Commute UX: Telephone Dialog System for Location-based Services

Ivan Tashev, Michael Seltzer, Yun-Cheng Ju, Dong Yu, Alex Acero
Overview

- Commute UX project
- System architecture
- Understanding locations
- Rendering locations
- Deployment and results
- Demo
- Conclusions and future work
Commute UX project

- Aiming to improve the driver’s experience during commute time
  - 24 minutes average commute time one way
  - 70% drive alone (numbers for USA)

- Why speech
  - Driving is hands-busy and eyes-busy
  - Speech is the safest communication channel

- First system: telephone info line
  - Able to reach large number of users
  - Free phrase queries
  - Traffic, gas prices, and weather

- Understanding and conveying locations – critical for the system
System Architecture (2)

- **Speech Recognizer**
  - Task dependent, slot based LM
  - Filler word N-gram and garbage collection

- **Semantic parser**
  - Task classification and task specific semantic slots

- **Dialog manager**
  - Two level state machine

- **Context manager**
  - Plays key role in the usability of the system
  - Updates the LM and the semantic model
System Architecture (3)

- Information Retriever
  - Connection to the backend database: geographic, personal, real-time info
- Response Manager
- Real-time updaters
- Web site
Understanding locations

- Understanding location is a complex problem
  - Large vocabulary, complex phrases
  - Ways to handle: limit the scope, n-best, TF-IDF
- Locations, locations:

| Type            | Example                                                                 |
|-----------------|--------------------------------------------------------------------------|
| City            | Kirkland                                                                |
| ZIP             | 98034                                                                   |
| Neighborhood    | Juanita                                                                 |
| Intersection    | Corner of 150\textsuperscript{th} Av NE and 36\textsuperscript{th} St.  |
| Zone            | Within 1 mile from <point>                                              |
| Road            | I-405 northbound                                                        |
| Section         | between exits 10 and 22 of I-405                                       |
Understanding locations: context

- Types of locations to understand is application and context dependent
- Location as context
  - Keep it once entered
  - Use it in the next phases of the dialogue
- Example:
  Q: “How is the traffic between Bellevue and Seattle?”
  A: “The traffic between Bellevue and Seattle via I-90 is ...”
  Q: “How about via 520?”
  A: “The traffic between Bellevue and Seattle via SR520 is ...”
Understanding locations: Intersections

- We often use intersections when conveying locations to each other
  - Intersections are readily known
  - People often don’t know exact address

Intersection understanding:

For more details see our Interspeech 2007 paper
Understanding locations: Personalization

- Website integration for **Personal Points of Interest (PPOI)**
- Use like any other location
  - “How’s the traffic from work to Julie’s school”
- Example points of interest

| Delete | Edit | Description | Name                      | Address            | City           | Zip  |
|--------|------|-------------|---------------------------|--------------------|----------------|------|
|        |      | art class   | art by fire               | 5465 Leary Ave NW  | Seattle        | 98107|
|        |      | glass museum| Tacoma Museum of Glass    | 1801 E Dock St    | Tacoma         | 98402|
|        |      | Gym         | Pro Club                  | 4455 148th Ave NE | Bellevue       | 98007|
|        |      | Home        | Home                      | 424 Belmont Ave E | Seattle        | 98102|
|        |      | Julie's school| washington middle school | 2102 S Jackson St | Seattle        | 98144|
|        |      | Red Hook    | Red hook Brewery          | 14300 NE 145th St | Woodinville    | 98072|
|        |      | Work        | Work                      | 4062 148th Ave NE | Redmond        | 98052|

↓ Add new locations
Rendering locations

- Rendering the location is critical for the usability of the information system
  - Difficulties due to TTS quality
  - More difficult in noisy conditions
  - The driver is under cognitive load
  - Short term memory in humans is limited

- **Four ways to render a location**

| Rendering Type       | Description                                                                 |
|----------------------|-----------------------------------------------------------------------------|
| Address only         | 14803 Northeast 51st Street                                                 |
| Address & POI        | 251 Rainier Avenue North, near Renton Chamber of Commerce                   |
| Intersection only    | The corner of East Madison Street and 17th Avenue                           |
| Intersection & POI   | The corner of NE Woodinville Road and 131st Avenue, near City Hall          |

- **Final normalization and conversion**
Rendering locations

- Rendering the location is critical for the usability of the information system.
- Difficulties due to TTS quality
- More difficult in noisy conditions
- The driver is under cognitive load
- Short term memory in humans is limited

Four ways to render a location:

- Final normalization and conversion
- Address only
- Address & POI
- Intersection only
- Intersection & POI

| Intersection & POI | The corner of NE Woodinville Road and 131st Avenue, near City Hall |
|--------------------|------------------------------------------------------------------|
|                    | The corner of NE Woodinville Way NE and 17th Avenue, near Renton Chamber of Commerce |
|                    | The corner of East Madison Street and 17th Avenue, near City Hall |
|                    | The corner of NE Woodinville Road and 131st Avenue, near City Hall |

