Some aspects of the ICSI 1998 Broadcast News effort

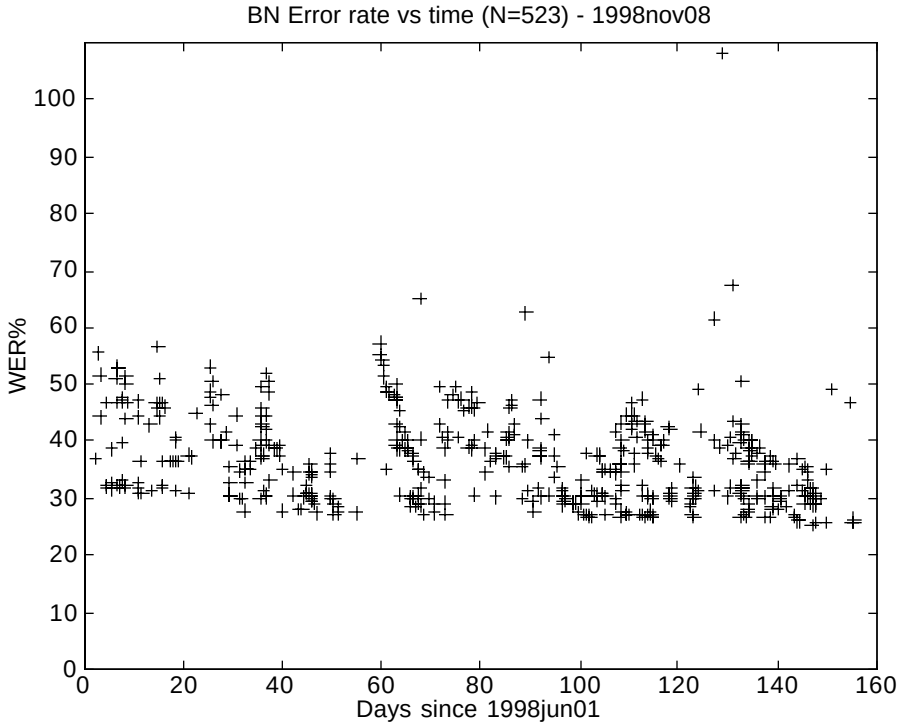
Dan Ellis
International Computer Science Institute, Berkeley CA
<dpwe@icsi.berkeley.edu>

Outline

1. **Overview**
2. **Feature choice**
3. **Net size and training size**
4. **Whole-utterance features**
5. **Gender-dependence**

Overview

- **6 months, scores of trainings, 100s of decodes**



- **Variables:**

  feature, netsize, trnset, labels, testset, pruning, phone models, lexicon, LM, ...

Feature choice

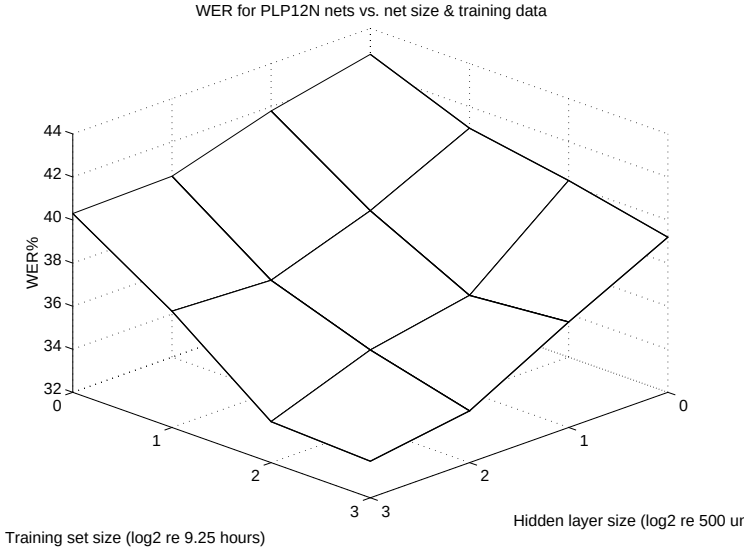
- **Cambridge: normalized PLP**
- **Band-limited to 4kHz for 'difference'**
- **Rasta performed poorly (16ms windows)**
- **Search over deltas, context window**
- **plp12N-8k best alone**
- **msg1N-8k best for combination with RNN**

