Introduction to Audio Content Analysis

Module A.1: Fundamentals — Digitization

alexander lerch

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Section A.1

- **lecture content**
  - discretization of signals in time and amplitude
  - ambiguity and aliasing
  - sampling theorem
  - properties of the quantization error

- **learning objectives**
  - summarize the principle of discretization
  - describe the implications of the sample theorem

corresponding textbook section

Section A.1

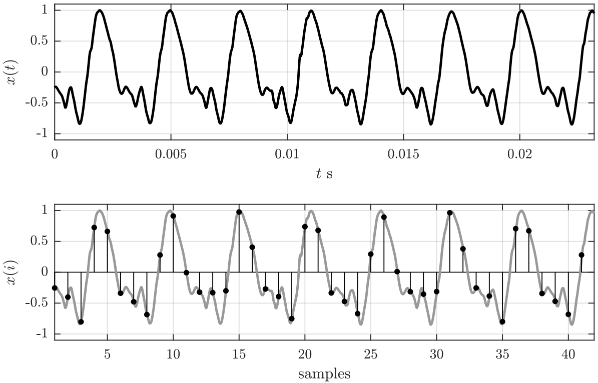
- **lecture content**
  - discretization of signals in time and amplitude
  - ambiguity and aliasing
  - sampling theorem
  - properties of the quantization error
- **learning objectives**
  - summarize the principle of discretization
  - describe the implications of the sample theorem

digital signals are represented with a limited number of values

$\Rightarrow$

1. **sampling**: time discretization
   continuous time $\mapsto$ discrete equidistant points in time
2. **quantization**: amplitude discretization
   continuous amplitude $\mapsto$ discrete, pre-defined, set of values

sampling
basic concept



- f_S [Hz]: number of samples per second
- $T_S = {}^1/_{f_S}$ [s]: distance between two neighboring samples

matlab source: plotSampling01.m

sampling
sampling frequencies

What are typical sample rates

sampling
sampling frequencies

What are typical sample rates

- 8–16 kHz: speech (phone)
- 44.1–48 kHz: (consumer) audio/music
- $>$48 kHz: production audio

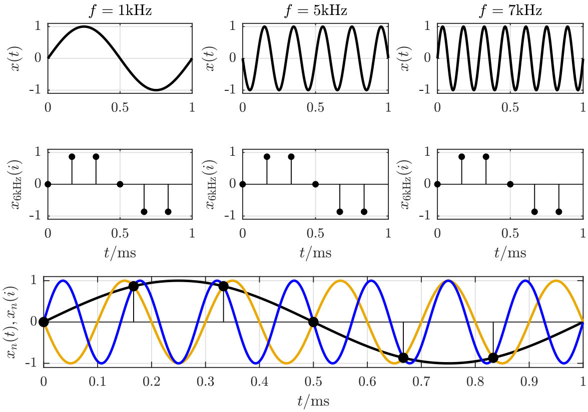
sampling
sampling frequencies

What are typical sample rates

- 8–16 kHz: speech (phone)
- 44.1–48 kHz: (consumer) audio/music
- $>$48 kHz: production audio

f_S	44.1 kHz	32 kHz	22.05 kHz	16 kHz	8 kHz	6 kHz

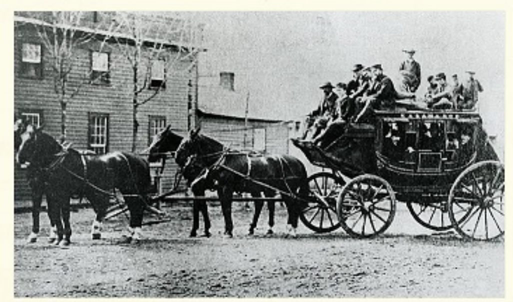
sampling
sampling ambiguity



matlab source: plotSampling02.m

sampling

sampling ambiguity — wagon-wheel effect



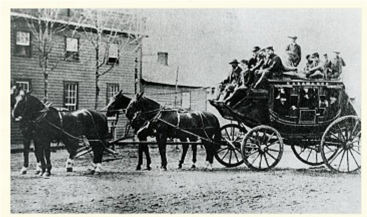
flickr.com/photos/fotoguy49057/12209056184

sampling

sampling ambiguity — wagon-wheel effect

compare speed of wheel (spokes) f_wheel between real world and video recording for an accelerating stage coach

1. $f_\mathrm{wheel} < \frac{f_\mathrm{S}}{2}$
   speeding up
2. $\frac{f_\mathrm{S}}{2} < f_\mathrm{wheel} < f_\mathrm{S}$
   slowing down
3. $f_\mathrm{wheel} = f_\mathrm{S}$:
   standing still
4. ...

