Multiband With Contaminated Training Data
Results on AURORA 2

TCTS
Faculté Polytechnique de Mons
Belgium

POLY
TECH.
MONS

INTRODUCTION

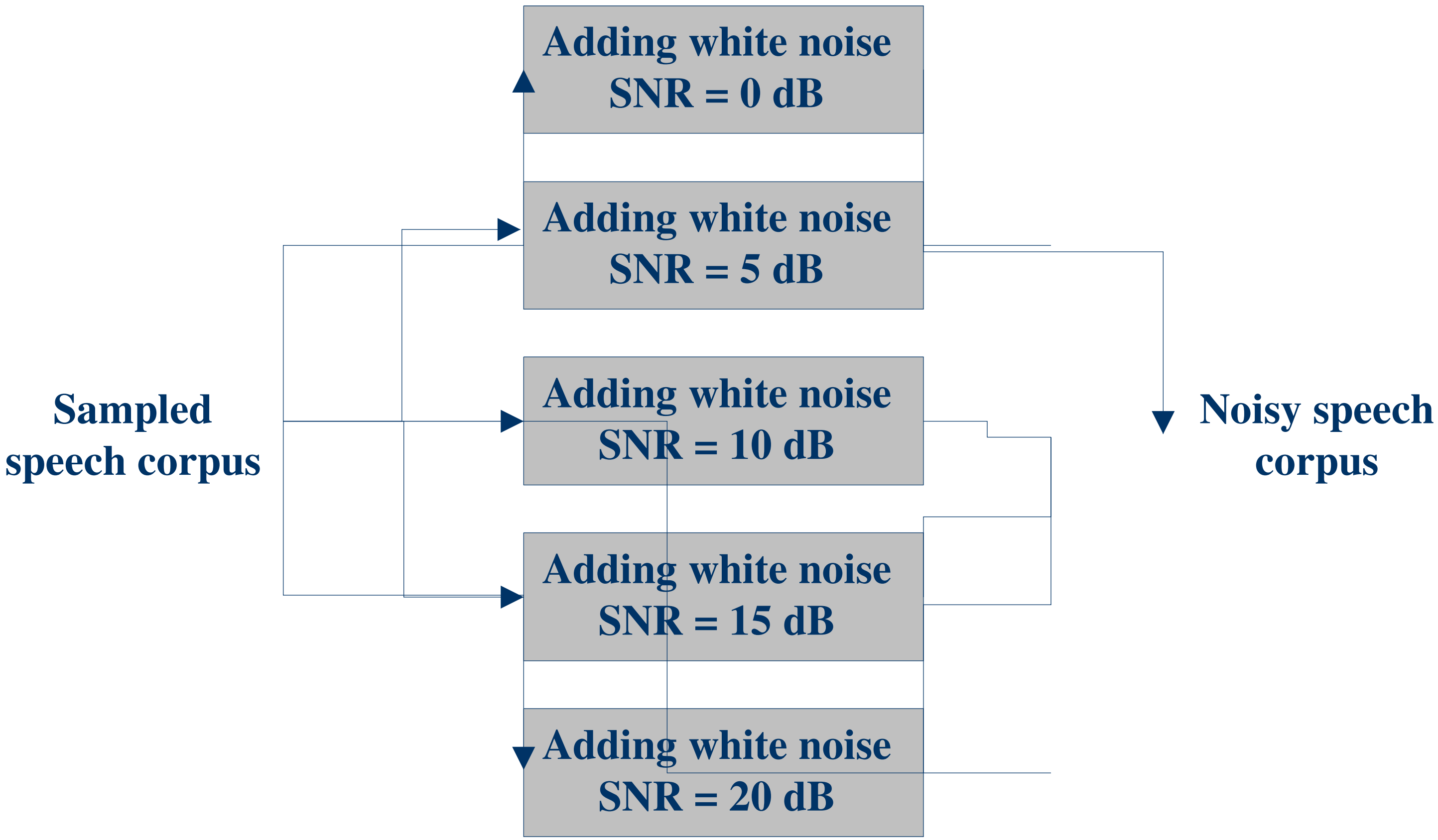
- The noise contamination of speech corpus leads to quasi–optimal performance when test noise conditions match training noise condition.

