Pattern Recognition and Machine Learning

Part 1: Introduction and Motivation

Gerhard Schmidt

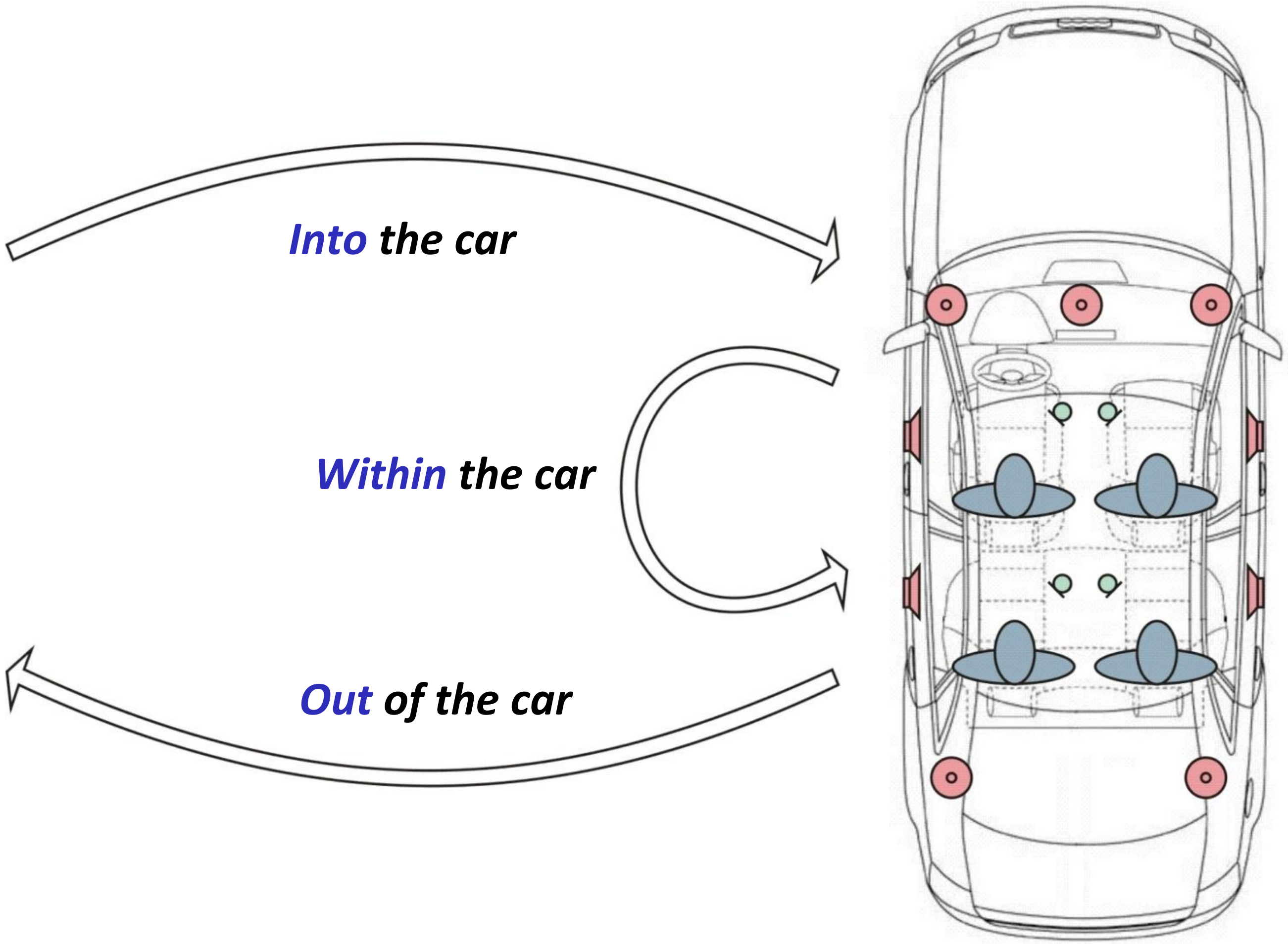
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering
Institute of Electrical and Information Engineering
Digital Signal Processing and System Theory

Contents

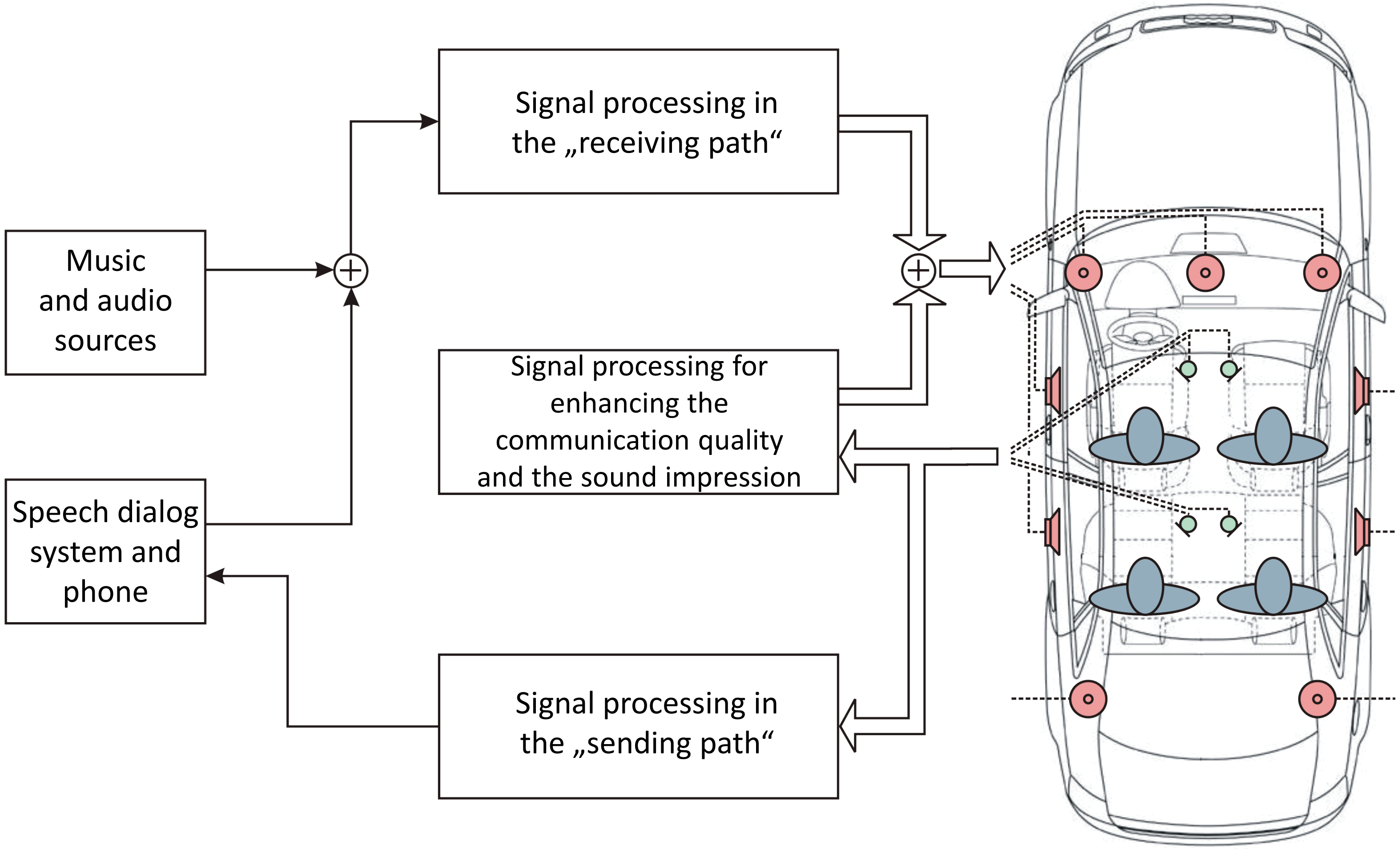
Introduction

- Speech and audio signal paths in a car
- Basics on pattern recognition
- Contents of the lecture
- Boundary conditions of the lecture (exercises, exam, etc.)
- Notation used in the lecture
- Literature
- Automotive examples (with some "medical touch")

