Speech enabled Self Help System for Insurance Agents

Workshop on Multimedia Workshop on Multimedia Applications for Enterprises

Sunil Kopparapu, Ambikesh Shukla, Sumitra Das, Akhilesh Srivastava

SunilKumar.Kopparapu@TCS.COM

TCS Innovation Lab - Mumbai

Advanced Technology Applications Group,

Yantra Park, Thane (West), Maharastra

May 2007

Overview

- Background
- Self Help System (Block Diagram)
- Essential Components
- System Functional View - *Need for integrating speech and NL*
- Ideal System
- Requirements of a self help system
- Advantages
- Demonstration

Background

- Insurance agents act as intermediaries between insurance company and their clients.

Background

- Insurance agents act as intermediaries between insurance company and their clients.
- Insurance agents keep track of information of their clients (policy status, maturity status, change of address request) by being in touch with the insurance companies.

Background

- Insurance agents act as intermediaries between insurance company and their clients.
- Insurance agents keep track of information of their clients (policy status, maturity status, change of address request) by being in touch with the insurance companies.
- Call center provided by insurance companies cater to (dynamic) information needed by agents

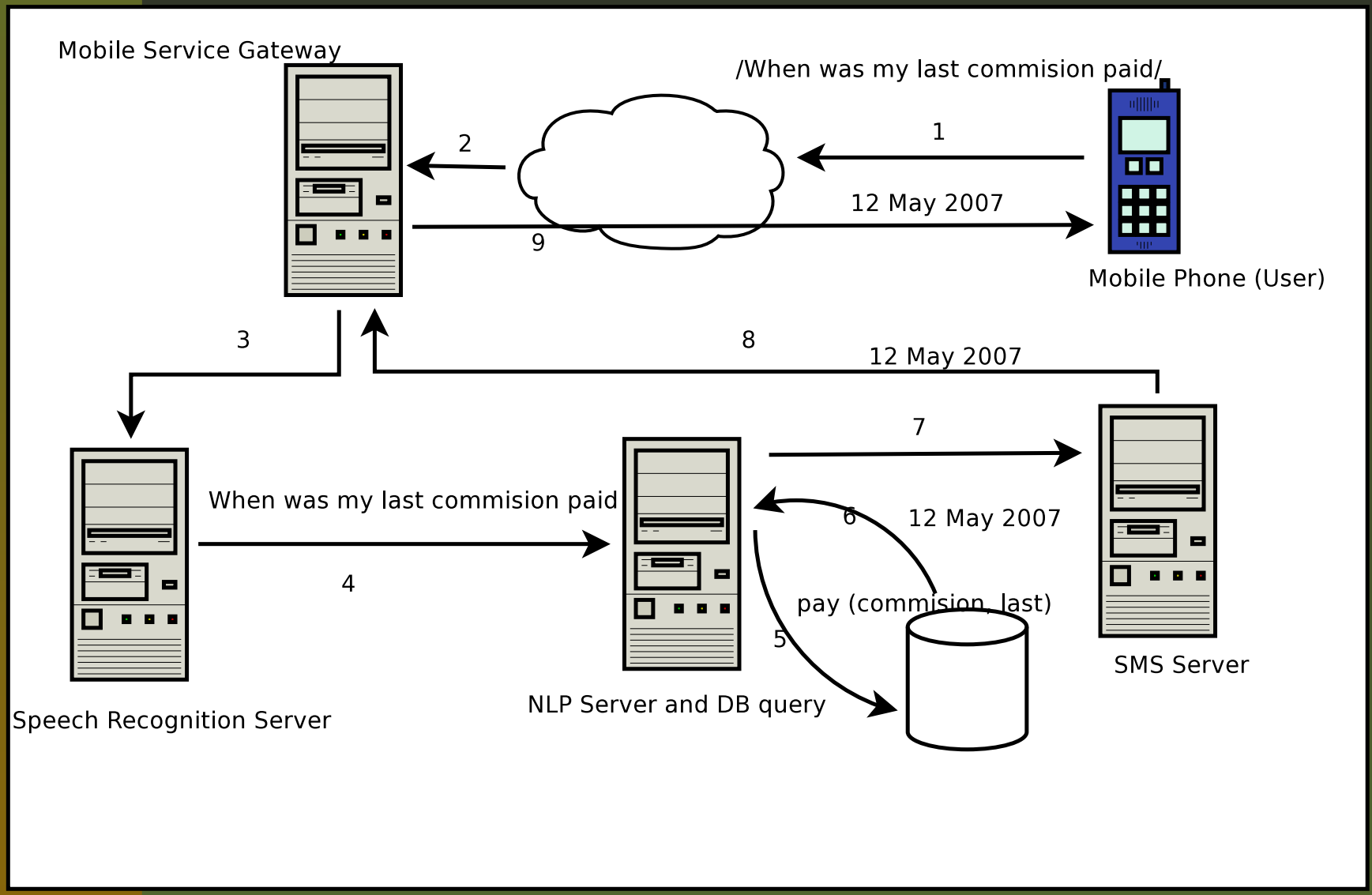
Background

- Insurance agents act as intermediaries between insurance company and their clients.
- Insurance agents keep track of information of their clients (policy status, maturity status, change of address request) by being in touch with the insurance companies.
- Call center provided by insurance companies cater to (dynamic) information needed by agents
- Insurance companies spend **time** and **money** to maintain a people driven call center.

Background

- Insurance agents act as intermediaries between insurance company and their clients.
- Insurance agents keep track of information of their clients (policy status, maturity status, change of address request) by being in touch with the insurance companies.
- Call center provided by insurance companies cater to (dynamic) information needed by agents
- Insurance companies spend **time** and **money** to maintain a people driven call center.
