Experiments with Multisource Decoding and '*A priori*' Fragments

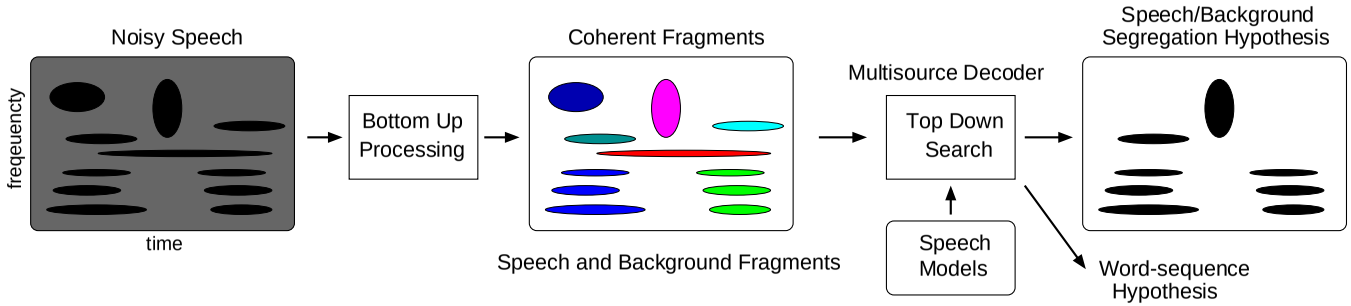
Speech and Hearing Research Group,

Dept. Computer Science,

University of Sheffield, UK

June 6, 2002

The Multisource System

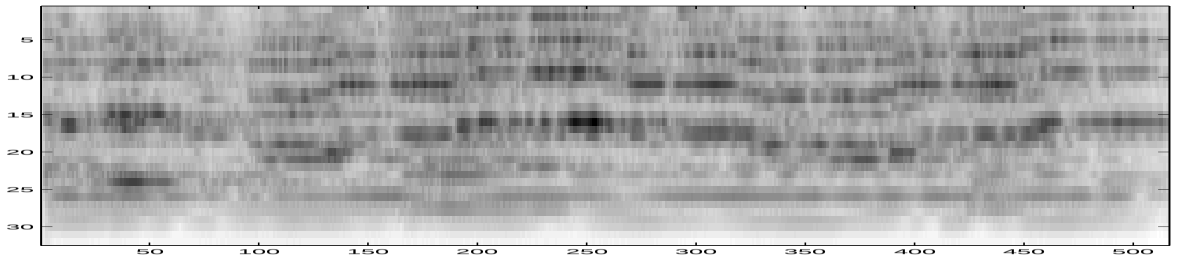


Testing issues:

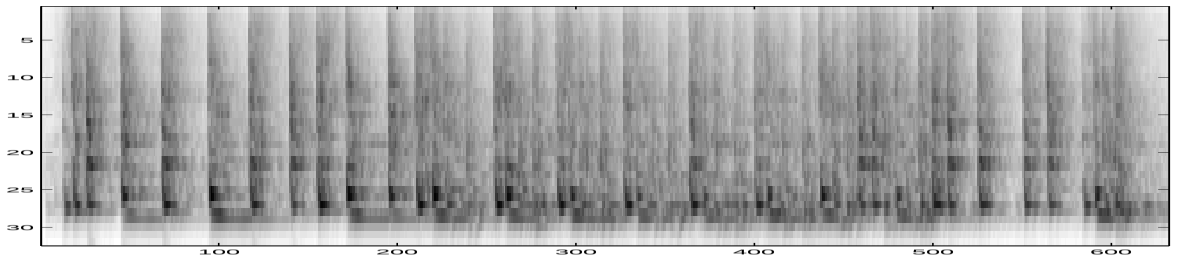
- Need highly non-stationary test data to properly test the approach
- Need a strategy for allowing back-end to be tested in isolation of front-end

