DORSA Song classification

Friedhelm Schwenker [1]

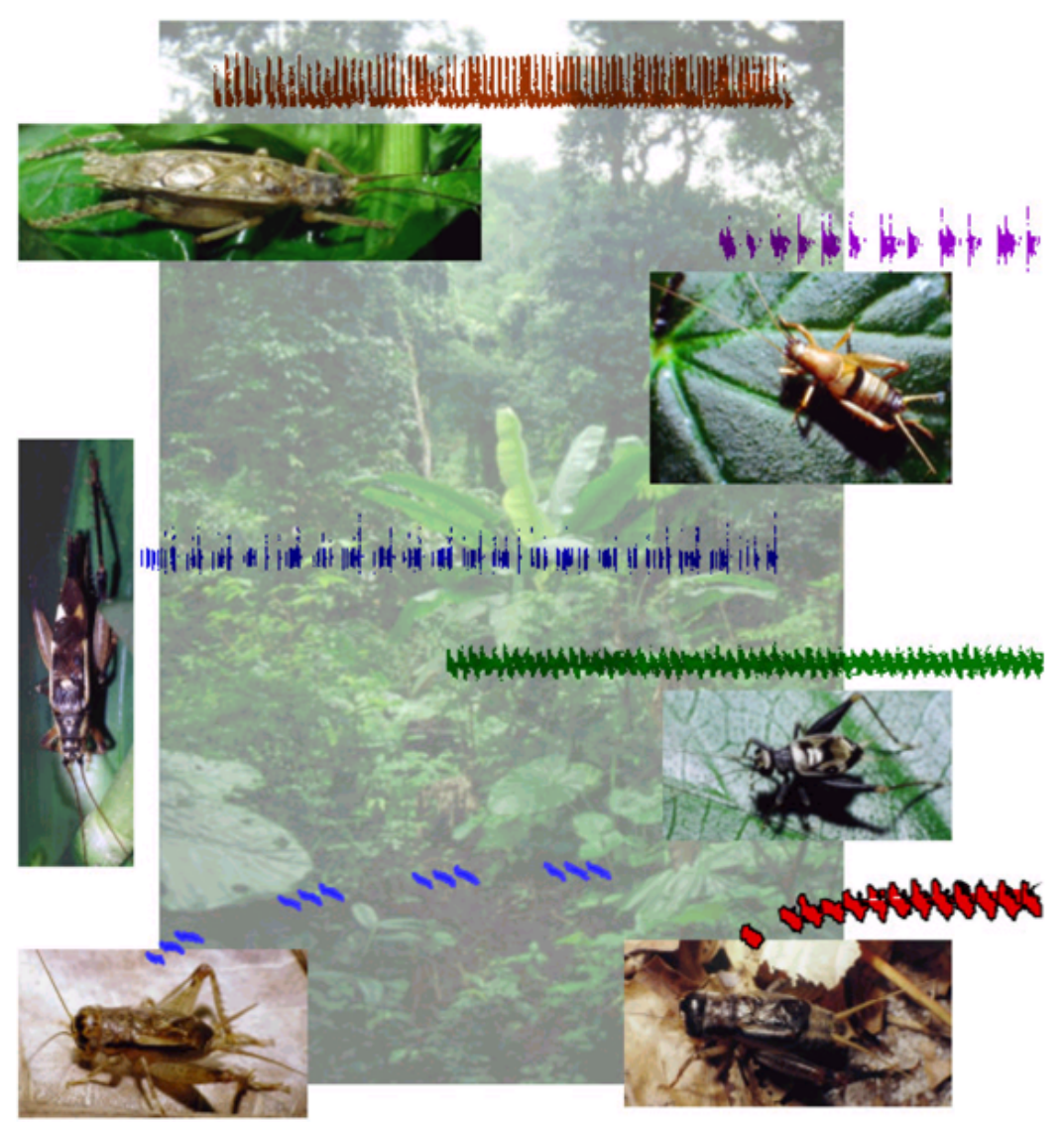
[1]Department of Neural Information Processing
University of Ulm

First European Workshop on
Animal Sound Research and Libraries

Statistical Pattern Recognition Problems

- Recognition of visual objects from camera images (OCR, faces recognition)
- Medical diagnosis and bio informatics
- Speaker recognition (Who is speaking?)
- Speech recognition/understanding (What is the content?)
- Recognition of emotions in speech (How?)
- Language recognition (What language?)
- ....
- Bio-acoustic pattern recognition

Bio-acoustic pattern recognition



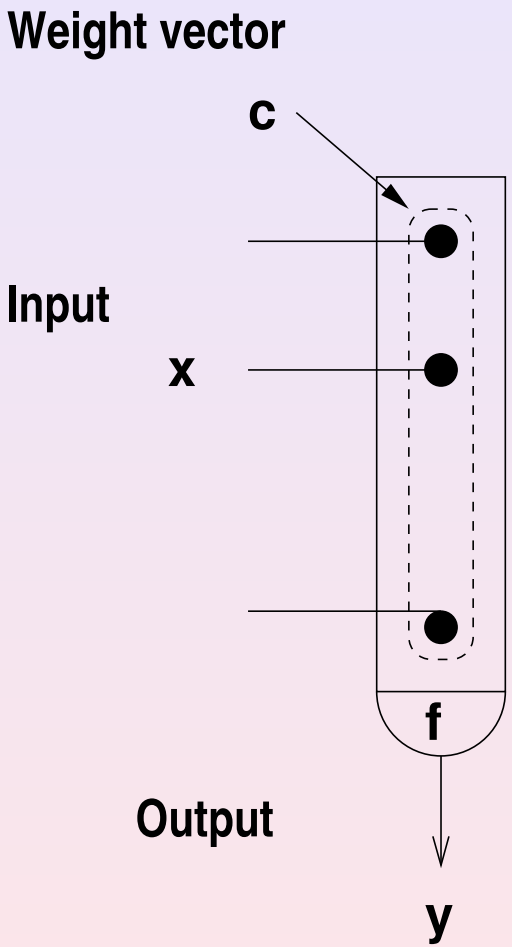
Outline

- **Artificial neural networks**
- Sensor-fusion in neural networks
- Multiple classifier systems
- Recognition of bio-acoustic time series

Neural network properties

	Von Neumann Computer	Biologial neural net
Processor	complex	simple
	high speed	low speed
	1 or a few	large number
Computing	centralized	distributed
	sequential	parallel
	by programs	**by learning from data**
Memory	localized	distribute
	addressable by keys	addressable by content
	not faulttolerant	faulttolerant

Neural Models



- **Linear neuron**

$$y = \langle x, c \rangle = \sum_{i=1}^{n} x_i c_i$$

- **Threshold neuron**

$$y = \begin{cases} 1 & \langle x, c \rangle \geq \theta \\ 0 & \text{sonst} \end{cases}$$

