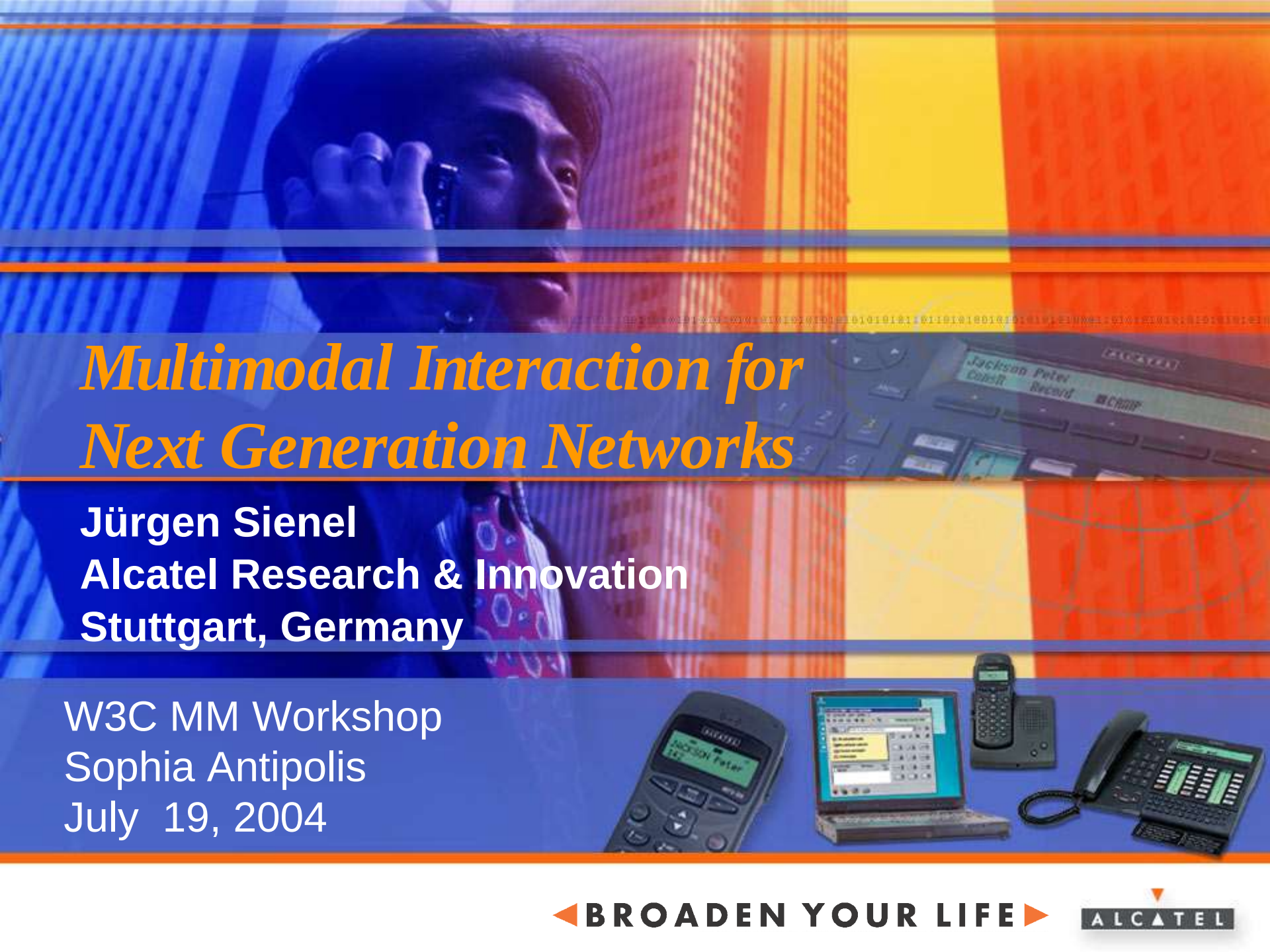


Multimodal Interaction for Next Generation Networks

Jürgen Sienel
Alcatel Research & Innovation
Stuttgart, Germany

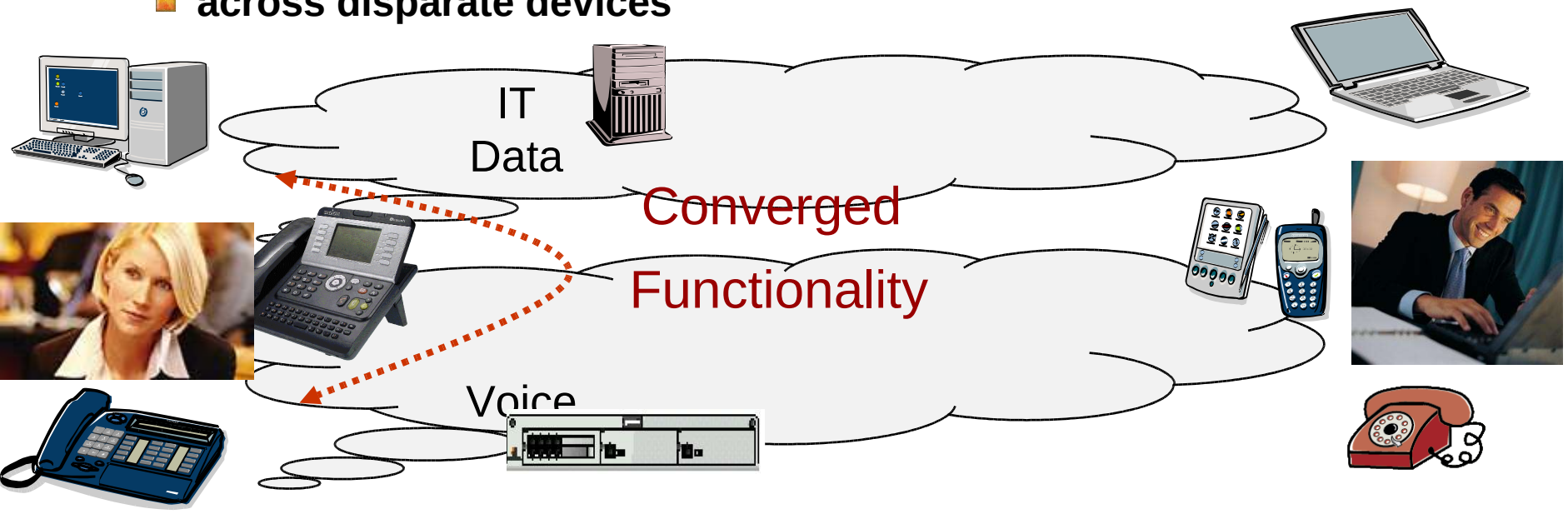
