Chapter 5.
Digital Audio Processing

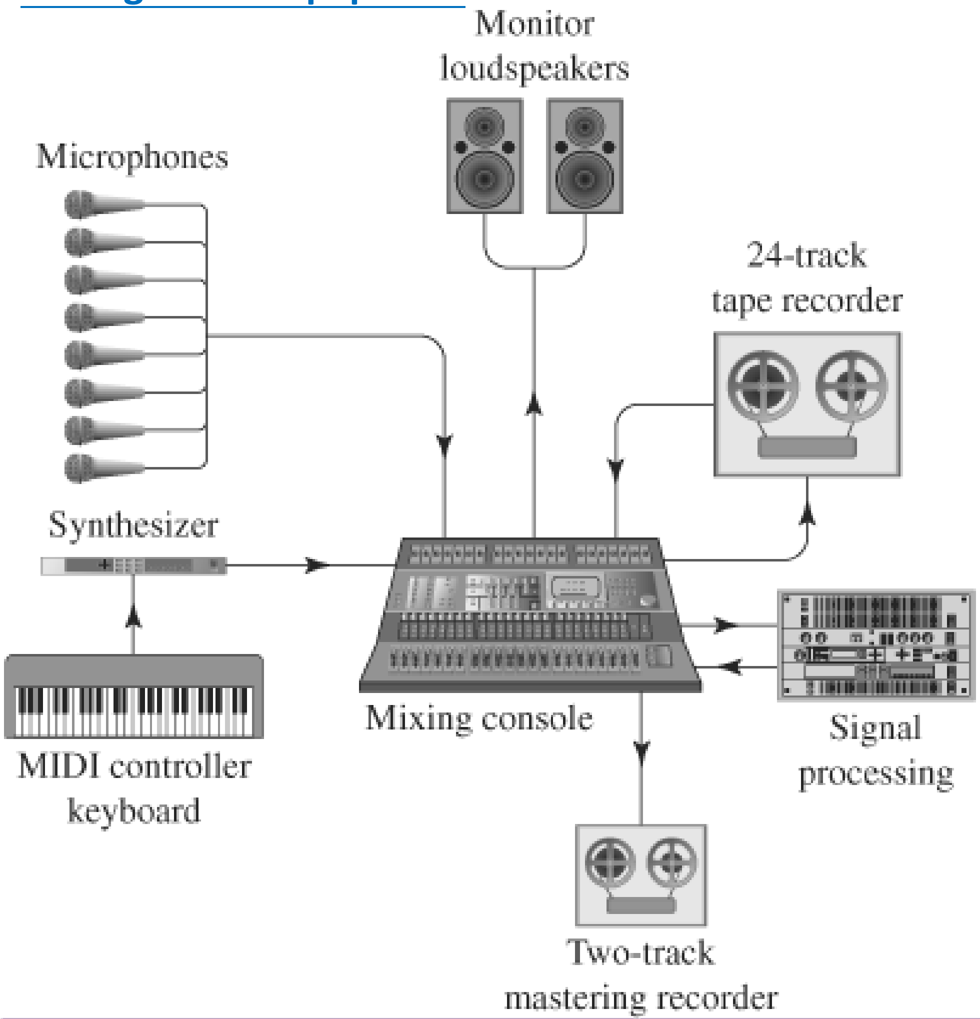
Part I: Sec. 5.1-5.3

Objectives

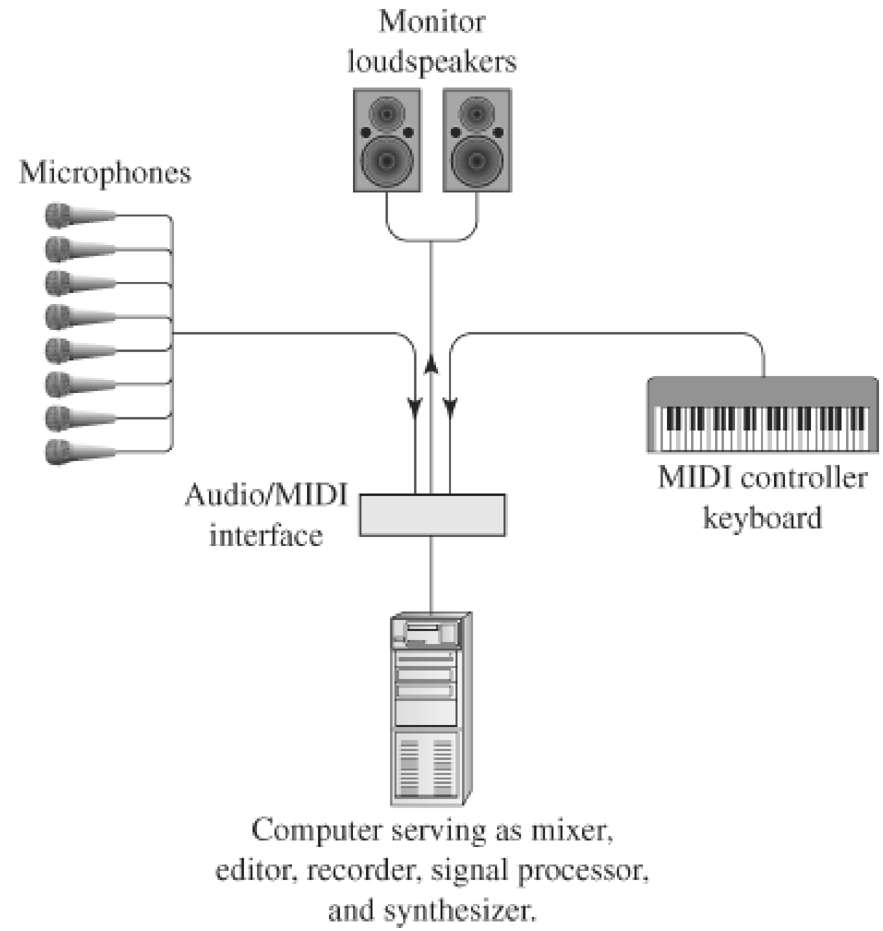
- Know the basic hardware and software components of a digital audio processing environment.
- Understand how normalization, compression, expansion, equalization, and reverb are applied and what they do to digital audio.
- Understand methods for audio restoration.
- Understand how filters are applied and how they work mathematically.
- Understand the concept and examples of time-based encoding for digital audio.
- Understand the concept and examples of perceptual encoding for digital audio.
- Understand the concept, implementation, and application of MPEG audio compression.

Digital Audio Work Environments

Analog sound equipment



Digital sound equipment



Digital Audio Work Environments

- Analog sound equipment
  - A mixer(mixing console) is used to gather inputs from microphones and instruments and dynamically adjust their amplitudes, equalize the frequency components, compress the dynamic range, apply special effects, and route the processed sound to outputs such as speakers.
  - Each input to the mixer is designated as a **channel.** Multiple channels can be gathered into a **bus.**
  - The signal can be split to multiple outputs so that speakers can be set up appropriately around an auditorium to create the desired sound environment.

Digital Audio Work Environments

- Digital sound equipment
  - Many hardware devices are absorbed into the software of the digital mixer.
  - There are an analog-to-digital converter(ADC) and a digital-to-analog converter(DAC) inside the digital mixer.
  - The audio/MIDI interface connects the hardware with the computer.

Sound Card

- Sound card is the most basic element of a digital audio workstation.
- A sound card
  - provides input jacks for microphones and external audio sources
  - converts sound from analog to digital form as it is being recorded, using an ADC
  - provides output jacks for headphones and external speakers
  - converts sound from digital to analog form as it is being played, using a DAC
  - synthesizes MIDI sound samples.

