



(11)

EP 4 135 350 A1

(12)

DEMANDE DE BREVET EUROPEEN

(43) Date de publication:
15.02.2023 Bulletin 2023/07

(51) Classification Internationale des Brevets (IPC):
H04S 7/00 ^(1980.01) **G10L 19/008** ^(2013.01)

(21) Numéro de dépôt: **22197901.6**

(52) Classification Coopérative des Brevets (CPC):
H04S 7/304; G10L 19/008; H04S 2400/11;
H04S 2420/01

(22) Date de dépôt: **07.12.2018**

(84) Etats contractants désignés:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO RS SE SI SK SM TR

(71) Demandeur: **Orange**
92130 Issy-les-Moulineaux (FR)

(72) Inventeur: **PALLONE, Grégory**
92326 Châtillon Cedex (FR)

(30) Priorité: **19.12.2017 FR 1762478**

(62) Numéro(s) de document de la (des) demande(s)
initiale(s) en application de l'article 76 CBE:
18833274.6 / 3 729 832

Remarques:

Cette demande a été déposée le 27.09.2022 comme
demande divisionnaire de la demande mentionnée
sous le code INID 62.

(54) **TRAITEMENT D'UN SIGNAL MONOPHONIQUE DANS UN DÉCODEUR AUDIO 3D RESTITUANT UN CONTENU BINAURAL**

(57) L'invention se rapporte à un procédé de traitement d'un signal monophonique audio dans un décodeur audio 3D comportant une étape de traitement de binauralisation des signaux décodés destinés à être restitué spatialement par un casque audio. Le procédé est tel que, à la détection (E200), dans un flux de données représentatif du signal monophonique, d'une indication de non-traitement de binauralisation associée à une information de position spatiale de restitution, le signal monophonique décodé est dirigé (O-E200) directement vers un module de mixage comportant un moteur de rendu stéréophonique prenant en compte l'information de position pour construire deux voies de restitution (E220) traitées directement par une étape de mixage direct (E230) sommant ces deux voies avec un signal binauralisé issu du traitement de binauralisation, pour être restitué (E240) sur le casque audio.

L'invention se rapporte également à un dispositif et décodeur mettant en oeuvre le procédé de traitement.

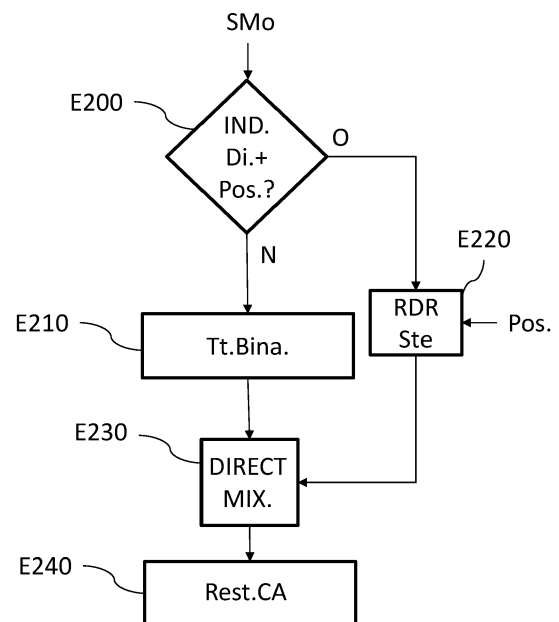


FIG. 2

Description

[0001] La présente invention se rapporte au traitement d'un signal audio dans un système de décodage audio 3D de type codec normalisé MPEG-H 3D audio. L'invention se rapporte plus particulièrement au traitement d'un signal monophonique destiné à être restitué sur un casque recevant par ailleurs des signaux audio binauraux.

[0002] Le terme binaural vise une restitution sur casque audio ou paire d'écouteurs, d'un signal sonore avec néanmoins des effets de spatialisation. Un traitement binaural de signaux audio, appelé par la suite binauralisation ou traitement de binauralisation, utilise des filtres HRTF (pour « Head Related Transfert Function » en anglais) dans le domaine fréquentiel ou HRIR, BRIR (Pour « Head Related Transfert Function », « Binaural Room Impulse Response » en anglais) dans le domaine temporel qui reproduisent les fonctions de transfert acoustiques entre les sources sonores et les oreilles de l'auditeur. Ces filtres servent à simuler des indices de localisation auditive qui permettent à un auditeur de localiser les sources sonores comme en situation d'écoute réelle.

[0003] Le signal de l'oreille droite est obtenu en filtrant un signal monophonique par la fonction de transfert (HRTF) de l'oreille droite et le signal de l'oreille gauche est obtenu en filtrant ce même signal monophonique par la fonction de transfert de l'oreille gauche.

[0004] Dans les codecs de type NGA (pour « Next Generation Audio » en anglais), tels que MPEG-H 3D audio décrit dans le document référencé ISO/IEC 23008-3 : « High efficiency coding and media delivery in heterogeneous environments - Part 3 : 3D audio » publié le 25/07/2014 ou encore AC4 décrit dans le document référencé ETSI TS 103 190 : « Digital Audio Compression Standard » publié en Avril 2014, les signaux reçus au décodeur sont dans un premier temps décodés puis subissent un traitement de binauralisation tel que décrit ci-dessus avant d'être restitués sur un casque audio. On s'intéresse ici au cas de la restitution sur casque audio, avec son spatialisé, c'est-à-dire à un signal binauralisé.

