NIST 2006 Speaker Recognition Evaluation

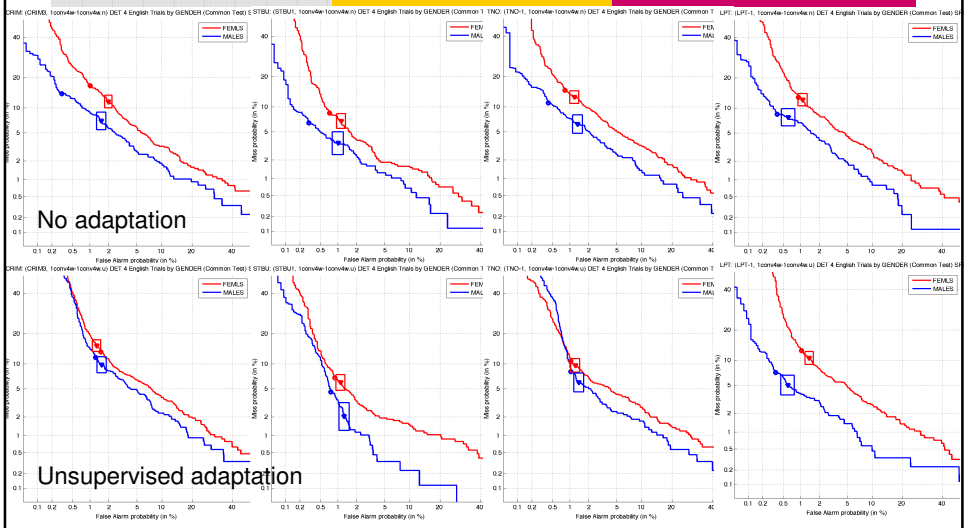
Additional Results

Alvin Martin and Audrey Le
www.nist.gov/speech/tests/spk

June 26-27th, 2006
San Juan, Puerto Rico

Performance by Sex

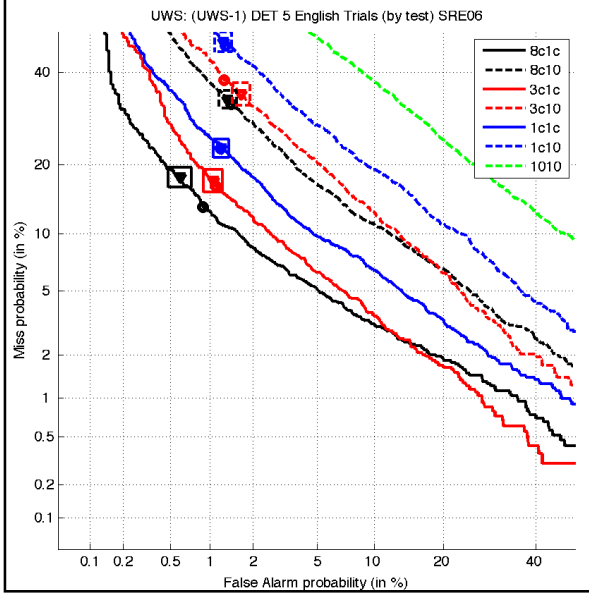
Common Condition 1conv4w-1conv4w



- Most sites had better performance on males – why?
- Unsupervised training seems to have lessened the performance advantage for males on three of the graphs shown and increased the advantage for one

Performance by Training/Test Duration

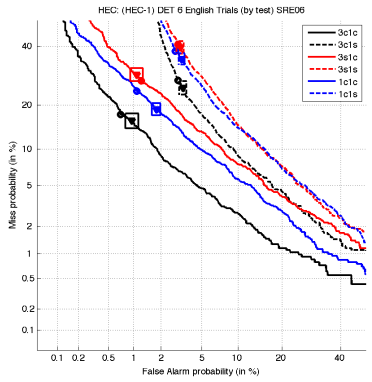
Common Condition Trials

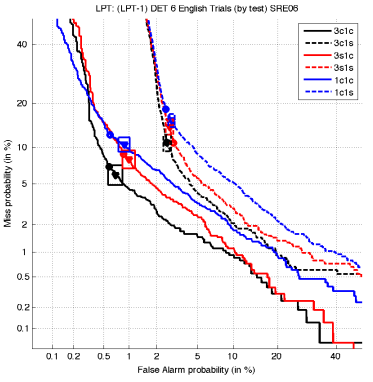


- Training could be 8conv4w, 3conv4w, 1conv4w, or 10sec4w
- Test could be 1conv4w or 10sec4w
- Performance differences were as expected
- Biggest differences occurred between 10sec and longer durations for test (or for both training and test)