W3C MM Workshop
Sophia Antipolis
July 19, 2004

Outline

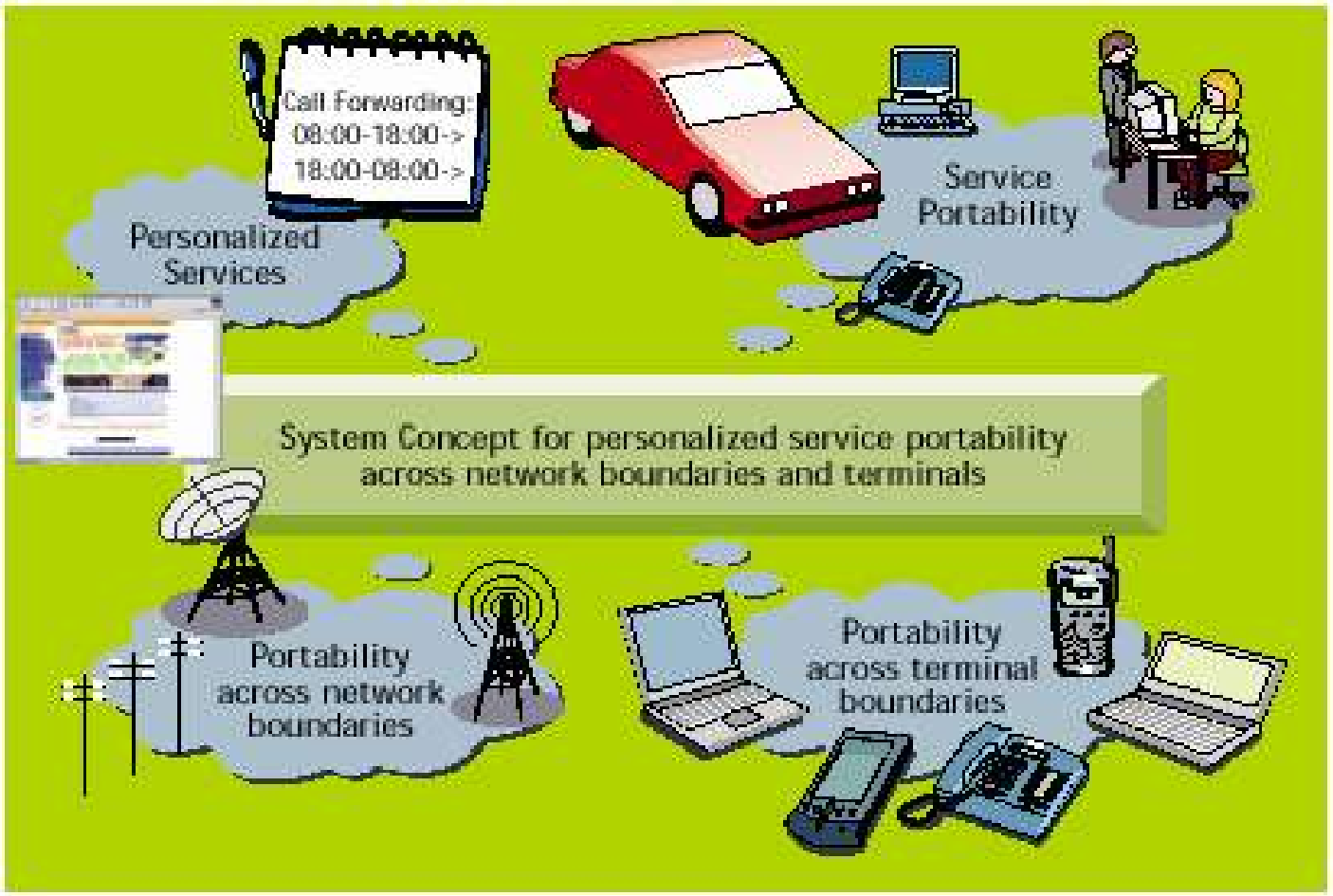
- Motivation
- Multimodal Applications
- Multimodal Architecture Approaches
- Standardisation Issues
- Conclusion