- 0.05 mi radius
Rendering locations

- Rendering the location is critical for the usability of the information system
  - Difficulties due to TTS quality
  - More difficult in noisy conditions
  - The driver is under cognitive load
  - Short term memory in humans is limited
- Four ways to render a location
  - Final normalization and conversion
Rendering locations

- Rendering the location is critical for the usability of the information system
  - Difficulties due to TTS quality
  - More difficult in noisy conditions
  - The driver is under cognitive load
  - Short term memory in humans is limited

- Four ways to render a location

| Type                  | Example                                      |
|-----------------------|----------------------------------------------|
| Address only          | 14803 Northeast 51\textsuperscript{st} Street |
| Address & POI         | 251 Rainier Avenue North, near Renton Chamber of Commerce |
| Intersection only     | The corner of East Madison Street and 17\textsuperscript{th} Avenue |
| Intersection & POI    | The corner of NE Woodinville Road and 131\textsuperscript{st} Avenue, near City Hall |

- Final normalization and conversion
Rendering locations: user study

- 40 users were asked to hear and type two addresses and to tell their preferences (133 trials)
- The accuracy is relatively the same, except for Intersection & POI case
- Users strongly prefer rendering with intersections
- Recommended order:
  - Intersection
  - Address
- “POIs are useful only when you know them”

| Question type           | Number | Sum  | Accuracy (%) |
|-------------------------|--------|------|--------------|
| Address only            | 67     | 57.5 | 85.82        |
| Address & POI           | 65     | 53.5 | 82.31        |
| Intersection only       | 65     | 54.0 | 83.08        |
| Intersection & POI      | 69     | 47.7 | 69.13        |

Recognition accuracy

User preferences
Deployment: calls analysis

- System demonstrated to ~800 Microsoft employees
- Total of 276 registered users
- Eight weeks period (March 12 – May 6, 2007)
- Total of 698 calls, ~12.5 per day
- 40 users accounted for 50% of the calls
- Total of 927 task attempts

| Task Type     | All | Registered | Non-registered |
|---------------|-----|------------|----------------|
| Traffic       | 3.56| 3.33       | 4.08           |
| Gas Prices    | 3.73| 3.54       | 4.14           |
| Weather       | 3.80| 3.61       | 4.41           |
| Total         | 3.65| 3.44       | 4.14           |

Average number of turns per task

Task completion ratio, %

September 1st 2007

SIGdial - Antwerp, Belgium
Deployment: user preferences

- Solicited 112 users and received 23 responses

| Statement                                                                 | Strongly Disagree | Disagree | Neutral | Agree | Strongly agree |
|---------------------------------------------------------------------------|-------------------|----------|---------|-------|----------------|
| I always ask for the same information from Commute UX                     |                   |          |         |       |                |
| I often use Personal Points of Interest other than work and home          |                   |          |         |       |                |
| I often use my Personal Points of Interest                                |                   |          |         |       |                |
| I know that I can give more information than the system asks for          |                   |          |         |       |                |
| I know that I can speak in a natural manner to Commute UX                 |                   |          |         |       |                |
| When I use Commute UX, I just answer the questions the system asks me     |                   |          |         |       |                |
| When I use Commute UX, I speak in a natural manner to ask for information |                   |          |         |       |                |
| With reasonable effort, I can get the information I expect from Commute UX|                   |          |         |       |                |
| Commute UX is easy to use                                                 |                   |          |         |       |                |
| Commute UX understands my speech                                          |                   |          |         |       |                |
| Commute UX is useful                                                      |                   |          |         |       |                |
Commute UX: demo

- Cheapest gas, city recognition
- Nearest gas station, intersection recognition
- Traffic, personal point of interest (home)
- Traffic, default route
- Traffic, via clause
- Weather, location passed as context

Demo 1

Demo 2
Conclusions and future work

- Created a telephone based free phrase dialog system for traffic, gas prices and weather
- Understanding and proper rendering of locations is critical for the system
- Using personal points of interest reduces the number of turns and increases the success ratio
- Future opportunities to improve the system through personalization, user adaptation, reducing the number of turns
Finally

Thank you for your attention! 😊

Feel free to give it a try:
   1-877-MSFT-511 (toll free in USA)
   Information for Washington state only
   Compare with 511 traffic info line

Comments: commuxfb@microsoft.com
Backup slides
System architecture (4)

- Built on top of Microsoft Speech Server
- SQL server powered backend
- Integrated into VS graphical design environment
Deployment: calls analysis (2)

- Peaks at morning and evening rush hours
- No substantial changes in the task proportion during the day

![Calls distribution during the day](image)

![Percentage of calls per task](image)
Deployment: calls analysis (3)

- Usage per day of the week

![Bar chart showing usage per day of the week]

**Usage per day of the week**

- **Average number of calls**
- **Day of the week**

- **September 1st 2007**
- **SIGdial - Antwerp, Belgium**