Internet Draft Rosenberg/Schulzrinne/Sinnreich
draft-rosenberg-sip-hearingimpaired-00.txt Columbia U./dynamicsoft/WCOM
July 13, 2000
Expires: January 2001

SIP Enabled Services to Support the Hearing Impaired

Abstract

This document outlines a set of services enabled by the Session Initiation Protocol (SIP), that allow for access to voice services by people who are hearing impaired. SIP has gained much attention as a tool for voice communications on the Internet. Therefore, considerations for universal access of its services are important. This document does not propose any extensions or new capabilities to SIP, but rather a set of services enabled by it.

1 Introduction

The Session Initiation Protocol (SIP) [1] is used to initiate, modify, and terminate interactive sessions between sets of users. Often, these sessions are voice sessions, described by the Session Description Protocol (SDP) [2]. Unfortunately, not everyone is able to participate in voice sessions. In particular, people who are hearing impaired often cannot act as senders or recipients on a voice session. Within the Public Switched Telephone System (PSTN), services have been defined that allow for access to ciruit switched voice services by the hearing impaired. We believe it is important to offer these kinds of services in an IP context. In fact, the flexibility of SIP affords us the ability to improve on these services, and offer more extensive forms of universal service access to the hearding impaired.

This document outlines a few possible services that enable universal access of voice sessions, initiated by SIP, to users who are hearing impaired. These services are generally enabled by baseline SIP [1], or through the use of the caller preferences specification [3]. No additional extensions are proposed here in order to support universal access.

2 Example Services and Call Flows

We provide the following examples services and accompanying call flows:

Redirect to IM: The caller has phone and IM client. The called party has a phone and IM client. The phone call is redirected to IM and both parties use IM to communicate.

One-way speech to text translation service: The caller has only a phone. The called party has a text terminal to receive and a phone to send. A relay service translates in one direction only from speech to text.

One-way speech to sign language translation service: The caller has just a phone. The called party has a video terminal to receive and a phone to send. A relay service translates in one direction only from speech to video, with the video being a sign language representation of the speech.

Two-way speech to text and text to speech with translation service: The caller only has a phone. The called party uses text both ways. A relay service translates in one direction from text to speech and from speech to text in the other direction. A computer can do the text to speech translation.

Hearing impaired calling party calling through relay: The caller has text only. The called party only has a phone. A relay service translates in one direction from text to speech and from speech to text in the other direction. A computer can do the text to speech translation.

Alerts are provided to the phone user that the other party is hearing impaired and if the case, a relay service is automatically inserted.

2.1 Redirect to IM

One advantage of providing voice services through the Internet is the access to other IP services that can be used in conjunction with voice. In support of the hearing impaired, Instant Messaging (IM) is particularly useful. IM allows for instantaneous text messaging between IP connected users. Recent work has specified how IM service can be enabled by SIP [4].

One way to use IM to support the hearing impaired is to redirect a voice call to an IM exchange (provided the caller supports IM). The service works as follows. A voice call is initiated by a PC or other terminal that supports IM. Indication of support for IM is done through the caller preferences specification [3], which allows the caller to indicate characteristics of URLs they are willing to be redirected to. In this case, they would indicate support of the MESSAGE method, used for instant messaging within SIP. Support for other instant messaging protocols, so long as they are described by standardized URL schemes, can also be indicated.

When the call arrives at the user agent of the hearing impaired user, the UA checks for support of instant messaging. If such support is indicated, the UAS sends a 302 (Use IM - Hearing Impaired) redirect, containing a URL to be used for IM. This redirect is forwarded back to the calling party, whose IM tool pops up with an IM filled in with the address of the called party. The two can then participate in a pure IM session.

The service can also be provided by an application server serving the hearing impaired user. The application server, upon receiving the INVITE, would initiate its own INVITE towards the hearing impaired user (without indicating any kind of media session). This has the effect of alerting (through a flashing light or some other means) that an incoming call is taking place. If accepted, the application server can then redirect the initial caller to send an IM to a preconfigured IM address.

Figure 1 contains a call flow for the service assuming it is being provided by the called UA.

```
             |                              |
             |    F1: INVITE                |
             | --------------------------> |
             |                              |
             |                              |
             |   F2: 200 OK                 |
             | <-------------------------- |
             |                              |
             |                              |
             |   F3: ACK                    |
             | --------------------------> |
             |                              |
             |                              |
             |                              |
             |                              |
             |    F4: MESSAGE               |
             | --------------------------> |
             |    F5: 200 OK                |
             | <-------------------------- |
             |                              |
             |    F6: MESSAGE               |
             | --------------------------> |
             |    F7: 200 OK                |
             | <-------------------------- |
             |                              |

          Caller                         Hearing
                                         Impaired
                                         User
```