The Noise Sources

- Violins



- Drums



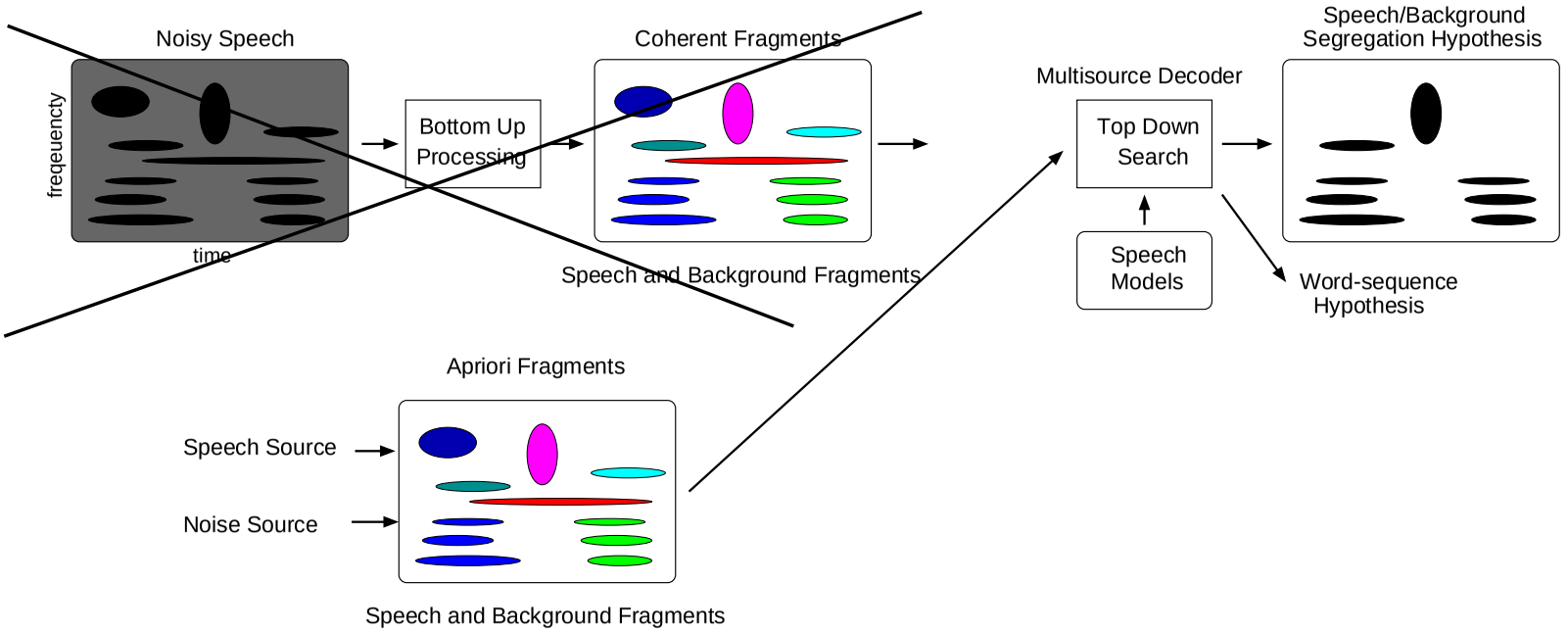
- Speech (AURORA utterances with opposing gender)

