

Groupe de travail Réseau

Request for Comments : 5574

Catégorie : Sur la voie de la normalisation

Traduction Claude Brière de L'Isle

G. Herlein, indépendant

J. Valin, Xiph.Org Foundation

A. Heggstad, Creytiv.com

A. Moizard, Antisip

juin 2009

Format de charge utile RTP pour codec Speex

Statut du présent mémoire

Le présent document spécifie un protocole de l'Internet sur la voie de la normalisation pour la communauté de l'Internet, et appelle à des discussions et suggestions pour son amélioration. Prière de se référer à l'édition en cours des "Protocoles officiels de l'Internet" (STD 1) pour voir l'état de normalisation et le statut de ce protocole. La distribution du présent mémoire n'est soumise à aucune restriction.

Notice de droits de reproduction

Copyright (c) 2009 IETF Trust et les personnes identifiées comme auteurs du document. Tous droits réservés.

Le présent document est soumis au BCP 78 et aux dispositions légales de l'IETF Trust qui se rapportent aux documents de l'IETF (<http://trustee.ietf.org/license-info>) en vigueur à la date de publication de ce document. Prière de revoir ces documents avec attention, car ils décrivent vos droits et obligations par rapport à ce document.

Le présent document peut contenir des matériaux provenant de documents de l'IETF ou de contributions à l'IETF publiées ou rendues disponibles au public avant le 10 novembre 2008. La ou les personnes qui ont le contrôle des droits de reproduction sur tout ou partie de ces matériaux peuvent n'avoir pas accordé à l'IETF Trust le droit de permettre des modifications de ces matériaux en dehors du processus de normalisation de l'IETF. Sans l'obtention d'une licence adéquate de la part de la ou des personnes qui ont le contrôle des droits de reproduction de ces matériaux, le présent document ne peut pas être modifié en dehors du processus de normalisation de l'IETF, et des travaux dérivés ne peuvent pas être créés en dehors du processus de normalisation de l'IETF, excepté pour le formater en vue de sa publication comme RFC ou pour le traduire dans une autre langue que l'anglais.

Résumé

Speex est un codec vocal en accès libre convenable pour l'usage d'application de type voix sur IP (VoIP, *Voice over IP*). Le présent document décrit le format de charge utile pour les flux binaires générés par Speex au sein d'un paquet RTP. Les détails nécessaires pour l'utilisation de Speex avec le protocole de description de session (SDP, *Session Description Protocol*) sont aussi inclus.

Table des matières

1. Introduction.....	2
2. Terminologie.....	2
3. Usage RTP pour Speex.....	2
3.1 Champs d'en-tête RTP Speex.....	2
3.2 Format de charge utile RTP pour Speex.....	2
3.3 Charge utile Speex.....	3
3.4 Exemple de paquet Speex.....	3
3.5 Plusieurs trames Speex dans un paquet RTP.....	3
4. Considérations relatives à l'IANA.....	4
4.1 Enregistrement du type de support.....	4
5. Usage de Speex par SDP	5
5.1 Exemple de prise en charge de tous les modes, mode 4 préféré.....	6
5.2 Exemple de prise en charge des seuls modes 3 et 5.....	6
5.3 Exemple avec débit binaire variable et bruit de confort.....	6
5.4 Exemple avec détection d'activité vocale.....	6
5.5 Exemple avec plusieurs taux d'échantillonnage.....	7
5.6 Exemple avec Ptime et plusieurs trames Speex.....	7
5.7 Exemple avec échange d'offre/réponse complet.....	7
6. Lignes directrices pour la mise en œuvre.....	7
7. Considérations sur la sécurité.....	7

8. Remerciements.....	8
9. Références.....	8
9.1 Références normatives.....	8
9.2 Références pour information.....	8
Adresse des auteurs.....	9

1. Introduction

Speex se fonde sur la technique de codage prédictif linéaire avec excitation par code (CELP, *Code Excited Linear Prediction*) [CELP] avec prise en charge de la bande étroite (8 kHz nominal) de la large bande (16 kHz nominal) ou de l'ultra large bande (32 kHz nominal). Les principales caractéristiques peuvent être résumées comme suit :

