

Union internationale des télécommunications

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BS.1196-8
(10/2019)

**Codage audio pour la
radiodiffusion numérique**

Série BS
Service de radiodiffusion sonore



Union
internationale des
télécommunications

Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse <http://www.itu.int/ITU-R/go/patents/fr>, où l'on trouvera également les Lignes directrices pour la mise en oeuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Egalement disponible en ligne: <http://www.itu.int/publ/R-REC/fr>)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	Service de radiodiffusion télévisuelle
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Emissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

Note: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.

Publication électronique
Genève, 2020

© UIT 2020

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

RECOMMANDATION UIT-R BS.1196-8*

Codage audio pour la radiodiffusion numérique

(Question UIT-R 19-1/6)

(1995-2001-2010-2012-02/2015-10/2015-2017-01/2019-10/2019)

Domaine d'application

La présente Recommandation spécifie les systèmes de codage des sources audio applicables à la radiodiffusion sonore et télévisuelle numérique. Elle fournit également les spécifications d'un système destiné à l'amélioration multicanal rétrocompatible de systèmes de radiodiffusion sonore et télévisuelle numérique.

Mots clés

Audio, codage audio, radiodiffusion, numérique, son, télévision, codec

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que les spécifications utilisateur en matière de systèmes de codage audio pour la radiodiffusion numérique sont énoncées dans la Recommandation UIT-R BS.1548;
- b) que la Recommandation UIT-R BS.775 décrit un système sonore multicanal avec ou sans image associée et qu'un système sonore multicanal de haute qualité utilisant une réduction efficace du débit binaire est un élément essentiel des systèmes de radiodiffusion numérique;
- c) que le système sonore évolué spécifié dans la Recommandation UIT-R BS.2051 comprend des configurations de canaux en trois dimensions et utilise des métadonnées statiques ou dynamiques pour commander les signaux basés sur un objet, basés sur une scène et basés sur un canal;
- d) que l'évaluation subjective des dégradations faibles dans les systèmes audio, y compris les systèmes audio multivoies, fait l'objet de la Recommandation UIT-R BS.1116;
- e) que l'évaluation subjective du niveau de qualité sonore intermédiaire des systèmes audio fait l'objet de la Recommandation UIT-R BS.1534 (MUSHRA);
- f) 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;
- g) que l'usage de techniques de codage de source audio similaires pour différents services pourrait permettre une plus grande souplesse d'utilisation des systèmes et une diminution du coût des récepteurs;
- h) que de nombreux services de radiodiffusion utilisent déjà des codecs audio appartenant aux familles MPEG-1, MPEG-2, MPEG-4, AC-3 et E-AC-3, ou en ont spécifié l'usage;
- i) que la Recommandation UIT-R R BS.1548 fournit une liste de codecs dont on a démontré la conformité aux spécifications du radiodiffuseur pour la contribution, la distribution et l'émission;
- j) que les radiodiffuseurs qui n'ont pas encore démarré de services doivent être à même de choisir le système le mieux adapté à leur application;

* Cette Recommandation doit être portée à l'attention de l'Organisation internationale de standardisation (ISO) et de la Commission électrotechnique internationale (CEI).

- k) que les radiodiffuseurs peuvent avoir besoin d'étudier la compatibilité avec les systèmes et équipements de radiodiffusion existants lorsqu'ils choisissent un nouveau système;
- l) que lors de la mise en place d'un système audio multicanal, il convient de prendre en compte les récepteurs monophoniques et stéréophoniques existants;
- m) qu'une extension multicanal rétrocompatible réalisée sur un système de codage audio existant peut apporter une meilleure efficacité en termes de débit binaire que la radiodiffusion simultanée;
- n) qu'un système de codage audio devrait de préférence pouvoir coder la parole et la musique avec la même haute fidélité,

recommande

1 d'utiliser, pour les nouvelles applications de radiodiffusion sonore ou télévisuelle numérique, lorsque la compatibilité avec les transmissions et les équipements hérités n'est pas requise, l'un des systèmes de codage audio à faible débit binaire suivants:

- HE-AAC étendu, tel que spécifié dans la norme ISO/CEI 23003-3:2012;
- E-AC-3, tel que spécifié dans la norme ETSI TS 102 366 (2014-08);
- AC-4, tel que spécifié dans la norme ETSI TS 103 190-1 v1.3.1 (2018-02) et dans la norme ETSI TS 103 190-2 v1.2.1 (2018-02);
- Profil LC de l'audio 3D MPEG-H, tel que spécifié dans la norme ISO/CEI 23008-3: 2019;
- DTS-UHD tel que spécifié dans la norme ETSI TS 103 491 v1.1.1 (2017-04).

NOTE 1 – HE-AAC étendu est un surensemble plus souple des systèmes MPEG-4 HE-AAC v2, HE-AAC et AAC-LC, qui comprend le système Discours unifié et codage audio (USAC) MPEG-D.

NOTE 2 – E-AC-3 est un surensemble plus souple du système AC-3.

NOTE 3 – Les spécifications AC-4, profil LC de l'audio 3D MPEG-H et DTS-UHD offrent des moyens pour prendre en charge le système sonore évolué spécifié dans la Recommandation UIT-R BS.2051, et les utilisateurs devraient se reporter à la Recommandation UIT-R BS.1548 pour ce qui concerne la conformité des codecs;

2 d'utiliser, pour les applications de radiodiffusion sonore ou télévisuelle numérique, lorsque la compatibilité avec les transmissions et les équipements existants est requise, l'un des systèmes de codage à faible débit binaire suivants:

- MPEG-1 de couche II, tel que spécifié dans la norme ISO/CEI 11172-3:1993;
- MPEG-2 de couche II à fréquence d'échantillonnage moitié, tel que spécifié dans la norme ISO/CEI 13818-3:1998;
- MPEG-2 AAC-LC ou MPEG-2 AAC-LC avec SBR, tel que spécifié dans la norme ISO/CEI 13818-7:2006;
- MPEG-4 AAC-LC, tel que spécifié dans la norme ISO/CEI 14496-3:2009;
- MPEG-4 HE AAC v2, tel que spécifié dans la norme ISO/CEI 14496-3:2009;
- AC-3, tel que spécifié dans la norme ETSI TS 102 366 (2014-08);

NOTE 4 – Le codec ISO/CEI 11172-3 peut parfois être désigné sous le nom de codec 13818-3 étant donné que la spécification de ce dernier se réfère à la norme 11172-3.

NOTE 5 – Il est recommandé de prendre en charge le système HE-AAC étendu tel que spécifié dans la norme ISO/CEI 23003-3:2012. Ce système intègre toutes les versions AAC susmentionnées et garantit donc, en employant un unique décodeur la compatibilité à l'échelle mondiale avec des systèmes de radiodiffusion futurs ou existants;

3 d'utiliser, pour l'extension multicanal rétrocompatible des systèmes de radiodiffusion sonore et télévisuelle numérique, les extensions audio multicanal décrites dans la norme ISO/CEI 23003-1:2007;

NOTE 6 – Etant donné que la technologie MPEG ambiophonique décrite dans la norme ISO/CEI 23003-1:2007 est indépendante de la technologie de compression (codeur de base) utilisée pour la transmission du signal rétrocompatible, les outils d'amélioration multicanal décrits peuvent être utilisés en combinaison avec l'un quelconque des systèmes de codage recommandés aux points 1 et 2 du *recommande*;

4 d'utiliser éventuellement, pour les liaisons de distribution et de contribution, l'un des codages suivants, avec un débit binaire au moins égal au débit indiqué 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:

- codage MPEG-1 de couche II, tel que spécifié dans la norme ISO/CEI 11172-3, à un débit d'au moins 180 kbit/s par signal audio mono;
- codage MPEG-4 AAC, tel que spécifié dans la norme ISO/CEI 14496-3, à un débit d'au moins 144 kbit/s par signal audio mono lorsqu'on utilise jusqu'à cinq codecs en série;
- codage AC-4 tel que spécifié dans les normes ETSI TS 103 190-1 v1.3.1 (2018-02) et ETSI TS 103 190-2 v1.2.1 (2018-02), à un débit d'au moins 128 kbit/s par signal audio mono lorsqu'on utilise jusqu'à cinq codecs en série;
- codage MPEG-H 3D Audio, tel que spécifié dans la norme ISO/CEI 23008-3, à un débit d'au moins 144 kbit/s par signal audio mono lorsqu'on utilise jusqu'à cinq codecs en série;

5 d'utiliser éventuellement, pour les liaisons de diffusion de commentaires, le système de codage de couche III de la norme ISO/CEI 11172-3, avec un débit binaire d'au moins 60 kbit/s, à l'exclusion des données auxiliaires, pour les signaux mono, 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;

6 d'utiliser une fréquence d'échantillonnage de 48 kHz pour les applications qui requièrent une haute qualité;

7 de ne pas appliquer d'accentuation au signal d'entrée du codeur audio à faible débit binaire, le codeur ne devant pas, lui non plus, en introduire,

recommande en outre

1 de se reporter à la version la plus récente de la Recommandation UIT-R BS.1548 pour toute information concernant les configurations des systèmes de codage qui se sont révélées être conformes aux prescriptions de qualité et autres prescriptions utilisateur pour la contribution, la distribution et l'émission;

2 de poursuivre les études requises concernant les exigences des systèmes sonores évolués spécifiés dans la version la plus récente de la Recommandation UIT-R BS.2051 et de mettre à jour la présente Recommandation une fois ces études achevées.

NOTE – Les Annexes 1 à 8 fournissent des informations sur les codecs mentionnés dans la présente Recommandation.

Annexe 1 (pour information)

MPEG-1 et MPEG-2, couches II et III audio

1 Codage

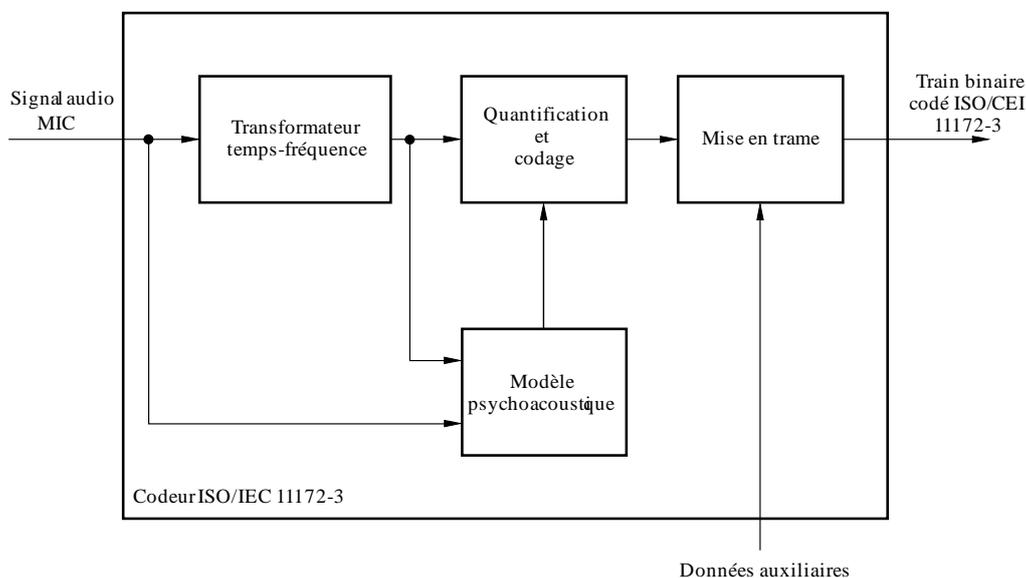
Le codeur traite le signal audionumérique et comprime le train binaire. L'algorithme de codage n'est pas normalisé et peut utiliser divers moyens pour effectuer le codage, par exemple l'estimation du seuil de masquage à l'écoute, la quantification et les facteurs d'échelle (Note 1 ci-après). Toutefois, 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. La transformation temps-fréquence donne une représentation filtrée et sous-échantillonnée du flux audio entrant. Les échantillons transformés 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 transformation 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 oeuvre 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 transformés. Là aussi, la fonction de transfert de ce bloc dépend du système de codage. Le bloc «mise en trame» assemble, pour la couche choisie, le train binaire proprement dit à partir des données de sortie des autres blocs (données d'allocation des bits, facteurs d'échelle, échantillons de sous-bande codés, etc.) 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 en fonction de la couche utilisée.

Couche I: Cette couche comprend la transformation 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'allocation 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 apporte un codage de l'allocation des bits, des facteurs d'échelle et des échantillons supplémentaires. Une trame de la couche II représente $3 \times 384 = 1\,152$ échantillons par voie.

Couche III: Cette couche apporte une résolution de fréquence plus élevée, qui repose sur un banc de filtres hybride (banc de filtres à 32 sous-bandes avec transformée en cosinus discrète modifiée de longueur variable). Elle ajoute une quantification non uniforme, une segmentation adaptative et un codage entropique des valeurs quantifiées. Une trame de la couche III représente 1 152 échantillons par voie.

