Announcements

- Midterms?
- Lab #5 due today
- Lab #6 Tomorrow
- Writing Assignment #6 due tomorrow
- Quiz #3 Friday (Lectures 14-17)

Objectives

- Representing sound in a digital form
- Recording and playing sound
- Music synthesis
The Nature of Sound

- Sound is an analog phenomenon
  - Rapid vibrations (pressure fluctuations) travel as waves through the air (or through some other medium)
- What is a wave?
  - A variation that travels through some medium
  - We hear sound when air vibrations make our eardrums vibrate

Animations courtesy of Dr. Dan Russell, Kettering University
http://www.kettering.edu/~drussell/Demos/waves-intro/waves-intro.html

Vibrations can be described by two numbers:
- The vibration’s size, or amplitude
- And its speed, or frequency

- Larger-amplitude vibrations = louder sound
- Higher-frequency vibrations = higher pitch
  - Frequency usually described in hertz (Hz) – vibrations per sec
  - Frequency of Middle C is about 262 Hz
    - (262 cycles per second)
  - Highest C on the piano (“C8”) is 4186 Hz (4.186 KHz)
  - Lowest C on the piano (“C1”) is 33 Hz
The Nature of Sound

- Human hearing is typically in the range of about 20 Hz to 20,000 Hz (20 KHz)
  - Dogs can hear higher frequencies
- A ‘pure’ sound will have a single frequency
  - Usually electronic sounds (‘beeps’)
- Real (natural) sounds are complex combinations of many amplitudes and frequencies

The Nature of Sound

- The vibrations of sound are often called a sound wave
- These vibrations can be plotted on a graph, or sonogram
- A pure sound is a simple sound wave with a single frequency
  - The frequency dictates the sound’s pitch
  - The amplitude dictates its loudness

The Nature of Sound

- A natural sound is much more complex, with a combination of frequencies and amplitudes:
Digitizing Sound

- Like any analog information, sound must be digitized before a computer can use it
- Like image data, there are two steps for digitizing sound:
  1. Sampling
  2. Quantizing

Digitizing Sound

- Consider some arbitrary sound wave
  - We want to digitize it – determine a bunch of numbers that will represent the original sound wave

Digitizing Sound

- Step 1 – Sampling:
  - Choose a sample rate (number of samples per second)
Digitizing Sound

- Step 2 – Quantizing
  - Determine an amplitude value for each sample

The series of amplitudes is a digital approximation of the original sound wave.

Compare the digitized sound wave to the original
- Does this example have enough samples (a high enough sampling rate)?
Digitizing Sound

- A higher sampling rate gives a digital signal that is a closer approximation to the original
  - “Higher fidelity” – more “truthful” to the original

Sampling Rate

- When sampling sound, we can choose the sampling rate
- We can hear up to about 20 KHz (20,000 cycles per sec)
  - For full fidelity, sampling rate must be high enough to reproduce frequencies up to 20 KHz
- Nyquist’s Theorem:
  - Sample rate must be at least twice the desired frequency
- So, many digital audio formats (including CDs) are recorded at 44,100 Hz (44.1 KHz)
  - Most MP3 files have sampling rates lower than this
- Lower sampling rates give poorer-quality sound
  - Equivalent of “telephone” quality might be about 6–8 KHz

Limitations of Digitized Sound

- Digital audio is subject to the limitations of
  - Sampling rate
    - Number of discrete samples per second
  - Resolution
    - The accuracy of the quantized amplitude due to the number of levels of quantization – the bit depth
  - Dynamic range
    - The range of quantized amplitude values
When quantizing sounds, we have a choice of number of levels to use for amplitude values. The number of bits used to store amplitude data is called the bit depth. Using more bits gives us more amplitude levels.  

- 8-bit digital sound represents amplitudes using 8-bit integers (2^8 levels: 256 levels).
- 16-bit digital sound has 65,536 levels (2^16 = 65,536).
- A higher bit depth—a larger number of levels—allows for a higher resolution of the sound.
- A wider spacing of those levels corresponds to a greater dynamic range.

Resolution and Sampling Rate

- 5 bits per sample
- 28 samples (both doubled)
Dynamic Range

- The *dynamic range* is the difference between the softest and loudest sounds that can be correctly reproduced
  - If a sound requires a larger amplitude than we can store, we lose information (*clipping* occurs)

Audio Compression

- Both lossless and lossy compression techniques are used on audio files
  - Four minutes of CD audio (44.1 kHz, 16-bit stereo) requires about 40MB of disk space
    - WAV files use an uncompressed audio format
  - In an MP3 file encoded at 128 kbps, a four-minute song requires less than 4MB of space
    - MP3 ("MPEG-1 Audio Layer 3") uses lossy compression
  - Similar to JPEG, the MP3 format takes advantage of the sensitivity of human hearing
    - Resolution is reduced in parts of the sound that are less likely to be noticed

