International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

H.248.74

(04/2016)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Infrastructure of audiovisual services – Communication procedures

# Gateway control protocol: Media resource control enhancement packages

Recommendation ITU-T H.248.74

ITU-T H-SERIES RECOMMENDATIONS

**AUDIOVISUAL AND MULTIMEDIA SYSTEMS**

| | |
|---|---|
| CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS | H.100–H.199 |
| INFRASTRUCTURE OF AUDIOVISUAL SERVICES | |
| General | H.200–H.219 |
| Transmission multiplexing and synchronization | H.220–H.229 |
| Systems aspects | H.230–H.239 |
| **Communication procedures** | **H.240–H.259** |
| Coding of moving video | H.260–H.279 |
| Related systems aspects | H.280–H.299 |
| Systems and terminal equipment for audiovisual services | H.300–H.349 |
| Directory services architecture for audiovisual and multimedia services | H.350–H.359 |
| Quality of service architecture for audiovisual and multimedia services | H.360–H.369 |
| Telepresence | H.420–H.429 |
| Supplementary services for multimedia | H.450–H.499 |
| MOBILITY AND COLLABORATION PROCEDURES | |
| Overview of Mobility and Collaboration, definitions, protocols and procedures | H.500–H.509 |
| Mobility for H-Series multimedia systems and services | H.510–H.519 |
| Mobile multimedia collaboration applications and services | H.520–H.529 |
| Security for mobile multimedia systems and services | H.530–H.539 |
| Security for mobile multimedia collaboration applications and services | H.540–H.549 |
| Mobility interworking procedures | H.550–H.559 |
| Mobile multimedia collaboration inter-working procedures | H.560–H.569 |
| BROADBAND, TRIPLE-PLAY AND ADVANCED MULTIMEDIA SERVICES | |
| Broadband multimedia services over VDSL | H.610–H.619 |
| Advanced multimedia services and applications | H.620–H.629 |
| Ubiquitous sensor network applications and Internet of Things | H.640–H.649 |
| IPTV MULTIMEDIA SERVICES AND APPLICATIONS FOR IPTV | |
| General aspects | H.700–H.719 |
| IPTV terminal devices | H.720–H.729 |
| IPTV middleware | H.730–H.739 |
| IPTV application event handling | H.740–H.749 |
| IPTV metadata | H.750–H.759 |
| IPTV multimedia application frameworks | H.760–H.769 |
| IPTV service discovery up to consumption | H.770–H.779 |
| Digital Signage | H.780–H.789 |
| E-HEALTH MULTIMEDIA SERVICES AND APPLICATIONS | |
| Personal health systems | H.810–H.819 |
| Interoperability compliance testing of personal health systems (HRN, PAN, LAN, TAN and WAN) | H.820–H.859 |
| Multimedia e-health data exchange services | H.860–H.869 |

*For further details, please refer to the list of ITU-T Recommendations.*

# Recommendation ITU-T H.248.74

## Gateway Control Protocol: Media resource control enhancement packages


### Summary

Recommendation ITU-T H.248.74 provides guidelines for interworking the media resource control protocol (MRCP) version 2 (outlined in IETF RFC 6787) to the ITU-T H.248-series of Recommendations. Packages are defined to support functionality required by MRCP but not already supported by the ITU-T H.248-series of Recommendations.

This new functionality includes:

– A package to detect the start of media inputs when recording is specified;

– A package to enable the media gateway controller (MGC) to trim a period of time off the end of a recorded media file;

– A package that enables an MGC to control the sensitivity associated with a recording;

– An enhanced automatic speech recognition (ASR) package to enable the "Confidence Threshold", "N Best List" and the "Input Waveform URI" to be specified;

– An enhanced text to speech (TTS) package to enable the "Play Offset" and "Play Length" to be specified;

– A Play Offset Control package to enable the MGC to offset an already playing signal.



### History

| Edition | Recommendation | Approval | Study Group | Unique ID* |
|---|---|---|---|---|
| 1.0 | ITU-T H.248.74 | 2016-04-13 | 16 | 11.1002/1000/12635 |

### Keywords

Automatic speech recognition; media gateway protocol; media playout; media recording; media resource control enhancement packages; MEGACO; package; RTSP; speech; text to speech.


---

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

## FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

## NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

## INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

## Table of Contents

# Recommendation ITU-T H.248.74

## Gateway Control Protocol: Media resource control enhancement packages

### 1 Scope

[b-IETF RFC 4313] outlines the needs and requirements for a protocol to control the distributed media processing resources, including automatic speech recognition (ASR), speaker identification/speaker verification (SI/SV), and text to speech (TTS) resources. The media resource control protocol (MRCP) [IETF RFC 6787] is a protocol that is designed to allow a client device to control media processing resources in a network.

These media resources may reside in a decomposed media gateway (MG) utilising [ITU-T H.248.1] as a media gateway control protocol. In such a scenario there is an interworking between MRCP and ITU-T H.248.

This Recommendation provides an analysis of the interaction between the various MRCP methods, headers and events and the ITU-T H.248-series of Recommendations. It does not provide a complete interworking specification in terms of call flows due to the complexity of scenarios arising from media resource control. Many of the functions can be achieved through the mechanisms defined in the existing ITU-T H.248-Series of Recommendations (i.e., [ITU-T H.248.9] and [ITU-T H.248.66]). However, some functions require the definition of additional ITU-T H.248 packages. This Recommendation details these enhancements.

### 2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T H.248.1] Recommendation ITU-T H.248.1 (2013), *Gateway control protocol: Version 3*.

[ITU-T H.248.9] Recommendation ITU-T H.248.9 (2009), *Gateway control protocol: Advanced media server packages*.

[ITU-T H.248.66] Recommendation ITU-T H.248.66 (2016), *Gateway control protocol: Packages for RTSP and H.248 interworking*.

[IETF RFC 6787] IETF RFC 6787 (2012), *Media Resource Control Protocol Version 2 (MRCPv2)*.

### 3 Definitions

#### 3.1 Terms defined elsewhere

None.

#### 3.2 Terms defined in this Recommendation

**3.2.1 ITU-T H.248 interworking**: The process of mapping protocol elements and functionality between the ITU-T H.248-series of Recommendations and another protocol.

**3.2.2 Modify.req**: A Modify command request as defined by [ITU-T H.248.1].

**3.2.3 utterance**: In the general sense, meaningful user input in any modality, e.g., speech, dual tone multi frequency (DTMF).

## 4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AAS Advanced Audio Server

ASR Automatic Speech Recognition

DTMF Dual Tone Multi Frequency

EMMA Extensible Multimodal Annotation markup language

ID Identity

MG Media Gateway

MGC Media Gateway Controller

MRCP Media Resource Control Protocol

NL Natural Language

NLSML Natural Language Semantics Markup Language

SI Speaker Identification

SRGS Speech Recognition Grammar Specification

SSML Speech Synthesis Markup Language

SV Speaker Verification

TTS Text To Speech

URI Uniform Resource Identifier

## 5 Conventions

None.

## 6 MRCP and ITU-T H.248 interworking

This clause takes the main MRCP functions: speech synthesizer control, automatic speech recognition, recording and speaker verification and identification, and describes how the protocol elements associated with these functions may be interworked with various ITU-T H.248-series Recommendations.

### 6.1 Speech synthesizer control: TTS

The advanced audio server (AAS) package for text to speech (TTS) enhancement of [ITU-T H.248.9] provides the basic functions for the MGC to control a TTS server.

Table 1 lists all the MRCP protocol elements for the speech synthesizer resource defined in [IETF RFC 6787] and their corresponding ITU-T H.248 equivalents.

**Table 1 – Speech synthesizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| TTS | Section 8 of [IETF RFC 6787]: Speech Synthesizer Resource | Clause 13 of [ITU-T H.248.9]: Advanced Audio Server base package for TTS enhancement |

**Table 1 – Speech synthesizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| Methods (See the definitions of the methods in section 8.6-8.14 of [IETF RFC 6787]) | Method: SPEAK<br>Provides the speech text and initiates speech synthesis and streaming. | Supported by the signals defined in AAS base package for TTS enhancement in clause 13 of [ITU-T H.248.9]. |
| | Method: STOP<br>Stop speaking. | The TTS output may be stopped by removing the signal that ordered the play out from a signals descriptor via normal [ITU-T H.248.1] procedures. |
| | Method: BARGE-IN-OCCURED<br>Interrupts the speaking when a barge-in-able event is detected. | A barge-in-event may be detected by the MGC and subsequently a Modify.req command can be used by the MGC to interrupt the signal (e.g., by removing the playing signal and issuing a new signal). |
| | Method: PAUSE<br>Pauses speech output. | This is supported by the Signal Pause package in clause 13 of [ITU-T H.248.66]. |
| | Method: RESUME<br>It tells a paused synthesizer resource to resume speaking. | This is supported by the Signal Pause package in clause 13 of [ITU-T H.248.66]. |
| | Method: CONTROL<br>Requests the synthesizer to jump forward or backward in what it is speaking, change speaker rate, speaker parameters, etc. | The control indications can be specified via a Modify.req command. See clauses 14 and 15 of [ITU-T H.248.66] for the speed/scale adjustment.<br>The ability to jump from either the beginning of the announcement or the current position is supported by the "Enhanced TTS" package in clause 11 and the "Play Offset Control" package in clause 12. |
| | Method: SPEAK-COMPLETE<br>An event message to indicate that the corresponding "SPEAK" request was completed. | The "Signal Completion" event from the "Generic" package in clause E.1 of [ITU-T H.248.1] is used to notify that the signal completion has occurred. |
| | Method: SPEECH-MARKER<br>An event message to indicate that the synthesizer has encountered a marker tag in the current speech markup. | This is similar to the usage of PLAY_NOTIFY in [ITU-T H.248.66]. Text marker detection is supported by the "Mark Detection" event in clause 13.2.2 of [ITU-T H.248.9]. |
| | Method: DEFINE-LEXICON<br>It provides a lexicon and tells the server to load, unload, activate or deactivate the lexicon. | Lexicon is supported as part of the speech synthesis markup language (SSML) format when carried in the Play script signal in clause 13.3.2 of [ITU-T H.248.9]. |
| Headers (See the definitions of the headers in clause 8.4 of [IETF RFC 6787]) | Header: Jump-Size<br>Used in a "SPEAK" request: as a desired offset into the synthesized speech.<br>Used in an active "Control" request: to control the amount to jump forward or backward. | The "SPEAK" request is supported by the parameter "Offset" of the Advanced audio server base package in clause 8 of [ITU-T H.248.9].<br>The "Control" request is supported by [ITU-T H.248.66].<br>[ITU-T H.248.66] supports text based playback control. |

**Table 1 – Speech synthesizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Header: Kill-On-Barge-In<br>Enables kill-on-barge-in support. If enabled, the client can notify the BARGE-IN-OCCURRED event to the server. | As [ITU-T H.248.1] follows a master-slave model, the MGC can stop or interrupt the TTS signal at any time. |
| | Header: Completion Cause<br>It indicates the reason the "SPEAK" request completed. | To indicate the completion cause Termination method, the "Signal Completion" event (Clause E.1 of [ITU-T H.248.1]) can be used. |
| | Header: Completion Reason<br>It contains the reason text behind the "SPEAK" request completion. | See "Completion Cause" row above. |
| | Header: Fetch Hint<br>It controls the URI access properties when the server needs to fetch some documents. | The MRCP "Prefetch" element is not applicable to ITU-T H.248. In ITU-T H.248, it is the MGC that controls when downloading is appropriate. |
| | Header: Audio Fetch Hint<br>It controls the URI access properties of audio files. | This is not applicable to ITU-T H.248. |
| | Header: Speak Restart<br>It controls the restart behaviours. | This is supported by the "Playback Relative Scale Adjustment" package in [ITU-T H.248.66]. |
| | Header: Speak Length<br>It controls the maximum length of speech to speak. | ITU-T H.248 allows the specification of signal duration which is effectively the length of the playout in milliseconds.<br>The "Enhanced TTS" package in clause 11 allows the specification of length in terms of grammar (e.g., word/sentence/paragraph). |
| | Header: Speaker Profile<br>To specify a URI which references the profile of the speaker. | [ITU-T H.248.9] allows the MGC to specify a collection of voice parameters by referencing provisioned segment specifications. |
| | Header: Voice-Parameter<br>It contains a set of parameters, including voice gender, voice age, voice variant and voice name, which define the voice of the speaker. | This is supported by the use of an SSML script. The MGC utilises SSML scripts in [ITU-T H.248.9]. |
| | Header: Prosody-Parameters<br>It contains a set of parameters, including pitch, timing, pausing, speaking rate, emphasis on words and many other features, which define the prosody of the speech. | This is supported by the use of an SSML script. The MGC utilises SSML scripts in [ITU-T H.248.9]. |
| | Header: Speech Language<br>Specifies the default language of the speech data. | This is supported by the use of an SSML script. The MGC utilises SSML scripts in [ITU-T H.248.9]. |
| | Header: Load-Lexicon<br>Header: Lexicon-Search-Order | This is supported by the use of an SSML script. The MGC utilises SSML scripts in |

**Table 1 – Speech synthesizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | These relates to the usage of the lexicons. | [ITU-T H.248.9].The order is given by place in the script. |
| | Header: Speech Marker<br>It contains the marker tag and timestamp information related to the bookmarker that the playout encounters. | Text marker detection is supported by the "Mark Detection" event in clause 13.2.2 of [ITU-T H.248.9]. |

## 6.2 Automatic speech recognition: ASR

In MRCP, the "recog-only" method and the "enrolment" method are defined for speech recognizer functionality. It is optional for a recognizer resource to support voice enrolled grammars and the enrolment related parameters. [ITU-T H.248.9] provides the ASR package and AAS Digit Collection package to implement the voice and digit recognition function.