[0005] Les codecs cités prévoient donc la possibilité d'une restitution sur plusieurs des haut-parleurs virtuels grâce à l'écoute d'un signal binauralisé sur casque mais prévoient également la possibilité d'une restitution sur plusieurs haut-parleurs réels, d'un son spatialisé.

[0006] Dans certains cas, est associée avec le traitement de binauralisation, une fonction de traitement de suivi de la tête de l'auditeur (« Head tracking » en anglais) que l'on nommera rendu dynamique, par opposition au rendu statique. Ce traitement permet de prendre en compte le mouvement de la tête de l'auditeur pour modifier la restitution sonore sur chaque oreille afin de garder la restitution de la scène sonore stable. En d'autres termes, l'auditeur percevra les sources sonores au même endroit dans l'espace physique s'il bouge ou s'il ne bouge pas la tête.

[0007] Ceci peut être important pour la visualisation et l'écoute associée d'un contenu vidéo 360°.

[0008] Cependant, pour certains contenus, il n'est pas souhaitable qu'ils soient traités par ce type de traitement. En effet, dans certains cas, lorsque le contenu a été créé spécifiquement pour un rendu binaural, par exemple si les signaux ont été enregistrés directement par une tête artificielle ou déjà traités par un traitement de binauralisation, alors ils doivent être restitués directement sur les écouteurs du casque. Ces signaux ne nécessitent pas de traitement de binauralisation supplémentaire.

[0009] De même, un producteur de contenu peut souhaiter qu'un signal sonore soit restitué de façon indépendante de la scène sonore, c'est-à-dire qu'il soit perçu comme un son à part de la scène sonore, par exemple comme dans le cas d'une voix « OFF ».

[0010] Ce type de restitution peut permettre par exemple de donner des explications sur une scène sonore par ailleurs restituée. Par exemple, le producteur de contenu peut souhaiter que le son soit restitué sur une seule oreille pour pouvoir obtenir un effet volontaire de type « oreillette », c'est-à-dire que le son n'est entendu que d'une seule oreille. On peut souhaiter également que ce son reste en permanence uniquement sur cette oreille même si l'auditeur bouge sa tête, ce qui est le cas dans l'exemple précédent. Le producteur de contenu peut également souhaiter que ce son soit restitué à une position précise dans l'espace sonore, par rapport à une oreille de l'auditeur (et pas uniquement à l'intérieur d'une seule oreille) et ce, même s'il bouge la tête.

[0011] Un tel signal monophonique décodé et mis en entrée d'un système de restitution d'un codec de type MPEG-H 3D audio ou AC4, sera binauralisé. Le son sera alors réparti sur les deux oreilles (même s'il sera moins fort dans l'oreille contra-latérale) et si l'auditeur bouge sa tête, il ne percevra pas le son de la même façon sur son oreille, puisque le traitement de suivi de la tête, s'il est mis en oeuvre, fera en sorte que la position de la source sonore reste la même que dans la scène sonore initiale : selon la position de la tête, le son apparaîtra donc plus fort dans l'une ou l'autre des oreilles.

[0012] Dans une proposition de modification du codec MPEG-H 3D audio, une contribution référencée « ISO/IEC JTC1/SC29/WG11 MPEG2015/M37265 » d'Octobre 2015 propose d'identifier les contenus qui ne doivent pas être altérés par la binauralisation.

[0013] Ainsi, une identification « Dichotic » est associée aux contenus ne devant pas être traités par binauralisation.

[0014] Tous les éléments audio seront alors binauralisés sauf ceux référencés « Dichotic ». « Dichotic » signifie que l'on a un signal différent sur chacune des oreilles.

[0015] De la même façon, dans la norme AC4, un bit d'information indique qu'un signal est déjà virtualisé. Ce bit permet la désactivation du post-traitement. Les contenus ainsi identifiés sont des contenus déjà formatés pour le casque audio, c'est à dire en binaural. Ils comportent deux canaux.

[0016] Ces méthodes ne traitent pas du cas d'un signal

monophonique pour lequel, le producteur de la scène sonore ne désire pas de binauralisation.

[0017] Ceci ne permet pas de restituer un signal monophonique de façon indépendante de la scène sonore, à une position précise par rapport à une oreille d'un auditeur qu'on appellera en mode « oreillette ». En utilisant les techniques de l'état de l'art à deux canaux, une solution serait de créer un contenu à 2 canaux constitué d'un signal dans une des voies et d'un silence dans l'autre voie pour une restitution souhaitée sur une seule oreille ou bien de créer un contenu stéréophonique prenant en compte la position spatiale souhaitée et d'identifier ce contenu comme ayant déjà été spatialisé avant de le transmettre.

[0018] Cependant ce type de traitement crée de la complexité par la création de ce contenu stéréophonique et demande un débit supplémentaire de transmission de ce contenu stéréophonique.

