



Multimodal Design

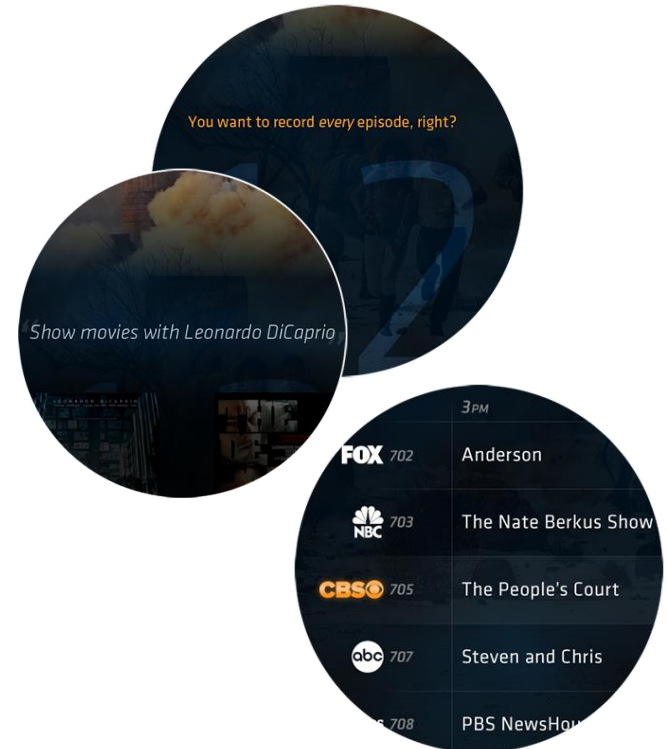
March 2014

Multimodal

Multiple interaction modes, typically allowing either more than one discrete way to interact (i.e. speech or text) or combinations of discrete modes (i.e. speech-in, text-out)

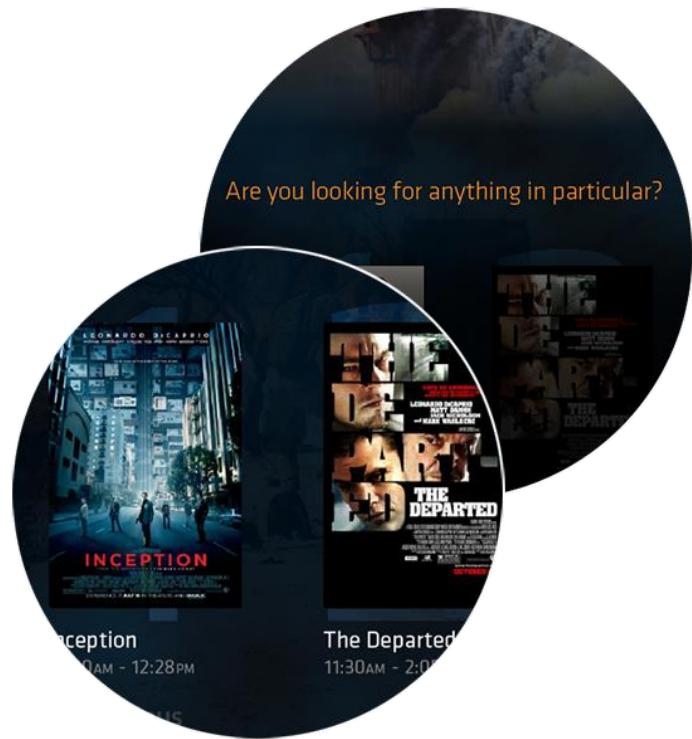
Speech and Graphical Interfaces

- Speech Interfaces
 - Application-Initiated
 - Accomplish well-defined tasks
 - Perform simple disambiguation
 - Eyes-free, hands-free environments
 - User-Initiated
 - Provide complex input
 - Negotiate numerous variables
 - Allows various paths
- Graphical Interfaces
 - Display and explore lots of output
 - Present, review complex information
 - Select from ambiguous alternatives



