Soft Decisions in Missing Data ASR

Speech and Hearing Research Group,

Dept. Computer Science,

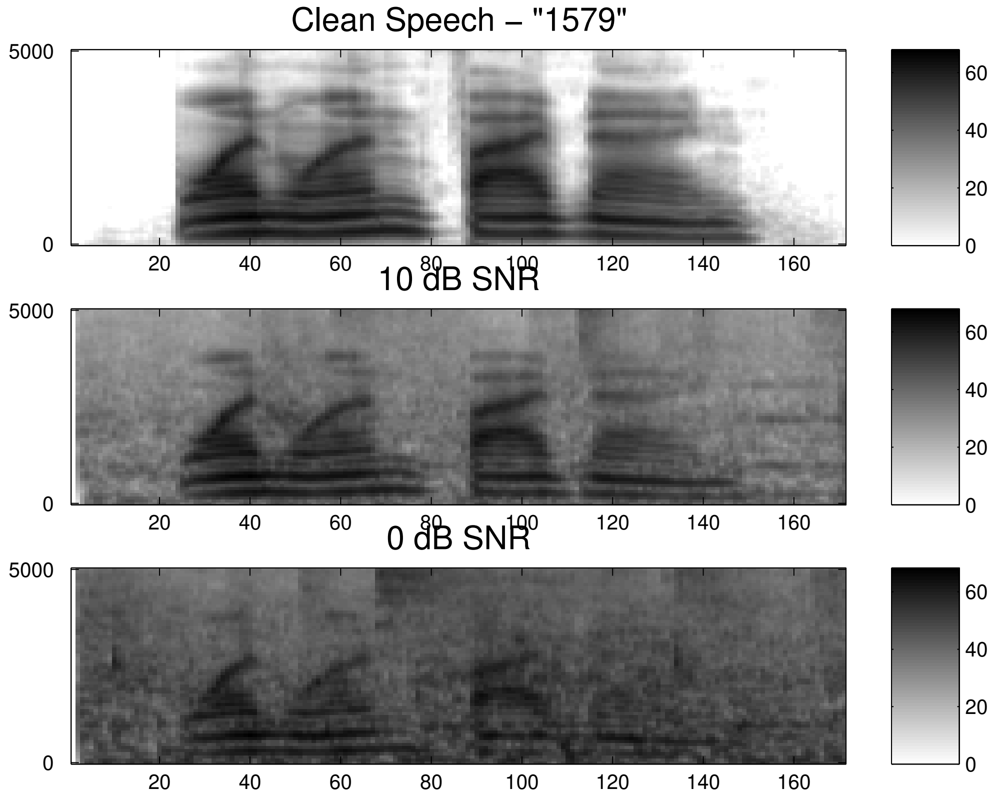
University of Sheffield, UK

September 14, 2000

Soft Decisions in Missing Data ASR

- The Missing Data approach to ASR
- Introducing Soft Decision
- Experiments with TIDigits
- Conclusions

Missing data ASR



The Missing Data Approach to ASR

- Generate spectro-temporal representation of noisy signal.
- Generate missing-data mask - i.e. labelling of each point as either clean speech or missing data.
- Perform standard ASR **with models trained on clean speech.**

  ASR employs adapted probability calculations:

  – For clean speech points employ likelihood as usual.

  – For missing data points employ ‘induction constraint’.

Generating the missing data mask.

