

CT-aacPlus

Codage audio

Une technologie de pointe

Martin Dietz et Stefan Meltzer
Coding Technologies, Allemagne

Le système CC-aacPlus associe la technologie SBR, mise au point par Coding Technologies, à celle du codage audio avancé (AAC) MPEG qui constitue l'un des systèmes de codage audio perceptuel les plus puissants. Capable de distribuer des signaux sonores de haute qualité à des débits aussi bas que 24 kbit/s en mono et 48 kbit/s en stéréo, il sera utilisé par Digital Radio Mondiale (DRM) pour acheminer sous les 30 MHz une qualité sonore au moins équivalente à celle des services MF actuels.

Le codage audio avancé (*Advanced Audio Coding*, AAC) est l'un des plus puissants algorithmes actuels de codage perceptuel traditionnel. Associée à la technologie SBR (*Spectral Band Replication*), l'efficacité de codage du système AAC peut encore être augmentée d'au moins 30%, offrant ainsi la même qualité sonore à un débit réduit d'autant.

Le système de transmission Digital Radio Mondiale [1] va appliquer la combinaison AAC / SBR, appelée CT-aacPlus (*Coding Technologies advanced audio coding Plus*), dans les bandes de fréquences situées en dessous des 30 MHz, pour une qualité sonore quasiment équivalente à celle de la

modulation de fréquence à des débits de l'ordre de 20 kbit/s par canal audio.

La technologie SBR est le développement le plus récent du codage audio. Elle peut être combinée à presque tous les systèmes de codage classiques et permet un accroissement moyen de l'efficacité du débit binaire d'au moins 30%. Il est possible d'associer les systèmes de codage de manière compatible permettant ainsi de faire évoluer en douceur les systèmes actuels basés sur des codeurs audio traditionnels.

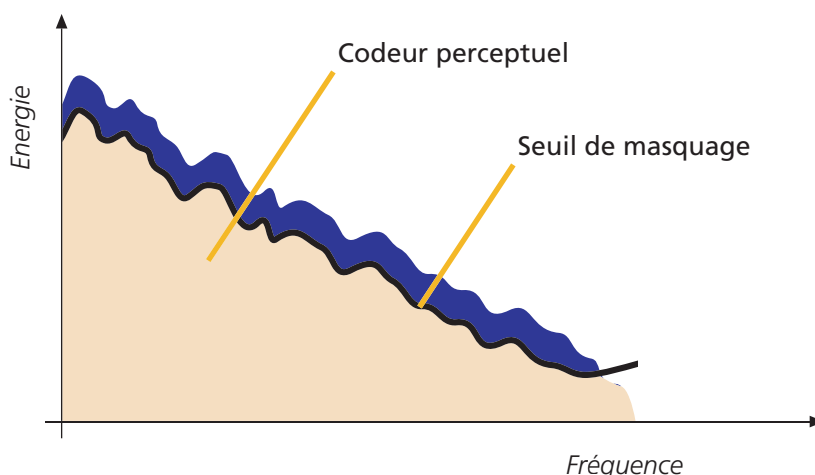
Les scénarios de transition appropriés permettront d'introduire la nouvelle technologie d'un façon presque insens-

ble pour ensuite profiter pleinement de tous les avantages de l'amélioration de l'efficacité du débit propre à la technologie SBR.

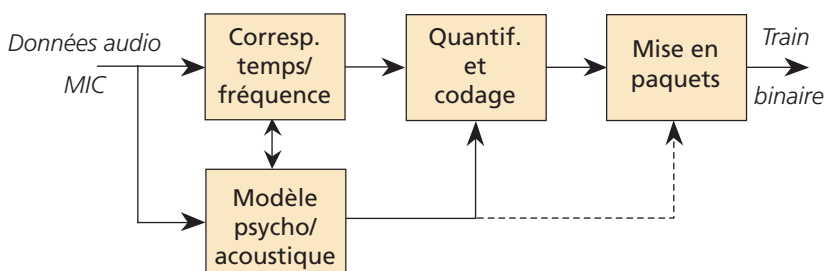
Codage audio perceptuel

Les recherches sur les codeurs audio perceptuels ont débuté il y a une vingtaine d'années. L'étude du système auditif humain a révélé que l'audition est essentiellement basée sur une analyse spectrale à court terme du signal sonore. On a constaté un effet appelé

