

A decorative graphic on the left side of the slide, consisting of a vertical black line intersecting a horizontal black line. To the left of the vertical line are three overlapping squares: a blue one on top, a red one on the left, and a yellow one on the bottom. To the right of the vertical line are three vertical bars of varying heights and colors: a blue one, a white one, and a yellow one.

How to Wreck a Nice Beach

Theory and Practice

Paul Hsu

CSAIL Spoken Language Systems

March 6, 2007

Speech Recognition Today



■ Dictation

- Transcribe spoken words to text
- Support punctuation and correction
- Dragon NaturallySpeaking (2004)



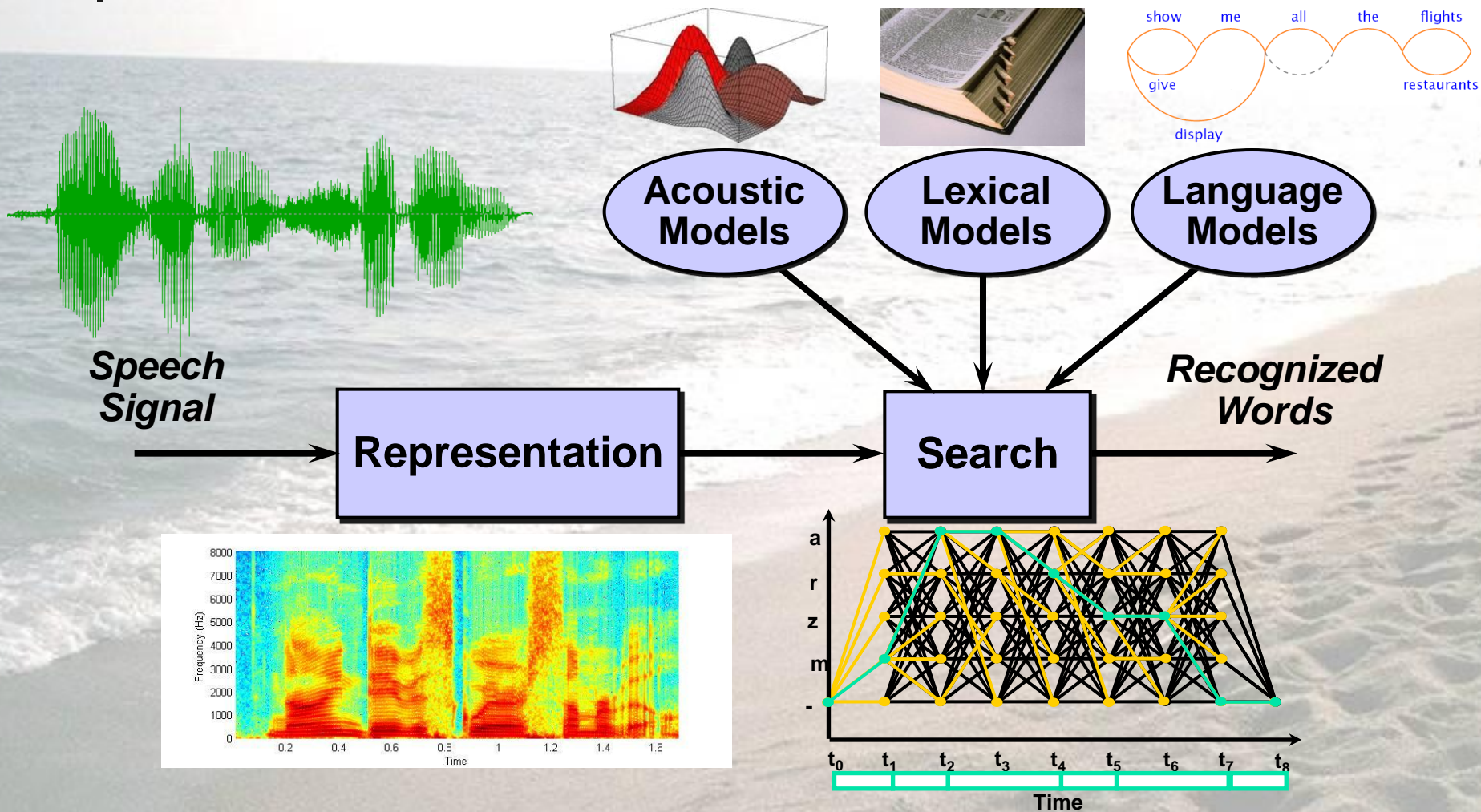
■ Interactive Voice Response

- System-initiated dialog
- Saturday Night Live Mock (2005)

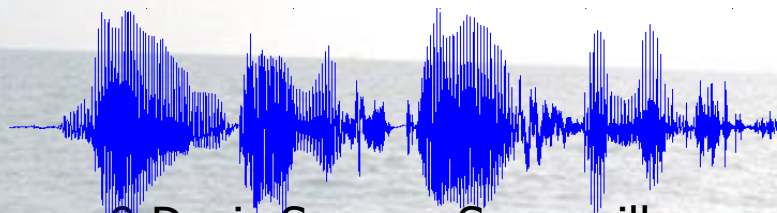
A photograph of a beach scene. The ocean is on the left, with waves crashing onto the shore. The sand is on the right, with several footprints visible. The sky is bright and overcast. The word "Theory" is written in a large, blue, sans-serif font in the center of the image.

