A Prototype Voice-Response Questionnaire for the U.S. Census

Ronald A. Cole
David G. Novick
Mark Fanty
Pieter Vermeulen
Stephen Sutton
Dan Burnett
Johan Schalkwyk

Why automate the census?

- Increase responsiveness
  - Include illiterates
  - Non-English speakers
  - Disadvantaged persons
- Decrease cost
- Increase accuracy

Goals of Study

- Demonstrate feasibility of a telephone voice questionnaire for the year 2000 census
- Produce prototype systems for English and Spanish
- Develop a "Census Speech Corpus" and place in the public domain
- Create a testbed for future research
- Compare technologies (OGI, CMU)
- Proceed to 1995 trial

Census Task

Obtain the following information:

- Name
- Gender
- Marital status
- Date of birth
- Hispanic origin
- Race

Nature of Problem

- Telephone speech
- Speaker variability
- Background noise
- Creative, unexpected, unresponsive answers

Self-evaluating Questionnaire:

The Caller Reviews, Comments & Corrects

- System summarizes recognized information:
  "We recognized that you are a white male of non-hispanic origin born on July 1, 1946. Is this correct? Please say yes or no."
- Identify errors:
  If no ***, "Which of the following are incorrect? Race, sex, Hispanic origin, or date of birth?"***
- Repair dialogue corrects errors

Stages of Project

- Design and Test Protocols
- Data Collection and Transcription
- Develop Spoken Language System
- Evaluate and Refine System

Protocol Design

Design a protocol that is:

- Easy to use
- Produces (only) the desired responses

Good vs. Bad Prompts

- *"What is your marital status?"*
  "Unhappy"
  "On hold"

- *"Have you ever been married? Please say yes or no."*

If yes recognized:

"Are you now married, widowed, divorced, or separated?"

Data Collection

- Callers solicited by Census from 12 geographical regions
- Callers call 800-number
- Answer Census questions +evaluation questions
- Speech recognition used to complete questionnaire
- >4,000 complete calls

Transcription of Calls

- Exact Word-level transcription
  - boundaries of target words marked
  - includes pauses
  - background noises
- Behavioral Coding Scheme
  - 10 levels of compliance
  - used for analyzing prompts
- 1200 calls transcribed and given behavior codes

Time-aligned Word-level Labeling

1. Aligned to the waveform
2. Mark beginning and ending of each word
3. Label any non-speech information

Behavioral Coding Scheme

Examples of behavioral codes for marital status question:

Question: "Have you ever been married?"

AA1	'yes'
AA2	'yes I am'
AA3	'unfortunately'
QA	'well, sort of'
IA1	'have you?'
IA2	<says nothing>
IA3	<garbled speech>
RC	'does common law marriage count?
IN	'..s' (interrupts)
DK	'I'm not sure'
RF	'that's none of your business'
O	'what was the question again?'

Percentage of Census Responses Which are Informative

Control Flow: Current System

Dialogue Control

- Generates prompt for the current state (e.g., "Are you female or male?")
- Passes caller's response to speech recognizer with recognition grammar for current state

