INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2022-2024

SG16-TD129/WP1

STUDY GROUP 16

Original: English

Question(s): 21/16 Geneva, 10-21 July 2023

TD

Source: Editor F.ACSMRF

Title: F.ACSMRF "Requirements and framework for advanced call services with media enhancements" (New): Initial draft (Geneva, 10-21 July 2023)

Contact: Zhi Qiao, China Unicom, China — Tel: +86 18601909300, E-mail: qiaozhi3@chinaunicom.cn

Contact: Jianxin Liao, Beijing University of Posts and Telecommunications, China — Tel: +86 13901234784, E-mail: liaojx@bupt.edu.cn

Contact: Chuanyang Miao, ZTE Corporation, China — Tel: +86-25-88014611, E-mail: miao.chuanyang@zte.com.cn

Contact: Qiwei Shen, Beijing University of Posts and Telecommunications, China — Tel: +86 13911283993, E-mail: shenqw@bupt.edu.cn

Contact: Shenxiu Du, China Mobile, China — Tel: +86 15010253388, E-mail: dushenxiu@chinamobile.com

Abstract: This TD contains the draft Recommendation F.ACSMRF "Requirements and framework for advanced call services with media enhancements" . This updated draft is based on the discussion of the outcome of contribution (SG16-C188-R1) in SG16 meeting (Geneva, 10-21 July 2023).

NOTE – Editorial adjustments performed by TSB.

This document contains the initial text for this new work item, based on the discussions of the following contributions.

Document	Source	Title	Agreements
SG16-C188-R1	Beijing UPT (China), China Mobile., China Unicom, ZTE	New: F.ACSMRF: Proposal for a new work item to study the requirements and framework for advanced call services with extra media	A change in the title is introduced, from "Requirements and framework for advanced call services with extra media" to "Requirements and framework for advanced call services with media enhancements".

CONTENTS

Page

1 Scope 4

2 References 4

3 Definitions 4
- 3.1 Terms defined elsewhere 4
- 3.2 Terms Defined within this document 4

4 Abbreviations 5

5 Convention 5

6 Introduction 5

7 System framework 6

8 Scenarios types 8
- 8.1 Overlay media 8
- 8.2 Media injection 9
- 8.3 Media immerse 9

9 Requirements 10
- 9.1 General requirements 10
- 9.2 Advanced call platform requirements 10
- 9.3 Terminal device requirements 10

10 Security considerations 10

Appendix I Scenarios of advanced call services with media enhancements 12
- I.1 Overlay media 12
- I.2 Media injection 13
- I.3 Media immerse 14

Bibliography 15

List of Tables

Page

No table of figures entries found.

List of Figures

Page

Figure 6-1 – Framework of advanced call services with media enhancements 8

Page

Figure 8-1 – The schematic of overlay media type 8
Figure 8-2 – The schematic of media injection type 9
Figure 8-3- The schematic of media immerse type 10
Figure I.1 – Typical scenarios of overlay media type 13
Figure I.2 – Typical scenarios of media injection type 14
Figure I.3 – Typical scenarios of Media immersion type 14

Draft Recommendation ITU-T F.ACSMRF

Requirements and ~~F~~framework for ~~A~~advanced ~~C~~call ~~S~~services with ~~Extra Media~~media enhancements

Summary

Advanced call services with media enhancements are an extension of the existing audio and video call services, allowing transmission of enhanced media content during a call. The enhanced media content is transmitted through a newly established logical pipe. Enhanced media content includes audio, video, text and other content.Advanced call services with media enhancements enable a visual, multimedia-rich, highly perceptible, and fully interactive experience to users,providing users with more interactive experiences.

This document provides the requirements and framework for advanced call services with media enhancements..~~[To be provided before Consent]~~

Commented [TSB1]: Editor: In English, space is necessary after punctuation e.g. . , ; :) (

Keywords

~~Advanced Call Service; 5G; Extra Media; XR; AI~~ ~~Advanced Call Service; Extra Media;~~ advanced call services; media enhancements;

1 Scope

This draft Recommendation identifies the requirements and framework for advanced call services with ~~extra media~~media enhancements. ~~-~~In particular, this Recommendation covers the following aspects for advanced call services with media enhancements:

– System framework of advanced call services with media enhancements

– Scenarios types of advanced call services with media enhancements

– Requirements of advanced call services with media enhancements

Scenarios of advanced call services with media enhancements are provided in Appendix I.

