Decoding Speech in the Presence of Other Sound Sources:

Integrating CASA and Missing Data ASR

Speech and Hearing Research Group,

Dept. Computer Science,

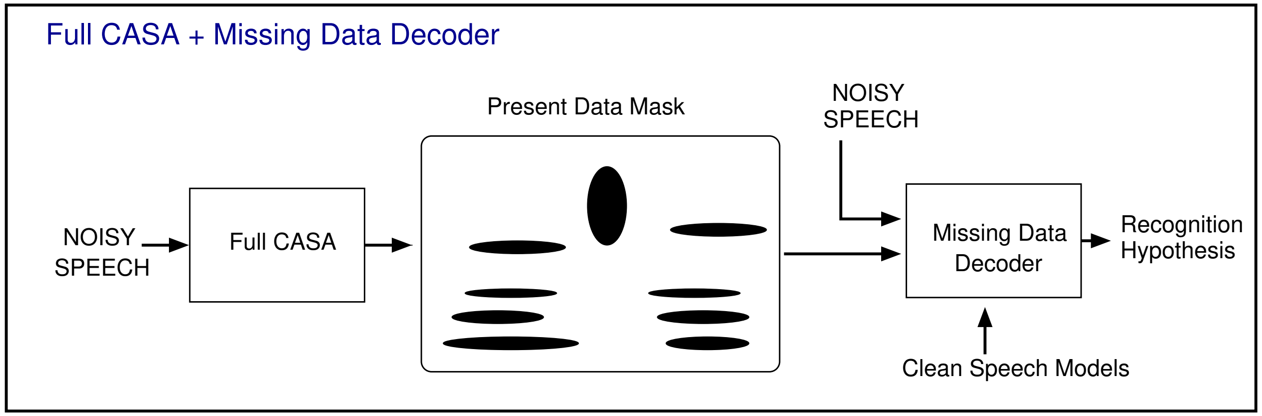
University of Sheffield, UK

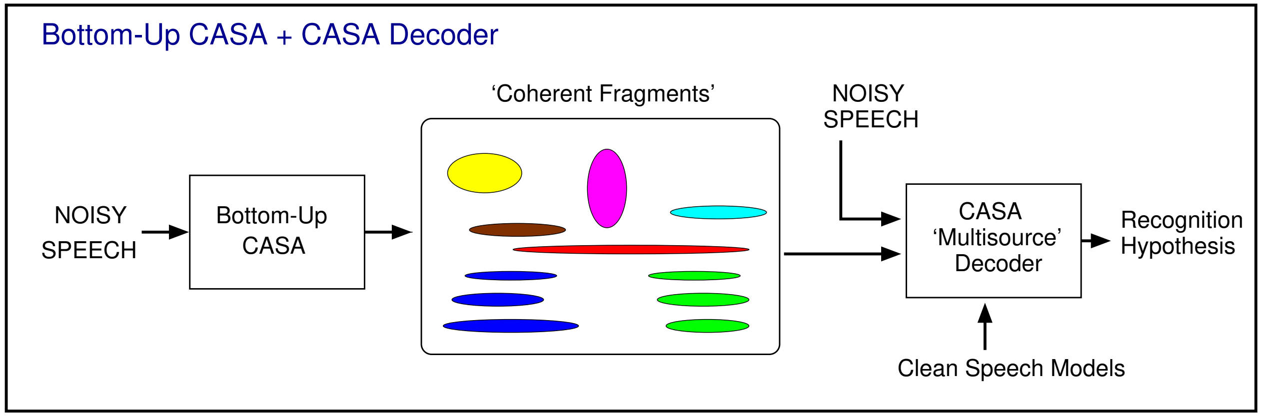
September 14, 2000

Decoding Speech in the Presence of Other Sound Sources

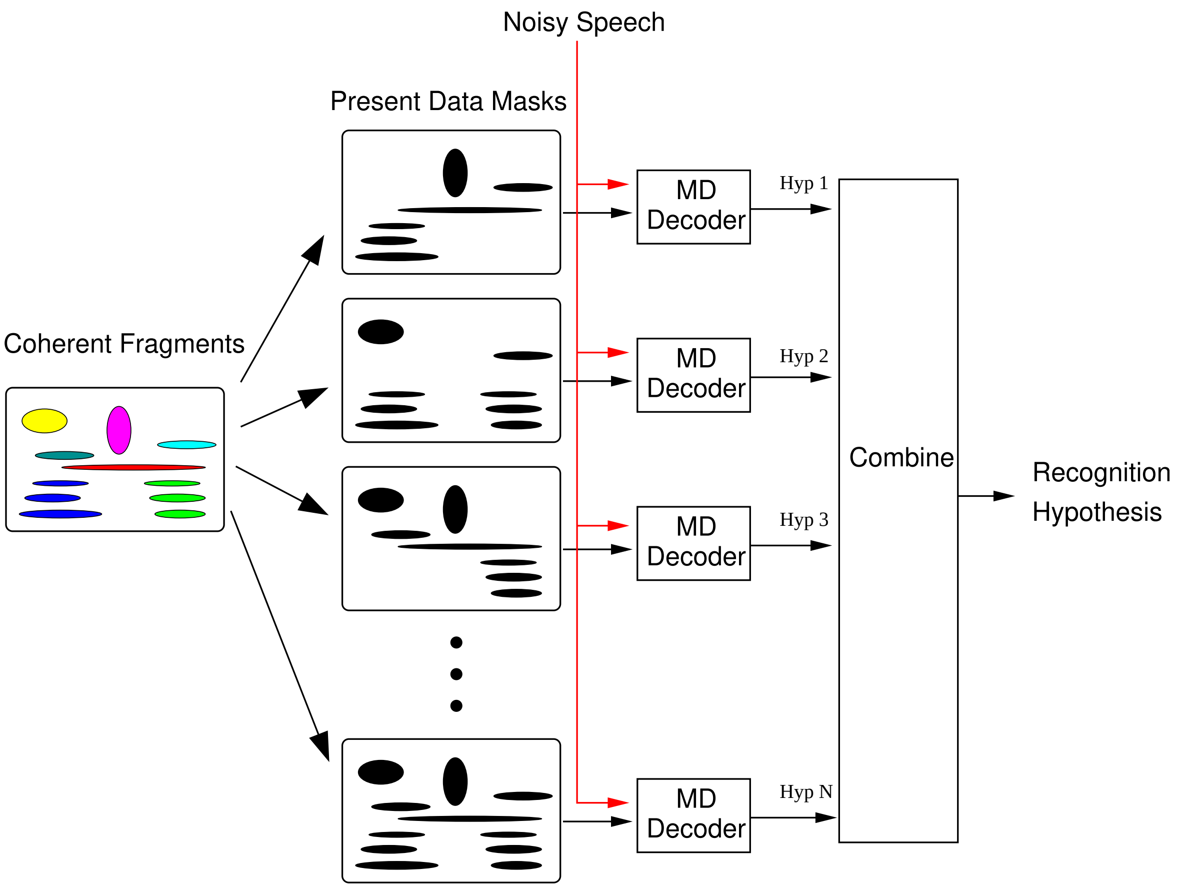
- Integrating CASA and Missing Data ASR
- The Multisource Decoder
- Some Preliminary Experiments with TIDigits
- Future Work

