



# **The State-of-the-Art of, and the Market for, Speech Technologies**

**Stuart R. Patterson  
LangTech  
November 24, 2003**

**ScanSoft®**  
**Productivity**  
**Without Boundaries™**

# Agenda

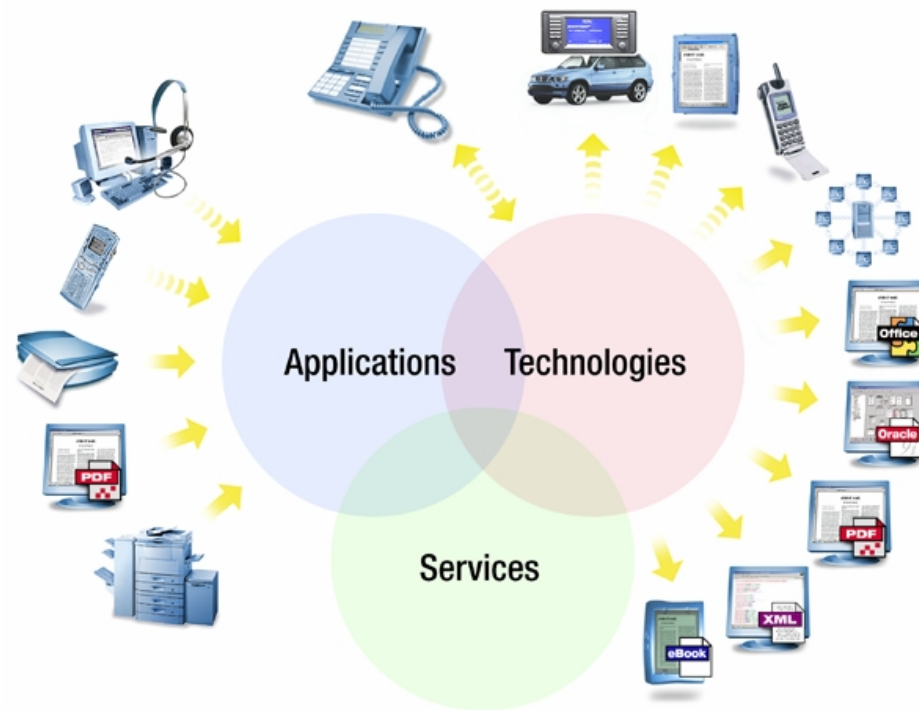
---

- ♦ Update on ScanSoft
- ♦ Converging Technologies
- ♦ Converging Markets
- ♦ Future Applications ... Future Solutions

# ScanSoft, Inc.

---

The leading supplier of speech and imaging solutions that increase productivity, reduce costs and improve customer service



# ScanSoft History

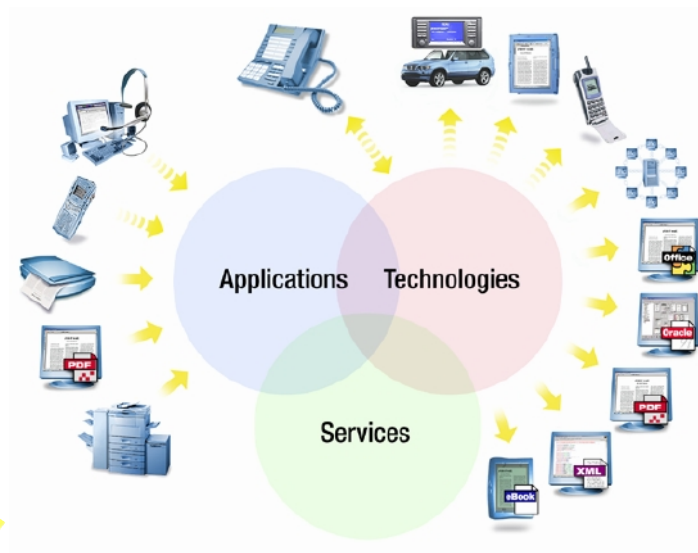


# ScanSoft at a Glance

---

- ◆ **#1 in all major product categories**
- ◆ **750 employees**
- ◆ **15 offices worldwide**
- ◆ **200+ patents or patents pending**
- ◆ **15 million registered users**
- ◆ **48 languages for speech recognition**
- ◆ **22 languages for synthesized speech**
- ◆ **9 languages supported by productivity applications**
- ◆ **2,000+ channel and platform partners**

