# Digital Audio Technologies, Audio Mixing and Sound Effects

Titus Neal

Nigel Mccarty

First Edition, 2012

ISBN 978-81-323-1473-8



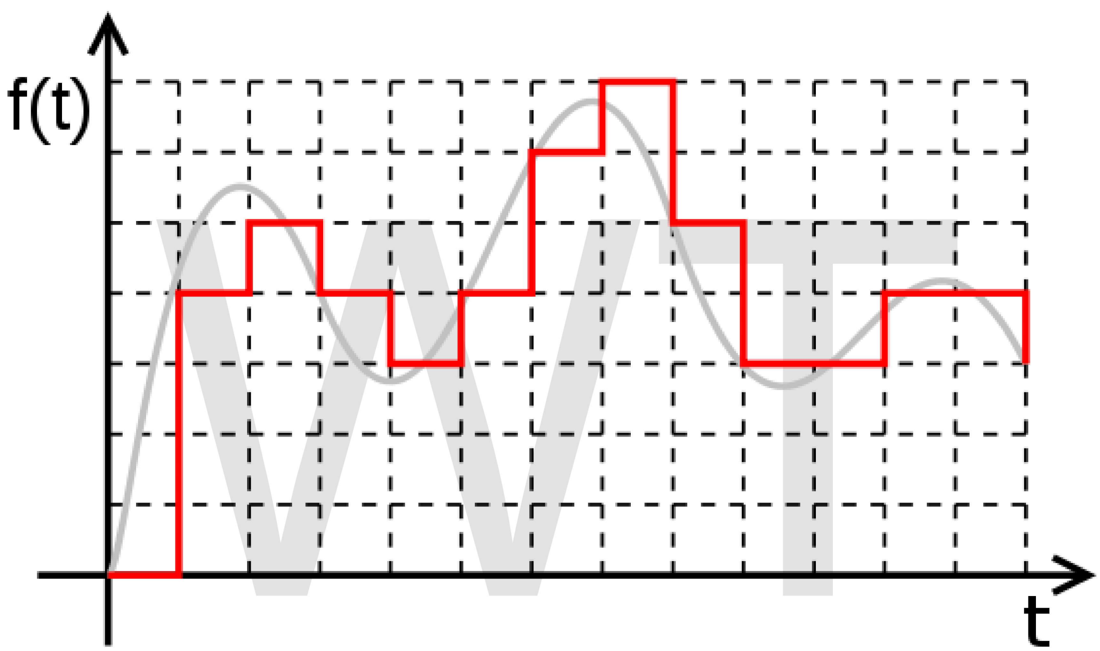


*Published by:*
**College Publishing House**
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

# Table of Contents

# Chapter 1

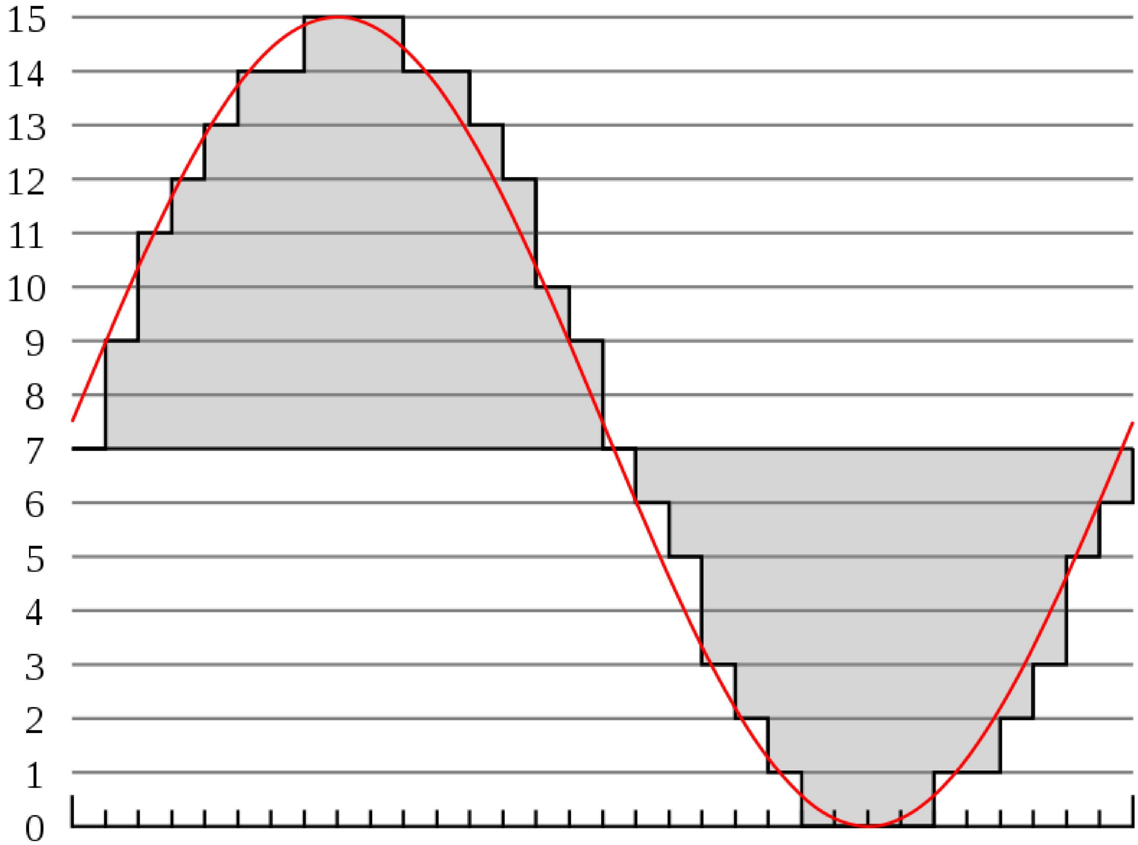
# Digital Audio




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.

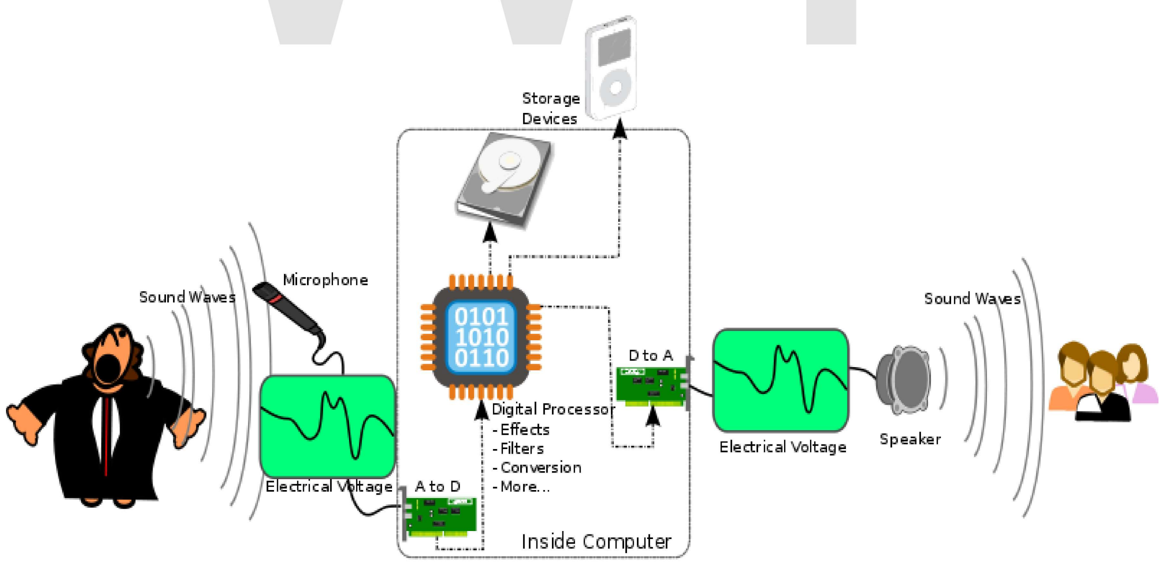
In an analog audio system, sounds begin as physical waveforms in the air, are transformed into an electrical representation of the waveform, via a transducer (for example, a microphone), and are stored or transmitted. To be re-created into sound, the process is reversed, through amplification and then 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<br>optional | 1 or 2<br>3 to 8 | 18.0 Mbit/s<br>18.0 | optional | 1 to 8 | 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.

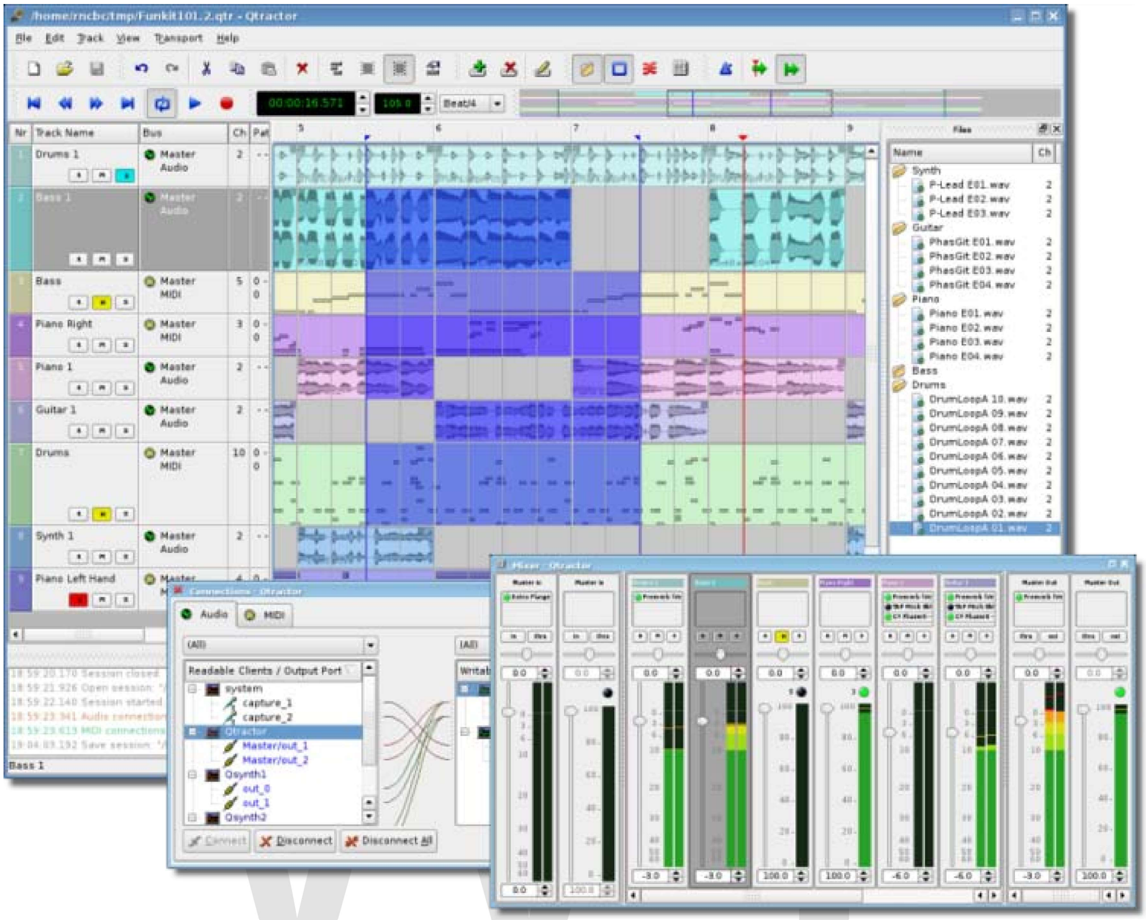
By the late 1980s, a number of consumer level computers such as the Apple Macintosh and Atari ST began to have enough power to handle the task of digital audio editing. Macromedia's Soundedit, with Microdeal's Replay Professional and Digidesign's "Sound Tools" and "Sound Designer" were used to edit audio 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)

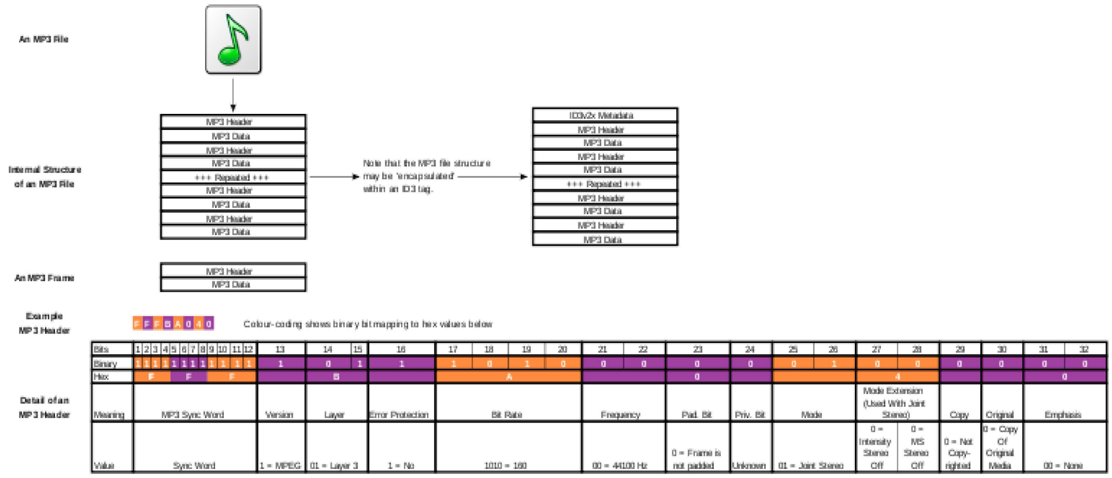
| MPEG-1 Audio Layer III | MPEG-2 Audio Layer III | nonstandard proprietary MPEG-2.5 Audio Layer III |
|---|---|---|
| - | - | 8000 Hz |
| - | - | 11025 Hz |
| - | - | 12000 Hz |
| - | 16000 Hz | - |
| - | 22050 Hz | - |
| - | 24000 Hz | - |
| 32000 Hz | - | - |
| 44100 Hz | - | - |
| 48000 Hz | - |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

## 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.

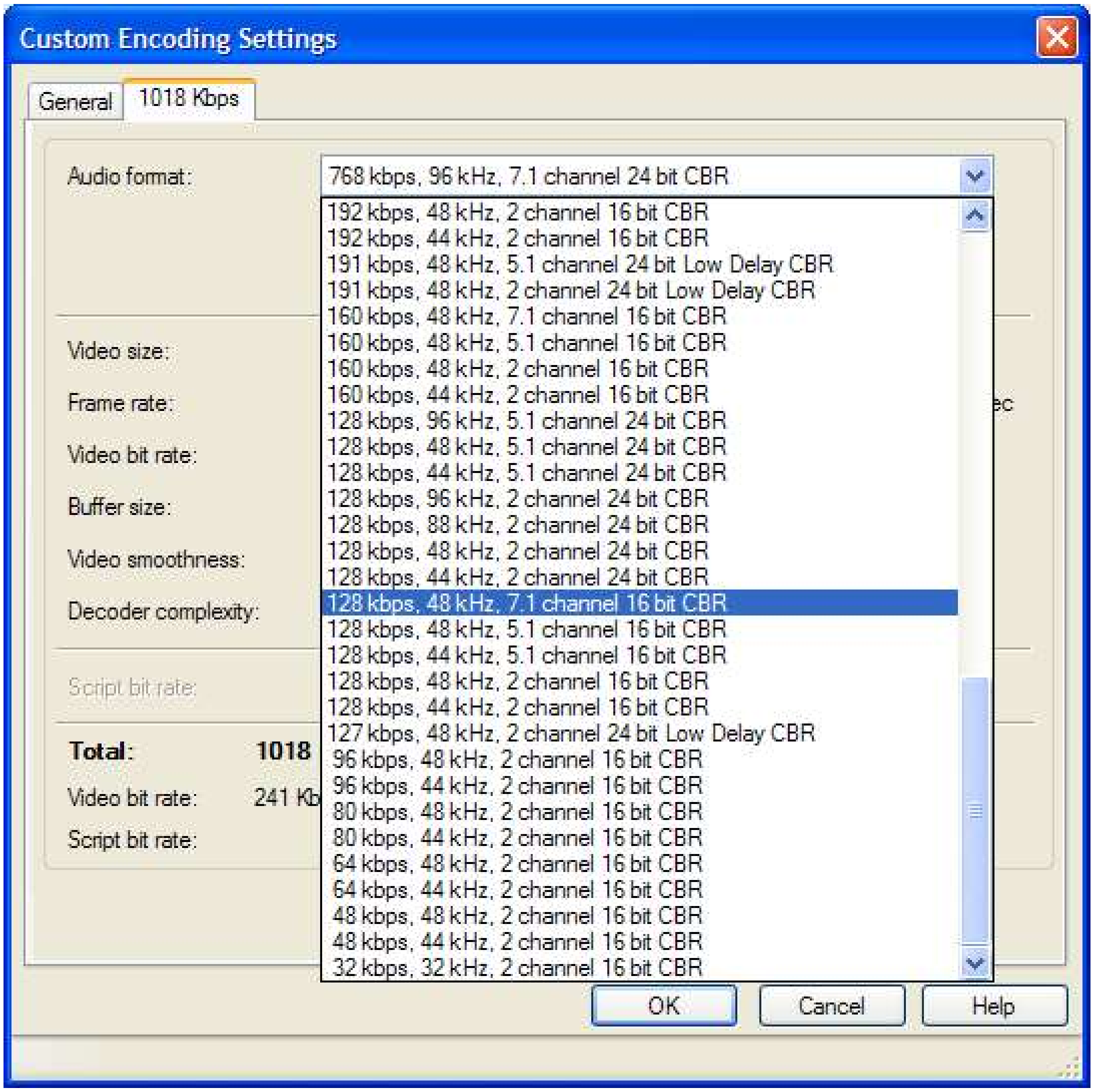
WMA can encode audio signals sampled at up to 48 kHz with up to two discrete channels (stereo). WMA 9 introduced variable bit rate (VBR) and average bit rate (ABR) coding techniques into the MS encoder although both were technically supported by the original format. WMA 9.1 also added support for low-delay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

