

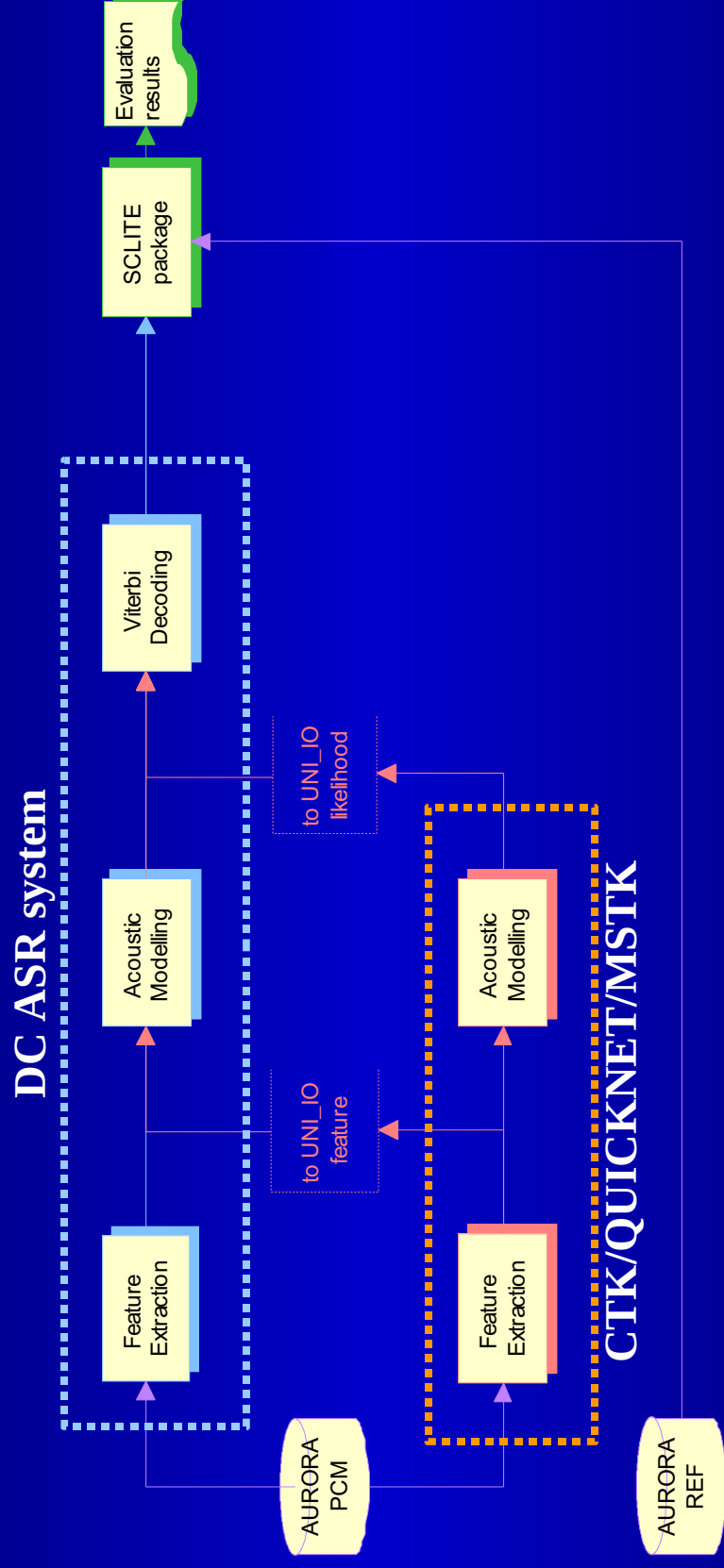
# Progress in the framework of the RESPITE project at DaimlerChrysler Research & Technology