- An **automated self help system** to cater to the insurance agents makes economic sense.

Block Diagram: Self help system



Essential Components

- **Speech Recognition Engine**
  converts speech query into text query

Essential Components

- **Speech Recognition Engine**
  converts speech query into text query
- **Natural Language Processing**
  any errors due to speech recognition process handled using key concept key word based processing plus Error Handling - what to do when *exact* response is not found in the database?

Essential Components

- **Speech Recognition Engine**
  converts speech query into text query
- **Natural Language Processing**
  any errors due to speech recognition process handled using key concept key word based processing plus Error Handling - what to do when *exact* response is not found in the database?
- **Database search**
  text query is converted into SQL query to fetch a text response

Essential Components

- **Speech Recognition Engine**
  converts speech query into text query
- **Natural Language Processing**
  any errors due to speech recognition process handled using key concept key word based processing plus Error Handling - what to do when *exact* response is not found in the database?
- **Database search**
  text query is converted into SQL query to fetch a text response
- **SMS server**
  text response sent to a mobile number

Functional View (1)

Let $\mathcal{S}$ represent speech recognition and $\mathcal{N}$ be the natural language processing engine

- $\mathcal{S}$: time sequence $\rightarrow$ string sequence
- $\mathcal{N}$: string sequence $\rightarrow$ string sequence

Let x_t represents the spoken query (corresponding to say the string x_s). Then

- $\mathcal{S}(x_t) = x_{s'}$ (speech engine)
- $\mathcal{N}(x_{s'}) = x_{s''}$ (NL processing)

The idea is to build $\mathcal{S}$ and $\mathcal{N}$ such that their combined (sequential?) effort, namely, $\mathcal{N}(\mathcal{S}(x_t)) = x_{s''}$ is such that $x_{s''} \approx x_s$.

Functional View (2)

Observe,

- $\mathcal{S}$ uses acoustic models and language grammar (tightly coupled) to convert x_t to $x_{s'}$
- $\mathcal{N}$ operates on $x_{s'}$ and uses only language grammar to convert it into $x_{s''}$.

Do we we really need both $\mathcal{S}$ and $\mathcal{N}$? Why not .....

