

Groupe de travail Réseau
Request for Comments : 4060
 Catégorie : Sur la voie de la normalisation
 Traduction Claude Brière de L'Isle

Q. Xie, Motorola
 D. Pearce, Motorola
 mai 2005

Formats de charge utile RTP pour les normes européennes ES 202 050, ES 202 211, et ES 202 212 de codage réparti de reconnaissance de la parole de l'Institut européen des normes de télécommunications (ETSI)

Statut du présent mémoire

Le présent document spécifie un protocole de l'Internet en cours de 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 Copyright

Copyright (C) The Internet Society (2005).

Résumé

Le présent document spécifie les formats de charge utile RTP pour encapsuler les flux de caractéristiques de traitement de signal des normes européennes ES 202 050 "Frontal évolué de reconnaissance répartie de la parole" (AFE, *Advanced Front-end*) ES 202 211 "Frontal étendu de reconnaissance répartie de la parole" (XFE, *Extended Front-end*) et ES 202 212 "Frontal étendu évolué de reconnaissance répartie de la parole" (XAFE, *Extended Advanced Front-end*) de l'Institut Européen des normes de télécommunications (ETSI) pour les systèmes répartis de reconnaissance de la parole (DSR, *distributed speech recognition*).

Table des matières

1. Introduction.....	1
1.1 Conventions et acronymes.....	2
2. Codecs DSR frontaux d'ETSI.....	2
2.1 Codec DSR frontal évolué de l'ES 202 050.....	2
2.2 Codec DSR frontal étendu de l'ES 202 211.....	3
2.3 Codec DSR frontal étendu évolué de l'ES 202 212.....	3
3. Formats de charge utile de DSR RTP.....	3
3.1 Considérations communes aux trois formats de charge utile de DSR RTP.....	3
3.2 Format de charge utile pour l'ES 202 050 DSR.....	4
3.3 Format de charge utile pour l'ES 202 211 DSR.....	5
3.4 Format de charge utile pour DSR de l'ES 202 212.....	7
4. Considérations relatives à l'IANA.....	9
4.1 Transposition des paramètres MIME en SDP.....	9
4.2 Usage en offre/réponse.....	10
4.3 Contrôle encombrement.....	10
5. Considérations sur la sécurité.....	10
6. Remerciements.....	10
7. Références.....	10
7.1 Références normatives.....	10
7.2 Références pour information.....	11
Adresse des auteurs.....	11
Déclaration complète de droits de reproduction.....	11

1. Introduction

La technologie de reconnaissance répartie de la parole (DSR, *Distributed speech recognition*) est destinée à un appareil distant agissant comme client mince (autrement dit, le frontal) pour communiquer avec un serveur de reconnaissance vocale (autrement dit un moteur vocal) sur une connexion réseau pour obtenir des services de reconnaissance vocale. On trouvera plus de détails sur la DSR dans l'Internet dans la [RFC3557].

Pour réaliser l'interopérabilité entre les différents appareils clients et moteurs vocaux, la première norme d'ETSI sur le frontal DSR, l'ES 201 108 a été publiée début 2000 [ES201108]. Une mise en paquet RTP pour les trames conformes à l'ES 201 108 est définie dans la [RFC3557] par l'IETF.

Dans [ES202050], ETSI produit une autre norme pour un frontal DSR évolué qui fournit des performances de reconnaissance substantiellement améliorées quand du bruit de fond est présent. Les codecs dans l'ES 202 050 utilisent un format de trame légèrement différent de celui de l'ES 201 108 et donc les deux n'inter-opèrent pas.

La mise en paquet RTP pour le frontal conforme à l'ES 202 050 défini dans le présent document utilise la même présentation de format de paquet RTP que celle définie dans la [RFC3557]. Les différences sont dans la définition binaire de la trame de codec de DSR et l'enregistrement du type MIME de charge utile.

Les deux normes suivantes, ES 202 211 et ES 202 212, fournissent des extensions à chacune des normes de frontal DSR. Les extensions permettent de reconstruire l'onde vocale pour l'audition humaine et peuvent aussi être utilisées pour améliorer les performances de reconnaissance pour les tonalités de langages. Ceci est fait par l'envoi d'informations tonales et vocales supplémentaires pour chaque trame avec les caractéristiques de reconnaissance.

Le format de paquet RTP pour ces normes étendues est aussi défini dans le présent document.

On notera que les performances de la plupart des appareils de reconnaissance sont extrêmement sensibles aux pertes de trame consécutives et les appareils de reconnaissance DSR n'y font pas exception. Si on s'attend à ce qu'une session de DSR sur RTP subisse un fort taux de perte de paquet entre le frontal et le moteur vocal, on devrait envisager de limiter le nombre maximum de trames DSR autorisées dans un paquet, ou d'employer d'autres techniques de gestion des pertes, comme la FEC ou l'entrelacement, pour minimiser les chances de perte de trames consécutives.

1.1 Conventions et acronymes

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 le BCP 14, [RFC2119].