[0019] Il existe donc un besoin d'offrir une solution qui permette de faire transiter un signal qui sera restitué à une position précise par rapport à une oreille d'un porteur de casque audio de façon indépendante d'une scène sonore restituée par ce même casque, tout en optimisant le débit du codec utilisé.

[0020] La présente invention vient améliorer la situation.

[0021] Elle propose à cet effet, un procédé de traitement d'un signal monophonique audio dans un décodeur audio 3D comportant une étape de traitement de binauralisation des signaux décodés destinés à être restitué spatialement par un casque audio. Le procédé est tel que, à la détection, dans un flux de données représentatif du signal monophonique, d'une indication de non-traitement de binauralisation associée à une information de position spatiale de restitution, le signal monophonique décodé est dirigé directement vers un module de mixage comportant un moteur de rendu stéréophonique prenant en compte l'information de position pour construire deux voies de restitution traitées par une étape de mixage direct sommant ces deux voies avec un signal binauralisé issu du traitement de binauralisation, pour être restitué sur le casque audio.

[0022] Ainsi, il est possible de spécifier qu'un contenu monophonique doit être restitué à une position spatiale précise par rapport à une oreille d'un auditeur et qu'il ne subisse pas de traitement de binauralisation de façon à ce que ce signal restitué puisse avoir un effet « oreillette », c'est-à-dire qu'il soit entendu par l'auditeur à une position déterminée par rapport à une oreille, à l'intérieur de la tête de la même façon qu'un signal stéréophonique et ceci même si la tête de l'auditeur bouge.

[0023] En effet, les signaux stéréophoniques sont caractérisés par le fait que chaque source sonore se trouve présente dans chacune des 2 voies de sortie (gauche et droite) avec une différence d'intensité (ou ILD pour « Interaural Level Différence ») et parfois de temps (ou ITD pour « Interaural Time Différence ») entre les voies. Lors d'une écoute au casque d'un signal stéréophonique,

les sources sont perçues à l'intérieur de la tête, à un endroit se situant entre l'oreille gauche et l'oreille droite, dépendant de l'ILD et/ou de l'ITD. Les signaux binauraux s'opposent aux signaux stéréophoniques en ce que les sources se voient appliquer un filtre reproduisant le trajet acoustique de la source à l'oreille de l'auditeur. Lors d'une écoute au casque d'un signal binaural, les sources sont perçues en dehors de la tête, à un endroit se situant sur une sphère, dépendant du filtre utilisé.

[0024] Les signaux stéréophoniques et binauraux se rapprochent en ce qu'ils sont constitués de 2 voies gauche et droite, et se distinguent par le contenu de ces 2 voies.

[0025] Ce signal mono (pour monophonique) restitué vient alors en superposition aux autres signaux restitués qui forment une scène sonore 3D.

[0026] Le débit nécessaire pour indiquer ce type de contenu est optimisé puisqu'il ne suffit de coder qu'une indication de position dans la scène sonore en plus de l'indication de non-binauralisation pour informer le décodeur du traitement à effectuer, contrairement à une méthode qui nécessiterait d'encoder, transmettre puis décoder un signal stéréophonique prenant en compte cette position spatiale.

[0027] Les différents modes particuliers de réalisation mentionnés ci-après peuvent être ajoutés indépendamment ou en combinaison les uns avec les autres, aux étapes du procédé de traitement défini ci-dessus.

[0028] Dans un mode de réalisation particulier, l'information de position spatiale de restitution est une donnée binaire indiquant une seule voie du casque audio de restitution.

[0029] Cette information ne nécessite qu'un bit de codage, ce qui permet encore de restreindre le débit nécessaire.

[0030] Dans ce mode de réalisation, seule la voie de restitution correspondant à la voie indiquée par la donnée binaire est sommée à la voie correspondante du signal binauralisé à l'étape de mixage direct, l'autre voie de restitution étant de valeur nulle.

[0031] La sommation ainsi effectuée est simple à mettre en oeuvre et apporte l'effet « oreillette » désiré, de superposition du signal mono à la scène sonore restituée.

[0032] Dans un mode de réalisation particulier, le signal monophonique est un signal de type canal dirigé vers le module de mixage avec l'information de position spatiale de restitution.

[0033] Ainsi, le signal monophonique ne subit pas d'étape de traitement de binauralisation et n'est pas traité comme les signaux de type canal habituellement traités par les méthodes de l'état de l'art. Ce signal est traité par un module de mixage comportant un moteur de rendu stéréophonique différent de celui existant pour les signaux de type canal. Ce moteur de rendu consiste à dupliquer le signal monophonique sur les 2 voies, en appliquant des facteurs fonctions de l'information de position spatiale de restitution, sur les deux canaux.

[0034] Ce moteur de rendu stéréophonique peut par ailleurs être intégré au moteur de rendu canal avec un traitement différencié selon la détection faite pour le signal à l'entrée de ce moteur de rendu ou au module de mixage direct sommant les voies issues de ce moteur de rendu stéréophonique au signal binauralisé issu du module de traitement de binauralisation.