Digital Audio Processing Software

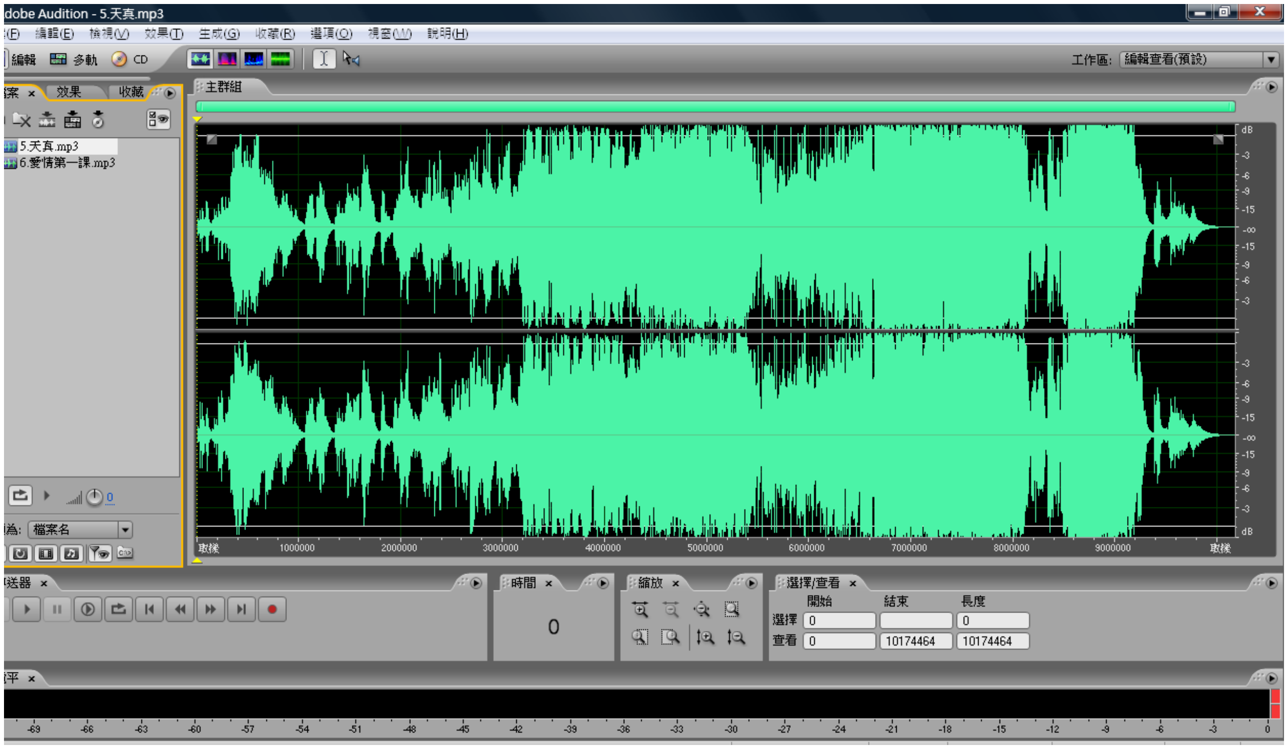
- Generally, digital audio processing softwares have the following features:
  - the ability to import and save audio files in a variety of formats
  - an interface (called transport controls) for recording and playing sound
  - a waveform view that allows you to edit the wave, often down to the sample level
  - multitrack editors
  - audio restoration tools to remove hisses, clicks, pops, and background noise

Digital Audio Processing Software

- the ability to take input from or direct output to multiple channels
- special effects such as reverb, panning, or flange
- controls for equalizing and adjusting volume and dynamic range
- frequency filters
- the ability to handle the MIDI format along with digital audio and to integrate the two types of data into one audio file
- the ability to record samples and add to the bank of MIDI patches
- compression codecs

Waveform View

- In digital audio processing programs, you often have a choice between working in a ***waveform view*** or a ***multitrack view***.



Waveform view (Adobe Audition)

Waveform View

- Sometimes also called the **sample editor**.
- You can view and edit a sound wave down to the level of individual sample values.
- The waveform view is where you apply effects and processes that cannot be done in real-time, primarily because these processes require that the whole audio file be examined before new values can be computed.
- Sample values are altered.

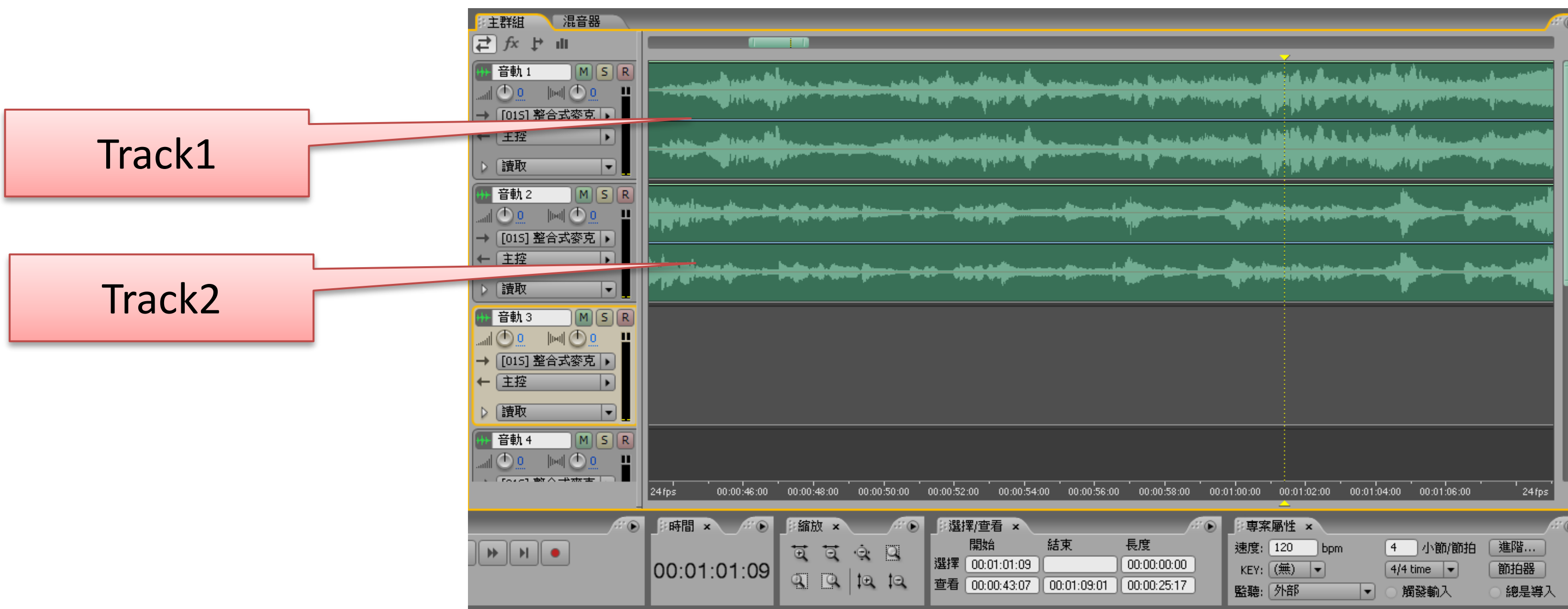
Waveform View

- The standard representation of time for digital audio and video is **SMPTE** (*Society of Motion Picture and Television Engineers*)
- The timeline is divided into hours, minutes, seconds, and frames, denoted as *h : m: s : f*
- If the file is a video, then the frame number is associated with the video. Otherwise(pure audio file), the programs provide some choices such as 24fps or 30 fps.
- 1:2:3:4 →

  1st hour, 2nd minute, 3rd second, 4th frame

Multitrack View

- The multitrack view allows you to record different sounds, musical instruments, or voices on separate tracks so that you can work with these units independently.