Converged Functionality

- **Access Information Services through Communication Networks**
  - **to deliver next generation services,**
  - **across the domains of enterprise, fixed and mobile,**
  - **across disparate devices**



Motivation
Mobile Environment



Human – Machine Communication

Customizable (operator, user), adaptive to user profile, preference and terminal capability

Multimodal User Interaction

Fixed

Voice/ Graphic

Mobile

Voice

Car

Voice/ Graphic

Home

Graphic

Public Areas

Multimodal Interaction
Reasons

- Human perception allows the parallel processing of multiple input channels

- Higher „Bandwidth“ of communication (Non-verbal)
- Concentration on strength of each modality

- Selection of most appropriate modality depending on
  - environment, e.g. noisy
  - context, e.g. driving in a car
  - complexity of task, e.g. directory assistance
  - device capability , e.g. small displays
  - preferences and disabilities of the user, e.g. visually impaired

Multimodal Applications Operators Visions

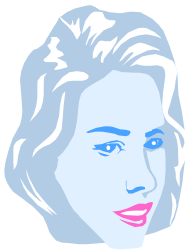
Enablers	Application Area	Services / Features
Automatic Speech Recognition	Telephone Services	Voice-activated dialing, Call Handling
Text-to-Speech	Information Services	Voice Portals, Wireless Web, Telematics
Web Interfaces	Messaging	Handling of Voice mail, email and UM, IM
Multimodal Interaction	Operator Services	Voice deputy, Directory Assistance
User Identification	Enterprise Applications	Call/Contact center
	Mobile Commerce	Multi Modal Event Notification, Mobile transactions
	Security Services	Speaker verification, Biometrics