# ScanSoft Targeted Solutions



## Productivity Applications

- Document conversion
- Digital paper
- Dictation
- eForms

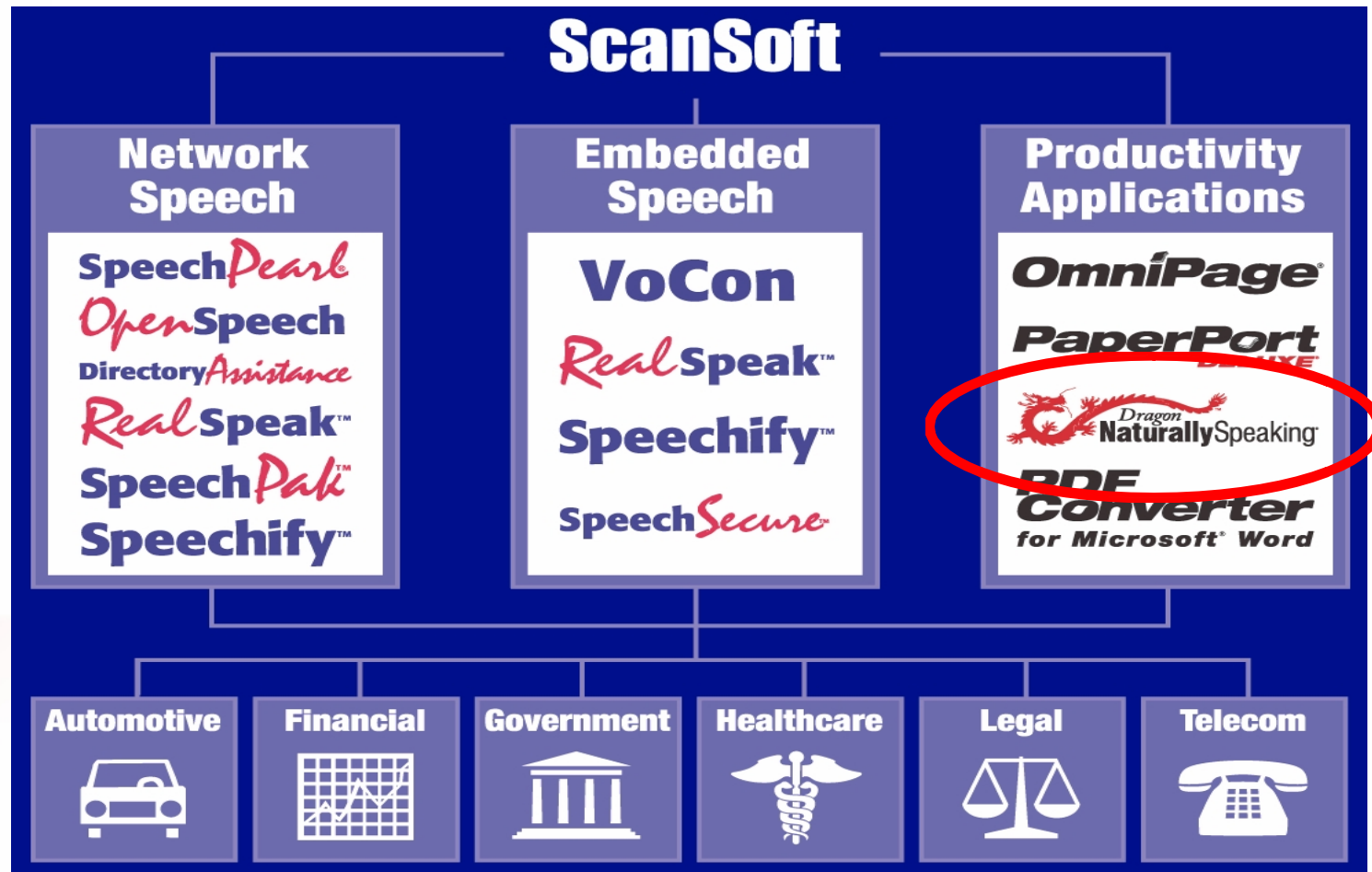
## Network Speech

- Speech recognition
- Text-to-speech
- Speaker verification
- Packaged applications

## Embedded speech

- Speech recognition
- Embedded TTS
- Speaker verification

# Bringing Speech to 3 Major Markets



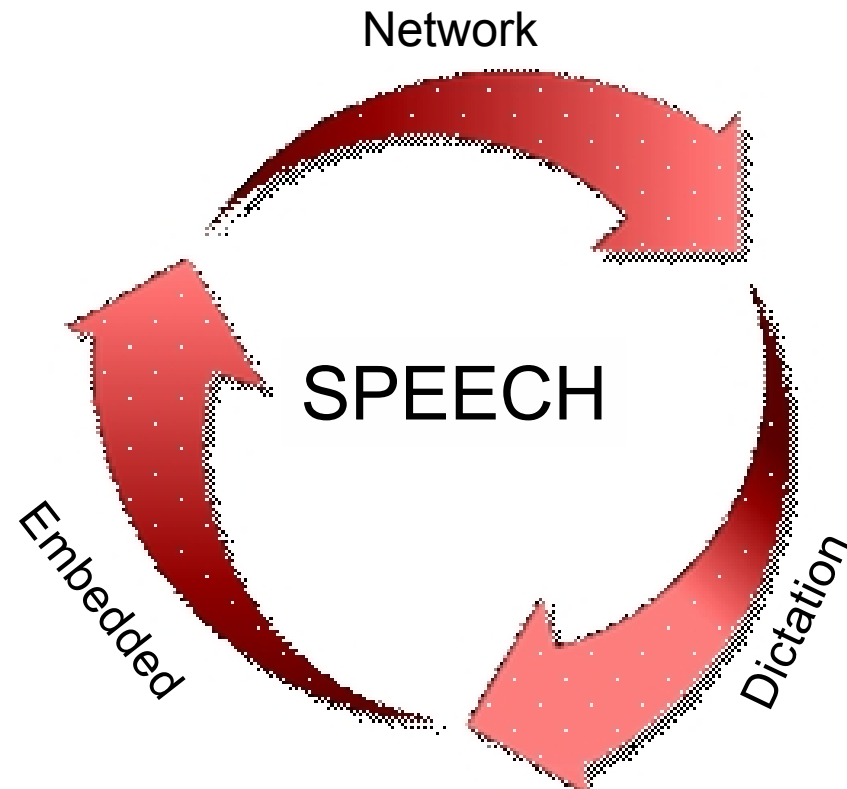


---

# Converging Technologies

# Range of Speech Technologies

---



# State-of-the-Art of ASR Technology

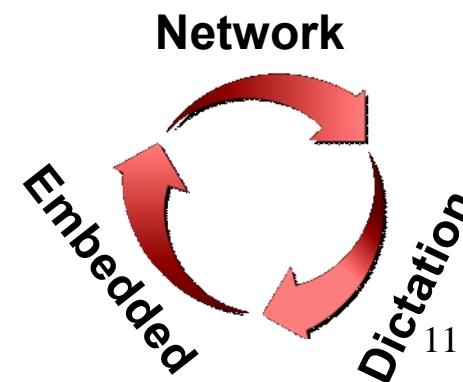
---

- ♦ Network ASR
  - ♦ Merged capabilities of OpenSpeech (OSR) and SpeechPearl
    - Platform and market coverage of OSR
    - Language coverage of SpeechPearl
- ♦ Embedded ASR
  - ♦ Vocon 3200 for high-end capabilities
  - ♦ Vocon SF for DSP-based implementations
- ♦ Dictation
  - ♦ Dragon tailored for vertical applications