- We observe that, in narrow frequency bands, the noise characteristics basically differ by their level only.

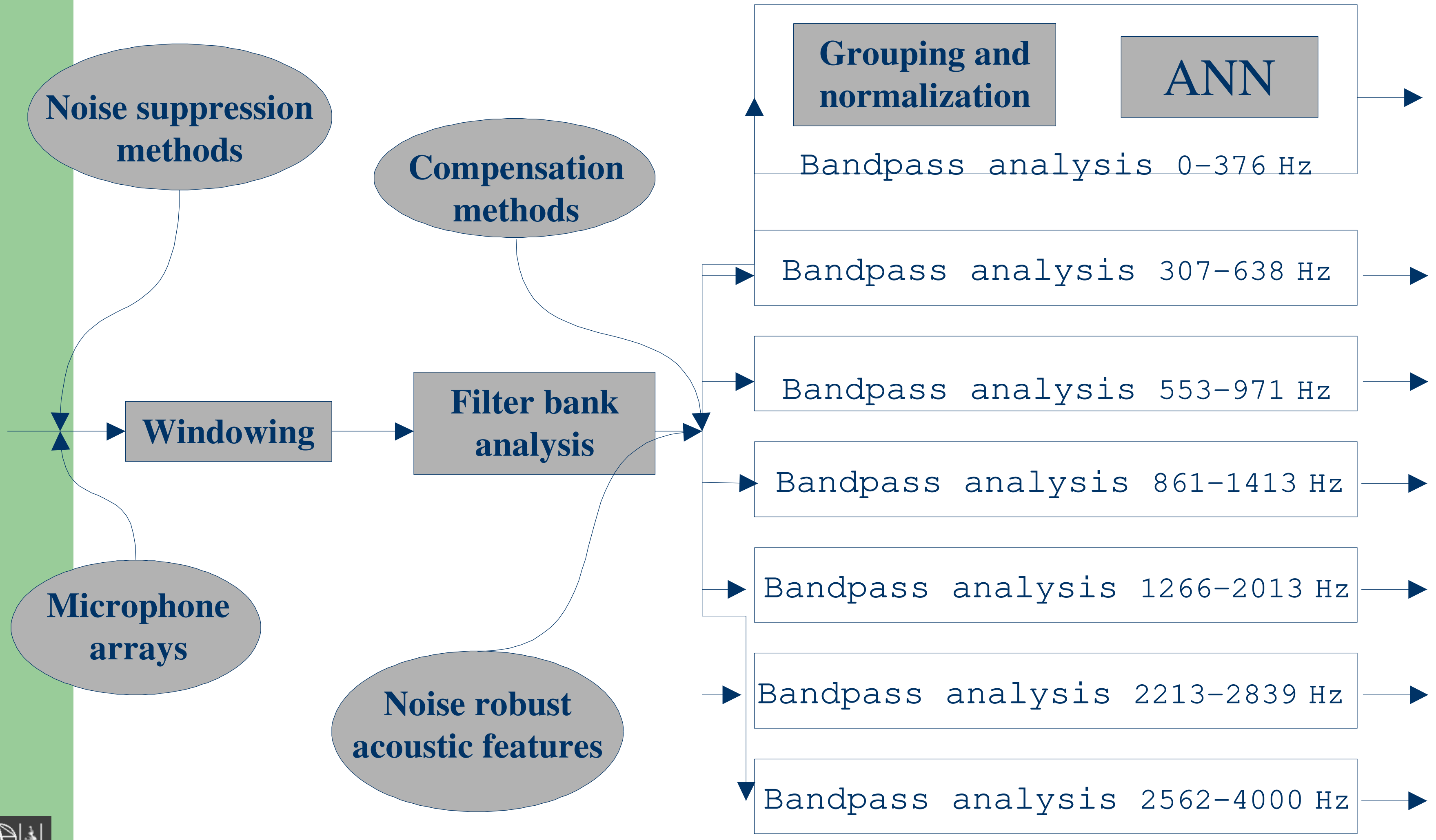
- Combining the multiband approach and the training data contamination can lead to models robust models for any kind of noises.

- We train models in each subband from data corrupted by white noise at different SNR. Subbands are then recombined using a MLP.