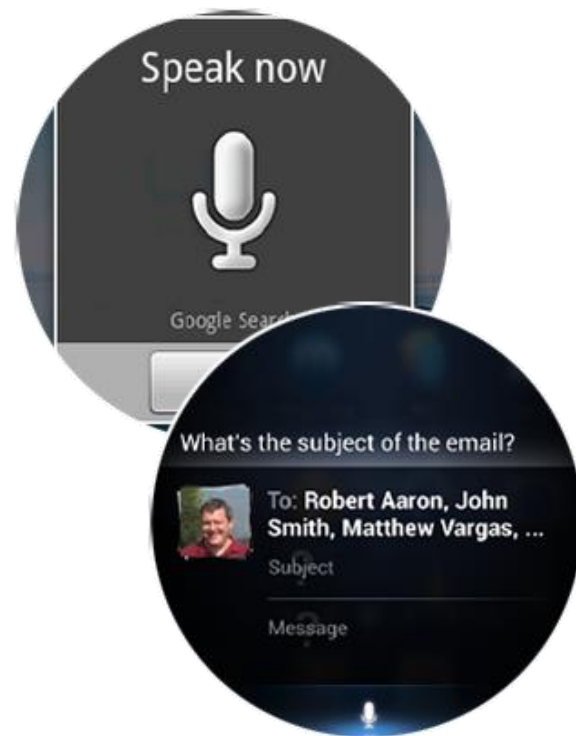
Fundamentals of Multimodal Speech Design

- Be aware of available modalities
- Leverage and foreground modal strengths
- Ensure interoperability, with modes working in concert
- Speech can be leveraged as an interface, not just an input
- Create a conversational affect



Input vs. Interaction

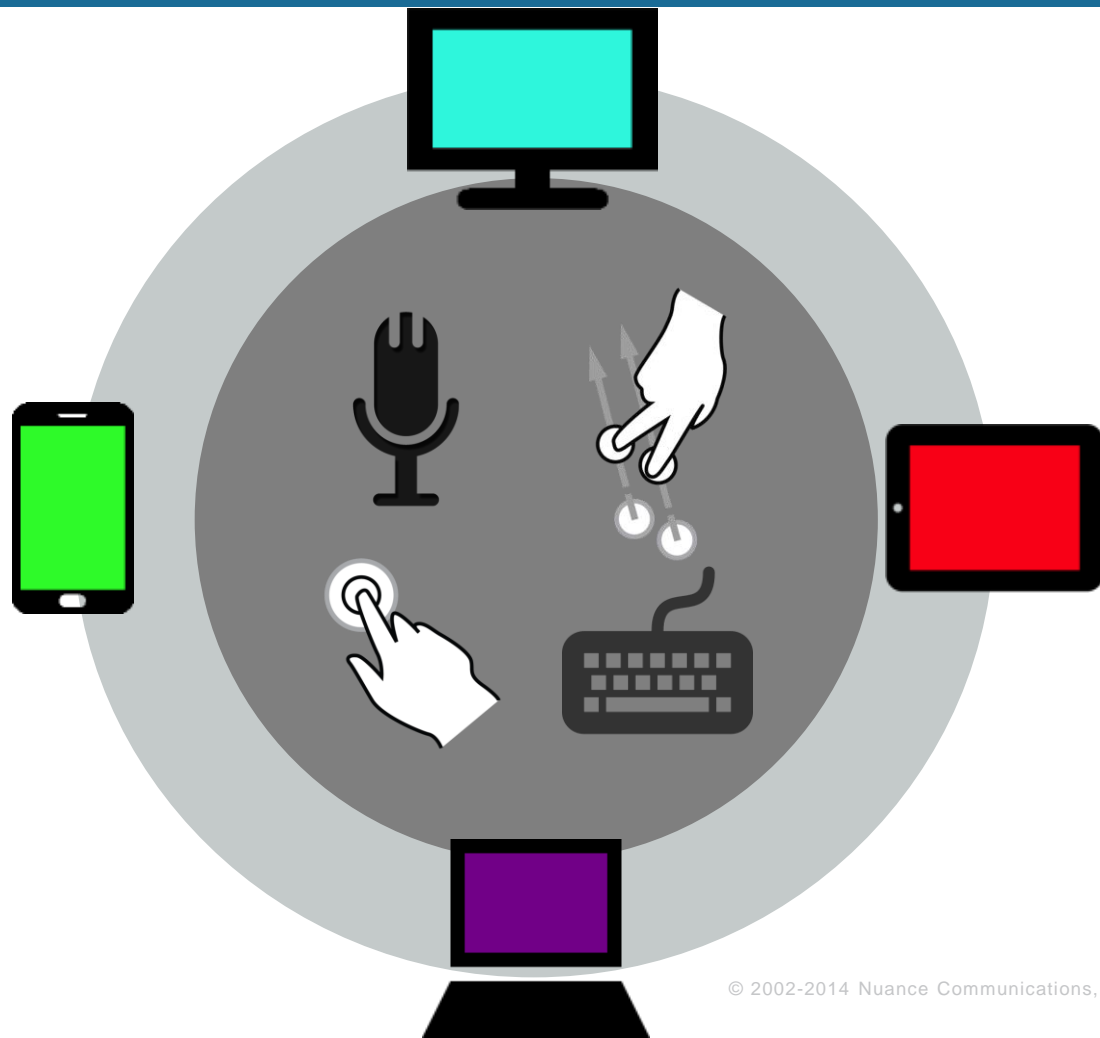
- Leverage users' experience
- Eliminates the learning curve
- Reduces users' cognitive load
- Accommodates unresolved queries
- Provides richer feedback



