
ThisI progress report

Dan Ellis

International Computer Science Institute, Berkeley CA

<dpwe@icsi.berkeley.edu>

Outline

- 1 ThisIGui enhancements
- 2 ThisI MSG-MLP acoustic model
- 3 MLP-based speaker ID
- 4 Speech/nonspeech discrimination
- 5 Quicknet enhancements etc.



1

ThisIRGui enhancements

- **Spoken query input:** SPRACHdemo/AbbotDemo
- **NLP integration:** Thomson's prolog lattice parser
- **Faster:** SRT files parsed & saved (+ Tcl8)
- **Better:** supports Real Audio, more status feedback

The screenshot shows the 'ThisIR demo' application window. The interface includes a menu bar (File, Options), a control panel on the left with buttons for recording and playing speech, and a main display area. The main display area shows the entered query 'a giuliani is a elections', the search results for 'giuliani elections', and a parse tree for the recognized query 'i'm working on giuliani's election'.

Enter query: a giuliani is a elections

Results for: giuliani elections

Program	Date	Offset	Context
PRI The World	1997oct16	00:33	new york mayor rudolph giuliani h:
CNN The World Today	1997sep09	52:32	a race against mayor rudy giuliani
CNN Early Prime	1997jun27	30:33	the new york mayor rudy giuliani e
PRI The World	1997oct15	58:53	the new york mayor rudolph giulian
CNN Primetime News	1997sep18	02:52	last year's teamsters presidential e
CNN Early Prime	1997jun27	43:04	gun violence is what mayor giuliar
PRI The World	1997oct23	28:47	polls have closed in local election
CNN The World Today	1997aug23	06:27	the white house official tells c. n. n
CNN Primetime News	1997sep23	05:34	police eight yards mayor rudy giul

Program: PRI The World **Date:** 1997oct16 **File:** eh971016 **Stop playback**

00:01 terrorist attacks one hundred and fifteen people died in two explosions at the israeli embassy and jewish community center the only suspect arrested so far are for argentine police officers president clinton told the families that the u. s. would offer any assistance necessary to solve the crimes publicly white house officials rejected the views of many argentinians that president carlos menem who has yet to meet with the victims' families himself has done too little to solve the murders mar alliance and n. p. r. news blame insiders

00:33 new york mayor rudolph giuliani has filed a lawsuit challenging the constitutionality of the line i am veto the suit argues that the new party shifts power to tax and spend from congress to the president and p. r.'s elizabeth arnold reports the new york mayor's interest centers on medicaid funding provisions vetoed by president clinton in his first use of the alignment party state officials estimate two point six billion dollars are at stake last august president clinton struck three item's from the bill including a provision that would have spared the state of new york from having to return the two point six billion dollars in medicaid eight hundred c. from the

Recog: i'm working on giuliani's election
Parsed: i am working on a giuliani is a elections
Keywds: a giuliani is a elections

```

graph TD
    be[be] --- vp7[vp7]
    be --- keyw[keyw]
    vp7 --- aux3_p[aux3_p]
    vp7 --- ger1[ger1]
    aux3_p --- np1[np1]
    aux3_p --- am[am]
    np1 --- pronoun_pers[pronoun_pers]
    ger1 --- working[working]
    keyw --- a[a]
    keyw --- giuliani[giuliani]
    keyw --- is[is]
  
```



MSG-MLP acoustic model

- **From SPRACH: combined models good**
 - especially plp-RNN and msg-MLP
 - e.g. WER: 27.2% + 29.7% → 24.9%
- **Obtained 50hr BBC training set from Softsound**
 - trained (28x9):8000:42 multi-layer perceptron
 - used TetraSPERT = 175h train, 375 MCUPs
 - feature calc: 0.2 xRT; fwd pass: 0.3-1.6-2.1 xRT
- **Results (euro99Eval test set):**
 - RNN baseline: 29.2%
 - msg1N-8k alone: 35.5%
 - Posterior combination: 28.7%
- **Why less benefit than SPRACH?**
 - not enough training data?
 - no msg-based realignment?
 - less telephone-bandwidth data?
 - (bugs?)



MLP-based speaker ID

(par Dominique GENOUD, ex-IDIAP)

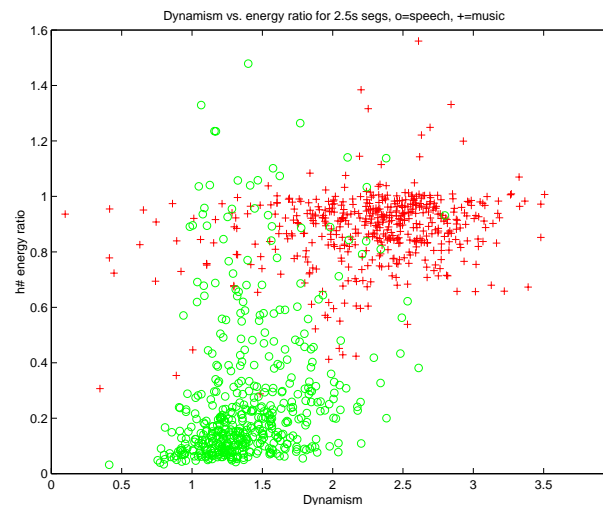
- **How to use hybrid systems for speaker ID?**
 - train speaker-dependent nets? too little data
 - specialize SI nets with a little SD data and compare posteriors?
- **But nets are *discriminative*...**
 - speaker-detection nets have *two* outputs per phone: speaker's phone, rest-of-world phone
 - train on 15 min of speaker + 15 min of others
 - sum posteriors on Viterbi path for each half
 - EER on 12 speakers (from BN) ~ 9% (c/w ?)
- **Applications**
 - speaker ID - but have to gather training data
 - speaker-adapted recognition:
SD-trained nets have ~ 20% RER reduction



Speech/nonspeech discrimination

(with Gethin Williams of SU)

- **Posteriors features to detect ‘decodable’ segs**
- **4 statistics based on acoustic model outputs:**
 - avg per-frame entropy
 - |first-order diff|²
 - energy ratio of h#
 - phone var’ce template
- **Test on Scheirer/Slaney music+speech data:**
 - classif err: 0% (15s segs); 1.3% (2.5s segs)
 - use to discard non-speech before decode



Quicknet enhancements

- **New release, quicknet v0_97**
- **MultiSPERT support integrated**
 - general client-server structure for other CPUs?
- **Online delta calculation**
 - saves disk space
 - waiting on Torrent-native convolution
- **Online per-utterance normalization**
 - two-pass - bad with online deltas
- **Also:**
 - new RLE-compressed ilab label format (1/30th)
 - full support for pre/lna input
 - bug fixes



Other news

- **Fabio CRESTANI visiting ICSI**
 - ex-Glasgow IR
 - IR for spoken documents, PDA applications
- **Multimedia indexing project with UCB EE (Avideh Zakhor)**
 - Build indexes for video URLs found on the web
 - testing BN recognizers on decompressed audio
- **Segmentation-as-decoding**
 - MDL-style criterion: next segment merges into this model, or starts a new model?
 - long, stable segments become easier to test
 - incremental decoder-like algorithm:
lookahead window, alternate hypotheses

