Chapter 5.5
Audio Programming

Audio Programming

Audio in games is more important than ever before

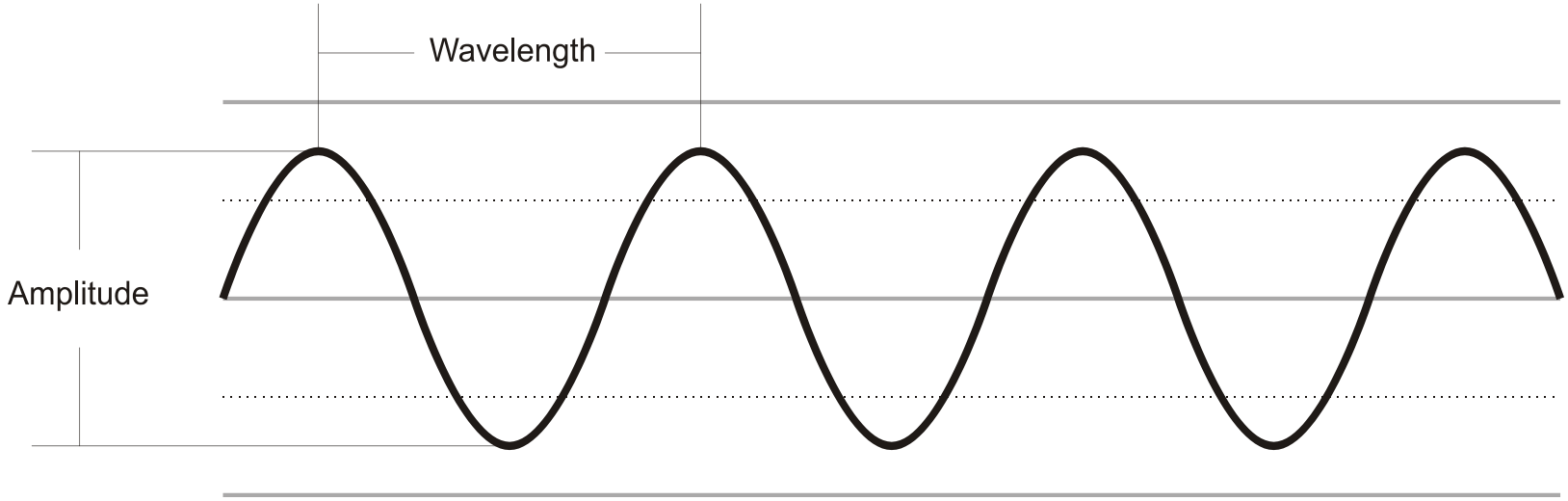
Programming Basic Audio

- Most gaming hardware has similar capabilities (on similar platforms)
  - Mostly programming interfaces differ
- Learning fundamental concepts of audio programming is important

API Choices

- DirectSound (part of DirectX API)
  - Only available on Windows platforms
- OpenAL
  - Newer API
  - Available on multiple platforms
- Proprietary APIs
  - Typically available on consoles
- 3rd Party Licensable APIs
  - Can offer broad cross-platform solutions