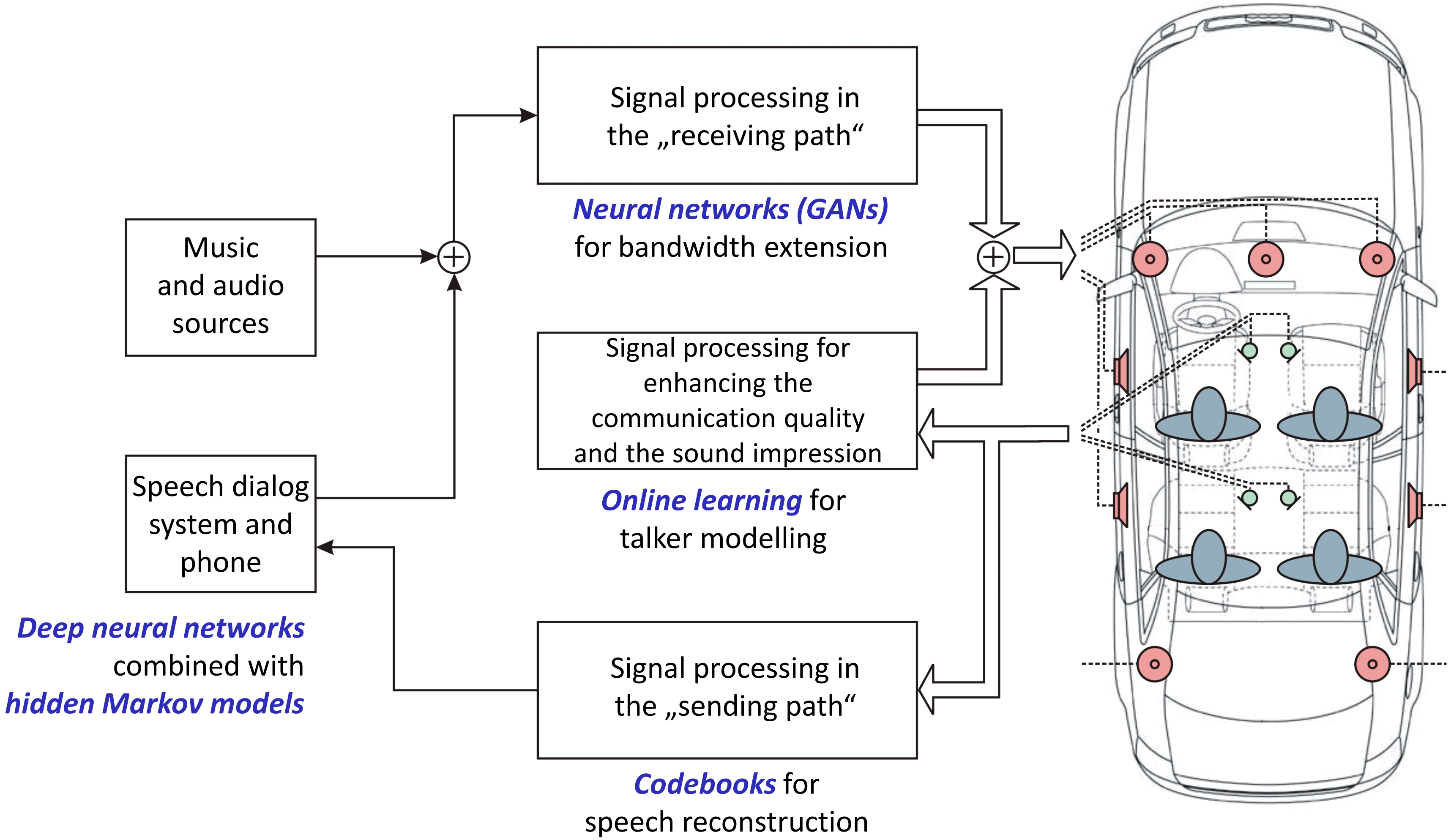
Speech and Audio Signal Paths in a Car



Speech and Audio Signal Paths in a Car



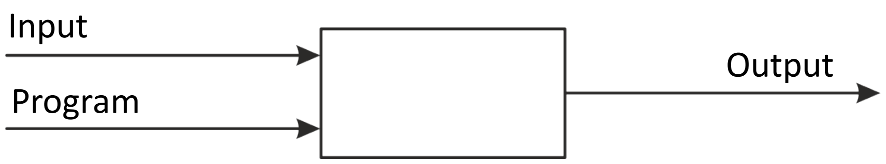
Speech and Audio Signal Paths in a Car



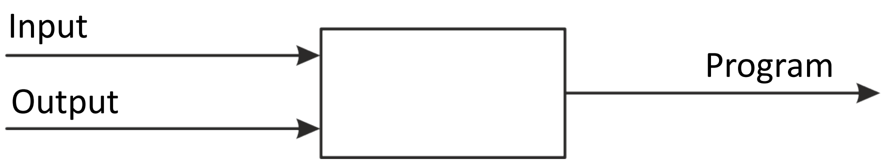
Christian-Albrechts-Universität zu Kiel

Some Basics on the Ingredients of Pattern Recognition

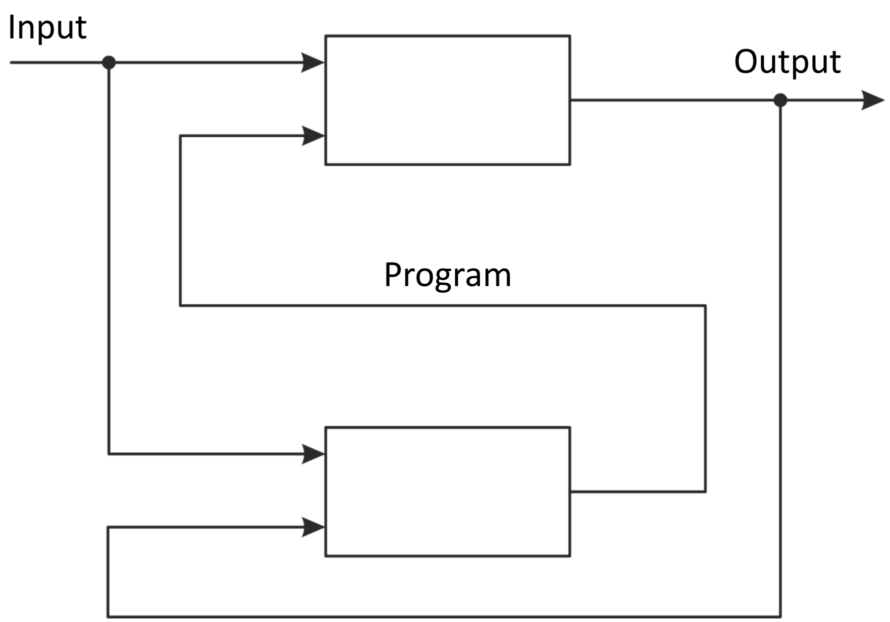
Basic signal processing:



Machine learning / pattern recognition:

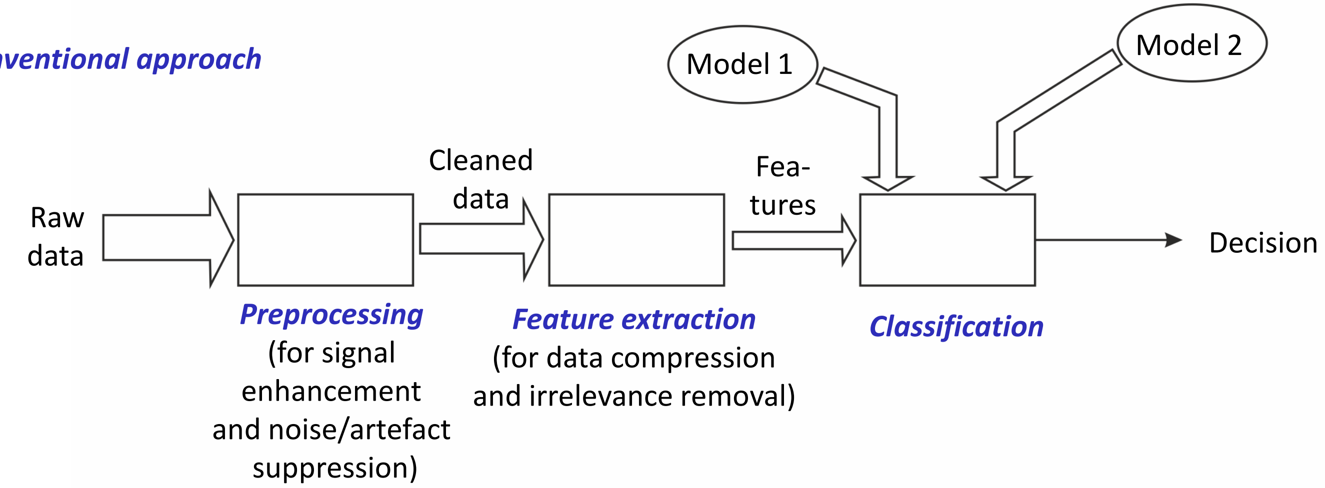


"Fully installed" systems:



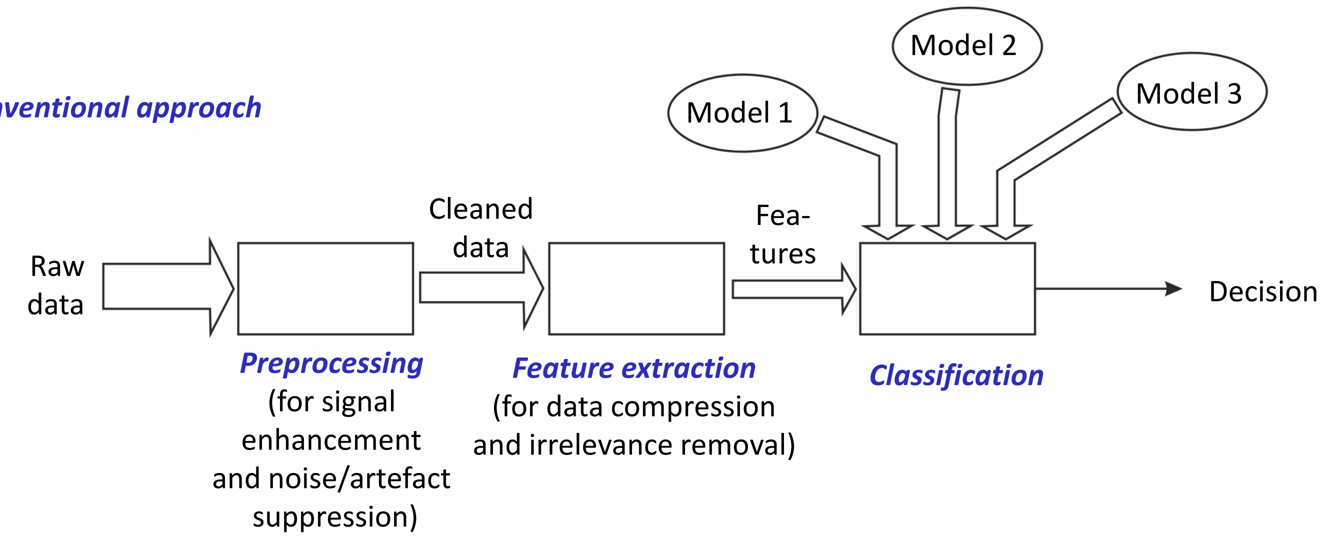
Some Basics on the Ingredients of Pattern Recognition – Conventional versus "End to End"