Dr-Ing. Fritz Class and Joan Marí  
Martigny, Jan. 2002

# Contents

- DaimlerChrysler off-line demonstrator
  - Block-diagram of our off-line demonstrator
  - Next evaluation experiments using our demonstrator
- On-going research in Discriminative Feature Extraction
  - TANDEM acoustic modelling
  - Linear-Discriminant-Analysis-based (LDA) front-end
  - Quadratic-Discriminant-Analysis-based (QDA) front-end
  - A two-layer perceptron to generate state-posteriors from QDA features (RBFs)
  - Results

# DC off-line demonstrator: block-diagram



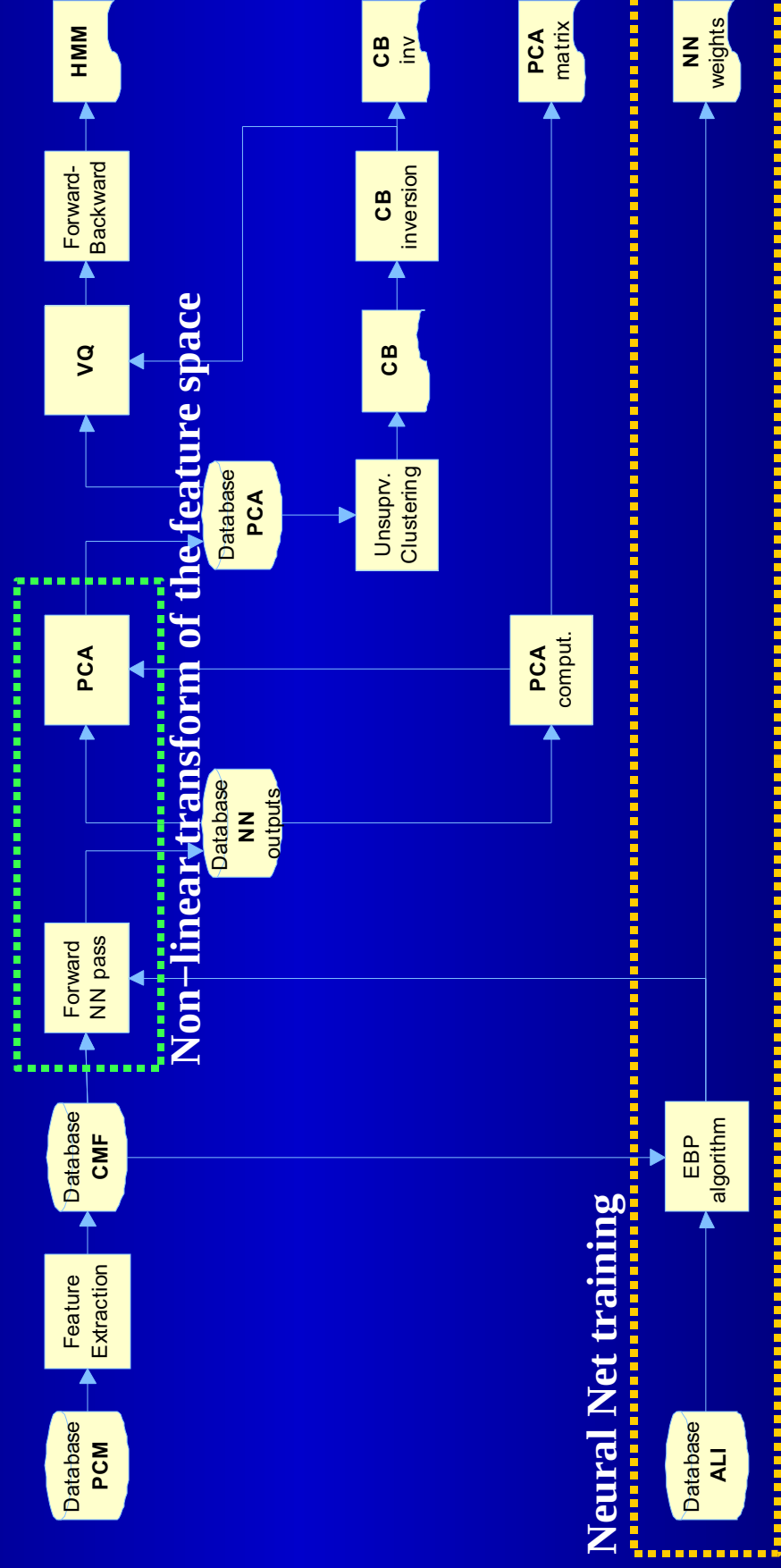
## **DC off-line demonstrator: next steps**

