ITU-T Technical Paper

(07/2020)

FSTP.ACC-WebVRI

Guideline on web-based remote sign language interpretation or video remote interpretation (VRI) system

Technical Paper ITU-T FSTP.ACC-WebVRI

Guideline on web-based remote sign language interpretation or video remote interpretation (VRI) system

Summary

Due to the COVID-19 pandemic, the practice of physical distancing makes it difficult for a sign language interpreter to accompany a deaf or a hard of hearing person when the latter visits places such as a government agency, a school, a meeting, or a hospital. It is now almost imperative that a remote sign language interpretation, or a video remote interpretation (VRI) be implemented.

During the time of physical distancing when almost any schooling and medical consultation needs to be done remotely, a non-interoperable VRI system for deaf and hard of hearing persons will exclude them from important social services. It is therefore important to have a standard guideline for a VRI or VRI system, which considers interoperability and future effectiveness.

Considering the immediacy of the need as well as the cost of system introduction and practicality of the implementation, such a guideline is most likely to be based on web-based technologies.

This Technical Paper describes a web-based VRI, based on Web real time communication (RTC), and describes how it can be used in a scenario where community sign language interpreters can participate, as well as ways in which other remote services, online medical treatment and distance education, can harmonize with the Web-based VRI system.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1 of the ITU-T Technical Paper on "*Guideline on Web-based remote sign language interpretation (VRI)*" approved at the ITU-T Study Group 16 virtual meeting held 22 June – 3 July 2020.

Keywords

Accessibility, disaster resilience, distance learning, online medicine, remote sign language interpretation, total conversation, video remote interpretation, VRI, WebRTC.

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.

© ITU 2020

Table of Contents

List of Figures

Technical Paper ITU-T FSTP.ACC-WebVRI

Guideline on web-based remote sign language interpretation or video remote interpretation (VRI) system

1.　　Scope

This Technical Paper publication is of a non-normative nature and describes a guideline on introducing a web-based remote sign language interpretation system, or video remote interpretation (VRI) system. It describes the general requirements of a VRI system, terminals, and VRI functional components. It introduces a VRI system based on Web real time communication (RTC), and describes how it can be used in a scenario where community sign language interpreters can participate, as well as ways in which other remote services such as online medical treatment and distance education, can harmonize with the web-based VRI system.

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 E.800 (09/2008)]

[ITU-T F.703 (11/2000)]

[ITU-T F.742 (09/2005)]

[ITU-T F.791 (08/2018)]

[ITU-T F.930 (03/2018)]

[ITU-T G.711 (11/1988)]

[ITU-T H.264 (V14) (08/2021)]

[ITU-T H.265 (V7) (11/2019)]　　Recommendation ITU-T H.265 (V7) (11/2019) | ISO/IEC 23008-2, *Recommendation ITU-T H.265 (2019), High efficiency video coding.*.

[ITU-T H.702 (08/2020)]

[ITU-T V.34 (10/1996)]

[ISO 717-1]　　ISO 717-1, *Acoustics—Rating of sound insulation in buildings and of building elements—Part 1: Airborne sound insulation*, Fourth edition.

[ISO 9241-210]　　ISO 9241-210, *Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems*, Second edition.

[ISO/IEC 18004]　　ISO/IEC 18004, *Information technology—Automatic identification and data capture techniques—QR code bar code symbology specification*, Fourth edition.

[IETF RFC 1738]	IETF RFC 1738 (1994), *Uniform Resource Locators (URL)*.
[IETF RFC 1983]	IETF RFC 1983 (1996), *Internet Users' Glossary*.
[IETF RFC 2818]	IETF RFC 2818 (2000), *HTTP Over TLS*.
[IETF RFC 3551]	IETF RFC 3551 (2003), *RTP Profile for Audio and Video Conferences with Minimal Control*.
[IETF RFC 3711]	IETF RFC 3711 (2004), *The Secure Real-time Transport Protocol (SRTP)*.
[IETF RFC 4347]	IETF RFC 4347 (2006), *Datagram Transport Layer Security*.
[IETF RFC 4733]	IETF RFC 4733 (2006), *RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals*.
[IETF RFC 5246]	IETF RFC 5246 (2008), *The Transport Layer Security (TLS) Protocol Version 1.2*.
[IETF RFC 5321]	IETF RFC 5321 (2008), *Simple Mail Transfer Protocol*.
[IETF RFC 5322]	IETF RFC 5322 (2008), *Internet Message Format*.
[IETF RFC 5764]	IETF RFC 5764 (2010), *Datagram Transport Layer Security (DTLS) Extension to Establish Keys for the Secure Real-time Transport Protocol (SRTP)*.
[IETF RFC 6101]	IETF RFC 6101 (2011), *The Secure Sockets Layer (SSL) Protocol Version 3.0*.
[IETF RFC 6347]	IETF RFC 6347 (2012), *Datagram Transport Layer Security Version 1.2*.
[IETF RFC 6716]	IETF RFC 6716 (2012), *Definition of the Opus Audio Codec*.
[IETF RFC 768]	IETF RFC 768 (1980), *User Datagram Protocol*.
[IETF RFC 7874]	IETF RFC 7874 (2016), *WebRTC Audio Codec and Processing Requirements*.
[IETF RFC 793]	IETF RFC 793 (1981), *Transmission Control Protocol*.
[IETF RFC 8445]	IETF RFC 8445 (2018), *Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal*.
[W3C-WebRTC]	W3C-WebRTC, W3C *WebRTC 1.0, Real-time Communication Between Browsers*.

3. Definitions

3.1. Terms defined elsewhere

This Technical Paper uses the following terms defined elsewhere:

3.1.1. distance learning: [ITU-T F.742 (09/2005)]: Learning experiences and environments distributed over space and time (asynchronous learning). In [ITU-T F.742 (09/2005)], it refers to distance learning using telecommunication services over telecommunication networks.

3.1.2. quality of service: [ITU-T E.800 (09/2008)]: Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.

3.1.3. sign language: [ITU-T F.930 (03/2018)]: A natural language that, instead of relying on acoustically conveyed sound patterns, uses signs made by moving the hands combined with facial expressions and postures of the body to convey meaning It is also called signed language or simply visual signing.

3.1.4. sign language interpretation: [ITU-T F.791 (08/2018)]: Synchronized showing of an interpreter who uses sign language to convey the main audio content and dialogue to people who use sign language.

