sampling frequency be in the relationship of a rational number with the horizontal frequency of the video signal. The reasons are as follows.

a) No problems are expected when digital audio instruments are used in connection with video instruments.

b) Some video instruments (for example, VTR) can be used as part of the digital audio system.

c) Digitalized audio signals can easily be converted into television format signals and broadcast using television channels.

d) In the future video signals will be digitalized.

4) To convert the digitalized audio signal into the television format signal, it is necessary that the sampling frequency f_s and the horizontal frequency f_H be in the following relationship:

$$\frac{f_s}{Kf_H} = \frac{L - B}{L} = \frac{n}{m}$$

where K is the number of digitalized audio samples per horizontal line, L the total number of horizontal lines, B the number of horizontal lines in a vertical blanking period, and m and n are integers to be chosen (preferably small).

From the viewpoint of VTR application, $K = 3$, $B > 28$ are the conditions, and we get the optimum sampling frequency as 44.05594 kHz (45 kHz is too high).

Based upon the PAL standard, the sampling frequency becomes 44.100 kHz. The frequency difference between the NTSC block and the PAL block causes a timing error of 3 seconds per hour in the exchange of recorded programs. However, this error can be compensated in the digital domain by means of the sample-elimination (or doubling) technique without audible errors.

APPENDIX

REEXAMINATION OF HUMAN AUDIBLE BANDWIDTH

The study of human audible bandwidth dates back to the early experiment of W. B. Snow [1], followed by Garnet and Kerney [2] and Olson [3]. Those studies reached the result that the band limit of 40 Hz to 14 Hz is sufficient to reproduce musical programs without perceptible tone changes.

Recent improvements in audio devices have expanded the recording/reproducing frequency range, and developments of electronic musical instruments have increased the intensity of high-frequency components in the musical program. These improvements raise doubts about the correctness of the above-mentioned studies, given today's situation.

Current engineering of digital audio requires sampling and low-pass filtering, which has inevitable band-limit problems. With this background, let us reexamine the bandwidth of human audibility. In this appendix we discuss the effects of high-frequency cutoff.

Experiments

We designed the experiments to make the following items clear. (Note that in most digital audio systems the high end of the recording bandwidth is set at 20 kHz.)

1) Is the cutoff frequency of 20 kHz sufficient to keep original tone quality?

2) What frequency is the critical cut-off point with respect to human audibility?

For this experiment we used the blind A–B comparison method. The setup is shown in Fig. 1. The judging process is as follows.

According to Fig. 1, let the low-pass mode be A and the all-pass mode B. The listener is handed a blind switch box having 20 preset combinations of A and B and a status change switch. If the combination is A–A, he listens to the same sound at every status change. If the combination is A–B, he listens to different sounds in turn. He operates the combination selection and status switch by himself and completes the examination of 20 combinations.

According to the probability theory, if he selects 15 counts correctly, his discriminability is supported by more than 95% reliability.

In this examination the loudspeaker frequency response and the frequency analysis of the program source are important. We used a pair of specially designed loudspeakers; their frequency responses are shown in Figs. 2 and 3.

Prior to the examination we checked the discriminability which might be program-source dependent using white noise, real cymbal-crash sound (picked up with a calibrated condenser microphone), and prerecorded music. We found that the discriminability does not depend upon the kind of program source but upon the abundance of its high-frequency components.

Taking into account the statistical treatment discussed previously, we used the program source of rock music in which the Moog synthesizer is tweedling and the cymbals are crashing.

Fig. 4 shows a comparative display which proves the abundance of high-frequency signal components. For this experiment we made four pairs of seventh-order Chebyscheff-type low-pass filters having the cut-off frequencies of 14 kHz, 16 kHz, 18 kHz, and 20 kHz. As an example, the amplitude and group-delay characteristics and the circuit diagram are shown in Fig. 5 ($F_c = 20$ kHz). The cutoff is sharp enough, and the group delay increases remarkably at the cut-off point.