Theory

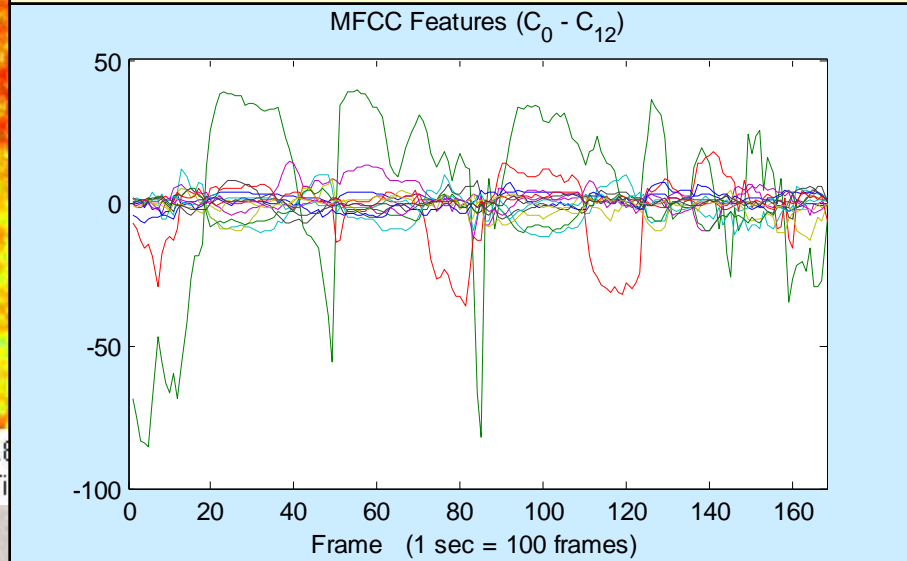
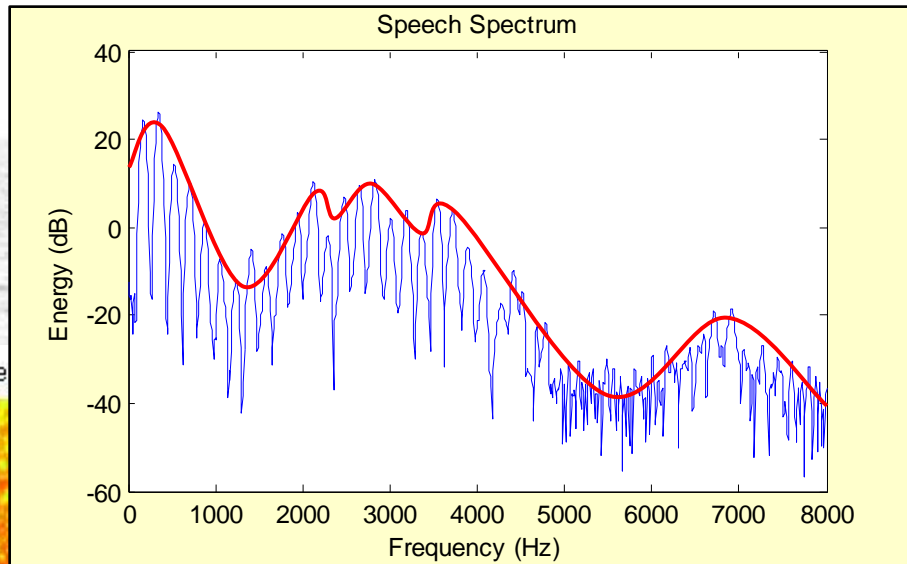
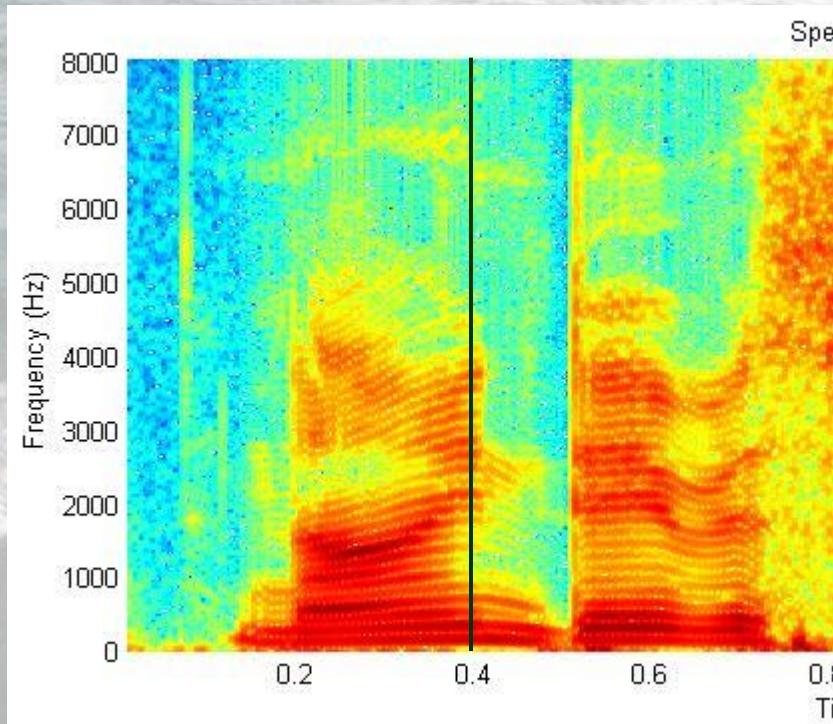
Speech Recognition Overview