Automatic speech recognition functionality in ITU-T H.248 utilises the Extensible MultiModel Annotation markup language (EMMA) to return the results of recognition. MRCP utilises the Natural Language Semantics Markup Language (NLSML) (see section 16 of [IETF RFC 6787] to return the results of recognition. However, EMMA is unable to return Enrolment results. Rather than requiring a new scheme to be supported by an ITU-T H.248 MG, EMMA is utilised to return the results. Where additional enrolment information is needed this Recommendation defines additional Events/ObservedEvents to support these.

Table 2 shows all the protocol elements listed in the speech recognizer part of MRCP and their corresponding ITU-T H.248 equivalents.

**Table 2 – Speech recognizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| ASR | Section 9 of [IETF RFC 6787]: Speech Recognizer Resource | Clause 12 of [ITU-T H.248.9]: ASR package |
| Recognizer Methods (See the definitions of the methods in section 9.8-9.20 of [IETF RFC 6787]) | Recog-only-method: DEFINE-GRAMMAR<br>Provides one or more grammars and requests the server to access, fetch, and compile the grammars as needed. | This is supported by the "Recognition Grammar Script/Identifier", "Recognition Grammar Script Type" and "Recognition Grammar Script Format" parameters of the ASR signal defined in clause 12 of [ITU-T H.248.9]. |
| | Recog-only-method: RECOGNIZE<br>Start recognition and provides it with one or more grammar references for grammars. | This is supported by the ASR recognition signals in clauses 12.3.1 and 12.3.2 of [ITU-T H.248.9]. |
| | Recog-only-method: INTERPRET<br>Takes as input an interpret-text header containing the text for which the semantic interpretation is desired. | This is supported through the use of speech recognition grammar specification (SRGS) with the ASR recognition signals in clauses 12.3.1 and 12.3.2 of [ITU-T H.248.9]. For further information regarding interpretation capabilities of SRGS see clause 1.4 of [b-W3C SRGS]. |

**Table 2 – Speech recognizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Recog-only-method: GET-RESULT<br>Retrieves results for a completed recognition. | The MGC receives the results of the ASR via the "Automatic Speech Recognition Success" event in clause 12.2.2 of [ITU-T H.248.9].<br>When "Voice enrolment" is used, further results may be returned via the "Enrolment Results" event in the "Voice Enrolled Grammar" package (see clause 13.2.2). |
| | Recog-only-method: START-INPUT-TIMERS<br>Starts the timers when the client wants. | The MGC can set the timers using the relevant parameters of the ASR recognition signals in clauses 12.3.1 and 12.3.2 of [ITU-T H.248.9]. |
| | Recog-only-method: STOP<br>Stops recognition if a request is active. | The ASR output may be stopped by removing the signal that ordered the play out from a signals descriptor via normal [ITU-T H.248.1] procedures. |
| | Enrollment-method: START-PHRASE-ENROLLMENT<br>Starts a new phrase enrolment session. | This is supported by the "Enrolment Start" signal in "Voice Enrolled Grammar" package (clause 13.3.2). |
| | Enrollment-method: ENROLLMENT-ROLLBACK<br>Discards the last live utterance from the RECOGNIZE operation. | This is supported by the "Enrolment Rollback" signal in "Voice Enrolled Grammar" package (clause 13.3.3). |
| | Enrollment-method: END-PHRASE-ENROLLMENT<br>Ends an active phrase enrolment session. | This is supported by the "End Phrase Enrolment" signal in "Voice Enrolled Grammar" package (clause 13.3.4). |
| | Enrollment-method: MODIFY-PHRASE<br>Is used to change the phrase ID, natural language phrase (phrase-NL) and/or weight for a given phrase in a personal grammar. | This is supported by the "Modify Phrase" signal in "Voice Enrolled Grammar" package (clause 13.3.5). |
| | Enrollment-method: DELETE-PHRASE<br>Is used to delete a phase in a personal grammar. | This is supported by the "Delete Phrase" signal in "Voice Enrolled Grammar" package (clause 13.3.6). |
| Recognizer events | Recognizer-event: START-OF-INPUT<br>Indicates that the recognition resource has detected speech or a DTMF digit in the media stream. | Dual tone multi frequency (DTMF) detection is supported by [ITU-T H.248.1]. The detection of media input is supported by the "Media Start" package in clause 7. |
| | Recognizer-event: RECOGNITION-COMPLETE<br>Indicates that the recognition is completed. | This is supported by the "Automatic Speech Recognition Failure" and "Automatic Speech Recognition Success" events of the ASR package (clause 12 of [ITU-T H.248.9]. |

**Table 2 – Speech recognizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | | When "Voice enrolment" is used further results may be returned via the "Automatic Speech Recognition Failure" event (clause 13.2.1) and the "Enrolment Results" event (clause 13.2.2) in the "Voice Enrolled Grammar" package. |
| | Recognizer-event: INTERPRETATION-COMPLETE<br>Indicates that the INTERPRET operation is complete. | This is supported by the "Automatic Speech Recognition Failure" and "Automatic Speech Recognition Success" events of the ASR package (clause 12 of [ITU-T H.248.9]. An EMMA script [b-W3C EMMA] may contain interpreted results. |
| Headers | Header: Confidence Threshold<br>Used to tell the recognizer resource what confidence level the client considers a successful match. | The confidence score can be carried in the EMMA format result. The threshold is supported by the "Enhanced ASR" package in clause 10. |
| | Header: Sensitivity Level<br>Allows the client to set the sensitivity level for the recognizer. | This is supported by the "Recognition Sensitivity" parameters associated with signals in the Automatic Speech Recognition package" (clause 12.3 of [ITU-T H.248.9]). |
| | Header: Speed vs Accuracy<br>Indicates the performance or precision of the recognition. | This is supported by the "Recognition Precision" parameters associated with signals in the Automatic Speech Recognition package" (clause 12.3 of [ITU-T H.248.9]). |
| | Header: N Best List Length<br>Asks the recognition resource to send more than 1 alternative. | This supported by the "Enhanced ASR" package in clause 10. EMMA can also support multiple recognition results. |
| | Header: Input Type<br>Specifies what the input that caused the barge-in event was, DTMF or speech. | Via the received ObservedEvent the MGC will know the type of event that was generated by the MG. |
| | Header: No Input Timeout<br>Indicates the time to wait for input before a timeout occurs. | This is supported by the "Waiting Time for Input" parameters associated with signals in the "Automatic Speech Recognition" package (clause 12.3 of [ITU-T H.248.9]). |
| | Header: Recognition Timeout<br>Specifies a certain period of time for the detection of no match when recognition was already started. | This is supported by the "Maximum Recognition Time" parameters associated with signals in the "Automatic Speech Recognition" package (clause 12.3 of [ITU-T H.248.9]). |
| | Header: Waveform URI<br>Records the incoming audio stream into a stored file. | This is supported by the "Record File" parameter of the "Automatic Speech Recognition Start With Grammar Script" signal (clause 12.3.1.1.11 of [ITU-T H.248.9]), and also the "Recording Identifier" parameter of the AAS Recording package (clause 10.3.1.1.9 of [ITU-T H.248.9]).<br>When used with enrolment the WaveForm URI may also be returned in response to a request to end the enrolment session. This is supported by the "Recorded File" |

**Table 2 – Speech recognizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | | parameter of the "Voice Enrolled Grammar" package (clause 13.3.4.1.2). |
| | Header: Media Type<br>Specifies the Media Type in which to store captured, audio or video. | The media type is specified by the file extension of the media file used to store the recording. The properties of the local and remote descriptors also specify whether media is audio or video. |
| | Header: Input-Waveform-URI<br>Specifies a URI pointing to audio content to be processed. | [ITU-T H.248.9] only supports user input voice recognition, not the URI form. The "Enhanced ASR" package in clause 10 supports an URI form. |
| | Header: Failed URI<br>Indicates the URI that the server failed to access. | Due to the [ITU-T H.248.1] command request and response mechanism the MGC knows the URI that it indicated to the MG. |
| | Header: Failed URI Cause<br>Indicates the specific URI or protocol specific response code for the URI in the Failed-URI header. | The "Enhanced ASR" package in clause 10 contains an event to indicate the cause. |
| | Header: Completion Cause<br>Indicates the reason behind the Recognize method completion. | Supported by the "Automatic Speech Recognition Failure" event in clause 12.2 of [ITU-T H.248.9]. |
| | Header: Completion Reason<br>This contains the reason text behind the "SPEAK" request completion. | Supported by the "Automatic Speech Recognition Failure" event in clause 12.2 of [ITU-T H.248.9]. |
| | Header: Recognizer Context Block<br>Contains the recognizer context block. | This is a vendor specific element, which can be implemented by the vendor's own packages. |
| | Header: Start Input Timers<br>Indicates whether or not to start the no-input timer. | If an MGC does not want to start the timers, it can issue a speech recognition signal (see clause 12.3 of [ITU-T H.248.9]) without the timer parameters. |
| | Header: Speech Complete Timeout<br>Specifies the length of silence required following user speech before the speech recognizer finalizes a result. | This is supported by the "Post Speech Timer" parameters with signals in the "Automatic Speech Recognition" package (clause 12.3 of [ITU-T H.248.9]). |
| | Header: Speech Incomplete Timeout<br>A timer to detect an incomplete match. | This is supported by the "Post Speech Timer" parameters associated with signals in the "Automatic Speech Recognition" package (clause 12.3 of [ITU-T H.248.9]). This package makes no distinction between incomplete or complete speech. |
| | Header: DTMF Interdigit Timeout<br>A timer to detect inter-digit timeout. | The inter-digit timer is supported via the digit map syntax in the "Digit Map" parameter in the "AAS Digit Collection" package (clause 9.3.1.1.10 of [ITU-T H.248.9]). |

**Table 2 – Speech recognizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Header: DTMF Term Timeout<br>Specifies the terminating timeout to use when recognizing DTMF input. | The [ITU-T H.248.1] digit map syntax specifies this capability. |
| | Header: DTMF-Term-Char<br>Indicates the terminating DTMF character. | This is supported by the "End Input Key" parameter of "AAS Digit Collection" package (clause 9.3.1.1.19 of [ITU-T H.248.9]). |
| | Header: Save Waveform<br>A Boolean header to request the recognizer resource to save the audio input. | This is supported by via the signals in the "AAS Recording" package (clause 10.3 of [ITU-T H.248.9]). |
| | Header: New Audio Channel<br>A Boolean header to indicate that further input audio comes from a different audio source, channel or speaker. | The MGC can issue an appropriate "Automatic Speech Recognition" package signal to initiate recording from a new source. The MGC may also be required to update the appropriate stream descriptors for a new source. |
| | Header: Speech-Language<br>Specifies the language of recognition grammar data. | This is supported by the "Recognition Grammar Script" parameter (clause 12.3.1.1.2 of [ITU-T H.248.9] that enables the use of SRGS. SRGS supports the specification of the language. |
| | Header: Ver-Buffer-Utterance<br>Buffers the utterance associated with this recognition request into a buffer available to a co-resident verifier resource. | This is supported via the "Ver Buffer Utterance" parameter in clause 10.3.1.1.4. |
| | Header: Recognition-Mode<br>Specifies what mode the RECOGNIZE method will operate in: "normal" or "hotword". | This is supported by the "Recognition Mode" parameters associated with signals in the "Automatic Speech Recognition" package (clause 12.3 of [ITU-T H.248.9]). |
| | Header: Cancel-If-Queue<br>Specifies whether the server should terminate the processing Recognize request if another request is received. | KeepActive flag defined in [ITU-T H.248.1] can be used to control the playing signal. |
| | Header: Hotword-Max-Duration<br>Specifies the maximum length of an utterance (in seconds) that will be considered for "hotword" recognition. | This is supported by the "Recognition Grammar Script" parameter (clause 12.3.1.1.2 of [ITU-T H.248.9] that enables the use of SRGS. SRGS supports the specification of the maximum "hot word" duration. |
| | Header: Hotword-Min-Duration<br>Specifies the minimum length of an utterance (in seconds) that will be considered for "Hotword" recognition. | This is supported by the "Recognition Grammar Script" parameter (clause 12.3.1.1.2 of [ITU-T H.248.9] that enables the use of SRGS. SRGS supports the specification of the minimum "hotword" duration. |