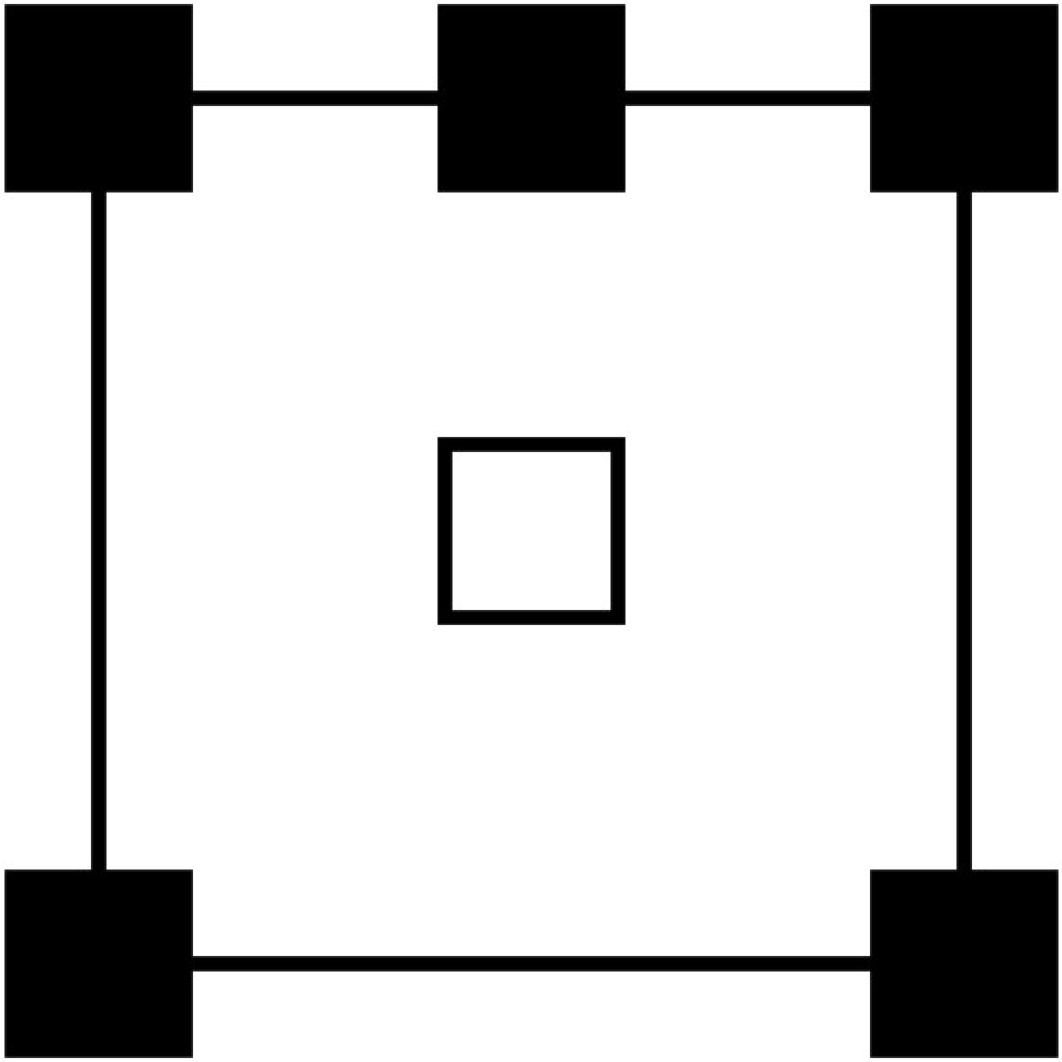
**Windows Media Audio Professional** (WMA Pro) is an improved lossy codec closely related to WMA standard. It retains most of the same general coding features, but also features improved entropy coding and quantization strategies as well as more efficient stereo coding. Notably, many of the WMA standard's 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

# Audio Mixing

Digital Mixing Console Sony DMX R-100 used in project studios

In audio recording, **audio mixing** is the process by which a multitude of recorded sounds are combined into one or more channels, most commonly two-channel stereo. In the process, the source signals' level, frequency content, dynamics and panoramic position are manipulated and effects such as reverb may be added. This practical, aesthetic or otherwise creative treatment is done in order to produce a mix that is more appealing to listeners.

Audio mixing is done in studios as part of an album or single making. The mixing stage often follows the multitrack recording stage and the final mixes are normally submitted to a mastering engineer. The process is generally carried out by a mix engineer, also called **mixing engineer**, or **mixer**, though sometimes it is the musical producer, or even the artist, who mixes the recorded material.

Prior to the emergence of DAWs (Digital Audio Workstations), the process of mixing used to be carried out on a device known as an *audio mixer*, *sound board*, *desk*, or *mixing console*. Nowadays, more and more engineers and independent artists are using a personal computer for the process (commonly referred to as mixing *in-the-box*).

## *The role of audio mixing*

The role of music producer is not necessarily a technical one, with the physical aspects of recording being assumed by the audio engineer, and so producers often leave the similarly technical mixing process to a specialist audio mixer. Even producers with a technical background may prefer that a mixer comes in to take care of the final stage of the production process. Noted producer and mixer Joe Chiccarelli has said that it is often better for a project that an outside person comes in because:

"when you're spending months on a project you get so mired in the detail that you can't bring all the enthusiasm to the final [mixing] stage that you'd like. [You] need somebody else to take over those responsibilities so that you can sit back and regain your objectivity."

However, as Chicarelli explains, sometimes limited budgets dictate that a producer takes care of the mixing as well.

## *History*

Mixing as we know it today emerged with the introduction of commercial multitrack tape machines, most notably the 8-track recorders that were introduced during the 1960s. The ability to record sounds into a multitude of channels meant that treating these sounds can be postponed to a later stage - the mixing stage.

In the 1980s, home recording and mixing began to take market share from recording studios. The 4-track Portastudio was introduced in 1979. Using one, Bruce Springsteen released the album *Nebraska* in 1982. The Eurythmics topped the charts in 1983 with the song "Sweet Dreams (Are Made of This)", recorded by bandmember Dave Stewart on a makeshift 8-track recorder. In the mid-to-late 1990s, computers replaced tape-based recording for most home studios, with the Power Macintosh proving popular. At the same time, digital audio workstations (DAW), first used in the mid-1980s, began to replace tape in many professional recording studios.

## *Equipment*

### Mixers

A mixer, or mixing console, or mixing desk, or mixing board, or software mixer is the operational heart of the mixing process. Mixers offer a multitude of inputs, each is fed by a track from a multitrack recorder; mixers would normally have 2 main outputs (in the case of two-channel stereo mixing) or 8 (in the case of surround).

Mixers offer three main functionalities:

- **Mixing -** summing signals together, which is normally done by a dedicated summing amplifier or in the case of digital by a simple algorithm.

- **Routing -** allows the routing of source signals to internal buses or external processing units and effects.
- **Processing -** many mixers also offer on-board processors, like equalizers and compressors.

Simple mixing console

## Outboard gear and plugins

Outboard gear (analog) and software plugins (digital) can be inserted to the signal path in order to extend processing possibilities. Outboard gear and plugins fall into two main categories:

- **Processors -** these devices are normally connected in series to the signal path, so the input signal is replaced with the processed signal (e.g. equalizers).
- **Effects -** while an effect can be considered as any unit that affects the signal, the term is mostly used to describe units that are connected in parallel to the signal path and therefore they add to the existing sounds, but do not replace them. Examples would include reverb and delay.

Common classes:

- **Processors**:
    - **Faders -** used to attenuate or boost the level of signals.

    - **Pan pots -** used to pan signal to the left or right and in surround also back and front.
    - **Equalizers -** used to manipulate the frequency content of signals.
    - **Compressors -** used to manipulate the dynamic content of signals. Among many applications they can even the level fluctuations of a singer, or reshape dynamic envelopes of percussive instruments (e.g. adding attack to a snare).
    - **Gates -** used mainly to attenuate low-level signals, for example, the kick spill on a snare recording.
- **Effects**:
    - **Reverbs -** used to simulate the boundary reflection created in a real room, but that adding a sense of space to otherwise 'dry' recordings.
    - **Delays -** most commonly used to add distinct echoes as a creative effect.

## *Mixing Domains*

The process of mixing often accounts for a few mixing domains:

- **Level -** concerned with the relative level between instruments and their dynamics.
- **Frequency -** concerned with the spectral content of the various instruments and the overall mix.
- **Space -** concerned with the spatial aspect of the various instruments. The space domain is often further subdivided into two sub-domains:
    - **Stereo -** concerned with the horizontal panoramic aspects of instruments.
    - **Depth -** concerned with the front-back aspects of instruments.

## *Mixing in Surround*

Mixing in surround is very similar to mixing in stereo except that there are more speakers, placed to 'surround' the listener. The same mixing domains mentioned above are involved, but instead of stereo's horizontal panoramic aspects, and depth's front-back aspects, mixing in surround lets the mix engineer pan sources within a much more three dimensional environment. In a surround mix, sounds can appear to originate from any direction.

There are two common ways to approach mixing in surround:

- **Expanded Stereo -** With this approach, the mix will still sound very much like an ordinary stereo mix. Most of the sources such as the instruments of a band, the vocals, and so on, will still be panned between the left and right speakers, but lower levels might also be sent to the rear speakers in order to create a wider stereo image, while lead sources such as the main vocal might be sent to the center speaker. Additionally, reverb and delay effects will often be sent to the rear speakers to create a more realistic sense of space. In the case of mixing a live recording that was performed in front of an audience, signal recorded by microphones aimed at, or placed among the audience will also often be sent to the rear speakers to make the listener feel as if he or she is in the crowd.

- **Complete Surround / All Speakers Are Treated Equally -** Instead of following the traditional ways of mixing in stereo, this much less conservative approach lets the mix engineer do anything he or she feels like. Instruments can appear to originate from anywhere, or even spin around the listener. When done tastefully, interesting sonic experiences can be achieved.

Naturally, these two approaches can be combined any way the mix engineer sees fit. Recently, a third approach, or method of mixing in surround was developed by surround mix engineer Unne Liljeblad.

- **MSS - Multi Stereo Surround -** This approach treats the speakers in a surround sound system as a multitude of stereo pairs. For example, a stereo recording of a piano, created using two microphones in an ORTF configuration, might have its left channel sent to the Left Rear Speaker and its right channel sent to the Center Speaker. The piano might also be sent to a reverb having its left and right outputs sent to the Left Front Speaker and Right Rear Speaker respectively. Additional elements of the song, such as an acoustic guitar recorded in stereo, might have its left and right channels sent to the Left Front Speaker and the Right Rear Speaker with a reverb returning to the Left Rear Speaker and the Center Speaker. Thus, multiple clean stereo recordings surround the listener without the smearing comb filtering effects that often occurs when the same or similar sources are sent to multiple speakers.

## Chapter 12

# Automixer

Microphones at a press conference being processed through a Dugan E-1 automixer which has been placed on top of the regular audio mixer. San Francisco mayor Gavin Newsom is speaking at a lectern, while golfers Fred Couples and Greg Norman are seated on stage. Five of eight automixer inputs have been muted and are showing red LEDs. The active input is showing full gain with a ladder of green LEDs

In professional audio, an **automixer** is a hardware or software device that balances multiple sound sources, usually microphones, based on each source's level, quickly and dramatically attenuating inactive inputs on the fly to deliver a more focused and intelligible mix that has less hiss, rumble, reverberation and noise. Automatic microphone mixers use a variety of protocols that allow increased gain before feedback for live sound reinforcement as well as reducing comb filtering between multiple microphones for recorded and broadcast applications.

Invented by Dan Dugan in 1976, automixers are typically used to mix panel discussions on television shows and at conferences and seminars. They can also be used to mix actors' wireless microphones in theater productions and musicals. They are frequently employed in commercial sound systems such as in courtrooms and city council chambers where it is not expected that a live sound operator will be present to mix the microphones. Wherever automixers are used in live sound reinforcement, their main benefit is that they work to maintain a steady limit on the overall signal level of the microphones; if a public address system is set up so that one microphone will not feed back, then, in general, multiple microphones will not feed back if they are automixed. The equivalent number of open mics (NOM) present at the output of the automixer is kept low, regardless of the actual number of open mics.

A skilled audio mix operator can greatly enhance the performance of a sound reinforcement system but will never be able to anticipate with perfect accuracy which participant will speak next in a free-wheeling discussion. Sudden interjections by panelists may be lost completely, or the beginning of a word may be absent until the operator responds as quickly as humanly possible to fade up their audio signal (this loss of the beginning is called *upcut*). A properly adjusted automixer can help in avoiding lost words or phrases due to upcut mistakes or lapses of attention.

## History

Frank J. Clement and Bell Labs received a patent in 1969 for a multiple station conference telephone system that switched its output to the loudest input. The next year, Emil Torick and Richard G. Allen were granted a patent for an "Automatic Gain Control System with Noise Variable Threshold", an adaptive threshold circuit invention with its patent assignation going to Columbia Broadcasting System.

Some systems using electro-mechanical switching in regard to microphone activation were engineered in the late 1960s and early 1970s. Peter W. Tappan and Robert F. Ancha devised a system of seat sensors that would activate one of 350 hidden microphones at the Seventeenth Church of Christ, Scientist in Chicago in 1970. From approximately 1968, Ken Patterson and Diversified Concepts developed a hardware system that could detect the "Number of Open Microphones" (NOM) and attenuate the master output by an amount which increased with a higher number of microphones in use. This latter system was public domain.

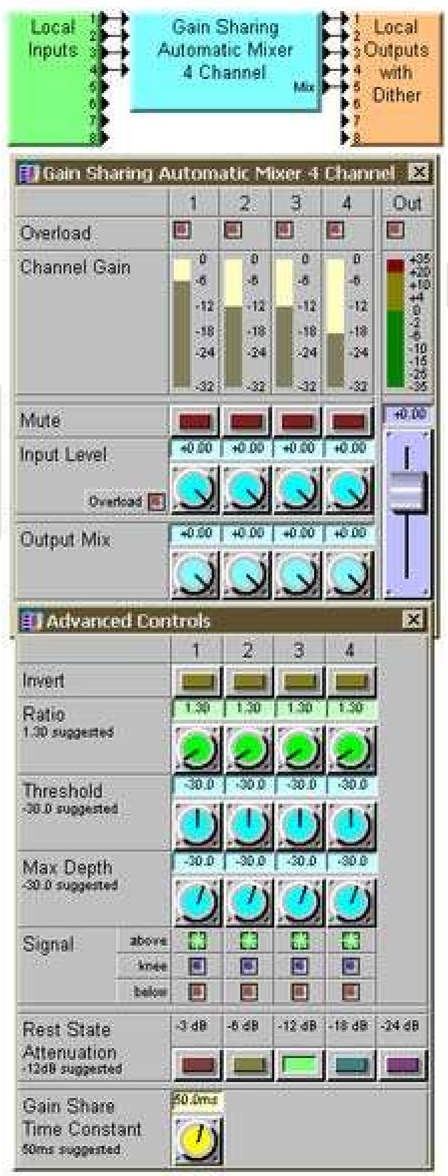
In 1971, Gregory Maston of Bell Labs filed for a patent involving a circuit that could switch between several audio sources based on their levels. The loudest one was latched into the mix. This system did not ramp switched signals smoothly in and out and did not maintain a constant ambience. It was intended for speakerphone conferencing applications. In 1972, Keith A. T. Knox with the British Post Office Corporation developed an adaptive threshold gate circuit intended for speakerphone usage. The system used a second microphone somewhat near the first to sense ambient noise level.


Dan Dugan's first automixers

Dan Dugan showed his first "Adaptive Threshold Automatic Microphone Mixing System" in 1974 at the 49th Audio Engineering Society (AES) meeting in New York, and was granted a patent for a control apparatus for sound reinforcement systems which sensed ambient sound level in the environment of a theater to control each microphone's individual level. In 1976, Dugan was granted a patent for an automatic microphone mixing process whereby the total gain of the system remains constant. He began manufacturing his first automixer system, the Model A, based on his two patents. Dugan built 60 units, with the first, hand-assembled one taken to Bell Labs to be installed in their conference room for Harvey Fletcher. The algorithm was elegantly simple: *Each individual input channel is attenuated by an amount, in dB, equal to the difference, in dB, between that channel's level and the sum of all channel levels.* Dugan licensed the system to Altec who released several automixer models including the 1674A, -B and -C series and the 1684A, all focusing on speech applications. (The 1684A became an Electrovoice product and is currently administered by their Commercial division.) The earliest Altec product implementation was regarded as inferior within the commercial audio contractor industry, and other manufacturers began to design their own automixer products.

In 1978, Richard W. Peters of Industrial Research Products (IRP) was granted an improvement patent entitled "Priority mixer control". IRP released the Voice-Matic series of 4x1 and 8x1 automatic mixers using "Dynamic Threshold Sensing" that weighed a combination of the amplitude and history of the signal to determine channel access. The master output was attenuated at the rate of 3 dB for every doubling of NOM. This master output reduction was the solution used by Yamaha Pro Audio two decades later in their DME series of digital signal processing (DSP) products, incorporating an automixer function which was otherwise an 8- or 16-channel noise gate.

Eugene Campbell and Terrance Whittemore of Colorado were granted a patent in 1982 for an automatic microphone mixing algorithm that allowed for musical performance mixing that would not be dominated by the loudest vocalist or instrumentalist.




Graphic user interface for a digital automixer that uses a gain-sharing protocol. Controls include threshold, depth, polarity inversion and muting for each input, as well as volume controls for the four inputs, the four individual outputs and the full mix output

Stephen D. Julstrom of Shure Brothers, Inc. (Evanston, Illinois) was granted a patent in 1987 for a teleconferencing system that used special directionally gated microphones mixed automatically and sent to a distant party via telephone line. The return signal from the distant party was compared in strength to the local mix to determine which party was to be most prominent in the overall mix. Any interrupting party was given priority. Four years later, Shure would introduce the AMS4000 and AMS8000 automixers for sound reinforcement; mixers which required the use of special directional condenser microphones of the Shure AMS Series.

