

IBM

Distributed Speech Recognition - a Doorway to New Speech-Enabled Services

Zohar Sivan

Manager, Media Services and Technologies

IBM Research Lab in Haifa



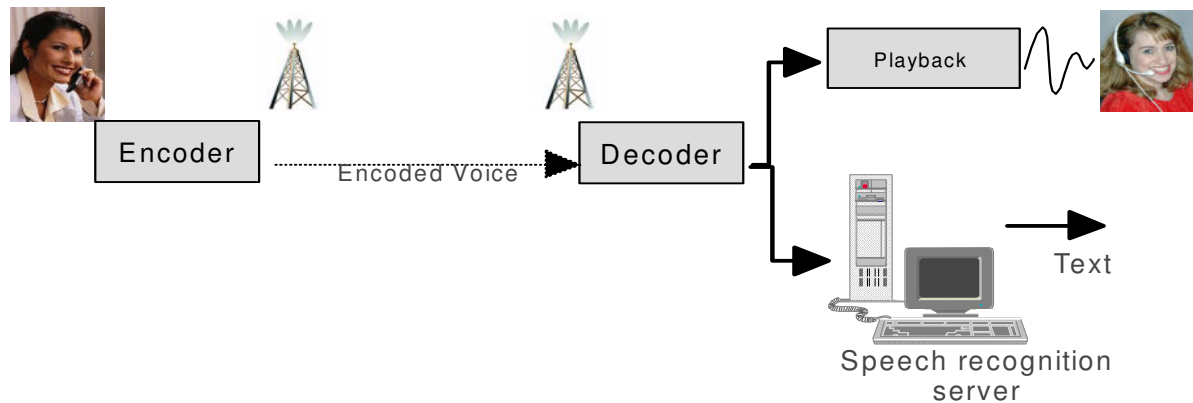
Background

- ◇ What are DSR and DSR-optimized Codecs ?
 - ◇ Speech recognition Front-End and Back-End are carried out at separate locations.
 - ◇ DSR Codecs are optimized for best speech recognition accuracy, by encoding and sending speech recognition features
- ◇ What type of services does DSR address ?
 - ◇ DSR technology is most suitable for the following scenario:
 - ◇ Devices with limited resources - Local ASR cannot be used
 - ◇ Advanced speech or speaker recognition applications/services, where accuracy is crucial
 - ◇ Noisy communication channel and acoustic environment
 - ◇ Multimodal applications
 - ◇ Very common scenario for wireless communication from a mobile phone or PDA

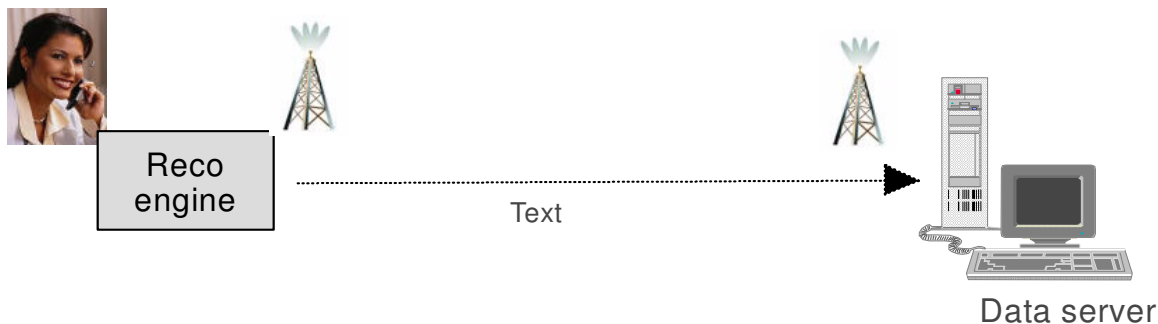


Distributed Speech Recognition (DSR)

Existing schemes and their drawbacks:



Server ASR:
Low bit-rate speech coding and channel errors degrade recognition accuracy



