

RECOMMANDATION UIT-R BS.1115-1^{*,**}**Codage audio à faible débit binaire**

(Question UIT-R 19/6)

(1994-2005)

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que des services audionumériques de haute qualité – disques compacts (CD), cassettes et radiodiffusion – sont offerts au grand public;
- b) que les caractéristiques fondamentales d'un système de radiodiffusion sonore numérique sont données dans les Recommandations UIT-R BS.774 et UIT-R BO.789;
- c) que les spécifications utilisateur en matière de systèmes de codage audio pour la radiodiffusion numérique sont énoncées dans la Recommandation UIT-R BS.1548;
- d) que la qualité audio de base et l'effet stéréophonique requis pour la radiodiffusion sonore numérique doivent être pratiquement indifférenciables de celle d'un disque compact;
- e) que les systèmes de codage à faible débit binaire utilisés dans les connexions de programme (liaisons de contribution, de distribution et de diffusion de commentaires) qui alimentent le système de radiodiffusion sonore numérique doivent avoir des caractéristiques telles qu'ils puissent fournir à l'émetteur de radiodiffusion sonore numérique des signaux audio de très grande qualité;
- f) que, dans le cas des liaisons de contribution et de distribution, cela suppose un niveau de qualité supérieure à celui d'un disque compact en raison des contraintes liées aux marges de traitement et de chaînage et (dans le cas de liaisons de contribution) à la marge de saturation ou «marge totale»;
- g) que, dans le cas des liaisons de diffusion de commentaires, cela suppose un niveau de qualité qui permette d'offrir à l'auditeur des signaux vocaux d'excellente qualité mais que, lorsque des liaisons de diffusion de commentaires acheminent des programmes musicaux, il faut s'attendre à un niveau de qualité moindre;
- h) que le codage à faible débit binaire entraîne en général un retard de plusieurs dizaines de millisecondes, ce qui peut être important pour des manœuvres d'exploitation comme la mise en œuvre de circuits d'ordre hors émission;
- j) que le Secteur des radiocommunications de l'UIT a testé un système de codage à faible débit binaire pour des signaux audio de haute qualité et a obtenu une qualité satisfaisante pour un certain nombre d'applications (voir l'Appendice 6 et la Recommandation UIT-R BS.1548);

* La Commission d'études 6 des radiocommunications a apporté des modifications rédactionnelles à cette Recommandation en 2003 conformément à la Résolution UIT-R 44.

** Cette Recommandation doit être portée à l'attention des Commissions d'études 9 et 15 de la normalisation des télécommunications, de l'Organisation internationale de normalisation/Commission électrotechnique internationale (ISO/CEI) (JTC 1/SC 29/WG11), de l'Union européenne de radio-télévision (UER) et de l'AES (Audio Engineering Society).

k) que les méthodes objectives classiques (qui permettent, par exemple, de mesurer le rapport signal/bruit et la distorsion) risquent de ne plus convenir pour évaluer la qualité des systèmes à schémas de codage audio à faible débit binaire et que la méthode de mesure objective perceptuelle énoncée dans la Recommandation UIT-R BS.1387 pourrait finalement compléter ou remplacer les méthodes objectives classiques d'évaluation et compléter les méthodes d'évaluation subjectives,

recommande

1 d'utiliser, pour les applications de radiodiffusion sonore numérique énumérées aux points 4 à 6 du *recommande*, le codage audio à faible débit binaire décrit dans les Normes ISO/CEI 11172-3:1993 (voir la Note 1), ISO/CEI 13818-7:2003 ou AC-3, conformément à l'Annexe 2 de la Recommandation UIT-R BS.1196 (une brève description de ces normes est donnée dans les Appendices 1, 3 et 4);

NOTE 1 – Le codec ISO/CEI 11172-3 peut parfois être désigné comme le codec 13818-3 étant donné que ce dernier se réfère à la Norme 11172-3.

2 d'utiliser une fréquence d'échantillonnage de 48 kHz ou de 32 kHz pour toutes les applications;

3 de ne pas appliquer de préaccentuation au signal d'entrée du codeur audio à faible débit binaire, le codeur ne devant pas, lui aussi, en introduire;

4 d'utiliser, pour des émissions de radiodiffusion sonore numérique (voir la Note 1), les codecs suivants:

- le système de codage de Couche II de l'ISO/CEI 11172-3, avec un débit binaire de 128 kbit/s pour un signal mono et de 256 kbit/s (c'est-à-dire 2×128 kbit/s) pour un signal stéréo et un codage indépendant de la voie gauche et de la voie droite du signal stéréo (voir la Note 2);
- le système de codage ISO/CEI 13818-7, avec un débit binaire d'au moins 144 kbit/s pour un signal stéréo;
- le système de codage AC-3, avec un débit binaire d'au moins 192 kbit/s pour un signal stéréo.

NOTE 1 – L'Appendice 1 à l'Annexe 2 de la Recommandation UIT-R BS.1548 contient des informations sur les systèmes de codage qui se sont révélés être conformes aux prescriptions de qualité et autres pour les applications de diffusion.

NOTE 2 – On prévoit que les radiodiffuseurs souhaiteront disposer, dans le signal émis, d'une certaine capacité pour des données auxiliaires. Il faut toutefois qu'ils soient conscients que la réduction du débit binaire audio affecte en général la qualité audio.

5 d'utiliser, pour les liaisons de distribution et de contribution, le système de codage de Couche II de l'ISO/CEI 11172-3, avec un débit binaire d'au moins 180 kbit/s par signal audio (c'est-à-dire par signal mono ou par composante d'un signal stéréo dont les deux voies ont été codées indépendamment), à l'exclusion des données auxiliaires (voir l'Appendice 2) (voir la Note 2); mais dans le cas particulier d'une liaison de distribution unique sans raccordement en cascade ultérieur (voir la Note 3), un débit binaire d'au moins 120 kbit/s par signal audio peut être utilisé;

NOTE 1 – L'Appendice 1 à l'Annexe 1 de la Recommandation UIT-R BS.1548 contient des informations sur les systèmes de codage qui se sont révélés être conformes aux prescriptions de qualité et autres pour les applications de contribution et de distribution.

NOTE 2 – Lorsqu'on appliquera la présente Recommandation, il faudra tenir compte du fait que dans certains réseaux de télécommunication, il n'est pas possible d'utiliser sans restriction des voies à 64 kbit/s.

NOTE 3 – Lorsque, par exemple, un émetteur analogique est alimenté par une liaison de distribution unique ou lorsqu'un signal codé au studio selon la norme finale d'émission est envoyé à l'émetteur sans codage ou décodage ultérieur.

