Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 17, 2018

R. Gellens
Core Technology Consulting

October 14, 2017

Negotiating Human Language in Real-Time Communications
draft-ietf-slim-negotiating-human-language-16

Abstract

Users have various human (natural) language needs, abilities, and preferences regarding spoken, written, and signed languages. This document adds new SDP media-level attributes so that when establishing interactive communication sessions ("calls"), it is possible to negotiate (communicate and match) the caller's language and media needs with the capabilities of the called party. This is especially important with emergency calls, where a call can be handled by a call taker capable of communicating with the user, or a translator or relay operator can be bridged into the call during setup, but this applies to non-emergency calls as well (as an example, when calling a company call center).

This document describes the need and a solution using new SDP media attributes.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 17, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 3
1.1. Applicability 4
2. Terminology 4
3. Desired Semantics 5
4. The existing 'lang' attribute 5
5. Solution 5
5.1. Rationale 5
5.2. The 'hlang-send' and 'hlang-recv' attributes 6
5.3. No Language in Common 7
5.4. Undefined Combinations 7
5.5. Examples 7
6. IANA Considerations 10
6.1. att-field Table in SDP Parameters 10
6.2. Warn-Codes Sub-Registry of SIP Parameters 11
7. Security Considerations 11
8. Privacy Considerations 11
9. Changes from Previous Versions 11
9.1. Changes from draft-ietf-slim-...-04 to draft-ietf-slim-...-06 11
9.2. Changes from draft-ietf-slim-...-02 to draft-ietf-slim-...-03 12
9.3. Changes from draft-ietf-slim-...-01 to draft-ietf-slim-...-02 12
9.4. Changes from draft-ietf-slim-...-00 to draft-ietf-slim-...-01 12
9.5. Changes from draft-gellens-slim-...-03 to draft-ietf-slim-...-00 12
9.6. Changes from draft-gellens-slim-...-02 to draft-gellens-slim-...-03 12
9.7. Changes from draft-gellens-slim-...-01 to draft-gellens-slim-...-02 13
9.8. Changes from draft-gellens-slim-...-00 to draft-gellens-slim-...-01 13
9.9. Changes from draft-gellens-mmusic-...-02 to draft-gellens-slim-...-00 13
9.10. Changes from draft-gellens-mmusic-...-01 to -02 13

9.11. Changes from draft-gellens-mmusic-...-00 to -01 13
9.12. Changes from draft-gellens-...-02 to draft-gellens-mmusic-...-00 14
9.13. Changes from draft-gellens-...-01 to -02 14
9.14. Changes from draft-gellens-...-00 to -01 15
10. Contributors 15
11. Acknowledgments 15
12. References 15
12.1. Normative References 15
12.2. Informational References 16
Author's Address 16

1. Introduction

A mutually comprehensible language is helpful for human communication. This document addresses the negotiation of human (natural) language and media modality (spoken, signed, written) in real-time communications. A companion document [RFC8255] addresses language selection in email.

Unless the caller and callee know each other or there is contextual or out-of- band information from which the language(s) and media modalities can be determined, there is a need for spoken, signed, or written languages to be negotiated based on the caller's needs and the callee's capabilities. This need applies to both emergency and non-emergency calls. For example, it is helpful for a caller to a company call center or a Public Safety Answering Point (PSAP) to be able to indicate preferred signed, written, and/or spoken languages, and for the callee to be able to indicate its capabilities in this area, allowing the call to proceed using the language(s) and media forms supported by both.

For various reasons, including the ability to establish multiple streams using different media (e.g., voice, text, video), it makes sense to use a per-stream negotiation mechanism known as the Session Description Protocol (SDP). Utilizing SDP enables the solution described in this document to be applied to all interactive communications negotiated using SDP, in emergency as well as non-emergency scenarios.

By treating language as another SDP attribute that is negotiated along with other aspects of a media stream, it becomes possible to accommodate a range of users' needs and called party facilities. For example, some users may be able to speak several languages, but have a preference. Some called parties may support some of those languages internally but require the use of a translation service for others, or may have a limited number of call takers able to use certain languages. Another example would be a user who is able to

speak but is deaf or hard-of-hearing and and desires a voice stream to send spoken language plus a text stream to receive written language. Making language a media attribute allows the standard session negotiation mechanism to handle this by providing the information and mechanism for the endpoints to make appropriate decisions.

