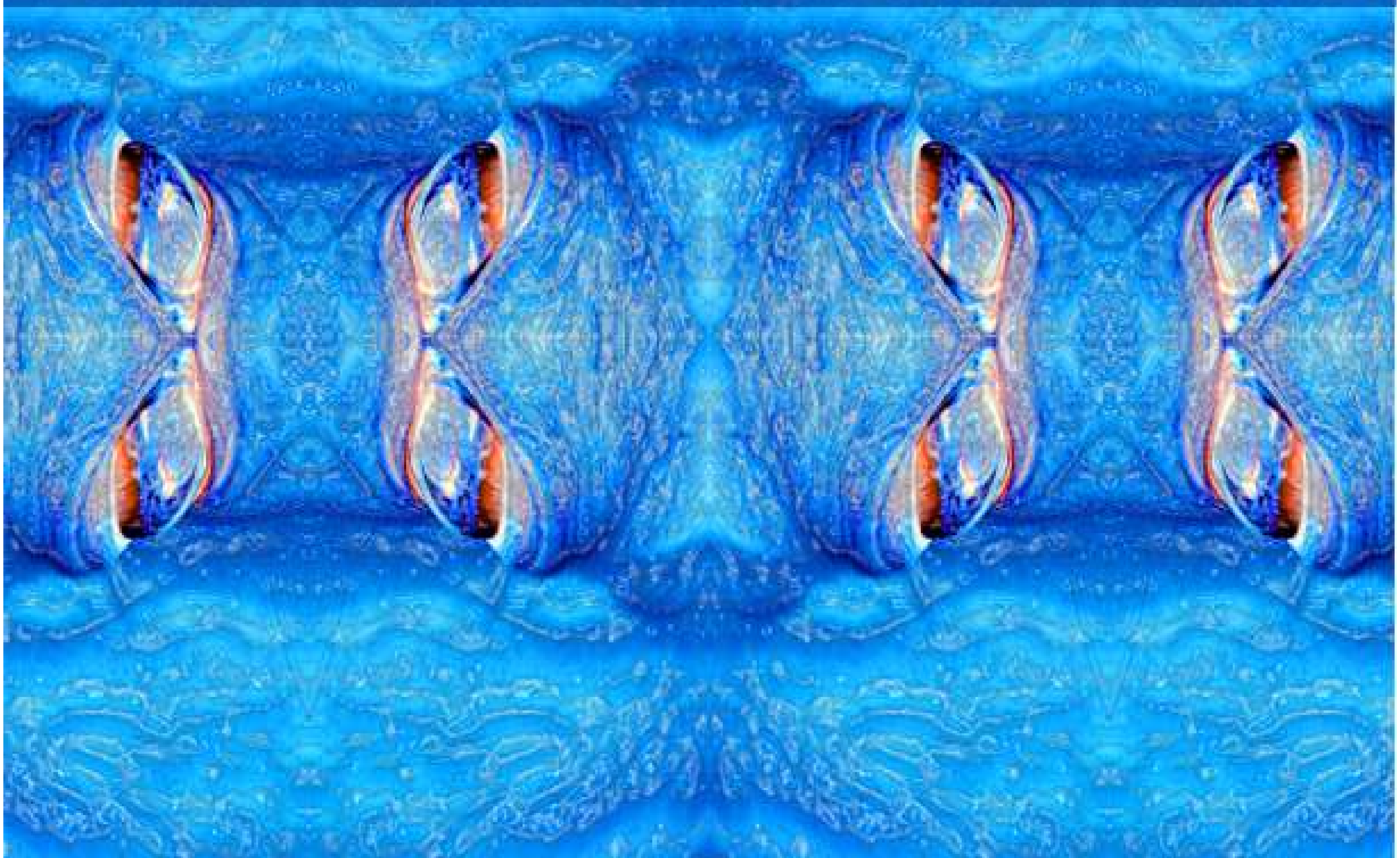
Digital Audio Technologies



Nigel Mccarty

First Edition, 2012

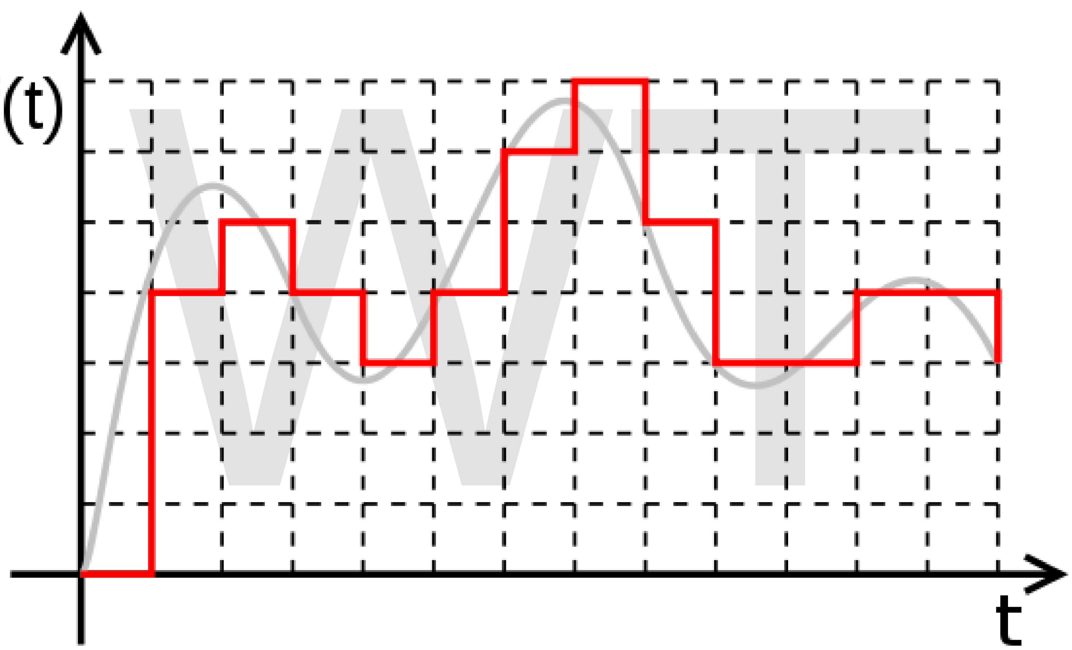
ISBN 978-81-323-3039-4

Published by:
Research World
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

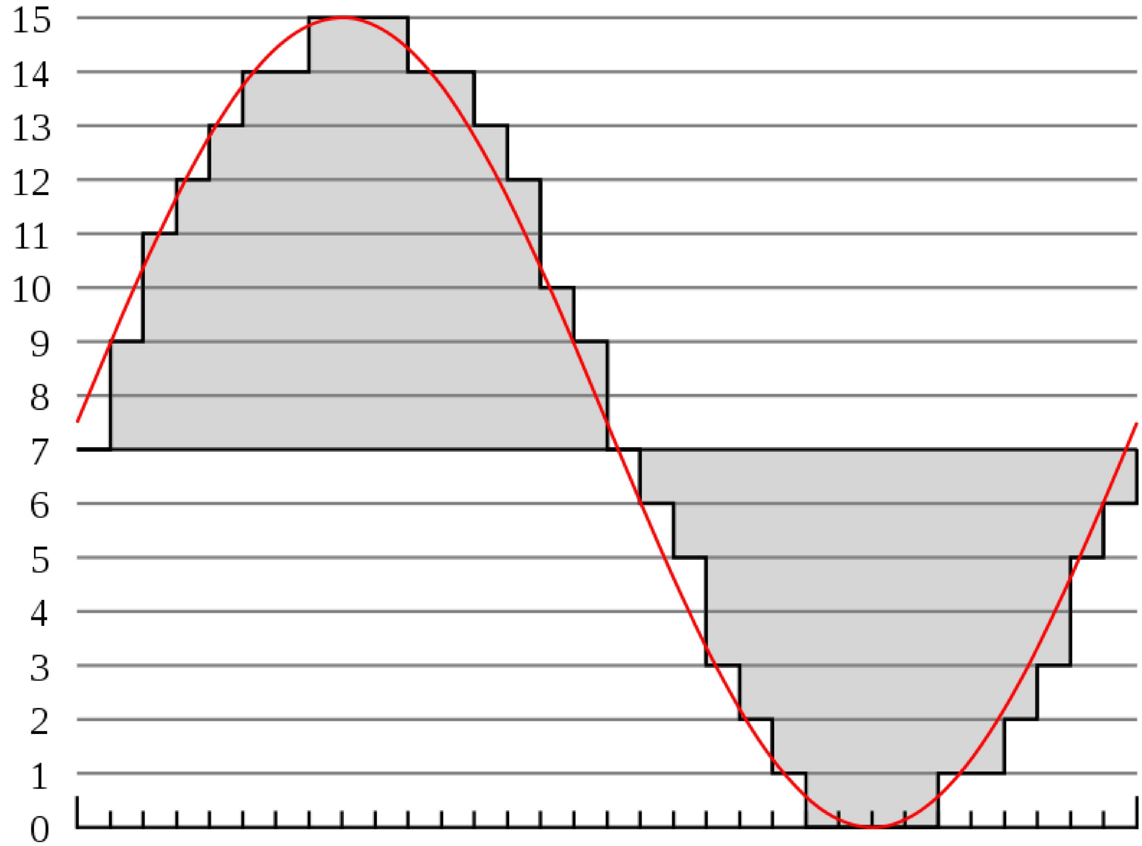
Digital Audio



A sound wave, in gray, represented digitally, in red (after quantization and zero-order hold).

Digital audio is the result of sound reproduction, using pulse-code modulation and digital signals. This includes analog-to-digital conversion (ADC), digital-to-analog conversion (DAC), storage, and transmission. In effect, the system commonly referred to as digital is in fact a discrete-time, discrete-level analog of a previous electrical analog. While modern systems can be quite subtle in their methods, the primary usefulness of a digital system is the ability to store, retrieve and transmit signals without any loss of quality.

Overview of digital audio



Sampling and 4-bit quantization of an analog signal (red) using Pulse-code modulation

Digital audio has emerged because of its usefulness in the recording, manipulation, mass-production, and distribution of sound. Modern distribution of music across the Internet via on-line stores depends on digital recording and digital compression algorithms. Distribution of audio as data files rather than as physical objects has significantly reduced the cost of distribution.

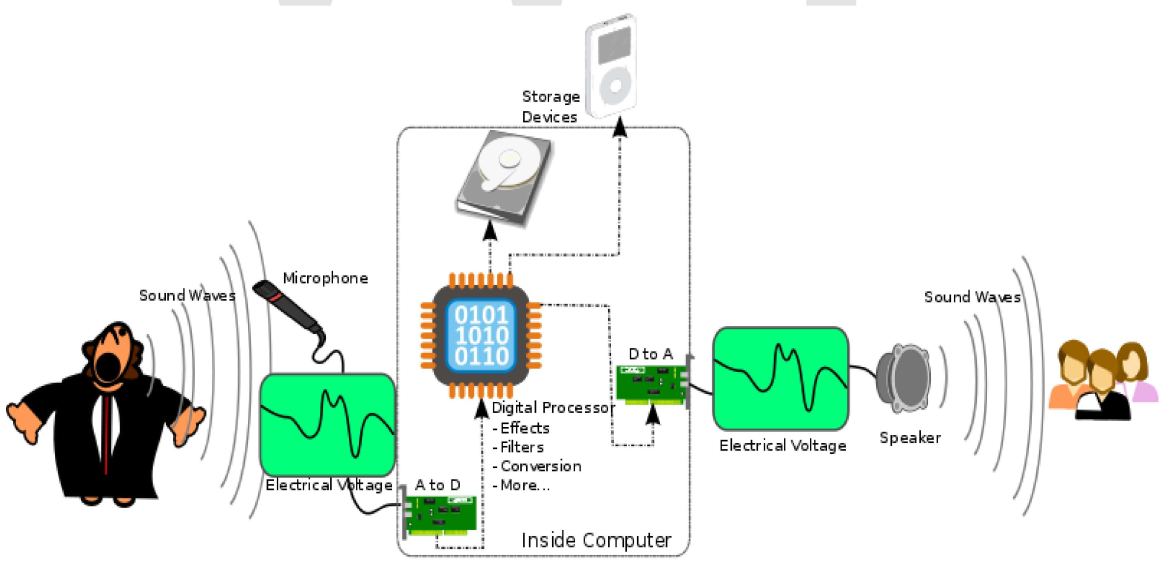
In an analog audio system, sounds begin as physical waveforms in the air, are transformed into an electrical representation of the waveform, via a transducer (for example, a microphone), and are stored or transmitted. To be re-created into sound, the process is reversed, through amplification and then conversion back into physical waveforms via a loudspeaker. Although its nature may change, analog audio's fundamental wave-like characteristics remain the same during its storage, transformation, duplication, and amplification.

Analog audio signals are susceptible to noise and distortion, unavoidable due to the innate characteristics of electronic circuits and associated devices. In the case of purely analog recording and reproduction, numerous opportunities for the introduction of noise and distortion exist throughout the entire process. When audio is digitized, distortion and noise are introduced only by the stages that precede conversion to digital format, and by the stages that follow conversion back to analog.

The digital audio chain begins when an analog audio signal is first sampled, and then (for pulse-code modulation, the usual form of digital audio) it is converted into binary signals—‘on/off’ pulses—which are stored as binary electronic, magnetic, or optical signals, rather than as continuous time, continuous level electronic or electromechanical signals. This signal may then be further encoded to allow correction of any errors that might occur in the storage or transmission of the signal, however this encoding is for error correction, and is not strictly part of the digital audio process. This "channel coding" is essential to the ability of broadcast or recorded digital system to avoid loss of bit accuracy. The discrete time and level of the binary signal allow a decoder to recreate the analog signal upon replay. An example of a channel code is Eight to Fourteen Bit Modulation as used in the audio Compact Disc (CD).

Conversion process

A digital audio system starts with an ADC that converts an analog signal to a digital signal. The ADC runs at a sampling rate and converts at a known bit resolution. For example, CD audio has a sampling rate of 44.1 kHz (44,100 samples per second) and 16-bit resolution for each channel. For stereo there are two channels: 'left' and 'right'. If the analog signal is not already bandlimited then an anti-aliasing filter is necessary before conversion, to prevent aliasing in the digital signal. (Aliasing occurs when frequencies above the Nyquist frequency have not been band limited, and instead appear as audible artifacts in the lower frequencies).



The lifecycle of sound from its source, through an ADC, digital processing, a DAC, and finally as sound again.

The digital audio signal may be stored or transmitted. Digital audio storage can be on a CD, a digital audio player, a hard drive, USB flash drive, CompactFlash, or any other digital data storage device. The digital signal may then be altered in a process which is called digital signal processing where it may be filtered or have effects applied. Audio data compression techniques — such as MP3, Advanced Audio Coding, Ogg Vorbis, or FLAC — are commonly employed to reduce the file size. Digital audio can be streamed to other devices.

The last step is for digital audio to be converted back to an analog signal with a DAC. Like ADCs, DACs run at a specific sampling rate and bit resolution but through the processes of oversampling, upsampling, and downsampling, this sampling rate may not be the same as the initial sampling rate.

History of digital audio use in commercial recording

Pulse-code modulation was invented by British scientist Alec Reeves in 1937 and was used in telecommunications applications long before its first use in commercial broadcast and recording. Commercial digital recording was pioneered in Japan by NHK, and Nippon Columbia (a.k.a. Denon) in the 1960's. The first commercial digital recordings were released in 1971.

The BBC also began experimenting with digital audio in the 1960s. By the early 1970s they had developed a 2-channel recorder and in 1972 they deployed a digital audio transmission system linking their broadcast center to their remote transmitters.

The first 16-bit PCM recording in the United States was made by Thomas Stockham at the Santa Fe Opera in 1976 on a Soundstream recorder. In 1978, an improved version of the Soundstream system was used by Telarc to produce several classical recordings. At the same time 3M was well along in development of their digital multitrack recorder based on BBC technology. The first all-digital album recorded on this machine was Ry Cooder's "Bop 'Til You Drop" which was released in 1979. In a crash program started in 1978, British record label Decca developed their own 2-track digital audio recorders. Decca released the first European digital recording in 1979.

Helped along by introduction of popular digital multitrack recorders from Sony and Mitsubishi in the early 1980's, digital recording was soon embraced by the major record companies. With the introduction of the CD by Sony and Philips in 1982, digital audio was embraced by consumers as well.

Digital audio technologies

Digital audio broadcasting

- Digital Audio Broadcasting (DAB)
- HD Radio
- Digital Radio Mondiale (DRM)

- In-band on-channel (IBOC)

Storage technologies:

- Digital audio player
- Digital Audio Tape (DAT)
- Compact Disc (CD)
- Hard disk recorder
- DVD Audio
- MiniDisc
- Super Audio CD
- Various audio file formats

Digital audio interfaces

Audio-specific interfaces include:

- AC'97 (Audio Codec 1997) interface between Integrated circuits on PC motherboards
- Intel High Definition Audio A modern replacement for AC'97
- ADAT interface
- AES3 interface with XLR connectors
- AES47, Professional AES3-style digital audio over Asynchronous Transfer Mode networks
- I²S (Inter-IC sound) interface between Integrated circuits in consumer electronics
- MADI Multichannel Audio Digital Interface
- MIDI low-bandwidth interconnect for carrying instrument data; cannot carry sound but can carry digital sample data in non-realtime
- S/PDIF, either over coaxial cable or TOSLINK
- TDIF, TASCAM proprietary format with D-sub cable
- A2DP via Bluetooth

Naturally, any digital bus (e.g., USB, FireWire, and PCI) can carry digital audio. Also, several interfaces are engineered to carry digital video and audio together, including HDMI and DisplayPort.

In professional architectural or installation applications, many digital audio networking protocols and interfaces exist.

Chapter 2

Digital Mixing Console

In professional audio, a **Digital Mixing Console (DMC)**, is an electronic device for combining, routing, and changing the dynamics of digital audio samples. The digital audio samples are summed to produce a combined output. A professional digital mixing console is a dedicated desk or control surface produced exclusively for the task. Computers with specialized outboard gear may serve the function, with less control capability—fewer independent fader moves can be initiated at the same time.

Yamaha M7CL in place for a live production

Uses

Digital mixing consoles are typically used in recording studios, public address systems, sound reinforcement systems, broadcasting, television, and film post-production.

Solutions to common sound system problems

Assuming that an institution has an adequate sound system, a common problem is often the person operating the system! An institution can spend thousands on a state of the art sound system but it is only as good as the person operating it. If the institution is fortunate enough to have a professional operator, he or she cannot always be there for every event.

A second major problem is the improper location of the equipment. The sound operator is often secluded in a closed in room, behind glass, etc. No sound operator can properly adjust a live sound system unless he or she can hear exactly like the majority of the audience.

The best solution to offset operator problems is to automate whenever possible. The advent of modern digital computer technology has now made it possible to install sound system components that will almost operate themselves. As one upgrades or installs a completely new sound system, they should try to obtain items that require as little hands-on, human operation, as possible.

Some digital mixers have analog style control features. Those older style manual slider and knob controls make the mixer more user friendly. Those controls however, are commanding a totally digital platform, much like desktop icons or shortcut icons control programs on a personal computer.

A digital mixer can offset the lack of sound operator expertise because it remembers what a person that knew what they were doing, told it to do. For example: A knowledgeable person can adjust all of the microphone settings, monitors, etc., for a given event. After everything is properly adjusted, that setup is assigned a name and stored in the memory. Afterward, a less knowledgeable operator can simply look on the computer screen or touch screen to recall that setting. Then, like some invisible man was operating the system, all of the controls move to their proper positions. Likewise, setups for any other events can be stored and instantly recalled.

One can easily program many different pre-set configurations or "snapshots," into the mixing console. Once a stored setting is recalled, the operator can still make manual volume adjustments, etc., without affecting the stored program. In other words, they can change a lot of stuff and all one has to do is hit the recall and the mixer automatically returns to all of the correct startup settings.

Digidesign's Venue Profile mixer on location at a corporate event. This mixer allows plugins from third-party vendors. A Smart software screen is partially shown on the right—Smart allows plugins, too.

Another advantage of this digital or computer type of mixing console is the abundance of control features that it provides for each microphone channel. It would take racks full of expensive gates, compressor limiters, equalizers, feedback controllers and other signal processing gear, to do what a modern digital mixer does internally. The elimination of all these outboard rack pieces, with all of their switches, knobs and connections, will make a system even less vulnerable to failure and outside interference.

Third-party plugins can add functionality in a digital mixer. Plugins allow for further expansion of the mixer's onboard equalization, compression and reverberation effects.

When mixing for recording purposes one can rely on headphones alone. The opposite is the case for live sound mixing. No sound operator can properly adjust a live sound system unless he can hear exactly like the majority of the audience. For issues related to space, appearance and security, one cannot always locate their sound control equipment, out in the middle of their auditorium. This is even more of a problem if the auditorium is a multi-use building that is often converted for other events.

A digital mixer can also solve this problem. A sound operator can operate the whole sound system from a laptop or desktop computer. With the proper setup, it can even be done by a wireless laptop configuration for increased mobility. In fact, many of the digital mixer's functions are easier to operate from the computer screen than the actual mixing console.

