Lecture

Foundations of Audio Signal Processing

MA-INF 2113

Winter Term 2025/26

§1 Introduction and Motivation

Prof. Dr. Frank Kurth

Fraunhofer FKIE

UNIVERSITÄT BONN

Audio Signal Processing...

UNIVERSITÄT BONN

...typically deals with the tasks of

- representing,
- analyzing / obtaining information from,
- processing / transforming, or
- generating / synthesizing

acoustic data on a computer.

In our research (@ U Bonn), we are mainly interested in the first three topics.

Audio Signals & Human perception

UNIVERSITÄT BONN

Roughly, we are interested in

„information a human can perceive with his auditory system“,

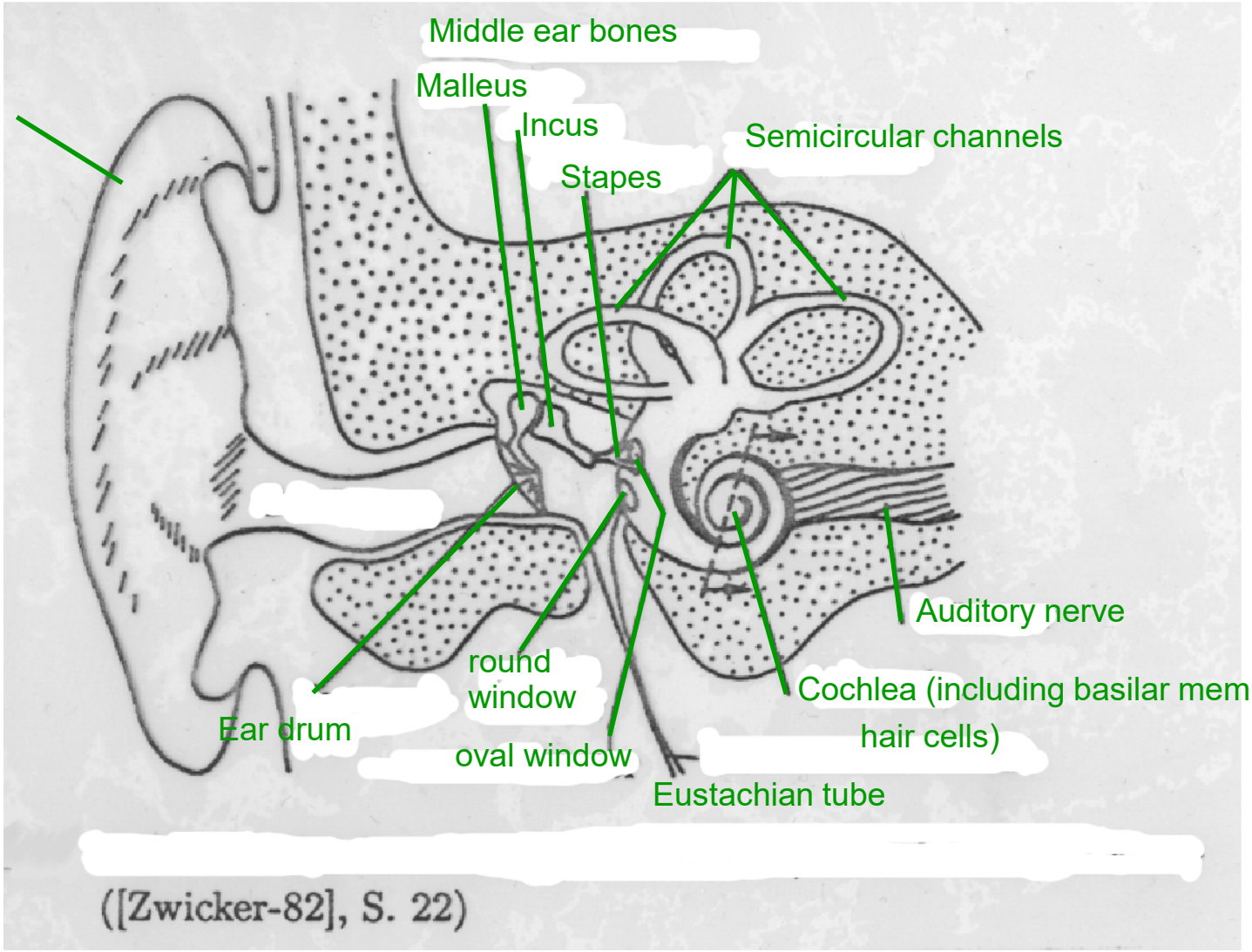
the *Human Auditory System* (for short: HAS).

We simply speak of *audio information* or *audio data*.

Do you know some examples?

