Identification of Musical Instruments by means of the Hough-Transformation

Christian Röver[1], Frank Klefenz[2] and Claus Weihs[1]

[1] Fachbereich Statistik
Universität Dortmund
44221 Dortmund, Germany
roever@statistik.uni-dortmund.de

[2] Fraunhofer-Institut für Digitale Medientechnologie
Langewiesener Straße 22
98693 Ilmenau, Germany

March 9, 2004

Overview

1. the Hough-transform
2. application to sound data
3. resulting data format
4. classification approaches
5. results

The Hough-transform

- originally developed for **image processing**[1]: detection of **straight lines**, later generalized to **arbitrary functions/shapes**[2]

- similar to **regression**
  - **robust**
  - **simultaneous** fitting of several lines possible

---

[1] Hough, P.V.C. (1959): Machine analysis of bubble chamber pictures. In: *International conference on high-energy accelerators and instrumentation*. Genève, 554-556.

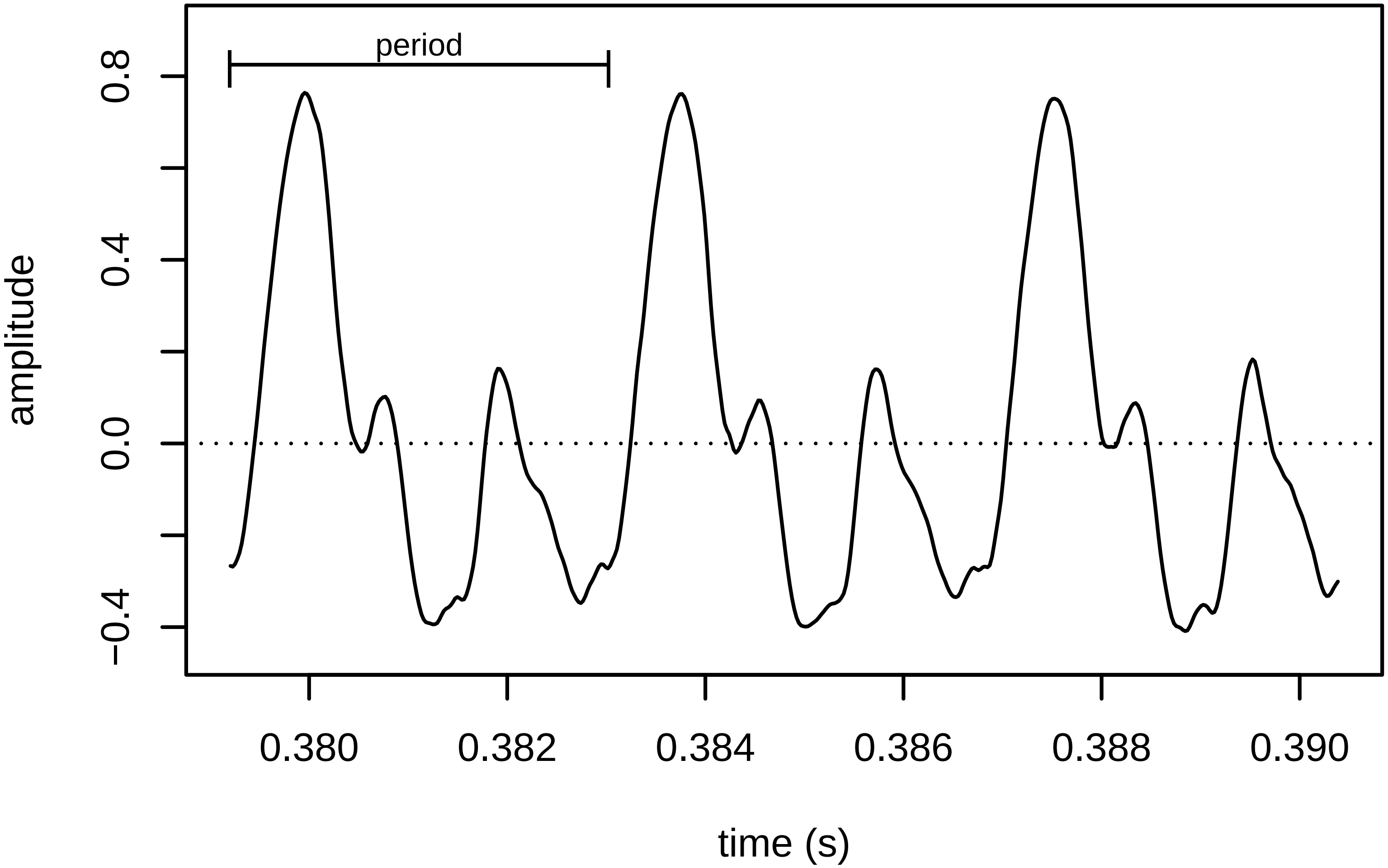
[2] Shapiro, S.D. (1978): Feature Space Transforms for Curve Detection. *Pattern Recognition, 10, 129–143*.

