# RESPITE: Tandem & multistream research

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

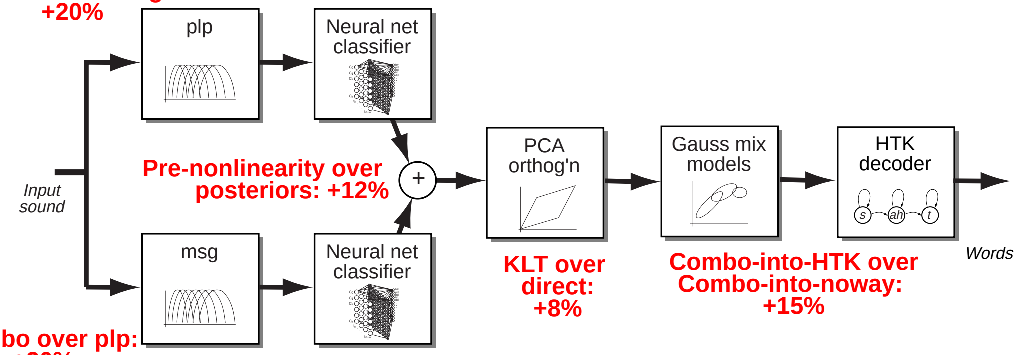
## Outline

1. **Tandem & LVCSR**
2. **Mutual information for multistream design**
3. **Other multistream work at ICSI**
4. **Other projects:**
   - **Meeting recorder**
   - **LabROSA**
   - **CRAC workshop**

1

# Recent Tandem work

Combo over msg: +20%



Input sound

plp

msg

Neural net classifier

Neural net classifier

Pre-nonlinearity over posteriors: +12%

+

PCA orthog'n

Gauss mix models

HTK decoder

Words

KLT over direct: +8%

Combo-into-HTK over Combo-into-noway: +15%

Combo over plp: +20%

Combo over mfcc: +25%

NN over HTK: +15%

Tandem over HTK: +35%

Tandem over hybrid: +25%

Tandem combo over HTK mfcc baseline: +53%

- **Aurora 2000 (mismatched test conditions)**
  - normalization much more important: online?
  - baseline WER ratio (smaller is better):

| *System* | *Matched test* | *Medium mismatch* | *High mismatch* |
|---|---|---|---|
| plp, utt-norm | 78% | 69% | 63% |
| tandem, utt-norm | 63% | 73% | 52% |
| tandem, onl-norm | 74% | 81% | 64% |

(Pratibha Jain, OGI)

# Tandem for LVCSR

- **DARPA SPINE task (spont. noisy) (e.g.)**
- **Collaboration with OGI & CMU**
  - tandem needs GMM-HMM expertise!
- **Tight timescale**
  - Tandem system not optimized, one stream
- **Evaluation submitted, results not yet official**
  - unofficial WERs:

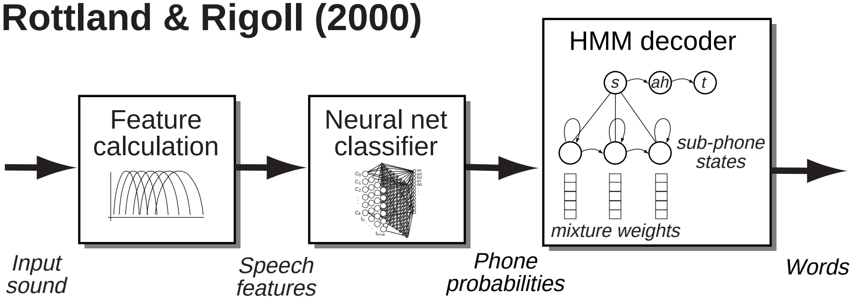
| | |
|---|---|
| MFCC/SPHINX: | 35% |
| Tandem/SPHINX: | 30.1% |
| full-up CMU (ROVER+MLLR): | 26.5% |
| CMU + Tandem (ROVER): | 25.7% |

- **Conclusions:**
  - Tandem from CI labels still tractable for LV
  - improvements may not be so dramatic