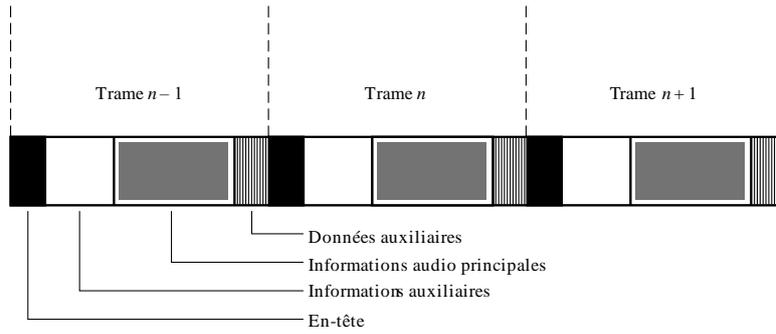
Quatre modes différents sont possibles pour chacune des couches:

- 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 en exploitant la redondance et la présence d'éléments non pertinents pour la stéréo). Le mode stéréo combiné peut servir à améliorer la qualité audio aux faibles débits binaires et/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 binaire codé est composé de trames successives. Les champs contenus dans la trame diffèrent selon la couche, comme indiqué ci-dessous:

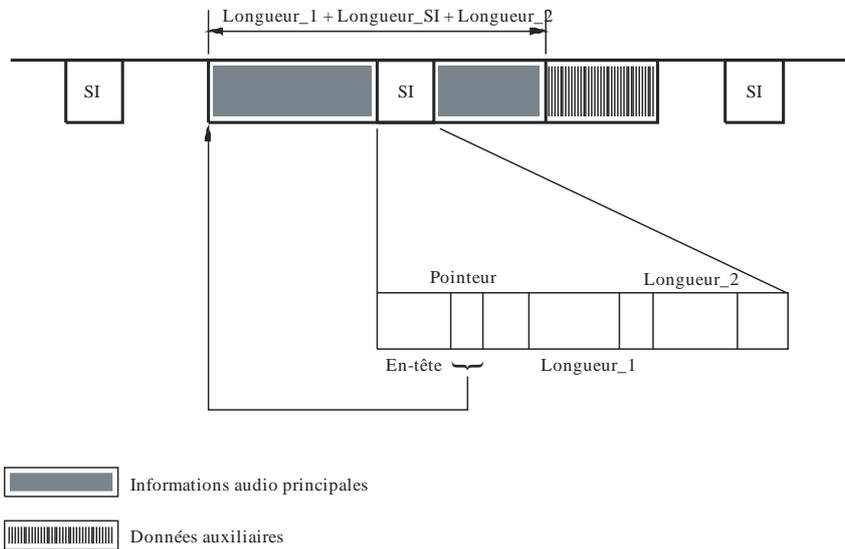
FIGURE 2
Format du train binaire de Couche II ISO/CEI 11172-3



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'allocation des bits et les facteurs d'échelle	
Informations 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	BS.1196-02

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



Layer III:

Information auxiliaire (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;
Informations audio principales:	partie du train binaire contenant les signaux audio codés;
Données auxiliaires:	partie du train binaire contenant des données définissables par l'utilisateur.

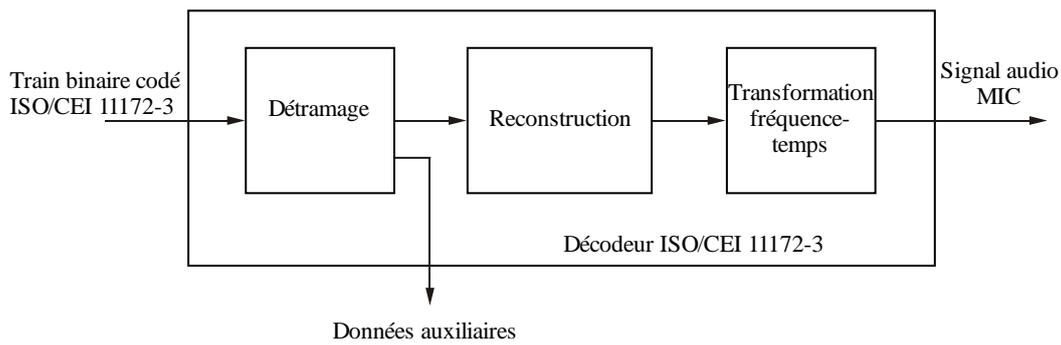
4 Décodage

Le décodeur accepte les trains binaires audio comprimés 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 le train binaire audio codé. Le mécanisme de détramage et de décodage du train binaire effectue en option une détection d'erreur si le codeur applique un mécanisme de correction d'erreur. Le train binaire est détramé de manière à retrouver les différents éléments d'information comme l'en-tête de trame audio, l'allocation des bits, les facteurs d'échelle, les échantillons transformés et, éventuellement, les données auxiliaires. Le processus de reconstruction reconstruit la version quantifiée de l'ensemble des échantillons transformés. La transformation fréquence-temps retransforme ces échantillons en signaux audio au format MIC linéaire.

FIGURE 4

Schéma fonctionnel du décodeur



BS.1196-04

Annexe 2 (pour information)

MPEG-2 et MPEG-4 AAC audio

1 Introduction

La norme ISO/CEI 13818-7 décrit la norme MPEG-2 audio non rétrocompatible, 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 produire une qualité audio élevée. La mémoire et la puissance de traitement nécessaires sont supérieures à celles requises 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 au profil LC sont inférieures à celles requises pour 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 du bruit dans le domaine temporel (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 avec outil de commande de gain. Il permet de choisir certaines bandes de fréquences à décoder, ce qui permet de limiter les éléments matériels du décodeur. 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 1), et un maximum de 48 canaux audio. Le Tableau 2 présente les configurations de canaux par défaut, à savoir les configurations monocanal, bicanal, à cinq canaux (trois à l'avant/deux à l'arrière) et à cinq canaux plus un 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 autorise une certaine souplesse dans la disposition multicanal des haut-parleurs. La réduction par mixage du nombre de canaux («*down-mix*») est également prise en charge. L'utilisateur peut désigner un coefficient de réduction pour effectuer une conversion multicanal/bicanal. Il est par conséquent possible de régler la qualité du son au moyen d'un dispositif de lecture comportant seulement deux canaux.

TABLEAU 1

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 2
Configurations de canaux par défaut⁽¹⁾

Valeur ⁽²⁾	Nombre de haut-parleurs	Éléments de syntaxe audio, énumérés par ordre de réception	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) ⁽³⁾	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 ⁽⁴⁾
1	1	single_channel_element	M+000	Mono
2	2	channel_pair_element	M+030, M-030	Gauche, droit
3	3	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Gauche, droit
4	4	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Gauche, droit
		single_channel_element()	M+180	Ambiance mono
5	5	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Gauche, droit
		channel_pair_element()	M+110, M-110	Ambiophonique gauche, Ambiophonique droit
6	5 + 1	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Gauche, droit
		channel_pair_element()	M+110, M-110	Ambiophonique gauche, Ambiophonique droit
		lfe_element()	LFE1	Effets basse fréquence
7	7 + 1 Avant	single_channel_element()	M+000	N/D ⁽⁵⁾
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+045, M-045	
		channel_pair_element()	M+110, M-110	
		lfe_element()	LFE1	
8-10	–	–	réservé	–
11	6 + 1	single_channel_element()	M+000	N/D
		channel_pair_element()	M+030, M-030	
		channel_pair_element()	M+110, M-110	
		single_channel_element()	M+180	
		lfe_element()	LFE1	
12	7 + 1 Arrière	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Gauche, droit
		channel_pair_element()	M+090, M-090	Ambiophonique gauche, Ambiophonique droit
		channel_pair_element()	M+135, M-135	Ambiophonique arrière gauche, Ambiophonique arrière droit
		lfe_element()	LFE1	Effet basse fréquence
13	22 + 2	single_channel_element()	M+000	Avant central
		channel_pair_element()	M+030, M-030	Avant gauche central, avant droit central
		channel_pair_element()	M+060, M-060	Avant gauche, avant droit
		channel_pair_element()	M+090, M-090	Latéral gauche, latéral droit
		channel_pair_element()	M+135, M-135	Arrière gauche, arrière droit
		single_channel_element()	M+180	Arrière central
		lfe_element()	LFE1	Effets basse fréquence-1

TABLEAU 2 (fin)

Valeur ⁽²⁾	Nombre de haut-parleurs	Éléments de syntaxe audio, énumérés par ordre de réception	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) ⁽³⁾	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 ⁽⁴⁾
		lfe_element()	LFE2	Effets basse fréquence-2
		single_channel_element()	U+000	Supérieur avant central
		channel_pair_element()	U+045, U-045	Supérieur avant gauche, supérieur avant droit
		channel_pair_element()	U+090, U-090	Supérieur latéral gauche, supérieur latéral droit
		single_channel_element()	T+000	Supérieur central
		channel_pair_element()	U+135, U-135	Supérieur arrière gauche, supérieur arrière droit
		single_channel_element()	U+180	Supérieur arrière central
		single_channel_element()	B+000	Inférieur avant central
		channel_pair_element()	B+045, U-045	Inférieur avant gauche, inférieur avant droit
14	7 + 1 Supérieur	single_channel_element()	M+000	Central
		channel_pair_element()	M+030, M-030	Gauche, droit
		channel_pair_element()	M+110, M-110	Ambiophonique gauche, Ambiophonique droit
		lfe_element()	LFE1	Effets basse fréquence
		channel_pair_element()	U+030, U-030	Supérieur avant gauche, supérieur avant droit
15	–	–	réservé	–

(1) La liste est tirée du Tableau 1.19 de la norme ISO/CEI 14496-3:2009/Amd.4:2013.

(2) La configuration de canaux de sortie audio est indiquée par un champ de quatre bits qui achemine une valeur de configuration de canaux définie dans la norme ISO/CEI 23001-8:2013 «Points de code indépendants du codage». La norme MPEG-2 s'applique pour les valeurs de configuration de canaux allant jusqu'à 7. La norme MPEG-4 AAC s'applique pour les valeurs de configuration de canaux allant jusqu'à 15.

(3) Identification des haut-parleurs par des étiquettes conformément à la Recommandation UIT-R BS.2051.

(4) Veuillez noter que l'étiquette et le nom d'un canal dépendent de la configuration effective du canal.

(5) N/D: non disponible; configuration de canal ne figurant pas dans la Recommandation UIT-R BS.2051 ou UIT-R BS.775.

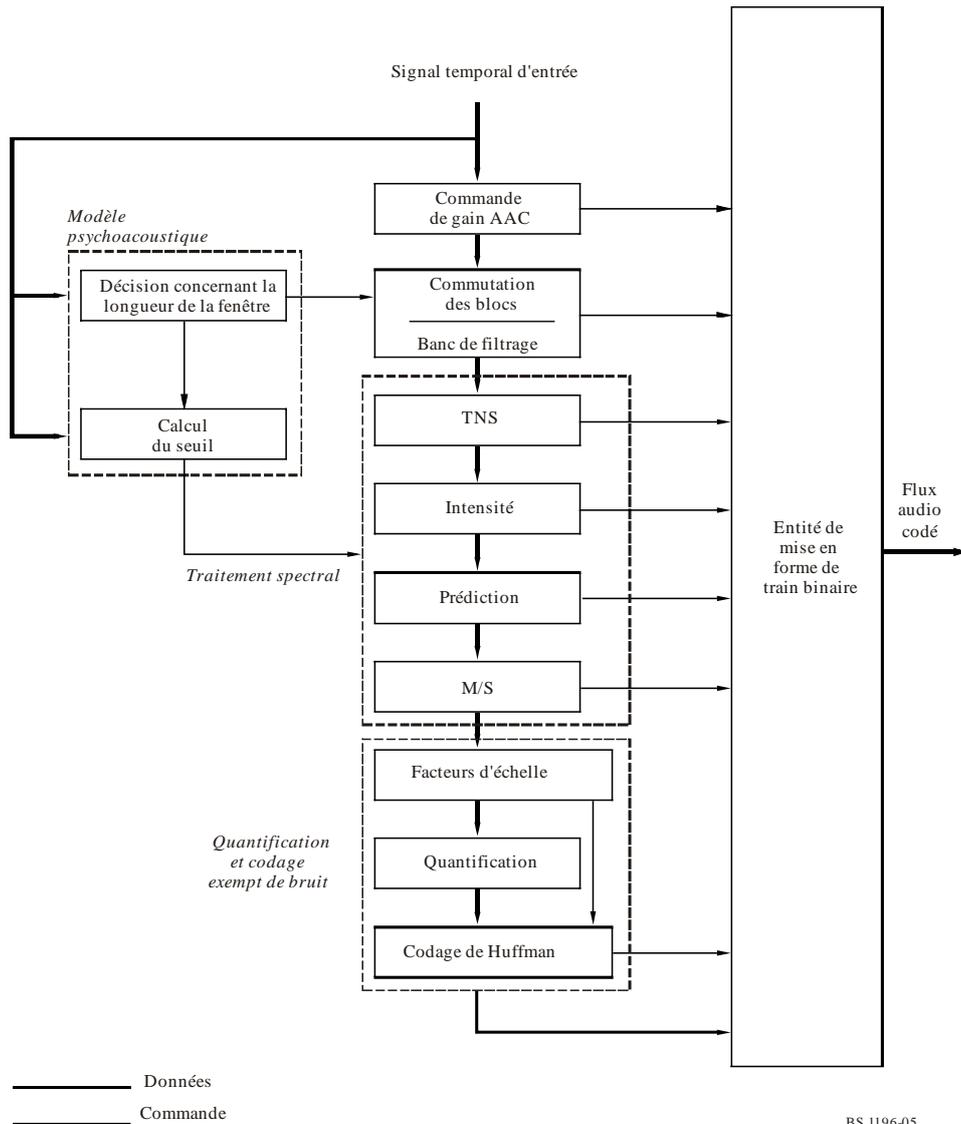
2 Codage

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

- Commande de gain: Commande partageant le signal d'entrée en quatre bandes de fréquences espacées de façon égale. La commande de gain est utilisée pour le profil SSR.
- Banc de filtres: Dispositif (TCDM, transformée en cosinus discrète modifiée) décomposant le signal d'entrée en composantes spectrales sous-échantillonnées avec une résolution en fréquence de 23 Hz et une résolution temporelle de 21,3 ms (128 composantes spectrales) ou avec une résolution en fréquence de 187 Hz et une résolution temporelle de 2,6 ms (1 024 composantes 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 du bruit dans le domaine temporel (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 qui sont émises. 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 composantes 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 un 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 se compose des coefficients spectraux quantifiés et codés et d'informations supplémentaires provenant de chaque outil.
- Modèle psychoacoustique: On calcule le seuil de masquage courant au moyen d'un modèle psychoacoustique à partir du signal d'entrée. Le modèle psychoacoustique utilisé est analogue au modèle psychoacoustique 2 de 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

