Multi-modal Multi-turn Applications
Usability Testing

Silke Witt, Ph.D.
Fluential
Multi-modal Multi-turn Applications

- Definition
- Platform
- Design Challenges
- Sample Application: Travel Domain
- Usability Options
- Lab-based Usability Results
What is a “Multi-modal Multi-turn Application?"

“ The situation where the user is provided with multiple modes for interacting with a system”

- Users can use one of multiple input modalities
- Users can switch the modality they are using
- Users can use multiple modalities simultaneously
- 1 interaction with the system can span multiple turns
- Output modality can be a combination of several modalities used simultaneously
Fluential’s Multi-modal Multi-turn Platform

**Targeted Speech Recognition**
- Domain specific vocabulary and domain modeling
- Reduce recognition error rate

**Personalization**
- Intelligent ranking
- Knows user preferences

**Natural Language Understanding**
- Understands user intention
- Contextual meaning
- Determines all salient information in a query

**Multi-topic conversation**
- Can switch between topics
- Remembers context

**Conversational Dialog**
- Determines user goal
- Identifies missing information
- Interacts with user to get info
- Unstructured conversation
Multi-modal Multi-turn Architecture

Mobile Application

Fluential API

- Start recog.
- Tex/Touch entry
- Key/ value pairs
- Data export API

Audio streaming

Interaction engine

User data & context knowledge

Domain specific ASR

NLU engine

Domain Knowledge Database

UI layer

User data & context knowledge

Tex/Touch entry

Key/ value pairs

Start recog.

Data export API
Documentation for Multi-Modal Multi-Turn Design

VUI Design
Flow charts

GUI Design
Wireframes

MM-MT Design
A combination
Multi-modal Multi-turn Use-case Documentation

AIR USE CASE – SEARCH CITY

1. I need to fly to New York on June 5th
   - Visual volume bars while user is speaking
   - Prompt sound (1) when mic. button is pressed

2. Wow – not waiting long at all
   - Waiting Sound (2) while Sofia is thinking

3. Finished sound (3)
   - Three airports serve New York. Which one do you want?
   - JFK
   - LaGuardia
   - Newark

4. “JFK” or user can say: “I want to fly into JFK”
   - Ok. Flying to JFK on June 5th. Where are you flying out of?

5. I’m flying out of Detroit and I want to come back on the thirteenth
   - Ok. Flying from Detroit to New York on June 5th, returning June 13th

6. Here’s what I found
   - Delta
   - Virgin
   - American

© 2012 Fluential Inc. All rights reserved
Travel Domain Application

- Spoken natural language interface
- Multiple simultaneous data input
- Understanding of relative dates, location, travellers
- GPS integration
Travel Domain Application

• Weather forecast and average information
• Search for flights and hotels
• Sort / Filter/ detailed view of search results:
  • “Show me only United flights that leave between 9 and 11am and sort by the cheapest”
Multi-modal Design and Usability Testing

Speech is an intuitive interface

A double-edged sword

Easy interface no need to figure out what hides behind buttons **BUT** how to tell users the system’s capabilities?

With speech, people expect understanding and intelligence as well

Speech to control mobile application is a relatively new paradigm for users
Multi-modal Usability Questions

Design guidelines exist for:

But what about Multi-modal, Multi-turn interaction design: Which guidelines from these three areas apply, which ones don’t?

- Match the modality, i.e. if voice in, then voice out?
- Use the most efficient modality? Voice in, Visual out?
- Let user set the preferred modality combination?
- Do the output via multiple modalities?
Multi-modal Usability Questions

**Microphone:**
- Where should it be placed?
- How should it work?
- What size, colors, sounds to go with it?

**Multiple output modality alignment:**
- Application is presenting information using both visual and voice interfaces
- Do these two need to be synchronized?
- Or can they complement each other? Listening is transient and takes longer than reading.

**User education:**
- How tell to users the system capabilities?
- Videoclip Tutorial
- Just-in-time hints
- Include information permanently on screen
Usability Options on MM-MT mobile applications

- Lab-based usability testing
- Focus groups
- Crowd-sourced usability testing

Increase in participant numbers
Decrease in cost
Usability Protocol

Lab-based usability

8 subjects

1 camera to record interaction, 1 camera to record screen

Challenge with crowd-based study was to that audio can’t be captures.
Overall Findings

• Users completed 84% of the tasks
• Overall positive user feedback:

  "The interface was simple. I liked the cute little face."

  "I like that the app speaks to me."

  "I liked the fact that no need to repeat like with so many voice applications"

The system is intelligent and helpful!

BUT we also learned a lot! The biggest take-away:

The multi-modal multi-turn functionality was working, but we don’t tell users to what the app can do or how to use it!
More Usability Findings

Focus of study was on general usability issues with the aim to derive design guidelines for this spoken multi-modal mobile applications.

Top 4 Take-away’s:

- Speech is an ‘invisible’ modality
- Evolving microphone behavior standard causes turn-taking issues
- IVR-legacy driven telegraphic speech causes errors
- Feedback: Apps needs to display details of each understood data point
Speech as an ‘invisible’ modality

Initial Tutorial:
- Current system only has a simple hint at the start of the application

IVR legacy:
- 7 out of 8 users needed verbal clarification after completing one task to switch from telegraphic speech to natural language

Just-in-time Hints:
- The error correction capabilities were not clear to all users

Multi-modality:
- People generally liked the combined audio and text confirmation, but wanted it to be more explicit
- One user only used free from text entry even though the usability was clearly about a ‘speech’ usability
Microphone

Current Functionality:
Tap to start record, tap to end or timeout to stop

- Default screen
- "July 5th through the 17th"
- Processing animation

Tap to Start
- Multiple users started speaking without pressing the mic button

Tap to Stop
- Only 1 user learnt how to tap to stop speech

Color Schema
- Request for brighter color to stand more out

Sound Animation
- Percolator sound during wait was liked.
- Users also wanted ‘click’ for start and stop
Thank you!

Contact info: Silke Witt, switt@fluential.com