Oregon Graduate Institute
of
Science & Technology

Spoken Letter Recognition

Ron Cole

Mark Fanty

OGI Speech Group

Alphabet Recognition Over Telephone Lines

PROJECT GOALS:

- Retrieve correct name
- Spelled with pauses between letters
- By any speaker
- From database of 50,000 names
- By July 1, 1991

Why English Letters?

- **Well-Defined, Manageable Problem**
- **Useful Real-World Applications**
- **Well Studied: Can Compare our Approach to Other Technologies**
- **Requires Solutions to Fundamental Problems**

Difficult Discriminations

- **E set: B D E P T G V Z C**
- **A set: J K**
- **EH set: M N F S**
- **AH set: I Y R**

Ambiguous Letter Sequences

HE	AG
HA	AJ
SC	SE
J	JE
E	EE
FV	FE

Telephone Speech: Additional Problems

- Limited bandwidth (F vs S)
- Greater speaker variability
- Greater variability in background noise
- Channel distortion (echo, crosstalk)
- Different handsets

Bandwidth limitation
Formant attenuation

spectrogram pasteup
of normal E, missing f1 and missing
f2

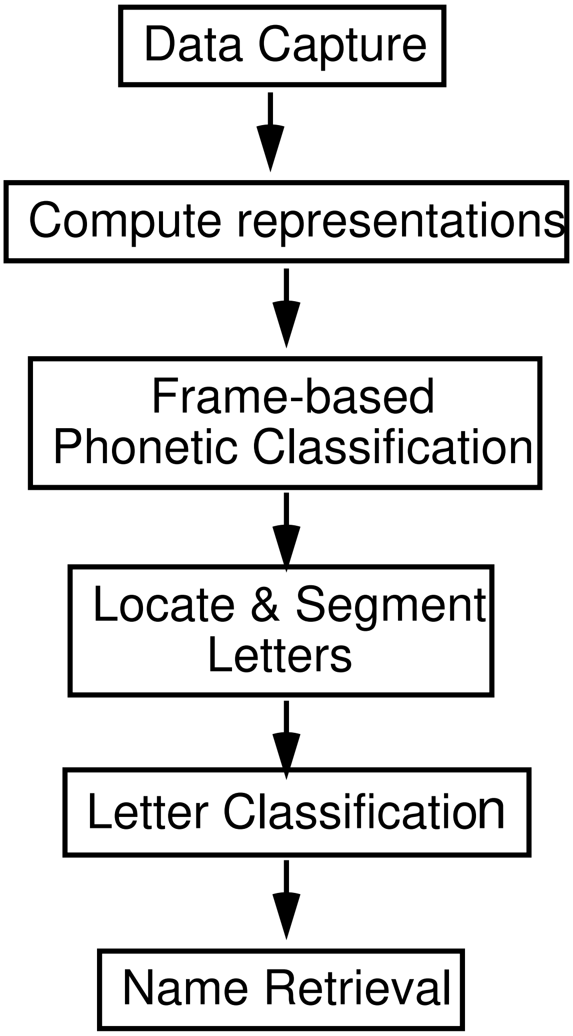
Database Collection

- About 4,000 speakers provided telephone speech samples
- Automatic data collection (800 and local numbers) using Gradient Technology box
- Speakers:
  - -Said Last Name
  - -Said 1st Name
  - -Said city where grew up
  - -Said city from where called
  - -Spelled last name naturally
  - -Spelled 1st name w/pauses
  - -Spelled last name w/pauses
  - -Answered yes/no questions
- Additional 1,000 speakers:
  - - Spelled last name
  - - Said last name
  - - Said city names
  - - Said "capital, hyphen, apostrophe"
  - - Said alphabet with pauses between letters

Human Perceptual Experiment

- 10 listeners identified 3197 letters from 100 alphabets and 100 spelled names
- Performance ranged from 89.9% to 95.1% (average: 92.9%)
- Compares to approx. 99% recognition of microphone-quality speech

