

**GENIVI Alliance**

GENIVI Document CS000XX

SpeechOutputService

Component Specification

Not Accepted Version 1.0.0

**25-01-2017**

**Sponsored by:**

GENIVI Alliance

true

NotAccepted

**Abstract:**

This document provides a component specification for the Speech Output Service.This document provides the Component Specification for the NavigationCore

**Keywords:**

Navigation, LocationInput, Routing, Guidance

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.Copyright © 2017, trueCompany ABC, Company XYZ.

All rights reserved.

The information within this document is the property of the true and its use and disclosure are restricted. Elements of GENIVI Alliance specifications may be subject to third party intellectual property rights, including without limitation, patent, copyright or trademark rights (and such third parties may or may not be members of GENIVI Alliance). GENIVI Alliance true not responsible and shall not be held responsible in any manner for identifying, failing to identify, or for securing proper access to or use of, any or all such third party intellectual property rights.true

“”“”

GENIVI and the GENIVI Logo are trademarks of GENIVI Alliance in the U.S. and/or other countries. Other company, brand and product names referred to in this document may be trademarks that are claimed as the property of their respective owners.

trueThis work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

http://creativecommons.org/licenses/by-sa/4.0

The above notice and this paragraph must be included on all copies of this document that are made.

GENIVI Alliance

2400 Camino Ramon, Suite 375

San Ramon, CA 94583, USA

Revision History

Document revision history

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Author | Description |
| *2015-01-02* | *0.1* | *David Kämpf* | *Initial revision. Containing the results of the Speech F2F in Erlangen 12/2014.* |
| *2015-01-05* | *0.1.1* | *Mario Thielert* | *Minor corrections on initial revision* |
| *2015-01-05* | *0.1.2* | *David Kämpf* | *Synchronized with new UML model. Minor additions and fixes.* |
| 2017-02-22 | 0.1.3 | Philippe Colliot | Refine document to allow API generation from the Franca file |

Table of Contents

[1.1System Overview 5](#__RefHeading___Toc5043_724390506)

[1.2Subsystem Speech Overview 5](#__RefHeading___Toc5045_724390506)

[1.3Component Overview 5](#__RefHeading___Toc5047_724390506)

[1.4Document Overview 6](#__RefHeading___Toc5049_724390506)

[2References 7](#__RefHeading___Toc5051_724390506)

[3Glossary 8](#__RefHeading___Toc5053_724390506)

[4Requirements 9](#__RefHeading___Toc5055_724390506)

[4.1Functional Requirements 9](#__RefHeading___Toc5057_724390506)

[4.2Non Functional Requirements 13](#__RefHeading___Toc5059_724390506)

[5Constraints and Assumptions 14](#__RefHeading___Toc5061_724390506)

[6Architecture 15](#__RefHeading___Toc5063_724390506)

[6.1Architecture Overview 15](#__RefHeading___Toc5065_724390506)

[6.1.1Component Interfaces 15](#__RefHeading___Toc5067_724390506)

[6.1.2Component Dependencies 16](#__RefHeading___Toc5069_724390506)

[6.1.3Component Traceability 16](#__RefHeading___Toc5071_724390506)

[6.2SpeechOutputService Details 16](#__RefHeading___Toc5073_724390506)

[6.2.1Responsibility and Features 16](#__RefHeading___Toc5075_724390506)

[6.2.2Provided Interfaces 17](#__RefHeading___Toc5077_724390506)

[6.2.3Required Interfaces 17](#__RefHeading___Toc5079_724390506)

[7Collaboration 18](#__RefHeading___Toc5081_724390506)

[7.1Use Case Realization: Play Prompts sequentially 18](#__RefHeading___Toc5083_724390506)

[7.2Use Case Realization: Add multiple Text Chunks 18](#__RefHeading___Toc5085_724390506)

[7.3Use Case Realization: Navigation and Reader Application trying to prompt 20](#__RefHeading___Toc5087_724390506)

[8Interfaces 21](#__RefHeading___Toc5089_724390506)