Figure 1: Redirecting to an IM

Message F1 is:

```
INVITE sip:hiu@example.com SIP/2.0
Via: SIP/2.0/UDP a.example.com
From: sip:caller@example.com
```

```
To: sip:hiu@example.com
Call-ID: 9asdg9a7@1.2.3.4
CSeq: 1 INVITE
Contact: sip:caller@a.example.com
Accept-Contact: *;methods=''MESSAGE,SUBSCRIBE''
Content-Type: application/sdp
Content-Length: XX

<SDP>
```

Message F2 is:

```
SIP/2.0 300 Try IM
Via: SIP/2.0/UDP a.example.com
From: sip:caller@example.com
To: sip:hiu@example.com;tag=9ajsd9aumlaa
Call-ID: 9asdg9a7@1.2.3.4
CSeq: 1 INVITE
Contact: sip:hiu@example.com;method=MESSAGE
```

2.2 One-way Speech-to-text Translation Service

An alternative approach is to use a relay, which is a person who can listen to the calling party, type up the text, and send it to the hearing impaired user either through instant messages or through text over RTP [5].

In one variant on this service, a call is made to a hearing impaired person. If the hearing impaired user wishes to accept the call, they send a 183 (Using a Translator for Hearing Impaired) response to the call.

The provisional response to the caller is used by the client to alert the caller to the fact that the called party is hearing disabled and that a relay service will be part of the call. This is useful to help the caller to tune the speaking style, so as to adjust for such a type of conmmunication.

Then, after sending the 183, using the third party call control mechanisms [6], the called party launches a call to a translator, with that INVITE containing SDP that indicates support for only the RTP payload format for text messages. The response from the

translator (presumably accepting the call), contains SDP where the translator expects to receive audio to be translated to text. When this 200 OK arrives at the hearing impaired user, that SDP is placed into the 200 OK of the call. The result is that the caller will be sending media to the translator, and the hearing impaired user will receive a textual version of it over RTP. However, the hearing impaired user sends audio directly to the caller. Clearly, this service only works for users who are hearing impaired but not speech impaired. When this is the case, it has the advantage of sending the speech directly between the participants in the direction that is possible, reducing latency. Such an asymmetric service is not readily supported within the PSTN.

The call flow for this service is depicted in Figure 2.

Message F1 is:

```
INVITE sip:hiu@example.com SIP/2.0
Via: SIP/2.0/UDP a.example.com
From: sip:caller@example.com
To: sip:hiu@example.com
Call-ID: 9asdg9a7@1.2.3.4
CSeq: 1 INVITE
Contact: sip:caller@a.example.com
Accept-Contact: *;methods=''MESSAGE,SUBSCRIBE''
Content-Type: application/sdp
Content-Length: XX

<SDP 1>
```

message F2 is:

```
SIP/2.0 183 Using Translator for Hearing Impaired... Please Wait
Via: SIP/2.0/UDP a.example.com
From: sip:caller@example.com
To: sip:hiu@example.com;tag=9ajsd9aumlaa
Call-ID: 9asdg9a7@1.2.3.4
CSeq: 1 INVITE
```

message F3 is:

```
INVITE sip:speech2txt@example.com SIP/2.0
Via: SIP/2.0/UDP b.example.com
From: sip:hiu@example.com
To: sip:speech2txt@example.com
Call-ID: 88725392k@4.3.2.1
CSeq: 7 INVITE
Contact: sip:hiu@b.example.com
Content-Type: application/sdp
Content-Length: XX

<SDP 2 with text RTP payload format as only codec>
```

message F4 is:

```
SIP/2.0 200 OK - translating
Via: SIP/2.0/UDP b.example.com
From: sip:hiu@example.com
To: sip:speech2txt@example.com;tag=1238827819
Call-ID: 88725392k@4.3.2.1
CSeq: 7 INVITE
Contact: sip:speech2txt@c.example.com
Content-Type: application/sdp
Content-Length: XX

<SDP 3>
```

message F5 is:

```
SIP/2.0 200 OK
Via: SIP/2.0/UDP a.example.com
From: sip:caller@example.com
To: sip:hiu@example.com;tag=9ajsd9aumlaa
Call-ID: 9asdg9a7@1.2.3.4
CSeq: 1 INVITE
Content-Type: application/sdp
Content-Length: XX

<SDP 3>
```

```
        |                        |                         |
        |   F1: INVITE           |                         |
        | ---------------------> |                         |
        |                        |                         |
        |   F2: 183              |                         |
        | <--------------------- |                         |
        |                        |   F3: INVITE            |
        |                        | ----------------------> |
        |                        |                         |
        |                        |   F4: 200 OK            |
        |                        | <---------------------- |
        |   F5: 200 OK           |                         |
        | <--------------------- |                         |
        |                        |                         |
        |                        |                         |
        |   F6: ACK              |                         |
        | ---------------------> |                         |
        |                        |   F7: ACK               |
        |                        | ----------------------> |
        |                        |                         |
        |                        |                         |
        |                        |                         |
        |                        |                         |
        |    RTP (audio)         |                         |
        | ---------------------------------------------->  |
        | <--------------------- |                         |
        |                        |                         |
        |                        |                         |
        |                        |    RTP (text)           |
        |                        | <---------------------- |
        |                        |                         |
        |                        |                         |
        |                        |                         |
        |                        |                         |
        |                        |                         |
        |                        |                         |

      Caller                  Hearing                   Translator
                              Impaired
                              User
```

