Multimodal and Multi-device Services Requirements

Candidate Version 1.1 – 13 Nov 2003

Open Mobile Alliance

OMA-RD-Multimodal_Multi-device_Services-V1_1-20031113-C

Contents

1. SCOPE (INFORMATIVE) 4
2. REFERENCES 5
   2.1 NORMATIVE REFERENCES 5
   2.2 INFORMATIVE REFERENCES 5
3. TERMINOLOGY AND CONVENTIONS 6
   3.1 CONVENTIONS 6
   3.2 DEFINITIONS 6
   3.3 ABBREVIATIONS 7
4. INTRODUCTION (INFORMATIVE) 8
   4.1 MOTIVATION AND VALUE PROPOSITION 8
   4.2 RELATED STANDARD AND INDUSTRY ACTIVITIES 10
   4.3 OBJECTIVES OF THE PRESENT REQUIREMENT SPECIFICATION 11
5. DESCRIPTION (INFORMATIVE) 12
   5.1 MULTIMODAL AND MULTI-DEVICE APPLICATIONS 12
   5.2 USE CASES 12
      5.2.1 Generic Examples 12
      5.2.2 Detailed use cases 14
   5.3 MULTIMODAL CONCEPTUAL MODEL 25
   5.4 SYSTEM CONFIGURATIONS 28
      5.4.1 Modular and interoperable systems 28
6. REQUIREMENTS (NORMATIVE) 30
   6.1 HIGH LEVEL REQUIREMENTS 30
      6.1.1 User-centric high level requirements 30
      6.1.2 Service-centric high level requirements 30
      6.1.3 OMA-Centric high level requirements 31
   6.2 SECURITY 31
   6.3 PRIVACY 32
   6.4 CHARGING 32
   6.5 ADMINISTRATION 33
   6.6 TERMINAL DEVICES AND SMARTCARDS 33
      6.6.1 Terminal devices 33
      6.6.2 Smartcards 33
   6.7 PLATFORMS 34
   6.8 NETWORK INTERFACES 34
   6.9 USABILITY 34
   6.10 INTEROPERABILITY 35
   6.11 APPLICATION DEVELOPMENT 36
APPENDIX A. CHANGE HISTORY (INFORMATIVE) 38
   A.1 APPROVED VERSION HISTORY 38
   A.2 DRAFT/CANDIDATE VERSION 1.0 HISTORY 38
      A.2.1 Document History 39
APPENDIX B. (NORMATIVE) 40
   B.1: W3C MULTIMODAL INTERACTION ACTIVITY 40
   B.2: VOICE APPLICATIONS 40
   B.3: DEVICE-INDEPENDENT AUTHORING 40

1. Scope (informative)

This document describes the requirements for multimodal and multi-devices services in the scope of the OMA architecture.

Such services enable access to mobile services through different modalities (e.g. keypad, GUI, Voice, handwriting) or devices.

The requirements are primarily viewed from an end user point of view. Implications are derived in terms of requirements on network operators, service providers and terminals manufacturers.

The requirements apply only to supporting the applications or services for which multimodal or multi-device interactions make sense and are desired. Applications and services may otherwise be designed without providing such a user experience.

2. References

2.1 Normative References

[CREQ] "Specification of WAP Conformance Requirements". Open Mobile Alliance™. OMA-WAP-221-CREQ-20010425-a.

[RFC2119] "Key words for use in RFCs to Indicate Requirement Levels". S. Bradner. March 1997. URL:http://www.ietf.org/rfc/rfc2119.txt

[RFC2234] "Augmented BNF for Syntax Specifications: ABNF". D. Crocker, Ed., P. Overell. November 1997. URL:http://www.ietf.org/rfc/rfc2234.txt

2.2 Informative References

[SES3GPP] 3GPP TR 22.977 – Feasibility Study for Speech Enabled Services (Release 6).

[SMIL] W3C SMIL 2.0. URL: http://www.w3.org/TR/smil20

[SPEECHSC] IETF SPEECHSC – Speech Service Control working group

[SRF3GPP] 3 GPP TS 22.243 – Stage-1 - Speech Recognition Framework for Automated Voice Services (Release 6).

[WAPARCH] "WAP Architecture". Open Mobile Alliance™. OMA-WAP-210-WAPArch-20001130-p.

[W3CDI] W3C Device Independent Activity, URI: http://www.w3c.org/2001/di/.

[W3CMMI] W3C Multimodal and multi-device Activity, URI: http://www.w3.org/2002/mmi/.

Requirements: http://www.w3.org/TR/mmi-reqs/

Framework: http://www.w3.org/TR/mmi-framework/

3. Terminology and Conventions

3.1 Conventions

The key words “MUST”, “MUST NOT”, “REQUIRED”, “SHALL”, “SHALL NOT”, “SHOULD”, “SHOULD NOT”, “RECOMMENDED”, “MAY”, and “OPTIONAL” in this document are to be interpreted as described in [RFC2119].

All sections and appendixes are normative, unless they are explicitly indicated to be informative.

3.2 Definitions

Automated Voice Services	Voice applications that provide a voice interface driven by a voice dialog manager to drive the conversation with the user in order to complete a transaction and possibly execute requested actions. It relies on speech recognition engines to map user voice input into textual or semantic inputs to the dialog manager and mechanisms to generate voice or recorded audio prompts (text-to-speech synthesis, audio playback). It is possible that it relies on additional speech processing (e.g. speaker verification). Typically telephony-based automated voice services also provide call processing and DTMF recognition capabilities. Examples of traditional automated voice services are traditional IVR (Interactive Voice Response Systems) and VoiceXML Browsers.
Conventional Codec	The module in a terminal that encodes the speech input waveform, similar to the encoder in a vocoder e.g. EFR, AMR.
Channel	A particular user agent (browser), device, or a particular modality.
DSR Optimised Codec	The module in a mobile terminal which takes speech input, extracts acoustic features and encodes them with a scheme optimised for speech recognition. This module is similar to the the conventional codec (e.g. AMR). On the server-side, the uplink encoded stream can be directly consumed by speech engines without having to be converted to a waveform.
Declarative	Based on markup/XML, as opposed to scripts or other procedural / imperative code.
Disconnected mode	The multimodal applications can maintain their state during a session. If the connection to the back-end service is interrupted they can operate in an off-line mode for the parts of the interaction that can be completed off-line and then update their state when the connection is restored.
Haptic interface	An interface that returns information to the user in the form of felt sensation. Haptic interfaces attempt to improve the lack of stimulus for the sense of touch in human-computer interaction by rendering information to the user through force feedback, heat, pressure or texture.
Mono-modal application	Application designed for access through only one channel or channel type (e.g. WAP, Web or Voice exclusively).
Multi-channel application	Applications designed for ubiquitous access through different channels, one channel at a time. No particular attention is paid to synchronization or coordination across different channels.
Multi-device applications	An application that supports the capability to interact with a particular application over a number of physical devices with user agents being synchronized with multi-device services. These browsers may support the same (e.g. GUI) or different modalities.
Multimodal application	An application that supports more than one interaction mode by relying on a combination of multiple input (e.g. key, stylus, voice, …) to access and manipulate information on the move and to enable the most convenient output (display, tactile, audio) at the discretion of the user/ terminal capability.
Multimodal browser	“Declarative” user agent that renders different modalities (e.g. GUI browser and Speech Browser) simultaneously available (concurrently or sequentially) and synchronized for the user to interact with the application. These different user agents may be located on the same device or distributed on among terminals or across the mobile network(s).
Speech Recognition Framework	A generic framework to distribute the audio sub-system and the speech services by sending encoded speech between the client and the server. For the uplink, it can rely on conventional (AMR) or on DSR optimised codecs where acoustic features are extracted and encoded on the terminal

Tactile interface An interface that allows a user to enter input through a touch-enabled display, the manipulation of a joystick or keypad-based controls or the use of a data-glove

Turn In a conversational interface, a turn is the time-period during which one party will communicate information before another party responds. This information may be presented using multiple modalities (voice, gestures, etc)

User agent The component that renders the presentation data into physical effects that can be perceived and interacted with by the user. For a given modality this may be a separate browser or platform or one or multiple components internal to a browser or platform.

3.3 Abbreviations

AMR Adaptive Multi Rate
DOM Document Object Model
DSR Distributed Speech Recognition
DSV
DTMF Dual Tone Multi-Frequency
GUI Graphical User Interface
IVR Interactive Voice Response System
MAE Mobile Application Environment
SES Speech Enabled Services
SRF Speech Recognition Framework
STQ Speech Processing, Transmission and Quality Aspects
URI Uniform Resource Identifier
WSIA Web Services for Interactive Applications
WSRP Web Services for Remote Portlets

4. Introduction (informative)

4.1 Motivation and value proposition

This specification defines the requirements for the support of multimodal and multi-device applications by the OMA architecture framework.

Online access to information is fast becoming a must-have. Along with this trend comes a new usage model for information access, particularly in mobile environments. Information appliances such as navigation systems in cars are standard in high-end cars already and will penetrate lower-end vehicles soon. Data access using mobile phones, though limited, has significant momentum and will become soon widespread. In this new computing paradigm, a person will expect to have access to information and interactions in a seamless manner in many environments, be it in the office, at home, in the car, often on several different devices. These new access methods have compelling advantages, such as mobile accessibility, low cost, ease of use, and mass-market penetration. They also have their limitations - in particular, it is hard to enter and access data using small devices, speech recognition can introduce mistakes that can sometimes be repeating and therefore blocking the transaction, one interaction mode does not suit all circumstances, and so on.

Mobile access usability often fails miserably; accomplishing even the simplest of tasks takes much too long to provide any user satisfaction. It is thus essential for the widespread acceptance of this computing paradigm to provide an efficient and usable interface on the different device platforms that people are expected to use to access and interact with information.

Indeed, today, mobile users are experiencing the following pain points:

- **It is hard to enter and access data using small devices** (GUI limitation for mobile devices)
  - Tiny keypads are difficult to use to enter long input. Thumb typing is challenging.
  - However, the displays, even if limited, are great to browse lists and images.
    - The capability to visualize images is a great advantage over text-only or voice-only channels that can't present an image to the user. On the other hand, limited display capability still limit the user experience and may sometimes prevent to convey some information or support particular interactions.
- **Speech Recognition still makes mistakes**
  - Speech recognition mistakes can be painful and even block usage of the application if repeated. For example, if a name dialling fails periodically, the lack of reliability often leads to abandonment by the user.
  - Limitation of speech and voice solutions that may appear unreliable.
- **Voice is serial** (Limitation of speech and voice solutions.)
  - It is difficult to manage long output.
  - However, speech is great for complex input.
- **One interaction mode does not suit all circumstances**
  - Each mode or device has its pros and cons depending on the end user's environment or preferences.
- **Accessibility**
  - Multiple modes on a device will be of enormous help to people with disabilities to choose a mode that suits them.