Feature	Elements	WER% alone	WER% RNN combo
RNN baseline			33.2
plp12N	13	36.7	31.1
ras12+dN	26	44.4	32.5
msg1N	28	39.4	29.9

(2000HU, 37h trainset, align2 labels, 7hyp decode)

Net and training set: Size matters

- **Huge amount of training data available**
  - 74h (16M training patterns @ 16ms) → 142h
- **Search over net size / training set size**



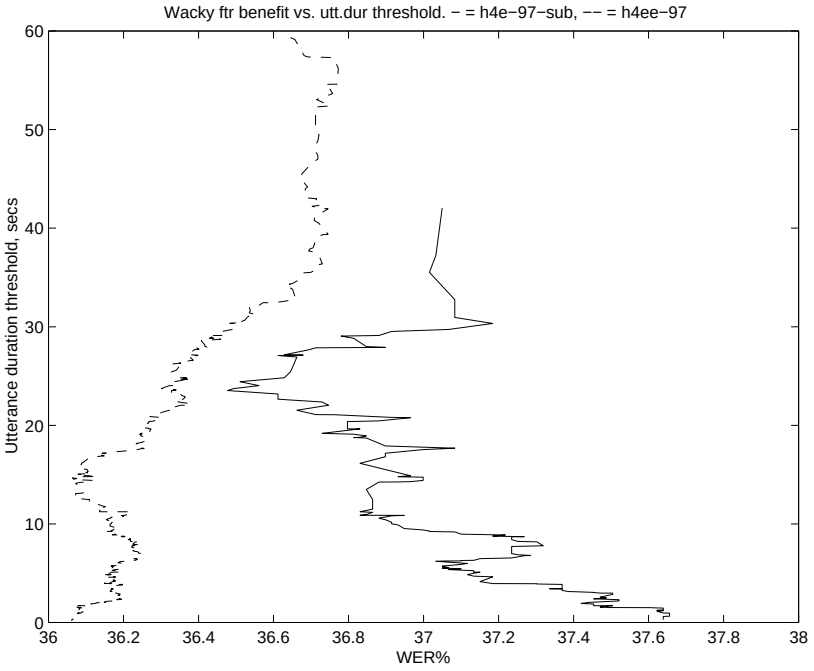
HUs	Trnset	Labels	Trn time	WER%	ComboWER %
2000	37h	align2	4 days	38.6	30.4
4000	74h	align2	7 days	35.3	29.3
8000	142h	align4	21 days	31.6	26.8

(msg1N, 7hyp decode)

Whole-utterance features

- **BN groups have focussed on adaptation & normalization**
  - VTLN, MLLR, SAT
- **Maybe do similar thing with extra net inputs**

→ **Whole-utterance pre-normalization feature-dimension variances as constant inputs to net**



Gender dependence (GD)

- **Train separate nets on Female/Male data**
  - males represented 2:1 in BN
  - oracle labels?
- **Encouraging results:**

Net	F% (2224)	M% (3711)	WER%(5938)
2000HU/25h UF	28.3	53.1	43.8
2000HU/25h UM	42.1	33.3	36.6
4000HU/50h U	29.6	33.6	32.1
Oracle best	25.7	30.5	28.7
Combination scheme	27.7	32.2	30.5

- **Best practical scheme**
  - use classifier entropy to choose M or F net
  - use decoder likelihood to choose GD or GI

Eval results

Group	Full 1	Full 2	10x 1	10x 2
BBN	15.2	14.2	17.5	16.7
CU-HTK	14.2	13.5	16.8	15.5
Dragon	15.7	13.4	18.0	15.5
IBM	14.5	12.4	21.2	17.8
LIMSI	13.8	13.3		
OGI/Fonix	27.9	23.6		
Philips/RWTH	18.5	16.8		
SPRACH	21.7	20.0	26.2	23.8
SRI	22.1	20.1	23.4	22.2