Draft 3G TS 22.243 1.5.0 (2002-6)

Technical Specification

S1-021277

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects
Service aspects; Stage 1
Speech Recognition Framework for Automated Voice Services
(3G TS 22.243 Release 6)

Deleted: 4
Deleted: 5
Deleted: 1
Deleted: Distributed
Deleted: Based
Deleted: 5

The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.

The present document has not been subject to any approval process by the 3GPP Organisational Partners and shall not be implemented.
This Specification is provided for future development work within 3GPP only. The Organisational Partners accept no liability for any use of this Specification.
Specifications and reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organisational Partners' Publications Offices.

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: 1

Keywords

UMTS, service, user profile

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

Copyright Notification

No part may be reproduced except as authorized by written permission
The copyright and the foregoing restriction extend to reproduction in all media.

© 2002, 3GPP Organizational Partners (ARIB, CWTS, ETSI, T1, TTA, TTC).
All rights reserved.

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65Release 6

Inserted: Release 65

Contents

Contents 3

Foreword 4

Introduction 5

1. Scope 6

2. References 6

2.1 Informative References 7

3. Definitions and abbreviations 7

3.1 Definitions 7

3.2 Abbreviations 8

4. Requirements 9

4.1 Initiation 9

4.2 Information during the speech recognition session 9

4.3 Control 9

4.4 User Perspective (User Interface) 10

5 UE and network capabilities 10

6. Administration 10

6.1 Authorization 10

6.2 Deauthorization 11

6.3 Registration 11

6.4 Deregistration 11

6.5 Activation 11

6.6 Deactivation 11

7. Service Provisioning 11

8. Security 11

9. Privacy 12

10. Charging 12

11. Roaming 12

12. Interaction with other services 12

Annex A (informative): Speech recognition Framework-based automated voice service examples 13

Annex B (informative): Change History 14

Deleted: Contents 3¶
Foreword 5¶
Introduction 6¶
1. Scope 7¶
2. References 8¶
2.1 Informative References 8¶
3. Definitions and abbreviations 8¶
3.1 Definitions 8¶
3.2 Abbreviations 9¶
4. Requirements 10¶
4.1 Initiation 10¶
4.2 Information during the speech recognition session 10¶
4.3 Control 11¶
4.4 User Perspective (User Interface) 11¶
5 UE and network capabilities 11¶
6. Administration 11¶
6.1 Authorization 11¶
6.2 Deauthorization 12¶
6.3 Registration 12¶
6.4 Deregistration 12¶
6.5 Activation 12¶
6.6 Deactivation 12¶
7. Service Provisioning 12¶
8. Security 12¶
9. Privacy 13¶
10. Charging 13¶
11. Roaming 13¶
12. Interaction with other services 13¶
Annex A (informative): Speech recognition service examples 15¶
Annex B (informative): Change History 16¶
Contents 3¶ ... [1]

Inserted: Contents 3¶
Foreword 5¶
Introduction 6¶ ... [2]

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65Release 6

Inserted: Release 65

Foreword

This Technical Specification has been produced by the 3GPP.

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of this TR, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the specification.

Introduction

Forecasts show that speech-driven services will play an important role on the 3G market. People want the ability to access information while on the move and the small portable mobile devices that will be used to access this information need improved user interfaces using speech input. At present, however, the complexity of medium and large vocabulary speech recognition systems is beyond the memory and computational resources of such devices. Also associated delay to download speech data files (e.g. grammars, acoustic models, language models, vocabularies etc..) may be prohibitive. Eventually, it may not always be acceptable for the speech service providers to allow download of these speech data files if they contained confidential information (password (security issue), customer names and address (privacy issue)) or intellectual properties; for example a well crafted speech grammar is often considered by speech service providers as a trade secret.

Server-side processing of the combined speech and DTMF input and speech output, can overcome these constraints by taking full advantage of memory and processing power as well as specialized speech engines and data files. However, the distortions introduced by the encoding used to send the audio between the client and the server as well as additional network errors can significantly degrade the performance of the speech engines; therefore also limiting the achievable speech functionalities. A server-side speech service is generally equivalent to a phone call to an automatic service. As for any other telephony service, DTMF is a feature that should always be considered as needed.