As the mobile internet access gains acceptance, the foreseen evolution of mobile devices does not immediately address user experience challenges:

- Devices are getting smaller, not larger
- Devices and applications are becoming more complex
- Adding color, animation, camera, etc. does not simplify the user interface
- Voice-only access is often not yet reliable enough for mobile users.
- CRMs (Customer Relation Management) / IVRs (Interactive Voice Responses) are mostly not yet web-centric.
- Etc…

We can expect and already observe a trend towards a new frontier of interactive services: **multimodal and multi-device services**.

Indeed, today, the ability to access information while on the move and the pain incurred to enter data using small portable mobile devices needs some immediate relief. Wireless browsers and voice user interfaces all have advantages and disadvantages.

With the provision of multimodal access, multi-device capabilities, and synchronized content and services, users will be able to freely choose the right mode and most appropriate device for a particular task and particular situation and therefore they can freely take advantage of the modality or device that is the best suited for a particular interaction in a particular situation. This is the main value proposition of multimodal and multi-device services.

The user experience is thus optimised by selecting the most appropriate access mechanism for a particular interaction in a particular situation/context. The user can do this selection when the application makes supplementary use of the different available modalities or devices, or rather provided by the service when the use is fundamentally complementary.

By letting the user switch modalities, the user can also for example correct speech recognition errors therefore significantly increasing the reliability of voice applications.

Multimodal and multi-device applications enable reliable access and interaction to information anytime, anywhere, anyhow, independently of the activity or environment, especially when using small portable mobile devices.

Multimodal and multi-device applications provide a compelling user interface which exploits the advantages of each modality or devices types without being constrained by their limitations: for example, talking is easier than typing, but reading is faster in some conditions are more appropriate than listening, etc....

Multimodal and multi-device services fundamentally address the hierarchy of user's needs illustrated in Figure 1.

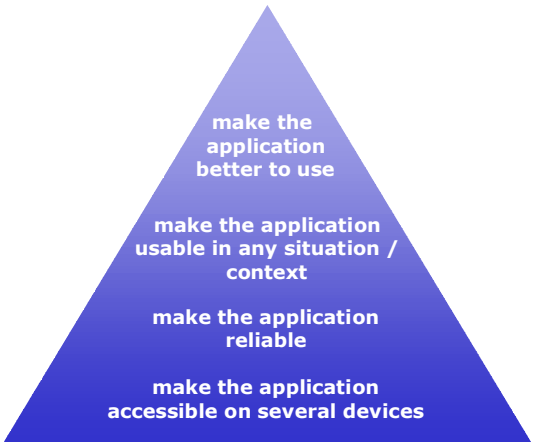

Figure 1 – Hierarchy of needs that multimodal and multi-device services should target to satisfy. Context designates in this figure any external environment variable.

At the most primitive level, users need to get their applications on a variety of performance enhancing devices. As the proliferation of these anywhere applications grow, reliability, independence of user situation, and then ease of use will become the necessary ingredients to increase the user efficiency. The continuing expansion of application capabilities will lead toward the ability to create and deliver personalized tailored solutions that fulfill user needs.

Note also that GUI is fundamentally a multimodal and multi-device user interface that combines output on a screen and input on a keyboard, keypad (+ pointer), or touchscreen (ink input). This illustrates that the notion of multimodality may change in time as particular multimodal interactions become represented into a limited set of programming patterns.

The need for multi-channel, multi-device and multimodal interfaces will be exacerbated by the recent legislation initiatives to provided accessible services. Mobile service providers (operators, terminal vendors and content / application providers) may soon be liable for not providing appropriate mechanism to allow people with disabilities to access their services. In the mobile spaces, this adds additional levels of complexity over the wired internet or voice / telephony space. Supplementary multimodal applications is often an efficient way to achieve accessibility; for large sets of supported disabilities and alternative access mechanism and modalities. For a while, accessibility support should also include support of TTY communications.

4.2 Related standard and industry activities

Multimodal and multi-device interaction is gaining great interest in the industry. Different standard bodies are addressing it or related issues:

- W3C:
  - Multimodal interaction activity: expected to address aspects of the authoring of multimodal, multi-device and conversational applications as well as some associated exchange formats. This also includes specification of support for handwriting-based user interfaces.
    - The W3C Multimodal Interaction Requirements document.
    - Framework for multimodal interactions
    - APIs / interfaces to support synchronization of different modalities; independently of the authoring or deployment model
  - Voice Browsing: specification of voice channel (VoiceXML, CCML, data file XML specifications, …)
  - Device Independent Activity: specification of delivery context and device independent authoring (key for multichannel authoring and support of dynamic modalities)
  - XML events
  - DOM
  - XHTML
  - CSS / XSL
  - SMIL: multimedia synchronization
  - Web services and XML protocols (SOAP etc…)
- 3GPP:
  - SES – Speech Enabled Services, covering multimodal and multi-device services. The scope of SES is illustrated in Figure 2. The following specifications have been defined.
    - 3GPP TR 22.977 – Feasibility Study for Speech Enabled Services (Release 6).
    - 3 GPP TS 22.243 – Stage-1 - Speech Recognition Framework for Automated Voice Services (Release 6).

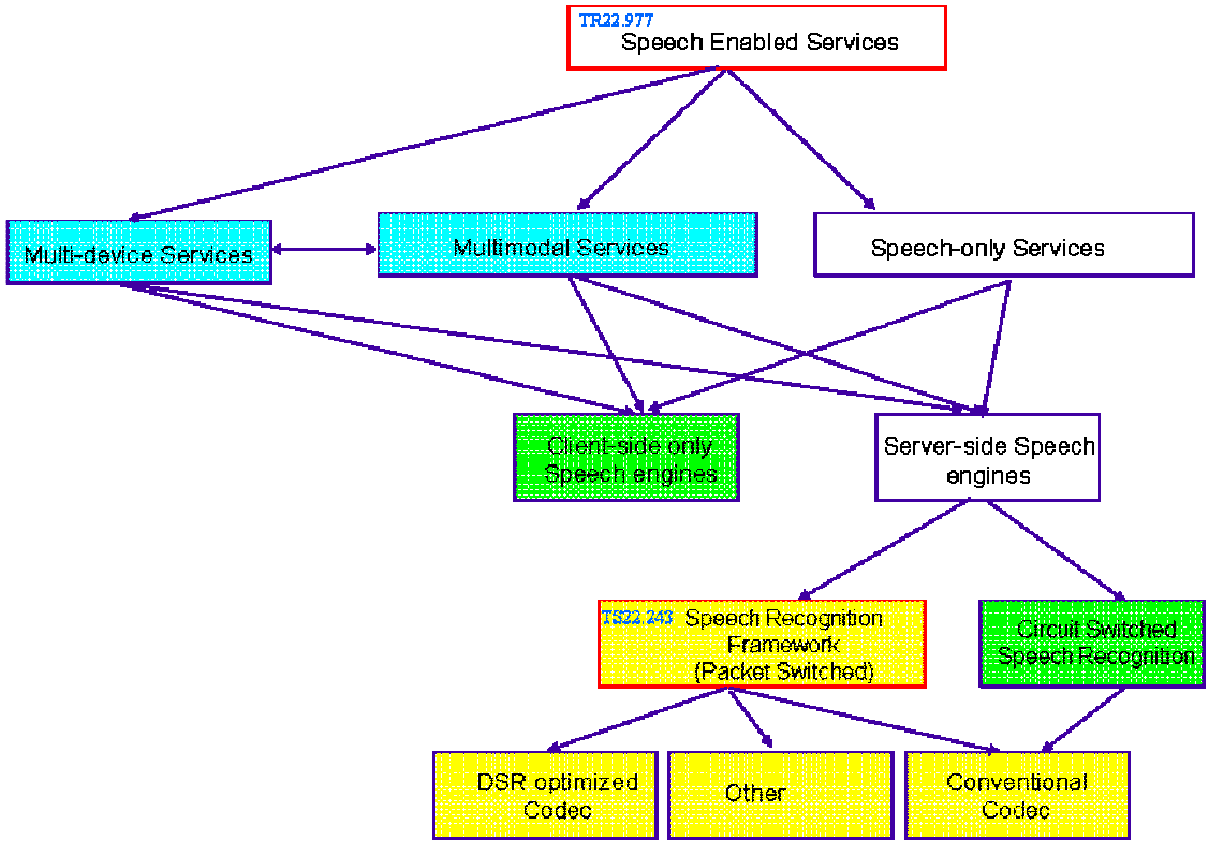

Figure 2 – SES at 3GPP

- ETSI:
  - STQ DSR (Aurora) – Promotion of SRF as a DSR framework (protocols stack and codecs) and a multimodal reference architecture that can take advantage in distributed case of the DSR framework.
- IETF:
  - SPEECHSC: The IETF speechsc group is aiming at a protocol for distributed control of speech servers. Theses speech remote control protocols are good candidates to enable several thin client configurations of multimodal browsers that are most relevant for mobile / 3G deployments.
  - AVT: payloads for different codecs (AMR, DSR) for speech and multimodal services.
- OASIS
  - WSIA and WSRP that address web services user interfaces, coordination between web services and usage in portals.
    - These activities address the issues of binding a user interface to web services and perfoming aggregation and coordination of web services with a user interface. These capabilities are directly relevant to:
      - Building multi-modal or multi-device applications by aggregation
    - The web service coordination activities and relationship to user interface may share technologies with approaches to consider to support multimodal and multi-device services.
- ITU:
  - SG16: that address DSR and DSV

In addition different other standard bodies, industry consortium are paying close interest to multimodal:
- VoiceXML Forum
- SALT Forum
- NCITS V2

OMA has initiated activities to support multimodal synchronization by the terminal (MAE WG) and as use case targeted by the OMA architecture.

4.3 Objectives of the present requirement specification

The present specification provides user-centric requirements for multimodal and multi-device services as well as voice enablement of OMA applications.

It then derives requirements on the OMA architecture in terms of features and capabilities on the terminal, network services, server and service authoring and deployment.

The assumptions for the design point aim at achieving interoperability, longevity and extensibility of OMA multimodal and multi-device services across a wide range of device capabilities, application capabilities, network capabilities and underlying deployment configuration.

An important objective of the specification will be to provide an end-to-end specification of mobile multimodal and multi-device services. This includes the provision of an overall architecture and content authoring specifications. Re-use and profiling of specifications produced by other standard bodies and fora mentioned above is intended, whenever appropriate (technically, time-wise etc.).

5. Description (informative)

5.1 Multimodal and multi-device applications

Based on the previous discussions, multimodal and multi-device services appear as a user interface service that aims at improving the user experience and increasing the reliability of the access to any application. Therefore, any application or service will greatly profit from multimodal or multi-device capabilities.

As most mobile applications, the multimodal and multi-device applications expected to present the most value for user are applications that:

- Are time-sensitive: the access to the information or transaction must take place now, not later.
- Enable communication
- Are location-specific, -dependent or -related.
- Provide entertainment.