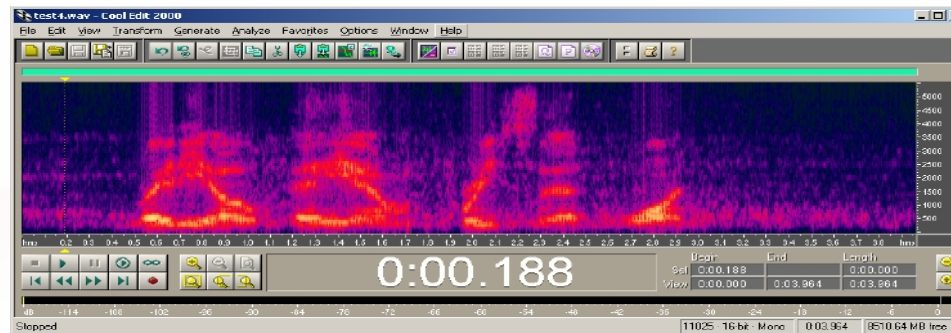
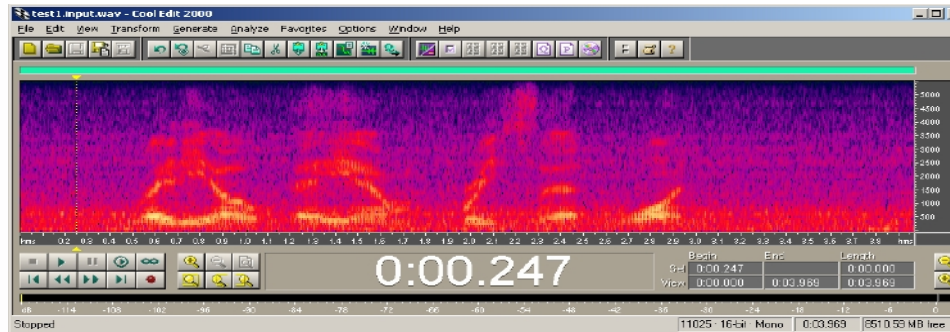
# Leveraging ASR Technologies

---

- ♦ Network → Embedded
  - ♦ High-end, large vocabulary, grammar-based ASR
  - ♦ Dialog tools
- ♦ Embedded → Network
  - ♦ Efficient resource use
  - ♦ Noise robustness
- ♦ Dictation → Network, Embedded
  - ♦ Open, dictation grammars
  - ♦ High performance, acoustic modeling and adaptation
- ♦ Network, Embedded → Dictation
  - ♦ Noise robustness
  - ♦ Processing efficiency







# Cross-Fertilization in R&D



**Work to improve functionality in Automotive applications benefits core ASR engine**

# State-of-the Art in Text-To-Speech

---

- ♦ Becoming Increasingly Important
  - ♦ Access to wide range of context (news, email,...)
  - ♦ Large applications (name & address capture, directory assistance, etc.)
- ♦ Technology advancements
  - ♦ 20 years ago: making it understandable 
  - ♦ Today: making it natural 
  - ♦ Next: personality  

# What's Coming In Text-to-Speech?

---

- ♦ Continued quality improvements
- ♦ Range of Products for Network, Desktop, and Embedded Applications
  - ♦ Call center automation
  - ♦ DA
  - ♦ Email / Unified Messaging
  - ♦ SMS-to-voice
  - ♦ Automotive Applications
  - ♦ Assistive technologies
- ♦ Application-specific quality improvements
  - ♦ Scansoft-supplied Blades
  - ♦ Tools for tuning
  - ♦ Mix of TTS and human recordings

# Authentication methodologies

---

T345n#iq3P,  
4590, Jones

- ♦ **Something you Know ... but can forget!**

Password, PIN, “mother’s maiden name”....



- ♦ **Something you Have ... but can be misplaced or stolen!**

Physical Key, Token, Mag Card, ....