[0035] Dans un mode de réalisation lié au signal de type canal, l'information de position spatiale de restitution est une donnée de différence interaurale de niveau sonore de type ILD ou plus généralement une information de rapport de niveau entre les voies gauche et droite.

[0036] Dans un autre mode de réalisation, le signal monophonique est un signal de type objet associé à un ensemble de paramètres de restitution comprenant l'indication de non-binauralisation et l'information de position de restitution, le signal étant dirigé vers le module de mixage avec l'information de position spatiale de restitution.

[0037] Dans cet autre mode de réalisation, l'information de position spatiale de restitution est par exemple une donnée d'angle azimut.

[0038] Cette information permet de donner une position de restitution par rapport à une oreille du porteur du casque audio de façon à ce que ce son soit restitué en superposition d'une scène sonore.

[0039] Ainsi, le signal monophonique ne subit pas d'étape de traitement de binauralisation et n'est pas traité comme les signaux de type objet habituellement traités par les méthodes de l'état de l'art. Ce signal est traité par un module de mixage comportant un moteur de rendu stéréophonique différent de celui existant pour les signaux de type objet. L'indication de non-traitement de binauralisation ainsi que l'information de position de restitution sont comprises dans les paramètres de restitution (Metadata) associés au signal de type objet. Ce moteur de rendu peut par ailleurs être intégré au moteur de rendu objet ou au module de mixage direct sommant les voies issues de ce moteur de rendu stéréophonique au signal binauralisé issu du module de traitement de binauralisation.

[0040] La présente invention se rapporte aussi à un dispositif de traitement d'un signal monophonique audio comportant un module de traitement de binauralisation de signaux décodés destinés à être restitués spatialement par un casque audio. Ce dispositif est tel qu'il comporte :

- un module de détection apte à détecter, dans un flux de données représentatif du signal monophonique, une indication de non-traitement de binauralisation associée à une information de position spatiale de restitution ;
- un module de redirection, dans le cas d'une détection positive par le module de détection, apte à diriger le signal monophonique vers un module de mixage ;
- un module de mixage comportant un moteur de rendu stéréophonique apte à prendre en compte l'infor-

mation de position pour construire deux voies de restitution, le module de mixage étant apte à traiter directement les deux voies de restitution en les sommant avec un signal binauralisé issu du module de traitement de binauralisation, pour être restitué sur le casque audio.

[0041] Ce dispositif présente les mêmes avantages que le procédé décrit précédemment, qu'il met en oeuvre.

[0042] Le moteur de rendu stéréophonique étant intégré dans le module de mixage direct, ce n'est qu'au module de mixage direct que les voies de restitution sont construites, seule l'information de position étant alors transmise avec le signal mono jusqu'au module de mixage direct. Ce signal peut être de type canal ou de type objet.

[0043] Dans un mode de réalisation, le signal monophonique est un signal de type canal et le moteur de rendu stéréophonique est intégré à un moteur de rendu canal construisant par ailleurs des voies de restitution pour des signaux à plusieurs canaux.

[0044] Dans un autre mode de réalisation, le signal monophonique est un signal de type objet et le moteur de rendu stéréophonique est intégré à un moteur de rendu objet construisant par ailleurs des voies de restitution pour des signaux monophoniques associées à des ensembles de paramètres de restitution.

[0045] La présente invention vise un décodeur audio comportant un dispositif de traitement tel que décrit ainsi qu'un programme informatique comportant des instructions de code pour la mise en oeuvre des étapes du procédé de traitement tel que décrit, lorsque ces instructions sont exécutées par un processeur.

[0046] Enfin l'invention se rapporte à un support de stockage, lisible par un processeur, intégré ou non au dispositif de traitement, éventuellement amovible, mémorisant un programme informatique comportant des instructions pour l'exécution du procédé de traitement tel que décrit précédemment.

[0047] D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description suivante, donnée uniquement à titre d'exemple non limitatif, et faite en référence aux dessins annexés, sur lesquels :

- la figure 1 illustre un décodeur de type MPEG-H 3D audio tel qu'il existe dans l'état de l'art ;
- La figure 2 illustre les étapes d'un procédé de traitement selon un mode de réalisation de l'invention ;
- la figure 3 illustre un décodeur comportant un dispositif de traitement selon un premier mode de réalisation de l'invention ;
- la figure 4 illustre un décodeur comportant un dispositif de traitement selon un deuxième mode de réalisation de l'invention ; et
- la figure 5 illustre une représentation matérielle d'un dispositif de traitement selon un mode de réalisation

de l'invention.

[0048] La **figure 1** illustre schématiquement un décodeur tel que normalisé dans la norme MPEG-H 3D audio selon le document référencé ci-dessus. Le bloc 101 est un module de décodage cœur qui décode à la fois des signaux audio multicanaux (Ch.) de type « canal », des signaux audio monophoniques de type « objet » (Obj.) associés à des paramètres de spatialisation (« Metadata ») (Obj.MeDa.) et des signaux audio en format audio ambiophonique d'ordre supérieur (HOA) (HOA pour « Higher Order Ambisonic » en anglais).