Conventional approach



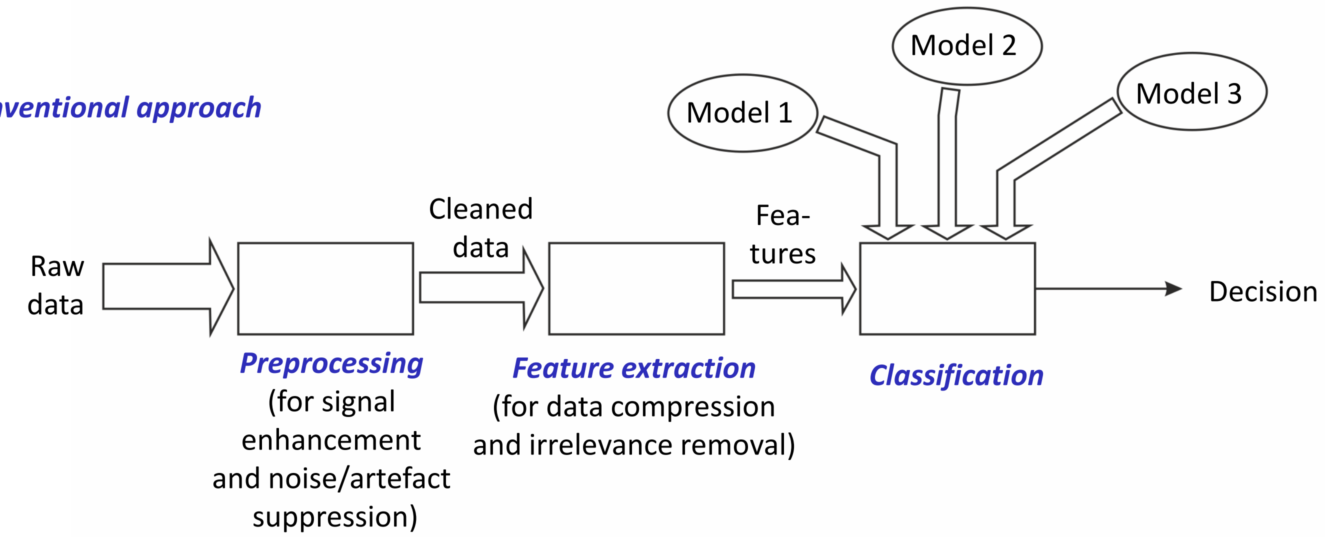
Some Basics on the Ingredients of Pattern Recognition – Conventional versus “End to End”

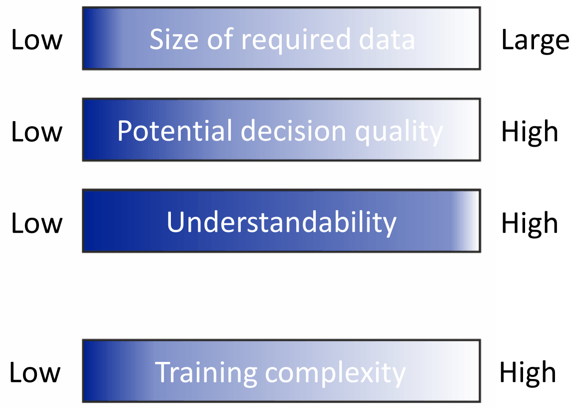
Conventional approach



Some Basics on the Ingredients of Pattern Recognition – Conventional versus "End to End"

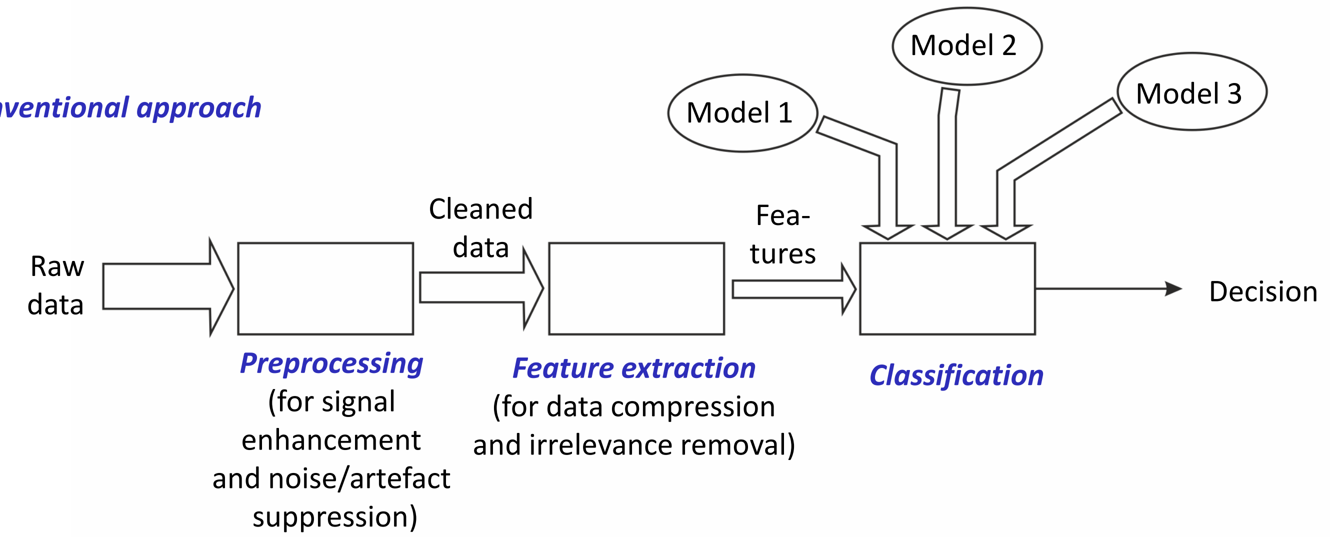
Conventional approach

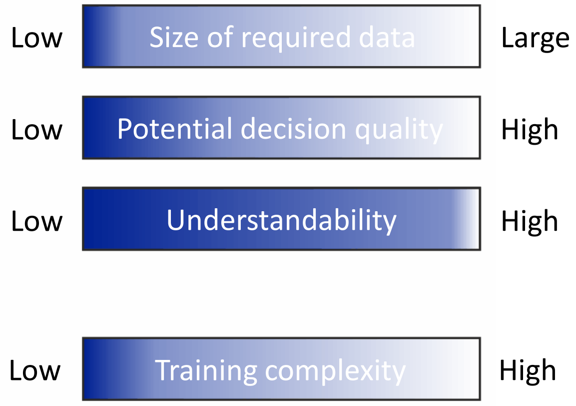




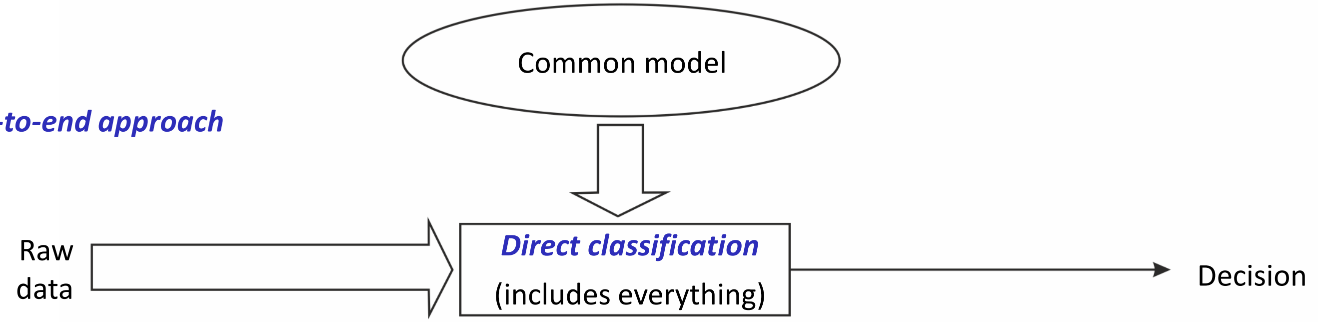
Some Basics on the Ingredients of Pattern Recognition – Conventional versus "End to End"

