Sound

Basics

✧ Sound is a sequence of waves of pressure which propagates through compressible media such as air or water.
✧ Digital representation of an analog signal
  ✧ Sampling
  ✧ Quantization
✧ Parameters:
  ✧ Sampling Rate (Samples per Second)
  ✧ Quantization Levels (Bits per Sample)
✧ This is a form of coding too:
  ✧ Pulse-code modulation (PCM)
Pulse-code Modulation (PCM)

Why Call it PCM?

4-bit PCM

<table>
<thead>
<tr>
<th>DECIMAL NUMBER</th>
<th>BINARY EQUIVALENT</th>
<th>PULSE-CODE WAVE FORMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$2^3$ $2^2$ $2^1$ $2^0$</td>
<td>$2^3$ $2^2$ $2^1$ $2^0$</td>
</tr>
<tr>
<td>0</td>
<td>0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 0 0 1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0 0 1 0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 1 0 0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0 1 1 0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 1 1 1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1 0 0 1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1 0 1 0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1 1 0 1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1 1 1 0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1 1 1 0</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1 1 1 1</td>
<td></td>
</tr>
</tbody>
</table>
Audio: Sampling & Quantization

How to choose proper...

- **Sampling Rate**
  - 8 Khz?
- **Quantization Level**
  - 8 bit/sample?

- Low Sampling rate: Aliasing, Low quality in reconstruction
- High Sampling rate: Data Redundancy, High storage, High processing power consumption
- Low Quantization Levels: Large Q. noise, low SNR.
- High Quantization levels: More bits required, High storage, High processing power consumption

- **Bit per Second for 8000 Hz 8 bit PCM**
  - 64 kbit/s
Audio, Sampling

- **Human Hearing Frequency Range**
  - **20 Hz to 20 kHz**
  - Most people will find that their hearing is most sensitive around 1-4 kHz and that it is less sensitive at high and low frequencies.
  - Play with “Audacity” tone generator to test your hearing

Play with Audacity tone generator to test your hearing: audacity.sourceforge.net
Audio, Sampling

Test your own hearing range

3 Sec. tones with different frequencies.

10 Hz  20 Hz  30 Hz  40 Hz  50 Hz  75 Hz  100 Hz  200 Hz  500 Hz

1 KHz  2 KHz  4 KHz  8 KHz  12 KHz  14 KHz  15 KHz  16 KHz  18 KHz
Audio, Sampling

Hearing Range

<table>
<thead>
<tr>
<th>Animal</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferret</td>
<td>16-44,000</td>
</tr>
<tr>
<td>Cow</td>
<td>23-35,000</td>
</tr>
<tr>
<td>Cat</td>
<td>45-64,000</td>
</tr>
<tr>
<td>Guinea Pig</td>
<td>54-50,000</td>
</tr>
<tr>
<td>Horse</td>
<td>55-33,500</td>
</tr>
<tr>
<td>Human</td>
<td>64-23,000</td>
</tr>
<tr>
<td>Dog</td>
<td>67-45,000</td>
</tr>
<tr>
<td>Sheep</td>
<td>100-30,000</td>
</tr>
<tr>
<td>Raccoon</td>
<td>100-40,000</td>
</tr>
<tr>
<td>Gerbil</td>
<td>100-60,000</td>
</tr>
<tr>
<td>Rat</td>
<td>200-76,000</td>
</tr>
<tr>
<td>Hedgehog</td>
<td>250-45,000</td>
</tr>
<tr>
<td>Rabbit</td>
<td>360-42,000</td>
</tr>
<tr>
<td>Possum</td>
<td>500-64,000</td>
</tr>
<tr>
<td>Mouse</td>
<td>1,000-91,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Animal</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bat</td>
<td>2000-110,000</td>
</tr>
<tr>
<td>Chinchilla</td>
<td>90-22,000</td>
</tr>
<tr>
<td>Elephant</td>
<td>16-12,000</td>
</tr>
<tr>
<td>Fur Seal</td>
<td>800-50,000</td>
</tr>
<tr>
<td>Beluga Whale</td>
<td>1000-123,000</td>
</tr>
<tr>
<td>Cal Sea Lion</td>
<td>450-50,000</td>
</tr>
<tr>
<td>Harp Seal</td>
<td>950-85,000</td>
</tr>
<tr>
<td>Harbour Porpoise</td>
<td>550-105,000</td>
</tr>
<tr>
<td>Killer Whale</td>
<td>800-13,500</td>
</tr>
<tr>
<td>Bottlenose Dolphin</td>
<td>90-105,000</td>
</tr>
</tbody>
</table>
Frequency Allocations

Radio Frequency Bands

9 KHz thr. 300 GHz
AM Radio
535 KHz thr. 1.6 MHz
FM Radio
88 MHz thr. 108 MHz
TV
Various bands from
54 MHz thr. 700 MHz
GSM (Global System for Mobile Communications)
Mostly 900 MHz and 1800 MHZ

United States radio spectrum frequency allocations chart as of 2011.
Amplitude Modulation (AM) conveys information over a carrier wave (the transmitted signal) by varying the amplitude (strength) of the carrier in relation to the information being sent (carrier's frequency remains constant).

