Distributed Speech Recognition

David Pearce
Motorola Labs
bdp003@motorola.com

Voice & Multimodal

timodal-enabled Services

User enters commands via:

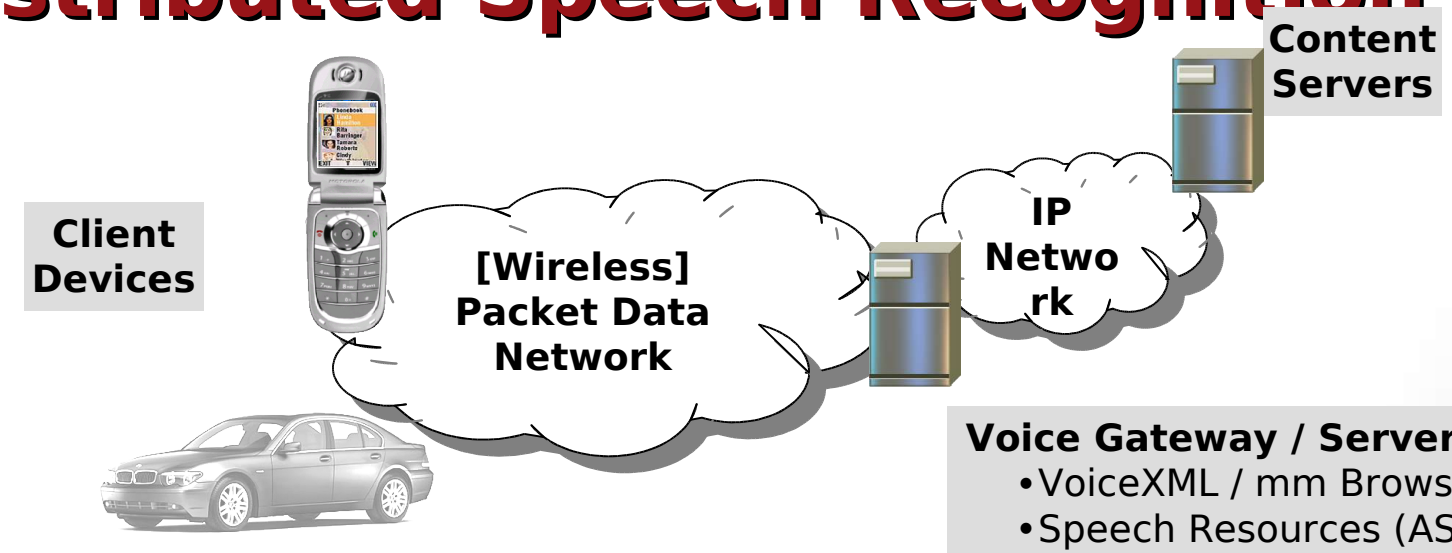
SPEECH | KEYPAD

System

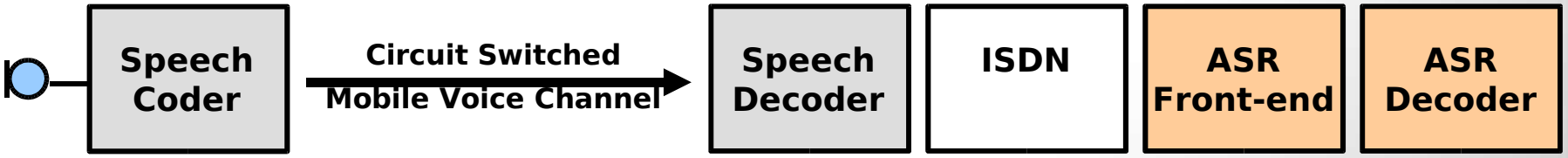
TEXT | GRAPHICS

SPEECH | SOUNDS

Distributed Speech Recognition

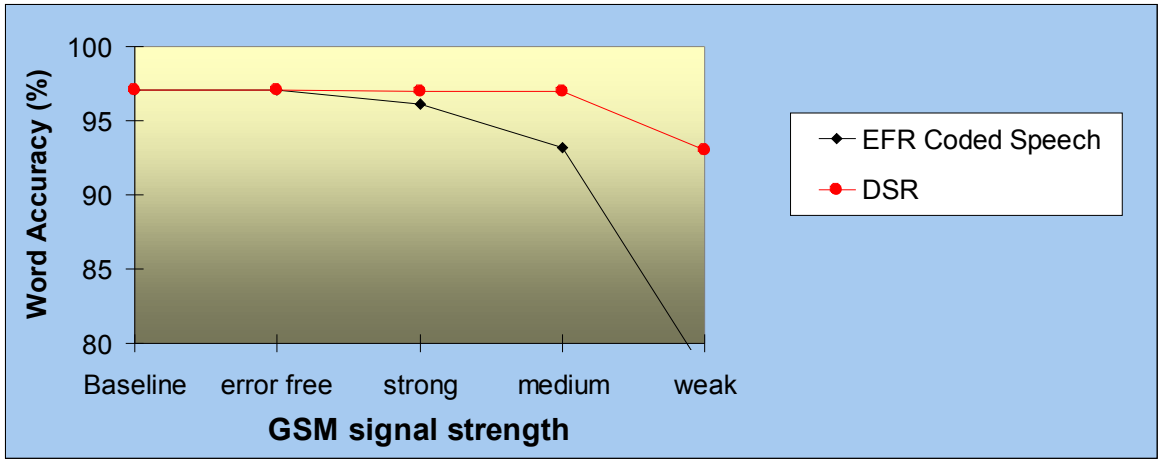


Conventional



DSR