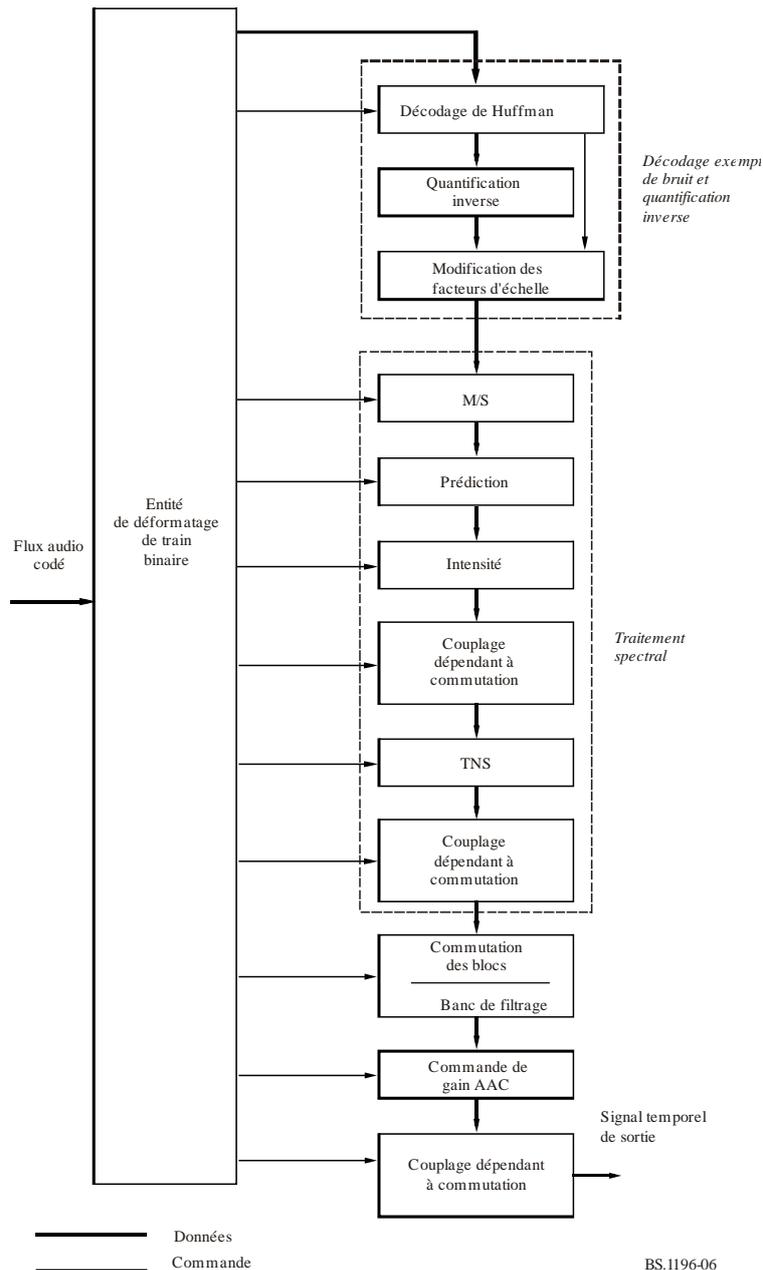


3 Décodage

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.

FIGURE 6

Schéma fonctionnel d'un décodeur MPEG-2 AAC

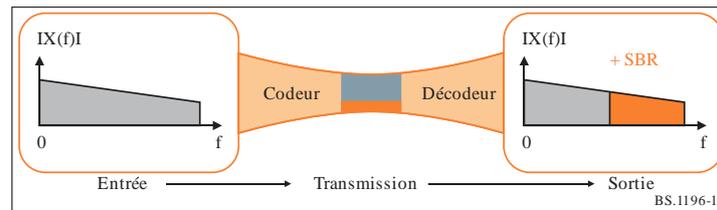


Le décodeur a pour fonctions de trouver la description des spectres audio quantifiés dans le train binaire, de décoder les valeurs quantifiées et les autres informations relatives à la reconstruction, de reconstruire les spectres quantifiés, de traiter les spectres reconstruits au moyen des outils actifs 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 l'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 «traverser» (*pass through*) est retenue et, dans tous les cas où une opération spectrale est omise, les spectres en entrée traversent directement l'outil sans modification.

4 HE AAC et reconstruction de bande spectrale

Le profil HE AAC (*High Efficiency AAC*, AAC à haute efficacité) apporte la reconstruction de bande spectrale (SBR, *spectral band replication*). La SBR est une méthode très efficace de codage des hautes fréquences utilisée dans les algorithmes de compression audio. Elle renforce l'efficacité des codecs audio et vocaux de faible débit binaire en augmentant la largeur de bande audio à un débit donné ou en améliorant l'efficacité du codage à un niveau de qualité donné.

Seule est codée et émise la partie inférieure du spectre, celle à laquelle l'oreille humaine est la plus sensible. Au lieu de transmettre la partie supérieure du spectre, ce système utilise la SBR comme un processus de post-décodage pour reconstruire les fréquences supérieures à partir d'une analyse des fréquences inférieures transmises. Pour garantir l'exactitude de la reconstruction, des paramètres utilisés par la SBR sont transmis dans le train binaire à un débit très faible.

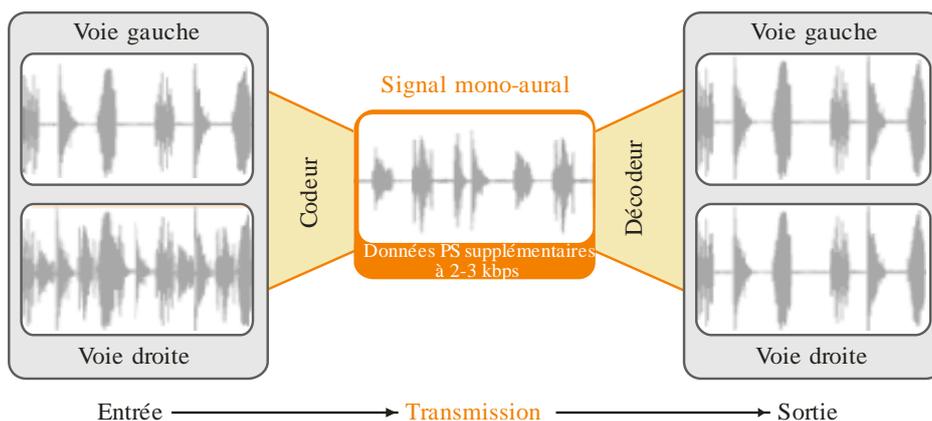


Le train binaire HE AAC est une version améliorée du train binaire audio AAC. Les données SBR supplémentaires sont insérées dans l'élément de remplissage AAC (élément «FIL»), ce qui garantit une compatibilité avec la norme AAC. La technologie HE AAC est un système à débit dual. Le train binaire audio AAC pur rétrocompatible est transmis à une fréquence d'échantillonnage égale à la moitié de celle correspondant aux améliorations SBR. Ainsi, un décodeur AAC, qui n'est pas conçu pour décoder les données d'amélioration SBR, produit en sortie un signal temporel dont la fréquence d'échantillonnage est égale à la moitié de celle produite par un décodeur HE AAC.

5 HE AAC version 2 et stéréo paramétrique

Le HE AAC v2 est une extension du HE AAC qui offre la stéréo paramétrique (PS, *parametric stereo*) en vue de renforcer l'efficacité de la compression audio pour les signaux stéréo à faible débit binaire.

Le codeur analyse le signal audio stéréo et construit une représentation paramétrique de l'image stéréo. Il n'est dès lors plus nécessaire de transmettre les deux canaux et seule une représentation mono-aurale du signal stéréo d'origine est codée. Les paramètres nécessaires à la reconstruction de l'image stéréo sont transmis avec le signal.



En conséquence, la qualité audio perçue d'un train binaire de faible débit (par exemple, 24 kbit/s) intégrant une stéréo paramétrique est bien supérieure à la qualité d'un train binaire analogue sans stéréo paramétrique.

Le train binaire HE AAC v2 est construit sur le train binaire HE AAC. Les données supplémentaires pour la stéréo paramétrique sont insérées dans l'élément d'extension SBR du train HE AAC mono, ce qui garantit une compatibilité avec le HE AAC et l'AAC.

Un décodeur HE AAC, qui n'est pas conçu pour décoder les données d'amélioration liées à la stéréo paramétrique, produit un signal de sortie monophonique à pleine largeur de bande. Un décodeur AAC pur, qui n'est pas conçu pour décoder les données d'amélioration SBR, produit en sortie un signal temporel à la moitié de la fréquence d'échantillonnage.

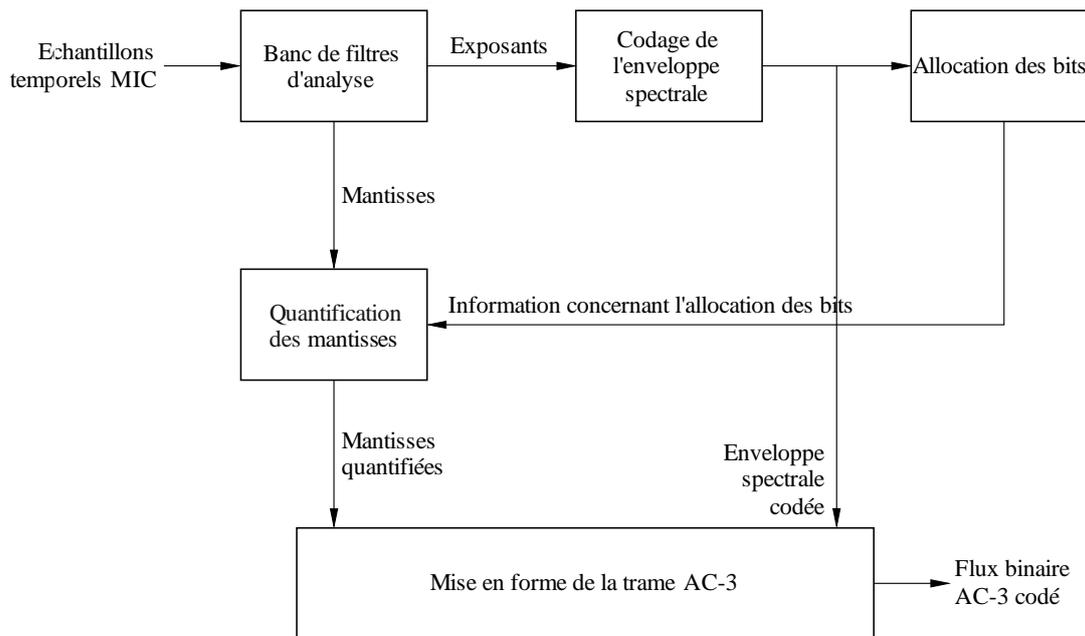
Annexe 3 (pour information)

AC-3 et E-AC-3 audio

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 débit binaire en entrée sur débit binaire en 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, en partant d'une séquence d'échantillons temporels MIC pour obtenir une séquence de blocs de coefficients de fréquence. Ceci est réalisé par le banc de filtres d'analyse. Des blocs de 512 échantillons temporels avec chevauchement sont multipliés par une fenêtre temporelle et subissent une transformation vers le domaine fréquentiel. En raison du recouvrement des blocs, chaque échantillon MIC entrant se trouve représenté dans deux blocs transformés consécutifs. Le domaine fréquentiel peut alors être réduit dans un rapport 2 de manière à ce 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 approximative du spectre du signal. Cette représentation, appelée enveloppe spectrale, est utilisée par le sous-programme de base d'allocation des 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 mises sous forme d'une trame AC-3. Le flux binaire AC-3 est une séquence de trames AC-3.

FIGURE 7
Codeur AC-3



BS.1196-07

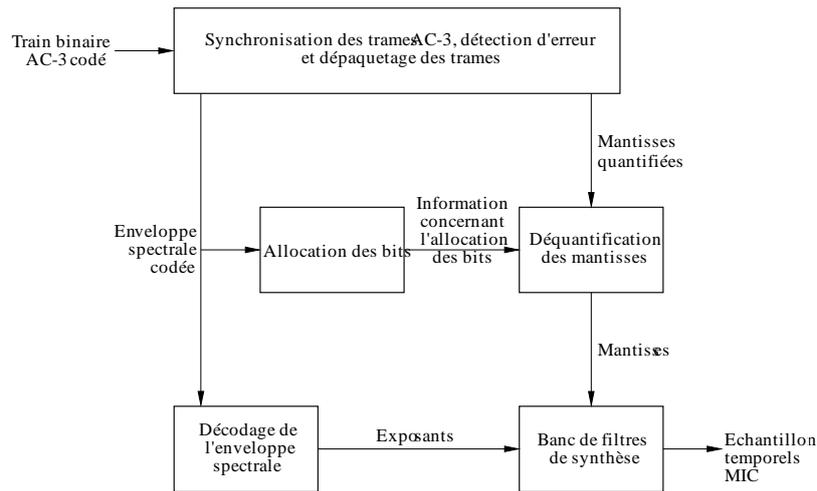
En réalité, le fonctionnement du codeur AC-3 est plus complexe que ne l'indique la Fig. 7. En effet:

- un en-tête de trame est ajouté, contenant les informations nécessaires pour la synchronisation et le décodage du train binaire (débit binaire, fréquence d'échantillonnage, nombre de canaux codés, etc.);
- des codes de détection d'erreur sont insérés pour 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 de façon 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 allocation des bits plus complexe peut être réalisée et les paramètres du sous-programme central d'allocation peuvent être modifiés à des fins d'optimisation;
- les canaux peuvent être couplés entre eux aux fréquences hautes en vue d'améliorer le gain de codage pour un fonctionnement à des débits 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 et d'améliorer les résultats dans le cas où le signal bicanal est décodé avec un décodeur ambiophonique matriciel.

2 Décodage

Le décodage est en substance le processus inverse du codage. Le décodeur présenté à la Fig. 8 doit se synchroniser sur le train binaire codé, détecter les erreurs et dépaqueter les différents types de données, notamment l'enveloppe spectrale codée et les mantisses quantifiées. Le sous-programme d'allocation des bits est exécuté et les résultats sont utilisés pour désassembler et déquantifier les mantisses et l'enveloppe spectrale est décodée de manière à produire les exposants. Les exposants et mantisses sont reconvertis dans le domaine temporel afin d'obtenir les échantillons temporels MIC décodés.

