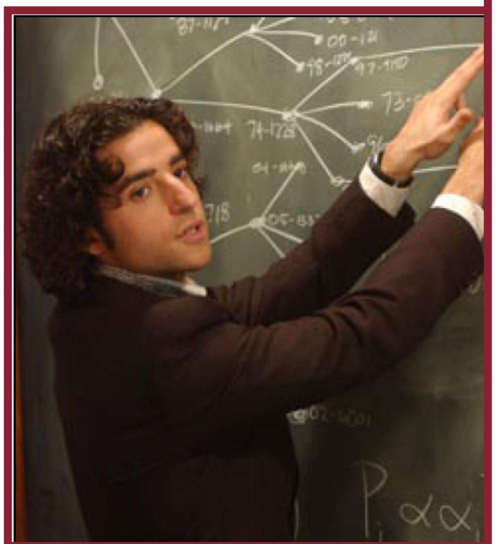
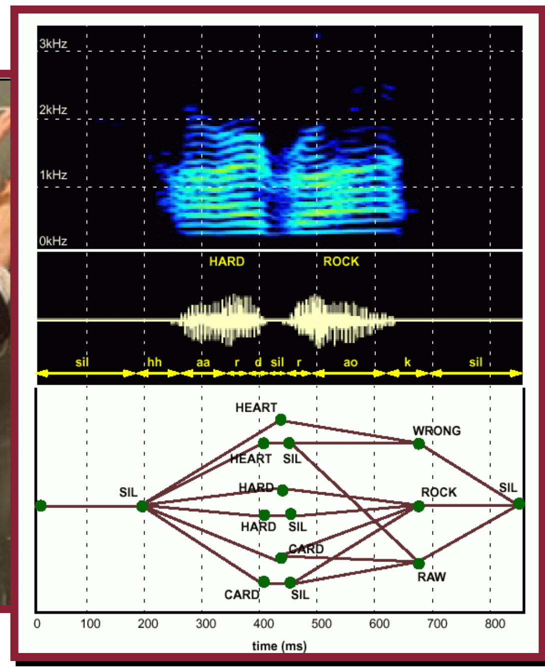
National Security by the *Numb3rs:*

The Important Role of Mathematics in Human Language Technology





Joseph Picone, PhD

Department of Electrical and Computer Engineering

Mississippi State University

Abstract and Biography

ABSTRACT:

In elementary school, my junior high school math and science teachers used to argue about whether "Math is the Queen of Science." In high school, it was fashionable for students to lose interest in math in favor of law and medicine. In college, you are often told there is no future in math other than teaching.

However, a day does not go by that math doesn't play a vital role in every aspect of life. This is particularly true for national security, where mathematics and human language form the basis for our worldwide leadership in information technology. Television shows such as Numb3rs are popularizing the critical role that mathematics plays in solving crimes and understanding human behavior, creating renewed appreciation for those who pursue careers in math.

In this talk, we will discuss ways that math has made a difference in applications of human language technology in the defense industry. We will discuss how human language can be decoded by computer using mathematical models that extract language, speaker identity, and even the words that are spoken.

BIOGRAPHY:

Dr. Joseph Picone graduated from Fenwick High School in Chicago, Illinois in 1975, where he first dreamed of getting paid for solving calculus problems. He subsequently received his Ph.D. in Electrical Engineering from Illinois Institute of Technology in 1983, where he first dreamed of getting paid for solving speech research problems using calculus.

Dr. Picone is currently a Professor in the Department of Electrical and Computer Engineering at Mississippi State University, where he does get paid for solving speech research problems using calculus. His principal research interests are the development of new statistical modeling techniques for speech recognition. He has previously been employed by Texas Instruments and AT&T Bell Laboratories, where he also solved speech research problems using calculus.

Dr. Picone is currently on assignment at the Department of Defense where he directs applied research and development of technology for information retrieval from voice. After more than 20 years, he is still having fun every day, and getting paid, for solving speech research problems using calculus.

URL: http://www.ece.msstate.edu/research/isip/publications/seminars/external/2006/msms/

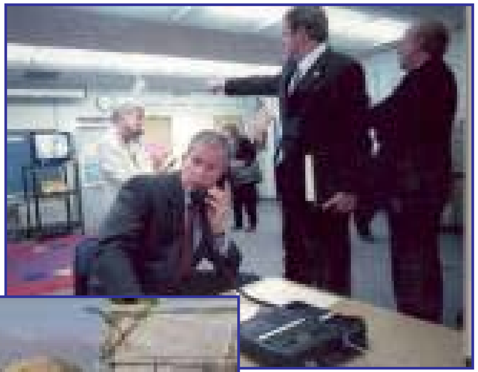
National Security Agency (www.nsa.gov)

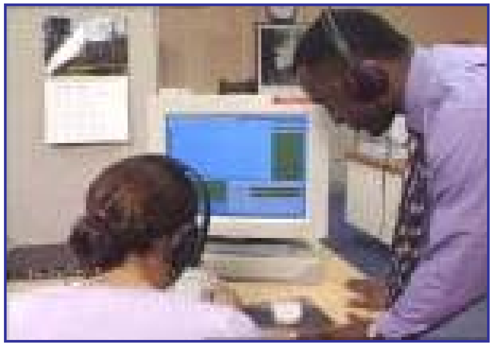
- On November 4, 1952, President Harry S. Truman signed National Security Council Intelligence Directive no. 9.
- For more than 50 years, NSA has diligently conducted its two-fold mission to provide and protect vital communication information.

- The National Cryptologic Museum and the National Vigilance Park honor those who have “served in silence”

Codemaking and Codebreaking

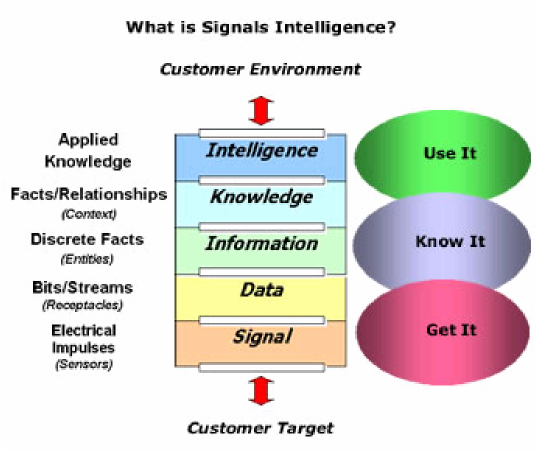
- Mathematicians, linguists, engineers, and computer scientists focus on two core missions: Information Assurance (codemaking) and Signals Intelligence (codebreaking).
- Information Assurance prevents America's adversaries from exploiting sensitive U.S. government communications.
- Signals Intelligence involves exploiting foreign adversaries' communications.





- NSA customers use this information to counter terrorism, prevent the proliferation of weapons of mass destruction and thwart the flow of narcotics into our country.
- NSA's missions depend on the expertise of a joint military and civilian workforce.