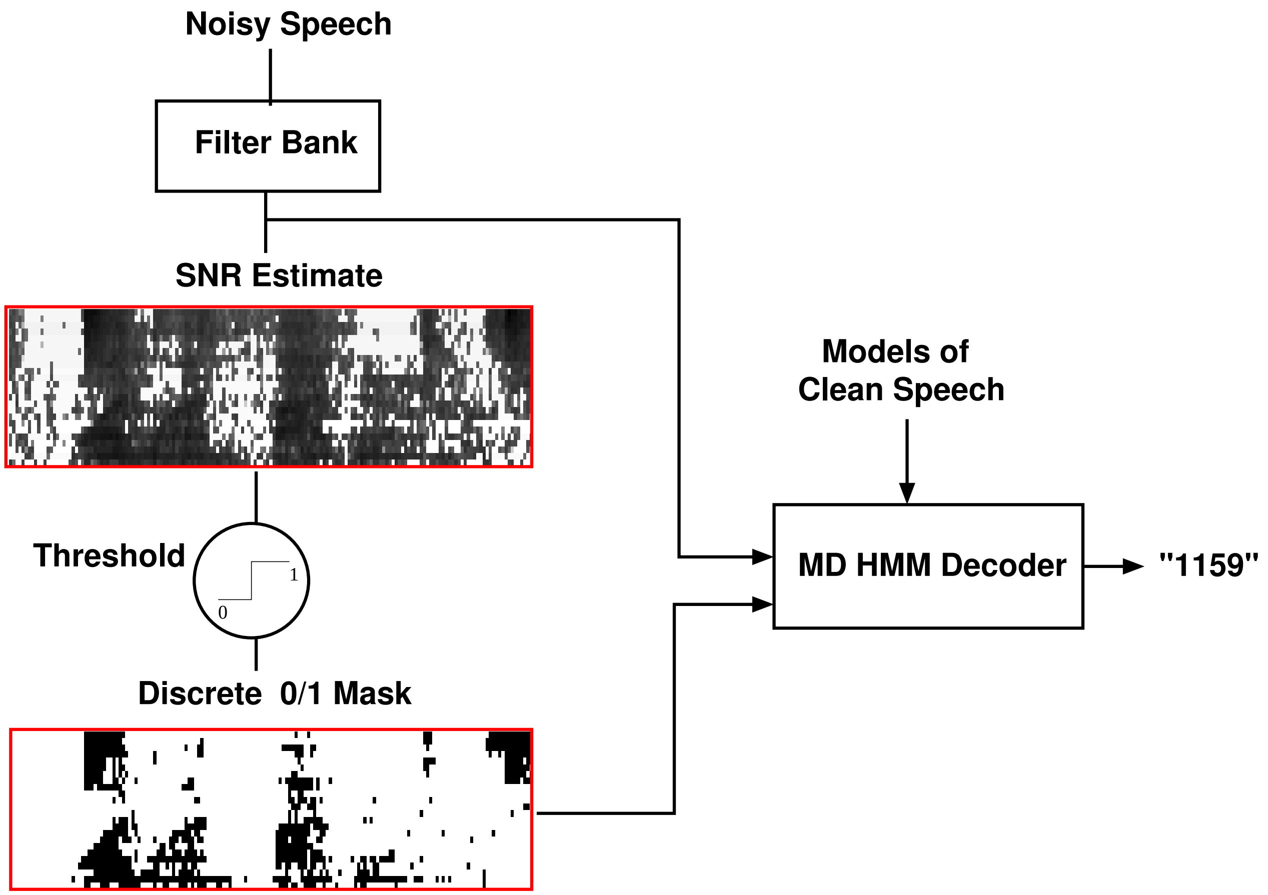
- **Using simple noise estimation techniques.**

  Generate signal and noise estimates and mark regions as reliable where the *estimated local* SNR is higher than some fixed threshold.