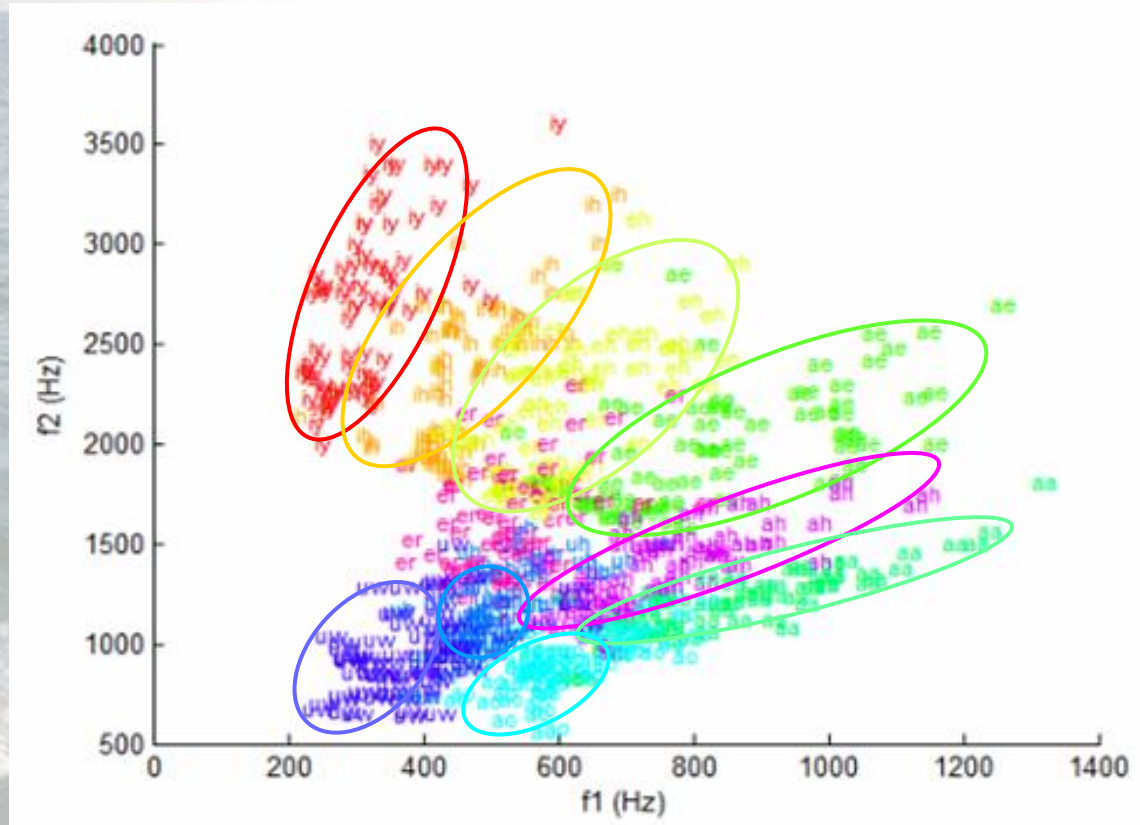
Speech Signal Processing



9 Davis Square, Somerville



Acoustic Modeling



Techniques

- Pattern match
- Dim reduction

Challenges

- Lots of overlap
- Data annotation
- Speaker / Accent
- Noise



Lexical Modeling

a	(ax ey)
...	
beach	b iy ch
...	
nice	n (iy ay) s
...	
recognize	r eh k ax gd n ay z
...	
speech	s p- iy ch
...	
stata	s t- (ey aa) tf ax
...	
tomato	t ax m (ey aa) tf ow
...	
wreck	r eh kd
...	

Techniques

- Dictionary
- Pron generation

Challenges

- Missing words
 - 6.Insight
- Pron variation
 - Nice
 - Stata

Language Modeling

Purpose

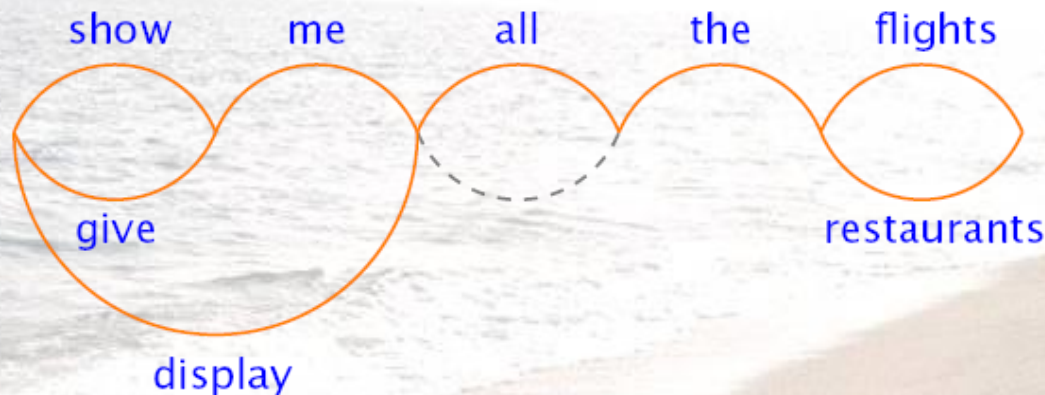
- Constrain word order
- Assign probability
 1. recognize speech
 2. wreck a nice beach