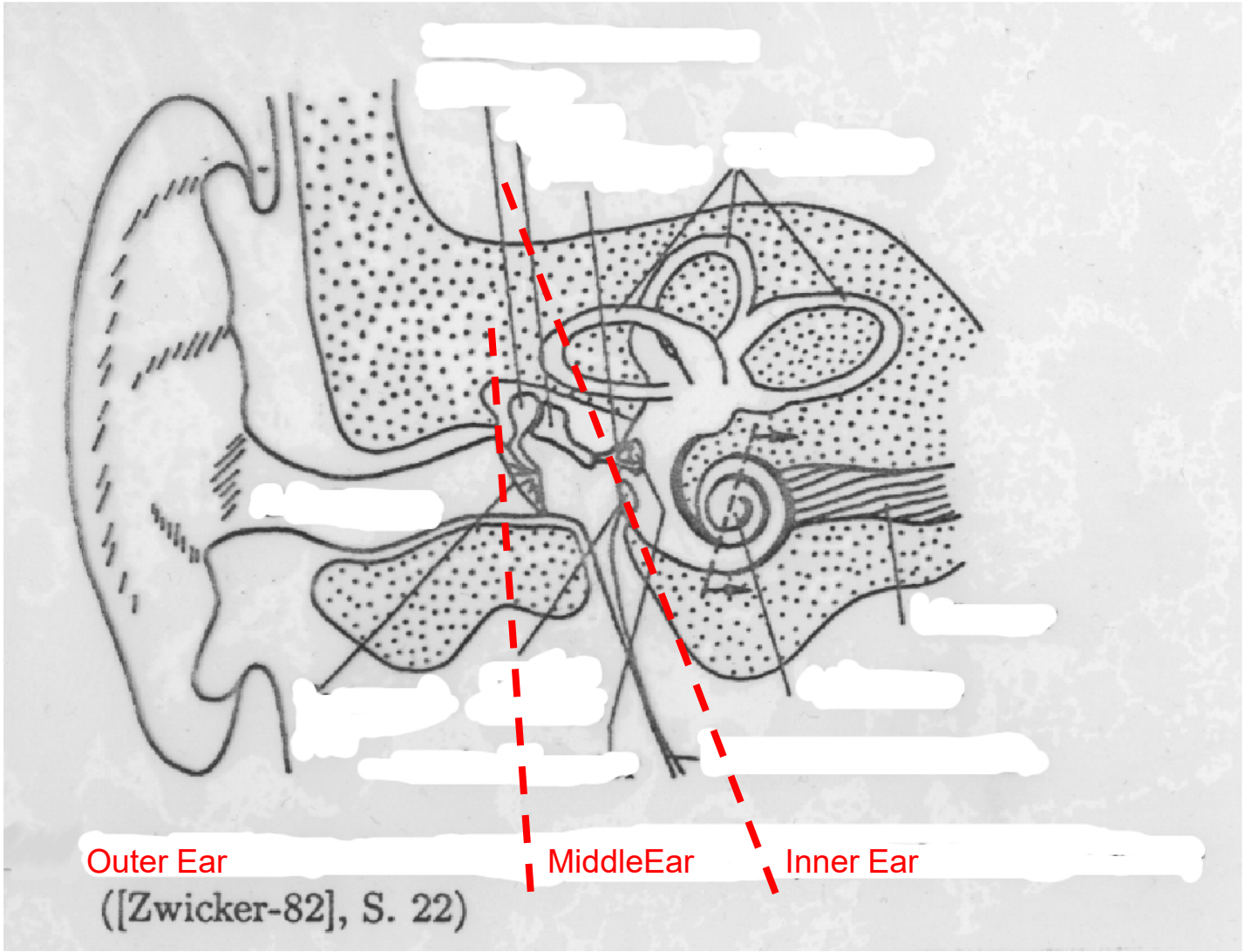
The Human Auditory System

UNIVERSITÄT BONN



The Human Auditory System

UNIVERSITÄT BONN



([Zwicker-82], S. 22)

Let‘s listen to some audio signals

UNIVERSITÄT BONN

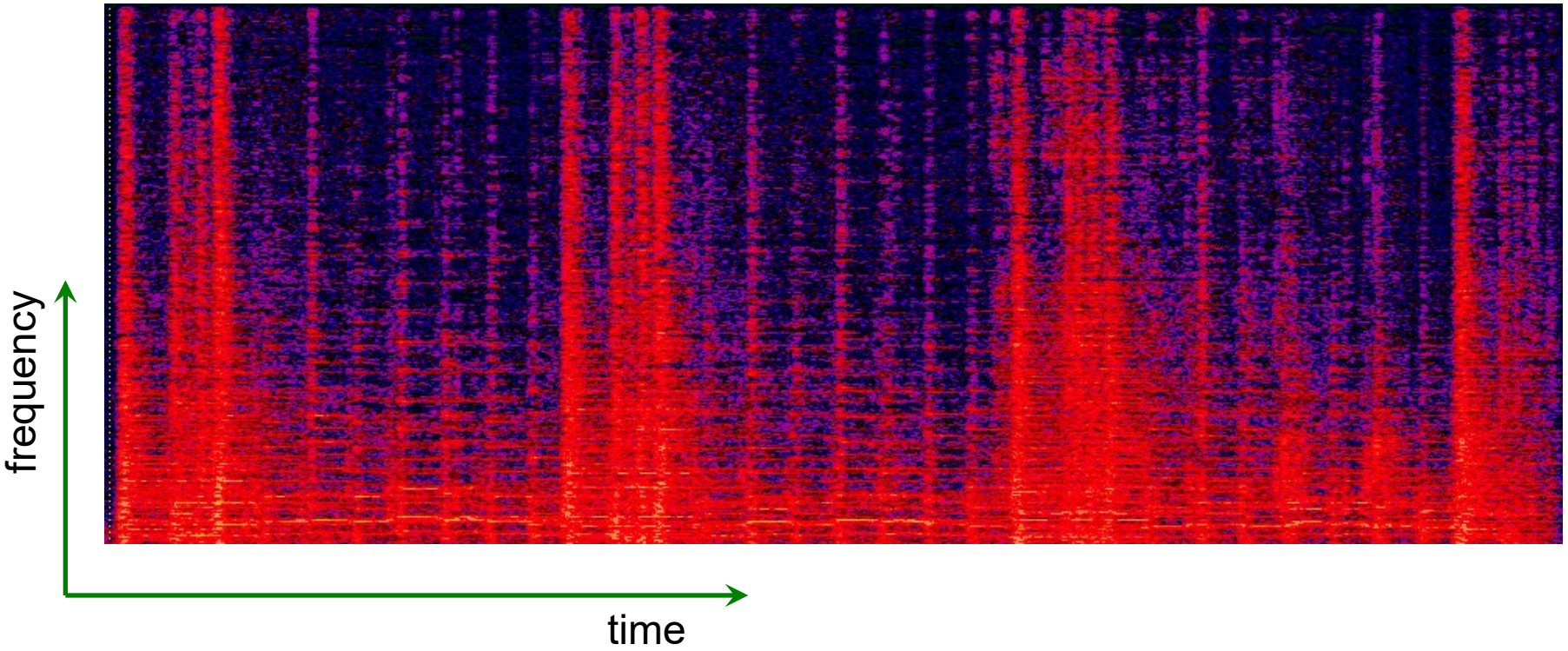
- Speech
- Music
- Animal sounds
- Environmental sounds
- Just noise (what is noise, anyway?)
- …



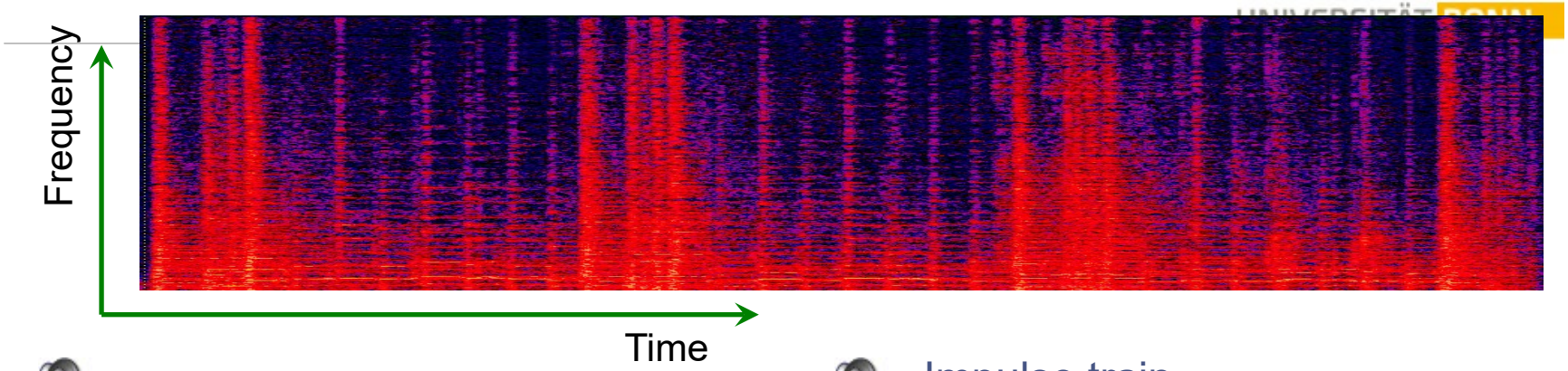


Quiz: Can you *see* the audio example?