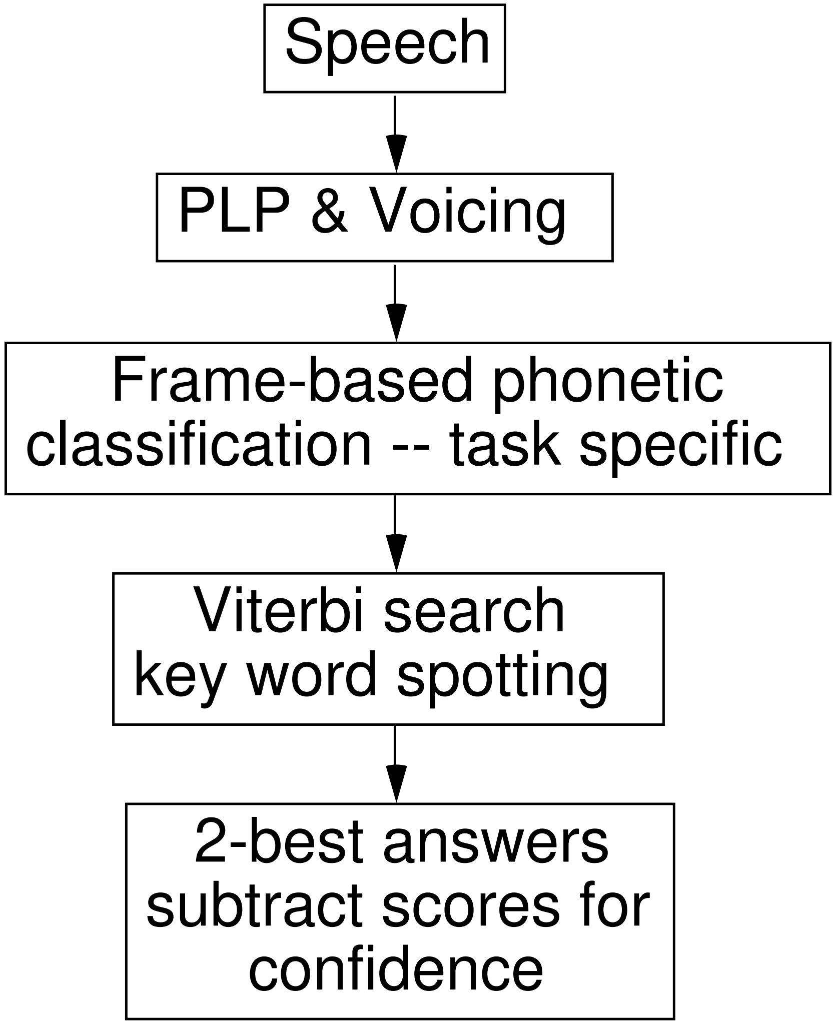
Speech Module

- Computes PLP features every 6 msec
- Neural network uses PLP features to assign 39 phoneme scores to each frame
- Viterbi search finds best matching legal word sequence
    - word models have alternate pronunciations
    - garbage models match noise and out of vocabulary words
- Return recognized word sequence to dialogue control

Dialogue Control

- Change state based on caller's response (e.g., if no response, stay in state to repeat same question)

Recognition



Speech

PLP & Voicing

Frame-based phonetic classification -- task specific

Viterbi search
key word spotting

2-best answers
subtract scores for confidence

Census Word Spotting

Digital signal

Endpoint detection

Voicing classifier

PLP analysis

Phoneme classification of 6-msec frames (phoneme set customized to the task)

Search phoneme probability matrix for 2 best scoring target words

Confidence is average frame score oftop word divided by average frame score of second word

Census Word Spotting

Grammer:

ANY ⟶ <Target Word> ⟶ ANY

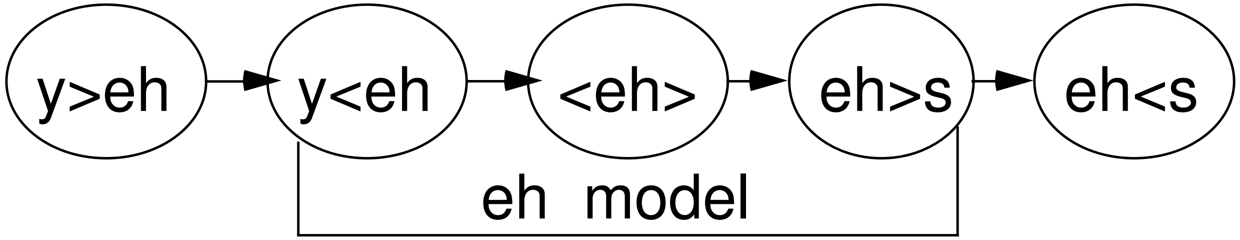
ANY = .pau (silence)
s (hiss)
.garbage (4th largest phoneme score for each frame)

