Hands On: Multimedia Methods for Large Scale Video Analysis (Lecture)

Dr. Gerald Friedland, fractor@icsi.berkeley.edu
Today

- Answers to Questions
Today

- Answers to Questions
- What is needed for large scale video analysis?
Today

• Answers to Questions
• What is needed for large scale video analysis?
• Start: Sound
Today

• Answers to Questions
• What is needed for large scale video analysis?
• Start: Sound
  • What is it?
Today

• Answers to Questions
• What is needed for large scale video analysis?
• Start: Sound
  • What is it?
  • How is it recorded and stored?
Today

• Answers to Questions
• What is needed for large scale video analysis?
• Start: Sound
  • What is it?
  • How is it recorded and stored?
  • What are it’s most important properties (to us)?
FAQ
FAQ

- Can we use YouTube videos
FAQ

• Can we use YouTube videos
  ➡ You can use the YouTube API. Downloading YouTube videos is at your own risk.
FAQ

- Can we build a product as our project?
FAQ

• Can we build a product as our project?
  ➡ Sure.
FAQ
FAQ

• How is the book related to the class?
FAQ

- How is the book related to the class?
  ➔ The book is additional material that is especially good to get a big picture overview of the multimedia field.
FAQ

• How is the book related to the class?
  ➡ The book is additional material that is especially good to get a big picture overview of the multimedia field.
  ➡ The book is NOT a substitute for the lectures.
FAQ

• Does the project need to include work that includes all media?
FAQ

• Does the project need to include work that includes all media?
  ➡ No. It’s your own choice how many different media streams to include.
FAQ
FAQ

• Are we able to work on the project if we have no prior knowledge in multimedia content analysis?
FAQ

• Are we able to work on the project if we have no prior knowledge in multimedia content analysis?

➡ Yes. Remember to concentrate on using available tools, such as feature extractors, machine learning engines, and scoring scripts. All of these are already installed at ICSI.
FAQ
FAQ

• How can the class teach all of computer vision, acoustic processing, NLP, and the other topics? This is too much!
FAQ

• How can the class teach all of computer vision, acoustic processing, NLP, and the other topics? This is too much!

➡ True. The class will only give an introduction to these topics as the focus is on scalability. However, this will be enough to create your own project using available tools. If you have further expertise you are free to apply it!
FAQ
FAQ

• I have a project idea. Can we talk?
FAQ

• I have a project idea. Can we talk?
  ➡Yes, please. Best thing is to follow the following guideline:
    1) Send me an abstract in an email
    2) Schedule a time to talk
    3) Start your project.
What’s needed for large scale video analysis
What’s needed for large scale video analysis

- Depends on the task
What’s needed for large scale video analysis

- Depends on the task
- Available data:
  - video (images, optical flow)
  - audio (speech, music, noise)
  - human generated metadata (tags, title, description)
  - sensoric metadata (GPS, focus, camera-type)
What’s needed for large scale video analysis

- Depends on the task
- Available data:
  - video (images, optical flow)
  - audio (speech, music, noise)
  - human generated metadata (tags, title, description)
  - sensoric metadata (GPS, focus, camera-type)
- Problem: All data streams uncertain for most tasks!
Current Situation

- Multimedia Computing
- Computer Vision
- Speech Processing
- Music Processing
- Computer Listening
- Natural Language Processing
- CASA

Area being worked on

Area not being worked on
What are the next lectures about?

- Computer Vision
- Multimedia Computing
- Natural Language Processing
- Speech Processing
- Computer Listening
- Music Processing
- CASA

Next Lectures

- Area being worked on
- Area not being worked on
Introduction to Sound
Introduction to Sound

• What is sound?
Introduction to Sound

• What is sound?
• How is it recorded and stored?
Introduction to Sound

• What is sound?
• How is it recorded and stored?
• What are it’s most important properties (to us)?
Introduction to Sound

- What is sound?
- How is it recorded and stored?
- What are its most important properties (to us)?
- Introduction to features 1
Introduction to Sound

- What is sound?
- How is it recorded and stored?
- What are it’s most important properties (to us)?
- Introduction to features 1
- Some frameworks and tools to work with sound
What is Sound?
What is Sound?