Examples of such multimodal or multi-device services include (the list is non-exhaustive):

- Communication assistance (Name dialling, Service Portal, Directory assistance)
- Information retrieval (e.g., obtaining stock-quotes, checking local weather reports, flight schedules, movie/concert show times and locations)
- M-Commerce and other transactions (e.g., buying movie/concert tickets, stock trades, banking transactions)
- Personal Information Manager (PIM) functions (e.g., making/checking appointments, managing contacts list, address book, etc.)
- Messaging (IM, unified messaging, etc…)
- Information capture (e.g. dictation of short memos)
- Navigation (car navigation service)
- Games
- Location – based applications
- Adhoc, dynamic and nomadic network applications (e.g. Dynamic use of kiosks / wall mounted displays / phones / PDAs).
  - This kind of applications illustrates why multimodal and multi-device user interface can be considered as the core user interface of mobile computing. Possibilities are endless and such capability will be demanded and expected by mobile users.
- Remote-control application (multimodal remote control, multi-device remote console)
- Accessibility to mobile services.
- Multi-user applications (games, collaborations).

5.2 Use Cases

Numerous use cases can be considered, associated to numerous different mobile applications and services.

The Architecture framework working group has developed a use case for multimodal and multi-device services.

5.2.1 Generic Examples

5.2.1.1 Presenting complementary information on different output modes:

When using a mobile device to request audio information about the local weather forecast, a picture could be sent to the mobile to complement the spoken forecast. When asking for walking directions to a nearby restaurant, a map could be displayed. For an incoming call, the mobile display could show a photograph of the caller.

5.2.1.2 Allowing switching between different modes depending on the context:

In a noisy room, speech recognition may not work, or the user may be unable or speak. In this example, it would be advantageous to use keypad or handwriting inputs instead of speech input. Conversely, the user may be comfortable with

looking at a form on the display, but choose to use speech to fill in text fields, rather than struggling with awkward keypad inputs.

5.2.1.3 Allowing switching between different configurations:

- Switch between a fat client configuration with limited speech capabilities and a thin client configuration with rich speech capabilities.
- Switch, as the user moves away, in a multi-device case between a proximity-based synchronization (e.g. between terminal and kiosk) to a synchronization in the network between a voice interface and the terminal.
- Switches to support changes from connected to disconnected mode as discussed in section 6.6.1.

5.2.1.4 Biometric Authentication:

The end-user would log into their device using biometrics and then utilize tactile, voice, handwriting, gestural, or other mode for controlling their device. Biometrics would also be tied into any security requirements on applications that require security. The ability to change from one mode to another while in a session (either voice or data) enables the end-user to utilize any mode the situation requires.

5.2.1.5 Future multimodal devices will have adaptive interfaces:

These are interfaces that are able to adapt themselves to the user and interaction - specific to the end-user's profile. Having one mode does not allow for this scenario.

5.2.1.6 Multi-user interactions

Multimodal and multi-device interfaces may allow for many users. Accordingly, multiple users can interact with a same application on different devices or using different modalities. This is an extension of the previous examples where the different interactions channels may be used by different users. Useful applications includes:

- Games
- Shared collaborative applications (Calendering, white board/shared presentations).

5.2.1.7 Multi-device interactions

Most people have several devices. These devices provide varying types of functionality to users. The ability to use the best of the best when these devices are together would enable the user to use the largest display if need be, or the best speaker, or handwriting - simply using the best interface for the situation at hand.

Multi-device interactions enable user to select the device or devices that are the most appropriate for a given transaction in a given context. In the mobile space, this is especially usefule when it allows user to handle different mobile devices (e.g. phones and PDAs) or interface with other systems interconnected voia proximity networks or via a combination or wireless and wired network.

Use cases include:

- Remote control capabilities
- Proximity payments and transactions systems
- Kiosk / wall screen combinations with mobile terminals
- Accessibility device as discussed in section 5.2.2.6 about Figure 5.

Multi-device interactions are very closely and treated similarly to multimodal interactions because they provide the same value proposal to allow the user to select the most appropriate channel of interaction (modality or device) to perform a particular interaction in a particular context.

5.2.2 Detailed use cases

The use cases presented here are classified into the following categories:

- Multimodal Content Use Cases
- Multimodal Automobile Use Cases
- Multimodal Call Center Use Cases
- Multimodal Use Cases with Local and Remote Functionality
- Multimodal Application Download Use Cases
- Multimodal Accessibility Use Cases

In the sub-clauses that follow describing the use cases, further flows may be required where they are required to meet functional, security, usability or business needs. For the sake of clarity these have been omitted.

5.2.2.1 Multimodal Content Use Cases

5.2.2.1.1 Short Description

End User/Subscriber navigates to their favorite multimodal content site by speaking the shortcut names of the site to their multimodal enabled mobile terminal. The user receives detailed information via GUI, voice, or other modality. The user may then browse various options and download content. This service may require a user to subscribe to a service offering the multimodal content download service

5.2.2.1.2 Actors

- End User/Subscriber. The User/Subscriber is authorized to define and change user preference parameters and content selection.
- Network operator or authorized agent of the network operator
- Service provider which has an established billing agreement with operator
- Content provider which has an established agreement with service provider

5.2.2.1.2.1. Actor Specific Issues

- End User/Subscriber: The User is the consumer of the service.
- Service Provider: The user has subscribed to a service provider offering the multimodal content service. Service Provider has established billing agreement with the operator.
- Network Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Content provider: Content Provider produces the content for the service. It has established an agreement with Service Provider

5.2.2.1.2.2. Actor Specific Benefits

- End User/Subscriber. The User benefits from more natural interaction with the service.
- Operator: Operator benefits from a share of the money from the user. It also benefits from the increased data flow due to the increase in the usage of such service.
- Service Provider: Service provider gets a share of the money from the user for every user session. The increased usability experience of the user, including the ability to use the service from any place, increases the monetary benefits for the movie service provider.
- Content provider: Content provider also benefits from generating new content for multimodal services.

5.2.2.1.3 Pre-conditions

- User is the Subscriber and has purchased a service contract with the Network Operator. The end user may or may not subscribe to a multimodal content service.
- Authorized agent (e.g. a content or service provider) has a multimodal system to distribute multimodal content to subscribers.
- Device is capable of interfacing with a multimodal content service and is provisioned with parameters necessary to obtain the services the User/Subscriber has purchased. The device also has multiple modes of output methods available to the user.

5.2.2.1.4 Post-conditions

- Device is provisioned with parameters necessary to obtain the services the User/Subscriber has purchased.
- Network and service provider end points recognize the device as having authorization to use, download, or browse the purchased multimodal services and content.
- Device and all purchased multimodal content are fully operational.

5.2.2.1.5 Normal Flow

1. Device is discovered by local network infrastructure with multimodal interaction capabilities.
2. The multimodal system queries for device and user information, including but not limited to identification of subscription information, device type, and user preferences.
3. A trusted relationship with the service/content provider is established.
4. Device type, capabilities and user preferences are sent to the multimodal system.
5. User browses available multimodal content.
6. User selects desired content and a request for content is sent to the multimodal system.
7. Multimodal content is sent to Device and presented using the appropriate modality.
8. The delivery of the requested multimodal content is confirmed.

5.2.2.1.6 Alternative Flow

Alternatives to the flow described in section 5.2.2.1.5 include, the user may initiate the multimodal content service or the subscribed service may be initiated by the network, service or content provider.

The browsing or output modalities for the content may be in the form of a single output mode, or multiple sequential or composite modes.

5.2.2.1.7 Operational and Quality of Experience Requirements

- Each user is identified by the normal means used for identification for other services.
- The terminal must be capable of providing the service obeying the security and privacy requirements.
- Users shall be able to use the service when roaming to another operator's network.
- There is good voice quality for the feedback from the service
- The round trip time experienced by a user during multimodal interaction shall remain acceptable.

5.2.2.2 Multimodal Automobile Use Cases

5.2.2.2.1 Short Description

- End User/Subscriber connects to service to receive driving directions and inquire about points of interest along the travel route such as restaurants and gas stations. This service may require a user to subscribe in order to access the information or it can be part of a service plan of automobile ownership. The user can initiate the connection to service by speaking a shortcut name for the service.
- End User/Subscriber connects to voicemail service. This service may require a user to subscribe in order to access the information or it can be part of a service plan of automobile ownership. The user can initiate the connection to service by speaking a shortcut name for the service.

Note #1: The automobile scenarios offer two complementary modes of operation to provide an enhanced multimodal experience when the End User (assumed to be the driver) is not engaged in driving, i.e., the car is parked and also to address the concerns for hands-busy/eyes-busy operation when the car is in motion.

Note #2: The device is connected to the automobile display and audio equipment to enhance the quality of its input and output functionality

5.2.2.2.2 Actors

- End User/Subscriber. The User/Subscriber is authorized to define and change user preference parameters affecting the behavior of the service access
- Network operator or authorized agent of the network operator
- Service provider that has an established billing agreement with operator
- Content provider that has an established billing agreement with service provider

5.2.2.2.2.1. Actor Specific Issues

- End User/Subscriber: The User is the consumer of the service.
- Service Provider: The user has subscribed to a service provider offering the multimodal content service. Service Provider has established billing agreement with the operator.
- Network Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Content provider: Content Provider produces the content for the service. It has established an agreement with Service Provider

5.2.2.2.2.2. Actor Specific Benefits

- End User/Subscriber. The User benefits from: more
- More natural interaction with the service.
    - Timely access to services and information while driving
    - Interface alternatives that address safety and efficiency while driving
- Operator: Operator benefits from a share of the money from the user. It also benefits from the increased data flow due to the increase in the usage of such service.
- Service Provider: Service provider gets a share of the money from the user for every user session. The increased usability experience of the user, including the ability to use the service from any place, increases the monetary benefits for the movie service provider.
- Content provider: Content provider also benefits from generating new content for multimodal services.

5.2.2.2.3 Pre-conditions

- User is the Subscriber and has purchased a service contract with the Network Operator. The end user may or may not subscribe to a multimodal content service (driving directions)
- User is the Subscriber and has purchased a service contract with the Network Operator. The end user must subscribe to a multimodal content service (voicemail)
- Authorized agent (e.g. a content or service provider) has a multimodal system to distribute multimodal content to subscribers.
- Device is capable of interfacing with a multimodal content service and is provisioned with parameters necessary to obtain the services the User/Subscriber has purchased. The device also has multiple modes of output methods available to the user. The device may also be able to interface with the automobile display and audio equipment for higher quality output methods
- Additional services may be required such as location services and GPS.

5.2.2.2.4 Post-conditions

- Device is provisioned with parameters necessary to obtain the services the User/Subscriber has purchased.
- Network and service provider end points recognize the device as having authorization to use, download, or browse the purchased multimodal services and content.
- Device and all purchased multimodal content are fully operational.

5.2.2.2.5 Normal Flow

The user is initially not connected to any service. The user initiates the service connection either by dialing into the service or by pressing button on steering wheel (which effectively dials into the service).TBD

- The automobile can transmit its current GPS coordinates to provide user's current location
- User says destination address
- Service responds with confirmation of destination address, calculates the route and the final driving directions are sent to the automobile multimodal system.
- As the user continues driving the turn-by-turn directions are played back through the voice/audio output interface of the system.
- The user asks for a gas station along the route
- The system calculates directions to nearest gas station along the original trip route and sends these directions to the automobile multimodal system.
- Directions to the nearest gas stations are played back through the voice/audio output interface of the system
- When the user arrives at the gas station and the car is stationary, the user can review a map of the area with the overlaid driving directions on the device display.
- The user can interact with the device using visual and voice input to change the details of the trip before getting back on the road.