The term "negotiation" is used here rather than "indication" because human language (spoken/written/signed) can be negotiated in the same manner as media (audio/text/video) and codecs. For example, if we think of a user calling an airline reservation center, the user may have a set of languages he or she speaks, with perhaps preferences for one or a few, while the airline reservation center will support a fixed set of languages. Negotiation should select the user's most preferred language that is supported by the call center. Both sides should be aware of which language was negotiated. This is conceptually similar to the way other aspects of each media stream are negotiated using SDP (e.g., media type and codecs).

Since this is a protocol mechanism, the user equipment (UE client) needs to know the user's preferred languages; while this document does not address how clients determine this, reasonable techniques could include a configuration mechanism with a default of the language of the user interface; in some cases, a UE could tie language and media preferences, such as a preference for a video stream using a signed language and/or a text or audio stream using a written/spoken language.

1.1. Applicability

Within this document, it is assumed that the negotiating endpoints have already been determined, so that a per-stream negotiation based on the Session Description Protocol (SDP) can proceed.

When setting up interactive communications sessions it is necessary to route signaling messages to the appropriate endpoint(s). This document does not address the problem of language-based routing.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Desired Semantics

The desired solution is a media attribute (preferably per direction) that may be used within an offer to indicate the preferred language(s) of each (direction of a) media stream, and within an answer to indicate the accepted language. The semantics of including multiple languages for a media stream within an offer is that the languages are listed in order of preference.

(Negotiating multiple simultaneous languages within a media stream is out of scope of this document.)

4. The existing 'lang' attribute

RFC 4566 [RFC4566] specifies an attribute 'lang' which appears similar to what is needed here, but is not sufficiently specific or flexible for the needs of this document. In addition, 'lang' is not mentioned in [RFC3264] and there are no known implementations in SIP. Further, it is useful to be able to specify language per direction (sending and receiving). This document therefore defines two new attributes.

5. Solution

An SDP attribute (per direction) seems the natural choice to negotiate human (natural) language of an interactive media stream, using the language tags of BCP 47 [RFC5646].

5.1. Rationale

The decision to base the proposal at the media negotiation level, and specifically to use SDP, came after significant debate and discussion. From an engineering standpoint, it is possible to meet the objectives using a variety of mechanisms, but none are perfect. None of the proposed alternatives was clearly better technically in enough ways to win over proponents of the others, and none were clearly so bad technically as to be easily rejected. As is often the case in engineering, choosing the solution is a matter of balancing trade-offs, and ultimately more a matter of taste than technical merit. The two main proposals were to use SDP and SIP. SDP has the advantage that the language is negotiated with the media to which it applies, while SIP has the issue that the languages expressed may not match the SDP media negotiated (for example, a session could negotiate a signed language at the SIP level but fail to negotiate a video media stream at the SDP layer).

The mechanism described here for SDP can be adapted to media negotiation protocols other than SDP.

5.2. The 'hlang-send' and 'hlang-recv' attributes

This document defines two media-level attributes starting with 'hlang' (short for "human interactive language") to negotiate which human language is selected for use in each interactive media stream. There are two attributes, one ending in "-send" and the other in "-recv", registered in Section 6. Each can appear in offers and answers for media streams.

In an offer, the 'hlang-send' value is a list of one or more language(s) the offerer is willing to use when sending using the media, and the 'hlang-recv' value is a list of one or more language(s) the offerer is willing to use when receiving using the media. The list of languages is in preference order (first is most preferred). When a media is intended for interactive communication using a language in one direction only (such as a user sending using text and receiving using audio), either hlang-send or hlang-recv MAY be omitted. When a media is not primarily intended for language (for example, a video or audio stream intended for background only) both SHOULD be omitted. Otherwise, both SHOULD have the same value. The two SHOULD NOT be set to languages which are difficult to match together (e.g., specifying a desire to send audio in Hungarian and receive audio in Portuguese will make it difficult to successfully complete the call).

In an answer, 'hlang-send' is the language the answerer will send if using the media for language (which in most cases is one of the languages in the offer's 'hlang-recv'), and 'hlang-recv' is the language the answerer expects to receive if using the media for language (which in most cases is one of the languages in the offer's 'hlang-send').

Each value MUST be a list of one or more language tags per BCP 47 [RFC5646], separated by white space. BCP 47 describes mechanisms for matching language tags. Note that [RFC5646] Section 4.1 advises to "tag content wisely" and not include unnecessary subtags.

When placing an emergency call, and in any other case where the language cannot be inferred from context, in an offer each media stream primarily intended for human language communication SHOULD specify both (or for asymmetrical language use, one of) the 'hlang-send' and 'hlang-recv' attributes.