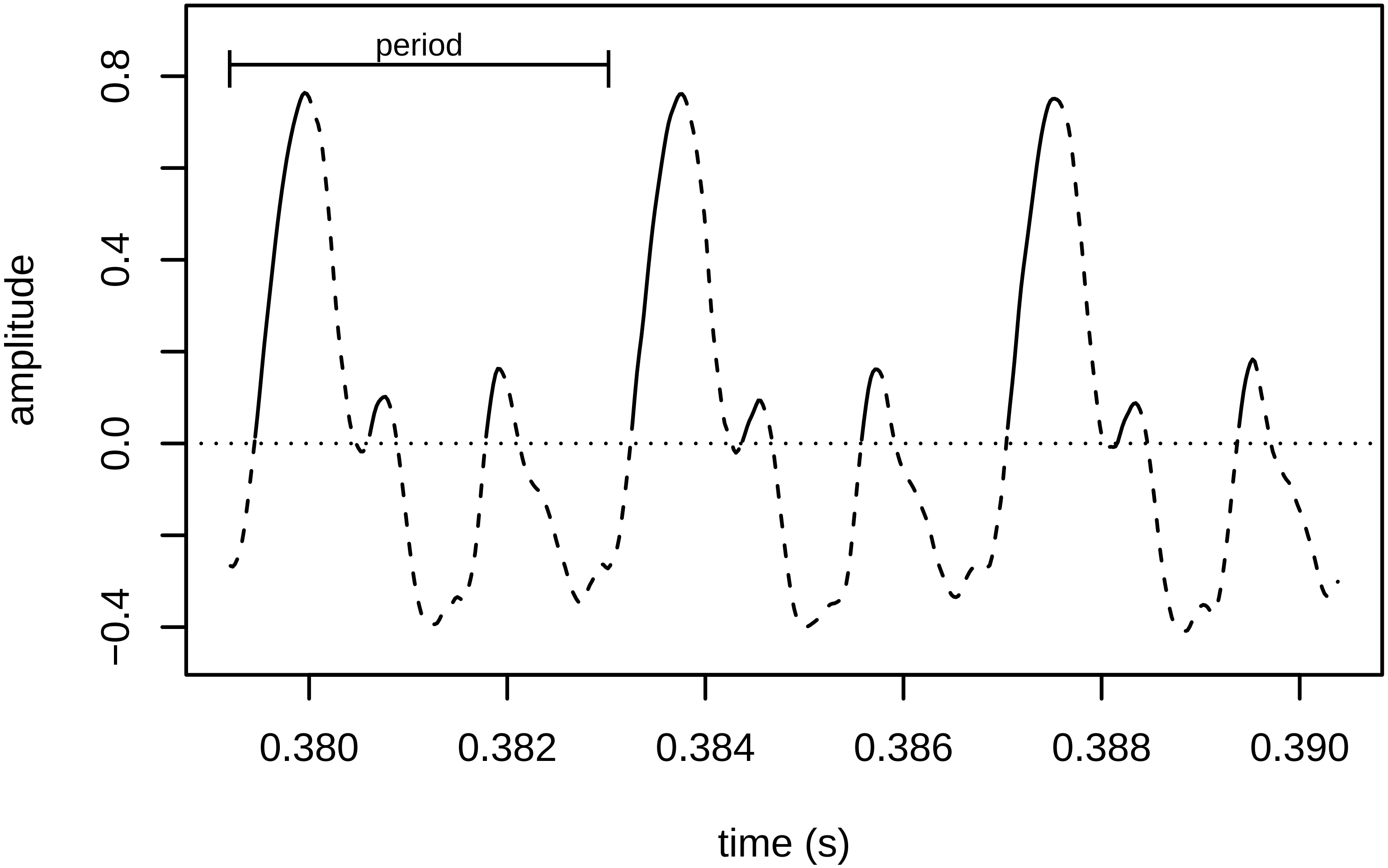
Audio data

- apply to (digital) audio data
- **motivation**: characterize sounds by **oscillation pattern**

➜ does that lead to useful sound characterization?