- o Logiciel libre/accès gratuit
- o Intégration du large bande et de la bande étroite dans le même flux binaire
- o Large gamme de débits binaires disponibles
- o Commutation dynamique de débit binaire et débit binaire variable (VBR)
- o Détection d'activité vocale (VAD, intégrée avec VBR)
- o Complexité variable

Le codec Speex prend en charge une large gamme de débits binaires de 2,15 kbit/s à 44 kbit/s. Dans certains cas cependant, il peut ne pas être possible à une mise en œuvre d'inclure la prise en charge de tous les débits (par exemple, à cause de contraintes de bande passante, de RAM ou de CPU). Dans ces cas, pour être conforme à la présente spécification, les mises en œuvre DOIVENT prendre en charge au moins le codage en bande étroite (8 kHz) et le décodage à 8 kbit/s (bande étroite mode 3). La prise en charge de la bande étroite à 15 kbit/s (bande étroite mode 5) est RECOMMANDÉE et la prise en charge du large bande à 27,8 kbit/s (large bande mode 8) est aussi RECOMMANDÉE. Le taux d'échantillonnage DOIT être 8, 16 ou 32 kHz. La présente spécification définit seulement un seul canal audio (mono).

2. Terminologie

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" en majuscules dans ce document sont à interpréter comme décrit dans la [RFC2119] et indiquent les niveaux d'exigence pour les mises en œuvre conformes.

3. Usage RTP pour Speex

3.1 Champs d'en-tête RTP Speex

L'en-tête RTP est défini dans la spécification de RTP [RFC3550]. Ce paragraphe définit comment les champs dans l'en-tête RTP sont utilisés.

Type de charge utile (PT) : L'allocation d'un type de charge utile RTP pour ce format de paquet sort du domaine d'application du présent document ; il est spécifié par le profil RTP dans lequel ce format de charge utile est utilisé, ou signalé dynamiquement hors bande (par exemple, en utilisant SDP).

Bit Marqueur (M) : le bit M est réglé à un sur le premier paquet envoyé après une période de silence, durant laquelle des paquets n'ont pas été transmis de façon contiguë.

Bit Extension (X) : défini par le profil RTP utilisé.

Horodatage : mot de 32 bits qui correspond à l'instant d'échantillonnage pour la première trame dans le paquet RTP.

3.2 Format de charge utile RTP pour Speex

La charge utile RTP pour Speex a le format montré à la Figure 1. Aucun champ d'en-tête supplémentaire spécifique de ce format de charge utile n'est exigé. Pour le transport fondé sur RTP d'audio codé en Speex, l'en-tête RTP standard [RFC3550] est suivi de un ou plusieurs blocs de données de charge utile. Un bourrage terminal facultatif peut aussi être utilisé.

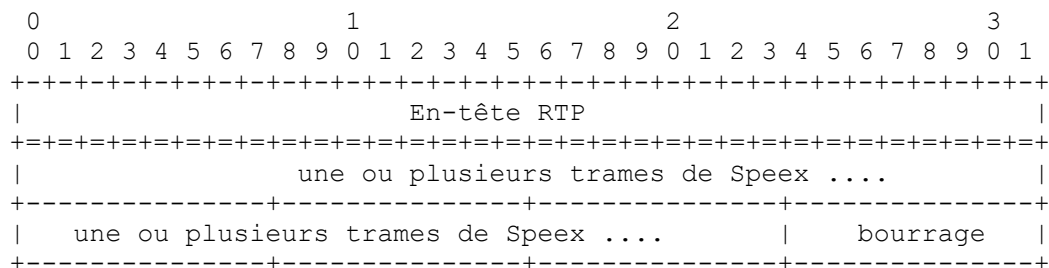


Figure 1 : Charge utile RTP pour Speex

3.3 Charge utile Speex

Pour les besoins de la mise en paquets du flux binaire dans RTP, il est seulement nécessaire de considérer la séquence de bits telle que sortie par le codeur Speex [SPEEX], et de présenter la même séquence au décodeur. Le format de charge utile décrit ici maintient cette séquence.