**Table 2 – Speech recognizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Header: Interpret-Text<br>Provides a pointer to the text for which a natural language interpretation is desired. | This is supported through the use of SRGS with the "ASR recognition with grammar identifier" signal in clause 12.3.2 of [ITU-T H.248.9]. For further information regarding interpretation capabilities of SRGS see clause 1.4 of [b-W3C SRGS]. |
| | Header: DTMF-Buffer-Time<br>Specifies the size in time of the "type-ahead" buffer for the recognizer. | The Advanced Audio Server packages defined in [ITU-T H.248.9] supports type-ahead by default. Type-ahead can be turned off by specifying that digit collection begins only after the initial prompt has been played out. The MGC can turn on and off the Event buffer (i.e., by adding and removing a digit detection event) at the required time interval. |
| | Header: Clear-DTMF-Buffer<br>Clears the DTMF "type-ahead" buffer before starting the recognition. | This is supported by the "Clear Digit Buffer" parameter of "AAS Digit Collection" package (clause 9.3.1.1.8 of [ITU-T H.248.9]). |
| | Header: Early-No-Match<br>Indicates to the recognizer whether or not to do early matching. | Enable/disable early matching is not supported via ITU-T H.248. It is the MG that decides whether or not to use early matching. |
| | Header: Num-Min-Consistent-Pronunciations<br>Specifies the minimum number of consistent pronunciations that must be obtained to voice enrol a new phrase. | This is supported by the "NumMinConsistentPronunciations" parameter of the "Voice Enrolled Grammar" package (clause 13.3.2.1.2). |
| | Header: Consistency-Threshold<br>Specifies how similar to a previously enrolled pronunciation of the same phrase an utterance needs to be in order to be considered "consistent." | This is supported by the "Consistency Threshold" parameter of the "Voice Enrolled Grammar" package (clause 13.3.2.1.3). |
| | Header: Clash-Threshold<br>Specifies how similar the pronunciations of two different phrases can be before they are considered to be clashing. | This is supported by the "Clash Threshold" parameter of the "Voice Enrolled Grammar" package (clause 13.3.2.1.4). |
| | Header: Personal-Grammar-URI<br>Specifies the speaker-trained grammar to be used or referenced during enrolment operations. | This is supported by the "Personal Grammar URI" parameter of the "Voice Enrolled Grammar" package (clause 13.3.2.1.1). |
| | Header: Enroll-Utterance<br>Enables the addition of the collected utterance to the personal grammar. | This is supported by the "Enrol Utterance" parameter of the "Voice Enrolled Grammar" package (clause 13.3.1.1.3). |

**Table 2 – Speech recognizer related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Header: Phrase-Id<br>Identifies a phrase in an existing personal grammar for which enrolment is desired. | This is supported by the "Phrase Identity" parameter defined in the "Voice Enrolled Grammar" package (clause 13). With which signal/event the parameter is used is dependent on the MRCP method. |
| | Header: Phrase-NL<br>Specifies the interpreted text to be returned when the phrase is recognized. | This is supported by the "PhraseNL" parameter defined in the "Voice Enrolled Grammar" package (clause 13). With which signal the parameter is used is dependent on the MRCP method. |
| | Header: Weight<br>Represents the occurrence likelihood of a phrase in an enrolled grammar. | This is supported by the "Weight" parameter defined in the "Voice Enrolled Grammar" package (clause 13). With which signal the parameter is used is dependent on the MRCP method. |
| | Header: Save-Best-Waveform<br>Request the recognizer resource to save the audio stream for the best repetition of the phrase that was used during the enrolment session. | This is supported by the "Save Best Waveform" parameter of the "Voice Enrolled Grammar" package (clause 13.3.2.1.8). |
| | Header: New-Phrase-Id<br>Replaces the ID used to identify the phrase in a personal grammar. | This is supported by the "New Phrase Id" parameter of the "Voice Enrolled Grammar" package (clause 13.3.5.1.3). |
| | Header: Confusable-Phrases-URI<br>Specifies a grammar that defines invalid phrases for enrolment. | This is supported by the "Confusable phrases URI" parameter of the "Voice Enrolled Grammar" package (clause 13.3.1.1.2). |
| | Header: Abort-Phrase-Enrolment<br>Aborts the phrase enrolment session. | This is supported by the "Abort Phrase Enrolment" parameter of the "Voice Enrolled Grammar" package (clause 13.3.4.1.1). |

## 6.3 Recording

In MRCP, the recorder resource captures received audio and video and stores it as content pointed to by a URI. [ITU-T H.248.9] provides recording functionality by means of the AAS Recording package and the Multimedia Recording package.

Table 3 shows all the protocol elements listed in the recording part of [IETF RFC 6787] and their corresponding ITU-T H.248 equivalents.

**Table 3 – Recording related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| Recording | Section 10 of [IETF RFC 6787]: Recorder Resource. | Clause 10 of [ITU-T H.248.9]: AAS Recording package<br>Clause 15 of [ITU-T H.248.9]: Multimedia Recording package. |

**Table 3 – Recording related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| Recorder method | recorder-method: RECORD<br>Places the recorder resource in the Recording state. | Supported by the "PlayRecord" signals defined in the "AAS Recording" (clause 10.3.1 of [ITU-T H.248.9]) and "Multimedia Recording" (clause 15.3.1 of [ITU-T H.248.9] packages. |
| | recorder-method: STOP<br>Moves the recorder from the recording state back to the idle state. | The recording may be stopped by removing the signal that ordered the play out from a signals descriptor via normal [ITU-T H.248.1] procedures. |
| | recorder-method: START-INPUT-TIMERS<br>Starts the timers when the client wants. | The MGC can set the timers using the relevant parameters of the "Play Record" signal in the "AAS Recording" or "Multimedia Recording" packages. |
| Recorder events | recorder-event: START-OF-INPUT<br>This event is always returned by the recording resource when speech has been detected. | This is supported by the "Media Start" package in clause 7. |
| | recorder-event: RECORD-COMPLETE<br>Indicates the completion of the recording. | This is supported by the "Audio operation failure" and the "PlayRecord success" events of the "AAS Recording" (clause 10.2 of [ITU-T H.248.9]) package. |
| Headers | Header: Sensitivity Level<br>Allows the client to set the variable level of sound sensitivity for the recorder. | This is supported by the "Enhanced Recording" package in clause 9. |
| | Header: No Input Timeout<br>Indicates whether or not to start the no-input timer. | This is supported by the "Pre-Speech Timer" parameter of the "Play Record" signal in the "AAS Recording" (clause 10.3.1.1.6 of [ITU-T H.248.9]) or "Multimedia Recording" (clause 15.3.1.1.6 of [ITU-T H.248.9]) packages. |
| | Header: Completion Cause<br>Indicates the reason behind the Record method completion. | This is supported by the events defined in the "AAS Recording" package (clause 10 of [ITU-T H.248.9]. |
| | Header: Completion Reason<br>Contains the reason text behind the Record request completion. | This is supported by the events defined in the "AAS Recording" package (clause 10 of [ITU-T H.248.9]. |
| | Header: Failed URI<br>Contains the URI that the server failed to access. | Due to the [ITU-T H.248.1] command request and response mechanism the MGC knows the URI that it indicated to the MG. |
| | Header: Failed URI Cause<br>The URI specific or protocol specific response code for the URI in the Failed-URI header. | The "Enhanced ASR" package in clause 10 contains an event to indicate the cause. |
| | Header: Record URI | The URI is supported by the "Recording Id" parameter of the "Play Record success" event in the "AAS Recording" (clause 10.2.2.2.4 of [ITU-T H.248.9]). |

**Table 3 – Recording related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Specifies the URI where the speech is stored, the size and duration of the record. | The duration is supported by the "Record duration" parameter of the "Play Record success" event in the "AAS Recording" (clause 10.2.2.2.5 of [ITU-T H.248.9]).<br>The size is supported by the "Recording Length" parameter of the "Play Record success" event in the "AAS Recording" (clause 10.2.2.2.7 of [ITU-T H.248.9]). |
| | Header: Media Type<br>Specifies to the server the Media Type of the captured audio or video. | The media type was specified by the file extension of the media file used to store the recording. |
| | Header: Max Time<br>Specifies the maximum length of the recording in milliseconds. | This is supported by the "Record Length Timer" parameter of the "Play Record" signal in the "AAS Recording" (clause 10.3.1.1.8 of [ITU-T H.248.9]) or "Multimedia Recording" (clause 15.3.1.1.8 of [ITU-T H.248.9]) packages. |
| | Header: Trim-Length<br>Specifies the length of audio to be trimmed from the end of the recording after recording has stopped. | This is supported by the "Trim" package in clause 8. |
| | Header: Final Silence<br>Specifies the length of silence in the audio that is to be interpreted as the end of the recording. | This is supported by the "Post Speech Timer" parameter of the "Play Record" signal in the "AAS Recording" (clause 10.3.1.1.7 of [ITU-T H.248.9]) or "Multimedia Recording" (clause 15.3.1.1.7 of [ITU-T H.248.9]) packages. |
| | Header: Capture On Speech<br>Indicates whether to wait for the end-pointing functionality to detect speech before it starts capturing. | This is supported by the setting a value of "0" for the "Pre-Speech Timer" parameter of the "Play Record" signal in the "AAS Recording" (clause 10.3.1.1.6 of [ITU-T H.248.9]) or "Multimedia Recording" (clause 15.3.1.1.6 of [ITU-T H.248.9]) packages.. |
| | Header: Ver-Buffer-Utterance<br>Buffers the utterance associated with this recording request into the verification buffer. | This is supported via the "Ver Buffer Utterance" parameter in clause 10.3.1.1.4. |
| | Header: Start Input Timers<br>A Boolean header to indicate whether to enable the no-input timer. | If an MGC does not want to start the timers, it can issue a "Play Record" signal (see clauses 10.3.1 of [ITU-T H.248.9] and 10.3.1 of [ITU-T H.248.9]) without the timer parameters. |
| | Header: New Audio Channel<br>A Boolean header to indicate that further input audio comes from a different audio source, channel or speaker. | The MGC can issue an appropriate "AAS Recording" package signal to initiate recording from a new source. The MGC may also be required to update the appropriate stream descriptors for a new source. |

## 6.4 Speaker verification and identification

Table 4 shows all the protocol elements listed in the speaker verification and identification part of MRCP and their corresponding ITU-T H.248 equivalents.