➜ check by trying to recognize sounds


period
amplitude
−0.4
0.0
0.4
0.8
0.380
0.382
0.384
0.386
0.388
0.390
time (s)


period
amplitude
0.8
0.4
0.0
−0.4
0.380
0.382
0.384
0.386
0.388
0.390
time (s)

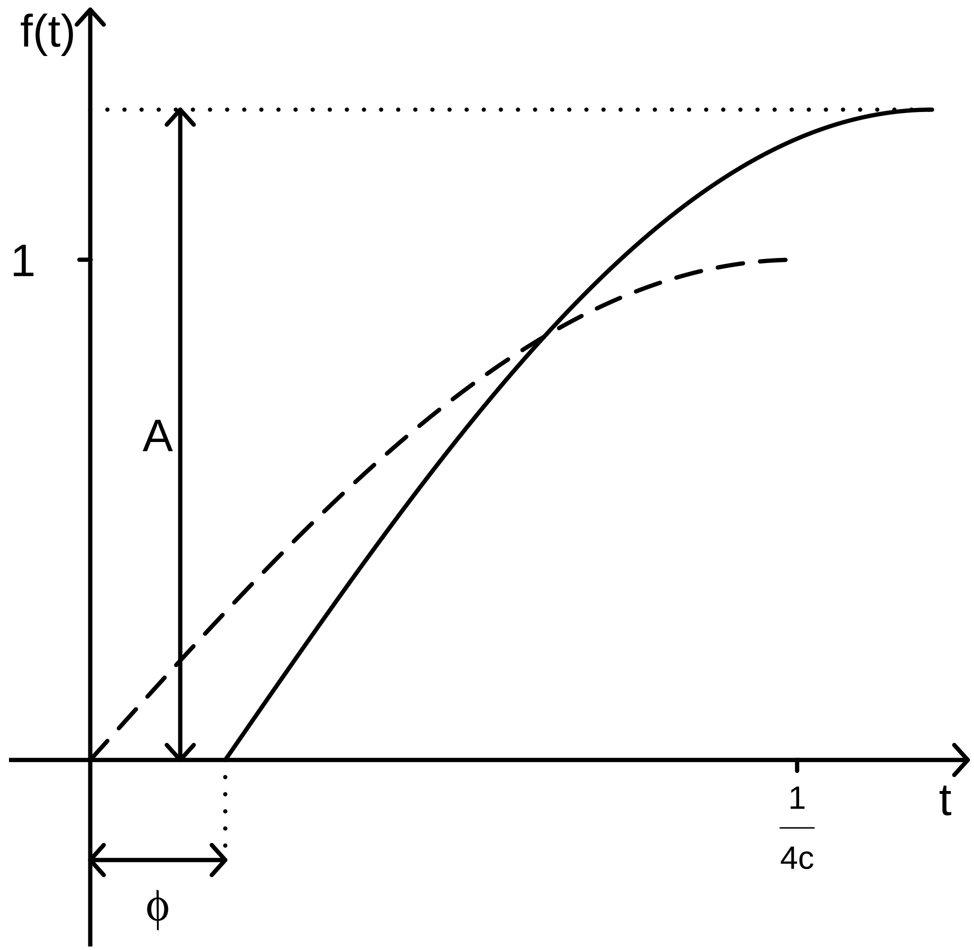
Transform parameter setting

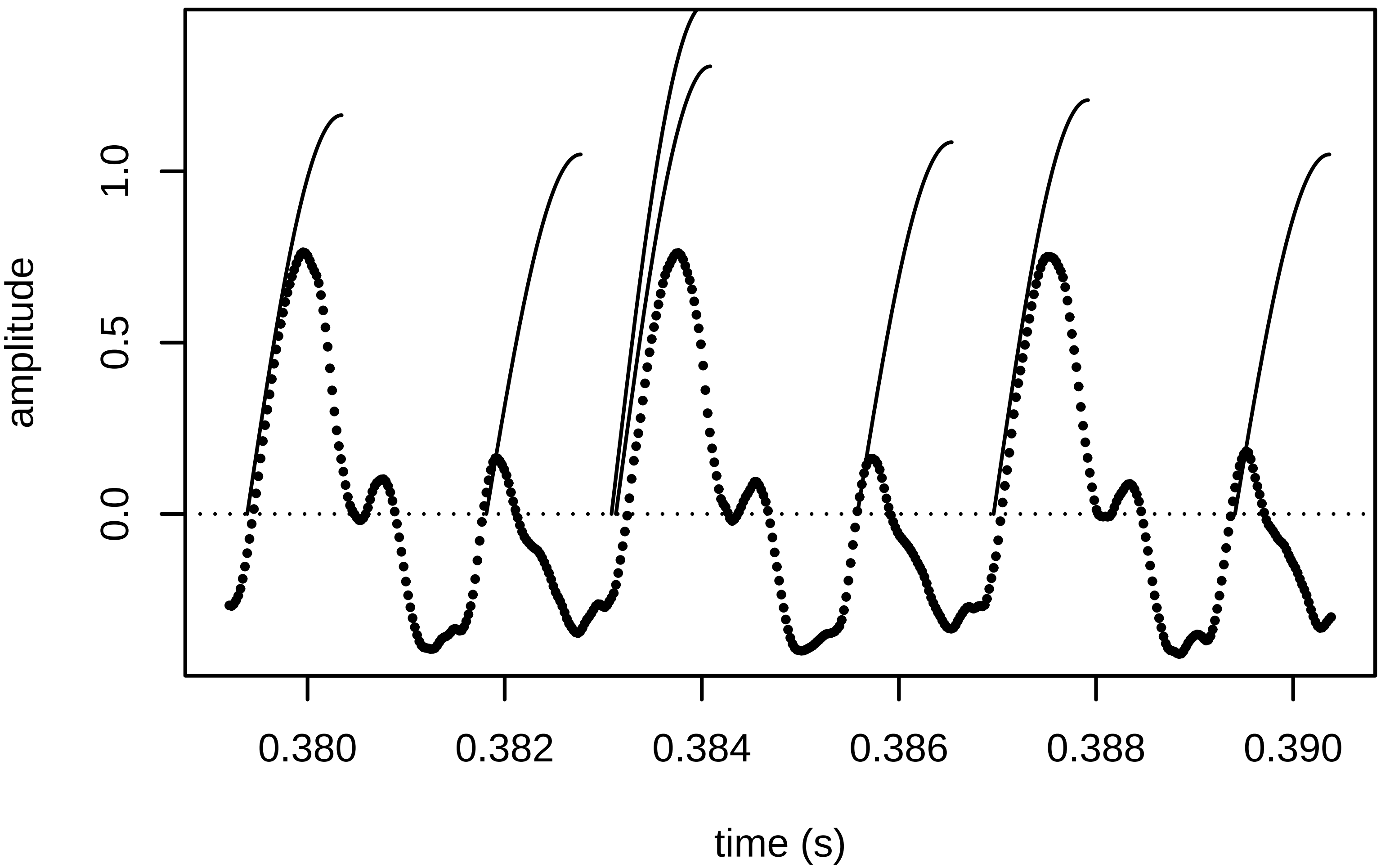
- focus on **signal edges**
- fit a **sinusoidal function** to sound samples:

$$f(t) = A \cdot \sin(2\pi c \cdot t - \varphi) \qquad (\varphi \leq t \leq \varphi + \tfrac{1}{4c})$$

$A \geq 1$:	amplitude	$\longrightarrow$	**slope**
$\varphi \geq 0$:	phase difference	$\longrightarrow$	**time**
c :	center frequency	(fixed)	

