SPRACH - WP 6 & 8:
Software engineering work at ICSI

March 1998

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

1. **Hardware: MultiSPERT**
2. **Software: speech & visualization tools**
3. **Packaging: SPRACHworks**

1 Hardware: MultiSPERT

- **Multiple SPERT boards on single host**
  → "MultiSPERT"
- **New software, firmware, hardware**
  - for nets: 'pattern parallel' and 'network parallel'
- **Fastest performance: 500 MCUPS**
  (5 boards, 10x previous single-board)
- **Current neural-net trainings:**
  8000 hidden units, 3.2M params, 18M frames
  → ~10^{15} ops per training iteration
  - this size not previously possible

2 Software: Speech tools

- **New components:**

  `feacalc`: front-end feature calculation
  - uses core RASTA-PLP routines from **`rasta`**
  - mnemonic options ("**`-L`**" → "**`-rasta log`**")
  - comprehensive file format support

  `featools`: 'scaffolding' for novel features
  - handles finding input files & assembling output
  - user provides (chain of) online-feature applets

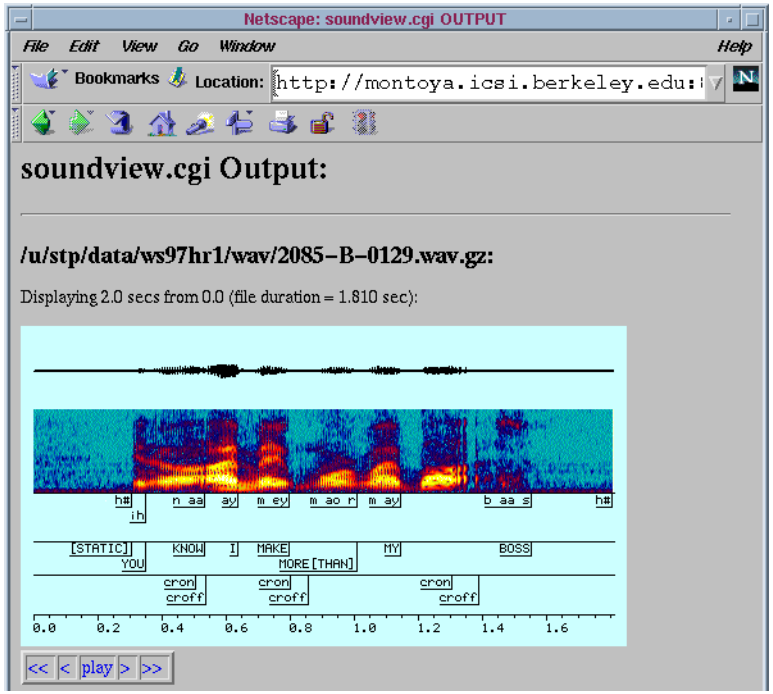
  `pfile_utils`: feature archive manipulation
  - merging, splitting, rearranging, summary stats
  - **`pfile_gaussian, pfile_klt...`**

Visualization & User Interface tools

- **ICSI is developing a toolkit of Tcl/Tk modules for visualization & demos**
- **Current pieces**
  - visualization of signals, features, probabilities, label alignments
  - interactive & file-based audio input/output
  - web-interface tools
- **Some current applications**
  - **`sgramImg.cgi`**
    on-demand spectrogram gifs
  - **`recogviz`**
    comparing different ASR configs
  - **`berpdemo98`**
    speech application plus graphics

sgramImg.cgi

- **CGI script: spectrogram GIF on-demand**
  - point to in <img src..> tag of other pages