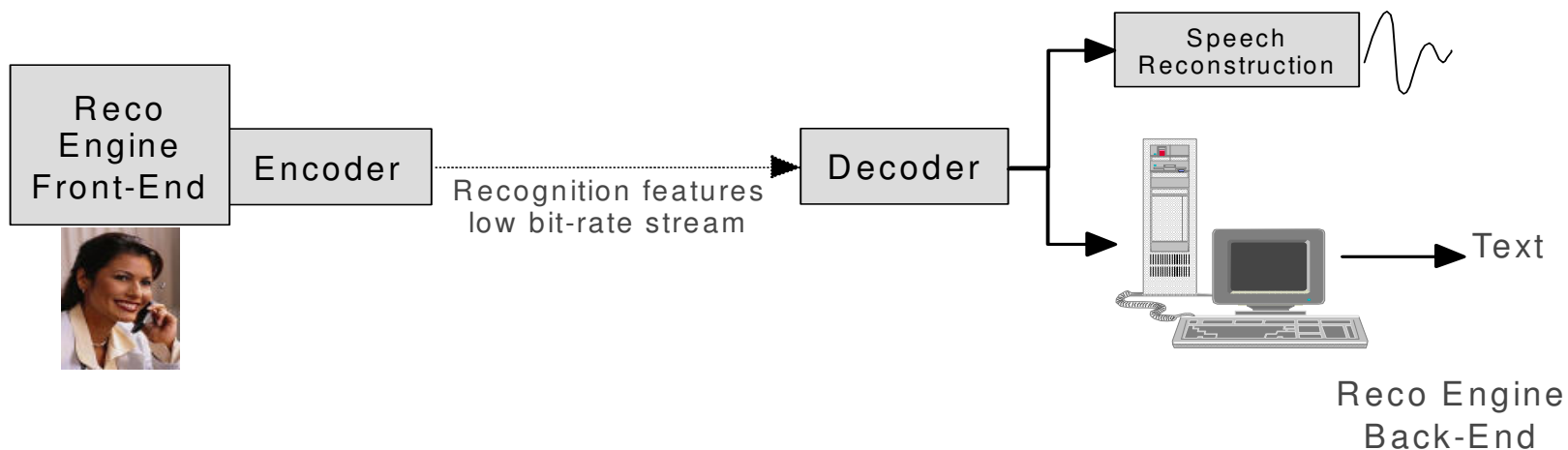
Embedded ASR:

- Requires a "fat" client.
- Limited vocabulary size
- Accuracy issues
- No server playback
- Complex upgrade
- Small channel utilization



The DSR Solution

- ◆ Recognition features extraction (Front-End) on the client device
- ◆ Compressed features are sent over error protected data channel (4.8-5.8 kbps)
- ◆ Features decompression & error mitigation at server, lost packets are recovered by features interpolation
- ◆ Playback at the server side is enabled by a new technique for speech reconstruction from recognition features





Benefits of DSR for Mobile Speech-enabled Applications

- ◆ Higher recognition accuracy
 - ◆ Server-based recognition can use more memory/computation resources
 - ◆ Minimal impact of speech codec & channel errors
- ◆ Small footprint on device
 - ◆ Feature extraction and encoding require small resources, independent of the recognition task complexity
 - ◆ Sophisticated applications (NLU, dictation) can be run on low-end embedded devices
- ◆ Business value
 - ◆ Enables Multimodal applications by utilizing a single channel for both voice and data
 - ◆ Ability to provide advanced speech-enabled services to a multitude of low-end devices in various environments
 - ◆ Bridges mismatch of applications and device lifecycle