- Evaluate results of IDIAP Multi-Stream toolkit on the AURORA 2000 database and compare them with those of SPRACHcore and CTK toolkits
- Determine, given the results of the previous evaluation and system requirements, which is the desirable technique for our purposes
- Using our own in-car american english database compare our baseline system with the selected optimum technique

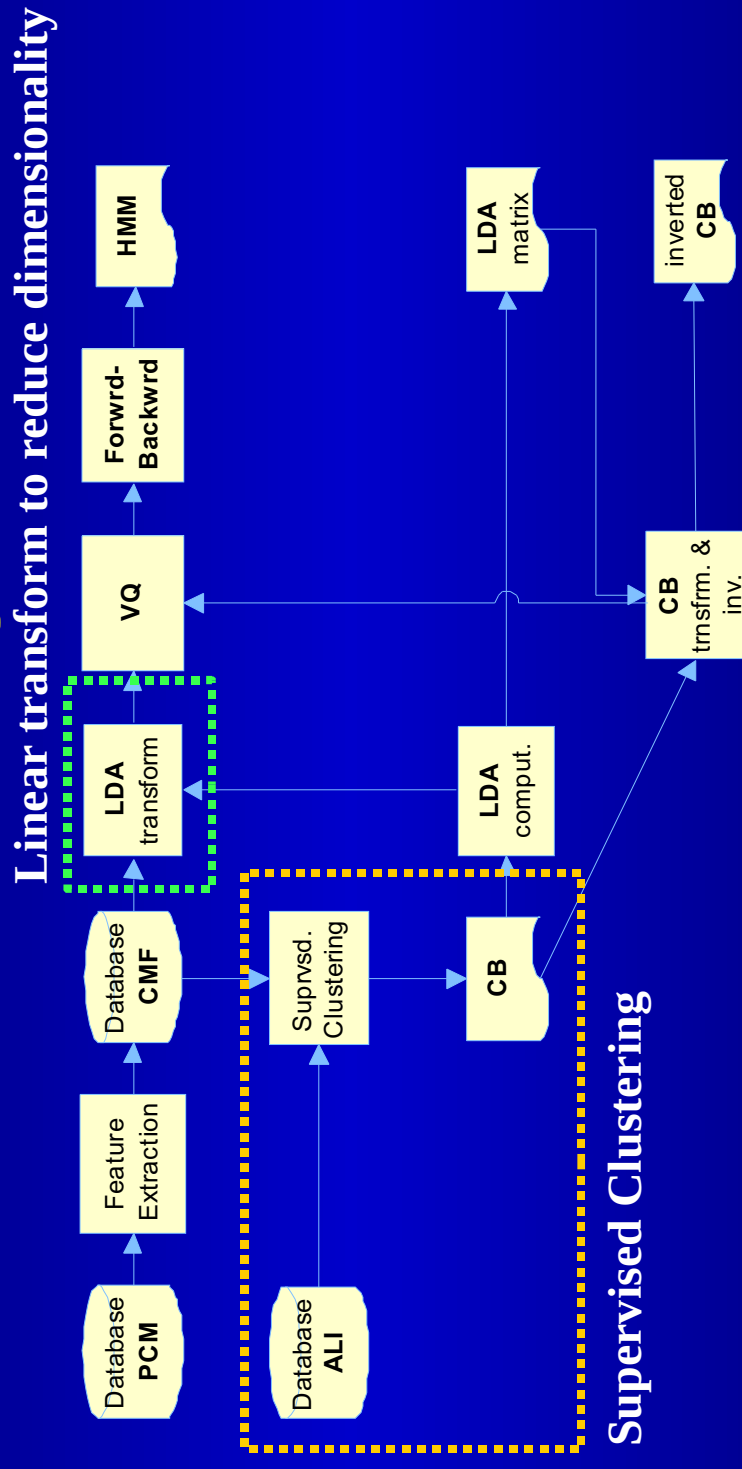
# Contents

- DaimlerChrysler off-line demonstrator