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; all 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.

[TBD]

3 Definitions

3.1 Terms defined elsewhere

This document uses the following terms defined elsewhere

[TBD]

3.2 Terms Defined within this document

This document defines the following terms

[TBD]

4 Abbreviations

IMS	Internet Protocol Multimedia Subsystem
AST	Automatic Speech Recognition
TTS	Text To Speech
AI	Artificial Intelligence
XR	eXtended Reality
AR	Augmented Reality

5 Convention

[TBD]

6 Introduction

The existing typical call establishes a connection between the calling party and the called party. In the case of audio calls, the connection includes an audio channel for transmitting audio contents. For video calls, an additional video channel is established within the connection to transmit the video contents captured by the camera. During a video call, it is possible to be requested to release connection resources to downgrade to a voice call, and a voice call can also be requested to reapply connection resources to re-upgrade to a video call.

Advanced call services with ~~extra media~~media enhancements ~~is~~are an extension of the existing call services, allowing the transmission of ~~additional~~~~extra~~ enhanced media ~~information~~ content during a call. The enhanced ~~additional~~~~extra~~ media content is transmitted through a newly established logical pipe, which is logically independent of the audio and video channels established for the call. Enhanced ~~Additional~~~~Extra~~ media content includes audio, video, text and other content. The contents from logical pipe and channel are rendered or composed in either terminal devices or platform. The new contents after media processing are displayed to users, making the call service more immersive and enjoyable.

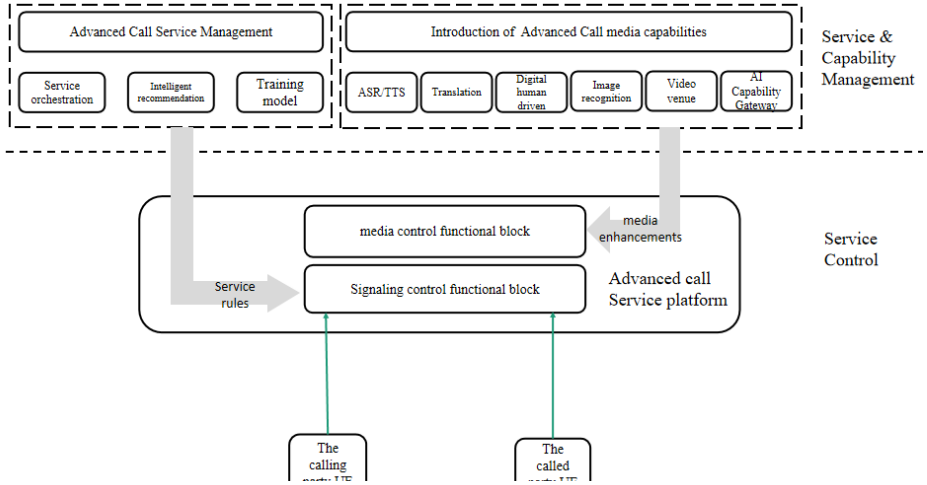
Advanced call services with ~~extra media~~media enhancements ~~is~~are an upgrade to the existing call services, which leveraging 5G network technology and incorporating AI and XR technologies. It is introduced virtual avatars, AR specific effects, AI voice recognition, screen sharing, and other features into the call. It is provided users with enhanced communication and video interaction, enabling real-time multimedia sharing, instant interactions. It is improved personalized socializing, efficient business communication, and convenient interactions, making calls more interesting, and providing users with a more abundant communication experience.