- **Sigmoidal neuron**

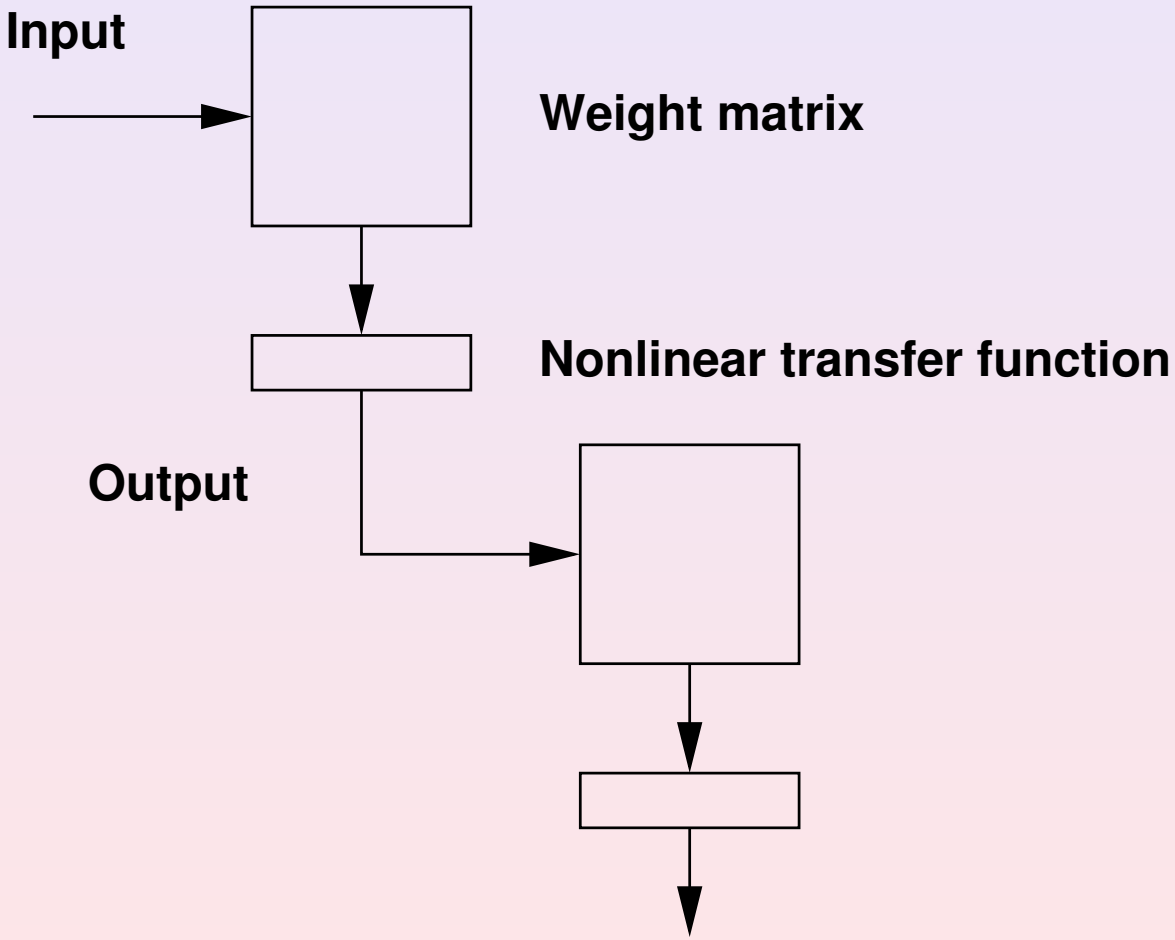
$$y = f(\langle x, c \rangle - \theta), \quad f(s) = \frac{1}{1 + \exp(-\beta s)}$$

- **RBF neuron**

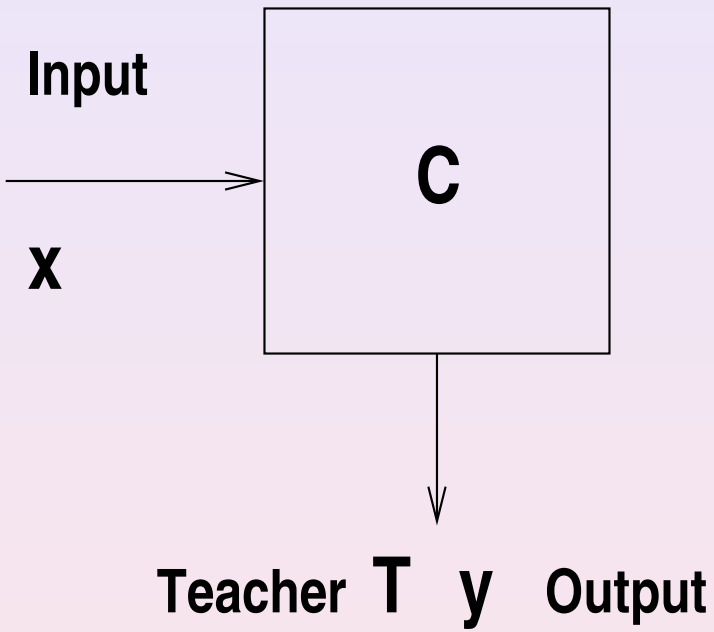
$$y = f(\|x - c\|_2), \quad f(r) = \exp(-\frac{r^2}{2\sigma^2})$$

Layered Networks

Layered neural networks (single or multilayer perceptrons, radial basis function networks) are widely used in pattern recognition and regression applications.

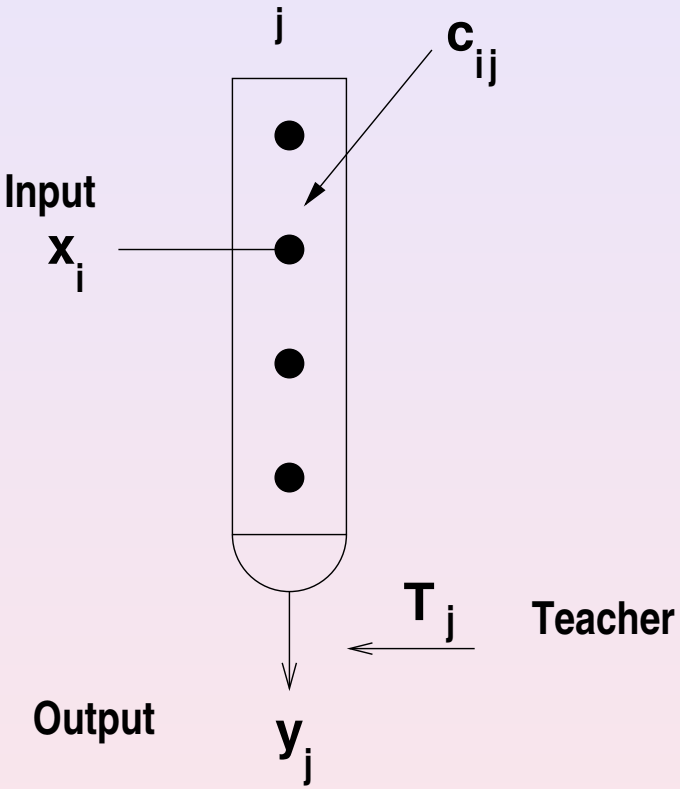


Learning in artificial neural nets



- Mapping $F_C : X \rightarrow Y$, connectivity matrix C **learnt by examples**
- Data $x \in X$ or $(x, T) \in X \times Y$
- Different types of target function $E(C)$.
- Optimising $E(C)$ leads to learning rules for C.