Les acronymes suivants sont utilisés dans le présent document :

DSR (*Distributed Speech Recognition*) : reconnaissance répartie de la parole

ETSI (*European Telecommunications Standards Institute*) : Institut européen des normes de télécommunications

FP (*Frame Pair*) : paire de trames

DTX (*Discontinuous Transmission*) : transmission discontinue

VAD (*Voice Activity Detection*) : détection d'activité vocale

2. Codecs DSR frontaux d'ETSI

Certaines caractéristiques pertinentes des codecs DSR frontaux des normes ES 202 050 évolué, ES 202 211 étendu, et ES 202 212 étendu évolué sont résumées ci-dessous.

2.1 Codec DSR frontal évolué de l'ES 202 050

Le calcul frontal est un schéma fondé sur la trame qui produit un vecteur de sortie toutes les 10 ms. Dans l'extraction de caractéristiques de frontal, la réduction de bruit est effectuée d'abord par un filtrage de Wiener. Ensuite, le traitement de forme d'onde est appliqué au signal nettoyé du bruit et les caractéristiques mel-cepstral sont calculées. À la fin, une égalisation aveugle est appliquée aux caractéristiques cepstral. L'algorithme de frontal produit à ce résultat une représentation mel-cepstral dans le même format que ES 210 108, c'est-à-dire, 12 coefficients cepstral [C1 - C12], C0 et log énergie. La détection d'activité vocale (VAD, *Voice activity detection*) pour le classement de chaque trame comme parole ou non parole est aussi mise en œuvre dans l'extraction de caractéristiques. Les informations de VAD sont incluses dans le format de charge utile pour chaque paire de trames à envoyer au moteur de reconnaissance distant au titre de la charge utile. Ces informations peuvent facultativement être utilisées par le moteur de reconnaissance receveur pour éliminer les trames qui ne sont pas de parole. Le frontal prend en charge trois taux d'échantillonnage bruts : 8 kHz, 11 kHz, et 16 kHz (Noter qu'à la différence de certains autres codecs vocaux, la taille de trame caractéristique de DSR présentée à la mise en paquet RTP ne dépend pas du nombre d'échantillons de parole utilisés dans chaque trame d'échantillons de 10 ms. Cela va devenir plus évident dans les paragraphes suivants).

Après le calcul de la représentation mel-cepstral, la représentation est d'abord quantifiée via la quantification de vecteur partagé pour réduire le taux de données du flux codé. Ensuite, les vecteurs quantifiés provenant de deux trames consécutives sont mis dans une FP, comme décrit plus en détails au paragraphe 3.2.

2.2 Codec DSR frontal étendu de l'ES 202 211

Certaines caractéristiques pertinentes des codecs DSR frontaux étendus de la norme ES 202 211 sont résumées ci-dessous.

ES 202 211 est une extension de la norme ES 201 108 de frontal DSR mel-cepstrum [ES201108]. Le frontal mel-cepstrum fournit les caractéristiques pour la reconnaissance de la parole mais elles ne sont pas disponibles pour l'écoute humaine. L'objet de l'extension est de permettre la reconstruction de la forme d'onde de la parole à partir de ces caractéristiques afin qu'elle puisse être répétée. La partie extraction de caractéristiques de frontal du traitement est exactement la même que pour ES 201 108. Pour permettre la reconstruction de la parole des informations supplémentaires sur les fréquences fondamentales (perçues comme un intervalle) et la classe vocale (par exemple, non parole, vocal, non vocal et mixte) sont nécessaires. Ces informations supplémentaires sont fournies par les algorithmes de traitement de frontal étendu du côté appareil. Elles sont compressées et transmises au serveur avec les caractéristiques du frontal. Ces informations supplémentaires peuvent aussi être utiles pour améliorer les performances de reconnaissance de la parole avec des langues tonales comme le mandarin, le cantonais et le thaï.

Les pleines informations sur les algorithmes de traitement du signal côté client utilisés dans la norme sont décrites dans la norme [ES202211].

Les informations supplémentaires de fréquence fondamentale et de classe vocale sont compressées pour chaque paire de trame. L'écartement pour la première trame de la FP est quantifié sur 7 bits et la seconde trame est quantifiée de façon différentielle sur 7 bits. La classe vocale est indiquée par un bit pour chaque trame. Le total des informations d'extension pour une paire de trames consiste donc en 14 bits plus deux bits supplémentaires de CRC de protection d'erreur calculé seulement sur ces bits d'extension.

Les informations totales pour la paire de trames sont constituées de 92 bits pour les deux trames compressées de caractéristiques de frontal (incluant 4 bits pour leur CRC) plus 16 bits pour l'extension (incluant 2 bits pour leur CRC) et 4 bits de bourrage nul pour donner un total de 14 octets par paire de trames. Comme pour ES 201 208, la paire de trame étendue correspond aussi à 20 ms de parole. Le frontal étendu prend en charge trois taux d'échantillonnage brut : 8 kHz, 11 kHz, et 16 kHz.

Les vecteurs quantifiés provenant de deux trames consécutives sont mis dans une FP, comme décrit plus en détails au paragraphe 3.3.