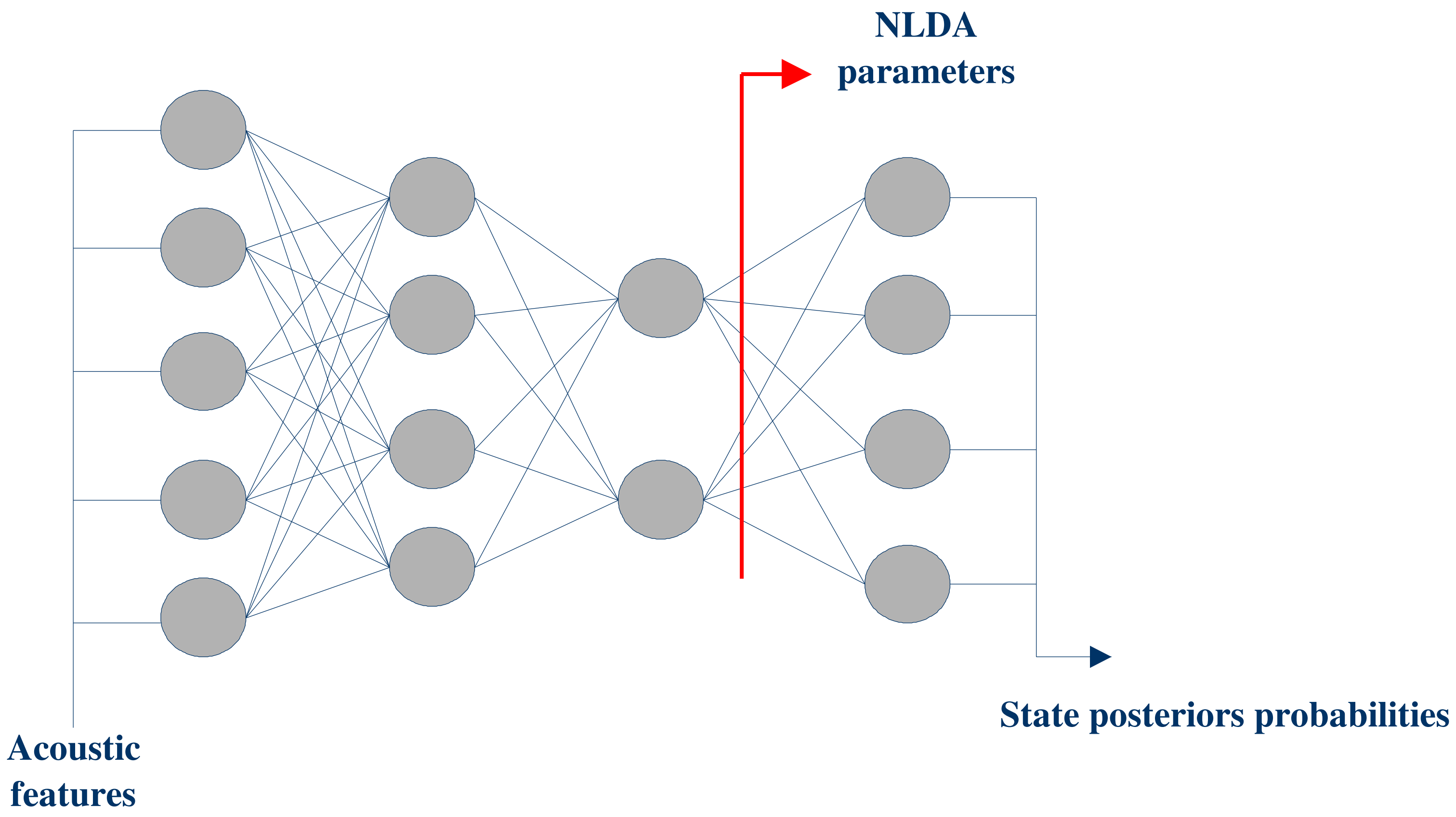
CONTAMINATED TRAINING CORPUS



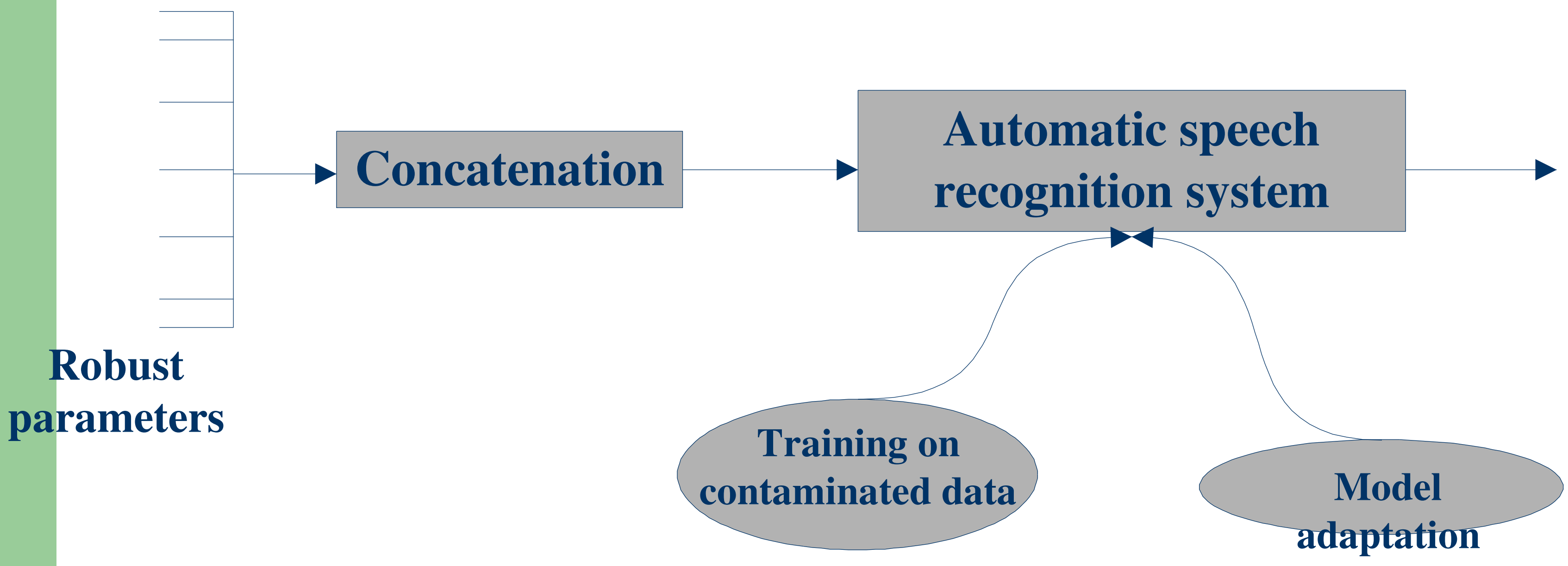
MULTIBAND ANALYSIS



NONLINEAR DISCRIMINANT ANALYSIS



ROBUST ASR


Concatenation
Automatic speech recognition system
Robust parameters
Training on contaminated data
Model adaptation

AURORA 2

Clean training set: 8440 utterances

Multi–condition training set: 8440 utterances

Contaminated training set: 8440 utterances corrupted by white noise + 4220 clean utterances.

Test set 'a': 4 different kinds of noises matching the multi–condition training set covering SNR from clean speech to –5 dB.

Acoustic models: Hybrid HMM/MLP trained on Daimler–Chrysler word models (127 HMM states).