Using dual DMCs to improve live recording

In truly professional broadcast and recording applications, one does not use what is referred to as the house mix for high quality audio recordings. There is a simple reason for this. When one is engineering live sound for any auditorium, one must deal with the acoustic parameters of that particular auditorium. This requires various adjustments of equalization, bass, treble, volume, etc. While those adjustments may enhance the sound quality in the auditorium, they are not necessarily needed for the recording. In fact, those house mix adjustments often diminish the quality of the recorded sound. Once the bass, treble, volume and other effects of the house mix are added to the recording mix, it is most difficult to correct. In the reverse, adjustments and signal processing effects that are often used to enhance a recording mix are not always needed in the house mix.

Separating the house mix from the recording mix easily solves this problem. This is done with two mixing consoles and a transformer based microphone splitter and snake system. In simple terms, the wires from the house microphones are divided with transformers and sent to two separate mixers at the same time. Then, the operator of either mixer, house or recording, can access each of the house microphones as if they were the only one using it. The house mix adjustments will not affect the recording mix and vice versa. The outputs of the house mixer are designated to the live sound amplifiers and speakers. The outputs of the recording mixer are designated to the recording and/or broadcast equipment.

Advantages and disadvantages

Advantages

- There is no added noise, (unintentional) distortion, or other signal degradation while the signal is in the digital domain, between the output of the analog to digital converter (ADC) and the input to the digital to analog converter (DAC).
- Aux sends can be mixed on the main faders rather than on a row of potentiometers.
- Signal routing is often much more flexible than with an analog-based console.
- The setup of the console can be saved and loaded at will. This is particularly useful in live events where a setup for each band can be largely prepared in advance, saved, and then loaded as needed.
- There are typically many on-board effects and virtual signal processors available, eliminating the need for additional hardware modules, and the associated cost, size, weight, cabling, signal quality issues, etc.

Disadvantages

- There is an analog to digital conversion, then processing of the signal, then again digital to analog conversion, which degrades the sound quality. This is subject to debate, since the quality degradation is not always noticeable.
- The number of faders is often less than the number of input channels. The extra input channels are not accessible until a bank of faders is switched to control them.
- Digital conversion and processing adds latency, or delay, into the signal.
- The act of making adjustments is often slower for compact digital mixers which require the user to page through one or more layers of commands before reaching the desired control.

Popular product examples

Klotz Digital's Vadis DC II mixer in use at a Virgin Radio outside broadcast

- - Allen & Heath iLive
- - Cadac Electronics S-Digital
- - DiGiCo D1 Live, D5 Live, D5T, SD7, SD8, SD8-24, SD9
- - Avid D-Show Venue, D-Show Profile, and SC 48

- - Innovason SY48 and SY80
- - Klotz Digital's AEON and D.C.II, in the radio broadcast segment
- - Mackie TT24
- - Midas XL8, PRO3, PRO6, PRO9, and the VeniceF series (analog console with digital I/O capabilities)
- - PreSonus StudioLive 16.4.2 and 24.4.2
- - RML Labs Software Audio Console (SAC)
- - Roland M-380, M-400, Edirol M-16DX
- - Soundcraft Si1, Si2, Si3, Vi4 and Vi6
- - Studer Vista 8
- - Tascam DM3200 and DM4800
- - Yamaha 01V, LS9, M7CL, DM1000, DM2000, PM5D and PM1D

Chapter 3

Dolby Digital Plus and Digital Sound Revolution

Dolby Digital Plus

Dolby Digital Plus logo

Dolby Digital Plus (**DD+** or **E-AC-3** (Enhanced AC-3), and sometimes incorrectly as **EC-3**) is a digital audio compression scheme. It is an incompatible development of the technologies used in the earlier Dolby Digital system. E-AC-3 has a number of improvements aimed at increasing quality at a given bitrate compared with legacy Dolby Digital (AC-3). While legacy AC-3 supports up to 5 full-range audio channels at a coded bitrate of 640 kbit/s, E-AC-3 supports up to 13 full range audio channels at a coded bitrate of 6.144 Mbit/s peak.

Dolby Digital Plus bitstreams are not backward compatible with legacy Dolby Digital decoders, and decoders that output audio over legacy S/PDIF connections must transcode the bitstreams to an older format such as PCM, AC-3, or DTS.

Codec changes

The codec used by Dolby Digital Plus is based on the original Dolby Digital codec, but with several enhancements to improve coding efficiency:

- Transient pre-noise processing - to reduce "pre-noise" artifacts before sharp transients.

- Enhanced channel coupling – which maintains phase relationships between channels, and improves performance of matrix decoders.
- Adaptive hybrid transform processing – an improved bit allocation and quantization algorithm.

Dolby claims that these changes can result in bitrate improvements of up to 50% while still allowing for the signal to be efficiently converted to Dolby Digital for backwards compatibility.

Specifications

Dolby Digital Plus is capable of the following:

- Coded bitrate: 0.032 to 6.144 Mbit/s
- Audio Channels: 1.0 to 13.1 (i.e. from mono to 13 full range channels and a low frequency effects channel)
- Sample rate: 32, 44.1 or 48 kHz
- Bit depth: up to 24 bits per channel.

The full set of technical specifications for E-AC-3 is published in Annex E of ATSC A/52B, as well as in ETSI TS 102 366 V1.2.1 (2008-08).

Physical transport

As of 2007, HDMI 1.3 is the only means to transport a raw DD+ bitstream between two pieces of consumer equipment. The older and more widespread S/PDIF interface cannot directly transport DD+ bitstreams. A number of methods of transcoding exist to convert an E-AC-3 bitstream into a S/PDIF compatible bitstream.

HD DVD and Blu-ray

The maximum number of discrete coded channels is the same for both formats: 7.1. However, HD DVD and Blu-ray impose different technical constraints on the supported audio-codecs. Hence, the usage of DD+ differs substantially between HD DVD and Blu-ray.

Dolby Digital (AC-3) and Dolby Digital Plus (E-AC-3) bitrate comparison

Codec	HD DVD			Blu-ray		
	Decoding	Channels	Bitrate	Decoding	Channels	Bitrate
AC-3	mandatory	1 to 5.1	504 kbit/s	mandatory	1 to 5.1	640 kbit/s
E-AC-3	mandatory	1 to 7.1	3.0 Mbit/s	optional, available for rear channels only	6.1 to 7.1	1.7 Mbit/s
TrueHD	mandatory	1 or 2	18.0	optional	1 to 8	18.0

optional	3 to 8	Mbit/s 18.0 Mbit/s		Mbit/s

On HD DVD, DD+ is designated a mandatory audio codec. An HD DVD movie may use DD+ as the primary (or only) audio track. An HD DVD player is required to support DD+ audio by decoding and outputting it to the player's output jacks. As stored on disc, the DD+ bitstream can carry for any number of audio channels up to the maximum allowed, at any bitrate up to 3.0 Mbit/s.

On Blu-ray Disc, DD+ is an optional codec, and is deployed as an extension to a "core" AC-3 5.1 audiotrack. The AC-3 core is encoded at 640 kbit/s, carries 5 primary channels (and 1 LFE), and is independently playable as a movie audio track by any Blu-ray player. The DD+ extension bitstream is used on players that support it by replacing the rear channels in the 5.1 setup with higher fidelity versions, along with providing a possible channel extension to 6.1 or 7.1. The complete audio track is allowed a combined bitrate of 1.7 Mbit/s: 640 kbit/s for the AC-3 5.1 core, and 1 Mbit/s for the DD+ extension. During playback, both the core and extension bitstreams contribute to the final audio-output, according to rules embedded in the bitstream metadata.

Media players and downmixing

Generally, a Dolby Digital Plus bitstream can only be transported over an HDMI 1.3 or greater link. Older receivers support earlier versions of HDMI, or only have support for the S/PDIF system for digital audio, or analog inputs.

For non-HDMI 1.3 links, the player can decode the audio and then transmit it via a variety of different methods.

- Earlier versions of HDMI, such as HDMI 1.1, support PCM audio, where the player decodes the audio and transmits it losslessly as PCM over HDMI to the receiver.
- Some receivers and players support analog surround sound, and the player can decode the audio, and transmit it to the receiver as analog audio.

Most receivers and players support S/PDIF. This lower bandwidth digital connection is not capable of transmitting lossless audio with more than two channels, but a player can transmit a S/PDIF compatible audio stream to the receiver in one of the following ways:

- Blu-ray disc players can take advantage of the legacy 5.1 AC-3 bitstream embedded in the E-AC-3 bitstream, transmitting just the AC-3 bitstream with no modifications.
- Players supporting the HD DVD standard can transcode the decoded audio into another format. Depending upon the method and options available to the player, this can be done with relatively little quality loss. Dolby's reference decoder, available to all licensees, exploits the common heritage between AC-3 and DD+

by performing the operation in the frequency domain. Hybrid re-compression avoids unnecessary end-to-end decompression and subsequent recompression (DD+ → LPCM → AC-3.) In addition to AC-3, some HD DVD players transcode audio compatible with S/PDIF into 1.5 Mbit/s DTS audio.

Should the player need to decode the audio for a non-HDMI 1.3 receiver, the results should be predictable. The DD+ specification explicitly defines downmixing modes and mechanics, so any source soundfield (up to 14.1) can be reproduced predictably for any listening environment (down to a single channel).

Digital Sound Revolution

The **digital sound revolution** (or **digital audio revolution**) refers to the widespread adoption of digital audio technology in the computer industry beginning in the 1980s.

Prior methods

Software-based pulse-width modulation

Some of the first computer music was created in 1961 by LaFarr Stuart, who wrote software to modulate the duration of and between pulses (pulse-width modulation or "PWM", via a process now often referred to as "bit-banging") on a bus line that had been connected to an amplified speaker originally installed to monitor the functioning of Iowa State University's CYCLONE computer, a derivative of the Illiac. The entire computer was used to create simple, recognizable tunes using digital audio. A recording of an interview with Mr. Stuart and his computer music was broadcast nationally on the NBC radio network program Monitor (NBC Radio) on February 10, 1962.

The speakers in the IBM PC (released in 1981) and its successors may be used to create sounds and music using a similar mechanism.

FM synthesis

The first specialized audio circuits in computers included simple analog oscillators that could be set to desired frequencies, generally approximating tones along the musical scale. A base frequency was then modulated with analog filters to create desired effects; this process of audio waveform synthesis using frequency modulation is usually referred to as FM synthesis.

Early integrated circuit devices to incorporate FM synthesis methods include the Atari POKEY custom application-specific integrated circuit or "ASIC" (U.S. Patent 4,314,236 issued February 2, 1982) in the Atari 800 and the MOS Technology 6581/8580 SID chip (U.S. Patent 4,677,890, filed on February 27, 1983 and issued on July 7, 1987) used in the Commodore 64. The Yamaha OPL2 chip set (YM3812 and external digital-to-analog converter) was included on the AdLib sound card (1987), on the Creative Technology

Sound Blaster (1989) and (in pairs, to create stereo sound) on the Media Vision Pro AudioSpectrum (1991); these were replaced by the next generation Yamaha OPL3 chip set on the Pro AudioSpectrum 16 and Sound Blaster 16.

Digital-to-analog converters

As they became more cost-effective, digital-to-analog converter (often called "D-to-A"--abbreviated "D/A", or "DAC") integrated circuits augmented and ultimately replaced FM synthesis devices. These devices enabled computers to play digital audio using an encoding technique known as pulse-code modulation ("PCM"). Unlike pulse-width modulation ("PWM"), which turns a signal on and off, pulse-code modulation also allows the level of a signal to be set to several intermediate levels; in this regard, PWM and PCM are similar to black-and-white and grayscale images, respectively.

Digital audio compact discs (using PCM) were introduced in 1982. Starting in 1985, the medium was adapted for the storage of computer data via the Yellow Book and the High Sierra Format, which became the basis for the 1988 CD-ROM data standard ISO 9660.

The Apple Macintosh (1984) and Atari ST (1985) could produce digital audio via software. Without dedicated audio hardware, digital audio on these machines were usually limited to title screens in games (at higher sampling rates) or games which did not feature heavy animation which left enough CPU time to play lower quality samples.

The first computer to feature a digital sound processor was the Commodore Amiga released in 1985. The MOS Technology 8364 Paula chip featured four independent 8-bit D/A converters. The Paula features four mono audio channels, or two combined stereo channels. This meant for the first time a computer could play digital samples from memory with virtually no CPU usage, or any clever software tricks.

In 1989, the Creative Technology Sound Blaster featured a processor and digital-to-analog converter, and incorporated the Yamaha OPL2 chip set FM synthesis devices for compatibility with the AdLib sound card. In 1991, Media Vision introduced the original Pro AudioSpectrum, which offered similar functionality but introduced stereo sound, an audio mixer and CD-ROM interface (SCSI and many variants); its 16-bit successor, the Pro AudioSpectrum 16, offered CD-quality sound via its 16-bit compressor-decompressor ("CODEC").

In 1997, Intel Corporation created its Audio CODEC standard AC'97, which was superseded in 2004 by Intel High Definition Audio (HD Audio).

Compression

High fidelity audio hardware became inexpensive faster than data storage media, driving the development of compression techniques.

A popular early variant of pulse-code modulation ("PCM") was a compressed version called adaptive differential pulse-code modulation ("ADPCM").

Sound module files (originally Amiga .MOD files) enabled music to be created and shared via compact files and played back with high quality (using four channels, each at half the sampling rate of audio compact discs). Soon after the release of its Pro AudioSpectrum 16, Media Vision included with it a MOD file player and sample music files.

In the late 1990s, the MP3 format emerged, allowing music to be stored in relatively small files by using high compressions rates through a predictive synthesis technique. Modern computer CD-ROM drives allowed the redbook audio to be read in digital format (versus earlier drives that merely output analog audio), which allows entire volumes of music to be copied and encoded many times faster than normal playback speed.

Non-moving storage

After the year 2000, strong demand for small portable music players such as Apple's iPods drove competition in component sales, resulting in data storage devices becoming increasingly inexpensive. Ultimately, non-volatile semiconductor-based storage devices became less expensive than fixed hard disk drives.

Online music distribution

The popularity of high-quality compressed music and the widespread availability of Internet access enabled widespread copyright infringement (most notably through Napster) followed by widespread legitimate sales of music online through the Apple iTunes Music Store, Amazon.com, Walmart.com and others.

Chapter 4

Digital Audio Workstation

A **digital audio workstation** (**DAW**) is a music sequencer designed solely or primarily for recording, editing and playing back digital audio.

DAWs were originally tape-less, microprocessor-based systems such as the Synclavier and Fairlight CMI. Modern DAWs are software running on computers with audio interface hardware.

Integrated

An integrated DAW consists of a mixing console, control surface, audio converter, and data storage in one device. Integrated DAWs were more popular before personal computers became powerful enough to run DAW software. As computer power increased and price decreased, the popularity of the costly integrated systems dropped. However, systems such as the Orban Audicy once flourished in the radio and television markets. Today, some systems still offer computerless arranging and recording features with a full graphical user interface.

Software

A computer-based DAW has four basic components: a computer, an ADC-DAC (also called a sound card, audio interface, etc.), a digital audio editor software, and at least one input device for adding or modifying musical note data (this could be as simple as a mouse, and as sophisticated as a MIDI controller keyboard, or an automated fader board for mixing track volumes, etc.). The computer acts as a host for the sound card and software and provides processing power for audio editing. The sound card (if used) or external audio interface typically converts analog audio signals into digital form, and for playback converting digital to analog audio; it may also assist in further processing the audio. The software controls all related hardware components and provides a user interface to allow for recording, editing, and playback. Most computer-based DAWs have extensive MIDI recording, editing, and playback capabilities, and some even have minor video-related features.

Simple smartphone-based DAWs, called Mobile Audio Workstation (MAWs), are also available, used for example by journalists for recording and editing on location.

Common functionality

As software systems, DAWs could be designed with any user interface, but generally they are based on a multitrack tape recorder metaphor, making it easier for recording engineers and musicians already familiar with using tape recorders to become familiar with the new systems. Therefore, computer-based DAWs tend to have a standard layout which includes transport controls (play, rewind, record, etc.), track controls and/or a mixer, and a waveform display. In single-track DAWs, only one (mono or stereo form) sound is displayed at a time.

Multitrack DAWs support operations on multiple tracks at once. Like a mixing console, each track typically has controls that allow the user to adjust the overall volume and stereo balance (pan) of the sound on each track. In a traditional recording studio additional processing is physically plugged in to the audio signal path, a DAW however can also route in software or uses software plugins to process the sound on a track.

DAWs are capable of many of the same functions as a traditional tape-based studio setup, and in recent years have almost completely replaced them. Modern advanced recording studios may have multiple types of DAWs in them and it is not uncommon for a sound engineer and/or musician to travel with a portable laptop-based DAW, although interoperability between different DAWs is poor.

Perhaps the most significant feature available from a DAW that is not available in analogue recording is the ability to 'undo' a previous action. Undo makes it much easier to avoid accidentally permanently erasing or recording over a previous recording. If a mistake is made, the undo command is used to conveniently revert the changed data to a previous state. Cut, Copy, Paste, and Undo are familiar and common computer commands and usually available in DAWs in some form.

Commonly DAWs feature some form of automation, often performed through "envelopes". Envelopes are procedural line segment-based or curve-based interactive graphs. The lines and curves of the automation graph are joined by or comprise adjustable points. By creating and adjusting multiple points along a waveform or control events, the user can specify parameters of the output over time (e.g., volume or pan). Automation data may also be directly derived from human gestures recorded by a control surface or controller. MIDI is a common data protocol used for transferring such gestures to the DAW.

MIDI recording, editing, and playback is increasingly incorporated into modern DAWs of all types, as is Synchronization with other audio and/or video tools.

Early history

The earliest attempts at creating digital audio workstations in the 1970s and 80s were limited by factors such as the high price of storage, vastly smaller slower processing and disk speeds available. But in the face of this, the company Soundstream (who previously came to prominence in the early days of digital audio by releasing one of the first commercially available digital audio tape recorders in 1977), built what could be considered the first digital audio workstation in 1978, using some of the most current computer hardware of the time. The *Digital Editing System*, as Soundstream called it, consisted of a DEC PDP-11/60 minicomputer running a custom software package called DAP (Digital Audio Processor), a Braegen 14"-platter hard disk drive, a storage oscilloscope to display audio waveforms to be edited, a video display terminal for controlling the system, and interface cards that plugged into the PDP-11's Unibus slots (the *Digital Audio Interface*, or *DAI*) that provided analog and digital audio input and output for interfacing to both Soundstream's digital recorders and conventional analog tape recorders as well. The DAP software could perform edits to the audio recorded on the system's hard disks, as well as provide effects such as crossfades.

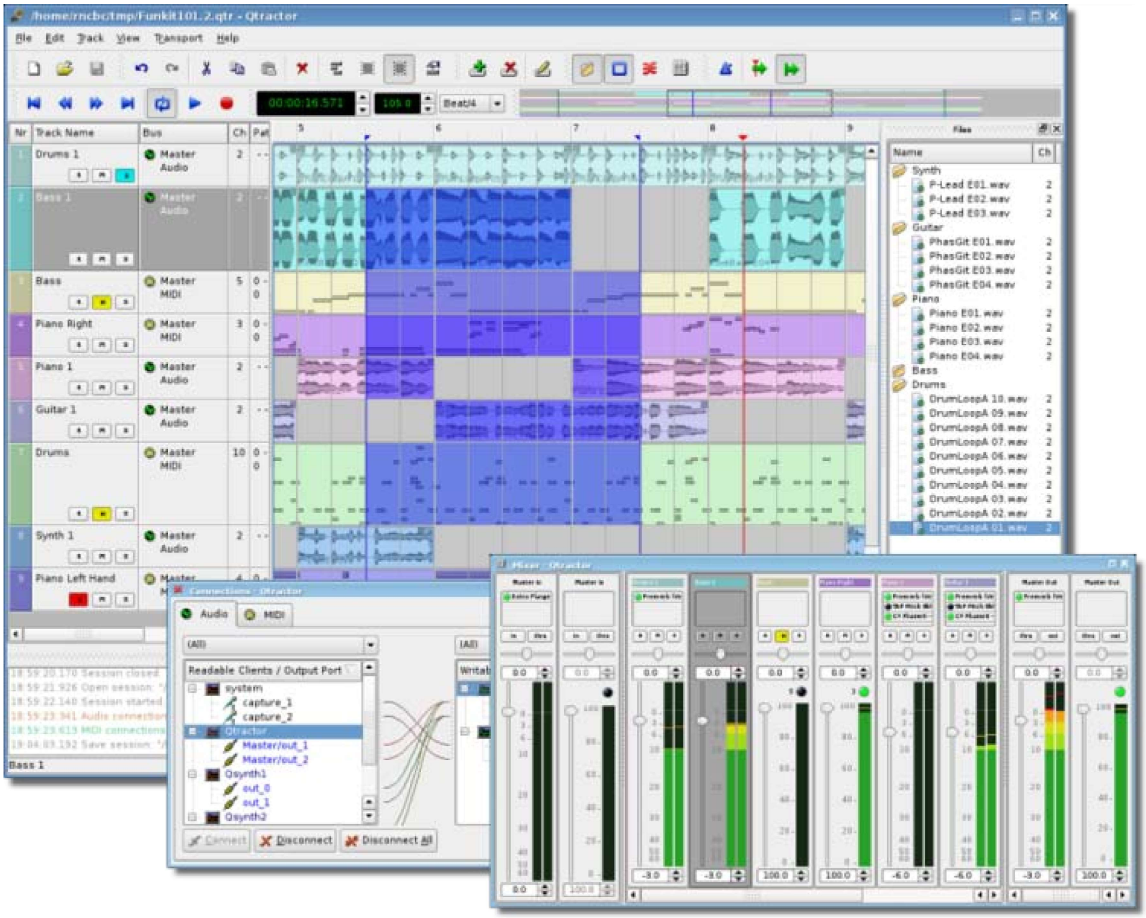
By the late 1980s, a number of consumer level computers such as the Apple Macintosh and Atari ST began to have enough power to handle the task of digital audio editing. Macromedia's Soundedit, with Microdeal's Replay Professional and Digidesign's "Sound Tools" and "Sound Designer" were used to edit audio samples for sampling keyboards like the E-mu Emulator II and the Akai S900, and soon went on to be used for simple two track audio editing and CD mastering purposes. In 1994, a company in California called OSC produced a 4 track editing-recorder application called DECK that ran on Digidesign's hardware system, and which was employed in the production of The Residents' "Freakshow" LP; this combination of audio software and hardware was one of the earliest examples of what we today would call a DAW.