masquage : le système auditif humain ne peut pas percevoir les déformations masquées par un signal plus puissant à une fréquence proche. De ce fait, en regardant le spectre à court terme, on peut calculer ce qu'on a baptisé un « seuil de masquage » pour ce spectre. Les déformations situées en dessous du seuil sont, dans l'idéal, inaudibles. On a alors étudié comment calculer le seuil de masquage (« modèle psycho-acoustique ») et comment traiter le signal de manière à ce qu'il ne contienne que les informations audibles. Théoriquement, un codec audio applique une compression telle que la déformation introduite intervient exactement au-dessous du seuil de masquage. La figure ci-dessus illustre le bruit de quantification produit par un codeur perceptuel idéal.



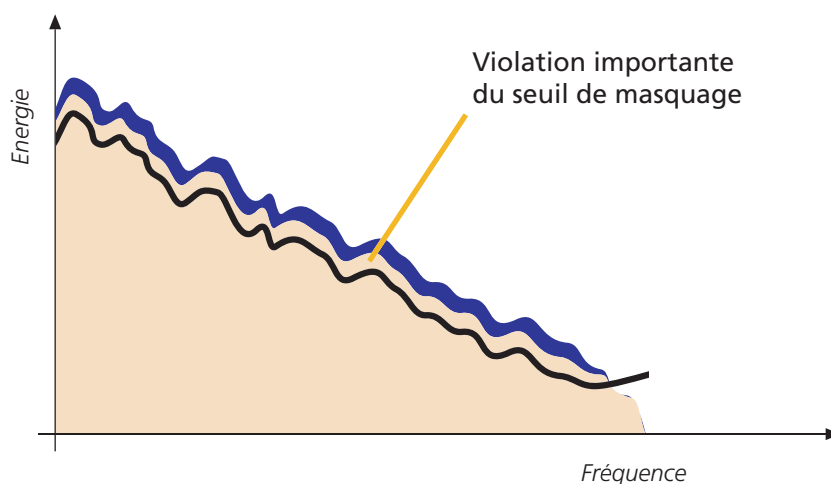
Codeur perceptuel idéal



Principe des codeurs perceptuels traditionnels

Ces recherches sont à l'origine des codecs audio perceptuels traditionnels actuels qui sont basés sur des codecs temporels tels que, le MPEG Layer 2, le Dolby AC-3, le MP3, le Sony Atrac, le Lucent PAC et le MPEG AAC. Tous ces codecs reposent sur le même principe, illustré ci-contre.

Le signal audio est transformé bloc par bloc dans le domaine fréquentiel à l'aide d'un banc de filtrage ou une transformée. Le spectre à courte durée ainsi obtenu est alors quantifié de manière à ne pas dépasser le seuil de masquage calculé à l'aide du modèle psycho-acoustique. Ce spectre quantifié est codé et converti en une salve binaire. Le décodeur exécute les étapes inverses pour traiter le signal, mais n'utilise généralement pas de modèle psycho-acoustique.



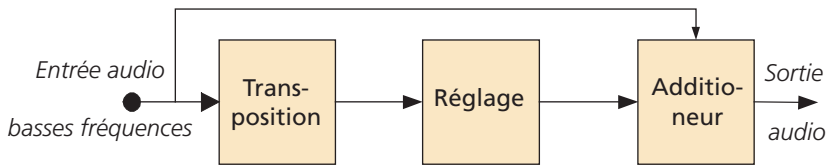
Codage hors limite de la forme d'onde

Bien que les codecs temporels perceptuels classiques parviennent déjà à un niveau de compression important, ils ne sont pas encore assez puissants pour satisfaire les besoins des systèmes utilisant des lignes téléphoniques analogiques ou numériques, ni des systèmes radio ou la radiodiffusion. La dernière figure ci-dessus illustre ce qui se passe lorsqu'on accroît encore le taux de compression dans un des codecs : la déformation qu'il introduit dépasse alors le seuil de masquage et produit des défauts perceptibles.

La principale méthode appliquée pour résoudre ce problème dans les codecs temporels perceptuels traditionnels

consiste à limiter la largeur de la bande audio. De ce fait, le reste du spectre dispose de davantage d'informations, ce qui se traduit par un signal propre, mais qui sonne creux. Il existe une autre méthode, appelée *stéréo d'intensité*, mais elle ne peut s'utiliser

que pour les signaux stéréo. Avec la stéréo d'intensité, on ne transmet qu'un canal et des informations panoramiques, au lieu d'un canal gauche et d'un canal droit. Cette méthode ne permet cependant qu'une augmentation limitée de l'efficacité de la com-



Régénération de la haute fréquence dans le décodeur

pression étant donné que souvent l'image stéréophonique est détruite.