Clients acting on behalf of end users are expected to set one or both 'hlang-send' and 'hlang-recv' attributes on each media stream primarily intended for human communication in an offer when placing an outgoing session, and either ignore or take into consideration the attributes when receiving incoming calls, based on local

configuration and capabilities. Systems acting on behalf of call centers and PSAPs are expected to take into account the attributes when processing inbound calls.

Note that media and language negotiation might result in more media streams being accepted than are needed by the users (e.g., if more preferred and less preferred combinations of media and language are all accepted). This is not a problem.

5.3. No Language in Common

A consideration with the ability to negotiate language is if the call proceeds or fails if the callee does not support any of the languages requested by the caller. This document does not mandate either behavior.

If the call is rejected due to lack of any languages in common, it is suggested to use SIP response code 488 (Not Acceptable Here) or 606 (Not Acceptable) [RFC3261] and include a Warning header field [RFC3261] in the SIP response. The Warning header field contains a warning code of [TBD: IANA VALUE, e.g., 308] and a warning text indicating that there are no mutually-supported languages; the text SHOULD also contain the supported languages and media.

Example:

```
Warning: [TBD: IANA VALUE, e.g., 308] proxy.example.com
   "Incompatible language specification: Requested languages not
   supported. Supported languages are: es, en; supported media
   are: audio, text."
```

5.4. Undefined Combinations

The behavior when specifying a non-signed language tag for a video media stream, or a signed language tag for an audio or text media stream, is not defined in this document.

The problem of knowing which language tags are signed and which are not is out of scope of this document.

5.5. Examples

Some examples are shown below. For clarity, only the most directly relevant portions of the SDP block are shown.

An offer or answer indicating spoken English both ways:

```
m=audio 49170 RTP/AVP 0
a=hlang-send:en
a=hlang-recv:en
```

An offer indicating American Sign Language both ways:

```
m=video 51372 RTP/AVP 31 32
a=hlang-send:ase
a=hlang-recv:ase
```

An offer requesting spoken Spanish both ways (most preferred), spoken Basque both ways (second preference), or spoken English both ways (third preference):

```
m=audio 49250 RTP/AVP 20
a=hlang-send:es eu en
a=hlang-recv:es eu en
```

An answer to the above offer indicating spoken Spanish both ways:

```
m=audio 49250 RTP/AVP 20
a=hlang-send:es
a=hlang-recv:es
```

An alternative answer to the above offer indicating spoken Italian both ways (as the callee does not support any of the requested languages but chose to proceed with the call):

```
m=audio 49250 RTP/AVP 20
a=hlang-send:it
a=hlang-recv:it
```

An offer or answer indicating written Greek both ways:

```
m=text 45020 RTP/AVP 103 104
a=hlang-send:gr
a=hlang-recv:gr
```

An offer requesting the following media streams: video for the caller to send using Argentine Sign Language, text for the caller to send using written Spanish (most preferred) or written Portuguese, audio for the caller to receive spoken Spanish (most preferred) or spoken Portuguese:

```
m=video 51372 RTP/AVP 31 32
a=hlang-send:aed

m=text 45020 RTP/AVP 103 104
```

```
a=hlang-send:sp pt

m=audio 49250 RTP/AVP 20
a=hlang-recv:sp pt
```

An answer for the above offer, indicating text in which the callee will receive written Spanish, and audio in which the callee will send spoken Spanish. The answering party had no video capability:

```
m=video 0 RTP/AVP 31 32
m=text 45020 RTP/AVP 103 104
a=hlang-recv:sp

m=audio 49250 RTP/AVP 20
a=hlang-send:sp
```

An offer requesting the following media streams: text for the caller to send using written English (most preferred) or written Spanish, audio for the caller to receive spoken English (most preferred) or spoken Spanish, supplemental video:

```
m=text 45020 RTP/AVP 103 104
a=hlang-send:en sp

m=audio 49250 RTP/AVP 20
a=hlang-recv:en sp

m=video 51372 RTP/AVP 31 32
```

An answer for the above offer, indicating text in which the callee will receive written Spanish, audio in which the callee will send spoken Spanish, and supplemental video:

```
m=text 45020 RTP/AVP 103 104
a=hlang-recv:sp

m=audio 49250 RTP/AVP 20
a=hlang-send:sp

m=video 51372 RTP/AVP 31 32
```

Note that, even though the examples show the same (or essentially the same) language being used in both directions (even when the modality differs), there is no requirement that this be the case. However, in practice, doing so is likely to increase the chances of successful matching.