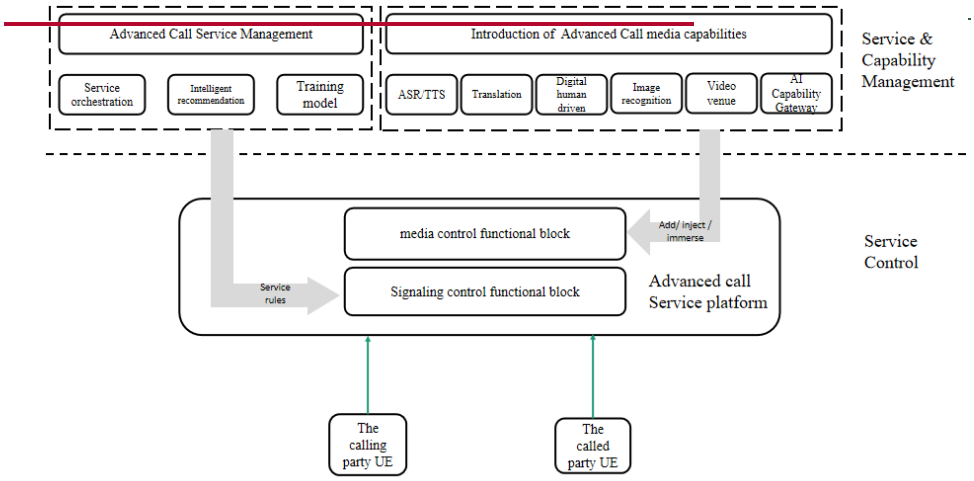
Advanced call services transmit enhanced ~~extra~~ media content through new logical pipes including audio pipe, video pipe and data pipe. And there is a new platform which is called advanced call platform processes signaling and media to implement service function. The directions of contents transmission in logical pipe can be from the calling party to the called party, from the called party to the calling party, from the calling party to the platform, from the called party to the platform, or from the platform to the calling party or the called party. Each logical pipe can be transmitted different content, which may or may not be related to the content transmitted in the audio or video channel originally established by two calling parties in call process. According to the contents of services, Advanced call services with ~~extra media~~media enhancements can be categorized into three types with specific service scenarios as shown below, including overlay media, injection media and Immerse Media.In a typical media processing service flow from end to end, video contents are captured by the content provider firstly. After the video is uploaded to the cloud server, media processing services begins to work. In order to adapt to different network environments and

terminals, video will be usually transcoded in the cloud. Some personalized value-added services are provided, which are including data statistics functions, content review functions and real-time watermarking functions and so on. After media processing, the video contents are ingested to CDN, contents are distributed and speeded up to the terminal devices. Video contents are decoded and played on the terminal devices.

7 System framework

Advanced call services with ~~extra media~~media enhancements ~~is~~are based on the services and multimedia content services and ecology. Advanced call service platform, mobile core network, and terminal devices work together to provide the services. Service management formulates service rules for the advanced call Service platform through various methods such as service orchestration，intelligent recommendation and model training. The introduction of media capabilities provides various capabilities such as AST(Automatic Speech Recognition)/TTS(Text To Speech)，translation and AI capability engine for the platform's media control. A ~~An~~ ~~f~~framework of advanced call services with ~~extra media~~media enhancements ~~system~~ is shown in Figure 6-1.


Advanced Call Service Management
Service orchestration
Intelligent recommendation
Training model
Introduction of Advanced Call media capabilities
ASR/TTS
Translation
Digital human driven
Image recognition
Video venue
AI Capability Gateway
Service & Capability Management
media control functional block
media enhancements
Service rules
Signaling control functional block
Advanced call Service platform
Service Control
The calling
The called


Advanced Call Service Management
Service orchestration
Intelligent recommendation
Training model
Introduction of Advanced Call media capabilities
ASR/TTS
Translation
Digital human driven
Image recognition
Video venue
AI Capability Gateway
Service & Capability Management
media control functional block
Add/ inject / immerse
Service rules
Signaling control functional block
Advanced call Service platform
Service Control
The calling party UE
The called party UE

