

MMUSIC Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 13, 2014

R. Gellens
Qualcomm Technologies Inc.
July 14, 2013

Negotiating Human Language Using SDP
draft-gellens-mmusic-negotiating-human-language-01

Abstract

Users have various human (natural) language needs, abilities, and preferences regarding spoken, written, and signed languages. When establishing interactive communication "calls" there needs to be a way to communicate and ideally match (i.e., negotiate) the caller's language preferences with the capabilities of the called party. This is especially important with emergency calls, where a call can be routed to a Public Safety Answering Point (PSAP) or 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 expected use, and describes a solution using new SDP stream attributes plus an optional SIP "hint."

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 January 13, 2014.

Copyright Notice

Copyright (c) 2013 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	2
2. Terminology	4
3. Expected Use	5
4. Example Use Cases	5
4.1. Emergency Call from English Speaker in Spain	5
4.2. Emergency Call from Spanish/English Speaker in France	5
4.3. Call to Call Center from Russian Speaker in U.S.	6
4.4. Emergency Call from speech-impaired caller in the U.S.	6
4.5. Emergency Call from deaf caller in the U.S.	6
5. Desired Semantics	7
6. The existing 'lang' attribute	7
7. Proposed Solution	8
7.1. New 'humintlang-send' and 'humintlang-recv' attributes	8
7.2. Advisory vs Required	10
7.3. SIP "hint"	10
7.4. Silly States	11
8. IANA Considerations	11
9. Security Considerations	11
10. Changes from Previous Versions	12
10.1. Changes from draft-gellens-...-00 to -01	12
10.2. Changes from draft-gellens-...-01 to -02	12
10.3. Changes from draft-gellens-...-02 to draft-gellens-mmusic-...-00	12
10.4. draft-gellens-mmusic-...-00 to -01	13
11. Acknowledgments	13
12. References	13
12.1. Normative References	13
12.2. Informational References	14
Author's Address	14

1. Introduction

When setting up interactive communication sessions (using SIP or other protocols), human (natural) language negotiation may be needed. When the caller and callee know each other or where context or out of band information implies the language, such negotiation is typically not needed. In other cases, there is a need for spoken, signed, or written languages to be negotiated based on the caller's preferences and the callee's capabilities. This need applies to both emergency and non-emergency calls. 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, in this case, SDP.

This approach has a number of benefits, including that it is generic (applies to all interactive communications negotiated using SDP) and not limited to emergency calls. In some cases such a facility isn't needed, because the language is known from the context (such as when a caller places a call to a sign language relay center, to a friend, or colleague). But it is clearly useful in many other cases. For example, someone calling a company call center or a Public Safety Answering Point (PSAP) should be able to indicate if one or more specific signed, written, and/or spoken languages are preferred, the callee should be able to indicate its capabilities in this area, and the call proceed using in-common language(s) and media forms.

Since this is a protocol mechanism, the user equipment (UE client) needs to know the user's preferred languages; a reasonable technique 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.

Including the user's human (natural) language preferences in the session establishment negotiation is independent of the use of a relay service and is transparent to a voice service provider. For example, assume a user within the United States who speaks Spanish but not English places a voice call using an IMS device. It doesn't matter if the call is an emergency call or not (e.g., to an airline reservation desk). The language information is transparent to the IMS carrier, but is part of the session negotiation between the UE and the terminating entity. In the case of a call to e.g., an airline, the call can be automatically routed to a Spanish-speaking agent. In the case of an emergency call, the Emergency Services IP network (ESInet) and the PSAP may choose to take the language and media preferences into account when determining how to route and process the call (i.e., language and media needs may be considered within policy-based routing (PBR)).

By treating language as another 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 requires a voice stream plus a text stream (known as voice carry over). 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.

Regarding relay services, in the case of an emergency call requiring sign language such as ASL, there are two common approaches: the caller initiates the call to a relay center, or the caller places the call to emergency services (e.g., 911 in the U.S. or 112 in Europe). In the former case, the language need is ancillary and supplemental. In the latter case, the ESInet and/or PSAP may take the need for sign language into account and bridge in a relay center. In this case, the ESInet and PSAP have all the standard information available (such as location) but are able to bridge the relay sooner in the call processing.

