Looking back to move forward:

What dialog research can tell us

Marie Meteer, Ph.D.
Adjunct Professor Brandeis University

April 15, 2013
Characteristics of an “Advanced Dialog System”

- Mixed initiative
  - User can change the topic or revisit previous dialog elements
  - System can take control to get more clarifying information
  - User can say anything at any point and get some intelligent response

- Context aware
  - “Remembers” what the user has already said and uses that information
  - Recognizes the users goals based on what has happened so far
  - Tracks the users “focus”
More Characteristics of Advanced Dialog

- Multimodal
  - Input and output can be audio/visual/tactile
- “Natural” interaction, variability of expression
- “Self aware”
  - Knows when to ask for clarification and when to move forward
- Helpful
  - Offers additional information or incorporates multiple steps
- Explore some of these concepts in the Advance Dialog Workshop up next in this room!
Today’s standard IVR approach

Finite state grammar defines all word sequences and meanings

Context: Defined by call flow
Next action Defined by call flow

Voice Data

Dictionary

Grammar

Call Flow

Actions / Backend

Acoustic model

Prerecorded speech or TTS

Task Manager

Voice Images

Context: Defined by call flow
Next action Defined by call flow

Finite state grammar defines all word sequences and meanings

Prerecorded speech or TTS

Task Manager
“NextGen” adds NLU

- Dictionary
- Acoustic model
- SLM, Topic ID, Partial Parsing
- Call Flow
- Actions / Backend

- Voice Data
- Voice Images

- SR with SLM
- SSM: one label per model
- NLU: Grammar fills slot values

Context: Defined by call flow
Next action Defined by call flow

Prerecorded speech or TTS

Task Manager
Voice Search

Voice Data

Voice Images

Dictionary

Acoustic model

SLM

Info Extraction

SR with SLM

Search term extraction

Form query

No context

Format results

Search Engine

Actions / Backend

Prerecorded speech or TTS, Text to screen
Dialogue System Architecture

- Acoustic model
- Grammar/SLM
- Parser/Topic ID/NLU
- Dialog strategies Task models
- Actions / Backend

Voice Data
- Speech Recognition
- Natural Language Understanding
- Dialog Manager
  - Context update
  - Response Strategy Selection
- Task Manager

Voice Images
- Speech Synthesis (TTS)
- Natural Language Generation
Example

- When is my Visa bill due
  - Monday, April 15th
- How much do I owe?
  - Your balance is $583.24 and the minimum payment is twenty five dollars.
- Would you like to pay your bill?
- How much do I have in my checking account?
  - You have 164.25 in your checking account
- How about my savings
  - You have 2019.97 in your savings account
- Transfer $500 from my savings to my checking
  - $500 has been transferred from your savings to your checking account
- OK. Now pay the minimum

Still asking about Visa bill

Same goal, different account

Back to Bill Payment.
How do we move forward?

- What goes in the “Discourse Manager” box?
  - Context created by the discourse so far
  - Knowledge of the tasks the system is capable of
  - Knowledge of the input and output modalities and how they relate
Discourse Structure  
Grosz and Sidner, 1986  

- Three components  
  - Linguistic: sentence, phrase structure, cue phrases  
  - Attentional: Focus of attention  
  - Intentional: Speaker’s purpose and the relations between those purposes  

- “Structure” is key  
  - Originally described as a stack  
  - Really needs to be a persistent representation  
    - Once said, things never really go away
What is “Intentional” Structure

- **What are they trying to do?**
  - Intention, plans, goals

- **DAMSL (Dialog Act Markup...)**
  - Dimensions in Dialogue Act Annotation

- **Tasks**
  - Ravenclaw
DESIGN... for Task Completion

Use Case

UX

Task

Conversations

Dialogs
Collections/Error/Feedback

Intent

NLU

Concepts
Entities

Slots
What is “Attentional” Structure

What are they talking about?

• Focus, anaphora

• Centering
Referring Expressions

- Referring” goes beyond objects
  - “According to Doug, Sue just bought a 1962 Ford Falcon”
    - But that turned out to be a lie (a speech act)
  - But that was false (proposition)
  - That struck me as a funny way to describe the situation (manner of description)
  - That caused Sue to become rather poor (event)

- And beyond speech
  - Is that one open tonight?
  - Is it open tomorrow?
Intentional and Attentional Structure

**Intentional**
- Pymt_info
  - get date
  - provide date
- PayBill
  - get amt
  - provide amt

**Attentional**
- [visa] [when?]
- [visa] [D 4/15] [amt?]
- {OBJ visa due 415 bal: $A Min: $B}
- [visa]
- [Bal: $A] [Min: $B]
- [CA]
  - {ACCT ch bal: $C}
  - {ACCT sav bal: $D}
- [SA]
  - Transfer$
  - PayBill
  - checking
  - savings
  - provide amt
Where in the Architecture

I’d like to **get**, **get my bank balance**, uh **checking balance**

Task = GetBalance  Account = Checking

- Speech recognition uses a statistical language model
  - User can express their requests many different ways
  - Does not have to fit a specific “grammar”

- **Statistical task / intention recognition**
  - System recognizes what the user is trying to do
    - Get balance, pay bill, view transactions

- **Information extraction**
  - System picks out the specific pieces of information to accomplish the task
    - Credit card name, date, account type
And what about that statistical stuff?

- Should we be using statistical models at the discourse level?
  - YES!

- BUT:
  - What you are modeling is critical
  - Can’t build a model without data!!!

- Bootstrap
  - Build in targeted behavior
  - Never throw out your data
  - Beg, borrow and steal*
    - *with permission, of course
Bootstrapping: 1st level
Know your App

Out
But predictable
Define specific repair strategies

In the App
Define Intentional Attentional Structure

HUH?
Define Repair Strategies

© MM Consulting 2010
What can research tell us?

- What information to model
  - What’s working, what’s not
  - What’s coming down the line
  - SIGDIAL:  [www.sigdial.org](http://www.sigdial.org)

- How do I find people that can do this stuff?
  - Brandeis University Computational Linguistics MA
  - [www.brandeis.edu/programs/comp-linguistics/](http://www.brandeis.edu/programs/comp-linguistics/)
  - Or contact me at mmeteer@mac.com

**Thanks!**