Signals Intelligence

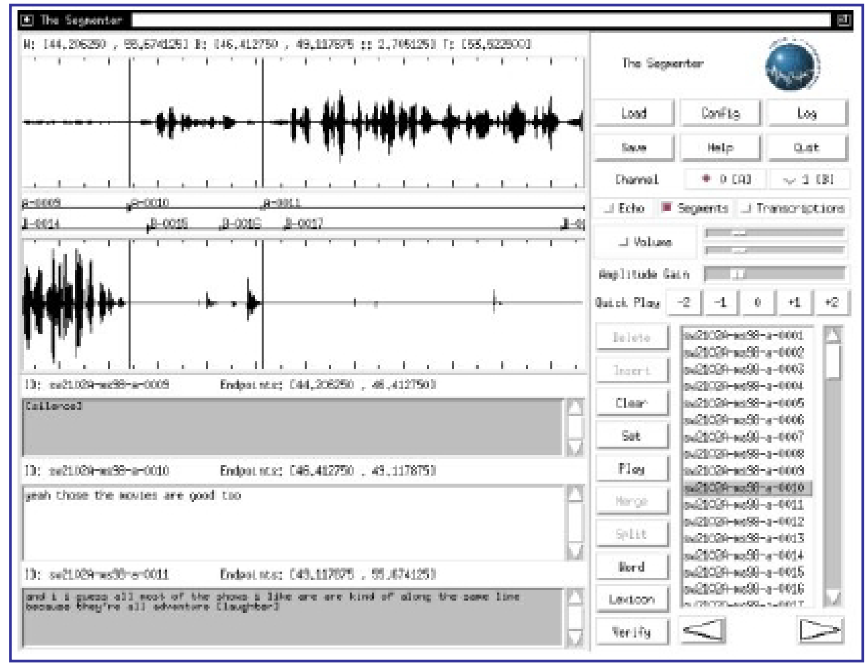


- The National Security Agency collects, processes and disseminates foreign Signals Intelligence (SIGINT).
- "Knowledge Is Power"
- SIGINT is derived from three sources:
  - Human Intelligence (HUMINT) is primarily the responsibility of the CIA and DIA
  - Imagery Intelligence (IMINT) belongs to NGA
  - Military Intelligence and Measurement and Signature Intelligence (MASINT) belongs to DIA.

- The many languages used in the nations and regions of the world that are of interest to our military and national leaders require NSA to maintain a wide variety of language capabilities.
- Successful SIGINT depends on the skills of not only language professionals but those of mathematicians, analysts, and engineers, as well.

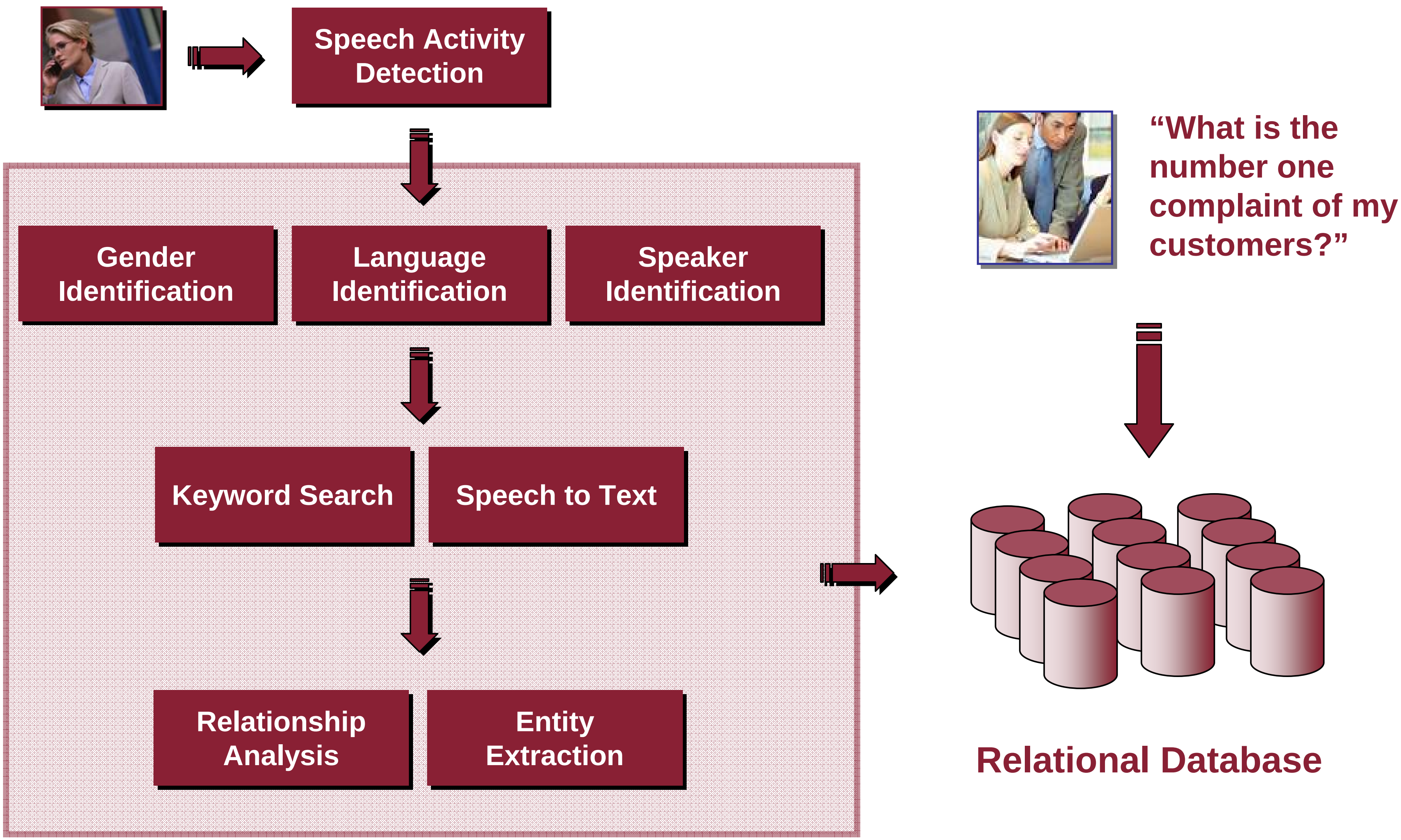
Information Retrieval From Voice

- Metadata extraction from conversational speech
- Automatic gisting and intelligence gathering
- Speech to text is the core technology challenge
- Machines vs. humans



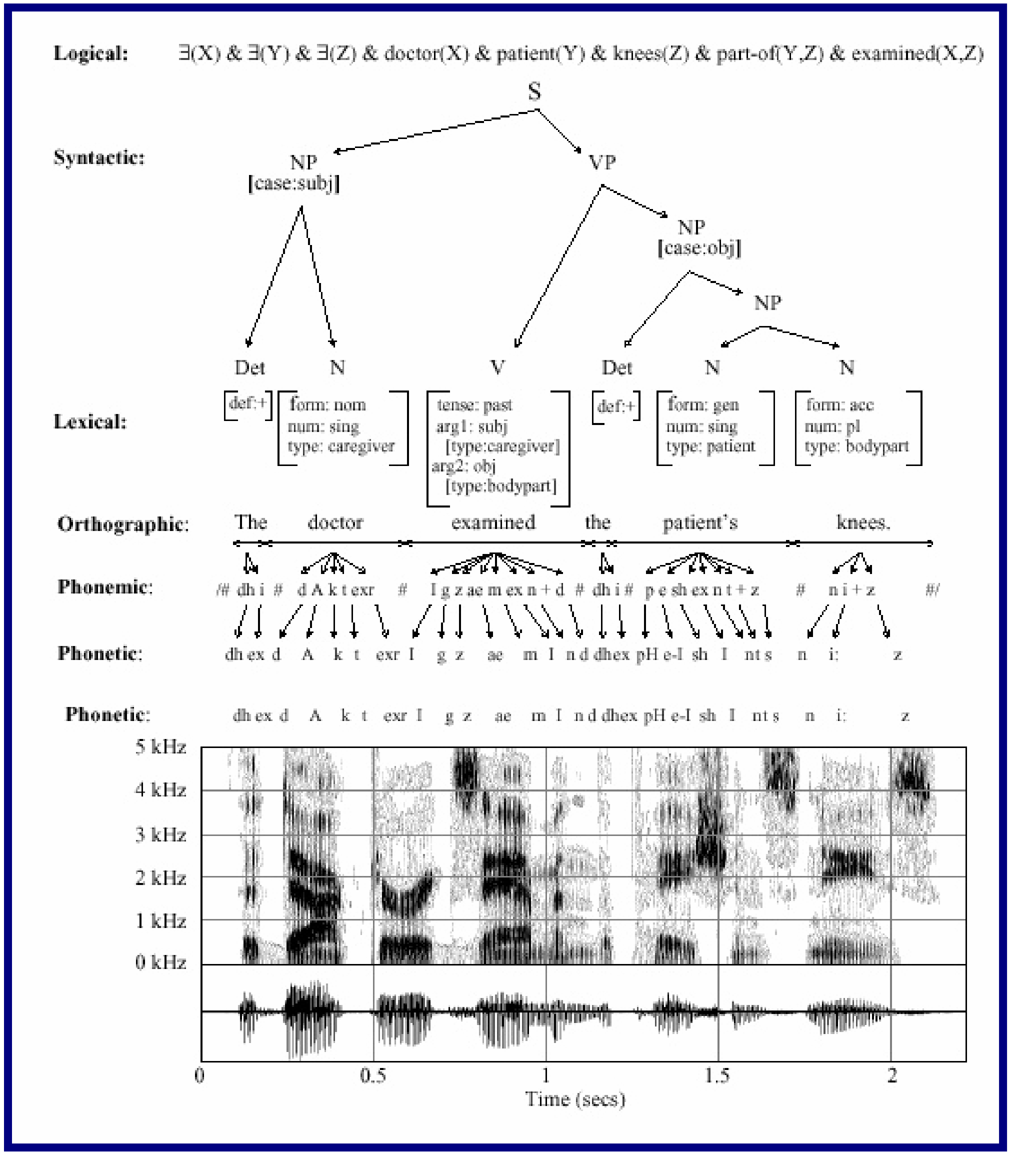
- Real-time audio indexing
- Time-varying channel
- Dynamic language model
- Multilingual and cross-lingual