Les paramètres reçus par le serveur distant de la charge utile DSR étendue de RTP spécifiée ici peuvent être utilisés pour synthétiser une forme d'onde de parole intelligible pour la réponse. Les algorithmes pour le faire sont décrits dans la spécification [ES202211].

2.3 Codec DSR frontal étendu évolué de l'ES 202 212

ES 202 212 est l'extension pour le frontal DSR évolué de la norme [ES202050]. Il fournit les mêmes capacités que le frontal mel-cepstrum étendu décrit au paragraphe 2.2 mais pour le frontal DSR évolué.

3. Formats de charge utile de DSR RTP

3.1 Considérations communes aux trois formats de charge utile de DSR RTP

Les trois formats de charge utile RTP de DSR définis dans ce document partagent les considérations ou comportements suivants.

3.1.1 Nombre de FP dans chaque paquet RTP

Tout nombre de FP PEUT être agrégé dans une charge utile RTP et elles DOIVENT être consécutives dans le temps. Cependant, la charge utile RTP DEVRAIT toujours rester plus petite que la MTU afin d'éviter la fragmentation IP et

DEVRAIT suivre les recommandations du paragraphe 3.1 de la [RFC3557] lors de la détermination du nombre approprié de FP dans une charge utile RTP.

3.1.2 Prise en charge de la transmission discontinue

Les considérations décrites au paragraphe 3.2 de la [RFC3557] s'appliquent aux trois charges utiles DSR RTP définies dans le présent document.

3.1.3 Usage de l'en-tête RTP

Le format de l'en-tête RTP est spécifié dans la [RFC3550]. Les trois formats de charge utile définis ici utilisent les champs de l'en-tête en cohérence avec la présente spécification.

L'horodatage RTP correspond à l'instant d'échantillonnage du premier échantillon codé pour la première FP dans le paquet. La fréquence de l'horloge d'horodatage est la même que la fréquence d'échantillonnage, de sorte que l'unité d'horodatage est en échantillons.

Comme défini par les trois codecs frontaux, la durée d'une FP est 20 ms, correspondant à 160, 220, ou 320 échantillons codés avec un taux d'échantillonnage de 8, 11, ou 16 kHz utilisé respectivement au frontal. Donc, l'horodatage est augmenté, respectivement de 160, 220, ou 320 pour chaque FP consécutive.

La charge utile DSR pour les trois codecs frontaux est toujours un nombre entier d'octets. Si du bourrage supplémentaire est requis pour un autre objet, le bit P peut alors être établi dans l'en-tête RTP et le bourrage ajouté comme spécifié dans la [RFC3550].

Le bit marqueur (M) d'en-tête RTP DOIT être réglé selon les règles générales des codecs audio, comme défini au paragraphe 4.1 de la [RFC3551].

Le présent document ne spécifie pas d'allocation d'un type de charge utile RTP pour ces trois nouveaux formats de paquet. Il est prévu que le profil RTP sous lequel un de ces formats de charge utile va être utilisé pour allouer un type de charge utile pour ce codage, ou va spécifier le type de charge utile à allouer de façon dynamique.

3.2 Format de charge utile pour l'ES 202 050 DSR

Un datagramme de charge utile RTP de DSR ES 202 050 utilise exactement la même disposition que définie à la Section 3 de la [RFC3557], c'est-à-dire, un en-tête RTP standard suivi par une charge utile DSR contenant une série de FP de DSR.

La taille de chaque FP ES 202 050 FP reste de 96 bits ou 12 octets, comme défini dans les paragraphes suivants. Cela assure qu'une charge utile de DSR RTP va toujours se terminer sur une limite d'octet.

3.2.1 Formats de paire de trame

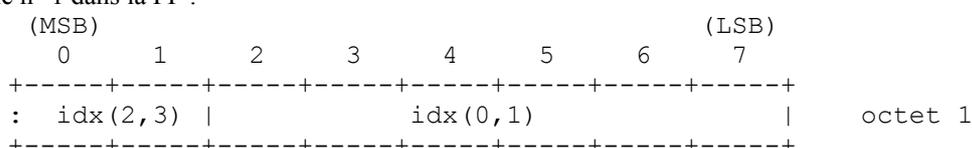
3.2.1.1 Format des FP de parole et non parole

La trame suivante de mel-cepstral DOIT être utilisée, comme défini dans [ES202050]:

Les paires de trames quantifiées de 10 ms de mel-cepstral DOIVENT être groupées et protégées par un CRC de 4 bits formant une FP de 92 bits. À la fin, chaque FP DOIT être bourrée avec 4 zéros sur les quatre bits de poids fort (*MSB*) du dernier octet afin d'aligner la FP sur la limite d'octet.