Many major recording studios finally "went digital" after Digidesign introduced its Pro Tools software, modelled after the traditional method and signal flow present in almost all analog recording devices. At this time, most of the DAWs were Apple Mac based (e.g. Pro Tools, Studer Dyaxis, Sonic Solutions). Around 1992, the first Windows based DAWs started to emerge from companies such as Soundscape Digital Technology (which was later acquired by Mackie then by SSL), SADiE, Echo Digital Audio and Spectral Synthesis. All the systems at this point used dedicated hardware for their audio processing.

The first Windows based software-only product, introduced in 1993, was Samplitude Studio (which already existed in 1992 as an audio editor for the Commodore Amiga).

In 1996, German company Steinberg introduced Cubase VST, which could record and play back up to 32 tracks of digital audio on an Apple Macintosh without need of any external DSP hardware. Cubase not only modelled a tape-like interface for recording and editing, but also modelled the entire mixing desk and effects rack common in analog studios. This revolutionised the DAW world, both in features and price tag, and was quickly imitated by most of the other modern DAW systems we know today.

Free and open source software



Qtractor screenshot

There are many free and open-source software programs that can facilitate a DAW. These are often designed to run on a variety of operating systems and are usually developed non-commercially.

The development of digital audio for Linux and BSD fostered technologies such as ALSA, which drives audio hardware, and JACK. JACK allows any JACK-aware audio software to connect to any other audio software running on the system, such as connecting an ALSA- or OSS-driven soundcard to a mixing and editing front-end, like Ardour or Rosegarden. In this way, JACK acts as a virtual audio patch bay, and it can be configured to use a computer's resources in real time, with dedicated memory, and with various options that minimize the DAW's latency. This kind of abstraction and configuration allows DJs to use multiple programs for editing and synthesizing audio streams, or multitasking and duplexing, without the need for analogue conversion, or asynchronous saving and reloading files, and ensures a high level of audio fidelity.

- Audacity is a free and open-source digital audio editor that can run on Mac OS X, Microsoft Windows, and Linux; it is particularly popular in the podcast community, and also has a large following among the visually-impaired due to its keyboard interface.
- Macaw is a public domain source code music program that runs on Microsoft Windows. It features a large number of built in instrument synthesizers and effects, and it can load and play SoundFonts.
- Rosegarden is a multi-featured audio application that includes audio mixing plugins, a notation editor, and MIDI matrix editor. The MusE Sequencer is a similarly featured audio application that includes an audio mixer and a MIDI sequencer.

Other open-source programs include virtual synthesizers and MIDI controllers, such as those provided by FluidSynth and TiMidity. Both can load SoundFonts to expand the voices and instruments available for synthesis and expand the ports and channels available to synthesizers. Such virtualization allows users to expand the traditional limitations of ADC-DAC hardware.

The Linux Audio Development (LAD) mailing list is a major driving force in developing standards, such as the LADSPA, DSSI and LV2 plugin architectures. The Virtual Studio Technology (VST) plugin standard is supported as an option by some such programs but is generally implemented as a separate plugin, not a built-in option, due to Steinberg's licensing scheme. Among others, the creators of Audacity provide an optional, somewhat minimalist, VST-to-LADSPA bridge plugin for their software, but it is a separate download.

Chapter 5

MP3

MPEG-1 or MPEG-2 Audio Layer III, more commonly referred to as **MP3**, is a patented digital audio encoding format using a form of lossy data compression. It is a common audio format for consumer audio storage, as well as a de facto standard of digital audio compression for the transfer and playback of music on digital audio players.

MP3 is an audio-specific format that was designed by the Moving Picture Experts Group as part of its MPEG-1 standard and later extended in MPEG-2 standard. The first MPEG subgroup - *Audio* group was formed by several teams of engineers at Fraunhofer IIS, University of Hannover, AT&T-Bell Labs, Thomson-Brandt, CCETT, and others. MPEG-1 Audio (MPEG-1 Part 3), which included MPEG-1 Audio Layer I, II and III was approved as a committee draft of ISO/IEC standard in 1991, finalised in 1992 and published in 1993 (ISO/IEC 11172-3:1993). Backwards compatible MPEG-2 Audio (MPEG-2 Part 3) with additional bit rates and sample rates was published in 1995 (ISO/IEC 13818-3:1995).

The use in MP3 of a lossy compression algorithm is designed to greatly reduce the amount of data required to represent the audio recording and still sound like a faithful reproduction of the original uncompressed audio for most listeners. An MP3 file that is created using the setting of 128 kbit/s will result in a file that is about 11 times smaller than the CD file created from the original audio source. An MP3 file can also be constructed at higher or lower bit rates, with higher or lower resulting quality.

The compression works by reducing accuracy of certain parts of sound that are considered to be beyond the auditory resolution ability of most people. This method is commonly referred to as perceptual coding. It uses psychoacoustic models to discard or reduce precision of components less audible to human hearing, and then records the remaining information in an efficient manner.

History

Development

The MP3 lossy audio data compression algorithm takes advantage of a perceptual limitation of human hearing called auditory masking. In 1894, Alfred Marshall Mayer reported that a tone could be rendered inaudible by another tone of lower frequency. In 1959, Richard Ehmer described a complete set of auditory curves regarding this phenomenon. Ernst Terhardt *et al.* created an algorithm describing auditory masking with high accuracy. This work added to a variety of reports from authors dating back to Fletcher, and to the work that initially determined critical ratios and critical bandwidths.

The psychoacoustic masking codec was first proposed in 1979, apparently independently, by Manfred R. Schroeder, et al. from AT&T-Bell Labs in Murray Hill, NJ, and M. A. Krasner both in the United States. Krasner was the first to publish and to produce hardware for speech (not usable as music bit compression), but the publication of his results as a relatively obscure Lincoln Laboratory Technical Report did not immediately influence the mainstream of psychoacoustic codec development. Manfred Schroeder was already a well-known and revered figure in the worldwide community of acoustical and electrical engineers, but his paper was not much noticed, since it described negative results due to the particular nature of speech and the linear predictive coding (LPC) gain present in speech. Both Krasner and Schroeder built upon the work performed by Eberhard F. Zwicker in the areas of tuning and masking of critical bands, that in turn built on the fundamental research in the area from Bell Labs of Harvey Fletcher and his collaborators. A wide variety of (mostly perceptual) audio compression algorithms were reported in IEEE's refereed Journal on Selected Areas in Communications. That journal reported in February 1988 on a wide range of established, working audio bit compression technologies, some of them using auditory masking as part of their fundamental design, and several showing real-time hardware implementations.

The immediate predecessors of MP3 were "Optimum Coding in the Frequency Domain" (OCF), and Perceptual Transform Coding (PXFM). These two codecs, along with block-switching contributions from Thomson-Brandt, were merged into a codec called ASPEC, which was submitted to MPEG, and which won the quality competition, but that was mistakenly rejected as too complex to implement. The first practical implementation of an audio perceptual coder (OCF) in hardware (Krasner's hardware was too cumbersome and slow for practical use), was an implementation of a psychoacoustic transform coder based on Motorola 56000 DSP chips.

MP3 is directly descended from OCF and PXFM. MP3 represents the outcome of the collaboration of Karlheinz Brandenburg, working as a postdoc at AT&T-Bell Labs with James D. (JJ) Johnston of AT&T-Bell Labs, collaborating with the Fraunhofer Society for Integrated Circuits, Erlangen, with relatively minor contributions from the MP2 branch of psychoacoustic sub-band coders.

MPEG-1 Audio Layer 2 encoding began as the Digital Audio Broadcast (DAB) project managed by Egon Meier-Engelen of the *Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt* (later on called *Deutsches Zentrum für Luft- und Raumfahrt*, German Aerospace Center) in Germany. The European Community financed this project, commonly known as EU-147 (or Eureka 147), from 1987 to 1994 as a part of the EUREKA research program. MUSICAM Audio Coding was developed as part of the Eureka 147 project and has been subject to the standardization process within the ISO/Moving Pictures Expert Group (MPEG).

As a doctoral student at Germany's University of Erlangen-Nuremberg, Karlheinz Brandenburg began working on digital music compression in the early 1980s, focusing on how people perceive music. He completed his doctoral work in 1989 and became an assistant professor at Erlangen-Nuremberg. While there, he continued to work on music compression with scientists at the Fraunhofer Society (in 1993 he joined the staff of the Fraunhofer Institute).

Standardization

In 1991, there were only two proposals available that could be completely assessed for an MPEG audio standard: Musicam (Masking pattern adapted Universal Subband Integrated Coding And Multiplexing) (a.k.a. *MUSICAM*) and ASPEC (Adaptive Spectral Perceptual Entropy Coding). The Musicam technique, as proposed by Philips (The Netherlands), CCETT (France) and Institut für Rundfunktechnik (Germany) was chosen due to its simplicity and error robustness, as well as its low computational power associated with the encoding of high quality compressed audio. The Musicam format, based on sub-band coding, was the basis of the MPEG Audio compression format (sampling rates, structure of frames, headers, number of samples per frame).

Much of its technology and ideas were incorporated into the definition of ISO MPEG Audio Layer I and Layer II and the filter bank alone into Layer III (MP3) format as part of the computationally inefficient hybrid filter bank. Under the chairmanship of Professor Musmann (University of Hannover) the editing of the standard was made under the responsibilities of Leon van de Kerkhof (Layer I) and Gerhard Stoll (Layer II).

ASPEC was the joint proposal of AT&T Bell Laboratories, Thomson Consumer Electronics, Fraunhofer Society and CNET. It provided the highest coding efficiency.

A working group consisting of Leon van de Kerkhof (The Netherlands), Gerhard Stoll (Germany), Leonardo Chiariglione (Italy), Yves-François Dehery (France), Karlheinz Brandenburg (Germany) and James D. Johnston (USA) took ideas from ASPEC, integrated the filter bank from Layer 2, added some of their own ideas and created MP3, which was designed to achieve the same quality at 128 kbit/s as MP2 at 192 kbit/s.

All algorithms for MPEG-1 Audio Layer I, II and III were approved in 1991 and finalized in 1992 as part of MPEG-1, the first standard suite by MPEG, which resulted in the international standard **ISO/IEC 11172-3** (a.k.a. *MPEG-1 Audio* or *MPEG-1 Part 3*), published in 1993. Further work on MPEG audio was finalized in 1994 as part of the second suite of MPEG standards, MPEG-2, more formally known as international standard **ISO/IEC 13818-3** (a.k.a. *MPEG-2 Part 3* or backwards compatible *MPEG-2 Audio* or *MPEG-2 Audio BC*), originally published in 1995. MPEG-2 Part 3 (ISO/IEC 13818-3) defined additional bit rates and sample rates for MPEG-1 Audio Layer I, II and III. The new sampling rates are exactly half that of those originally defined for MPEG-1 Audio. MPEG-2 Part 3 also enhanced MPEG-1's audio by allowing the coding of audio programs with more than two channels, up to 5.1 multichannel. There is also *MPEG-2.5* audio, a proprietary unofficial extension developed by Fraunhofer IIS. It enables MP3 to work satisfactorily at very low bitrates and added lower sampling frequencies. MPEG-2.5 was not developed by MPEG and was never approved as an international standard.

MPEG Audio Layer III versions

Version	International Standard[*]	First public release date (First edition)	Latest public release date (edition)
MPEG-1 Audio Layer III	ISO/IEC 11172-3 (MPEG-1 Part 3)	1993	
MPEG-2 Audio Layer III	ISO/IEC 13818-3 (MPEG-2 Part 3)	1995	1998
MPEG-2.5 Audio Layer III	nonstandard, proprietary		

- Note: The ISO standard ISO/IEC 11172-3 (a.k.a. MPEG-1 Audio) defined three formats: the MPEG-1 Audio Layer I, Layer II and Layer III. The ISO standard ISO/IEC 13818-3 (a.k.a. MPEG-2 Audio) defined extended version of the MPEG-1 Audio - MPEG-2 Audio Layer I, Layer II and Layer III. MPEG-2 Audio (MPEG-2 Part 3) should not be confused with MPEG-2 AAC (MPEG-2 Part 7 - ISO/IEC 13818-7).

Compression efficiency of encoders is typically defined by the bit rate, because compression ratio depends on the bit depth and sampling rate of the input signal. Nevertheless, compression ratios are often published. They may use the Compact Disc (CD) parameters as references (44.1 kHz, 2 channels at 16 bits per channel or 2×16 bit), or sometimes the Digital Audio Tape (DAT) SP parameters (48 kHz, 2×16 bit). Compression ratios with this latter reference are higher, which demonstrates the problem with use of the term *compression ratio* for lossy encoders.

Karlheinz Brandenburg used a CD recording of Suzanne Vega's song "Tom's Diner" to assess and refine the MP3 compression algorithm. This song was chosen because of its nearly monophonic nature and wide spectral content, making it easier to hear imperfections in the compression format during playbacks. Some jokingly refer to Suzanne Vega as "The mother of MP3". Some more critical audio excerpts (glockenspiel, triangle, accordion, etc.) were taken from the EBU V3/SQAM reference compact disc and have been used by professional sound engineers to assess the subjective quality of the MPEG Audio formats. This particular track has an interesting property in that the two channels are almost, but not completely, the same, leading to a case where Binaural Masking Level Depression causes spatial unmasking of noise artifacts unless the encoder properly recognizes the situation and applies corrections similar to those detailed in the MPEG-2 AAC psychoacoustic model.

Going public

A reference simulation software implementation, written in the C language and later known as *ISO 11172-5*, was developed (in 1991–1996) by the members of the ISO MPEG Audio committee in order to produce bit compliant MPEG Audio files (Layer 1, Layer 2, Layer 3). It was approved as a committee draft of ISO/IEC technical report in March 1994 and printed as document CD 11172-5 in April 1994. It was approved as a draft technical report (DTR/DIS) in November 1994, finalized in 1996 and published as international standard ISO/IEC TR 11172-5:1998 in 1998. The reference software in C language was later published as a freely available ISO standard. Working in non-real time on a number of operating systems, it was able to demonstrate the first real time hardware decoding (DSP based) of compressed audio. Some other real time implementation of MPEG Audio encoders were available for the purpose of digital broadcasting (radio DAB, television DVB) towards consumer receivers and set top boxes.

On July 7, 1994, the Fraunhofer Society released the first software MP3 encoder called l3enc. The filename extension *.mp3* was chosen by the Fraunhofer team on July 14, 1995 (previously, the files had been named *.bit*). With the first real-time software MP3 player Winplay3 (released September 9, 1995) many people were able to encode and play back MP3 files on their PCs. Because of the relatively small hard drives back in that time (~ 500–1000 MB) lossy compression was essential to store non-instrument based music for playback on computer.

Internet

From the second half of 1994 through the late 1990s, MP3 files began to spread on the Internet. The popularity of MP3s began to rise rapidly with the advent of Nullsoft's audio player Winamp, released in 1997. In 1998, the first portable solid state digital audio player MPMan, developed by SaeHan Information Systems which is headquartered in Seoul, South Korea, was released and the Rio PMP300 was sold afterwards, despite legal suppression efforts by the RIAA.

In November 1997, the website mp3.com was offering thousands of MP3s created by independent artists for free. The small size of MP3 files enabled widespread peer-to-peer file sharing of music ripped from CDs, which would have previously been nearly impossible. The first large peer-to-peer filesharing network, Napster, was launched in 1999.

The ease of creating and sharing MP3s resulted in widespread copyright infringement. Major record companies argue that this free sharing of music reduces sales, and call it "music piracy". They reacted by pursuing lawsuits against Napster (which was eventually shut down and later sold) and against individual users who engaged in file sharing.

Despite the popularity of the MP3 format, online music retailers often use other proprietary formats that are encrypted or obfuscated in order to make it difficult to use purchased music files in ways not specifically authorized by the record companies. Attempting to control the use of files in this way is known as Digital Rights Management. Record companies argue that this is necessary to prevent the files from being made available on peer-to-peer file sharing networks. This has other side effects, though, such as preventing users from playing back their purchased music on different types of devices. However, the audio content of these files can usually be converted into an unencrypted format. For instance, users are often allowed to burn files to audio CD, which requires conversion to an unencrypted audio format.

Unauthorized MP3 file sharing continues on next-generation peer-to-peer networks. Some authorized services, such as Beatport, Bleep, Juno Records, eMusic, Zune Marketplace, Walmart.com, Rhapsody, the legal incarnation of Napster, and Amazon.com sell unrestricted music in the MP3 format.

Encoding audio

The MPEG-1 standard does not include a precise specification for an MP3 encoder, but does provide example psychoacoustic models, rate loop, and the like in the non-normative part of the original standard. At present, these suggested implementations are quite dated. Implementers of the standard were supposed to devise their own algorithms suitable for removing parts of the information from the audio input. As a result, there are many different MP3 encoders available, each producing files of differing quality. Comparisons are widely available, so it is easy for a prospective user of an encoder to research the best choice. It must be kept in mind that an encoder that is proficient at encoding at higher bit rates (such as LAME) is not necessarily as good at lower bit rates.

During encoding, 576 time-domain samples are taken and are transformed to 576 frequency-domain samples. If there is a transient, 192 samples are taken instead of 576. This is done to limit the temporal spread of quantization noise accompanying the transient.

Decoding audio

Decoding, on the other hand, is carefully defined in the standard. Most decoders are "bitstream compliant", which means that the decompressed output – that they produce from a given MP3 file – will be the same, within a specified degree of rounding tolerance, as the output specified mathematically in the ISO/IEC high standard document (ISO/IEC 11172-3). Therefore, comparison of decoders is usually based on how computationally efficient they are (i.e., how much memory or CPU time they use in the decoding process).

Audio quality

When performing lossy audio encoding, such as creating an MP3 file, there is a trade-off between the amount of space used and the sound quality of the result. Typically, the creator is allowed to set a bit rate, which specifies how many kilobits the file may use per second of audio. The higher the bit rate, the larger the compressed file will be, and, generally, the closer it will sound to the original file.

With too low a bit rate, compression artifacts (i.e. sounds that were not present in the original recording) may be audible in the reproduction. Some audio is hard to compress because of its randomness and sharp attacks. When this type of audio is compressed, artifacts such as ringing or pre-echo are usually heard. A sample of applause compressed with a relatively low bit rate provides a good example of compression artifacts.

Besides the bit rate of an encoded piece of audio, the quality of MP3 files also depends on the quality of the encoder itself, and the difficulty of the signal being encoded. As the MP3 standard allows quite a bit of freedom with encoding algorithms, different encoders may feature quite different quality, even with identical bit rates. As an example, in a public listening test featuring two different MP3 encoders at about 128 kbit/s, one scored 3.66 on a 1–5 scale, while the other scored only 2.22.

Quality is dependent on the choice of encoder and encoding parameters.

The simplest type of MP3 file uses one bit rate for the entire file — this is known as Constant Bit Rate (CBR) encoding. Using a constant bit rate makes encoding simpler and faster. However, it is also possible to create files where the bit rate changes throughout the file. These are known as Variable Bit Rate (VBR) files. The idea behind this is that, in any piece of audio, some parts will be much easier to compress, such as silence or music containing only a few instruments, while others will be more difficult to compress. So, the overall quality of the file may be increased by using a lower bit rate for the less complex passages and a higher one for the more complex parts. With some encoders, it is possible to specify a given quality, and the encoder will vary the bit rate accordingly. Users who know a particular "quality setting" that is transparent to their ears can use this value when encoding all of their music, and not need to worry about performing personal listening tests on each piece of music to determine the correct bit rate.

Perceived quality can be influenced by listening environment (ambient noise), listener attention, and listener training and in most cases by listener audio equipment (such as sound cards, speakers and headphones).

A test given to new students by Stanford University Music Professor Jonathan Berger showed that student preference for MP3 quality music has risen each year. Berger said the students seem to prefer the 'sizzle' sounds that MP3s bring to music.

Bit rate

Several bit rates are specified in the MPEG-1 Audio Layer III standard: 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256 and 320 kbit/s, and the available sampling frequencies are 32, 44.1 and 48 kHz. Additional extensions were defined in MPEG-2 Audio Layer III: bit rates 8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 144, 160 kbit/s and sampling frequencies 16, 22.05 and 24 kHz.

A sample rate of 44.1 kHz is almost always used, because this is also used for CD audio, the main source used for creating MP3 files. A greater variety of bit rates are used on the Internet. The rate of 128 kbit/s is commonly used, at a compression ratio of 11:1, offering adequate audio quality in a relatively small space. As Internet bandwidth availability and hard drive sizes have increased, higher bit rates up to 320 kbps are widespread.

Uncompressed audio as stored on an audio-CD has a bit rate of 1,411.2 kbit/s, so the bitrates 128, 160 and 192 kbit/s represent compression ratios of approximately 11:1, 9:1 and 7:1 respectively.