FIGURE 8
Décodeur AC-3



BS.1196-08

En réalité, le fonctionnement du décodeur AC-3 est plus complexe que ne l'indique la Fig. 8. En effet:

- en cas de détection d'erreur, le décodeur peut masquer l'erreur ou insérer un silence;
- les canaux dont les contenus à haute fréquence ont été couplés doivent être dégroupés;
- une opération de dématricage doit être exécutée (dans le mode bicanal) lorsque les canaux ont été rematriculés;
- la résolution du banc de filtres de synthèse doit être modifiée de manière dynamique comme l'a été la résolution du banc de filtres d'analyse du codeur pendant le processus de codage.

3 E-AC-3

L'AC-3 amélioré (E-AC-3, *Enhanced AC-3*) ajoute plusieurs outils de codage et fonctions au codec AC-3 de base décrit ci-dessus. Les outils de codage supplémentaires améliorent l'efficacité du codage et permettent un fonctionnement à des débits binaires inférieurs, et les fonctions supplémentaires apportent une plus grande souplesse au niveau applicatif.

Outils de codage supplémentaires:

- Transformée hybride adaptative – Couche supplémentaire appliquée dans le banc de filtres d'analyse/de synthèse pour une résolution spectrale plus fine (1/6 de l'AC-3).
- Traitement du pré-bruit transitoire – Outil supplémentaire pour réduire le pré-bruit transitoire.
- Extension spectrale – Synthèse des composantes hautes fréquences par le décodeur reposant sur des informations auxiliaires créées par le codeur.
- Couplage amélioré – Gestion de la phase et de l'amplitude dans le couplage des canaux.

Fonctions supplémentaires:

- Granularité du débit binaire plus fine.
- Débit binaire maximum supérieur (3 Mbit/s).
- Les sous-flux peuvent acheminer des canaux audio supplémentaires, par exemple des canaux 7.1, ou des pistes de commentaires.

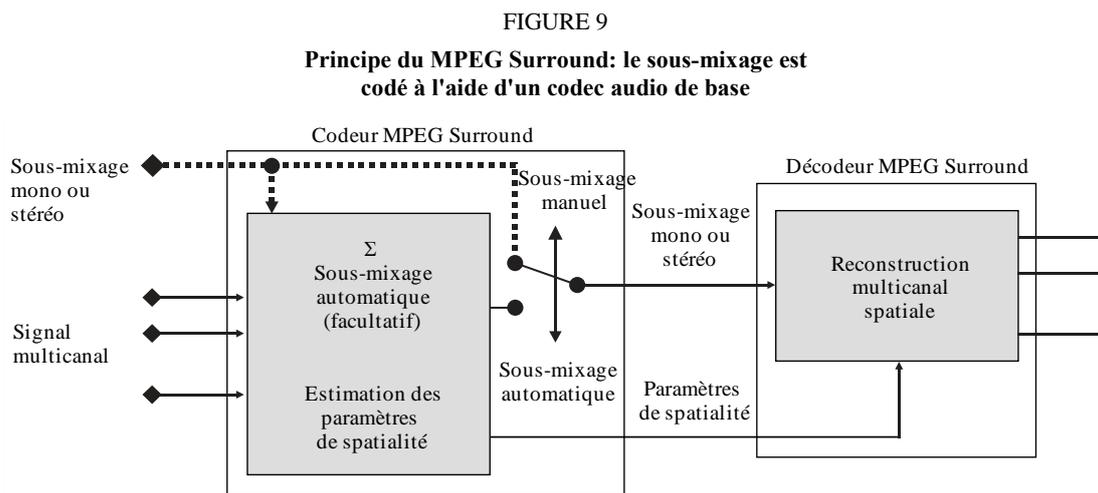
Annexe 4 (pour information)

MPEG Surround

1 Introduction

La norme ISO/CEI 23003-1 ou technologie MPEG Surround est une méthode extrêmement efficace pour le codage du son multicanal, qui permet la transmission du son ambiophonique ou «*surround*» à des débits binaires couramment utilisés pour le codage du son mono ou stéréo. Cette technologie permet de représenter un signal audio multicanal à N canaux grâce à un sous-mixage (*downmix*) $M < N$ et des données de contrôle supplémentaires. Dans les modes de fonctionnement standard, un codeur MPEG Surround crée, à partir d'une entrée audio multicanal, un sous-mixage mono ou stéréo. Le sous-mixage obtenu est codé à l'aide d'un codec audio de base standard, par exemple l'un des systèmes de codage recommandés aux *recommande* 1 et 2. En plus du sous-mixage, le MPEG Surround produit une description paramétrique de l'image spatiale du signal audio multicanal. Cette description est ajoutée au codec audio de base sous la forme d'un flux de données auxiliaires, tout en assurant la rétrocompatibilité. Les décodeurs mono ou stéréo d'ancienne génération ne tiennent pas compte de ces données auxiliaires et lisent le sous-mixage mono ou stéréo. Les décodeurs compatibles MPEG Surround commencent par décoder le sous-mixage mono ou stéréo, puis utilisent les paramètres de description de l'image spatiale extraits du flux de données auxiliaires pour produire un signal audio multicanal de haute qualité.

La Fig. 9 illustre le principe du MPEG Surround.



BS.1196-09

Grâce au MPEG Surround, les services existants peuvent être facilement mis à jour en vue d'offrir un son ambiophonique tout en restant rétrocompatibles. Alors que le décodeur stéréo d'un appareil d'ancienne génération ignore les données relatives au MPEG Surround et joue le signal stéréo sans dégradation de la qualité, un décodeur compatible MPEG Surround fournit un son multicanal de haute qualité.

2 Codage

Le codeur MPEG Surround a pour fonction de représenter un signal d'entrée multicanal sous forme de signal mono ou stéréo rétrocompatible, auquel sont associés des paramètres de spatialité permettant de reconstruire un signal de sortie multicanal qui ressemble aux signaux d'entrée d'origine du point de vue de la perception. Un sous-mixage extérieur (sous mixage dit «artistique») peut être utilisé à la place du sous-mixage produit automatiquement. Le sous-mixage doit préserver les caractéristiques de spatialité du signal sonore en entrée.

Le MPEG Surround repose sur la technologie stéréo paramétrique, qui, combinée au HE AAC, a donné la norme HE AAC v2. En combinant plusieurs modules de stéréo paramétrique et d'autres modules récemment développés, plusieurs structures prenant en charge des combinaisons différentes de nombre d'entrées et de canaux de sous-mixage ont été définies. Par exemple, pour un signal d'entrée multicanal 5.1, trois configurations sont possibles: une configuration pour les systèmes reposant sur le sous-mixage stéréo (configuration 525) et deux configurations différentes pour les systèmes reposant sur le sous-mixage mono (configurations 515₁ et 515₂ utilisant des cascades de modules différentes).

Le MPEG Surround intègre un certain nombre d'outils qui offrent des fonctions permettant une large application de la norme. L'une des fonctions essentielles est la capacité à faire varier la qualité de l'image spatiale de façon progressive, partant d'un surdébit spatial très faible jusqu'à une représentation du son transparente. Autre caractéristique importante, l'entrée du décodeur peut être rendue compatible avec les technologies ambiophoniques matricielles existantes.

Ces fonctions, ainsi que d'autres, sont assurées par les outils de codage suivants, qui occupent une place importante dans cette technologie:

- Codage résiduel: Outre les paramètres de spatialité, des signaux résiduels peuvent être acheminés à l'aide d'une technique de codage hybride. Ces signaux se substituent à une partie des signaux décorrélés (qui font partie des modules de stéréo paramétrique). Les signaux résiduels sont codés en passant du domaine FMQ (filtres miroirs en quadrature) au domaine TCDM (transformée en cosinus discrète modifiée) puis en codant les coefficients TCDM à l'aide de la norme AAC.
- Compatibilité matricielle: De manière facultative, le sous-mixage stéréo peut être prétraité pour le rendre compatible avec les technologies ambiophoniques matricielles d'ancienne génération en vue d'assurer une rétrocompatibilité avec les dispositifs qui ne peuvent décoder que le flux binaire stéréo, mais qui sont équipés d'un décodeur ambiophonique matriciel.
- Signaux de sous-mixage arbitraires: Le système MPEG Surround est capable de gérer, outre les sous-mixages générés par un codeur, les sous-mixages dits «artistiques» fournis au codeur en plus du signal multicanal d'origine.
- MPEG Surround sur MIC: En règle générale, les paramètres de spatialité du MPEG Surround sont acheminés dans la partie du système de compression audio sous-jacent réservée aux données auxiliaires. Dans le cas des applications pour lesquelles le sous-mixage est transmis au format MIC, le MPEG Surround prend également en charge une méthode permettant d'acheminer les paramètres de spatialité sur des canaux audio non compressés. La technologie sous-jacente est désignée sous le nom de «*buried data*» ou «données enfouies».

3 Décodage

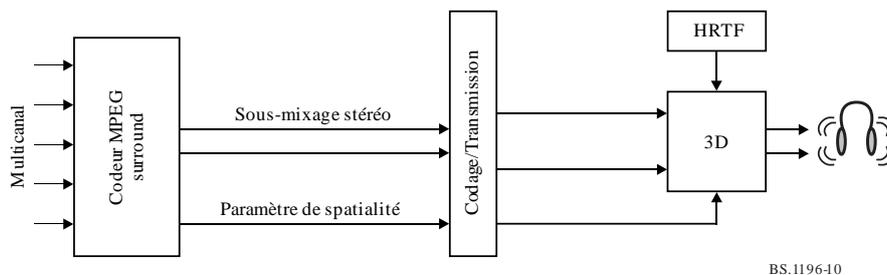
Outre la restitution vers une sortie multicanal, le décodeur MPEG Surround prend en charge la restitution vers d'autres configurations de sortie:

- Surround virtuel: Le système MPEG Surround peut exploiter les paramètres de spatialité pour restituer le sous-mixage vers une sortie surround virtuelle stéréo connectée à un casque

d'ancienne génération. La norme ne spécifie pas la fonction de transfert anatomique de l'individu (HRTF, *head-related transfer function*), mais seulement l'interface avec cette fonction, ce qui permet d'adapter la mise en oeuvre au cas d'utilisation. Le traitement surround virtuel peut être appliqué à la fois dans le décodeur et dans le codeur, ce dernier offrant la possibilité d'un rendu ambiophonique virtuel sur le sous-mixage, sans qu'un décodeur MPEG Surround soit nécessaire. Le décodeur MPEG Surround peut toutefois annuler le traitement surround virtuel du sous-mixage et appliquer un autre traitement surround virtuel. Le principe de base est illustré à la Fig. 10.

FIGURE 10

Décodage surround virtuel du MPEG Surround



BS.1196-10

- Mode matriciel amélioré: Dans le cas d'un signal stéréo d'ancienne génération, le MPEG Surround est capable d'estimer, à partir du sous-mixage, les informations auxiliaires de spatialité non contenues dans le signal et de créer ainsi une sortie multicanal offrant une qualité supérieure au système surround matriciel conventionnel.
- Elagage: Du fait de la structure sous-jacente, le décodeur MPEG Surround peut restituer sa sortie vers des configurations pour lesquelles les canaux sont en nombre inférieur à ceux présents dans l'entrée multicanal du codeur.

4 Profils et niveaux

Le décodeur MPEG Surround peut être mis en oeuvre sous deux versions: l'une de qualité supérieure et l'autre de faible puissance. Bien que les signaux en sortie soient différents, les deux versions fonctionnent sur le même flux binaire.

Le profil de référence du MPEG Surround définit six niveaux hiérarchiques, qui diffèrent par le nombre de canaux en entrée et en sortie, par l'intervalle des fréquences d'échantillonnage et par la largeur de bande de décodage du signal résiduel. Pour garantir un décodage correct, le niveau du décodeur doit être supérieur ou égal au niveau du flux binaire. De plus, les décodeurs de niveau 1, 2 et 3 sont capables de décoder tous les flux binaires de niveau 2, 3 et 4, avec éventuellement une légère baisse de qualité due aux limites du décodeur. Par ailleurs, la qualité et le format de la sortie d'un décodeur MPEG Surround dépendent de sa configuration. A noter que les aspects liés à la configuration du décodeur sont totalement indépendants des différents niveaux définis dans ce profil.

5 Interconnexion avec les codecs audio

Le MPEG Surround fonctionne comme une extension de pré- et post-traitement au-dessus des systèmes de codage audio d'ancienne génération. Il est donc capable de s'adapter à tout codeur audio de base, ou presque. Le MPEG Surround adopte une mise en trames très souple, qui s'adapte à une large gamme de codeurs. Il fournit en outre les moyens d'optimiser la connexion avec les codeurs utilisant déjà des outils paramétriques (par exemple, la reconstruction de bande spectrale).

Annexe 5 (pour information)

Codage AAC à haute efficacité étendu (HE AAC étendu)

1 Introduction

Le profil HE AAC étendu est spécifié dans le cadre de la norme ISO/CEI 23003-3, Discours unifié et codage audio (USAC) MPEG-D. L'USAC est une norme de codage audio qui permet le codage de la parole, de contenus audio et de tout mélange de parole et de contenus audio en garantissant une qualité audio constante pour tous les équipements sonores qui emploient une large gamme de débits binaires. Il prend en charge le codage dans un seul canal ou dans plusieurs canaux, à des débits binaires élevés, et offre une perception «transparente» de la qualité. Il permet également un codage très efficace à des débits binaires très faibles tout en préservant la largeur de bande audio dans sa totalité.

Alors que les codecs audio précédents présentaient des avantages et des inconvénients lors du codage soit de la parole soit de contenus audio, l'USAC peut coder tous les contenus avec la même haute fidélité, indépendamment du type de contenu.