**Table 4 – Speaker verification and identification related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| Speaker Verification and Identification | Section 11 of [IETF RFC 6787]: Speaker Verification and Identification. | Clause 14 of this Recommendation. |
| Speaker Verification Methods (See the definitions of the methods in section 11.6-11.18 of [IETF RFC 6787]) | Method: START-SESSION<br>Starts a Speaker Verification or Identification session. | This is supported by the "Start Session" signal in "Speaker Verification and Identification" package (clause 14.3.1). |
| | Method: END-SESSION<br>Terminates an ongoing verification session and releases the verification voiceprint resources. | This is supported by the "End Session" signal in "Speaker Verification and Identification" package (clause 14.3.10). |
| | Method: QUERY-VOICEPRINT<br>Gets status information on a particular voiceprint and can be used to ascertain if a voiceprint or repository exists and if it contains trained voiceprints. | This is supported by the "Query Voiceprint" signal in "Speaker Verification and Identification" package (clause 14.3.4). |
| | Method: DELETE-VOICEPRINT<br>Removes a voiceprint from a repository. | This is supported by the "Delete Voiceprint" signal in "Speaker Verification and Identification" package (clause 14.3.5). |
| | Method: VERIFY<br>Requests the verifier resource to either train/adapt the voiceprint or to verify/identify a claimed identity. | This is supported by the "Verify" signal in "Speaker Verification and Identification" package (clause 14.3.2). |
| | Method: VERIFY-FROM-BUFFER<br>Directs the verifier resource to verify buffered audio against a voiceprint. | This is supported by the "Verify from Buffer" signal in "Speaker Verification and Identification" package (clause 14.3.3). |
| | Method: VERIFY-ROLLBACK<br>Discards the last buffered utterance or discards the last live utterances. | This is supported by the "Verify Rollback" signal in "Speaker Verification and Identification" package (clause 14.3.6). |
| | Method: STOP<br>Stops the VERIFY or VERIFY-FROM-BUFFER request if one is active. | This is supported by the "Stop" signal in "Speaker Verification and Identification" package (clause 14.3.7). |
| | Method: START-INPUT-TIMERS<br>Starts the timers when the client wants. | The MGC can set the timers using the relevant parameters of the Speaker Verification signals in clauses 14.3.1. |
| | Method: CLEAR-BUFFER<br>Clears the verification buffer. | This is supported by the "Clear Buffer" signal in "Speaker Verification and Identification" package (clause 14.3.8). |
| | Method: GET-INTERMEDIATE-RESULT | This is supported by the "Get Intermediate Result" signal in |

**Table 4 – Speaker verification and identification related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Polls for intermediate results of a verification request that is in progress. | "Speaker Verification and Identification" package (clause 14.3.9). |
| Verification events | Verification-Event: VERIFICATION-COMPLETE<br>Indicates that the recognition resource has detected speech or a DTMF digit in the media stream. | This is supported by the "Verification Failure" and "Verification Results" events of the Speaker Verification and Identification package in clause 14. |
| | Verification-Event: START-OF-INPUT<br>Indicates that the speech has been detected. | The detection of media input is supported by the "Media Start" package in clause 7. |
| Headers | Header: Repository-URI<br>Specifies the voiceprint repository to be used or referenced during speaker verification or identification operations. | This is supported by the "Repository-URI" parameter of the "Start Session", "Query Voiceprint" and "Delete Voiceprint" signals in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Voiceprint-Identifier<br>Specifies the claimed identity for verification applications. | This is supported by the "Voiceprint Identifier" parameter of the "Start Session", "Query Voiceprint" and "Delete Voiceprint" signals in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Verification-Mode<br>Specifies the mode of the verifier resource. | This is supported by the "Verification Mode" parameter of the "Start Session" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Adapt-Model<br>Indicates the desired behaviour of the verifier resource after a successful verification operation. | This is supported by the "Adapt Model" parameter of the "Start Session" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Abort-Model<br>Indicates the desired behaviour of the verifier resource upon session termination. | This is supported by the "Abort Model" parameter of the "End Session" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Min-Verification-Score<br>Determines the minimum verification score for which a verification decision of "accepted" may be declared. | This is supported by the "Min Verification Score" parameter of the "Start Session" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Num-Min-Verification-Phrases<br>Specifies the minimum number of valid utterances before a positive decision is given for verification. | This is supported by the "Num Min Verification Phrases" parameter of the "Start Session" signal in the "Speaker Verification and Identification" package in clause 14. |

**Table 4 – Speaker verification and identification related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Header: Num-Max-Verification-Phrases<br>Specifies the number of valid utterances required before a decision is forced for verification. | This is supported by the "Num Max Verification Phrases" parameter of the "Start Session" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: No-Input-Timeout<br>Sets the length of time from the start of the verification timers (see START-INPUT-TIMERS) until the declaration of a no-input event in the VERIFICATION-COMPLETE event. | This is supported by the "Waiting Time for Input" parameter of the "Verify" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Save-Waveform<br>Requests the verifier resource to save the audio stream that was used for verification/identification. | This is supported by the "Save Waveform" parameter of the "Verify" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Media Type<br>Specifies the Media Type of the captured audio or video. | The media type is specified by the file extension of the media file used to store the recording. The properties of the local and remote descriptors also specify whether media is audio or video. |
| | Header: Waveform URI<br>Specifies the URI of the recorded incoming audio streams. | This is supported by the "Waveform URI" parameter of the "Verification Results" event in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Voiceprint-Exists<br>Indicates the status of the voiceprint. | This is supported by the "Voiceprint Exist" parameter of the "Query Voiceprint" and "Delete Voiceprint" signals in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Ver-Buffer-Utterance<br>Indicates that this utterance could be later considered for Speaker Verification. | This is supported by the "Ver Buffer Utterance" parameter of the "Verify" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Input-Waveform-URI<br>Specifies a URI pointing to audio content to be processed. | This is supported by the "Input Waveform URI" parameter of the "Verify" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Completion Cause<br>Indicates the cause of VERIFY or VERIFY-FROM-BUFFER method completion. | Supported by the "Verification Failure" and "Verification Results" events in clause 14. |
| | Header: Completion Reason<br>Contains the reason text behind the VERIFY request completion. | Supported by the "Verification Failure" and "Verification Results" events in clause 14. |

**Table 4 – Speaker verification and identification related interworking**

| Item | MRCP | ITU-T H.248 interworking |
|---|---|---|
| | Header: Speech Complete Timeout<br>Specifies the length of silence required following user speech before it finalizes a result. | This is supported by the "Post Speech Timer" parameter of the "Verify" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: New Audio Channel<br>A Boolean header to indicate that further input audio comes from a different audio source, channel or speaker. | The MGC can issue an appropriate "Verify" signal to initiate verify from a new source. The MGC may also be required to update the appropriate stream descriptors for a new source. |
| | Header: Abort-Verification<br>Indicates whether to abort a VERIFY method in progress. | This is supported by the "Abort Verification" parameter of the "Stop" signal in the "Speaker Verification and Identification" package in clause 14. |
| | Header: Start Input Timers<br>Indicates whether to start the no-input timer. | If an MGC does not want to start the timers, it can issue a verify signal without the timer parameters. |

## 7 Media start package

**Package name**: Media Start package

**Package ID**: mstart (0x00fa)

**Description**: When used in conjunction with an ITU-T H.248 signal that enables recording (e.g., the "Play Record" signal from [ITU-T H.248.9]). This package enables an MGC to be notified that the MG has detected the start of media input (e.g., speech).

**Version**: 1

**Extends**: None

### 7.1 Properties

None.

### 7.2 Events

#### 7.2.1 Start of media

**Event name**: Start of Media

**Event ID**: som (0x0001)

**Description**: This event is enabled to detect the start of media input. The MG shall notify the ObservedEvent to the MGC when the MG detects the start of the input media stream.

##### 7.2.1.1 EventDescriptor parameters

None.

##### 7.2.1.2 ObservedEventDescriptor parameters

None.

### 7.3 Signals

None.

### 7.4 Statistics

None.

### 7.5 Error codes

None.

### 7.6 Procedures

The "Start of Media" (*mstart/som*) event may be used when the MG performs a recording operation. The use of the *mstart/som* event by itself does not instantiate recording. An appropriate recording signal (i.e., the "Play Record" signal from [ITU-T H.248.9]) shall be used to initiate recording. If the *mstart/som* event is set, once the recording has started (via a signal) the MG shall detect the start of the media stream (i.e., user speech). Once the MG detects the start of media input by the user, the *mstart/som* ObservedEvent is notified to the MGC.

## 8 Trim package

**Package name**: Trim package

**Package ID**: trim (0x00fb)

**Description**: An MGC may request an MG to record media to a certain file. Several methods exist for the determining when to stop recording to the file. One such method is to detect the end of input through a post speech timer. Thus in some circumstances the recorded media file may have superfluous data for recording silence periods. This package allows the MGC to "trim" or remove data associated with this superfluous period of time.

This package may be used in any situation where a time period needs to be trimmed from the end of a media file. See section 10.4.10 of [IETF RFC 6787] for MRCP related usage.

**Version**: 1

**Extends**: None

### 8.1 Properties

None.

### 8.2 Events

None.

### 8.3 Signals

#### 8.3.1 Trim

**Signal name**: Trim

**Signal ID**: trim (0x0001)

**Description**: This signal requests the MG to remove a number of milliseconds from the end of a recorded media file.

**Signal Type**: Brief

**Duration**: Not applicable to BR signals.

#### 8.3.1.1 Additional parameters

##### 8.3.1.1.1 Media file

**Parameter name**: Media File

**Parameter ID**: mf (0x0001)

**Description**: This parameter indicates the URI of the media file on which the trim operates.

**Type**: String

**Optional**: No

**Possible values**: A physical segment identifier satisfying the syntax of clause 6.2.5.2 of [ITU-T H.248.9]. If the identifier is an http:// URI, it must not have a query part.

**Default**: None

##### 8.3.1.1.2 Trim length

**Parameter name**: Trim Length

**Parameter ID**: tl (0x0002)

**Description**: This parameter contains the length of time to be trimmed from the end of the media file. It is specified in millisecond units. See section 10.4.10 of [IETF RFC 6787] for more information.

**Type**: Integer

**Optional**: No

**Possible values**: Any positive value (in milliseconds)

**Default**: 0

### 8.4 Statistics

None.

### 8.5 Error codes

None.

### 8.6 Procedures

When recording a playing media stream, the MGC may have to stop the recording and trim the recorded file if the MGC does not want to record the contents any longer. In such cases, the MGC sends the Trim (*trim/trim*) signal with Trim Length (*tl*) and Media File (*mf*) parameters, which indicates the MG to remove X end ms specified by the *tl* parameter from the media file. The MG then trims X ms off the end of the file identified by the *mf* parameter.

If the MG is unable to find the media file associated with the *mf* parameter it shall return error code 606 "Unknown segment ID". If the MG is unable to trim the required time from the media file it shall return error code 449 "Unsupported or Unknown Parameter or Property Value".

Using the *trim/trim* signal to remove all the time associated with the media file is not used to remove the media file. Management of the media files is achieved through the use of the "Advanced audio server segment management package" [ITU-T H.248.9].

## 9 Enhanced recording package

**Package name**: Enhanced Recording package

**Package ID**: eaasrec (0x00fc)

**Description**: This package enables the MGC to associate a recording sensitivity metric with the "Play Record" operation. The sensitivity metric allows the MGC to set the level of sensitivity if the MG supports variable sound sensitivity. By supporting variable sound sensitivity, the MG can filter out background noise in order to detect speech.

It also allows the MGC to indicate if the recording is placed into a buffer for further processing.

**Version**: 1

**Extends**: aasrec (0x0035) version 1.

### 9.1 Properties

None.

### 9.2 Events

None.

### 9.3 Signals

#### 9.3.1 PlayRecord

**Signal name**: PlayRecord

**Signal ID**: playrec (0x0002)

**Description**: As per clause 10.3.1 of [ITU-T H.248.9], with the addition of a sensitivity parameter and a parameter to buffer the recording.

**Signal Type**: Defaults to

**Duration**: Defaults to 30000 (5 minutes) or as provisioned for the termination.

##### 9.3.1.1 Additional parameters

###### 9.3.1.1.1 Recording sensitivity

**Parameter name**: Recording Sensitivity

**Parameter ID**: rs (0x0034)

**Description**: To filter out background noise so that it is not mistaken for speech, the MG, acting as a recorder, may support a variable level of sound sensitivity. The sensitivity level is an integer value between 0 and 100, if the sensitivity is set higher, the recording is less affected by the environment. A higher value will result in speech with less background noise being recorded. See section 10.4.1 of [IETF RFC 6787] for more information.

**Type**: Integer

**Optional**: Yes

**Possible values**: 0-100

**Default**: Provisioned

#### 9.3.1.1.2 Ver buffer utterance

**Parameter name**: Ver Buffer Utterance

**Parameter ID**: vbu (0x0035)

**Description**: This parameter indicates to the MG to buffer the utterance associated with this recording. For more information, see section 9.4.25 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Buffer Recording

Off (False): Do not Buffer Recording

**Default**: Off

### 9.4 Statistics

None.

### 9.5 Error codes

None.

### 9.6 Procedures

The procedures are as per clause 10.5 of [ITU-T H.248.9], however the data that is recorded from the media stream is subject to a sensitivity metric as described above.

When recording and verification on the MG share the same buffer resource, the MGC may use the "Ver Buffer Utterance" parameter to indicate to the MG to buffer the utterance associated with the ASR signal into a buffer. The buffered utterance can be used in the verification operations. For further procedures on the use of the buffer see clause 14.6.

## 10 Enhanced ASR package

**Package name**: Enhanced ASR package

**Package ID**: easr (0x00fd)

**Description**: This package provides an expanded set of automatic speech recognition capabilities over the base ASR package. These capabilities include:

"Confidence Threshold" used to control the confidence level required for recognition.

"N Best List" enables the MG to return multiple recognition results if the MG detects alternatives.

"Input Waveform URI" enables the MGC to request that the MG parse an existing media file rather than real time speech input.

**Version**: 1

**Extends**: asr (0x00a6) version 1

## 10.1 Properties

None.