Frequency Modulation (FM) works by varying carrier's instantaneous frequency.

Phase Modulation (PM) represents information as variations in the instantaneous phase of a carrier wave.

\[ y(t) = A_c \sin(\omega_c t + m(t) + \phi) \]

PM Example
Audio, Sampling

Human Vocal Range

- **Normal**: 80 Hz to 1100 Hz
- **Guinness Book of Records**
  - **Female**: Georgia Brown
    - Eight octaves
    - G2 (97.9989 Hz) thr. G10 (25087.7150 Hz)
  - **Male**: Tim Storms
    - Ten octaves (0.7973 Hz thr. 807.3 Hz)

Octave: In music, an octave is the interval between one musical pitch and another with half or double its frequency.

Play and see! Low freq. voice of Tim Storms, as you may can’t hear it.

www.guinnessworldrecords.com/world-records/1000/greatest-vocal-range-female

www.guinnessworldrecords.com/world-records/3000/greatest-vocal-range-male
Audio, Sampling

Common Sampling Rates

- **8,000 Hz** - Telephone, adequate for human speech.
- **11,025 Hz** – lower quality PCM (one quarter the sampling rate of audio CDs).
- **22,050 Hz** – Radio.
- **32,000 Hz** - miniDV digital video camcorder, DAT (LP mode).
- **44,100 Hz** - Audio CD, also most commonly used with MPEG-1 audio (VCD, SVCD, MP3) (Originally chosen by Sony, 1979).
- **48,000 Hz** - Digital sound used for miniDV, digital TV, DVD, DAT, films and professional audio.
- **96,000 or 192,000 Hz** - DVD-Audio, some LPCM DVD tracks, BD-ROM (Blu-ray Disc) audio tracks, and HD-DVD (High-Definition DVD) audio tracks.
- **2.8224 MHz** - Super Audio CD (SACD), 1-bit sigma-delta modulation process known as Direct Stream Digital (DSD), co-developed by Sony and Philips.
- **5.6448 MHz** - Double-Rate DSD, 1-bit Direct Stream Digital at 2x the rate of the SACD. Used in some professional DSD recorders (128 * 44100 Hz).
- **DXD** - 24-bit sampled at 352.8 kHz, suited for editing, eq. with 8.4672 MHz 1-bit DSD
Pulse-code Modulation (PCM)

Direct Stream Digital

- The trademark name used by Sony and Philips.
- Uses pulse-density modulation encoding
Audio, Quantization

Uniform Quantizer, Midtread

- Simple and popular
- Midtread
  - Odd number of reconstruction levels (N) (quantizing levels)
  - Here N = 9
Audio, Quantization

Uniform Quantizer, Midrise

- **Simple and popular**
- **Midrise**
  - Even number of reconstruction levels
- **Here N = 8**

Quantization error for bounded input
Audio, Quantization

Quantization Levels, SNR in dB

- Want to prevent human ear fatigue by minimizing quantization noise

\[
\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}}
\]

\[
\text{SNR}_{\text{dB}} = 10\log_{10} \left( \frac{P_{\text{signal}}}{P_{\text{noise}}} \right)
\]

- Signal-to-Noise Ratio (SNR) = 6.02*B dB
- SNR is approximately 6 dB per bit.
  - 16-bit => 96 dB
  - Above 36 dB is required

Horizontal axis: Power ratio in linear scale
Vertical axis: Power ratio in dB
Audio, Quantization

dB in Sound Pressure Level (SPL)

- dB in SNR not to be confused with dB in SPL
- SPL: A logarithmic measure of the effective sound pressure of a sound relative to a reference value.
- The commonly used reference sound pressure: 20 micropascals (µPa) (roughly the sound of a mosquito flying 3 m away)

<table>
<thead>
<tr>
<th>Sound</th>
<th>SPL (Pa)</th>
<th>SPL (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jet engine at 30 m</td>
<td>632</td>
<td>150</td>
</tr>
<tr>
<td>Passenger car at 10 m</td>
<td>$2 \times 10^{-2} - 2 \times 10^{-1}$ Pa</td>
<td>60 – 80</td>
</tr>
<tr>
<td>Normal conversation at 1 m</td>
<td>$2 \times 10^{-3} - 2 \times 10^{-2}$ Pa</td>
<td>40 – 60</td>
</tr>
<tr>
<td>Calm breathing</td>
<td>$2 \times 10^{-4} - 6.32 \times 10^{-4}$ Pa</td>
<td>10</td>
</tr>
</tbody>
</table>
Audio, Quantization

**dBA in Sound Pressure Level (SPL)**

The human ear responds more to frequencies between 500 Hz and 8 kHz and is less sensitive to very low-pitch or high-pitch noises. The frequency weightings used in sound level meters are often related to the response of the human ear, to ensure that the meter is **measuring** pretty much **what you actually hear**.

