Tandem modeling investigations

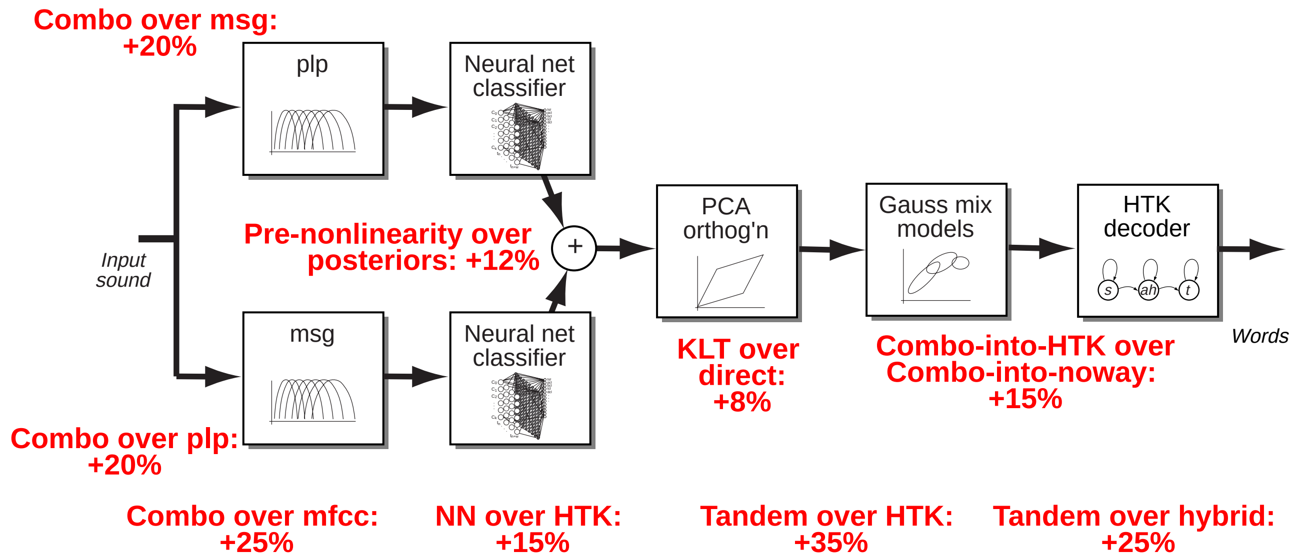
Dan Ellis
International Computer Science Institute, Berkeley CA
<dpwe@icsi.berkeley.edu>

Outline

1. **What makes Tandem successful?**
2. **Can we make Tandem better?**
3. **Does Tandem work with LVCSR tricks?**

1 What makes Tandem work?

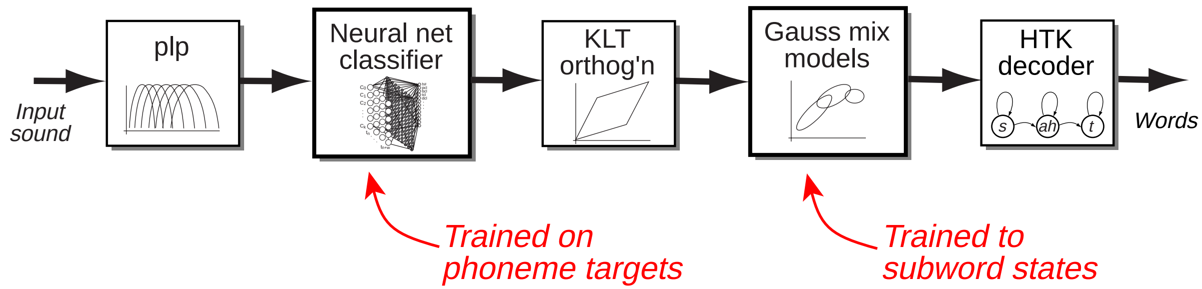
(with Manuel Reyes)



Tandem combo over HTK mfcc baseline: +53%

- **Model diversity?**
  - try a phone-based GMM model
  - try training the NN model to HTK state labels
- **Discriminative network training?**
  - (try posteriors from GMM & Bayes)

Phone vs. word models



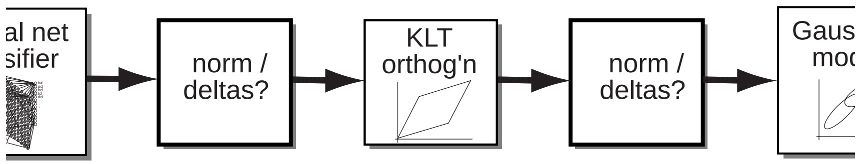
- **Try a phone-based HTK model (instead of whole-word models)**
- **Try training NN model to subword-state labels**
  - 181 net outputs; reduce to 40 in KLT
- **Results (Aurora2k, HTK-baseline WER ratio):**

System	*test A: matched*	*test B: var noise*	*test C: var chan*
Tandem PLP baseline	63.5%	70l.3%	59.5%
Phone-based HTK sys	63.6%	72.5%	61.5%
Subword-based NN sys	63.1%	**62.8%**	**55.1%**

