

Specification for Speech Links Protocol from SpeechWorks
Version 1.0

July 15, 2000

Copyright © SpeechWorks International, Inc. All Rights Reserved.

Table of Contents

Table of Contents	1
Introduction.....	3
Overview of speech link Functionality	3
Definitions	3
Caller.....	3
Initiator	3
Recipient	3
Setup	3
PSTN.....	3
Termination.....	3
Transfer.....	3
Transfer Target	3
Transferee	4
Transferor	4
Overview of speech link Operation	4
speech links Protocol	5
speech link Addresses	5
Call Setup - Simple	5
Limitations	6
Call Setup – With Handshake	6
Limitations	7
Call Termination	7
Call Transfer	8
Limitations	9
Call Transfer with Return	9
Message Formats	9
Message Body.....	10
speech links XML Tags	10
Sample Call Setup Session.....	11
Summary of Recipient Requirements	12
Call Setup.....	13
Call Termination	15
Transferring a Call	16
Summary of Initiator Requirements.....	17
Call Setup.....	17
Call Termination	18
Call Transfer	18

Introduction

Overview of speech link Functionality

The speech link Protocol is an application-layer control protocol for transferring callers between cooperating speech applications, pre-existing interactive voice response (IVR) applications based on dual-tone multi-frequency (DTMF, or TouchTone™) input, and human agents both in call centers and in other locations. The voice path of the call can be carried via the Public Switched Telephone Network (PSTN), a Voice over IP (VoIP) network, or a combination of the two. The control of the call is handled via an IP network, usually the internet. A connection managed via the speech link Protocol is referred to as a “speech link”.

A speech link uses the Session Initiation Protocol (SIP - IETF RFC2543) to implement call control. Information about the call is exchanged between cooperating applications using MIME-encoded user data in the SIP message body, and within the message body information is specified via XML and follows the XML syntax. The speech link Protocol uses specific XML tags itself, and silently ignores other tags, allowing cooperating applications to exchange information through the same mechanism.

Definitions

Caller

Typically a person communicating using a telephone or similar device to a speech application.

Initiator

The party initiating a speech link. The Initiator must already be in communication with the caller and know the speech link Address of the Recipient. The Initiator must be able to transfer calls on behalf of a Recipient.

Recipient

The party receiving a call from an Initiator. Recipients can become Initiators in their own right, bridging the call between the party that initiated the call with them, and the new Recipient. Recipients can also request that the Initiator transfer the call on their behalf.

Setup

The process of an Initiator and a Recipient exchanging SIP messages and PSTN signaling to establish a voice path between the Caller and the Recipient.

PSTN

The Public Switched Telephone Network.

Termination

The process of concluding a call between an Initiator and a Recipient.

Transfer

The process of transferring a Caller from one Recipient to another one, via the Initiator.

Transfer Target

The party to whom a Caller is being transferred.

Transferee

The party transferring a Caller on behalf of a previous Recipient (the Transferor) to a new one (the Transfer Target)

Transferor

The Recipient requesting that the Initiator act as a Transferee in transferring a Caller to a new Recipient (the Transfer Target).

Overview of speech link Operation

A speech link requires that a Caller (usually a person) be connected to or through a voice network connection that is managed by an intelligent application. This application can act as the Initiator of a speech link to connect the caller to another application or person.

To establish a speech link, the Initiator must know the “speech link Address” of the intended Recipient, identified by a speech link URI, which may be the same as the Recipient’s SIP URI.

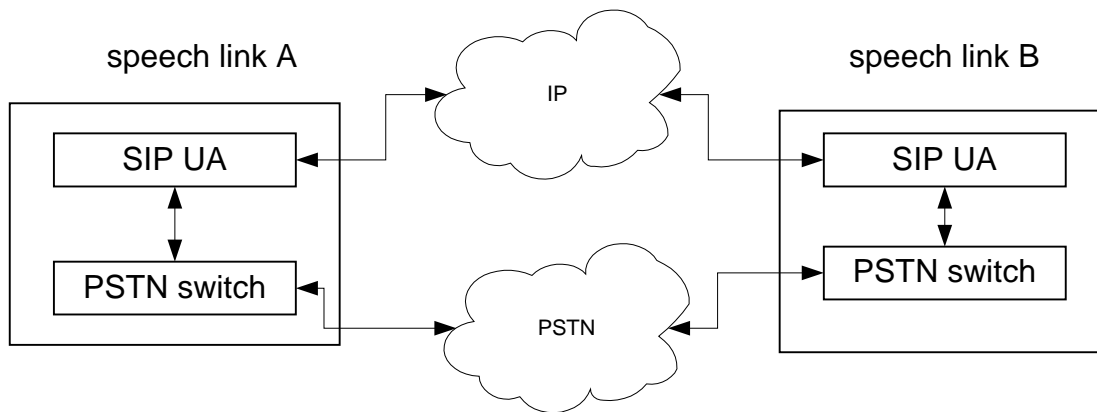
By exchanging IP messages in the speech links Protocol with each other, the Initiator and the Recipient establish a voice connection between the Caller and the Recipient. In the case of a PSTN connection, this is usually “bridged” (or “hairpinned”, or “tromboned”) between the Initiator and both the Caller and Recipient.