Techniques

- Context-free grammar
- N-gram

Challenges

- Data sparsity
- Domain adaptation



0.036 the

0.026 a

0.018 of

...

0.007 good

...

0.005 day

...

0.003 morning

...

...

0.011 a good

0.003 a morning

...

0.086 good morning

0.026 good day

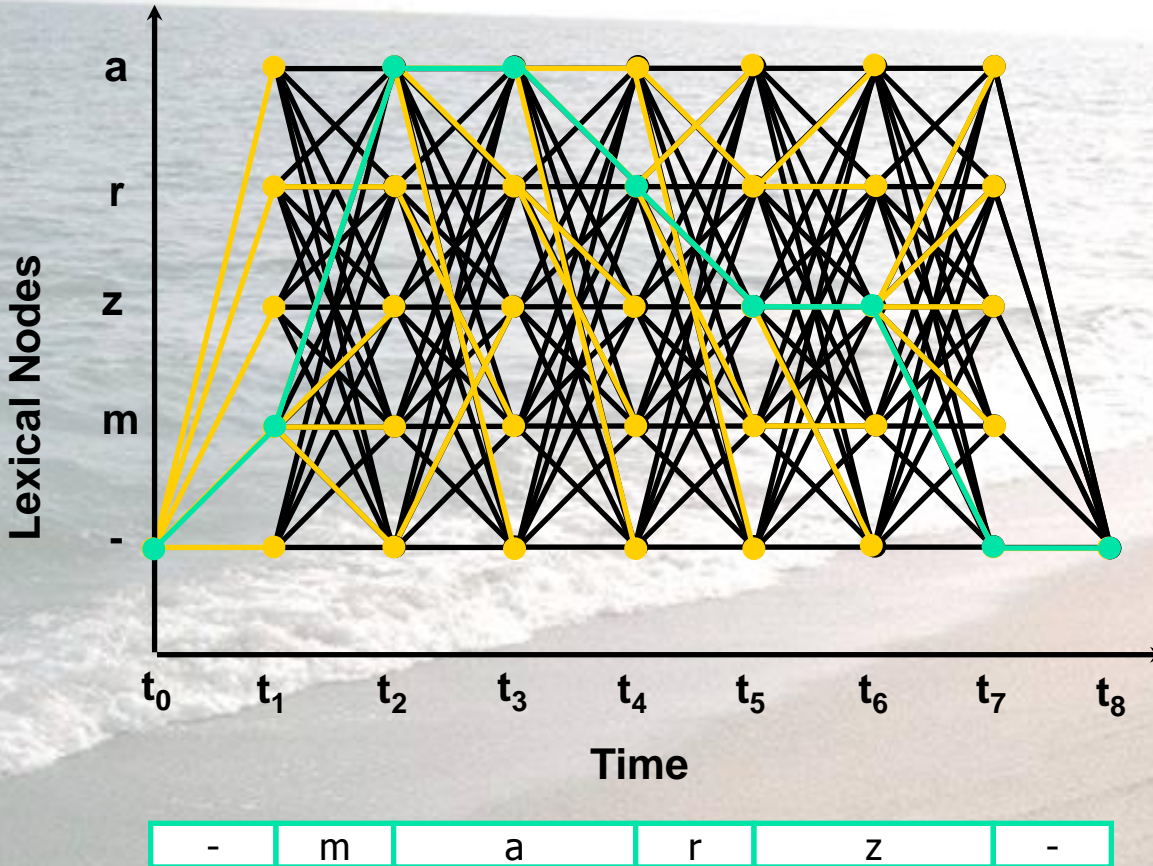
...