.garbage is intended to match extraneous speech

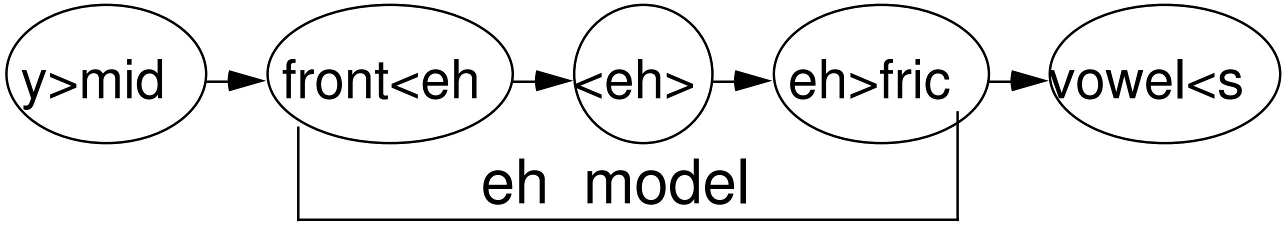
Context-Dependent Modeling

yes /y eh s/

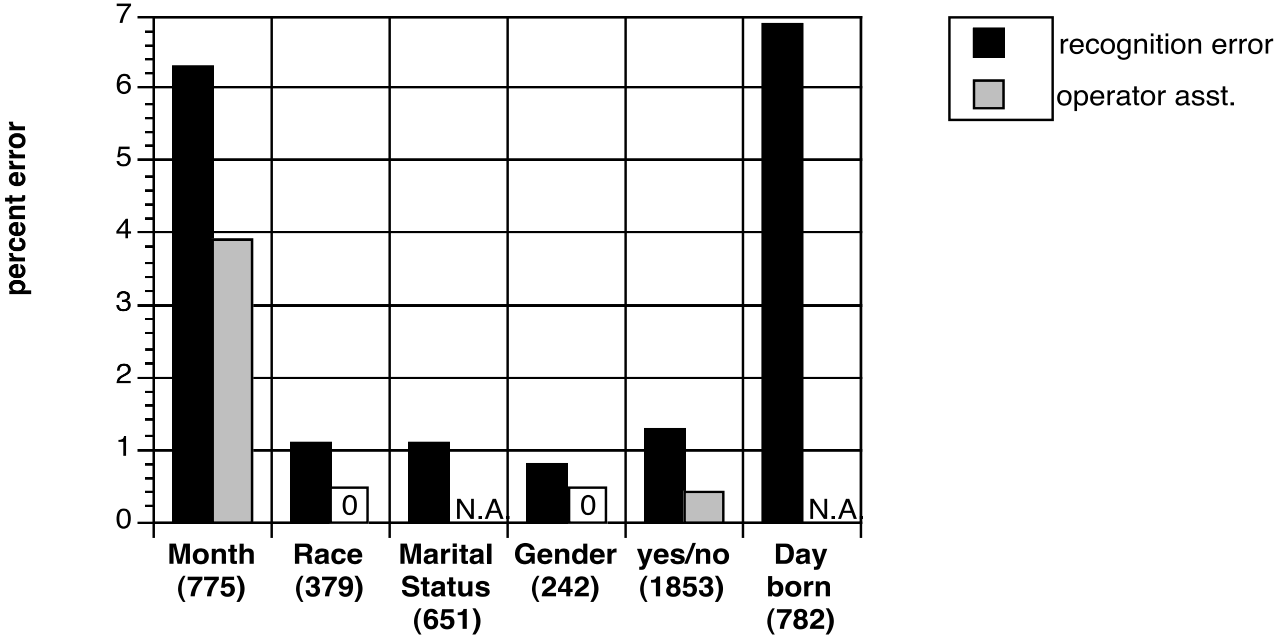
Task dependent



Task independent



Recognition Performance



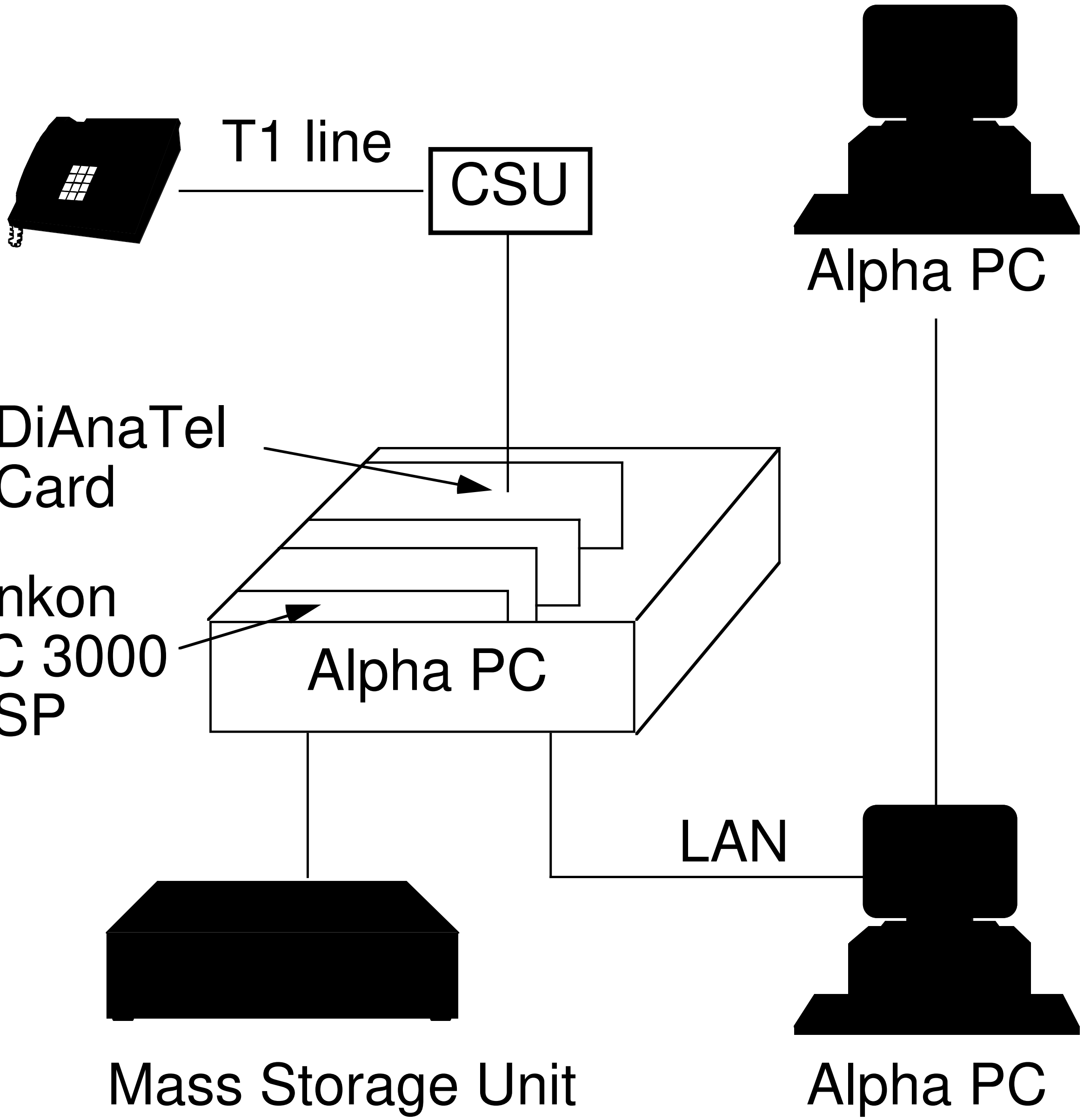
5% of calls needing review

Example: Yes/ No: 1829/1853