Recipients can ask the Initiator to transfer the Caller to another Recipient via a speech link Address, and optionally can request the Initiator to return the Caller if the second Recipient terminates the call.

speech links Protocol

While SIP is responsible for the session management, it is not responsible for the actual transmission of the voice data. The initial implementation of speech links uses the PSTN (Public Switched Telephone Network) for carrying the voice data.

The following diagram shows the connections between two speech link sites and their components.



The calling speech link site (the “Initiator”) uses SIP messages to reach the called speech link site (the “Recipient”) over IP. In the process of establishing the SIP connection, the called speech link sends a PSTN phone number to the calling speech link. To do this, the called speech link sends a SIP provisional response with the phone number to dial included in the body of the message. Once the calling speech link has received the phone number and has established the connection with the called speech link through PSTN, the session is established.

This is a major difference with regular SIP-to-SIP dialing. In a regular SIP-to-SIP call, the media path is established through IP when the call is answered (200 OK response sent by the callee). Also, with regular SIP-to-SIP dialing, there usually is an IP data path created between the calling User Agent (UA) and the called UA. In speech link framework using the PSTN, there is no IP data path created when the connection is established.

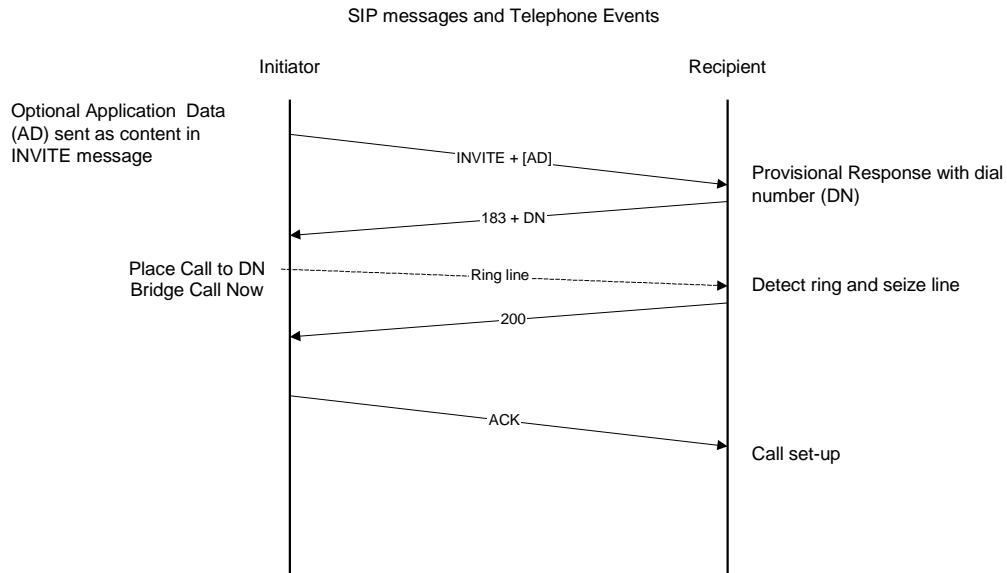
Speech link Addresses

To establish a speech link, the Initiator must know the “speech link Address” of the intended Recipient, identified by a speech link URI, which may be the same as the Recipient’s SIP URI. Typically, this is of the form sip:[user@host](#), or sip:[speech link@host](#), where the first part is either an individual or group name or telephone number, and the second part is either a domain name or a numeric network address.

Call Setup - Simple

The simplest form of call setup can be used to establish a speech link to any directly dialed telephone number on the PSTN that has been established with a specific speech link address.

A normal call follows the sequence of messages shown in the state transition diagram below.



1. The Initiator sends a SIP INVITE, optionally including application data in the message body of the SIP message.
2. The Recipient returns a provisional SIP response (183 Session Progress). The Recipient must return an E.164 dial number (DN) in the body of the response.
3. When the Initiator receives a provisional return-code that contains the DN, it dials the number specified in the body of the message and connects the voice path of this call leg to the caller.
4. The Recipient detects ring and seizes the line. Before or after seizing the line it must send the SIP 200 OK response, indicating the success result for SIP.
5. The Initiator waits for both a return-code 200 and detection of the line being answered using call progress detection. After the Initiator detects both events, it immediately sends an ACK to acknowledge the call set-up. When the Recipient receives the ACK the call is fully set-up.

Limitations

This simple protocol does not ensure that the user data in the message body is associated with the PSTN call that the Recipient answers. It is possible that another non-speech link caller might originate a call to the DN that would ring the line between steps 2 and 4 of the sequence above.