Supervised learning



- Output y_j, teaching signal T_j.
  c_{ij} adapted, such that $y_j \approx T_j$.
- Example: *Delta-rule*

$$\Delta c_{ij} \sim x_i \left(T_j - y_j \right)$$

- Delta-rule minimises

$$E(c) = \| T - y \|^2$$

Model complexity and training data

- Artificial neural networks can solve complex tasks, e.g. high-dimensional input (many input variables), high-dimensional output (multi-class problems).
- Large networks (with many parameters) are needed to achieve good approximations.
- Size of the training set grows with the number of free parameters

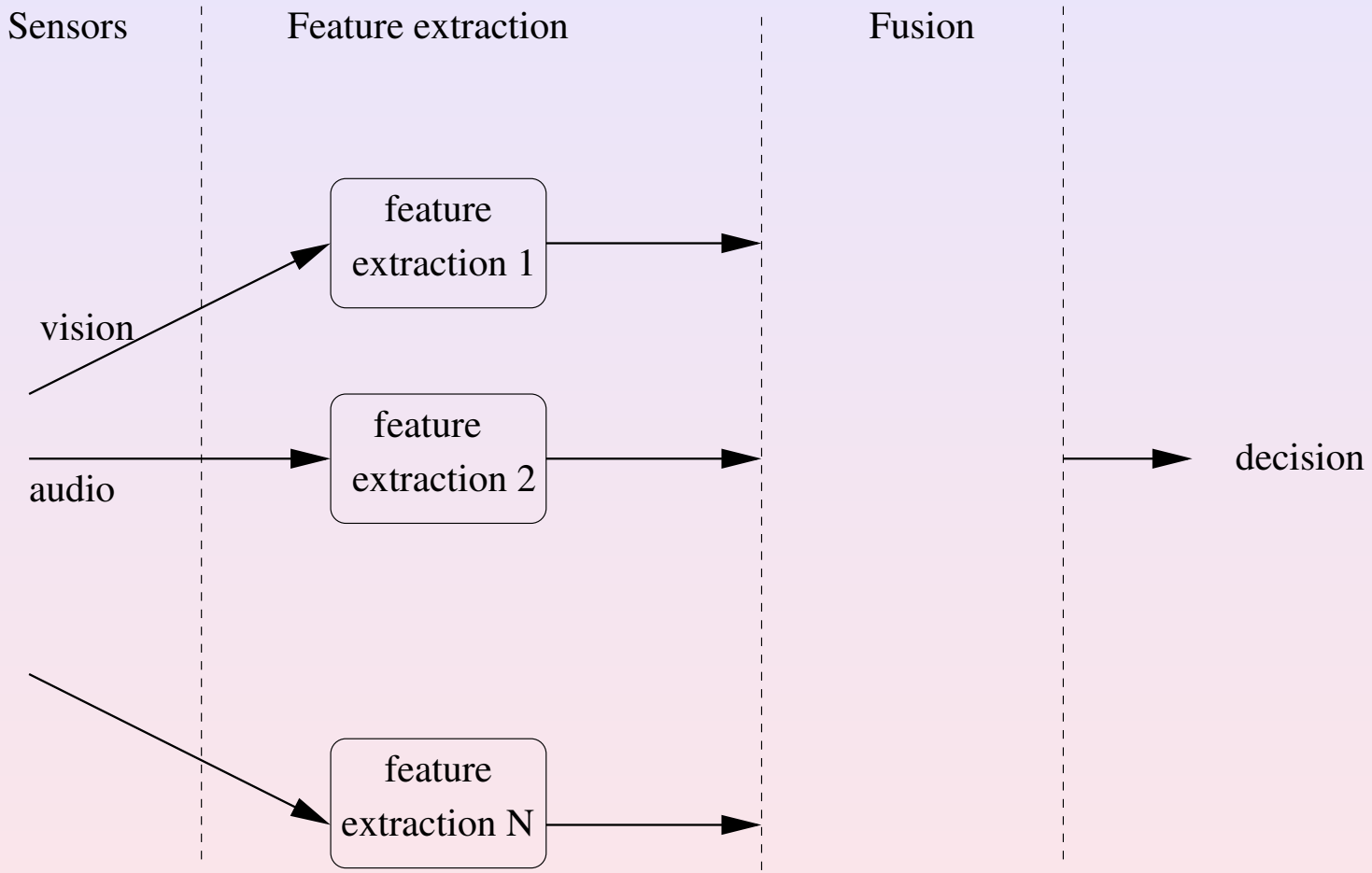
$$M_{\epsilon,\delta} = O\left(\frac{\text{VCdim}}{\epsilon} \log \frac{1}{\epsilon} + \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right)\right), \quad \text{VCdim} = O(W \log(K))$$

  W (K) number of weights (units), ϵ error, $1 - \delta$ confidence.
- Typically the training data set is to small.
- Possible approach: **Decomposition of the learning task using sensor fusion**.