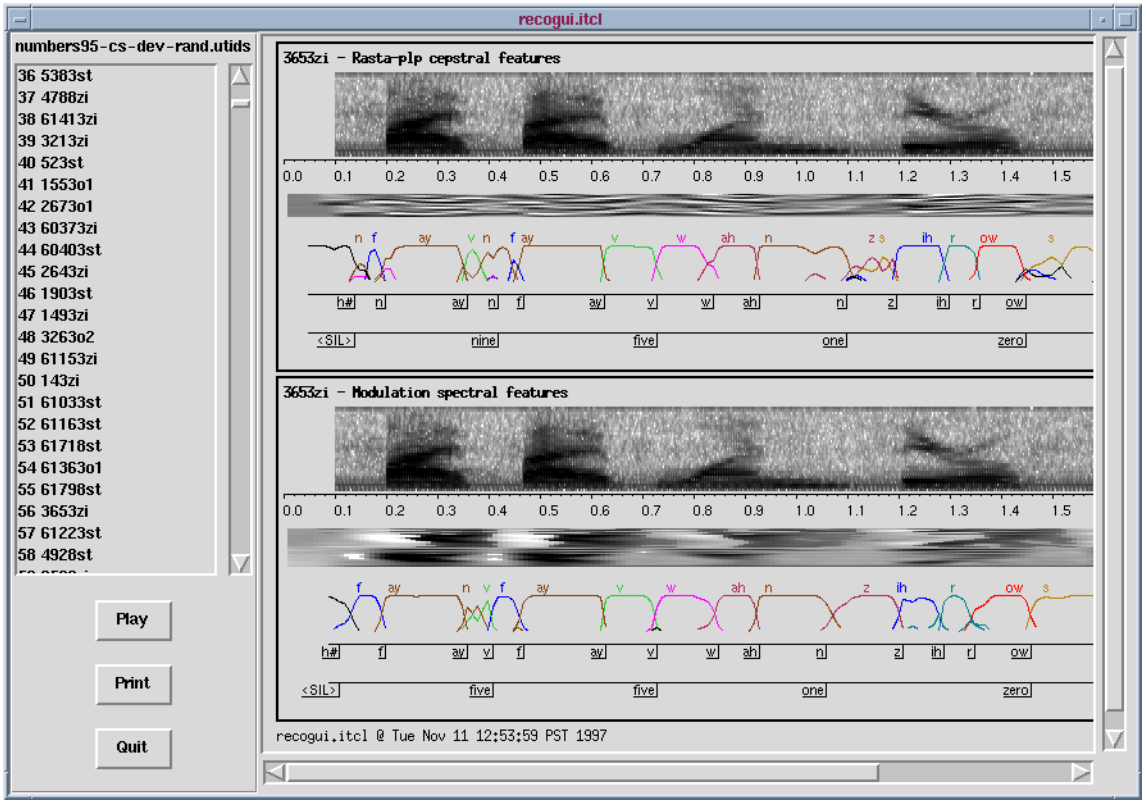


- **Automatically includes xlabel annotation**
- **Try it, e.g. links from:**

http://www.icsi.berkeley.edu/~dpwe/research/etc/phnless.html

recogviz

- **Motivation: compare recognition techniques**
  - at each stage in process (signal, features, probs)



- **Easily to incorporate novel features etc.**

berpdemo98

- **Existing application plus recogviz modules**

- **Demo available...**

3 Packaging: SPRACHworks

- **Original SPRACHworks goal:**
  Source-level distribution of all partner tools
  - integrated: compatibility between tools
  - portable: easily compiled
  - ready-to-run: demos to play with
- **HUGE amount of work!**
- **Reduced scope:**
  - several separate but conformal packages (ICSI, Cambridge, FPMs/Strut)
  - incremental incorporation of packages
- **Status:**
  - **first collection:** April ’97 (Cambridge)
  - **latest:** single ‘make’ for complete speech demo (“SPRACHcore” = 19 packages)
  - ICSI-developed, installed at FPMs & IDIAP

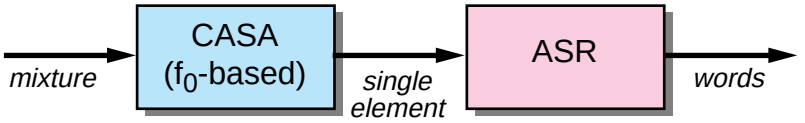
Future work

- **Tools**
  - language modelling, two-pass decoder, etc...
  - formal modular structure for visualization
    $\rightarrow$ ‘drop-in’ demos
- **Packaging**
  - include more packages: ASR training
  - tutorial documentation
  - more platforms?

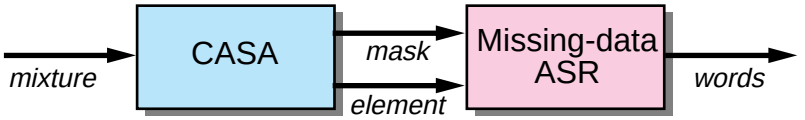
Combined speech/nonspeech analysis:

How to combine ASR with Computational Auditory Scene Analysis (CASA) for nonspeech transients?