This document describes a generic speech recognition framework to distribute the audio sub-system and the speech services by sending encoded speech and meta-information between the client and the server using IMS. Instead of using a voice channel as in today's server-based speech services, an error-protected data channel will be used to transport encoded speech from the client audio sub-system (terminal client) to remote speech engines (on server) for processing (e.g. speech recognition, speaker recognition,). The speech recognition framework will also enable downlink data streaming of voice and recorded audio prompt generated by server to the terminal client audio subsystem. The speech recognition framework may use conventional codecs like AMR or Distributed Speech Recognition (DSR) optimized codecs. Appropriate uplink DSR codecs optimized for speech recognition may alleviate some of the problems identified above.

The speech recognition framework will provide users with a high performance distributed speech interface to server-based automatic speech services with communication, information access or transactional purposes.

The types of supported user interfaces include those that are voice only, for example, automatic speech access to information, such as a voice portal described in this section. These typically support combined speech or DTMF input.

In the future, a new range of multi-modal applications is also envisaged incorporating different modes of input (e.g. speech, keyboard, pen) and speech and visual output.

A 3GPP speech recognition framework enables the use of conventional codecs (e.g. AMR) or DSR optimized codecs to distribute in the network the speech engines that process speech input or generate speech output. It includes:

- At least a default uplink DSR optimized codec specification.
- A stack of speech recognition protocols (based on SIP/SDP and RTP and consistent with IMS SIP sessions) to support:
  - Establishment of uplink and downlink sessions, along with codec negotiation (using SIP/SDP)
  - Transport of speech recognition payload (uplink) with conversational QoS (RTP)
  - Support of transport (also at conversational QoS) of meta-information required for the deployment of speech recognition applications between the terminal and speech engines (meta-information may include terminal events and settings, audio sub-system parameters and settings, etc.).
    - Examples of such meta-information could include: speech/no-speech, barge-in information, terminal events and settings

For example, ETSI STQ Aurora has developed such a DSR framework of protocols (e.g. IETF AVT DSR) compatible with 3GPP IMS (SIP, SDP, RTP) to exchange audio, control and speech meta-information as well as with the use of AMR as uplink codec.

It shall be possible to recommend a codec to be supported by default to deploy services that rely on the 3GPP speech recognition framework.

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: are
Deleted: s
Deleted: In t
Deleted: , Distributed Speech Recognition (DSR) is
Deleted: The additional use of appropriate uplink DSR codecs optimized for speech recognition overcomes the problems identified above.
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: minimize the negative effect of encoding and network losses on the speech engine performance. It
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: Examples of such meta-information could include: speech/no-speech, barge-in information, terminal events and settings, audio sub-system acquired parameters (e.g. noise level, ..) and settings, audio sub-system control and application specific exchanges

To that effect, the specifications will consider DSR optimized codecs. ETSI has published DSR optimized codecs specifications (ETSI ES 201 108 [4] & ETSI ES 202 050 [9]).

References [7,8] present early comparisons of the performances obtained with such DSR optimised encoding schemes with respect to conventional (e.g. AMR) codecs [7,8].

1. Scope

This Technical Specification defines the stage one description of the Speech Recognition Framework for Automated Voice Services. Stage One is the set of requirements for data seen primarily from the user's and service providers' points of view.

This TS includes information applicable to network operators, service providers, terminal and network manufacturers.

This TS contains the core requirements for the Speech Recognition Framework for automated voice services.

The scope of this Stage 1 is to identify the requirements for 3G networks to support the deployments of a speech recognition framework - based automated voice services and therefore to introduce a 3GPP speech recognition framework. The Speech Recognition Framework for automated voice services is an optional feature in a 3GPP system.

Figure 1 positions the speech recognition framework with respect to speech and multimodal and multidevice services.