- ♦ **Something you Are ... forever & always!**

**Voice**, Finger print, Retina, ....



---

# Converging Markets

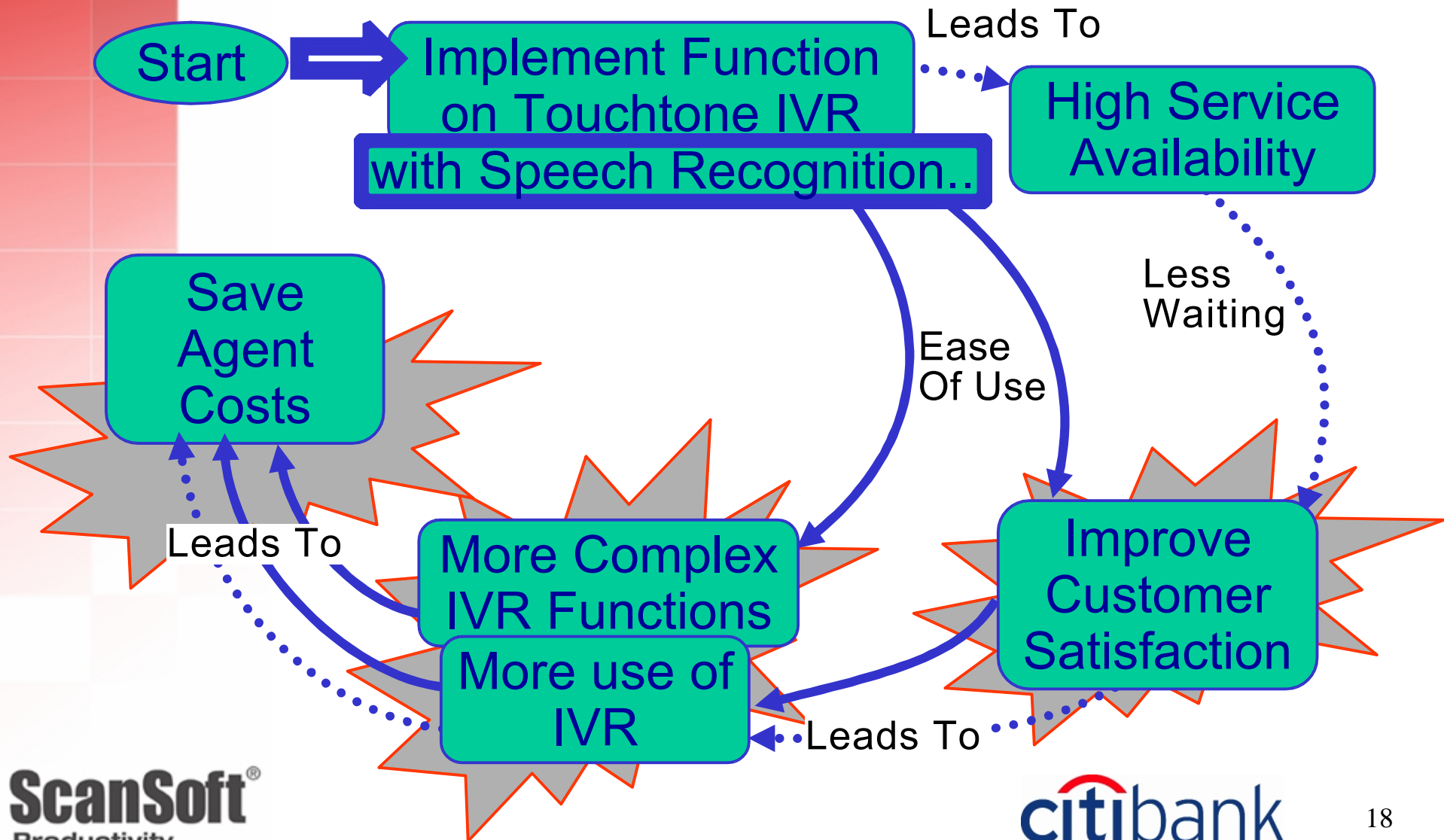
# The Evolution of Self-Service

Degree of Power  
and Control



Ease of Access and Use

# Economic View: A Virtuous Circle



# Current Dialog Technologies

---

- ♦ Mainly telephone-based
- ♦ What's good about this:
  - ♦ Ubiquitous devices
  - ♦ Speech is natural interface
- ♦ What's bad about this:
  - ♦ Speech-only interfaces challenging
  - ♦ Telephony infrastructure not ideal for speech recognition

# Next Generation Mobile Devices

- ♦ Capabilities
  - ♦ Available local computation
  - ♦ High quality displays
  - ♦ Wireless networking
- ♦ Multi-Modal
  - ♦ Speech + Display + Stylus
- ♦ Distributed Speech Processing
  - ♦ Local speech for local functions
  - ♦ Distributed speech for network applications