Conventional approach

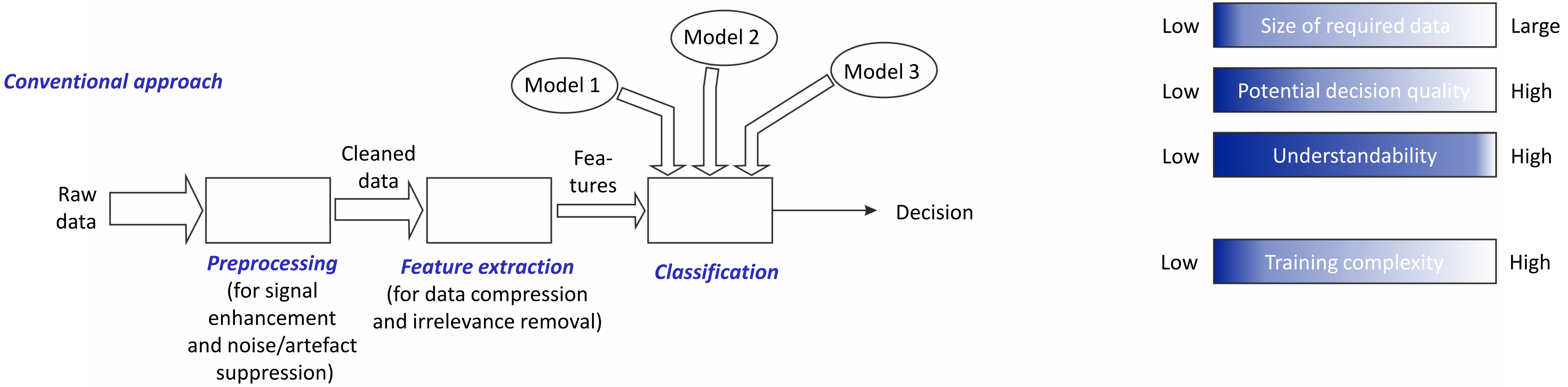


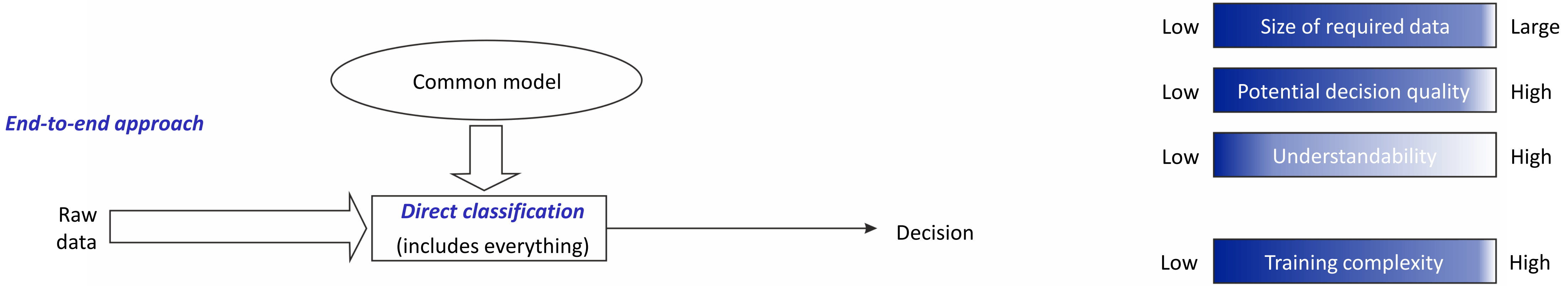


End-to-end approach



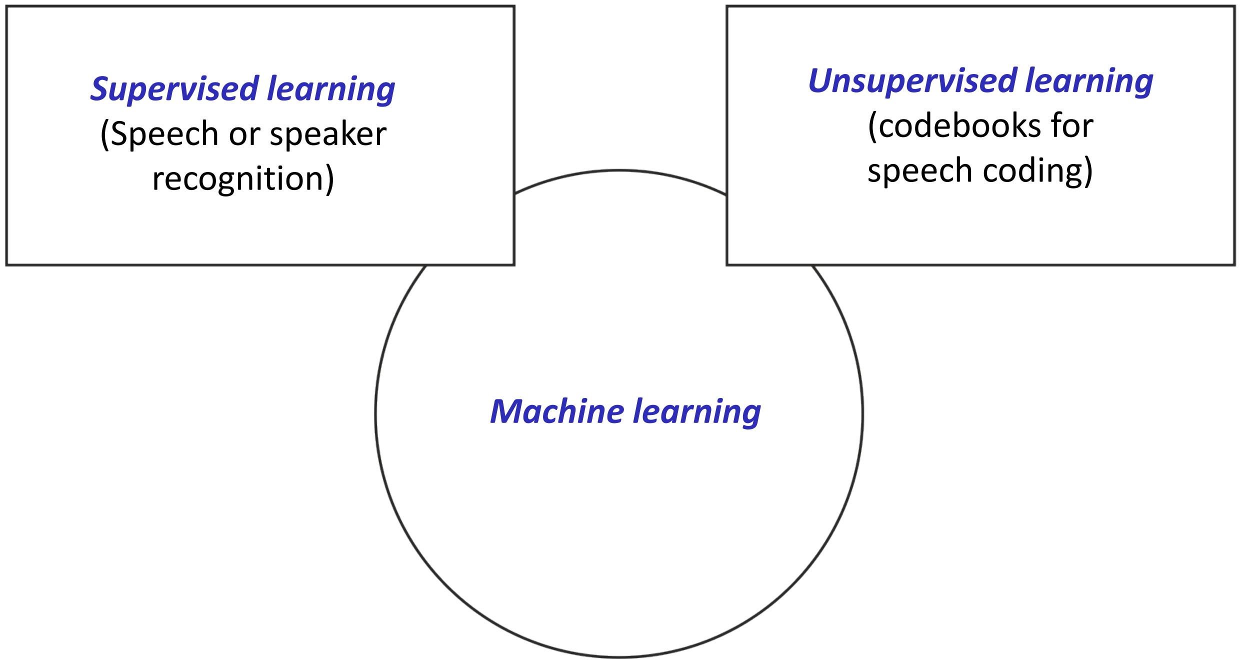
Some Basics on the Ingredients of Pattern Recognition – Conventional versus “End to End”





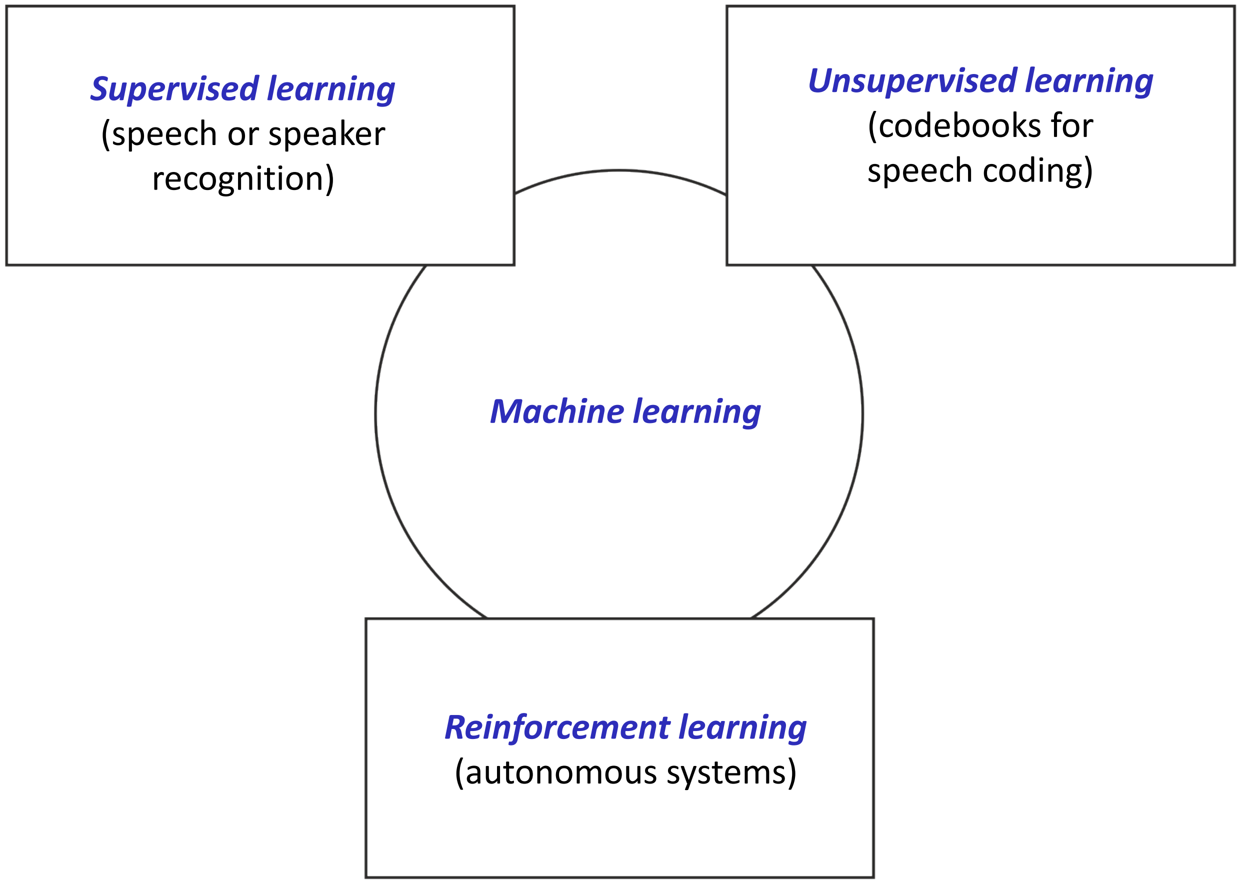
The Different Types of Machine Learning

Supervised versus unsupervised learning:



The Different Types of Machine Learning

Supervised/unsupervised learning versus Reinforcement learning:



Contents of the Lecture (Entire Term)

Preprocessing for improving the „noise robustness“

- Single-channel noise suppression and cost functions
- Beamforming

Data compression

- Feature extraction

Pattern recognition

- Codebooks
- Gaussian mixture models (GMMs)
- Neural networks (NNs)
- Hidden Markov models (HMMs)
- Explainable artificial intelligence (explainable AI)

Exercise

- Python
- TensorFlow

(Programing environment and example data bases available via the our website)

Boundary Conditions of the Lecture

- ***ECTS points***
  - 5 credit points
- ***Oral examination***
  - about 45 minutes per student
  - After the term
- ***Talks (part of the exercise)***
  - About 10 minutes talk plus 5 minutes discussion
  - Topics are available from now on
- ***Lecture slides***
  - In the internet via ***dss-kiel.de***

Notation – Part 1

Scalars:

- Signals: $x(n)$
- Impulse responses (time-variant): $h_i(n)$ ← *Coefficient index*
- Example for a (real) convolution: $y(n) = \sum_{i=0}^{N-1} x(n-i)\, h_i(n)$

Vectors:

Boldface and lowercase

- Signal vectors: $\boldsymbol{x}(n) = \left[x(n),\, x(n-1),\, \dots,\, x(n-N+1)\right]^{\mathrm{T}}$
- Impulse response vectors (time-variant) : $\boldsymbol{h}(n) = \left[h_0(n),\, h_1(n),\, \dots,\, h_{N-1}(n)\right]^{\mathrm{T}}$
- Example for a real convolution: $y(n) = \boldsymbol{x}^{\mathrm{T}}(n)\, \boldsymbol{h}(n) = \boldsymbol{h}^{\mathrm{T}}(n)\, \boldsymbol{x}(n)$