6. IANA Considerations

6.1. att-field Table in SDP Parameters

IANA is kindly requested to add two entries to the 'att-field (media level only)' table of the SDP parameters registry:

Attribute Name: hlang-recv

Contact Name: Randall Gellens

Contact Email Address: rg+ietf@randy.pensive.org

Attribute Value: hlang-value

Attribute Syntax:

```
   hlang-value =  Language-Tag *( SP Language-tag )

                        ; Language-Tag as defined in BCP 47

   SP          =  1*" " ; one or more space (%x20) characters
```

Attribute Semantics: Described in Section 5.2 of TBD: THIS DOCUMENT

Usage Level: media

Mux Category: NORMAL

Charset Dependent: No

Purpose: See Section 5.2 of TBD: THIS DOCUMENT

O/A Procedures: See Section 5.2 of TBD: THIS DOCUMENT

Reference: TBD: THIS DOCUMENT

Attribute Name: hlang-send

Contact Name: Randall Gellens

Contact Email Address: rg+ietf@randy.pensive.org

Attribute Value: hlang-value

Attribute Semantics: Described in Section 5.2 of TBD: THIS DOCUMENT

Usage Level: media

Mux Category: NORMAL

Charset Dependent: No

Purpose: See Section 5.2 of TBD: THIS DOCUMENT

O/A Procedures: See Section 5.2 of TBD: THIS DOCUMENT

Reference: TBD: THIS DOCUMENT

6.2. Warn-Codes Sub-Registry of SIP Parameters

IANA is requested to add a new value in the warn-codes sub-registry of SIP parameters in the 300 through 329 range that is allocated for indicating problems with keywords in the session description. The reference is to this document. The warn text is "Incompatible language specification: Requested languages not supported. Supported languages and media are: [list of supported languages and media]."

7. Security Considerations

The Security Considerations of BCP 47 [RFC5646] apply here. In addition, if the 'hlang-send' or 'hlang-recv' values are altered or deleted en route, the session could fail or languages incomprehensible to the caller could be selected; however, this is also a risk if any SDP parameters are modified en route.

8. Privacy Considerations

Language and media information can suggest a user's nationality, background, abilities, disabilities, etc.

9. Changes from Previous Versions

RFC EDITOR: Please remove this section prior to publication.

9.1. Changes from draft-ietf-slim-...-04 to draft-ietf-slim-...-06

- Deleted Section 3 ("Expected Use")
- Reworded modalities in Introduction from "voice, video, text" to "spoken, signed, written"
- Reworded text about "increasingly fine-grained distinctions" to instead merely point to BCP 47 Section 4.1's advice to "tag content wisely" and not include unnecessary subtags

- Changed IANA registration of new SDP attributes to follow RFC 4566 template with extra fields suggested in 4566-bis (expired draft)
- Deleted "(known as voice carry over)"
- Changed textual instanced of RFC 5646 to BCP 47, although actual reference remains RFC due to xml2rfc limitations

9.2. Changes from draft-ietf-slim-...-02 to draft-ietf-slim-...-03

- Added Examples
- Added Privacy Considerations section
- Other editorial changes for clarity

9.3. Changes from draft-ietf-slim-...-01 to draft-ietf-slim-...-02

- Deleted most of Section 4 and replaced with a very short summary
- Replaced "wishes to" with "is willing to" in Section 5.2
- Reworded description of attribute usage to clarify when to set both, only one, or neither
- Deleted all uses of "IMS"
- Other editorial changes for clarity

9.4. Changes from draft-ietf-slim-...-00 to draft-ietf-slim-...-01

- Editorial changes to wording in Section 5.

9.5. Changes from draft-gellens-slim-...-03 to draft-ietf-slim-...-00

- Updated title to reflect WG adoption

9.6. Changes from draft-gellens-slim-...-02 to draft-gellens-slim-...-03

- Removed Use Cases section, per face-to-face discussion at IETF 93
- Removed discussion of routing, per face-to-face discussion at IETF 93

9.7. Changes from draft-gellens-slim-....-01 to draft-gellens-slim-....-02

- Updated NENA usage mention
- Removed background text reference to draft-saintandre-sip-xmpp-chat-04 since that draft expired

9.8. Changes from draft-gellens-slim-....-00 to draft-gellens-slim-....-01

- Revision to keep draft from expiring

9.9. Changes from draft-gellens-mmusic-....-02 to draft-gellens-slim-....-00

- Changed name from -mmusic- to -slim- to reflect proposed WG name
- As a result of the face-to-face discussion in Toronto, the SDP vs SIP issue was resolved by going back to SDP, taking out the SIP hint, and converting what had been a set of alternate proposals for various ways of doing it within SIP into an informative annex section which includes background on why SDP is the proposal
- Added mention that enabling a mutually comprehensible language is a general problem of which this document addresses the real-time side, with reference to [RFC8255] which addresses the non-real-time side.

