Review of SPRACH/Thisl meetings Cambridge UK, 1998sep03/04

Dan Ellis

International Computer Science Institute, Berkeley CA

<dpwe@icsi.berkeley.edu>

Outline

1 **SPRACH overview**

2 **The SPRACH Broadcast News effort**

3 **SPRACH final review meeting**

4 **Thisl overview**

5 **Thisl technical highlights**

SPRACH

1

- **EU-funded basic-research project**
  - 1996-1998 (follow-on to 'Wernicke' project)

- **Participants:**
  - FPMons, Cambridge, Sheffield, ICSI, INESC

- **Motto:**

  "Let a hundred flowers bloom, let a hundred schools of thought compete."
  - MLP, RNN, multistream, mod. spec., syllables...

- **Status:**

  March 1998: Enough horticulture, time to focus & collaborate

  $\rightarrow$Broadcast News 1998 evaluation submission

SPRACH: Technical highlights

- **Focus on BN'98 system**
  - evaluation period: 1998oct13 - 1998nov10

- **Acoustic models:**
  - Cambridge RNNs (2 or 4) trained on 74h
  - ICSI MLP: 4000 or 8000 HU, 74 or 150h
  - Sheffield: best combo is 2xRNN+MLP, entropy weighted over 5s window ($27\% \rightarrow 26.2\%$)
  - Context-dependent system? $p(q,C \mid X) = p(q \mid X) \cdot p(C \mid q, X)$
  - Mons: NLDA: (26x9):600:**36**:54 begets (**36**x9):600:54 (WER 43.3% c/w 47.0%)

Other BN'98 issues

- **Segmentation**
  - CMU system hurt CUCON badly in 1997
  - Cambridge (Tony) has new system based on GMM mixture weights & clustering
  - Non-segmenting (chronos, online norm)?

- **Pronunciations & grammar**
  - Expanded dictionary from acoustics-derived decision-tree rewrites (Eric)
  - 65k wordlist based on training corpus, recent stories
  - Multiwords from top bigrams, baseform variation
  - Variety of large text corpora for training trigram & 4-gram LMs

- **Spoke systems**
  - 1x, 10x, 50x systems (lattice rescore, dyn pron)
  - named entities

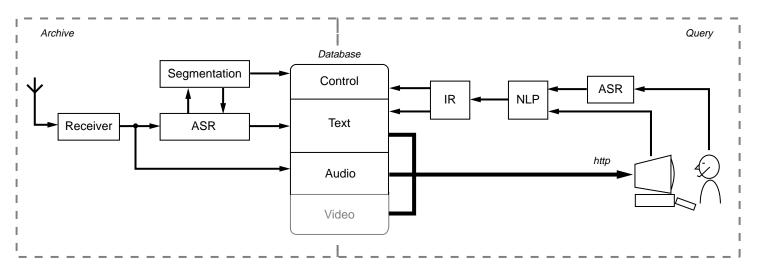# SPRACH final review

- **To be held Lisbon, 1998dec16**

- **Review 1998 BN system**
  - emphasize collaboration

- **Also: demonstrate recognition systems**
  - Portuguese: 22h read speech, 26x7:2000:39
  - English, French, Portuguese
    in BeRP-like front-end
  - Real-time decoding of CNN?

ThisI overview

- **EU funded project, conceived as 'applied'**

- **"THematic Indexing of Spoken Language"**
  - specifically: newsroom access to archive of news broadcasts
  - BBC is key industrial collaborator
  - also, French system?

- **Next review is 'workshop' at Cambridge, 1999apr**

ThisI: Technical developments

- **Training data**
  - BBC has transcribed & aligned 20 hours

- **Archive data**
  - 15 hr/week recorded since 1998jan = ~20GB

- **Information Retrieval (IR) query engine**
  - Public-domain versions (PRISE)
    used in TREC/MUC
  - Sheffield have developed 'homebrew'
  - Query expansion, discrimination

- **Demo system**
  - Softsound (Cambridge) to train RNN recognizer
    (derived from WSJCAM0)
  - Database building: run recognizer on archive
  - Sheffield have basic CGI-based demo
  - ICSI to enhance demo front-end

More Thisl highlights

- **French baseline BN-style system**
  - DanG@IDIAP: new prons gave 10% abs to 27%
  - IDIAP: 12h BREF training data c/w 200h LIMSI

- **IDIAP: Confidence scores via N x 1-best**
  - e.g. unigram, bigram, trigram, diff LMSF
  - idea is to analyze occurrences of words...

- **FPMs: speaker tracking**
  - mixture models for 1s each side of boundary
  - distance metric between mixtures
  - slide for best boundary

- **Softsound:**
  - adaptive beam in Chronos decoder
  - teletext for automatic pronunciation checking
  - 'confidence-based' training rejects poor segs