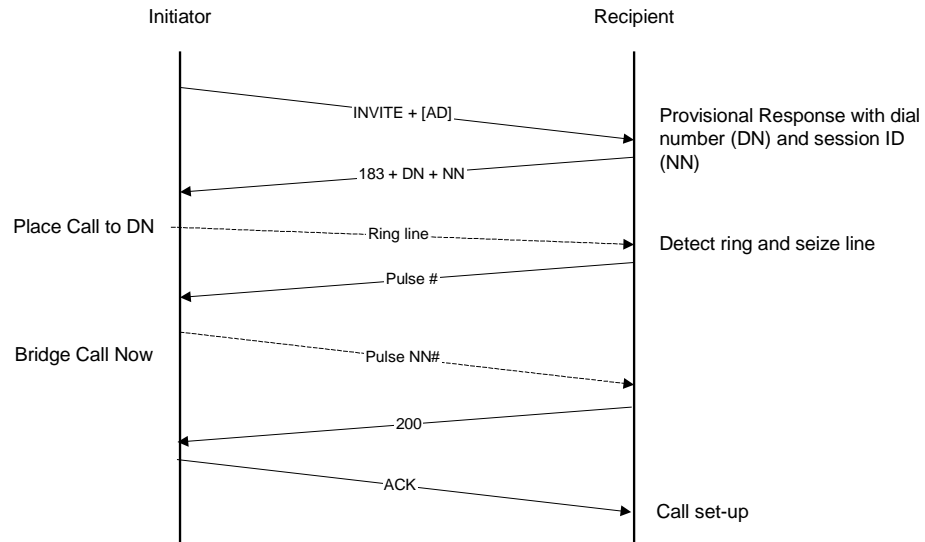
The caller will also hear all the progress signaling of the call leg being set up – busy signals, ringing, Special Information Tones, etc.

In some network environments, call progress detection is dependent on the Initiator recognizing something that sounds like speech or an answering machine on the Recipient side. Should it fail to detect this, it may not be able to determine when, or even if the call has definitely been completed successfully. This may affect the accuracy of billing and reporting, and may even leave a caller listening to “dead air” indefinitely.

Call Setup – With Handshake

This more complex form of call setup eliminates the limitations of the simple method described above, but requires coordination by the Recipient of the SIP and PSTN connections. It also supports managing a group of incoming lines with a single dialed number. A normal call using this method looks like this:

SIP messages and Telephone Events



1. The Initiator sends a SIP INVITE, optionally including application data in the message body of the SIP message (same as above)
2. The Recipient returns a provisional SIP response (183 Session Progress). The Recipient must return an E.164 dial number (DN) in the body of the response, and also supplies a sequence number (NN) which must be temporally unique for the Recipient speech link Address. The presence or absence of this tag distinguishes the type of call setup the Recipient supports.
3. The Initiator dials the DN specified in the body of the message, but does NOT connect the voice path of this call leg to the caller yet.
4. The Recipient detects ring and seizes the line, and sends a DTMF # on the line.
5. The Initiator waits for a # sign to detect call progress. On receipt of the # sign it sends NN# down the telephone line.
6. After receiving a valid NN#, the Recipient must send the SIP 200 OK response, indicating the success result for SIP.
7. When it receives the 200 OK, the Initiator connects the caller voice path and immediately sends an ACK to acknowledge the call set-up.

Initiators must support both the simple Call Setup, and the Handshake version.

Limitations

This protocol is only suitable for automated systems at both the Initiator and Recipient. Should a human caller reach the designated DN directly, they will hear only a DTMF #. It would be possible for the Recipient to combine this signal with a voice prompt message, and perhaps transfer the call if no NN# is received within a specific time limit.

Call Termination

During an active speech link call, if the caller terminates the connection (hangs up), the Initiator must disconnect the PSTN link to the Recipient, and send an SIP BYE request. The voice path disconnect and the BYE request may occur in either order. The Recipient must respond with a 200 OK message, and of course terminate the PSTN connection on its side as well. If the Recipient wishes to return application