Integrating CASA and Missing Data ASR



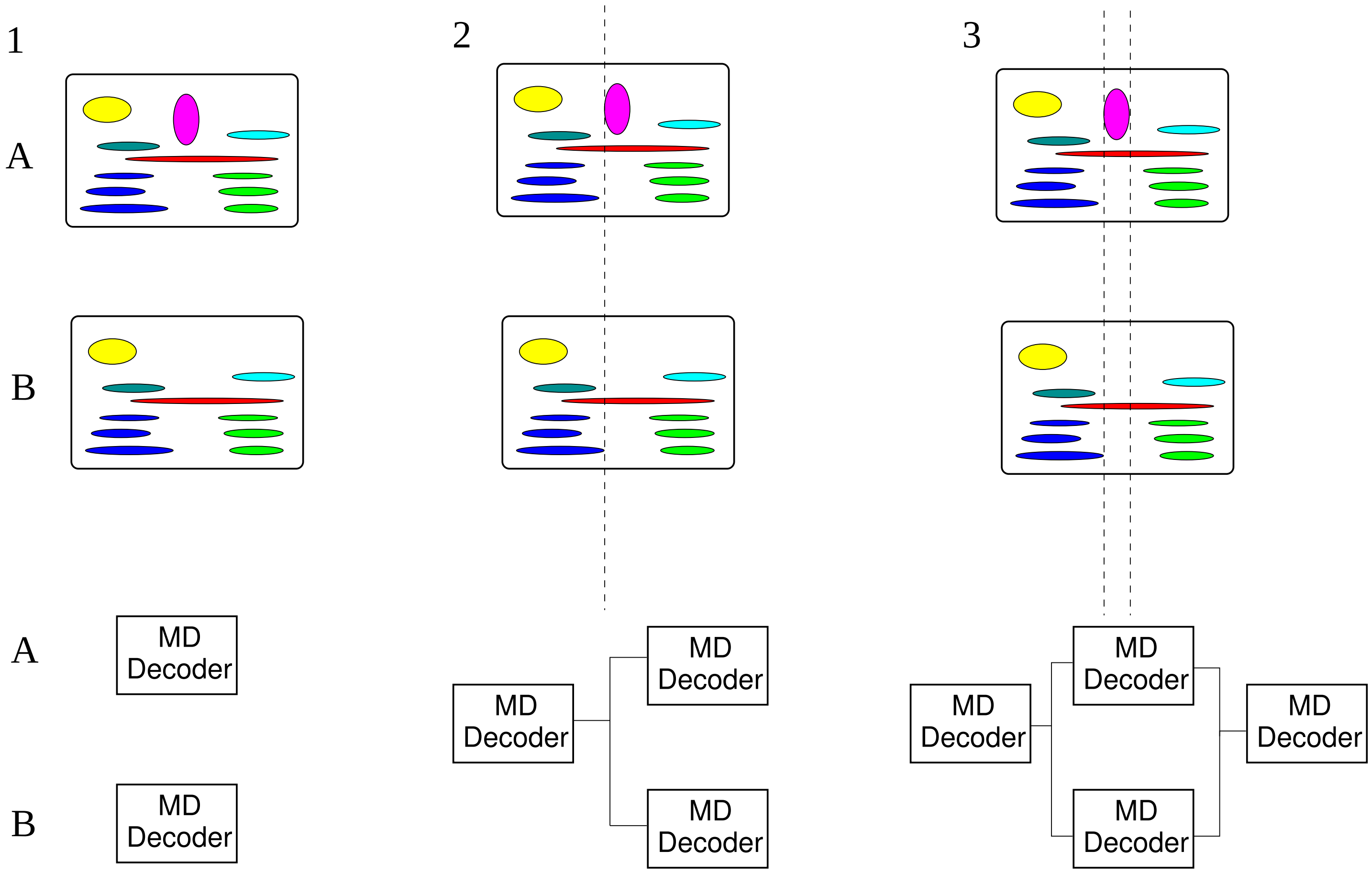


The Brute Force Approach



But ... must decode 2^N possible scenes. N can be quite large.

How can we reduce the computation required?



