Presented by

Dr.-Ing. Thomas Netzel

Application of Neural Networks in sound Recognition

Agenda

- **Introduction**
  - Problem: Automatic sound recognition during sleep
  - Acoustic properties of snoring
  - The human sound production process
- **Signal processing**
  - Signal pre-processing
  - Speech analysis and feature detection
- **Sound recognition using neural networks**
  - Topology and characteristics of neural networks
  - Learning method for sound recognition
- **Results**
- **Summary and Conclusion**

Introduction: Problem: Automatic sound recognition during sleep

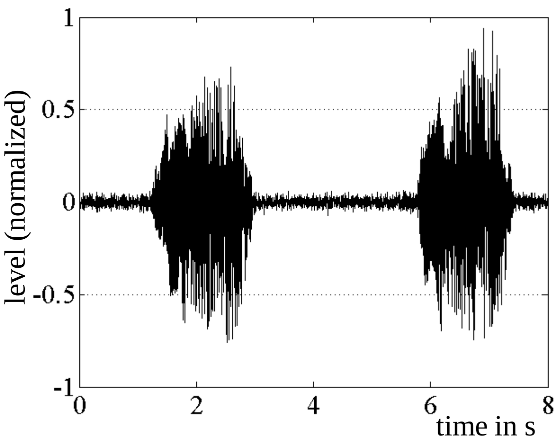
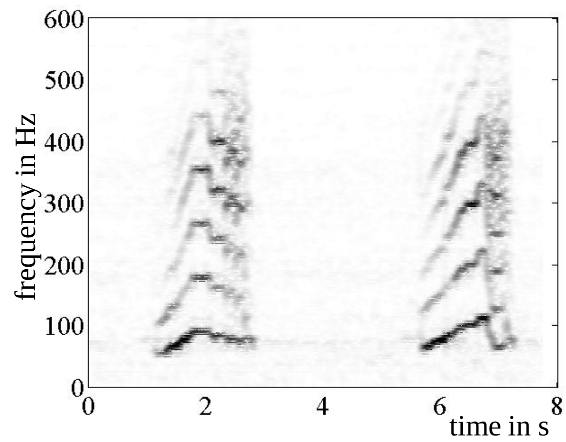
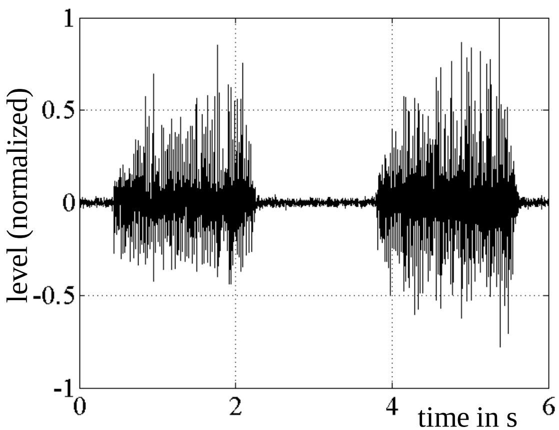
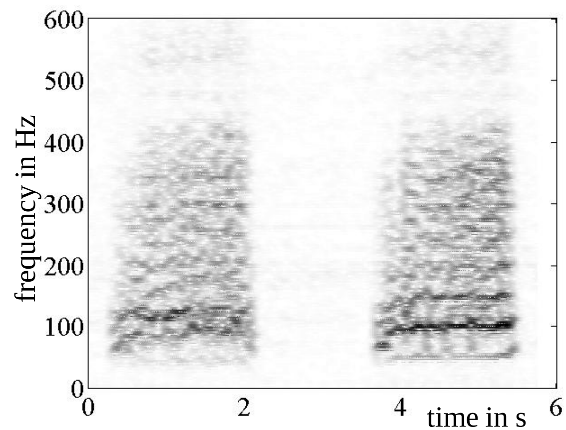
Automatic sound recognition and determination of:

- Snore
- Artefacts (cough, harrumps, noises, swallowing)

Side conditions:

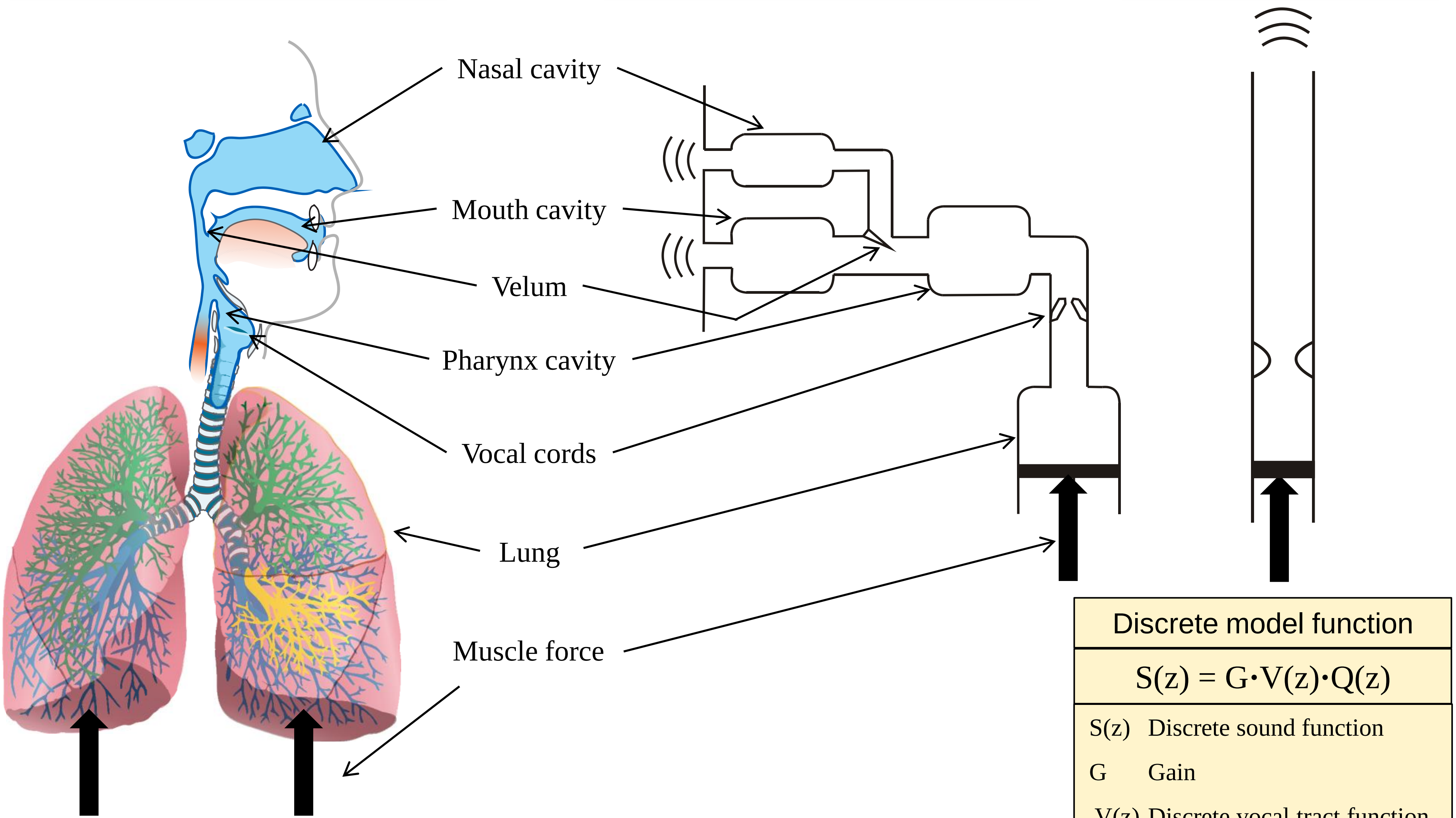
- Sound recognition during respiration
- Application of pressure sensor in respirator
    - Limited sample rate

Introduction: Acoustic properties of snoring

	Time signal	Spectrogram	properties
Voiced snoring			Fundamental frequency Harmonics (formants) Most energy < 500 Hz
Unvoiced snoring			No fundamental frequency No harmonics