## 10.2 Events

### 10.2.1 Automatic speech recognition failure

**Event name**: ASR Failure

**Event ID**: asrfail (0x0001)

**Description**: As per clause 12.2.1 of [ITU-T H.248.9] with the additional failure code:630 URI access failed

## 10.3 Signals

### 10.3.1 Automatic speech recognition start with grammar script

**Signal name**: ASR Recognition With Grammar Script

**Signal ID**: asrwgs (0x0001)

**Description**: As per clause 12.3.1 of [ITU-T H.248.9]

**Signal Type**: Brief

**Duration**: N/A

#### 10.3.1.1 Additional parameters

##### 10.3.1.1.1 Confidence threshold

**Parameter name**: Confidence Threshold

**Parameter ID**: ct (0x002d)

**Description**: This parameter contains the speech recognition confidence level. The value is an integer between 0 and 100. A value of 0 means minimum confidence is needed for recognition, and a value of 100 requires maximum confidence. See section 9.4.1 of [IETF RFC 6787] for more information.

**Type**: Integer

**Optional**: Yes

**Possible values**: 0-100

**Default**: Provisioned

##### 10.3.1.1.2 N best list length

**Parameter name**: N Best List Length

**Parameter ID**: nbll (0x002e)

**Description**: This parameter specifies the maximum number of the alternative matches. The MGC, by setting this parameter, can ask the MG to send it more than 1 alternative. All alternatives must still be above the Confidence Threshold. See section 9.4.4 of [IETF RFC 6787] for more information.

**Type**: Integer

**Optional**: Yes

**Possible values**: Any positive value

**Default**: 1

#### 10.3.1.1.3 Input waveform URI

**Parameter name**: Input Waveform URI

**Parameter ID**: iwu (0x002f)

**Description**: This parameter specifies a URI pointing to audio content to be processed. This parameter enables the URI form voice recognition. See section 9.4.8 of [IETF RFC 6787] for more information.

**Type**: String

**Optional**: Yes

**Possible values**: One or more segment identifiers adhering to the syntax described in clause 6 of [ITU-T H.248.9].

**Default**: None

#### 10.3.1.1.4 Ver buffer utterance

**Parameter name**: Ver Buffer Utterance

**Parameter ID**: vbu (0x002g)

**Description**: This parameter indicates the MG to buffer the utterance associated with this recognition. For more information, see section 9.4.25 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Buffer Utterance
Off (False): Do not Buffer

**Default:** Off

### 10.4 Statistics

None.

### 10.5 Error codes

None.

### 10.6 Procedures

The procedures are as per clause 12.5 of [ITU-T H.248.9] with the additional procedures.

The MGC may use the parameter "Confidence Threshold" to indicate the criterion for a successful recognition to the MG.

For ASR the MG matches an incoming stream with the supplied grammar, when the parameter "N Best List Length" is set, the MG may come up with more than one alternative match. The parameter "N Best List Length" specifies the maximum number of the recognition results that the MG should report. If this parameter is not indicated, by default, the MG returns only the best match above the confidence threshold.

The parameter "Input Waveform URI" specifies a URI pointing to audio content to be processed by the MG. Based on the setting of the parameter, the MGC and the MG are able to support the URI based recognition. If the MG fails to access the specified URI, the MG shall report the error code 630 to the MGC. When the parameter is provided the MG shall process the media indicated by the URI instead of waiting for external speech input.

When the recognition and verification on the MG share the same buffer resource, the MGC may use the "Ver Buffer Utterance" parameter to indicate the MG to buffer the utterance associated with the ASR signal into a buffer. The buffered utterance can be used in the verification operations. For further procedures on the use of the buffer see clause 14.6.

## 11 Enhanced TTS package

**Package name**: Enhanced TTS package

**Package ID**: etts (0x00fe)

**Description**: This package provides an expanded set of text to speech capabilities over the base TTS package. These capabilities include:

"Play Offset" enables the MGC to indicate that speech output should begin from a certain offset from the start of the speech markup.

"Play Length" enables the MGC to indicate the length of speech playout in terms of grammar (e.g., words, sentences) rather than time.

**Version**: 1

**Extends**: aastts (0x00a8) version 1

### 11.1 Properties

None.

### 11.2 Events

None.

### 11.3 Signals

#### 11.3.1 Play segment identifier

**Signal name**: Play Segment Identifier

**Signal ID**: playsid (0x0001)

**Description**: Plays one or more TTS Segment Identifier(s) with additional functionality to allow playout from a certain offset (jump) within a segment. See section 8.4.1 of [IETF RFC 6787] for more information.

**Signal Type**: Defaults to BR (play continues until the specified or default number of iterations is completed).

**Duration**: Not applicable to BR signals.

##### 11.3.1.1 Additional parameters

###### 11.3.1.1.1 Play offset

**Parameter name**: Play Offset

**Parameter ID**: po (0x0026)

**Description**: The Play Offset parameter can be used to indicate the desired offset into the synthesized speech.

**Type**: Integer

**Optional**: Yes

**Possible values**: 0, positive, or negative offset units

**Default**: 0 (No offset)

#### 11.3.1.1.2 Play unit

**Parameter name**: Play Unit

**Parameter ID**: pu (0x0027)

**Description**: This parameter is used to indicate the unit of measure used in the Play Offset and Play Length parameters in this signal.

**Type**: Enumeration

**Optional**: Yes

**Possible values**: Millisecond (0x0001)
Word (0x0002)
Sentence (0x0003)
Paragraph (0x0004)

**Default**: Millisecond

#### 11.3.1.1.3 Play length

**Parameter name**: Play Length

**Parameter ID**: pl (0x0028)

**Description**: The Play Length parameter indicates the length of speech to speak in terms of units defined by the offset unit parameter.

**Type**: Integer

**Optional**: Yes

**Possible values**: Any positive integer

**Default**: 0 (The entire script is played)

### 11.4 Statistics

None.

### 11.5 Error codes

None.

### 11.6 Procedure

The procedures are per clause 13.6 of [ITU-T H.248.9] with the following additions.

If the MGC needs to instruct the MG to play from an offset, it must provide the Play Offset (*po*) parameter to the MG. The MG shall then begin speaking from this position into the speech markup. The MGC may also use the Play Length (*pl*) parameter to specify the length of speech playout. For both of these parameters, the Play Unit (*pu*) parameter defines the unit of measure.

The use of the "Play Offset" and "Play Length" capabilities applies to the cases where the MGC issues a Play Segment Identifier signal. However, if the MGC issues a Play Script signal, the MGC can formulate the speech markup scripts so that it only sends the specific texts that it wants the MG to process.

To support other playback controls, the general packages in [ITU-T H.248.66], including Signal Pause package, Data Delivery Speed Adjustment package, and Playback Relative Scale Adjustment package, can also be used.

## 12 Play offset control package

**Package name**: Play Offset Control package

**Package ID**: poc (0x00ff)

**Description**: This package allows the MGC to request the MG to alter the play offset of a particular signal. This is a generic package that may be applied to a diverse range of signals related to media play out. See section 8.4.1 of [IETF RFC 6787] for more information.

**Version**: 1

**Extends**: None

### 12.1 Properties

None.

### 12.2 Events

None.

### 12.3 Signals

#### 12.3.1 Jump

**Signal name**: Jump

**Signal ID**: jump (0x0001)

**Description**: This Signal requests that the MG alter the playout of a particular signal to a new location indicated by an offset (i.e., a jump operation).

**Signal Type**: Brief

**Duration**: Not applicable to BR signals.

##### 12.3.1.1 Additional parameters

###### 12.3.1.1.1 Signal identity

**Parameter name**: Signal Identity

**Parameter ID**: sigid (0x0001)

**Description**: The Signal Identity parameter allows an MGC to request a jump operation on the playout of a signal with the indicated signal identity.

**Type**: Octet String

**Optional**: Yes

**Possible values**: For **text encoding** a package Name/Signal ID encoded as per **`pkgdName`** in clause B.2 of [ITU-T H.248.1].

For **binary encoding** a package Name/Signal ID encoded as per **`SignalName`** in clause A.2 of [ ITU-T H.248.1].

**Default**: None

#### 12.3.1.1.2 Signal list identity

**Parameter name**: Signal List Identity

**Parameter ID**: listid (0x0002)

**Description**: The Signal List Identity parameter allows an MGC to request a jump operation on the playout of a Signal List with the indicated identity.

**Type**: Integer

**Optional**: Yes

**Possible values**: 0 to 65535

**Default**: None

#### 12.3.1.1.3 Jump value

**Parameter name**: Jump Value

**Parameter ID**: jv (0x0003)

**Description**: This parameter indicates the offset value that the jump operation takes.

**Type**: Integer

**Optional**: No

**Possible values**: 0, positive or negative offset units

**Default**: None

#### 12.3.1.1.4 Value unit

**Parameter name**: Value Unit

**Parameter ID**: vu (0x0004)

**Description**: This parameter is used to indicate the unit of measure used by the Jump Value parameter in this signal.

**Type**: Enumeration

**Optional**: Yes

**Possible values**:
Millisecond (0x0001)
Word (0x0002)
Sentence (0x0003)
Paragraph (0x0004)
Frame (0x0005)
Byte (0x0006)

**Default**: Provisioned

### 12.4 Statistics

None.

### 12.5 Error codes

#### 12.5.1 Unsupported offset unit or value

**Error code #**: 480

**Name**: Unsupported offset unit or value

**Definition**: The MG has been unable to alter the offset of the signal based on the indicated offset unit and value. The offset is not applied for the specific signal. For example: this may result when playing a TTS signal, an MG may not understand to jump the specified number of milliseconds, or an MG may not be able to jump two paragraphs in a normal audio announcement.

**Error text in the error descriptor**: None.

**Comment**: The signal referenced by the Jump signal continues to play out.

### 12.6 Procedures

In order to alter the playout position of signal, the MGC shall send the "Play Offset Control package Jump" (*poc/jump*) signal to the MG. The MGC shall include existing Signals/Signal Lists as well as the *poc/jump* signal in the Signals Descriptor in order to maintain the playout of the signal. This makes use of the ITU-T H.248 capability where multiple signals are played simultaneously (see clause 7.1.11 of [ITU-T H.248.1]).

The MG shall jump (otherwise known as offset) the playout position according to the Jump Value (*jv*) parameter and the Value Unit (*vu*) parameter from the MGC. The support of the text format units in the *vu* parameter enables the text-based play control of the MGC and the MG. See section 8.4.1 of [IETF RFC 6787] for an example of how Jump adjustment is used.

If the MG is unable to support the *jv* or *vu* parameter values provided by the MGC, or if the indicated values go beyond the range of the particular signal, the MG may return error code 449 "Unsupported or Unknown Parameter or Property Value" in the response. If the *vu* parameter is supported by the MG but not applied to the particular signal, the MG would return error code 480 "Unsupported offset unit or value" to the MGC.

## 13 Voice enrolled grammar package

**Package name**: Voice Enrolled Grammar package

**Package ID**: veg (0x0102)

**Description**: This package allows an MG to support Voice Enrolled Grammars. With this functionality, enrolment is performed using a person's voice. For example, a list of contacts can be created and maintained by recording the person's names using the caller's voice. This technique is sometimes also called speaker-dependent recognition.

**Version**: 1

**Extends**: easr (0x00fd) version 1.

### 13.1 Properties

None.

### 13.2 Events

#### 13.2.1 Automatic speech recognition failure

**Event name**: Automatic Speech Recognition Failure

**Event ID**: asrfail (0x0001)

**Description**: As per clause 10.2.1 with the additional failure code for enrolment:

631 Enrolment error or timeout

### 13.2.2 Enrolment results

**Event name**: Enrolment Results

**Event ID**: enrres (0x0004)

**Description**: This contains the results of the enrolment session. See section 9.7 of [IETF RFC 6787] for more information.

#### 13.2.2.1 EventDescriptor parameters

None.

#### 13.2.2.2 ObservedEventDescriptor parameters

##### 13.2.2.2.1 Number of clashes

**Parameter name**: Number of Clashes

**Parameter ID**: numcla (0x0001)

**Description**: This parameter contains the number of clashes that this pronunciation has with other pronunciations in an active enrolment session. For more information, see section 9.7.1 of [IETF RFC 6787].

**Type**: Integer

**Optional**: No

**Possible values**: 0 upwards

**Default**: None

##### 13.2.2.2.2 Number of good repetitions

**Parameter name**: Number of good repetitions

**Parameter ID**: numgrep (0x0002)

**Description**: This parameter contains the number of consistent pronunciations obtained so far in an active enrolment session. For more information, see section 9.7.2 of [IETF RFC 6787].

