

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

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/>.

© ITU 2016

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

	Page
1 Scope.....	1
2 References.....	1
3 Definitions	1
3.1 Terms defined elsewhere	1
3.2 Terms defined in this Recommendation.....	1
4 Abbreviations and acronyms	2
5 Conventions	2
6 MRCP and ITU-T H.248 interworking	2
6.1 Speech synthesizer control: TTS	2
6.2 Automatic speech recognition: ASR	5
6.3 Recording	12
6.4 Speaker verification and identification.....	14
7 Media start package	18
7.1 Properties.....	18
7.2 Events	18
7.3 Signals	19
7.4 Statistics.....	19
7.5 Error codes.....	19
7.6 Procedures	19
8 Trim package	19
8.1 Properties.....	19
8.2 Events	19
8.3 Signals	20
8.4 Statistics.....	20
8.5 Error codes.....	20
8.6 Procedures	21
9 Enhanced recording package	21
9.1 Properties.....	21
9.2 Events	21
9.3 Signals	21
9.4 Statistics.....	22
9.5 Error codes.....	22
9.6 Procedures	22
10 Enhanced ASR package.....	23
10.1 Properties.....	23
10.2 Events	23
10.3 Signals	23

		Page
	10.4 Statistics.....	25
	10.5 Error codes.....	25
	10.6 Procedures	25
11	Enhanced TTS package	25
	11.1 Properties.....	25
	11.2 Events	26
	11.3 Signals	26
	11.4 Statistics.....	27
	11.5 Error codes.....	27
	11.6 Procedure	27
12	Play offset control package.....	27
	12.1 Properties.....	27
	12.2 Events	27
	12.3 Signals	28
	12.4 Statistics.....	29
	12.5 Error codes.....	29
	12.6 Procedures	29
13	Voice enrolled grammar package	30
	13.1 Properties.....	30
	13.2 Events	30
	13.3 Signals	32
	13.4 Statistics.....	39
	13.5 Error codes.....	39
	13.6 Procedures	40
14	Speaker verification and identification package.....	41
	14.1 Properties.....	41
	14.2 Events	41
	14.3 Signals	45
	14.4 Statistics.....	53
	14.5 Error codes.....	53
	14.6 Procedures	54
	Bibliography.....	56

iv

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 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 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 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 Pauses speech output.	This is supported by the Signal Pause package in clause 13 of [ITU-T H.248.66].
	Method: RESUME 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 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. 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 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 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 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 Used in a "SPEAK" request: as a desired offset into the synthesized speech. 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]. The "Control" request is supported by [ITU-T H.248.66]. [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 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 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 It contains the reason text behind the "SPEAK" request completion.	See "Completion Cause" row above.
	Header: Fetch Hint 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 It controls the URI access properties of audio files.	This is not applicable to ITU-T H.248.
	Header: Speak Restart It controls the restart behaviours.	This is supported by the "Playback Relative Scale Adjustment" package in [ITU-T H.248.66].
	Header: Speak Length 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. The "Enhanced TTS" package in clause 11 allows the specification of length in terms of grammar (e.g., word/sentence/paragraph).
	Header: Speaker Profile 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 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 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 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 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 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 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 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 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 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]. 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 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 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 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 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 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 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).
Recognizer events	Recognizer-event: START-OF-INPUT 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 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 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 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 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 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 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 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 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 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 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]). 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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]). 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 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 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 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 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 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 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 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 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.