Introduction: The human sound production process

Human vocal mechanism | **Model vocal mechanism** | **Simplified model**



Discrete model function

$$S(z) = G \cdot V(z) \cdot Q(z)$$

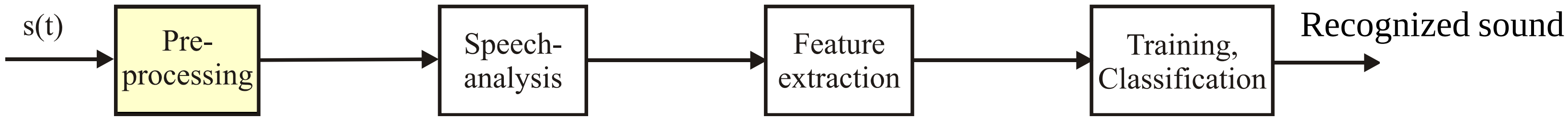
$S(z)$	Discrete sound function
G	Gain
$V(z)$	Discrete vocal tract function
$Q(z)$	Discrete noise function

Agenda

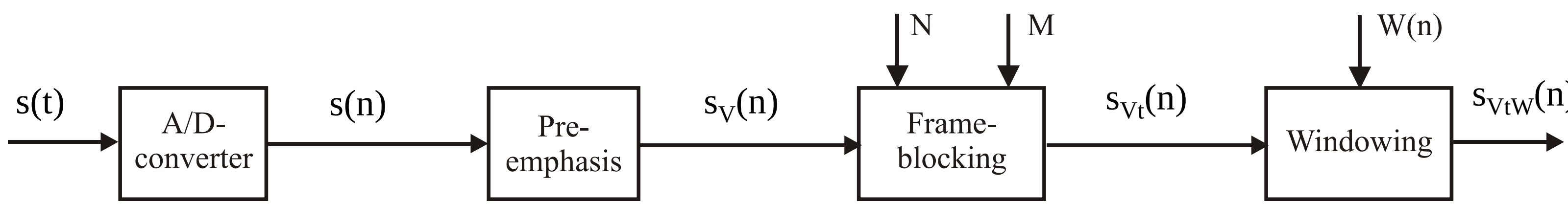
- **Introduction**
  - Problem: Automatic sound recognition during sleep
  - Acoustic properties of snoring
  - The human sound production process
- **Signal processing**
  - Signal pre-processing
  - Speech analysis and feature detection
- **Sound recognition using neural networks**
  - Topology and characteristics of neural networks
  - Learning method for sound recognition
- **Results**
- **Summary and Conclusion**

Signal processing: Signal pre-processing

Sound recognition system



Signal pre-processing



FIR-Filter 1. order

- Spectral flattening
- Reduce influence of limited precision

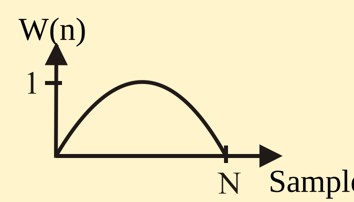
Frame blocking of N samples

- Basis short term analysis

Separation of adjacent frames by M Samples

- Smooth spectral transitions

$s_{VtW}(n) = W(n)\, s_{Vt}(n)$



Minimize leakage effect

Hamming window

$$W(n) = 0{,}54 - 0{,}46 \cos\left(\frac{2\pi n}{N-1}\right)$$

$$0 \le n \le N-1$$

Signal processing: Speech analysis

Sound recognition system

s(t) → Pre-processing → Speech-analysis → Feature extraction → Training, Classification → Recognized sound

Linear Predictive Coding (LPC)

Model

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gq(n)$$

$$S(z) = \left[\frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \right] G\, Q(z)$$

$$S(z) = G\, V(z)\, Q(z)$$

$$V(z) = \frac{1}{A(z)}$$

Prediction

$$\tilde{s}(n) = \sum_{i=1}^{p} a_i s(n-i)$$

Error minimization

$$e = s(n) - \tilde{s}(n)$$

$$E = \sum_{m=1}^{N-1} \left[s(m) - \sum_{i=1}^{p} a_i\, s(m-i) \right]^2 \rightarrow \min$$

