
THISL & RESPITE progress reports

Dan Ellis

International Computer Science Institute, Berkeley CA

<dpwe@icsi.berkeley.edu>

Outline

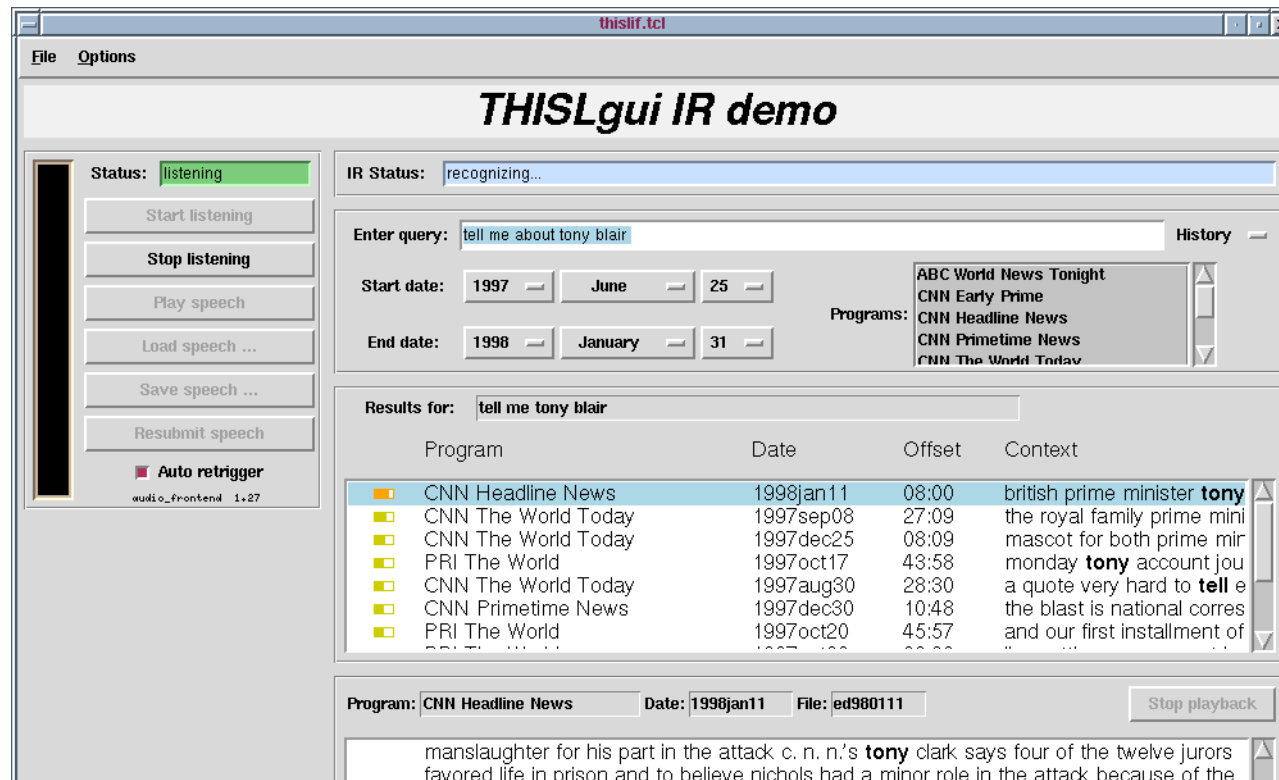
- 1 Thisl progress:
GUI, SQL evaluation, exotic data**
- 2 Respite progress:
Choosing streams, NN feature extraction,
Aurora evaluation results**