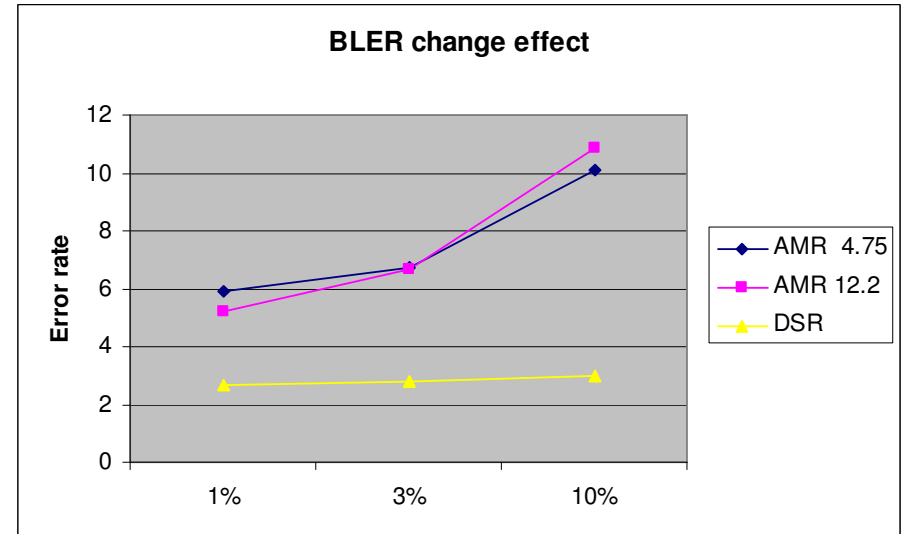
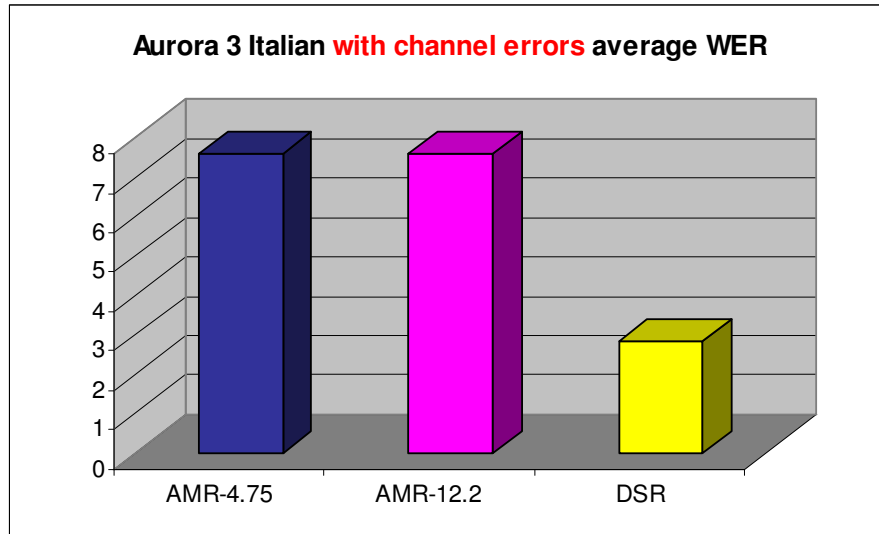


DSR Standardization

- ◆ ETSI/Aurora
 - ◆ Participants: Motorola, Alcatel, FT, BT, Siemens, Nokia, Ericsson, Qualcomm, IBM, Scansoft, Nuance, OGI, HP, Intel
 - ◆ Standardization of 4 DSR optimized Codecs (front-ends):
 - ◆ Basic front-end (2000), Advanced front-end (2002)
 - ◆ **Extended basic/advanced front-ends (November 2003) by IBM and Motorola**
- ◆ 3GPP
 - ◆ Main players: Motorola, Nokia, Ericsson, Alcatel, Siemens
 - ◆ SA4: Codec Work to Support Speech Recognition Framework for Automated Voice Services:
 - ◆ **February 2004 - DSR to be recommended !**
 - ◆ **Average improvements in speech recognition accuracy of over 30% compared to GSM-AMR**