Video from ViHart (Youtube):
What is Sound?

Video from ViHart (Youtube):
http://www.youtube.com/watch?v=i_0DXxNeaQ0
What is Sound?
What is Sound?

“a traveling wave which is an oscillation of pressure transmitted through a solid, liquid, or gas, composed of frequencies within the range of hearing and of a level sufficiently strong to be heard, or the sensation stimulated in organs of hearing by such vibrations.” (AHD)
Visualizations of Sound
Visualizations of Sound

Time Domain aka Amplitude Space aka Waveform
Visualizations of Sound
Visualizations of Sound

Frequency Domain aka Fourier Space aka Spectrum
Visualizations of Sound
Visualizations of Sound

Spectrogram
Hearing Spectrum

Source: http://sound.westhost.com/articles/fadb.htm
dB SPL?

- decibel Sound Pressure Level
- NOT a physical unit, only a scale

\[
L_p = 10 \log_{10} \left( \frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \right) = 20 \log_{10} \left( \frac{p_{\text{rms}}}{p_{\text{ref}}} \right) \text{ dB},
\]

where \( p_{\text{ref}} \) is the reference sound pressure and \( p_{\text{rms}} \) is the rms sound pressure being measured.
bits -> db Range (Cheat sheet)

- 8 bits -> 48 dB SPL
- 11 bits -> 66 dB SPL
- 16 bits -> 96 dB SPL
- 24 bits -> 144 dB SPL
Figure 1. Sound pressure A-weighting scheme according to IEC 61672:2003.

Sound pressure levels weighted by the A scheme are usually labelled as dBA or dB(A). Please note that dB and dBA, like a percent symbol '%', define ratios and not physical units of measurement. A value of 10 dB can refer to completely different sound pressure levels depending on the reference. Also there are no physical units associated with dB.

Observed Properties of Sound

As explained in the previous paragraph, sound is a pressure wave traveling through a medium. In practice, sounds are not exclusively traveling in a homogenous medium from a source to exhaustion. The environment is filled with objects, sometimes sounds are produced in a closed room, and sounds pressure waves may collide with other sounds. The resulting effects of these conditions play a large role when designing multimedia systems. Also, the effects on sound are more significant than on light waves. The three most important ones are echo, reverberation, and interference.

An echo is a reflection of sound, arriving at the listener some time after the original sound. Typical examples are the echo produced by the bottom of a well, by a building, or by the walls of an enclosed room. Sounds is very easily reflected by most materials so echos are always present in every environment. A true echo is a single reflection of the sound source. Mostly, however, many echoes form reverberation. The time delay is the extra distance divided by the speed of sound. When dealing with audible frequencies, the human ear cannot distinguish an echo from the original...
How is Sound Recorded?

Not surprisingly today's sound recording still obeys the same principles with two main exceptions: First, the sound waves are converted to electrical waves by a microphone and second, most of today's storage media is digital, i.e. sound waves are converted into binary numbers before they are imprinted on the medium. The media themselves, such as CDROM or DAT are a bit more sophisticated than Edison's cylinders. Having said that, we are currently observing the replacement of all of these specialized media with generic media, such as harddisks and flash memory. We therefore decided not to explain the technical details of these, the reader is referred to the bibliography for further information. The next paragraphs, however, will explain the governing principles of modern sound processing.

Microphones

A microphone is an acoustic sensor that converts sound into an electrical signal. The general principle is that sound pressure is inflicted on a membrane which varies it's electrical resistance according to the movement. Most microphones in use today for audio use electromagnetic induction (dynamic microphone) by letting the membrane swing a magnetic field produced by a coil, capacitance change (condenser microphone) by letting the membrane be part of a capacitor which varies capacity with movement, or piezoelectric generation (piezo crystals emit electricity when under pressure). Some modern microphones use light modulation to produce the electric signal by "watching" the mechanical vibration (laser microphones). A single dynamic membrane will not respond linearly to all audio frequencies. Some microphones for this reason utilize multiple membranes for the different parts of the audio spectrum and then combine the resulting signals. The different microphone types have different electrical properties. A complete
How is Sound Recorded?