**Type**: Integer

**Optional**: No

**Possible values**: 0 upwards

**Default**: None

##### 13.2.2.2.3 Number of repetitions still needed

**Parameter name**: Number of repetitions still needed

**Parameter ID**: numrepsn (0x0003)

**Description**: This parameter contains the number of consistent pronunciations that must still be obtained before the new phrase can be added to the enrolment grammar. For more information, see section 9.7.3 of [IETF RFC 6787].

**Type**: Integer

**Optional**: No

**Possible values**: 0 upwards

**Default**: None

#### 13.2.2.2.4 Consistency status

**Parameter name**: Consistency Status

**Parameter ID**: consta (0x0004)

**Description**: This parameter indicates how consistent the repetitions are when learning a new phrase. For more information, see section 9.7.4 of [IETF RFC 6787].

**Type**: Enumeration

**Optional**: No

**Possible values**: "con" (0x0001): Consistent
"incon" (0x0002): Inconsistent
"undec" (0x0003): Undecided

**Default**: None

#### 13.2.2.2.5 Clashing phrase identities

**Parameter name**: Clashing Phrase Identities

**Parameter ID**: clapi (0x0005)

**Description**: This parameter contains the phrase Ids of clashing pronunciation(s), if any. For more information, see section 9.7.5 of [IETF RFC 6787].

**Type**: Sub-list of String

**Optional**: Yes

**Possible values**: See element "clash-phrase-ids" section 16.2 of [IETF RFC 6787].

**Default**: None

#### 13.2.2.2.6 Transcriptions

**Parameter name**: Transcriptions

**Parameter ID**: trans (0x0006)

**Description**: This parameter contains the transcriptions returned in the last repetition of the phrase being enrolled. For more information, see section 9.7.6 of [IETF RFC 6787].

**Type**: Sub-list of String

**Optional**: Yes

**Possible values**: See element "transcriptions" section 16.2 of [IETF RFC 6787].

**Default**: None

#### 13.2.2.2.7 Confusable phrases

**Parameter name**: Confusable phrases

**Parameter ID**: conph (0x0007)

**Description**: This parameter contains a list of phrases from a command grammar that are confusable with the phrase being added to the personal grammar. For more information, see section 9.7.7 of [IETF RFC 6787].

**Type**: Sub-list of String

**Optional**: Yes

**Possible values**: See element "confusable-phrases" section 16.2 of [IETF RFC 6787].

**Default**: None

## 13.3 Signals

### 13.3.1 Automatic speech recognition start with grammar script

**Signal name**: ASR Recognition With Grammar Script

**Signal ID**: asrwgs (0x0001)

**Description**: As per clause 10.3.1 with the addition of functionality for the MGC to indicate that a collected utterance must be added to the personal grammar being enrolled. The package also allows a set of invalid phrases to be provided to the MG.

**Signal Type**: Brief

**Duration**: N/A

#### 13.3.1.1 Additional parameters

##### 13.3.1.1.1 Enrol utterance

**Parameter ID**: eu (0x0030)

**Description**: This parameter indicates whether the collected utterance is added to the personal grammar being enrolled.

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Enrol Utterance
Off (False): Do not enrol utterance

**Default**: Off

##### 13.3.1.1.2 Confusable phrases URI

**Parameter name**: Confusable phrases URI

**Parameter ID**: conph (0x0031)

**Description**: This parameter specifies a grammar that defines invalid phrases for enrolment. For more information, see section 9.4.44 of [IETF RFC 6787].

**Type**: String

**Optional**: Yes

**Possible values**: A physical segment identifier satisfying the syntax of clause 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

### 13.3.2 Enrolment start

**Signal name**: Enrolment Start

**Signal ID**: enrst (0x0003)

**Description**: This signal triggers a voice enrolled grammar session. For further information see section 9.15 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 13.3.2.1 Additional parameters

##### 13.3.2.1.1 Personal grammar URI

**Parameter name**: Personal Grammar URI

**Parameter ID**: pgURI (0x0001)

**Description**: This specifies a URI to be assigned to the grammar that is used during enrolment to store the personal list of phrases. For further information see section 9.4.37 of [IETF RFC 6787].

**Type**: String

**Optional**: Yes

**Possible values**: A physical segment identifier satisfying the syntax of clause 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

##### 13.3.2.1.2 Minimum number of consistent pronunciations

**Parameter name**: NumMinConsistentPronunciations

**Parameter ID**: nmcp (0x0002)

**Description**: This specifies the minimum number of consistent pronunciations that must be obtained to voice enrol a new phrase. For further information see section 9.4.34 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: 1 upwards

**Default**: Provisioned

##### 13.3.2.1.3 Consistency threshold

**Parameter name**: Consistency Threshold

**Parameter ID**: cont (0x0003)

**Description**: This allows the MGC to specify how similar to a previously enrolled pronunciation of the same phrase an utterance needs to be in order to be considered "consistent". The recognizer may support a variable level of consistency threshold. The level is an integer value between 0 and 100. For more information, see section 9.4.35 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: 0-100

**Default**: Provisioned

#### 13.3.2.1.4 Clash threshold

**Parameter name**: Clash Threshold

**Parameter ID**: clat (0x0004)

**Description**: This allows the MGC to specify how similar the pronunciations of two different phrases can be before they are considered to be clashing. The recognizer may support a variable level of clash threshold. The level is an integer value between 0 and 100. See clause 9.4.36 of [IETF RFC 6787] for more information.

**Type**: Integer

**Optional**: Yes

**Possible values**: 0-100

**Default**: Provisioned

#### 13.3.2.1.5 Phrase identity

**Parameter name**: Phrase Identity

**Parameter ID**: phaid (0x0005)

**Description**: This allows the MGC to indicate the phrase in an existing personal grammar for which enrolment is desired. For more information, see section 9.4.39 of [IETF RFC 6787].

**Type**: String

**Optional**: Yes

**Possible values**: Non-empty string

**Default**: None

#### 13.3.2.1.6 Natural language phrase

**Parameter name**: PhraseNL

**Parameter ID**: phaNL (0x0006)

**Description**: This allows the MG to indicate the interpreted text to be returned when the phrase is recognized. For further information see section 9.4.40 of [IETF RFC 6787].

**Type**: String

**Optional**: Yes

**Possible values**: Non-empty string

**Default**: None

#### 13.3.2.1.7 Weight

**Parameter name**: Weight

**Parameter ID**: wei (0x0007)

**Description**: The value of this parameter represents the likelihood of occurrence of a phrase in an enrolled grammar. The type of this parameter is an

integer value between 0 and 100. See section 9.4.41 of [IETF RFC 6787] for more information.

**Type**: Integer

**Optional**: Yes

**Possible values**: 0-100

**Default**: Provisioned

#### 13.3.2.1.8 Save best waveform

**Parameter name**: Save Best Waveform

**Parameter ID**: sbw (0x0008)

**Description**: This allows the MGC to indicate whether to save the audio stream for the best repetition of the phrase that was used during the enrolment session. For more information, see section 9.4.42 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On/Off

**Default**: None

### 13.3.3 Enrolment rollback

**Signal name**: Enrolment Rollback

**Signal ID**: enrrol (0x0004)

**Description**: This signal indicates that the MG shall discard the last utterance collected via an automatic speech recognition signal (e.g., an asrwgs signal). For more information, see section 9.16 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 13.3.3.1 Additional parameters

None.

### 13.3.4 End phrase enrolment

**Signal name:** End Phrase Enrolment

**Signal ID**: epenr (0x0005)

**Description**: This signal indicates to the MG to commit the new phrase in the grammar associated with the enrolment session or abort the phrase enrolment session. For more information, see section 9.17 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 13.3.4.1 Additional parameters

#### 13.3.4.1.1 Abort phrase enrolment

**Parameter name**: Abort Phrase Enrolment

**Parameter ID**: ape (0x0001)

**Description**: This parameter indicates whether to abort the phrase enrolment session.

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Abort phrase enrolment
Off (False): Do not abort phrase enrolment

**Default**: Off

#### 13.3.4.1.2 Recorded file

**Parameter name**: Recorded File

**Parameter ID**: rf (0x0002)

**Description**: By wildcarding "CHOOSE" this parameter the MGC is able to determine the location of where the audio stream associated with the enrolled grammars is saved.

NOTE – In MRCP this is known as the "Waveform-URI". For further information see section 9.4.8 of [IETF RFC 6787].

**Type**: String

**Optional**: Yes

**Possible values**: The record file specification conforming to the syntax described in clause 6 of [ITU-T H.248.9].

**Default**: None

### 13.3.5 Modify phrase

**Signal name**: Modify Phrase

**Signal ID**: modp (0x0006)

**Description**: This signal allows the MGC to change the phrase ID, NL phrase and/or weight for a given phrase in a personal grammar. See section 9.18 of [IETF RFC 6787] for more information.

**Signal Type**: Brief

**Duration**: N/A

#### 13.3.5.1 Additional parameters

#### 13.3.5.1.1 Personal grammar URI

**Parameter name**: Personal Grammar URI

**Parameter ID**: pgURI (0x0001)

**Description**: Specifies the URI of the personal grammar that contains the phrase.

**Type**: String

**Optional**: No

**Possible values**: A physical segment identifier satisfying the syntax of clause 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

#### 13.3.5.1.2 Phrase Id

**Parameter name**: Phrase Id

**Parameter ID**: phaid (0x0002)

**Description**: Indicates the phrase to be modified.

**Type**: String

**Optional**: No

**Possible values**: Non-empty string

**Default**: None

#### 13.3.5.1.3 New phrase Id

**Parameter name**: New Phrase Id

**Parameter ID**: npid (0x0003)

**Description**: Indicates the new phrase ID to which the MG requires the existing phrase to be changed.

**Type**: String

**Optional**: Yes

**Possible values**: Non-empty string

**Default**: None

#### 13.3.5.1.4 Natural language phrase

**Parameter name**: PhraseNL

**Parameter ID**: phaNL (0x0004)

**Description**: To indicate the interpreted text to be returned when the phrase is recognized.

**Type**: String

**Optional**: Yes

**Possible values**: Non-empty string

**Default**: None

#### 13.3.5.1.5 Weight

**Parameter name**: Weight

**Parameter ID**: wei (0x0005)

**Description**: The value of this parameter represents the occurrence likelihood of a phrase in an enrolled grammar. The type of this parameter is an integer value between 0 and 100.

**Type**: Integer

**Optional**: Yes

**Possible values**: 0-100

**Default**: None

### 13.3.6 Delete phrase

**Signal name**: Delete Phrase

**Signal ID**: delp (0x0007)

**Description**: This signal allows an MGC to delete a phase in a personal grammar added through voice enrolment or text enrolment. See section 9.19 of [IETF RFC 6787] for more information.

**Signal Type**: Brief

**Duration**: N/A

#### 13.3.6.1 Additional parameters

##### 13.3.6.1.1 Personal grammar URI

**Parameter name**: Personal Grammar URI

**Parameter ID**: pgURI (0x0001)

**Description**: Specifies the URI of the personal grammar that contains the phrase.

**Type**: String

**Optional**: No

**Possible values**: A physical segment identifier satisfying the syntax of clause 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

##### 13.3.6.1.2 Phrase Id

**Parameter name**: Phrase Id

**Parameter ID**: phaid (0x0002)

**Description**: Indicates the phrase to be deleted.

**Type**: String

**Optional**: No

**Possible values**: non-empty string

**Default**: None

## 13.4 Statistics

None.

## 13.5 Error codes

### 13.5.1 Invalid action for enrolment session state

**Error code #**: 483

**Name**: Invalid action for Enrolment Session state

**Definition**: The MGC has performed an invalid action due to the voice enrolled grammar session state on the Termination/Stream. The action is disregarded and, as a result, the voice enrolled grammar session remains in the same state.

**Error text in the error descriptor**: None.

**Comment**: None.

## 13.6 Procedures

### 13.6.1 Initiating an enrolment session

To initiate an enrolment session, the MGC shall send the "Enrolment Start" (*veg/enrst*) signal with the required parameters to the MG. Once received the MG shall establish an enrolment session during which any signal defined by the Voice Enrolled Grammar package will be deemed to operate on the session. As such, there shall be only one session per stream active on a Termination at any one time. If the MGC tries to initiate a subsequent enrolment session before ending the ongoing session the MG shall respond with error code 483 "Invalid action for Enrolment Session state".

The MGC shall also set the "Enrolment Results" (*veg/enrres*) event in order to receive the results of the enrolment.

Section 9.15 of [IETF RFC 6787] indicates how the different MRCP equivalent parameters are used.