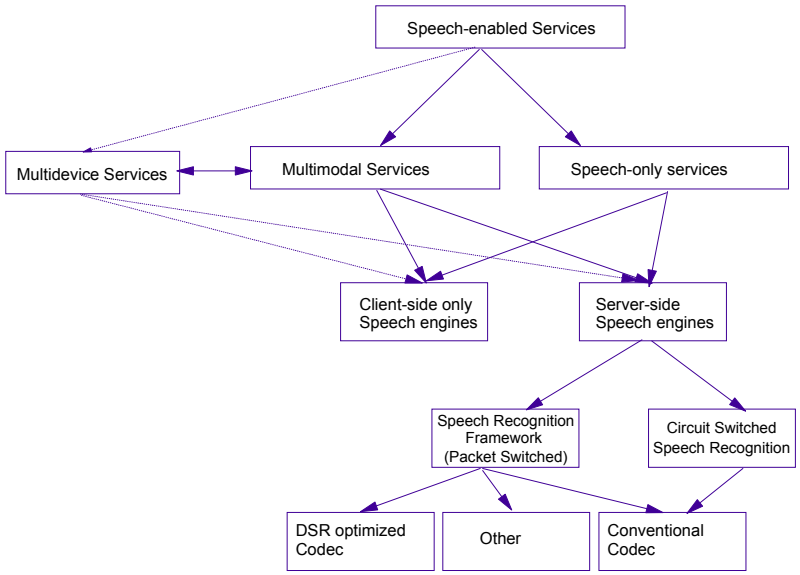

Figure 1 - It positions the speech recognition framework among the different technical solutions that can support speech enabled services.

2. References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 ... [3]
Inserted: Release 65
Inserted: Draft 3G TS 22.243 ... [4]
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: and DSR optimised codecs
Deleted:
Deleted: ¶
Deleted: Studies show that
Deleted: improves the speech ... [5]
Deleted: speech recognition spe ... [6]
Inserted: speech recognition sp ... [7]
Deleted: Distributed
Deleted: Based
Deleted: Distributed
Deleted: Based
Deleted: A
Deleted: s
Inserted: s
Deleted: V
Deleted: oice
Deleted: S
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: A
Inserted: Automated
Deleted: s
Inserted: s
Deleted: V
Inserted: Vvoice
Deleted: S
Inserted: Services
Deleted: DSR
Deleted: DSR
Deleted: S
Inserted: Speech recognition

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.

[1] 3GPP TS 21.133: "3G Security; Security Threats and Requirements"

[2] 3GPP TS 21.905: "Vocabulary for 3GPP Specifications"

[3] 3GPP TR 22.941: "IP Based Multimedia Services Framework"

[4] ETSI ES 201 108 v1.1.2 Distributed Speech Recognition: Front-end Feature Extraction Algorithm; Compression Algorithm", April 2000.

[5] 3GPP TS 22.105: "Services and Service Capabilities"

[6] 3GPP TS 22.228: "Service Requirements for IP multimedia Sub System"

2.1 Informative References

[7] AU/410/02 "Speech recognition performance comparison between AMR speech coding and the Advanced DSR Front-End (ETSI ES 202 050)", Motorola, April 2002

[8] AU/411/02 "Speech recognition performance comparison between AMR speech coding and the DSR Front-End (ETSI ES 201 108)", France Telecom, April 2002

[9] ETSI ES 202 050 v0.0.0 "Distributed Speech Recognition: Advanced Front-end Feature Extraction Algorithm; Compression Algorithm", standard selected; document in preparation.

[10] TSG T2010705: Stéphane H. Maes, "Support of Multi-modal and Multi-device browsers applications by 3GPP".

[11] TSG S1-020769: Stéphane H. Maes, Multi-modal and multi-device automated services

3. Definitions and abbreviations

Definitions and abbreviations used in the present document are listed in TR 21.905 [2]. For the purposes of this document the following definitions and abbreviations apply:

3.1 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.

DSR Optimised Codec: The module in UE which takes speech input, extracts acoustic features and encodes them with a scheme optimised for speech recognition. This module is similar to the encoder in a vocoder (such as EFR, AMR). On the server-side, the uplink encoded stream can be directly consumed by speech engines without having to be converted to a waveform.

Deleted: **Distributed Speech Recognition**: 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 typically relies on DSR codecs optimised for speech recognition: acoustic features are extracted and encoded on the terminal. ¶
Distributed Speech Recognition Based Automated Voice Service: An Automated Voice Service utilising Distributed Speech Recognition. In such a case the user voice input is captured and encoded, with a DSR codec optimised for speech recognition. The encoded speech is streamed uplink to server-side speech engines that process it. The application dialog manager generates prompts that are streamed downlink to the terminal.¶
Deleted: istributed
Deleted: peech
Deleted: ecognition

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 (ASR) or on DSR optimised codecs where acoustic features are extracted and encoded on the terminal.

Speech Recognition Framework-based Automated Voice Service: An automated voice service utilising the speech recognition framework to distribute the speech engines from the audio sub-system. In such a case the user voice input is captured and encoded, with a conventional or a DSR optimised for speech recognition as negotiated at session initiation. The encoded speech is streamed uplink to server-side speech engines that process it. The application dialog manager generates prompts that are streamed downlink to the terminal.

SRF Call: An uninterrupted interaction of a user with an application that relies on SRF-based automated voice services.