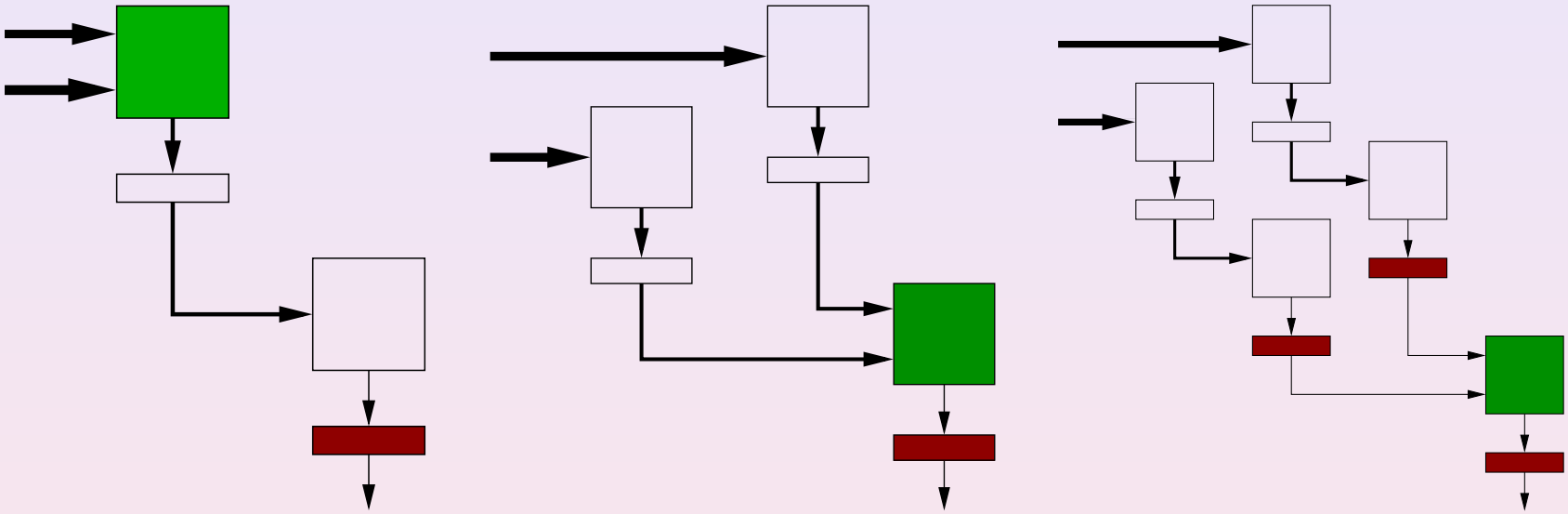
Sensor-fusion in biological neural networks

- Localisation of an object: Stereoscopic vision (to calculate of depth information)
- Recognition of speech in a noisy environment (cocktail party): Audio-visual speech recognition
- Grasping an object: (stereo) vision and tactiation
- Object falling on one's foot: (stereo) vision, audition, and tactition.

Fusion of sensor information



Early fusion • Mid-level fusion • Late fusion



Audio-Visual Speech Recognition

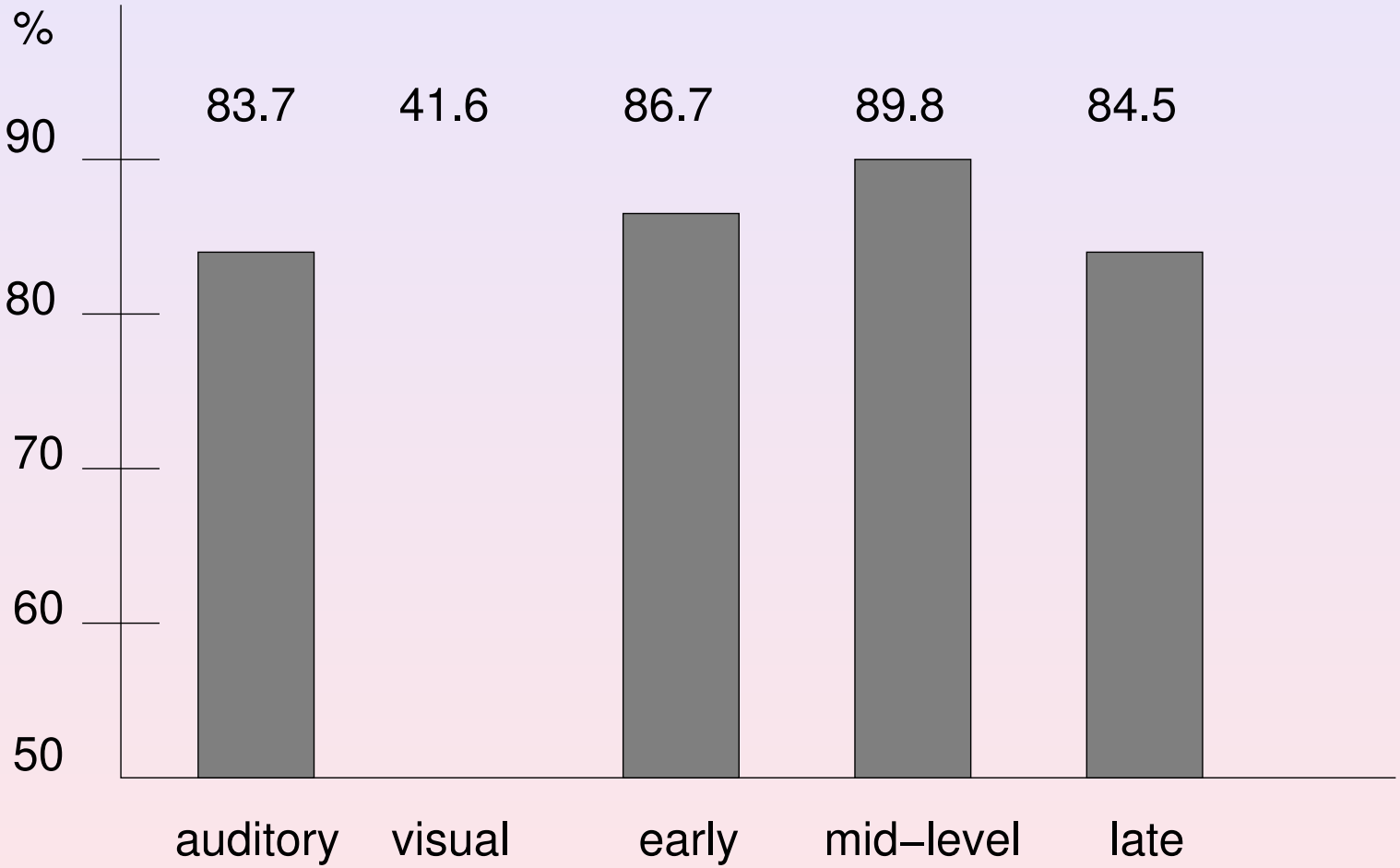
Visual features are invariant against acoustic noise perturbation.

- Combination of visual and audio features can improve recognition performance:
  - Visual input important
    Example: ***/m/*** and ***/n/***
  - Auditory input important
    Beispiel: ***/p/*** und ***/b/***
- McGurk-MacDonald-effect (an auditory illusion)

Input		recognized
visual	auditory	
ga	ba	da
ba	pa	ta

Results on AVSR

Classification rates (word level) for MLPs.