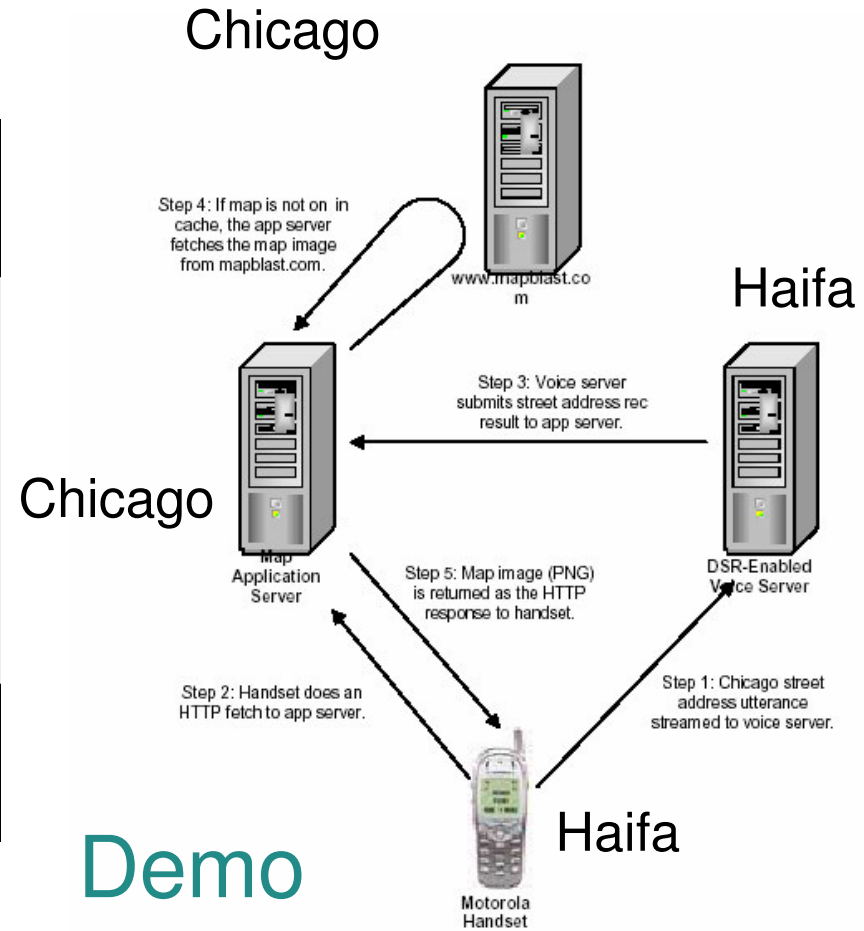
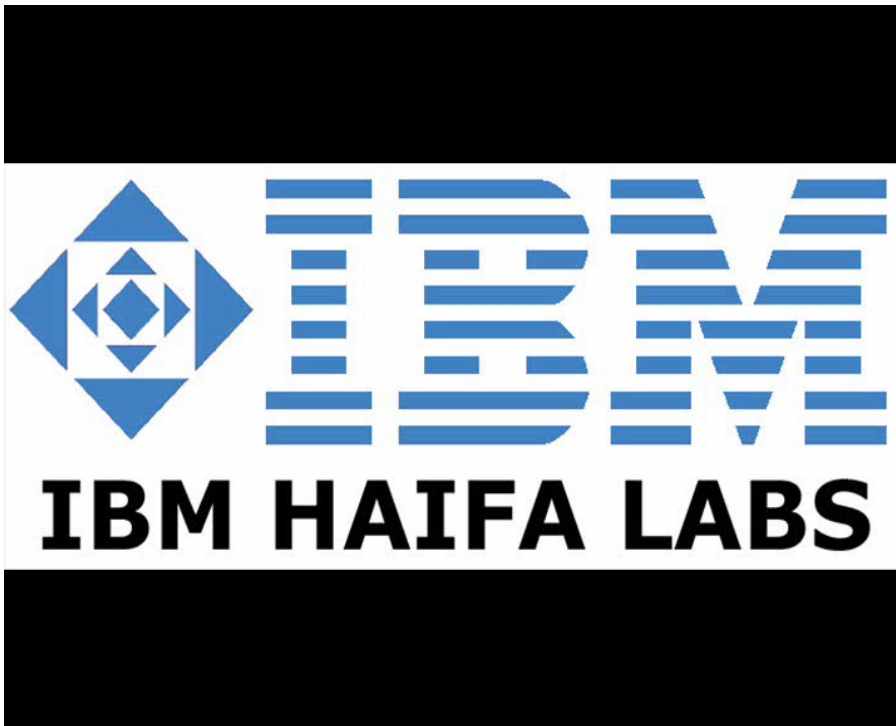