0.149 of a

0.057 of day

...

Search



Techniques

- Dyn programming
- A* search backtrace
- Pruning

Challenges

- Huge search space

A photograph of a beach scene. The ocean is on the left, with waves crashing onto the shore. The sand is on the right, with several footprints visible. The sky is bright and overcast. The word "Practice" is written in a large, blue, sans-serif font across the center of the image.

Practice

Command & Control

Microsoft Windows Vista Speech Recognition



Features

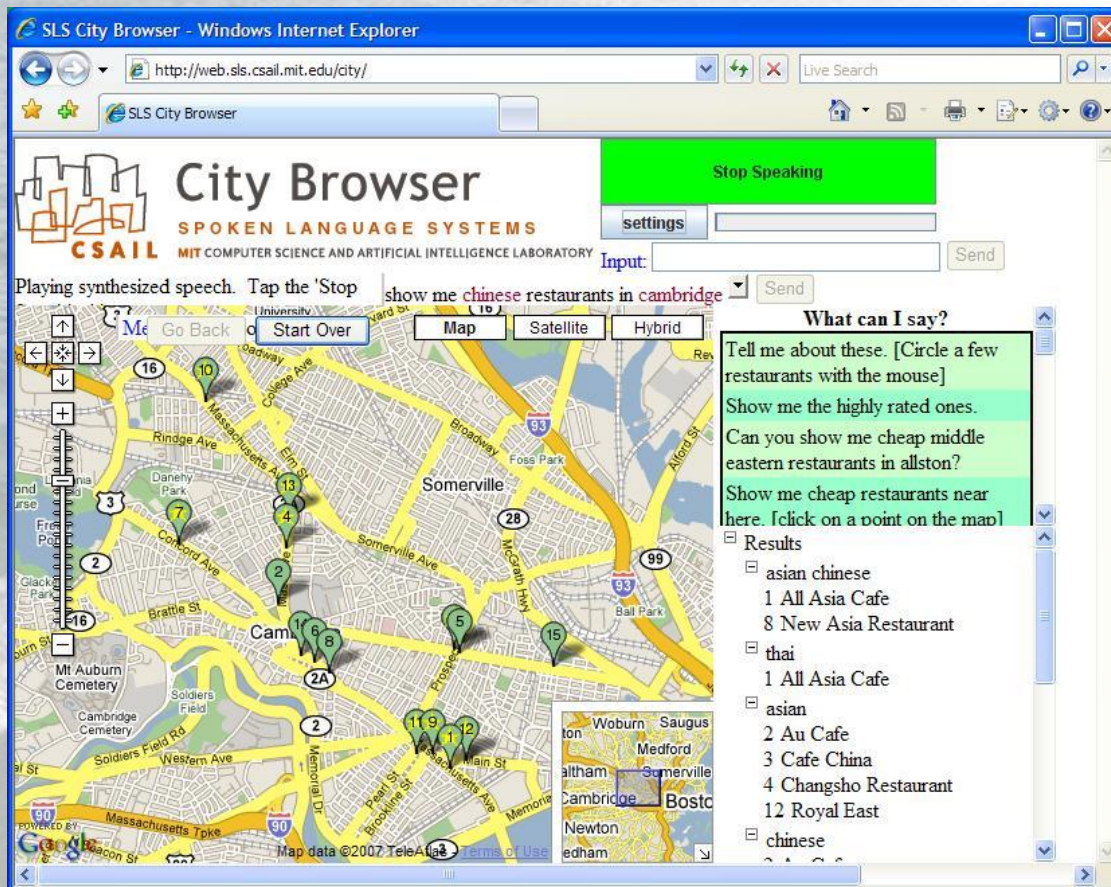
- Control PC apps
- Dictate documents
- Accessibility

