

Research in the Voice Space: What's Coming Next?

Robert Dale
rdale@language-technology.com

The Aim of This Talk

- To provide a 'heat map' of new ideas and developments in spoken language dialog systems research

But First ...

- How do you think speech apps will be different in 5 years' time? 10 years' time?

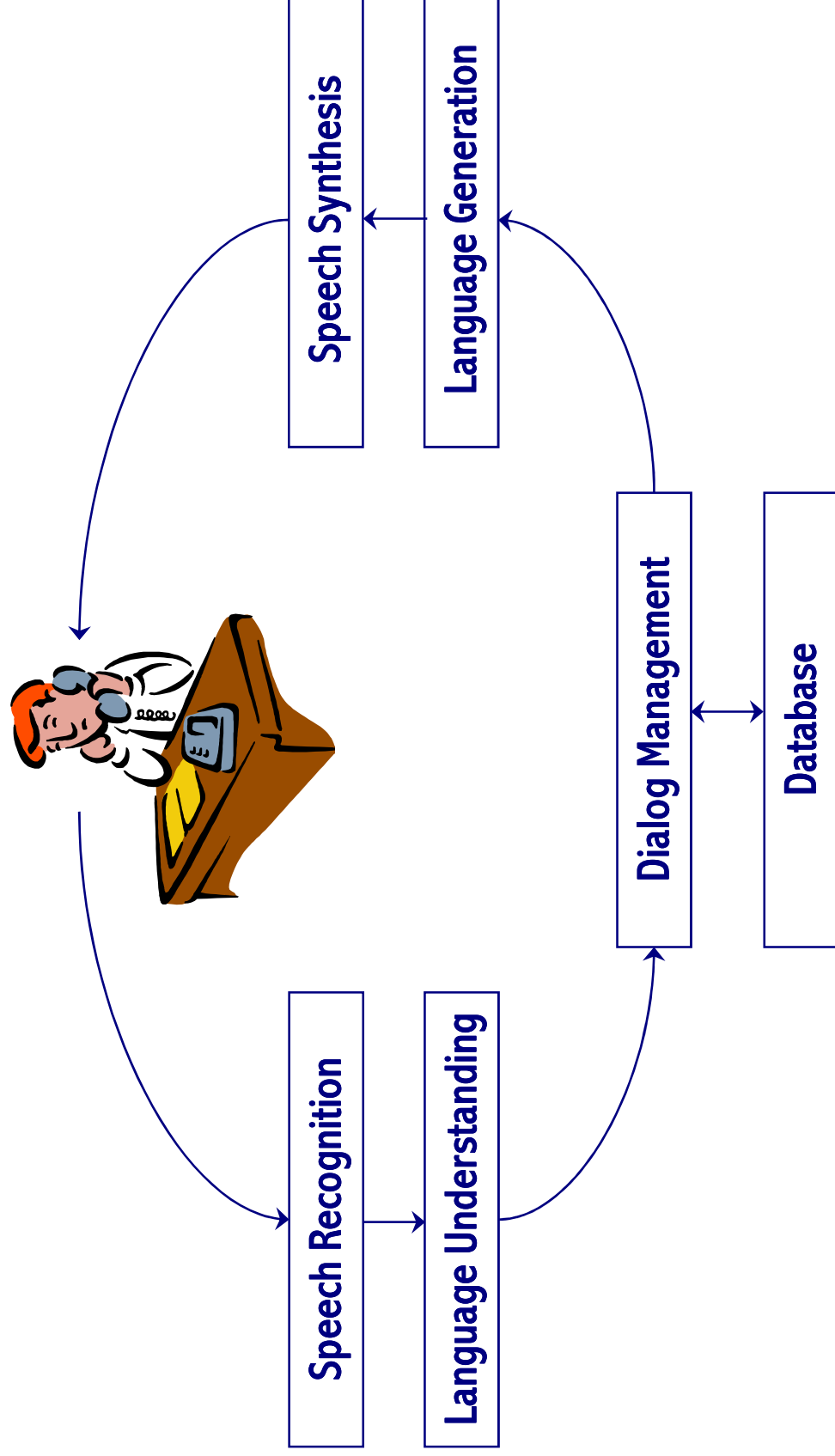
Outline

- **Research and Practice in Dialog Systems**
- **Hot Spot #1: Incremental Utterance Interpretation**
- **Hot Spot #2: Adaptation and Alignment**
- **Hot Spot #3: Embodied Conversational Agents**
- **Hot Spot #4: Advanced Dialog Modelling Techniques**
- **Other Ideas to Watch**
- **Finding Out More**