Core Technologies in Information Retrieval From Voice

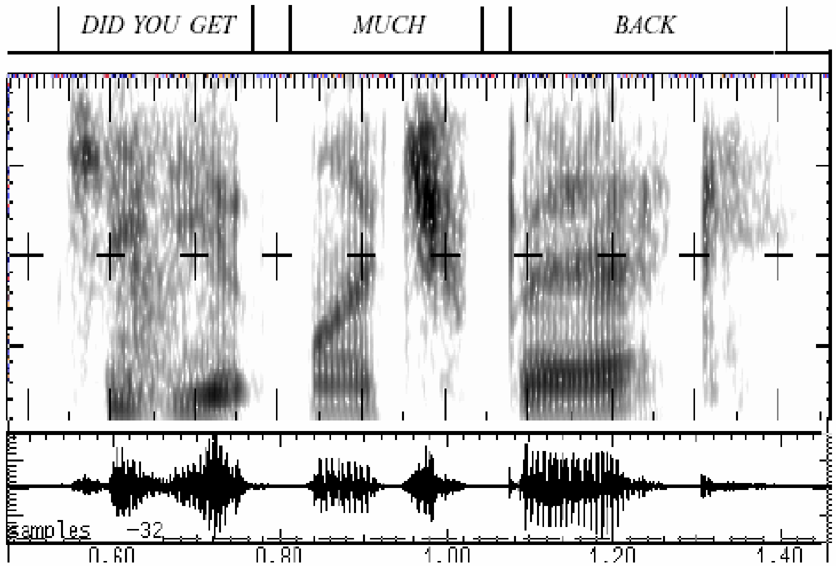


Information Retrieval Requires Understanding

- **Traditional Output:**
  - best word sequence
  - time alignment of information
- **Other Outputs:**
  - word graphs
  - N-best sentences
  - confidence measures
  - metadata such as speaker identity, accent, and prosody
- **Applications:**
  - Information localization
  - data mining
  - emotional state
  - stress, fatigue, deception

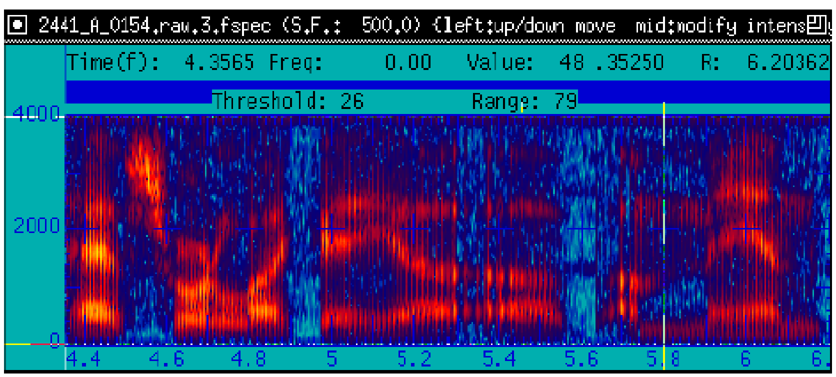


Closer Look: What Makes Speech to Text So Hard?



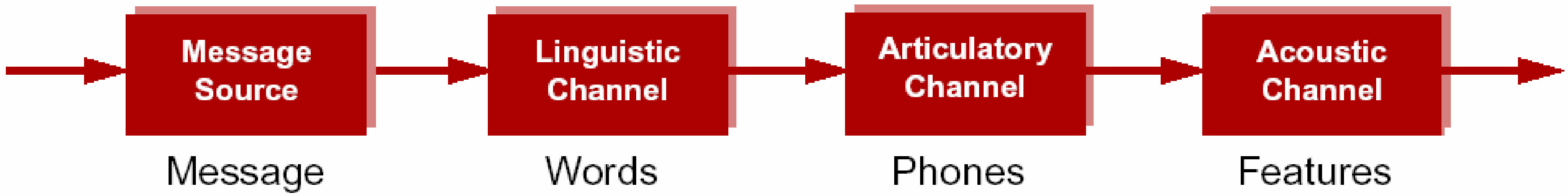
- "Did you get" is reduced such that the resulting word is pronounced "jyuge."
- Phoneme deletion rate: ~12%
  Syllable deletion rate: ~1%
- Predicting pronunciations of words is crucial!

"have sort of like a a a manpower"



- Conversational speech defies conventional grammatical structure.
- Constrained interfaces have failed!

Statistical Approach: Noisy Communication Channel Model



Bayesian formulation for speech recognition:

$$P(W|A) = P(A|W)P(W)/P(A)$$

Objective: minimize the word error rate by maximizing $P(W|A)$

Approach: maximize $P(A|W)$ (training)