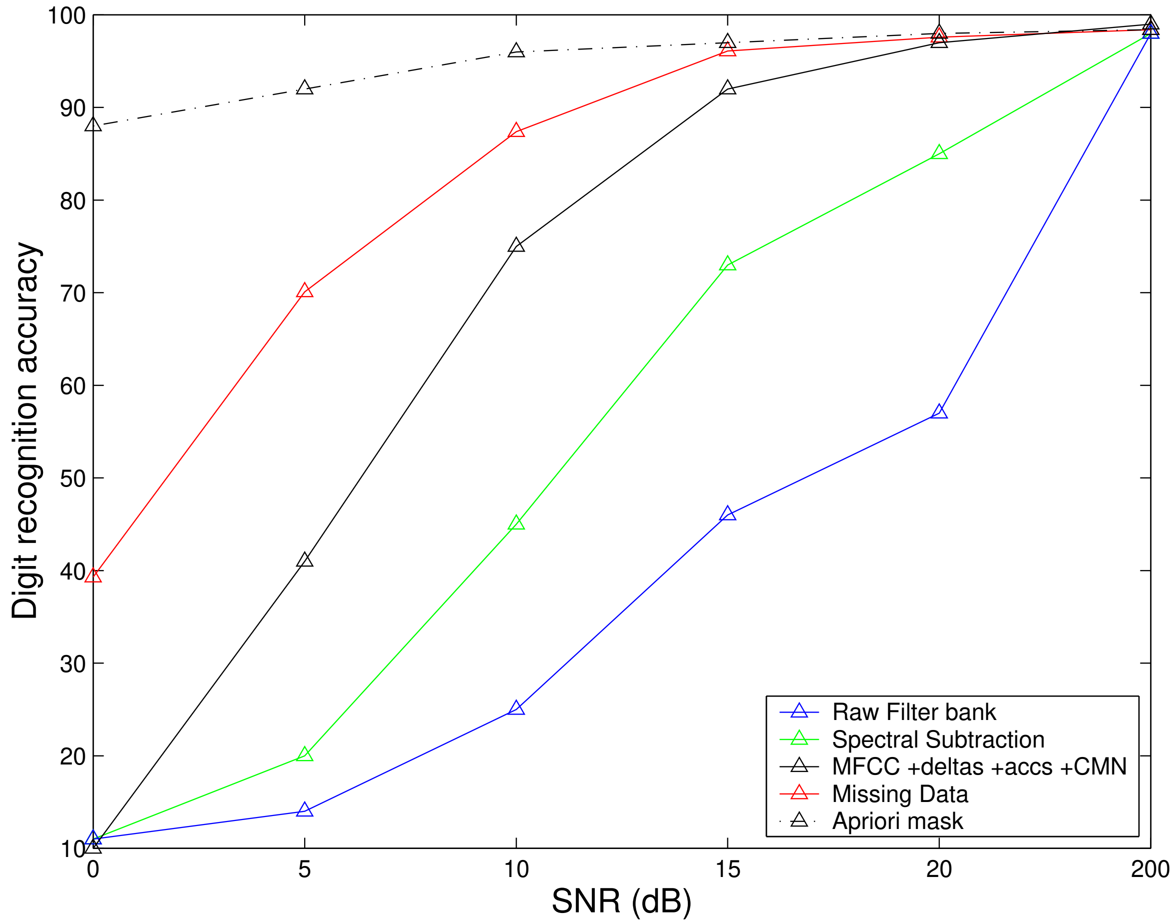
- **Cheating** - using a priori knowledge of the noise signal.

  Compare noise and signal before mixing and mark regions as reliable where the *true local* SNR is high.