[0049] Un signal de type canal est décodé et traité par un moteur de rendu canal 102 (« Channel renderer » en anglais, encore appelé « Format Converter » dans MPEG-H 3D Audio) afin d'adapter ce signal canal au système de restitution audio. Le moteur de rendu canal connaît les caractéristiques du système de restitution et fournit ainsi un signal par voie de restitution (Rdr.Ch.) pour alimenter soit des haut-parleurs réels soit des haut-parleurs virtuels (qui seront alors binauralisés pour un rendu au casque).

[0050] Ces voies de restitutions sont mixées par le module de mixage 110, à d'autres voies de restitutions issues des moteurs de rendu objet 103 et HOA 105 décrits ultérieurement.

[0051] Les signaux de type objet (Obj.) sont des signaux monophoniques associés à des données (« Metadata ») comme des paramètres de spatialisation (angles azimut, élévation) qui permettent de positionner le signal monophonique dans la scène sonore spatialisée, des paramètres de priorité ou des paramètres de volume sonore. Ces signaux objet sont décodés ainsi que les paramètres associés, par le module de décodage 101 et sont traités par un moteur de rendu objet 103 (« Object Renderer » en anglais) qui, connaissant les caractéristiques du système de restitution, adapte ces signaux monophoniques à ces caractéristiques. Les différentes voies de restitution (Rdr.Obj.) ainsi créées sont mixées avec les autres voies de restitution issues des moteurs de rendu canal et HOA, par le module de mixage 110.

[0052] De la même façon, les signaux de type ambiophonique (HOA pour « Higher Order Ambisonic » en anglais) sont décodés et les composantes ambiophoniques décodées sont mis en entrée d'un moteur de rendu ambiophonique 105 (« HOA renderer » en anglais) pour adapter ces composantes au système de restitution sonore.

[0053] Les voies de restitution (Rdr.HOA) créées par ce moteur de rendu HOA sont mixées en 110 avec les voies de restitution créées par les autres moteurs de rendu 102 et 103.

[0054] Les signaux à la sortie du module de mixage 110 peuvent être restitués par des hauts parleurs réels HP situés dans une pièce de restitution. Dans ce cas, les signaux en sortie du module de mixage peuvent alimenter directement ces haut-parleurs réels, une voie cor-

respondant à un haut-parleur.

[0055] Dans le cas où les signaux en sortie du module de mixage sont à restituer sur un casque audio CA, alors ces signaux sont traités par un module de traitement de binauralisation 120 selon des techniques de binauralisation décrits par exemple dans le document cité pour la norme MPEG-H 3D audio.

[0056] Ainsi, tous les signaux destinés à être restitués sur un casque audio, sont traités par le module de traitement de binauralisation 120.

[0057] La **figure 2** décrit à présent les étapes d'un procédé de traitement selon un mode de réalisation de l'invention.

[0058] Ce procédé concerne le traitement d'un signal monophonique dans un décodeur audio 3D. Une étape E200 détecte si le flux de données (SMo) représentatif du signal monophonique (par exemple le bitstream à l'entrée du décodeur audio) comporte une indication de non-traitement de binauralisation associée à une information de position spatiale de restitution. Dans le cas contraire (N à l'étape E200), le signal doit être binauralisé. Il est traité par un traitement de binauralisation, à l'étape E210, avant d'être restitué en E240 sur un casque audio de restitution. Ce signal binauralisé peut être mixé avec d'autres signaux stéréophoniques issus de l'étape E220 décrite ci-dessous.

[0059] Dans le cas où le le flux de données représentatif du signal monophonique comporte à la fois une indication de non-binauralisation (Di.) et une information de position spatiale de restitution (Pos.) (O à l'étape E200), le signal monophonique décodé est dirigé vers un moteur de rendu stéréophonique pour être traité par une étape E220.

[0060] Cette indication de non-binauralisation peut être par exemple comme dans l'état de l'art, une identification « Dichotic » donnée au signal monophonique ou une autre identification comprise comme une instruction de ne pas traiter le signal par un traitement de binauralisation. L'information de position spatiale de restitution peut être par exemple un angle azimut indiquant la position de restitution du son par rapport à une oreille, droite ou gauche, ou encore une indication de différence de niveau entre les voies gauche et droite comme une information d'ILD permettant de répartir l'énergie du signal monophonique entre les voies gauche et droite, ou encore simplement l'indication d'une seule voie de restitution, correspondant à l'oreille droite ou gauche. Dans ce dernier cas, cette information est une information binaire qui ne nécessite que très peu de débit (1 seul bit d'information).

[0061] A l'étape E220, l'information de position est prise en compte pour construire deux voies de restitution pour les deux écouteurs du casque audio. Ces deux voies de restitution ainsi construites sont traitées directement par une étape de mixage direct E230 sommant ces deux voies stéréophoniques avec les deux voies du signal binauralisé issues du traitement de binauralisation E210.

[0062] Chacune des voies de restitution stéréophoni-

que est alors sommée avec la voie correspondante du signal binauralisé.

[0063] Suite à cette étape de mixage direct, les deux voies de restitution issues de l'étape de mixage E230 sont restituées en E240 sur le casque audio CA.