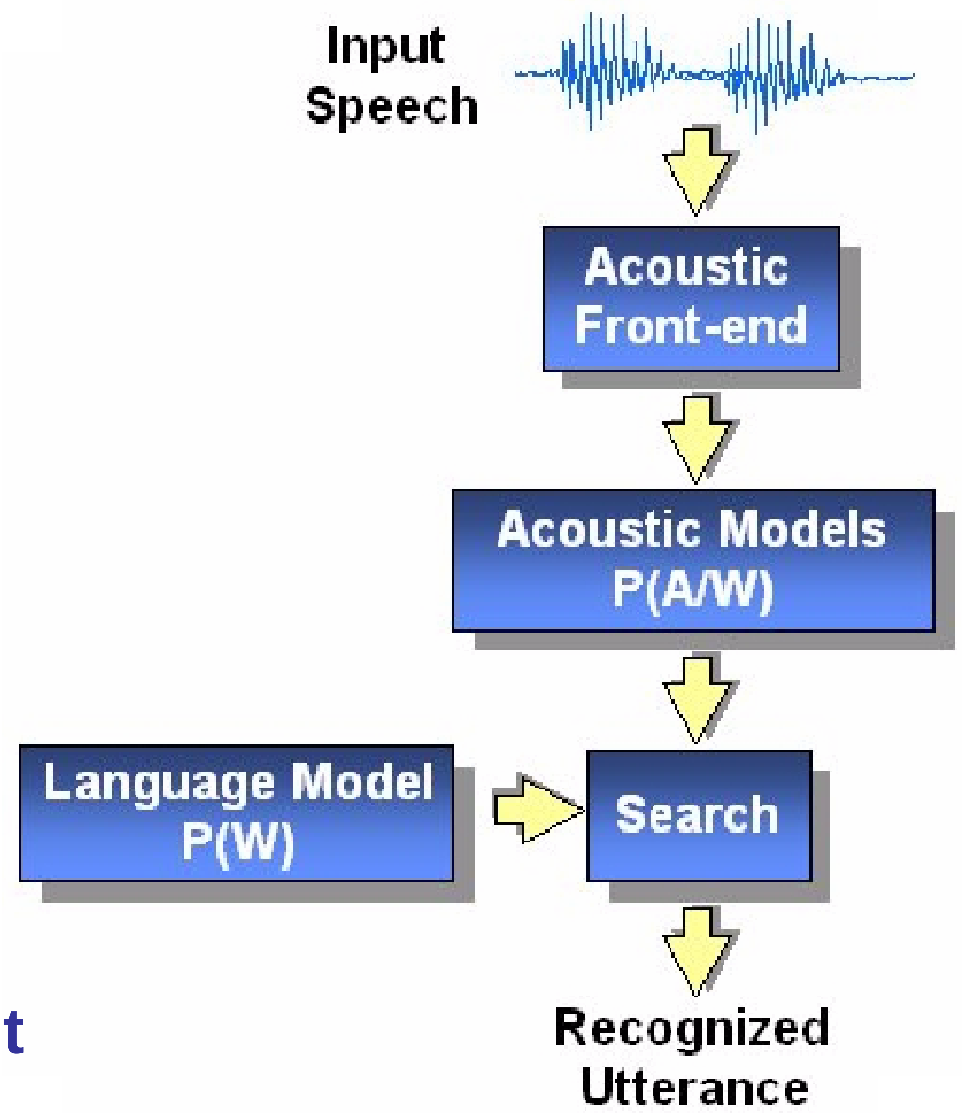
- $P(A|W)$: acoustic model (hidden Markov models, Gaussians)
- $P(W)$: language model (finite state machines, N-grams)
- $P(A)$: acoustics (ignore during maximization)

Speech to Text Architectures

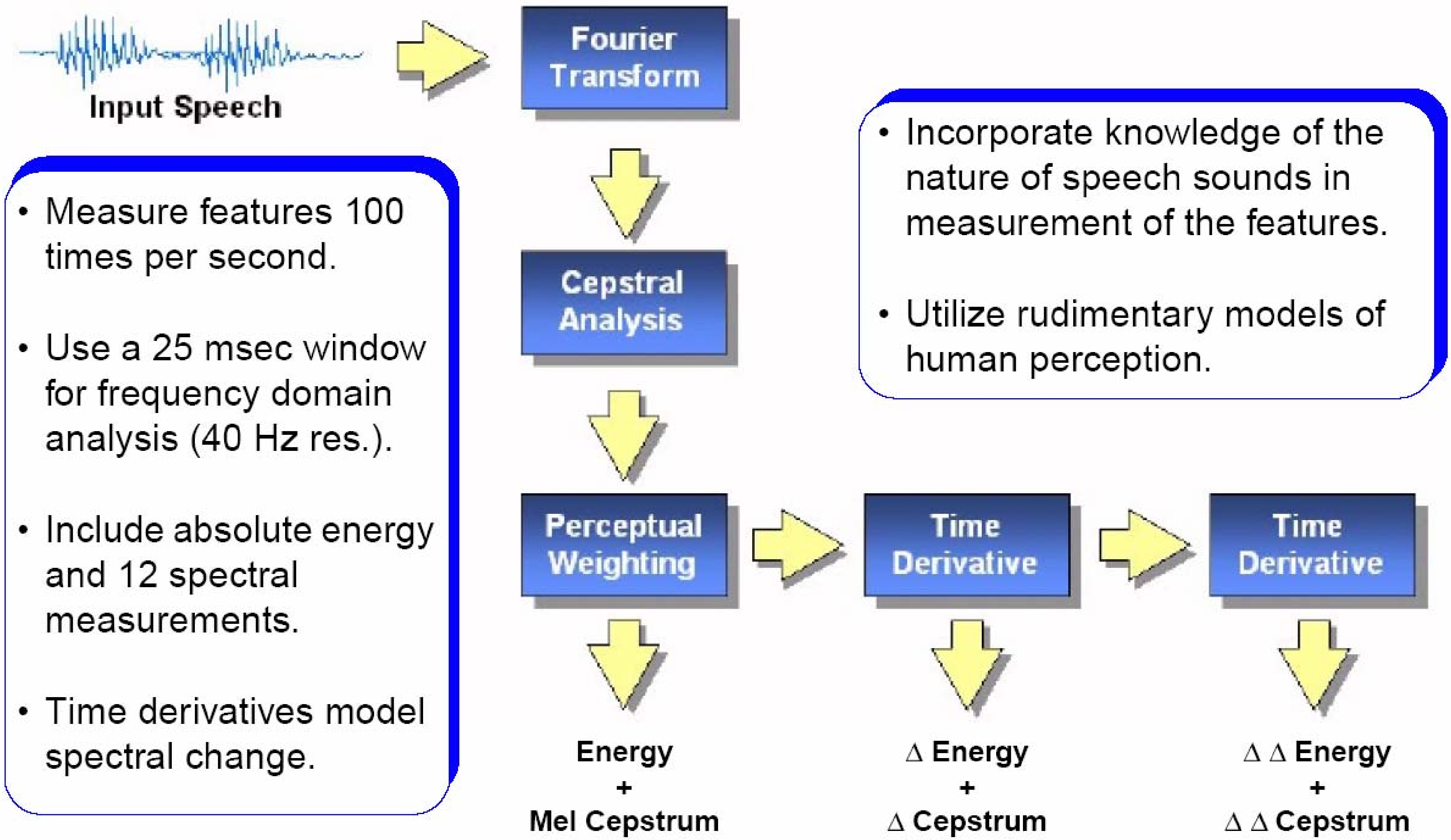
Core components:

- transduction
- feature extraction
- acoustic modeling (hidden Markov models)
- language modeling (statistical N-grams)
- search (Viterbi beam)
- knowledge sources

Classic engineering approach: decompose a problem into smaller disjoint problems. Does this work?