## Not just for high-end cars

---



Prius multi-information display panel

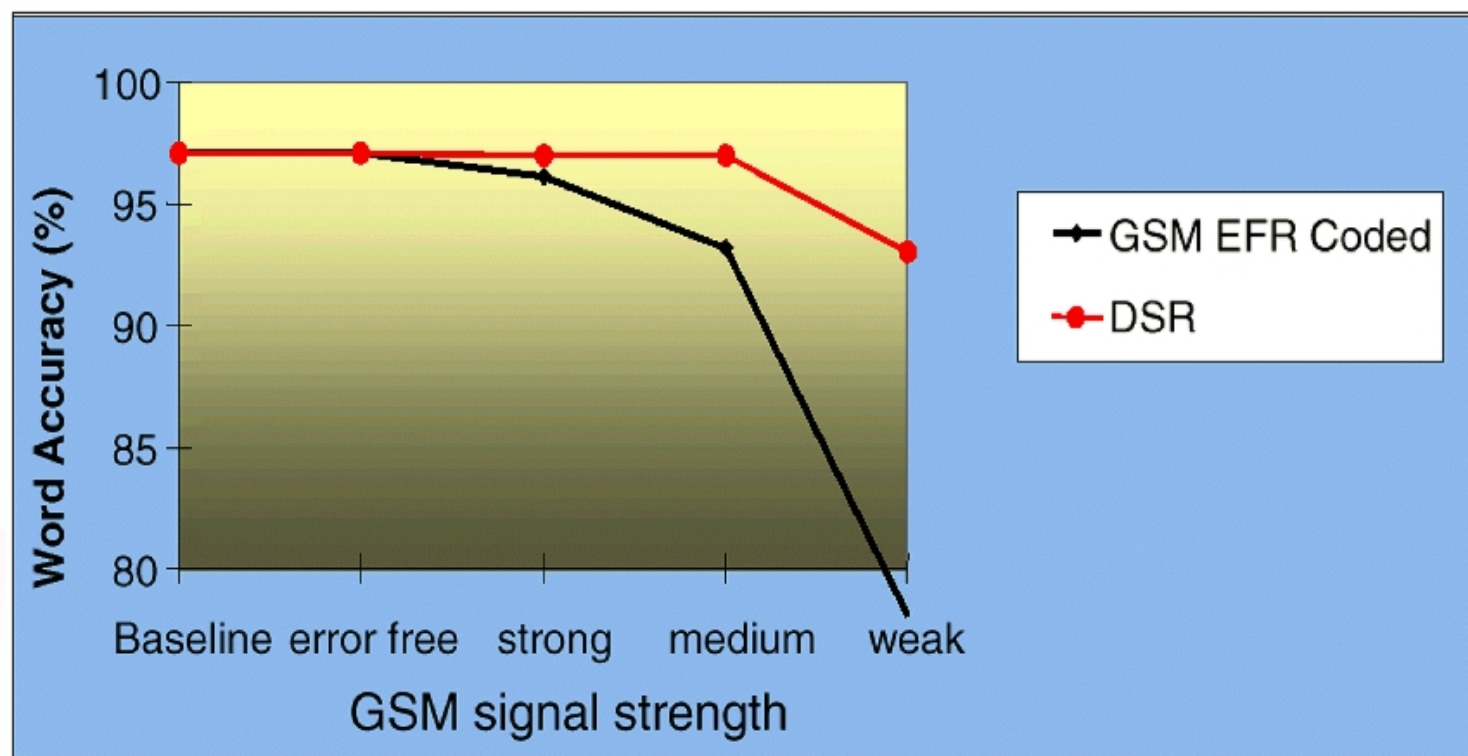
\_prius

# Distributed Speech Recognition

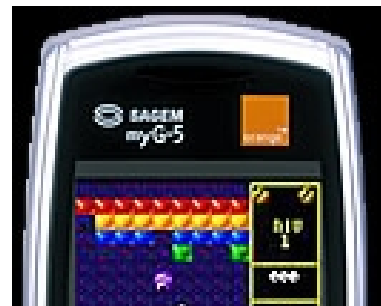
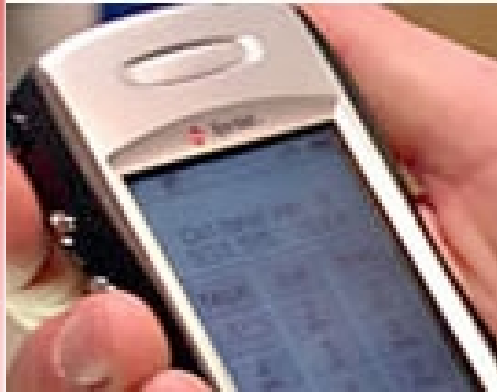
- ◆ Improved recognition performance
  - ◆ Local noise cancellation
  - ◆ Higher speech bandwidth
- ◆ Performance gains
  - ◆ Lower bitrate
  - ◆ Lower latency
- ◆ Share front-end with embedded ASR
- ◆ Consistency in UI of Embedded ASR and Network ASR



# DSR Performance With Network Errors



# Mobile Phones + PDA's



**ScanSoft**  
Productivity  
Without Boundaries™

# Speech as Interface for Mobile Phones

---

- ♦ Phone-based applications
  - ♦ Voice Dialing
  - ♦ PIM, PDA, SMS, Email
  - ♦ Controlling phone
- ♦ Used as Interface to network-based services
  - ♦ Plain old telephone
  - ♦ Distributed Speech Recognition
  - ♦ Multimodal



## Where We're Heading ...

---

- ♦ More and more demand for mobile access to information and transactions
- ♦ Wide range of devices, applications, situations
- ♦ Speech input and output is an important mode of interaction
- ♦ Integrated set of speech and dialog technologies, products and services

# Converging Markets and Technologies

---

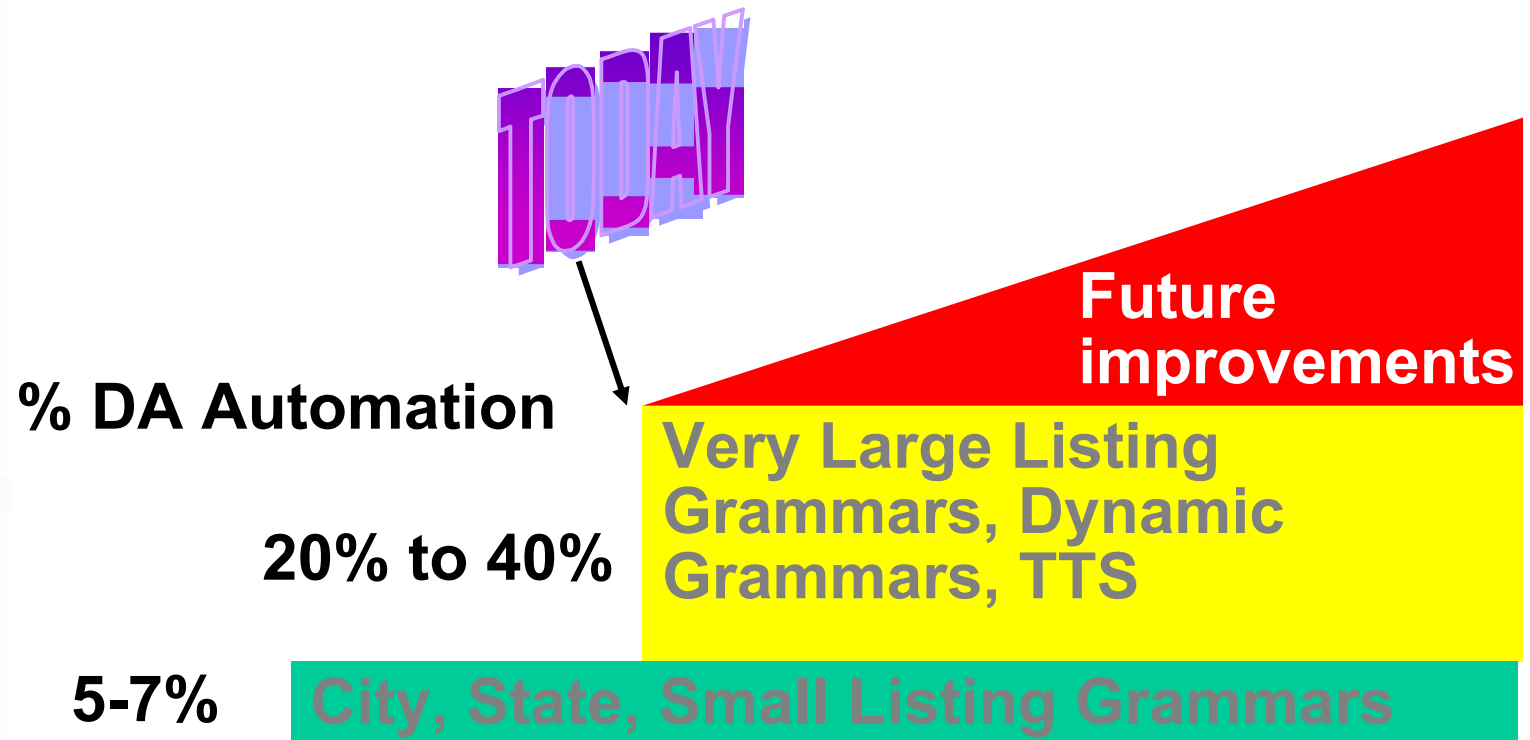
- ♦ Mobile device/Network convergence
  - More advanced devices + data networks
  - = MultiModal + Distributed Speech Processing
- ♦ Dictation/Open grammars for network
  - ♦ Transcripts of voicemails
  - ♦ SMS dictation
- ♦ Dictation for embedded
  - ♦ Processing available soon
  - ♦ Dictation of SMS, Voicemail, etc.