Une trame Speex normale, codée au débit binaire maximum, fait approximativement 110 octets et le nombre total de trames Speex DEVRAIT être gardé inférieur à la MTU de chemin MTU pour prévenir la fragmentation. Les trames Speex NE DOIVENT PAS être fragmentées sur plusieurs paquets RTP.

Les trames Speex doivent être placées en commençant par la trame la plus ancienne et en continuant ensuite consécutivement dans le temps.

Un paquet RTP PEUT contenir des trames Speex de même débit binaire ou de débits binaires variables, car le débit binaire pour une trame est convoyé dans la bande avec le signal.

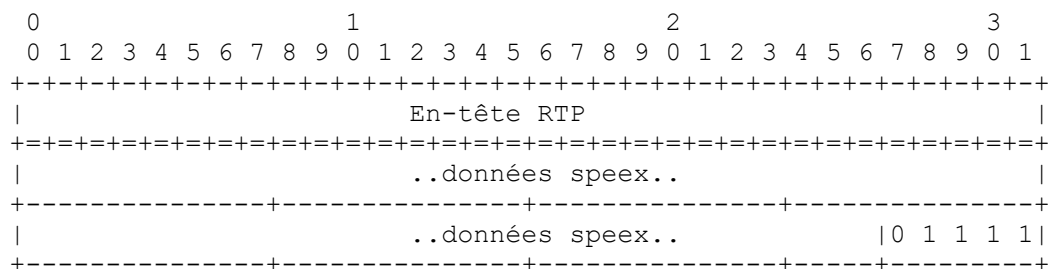
L'algorithme de codage et décodage peut changer le débit binaire à toute limite de trame de 20 ms, avec la notification de changement de débit binaire fournie dans la bande avec le flux binaire. Chaque trame contient les informations de taux d'échantillonnage (bande étroite, large bande, ou ultra large bande) et "mode" (débit binaire) dans le flux binaire. Aucune notification hors bande n'est exigée pour que le décodeur traite les changements de débit binaire envoyés par le codeur.

Le taux d'échantillonnage DOIT être 8 000 Hz, 16 000 Hz, ou 32 000 Hz.

La charge utile RTP DOIT être bourrée pour fournir un nombre entier d'octets comme longueur de charge utile. Ces bits de bourrage sont alignés sur le bit de moindre poids (LSB, *Least Significant Bit*) dans l'ordre des octets du réseau et consiste en un 0 suivi par des uns (jusqu'à la fin de l'octet). Ce bourrage est seulement exigé pour la dernière trame du paquet, et seulement pour assurer que le contenu du paquet se termine sur une limite d'octet.

3.4 Exemple de paquet Speex

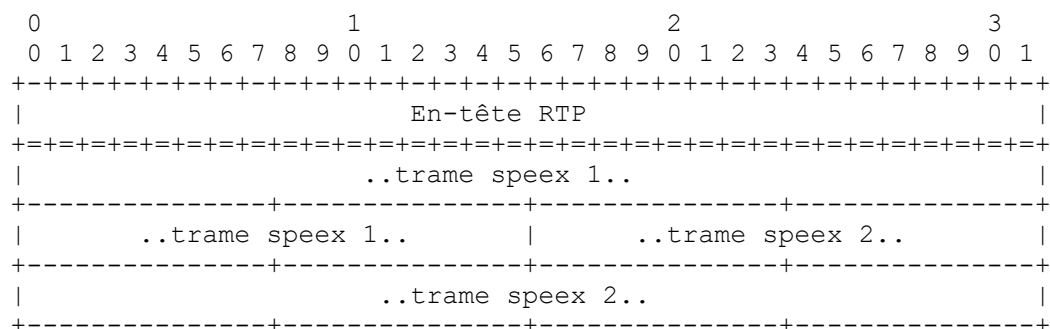
Dans l'exemple ci-dessous, on a une seule trame Speex avec 5 bits de bourrage pour assurer que la taille de paquet tombe sur une limite d'octet.