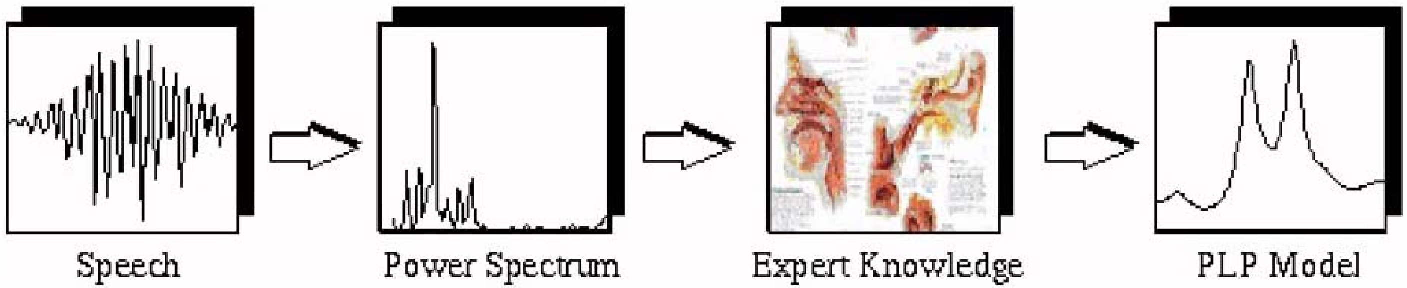


Feature Extraction in Speech to Text

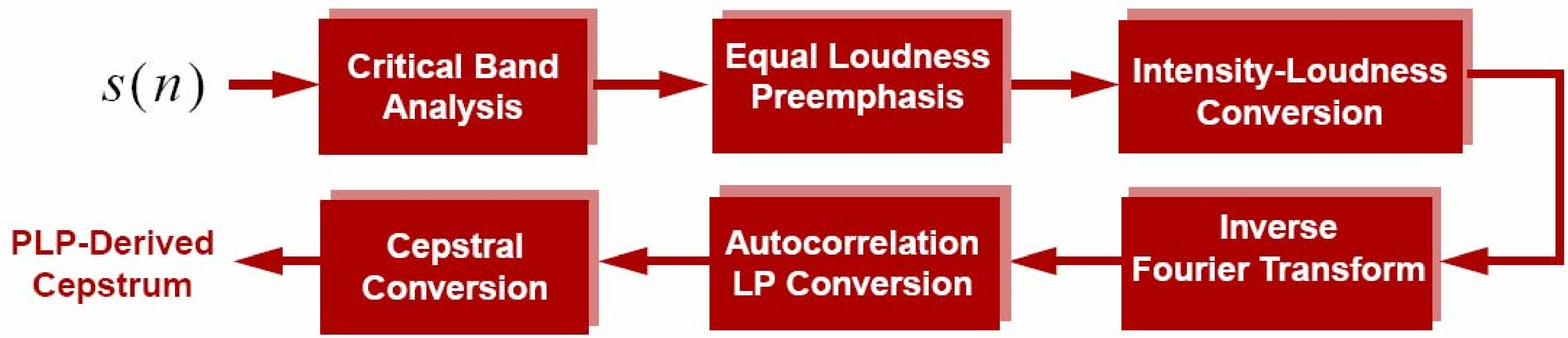


Adding More Knowledge to the Front End

- Incorporate more knowledge about the physics of speech:



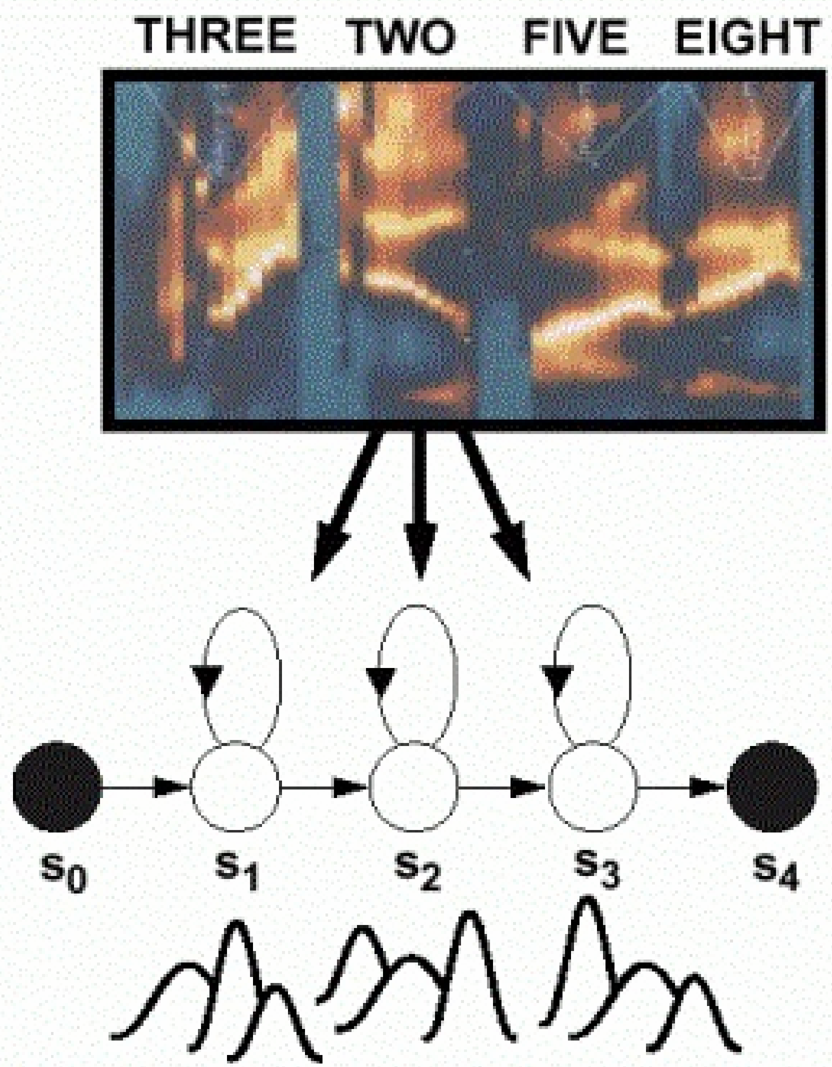
- Processing steps are similar to conventional analysis:



- Word error rate (WER) reduction is very small.

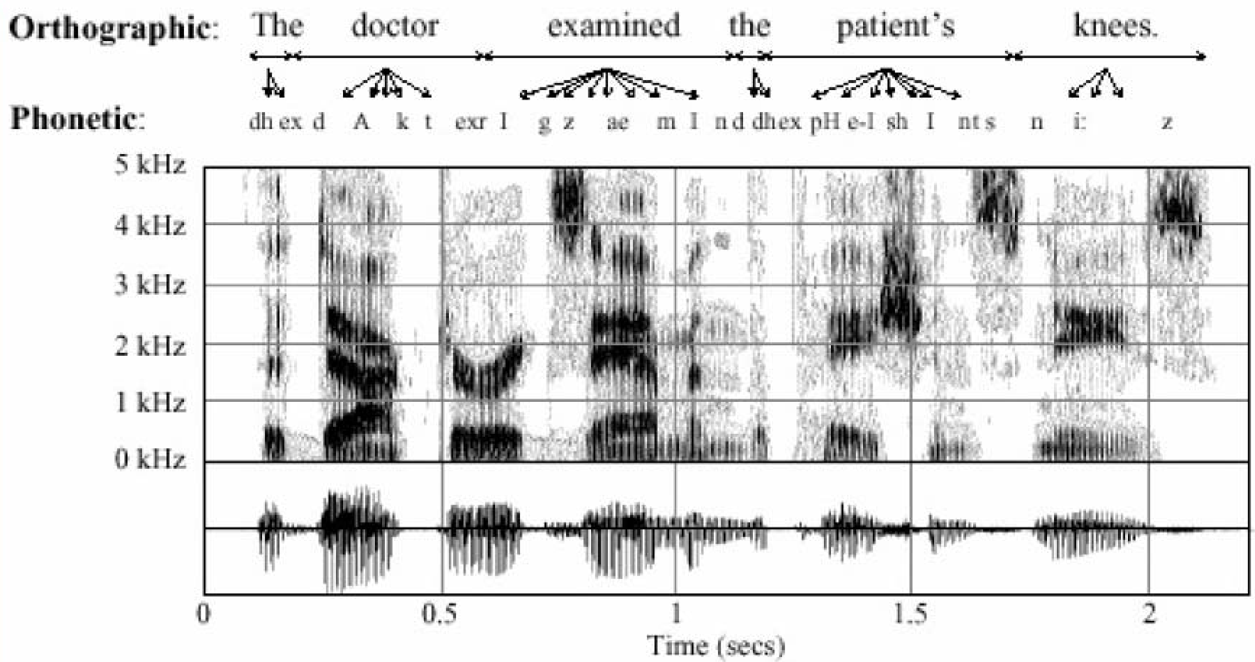
Acoustic Modeling: Hidden Markov Models

- Acoustic models encode the temporal evolution of the features (spectrum).
- Gaussian mixture distributions are used to account for variations in speaker, accent, and pronunciation.
- Phonetic model topologies are simple left-to-right structures.
- Skip states (time-warping) and multiple paths (alternate pronunciations) are also common features of models.
- Sharing model parameters is a common strategy to reduce complexity.