$$f(t) = A \cdot \sin(2\pi c \cdot t - \varphi)$$





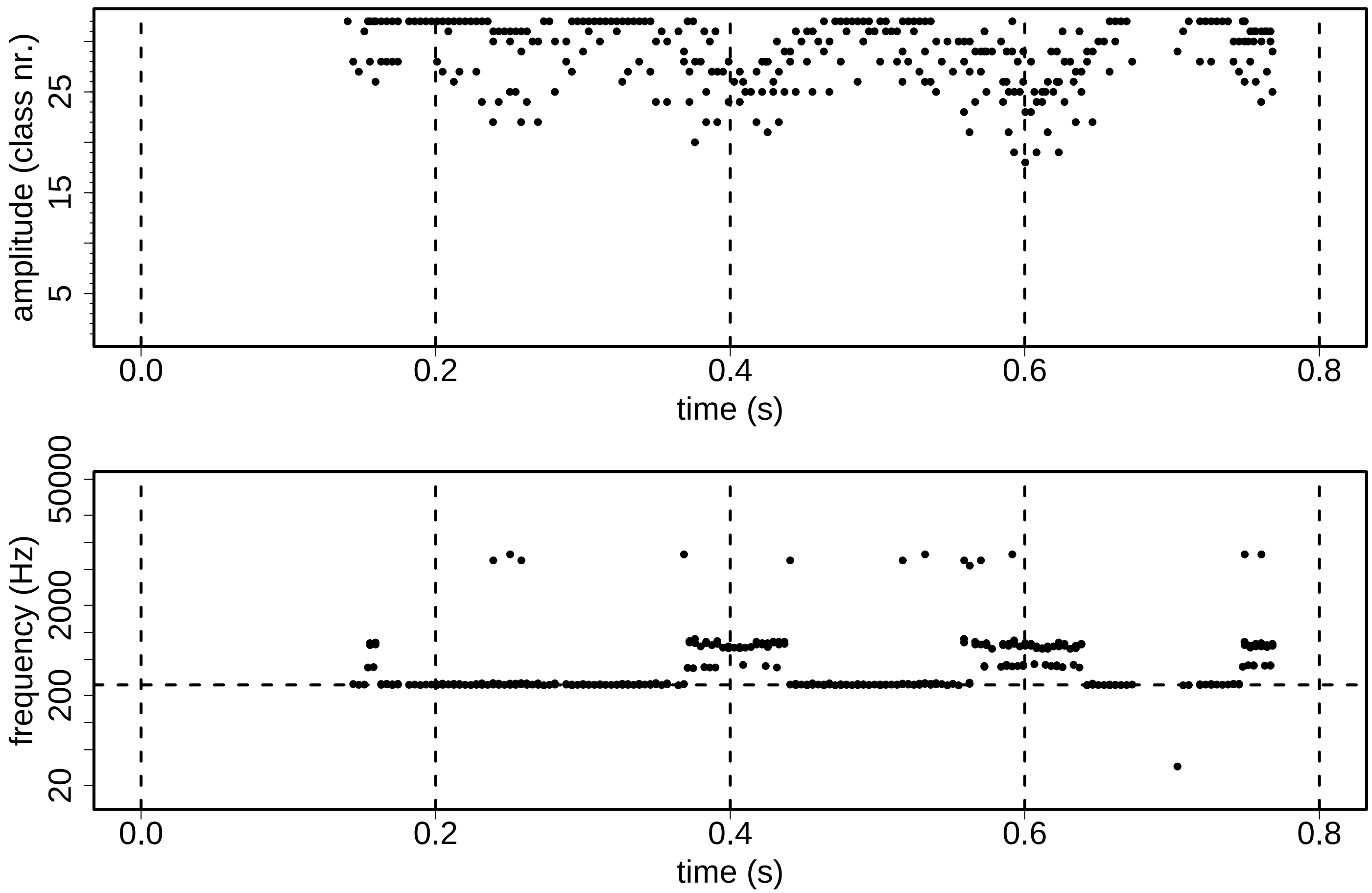
amplitude
0.0
0.5
1.0
0.380
0.382
0.384
0.386
0.388
0.390
time (s)