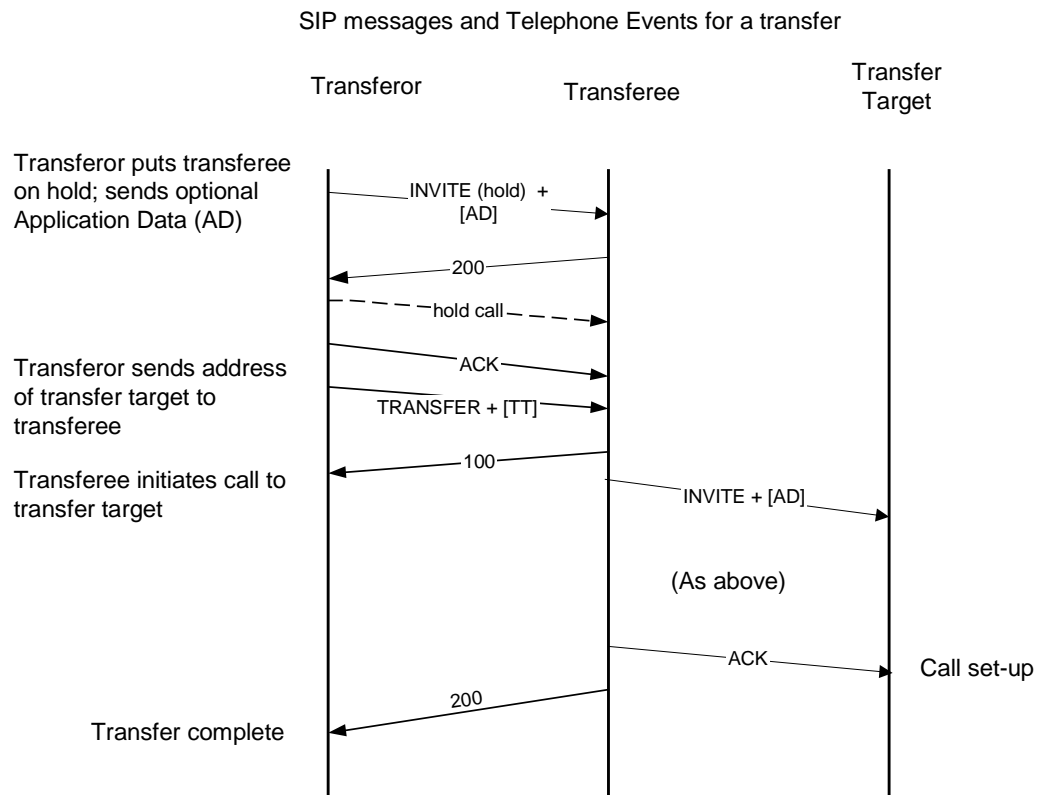
information, it must issue a BYE of its own prior to sending the 200 OK response; the Initiator must in turn respond to this BYE with a 200 OK.

The Recipient can also send a BYE request to the Initiator at any time during an active speech link, and drop its PSTN connection. The BYE and the PSTN disconnect can occur in either order. The Initiator must respond with a 200 OK, but need not send its own BYE.

Call Transfer

After a call has been set up and is active, a speech link Recipient (the Transferor) can request the Initiator (the Transferee) to transfer the call to another speech link Recipient (the Transfer Target). This is an alternate way of terminating the call leg to the Transferor. It will often be more cost-effective than the Recipient setting up a second call as an Initiator and bridging the two call legs.

Transferring a call from one speech link enabled site to another speech link enabled site is accomplished



using the REFER, INVITE and ACK messages, as defined in the July 2000 issue of the SIP Call Control Transfer Internet Draft. The steps in a successful transfer are shown below:

1. The Transferor starts the transfer by putting the Transferee on hold, it does so by sending an SIP INVITE request with a modified session description. In keeping with the SIP convention (B.6 of RFC2453), the session description is the same as in the original invitation, but the DN is set to 0. It may optionally include application data that will be forwarded on to the Transfer Target.
2. The Transferee responds with the SIP 200-OK Response and suspends the voice path to the Transferor.
3. The Transferor sends an ACK to complete the INVITE sequence, and then sends a REFER request specifying the speech link Address of the Transfer Target.

4. The Transferee sends an SIP 100-Trying Response to the Transferor, and sends an INVITE to the Transfer Target to begin a Call Setup. The INVITE must include all call data provided by the Transferor in Step 1, and optionally additional data provided by the Transferee. The rest of the Call Setup sequence proceeds as described above.
5. When the Call Setup is complete, the Transferee sends a 200-OK Response to the Transferor, and initiates a Call Termination sequence with a BYE Request. The rest of this sequence continues as described above.

Limitations

Only a speech link Recipient currently active on a call can initiate a Transfer.

A speech link does not support the Consultation Hold or Protected Transfer Target methods of using the SIP Transfer Request.

Call Transfer with Return

The Transferor may optionally request that the call be returned when the Transfer Target would otherwise terminate the call. To do so, it simply specifies a speech link Address to return the call to as part of the INVITE (hold) Request that initiates the transfer. [Although it might be clearer to specify this option as part of the REFER Request, the current RFC does not support a message body as part of this message.]

This speech link “Return Address” must be passed to the Transfer Target by the Transferee to inform it that a transfer has been requested; the rest of the protocol for setting up the transfer is the same as the simple case above.

The Transferee is responsible for initiating a Call Setup with the Transferor when the call with the Transfer Target is terminated, even if the PSTN connection to the caller has also been disconnected. The SIP Call-ID remains the same as in the initial Call Setup, and additional or changed application data may be sent on the Invite, including all application data provided by the Transfer Target as part of its BYE Request.

In the case of reporting a disconnect, the Transferor sets the CO (connected) tag in the message body to “FALSE”. If it is absent, the value “TRUE” is assumed.

Should the returning Call Setup fail, the Transferee may continue the call with the caller, or may disconnect the caller.

Message Formats

In SIP, data that is standard to SIP, such as the Call-ID, are sent in the message header. speech links specific data, such as the DN or the NN, and application data are passed in the message body.