Le diagramme suivant montre une trame complète ES 202 050 :

Trame n° 1 dans la FP :



```

:      idx(4,5)      |      idx(2,3) (suite) :      octet 2
+-----+-----+-----+-----+-----+
|      idx(6,7)      |      idx(4,5) (suite) |      octet 3
+-----+-----+-----+-----+-----+
idx(10,11) | VAD |      idx(8,9)      |      octet 4
+-----+-----+-----+-----+-----+
:      idx(12,13)    |      idx(10,11) (suite) :      octet 5
+-----+-----+-----+-----+-----+
|      idx(12,13)    |      (suite) :      octet 6/1
+-----+-----+-----+-----+-----+

```

Trame n° 2 dans la FP :

```

(MSB)                                (LSB)
  0      1      2      3      4      5      6      7
+-----+-----+-----+-----+-----+
:      idx(0,1)      |      octet 6/2
+-----+-----+-----+-----+-----+
|      idx(2,3)      |      idx(0,1) (suite) |      octet 7
+-----+-----+-----+-----+-----+
:  idx(6,7) |      idx(4,5)      |      octet 8
+-----+-----+-----+-----+-----+
:      idx(8,9)      |      idx(6,7) (suite) :      octet 9
+-----+-----+-----+-----+-----+
|      idx(10,11)    |      VAD |      idx(8,9) (suite) |      octet 10
+-----+-----+-----+-----+-----+
|      idx(12,13)    |      octet 11
+-----+-----+-----+-----+-----+

```

CRC pour la trame n° 1 et la trame n° 2 et bourrage dans la FP :

```

(MSB)                                (LSB)
  0      1      2      3      4      5      6      7
+-----+-----+-----+-----+-----+
|  0  |  0  |  0  |  0  |      CRC      |      octet 12
+-----+-----+-----+-----+-----+

```

Le CRC de 4 bits dans la FP DOIT être calculé en utilisant la formule (incluant les règles d'ordre des bits) définie au paragraphe 7.2 de [ES202050].

Donc, chaque FP représente 20 ms de parole d'origine. Noter que chaque FP DOIT être bourrée avec 4 zéros aux 4 bits de poids fort du dernier octet afin d'aligner la FP sur une limite d'octet, comme montré ci-dessus. Cela fait une taille totale de 96 bits pour la FP, soit 12 octets. Noter que ce bourrage est distinct du bourrage indiqué par le bit P dans l'en-tête RTP.

La définition des indices et du fanion "VAD" est décrite dans [ES202050] et leur valeur est réglée et examinée seulement par les codecs dans le client de frontal et le reconnaisseur.

3.2.1.2 Format de FP nulle

Des FP nulles sont envoyées pour marquer la fin d'un segment de transmission. Les détails sur le segment de transmission et l'utilisation des FP nulles se trouvent dans la [RFC3557].

Une FP nulle pour le codec frontal ES 202 050 est définie en réglant le contenu des deux premières trames de la FP à nul c'est-à-dire, en remplissant de zéros les 88 premiers de la FP). Le CRC de 4 bits DOIT être calculé de la même façon que décrit au paragraphe 7.2.4 de [ES202050], et 4 zéros DOIVENT être bourrés à la fin de la FP Nulle afin de l'aligner sur la limite d'octet.

3.3 Format de charge utile pour l'ES 202 211 DSR

Un datagramme de charge utile RTP de DSR ES 202 211 est très similaire à celui défini à la Section 3 de la [RFC3557], c'est-à-dire, un en-tête RTP standard suivi par une charge utile DSR contenant une série de FP de DSR.

La taille de chaque FP ES 202 211 est de 112 bits ou 14 octets, comme défini dans les paragraphes suivants. Cela assure qu'une charge utile RTP de DSR va toujours se terminer sur une limite d'octet.

3.3.1 Formats de paire de trames

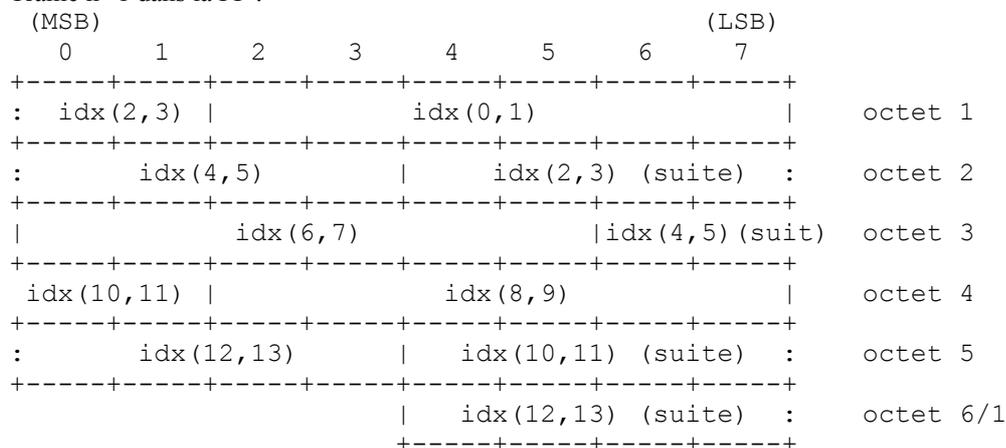
3.3.1.1 Format des FP de parole et non parole

La trame mel-cepstral suivante DOIT être utilisée, comme défini au paragraphe 6.2.4 de [ES202211]:

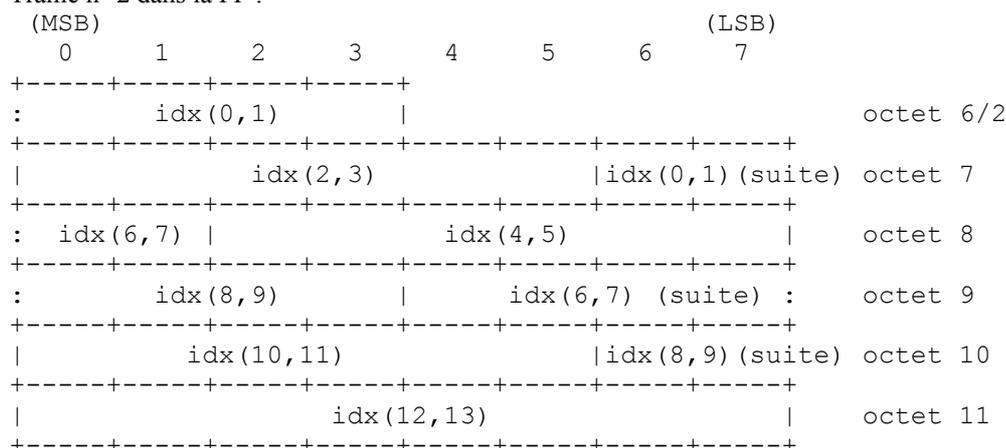
Suivent immédiatement deux trames (trame n°1 et trame n°2) d'indices de codets (soit 88 bits), puis un CRC de 4 bits calculé sur ces 88 bits. Suit ensuite les indices d'écartement (*pitch*) de la première trame (Pidx1 : 7 bits) et de la seconde trame (Pidx2 : 5 bits) de la paire de trames. Les indices de classe des deux trames dans la paire de trames de 1 bit chacun (Cidx1 et Cidx2) suivent ensuite. Finalement, un CRC de 2 bits calculé sur les bits d'écartement et de classe (total : 14 bits) de la paire de trames est inclus (PC-CRC). Le nombre total de bits dans un paquet de paire de trames est donc $44 + 44 + 4 + 7 + 5 + 1 + 1 + 2 = 108$. À la fin, chaque FP DOIT être bourrée de 4 zéros sur les 4 bits de poids fort du dernier octet afin d'aligner la FP sur la limite d'octet.

Le diagramme suivant montre une FP ES 202 211 complète :

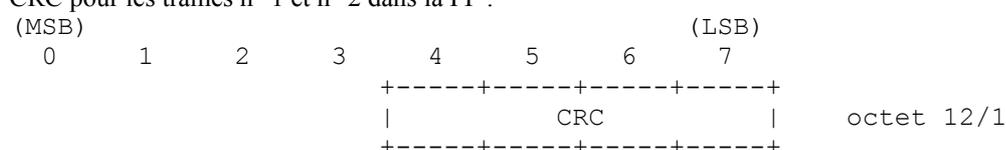
Trame n° 1 dans la FP :



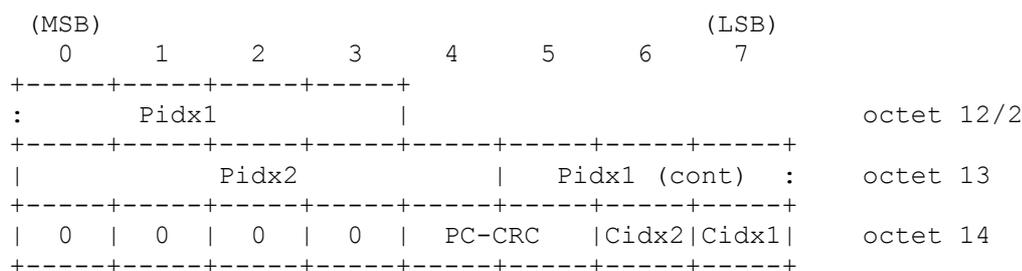
Trame n° 2 dans la FP :



CRC pour les trames n° 1 et n° 2 dans la FP :



Informations d'extension et bourrage dans la FP :



Le CRC de 4 bits et le PC-CRC de 2 bits dans la FP DOIVENT être calculés en utilisant la formule (incluant les règles d'ordre des bits) définies au paragraphe 6.2.4 de [ES202211].

Donc, chaque FP représente 20 ms de parole originale. Noter, comme montré ci-dessus, que chaque FP DOIT être bourrée avec 4 zéros sur les 4 bits de MSB (*poinds fort*) du dernier octet afin d'aligner la FP sur la limite d'octet. Cela fait une taille totale de FP de 112 bits, soit 14 octets. Noter que ce bourrage est distinct du bourrage indiqué par le bit P dans l'en-tête RTP.

3.3.1.2 Format de FP Nulle

Une FP Nulle pour le codec frontal ES 202 211 est définie en réglant tous les 112 bits de la FP à zéro. Les FP Nulles sont envoyées pour marquer la fin d'un segment de transmission. Les détails sur le segment de transmission et l'utilisation des FP Nulles se trouvent dans la [RFC3557].

3.4 Format de charge utile pour DSR de l'ES 202 212

Similaire aux autres schémas de codage de frontal DSR d'ETSI, le flux codé de caractéristiques de DSR de l'ES 202 212 est transmis dans une séquence de FP, où chaque FP représente deux trames consécutives de voix originale.