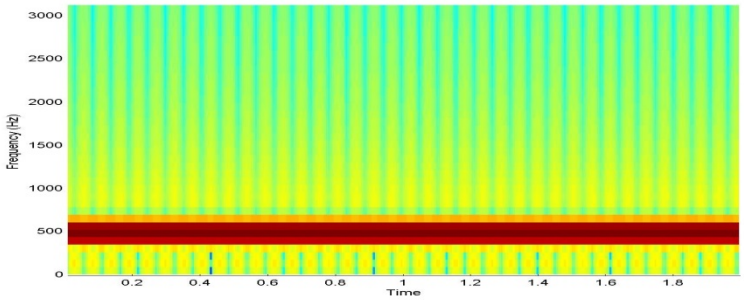
UNIVERSITÄT BONN

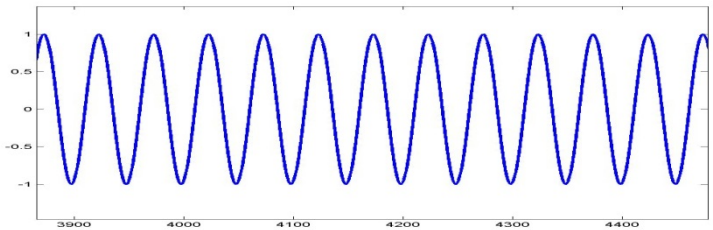


Spectrogram / WFT (§4)