SRF Session: Exchange of audio and meta-information, explicitly negotiated and initiated by the SRF session control protocols, between terminal (audio-sub-systems) and SRF-based automated voice services. Sessions last until explicitly terminated by the control protocols.

SRF User Agent: a process within a terminal that enables the user to select a particular SRF-based automated voice service or to enter the address of a SRF-based automated voice service. The user agent converts the user input or selection into a SIP IMS session initiation with the corresponding SRF-based automated voice service. The user agent can also terminate the session with the service when the user device to disconnect.

Text-to-Speech Synthesis: A technology to convert text in a given language into human speech in that particular language.

Meta information: Non-speech data that may be required to facilitate and enhance the server-side processing of the input speech and facilitate the dialog management in an automated voice service. These may include keypad events over-riding spoken input, notification that the UE is in hands-free mode, client-side collected information, etc….

Dialog manager: A technology to drive a dialog between user and automated voice services. For example a VoiceXML voice browser is essentially a dialog manager programmed by VoiceXML that drives speech recognition and text-to-speech engines.

3.2 Abbreviations

For the purposes of this document the following abbreviations apply:

AMR – Adaptive Multi Rate

DSR – Distributed Speech Recognition

DTMF – Dual Tone Multi-Frequency

IETF – Internet Engineering Task Force

IMS – IP Multimedia Subsystem

IVR – Interactive Voice Response system

PCM – Pulse Coded Modulation

PIM - Personal Information Manager

SIP – Session Initiation Protocol

SRF – Speech Recognition Framework

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: s
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: DSR
Deleted: A
Deleted: DSR
Deleted: A
Deleted: DSR
Deleted: A
Deleted: V
Deleted: S

URI – Uniform Resource Identifier

4. Requirements

The following list gives the high level requirements for the SRF-based automated voice services:

- Users of the SRF-based automated voice service shall be able to initiate voice communication, access information or conduct transactions by voice commands using speech recognition. Examples of SRF-based automated voice services are provided in Appendix A.

The speech recognition framework for automated voice services will be offered by the network operators and will bring value to the network operator by the ability to charge for the SRF-based automated voice services.

This service shall be offered over the IMS. The protocols used for the meta information and front-end parameters (from terminal to server) and associated control and application specific information will be based on those in IMS.

4.1 Initiation

It shall be possible for a user to initiate a connection to the SRF-based automatic voice services by entering the identity of the service. Most commonly, when used as a voice service, this will be performed by entering a phone number. However, particular terminals may offer a user agent that accepts other addressing schemes to be entered by the user: IP address, URI, e-mail address possibly associated to a protocol identifier. This is particularly important for multi-modal usages.

In all cases, the terminal will convert the address entered by the user to initiate a session via the SIP IMS session initiation protocol and establish the different SRF protocols. During this initiation of the SRF session, it shall be possible to negotiate the uplink and downlink codecs. The terminal may support a DSR optimized codec as a default uplink codec.

4.2 Information during the speech recognition session

Codec negotiation during a SRF session should be optionally supported.

This may be motivated by the expected or observed acoustic environment, the service package purchased by the user, the user profile (e.g. hands-free as default) or service need. The user speaks to the service and receives output back from the automated voice service provider as audio (recorded 'natural' speech) or Text-to-Speech Synthesis. The output from the server can be provided in the downlink as a streaming service or by using conversational speech codec.

Additional control and application specific information shall be exchanged during the session between the client and the service. Accordingly some terminals shall support sending additional data to the service (e.g. keypad information and other terminal events) and receiving data feedback that shall be displayed on the terminal screen.

Dynamic payload switches within a session may be considered to transport meta-information.

4.3 Control

It shall be possible to use SRF sessions in order to provide access to SRF-based automated voice services. For example applications might use a SRF session to access and navigate within and between the various SRF-based automated voice services by spoken commands or pressed keypads.

It shall be possible for network operators to control access to SRF-based automated voice services based on subscription profile of the callers.