Performance by Channel Condition

Common Condition Trials



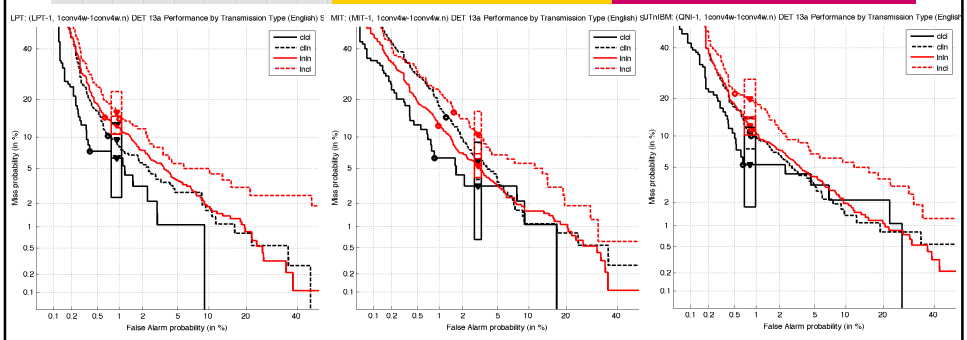


- Training could be 3conv separate, 3conv summed, or 1conv separate
- Test could be 1conv separate or 1conv summed
- Test condition had greater effect on performance than training condition
- Use of summed channel data hurt performance compared with separate channel, as expected
- Results for 3conv summed vs. 1conv separate show some variation

Performance by Transmission Type

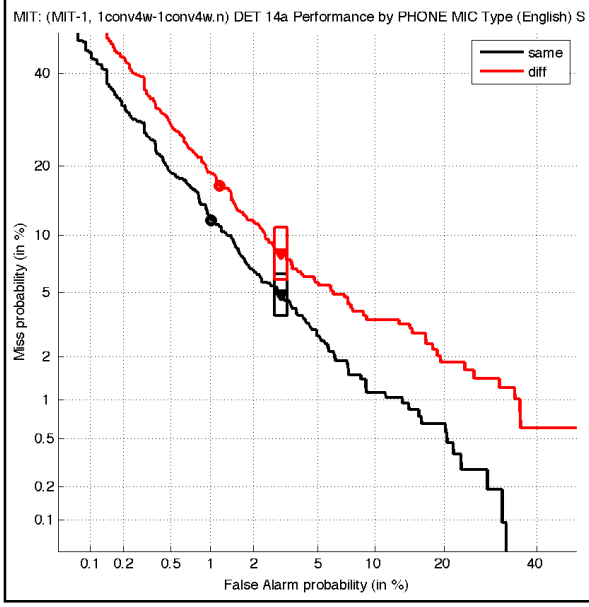
- Transmission type (reported by caller) could be
  - Cellular
  - Cordless
  - Landline
- We concatenated cordless and landline to limit the number of conditions and increase the trial counts for each condition

Performance by Transmission Type



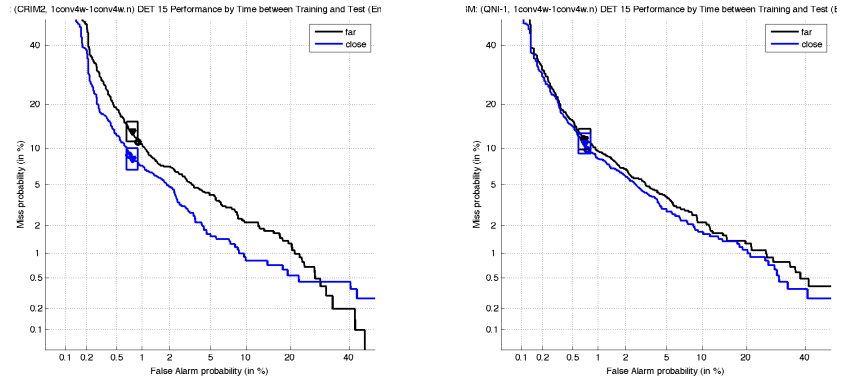
- Plots show effects of training/test being cellular/landline for common condition target trials
- Advantage for cellular over landline is surprising
- As expected, matched conditions largely do better
- Training condition appears to matter more than test

Performance by Handset Type



- Handset types (reported by callers) were
  - Speaker-phone
  - Ear-bud
  - Head-set
  - Hand-held
- Limited data for many types made performance comparison difficult
- Here we simply examine same vs. different type between training and test for target trials in the common condition
- Results are no surprise – having the same type of handset in both training and test helps performance

Time Between Training and Test



- Last year performance was found to vary greatly as a function of the time between the training and test recordings of target trials. This seemed quite unreasonable. LDC indicated no procedural might account for it. It remains a mystery
- This year the performance differences were small, as expected
- Target trials are divided between those where the recording interval exceeded, or not, 5 days