3.5 Plusieurs trames Speex dans un paquet RTP

Voici un exemple de deux trames Speex contenues dans un paquet RTP. La longueur de la trame Speex dans cet exemple tombe sur une limite d'octet de sorte qu'il n'y a pas de bourrage.

Le décodeur Speex [SPEEX] peut détecter le débit binaire d'après la charge utile et est chargé de détecter les limites de 20 ms entre chaque trame.



4. Considérations relatives à l'IANA

Le présent document définit le type de support Speex.

4.1 Enregistrement du type de support

Ce paragraphe décrit les types de supports et les noms associés à ce format de charge utile. Il enregistre les types de supports, conformément à la [RFC4288].

4.1.1 Enregistrement du type de support Audio/Speex

Nom de type de supports : audio

Nom de sous type de supports : speex

Paramètres exigés :

rate : débit d'horloge d'horodatage RTP, qui est égal au taux d'échantillonnage en Hz. Le taux d'échantillonnage DOIT être 8000, 16000, ou 32000.

Paramètres facultatifs :

ptime : DEVRAIT être un multiple de 20 ms [RFC4566]

maxptime : DEVRAIT être un multiple de 20 ms [RFC4566]

vbr : débit binaire variable - "on", "off", ou "vad" ("off" par défaut). Si "on", le débit binaire variable est activé. Si "off", il est désactivé. Si il est réglé à "vad", alors un débit binaire constant est utilisé, mais le silence va être codé avec des trames courtes spéciales pour indiquer une absence de voix pour cette période. Ce paramètre est une préférence pour le codeur.

cng : génération de bruit de confort - "on" ou "off" ("off" par défaut). Si c'est "off" les trames de silence vont être silencieuses ; si c'est "on", ces trames vont alors être remplies avec du bruit de confort. Ce paramètre est une préférence pour le codeur.

mode : liste séparée de virgules des modes de décodage Speex pris en charge, par ordre de préférence. Le premier est le préféré et le reste est en ordre de préférence décroissante. Les modes valides sont différents pour la bande étroite et pour la large bande, et sont définis comme suit :

* {1,2,3,4,5,6,7,8,any} pour la bande étroite,

* {0,1,2,3,4,5,6,7,8,9,10,any} pour la large bande et l'ultra large bande.

Le paramètre "mode" peut contenir plusieurs valeurs. Dans ce cas, la partie distante DEVRAIT configurer son codeur en utilisant le premier mode pris en charge fourni. Quand "any" est utilisé, l'offreur indique qu'il prend en charge tous les modes de décodage. La valeur du paramètre "mode" DOIT toujours être entre guillemets. Si le paramètre "mode" n'est pas fourni, la valeur de mode est considérée être équivalente à "mode="3,any"" en bande étroite et à "mode="8,any"" en large bande et ultra large bande. Noter que chaque trame Speex contient bien le mode (ou débit binaire) qui devrait être utilisé pour la décodage. Donc, une application DOIT être capable de décodage toute trame Speex

sauf si le SDP spécifie clairement que certains modes ne sont pas pris en charge (par exemple, en n'incluant pas "mode="any"). Indiquer la prise en charge d'un ensemble donné de modes de décodage implique aussi que la mise en œuvre prend en charge les mêmes modes de codage.

Considérations de codage : ce type de support est tramé et binaire, voir le paragraphe 4.8 de la [RFC4288].

Considérations sur la sécurité : voir la Section 6.

Considérations d'interopérabilité : aucune.

Spécification publiée : RFC 5574.

Applications qui utilisent ce type de support : applications de flux audio et de conférence.

Informations supplémentaires : aucune.

Personne et adresse de messagerie à contacter pour plus d'informations : Alfred E. Heggstad: aeh@db.org

Usage prévu : COMMUN

Restrictions d'usage : ce type de support dépend du tramage RTP, et donc est seulement défini pour le transfert via RTP [RFC3550]. Le transport dans d'autres protocoles de tramage n'est pas défini pour le moment.