1. isolate $\mathcal{S}$ and $\mathcal{N}$ - language processing only in $\mathcal{N}$ (or)
2. combine everything into $\mathcal{S}$ and do away with $\mathcal{N}$

Functional View (3)

- Grammar used in $\mathcal{S}$ is
  - coupled with the acoustic models,
  - degree of configurability is limited (speech to text)
  - (but) necessary to perform *reasonable* recognition
- Relatively high degree of configurability possible in $\mathcal{N} : x_s \longrightarrow x_s$ (text to text)

$\Rightarrow \mathcal{N}$ is necessary; Need both $\mathcal{S}$ and $\mathcal{N}$ to coexist with (minor?) overlap of functionality to produce better (user) interfaces.

Combining $\mathcal{S}$ and $\mathcal{N}$

Ideal system $x_t \rightarrow x_s$.

- $\mathcal{S}_1 \rightarrow$ allows you to speak anything (dictation system); $\mathcal{S}_1(x_t) = x^1_{s'}$ $(\uparrow)$
- $\mathcal{S}_2 \rightarrow \mathcal{S}_1$ + configured to a particular person (person dependent); $\mathcal{S}_1(x_t) = x^2_{s'}$ $(\leftrightarrow)$
- $\mathcal{S}_3 \rightarrow \mathcal{S}_2$ + allows you to speak within a restricted grammar; $\mathcal{S}_3(x_t) = x^3_{s'}$ $(\downarrow)$

Clearly, $d(x^1_{s'}, x_s) > d(x^2_{s'}, x_s) > d(x^3_{s'}, x_s)$. For the system to perform *well* the contribution of $\mathcal{N}$ would vary (generate $x^1_{s''}$, $x^2_{s''}$, $x^3_{s''}$, such that $d(x^1_{s''}, x_s) \approx d(x^2_{s''}, x_s) \approx d(x^3_{s''}, x_s) \approx 0$).

Any Combinination of $\mathcal{S}$ and $\mathcal{N}$?

Will any $\mathcal{S}$ work? What if the performance of $\mathcal{S}$ is *poor*? What is the requirement?
Only if $d(x^1_{s'}, x_s) < \epsilon$ or $\mathcal{S}(x_t) - x_s < \epsilon$ will $\mathcal{N}$ have a role to perform. (not hard to guess)

- What is this minimum ϵ? Minimum accuracy of $\mathcal{S}$?
- Can we determine it? Quantify it?
- Is it dependent on the domain? problem?

More Questions

- Should $\mathcal{S}$ and $\mathcal{N}$ operate in sequence?
- Are there any other ways of combining $\mathcal{S}$ and $\mathcal{N}$?

Ideal System

- Open Speech - Dictation like (free speech - speak without constraints)
- Speaker independent (different accents, dialects, age, gender)
- Environment independent (office, public telephone)

For ideal system

- User experience is good but
- Speech Recognition accuracies are poor

So ....

Non-ideal but ...

- limit domain (queries related to specific domain insurance agent self help)
- tune (constrain?) the system (make use of a priori information on expected queries)

Appreciate

- Speech recognition not always accurate
- no need for exact recognition of what is spoken; may be (just) the key-concepts and keywords are sufficient

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent
  - on behalf of their clients (example, *What is the maturity value of the policy xyz*) and

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent
  - on behalf of their clients (example, *What is the maturity value of the policy xyz*) and
  - themselves (example, *When was my last commission paid?*).