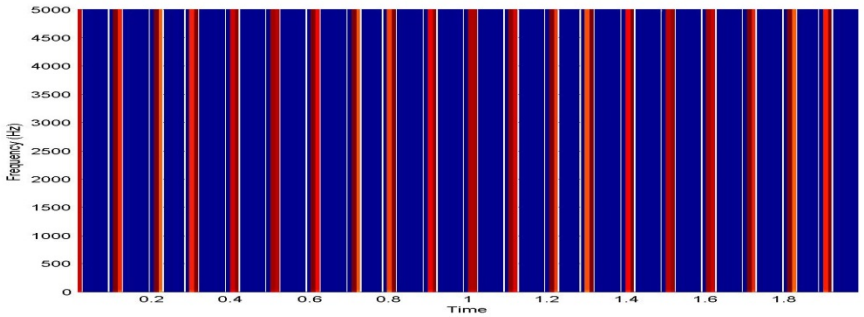


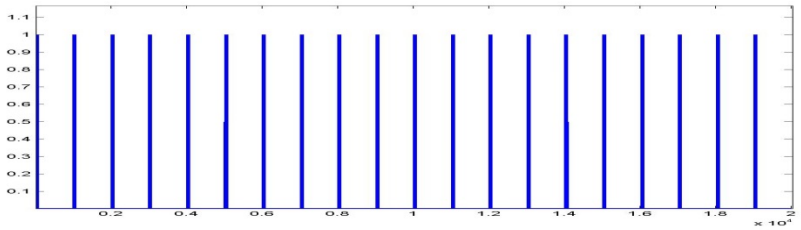
Sinusoid, e.g: 440 Hz





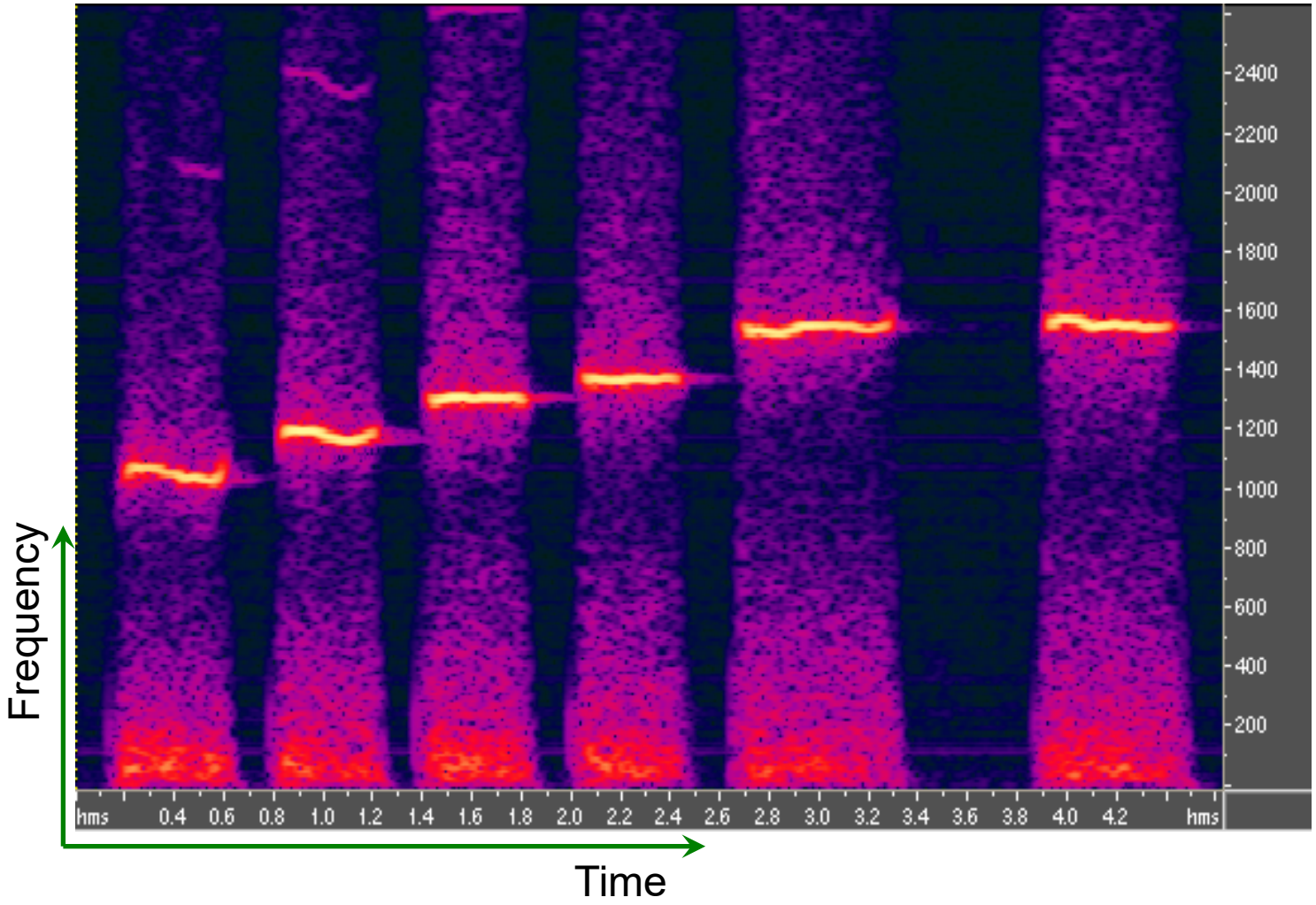
Impulse train





„A melody“ – Sequence of sinusoids

UNIVERSITÄT BONN



Representations of Audio

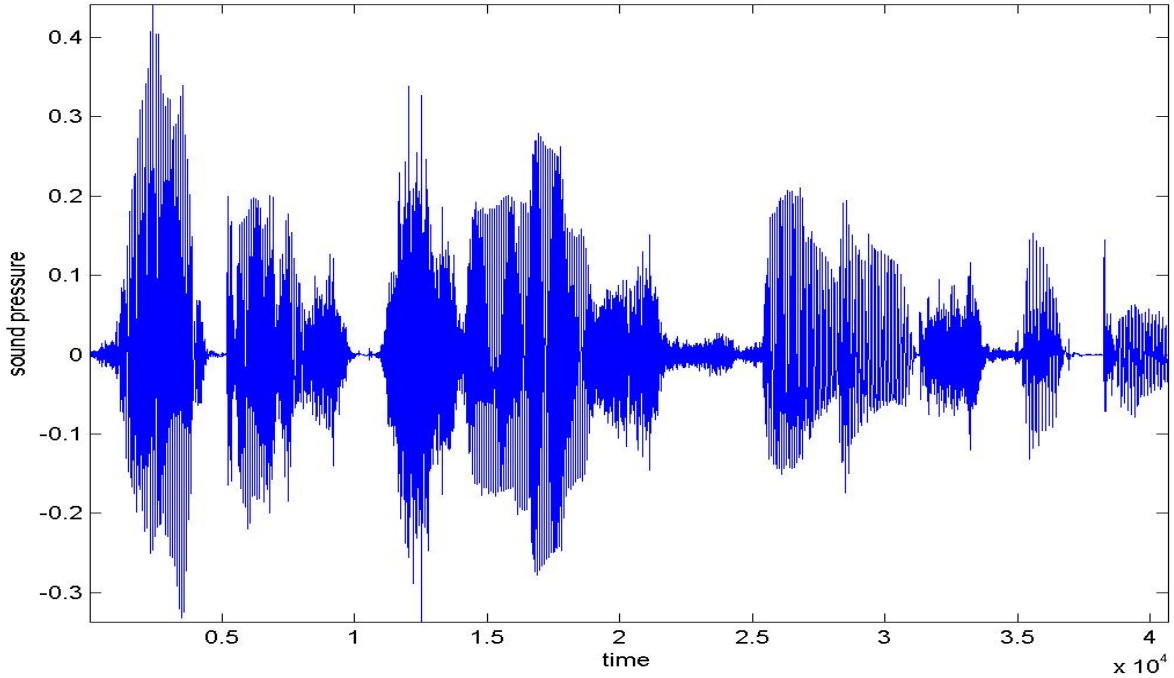
UNIVERSITÄT BONN

What kind of representations of audio data do you know?

- Representations of the physical waveform
  - Audio CD or DVD
  - Analog tape
  - Record (LP, phonograph/gramophone/vinyle record)
  - .wav – file on a computer
  - compressed format like .mpeg, .mp3, .aac

  Those are signal-based representations.
- Meta-formats which can be used to create audio:
  - Texts
  - Sheet music
  - MIDI / MusicXML

Audio Signals



Physical waveform of the audio – sound pressure level over time.

Mathematically, a signal is a function

$$f : R \rightarrow R$$

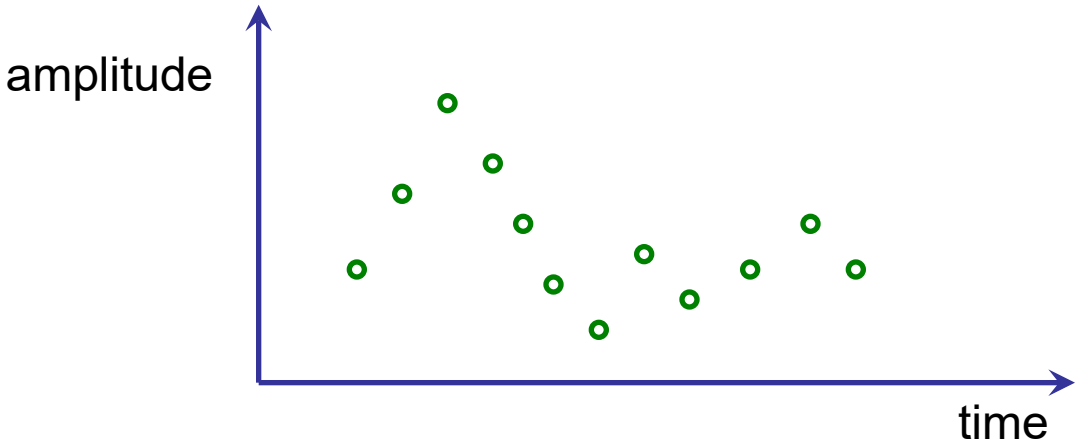
mapping time- to sound pressure-values. (R ~ real numbers)

On a Computer: Discrete Signals

UNIVERSITÄT BONN

To store audio signals on a computer, the domain and range have to be discretized. A signal now is an integer-valued function

$$x : Z \to Z$$



A signal ist mostly notated as a *sequence* of integer values, e.g.

..., -1, -2, -5, -3, 1, 2, 4, 7, 2, -1,...

could be contents of a .wav-File.

(Here: 1-channel- or monophonic (*mono-)* signal.)

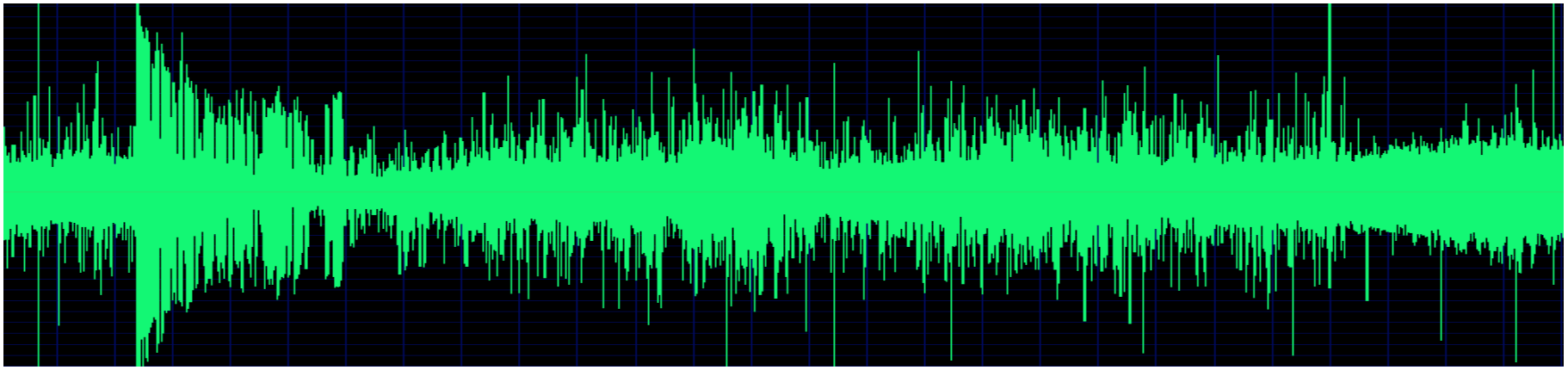
Fundamental area: Signal Processing

UNIVERSITÄT BONN

- How to...
  - formally represent audio (as a signal)
  - process/manipulate/transform signals
- Most work in Audio – at least as long as processing of the „physical“ representations is involved – is based on *Signal Processing*.
- On a computer, *Digital Signal Processing*.
- A major part of this lecture will hence be devoted to an introduction to digital signal processing.
- But let‘s first consider some „real-world“ tasks in audio processing!