Constructing the test set

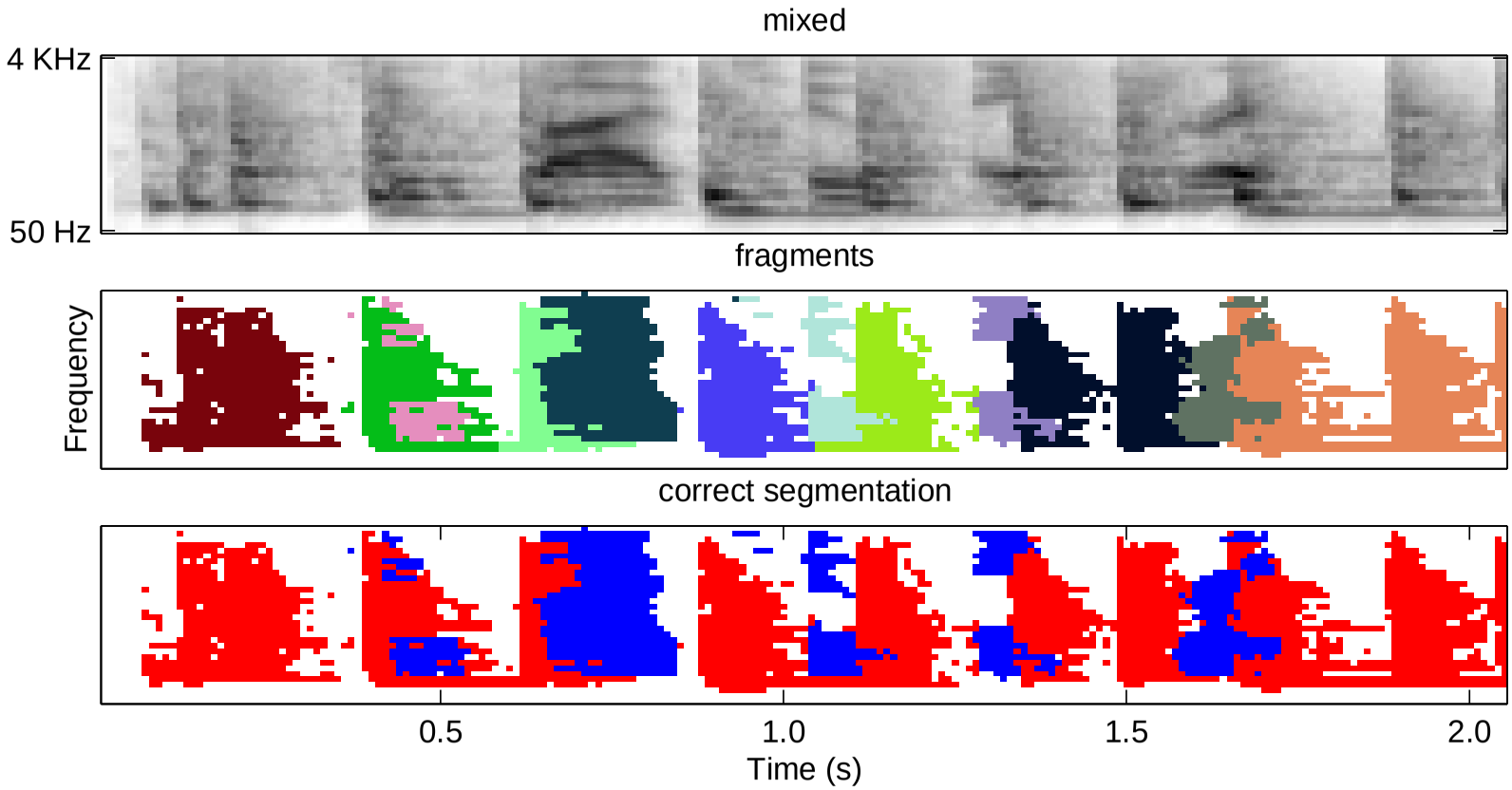
- Aurora test set A clean utterances ordered by length
- 318 M/F pairs of matched length identified
- i.e. 318 target utterances, and 318 masking utterances
- 10 second Drum and Violin extracts downsampled to 8KHz and filtered with G712 filter
- Drum and Violin masking noises for each of the 318 targets cut from the 10 second extracts
- AURORA targets + masking noise mixed so that SNR averages at 0dB during target speech