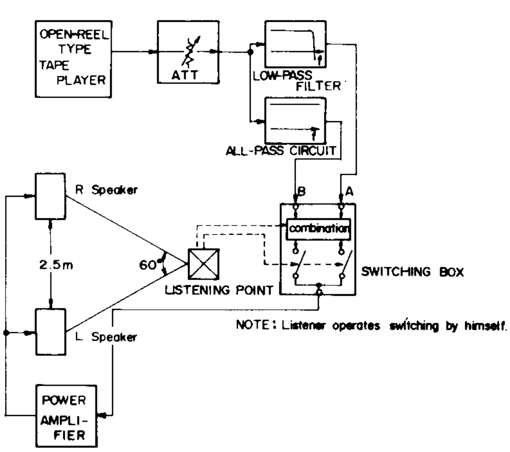
Results

We asked about 30 audio engineers to be judges. They have a wide experience in listening to sounds and are sensitive to musical material. For example, some of them are loudspeaker designers, amplifier designers, mastering operators, tape-recorder designers, some of them play instruments, etc.

First we carried out the 20-kHz cut-off examination, and next we did the 14-kHz cut-off examination. Further examinations of 16-kHz cutoff and 18-kHz cutoff were limited to those who selected more than 13 counts correctly. The results are shown in Fig. 6.

At 20-kHz cutoff, although it includes the effect caused by the phase characteristics of the low-pass filter, none of them could discriminate, the average correct count being 10 (random choice).

This suggests that hearing the effect of a 20-kHz cutoff is rare, and if the effect is heard, it may not deteriorate the



0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
B	B	A	A	A	B	B	A	A	B	A	B	B	A	B	A	A	B	B	A
A	B	A	A	A	A	A	B	B	B	B	B	B	A	A	B	A	B	A	B

An example of combinations

Fig. 1. Test setup for determining audibility of change of bandwidth due to low-pass filter.

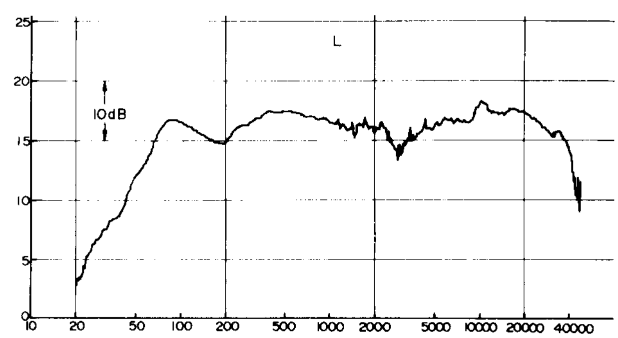


Fig. 2. Frequency response of left-channel loudspeaker.

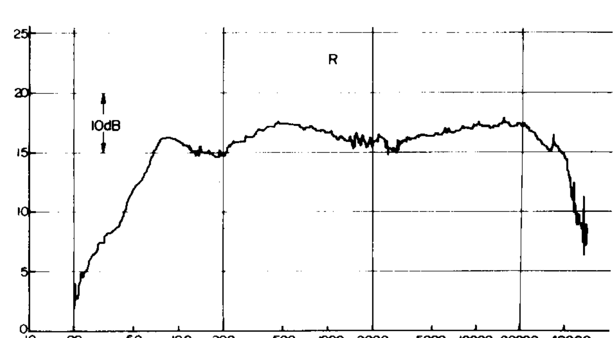


Fig. 3. Frequency response of right-channel loudspeaker.