Recognition: STRUT Viterbi decoder, no syntax

TEST CONDITIONS

Clean training set/J–RASTA

MLP: (15*13) x 1000 x 127 = 323,195 parameters

Multi–condition training set/J–RASTA

MLP: (15*13) x 1000 x 127 = 323,195 parameters

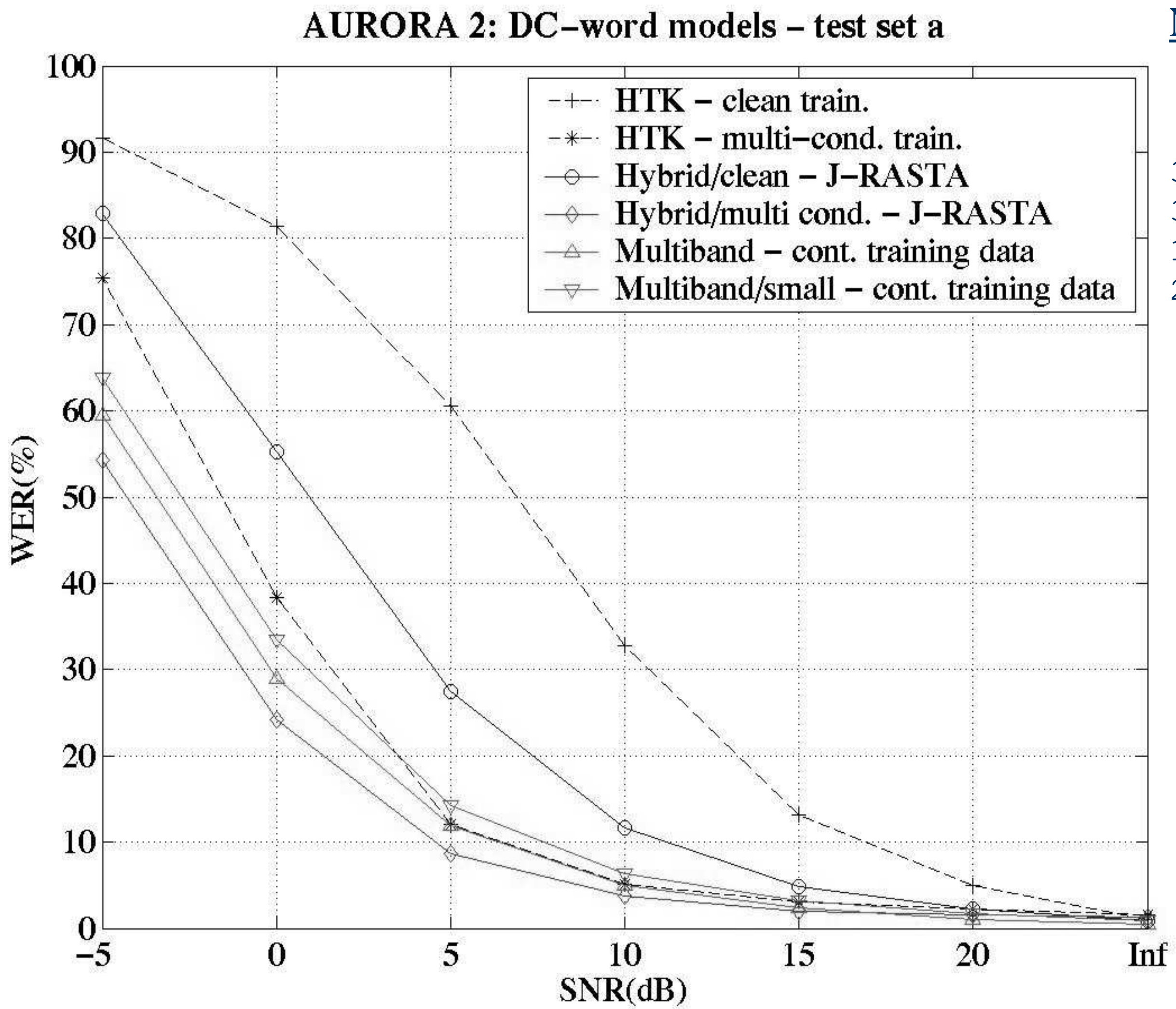
Contaminated training set/multiband

–7 subbands (15*4) x 1000 x 30 x 127
Recombination MLP: (3*210) x 1000 x 127
Total: 1,531,185 parameters

–7 subbands (15*4) x 150 x 30 x 127
Recombination MLP: 210 x 500 x 127
Total: 285,565 parameters

POLY TECH. MONS

RESULTS



Number of parameters

323,195
323,195
1,531,185
285,565

CONCLUSIONS

The combination of the multiband paradigm and training data contamination has been tested on the reference task: AURORA 2.

We got up to 57% relative improvement compared to robust features such as J–RASTA PLP features.

Compared to matching noise condition training, WER are only 10% (relative) higher.

Test with a very « light » system led to a small degradation of recognition performance.

POLY
TECH.
MONS