3GPP DSR Evaluation - IBM Results

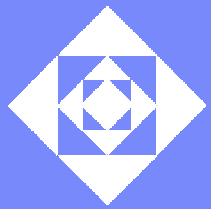




DSR Demonstration



Demo



Use Cases for DSR technology



Mobile Workforce - Travel Management

- ◆ Application
 - ◆ Mobile Travel Expense
- ◆ Application Provider: SAP
- ◆ Configuration
 - ◆ GSM / iDEN / CDMA Mobile Phone (optional camera)
 - ◆ X+V Multimodal Phone Browser
 - ◆ 2.5G Network with DSR
- ◆ Sample Interaction
 - ◆ Employee traveling on business launches travel expense applet on mobile phone, and enters amounts & dates for expenses (meals, taxi, etc.) on-the-spot, having them immediately recorded in back-end travel expense database.
- ◆ Value Add
 - ◆ Allows user to enter travel and expense information quickly while on the road
 - ◆ Allows entering data either verbally or by text, with both visual and aural confirmations
 - ◆ Camera-phone can enable immediate photo recording for receipts, etc.





Mobile Consumer

- ◇ Applications
 - ◇ Yellow Pages, movie schedules, weather, news, financial, etc.
- ◇ Application Provider: Wireless Carriers
- ◇ Configuration
 - ◇ GSM / iDEN / CDMA mobile phone
 - ◇ X+V-enabled phone browser
 - ◇ Server-based X+V
 - ◇ 2.5G Network with DSR
- ◇ Sample Interaction
 - ◇ The user presses an action button to load the appropriate application and then utters:
 - ◇ "What action movies are playing near me at 8 o'clock tonight?" The non-multimodal query would involve navigating at least 4 WAP submenus: application -> category -> location -> temporal
 - ◇ "What's the forecast for New York City tomorrow?"
 - ◇ "What did IBM close at today?"
- ◇ Value Add
 - ◇ Reaches large numbers of consumer clients (600m+)
 - ◇ Allows easy information access for consumers who typically ignore data functionality due to interaction issues, high value our customer's customer
 - ◇ Centralized management
 - ◇ Drives additional data usage when majority only use voice services
 - ◇ New capabilities open to application providers





Public Sector - Emergency Response

- ◆ Application
 - ◆ Mobile Emergency Response
- ◆ Application Provider: SAP
- ◆ Configuration
 - ◆ iDEN Mobile Phone (optional GPS)
 - ◆ X+V Multimodal Phone Browser
 - ◆ 2.5G Network with DSR
- ◆ Sample Interaction
 - ◆ Emergency responder arrives at the scene of an accident, uses the mobile emergency response application to request personnel & equipment, and goes through a checklist of tasks to be performed for the particular situation.
- ◆ Value Add
 - ◆ Allows responder to do voice-driven resource management tasks using a small mobile device in a hands-busy, eyes-busy emergency situation.
 - ◆ Visual output allows maps & detailed resource information to be displayed, complementing audio output.
 - ◆ GPS phone capability can allow resource tracking & positioning





Mobile Workforce - Sales Force Automation

- ◆ Application
 - ◆ Mobile Sales Management
- ◆ Application Provider: SAP
- ◆ Configuration
 - ◆ GSM / iDEN / CDMA Mobile Phone
 - ◆ X+V Multimodal Phone Browser
 - ◆ 2.5G Network with DSR
- ◆ Sample Interaction
 - ◆ A sales manager uses voice queries to easily track and approve orders, authorize quotes, and check inventories of items. Relevant sales information is displayed on the phone browser with aural confirmations.
- ◆ Value Add
 - ◆ Allows sales professional to quickly make queries and transactions using just a mobile phone, without need for desktop or laptop access.
 - ◆ Allows voice-driven transactions in hands-busy, eyes-busy situations (e.g. inventory management), with visual feedback when required for detailed transaction information.





Mobile Workforce - Contact Management

- ◆ Application
 - ◆ Mobile Directory Lookup / Bluepages
- ◆ Application Provider: (IBM Internal)
- ◆ Configuration
 - ◆ GSM / iDEN / CDMA Mobile Phone
 - ◆ X+V Multimodal Phone Browser
 - ◆ 2.5G Network with DSR
- ◆ Sample Interaction
 - ◆ A mobile employee enters the name of a co-worker or customer by voice, and receives detailed contact information (including photo) from directory back-end database. A phone call can then be immediately placed to the contact using either a voice command or button on the phone.
- ◆ Value Add
 - ◆ Allows mobile employee to quickly retrieve contact information by voice when on the road, using only a mobile phone.
 - ◆ Allows photos and other detailed contact information to be displayed on the phone, complementing audio output.