- **Diversity doesn't help**
  - subword units may be good for NN

2 Enhancements to Tandem-Aurora

- **More tandem-feature-domain processing:**



- **Results (HTK baseline WER ratio):**

System	*test A: matched*	*test B: var noise*	*test C: var chan*
PLP: Tandem baseline	63.5%	70l.3%	59.5%
PLP: norm - KLT	72.6%	71.2%	63.6%
PLP: KLT - norm	57.8%	58.8%	51.3%
PLP: KLT - delta	59.0%	60.2%	52.9%
PLP: KLT - delta - norm	58.1%	59.9%	48.9%
PLP: delta - KLT - norm	**54.7%**	**53.6%**	**46.9%**

  - delta-KLT-norm: 80% Tdm baseline WER

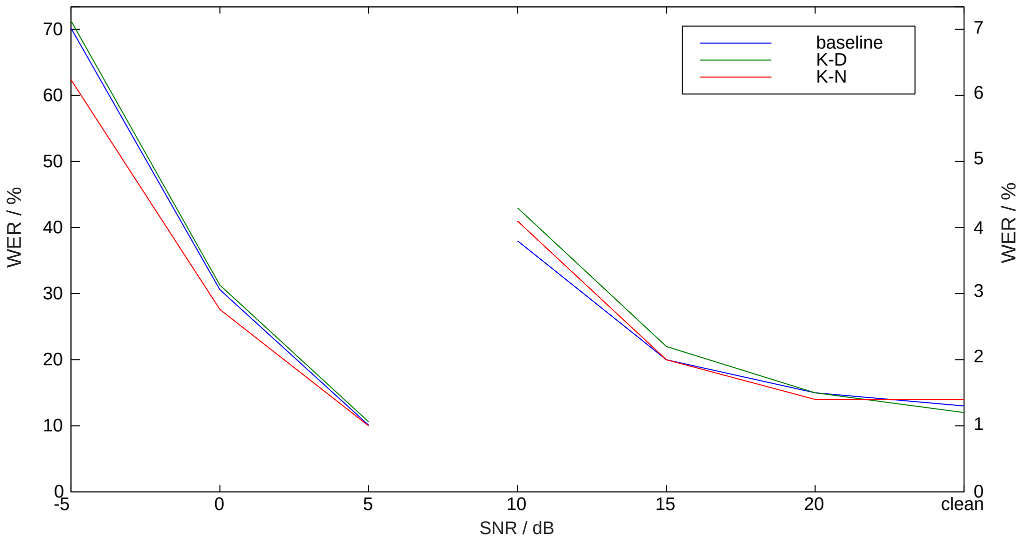
Best effort Tandem system

- **Deltas & norms help PLP:**
  try on combo (PLP+MSG) system:

System	*test A: matched*	*test B: var noise*	*test C: var chan*
PLP+MSG: baseline	51.1%	52.0%	45.6%
PLP+MSG: dlt-KLT-nrm	50.9%	50.5%	43.6%
PLP+MSG: KLT-nrm	**48.3%**	**49.5%**	**39.4%**

  - deltas *hurt* for MSG: features too sluggish?

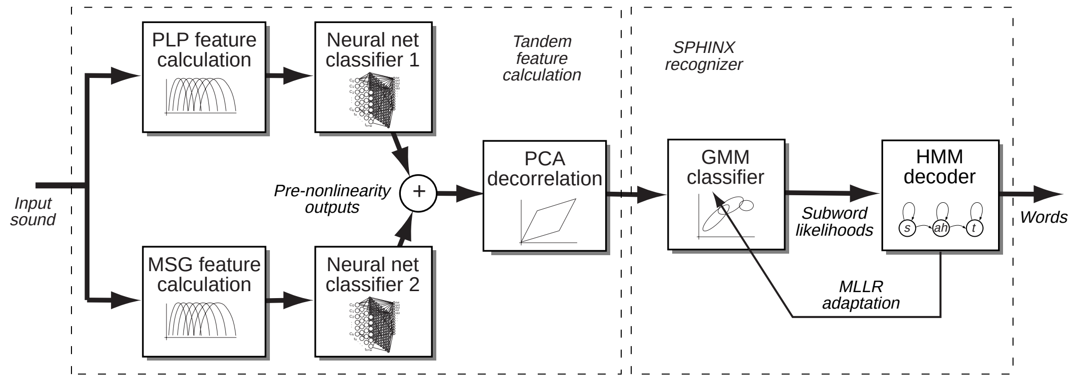
- **Deltas help clean, norms help noisy:**



Tandem for LVCSR: the SPINE task

(with Rita Singh/CMU & Sunil Sivadas/OGI)

- **Noisy spontaneous speech, ~5000 word vocab**
- **Recognition:**



  - same tandem features
  - NN training from Broadcast News boot + iterate
  - GMM-HMM has context-dependence, MLLR

SPINE-Tandem results

- **Evaluation WER results:**

Features (dimensions)	*CI system*	*CD system*	*CD + MLLR*
MFCC + d + dd (39)	69.5%	35.1%	33.5%
Tandem features (56)	47.6%	35.7%	32.8%

  - much better for CI systems
  - differences evaporate with CD, MLLR

- **Not quite fair:**
  - CD senones optimized for MFCC
  - worth 2-3% absolute?

- **Not unexpected:**
  - NN confounds CD variants
  - Tandem ‘space’ very nonlinear - bad for MLLR

- **Any hope?**
  - more training data / train CD classes / ...