6 d'utiliser, pour les liaisons de diffusion de commentaires, le système de codage de Couche III de l'ISO/CEI 11172-3, avec un débit binaire d'au moins 60 kbit/s, à l'exclusion des données auxiliaires, pour les signaux mono (voir la Note 1) et un débit d'au moins 120 kbit/s, à l'exclusion des données auxiliaires, pour les signaux stéréo, en utilisant un codage stéréo combiné (voir l'Appendice 2) (voir la Note 2);

NOTE 1 – Selon des résultats d'essais obtenus en 1993 avec écoute au casque pour évaluer la qualité d'un codec commentaire mono Couche III à 60 kbit/s, celui-ci ne donnerait pas la qualité requise pour un signal vocal. La présente note sera maintenue dans la Recommandation tant que la qualité du codec commentaire ne satisfait pas les exigences de qualité.

NOTE 2 – Lorsqu'on appliquera la présente Recommandation, il faudra tenir compte du fait que dans certains réseaux de télécommunication, il n'est pas possible d'utiliser sans restriction des voies à 64 kbit/s.

7 de réduire au minimum le délai de mise en œuvre des codeurs et des décodeurs.

NOTE – Les Normes ISO/CEI 11172-3 (MPEG-1 audio) et 13818-7 (MPEG-2 AAC) sont disponibles sous forme électronique à l'adresse suivante: <http://www.iso.org/itu>.

Appendice 1

Description de la méthode de codage conforme à la Norme ISO/CEI 11172-3 Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s – Partie 3: Audio

1 Codage

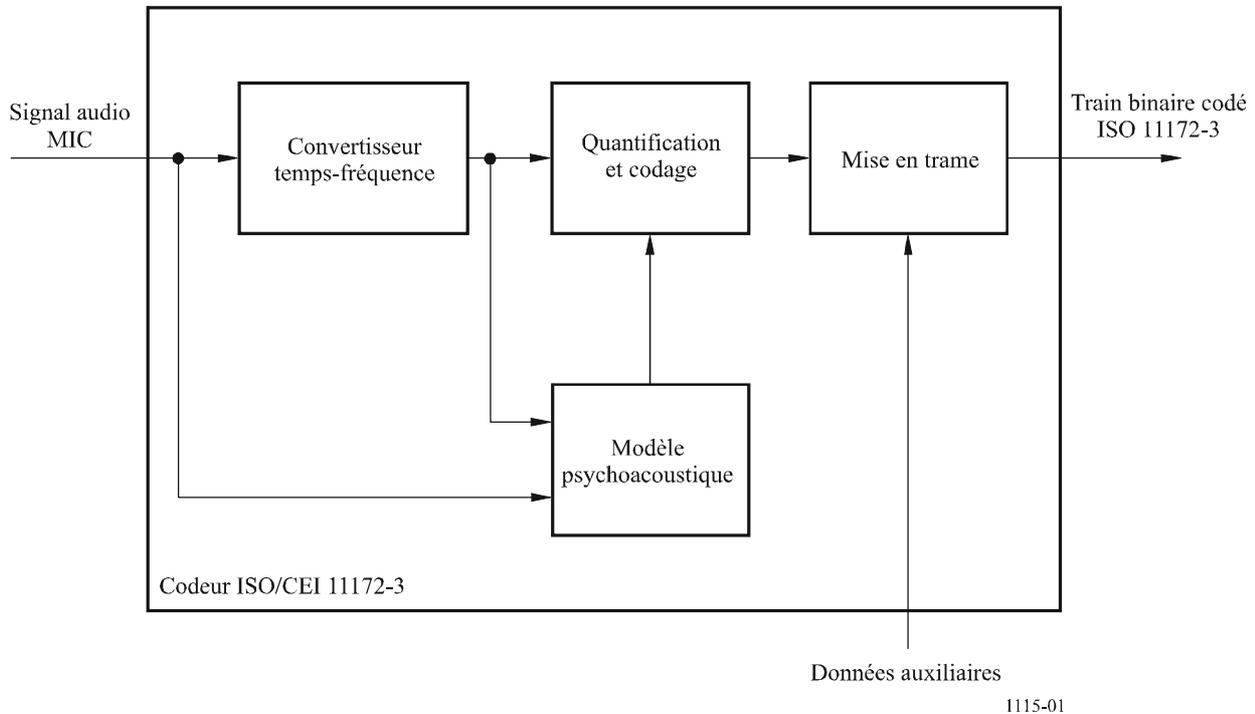
Le codeur traite le signal audionumérique et comprime le train binaire. L'algorithme de codage n'est pas normalisé et peut utiliser divers moyens pour effectuer le codage, par exemple l'estimation du seuil de masquage à l'écoute, la quantification et les facteurs d'échelle (Note 1). Toutefois, la sortie du codeur doit être telle qu'un décodeur conforme à la présente Recommandation puisse produire un signal audio convenant pour l'application souhaitée.

NOTE 1 – On obtiendra une norme minimale de qualité satisfaisante avec un codeur conforme à la description donnée dans les Annexes C et D de la Norme ISO/CEI 11172-3, 1993.

La Fig. 1 décrit un codeur type. Le codeur est alimenté par des échantillons de signaux audio. Le convertisseur temps-fréquence donne une représentation filtrée et sous-échantillonnée du train binaire audio entrant. Les échantillons convertis peuvent être appelés échantillons de sous-bande (comme dans la Couche I ou II, voir ci-après) ou échantillons de sous-bande transformés (comme dans la Couche III). Un modèle psychoacoustique, utilisant une transformée de Fourier rapide, parallèlement à la conversion temps-fréquence du signal audio, crée un ensemble de données permettant de commander la quantification et le codage. Ces données sont différentes selon les modalités concrètes de mise en œuvre du codeur. Une possibilité consiste à utiliser une évaluation du seuil de masquage pour commander le quantificateur. Le bloc de quantification et de codage génère un ensemble de symboles de codage à partir des échantillons entrants convertis. Là aussi, ce bloc dépend du système de codage. Le bloc «mise en trame» assemble le train binaire proprement

dit à partir des données de sortie des autres blocs et ajoute d'autres informations dans le champ des données auxiliaires (protection contre les erreurs par exemple), si nécessaire.

FIGURE 1
Schéma fonctionnel d'un codeur type



2 Couches

Selon l'application, on peut utiliser différentes couches du système de codage: la complexité du codeur s'accroît et sa qualité s'améliore selon les couches.

Couche I: Cette couche comprend la conversion de base des signaux audionumériques entrants en 32 sous-bandes, une segmentation fixe pour formater les données en blocs, un modèle psychoacoustique permettant de déterminer l'attribution adaptative des bits et une quantification utilisant la compression-extension et le formatage des blocs. Une trame de la Couche I représente 384 échantillons par voie.

Couche II: Cette couche assure un codage supplémentaire de l'attribution des bits et fournit des facteurs d'échelle et des échantillons. Une trame de la Couche II représente $3 \times 384 = 1\,152$ échantillons par voie.

Couche III: Cette couche introduit une résolution de fréquence plus élevée basée sur un ensemble de filtres hybride (un ensemble de 32 filtres de sous-bande avec transformation en cosinus discrète modifiée de longueur variable). Elle ajoute un quantificateur non uniforme, une segmentation adaptative et un codage entropique des valeurs quantifiées. Une trame de la Couche III représente 1 152 échantillons par voie.