Afin de parvenir à une qualité également bonne pour le codage de contenus audio ou de la parole, l'USAC utilise les techniques éprouvées de codage fondées sur la transformée en cosinus discrète modifiée (TCDM), connues depuis le codage audio MPEG-4 (MPEG-4 AAC, HE AAC, HE AAC v2), et les associe à des éléments spécialisés de codage de la parole tels que la prévision linéaire à excitation par séquence codée à structure algébrique (ACELP). Des outils paramétriques de codage tels que la reconstruction de bande spectrale (SBR) MPEG-4 et le MPEG-D MPEG Surround sont améliorés et étroitement intégrés dans le codec. Le résultat permet un codage très efficace, qui s'effectue jusqu'aux débits binaires les plus faibles.

Actuellement, la norme USAC spécifie deux profils:

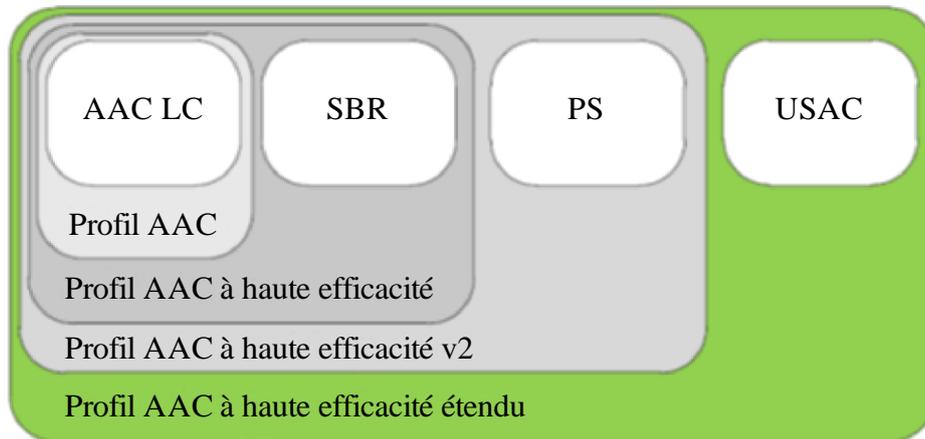
- *Profil de référence de l'USAC*

Le profil de référence de l'USAC fournit la fonctionnalité intégrale de la norme USAC tout en maintenant le degré de complexité des calculs globalement à un faible niveau. Les outils qui nécessitent trop de mémoire ou de puissance de traitement sont exclus.

- *Profil étendu HE AAC*

Ce profil, spécialement destiné aux applications qui doivent rester compatibles avec la famille de profils AAC existants (AAC, HE AAC et HE AAC v2), étend le profil existant HE AAC v2 grâce à l'ajout de capacités USAC. Il inclut le niveau 2 du *profil de référence de l'USAC*. En conséquence, les décodeurs du profil étendu HE AAC peuvent décoder tous les flux binaires HE AAC v2 ainsi que les flux binaires USAC (jusqu'à deux canaux).

FIGURE 11
Structure du profil AAC à haute efficacité étendu



BS.1196-11

L'USAC, qui prend en charge des fréquences d'échantillonnage allant de 7,35 kHz à 96 kHz, a montré qu'il fournissait une bonne qualité audio, pour une gamme de débits binaires allant de 8 kbit/s à des débits binaires, dont la perception est transparente. Cela a été prouvé dans le test de vérification (document MPEG2011/N12232) de l'ISO/CEI JTC 1/SC 29/WG 11 qui est joint au Document 6B/286(Rév.2).

La configuration des canaux peut être choisie librement. Treize configurations de canaux par défaut différentes peuvent être signalées efficacement pour les scénarios d'application les plus courants. Ces configurations par défaut comprennent toutes les configurations de canaux MPEG-4, telles que les configurations mono, stéréo, Surround 5.0 et 5.1, ou encore les configurations de haut-parleurs 7.1 ou 22.2.

2 Codage

Comme cela est courant dans la normalisation MPEG, la norme ISO/IEC 23003-3 ne spécifie que le processus de décodage pour les fichiers et les flux de données MPEG-D USAC. Elle ne contient pas de spécification normative du processus de codage.

Une structure type possible de codeur est représentée à la Fig. 12.

Le codeur comporte les outils de codage suivants:

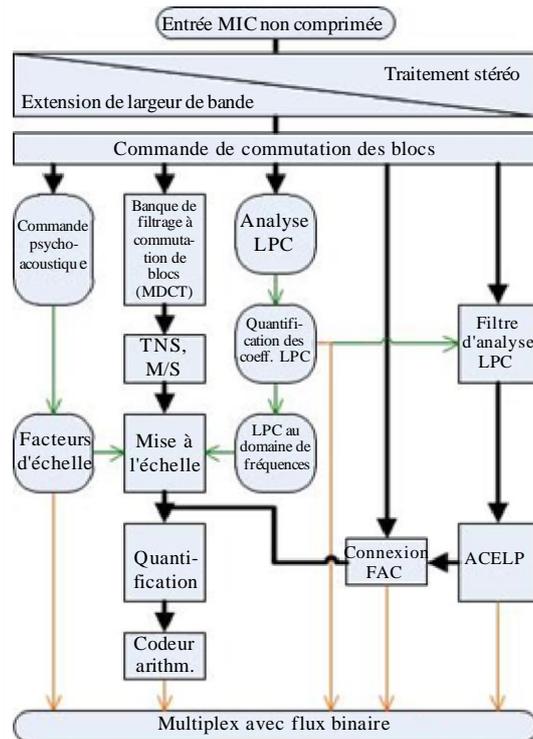
- Traitement stéréo: à des débits binaires faibles/intermédiaires, l'USAC emploie des technologies de codage stéréo paramétriques. Elles sont similaires en principe à l'outil PS décrit dans l'Appendice 2.5 de la norme ISO/CEI 23003-3, mais sont plutôt fondées sur le MPEG Surround décrit dans l'Annexe 4 et donc nommées MPEG Surround 2-1-2 (MPS 2-1-2). Le codeur extrait à partir du signal audio entrant une représentation paramétrique à haute efficacité de l'image stéréo. Ces paramètres sont transmis dans le flux binaire avec un signal de sous-mixage monophonique. En option, le codeur peut choisir d'émettre un signal résiduel qui modifie le processus de reconstruction du signal stéréo au niveau du décodeur. Le mécanisme de codage résiduel permet le passage graduel du codage stéréo entièrement paramétrique au codage stéréo à canaux entièrement discrets. L'outil MPS 2-1-2 est un élément intrinsèque du codec USAC. A des débits binaires plus élevés, auxquels le codage paramétrique et la prévision ACELP ne sont généralement pas actifs, le

codage stéréo peut être effectué uniquement dans le domaine TCDM au moyen d'une prévision stéréo à valeur complexe. C'est pourquoi cette méthode est appelée codage stéréo à prévision complexe. Il peut être considéré comme une généralisation du codage stéréo M/S classique.

- Extension de la largeur de bande: l'extension paramétrique de la largeur de bande est une version plusieurs fois améliorée de la reconstruction de bande spectrale (SBR) MPEG-4, qui est décrite dans l'Appendice 2.4 de la norme ISO/CEI 23003-3. Le codeur évalue l'enveloppe spectrale et la tonalité des bandes de fréquence audio plus élevées et transmet les paramètres correspondants au décodeur. Le codeur peut choisir entre deux types de réémetteurs différents (réémission en harmonie ou copie améliorée) et entre trois facteurs de réémission (1:2, 3:8, 1:4). L'outil SBR amélioré est un élément intrinsèque du codec USAC.
- Banc de filtrage, commutation des blocs: un banc de filtrage employant la transformée TCDM est l'élément essentiel sur lequel est fondé le codeur de base. En fonction du mécanisme de mise en forme du bruit de quantification, la résolution de la transformée peut être choisie parmi celles qui correspondent à 1024, 512, 256 ou 128 raies spectrales. En association avec le facteur de réémission SBR 3:8, la résolution peut être modifiée et ne correspondre qu'à $\frac{3}{4}$ des possibilités mentionnées ci-dessus, offrant une meilleure granularité temporelle, même à des taux d'échantillonnage plus faibles.
- Mise en forme du bruit dans le domaine temporel (TNS), codage stéréo M/S, quantification: ces outils ont été repris du codage AAC et sont employés de manière analogue à celle qui figure à l'Appendice 2.2 de la norme ISO/CEI 23003-3.
- Codeur arithmétique adaptatif en fonction du contexte: le codage sans bruit (c'est-à-dire entropique) des coefficients spectraux de la transformée TCDM est traité par un codeur arithmétique qui choisit ses tables de probabilité sur la base de raies spectrales codées précédemment.
- Contrôle psychoacoustique, mise à échelle au moyen d'un facteur: le modèle psychoacoustique fondé sur un facteur d'échelle est semblable à celui qui est utilisé dans le codage AAC (voir l'Appendice 2.2 de la norme ISO/CEI 23003-3).
- Mise à l'échelle sur la base de paramètres de codage prédictif linéaire (LPC): cet outil de mise en forme du bruit dans le domaine spectral peut être utilisé comme remplacement de l'outil susmentionné de mise à échelle au moyen d'un facteur. La version pondérée d'une représentation en termes de fréquence d'un ensemble de coefficients de filtrage LPC est appliquée aux coefficients spectraux de la transformée TCDM avant la quantification et le codage.
- Prévision ACELP: l'outil de codage ACELP utilise la représentation d'excitation du répertoire de séquences codées adaptatives/innovantes éprouvées que l'on connaît grâce aux codecs de la parole à la pointe de la technologie.
- Multiplex avec flux binaire: le flux binaire final est composé de divers éléments produits par les outils de codage.
- Correction FAC: l'outil de correction du repliement vers l'avant (FAC) fournit un mécanisme permettant de passer sans heurts d'un codage employant une transformée TCDM victime de repliement au codage ACELP dans le domaine temporel.

FIGURE 12

Schéma fonctionnel d'un codeur MPEG-D USAC



BS.1196-12

3 Décodage

La structure de base du décodeur MPEG-D USAC est représentée à la Fig. 13. Le processus de décodage consiste généralement en l'inverse du processus de codage.

4 Profils et niveaux

La norme MPEG définit actuellement deux profils qui emploient le codec USAC.

– *Profil de référence USAC*

Le profil de référence USAC contient le codec USAC complet à l'exception d'un nombre restreint d'outils pour lesquels la complexité de calcul dans le cas le plus défavorable est excessive. Ces outils ne sont pas décrits ci-dessus. Ce profil est un profil clair et autonome pour les applications et cas d'utilisation où la capacité de prise en charge de la famille de profils AAC (profil AAC, profils HE AAC, profil HE AAC v2) ne convient pas.

– *Profil HE AAC étendu*

Le profil HE AAC étendu contient tous les outils du profil AAC v2 à haute efficacité et est en tant que tel en mesure de décoder tous les flux de profils de la famille AAC. En outre, le profil intègre la capacité mono/stéréo du profil de référence USAC. En conséquence, ce profil constitue une évolution naturelle du profil HE AAC v2 parce que la partie mono/stéréo de l'USAC (lorsqu'il fonctionne à des débits faibles) fournit la valeur ajoutée d'une qualité de fonctionnement constante pour tous les types de contenus à des débits binaires faibles.

Annexe 6 (pour information)

Points de code indépendants du codage (CICP) pour le codage MPEG

1 Introduction

La norme ISO/CEI 23001-8:2013 décrit les aspects liés au codage des programmes audio qui sont indépendants de la représentation codée y compris la position et la disposition des systèmes de haut-parleurs. Les configurations de canaux par défaut comprennent les configurations de canaux spécifiées dans les Recommandations UIT-R BS.775 et UIT-R BS.2051. Le Tableau 3 donne toutes les configurations de canaux.

TABLEAU 3

Configurations de canaux et dispositions des haut-parleurs (Note 1)

Valeurs de configuration de canaux* ¹ (Note 1)	Nombre de haut-parleurs (Note 2)	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) (Note 3)	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 (Note 4)
0	Toutes les configurations		
1	1/0.0 (0+1+0)	M+000	Mono
2	2/0.0 (0+2+0)	M+030	Gauche
		M-030	Droite
3	3/0.0 (0+3+0)	M+000	Centre
		M+030	Gauche
		M-030	Droite

TABLEAU 3 (suite)

Valeurs de configuration de canaux* ¹ (Note 1)	Nombre de haut-parleurs (Note 2)	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) (Note 3)	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 (Note 4)
4	3/1.0 (0+4+0)	M+000	Centre
		M+030	Gauche
		M-030	Droite
		M+180	Ambiance mono
5	3/2.0 (0+5+0)	M+000	Centre
		M+030	Gauche
		M-030	Droite
		M+110	Ambiophonique gauche
6	3/2.1 (0+5+0)	M-110	Ambiophonique droit
		M+000	Centre
		M+030	Gauche
		M-030	Droite
		M+110	Ambiophonique gauche
		M-110	Ambiophonique droit
7	5/2.1 (0+7+0)	LFE1	Effets basse fréquence
		M+000	N/D* ²
		M+030	
		M-030	
		M+045	
		M-045	
		M+110	
		M-110	
LFE1			
8	1+1	Canal 1	N/D
		Canal 2	
9	2/1.0 (0+3+0)	M+030	Gauche
		M-030	Droite
		M+180	Ambiance mono
10	2/2.0 (0+4+0)	M+030	Gauche
		M-030	Droite
		M+110	Ambiophonique gauche
		M-110	Ambiophonique droit
11	3/3.1 (0+6+0)	M+000	N/D
		M+030	
		M-030	
		M+110	
		M-110	
		M+180	
LFE1			

TABLEAU 3 (suite)