A-Weighted frequency response

**dBZ**: means no weighting at all
Audio, Quantization

VSLM
Virtual Sound Level Meter (VSLM)

The MATLAB development of a virtual sound level meter for analyzing calibrated sound files.
Ref.: http://sourceforge.net/projects/vslm/
Audio, Quantization

Quantization Levels, SNR in dB

Sample output of SpeechNoise_T03.m (Play noisy speech with different SNR values)
Audio, Quantization

6 dB per bit rule of thump

**Average power of a process or signal:**

\[ \int_{-\infty}^{+\infty} (x - \mu_x)^2 p(x) \, dx = \sigma_x^2 \]

- **\( \mu_x \):** Mean
- **\( p(x) \):** Probability density function

\[ \text{Assumption: } e[n] \text{ is uniform over } (-\frac{\Delta}{2}, \frac{\Delta}{2}) \]

\[ \Delta = \frac{2X_m}{2^B} \]

\[ e[n] = \hat{x}[n] - x[n] \]

\[ -X_m < x[n] < X_m \]

\[ -\frac{\Delta}{2} < e[n] \leq \frac{\Delta}{2} \]

\[ \text{The probability density function of } e[n] \]

\[ SNR_{dB}(B) = 10 \log_{10} \left( \frac{\sigma_x^2}{\sigma_e^2} \right) \]

\[ = \left( 20 \log_{10} 2 \right) B + 10 \log_{10} \left( \frac{3\sigma_x^2}{X_m^2} \right) \]

\[ = 6.02 B + 10 \log_{10} \left( \frac{3\sigma_x^2}{X_m^2} \right) \]
Audio, Quantization

6 dB per bit rule of thump

\[
SNR_{dB} (B) = 6.02B + 10 \log_{10} \left( \frac{3\sigma_x^2}{X_m^2} \right)
\]

Example 1: SNR for Uniform Quantization of Uniformly-Distributed Input

\[
\sigma_x^2 = \frac{X_m^2}{3}
\]

\[
SNR_{dB} (B) = 6.02B
\]

Example 2: SNR for Uniform Quantization of Sinusoidal Input

\[
\sigma_x^2 = \frac{X_m^2}{2}
\]

\[
SNR_{dB} (B) = 6.02B + 1.76
\]

Example 3: SNR for Uniform Quantization of Gaussian Input

\[
\sigma_x^2 = \frac{X_m^2}{16}
\]

\[
SNR_{dB} (B) = 6.02B - 7.27
\]
Audio, Bit-rate

Good to Know

- The average person cannot tell the difference between a bit rate above 192 kbit/s and the original CD/WAV.
- Even if your headphones seal really well around your ears, they will probably only give you about 20 to 25 dB insulation from the external sound.

Noise level for 192 kbps audio is under -125 dB and certainly inaudible.

Meaning of this dB: Noise power after coding and decoding over original signal power in logarithmic scale.
Audio, Quantization

Sound Signal Histogram

What is a Histogram?
Histogram: To roughly assess the probability distribution of a given variable by depicting the frequencies of observations occurring in certain ranges of values.

Wonder how the histogram of a typical sound signal will look like? Is it uniform?

Histogram of the set \{1, 2, 2, 3, 3, 3, 4, 4, 5, 6\} with a bin width of 10.

<table>
<thead>
<tr>
<th>Data Range (bin)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>1</td>
</tr>
<tr>
<td>10-20</td>
<td>3</td>
</tr>
<tr>
<td>20-30</td>
<td>6</td>
</tr>
<tr>
<td>30-40</td>
<td>4</td>
</tr>
<tr>
<td>40-50</td>
<td>2</td>
</tr>
</tbody>
</table>
Audio, Quantization

Typical Speech Signal Waveform

See SpeechHistT01.m
Audio, Quantization

Typical Speech Signal Histogram

- 8-bins
- 16-bins
- 64-bins
- 256-bins

```
figure;
hist(x, 256);
axis([-1 1 -inf inf])
```
Audio, Quantization