# Current Tandem work

- **Aurora 2000: Cross-language**
  - training Finnish & Italian systems
  - union of all phone sets?
  - clustering of cross-language phones
- **Other targets for neural net training**
  - HMM states
  - articulatory targets
- **System variants**
  - 'mixture of posteriors'

**Rottland & Rigoll (2000)**

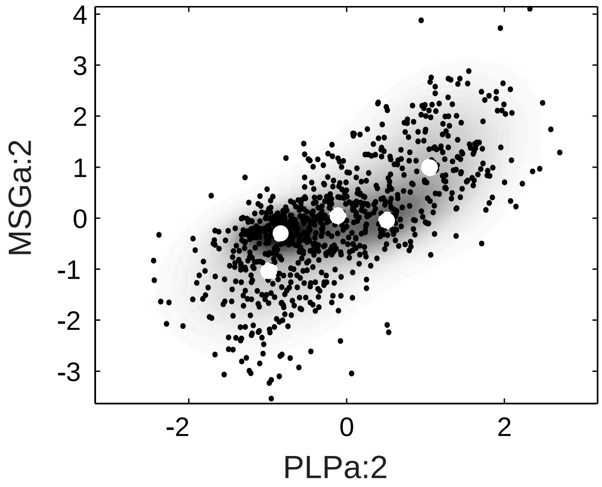


- **Transfer to DC**

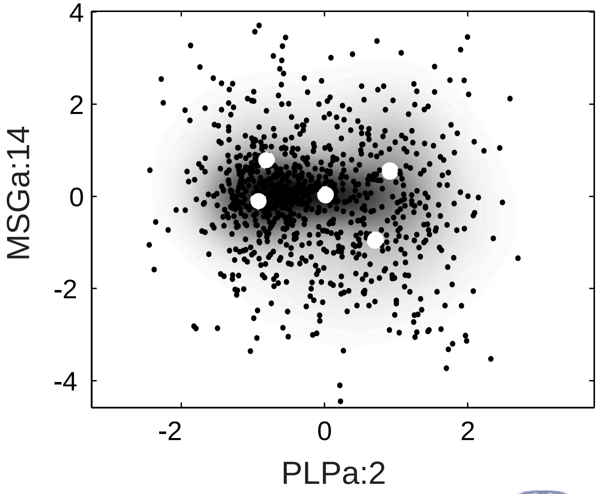
# 2 Mutual Info for multistream design

- **Combination best for *complementary* streams**
- **Try to predict by looking at Mutual Information:**
  - low *classification* MI implies different information
- **Can also use to choose combination point**
  - feature combination (concatenation) for streams with interdependence (*high* feature MI)
  - else posterior (post-classifier) combination