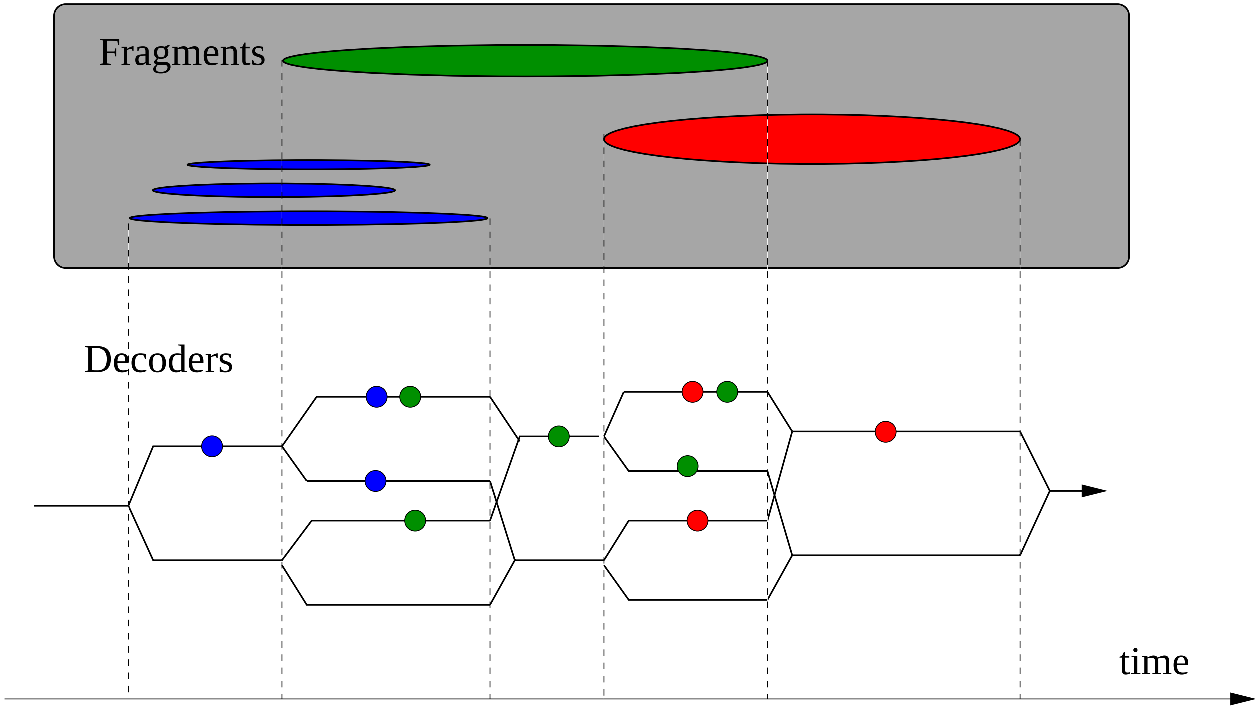
The Multisource decoder

Rather than decode all possible 2^N scene hypotheses independently, we can adapt the decoder architecture so that:

- new partial scene hypotheses are introduced (**'spawned'**) at the instant a new fragment starts.
- partial scene hypotheses are compared (**'merged'**) at the instant a fragment ends.

This reduces the number of simultaneous decodings to 2^M, where M is the maximum number of **simultaneous** fragments.

Spawning and Merging for Overlapping Fragments



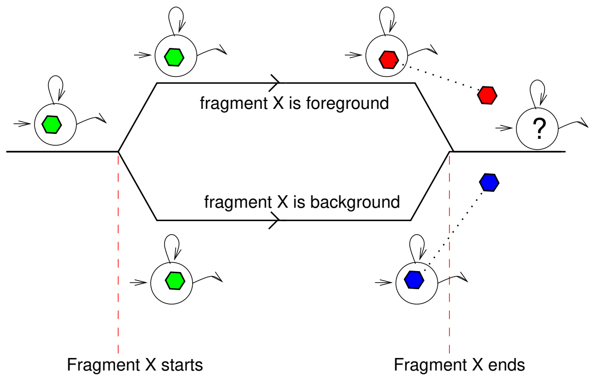
(Filled circles indicated which fragments are being considered as belonging to the speech source.)

The Multisource Decoder Implementation

Based on a standard token passing Viterbi decoder except:

- **When a new fragment starts we spawn a new decoder.**

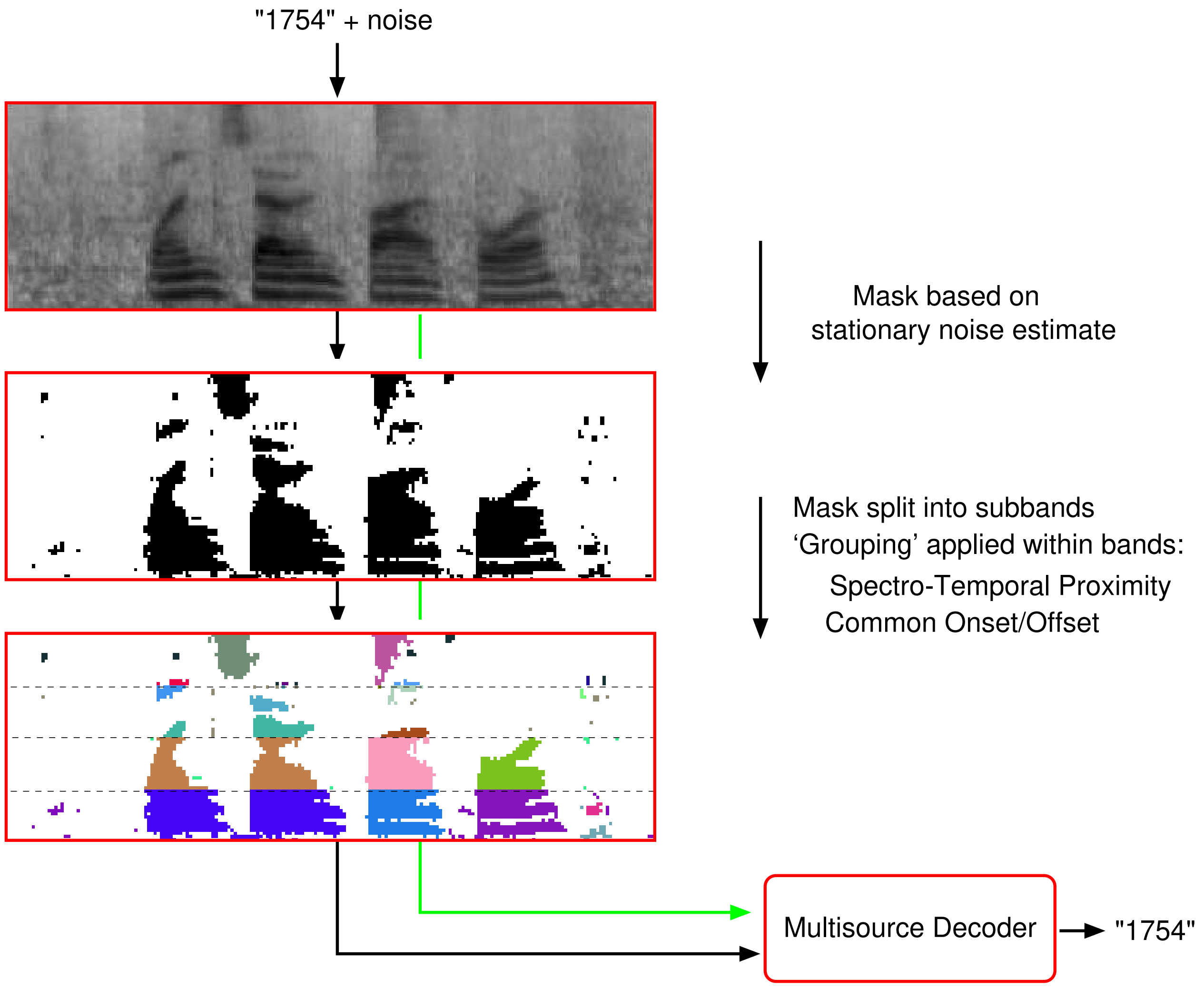
  i.e. every state and token in the existing decoder is copied.



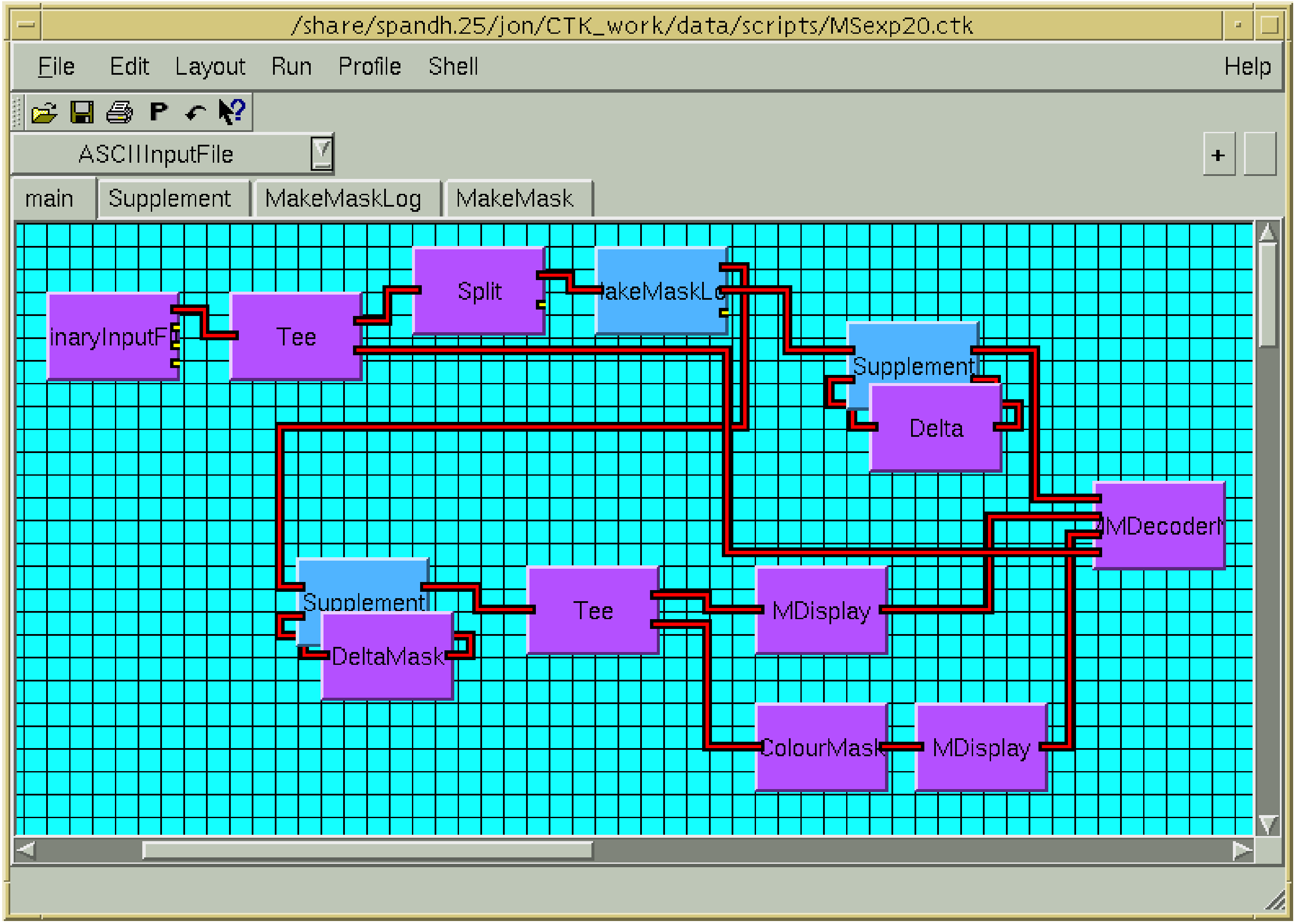
- **When a fragment ends we must merge decoders.**