[1.1The following pages describe the interfaces of the SpeechOutputService API 21](#__RefHeading___Toc5091_724390506)

[9Implementation 22](#__RefHeading___Toc5093_724390506)

[9.1Implementation details 22](#__RefHeading___Toc5095_724390506)

[9.2Usage examples 22](#__RefHeading___Toc5097_724390506)

[9.3Test Plan 22](#__RefHeading___Toc5099_724390506)

## System Overview

*Boiler plate, to be written, describing the overall GENIVI Software Platform.*

## Subsystem Speech Overview

The Speech Subsystem contains of 3 components:

* **Speech Dialog Service**
  + Modeling the User Interaction
  + Handling of resources
  + Interact with GUI
  + Interact with Business Logic
* **Speech Input Service** 
  + Integration of the Voice Recognizer
  + Resource handling
* **Speech Output Service**
  + Integration of TTS Engine

C:\Documents and Settings\Administrator\Desktop\Image2.EMF*.*

## Component Overview

The SpeechOutputService is encapsulating access to the TTS engine. The responsibilities of SpeechOutputService are:

* Provide access to the TTS for the SpeechDialogService
* Provide access to the TTS for applications
* Arbitrate conflicting requests to speak a prompt
* Encapsulate audio connection handling for the client application

In the current POC implementation SpeechOutputService’s interface is realized as C++ interface. Due to the fact that it is strictly asynchronous other implementations are feasible as well.

## Document Overview

This document is describing the SpeechOutputService abstract component version 0.1.2.

# References

The following standards and specifications contain provisions, which through reference in this document constitute provisions of this specification. All the standards and specifications listed are normative references. At the time of publication, the editions indicated were valid. All standards and specifications are subject to revision, and parties to agreements based on this specification are encouraged to investigate the possibility of applying the most recent editions of the standards and specifications indicated below.

1. GENIVI UML Model - https://svn.genivi.org/uml-model/genivi/trunk

# Glossary

|  |  |  |
| --- | --- | --- |
| Acronym | Term | Definition |
| RDF | Resource Description Framework |  |
| DB | Database |  |
| NBT | Next Big Thing |  |
| EWMH | Extended Window Manager Hints |  |
| KDE | K Desktop (Windows) Environment |  |
| ICCM | Inter-Client Communication Conventions Manual (of KDE) |  |
| PDA | Personal Digital Assistant |  |
| PIM | Personal Information Management |  |
| UUID | Unique ID (Identification) |  |
| CDDB | Compact Disc Database (now Gracenote) |  |
| IDE | Integrated Drive Electronics |  |
| SCSI | Small Computer Serial Interface |  |
| USB | Universal Serial Bus |  |
| UDF | Universal Disk Format (for optical media) |  |
| VFAT | Virtual File Allocation Table (Linux support for Microsoft FAT file systems) |  |
| TTS | Text to Speech Engine | An engine that is converting graphems (written text) into spoken output. |
| SPC | Speech |  |
| SOS | Speech Output Service |  |
| VR | Voice Recognizer |  |
| SW | Software |  |
| IPC | Inter Process Communication |  |

Table 1 – Acronym and Term Definitions

# Requirements

The information in this chapter is provided only for information purpose; this is not a normative part.