By making this facility part of the end-to-end negotiation, the question of which entity provides or engages the relay service becomes separate from the call processing mechanics; if the caller directs the call to a relay service then the human language negotiation facility provides extra information to the relay service but calls will still function without it; if the caller directs the call to emergency services, then the ESInet/PSAP are able to take the user's human language needs into account, e.g., by routing to a particular PSAP or call taker or bridging a relay service or translator.

The term "negotiation" is used here rather than "indication" because human language (spoken/written/signed) is something that can be negotiated in the same way as which forms of media (audio/text/video) or which codecs. For example, if we think of non-emergency calls, such as 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).

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. Expected Use

This facility is expected to be used by NENA and 3GPP. NENA is likely to reference it in NENA 08-01 (i3 Stage 3) in describing attributes of calls presented to an ESInet, and in that or other documents describing Policy-Based Routing (PBR) capabilities within a Policy-Based Routing Function. 3GPP is expected to reference this mechanism in general call handling and emergency call handling. Recent CRs introduced in SA1 have anticipated this functionality being provided within SDP.

4. Example Use Cases

4.1. Emergency Call from English Speaker in Spain

Someone who speaks only English is visiting Spain and places an emergency (112) call. The call offers an audio stream using English. The ESInet and PSAP have policy-based routing rules that take into account the SDP language request when deciding how to route and process the call. The ESInet routes the call to a PSAP within Spain where an English-speaking call taker is available, and the PSAP selects an English-speaking call taker to handle the call. The PSAP answers the offer with an audio stream using English. The call is established with an audio stream; the caller and call taker communicate in English.

Alternatively, the ESInet routes the call to a cooperating PSAP within the U.K. The PSAP answers the offer with an audio stream using English. The call is established with an audio stream; the caller and call taker communicate in English. (This approach is similar to that envisioned in REACH112 Total Conversation.)

4.2. Emergency Call from Spanish/English Speaker in France

Someone who speaks both Spanish and English (but prefers Spanish) is visiting France and places an emergency (112) call. The call offers an audio stream listing first Spanish (meaning most preferred) and then English. The ESInet and PSAP have policy-based routing rules that take into account the SDP language request when deciding how to route and process the call. The ESInet routes the call to a PSAP within France where a Spanish-speaking call taker is available, and the PSAP selects a Spanish-speaking call taker to handle the call. The PSAP answers the offer with an audio stream listing Spanish. The call is established with an audio stream; the caller and call taker

communicate in Spanish.

Alternatively, the ESInet routes the call to a cooperating PSAP in Spain or England. (This approach is similar to that envisioned in REACH112 Total Conversation.)

Alternatively, there is no ESInet or the ESInet does not take language into account in its PBR. The call is routed to a PSAP in France. The PSAP ignores the language information in the SDP offer, and answers the offer with an audio stream with no language or with French. The UE continues the call anyway. The call taker answers in French, the user tries speaking Spanish and perhaps English. The call taker bridges in a translation service or transfers the call to a multilingual call taker.

4.3. Call to Call Center from Russian Speaker in U.S.

A Russian speaker is visiting the U.S. and places a call to her airline reservation desk to inquire about her return flight. The airline call processing system takes into account the SDP language request and decides to route the call to its call center within Russia.

Alternatively, if the airline call processing system does not look at SDP, it uses the SIP "hint" if present.

4.4. Emergency Call from speech-impaired caller in the U.S.

Someone who uses English but is speech-impaired places an emergency (911) call. The call offers an audio stream listing English and a real-time text stream also using English. The ESInet and PSAP have policy-based routing rules that take into account the SDP language and media requests when deciding how to route and process the call. The ESInet routes the call to a PSAP with real-time text capabilities. The PSAP answers the offer with an audio stream listing English and a real-time text stream listing English. The call is established with an audio and a real-time text stream; the caller and call taker communicate in English using voice from the call-taker to the caller and text from the caller to the call taker. The audio stream is two-way, allowing the call taker to hear background sounds.