Étape suivante : le SBR

SBR, la technologie de répétition de la bande spectrale mise au point par Coding Technologies, est une technologie nouvelle procurant un accroissement considérable de l'efficacité du codage. La SBR est le fruit des derniers progrès réalisés par la recherche dans le domaine du codage audio. Ils ont montré qu'il était possible de représenter les hautes fréquences d'un signal audio de manière beaucoup plus efficace. Le principal effet utilisé est la forte corrélation entre le contenu en basses et hautes fréquences d'un signal.

Dans un système de codage basé sur la SBR, le codage temporel ne code que les fréquences les plus basses du signal. Ce contenu en basses fréquences est utilisé pour recréer le contenu en hautes fréquences dans le décodeur (ci-dessus).

Cette opération est réalisée avec des méthodes de transposition très pointues. Le contenu en hautes fréquences recréé subit quelques réglages en fré-

quence et dans le domaine temporel avant d'être combiné à la partie basses fréquences du signal.

Le dispositif illustré ci-dessous intègre le codec AAC et la technologie SBR. Côté codeur, le codage SBR est réalisé avant le codage AAC. Le signal à l'entrée du codeur AAC subit une limitation en largeur de bande puisque les hautes fréquences seront recréées par l'algorithme SBR. Quelques informations supplémentaires, nécessaires au décodeur SBR pour reconstruire la partie hautes fréquences, sont multiplexées dans le train binaire codé à l'intérieur du multiplexeur. Côté décodeur, la sortie basses fréquences du décodeur AAC est injectée dans l'unité SBR qui recrée les hautes fréquences pour produire un signal audio totalement reconstitué.

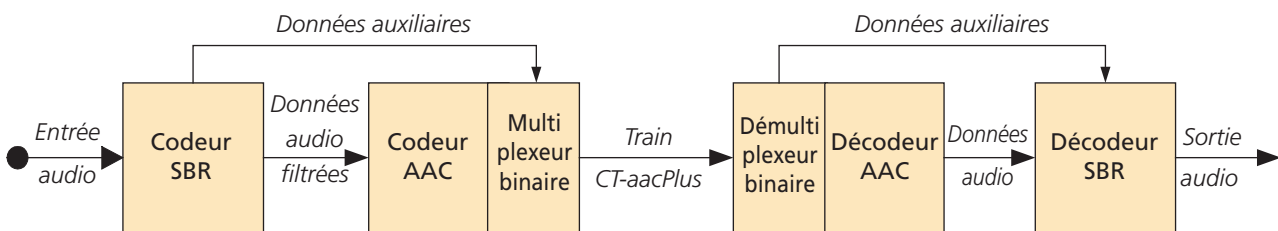
Le système CT-aacPlus offre une efficacité de compression bien supérieure à celle du système AAC isolé. En effet, la partie AAC du CT-aacPlus ne sert qu'à coder le contenu basses fréquences du signal. Les informations SBR n'ayant besoin que de très peu de données, le codec AAC dispose de beaucoup plus d'informations pour coder le contenu en basses fréquences (figure page suivante). De cette manière, le seuil de masquage n'est pas dépassé.

L'unité SBR recrée ensuite les hautes fréquences et associe les deux parties pour obtenir le signal total. Le train binaire qui en résulte comprend deux composantes, la partie AAC et la partie SBR. Le format binaire peut être conçu pour qu'un décodeur AAC dédié traite la partie AAC. Bien que le signal audio obtenu ait une largeur de bande limitée, cette possibilité permet d'introduire la SBR dans les systèmes actuels équipés de codecs audio perceptuels traditionnels.