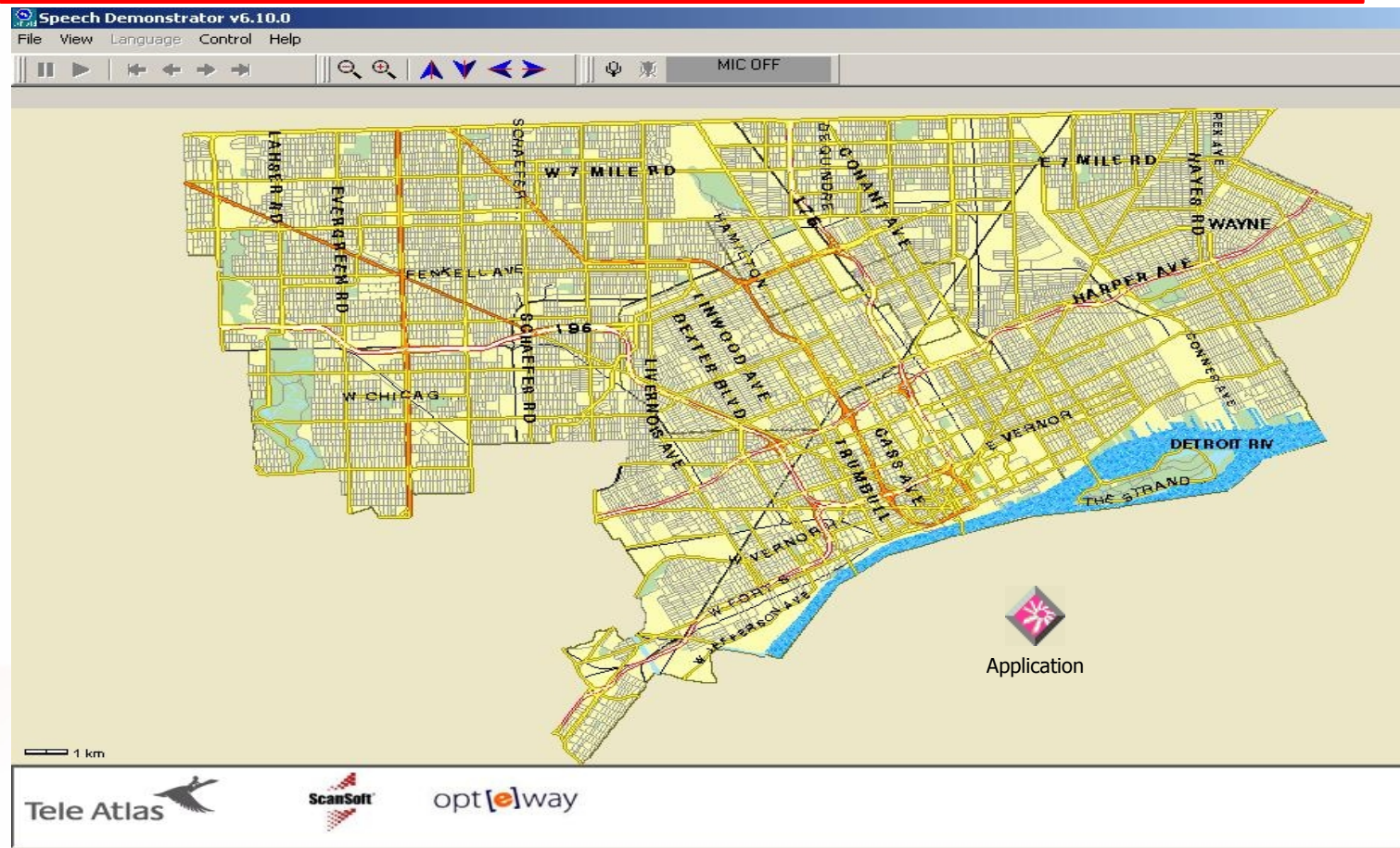
## Future Applications ... Future Solutions