Context-Dependent Acoustic Units

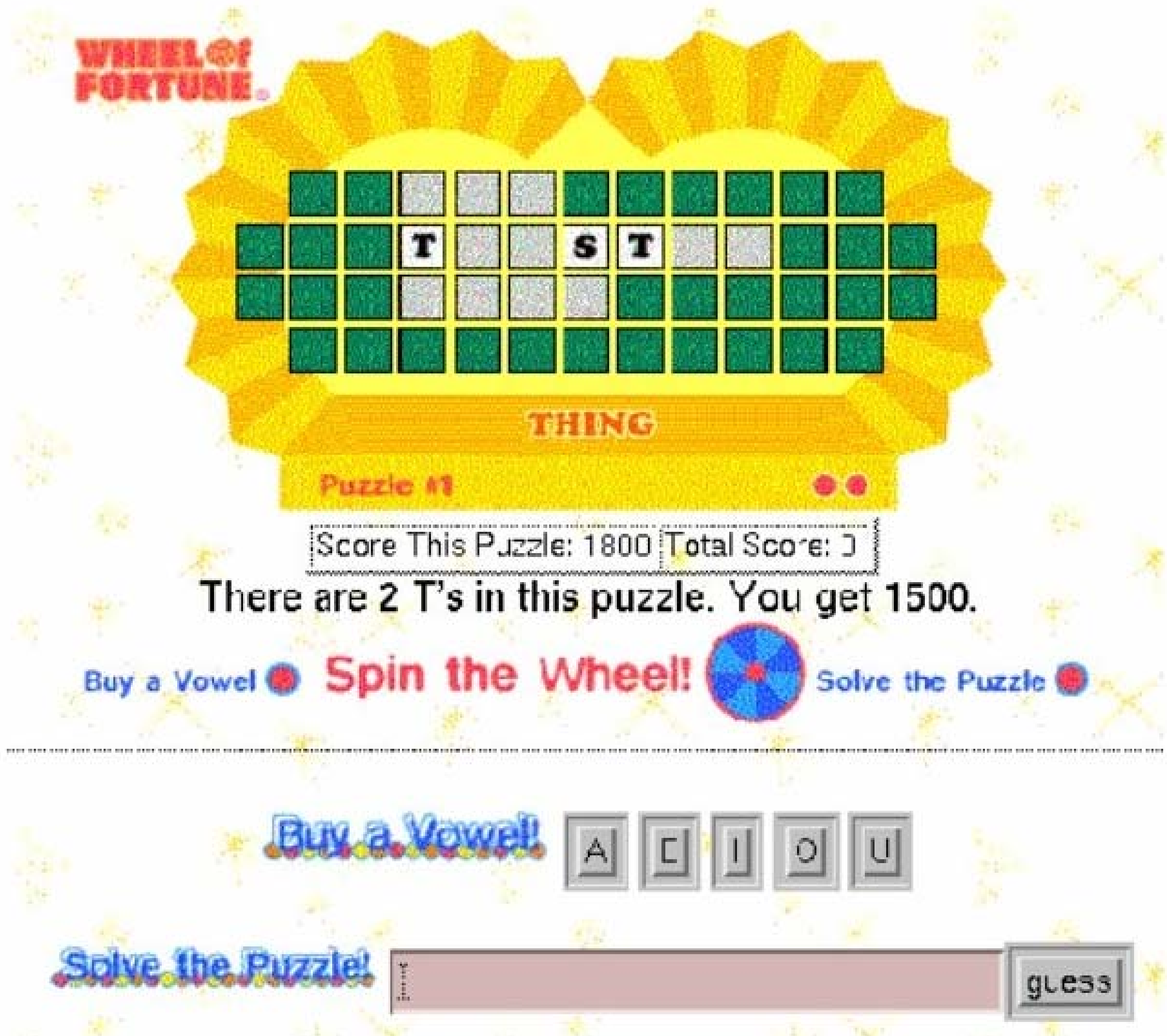
- **Phonetic units are preferred.**
- **Training does not require phonetic transcriptions.**
- **Many types of phonetic units.**
- **Cross-word units add complexity.**



[silence] BEAT IT [silence]

Monophone Modeling	sil	b	iy	t	ih	t	sil
Word internal triphone modeling	sil	b+iy	b-iy+t	iy-t	ih+t	ih-t	sil
Crossword triphone modeling	sil	sil-b+iy	b-iy+t	iy-t+ih	t-ih+t	ih-t+sil	sil

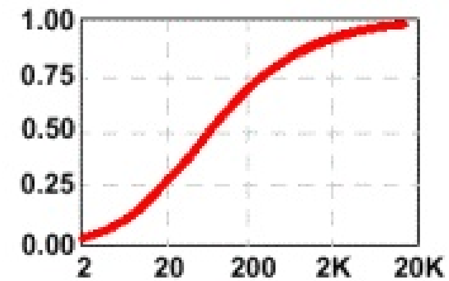
Language Modeling Involves Word Prediction



N-gram Modeling is Very Powerful

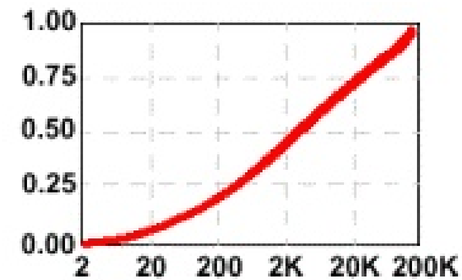
Unigrams (SWB):

- Most Common: I, and, the , you, a
- Rank-100: she, an, going
- Least Common: Abraham, Alastair, Acura



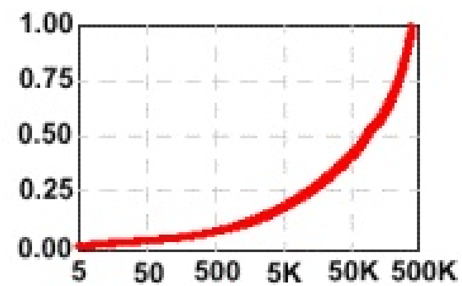
Bigrams (SWB):

- Most Common: “you know”, “yeah S!”, “!S um-hum”, “I think”
- Rank -100: “do it”, “that we”, “don’t think”
- Least Common: “raw fish”, “moisture content, “Reagan Bush”



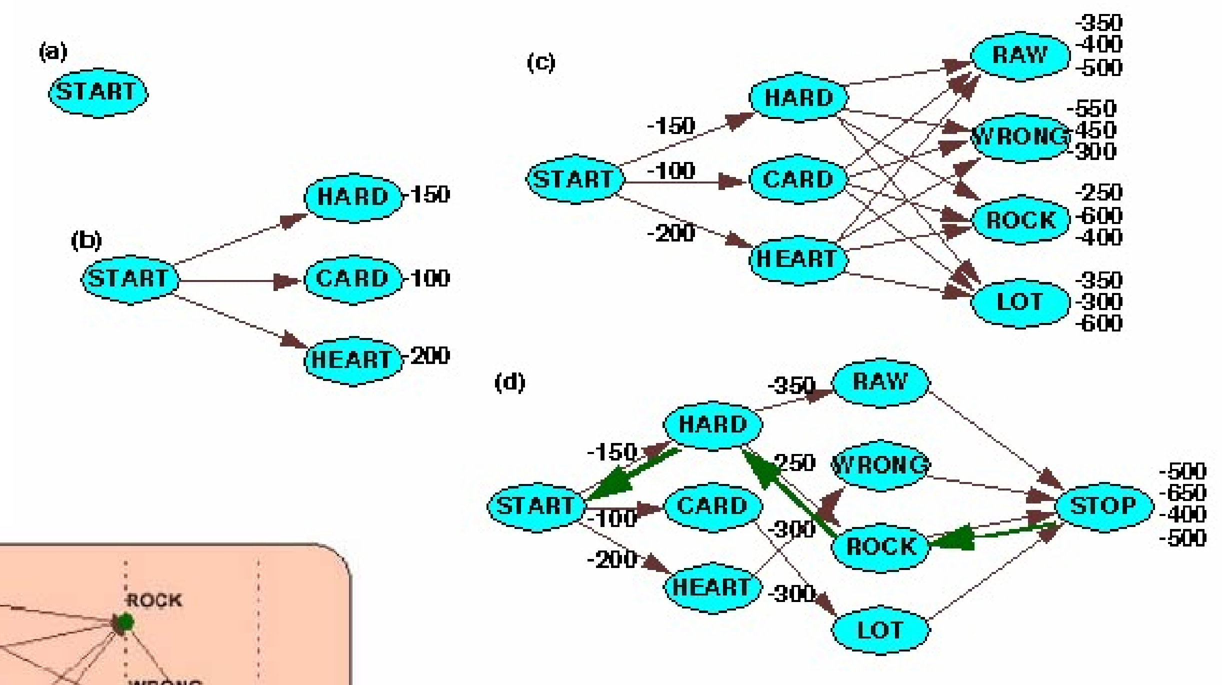
Trigrams (SWB):