Not surprisingly today's sound recording still obeys the same principles with two main exceptions: First, the sound waves are converted to electrical waves by a microphone and second, most of today's storage media is digital, i.e., sound waves are converted into binary numbers before they are imprinted on the medium. The media themselves, such as CDROM or DAT are a bit more sophisticated than Edison's cylinders. Having said that, we are currently observing the replacement of all of these specialized media with generic media, such as hard disks and flash memory. We therefore decided not to explain the technical details of these, the reader is referred to the bibliography for further information. The next paragraphs, however, will explain the governing principles of modern sound processing.

Microphones

A microphone is an acoustic sensor that converts sound into an electrical signal. The general principle is that sound pressure is inflicted on a membrane which varies its electrical resistance according to the movement. Most microphones in use today for audio use electromagnetic induction (dynamic microphone) by letting the membrane swing a magnetic field produced by a coil, capacitance change (condenser microphone) by letting the membrane be part of a capacitor which varies capacity with movement, or piezoelectric generation (piezo crystals emit electricity when under pressure). Some modern microphones use light modulation to produce the electric signal by "watching" the mechanical vibration (laser microphones). A single dynamic membrane will not respond linearly to all audio frequencies. Some microphones for this reason utilize multiple membranes for the different parts of the audio spectrum and then combine the resulting signals. The different microphone types have different electrical properties. A complete
How is Sound Recorded?

Not surprisingly today's sound recording still obeys the same principles with two main exceptions:

First, the sound waves are converted to electrical waves by a microphone and second, most of today's storage media is digital, i.e. sound waves are converted into binary numbers before they are imprinted on the medium. The media themselves, such as CDROM or DAT are a bit more sophisticated than Edison's cylinders. Having said that, we are currently observing the replacement of all of these specialized media with generic media, such as harddisks and flash memory. We therefore decided not to explain the technical details of these, the reader is referred to the bibliography for further information. The next paragraphs, however, will explain the governing principles of modern sound processing.

Microphones

A microphone is an acoustic sensor that converts sound into an electrical signal. The general principle is that sound pressure is inflicted on a membrane which varies its electrical resistance according to the movement. Most microphones in use today for audio use electromagnetic induction (dynamic microphone) by letting the membrane swing a magnetic field produced by a coil, capacitance change (condenser microphone) by letting the membrane be part of a capacitor which varies capacity with movement, or piezoelectric generation (piezo crystals emit electricity when under pressure). Some modern microphones use light modulation to produce the electric signal by “watching” the mechanical vibration (laser microphones). A single dynamic membrane will not respond linearly to all audio frequencies. Some microphones for this reason utilize multiple membranes for the different parts of the audio spectrum and then combine the resulting signals. The different microphone types have different electrical properties. A complete
How is Sound Recorded?

Not surprisingly today’s sound recording still obeys the same principles with two main exceptions: First, the sound waves are converted to electrical waves by a microphone and second, most of today’s storage media is digital, i.e. sound waves are converted into binary numbers before they are imprinted on the medium. The media themselves, such as CD-ROM or DAT are a bit more sophisticated than Edison’s cylinders. Having said that, we are currently observing the replacement of all of these specialized media with generic media, such as harddisks and flash memory. We therefore decided not to explain the technical details of these, the reader is referred to the bibliography for further information. The next paragraphs, however, will explain the governing principles of modern sound processing.