Multitrack View

- A **track** is a sequence of audio samples that can be played and edited as a separate unit.
- You can **mix down** the tracks, collapsing them all into one unit.
- Often, different tracks are associated with different instruments or voices.

Mastering

- **Mastering** is the process of preparing and transferring recorded audio from a source containing the final mix to a data storage device.
- For example, if the audio file is one musical piece to be put on a CD with others, mastering involves sequencing the pieces, normalizing their volumes with respect to one another so one doesn't sound much louder than another.

Channel

- A **channel** corresponds to a stream of audio data, both input and output.
- Recording on only one channel is called monophonic or simply **mono**. Two channels are **stereo**.
- Different channels are sent out through different speakers, giving the sound more dimension, as if it comes from different places.
- The popularity of multichannel audio is also growing.

Digital Audio File Types

- File formats differ in
  - How are the samples encoded?
  - What is the format of the data?
  - Is the file compressed?
  - …etc.
- Raw files have nothing but sample values in them. There's no header to indicate the sampling rate, sample size, or type of encoding.
- Sometimes different file types have the same file extension. We must identify the file type by reading the header of the file.

Digital Audio File Types

- Representative Audio File Formats

.rm	RealMedia	Supports streamed audio, which allows you to begin listening to the audio without having to download the entire file first.
.tta	True Audio	A free codec providing lossless compression at a ratio of about 1.5:1 or 2:1; popular for lossless compression of web-based audio; competitive with *.flac*.
A-law or μ-law *.wav*	CCITT standard G.711 standard formats	8-bit per sample files created from 16-bit per sample using A-law or μ-law encoding, achieving the equivalent of about 13-bit dynamic range (about 78 dB).
DVI/IMA *.wav*	International Multimedia Association version of ADPCM *.wav* format with ADPCM compression	Uses a different (faster) method than Microsoft ADPCM; a good alternative to MPEG with fast decoding and good quality of compressed audio.
Microsoft ADPCM *.wav*	Microsoft	Uses ADPCM; yields 4-bit per channel compressed data for 4:1 compression ratio.
Windows PCM *.wav*	Microsoft	Standard Windows *.wav* format for uncompressed PCM audio; supports both mono and stereo at a variety of bit depths and sample rates; follows RIFF (Resource Information File Format) specification, allowing for extra user information to be saved with the file.
.wma or *.asf*	Microsoft Windows Media (audio)	Microsoft proprietary codec; allows you to choose quality settings, including constant bit rate (CBR) vs. variable bit rate (VBR) and lossy vs. lossless compression; competitive with *.aac*; uses *.asf* file extension if encapsulated in Advanced Systems Format; supports streaming audio.

Digital Audio File Types

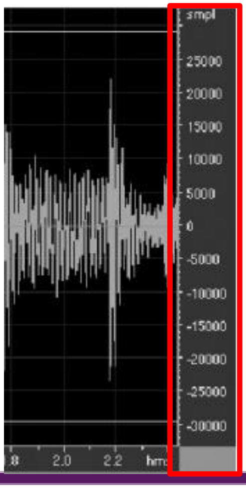
.rm	RealMedia	Supports streamed audio, which allows you to begin listening to the audio without having to download the entire file first.
.tta	True Audio	A free codec providing lossless compression at a ratio of about 1.5:1 or 2:1; popular for lossless compression of web-based audio; competitive with *.flac*.
A-law or μ-law *.wav*	CCITT standard G.711 standard formats	8-bit per sample files created from 16-bit per sample using A-law or μ-law encoding, achieving the equivalent of about 13-bit dynamic range (about 78 dB).
DVI/IMA *.wav*	International Multimedia Association version of ADPCM *.wav* format with ADPCM compression	Uses a different (faster) method than Microsoft ADPCM; a good alternative to MPEG with fast decoding and good quality of compressed audio.
Microsoft ADPCM *.wav*	Microsoft	Uses ADPCM; yields 4-bit per channel compressed data for 4:1 compression ratio.
Windows PCM *.wav*	Microsoft	Standard Windows *.wav* format for uncompressed PCM audio; supports both mono and stereo at a variety of bit depths and sample rates; follows RIFF (Resource Information File Format) specification, allowing for extra user information to be saved with the file.
.wma or *.asf*	Microsoft Windows Media (audio)	Microsoft proprietary codec; allows you to choose quality settings, including constant bit rate (CBR) vs. variable bit rate (VBR) and lossy vs. lossless compression; competitive with *.aac*; uses *.asf* file extension if encapsulated in Advanced Systems Format; supports streaming audio.

DBFS

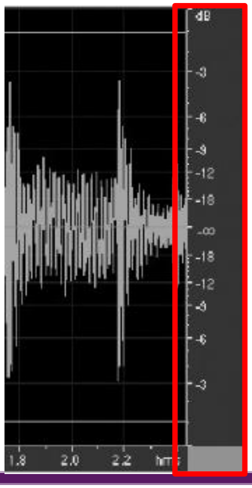
- In audio processing softwares, the amplitude is often shown in **dBFS. (*decibels-full-scale*)**

$$\text{dBFS} = 20\log_{10}\left(\frac{|x|}{2^{n-1}}\right)$$

- The center of the amplitude axis is $-\infty$ dBFS, and above and below this axis the values progress to the maximum of 0 dBFS.



