Technical Specification Group Services and System Aspects
Meeting #18, New Orleans, USA, 9-12 December 2002

TSGS#17(02)0663

Source: **SA1**

Title: **Release 6 CR on TS 22.243 on Codecs used for speech recognition framework**

Document for: **Approval**

Agenda Item: **7.1.3**

SA Doc	Spec	CR	Rev	Phase	Cat	Subject	Old Vers	New Vers	SA1 Doc
SP-020663	22.243	002		Rel-6	F	CR on TS 22.243, Codecs used for speech recognition framework	6.0.0	6.1.0	S1-022162

TSG-SA1 SWG #18 **S1-022162**
Busan, South Korea, November 2002 **Agenda Item: 10.10**

CR-Form-v7

CHANGE REQUEST

⌘ **22.243** **CR** **002** ⌘ **rev** **-** ⌘ Current version: **6.0.0** ⌘

*For **HELP** on using this form, see bottom of this page or look at the pop-up text over the ⌘ symbols.*

Proposed change affects: UICC apps⌘ [ ] ME [X] Radio Access Network [ ] Core Network [X]

Title:	⌘	Codecs used for speech recognition framework
Source:	⌘	SA1 (Ericsson)
Work item code:	⌘	SRSES
Date:	⌘	31.10.2002
Category:	⌘	**F**
Release:	⌘	Rel-6

Use one of the following categories:
F *(correction)*
A *(corresponds to a correction in an earlier release)*
B *(addition of feature),*
C *(functional modification of feature)*
D *(editorial modification)*
Detailed explanations of the above categories can be found in 3GPP TR 21.900.

Use one of the following releases:

2	*(GSM Phase 2)*
R96	*(Release 1996)*
R97	*(Release 1997)*
R98	*(Release 1998)*
R99	*(Release 1999)*
Rel-4	*(Release 4)*
Rel-5	*(Release 5)*
Rel-6	*(Release 6)*

Reason for change:	⌘	The TS 22.243 contains in a couple of places discussion about enabling use of codecs for speech recognition framework. It is not clear from the TS if a default codec for this purpose should be considered. Since conventional codecs are already included in the standard we suggest that for Rel-6 the conventional codecs are considered as default and the DSR codecs as optional.
Summary of change:	⌘	Conventional codecs are default for speech recognition framework Rel-6, DSR codecs are optional.
Consequences if not approved:	⌘	Not clear what is default codec for speech recognition framework

Clauses affected:	⌘	Chapter 4, 5

Other specs affected: ⌘	Y	N		⌘
		X	Other core specifications	
		X	Test specifications	
		X	O&M Specifications	

Other comments: ⌘

How to create CRs using this form:
Comprehensive information and tips about how to create CRs can be found at http://www.3gpp.org/specs/CR.htm.
Below is a brief summary:

1) Fill out the above form. The symbols above marked ⌘ contain pop-up help information about the field that they are closest to.

2) Obtain the latest version for the release of the specification to which the change is proposed. Use the MS Word "revision marks" feature (also known as "track changes") when making the changes. All 3GPP specifications can be downloaded from the 3GPP server under ftp://ftp.3gpp.org/specs/ For the latest version, look for the directory name with the latest date e.g. 2001-03 contains the specifications resulting from the March 2001 TSG meetings.

3) With "track changes" disabled, paste the entire CR form (use CTRL-A to select it) into the specification just in front of the clause containing the first piece of changed text. Delete those parts of the specification which are not relevant to the change request.

4 Requirements

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

- Default uplink and downlink codec specifications.
- A stack of speech recognition protocols to support:
  - Establishment of uplink and downlink sessions, along with codec negotiation
  - Transport of speech recognition payload (uplink) with conversational QoS
  - 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 events, parameters and settings, etc.).

IMS provides a protocol stack (e.g. SIP/SDP, RTP and QoS), that may advantageously be used to implement such capabilities.

It shall be possible to recommend a codec to be supported by default to deploy services that rely on the 3GPP speech recognition framework. To that effect, the specifications will consider either conventional speech codecs (e.g. AMR) or optional DSR optimized codecs.

ETSI has published DSR optimized codecs specifications (ETSI ES 201 108 & ETSI ES 202 050 [7, 8, 9, 10]) and a payload format for transport of DSR data over RTP (IETF AVT DSR).

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 may be offered over a packet switched network; however in general this requires specification of a complete protocol stack. When this service is 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 can and shall 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 shall support a codec suitable for speech recognition ~~optimized codec~~ as a default uplink codec.

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 codec (conventional codec or optional DSR optimized codec).
- 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, 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 [4]

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