Digital Audio Technology Seminar Notes

Douglas Dixon
Manifest Technology® LLC

May 2005

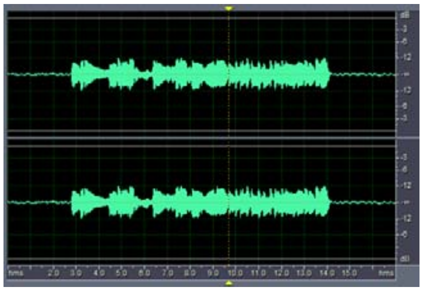
www.manifest-tech.com

Digital Audio

- **Audio Characteristics**
  - Synthesized / Sampled Waveforms
  - Sample Rate, Sample Size, Channels
- **Audio Compression**
  - Data Size, File Formats
- **Audio Processing**
  - Sample editing
  - Multi-track mixing and effects
  - Soundtrack generation, Looping
  - Composing

Audio Characteristics

- **Synthesized – MIDI**
- **Sampled – WAVE**



(Adobe Audition)

- **Waveforms**
  - Amplitude, Frequency
  - Loudness – Dynamic Range – dB
  - Analog to Digital
- **Sample Rate – Samples per second (8000, 44100)**
- **Sample Size – Bits per sample (8, 16)**
- **Channels – Number (mono, stereo, 5.1)**

Synthesized Audio – MIDI

- **MIDI – Musical Instrument Digital Interface**
  - Instrumental music, Orchestrate on synthesized instruments
  - Digital orchestra: Instruments, Notes, Duration
- **Process**
  - MIDI Sequencer – Sampler – Record performance (keyboard)
  - Compose / Record
  - Arrange and Orchestrate – Notation
  - MIDI Controller – Synthesizer – Play music
    - Instruments and sounds
- **Features**
  - Polyphony – Number of voices / instrument sounds (32, 128)
  - Multitimbral – Simultaneous channels (16), instruments / patches

Waveforms

- **Sound as Waves**
  - Vibration – Harmonic motion – Sound pressure
  - Amplitude – Height; Change in pressure from peak to trough
  - Cycle – Period; Time to repeat one amplitude cycle
  - Frequency – Cycles per second (Hz) (1/Period)
  - Phase – Distance along cycle (in 360 degrees)
  - Wavelength – Length of one cycle (in., cm.)
- **Visualizing Sound**
  - Regular = pure tone, single note, no overtones
  - Amplitude = volume, Frequency = pitch
  - Middle C = 200 Hz, Octave step to same note (2X)
- **Combining Waveforms**
  - Add or subtract for more complex waveform
  - In-phase – Add; Out of phase 180 degrees – Cancel
  - Decompose

Loudness – Dynamic Range

- **Hearing is non-linear**
  - Amplitude of sound wave vs. perceived loudness
  - Relative loudness related to ratio of intensities
  - Dynamic range of sound amplitudes – clipping
- **Decibels (dB)**
  - Sound pressure level
  - Level (dB) = 20 log10 (A1/A2)
  - Double amplitude = 6 dB
- **Signal to Noise Ratio**
  - Each sample bit = 6 dB SNR
  - 8-bit = 48 dB SNR
  - 16-bit = 96 dB SNR
  - Noise floor – Hiss, hum, static

Sound	~ dB	~ X
Human hearing threshold	0	
Faint Sounds	10 - 30	10 X
Moderate Sounds	40 - 50	
Loud Sounds	60 - 80	1 k
Extremely Loud Sounds	90 -110	100 k
Painful Sounds	120 – 140	1 m 10 m

Decibel Levels

- **Hearing Damage**
  - Sounds above ~ 85 dB
  - Power plus time
    - 84 dB for 8 hours
    - 100 dB for 30 min.
    - 110 dB for 1 min.
  - By 140 dB – Immed. Damage
- **Signs**
  - Must shout to communicate
  - Ears ringing
  - Sounds muffled, as in a barrel
  - Sounds distorted

(dangerousdecibels.org)