Sample value



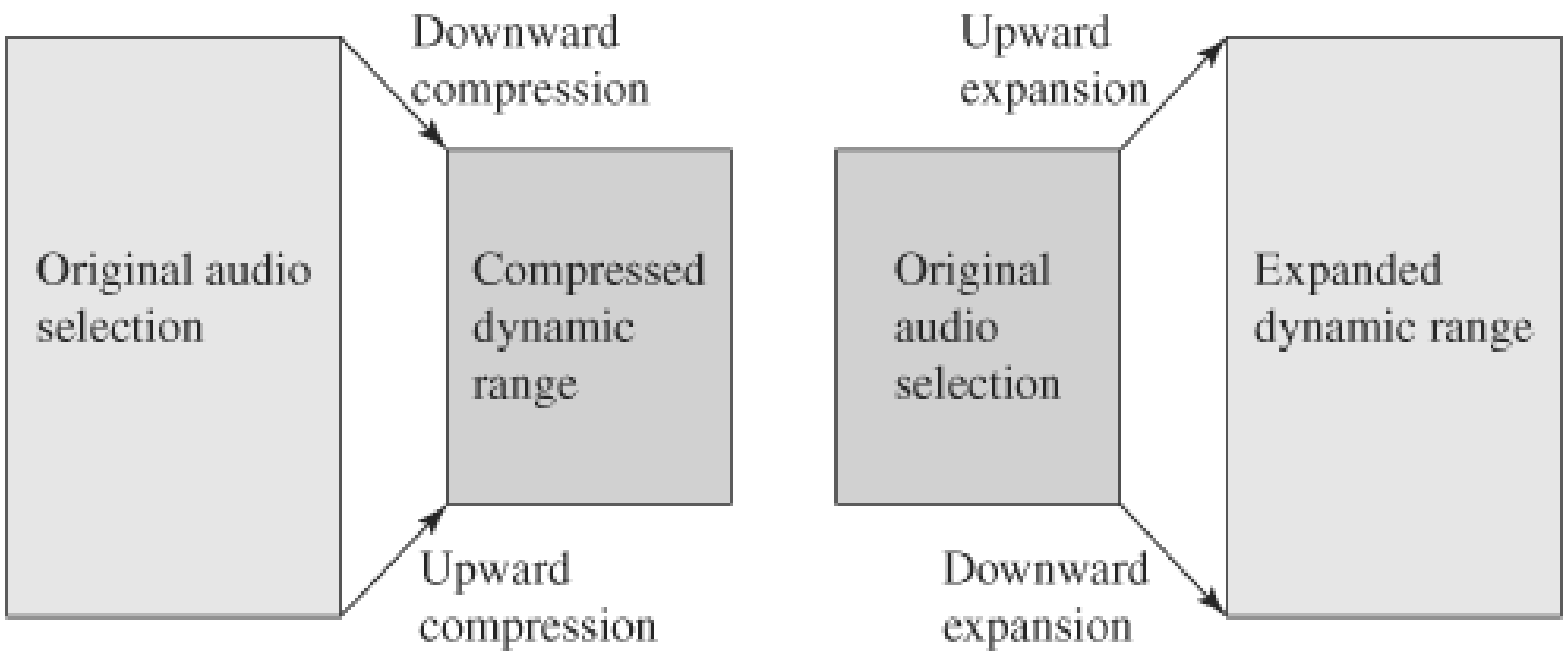
dBFS

Dynamic Processing

- **Dynamics processing** is the process of adjusting the dynamic range of an audio selection, either to reduce or to increase.
- An increase in amplitude is called **gain** or **boost**. A decrease in amplitude is called **attenuation** or, informally, a **cut**.
- We introduce 4 digital dynamics processing tools here: **hard limiting**, **normalization**, **compression**, and **expansion**.

Compression and Expansion

- Types of dynamic range compression and expansion

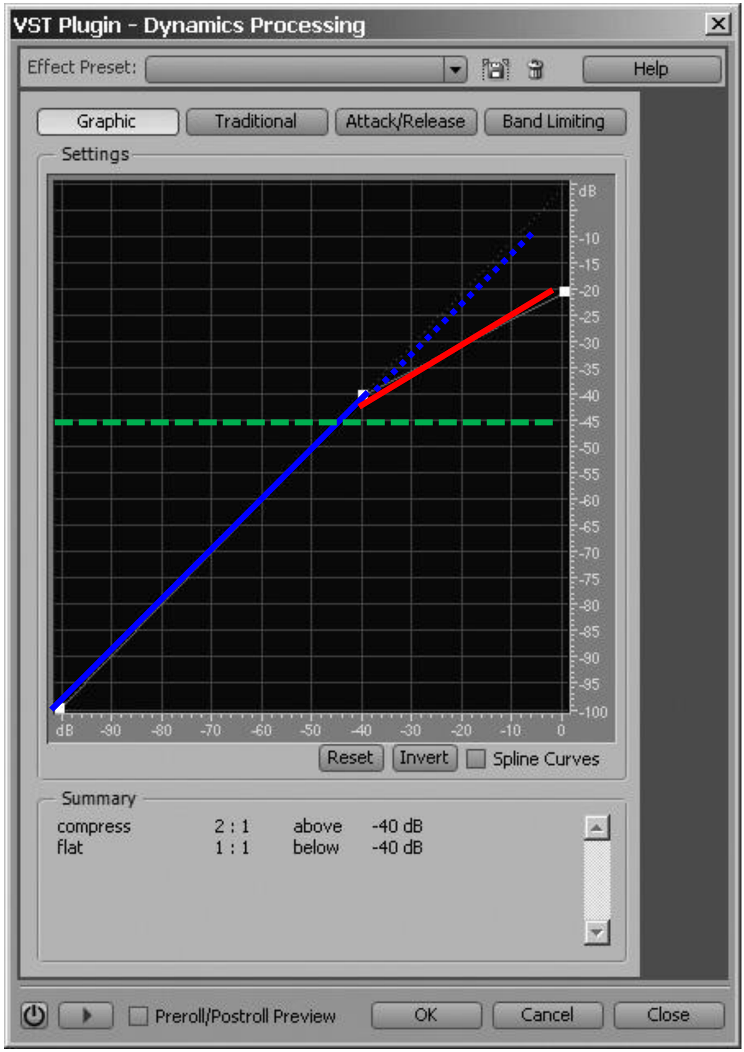


Compression and Expansion

- ***Downward compression*** lowers the amplitude of signals that are above a designated level, without changing the amplitude of signals below the designated level. It reduces the dynamic range.
- ***Upward compression*** raises the amplitude of signals that are below a designated level without altering the amplitude of signals above the designated level. It reduces the dynamic range.
- ***Upward expansion*** raises the amplitude of signals that are above a designated level, without changing the amplitude of signals below that level. It increases the dynamic range.
- ***Downward expansion*** lowers the amplitude of signals that are below a designated level without changing the amplitude of signals above this level. It increases the dynamic range.

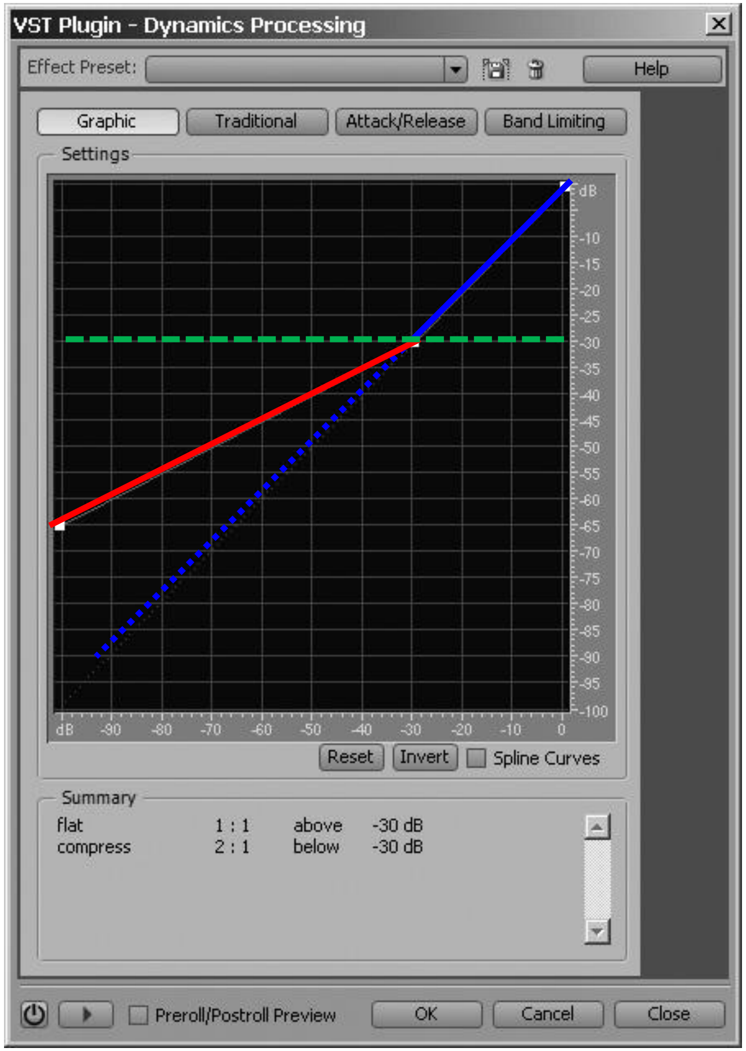
Compression and Expansion - Examples

- Downward compression:
- Amplitudes higher than -40dB is lowered by a 2 : 1 ratio.