Un datagramme de charge utile RTP de DSR ES 202 212 est très similaire à celui défini à la Section 3 de la [RFC3557], c'est-à-dire, un en-tête RTP standard suivi par une charge utile DSR contenant une série de FP DSR.

La taille de chaque FP ES 202 212 est de 112 bits, soit 14 octets, comme défini dans les paragraphes qui suivent. Cela assure qu'une charge utile de DSR ES 202 212 va toujours se terminer sur une limite d'octet.

3.4.1 Formats de paire de trames

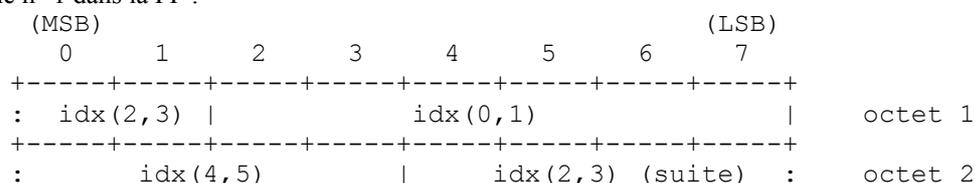
3.4.1.1 Format de FP de parole et non parole

La trame mel-cepstral suivante DOIT être utilisée, comme défini au paragraphe 7.2.4 de [ES202212] :

Immédiatement à la suite de deux trames (trame n° 1 et trame n° 2) d'indices de codets (soit 88 bits), il y a un CRC de 4 bits calculé sur ces 88 bits. Les indices d'écartement de la première trame (Pidx1 : 7 bits) et de la seconde trame (Pidx2 : 5 bits) de la paire de trames suivent. Les indices de classe des deux trames dans la paire de trames de 1 bit chacun suivent ensuite (Cidx1 et Cidx2). Finalement, un CRC de 2 bits (PC-CRC) calculé sur les bits d'écartement et de classe (total : 14 bits) de la paire de trames est inclus. Le nombre total de bits dans un paquet de paire de trames est donc de $44 + 44 + 4 + 7 + 5 + 1 + 1 + 2 = 108$. À la fin, chaque FP DOIT être bourrée de 4 zéros aux 4 bits de poids fort du dernier octet afin d'aligner la FP sur la limite d'octet. Le bourrage amène la taille totale d'une FP à 112 bits, soit 14 octets. Noter que ce bourrage est distinct du bourrage indiqué par le bit P dans l'en-tête RTP.

Le diagramme suivant montre une FP ES 202 212 complète :

Trame n° 1 dans la FP :



```

+-----+-----+-----+-----+-----+-----+-----+-----+
|           idx(6,7)           |idx(4,5) (suite)  octet 3
+-----+-----+-----+-----+-----+-----+-----+-----+
idx(10,11) | VAD |           idx(8,9)           |           octet 4
+-----+-----+-----+-----+-----+-----+-----+-----+
:           idx(12,13)           |   idx(10,11) (suite)   :           octet 5
+-----+-----+-----+-----+-----+-----+-----+-----+
|           idx(12,13) (suite)   :           octet 6/1
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Trame n° 2 dans la FP :

```

(MSB)                                     (LSB)
 0      1      2      3      4      5      6      7
+-----+-----+-----+-----+-----+-----+-----+-----+
:           idx(0,1)           |                               octet 6/2
+-----+-----+-----+-----+-----+-----+-----+-----+
|           idx(2,3)           |idx(0,1) (suite)  octet 7
+-----+-----+-----+-----+-----+-----+-----+-----+
:   idx(6,7) |           idx(4,5)           |           octet 8
+-----+-----+-----+-----+-----+-----+-----+-----+
:           idx(8,9)           |   idx(6,7) (suite)   :           octet 9
+-----+-----+-----+-----+-----+-----+-----+-----+
|           idx(10,11)         | VAD |idx(8,9) (suite) octet 10
+-----+-----+-----+-----+-----+-----+-----+-----+
|           idx(12,13)         |                               octet 11
+-----+-----+-----+-----+-----+-----+-----+-----+

```

CRC pour les trames n° 1 et n° 2 dans la FP :