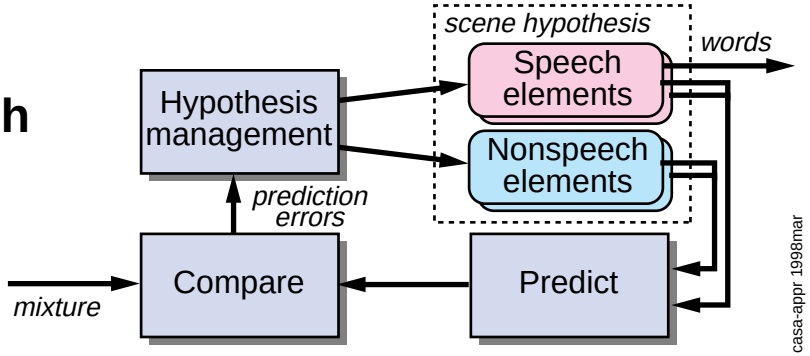
- **Preprocessor** (Weintraub'85, Nakatani/Okuno'97)



- **Use extra CASA info in ASR** (Cooke,Morris & Green'97)



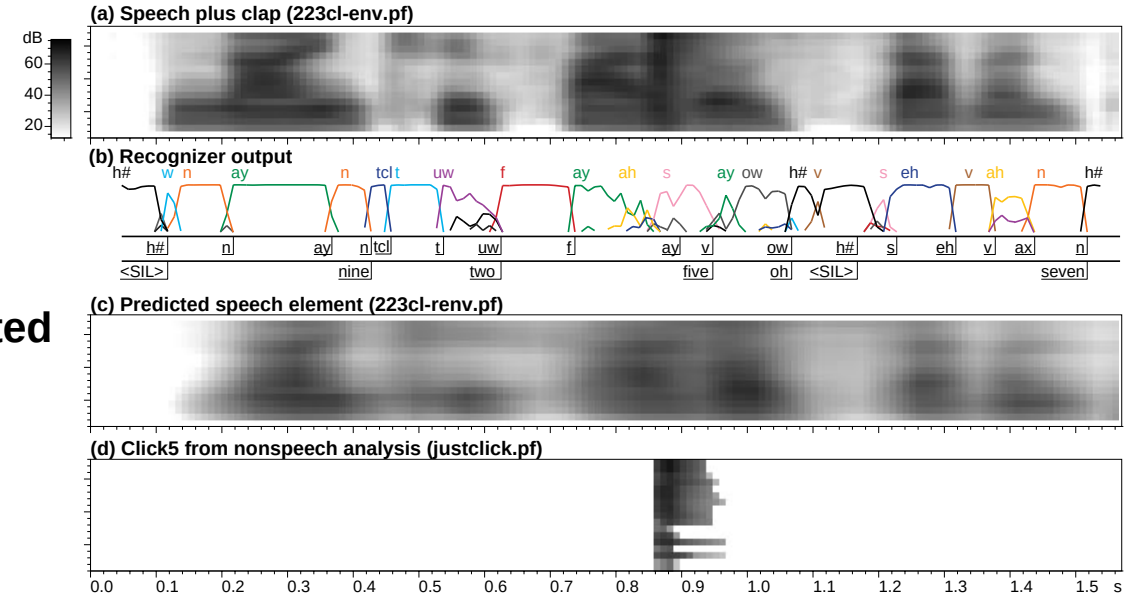
- **Simultaneous search for speech & nonspeech elements**



  - **Difference between *observations* & *predictions* drives analysis**

Speech/nonspeech: preliminary results

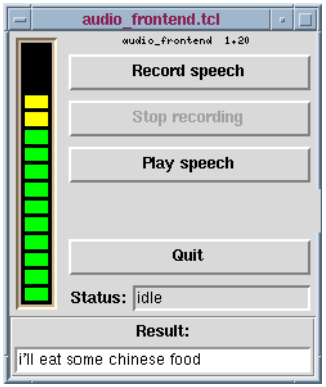
- **Original speech +clap**
- **Bootstrap recognition**
- **Reconstructed speech element**
- **Recovered nonspeech**



