



ASRU Panel: Recording situated human communications

Alex Acero
Microsoft Research



Recording Situated Human communication

- Interested in human-machine communication
- Not in human-human communication
 - Unless that helps us build better systems

Recording Situated Human communication

- Not interested in recording per se
 - Unless that helps us build better systems
- There's nothing like **matched data**:
 - Don't collect lab data since it doesn't match real data
 - Build a real system and record those interactions
- There's no data like **more data**:
 - Build a useful system and you'll get lots of data
 - But building such a system is lots of work ☹️

Recording **Situated** Human communication

- Situated => in context
 - Not in a simulated context but a real application
 - How to bootstrap the process?



Natural Language

WindowsAirlinesApp - Microsoft Visual Studio

File Edit View Project Build Debug Data Tools Window Community Help

Debug Mixed Platforms

Task1.saml* Class1.cs Start Page

Saml Comp...
Pointer
CallerID
CallTransfer
CodeBlock
ExitPoint
MileStone
NLU
OfflineOpe...
PlayPrompt
RecordRep...
Scheduler
VoiceMenu
Task1
General

There are no usable controls in this group. Drag an item onto this text to add it to the toolbox.

```
graph TD; EntryPoint --> MainMenu[NLU: MainMenu]; MainMenu -- Reservations --> Itinerary[NLU: Itinerary]; MainMenu -- Flight Status --> FlightInformation[NLU: FlightInformation]; Itinerary -- Itinerary --> NumberOfTickets[NLU: NumberOfTickets]; FlightInformation -- Itinerary --> FlightStatusReport[PlayPrompt: FlightStatusReport]; MainMenu -- Announcement --> Announcement[PlayPrompt: Announcement]; FlightInformation -- BaggageClaim --> BaggageClaim[PlayPrompt: BaggageClaim]; NumberOfTickets -- Tickets --> ReservationSummary[PlayPrompt: ReservationSummary]; BaggageClaim -- Always --> RecordReplay1[RecordReplay: RecordReplay1];
```

Solution Explorer - ...

Properties

Microsoft.Tunes.NLU

AllowBarge	True
ClassifierU	Task1_Main
HelpDtmfK	
HelpPhrase	
Recognitio	(Collection)
RejectionT	0
RepeatDtn	
RepeatPhr	
SemanticSl	(Collection)
W3CGramr	

RecognitionBranches
Recognition Branch Collection

Rebuild All succeeded

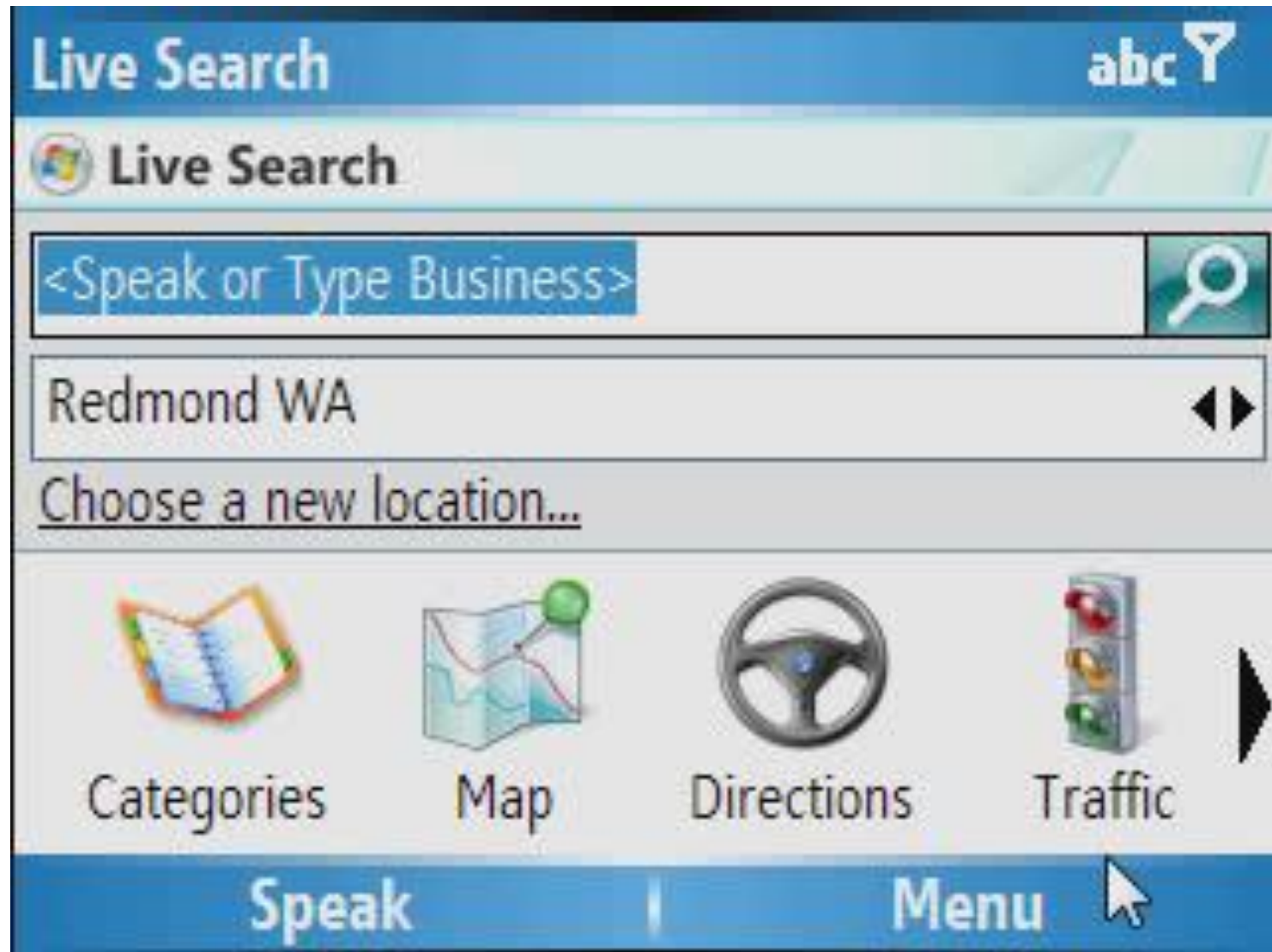
Voice access to phones



Multimodal Systems

- Users will use speech only if it's more convenient than the alternatives:
 - Users with data-enabled phones can type their request too and will use speech if it's more accurate.
 - An image is worth a thousand words. Looking at a screen with a map is more useful than TTS

Multimodal voice search



Multimodality Benefits

- Compared to speech-only:
 - User sees system response more quickly
 - User sees what system understood
 - User can know what system expects
- Compared to GUI-only:
 - Faster entry
 - Better use of small screen

Summary

- Not just recording the communication but
- Conducting usability studies
 - For novice users
 - And longitudinal studies for habitual users
- There's no data like real (matched) data
- There's no data like more data

Thank you