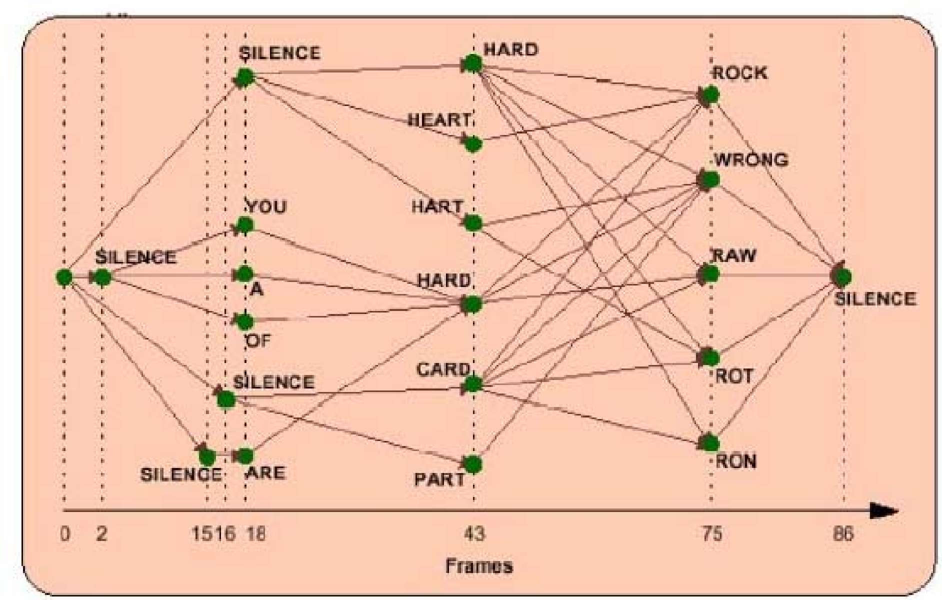
- Most Common: “!S um-hum S!”, “a lot of”, “I don’t know”
- Rank-100: “it was a”, “you know that”
- Least Common: “you have parents”, “you seen Brooklyn”



Search Involves Topology and Graph Theory

- **breadth-first**
- **time synchronous**
- **beam pruning**

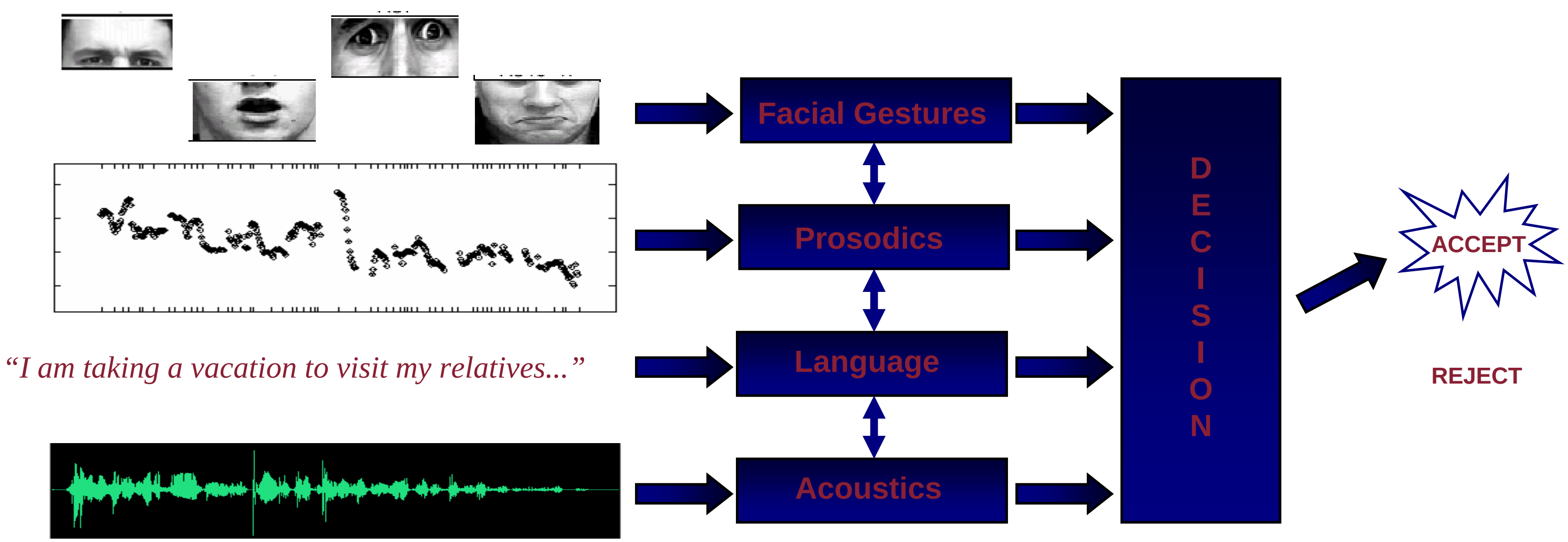




- **supervision**
- **word prediction**
- **natural language**

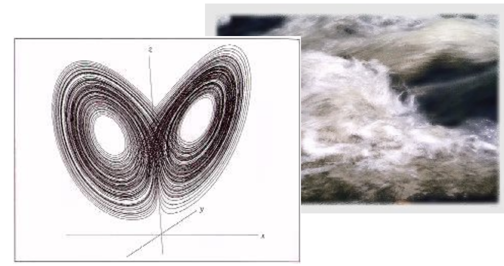
Detection of Stress and Fatigue Detection

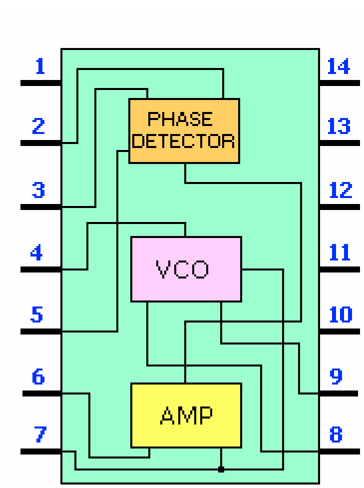
- **Recognition of emotion, stress, fatigue, and other voice qualities are possible from enhanced descriptions of the speech signal**
- **Applications range from call centers to homeland security**



Nonlinear Statistical Modeling of Speech

"Though linear statistical models have dominated the literature for the past 100 years, they have yet to explain simple physical phenomena."