[0064] Dans un mode de réalisation où l'information de position spatiale de restitution est une donnée binaire indiquant une seule voie du casque audio de restitution, cela veut dire que le signal monophonique doit être restitué uniquement sur un écouteur de ce casque. Les deux voies de restitution construites à l'étape E220 par le moteur de rendu stéréophonique sont constituées d'une voie comportant le signal monophonique, l'autre voie étant nulle, et donc possiblement absente.

[0065] A l'étape de mixage direct E230, une seule voie est donc sommée avec la voie correspondante du signal binauralisé, l'autre voie étant nulle. Cette étape de mixage est donc simplifiée.

[0066] Ainsi, l'auditeur muni du casque audio entend d'une part, une scène sonore spatialisée provenant du signal binauralisé, cette scène sonore est entendue par lui au même endroit physique même s'il bouge la tête dans le cas d'un rendu dynamique et d'autre part, un son positionné à l'intérieur de la tête, entre une oreille et le centre de la tête, qui se superpose à la scène sonore de façon indépendante, c'est-à-dire que si l'auditeur bouge la tête, ce son sera entendu à la même position par rapport à une oreille.

[0067] Ce son est donc perçu en superposition des autres sons binauralisés de la scène sonore, et agira par exemple comme une voix « OFF » à cette scène sonore.

[0068] L'effet « oreillette » est alors réalisé.

[0069] La **figure 3** illustre un premier mode de réalisation d'un décodeur comportant un dispositif de traitement mettant en oeuvre le procédé de traitement décrit en référence à la figure 2. Dans cet exemple de réalisation, le signal monophonique traité par le procédé mis en oeuvre est un signal de type canal (Ch.).

[0070] Les signaux de type objet (obj.) et de type HOA (HOA) sont traités de la même façon par les blocs respectifs 303, 304 et 305 que les blocs 103, 104 et 105 décrits en référence à la figure 1. De la même façon, le bloc de mixage 310 effectue un mixage tel que décrit pour le bloc 110 de la figure 1.

[0071] Le bloc 330 recevant les signaux de type canal traite différemment un signal monophonique comportant une indication de non-binauralisation (Di.) associée à une information de position spatiale de restitution (Pos.) qu'un autre signal ne comportant pas ces informations, en particulier un signal multicanal. Pour ces signaux ne comportant pas ces informations, ils sont traités par le bloc 302 de la même façon que le bloc 102 décrit en référence à la figure 1.

[0072] Pour un signal monophonique comportant l'indication de non-binauralisation associée à une information de position spatiale de restitution, le bloc 330 agit comme un routeur ou interrupteur et dirige le signal monophonique décodé (Mo.) vers un moteur de rendu sté-

réophonique 331. Ce moteur de rendu stéréophonique reçoit par ailleurs, du module de décodage, l'information de position spatiale de restitution (Pos.). Avec cette information, il construit deux voies de restitution (2 Vo.), correspondants aux voies gauche et droite du casque audio de restitution, pour que ces voies soient restituées sur le casque audio CA.

[0073] Dans un exemple de réalisation, l'information de position spatiale de restitution est une information de différence interaurale de niveau sonore entre les voies gauche et droite. Cette information permet de définir un facteur à appliquer à chacune des voies de restitution pour respecter cette position spatiale de restitution.

[0074] La définition de ces facteurs peut s'effectuer comme dans le document référencé MPEG-2 AAC: ISO/IEC 13818-4:2004/DCOR 2, AAC dans la section 7.2 décrivant l'intensité stéré.

[0075] Avant d'être restituées sur le casque audio, ces voies de restitution sont ajoutées aux voies d'un signal binauralisé issu du module de binauralisation 320 qui effectue un traitement de binauralisation de la même façon que le bloc 120 de la figure 1.

[0076] Cette étape de sommation des voies s'effectue par le module de mixage direct 340 qui somme la voie gauche issue du moteur de rendu stéréophonique 331 à la voie gauche du signal binauralisé issu du module de traitement de binauralisation 320 et la voie droite issue du moteur de rendu stéréophonique 331 à la voie droite du signal binauralisé issu du module de traitement de binauralisation 320, avant la restitution sur le casque CA.

[0077] Ainsi, le signal monophonique ne passe pas par le module de traitement de binauralisation 320, il est transmis directement au moteur de rendu stéréophonique 331 avant d'être mixé directement à un signal binauralisé.

[0078] Ce signal ne subira donc pas non plus de traitement de suivi de la tête. Le son restitué sera donc à une position de restitution par rapport à une oreille de l'auditeur et restera à cette position même si l'auditeur bouge sa tête.

[0079] Dans ce mode de réalisation, le moteur de rendu stéréophonique 331 peut être intégré au moteur de rendu canal 302. Dans ce cas, ce moteur de rendu canal met en oeuvre à la fois l'adaptation des signaux de type canal classiques, comme décrit à la figure 1 et la construction des deux voies de restitution du moteur de rendu 331 comme expliqué ci-dessus en recevant l'information de position spatiale de restitution (Pos.). Seules les deux voies de restitution sont alors redirigées vers le module de mixage direct 340 avant restitution sur le casque audio CA.