At the 87th AES Convention in 1989, Dugan introduced the idea of using an automixer inserted within target channels on a professional audio mixer. Each microphone's signal would be interrupted inside its channel strip and sent to a variable gain circuit contained within the automixer. The signal would then be returned to the mixer at a level consistent with the Dugan algorithm. This became the Dugan Model D automixer.

In 1991, Dugan's patent expired. Competing manufacturers began to bring the Dugan algorithm directly to their product designs. In 1993, Travis M. Sims, Jr. of Lectrosonics (Rio Rancho, New Mexico) was granted a patent for a sound system with rate controlled, variable attenuation of microphone inputs, including the Dugan algorithm as well as loudspeaker zone attenuation when in close proximity to an active microphone. The loudspeaker zone part of the patent cited a 1985 patent for proportional amplification by Eugene R. Griffith, Jr. of LVW Systems of Colorado Springs, a commercial audio contractor. In 1995, Sims and Lectrosonics gained another patent for an "Adaptive proportional gain audio mixing system" which incorporated a number of ideas including the Dugan algorithm for maintaining a constant total gain of all the inputs.

In 1996, Dugan came out with the Model D-1, a speech-only economy model that did not offer the music system of the Model D.

In 1997, John H. Roberts of Peavey Electronics was granted a patent for an automatic mixer priority circuit, enabling a hierarchy of logic weighting that allowed selected signals to push forward in the mix when they are in use, while still maintaining the useful constant unity, gain-sharing relationship first described by Dugan. The hierarchy enabled a host, moderator or chairperson to speak over other participants and retain control of the discussion. Peavey's Architectural Acoustics division used three levels of hierarchy in their 1998 "Automix 2" product, placing the first- and second-most influentially weighted sources at inputs 1 and 2, respectively.

Dan Dugan licensed his system to Protech Audio (Indian Lake, New York) in 1997, yielding the Protech 2000 model series.

In 2004, the first standard audio mixer incorporating an eight-channel automixer section was released by Peavey in their Sanctuary Series, and in 2006 the similar HP-W was introduced by Crest. Both mixers were aimed at the House of Worship market, adding functions that ease the process of audio mixing to religious organizations.

In 2007, Mark W. Gilbert and Gregory H. Canfield of Shure (Niles, Illinois) were granted a patent for a digital microphone automixer system that used time of arrival as its main decision-making criteria.

Generations of Dugan's insertable automixers

In February 2011, Dugan announced an automixer card to plug in the accessory slot of a Yamaha digital mixing console such as the LS-9, M7CL or PM-5D. This card, the Dugan-MY16, can mix 16 channels of microphone inputs at 44.1–48 kHz or 8 channels at 88.2–96 kHz sampling rates. Channels to be automixed are assigned in the mixer's graphic user interface, and can then be controlled by a common web browser interface affecting only the Dugan-MY16 card, allowing remote control with an iPad, touchscreen computer or laptop over wireless network.

## Related applications

- Speech intelligibility enhancement, James M. Kates of Signatron (1984). This system uses Dugan's automatic mixing algorithm to reconstitute several spectral regions of a signal that has been divided into frequency bands for short-time spectral analysis in order to achieve greater intelligibility of spoken consonants.

- Secure conferencing, patent by Raoul E. Drapeau (1993). An automixing algorithm attempts to mask incidental speech that is below automix threshold but which can be audible in the mix. The automix circuitry indicates which sources are active, and whether masking of low-level signals is occurring.

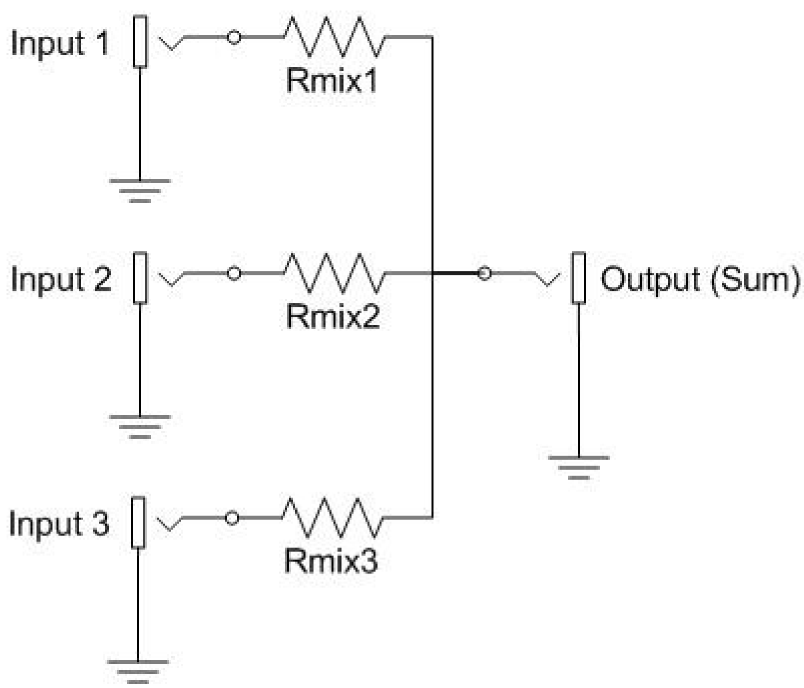
## Automixer manufacturers and products

- AKG Acoustics; AS8
- APB Dynasonics; ProSpec Auto-Mixer
- Audio-Technicaa; AT-MX341a SmartMixer.
- beyerdynamic; MCS 100
- Biamp Systems; Audia, Nexia, and AutoTwo
- Crest Audio; HP-W
- Crown; USM-810
- Gentner, Comrex, ClearOne; Converge Series, XAP Series
- Dan Dugan Sound Design; Models A, D, D-1, D-2, D-3, E and E-1
- Industrial Research Products; Voice-Tech
- Intelix; AMIX Series
- Ivie; AudioNet Automatic Matrix Mixers
- Lectrosonics; LecNet2 Series
- Peavey Electronics; Sanctuary, Mediamatrix, NION, Automix 4, Automix 2, etc.
- Rane; RPM 88
- Shure; SCM410, SCM810, FP410
- Symetrix; SymNet Automixing
- TOA; AX-1000A, 9000 Series Digital Matrix Mixer/Amplifier
- Yamaha Pro Audio; Dugan-MY16 card, for the accessory slot of the LS-9, M7CL, or PM-5D

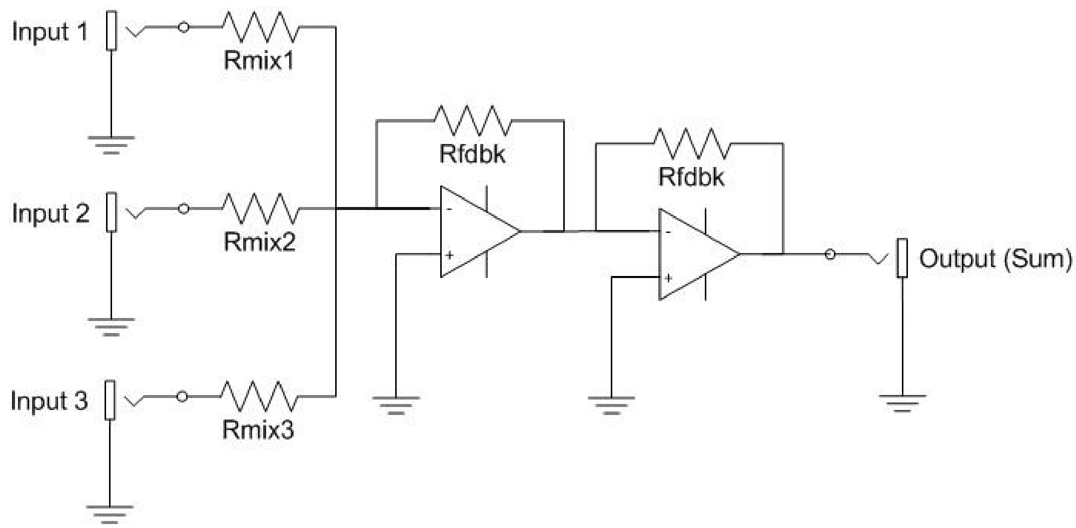
# Chapter 13

# Electronic Mixer and Beatmatching

## Electronic mixer




A simple three-channel passive additive mixer. More channels can be added by simply adding more input jacks and mix resistors.




A "virtual ground" active additive mixer. The buffer amplifiers serve to reduce crosstalk and distortion.

An **electronic mixer** is a device that combines two or more electrical or electronic signals into one or two composite output signals. There are two basic circuits that both use the term *mixer*, but they are very different types of circuits: additive mixers and multiplying mixers.

Additive mixers add two or more signals together, and this terminology ("mixer") is only used in the realm of audio electronics where audio mixers are used to add together audio frequency signals such as voice signals, music signals, and sound effects.

Multiplying mixers multiply together two time-varying input signals instantaneously (instant-by-instant). If the two input signals are both sinusoids of specified frequencies $f_1$ and $f_2$, then the output of the mixer will contain two new sinsoids that have the sum $f_1 + f_2$ frequency and the difference frequency absolute value$\{f_1 - f_2\}$.

Note: Any nonlinear electronic block driven by two signals with frequencies f1 and f2 would generate intermodulation (mixing) products. A multiplier (which is a nonlinear device too) will generate ideally only the sum and difference frequencies, whereas a arbitrary nonlinear block would generate also signals at e.g. 2*f1-3*f2, etc. Therefore in the past often more or less normal nonlinear amplifiers or just single diodes have been used as mixers, instead of a more complex multiplier. A multiplier has usually the advantage of rejecting - at least partly - undesired higher-order intermodulations and larger conversion gain.

## *Additive mixers*

Additive mixers add two or more signals, giving out a composite signal that contains the frequency components of each of the source signals. The simplest additive mixers are simple resistor networks, and thus purely passive, while more complex mixers employ active components such as buffer amplifiers for impedance matching and better isolation.

## *Product Mixers*

Ideal product mixers act as signal multipliers, producing an output signal equal to the product of the two input signals. Product mixers are often used in conjugation with an oscillator in the communications field to modulate signal frequencies. Product mixers can either up-convert or down-convert an input signal frequency, but they are more commonly used to down-convert to a lower frequency to allow for simpler filter designs, as done in superheterodyne receivers. In many typical circuits, the single output signal actually contains multiple waveforms, namely those at the sum and difference of the two input frequencies and harmonic waveforms. The output signal may be obtained by removing the other signal components with a filter.

Product mixers have been implemented in a wide variety of ways. The most popular are Gilbert cell mixers, diode mixers, diode ring mixers (ring modulation) and switching mixers. Diodes mixers take advantage of the non-linearity of diode devices to produce the desired multiplication in the squared term. It is a very inefficient method as most of the power output is in other unwanted terms which need filtering out. Inexpensive AM radios still use diode mixers.

Electronic mixers are usually made with transistors and/or diodes arranged in a balanced circuit or even a double-balanced circuit. These are readily manufactured by using the technology of either monolithic integrated circuits or hybrid integrated circuits. These are designed for a wide variety of frequency ranges, and they are mass produced to tight toleraces by the hundreds of thousands. These mixers, especially the double-balanced variety, and be bought in large numbers at prices ranging from a dime to a quarter apiece.

These double-balanced mixers are very widely used in microwave communication systems, satellite communication systems, and ultrahigh frequency (UHF) communications transmitters and receivers, and in radar systems transmitters and receivers.

*Gilbert cell* mixers are just an arrangement of transistors that multiplies the two signals. The switching mixers (below) pass more power and usually insert less distortion.

*switching mixers* use arrays of field effect transistors or (in older days) vacuum tubes). These are used as electronic switches, to permit the signal to go one direction, then the other. They are controlled by the signal being mixed. They are especially popular with digitally-controlled radios.

# Beatmatching

**Beatmatching** is a disc jockey technique of pitch shifting or timestretching a track to match its tempo to that of the currently playing track e.g. the kicks and snares in two house records hit at the same time when both records are played simultaneously. Beatmatching is a component of mixing which employs beatmatching combined with equalization, attention to phrasing and track selection in an attempt to make a single mix that flows together and has a good structure.

The technique was developed to keep the people from leaving the dancefloor at the end of the song. These days it is considered basic among DJs in electronic dance music genres, and it is standard practice in clubs to keep the constant beat through the night, even if DJs change in the middle.

Beatmatching is no longer considered a novelty, and new digital mixers have made the technique much easier to master.

## *Technique*

The beatmatching technique consists of the following steps:

1. While a record is playing, beatmatch a new record to it, using headphones for monitoring. Use gain (or *trim*) control on the mixer to match the levels of the two records.
2. Restart and slip-cue the new record at the right time, begin the new record on beat with the record currently playing. Pay attention to track structures; careful phrasing can make the mix seamless.
3. If the beat on the new record hits before the beat on the current record then the new record is too fast, reduce the pitch and manually slow the speed of the new record to bring the beats back in sync.
4. If the beat on the new record hits after the beat on the current record then the new record is too slow, increase the pitch and manually increase the speed of the new record to bring the beats back in sync.
5. Continue this process until the two records are in sync with each other, it can be difficult to sync the two records perfectly, so manual adjustment of the records is necessary to maintain the beat synchronization.
6. Before fading in the new track, check that the beats of two tracks match by listening to both channels together in the headphones, as the sound from the speakers can reach you with a delay.
7. Gradually, fade in parts of the new track while fading out the old track. While in the mix, ensure that the tracks are still synchronized, adjusting the records if needed.

## Pitch and tempo

The pitch and tempo of a track are normally linked together: spin a disc 5% faster and both pitch and tempo will be 5% higher. However, some modern DJ software can change pitch and tempo independently using time-stretching and pitch-shifting, allowing harmonic mixing. This technique is referred to as beatmatching.

## *History*

Beatmatching was invented by Francis Grasso in the late 1960s and early 1970s. Initially he was counting the tempo with a metronome and looking for records with the same tempo. Later a mixer was built for him by Alex Rosner which let him listen to any channel in the headphones independently of what was playing on the speakers; this became the defining feature of DJ mixers. That and turntables with pitch control enabled him to mix tracks with different tempo by changing the pitch of the *cued* track to match its tempo with the track being played by ear. Essentially, the technique he originated hasn't changed since.

These days beatmatching is considered central to DJing, and features making it possible are a requirement for DJ-oriented players. In 1978, the Technics SL-1200MK2 turntable was released, whose comfortable and precise sliding pitch control and high torque direct drive motor made beatmatching easier and it became the standard among DJs. With the advent of the compact disc, DJ-oriented Compact Disc players with pitch control and other features enabling beatmatching (and sometimes scratching), dubbed CDJs, were introduced by various companies. More recently, software with similar capabilities has been developed to allow manipulation of digital audio files stored on computers using turntables with special vinyl records (e.g. Final Scratch, M-Audio Torq, Serato Scratch Live) or computer interface (e.g. Traktor DJ Studio, Mixxx, Virtual DJ). Other software including algorithmic beatmatching is Ableton Live, which allows for realtime music manipulation and deconstruction, or Mixmeister, a DJ Mixset creation tool. Freeware software such as Rapid Evolution can detect the beats per minute and determine the percent BPM difference between songs.

The change from pure hardware to software is on the rise, and big DJs are introducing new equipment to their kits such as the laptop, and dropping the difficulty of carrying hundreds of CDs with them. The creation of the mp3-player allowed DJs to have an alternative tool for DJIng. Limitations with mp3-player DJing equipment has meant that only second generation equipment such as the IDJ2 or the Cortex Dmix-300 have the pitch control that alters tempo and allows for beatmatching on a digital music player. However, recent additions to the Pioneer CDJ family, such as the CDJ-400, allow mp3-player and other digital storage devices (such as external hard drives and USB memory sticks) to be connected to the CDJ device via USB. This allows the DJ to make use of the beatmatching capabilities of the CDJ unit whilst playing digital music files from the mp3-player or other storage device.

# Chapter 14

# Fade and Live Sound Mixing

## Fade

Audio mixer faders at the Bull & Gate pub in Kentish Town, North London

In audio engineering, a **fade** is a gradual increase or decrease in the level of an audio signal. The term can also be used for film cinematography or theatre lighting, in much the same way.

A recorded song may be gradually reduced to silence at its end (**fade-out**), or may gradually increase from silence at the beginning (**fade-in**). For example, the songs "Bitter Sweet Symphony" by The Verve and "Turn to Stone" by Electric Light Orchestra fade in from the beginning, while the songs "Born to Be Wild" by Steppenwolf, "Boogie Oogie Oogie" by A Taste of Honey, and "Hey Jude" by The Beatles fade out. However, "Born to be Wild" and "Boogie Oogie Oogie" fade out in a matter of seconds, whereas "Hey Jude" takes over 2 minutes to completely fade out. "Goodbye Stranger" by Supertramp takes about a minute to fade out. Fading-out can serve as a recording solution for pieces of music that contain no obvious ending.

Though relatively rare, songs can fade out, then fade back in. Some examples of this are "Helter Skelter" and "Strawberry Fields Forever" by The Beatles, "Suspicious Minds" by Elvis Presley, "Thank You" by Led Zeppelin, "Undercover of the Night" by The Rolling Stones, and "Bop Gun (Endangered Species)" by Parliament.

The term *fade* is also used in multi-speaker audio systems to describe the balancing of power between front and rear channels.

## *Origins and early examples*

"Neptune," part of the orchestral suite *The Planets*, written by Gustav Holst between 1914 and 1916, was the first piece of music to have a fade-out ending. Holst stipulates that the women's choruses are "to be placed in an adjoining room, the door of which is to be left open until the last bar of the piece, when it is to be slowly and silently closed", and that the final bar (scored for choruses alone) is "to be repeated until the sound is lost in the distance". Although commonplace today, the effect bewitched audiences in the era before widespread recorded sound—after the initial 1918 run-through, Holst's daughter Imogen (in addition to watching the charwomen dancing in the aisles during "Jupiter") remarked that the ending was "unforgettable, with its hidden chorus of women's voices growing fainter and fainter... until the imagination knew no difference between sound and silence".