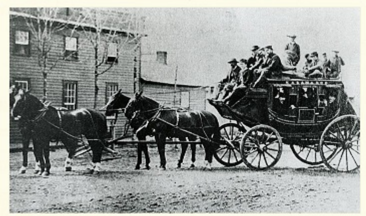


sampling

sampling ambiguity — wagon-wheel effect

compare speed of wheel (spokes) f_wheel between real world and video recording for an accelerating stage coach

1. $f_\mathrm{wheel} < \frac{f_\mathrm{S}}{2}$
   speeding up
2. $\frac{f_\mathrm{S}}{2} < f_\mathrm{wheel} < f_\mathrm{S}$
   slowing down
3. $f_\mathrm{wheel} = f_\mathrm{S}$:
   standing still
4. ...



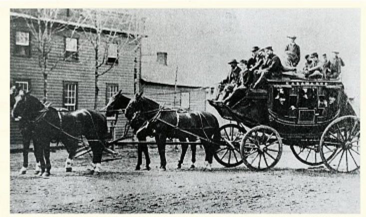
sampling

sampling ambiguity — wagon-wheel effect

Georgia Tech Center for Music Technology College of Design

compare speed of wheel (spokes) f_{wheel} between real world and video recording for an accelerating stage coach

1. $f_{\mathrm{wheel}} < \frac{f_{\mathrm{S}}}{2}$
   speeding up
2. $\frac{f_{\mathrm{S}}}{2} < f_{\mathrm{wheel}} < f_{\mathrm{S}}$
   slowing down
3. $f_{\mathrm{wheel}} = f_{\mathrm{S}}$:
   standing still
4. ...

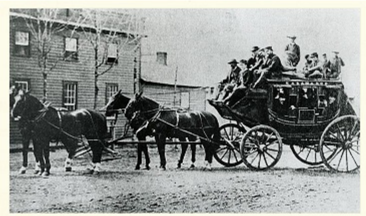


sampling

sampling ambiguity — wagon-wheel effect

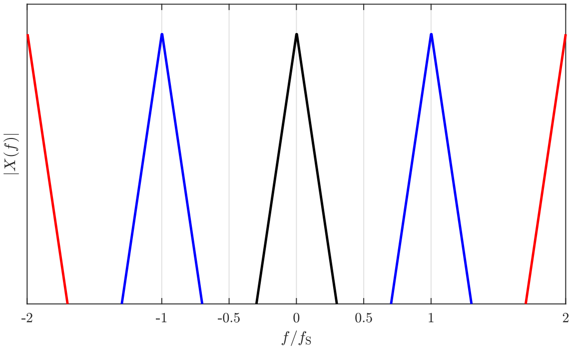
compare speed of wheel (spokes) f_wheel between real world and video recording for an accelerating stage coach

1. $f_\mathrm{wheel} < \frac{f_\mathrm{S}}{2}$
   speeding up
2. $\frac{f_\mathrm{S}}{2} < f_\mathrm{wheel} < f_\mathrm{S}$
   slowing down
3. $f_\mathrm{wheel} = f_\mathrm{S}$:
   standing still
4. ...



video example: youtu.be/QYYK4tICMIY

digital signals
sampling ambiguity — spectral domain



matlab source: matlab/animateSampling.m

sampling theorem

A sampled signal can be reconstructed without loss of information if the sample rate f_S is higher than twice the bandwidth f_max of the original audio signal.

$$f_\mathrm{S} > 2 \cdot f_\mathrm{max}$$

$f_\mathrm{S}/2$ is also referred to as the *Nyquist*[1]-rate

[1] Harry Nyquist, 1889–1976

digital signals

quantization

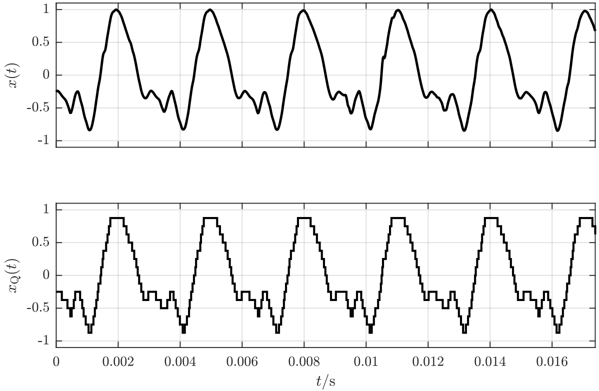
- continuous amplitude values are mapped to pre-defined, equidistant set of values
- signal stored in binary ⇒ # quantization steps equals **power of 2**
- **example: 4-bit quantization**
    - *word length*: $w = \log_2(\mathcal{M}) = 4\,\mathrm{bit}$
    - *number of quantization steps*: $\mathcal{M} = 2^w = 16$