Compression and Expansion - Examples

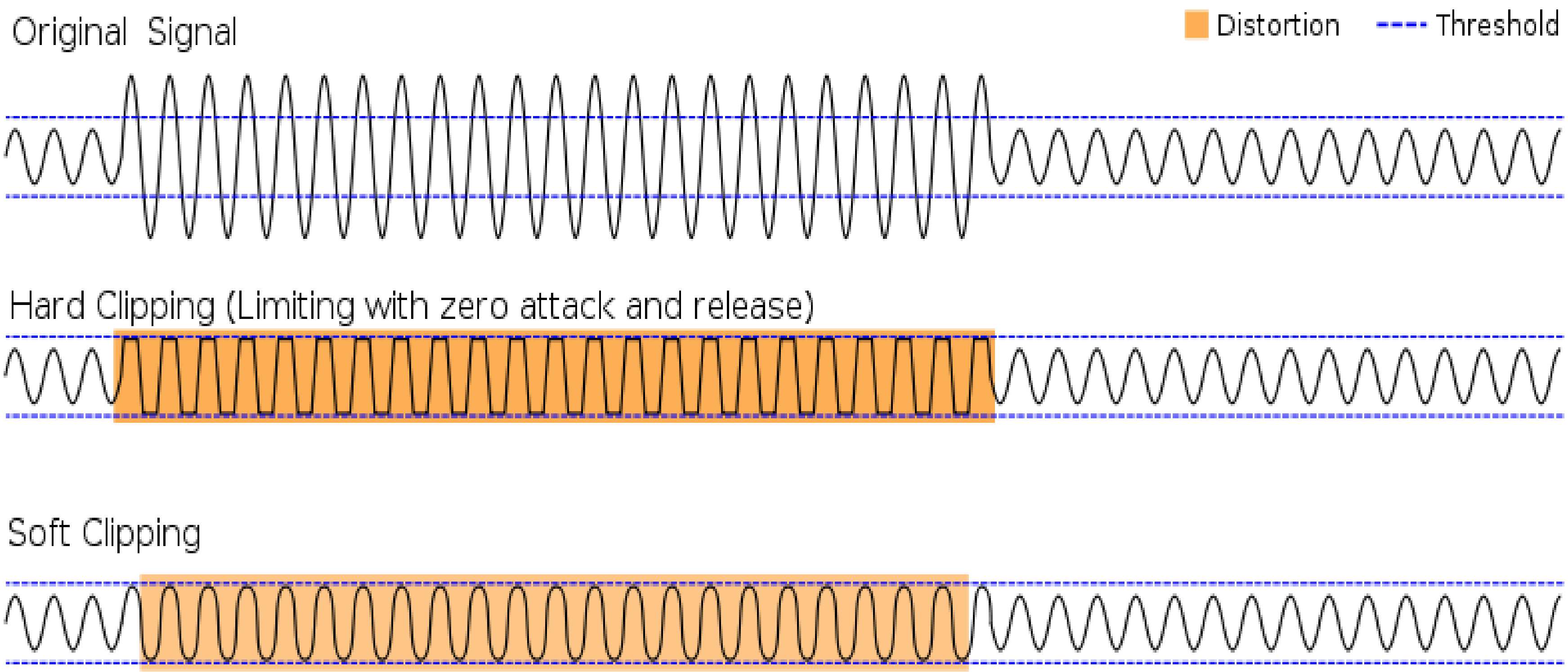
- Upward compression:
- Amplitudes higher than -30dB is lowered by a 2 : 1 ratio.



Limiting

- **Audio limiting** limits the amplitude of an audio signal to a designated level.
- Hard limiting (clipping)
  - cuts amplitudes of samples to a given maximum and/or minimum level.
- Soft limiting
  - audio signals above the designated amplitude are recorded at lower amplitude.

Limiting



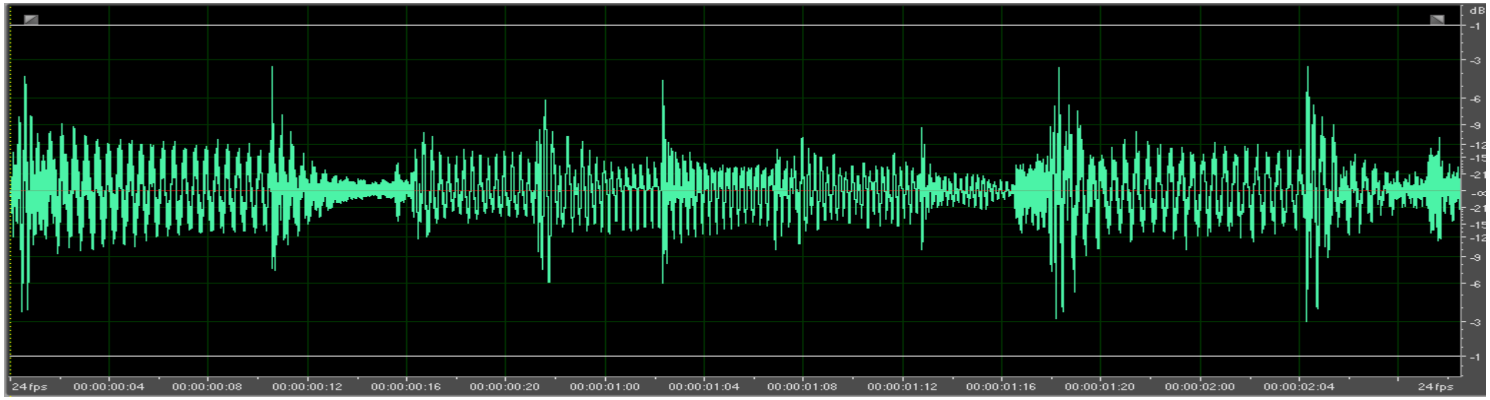
http://en.wikipedia.org/wiki/File:Clipping_compared_to_limiting.svg

Normalization

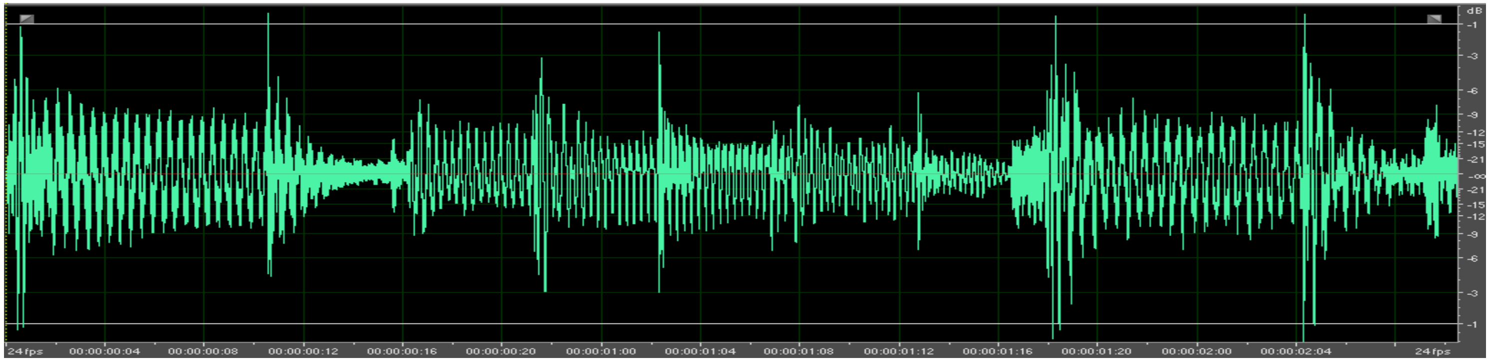
- Often, normalization is used to increase the perceived loudness of a piece after the dynamic range of the piece has been compressed.
- Normalization steps:
  1. Find the highest amplitude sample in the audio selection.
  2. Determine the gain needed in the amplitude to raise the highest amplitude to maximum amplitude.
  3. Raise all samples in the selection by this amount.