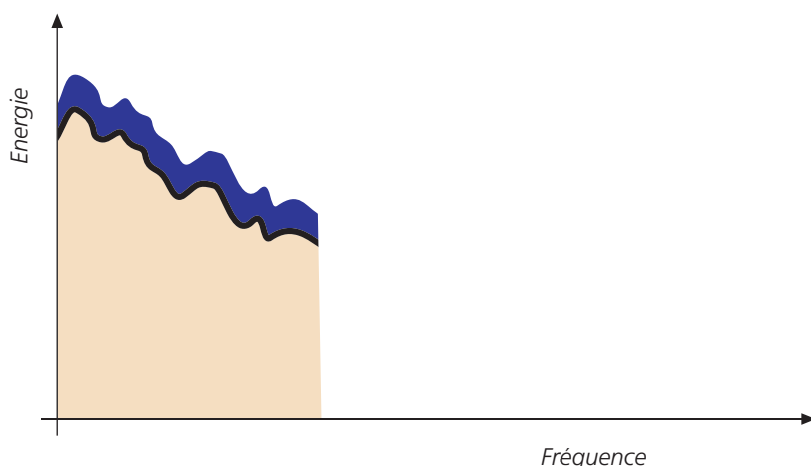
La technologie SBR peut s'utiliser avec tous les codecs audio produisant un signal propre en fréquences basses. Le format **MP3Pro**, combinaison du MPEG Layer-3 (MP3) et de la SBR, est un autre exemple d'application de cette technologie (voir l'encadré page suivante).

Codecs SBR améliorés et codecs traditionnels

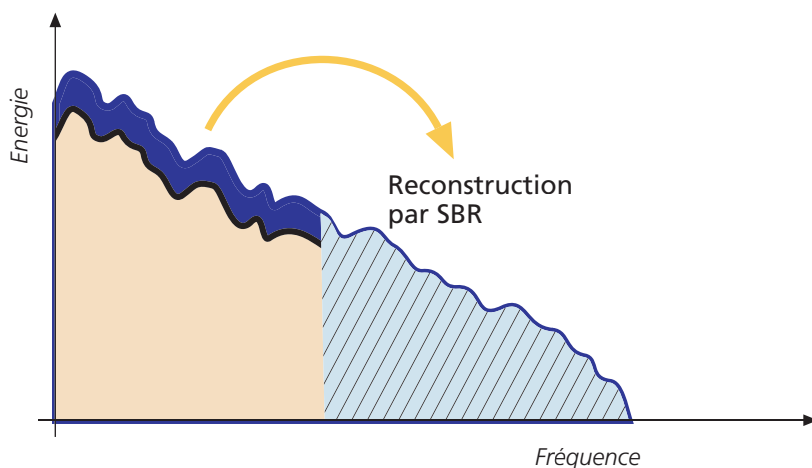
Pendant la sélection de l'algorithme de codage de source, le consortium DRM a effectué plusieurs tests d'écoute. L'un d'eux, réalisé à la BBC, qui comparait les systèmes AAC et AAC + SBR en 24 kbit/s mono, est représenté plus loin. Ces essais ont suivi la méthode MUSHRA [2]. L'essai utilise plusieurs signaux « d'ancrage » ou de référence tels que les versions à largeur de bande limitée (3,4 kHz et 7 kHz) du signal original, ainsi que le signal original lui-même. Ces signaux de référence accroissent la fiabilité des résultats. Sur l'échelle MUSHRA, le signal original est normalisé à 100.



Codage CT-aacPlus



Partie basses fréquences codée en AAC



La reconstruction des hautes fréquences par SBR

- Les essais effectués à la BBC ont porté sur les signaux suivants :
- signal AAC pur, c'est-à-dire MPEG-4 AAC à 24 kbit/s ;
 - signal de référence à 3,5 kHz et 7 kHz ;
 - signal AAC SBR, c'est-à-dire CT-aacPlus à 24 kbit/s ;

- AAC SBR principal – la partie AAC du signal CT-aacPlus – à 24 kbit/s ;
- AAC large bande, c'est-à-dire MPEG-4 AAC à 32 kbit/s ;
- référence cachée – le signal original, sans limitation de largeur de bande.

Les résultats font clairement apparaître la supériorité du CT-aacPlus par rapport à l'AAC du point de vue de l'efficacité de codage. Le CT-aacPlus à 24 kbit/s améliore l'efficacité de codage de plus de 30% par rapport à l'AAC à 32 kbit/s.