The technique of ending a spoken or musical recording by fading out the sound goes back to the earliest days of recording. In the era of mechanical (pre-electrical) recording, this could only be achieved by either moving the sound source away from the recording horn, or by gradually reducing the volume at which the performer/s were singing, playing or speaking. With the advent of electrical recording, smooth and controllable fadeout effects could be easily achieved by simply reducing the input volume from the microphones using the fader on the mixing desk.

No single recording can be reliably identified as "the first" to use the technique. In 2003, on the (now-defunct) website *Stupid Question*, John Ruch listed the following recordings as possible contenders:

Bill Haley’s cover version of "Rocket 88" (1951) fades out to indicate the titular car driving away. There are claims that The Beatles’ "Eight Days a Week" (recorded 1964) was the first song to use the reverse effect—a fade-in.

The earliest such recording anybody could name for me is an 1894 78 rpm record called "The Spirit of ’76", a narrated musical vignette with martial fife-and-drum that gets louder as it 'nears' the listener and quieter as it 'moves away'.

The fade-out as a simulation of a moving sound source seems to continue right up to "Rocket 88". But other examples aren’t so obvious (though fade-out may always imply that the song continues forever and we’re only passing by it for a few minutes).

The oldest true songs with fade-out pointed out to me by 78 record fans bear no obvious relationship to movement. One is "Barkin’ Dog" (1919) by the Ted Lewis Jazz Band. Another contender is "America" (1918), a patriotic piece by the chorus of evangelist Billy Sunday.

By the early 1930s longer songs were being put on both sides of records, with the piece fading out at the end of Side One and fading back in at the beginning of Side Two. Records at the time held only about two to five minutes of music per side. The segue allowed for longer songs (such as Count Basie’s "Miss Thing"), symphonies and live concert recordings.

However, shorter songs continued to use the fade-out for unclear reasons—for example, Fred Astaire’s movie theme "Flying Down to Rio" (1933). Even using fade-out as a segue device doesn’t seem obvious, though we certainly take it for granted today.

As a film buff, I have a gut feeling that movies were an influence here. Fade-ins and fade-outs are cinematic devices that begin and end scenes—film language that developed at the same time as these early recordings. The term 'fade-out' itself is of cinematic origin, appearing in print around 1918. And jazz, a favorite of early records, was a popular subject of early movies, too.

## Fader

A **fader** is any device used for fading, especially when it is a knob or button that *slides* along a track or slot. A knob which *rotates* is usually not considered a fader, although it is electrically and functionally equivalent. A fader can be either analogue, directly controlling the resistance or impedance to the source (e.g. a Potentiometer); or digital, numerically controlling a digital signal processor (DSP). Digital faders are also referred to as *virtual* faders, since they can be viewed on the screen of a digital audio workstation. Modern high end digital mixers often have piezo-electric actuators attached to the faders such that they can be multi-use and will jump to the correct position for the selected function and/or saved setting.

## *Crossfading*

A **crossfader** on a dj mixer essentially functions like two faders connected side-by-side, but in opposite directions. It allows a DJ to fade one source out while fading another source in at the same time. This is extremely useful when beatmatching two sources of audio (or more, where channels can be mapped to one of the two sides of the crossfader individually) such as phonograph records, compact discs or digital sources.

The technique of **crossfading** is also used in audio engineering as a mixing technique, particularly with instrumental solos. A mix engineer will often record two or more takes of a vocal or instrumental part and create a final version which is a composite of the best passages of these takes by crossfading between each track.

In the perfect case the crossfader would keep constant output level. However, there's no standard on how this should be achieved. Many DJ equipment manufacturers offer different mixers for different purposes (e.g. scratching, beatmixing, cut mixing, etc.). High-end mixers often have crossfade curve switches allowing the DJ to select the type of crossfade necessary. Experienced DJs are also able to crossfade between tracks using the channel faders.

There are many software applications that feature virtual crossfaders. For instance, burning-software for the recording of audio-CDs.

## *Pre-fader, post-fader*

On a mixer with auxiliary send mixes, the send mixes are configured **pre**-fader or **post-**fader.

If a send mix is configured **pre-**fader, then changes to the main channel strip fader does not affect the send mix. In live sound reinforcement, this is useful for stage monitor mixes where changes in the Front of House channel levels would distract the musicians. In recording and post production, configuring a send to be **pre-**fader allows the amount of audio sent to the aux bus to remain unaffected by the individual track fader.

If a send mix is configured **post-**fader, then the level sent to the send mix follows changes to the main channel strip fader. This is useful for reverberation and other signal processor effects.

# Live sound mixing

A monitor engineer and console at an outdoor event

**Live sound mixing** is the art of combining and processing a number of audio signals together to create a "mix" that the audience or performers at a live show hear. There can be a variety of different mixes required, depending on the performance requirements. Three types are: Front of House (FOH), which is primarily for the audience; monitor, which is exclusively for the performer(s); and recording or broadcast (cue), for special purposes.

Whenever sound reinforcement is needed for a live performance of either music, theater, spoken word, or sporting events, a specialized sound system is required. The primary goal is to cover the audience area and stage with a sufficiently amplified signal. The stage or monitor mix is necessary to enable performers to hear themselves and any other parts of the performance as needed. Also, the proper monitor mix can minimize time delays on large stages to help synchronize the performance. In addition, the stage mix can overcome the level of the house sound which can be confusing to listen to on the stage.

The source of sounds for a live mix can be electronic musical instruments, acoustic instruments, playback of pre-recorded sounds and music, voices, other sounds ambience, and/or sound effects. This part of the sound system generally comprises a number of microphones on the stage, to pick up acoustic sounds, and/or a wide variety of other electronic signals.

If the mixing is to occur at a distance from the stage, it is customary for the individual signals to be balanced, low impedance in order to have noise immunity and retain their frequency spectrum. Widely differing levels can be accommodated in modern sound reinforcement systems. An additional requirement is to run the signals with standardized connectors and wiring.

## *Equipment*

A mixing board, a number of speakers (passive or active), power amplifiers, a number of audio processing devices, and the cabling, rigging, and power system to connect all of these components is usually what makes up a complete Sound Reinforcement System. Having the sound mixed or manipulated in real time is required as things are happening live and need constant minor adjustment. Some performers prefer to have the interactions of live musicians translated to the audience directly. An example of this is the old style bluegrass group using only one microphone. The musicians balance their ensemble sound by ear, and move toward the mic to emphasize solos. On the other end of the spectrum are musical or dramatic productions which can have many dozens of individual sources and dozens of sub-mixes out to dozens of speaker systems to deliver the proper mix to each of the performers.

A live sound engineer can mix the sound from the audience position, from a specialized control room, from the stage, or a remote truck, depending on the performance requirements. A trend in large scale theatrical productions is to minimize or eliminate the amount of sound equipment in the audience area so as to retain more seats for the audience. Digital measurement systems such as Smaart, Spectrafoo and Meyer SIM combined with test microphones in the audience area can be used to help the operator monitor the output of multiple loudspeaker zones. Digital control systems can be used for to implement indicated adjustments. For larger and more complex sound systems, more engineers and technicians can be required. The two primary engineers are the Front of House (FOH) engineer and the Monitor Engineer. The Front of House engineer mixes the sound that the audience hears in the house and the Monitor engineer mixes the sound that the performers hear on stage. A live sound engineer refers to a person that is experienced in the set up and operation of a sound reinforcement system.

Equipment for a touring act is often packed in heavy-duty reusable road cases to prevent damage during transportation, load-in, and load-out. Some items might be packed partially assembled and transported in custom-built cases or transport frames, to minimize set-up and tear-down time.

## *Audio engineer positions*

### Monitor engineer

The monitor engineer's role is most essential at music events, as opposed to spoken word events. In most cases, each performer on stage has their own individual mix that is custom tailored by the monitor engineer to suit their audio needs. The monitor engineer is then faced with the challenge of pleasing anywhere from four to ten or more musicians with a good mix. At shows with a separate monitor mix position, that mixer is typically located just off-stage, to provide easier communication between the performers and the monitor engineer. Though monitor speakers are still in use today, the newest monitor system is what is known as an In Ear Monitor (IEM) system. In Ear Monitors look somewhat like hearing aids, and they are basically a pair of headphones that are custom molded for the musicians' ears and therefore greatly reduces the outside noise that they hear. This isolation protects the musicians' ears from being damaged from the long durations of high volumes that they are subjected to on a large stage. It also allows them to hear their individual mix with more clarity. At the largest and highest budgeted of concert events, each musician is hearing their own individual in ear mix. This involves much more than simply mixing the sound, but requires a great deal of additional audio processing to increase the quality of the performer's mix. Large shows will often use a mixer that is specifically designed for monitor applications.

### Front of house engineer

Two FOH consoles at an outdoor event. Each console is typically dedicated to a single band or artist. The telephone handset on the right is part of a closed-circuit intercom system to allow the FOH engineer to communicate with the monitor engineer.

The front of house engineer controls the mix for the audience, and most often operates from the middle of the audience or at the last few rows of the audience from an equipment area known as the "**F**ront **O**f **H**ouse Position" or "FOH". A front of house engineer will often use a variety of processors and effects to provide a particular style to the mix. As with the monitor engineer, front of house engineers are constantly listening to the overall blend in order to make decisions about adjusting the volume and frequency of each instrument or voice on stage. The front of house engineer often makes decisions about which effects devices to use and adjusts their relative levels and blends to meet his or her interpretation of the musical requirements of the song. For smaller shows such as bar and smaller club gigs, it is common for the monitors to be mixed from the Front of House position, and the number of individual monitor mixes could be limited by the capabilities of the Front of House mixer.

Other crew members, such as the lighting console operator, might also work from the FOH position, since they need to be able to see the show from the audience's perspective.

### Additional Considerations for Large Shows

For shows that have separate Front of House and monitor mix positions, the audio snake is often designed to provide one or more 'splits' of the audio signals coming from the stage inputs. One split will go to the Front of House mixer, and the other will go to the monitor mixer. In such cases, the snake might be configured with switchable or permanent ground lifts or transformers to isolate the splits and help guard against electrical and RFI noise from being introduced into the system. This is important to take into consideration, as the isolated/lifted path(s) cannot supply phantom power to devices on stage. This is to ensure that phantom power comes from only one place. Very large snakes could also have additional splits for multitrack recorders, broadcast trucks, etc. Some acts use digital snakes to improve routing/control flexibility and save weight, compared to an analog snake and multicore cable.

### Crew Communications

For shows in larger venues, where line-of-sight communication between crew members is often not practical, communications are often accomplished with walkie-talkies. Additionally, a hard-line communications system, such as a closed-circuit intercom, might be in use to allow direct communication between the FOH and monitor engineers.

## *Set up, tear down, and technical rehearsals*

The other duty that the live sound engineer serves is the set up and the tear down (removal or striking) of these sound reinforcement systems. For large tours and events, this is often a long (sometimes multiple day) and strenuous process. This will involve unloading the equipment, moving it all into the venue, setting up the systems and then sound-checking. For larger events the engineer will be assisted by a number of audio technicians some of whom may be responsible for maintaining the system during the show while the mix engineer focuses on the sound of the show. After the show is done, the live sound engineers and techs must tear down and load out the sound system for the next show on the tour. The tear down will often take significantly less time than the set up, because there is a much more obvious end objective (i.e. having all the equipment packed into the trucks). Very large touring acts might have two complete sets of equipment that are used at alternating venues on the tour, due to the amount of time needed for the set-up and tear-down of each show. This is done to maximize the number of shows that can be performed in a given amount of time on the tour.

Larger shows and touring acts might have a few days to a few weeks of technical rehearsals. This is time before the start of a tour or a performance run that is used to work out many of the technical issues related to the show. Technical rehearsals are used both by the performers to hone portions of their performances, and also by the engineers, technicians and crew to resolve issues, refine cues, refine the configuration of show control systems, and other logistical details, such as finding the most efficient way to pack the equipment into trucks for transportation to the next venue, if the show is a touring act. In general, the larger and more complex a show is, the more time is allotted for technical rehearsals, though scheduling constraints might impose firm limits on the amount of rehearsal time that is available.

Live sound mixing is an art form in its own right as there are a number of different ways that the mix can be done and a number of different ways that the final mix can sound. The live sound engineer very often has at least a basic musical understanding so that he can make the proper decisions on how to mix different types of music and different types of songs at a concert.

## Chapter 15

# Mixing Console

In professional audio, a **mixing console**, or **audio mixer**, also called a **sound board**, **mixing desk**, or **mixer** is an electronic device for combining (also called "mixing"), routing, and changing the level, timbre and/or dynamics of audio signals. A mixer can mix analog or digital signals, depending on the type of mixer. The modified signals (voltages or digital samples) are summed to produce the combined output signals.

Mixing consoles are used in many applications, including recording studios, public address systems, sound reinforcement systems, broadcasting, television, and film post-production. An example of a simple application would be to enable the signals that originated from two separate microphones (each being used by vocalists singing a duet, perhaps) to be heard through one set of speakers simultaneously. When used for live performances, the signal produced by the mixer will usually be sent directly to an amplifier, unless that particular mixer is "powered" or it is being connected to powered speakers.

BBC Local Radio Mark III radio mixing desk

### *Structure*

Yamaha 2403 audio mixing console in a 'live' mixing application

A typical analog mixing board has three sections:

- Channel inputs
- Master controls
- Audio level metering

The channel inputs are replicated monaural or stereo input channels with pre-amp controls, channel fader and pan, sub-group assignment, equalization and auxiliary mixing bus level controls. The master control section has sub-group faders, master faders, master auxiliary mixing bus level controls and auxiliary return level controls. In addition it may have solo monitoring controls, a stage talk-back microphone control, muting controls and an output matrix mixer. On smaller mixers the inputs are on the left of the mixing board and the master controls are on the right. In larger mixers, the master controls are in the center with inputs on both sides. The audio level meters may be above the input and master sections or they may be integrated into the input and master sections themselves

## *Channel input strip*

The input strip is usually separated into these sections:

- Input jacks / microphone preamplifiers
- Basic input controls
- Channel EQ (High, Mids and low)
- Routing Section including Direct Outs, Aux-sends, Panning control and Subgroup assignments
- Input Faders

On the Yamaha Console above, these sections are color coded for quick identification by the operator. Each signal that is input into the mixer has its own *channel*. Depending on the specific mixer, each channel is stereo or monaural. On most mixers, each channel has an XLR input, and many have RCA or quarter-inch Jack plug line inputs.

### Basic input controls

Below each input, there are usually several rotary controls (knobs, pots). The first is typically a *trim* or *gain* control. The inputs buffer the signal from the external device and this controls the amount of amplification or attenuation needed to bring the signal to a nominal level for processing. This stage is where most noise of interference is picked up, due to the high gains involved (around +50 dB, for a microphone). Balanced inputs and connectors, such as XLR or Tip-Ring-Sleeve (TRS) quarter-inch connectors, reduce interference problems.

There may be *insert* points after the buffer/gain stage, which send to and return from external processors which should only affect the signal of that particular channel. Insert points are most commonly used with effects that control a signal's amplitude, such as noise gates, expanders, and compressors.

### Auxiliary send routing

The *Auxiliary send* routes a split of the incoming signal to an auxiliary bus which can then be used with external devices. *Auxiliary sends* can either be pre-fader or post-fader, in that the level of a pre-fade send is set by the *Auxiliary send* control, whereas post-fade sends depend on the position of the channel fader as well. *Auxiliary sends* can be used to send the signal to an external processor such as a reverb, which can then be routed back through another channel or designated auxiliary returns on the mixer. These will normally be post-fader. Pre-fade *auxiliary sends* can be used to provide a monitor mix to musicians onstage, this mix is thus independent of the main mix.

Allen & Heath Mixing desk used for live performances

## Channel equalization

Further channel controls affect the equalization (EQ) of the signal by separately attenuating or boosting a range of frequencies, e.g., bass, midrange, and treble. Most large mixing consoles (24 channels and more) usually have sweep equalization in one or more bands of its parametric equalizer on each channel, where the frequency and affected bandwidth of equalization can be selected. Smaller mixing consoles have few or no equalization controls. Care must be taken not to add too much EQ to a signal that is already close to clipping; additional energy will overdrive the channel.

Some mixers have a general equalization control (either graphic or parametric) at the output.

### Subgroup and mix routing

Each channel on a mixer has an audio taper pot, or potentiometer, controlled by a sliding volume control (*fader*), that allows adjustment of the level, or amplitude, of that channel in the final *mix*. A typical mixing console has many rows of these sliding volume controls. Each control adjusts only its respective channel (or one half of a stereo channel); therefore, it only affects the level of the signal from one microphone or other audio device. The signals are summed to create the main *mix*, or combined on a *bus* as a submix, a group of channels that are then added to get the final mix (for instance, many drum mics could be grouped into a bus, and then the proportion of drums in the final mix can be controlled with one bus fader).

There may also be *insert* points for a certain bus, or even the entire mix.

## *Master output controls*

Subgroup and main output fader controls are often found together on the right hand side of the mixer or, on larger consoles, in a center section flanked by banks of input channels. Matrix routing is often contained in this master section, as are headphone and local loudspeaker monitoring controls. Talkback controls allow conversation with the artist through their wedges, headphones or IEMs (in-ear monitor). A test tone generator might be located in the master output section. Aux returns such as those signals returning from outboard reverb devices are often in the master section.

## *Metering*

Finally, there are usually one or more VU or peak meters to indicate the levels for each channel, or for the master outputs, and to indicate whether the console levels are overmodulating or clipping the signal. Most mixers have at least one additional output, besides the main mix. These are either individual bus outputs, or *auxiliary outputs*, used, for instance, to output a different mix to on-stage monitors. The operator can vary the mix (or levels of each channel) for each output.

As audio is heard in a logarithmic fashion (both amplitude and frequency), mixing console controls and displays are almost always in decibels, a logarithmic measurement system. This is also why special audio taper pots or circuits are needed. Since it is a relative measurement, and not a unit itself (like a percentage), the meters must be referenced to a nominal level. The "professional" nominal level is considered to be +4 dBu. The "consumer grade" level is −10 dBV.

## Hardware routing and patching

For convenience, some mixing consoles include inserts or a patch bay or patch panel. Patch bays are mainly used for recording mixers.

## Other features

Most, but not all, audio mixers can

- add external effects.
- use monaural signals to produce stereo sound by adjusting the position of each signal on the sound stage (pan and balance controls).
- provide phantom power (typically 48 volts) required by some microphones.
- create an audible tone via an oscillator, usually at 440 Hz, 1 kHz, or 2 kHz

Some mixers can

- add effects internally.
- read and write console automation.
- be interfaced with computers or other recording equipment (to control the mixer with computer presets, for instance).
- control or be controlled by a Digital Audio Workstation via Midi or proprietary commands.
- be powered by batteries.

## Digital versus analog

Digidesign's Venue Profile mixer on location at a corporate event. This digital mixer allows plugins from third-party vendors

Digital mixing console sales have increased dramatically since their introduction in the 1990s. Yamaha sold more than 1000 PM5D mixers by July, 2005, and other manufacturers are seeing increasing sales of their digital products. Digital mixers are more versatile than analog ones and offer many new features, such as the ability to save multiple mute groups, multiple VCA groups and channel settings into a scene and reconfigure signal routing at the touch of a button. The faders can be "swapped" or "flipped" to show aux send levels; a feature very useful in mixing artists' monitors. In addition, digital consoles often include a range of special effects such as parametric EQ, compression, gating, reverb, automatic feedback reduction, tap delay and straight delay. Some products are expandable via third-party software features (called plugins) that add further reverb, compression, delay and tone-shaping tools. Several digital mixers include spectrograph and real time analyzer functions. A few incorporate loudspeaker management tools such as crossover filtering and limiting. Digital signal processing can perform automatic mixing for some simple applications, such as courtrooms, conferences and panel discussions, but at this time no digital mixer in live audio includes automixing. Consoles with motorized faders can read and write console automation.

Digital mixers can be designed to be quieter than most analog mixers, as digital mixers often incorporate very low threshold noise gates to stop inactive mix bus background hiss from summing with active signals. Digital circuitry is more resistant to outside interference from radio transmitters such as walkie-talkies and cell phones.

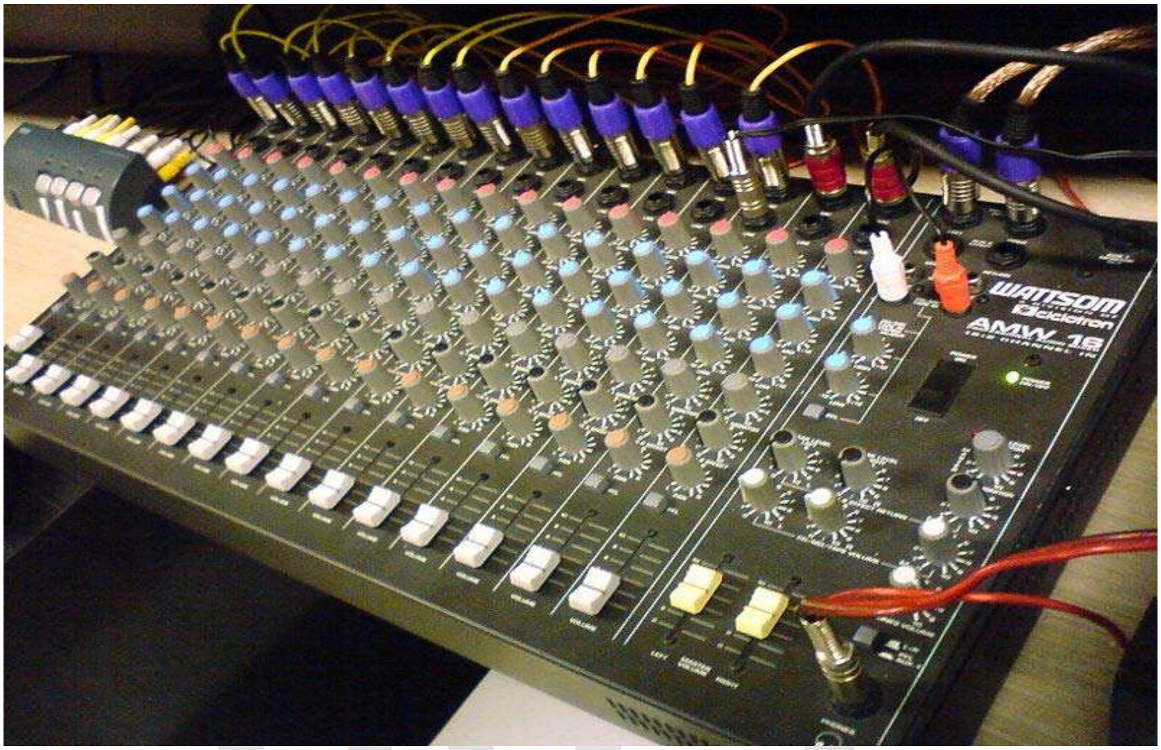
## Propagation delay

Digital mixers have an unavoidable amount of latency or propagation delay, ranging from 1.5 ms to as much as 10 ms, depending on the model of digital mixer and what functions are engaged. This small amount of delay isn't a problem for loudspeakers aimed at the audience or even monitor wedges aimed at the artist, but can be disorienting and unpleasant for IEMs (In ear monitors) where the artist hears their voice acoustically in their head *and* electronically amplified in their ears but delayed by a couple of milliseconds.

Every analog to digital conversion and digital to analog conversion within a digital mixer entails propagation delay. Audio inserts to favorite external analog processors make for almost double the usual delay. Further delay can be traced to format conversions such as from ADAT to AES3 and from normal digital signal processing steps.

Within a digital mixer there can be differing amounts of latency, depending on the routing and on how much DSP is in use. Assigning a signal to two parallel paths with significantly different processing on each path can result in extreme comb filtering when recombined. Some digital mixers incorporate internal methods of latency correction so that such problems are avoided.

## Ease of use




16-channel mixing console with compact short-throw faders

Analog consoles remain popular due to their continuing to have one knob, fader or button per function, a reassuring feature for the user. This takes up more physical space but allows more rapid response to changing performance conditions. Most digital mixers take advantage of the technology to reduce the physical space requirements of their product, entailing compromises in user interface such as a single shared channel adjustment area that is selectable for only one channel at a time. Additionally, most digital mixers have virtual pages or layers which change the fader banks into separate controls for additional inputs or for adjusting equalization or aux send levels. This layering can be confusing for operators.

Analog consoles make for simpler understanding of hardware routing. Many digital mixers allow internal reassignment of inputs so that convenient groupings of inputs appear near each other at the fader bank, a feature that can be disorienting for persons having to make a hardware patch change.

On the other hand, many digital mixers allow for extremely easy building of a mix from saved data. USB flash drives and other storage methods are employed to bring past performance data to a new venue in highly portable manner. At the new venue, the traveling mix technician simply plugs the collected data into the venue's digital mixer and quickly makes small adjustments to the local input and output patch layout, allowing for full show readiness in very short order.

Some digital mixers allow offline editing of the mix, a feature that lets the traveling technician use a laptop to make anticipated changes to the show while *en route*, further shortening the time it takes for the sound system to be ready for the artist.

## Sound quality

Both digital and analog mixers rely on analog microphone preamplifiers, a high-gain circuit that is the origin of much of the perceived character of sound quality in an audio mixer. In this respect, both formats are on par with each other. In a digital mixer, the microphone preamplifier is followed by an ADC which quantizes the audio stream. Ideally, this process is carefully engineered to deal gracefully with overloading and clipping while delivering an accurate digital stream over the linear dynamic range. Further processing and mixing of digital streams within a mixer need to avoid clipping and truncation if maximum audio quality is desired.

Analog mixers, too, must deal gracefully with overloading and clipping at the microphone preamplifier and as well as avoiding overloading of mix buses. Background hiss in an analog mixer is always present, though good gain stage management minimizes its audibility. Idle subgroups left "up" in a mix will add their background hiss to the main outputs; many digital mixers avoid this problem by low-level gating.

Many electronic design elements combine to affect perceived sound quality, making the global "analog mixer vs. digital mixer" question difficult to answer. Controlled ABX double-blind listening tests have not been published at this date; no conclusive answer can be reached. Experienced live sound professionals agree that microphones and loudspeakers (with their innate higher distortion levels) are a much greater source of coloration of sound than the choice of mixer. The mix style of the person mixing is also more important than the make and model of audio console. Analog and digital mixers both have been associated with extremely high-quality concert performances and studio recordings.

## Remote control

Analog mixing in live sound has had the option since the 1990s of using wired remote controls for certain digital processes such as monitor wedge equalization and parameter changes in outboard reverb devices. That concept has expanded until wired and wireless remote controls are being seen in relation to entire digital mixing platforms. It's possible to set up a sound system and mix via wireless (or wired) laptop, touchscreen or tablet, especially if the performance requires no unpredictable fast responses to multiple changing conditions on stage. Computer networks can connect digital system elements for expanded monitoring and control, allowing the system technician to make adjustments to distant devices during the performance. The use of remote control technology can be utilized to reduce "seat-kills", allowing more paying customers into the performance space.

## Virtual mixing

Increasingly, the mixing process can be performed on screen, using computer software and associated input, output and recording hardware. The traditional large control surface of the mixing console is not utilized, saving space at the engineer's mix position. Some virtual mixing (such as the Gamble DCX) uses digital controls of analog audio circuitry, but most virtual mixers are fully digital so as to save cost and physical space. In the virtual studio, there is either no normal mixer fader bank at all or there is a compact group of motorized faders designed to fit into a small space and connected to the computer via USB or Firewire. Many project studios use such a space-efficient solution, as the mixing room at other times can serve as business office, media archival, etc. Virtual mixing is heavily integrated as part of a digital audio workstation.

## *Applications*

A Behringer EuroRack UB1002FX in a DJ setup

Dub producers/engineers such as Lee "Scratch" Perry were perhaps the first musicians to use a mixing board as a musical instrument.

Public address systems will use a mixing console to set microphones for different speakers to the correct level, and can add in recorded sounds into the mix. A major requirement is to minimise audio feedback.

Most bands will use a mixing console to combine musical instruments and vocals to the correct level.

Radio broadcasts use a mixing desk to select audio from different sources, such as CD players, telephones, remote feeds, or prerecorded advertisements.

Noise music musicians such as Merzbow or Wolf Eyes may create feedback loops within mixers, creating an instrument known as a no-input mixer. The tones generated from a no-input mixer are created by connecting an output of the mixer into an input channel and manipulating the pitch with the mixer's dials.

Chapter 16

# 3D Audio Effect and Audio Feedback

## 3D audio effect

**3D audio effects** are a group of sound effects that attempt to widen the stereo image produced by two loudspeakers or stereo headphones, or to create the illusion of sound sources placed anywhere in 3 dimensional space, including behind, above or below the listener.

There are several types of **3D audio effects**:

- Those that only widen the stereo image by modifying phase information.
- Those that can place sounds outside the stereo basis.
- Those that include a complete 3D simulation.

### *Stereo widening*

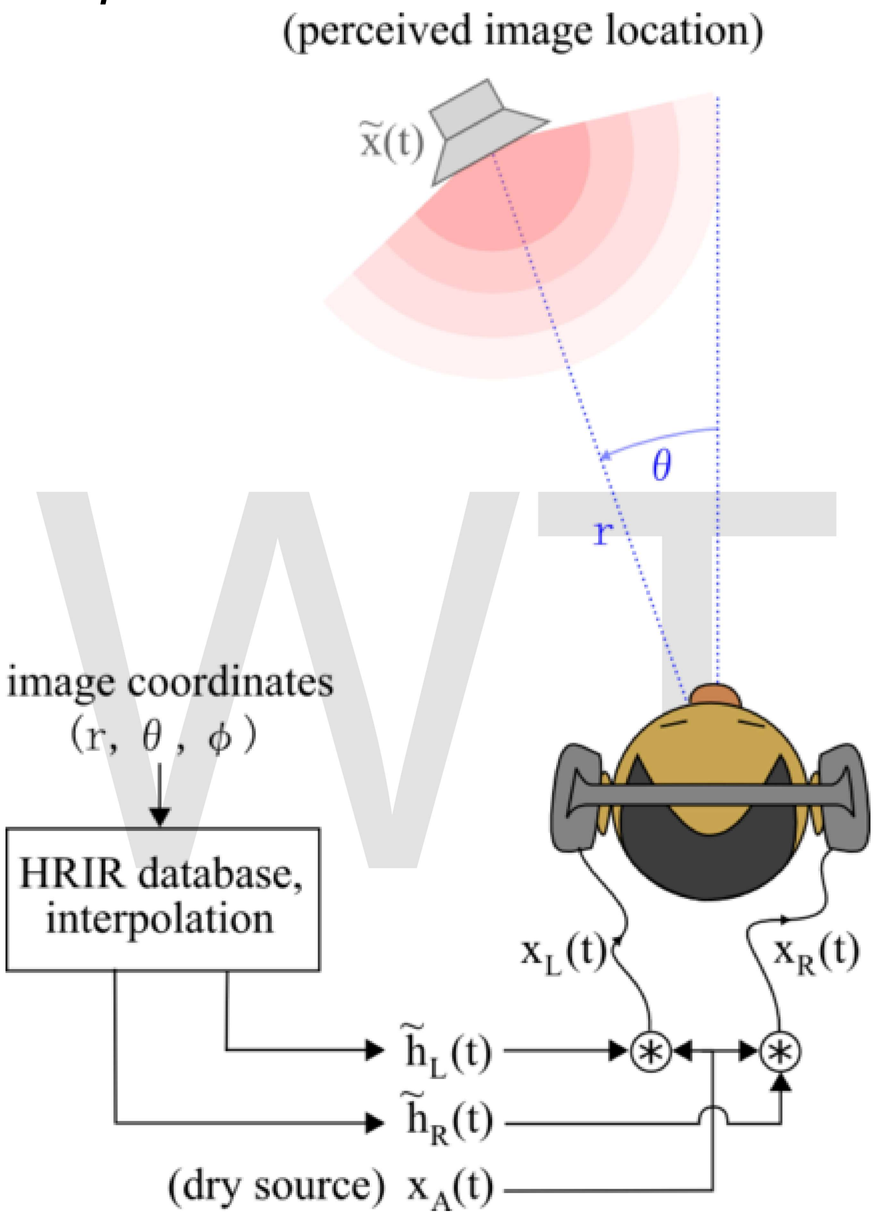
Widening of the stereo image can be achieved by manipulating the relationship of the side signal S and the center signal C: $C = \frac{L+R}{2}; S = \frac{L-R}{2}$ . A positive part of the side signal S is now fed into the left channel and a part with its phase inverted to the right channel. Some boomboxes feature such a process.

Another way of looking at this same effect, without extrapolating a center and side signal from the left and right signals, is to simply add the left signal, slightly attenuated and phase inverted, into the right channel and vice-versa. Taking this a step further, a small delay (20-100ms) can be added to the inverted signal before mixing it back in to the original for output, adding a slight reverberation to the effect.

### *Placement of sounds outside the stereo basis*

By manipulating parts of the sound according to psychoacoustic findings in phase and sound, it is possible to create sounds beyond the stereo basis. Effects from QSound Labs have been used on albums from Sting and Madonna in the beginning of the 1990s, as well as in the videogame Super Street Fighter II. Similarly, the pioneering work of researchers (Sibbald et al.) at EMI Central Research Labs in England in the 1980s, and later with Sensaura, produced "3D Audio" CDs...

### *Complete 3D positional audio*




A sound is placed in the horizontal plane by processing the sound with recorded head-related impulse responses.

The 3D simulation is the most advanced group of **3D audio effects**. Using head-related transfer functions and reverberation, the changes of sound on its way from the source (including reflections from walls and floors) to the listener's ear can be simulated. These effects include localization of sound sources behind, above and below the listener.

Some 3D technologies also convert binaural recordings to stereo recordings. MorrowSoundTrue3D converts binaural, stereo, 5.1 and other formats to 8.1 single and multiple zone 3D sound experiences in realtime.

3D Positional Audio effects emerged in the 1990s in PC and Game Consoles. As a medium, interactive games would benefit perhaps more than any other. However, although some technologies do seem to work better than others, 3D sound in games is still quite unconvincing, especially over speakers.

3D audio techniques have also been incorporated in music and video-game style music video arts. The Audioscape research project, provides musicians with a real-time 3D audiovisual content authoring and rendering environment, suitable for live performance applications.

A site with animations and theory of a system using HRTF's to create 3D Audio: ISVR Virtual Acoustics.

True representation of the elevation level for 3D loudspeaker reproduction become possible by the Ambisonics and Wave field synthesis Principle, MorrowSound True3D and A&G 3D-EST.

3D-EST is a new approach to 3D soundscape recording and reproduction that does not use HRTF or reverberation.

## *3-D audio presentations*

Some amusement parks have created attractions based around the principles of 3-D audio. One example is *Sounds Dangerous!* at Disney's Hollywood Studios at the Walt Disney World Resort in Florida. Guests wear special earphones as they watch a short film starring comedian Drew Carey. At a point in the film, the screen goes dark while a 3-D audio sound-track immerses the guests in the ongoing story. To ensure that the effect are heard properly, the earphone covers are color-coded to indicate how they should be worn. This is not a generated effect but a binaural recording.

MorrowSoundTrue3D soundscapes include Torino Winter Olympics, ProFootball Hall of Fame, Great Lakes Children's Museum, NokiaWorld 2008 Barcelona, Denver Museum Nature and Science Gates Planetarium, New York Historical Society, Copenhagen International Theatre, Gallery Rachel Haferkamp Köln, Muu Gallery Helsinki, New Sounds New York, ZHDK Zurich, OKKO Design Stockholm, BAFTA Awards London, Collection of Diana Zlotnick Studio City, CA, as well as Ecsite, AAM, ASTC and IPS conventions. These range from single 8.1 to 64.3 True3D installations, some interactive.

