Introduction to Audio Content Analysis

Module 3.2: Fundamentals — Blocking

alexander lerch

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Section 3.2

- **lecture content**
  - splitting the audio signal into blocks
  - block length and hop size
- **learning objectives**
  - describe the reasons for blocking
  - summarize the principle using the correct terminology

corresponding textbook section

Section 3.2

- **lecture content**
  - splitting the audio signal into blocks
  - block length and hop size

- **learning objectives**
  - describe the reasons for blocking
  - summarize the principle using the correct terminology

block based processing
introduction

- typical audio applications **process blocks** of audio data
- instead of having a function called per sample, it is called with block of samples

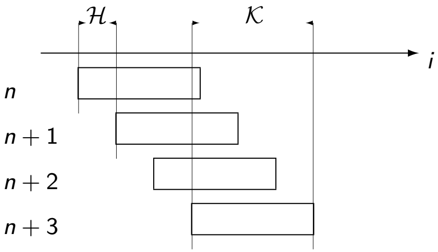
- **reasons**:
  - block based processing methods such as the Short-Time Fourier Transform
  - audio hardware characteristics (real-time systems)
  - efficiency (memory allocation, SIMD)
- typical block lengths:
  - 1...thousands of samples
  - often powers of 2 (1024, 2048, ...)

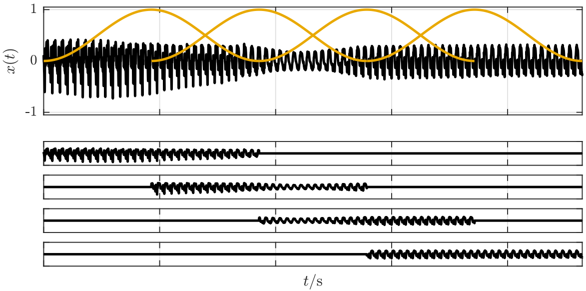
block based processing
introduction

- typical audio applications **process blocks** of audio data
- instead of having a function called per sample, it is called with block of samples

- **reasons**:
  - block based processing methods such as the Short-Time Fourier Transform
  - audio hardware characteristics (real-time systems)
  - efficiency (memory allocation, SIMD)
- typical block lengths:
  - 1...thousands of samples
  - often powers of 2 (1024, 2048, ...)

block based processing
description





- $\mathcal{K}$: block length
- $\mathcal{H}$: hop size
- n: block index
- i: sample index

matlab source: plotBlocking.m

block based processing
terms and definitions

- **block boundaries**:

$$
\begin{aligned}
i_\mathrm{s}(n) &= i_\mathrm{s}(n-1) + \mathcal{H} \\
i_\mathrm{e}(n) &= i_\mathrm{s}(n) + \mathcal{K} - 1
\end{aligned}
$$

- **overlap ratio**:

$$o_\mathrm{r} = \frac{\mathcal{K} - \mathcal{H}}{\mathcal{K}}$$

- **time stamp**:

$$t_\mathrm{s}(n) = \frac{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1}{2 \cdot f_\mathrm{S}} + \frac{i_\mathrm{s}(n)}{f_\mathrm{S}} = \frac{\mathcal{K}}{2 \cdot f_\mathrm{S}} + \frac{i_\mathrm{s}(n)}{f_\mathrm{S}}$$

- $\mathcal{K}$: block length
- $\mathcal{H}$: hop size
- n: block index
- i: sample index
- f_S: sample rate

summary
lecture content

- **audio signal is typically split into blocks**
- each block processed individually
- **terms**:
    - *block length*:
        - minimum: 1
        - typical: 256 . . . 16384
    - *hop size*:
        - minimum: 1
        - maximum: block length
        - typical: half of block length
    - block *time stamp*:
        - typically refers to either start or middle of block