Auteur : Alfred E. Heggstad

Contrôleur des changements : groupe de travail Transport audio/vidéo de l'IETF sur délégation de l'IESG.

5. Usage de Speex par SDP

Les informations portées dans la spécification de type de support ont une transposition spécifique en champs dans le protocole de description de session (SDP) [RFC4566], qui est couramment utilisée pour décrire les sessions RTP. Quand SDP est utilisé pour spécifier des sessions employant le codec Speex, la transposition est comme suit :

- o Le type de support ("audio") va dans le "m=" SDP comme nom de support.
- o Le sous type de support ("speex") va dans le "a=rtpmap" SDP comme nom de codage. Le paramètre exigé "rate" va aussi dans un "a=rtpmap" comme débit d'horloge.
- o Les paramètres "ptime" et "maxptime" vont respectivement dans les attributs SDP "a=ptime" et "a=maxptime".
- o Tous les paramètres restants vont dans l'attribut SDP "a=fmtp" en les copiant directement de la chaîne Type de support comme une liste séparée par des points virgules de paires de paramètre=valeur.

Les tableaux ci-dessous incluent l'équivalence entre modes et débits binaires pour la bande étroite, la large bande, et l'ultra large bande. Aussi, le réglage "qualité Speex" correspondant (voir SPEEX_SET_QUALITY dans le manuel de codec Speex [SPEEX]) est inclus comme indication.

Mode	Qualité Speex	Débit binaire
1	0	2,15 kbit/s
2	2	5,95 kbit/s
3	3 ou 4	8,00 kbit/s
4	5 ou 6	11,0 kbit/s
5	7 ou 8	15,0 kbit/s
6	9	18,2 kbit/s
7	10	24,6 kbit/s
8	1	3,95 kbit/s

Tableau 1 : Mode contre débit binaire pour bande étroite

Mode	Qualité Speex	Débit large bande	Débit ultra large bande
0	0	3,95 kbit/s	5,75 kbit/s
1	1	5,75 kbit/s	7,55 kbit/s
2	2	7,75 kbit/s	9,55 kbit/s
3	3	9,80 kbit/s	11,6 kbit/s
4	4	12,8 kbit/s	14,6 kbit/s
5	5	16,8 kbit/s	18,6 kbit/s
6	6	20,6 kbit/s	22,4 kbit/s
7	7	23,8 kbit/s	25,6 kbit/s
8	8	27,8 kbit/s	29,6 kbit/s
9	9	34,2 kbit/s	36,0 kbit/s
10	10	42,2 kbit/s	44,0 kbit/s

Tableau 2 : Mode contre débit binaire pour large bande et ultra large bande

Les paramètres Speex indiquent les capacités de décodage de l'agent, et ce que l'agent préfère recevoir.

Les paramètres Speex dans un échange d'offre/réponse SDP sont complètement orthogonaux, et il n'y a pas de relation entre l'offre et la réponse SDP.

Plusieurs paramètres spécifiques de Speex peuvent être donnés dans une seule ligne a=fmtp pourvu qu'ils soient séparés par un point-virgule :

```
a=fmtp:97 mode="1,any";vbr=on
```

Des exemples de description de session SDP utilisant des codages Speex suivent.

5.1 Exemple de prise en charge de tous les modes, mode 4 préféré

L'offreur indique qu'il souhaite recevoir un flux Speex à 8000 Hz, et souhaite recevoir le "mode 4" Speex. Il est important de comprendre que tout autre mode pourrait quand même être envoyé par la partie distante : l'appareil pourrait avoir une limitation en bande passante ou pourrait seulement être capable d'envoyer le "mode=3". Donc, les applications qui prennent en charge tous les modes de décodage DEVRAIENT inclure "mode=any" comme montré dans l'exemple ci-dessous :

```
m=audio 8088 RTP/AVP 97
a=rtpmap:97 speex/8000
a=fmtp:97 mode="4,any"
```