Quatre modes différents sont possibles pour chacune des couches:

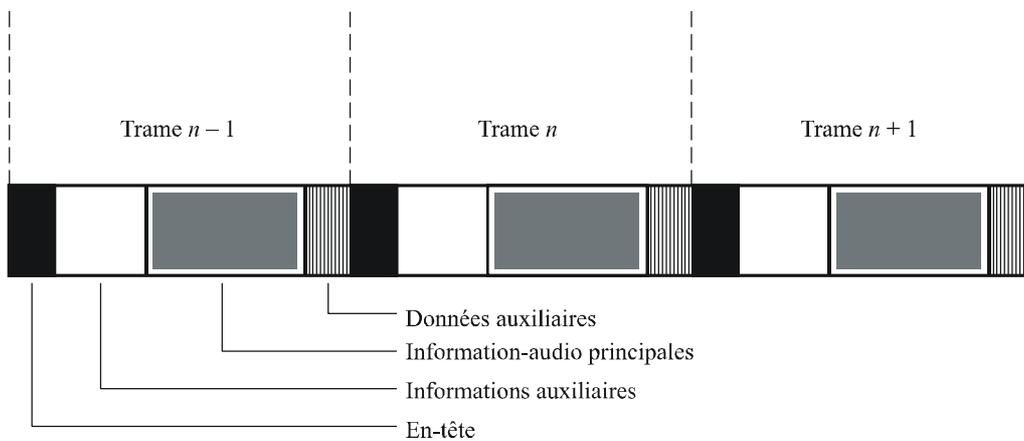
- voie unique;
- voie double (deux signaux audio indépendants codés à l'intérieur d'un seul train binaire, application bilingue par exemple);
- stéréo (les signaux gauche et droit d'une paire stéréophonique sont codés à l'intérieur d'un même train binaire);

- stéréo combiné (les signaux gauche et droit d'une paire stéréophonique sont codés dans un seul et même train binaire; la différence et la redondance stéréophoniques sont exploitées). Le mode stéréo combiné peut servir à améliorer la qualité audio aux faibles débits binaires ou à réduire le débit binaire des signaux stéréophoniques.

3 Format du train binaire codé

Les Fig. 2 et 3 donnent, respectivement pour la Couche II et la Couche III, une description schématique du train binaire ISO/CEI 11172-3. Un train de bits codé consiste en des trames successives. Selon la couche, une trame comprend les champs ci-après.

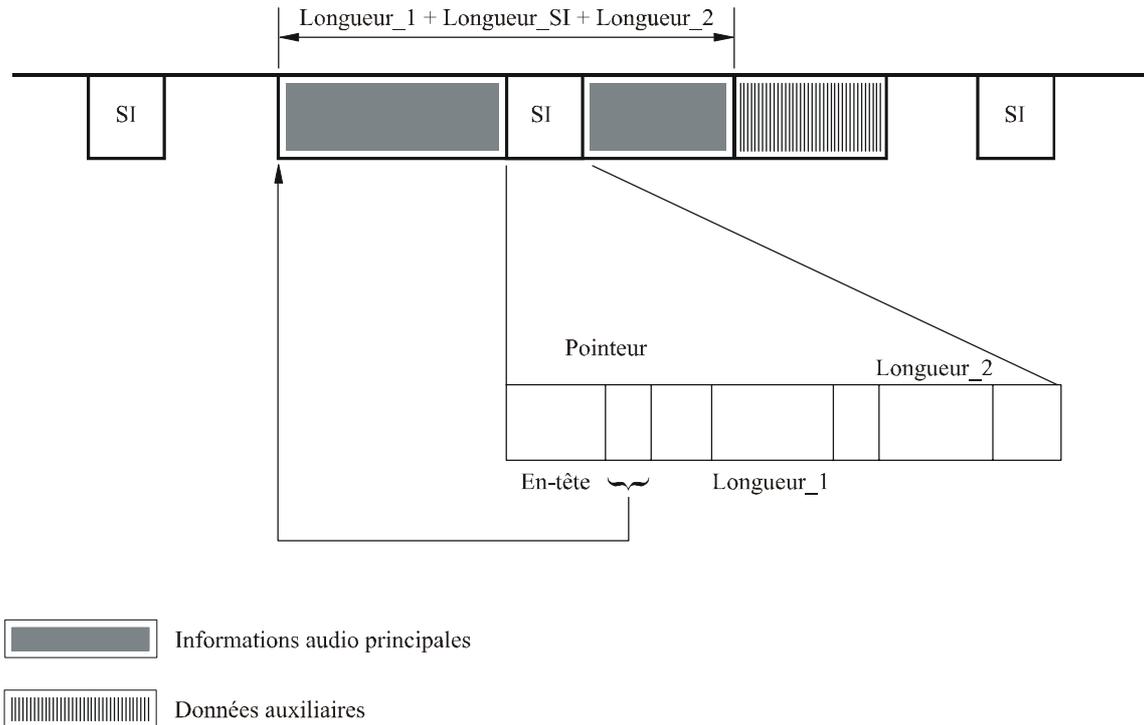
FIGURE 2
Format du train binaire de Couche II ISO/CEI 11172-3



Couche II:

En-tête:	partie du train binaire contenant les informations de synchronisation et d'état
Informations auxiliaires:	partie du train binaire contenant les informations sur l'attribution des bits et les facteurs d'échelle
Information audio principales:	partie du train binaire contenant les échantillons de sous-bande codés
Données auxiliaires:	partie du train binaire contenant des données définissables par l'utilisateur

FIGURE 3
Format du train binaire de Couche III ISO/CEI 11172-3



Couche III:

Informations auxiliaires (SI):	partie du train binaire contenant l'en-tête, le pointeur, la longueur_1 et la longueur_2, les informations sur les facteurs d'échelle, etc.
En-tête:	partie du train binaire contenant les informations de synchronisation et d'état
Pointeur:	indique le début des informations audio principales
Longueur_1:	longueur de la première partie des informations audio principales
Longueur_2:	longueur de la seconde partie des informations sur les principales données audio
Information audio principales:	partie du train binaire contenant les signaux audio codés
Données auxiliaires:	partie du train binaire contenant des données définissables par l'utilisateur

1115-03

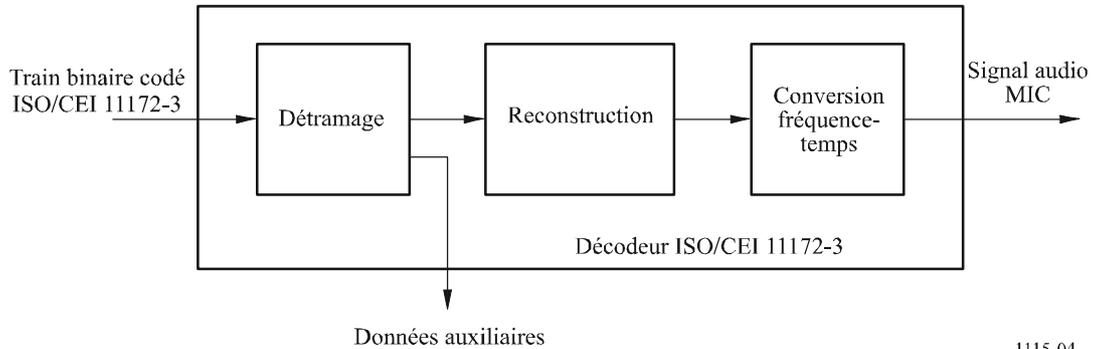
4 Décodage

Le décodeur accepte le train binaire audio comprimé dans la syntaxe définie dans la Norme ISO/CEI 11172-3, décode les éléments de données et utilise les informations pour produire le signal audionumérique de sortie.