Deleted: Draft 3G TS 22.243 ... [8]
Deleted: Draft 3G TS 22.243 ... [9]
Inserted: Draft 3G TS 22.24 ... [10]
Deleted: Draft 3G TS 22.243 ... [11]
Inserted: Draft 3G TS 22.24 ... [12]
Inserted: Draft 3G TS 22.24 ... [13]
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: Universal
Deleted: A…V…S…,…includ ... [14]
Deleted: based automated services
Inserted: a speech recognition ... [15]
Deleted: A
Deleted: s
Inserted: s
Deleted: V
Deleted: oice
Deleted: S
Deleted: distributed … (DSR) ... [16]
Deleted: A…V…S…¶ ... [17]
Deleted: For the sake of brevit ... [18]
Deleted: DSR…-based ... [19]
Deleted: A…V ... [20]
Deleted: DSR
Deleted: DSR-based
Deleted: A…V…S ... [21]
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: shall
Deleted: standard DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: ¶
Deleted: At initiation of the DS ... [22]
Inserted: speech recognition . ... [23]
Deleted: DSR
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: is not required but … ... [24]
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: speech

4.4 User Perspective (User Interface)

The user's interface to this service shall be via the UE. User can interact by spoken and keypad inputs. The UE can have a visual display capability. When supported by the terminal, the server-based application can display visual information (e.g., stock quote figures, flight gates and times) in addition to audio playback (via recorded speech or text-to-speech synthesis) of the information.

These constitute examples of multi-modal automated services, including different configurations (sequential multi-modal, thin client multi-modal configuration, multi-device configurations, etc…) as discussed in [10,11].

5 UE and network capabilities

In addition to the capabilities required for IMS Basic Voice session (such as the default voice codec that will be used for the downlink audio prompt stream), the following SRF-based automated voice service-specific capabilities shall be required in the UE and network:

- A default uplink SRF codec as will be determined by the 3GPP SRF specification
- Possible support of an uplink DSR optimised codec [4].
- A downlink conventional codec and downlink streaming capabilities (simultaneous with uplink)
- The capability to transmit keypad information from the client to the server (e.g., either DTMF or the keypad string)

It shall be possible to enable application specific information exchanges between the client and the server (e.g. client events (e.g. barge-in events), display information, multi-modal synchronization information, etc…), in the form of speech meta-information. It shall be possible to enable these exchanges with conversational QoS.

SRF shall be supported by an uplink bandwidth of 9.6 kbits/s for the payload and QoS (Quality of Service) for conversational class services as specified in TS 22.105 [5]

It shall be possible for the network to distinguish a SRF session from a basic voice session (e.g. for charging purposes).

6. Administration

SRF-based automated voice services may be provided by the network operator (home or visited) or by third parties. See appendix A for examples of such services.

The administration of the SRF-based automated voice services shall be under the control of the network operator. But when decided to do so by the network operator, it should be possible to the third part providers to administer the SRF-based automated voice services them selves through the gateway that they would connect to IMS. In such case, the third part provider performs all the administrative steps and no registration would be required with the network operator.

6.1 Authorization

Authorization for use of SRF-based automated voice services shall be under the control of the network operator. It shall require authorization of the connection to IMS.
The network operators can provide automated voice services or they can only provide the network that connects users to SRF-based automated voice services provided by third party application service providers. The network operator shall be able to permit or prevent access to a third party service. This requirement shall be treated as equivalent to allowing or prevent access to some phones numbers (IMS voice sessions) or internet services (domains for example in WAP data access).

The network operator shall be able to administer the authorization of a SRF-based automated voice service on a user basis as well as on a service basis (e.g. to authorize access to all users or prevent access to all users).

It shall be possible for the operator to provide for the user:

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: DSR
Deleted: speech recognition
Deleted: DSR
Inserted: speech recognition
Formatted: Bullets and Numbering
Deleted: A
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: DSR
Deleted: Speech recognition
Inserted: Speech recognition
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: DSR
Deleted: A
Inserted: A
Deleted: a
Deleted: speech
Deleted: DSR
Deleted: speech
Deleted: DSR
Deleted: DSR
Deleted: speech recognition-based
Inserted: speech recognition-based
Deleted: DSR
Deleted: DSR
Deleted: n
Inserted: n
Deleted: DSR

- Authorization to access a particular "address" (e.g. 3rd party SRF-based automated voice service)
- Authorization to use a service that the operator authorize access to when the 3rd party operator wishes to rely on the operator to control this access

It shall be possible for the third party provider to authorize usage of its services based on the identity of the user.

6.2 Deauthorization

Deauthorization for use of the SRF-based automated voice services shall be under the control of the network operator as for the authorization described in section 6.1.

6.3 Registration

Authorized SRF-based automated voice service register their address with the IMS upon authorization of the service authorized SRF-based automated voice service can then be reached by the user (by entering address and initiated a SRF session).