Non-standard bit rates up to 640 kbit/s can be achieved with the LAME encoder and the freeformat option, although few MP3 players can play those files. According to the ISO standard, decoders are only required to be able to decode streams up to 320 kbit/s.

MPEG-1 and MPEG-2 Audio Layer III
available bit rates (kbit/s)

MPEG-1 Audio Layer III	**MPEG-2 Audio Layer III**	**nonstandard proprietary MPEG-2.5 Audio Layer III**
-	8	8
-	16	16
-	24	24
32	32	32
40	40	40
48	48	48
56	56	56
64	64	64
80	80	80
96	96	96
112	112	112
128	128	128
-	144	144
160	160	160
192	-	-
224	-	-

256	-	-
320	-	-

MPEG-1 and MPEG-2 Audio Layer III available sampling rates (Hz)

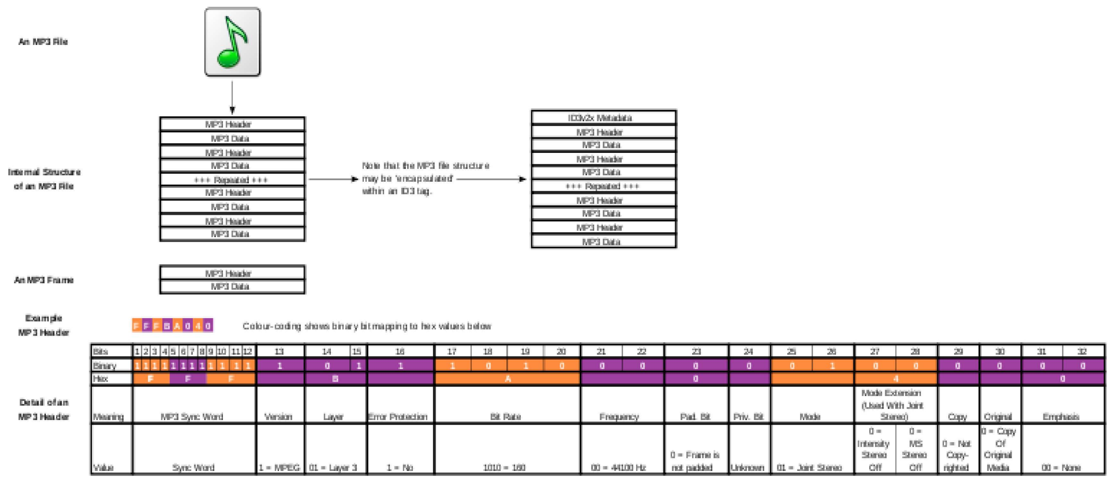
MPEG-1 Audio Layer III	**MPEG-2 Audio Layer III**	**nonstandard proprietary MPEG-2.5 Audio Layer III**
-	-	8000 Hz
-	-	11025 Hz
-	-	12000 Hz
-	16000 Hz	-
-	22050 Hz	-
-	24000 Hz	-
32000 Hz	-	-
44100 Hz	-	-
48000 Hz	-	-

VBR

MPEG audio may use variable bitrate (VBR), accomplished via bitrate switching on a per-frame basis, but only layer III decoders must support it. VBR is used when the goal is to achieve a fixed level of quality. The final file size of a VBR encoding is less predictable than with constant bitrate. Average bitrate is VBR implemented as a compromise between the two – the bitrate is allowed to vary for more consistent quality, but is controlled to remain near an average value chosen by the user, for predictable file sizes. Although an MP3 decoder must support VBR to be standards compliant, historically some decoders have bugs with VBR decoding, particularly before VBR encoders became widespread.

Layer III audio can also use a "bit reservoir", a partially full frame's ability to hold part of the next frame's audio data, allowing temporary changes in effective bitrate, even in a constant bitrate stream.

File structure



An MP3 file is made up of multiple MP3 frames, which consist of a header and a data block. This sequence of frames is called an elementary stream. Frames are not independent items ("byte reservoir") and therefore cannot be extracted on arbitrary frame boundaries. The MP3 Data blocks contain the (compressed) audio information in terms of frequencies and amplitudes. The diagram shows that the MP3 Header consists of a sync word, which is used to identify the beginning of a valid frame. This is followed by a bit indicating that this is the MPEG standard and two bits that indicate that layer 3 is used; hence MPEG-1 Audio Layer 3 or MP3. After this, the values will differ, depending on the MP3 file. *ISO/IEC 11172-3* defines the range of values for each section of the header along with the specification of the header. Most MP3 files today contain ID3 metadata, which precedes or follows the MP3 frames; as noted in the diagram.

Design limitations

There are several limitations inherent to the MP3 format that cannot be overcome by any MP3 encoder. Newer audio compression formats such as Vorbis, WMA Pro and AAC are generally void of a number of these limitations. In technical terms, some limitations include:

- Time resolution can be too low for highly transient signals and may cause smearing of percussive sounds.
- Due to the tree structure of the filter bank, pre-echo problems are made worse, as the combined impulse response of the two filter banks does not, and cannot, provide an optimum solution in time/frequency resolution.
- The combining of the two filter banks' outputs creates aliasing problems that must be handled partially by the "aliasing compensation" stage; however, that creates excess energy to be coded in the frequency domain, thereby decreasing coding efficiency.

- Frequency resolution is limited by the small long block window size, which decreases coding efficiency.
- There is no scale factor band for frequencies above 15.5/15.8 kHz.
- Joint stereo is done only on a frame-to-frame basis.
- Internal handling of the bit reservoir increases encoding delay.
- Encoder/decoder overall delay is not defined, which means there is no official provision for gapless playback. However, some encoders such as LAME can attach additional metadata that will allow players that can handle it to deliver seamless playback.
- The data stream can contain an optional checksum, but the checksum *only* protects the header data, not the audio data.

ID3 and other tags

A "tag" in an audio file is a section of the file that contains metadata such as the title, artist, album, track number or other information about the file's contents. The MP3 standards do not define tag formats for MP3 files, nor is there a standard container format that would support metadata and obviate the need for tags.

However, several *de facto* standards for tag formats exist. As of 2010, the most widespread are ID3v1 and ID3v2, and the more recently introduced APEv2. These tags are normally embedded at the beginning or end of MP3 files, separate from the actual MP3 frame data. MP3 decoders normally either read info from the tags, or just treat them as ignorable, non-MP3 junk data.

Playing & editing software often contains tag editing functionality, but there are also tag editor applications dedicated to the purpose.

Aside from metadata pertaining to the audio content, tags may also be used for DRM.

Volume normalization

Volume normalization or "volume levelling" is a 'non standardized frame, mp3 related or not' with metadata values (i.e. TAG-frame) metering the amplitude of the encoded data. The APE structure defined by independent programmer Frank Klemm for proprietary program MediaMonkey of Matthew T. Ashland has a levelling field to level out the output.

Background

Since volume levels of different audio sources can vary greatly, it is sometimes desirable to adjust the playback volume of audio files such that a consistent average volume is perceived. The idea is to control the *average* volume across multiple files, not the volume *peaks* in a single file. This *gain normalization*, while similar in purpose, is distinct from dynamic range compression (DRC), which is a form of normalization used in audio mastering. Gain normalization may defeat the intent of recording artists and audio engineers who deliberately set the volume levels of the audio they recorded.

Implementation

A standard for storing the average volume of an MP3 file in its metadata tag, enabling a specially designed player to automatically adjust the overall playback volume for each file, was proposed as Replay Gain function. The less popular and lesser implementation of such is "Replay Gain", which is not MP3-specific. When used on MP3 files, it is recorder differently in the file by different encoders. As of 2008 there was few 'Replay Gain'-ware players to support all implementations.

Licensing and patent issues

Many organizations have claimed ownership of patents related to MP3 decoding or encoding. These claims have led to a number of legal threats and actions from a variety of sources, resulting in uncertainty about which patents must be licensed in order to create MP3 products without committing patent infringement in countries that allow software patents.

The various MP3-related patents expire on dates ranging from 2007 to 2017 in the U.S. The initial near-complete MPEG-1 standard (parts 1, 2 and 3) was publicly available on December 6, 1991 as ISO CD 11172. In the United States, patents cannot claim inventions that were already publicly disclosed more than a year prior to the filing date, but for patents filed prior to June 8, 1995, submarine patents made it possible to extend the effective lifetime of a patent through application extensions. Patents filed for anything disclosed in ISO CD 11172 a year or more after its publication are questionable; if only the known MP3 patents filed by December 1992 are considered, then MP3 decoding may be patent free in the US by December 2012.

Technicolor (formerly called Thomson Consumer Electronics) claims to control MP3 licensing of the Layer 3 patents in many countries, including the United States, Japan, Canada and EU countries. Technicolor has been actively enforcing these patents.

MP3 license revenues generated about €100 million for the Fraunhofer Society in 2005.

In September 1998, the Fraunhofer Institute sent a letter to several developers of MP3 software stating that a license was required to "distribute and/or sell decoders and/or encoders". The letter claimed that unlicensed products "infringe the patent rights of Fraunhofer and Thomson. To make, sell and/or distribute products using the [MPEG Layer-3] standard and thus our patents, you need to obtain a license under these patents from us."

However, there exist both free and/or proprietary alternatives, with free formats such as Vorbis, FLAC, and others. Microsoft's usage of its own proprietary Windows Media format allows it to avoid licensing issues associated with these patents by avoiding usage of the MP3 format entirely. Until the key patents expire, unlicensed encoders and players could be infringing in countries where the patents are valid.

In spite of the patent restrictions, the perpetuation of the MP3 format continues. The reasons for this appear to be the network effects caused by:

- familiarity with the format
- the large quantity of music now available in the MP3 format
- the wide variety of existing software and hardware that takes advantage of the file format and does not support the alternatives
- the lack of DRM restrictions, which makes MP3 files easy to edit, copy and play in different portable digital players (Samsung, Apple, Creative, etc.)
- the majority of home users not knowing or not caring about the patents' existence and often not considering such legal issues when choosing their music format for personal use

Additionally, patent holders declined to enforce license fees on free and open source decoders, which allows many free MP3 decoders to develop. Thus, while patent fees have been an issue for companies that attempt to use MP3, they have not meaningfully impacted personal use.

Sisvel S.p.A. and its U.S. subsidiary Audio MPEG, Inc. previously sued Thomson for patent infringement on MP3 technology, but those disputes were resolved in November 2005 with Sisvel granting Thomson a license to their patents. Motorola also recently signed with Audio MPEG to license MP3-related patents.

In September 2006, German officials seized MP3 players from SanDisk's booth at the IFA show in Berlin after an Italian patents firm won an injunction on behalf of Sisvel against SanDisk in a dispute over licensing rights. The injunction was later reversed by a Berlin judge, but that reversal was in turn blocked the same day by another judge from the same court, "bringing the Patent Wild West to Germany" in the words of one commentator.

In February 2007, Texas MP3 Technologies sued Apple, Samsung Electronics and Sandisk in eastern Texas federal court, claiming infringement of a portable MP3 player patent that Texas MP3 said it had been assigned. Apple and Sandisk both settled the claims against them in January 2009. Samsung settled as well.

Alcatel-Lucent has asserted several MP3 coding and compression patents, allegedly inherited from AT&T-Bell Labs, in litigation of its own. In November 2006 (prior to the companies' merger), Alcatel sued Microsoft for allegedly infringing seven patents. On February 23, 2007, a San Diego jury awarded Alcatel-Lucent US $1.52 billion in damages for infringement of two of them. The court subsequently tossed the award, however, finding that one patent had not been infringed and that the other was not even owned by Alcatel-Lucent; it was co-owned by AT&T and Fraunhofer, who had licensed it to Microsoft, the judge ruled. That defense judgment was upheld on appeal in 2008.

Alternative technologies

Many other lossy and lossless audio codecs exist. Among these, mp3PRO, AAC, and MP2 are all members of the same technological family as MP3 and depend on roughly similar psychoacoustic models. The Fraunhofer Gesellschaft owns many of the basic patents underlying these codecs as well, with others held by Dolby Labs, Sony, Thomson Consumer Electronics, and AT&T. In addition, there is also the open source file format Vorbis that has been available free of charge and without any known patent restrictions.

Chapter 6

Digital Audio Tape

A 90-minute DAT cartridge, size compared to a AAA (LR03) battery.

Digital Audio Tape (**DAT** or **R-DAT**) is a signal recording and playback medium developed by Sony and introduced in 1987. In appearance it is similar to a compact audio cassette, using 4 mm magnetic tape enclosed in a protective shell, but is roughly half the size at 73 mm × 54 mm × 10.5 mm. As the name suggests, the recording is digital rather than analog. DAT has the ability to record at higher, equal or lower sampling rates than a CD (48, 44.1 or 32 kHz sampling rate respectively) at 16 bits quantization. If a digital source is copied then the DAT will produce an exact clone, unlike other digital media such as Digital Compact Cassette or non-Hi-MD MiniDisc, both of which use lossy data compression.

Like most formats of videocassette, a DAT cassette may only be recorded on one side, unlike an analog compact audio cassette.

Although intended as a replacement for audio cassettes, the format was never widely adopted by consumers because of issues of expense and concerns about unauthorized digital quality copies. The format saw moderate success in professional markets and as a computer storage medium. As Sony has ceased production of new recorders, it will become more difficult to play archived recordings in this format unless they are copied to other formats or hard drives.

History

Development

The technology of DAT is closely based on that of video recorders, using a rotating head and helical scan to record data. This prevents DATs from being physically edited in the cut-and-splice manner of analog tapes, or open-reel digital tapes like ProDigi or DASH.

The DAT standard allows for four sampling modes: 32 kHz at 12 bits, and 32 kHz, 44.1 kHz or 48 kHz at 16 bits. Certain recorders operate outside the specification, allowing recording at 96 kHz and 24 bits (HHS). Some machines aimed at the domestic market did not operate at 44.1 kHz when recording from analog sources. Since each recording standard uses the same tape, the quality of the sampling has a direct relation to the duration of the recording – 32 kHz at 12 bits will allow six hours of recording onto a three hour tape while HHS will only give 90 minutes from the same tape. Included in the signal data are subcodes to indicate the start and end of tracks or to skip a section entirely; this allows for indexing and fast seeking. Two-channel stereo recording is supported under all sampling rates and bit depths, but the R-DAT standard does support 4-channel recording at 32 kHz.

DAT "tapes" are between 15 and 180 minutes in length, a 120-minute tape being 60 meters in length. DAT "tapes" longer than 60 meters tend to be problematic in DAT recorders due to the thinner media. DAT machines running at 48 kHz and 44.1 kHz sample rates transport the tape at 8.15 mm/s. DAT machines running at 32 kHz sample rate transport the tape at 4.075 mm/s.

Predecessor formats

DAT was not the first digital audio tape; pulse-code modulation (PCM) was used in Japan by Denon in 1972 for the mastering and production of analogue phonograph records, using a 2-inch Quadruplex-format videotape recorder for its transport, but this was not developed into a consumer product. Denon's development dated from its work with Japan's NHK Broadcasting; NHK developed the first high-fidelity PCM audio recorder in the late 1960's. Denon continued development of their PCM recorders that used professional video machines as the storage medium, eventually building 8-track units used for, among other productions, a series of jazz records made in New York in the late 1970's.

In 1976, another digital audio tape format was developed by Soundstream, using 1 in (25.4 mm) wide reel-to-reel tape loaded on an instrumentation recorder manufactured by Honeywell acting as a transport, which in turn was connected to outboard digital audio encoding and decoding hardware of Soundstream's own design. Soundstream's format was improved through several prototypes and when it was developed to 50khz sampling rate at 16 bits, it was deemed good enough for professional classical recording by the company's first client, Telarc Records of Cleveland, Ohio. Telarc's April, 1978 recording of the Holst Suites for Band by Fred Fennell and the Cleveland Wind Ensemble was a landmark release, and ushered in digital recording for America's classical music labels. Soundstream's system was also used by RCA.

Starting in 1978, 3M introduced its own line and format of digital audio tape recorders for use in a recording studio. One of the first prototypes of 3M's system was installed in the studios of Sound 80 in Minneapolis, Minnesota. This system was used in June 1978 to record Aaron Copland's "Appalachian Spring" by the St. Paul Chamber Orchestra conducted by Dennis Russell Davies. That record was the first Grammy-winning digital recording. The production version of the 3M Digital Mastering System was used in 1979 to record the first all-digital rock album, Ry Cooder's "Bop Till You Drop," made at Warner Brothers Studio in California.

The first consumer-oriented PCM format used consumer video tape formats (Beta and VHS) as the storage medium. These systems used the EIAJ digital format, which sampled at 44.056khz at 14 bits. The Sony PCM-F1 system debuted in 1981, and Sony from the start offered the option of 16-bit wordlength. Other systems were marketed by Akai, JVC, Nakamichi and others. Panasonic, via its Technics division, briefly sold a digital recorder that combined a EIAJ digital adapter with a VHS video transport, the SV-P100. These machines were marketed by consumer electronics companies to consumers, but they were very pricey compared to cassette or even reel-to-reel decks of the time. They did catch on with the more budget conscious professional recordists, and some boutique-label professional releases were recorded using these machines.

Starting in the early 1980's, professional systems using a PCM adaptor were also common as mastering formats. These systems digitized an analog audio signal and then encoded the resulting digital stream into an analog video signal so that a conventional VCR could be used as a storage medium.

One of the most significant examples of a PCM adaptor-based system was the Sony PCM-1600 digital audio mastering system, introduced in 1978. The PCM-1600 used a U-Matic-format VCR for its transport, connected to external digital audio processing hardware. It (and its later versions such as the PCM-1610 and 1630) was widely used for the production and mastering of some of the first Digital Audio CDs in the early 1980s. Once CD's were commercially introduced in 1983, tapes recorded on the PCM-1600 were sent to the CD pressing plants to be used to make the glass master disc for CD replication.

Other examples include dbx, Inc.'s Model 700 system, which, similar to modern Super Audio CDs, used a high sample-rate delta-sigma modulation rather than PCM; Decca's 1970s PCM system, which used a videotape recorder manufactured by IVC for a transport; and Mitsubishi's X-80 digital recorder, a 6.4 mm (¼ in) open reel digital mastering format that used a very unusual sampling rate of 50.4 kHz.

For high-quality studio recording, all of these formats were effectively made obsolete in the early 1980s by two competing reel-to-reel formats with stationary heads: Sony's DASH format and Mitsubishi's continuation of the X-80 recorder, which was improved upon to become the ProDigi format. (In fact, one of the first ProDigi-format recorders, the Mitsubishi X-86C, was playback-compatible with tapes recorded on an X-80.) Both of these formats remained popular as an analog alternative until the early 1990s, when hard disk recorders rendered them obsolete.

R-DAT and S-DAT (DCC)

The DAT recorder mechanism was considerably more complex and expensive than an analogue cassette deck mechanism due to the rotary helical scan head, therefore Philips & Panasonic Corporation developed a rival digital tape recorder system with a stationary head based on the analogue compact cassette. The DCC was cheaper and simpler mechanically than DAT, but did not make perfect digital copies as it used a lossy compression technique called PASC. (Lossy compression was necessary to reduce the data rate to a level that a stationary head could cope with). DCC was never a competitor to DAT in recording studios because DAT was already established and as it was launched at the same time as Sony's Minidisc format (which has random access and editing features) it was not successful with consumers either. However, DCC proved that high quality digital recording could be achieved with a cheap simple mechanism using stationary heads.

Anti-DAT lobbying

In the late 1980s, the Recording Industry Association of America unsuccessfully lobbied against the introduction of DAT devices into the U.S. Initially, the organization threatened legal action against any manufacturer attempting to sell DAT machines in the country. It later sought to impose restrictions on DAT recorders to prevent them from being used to copy LPs, CDs, and prerecorded cassettes. One of these efforts, the Digital Audio Recorder Copycode Act of 1987 (introduced by Sen. Al Gore and Rep. Waxman), instigated by CBS Records president Walter Yetnikoff, involved a technology called CopyCode and required DAT machines to include a chip to detect attempts to copy material recorded with a notch filter, meaning that copyrighted prerecorded music, whether analog or digital, would have distorted sound. A National Bureau of Standards study showed that not only were the effects plainly audible, but that it was not even effective at preventing copying.

This opposition by CBS softened after Sony, a DAT manufacturer, bought CBS Records in January 1988. By June 1989, an agreement was reached, and the only concession the RIAA would receive was a more practical recommendation from manufacturers to Congress that legislation be enacted to require that recorders have a Serial Copy Management System to prevent digital copying for more than a single generation. This requirement was enacted as part of the Audio Home Recording Act of 1992, which also imposed taxes on DAT recorders and blank media.

Uses of DAT

Professional recording industry

DAT was widely used in the professional audio recording industry in the 1990s, and is still used to some extent as of 2010, particularly with regard to archives created in the '90s, although most labels have a program in place to transfer these tapes to a hard disk-based database. DAT is used professionally due to its lossless encoding, which allows a master tape to be created that is secure and does not induce tape noise (hiss) into the recording. In the correct setup, a DAT recording can be created without even having to be decoded to analogue until the final output stage, since digital multi-track recorders and digital mixing consoles can be used to create a fully digital chain. In this configuration, it is possible for the audio to remain digital from the first AD converter after the mic preamp until it is in a CD player.

DATs were also frequently used by radio broadcasters. Until recently, they were still used by the BBC as an emergency broadcast that would initiate if the player detected dead air for longer than a pre-determined time. This would mean that if for any reason the broadcast from the studio stopped, the DAT would continue broadcast until normal service could be resumed.