  For each state, pair up the tokens in the two decoders. New decoder takes highest scoring token of each pair.

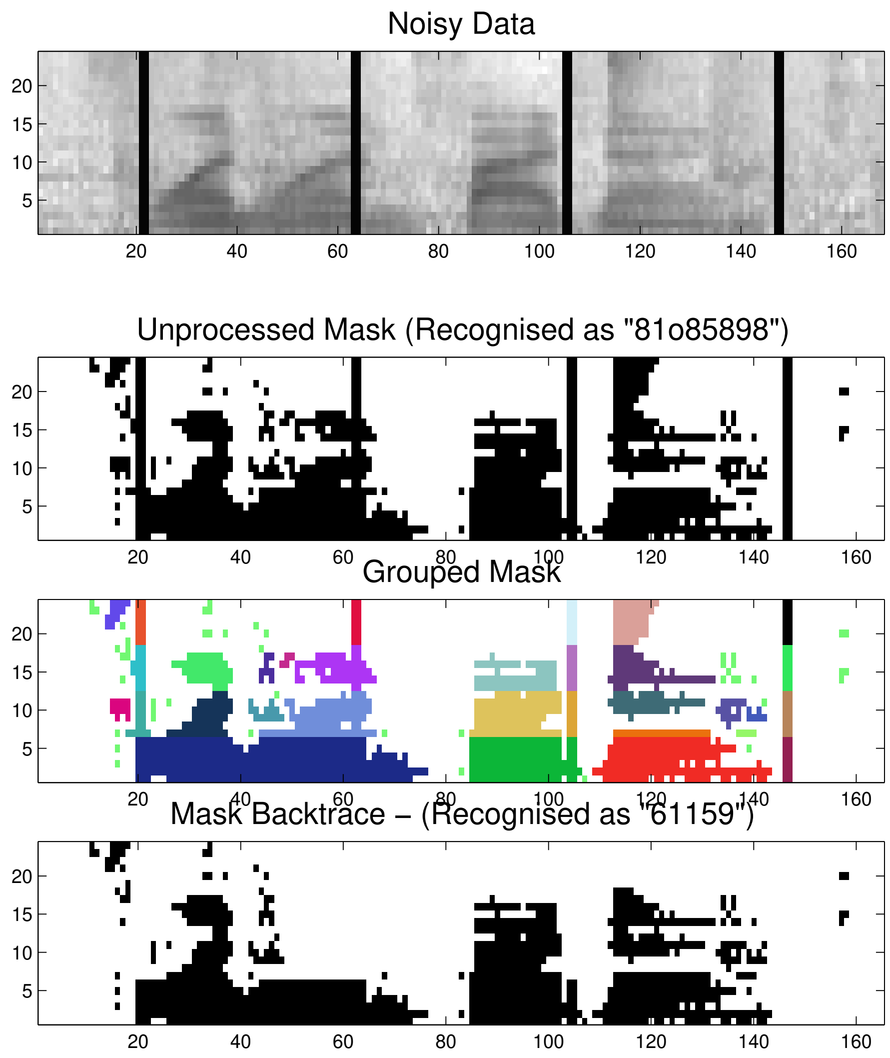
A Baseline System



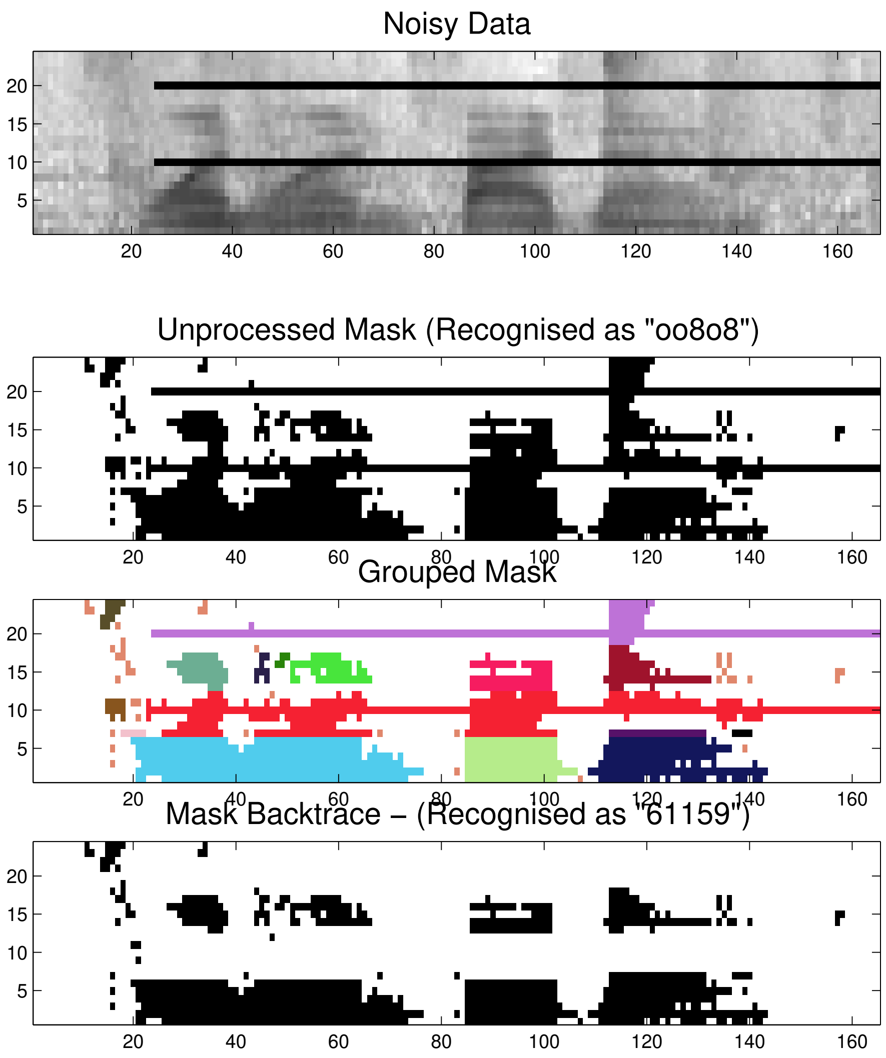
System as Implemented using the CASA Toolkit



