

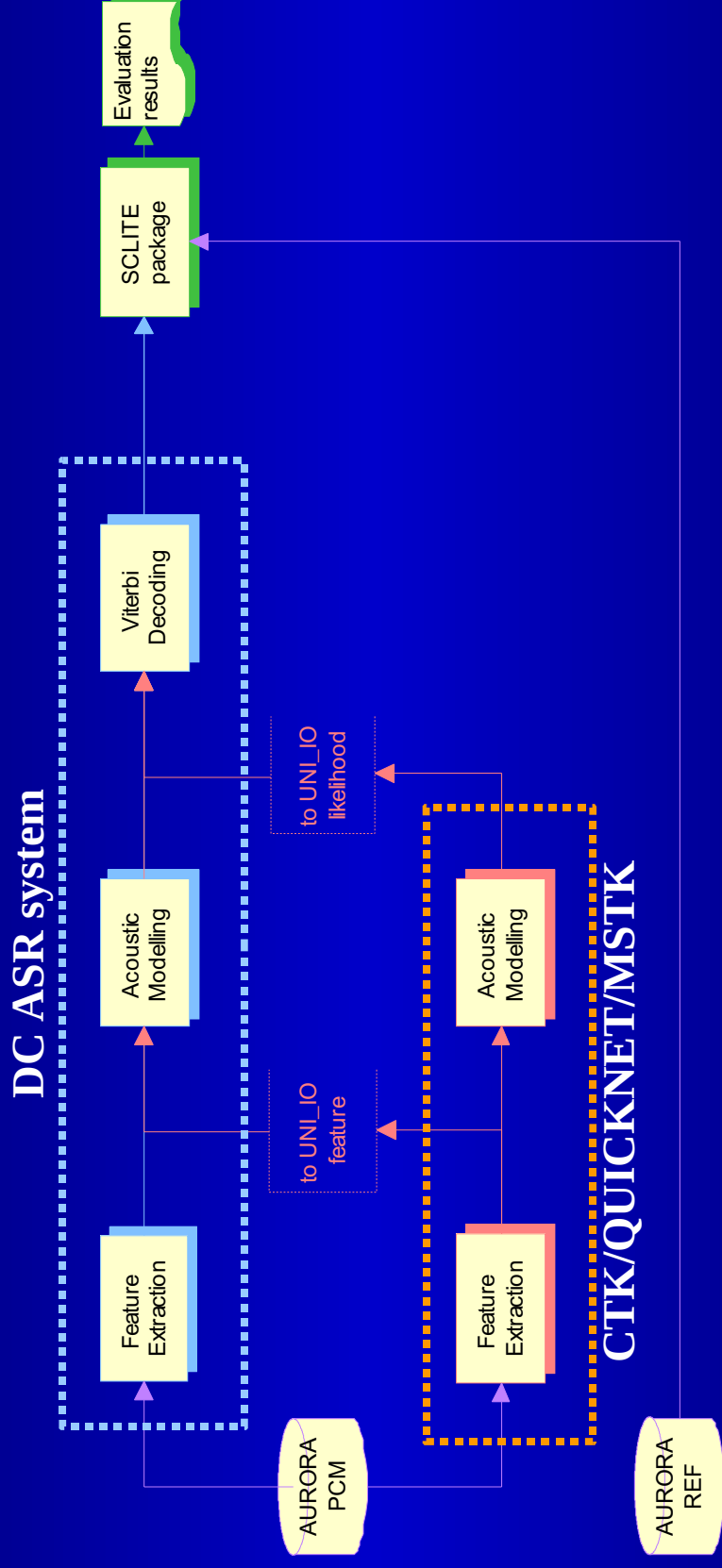
Progress in the framework of the RESPITE project at DaimlerChrysler Research & Technology

Dr-Ing. Fritz Class and Joan Marí
Sheffield, June 2002

Contents

- DaimlerChrysler off-line demonstrator
 - Block-diagram of our off-line demonstrator
 - Evaluation experiments using our demonstrator
- On-going research in Discriminative Feature Extraction
 - TANDEM acoustic modelling
 - Clustering of HMM-states to define discriminative feature space
- British-English recognizer
- „Online demonstrator“