```

(MSB)                                     (LSB)
 0      1      2      3      4      5      6      7
+-----+-----+-----+-----+-----+-----+-----+-----+
|                               |           CRC           |           octet 12/1
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Information d'extension et bourrage dans la FP :

```

(MSB)                                     (LSB)
 0      1      2      3      4      5      6      7
+-----+-----+-----+-----+-----+-----+-----+-----+
:           Pidx1           |                               octet 12/2
+-----+-----+-----+-----+-----+-----+-----+-----+
|           Pidx2           |   Pidx1 (suite)   :           octet 13
+-----+-----+-----+-----+-----+-----+-----+-----+
| 0 | 0 | 0 | 0 | PC-CRC |Cidx2|Cidx1|           octet 14
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Les indices de codets, le fanionVAD, l'indice d'écartement, et l'indice de classe sont spécifiés à la Section 6 de [ES202212]. Le CRC de 4 bits et le PC-CRC de 2 bits dans la FP DOIVENT être calculés en utilisant la formule (incluant les règles d'ordre des bits) définie au paragraphe 7.2.4 de [ES202212].

3.4.1.2 Format de FP nulle

Une FP Nulle pour le codec frontal ES 202 212 est définie en réglant tous les 112 bits de la FP à zéro. Les FP Nulles sont envoyées pour marquer la fin d'un segment de transmission. Les détails sur le segment de transmission et l'utilisation des FP Nulles se trouvent dans la [RFC3557].

4. Considérations relatives à l'IANA

Pour chacun des trois codecs frontaux de DSR ETSI couverts dans le présent document, un nouvel enregistrement de sous type MIME a été enregistré par l'IANA pour le type de charge utile correspondant, comme décrit ci-dessous.

Nom de type de support : audio

Nom de sous type de support :

dsr-es202050 (pour le frontal ES 202 050)

dsr-es202211 (pour le frontal ES 202 211)

dsr-es202212 (pour le frontal ES 202 212)

Paramètres exigés : aucun

Paramètres facultatifs :

débit : indique le taux d'échantillonnage de la parole. Les valeurs valides incluent : 8000, 11000, et 16000. Si ce paramètre est absent, on suppose de taux d'échantillonnage de 8000.

maxptime : voir la [RFC3267]. Si ce paramètre est absent, maxptime est supposé être 80 ms.

Note : comme les performances de la plupart des appareils de reconnaissance de la parole sont extrêmement sensibles aux pertes consécutives de FP, si l'utilisateur du format de charge utile s'attend à un fort ratio de perte de paquets pour la session, il devrait explicitement choisir une valeur de maxptime pour la session qui soit plus courte que la valeur par défaut.

ptime : voir la [RFC2327].

Considérations de codage : ces types sont définis pour le transfert via RTP [RFC3550] comme décrit à la Section 3 de la RFC 4060.

Considérations de sécurité : voir la Section 5 de la RFC 4060.

Adresse de la personne & de messagerie à contacter pour plus d'informations : Qiaobing.Xie@motorola.com

Usage prévu : COMMUN. Il est prévu que de nombreuses applications VoIP (ainsi que d'applications mobiles) vont utiliser ce type.

Auteur : Qiaobing.Xie@motorola.com

Contrôleur des changements : groupe de travail IETF Transport Audio/Vidéo

4.1 Transposition des paramètres MIME en SDP

Les informations portées dans la spécification de type de support MIME ont une transposition spécifique en les champs du protocole de description de session (SDP) [RFC2327], qui est couramment utilisé pour décrire les sessions RTP. Quand SDP est utilisé pour spécifier des sessions employant le codec DSR ES 202 050, ES 202 211, ou ES 202 212, la transposition est la suivante :

- o Le type MIME ("audio") va dans SDP "m=" comme nom de support.
- o Le sous type MIME ("dsr-es202050", "dsr-es202211", ou "dsr-es202212") va dans SDP "a=rtpmap" comme nom de codage.
- o Le paramètre facultatif "taux" va aussi dans "a=rtpmap" comme débit d'horloge. Si aucun taux n'est donné, la valeur par défaut (c'est-à-dire, 8000) est alors utilisée dans SDP.
- o Les paramètres facultatifs "ptime" et "maxptime" vont respectivement dans les attributs SDP "a=ptime" et "a=maxptime".

Exemple d'usage du DSR ES 202 050 :

```
m=audio 49120 RTP/AVP 101
a=rtpmap:101 dsr-es202050/8000
a=maxptime:40
```

Exemple d'usage du DSR ES 202 211 :

```
m=audio 49120 RTP/AVP 101
a=rtpmap:101 dsr-es202211/8000
a=maxptime:40
```

Exemple d'usage du DSR ES 202 212 :