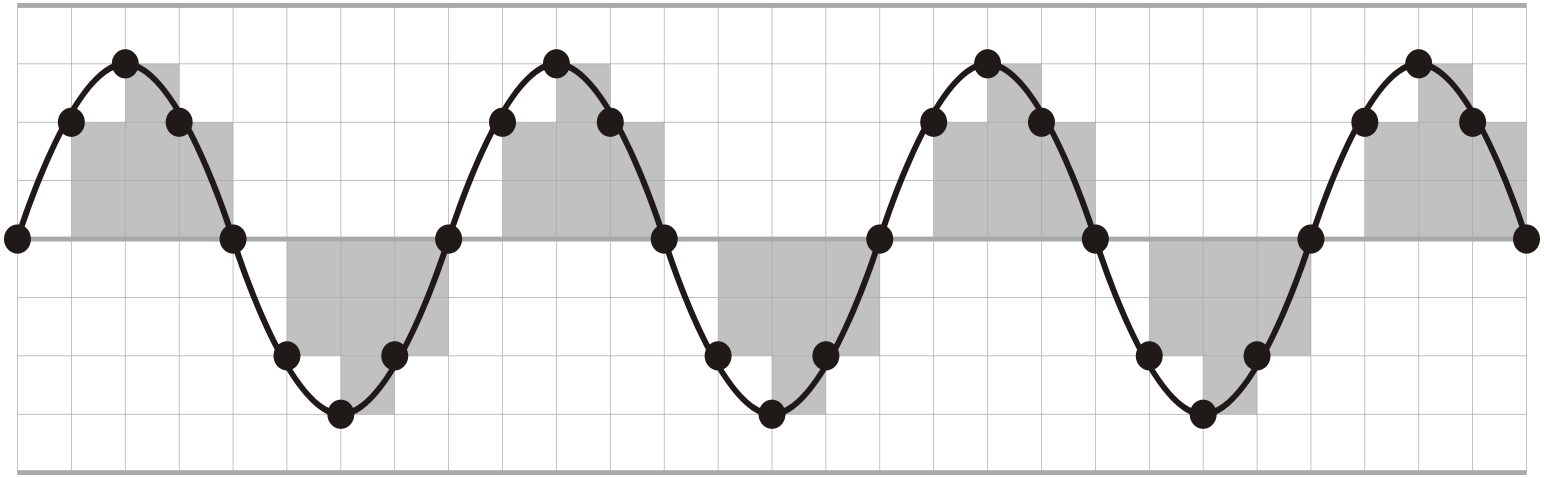
Analog Sound Wave



Basic Audio Terminology and Physics

- **Amplitude**
  - Measurement of a sound wave's pressure
- **Frequency**
  - Measurement of the interval between wave cycles, typically measured in Hertz
- **Pitch**
  - The perception of frequency
- **Tuning**
  - Musical distribution of frequencies over keys
- **Decibel**
  - Measures sound amplitude