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent
  - on behalf of their clients (example, *What is the maturity value of the policy xyz*) and
  - themselves (example, *When was my last commission paid?*).
- System to cater to

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent
  - on behalf of their clients (example, *What is the maturity value of the policy xyz*) and
  - themselves (example, *When was my last commission paid?*).
- System to cater to
  - different accents

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent
  - on behalf of their clients (example, *What is the maturity value of the policy xyz*) and
  - themselves (example, *When was my last commission paid?*).
- System to cater to
  - different accents
  - different ways in which same queries can be asked

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent
  - on behalf of their clients (example, *What is the maturity value of the policy xyz*) and
  - themselves (example, *When was my last commission paid?*).
- System to cater to
  - different accents
  - different ways in which same queries can be asked
  - different complexity of queries

Insurance Agent Self Help System

- System should cater to the different kinds of information sought by insurance agent
  - on behalf of their clients (example, *What is the maturity value of the policy xyz*) and
  - themselves (example, *When was my last commission paid?*).
- System to cater to
  - different accents
  - different ways in which same queries can be asked
  - different complexity of queries
  - Response as a SMS (information record)

Need: Enable Natural English Query

System allows querying in several different ways (natural language).

- Surrender value of policy xyz?
- What is the surrender value of policy xyz?
- Can you tell me surrender value of policy xyz?
- Please let me know the surrender value of policy xyz?
- Please tell me surrender value of policy xyz?
- Tell me surrender value of policy xyz?
- My policy is xyz. What is its surrender value?

... surrender_value .... policy ... xyz....

Advantages of speech self help system

- 24×7 (support for frequent asked queries)

Advantages of speech self help system

- 24×7 (support for frequent asked queries)
- Automated self service (reducing call volume)

Advantages of speech self help system

- 24×7 (support for frequent asked queries)
- Automated self service (reducing call volume)
- Accuracy and consistency in call experience

Advantages of speech self help system

- 24×7 (support for frequent asked queries)
- Automated self service (reducing call volume)
- Accuracy and consistency in call experience
- Automated speech and text (SMS) response (provides a choice)