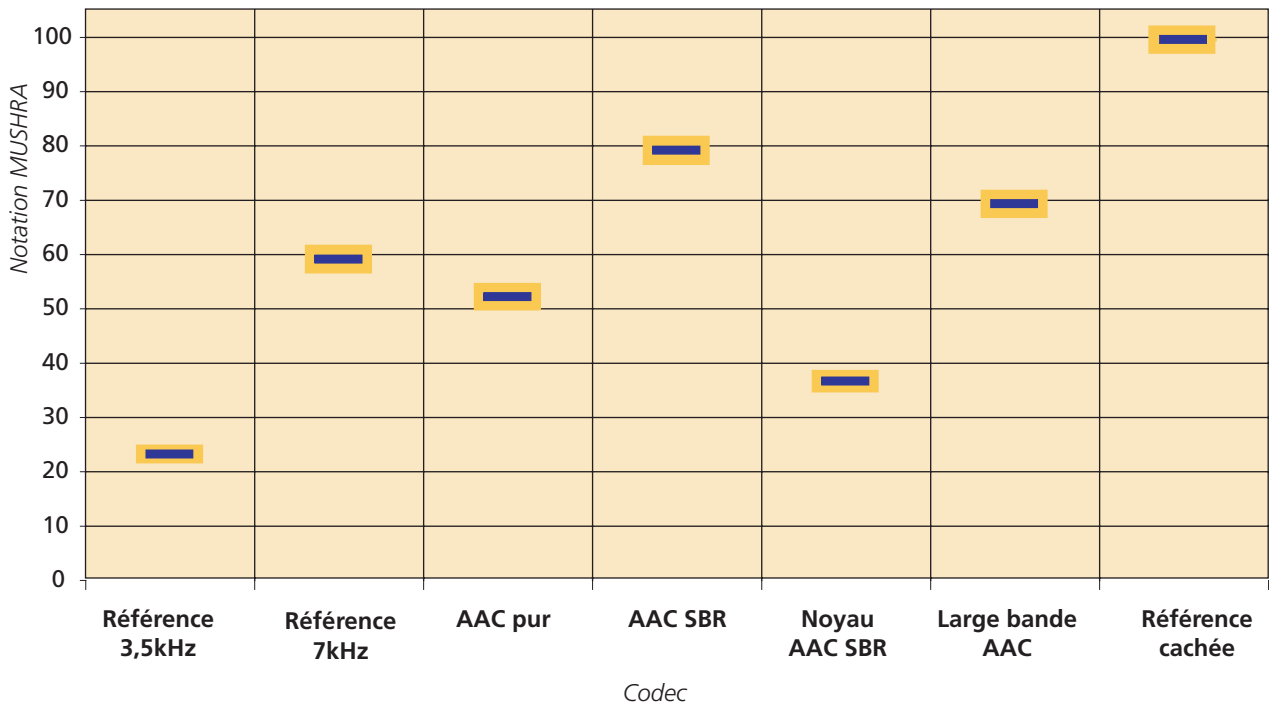
Un autre essai MUSHRA, effectué à l'aide de plusieurs codecs disponibles dans le commerce, a été réalisé par l'UER. Les premiers résultats ont été annoncés lors d'une récente conférence de l'AES au Royaume-Uni [3]. Il s'agissait dans cet essai de comparer le MP3Pro et le CT-aacPlus au MP3, à l'AAC, au WMA 8.0 et au Real Audio en stéréo à 48 kbit/s. Les résultats sont représentés à la page suivante.

On peut interpréter l'essai de l'UER de la manière suivante : la référence cachée (Ref.) obtient une note en dessous de 100, ce qui signifie que certains auditeurs ne pourront pas la différencier d'une très bonne version codée (très vraisemblablement produite par le codec CT-aacPlus) pour certains éléments. C'est le codec CT-aacPlus qui obtient les meilleurs résultats, avec une note moyenne d'environ 88. Ensuite vient le MP3Pro, suivi du MPEG-4 AAC.

MP3Pro et CT-aacPlus

La première technologie mise au point et commercialisée par Coding Technologies fut la technologie appelée *Spectral Band Replication* (SBR). La combinaison du MP3 et de la SBR a été baptisée MP3Pro et commercialisée par Coding technologies et son partenaire de licence, Thomson Multimedia, en juin 2001. Thomson/RCA, Philips, MusicMatch, STMicroelectronics, Texas Instruments, Gracenote, Syntrillium, Ahead et bien d'autres fabricants ont déjà lancé ou annoncé des produits utilisant MP3Pro.