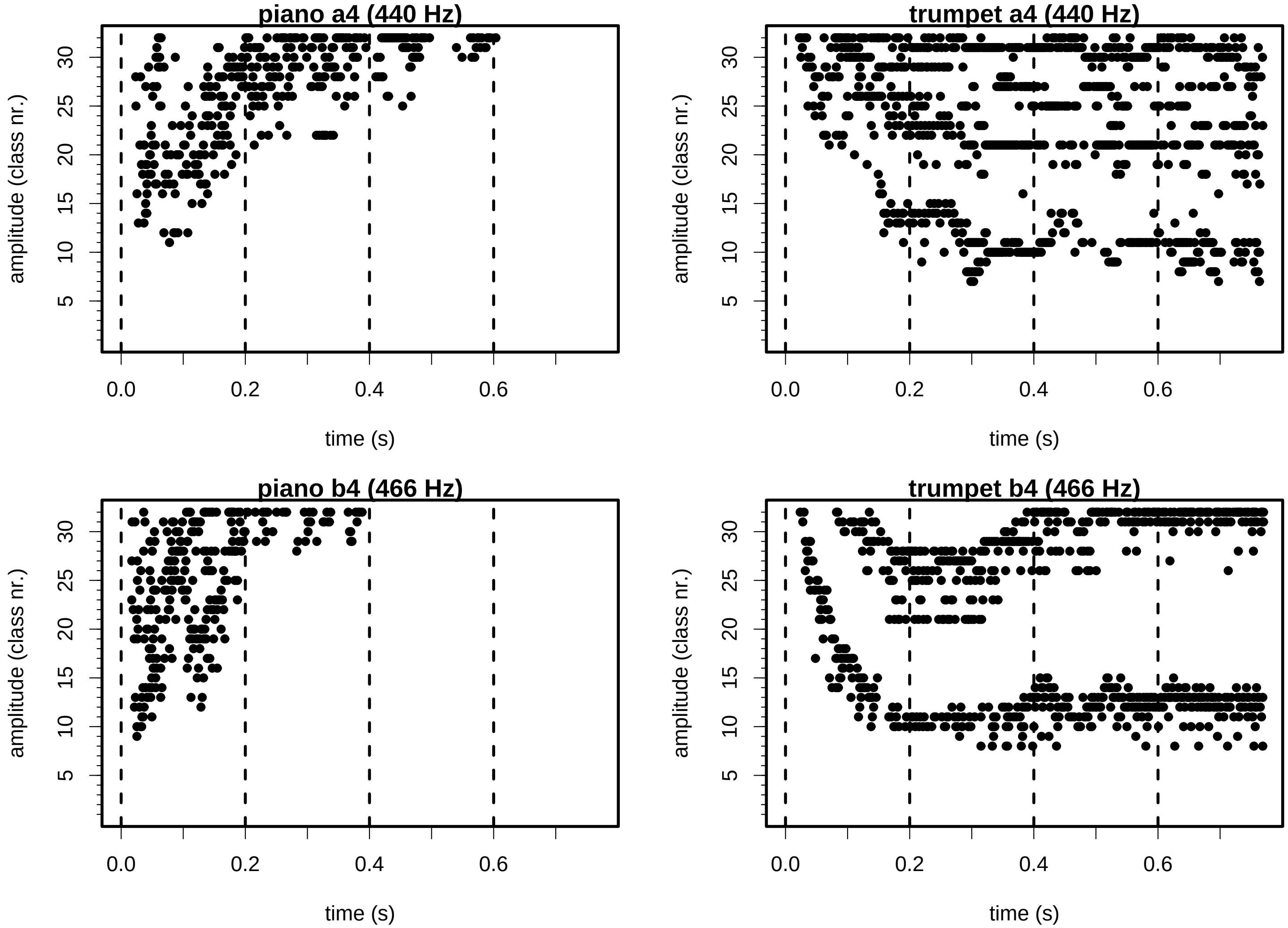
Resulting data

- transformed sound is another **time series**:

	phase difference φ		amplitude A	
Nr.	sample	seconds	class-nr.	value
⋮	⋮	⋮	⋮	⋮
104	16731	0.3793881	28	1.163636
105	16838	0.3818141	31	1.049180
106	16894	0.3830841	22	1.488372
107	19896	0.3831291	25	1.306122
108	17004	0.3855781	30	1.084746
109	17065	0.3869611	27	1.207547
110	17173	0.3894101	31	1.049180
⋮	⋮	⋮	⋮	⋮

Amplitude (A) and Frequency $\left(\frac{1}{\varphi_t - \varphi_{t-1}}\right)$ over time




piano a4 (440 Hz)
trumpet a4 (440 Hz)
piano b4 (466 Hz)
trumpet b4 (466 Hz)
amplitude (class nr.)
time (s)
5
10
15
20
25
30
0.0
0.2
0.4
0.6