matlab source: plotQuantization.m

digital signals

quantization

- continuous amplitude values are mapped to pre-defined, equidistant set of values
- signal stored in binary ⇒ # quantization steps equals **power of 2**
- **example: 4-bit quantization**
  - *word length*: $w = \log_2(\mathcal{M}) = 4\,\mathrm{bit}$
  - *number of quantization steps*: $\mathcal{M} = 2^w = 16$



matlab source: plotQuantization.m

What are typical wordlengths?

What are typical wordlengths?

- 8 bit: speech
- 12–14 bit: low quality audio/music
- 16 bit: (consumer) audio/music
- >16 bit: production audio

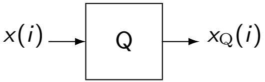
What are typical wordlengths?

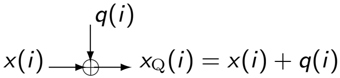
- 8 bit: speech
- 12–14 bit: low quality audio/music
- 16 bit: (consumer) audio/music
- $>$16 bit: production audio

w	16 bit	12 bit	8 bit	4 bit	2 bit

digital signals

quantization error



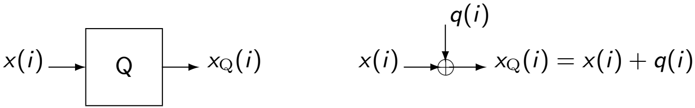


model for quantization:
quantization noise q is added to input signal x

$$\begin{aligned} x_Q(i) &= x(i) + q(i) \\ q(i) &= x(i) - x_Q(i) \end{aligned}$$

digital signals

quantization error



model for quantization:
quantization noise q is added to input signal x

$$
\begin{aligned}
x_\mathrm{Q}(i) &= x(i) + q(i) \\
q(i) &= x(i) - x_\mathrm{Q}(i)
\end{aligned}
$$

digital signals
quantization error magnitude

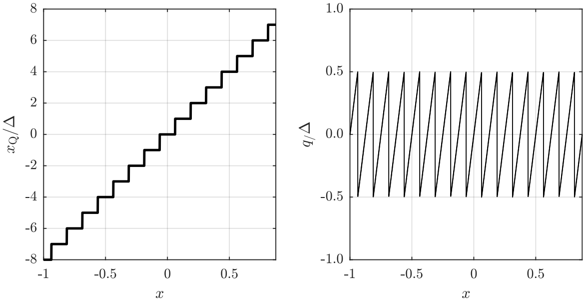
What is the maximum amplitude of the quantization error?

matlab source: plotQuantizationError.m

digital signals

quantization error magnitude

What is the maximum amplitude of the quantization error?



matlab source: plotQuantizationError.m

digital signals
quantization error properties

Under the assumption that the signal has a variance much higher than the quantization step size (no derivation), we find that the quantization error

- is white noise and uncorrelated to signal,
- is uniformly distributed, and
- its power W_Q is directly related to the wordlength.

The quantizer quality is usually given by its *Signal-to-Noise Ratio (SNR)*

$$SNR = 10 \cdot \log_{10}\left(\frac{W_\mathrm{S}}{W_\mathrm{Q}}\right) \; [dB]$$

digital signals

quantization error properties

Under the assumption that the signal has a variance much higher than the quantization step size (no derivation), we find that the quantization error

- is white noise and uncorrelated to signal,
- is uniformly distributed, and
- its power W_Q is directly related to the wordlength.