Valeurs de configuration de canaux* ¹ (Note 1)	Nombre de haut-parleurs (Note 2)	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) (Note 3)	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 (Note 4)
12	3/4.1 (0+7+0)	M+000	Centre
		M+030	Gauche
		M-030	Droit
		M+090	Ambiophonique latéral gauche
		M-090	Ambiophonique latéral droit
		M+135	Ambiophonique arrière gauche
		M-135	Ambiophonique arrière droit
		LFE1	Effets basse fréquence
13	11/11.2 (9+10+3)	M+000	Avant central
		M+030	Avant gauche central
		M-030	Avant gauche central
		M+060	Avant gauche
		M-060	Avant droit
		M+090	Latéral gauche
		M-090	Latéral droit
		M+135	Arrière gauche
		M-135	Arrière droit
		M+180	Arrière central
		LFE1	Effets basse fréquence -1
		LFE2	Effets basse fréquence -2
		U+000	Supérieur avant central
		U+045	Supérieur avant gauche
		U-045	Supérieur avant droit
		U+090	Supérieur latéral gauche
		U-090	Supérieur latéral droit
		T+000	Supérieur central
		U+135	Supérieur arrière gauche
		U-135	Supérieur arrière droit
		U+180	Supérieur arrière central
B+000	Inférieur avant central		
B+045	Inférieur avant gauche		
U-045	Inférieur avant droit		
14	5/2.1 (2+5+0)	M+000	Centre
		M+030	Gauche
		M-030	Droite
		M+110	Ambiophonique gauche
		M-110	Ambiophonique droit
		LFE1	Effets basse fréquence
		U+030	Supérieur avant gauche
		U-030	Supérieur avant droit

TABLEAU 3 (suite)

Valeurs de configuration de canaux* ¹ (Note 1)	Nombre de haut-parleurs (Note 2)	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) (Note 3)	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 (Note 4)
15	5/5.2 (3+7+0)	M+000	Centre
		M+030	Gauche
		M-030	Droite
		M+090	Latéral gauche
		M-090	Latéral droit
		M+135	Arrière gauche
		M-135	Arrière droit
		U+045	Supérieur gauche
		U-045	Supérieur droit
		UH+180	Supérieur central
		LFE1	Effets basse fréquence
		LFE2	Effets basse fréquence
16	5/4.1 (4+5+0)	M+000	Centre
		M+030	Gauche
		M-030	Droite
		M+110	Ambiophonique gauche
		M-110	Ambiophonique droit
		LFE1	Effets basse fréquence
		U+030	Supérieur avant gauche
		U-030	Supérieur avant droit
		U+110	Supérieur arrière gauche
		U-110	Supérieur arrière droit
17	6/5.1 (6+5+0)	M+000	N/D
		M+030	
		M-030	
		M+110	
		M-110	
		LFE1	
		U+000	
		U+030	
		U-030	
		U+110	
		U-110	
		T+000	

TABLEAU 3 (suite)

Valeurs de configuration de canaux* ¹ (Note 1)	Nombre de haut-parleurs (Note 2)	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) (Note 3)	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 (Note 4)
18	6/7.1 (6+7+0)	M+000	N/D
		M+030	
		M-030	
		M+110	
		M-110	
		M+150	
		M-150	
		LFE1	
		U+000	
		U+030	
		U-030	
		U+110	
		U-110	
T+000			
19	5/6.1 (4+7+0)	M+000	Centre
		M+030	Gauche
		M-030	Droit
		M+090	Ambiophonique latéral gauche
		M-090	Ambiophonique latéral droit
		M+135	Ambiophonique arrière gauche
		M-135	Ambiophonique arrière droit
		LFE	Effets basse fréquence
		U+045	Supérieur avant gauche
		U-045	Supérieur avant droit
		U+135	Supérieur arrière gauche
		U-135	Supérieur arrière droit
20	7/6.1 (4+9+0)	M+000	Centre
		M+SC	Gauche de l'écran
		M-SC	Droite de l'écran
		M+030	Gauche
		M-030	Droite
		M+090	Ambiophonique latéral gauche
		M-090	Ambiophonique latéral droit
		M+135	Ambiophonique arrière gauche
		M-135	Ambiophonique arrière droit
		LFE	Effets basse fréquence
		U+045	Supérieur avant gauche
		U-045	Supérieur avant droit
		U+135	Supérieur arrière gauche
U-135	Supérieur arrière droit		

TABLEAU 3 (*fin*)

Valeurs de configuration de canaux* ¹ (Note 1)	Nombre de haut-parleurs (Note 2)	Correspondance par défaut entre l'élément de syntaxe et le(s) haut-parleur(s) (Note 3)	Nom de canal spécifié dans la Recommandation UIT-R BS.775 ou UIT-R BS.2051 (Note 4)
21-63	Réservé		

*¹ La configuration de canaux de sortie audio est indiquée par un champ de six bits qui achemine une valeur de configuration de canaux définie dans la norme ISO/CEI 23001-8:2013 «Points de code indépendants du codage».

*² N/D: non disponible; configuration de canaux ne figurant pas dans la Recommandation UIT-R BS.2051 ou UIT-R BS.775.

NOTE 1 – La liste est tirée du Tableau 8 de la norme ISO/CEI 23001-8:2013/Amd.1:2015.

NOTE 2 – La notion de nombre de haut-parleurs est donnée selon la convention «Haut-parleurs avant/haut-parleurs ambiophoniques. Haut-parleurs LFE» et entre crochets «Haut-parleurs couche supérieure + haut-parleurs couche intermédiaire + haut-parleurs couche inférieure», les haut-parleurs LFE étant exclus.

NOTE 3 – Identification des haut-parleurs par des étiquettes conformément à la Recommandation UIT-R BS.2051.

NOTE 4 – L'étiquette et le nom d'un canal dépendent de la configuration effective du canal.

Annexe 7 (pour information)

AC-4

1 Introduction

AC-4 est un système de codage évolué pour la radiodiffusion numérique qui utilise un algorithme de compression numérique et divers outils de codage paramétrique dont le but est d'améliorer l'efficacité et la fonctionnalité. AC-4 prend en charge, en natif, les formats d'entrée et de sortie basés sur un canal et basés sur un objet.

Le Tableau 4 fournit une liste des formats de canaux que le système AC-4 prend en charge en natif, ce qui comprend les configurations de canal requises pour l'émission définies dans la Recommandation UIT-R BS.1548. AC-4 prend aussi en charge le codage de formats de nombre de canaux plus élevé qui sont éventuellement utilisés pour mettre en œuvre les systèmes audio évolués de la Recommandation UIT-R BS.2051.

TABLEAU 4

Résumé des formats de canal pris en charge en natif

Format	Nombre de canaux	Note
Mono (format 1/0 ⁽¹⁾)	1	
Stéréo (format 2/0 ⁽¹⁾)	2	
3.0 (format 3/0 ⁽¹⁾)	3	
5.0/5.1 (format 3/2 ⁽¹⁾)	5/6	
7.0/7.1 (Système I ⁽²⁾)	7/8	Signalisation de trois configurations de haut-parleur différentes
7.0.4/7.1+4 (Système J ⁽²⁾)	11/12	Signalisation de sous-ensembles contenant un plus petit nombre de canaux
9.0.4/9.1.4	13/14	Signalisation de sous-ensembles contenant un plus petit nombre de canaux
22.2 (Système H ⁽²⁾)	24	

⁽¹⁾ Spécifié dans la Recommandation UIT-R BS.775.

⁽²⁾ Spécifié dans la Recommandation UIT-R BS.2051.

Le système AC-4 peut coder de 1 à 22+2 canaux de source audio MIC en un train binaire série dont le débit est compris entre 24 kbit/s et 1 536 kbit/s. Outre les représentations basées sur un canal, AC-4 prend aussi en charge le codage des objets audio dynamiques. On trouvera une description complète de la syntaxe du train binaire AC-4 dans le document ETSI TS 103 190-2.

Le Tableau 5 propose des débits binaires pour diverses configurations de canaux qui répondent aux exigences de la Recommandation UIT-R BS.1548.

TABLEAU 5

Résumé des débits binaires applicables à certaines configurations de canaux

Mode	Débits binaires
2.0 Stéréo	96 kbit/s
5.1 ambiophonique	192 kbit/s
22.2 ambiophonique	1 536 kbit/s

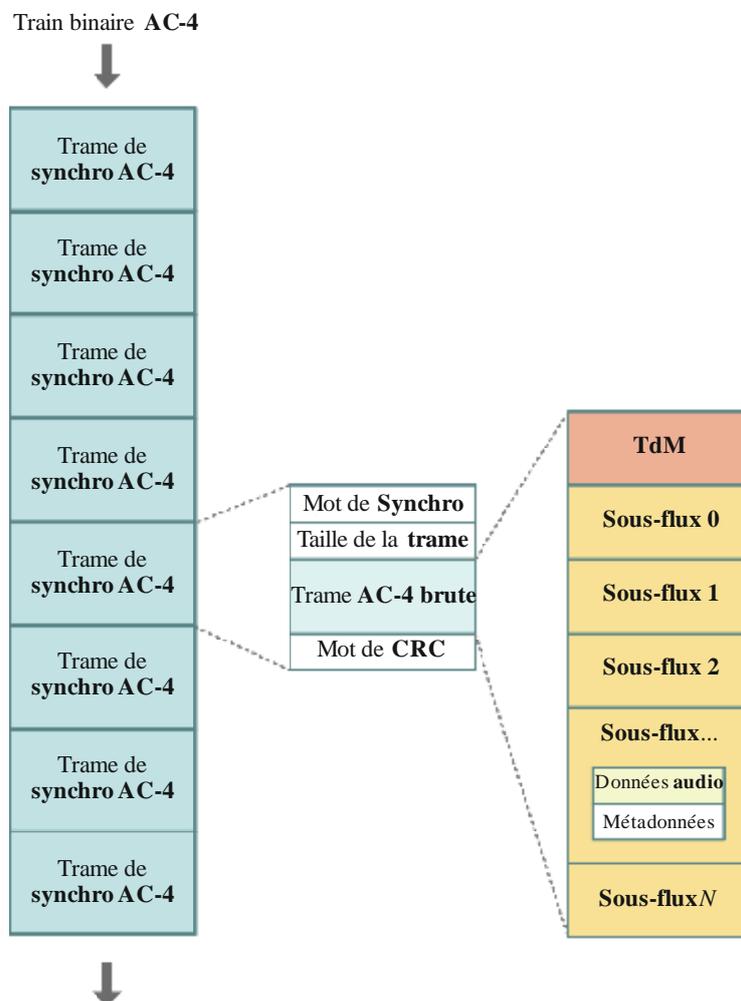
AC-4 prend également en charge en natif les fonctions suivantes relatives au système:

- Gestion intelligente du niveau sonore conforme aux Recommandations UIT-R BS.1770 et UIT-R BS.1771, y compris la signalisation permettant d'indiquer la conformité avec les réglementations internationales et plusieurs réglementations régionales en vigueur en matière de niveau sonore.
- Prise en charge du codage et du décodage des représentations audio basées sur un canal et basées sur un objet.

- Prise en charge des environnements d'écoute tel qu'exigé dans la Recommandation UIT-R BS.1909, à savoir les «environnements habitation» et les «environnements mobiles». Y compris la régulation évoluée de la dynamique applicable à un large éventail de types d'appareils pour les environnements «habitation» et «mobiles».
- Amélioration du dialogue.
- Codage synchrone des trames vidéo, qui permet d'aligner les trames audio et les trames vidéo.
- Prise en charge, en natif, de l'acheminement et de la signalisation de données auxiliaires évoluées ou de métadonnées.

Le train binaire AC-4, illustré à la Fig. 14, est composé de trames de synchronisation AC-4, qui commencent par un mot de synchronisation et se terminent par un mot de contrôle de redondance cyclique (CRC). Le mot de synchronisation permet au décodeur d'identifier facilement le début d'une trame AC-4, tandis que le mot de CRC lui permet de détecter d'éventuelles erreurs dans le train binaire et d'effectuer si besoin un masquage d'erreur. La trame du codec proprement dite ou trame «AC-4 brute» consiste en une TdM (table des matières) et au moins un sous-flux.

FIGURE 14
Syntaxe du train binaire de haut niveau



RS 1196-14

Chaque sous-flux contient les données audio codées ainsi que les métadonnées (données auxiliaires) associées. La TdM contient les informations nécessaires pour décoder le ou les sous-flux.

2 Codage

Le codeur AC-4 n'est pas spécifié par une norme, mais par toute une série d'outils de codage qui sont pris en charge de façon à ce qu'un train binaire codé conforme puisse être généré.

Le système AC-4 utilise une quantification et un codage dans le domaine transformée au moyen d'une transformée en cosinus discrète modifiée (TCDM) avec commutation de fenêtres flexible et codage paramétrique dans un banc de filtres miroirs en pseudo-quadrature. AC-4 prend en charge le codage de 1 à 22+2 canaux audio MIC en un train binaire série à des débits compris entre 24 kbit/s et 1 536 kbit/s. Pour le codage de canaux stéréo 2.0 et 5.1, les débits binaires de 96 kbit/s et 192 kbit/s répondent aux exigences de qualité de fonctionnement spécifiées dans la Recommandation UIT-R BS.1548. AC-4 prend en charge trois modes d'allocation de bits: débit binaire constant, débit binaire moyen et débit binaire variable.