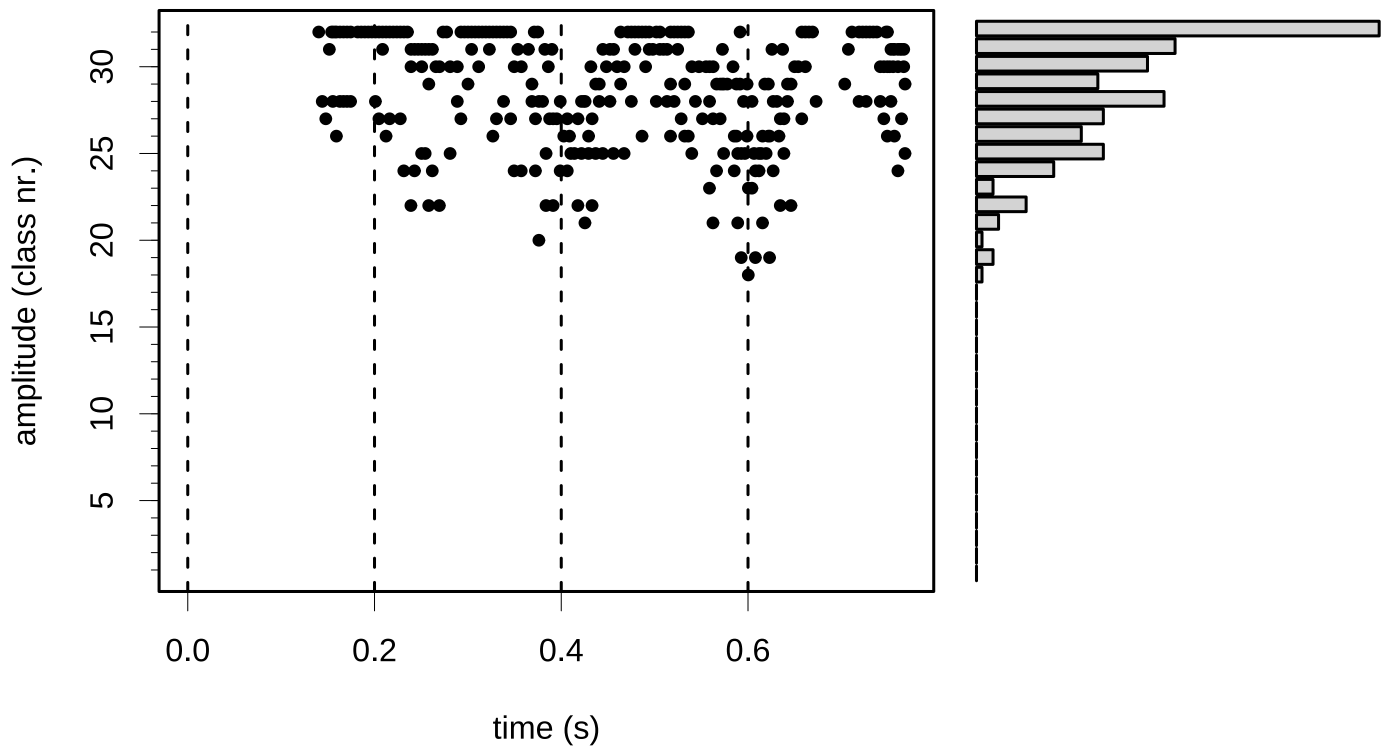
Classification

➜ **How** can we use transformed data for classification?

- **first approach**:
  do **frequencies** of the 32 possible **amplitude values** yield a sufficient ('spectrum-like') sound characterization?

- **second approach**:
  derive **characterizing variables**
  - **characterize** (marginal) **distributions** of amplitudes and frequencies
  - **characterize** distribution over time: **autocorrelation** and **trend**
  - **. . .**

First approach



➜ 32 variables + pitch = **33 total**

Second approach

- transform **durations** between signal edges into **frequencies**
- **mean** amplitude, **mean** frequency
- amplitude **trend** over time
- **autocorrelation** of amplitudes
- . . .

➜ **62 variables** total

Data

- investigated **data set**[3]: 1987 digitized sounds
  (CD-quality — 44.1 kHz, 16 bit, mono)
  pitches are given

- 62 sequences of ≈32 sounds

- sequences of sounds by same or similar instruments were grouped together (e.g. piano at different volumes or bassoon and contrabassoon)

➜ **25** instrument **classes**

---

[3] Opolko, F., Wapnick, J.: McGill University Master Samples (*CD-Set*). 1987. See `http://www.music.mcgill.ca/resources/mums/html/`

Applied methods

- **LDA**: Linear Discriminant Analysis
- **QDA**: Quadratic Discriminant Analysis
- **naive Bayes**
- **RDA**: Regularized Discriminant Analysis
- **Support Vector Machine**
- **Classification Tree**
- **k-NN**: k-Nearest-Neighbour

Regularized Discriminant Analysis (RDA)[4]

- QDA-like; covariance matrix is **manipulated** using **two parameters**
- only **one** of them improved classification
- **class** k **covariance matrix** estimate reduces to:

$$\hat{\Sigma}_k^{\mathrm{RDA}} = \lambda \hat{\Sigma}^{\mathrm{LDA}} + (1-\lambda)\hat{\Sigma}_k^{\mathrm{QDA}} \qquad (0 \leq \lambda \leq 1)$$