Dynamic Processing - Example

- Bossa.wav - original

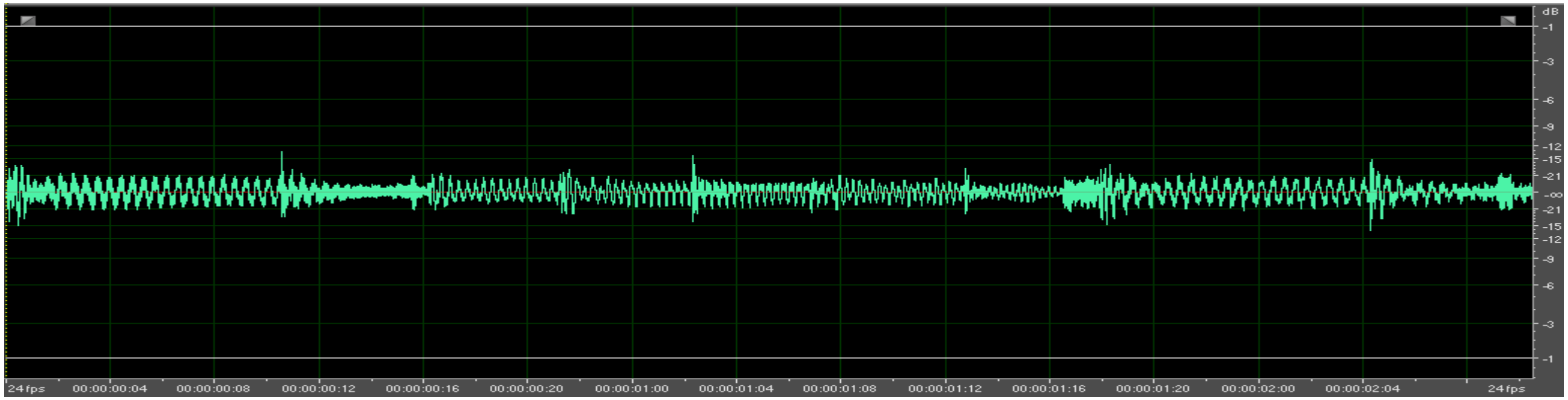


- Bossa.wav - normalized

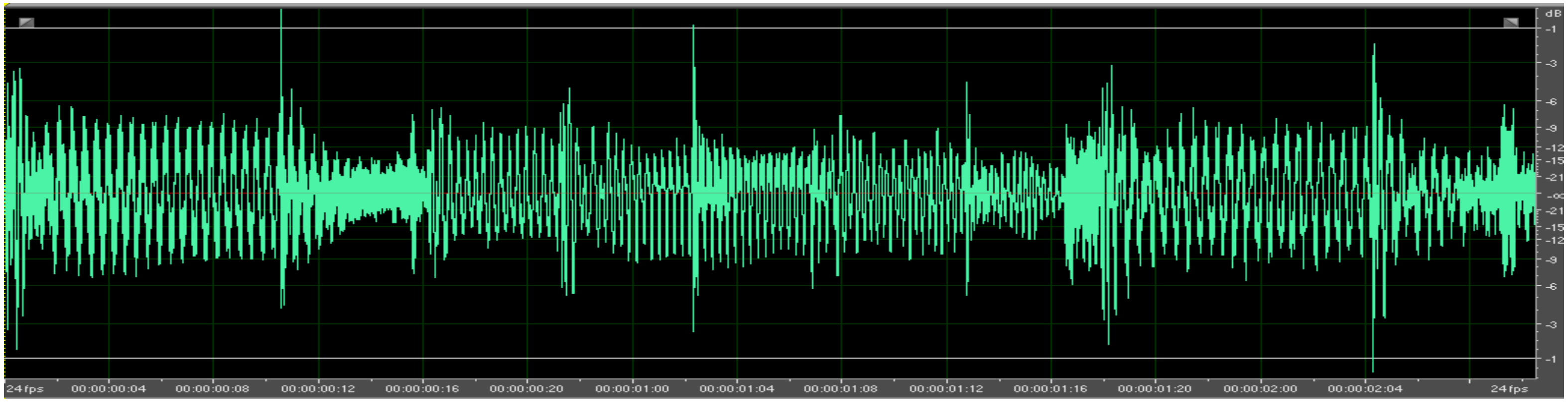


Dynamic Processing - Example

- Bossa.wav - compressed



- Bossa.wav – compressed + normalized

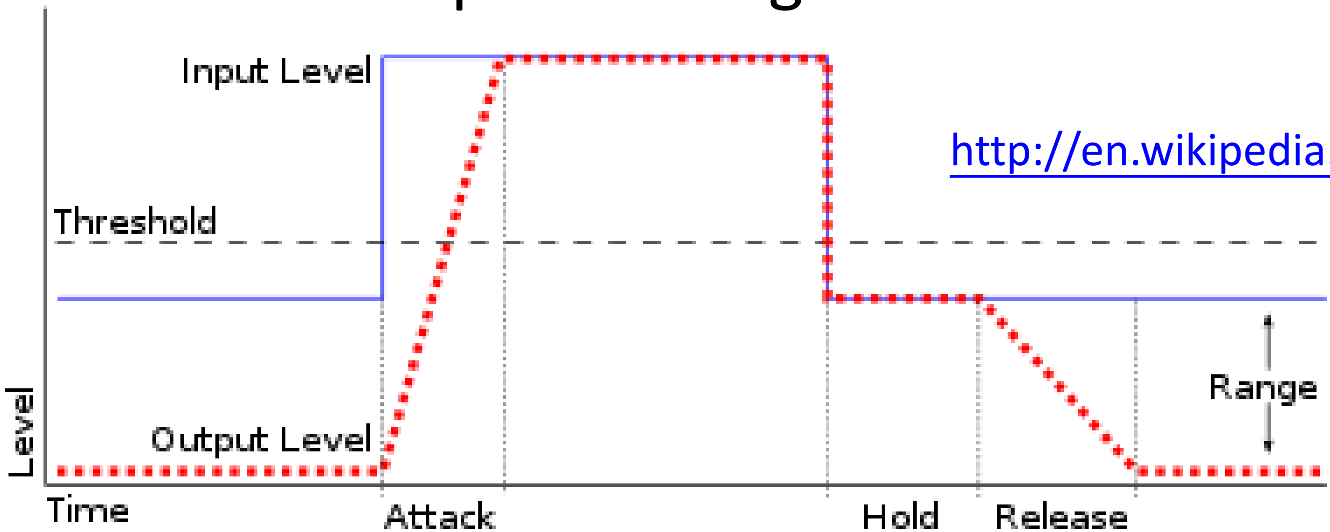


Audio Restoration

- In this section, we introduce three basic types of audio restoration to alleviate the background noise that arises from the microphone, air, disk …etc.
  - Noise gating
  - Noise reduction
  - click and pop removal

Noise Gating

- A **noise gate** allows a signal to pass through only when it is above a set threshold.
- It is used when the level of the signal is above the level of the noise. It does not remove noise from the signal. When the gate is open, both the signal and the noise will pass through.