## Functional Requirements

|  |  |  |  |
| --- | --- | --- | --- |
| **Requirement ID** | **Requirement** | **Priority** | **Rationale** |
| SW-SPC-SOS-071 | After loading of languages or voices, the TTS shall support a signal indicating the end of loading. | Medium |  |
| SW-SPC-SOS-064 | At system SW upgrade of the Speech Output component, TTS engine, languages and prerecordings shall be upgradable separately. | Medium |  |
| SW-SPC-SOS-039 | Context Dependent Disambiguation of Abbreviations: TTS shall disambiguate frequently used abbreviations based on their context. Abbreviations may have a different meaning and pronounciation based on their context. E.g. compare "St. Martin" versus "main | Medium |  |
| SW-SPC-SOS-059 | Continuity of Speech Output Within a text to be output by TTS, no unnatural pauses shall occur effected by operational constraints. | Medium |  |
| SW-SPC-SOS-036 | Control of Speaking Rate: TTS shall enable the user to adjust the average speaking rate of TTS for a specific voice according to his personal preferences. Such an adjustment is permanent and survives power-off. | Medium |  |
| SW-SPC-SOS-037 | Control of Voice Pitch: TTS shall enable the user or an application to adjust the pitch level of a voice. TTS shall accept permanent voice-specific adjustments by a user and temporary adjustments by an application. | Medium |  |
| SW-SPC-SOS-041 | Customer specific dictionary extensions shall be permanent and survive power-off. | Medium |  |
| SW-SPC-SOS-072 | During loading of languages or voices, the system shall supply an indicator for the status of the loading process, to be potentially used by the HMI to show a progress bar. | Medium |  |
| SW-SPC-SOS-018 | Identification of Foreign Words in a Text: TTS should identify the language origin of foreign language words in a native language text, and select language specific conversion rules that lead to an adequate pronounciation. | Medium |  |
| SW-SPC-SOS-017 | Identification of Text Language: TTS should identify the language in which a text is written, allowing to select an appropriate TTS module for the identified language. | Medium |  |
| SW-SPC-SOS-055 | In case of an upgrade at a major change level including modification of user-specific data, the TTS module shall provide an additional function allowing transformation of the user-specific data into a re-usable form, with an acceptable processing time an | Medium |  |
| SW-SPC-SOS-021 | In case of phonetic input, TTS shall map any phonemes from foreign languages to appropriate phonemes of the active language. | Medium |  |
| SW-SPC-SOS-019 | Language tagging: TTS should support language tags for words/phrases in the text passed to the TTS. | Medium |  |
| SW-SPC-SOS-024 | Modification of Voices at Runtime: The TTS shall support modification of the parameters of the currently used voice at runtime, to allow the adaptation of the voice (e.g. speed and pitch) to the preferences of the customer or user. | Medium |  |
| SW-SPC-SOS-016 | Multi-Linguality: The TTS shall be able to read text that contains different languages with adequate pronounciation. | Medium |  |
| SW-SPC-SOS-023 | Multiple Voices per Language: At least for the primary languages, TTS shall allow selection between at least one female and one male voice. | Medium |  |
| SW-SPC-SOS-020 | Pronounciation of Non-Native Phonemes: The TTS shall be able to speak a text which includes phonemes that are not part of the active language. | Medium |  |
| SW-SPC-SOS-006 | Reading of Abbreviations: The TTS module shall support pronounciation of common abbreviations. | Medium |  |
| SW-SPC-SOS-005 | Reading of Mixed Plain and Phonetic Text: The TTS module shall support mixing of text in phonetic form and in graphemic form into application-specific messages, if sufficient application-specific information is available to create an overall message inton | Medium |  |
| SW-SPC-SOS-007 | Reading of Numbers: The Speech Output component shall support tags for numbers embedded into speakble text to enforce the way numbers are spoken. E.g. Numbers can be spoken as numerical value (nine hundred eleven) or as digit sequence (nine one one). | Medium |  |