5.2.2.2.6 Alternative Flow

Alternatives to the flow described in section 5.2.2.2.5 include, the user may initiate the multimodal content service or the subscribed service may be initiated by the network, service or content provider may initiate the subscribed service.

The user may initiate the service with the device when the user is not in the car using a combination of visual and voice multimodal interface. The user then gets in the car, and the interface changes to a voice-only interface to continue the interaction.

The browsing or output modalities for the content may be in the form of a single output mode, or multiple sequential or synchronized concurrent modes.

Voice operation of the device can is fully complementary to other multimodal sequential or composite input concurrent modes so as to satisfy the automobile hands/eyes busy environment. When the car is parked, the service can be accessed with a combination of voice and visual modes for input and output. When the car is in motion, the service is accessible only through a voice interface both for input and output. As an alternative, it is possible to suggest a secondary channel for output that uses a graphics display but the primary output uses voice and audio (Head-up display (HUD)).

5.2.2.2.7 Operational and Quality of Experience Requirements

- Each user is identified by the normal means used for identification for other services.
- The terminal must be capable of providing the service obeying the security and privacy requirements.
- Users shall be able to use the service when roaming to another operator's network.
- There is good voice quality for the feedback from the service
- The round trip time experienced by a user during multimodal interaction shall remain acceptable

5.2.2.3 Multimodal Call Center Use Cases

5.2.2.3.1 Short Description

This example describes the interaction of a user with a movie reservation service. A user session is initiated in the beginning and the user is authenticated. Based on user's interaction, video is also pushed to the terminal.

5.2.2.3.2 Actors

- User: The user is the consumer of the service.
- Service Provider: The user has subscribed to a service provider offering the call center reservation service. Service Provider has established billing agreement with the operator
- Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Content provider: Content Provider produces the content for the service. It has established an agreement with Service Provider

5.2.2.3.2.1. Actor Specific Issues

- End User/Subscriber: The User is the consumer of the service.
- Service Provider: The user has subscribed to a service provider offering the multimodal content service. Service Provider has established billing agreement with the operator.
- Network Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Content provider: Content Provider produces the content for the service. It has established an agreement with Service Provider

5.2.2.3.2.2. Actor Specific Benefits

- End User/Subscriber. The User benefits from more natural interaction with the service.
- Operator: Operator benefits from a share of the money from the user. It also benefits from the increased data flow due to the increase in the usage of such service.

- Service Provider: Service provider gets a share of the money from the user for every user session. The increased usability experience of the user, including the ability to use the service from any place, increases the monetary benefits for the movie service provider.
- Content provider: Content provider also benefits from generating new content for multimodal services.

5.2.2.3.3 Pre-conditions

- User is the Subscriber and has purchased a service contract with the Network Operator. The end user may or may not subscribe to a multimodal content service.
- Service Provider: The user has subscribed to a service provider offering the call center reservation service. Service Provider has established billing agreement with the operator
- Network Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Content provider: Content Provider produces the content for the service. It has established an agreement with Service Provider

5.2.2.3.4 Post-conditions

- The user has reserved a ticket for seeing the movie. The multimodal service session is now closed.

5.2.2.3.5 Normal Flow

Here the interaction is basically sequential interaction (only one modality at a time) and directed by the machine.

1. The user is initially not connected to any service. The user calls the “reservations” service, and validates their user name and password.
2. The user says: “movies in Atlanta today evening”.
3. Service: A list of movie names that have free seats available is displayed on her terminal. The user is requested to click any movie link for more information.
4. The user clicks on one of the movies to get more information about the movie.
5. The user is asked whether they would like to watch the trailer.
6. When the user says “yes”, a 2 minute movie trailer (video) is downloaded to their terminal.
7. The user decides to see the movie and clicks the link provided for a map of the seat.
8. The system shows a map of the seats.
9. The user selects a seat in the back row by saying the number of the seat they wish and completes the reservation.
10. A movie e-ticket is sent to the user as an SMS.

5.2.2.3.6 Alternative Flow

Here the interaction is basically composite interaction (multiple modalities at a time) and mixed initiative (directed by both machine and the user).

1. The user is initially not connected to any service. The user calls the “reservations” service, and validates their user name and password.
2. The user says: “movies in Atlanta today evening”.
3. Service: A list of movie names that have free seats available is displayed on her terminal. The user is requested to click any movie link for more information.

4. The user clicks on one of the movies and says "I'd like to see the trailer of this movie"

5. A 2 minute movie trailer (video) is downloaded to their terminal.

6. The user decides to see the movie and clicks the link provided for a map of the seat.

7. The system shows a map of the seats.

8. The user selects a seat in the back row by saying the number of the seat the user wishes and completes the reservation.

9. A movie e-ticket is sent to the user as an SMS.

5.2.2.3.7 Operational and Quality of Experience Requirements

- Each user is identified by the normal means used for identification for other services.
- The terminal must be capable of providing the service obeying the security and privacy requirements.
- Users shall be able to use the service when roaming to another operator's network.
- There is good voice quality for the feedback from the service
- The round trip time experienced by a user during multimodal interaction shall remain acceptable.

5.2.2.4 Multimodal Use Cases with Local and Remote Functionality

5.2.2.4.1 Short Description

Depending on the capabilities of the multimodal device local functionality can be deployed for multi modal services and applications. The use case described in the following aimes at name dialing functionality, where in a first step local address book entries of the device are utilized, and if no appropriate entry is found, a remote services is contacted for performing exhaustive searches on public or company specific services.

The use case represents a category of multimodal services, that extend local device services and applications by comprehensive server-based applications, offering a homogenious user interaction across device boundaries.

5.2.2.4.2 Actors

- End User/Subscriber. The user has an agreement to contact a company specific directory service or to use a public directory service
- Network operator or authorized agent of the network operator
- Service provider which has an established billing agreement with operator
- Content provider which has an established agreement with service provider

5.2.2.4.2.1. Actor Specific Issues

- End User/Subscriber: The User is the consumer of the service.
- Network Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Service Provider: The user has subscribed to a service provider offering the multimodal directory service. Service Provider has established billing agreement with the network operator.
- Content provider: Content Provider produces and maintains the address database for the service. It has established an agreement with Service Provider

5.2.2.4.2.2. Actor Specific Benefits

- End User/Subscriber. The User benefits from the extension of the service regarding comprehensiveness of the functionality, easier access to address data and more multimodal address data provision like images, location maps etc. Users can decide themselves upon the free local, but restricted services or the comprehensive remote services with costs.

- Operator: Operator benefits from the increased data flow, arising from the increase in the usage of such service as well as from the increase in multimodal content.
- Service Provider: Service provider gets a share of the money from the user for every user session. The increased usability experience of the user, including the ability to use the service from any place, increases the monetary benefits for the service provider.
- Content provider: Content provider have the opportunity to offer their databases also to multimodal services.

5.2.2.4.3 Pre-conditions

- User is the Subscriber and has purchased an air time service contract with the Network Operator. The end user may or may not subscribe to a multimodal name dialing service. In case of a company directory service, the user has to be authorized to access the enterprise network.
- The content provider may extend the directory address data by multimodal content e.g. images of users
- The service provider has a multimodal system to distribute multimodal content to subscribers.
- Device is capable of performing multimodal access to the local address database for name dialing purposes. Furthermore the device is provisioned with parameters necessary to access remote directory services. The device has the same modes of input and output methods available to the user for the local and remote services.

5.2.2.4.4 Post-conditions

- Device is provisioned with parameters necessary to obtain the services the User/Subscriber has purchased.
- Network and service provider end points recognize the device as having authorization to use, download, or browse the purchased multimodal services and content.
- Device and all purchased multimodal content are fully operational.

5.2.2.4.5 Normal Flow

1. User activates the voice dialing function on multi modal phone and speaks name of recipient
2. The mobile device performs a search in the local address book
3. If the entry is not found in the local address book, a remote directory services is called
4. The remote service has the same look-and-feel, so that the user is subtly moved from a local to an remote interaction
5. Directed dialogs and helps are used to narrow down the search using aural and visual prompts
6. When the user finds the right entry, he is directly connected to the recipient and in parallel the address data are sent by SMS to the user's phone
7. The user can enter the address data in the local address book for a subsequent call

5.2.2.4.6 Alternative Flow

Alternatively to the flow described, the user may access a remote company directory service. Here, the user will be authenticated to enable the enterprise network access. Then he can access the multimodal service for company address data.

5.2.2.4.7 Operational and Quality of Experience Requirements

- The user shall be able to select and change access parameters of remote address services
- In case of protected remote services like enterprise directories the user shall be able to be identified by the normal means used for identification for other services.
- The terminal must be capable of providing the service obeying the security and privacy requirements.

- Users shall be able to use the service when roaming to another operator's network.
- The transfer time from the local to the remote service experienced by a user during the multimodal interaction shall remain acceptable.

5.2.2.5 Multimodal Application Download Use Cases

5.2.2.5.1 Short Description

The download of applications enables the user to gain new functionality for his device. Thus, multimodal applications like game, tools, office applets as well as interaction and communication support allow new experiences and latest innovation regarding multimodality. Downloaded multimodal applications exploit local capabilities of mobile devices for providing multimodal interaction of the user with the device. The application itself can of course make use of remote content and services.

5.2.2.5.2 Actors

- The End User/Subscriber is authorized to define and change user preference parameters and application selection
- Network operator or authorized agent of the network operator
- Service provider which has an established billing agreement with operator for offering application download services
- Application provider which has an established agreement with service provider to provide multimodal applications to the application download service provider

5.2.2.5.2.1. Actor Specific Issues

- End User/Subscriber: The User is the consumer of the service.
- Service Provider: The user has subscribed to a service provider offering the multimodal content service. Service Provider has established billing agreement with the operator.
- Network Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Application provider: Application Provider maintains the applications for the service. It has established an agreement with Service Provider

5.2.2.5.2.2. Actor Specific Benefits

- End User/Subscriber. The User benefits from new and changing multimodal applications selectable and possibly customizable according to his needs. The ability to use the service from any place, allows the user to download new multimodal functionality during travelling and on holidays.
- Operator: Operator benefits from a share of the money from the user. It also benefits from the increased data flow due to the increase in the usage of such service. Especially in case of an access of the downloaded application to further services, this generates additional revenue during the application lifetime.
- Service Provider: Service provider gets a share of the money from the user for every download.
- Application provider: Application provider also benefits from generating new multimodal applications.

5.2.2.5.3 Pre-conditions

- User is the Subscriber and has purchased an air time service contract with the Network Operator. The end user may or may not subscribe to a multimodal application download service.
- The service provider has a download system to offer multimodal applications to users.
- Device is capable of executing multimodal applications and provides the supported device modalities to the multimodal application.

5.2.2.5.4 Post-conditions

- Device is provisioned with parameters necessary to obtain the download service the User/Subscriber wants to invoke.
- Network and service provider end points recognize the device as having authorization to use and download the purchased multimodal application.
- Device and all purchased multimodal applications are fully operational.