Amateur and home use

DAT was envisaged by proponents as the successor format to analogue audio cassettes in the way that the compact disc was the successor to vinyl-based recordings; however, the technology was never as commercially popular as CD. DAT recorders have remained relatively expensive, and commercial recordings are generally not made available on the format. However, DAT was, for a time, popular for making and trading recordings of live music, since available DAT recorders predated affordable CD recorders.

In the U.S., the RIAA and music publishers continued to lobby against DAT, arguing that consumers' ability to make perfect digital copies of music would destroy the market for commercial audio recordings. The opposition to DAT culminated in the passage of the resulting Audio Home Recording Act of 1992, which, among other things, effectively imposed a tax on DAT devices and blank media.

Computer data storage medium

The format was designed for audio use, but through the ISO Digital Data Storage standard it has been adopted for general data storage, storing from 1.3 to 80 GB on a 60 to 180 meter tape depending on the standard and compression. It is a sequential-access medium and is commonly used for backups. Due to the higher requirements for capacity and integrity in data backups, a computer-grade DAT was introduced, called DDS (Digital Data Storage). Although functionally similar to audio DATs, only a few DDS and DAT drives (in particular, those manufactured by Archive for SGI workstations) are capable of reading the audio data from a DAT cassette. SGI DDS4 drives no longer have audio support; SGI removed the feature due to "lack of demand".

Future of DAT

In November 2005, Sony announced that its remaining DAT machine models would be discontinued the following month. Sony has sold around 660,000 DAT products since its introduction in 1987. However, the DAT format still finds regular use in film and television recording, principally due to the support in some recorders for SMPTE time code synchronization. It is slowly being superseded by modern hard disk recording equipment which offers much more flexibility and storage. In 2004, Sony introduced the Hi-MD Walkman with the ability to record in linear PCM. Hi-MD has found some favor as a disc-based DAT alternative for field recordings and general portable playback.

Archived audio problem

The discontinuation of DAT player production lead to a significant problem regarding audio archives, since a tremendous number of recordings from the mid-80s until about 2000 exist solely on DAT.

Chapter 7

Windows Media Audio

Windows Media Audio (**WMA**) is an audio data compression technology developed by Microsoft. The name can be used to refer to its audio file format or its audio codecs. It is a proprietary technology that forms part of the Windows Media framework. WMA consists of four distinct codecs. The original WMA codec, known simply as *WMA*, was conceived as a competitor to the popular MP3 and RealAudio codecs. *WMA Pro*, a newer and more advanced codec, supports multichannel and high resolution audio. A lossless codec, *WMA Lossless*, compresses audio data without loss of audio fidelity (the regular WMA format is not lossless). And *WMA Voice*, targeted at voice content, applies compression using a range of low bit rates.

Development history

The first WMA codec was based on earlier work by Henrique Malvar and his team which was transferred to the Windows Media team at Microsoft. Malvar was a senior researcher and manager of the Signal Processing Group at Microsoft Research, whose team worked on the *MSAudio* project. The first finalized codec was initially referred to as *MSAudio 4.0*. It was later officially released as *Windows Media Audio*, as part of Windows Media Technologies 4.0. Microsoft claimed that WMA could produce files that were half the size of equivalent-quality MP3 files; Microsoft also claimed that WMA delivered "near CD-quality" audio at 64 kbit/s. The former claim however was rejected by some audiophiles. RealNetworks also challenged Microsoft's claims regarding WMA's superior audio quality compared to RealAudio.

Newer versions of WMA became available: *Windows Media Audio 2* in 1999, *Windows Media Audio 7* in 2000, *Windows Media Audio 8* in 2001, and *Windows Media Audio 9* in 2003. Microsoft first announced its plans to license WMA technology to third-parties in 1999. Although earlier versions of Windows Media Player played WMA files, support for WMA file creation was not added until the seventh version. In 2003, Microsoft released new audio codecs that were not compatible with the original WMA codec. These codecs were *Windows Media Audio 9 Professional*, *Windows Media Audio 9 Lossless*, and *Windows Media Audio 9 Voice*.

Container format

A WMA file is in most circumstances contained in the Advanced Systems Format (ASF), a proprietary Microsoft container format for digital audio or digital video.. The ASF container format specifies how metadata about the file is to be encoded, similar to the ID3 tags used by MP3 files. Metadata may include song name, track number, artist name, and also audio normalization values. This container can optionally support digital rights management (DRM) using a combination of elliptic curve cryptography key exchange, DES block cipher, a custom block cipher, RC4 stream cipher and the SHA-1 hashing function.

Codecs

Each WMA file features a single audio track in one of following codecs: WMA, WMA Pro, WMA Lossless, or WMA Voice. These codecs are technically distinct and mutually incompatible. Each codec is further explained below.

Windows Media Audio

Windows Media Audio (WMA) is the most common codec of the four WMA codecs. Colloquial usage of the term *WMA*, especially in marketing materials and device specifications, usually refers to this codec only. The first version of the codec released in 1999 is regarded as WMA 1. In the same year, the bit stream syntax, or compression algorithm, was altered in minor ways and became WMA 2. Since then, newer versions of the codec were released, but the decoding process remained the same, ensuring compatibility between codec versions. WMA is a lossy audio codec based on the study of psychoacoustics. Audio signals that are deemed to be imperceptible to the human ear are encoded with reduced resolution during the compression process.

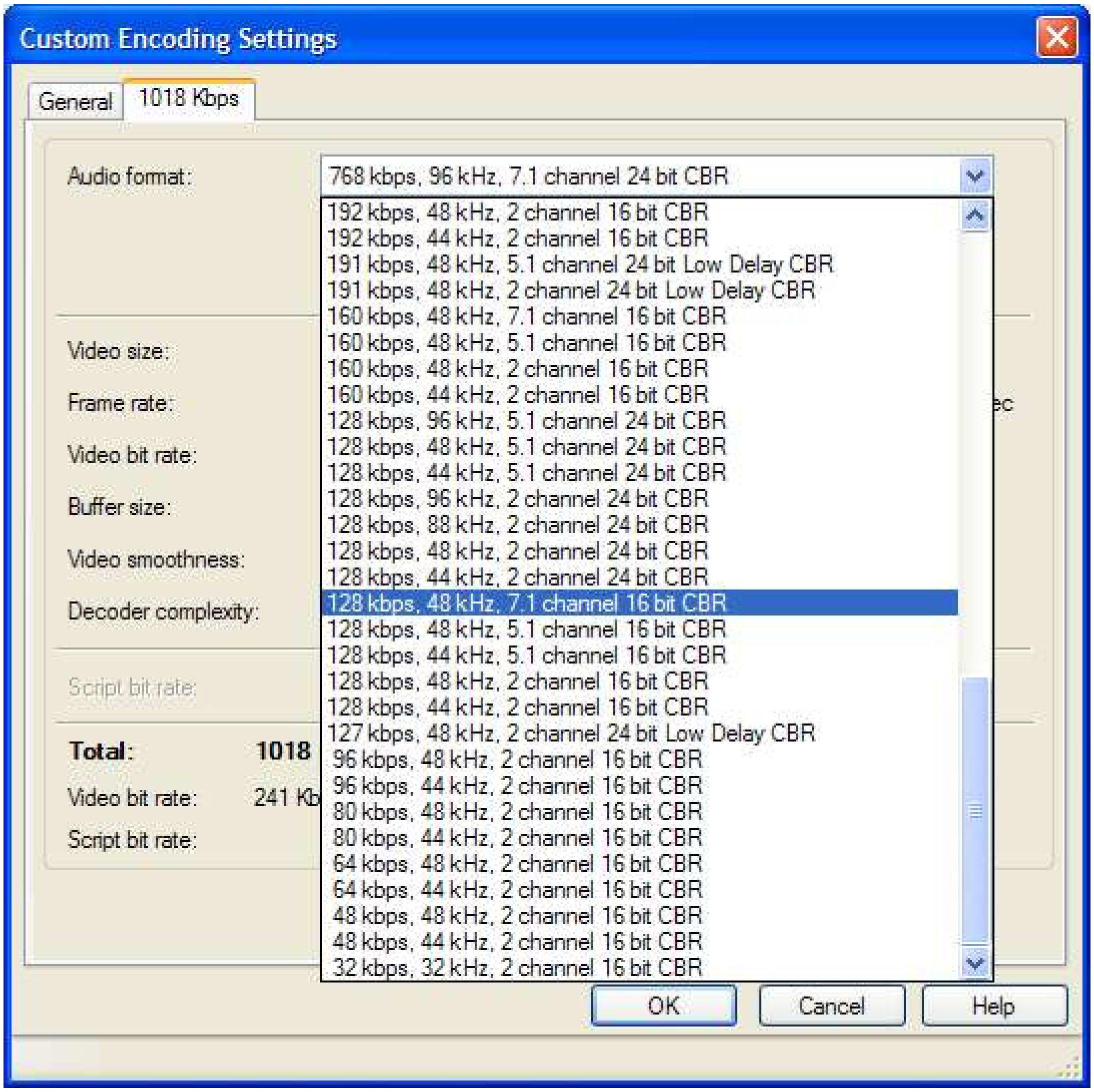
WMA can encode audio signals sampled at up to 48 kHz with up to two discrete channels (stereo). WMA 9 introduced variable bit rate (VBR) and average bit rate (ABR) coding techniques into the MS encoder although both were technically supported by the original format. WMA 9.1 also added support for low-delay audio, which reduces latency for encoding and decoding.

Fundamentally, WMA is a transform coder based on modified discrete cosine transform (MDCT), somewhat similar to AAC, Cook and Vorbis. The bit stream of WMA is composed of superframes, each containing 1 or more frames of 2048 samples. If the bit reservoir is not used, a frame is equal to a superframe. Each frame contains a number of blocks, which are 128, 256, 512, 1024, or 2048 samples long after being transformed into the frequency domain via the MDCT. In the frequency domain, masking for the transformed samples is determined, and then used to requantize the samples. Finally, the floating point samples are decomposed into coefficient and exponent parts and independently huffman coded. Stereo information is typically mid/side coded. At low bit rates, line spectral pairs (typically less than 17 kbit/s) and a form of noise coding (typically less than 33 kbit/s) can also be used to improve quality.

Like AAC and Ogg Vorbis, WMA was intended to address perceived deficiencies in the MP3 standard. Given their common design goals, it's not surprising that the three formats ended up making similar design choices. All three are pure transform codecs. Furthermore the MDCT implementation used in WMA is essentially a superset of those used in Ogg and AAC such that WMA iMDCT and windowing routines can be used to decode AAC and Ogg Vorbis almost unmodified. However, quantization and stereo coding is handled differently in each codec. The primary distinguishing trait of the WMA Standard format is its unique use of 5 different block sizes, compared to MP3, AAC, and Ogg Vorbis which each restrict files to just two sizes. WMA Pro extends this by adding a 6th block size used at 88.1/96 kHz sampling rate.

Certified PlaysForSure devices, as well as a large number of uncertified devices, ranging from portable hand-held music players to set-top DVD players, support the playback of WMA files. Most PlaysForSure-certified online stores distribute content using this codec only. In 2005, Nokia announced its plans to support WMA playback in future Nokia handsets. In the same year, an update was made available for the PlayStation Portable (version 2.60) which allowed WMA files to be played on the device for the first time.

Windows Media Audio Professional



Screenshot of Windows Media Encoder 9 Series, displaying new encoding options for Windows Media Audio 10 Professional.

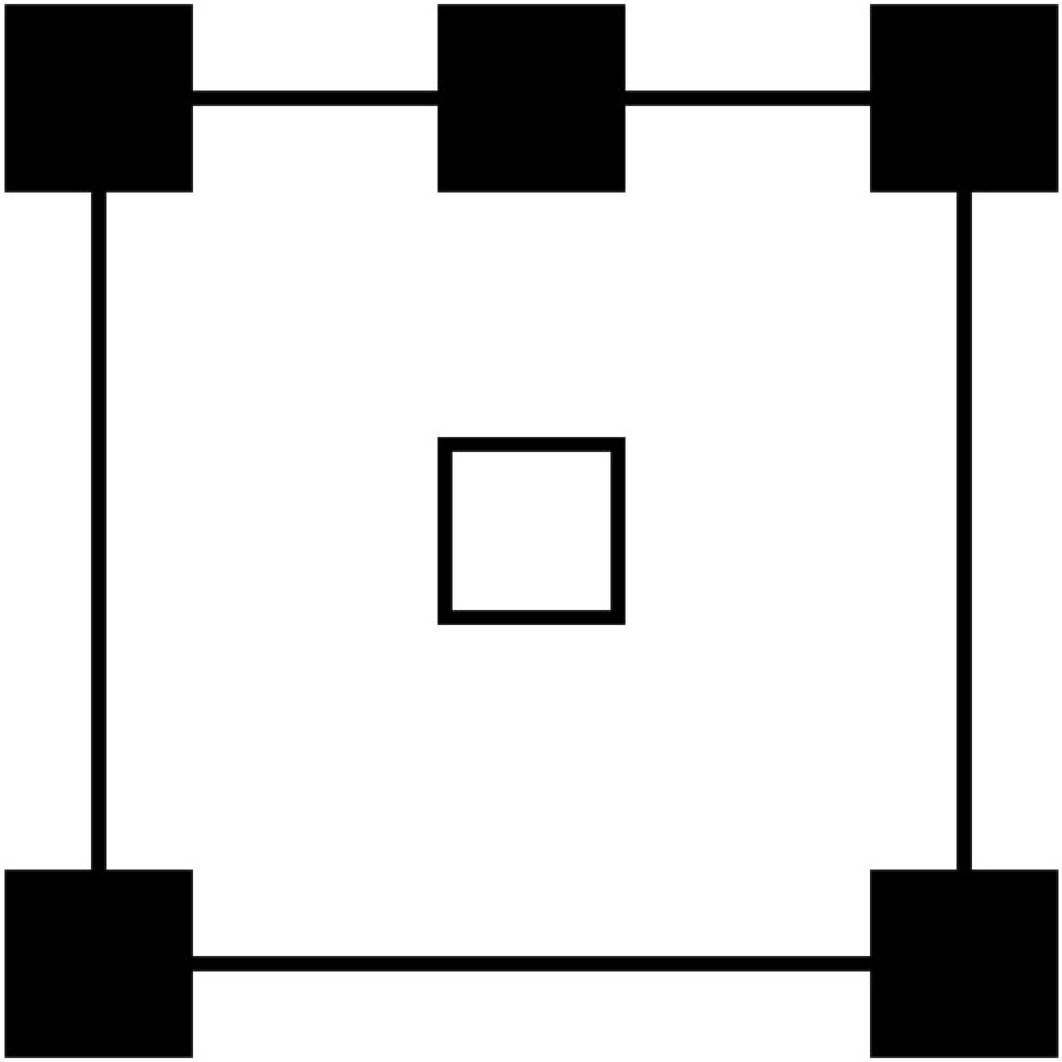
Windows Media Audio Professional (WMA Pro) is an improved lossy codec closely related to WMA standard. It retains most of the same general coding features, but also features improved entropy coding and quantization strategies as well as more efficient stereo coding. Notably, many of the WMA standard's low bitrate features have been removed, as the core codec is designed for efficient coding at most bitrates. Its main competitors include AAC, HE-AAC, Vorbis, Dolby Digital, and DTS. It can support audio resolutions of up to 96 kHz and up to eight discrete channels (7.1 channel surround). WMA Pro also supports dynamic range compression, which reduces the volume difference between the loudest and quietest sounds in the audio track. According to Microsoft's Amir Majidimehr, WMA Pro can technically go beyond 7.1 surround sound and support "an unlimited number of channels."

The codec's bit stream syntax was frozen at the first version, WMA 9 Pro. Later versions of WMA Pro introduced low-bit rate encoding, low-delay audio, frequency interpolation mode, and an expanded range of sampling rate and bit-depth encoding options. A WMA 10 Pro file compressed with frequency interpolation mode comprises a WMA 9 Pro track encoded at half the original sampling rate, which is then restored using a new compression algorithm. In this situation, WMA 9 Pro players which have not been updated to the WMA 10 Pro codec can only decode the lower quality WMA 9 Pro stream. Starting with WMA 10 Pro, eight channel encoding starts at 128 kbit/s, and tracks can be encoded at the native audio CD resolution (44.1 kHz, 16-bit), previously the domain of WMA Standard.

Despite a growing number of supported devices and its superiority over WMA, WMA Pro still has little hardware and software support. Some notable exceptions to this are the Microsoft Zune (limited to stereo), Xbox 360, Windows Mobile-powered devices with Windows Media Player 10 Mobile, newer Toshiba Gigabeat and Motorola devices, and devices running recent versions of the Rockbox alternative firmware. In addition, WMA Pro is a requirement for the WMV HD certification program. On the software side, Verizon utilizes WMA 10 Pro for its V CAST Music Service, and Windows Media Player 11 has promoted the codec as an alternative to WMA for copying audio CD tracks. WMA Pro is supported in Silverlight as of version 2 (though only in stereo mode). In the absence of the appropriate audio hardware, WMA Pro can automatically downmix multichannel audio to stereo or mono, and 24-bit resolution to 16-bit during playback.

A notable example of WMA Pro being used instead of WMA Standard is the NBC Olympics website which uses WMA 10 Pro in its low-bitrate mode at 48 kbit/s.

Windows Media Audio Lossless



Label for 5.1 surround sound, the maximum channel configuration for Windows Media Audio Lossless.

Windows Media Audio Lossless (WMA Lossless) is a lossless audio codec that competes with ATRAC Advanced Lossless, Dolby TrueHD, DTS-HD Master Audio, Apple Lossless, Shorten, Monkey's Audio, FLAC, and WavPack (the last two have the advantage of being open source software and available for nearly any operating system). Designed for archival purposes, it compresses audio signals without loss of quality from the original using VBR. When decompressed, the audio signal is an exact replica of the original. The first version of the codec, WMA 9 Lossless, and its revisions support up to 96 kHz, 24-bit audio for up to 6 discrete channels (5.1 channel surround) with dynamic range compression control. The typical compression ratio for music varies between 1.7:1 and 3:1.

Hardware support for the codec is available on the Cowon A3 , Cowon S9, Bang & Olufsen Serenata, Sony Walkman NWZ-A and NWZ-S series, Zune 4, 8, 80 30, Zune 120 (with firmware version 2.2 or later) and the new Zune HD, Xbox 360, Windows Mobile-powered devices with Windows Media Player 10 Mobile, Toshiba Gigabeat S and V models, Toshiba T-400, the Meizu M3, and Best Buy's Insignia NS-DV, Pilot, and Sport music players. Contrary to some claims, the Archos make of media devices do not support WMA Lossless, nor does the SONOS system. Like WMA Standard, WMA Lossless is being used by a few online stores to distribute music online. Similar to WMA Pro, the WMA Lossless decoder can perform downmixing when capable audio hardware is not present.

Windows Media Audio Voice

Windows Media Audio Voice (WMA Voice) is a lossy audio codec that competes with Speex (used in Microsoft's own Xbox Live online service), ACELP, and other codecs. Designed for low-bandwidth, voice playback applications, it employs low-pass and high-pass filtering of sound outside the human speech frequency range to achieve higher compression efficiency than WMA. It can automatically detect sections of an audio track containing both voice and music and use the standard WMA compression algorithm instead. WMA Voice supports up to 22.05 kHz for a single channel (mono) only. Encoding is limited to constant bit rate (CBR) and up to 20 kbit/s. The first and only version of the codec is WMA 9 Voice.

Windows Mobile-powered devices with Windows Media Player 10 Mobile have native support for WMA 9 Voice playback. In addition, BBC World Service has employed WMA Voice for its Internet radio streaming service.

Sound quality

Microsoft claims that audio encoded with WMA sounds better than MP3 at the same bit rate; Microsoft also claims that audio encoded with WMA at lower bit rates sound better than MP3 at higher bit rates. Double blind listening tests with other lossy audio codecs have shown varying results, from failure to support Microsoft's claims about its superior quality to supremacy over other codecs. One independent test conducted in May 2004 at 128 kbit/s showed that WMA was roughly equivalent to LAME MP3; inferior to AAC and Vorbis; and superior to ATRAC3 (software version).

Some conclusions made by recent studies:

- At 32 kbit/s, WMA Standard was noticeably better than LAME MP3, but not better than other modern codecs in a collective, independent test in July 2004.
- At 48 kbit/s, WMA 10 Pro was ranked second after Nero HE-AAC and better than WMA 9.2 in an independent listening test organized and supported by

Sebastian Mares and Hydrogenaudio Forums in December 2006. This test, however, used CBR for WMA 10 Pro and VBR for the other codecs.
- At 64 kbit/s, WMA Pro outperformed Nero HE-AAC in a commissioned, independent listening test performed by the National Software Testing Labs in 2005. Out of 300 participants, "71% of all listeners indicated that WMA Pro was equal to or better than HE AAC."
- At 80 kbit/s and 96 kbit/s, WMA had lower quality than HE-AAC, AAC-LC, and Vorbis; near-equivalent quality to MP3, and better quality than MPC in individual tests done in 2005.
- At 128 kbit/s, there was a four-way tie between aoTuV Vorbis, LAME MP3, WMA 9 Pro and AAC in a large scale test in January 2006, with each codec sounding close to the uncompressed music file for most listeners.
- At 768 kbit/s, WMA 9 Pro delivered full-spectrum response at half the bit rate required for DTS in a comparative test done by EDN in October 2003. The test sample was a 48 kHz, 5.1 channel surround audio track.

