

# **Digital Formats for Library of Congress Collections**

## **Factors and Formats for Sound**

**DRAFT**

**Caroline R. Arms**

**Carl Fleischhauer**

**Office of Strategic Initiatives**

**Table of Contents**

Quality and Functionality Factors for Sound..... 3  
    Normal rendering for sound..... 3  
    Fidelity (support for high audio resolution)..... 3  
    Sound field (support for multichannel audio)..... 4  
    Functionality beyond normal sound rendering..... 4  
Sound Format Preferences in Summary ..... 6  
    Device-independent digital audio ..... 6  
    Note-based digital musical composition..... 7  
Sound Content Subcategories and Specific Format Preferences ..... 8  
    Table: Sound content categories and their significant characteristics ..... 8  
    Table: Sound content categories and format preferences ..... 10

## *Scope*

The discussion in this document concerns individual sound items. The discussion does not address compound documents that include images and other forms of expression.

## *Quality and Functionality Factors for Sound*

### **Normal rendering for sound**

Normal rendering for sound is playback in mono or stereo through one or two speakers (or equivalent headphones) using software providing user control over volume, tone, balance, etc., and the means to fast forward and to find a specific track or other segment, e.g., a chapter in a recorded book, or to go to a specific elapsed time with the file. Normal rendering would also allow playback through software that allows sound analysis and excerpting. Normal rendering must not be limited to specific hardware models or devices and must be feasible for current users and future users and scholars. This level of functionality is expected of any candidate digital format for preserving sound content, and is not mentioned as a factor for choosing among formats.

### **Fidelity (support for high audio resolution)**

*Fidelity* refers to the degree to which “high fidelity” content may be reproduced within this format. In this context, the term is meant broadly, referring to the factors that will influence a careful (even expert) listening experience. Strictly speaking, this factor is limited to formats that reproduce sound waveforms, where a real test of fidelity occurs when the reproduction is repurposed, e.g., a “master file” is used as the basis for the master for a new audio-CD music release. When discussing formats for note-based representations of music, e.g., MIDI (Musical Instrument Digital Interface), this analysis extends the use of *fidelity* to represent the accuracy or completeness of the notation and tonality data associated with a given content item.

For files that represent sound waveforms using linear PCM (Pulse Code Modulated) data, the two characteristics most often associated with fidelity are sampling frequency and word length (i.e., bit depth).<sup>1</sup> Other factors may also influence fidelity, such as the presence of distortion, watermarking, or--in compressed renderings derived from PCM files--audible artifacts that result from the application of compression. In general, uncompressed data offers the highest fidelity; however, lossy compression based on understanding of human perception of sound has proved to provide a high level of fidelity in normal playback conditions. New techniques for compression are being developed, tested, and standardized and specifications for what constitutes “high-quality” compression can be expected to change over time. As of 2003, the Library of Congress

---

<sup>1</sup> An alternate structure to PCM has been developed by Sony and is called by them DSD (Direct Stream Digital). It is a one-bit deep coding with a bit rate of 2.8224 megabits per second. Audio engineers also refer to this structure as pulse width modulation (PWM) or delta-sigma modulation. At this writing, DSD is exclusively heard on Sony SACDs (Super Audio Compact Disks), and no media-independent format is in use. Comments in the engineering press vary; some state a clear preference for high-resolution PCM sampling, while others politely say that both approaches will be acceptable to most audiophiles.

Motion Picture Broadcast and Recorded Sound Division, considers an MP3 file with compression to support a 128 Kbits/second playback data rate per channel (256 Kbits/second for stereo) as at the lower end of quality acceptable for published music. Lower-quality compression would be acceptable for culturally significant, but home-produced sound.

The effects of compression may be reflected in other factors relating to choice of digital formats. In general, compression detracts from transparency, desirable in any format chosen for long-term preservation. Also, constraints on data transfer rates may require balancing support for multichannel audio (surround sound) with fidelity limitations in individual channels.

The treatment of spoken word content or artificially generated speech in the context of telephony (and related activities) has a special character, using coding techniques such as  $\mu$ -law, LPC (linear predictive coding), and GSM (Global System for Mobile telecommunications) compression. Content from these modes of communication is unlikely to find its way into Library of Congress collections. Speech-compression formats are in use, however, for some digital recorded books and other recordings. For example, the downloadable book and radio broadcasts available from Audible, Inc. (<http://www.audible.com>) are offered in five formats, including formats based in speech compression and others based in the formats typically used for music, i.e., MP3. The Library's preference is likely to be for the higher quality version—MP3-based in the case of Audible, Inc.—rather than a lower-quality, speech-compressed version.

### **Sound field (support for multichannel audio)**

*Sound field* refers to the degree to which stereo, surround sound, and other multi-channel audio may be represented within the format, especially in channels beyond normal stereo rendering.

