VISIONS, TECHNOLOGY, AND BUSINESS OF CONVERSATIONAL MACHINES

Mazin Gilbert, Director, AT&T
August 7-10, 2006

Joint work with Roberto Pieraccini, Tell-Eureka
A brief history of spoken language technology
Talking Machines: First Steps into Spoken Language Technology

Von Kempelen (1791)

Joseph Faber (1835)

Homer Dudley
Bell Labs (1939)
The 60's to 90's: Technology Evolution

Template Matching

Isolated Words
Speaker Dependent

Connected Words
Speaker Independent

Context Dependent
Sub-Word Units

Stochastic Language
Models

Acoustic/Phonetic

Hidden Markov Models

The statistical approach becomes ubiquitous
The 90’s: the Birth of the Spoken Dialog Industry

[Diagram showing the development of the spoken dialog industry with companies like AT&T, IBM, Microsoft, Nuance, Tellme, Avaya, VoiceGenie, Genesys, and Convergys.]
Modern speech technology
The Speech Technology Chain

- Text-to-Speech Synthesis (TTS)
- Automatic Speech Recognition (ASR)
- Spoken Language Generation (SLG)
- Spoken Language Understanding (SLU)
- Dialog Management (DM)

Words → Meaning → Action → Words

Speech: Data, Rules

Data, Rules

Speech Recognition

Meaning

Dialog Management

Action

Words

Speech
Accurately and efficiently convert a speech signal into a text message independent of the device, speaker or the environment.
ASR - The Big Picture!

Input Speech

Feature Extraction → Acoustic Model $P(X|W)$
Decoding/PATTERN Classification

Language Model $P(W)$
Word Lexicon

Confidence Scoring

"Hello World" (0.9) (0.8)

Change AM for each new language

Change LM and Lexicon for each new language & app.
Machines are 5-50 times worse than humans on virtually any recognition task.
The Speech Technology Chain

- **TTS**: Text-to-Speech Synthesis
- **ASR**: Automatic Speech Recognition
- **SLG**: Spoken Language Generation
- **SLU**: Spoken Language Understanding
- **DM**: Dialog Management

Extract the meaning from recognized speech and interpret a user’s request.
Why is SLU a Difficult Problem?

Ways to say “question about my bill”

<table>
<thead>
<tr>
<th>Variations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>question about my bill</td>
</tr>
<tr>
<td>63</td>
<td>question on my bill</td>
</tr>
<tr>
<td>57</td>
<td>calling about my bill</td>
</tr>
<tr>
<td>43</td>
<td>talk to somebody about my bill</td>
</tr>
<tr>
<td>41</td>
<td>talk to someone about my bill</td>
</tr>
<tr>
<td>32</td>
<td>questions about my bill</td>
</tr>
<tr>
<td>30</td>
<td>problem with my bill</td>
</tr>
<tr>
<td>23</td>
<td>speak to someone about my bill</td>
</tr>
<tr>
<td>22</td>
<td>calling about a bill</td>
</tr>
<tr>
<td>20</td>
<td>calling about my phone bill</td>
</tr>
<tr>
<td>16</td>
<td>questions on my bill</td>
</tr>
<tr>
<td>16</td>
<td>question about a bill</td>
</tr>
<tr>
<td>15</td>
<td>talk about my bill</td>
</tr>
<tr>
<td>11</td>
<td>question about my phone bill</td>
</tr>
<tr>
<td>11</td>
<td>question about my billing</td>
</tr>
<tr>
<td>11</td>
<td>discuss my bill</td>
</tr>
<tr>
<td>10</td>
<td>speak with someone about my bill</td>
</tr>
<tr>
<td>10</td>
<td>calling about my billing</td>
</tr>
<tr>
<td>9</td>
<td>problem with my phone bill</td>
</tr>
<tr>
<td>9</td>
<td>calling about my telephone bill</td>
</tr>
<tr>
<td>8</td>
<td>speak to someone in billing</td>
</tr>
<tr>
<td>8</td>
<td>question about the bill</td>
</tr>
<tr>
<td>7</td>
<td>speak to somebody about my bill</td>
</tr>
<tr>
<td>7</td>
<td>speak to a billing</td>
</tr>
<tr>
<td>7</td>
<td>question on my phone bill</td>
</tr>
<tr>
<td>7</td>
<td>calling regarding my bill</td>
</tr>
<tr>
<td>7</td>
<td>calling concerning my bill</td>
</tr>
<tr>
<td>6</td>
<td>talk to somebody in billing</td>
</tr>
<tr>
<td>6</td>
<td>questions about my billing</td>
</tr>
<tr>
<td>6</td>
<td>question on my billing</td>
</tr>
<tr>
<td>6</td>
<td>problem with my billing</td>
</tr>
<tr>
<td>6</td>
<td>information about my bill</td>
</tr>
<tr>
<td>6</td>
<td>calling about my A T and T bill</td>
</tr>
<tr>
<td>5</td>
<td>talk to someone about my phone bill</td>
</tr>
<tr>
<td>5</td>
<td>talk to someone about a bill</td>
</tr>
<tr>
<td>5</td>
<td>talk to somebody about my billing</td>
</tr>
<tr>
<td>5</td>
<td>talk to somebody about a bill</td>
</tr>
<tr>
<td>5</td>
<td>speak to someone in the billing</td>
</tr>
<tr>
<td>5</td>
<td>speak to someone about a bill</td>
</tr>
<tr>
<td>5</td>
<td>questions on my billing</td>
</tr>
<tr>
<td>5</td>
<td>question on the bill</td>
</tr>
<tr>
<td>5</td>
<td>question on a bill</td>
</tr>
<tr>
<td>5</td>
<td>question my bill</td>
</tr>
<tr>
<td>5</td>
<td>calling in regards to my bill</td>
</tr>
<tr>
<td>5</td>
<td>calling about the bill</td>
</tr>
<tr>
<td>4</td>
<td>talk to someone about my telephone bill</td>
</tr>
<tr>
<td>4</td>
<td>talk to somebody about my account</td>
</tr>
<tr>
<td>4</td>
<td>talk to billing</td>
</tr>
<tr>
<td>4</td>
<td>speak with someone in billing</td>
</tr>
<tr>
<td>4</td>
<td>question about my telephone bill</td>
</tr>
<tr>
<td>4</td>
<td>information on my bill</td>
</tr>
<tr>
<td>4</td>
<td>calling regarding my statement</td>
</tr>
<tr>
<td></td>
<td>1 talk to someo- to someone about my moms telephone bill</td>
</tr>
<tr>
<td></td>
<td>1 question about the new A T and T billing</td>
</tr>
</tbody>
</table>