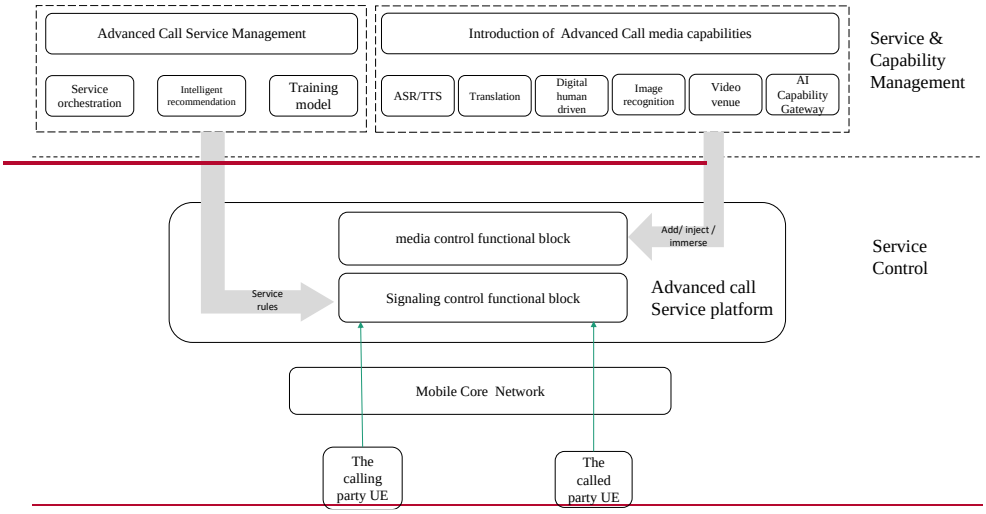


Figure 6-1 – Framework of advanced call services with ~~extra media~~media enhancements

8 Scenarios types

[Editor notes: the title of this chapter need to be considered to change in the future.]

~~7.1~~8.1 Overlay media

The typical scenarios of overlay media type include blur background, virtual background, early media continuation, XR effect, screen sharing, and white boarding. The schematic of overlay media type is depicted in Figure 8-1.

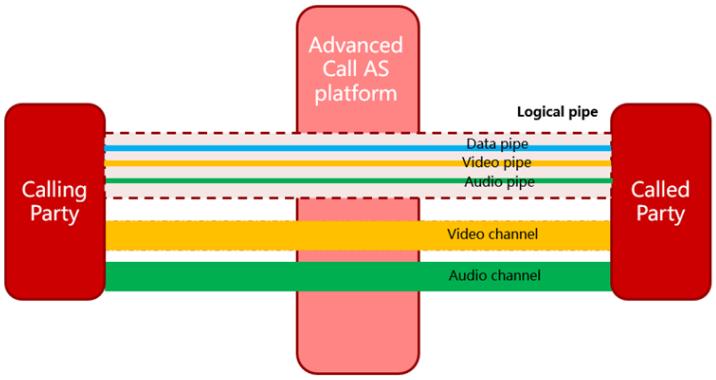


Figure 8-1 – The schematic of overlay media type

– An additional logical pipe is established. The audio/video/data contents transmitted in the logical pipe is generated from the base audio/video content transmitted in the video and audio channel. The contents transmitted in audio or video channel are injected to the logical pipe.

– The original audio/video contents transmitted in the video and audio channel cannot be changed by contents transmitted in logical pipe.

– Typical service scenario: speech-to-Text. The audio contents transmitted in audio channel are injected to the logical pipe and translated into texts.

~~7.2~~8.2 Media injection

The typical scenarios of injection media include speech-to-text and real-time translation. The schematic of media injection is depicted in Figure 8-2.