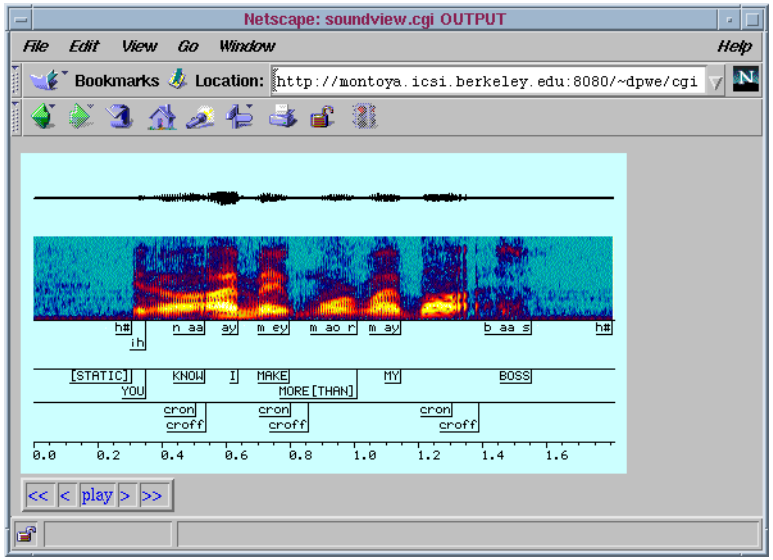
- **Issues:**
  - iterative refinement of each element
    $\rightarrow$ use nonspeech to ‘relax’ speech recognition
  - reconstructing speech features from ASR output
  - use f_0-based separation to bootstrap ASR; recognizer trained on periodic + noise

ICSI GUI tools in ThisL

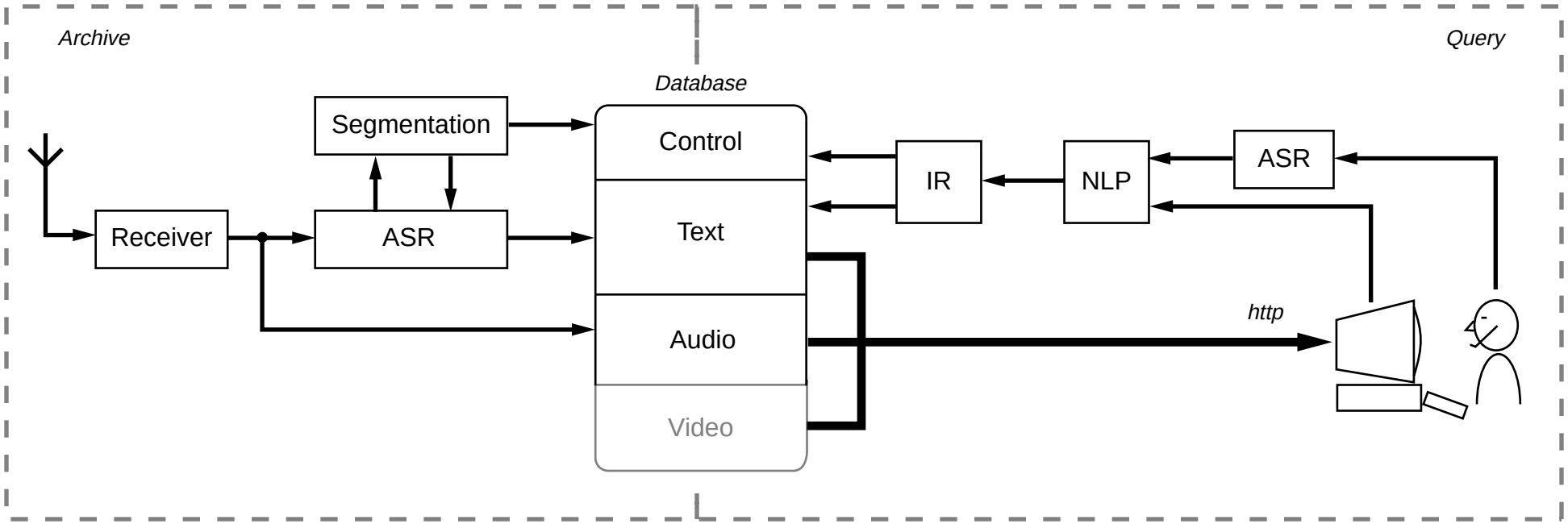
- **Interactive `audio_frontend` for commands**



- **Web-embeddable recognition display tools**



ThisL Overview



Domains:
- BBC news (3hr/day)
- NIST TREC
- French?

Issues:
- Size of episodes → decoder
- Size of database → speed
- Segmentation/side info
- Information Retrieval

Dan Ellis <dpwe@icsi.berkeley.edu> 1998apr02