Nick Cave's new novel The Death of Bunny Munro was recording in audiobook format using 3D audio.

# Audio feedback

**Audio feedback** (also known as the **Larsen effect** after the Danish scientist, Søren Larsen, who first discovered its principles) is a special kind of positive feedback which occurs when a sound loop exists between an audio input (for example, a microphone or guitar pickup) and an audio output (for example, a loudspeaker). In this example, a signal received by the microphone is amplified and passed out of the loudspeaker. The sound from the loudspeaker can then be received by the microphone again, amplified further, and then passed out through the loudspeaker again. This is a good example of positive feedback. The frequency of the resulting sound is determined by resonance frequencies in the microphone, amplifier, and loudspeaker, the acoustics of the room, the directional pick-up and emission patterns of the microphone and loudspeaker, and the distance between them.

## *History and theory*

The conditions for feedback follow the Barkhausen stability criterion, namely that, with sufficiently high gain, a stable oscillation can (and usually will) occur in a feedback loop whose frequency is such that the phase delay is an integer multiple of 360 degrees and the gain at that frequency is equal to 1. If the gain is increased until it is greater than 1 for some frequency, then it will be equal to 1 at a nearby frequency, and the system will start to oscillate at that frequency at the merest input excitation, that is to say: sound will be produced without anyone actually playing. This is the principle upon which electronic oscillators are based; although in that case the feedback loop is purely electronic, the principle is the same. If the gain is large, but slightly less than 1, then high-pitched slowly decaying feedback tones will be created, but only with some input sound.

The first academic work on acoustical feedback was done by Dr. C. Paul Boner, PhD., beginning in 1962. Dr. Boner reasoned that when feedback happened, it did so at one precise frequency. He also reasoned that you could stop it by inserting a very narrow notch filter at that frequency in the loudspeaker's signal chain. He worked with Gifford White, founder of White Instruments to hand craft notch filters for specific feedback frequencies in specific rooms. Dr. Boner was responsible for establishing basic theories of acoustic feedback, room-ring modes, and room-sound system equalizing techniques.

## *Prevention*

Most audio feedback results in a high-pitched squealing noise familiar to those who have listened to bands at house parties, and other locations where the sound setup is less than ideal. Usually this occurs when live microphones are pointed in the general direction of the output speakers.

## Distance

To keep the maximal loop gain under 1, the amount of sound energy that is fed back to the microphones has to be as small as possible. As sound pressure falls off with 1/r with respect to the distance r in free space or up to a distance known as reverberation distance in closed spaces (and the energy density with $1/r^2$), it is important to keep the microphones at a large enough distance from the speaker systems.

## Directivity

Additionally, the loudspeakers and microphones should have non-uniform directivity and should stay out of the maximum sensitivity of each other, ideally at a direction of cancellation. Public address speakers often achieve directivity in the mid and treble region (and good efficiency) via horn systems. Sometimes the woofers have a cardioid characteristic.

Professional setups circumvent feedback by placing the main speakers a far distance from the band or artist, and then having several smaller speakers known as *monitors* pointing back at each band member, but in the opposite direction to that in which the microphones are pointing. This allows independent control of the sound pressure levels for the audience and the performers.

If monitors are oriented at 180 degrees to the microphones that are their sources, the microphones should have a cardioid pickup pattern. Super- or hypercardioid patterns are suitable if the monitor speakers are located at a different angle on the back side of the microphones, they also better cancel reverberations coming from elsewhere. Almost all microphones for sound reinforcement are directional.

## Frequency response

Almost always, the natural frequency responses of sound reinforcement systems is not ideally flat. This leads to acoustical feedback at the frequency with the highest loop gain, which may be much higher than the average gain over all frequencies (resonance). It is therefore helpful to apply some form of equalization to reduce the gain of this frequency.

Feedback can be reduced manually by "ringing out" a microphone. The sound engineer can increase the level of a microphone or guitar pickup until feedback occurs. The engineer can then turn down frequency on a band equalizer preventing feedback at that pitch but allowing maximum volume. Professional sound engineers can "ring out" microphones and pick-ups by ear but most use a real time analyzer connected to a microphone to show the ringing frequency.

To avoid feedback, automatic anti-feedback devices can be used. (In the marketplace these go by the name "feedback destroyer" or "feedback eliminator".) Some of these work by shifting the frequency slightly, resulting in a "chirp"-sound instead of a howling sound due to the upshifting the frequency of the feedback. Other devices use sharp notch-filters to filter out offending frequencies. Adaptive algorithms are often used to automatically tune these notch filters.

## *Deliberate uses*

### Early examples in popular music

While audio feedback is usually undesirable, it has entered into musical history as a desired effect beginning in the 1950s with Albert Collins, Johnny "Guitar" Watson and Guitar Slim who all independently recorded and published music featuring that effect. According to Allmusic's Richie Unterberger, the very first use of feedback on a rock record is the song "I Feel Fine" by The Beatles, recorded in 1964. The Who's 1965 hits "Anyway, Anyhow, Anywhere" and "My Generation" featured feedback manipulation by Pete Townshend, with an extended solo in the former and the shaking of his guitar in front of the amplifier to create a throbbing noise in the latter. Canned Heat's Fried Hockey Boogie (off of their 1968 album Boogie with Canned Heat) also featured guitar feedback produced by Henry Vestine during his solo to create a highly amplified distorted boogie style of feedback.

Feedback was used extensively after 1965 by The Monks, Jefferson Airplane, The Velvet Underground and the Grateful Dead, who included in many of their live shows a segment named *Feedback*, a several-minutes long feedback-driven improvisation. Feedback has since become a striking characteristic of rock music, as electric guitar players such as Jeff Beck, Pete Townshend and Jimi Hendrix deliberately induced feedback by holding their guitars close to the amplifier. Lou Reed created his 1975 album *Metal Machine Music* entirely from loops of feedback played at various speeds. A perfect example of feedback can be heard on Jimi Hendrix's performance of Can You See Me? at the Monterey Pop Festival. The entire Guitar solo was created using amplifier feedback.

### Examples in modern classical music

Though closed circuit feedback was a prominent feature in many early experimental electronic music compositions, it was contemporary American composer Robert Ashley who first used acoustic feedback as sound material in his work *The Wolfman* (1964). Steve Reich makes extensive use of audio feedback in his work *Pendulum Music* (1968) by swinging a series of microphones back and forth in front of their corresponding amplifiers.

### Contemporary uses

Audio feedback became a signature feature of many underground rock bands during the 1980s. American noise-rockers Sonic Youth melded the rock-feedback tradition with a compositional/classical approach (notably covering Reich's "Pendulum Music"), and guitarist/producer Steve Albini's group Big Black also worked controlled feedback into the makeup of their songs. With the alternative rock movement of the 1990s, feedback again saw a surge in popular usage by suddenly mainstream acts like Nirvana, the Red Hot Chili Peppers, Rage Against the Machine and The Smashing Pumpkins.

## Marketing

The principle of feedback is used in many guitar sustain devices. Examples include handheld devices like the Ebow, built-in guitar pickups that increase the instrument's sonic sustain, string drivers mounted on a stand such as the Guitar Resonator, and sonic transducers mounted on the head of a guitar. Intended closed-circuit feedback can also be created by an effects unit, such as a delay pedal or effect fed back into a mixing console. The feedback can be controlled by using the fader to determine a volume level.

# Chapter 17

# Automatic Double Tracking

**Automatic double tracking (ADT)** (or alternatively **Artificial Double Tracking**) was an analogue recording technique designed to enhance the sound of voices or instruments during the recording process. It used tape delay to create a delayed copy of an audio signal which was then combined with the original. The effect was intended to simulate the sound of the natural doubling of voices or instruments achieved by doubletracking. The technique was originally developed in 1966 by engineers at Abbey Road Studios in London at the request of The Beatles.

## *Overview*

As early as the 1950s it was discovered that doubletracking the lead vocal in a song gave it a fuller, more appealing sound, especially for singers with weak or light voices. Use of this technique became possible due to the advent of magnetic tape for use in sound recording. Originally, a pair of single-track (or "mono") tape recorders were used to produce the effect; later, multitrack tape machines were used. Early exponents of this technique were Les Paul and Buddy Holly. Before the development of ADT, it was necessary to either record the vocal track twice on two different tracks of a multitrack tape, or to record the vocal first on one tape, then again on a second tape while simultaneously copying the first to the second—a process that could be both tedious and exacting, and might require several takes. After the development of ADT, this process became known as "manual doubletracking".

## *Ken Townsend*

ADT was invented especially for the Beatles during the spring of 1966 by Ken Townsend, a recording engineer employed at EMI's Abbey Road Studios, mainly at the instigation of John Lennon. Lennon hated the tedium of doubletracking during sessions and regularly expressed a desire for a technical alternative.

## *The doubletracking effect*

Doubletracking produces its effect due to it being impossible for a performer to sing or play the same part in exactly the same way twice, meaning that two different performances of the same part have to be recorded so that the fuller, "chorused" effect is created—if one simply plays back two copies of the same performance in perfect sync, the two sound images become one and no doubletracking effect is produced.

Townsend realised that, if two identical performances were played back with one of them slightly out of sync, the sound image would alter and widen, similarly to doubletracking. There was no reliable way that this effect could be achieved by simply copying a vocal track on to another deck and then playing it back with the master slightly out of sync; at the time, there was no technique for synchronising two different tape machines. The end result would be that the second tape deck would gradually drift further and further from the first.

## *Tape delay system*

Instead, Townsend came up with a system using tape delay, after similar principles already in place for echoes applied via tape during a song mixdown. In essence, Townsend's system added a second tape recorder to the regular setup. When mixing a song, its vocal track was routed from the recording head of the multitrack tape, which was before the playback head, and fed to the record head of the second tape recorder. An oscillator was used to vary the speed of the second machine, providing more or less delay depending on how fast or slow the second machine was run relative to the first. This signal was then routed from the playback head of the second machine to a separate fader on the mixer. This allowed the delayed vocal to be combined with the normal vocal, creating the double tracked effect.

## *Use by The Beatles*

The Beatles were thrilled by Townsend's technique and used it throughout the *Revolver* album, and on many of their subsequent recordings. It has been incorrectly claimed that the first use of ADT was on the first half of Lennon's vocal track on "Tomorrow Never Knows", but in fact this vocal track features manual doubletracking. However most of the doubletracked vocals heard on the rest of the album were created using ADT, while the group also used the technique on a number of the instrumental parts to colour the sounds – there is in fact more use of ADT on the mono version of the album than on the more widely known stereo version, with the lead guitar on "Taxman" and the backwards guitar on "I'm Only Sleeping" treated with the effect. ADT could not only be used to create a single double-tracked sound image; but when used on a stereo mix, the effect could be used to "split" the vocal between the two stereo channels, creating the impression of two different vocal parts on either side of the stereo picture. This technique was used on the stereo mixes of "I'm Only Sleeping", "Love You To", "And Your Bird Can Sing", and "Doctor Robert" (on "Here, There and Everywhere", the similar effect heard is actually two different vocals manually double-tracked and panned; on "Eleanor Rigby", the effect is obtained by a combination of manual double-tracking and ADT). This technique could also be applied to instrumental parts as well: on "Love You To", the same use of ADT was applied to the acoustic guitar track, giving the impression of multiple guitars panned left and right.

## *Flanging*

Lennon dubbed the technique "flanging" after producer George Martin jokingly told him it was produced using a "double-bifurcated sploshing flange". Only years later did Martin learn that another technique, also called flanging, was already in use (it is rumoured that recording engineer Joe Meek was the inventor of flanging in the 1950s; however, it is more often attributed to Townsend. Whether that attribution is valid or not is contested due to the close contact between George Martin, who misnomered ADT as flanging, and Townsend.) The term referred to an engineer's alternately pressing and releasing his finger against the flange (rim) of the supply reel on one of two synchronized tape machines as the same audio signal was combined and transferred to a third machine, slightly slowing the machine then allowing it to come back up to speed and in sync with the other, applying a "swooshing" comb filtering effect to the combined audio signal. Alternatively, the engineer could press the flange of one supply reel then the other to achieve a fuller effect.

An additional explanation for the pedigree of flanging has it named after Fred Flange, a pseudonym given to Matt Monro by Peter Sellers, who used a Monro recording to open his 1959 Sinatra parody album *Songs for Swingin' Sellers*. The album was produced by Martin, and presumably the connection with flanging comes from Monro's mimicking (double-tracking) Sinatra. Engineers at Abbey Road realised that the technique they had developed needed a proper technical name and eventually christened it ADT, short for "Artificial Double Tracking", although elsewhere the term "Automatic Double Tracking" became more common.

## *ADT versus manual doubletracking*

Townsend's process succeeded in simulating manual doubletracking quite effectively; however, attentive listeners can often tell the difference between ADT and "real" doubletracking, with the former having a synthetic quality to it and having none of the audible differences between the vocal tracks frequently present in the latter. Over the years, many artists, including the Beatles, continued to use both manual doubletracking, ADT, or a combination of both in different circumstances depending on the effects they wished to achieve, with each technique thought to have certain unique qualities of its own.

The Beatles used ADT widely in conjunction with manual double-tracking on all their subsequent albums, with the exception of *Let It Be*, which was initially intended to be an "honest" album utilising no technical artifice (ADT can still be heard on the finished album, however, due to Phil Spector treating a Hammond organ part with it on his mix of the title track). Some notable examples of ADT use by the Beatles in the years following *Revolver* include "Fixing a Hole" (where the bass guitar part is treated with ADT in an attempt to simulate a "fretless tone"), "Within You, Without You" (on which ADT is supposedly used on almost every vocal and instrumental part on the track), "I Am the Walrus" (which uses ADT in conjunction with equalisation to help simulate a "fake stereo" effect on the second half of the stereo mix, which was sourced from the mono mix, by splitting the entire mix between the channels), and the unusually wide ADT used on the lead vocal tracks on "Being for the Benefit of Mr. Kite" and "Blue Jay Way".

## *Other users of ADT*

Townsend's technique, and minor variations on it, quickly caught on and almost immediately began to be used by other artists and record producers. Former Beatles engineer Norman Smith used ADT extensively on Pink Floyd's debut album *The Piper at the Gates of Dawn*, recorded at Abbey Road in 1967. As well as using it for more conventional simulated doubletracking, Smith made much use of the technique to split Syd Barrett's vocals between the stereo channels. In some cases, Smith (or possibly Barrett himself) used such extraordinarily wide ADT in this way as to give the slightly disorientating impression of not so much doubletracking but two quite separate voices on either channel wildly out of time with each other – the best example of this is perhaps on "Bike". Similar effects were later used on some of Barrett's solo works, perhaps indicating his fondness for this unusual use of ADT. Pink Floyd themselves continued to use ADT on most, if not all, of their subsequent albums up until the 1980s, with one notable use being on "Alan's Psychedelic Breakfast", where a part of the drum track is treated with ADT.

In the U.S., Simon and Garfunkel began to use ADT on stereo mixes of their songs to split vocal tracks between the channels, examples of which include "Mrs. Robinson" and "Cecilla".

Gary Kellgren, Jimi Hendrix's engineer, made extensive use of ADT on all of Jimi's albums, frequently using the technique to split vocal, guitar, and even drum parts between the stereo channels.

## *Psychedelic music*

With the rise of psychedelic music, many artists used variations on Townsend's technique to create the "flanging" effect mentioned above, adding a slightly disorientating "swooshing" quality to instruments and voices (although in practice this effect is actually more similar to what today is called "phasing" rather than "flanging"). The Beatles themselves used this effect on "Lucy in the Sky with Diamonds" and more prominently on "Blue Jay Way". A notable example of this technique is "Itchycoo Park" by the Small Faces, where the effect is prominent almost throughout the entire track, particularly on the vocals, drums and cymbals during the chorus. Hendrix also utilised this technique extensively. An example of an ADT variation being used to create an effect more similar to what is considered "flanging" today (rather than phasing) is on the Beatles' *White Album* tracks "Cry Baby Cry" and "While My Guitar Gently Weeps" – the former features "flanging" on the acoustic guitar part, the latter on the lead guitar.

### Doubling echo

A similar technique to ADT is doubling echo, which uses short delays to mimic the doubletracking effect. Many effects units were developed to produce similar sounds, such as chorus, flangers, and phasers, all of which use an oscillating delay (or, in the phaser, a variable phase network).

### Arrival of digital technology

ADT became the standard recording studio technique for simulating doubletracking throughout the late 1960s and 1970s until the arrival of digital technology in the 1980s (although not all engineers could apparently figure out how to reproduce the effect successfully, with Jack Douglas recalling that he was at a loss when John Lennon asked him to use ADT on his vocals during a recording session in 1980 but was unable to adequately explain to his producer how the tape decks should be set up to create the effect). With the advent of digital recording, tape- and analog-based delay methods have not been much used, though many of these analog techniques are frequently emulated using comparable digital techniques, or in some cases plugins which are used to extend the capabilities of a Digital audio workstation. Although using digital delay to simulate double-tracking produces a very similar effect to ADT, some claim to be able to hear the difference between the two (certainly one can tell the difference between digital delay and manual double-tracking, as was the case with ADT in previous years – manual double-tracking continues to be used by a number of artists). Some musicians and engineers may casually use the term ADT to refer to any form of simulated doubletracking, including digital delay used in this manner, although strictly speaking the term should refer to the analogue technique. One of the very few examples of ADT being used in recent times is on the Beatles' *Anthology* albums from the mid-1990s, on which George Martin and Geoff Emerick decided to revive the analogue technique rather than simply use the modern digital alternatives in order to achieve a more authentic sound, feeling that ADT produced a warmer, less synthetic sound than digital delay and the latter would be inappropriate for use on recordings made on analogue equipment in the 1960s.