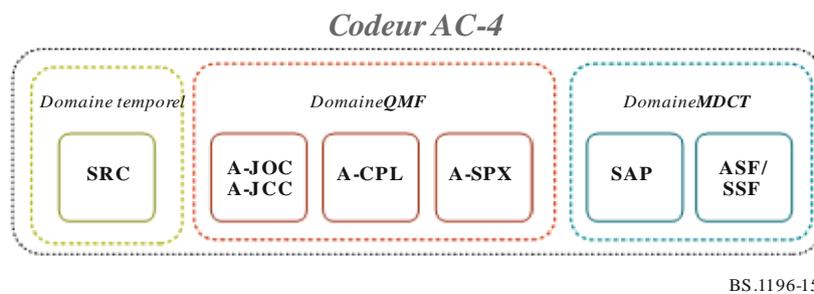
Le codeur AC-4 peut être mis en œuvre au moyen de divers outils de codage (voir la Fig. 15) pour améliorer l'efficacité et la souplesse/fonctionnalité en fonction du mode d'opération et/ou de l'application. L'ordre dans lequel les outils sont mis en œuvre va de gauche à droite: un signal audio MIC est présenté en entrée aux outils situés à gauche, et un train binaire codé AC-4 est produit en sortie, à droite. Description des outils de codage:

- SRC: Le codeur AC-4 doit intégrer un convertisseur de taux d'échantillonnage (SRC, *Sample Rate Converter*) pour prendre en charge une durée de trame qui coïncide avec une trame vidéo dans les modes de codage de trame synchrones. Selon le débit des trames, le signal d'entrée est converti en l'un des taux d'échantillonnage internes (46 034 Hz, 46 080 Hz, 48 000 Hz ou 51 200 Hz) utilisés par les outils de codage FMQ et TCDM qui suivent.
- A-JCC: L'outil de codage de canal conjoint évolué (A-JCC, *Advanced Joint Channel Coding*) effectue un sous-mixage des canaux immersifs présentés en entrée (plus de 5 canaux) vers un nombre inférieur de canaux, et code les paramètres associés. Les paramètres permettent la reconstruction de la totalité du canal d'entrée dans le décodeur.
- A-JOC: L'outil de codage d'objet conjoint évolué (A-JOC, *Advanced Joint Object Coding*) prend en entrée des objets audio et les code spatialement pour produire un nombre plus petit d'objets en sortie, réduisant ainsi le nombre de signaux codés TCDM. Les paramètres qui sont codés avec le nombre réduit d'objets de sortie permettent la reconstruction des objets dans le décodeur.
- A-CPL: L'outil de couplage évolué (A-CPL, *Advanced Coupling*) effectue un sous-mixage de 2 canaux en 1 canal, et code les paramètres associés qui permettent la reconstruction du signal d'entrée original à 2 canaux.
- A-SPX: L'outil d'extension spectrale évoluée (A-SPX, *Advanced Spectral Extension*) code les paramètres associés au contenu haute fréquence du signal d'entrée, puis procède à une extension spectrale dans le décodeur. Les paramètres sont constitués de l'enveloppe, de la tonalité et des mesures de bruit. Les résolutions spectrale et temporelle des paramètres peuvent être adaptées aux caractéristiques du signal d'entrée.
- SAP: L'outil de traitement audio stéréo (SAP, *Stereo Audio Processing*) effectue un codage de canal conjoint dans le domaine TCDM entre deux canaux d'entrée ou plus.
- ASF: L'outil d'étage d'entrée spectral audio (ASF, *Audio Spectral Front-end*) est un outil de quantification et de codage basé sur la transformée TCDM, qui utilise la commutation de fenêtre de transformée. Le module de commande de la commutation de fenêtre choisit la longueur de transformée optimale pour une trame en fonction du type de signal d'entrée. Les coefficients de la TCDM et des informations de commande additionnelles sont mémorisés après la quantification non linéaire et le codage sans bruit vers le train binaire. La répartition binaire dans le temps et la fréquence est contrôlée par un tampon de bits et un modèle de

perception. Le modèle de tampon binaire tient compte des bits utilisés par les autres outils du codeur et des métadonnées générales.

- SSF: L'outil d'étage d'entrée spectral de la parole (SSF, *Speech Spectral Front-end*) est un outil alternatif de quantification et de codage basé sur la transformée TCDM, qui est spécifique au codage de la parole au moyen de transformées courtes. Il effectue une quantification et un codage dans le domaine TCDM au moyen d'un prédicteur en sous-bandes. Les outils SSF et ASF s'excluent mutuellement; autrement dit, pour coder la transformée TCDM, on utilise soit l'outil SSF, soit l'outil ASF.

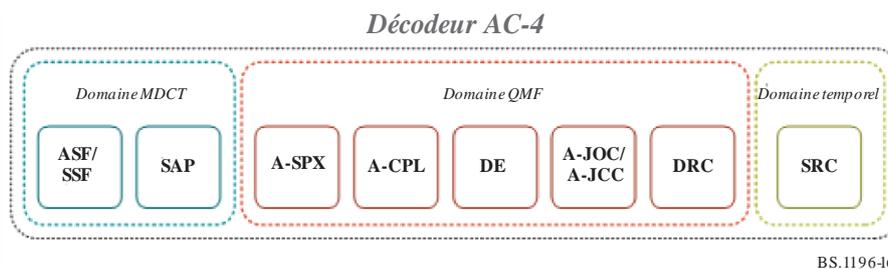
FIGURE 15
Outils de codage disponibles dans le système AC-4



3 Décodage

Le décodage est pratiquement le processus inverse du codage, ainsi que l'illustre la Fig. 16.

FIGURE 16
Outils de décodage



Le train binaire AC-4 et la conception du décodeur prennent en charge, de façon native, la mise en œuvre de décodeurs de moindre complexité afin d'être compatibles avec des dispositifs aux capacités limitées (téléphones portables/tablettes par exemple).

Deux méthodes sont utilisées pour mettre en œuvre cette fonctionnalité:

- 1) **Décodage de base (*Core Decoding*):** AC-4 prend en charge deux modes de décodage: le décodage complet (*Full Decoding*) et le décodage de base (*Core Decoding*). Le décodeur AC-4 peut fonctionner dans un mode de décodage de base, dans lequel un sous-ensemble de base des canaux du programme codé (représentant le contenu de tous les canaux d'entrée) est décodé et représente une reproduction compatible du programme avec une complexité calculatoire réduite.

- 2) Décodage échelonnable: Le système AC-4 permet la dimensionnabilité de la fréquence d'échantillonnage; il peut ainsi prendre en charge des fréquences d'échantillonnage plus élevées, en particulier 96 kHz et 192 kHz, de manière échelonnable. Les dispositifs qui ne gèrent que des sorties à 48 kHz se contentent de décoder la couche de base.

Outre la gestion du décodage du format de canal pris en charge de façon native (voir le Tableau 4), le système AC-4 prend en charge l'utilisation des données auxiliaires ou des métadonnées. Cela permet le sous-mixage de la sortie du canal décodé, depuis un nombre de canaux supérieur vers un nombre de canaux de sortie inférieur, comme cela est requis pour le dispositif de manière prévisible.

Annexe 8 (pour information)

Profil audio 3D MPEG-H de faible complexité

1 Introduction

L'audio 3D MPEG-H est une norme de codage audio conçue pour prendre en charge le codage audio sous la forme de canaux audio, d'objets audio ou d'une composante HOA (*Higher Order Ambisonics*). Cette norme propose des solutions pour la normalisation du niveau sonore et la commande de plage dynamique. Les types de contenu (canaux, objets ou HOA) peuvent être utilisés seuls ou en combinaison. L'utilisation de groupes de canaux audio, d'objets ou de HOA permet l'interactivité ou la personnalisation d'un programme, par exemple la sélection de différentes pistes de langue ou l'ajustement du gain ou de la position des objets au cours de la restitution dans le décodeur MPEG-H.

La spécification audio 3D MPEG-H est publiée dans la norme ISO/CEI 23008-3:2015. L'amendement 3, dans lequel sont spécifiés le profil de faible complexité (profil LC) de la norme audio 3D MPEG-H ainsi que des techniques additionnelles, a été publié début 2017.

Le profil LC de l'audio 3D MPEG-H peut prendre en charge jusqu'à 24 haut-parleurs en sortie et 56 canaux de base codec (dont 28 peuvent être décodés en même temps).

Exemples de configurations de haut-parleurs cibles:

- Stéréo 2.0 (format 2/0 spécifié dans la Recommandation UIT-R BS.775).
- Audio à canaux multiples 5.1 (format 3/2 spécifié dans la Recommandation UIT-R BS.775).
- Audio en immersion 10.2 (système F spécifié dans la Recommandation UIT-R BS.2051).
- Audio en immersion 22.2 (système H spécifié dans la Recommandation UIT-R BS.2051).

Cette norme peut être utilisée pour un large éventail d'applications, parmi lesquelles le stockage et la transmission de son ambiophonique et stéréo. Elle prend en charge l'interactivité et le son en immersion, ce qui est important pour satisfaire les exigences de la diffusion de média de prochaine génération, en particulier les nouveaux systèmes de radiodiffusion visuelle et les services de diffusion de divertissement en continu ainsi que les contenus et les services de réalité virtuelle.

Par exemple, dans la radiodiffusion visuelle, les commentaires ou les dialogues peuvent être envoyés sous forme d'objets audio, et combinés avec un socle de canaux immersifs dans le décodeur audio 3D MPEG-H. Cela permet de transmettre de façon efficace des dialogues en plusieurs langues et d'offrir à l'auditeur la possibilité de régler l'équilibre entre les dialogues et les autres éléments sonores, à sa convenance. Ce concept peut être étendu à d'autres éléments qui ne sont normalement pas présents

dans un signal de radiodiffusion, par exemple une description audio pour déficients visuels, les commentaires du metteur en scène, ou le dialogue des participants à un événement sportif.

Le profil LC de l'audio 3D MPEG-H prend en charge la gestion du niveau sonore conforme aux Recommandations UIT-R BS.1770 et UIT-R BS.1771, y compris la signalisation permettant d'indiquer la conformité avec les réglementations internationales et plusieurs réglementations régionales en matière de niveau sonore. Il prend également en charge la régulation évoluée de la dynamique applicable à un large éventail de types d'appareils pour les environnements «habitation» et «mobiles».

2 Codage

L'architecture de codec audio 3D MPEG-H est bâtie sur un codec perceptuel pour la compression des différentes classes de signaux d'entrée, qui repose sur le système «discours unifié et codage audio (USAC) MPEG». Le système USAC permet de compresser des signaux audio monocanal ou multicanaux, à des débits de 8 kbit/s par canal et plus.

Pour satisfaire aux nouvelles exigences qui sont apparues dans le contexte de l'audio 3D, cette technologie a été étendue pour inclure des outils qui exploitent spécifiquement les effets perceptuels de la reproduction 3D et améliorent ainsi l'efficacité du codage, par exemple:

- Le remplissage de trous intelligent (IGF, *Intelligent Gap Filling*) offre un remplissage de bruit amélioré. IGF est un outil qui restaure, de façon paramétrique, des portions du spectre émis au moyen d'informations adaptées issues de pavés spectraux qui sont adjacents en fréquence et en temps. L'assignation et le traitement de ces pavés spectraux sont contrôlés par le codeur sur la base d'une analyse du signal d'entrée. Ici, les trous spectraux peuvent être remplis avec des coefficients spectraux qui, d'un point de vue perceptuel, affichent une meilleure concordance que les séquences de bruit pseudo-aléatoire du remplissage de bruit conventionnel.
- Outre les améliorations en termes d'efficacité de codage, le noyau USAC-3D comporte de nouveaux mécanismes de signalisation pour la configuration du contenu/des haut-parleurs 3D et pour le type de signal acheminé dans le flux compressé (canal audio/objet audio/signal HOA).

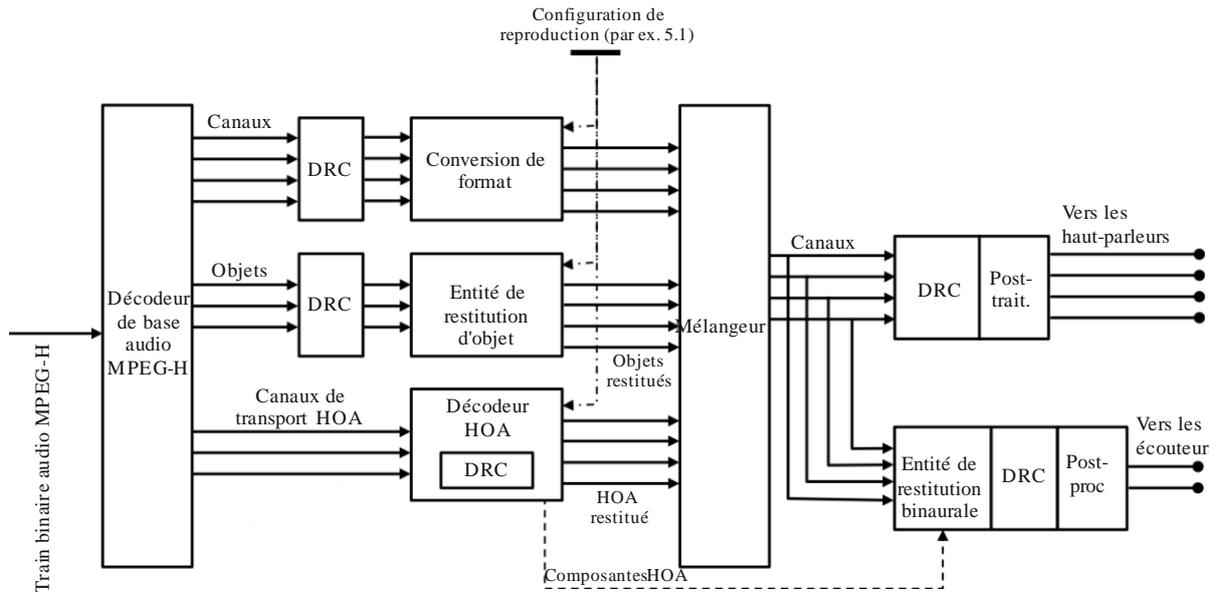
Un nouvel aspect de la conception de la charge utile audio comprimée réside dans l'amélioration du comportement, dans le cas de la commutation de débit instantanée ou de repérage rapide (*fast cue-in*) telle qu'elle apparaît dans le contexte de la diffusion en continu adaptative dynamique MPEG (DASH, *Dynamic Adaptive Streaming*). À cette fin, des trames dites «trames de diffusion immédiates» ont été ajoutées à la syntaxe; elles permettent une transition sans trou d'un flux à l'autre. Ce principe est particulièrement avantageux dans le cas de la diffusion adaptative sur réseaux IP.

3 Décodage

Un diagramme fonctionnel de plus haut niveau est présenté à la Fig. 17. Il illustre l'architecture du décodeur audio de profil LC pour l'audio 3D MPEG-H.