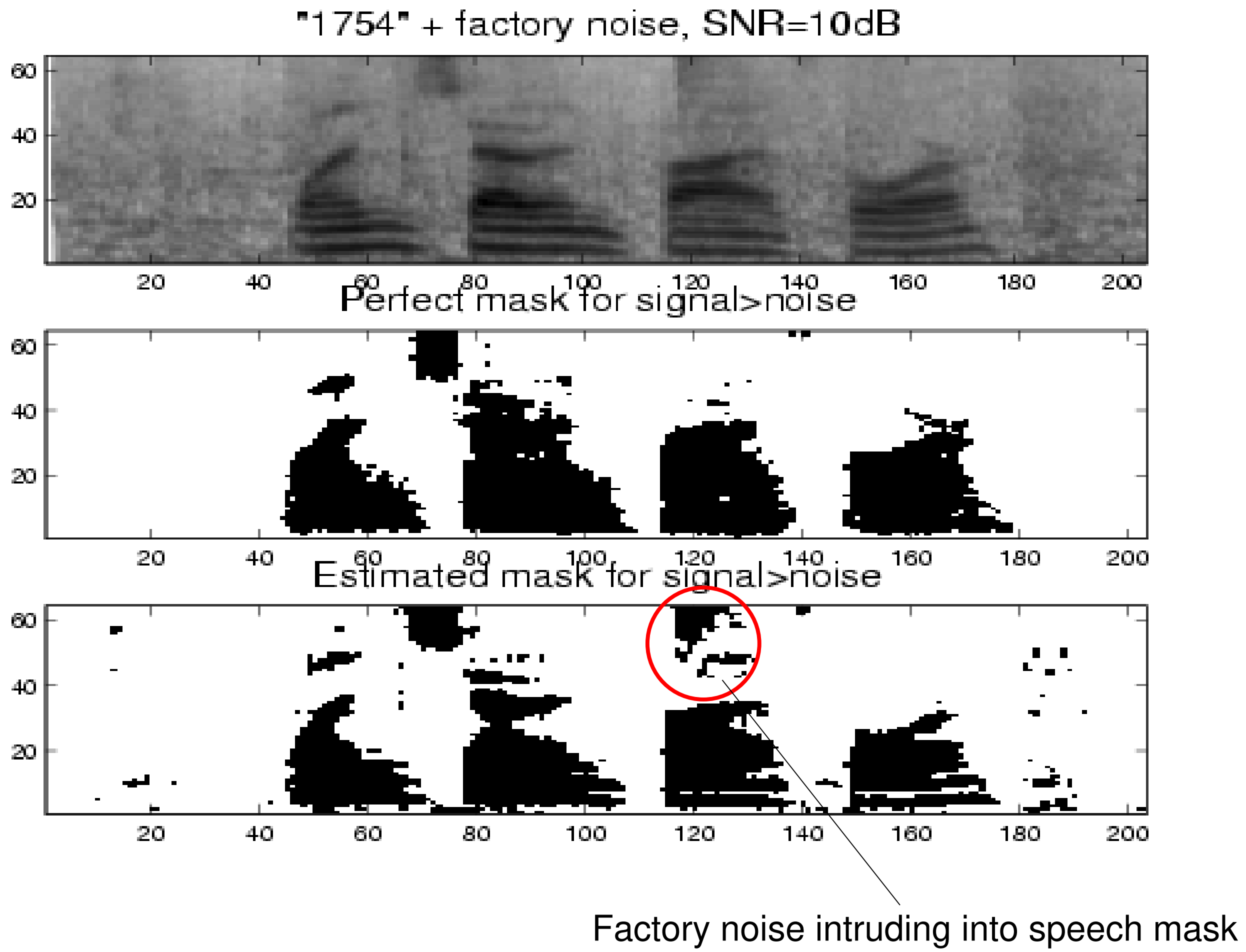
Missing Data System



Factory Noise Results



‘Perfect’ versus Estimated Masks for Factory Noise



Problem with the Discrete Missing Data Mask

Problem:

- Mask should indicate points where data is reliable
  i.e. where **true** SNR > fixed threshold.
- However, we can only directly calculate where
  estimated SNR > fixed threshold.
- Estimation error in SNR –> error in mask.

Soft Decisions in Missing Data

Solution:

Replace hard decision, **true SNR > threshold**, with soft probability **p(true SNR > fixed threshold)**.

i.e. Rather than using a **discrete 1/0 mask** we use a **fuzzy mask** containing values between 0.0 and 1.0.

These values are interpreted as "the probability that the data is reliable".

Making the Soft Mask

We need to find: $p(S > X|\hat{S})$

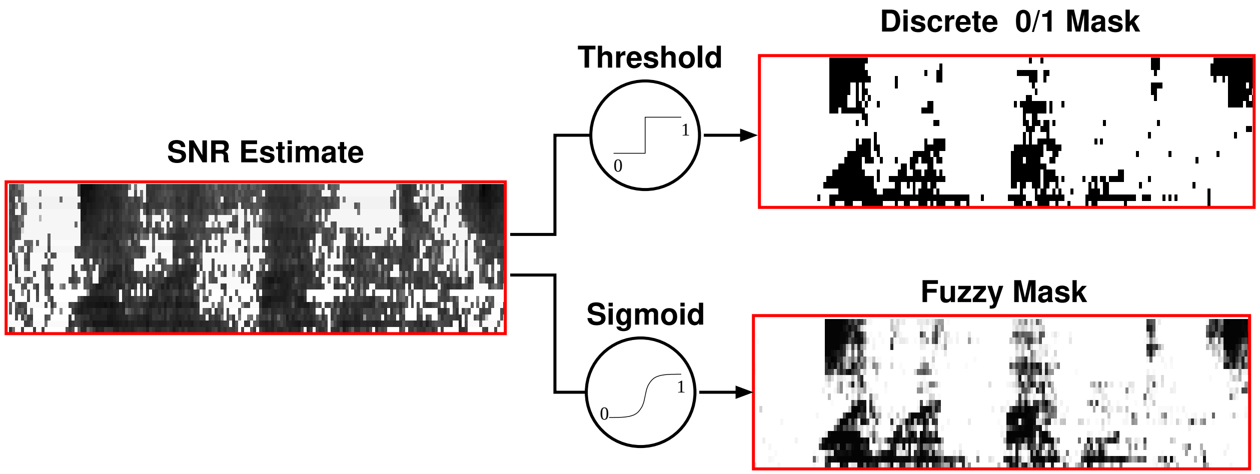
Where S is the true SNR, $\hat{S}$ is the estimated SNR and X is the fixed threshold.