1

This!: GUI

- Automatic voice detection
 - option to trigger on any utterance

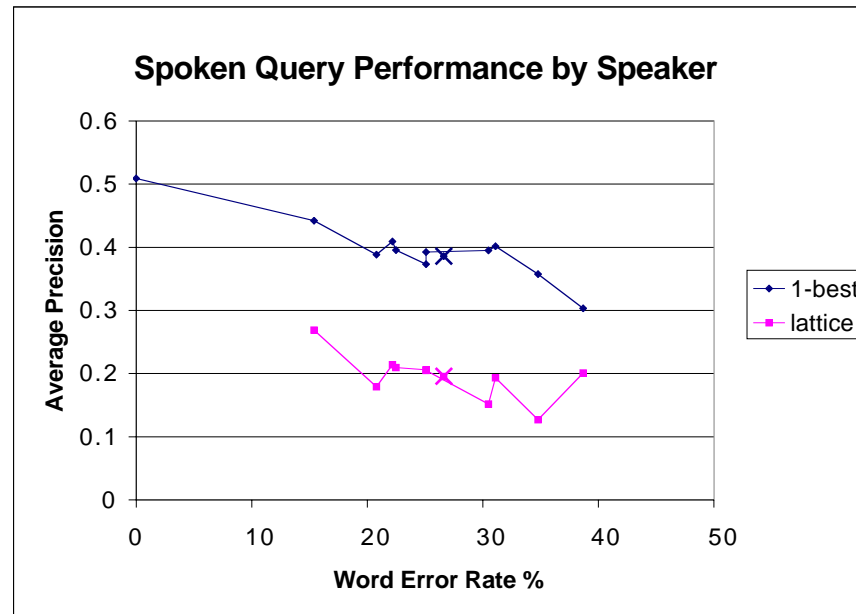


- Thomson NLP?
- Ship to Thomson/BBC?



Spoken Query Interface (SQI) Evaluation

- **Full IR processing from spoken queries**
 - 1-best hypotheses
 - all terms from (smallish) lattice (KW prec=16%)
 - (syntax not covered by Thomson NLP)



- **Spoken queries work OK**
 - dumb lattice processing does not



Exotic data evaluation

- **Range of non-news programmes from BBC**
 - natural history
 - interviews
 - features
- **Evaluation plan**
 - basic WER%: (not so bad)

<i>Data set</i>	<i>WER</i>	<i>OOV</i>	<i>Avg.Fr. Entropy</i>
Euro99Eval (6 news shows, 31k words)	29.2 ± 7.6 %	0.84%	1.14 ± 0.11
1999dec Exotic (13 varied files, 44k words)	38.9 ± 8.4 %	0.70%	1.25 ± 0.09

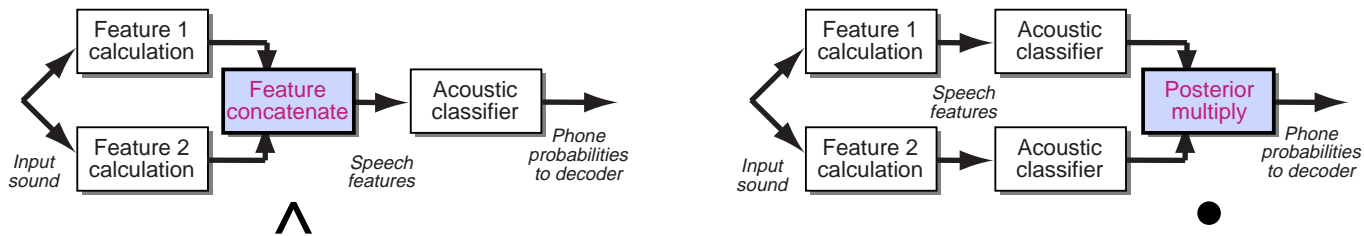
- separate acoustic & language model scores
(for recognized & forced alignments)
- Chase-style blame assignment



2

Combining feature streams

- **How to allocate feature dimensions to models?**
 - lower-dimension models train more quickly
 - higher-dimension models find more interactions



- **Variations of PLP & MSG for Aurora:**

<i>Features</i>	<i>Parameters</i>	<i>baseline WER ratio</i>
plp12•dplp12	136k	97.6%
plp12^dplp12	124k	89.6%
msg3a•msg3b	145k	101.1%
msg3a^msg3b	133k	85.8%
plp12•dplp12•msg3a•msg3b	281k	76.5%
plp12^dplp12^msg3a^msg3b	245k	74.1%
plp12^dplp12•msg3a^msg3b	257k	63.0%