Research and Practice in Voice Systems

- **The focus in commercial deployments:**
 - **Maximise usability and complete the call**
- **The focus in research labs:**
 - **Make dialogs more natural and human-like**
- **These may not be compatible goals:**
 - **See Bruce Balentine's "It's Better to Be a Good Machine Than a Bad Person"**

The Focus: Spoken Language Dialog Systems



Hot Topics We'll Be Looking At

- **SLDS technologies in development in research laboratories now and over the last 5 years that we might expect to find their way into applications in the next 5 years**

Hot Topics We Won't Be Looking At

- Dialog with robots
- Handling multiparty speech in meeting room environments
- In-car applications

Outline

- **Research and Practice in Voice Systems**
- **Hot Spot #1: Incremental Utterance Interpretation**
- **Hot Spot #2: Adaptation and Alignment**
- **Hot Spot #3: Embodied Conversational Agents**
- **Hot Spot #4: Advanced Dialog Modelling Techniques**
- **Other Ideas to Watch**
- **Finding Out More**

Status Quo

750ms min. silence



Problems:

- Psycholinguistic evidence
- Efficiency
- End-of-utterance

User

ASR



Parser



Dialogue manager



Generator



TTS

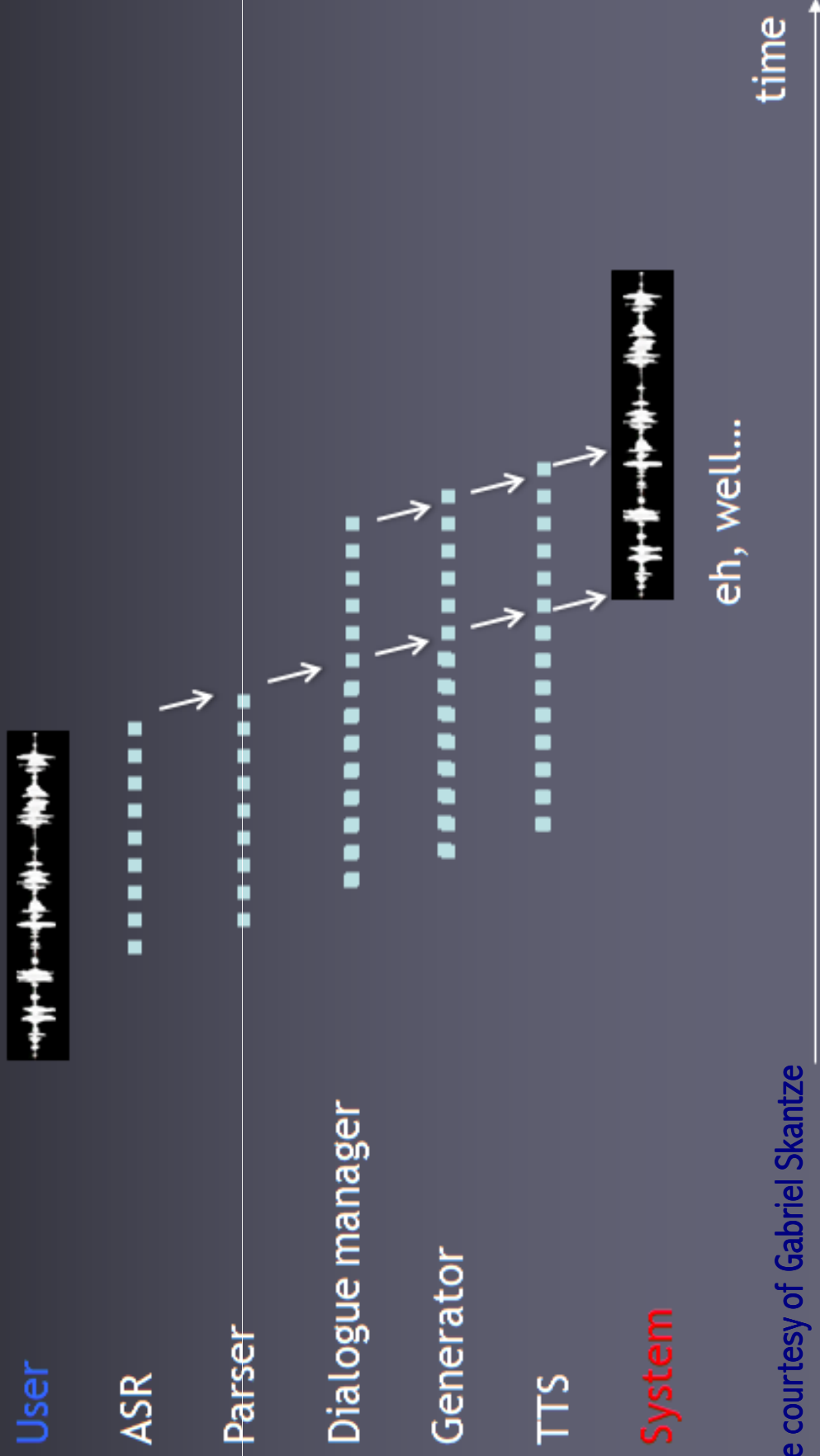


System



Incremental production

Starting to speak before knowing exactly what to say



The Realities of Conversational Behaviour

- We interpret and generate incrementally
- We are adept at knowing when to talk or interrupt
- We are able to see past disfluencies

Adding These Capabilities to Applications

- [The NUMBERS system](#) [Schlangen and Skantze 2009]

Prospects

- **Barge-in is the simplest possible form of 'incremental processing'**