Message Body

SIP uses MIME to describe the message body. In speech links, the message body is communicated using XML. So, following SIP, speech links messages always define the message body with the following header fields:

```
Content-Type: text/xml
Content-Length: <length>
```

An empty line (CR LF) indicates the end of the header fields, which is followed by <length> bytes of message body.

speech links XML Tags

The top level tag for speech link information is <session>. This tag can contain the following embedded tags:

ANI	Optional E.164 format number identifying caller's originating telephone number	Initiator	INVITE Request
DN	E.164 format telephone number to dial to establish voice path	Recipient	183-Session Progress Response
NN	Optional sequence number. Should be as few digits as possible to minimize call setup time, but allow enough values to avoid having multiple calls being set up with the same sequence number	Recipient	183-Session Progress Response
RT	Optional speech link Address to which the call is to be returned following termination by the Transfer Target	Transferor	INVITE (hold) Request
CO	Indicates whether caller is connected or not on a return from a transfer; value is TRUE (assumed if missing), or FALSE.	Transferee	INVITE Request (back to initial Transferor)
SC	Optional SpeechCookie tag reserved for use by the Speech Media Alliance	All	INVITE, BYE Requests

A speech link silently ignores other tags in the message body. This allows applications built on top of speech links to communicate information through the same mechanism.

A typical INVITE message body would look like:

```
<?xml version="1.0" standalone="yes"?>
<session>
<ANI>16174284444</ANI>
</session>
```

A typical message body of a Response to an INVITE would look like:

```
<?xml version="1.0" standalone="yes"?>
<session>
<DN>12125156368</DN>
<NN>12</NN>
</session>
```

Sample Call Setup Session

The following show a simple example of complete messages used to set up a call.

1. Initiator invites Recipient:

```
INVITE sip:UserB@there.com SIP/2.0
FROM: User A <sip:UserA@here.com>
To: User B <sip:UserB@there.com>
Call-ID: 12345600@here.com
Cseq: 1 INVITE
Contact: User A UserA@here.com
Content-Type: text/xml
Content-Length: 114
```

```
<?xml version="1.0" standalone="yes"?>
<session>
<ANI>15143234567</ANI>
</session>
```

2. Recipient sends a provisional response with DN and NN:

```
SIP/2.0 183 Session In Progress
From: User A <sip:UserA@here.com>
To: User B <sip:UserB@there.com>
Call-Id: 12345600@here.com
Cseq: 1 INVITE
Content-type: text/xml
Content-Length: 94
```

```
<?xml version="1.0" standalone="yes"?>
<session>
<NN>123</NN>
<DN>16173436781</DN>
</session>
```

3. Initiator dials the provided DN
4. Recipient answers phone call and pulses a DTMF #:
5. Initiator pulses the NN#
6. Recipient sends a final response:

```
SIP/2.0 200 OK
From: User A <sip:UserA@here.com>
To: User B <sip:UserB@there.com>
Call-ID: 12345600@here.com
Cseq      : 1 INVITE
Contact: User B <sip:UserB@there.com>
Content-Length: 0
```

7. Initiator sends the ACK request.

```
ACK sip:UserB@there.com
From: User A <sip:UserA@here.com>
```

To: User B <sip:UserB@there.com>
Call-ID: 12345600@here.com
Cseq: 1 ACK
Content-Length: 0