5.2.2.5.5 Normal Flow

1. User connects to and authenticates at Download Service
2. User selects gaming application to be downloaded
3. Download is started and the multimodal application is transmitted to the user's device
4. Downloaded Game is decompressed if necessary and becomes registered in mobile device
5. The game utilizes the multi-modal capabilities of the device and integrates the supported device modalities for the game
6. User starts game and controls it utilizing the modalities of the device such as speech and keypad
7. Gaming output is done via the device output modalities like sound and screen.

5.2.2.5.6 Alternative Flow

An alternative flow is that the user selects the gaming application via different multi modal interaction using speech to select gaming category and cursor to select particular game. Thus the interaction with the download service can also be performed with the same modalities as the latter application is handled.

5.2.2.5.7 Operational and Quality of Experience Requirements

- Each user is identified by the normal means used for identification for other services.
- The device must be capable of providing the service obeying the security and privacy requirements. The application must be executed with security constraints to guarantee the privacy of the user data and the stability of the user device.
- Users shall be able to use the service when roaming to another operator’s network.
- The device has to expose his multimodal capabilities to the multimodal application
- The device has to provide interfaces to the modalities to the multimodal application

5.2.2.6 Multimodal Accessibility Use Cases

5.2.2.6.1 Short Description

Visually impaired people can utilize a mode of input or output that allows for their impairment, for example a visually impaired person can access information through a limited set of devices according to their modalities. Speech, Braille, touch sensitive interfaces are examples of these interaction methods. Having multimodal interfaces allow the impaired population to utilize the modes that are appropriate to their situation. Another example could be head gestures where an individual who has limited arm movement could control the device using their head or hand gestures. This could simply be a yes or no situation with 2 modes - gestural and graphical output on the display with voice as another option if need be.

To illustrate the accessibility issues, a particular example is described. Numerous other accessibility use cases can be considered.

The described use case represents "Design for All" consideration, to allow a consistent service access even with modality constraints. Predefinition and restrictions of particular modalities should be in the responsibility of the user.

5.2.2.6.2 Actors

- End User/Subscriber. The user has an agreement to contact a multimodal news service
- Network operator or authorized agent of the network operator
- Service provider which has an established billing agreement with operator
- Content provider which has an established agreement with service provider

5.2.2.6.2.1. Actor Specific Issues

- End User/Subscriber: The User is the consumer of the service.
- Network Operator: Operator provides the billing for the service. Operator can also be the service provider.
- Service Provider: The user has subscribed to a service provider offering the multimodal news service. Service Provider has established billing agreement with the network operator.
- Content provider: Content Provider produces and maintains the news and information for the service. He has to take care to provide the same level of information for the different interaction modalities and to perform a consistent authoring of the content. The content provider has established an agreement with Service Provider

5.2.2.6.2.2. Actor Specific Benefits

- End User/Subscriber. The User benefits from the different modalities of the service as he can determine the most convenient modality without worrying about a potential information loss due to a missing modality.
- Operator: Operator benefits from the increased data flow due to new customers that might be focused on particular modalities and which would not use the service, if it was with only one specific modality.
- Service Provider: Service provider gets a share of the money from the user for every user session.
- Content provider: Content provider have the opportunity to offer their news and information databases also to multimodal services.

5.2.2.6.3 Pre-conditions

- User is the Subscriber and has purchased an air time service contract with the Network Operator. The end user may or may not subscribe to a multimodal news service.
- The content provider may extend the news by further multimodal content e.g. photos, interviews, graphs, maps to establish a comprehensive multimodal experience.
- The service provider has a multimodal system to distribute multimodal content to subscribers.
- Device is capable of performing being restricted to particular modalities by the user's profile. Furthermore the device is provisioned with parameters necessary to access remote news services.

5.2.2.6.4 Post-conditions

- Device is provisioned with parameters necessary to obtain the services the User/Subscriber has purchased.
- Network and service provider end points recognize the device as having authorization to use, download, or browse the purchased multimodal services and content.
- Device and all purchased multimodal content are fully operational.

5.2.2.6.5 Normal Flow

1. A visually impaired user has the modalities of the cell phone in the user's profile restricted to speech modalities
2. The user accesses the corporate news service via voice commands
3. User gets a choice of news by synthesized speech

4. The user enters the selection by a voice command

5. The news are prompted by audio

5.2.2.6.6 Alternative Flow

A non-impaired person accesses the same information via a visual browser and tactile interaction. He will get the same information by using a different modality.

5.2.2.6.7 Operational and Quality of Experience Requirements

- The user select the preferred modalities across service boundaries. Thus the user's profile on the mobile device is available to different services.
- The terminal must be capable of providing the service obeying the security and privacy requirements.
- Users shall be able to use the service when roaming to another operator's network.
- The device supports the restrictions to particular modelities.

5.3 Multimodal Conceptual Model

The following two figures illustrate conceptually, how multimodal applications (illustrated with voice and GUI (e.g. with screen output and keypad input) improve the user interface and increase reliability.

Figure 3 illustrates conceptually the evolution of the user experience from today with only mono-channel services (i.e. accessible through Voice, Web or WAP only) and multi-channel (accessible through several of these channels, one channel at a time) services, towards a multimodal or multi-device space where users can interact with more than one device or modality at the same time.

This can be extended to other modes of interaction like pen input, multimedia output, video input, haptic/tactile input and output, TTY communications, etc…

Similarly, a user may interact with the same application simultaneously or sequentially through different devices (e.g. phone, PDA, wall screen display, kiosk, etc…). Fundamentally this depends on the capabilities of the terminals and the fact that such a modality or device is supported by the way that the multimodal or multi-device application has been authored.

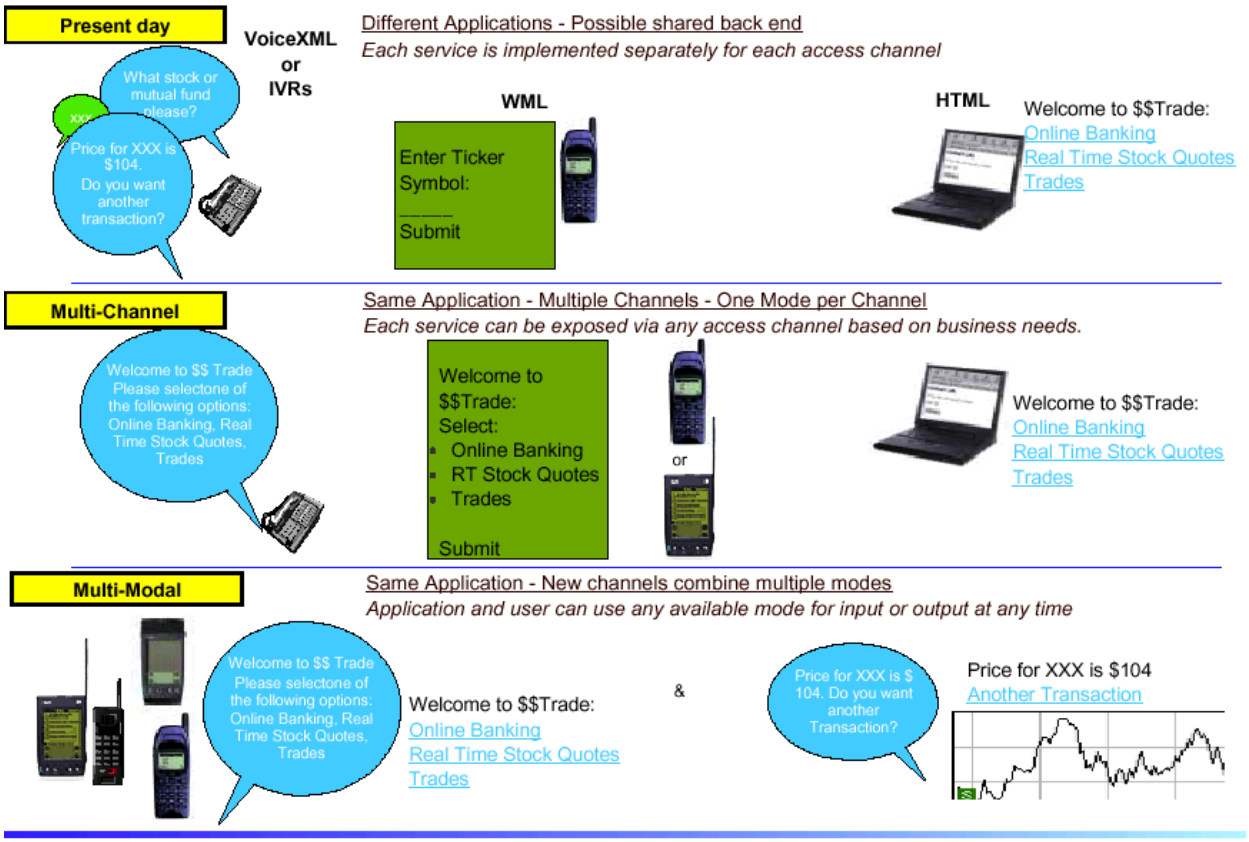

Figure 3 – Illustrates schematically the evolution of applications that are today designed for a single channel (monochannel) towards applications accessible through different channels, one channel (modality or device) at a time and eventually multimodal and multi-device applications where different devices or modalities can be used concurrently.

An example of a usage scenario for a multimodal air travel reservation service is summarized in Figure 4.

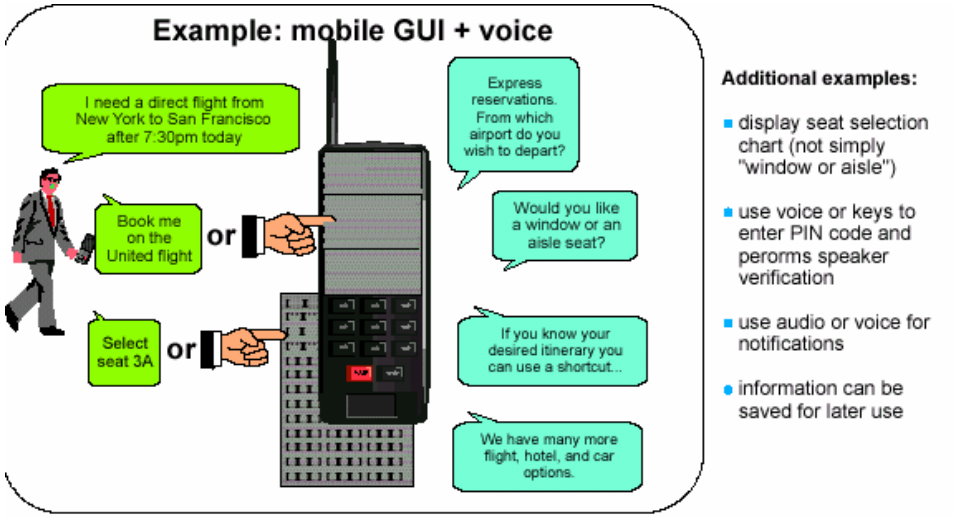

Figure 4 – Example of a multimodal usage scenario

Applications may be designed so that the user can at any time select the most appropriate modality (/ies) or device(s). Typically, in such cases, services make supplementary use of different modalities or devices: the same interactions (input or output) can be achieved in each modality or device stand-alone.