Our SNR is based on a stationary noise estimation.

So assuming noise has a Gaussian distribution with variance, v, estimation error will be Gaussian with mean 0, and variance v and:

$$p(S > X|\hat{S} = y) = erf(\frac{y-X}{v})$$

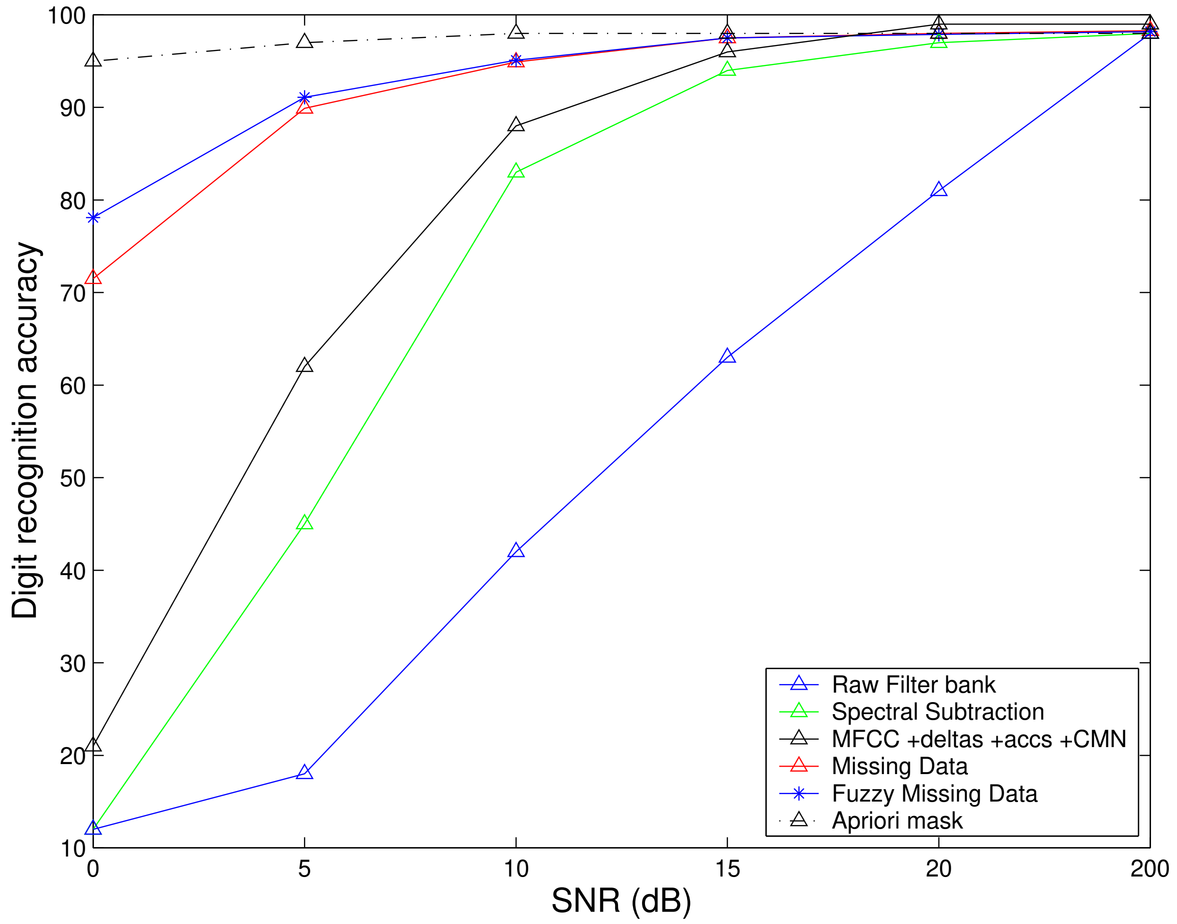