| SW-SPC-SOS-004 | Reading of Phonetic Text: The TTS module shall support reading of phonetic texts, optionally including prosody information, and shall generate output in form of an audio stream. | Medium |  |
| SW-SPC-SOS-003 | Reading of Plain Text: The TTS module shall support reading of plain ( = graphemic resp. orthographic) text, generating speech output in form of an audio stream. | Medium |  |
| SW-SPC-SOS-060 | Seamless Speech Output: Concatenated prerecorded prompts shall produce a continuous voice stream, without any unintended speech gaps.This requirement has to be fulfilled during the output of a prerecorded speech segment, during concatenation at the boun | Medium |  |
| SW-SPC-SOS-009 | Speech Output component shall process application specifc tags indicating a context (like E-Mail, TTS or Navigation) in order to enable the TTS to optimize text reading. | Medium |  |
| SW-SPC-SOS-038 | Standard Dictionary for Abbreviations: TTS shall contain a standard dictionary for correct pronounciation of unusual words and for standard abbreviations. | Medium |  |
| SW-SPC-SOS-040 | TTS shall allow extension of the standard dictionaries and abbreviations by customer specific entries. | Medium |  |
| SW-SPC-SOS-058 | TTS shall be able to output user specific prompts. | Medium |  |
| SW-SPC-SOS-057 | TTS shall be able to store and access a customer or user specific exception dictionary. | Medium |  |
| SW-SPC-SOS-035 | TTS shall enable the user or an application to adjust the volume of different TTS voices in relation to each other according to his personal preferences. Such a modification shall be permanent and survives power-off. | Medium |  |
| SW-SPC-SOS-029 | TTS shall offer a mechanism to specify/mark time positions within or at the end of an utterance | Medium |  |
| SW-SPC-SOS-043 | TTS shall offer an improved pronounciation of names, if a name is explicitly marked as name. | Medium |  |
| SW-SPC-SOS-044 | TTS shall offer an improved pronounciation of street names and location names, if the street name or location name is marked as address. | Medium |  |
| SW-SPC-SOS-042 | TTS shall offer the usual text specific preprocessing mechanisms to correctly pronounce various number formats, date, time, currencies, phone numbers etc. | Medium |  |
| SW-SPC-SOS-046 | TTS shall provide an improved pronounciation of a CD, DVD, song or movie title, if the text item is marked as a related title. | Medium |  |
| SW-SPC-SOS-045 | TTS shall provide an improved pronounciation of radio station names, if the name is marked as radio station name. | Medium |  |
| SW-SPC-SOS-031 | TTS shall support activation of a language that has been newly loaded into non-volatile memory. | Medium |  |
| SW-SPC-SOS-062 | TTS shall support activation of more than one language. | Medium |  |
| SW-SPC-SOS-032 | TTS shall support deactivation of the active language for instance for language resources exchange. | Medium |  |
| SW-SPC-SOS-002 | The G2P of the TTS module shall be aligned with the G2P used by the Speech Recognizer to pronounce the spoken text in a way that is suitable for speech recognition. | Medium |  |
| SW-SPC-SOS-048 | The TTS engine can be switched off. | Medium |  |
| SW-SPC-SOS-010 | The TTS module shall be able to correctly interpret or map the phoneme sets used by third party database providers like GraceNote or Navteq. | Medium |  |
| SW-SPC-SOS-008 | The TTS module shall support domain specific exception dictionaries. An exception dictionary bypasses the G2P processing of the passed word or abbreviation and provides the domain specific pronounciation. | Medium |  |
| SW-SPC-SOS-012 | The TTS module shall support reading of POI names provided in text form or phonetical form by navigation databases. | Medium |  |
| SW-SPC-SOS-011 | The TTS module shall support reading of location names or street names provided in text form or phonetical form by navigation databases. | Medium |  |
| SW-SPC-SOS-013 | The TTS module shall support reading of person names as stored in telephone databases or PDA contact databases. | Medium |  |
| SW-SPC-SOS-014 | The TTS module shall support reading of the contents of artist names, song titles, and album titles, as stored in the data fields of audio entertainment files like e.g. MP3-files or WMA-files. | Medium |  |