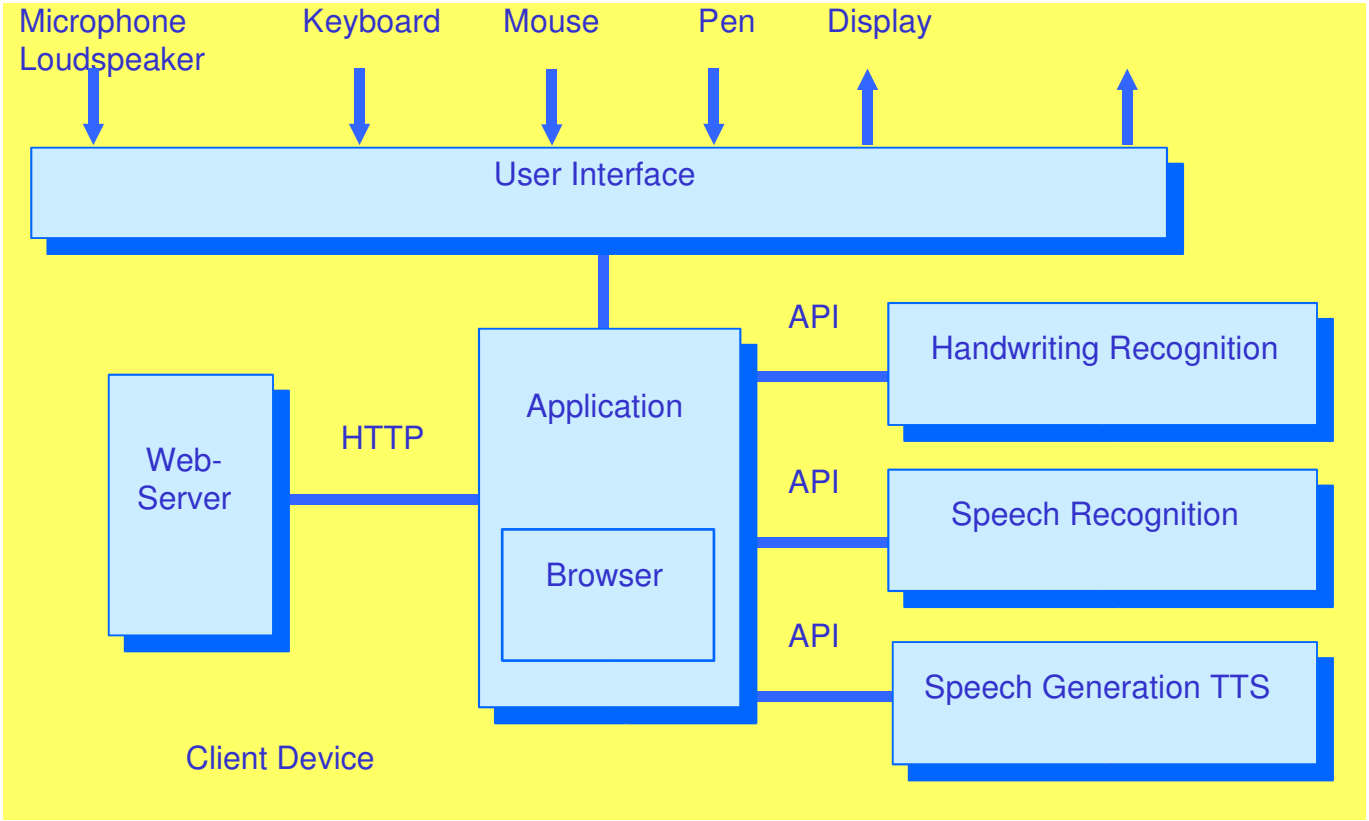
Multimodal Application
Instant Messaging

- Adaptation to terminal capability and user preference
- Flexible combination of visual and acoustical interaction
- Customization

Multi-modal	Multi-modal usage in Train (graphic only)	Multi-modal usage in Car (voice only)
Receiving a message		
Display: Alert on receiving a message Option: Voice: you got a message from Alain, would you like to see it? Speech input: Yes, No or Stylus : Icon press Display: message Option: Voice output	Display: Alert on receiving a message Stylus : Icon press Display: message	Voice: you got a message from Alain, would you like to hear it? Speech input: Yes, No Play Message