4.5. Emergency Call from deaf caller in the U.S.

A deaf caller who uses American Sign Language (ASL) places an emergency (911) call. The call offers a video stream listing ASL and an audio stream with no language indicated. The ESInet and PSAP have policy-based routing rules that take into account the SDP language and media needs when deciding how to route and process the call. The ESInet routes the call to a PSAP. The PSAP answers the offer with an

audio stream listing English and a video stream listing ASL. The PSAP bridges in a sign language interpreter. The call is established with an audio and a video stream.

5. Desired Semantics

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

(While it is true that a conversation among multilingual people often involves multiple languages, the usefulness of providing a way to negotiate this as a general facility is outweighed by the complexity of the desired semantics of the SDP attribute to allow negotiation of multiple simultaneous languages within an interactive media stream.)

6. The existing 'lang' attribute

RFC 4566 specifies an attribute 'lang' which sounds similar to what is needed here, the difference being that it specifies that 'a=lang' is declarative with the semantics of multiple 'lang' attributes being that all of them are used, while we want a means to negotiate which one is used in each stream. This difference means that the existing 'lang' attribute can't be used and we need to define a new attribute.

The text from RFC 4566 [RFC4566] is:

```
a=lang:<language tag>
```

This can be a session-level attribute or a media-level attribute. As a session-level attribute, it specifies the default language for the session being described. As a media-level attribute, it specifies the language for that media, overriding any session-level language specified. Multiple lang attributes can be provided either at session or media level if the session description or media use multiple languages, in which case the order of the attributes indicates the order of importance of the various languages in the session or media from most important to least important.

The "lang" attribute value must be a single [RFC3066] language tag in US-ASCII [RFC3066]. It is not dependent on the charset attribute. A "lang" attribute SHOULD be specified when a session is of sufficient scope to cross geographic boundaries where the language of recipients cannot be assumed, or where the session is in a different language from the locally assumed norm.

Note that there are existing examples of it being used in exactly the way we need. For example, draft-saintandre-sip-xmpp-chat-04 [I-D .saintandre-sip-xmpp-chat] contains an example where the initial invitation contains two 'a=lang' entries for a media stream (for English and Italian) and the OK accepts one of them (Italian), which matches what we need:

Example: (F1) SIP user starts the session

```
INVITE sip:juliet@example.com SIP/2.0
To: <sip:juliet@example.com>
From: <sip:romeo@example.net>;tag=576
Subject: Open chat with Romeo?
Call-ID: 742507no
Content-Type: application/sdp

c=IN IP4 s2x.example.net
m=message 7313 TCP/MSRP *
a=accept-types:text/plain
a=lang:en
a=lang:it
a=path:msrp://s2x.example.net:7313/ansp7lweztas;tcp
```

Example: (F2) Gateway accepts session on Juliet's behalf

```
SIP/2.0 200 OK
To: <sip:juliet@example.com>;tag=534
From: <sip:romeo@example.net>;tag=576
Call-ID: 742507no
Content-Type: application/sdp

c=IN IP4 x2s.example.com
m=message 8763 TCP/MSRP *
a=accept-types:text/plain
a=lang:it
a=path:msrp://x2s.example.com:8763/lkjh37s2s20w2a;tcp
```

The example serves as evidence of the need for an SDP attribute with the semantics as described in this document; unfortunately, the existing 'lang' attribute is not it.

7. Proposed Solution

An SDP attribute seems the natural choice to negotiate human (natural) language of an interactive media stream. The attribute value should be a language tag per RFC 5646 [RFC5646]

7.1. New 'humintlang-send' and 'humintlang-recv' attributes

Rather than re-use 'lang' we define two new media-level attributes starting with 'humintlang' (short for "human interactive language") to negotiate which human language is used in each (interactive) media stream. There are two attributes, one ending in "-send" and the other in "-recv" to indicate the language used when sending and receiving media:

a=humintlang-send:<language tag>

a=humintlang-recv:<language tag>

Each can appear multiple times in an offer for a media stream.