6.4 Deregistration

Disconnection from the IMS shall prevent the use of the SRF-based automated voice service. Deregistration may be decided by the third party provider.

6.5 Activation

Once authorized and registered a SRF-based automated voice service is deemed activated as for other IMS services.

6.6 Deactivation

Deactivation shall be done by deregistering the services (operator or service provider initiative) or by refusing to initiate a SRF session (service provider initiative).

7. Service Provisioning

The SRF-based automated voice services shall be able to be provisioned by either the network operator (roaming or home) or by a 3rd party service provider.
It shall be possible for network operators and 3rd party service providers to offer SRF-based automatic voice services by providing identity of service, such as a phone number, an IP address or a URI that the user can enter or select on the terminal.

8. Security

The "Security Threats and Requirements" specified in 21.133 [1] shall not be compromised.

It shall be possible to deny unauthorized access to 3GPP SRF-based automated voice services. An authorization may be based on the following,

- identity of the accessing user agent, server or device
- the destination user, device or user agent

Third parties shall have authorization from the User and PLMN Operators in order to access 3GPP SRF-based automated voice services.

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.24 ... [25]
Inserted: Release 65
Inserted: Draft 3G TS 22.24 ... [26]
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: DSR
Deleted: DSR
Deleted: speech recognition-based
Inserted: speech recognition-based
Deleted: DSR
Deleted: A
Deleted: DSR …DSR ... [27]
Deleted: speech recognition
Inserted: speech recognition
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: A…V ... [28]
Deleted: n
Inserted: n
Deleted: DSR
Deleted: A
Inserted: A
Deleted: a
Deleted: V
Deleted: DSR
Deleted: speech recognition
Deleted: DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: -based…A…V…S ... [29]
Deleted: DSR-based
Deleted: A…V…S ... [30]
Deleted: the …Distributed ... [31]
Deleted: Speech Recognition.. ... [32]
Deleted: the …Distributed ... [33]
Deleted: Speech Recognition.. ... [34]

It shall be possible to reconstitute PCM samples from DSR packets so that the user's spoken command can be transcribed at a later time, if required.

9. Privacy

For SRF-based automated voice services, privacy requirements shall be at least as good as for IMS voice or data sessions [6]:

- It shall be possible to encrypt speech and speech meta-data exchanges

- It shall be possible to prevent exchange of the user's true identity, location and other terminal or user related information when required.

SRF-based automated voice services, may imply that the service provider collects 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. The SRF-based automated voice services shall not add additional privacy risks.

10. Charging

The user can be charged for sessions with SRF-based automated voice services in a variety of ways. The following shall be possible:

a) By duration of session (including "one-off" charge/flat rate)

b) By data volume transferred (number of packets) or other similar criteria.

c) By subscription fees for the service (unlimited usage or unlimited usage up to a point and then per-use fees)

d) Free (e.g. with the service being subsidised by advertising revenue from advertisement spots). The advertisement spots may be inserted either at session start-up or close, or designed in such that system delay time is masked (e.g., while the user is waiting for the flight schedules to be returned, or a purchase transaction to be completed). The network operator will receive revenue from users directly as well as from the content and service providers who want their sites to be accessible via the automated voice service, and from advertisers. Advertising spots can be inserted at appropriate points during the session (e.g., at the beginning of the session, while the user is waiting for a system response, or at the end of a session).

SRF-based automated voice services shall be available to pre-paid and post-paid subscribers.

11. Roaming

The user shall be able to utilize SRF-based automated voice services when roaming in any IMS compatible mobile network.

The capabilities of the SRF-based automated voice service shall be available in the roamed-to network in the same manner as in the home network, within the limitation of the capabilities of the serving network.

12. Interaction with other services

No interaction with other services identified. When connected to the IMS, other IMS services are available to the user through the terminal.

Other services (non IMS voice etc..) may be available with or without disconnecting from the IMS.

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65
Inserted: Release 65
Deleted: Release 65Release 6
Inserted: Release 65
Deleted: DSR
Deleted: DSR a
Deleted: A
Inserted: A
Deleted: DSR speech recognition
Inserted: speech recognition-based
Deleted: the DSR
Deleted: speech recognition
Inserted: speech recognition
Deleted: V
Deleted: S
Deleted: DSR
Deleted: Speech recognition
Inserted: Speech recognition
Deleted: the DSR
Deleted: Speech recognition
Inserted: Speech recognition
Deleted: V
Deleted: S
Deleted: DSR-based
Deleted: V
Deleted: S

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65Release 6

Inserted: Release 65

Deleted: DSR

Annex A (informative): Speech recognition Framework-based automated voice service examples

Examples of Automated Voice Services include:

- 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)
- Multi-modal applications with a GUI user agent on the terminal synchronized with a DSR automated voice service.