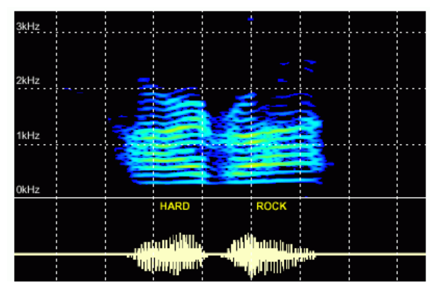




- **Motivated by a phase-locked loop analogy**
- **Application of principles of chaos and strange attractor theory to acoustic modeling in speech**
- **Baseline comparisons to other nonlinear methods**

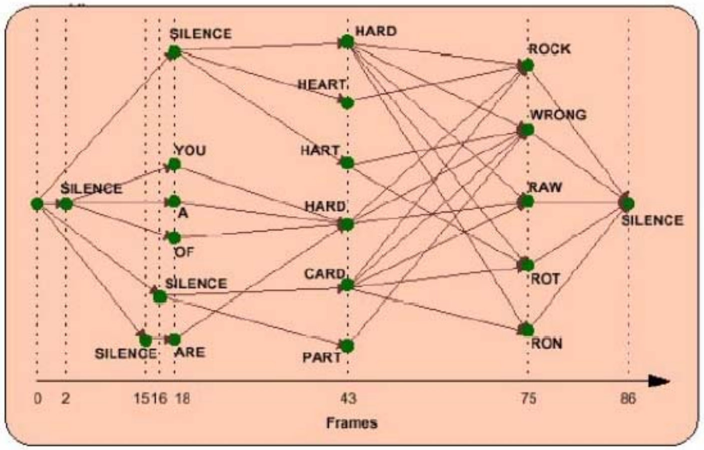
Expected outcomes:

- **Reduced complexity of statistical models for speech (two order of magnitude reduction)**
- **High performance channel-independent text-independent speaker verification/identification**



Summary

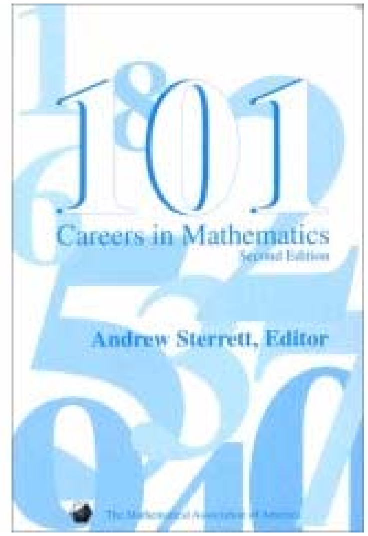
- Mathematical models of a speech signal drive our ability to extract information.
- Mathematical models of language/discourse drive our ability to understand speech.

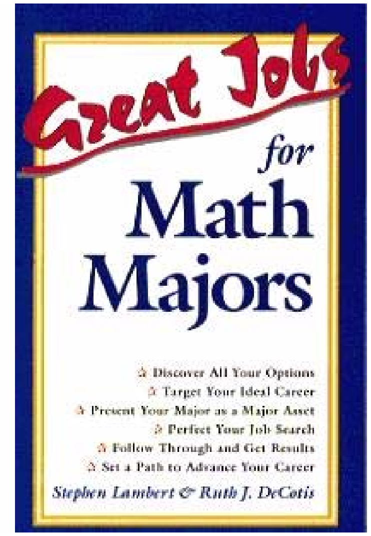


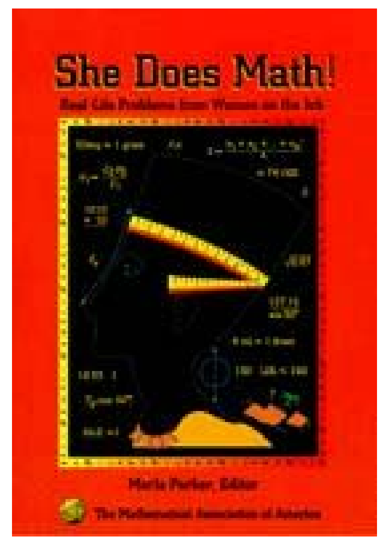
- It is still a challenge to use mathematics to model linguistic knowledge of the world's languages.
- The integration of these technologies are providing a wealth of applications, including smart portable devices and Internet-based information services.

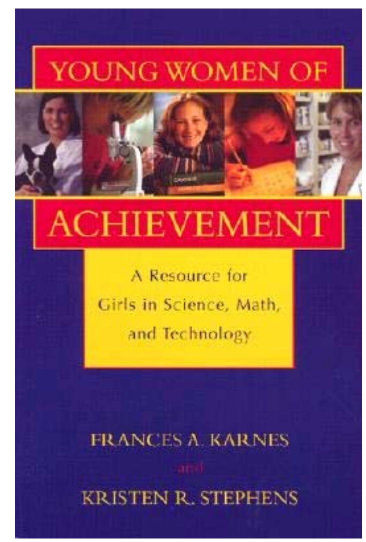
Appendix: To Learn More…

1. A. Sterrett, *101 Careers in Mathematics*, Mathematical Association of America (MAA), Washington, D.C., USA, ISBN: 0883857286, 2003.
2. S. Lambert and R. DeCotis, *Great Jobs for Math Majors*, McGraw-Hill, New York, New York, USA, ISBN: 0844264229, 1998.
3. *She Does Math!: Real-Life Problems from Women on the Job*, Mathematical Association of America (MAA), Washington, D.C., USA, ISBN: 0883857022, 1995.
4. F. Karnes and K. Stephens, *Young Women of Achievement: A Resource for Girls in Science, Math and Technology*, Prometheus Books, Amherst, New York, USA, ISBN: 1573929654, 2002.

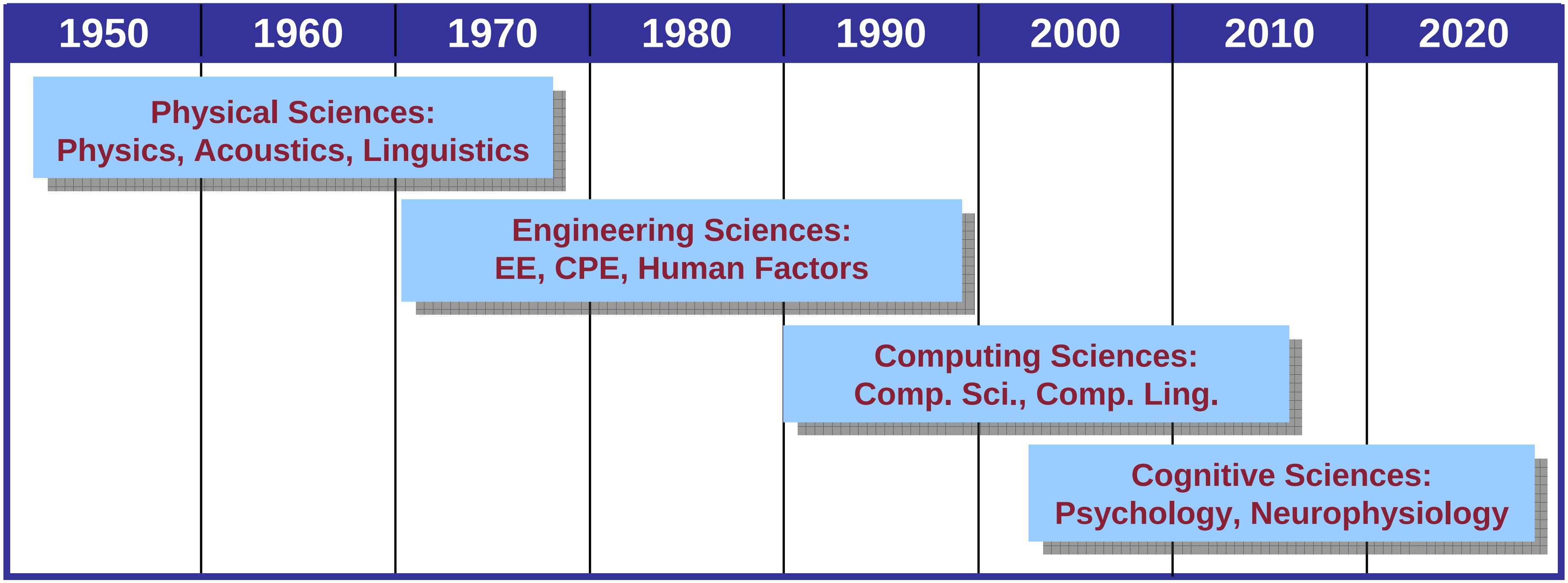








Appendix: Historical Perspective of Prominent Disciplines



Observations:

- **Field continually accumulating new expertise.**
- **As obvious mathematical techniques have been exhausted ("low-hanging fruit"), there will be a return to basic science (e.g., fMRI brain activity imaging).**