DC off-line demonstrator: block-diagram



DC off-line demonstrator: results on AURORA 2000

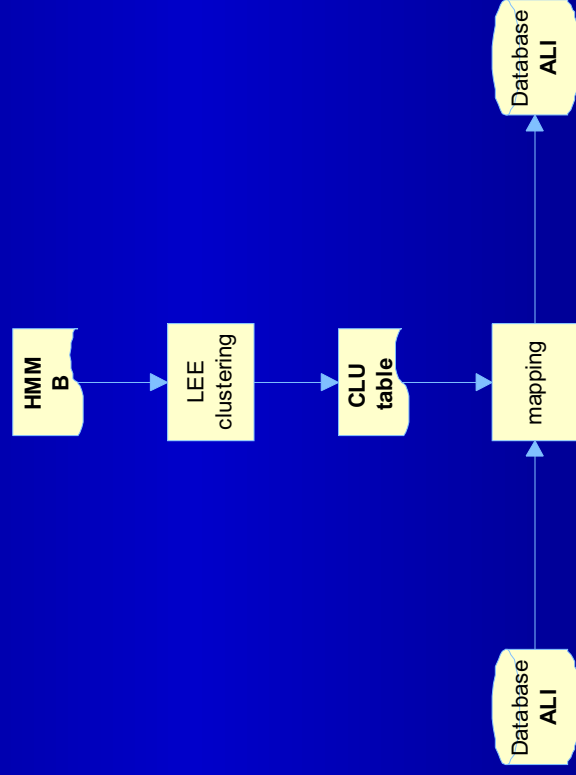
<i>recogniser</i>	<i>Average %WER on AURORA 2000</i>	<i>training set</i>
DC baseline	13,23	multicondition
FPM	11,8	clean + white noise
Sheffield	19,3	clean
ICSI	8,4	multicondition
IDIAP	?	?

Discriminative Feature Extraction: TANDEM Acoustic Modelling

- The TANDEM approach finds a feature space transform to reduce the dimensionality while preserving important classification information
 - Typically, dimensionality is reduced from N to n ($N \gg n$), where n is the number of phones in the digit set, and encode thus essential classification information
- Similar to LDA concept, but with a different criterion to be minimised and of course a non-linear mapping instead
 - Typically LDA transform is found by assigning to each HMM-state a high dimensional gaussian distribution, and minimising a criterion based on the inter- and intra-state scatter covariance matrices computed from the state-gaussians
- Both approaches can be theoretically linked using Bayes Classifier Theory

Discriminative Feature Extraction: Clustering HMM-states to define discriminative feature space

- The idea is to cluster HMM-states which have similar mixtures of gaussians, because then the state-likelihoods of those states for a given frame will be similar
- This target-clusters encode the essential classification information
- By applying a mapping, a new state-cluster alignment is obtained, which can be used to train NN for discriminative feature extraction



Discriminative Feature Extraction: results on AURORA 2000

<i>recogniser</i>	<i>Average %WER on AURORA 2000</i>
LDA DC baseline	12,85
PLP TANDEM (24 phonemes)	9,9
PLP Cofs NLDA (24 state-clusters)	10,28
PLP Cofs NLDA (32 state-clusters)	10,1

British-English recognizer

- newly developed with latest perceptions out of RESPITE work
- training with inhouse real car database;
- about 1000 native english speakers (60% male, 40% female)
- 16 khz sampling rate
- training:
 - vocabulary: commands for handling car functions like car-phone, audio (adio, cd, climate,...), navigation system; city and street names; digit strings; spelling; longer sentences (conversation)
 - pre-version; optimizations are under work
- tests:
 - test set 3800 utterances, not included in training set
 - digit strings (2300), spelling (1000), city/street names (500)
 - test vocabulary: 350 words
 - tests on commands are under work

British-English recognizer: pre-version results (% word error rate)

<i>recognizer:</i>	<i>old (2000)</i>	<i>new (May 2002)</i>
city-, street names	15.4	10.3
spelling	16.9	12.5
digit strings	2.9	2.6