European projects update

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

Outline

1. *Thisl* final year planning meeting
   (BBC, London, feb03)
2. Future PDA project discussion
   (Sheffield Univ., feb04)
3. *RESPITE* kickoff meeting
   (ICP, Grenoble, feb07-08)
This final year planning meeting

- This project:
  Using ASR (&c) to index BBC news archives

- ESCA workshop on Spoken Document Retrieval (SDR) - April, Cambridge
  - systems, IR/IE
  - demos, including thisIR

- Current actions:
  - finalize UKEng system to run on 1000hr archive (to demo in April)
  - ICSI to train MLP on BBC data (50hr)
  - segmentation? speaker tracking?

- Other highlights:
  - Latent Semantic Analysis with Self-Organizing Maps for SDR (Mikko Kurimo/IDIAP)
  - Confidence-based embedded training (Tony)
ThisI demo

- Stand-alone Tcl/Tk implementation
- doesn’t require httpd
- speech-input ready

Diana crash investigation tells C. N. a massive police report includes two witnesses saying they saw flat tire zigzagging out of this Paris tunnel right after the crash killed Diana, her friend and their driver. Witnesses also tell police the flat exhaust pipe appeared to be damaged in sought the tone in addition a large dog reportedly was seen in the fact that car purse police are searching for the flat they do believe that it played in time a scar.

10:11 where scores of witnesses and there was harsh testimony from Oklahoma city bombings of robbers and grieving family members prosecution has now rested in the penalty phase of Terry Nichols trial of C. N. national correspondent Tony Clark has on another difficult day in court the second floor courtroom echoed with anguished shots of Kathleen Turner she lost her four year old daughter and mother and father in law in the bombing facing at Terry Nichols and pounding on the

towel aside the heated on demands arise on the民众 to have the verdict taken from the jury.

European Trip Report - Dan Ellis 1999feb11 - 3
PDA proposal discussion

• New EU ‘Future & Emerging Tech’ funding program (Framework 5)
  - meeting to plan a possible project proposal

• Partners interested in speech-centered PDA:
  - Roger Tucker (HP Labs): audio info appliances
  - Tony Robinson (Cambridge): ASR, systems
  - Steve Renals (Sheffield): information access (IA)
  - IDIAP: multi-modal, Hervé’s ‘encapsulators’
  - FPMs: synthesis
  - ICSI: robust SR, UI/apps

• HP vision
  - ‘fat pen’ with mic & small display
  - dictaphone meets PDA
  - docks to PC
  - use pen-motion info?
  - access: skimming, summary, keyword search...
SpeechPDA proposal (cont’d)

• **Current application definition**
  - palm-style machine, docks, mixed online/offline
  - ASR for search, control, form-filling, summaries
  - voice notes/dictation as primary focus;
    ‘ambient’ recording as parallel track (shares IA)

• **Work packets**
  - ASR: wordspotting / robustness / efficiency
  - Info access: browse / skim / structure / SDR
  - System integration & user interface
  - ? other modalities (stylus, video?)
  - Evaluation: components / integrated system

• **Other observations**
  - meeting recorder doesn’t have to be palm-size
    (hierarchy of size/power trade-offs?)
  - using ASR (IA) independent of where it’s done
  - OK to have algorithms without a real prototype
RESPITE kickoff meeting

- **REcognition of Speech by Partial Information TEEchniques**
  - Auditory scene analysis etc. to find information
  - Multistream & missing-data to exploit it
  - new 3yr EU-funded project: Sheffield, IDIAP, FPMs, ICP Grenoble, ICSI, DaimlerChrysler, Matra

- **Rationalize work at partner labs:**
  - missing-data at SU, IDIAP, FPMs
  - multi-stream at IDIAP, ICSI, ICP

- **Baseline task:**
  - “Aurora” Distributed Speech Recognition task: TIDIGITS corrupted in various ways
  - HTK and/or comparable system configuration

- **CASA toolkit**
  - practical information for use in ASR
Issues in missing data (Sheff/IDIAP)

- **Input features tagged as present/missing**
  - e.g. by subband SNR, scene analysis

- **Classic: ‘Class imputation’**
  - integrate over missing data dimensions to evaluate output likelihoods:
    \[
    p(X|q) = \int p(X_{\text{present}}, X_{\text{missing}} | q) \, dX_{\text{missing}}
    \]
  - i.e. just skip dimensions of Gaussian
  - can use ‘upper bounds’ on spectral values

- **New: ‘Data imputation’**
  - use \( E[X_{\text{missing}} | X_{\text{present}}, q] \)
  - permits cepstra, deltas

- **What about connectionist systems?**
  - also permitted by data imputation
  - or Radial Basis Function neural networks?
  (Andy Morris/IDIAP)
Harmonicity labelling for multistream
(Herve Glotin, ICP/IDIAP)

- ‘Pitch pulse’ in envelope autocorrelation is correlated to subband SNR (for vowels)
- Use artificial mixtures to train $R_{xx}$→SNR map
- ‘Full combination’ multistream needs weights:
  - $p(q \mid a,b,c,d) = \sum_S p(S) \cdot p(q \mid S,a,b,c,d)$
    
    $S$ ranges over 16 possible combinations
  - uniform weighting is worse than best single $S$
  - $p(S) = p(SNR > \theta)$ gives best result:
    
    NB-noise-Num95: 15%FB → 13.3%WMB
Multistream vs. alternatives
(Andy Morris)

- Keep an eye on alternative techniques
  - e.g. noise robustness through spectral subtraction, microphone techniques

- Techniques may not combine additively
  - e.g. log Rasta vs. j-Rasta for full/multi-band:

\[
\begin{array}{ccc}
\text{SNR} & \text{WER} & \text{FB log Rasta} \\
\text{SNR} & \text{WER} & \text{MB log Rasta} \\
\end{array}
\]

\[
\begin{array}{ccc}
-30dB & 0% & j-Rasta \\
+30dB & 100% & \text{log Rasta} \\
\end{array}
\]