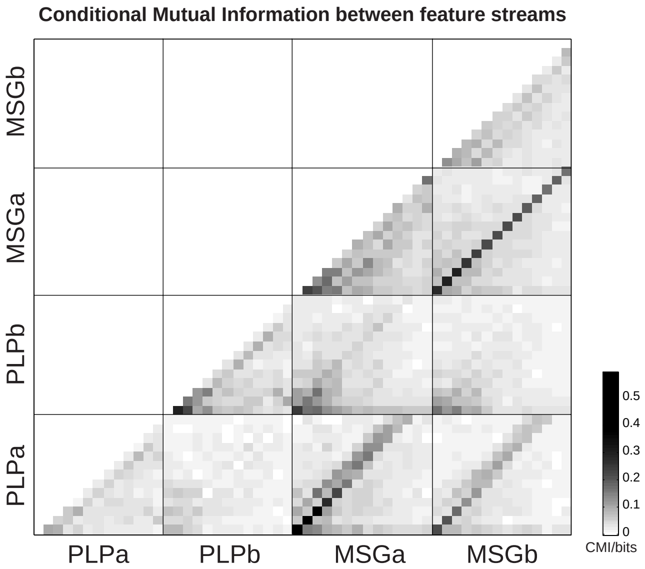
Hx = 7.30, Hy = 6.99, MI = 0.52



Hx = 7.17, Hy = 6.88, MI = 0.03



# MI for multistream: results

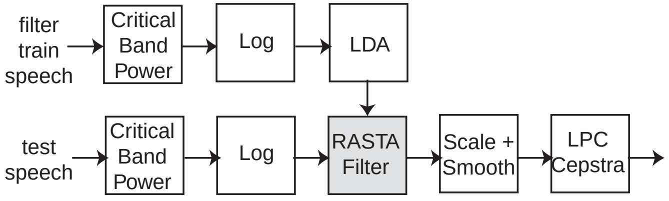


| *Stream 1* | *Stream 2* | *Feature CMI* | *Classif CMI* | *FC WER ratio* | *PC WER ratio* |
|---|---|---|---|---|---|
| PLPa | PLPb | 0.04 | 0.26 | 89.6% | 97.6% |
| MSGa | MSGb | 0.21 | 0.25 | 85.8% | 101.1% |
| PLPa | MSGb | 0.11 | 0.15 | 78.1% | 86.3% |
| PLPb | MSGa | 0.09 | 0.24 | 87.5% | 89.7% |

- **Low Classif. CMI correlates with good pairs**
- **PC vs. FC more complex than Feature CMI**

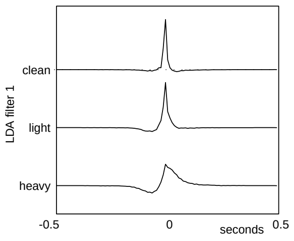
# 3 Other Multistream work: Multifeature combination (Mike Shire)

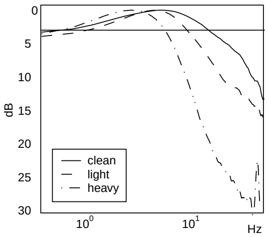
- **LDA design of condition-dependent features:**



filter train speech → Critical Band Power → Log → LDA

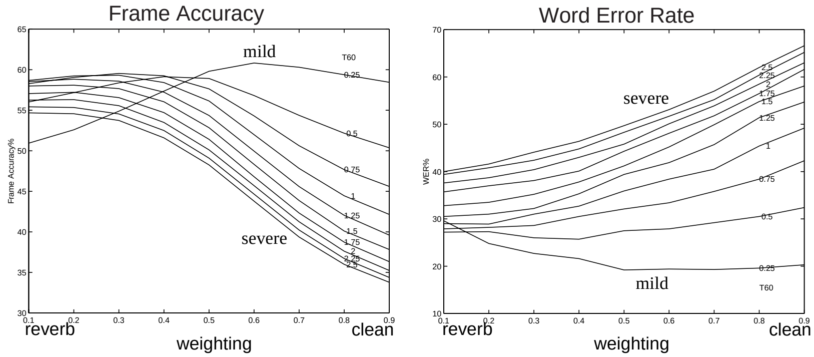
test speech → Critical Band Power → Log → RASTA Filter → Scale + Smooth → LPC Cepstra →