## Chapter 18

# Delay (Audio Effect)

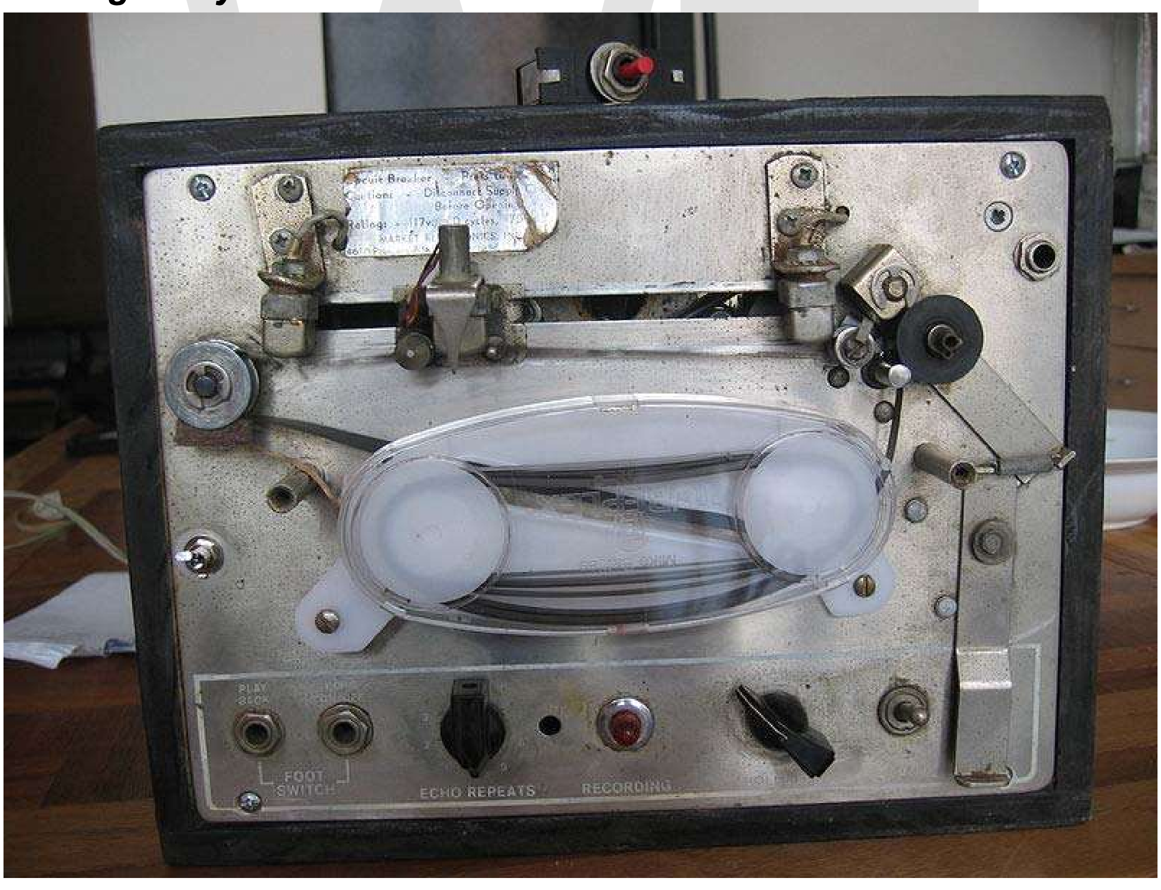
Various kind of delay effect units

**Delay** is an audio effect which records an input signal to an audio storage medium, and then plays it back after a period of time. The delayed signal may either be played back multiple times, or played back into the recording again, to create the sound of a repeating, decaying echo.

## *Early delay systems*

The first delay effects were achieved using tape loops improvised on reel-to-reel magnetic recording systems. By shortening or lengthening the loop of tape and adjusting the read and write heads, the nature of the delayed echo could be controlled. This technique was most common among early composers of Musique concrète (Pierre Schaeffer), and composers such as Karlheinz Stockhausen, who had sometimes devised elaborate systems involving long tapes and multiple recorders and playback systems, collectively processing the input of a live performer or ensemble. Audio engineers working in popular music quickly adapted similar techniques, to augment their use of plate reverb and other studio technologies designed to simulate natural echo.

## *Analog delay*




Echoplex EP-2

Before the invention of audio delay technology, music employing a delayed echo had to be recorded in a naturally reverberant space, often an inconvenience for musicians and engineers. The popularity of an easy-to-implement real-time echo effect led to the production of systems offering an all-in-one effects unit that could be adjusted to produce echoes of any interval or amplitude. The presence of multiple "taps" (playback heads) made it possible to have delays at varying rhythmic intervals; this allowed musicians an additional means of expression over natural periodic echoes.

Many delay processors based on analog tape recording, such as Ray Butts' Echosonic (1952), Mike Battle's Echoplex (1959), or the Roland Space Echo (1973), used magnetic tape as their recording and playback medium. Electric motors guided a tape loop through a device with a variety of mechanisms allowing modification of the effect's parameters. In the case of the popular Echoplex EP-2, the play head was fixed, while a combination record and erase head was mounted on a slide, thus the delay time of the echo was adjusted by changing the distance between the record and play heads. In the Space Echo, all of the heads are fixed, but the speed of the tape could be adjusted, changing the delay time. Thin magnetic tape was not entirely suited for continuous operation, however, so the tape loop had to be replaced from time to time to maintain the audio fidelity of the processed sounds.

The Binson Echorec, another popular unit, used a rotating magnetic drum as its storage medium. This provided an advantage over tape, as the durable drums were able to last for many years with little deterioration in the audio quality. Other devices used spinning magnetic discs, not entirely unlike those used in modern hard disk drives.

Robert Fripp used two Revox reel to reel tape recorders to achieve very long delay times for solo guitar performance. He dubbed this technology "Frippertronics", and used it in a number of recordings. John Martyn is widely acclaimed as the pioneer of the echoplex. Perhaps the earliest indication of his use can be heard on the songs Would You Believe Me and The Ocean on the album *Stormbringer* released in February 1970. This was a first taste of things to come from Martyn’s interest in electronics and the boundless possibilities of electric music. *Glistening Glyndebourne* on the album *Bless The Weather* (1971) showcased his developing technique of playing acoustic guitar through the echoplex to stunning effect. He later went on to experiment with a fuzz box, a volume/wah wah pedal and the echoplex on highly acclaimed *Inside Out* (1973) and *One World* (1977). Martyn is cited as an inspiration by many musicians including U2's The Edge.

Often incorporating vacuum tube-based electronics, surviving analog delay units are sought by modern musicians who wish to employ some of the timbres achievable with this technology.

Solid state delay units using analog bucket brigade delay circuits became available in the 1970s and were briefly a mainstream alternative to tape echo. Though solid state analog delays are less flexible than digital delays and generally have shorter delay times, several classic models such as the discontinued Boss DM-2 are still sought after for their "warmer", more natural echo quality and progressively decaying echos. Additionally, several companies make new analog delays. Old delay systems like the Roland Space Echo and Echoplex are still highly regarded and used with some frequency by modern bands.

## *Digital delay*


Ibanez DE-7 delay pedal

The availability of inexpensive digital signal processing electronics in the late 1970s and 1980s led to the development of the first digital delay effects. Initially, they were only available in expensive rack mounted units but eventually as costs came down and the electronics grew smaller, they became available in the form of foot pedals. The first digital delay offered in a pedal was the Boss DD-2 in 1984. Rack mounted delay units evolved into digital reverb units and on to digital multieffects units capable of more sophisticated effects than pure delay, such as reverb and Audio timescale-pitch modification effects.

Digital delay systems function by sampling the input signal through an analog-to-digital converter, after which the signal is passed through a series of digital signal processors that record it into a storage buffer, and then play back the stored audio based on parameters set by the user. The delayed ("wet") output may be mixed with the unmodified ("dry") signal after, or before, it is sent to a digital-to-analog converter for output.

Many modern digital delays present an extensive array of options, including a control over the time before playback of the delayed signal. Most also allow the user to select the overall level of the processed signal in relation to the unmodified one, or the level at which the delayed signal is fed back into the buffer, to be repeated again. Some systems today allow more exotic controls, such as the ability to add an audio filter, or to play back the buffer's contents in reverse.

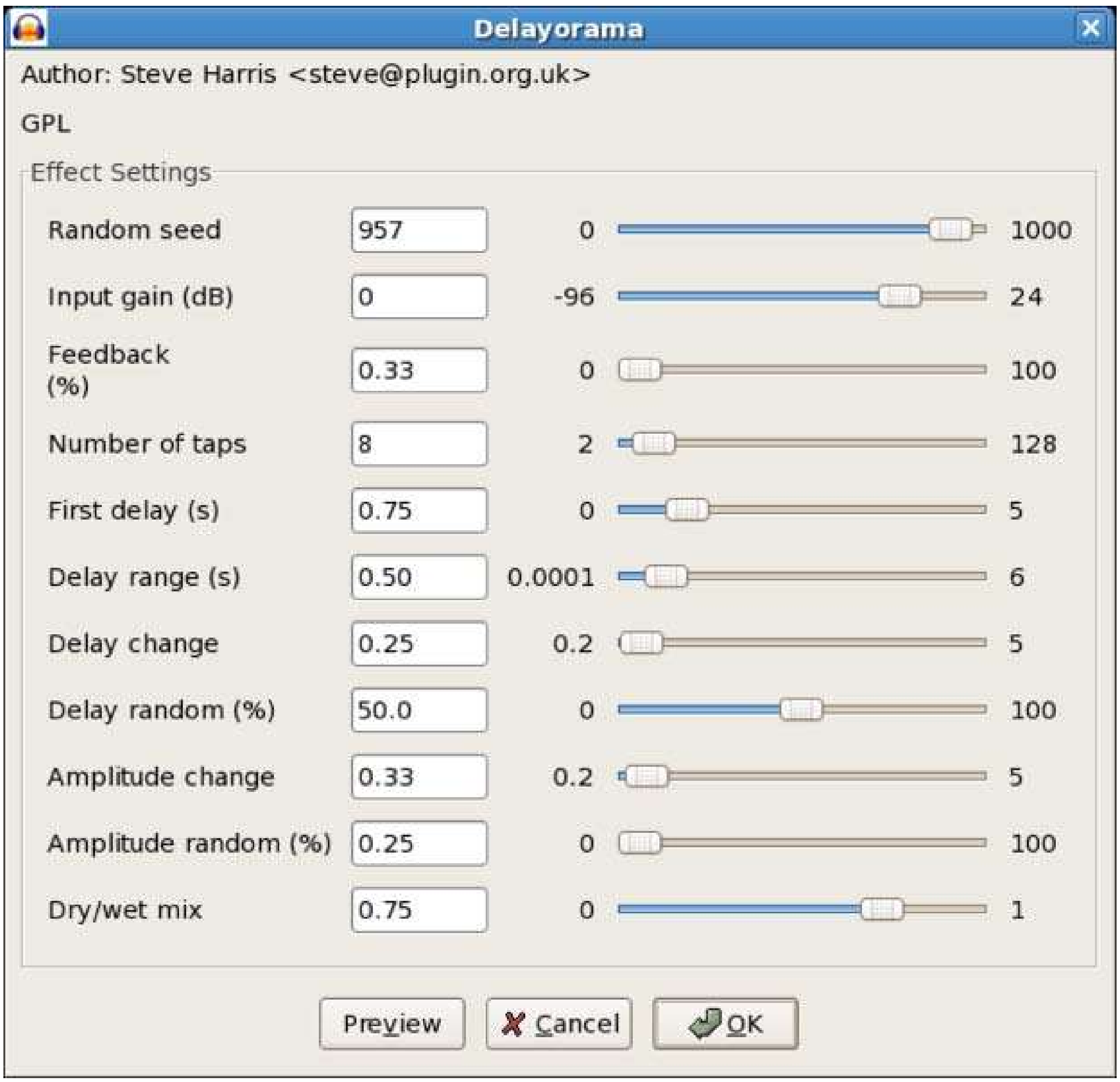
As digital memory became cheaper in the 80s, units like Lexicon PCM84, Roland SDE3000, TC Electronic 2290 offered above 3 seconds delay time, enough to create background loops, rhythms and phrases. The 2290 was upgradable to 32 seconds, and Electro-Harmonix offered a 16-second delay and looping machine.

A sample setup of loop music

### *From delay to loop*

While the mentioned long delay units were a bit clumsy to create loops, the Paradis LOOP Delay, created in 1992, was the first unit with dedicated looping functions Record, Overdub, Multiply, Insert, Replace, ... Gibson manufactured a slightly improved version as Echoplex Digital Pro until 2006. Its software Aurisis LOOP is also the last loop tool based on a continuous memory structure as used by tape and digital delays. Most following loopers repeat samples and thus have little in common with a digital delay, the exceptions being Maneco's early loopers, the Boss DD-20 in digital delay mode, and the Pigtronix echolution.

### *Computer software*




Steve Harris' Delayorama software

A natural development from digital delay-processing hardware was the appearance of software-based delay systems. In large part, this coincided with the popularity of both professional and consumer audio editing software. Software delays, in many cases, offer much greater flexibility than even the most recent digital hardware delays. Abundant system memory on modern personal computers offers practically limitless storage for the audio buffer, and the natural efficiency of audio delay algorithms has made the implementation trivial for delays offering shifting or random delay times, or the insertion of other audio effects during the feedback process. Many authors of software plugins have added functionality to emulate the sounds of the earlier analog units.

Software-based delays are most popular today among musicians in electronic genres or those who prefer to audition the effect in a digital audio editing and mixing environment.

### *Uses*

In popular and electronic music, electric guitarists use delay to produce densely overlaid textures of notes with rhythms complementary to the music. Vocalists and instrumentalists use it to add a dense or ethereal quality to their singing or playing. Extremely long delays of 10 seconds or more are often used to create loops of a whole musical phrase.

*Echoplex* is a term often applied to the use of multiple echoes which recur in approximate synchronization with a musical rhythm, so that the notes played combine and recombine in interesting ways. In fact, it was the name of a particular delay unit, the Maestro Echoplex.

*Doubling echo* is produced by adding short-range delay to a recorded sound. Delays of thirty to fifty milliseconds are the most common; longer delay times become slapback echo. Mixing the original and delayed sounds creates an effect similar to doubletracking, or unison performance.

*Slapback echo* uses a longer delay time (seventy-five to 250 milliseconds), with little or no feedback. The effect is characteristic of vocals on 1950s rock-n-roll records, particularly those issued by Sun. It is also sometimes used on instruments, particularly drums and percussion. Slapback was often produced by refeeding the output signal from the playback head of a tape recorder to its record head, the physical space between heads, the speed of the tape, and the chosen volume being the main controlling factors. Analog and later digital delay machines also easily produced the effect.

*Flanging*, *chorus* and *reverberation* (reverb) are all delay-based sound effects. With flanging and chorus, the delay time is very short and usually modulated. With reverberation there are multiple delays and feedback so that individual echoes are blurred together, recreating the sound of an acoustic space.

## Straight delay

In audio reinforcement, a straight delay is used to compensate for the passage of sound through the air. Unlike audio delay effects devices, straight delay is not mixed back in with the original signal. The delayed signal alone is sent to loudspeakers so that the speakers reinforce the stage sound at the same time or slightly later than the acoustic sound from the stage, approximately 1 millisecond of straight delay per foot of air or 3 milliseconds per meter, depending on the air temperature's effect on the speed of sound. Because of the Haas effect, this technique allows audio engineers to use additional speaker systems placed away from the stage and still give the illusion that all sound originates from the stage. The purpose is to deliver sufficient sound volume to the back of the venue without resorting to excessive sound volumes near the front.

Straight delay is also used in audio to video synchronization to align sound with visual media if the visual source is delayed. Visual media can become delayed by a number of mechanisms, in which case the associated audio should be delayed to match.

# Chapter 19

# Distortion (Music)


A well-used "Turbo Distortion" guitar effect pedal made by Boss

**Distortion**, **overdrive** and **fuzz** are effects applied to the electric guitar, the electric bass, and other amplified instruments such as the Hammond organ, synthesizers, harmonica and even vocals by electronically clipping the signal. This adds sustain and additional harmonics to the signal.

The most subtle types of distortion add a "warm" thickness to the original tone, used in electric blues, for instance, while more extreme types range from the noisy, buzzy sound of a 1960s fuzzbox to the screaming, "bite", "grit", and "crunch" of a late 1980s thrash-style distortion pedal and the hard-edged distortion featured in noise music, hardcore punk, industrial, grunge, and metal. A **fuzzbox** (or **fuzz box**) boosts and clips the signal sufficiently to turn a standard sine wave into a waveform much closer to a square wave. This gives a much more distorted and synthetic sound than a standard distortion or overdrive. Fuzz boxes also tend to have lower mid frequencies than other distortion types.

Distortion can be produced by many components of an instrument's signal path, including effects pedals, the pre-amplifier, power amplifier, speakers, or more recently, digital amplifier modelling devices and software. The distortion in guitar effects pedals such as the Ibanez/Maxon TS-9 and 808 Tube Screamer is produced by transistor or diode clipping. Many players use a combination of these to obtain their "signature" tone.

## *History*

In the early days of guitar amplification, amplifiers were primitive and low-fidelity, and distortion was inherent in the signal chain. Early examples of distortion were often the result of accidental damage to a guitar amplifier, its vacuum valves or loudspeakers. The earliest uses of intentional distortion that have been recorded were achieved through "doctoring" amplifiers and speakers. Guitarists would use a razor blade, screwdriver or pencil to poke holes into their speaker cones to create a distorted sound.

One notable example was Link Wray, who dislodged a tube by accident, and then took to doing so as a habit to get a noisy, dirty sound for his solos. During the recording of "Rocket 88", one of the early rock and roll songs, Ike Turner and the Kings of Rhythm guitarist Willie Kizart used an amplifier that had been damaged in transit, resulting in an early recorded example of guitar distortion. For the recording of "The Train Kept A-Rollin'" (1956) by the Johnny Burnette Trio, a valve fell out of the amplifier during a live performance. When a reviewer then raved about the crazy new sound, Paul Burlison, the guitar player, used the same tone in the recording studio.