In an offer, the 'humintlang-send' values constitute a list in preference order (first is most preferred) of the languages the offerer wishes to send, and the 'humintlang-recv' values constitute a list in preference order of the languages the offerer wishes to receive. In cases where the user wishes to use one media for sending and another for receiving (such as a speech-impaired user who wishes to send using text and receive using audio), one of the two will be unset. In other cases, both SHOULD have the same values in the same order. 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, 'humintlang-send' is the accepted language the answerer will send (which in most cases is one of the languages in the offer's 'humintlang-recv'), and 'humintlang-recv' is the accepted language the answerer expects to receive (which in most cases is one of the languages in the offer's 'humintlang-send').

Each value MUST be a language tag per RFC 5646 [RFC5646]. RFC 5646 describes mechanisms for matching language tags. While RFC 5646 provides a mechanism accommodating increasingly fine-grained distinctions, in the interest of maximum interoperability for real-time interactive communications, each 'humintlang-send' and 'humintlang-recv' value SHOULD be restricted to the largest granularity of language tags; in other words, it is RECOMMENDED to specify only a Primary-subtag and NOT to include subtags (e.g., for region or dialect) unless the languages might be mutually incomprehensible without them.

In an offer, each language tag value MAY have an asterisk appended as the last character (after the registry value). The asterisk indicates a request by the caller to not fail the call if there is no language in common. See Section 7.2 for more information and discussion.

When placing an emergency call, and in any other case where the language cannot be assumed from context, each media stream in an

offer SHOULD specify one or both 'humintlang-send' and 'humintlang-recv' attributes (to avoid ambiguity).

Note that while signed language tags are used with a video stream to indicate sign language, a spoken language tag for a video stream in parallel with an audio stream with the same spoken language tag indicates a request for a supplemental video stream to see the speaker.

Clients acting on behalf of end users are expected to set one or both 'humintlang-send' and 'humintlang-recv' attributes on each media stream in an offer when placing an outgoing session but ignore the attributes when receiving incoming calls. Systems acting on behalf of call centers and PSAPs are expected to take into account the values when processing inbound calls.

7.2. Advisory vs Required

One important consideration with this mechanism is if the call fails if the callee does not support any of the languages requested by the caller.

In order to provide for maximum likelihood of a successful communication session, especially in the case of emergency calling, the mechanism defined here provides a way for the caller to indicate a preference for the call failing or succeeding when there is no language in common. However, the callee is NOT REQUIRED to honor this preference. For example, a PSAP MAY choose to attempt the call even with no language in common, while a corporate call center MAY choose to fail the call.

The mechanism for indicating this preference is that, in an offer, if the last character of any of the 'humintlang-recv' or 'humintlang-send' values is an asterisk, this indicates a request to not fail the call (similar to SIP Accept-Language syntax). Either way, the called party MAY ignore this, e.g., for the emergency services use case, a PSAP will likely not fail the call.

7.3. SIP "hint"

SDP is used for stream negotiation, and emergency services based on NENA i3 have the ability to reference SDP within policy-based routing rules. However, it is possible that some entities wishing to take the caller's language needs into account may lack this ability. Accordingly, a SIP header is provided to supply a "hint" regarding the caller's language needs. This is merely a hint, not actual negotiation.

Protocols other than SIP may be used to establish interactive communication sessions; this document does not provide a "hint"

mechanism for any protocol other than SIP.

TBD: The SIP header used for the "hint" is either a new header or caller preferences (RFC 3841), presumably Accept-Contact (where the caller sets media and language and optionally required; it's not yet clear if this matches our desired semantics.

This SIP header is just a hint; it is RECOMMENDED to be sent; it is NOT REQUIRED to be sent or to be used on receipt.

In the case of spoken languages using an audio stream, this "hint" regarding language may be sufficient to allow the callee to optimally handle the call, but since the "hint" only deals with language, not media type, it is not sufficient when the caller requests non-audio media such as text or video.