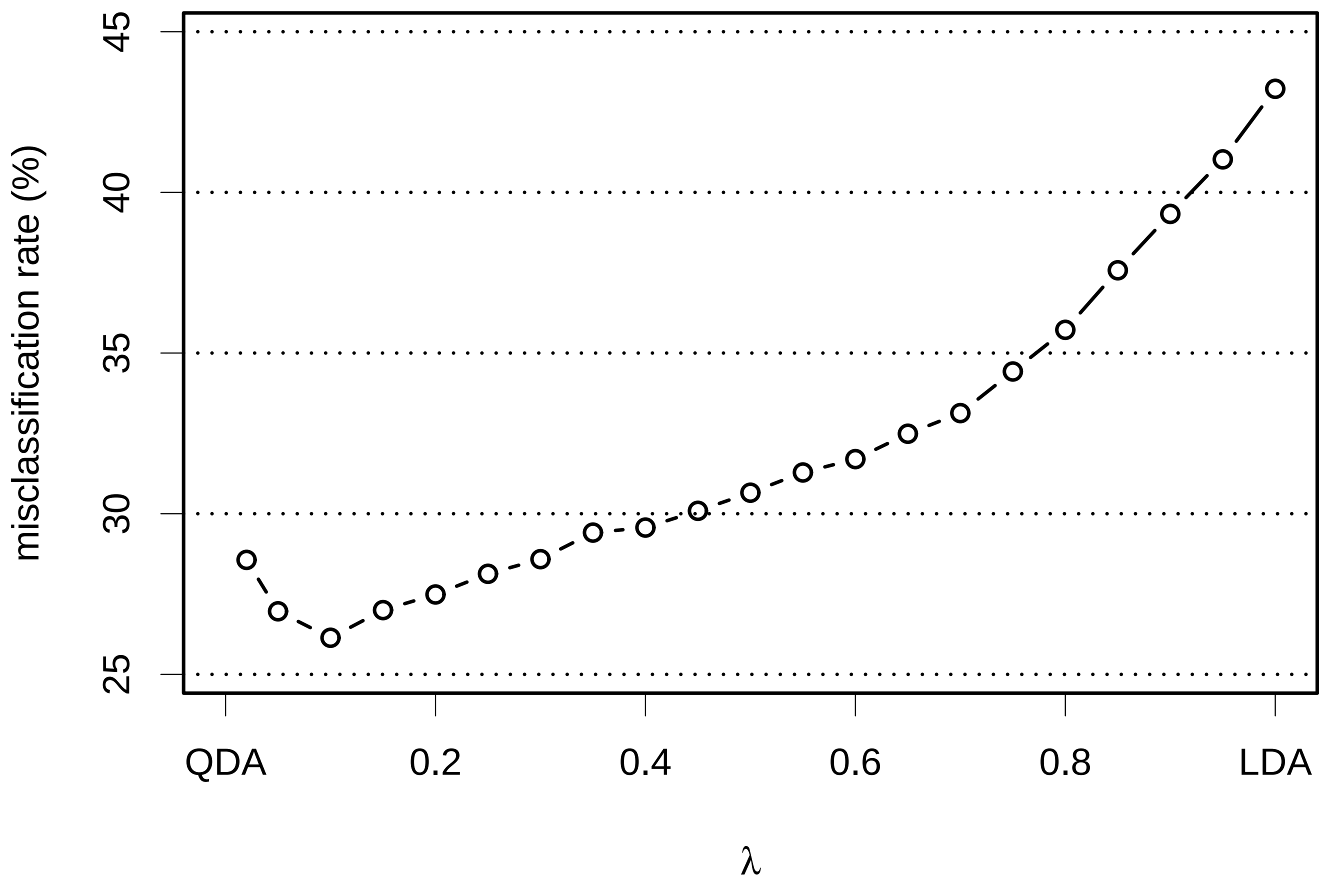
- $\lambda = 0 \quad \longrightarrow \quad$ **QDA**
  $\lambda = 1 \quad \longrightarrow \quad$ **LDA**

[4] Friedman, J.H. (1989): Regularized Discriminant Analysis. *Journal of the American Statistical Association, 84, No. 405, 165–175*

Variable selection

- necessary for second approach (not appropriate in first approach)
- performed **iteratively** in a **stepwise** manner:
  - **start** with pitch only
  - in every step **include** variable that leads to greatest **misclassification rate improvement**
  - **misclassification rate** estimated by **cross-validation**

RDA-parameter tuning



Applied methods

- **first approach** (amplitude frequencies):
  best result: **66% error rate** using **k-Nearest-Neighbour**

- **second approach** (characterizing variables):
  final result: **26.1% error rate** using **Regularized Discriminant Analysis (RDA)** with **11 variables** and $\lambda = 0.1$

Discriminating features

- **pitch**
- **waiting time** for first edge and sound **duration**
- signal edge **rate** (per second)
- **mean**, **variance** and **shape** of amplitude distribution
- **trend** of amplitudes
- **mean** and **variance** of frequency distribution
- **correlation** of amplitude and frequency

Comparing the results

- final **misclassification rate**: 26.1%
- misclassification rate by **guessing**: $\frac{24}{25} = 96\%$
- rates achieved by **humans**: $\approx 44\%$
- rates by **automatic recognition**[5]: ≈ 19 – 7.2%

[5] Bruderer, M.J. (2003): *Automatic recognition of musical instruments*, Masters Thesis, Ecole Polytechnique Federale de Lausanne.

Conclusions

➜ Hough-transformation yields useful **characterization** of a sound

➜ classification results achieved with RDA better than human, still worse than with other approaches
(comparable?)

- open questions:
  noise sensitivity?
  other transform parameter settings?
  . . .