| SW-SPC-SOS-015 | The TTS module shall support reading of the contents of artist names, song titles, and album titles, using enhanced information as provided by web-based databases like e.g. Gracenote. | Medium |  |
| SW-SPC-SOS-034 | The TTS shall deliver it's status. | Medium |  |
| SW-SPC-SOS-026 | The TTS shall provide a function to stop the running prompt. | Medium |  |
| SW-SPC-SOS-025 | The TTS shall start Text to Speech conversion after receiving an explicit start command. | Medium |  |
| SW-SPC-SOS-067 | The TTS system shall support selection of a specific voice by the customer, out of a set of one or more voices per language. | Medium |  |
| SW-SPC-SOS-056 | The language loading methods shall support the exchange of application-specific voice prompt sets without conflicts to other voice prompt sets of the used voice or voice style. | Medium |  |
| SW-SPC-SOS-022 | The mapping of phonemes of foreign languages to the active language shall be seamless, i.e. no foreign phonemes may bed dropped due to a missing mapping function. | Medium |  |
| SW-SPC-SOS-073 | The speech output component shall be compiant with GENIVI Audio architecture. | Medium |  |
| SW-SPC-SOS-053 | The speech output module shall be able to read aloud electronic mail and SMS messages. | Medium |  |
| SW-SPC-SOS-054 | The speech output module shall be able to read aloud the contents of fields in ID3 tags as provided by compressed music files. | Medium |  |
| SW-SPC-SOS-050 | The speech output module shall support output consisting of both TTS speech sequences and prompt speech sequences, mixed in arbitrary order. | Medium |  |
| SW-SPC-SOS-051 | The speech output module shall support output of announcements for navigation guidance messages. | Medium |  |
| SW-SPC-SOS-052 | The speech output module shall support output of announcements for speech dialogues. | Medium |  |
| SW-SPC-SOS-049 | The speech output module shall support output of prerecorded speech. | Medium |  |
| SW-SPC-SOS-047 | The speech output module shall support output of synthetic speech generated by TTS. | Medium |  |
| SW-SPC-SOS-079 | The speech output service shall return an error on a new prompter request while a TTS is active | Medium |  |
| SW-SPC-SOS-076 | The speech output system shall return an error if pause is called and no TTS is active | Medium |  |
| SW-SPC-SOS-077 | The speech output system shall return an error if resume is called and no TTS is active | Medium |  |
| SW-SPC-SOS-075 | The speech output system should provide the current state | Medium |  |
| SW-SPC-SOS-069 | The system shall enable the car manufacturer and the user to specify the voice of the active language. | Medium |  |
| SW-SPC-SOS-068 | The system shall enable the car manufacturer and the user to specify the voices to be loaded for each language. | Medium |  |
| SW-SPC-SOS-033 | The system shall enable the user to select the active language out of the set of loaded languages. | Medium |  |
| SW-SPC-SOS-001 | The system shall have a TTS Engine. | Medium |  |
| SW-SPC-SOS-065 | The system shall support a deletion mechanism to remove language subpackages that are not required any more. | Medium |  |
| SW-SPC-SOS-063 | The system shall support a language loading mechanism to replace the available languages. | Medium |  |
| SW-SPC-SOS-066 | The system shall support a version comparison mechanism to determine the languages that need to be updated to newer versions. | Medium |  |
| SW-SPC-SOS-061 | The system shall support output of recorded speech including signals other than speech like e.g. jingles or signal tones in a sound quality comparable to the quality of prerecorded speech. | Medium |  |
| SW-SPC-SOS-030 | Whenever TTS speech output reaches a marked position, TTS shall issue an event to the system, signalling that speech output has reached the defined time position. | Medium |  |
| SW-SPC-SOS-070 | While loading languages, voices or prompt sets, speech input and output shall be disabled for all applications. | Medium |  |
| SW-HAF-SPC-018 | The system generates a spoken feedback or guidance for the user. The audio data is either already stored on the system (pre-recording) or will be generated on the fly by a TTS engine. | Medium |  |