Digital Audio

- A good overview of digital audio:
Playing and Recording Sound

- Circuits built in to your computer handle the conversion between sound waves and digital files
  - Analog-to-digital converter (ADC)
    - A microphone turns the analog vibrations into an analog electrical signal
    - An ADC converts that analog electrical signal to a digital file (a series of numbers)
  - Digital-to-analog converter (DAC)
    - A speaker requires an analog electrical signal to generate sound
    - A DAC converts the digital file back into the analog electrical signal that drives the speaker

Editing Digitized Sound

- Digitized sound is just a bunch of numbers
  - Numbers can be changed by using math
- So, digital sound can be changed by applying mathematical formulas to the numerical sound data
  - Change the sound’s frequency (pitch)
  - Change the sound’s amplitude (volume)
  - Add echoes and other special effects
  - Remove noise
  - Isolate individual speakers or instruments
- New sounds can be created by combining or otherwise changing existing sounds
Synthesizing Music

- Music is a series of sounds
  - Each note has a particular pitch, or frequency
  - Each instrument has a particular sound quality
  - Natural musical sounds are complex wave forms with many frequencies combined
- Pure sounds – simple wave forms – can be used to make simple music
  - Analog electronic circuits can create simple wave forms
  - These circuits are known as oscillators

Simple Oscillators

- Different kinds of simple oscillations are possible
  - Each sounds slightly different
  - Some examples

  - Sine waves
  - Square waves
  - Sawtooth waves

Synthesizing Music

- Simple wave forms of different frequencies, shapes and amplitudes can be combined to synthesize different sounds
  - This combination of waves sounds like a string being plucked:
  - This is a completely synthetic sound, created from simple oscillators
Synthesizing Music

- Integrated synthesis starts with a natural sound
  - Often called “Sampling”
  - A sample of a natural sound is recorded
    - One note on a piano
    - One pluck of a guitar string
    - One beat on a drum or cymbal
  - Different notes are created from the sample by changing the frequency
    - Can assign the different notes to keys of a keyboard

Modifying Vocals

- Similar techniques can be applied to voices
  - Many vocal recordings are modified to ‘correct’ errors in the singer’s pitch
  - Software such as “Auto-Tune” by Antares Audio Technologies
    - or “Melodyne” by Celemony
  - Some ‘before’ and ‘after’ examples are at
    - [antarestech.com/products/avp.shtml](http://antarestech.com/products/avp.shtml)
  - Also sometimes used for special effects
    - (such as in Cher’s song *Believe*... or Daft Punk’s *One More Time*)

Modifying Vocals

- Clip From NPR’s Day to Day
- David Was discussing modified vocals in music
**MIDI Instruments and Devices**

- MIDI – a standard interface between electronic musical instruments, synthesizers and keyboards
  - *Musical Instrument Digital Interface*
- MIDI-capable devices can communicate with each other
- Advantages to MIDI:
  - MIDI is just the instructions for the sounds, not the entire recorded sound (much smaller file size)
  - Sound is created on the output device – with whatever quality is available (flexibility)
  - Files can be easily edited and mixed for multiple tracks

---

**Computer Speech**

- Much current research on natural language processing
- Three separate issues:
  1. Speech synthesis (making a computer speak understandably)
  2. Speech recognition (making a computer recognize spoken words)
  3. Language comprehension (making a computer understand natural language)
- Reasons why computer speech might be valuable
  - Interfaces for the disabled
  - Voice-activated commerce and information lines
  - Informational announcements and warnings
  - Etc.

---

**Speech Synthesis**

- Speech synthesis is the easiest of the three problems
  - The input is text characters (unambiguous and consistent)
- One ‘obvious’ approach:
  - a recorded ‘dictionary’ of every possible word
  - Takes up a lot of space
  - Does not allow for context-dependent pronunciation differences
  - Homographs – how to choose?
  - Limited vocabulary
    - Especially problematic with names, jargon, technical terms, etc.
  - Poor performance when each recorded word requires individual access from disk
Speech Synthesis

- A better approach: phonemes
  - English uses only about 50 phonemes for all possible words
  - Not many more are needed to accommodate most other languages
  - Can be quickly accessed
- Still have to look at the context of the words to determine the correct pronunciation and inflection
  - Tough, through, bough, cough, thorough
  - Dr. (doctor) vs. Dr. (drive)
  - Desert vs. desert vs. dessert
  - “Would you desert me in the desert with nothing but dessert?”
  - Inflection changes at sentence beginning or end
- Quick quiz: What is the pronunciation of “ghoti”? 
  - “gh” as in “cough”, “o” as in “women”, “ti” as in “nation”

Phonemes (American English)