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Draft 3G TS 22.243 1.45.0 (2002-516)

Inserted: Draft 3G TS 22.243 1.45.0 (2002-516)

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65

Inserted: Release 65

Deleted: Release 65Release 6

Inserted: Release 65

Annex B (informative): Change History

Change history											
TSG SA#	**SA Doc.**	**SA1 Doc**	**Spec**	**CR**	**Rev**	**Rel**	**Cat**	**Subject/Comment**	**Old**	**New**	**Work Item**
			22.243					Initial draft based on content of TR 22.941		0.0.1	
								Output of ad-hoc drafting session		0.0.2	
								Presented to TSG SA#14 for information	0.0.2	1.0.0	
								Reviewed in SWG DSR 14 Jan 2002	1.0.0	1.1.0	
								Reviewed again in SWG DSR 14 Jan 2002	1.1.0	1.2.0	
								Contribution to 1.2.0 to address open issues	1.2.0	1.3.0	
								Update to 1.3.0 to address comments at DSR SWG (Sophia Antipolis)	1.3.0	1.4.0	
								Update to 1.4.0 to reflect the new SRF terminology.	1.4.0	1.5.0	

Deleted: provide more generic terminology

Deleted: to

Inserted: to reflect the new SRF terminology. ... [35]

SA1 Rapporteur: Jerome Vogedes, Motorola, Inc.

Email: Jerome.Vogedes@motorola.com Telephone: +18475231796

Page 3: [1] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 10:52 PM

Contents 3
Foreword 5
Introduction 6
1. Scope 7
2. References 8
2.1 Informative References 8
3. Definitions and abbreviations8
3.1 Definitions 8
3.2 Abbreviations 9
4. Requirements 10
4.1 Initiation 10
4.2 Information during the speech recognition session 10
4.3 Control 11
4.4 User Perspective (User Interface) 11
5 UE and network capabilities 11
6. Administration 11
6.1 Authorization 11
6.2 Deauthorization 12
6.3 Registration 12
6.4 Deregistration 12
6.5 Activation 12
6.6 Deactivation 12
7. Service Provisioning 12
8. Security 12
9. Privacy 13
10. Charging 13
11. Roaming 13
12. Interaction with other services 13
Annex A (informative): Speech recognition service examples 15
Annex B (informative): Change History 16
Contents 3
Foreword 4
Introduction 5
1. Scope 6
2. References 6
3. Definitions and abbreviations 7
3.1 Definitions 7
3.2 Abbreviations 7
4. Requirements 8
4.1 Initiation 8
4.2 Information during the DSR session 8
4.3 Control 9
4.4 User Perspective (User Interface) 9
5 UE and network capabilities 9

6. Administration 9
6.1 Authorization 10
6.2 Deauthorization 10
6.3 Registration 10
6.4 Deregistration 10
6.5 Activation 10
6.6 Deactivation 10
7. Service Provisioning 10
8. Security 11
9. Privacy 11
10. Charging 11
11. Roaming 12
12. Interaction with other services 12
Annex A (informative): DSR service examples 13
Annex B (informative): Change History 14

Page 3: [2] Inserted Motorola PC 6/23/2002 11:32 PM

Contents 3
Foreword 5
Introduction 6
1. Scope 7
2. References 8
2.1 Informative References 8
3. Definitions and abbreviations 8
3.1 Definitions 8
3.2 Abbreviations 9
4. Requirements 10
4.1 Initiation 10
4.2 Information during the speech recognition session 10
4.3 Control 11
4.4 User Perspective (User Interface) 11
5 UE and network capabilities 11
6. Administration 11
6.1 Authorization 11
6.2 Deauthorization 12
6.3 Registration 12
6.4 Deregistration 12
6.5 Activation 12
6.6 Deactivation 12
7. Service Provisioning 12
8. Security 12
9. Privacy 13
10. Charging 13
11. Roaming 13
12. Interaction with other services 13
Annex A (informative): Speech recognition service examples 15
Annex B (informative): Change History 16