It is also possible that the application is designed to impose on the user what modality (/ies) or interactions must be used for a particular interaction. Typically, in such cases, services make complementary use different modalities or devices: the interactions available to the user differ per modality.

Eventually, it is possible that complementary or supplementary use of the application is (dynamically) decided by the user (preferences), service provider or application design.

In general the choice for a particular modality or device is motivated by:

- The available modalities or devices
- The nature of the interaction (what kind of input or what kind of output, can it be publicly displayed etc…)
- The situation (what is the user's activity (driving, sitting, walking, …), the environment (noisy, private, public, …) etc.
- The need to circumvent a modality to correct or overcome a limitation of that modality (e.g. repeating or blocking speech recognition mistakes or handwriting recognition mistakes, interactions that have no or very little meaning in a particular modality or can not be executed on a particular device – like drawing an input by voice…).

Multimodal and multi-device services may differ by the granularity of the multimodal synchronization. The multimodal synchronization granularity defines the way that an input (or output update) in one modality is reflected in the output in another modality or device as well as way that it may be combined across modalities.

The following levels of multimodal synchronization granularity can be considered:

- **Event-level synchronization:** User inputs (and outputs) in one modality are captured at the level the individual DOM events and immediately reflected in the other modality.
- **Form-level input synchronization**: Inputs (and outputs) in one modality are reflected in the other only after reaching a particular point in presentation (e.g. completing a certain amount of fields in a form).
- **Field-level input synchronization**: Inputs (and outputs) in one modality are reflected in the other after the user changes focus (e.g. move from input field to input field) or completes the interaction (e.g. complete a select in a menu).
- **Page-level**: Inputs (and outputs) in one modality are reflected in the other only after submission of the page.

In addition, the following levels should also be considered:

- **Event-level** input synchronization with output media (e.g. SMIL 2.0 [SMIL])
- **Media synchronization:** Synchronization between output media (e.g. as specified by SMIL 2.0 [SMIL])
- **Session level:** Suspend and resume behavior; an application suspended in one modality can be resumed in the same or another modality. This includes the notion of **sequential** multimodal or multi-device interactions, where one modality or device is active at a time for input (i.e. available for input and/or output or used for input and/or output).

Inputs can be:

- Sequential: only in one modality at a time.
- Limited to one modality per turn.
- Composite: the interaction events can come from different modalities and they can be combined into a single input for the turn): e.g. user says "open this", while selecting on a screen.

There are different ways of navigating through an application:

- Machine driven (e.g. IVR (Interactive Voice Response System) driven dialog through a pre-established sequence of menus).
- User-driven (e.g. GUI application)
- Mixed initiative: user as well as system can drive the interaction in a natural manner. Typically, the system is also able to derive the intent of the user and combine inputs in a same turn based on the inputs, context (i.e. state of the application, history of state, inputs and output) and external knowledge sources (e.g. other application or information at the level of the business logic (e.g. are there flights that match such a query; does such a date exist and make sense (e.g. a return date before departure date would not make sense)).

In addition, multi-modal applications offer a simple and convenient solution to the problem of accessibility to applications and services that can be considered as an alternative or complementary approach to the adoption of a particular authoring approach like device-independent authoring [W3CDI]. Users with disabilities may rely on the following usage scenarios:

- When the services make supplementary use of the different available modalities or devices, it is possible for a disabled user to select to use only or mostly a particular modality (e.g. voice for a eye impaired user or GUI for a

hearing impaired user, etc…): multi-modal and multi-device services can be made implicitly accessible and therefore comply with the upcoming accessibility regulations.

- If the user possess a personal device tuned to his preferences or that enable him to overcome some disabilities, he or she can use this device to interact with the a terminal and services in multi-device mode; where the dedicated device acts as remote control for the terminal. A particular implementation is illustrated in Figure 5.

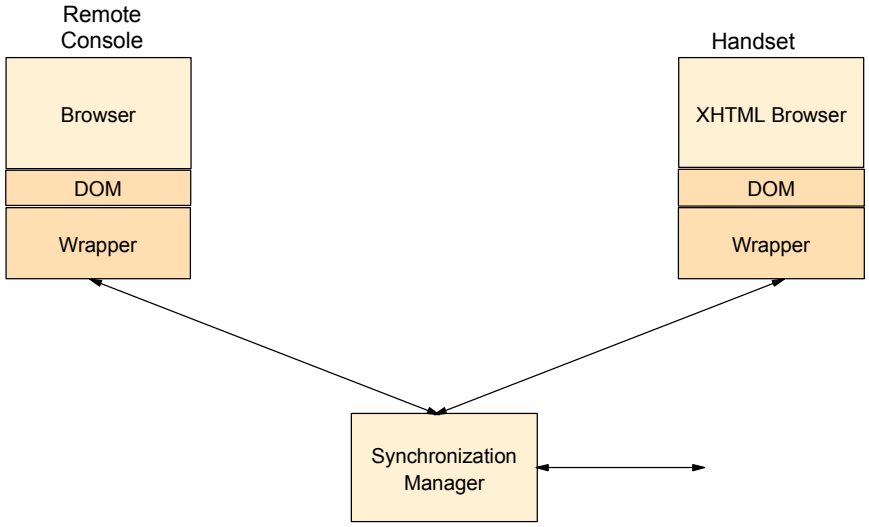

Figure 5 – Accessibility to wireless applications on handset or other terminal / UE using a dedicated remote console in multi-device configuration. The figure illustrates a possible deployment architecture.

5.4 System configurations

5.4.1 Modular and interoperable systems

Users can access and interact with mobile services through several devices or modalities.

The available modalities and devices as well as the user preferences may change from one session to another. They may also dynamically change during a same session.

Some modalities (e.g. speech, handwriting recognition, video input recognition etc…) and capabilities (e.g. conversational dialog management), may require distribution of:

- The modality-specific presentation (i.e. user interface) (e.g. a server-side voice browser)
- The dialog management (e.g server-side synchronization and dialog management)
- The processing of input and output (e.g. server-side speech engines)

For example, speech-driven applications include simple terminal based applications like voice dialling and command and control applications with limited vocabularies that facilitate the speech recogniser to be implemented solely in the terminal. However, more demanding applications like dictation, time-table enquiry systems, street navigation etc require complex speech recognition systems that would need lots of memory and computational resources - items that are scarce in today's portable devices. Hence these applications require part or whole of the speech recognition process to be carried out in the network, which can accommodate bigger and more complex computational devices. Other considerations may also lead to prefer using server-side processing: when the task is too complex for the local engine, when the task requires a specialized engine, when it would not be possible to download the speech data files (grammars etc…) without introducing significant delays or taking too much bandwidth or when intellectual property (e.g. proprietary grammars), security or privacy considerations (e.g. it would be inappropriate to download a grammar or vocabulary file that contains the names of the customers of a bank or the password grammars) make it not appropriate to download such data files on the client or to perform the processing on the client.

Similar considerations may influence the location of modality specific browsers and the synchronization and dialog management components.

Several configurations may adequately support a particular service. The need to support different multimodal or multi-device services may require support for several configurations.

On the other hand, with mobile applications, it may also be important to support some capabilities in disconnected mode.

In general, fat terminals with most of the functions (except some input and output processing) located on the terminal will be of great interest. However, thin client solutions, with the ability to use existing mobile terminal with minimal feature additions will be key to the success of mobile services and to a wide adoption of multimodal and multi-device services. In such cases some of the modality, processing or management is performed on the server.

It is also possible to distribute the processing or the modality or device-specific presentation across several devices. This may also dynamically change in time and may or may not involve server modules.

The different configurations and functions is a function of the devices / channels, tasks and environment. Configurations may require registration and negotiation or dynamic provisioning of the device when accessing a multimodal or multi-device service. It also depends on having appropriate mechanisms to query, or examine the device capabilities and configuration modes. This is related to and should be considered by on-going work items and specification work on:

- Device management
- UAprof.

6. Requirements (normative)

The requirements do not assume or imply the technology or implementation of the requirements.

6.1 High level requirements

6.1.1 User-centric high level requirements

- A user MUST be able to use several modalities to interact with the same application (UC 5.2.2.2, UC 5.2.2.5) .
- A user SHOULD be able to use several devices to interact with the same application
- When supported by the application, the user MUST be able to select the preferred mode of interaction for any particular interaction in a particular situation (supplementary use of modalities).(UC 5.2.2.2, UC 5.2.2.3, UC 5.2.2.6)
- The user MUST be able to switch back and forth between the modalities supported by a service while in a session. (UC 5.2.2.2)
- The user MUST be able to control the availability of devices and modalities for the service, when supported by the application. (UC 5.2.1.7)
- Access to multimodal and multi-device services MUST NOT compromise security and privacy.
- Delays in presentation updates as a result of a user interaction MUST be as small as possible.
- The multimodal service specification MUST support the following modalities:
    - GUI (Screen, keyboard or keypad) (UC 5.2.2.2, UC 5.2.2.3)
    - Voice (UC 5.2.2.2, UC 5.2.2.3)
    - TTY
- The multimodal service specification SHOULD support:
    - Handwriting and gestures
- Extensibility for other modalities SHOULD be provided in order to cope with emerging interaction modalities.
- The supported configurations MUST include multi-device support. (UC 5.2.1.7)

6.1.2 Service-centric high level requirements

- Multimodal and multi-device services SHOULD support a wide range of multimodal capabilities (synchronization granularity, available modalities and devices, etc…).
- Multimodal and multi-device services SHOULD be adaptive to the multimodal and multi-device capabilities. (UC 5.2.1.7)
- It MUST be possible to deploy services that make complementary use of modalities or devices and therefore impose to the user the modality to use for a particular interaction in a particular situation. (UC 5.2.2.2, UC 5.2.2.3)
- Depending on the user's preferences or network provider settings, the multimodal or multi-device services MUST support automatic configuration established upon connection to the network or manual requests of configuration by the user. (UC 5.2.2.1, UC 5.2.2.6)
- The deployment of declarative multimodal or multi-device services SHOULD support the programming model developed by the W3C MMI working group. However, it MAY be optimized for mobile deployments.

- Until such a specification is produced, user agents MAY interoperate with services authored with the different authoring approaches considered by the W3C MMI.
  - It is however recommended to aim at supporting the W3C MI specification as main target.
  - If a W3C MMI authoring specification is unavailable or if it is inadequate to support the requirements enumerated in the present document, a multimodal and multi-device authoring specification MUST be provided.
  - It is hoped that this would be limited to the work on the specification of a profile.

- Non-declarative multimodal and multi-device services MAY be supported.