Approaches
Multimodal Browser





Multimodal Browser

Some pros and cons

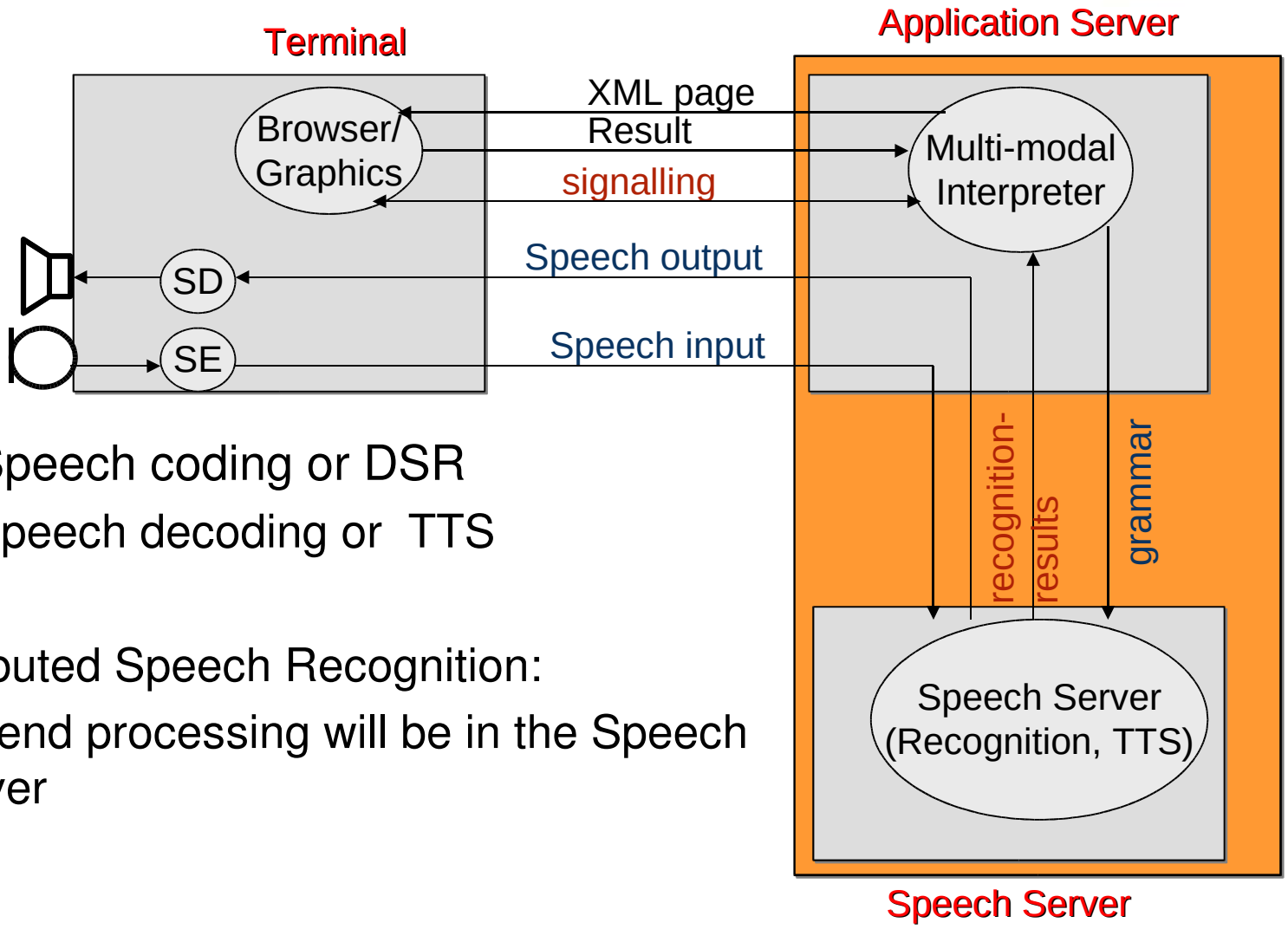
Pros:

- All functionality in one device
- Just one handler for a document
- Easy synchronisation methods of graphics and voice
- Direct interpretation and handling of sensors

Cons:

- Limited resources on mobile devices
- Dependent on the device
- Multilinguality may be missing
- Interaction Management has no deeper application knowledge, can only interpret the document
- Transfer of application data (e.g. grammars) might be more expensive than transfer of speech

Multimodal Architectures
Server based Approach



SE: Speech coding or DSR

SD: Speech decoding or TTS

Distributed Speech Recognition:

Backend processing will be in the Speech Server

Server based Approach

Some pros and cons

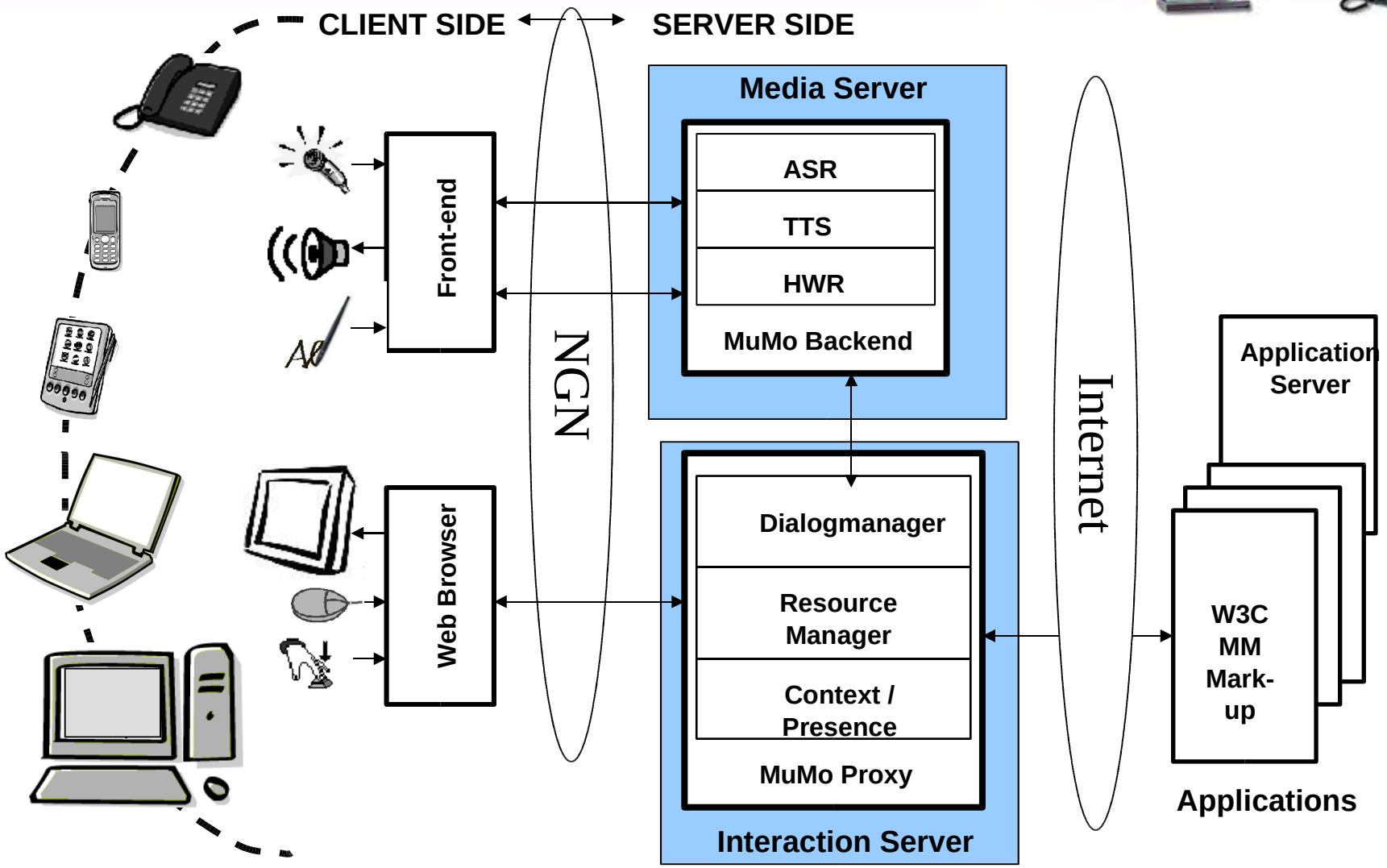
Pros:

- Exact knowledge of the application
- Handling of meta dialog
- Storage of voice records for security reasons (banking application)
- Easy support of multilingual applications

Cons:

- How to get detailed knowledge about the class of device
- Direct interpretation and handling of sensor data in terminal
- Harder to synchronise (delays)
- No sharing of ASR and TTS resources

Approaches
Distributed Architecture



Multimodal Browser
Some pros and cons

Pros:

- On demand functionality (use of local functionality where possible)
- Storage of voice records for security reasons (banking application)
- Could support of multilingual applications
- Better interpretation and handling of sensors and device capabilities
- Optimised network traffic
- May support multiple devices

Cons:

- Complex Synchronisation
- Interaction Management has no deeper application knowledge, can only interpret the document
- Higher standardisation effort needed
- Architecture may be not transparent for application developer

Requirements for Standardisation

- Multimodal Framework and Components
- System and Environment Definition
- Result proposition (EMMA)
- Support of Distributed Processing (DOM)

- Interface to Media Processing Modules (SpeechSc)
- Improved device descriptions and presence (DI, OMA)

- WebService Interface for component binding

- Interface on parallel devices
- Definition of modality independent dialogs and content

Conclusion

- Next Generation Networks will provide converged IT and communication access to a set of existing and new services and application
- Independence from the end-user device is must
- Multimodal Interfaces support the usability of such services and devices
- A network centric architecture offering On-Demand capabilities can support the multi device access
- Standardisation has to be continued, more interaction between the organisations might be needed to fulfil the common vision

www.alcatel.com