  - Block-diagram of our off-line demonstrator
  - Next evaluation experiments using our demonstrator
- On-going research in Discriminative Feature Extraction
  - TANDEM acoustic modelling
  - Linear-Discriminant-Analysis-based (LDA) front-end
  - Quadratic-Discriminant-Analysis-based (QDA) front-end
  - A two-layer perceptron to generate state-posteriors from QDA features (RBFs)
  - Results and Conclusions

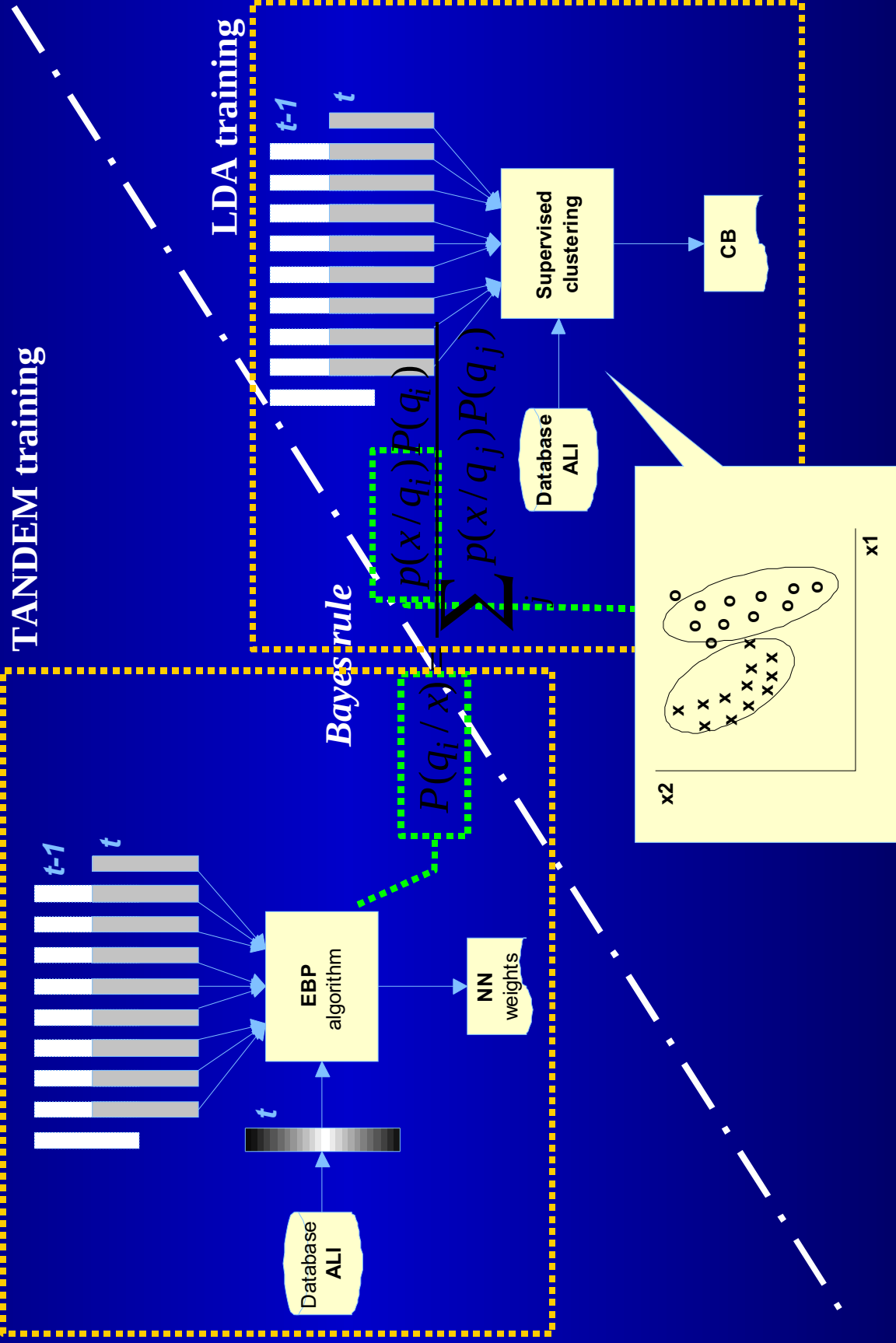
# Discriminative Feature Extraction: TANDEM training