The sound field factor pertains to audio file formats in which multi-channel audio is represented in multiplexed or interleaved structures. (This is distinct from multi-file structures, e.g., SMIL, in which multiple files are synchronized for playback.) At this writing, in addition to stereo, digital audio content is being offered in structures referred to as 5.1 (for loudspeakers at front left, front right, back left, back right, and low frequency/woofer), 7.1 (same with two added side-center loudspeakers), and so on. In many cases, the practical constraints of digital data transfer rates during real-time playback force the creators of these formats to limit the fidelity of some channels, typically channels other than front left and front right.

### **Functionality beyond normal sound rendering**

*To be added: discussion of notation-based representations, such as MIDI.*

Albums of recorded music, recorded books, and other multipart works will benefit from the inclusion of metadata, track indication, markup, or other elements that provide:

- Playlists or tables of contents

- Identification ,description, and location for chapters, sections, illustrations, individual musical selections, and the like
- Descriptive information for display by players or play software, e.g., title, names of authors, performers, narrators; names of chapters or sections; additional information of the types familiar to library users

## ***Sound Format Preferences in Summary***

### **Media-independent digital sound recordings**

#### **General**

Fidelity characteristics (bitstream encoding) should be used as the primary consideration; choice of file formats as secondary. This section deals only with media-independent digital audio. For audio in physical formats intended for consumer audio equipment, see section on Phonorecords.

#### **Bitstream encoding for sound recordings (relates to fidelity)**

- Higher sampling rate (usually expressed as kHz, e.g., 96kHz) preferred over lower sampling rate.
- 24-bit sample word-length preferred over shorter
- Linear PCM (uncompressed) preferred over compressed (lossy or lossless)
- Higher data rate (e.g. 128 K bits per second) preferred over lower data rate for same compression scheme and sampling rate.
- AAC compression preferred over MPEG-layer 2 (MP3)
- Surround sound (5.1 or 7.1) encoding only necessary if essential to creator's intent. In other cases, uncompressed encoding in stereo is preferred.

#### **File type**

- Not copy-protected rather than copy-protected
- Relatively complete descriptive and technical metadata rather than minimal
- Acceptable file formats for mono/stereo, other than speech
  - Broadcast WAVE with Linear PCM encoding
  - WAVE with Linear PCM encoding
  - AIFF with Linear PCM encoding
  - MPEG-4 with AAC encoding
  - MPEG-2 with AAC encoding
  - MP3
  - WMA with Microsoft CODEC encoding
- Acceptable file formats for surround sound
  - MPEG-4 with AAC encoding
  - WMA with Microsoft CODEC encoding

## **Note-based digital musical composition**

### **General**

- With associated sounds (either through General MIDI specifications or downloadable sounds) rather than without.

### **Bitstream encoding**

- MIDI preferred over all other encodings. Digital compositions in other note-based formats should be submitted as audio files.
- General MIDI Level 1 preferred over General MIDI Level 2 for sequences
- DLS Level 2 (or 2.1) preferred over Level 1 for customized sounds

### **File type**

Order of preference

- XMF (eXtensible Music Format)
- SMF (Standard Midi File)

### **Recorded books**

The preference for this subcategory are generally identical to those for device-independent digital sound recordings. The exception is that preferred recorded book formats (when in a capable player, hardware or software) will support a variety of end-user functions, including:

- Bookmarking set by the user
- Automatic holding the last position (where play left off)
- Display of time elapsed, time remaining, and the ability to go to a specified time
- Support for navigation, e.g., to chapters, sections, or illustrations
- Display of descriptive information, e.g., title, name of author, name of narrator, name of chapter titles
- Re-read capability (repeat last sentence or paragraph)

*Sound Content Subcategories and Specific Format Preferences*

**Table: Sound content subcategories and their significant characteristics**

Description	Fidelity (resolution) [1]	Sound field (beyond stereo) [2]	Rendering expectations beyond normal [4]	Special functionality required by custodians [5]	Special functionality expected by end users	Effect of technical protection [6]
<b>S1</b> Audio, surround sound	Very important	Retain with minimal change [3]	Surround sound, multiple speakers	Downsample, take excerpts, etc., without artifacting.		Must not affect fidelity or normal rendering
<b>S2</b> Audio, mono or stereo Includes most audio content, such as: - Audio content registered for copyright or requested through mandatory deposit; - Collections of digital audio acquired from collectors, producing organizations, etc. - Home-produced sound as might be received from individuals contributing to oral history collections; - Music created or received in note-based form (e.g. MIDI) and preserved as audio	Very important or important, depending on item			Downsample, take excerpts, etc., without artifacting		Must not affect fidelity or normal rendering
<b>S3</b> Audio, streamed webcast, harvested in bulk	Less important					Must not affect normal rendering
<b>S4</b> Audio incidental to Web harvesting (e.g., background audio) [7]	Not important					Not important
<b>S5</b> Note-based representation for generating music (e.g., MIDI) when special functionality is required [8]	Retain precision of original	N/A	Through specialized performance software	Retain functionality of original via performance & composition software [8]	Retain functionality of original via performance & composition software	Must not affect functionality for end users
<b>S6</b> Synthetic encoding for non-music (e.g., sound effects, voice for phone directory						

	assistance) [For future consideration.]					
S7	Recorded books [9]					Various [10]