The quantizer quality is usually given by its *Signal-to-Noise Ratio (SNR)*

$$SNR = 10 \cdot \log_{10}\left(\frac{W_\mathrm{S}}{W_\mathrm{Q}}\right) \; [dB]$$

digital signals
quantization: SNR

Georgia Tech | Center for Music Technology
College of Design

signal-to-noise ratio (quantizer)

$$SNR = 6.02 \cdot w + c_\mathrm{S} \quad [dB]$$

- every additional bit adds app. 6 dB SNR
- constant c_S depends on *signal* (scaling and PDF)

- square wave (full scale): $c_\mathrm{S} = 10.80\,\mathrm{dB}$
- sinusoidal wave (full scale): $c_\mathrm{S} = 1.76\,\mathrm{dB}$
- rectangular PDF (full scale): $c_\mathrm{S} = 0\,\mathrm{dB}$
- Gaussian PDF (full scale $= 4\sigma_g$): $c_\mathrm{S} = -7.27\,\mathrm{dB}$

signal-to-noise ratio (quantizer)

$$SNR = 6.02 \cdot w + c_\mathrm{S} \quad [dB]$$

- every additional bit adds app. 6 dB SNR
- constant c_S depends on *signal* (scaling and PDF)

- square wave (full scale): $c_\mathrm{S} = 10.80\,\mathrm{dB}$
- sinusoidal wave (full scale): $c_\mathrm{S} = 1.76\,\mathrm{dB}$
- rectangular PDF (full scale): $c_\mathrm{S} = 0\,\mathrm{dB}$
- Gaussian PDF (full scale $= 4\sigma_g$): $c_\mathrm{S} = -7.27\,\mathrm{dB}$

- when represented as integer, different wordlengths lead to different maximum amplitude ranges
- most common: normalize to the absolute maximum integer value and represent the signal in **floating point format**

⇒ signal amplitude:

$$-1 \leq x_\mathrm{Q} < 1$$

⇒ level:

$$\text{max. amplitude} \mapsto 0 dBFS$$

- floating point representation

$$x_\mathrm{Q} = M_\mathrm{G} \cdot 2^{E_\mathrm{G}}$$

- internal float point representation usually treated as signal being **not quantized**

digital signals
amplitude in DSP

- when represented as integer, different wordlengths lead to different maximum amplitude ranges
- most common: normalize to the absolute maximum integer value and represent the signal in **floating point format**

⇒ signal amplitude:

$$-1 \leq x_\mathrm{Q} < 1$$

⇒ level:

$$\text{max. amplitude} \mapsto 0dBFS$$

- floating point representation

$$x_\mathrm{Q} = M_\mathrm{G} \cdot 2^{E_\mathrm{G}}$$

- internal float point representation usually treated as signal being **not quantized**

digital signals
amplitude in DSP

- when represented as integer, different wordlengths lead to different maximum amplitude ranges
- most common: normalize to the absolute maximum integer value and represent the signal in **floating point format**

$\Rightarrow$ signal amplitude:

$$-1 \leq x_\mathrm{Q} < 1$$

$\Rightarrow$ level:

$$\text{max. amplitude} \mapsto 0 dBFS$$

- floating point representation

$$x_\mathrm{Q} = M_\mathrm{G} \cdot 2^{E_\mathrm{G}}$$

- internal float point representation usually treated as signal being **not quantized**

digital signals
amplitude in DSP

- when represented as integer, different wordlengths lead to different maximum amplitude ranges
- most common: normalize to the absolute maximum integer value and represent the signal in **floating point format**

⇒ signal amplitude:

$$-1 \leq x_\mathrm{Q} < 1$$

⇒ level:

$$\text{max. amplitude} \mapsto 0dBFS$$

- floating point representation

$$x_\mathrm{Q} = M_\mathrm{G} \cdot 2^{E_\mathrm{G}}$$

- internal float point representation usually treated as signal being **not quantized**

digital signals
amplitude in DSP

- when represented as integer, different wordlengths lead to different maximum amplitude ranges
- most common: normalize to the absolute maximum integer value and represent the signal in **floating point format**

⇒ signal amplitude:

$$-1 \leq x_\mathrm{Q} < 1$$

⇒ level:

$$\text{max. amplitude} \mapsto 0 dBFS$$

- floating point representation

$$x_\mathrm{Q} = M_\mathrm{G} \cdot 2^{E_\mathrm{G}}$$

- internal float point representation usually treated as signal being **not quantized**

summary
lecture content

- continuous signal is sampled to be **discrete in time**
  - number of samples per second is called sampling rate or sampling frequency
- continuous signal is quantized to be **discrete in amplitude**
  - number of quantization steps equals $2^{\mathrm{wordlength}}$
- **sampling theorem**
  - sampled signal can be reconstructed without loss of information if the sample rate f_S is higher than twice the bandwidth f_max of the original audio signal
  - otherwise reconstruction is ambiguous and aliasing occurs
- **quantization error properties**
  - maximum amplitude is half the step size
  - number of steps depends on wordlength
- **SNR**
  - SNR depends on input signal characteristic and wordlength
  - SNR increases linearly (6 dB/bit) with wordlength