[0080] Dans une variante de réalisation, le moteur de rendu stéréophonique 331 est intégré au module de mixage direct 340. Dans ce cas, le module de routage 330, dirige le signal monophonique décodé (pour lequel il a été détecté l'indication de non-binauralisation et l'information de position spatiale de restitution) vers le module de mixage direct 340. D'autre part, l'information de

position spatiale de restitution (Pos.) décodée est transmise également au module de mixage direct 340. Ce module de mixage direct comportant alors le moteur de rendu stéréophonique, met en oeuvre la construction des deux voies de restitution prenant en compte l'information de position spatiale de restitution ainsi que le mixage de ces deux voies de restitution avec les voies de restitution d'un signal binauralisé issu du module de traitement de binauralisation 320.

[0081] La figure 4 illustre un deuxième mode de réalisation d'un décodeur comportant un dispositif de traitement mettant en oeuvre le procédé de traitement décrit en référence à la figure 2. Dans cet exemple de réalisation, le signal monophonique traité par le procédé mis en oeuvre est un signal de type objet (Obj.).

[0082] Les signaux de type canal (Ch.) et de type HOA (HOA) sont traités de la même façon par les blocs respectifs 402 et 405 que les blocs 102 et 105 décrits en référence à la figure 1. De la même façon, le bloc de mixage 410 effectue un mixage tel que décrit pour le bloc 110 de la figure 1.

[0083] Le bloc 430 recevant les signaux de type objet (Obj.) traite différemment un signal monophonique pour lequel il a été détecté une indication de non-binauralisation (Di.) associée à une information de position spatiale de restitution (Pos.) qu'un autre signal monophonique pour lequel ces informations n'ont pas été détectées.

[0084] Pour ces signaux monophoniques pour lequel ces informations n'ont pas été détectés, ils sont traités par le bloc 403 de la même façon que le bloc 103 décrit en référence à la figure 1 en utilisant les paramètres décodés du bloc 404 décodant les Metadata de la même façon que le bloc 104 de la figure 1.

[0085] Pour un signal monophonique de type objet pour lequel il a été détecté l'indication de non-binauralisation associée à une information de position spatiale de restitution, le bloc 430 agit comme un routeur ou interrupteur et dirige le signal monophonique décodé (Mo.) vers un moteur de rendu stéréophonique 431.

[0086] L'indication de non-binauralisation (Di.) ainsi que l'information de position spatiale de restitution (Pos.) sont décodées par le bloc de décodage 404 des metadata ou paramètres associés aux signaux de type objet. L'indication de non-binauralisation (Di.) est transmise au bloc de routage 430 et l'information de position spatiale de restitution est transmise au moteur de rendu stéréophonique 431.

[0087] Ce moteur de rendu stéréophonique recevant ainsi l'information de position spatiale de restitution (Pos.), construit deux voies de restitution, correspondants aux voies gauche et droite du casque audio de restitution, pour que ces voies soient restituées sur le casque audio CA.

[0088] Dans un exemple de réalisation, l'information de position spatiale de restitution est une information d'angle azimut définissant un angle entre la position de restitution voulue et le centre de la tête de l'auditeur.

[0089] Cette information permet de définir un facteur

à appliquer à chacune des voies de restitution pour respecter cette position spatiale de restitution.

[0090] Les facteurs de gains pour les voies gauche et droite peuvent être calculés de la manière présentée dans le document intitulé « Virtual Sound Source Positioning Using Vector Base Amplitude Panning » de Ville Pulkki dans J. Audio Eng. Soc., Vol.45, No.6, de Juin 1997.

[0091] Par exemple, les facteurs de gain du moteur de rendu stéréophonique peuvent être donnés par:

$$g_1 = (\cos O \cdot \sin H + \sin O \cdot \cos H) / (2 \cdot \cos H \cdot \sin H)$$

$$g_2 = (\cos O \cdot \sin H - \sin O \cdot \cos H) / (2 \cdot \cos H \cdot \sin H)$$

[0092] Où g_1 et g_2 correspondent aux facteurs pour les signaux des voies gauche et droite, O est l'angle entre la direction frontale et l'objet (nommé azimut), et H est l'angle entre la direction frontale et la position du haut-parleur virtuel (correspondant au demi-angle entre les haut-parleurs), fixé par exemple à 45° .

[0093] Avant d'être restituées sur le casque audio, ces voies de restitution sont ajoutées aux voies d'un signal binauralisé issu du module de binauralisation 420 qui effectue un traitement de binauralisation de la même façon que le bloc 120 de la figure 1.

[0094] Cette étape de sommation des voies s'effectue par le module de mixage direct 440 qui somme la voie gauche issue du moteur de rendu stéréophonique 431 à la voie gauche du signal binauralisé issu du module de traitement de binauralisation 420 et la voie droite issue du moteur de rendu stéréophonique 431 à la voie droite du signal binauralisé issu du module de traitement de binauralisation 420, avant la restitution sur le casque CA.

[0095] Ainsi, le signal monophonique ne passe pas par le module de traitement de binauralisation 420, il est transmis directement au moteur de rendu stéréophonique 431 avant d'être mixé directement à un signal binauralisé.