Uniform Quantizer, Midtread

Original sound

Quantized signal

3-bit Quantized

Original (17700:18200)

\[
aQ\_Partition = [-0.8750\times10^{-1}, -0.6250\times10^{-1}, -0.3750\times10^{-1}, -0.1250\times10^{-1}, 0.1250\times10^{-1}, 0.3750\times10^{-1}, 0.6250\times10^{-1}] ;
\]

\[
aQ\_Codebook = [-1.000\times10^{-1}, -0.750\times10^{-1}, -0.500\times10^{-1}, -0.250\times10^{-1}, 0.000\times10^{0}, 0.250\times10^{-1}, 0.500\times10^{-1}, 0.750\times10^{-1}] ;
\]

See SpeechQuantizationT02.m
Audio, Quantization

Uniform Quantizer, Midrise

See SpeechQuantizationT01.m

\[
aQ_{\text{Partition}} = [-0.7500 \quad -0.5000 \quad -0.2500 \quad 0 \quad 0.2500 \quad 0.5000 \quad 0.7500];
\]

\[
aQ_{\text{Codebook}} = [-0.8750 \quad -0.6250 \quad -0.3750 \quad -0.1250 \quad 0.1250 \quad 0.3750 \quad 0.6250 \quad 0.8750];
\]
Audio, Quantization

Nonuniform Quantization

Typical speech signal and its histogram

(a) Uniform and (b) non-uniform quantization $Q(x)$ and quantization error $q(x)$
Audio, Quantization

Uniform Quantizer, Midtread

Original sound

Quantized signal

3-bit Quantized

Original (17700:18200)

See SpeechQuantizationT02.m

aQ_Partition = [-0.8750 -0.6250 -0.3750 -0.1250 0.1250 0.3750 0.6250];
aQ_Codebook = [-1.0000 -0.7500 -0.5000 -0.2500 0 0.2500 0.5000 0.7500];
Audio, Quantization

Non-Uniform Quantizer

Play and compare quantized speech in MATLAB

See SpeechQuantizationT03.m

\[
aQ\_Partition = [-0.5 -0.25 -0.1 -0.05 0.05 0.1 0.25];
\]
\[
aQ\_Codebook = [-0.6 -0.32 -0.17 -0.075 0 0.075 0.17 0.32];
\]
### Audio, Quantization

<table>
<thead>
<tr>
<th>Uniform 3-bit</th>
<th>SNR in dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>MidRise</td>
<td>2.9701</td>
</tr>
<tr>
<td>MidTread</td>
<td>14.8078</td>
</tr>
<tr>
<td>MidRise</td>
<td>15.8282</td>
</tr>
<tr>
<td>MidTread</td>
<td>25.0756</td>
</tr>
<tr>
<td>MidRise</td>
<td>28.4824</td>
</tr>
<tr>
<td>MidTread</td>
<td>36.167</td>
</tr>
</tbody>
</table>

Sample output of SpeechQuantizationT01.m thr. SpeechQuantizationT03.m
(Uniform and Nonuniform Speech Quantization)
Audio, Quantization

u-law, a-law

- Nonuniform quantizers: Difficult to make, Expensive.
- Solution: Companding → Uniform Q. → Expanding
Audio, Quantization

u-law, a-law

\[ x[n] = \text{speech /song/} \]

\[ y[n] = C(x[n]) \]
Companded Signal

Segment of \( x[n] \)

Segment of \( y[n] \)
Companded Signal
Audio, Quantization

u-law, a-law

u-law
North America and Japan

\[ F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \quad -1 \leq x \leq 1 \]

a-law
Europe

\[ F(x) = \text{sgn}(x) \begin{cases} \frac{Ax}{1+\ln(A|x|)} & \frac{1}{A} \leq |x| \leq 1 \\ 1+\ln(A|x|) & |x| < \frac{1}{A} \end{cases} \]
MATLAB code or GUI implementation (Take a look at Speech noise test MATLAB codes to have sample input signal and to find out more about how to plot and play the sounds.

Make a plot and show how SNR changes with different values for Mu and B.
Human Auditory System

How ear works: Ear structure

How Does Hearing Work?

Excessive noise exposure is one of the leading causes of hearing loss. The tiny hair cells in the inner ear are easily damaged by loud noise and once you lose them, they never grow back!
Human Auditory System

How ear works: Filter banks

Play “How the ear works” animation.

Filter bank frequencies on the cochlea.

Inside the cochlea.
Audio, Sampling

Aliasing due to low sampling

Properly sampled image of brick wall.

Spatial aliasing in the form of a Moiré pattern.

Two different sinusoids that fit the same set of samples.
Thank You

Next Session: Speech II

1. http://ce.sharif.edu/~m_amiri/