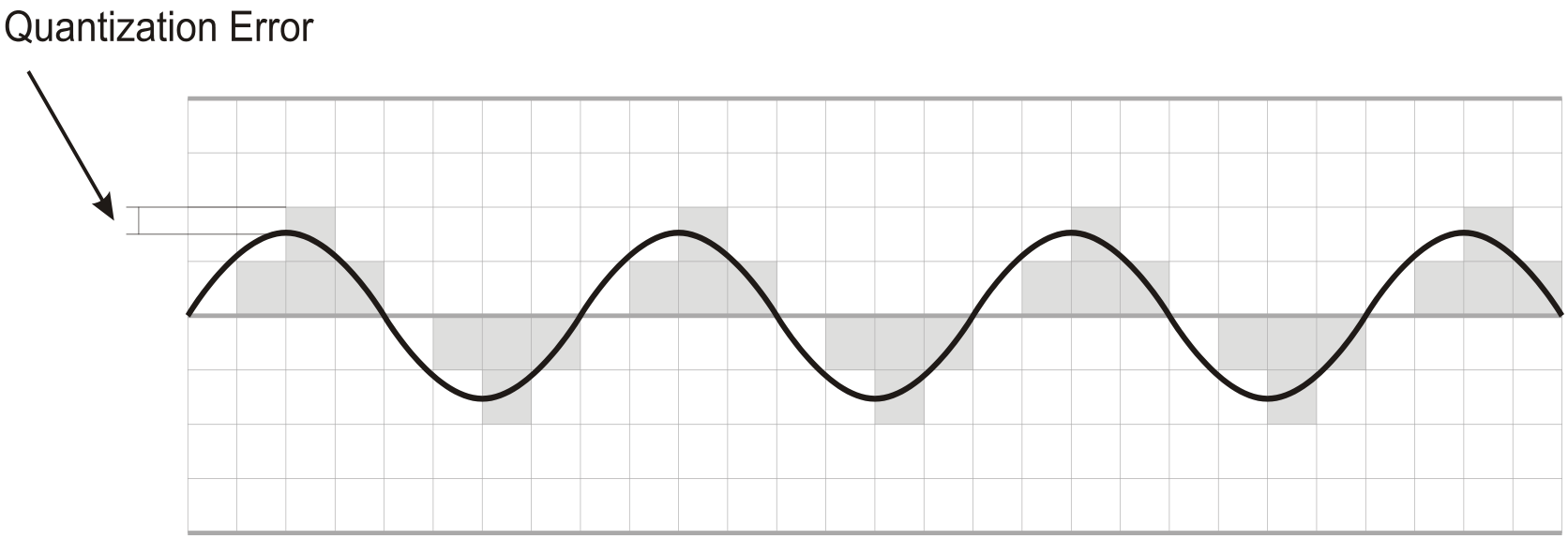
Digital Representation of a Sound Wave



Digital Representation of a Sound Wave

- Most common technique known as *sampling*
  - Sampling involves measuring the amplitude of the analog wave file at discrete intervals
  - The frequency of sampling is known as *sampling rate*
  - Each sample is typically stored in a value ranging from 4 to 24 bits in size
  - The size of the sample value in bits is known as the 'bit depth'
  - Music CDs have a sample rate and bit depth of 44.1 kHz (samples/sec) and 16 bits (sample size)

Quantization Error in Sampling



Bit Depth and Signal Noise

- Bit depth of sample data affects signal noise
  - Signal to noise ratio = number of available bits / 1
  - For example, 8-bit samples have a 256:1 SNR (~48 dB), and 16-bit samples have a 65,536:1 SNR (~96 dB)
  - Decibel ratio is calculated using 10 x $\log_{10}$ (ratio) or 8.685890 x log *e* (ratio)

Sampling Frequency and Frequency Reproduction

- Sampling frequency affects range and quality of high-frequency reproduction
- Nyquist Limit
  - Frequencies up to one-half the sampling rate can be reproduced
  - Audio quality degrades as frequency approaches this limit

Modern Audio Hardware

- Samples are piped into sound “channels”
  - Often a hardware pipeline from this point
- Various operations, such as volume, pan, and pitch may be applied
- 3D sounds may apply HRTF algorithms and/or mix the sound into final output buffers.

Sound Playback Techniques

- Two basic playback methods:
  1. Play sample entirely from memory buffer
  2. Stream data in real-time from storage medium
     - Streaming is more memory efficient for very large audio files, such as music tracks, dialogue, etc
     - Streaming systems use either a circular buffer with read-write pointers, or a double-buffering algorithm

Sample Playback and Manipulation

- Three basic operations you should know
  - **Panning** is the attenuation of left and right channels of a mixed sound
    - Results in spatial positioning within the aural stereo field
  - **Pitch** allows the adjustment of a sample's playback frequency in real-time
  - **Volume** control typically attenuates the volume of a sound
    - Amplification is generally never supported