# Application Focus: Directory Assistance

Improvements in ASR and TTS are driving the adoption of speech in DA



# SpeechPak Voice Destination Entry (VDE)



- SpeechPak VDE on SH4 / QNX in development
- Large Vocabulary Recognition of 10K words or greater
- Spelling Fallback

# Engineering Focus

---

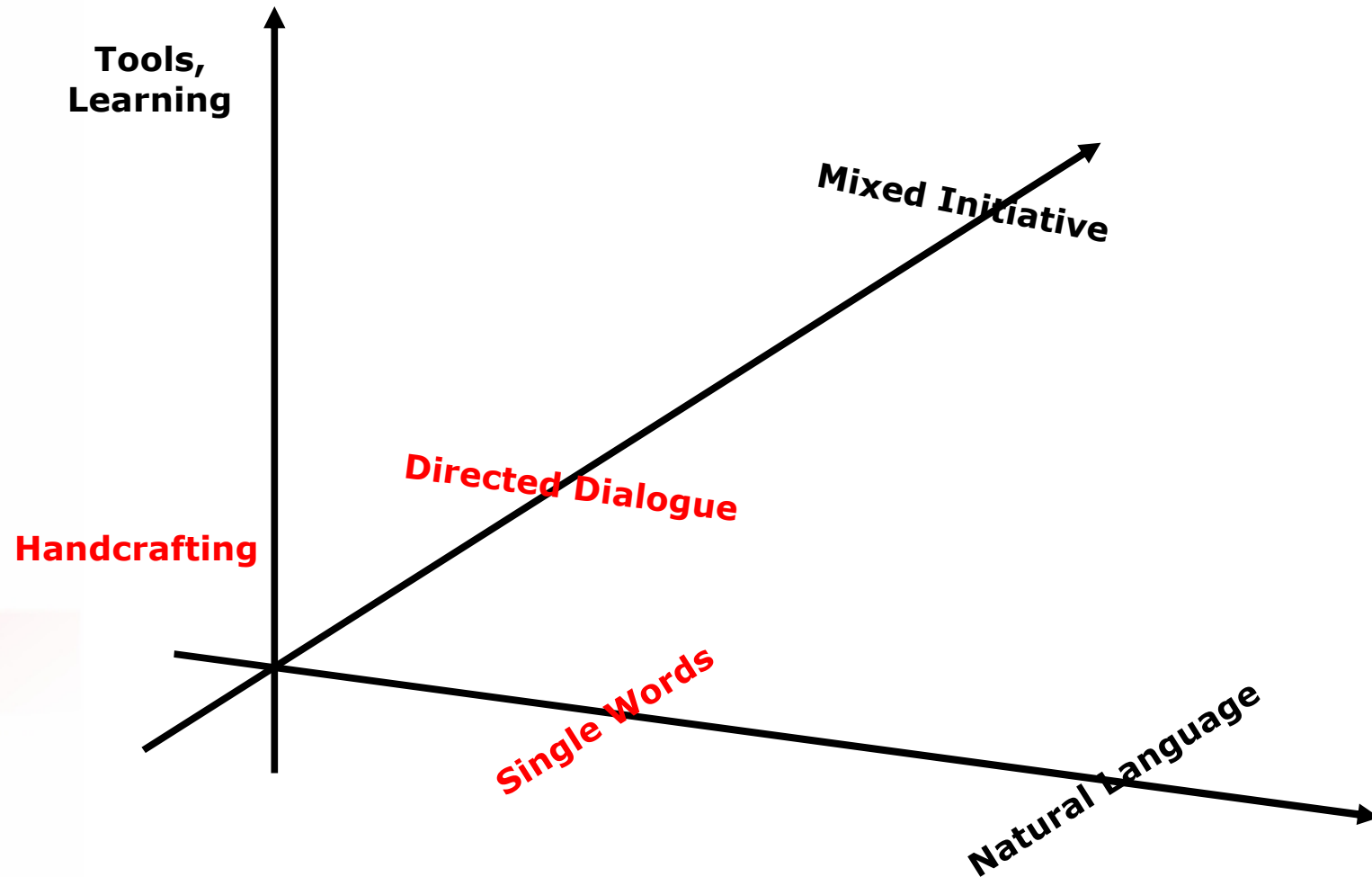
- ♦ Constant Improvement of Technologies
- ♦ Better Tools
- ♦ Process/Design

# System vs Human Performance

---

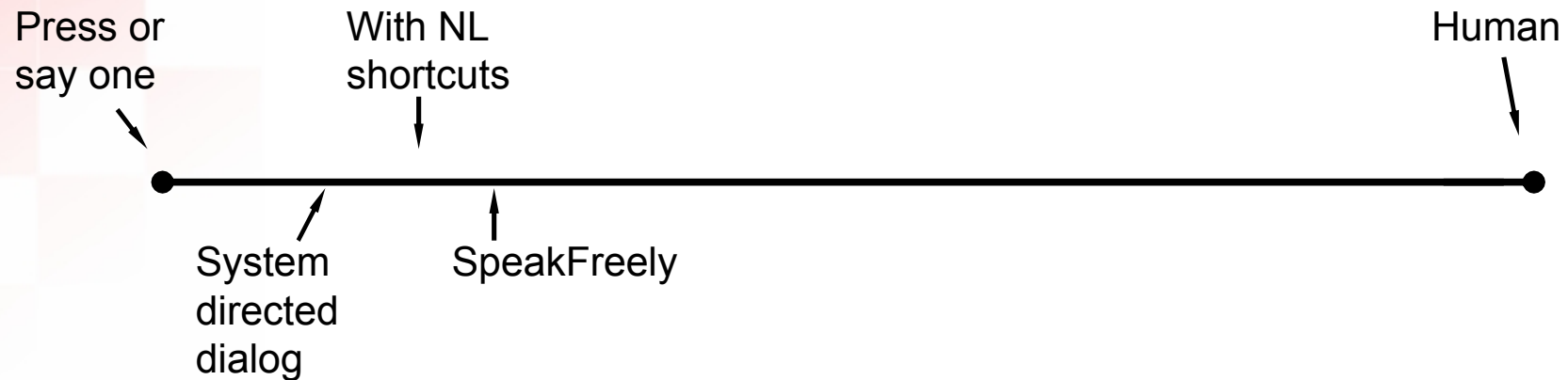
- ♦ Handling out of domain
- ♦ Turn taking
- ♦ Error correction strategies
- ♦ Dynamic Prompting
- ♦ Knowing when dialogs are going wrong
- ♦ Use of extra-linguistic information