7.4. Silly States

It is possible to specify a "silly state" where the language specified does not make sense for the media type, such as specifying a signed language for an audio media stream.

An offer MUST NOT be created where the language does not make sense for the media type. If such an offer is received, the receiver MAY reject the media, ignore the language specified, or attempt to interpret the intent (e.g., if American Sign Language is specified for an audio media stream, this might be interpreted as a desire to use spoken English).

A spoken language tag for a video stream in conjunction with an audio stream with the same language might indicate a request for supplemental video to see the speaker.

8. IANA Considerations

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

Type	Name	Reference
att-field (media level only)	humintlang-send	(this document)
att-field (media level only)	humintlang-recv	(this document)

9. Security Considerations

The Security Considerations of RFC 5646 [RFC5646] apply here (as a use of that RFC). In addition, if the 'humintlang-send' or 'humintlang-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.

10. Changes from Previous Versions

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

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

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

- o Updated text for (possible) new attribute "humintlang" to reference RFC 5646
- o 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.
- o Added clarifying text for (possible) new attribute "humintlang" to attempt to better describe the role of language tags in media in an offer and an answer.

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

- o Updated text to refer to RFC 5646 rather than the IANA language subtags registry directly.
- o Moved discussion of existing 'lang' attribute out of "Proposed Solution" section and into own section now that it is not part of proposal.
- o Updated text about existing 'lang' attribute.
- o Added example use cases.
- o Replaced proposed single 'humintlang' attribute with 'humintlang-send' and 'humintlang-recv' per Harald's request/information that it was a misuse of SDP to use the same attribute for sending and receiving.
- o Added section describing usage being advisory vs required and text in attribute section.
- o Added section on SIP "hint" header (not yet nailed down between new and existing header).

- o Added text discussing usage in policy-based routing function or use of SIP header "hint" if unable to do so.
- o Added SHOULD that the value of the parameters stick to the largest granularity of language tags.
- o Added text to Introduction to be try and be more clear about purpose of document and problem being solved.
- o Many wording improvements and clarifications throughout the document.
- o Filled in Security Considerations.
- o Filled in IANA Considerations.
- o 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.

10.4. draft-gellens-mmusic-...-00 to -01

- o Relaxed language on setting -send and -receive to same values; added text on leaving on empty to indicate asymmetric usage.
- o 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.

11. Acknowledgments

Many thanks to Bernard Aboba, Harald Alvestrand, Flemming Andreassen, Francois Audet, Eric Burger, Keith Drage, Doug Ewell, Christian Groves, Gunnar Hellstrom, Andrew Hutton, Hadriel Kaplan, Ari Keranen, John Klensin, Paul Kyzivat, John Levine, Alexey Melnikov, James Polk, Pete Resnick, 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, March 1997.
- [RFC4566] Handley, M., Jacobson, V. and C. Perkins, "SDP: Session Description Protocol", RFC 4566, July 2006.
- [RFC5646] Phillips, A. and M. Davis, "Tags for Identifying Languages", BCP 47, RFC 5646, September 2009.

12.2. Informational References

[I-D.iab-privacy-considerations]

Cooper, A., Tschofenig, H., Aboba, B., Peterson, J.,
Morris, J., Hansen, M. and R. Smith, "Privacy
Considerations for Internet Protocols", Internet-Draft
draft-iab-privacy-considerations-03, July 2012.

[I-D.saintandre-sip-xmpp-chat]

Saint-Andre, P., Gavita, E., Hossain, N. and S. Loreto,
"Interworking between the Session Initiation Protocol
(SIP) and the Extensible Messaging and Presence Protocol
(XMPP): One-to-One Text Chat", Internet-Draft draft-
saintandre-sip-xmpp-chat-04, October 2012.

[RFC3066] Alvestrand, H., "Tags for the Identification of
Languages", RFC 3066, January 2001.

Author's Address

Randall Gellens
Qualcomm Technologies Inc.
5775 Morehouse Drive
San Diego, CA 92121
US

Email: rg+ietf@qti.qualcomm.com