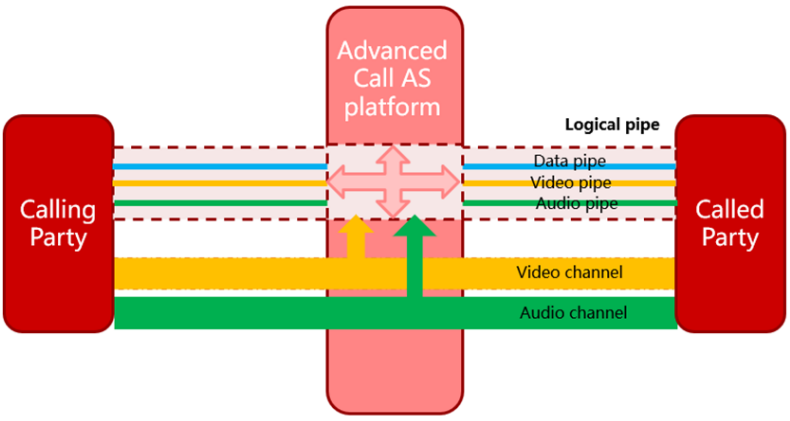


Figure 8-2 – The schematic of media injection type

– An additional logical pipe is established. The audio/video/data contents are transmitted in the logical pipe is generated from the base audio/video content transmitted in the video and audio channel. The contents transmitted in audio or video channel are injected to the logical pipe.

– The base audio/video content transmitted in the video and audio channel cannot be changed by logical pipe.

– Typical service scenario: Speech-to-Text. The audio contents transmitted in audio channel are injected to the logical pipe and translated into texts.

~~7.3~~8.3 Media immerse

The typical scenarios of media immerse include digital avatar and immersive meeting， utilizes the new media pipe to transmit new content and interactive video (data). The immerse media pipe is depicted in Figure 8-3.

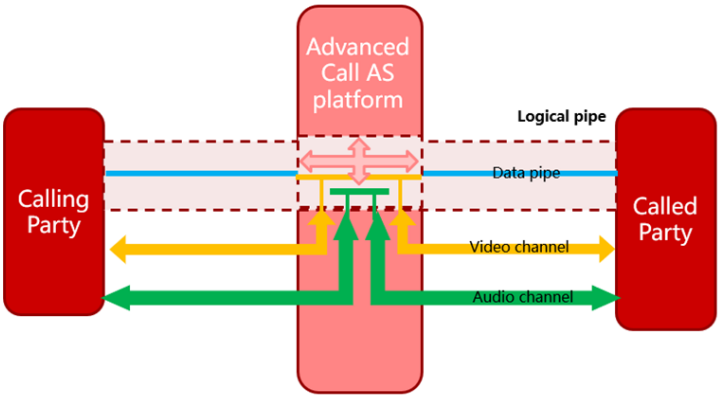


Figure 8-3- The schematic of media immerse type

– An additional logical pipe is established. The audio/video/data contents transmitted in the logical pipe are mixture of the user's original audio/video contents.

– The contents transmitted in the original audio or video channel are replaced and substituted by the new contents transmitted in logical pipes which is between the terminal devices and the advanced call services platform. The new contents are transmitted in logical pipe from advanced call AS platform to terminal devices.

– Typical service: immersive meeting. Advanced call service platform combines the user's audio/video contents with the virtual meeting background, generating new virtual meeting contents. The mixed contents are sent to users through the additional logical pipe. The original video and audio channel are replaced by logical pipe.

~~8~~9 Requirements

This clause addresses the requirements for advanced call services with media enhancements.~~media processing service~~.

[TBD]

~~8.1~~9.1 General requirements

[TBD]

~~8.2 Network Requirements~~

~~[TBD]~~

~~8.3~~9.2 Advanced call platform requirements

[TBD]

~~8.4~~9.3 Terminal device requirements

[TBD]

~~9~~10 Security considerations

[TBD]

Appendix I

Scenarios of advanced call services with ~~extra media~~media enhancements

This ~~clause~~ appendix describes ~~the~~ typical application scenarios of advanced call services with ~~extra media~~media enhancements.

I.1 Overlay media

Voice between A and B is delivered transparently on audio pipe. Typical scenarios are illustrated in Figure I.1 and described as follows.

Blur background

– Blur background refers to the process of blurring the user's background environment during a call.

Virtual background

– Virtual background refers to replacing the user's background environment with an image during a call. It supports users selecting virtual backgrounds from a pre-existing library and also allows users to upload their own images to define their virtual backgrounds.