Matrices:

$$\boldsymbol{A}(n) = \begin{bmatrix} a_{00}(n) & a_{01}(n) & \dots & a_{0N}(n) \\ a_{10}(n) & a_{11}(n) & \dots & a_{1N}(n) \\ \vdots & \vdots & & \vdots \\ a_{M0}(n) & a_{M1}(n) & \dots & a_{MN}(n) \end{bmatrix}$$

Boldface and uppercase

Notation – Part 2

Random variables and processes:

- Notation: $x(n),\ x_1(n),\ x_2(n)$

  No differences between deterministic signals and random processes – different writing styles: $\boldsymbol{x}(\eta, n),\ x(\omega, n),\ X(n)$

- Probability density function: $f_x(x, n),\ f_{x_1 x_2}(x_1, x_2, n_1, n_2)$

- Stationary random processes:

$$\begin{aligned} f_x(x,\, n) &= f_x(x,\, n + n_0) \;=\; f_x(x) \\ f_{x_1 x_2}(x_1,\, x_2,\, n_1,\, n_2) &= f_{x_1 x_2}(x_2,\, x_2,\, n_1 + n_0,\, n_2 + n_0) \\ &= f_{x_1 x_2}(x_1,\, x_2,\, n_2 - n_1) \end{aligned}$$

- Expected values of stationary random processes:

$$\mathrm{E}\Big\{x(n)\Big\} = \int\limits_{x=-\infty}^{\infty} x\, f_x(x)\, dx = m_x^{(1)} = m_x$$

$$\mathrm{E}\Big\{x^2(n)\Big\} = \int\limits_{x=-\infty}^{\infty} x^2\, f_x(x)\, dx = m_x^{(2)}, \qquad \mathrm{E}\Big\{g\big(x(n)\big)\Big\} = \int\limits_{x=-\infty}^{\infty} g(x)\, f_x(x)\, dx$$

Notation – Part 3

Auto and cross correlation for real, stationary random processes:

- Auto-correlation function:

$$\mathrm{E}\Big\{x(n)\,x(n+l)\Big\} \;=\; s_{xx}(l)$$

- Cross-correlation function:

$$\mathrm{E}\Big\{x(n)\,y(n+l)\Big\} \;=\; s_{xy}(l)$$

- (Auto) power spectral density:

$$S_{xx}(\Omega) \;=\; \sum_{l=-\infty}^{\infty} \mathrm{E}\Big\{x(n)\,x(n+l)\Big\}\,e^{-j\Omega\,l} \;=\; \sum_{l=-\infty}^{\infty} s_{xx}(l)\,e^{-j\Omega\,l}$$

- (Cross) power spectral density:

$$S_{xy}(\Omega) \;=\; \sum_{l=-\infty}^{\infty} \mathrm{E}\Big\{x(n)\,y(n+l)\Big\}\,e^{-j\Omega\,l} \;=\; \sum_{l=-\infty}^{\infty} s_{xy}(l)\,e^{-j\Omega\,l}$$

Notation – Part 4

Stationary white noise:

- Auto-correlation function:

$$s_{xx}(l)\Big|_{\text{white noise}} = \begin{cases} \sigma_x^2, & \text{if } l = 0, \\ 0, & \text{else.} \end{cases}$$

- Auto power spectral density:

$$S_{xx}(\Omega)\Big|_{\text{white noise}} = \sigma_x^2$$

Application Examples

Just a few examples for pattern recognition and machine learning from our automotive branch:

- Hands-free telephony (cleaning) and dialog systems (recognition)
- Siren detection (detection)
- Bandwidth extension (estimation)
- Prediction of steering movements (recognition/prediction)

Application Examples

A few words before we start:

- Pure research projects have to be high-risk ones to get funding.
- In the following you will see some of the current projects of the DSS group. Some of them are low-risk ones (e.g. Master thesis), some stem out of a high risk class.
- Partly pattern recognition and machine learning is in the focus, partly also new sensor concepts.
- It can not always be guaranteed that the project goals will / can be reached.

Application Examples – Hands-free Telephony and Speech Dialog Systems

Problem:

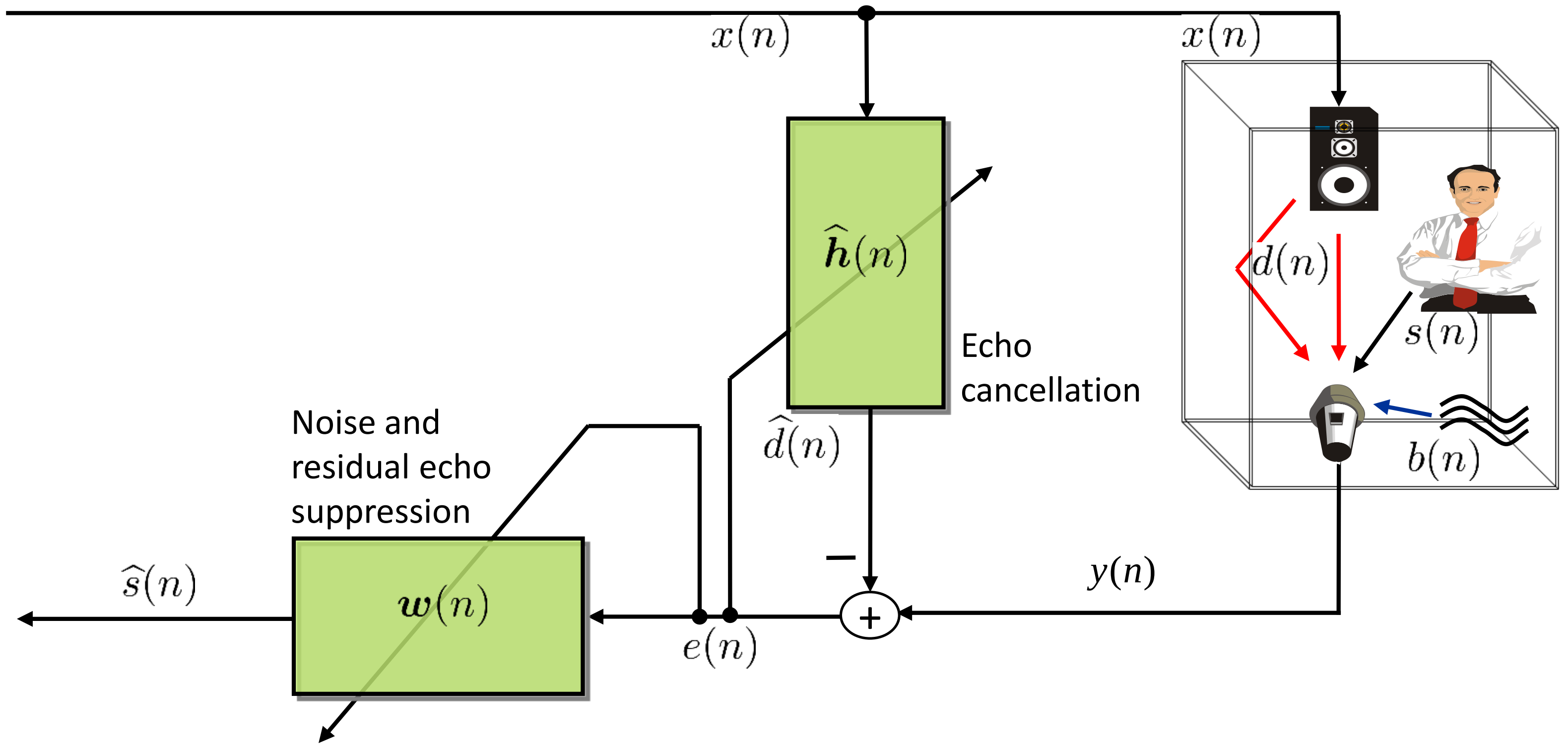
- Speech recordings in a car are superposed with several distortions:
  - Echo components (remote partner, dialog system output)
  - Background noise
  - Embedded speech recognition

Solution:

- Apply adaptive filters to cancel and/or suppress distortions.
- Details will be explained in the next two weeks.
- Further details are in taught in the lecture "Adaptive Filters".

Application Examples – Hands-free Telephony and Speech Dialog Systems

Hands-free telephony – a basic system:

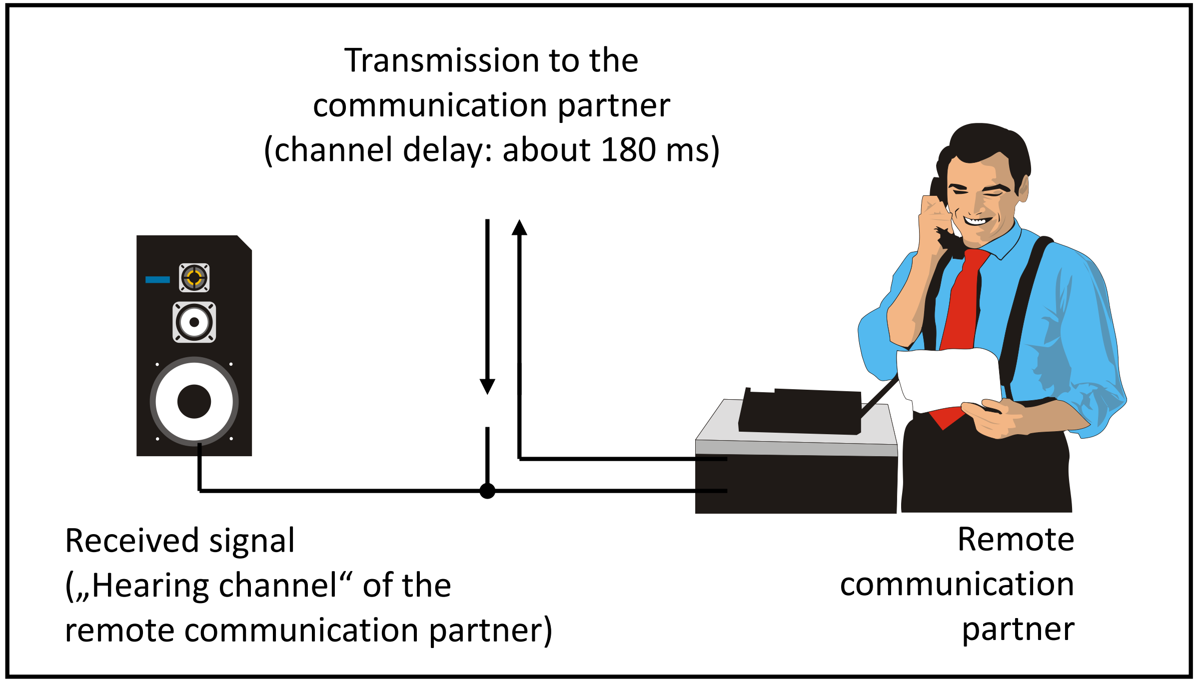


Application Examples – Hands-free Telephony and Speech Dialog Systems

Stereo signals (16 kHz):

Left:	Right:
Received signal ...	Sent signal ...