Figure 2: One Way Translation Service

message F6 and F7 are standard ACK messages, not shown.

2.3 One-way Speech-to-Sign-Language Translation Service

from a normal phone, makes a call to a hearing impaired user. The hearing impaired user establishes a connection with a translator service that will listen to speech and "convert" it to sign language. The sign language is sent to the hearing impaired used through a video stream.

This service is accomplished identically to the one way speech to text translation service. The call flow is the same as listed in Figure 2. The only difference is that the SDP which indicates text, will instead indicate video. The RTP stream marked as containing text, will instead contain video.

2.4 Two-way speech to text and text to speech with translation service

The service in the previous section can be extended to include one relay for speech to text and another that does text to speech (where the text is typed by the speech impaired user). The text to speech translation can be done by a computer. If people are used to translate in both directions, these translators may be the same person, but they need not be. This has the interesting effect of introducing some form of privacy. With two different translators, neither is privy to the complete conversation, and in all likelihood, would not be able to ascertain what is actually being talked about.

A call flow for this variant on the service is shown in Figure 3.

Messages F1, F2, F3 and F4 are the same as above. F5 is a standard ACK. F6 is:

```
INVITE sip:text2speech@example.com SIP/2.0
Via: SIP/2.0/UDP b.example.com
From: sip:hiu@example.com
To: sip:text2speech@example.com
Call-ID: 87765448902@4.3.2.1
CSeq: 88 INVITE
Contact: sip:hiu@b.example.com
Content-Type: application/sdp
Content-Length: XX

<SDP 1>
```

and F7 looks like:

```
SIP/2.0 200 OK
Via: SIP/2.0/UDP b.example.com
From: sip:hiu@example.com
To: sip:text2speech@example.com;tag=9asdgnzli98a0
Call-ID: 87765448902@4.3.2.1
CSeq: 88 INVITE
Contact: sip:text2speech@d.example.com
Content-Type: application/sdp
Content-Length: XX

<SDP 4 w/ RTP payload type for text>
```

F8 is a standard ACK. F9 looks like F5 from the asymetric version of the service.

Our approach also has the advtange that any application service provider can be used for these translation services. Different providers can be used for each direction, and these providers do not need to be affiliated in any way with the ISP providing IP services for the hearing impaired user. This provides for greater competition, and thus improved service.

This approach also has the advantage of allowing one direction (speech to text), the other direction (text to speech), or both, to be performed by automated systems. For example, text to speech technology is fairly robust, and could be used in one direction, whereas a human operator could be used in the reverse (speech to text) direction, since speech recognition is not that robust. The call flow is completely identical, independently of whether the translation is done by human or machine. A machine would simply answer all calls to a specific address (sip:translator@asp.com), and echo the media (text or speech) back to the caller after conversion (conversion direction would be determined by the media capabilities indicated in the INVITE). In fact, there are other applications for such conversion systems. Providers of them could not only enable services for the hearing impaired, but other applications as well. Examples include voice browsing of the web, email to speech readout over phones, and instant message to voicemail services. In fact, the opposite direction is quite likely - providers that perform these services can reuse their systems, without any work, to also provide services to the hearing impaired.

2.5 Hearing Impaired Calling Party through Relay

In this section, we consider a relay where the calling party is hearing impaired.

```
        |                        |                         |        |
        |   F1: INVITE           |                         |        |
        | ---------------------> |                         |        |
        |                        |                         |        |
        |   F2: 183              |                         |        |
        | <--------------------- |                         |        |
        |                        |   F3: INVITE            |        |
        |                        | ----------------------> |        |
        |                        |                         |        |
        |                        |   F4: 200 OK            |        |
        |                        | <---------------------- |        |
        |                        |                         |        |
        |                        |   F5: ACK               |        |
        |                        | ----------------------> |        |
        |                        |                         |        |
        |                        |   F6: INVITE            |        |
        |                        | -------------------------------> |
        |                        |                         |        |
        |                        |   F7: 200 OK            |        |
        |                        | <------------------------------- |
        |                        |                         |        |
        |                        |   F8: ACK               |        |
        |                        | ------------------------------- >|
        |                        |                         |        |
        |   F9: 200 OK           |                         |        |
        | <--------------------- |                         |        |
        |                        |                         |        |
        |                        |                         |        |
        |   F10 ACK              |                         |        |
        | ---------------------> |                         |        |
        |                        |  RTP (speech)           |        |
        |------------------------------------------------->|        |
        |                        |<------------------------|        |
        |                        |     RTP (text)          |        |
        |                        |                         |        |
        |                        |                         |        |
        |                        |     RTP (text)          |        |
        |                        |--------------------------------->|
        |<-----------------------+----------------------------------|
        |         RTP (Speech)   |                         |        |
        |                        |                         |        |
        |                        |                         |        |

      Caller                   Hearing                   STT       TTS
                               Impaired
```