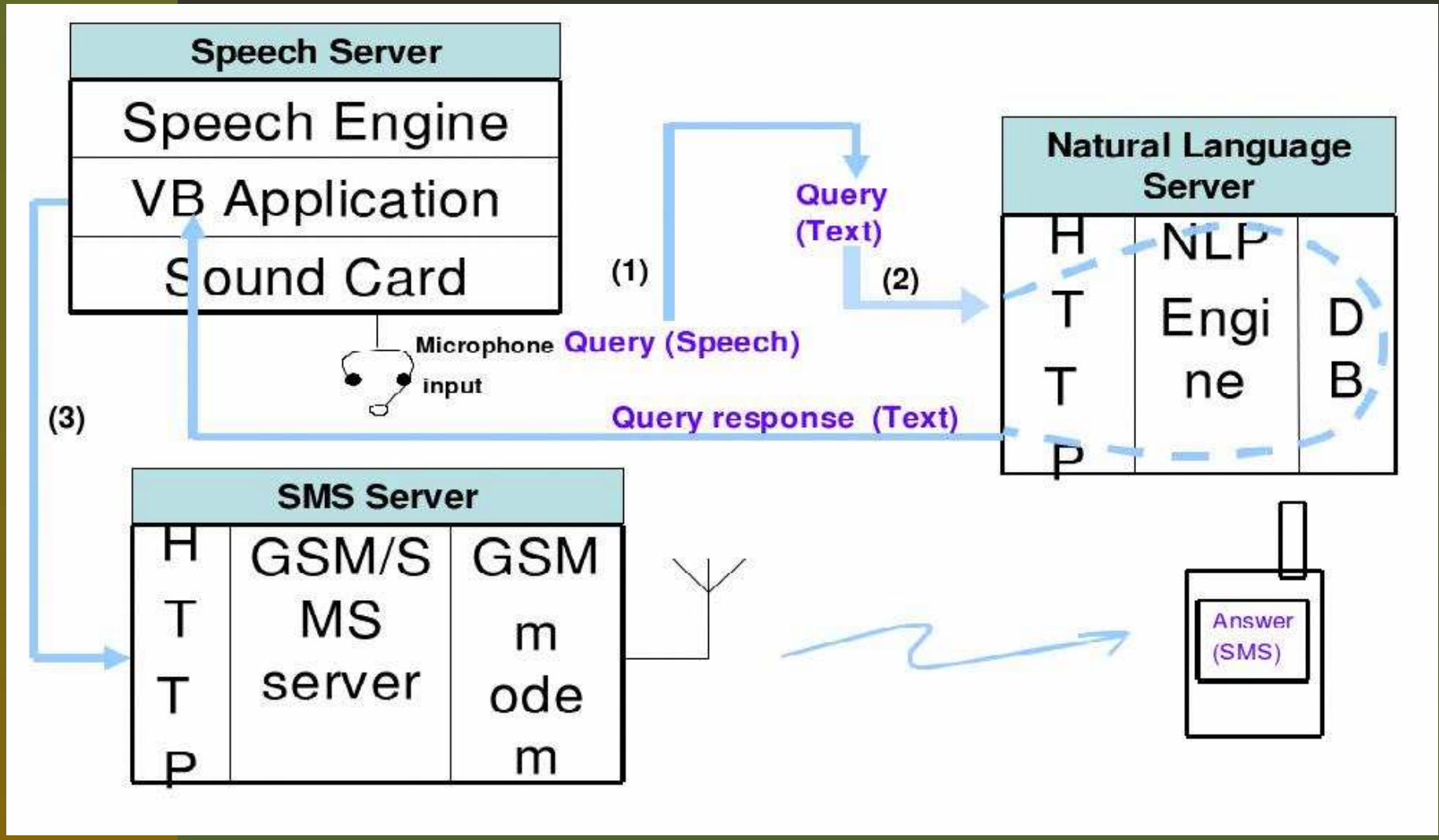
Advantages of speech self help system

- 24×7 (support for frequent asked queries)
- Automated self service (reducing call volume)
- Accuracy and consistency in call experience
- Automated speech and text (SMS) response (provides a choice)
- Ability to handle multiple levels of complexity

Advantages of speech self help system

- 24×7 (support for frequent asked queries)
- Automated self service (reducing call volume)
- Accuracy and consistency in call experience
- Automated speech and text (SMS) response (provides a choice)
- Ability to handle multiple levels of complexity
- Wider opportunities in Tele-servicing (New Business?)

Demonstration: Block Diagram



Demonstration

- System uses

Demonstration

- System uses
  - Speech Recognition ($\mathcal{S}$): Microsoft SAPI speech recognition engine SDK

Demonstration

- System uses
  - Speech Recognition ($\mathcal{S}$): Microsoft SAPI speech recognition engine SDK
  - Natural Language Processing ($\mathcal{N}$): Built in-house

Demonstration

- System uses
  - Speech Recognition ($\mathcal{S}$): Microsoft SAPI speech recognition engine SDK
  - Natural Language Processing ($\mathcal{N}$): Built in-house
  - SMS Server: Kannel (free ware)

Demonstration

- System uses
  - Speech Recognition ($\mathcal{S}$): Microsoft SAPI speech recognition engine SDK
  - Natural Language Processing ($\mathcal{N}$): Built in-house
  - SMS Server: Kannel (free ware)
- Three versions (processing distributed in $\mathcal{S}$ and $\mathcal{N}$)

Demonstration

- System uses
  - Speech Recognition ($\mathcal{S}$): Microsoft SAPI speech recognition engine SDK
  - Natural Language Processing ($\mathcal{N}$): Built in-house
  - SMS Server: Kannel (free ware)
- Three versions (processing distributed in $\mathcal{S}$ and $\mathcal{N}$)
  - $\mathcal{S}_1$ has very constrained grammar; very less processing in $\mathcal{N}$