Willie Johnson's playing on Howlin' Wolf's Memphis recordings of 1951-2 is marked by a consistent use of deliberate distortion, creating a raucous, menacing sound that complements Howlin' Wolf's singing. Another early user of valve overdrive was Chuck Berry who at the start of his career played through small valve amplifiers, the only ones he could afford. Because of their low output they were easy to overdrive, giving his guitar tone the warm sound heard on his first hit "Maybellene". On later recordings he was able to afford better and larger amps and consequently his tone became cleaner.

Leo Fender of Fender guitars and amplifiers observed these trends and engineered many of his amplifiers to "compress" and/or "overdrive" slightly without drastically distorting the signal. The early Fender "Tweed" and "Blackface" amplifiers are considered a good example of clean electric guitar tone. Many later amplifiers are based on these designs.

Significantly, Jim Marshall of Marshall Amplifiers copied the Fender Bassman using parts available in the United Kingdom, creating an amplifier with significant overdrive that quickly caught on in the local music scene and laid the foundation for the powerful, thick "Marshall Sound" that can be heard on so many early Hard Rock albums.

Nashville session musician Grady Martin produced the fuzz sound in 1961 during a recording session for Marty Robbins' "Don't Worry" due to a faulty recording console preamplifier circuit. In 1962, The Ventures, having heard the guitar tone on "Don't Worry", asked friend Red Rhodes, a steel player and electronics wizard, how they could reproduce the sound. A few months later, Rhodes presented them with a custom fuzz box, reportedly the first, which The Ventures used to record "2000 Pound Bee." The song charted in December 1962 and is identified by multiple sources, including *The VH-1 Music First Rock Stars Encyclopedia*, as the first single to use actual guitar fuzz box (the story was in the April 2007 issue of *Guitar Buyer* magazine in an article titled, "Caught By The Fuzz"). Fuzzboxes became popular as a much easier way to create a distorted sound.

Distortion gained wider popularity after a distorted sound was popularised by Dave Davies of The Kinks who played through a small amp whose speaker cone had been slashed with a razor blade, used to record "You Really Got Me", the band's first number one single and the first popular rock & roll song using a distorted power chord riff.

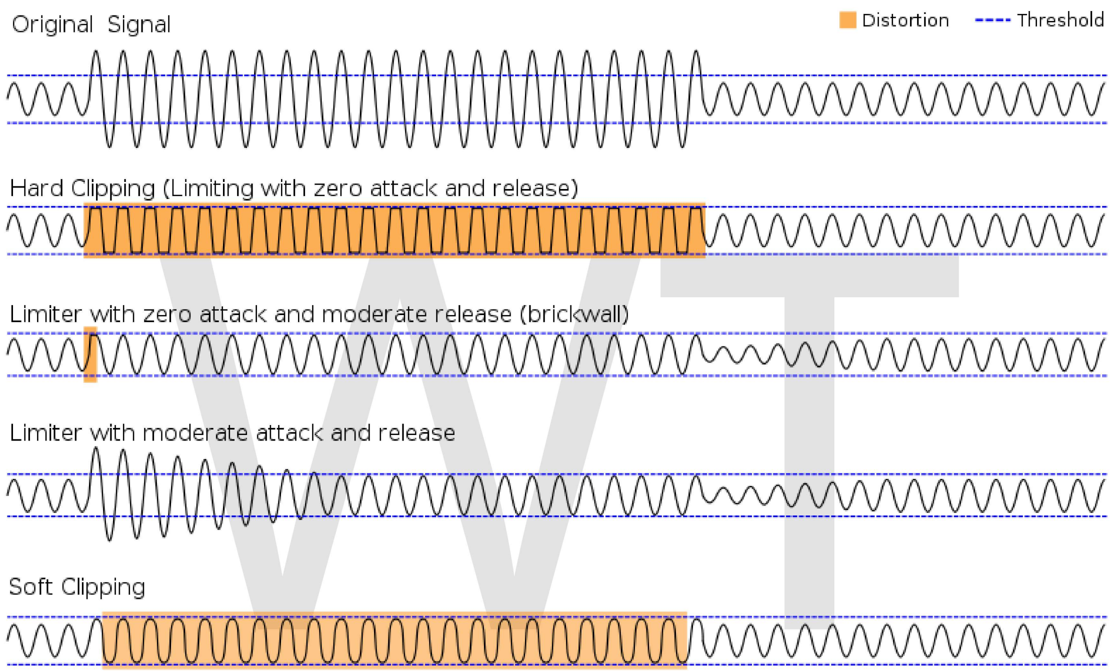
The first purpose-designed commercial distortion circuit was marketed by Maestro as the "Fuzz Tone" Model FZ-1. In May 1965 Keith Richards used a Gibson Maestro Fuzz-Tone to record "(I Can't Get No) Satisfaction". The song's success so boosted sales of the device that all available stock had sold out by the end of 1965. Jimi Hendrix is particularly associated with the increased use of such effects, many designed or modified by guitar technologist Roger Mayer.

Other examples of fuzzboxes include the highly-sought Mosrite FuzzRITE, the Fuzz Face (originally made by the Arbiter Group) used by Jimi Hendrix, the Big Muff Pi (made by Electro-Harmonix) and the Vox Tone Bender, used by Paul McCartney on George Harrison's composition Think for Yourself, and many other Beatles recordings. Pete Townshend used a Univox Super Fuzz pedal starting from 1968 and used on many recordings and stage shows by The Who (being his only pedal for concerts from 1968–1978).

Early fuzzboxes used germanium transistors but by the end of the 1960s, these were replaced by silicon transistors, which are more uniform in performance and less affected by changes in temperature and by source impedance) - low impedance signals that pass through germanium-equipped fuzzes tend to suffer from a pronounced drop in volume and bass response. In the 2000s, many boutique guitar effects builders offer fuzz pedals with germanium transistors for a "retro" sound. Some units employ both silicon and germanium transistors.

## *Physics of clipping*

As clipping is a non-linear process, intermodulation will occur, leading to the generation of an output signal rich in extra harmonics of the input signal. Intermodulation distortion also produces frequency components at the various sums and differences of the frequency components of the input signal. In general, these components will not be harmonically related to the input signal, leading to dissonance. To reduce unwanted dissonance, simple power chords (root, fifth, and octave) are often used when using fuzzboxes, rather than triads (root, third, and fifth) or four-note chords (root, third, fifth, and seventh).




Waveform plot showing the different types of clipping. Valve overdrive is a form of soft limiting, while transistor clipping or extremely overdriven valves resemble hard clipping.

Literally, the word distortion refers to any aberration of the waveform of an electronic circuit's output signal from its input signal. In the context of musical instrument amplification, it refers to various forms of clipping, which is the truncation of the part of an input signal that exceeds certain voltage limits. Because both valves and transistors behave linearly within a certain voltage region, distortion circuits are finely tuned so that the average signal peak just barely pushes the circuit into the clipping region, resulting in the softest clip and the least harsh distortion. Because of this, as the guitar strings are plucked harder, the amount of distortion and the resulting volume both increase, and lighter plucking cleans-up the sound.

## Valve overdrive

Before the widespread adoption of the transistor, the traditional way to create gain (amplification) and distortion was through vacuum valves (called "tubes" in North America). A vacuum valve has a maximum input voltage determined by its bias and a minimum input voltage determined by its supply voltage. When any part of the input waveform approaches these limits, the valve's amplification becomes less linear, meaning that smaller voltages get amplified more than the large ones. This causes the peaks of the output waveform to be compressed, resulting in a waveform that looks "squashed". This is known as "soft clipping", and emphasizes the lower rather than higher-order harmonics that add to the warmth and richness of the guitar's tone. If the valve is driven harder, the compression becomes more extreme and the peaks of the waveforms are clipped. This adds additional odd-order harmonics, creating a "dirty" or "gritty" tone. The harmonics are entirely odd for a symmetrically-distorted signal. Asymmetrical distortion is also typical of valve amplifiers, where the positive swing of the AC signal may be distorted differently than the negative swing. Asymmetrical distortion results in even harmonics, which add a unique tone to the valve-generated distortion.

Valve distortion is commonly referred to as overdrive, as it is attained by driving the valves in an amplifier at a higher level than can be handled cleanly. Multiple stages of valve gain/clipping can be "cascaded" to produce a thicker and more complex distortion sound. In some modern valve effects, the "dirty" or "gritty" tone is actually achieved not by high voltage, but by running the circuit at voltages that are too low for the circuit components, resulting in greater non-linearity and distortion. These designs are referred to as "starved plate" configurations, and result in an "amp death" sound.



## Transistor clipping

Transistor clipping stages, on the other hand, behave far more linearly within their operating regions, and thus faithfully amplify the instrument's signal until the input voltage falls outside its operating region, at which point the signal is clipped without compression, known as "hard clipping" or limiting. This type of distortion tends to produce odd harmonics at higher orders due to the sharpness of the 'corners' put into the time domain signal. Electronically, this is usually achieved by either amplifying the signal to a point where it must be clipped to the supply rails, or by clipping the signal across diodes. Many solid state distortion devices attempt to emulate the sound of overdriven vacuum valves.

## *Approaches*

Guitar distortion can be produced by many components of the guitar's signal path, including effects pedals, the pre-amplifier, power amplifier, and speakers. Many players use a combination of these to obtain their "signature" tone.

## Overdrive/distortion pedals


The Ibanez TS-9 Tube Screamer, a popular overdrive pedal

Because they are often designed to operate with low voltages (such as 9 volt batteries), overdrive and distortion pedals typically use transistors to generate distortion. Classic examples include the Ibanez Tube Screamer and the Electro-Harmonix Big Muff. A few more modern effects pedals incorporate valves; usually these still run at voltages that are below the valve's design specifications, resulting in a "starved plate" configuration that some people feel generates harsh and buzzy distortion. Distortion pedals usually also provide signal gain, which can be used to drive the input stage of the pre-amplifier harder, resulting in further distortion and, in some cases, higher volume.

## Pre-amplifier distortion

The pre-amplifier section of a guitar amplifier serves to amplify a weak instrument signal to a level that can drive the power amplifier. It often also contains circuitry to shape the tone of the instrument, including equalization and gain controls. Often multiple cascading gain/clipping stages are employed to generate distortion. Because the first component in a valve amplifier is a valve gain stage, the output level of the preceding elements of the signal chain has a strong influence on the distortion created by that stage. The output level of the guitar's pickups, the setting of the guitar's volume knob, how hard the strings are plucked, and the use of volume-boosting effects pedals can drive this stage harder and create more distortion.

During the 1980s and 1990s, many amps featured a "master volume" control, essentially an adjustable attenuator between the preamp section and the power amp that conveniently enables the generation of high distortion levels in the guitar amp's preamp section while diverting most of the resulting signal away from the power valves, keeping the output volume at manageable levels. However, this also results in the power valves being operated well within their linear region, reducing the distortion that they add to the output signal.

Solid-state gain/clipping stages are also employed in many amplifiers. Some amplifiers (notably the Marshall JCM900) utilize hybrid designs that employ both valve and solid-state components.

### Power amplifier distortion

A pair of 6L6GC power valves, often used in American-made amplifiers

Power valves can be overdriven in the same way that pre-amplifier valves can, but because these valves are designed to output more power, the distortion and character they add to the guitar's tone is unique. During the 1960s to early 1970s, distortion was primarily created by overdriving the power valves. Because they have become accustomed to this sound, many guitar players favour this type of distortion, and thus set their amps to maximum levels in order to drive the power section hard. Many valve-based amplifiers in common use have a balanced configuration in their power section, with matched pairs of tubes driving the output transformer in opposite phases, power amplifier distortion is entirely symmetric, generating only odd-order harmonics and does not generate even-order harmonics (unless the tubes are not a matched pair are poorly matched to each other and so have very different transfer characteristics).

Because driving the power valves this hard also means maximum volume, which can be difficult to manage in a small recording or rehearsal space, many solutions have emerged that in some way divert some of this power valve output from the speakers, allow the player to generate power valve distortion without excessive volume. These include built-in or separate power attenuators and power-supply-based power attenuation. Lower-power valve amps (such as a quarter-watt or less), speaker isolation boxes, and low-efficiency guitar speakers are also used to tame the volume.

Although traditional amplifiers were complete circuits including both preamp and power amp, power-valve distortion can also be produced in a dedicated rackmount valve power amp. A modular rackmount setup often involves a rackmount preamp, a rackmount valve power amp, and a rackmount dummy load to attenuate the output to desired volume levels. Some effects pedals internally produce power-valve distortion, including an optional dummy load for use as a power-valve distortion pedal. Such effects units can use a preamp valve such as the 12AX7 in a power-valve circuit configuration (as in the Stephenson's Stage Hog), or use a conventional power valve, such as the EL84 (as in the H&K Crunch Master compact tabletop unit). However, because these are usually placed before the pre-amplifier in the signal chain, they contribute to the overall tone in a different way.

A Direct Inject signal can capture the power-tube distortion sound without the direct coloration of a guitar speaker and microphone. This DI signal can be blended with a miked guitar speaker, with the DI providing a more present, immediate, bright sound, and the miked guitar speaker providing a colored, remote, darker sound. The DI signal can be obtained from a DI jack on the guitar amp, or from the Line Out jack of a power attenuator.

## Output transformer distortion

The output transformer sits between the power valves and the speaker, serving to match impedance and voltage. When a transformer's ferromagnetic core becomes electromagnetically saturated, it will clip symmetrically in a correctly-operating amplifier, since there is negligible net DC current in either coil. This adds additional odd-order distortion to the signal delivered to the speakers.

## Power supply "sag"

Early valve amplifiers usually used unregulated power supplies. This was due to the high cost associated with high-quality high-voltage power supplies. The typical anode supply was simply a rectifier, an inductor and a filter capacitor. When the valve amplifier was operated at high volume, the power supply voltage would dip, reducing power output and causing signal attenuation and compression. This dipping effect is known as "sag", and is sought after by some electric guitarists. Sag only occurs in Class AB amplifiers. This is because, technically, sag results from more current being drawn from the power supply, causing a greater voltage drop over the rectifier valve. In a Class A amplifier, current draw is constant, so sag does not occur.

As this effect is more pronounced with higher input signals, the harder "attack" of a note will be compressed more heavily than the lower-voltage "decay", making the latter seem louder and thereby improving sustain. Additionally, because the level of compression is affected by input volume, the player can control it via their playing intensity: playing harder results in more compression or "sag". In contrast, modern amplifiers often use high-quality, well-regulated power supplies. In theory, these keep the supply voltage constant, but in reality there is still some small variation, largely due to resistive losses in the cabling from the power supply to the gain stage.

## Speaker distortion

Guitar loudspeakers are designed differently from high fidelity stereo speakers or PA system speakers. While hi-fi and PA speakers are designed to reproduce the sound with as little distortion as possible, guitar speakers are usually designed so that they will shape or colour the tone of the guitar, either by enhancing some frequencies or attenuating unwanted frequencies. As well, when the power delivered to a guitar speaker approaches its maximum rated power, the speaker's performance becomes less linear, causing the speaker to "break up", adding further distortion and colouration to the signal. Some speakers are designed to have lots of clean headroom, while others are designed to break up early to deliver grit and growl.

## Amp modeling for distortion emulation

Guitar amp modeling devices and software can reproduce various guitar-specific distortion qualities that are associated with a range of popular "stomp box" pedals and amplifiers. Amp modeling devices typically use digital signal processing to recreate the sound of plugging into analogue pedals and over driven valve amplifiers. The most sophisticated devices allow the user to customize the simulated results of using different preamp, power-tube, speaker distortion, speaker cabinet, and microphone placement combinations. For example, a guitarist using a small amp modeling pedal could simulate the sound of plugging their electric guitar into a heavy vintage valve amplifier and a stack of 8 X 10" speaker cabinets. Some modeling devices allow even more detailed simulation, such as the different tonal effect that would occur from miking the speaker cabinets with a cardioid microphone or a ribbon mic.

Amplifier and distortion modeling can be accomplished by using different methods: real-time software running on a computer: hardware such as a compact pedal, oversize pedal, rack mount processor, desktop or floor processor; guitar amp heads, including hybrid valve amps that use both analog and digital technology. As sound of the highest-end modeling devices can be very convincing, these processors are widely used for live performances, because it reduces the amount of heavy, vintage amplifiers that guitarists have to transport. Some smaller independent recording studios use modeling software or processors, because they allow performers to produce widely-used "classic tones" without having to purchase or rent expensive vintage equipment or record at high volume levels. Digital modeling devices may not be able to recreate all the subtle aspects of the sound of vintage, over driven valve amplifiers, because this sound is the product of a range of non-linear and random factors, ranging from the heat of the vacuum valves to the age and condition of the speakers. As a result, professional musicians tend to use actual vintage valve amps for recordings, because the recorded sound will have to stand up to greater scrutiny from listeners and critics. Another aspect in this discussion is the latency of the digital modelling device output signals which can influence the players feeling while playing.

### *Voicing with equalization*

Rock guitar distortion is obtained and shaped at various points in the signal processing chain, including multiple stages of preamp distortion, power valve distortion, output and power transformer distortion, and guitar speaker distortion. Much of the distortion character or voicing is controlled by the frequency response curve before and after each distortion stage. This dependency of distortion voicing on frequency response can be heard in the effect that a wah pedal has on the subsequent distortion stage, or by using an EQ pedal to favor the bass or treble components of the guitar pickup signal prior to the first distortion stage. Some guitarists place an equalizer pedal after the distortion effect, to emphasize or de-emphasize different frequencies and create different tonal coloration.

A guitar amplifier's tone controls shape a different power-valve distortion voicing if the tone controls are set to emphasize the bass or treble. Extreme settings are most popular in heavy metal. Increasing the bass and treble while reducing or eliminating the centre midrange (750 Hz) results in what is popularly known as a "scooped" sound (since the midrange frequencies are "scooped" out). James Hetfield of Metallica used this aggressive-sounding tone on many songs on Metallica's first four studio albums. Conversely, decreasing the bass while increasing the midrange and treble creates a punchy, harsher sound. Kerry King and Jeff Hanneman of Slayer have both used midrange-heavy tones since the mid-1980s.