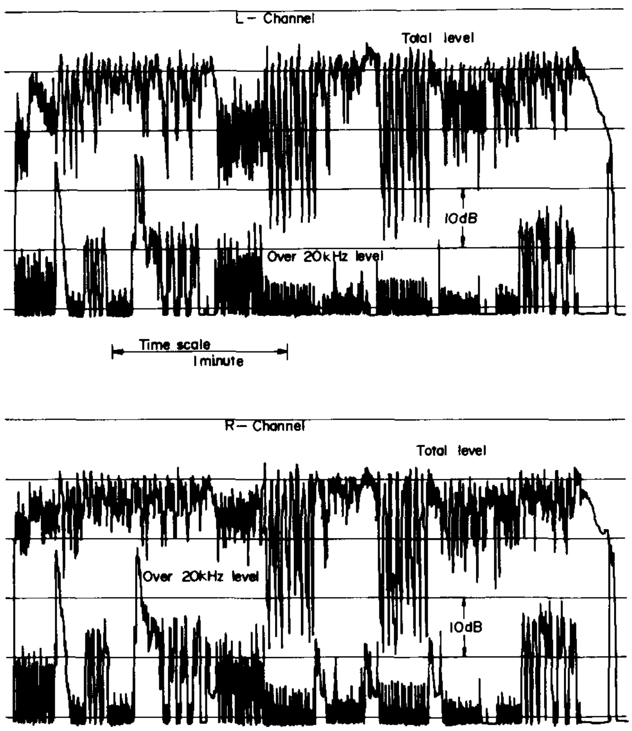


Fig. 4. Comparison of level of signals above 20 kHz with total signal level.

sound quality. Also from Fig. 6, the critical cutoff is considered to be around 15 kHz, which proves the earlier results by Snow *et al.* [1].

Conclusion

1) 20-kHz cutoff is considered to be quite sufficient, without deteriorating sound quality (including the effect caused by the phase characteristics of low-pass filters).

2) The critical cut-off frequency is considered to be about 15 kHz.

3) In designing digital audio systems, the sampling frequency should preferably be set to cover the 20-kHz bandwidth. The sampling frequency of 44.05594 kHz satisfies this requirement.

Human engineering is involved in the discussion of digital audio technology. It requires a reexamination of human audibility in today's recording which reflects

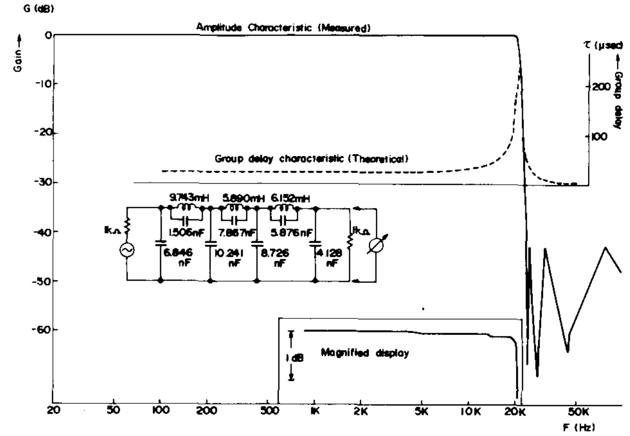


Fig. 5. Frequency response and group delay of the 20-kHz low-pass filter used in this experiment.

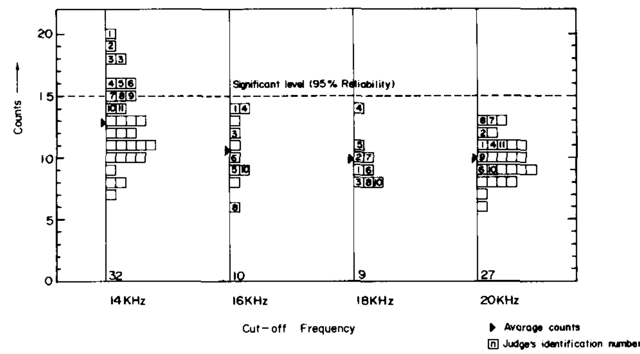


Fig. 6. Number of correct judgments of the effect of the insertion of a low-pass filter of the given cut-off frequency.

highly advanced playback equipment, modern rock music, synthesizers, etc. Authors are attempting to clarify this problem through studies in various phases.

REFERENCES

[1] W. B. Snow, "Audible Frequency Ranges of Music, Speech, and Music," *J. Acoust. Soc. Am.*, vol. 3, Pt. 1, pp. 155–166 (July 1931).

[2] D. K. Gannett and I. Kerney, "The Discernability of Changes in Program Bandwidth," *Bell Syst. Tech. J.*, vol. 23, pp. 1–10 (Jan. 1944).

[3] H. F. Olson, "Frequency Range Preference for Speech and Music," *J. Acoust. Soc. Am.*, vol. 19, pp. 549–555 (July 1974).