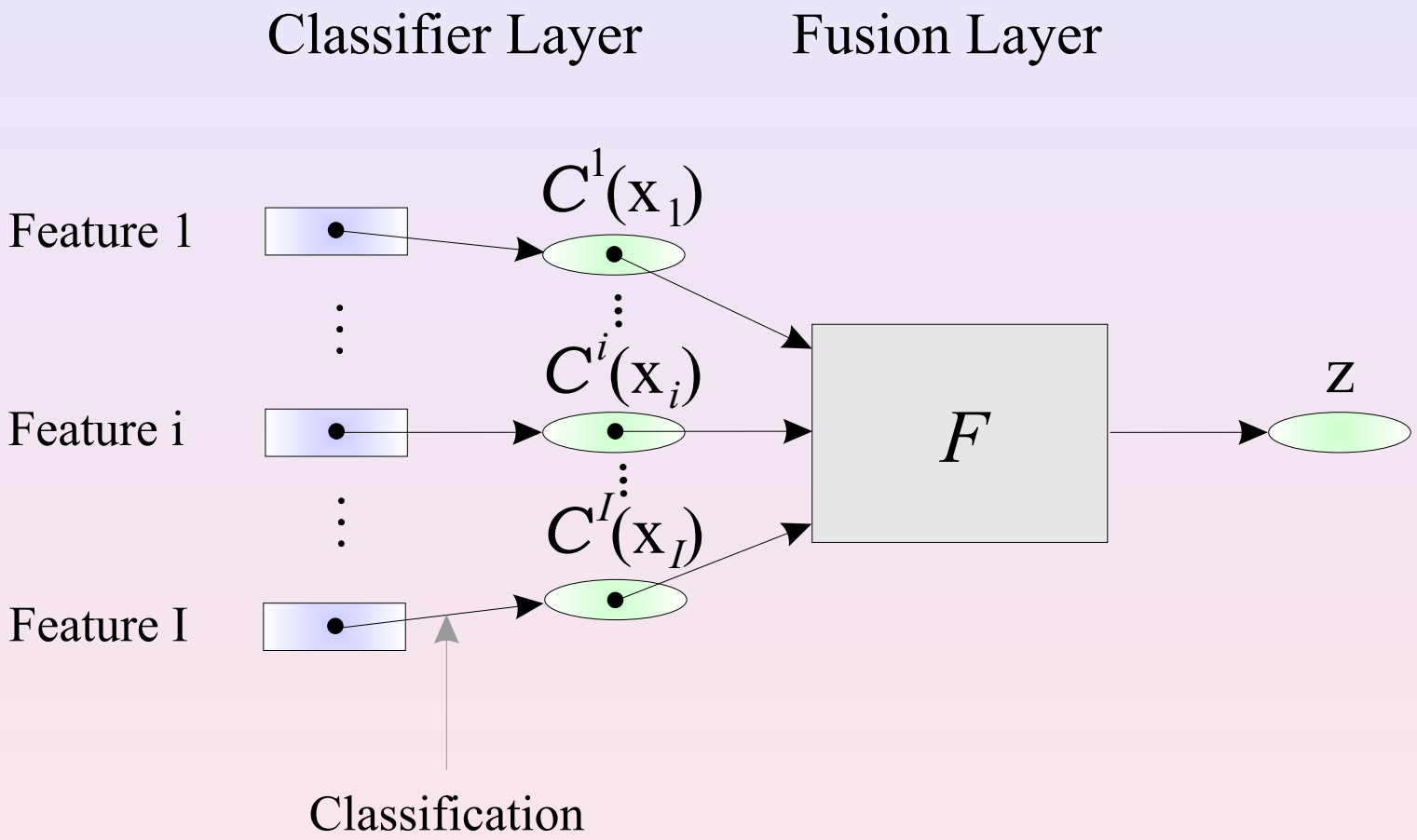


Types of MCS

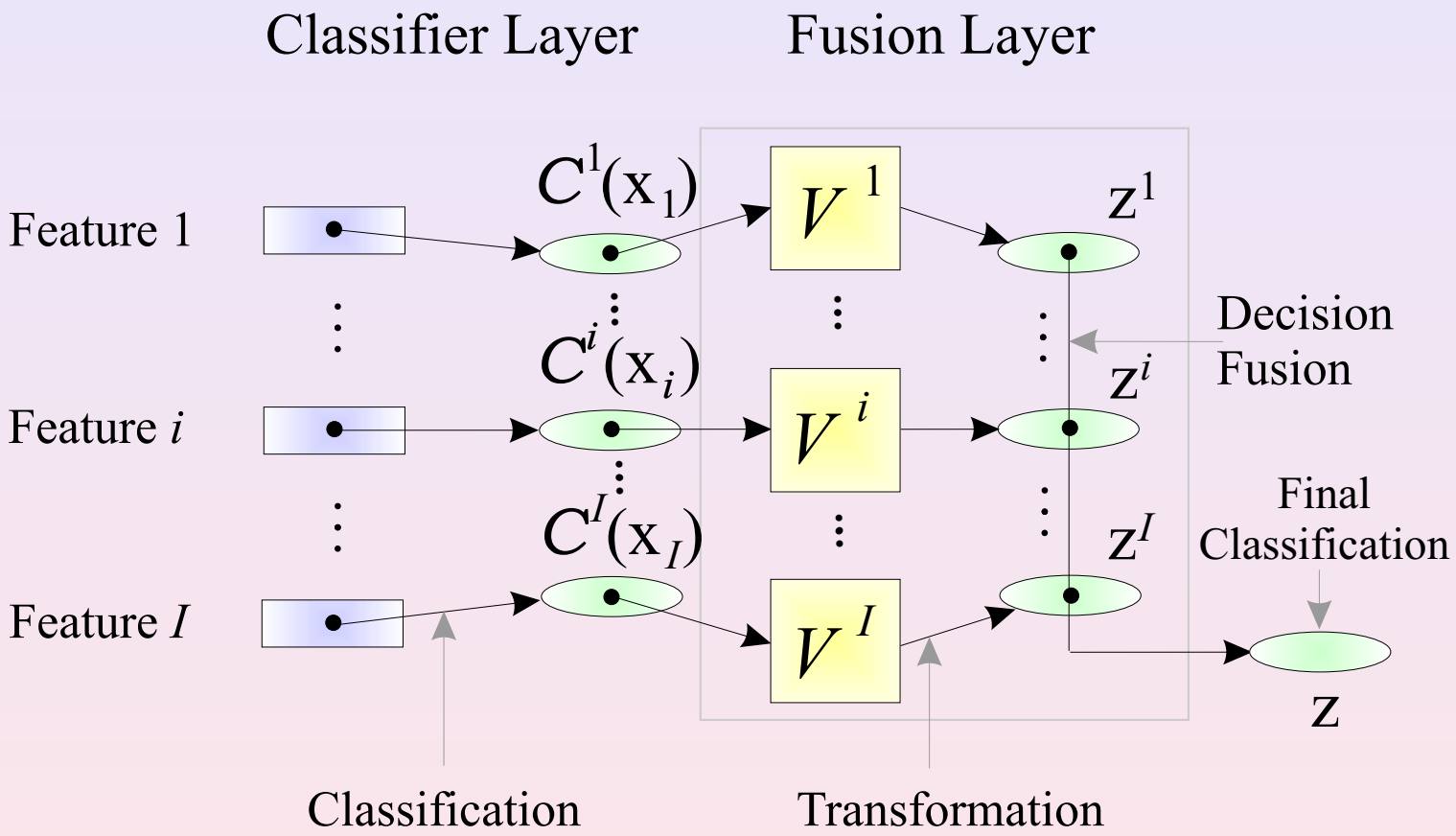
Multiple classifier system (MCS) is an **ensemble** or **sensor fusion network** of many (more than one) classifiers together with a **combination/fusion** mapping (**late fusion**).
Examples:

- Different classifier architectures (ANNs, Decision Trees, SVMs,...)
- Different data sets (Bagging, Boosting,...)
- **Classification and fusion of different feature spaces**

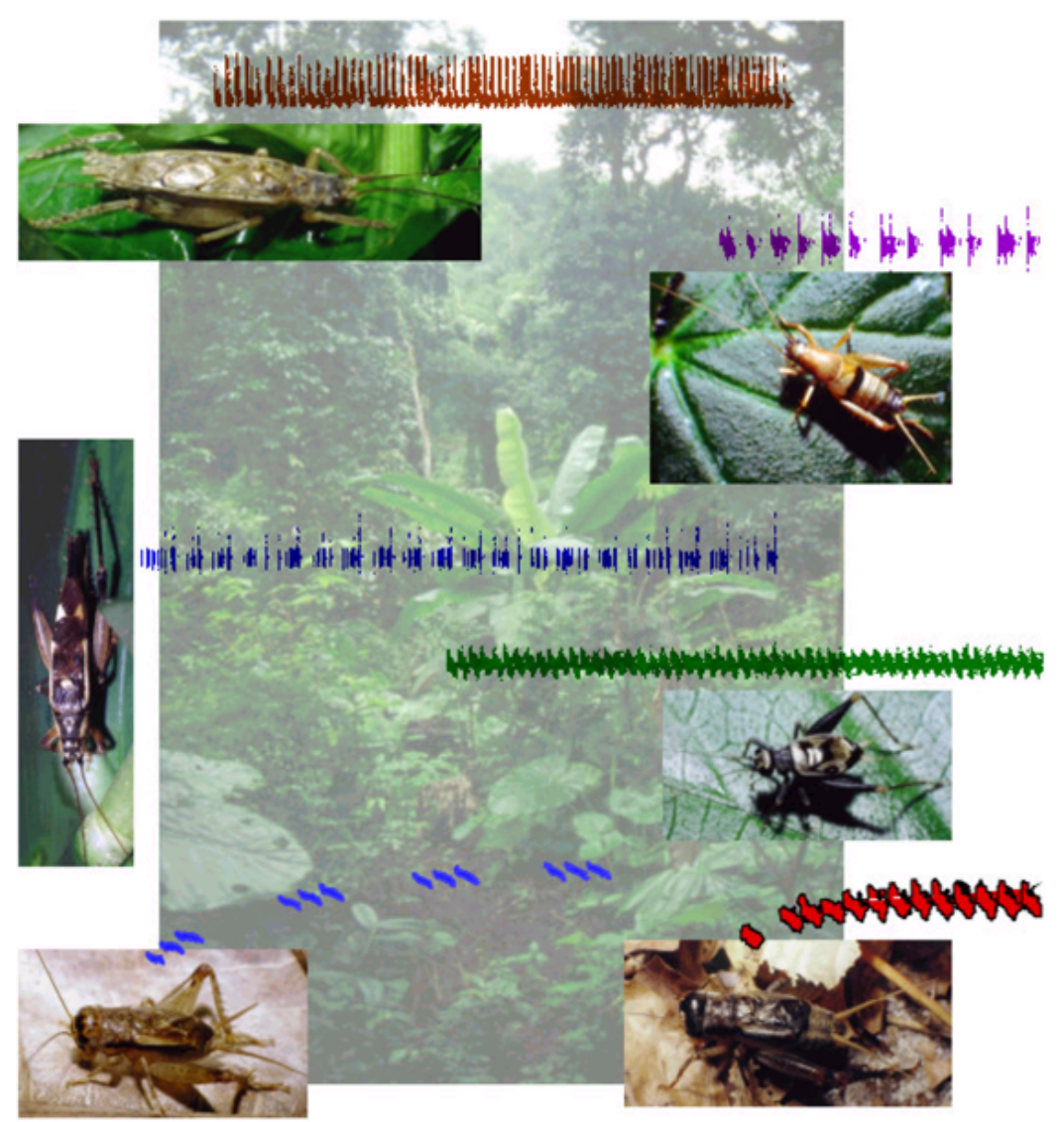
MCS architecture



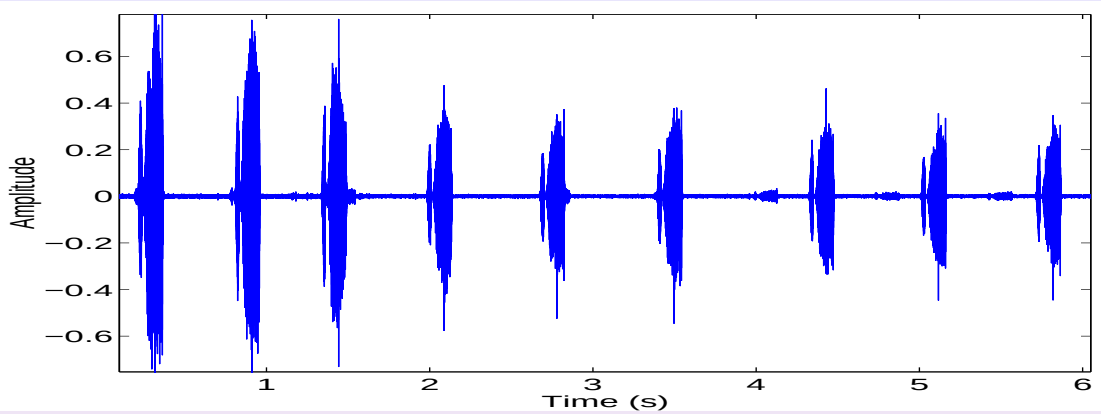
Adaptive multiple classifier systems

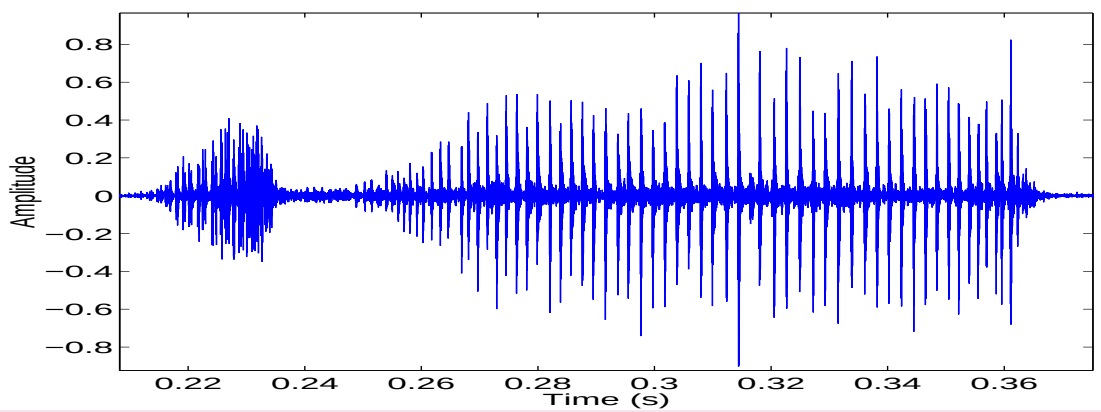


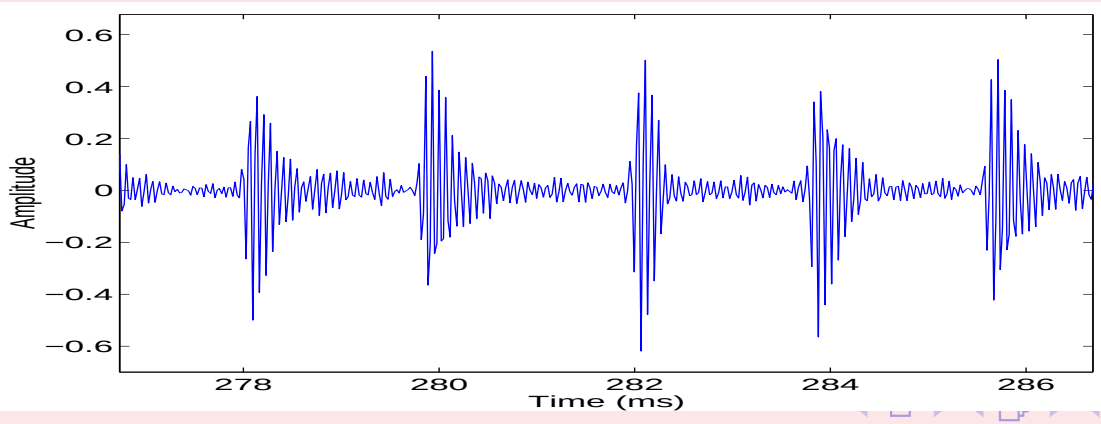
Bio-acoustic pattern recognition



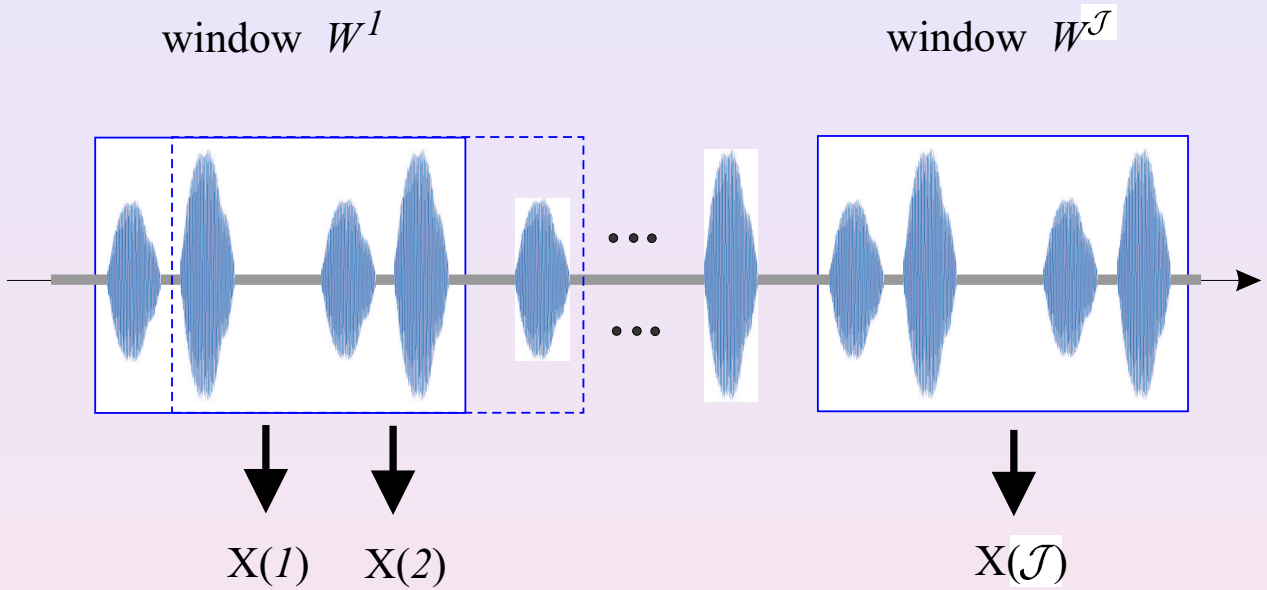
Example : *Ephippiger*







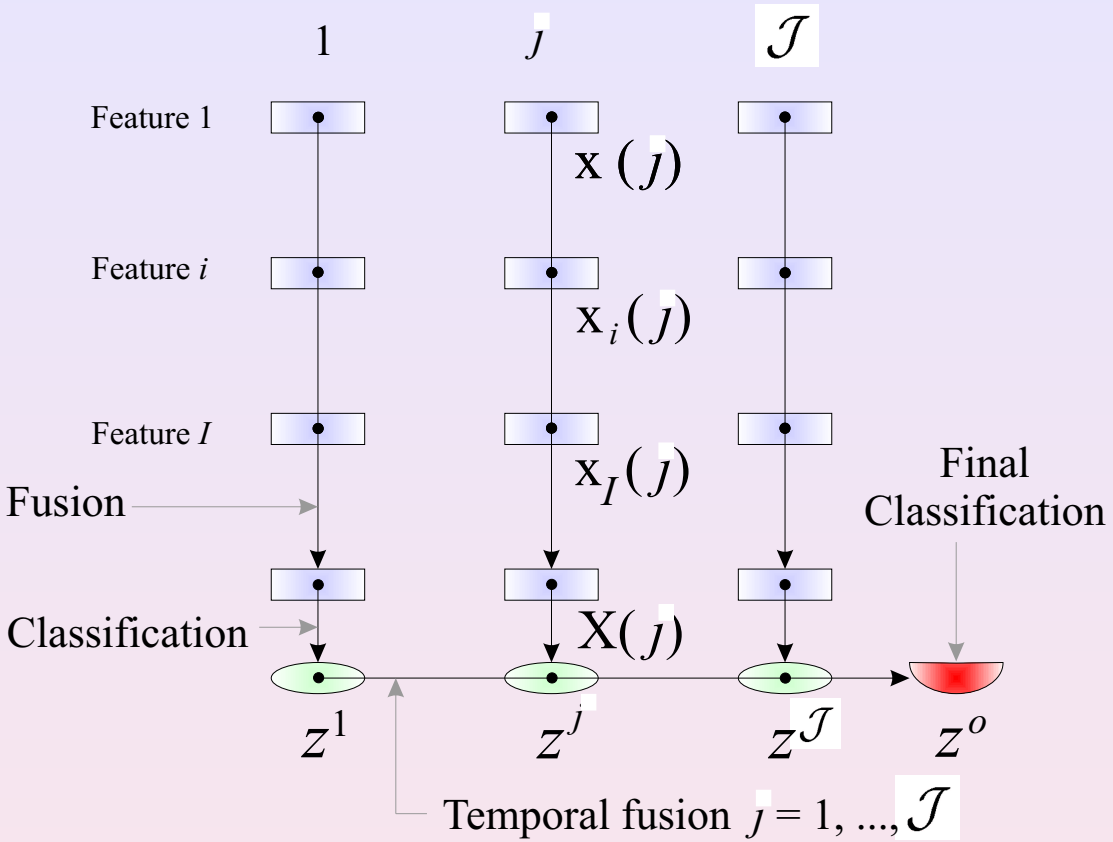
Extraction of Local Features in Time series



Signal $s(t)_{t=1}^{T}$

I local features $X(\jmath) = (\mathbf{x}_1(\jmath) \ldots \mathbf{x}_I(\jmath))$

FCT-Architecture

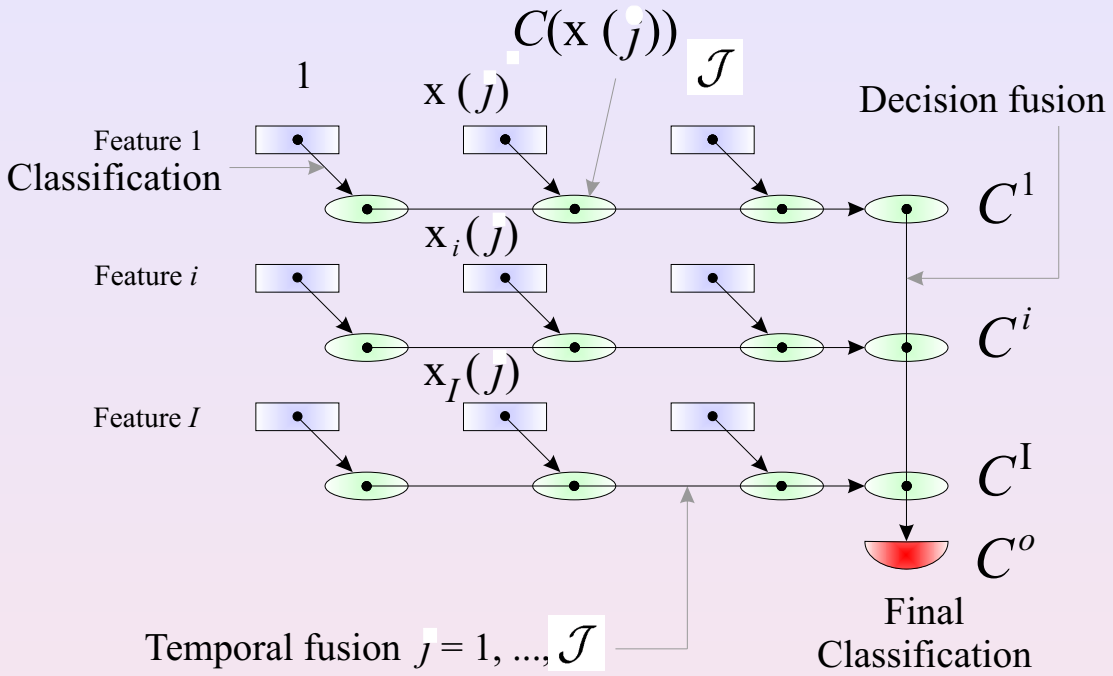