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

ETSI

DSR Advanced front-end (Oct 2002)
DSR Extended Advanced Front-end (Nov 2003)

3GPP
A GLOBAL INITIATIVE

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

IETF

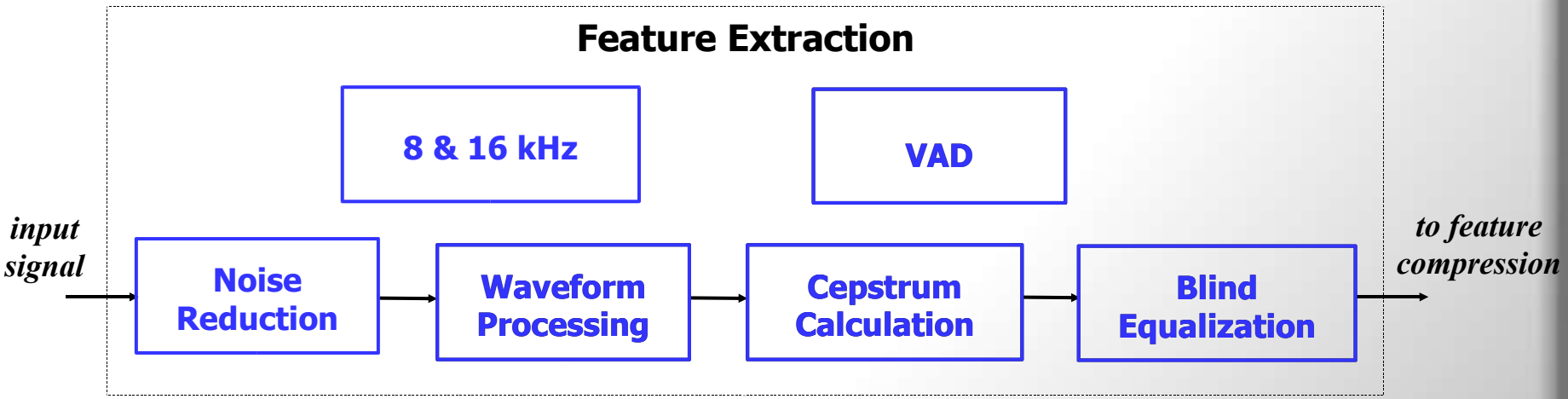
RTP payload formats for DSR
Specifications standardised rfc4060

3GPP2

Speech Enabled Services
New Work Item (Approved Jan 2005)