Sound	~ dB
Human hearing threshold	0
Faint Sounds	
Quiet recording studio	10
Rustling leaves	20
Whisper	30
Moderate Sounds	
Quiet room	40
Moderate rainfall	50
Loud Sounds	
Normal conversation, Dishwasher	60
Busy traffic, Vacuum cleaner	70
Alarm clock	80
Extremely Loud Sounds	
Car horn, Lawnmower	90
Snowmobile, Chainsaw	100
Rock concert	110+
Painful Sounds	
Jet plane takeoff	120
Gunshot, Air raid siren	140

Analog to Digital Audio

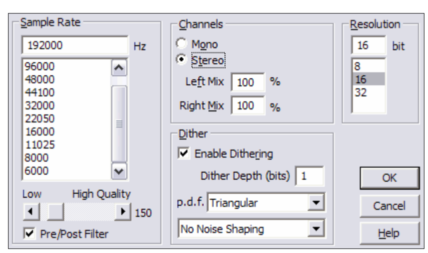
- **Analog**
  - Microphone – Converts sound pressure waves to voltage
  - Transmit – On wire as voltage (high positive, low negative)
  - Store – On tape (magnetic strength), vinyl records (groove height)
  - Speaker – Converts voltage to sound pressure wave
- **Digital – Sampled**
  - Analog to digital conversion (A-to-D)
    - Convert continuous waveform to discrete digital samples
    - Capture, Digitize, Sampling
    - Clipping – Beyond range, min/max
  - Sample rate, sample size
    - Nyquist rate – Sampling rate more than double frequency
    - Human voice < 5 kHz, Human hearing < 20 kHz
    - Avoid aliasing – Speech ~ 8k, CD audio ~ 44k

Sample Rates

5,000	**- Highest human voice**
8,000	Speech – Telephony – U-LAW
11,025	Quarter CD
16,000	G.722 compression standard.
18900	CD-ROM/XA standard
20,000	**- Limit of human hearing (17k)**
22,050	Half CD
32,000	**DV; Used in digital radio and other TV**
37800	CD-ROM/XA
44,056	Prof. audio, integral samples in video frame
44,100	**CD Audio**
48,000	**DV; DVD-Video; DAT (Digital Audio Tape)**
96,000	**CD Audio, AAC, DVD PCM**

Sample Size

- **Capture Analog Amplitude**
  – Higher fidelity, Greater dynamic range, Lower noise floor,
- **Sample format**
  – Bit depth / resolution
  – Quantization error – Distortion, lose fidelity, add noise
  – 8-bit – Voice (dither, some hiss)
  – 16-bit – Music (byte order)
  – 24-bit – DVD-Audio
  – Float – Processing



(Adobe Audition)

Audio Data Sizes

- **CD audio = 635 MB / hour (size of CD)**
- **MP3 = 1 MB / minute, 60 MB / hour (10 X); Another 1/2 to 1/3**

Raw Audio Rates	Samp/s	Bits	Chan	Bits/sec	KB/sec	MB/min	MB/Hr
Voice – Phone	8,000	8	1	64 K	8	0.5	29
CD Music (stereo)	44,100	16	2	1.4 M	176	10	635

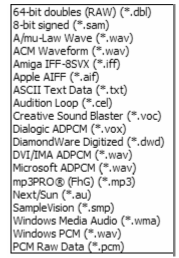
Audio Rates - Stereo	Kbps	MB/min	/song	MB/hr	%
CD Audio - uncompressed	**1411**	**10.5**	**42**	**635**	--
High-qual MP3, WMA, AAC	192	1.44	6	86	14%
Downloaded music, MP3	**128**	**1**	**4**	**58**	**9%**
High-quality streaming, portables	**64**	**0.5**	**2**	**29**	**5%**
CD stereo quality, WMA, AAC	**48**	**0.33**	**1.4**	**22**	**3%**
Low rate - music	20	0.15	0.6	9	1.4%
Low rate - voice	8	0.06	0.2	4	0.6%