Le décodeur est alimenté par les données relatives au train binaire audio. Le mécanisme de déramage et de décodage du train binaire effectue en option une détection d'erreur si le codeur comporte un mécanisme de correction d'erreur. Le train binaire est déramé de manière à retrouver les différents éléments d'information comme l'en-tête de trame audio, l'attribution des bits, les facteurs d'échelle, les échantillons convertis et, éventuellement, les données auxiliaires. Le

processus de reconstruction reconstruit la version quantifiée de l'ensemble des échantillons convertis. La conversion fréquence-temps retransforme ces échantillons convertis en signaux audio MIC uniformes.

FIGURE 4
Schéma fonctionnel du décodeur



1115-04

Appendice 2

Directives d'exploitation des codecs ISO/CEI 11172-3 à des débits binaires spécifiques

Le présent Appendice a pour but de donner des directives concernant des applications pour lesquelles il serait nécessaire d'exploiter un codec ISO/CEI 11172-3 à un débit binaire spécifique qui ne figure pas explicitement dans la Norme ISO/CEI (Tableau 1). Le choix de ces débits binaires spécifiques peut dépendre du débit binaire disponible sur une voie de transmission et du débit binaire minimal approprié que donne la Recommandation.

On trouvera un exposé plus complet dans la Recommandation UIT-T J.52: Transmission numérique de signaux radiophoniques de haute qualité utilisant un, deux ou trois canaux à 64 kbit/s par signal monophonique (et jusqu'à 6 par signal stéréophonique).

Les débits binaires ci-après figurent explicitement dans la Norme ISO/CEI.

TABLEAU 1

Couche II (kbit/s)	Couche III (kbit/s)
32	32
48	40
56	48
64	56
80	64
96	80
112	96
128	112
160	128
192	160
224	192
256	224
320	256
384	320

Les chiffres donnés dans le Tableau 1 indiquent le débit binaire total par programme audio indépendamment du mode. Il existe un format supplémentaire, le format libre, qui peut être utilisé pour des débits binaires fixes ne figurant pas dans le Tableau, le maximum étant de 384 kbit/s pour la Couche II et de 320 kbit/s pour la Couche III.

Une trame audio de Couche II ou de Couche III ISO/CEI 11172-3 correspond toujours à 1 152 échantillons de signaux MIC audio entrants par voie. La durée de la trame est de 24 ms pour une fréquence d'échantillonnage de 48 kHz et de 36 ms pour une fréquence d'échantillonnage de 32 kHz. Une trame contient toujours un nombre entier d'octets. On peut calculer le nombre d'octets en multipliant la durée de la trame par le débit binaire et en divisant par 8. Le débit binaire et la fréquence d'échantillonnage sont indiqués dans l'en-tête, au début de la trame audio.

On peut utiliser les trois méthodes ci-après pour obtenir un débit binaire audio qui ne figure pas dans le Tableau.

1 Format libre

L'état «format libre» doit être indiqué dans l'en-tête. Dans ce cas, le décodeur mesure au début la distance entre deux mots de synchronisation consécutifs, après quoi on peut utiliser un mécanisme de synchronisation inertielle comme dans le cas d'un débit binaire défini. Il est possible de calculer la longueur de la trame en octets à l'aide de la formule donnée ci-après. Le codeur doit tenir compte de ce chiffre.

Par exemple, si l'on veut un débit binaire de 240 kbit/s pour un programme stéréo avec une fréquence d'échantillonnage de 48 kHz, la longueur de la trame sera $0,024 \times 240\,000 : 8 = 720$ octets. Pour les débits binaires qui ne donneraient pas un nombre entier d'octets dans la trame, on peut mettre en œuvre le débit binaire voulu en procédant à un remplissage, comme cela est expliqué au § 2.4.2.3 de la Norme ISO/CEI 11172-3.

Cette méthode peut être utilisée pour les Couches II et III.

2 Utilisation du champ des données auxiliaires (pour la Couche II)

On peut choisir l'un des débits binaires figurant dans la liste (supérieur ou égal au débit binaire requis). En resserrant l'attribution des bits, le codeur peut être positionné de façon à réserver un certain nombre de bits pour les données auxiliaires de sorte que le débit binaire audio véritable est égal au débit binaire requis. Les données auxiliaires pour la Couche II sont situées à l'extrémité de chaque trame audio.

Les informations binaires sur les données auxiliaires peuvent être supprimées ou «écrasées» par réécriture avant la transmission. Si elles sont supprimées, elles sont remplacées par des bits fictifs à insérer avant décodage.

Par exemple, si l'on souhaite un débit binaire de 240 kbit/s pour un programme stéréo avec une fréquence d'échantillonnage de 48 kHz, on peut choisir un débit binaire de 256 kbit/s et l'indiquer dans l'en-tête de la trame conformément à la Norme ISO/CEI. Le codeur doit être positionné de façon à réserver 48 octets par trame pour les données auxiliaires, ce qui correspond à une capacité de données auxiliaires de 16 kbit/s.

3 Commutation de débit binaire dynamique (pour la Couche III)

La commutation dynamique de débit binaire de la Couche III peut également servir pour les débits binaires qui ne figurent pas dans le Tableau 1. Si, par exemple, on a besoin d'un débit binaire de 120 kbit/s, un codeur de Couche III peut fonctionner dans un mode où des débits binaires de 128 et 112 kbit/s explicitement définis dans la Norme ISO/CEI sont utilisés en alternance. Dans ce mode, le codeur doit être positionné sur un débit binaire qui est la moyenne des deux débits binaires alternés. La technique «tampon» (réserve de bits) de Couche III permet de maintenir un débit binaire constant pour le processus de codage. Avec cette méthode, l'intervalle binaire entre mots de synchronisation oscille entre deux valeurs.

Appendice 3

Description de la méthode de codage ISO/CEI 13818-7 Codage générique des images animées et du son associé – Partie 7: Codage du son avancé (AAC)

Voir la Norme ISO/CEI 13818-7.

1 Introduction

La Norme ISO/CEI 13818-7 décrit la norme non rétrocompatible avec la Norme MPEG-2 audio, désignée sous le nom de codage audio avancé (AAC) MPEG-2, norme applicable à la diffusion multicanal de qualité supérieure à celle que l'on obtient lorsque la rétrocompatibilité avec la Norme MPEG-1 est imposée.