Soft Decisions in Missing Data



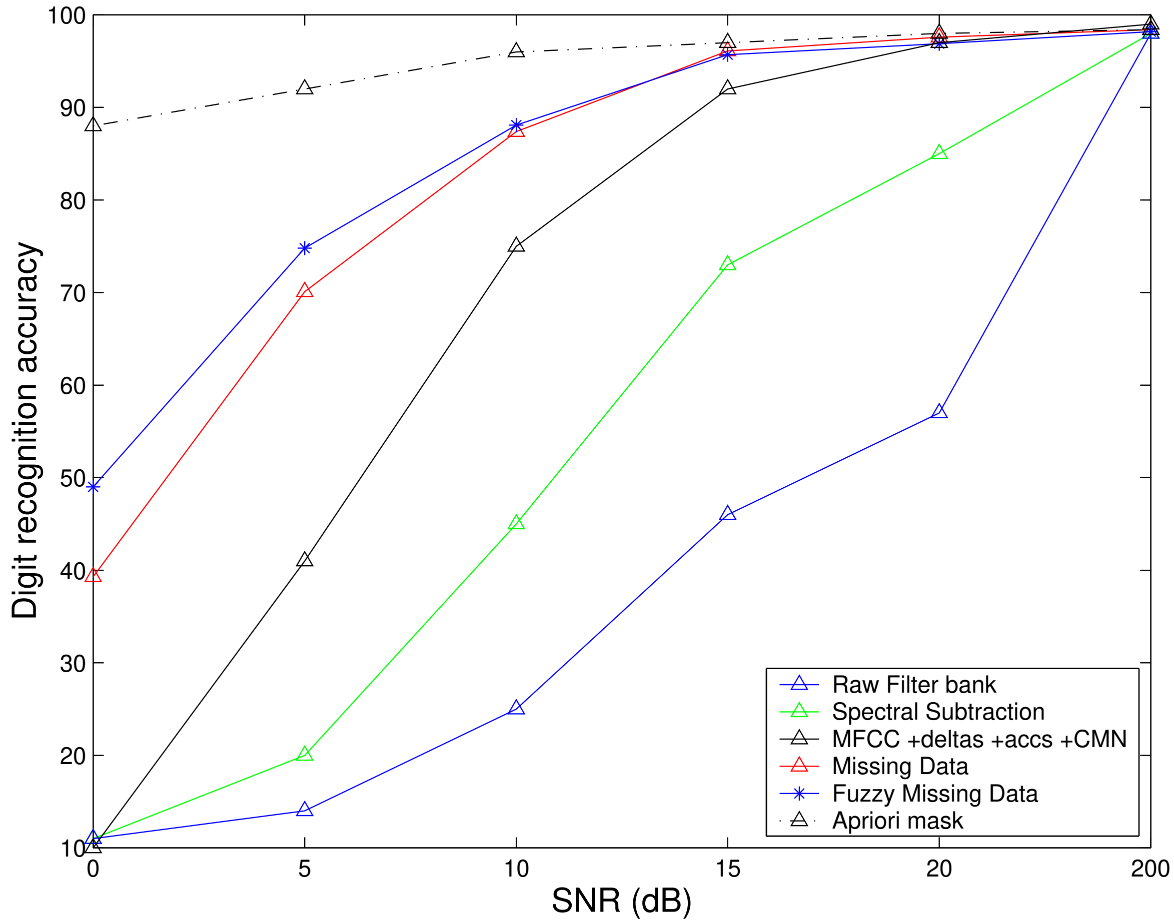
Soft mask values are interpreted as "the probability that the data is reliable".

So rather than use the *present data likelihood* **OR** *the missing data 'induction constraint'*, every point uses weighted sum of **BOTH** terms.