Challenges

- Constrained cmds
- User training

Interactive Dialog Systems

SLS City Browser <http://web.sls.csail.mit.edu/city/>



Features

- Restaurants, POI
- Free-form dialog
- Query refinement
- Multimodal control

Challenges

- Labor intensive
- Data collection

Audio Indexing & Search

SLS Lecture Browser <http://web.sls.csail.mit.edu/lectures/>

MIT Lecture Browser - Windows Internet Explorer

http://web.sls.csail.mit.edu/lectures/

MIT Lecture Browser

Lecture Browser
SPOKEN LANGUAGE SYSTEMS

CSAIL MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

Search for words: and/or pick a category:

speech Any category Search

Examples: violin, "solar system", wine AND glass

7. The Sounds of Music
Family Adventures in Science and Technology 1:43:06
Program, September 28, 2003 (Walter Lewin)

delegation the citizens met in washington state ... was there was beautiful the occasion start and speech you can enter altogether appropriate ... there's was the grand opening of the tacoma narrows

8. The Birth and Death of Stars
May 7, 2003 (Walter Lewin) 1:13:19

alright ... to the beach the right ... attitude toward ... so string and the wind instruments they resonate ... a very unique frequency is you all extras now we call the harmonics natural frequencies ... but all objects have resonance frequencies and not just one but a whole zoo um-hum off resonance frequencies which sometimes ... are very mysterious and not so easy to predict ... as an example i have here a minus ... and that is a very complicated piece of glass and i can make this glass itself talking about p a and now but about the glass itself

Features

- Keyword search
- Topic segmentation
- Lecture transcript
- A/V navigation

Challenges

- Disfluencies
- Jargons

Mobile Speech Recognition

SLS Pocket SUMMIT Speech Recognizer



Features

- Small-footprint
- Low CPU/memory

Challenges

- Noise robustness
- Limited grammar

A photograph of a beach at sunset or sunrise. The sky is a pale, hazy yellow. The ocean is a deep blue-grey, with white foam from waves crashing onto the shore. The sand is a warm, golden-brown color, and there are several footprints visible in the foreground. The word "Challenges" is written in a large, bold, blue font across the center of the image.