$$\frac{\partial E}{\partial a_i} = 0 \qquad i = 1, 2, \cdots, p$$

$$\boxed{\underline{\Phi}_{yy} \cdot \underline{a} = \underline{\varphi}_{yy}}$$

Predictor coefficients a_i per short term analysis window

Signal processing: Feature detection

Sound recognition system

s(t) → Pre-processing → Speech-analysis → Feature extraction → Training, Classification → Recognized sound

$$s_{norm}(n) = v(n) * q(n)$$

$$F^{-1}\left\{\ln\left[S_{Norm}\left(e^{j\omega}\right)\right]\right\} = F^{-1}\left\{\ln\left[V\left(e^{j\omega}\right)\right]\right\} + F^{-1}\left\{\ln\left[Q\left(e^{j\omega}\right)\right]\right\}$$

Cepstrum c_{xx}

$$c_{xx}(q) = F^{-1}\left[\ln\left|S\left(e^{j\omega}\right)\right|^2\right]$$

$$c_{xx}(q) = \begin{cases} \ln(G), & q = 0 \\ a_q + \sum_{k=1}^{q-1}\left(\frac{k}{q}\right) a_{q-k}\, c_{xx}(k), & q = 1, 2, \cdots, p \end{cases}$$

Liftered cepstrum

$$c_{xx}^{L}(q) = f^{L}(q)\, c_{xx}(q)$$

$$f^{L}(q) = \begin{cases} 1 + \frac{p}{2}\sin\left(\frac{\pi q}{p}\right), & q = 0, 1 \cdots, p \\ 0, & q > p \end{cases}$$

Cepstrum time derivative

$$\frac{\partial c_{xx}^{L}(q,t)}{\partial t} = \dot{c}_{xx}^{L}(q,t)$$

Feature vector

$$\underline{\rho} = \begin{bmatrix} G \\ c_{xx}^{L}(1,t) \\ \vdots \\ c_{xx}^{L}(p,t) \\ \dot{c}_{xx}^{L}(1,t) \\ \vdots \\ \dot{c}_{xx}^{L}(p,t) \end{bmatrix}$$

Agenda

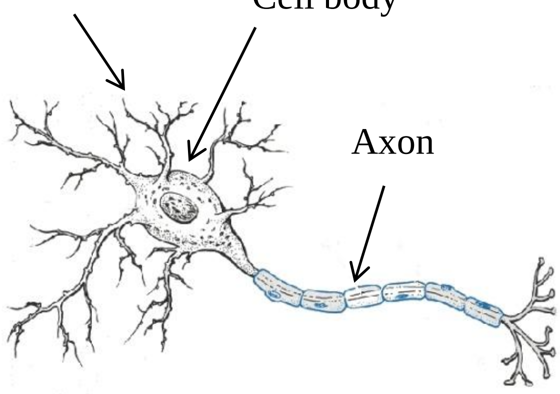
- **Introduction**
  - Problem: Automatic sound recognition during sleep
  - Acoustic properties of snoring
  - The human sound production process
- **Signal processing**
  - Signal pre-processing
  - Speech analysis and feature detection
- **Sound recognition using neural networks**
  - Topology and characteristics of neural networks
  - Learning method for sound recognition
- **Results**
- **Summary and Conclusion**