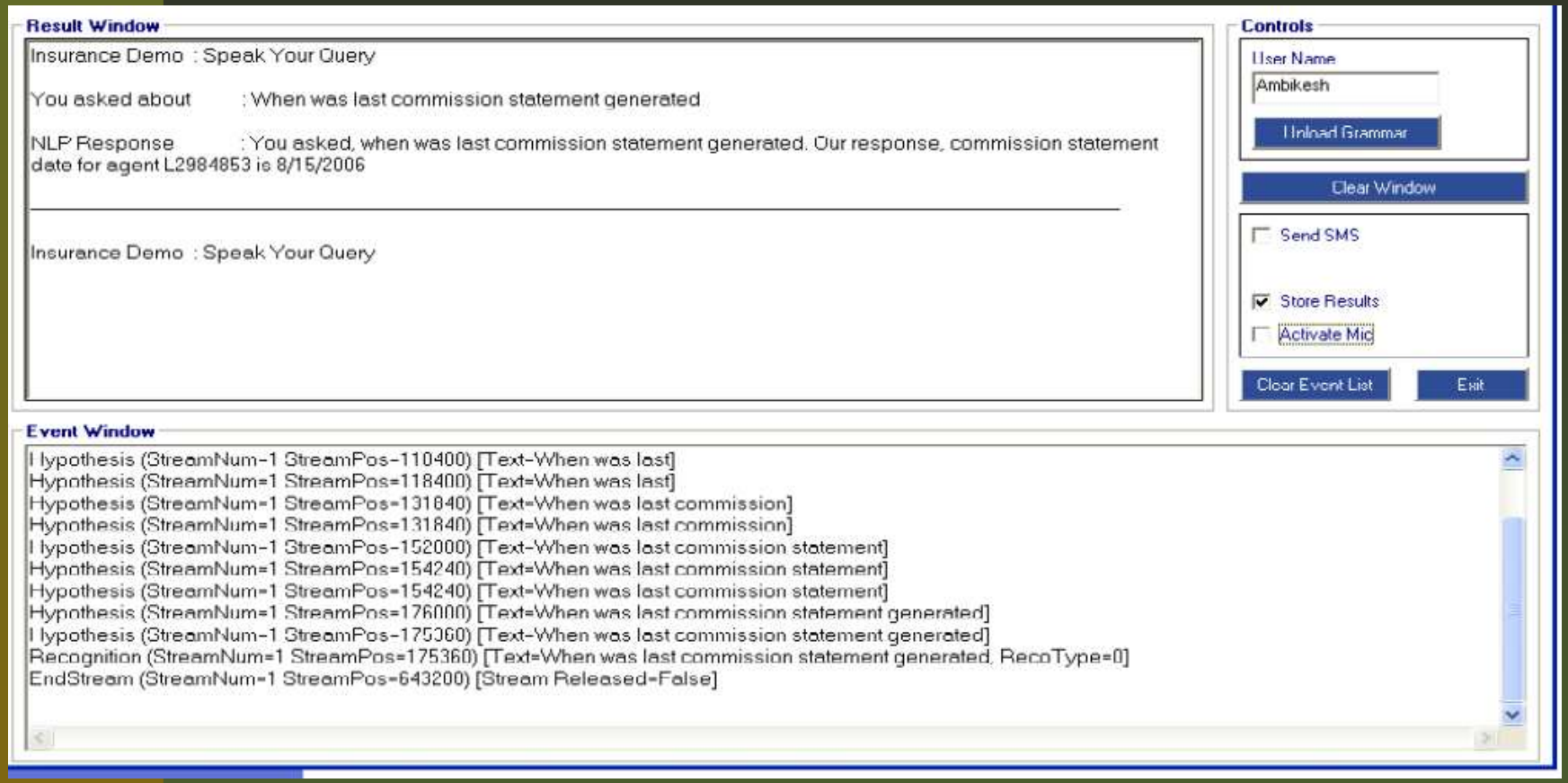
Demonstration

- System uses
  - Speech Recognition ($\mathcal{S}$): Microsoft SAPI speech recognition engine SDK
  - Natural Language Processing ($\mathcal{N}$): Built in-house
  - SMS Server: Kannel (free ware)
- Three versions (processing distributed in $\mathcal{S}$ and $\mathcal{N}$)
  - $\mathcal{S}_1$ has very constrained grammar; very less processing in $\mathcal{N}$
  - $\mathcal{S}_2$ has liberal grammar; more processing in $\mathcal{N}$