Motivating examples: Audio processing tasks
(1) Audio detection

UNIVERSITÄT BONN



Given: An audio signal and a target signal type

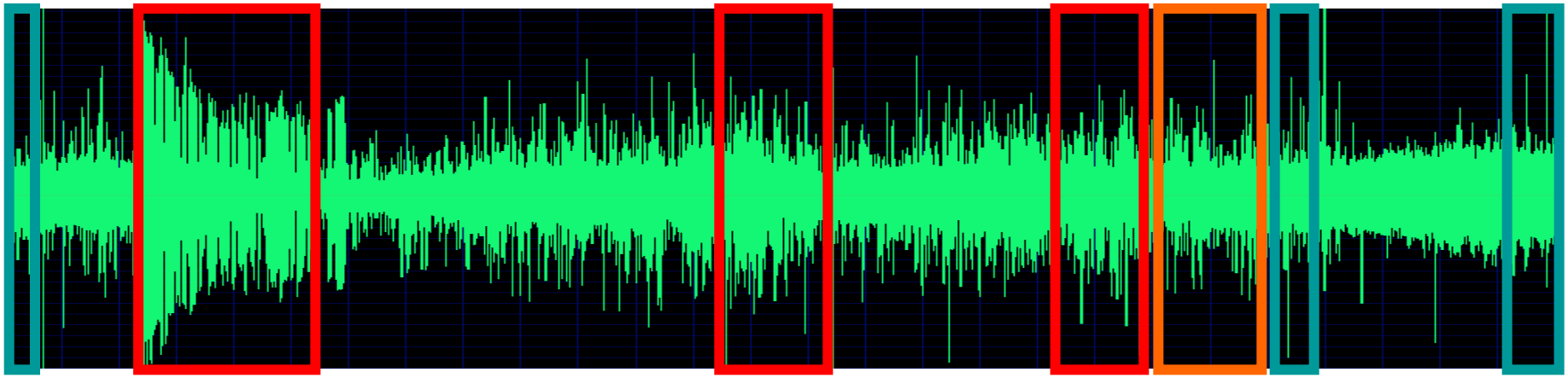
Task: Return a list of (temporal) signal segments containing a signal of the targeted type

Side conditions:

- Return all such segments (no „false negatives“).
- Do not return wrong segments (no „false positives“).

Motivating examples: Audio processing tasks
(1) Audio detection

UNIVERSITÄT BONN

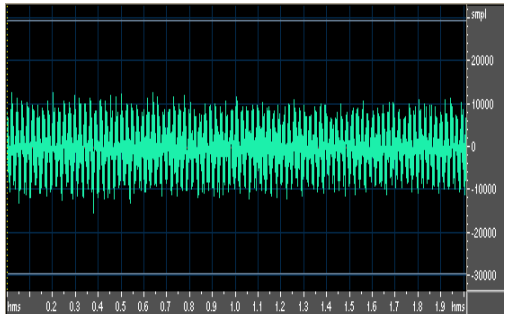


„Speech“ Segments

Also „Speech“ Segments, but weak quality

For comparison: background noise

Motivating examples: Audio processing tasks
(2) Audio classification



Classes:

{noise, speech, music, Vuvuzela}

„Vuvuzela“

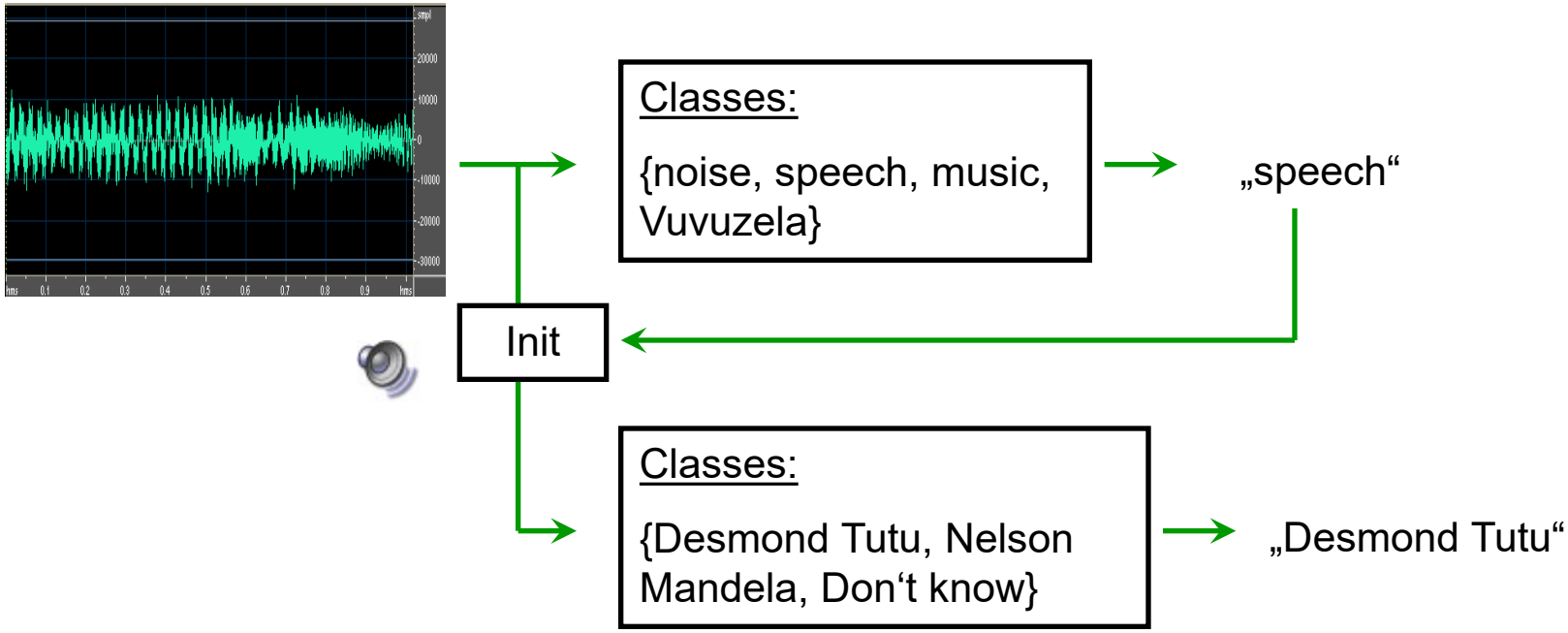
Given: Audio signal x; N different signal types (called „classes“)

(Class examples: „Peters voice“, „Carols voice“, „background noise“, „don‘t know“)

Task: Assign to x the correct class.

Side conditions: Typically, x will be of short duration.