Criticism of claimed quality

Microsoft's claims of WMA sound quality have frequently drawn complaints. "Some audiophiles challenge Microsoft's claims regarding WMA's quality," according to a published article from EDN. Another article from MP3 Developments wrote that Microsoft's claim about CD-quality audio at 64 kbit/s with WMA was "very far from the truth." At the early stages of WMA's development, a representative from RealNetworks claimed that WMA was a "clear and futile effort by Microsoft to catch up with RealAudio 8."

Microsoft has sometimes claimed that the sound quality of WMA at 64 kbit/s equals or exceeds that of MP3 at 128 kbit/s (both WMA and MP3 are considered near-transparent at 192 kbit/s by most listeners). In a 1999 study funded by Microsoft, National Software Testing Laboratories (NSTL) found that listeners preferred WMA at 64 kbit/s to MP3 at 128 kbit/s (as encoded by MusicMatch Jukebox). However, a September 2003 public listening test conducted by Roberto Amorim found that listeners preferred 128 kbit/s MP3 to 64 kbit/s WMA audio with greater than 99% confidence. This conclusion applied equally to other codecs at the same bitrate, leading him to conclude that:

> “ No codec delivers the marketing plot of same quality as MP3 at half the bitrates. ”

It is important to note that both MP3 and WMA encoders have undergone active development and improvement for many years, so their relative quality may change over time.

A July 2007 public listening test by Sebastian Mares found that 64 kbit/s HE-AAC audio (encoded by Nero Digital) was statistically tied with 64 kbit/s WMA Pro audio, in terms of listener preference.

Players

Apart from Windows Media Player, most of the WMA compression formats can be played using ALLPlayer, VLC media player, MPlayer, RealPlayer, Winamp, Zune Software (with certain limitations—DSP plugin support and DirectSound output is disabled using the default WMA plugin), and many other software media players. The Microsoft Zune media management software supports most WMA codecs, but uses a variation of Windows Media DRM which is used by PlaysForSure.

The FFmpeg project has reverse-engineered and re-implemented the WMA codecs (except WMA Lossless) to allow their use on POSIX-compliant operating systems such as Linux. The rockbox project further extended this codec to be suitable for embedded cores, allowing playback on portable MP3 players and cell phones running open source software. RealNetworks has announced plans to support playback of DRM-unprotected WMA files in RealPlayer for Linux. On the Macintosh platform, Microsoft released a PowerPC version of Windows Media Player for Mac OS X in 2003, but further development of the software has ceased. Microsoft currently endorses the third-party Flip4Mac WMA, a QuickTime component that allows Macintosh users to play WMA files in any player that uses the QuickTime framework. Flip4Mac, however, does not currently support the Windows Media Audio Voice codec.

Encoders

Software that can export audio in WMA format include Windows Media Player, Windows Movie Maker, Microsoft Expression Encoder, Sony Sound Forge, GOM Player, RealPlayer, Adobe Premiere Pro, Adobe Audition, and Adobe Soundbooth. Microsoft Office OneNote supports encoding in all WMA codecs, and Windows Media Encoder supports all available bit rate and resolution options as well. Open source players like VLC media player can also do some encoding.

Digital rights management

The WMA codecs are most often used with the ASF container format, which has an optional DRM facility. Windows Media DRM, which can be used in conjunction with WMA, supports time-limited music subscription services such as those offered by unlimited download services, including MTV's URGE, Napster, Rhapsody, Yahoo! Music Unlimited, and Virgin Digital. Windows Media DRM, a component of PlaysForSure and Windows Media Connect, is supported on many modern portable audio devices and streaming media clients such as Roku, SoundBridge, Xbox 360, and Wii. Players that support the WMA format but not Windows Media DRM list protected titles as unplayable.

Chapter 8

dbx Model 700 Digital Audio Processor and Quantization (Sound Processing)

dbx Model 700 Digital Audio Processor

The **dbx Model 700 Digital Audio Processor** was a professional audio ADC/DAC combination unit, which digitized a stereo analog audio input into a bitstream, which was then encoded and encapsulated in an analog composite video signal, for recording to tape using a VCR as a transport. Unlike other similar pieces of equipment like the Sony PCM-F1, the Model 700 used a technique called *Companded Predictive Delta Modulation,* rather than the now-common pulse-code modulation. At the time of its introduction in the mid-1980s the device was the first commercial product to use this method, although it had been proposed in the 1960s and prototyped in the late '70s.

History

Unlike the many digital recording formats that would follow (e.g. DAT and ADAT), the Model 700 had no capability for storage on its own, and relied on an analog recording medium supplied by the user. In general, any high-quality VHS VCR would do, although 3/4" U-matic or Beta decks could also have been used. If viewed on a monitor, the output stream of a Model 700 looked like analog TV "static" or noise, with slight black bars running down either side.

Early on, the machine was hailed as "the best recording device you can buy," and Stereophile Magazine reviewed it positively. Many people liked the format because it offered more dynamic range than analog tape, but without the "hard clipping" inherent in PCM audio recorders of the time. The Model 700 had been designed from the beginning to have many 'tape-like' characteristics, including "soft saturation," and at a time when most professional and amateur recordists were used to analog tape, this was considered a significant feature. It also offered 14dB more dynamic range than 44.1kHz/16b audio, and because of its very high sample rate (644kHz), it did not contain the same anti-aliasing filters necessary in PCM recorders at the time, which were thought to cause undesirable harmonic interference.

The device sold for $4,600 in 1986, and that was without a video recorder on which to store the output, putting it out of the reach of all but the most wealthy home users. However, its target market was professional and studio users, and here it enjoyed relative popularity for a short amount of time as a mastering or mixdown recorder, recording the final output from a multitrack system.

The Model 700 was available in several different versions. In its most basic incarnation, it had two line-level, balanced inputs. One popular upgrade was the addition of one or two microphone preamps, which were installed on removable cards into slots in the machine. These allowed stereo recording directly into the Model 700, bypassing a mixing console. Since the recorder had a significantly lower noise floor than most mixers of the same era, this method made the best use of the system's available dynamic range. Another, much more rare accessory was the *Model 700D Disc Mastering Delay*. This was a device used for mastering vinyl records, and which attached to a proprietary 25-pin digital output on the back of the Model 700 recorder. Because of the nature of vinyl records (which rotate at a constant angular velocity but at a changing linear velocity with respect to the needle as it moves inwards), it is necessary to send the audio signal to the computer controlling the movement (pitch) of the cutting lathe. In this way, during quiet passages the pitch is increased thus enabaling more grooves per inch. The computer needs audio signal that is exactly one rotation ahead of the actual audio fed to the cutting lathe, so that it can move the cutting head rapidly forward before a loud passage is grooved. The disc mastering delay achieved this just like the analog mastering tape decks used an extra playback head.

The Model 700 was developed and sold by the dbx corporation of Newton, Massachusetts, better known for their system of noise reduction for analog tape.

Technical specifications

- Dynamic Range: 110dB typical with A-weighted noise 20Hz-20kHz; >105dB unweighted
- Frequency Response: 20Hz-20kHz, sine or pink noise, 100mV, reference record position
- THD: less than 0.05%, 1V input, 1 kHz
- Wow and flutter: less than 0.01% unweighted; 0.006% wrms
- Anti-aliasing filters: -3dB at 37 kHz
- Sampling Rate: 644 kHz
- Bit Rate: 644 kbit/s
- Mic Pre: adds less than 1dB noise, 100 to 1k-ohm impedance
- Max In/Out Levels: +24dBm

Theory of operation

The Model 700 converted analog audio into digital data using a type of delta-sigma modulation, called "Companded Predictive Delta Modulation," or "CPDM" (both trademarked). In a traditional, single-integrated delta-sigma ADC, the voltage of an input signal is compared to the output of an integrator. If the input signal is higher than the integrator's output, a 1 is recorded, and the integrator is given a command to increase by a certain amount. On each clock cycle, the comparison is repeated (and another 1 is recorded) until the integrator's output exceeds the input voltage, at which point a 0 is recorded and the integrator is told to decrease. In this way, the integrator attempts to follow the input signal as closely as it can. When fed a constant-voltage signal (or when the input is removed completely), the output will "idle" and produce a stream of alternating 1s and 0s. A decoder listening to this stream would produce a small sinusoidal or triangle-wave output, even though the correct output should be flat: this is a form of quantization error.

Where the Model 700 differs from classical delta-sigma modulation is in its replacement of the single integrator with a complex system of comparators and high-order linear prediction filters. This was done in order to reduce the quantization error, and is accomplished in part by changing the effective "step size" of the encoder based on previously recorded information. (Thereby increasing or decreasing the slew rate.) The system also has two analog pre-processing steps which compress the input signal in both the amplitude and frequency domain, in order to more closely match the abilities of the encoder. This compression is done adaptively based on previously encoded signal, and is reversed on the decoding end.

Quantization (Sound Processing)

In signal processing and digital audio, quantization is the process of approximating a continuous range of values (or a very large set of possible discrete values) by a relatively-small set of discrete symbols or integer values.

After sampling, sound signals are usually represented by one of a fixed number of values, in a process known as pulse-code modulation (PCM). Some specific issues related to quantization of audio signals follow.

Audio quantization

Telephony applications frequently use 8-bit quantization. That is, values of the analogue waveform are rounded to the closest of 256 distinct voltage values represented by an 8-bit binary number. This crude quantization introduces substantial quantization noise into the signal, but the result is still more than adequate to represent human speech.

By comparison, compact discs use a 16-bit digital representation, allowing 65,536 distinct voltage levels. This is far better than telephone quantization, but CD audio representing low signal levels would still sound noticeably 'granular' because of the quantizing noise. However, sometimes an addition of a small amount of noise is added to the signal before digitization. This deliberately-added noise is known as dither. Adding dither eliminates this granularity, and gives very low distortion, but at the expense of a small increase in noise level. Measured using ITU-R 468 noise weighting, this is about

66dB below alignment level, or 84dB below FS (full scale) digital, which is somewhat lower than the microphone noise level on most recordings, and hence of no consequence.

Optimizing dither waveforms

In a seminar paper published in the AES Journal, Lipshitz and Vanderkooy pointed out that different noise types, with different probability density functions (PDFs) behave differently when used as dither signals, and suggested optimal levels of dither signal for audio. Gaussian noise requires a higher level for full elimination of distortion than rectangular PDF or triangular PDF noise. Triangular PDF noise has the advantage of requiring a lower level of added noise to eliminate distortion and also minimizing 'noise modulation'. The latter refers to audible changes in the residual noise on low-level music that are found to draw attention to the noise.

Noise shaping for lower audibility

An alternative to dither is noise shaping, which involves a feedback process in which the final digitized signal is compared with the original, and the instantaneous errors on successive past samples integrated and used to determine whether the next sample is rounded up or down. This smooths out the errors in a way that alters the spectral noise content. By inserting a weighting filter in the feedback path, the spectral content of the noise can be shifted to areas of the 'equal-loudness contours' where the human ear is least sensitive, producing a lower subjective noise level (-68/-70dB typically ITU-R 468 weighted).

24-bit quantization

24-bit audio is sometimes used undithered, because for most audio equipment and situations the noise level of the digital converter can be louder than the required level of any dither that might be applied.

There is some disagreement over the recent trend towards higher bit-depth audio. It is argued by some that the dynamic range presented by 16-bit is sufficient to store the dynamic range present in almost all music. In terms of pure data storage this is often true, as a high-end system can extract an extremely good sound out of the 16-bits stored in a well-mastered CD. However, audio with very loud and very quiet sections can require some of the above dithering techniques to fit it into 16-bits. This is not a problem for most recently produced popular music, which is often mastered so that it constantly sits close to the maximum signal; however, higher resolution audio formats are already being used (especially for applications such as film soundtracks, where there is often a very wide dynamic range between whispered conversations and explosions).

For most situations the advantage given by resolution higher than 16-bit is mainly in the processing of audio. No digital filter is perfect, but if the audio is upsampled and the audio is done in 24-bit or higher, then the distortion introduced by filtering will be much quieter (as the errors always creep into the least significant bits) and a well-designed filter can weight the distortion more towards the higher inaudible frequencies (but a sample rate higher than 48kHz is needed so that these inaudible ultrasonic frequencies are available for soaking up errors).

There is also a good case for 24-bit (or higher) recording in the live studio, because it enables greater headroom (often 24dB or more rather than 18dB) to be left on the recording without encountering quantization errors at low volumes. This means that brief peaks are not harshly clipped, but can be compressed or soft-limited later to suit the final medium.

Environments where large amounts of signal processing are required (such as mastering or synthesis) can require even more than 24 bits. Some modern audio editors convert incoming audio to 32-bit (both for an increased dynamic range to reduce clipping, and to minimize noise in intermediate stages of filtering), and some DAW environments (such as recent versions of REAPER and SONAR) use 64-bit audio for their underlying engine.

Chapter 9

Audio Compression (Data)

Audio compression is a form of data compression designed to reduce the transmission bandwidth requirement of digital audio streams and the storage size of audio files. Audio compression algorithms are implemented in computer software as audio codecs. Generic data compression algorithms perform poorly with audio data, seldom reducing data size much below 87% from the original, and are not designed for use in real time applications. Consequently, specifically optimized audio lossless and lossy algorithms have been created. Lossy algorithms provide greater compression rates and are used in mainstream consumer audio devices.

In both lossy and lossless compression, information redundancy is reduced, using methods such as coding, pattern recognition and linear prediction to reduce the amount of information used to represent the uncompressed data.

The trade-off between slightly reduced audio quality and transmission or storage size is outweighed by the latter for most practical audio applications in which users may not perceive the loss in playback rendition quality. For example, one Compact Disc holds approximately one hour of uncompressed high fidelity music, less than 2 hours of music compressed losslessly, or 7 hours of music compressed in the MP3 format at medium bit rates.

Lossless audio compression

Lossless audio compression produces a representation of digital data that can be expanded to an exact digital duplicate of the original audio stream. This is in contrast to the irreversible changes upon playback from lossy compression techniques such as Vorbis and MP3. Compression ratios are similar to those for generic lossless data compression (around 50–60% of original size), and substantially less than for lossy compression, which typically yield 5–20% of original size.

Applications

The primary application areas of lossless encoding are:

Archives
: For archival purposes it is generally desired to preserve the source material exactly (i.e. at 'best possible quality').

Editing
: Audio engineers use lossless compression for audio editing to avoid digital generation loss.

High fidelity playback
: Audiophiles prefer lossless compression formats to avoid compression artifacts.

Mastering of casual-use audio media
: High quality master copies of recordings are used to produce lossily compressed versions for digital audio players. As formats and encoders improve, updated lossily compressed files may be generated from the lossless master.

As file storage and communications bandwidth have become less expensive and more available, lossless audio compression has become more popular.

Formats

Shorten was an early lossless format; newer ones include Free Lossless Audio Codec (FLAC), Apple's Apple Lossless, MPEG-4 ALS, Windows Media Audio 9 Lossless (WMA Lossless), Monkey's Audio, and TTA.

Some audio formats feature a combination of a lossy format and a lossless correction; this allows stripping the correction to easily obtain a lossy file. Such formats include MPEG-4 SLS (Scalable to Lossless), WavPack, and OptimFROG DualStream.

Some formats are associated with a technology, such as:

- Direct Stream Transfer, used in Super Audio CD
- Meridian Lossless Packing, used in DVD-Audio, Dolby TrueHD, Blu-ray and HD DVD

Difficulties in lossless compression of audio data

It is difficult to maintain all the data in an audio stream and achieve substantial compression. First, the vast majority of sound recordings are highly complex, recorded from the real world. As one of the key methods of compression is to find patterns and repetition, more chaotic data such as audio doesn't compress well. In a similar manner, photographs compress less efficiently with lossless methods than simpler computer-generated images do. But interestingly, even computer generated sounds can contain very complicated waveforms that present a challenge to many compression algorithms. This is due to the nature of audio waveforms, which are generally difficult to simplify without a (necessarily lossy) conversion to frequency information, as performed by the human ear.

The second reason is that values of audio samples change very quickly, so generic data compression algorithms don't work well for audio, and strings of consecutive bytes don't generally appear very often. However, convolution with the filter [-1 1] (that is, taking the first derivative) tends to slightly whiten (decorrelate, make flat) the spectrum, thereby allowing traditional lossless compression at the encoder to do its job; integration at the decoder restores the original signal. Codecs such as FLAC, Shorten and TTA use linear prediction to estimate the spectrum of the signal. At the encoder, the estimator's inverse is used to whiten the signal by removing spectral peaks while the estimator is used to reconstruct the original signal at the decoder.

Evaluation criteria

Lossless audio codecs have no quality issues, so the usability can be estimated by

- Speed of compression and decompression
- Degree of compression
- Robustness and error correction
- Product support

Lossy audio compression

Lossy audio compression is used in a wide range of applications. In addition to the direct applications (mp3 players or computers), digitally compressed audio streams are used in most video DVDs; digital television; streaming media on the internet; satellite and cable radio; and increasingly in terrestrial radio broadcasts. Lossy compression typically achieves far greater compression than lossless compression (data of 5 percent to 20 percent of the original stream, rather than 50 percent to 60 percent), by discarding less-critical data.

The innovation of lossy audio compression was to use psychoacoustics to recognize that not all data in an audio stream can be perceived by the human auditory system. Most lossy compression reduces perceptual redundancy by first identifying sounds which are considered perceptually irrelevant, that is, sounds that are very hard to hear. Typical examples include high frequencies, or sounds that occur at the same time as louder sounds. Those sounds are coded with decreased accuracy or not coded at all.

If reducing perceptual redundancy does not achieve sufficient compression for a particular application, it may require further lossy compression. Depending on the audio source, this still may not produce perceptible differences. Speech for example can be compressed far more than music. Most lossy compression schemes allow compression parameters to be adjusted to achieve a target rate of data, usually expressed as a bit rate. Again, the data reduction will be guided by some model of how important the sound is as perceived by the human ear, with the goal of efficiency and optimized quality for the target data rate. (There are many different models used for this perceptual analysis, some better suited to different types of audio than others.) Hence, depending on the bandwidth and storage requirements, the use of lossy compression may result in a perceived reduction of the audio quality that ranges from none to severe, but generally an obviously audible reduction in quality is unacceptable to listeners.

Because data is removed during lossy compression and cannot be recovered by decompression, some people may not prefer lossy compression for archival storage. Hence, as noted, even those who use lossy compression (for portable audio applications, for example) may wish to keep a losslessly compressed archive for other applications. In addition, the technology of compression continues to advance, and achieving a state-of-the-art lossy compression would require one to begin again with the lossless, original audio data and compress with the new lossy codec. The nature of lossy compression (for both audio and images) results in increasing degradation of quality if data are decompressed, then recompressed using lossy compression.

Coding methods

Transform domain methods

In order to determine what information in an audio signal is perceptually irrelevant, most lossy compression algorithms use transforms such as the modified discrete cosine transform (MDCT) to convert time domain sampled waveforms into a transform domain. Once transformed, typically into the frequency domain, component frequencies can be allocated bits according to how audible they are. Audibility of spectral components is determined by first calculating a masking threshold, below which it is estimated that sounds will be beyond the limits of human perception.

The masking threshold is calculated using the absolute threshold of hearing and the principles of simultaneous masking - the phenomenon wherein a signal is masked by another signal separated by frequency - and, in some cases, temporal masking - where a signal is masked by another signal separated by time. Equal-loudness contours may also be used to weight the perceptual importance of different components. Models of the human ear-brain combination incorporating such effects are often called psychoacoustic models.

Time domain methods

Other types of lossy compressors, such as the linear predictive coding (LPC) used with speech, are *source-based coders*. These coders use a model of the sound's generator (such as the human vocal tract with LPC) to whiten the audio signal (i.e., flatten its spectrum) prior to quantization. LPC may also be thought of as a basic perceptual coding technique; reconstruction of an audio signal using a linear predictor shapes the coder's quantization noise into the spectrum of the target signal, partially masking it.

Applications

Due to the nature of lossy algorithms, audio quality suffers when a file is decompressed and recompressed (digital generation loss). This makes lossy compression unsuitable for storing the intermediate results in professional audio engineering applications, such as sound editing and multitrack recording. However, they are very popular with end users (particularly MP3), as a megabyte can store about a minute's worth of music at adequate quality.

Usability

Usability of lossy audio codecs is determined by:

- Perceived audio quality
- Compression factor
- Speed of compression and decompression
- Inherent latency of algorithm
- Product support

Lossy formats are often used for the distribution of streaming audio, or interactive applications (such as the coding of speech for digital transmission in cell phone networks). In such applications, the data must be decompressed as the data flows, rather than after the entire data stream has been transmitted. Not all audio codecs can be used for streaming applications, and for such applications a codec designed to stream data effectively will usually be chosen.

Latency results from the methods used to encode and decode the data. Some codecs will analyze a longer segment of the data to optimize efficiency, and then code it in a manner that requires a larger segment of data at one time in order to decode. (Often codecs create segments called a "frame" to create discrete data segments for encoding and decoding.) The inherent latency of the coding algorithm can be critical; for example, when there is two-way transmission of data, such as with a telephone conversation, significant delays may seriously degrade the perceived quality.

In contrast to the speed of compression, which is proportional to the number of operations required by the algorithm, here latency refers to the number of samples which must be analysed before a block of audio is processed. In the minimum case, latency is 0 zero samples (e.g., if the coder/decoder simply reduces the number of bits used to quantize the signal). Time domain algorithms such as LPC also often have low latencies, hence their popularity in speech coding for telephony. In algorithms such as MP3, however, a large number of samples have to be analyzed in order to implement a psychoacoustic model in the frequency domain, and latency is on the order of 23 ms (46 ms for two-way communication).