Sound recognition using neural networks
Topology of neural networks

Sound recognition system

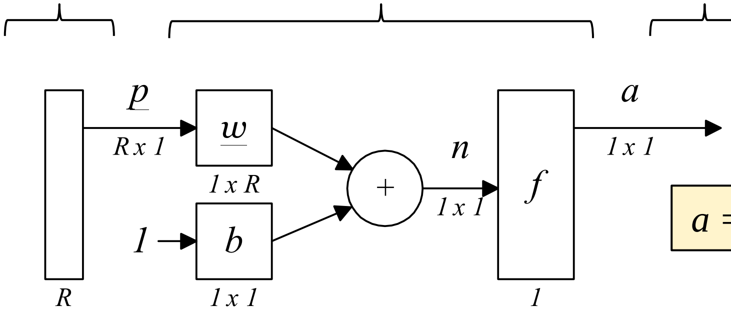
Topology of neural networks

Human neuron



- human brain: 10^{11} neurons
- 10^4 connections (synapses) per neuron

Neuron model

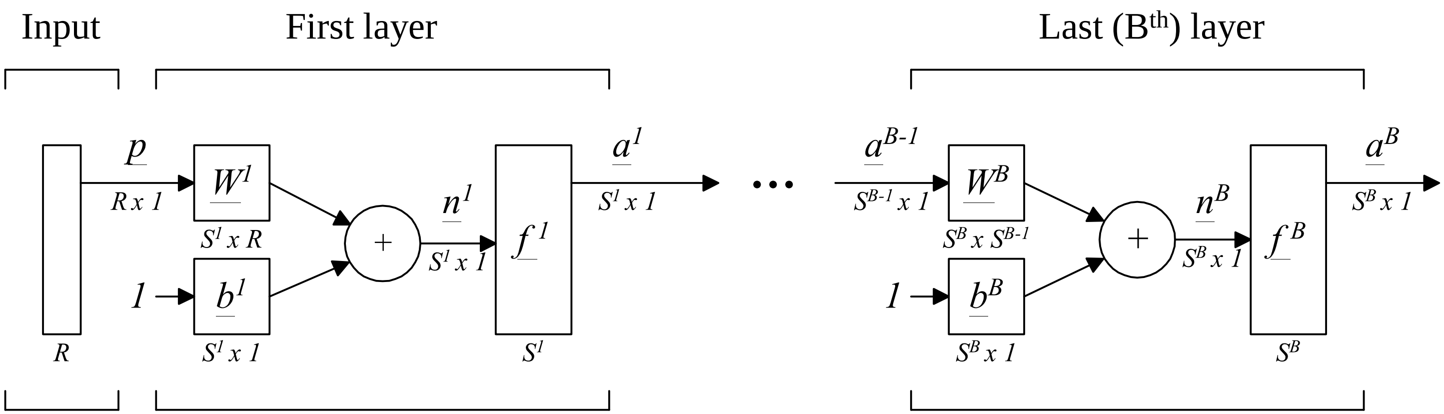


$a = f(\underline{w}\,\underline{p} + b)$

R	Inputs (features)	b	Bias
$\underline{p}$	Input vector	f	Transfer function
$\underline{w}$	Weight vector	a	Output

Sound recognition using neural networks
Topology and characteristics of neural networks (feed forward)

Neural network with B layers



R	Inputs (features)
$\underline{p}$	Input vector
$\underline{w}$	Weight vector
b	Bias
n	Input of $\underline{f}$
f	Transfer function
a	Output

$$\underline{a}^1 = \underline{f}^1(\underline{W}^1\underline{p} + \underline{b}^1)$$

$$\underline{a}^B = \underline{f}^B(\underline{W}^B\underline{a}^{B-1} + \underline{b}^B)$$

Output	$\underline{a}^B = \underline{f}^B \{ \underline{W}^B \underline{f}^{B-1} \{ \cdots [\underline{W}^2 \underline{f}^1 (\underline{W}^1 \underline{p} + \underline{b}^1) + \underline{b}^2] \cdots \} + \underline{b}^B \}$

Charakteristics of neural networks

- Approximation of non-linear functions
- Robust classification of unknown signals
- Easy train process
- Combination of different knowledge sources
  - Acoustic, lexical, pragmatic features, ...

- Pattern recognition and completion
- Data based modelling and prediction
- Contoller design and optimisation
- Sensor fusion
- Automatic sound recognition

Sound recognition using neural networks

Learning method for sound recognition

Moment gradient method with adaptive learning rate

NN Output	$\underline{a}^{B} = \underline{f}^{B}\{\underline{W}^{B}\underline{f}^{B-1}\{\cdots[\underline{W}^{2}\underline{f}^{1}(\underline{W}^{1}\underline{p}+\underline{b}^{1})+\underline{b}^{2}]\cdots\}+\underline{b}^{B}\}$

Trained

Gradient method

- Training with known Input/Output Data
- Calculation of error function

$$E^{k} = \left[\underline{a}^{B}_{Soll}(k) - \underline{a}^{B}(k)\right]^{T}\left[\underline{a}^{B}_{Soll}(k) - \underline{a}^{B}(k)\right]$$