Advantages of Multimodal Design

- Frictionless experience
 - Simplifies and removes barriers to task completion
- User agency
 - Users have freedom to work across modalities
- Task efficiency
 - Modes working in concert are more effective than one
- Flexibility
 - System adapts to the environment and user's needs





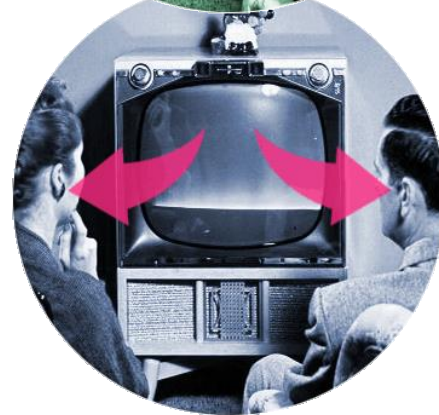
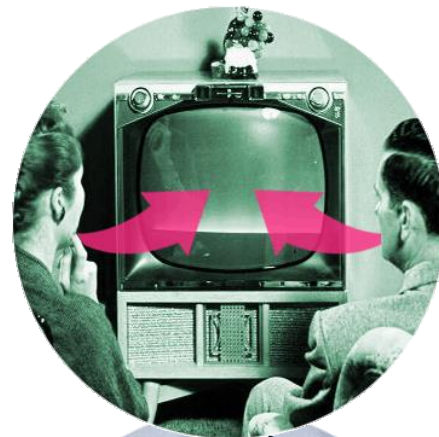
Utilizing Speech

- Advantages

- Allows expression of complex information
- Flattens hierarchies, streamlines navigation
- Simplifies obscure technical tasks
- Mitigates some accessibility issues

- Challenges

- Speech makes the interaction public
- Environment not always conducive to speech
- Alone, introduces other accessibility issues



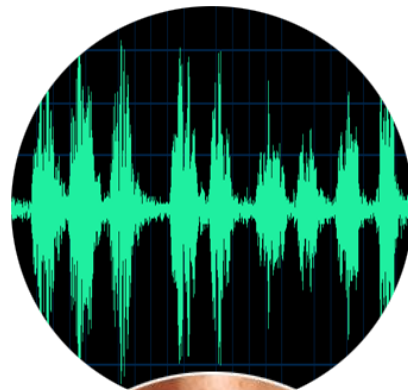
Interacting with the System

- Design for accessibility
 - Form factor
 - Functionality
 - Environment
- Touch to talk
 - Tapping an explicit microphone button
- Wake-up Word
 - Always on, always listening
- Gestures
 - Raising hand in front of a TV
 - Flicking from a smart watch bezel



Maximize Interface Transparency

- Structure the interaction
 - Moderate the dialog – based on context, user behaviors, preferences – to avoid distraction
- Adaptive dialog
 - Taper prompting to pace the experience and dilate it to provide instruction as needed
- Match modality with the user
 - The dialog should take its cues from the user, advancing and receding based on behavior