9.10. Changes from draft-gellens-mmusic-....-01 to -02

- Added clarifying text on leaving attributes unset for media not primarily intended for human language communication (e.g., background audio or video).
- Added new section ("Alternative Proposal: Caller-prefs") discussing use of SIP-level Caller-prefs instead of SDP-level.

9.11. Changes from draft-gellens-mmusic-....-00 to -01

- Relaxed language on setting -send and -receive to same values; added text on leaving on empty to indicate asymmetric usage.
- Added text that clients on behalf of end users are expected to set the attributes on outgoing calls and ignore on incoming calls while systems on behalf of call centers and PSAPs are expected to take the attributes into account when processing incoming calls.

9.12. Changes from draft-gellens-...-02 to draft-gellens-mmusic-...-00

- Updated text to refer to RFC 5646 rather than the IANA language subtags registry directly.
- Moved discussion of existing 'lang' attribute out of "Proposed Solution" section and into own section now that it is not part of proposal.
- Updated text about existing 'lang' attribute.
- Added example use cases.
- Replaced proposed single 'hlang' attribute with 'hlang-send' and 'hlang-recv' per Harald's request/information that it was a misuse of SDP to use the same attribute for sending and receiving.
- Added section describing usage being advisory vs required and text in attribute section.
- Added section on SIP "hint" header (not yet nailed down between new and existing header).
- Added text discussing usage in policy-based routing function or use of SIP header "hint" if unable to do so.
- Added SHOULD that the value of the parameters stick to the largest granularity of language tags.
- Added text to Introduction to be try and be more clear about purpose of document and problem being solved.
- Many wording improvements and clarifications throughout the document.
- Filled in Security Considerations.
- Filled in IANA Considerations.
- Added to Acknowledgments those who participated in the Orlando ad-hoc discussion as well as those who participated in email discussion and side one-on-one discussions.

9.13. Changes from draft-gellens-...-01 to -02

- Updated text for (possible) new attribute "hlang" to reference RFC 5646

- Added clarifying text for (possible) re-use of existing 'lang' attribute saying that the registration would be updated to reflect different semantics for multiple values for interactive versus non-interactive media.

- Added clarifying text for (possible) new attribute "hlang" to attempt to better describe the role of language tags in media in an offer and an answer.

9.14. Changes from draft-gellens-...-00 to -01

- Changed name of (possible) new attribute from 'humlang" to "hlang"
- Added discussion of silly state (language not appropriate for media type)
- Added Voice Carry Over example
- Added mention of multilingual people and multiple languages
- Minor text clarifications

10. Contributors

Gunnar Hellstrom deserves special mention for his reviews and assistance.

11. Acknowledgments

Many thanks to Bernard Aboba, Harald Alvestrand, Flemming Andreasen, Francois Audet, Eric Burger, Keith Drage, Doug Ewell, Christian Groves, Andrew Hutton, Hadriel Kaplan, Ari Keranen, John Klensin, Paul Kyzivat, John Levine, Alexey Melnikov, James Polk, Pete Resnick, Natasha Rooney, Brian Rosen, Peter Saint-Andre, and Dale Worley for reviews, corrections, suggestions, and participating in in-person and email discussions.

12. References

12.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, DOI 10.17487/RFC3261, June 2002, <https://www.rfc-editor.org/info/rfc3261>.

[RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", RFC 4566, DOI 10.17487/RFC4566, July 2006, <https://www.rfc-editor.org/info/rfc4566>.

[RFC5646] Phillips, A., Ed. and M. Davis, Ed., "Tags for Identifying Languages", BCP 47, RFC 5646, DOI 10.17487/RFC5646, September 2009, <https://www.rfc-editor.org/info/rfc5646>.

12.2. Informational References

[RFC3264] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", RFC 3264, DOI 10.17487/RFC3264, June 2002, <https://www.rfc-editor.org/info/rfc3264>.

[RFC8255] Tomkinson, N. and N. Borenstein, "Multiple Language Content Type", RFC 8255, DOI 10.17487/RFC8255, October 2017, <https://www.rfc-editor.org/info/rfc8255>.

Author's Address

Randall Gellens
Core Technology Consulting

Email: rg+ietf@coretechnologyconsulting.com
URI: http://www.coretechnologyconsulting.com