Weight update

- Weight update based on:
  - Error function
  - Applied momentum ζ and learning rate α

$$\underline{W}^{r}(k) = \underline{W}^{r}(k-1) + \Delta\underline{W}^{r}(k) \qquad \underline{b}^{r}(k) = \underline{b}^{r}(k-1) + \Delta\underline{b}^{r}(k)$$

Learning rate adaptation

- Learning rate adaptation after training loop with all training data based on perfomance Index

$$EE^{k} = \sum_{j=1}^{Q_T} E^{k}\left[\underline{a}^{B}_{Soll,j}(k), \underline{a}^{B}(k)\right]^{T}\left[\underline{a}^{B}_{Soll,j}(k) - \underline{a}^{B}(k)\right]$$

Stop Learning

- Stop training with minimum performance index EE

$\underline{p}$	Input vector
$\underline{W}$	Weight matrix
$\underline{b}$	Bias vector
$\underline{f}$	Transfer functions
$\underline{a}$	Output vector
ζ	Momentum
α	Learning rate
EE	Perform. Index
Q^{T}	Number of Learningdata

Agenda

- **Introduction**
  - Problem: Automatic sound recognition during sleep
  - Acoustic properties of snoring
  - The human sound production process
- **Signal processing**
  - Signal pre-processing
  - Speech analysis and feature detection
- **Sound recognition using neural networks**
  - Topology and characteristics of neural networks
  - Learning method for sound recognition
- **Results**
- **Summary and Conclusion**

Results: Methods

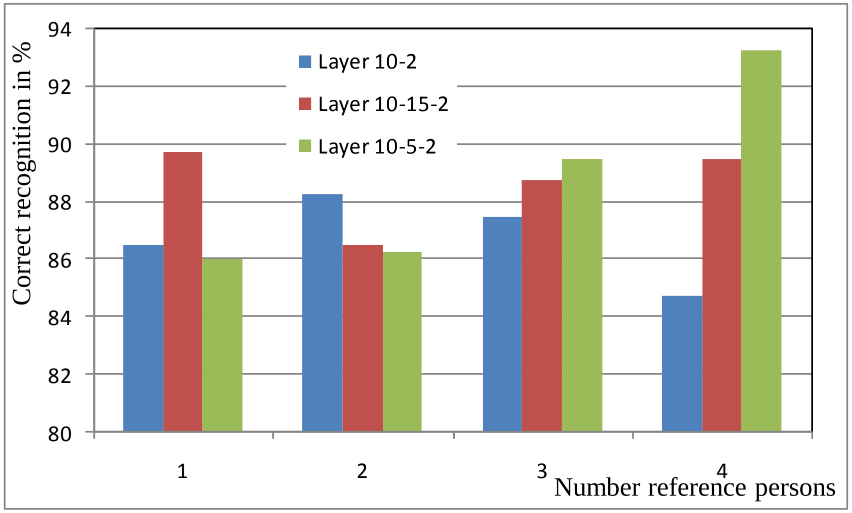
Feature vector	$\underline{\rho} = \begin{bmatrix} G & c_{xx}^{L}(1,t) & \cdots & c_{xx}^{L}(6,t) & \dot{c}_{xx}^{L}(1,t) & \cdots & \dot{c}_{xx}^{L}(3,t) \end{bmatrix}^{T}$

Neural network	
Topology	➢ 3 networks with 2 and 3 layers: ➢ 10-2, 10-15-2, 10-5-2 ➢ 10 input, 2 outputs ➢ Transfer functions: Log-Sigmoid $f(n) = \dfrac{1}{1+e^{-n}}$
Learning	➢ 20 snores $\underline{a}_{Soll,snore}^{B} = \begin{bmatrix} 1 & 0 \end{bmatrix}^{T}$ ➢ 20 artefact $\underline{a}_{Soll,artefact}^{B} = \begin{bmatrix} 0 & 1 \end{bmatrix}^{T}$ ➢ 1-4 Reference persons $\alpha(0) = 0{,}01$ $\zeta = 0{,}9$ $\eta = 1{,}04$ $l = 5$
Test	➢ 20 snores ➢ 160 artefacts