Microphones

A microphone is an acoustic sensor that converts sound into an electrical signal. The general principle is that sound pressure is inflicted on a membrane which varies its electrical resistance according to the movement. Most microphones in use today for audio use electromagnetic induction (dynamic microphone) by letting the membrane swing a magnetic field produced by a coil, capacitance change (condenser microphone) by letting the membrane be part of a capacitor which varies capacity with movement, or piezoelectric generation (piezo crystals emit electricity when under pressure). Some modern microphones use light modulation to produce the electric signal by “watching” the mechanical vibration (laser microphones). A single dynamic membrane will not respond linearly to all audio frequencies. Some microphones for this reason utilize multiple membranes for the different parts of the audio spectrum and then combine the resulting signals. The different microphone types have different electrical properties. A complete...
How is Sound Recorded?

Not surprisingly, today's sound recording still obeys the same principles with two main exceptions: first, the sound waves are converted to electrical waves by a microphone; and second, most of today's storage media is digital, i.e., sound waves are converted into binary numbers before they are imprinted on the medium. The media themselves, such as CD-ROM or DAT, are a bit more sophisticated than Edison's cylinders. Having said that, we are currently observing the replacement of all of these specialized media with generic media, such as hard disks and flash memory. We therefore decided not to explain the technical details of these; the reader is referred to the bibliography for further information.

The next paragraphs will explain the governing principles of modern sound processing.

**Microphones**

A microphone is an acoustic sensor that converts sound into an electrical signal. The general principle is that sound pressure is inflicted on a membrane which varies its electrical resistance according to the movement. Most microphones in use today for audio use electromagnetic induction (dynamic microphone) by letting the membrane swing a magnetic field produced by a coil, capacitance change (condenser microphone) by letting the membrane be part of a capacitor which varies capacity with movement, or piezoelectric generation (piezo crystals emit electricity when under pressure). Some modern microphones use light modulation to produce the electric signal by "watching" the mechanical vibration (laser microphones). A single dynamic membrane will not respond linearly to all audio frequencies. Some microphones for this reason utilize multiple membranes for the different parts of the audio spectrum and then combine the resulting signals. The different microphone types have different electrical properties. A complete...
How is Sound Recorded?

Not surprisingly today's sound recording still obeys the same principles with two main exceptions: First, the sound waves are converted to electrical waves by a microphone and second, most of today's storage media is digital, i.e. sound waves are converted into binary numbers before they are imprinted on the medium. The media themselves, such as CD-ROM or DAT are a bit more sophisticated than Edison's cylinders. Having said that, we are currently observing the replacement of all of these specialized media with generic media, such as harddisks and flash memory. We therefore decided not to explain the technical details of these, the reader is referred to the bibliography for further information. The next paragraphs, however, will explain the governing principles of modern sound processing.

Microphones

A microphone is an acoustic sensor that converts sound into an electrical signal. The general principle is that sound pressure is inflicted on a membrane which varies its electrical resistance according to the movement. Most microphones in use today for audio use electromagnetic induction (dynamic microphone) by letting the membrane swing a magnetic field produced by a coil, capacitance change (condenser microphone) by letting the membrane be part of a capacitor which varies capacity with movement, or piezoelectric generation (piezo crystals emit electricity when under pressure). Some modern microphones use light modulation to produce the electric signal by "watching" the mechanical vibration (laser microphones). A single dynamic membrane will not respond linearly to all audio frequencies. Some microphones for this reason utilize multiple membranes for the different parts of the audio spectrum and then combine the resulting signals. The different microphone types have different electrical properties. A complete...
How is Sound Recorded?

Not surprisingly today's sound recording still obeys the same principles with two main exceptions:

First, the sound waves are converted to electrical waves by a microphone and second, most of today's storage media is digital, i.e. sound waves are converted into binary numbers before they are imprinted on the medium. The media themselves, such as CD-ROM or DAT are a bit more sophisticated than Edison's cylinders. Having said that, we are currently observing the replacement of all of these specialized media with generic media, such as harddisks and flash memory. We therefore decided not to explain the technical details of these, the reader is referred to the bibliography for further information. The next paragraphs, however, will explain the governing principles of modern sound processing.