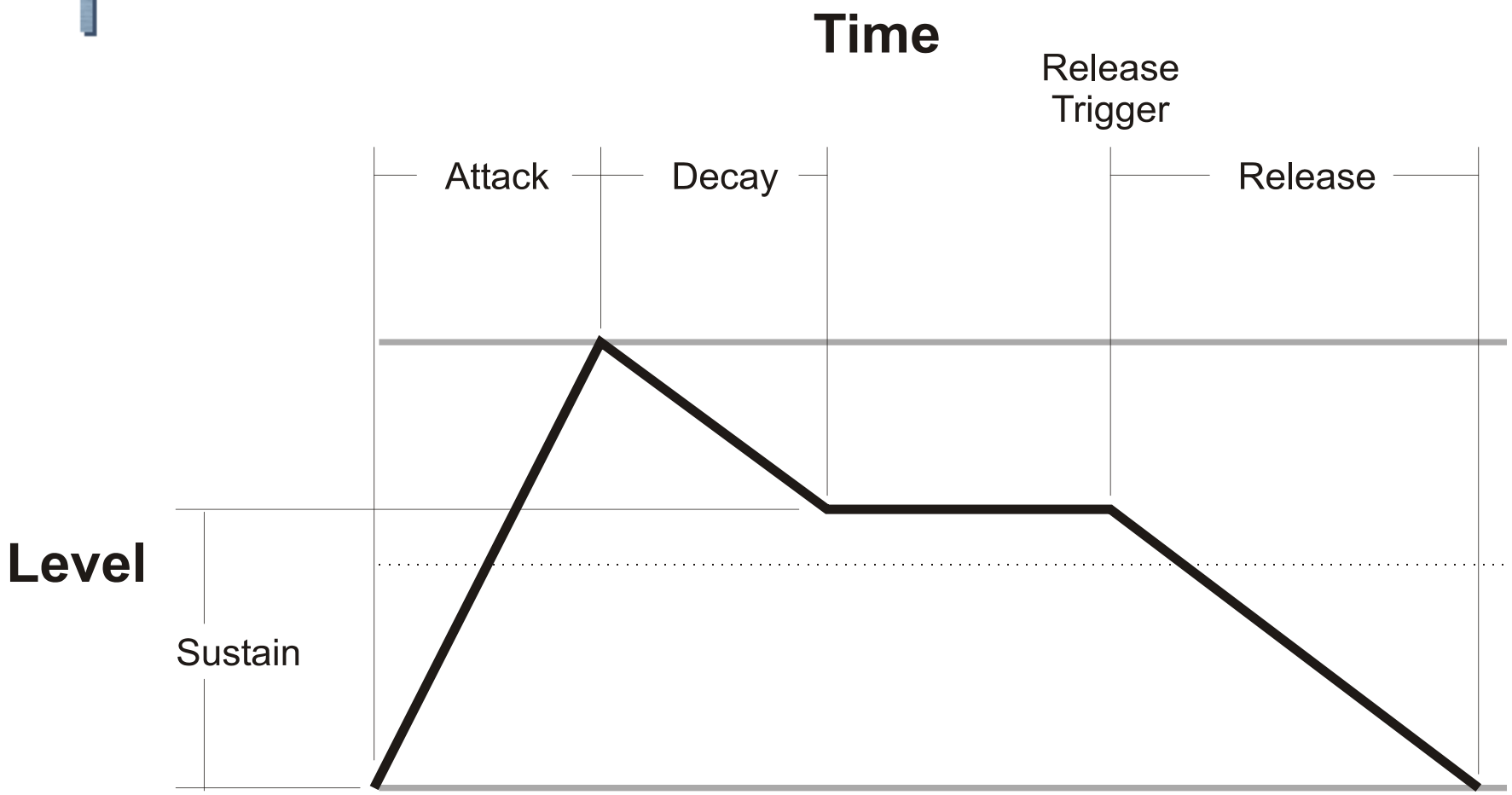
Compressed Audio Format

- Compressed audio formats allow sound and music to be stored more compactly
  - Bit reduction codecs generally are lightweight
    - ADPCM compression is implemented in hardware on all the major current video game console systems
  - Psycho-acoustic codecs often have better compression
    - Require substantially more computational horsepower to decode

MP3, Ogg Vorbis, Licensing & Patent Issues

- The MP3 format is patented
  - Any commercial game is subject to licensing terms as determined by Fraunhofer & Thompson Multimedia, the holders of the patents
- Ogg Vorbis is similar to MP3 in many ways
  - Open source and patent-free (royalty-free)
- Be aware of patent and license restrictions when using 3rd party software

ADSR Envelope (Attack, Decay, Sustain, Release)