Using A Priori Test Fragments

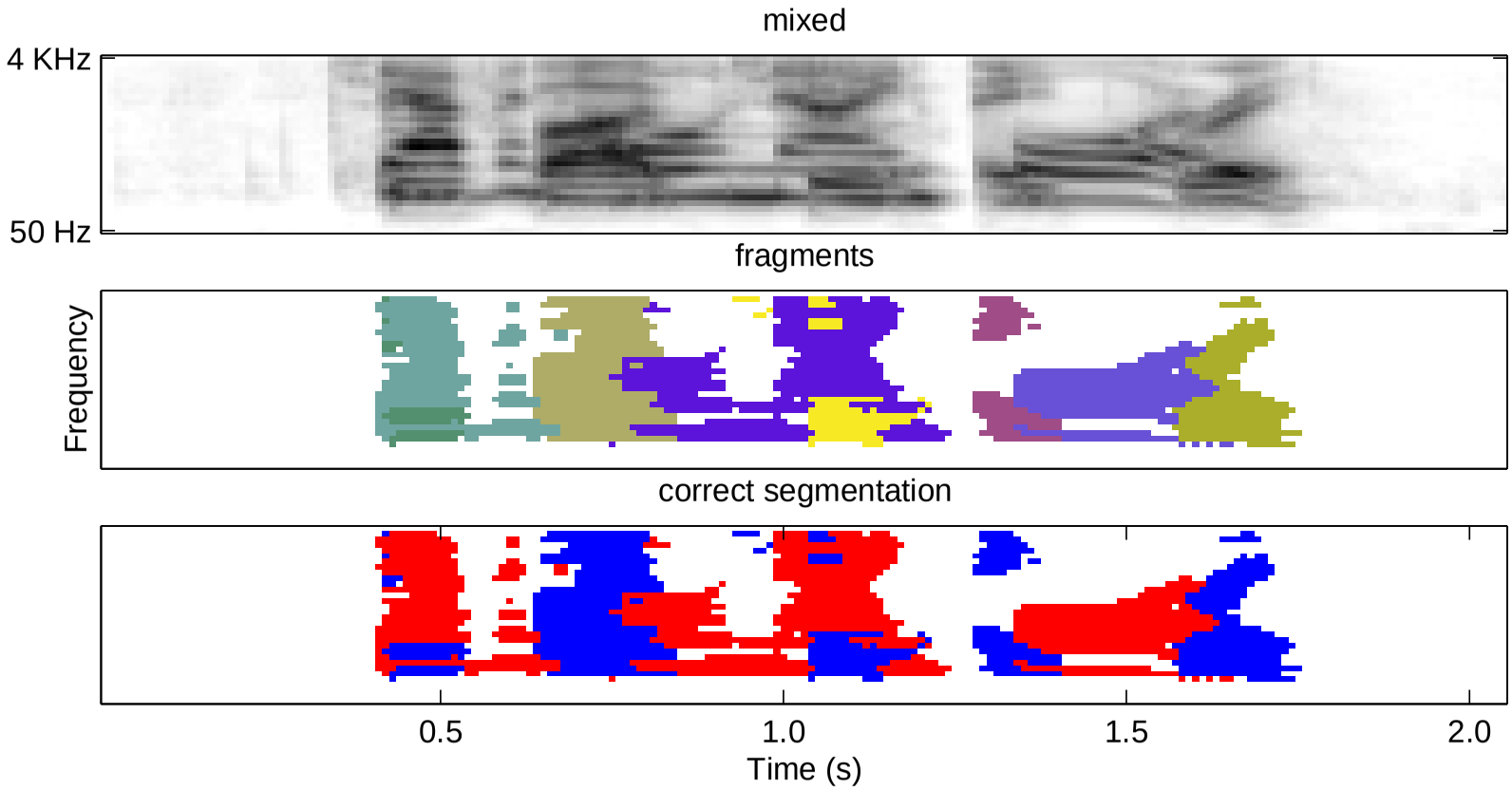
Use knowledge of signals prior to mixing to mark out a set of 'ideal' test fragments. i.e. Each fragment contains energy from either the target or the mask.