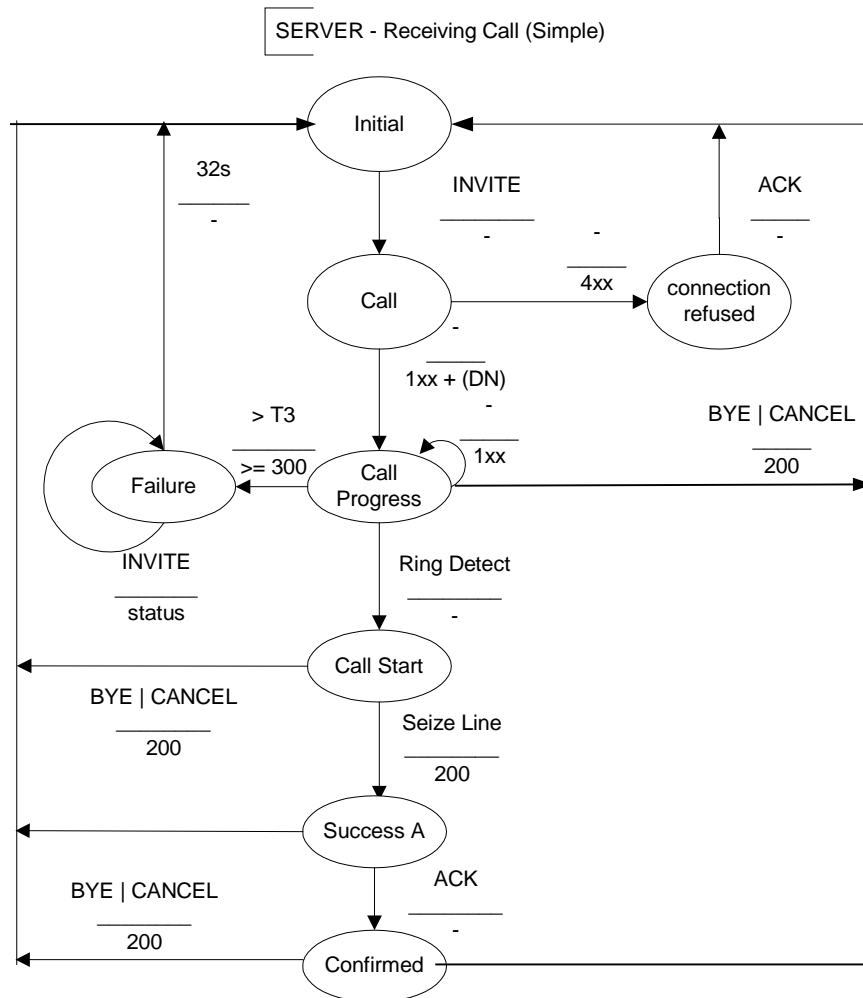
Summary of Recipient Requirements

The speech link protocol is designed to keep the implementation requirements for a Recipient as simple as possible to encourage widespread adoption. This section summarizes these minimum requirements, which should be possible with a small application using a minimally compliant SIP server (A.2 of RFC 2543).

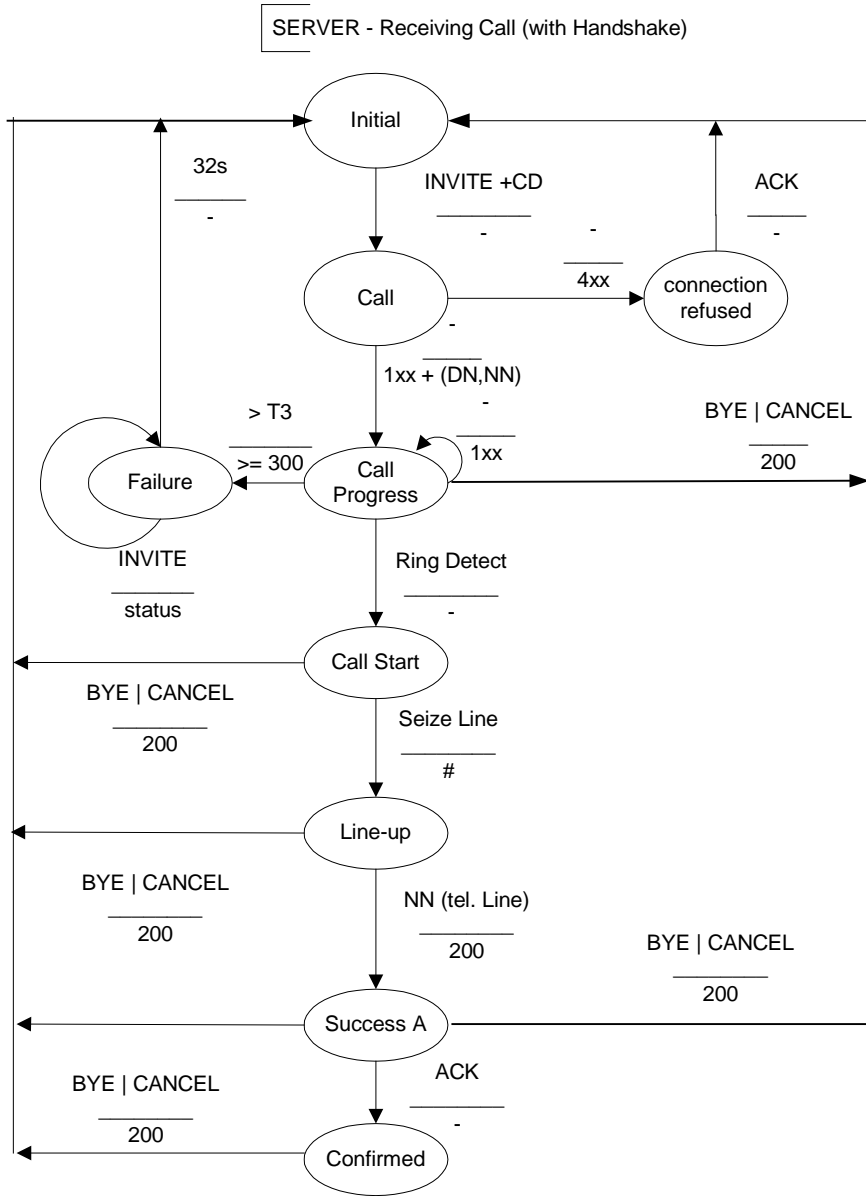
In the following diagrams, each state transition is labeled by an event that causes the transition (above the horizontal line), and the message that is transmitted in response (below the line).

Call Setup

The Recipient is responsible for aborting the call setup if it does not receive an incoming call within timer T3 of the INVITE Request.