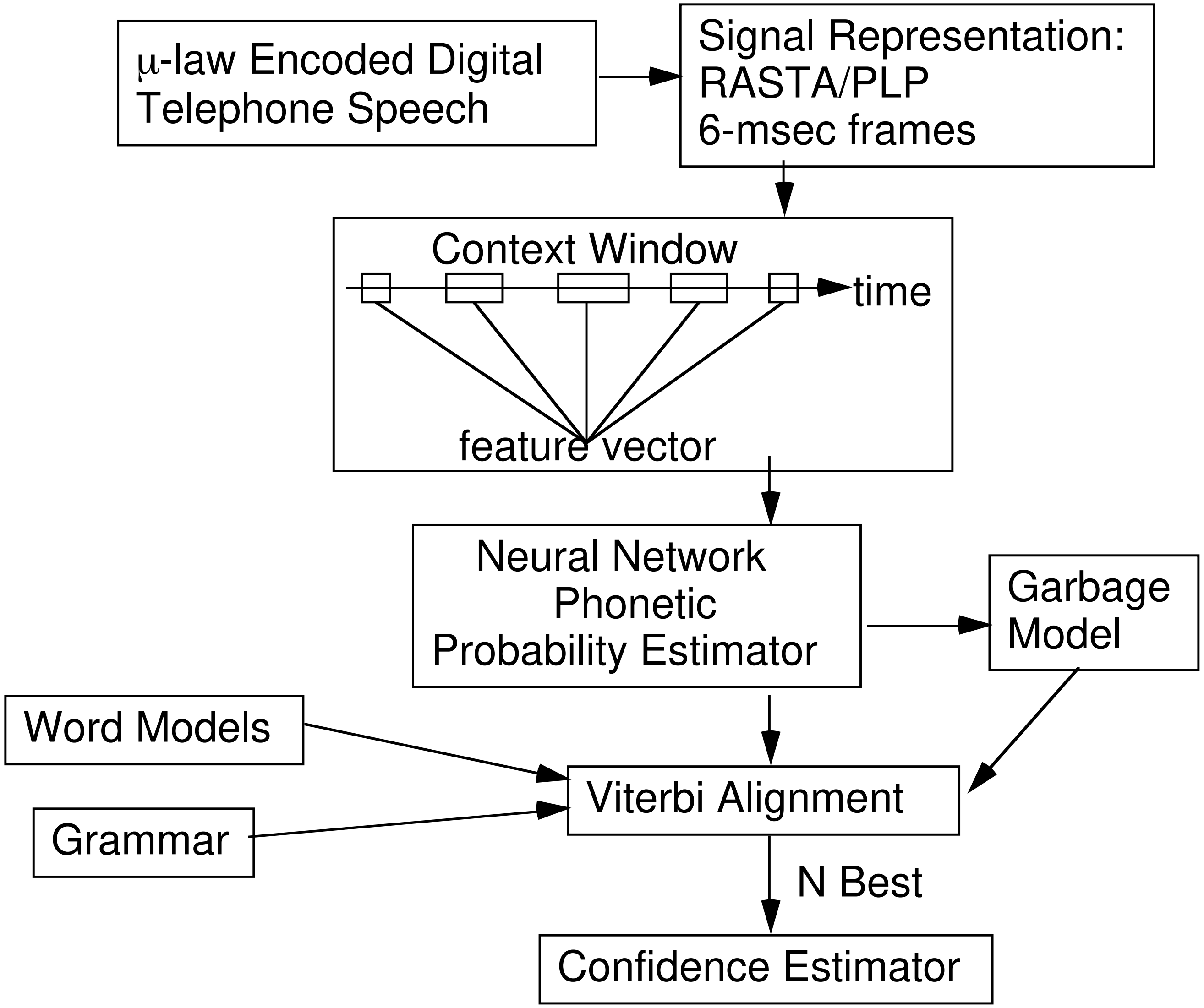
	accept	reject
correct	1737	92
incorrect	8	16

Phase 2 System Capabilities
1995 Pre-test

- Prototype systems for 24 simultaneous calls
- Spanish & English
- Accurate, robust, graceful dialogue, self evaluating
- Operator hand off


T1 line
CSU
Alpha PC
DiAnaTel
Card
Linkon
FC 3000
DSP
Alpha PC
LAN
Mass Storage Unit
Alpha PC

Census System Components



μ-law Encoded Digital Telephone Speech

Signal Representation:
RASTA/PLP
6-msec frames

Context Window

Neural Network
Phonetic
Probability Estimator

Garbage Model

Word Models

Grammar

Viterbi Alignment

N Best

Confidence Estimator