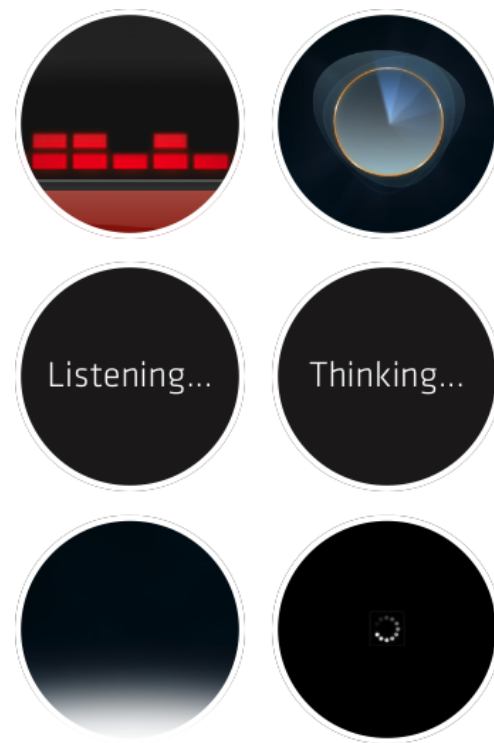
Navigation

- Adaptive UI
 - VUI and GUI should accommodate multiple paths to task resolution
- Deemphasize Hierarchy
 - The structure of the application is implicit in the VUI, but explicit in the GUI
- Visual reinforcement
 - Cue context and available options using icons, transformations, and subtle enumeration



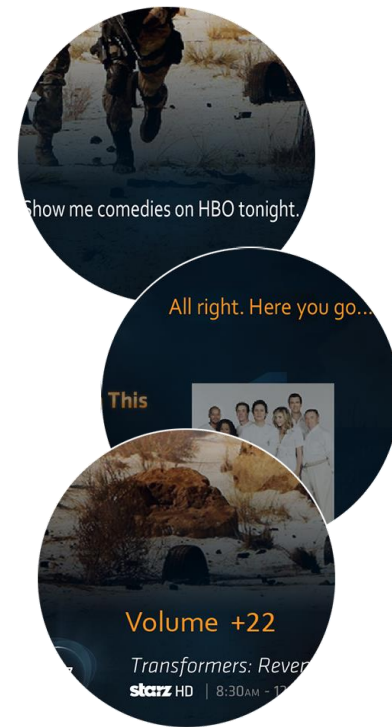
Feedback and State Awareness

- Conversational Cues
 - Dialog itself should conform to normative conversational behavior
- Visual and Audio Cues
 - Absent non-verbal cues – like facial expressions and gestures – visual and audio feedback provide help the user understand context
- Two is Better Than One
 - Dialog and visual feedback work together to reinforce state awareness



Dialog & Transcription

- Advantages
 - Since speech itself is fleeting, visually representing the dialog can help users understand context
- Dialog Design
 - Dialog should be designed – using discourse markers, acknowledgements, etc. – so that the user can easily interpret standard cues
- Focus on the Content
 - The literal meaning of what the user says is less important than the semantic content of their input



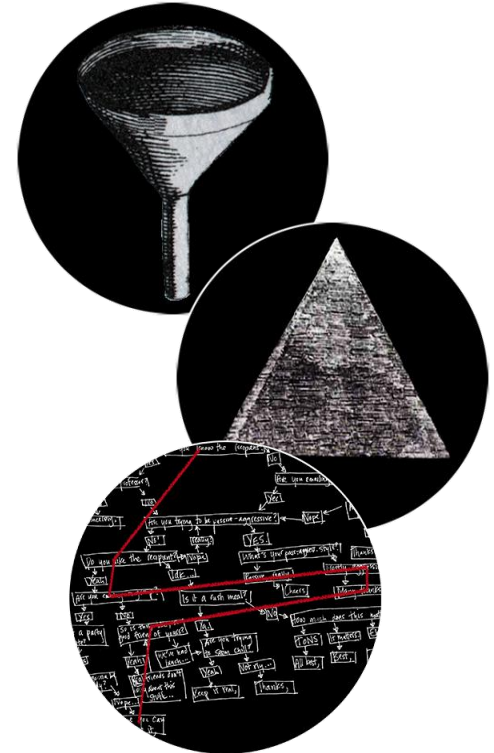
Environment

- Assume attention is divided
 - Mobile and on-the-go
 - TV viewing environment is often shared, distracting
 - Multi-tasking prevalent in desktop interactions
 - Driver always focused on the road and driving
- Assume the stakes are low
 - Relax constraints and reduce confirmations
- Except when they're not
 - Be careful when there are monetary transactions or public-facing information like social network posts involved