3.1.5. sound transmission class: [b-ASTM E413-16]: An integer rating of how well a building partition attenuates airborne sound.

3.1.6. user experience: [ITU-T F.791 (08/2018)]: Person's perceptions and responses resulting from the use or anticipated use of a product, system or service, including navigation of physical and virtual environment.

NOTE 4: Adapted from [ISO 9241-210].

NOTE 1 – User experience includes all the user's emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments that occur before, during and after use.

NOTE 2 – User experience is a consequence of brand image, presentation, functionality, system performance, interactive behaviour and assistive capabilities of the interactive system, the user's internal and physical state resulting from prior experiences, attitudes, skills and personality, as well as the context of use.

NOTE 3 – Usability, when interpreted from the perspective of the user's personal goals, can include the kind of perceptual and emotional aspects typically associated with user experience. Usability criteria can be used to assess aspects of user experience.

3.1.7. video relay (or video-to-speech relay): [ITU-T F.930 (03/2018)]: A telecommunications relay service that allows communication by individuals with speech and hearing disabilities. Visual information is converted by a communication assistant (CA) to verbal information over a voice telecommunication service.

NOTE – Video relay allows persons with hearing or speech disabilities who use sign language to communicate with voice telephone users through video equipment. The video link allows the CA to view and interpret the party's signed conversation (or visual communication) and relay the conversation back and forth with a voice caller.

3.2.   Terms defined in this Technical Paper

This Technical Paper defines the following terms:

3.2.1. key performance indicators: A set of measurement criteria that helps define the degree of achievement of a certain business or service.

3.2.2. online medical examination recommendation: Medical examination recommendation with the minimum necessary medical judgment according to the condition of the patient, based on the complaints of symptoms from patients and the collected information, via e.g., interviews on mental and physical conditions, the suspected diseases are judged, the names of the diseases are listed, and the appropriate department to be consulted is selected.

3.2.3.   video remote interpretation (VRI): A system or service that provides a remote sign language interpretation between a sign language user and a hearing person using a telecommunication means.

NOTE – Video remote interpretation (VRI) can also mean any remote interpretation, but the use in this Recommendation is for the case where sign language interpretation is involved.

3.2.4.   VRI agent: The sign language interpreter that performs the remote sign language interpretation in a video remote interpretation (VRI) session.

3.2.5.   VRI client: An entity that requests a video remote interpretation (VRI) session, accepts the VRI notification, and further transmits VRI session data to a VRI agent using a VRI client terminal, thus initiating the VRI session.

3.2.6.   VRI coordinator: An entity that accepts the request from the video remote interpretation (VRI) client, contacts a VRI agent, arranges a VRI session. It further sends VRI notification to the VRI client.

3.2.7.   VRI operation: An act of providing a remote sign language interpretation service using video and telecommunication.

3.2.8.   VRI session: A period that is marked by the beginning and the end of video remote interpretation (VRI) operation.

NOTE – The beginning is usually pre-arranged, except for the time of emergency.

4.   Abbreviations and acronyms

This Technical Paper uses the following abbreviations and acronyms:

AES	Advanced Encryption Standard
AFCEA	Armed Forces Communications and Electronics Association
AV1	AOMedia Video 1
DTLS	Datagram Transport Layer Security
HTML	Hypertext Mark-Up Language
HTTP	Hypertext Transfer Protocol
ICE	Interactive Connectivity Establishment
ICT	Information Communication Technology
IP	Internet Protocol
KPI	Key Performance Indicator
MD	Medical Doctor
NAT	Network Address Translator
P2P	Peer-to-Peer
PWD	Persons With Disabilities
QoS	Quality of Service
RTC	Real Time Communication
RTT	Real-Time Text
SL	Sign Language
SRTP	Secure Real-time Transport Protocol

STC	Sound Transmission Class
STUN	Session Traversal Utilities for NAT
TCP	Transmission Control Protocol
TRS	Telecommunication Relay Service
TURN	Traversal Using Relay around NAT
UDP	User Datagram Protocol
URL	Universal Resource Locator
VRI	Video Remote Interpretation
VRS	Video Relay Service

5.　　Background

Due to the COVID-19 pandemic, the practice of physical distancing makes it difficult for a sign language interpreter to accompany a deaf or a hard of hearing person when the latter visits places such as a government agency, a school, a meeting, or a hospital. It is now almost imperative that a remote sign language interpretation, or a video remote interpretation (VRI) system be implemented and made available to deaf and hard of hearing persons, not only for access to communication but also for protecting the lives of sign language interpreters.

However, there are many video conferencing tools and it may be difficult immediately for (local) governments and communities to choose a VRI system off-the-shelf. Moreover, if each local government or community introduces a system based on its own proprietary specifications, it will hinder interoperability among the systems, especially during emergencies and will hinder cooperation in the medical, emergency, and welfare fields.

Especially during the time of physical distancing when almost any schooling and medical consultation need to be done remotely, a non-interoperable VRI system for deaf and hard of hearing persons will exclude them from important social services. It is therefore important to have a standard guideline for a VRI or VRI system, which considers interoperability and future effectiveness.

Considering the immediacy of the need as well as the cost of system introduction and practicality of the implementation, such a guideline is most likely to be based on web-based technologies.

6.　　Video remote interpretation (VRI) system

Figure 1 shows the general case of a web VRI system in a case where both the person with disabilities (PWD) and the hearing person are in the same location.
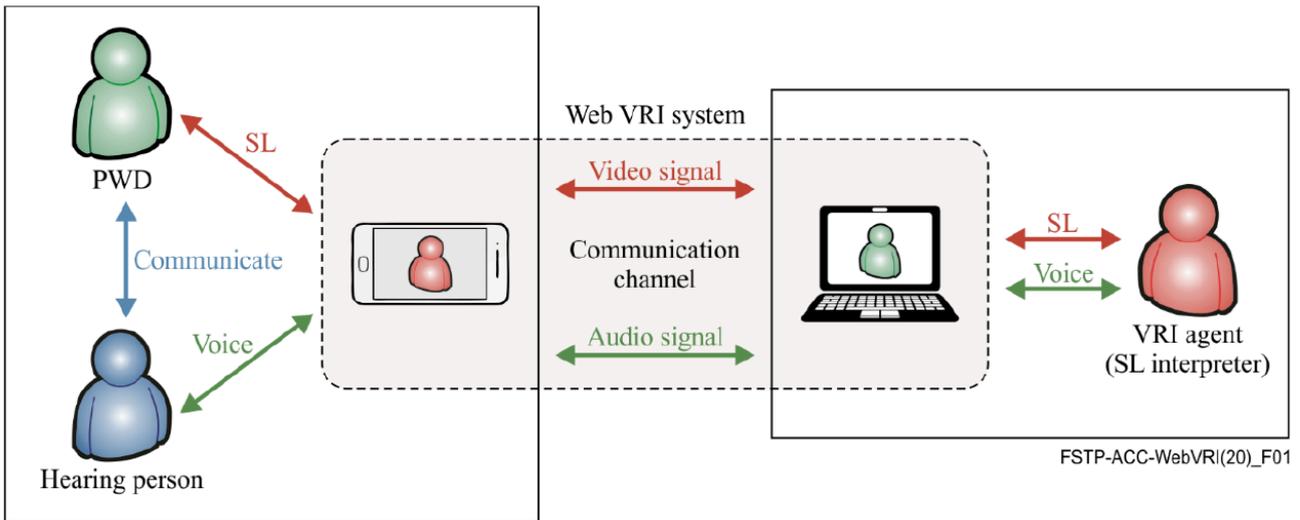
Figure 1 — Both PWD and hearing person (e.g., an MD) are in the same location

The VRI service also includes the case where a PWD user and a hearing user both participate in an online conversation, such as an online medical consultation, and a VRI is provided for the PWD. This case is shown in Figure 2.
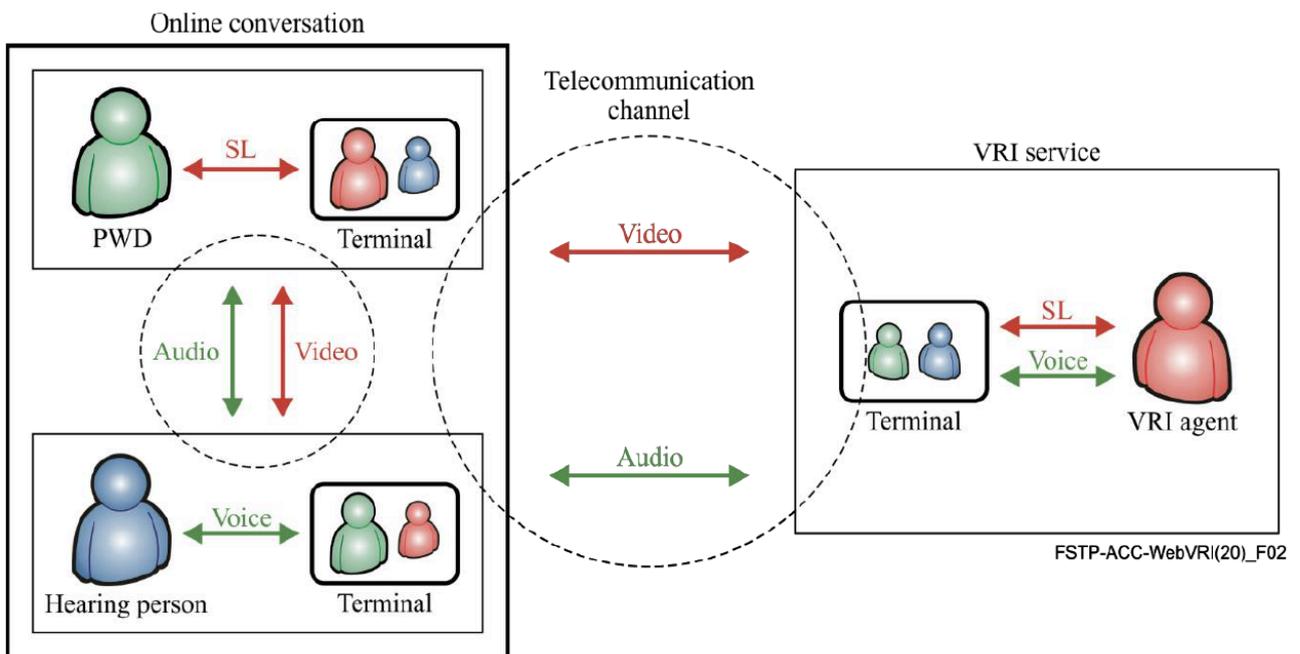
Figure 2 — PWD and hearing person are in
different locations (e.g., telemedicine, tele-education)

7. General requirements

The following is a non-exhaustive list of general requirements:

R1: It is recommended that the system can operate in accordance with the workflow of sign language interpreters in the community;

R2: It is required that the system allows the sign language interpreter to perform remote sign language interpretation at a safe place;

R3: It is recommended that the system allows the deaf and hard of hearing persons to use their own smartphones or tablets;

R4: It is recommended that the system is based on standards and not be implementable by a single terminal equipment vendor in order to avoid vendor lock-in and interfering with fair competition;

R5: It is required that the system protects confidentiality of privacy and interpreted content during the remote sign language interpretation;

R6: It is required that the system does not impose any complications that might increase vulnerabilities of deaf and hard of hearing persons in operating terminals and systems. (e.g., App download, etc.);

R7: It is recommended that the system can meet the advances in communication technology;

R8: It is recommended that the system is portable and resistant to disasters;

R9: It is recommended that the system does not require many complicated processes such as application download, installation, ID registration, etc.

NOTE – Downloading and installing special applications is not desirable from the viewpoint of safety and ease of use by users. One of the problems from the viewpoint of safety when downloading an app is that the reliability of the app may be unclear. Downloading and installing a new app has the risk of potentially introducing malware. Especially in the case of a smartphone, the user does not know what kind of software is installed in the terminal along with the application. Also, it may not be known where the download source is; there are many cases where malicious apps are downloaded due to spoofing. See [b-WebRTC-Security] for more information.

8. Web-based video technology

In order to meet the general requirement "R9: It is recommended that the system does not require application download, installation, ID registration, etc." mentioned in clause 7, it is recommended to use the method to connect users to a website and to accept and implement services through this website through WebRTC

8.1. WebRTC

Web real time communication (WebRTC) refers to a set of standard protocols centred on [W3C-WebRTC] that enables peer-to-peer (P2P) sharing of audio and video signals between browsers. It includes JavaScript APIs and protocols such as [IETF RFC 7874], [IETF RFC 3551], [IETF RFC 4733] and [ITU-T G.711 (11/1988)]. P2P is a communication method characterized in that when communicating between multiple terminals, peers and peers communicate with each other, and tasks or workloads are equally divided among the peers. In WebRTC, the browser is the peer.

The WebRTC standard roughly consists of two technologies: media acquisition and P2P connection. Media acquisition mainly involves video cameras and microphones. P2P communication is specified in [W3C-WebRTC]. For the WebRTC software architecture, see [b-WebRTC-Overview].

8.2. VRI sequence

Figure 3 shows the sequence of a VRI session using WebRTC.

FSTP-ACC-WebVRI(20)_F03
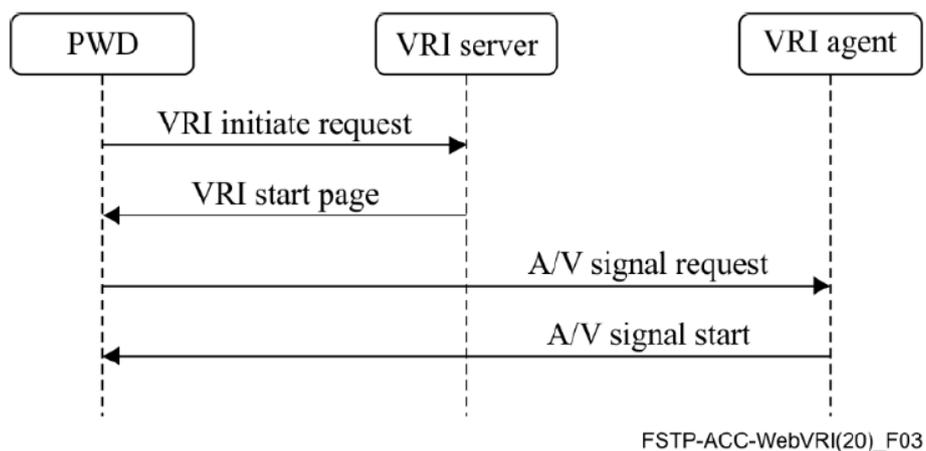
Figure 3 — Sequence of a VRI session

8.2.1. Media acquisition

It is recommended that the following codecs be used.

8.2.1.1. Audio codecs
– [IETF RFC 6716]
– [ITU-T G.711 (11/1988)]

8.2.1.2. Video codecs
– VP8
– VP9
– [ITU-T H.264 (V14) (08/2021)]
– [ITU-T H.265 (V7) (11/2019)]
– [b-AV1]

8.2.2. Security

The WebRTC part must be protected according to the following WebRTC standards:

8.2.2.1. Datagram transport layer security (DTLS)

Datagram transport layer security (DTLS) is a communication protocol designed to protect the privacy of data and prevent eavesdropping and tampering, as defined by [IETF RFC 4347] and [IETF RFC 6347]. It is based on the transport layer security (TLS) protocol [IETF RFC 5246] for the security of communication networks. The main difference between DTSL and TLS is that DTLS uses user datagram protocol (UDP) [IETF RFC 768] and TLS uses transmission control protocol (TCP).

8.2.2.2. Secure real-time transport protocol (SRTP)

SRTP is an encrypted RTP that is used to send and receive encrypted audio and video. SRTP is defined in [IETF RFC 3711].

8.2.2.3. Encryption

In the case of encryption, WebRTC uses the standard encryption algorithm, which is widely used internationally. Accordingly, advanced encryption standard (AES) shall be used.

8.3. Quality of service (QoS)

A VRI system must take quality of service quality of service (QoS) into consideration. QoS KPIs are for future study.

9. General requirements of web VRI service operation

Figure 4 shows that service architecture for the web-based remote sign language interpretation service, or web VRI service for short.
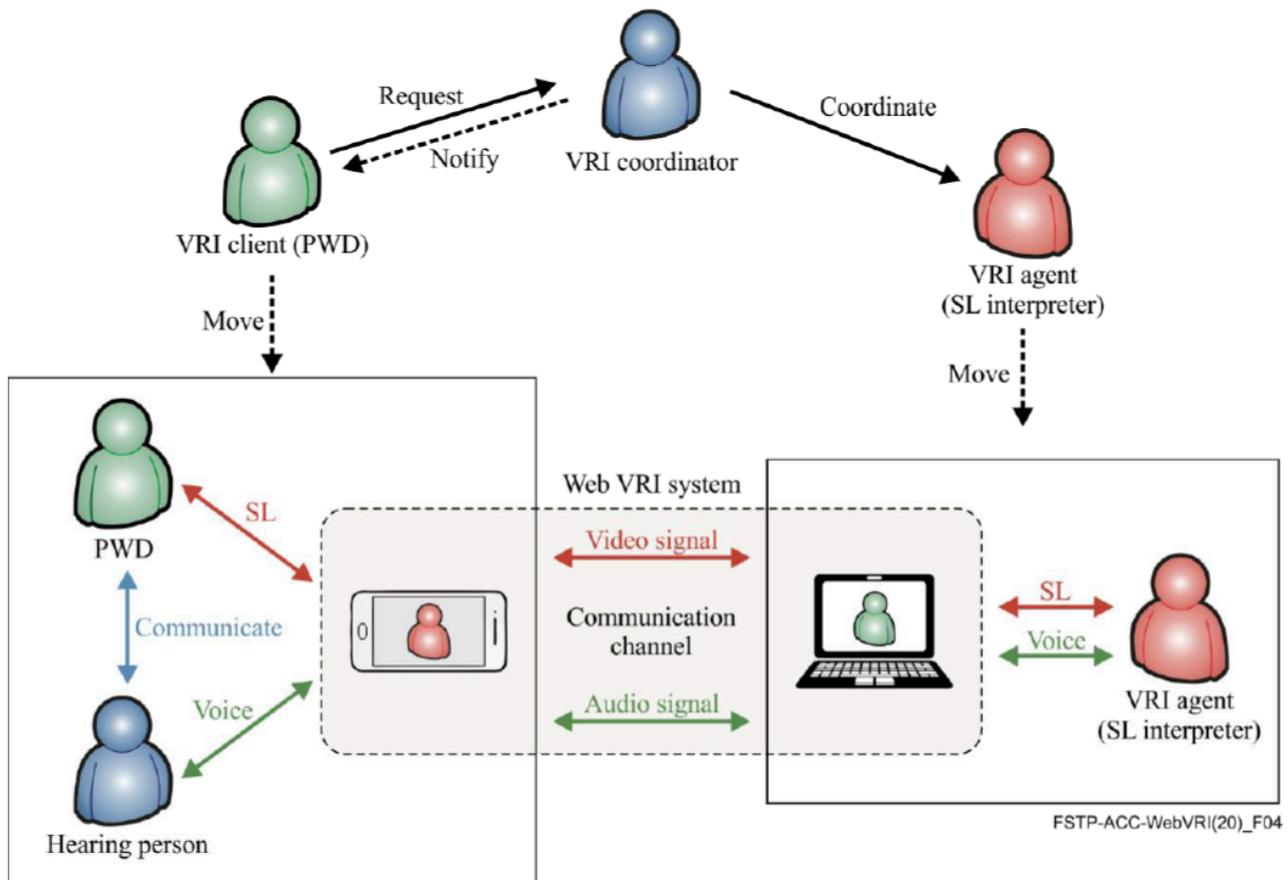
Figure 4 — Remote sign language interpretation service architecture

The architecture includes three VRI functional components: VRI client function, VRI coordinator function and VRI agent function. This clause describes these functions.

9.1. VRI functions

9.1.1. VRI client function

An entity that requests a VRI session, accepts the VRI notification, and further transmits VRI session data to a VRI agent using a VRI client terminal, thus initiating the VRI session. This is the function that the deaf or hard of hearing person plays.

9.1.2. VRI coordinator function

The VRI coordinator function accepts the VRI request from the client, contacts a VRI agent, and arranges the VRI session. The VRI coordinator function includes the VRI notification function, and, using the VRI notification function, it generates and returns the VRI notification to the VRI client.

NOTE – If the VRI session is expected to be long, more than one sign language interpreter must be assigned.

9.1.3. VRI agent function

The VRI agent function is the function to perform VRI operation via the VRI system.

A VRI agent waits at the booth at a predetermined time for his/her safety as well as for protecting the confidentiality of the contents of the VRI session, and performs sign language interpretation remotely. The VRI agent terminal receives the VRI session data, whereupon the VRI session is started.

NOTE – In the event of an emergency, there is a possibility that the VRI agent operates at home. In this case, it is necessary to provide conditions such that the confidentiality can be sufficiently protected, such as by having a home-based booth function. Refer to clause 10.4.1 for the requirements on a booth.

9.2. VRI notification function

The VRI notification function is part of the VRI coordinator function. After adjusting the VRI request, the confirmed result is stored as a VRI notification data item, and then as a VRI notification form. A function that generates and returns to the VRI client.

9.2.1. VRI notification data format and communication format

This clause describes the data format and communication format of the dispatch notice sent by the VRI coordinator to the dispatch requester after adjusting the VRI service provider.

9.2.2. VRI notification data item

It is recommended that at least the following data items are included in the VRI notification data:
– date and time the VRI request was received
– scheduled date and time for the VRI session
– planned place for VRI
– name(s) of the VRI agent(s) (sign language interpreter(s))
– VRI session data

9.2.3. VRI session data

The VRI session data includes the information necessary for connecting the client's VRI receiving terminal to the VRI agent's terminal.

9.2.4. VRI session data structure

The following metadata is stored in the VRI session data.
– VRI scheduled date and time
– VRI server URL

9.2.5. VRI session data encoding format

For the convenience of the VRI client, encode the VRI session data by the following method. Also, it is desirable to use both methods to ensure that the VRI client can obtain the notification.
– QR code: 2D barcode according to [ISO/IEC 18004]
– URL: Data from [IETF RFC 1738].

9.2.6. VRI session data transmission method

The following methods are used to notify the VRI notification:
– Facsimile: Method according to [ITU-T V.34 (10/1996)]
– E-mail: Method according to [IETF RFC 5321] and [IETF RFC 5322].

An example of the above implementation is shown in Figure I.1.

10. General requirements on terminals

10.1. Requirements for VRI client terminal function

The following describes the recommended functional requirements for terminals that can connect to the VRI client terminal system and receive a VRI service. Figure 5 shows the recommended functional specification for VRI terminal equipment.
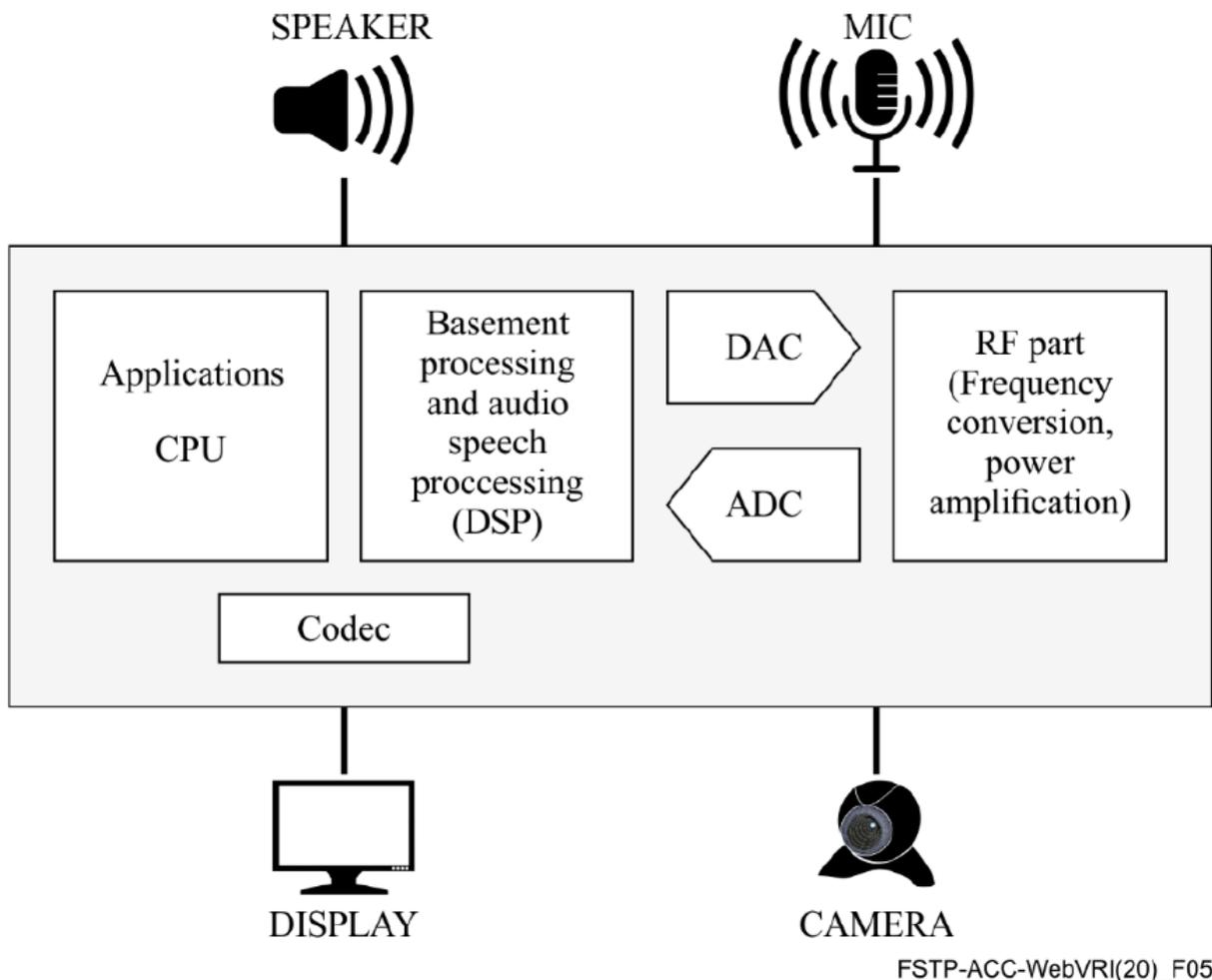
Figure 5 — Recommended functional specification for VRI terminal equipment

10.1.1. IP connection function

The VRI client terminal must have the function of connecting to the IP network.

10.1.2. Wireless communication function

The VRI client terminal must have a wireless communication function.

10.1.3. Audio input function (microphone)

The VRI client terminal must have an input function for voice. It may be used as a headset together with headphones or earphones. The performance of voice input shall be specified separately.

It is required that the VRI client terminal owned by deaf people not be set to the mute mode.

10.1.4. Audio output function (speaker)

The VRI client terminal must have the function of reproducing voice data and transmitting it as sound to the outside. In this case, it may be used as a headset together with the microphone.

It is required that the VRI client terminal owned by deaf people not be set to the mute mode.

10.1.5. Camera

The VRI client terminal must have the function of converting the video data into data for communication. Built-in type is preferred.

The requirements for camera performance will be specified separately.

10.1.6. Video display function

The VRI client terminal is required to have the function to acquire the video data from the communication channel, render it as video and display it. The video display should have sufficient size and resolution so that the sign language is understandable. The requirements for the performance of the video display function will be specified separately.

10.2. Requirements for VRI agent terminal user interface

The user interface of the VRI client terminal shall implement the following conditions in consideration of the ease of use by PWD. For user experience, standards such as [ISO 9241-210] should be consulted.

NOTE – However, it is possible to add designs such as logos according to the characteristics of each area, as long as the operability is not hindered.

10.2.1. Minimal motion

There should be as few screen operations as possible before the video of the VRI is displayed.

10.2.2. Use of symbolism

It is recommended that the information about the operation on the screen be understandable without reading text, and the use of images such as icons is recommended.

10.2.3. Simplicity of displayed information

It is recommended to avoid providing too much unnecessary information and that the minimum essential information is presented so that the operations on the screen are easy to understand.

10.2.4. Ease of operation

It is recommended that the user interface allows exchange of necessary information without complicated operations.

10.2.5. Responsive reaction

It is required that the user interface is responsive and returns an easy-to-understand response to the user's operation: e.g., it should be clear whether the operation has been accepted, rejected, is on hold, or has ended.

10.3. Requirements for VRI agent terminal function

The following describes the specifications for a terminal that allows the VRI agent to connect to a VRI system and provide a VRI session.

10.3.1. IP connection function

The VRI agent terminal for the VRI must have the function of connecting to the IP network.

10.3.2. Wireless communication function

The VRI agent terminal for the VRI preferably has a wireless communication function.

10.3.3. Audio input function (microphone)

The VRI agent terminal must have an input function for voice. The performance of voice input shall be specified separately. Use of external input microphone or headset type microphone is preferable.

10.3.4. Audio output function (speaker)

The VRI agent terminal must have the function of playing voice data and transmitting it outside. In this case, it is assumed that the headphones and earphones are used together. A headset type speaker is desirable.

10.3.5. Camera

The VRI agent terminal must have the function of converting the video data into data for communication. It must have sufficient performance so that the video display on the VRI client terminal is clear.

10.3.6. Video display function

The VRI agent terminal shall have the function to acquire the video data from the communication channel, render it as video and display it. The video display of the VRI agent must be large enough and capable of reading the sign language of the VRI client.

10.4. Requirements on the instalment of a VRI agent terminal

The VRI agent terminal and its peripherals should be installed in a booth so that the VRI agent's safety and the contents of the interpretation can be kept secret. Only one sign language interpreter is generally expected to be accommodated in the booth at a time. However, it is possible to use one that can accommodate more than one person, such as a supervisor.

The booths should be distributed among multiple locations with at least two locations to ensure the VRI service would have high availability and robustness.

10.4.1. Booth

The booth must have sufficient insulation for soundproofing to maintain confidentiality. It is recommended that the has the sound insulation performance equivalent to that of sound transmission class STC 35 or higher, as defined in [b-ASTM E413-16] or its equivalent by [ISO 717-1].

11. Coordination with other online conversational services

11.1. VRI and Telemedicine

Under COVID-19, medical treatment online, including by telephone, is recommended to prevent infections. For example, it is expected that a regular patient with chronic disease can receive medical examination and consultation (and treatment) as well as prescriptions via telephones and other information communication technology (ICT) devices. Therefore, it is important that a VRI system can be used alongside telemedicine services.

11.1.1. Telemedicine and online medical services

Figure 6 shows the demarcation among online medical treatment, online medical examination recommendation, and telehealth consultation in the general context of telemedicine.
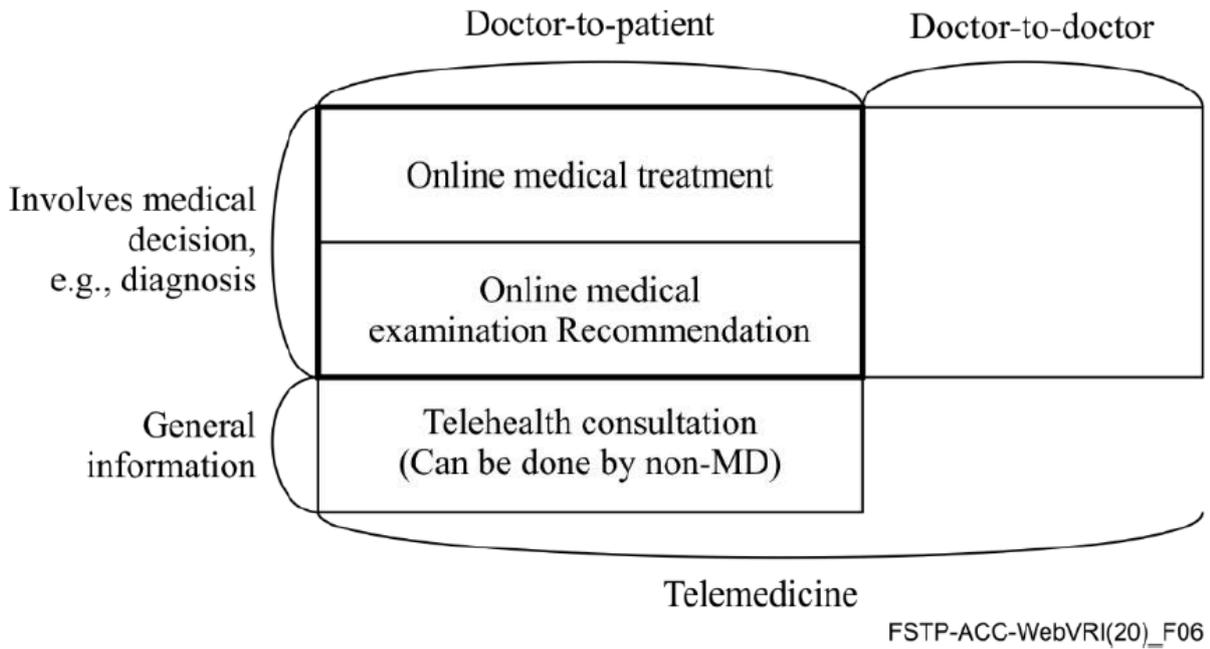
Figure 6 — Demarcation between telemedicine and online medical services

As can be seen from this figure, telemedicine includes both DtoD and DtoP, while online medical services are DtoP. VRI service is mainly required for DtoP, such as online medical treatment.

11.1.2.   Online medical treatment and VRI

Figure 7 shows the sequence diagram that can be considered for cooperation between the VRI and online medical care.
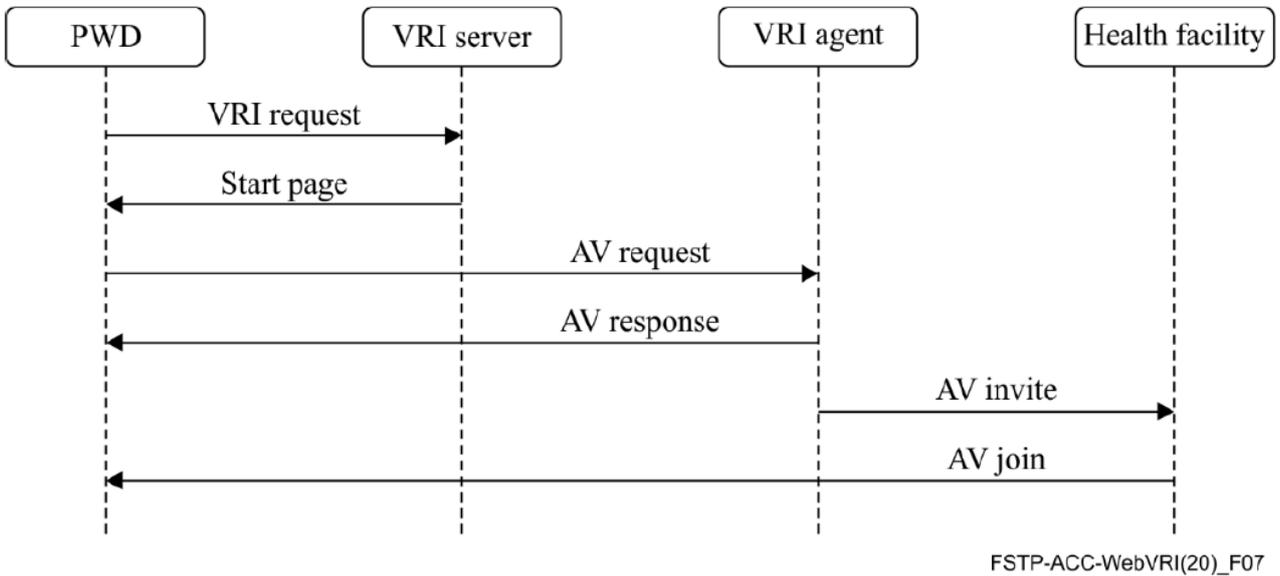
Figure 7 — VRI Sequence with online medical treatment

It is assumed that the VRI client calls the VRI service and starts the VRI session. It is also assumed that the medical institution has a communication environment capable of receiving the communication requests (voice / video invitation) from the VRI agent.

The VRI client terminal must be able to display and play two sets of audio and video signals: the VRI agent and the medical personnel. Also, it is usually desirable to be able to display a video monitoring

the client. In addition, it is desirable that the image of the VRI agent is at least as large as the medical personnel so that the sign language can be easily read.

Further details of the interaction between a web-VRI system and online medical treatment are for future study.

11.2. Online education and distance learning

When a VRI is provided to an online class, the method of VRI operation differs depending on the tool of the class. Therefore, it is desirable to formulate rules (protocol) for VRI operation methods for each tool.

It is also desirable to take into account the guidelines for providing accessibility at remote international conferences such as [b-ITU-RemPart].

Further, according to the general requirements of distance learning, the VRI system is recommended to meet at least the following requirements for cooperation with distance learning.
–   A flexible architecture be used so that terminals, software, communication networks, etc. can be selected from multiple perspectives.
–   Based on standard software that can be used in multiple fields, not as special specifications for schools.
–   Capable of introducing cloud computing, not restricted by hardware renewal cycle.
–   Able to introduce a thin client environment.
–   Capable of using secure browsers.
–   It should be possible to limit connection only with applications with a file encryption function.

In the case of an online class, it may be difficult for a VRI client to receive a VRI service on a smartphone. It is also possible to use the IPTV set-top box or TV receiver, as defined in [ITU-T H.702 (08/2020)] as a terminal for VRI.

In addition, when captioning is provided, it is necessary to guarantee quality by humans so that text is not generated using only automatic speech recognition, as shown in [b-WFD-ASR].

11.3. Disaster resilience and VRI

It has been pointed out that the VRI is effective for the deaf and hard of hearing people and others in the disaster-stricken area to obtain information.

Taking COVID-19 into account, evacuation shelters are required to take measures against infectious diseases, and disaster prevention manuals in various places should be revised accordingly. As SL interpreters are most likely unavailable on site in such an event, VRI service must be included in disaster prevention measures.

11.4. Total conversation and VRI

As described in [ITU-T F.703 (11/2000)], it is recommended that the system described here be extended to include not only sign language but also communication support using text and images by total conversation. It is recommended that this Technical Paper be applied to communication support for hard of hearing people who do not use sign language. In such a case, as stated in [b-WFD-ASR], it is required that the quality of the captioned text be checked and guaranteed by humans rather than using only automatic speech recognition.

Appendix I

Implementation example

(This appendix does not form an integral part of this Technical Paper.)
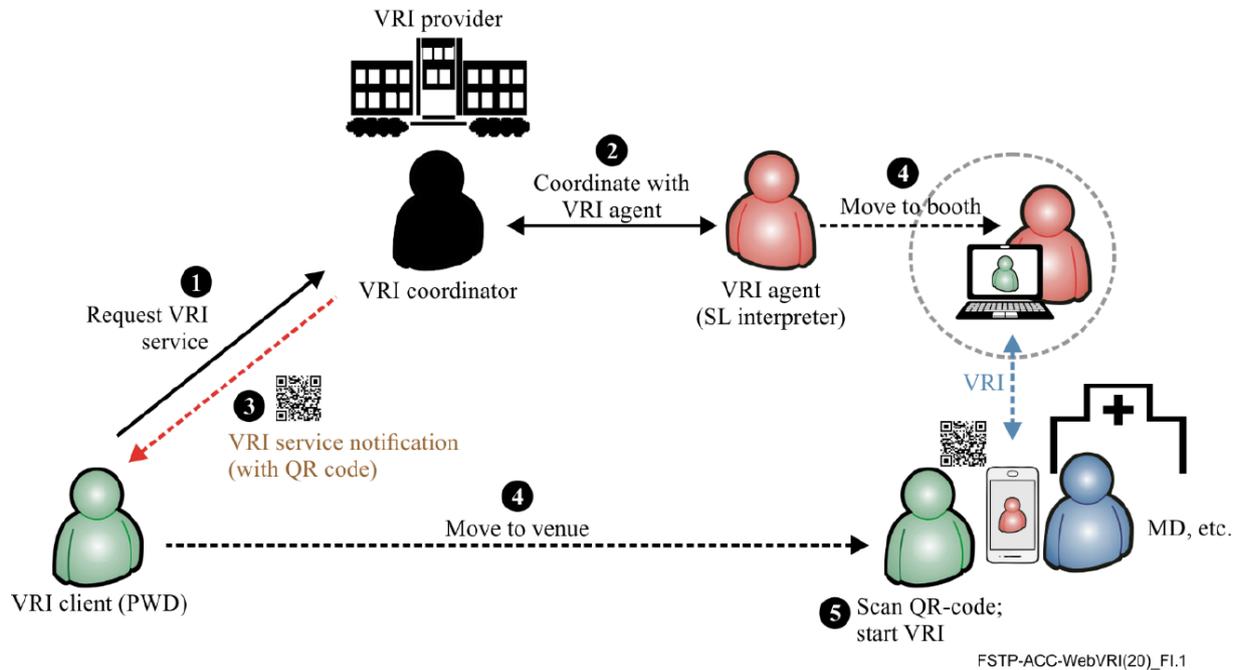
Figure I.1 — Example of web VRI service

The following is the VRI workflow based on this guideline, as depicted in Figure I.1:

1) The VRI client executes a VRI request.
2) After receiving the VRI request from the VRI client, the VRI coordinator contacts and coordinates with the registered VRI agent.
3) The result adjusted and confirmed by the VRI coordinator is generated as a VRI notification (with QR code) using the VRI notification function and sent to the VRI client by email or fax.
4) The VRI client receives the VRI notification and moves to the place where the VRI session was requested at the requested time.
   The VRI agent moves to the booth and waits at the assigned time.
5) The VRI client accesses the URL in the QR code or email on his or her terminal, and the VRI agent starts the VRI session.

Bibliography

[b-ASTM E413-16]   b-ASTM E413-16, ASTM International *Classification for Rating Sound Insulation.*

[b-AV1]   b-AV1, *AV1 Bitstream & Decoding Process Specification.* https://aomediacodec.github.io/av1-spec/av1-spec.pdf.

[b-ITU-RemPart]   b-ITU-RemPart, ITU-T FSTP-ACC-RemPart, *Guidelines for supporting remote participation in meetings for all.*

[b-ITU-T H.Sup1]   b-ITU-T H.Sup1, ITU-T H-Series Supplement 1 (1999), *Application profile – Sign language and lip-reading real-time conversation using low bit rate video communication.*

[b-ITU-T L.Sup35]   b-ITU-T L.Sup35, ITU-T L-Series Supplement 35 (2006) *Framework of disaster management for network resilience and recovery.*

[b-WFD-ASR]   b-WFD-ASR, WFD and IFHOH Joint Statement: *Automatic Speech Recognition in Telephone Relay Services and in Captioning Services.*

[b-WebRTC-DOD]   b-WebRTC-DOD, Steven Boberski (2019), How WebRTC Can Benefit the Department of Defense, SIGNAL Media, January 30, published by *Armed Forces Communications and Electronics Association*, available at: https://www.afcea.org/content/how-webrtc-can-benefit-department-defense.

[b-WebRTC-Overview] b-WebRTC-Overview, Overview: Real Time Protocols for Browser-based Applications. H. Alvestrand. IETF. 14 February 2014. Active Internet-Draft. URL: https://tools.ietf.org/html/draft-ietf-rtcweb-overview.

[b-WebRTC-Security] b-WebRTC-Security, NTT Communications, A Study of WebRTC Security, available at: https://webrtc-security.github.io/.