Fusion: $X(j) = (\mathbf{x}_1(j) \ldots \mathbf{x}_I(j)) \in \mathbb{R}^{\Phi}, \Phi = \sum_{i=1}^{I} d_i$

Classification: $C^j := \mathcal{C}^j(X(j))$

Temporal fusion: $C^o := \mathcal{F}(C^1 \ldots C^{\mathcal{J}})$

CDT-Architecture



Classification: $\mathcal{C}^i : \mathbb{R}^{d_i} \rightarrow \Delta, i = 1 \ldots I$

$\mathcal{C}^1(\mathbf{x}_1(J)) \ldots \mathcal{C}^I(\mathbf{x}_I(J))$

Decision fusion: $C^J := \mathcal{F}(\mathcal{C}^1(\mathbf{x}_1(J)) \ldots \mathcal{C}^I(\mathbf{x}_I(J)))$

Temporal fusion: $C^o := \mathcal{F}(C^1 \ldots C^{\mathcal{J}})$

Results for *cricket songs*

- Crossvalidation experiments (mean error rates) of 28 cricket species with 4 to 6 animals per species.
- Radial-Basis-Function Networks as first level classifiers.
- Extracted features: pulse length, pulse distance, energy contour,
- Averaged fusion lead to an error ≥ 0.1

Algorithm	ρ =0.0	ρ =0.2	ρ =0.4	ρ =0.6	ρ =0.8	ρ =1.0
DT	8.61	7.88	8.03	7.74	7.59	7.59
Multiple DT	8.32	8.03	7.15	6.86	6.86	6.72
Cluster DT	8.61	7.30	7.15	7.15	7.30	7.30

Summary

- ANN and decomposition of learning tasks
- Multiple classifier systems with static and adaptive fusion mappings.
- Fixed fusion mappings (averaging) are simple— adaptive fusion (decision templates or neural networks) provide accurate classification.
- Multi-layer neural networks support flexible fusion schemes (early, mid-level, late, multi-level fusion).
- MCS-architectures (with temporal fusion) for time series classification.
- Application of MCS to the classification of bio-acoustic time series for automated bio-acoustic monitoring.