Total 1083 variations in 1912 matches
Knowledge Sources for SLU

- Lexical
- Syntactic
- Semantic
- Pragmatic
- Acoustic/Phonetic

Enabling Applications:
- Call routing
- Problem solving
- Customer care
- Speech Translation
- Speech Data Mining
SLU - The Big Picture!

From ASR/DM (text, lattices, n-best, history)

- Text Normalization
  - Morphology, Synonyms

- Parsing/Decoding
  - Extracting named entities, semantic concepts, syntactic tree

- Interpretation
  - Slot filling, reasoning, task knowledge representation

To DM (concepts, entities, parse tree)
The Speech Technology Chain

Manage elaborate exchanges with the user, providing access to information
The Dialog Flow

- Context Interpretation
- Dialog Strategies
- Backend

Observation → Dialog State → Action → Transition
Mixed-Initiative Dialog

Who manages the dialog?

User

System

How may I help you?
I need to travel from Chicago to Newark tomorrow night

Please say just your departure city.
Chicago
The Speech Technology Chain

DM

SLU

TTS

ASR

SLG

Data, Rules

Speech

Speech

Words

Words

Action

Meaning

Text-to-Speech Synthesis

Automatic Speech Recognition

Spoken Language Generation

Spoken Language Understanding

Dialog Management

Translate the action of the DM into a textual representation
The Speech Technology Chain

Text-to-Speech Synthesis

Spoken Language Generation

Data, Rules

Speech

Words

Action

Meaning

Dialog Management

TTS

SLG

DM

SLU

ASR

Automatic Speech Recognition

Spoken Language Understanding

Speech

Provide completely natural, high intelligibility speech from text for any talker, language or accent
Concatenative Synthesis

Text Analysis, Letter-to-Sound, Prosody

Alphabetic Characters

Assemble Units that Match Input Targets

Phonetic Symbols, Prosody Targets

Speech Waveform Modification and Synthesis

Change Sound Store for each new voice and/or language

Dictionary and Rules

Store of Sound Units

Change Front-End for each new language

Text

Speech
Conversational Machines
Lessons Learned

• "There is no data like more data", but data is expensive to collect and label, and typically unavailable in large quantities for every speaker, language and environment.

• Significant resources and expertise are necessary for creating, maintaining and customizing conversational machines.

• Speech input/output is insufficient for accommodating for system failures and for creating complex automated applications for anyone and anywhere.
Multimodal Technology Components

- Visual
- TTS (Text-to-Speech Synthesis)
- ASR (Automatic Speech Recognition)
- SLU (Spoken Language Understanding)
- SLG (Spoken Language Generation)
- DM (Dialog Management)

Speech, Words, Action, Meaning, Data, Rules
Commercial Spoken Dialog Systems
The Speech Application Lifecycle

1. Requirements
2. VUI design
3. Usability testing
4. VUI development
5. Speech science
6. High level system design
7. System engineering
8. Integration
9. Partial deployment
10. Full deployment

- Analyst
- VUI Designer
- Speech Scientist
- VUI Designer
- Architect, App Developer Engineer
- Architect, App Developer Engineer
The Voice Web

Voice Browser

Telephony Platform

ASR

TTS

MRCP

Telephone

Web Server

Internet

VoiceXML

/SALT

SSML, SRGF, EMMA
The Speech Technology Market

- Server-based Telephony
- Conversational
- Desktop Dictation
- Embedded Car Cell
- Multimodal/Multimedia
- Security
- Call Center Automation
- Speech to Speech Translation
- Entertainment
Business in Conversational Technology

- Return on Investment (ROI)
  - Reduce cost
  - Enable self service options
  - New revenue opportunities

- Customer Retention
  - Better user interface
  - Reduce waiting time for callers
  - Reduce misrouting

- Branding
  - Project a new image and brand awareness
  - Use of persona
VISIONS, TECHNOLOGY, AND BUSINESS OF CONVERSATIONAL MACHINES