Example fragments for Speech + Drums



Example fragments for Speech + Speech

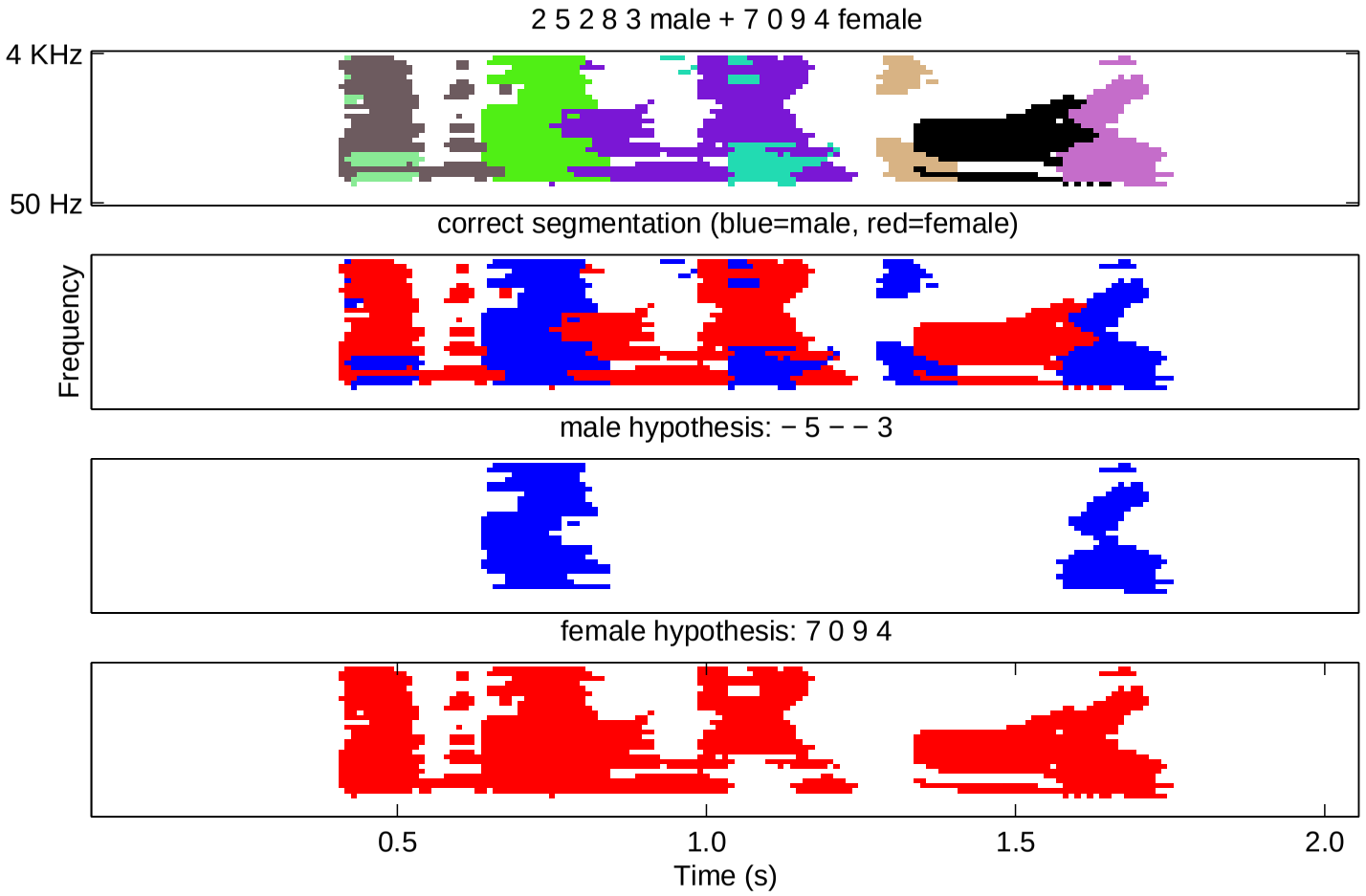


Recognition Results

	speech	violins	drums
Standard	28.6	7.9	-8.0
Soft MD	30.1	54.3	47.0
Adaptive	29.4	76.7	45.0
a priori MD	94.8	94.0	94.4
fragments	*42.4*	65.4	58.6

i.e. disappointing results - insufficient information in speech models to organise the fragments.