Speech encoding

Speech encoding is an important category of audio data compression. The perceptual models used to estimate what a human ear can hear are generally somewhat different from those used for music. The range of frequencies needed to convey the sounds of a human voice are normally far narrower than that needed for music, and the sound is normally less complex. As a result, speech can be encoded at high quality using relatively low bit rates.

This is accomplished, in general, by some combination of two approaches:

- Only encoding sounds that could be made by a single human voice.
- Throwing away more of the data in the signal—keeping just enough to reconstruct an "intelligible" voice rather than the full frequency range of human hearing.

Perhaps the earliest algorithms used in speech encoding (and audio data compression in general) were the A-law algorithm and the µ-law algorithm.

History

Solidyne 922: The world's first commercial audio bit compression card for PC, 1990

A literature compendium for a large variety of audio coding systems was published in the IEEE Journal on Selected Areas in Communications (JSAC), February 1988. While there were some papers from before that time, this collection documented an entire variety of finished, working audio coders, nearly all of them using perceptual (i.e. masking) techniques and some kind of frequency analysis and back-end noiseless coding. Several of these papers remarked on the difficulty of obtaining good, clean digital audio for research purposes. Most, if not all, of the authors in the JSAC edition were also active in the MPEG-1 Audio committee.

The world's first commercial broadcast automation audio compression system was developed by Oscar Bonello, an Engineering professor at the University of Buenos Aires. In 1983, using the psychoacoustic principle of the masking of critical bands first published in 1967, he started developing a practical application based on the recently developed IBM PC computer, and the broadcast automation system was launched in 1987 under the name Audicom. 20 years later, almost all the radio stations in the world were using similar technology, manufactured by a number of companies.

Chapter 10

Audio bit depth and Audio Converter

Audio bit depth

In digital audio, **bit depth** describes the number of bits of information recorded for each sample. Bit depth directly corresponds to the resolution of each sample in a set of digital audio data. Common examples of bit depth include CD quality audio, which is recorded at 16 bits, and DVD-Audio, which can support up to 24-bit audio.

Digital audio

A set of digital audio samples contains data that, when converted into an analog signal, provides the necessary information to reproduce the sound wave. In pulse-code modulation (PCM) sampling, the bit depth will limit quantities such as dynamic range and signal-to-noise ratio. The bit depth will not limit frequency range, which is limited by the sample rate.

By increasing the sampling bit depth, smaller fluctuations of the audio signal can be resolved (also referred to as an increase in dynamic range). The 'rule-of-thumb' relationship between bit depth and dynamic range is, for each 1-bit increase in bit depth, the dynamic range will increase by 6 dB. 24-bit digital audio has a theoretical maximum dynamic range of 144 dB, compared to 96 dB for 16-bit; however, current digital audio converter technology is limited to dynamic ranges of about 120 dB (20-bit) because of 'real world' limitations in integrated circuit design.

Technically speaking, bit depth is only meaningful when applied to pure PCM devices. Non-PCM formats, such as DSD or lossy compression systems like MP3, have bit depths that are not defined in the same sense as PCM. This is particularly true for lossy audio compression, where bits are allocated to other types of information, and the bits actually allocated to individual samples are allowed to fluctuate within the constraints imposed by the allocation algorithm.

Dynamic range

The importance of bit depth in PCM audio is that it determines the maximum possible dynamic range of the signal, or the difference between the loudest possible sounds and the lowest possible noise. For a typical PCM recording, in which no noise shaping is employed and the frequency range extends most of the way to the Nyquist limit, the dynamic range in decibels is equal to 1.76 + 6.02 * bits. This formula is often simplified to 6 dB per bit, which yields the common value of 96 dB for 16-bit CD audio.

It should be restated that this is only valid for PCM sampling without post-processing. Systems such as DSD use a different modulation technique where the signal-to-noise ratio is not determined exclusively by the bit depth and the audio band does not extend to the Nyquist frequency.

What is a 'bit' of data?

In computing parlance, *bit* is the abbreviation for a single *binary digit*, represented by a 0 or a 1. A *word* is a binary number with more than one digit. Binary numerics are base-2; thus, each digit can only be a *0* or a *1*. In comparison, traditional decimal numerics are base-10, having digits that can only be *0* through *9*. For example, the 16-bit binary number *0110111110111010* is equivalent to the 5-digit decimal number 28602. The number of bits per word is simply how many digits there are in the corresponding number. The words in commonly used PCM digital audio formats are 8, 16 or 24 bits long. Larger words have higher resolution. The resolution of a 16-bit system can be calculated by using 2^{16} which gives a value of 65,536. A 24 bit system (2^{24}) has a resolution of 16,777,216.

Bit rate

Bit rate refers to the amount of data, specifically bits, transmitted or received per second.

One of the most common bit rates given is that for compressed audio files. For example, an MP3 file might be described as having a bit rate of 160 kbps or 160 kbit/s or 160000 bits/second. This indicates the amount of compressed data needed to store one second of music.

The standard audio CD is said to have a data rate of 44.1 kHz/16, meaning that the audio data was sampled 44,100 times per second, with a bit depth of 16. CD tracks are usually stereo, using a left and right track, so the amount of audio data per second is double that of mono, where only a single track is used. The bit rate is then 44100 samples/second x 16 bits/sample x 2 = 1,411,200 bit/s or 1.4 Mbit/s.

This explains why, for example, a Minidisc recorder, which uses ATRAC compression, can store files lasting twice as long on a disc, if the default, recording in 2 channel stereo, is set to single channel mono recording.

To fully define a sound file's digital audio bit rates: the format of the data, the sampling rate, word size (bit depth), and the number of channels (e.g. mono, stereo, four-track), must be known.

Calculating values

An audio file's bit rate can be calculated given sufficient information. Given any three of the following four values, the fourth can be calculated.

Bit rate = (sampling rate) x (bit depth) x (number of channels)

E.g., for a recording with a 44.1 kHz sampling rate, a 16 bit depth, and 2 channels (stereo):
44100 x 16 x 2 = 1411200 bits per second, or *1411.2 kbit/s*

The eventual file size of an audio recording can also be calculated using a similar formula:

File Size (Bytes) = (sampling rate) x (bit depth) x (number of channels) x (seconds) / 8

E.g., a 70 minutes long CD quality recording will take up 740880000 Bytes, or 740MB:

44100 x 16 x 2 x 4200 / 8 = 740880000 Bytes

Audio Converter

In signal processing, an **audio converter** or **digital audio converter** is a type of electronic hardware technology which converts an analog audio signal to a digital audio format, either on the input (Analog-to-digital converter or ADC), or the output (Digital-to-analog converter, or DAC). They are common in numerous technologies —notably in computer sound cards, digital cellular phones, portable recording devices, and digital audio workstations (DAW). Once converted to digital format, digital audio signals and file formats can be processed in any of a number of ways as allowed by software — including converting to audio CD or MP3 formats.

Different types of converter units can operate at different resolutions which largely determines the resulting sound "quality." Depending on their quality and cost, converters also differ in their handling of:

- electronic (radio) interference
- signal-to-noise ratio (SNR)
- electronic noise shielding and rejection
- digital noise floor handling,

- oversampling, (input and/or output)
- simultaneous playback and recording (full-duplex)
- clock jitter

"Resolution" generally refers to both bit and sample rates (though it may occasionally refer specifically to the sampling rate; being the more variable of the two.) For example, an inexpensive brand consumer sound card for computer has a typical operating range of around CD quality — 44,100 samples per second, (44.1 kHz), with 65,536 allowable values for volume (16 bits, giving 2^{16} values).

Digital audio basics

A single ("mono") channel (or "track") of digital audio can be visualized as a waveform, and is composed of a series of sampled sound pressures. Each sample has been quantized to one out of a set of discrete values depending on the bit depth. The playback of the waveform reproduces the sound pressure changes in sequence, producing complex sounds.

"Bit depth" determines the number of alternative values for sampled sound pressure that can be represented. The "sampling rate" (or *frequency*) refers to the number of samples (or "snapshots") per second, measured in hertz (Hz).

A stereo "track" is simply two isolated mono tracks within the same file, which are played back simultaneously. The illusion of spatial sound can be created through differences between the left and right channels. High-quality motion picture sound formats, like Dolby Surround and DTS simply use more mono channels to create more complex illusions of spatial depth.

Sound quality

Increasing the I/O capacity for either or both bit resolution and sampling rate will increase the quality of the sound. Multiple tracks can enhance the experience of spatial sound and sound "quality."

"Professional" sound quality refers to highest current sound qualities for available hardware. In the past, very high resolution hardware was also highly specialized and expensive. Today, many commonly available technologies are sufficient for generating adequate or above average quality digital audio signals for use in various recording or transmission applications.

Currently it's possible to purchase hardware using reasonably high quality digital converters for well under 1000 USD. In professional uses, extremely high fidelity converters are typically sold as components within specialized rack mount multi-channel (multitrack) recording units. These typically connect to a computer via parallel, proprietary or specialized (i.e. PCI card), fiber optic, USB, or Firewire connections. USB

2 and Firewire are the most common, though Firewire has lower latency and thus is more suited for multitrack applications.

At the barebones level, audio card converters take a nominal line-level input signal and is designed to be a transparent link between the mixing board and the computer. Once converted to digital format, the sound files can be processed in any of a number of ways as allowed by software, including converting to audio CD or MP3 formats.

Both the recording and playback of sound must consider the nature of the digital medium to produce undesirable noise. The noise produced by analog tape, by comparison can be pleasing, as it is of sufficient resolution to produce "warm" natural sounding random noise. Digital recording has traditionally had the problem of sounding "cold," mainly due to the quantization effect of digitizing a linear wave into a segmented series of adjacent samples. In the case of long wave sounds, such as bass, the wave is smooth, and therefore its playback does not "jump" from a low volume bit to an adjacent high volume bit. However, high pitched sounds like cymbal crashes and similar "white" noise can only be represented by a virtually random array of bits. This random array exposes the weakness of digital playback to produce digital artifacts in the form of a high-pitched "harshness."

While better resolutions produce better sound and less digital harshness, an important function of digital converters is to minimize the known problems endemic to the digital format. A common solution was to use separate vacuum tube amplifiers to "warm" the sound at both input output. This is generally expensive, and normalizes the sound to the range given by the tube amp, which negates the benefits of digital's deep low and clear high frequencies. The main solution has generally been to use an oversampling process, which can differ substantially depending on its application for input or output. Oversampling is the processing of sound input at higher sampling rates than the hardware actually performs. On input, this produces a digital wave which is still based on the input, but altered to record substantially fewer jumps which would produce audible artifacts. On the output, oversampling serves the same role as using a tube amp, whereby the adjacent bits are re-quantized at higher resolution to produce a smoother wave.

Given the capabilities of the digital realm to transfer exact files without analog loss, it is rare that multitrack digital to analog (DAC) converters are needed, with the exception of basic playback monitoring of stereo or specialized surround sound formats.

Chapter 11

Non-Linear Editing System

In video, a **non-linear editing system** (**NLE**) is a video editing (NLVE) or audio editing (NLAE) system which can perform random access on the source material. It is named in contrast to 20th century methods of linear tape and film editing.

Non-linear editing

Non-linear editing for films and television postproduction is a modern editing method which enables direct access to any frame in a digital video clip, without needing to play or scrub/shuttle through adjacent footage to reach it, as was necessary with historical videotape editing systems. It is the most natural approach when all assets are available as files on disks rather than recordings on reels or tapes, while linear editing is related to the need to sequentially view a film or read a tape to edit it. On the other hand, the NLE method is similar in concept to the "cut and paste" technique used in film editing. However, with the appropriation of non-linear editing systems, the destructive act of cutting of film negatives is eliminated. Non-linear, non-destructive methods began to appear with the introduction of digital video technology. It can also be viewed as the audio/video equivalent of word processing, which is why it is called desktop editing in the consumer space.

Video and audio data are first captured to hard disks or other digital storage devices. The data is either recorded directly to the storage device or is imported from another source. Once imported they can be edited on a computer using any of a wide range of software.

In non-linear editing, the original source files are not lost or modified during editing. Professional editing software records the decisions of the editor in an edit decision list (EDL) which can be interchanged with other editing tools. Many generations and variations of the original source files can exist without needing to store many different copies, allowing for very flexible editing. It also makes it easy to change cuts and undo previous decisions simply by editing the edit decision list (without having to have the actual film data duplicated). Generation loss also controlled, due to not having to repeatedly re-encode the data when different effects are applied.

Compared to the linear method of tape-to-tape editing, non-linear editing offers the flexibility of film editing, with random access and easy project organization. With the edit decision lists, the editor can work on low-resolution copies of the video. This makes it possible to edit both standard-definition broadcast quality and high definition broadcast quality very quickly on normal PCs which do not have the power to do the full processing of the huge full-quality high-resolution data in real-time.

The costs of editing systems have dropped such that non-linear editing tools are now within the reach of home users. Some editing software can now be accessed free as web applications; some, like Cinelerra (focused on the professional market) and Blender3D, can be downloaded free of charge; and some, like AVS Video Editor, Microsoft's Windows Movie Maker or Apple Inc.'s iMovie, come included with the appropriate operating system.

A computer for non-linear editing of video will usually have a video capture card to capture analog video and/or a FireWire connection to capture digital video from a DV camera, with its video editing software. Modern web based editing systems can take video directly from a camera phone over a GPRS or 3G mobile connection, and editing can take place through a web browser interface, so strictly speaking a computer for video editing does not require any installed hardware or software beyond a web browser and an internet connection.

Various editing tasks can then be performed on the imported video before it is exported to another medium, or MPEG encoded for transfer to a DVD.

History

The first truly non-linear editor, the CMX 600, was introduced in 1971 by CMX Systems, a joint venture between CBS and Memorex. It recorded & played back black-and-white analog video recorded in "skip-field" mode on modified disk pack drives the size of washing machines. These were commonly used to store data digitally on mainframe computers of the time. The 600 had a console with 2 monitors built in. The right monitor, which played the preview video, was used by the editor to make cuts and edit decisions using a light pen. The editor selected from options which were superimposed as text over the preview video. The left monitor was used to display the edited video. A Digital PDP-11 computer served as a controller for the whole system. Because the video edited on the 600 was in black and white and in low-resolution "skip-field" mode, the 600 was suitable only for offline editing.

Various approximations of non-linear editing systems were built in the '80s using computers coordinating multiple laser discs, or banks of VCRs. One example of these tape & disc-based systems was Lucasfilm's EditDroid, which used several laserdiscs of the same raw footage to simulate random-access editing (a compatible system was developed for sound post production by Lucasfilm called SoundDroid--one of the earliest digital audio workstations).

The term "nonlinear editing" or "non-linear editing" was formalized in 1991 with the publication of Michael Rubin's *Nonlinear: A Guide to Digital Film and Video Editing* (Triad, 1991) -- which popularized this terminology over other language common at the time, including "real time" editing, "random-access" or "RA" editing, "virtual" editing, "electronic film" editing, and so on. The handbook has remained in print since 1991, currently in its 4th edition (Triad, 2000).

Computer processing advanced sufficiently by the end of the '80s to enable true digital imagery, and has progressed today to provide this capability in personal desktop computers.

An example of computing power progressing to make non-linear editing possible was demonstrated in the first all-digital non-linear editing system to be released, the "Harry" effects compositing system manufactured by Quantel in 1985. Although it was more of a video effects system, it had some non-linear editing capabilities. Most importantly, it could record (and apply effects to) 80 seconds (due to hard disk space limitations) of broadcast-quality uncompressed digital video encoded in 8-bit CCIR 601 format on its built-in hard disk array.

Non-linear editing with computers as we know it today was first introduced by Editing Machines Corp. in 1989 with the EMC2 editor; a hard disk based non-linear off-line editing system, using half-screen resolution video at 15 frames per second. A couple of weeks later that same year, Avid introduced the Avid/1, the first in the line of their Media Composer systems. It was based on the Apple Macintosh computer platform (Macintosh II systems were used) with special hardware and software developed and installed by Avid. The Avid/1 was not the first system to introduce modern concepts in non-linear editing such as timeline editing and clip bins — both of these were pioneered in Lucasfilm's EditDroid in the early 1980s.

The video quality of the Avid/1 (and later Media Composer systems from the late 80s) was somewhat low (about VHS quality), due to the use of a very early version of a Motion JPEG (M-JPEG) codec. But it was enough to be a very versatile system for offline editing, to revolutionize video and film editing, and quickly become the dominant NLE platform.

The NewTek Video Toaster Flyer included non-linear editing capabilities in addition to processing live video signals. The Flyer made use of hard drives to store video clips and audio, and allowed complex scripted playback. The Flyer was capable of simultaneous dual-channel playback, which allowed the Toaster's Video switcher to perform transitions and other effects on Video clips without the need for rendering. The Flyer portion of the Video Toaster/Flyer combination was a complete computer of its own, having its own Microprocessor and Embedded software. Its hardware included three embedded SCSI controllers. Two of these SCSI buses were used to store video data, and the third to store audio. The Flyer used a proprietary Wavelet compression algorithm known as VTASC, which was well regarded at the time for offering better visual quality than comparable Motion JPEG based non-linear editing systems.

Until 1993, the Avid Media Composer could only be used for editing commercials or other small content projects, because the Apple Macintosh computers could access only 50 gigabytes of storage at one time. In 1992, this limitation was overcome by a group of industry experts led by Rick Eye a Digital Video R&D team at the Disney Channel. By February 1993, this team had integrated a long form system which gave the Avid Media Composer Apple Macintosh access to over 7 terabytes of digital video data. With instant access to the shot footage of an entire movie, long form non-linear editing (Motion Picture Editing) was now possible. The system made its debut at the NAB conference in 1993, in the booths of the three primary sub-system manufacturers, Avid, Silicon Graphics and Sony. Within a year, thousands of these systems replaced a century of 35mm film editing equipment in major motion picture studios and TV stations world wide, making Avid the undisputed leader in non-linear editing systems for over a decade.

Although M-JPEG became the standard codec for NLE during the early 1990s, it had drawbacks. Its high computational requirements ruled out software implementations, leading to the extra cost and complexity of hardware compression/playback cards. More importantly, the traditional tape workflow had involved editing from tape, often in a rented facility. When the editor left the edit suite they could take their confidential video tapes with them. But the M-JPEG data rate was too high for systems like Avid on the Mac and Lightworks on PC to store the video on removable storage, so these used fixed hard disks instead. The tape paradigm of keeping your (confidential) content with you was not possible with these fixed disks. Editing machines were often rented from facilities houses on a per-hour basis, and some productions chose to delete their material after each edit session, and then recapture it the next day, in order to guarantee the security of their content. In addition, each NLE system had storage limited by its hard disk capacity.

These issues were addressed by a small UK company, Eidos plc. Eidos chose the new ARM-based computers from the UK and implemented an editing system, launched in Europe in 1990 at the International Broadcasting Convention. Because it implemented its own compression software designed specifically for non-linear editing, the Eidos system had no requirement for JPEG hardware and was cheap to produce. The software could decode multiple video and audio streams at once for real-time effects at no extra cost. But most significantly, for the first time, it allowed effectively unlimited quantities of cheap removable storage. The Eidos Edit 1, Edit 2, and later Optima systems allowed the editor to use *any* Eidos system, rather than being tied down to a particular one, and still keep his data secure. The Optima software editing system was closely tied to Acorn hardware, so when Acorn stopped manufacturing the Risc PC in the late 1990s, Eidos discontinued the Optima system.

In the early 1990s a small American company called Data Translation took what it knew about coding and decoding pictures for the US military and large corporate clients and threw $12m into developing a desktop editor which would use its proprietary compression algorithms and off-the-shelf parts. Their aim was to 'democratize' the desktop and take some of Avid's market. In August 1993 Media 100 entered the market and thousands of would-be editors had a low-cost, high-quality platform to use.

Around the same period of time there were two other competitors, providing non-linear systems that required special hardware often cards that had to be added to the computer system. Fast Video Machine was a PC based system that first came out as an offline system and later became more online capable. Immix Video Cube was also a contender for Media Production companies. The Immix Video Cube had a control surface with faders to allow mixing and shuttle controls without the purchase of third party controllers. Data Translation's Media 100 came with 3 different JPEG codecs for different types of graphics of video and many resolutions. The Media 100 system kept increasing its video resolution via software upgrades rather than hardware. This because the core cards had enough processing power to be expanded to resolutions as high as Avid systems at the upper end of the Avid product line. Cards at that time had CPU's on the cards, for example a 68000 processor, which were as powerful as the cards inside the Macintosh systems that hosted the application. These other companies caused tremendous downward market pressure on Avid. Avid was forced to continually offer lower priced system to compete with the Media 100 and other systems.

Inspired by the success of Media 100, members of the Premiere development team left Adobe to start a project called "Keygrip" for Macromedia. Difficulty raising support and money for development led the team to take their non-linear editor to NAB. After various companies made offers, Keygrip was purchased by Apple as Steve Jobs wanted a product to compete with Adobe Premiere in the desktop video market. At around the same time, Avid — now with Windows versions of its editing software — was considering abandoning the Macintosh platform. Apple released Final Cut Pro in 1999, and despite not being taken seriously at first by professionals, it has evolved into a serious competitor to Avid.