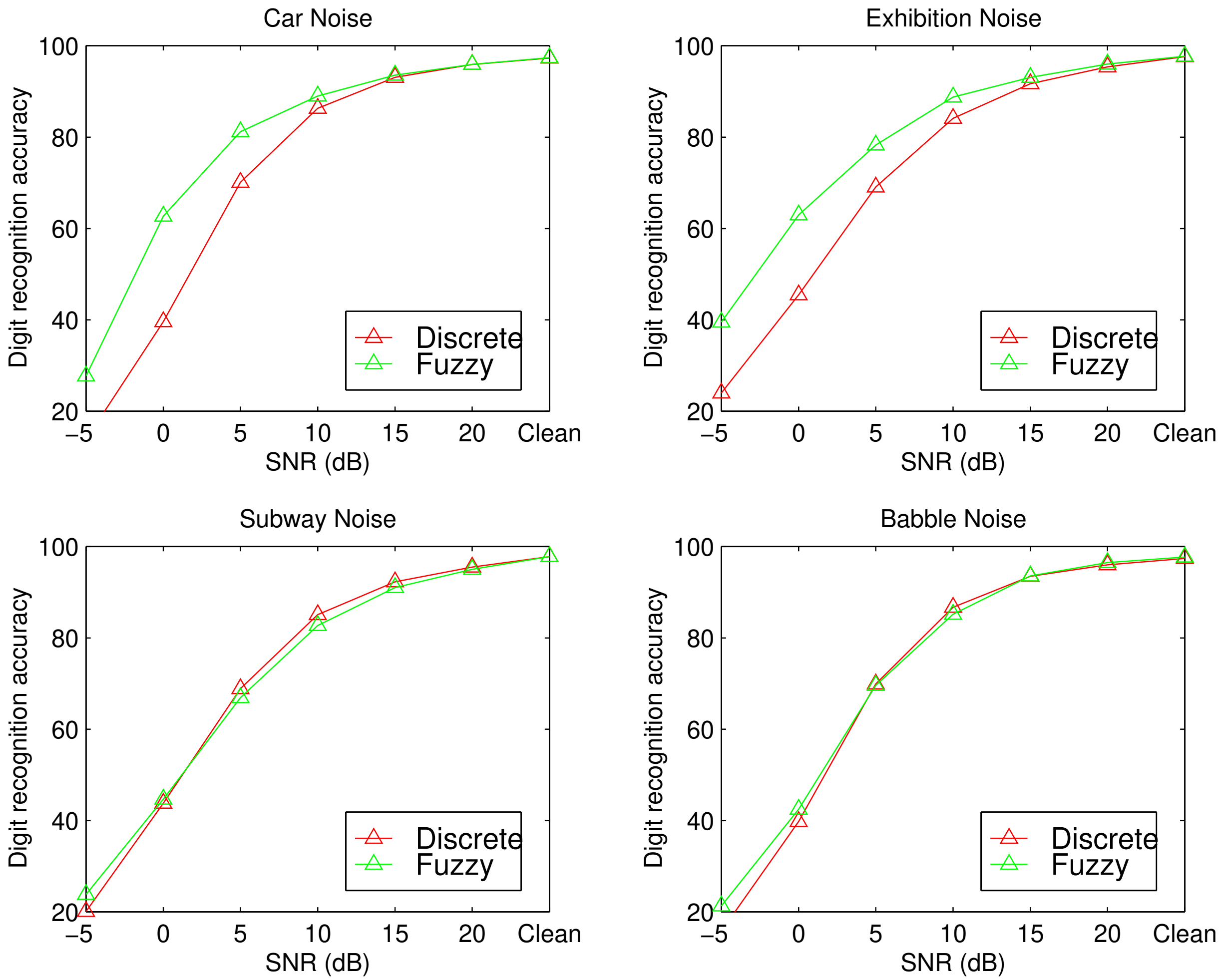
Lynx Helicopter Noise Results - Fuzzy Masks



Factory Noise Results - Fuzzy Masks



Aurora Results



(24 channel filter bank + deltas)

Conclusions and Further Work

- Hard errors in (1/0) missing data masks compromise recognition performance.
- Problem can be side stepped using a fuzzy mask.
- Have demonstrated that fuzzy masks can improve noisy TIDigit ASR at low SNRs.
- Best mapping for producing fuzzy mask is noise-type dependent.

Ideas for further work: Use noise variance estimates to derive mappings? Online adaptation of mapping? Separate mapping in each frequency channel?