# Discriminative Feature Extraction: LDA training



# Discriminative Feature Extraction: LDA





# Discriminative Feature Extraction: QDA

TANDEM can be interpreted as a kind of non-linear feature extraction  
TANDEM features are obtained from log-posteriors



Applying Bayes rule as in the previous slide

$$\log P(q_i / x) = \log p(x / q_i) - \log \sum_j p(x / q_j) P(q_j) + \log P(q_i)$$

A quadratic equation is obtained

$$\log P(q_i / x) \propto \frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) - \log p(x)$$

Key questions at this point are:

- Is one gaussian per cluster enough ?
- How many classes should be used ?
- Is the gaussianity assumption always a good one?

# Discriminative Feature Extraction: RBFs

A compromise between connectionist and parametric modelling are RBFs

Returning back to the Bayes rule

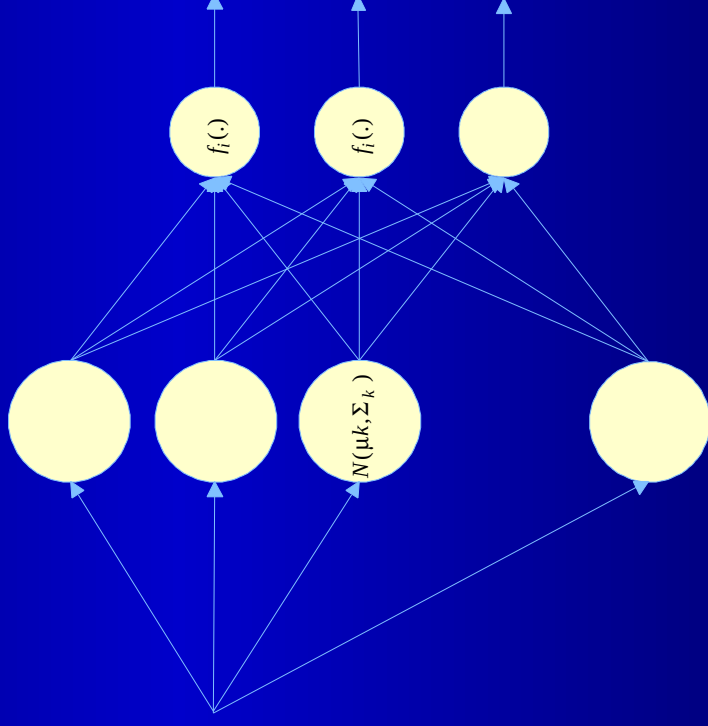
$$P(q_i / x) = \frac{p(x / q_i) P(q_i)}{\sum_j p(x / q_j) P(q_j)}$$

