Perspectives on the use of Distributed Speech Recognition for in-car Telematics

David Pearce
Motorola

"The Fully Networked Car, A Workshop on ICT in Vehicles"
ITU-T Geneva, 2-4 March 2005
Embedded vs Connected

Wireless Network

The Fully Networked Car, A Workshop on ICT in Vehicles
ITU-T Geneva, 2-4 March 2005
Key Elements of a Multimodal Architecture

Voice Gateway(s)

Distributed Speech Recognition

VoiceXML

Content Services

Voice Browser Control Protocol

Wireless packet data

Vehicle & Devices

Java and/or xHTML Browser

Wireless packet data
The Fully Networked Car, A Workshop on ICT in Vehicles
ITU-T Geneva, 2-4 March 2005

Distributed Speech Recognition

Client Devices

[Wireless] Packet Data Network

IP Network

Content Servers

Voice Gateway / Server:
- VoiceXML / SALT / X+V Browser
- Speech Resources (ASR, TTS, etc.)

Conventional

Speech Coder 

Circuit Switched Mobile Voice Channel

Speech Decoder

ISDN

ASR Front-end

ASR Decoder

DSR

ASR Front-end

Packet Data Channel e.g. GPRS or CDMA 1x

ASR Decoder
Benefits of DSR

- Improves performance over wireless channels
  - Minimises impact of codec & channel errors
  - Consistent performance over coverage area
- Improved performance in background noise
  - 53% reduction in error rate
- Ease of integration of combined speech and data applications
  - Use packet data channel for both DSR and other data
DSR Standards

DSR Advanced front-end (Oct 02)
DSR Extended Advanced Front-end (Nov 03)

Speech Enabled Services
Fixed point DSR standard created
DSR selected as the recommended codec for SES
(Approved June 04)

IETF
RTP payload formats for DSR
Specifications standardised at rfc

3GPP2
Speech Enabled Services
New Work Item (Approved Jan 05)
## Results of ASR vendor evaluations in 3GPP

<table>
<thead>
<tr>
<th></th>
<th>Number of db tested</th>
<th>AMR4.75 Average Absolute Performance</th>
<th>DSR Average Absolute Performance</th>
<th>Average Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digits</strong></td>
<td>11</td>
<td>13.2</td>
<td>7.7</td>
<td>39.9%</td>
</tr>
<tr>
<td><strong>Sub-word</strong></td>
<td>5</td>
<td>9.1</td>
<td>6.5</td>
<td>30.0%</td>
</tr>
<tr>
<td><strong>Tone confusability</strong></td>
<td>1</td>
<td>3.6</td>
<td>3.1</td>
<td>14.8%</td>
</tr>
<tr>
<td><strong>Channel errors</strong></td>
<td>4</td>
<td>6.1</td>
<td>2.4</td>
<td>52.8%</td>
</tr>
<tr>
<td><strong>Weighted Average</strong></td>
<td></td>
<td></td>
<td></td>
<td>36%</td>
</tr>
</tbody>
</table>

- Extensive testing on 21 different speech databases
  - Covering different languages, tasks and environments
- Tests performed with IBM and Scansoft commercial recognisers
- Results for low data-rate comparison for packet data (< 8kbit/s)
Packet Switched Channel Errors

- Aurora-3 Italian speech database
- GPRS network simulation for distribution of errors

3GPP Feb 2004
## Coded speech vs DSR

<table>
<thead>
<tr>
<th></th>
<th>DSR</th>
<th>AMR 4.75</th>
<th>Degradation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well matched</td>
<td>96.5</td>
<td>94.4</td>
<td>-57%</td>
</tr>
<tr>
<td>Med mismatch</td>
<td>90.4</td>
<td>83.9</td>
<td>-68%</td>
</tr>
<tr>
<td>High mismatch</td>
<td>88.6</td>
<td>76.8</td>
<td>-104%</td>
</tr>
<tr>
<td>Average</td>
<td>92.4</td>
<td>86.3</td>
<td>-73%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>DSR</th>
<th>EVRC</th>
<th>Degradation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well matched</td>
<td>96.5</td>
<td>90.6</td>
<td>-165%</td>
</tr>
<tr>
<td>Med mismatch</td>
<td>90.4</td>
<td>75.9</td>
<td>-151%</td>
</tr>
<tr>
<td>High mismatch</td>
<td>88.6</td>
<td>70.5</td>
<td>-160%</td>
</tr>
<tr>
<td>Average</td>
<td>92.4</td>
<td>80.4</td>
<td>-159%</td>
</tr>
</tbody>
</table>
Conclusions

- DSR provides substantial performance advantages for Voice driven Telematics services
- Standards for the DSR features and transport protocols for interoperability are complete
- Is there a need to incorporate in specific telematics standards?