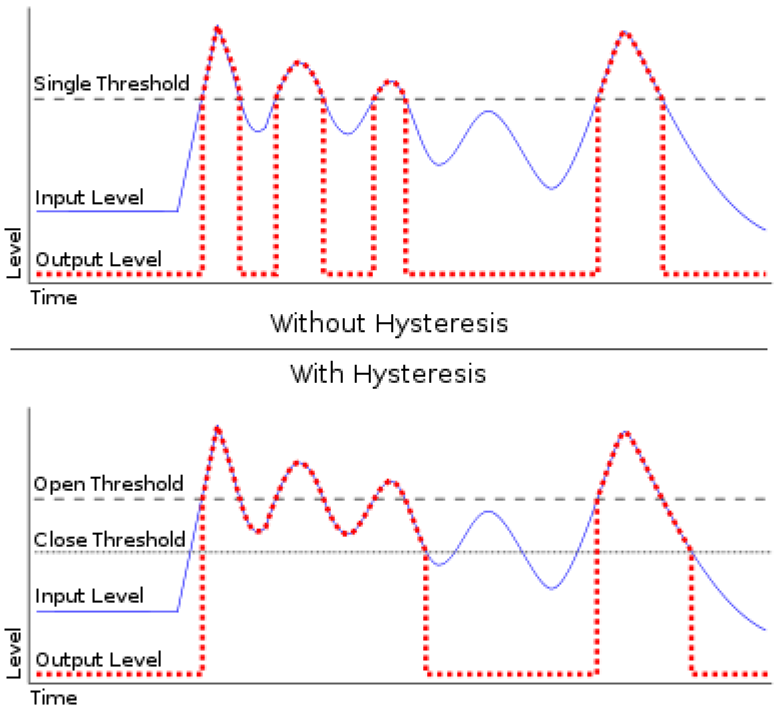
http://en.wikipedia.org/wiki/Noise_gate

Noise Gating

- **Reduction Level**: the amplitude to which you want the below-threshold samples to be reduced.
- **Attack**: the attack time indicates how quickly you want the gate to open when the signal goes above the threshold, like fade-in.
- **Release**: The release time indicates how quickly you want the gate to close, like fade-out.
- **Hold**: the amount of time the gate will stay open after the signal falls below the threshold.

Noise Gating

- If the signal keeps moving back and forth around the threshold, the gate will open and close continuously, creating a kind of **chatter.**
- The **hysteresis** control indicates the difference between the value **n** that caused the gate to open and the value **m** that will cause it to close again. If $n - m$ is large enough to contain the fluctuating signal, the noise gate won't cause chatter.

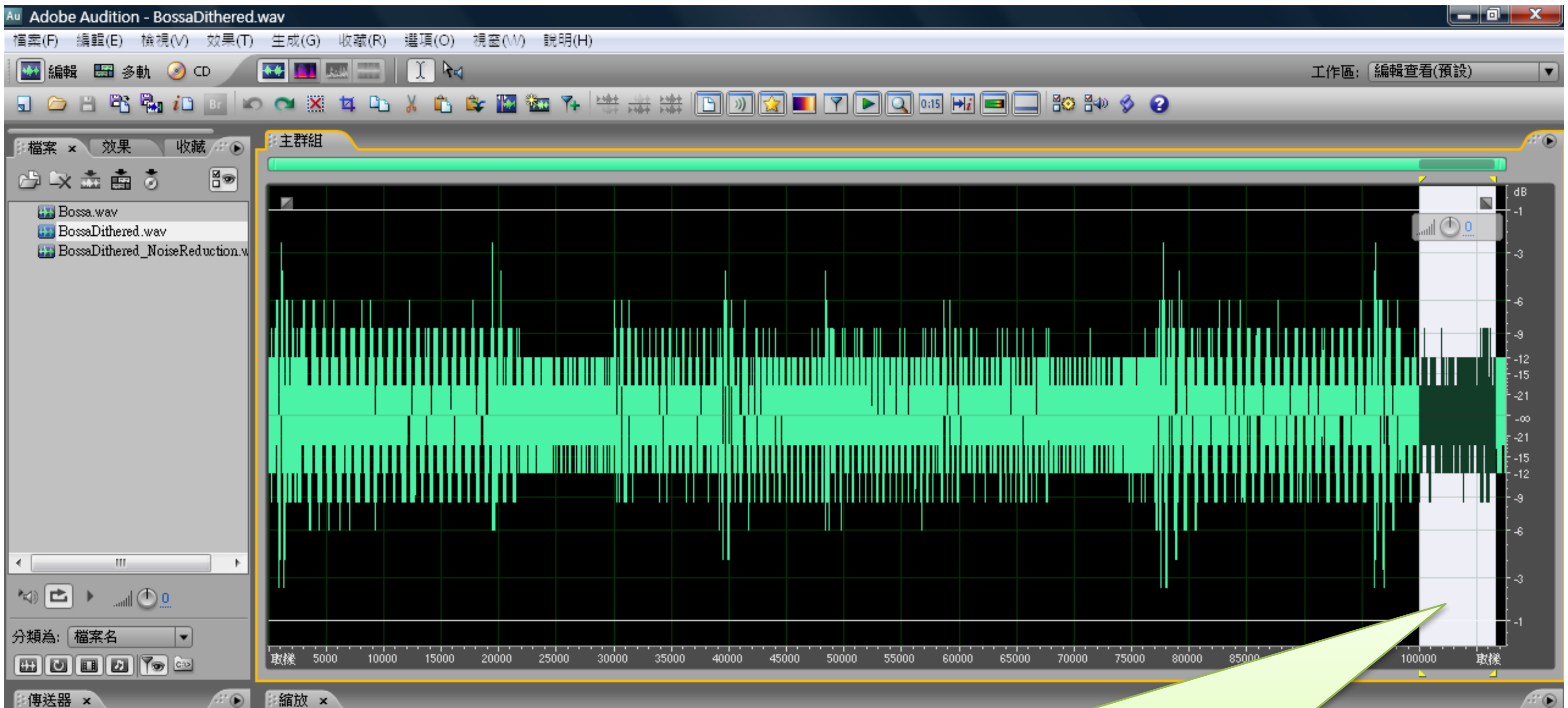


Noise Reduction

- Steps for noise reduction:
  1. Get a profile of the background noise. This can be done by selecting an area that should be silent, but that contains a hum or buzz.
  2. Determine the frequencies in the noise and their corresponding amplitude levels.
  3. The entire signal is processed in sections(FFT). The frequencies in each section are analyzed and compared to the profile, and if these sections contain frequency components similar to the noise, these can be eliminated below certain amplitudes.