DV

Another leap came in the late 1990s with the launch of DV-based video formats for consumer and professional use. With DV came IEEE 1394 (FireWire/iLink), a simple and inexpensive way of getting video into and out of computers. The video no longer had to be converted from an analog signal to digital data — it was recorded as digital to start with — and FireWire offered a straightforward way of transferring that data without the need for additional hardware or compression. With this innovation, editing become a more realistic proposition for standard computers with software-only packages. It enabled real desktop editing producing high-quality results at a fraction of the cost of other systems.

HD

More recently the introduction of highly compressed HD formats such as HDV has continued this trend, making it possible to edit HD material on a standard computer running a software-only editing application.

Avid is still considered the industry standard, with the majority of major feature films, television programs, and commercials created with its NLE systems. Final Cut Pro received an Technology & Engineering Emmy Award in 2002 and continues to develop a following.

Avid has held on to its market-leading position in the advent of cheaper software packages, notably Adobe Premiere in 1992 and Final Cut Pro in 1999. These three competing products by Avid, Adobe, and Apple are the foremost NLEs, often referred to as the A-Team. With advances in raw computer processing power, a number of new NLE software solutions have appeared on the market. An example of this is NewTek's software application SpeedEdit. Billed as the world's fastest video editing program, this application is an example of the continual streamlining and refinement of video editing software interfaces.

Since 2000, many personal computers include basic non-linear video editing software free of charge. This is the case of Apple iMovie for the Macintosh platform, PiTiVi for the Linux platform (it is installed by default on Ubuntu, the dominant desktop Linux distribution), and Windows Movie Maker for the Windows platform. This phenomenon has brought low-cost non-linear editing to consumers.

Quality

At one time, a primary concern with non-linear editing had been picture and sound quality. Storage limitations at the time required that all material undergo lossy compression techniques to reduce the amount of memory occupied.

Improvements in compression techniques and disk storage capacity have mitigated these concerns, and the migration to High Definition video and audio has virtually removed this concern completely. Most professional NLEs are also able to edit uncompressed video with the appropriate hardware.

Chapter 12

Multichannel Code

With the popularization of digital audio there is a Growing demand for compression and transmission techniques. The use of perceptual coding (based on psicoacustic model) has been a major breakthrough in Regarding the compressión digital audio. But it don't solve us the problem of multi-channel audio encoding.

The evolution of multichannel audio technology has gone Growing slowly beginning with the stereo going For existing systems 5.1 reaching systems of 10 or Even more channels. Besides these systems are no longer just for cinema or designed halls for reproduction, but actually they are used in what we call (home cinema).

The systems "home cinema" employ Dolby 5.1 which Does 5 channels and a sixth channel for the low frequencies. However, there are new applications that require the use of Many more channels. However, it seems important to emphasize the fact that all signs multichannel doesn't have the same Behavior. We could talk about two different categories:

The category 1 include Sound films intended to be reproduced in "home cinema 5.1" Systems or theatrical business. The Cross-correlation between channels tends to be high for symmetrical channels (L-Ls, R-Rs, C-R, C-L) but not among the rest of the channels.

The category 2 would those Live signals obtained using multiple Microphones to capture the acoustic properties of a room. The signals oriented for generate Acoustic fields, obtained by a linear grouping of Microphones, belong to this group. Such signals have a very high cross correlation between all channels.

Class 1 signals

To class 1 signals, the system most commonly used today is the Dolby Digital. This is used both in cinemas and in home cinema. His system compressión is dolby AC-3. The distribution of the speakers in a Dolby 5.1 is composed of three front channels (left, right and center), two surround channels (left and right) and a channel devoted to strengthening the bass effects. This channel is severely limited band of 20 to 120 Hz, while the other five have a frequency response of 20 Hz to 20kHz, so we are talking about a system of 5.1 channels.

The digital audio encoding that is used in the Compact Disc (16-bit PCM) achieves a dynamic range of 96 dB at the expense of working at a frequency 44.1 kHz with samples of 16-bit, which is a lot of data over to be stored or transmitted in a cost-effective manner, particularly in multichannel systems. That is why we need compression algorithm. The Dolby AC-3 achieves compression rates 10:1 also allowed for different bit rates depending on the number of channels encoded or quality required.

Dolby AC-3 has been designed to maximize the time and frequency masking characteristic of human hearing. This happens every signal to encode for a bank of filters, distributes bits that will be quantified with the espectral components of different bands in the light of the spectral characteristics of the encoded signal. An internal model that simulates the frequency masking and temporary hearing allows the encoder vary its resolution espectral-temporal depending on the nature of sound, in a way that ensures a minimum number of bits to describe each band signal in ensuring that the noise becomes totally masked. This model makes those masquerading frequency spectral components of the sound that will be masked by other are not encrypted. AC-3 also distributes the bits between the various channels so as to get a bit rate stable, allocating more bits to channels with a higher frequency content.

The algorithm AC-3 considers the six channels as a single entity by adding a single bit frame, which gets a bit rate less than separating each channel in a different frame.

The most important blocks of this algorithm are the following ones:

- Buffer-entry. - Filtering - Detector transient - Precombinación carrier

Buffer-entry

AC-3 is a block structured encoder, so that one or more blocks of samples of the signal in time are stored in the buffer for each channel input before proceeding with the prosecution. The blocks are usually composed of 512 samples.

Filtering

Input signals are individually filtered high pass at a frequency of 3 Hz to eliminate continuous component. The bass signal channel is also serious low-pass filtered at a frequency of 120 Hz.

Detector transient

We apply a band pass filter centred at high frequency that detects the presence of transients.

In the case of a signal that varies very quickly, such as the attack on a cymbal, we need a good temporal resolution of the same (which implies less spectral resolution), hence the block size must be to codify small for the quantization noise associated with this signal be temporarily confined in the vicinity of the same, so that this can be masked by noise that signal along the lines of masking temporary human ear.

It has imposed a limitation on the size variations that may suffer from the blocks in order to facilitate the process of consolidation; allowed eight different combinations of four types of window. Each of the eight combinations are identified by an ID Table The decoder must know at all times the kind of ID Table It is being used in the analysis of the signal, so that this information is multiplex together with the coefficients describing the signal. The information in Table ID Used in conjunction with their protection against mistakes is the 1% of the total bit rate.

Precombinación carrier .

In general the average bit multichannel systems is directly proportional to the square root of the number of channels. If we use 128 kbit / s to encode a single channel, an amount of 5.1 channels will require 128 • √ 5.1 = 289 kbit / s that can be transmitted using the speed with comfort typical working AC-3 (320 kbit / s). That is why most of the time will be sufficient to use as a method of compression algorithm of allocation of bits. However, when needed greater compression is also used method precombinación carrier.

This technique eliminates redundant information HF, and is based on the phenomenon psicoacustic that high frequencies in the human hearing is most sensitive to "surround" sound than the signal itself.

This behavior is used by the AC-3 separating the signals and high-frequency carrier envelope, so that information is encoded surround the more accurately the carrier.

The auditory impact is minimal, since the location of the sound is recorded on the envelope, which will combine sound in the ear by producing an effect equivalent to the original sound.

Besides all this, we take the high correlation that exists between channels using symmetric encryption difference and amount so we got also save more bit as symmetrical as channels són quite similar, we will need to encode only one, and the difference between this and the other channel.

Class 2 signals

KARUNEN-LOEVE Transform

For encoding multichannel signals for classe 2 exploit the properties of transform KARHUNEN-LOEVE. This is transformed into a product matrix type MxV = U. Where M is the matrix composed of the eigenvectors associated with the matrix covariances of V and U matrix signals uncorrelate that call matrix autochannels. V is the matrix that contains our multichannel signal to encode. And finally U és matrix output with our encoded signal.

KLT properties

What interesting of this operation is their properties. The first is that if we want to restore our original signals, we only have to multiply it by the matrix M but transposed. This greatly simplifies the decode time.

The second is that the U channels are ordered from highest energy to smaller energy. This is very useful for bit assignment for codification. The codification of the channels less energy Binary requires a bitrate much smaller than that of those more energy channels.

The third is that the signals obtained from the KLT retain the spectral characteristics and perceptive the audio signals. Therefore we can also use this property to apply perceptual coding.

Application of the properties for class 2 signals

Based on these properties we can reduce the bit rate. First seize the matrix M has its maximum values for diagonal. Since the diagonal is the cross-correlation of own channels. Therefore in the values that are far from the diagonal use fewer bits. The second property is the one that will take over as the last channels (which have a lower energy) is encode with less bits. This does not greatly affect the quality. And finally seize property to implement the third perceptual coding as can be for example encryption algorithm Advance Audio Codec (AAC). Using this technique can be achieved very high compression rates. But the only condition is that all channels have a high autocorrelation between all of them.

Chapter 13

ADAT and ADAT Lightpipe

ADAT

An ADAT XT 8-channel digital audio recorder

Alesis Digital Audio Tape or **ADAT** is a tape format used for simultaneously recording eight tracks of digital audio at once, onto Super VHS magnetic tape - a format similar to that used by consumer VCRs.

History

The product was announced in January 1991 at the NAMM convention in Anaheim, California. The first ADAT recorders shipped over a year later in February or March 1992. More audio tracks could be recorded by synchronizing up to 16 ADAT machines together, for a total of 128 tracks. While synchronization had been available in earlier machines, ADAT machines were the first to do so with sample-accurate timing - which in effect allowed a studio owner to purchase a 24-track tape machine eight tracks at a time. This capability and its comparatively low cost, originally introduced at $3995, were largely responsible for the rise of project studios in the 1990s.

Several versions of the ADAT machine were produced. The original ADAT (also known as "Blackface") and the ADAT XT recorded 16 bits per sample (ADAT Type I). A later generation of machines - the XT-20, LX-20 and M-20 - supports 20 bits per sample

(ADAT Type II). All ADAT machines use the same high quality S-VHS tape media. Tapes formatted in the older Type I style can be read and written in the more modern machines, but not the other way around. Later generations record at two sample rates, the 44.1 kHz and 48 kHz rates commonplace in the audio industry, although the original Blackface could only do 48 kHz. All models allow pitch control by varying the sample rate, and thus tape speed accordingly, so an original Blackface could record at 44.1 kHz (or another desired sample rate) if the pitch was lowered or raised by a specific amount.

With locate points it was possible to store sample exact positions on tape, making it easy to find specific parts of recordings. Using Auto Play and Auto Record functions made it possible to drop in recording at exact points, rather than relying on human ability to drop in at the right place.

ADAT machines could be controlled externally with the Alesis LRC (Little Remote Control), which could be attached to the ADAT with a 1/4" tip/sleeve plug, and featured the transport controls and most commonly used functions. Alternatively the BRC (Big Remote Control) could be used, which included many more features which the stand alone ADAT did not have, such as song naming, more locate points and MIDI Time Code synchronisation.

Current status

ADAT is a professional format, and while it has been replaced by the computer-based digital audio workstation, it is still used by some in the recording industry. It is also still in use for scientific work, and to drive laser light shows.

ADAT tapes are still available through pro audio retailers with products from Maxell, EMTEC (formerly the tape division of BASF) and HHB. HHB currently is the only company still manufacturing them.

Although it is a tape based format, the term ADAT now refers to its successor, the Alesis ADAT HD24. This is the next step in stand alone digital multitrack audio recorders, and features hard disk recording as opposed to the traditional tape based ADAT, which is now considered obsolete—although computer based recording holds back devices such as the HD24.

"ADAT" is also currently used as an abbreviation for the ADAT Lightpipe protocol, which transfers 8 tracks in a single fiber optic cable. The ADAT cable standard is no longer strictly tied to ADAT tape machines, and is now utilized by analog-to-digital converters, input cards for digital audio workstations, effects machines, etc. One of the original benefits of utilizing ADAT versus S/PDIF or AES3 was that a single cable could carry up to eight channels of audio. (AES10 (MADI) can now carry up to 64 channels.)

ADAT Lightpipe

The **ADAT Lightpipe**, officially the **ADAT Optical Interface**, is a standard for the transfer of digital audio between equipment. It was originally developed by Alesis but has since become widely accepted, with many third party hardware manufacturers including Lightpipe interfaces on their equipment. The protocol has become so popular that the term "ADAT" is now often used to refer to the transfer standard rather than the Alesis Digital Audio Tape itself.

Cables and interface

Lightpipe uses fiber optic cables (hence its name) to carry data, with Toslink connectors at either end, making them identical to S/PDIF optical cables. However, the data streams of the two protocols are totally incompatible. S/PDIF is mostly used for transferring stereo or multi-channel surround sound audio, whereas the ADAT optical interface supports up to 8 channels at 48 kHz, 24 bit. Recently, Lightpipe devices have been successfully interfaced via FireWire.

Data transfer

Lightpipe can carry eight channels of uncompressed digital audio at 24 bit resolution at 48,000 samples per second.

Although initially used for the transfer of digital audio between ADATs, the protocol was designed with future improvements in mind. All Lightpipe signals are transmitted at 24 bit resolution, no matter what the depth of the audio; information is contained within the Most Significant Bits and the rest of the bits remain a string of zeros. For example, if a 16 bit signal is sent via Lightpipe, the first sixteen bits contain the audio information while the other eight are simply occupied by zeros. The receiving device ignores information it cannot process. For example, a 20 bit signal going from a Type II ADAT to a Type I (which only operates at 16 bits) will simply ignore the bits below the sixteen MSBs.

Higher sample rates can be used with a proportionately reduced number of channels, although the original ADAT machines did not support this. The Alesis HD24 recorder demultiplexes high sample rates across multiple ADAT channels. A signal with a sample rate of 96 kHz is thus effectively split up into two signals of 48 kHz each, maintaining audio compatibility with the original ADATs.

Advantages

Lightpipe's main advantage is bit-transparent transfer of audio information. The lightpipe is "hot-pluggable", which means devices do not need to be turned off for plugging in or unplugging (although it is advisable to mute the receiving equipment, since there will be a large signal spike when the connection is made). The optical connect avoids groundloops, which can be troublesome in larger installations, and will not transfer any harmful electrical spikes from one device to the next.

Use in ADAT systems

Lightpipe was designed for use with the Alesis ADATs, and although extremely versatile, there are a few limitations. For straightforward digital audio transfer, the receiving device can synchronize to the lightpipe's embedded clock signal, achieving a 1:1 digital copy. For transport control, additional synchronization is needed between devices. (For example, using two ADAT machines at the same time to achieve 16-channel throughput would require better transport control; otherwise, the two ADAT machines would be very unlikely to play in sync.) Nine pin D connectors are used to transfer transport information. The Alesis ADAT HD24 also offers MIDI Time Code for synchronization with MIDI-enabled devices.

Lightpipe Bitstream

From the ADAT Lightpipe patent (US 5297181), the bitstream is transported in a frame of 256 bits. This frame is repeated at the desired sample rate (e.g. 48 kHz). Each of the eight channels contain 24 bits (MSB first), so 24 x 8 = 192 bits are allocated per frame. For synchronization, 6 sync (one) bits are interleaved per channel, adding 6 x 8 = 48 bits. The remaining 16 bits in the frame are allocated as: 4 user data bits, 2 sync bits and a frame sync pattern of 10 zero bits. The 10 zero bit pattern is unique in the frame and aids in synchronizing the bitstream at the ADAT receiver.

User data bit allocations

- User bit 0 is designated for Timecode transport
- User bit 1 is designated for MIDI data transport
- User bit 2 is designated for S/Mux indication (96 kHz sample rate mode)
- User bit 3 is reserved and set to 0

The transmission speed of the user bits is equal to the sampling rate (e.g. 48000 bits per second)

Competing Protocols

There are numerous digital audio transfer protocols. The most commonly used professional interface is AES3, developed by the Audio Engineering Society and the European Broadcasting Union, which transmits two channels of digital audio over a balanced XLR cable. S/PDIF (Sony/Philips Digital Interface) is the consumer version of this protocol, which uses either RCA leads or optical cables identical to lightpipe cables. MADI-X can carry 64 channels of audio at 48 kHz and 28 channels at 96 kHz, although it always requires a separate clock signal as it does not have one embedded.

However, recently, certainly in home and semi professional studios, USB and FireWire interfaces are the most popular means of transferring data. Their advantages over Lightpipe are large: compatibility is almost universal, all kinds of information can be transferred and a single cable can both send and receive data, whereas Lightpipe requires two separate leads for this. Yamaha's mLAN protocol exclusively uses the FireWire interface.

Chapter 14

Audio Interchange File Format

Audio Interchange File Format (**AIFF**) is an audio file format standard used for storing sound data for personal computers and other electronic audio devices. The format was co-developed by Apple Computer in 1988 based on Electronic Arts' Interchange File Format (IFF, widely used on Amiga systems) and is most commonly used on Apple Macintosh computer systems.

The audio data in a standard AIFF file is uncompressed pulse-code modulation (PCM). There is also a compressed variant of AIFF known as **AIFF-C** or **AIFC**, with various defined compression codecs.

Standard AIFF is a leading format (along with SDII and WAV) used by professional-level audio and video applications, and unlike the better-known lossy MP3 format, it is non-compressed (which aids rapid streaming of multiple audio files from disk to the application), and lossless. Like any non-compressed, lossless format, it uses much more disk space than MP3—about 10MB for one minute of stereo audio at a sample rate of 44.1 kHz and a sample size of 16 bits. In addition to audio data, AIFF can include loop point data and the musical note of a sample, for use by hardware samplers and musical applications.

The file extension for the standard AIFF format is **.aiff** or **.aif**. For the compressed variants it is supposed to be **.aifc**, but .aiff or .aif are accepted as well by audio applications supporting the format.

AIFF on Mac OS X

With the development of the Mac OS X operating system, Apple created a new type of AIFF which is, in effect, an alternative little-endian byte order format.

Because the AIFF architecture has no provision for alternative byte order, Apple used the existing AIFF-C compression architecture, and created a "pseudo-compressed" codec called **sowt** (**twos** spelled backwards). The only difference between a standard AIFF file and an AIFF-C/sowt file is the byte order; there is no compression involved at all.

Apple uses this new little-endian AIFF type as its standard on Mac OS X. When a file is imported to or exported from iTunes in "AIFF" format, it is actually AIFF-C/sowt that is being used. When audio from an audio CD disc is imported by dragging to the Mac OS X Desktop, the resulting file is also an AIFF-C/sowt. In all cases, Apple refers to the files simply as "AIFF", and uses the ".aiff" extension.

For the vast majority of users this technical situation is completely unnoticeable and irrelevant. The sound quality of standard AIFF and AIFF-C/sowt are identical, and the data can be converted back and forth without loss. Users of older audio applications, however, may find that an AIFF-C/sowt file will not play, or will prompt the user to convert the format on opening, or will play as static.

All traditional AIFF and AIFF-C files continue to work normally on Mac OS X (including on the new Intel-based hardware), and many third-party audio applications as well as hardware continue to use the standard AIFF big-endian byte order.

Note: As of Mac OS X version 10.4.9, the system will sometimes incorrectly display the AIFC icon for files with the **.aif** extension, whether or not the actual file format is AIFF or AIFF-C. This can be verified by opening the files in a hex editor and checking the FORM chunk's form type. This can sometimes happen when exporting files from QuickTime, and frequently happens when sending and receiving files between Windows and Mac computers or extracting files from an archive.

AIFF Apple Loops

Apple has also created another recent extension to the AIFF format in the form of Apple Loops used by GarageBand and Logic Audio, which allows the inclusion of data for pitch and tempo shifting by an application in the more common variety, and MIDI-sequence data and references to GarageBand playback instruments in another variety.

AppleLoops use the .aiff (or .aif) extension regardless of type.

Data format

An AIFF file is divided into a number of chunks. Each chunk is identified by a *chunk ID* more broadly referred to as FourCC.

Types of chunks found in AIFF files:

- Common Chunk (required)
- Sound Data Chunk (required)
- Marker Chunk
- Instrument Chunk

- Comment Chunk
- Name Chunk
- Author Chunk
- Copyright Chunk
- Annotation Chunk
- Audio Recording Chunk
- MIDI Data Chunk
- Application Chunk
- ID3 Chunk

AIFF-C common compression types

AIFF supports only uncompressed PCM data. AIFF-C also supports compression audio formats, that can be specified in the "COMM" chunk. The compression type is "NONE" for PCM audio data. The compression type is accompanied by a printable name. Common compression types and names include, but are not limited to:

AIFF-C common compression types

Compression Type	Compression Name	Data	Source
NONE	not compressed	PCM	Apple Computer, Inc.
fl32	32-bit floating point	IEEE 32-bit float	Apple Computer, Inc.
fl64	64-bit floating point	IEEE 64-bit float	Apple Computer, Inc.
alaw	ALaw 2:1	8-bit ITU-T G.711 A-law	Apple Computer, Inc.
ulaw	µLaw 2:1	8-bit ITU-T G.711 µ-law	Apple Computer, Inc.
ALAW	CCITT G.711 A-law	8-bit ITU-T G.711 A-law (64 kb/s)	SGI
ULAW	CCITT G.711 u-law	8-bit ITU-T G.711 µ-law (64 kb/s)	SGI
FL32	Float 32	IEEE 32-bit float	SoundHack & Csound
ADP4	4:1 Intel/DVI ADPCM		SoundHack
ima4	IMA 4:1		
ACE2	ACE 2-to-1		Apple IIGS ACE (Audio Compression/Expansion)
ACE8	ACE 8-to-3		

DWVW	Delta With Variable Word Width	TX16W Typhoon
MAC3	MACE 3-to-1	Apple Computer, Inc.
MAC6	MACE 6-to-1	Apple Computer, Inc.
Qclp	Qualcomm PureVoice	Qualcomm
QDMC	QDesign Music	QDesign
rt24	RT24 50:1	Voxware
rt29	RT29 50:1	Voxware