

STANDARDS AND INFORMATION DOCUMENTS

AES-R1-1997
s2008



AES project report for professional audio — Specifications for audio on high-capacity media

Users of this standard are encouraged to determine if they are using the latest printing incorporating all current amendments and editorial corrections. Information on the latest status, edition, and printing of a standard can be found at:
<http://www.aes.org/standards>

AUDIO ENGINEERING SOCIETY, INC.
551 Fifth Avenue, New York, NY 10176, US.



The AES Standards Committee is the organization responsible for the standards program of the Audio Engineering Society. It publishes technical standards, information documents and technical reports. Working groups and task groups with a fully international membership are engaged in writing standards covering fields that include topics of specific relevance to professional audio. Membership of any AES standards working group is open to all individuals who are materially and directly affected by the documents that may be issued under the scope of that working group.

Complete information, including working group scopes and project status is available at <http://www.aes.org/standards>. Enquiries may be addressed to standards@aes.org

The AES Standards Committee is supported in part by those listed below who, as Standards Sustainers, make significant financial contribution to its operation.



THE TELOS ALLIANCE



This list is current as of 2017/7/25

AES project report for professional audio — Specifications for audio on high-capacity media

Published by

Audio Engineering Society, Inc.

Copyright ©1998 by the Audio Engineering Society

Abstract

This document is the report of Task group SC-02-M, a task force within the Audio Engineering Society Standards Committee (AESSC), which studied the future of high-capacity audio media for over one year with input from more than 80 persons and organizations.

An AES report implies a consensus of those directly and materially affected by its provisions who have approved it as representing the views of their AESSC subcommittee but not of the AES as a whole. It is intended as a guide to aid the manufacturer, the consumer, and the general public. The existence of an AES report does not in any respect preclude anyone, whether or not he or she has approved the document, from manufacturing, marketing, purchasing, or using products, processes, or procedures not in agreement with the report. Publication does not assume any liability to any patent owner, nor does it assume any obligation whatever to parties using the report. This document is subject to periodic review and users are cautioned to obtain the latest edition.



Contents

Foreword.....3

Introduction.....4

Section 1 5

Principles.....5

Four Main Elements of Consideration 7

Tradeoffs in practical systems..... 16

Recommendations..... 16

Section 2 17

Production Technology 17

Release Media 19

DVD - Audio Example 22

Section 3 26

Summary 26

Bibliography 28



Foreword

[This foreword is not a part of *AES project report for professional audio — Specifications for audio on high-capacity media*, AES-R1-1997.]

Task group SC-02-M, a task force within the Audio Engineering Society Standards Committee (AESSC), studied the future of high-capacity audio media for over one year with input from more than 80 persons and organizations. The fundamental ingredients of the future of audio are the potential increase to the number of audio channels and improvements to audio coding methods including sample rate and word length. Any practical system will involve tradeoffs among these factors, and the task force discussed each of the ingredients in turn at length, to form the basis of recommendations for future work. In order to forward the goals of improved audio, the task force recommends that scientific listening tests be conducted of proposed new systems, so that in conjunction with the other technical factors that form any new system, a rational basis for the development of the industry can occur.

The Audio Engineering Society is unique in being international in scope, and in crossing boundaries between professional and consumer applications of audio. This places it in a position of responsibility to manage the hierarchical development of production, distribution, and reproduction oriented parts of the industry.

This document was prepared by the SC-02-M task group of the AESSC SC-02 Subcommittee on digital Audio. The task group was chaired by J. Eargle and T. Holman

This document is a report and not a standard of information document. It does not follow IEC and AES editorial style and has not proceeded beyond subcommittee approval in the AESSC consensus process.