- **More sophistication requires radical changes to the way we construct systems**

Outline

- **Research and Practice in Voice Systems**
- **Hot Spot #1: Incremental Utterance Interpretation**
- **Hot Spot #2: Adaptation and Alignment**
- **Hot Spot #3: Embodied Conversational Agents**
- **Hot Spot #4: Advanced Dialog Modelling Techniques**
- **Other Ideas to Watch**
- **Finding Out More**

Adapting to the User

- **Already happens in a trivial sense:**
 - **Dialog state represents a changing 'user state'**
- **A little more sophisticated:**
 - **VoiceXML's Form Interpretation Algorithm uses one-item-at-a-time as a fallback to mixed initiative**

Better Adaptation to the User

- Simple ideas:
 - Don't say what you can't recognize
 - Store a persistent user history

Even Better Adaptation

- What people do: alignment
- Places where alignment happens:
 - Speaking rate and volume
 - Word choice
 - Syntactic structure choice
 - Conceptualisation

Prospects

- **Simpler aspects of alignment are already technically feasible:**
 - **Choose prompts to play on the basis of which grammar rule was triggered by the input**
- **But where's the business case?**
 - **May provide a way of addressing customer dislike of default personas**

Outline

- **Research and Practice in Voice Systems**
- **Hot Spot #1: Incremental Utterance Interpretation**
- **Hot Spot #2: Adaptation and Alignment**
- **Hot Spot #3: Embodied Conversational Agents**
- **Hot Spot #4: Advanced Dialog Modelling Techniques**
- **Other Ideas to Watch**
- **Finding Out More**

Multimodal Dialog

- **Multimodal input**
 - **Touch and voice**
- **Multimodal output**
 - **Speech and image**

Embodied Conversational Agents

- GRETA [from Catherine Pelachaud, INRIA]