6.1.3 OMA-Centric high level requirements

- User agents SHOULD be able to interoperate with multimodal or multi-device services across a wide range of configurations:
  - Thin client configurations, where the user agent that support other modalities is on the server
  - Fat client configurations, where the user agent supports other modalities and is local to the terminal; (e.g. local speech recognition).
  - Hybrid or Distributed Configuration. Processing and functionality is distributed on the client and a remote server (e.g. DSR).
  - Configurations with multiple devices (fat client, thin client, or distributed configurations).
- Hybrid or distributed configurations MUST support the highest QoS for input and output interaction exchanges (multimodal exchanges).[1]
- Multimodal exchanges MUST be secure when distributed.
- A user MUST be able to move from local multimodal services (e.g. on fat clients) to remote multimodal services (UC 5.2.2.4) and vice versa[2].
- Modality interfaces to the modalities available on the device MUST be specified for mobile devices to enable download of multimodal applications (UC 5.2.2.5)
- Device capabilities descriptions (e.g. modality interfaces or supported configurations) MUST be specified for mobile devices to enable download of multimodal applications.

6.2 Security

In the case of distributed multi-modal or multi-device browser, the exchange of interaction events and client manipulation raise the following security issues that must be addressed:

- Interaction events and presentation manipulations can be intercepted by unauthorized third parties. This would enable reconstruction of the complete interaction with the application; especially in between submits to the backend. Any exchanged information such as temporarily selections etc would be accessible!
- Unauthorized third parties may be able to issue presentation manipulations that would affect the user agent.

Of course as for other application, authentication of users is expected to be required to allow access to confidential data (UC 5.2.2.4)

[1] These issues should be investigated by the Work Item on Quality of Experience.

[2] This requirement is to be understood as a MUST in between transaction and a SHOULD within transactions.

For example, security MUST NOT be compromised by multimodal synchronization or remote engine processing in distributed configurations.

Addressing this requirement MAY involve:
- Confidentiality and integrity of the exchanged information.
- That presentation manipulation are accepted only from trusted or authorized parties.
- Leverage of underlying network bearer security mechanisms.

In the case of client-based speech engines, additional security (and privacy) issues arise when the application is downloaded from a third party service provider:

- The speech data files (acoustic model grammars, language models, vocabularies, NL parser data files, etc…) sent to the client may contain proprietary or sensitive information (e.g. passwords, list of customers and associated input information, propprietary grammar, …).
    - The data files may be intercepted by un-authorized third parties or tampered with in the UE.
        - This may relate to the Digital Right Management work items.
- Results of some client-side engine sent across the network can be tampered with or intercepted on the UE or when transmitted.

These issues MUST be addressed by appropriate mechanisms or by requiring server-side engines when needed.

6.3 Privacy

Multimodal and multi-device privacy requirements MUST be at least as good as for other mobile services or voice sessions:

- It MUST be possible to encrypt modality-specific and multimodal synchronization exchanges.
- It MUST be possible to prevent exchange of the user's true identity, location and other terminal or user related information when required.

Multimodal and multi-device services may enable the service provider to collect information about the user or usage. This information should be treated according to the policies in place for data and voice (e.g. human to operator or human to automated service) services. Therefore, multimodal or multi-device services MUST NOT add additional privacy risks.

Also, interaction events enable reconstruction of the complete interaction with the application, including in between submission to the backend and therefore possibly beyond the knowledge or control of the user. This information or aspect of it may be considered as private by the user. Therefore, the multimodal synchronization SHOULD be associated to mechanisms that let the user specify the use that can be done of the information. Multimodal and multi-device services MUST produce similar schemas or mechanisms to describe their handling and use of the information or allow automation of the acceptance of privacy policies.

Trust and resolution mechanisms MUST be provided to enable the user to accept the particular service and configuration on the basis of the usage that will be made of such information or the management options provided to the user.

Privacy of user SHOULD NOT be threatened when exchanging speech data files (or other modality-specific data files) across the wireless network or by storing them on devices.

6.4 Charging

The charging of multimodal and multi-device services MUST support a variety of business models: the user can be charged for sessions with the speech or multimodal/multi-device service in a variety of ways.

In particular, the following charging models MUST be possible:
- By duration of session (including “one-off” charge/flat rate)
- By data volume transferred (number of packets) or other similar criteria.
- By subscription fees for the service (unlimited usage or unlimited usage up to a point and then per-use fees)
- Free (e.g. with the service being subsidised by advertising revenue from advertisement spots).

Speech or multimodal/multi-device services MUST be available to pre-paid and post-paid subscribers.

6.5 Administration

It MUST be possible for the provider of the multimodal or multidevice service to control access.

It MUST be possible that multimodal and multi-device services are provisioned by the network operator as well as third party service providers.

The administration of the multimodal and multi-device services (authorization, deauthorization, registration, deregistration, activation, deactivation, configuration, optimization) MUST be under the control of the network operator when needed and available to third party providers when authorized.

6.6 Terminal devices and smartcards

6.6.1 Terminal devices

- Local (e.g. fat client configurations) multimodal configurations depend only on the terminal; except when they rely on remote speech engine functions (or other modality-specific processing). Such implementations may not require any standardization other than support of the authoring format. However, terminal with fat client multimodal configuration capabilities MAY also support thin client configurations.[3]
- Terminals that are to support distributed configurations MUST support the multimodal synchronization interfaces and protocols.
- Devices MUST provide information about their multimodal capabilities; possibly in a given configuration (UC 5.2.2.6):
    - This MAY be in answer to a specific external query about the device capabilities.
    - This MAY be provided automatically by the device with all request or once / periodically within a session.
    - This may be provided by the device automatically upon changes to the device capabilities.
- The specifications SHOULD support connected/disconnected mode of operation where the device is sometimes connected through the network and sometime disconnected.
    - The specifications MAY support informing the user of the available (depending if the device is connected or disconnected) devices or modalities when picking the most appropriate mode of interaction.
    - It SHOULD be possible to seamlessly adapt the multimodal and multi-device behaviour based on the status of the network connectivity.
    - The specifications SHOULD support seamless switch between connected and disconnected mode and adaptation of the user interface.
- Devices that allow the execution of downloaded multimodal applications MUST provide modality interfaces to the modalities available on the device to multimodal applications (UC 5.2.2.5).

6.6.2 Smartcards

No requirement has been identified directly specific to smartcards other than the needs to:

- Identify the user and multimodal service subscription information
- Use, manage and provision user profile and preference

Such usages are not expected to introduce any new requirements on the smartcards.

[3] This issue also relates to the Work Item on Browser enhancement that should enable of multimodal and multi-device capabilities when considering the enhancements: e.g. synchronization interfaces, support of DOM, XML events, XForms, etc…

6.7 Platforms

- The specifications MUST support distributing the different components between the devices and/or servers.
- The specifications SHOULD support dynamic changes of configuration and change of the location of the synchronization management.
- The specifications MUST support dynamic provisioning of devices / channels to support a particular configuration when accessing a multimodal or multi-device service.
- Multimodal and multi-device services MUST be able to use a default or MAY advertise and negotiate a default or specific synchronization granularity level or coordination capability.
  - Default levels of synchronization granularity and coordination capabilities and behaviours MUST be specified.
  - Behaviours in case of incompatible capabilities MUST be specified.
  - The specifications MUST support the exchanges of such mechanisms.
- Multi-modal synchronization interfaces and protocols MUST enable handling and possible distribution of interaction events from one or multiple modalities or devices.
  - E.g. Interaction events MAY be represented as DOM events and XML events (distributed when needed)
- Multi-modal synchronization interfaces and protocols MUST enable manipulation, possibly remotely, of the presentation associated to each modality or device.
  - E.g. Presentation manipulation MAY be implemented via XML events and DOM (remote when needed) instructions.

6.8 Network interfaces

- Multimodal and multi-device services SHOULD be supported by Packet Switched Data networks.
- Multimodal and multi-device services MAY be supported by Circuit Switched Data networks.
- Remote speech engines MAY rely on a Speech Recognition Framework as defined by 3GPP [SRF3GPP] and Speech Engine Remote Control as defined by IETF [SPEECHSC].
- Multimodal synchronization protocols MUST support mechanisms to guarantee security and privacy as discussed in sections 6.2 and 6.3.
- The multimodal synchronization protocols SHOULD support mechanisms that enable the different components to identify and process input events and/or output manipulations and switch between modalities.
- Multimodal synchronization MUST have access to the highest possible quality of service available from the underlying network in order to minimize delays in the synchronization and confusing user interface behaviours.
- Multimodal synchronization protocols SHOULD provide mechanisms that enable handling synchronization delays introduced by the network.
- The infrastructure MUST support exchanges of device / channel capabilities (possibly in a given configuration) with other components:
  - This MAY be in a query response mode
  - This MAY be initiated by the device or server

6.9 Usability

- The use of device or modalities to interact with multimodal and multi-device services MUST comply with the user's preferences and/or environment; based on the available devices and modalities (UC 5.2.2.6).

- The specifications SHOULD support selection or changes of the configurations of the systems that are supported by the device and user agents.
    - It MUST enable dynamic re-configuration (e.g. adding or removing a device or modality; modifying the role of a device (e.g. synchronization management); changing usage preferences or multimodal synchronization granularity) during a session or while accessing a particular multimodal or multi-device service.
    - Distributed multimodal and multi-device configurations MUST support registration and configuration mechanisms.
        - At initiation or during a session
        - When re-configuring or disconnecting devices
- The user MUST be able to control/decide the set of the multiple devices or modalities combination and configure the system, within the limits of the configurations supported by a particular multimodal or multi-device service.
    - It MUST be possible for the user to determine what are allowable configurations and settings (coordination capability, supplementary versus complementary, etc) supported by a multimodal/multidevice service.
    - The user MUST be able to determine easily the current configuration and settings (e.g. multimodal synchronization granularity).
    - The user MUST be able to add or remove new modalities or devices or (re)configure them prior to accessing a multimodal and multi-device service or while interacting with it.
- The user SHOULD be able to initiate access to a multimodal or multi-device service from any of the modality or device; if supported by the authoring of the service.
- Access to a multimodal or multi-device service MAY trigger automated or manual configuration of the system..
- Users of multimodal and multi-device services MUST be able to initiate multimodal and multi-device sessions from one or multiple devices by providing (entering, selecting, speaking, etc) an addressing scheme.
- Conflicting interpretations of the user input (conflicting interaction events, dialog management issues with conversational multimodal applications) MAY be reported and left to the user to resolve, unless if specified otherwise by user profile.
- User interactions with multimodal and multi-device services MUST be secure and provide privacy guarantees:
    - Unsafe interactions (depending of nature of services, trust model, advertised privacy policy etc…) MUST be notified prior to execution
    - Whenever possible, more secure configurations or settings MAY be proposed to the user.
- Multimodal and multi-device systems SHOULD be able to access monochannel applications, provided that the session involves a user agent that supports the corresponding application / presentation format.
- Supplementary multimodal and multi-device services SHOULD be accessible as monochannel applications.
- The user MUST be able to select the most appropriate modality or device for a particular type of interaction at a particular moment and in a particular situation, when this is supported by the multimodal and multi-device service (e.g.. when the application supports supplementary use of modalities or devices).
- User SHOULD be able to use another modality to correct errors in one modality or device.