... of the remote communication partner



Initial filter convergence:

Adaptation at the beginning of the call

Double talk:

Both partners speak simultaneously

Enclosure dislocations:

Without Wiener filter

With Wiener filter

Application Examples – Hands-free Telephony and Speech Dialog Systems

Video from/with:

- Raymond Brückner (SVOX)
- Andreas Löw (SVOX)
- Patrick Langer (SVOX)

Link to video

Application Examples – Detection of Sirens and other “Outside Events”

Problem:

- The sound of sirens is sometimes hard to hear in a car (due to good damping of passenger compartments or due to loud music) and it’s hard to estimate the direction of the source.

Solution:

- Pattern recognition on the basis of microphone signals.
- Usually not much better as a human listener.

Application Examples – Detection of Sirens and other “Outside Events”

Problem:

- The sound of sirens is sometimes hard to hear in a car (due to good damping of passenger compartments or due to loud music) and it’s hard to estimate the direction of the source.

Solution:

- Pattern recognition on the basis of microphone signals.
- Usually not much better as a human listener.
- New sensors (outside microphones, in cooperation with a small company in Revensdorf called “mechakustik”)
  - Extremely high demands concerning robustness.
  - Large potential to be better than a human listener.

Application Examples – Bandwidth Extension for Speech Signals

Problem:

- Current connections have a limited bandwidth (telephone speech quality).
- New standards such as HD voice are better, but the are still limited (in terms of bandwidth and speech quality in general).

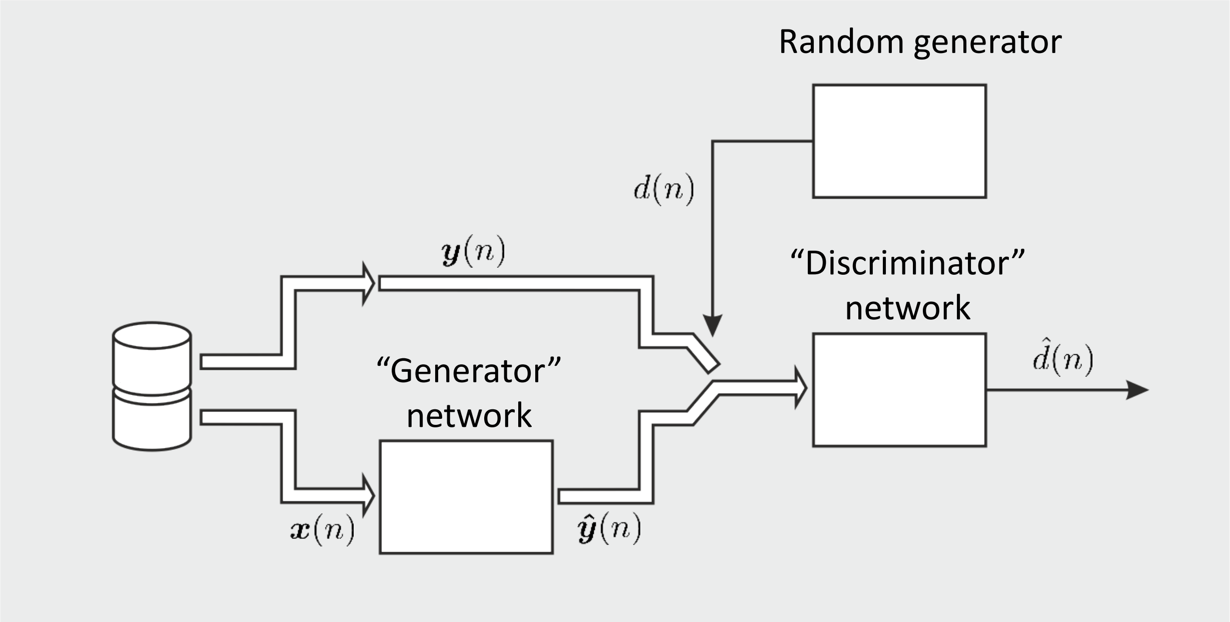
Solution:

- Estimation of the missing frequency components by means of pattern recognition and adding those on the receiver side.
- Usage of new network types (GAN approaches).

Application Examples – Bandwidth Extension for Speech Signals

Basic approach:

- Current connections have a limited bandwidth (telephone speech quality).
- New standards such as HD voice are better, but the are still limited (both, in terms of bandwidth and speech quality in general).

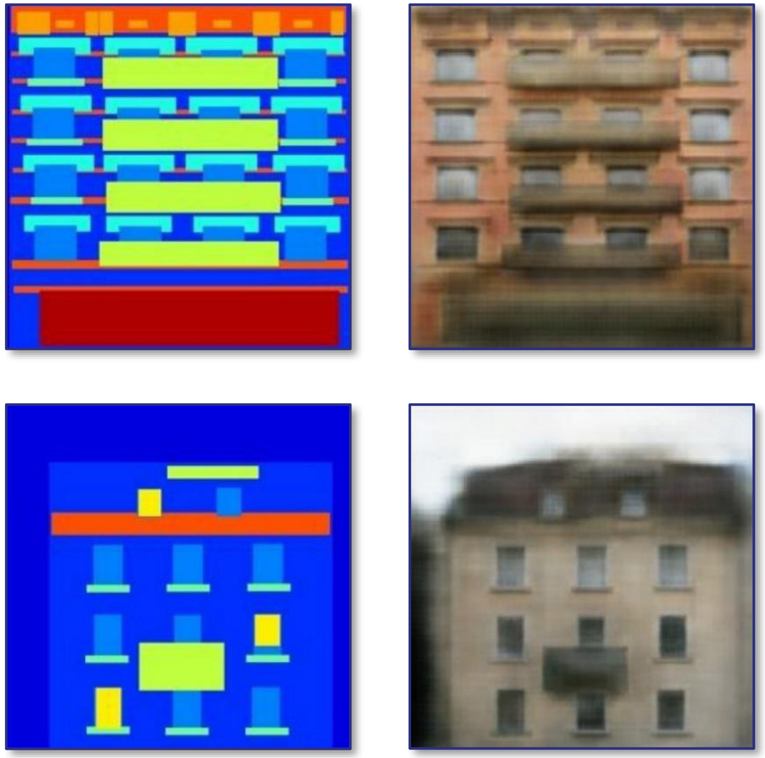


Application Examples – Bandwidth Extension for Speech Signals

Basic approach:

- Example for a so-called GAN approach:
  - Reconstruction of houses on the basis of rectangles
  - Conventional approach in the second column.

Source: P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros: ***Image-to-Image Translation with Conditional Adversarial Networks***, CoRR, vol. abs/1611.07004, 2016.



Input | Output of a conventionally trained network

Application Examples – Bandwidth Extension for Speech Signals

Basic approach:

- Example for a so-called GAN approach:
  - Reconstruction of houses on the basis of rectangles
  - Conventional approach in the second column.
  - GAN approach in the third column.

Source: P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros: ***Image-to-Image Translation with Conditional Adversarial Networks***, CoRR, vol. abs/1611.07004, 2016.