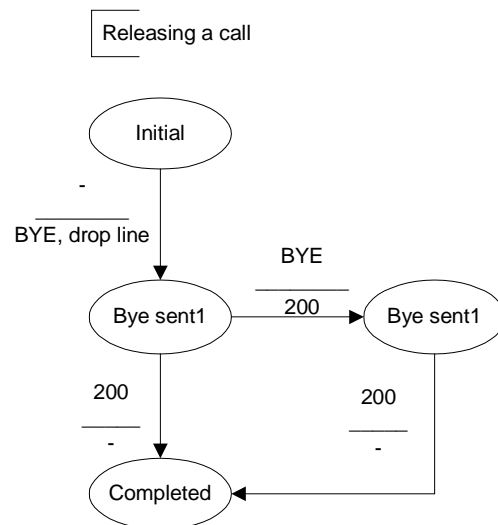
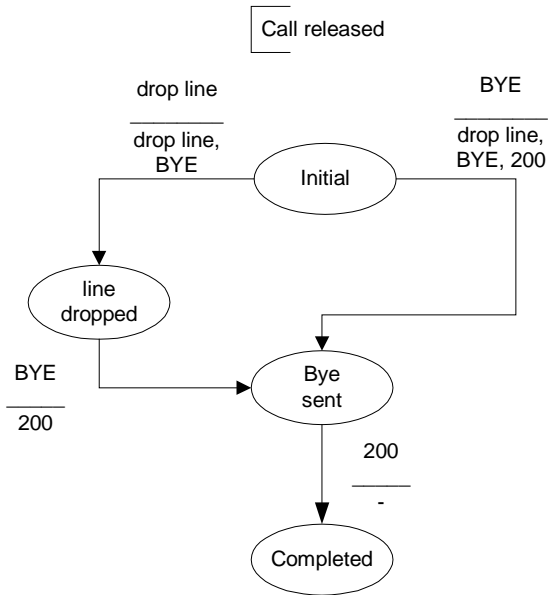


In the following case where the Recipient chooses to use the sequence number, it must also abort the set up if the sequence number is not provided in-band via DTMF, or if it does not match the one provided in the INVITE Request.



Call Termination

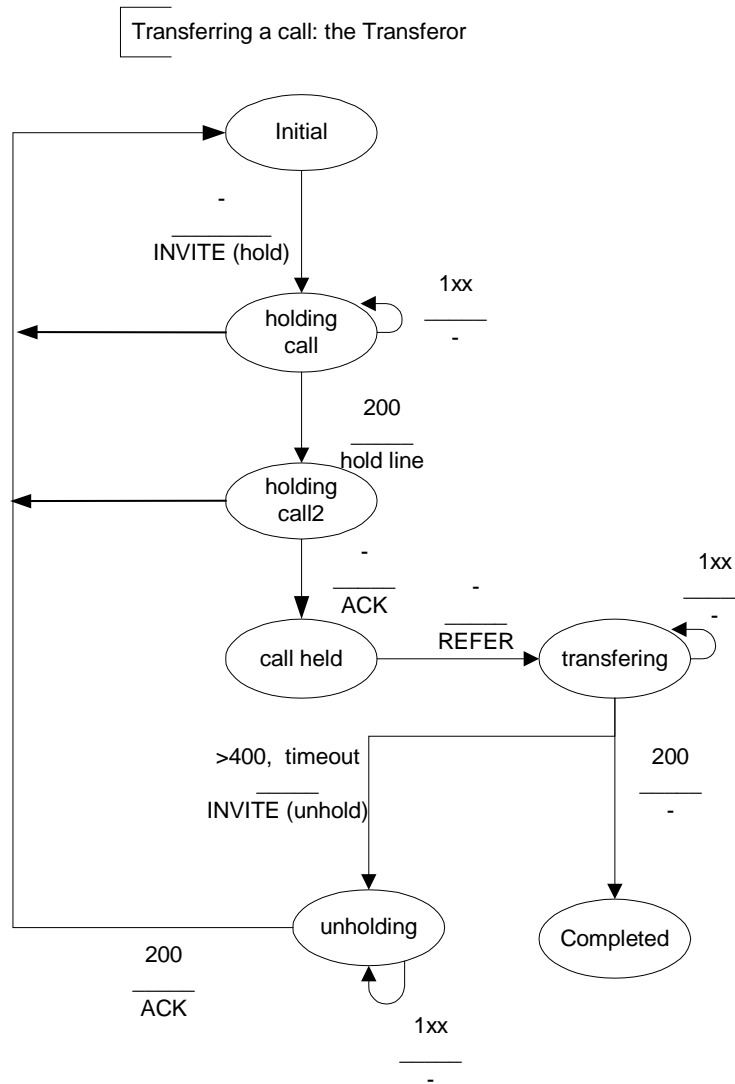
Call Termination is straightforward. The Recipient may optionally supply application data with the BYE Request; if it chooses not to do so, it need only respond to a BYE request from the Initiator. For a simple implementation, always generate a BYE in response to a detected line drop, or in response to a BYE from the Initiator. In either case, the Recipient needs to clean up a call if a 200-OK Response is not received from the Initiator in a reasonable time.