Microphones

A microphone is an acoustic sensor that converts sound into an electrical signal. The general principle is that sound pressure is inflicted on a membrane which varies it's electrical resistance according to the movement. Most microphones in use today for audio use electromagnetic induction (dynamic microphone) by letting the membrane swing a magnetic field produced by a coil, capacitance change (condenser microphone) by letting the membrane be part of a capacitor which varies capacity with movement, or piezoelectric generation (piezo crystals emit electricity when under pressure). Some modern microphones use light modulation to produce the electric signal by "watching" the mechanical vibration (laser microphones). A single dynamic membrane will not respond linearly to all audio frequencies. Some microphones for this reason utilize multiple membranes for the different parts of the audio spectrum and then combine the resulting signals. The different microphone types have different electrical properties. A complete...
Modern Microphone

Cross-Section of Dynamic Microphone

Source: http://www.mediacollege.com/audio/microphones/dynamic.html
Types of Microphones

- **Nearfield**: Close to sound source e.g., headset, boom microphone (movies, TV productions), singer microphones

- **Farfield**: Further away from sound source e.g., lapel microphone, stationary microphone, webcams, handheld cams.
Difference
Farfield/Nearfield

• Nearfield: More energy, less distortion, captures sound source well.

• Farfield: Captures environment with sound source, “better for forensics”, processing often slower.

Demo: http://www.icsi.berkeley.edu/Speech/mr/nearfar.html
Microphone Directionality

A microphone's directionality indicates how sensitive it is to sounds arriving at different angles about its central axis. The directionality of a microphone is usually visualized using a polar pattern. Polar patterns represent the location of points that produce the same signal level output in the microphone if a constant sound pressure level is generated from that point. Figure 4 shows some idealized example patterns. The patterns are considered idealized because in the real world, polar patterns are a function of frequency. Manufacturer’s diagrams therefore usually include multiple plots at different frequencies. Also, while an omnidirectional microphone’s response is generally considered to be a perfect sphere in three dimensions. In the real world, this is not the case.
Digitization of Sound

Digitizing is the representation of a signal by a discrete set of its samples. Instead of representing the sound signal by an electrical current proportional to its sound pressure, the signal is represented by on-off patterns that represent sample values of the analog signal at certain fixed points. The on-off patterns are much less susceptible to the distortions outlined above, especially copying is usually lossless. Conceptually, digitization works in two parts, illustrated in Figure 6.

- Discretization: The analog signal is read at regular time intervals (sampling rate), sampling the value of the signal at that point in time. One such reading is called a sample.
- Quantization: Samples are rounded to a fixed set of numbers (such as integers), a process known as quantization.

A series of quantized samples can be transformed back into an analog output that approximates the original analog representation by generating the signal represented by each sample. The sampling rate and the number of bits used to represent the sample values determine how close such an approximation to the analog signal a digitization will be.

The error introduced by the quantization is called quantization noise and affects how accurately the amplitude can be represented. Very few bits for the samples will result in the signal only being represented coarsely and will affect the perceived dynamic of the sound as well as introduce high-frequency artifacts. Typical bit representations for audio are 8, 16, and 24 bits.

The error introduced by the sampling rate is called discretization error and determines the maximum frequency that can be represented in the signal. This upper frequency limit is determined by the so-called Nyquist frequency. The Nyquist frequency, named after the Swedish-American engineer Harry Nyquist or the Nyquist–Shannon sampling theorem, is half the sampling rate.
Remember: Nyquist Limit!

Math: See Draft Chapter 3 of Friedland & Jain on mm-creole.org
Common Recording Resolutions

- 8000Hz, 8-bit log. companded ~ 11 bit uncompanded (a/µ-law): telephone
- 16000Hz, 16-bit linear: speech (Skype)
- 44100Hz, 16-bit linear, stereo: Compact Disk, many camcorders
- 48000Hz, 32-bit linear, stereo: Digital Audio Tape, Hard Disk Recorders
μ-law Companding

Graph showing the relationship between input and output amplitudes.
Next Week

• Intro to Audio Analysis