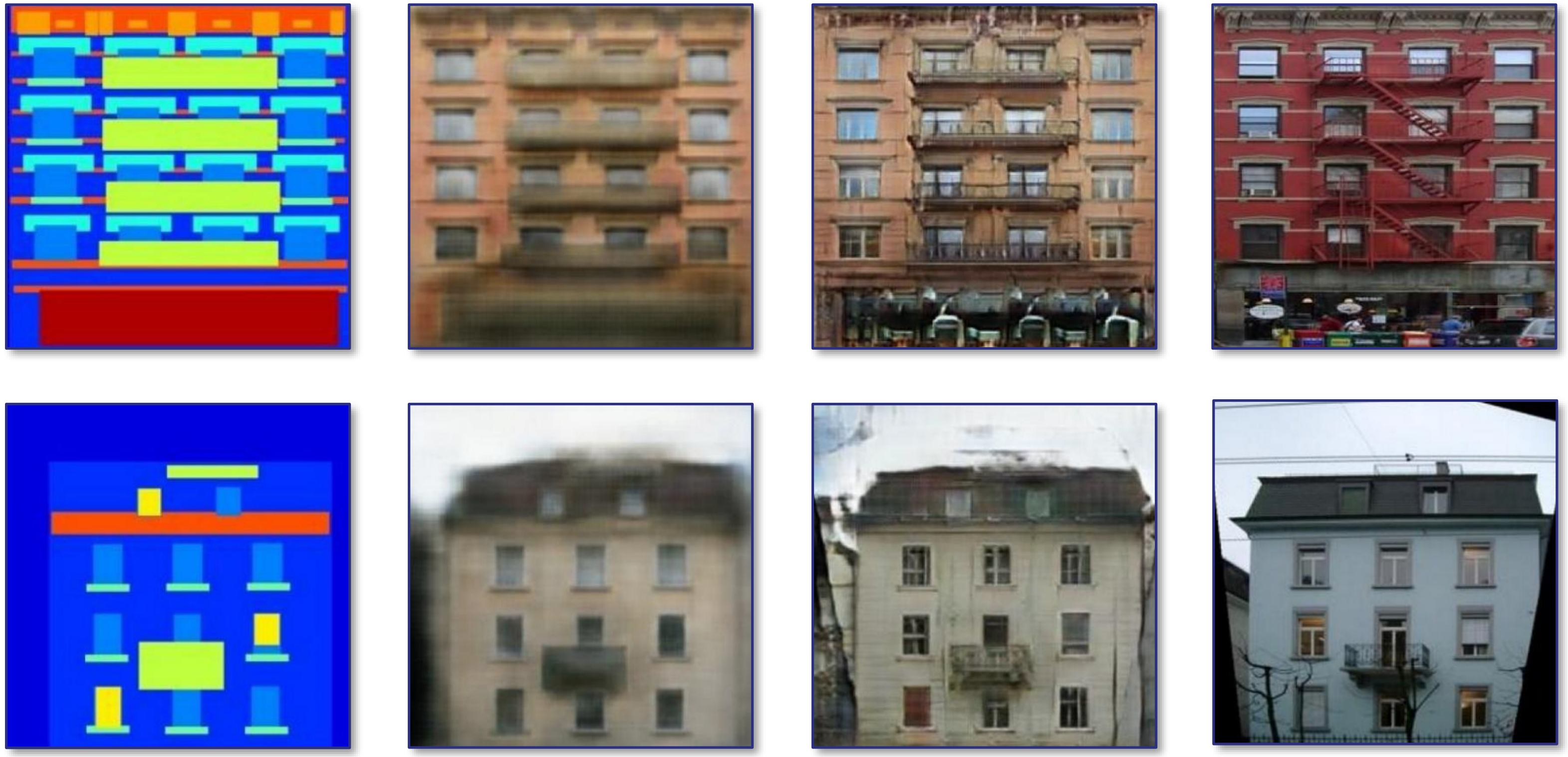
Input | Output of a conventionally trained network | Output of a conditional GAN approach

Application Examples – Bandwidth Extension for Speech Signals

Basic approach:

- Example for a so-called GAN approach:
  - Reconstruction of houses on the basis of rectangles
  - Conventional approach in the second column.
  - GAN approach in the third column.

Source: P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros: ***Image-to-Image Translation with Conditional Adversarial Networks***, CoRR, vol. abs/1611.07004, 2016.

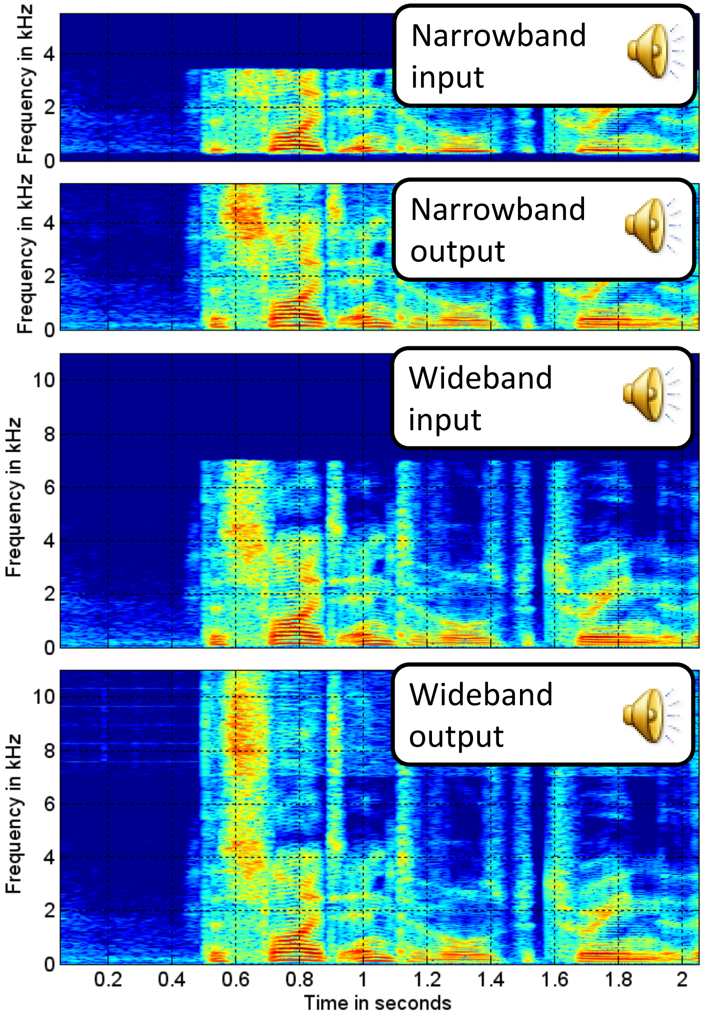


Input | Output of a conventionally trained network | Output of a conditional GAN approach | Original

Application Examples – Bandwidth Extension for Speech Signals

Audio examples:

- ***Narrowband*** connection (still the most often used connection):
  - Bandwidth of about 3.4 to 3.8 kHz
  - Extension to 5.5 to 8 kHz
- ***Wideband*** connection (HD voice, available today):
  - Bandwidth of about 7 kHz
  - Extension to 11 kHz



Application Examples – Prediction of Steering Movements

Problem:

- Contactless measurement of brain, heart, and muscle activities.
- Here first only a prediction of steering movements while driving a car.

Application Examples – Prediction of Steering Movements

Problem:

- Contactless measurement of brain, heart, and muscle activities.
- Here first only a prediction of steering movements while driving a car.

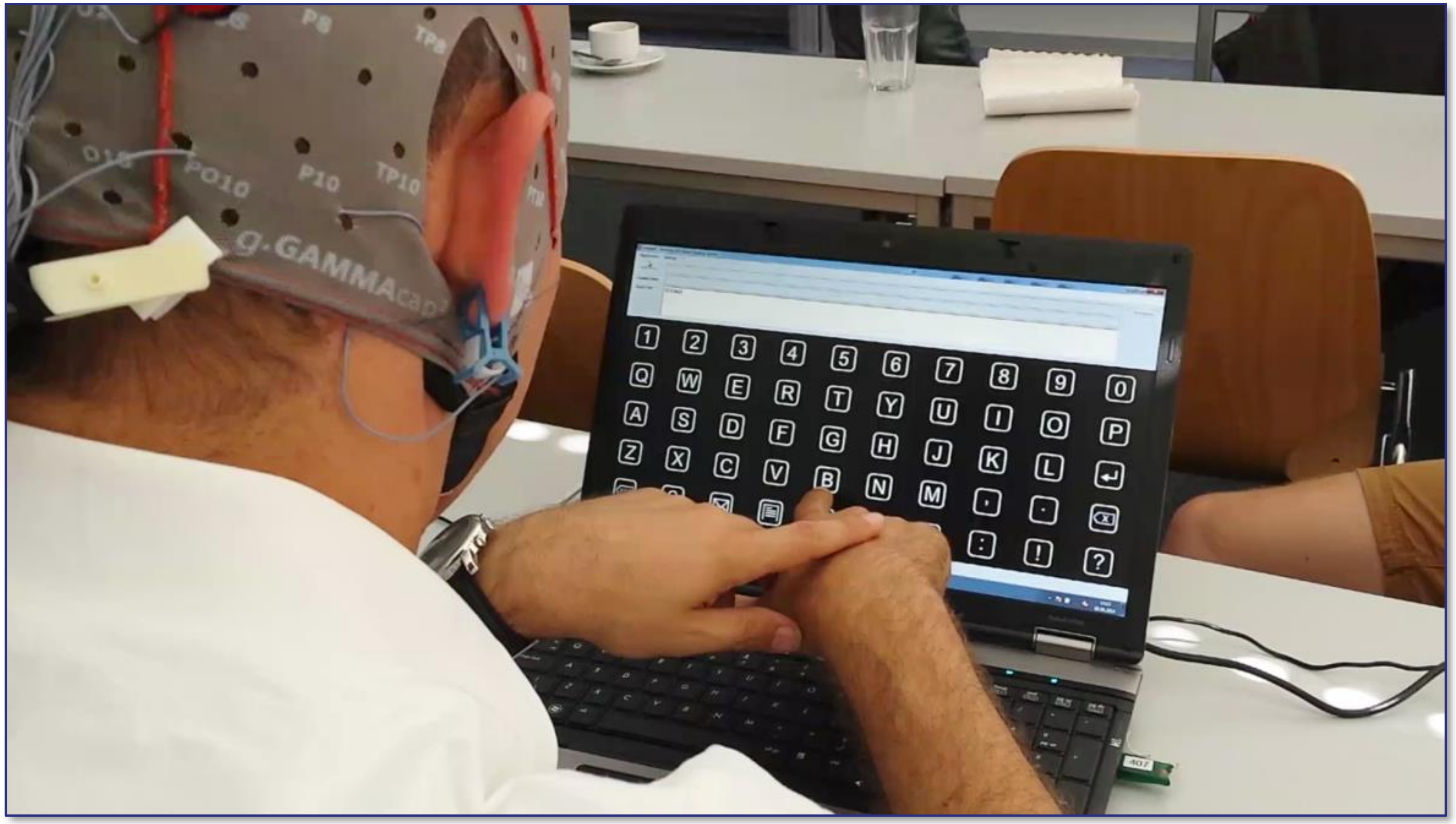
Solution:

- Recording of brain signals (and pattern recognition).