**Notes:**

1. For example, digital audio in the widely adopted linear PCM format associates fidelity (resolution) with sampling frequency and bit-depth.
2. Generally speaking, this characteristic is associated with surround sound, although it may also pertain to multi-channel audio (e.g., narration available in English and French). There is related interest in metadata that offers a map of the channels.
3. Reduction to two-loudspeaker rendering should be feasible with appropriate software. Normalization to stereo may be appropriate for content that has a more complex sound field originally but where this particular characteristic is not deemed necessary for retention.
4. Normal rendering means playback in mono or stereo through one or two speakers (or equivalent headphones) using software providing user control over volume, balance, fast forward, go-to-track, etc. Normal rendering would also allow playback through software that allows sound analysis and excerpting. Normal rendering must not be limited to specific hardware models or devices and must be feasible for current users and future users and scholars.
5. Normal functionality for custodians includes the ability to preserve digital content and provide service to users and designated communities now and decades. Thus custodians must be able to replicate the content on new media, migrate and normalize it in the face of changing technology, and disseminate it to users at a resolution consistent with network bandwidth constraints.
6. Technical protection must not prevent custodians from taking appropriate steps to preserve the digital content and make it accessible to future generations. See Notes 4 and 5.
7. In contrast, audio files harvested from the Web through a program targeted specifically at sound capture would be considered as S1 or S2.
8. For music composed using digital composition systems, guidelines will be necessary (by custodians and user communities) as to when the functionality inherent in a note-based representation is an essential characteristic and when the composition should be preserved as audio. The files of composers at the leading edge of digital composition will often be in non-standard note-based representations. These will require special consideration.
9. Recorded books in this table refers to commercially published recordings and not the specialized digital talking books for the blind and vision-impaired as specified in ANSI/NISO standard Z39.86.
10. Desired functions include bookmarking; holding last position (where play left off); display of time elapsed, time remaining, and the ability to go to a specified time; support for navigation, e.g., to chapters, sections, or illustrations; display of descriptive information, e.g., title, name of author, name of narrator, name of chapter titles; and re-read capability (repeat last sentence or paragraph).

**Table: Sound content subcategories and format preferences**

	Description	Preferred formats [1]		Acceptable formats	
		Encoding type	File type, subtype	Encoding type	File type, subtype
<b>S1</b>	Audio, surround sound [2]	5.1 or 7.1 surround, high-quality lossy [3]	WMA-Surround MPEG4-AAC-Surround		
<b>S2</b>	Audio, mono or stereo  Includes most audio content, such as: Audio content registered for copyright or requested through mandatory deposit; Collections of digital audio acquired from collectors, producing organizations, etc. Home-produced sound as might be received from individuals contributing to oral history collections; Music created or received in note-based form (e.g. MIDI) and preserved as audio	Linear PCM (i.e. uncompressed)	WAVE-PCM-BWF WAVE-PCM AIFF-PCM	High-quality lossy [3]  or  Low-quality lossy [3]	MP3 MPEG2-AAC MPEG4-AAC WMA-Compressed  MP3 MPEG2-AAC MPEG4-AAC WMA-Compressed
<b>S3</b>	Audio, streamed webcast, harvested in bulk	Low-quality lossy [3]	MP3 MPEG2-AAC MPEG4-AAC WMA-Compressed		
<b>S4</b>	Audio incidental to Web harvesting (e.g., background audio)	As available	Any		
<b>S5</b>	Note-based representation for generating music (e.g., MIDI) when special functionality is required	MIDI	MIDI		
<b>S6</b>	Synthetic encoding for non-music (e.g., sound effects, voice for phone directory assistance) <i>[For future consideration.]</i>		WMA		
<b>S7</b>	Recorded talking books				

**Notes:**

1. Other device-independent digital formats for sound exist and may be added as preferred or acceptable in the future. For example, another proposed approach to encoding audio bitstreams, sometimes referred to as Direct Stream Digital and claiming to result in higher fidelity, uses pulse density modulation (PDM) instead of pulse code modulation (PCM). This table excludes formats limited to particular tangible media as inappropriate as part of a general strategy for long-term preservation of digital content. Hence, the regular Audio CD is not listed; nor are surround-sound formats intended for home-theater use that are limited to certain tangible media (e.g., DVD-Audio & Super Audio CD).
2. Surround sound is important to retain when it is an important element of the artist's intent (e.g. sounds intended to move around in the performance space). However, for most conventional audio, reduction to stereo is appropriate.
3. The degree of compression and specific encoding algorithms applied to audio produce files at varying levels of quality. For example, in a current prototyping project, the Motion Picture, Broadcasting, and Recorded Sound Division sees an MP3 file (derived from a PCM bitstream at 44.1 kHz sampling) compressed for a 128Kbps per channel data rate as a reasonable "high quality" service version. The fidelity offered by such a file is roughly comparable to an audio CD played on normal consumer equipment in routine circumstances.