FIGURE 17

**Diagramme fonctionnel de plus haut niveau de l'architecture du
décodeur audio de profil LC pour l'audio 3D MPEG-H**



BS.1196-17

Les principaux composants sont un décodeur de base appelé USAC-3D, un ensemble de systèmes de restitution pour les différentes classes de signal et un mélangeur. Dans une première étape, les différents signaux de base sont convertis à partir de leur représentation en données comprimées au moyen d'un système appelé décodeur USAC-3D.

Les différentes classes de signaux (formes d'onde pour des signaux basés sur un canal et des signaux basés sur un objet ou des signaux basés sur des coefficients HOA) sont ensuite présentées à l'entrée de leurs systèmes de restitution respectifs, lesquels font correspondre ces signaux aux alimentations des haut-parleurs pour la configuration de reproduction particulière qui est disponible du côté du récepteur. Dès que tous les signaux restitués sont disponibles dans le format de reproduction, ils sont combinés dans une étape de mixage pour former le signal d'attaque d'un haut-parleur. Si une représentation binaurale est demandée, le signal est converti en une scène 3D virtuelle en vue de sa reproduction dans des écouteurs. Il est possible de transmettre n'importe quelle combinaison des différents types de signaux dans un flux MPEG-H unique, par exemple une combinaison de signaux basés sur un canal avec des signaux basés sur un objet ou une scène HOA avec des objets.

Les systèmes de restitution sont les suivants:

- Un convertisseur de format destiné à convertir les signaux basés sur un canal depuis leur format de haut-parleur de production vers la configuration de haut-parleur de reproduction.
- Un système de restitution d'objets destiné à placer des pistes basés sur un objet statiques ou dynamiques dans la configuration de reproduction.
- Un convertisseur HOA destiné à convertir la représentation HOA basée sur une scène en une configuration de reproduction réelle.
- Un système de restitution binaurale destiné à convertir une configuration de haut-parleurs virtuels en une sortie pour écouteurs.

De plus, la lecture et la restitution des différentes classes de signal peuvent être contrôlées par une interface utilisateur, si les métadonnées correspondantes indiquent que l'interactivité est disponible sur ces signaux.

Annexe 9 (pour information)

Système DTS-UHD

1 Introduction

Le système de codage DTS-UHD comprend un moteur de codage audio (ACE) qui offre une méthode efficace pour compresser les formes d'onde audio. Il prend en charge le codage des signaux audio basés sur un canal (CBA), des champs sonores ambisoniques et des signaux audio basés sur un objet (OBA). Le système DTS-UHD est défini dans la norme ETSI TS 104 491 et fait partie de la spécification ETSI TS 101 154 sur les flux de transport DVB MPEG 2. Le système offre une sensation d'immersion totale et permet, grâce à l'utilisation d'objets, de canaux, de champs sonores ou d'une combinaison de ces trois éléments tous ensemble, une interactivité et une personnalisation pour une expérience encore meilleure. Il assure en outre la prise en charge des services d'accessibilité pour les utilisateurs malvoyants ou malentendants.

Le système DTS-UHD prend en charge toutes les configurations de haut-parleurs spécifiées dans la Recommandation UIT-R BS.2051.

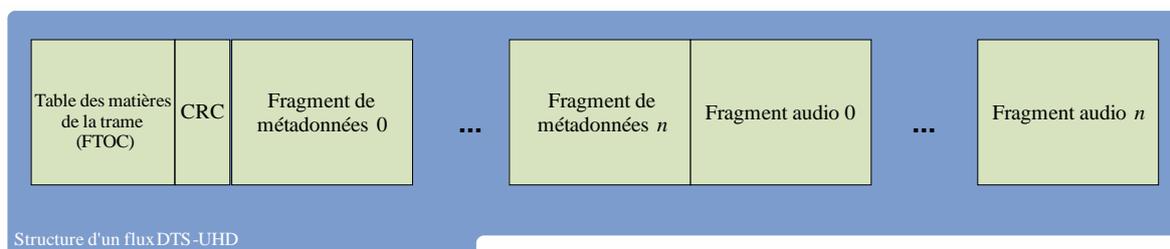
Tout en prenant pleinement en charge les signaux audio basés sur des canaux, le système DTS-UHD est aussi capable de coder jusqu'à 224 objets audio discrets, qui peuvent ensuite être organisés à l'intérieur d'un flux comprenant jusqu'à 32 groupes d'objets et 32 présentations.

Un flux DTS-UHD est composé d'une séquence de trames DTS-UHD contenant trois principaux éléments:

- La table des matières de la trame (FTOC) – cet élément permet aux décodeurs d'aller directement vers les éléments d'intérêt dans la trame.
- Les éléments fragments de métadonnées.
- Les éléments fragments audio.

La structure est présentée ci-dessous:

FIGURE 18
Structure d'un flux DTS-UHD



BS.1196-18

La table FTOC est présente dans chaque trame DTS-UHD et indique si la trame est une trame de synchronisation ou non, les trames de synchronisation fournissant tous les paramètres nécessaires pour lancer une session de décodage.

Les fragments audio transportent les formes d'onde compressées qui sont compressées et transportées de manière optimale; ces fragments peuvent être des trames de synchronisation ou des trames autres que de synchronisation. Un fragment audio est organisé en un certain nombre de sous-flux qui peuvent représenter des formes d'ondes mono, des formes d'onde stéréo et des formes d'onde LFE.

Un fragment de métadonnées décrit tous les aspects d'une composante audio, y compris le type de composante, les fragments audio nécessaires pour restituer la composante et des informations sur le volume sonore et la compression de plage dynamique. Un fragment de métadonnées pour un point d'accès (une trame de synchronisation DTS-UHD) renvoie à un fragment audio qui est également une trame de synchronisation et contient un ensemble complet des métadonnées nécessaires pour commencer le décodage.

2 Codeur

Le codeur ACE DTS-UHD n'est pas spécifié de manière normative dans les normes ETSI. Le codeur met en œuvre un ensemble complet d'outils capables de fournir de multiples flux binaires pour prendre en charge un certain nombre de configurations de haut-parleurs différentes.

La technologie de codage DTS-UHD maintient la qualité audio et peut produire des flux binaires conformes à l'émission de haute qualité spécifiée dans la Recommandation UIT-R BS.1548 pour les débits suivants:

Configuration des haut-parleurs	Débit (kbit/s)
0+2+0 (stéréo)	128
0+5+0 (5.1)	192
4+7+0 (7.1.4)	288

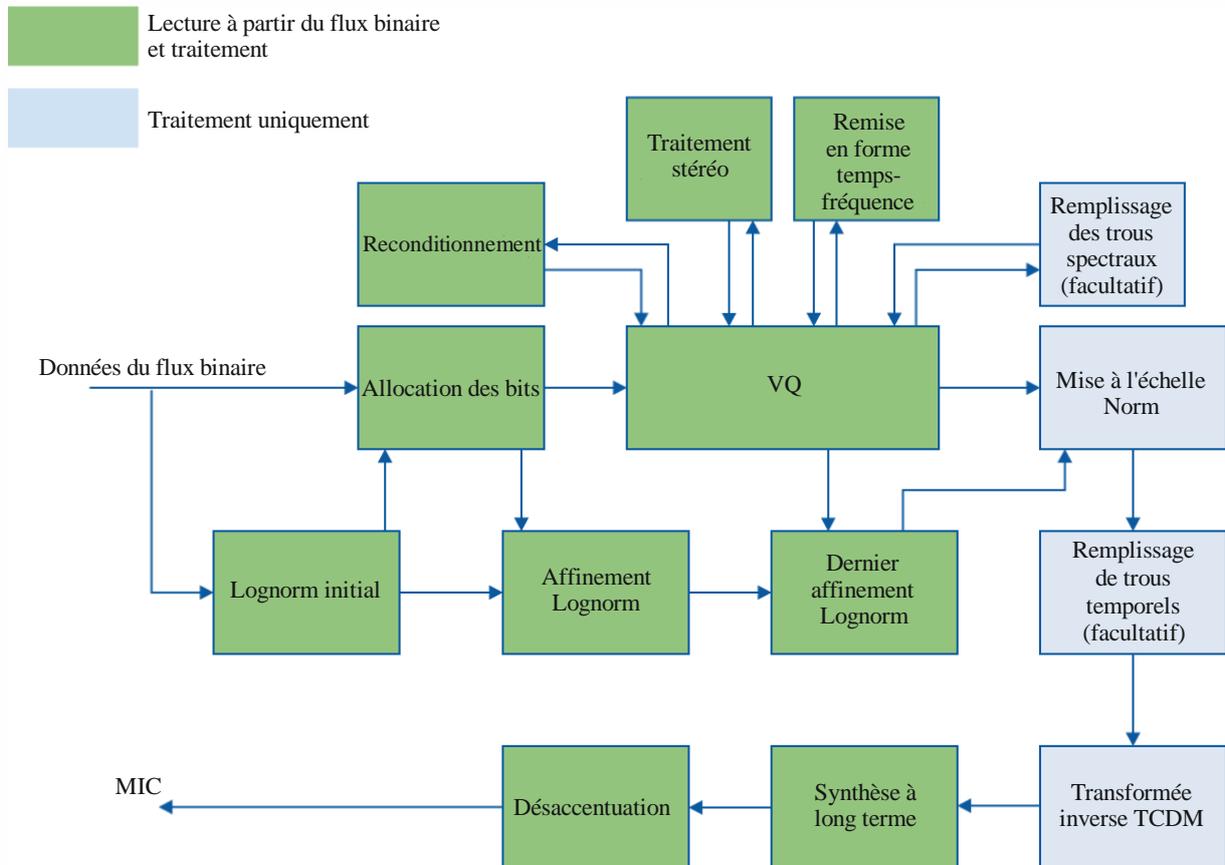
Le codeur DTS-UHD permet de maîtriser totalement le placement des trames de synchronisation afin que le service puisse contrôler les points d'entrée dans le service audio.

La fonctionnalité débit variable (VBR) du système DTS-UHD garantit un débit constant (CBR) sur chaque groupe GoF (groupe de trames – intervalle entre deux trames de synchronisation consécutives). Les trames de synchronisation sont définies par le codeur et l'intervalle de synchronisation peut être modifié de manière dynamique pendant le processus de codage selon les besoins. Le codeur DTS-UHD peut également définir le débit VBR de crête instantané maximal selon les besoins. Le modèle de contrôle du débit (BRC) du système DTS-UHD alloue de manière dynamique les bits à l'intérieur de chaque trame audio dans les objets, les canaux et les paires de canaux à l'intérieur d'un flux binaire. Le contrôle BRC consacre davantage de bits au codage du contenu le plus pertinent pour la perception dans chaque trame, tout en recherchant une qualité audio égale dans tous les objets et tous les canaux à l'intérieur d'une trame. Il est ainsi possible d'avoir une efficacité de codage améliorée avec un nombre accru d'éléments audio en faisant varier la répartition du contenu pertinent pour la perception dans les objets et les canaux. Le codeur alloue également de manière dynamique des bits aux trames audio avec pour objectif que la qualité audio perçue soit constante tout au long de l'axe du temps (compte tenu des contraintes concernant le débit CBR pour le groupe GoF et le débit de crête maximum spécifié). Il est à noter que la fonctionnalité BRC n'exige la transmission d'aucune information supplémentaire et que le décodeur n'est pas conscient de la fonctionnalité BRC. Chaque objet, canal ou paire de canaux continue de pouvoir être décodé de manière indépendante et «pleinement utilisable» en mode BRC.

3 Décodeur

Le décodeur ACE fonctionne sur des flux stéréo ACE, des flux mono ACE et des flux LFE ACE. Dans une trame ACE, les flux sont séparés selon ces trois classifications. Les flux stéréo comportent une optimisation supplémentaire qui utilise les informations communes aux deux canaux.

FIGURE 19
Décodeur ACE



BS.1196-19

Le décodage d'un flux ACE mono ou stéréo se déroule selon les étapes suivantes:

Lognorm initial: une première approximation de la puissance de la bande spectrale est extraite du flux binaire. Ces valeurs de puissance sont représentées dans le domaine logarithmique et appelées valeurs lognorm.

Allocation des bits: pour chaque bande spectrale, le nombre de bits disponibles pour le décodage normalisé de la bande spectrale est calculé. Ce calcul dépend des informations dans le flux binaire et (potentiellement) des valeurs lognorm initiales.

Affinement Lognorm: une petite partie des bits alloués est utilisée pour affiner les valeurs lognorm initiales.

VQ: (Quantification vectorielle), la majorité des bits alloués est utilisée pour reconstruire les valeurs normalisées des fréquences spectrales, une bande spectrale à la fois.

Reconditionnement: pour les bandes spectrales initialement «complètes» et des petites valeurs d'allocation des bits, les bandes non complètes reconstruites au départ sont reconditionnées afin de mieux s'approcher d'un spectre plein.

Traitement stéréo: pour les flux stéréo, les deux canaux reconstruits sont transformés en une représentation originale (non mixée).

Remise en forme temps-fréquence: les bandes spectrales sont prétraitées avec une simple transformée de Haar, ramenant toutes les bandes dans la même représentation temps-fréquence.

Remplissage des trous spectraux: les bandes avec l'énergie originale qui sont reconstruites avec des valeurs tout à zéro sont ensuite remplies de «bruit», généré à partir des valeurs des bandes reconstruites précédemment.

Dernier affinement Lognorm: tous les bits qui ont été laissés de côté après la quantification vectorielle sont utilisés pour faire un affinement final des valeurs lognorm.

Mise à l'échelle Norm: les bandes spectrales sont rétablies aux valeurs de puissance correspondant aux valeurs lognorm finales.

Remplissage des trous temporels: les bandes spectrales avec des valeurs tout à zéro sont potentiellement remplies avec du bruit de synthèse en fonction de la puissance de la bande spectrale des trames précédentes.

Transformée inverse TCDM: la représentation spectrale est transformée en représentation dans le domaine temporel (MIC).

Synthèse à long terme: un filtre de synthèse à long terme est utilisé pour reconstruire les composantes du signal qui ont été supprimées au moment du codage.