J. Nunn, Chairman
R. Finger, Vice-Chairman
1997-11-04



AES project report for professional audio — Specifications for audio on high-capacity media

Introduction

At the Fall 1995 meeting of the AESSC SC-02 Subcommittee on Digital Audio, held in New York, a task group, SC-02-M, Task Force on High-Capacity Digital Audio was formed. John Eargle and Tomlinson Holman were appointed co-chairs. The goals of the task force as described in AESSC project AES-X38, were to:

1. describe options for advanced audio systems, including utilization of high-capacity media and corresponding interconnection systems;
2. involve in its deliberations all interested members of the community, including artists, recording engineers, producers, record companies, development engineers, and manufacturers;
3. move towards goals 1 and 2 with all deliberate speed in order to have a significant influence on current developments.

This is a report of the task force including recommendations for further work. The first section of this report is a discussion of the various topics encountered and a summary of discussions had about the topics. The second section is the task force's document number D25, "Professional Technology and Release Media utilization of sample rate, word length, audio coding, and number of audio channels, with DVD-Audio as an example." The third section is the Task Force report to the SC-02 Subcommittee on Digital Audio, which details the operations of the task force.



Section 1

The topic of advanced audio systems is a broad and fascinating one, drawing on many disciplines across the board from traditional electronic engineering to psychoacoustics, for at the heart of development of the audio industry is the question "what advances are audible and to what degree?" To the extent that these questions could be answered, then a rational basis for development of an industry could be set forth. While these questions are clearly controversial, subject to ongoing debate, and perhaps will never be fully resolved, the task force took in information from experiments in the literature and carried out discussions among practitioners that add valuable insight to the debate.

New media certainly include the potential for increases in overall storage and transmission capacity, increases in the amount of information delivered for one program per unit of time, and increases in flexibility of media such as access to different programs or parts of programs. How these increases, and the potential tradeoffs among them, are to be used in practical media is the principal subject of debate.

Of course for any introduction of a new medium, many practical obstacles peculiar to that particular medium must be faced, but one goal of the task force was not to be too constrained to the requirements of just one new medium, but rather to step back and look at what factors may help unify multiple media and rationalize the practices in professional and consumer audio. On the other hand, if the task force worked without any practical constraints, then its efforts could easily be set aside as impractical. For these reasons, the discussion document that follows as the second section of this report is set out in three parts: Production Technology, Release Media, and DVD-Audio. In Production Technology the concentration is on professional practice, and especially how it may differ from consumer practice. Release Media concentrates on the desirable properties of a finished delivery medium, without specific requirements for bit rate and size capacity. The DVD-Audio section concludes with how most of the requirements set forth in the Release Media section can be accommodated within the constraints of one new medium.

Throughout the course of discussions, certain principles emerged about which there was relatively little controversy. Those are listed in the next sections, named Principles. Not surprisingly perhaps, these relatively non-controversial measures include those that require only small parts of any mass storage medium, or only low bit rates to implement. The "Big Four" items that do cause more controversy are: the method of coding the analog audio into digital; the sample rate and word length, both of which may interact with the method of coding; and the number of loudspeaker channels. Each of these will be discussed below.

Principles

1. Professional audio equipment should be designed considering that the ultimate release medium is likely to contain fewer audio channels than are used in the professional domain (i.e., multi-track production technique is commonplace). This requires for a rational system design that the **professional equipment be built to higher standards than the ultimate release medium** in order to preserve all of the quality potential of the release medium. (More than one professional channel will be combined into one release medium channel, and combining channels adds noise from the multiple channels in use together. Also headroom is affected as multiple channels added together may exceed the headroom of any one source channel.) For example, if the release medium is the 16-bit linear pulse-code modulation (PCM) conventional compact disc, and multi-track production techniques are in use, then a multi-track recorder to be used in delivering a 16-bit result in the release medium needs to have more than 16-bits of resolution, since combining channels of the source multi-track reduces the dynamic range of the result.
2. Professional audio equipment should operate in a domain which is either **linear PCM** based, or related to linear PCM in a straightforward manner. This will make the processing inside professional audio equipment simpler, and will minimize potential problems with cascading potentially different coding methods across multiple generations. More efficient and even potentially better psychoacoustic coding may exist than linear PCM and be put to use in release media, but that does not change the requirement in the professional domain for linear PCM or equivalent coding, due to its relative computational simplicity and attendant lower processing cost compared to other coding methods.
3. The program producer shall have **flexible use of the medium** within parameters which make hardware devices and corresponding software interchangeable within reason.