Motivating examples: Audio processing tasks
(2) Audio classification – 2-stage classifier



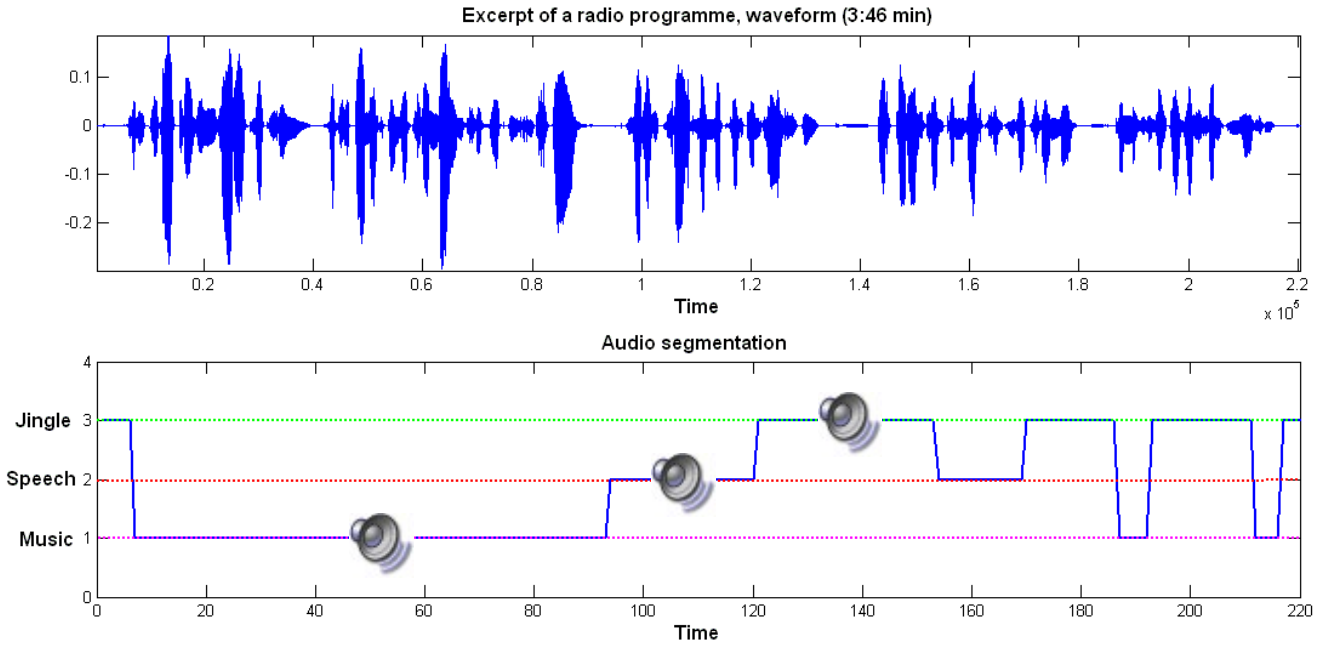
Given: Audio signal x; N different signal types (called „classes“)

(Class examples: „Peters voice“, „Carols voice“, „background noise“, „don‘t know“)

Task: Assign to x the correct class.

Side conditions: Typically, x will be of short duration.

Motivating examples: Audio processing tasks
(3) Audio segmentation



Given: A (typically long) audio signal $x = (x_1,..,x_n)$.

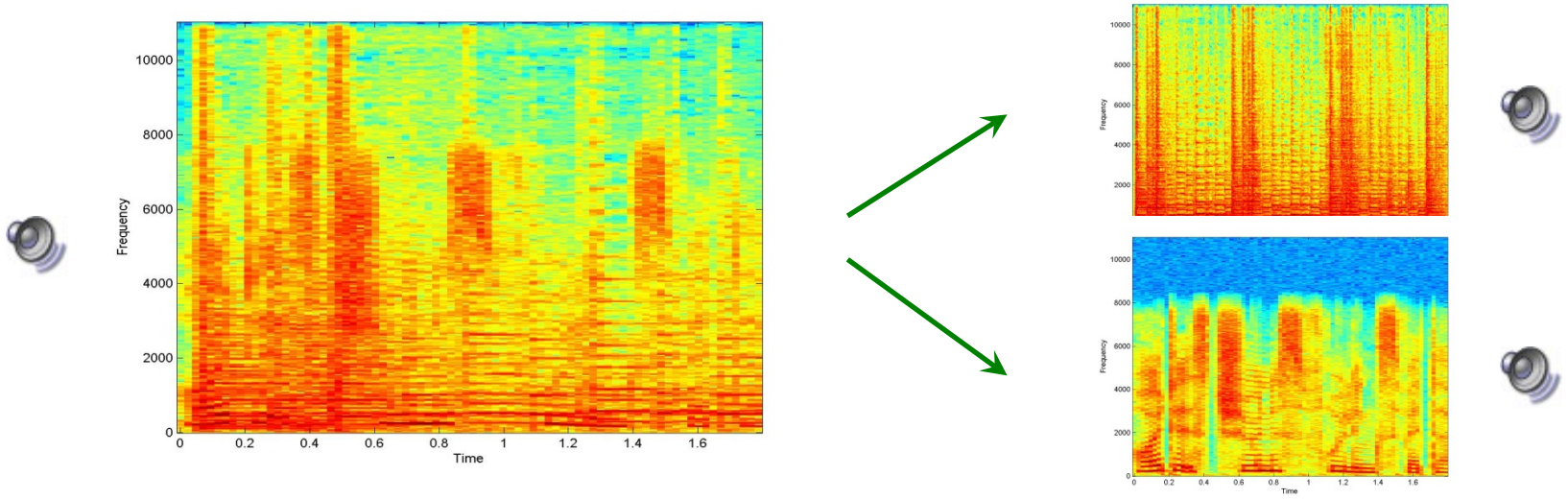
Task: Give a (temporal) partition of x in k segments $s_0,...,s_{k-1}$ where

$$s_i = (p_i, p_{i+1}-1) \text{ for integers } 1 = p_1 < p_2 < ... < p_k = n$$

Side conditions: For example,

- each segment should represent one of N given classes and
- pairs of subsequent segments should belong to two different classes.

Motivating examples: Audio processing tasks
(4) Source separation



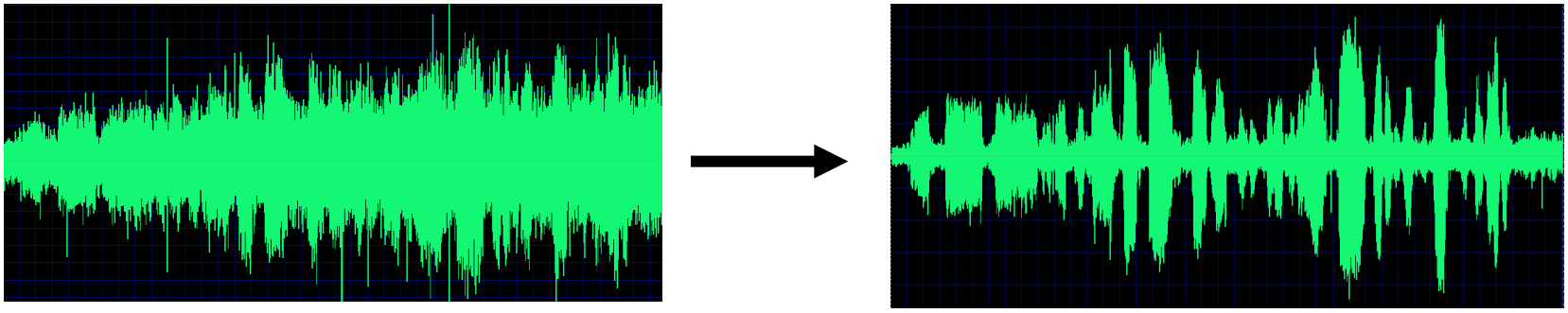
Given: N simultaneous audio recordings of a mix of M audio sources

Task: Return all M sources in M separate signals, each containing only one of the M sources.