5.2 Exemple de prise en charge des seuls modes 3 et 5

L'offreur indique le mode qu'il souhaite recevoir (Speex "mode 3"). Cette offre indique que le mode 3 et le mode 5 sont pris en charge et qu'aucun autre mode n'est pris en charge. La partie distante NE DOIT PAS configurer son codeur en utilisant un autre mode Speex.

```
m=audio 8088 RTP/AVP 97
a=rtpmap:97 speex/8000
a=fmtp:97 mode="3,5"
```

5.3 Exemple avec débit binaire variable et bruit de confort

L'offreur indique qu'il souhaite recevoir des trames de débit binaire variable avec bruit de confort :

```
m=audio 8088 RTP/AVP 97
a=rtpmap:97 speex/8000
a=fmtp:97 vbr=on;cng=on
```

5.4 Exemple avec détection d'activité vocale

L'offreur indique qu'il souhaite utiliser la suppression de silence. Dans ce cas, le paramètre vbr=vad va être utilisé :

```
m=audio 8088 RTP/AVP 97
a=rtpmap:97 speex/8000
a=fmtp:97 vbr=vad
```

5.5 Exemple avec plusieurs taux d'échantillonnage

L'offreur indique qu'il souhaite recevoir de l'audio Speex à 16000 Hz avec le mode 10 (42,2 kbit/s) ou, autrement, de l'audio Speex à 8000 Hz avec le mode 7 (24,6 kbit/s). L'offreur prend en charge le décodage de tous les modes.

```
m=audio 8088 RTP/AVP 97 98
a=rtpmap:97 speex/16000
a=fmtp:97 mode="10,any"
a=rtpmap:98 speex/8000
a=fmtp:98 mode="7,any"
```

5.6 Exemple avec Ptime et plusieurs trames Speex

L'attribut SDP "ptime" est utilisé pour noter l'intervalle de mise en paquets (c'est-à-dire, combien de millisecondes d'audio sont codées dans un seul paquet RTP). Comme Speex utilise des trames de 20 ms, les valeurs de ptime de multiples de 20 notent plusieurs trames Speex par paquet. Il est recommandé d'utiliser des valeurs de ptime qui sont un multiple de 20.

Si ptime contient une valeur qui n'est pas un multiple de 20, son interprétation interne devrait être arrondie au plus proche multiple de 20 avant de calculer le nombre de trames Speex. Par exemple, si l'attribut "ptime" est réglé à 30, l'interprétation interne devrait être arrondie à 40 et ensuite utilisée pour calculer deux trames Speex par paquet.

Dans l'exemple ci-dessous, la valeur de ptime est réglée à 40, indiquant qu'il y a deux trames dans chaque paquet.

```
m=audio 8088 RTP/AVP 97
a=rtpmap:97 speex/8000
a=ptime:40
```

Noter que le paramètre ptime s'applique à toutes les charges utiles mentionnées dans la ligne media et n'est pas utilisé au titre d'une directive a=fmtp.

Il faut veiller quand on règle la valeur de ptime à ce que la taille de paquet RTP n'excède pas la MTU du chemin.

5.7 Exemple avec échange d'offre/réponse complet

L'offreur indique qu'il souhaite recevoir de l'audio Speex à 16000 Hz ou, autrement, de l'audio Speex à 8000 Hz. L'offreur ne prend pas en charge TOUS les modes parce que aucun mode n'est spécifié.

```
m=audio 8088 RTP/AVP 97 98
a=rtpmap:97 speex/16000
a=rtpmap:98 speex/8000
```

Le répondeur indique qu'il souhaite recevoir de l'audio Speex à 8000 Hz, qui est le seul taux d'échantillonnage qu'il prend en charge. Le répondeur prend en charge TOUS les modes parce que aucun mode n'est spécifié.

```
m=audio 8088 RTP/AVP 99
a=rtpmap:99 speex/8000
```

6. Lignes directrices pour la mise en œuvre

Les mises en œuvre qui prennent en charge Speex sont responsables du décodage correct des trames Speex entrantes.