### 13.6.2 Recognising grammar

While the session is established, the MGC shall issue the "ASR Recognition with Grammar Script" (*veg/asrwgs*) signal with the appropriate parameters to the MG. The ASR procedures are as per clause 12.5 of [ITU-T H.248.9] with the addition of extra parameters defined by this (*veg*) package and the "Enhanced ASR" (easr) package.

The MG shall then enrol a new utterance in a grammar according to the supplied parameters. The MGC may issue the signal multiple times in order for the MG to learn the phrase and add it to a personal grammar.

Upon successful recognition an "ASR Success" ObservedEvent (see clause 12.2.2 of [ITU-T H.248.9] will be generated. As the enrolment session is still active, the "Enrolment Results" (*veg/enrres*) ObservedEvent shall also be generated.

### 13.6.3 Enrolment rollback

If the MGC requires that the last utterance detected by the MG be ignored it shall send the "Enrolment Rollback" (*veg/enrrol*) signal. As such, it only applies to one instance of detection; multiple detections cannot be deleted.

Section 9.16 of [IETF RFC 6787] discusses the effects of rollback.

### 13.6.4 Ending phrase enrolment

The MGC shall only send the "End Phrase Enrolment" (*veg/epenr*) signal during an active enrolment session and shall not send the signal whilst an automatic speech recognition signal (e.g., an asrwgs signal) is active. If it does, the MG shall respond with error code 483 "Invalid action for Enrolment Session state".

The MGC may issue the *veg/epenr* signal in order to commit the new phrase in the grammar when it has determined that the number of repetitions still needed is zero. This can be determined through the "Enrolment Results" (*veg/enrres*) Observed event.

If the MGC had previously requested to "Save Best Waveform" in the *veg/enrst* signal it shall include a wildcarded "CHOOSE" "Recorded File" parameter in the *veg/epenr* signal in order to receive the URI associated with the waveform.

If the MGC requires that the enrolment be aborted (i.e., no phrase is recorded in the personal grammar) it shall issue the *veg/epenr* signal with "Abort Phrase Enrolment" parameter set to "on".

If the MGC removes the Termination/Stream without issuing a *veg/epenr* signal, the MG shall consider this as aborting the enrolment session.

Section 9.17 of [IETF RFC 6787] provides further information regarding ending enrolment.

#### 13.6.5 Modifying a phrase

If the MG wants to change the phrase ID, natural language of the phrase and/or weight it shall send the "Modify Phrase" (*veg/modp*) signal with the appropriate parameters.

If the reference phrase does not exist error code 449 "Unsupported or Unknown Parameter or Property Value" is returned.

Section 9.18 of [IETF RFC 6787] provides further information regarding the modification of phrases in a personal grammar.

#### 13.6.6 Deleting a phrase

If the MG wants to delete a phrase, it shall send the "Delete Phrase" (*veg/delp*) signal with the appropriate parameters.

Section 9.19 of [IETF RFC 6787] provides further information regarding the deletion of phrases in a personal grammar.

## 14 Speaker verification and identification package

**Package name**: Speaker Verification and Identification package

**Package ID**: svi (0x0105)

**Description**: This package allows an MG to support Speaker Verification and Identification. With this functionality, verification or identification operations can be executed against live or buffered audio.

**Version**: 1

**Extends**: None

### 14.1 Properties

None.

### 14.2 Events

#### 14.2.1 Verification failure

**Event name**: Verification Failure

**Event ID**: vefail (0x0001)

**Description**: This event signifies the failure of a verification operation subsequent to the return of the response to the transaction that invoked it.

##### 14.2.1.1 EventDescriptor parameters

None.

##### 14.2.1.2 ObservedEventDescriptor parameters

###### 14.2.1.2.1 Return code

**Parameter name**: Return Code

**Parameter ID**: rc (0x0001)

**Description**: A return code indicating why a Speaker Verification operation failed.

**Type**: Integer

**Optional**: No

**Possible values**: As per clause 10.2.1 with the additional failure codes returned by the verification operation failure event:
632: Buffer Empty
633: Out of Sequence
634: Speech not usable to do the verification

**Default**: None

### 14.2.2 Verification results

**Event name**: Verification Results

**Event ID**: vere (0x0002)

**Description**: This event signifies the results of the verification operation, which may be intermediate ones or final ones based on the time that the event is triggered. See section 11.5 of [IETF RFC 6787] for more information.

#### 14.2.2.1 EventDescriptor Parameters

None.

#### 14.2.2.2 ObservedEventDescriptor parameters

##### 14.2.2.2.1 Voiceprint identifier

**Parameter name**: Voiceprint Identifier

**Parameter ID**: vpid (0x0001)

**Description**: This parameter contains the identity/identities of the voiceprint. For more information, see section 11.5.2.1 of [IETF RFC 6787].

**Type**: Sub-list of String

**Optional**: No

**Possible values**: See element "Voiceprint-Identifier" section 11.4.2 of [IETF RFC 6787]. If this parameter contains more than one item, the items in the following parameters must have the same order as this one.

**Default**: None

##### 14.2.2.2.2 Score type

**Parameter name**: Score Type

**Parameter ID**: st (0x0002)

**Description**: This parameter indicates the type of the verification match scores. It includes whether the other parameters from this signal relate to incremental or cumulative scores. For more information, see sections 11.5.2.2 and 11.5.2.3 of [IETF RFC 6787].

**Type**: Sub-list of Boolean

**Optional**: No

**Possible values**: On – Incremental
Off – Cumulative

NOTE – Incremental may only appear in the first position of the sub-list.

**Default**: None

#### 14.2.2.2.3 Decision

**Parameter name**: Decision

**Parameter ID**: dec (0x0003)

**Description**: This parameter indicates the verification decision. For more information, see section 11.5.2.4 of [IETF RFC 6787].

**Type**: Sub-list of Enumeration

**Optional**: No

**Possible values**: "ac" (0x0001): Accepted
"re" (0x0002): Rejected
"undec" (0x0003): Undecided

**Default**: None

#### 14.2.2.2.4 Utterance length

**Parameter name**: Utterance Length

**Parameter ID**: ul (0x0004)

**Description**: This parameter indicates the size in milliseconds of the last utterance or the cumulated set of utterances. For more information, see section 11.5.2.5 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: 0 upwards. (-1 indicates a NULL value; i.e., the parameter is not present for the voice print instance).

**Default**: None

#### 14.2.2.2.5 Device

**Parameter name**: Device

**Parameter ID**: dev (0x0005)

**Description**: This parameter indicates the apparent type of device used by the caller. For more information, see section 11.5.2.6 of [IETF RFC 6787].

**Type**: Sub-list of Enumeration

**Optional**: No

**Possible values**: "CP" (0x0001): Cellular phone
"EP" (0x0002): Electret phone
"CB" (0x0003): Carbon button phone
"UN" (0x0004): Unknown

NOTE – Electret and carbon button refer to microphone types.

**Default**: None

#### 14.2.2.2.6 Gender

**Parameter name**: Gender

**Parameter ID**: gen (0x0006)

**Description**: This parameter indicates the apparent gender of the speaker. For more information, see section 11.5.2.7 of [IETF RFC 6787].

**Type**: Sub-list of Enumeration

**Optional**: No

**Possible values**: "MA" (0x0001): Male
"FE" (0x0002): Female
"UN" (0x0003): Unknown

**Default**: None

#### 14.2.2.2.7 Adapted

**Parameter name**: Adapted

**Parameter ID**: ada (0x0007)

**Description**: This parameter indicates if the voiceprint has been adapted as a consequence of analysing the source utterances. For more information, see section 11.5.2.8 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True)
Off (False)

**Default**: None

#### 14.2.2.2.8 Verification score

**Parameter name**: Verification Score

**Parameter ID**: vs (0x0008)

**Description**: This parameter indicates the score of the last utterance as determined by verification. The score is an integer value between -100 and 100. For more information, see section 11.5.2.9 of [IETF RFC 6787].

**Type**: Sub-list of Integer

**Optional**: Yes

**Possible values**: -101 to 100 (-101 indicates a NULL value, i.e., the parameter is not present for the voice print instance).

**Default**: Provisioned

#### 14.2.2.2.9 Vendor specific results

**Parameter name**: Vendor Specific Results

**Parameter ID**: vsr (0x0009)

**Description**: This parameter contains implementation specific data, if any. For more information, see section 11.5.2.10 of [IETF RFC 6787].

**Type**: String

**Optional**: Yes

**Possible values**: See element "vendor-specific parameters" section 13.1.6 of [IETF RFC 6787].

**Default**: None

#### 14.2.2.2.10 Waveform URI

**Parameter name**: Waveform URI

**Parameter ID**: waURI (0x000a)

**Description**: Specifies the URI of the recorded audio stream during the verification.

**Type**: String

**Optional**: Yes. If the "Save Waveform" parameter is set to "True", this parameter must be returned.

**Possible values**: A physical segment identifier satisfying the syntax of 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

## 14.3 Signals

### 14.3.1 Start session

**Signal name**: Start Session

**Signal ID**: ss (0x0001)

**Description**: This signal triggers a speaker verification session. For further information see section 11.6 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.1.1 Additional parameters

##### 14.3.1.1.1 Repository URI

**Parameter name**: Repository URI

**Parameter ID**: reURI (0x0001)

**Description**: Specifies the URI of the voiceprint repository to be used or referenced during speaker verification or identification operations.

**Type**: String

**Optional**: No

**Possible values**: A physical segment identifier satisfying the syntax of 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

##### 14.3.1.1.2 Voiceprint identifier

**Parameter name**: Voiceprint Identifier

**Parameter ID**: vpid (0x0002)

**Description**: Specifies the claimed identity for verification applications.

**Type**: String

**Optional**: No

**Possible values**: See element "Voiceprint-Identifier" section 11.4.2 of [IETF RFC 6787].

**Default**: None

#### 14.3.1.1.3 Verification mode

**Parameter name**: Verification Mode

**Parameter ID**: vm (0x0003)

**Description**: This parameter specifies the mode of the verification. See section 11.4.3 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: No

**Possible values**: On – Train
Off – Verify

**Default**: None

#### 14.3.1.1.4 Adapt model

**Parameter name**: Adapt Model

**Parameter ID**: am (0x0004)

**Description**: This parameter indicates the desired behaviour of the verifier resource after a successful verification operation. See section 11.4.4 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Use collected audio to update the voiceprint
Off (False): Not update the voiceprint.

**Default**: Off

#### 14.3.1.1.5 Min verification score

**Parameter name**: Min Verification Score

**Parameter ID**: mvs (0x0005)

**Description**: This parameter determines the minimum verification score for which a verification decision of "accepted" may be declared. The score is an integer value between -100 and 100. For more information, see section 11.4.6 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: -100 to 100 inclusive

**Default**: Provisioned

#### 14.3.1.1.6 Num min verification phrases

**Parameter name:** Num Min Verification Phrases

**Parameter ID**: minvp (0x0006)

**Description**: This parameter specifies the minimum number of valid utterances before a positive decision is given for verification. For more information, see section 11.4.7 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: Any integer greater than 0

**Default**: 1

#### 14.3.1.1.7 Num max verification phrases

**Parameter name**: Num Max Verification Phrases

**Parameter ID**: maxvp (0x0007)

**Description**: This parameter specifies the number of valid utterances required before a decision is forced for verification. For more information, see section 11.4.8 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: Any integer greater than 0

**Default**: Provisioned

### 14.3.2 Verify

**Signal name**: Verify

**Signal ID**: verify (0x0002)

**Description**: This signal triggers the MG to do a Speaker Verify operation. This is either to train/adapt the voiceprint or to verify/identify a claimed identity. For more information see section 11.10 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.2.1 Additional parameters

#### 14.3.2.1.1 Waiting input time

**Parameter name**: Waiting Time for Input

**Parameter ID**: wit (0x0001)

**Description**: Defines the time to wait to detect user input, specified in units of 10 milliseconds. For more information, see section 11.4.9 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: 0 and up

**Default**: Provisioned

#### 14.3.2.1.2 Save waveform

**Parameter name**: Save Waveform

**Parameter ID**: sw (0x0002)

**Description**: This parameter indicates to the MG whether to save the audio stream that was used for verification/identification. For more information, see section 11.4.10 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Save the audio stream
Off (False): Do not save the audio stream

**Default**: Off

#### 14.3.2.1.3 Ver buffer utterance

**Parameter name**: Ver Buffer Utterance

**Parameter ID**: vbu (0x0003)

**Description**: This parameter indicates the MG to buffer the utterance associated with this verification. For more information, see section 11.4.14 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True)/Off (False)

**Default**: Off

#### 14.3.2.1.4 Input waveform URI

**Parameter name**: Input Waveform URI

**Parameter ID**: iwu (0x0004)

**Description**: This parameter specifies stored audio content that the MGC request the MG to fetch and process in the verification session. For more information, see section 11.4.15 of [IETF RFC 6787].