Telephone Directory Name Retrieval System



Data Capture

Compute representations

Frame-based
Phonetic Classification

Locate & Segment
Letters

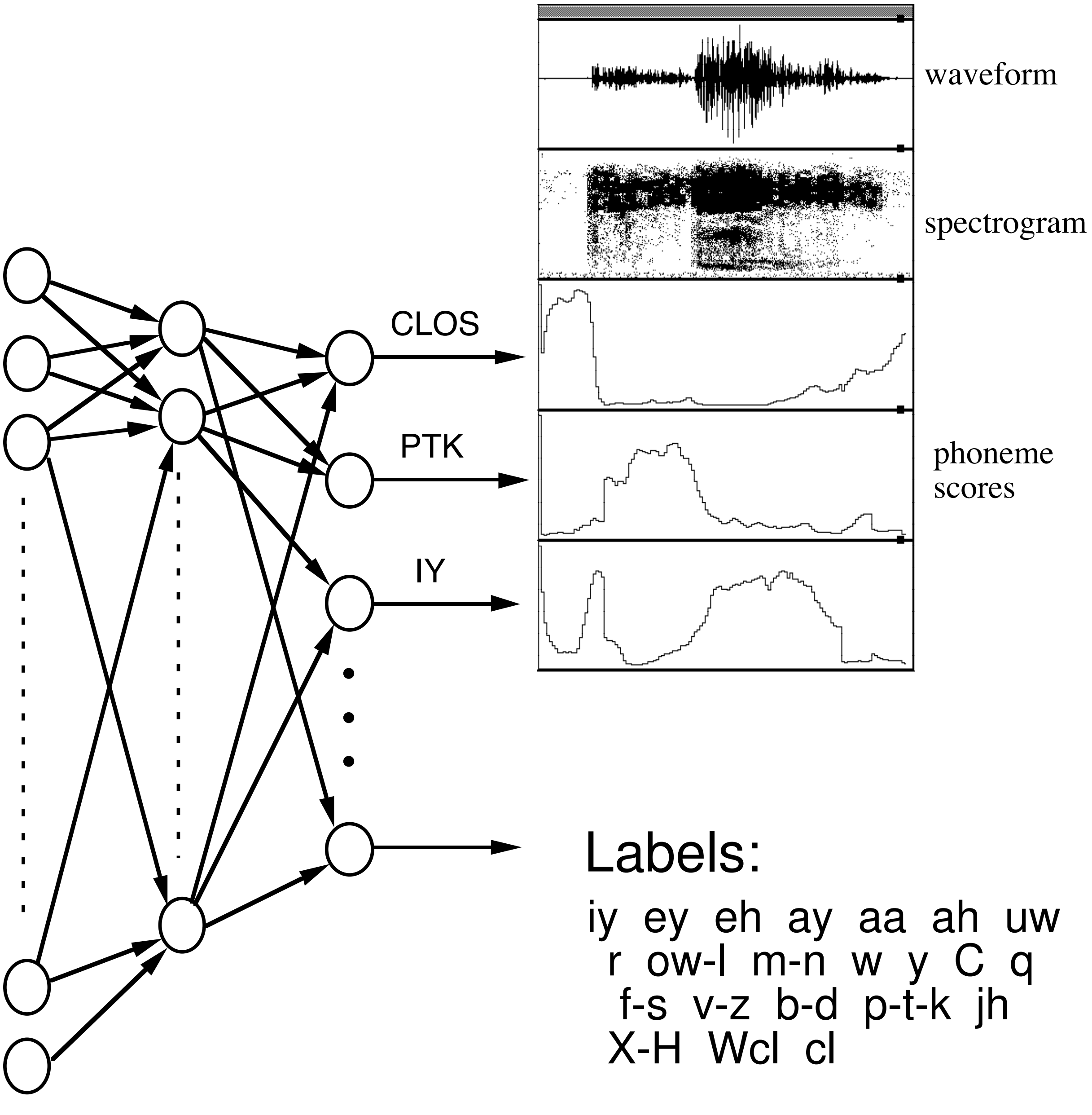
Letter Classification

Name Retrieval

Perceptual Linear Predictive (PLP) Speech Analysis

- Maps DFT to "auditory-like" spectrum
- Simulates known perceptual qualities of human hearing
  - Better frequency resolution at lower frequencies
  - Sounds perceived as louder in middle frequencies
  - Perceived intensity = cube root (Actual intensity)
- Auditory-like spectrum approximated with low-order all-pole model
- Order-7 PLP: 7 spectral coefficients, 1 power coefficient

Phonetic Classification

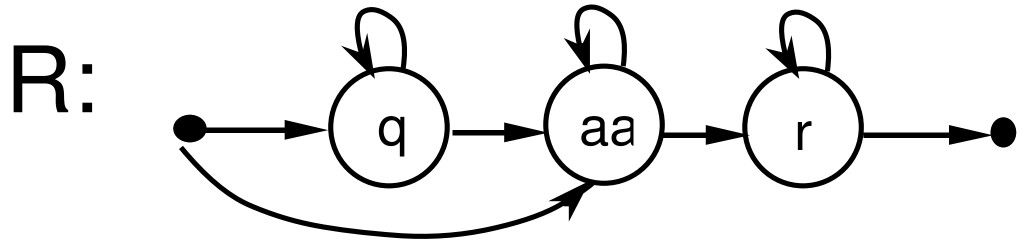


Labels:

iy ey eh ay aa ah uw
r ow-l m-n w y C q
f-s v-z b-d p-t-k jh
X-H Wcl cl

Viterbi Alignments

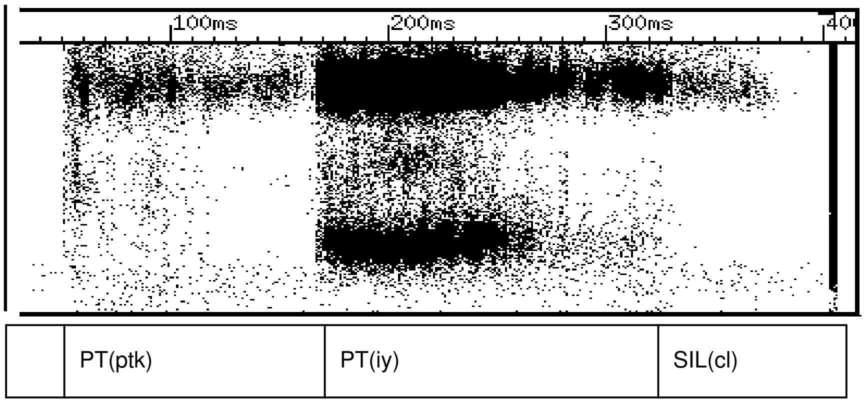
- Each letter has a simple pronunciation model



- Every state of the model corresponds to an output of the phonetic classifier
- There are no transition probabilities, but there are minimum and maximum durations
- The outputs of the classifier are used as probabilities
- A Viterbi search finds the highest-scoring sequence of letters and provides phonetic boundaries

Letter Classification

- Letters found by the Viterbi search are reclassified
- Feature extraction uses Viterbi segmentation
- Outputs are A-Z plus Noise

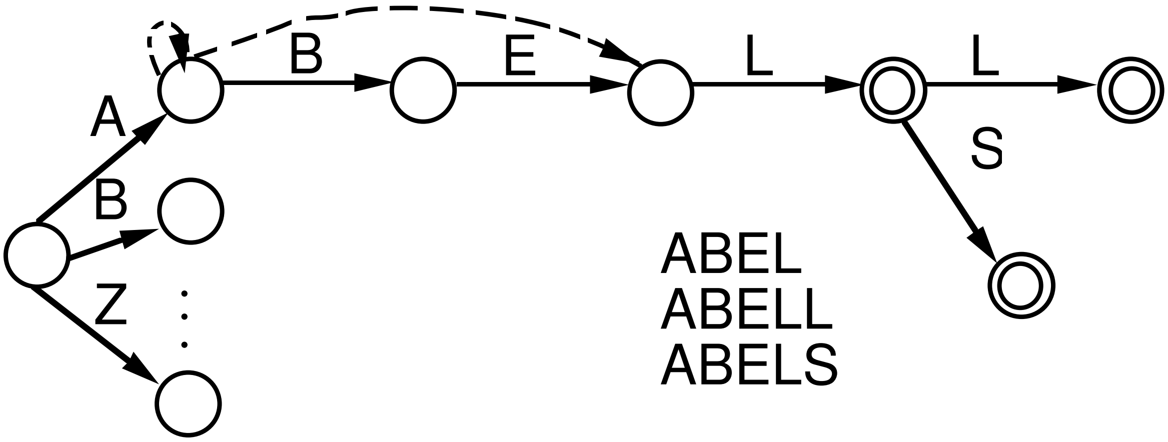


FEATURES

- PLP averaged over thirds of pre-son consonant
- PLP averaged over sevenths of sonorant
- PLP averaged over thirds of post-son region (up to 200 msec)
- Duration of pre-son consonant
- PLP □slices 6 & 10 msec after sonorant beginnin
- PLP □slices 6 & 10 msec before vowel-nasal boundary
- Zero crossing and amplitude profile of letter

Name Retrieval

- Output of letter classifier treated as a probability
- Letter insertions and deletions are allowed with penalty
- Names stored in tree for efficiency



System Performance

LETTER CLASSIFICATION:

- Training Set: 800 speakers
  - 400 alphabets from 400 speakers
  - 800 first & last names from 400 speakers
  - Total of 12,568 letters
- Testing Set: 400 speakers
  - 100 alphabets from 100 speakers
  - 300 last names
  - Total of 4,214 letters
- Results

Alphabet:	89.0% first choice .07% insertions
Names:	87.3% first choice 1.5% insertions

System Performance

NAME RETRIEVAL

- Three hundred spelled last names
- Searched database of 50,000 last names
- Results:

First Choice:	90.7%
Top Two	94.0%
Top Three	95.7%