User

This service works much like the one desribed above, relying on third party call control mechanisms. The caller sends an INVITE with SDP containing no codecs, targeted for the called party. If the called party accepts, the caller launches an INVITE to one or two translation services (depending on whether the caller is just hearing impaired, or both speech and hearing impaired). The INVITE to speech to text translation service contains SDP where the caller would like to receive the text; the response contains SDP that the caller places in the ACK to the called party. This connects the called party with the speech to text translator, with the resultant text being sent to the caller. If text to speech service is also needed, the caller places the SDP it received in the 200 OK from the called party into an INVITE to the translator. The response contains SDP with an address where the caller can send text.

Figure 4 shows a call flow using only speech to text translation services.

3 Security Considerations

Since the services described here rely on a person or machine to translate voice or text, there is an unavoidable trust relationship between the participants in the call and this service. As such, strict privacy of the conversation cannot be provided; the translator service needs to have access to the media stream. However, our approach of separating the text to speech and speech to text translator services affords some amount of privacy, as a single outside entity would not be privy to the entire conversation.

4 Acknowledgements

The authors would like to thank Vint Cerf/WCOM for encouraging this work and also to Teresa Hastings/WCOM. Both contributed to the initial discussions leading to this draft.

5 Authors Addresses

Jonathan Rosenberg
dynamicsoft
72 Eagle Rock Avenue
First Floor
East Hanover, NJ 07936
email: jdrosen@dynamicsoft.com

Henry Sinnreich

```
         |                       |                         |
         |   F1: INVITE no SDP   |                         |
         | --------------------> |                         |
         |                       |                         |
         |   F2: 200 OK  SDP1    |                         |
         | <-------------------- |                         |
         |                       |                         |
         |                       |                         |
         |     F3: INVITE        |                         |
         | ----------------------------------------------> |
         |                       |                         |
         |                       | F4: 200 OK SDP2         |
         | ----------------------------------------------- |
         |                       |                         |
         |     F5: ACK           |                         |
         | ----------------------------------------------> |
         |                       |                         |
         |                       |                         |
         |  F6: ACK SDP2         |                         |
         | --------------------> |                         |
         |                       |                         |
         |                       |                         |
         |   RTP (speech)        |                         |
         |---------------------->|   RTP (speech)          |
         |                       |------------------------>|
         |<------------------------------------------------|
         |   RTP (text)          |                         |
         |                       |                         |
         |                       |                         |
         |                       |                         |
         |                       |                         |
         |                       |                         |
         |                       |                         |

      Hearing                Called                      Speech to
      Impaired               Party                       Text Server
      Caller
```

Figure 4: Hearing Imaired Caller Call Flow

MCI Worldcom
400 International Parkway
Richardson, Texas 75081

email:henry.sinnreich@wcom.com

Henning Schulzrinne
Columbia University
M/S 0401
1214 Amsterdam Ave.
New York, NY 10027-7003
email: schulzrinne@cs.columbia.edu

6 Bibliography

[1] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: session initiation protocol," Request for Comments 2543, Internet Engineering Task Force, Mar. 1999.

[2] M. Handley and V. Jacobson, "SDP: session description protocol," Request for Comments 2327, Internet Engineering Task Force, Apr. 1998.

[3] H. Schulzrinne and J. Rosenberg, "SIP caller preferences and callee capabilities," Internet Draft, Internet Engineering Task Force, Mar. 2000. Work in progress.

[4] J. Rosenberg, R. Sparks, D. Willis, B. Campbell, H. Schulzrinne, J. Lennox, C. Huitema, B. Aboba, and D. Gurle, "SIP extensions for instant messaging," Internet Draft, Internet Engineering Task Force, June 2000. Work in progress.

[5] G. Hellstrom, "RTP payload for text conversation," Request for Comments 2793, Internet Engineering Task Force, May 2000.

[6] J. Rosenberg, H. Schulzrinne, and J. Peterson, "Third party call control in SIP," Internet Draft, Internet Engineering Task Force, Mar. 2000. Work in progress.