La technologie CT-aacPlus, combinaison des technologies AAC et SBR, est le codec audio le plus puissant disponible aujourd'hui. Il a été choisi pour des systèmes tels que XM Satellite Radio aux Etats-Unis et Digital Radio Mondiale (DRM) pour les transmissions en ondes courtes, moyennes et longues.



Essais de la BBC : comparaison entre AAC et CT-aacPlus, signal monophonique à 24 kbit/s

Le RealAudio (version 8.5), Windows Media (WMA) 8.5 et le MP3 se situent en dessous. Ce résultat constitue une confirmation de la qualité exceptionnelle des codecs renforcés par la technologie SBR, le MP3Pro et le CT-aacPlus.

Il convient aussi de noter que le WMA 8.5 ne fonctionne pas beaucoup mieux que le MP3, contrairement au battage publicitaire en sa faveur !

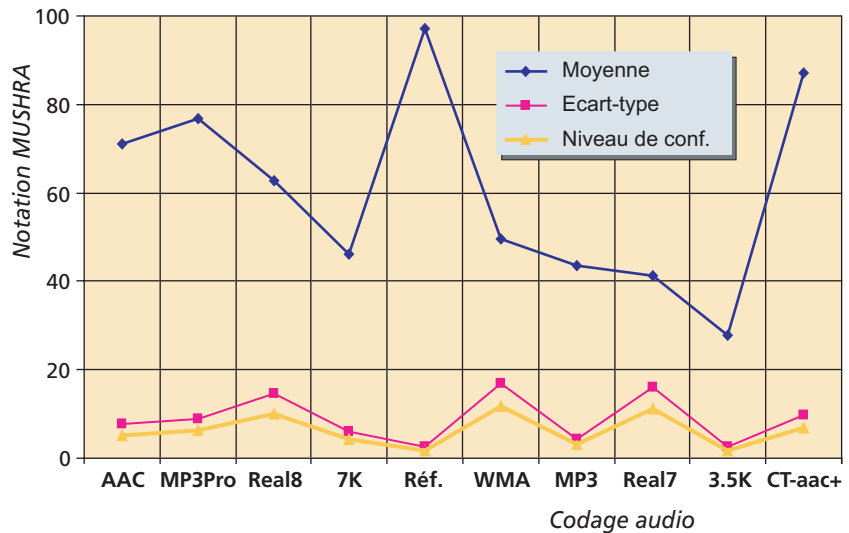
On obtient ainsi un débit binaire disponible pour le codage source compris entre 18 et 30 kbit/s, à peu près équivalent aux débits des modems analogiques. Il est donc essentiel de choisir la méthode de codage sonore la plus efficace pour parvenir à un son de qualité dans ces conditions.

Les excellentes performances du système CT-aacPlus permettent au système

DRM d'offrir une très haute qualité dans les bandes en MA. Dans un tel canal, on obtient un son mono de qualité et, si on dispose d'un débit suffisant, on peut même transmettre un certain niveau de stéréophonie. En regroupant deux canaux, ce qui sera peut-être possible à l'avenir si le cadre réglementaire le permet, les débits pourraient atteindre 40 - 48 kbit/s, permettant des programmes stéréo de haute qualité.

Le CT-aacPlus dans le système DRM

L'une des exigences majeures du développement du système DRM [1] était de conserver l'espace actuel des canaux MA. La largeur de spectre disponible pour chaque canal de transmission est de 9 ou 10 kHz (selon la région et la bande). Bien que le système applique des méthodes de modulation et de codage de voie très sophistiquées [4], les caractéristiques du canal n'autorisent que des indices d'occupation spectrale compris entre 2 et 3 bit/Hz (en dessous dans les cas très difficiles).



Essais de l'UER, signal stéréophonique à 48 kbit/s (résultats de l'IRT)



Martin Dietz, ingénieur en électricité de l'université Friedrich-Alexander de Erlangen (Allemagne), est depuis l'été 2000 président-directeur général de Coding Technologies, société germano-suédoise spécialisée dans les technologies de traitement amélioré du son et des signaux vocaux et plus particulièrement de codage audio. Après avoir obtenu son diplôme en 1992, il entre au Fraunhofer Institute for Integrated Circuits (FhG/IIS-A) à Erlangen comme chercheur dans le domaine du codage audio. De 1992 à 1995, il travaille à la normalisation de la couche 3 du MPEG (MP3) et à la mise en œuvre du premier décodeur intégré, utilisée aujourd'hui dans nombres de lecteurs MP3.