Chaque trame Speex contient toutes les informations nécessaires pour se décoder elle-même. En particulier, la valeurs de "mode" et de "ptime" proposées dans les contenus SDP NE DOIVENT PAS être utilisées pour le décodage : ces valeurs ne sont pas nécessaires pour décoder correctement un flux RTP Speex.

7. Considérations sur la sécurité

Les paquets RTP qui utilisent le format de charge utile défini dans la présente spécification sont soumis aux considérations sur la sécurité discutées dans la spécification de RTP [RFC3550], et tout profil RTP approprié. Cela implique que la confidentialité du flux de supports est réalisée par chiffrement. Parce que la compression de données utilisée avec ce format de charge utile est appliqué de bout en bout, le chiffrement peut être effectué après la compression afin qu'il n'y ait pas de conflit entre les deux opérations.

Une menace potentielle de déni de service existe pour les codages de données qui utilisent des techniques de compression qui ont une charge de calcul non uniforme sur l'extrémité receveuse. L'attaquant peut injecter dans le flux des datagrammes pathologiques qui sont complexes à décoder et causer la surcharge du receveur. Cependant, ce codage ne présente aucune non uniformité significative.

Comme avec tout protocole fondé sur IP, dans certaines circonstances, un receveur peut être surchargé simplement par la réception de trop de paquets, désirés ou non désirés. L'authentification de couche réseau peut être utilisée pour éliminer les paquets provenant de sources non désirées, mais le coût de traitement de l'authentification elle-même peut être trop élevé.

8. Remerciements

Les auteurs tiennent à remercier Equivalence Pty Ltd d'Australie de son assistance pour tenter de normaliser l'utilisation de Speex dans les applications H.323, et pour mettre en œuvre Speex dans leur pile open-source OpenH323. Les auteurs tiennent aussi à remercier Brian C. Wiles <brian@streamcomm.com> de StreamComm de son assistance au développement de la norme proposée pour l'utilisation de Speex dans les applications H.323.

Les auteurs voudraient aussi remercier les membres des communautés Speex et AVT de leurs apports : Ross Finlayson, Federico Montesino Pouzols, Henning Schulzrinne, Magnus Westerlund, Colin Perkins, et Ivo Emanuel Goncalves.

Merci aux anciens auteurs du présent document, Simon Morlat, Roger Hardiman, et Phil Kerr.

9. Références

9.1 Références normatives

- [RFC2119] S. Bradner, "[Mots clés à utiliser](#) dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997. DOI 10.17487/RFC2119, (MàJ par [RFC8174](#))
- [RFC3550] H. Schulzrinne, S. Casner, R. Frederick et V. Jacobson, "[RTP : un protocole de transport pour les applications en temps réel](#)", STD 64, DOI 10.17487/RFC3550, juillet 2003. (MàJ par [RFC7164](#), [RFC7160](#), [RFC8083](#), [RFC8108](#), [RFC8860](#))
- [RFC4566] M. Handley, V. Jacobson et C. Perkins, "SDP : [Protocole de description de session](#)", juillet 2006, DOI 10.17487/RFC4566. (P.S. ; remplacée par [RFC8866](#))

9.2 Références pour information

- [CELP] Schroeder, M. and B. Atal, "Code-excited linear prediction(CELP): High-quality speech at very low bit rates", Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol 10, pp. 937-940, 1985, <<http://www.ntis.gov/>>.
- [RFC4288] N. Freed et J. Klensin, "Spécifications du [type de support et procédures d'enregistrement](#)", [BCP 13](#), décembre

2005.

[SPEEX] Valin, J., "The Speex Codec Manual", <<http://www.speex.org/docs/>>.

Adresse des auteurs

Greg Herlein
Independent
2034 Filbert Street
San Francisco, California 94123
United States
mél : gherlein@herlein.com

Jean-Marc Valin
Xiph.Org Foundation
mél : jean-marc.valin@usherbrooke.ca

Alfred E. Heggstad
Creytiv.com
Biskop J. Nilssonsgt. 20a
Oslo 0659
Norway
mél : ah@db.org

Aymeric Moizard
Antisip
5 Place Benoît Crépu
Lyon, 69005
France
mél : jack@atosc.org