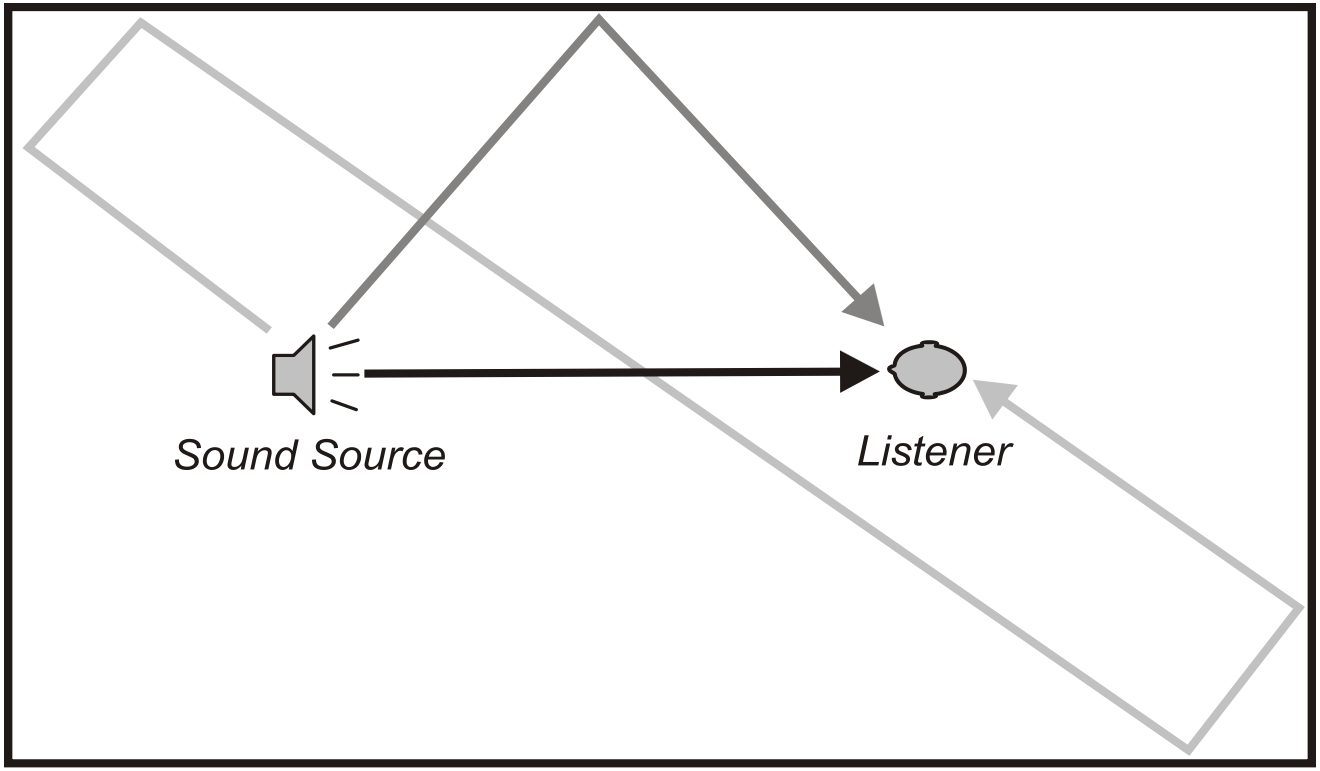
3D Audio

- Two sets of data required when working in world coordinates:
  - Listener Data
    - Composed of world position and orientation (virtual microphone in the world)
  - Source Data
    - Composed of sound position, orientation, velocity, etc (virtual sound source in the world)

Environmental Effects

- Environmental effects nearly always implemented in hardware
- Sound transmission is categorized in three ways
  - Direct transmission
  - Early reflections (echo)
  - Late reflections (reverberation)

Sound Transmission Categories



Direct Path

Echo

Reverberation

Environmental Effects Standards

- EAX 2.0 and beyond
  - EAX 2.0 developed by Creative Labs and released as an open standard
  - EAX 3.0 and 4.0 remain proprietary Creative Labs standards
- I3DL2
  - Open standard developed by IA-SIG, similar to EAX 2.0 in functionality

Programming Music Systems

- Two common music systems
  - MIDI-based systems
    - (Musical Instrument Digital Interface)
  - Digital audio streaming systems
    - (CD audio, MP3 playback, etc)

Advantages and Disadvantages of MIDI

- Actual music data size is negligible
- Easy to control, alter, and even generate in real-time
- High quality music is more difficult to compose and program
- Only effective if you can guarantee playback of a common instrument set

Other MIDI-based technologies to be aware of

- DLS (DownLoadable Sound) Format
  - A standardized format for instrument definition files
- iXMF (Interactive eXtensible Music Format)
  - New proposed standard for a container format for interactive music

Advantages / Disadvantages of Digital Audio Streams

- Superb musical reproduction is guaranteed
- Allows composers to work with any compositional techniques
- Some potential interactivity is sacrificed for expediency and musical quality
- Generally high storage requirements

A Conceptual Interactive Music Playback System

- Divide music into small two to eight-bar chunks that we'll call *segments*.
- A network of transitions from segment to segment (including loops and branches) is called a *theme*.
- Playing music is now as simple as choosing a theme to play. The transition map tracks the details.