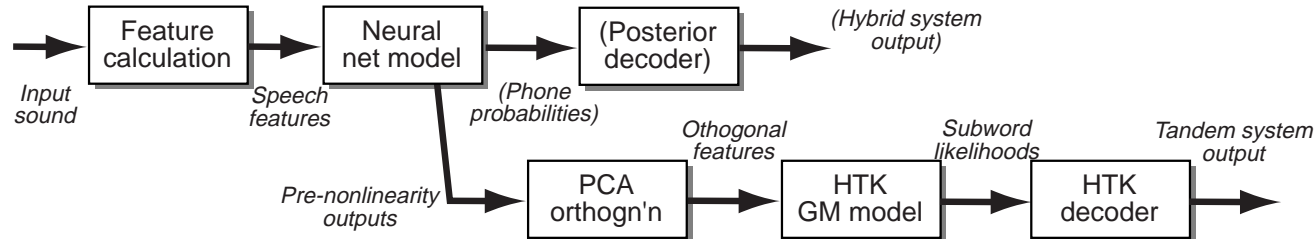
Choosing streams & combinations

- **Which combination methods?**
 - structural co-dependence is better modeled in a single feature space
 - orthogonal variability generalizes better with later combination
- **Which feature streams?**
 - best pairwise system was plp12^msg3b i.e. best single system plus worst!
 - combine streams with complementary information...
 - ... look at conditional mutual information?
 - ... of statistical model outputs?
 - ... compensating for baseline performance?



Tandem connectionist models

- **Posterior combination for HTK systems?**
- **Answer: use posteriors as HTK input features**



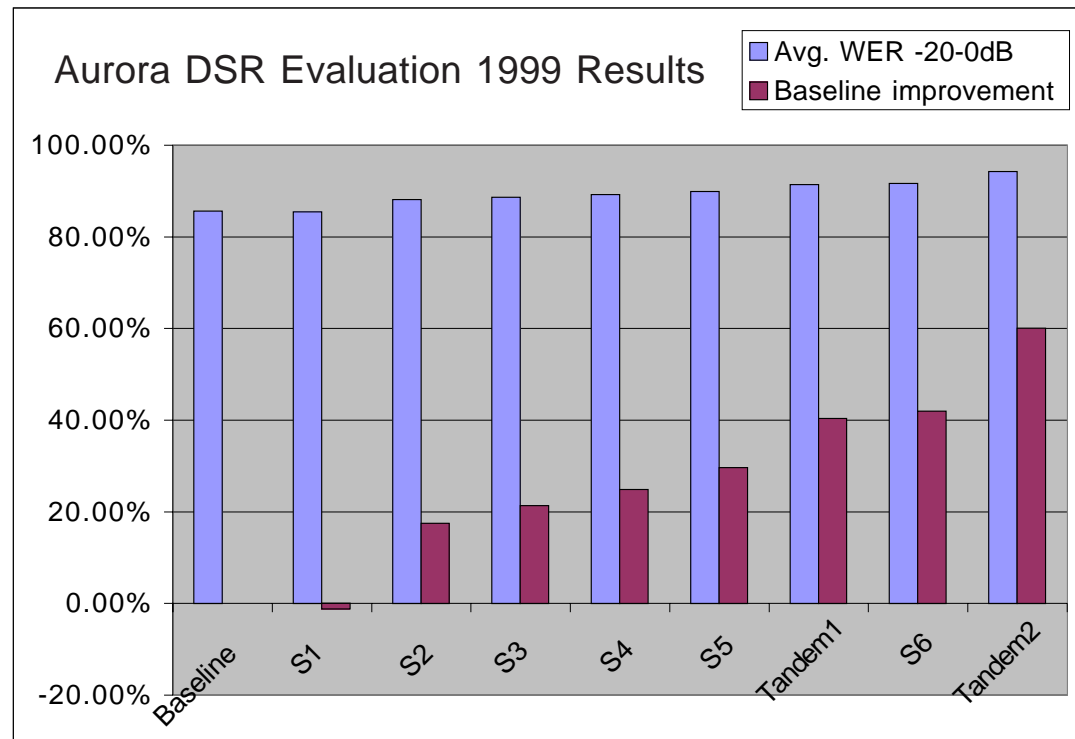
- (GMM system does not know they are phones)
- **Result: better performance than either alone!**
 - neural net has trained discriminatively
 - GMM HMMs learn context-dependent structure
 - extract complementary info from training data

<i>System-features</i>	<i>baseline WER ratio</i>
HTK-mfcc	100.0%
Hybrid-mfcc	84.6%
Tandem-mfcc	64.5%
Tandem-plp+msg	47.2%



Aurora “Distributed SR” evaluation

- 7 telecoms company submissions:



- Tandem systems from OGI-ICSI-Qualcomm
- **Best features for transmission?**
 - (filtered) subband energies may be sufficient