Le système AAC définit trois profils permettant d'obtenir un compromis entre, d'une part, la mémoire et la puissance de traitement nécessaires et, d'autre part, la qualité audio, à savoir:

- *Profil principal*
Le profil principal offre la qualité audio la plus élevée quel que soit le débit. Tous les outils à l'exception de la commande de gain peuvent être utilisés pour fournir une qualité audio élevée. La mémoire et la puissance de traitement nécessaires sont supérieures à celles offertes par le profil faible complexité (LC, *low complexity*). Un décodeur utilisant le profil principal peut décoder un train binaire codé selon le profil LC.
- *Profil faible complexité (LC, low complexity)*
La puissance de traitement et la mémoire nécessaires selon le profil LC sont inférieures à celles selon le profil principal, mais la qualité reste élevée. Dans le profil LC, ni la prédiction ni l'outil de commande de gain ne sont autorisés et la commande de mise en forme temporelle du bruit (TNS, *temporal noise shaping*) est limitée.
- *Profil taux d'échantillonnage adaptable (SSR, scalable sampling rate)*
Le profil SSR permet d'offrir un signal adaptable en fréquence à l'aide de l'outil de commande de gain. Il permet de choisir certaines bandes de fréquences à décoder, ce qui autorise le décodeur à utiliser un nombre limité d'éléments matériels. Par exemple, pour décoder seulement la bande de fréquences la plus basse à la fréquence d'échantillonnage de 48 kHz, le décodeur peut reproduire le signal audio d'une largeur de bande de 6 kHz avec une complexité de décodage minimale.

Le système AAC prend en charge 12 types de fréquences d'échantillonnage comprises entre 8 et 96 kHz (voir le Tableau 2), et un maximum de 48 canaux audio. Le Tableau 3 présente les configurations de canaux par défaut, à savoir les configurations monocanal, bicanal, cinq canaux (trois à l'avant/deux à l'arrière) et cinq canaux plus canal à effets basses fréquences (LFE, *low-frequency effects*) (largeur de bande < 200 Hz), etc. Outre ces configurations par défaut, il est possible de spécifier le nombre de haut-parleurs à chaque emplacement (à l'avant, sur les côtés et à l'arrière), ce qui permet une disposition multicanal souple des haut-parleurs. Une fonction de conversion d'un système multicanal en un simple système stéréo («*down-mix*») est également prise en charge. L'utilisateur peut en effet désigner un coefficient pour effectuer une conversion multicanal/simple stéréo. Il est par conséquent possible de contrôler la qualité du son au moyen d'un dispositif de lecture comportant seulement deux canaux.

TABLEAU 2
Fréquences d'échantillonnage prises en charge

Fréquence d'échantillonnage (Hz)
96 000
88 200
64 000
48 000
44 100
32 000
24 000
22 050
16 000
12 000
11 025
8 000

TABLEAU 3
Configurations de canaux par défaut

Nombre de haut-parleurs	Éléments de syntaxe audio, énumérés par ordre de réception	Placement correspondant par défaut
1	single_channel_element	Haut-parleur avant central
2	channel_pair_element	Haut-parleurs avant gauche et droit
3	single_channel_element()	Haut-parleur avant central
	channel_pair_element()	Haut-parleurs avant gauche et droit
4	single_channel_element()	Haut-parleur avant central
	channel_pair_element(),	Haut-parleurs avant gauche et droit
	single_channel_element()	Haut-parleur ambiophonique arrière
5	single_channel_element()	Haut-parleur avant central
	channel_pair_element()	Haut-parleurs avant gauche et droit
	channel_pair_element()	Haut-parleurs arrière ambiophoniques gauche et droit
5+1	single_channel_element()	Haut-parleur avant central
	channel_pair_element()	Haut-parleurs avant gauche et droit
	channel_pair_element()	Haut-parleurs arrière ambiophoniques gauche et droit
	Lfe_element()	Caisson de basses
7+1	single_channel_element()	Haut-parleur avant central
	channel_pair_element(),	Haut-parleurs latéraux gauche et droit
	channel_pair_element()	Haut-parleurs avant externes gauche et droit
	channel_pair_element()	Haut-parleurs arrière ambiophoniques gauche et droit
	lfe_element()	Caisson de basses

2 Codage

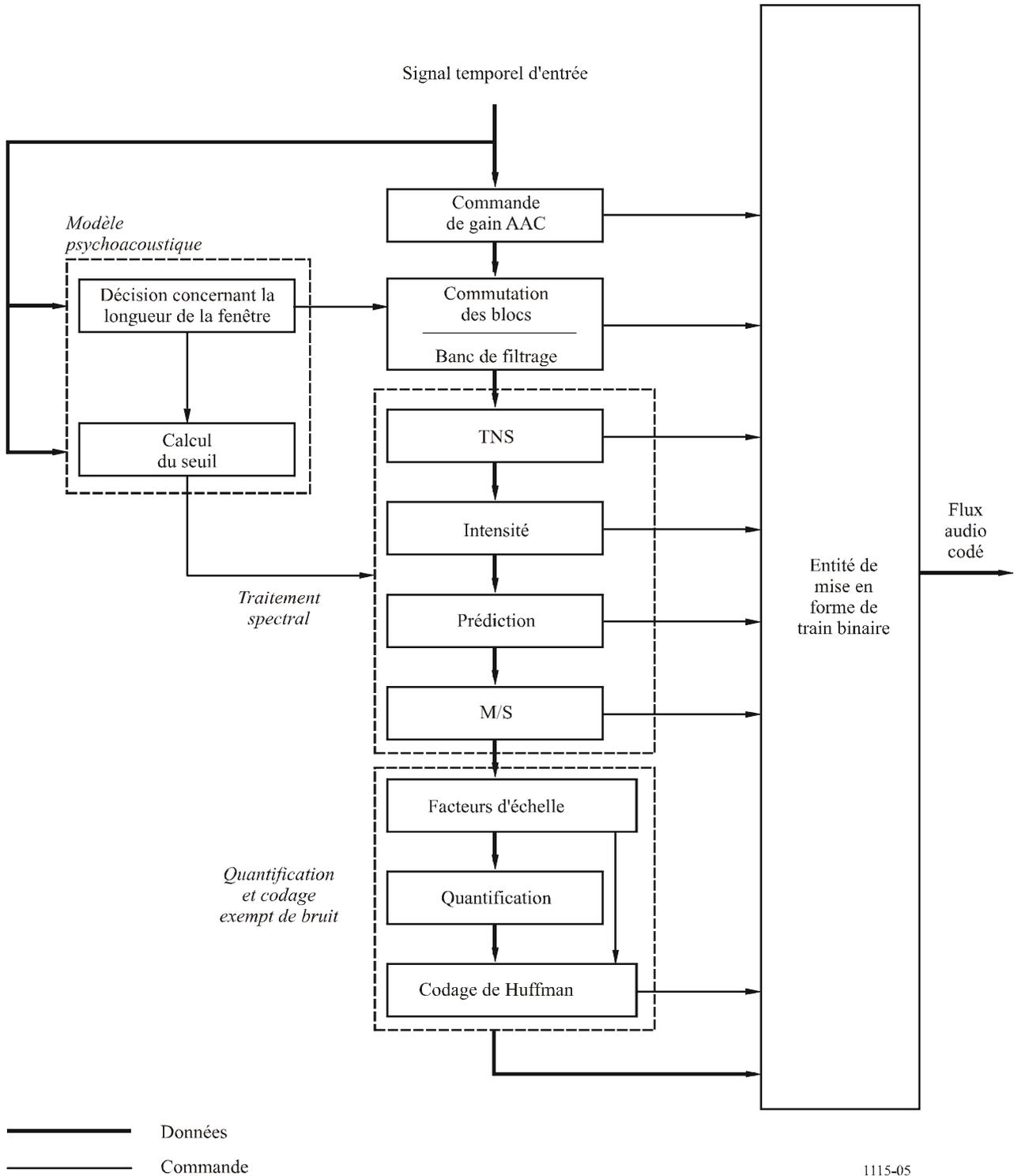
La structure de base du codeur MPEG-2 AAC est représentée dans la Fig. 5. Le système AAC est composé des outils de codage suivants:

- *Commande de gain*: Commande partageant le signal d'entrée en cinq bandes de fréquences espacées de façon égale. La commande de gain est utilisée pour le profil SSR.
- *Banc de filtrage*: Dispositif (MDCT, transformée en cosinus discrète modifiée) décomposant le signal d'entrée en raies spectrales sous-échantillonnées avec une résolution spectrale de 23 Hz et une résolution temporelle de 21,3 ms (128 raies spectrales) ou avec une résolution spectrale de 187 Hz et une résolution temporelle de 2,6 ms (1 024 raies spectrales) à une fréquence d'échantillonnage de 48 kHz. On choisit la taille de la fenêtre parmi deux tailles de fenêtre possibles.
- *Mise en forme temporelle du bruit (TNS, temporal noise shaping)*: L'opération TNS suit l'étape du banc de filtres d'analyse. La technique TNS permet au codeur de contrôler la structure fine temporelle du bruit de quantification.
- *Codage stéréo M/S (mid/side) et codage stéréo intensité*: Dans le cas de signaux audio multicanaux, le codage stéréo intensité et le codage stéréo M/S peuvent être appliqués. Dans le codage stéréo intensité, seule l'enveloppe spectrale est transmise de manière à réduire les informations de direction transmises. Dans le codage stéréo M/S, on peut

transmettre la somme normalisée (M pour *middle*) et les signaux de différence (S pour *side*) au lieu des signaux gauche et droit d'origine.

- *Prédiction*: Pour réduire la redondance des signaux stationnaires, on effectue la prédiction du domaine temporel entre les raies spectrales sous-échantillonnées des trames successives.
- *Quantification et codage exempt de bruit*: L'outil de quantification repose sur l'utilisation d'un quantificateur non uniforme avec une taille de pas de 1,5 dB. Le codage de Huffman s'applique au spectre quantifié, aux différents facteurs d'échelle et aux informations de direction.
- *Entité de mise en forme de train binaire*: On utilise enfin une entité de mise en forme de train binaire pour multiplexer le train binaire qui est composé des coefficients spectraux quantifiés et codés et d'informations supplémentaires provenant de chaque outil.
- *Modèle psychoacoustique*: On calcule le seuil de masquage actuel au moyen d'un modèle psychoacoustique à partir du signal d'entrée. Le modèle psychoacoustique utilisé est analogue au modèle psychoacoustique 2 conforme à la Norme ISO/CEI 11172-3. On emploie un rapport signal/masquage, calculé à partir du seuil de masquage et du niveau du signal d'entrée, au cours du processus de quantification afin de limiter le bruit de quantification audible et de choisir un outil de codage approprié.

FIGURE 5
Schéma fonctionnel d'un codeur MPEG-2 AAC



3 Décodage

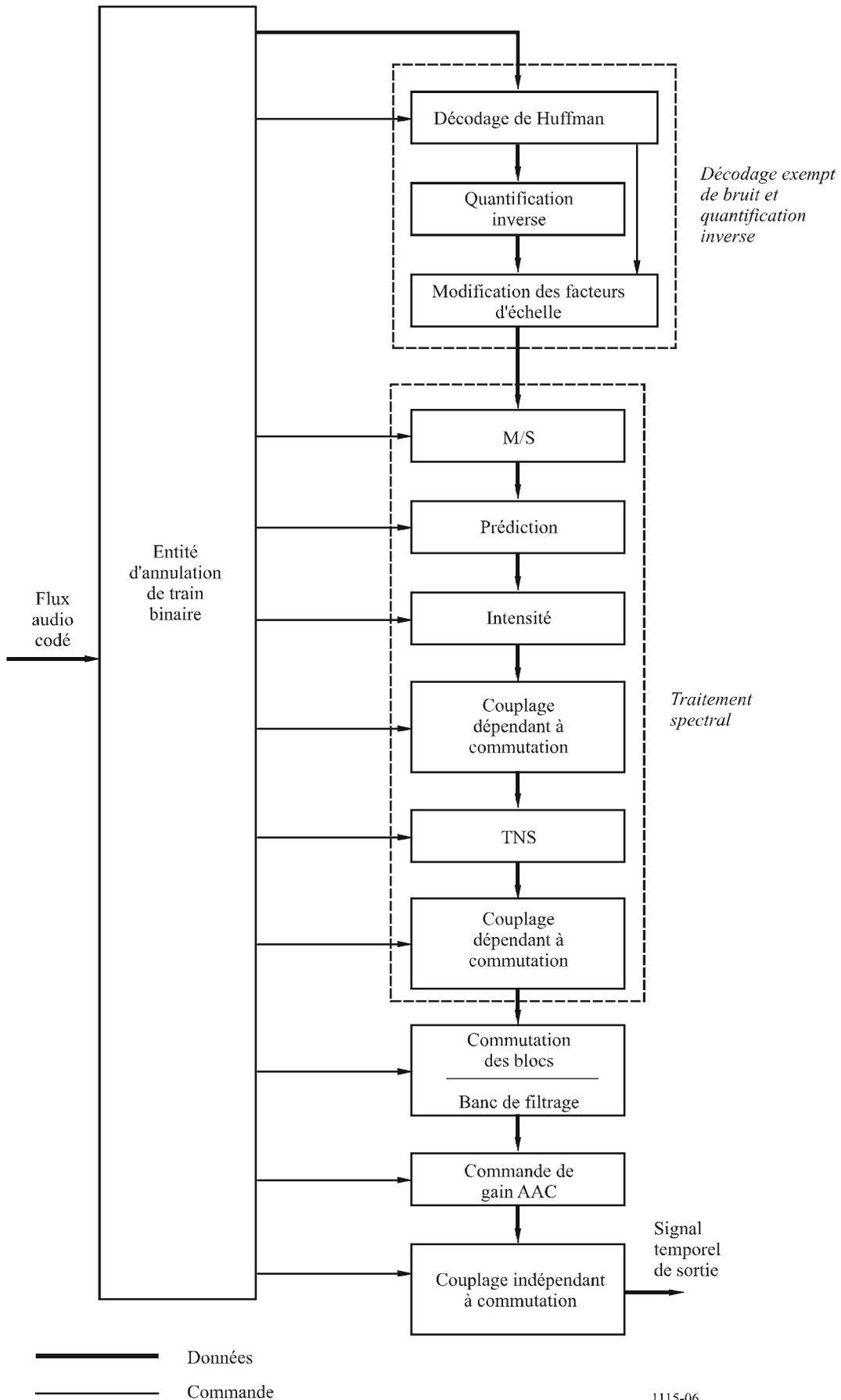
La structure de base du décodeur MPEG-2 AAC est représentée à la Fig. 6. Le processus de décodage consiste essentiellement en l'inverse du processus de codage.