Demonstration

- System uses
  - Speech Recognition ($\mathcal{S}$): Microsoft SAPI speech recognition engine SDK
  - Natural Language Processing ($\mathcal{N}$): Built in-house
  - SMS Server: Kannel (free ware)
- Three versions (processing distributed in $\mathcal{S}$ and $\mathcal{N}$)
  - $\mathcal{S}_1$ has very constrained grammar; very less processing in $\mathcal{N}$
  - $\mathcal{S}_2$ has liberal grammar; more processing in $\mathcal{N}$
  - $\mathcal{S}_3$ has **no** grammar; most processing in $\mathcal{N}$

Interface


Result Window
Insurance Demo : Speak Your Query
You asked about : When was last commission statement generated
NLP Response : You asked, when was last commission statement generated. Our response, commission statement date for agent L2984853 is 8/15/2006
Insurance Demo : Speak Your Query
Controls
User Name
Ambikesh
Unload Grammar
Clear Window
Send SMS
Store Results
Activate Mic
Clear Event List
Exit
Event Window
Hypothesis (StreamNum=1 StreamPos=110400) [Text=When was last]
Hypothesis (StreamNum=1 StreamPos=118400) [Text=When was last]
Hypothesis (StreamNum=1 StreamPos=131840) [Text=When was last commission]
Hypothesis (StreamNum=1 StreamPos=131840) [Text=When was last commission]
Hypothesis (StreamNum=1 StreamPos=152000) [Text=When was last commission statement]
Hypothesis (StreamNum=1 StreamPos=154240) [Text=When was last commission statement]
Hypothesis (StreamNum=1 StreamPos=154240) [Text=When was last commission statement]
Hypothesis (StreamNum=1 StreamPos=176000) [Text=When was last commission statement generated]
Hypothesis (StreamNum=1 StreamPos=175360) [Text=When was last commission statement generated]
Recognition (StreamNum=1 StreamPos=175360) [Text=When was last commission statement generated, RecoType=0]
EndStream (StreamNum=1 StreamPos=643200) [Stream Released=False]

$\mathcal{S}_1$ - constrained grammar

```
<GRAMMAR>
 <RULE NAME="S_1" TOPLEVEL="ACTIVE">
  <o> <RULEREF NAME="StartTag"/> </o>
  <RULEREF NAME="KeyConcept"/>
  <o> of <o> the </o> </o>
  <o> in <o> the </o> </o>
  <RULEREF NAME="KeyWord"/>
  <o> <RULEREF NAME="EndTag"/> </o>
 </RULE>

 <RULE NAME="StartTag">
  <P> What is the </P>
  <P> Please send me </P>
  <P> Can you please send me</P>
  <P> Can you tell me </P>
 </RULE>
</GRAMMAR>
```

$\mathcal{S}_1$ - constrained grammar - Example

StartTag	Keyword	Keyword	EndTag
What is the	surrender value	policy xyz	as of today
When was	death claim bonus		at the start of the year
Can you tell me		maturity value	during last year
What is my	last commission		at the start of last year

S_2 - liberal grammar

A keyword spotting (KWS) system

```
<GRAMMAR>
 <RULE NAME="S_2" TOPLEVEL="ACTIVE">
  <RULEREF NAME="DonotCare"/>
  <RULEREF NAME="KeyConcept"/>
  <RULEREF NAME="DonotCare"/>
  <RULEREF NAME="KeyWord"/>
  <RULEREF NAME="DonotCare"/>
 </RULE>
</GRAMMAR>
```

S_3 - no grammar

```
<GRAMMAR>
 <RULE NAME="S_3" TOPLEVEL="ACTIVE">
  <RULEREF NAME="DonotCare"/>
 </RULE>
</GRAMMAR>
```

Thank You

- Comments
- Suggestions
- Criticism
- Queries (in NL!)

SunilKumar.Kopparapu@TCS.Com
TCS Innovation Lab - Mumbai
Advanced Technology Application Group
Yantra Park, Thane (West)