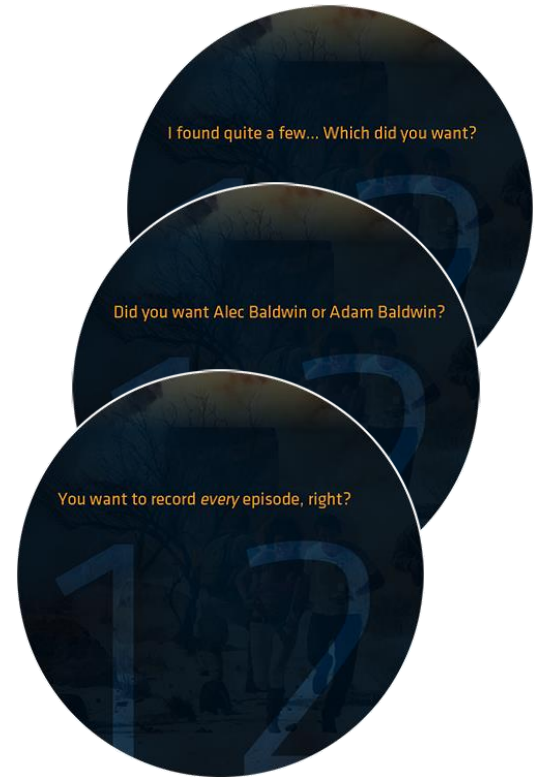
Content Discovery

- Search is a Funnel
 - Focused on winnowing to a single best option
 - Ease of refinement
- Browsing is a Pyramid
 - The goal is wide-ranging exploration of choices
- Users' intentions are mixed
 - The application must anticipate shifting goals, various strategies, and lateral exploration



Disambiguation: Different Contexts Require Tailored Strategies

- Manageable array
 - Differs based on form factor, interface modalities
 - Bring preferences to bear on organizing results
- Unmanageably large array
 - Filter based on preferences, user profile
 - Prompt for refinements
- No results
 - Relax search constraints
 - Prompt for refinements
 - Offer alternatives



Refinement

- Provide for complex input
 - Accommodate complex ‘multi-slot’ queries
- Prepare for refinement
 - Design dialogs to guide the user toward ‘task sufficiency’
- Handle a range of input
 - Articulating complex requests imposes a high cognitive load on the user



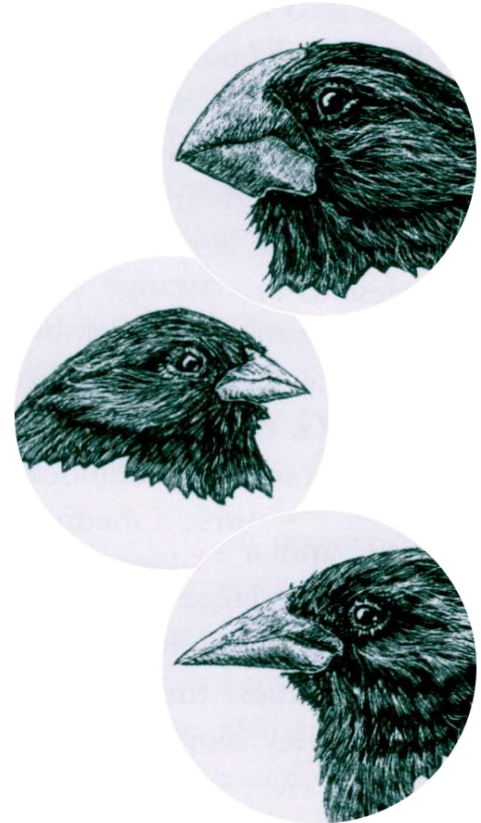
Command vs. Conversation

- Exploit the power of language
 - Command-based systems are blunt and unnatural – embrace semantic richness
- Leverage users' experiences
 - Most users have been carrying on conversations for a long time – they know how to do it
- Understand the limits
 - Avoids tasks that are difficult for users to articulate and fall back on other modes



Initiation and Assistance

- Adapt to user facility
 - Adjust the interaction – prompting, pace, strategy
– as users gain experience and expertise
- Out-of-the-box
 - Start out slow, thorough, and patient and gradually reduce support in favor of efficiency
- Situational instruction
 - Offer hints and use just-in-time help to provide instructions in the context of the user's task



Handling Errors

- Misunderstandings are inevitable
 - Just as in person-to-person communication, misinterpretations are a fact of life
- Fail gracefully and recover
 - Acknowledge mistakes and offer strategies to get back on track
- Contextual assistance
 - When problems arise, escalate and dilate to guide the user toward the goal



In Closing...

- Leverage and foreground modal strengths
- Ensure seamless interoperability
- Leverage users' experiences and natural instincts
- Utilize speech to significantly simplify otherwise complex interactions





Thank you