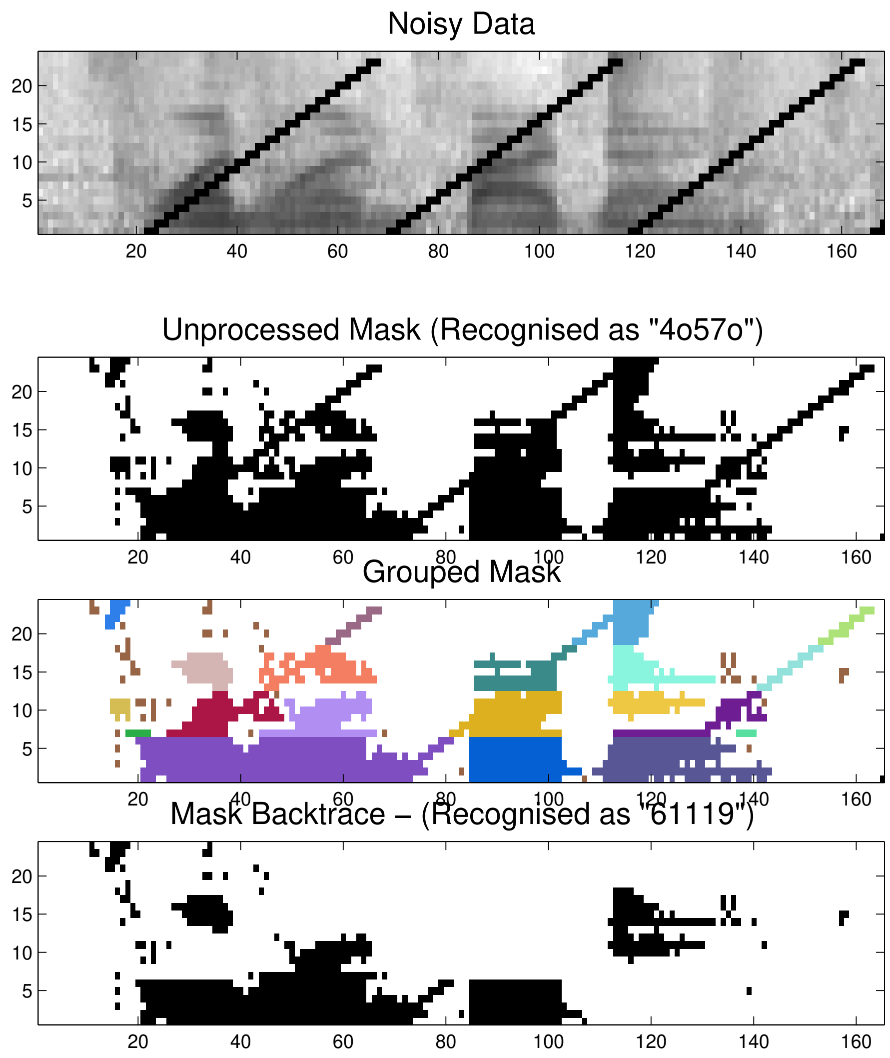
Example 1 - "1159" + broad band bursts



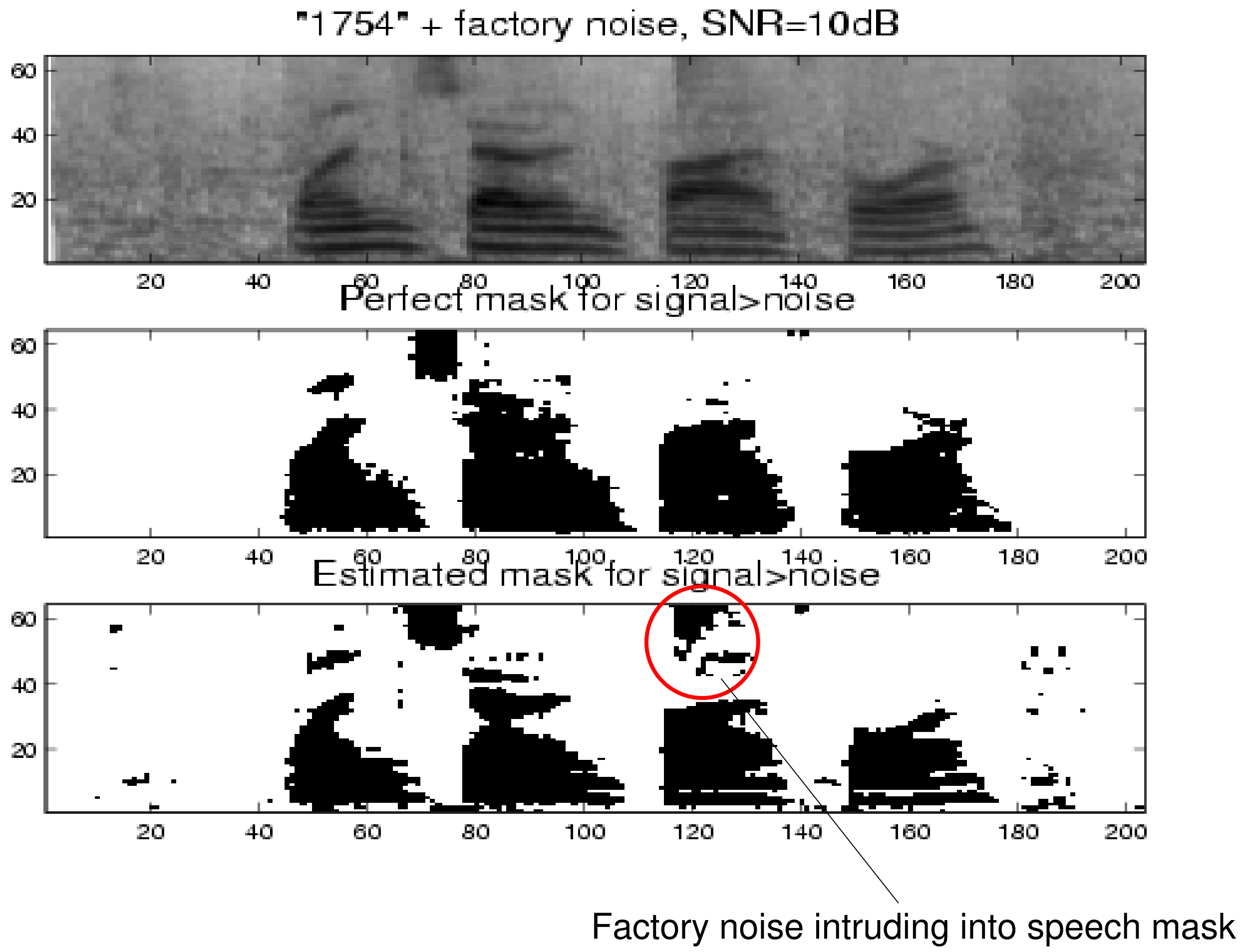
Example 2 - “1159” + unexpected tones



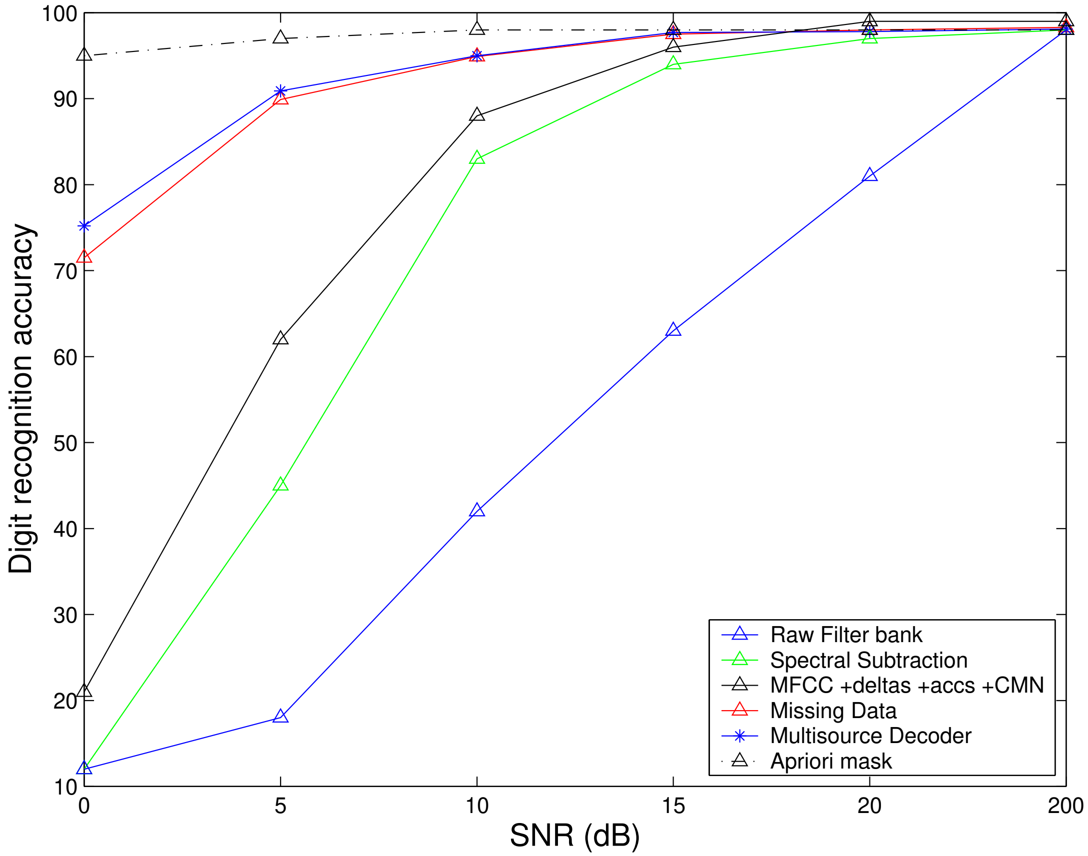
Example 3 - “1159” + chirps



‘Perfect’ versus Estimated Masks for Factory Noise

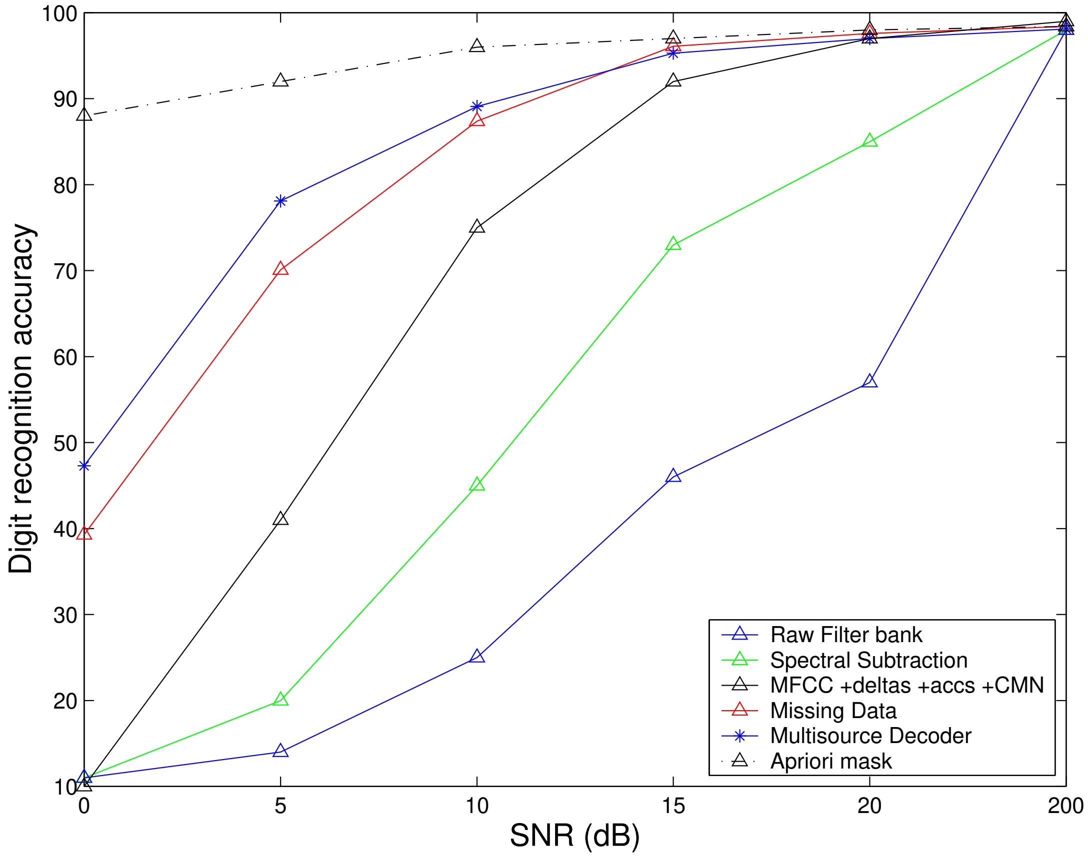


Lynx Helicopter Noise Results



(24 channel filterbank plus deltas)