Advanced Audio Programming

- 3D Audio Environmental Effects Integration
- Audio Scripting and Engine Integration
- Lip-sync Technology
- Advanced Voice Playback
- Voice Recognition

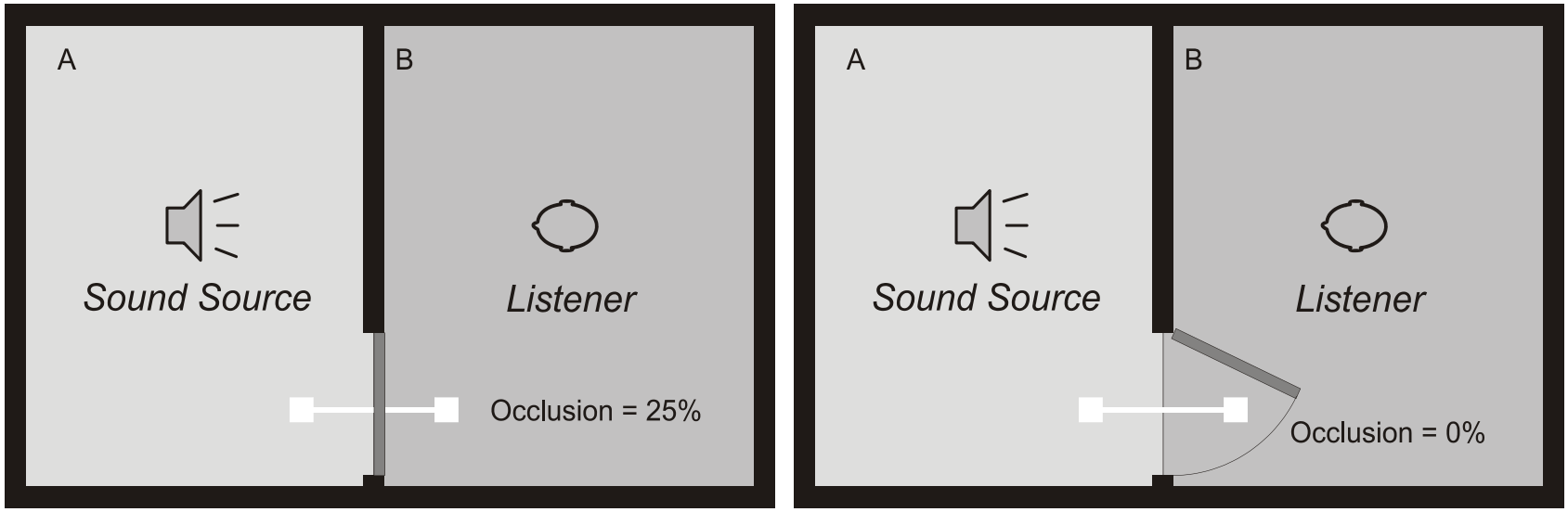
3D Audio Environmental Effects Integration

- Environmental effects should be driven by a room's shape and material composition.
  - Can determining the optimal effect settings be done automatically?
  - This may be important as game worlds become larger and more complex

3D Audio Environmental Effects Integration (cont)

- Sound occlusion and damping is a particularly difficult problem to solve
  - This is essentially a pathfinding problem for audio.
  - Doors can dynamically affect a sound's properties
  - Very few titles have even attempted a robust, general-purpose, and automated solution to these problems.

Dynamic Occlusion



Audio Scripting and Engine Integration

- Very little audio programming should be done by general game programmers
- Game Engine should offer robust support for audio triggers and scripts
- Engine should deal with audio scripts, not “sound files”
- Why is this so important?

Audio Scripting

- Many situations require much more information than can be embedded in a linear audio file
  - Sound Variation
  - Sound Repetition
  - Complex Sound Looping
  - Background Ambience

Lip-sync Technology

- Lip-sync technology is a blending of audio and visual techniques to create realistic-looking speech by in-game actors.
  - Simple techniques such as waveform amplitude measurement has worked previously, but…
  - In future titles, it will be considered inadequate.
  - Much work can still be done in this field.

Advanced Voice Playback

- Real-time spoken feedback is especially important in sports titles (simulated announcers)
- Game are reaching the limits of what current techniques (canned, prerecorded phrases combined in series) can provide.
- Again, this is an opportunity for future groundbreaking audio work.

Voice Recognition

- Spoken commands are much easier to deliver in certain situations.
- A great example of this? Squad-based tactical shooters.
- Current generation systems are still very error prone. A great opportunity for breakout audio technology.