Le décodeur a pour fonctions de trouver la description des spectres audio quantifiés dans le train binaire, de décoder les valeurs quantifiées et les autres informations relatives à la reconstruction, de reconstruire les spectres quantifiés, de traiter les spectres reconstruits au moyen des outils

fonctionnant dans le train binaire de manière à obtenir les spectres du signal réel tels que décrits par le train binaire d'entrée et, enfin, de convertir le domaine spectral en domaine temporel, avec ou sans outil optionnel de commande de gain. Après reconstruction initiale et mise à échelle du spectre reconstruit, on peut utiliser de nombreux outils optionnels pour modifier un ou plusieurs spectres afin de renforcer l'efficacité du codage. Pour chacun des outils optionnels fonctionnant dans le domaine spectral, l'option «transmettre» est retenue et, dans tous les cas où un mode spectral est omis, les spectres au niveau de l'entrée sont transmis directement, sans modification, au moyen de l'outil.

FIGURE 6

Schéma fonctionnel d'un décodeur MPEG-2 AAC



Appendice 4

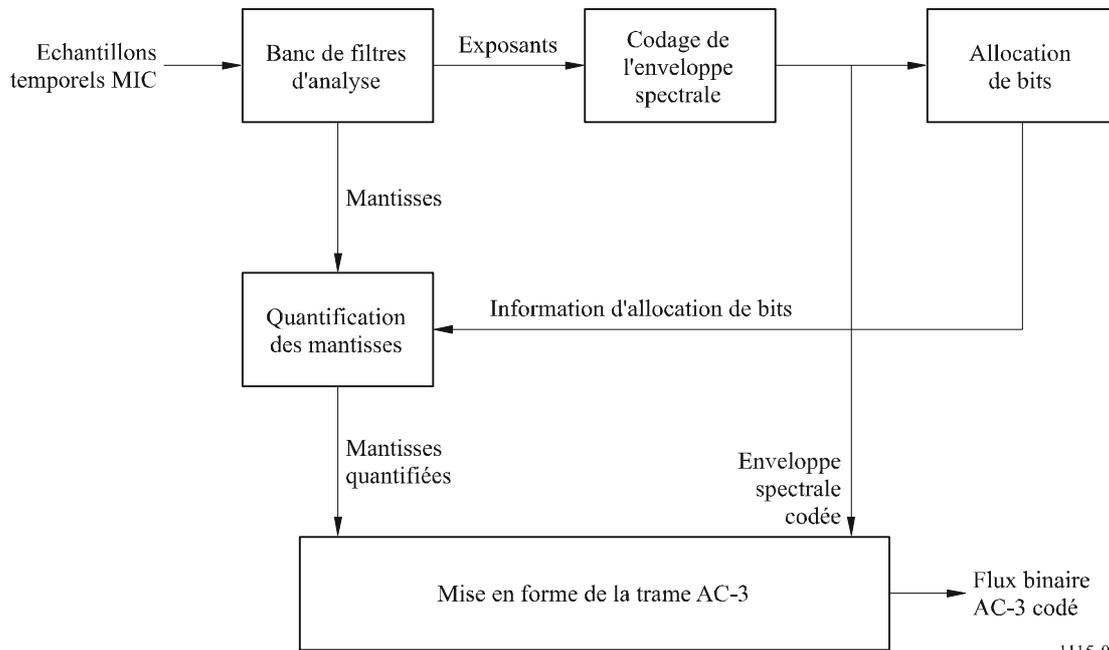
Description de la méthode de codage AC-3

Voir l'Annexe 2 de la Recommandation UIT-R BS.1196 et l'Annexe B de la Norme ATSC A/53B.

1 Codage

L'algorithme de compression numérique AC-3 peut coder de 1 à 5,1 canaux audio, à partir d'une représentation MIC, en un flux binaire série dont le débit est compris entre 32 et 640 kbit/s. L'algorithme AC-3 permet un gain de codage élevé (défini comme le rapport taux binaire d'entrée sur taux binaire de sortie) par quantification approximative d'une représentation du signal audio dans le domaine fréquentiel. La Fig. 7 représente ce processus sous forme synoptique. La première étape du processus de codage consiste à transformer la représentation du signal audio de la séquence d'échantillons temporels MIC en une séquence de blocs de coefficients de fréquence. Ceci est réalisé par le banc de filtres d'analyse. Les blocs de 512 échantillons temporels avec chevauchement partiel sont multipliés par une fenêtre temporelle et subissent une transformation vers le domaine fréquentiel. En raison du chevauchement, chaque échantillon MIC entrant est représenté dans deux blocs séquentiels transformés. Le domaine fréquentiel peut alors être réduit d'un facteur 2 de manière que chaque bloc ne contienne que 256 coefficients fréquentiels. Chaque coefficient fréquentiel est représenté en notation exponentielle binaire sous la forme d'un exposant binaire et d'une mantisse. L'ensemble des exposants est codé sous la forme d'une représentation de manière à représenter approximativement le spectre du signal. Cette représentation appelée enveloppe spectrale est utilisée par un sous-programme commun d'attribution de bits qui détermine le nombre de bits à utiliser pour coder chaque mantisse. L'enveloppe spectrale et les mantisses approximativement quantifiées de six blocs audio (1 536 échantillons audio) sont formatées en une trame AC-3. Le flux binaire AC-3 est une séquence de trames AC-3.