Gender Dependency

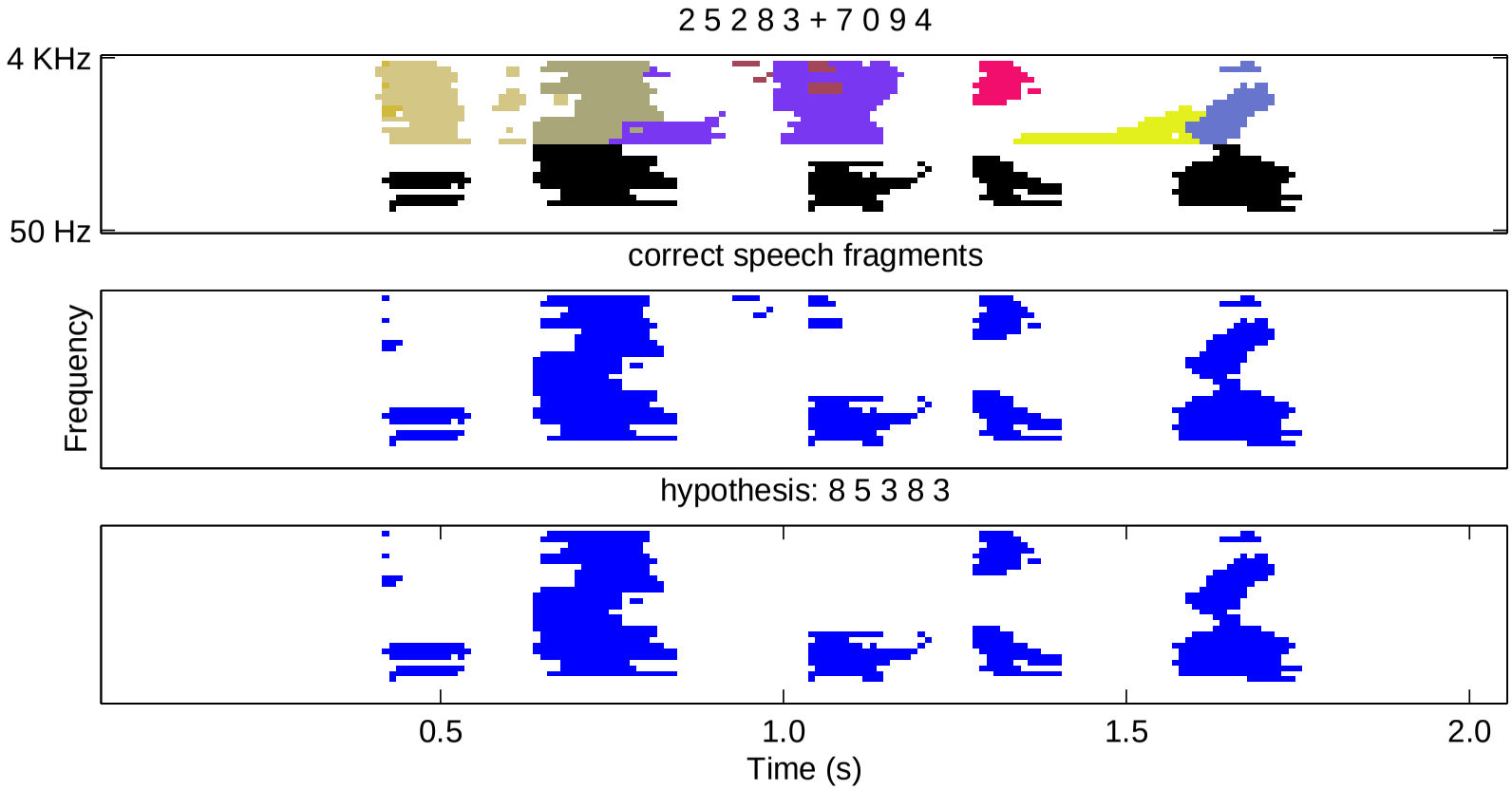


Awaiting results...

High Frequency Recruitment

The decoder is given the correct segmentation in the low frequency region.

Can it selectively recruit the correct high frequency fragments?



Hi Freq Recruitment Results

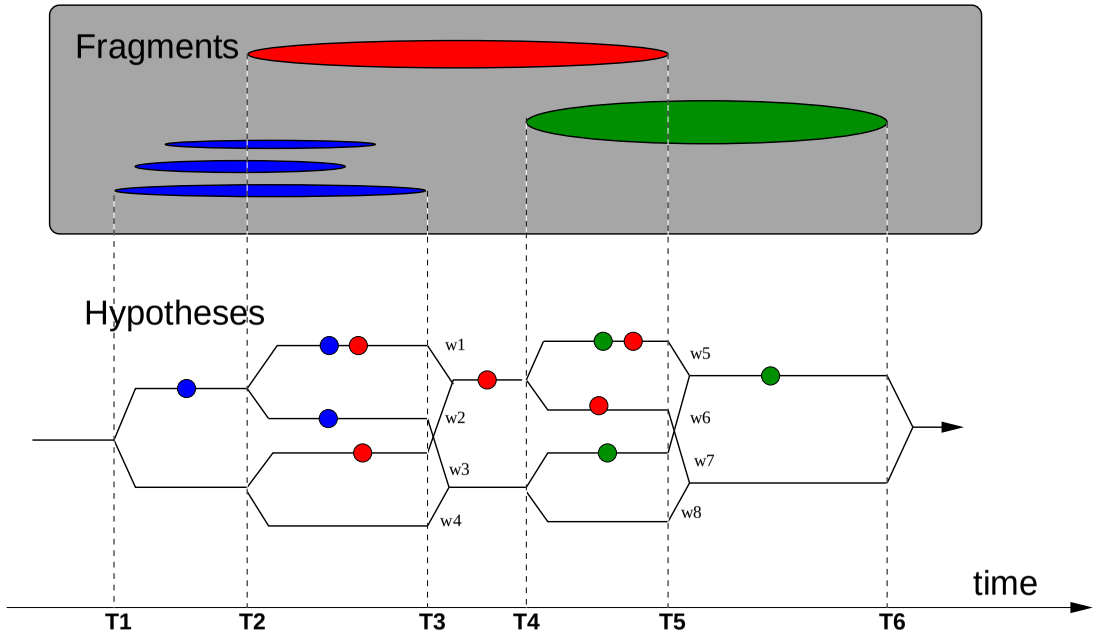
	speech	violins	drums
full apriori	94.8	94.4	94.4
low freq apriori	78.5	79.7	76.2
low freq + fragments	86.9	88.0	85.3

Kind of works... but need to check results are more than just chance.

Suggests that Multisource decoder may work if a subset of the fragments can be identified as speech prior to decoding.

Sequential Grouping

Modelling of primitive grouping forces that occur between fragments.



May be implemented by adding probabilities to decoding paths c.f. bigram/trigram language models.

Need to be careful to ensure we preserve Markov assumption. i.e. given the state the future must be independent of the past.

Work in progress...

Summary of Sheffield RESPITE work