ETSI TS 122 243 V13.0.0 (2016-01)

TECHNICAL SPECIFICATION

Digital cellular telecommunications system (Phase 2+);
Universal Mobile Telecommunications System (UMTS);
LTE;
Speech recognition framework for automated voice services;
Stage 1
(3GPP TS 22.243 version 13.0.0 Release 13)

Reference

RTS/TSGS-0122243vd00

Keywords

GSM,LTE,UMTS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (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 the present document, 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 document.

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 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. 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.

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.

1 Scope

The present document 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 Technical Specification includes information applicable to network operators, service providers, terminal and network manufacturers.

This Technical Specification 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 as part of speech-enabled services. The Speech Recognition Framework for automated voice services is an optional feature in a 3GPP system.

Figure 1 positions the Speech recognition Framework (SRF) with respect to other speech-enabled services as discussed in [6]. As illustrated, SRF is designed to support server-side speech recognition over packet switched network (e.g. IMS). As such SRF also enable configurations of multimodal and multi-device services that include distribute the speech engines.

Note that it is possible to design speech-enabled services that alternate or combine the use of client-side only engines and SRF.

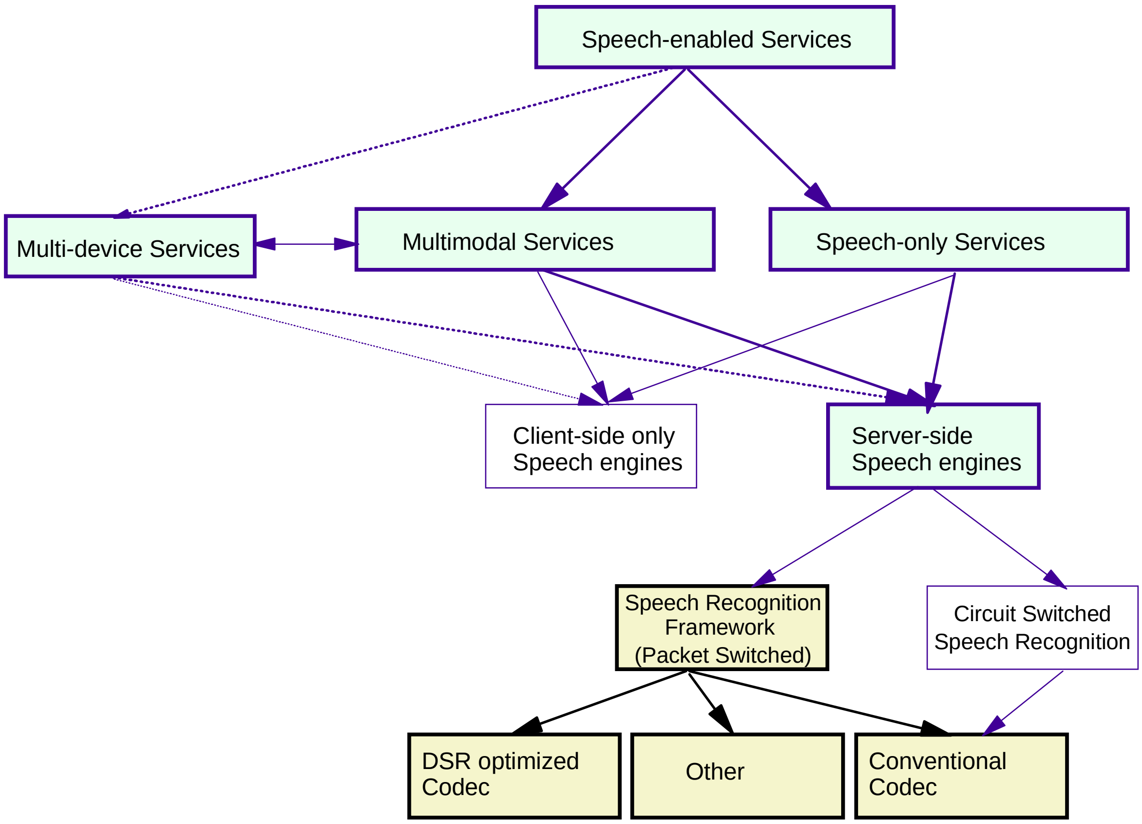

Figure 1 - Positions the scope of the speech recognition framework as part of general speech enabled services.