Most of the discussion about flexible use centered around the tradeoff amongst the number of audio channels, the method of coding, and the concomitant and associated items due to the method of coding, such as sample rate and word length if linear PCM coding is used.



A small part of the discussion described the potential for video accompanying audio, such as still pictures. Full-motion video already has media in the marketplace accompanied by multi-channel audio, so the concentration of the task force was on audio features rather than accompanying video features.

4. Where practical, release media shall have the **capability for both two-channel and multi-channel delivery modes simultaneously** to satisfy needs of the huge number of installed two-channel reproduction systems, thus providing backwards compatibility for these users while simultaneously supplying multi-channel or other higher capacity sound capability. This simultaneous two-channel and higher capacity capability is at the option of the producer, and may not be applicable to all high-capacity media.

Although "mixdown" to two channels from a greater number is possible in field equipment, even with "mixdown equations" specified by the producer from program to program, the mixing technique for two-channel recordings is sufficiently different from multi-channel recording that many program producers strongly supported this principle.

The basic problem with mixdown from a multi-channel release mix as a method to produce a two-channel output is that the producer has many more tools in the studio to make a mixdown than would ever be available in an end user mixdown, and since the bulk of the market is still two-channel, separate capability for two-channel and multi-channel or higher capacity audio on one medium was judged important.

5. Release media should carry information about the production in a useful form such as in a header.

For example, interchanging compact discs for program loudness is problematic. The difficulty is that, if properly made, the peaks of the program material on the disc just reach the peak-coded level, full scale. The loudness of programs depends though not on the peak level, but on a complex of characteristics involving the moment-by-moment frequency spectrum and its dynamic changes. "There are four basic psychological-physical relations that depict the perceptual attribute of sound called loudness:

1. The intensity levels required to make each tone or critical band of frequencies in the audible frequency range appear subjectively to be equally loud. [Explained by the equal-loudness contours.]
2. The duration required to achieve a stable perception of the loudness of a sound...
3. The growth of loudness as the bandwidth of the sound spectrum widens
4. The growth of loudness as the physical intensity of a given sound increases."

Karl D. Kryter, *The Effects of Noise on Man*, Academic Press, Inc., 1985., p. 113.

The result is that changing from one disc to another, with different ratios of loudness to peak level, means that the loudness varies from disc to disc, undesirable from the end user's point of view, although competent from the standpoint of each disc's utilization of the digital medium.

In the advanced television system (ATS) standard for the U.S., this problem has been solved by adding two header features present in each audio frame. The first is called dialog normalization, and the second, mix level. The first of these is not directly applicable to a wider range of program material encountered within the music industry because it relies on dialog to provide a standardized level. Nevertheless, the ideas presented in these two header features are useful in an audio-only medium.

First, the reference level in use for a particular program relative to digital full scale needs to be indicated to the playback equipment. This overcomes the problem that various program makers and different pieces of equipment use different reference levels ranging across at least -20 dB FS to -12 dB FS. By transmitting the reference level in use to the end-user equipment, valuable information is preserved. This information provides a means to make more highly compressed material that normally employs a higher reference level to play back at a more reasonable reproduction level (a lower level), and material with wider dynamic range, using a lower reference level, to be reproduced without modification.