We could express it as:

$$P(q_i / x) = f(w_{ik}, N(\mu_k, \Sigma_k))$$

Where  $f$  is the softmax function  
and  $N$  is the gaussian pdf

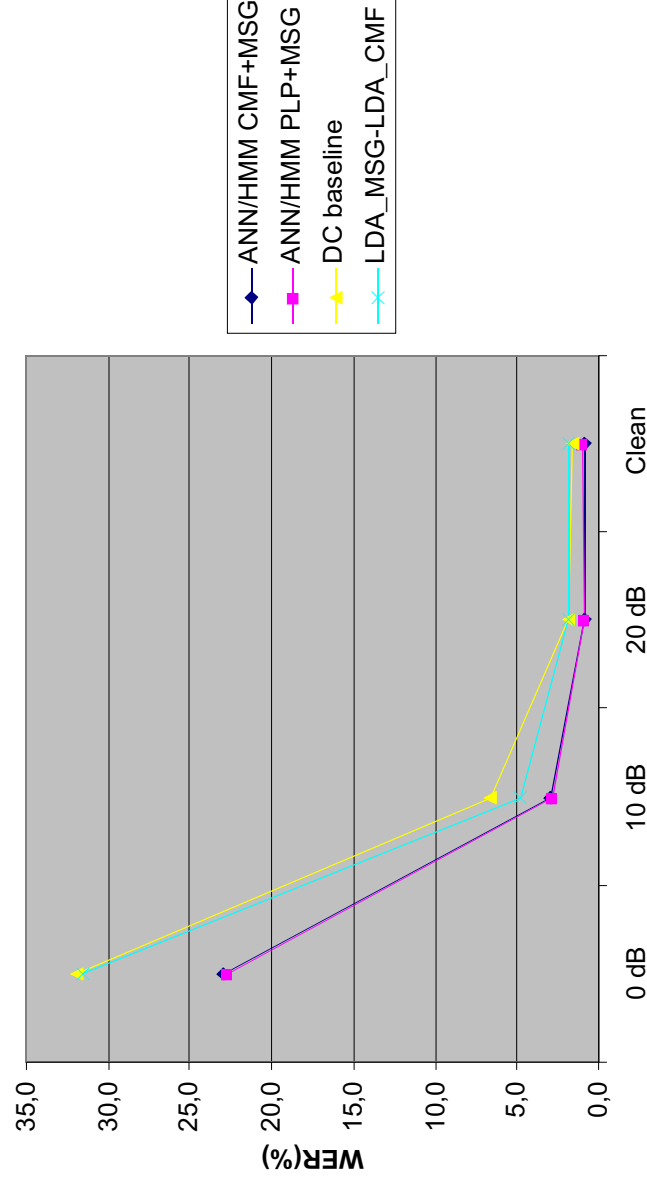
An RBF is thus obtained