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Examples</th>
<th>Synonyms</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>pie</td>
<td>pie, cat</td>
<td>pie, cat</td>
</tr>
<tr>
<td>t</td>
<td>tip</td>
<td>tip, pit</td>
<td>tip, pit</td>
</tr>
<tr>
<td>k</td>
<td>kill, cat</td>
<td>kill, cat</td>
<td>kill, cat</td>
</tr>
<tr>
<td>g</td>
<td>get</td>
<td>get, get</td>
<td>get, get</td>
</tr>
<tr>
<td>m</td>
<td>mama</td>
<td>mom, mama</td>
<td>mom, mama</td>
</tr>
<tr>
<td>n</td>
<td>noon, -en</td>
<td>noon, -en</td>
<td>noon, -en</td>
</tr>
<tr>
<td>b</td>
<td>bug, -en</td>
<td>bug, -en</td>
<td>bug, -en</td>
</tr>
<tr>
<td>v</td>
<td>vise, -en</td>
<td>vise, -en</td>
<td>vise, -en</td>
</tr>
<tr>
<td>s</td>
<td>she, she</td>
<td>she, she</td>
<td>she, she</td>
</tr>
<tr>
<td>z</td>
<td>zip</td>
<td>zip, -ep</td>
<td>zip, -ep</td>
</tr>
<tr>
<td>th</td>
<td>this, the</td>
<td>this, the</td>
<td>this, the</td>
</tr>
<tr>
<td>f</td>
<td>fish, feb</td>
<td>fish, feb</td>
<td>fish, feb</td>
</tr>
<tr>
<td>sh</td>
<td>she, she</td>
<td>she, she</td>
<td>she, she</td>
</tr>
<tr>
<td>ng</td>
<td>long</td>
<td>long, -ing</td>
<td>long, -ing</td>
</tr>
<tr>
<td>1</td>
<td>look, lo</td>
<td>look, lo</td>
<td>look, lo</td>
</tr>
<tr>
<td>r</td>
<td>ro, re</td>
<td>ro, re</td>
<td>ro, re</td>
</tr>
<tr>
<td>l</td>
<td>liv, liv</td>
<td>liv, liv</td>
<td>liv, liv</td>
</tr>
<tr>
<td>v</td>
<td>fall, fall</td>
<td>fall, fall</td>
<td>fall, fall</td>
</tr>
<tr>
<td>t</td>
<td>tap, tap</td>
<td>tap, tap</td>
<td>tap, tap</td>
</tr>
<tr>
<td>k</td>
<td>kick, -ik</td>
<td>kick, -ik</td>
<td>kick, -ik</td>
</tr>
<tr>
<td>g</td>
<td>gog</td>
<td>gog, -ik</td>
<td>gog, -ik</td>
</tr>
</tbody>
</table>

Speech Synthesis

- Most personal computers come with basic speech synthesis and speech recognition built-in
- Some interesting tools for trying out speech synthesis:
  - The ‘speech’ control panel
  - The ‘language bar’
- Start ➔ Programs ➔ Accessories ➔ Accessibility ➔ Narrator
- http://naturalvoices.att.com/demos/ (AT&T Research Labs)
Speech Recognition

• A much harder problem than speech synthesis
• Unlike text, the input — natural speech — is highly variable and ambiguous
  - Different speakers
  - Different inflections
  - Different emotions
  - Words run together
  - Homographs and homophones
    - E.g. “I scream” vs. “ice cream”
  - Punctuation, formatting, numbers, etc.

• Two major types of speech recognition
  1. Limited vocabulary
  2. Speaker dependent

Limited-Vocabulary Speech Recognition

• Designed to recognize speech from many different speakers and environments
• Useful in controlled situations
  - Voice-activated computers
  - Telephone response systems
  - Etc.
• Progress being made, but still limited use
**Speaker-Dependent Speech Recognition**

- Dictation systems
  - Commonly ‘trained’ to an individual’s voice
  - Compare possible words to a dictionary of real words
  - Specialized dictionaries required for recognizing technical or medical jargon or other specialized vocabularies
  - Performance can be improved by analyzing parts of speech (verbs, nouns, etc.) and the relative frequency of specific word pairs

**Speaker-Dependent Speech Recognition**

- Can try out speech recognition on your laptop:
  - Windows’ Language Bar
    - (Also includes a handwriting recognition system)

**Natural Language Comprehension**

- Natural languages are ambiguous
  - Whether from text or from recognized speech, can a computer
    - Understand the language?
    - Act on the content?
    - Translate between languages?
      - E.g. [www.google.com/language_tools](http://www.google.com/language_tools)
Natural Language Ambiguities

- **Lexical ambiguity**—created when words have multiple meanings
  - *Time flies like an arrow*
- **Syntactic ambiguity**—created when sentences can be constructed in various ways
  - *They are hunting dogs*
- **Referential ambiguity**—created when pronouns could be applied to multiple objects
  - *The brick fell on the computer but it is not broken*
- **Inflectional ambiguity**—different verbal inflections (not present in text-only forms)
  - *I never said she stole my money*
    - *I never said she stole my money* (I simply didn't even say it)
    - *I never said she stole my money* (I might have implied it in some way, but never explicitly).
    - *I never said she stole my money* (I never said she stole it).
    - *I never said she stole my money* (I didn't say it).
    - *I never said she stole my money* (I just said she probably borrowed it).
    - *I never said she stole my money* (I said she took something of mine, but not my money).