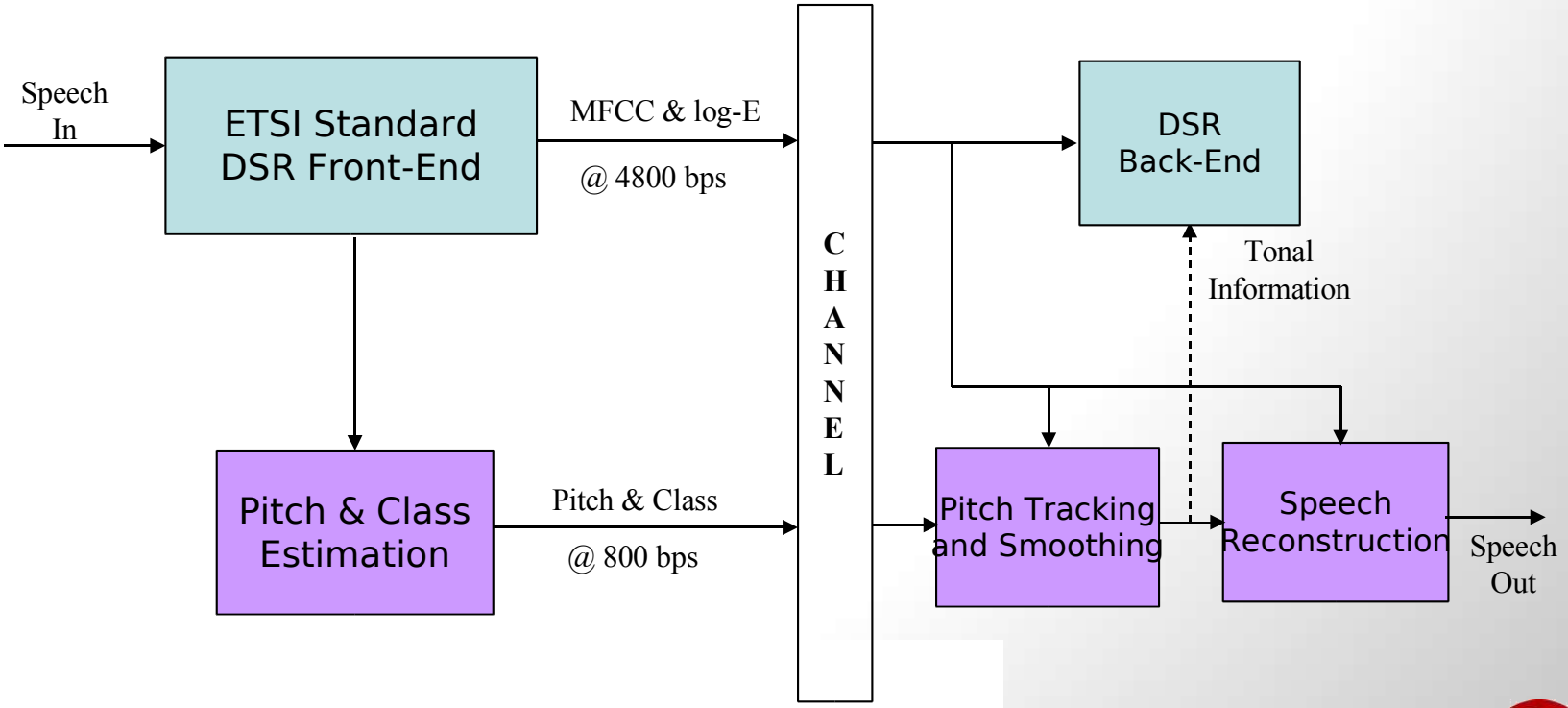
DSR Advanced Front-end (ES 202 050)

- Noise Robust Front-end
- Half error rate cf mel-cepstrum in background noise
  - Double Wiener filtering noise suppression
  - Waveform processing
  - Blind equalisation
- Representation: 12 cepstral coeffs, C0, logE
- Compression gives bit rate of 4.8kbit/s



DSR Extension (ES 202 212)

- Enables Speech waveform reconstruction at server for human listening
  - Adds 800bps containing pitch (**total 5.6kbps**):
  - Assists recogniser with tonal language recognition (e.g. Mandarin, Cantonese)