## Non Functional Requirements

There are currently no non-functional requirements for the SpeechOutputService*.*

# Constraints and Assumptions

*This section shall summarize the constraints and assumptions done in the project for the component.*

# Architecture

The information in this chapter is provided only for information and recommendation purpose; this is not a normative part.

## Architecture Overview



### Component Interfaces

C:\Documents and Settings\Administrator\Desktop\Image2.EMFThe main interface is the ISpeechOutputService which provides functions to:

* Open a prompter session
* Add text to the SpeechOutputService that will be spoken by the TTS
* Close a prompter session

In addition to that there is the IObserver interface that delivers status changes to the clients.

When opening a prompter session two paramaters will have to be provided: the connection type and the pre-processing tyoe:

* The connection type will be used to prioritize conflicting TTS requests, e.g. a Navigation prompt will not be interrupted by a Reader application trying to read out an E-Mail.
* The pre-processing type is configuring the pre-processor of the built in TTS engine in order to be optimized for a specific use case. Normally the pre-processor would select a specific set of rules that are application specific and would e.g in the case of navigation expand “in 100m turn right” to “in 100 meters turn right”.

Once the client has opened a TTS session with OpenPrompter it can speak by adding text chunks to the engine using the addTextChunk method.

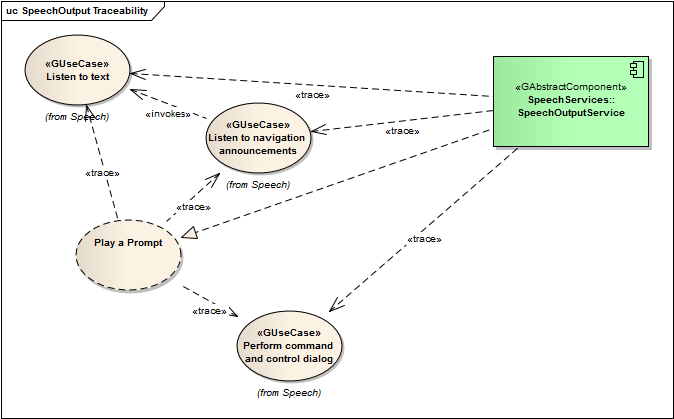
The client will receive receive status notifications concerning two major aspects:

* Status of the session (CS – Connection status), e.g. a notification the opening the prompter was successful
* Status of the TTS engine, e.g. a notfiction that the TTS engine is currently reading out text or has reached a specific position in text.

### Component Dependencies

SpeechOutputService has currently no dependencies.

### Component Traceability



Mainly the SpeechOutputService implements the “Play a prompt” Use Cases, be this navigation announcements, content from E-Mails or other messages and lists.

In addition to that SpeechOutputService traces all of the requirements in the section Speech Output as listed in chapter 4. Most of these requirements are satisfied by the TTS engine encapsulated by SpeechOutputService.

## SpeechOutputService Details

### Responsibility and Features

Responsibilities of the SpeechOutputService:

* Encapsulate the TTS engine and provide a vendor agnostic interface for applications
* Arbitrate concurrent access to the TTS engine from different applications
* Provide session handling to ease application development

### Provided Interfaces

SpeechOutputService provides ISpeechOutputService which is intended to provide control over the TTS engine for applications.

ISpeechOutputService provides methods to

* Control a session (openPrompter, closePrompter)
* Add text to the TTS buffer to be spoken
* Abort a running prompt

### Required Interfaces

SpeechOutputService in the current POC implementation requires the IObserver interface which is defining the callbacks that deliver status information to the clients.

Each client can register it’s callbacks at the SpeechOutputService and will be provided with information about:

* TTS status – providing information about the TTS engine, e.g. if TTS has started to put out a prompt
* Session status – providing information if the application was able to open a TTS session
* Buffer status – providing information about the text buffer

# Collaboration

## Use Case Realization: Play Prompts sequentially

The following sequence describes the “good case” of an application trying to put out a prompt. The application opens the session successfully with openPrompter and then adds one or multiple chunks of text that get spoken by the TTS engine. After the application is finished it closes the session with closePrompter.

C:\Documents and Settings\Administrator\Desktop\Image2.EMF

## Use Case Realization: Add multiple Text Chunks

A client application can add multiple text chunks to the TTS buffer that will be spoken in this order.

It is sometimes preferable to add multiple small chunks of text instead of one big chunk in order to optimize system latencies.

C:\Documents and Settings\Administrator\Desktop\Image2.EMF

## Use Case Realization: Navigation and Reader Application trying to prompt

The following sequence illustrates the behavior of SpeechOutputService in case of conflicting requests. The prompter type navigation has higher priority and thus gets access to the service.

C:\Documents and Settings\Administrator\Desktop\Image2.EMF

# Interfaces

## The following pages describe the interfaces of the SpeechOutputService API