Page 2: [3] Inserted David Pearce 7/1/2002 3:47 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 2: [4] Inserted Motorola PC 6/27/2002 3:52 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 6: [5] DeletedStéphane H. Maes - smaes@us.ibm.com 5/2/2002 12:27 PM

improves the speech recognition performances over other

Page 6: [6] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:12 AM

speech recognition speech recognition

Page 6: [7] Inserted Motorola PC 6/23/2002 10:26 PM

speech recognition speech recognition

Page 2: [8] DeletedStéphane H. Maes - smaes@us.ibm.com 7/1/2002 3:30 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 2: [9] DeletedStéphane H. Maes - smaes@us.ibm.com 7/1/2002 3:30 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 2: [10] Inserted David Pearce 7/1/2002 3:47 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 2: [11] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 10:41 AM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 2: [12] Inserted Motorola PC 7/1/2002 1:12 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 2: [13] Inserted Motorola PC 6/27/2002 3:52 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 9: [14] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:53 AM

A

Page 9: [14] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:53 AM

V

Page 9: [14] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:54 AM

S

Page 9: [14] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:29 AM

,

Page 9: [14] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:29 AM

including DSR a speech recognition framework

Page 9: [15] Inserted Motorola PC 6/23/2002 10:51 PM

a speech recognition framework

Page 9: [16] Deleted Motorola PC 6/23/2002 10:55 PM

distributed

Page 9: [16] Deleted Motorola PC 6/23/2002 10:55 PM

(DSR)

Page 9: [17] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:54 AM

A

Page 9: [17] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:54 AM

V

Page 9: [17] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:54 AM

S

Page 9: [17] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:29 AM

Page 9: [18] Deleted Motorola PC 6/23/2002 10:56 PM

For the sake of brevity, these services will be referred to as 'DSR-based Automated Voice Service' in the rest of this section.

Page 9: [19] Deleted Motorola PC 6/23/2002 10:55 PM

DSR

Page 9: [19] Deleted Motorola PC 6/23/2002 10:56 PM

-based

Page 9: [20] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:54 AM

A

Page 9: [20] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:54 AM

V

Page 9: [21] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:55 AM

A

Page 9: [21] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:55 AM

V

Page 9: [21] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:55 AM

S

Page 9: [22] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:34 AM

At initiation of the DSR speech recognition session it shall be possible to negotiate the uplink and downlink codecs. The terminal shall support a standard DSRspeech recognition optimized codec as a default uplink codec.

Page 9: [23] Inserted Motorola PC 6/23/2002 11:05 PM

speech recognition

Page 9: [23] Inserted Motorola PC 6/23/2002 11:06 PM

speech recognition

Page 9: [23] Inserted Motorola PC 6/23/2002 11:07 PM

a

Page 9: [24] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:35 AM

is not required but

Page 9: [24] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:35 AM

c

Page 2: [25] Inserted David Pearce 7/1/2002 3:47 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 2: [26] Inserted Motorola PC 6/27/2002 3:52 PM

Draft 3G TS 22.243 1.45.0 (2002-516)

Page 11: [27] Deleted Motorola PC 6/23/2002 11:15 PM

DSR

Page 11: [27] Deleted Motorola PC 6/23/2002 11:15 PM

DSR

Page 11: [28] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:58 AM

A

Page 11: [28] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:58 AM

V

Page 11: [29] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

-based

Page 11: [29] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

A

Page 11: [29] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

V

Page 11: [29] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

S

Page 11: [30] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

A

Page 11: [30] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

V

Page 11: [30] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

S

Page 11: [31] Deleted Motorola PC 6/23/2002 11:19 PM

the

Page 11: [31] Deleted Motorola PC 6/23/2002 11:18 PM

Distributed

Page 11: [32] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

Speech Recognition

Page 11: [32] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

Page 11: [32] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

B

Page 11: [32] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

V

Page 11: [32] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

S

Page 11: [33] Deleted Motorola PC 6/23/2002 11:20 PM

the

Page 11: [33] Deleted Motorola PC 6/23/2002 11:19 PM

Distributed

Page 11: [34] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

Speech Recognition

Page 11: [34] DeletedStéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:59 AM

Page 11: [34] Deleted Stéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

B

Page 11: [34] Deleted Stéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

V

Page 11: [34] Deleted Stéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:45 AM

S

Page 14: [35] Inserted Stéphane H. Maes - smaes@us.ibm.com 6/30/2002 11:48 AM

								Update to 1.4.0 to provide more generic terminologyto reflect the new SRF terminology.	1.4.0	1.5.0	