Second, an absolute level for loudness of reproduction should be provided. This can be accomplished by implementing mix level. Mix level is a parameter which enables the establishment of the gain structure necessary to reproduce the correct reproduced absolute sound pressure level (SPL). It gives the correspondence between absolute SPL and digital full scale, such as 105 dB for 0 dB FS. While the user is free to play the level anywhere he or she wants with a volume control of course, indicating the producer's intent with regard to playback level helps to realize the producer's intent either 1) by reproducing the original level or 2) if reproducing at a lower level applying an appropriate loudness compensation for the equal loudness contours of hearing. Such loudness controls in the past



have only been fixed approximations. By knowing both the intended original level, and the actual playback sound pressure level, a more exact compensation may be made.

Implementation of these two features, taken together, provides a means to solve the current perceived level problem of digital media: that of inconsistency of loudness from source to source and program to program, and provides a means for more correct reproduction of the program material.

6. There is very great utility to have a **target reproduction system** which actual systems in the field may emulate to a larger or smaller degree. Notating the target reproduction system as **Rosetta Stone data** may be used by the playback system to make adjustments so that the intended reproduction is more likely to take place.

Originally the thought behind this was the same as Richard Heyser's, "A Rosetta Stone for Audio," *JAES*. [...] Heyser's idea was to preserve the acoustics of the original space through a recording of a test signal in the space, in order that some future generation be able to better reproduce the acoustics of the original space over the playback system in a more sophisticated manner than available today.

Upon thinking about this however, the notion emerged that Heyser's idea was not sufficiently general, in the sense that only a small percentage of recordings are made under such conditions (natural acoustic spaces) without manipulation, and that the whole of audio recording engineering needs to be accommodated. Then the attention turned to playback systems. The film industry has benefited with better interchangeability of its program material through the promulgation of standards on the frequency response and loudness of monitoring. Journal papers in *JAES* have shown unequivocally that mixing engineers "mix with their ears" thereby winding up correcting for errors in their monitor systems. With the errors (in particular frequency response errors) removed, more consistent results are produced.

While the larger and more diffuse audio business in general probably cannot come to one set of monitoring standards like the film industry, it should be possible to notate what the producer's intent with the program material is: how loudly it is to be played and over a sound system having what kind of frequency response, including response directed at the listener (listening window response), and power response (response at all angles).

7. Two types of information content in addition to audio are recognized as useful: header information such as tables of contents, Rosetta Stone data and the like; and embedded information such as the score for a musical work, musical instrument digital interface (MIDI) sequence, or the like.

8. One fact is that media being introduced today and tomorrow can make greater use of such features as header-descriptor files and customized formats than could previous digital media since computer control over products has become the norm in the time since the first digital systems appeared on the marketplace. Thus, flexible use of new media is made practical through ongoing developments in the parallel field of computer control.

Four Main Elements of Consideration

I. Audio Coding

Linear PCM has served the audio industry long and well. It is well understood, relatively easy to use in processes such as equalization, and relatively inexpensive in terms of signal processing power needed to manipulate it. On the other hand, linear PCM may not be the most efficient code to use in terms of the highest perceived audio quality for a given bit capacity or bit rate, since it has not been optimized psychoacoustically, in general.

Linear PCM can be made more efficient with the use of special methods of coding such that the total storage capacity, or the bit rate, or both, can be reduced with no loss of information. Lossless coding provides more or less robust mathematical means to reduce the storage or bit rate requirements, while still performing identically to linear PCM in ultimate performance, since the bits intended to be applied to the output digital-to-analog converter are identical by definition, whether they have been subject to such lossless coding or not.

Developments in the field of low-bit-rate coding made for more efficiency in coding employ lossy compression wherein a psychoacoustic model is used to eliminate information content thought to be inaudible, while retaining all audible features of the signal. It must be pointed out that such psychoacoustic models represent, at their best, averages of large populations, and that there may be specialized populations that may hear audible artifacts of such processes if they are used nearly to the point of audibility for the average population. Also, they are based on masking at specific sound pressure levels, and reproduction at an unintended level could produce audible artifacts.