FIGURE 7
Codeur AC-3



1115-07

En réalité, le fonctionnement du codeur AC-3 est plus complexe que ne l'indique la Fig. 7 car:

- un en-tête de trame est ajouté, contenant les informations nécessaires pour la synchronisation, le codage et le décodage du flux binaire (taux binaire, fréquence d'échantillonnage, nombre de canaux codés, etc.);
- des codes de détection d'erreur sont insérés de manière à permettre au décodeur de vérifier qu'une trame reçue ne contient pas d'erreurs;
- la résolution spectrale du banc de filtres d'analyse peut être modifiée d'une manière dynamique pour assurer une meilleure adéquation des caractéristiques temps/fréquence de chaque bloc audio;
- l'enveloppe spectrale peut être codée avec une résolution temps/fréquence variable;
- une attribution de bits plus complexe peut être réalisée et les paramètres du sous-programme central d'attribution de bits peuvent être modifiés afin d'obtenir une attribution de bits plus optimisée;
- les canaux peuvent être couplés entre eux aux fréquences hautes afin de réaliser un gain de codage plus élevé pour un fonctionnement à des taux binaires plus faibles;
- dans le mode bicanal, un processus de rematriçage peut être exécuté de manière sélective afin d'obtenir un gain de codage supplémentaire dans le cas où le signal bicanal est décodé avec un décodeur ambiophonique matriciel.

2 Décodage

Le processus de décodage consiste essentiellement en l'inverse du processus de codage. Le décodeur présenté à la Fig. 8 doit se synchroniser sur le flux binaire codé, détecter les erreurs et annuler la mise en forme des différents types de données tels que l'enveloppe spectrale codée et les mantisses quantifiées. Le sous-programme d'attribution des bits est exécuté et les résultats sont utilisés pour désassembler et déquantifier les mantisses et l'enveloppe spectrale est décodée de

La Recommandation UIT-R BS.1387 spécifie une méthode de mesure objective de la qualité du son perçu.

Appendice 6

Méthode de sélection des codecs audio à faible débit binaire recommandés dans les essais réalisés en 1992-1993

On a effectué des tests subjectifs pour diverses applications à une voie et conventionnelles à deux voies et les résultats ont été comparés aux spécifications. Des études sont en cours sur des systèmes de codage multivoies.

Pour les trois applications émission, distribution et contribution, la qualité de base audio des codecs reproduite après décodage doit être équivalente à celle d'un disque compact, c'est-à-dire subjectivement indifférenciable pour la plupart des échantillons de programme audio. La qualité de l'image et l'effet stéréophonique doivent être maintenus.

Pour l'émission, les échantillons les plus critiques pour les codecs doivent être tels que la dégradation soit «perceptible mais non gênante» (note 4 sur l'échelle de dégradation). (Les essais subjectifs ont été faits avec l'échelle de dégradation à cinq notes (Recommandation UIT-R BS.562).) Au cours d'essais effectués en 1992, les systèmes suivants ont satisfait à ces exigences:

2 voies indépendantes à 128 kbit/s (voir la Note 1):

- ISO/CEI 11172-3 Couche II;
- ISO/CEI 11172-3 Couche III;
- Dolby AC-2.

stéréo combiné à 192 kbit/s:

- ISO/CEI 11172-3 Couche II;
- ISO/CEI 11172-3 Couche III.

Pour les applications de distribution, le signal après avoir transité dans trois codecs en série à 120 kbit/s par voie indépendante devrait se voir attribuer une note inférieure d'au plus 0,5 à celle du signal source d'origine (voir la Note 1). Une réduction de gain de 0,1 dB dans le domaine linéaire MIC a permis de restructurer les codes binaires. Le codec de Couche II de l'ISO/CEI 11172-3 a été le seul système à satisfaire aux exigences et sert donc de base pour ce type d'application.

NOTE 1 – A ce débit binaire, les codecs ont été notés 4,5 ou plus pour les échantillons les plus critiques. Etant donné que la méthode d'essai continue d'évoluer, on obtiendra peut-être d'autres notes pour les essais futurs

Pour les applications de contribution, le signal après avoir transité dans cinq codecs en série à 180 kbit/s par voie indépendante doit être indifférenciable du signal source d'origine. Le montage en série a été effectué avec une résolution linéaire MIC de 18 bits. Une réduction de gain de 0,1 dB dans le domaine linéaire MIC a permis de restructurer les codes binaires. Le codec de Couche II de l'ISO/CEI 11172-3 a été le seul système à satisfaire aux exigences et sert donc de base pour ce type d'application.

Pour les applications de commentaires, la qualité de base audio des signaux vocaux reproduits après décodage doit être équivalente à celle d'un signal d'origine linéaire MIC de 14 bits. Pendant les essais en conditions réelles, on a utilisé un format linéaire MIC à 16 bits. Des niveaux de dégradation perceptibles sont tolérés pour des séquences musicales ayant la qualité d'un disque

compact. Au cours d'essais effectués en 1992 avec des haut-parleurs seulement, seul le codec de Couche III ISO/CEI 11172-3 a fonctionné correctement pour des signaux vocaux et a été, en moyenne, le meilleur moyen pour les signaux musicaux. Pour les signaux vocaux, les notes ont toujours été supérieures à 4 (dégradation perceptible mais non gênante) en mode mono (60 kbit/s) et stéréo (stéréo combiné 120 kbit/s). Ce codec sert donc de base à la Recommandation pour ce type d'application. Au cours d'essais ultérieurs en 1993, les résultats obtenus avec écoute au casque pour évaluer la qualité d'un codec de commentaire mono Couche III à 60 kbit/s ont donné une note inférieure à 4, ce qui ne satisfait pas à la qualité exigée pour un signal vocal. Il faut continuer les études à 60 kbit/s.

Il ressort de l'évaluation de complexité que les décodeurs de Couche II ISO/CEI 11172-3 et Dolby AC-2 sont les moins complexes. Sur la base des considérations suivantes:

- décodeur de faible complexité,
- similitude avec les applications de distribution et de contribution,
- souplesse afin d'améliorer le codeur,
- codec testé aux deux débits binaires,

le format et le décodeur de Couche II ISO/CEI 11172-3 ont été choisis comme base de la Recommandation pour l'émission à 2×128 kbit/s. Au cours d'essais effectués en 1992 à 192 kbit/s, le décodeur de Couche II ISO/CEI 11172-3 n'a satisfait qu'occasionnellement aux exigences fixées et on a décidé qu'il fallait apporter des améliorations à ce débit binaire pour qu'il respecte pleinement les critères concernant la qualité audio de base (comme cela est le cas pour le décodeur de Couche III ISO/CEI 11172-3). D'autres essais effectués en 1993 n'ont révélé aucune amélioration à ce stade. Il faudra procéder à des études complémentaires à 192 kbit/s en mode stéréo combiné pour réexaminer l'emploi de ce débit binaire. Les administrations sont priées de présenter des contributions sur ce sujet.

En 1993, on a procédé à des essais de vérification de réseau sur une chaîne de radiodiffusion complète qui comprenait cinq codecs de contribution à 180 kbit/s en série, trois codecs de distribution à 120 kbit/s en série et un codec d'émission. Les résultats de ces essais ont montré que la qualité audio de base à l'extrémité de cette chaîne n'était pas satisfaisante.

Des essais ont aussi été effectués avec huit codecs à 180 kbit/s en série; il s'est avéré que cette configuration offrait une qualité acceptable.
