

---

---

# European projects update

Dan Ellis

International Computer Science Institute, Berkeley CA

<dpwe@icsi.berkeley.edu>

## Outline

1. ***This!* 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 this!IR
- **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)



# This! demo

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

The screenshot shows a window titled 'thislui.tcl' with a menu bar containing 'File' and 'Options'. The main title is 'This!IR demo'. Below the title is a search input field labeled 'Enter query:' containing the text 'terry nichols family'. Below that is a 'Results for:' field also containing 'terry nichols family'. A table displays search results with columns for Program, Date, Time, and Context. The second row is highlighted in blue. Below the table are input fields for 'Program:', 'Date:', 'Time:', and 'File:'. The 'Program:' field contains 'CNN Primetime News', 'Date:' contains '1997dec31', and 'File:' contains 'em971231'. The main text area shows a snippet of text with the search terms highlighted in bold.

Program	Date	Time	Context
CNN The World Today	1997dec26	09:02	tony clark among the pristine <b>terry nichol</b>
CNN Primetime News	1997dec31	10:11	bombings of robbers and grieving <b>family</b>
CNN The World Today	1997dec26	31:15	the penalty phase of <b>terry nichols</b> oklaho
CNN Headline News	1998jan05	02:37	the sentencing phase of the <b>terry nichols</b>
CNN Headline News	1998jan04	07:02	in the sentencing phase of <b>terry nichols</b>

Program: CNN Primetime News Date: 1997dec31 Time: File: em971231

diana crash investigation tells c. n. n. a massive police report includes two witnesses saying they saw fiat uno zigzag ing out of this paris tunnel right after the crash killed diana her friend and their driver witnesses also tell police the fiat 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 fiat they do believe that it played a role in time a scar

10:11 where scores of witnesses and there was harsh and testimony from oklahoma city bombings of robbers and grieving **family** members prosecution has now rested in the penalty phase of **terry nichols** trial c. n. n. national correspondent tony clark has more on another difficult day in court the second floor courtroom echoed with the anguish shots of kathleen trainer she lost her four year old daughter and mother and father in law in the bombing flaring at **terry nichols** and pounding on the witness stand she shouted my daughter's going on she was taken from me she was taken from my



---

---

## 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 TEchniques**
  - 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:
$$p(X|q) = \int p(X_{present}, X_{missing} | q) dX_{missing}$$
  - i.e. just skip dimensions of Gaussian
  - can use 'upper bounds' on spectral values
- **New: 'Data imputation'**
  - use  $E[X_{missing} | X_{present}, q]$
  - 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  $R_{xx} \rightarrow \text{SNR}$  map**
- **‘Full combination’ multistream needs weights:**
  - $p(q | a,b,c,d) = \sum_S p(S) \cdot p(q | S,a,b,c,d)$   
 $S$  ranges over 16 possible combinations
  - uniform weighting is worse than best single  $S$
  - $p(S) = p(\text{SNR} > \theta)$  gives best result:  
NB-noise-Num95: 15%FB  $\rightarrow$  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:

