Network Working Group
Request for Comments: 3351
Category: Informational

N. Charlton
Millpark
M. Gasson
Koru Solutions
G. Gybels
M. Spanner
RNID
A. van Wijk
Ericsson

August 2002

User Requirements for the Session Initiation Protocol (SIP) in Support of Deaf, Hard of Hearing and Speech-impaired Individuals

Status of this Memo

Copyright Notice

Abstract

This document presents a set of Session Initiation Protocol (SIP) user requirements that support communications for deaf, hard of hearing and speech-impaired individuals. These user requirements address the current difficulties of deaf, hard of hearing and speech-impaired individuals in using communications facilities, while acknowledging the multi-functional potential of SIP-based communications.

A number of issues related to these user requirements are further raised in this document.

Also included are some real world scenarios and some technical requirements to show the robustness of these requirements on a concept-level.

Table of Contents

1. Terminology and Conventions Used in this Document

In this document, the key words "MUST", "MUST NOT","REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in BCP 14, RFC2119[1] and indicate requirement levels for compliant SIP implementations.

For the purposes of this document, the following terms are considered to have these meanings:

Abilities: A person’s capacity for communicating which could include a hearing or speech impairment or not. The terms Abilities and Preferences apply to both caller and call-recipient.

Preferences: A person’s choice of communication mode. This could include any combination of media streams, e.g., text, audio, video.

The terms Abilities and Preferences apply to both caller and call-recipient.

Relay Service: A third-party or intermediary that enables communications between deaf, hard of hearing and speech-impaired people, and people without hearing or speech-impairment. Relay Services form a subset of the activities of Transcoding Services (see definition).

Transcoding Services: A human or automated third party acting as an intermediary in any session between two other User Agents (being a User Agent itself), and transcoding one stream into another (e.g., voice to text or vice versa).

Textphone: Sometimes called a TTY (teletypewriter), TDD (telecommunications device for the deaf) or a minicom, a textphone enables a deaf, hard of hearing or speech-impaired person to place a call to a telephone or another textphone. Some textphones use the V.18[3] protocol as a standard for communication with other textphone communication protocols world-wide.

User: A deaf, hard of hearing or speech-impaired individual. A user is otherwise referred to as a person or individual, and users are referred to as people.

Note: For the purposes of this document, a deaf, hard of hearing, or speech-impaired person is an individual who chooses to use SIP because it can minimize or eliminate constraints in using common communication devices. As SIP promises a total communication solution for any kind of person, regardless of ability and preference, there is no attempt to specifically define deaf, hard of hearing or speech-impaired in this document.

2. Introduction

The background for this document is the recent development of SIP[2] and SIP-based communications, and a growing awareness of deaf, hard of hearing and speech-impaired issues in the technical community.

The SIP capacity to simplify setting up, managing and tearing down communication sessions between all kinds of User Agents has specific implications for deaf, hard of hearing and speech-impaired individuals.

As SIP enables multiple sessions with translation between multiple types of media, these requirements aim to provide the standard for recognizing and enabling these interactions, and for a communications model that includes any and all types of SIP-networking abilities and preferences.

3. Purpose and Scope

The scope of this document is firstly to present a current set of user requirements for deaf, hard of hearing and speech-impaired individuals through SIP-enabled communications. These are then followed by some real world scenarios in SIP-communications that could be used in a test environment, and some concepts of how these requirements can be developed by service providers and User Agent manufacturers.

These recommendations make explicit the needs of a currently often disadvantaged user-group and attempt to match them with the capacity of SIP. It is not the intention here to prioritize the needs of deaf, hard of hearing and speech-impaired people in a way that would penalize other individuals.

These requirements aim to encourage developers and manufacturers world-wide to consider the specific needs of deaf, hard of hearing and speech-impaired individuals. This document presents a world-vision where deafness, hard of hearing or speech impairment are no longer a barrier to communication.

4. Background

Deaf, hard of hearing and speech-impaired people are currently often unable to use commonly available communication devices. Although this is documented[4], this does not mean that developers or manufacturers are always aware of this. Communication devices for deaf, hard of hearing and speech-impaired people are currently often primitive in design, expensive, and non-compatible with progressively designed, cheaper and more adaptable communication devices for other individuals. For example, many models of textphone are unable to communicate with other models.

Additionally, non-technical human communications, for example sign languages or lip-reading, are non-standard around the world.

There are intermediary or third-party relay services (e.g. transcoding services) that facilitate communications, uni- or bi-directional, for deaf, hard of hearing and speech-impaired people. Currently relay services are mostly operator-assisted (manual), although methods of partial automation are being implemented in some areas. These services enable full access to modern facilities and conveniences for deaf, hard of hearing and speech-impaired people. Although these services are somewhat limited, their value is undeniable as compared to their previous complete unavailability.

Yet communication methods in recent decades have proliferated: email, mobile phones, video streaming, etc. These methods are an advance in the development of data transfer technologies between devices.

Developers and advocates of SIP agree that it is a protocol that not only anticipates the growth in real-time communications between convergent networks, but also fulfills the potential of the Internet as a communications and information forum. Further, they agree that these developments allow a standard of communication that can be applied throughout all networking communities, regardless of abilities and preferences.

5. Deaf, Hard of Hearing and Speech-impaired Requirements for SIP

Introduction

The user requirements in this section are provided for the benefit of service providers, User Agent manufacturers and any other interested parties in the development of products and services for deaf, hard of hearing and speech-impaired people.

The user requirements are as follows:

5.1 Connection without Difficulty

This requirement states:

Whatever the preferences and abilities of the user and User Agent, there SHOULD be no difficulty in setting up SIP sessions. These sessions could include multiple proxies, call routing decisions, transcoding services, e.g., the relay service Typetalk[5] or other media processing, and could include multiple simultaneous or alternative media streams.

This means that any User Agent in the conversation (including transcoding services) MUST be able to add or remove a media stream from the call without having to tear it down and re-establish it.

5.2 User Profile

This requirement states:

Deaf, hard of hearing and speech-impaired user abilities and preferences (i.e., user profile) MUST be communicable by SIP, and these abilities and preferences MUST determine the handling of the session.

The User Profile for a deaf, hard of hearing or speech-impaired person might include details about:

- How media streams are received and transmitted (text, voice, video, or any combination, uni- or bi-directional).

- Redirecting specific media streams through a transcoding service (e.g., the relay service Typetalk)

- Roaming (e.g., a deaf person accessing their User Profile from a web-interface at an Internet cafe)

- Anonymity: i.e., not revealing that a deaf person is calling, even through a transcoding service (e.g., some relay services inform the call-recipient that there is an incoming text call without saying that a deaf person is calling).

  Part of this requirement is to ensure that deaf, hard of hearing and speech-impaired people can keep their preferences and abilities confidential from others, to avoid possible discrimination or prejudice, while still being able to establish a SIP session.

5.3 Intelligent Gateways

This requirement states:

SIP SHOULD support a class of User Agents to perform as gateways for legacy systems designed for deaf, hard of hearing and speech-impaired people.

For example, an individual could have a SIP User Agent acting as a gateway to a PSTN legacy textphone.

5.4 Inclusive Design

This requirement states:

Where applicable, design concepts for communications (devices, applications, etc.) MUST include the abilities and preferences of deaf, hard of hearing and speech-impaired people.

Transcoding services and User Agents MUST be able to connect with each other regardless of the provider or manufacturer. This means that new User Agents MUST be able to support legacy protocols through appropriate gateways.

5.5 Resource Management

This requirement states:

User Agents SHOULD be able to identify the content of a media stream in order to obtain such information as the cost of the media stream, if a transcoding service can support it, etc.

User Agents SHOULD be able to choose among transcoding services and similar services based on their capabilities (e.g., whether a transcoding service carries a particular media stream), and any policy constraints they impose (e.g., charging for use). It SHOULD be possible for User Agents to discover the availability of alternative media streams and to choose from them.

5.6 Confidentiality and Security

This requirement states:

All third-party or intermediaries (transcoding services) employed in a session for deaf, hard of hearing and speech-impaired people MUST offer a confidentiality policy. All information exchanged in this type of session SHOULD be secure, that is, erased before confidentiality is breached, unless otherwise required.

This means that transcoding services (e.g., interpretation, translation) MUST publish their confidentiality and security policies.

6. Some Real World Scenarios

These scenarios are intended to show some of the various types of media streams that would be initiated, managed, directed, and terminated in a SIP-enabled network, and shows how some resources might be managed between SIP-enabled networks, transcoding services and service providers.

To illustrate the communications dynamic of these kinds of scenarios, each one specifically mentions the kind of media streams transmitted, and whether User Agents and Transcoding Services are involved.

6.1 Transcoding Service

In this scenario, a hearing person calls the household of a deaf person and a hearing person.

1. A voice conversation is initiated between the hearing participants:

   (Person A) <-----Voice ---> (Person B)

2. During the conversation, the hearing person asks to talk with the deaf person, while keeping the voice connection open so that voice to voice communications can continue if required.

3. A Relay Service is invited into the conversation.

4. The Relay Service transcodes the hearing person’s words into text.

5. Text from the hearing person’s voice appears on the display of the deaf person’s User Agent.

6. The deaf person types a response.

7. The Relay Service receives the text and reads it to the hearing person:

```
(         ) <------------------Voice----------------> (         )
(Person A ) -----Voice---> ( Voice To Text  ) -Text-> (Person B )
(         ) <----Voice---- (Service Provider) <-Text- (         )
```

8. The hearing person asks to talk with the hearing person in the deaf person’s household.

9. The Relay Service withdraws from the call.

6.2 Media Service Provider

In this scenario, a deaf person wishes to receive the content of a radio program through a text stream transcoded from the program's audio stream.

1. The deaf person attempts to establish a connection to the radio broadcast, with User Agent preferences set to receiving audio stream as text.

2. The User Agent of the deaf person queries the radio station User Agent on whether a text stream is available, other than the audio stream.

3. However, the radio station has no text stream available for a deaf listener, and responds in the negative.

4. As no text stream is available, the deaf person's User Agent requests a voice-to-text transcoding service (e.g., a real-time captioning service) to come into the conversation space.

5. The transcoding service User Agent identifies the audio stream as a radio broadcast. However, the policy of the transcoding service is that it does not accept radio broadcasts because it would overload their resources far too quickly.

6. In this case, the connection fails.

Alternatively, continuing from 2 above:

3. The radio station does provide text with their audio streams.

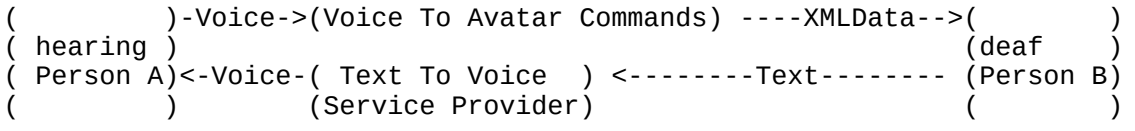
4. The deaf person receives a text stream of the radio program.

Note: To support deaf, hard of hearing and speech-impaired people, service providers are encouraged to provide text with audio streams.

6.3 Sign Language Interface

In this scenario, a deaf person enables a signing avatar (e.g., ViSiCAST[6]) by setting up a User Agent to receive audio streams as XML data that will operate an avatar for sign-language. For outgoing communications, the deaf person types text that is transcoded into an audio stream for the other conversation participant.

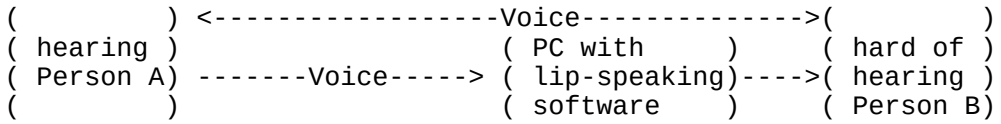
For example:



6.4 Synthetic Lip-speaking Support for Voice Calls

In order to receive voice calls, a hard of hearing person uses lip-speaking avatar software (e.g., Synface[7]) on a PC. The lip-speaking software processes voice (audio) stream data and displays a synthetic animated face that a hard of hearing person may be able to lip-read. During a conversation, the hard of hearing person uses the lip-speaking software as support for understanding the audio stream.

For example:



6.5 Voice Activated Menu Systems

In this scenario, a deaf person wishing to book cinema tickets with a credit card, uses a textphone to place the call. The cinema employs a voice-activated menu system for film titles and showing times.

1. The deaf person places a call to the cinema with a textphone:

   (Textphone) <-----Text ---> (Voice-activated System)

2. The cinema’s voice-activated menu requests an auditory response to continue.

3. A Relay Service is invited into the conversation.

4. The Relay Service transcodes the prompts of the voice-activated menu into text.

5. Text from the voice-activated menu appears on the display of the deaf person’s textphone.

6. The deaf person types a response.

7. The Relay Service receives the text and reads it to the voice-activated system:

```
(           )         (Relay Service   )          (                )
( deaf      ) -Text-> (Provider        ) -Voice-> (Voice-Activated)
( Person A  ) <-Text- (Text To Voice   ) <-Voice- (System         )
```

8. The transaction is finalized with a confirmed booking time.

9. The Relay Service withdraws from the call.

6.6 Conference Call

A conference call is scheduled between five people:

- Person A listens and types text (hearing, no speech)
- Person B recognizes sign language and signs back (deaf, no speech)
- Person C reads text and speaks (deaf or hearing impaired)
- Person D listens and speaks
- Person E recognizes sign language and reads text and signs

A conference call server calls the five people and based on their preferences sets up the different transcoding services required. Assuming English is the base language for the call, the following intermediate transcoding services are invoked:

- A transcoding service (English speech to English text)
- An English text to sign language service
- A sign language to English text service
- An English text to English speech service

Note: In order to translate from English speech to sign language, a chain of intermediate transcoding services was used (transcoding and English text to sign language) because there was no speech-to-sign language available for direct translation. Accordingly, the same applied for the translation from sign language to English speech.

Remarks: - Some services might be shared by users and/or other services.

- Person E uses two parallel streams (SL and English Text). The User Agent might perform time synchronisation when displaying the streams. However, this would require synchronisation information to be present on the streams.

- The session protocols might support optional buffering of media streams, so that users and/or intermediate services could go back to previous content or to invoke a transcoding service for content they just missed.

- Hearing impaired users might still receive audio as well, which they will use to drive some visual indicators so that they can better see where, for instance, the pauses are in the conversation.

7. Some Suggestions for Service Providers and User Agent Manufacturers

This section is included to encourage service providers and user agent manufacturers in developing products and services that can be used by as wide a range of individuals as possible, including deaf, hard of hearing and speech-impaired people.

- Service providers and User Agent manufacturers can offer to a deaf, hard of hearing and speech-impaired person the possibility of being able to prevent their specific abilities and preferences from being made public in any transaction.

- If a User Agent performs auditory signalling, for example a pager, it could also provide another signalling method; visual (e.g., a flashing light) or tactile (e.g., vibration).

- Service providers who allow the user to store specific abilities and preferences or settings (i.e., a user profile) might consider storing these settings in a central repository, accessible no matter what the location of the user and regardless of the User Agent used at that time or location.

- If there are several transcoding services available, the User Agent can be set to select the most economical/highest quality service.

- The service provider can show the cost per minute and any minimum charge of a transcoding service call before a session starts, allowing the user a choice of engaging in the service or not.

- Service providers are encouraged to offer an alternative stream to an audio stream, for example, text or data streams that operate avatars, etc.

- Service providers are encouraged to provide a text alternative to voice-activated menus, e.g., answering and voice mail systems.

- Manufacturers of voice-activated software are encouraged to provide an alternative visual format for software prompts, menus, messages, and status information.

- Manufacturers of mobile phones are encouraged to design equipment that avoids electro-magnetic interference with hearing aids.

- All services for interpreting, transliterating, or facilitating communications for deaf, hard of hearing and speech-impaired people are required to:

- Keep information exchanged during the transaction strictly confidential

- Enable information exchange literally and simply, without deviating and compromising the content

- Facilitate communication without bias, prejudice or opinion

- Match skill-sets to the requirements of the users of the service

- Behave in a professional and appropriate manner

- Be fair in pricing of services

- Strive to improve the skill-sets used for their services.

- Conference call services might consider ways to allow users who employ transcoding services (which usually introduce a delay) to have real-time information sufficient to be able to identify gaps in the conversation so they could inject comments, as well as ways to raise their hand, vote and carry out other activities where timing of their response relative to the real-time conversation is important.

8. Acknowledgements

The authors would like to thank the following individuals for their contributions to this document:

David R. Oran, Cisco
Mark Watson, Nortel Networks
Brian Grover, RNID
Anthony Rabin, RNID
Michael Hammer, Cisco
Henry Sinnreich, Worldcom
Rohan Mahy, Cisco
Julian Branston, Cedalion Hosting Services
Judy Harkins, Gallaudet University, Washington, D.C.
Cary Barbin, Gallaudet University, Washington, D.C.
Gregg Vanderheiden, Trace R&D Center University of Wisconsin-Madison
Gottfried Zimmerman, Trace R&D Center University of Wisconsin-Madison

Security Considerations

This document presents some privacy and security considerations. They are treated in Section 5.6 Confidentiality and Security.

Normative References

[1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[2] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M. and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

Informational References

[3] International Telecommunication Union (ITU), "Operational and interworking requirements for DCEs operating in the text telephone mode". ITU-T Recommendation V.18, November 2000.

[4] Moore, Matthew, et al. "For Hearing People Only: Answers to Some of the Most Commonly Asked Questions About the Deaf Community, Its Culture, and the Deaf Reality". MSM Productions Ltd., 2nd Edition, September 1993.

[5] http://www.typetalk.org.

[6] http://www.visicast.co.uk.

[7] http://www.speech.kth.se/teleface.

Authors' Addresses

Nathan Charlton
Millpark Limited
52 Coborn Road
London E3 2DG
Tel: +44-7050 803628
Fax: +44-7050 803628
EMail: nathan@millpark.com

Mick Gasson
Koru Solutions
30 Howland Way
London SE16 6HN
Tel: +44-20 7237 3488
Fax: +44-20 7237 3488
EMail: michael.gasson@korusolutions.com

Guido Gybels
RNID
19-23 Featherstone Street
London EC1Y 8SL
Tel: +44-20 7296 8000
Textphone: +44-20 7296 8001
Fax: +44-20 7296 8199
EMail: Guido.Gybels@rnid.org.uk

Mike Spanner
RNID
19-23 Featherstone Street
London EC1Y 8SL
Tel: +44-20 7296 8000
Textphone: +44-20 7296 8001
Fax: +44-20 7296 8199
EMail: mike.spanner@rnid.org.uk

Arnoud van Wijk
Ericsson EuroLab Netherlands BV
P.O. Box 8
5120 AA Rijen
The Netherlands
Fax: +31-161-247569
EMail: Arnoud.van.Wijk@eln.ericsson.se

Comments can be sent to the SIPPING mailing list.

Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.