Remarks:

- Typically, N < M.
- Example N = 1 is very hard
- Is the problem well-posed? Assume 15 = a+b - what are a, b?
- ... as sources are generally temporally correlated, there is some hope that the separation problem can be solved for particular cases.

Motivating examples: Audio processing tasks
(5) Noise reduction



Given: A target signal corrupted by noise.

Task:

- Reduce the amount of noise or
- Extract the target signal or
- Increase the intelligibility of the target signal.

Side conditions: Do not reduce the „quality“ of the target signal.

Motivating examples: Audio processing tasks
(6) Audio compression

UNIVERSITÄT BONN

Given: A target signal x, represented using N bytes.

Task: Convert the signal to a compressed form c(x) requiring M << N bytes.

Side conditions: It should be possible to ...

- (perfectly) reconstruct x from c(x) [lossless compression]

x = (1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 5, 5, 5, 5, 3, 3, 3, 2, 2, 1, 1, 1, 1)

c(x) = ((1,6), (2,3), (3,2), (5,4), (3,3), (2,2),(1,4)

removes redundance,

or

- perceptually reconstruct x from c(x) [lossy compression]

x given as a .wav – file „16 bit, 44100 Hz“ (e.g. 544 kB)

c(x) obtained as a .mp3 – file „@128 kbps“ (e.g. 102 kB)

removes irrelevance.

Motivating examples: Audio processing tasks
(7) Audio identification

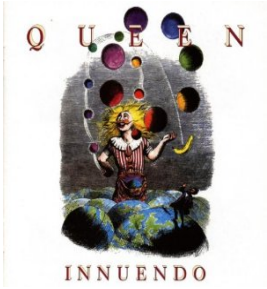
Title: Innuendo

Artist: Queen

Album: Innuendo

Issued: Feb. 4th, 1991

Label: Parlophone (EMI)



Given: A short music signal, e.g. recorded in a car using a mobile phone.

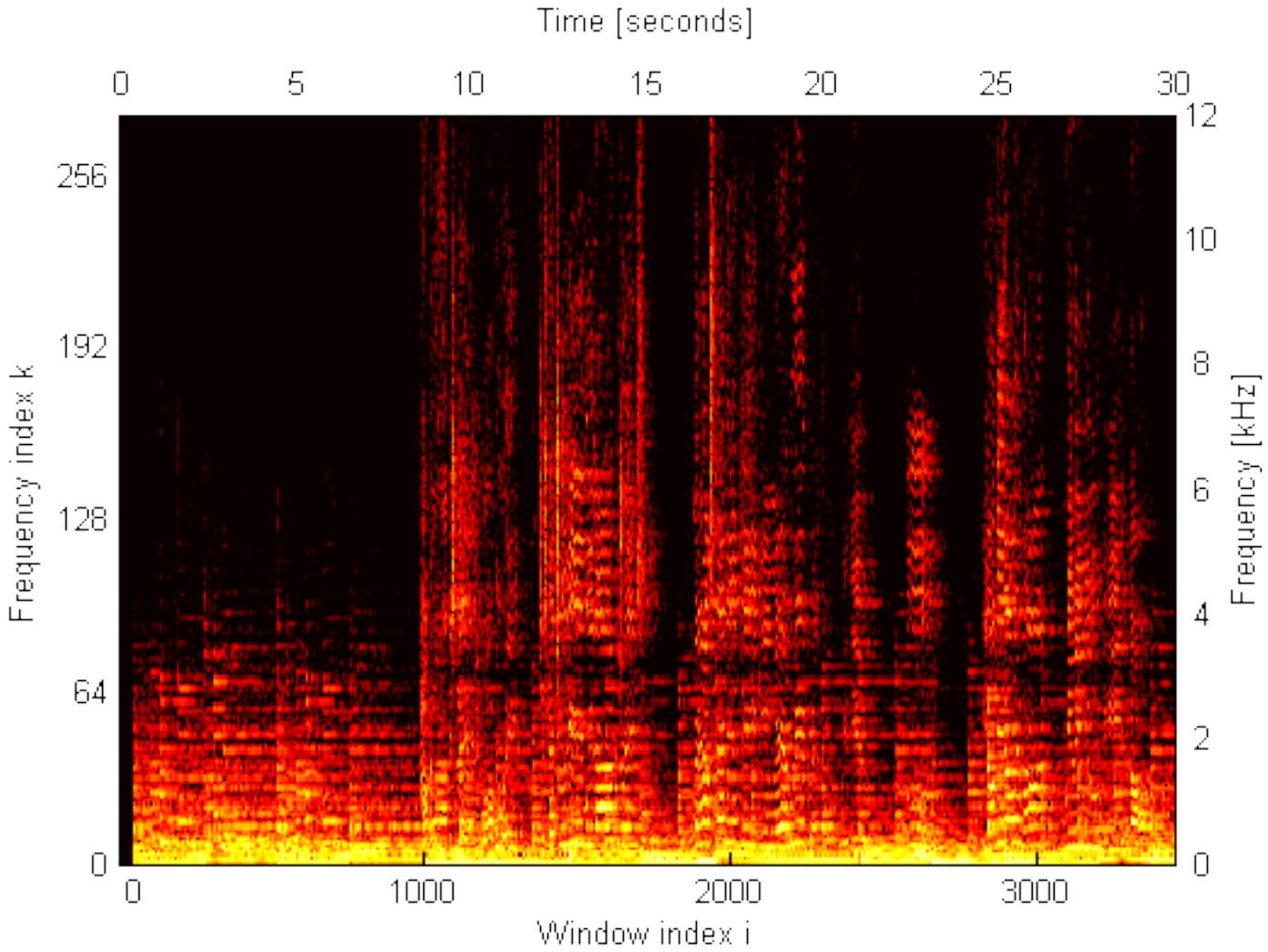
Task: Find out the name of the piece of music, the interpretor, and the name of the CD.

Side conditions:

- This service should be available as a mobile phone application.
- The task should be solved very fast.
- It should work on a very short recording and in a very noisy environment.