Audio Compression

- **Compression**
  - Codec – Coder / Decoder – Algorithm
  - Lossless vs. Lossy (Perceptual)
  - Compression ratio: Lossless 2:1, Lossy 11:1 (MP3 - 1 MB/4 min.)
  - Constant / Variable Bit Rate (CBR, VBR)
  - DCT – Discrete Cosine Transform – Waves as weighted sum of cosines
  - DRM – Digital Rights Management – Copy protecxtion
- **PCM – Pulse Code Modulation - Uncompressed**
- **ADPCM – Adaptive Differential Pulse Code Modulation**
  - Compressed – 16-bit sound data into 4-bit differences
- **Voice Compression**
  - ITU-T G.711, u-law, A-law; CCITT G.721, G.723, ITU-T G.726
  - LPC – Linear Predictive Coding -> GSM
    - Synthetic speech – Fit analytic vocal tract model
  - CELP – Code Excited Linear Predictor – LPC + errors

Audio File Formats

- **Platform Formats**
  - AIFF – Audio Interchange File Format – Macintosh
    AIFC – Compressed
  - WAVE – Windows – Typically PCM; Lossy ADPCM – 4:1
  - AU, SND – Sun – mu-law – 2:1

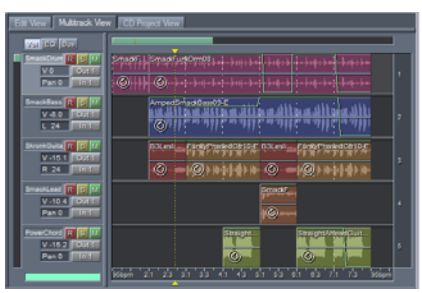


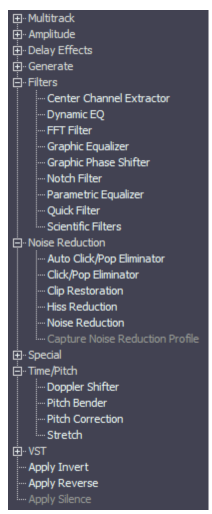
(Adobe Audition)

- **Web Formats**
  - MP3 – MPEG layer 3 - 32-320 kbps - target 64 kbps
  - AAC – Advanced Audio Coding – MPEG-2 (iTunes)
  - RA, RM – Real Audio
  - ASF, WMA – Windows Media
- **Other Formats**
  - Ogg Vorbis – Open source, royalty free; Vorbis music not voice
  - ATRAC – Adaptive Transform Acoustic Encoding
    - Sony MiniDisc; ATRAC3 split frequency bands

Audio Processing

- **Sample editing**
  - Waveform editing
- **Multi-track mixing and effects**
  - VU meter, cross-fade, gang
  - Location: Pan / Balance channels
    Surround-sound location
  - Sweeten
- **Looping**
- **Composing**
  - Sound track generation





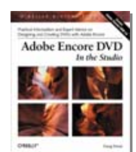
(Adobe Audition)

Waveform Processing

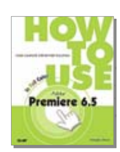
- **Amplitude – Levels – Normalize**
  - Dynamic range compress – Low to high, sound louder, headroom
  - Gain – Amplify, attenuate
- **Equalizer – Boost / reduce amplitude in freq. bands**
  - Graphic equalizer – Graphical view, sliders
- **Transforms**
  - Decompose to frequency components, filter bands
  - Fourier – Periodic wave is sum of harmonic frequencies (sine)
- **Filters – De-Esser hiss, Hum, Low pass, High pass**
- **Delays – Reverb, Echos, Chorus, Sound space**
- **Noise Reduction – Click/Pops, Hiss**
- **Time/Pitch – Pitch correction**

For More Information

Manifest Technology®
www.manifest-tech.com



The Manifest Technology site by Douglas Dixon contains over 150 articles and technical references on multimedia technology, especially digital video editing and DVD authoring.



Manifest Technology is a registered trademark of Douglas Dixon