Prospects

- **Only appropriate for situations where you can look at a screen**

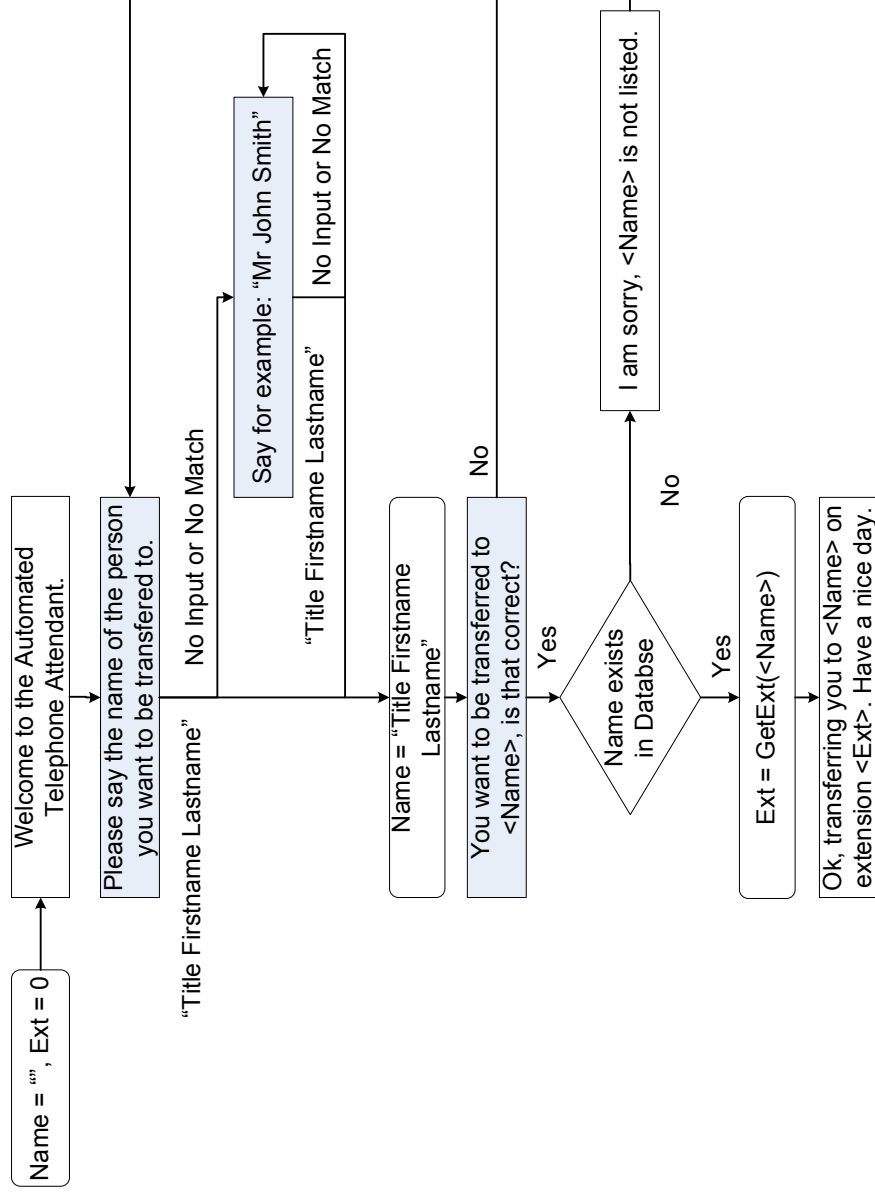
Outline

- **Research and Practice in Voice Systems**
- **Hot Spot #1: Incremental Utterance Interpretation**
- **Hot Spot #2: Adaptation and Alignment**
- **Hot Spot #3: Embodied Conversational Agents**
- **Hot Spot #4: Advanced Dialog Modelling Techniques**
- **Other Ideas to Watch**
- **Finding Out More**

Current Dialog Models

- **Dialog management = deciding what to do next**
- **The state-of-the-art: scripted dialogs via VoiceXML**
- **The bottom line: all deployed systems adopt a finite-state model of dialog**

Finite-State Dialog Models



Inference-Based Approaches

- What drives the dialog forward is an independent data structure:
 - An information state
 - An agenda
 - A system goal
- Consequence: the system has to reason about what to do next

Inference-Based Approaches

- **The problem:**
 - These systems are hard and expensive to engineer
- **A solution:**
 - Let the dialog system learn how to respond in different circumstances

Markov Decision Processes

- Dialogue management construed as a reinforcement learning problem:
 - Learn the appropriate sequence of actions that minimises some cost function
- First appeared in the literature over 10 years ago
- A very active area in the last 3-4 years
- Problem: Where do you get the data to learn from?
- One solution: Simulated users

Prospects

- **In theory, this approach provides a solution to the hard problem of designing effective dialogs**
- **In practice, developers and customers are likely to be uncomfortable adopting systems that are not under their control**

Outline

- **Research and Practice in Voice Systems**
 - **Hot Spot #1: Incremental Utterance Interpretation**
 - **Hot Spot #2: Adaptation and Alignment**
 - **Hot Spot #3: Embodied Conversational Agents**
 - **Hot Spot #4: Advanced Dialog Modelling Techniques**
- **Other Ideas to Watch**
- **Finding Out More**

Other Ideas to Watch

- Emotion detection
- Speech Graffiti

Emotion Detection

- 'Speech analytics'
- The state of the art:
 - Tracks volume and pitch to determine emotional state
 - Correlation between these features and the actual emotional state is not a simple mapping