Early media continuation

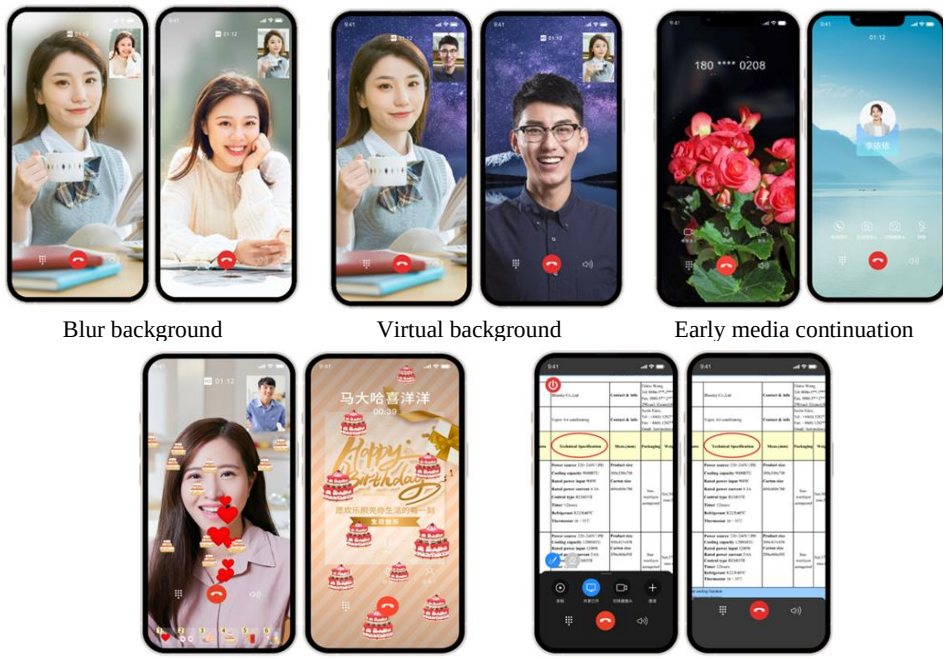
– Early media continuation refers to extending the content from the early media phase (e.g., video ring back tones) into the call and playing it as a video background silently before and after dialing and answering.

XR effects

– XR effects refer to the capability for users to overlay facial expression stickers on the call interface and choose voice expression effects.

Screen sharing and white board

– Screen sharing and white board refer to the capability of displaying the current content on the user's phone screen to the other party's call interface during the call. Users can also use features such as drawing tools and erasers to annotate and doodle, and the annotations will be displayed in real-time on the other party's call interface.



Blur background Virtual background Early media continuation

XR effects Screen sharing and white board

Figure I.1 – Typical scenarios of overlay media type

I.2 Media injection

New media pipe is added to deliver new contents which is related to the voice on the audio pipe. The new contents could be subtitles. Typical scenarios are following.

Speech-to-text transcription

– Speech-to-Text transcription refers to the content of the voice call is intelligently recognized during the call. And it is converted into subtitles and superimposed on the user's call interface. Transcription is supported dialects.

Real-time translation

– Real-time Translation in communication enables one- or multi-party free conversations across different languages, see Figure I.2. In point-to-point audio and video calls in different languages, Real-time translating converts the speech into text. The text translated from the local speech can be displayed on the remote terminal side during the conversation. Translating capability can invoke multiple third-party translation engines.

Figure I.2 – Typical scenarios of media injection type

I.3 Media immerse

New media pipes are added in which the new contents and interactive video (data) among the calling party, the called party and the advanced call platform are delivered. Typical scenarios are illustrated in Figure I.3 and described as follows.

Digital Avatar

– Digital avatar refers to replacing the user's portrait with a pre-set digital avatar image to present during the call.

Immersive Meeting

– Immersive meeting refers to blending pre-set digital avatars of users with virtual meeting scene to create immersive virtual video conferences.

Digital Avatar Immersive Meeting

Figure I.3 – Typical scenarios of Media immersion type

Bibliography

[TBD]