- **Combine various conditions, test on all:**

**LDA-RASTA-PLP: Combining CLEAN and REVERB**



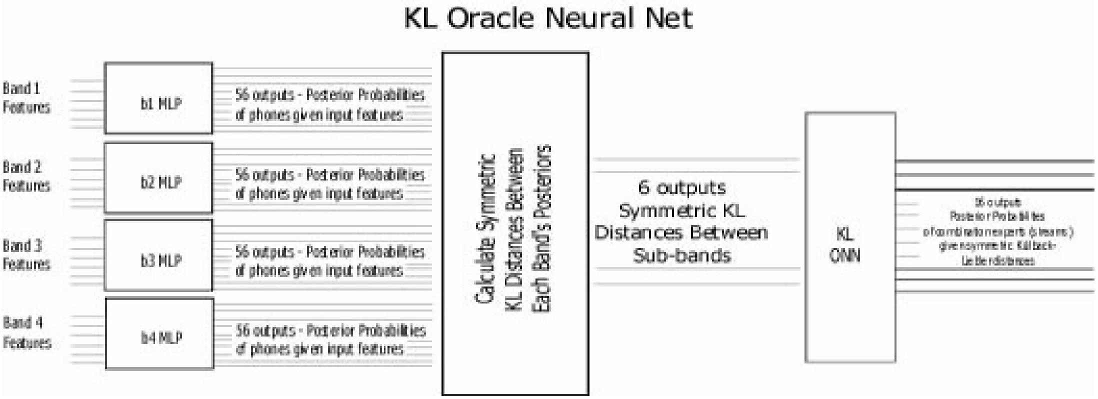
Frame Accuracy

Word Error Rate

# ‘Oracle nets’ for FC multistream (Barry Chen)

- **4 bands → 15 combinations (+priors): (smoothed) ‘oracle’ choice halves WER**

- **Can we train a net to make ‘oracle’ choice?**
  - based on KL distance between posteriors?


KL Oracle Neural Net
Band 1 Features
Band 2 Features
Band 3 Features
Band 4 Features
b1 MLP
b2 MLP
b3 MLP
b4 MLP
56 outputs - Posterior Probabilities of phones given input features
56 outputs - Posterior Probabilities of phones given input features
56 outputs - Posterior Probabilities of phones given input features
56 outputs - Posterior Probabilities of phones given input features
Calculate Symmetric KL Distances Between Each Band's Posteriors
6 outputs Symmetric KL Distances Between Sub-bands
KL ONN
15 outputs Posterior Probabilities of combination experts (streams) given symmetric Kullback-Leibler distances

| *System* | *Word Error Rate* |
|---|---|
| best net (4 band) | 5.1% |
| phone-smoothed oracle | 2.7% |
| KL oracle-net weighted streams | 4.9% |

- **Not much help in practice...**

# 4 Other projects: Meeting recorder

- **ASR in conventional meeting environments**
  - for transcription/summarization/retrieval
  - distant acoustics!
  - informal, overlapped speech (c/w ShATR)
- **Data collection:**

  - wired room at ICSI
  - other systems at UW ...

# Meeting Recorder (cont’d)

- **Preliminary analysis**
  - transcription & forced alignment (IBM)
  - ground truth in turns/overlaps
  - preliminary distant-mic recordings
- **Research areas**
  - meeting dialog: overlaps, turns etc.
  - language modeling for meetings
  - feature design for distant acoustics
- **Future support**
  - DARPA ‘ROAR’ program?

# LabROSA:

## The Laboratory for Recognition and Organization of Speech and Audio

- **New research group at Columbia University in the City of New York**
  - existing EE dept. signal processing group
  - addition of speech/audio for true multimedia
- **Research: extracting information from sound**
  - real-world ASR
  - higher-order: speaker ID, dialog structure
  - nonspeech: events, acoustic environment ID
- **Recruiting students**

  http://www.ctr.columbia.edu/~dpwe/LabROSA/

# CRAC2001:

## “Consistent and Reliable Acoustic Cues for speech and sound analysis”

http://www.ee.columbia.edu/CRAC2001/

- **RESPITE Contractual Obligation Workshop:**
  - Identifying sources/info (CASA, BSS, SNR est)
  - Robust ASR (MD, MS, compensation)
  - Nonspeech, music applications
  - Psychoacoustics of perception in noise
  - Combinations
- **Satellite event at Eurospeech-2001, Aarhus**
  - held on Sunday 2000-09-02 (day before) at Eurospeech location
  - separate registration
- **Workshop structure**
  - lecture + posters, am + pm, discussion
  - limit to ~ 40 participants

# CRAC2001 (cont'd)

- **Organizing committee**
  - Dan & Martin, co-chairs
  - Fred, Phil, Andy
  - Andrzej Drygajlo (EPFL) & H. Okuno (CASA)
- **Timetable:**
  - CFP: imminent
  - Abstracts: April 30th, 2001
- **Actions:**
  - help with publicity
  - plan your submission!