# Discriminative Feature Extraction: results

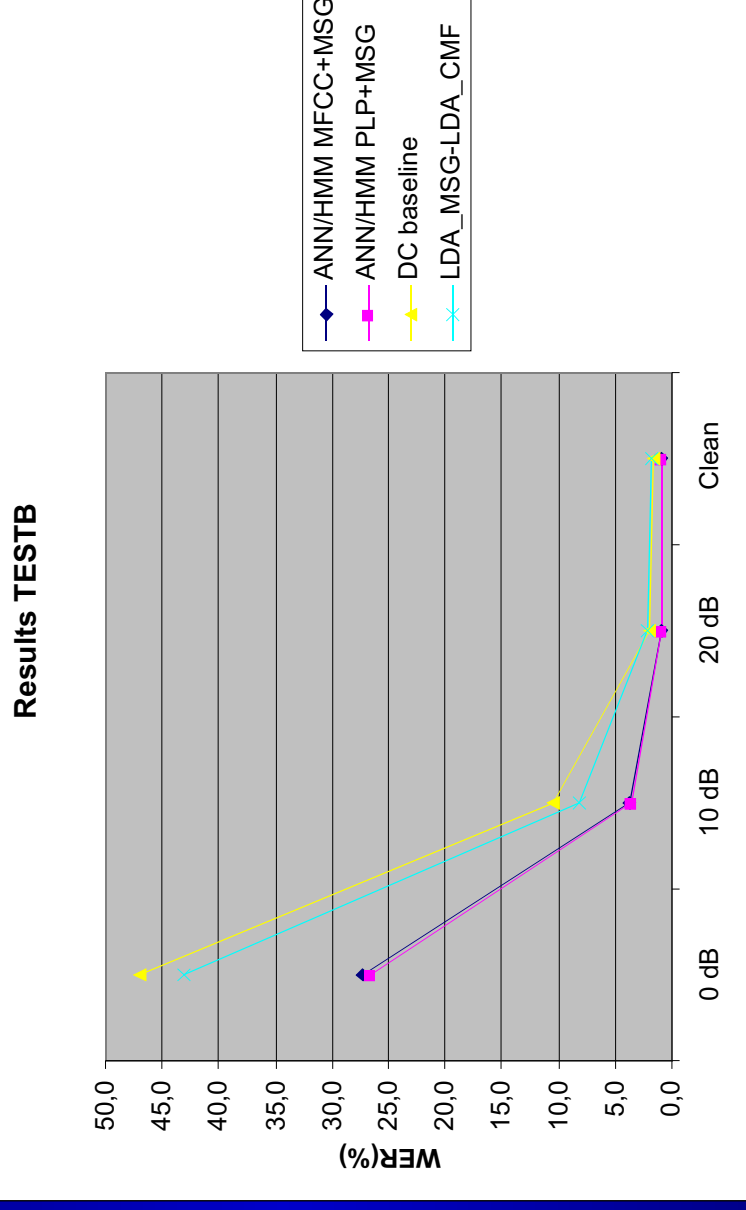
## Recognition results on AURORA 2000

Results TESTA



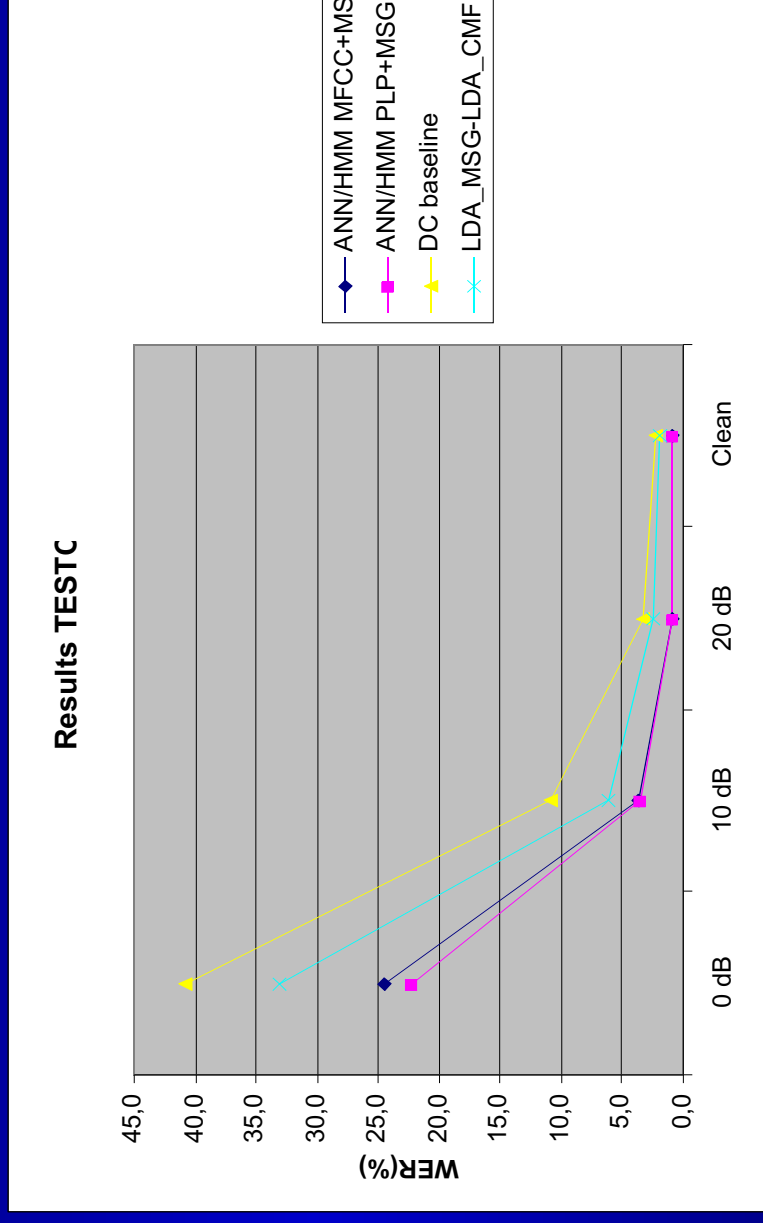
# Discriminative Feature Extraction: results