Transferring a Call

Receiving a call transfer as the Transfer Target is the same as being the Recipient of a Call Setup. By examining the RTN tag in the message body, the Recipient can determine if a return was requested, but need not do anything with this information.

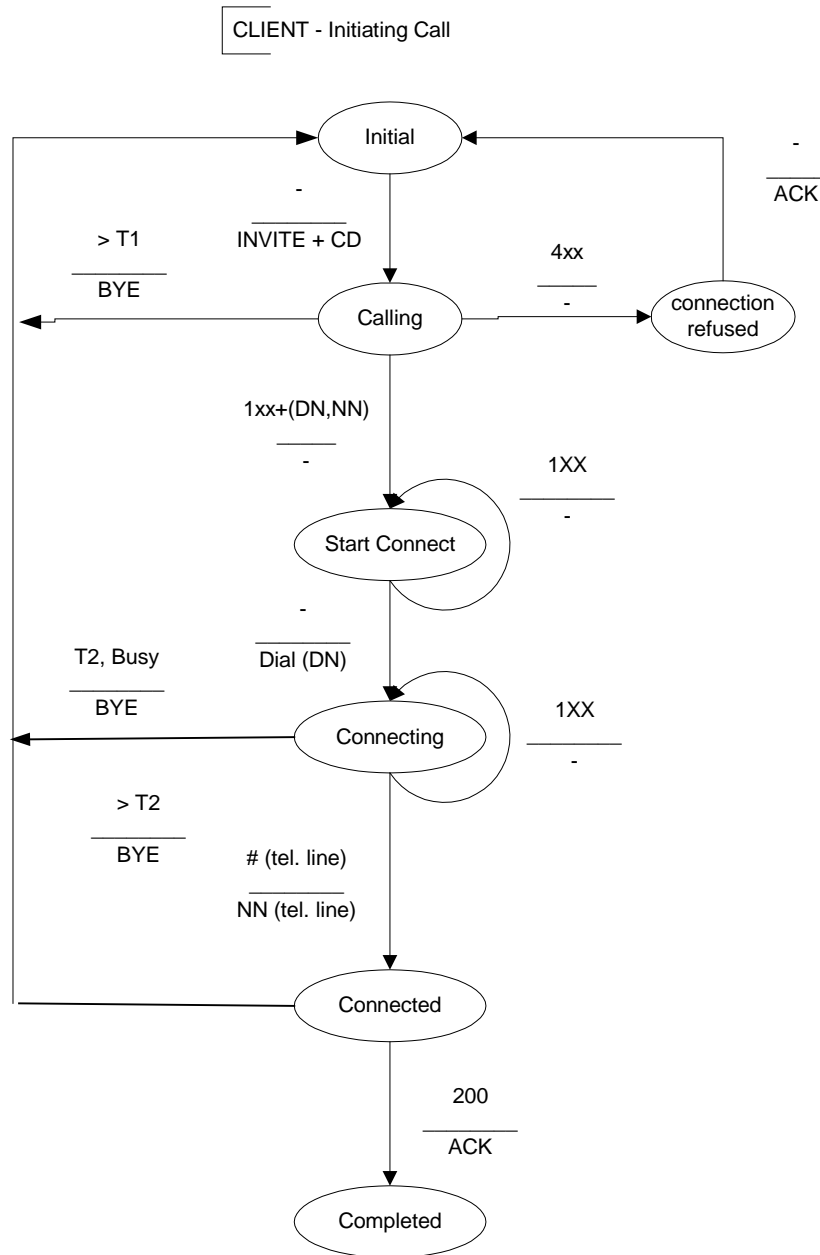
Initiating a transfer as the Transferor is an optional feature for a Recipient. If implemented with the Return capability, the Recipient must maintain a list of Call-Ids for which returns are expected, and match these when invitations are received. The Recipient is also responsible for “timing these out” after some extended period of time when it concludes that the required INVITE has been lost by the Transferee. It must also detect the CO tag in the return INVITE, and if set to FALSE immediately refuse the connection.



Summary of Initiator Requirements

Call Setup

The Initiator must support call progress detection, implementing at a minimum a timer T2 that expires after it has placed a call but not yet received indication of an answer from the Recipient. In addition, it must time out its initial INVITE after T1. In the diagram below, if the “NN” tag is not present in the message body, the “Connecting” state proceeds directly to “Completed” upon receipt of the 200-OK Response.



Call Termination

Call termination requirements for the Initiator are the same as for the Recipient.

Call Transfer

The Initiator has significant additional responsibilities as the Transferee in support Call Transfer. In addition to supporting the protocol described below, the Transferee must also maintain a return stack of at least 32 nested return locations for Transferors who request returns. In case of a normal termination by a Transfer Target, it must set up a call to the top of this return stack. In the case of a disconnect by the caller, it must loop through all return locations, and execute a Call Setup with the value of the CO tag set to FALSE.