Noise Reduction - Example

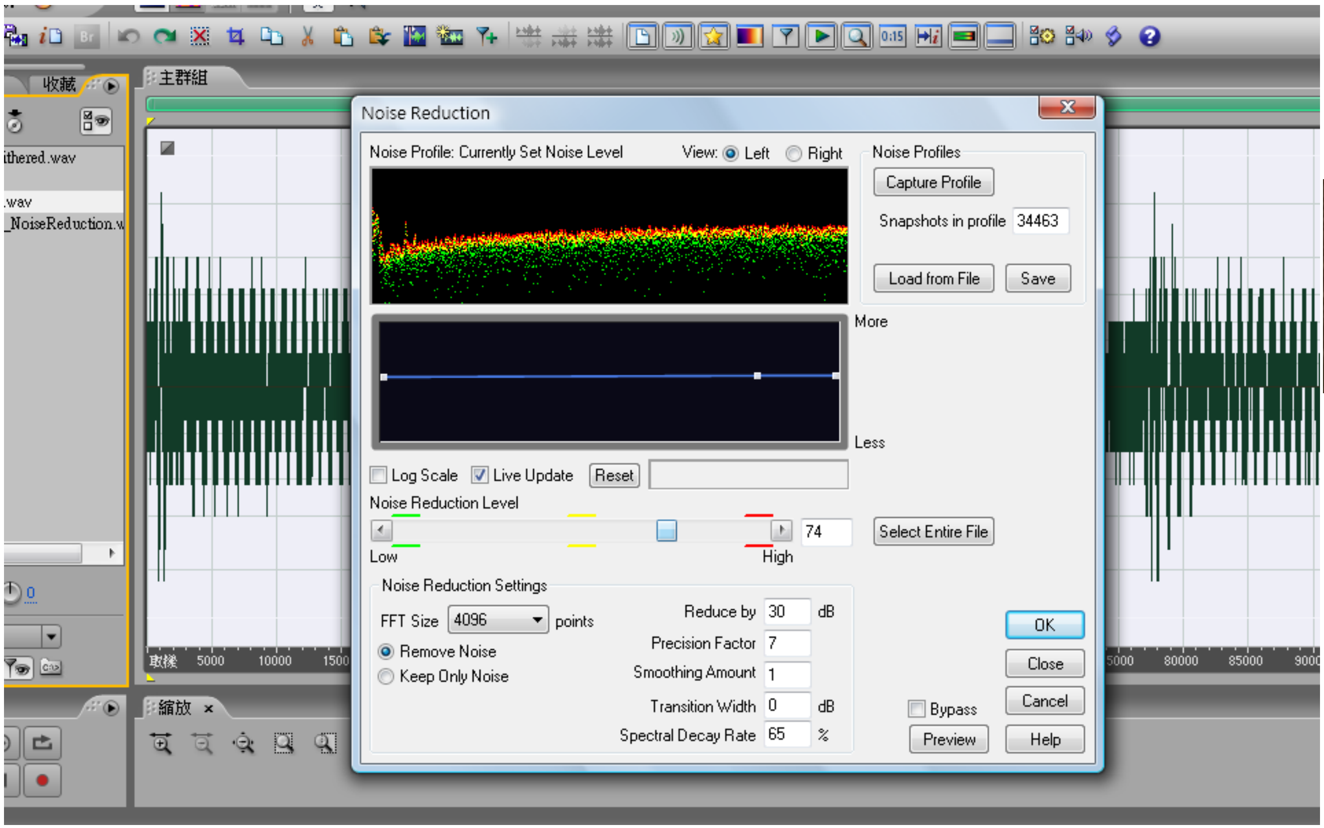
- Reduce the noise of 'Bossa_dithered.wav'



The area that should be silent

Noise Reduction - Example

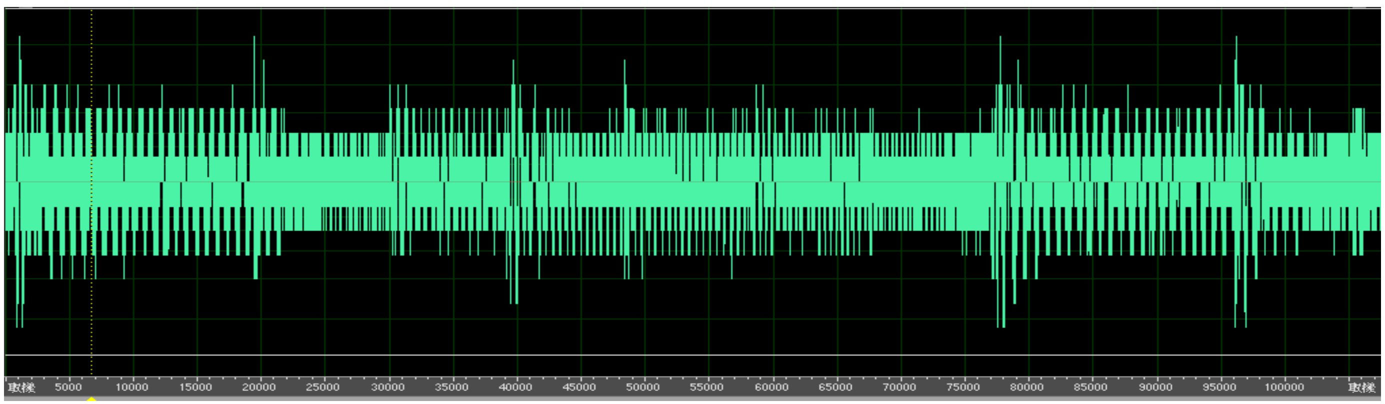
- Perform noise reduction



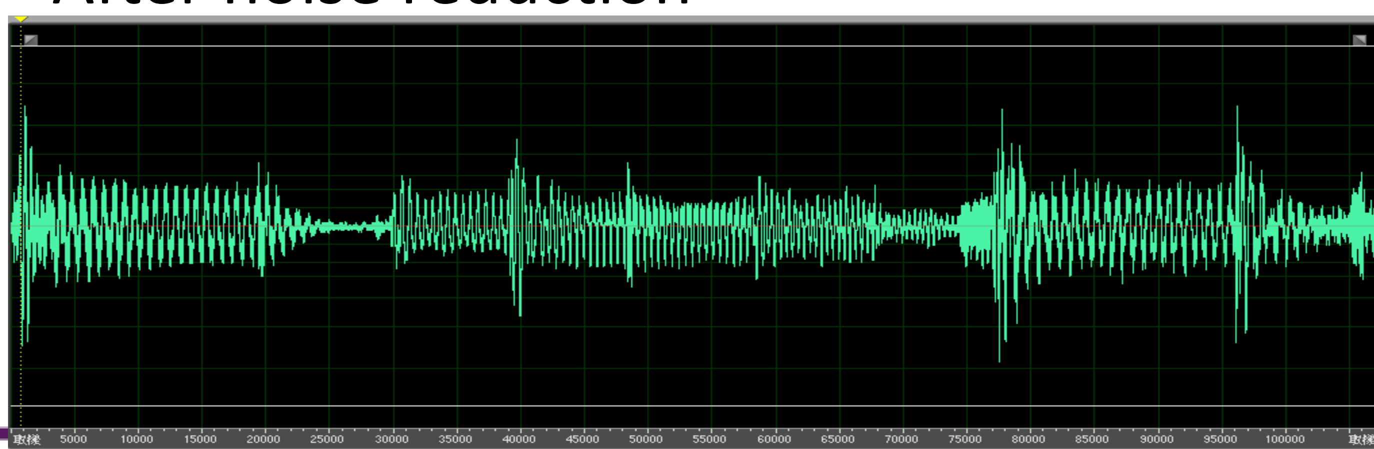
Red: original audio
Yello: processed audio
Green: noise floor

Noise Reduction - Example

- Before noise reduction



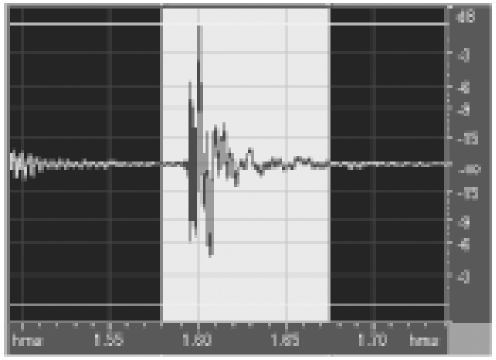
- After noise reduction



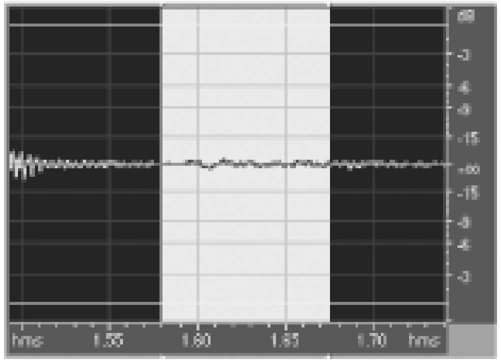
Click and Pop Removal

- A click or pop eliminator can look at a selected portion of an audio file, detect a sudden amplitude change, and eliminate this change by interpolating the sound wave between the start and end point of the click or pop.

An unwanted click in an audio file



Zooming in and selecting the click



The click removed