Factory Noise Results



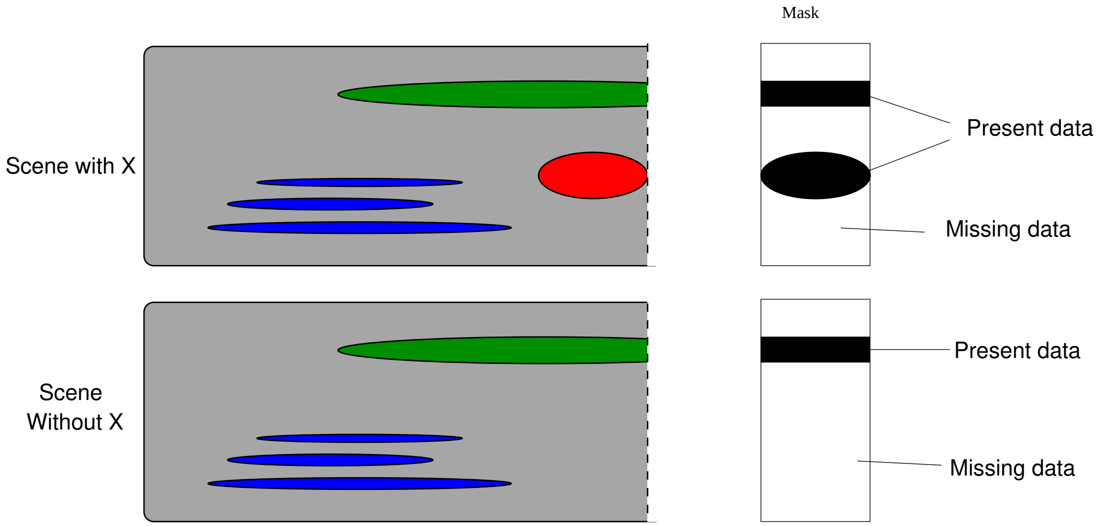
(24 channel filterbank plus deltas)

Summary

- A multisource decoder has been designed to interface between CASA and missing data ASR.
- The decoder has been applied (but without real CASA) to a noisy digit recognition.
- The techinique produces significant increases in recognition accuracy at low SNR.
- Next step - to use CASA processes to locate true coherent fragments.

The Merging Problem

The decoders being merged have 'seen' different amounts of data.



Probability calculations need 'normalising' so that missing data and present data components are comparable.

Divide by integration range to turn missing data probability into the **average likelihood of the missing component.**