6.10 Interoperability

- The specification MUST address the following issues: “interoperability, extensibility and longevity”..
    - Across configurations:
        - For example, a same server middleware, user agent or device can be used in different configurations.
    - Across capabilities (levels of synchronization granularities, conversational capabilities, network capability, new modalities, …) of the runtime and infrastructure.
        - For example, the same server middleware, user agent or device can be used to support different capabilities based on the capability of the network and terminal or based on the quality of service available from the network.

        - Multimodal and multi-device services MUST be able to support sequential multimodal or multi-device interactions for example for deployment on network or access through terminal that do not support voice and data simultaneously. In this case, on-going interaction with an application in one modality or on one device can be suspended and resumed on another device or using another modality.

    - Across different authoring approaches and features of the multimodal and multi-device services:

        - For example, a same server middleware, user agent or device can be used with different authoring approaches or features of multimodal and multi-device services.

- The end-to-end specifications MUST support the widest set of relevant configurations of user agents to access multimodal and multi-device services.

- Multimodal and multi-device services MUST be interoperable across configurations.

- Multimodal and multi-device services MUST be interoperable across different level of the synchronization granularity.

- Multimodal and multi-device services MUST be interoperable across different type of navigation and dialog management capabilities.

- The multimodal and multi-device specifications MUST be consistent with specifications produced by other related standard bodies.

- Multimodal synchronization protocols MUST be able to accommodate user agents that render declaratively programmed multimodal and multi-device services as well as non-declaratively programmed multimodal and multi-device services; if such services are introduced.

6.11 Application development

- There MUST be authoring language specifications and guidelines to develop multimodal and multi-device services. These specifications and guidelines MUST be consistent with the one produced by W3C MMI.

- These multimodal authoring language specifications SHOULD specify all what author need to write multimodal and multi-device services and applications that rely on them.

- The multimodal and multi-device authoring language specifications MUST support the deployment of applications with the wide range of multimodal capabilities discussed in this document.

    - Supplementary and complementary user interfaces

    - Different deployment configurations

    - Different multimodal synchronization granularities

    - Different modalities, devices

    - Dynamic changes of some of the above

- The multimodal and multi-device authoring language specifications MUST let the author develop applications customized for a particular set of capabilities.

- The multimodal and multi-device authoring language specifications MUST let the author develop applications written for a range of capabilities.

    - The author SHOULD then be able to also customize for particular capabilities.

- The multimodal and multi-device authoring language specifications MUST let the author impose or forbid configurations or capabilities and impose behaviour. E.g.:

    - Impose a modality switch

  - Indicate a system-initiated reconnect (e.g. system initiated voice call or GUI push)
  - Indicate where and what particular type of processing should take place (e.g. input processing by speech recognition engine or output processing by Text-to-Speech engine).
  - Program such processing.
  - Distribute at will the processing or indicate how the distribution is selected.

- The multimodal and multi-device authoring language specifications MUST let the author indicate to the system that the application relies on the behaviour provided by default by the deployment conditions.
- Declarative multimodal and multi-device authoring language specifications MUST be available
- Non-declarative multimodal and multi-device authoring language specifications SHOULD be available.
- There MUST be authoring language specifications and guidelines to develop multimodal and multi-device applications for the download to devices and for local execution on the user's device (see UC 5.2.2.5).

Appendix A. Change History (Informative)

A.1 Approved Version History

Reference	Date	Description
n/a	n/a	No prior version

A.2 Draft/Candidate Version 1.0 History

Document Identifier	Date	Section	Description
First Verions (0.1)	9/18/02	All	.First verion: OMA-REQ-2002-0067-Reqs-v0.1-20020918_Multimodal_Multi-device_Services-d
Version 0.2	10/27/02	All	Retention only of the Requirement aspects.
		All	Editorial version / dates / doc number updates
		Section 5.3.2	Removed (more adapted to a TR or output spec)
		Appendix A	Removed (more adapted to a TR or output spec)
		Sections 6, 6.1.3, 6.6.1, 6.7, 6.8, 6.10	Remove mentions of configurations described in section 5.3.2 and execution model described in appendix A
		Sections 6.7 and 6.9	Editorial updates (synchronization manager changed into synchronization management and re-phrasing of plausible technology directions.
Version 0.3	11/10/02	Sections 6.1.2 and 6.11	Additional authoring requirements
		All	Editorial and numbering changes
Version 0.4	12/31/02	Sections 2. 4.1 4.2 5.2 6. 6.1.2 Appendix A	 -Added references - Editorial updates - Editorial updates - Added reference to Arch framework MMI use case - Editorial updates - Editorial updates - Added section on Voice-only and DI authoring requirements
Version 0.5	1/24/03	All	Submitted to OMA REQ for review at Long Beach OMA Meeting. • Updates as prescribed in OMA-REQ-2003-0051-Review_report_Multimodal_Multi-device_reqs_20030123. • Additional editorial updates • Status review report: • OMA-REQ-2003-0080-Review_report_Multimodal_Multi-device_reqs_20030124_status_V0_2
Version 0.6	2/22/03	All	Submitted to OMA REQ for review at Cambridge OMA REQ Face-to-Face meeting • Address all the comments from Nokia (Feb 4, 2003 – Mika J. Laitinen) raised about version 0.5.
Version 0.7	5/31/03	Section 5	• Updates to reflect inputs: o OMA-REQ-2003-0228-MMReqInput1.zip o OMA-REQ-2003-0228r1-MMUI_UC_L0.zip o OMA-REQ-2003-0318-MMUI_UC_L0.zip
Version 1.0	6/1/03	Numbering	• Version 0.7 is renumbered as Version 1.0
	8/29/03	5.2.2	Incorporate OMA-REQ-2003-0365-MMUseCases.
	9/10/03	All	Incorporate OMA-REQ-2003-0532R01
	9/28/03	All	Update of the version agreed by REQ at the OMA Berlin Meeting to the latest OMA_RD template
Version 1.1	11/5/03	All	Update to accommodate R&A comments

Document Identifier	Date	Section	Description
	11/9/03	All	Update to accommodate IBM comments to OMA-REQ-2003-0696 as discussed on REQ mailing list.
	11/12/03	5 , 6.11	Update to accommodate Siemens comments
Candidate Version OMA-RD-Multimodal_Multi-device_Services-V1_1	13 Nov 2003	n/a	Status changed to Candidate by TP TP ref # OMA-TP-2003-0624-LATE-Change-Request-for-Multimodal-RD

A.2.1 Document History

Document History	
OMA-RD- Multimodal_Multi-device_Services-V1_1-20031112-D	• Editorial indent changes for some headers in section 5 • Changes of authoring specification to authoring language specifications in section 6.11
OMA-RD- Multimodal_Multi-device_Services-V1_1-20031109-D	• Reconcile proposal with IBM comments to OMA-REQ-2003-0696.
OMA-RD- Multimodal_Multi-device_Services-V1_1-20031105-d	• Incorporates proposal for changes following the CRs and Comments received during R&A of OMA-RD-Multimodal_Multi-device_Services-V1_0-20030928-d: o OMA-REQ-2003-0696 o Telefonica comments o IBM comments
OMA-RD-Multimodal_Multi-device_Services-V1_0-20030928-d	• Update of the version agreed by REQ at the OMA Berlin Meeting to the latest OMA_RD template
OMA-RD-Multimodal_Multi-device_Services-V1_0-20030910-D	• Incorporated OMA-REQ-2003-0532R01 • Agreed by OMA-REQ in Berlin on 10 Sept 2003.
OMA-REQ-2003-0525-v1_0-20030829_Multimodal_Multi-device_Services-d	• Incorporates OMA-REQ-2003-0365-MMUseCases as agreed at Atlanta Face To Face.
OMA-REQ-2003-0374R1-v1_0-20030531_Multimodal_Multi-device_Services-d	• Renumbered as V1:
OMA-REQ-2003-0374-v0.7-20030531_Multimodal_Multi-device_Services-d	• Updates to reflect inputs: o OMA-REQ-2003-0228-MMReqInput1.zip o OMA-REQ-2003-0228r1-MMUI_UC_L0.zip o OMA-REQ-2003-0318-MMUI_UC_L0.zip
OMA-REQ-2003-0080R1-v0.6-20030222_Multimodal_Multi-device_Services-d	• Updates to address Nokia's comments dated Feb 2, 2004.
OMA-REQ-2003-0080-v0.5-20030124_Multimodal_Multi-device_Services-d	• Updates based on Dallas REQ FTF review: o Address most of the issues identified in OMA-REQ-2003-0051-Review_report_Multimodal_Multi-device_reqs_20030123 • Additional inputs collected in between. • Presented to Long Beach REQ
OMA-REQ-2003-0031-v0.4-20021231_Multimodal_Multi-device_Services-d	• Third draft based on OMA REQ and OMA MAE comments • Submitted to Dallas OMA REQ
OMA-REQ-2002-0136-R1-v0.3-20021110_Multimodal_Multi-device_Services-d	• Update based on received comments: additional requirements.
OMA-REQ-2002-0136-v0.2-20021027_Multimodal_Multi-device_Services-d	• Second Draft as action item from OMA Requirements meeting in Sophia Antipolis, October 1-2, 2002. • Input to OMA REQ for Hawaii meeting (October Nov 10 to 15, 2002); presented for input from OMA REQ group. **Editor:** Stéphane H. Maes, stephane.maes@oracle.com
OMA-REQ-2002-0067-Reqs-v0.1-20020918_Multimodal_Multi-device_Services-d	• First Draft for OMA Requirements meeting in Sophia Antipolis, October 1-2, 2002. Proposed as input document. • This document with some extensions is the staring point of the MAE requirement activity on Multimodal and Multi-device services (OMA-Multimodal_Multi-device-REQS-V0.2-20021007-d) **Editor:** Stéphane H. Maes Sources: IBM, Motorola, Oracle

Appendix B. (Normative)

B.1 W3C Multimodal Interaction Activity

- The OMA multimodal and multi-device specifications SHOULD follow the W3C Multimodal Interaction Requirements [W3CMMI].

B.2 Voice Applications

As illustrated in the present document; Voice applications are of great interest for mobile applications; especially as voice-only or sequential (suspend and resume) multimodal applications.

- The OMA multimodal and multi-device specifications SHOULD include:
  - Voice-only enablers:
    - Server-side
    - Client-side
    - Hybrid
  - Sequential enablers
    - Server-side
    - Client-side
    - Hybrid
- The OMA multimodal and multi-device specifications SHOULD include
  - Authoring of Voice-only applications
  - Authoring of sequential voice / GUI applications

  This is in addition to generic multi-modal authoring as discussed in section 6.11.

- Authoring of Voice-only applications MUST be interoperable with existing Voice standards:
  - VoiceXML for Voice-only,
  - W3C MMI authoring specifications (if any) for multimodal.

B.3 Device-Independent Authoring

As discussed in section 6.11, authoring of application to be served to a wide range combinations of devices and modalities with different synchronization capabilities require the development of recommendations for device / channel-independent applications; designed to author from the onset for a wide range of devices and modalities.

- The OMA multimodal and multi-device specifications SHOULD address device-independent authoring.
  - Such specifications SHOULD be compatible with W3C work (HTML, XForms, CSS, DI).

This capability would be valuable to address the wide variety of mobile devices; independently of multimodal considerations.