De 1995 à 1997, il dirige le groupe de recherche de IIS chargé du développement du système ISO/MPEG AAC. En 1998, il est nommé à la tête du département audio, chargé de la R&D sur les algorithmes audio MP3, AAC et MPEG-4. Il a été honoré par l'Audio Engineering Society en décembre 2001 pour ses travaux dans le domaine du codage audio.

Stefan Meltzer, ingénieur en électricité de l'université Friedrich-Alexander, est vice-président chargé du développement commercial de Coding Technologies à Nuremberg. Diplômé en 1990, il entre au Fraunhofer Institute for Integrated Circuits et travaille tout d'abord dans le domaine de la conception des circuits intégrés. En 1995, il est nommé responsable de projet pour le développement du système WorldSpace Satellite Broadcasting. De 1998 à 2000, il dirige l'équipe de développement du système de radiodiffusion XM Satellite Radio.



Pour des cas particuliers, par exemple des transmissions particulièrement robustes, des diffusions simultanées analogiques/numériques ou des transmissions multilingues, on peut utiliser le codage AAC ou des signaux vocaux au débit le plus bas. Pour les émissions vocales à débit ultra-bas, souvent associées à un programme audio normal, c'est le codec vocal HVXC à 2 kbit/s qui convient. On prévoit de combiner les codeurs de signaux vocaux et la technologie SBR.

Le système DRM utilise le codage MPEG-4 AAC, supportant bien les erreurs, pour améliorer la robustesse en cas d'erreurs de transmission. Un format supertrame a été mis au point pour prendre en charge les propriétés de la partie codage de voie et modulation et fournir une trame commune aux différents codecs utilisés dans le système DRM.

Le codage audio CT-aacPlus est également utilisé par XM Satellite Radio [5]. XM Satellite Radio est l'un des deux fournisseurs américains de radio par satellite sur abonnement. A l'aide de deux satellites géostationnaires et d'un réseau d'émetteurs terrestres, il fournit, depuis novembre 2001, cent canaux audio différents sur l'ensemble du territoire des Etats-Unis. Son principal marché est constitué par les autoradios mais il devrait s'étendre à la réception domestique. Pour être fiable, le système recourt à la diversité et à la redondance. Etant donné la quantité limitée de spectre disponible pour ce service, il était impératif

d'utiliser le codage le plus efficace possible pour atteindre l'objectif de cent canaux. Les codecs traditionnels n'étant pas en mesure de le réaliser avec une qualité audio satisfaisante, XM Satellite Radio a choisi le système CT-aacPlus.

La nouvelle technologie d'extension de la largeur de bande SBR permet d'améliorer considérablement les technologies modernes de codage audio telles que le MP3 et l'AAC. Les études préliminaires ont montré qu'elle peut accroître de 30% l'efficacité du codage.

Le système CT-aacPlus, qui allie les technologies AAC et SBR, est incontestablement le codec audio le plus puissant aujourd'hui. Utilisé dans les systèmes DRM et XM Satellite Radio, il peut aussi s'étendre à d'autres codecs employés en radiodiffusion numérique, dans les communications mobiles et les applications d'émission en continu sur Internet.

Bibliographie

- [1] Site de DRM : <http://www.drm.org/plan-site.htm>
- [2] G. Stoll et F. Kozamernik: **EBU listening tests on Internet audio codecs**
EBU Technical Review n° 283, juin 2000

http://www.ebu.ch/trev_283-kozamernik.pdf

- [3] Martin Link, IRT GmbH: **Internet Audio Quality and the MUSHRA Test**
Conférence de l'AES UK : Audio Delivery: the changing home experience
<http://www.aes.org/sections/uk/conference/02program.html>
- [4] Jonathan Stott: **Principales caractéristiques techniques du DRM UER** – Revue technique, sélection 2001.
http://www.ebu.ch/trev_286-stott_fr.pdf
- [5] XM Satellite Radio : <http://www.xmradio.com/>
- [6] DRM Source Coding Group, 2001 : **Report on Subjective Listening Tests of SBR-LC, an AAC-based Audio Bandwidth Widening Tool**
<http://www.drm.org/plan-site.htm>

Les auteurs remercient leurs collègues suédois et allemands de Coding Technologies. Ils rendent également hommage au travail ardu et de qualité de leurs collègues de DRM.

©UER Publié en anglais : juillet 2002