## Recognition results on AURORA 2000



# Discriminative Feature Extraction: Results

## Recognition results on AURORA 2000



# Discriminative Feature Extraction: results

## Reduction of the dimensionality of the Neural Net

| TESTB-STREET WORD MLP (MFCC) |      |       |       |       |         |         |
|------------------------------|------|-------|-------|-------|---------|---------|
| Double Delta                 |      |       |       |       |         |         |
| MLP Type                     | 0 dB | 10 dB | 20 dB | clean | Average | Weights |
| (3x9x11)=297 + 480 + 127     | 37,5 | 4,4   | 1,8   | 2,4   | 11,5    | 203.520 |
| (3x9x11)=297 + 254 + 127     | 36,0 | 7,1   | 2,9   | 2,7   | 12,2    | 107.696 |
| Delta                        |      |       |       |       |         |         |
| MLP Type                     | 0 dB | 10 dB | 20 dB | clean | Average |         |
| (2x9x11)=198 + 480 + 127     | 35,2 | 6,9   | 2,6   | 2,6   | 11,8    | 156.000 |
| (2x9x11)=198 + 254 + 127     | 36,6 | 7,5   | 2,7   | 2,9   | 12,4    | 82.550  |
| No Delta                     |      |       |       |       |         |         |
| MLP Type                     | 0 dB | 10 dB | 20 dB | clean | Average |         |
| (17 x 11)=187 + 480 + 127    | 36,5 | 7,8   | 3,1   | 3,1   | 12,6    | 150.720 |
| (17 x 11)=187 + 254 + 127    | 37,4 | 8,5   | 3,0   | 3,3   | 13,1    | 79.756  |
| (13 x 11)=143 + 480 + 127    | 37,7 | 8,3   | 2,7   | 2,8   | 12,9    | 129.600 |
| (13 x 11)=143 + 254 + 127    | 39,1 | 8,3   | 3,1   | 3,3   | 13,5    | 68.580  |
| (9 x 11)=99 + 480 + 127      | 39,1 | 8,1   | 2,9   | 2,8   | 13,2    | 108.480 |
| (9 x 11)=99 + 254 + 127      | 40,1 | 8,7   | 3,1   | 3,1   | 13,8    | 57.404  |

## **Discriminative Feature Extraction: conclusions**

- TANDEM acoustic modelling can be performed with discriminant parametric models too (QDA)
- As a compromise between connectionist and parametric modelling RBFs can be used for TANDEM
- Concatenation of LDA-PLP and LDA-MSG features results in a slight improvement to our baseline LDA system
- Word-based Hybrid ANNN/HMMs are the best performing