Challenges

Noise Robustness

Microphone quality

- Close-Talking Headset
- Bluetooth Headset
- Mounted GPS



Environmental/Background noise

- Music
- Babble
- Heating Vent





Adaptation

Speaker Adaptation

- Gender 📣 📣
- Accent 📣

Domain Adaptation

- GPS navigation 📣 📣 📣
- Lecture transcription 📣



Application Diversity

Labor Intensive → Few Applications

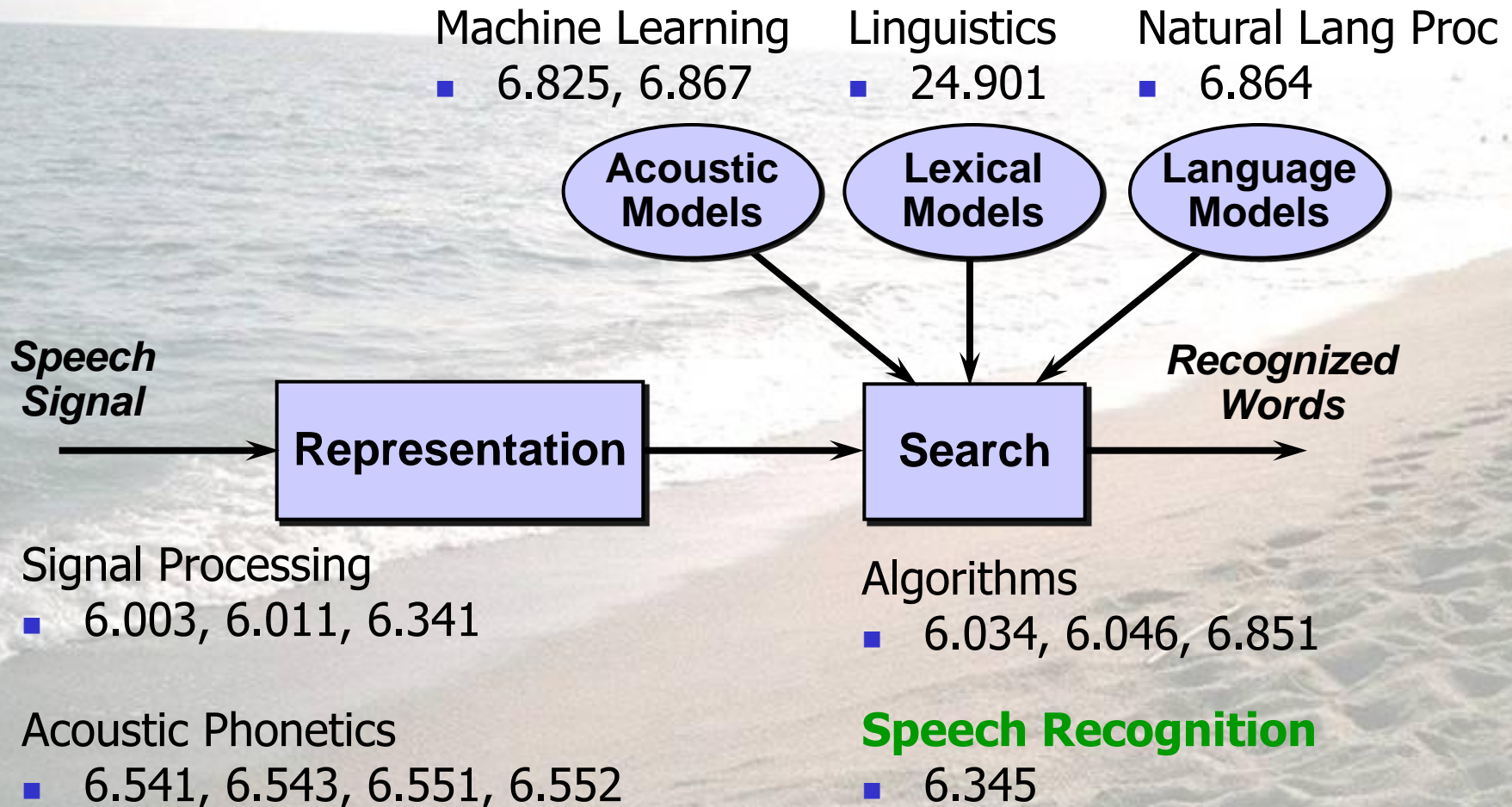
- Weather
- Flight Reservation
- Restaurants

Can the system *automatically* generate spoken dialogue systems via user feedback?

A photograph of a beach with waves crashing onto the shore. The word "Resources" is overlaid in a large, blue, sans-serif font. The background shows the ocean, a sandy beach, and a bright sky.

Resources

Related Courses





Research Groups @ MIT

CSAIL Spoken Language Systems Group

- PIs – James Glass, Stephanie Seneff, Victor Zue
- Research – Speech recognition and dialog systems
- <http://groups.csail.mit.edu/sls/>

RLE Speech Communications Group

- PIs – Kenneth Stevens, Stefanie Shattuck-Hufnagel
- Research – Speech production and perception
- <http://www.rle.mit.edu/speech/>



External Opportunities

Companies & Research Labs (alphabetical order)

- AT&T
- BBN
- Google
- IBM
- Microsoft
- Nuance
- SRI
- VoiceSignal Technology
- Yahoo
- ...



Conclusion

To wreck a nice beach, you need:

- Shovel
- Bulldozer
- ...

Questions?

Paul Hsu bohsu@mit.edu 32-G442