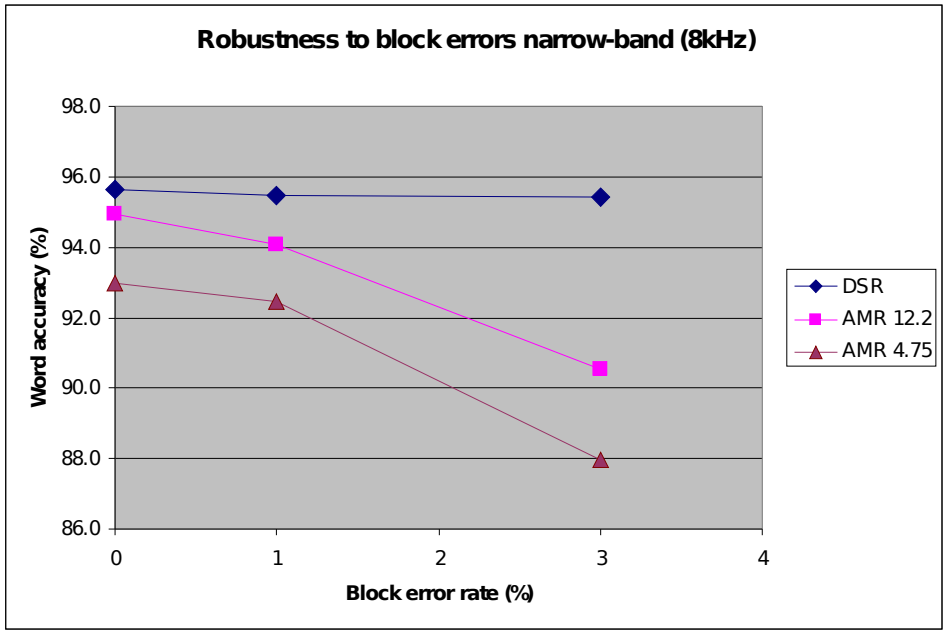


Results of ASR vendor evaluations in 3GPP

8 kHz	Number of db tested	AMR4.75 Average Absolute Performance	DSR Average Absolute Performance	Average Improvement
Digits	11	13.2	7.7	**39.9%**
Sub-word	5	9.1	6.5	**30.0%**
Tone confusability	1	3.6	3.1	**14.8%**
Channel errors	4	6.1	2.4	**52.8%**
Weighted Average				**36%**

- **Extensive testing on 21 different speech databases**
  - Covering different languages, tasks and environments
- **Tests performed with IBM and Scansoft commercial recognisers**
- **Results above are 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

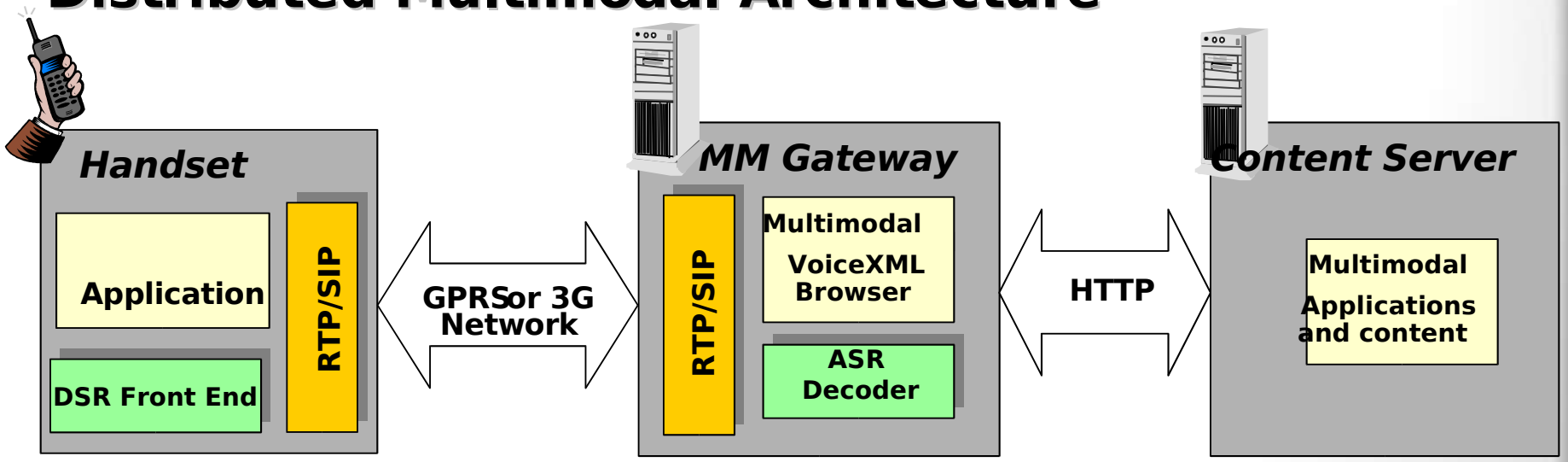
3GPP Feb 2004

Coded speech vs DSR (Aurora-3 Italian)

	DSR	AMR 4.75	Degradation
Well matched	**96.5**	**94.4**	**-57%**
Med mismatch	**90.4**	**83.9**	**-68%**
High mismatch	**88.6**	**76.8**	**-104%**
Average	92.4	86.3	-73%

	DSR	EVRC	Degradation
Well matched	**96.5**	**90.6**	**-165%**
Med mismatch	**90.4**	**75.9**	**-151%**
High mismatch	**88.6**	**70.5**	**-160%**
Average	92.4	80.4	-159%

Distributed Multimodal Architecture



Handset device

- Input modalities (i.e., DSR, keypad input, pen entry)
- Output media (e.g., Visual rendering, Decoded speech output)
- Application Environment (Java or WAP Browser)
- Protocols (SIP / RTP, Multimodal remote control)

Multimodal Gateway

- DSR Decoder
- Multimodal VoiceXML browser
- Protocols

Applications and content

- Content authoring
- Content delivery