Speech Analytics Vendors

- **Autonomy etalk:** <http://www.etalik.com>
- **Nemesysco:** <http://www.nemesysco.com>
- **VoiceSense:** <http://www.voicesense.com>

Speech Grafitti

- The Universal Speech Interface
- The ETSI Standards

Universal Speech Interface Keywords

- **Goal: around 10 keywords**
 - **NOW WHAT?**
 - **GO!**
 - **SCRATCH THAT**
 - **START OVER**
 - **REPEAT**
 - **RESTATE**
 - **MORE, NEXT, PREVIOUS, STOP**

A Sample Dialog with USI

- U1 Neighborhood is Shadyside
S2 Shadyside
U3 Now what?
S4 Distance to campus is, rent is, furnished is ...
U5 Rent is less than 700 dollars
S6 Less than 700 dollars
U7 Furnished is now what?
S8 Furnished can be yes, no, partially furnished
U9 Furnished is no, number of bedrooms is two. Go!
S10 Unfurnished, two bedrooms. 4 matches. At 5555 Forbes Ave, apartment, \$650, 0.5 miles from campus
U11 Start over
S12 Starting over
- U13 Neighborhood is Regent Square
S14 Wilkinsburg
U15 Scratch that. Neighborhood is Regent Square
S16 Regent Square
U17 Distance to campus is less than 3 miles, number of bedrooms is at least 2
S18 Hmm... at least 2 bedrooms
U19 Distance to campus is less than 3 miles
S20 Less than three miles
U21 Rent is less than \$1200
S22 Hmm... I didn't understand 'rent is less than Oakland'
U23 Rent is under \$1200
S24 Less than \$1200

ETSI

- **European Telecommunications Standards Institute**
- **Generic spoken command vocabulary for ICT devices and services**
- **Requirements:**
 - **for users, the command words should be both intuitive and easy to remember**
 - **a speech recognition system requires commands to be easily discriminable.**

Context Independent Common Commands

Index	ICT device/service function and/or functions	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
1.1	List commands and/or functions	Options	Choix	Menü	Menù, Lista comandi	Opciones	Request for listing of available words (optionally with their functionality)
1.2	Terminate service	Goodbye	Quitter, Au revoir	Beenden	Spegni, Fine	Salir	End call, get off line, end session
1.3	Go to top level of service	Main menu	Menu principal	Haupt-menü	Menù principale	Inicio, Menù principal	Leave sub-application, go to main menu or application
1.4	Enter idle mode	Standby	Veille	Stand-by	Sospendi, Stand-by	Espera	Put the Automatic Speech Recognition (ASR) into monitoring mode for a wake-up command
1.5	Transfer to human operator	Operator	Assistance	Hotline, Service	Operatore, Assistenza	Operador	Leave the speech recognition mode and transfer to a human attendant, an operator, in telecommunications-specific contexts. This command should also be used when offering relay services
(See note)							
1.6	Go back to previous node or menu	Go back	Retour	Zurück	Indietro	Atrás	Navigate backwards in a dialogue structure (can also be used to cancel a forced choice operation)
NOTE:	The command "Helpdesk" would be recommended for IT-specific contexts. However, it conflicts with the common command 2.1 "Help" in several languages and causes recognition conflicts from the ASR point of view.						

Context Dependent Common Commands

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
2.1	Help	Help	Aide	Hilfe	Aiuto	Ayuda	Provide context dependent explanations and guidance (may provide more detailed help on repetition of the command)
2.2	Read prompt again	Repeat	Répéter	Wiederholen	Ripeti	Repetir	Repetition of the last acoustic feedback message

Outline

- **Research and Practice in Voice Systems**
- **Hot Spot #1: Incremental Utterance Interpretation**
- **Hot Spot #2: Adaptation and Alignment**
- **Hot Spot #3: Embodied Conversational Agents**
- **Hot Spot #4: Advanced Dialog Modelling Techniques**
- **Other Ideas to Watch**
- **Finding Out More**

Finding Out More

- **SigDial: ACL Special Interest Group on Discourse and Dialogue**
 - <http://www.sigdial.org/>
- **SemDial: Workshop Series on the Semantics and Pragmatics of Dialogue**
 - <http://www.illc.uva.nl/semdial/>
- **Interspeech annual conference**
 - <http://www.isca-speech.org/conferences.html>