# Improvements Along 3 Dimensions

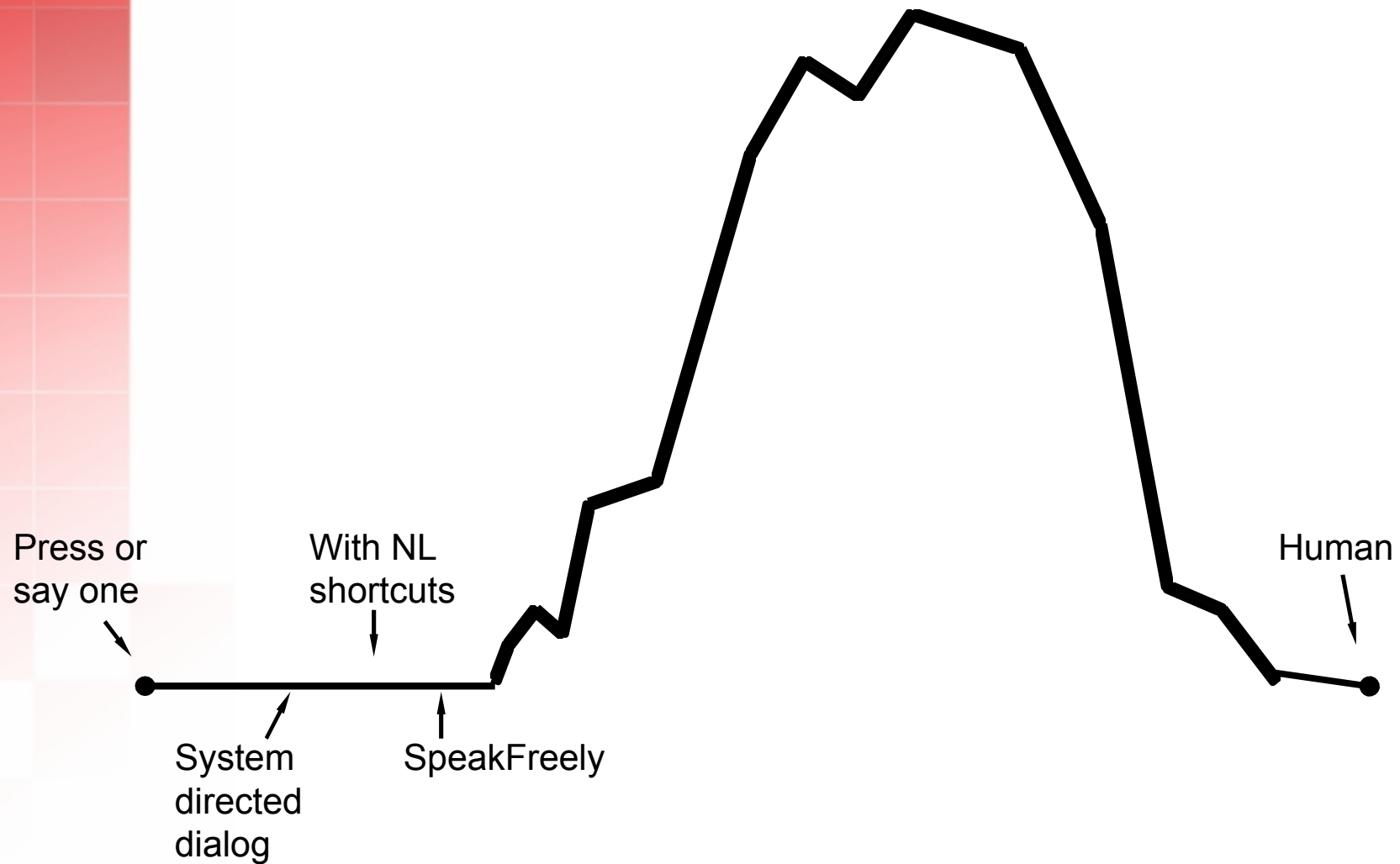


# Range of Dialog Complexity

---



# Range of Dialog Complexity



# User Expectations

---

- ♦ How can the user know how to interact with the system?
- ♦ To go beyond system directed:
  - ♦ Users need to develop expectation of interaction
  - ♦ Repeat usage of individual systems
  - ♦ Common aspects of interaction between systems

# “Branding ... Over the Phone”



**Boomer  
Outdoorsman**



**Gen-X/Y**



**Generic Prof'l.**



**Branded**

# Summary: Evolution of Speech

---

- ♦ Near-term
  - ♦ Integrated products
  - ♦ Applications/tools
  - ♦ Making use of technology synergies
- ♦ Market evolution
  - ♦ Increased device capabilities
  - ♦ Telephone -> distributed speech, multi-modal
- ♦ Speech and language as ubiquitous interface
  - ♦ Users will be encountering more and more language-based interfaces
  - ♦ Evolution of model of human machine dialog



---

# Thank You!