Application Examples – Detection of Sirens and other “Outside Events”

Basic idea:

- Usage of brain-computer interfaces (BCIs).



Application Examples – Prediction of Steering Movements

Problem:

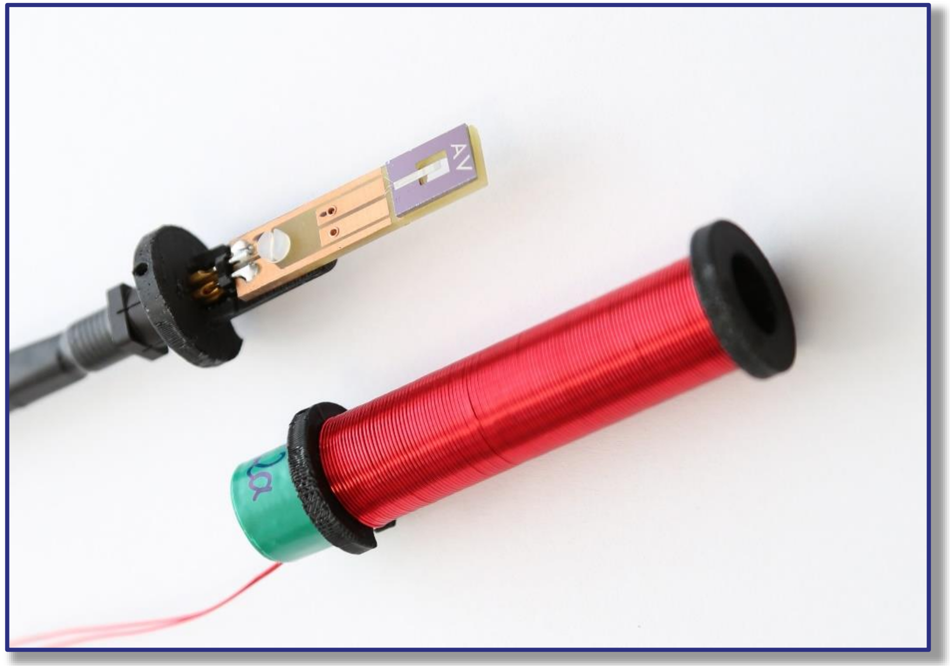
- Contactless measurement of brain, heart, and muscle activities.
- Here first only a prediction of steering movements while driving a car.

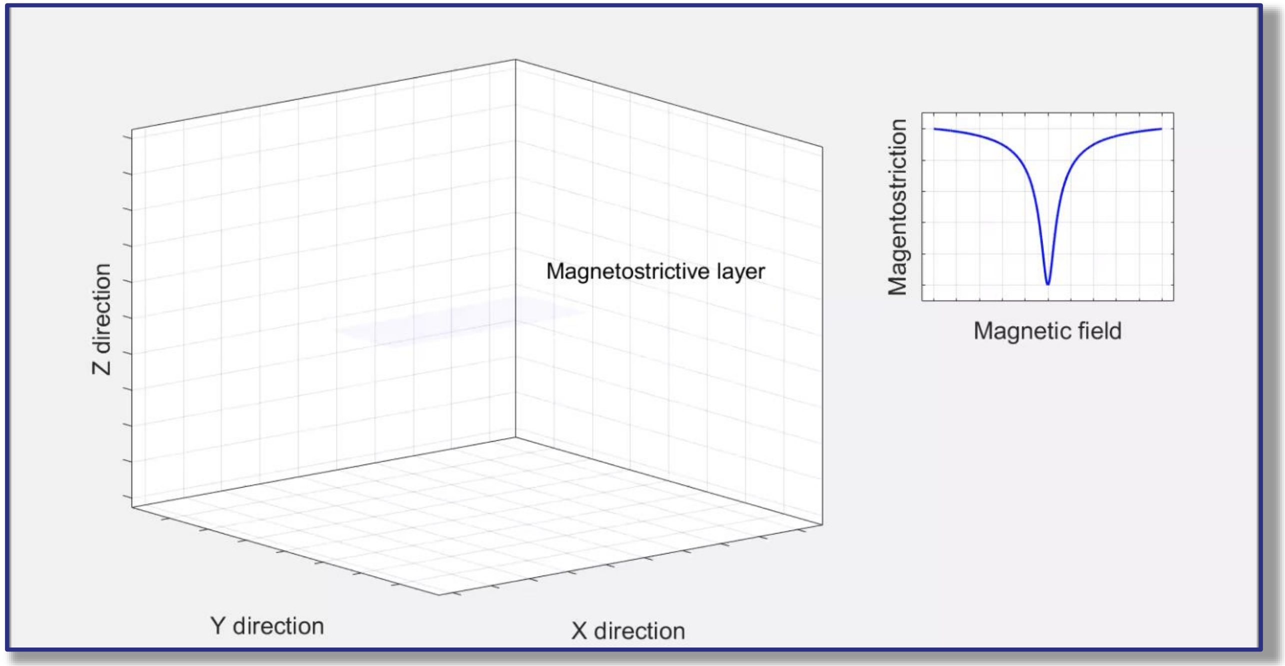
Solution:

- Recording of brain signals (and pattern recognition) …
- … not electrically, but magnetically.

Application Examples – Prediction of Steering Movements

Magnetoelectric sensors:





- Sensors e.g. in cantilever structure (or as SAW sensors)
- Magnetostrictive layer (FeCoSiB)
- Piezo layer (AlN)

- Basic principle of cantilever-based magnetoelectric sensors

Application Examples – Prediction of Steering Movements

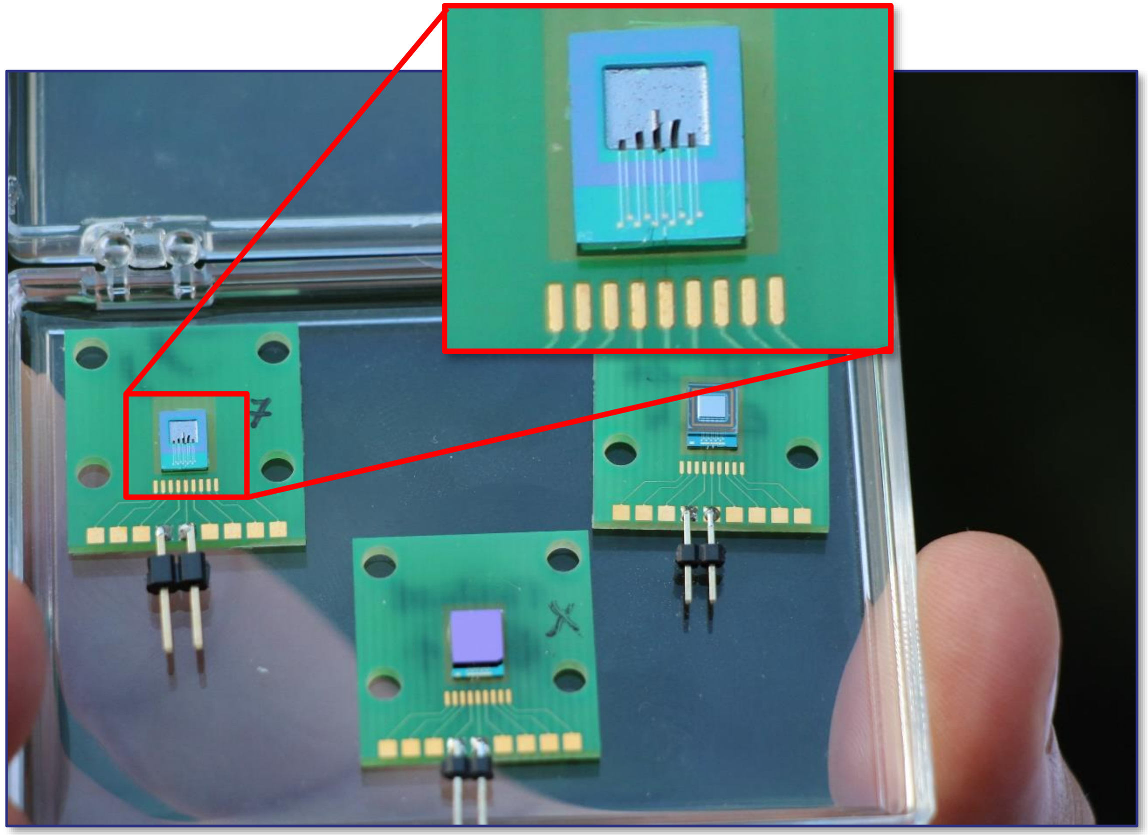
Magnetoelectric sensors:

- Collaborative research center 1261 at our university.
- Extremely sensitive magnetic sensors, that allow for unshielded and non-cooled magnetic sensor systems.
- For medical purposes a limit of detection in the fT regime are required, currently we achieve only pT.
- However, alternative sensor systems are available already right now.

Application Examples – Prediction of Steering Movements

Magnetoelectric sensors:

- Collaborative research center 1261 at our university.
- Extremely sensitive magnetic sensors, that allow for unshielded and non-cooled magnetic sensor systems.
- For medical purposes a limit of detection in the fT regime are required, currently we achieve only pT.
- However, alternative sensor systems are available already right now.
- In cooperation with the ISIT in Itzehoe (also part of the CRC 1261), very small magnetic sensors can be produced.



Summary and Outlook

Summary:

- Speech and audio signal paths in a car
- Basics on pattern recognition
- Contents of the lecture
- Boundary conditions of the lecture (exercises, exam, etc.)
- Notation used in the lecture
- Literature
- Automotive examples (with some “medical touch”)

Next part:

- Cost functions and noise suppression