**Type**: String

**Optional**: Yes

**Possible values**: One or more segment identifiers adhering to the syntax described in clause 6 of [ITU-T H.248.9].

**Default**: None

#### 14.3.2.1.5 Post speech timer

**Parameter name**: Post Speech Timer

**Parameter ID**: pst (0x0005)

**Description**: The amount of silence necessary after the end of the speech, specified in units of 10 milliseconds. For more information, see section 11.4.18 of [IETF RFC 6787].

**Type**: Integer

**Optional**: Yes

**Possible values**: 0 upwards

**Default**: Provisioned

### 14.3.3 Verify from buffer

**Signal name**: Verify from Buffer

**Signal ID**: verifybu (0x0003)

**Description**: This signal triggers the MG to perform speaker verification on audio from the buffer. This is either to train/adapt the voiceprint or to verify/identify a claimed identity. See section 11.11 of [IETF RFC 6787] for more information.

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.3.1 Additional parameters

None.

### 14.3.4 Query voiceprint

**Signal name**: Query Voiceprint

**Signal ID**: quvp (0x0004)

**Description**: This signal allows the MG to check if a voiceprint or repository exists and if it contains trained voiceprints. See section 11.8 of [IETF RFC 6787] for more information.

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.4.1 Additional parameters

##### 14.3.4.1.1 Repository URI

**Parameter name**: Repository URI

**Parameter ID**: reURI (0x0001)

**Description**: Specifies the URI of the voiceprint repository to be referenced.

**Type**: String

**Optional**: No

**Possible values**: A physical segment identifier satisfying the syntax of 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

##### 14.3.4.1.2 Voiceprint identifier

**Parameter name**: Voiceprint Identifier

**Parameter ID**: vpid (0x0002)

**Description**: Specifies the claimed identity that the MGC wants to query.

**Type**: String

**Optional**: No

**Possible values**: See element "Voiceprint-Identifier" section 11.4.2 of [IETF RFC 6787].

**Default**: None

#### 14.3.4.1.3 Voiceprint exist

**Parameter name**: Voiceprint Exist

**Parameter ID**: vpe (0x0003)

**Description**: This parameter indicates the status of the voiceprint. For more information, see section 11.4.13 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: No

**Possible values**: On (True): Exist
Off (False): Does not exist

**Default**: None

### 14.3.5 Delete voiceprint

**Signal name**: Delete Voiceprint

**Signal ID**: devp (0x0005)

**Description**: This signal allows an MGC to delete a voiceprint from a repository. See section 11.9 of [IETF RFC 6787] for more information.

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.5.1 Additional parameters

#### 14.3.5.1.1 Repository URI

**Parameter name**: Repository URI

**Parameter ID**: reURI (0x0001)

**Description**: Specifies the URI of the repository that contains the voiceprint.

**Type**: String

**Optional**: No

**Possible values**: A physical voiceprint repository identifier satisfying the syntax of clause 6.2.5.2 of [ITU-T H.248.9].

**Default**: None

#### 14.3.5.1.2 Voiceprint identifier

**Parameter name**: Voiceprint Identifier

**Parameter ID**: vpid (0x0002)

**Description**: This parameter indicates the voiceprint to be deleted.

**Type**: String

**Optional**: No

**Possible values**: See element "Voiceprint-Identifier" section 11.4.2 of [IETF RFC 6787].

**Default**: None

#### 14.3.5.1.3 Voiceprint exist

**Parameter name**: Voiceprint Exist

**Parameter ID**: vpe (0x0003)

**Description**: This parameter indicates the status of the voiceprint at the moment the delete execution started. For more information, see section 11.4.13 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: No

**Possible values**: On (True): Exist
Off (False): Does not exist

**Default**: None

### 14.3.6 Verify rollback

**Signal name**: Verify Rollback

**Signal ID**: verol (0x0006)

**Description**: This signal indicates that the MG shall discard the last utterance collected via "Verify" signal. For more information, see section 11.12 of [IETF RFC 6787].

NOTE – This has no effect on a "Verify from Buffer" signal.

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.6.1 Additional parameters

None.

### 14.3.7 Stop

**Signal name**: Stop

**Signal ID**: stop (0x0007)

**Description**: This signal stops the Verify or Verify from Buffer signal if one is active. For further information see section 11.13 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.7.1 Additional parameters

#### 14.3.7.1.1 Abort verification

**Parameter name**: Abort Verification

**Parameter ID**: av (0x0001)

**Description**: This parameter indicates whether to discard the ongoing verification results. If this parameter is set to "Off", the MG triggers the notification of the "Verification Results" event. For further information see section 11.4.5 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Discard the results
Off **(**False): Return the current verification results

**Default**: On

### 14.3.8 Clear buffer

**Signal name**: Clear Buffer

**Signal ID**: cb (0x0008)

**Description**: This signal indicates the MG to clear the verification buffer. For more information, see section 11.17 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.8.1 Additional parameters

None.

### 14.3.9 Get intermediate result

**Signal name**: Get Intermediate Result

**Signal ID**: gir (0x0009)

**Description**: This signal allows an MGC to request the intermediate results of a verification request that is in progress. Based on the receiving of this signal, the MG triggers the notification of the "Verification Results" event. See section 11.18 of [IETF RFC 6787] for more information.

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.9.1 Additional parameters

None.

### 14.3.10 End session

**Signal name**: End Session

**Signal ID**: es (0x000a)

**Description**: This signal ends a speaker verification session. For further information see section 11.7 of [IETF RFC 6787].

**Signal Type**: Brief

**Duration**: N/A

#### 14.3.10.1 Additional parameters

##### 14.3.10.1.1 Abort model

**Parameter name**: Abort Model

**Parameter ID**: am (0x0001)

**Description**: This parameter indicates the desired behaviour of the verifier resource upon session termination. For further information see section 11.4.21 of [IETF RFC 6787].

**Type**: Boolean

**Optional**: Yes

**Possible values**: On (True): Discard any pending changes to the voiceprint
Off (False): Commit any pending changes to the voiceprint.

**Default**: Off

### 14.4 Statistics

None.

### 14.5 Error codes

#### 14.5.1 Invalid action for verification session state

**Error code #**: 484

**Name:** Invalid action for Verification Session state

**Definition**: The MGC has performed an invalid action due to the speaker verification session state on the Termination/Stream. The action is disregarded and, as a result, the verification session remains in the same state.

**Error text in the error descriptor**: None

**Comment**: None.

### 14.6 Procedures

#### 14.6.1 Verification session control

To initiate a speaker verification session, the MGC shall send the "Start Session" (*svi/ss*) signal with the required parameters to the MG. Once received the MG shall establish a speaker verification session and get the voiceprint specified by the "Repository URI" and "Voiceprint Identifier" parameters. During the session, any signal defined by the speaker verification and identification package will be deemed to operate on it. As such, there shall be only one session per stream active on a Termination at any one time.

The MGC should set the "Verification Results" (*svi/vere)* event in order to receive the results of subsequent verification requests.

If the MGC tries to initiate a subsequent verification session before ending the ongoing session the MG shall respond with error code 484 "Invalid action for Verification Session state".

To end a speaker verification session, the MGC shall send the "End Session" (*svi/es*) signal to the MG. Based on this signal, the MG shall terminate the verification session and respond to the MGC.

Sections 11.6 and 11.7 of [IETF RFC 6787] indicate how the different MRCP equivalent parameters are used.

#### 14.6.2 Speaker verification

When receiving the "Verify" (*svi/verify*) signal sent by the MGC, the MG performs the speaker verification operation based on the instructions carried in the signal. This is either to train/adapt the voiceprint or to verify/identify a claimed identity, which depends on the value of the "Verification Mode" parameter of the *svi/ss* signal.

The MG shall then determine the results of the verification, and notify the results to the MGC via the "Verification Results" (*svi/vere*) event. The event may contain the results associated with several voice prints. Different results may be returned for the first and subsequent voice prints. The parameters of the *svi/vere* event are structured as "sub-list" in order to return multiple results. The

"Voice Print Identifier" (*vpid*) parameter is the master parameter. The other "sub-list" parameters in the event must have the same number of elements as the *vpid* signal. If the particular parameter is not applicable to the voice print instance, the element of the sub-list associated with that instance is assigned a default or NULL value.

If the MGC wants to perform speaker verification on audio from the buffer, it shall issue the "Verify from Buffer" (*svi/verifybu*) signal to the MG.

The MGC can instruct the "Get Intermediate Result" (*svi/gir*) signal to poll for intermediate results of a verification request that is in progress. When receiving this signal, the MG triggers the notification of the *svi/vere* event to the MGC.

When the MGC wants to stop the active verification operation, it shall send the "Stop" (*svi/stop*) signal to the MG. The "Abort Verification" parameter carried in the signal indicates whether or not to discard the ongoing verification results. If the parameter is set to "Off", the MG triggers the notification of the *svi/vere* event.

Speaker identification can be regard as a special type of speaker verification. If the MGC wants to do speaker identification, it may issue the *svi/verify* signal multiple times in order for the MG to verify the alternative voiceprints.

Sections 11.10, 11.11, 11.13, 11.15 and 11.18 of [IETF RFC 6787] indicate how the different MRCP methods operate.

### 14.6.3 Voiceprint operation

The MGC shall issue the "Query Voiceprint" (*svi/quvp*) signal to the MG to check the status of the specified voiceprint. The MGC shall send the signal fully specifying "Repository URI" and "Voiceprint Identifier" whilst setting the "Voiceprint Exist" to CHOOSE. The MG will then respond the query result with true/false value for the "Voiceprint Exist" parameter.

When the MGC wants to delete a specified voiceprint, it shall issue the "Delete Voiceprint" (*svi/devp*) signal to the MG. Using a "CHOOSE" wildcard with the parameter "Voiceprint Exist" specified to the MGC can get the status of the voiceprint at the moment the delete execution starts. The MG will then delete the voiceprint specified by the parameters carried in the signal, and respond the result to the MGC.

Sections 11.8 and 11.9 of [IETF RFC 6787] specifies the effects of voiceprint query and deletion.

### 14.6.4 Verify rollback

If the MGC requires that the last live utterance detected by the MG be ignored it shall send the "Verify Rollback" (*svi/*verol) signal. As such, it only applies to one instance of detection; multiple detections cannot be deleted. The MG shall delete the last detected live utterance, and respond the result to the MGC.

Section 11.12 of [IETF RFC 6787] discusses the effects of rollback.

### 14.6.5 Clear buffer

The MGC shall send the "Clear Buffer" (*svi/cb*) signal to clear the verification buffer. This buffer is used to buffer speech during a recognition, recording or verification operations that may later be used in the "Verify from Buffer" signal. The MG shall clear the buffer and respond to the MGC.

Section 11.17 of [IETF RFC 6787] discusses the background of the buffer.

# Bibliography

[b-IETF RFC 4313] IETF RFC 4313 (2005), *Requirements for Distributed Control of Automatic Speech Recognition (ASR), Speaker Identification/Speaker Verification (SI/SV), and Text-to-Speech (TTS) Resources*.

[b-W3C EMMA] W3C Working Draft (2007), *EMMA: Extensible Multimodal Annotation markup language*.

[b-W3C SRGS] W3C Recommendation (2004), *Speech Recognition Grammar Specification*.

## SERIES OF ITU-T RECOMMENDATIONS

| | |
|---|---|
| Series A | Organization of the work of ITU-T |
| Series D | General tariff principles |
| Series E | Overall network operation, telephone service, service operation and human factors |
| Series F | Non-telephone telecommunication services |
| Series G | Transmission systems and media, digital systems and networks |
| **Series H** | **Audiovisual and multimedia systems** |
| Series I | Integrated services digital network |
| Series J | Cable networks and transmission of television, sound programme and other multimedia signals |
| Series K | Protection against interference |
| Series L | Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant |
| Series M | Telecommunication management, including TMN and network maintenance |
| Series N | Maintenance: international sound programme and television transmission circuits |
| Series O | Specifications of measuring equipment |
| Series P | Terminals and subjective and objective assessment methods |
| Series Q | Switching and signalling |
| Series R | Telegraph transmission |
| Series S | Telegraph services terminal equipment |
| Series T | Terminals for telematic services |
| Series U | Telegraph switching |
| Series V | Data communication over the telephone network |
| Series X | Data networks, open system communications and security |
| Series Y | Global information infrastructure, Internet protocol aspects and next-generation networks, Internet of Things and smart cities |
| Series Z | Languages and general software aspects for telecommunication systems |

Printed in Switzerland
Geneva, 2016