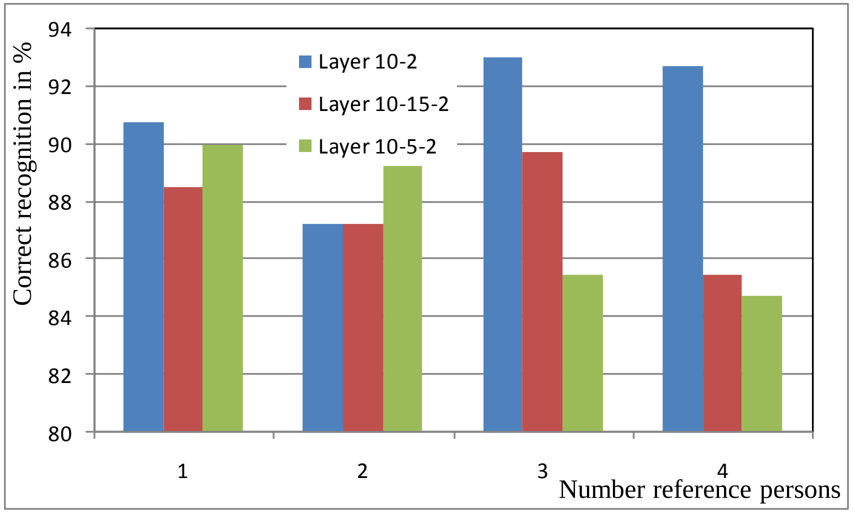
f	Transfer functions
n	Input vector of f
α	Learning rate
ζ	Momentum
η	Learning threshold
l	Stopp threshold

Results: Correct recognition

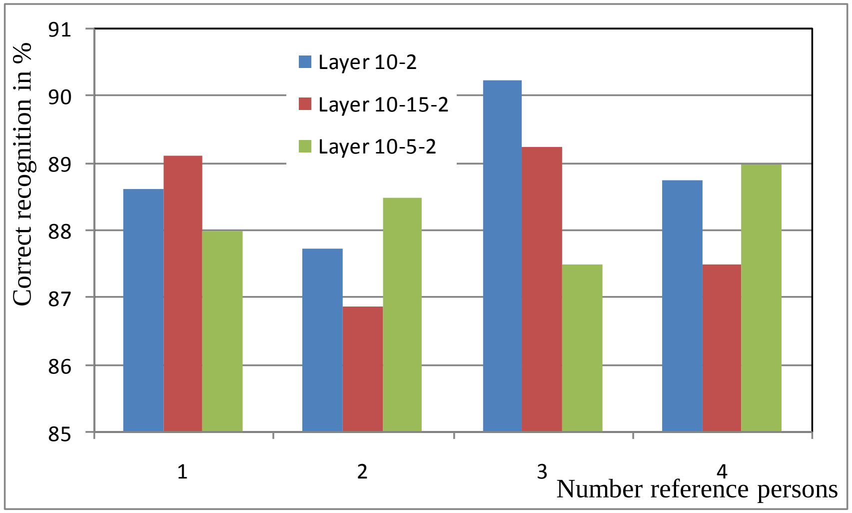
Snoring



Artefacts



Mixed sound



Snoring

- Increased detection rate ~ number reference person for 3 layer NN

Artefacts

- Decreased detection rate ~ number reference person for 3 layer NN
- Better detection with 2 layer NN

Agenda

- **Introduction**
  - Problem: Automatic sound recognition during sleep
  - Acoustic properties of snoring
  - The human sound production process
- **Signal processing**
  - Signal pre-processing
  - Speech analysis and feature detection
- **Sound recognition using neural networks**
  - Topology and characteristics of neural networks
  - Learning method for sound recognition
- **Results**
- **Summary and Conclusion**

Summary

- **Introduction in human sound signal processing**
  - Modelling
  - Signal pre-processing
  - Speech analysis and feature detection
- **Sound recognition using neural networks**
  - Topology and characteristics of neural networks
  - Learning method for sound recognition
- **Example: Automatic sound recognition during sleep**
  - Topology, learning
  - Classification results

Conclusion

- **Computer power consuming recognition task**
  - Realtime feature detection + classification (parallel computing)
- **Neural networks are reliable in sound recognition**
  - Easy implementation and training
  - Robust classification of unknown sound signals
  - Difficult optimization

Broader application of neural networks with falling computer prizes is expected.