```
m=audio 49120 RTP/AVP 101
a=rtpmap:101 dsr-es202212/8000
a=maxptime:40
```

4.2 Usage en offre/réponse

Tous les paramètres SDP dans ces formats de charge utile sont déclaratifs, et toutes les valeurs raisonnables sont supposées être prises en charge. Donc, l'usage standard de l'offre/réponse décrit dans la [RFC3264] devrait être suivi.

4.3 Contrôle d'encombrement

Le contrôle d'encombrement pour RTP DOIT être utilisé en accord avec la [RFC3550], et dans tout profil RTP applicable, par exemple, de la [RFC3551].

5. Considérations sur la sécurité

Les mises en œuvre qui utilisent le format de charge utile défini dans la présente spécification sont sujets aux considérations de sécurité discutées dans la spécification de RTP [RFC3550] et tout profil RTP, par exemple, la [RFC3551]. Cette charge utile ne spécifie aucun service de sécurité différent.

6. Remerciements

La conception présentée ici se fonde sur celle de la [RFC3557]. Les auteurs tiennent à remercier Magnus Westerlund et d'autres pour leur relecture et leurs commentaires.

7. Références

7.1 Références normatives

- [ES202050] European Telecommunications Standards Institute (ETSI) Standard ES 202 050, "Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Advanced Front-end Feature Extraction Algorithm; Compression Algorithms", <http://pda.etsi.org/pda/>.
- [ES202211] European Telecommunications Standards Institute (ETSI) Standard ES 202 211, "Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Extended front-end feature extraction algorithm; Compression algorithms; Back-end speech reconstruction algorithm", <http://pda.etsi.org/pda/>.
- [ES202212] European Telecommunications Standards Institute (ETSI) Standard ES 202 212, "Speech Processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Extended advanced front-end feature extraction algorithm; Compression algorithms; Back-end speech reconstruction algorithm", <http://pda.etsi.org/pda/>.
- [RFC2119] S. Bradner, "[Mots clés à utiliser](#) dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997. (MàJ par [RFC8174](#))
- [RFC2327] M. Handley et V. Jacobson, "SDP : [Protocole de description de session](#)", avril 1998. (*Obsolète; voir [RFC4566](#)*)
- [RFC3264] J. Rosenberg et H. Schulzrinne, "[Modèle d'offre/réponse](#) avec le protocole de description de session (SDP)", juin 2002. (*P.S. ; MàJ par [RFC8843](#)*)
- [RFC3267] J. Sjöberg et autres, "Format de charge utile et format de mémorisation de fichier pour les codecs audio AMR et AMR-WB dans RTP", juin 2002. (*Obsolète, voir [RFC4867](#)*) (*P.S.*)
- [RFC3550] H. Schulzrinne, S. Casner, R. Frederick et V. Jacobson, "[RTP : un protocole de transport pour les applications en temps réel](#)", STD 64, juillet 2003. (MàJ par [RFC7164](#), [RFC7160](#), [RFC8083](#), [RFC8108](#), [RFC8860](#))
- [RFC3551] H. Schulzrinne et S. Casner, "[Profil RTP pour conférences audio](#) et vidéo avec contrôle minimal", STD 65, juillet 2003. (MàJ par [RFC8860](#))
- [RFC3557] Q. Xie, éd., "[Format de charge utile RTP](#) pour la norme européenne ES 201 108 "Codage réparti de reconnaissance vocale" de l'Institut européen des normes de télécommunications (ETSI)", juillet 2003. (*P.S.*)

7.2 Références pour information

[ES201108] European Telecommunications Standards Institute (ETSI) Standard ES 201 108, "Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Front-end Feature Extraction Algorithm; Compression Algorithms", <http://pda.etsi.org/pda/>.

Adresse des auteurs

Qiaobing Xie
Motorola, Inc.
1501 W. Shure Drive, 2-F9
Arlington Heights, IL 60004
US
téléphone : +1-847-632-3028
mél : qxie1@email.mot.com

David Pearce
Motorola UK Research Laboratory
Viables Industrial Estate - Jays Close
Basingstoke, HANTS RG22 4PD
UK
téléphone : +44 (0)1256 484 436
mél : bdp003@motorola.com

Déclaration complète de droits de reproduction

Copyright (C) The IETF Trust (2006).

Le présent document est soumis aux droits, licences et restrictions contenus dans le BCP 78, et à www.rfc-editor.org, et sauf pour ce qui est mentionné ci-après, les auteurs conservent tous leurs droits.

Le présent document et les informations contenues sont fournis sur une base "EN L'ÉTAT" et le contributeur, l'organisation qu'il ou elle représente ou qui le/la finance (s'il en est), la INTERNET SOCIETY et la INTERNET ENGINEERING TASK FORCE déclinent toutes garanties, exprimées ou implicites, y compris mais non limitées à toute garantie que l'utilisation des informations encloses ne viole aucun droit ou aucune garantie implicite de commercialisation ou d'aptitude à un objet particulier.

Propriété intellectuelle

L'IETF ne prend pas position sur la validité et la portée de tout droit de propriété intellectuelle ou autres droits qui pourrait être revendiqués au titre de la mise en œuvre ou l'utilisation de la technologie décrite dans le présent document ou sur la mesure dans laquelle toute licence sur de tels droits pourrait être ou n'être pas disponible ; pas plus qu'elle ne prétend avoir accompli aucun effort pour identifier de tels droits. Les informations sur les procédures de l'ISOC au sujet des droits dans les documents de l'ISOC figurent dans les BCP 78 et BCP 79.

Des copies des dépôts d'IPR faites au secrétariat de l'IETF et toutes assurances de disponibilité de licences, ou le résultat de tentatives faites pour obtenir une licence ou permission générale d'utilisation de tels droits de propriété par ceux qui mettent en œuvre ou utilisent la présente spécification peuvent être obtenues sur répertoire en ligne des IPR de l'IETF à <http://www.ietf.org/ipr>.

L'IETF invite toute partie intéressée à porter son attention sur tous copyrights, licences ou applications de licence, ou autres droits de propriété qui pourraient couvrir les technologies qui peuvent être nécessaires pour mettre en œuvre la présente norme. Prière d'adresser les informations à l'IETF à ietf-ipr@ietf.org.

Remerciement

Le financement de la fonction d'édition des RFC est fourni par l'activité de soutien administratif (IASA) de l'IETF.