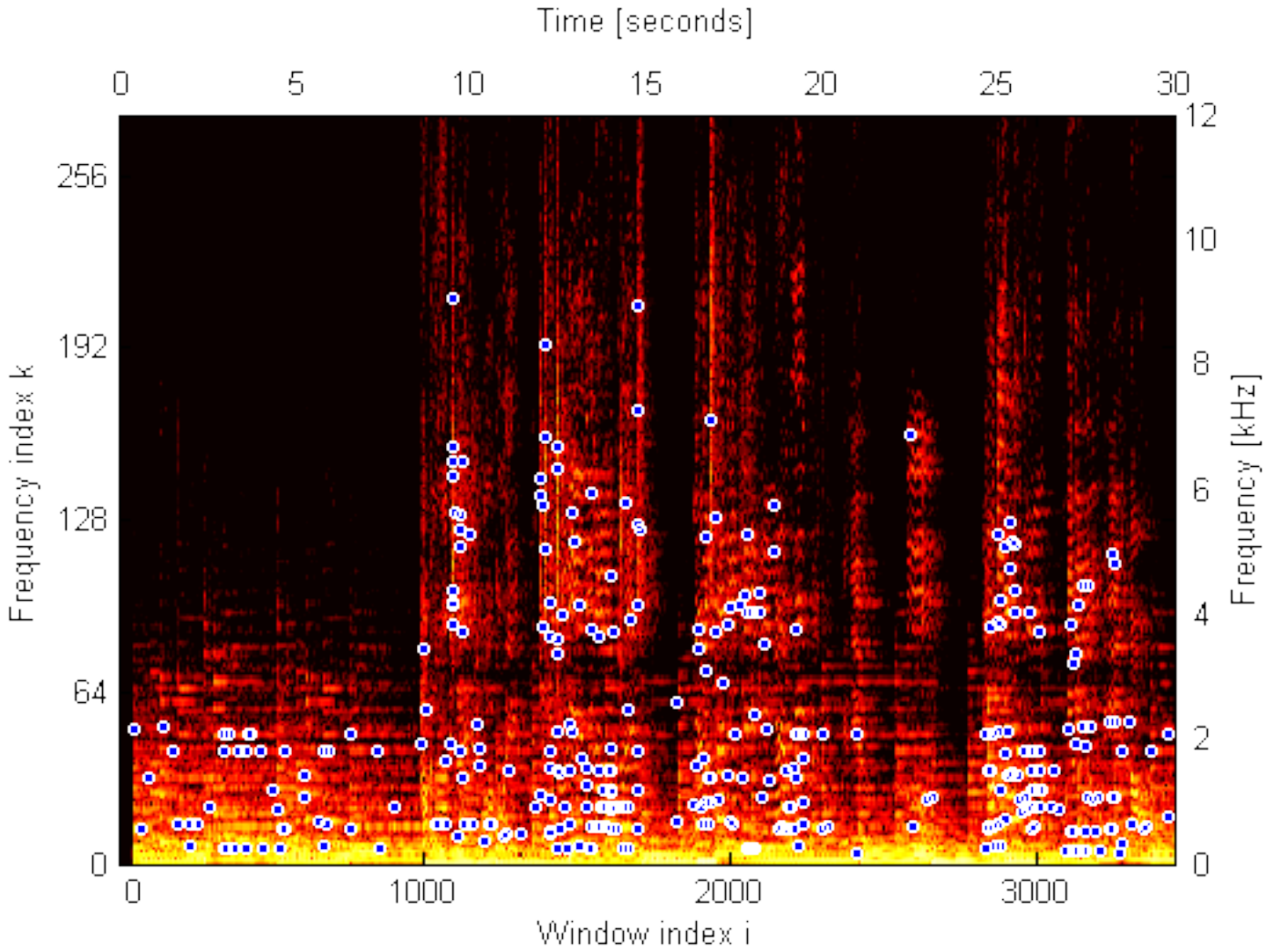
Music representation as a spectrogram

UNIVERSITÄT BONN



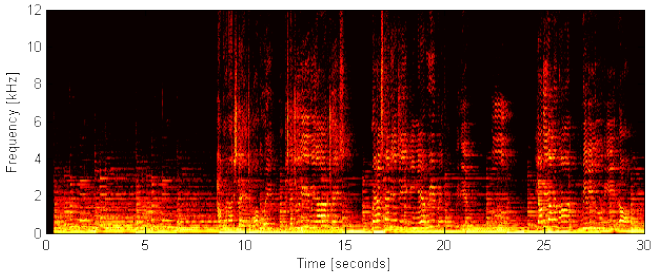
Computation of a digital fingerprint

UNIVERSITÄT BONN



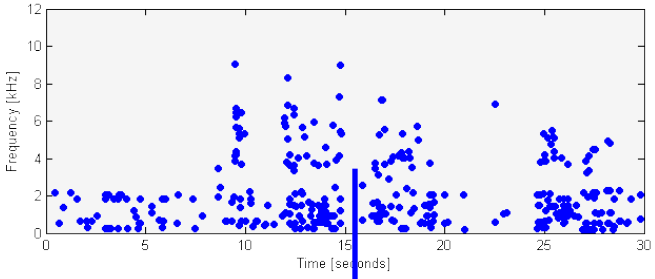
Audio identification ~ data base search

Document

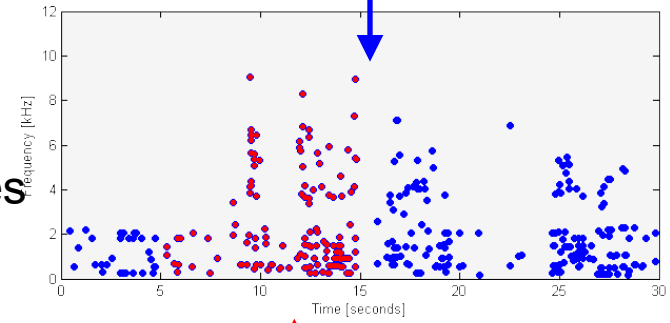


Feature extraction

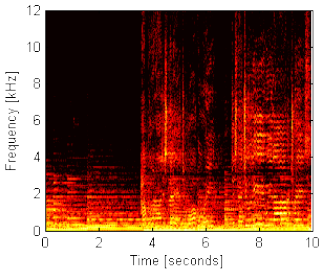
Indexing



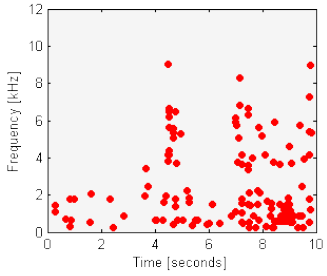
Comparison of feature sequences

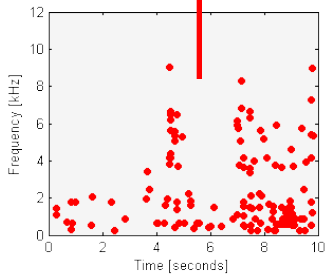


Query



Feature extraction





Match of time-shifted query sequence

Motivating examples: Audio processing tasks
(7) Audio identification

UNIVERSITÄT BONN

[2001] AudioID Fraunhofer IIS

Philips Audio Hashing

Audentify Uni Bonn

[2002] Launch of Shazam service

[2009] Shazam as an app

[2011] Top Apps (Apple)

SoundHound (top @paid)

Shazam (4 th @free)

[2012] Titles available

10 Mio. (Shazam)

28 Mio. (Gracenote)

Foundations of Signal Processing
Contents

UNIVERSITÄT BONN