[0096] Ce signal ne subira donc pas non plus de traitement de suivi de la tête. Le son restitué sera donc à une position de restitution par rapport à une oreille de l'auditeur et restera à cette position même si l'auditeur bouge sa tête.

[0097] Dans ce mode de réalisation, le moteur de rendu stéréophonique 431 peut être intégré au moteur de rendu objet 403. Dans ce cas, ce moteur de rendu objet met en oeuvre à la fois l'adaptation des signaux de type objet classiques, comme décrit à la figure 1 et la construction des deux voies de restitution du moteur de rendu 431 comme expliqué ci-dessus en recevant l'information de position spatiale de restitution (Pos.) du module de décodage 404 des paramètres. Seules les deux voies de restitution (2Vo.) sont alors redirigées vers le module de mixage direct 440 avant restitution sur le casque audio CA.

[0098] Dans une variante de réalisation, le moteur de rendu stéréophonique 431 est intégré au module de mixage direct 440. Dans ce cas, le module de routage 430, dirige le signal monophonique décodé (Mo.) (pour lequel il a été détecté l'indication de non-binauralisation et l'information de position spatiale de restitution) vers le module de mixage direct 440. D'autre part, l'information de position spatiale de restitution (Pos.) décodée est transmise également au module de mixage direct 440 par le module de décodage des paramètres 404. Ce module de mixage direct comportant alors le moteur de rendu stéréophonique, met en oeuvre la construction des deux voies de restitution prenant en compte l'information de position spatiale de restitution ainsi que le mixage de ces deux voies de restitution avec les voies de restitution d'un signal binauralisé issu du module de traitement de binauralisation 420.

[0099] La figure 5 illustre à présent un exemple de réalisation matérielle d'un dispositif de traitement apte à mettre en oeuvre le procédé de traitement selon l'invention.

[0100] Le dispositif DIS comporte un espace de stockage 530, par exemple une mémoire MEM, une unité de traitement 520 comportant un processeur PROC, piloté par un programme informatique Pg, stocké dans la mémoire 530 et mettant en oeuvre le procédé de traitement selon l'invention.

[0101] Le programme informatique Pg comporte des instructions de code pour la mise en oeuvre des étapes du procédé de traitement au sens de l'invention, lorsque ces instructions sont exécutées par le processeur PROC, et notamment, à la détection, dans un flux de données représentatif du signal monophonique, d'une indication de non-traitement de binauralisation associée à une information de position spatiale de restitution, une étape de direction du signal monophonique décodé vers un moteur de rendu stéréophonique prenant en compte l'information de position pour construire deux voies de restitution traitées directement par une étape de mixage direct sommant ces deux voies avec un signal binauralisé issu du traitement de binauralisation, pour être restitué sur le casque audio.

[0102] Typiquement, la description de la figure 2 reprend les étapes d'un algorithme d'un tel programme informatique.

[0103] A l'initialisation, les instructions de code du programme Pg sont par exemple chargées dans une mémoire RAM (non représentée) avant d'être exécutées par le processeur PROC de l'unité de traitement 520. Les instructions de programme peuvent être mémorisées sur un support de stockage tel qu'une mémoire flash, un disque dur ou tout autre support de stockage non-transitoire.

[0104] Le dispositif DIS comporte un module de réception 510 apte à recevoir un flux de données SMO représentatif notamment d'un signal monophonique. Il comprend un module de détection 540 apte à détecter, dans ce flux de données, une indication de non-traitement de binauralisation associée à une information de position

spatiale de restitution. Il comprend un module de direction 550, dans le cas d'une détection positive par le module de détection 540, du signal monophonique décodé vers un moteur de rendu stéréophonique 560, le moteur de rendu stéréophonique 560 étant apte à prendre en compte l'information de position pour construire deux voies de restitution.

[0105] Le dispositif DIS comporte également un module de mixage direct 570 apte à traiter directement les deux voies de restitution en les sommant avec les deux voies d'un signal binauralisé issu d'un module de traitement de binauralisation. Les voies de restitution ainsi obtenues sont transmises à un casque audio CA via un module de sortie 560, pour être restituées.

[0106] Ces différents modules sont tels que décrits en référence aux figures 3 et 4 selon les modes de réalisation.

[0107] Le terme module peut correspondre aussi bien à un composant logiciel qu'à un composant matériel ou un ensemble de composants matériels et logiciels, un composant logiciel correspondant lui-même à un ou plusieurs programmes ou sous-programmes d'ordinateur ou de manière plus générale à tout élément d'un programme apte à mettre en oeuvre une fonction ou un ensemble de fonctions telles que décrites pour les modules concernés. De la même manière, un composant matériel correspond à tout élément d'un ensemble matériel (ou hardware) apte à mettre en oeuvre une fonction ou un ensemble de fonctions pour le module concerné (circuit intégré, carte à puce, carte à mémoire, etc.)

[0108] Le dispositif peut être intégré dans un décodeur audio tel que décrit en figure 3 ou 4 et peut être intégré par exemple dans des équipements multimédia de type décodeur de salon, "set top box" ou lecteur de contenu audio ou vidéo. Ils peuvent également être intégrés dans des équipements de communication de type téléphone mobile ou passerelle de communication.

Revendications

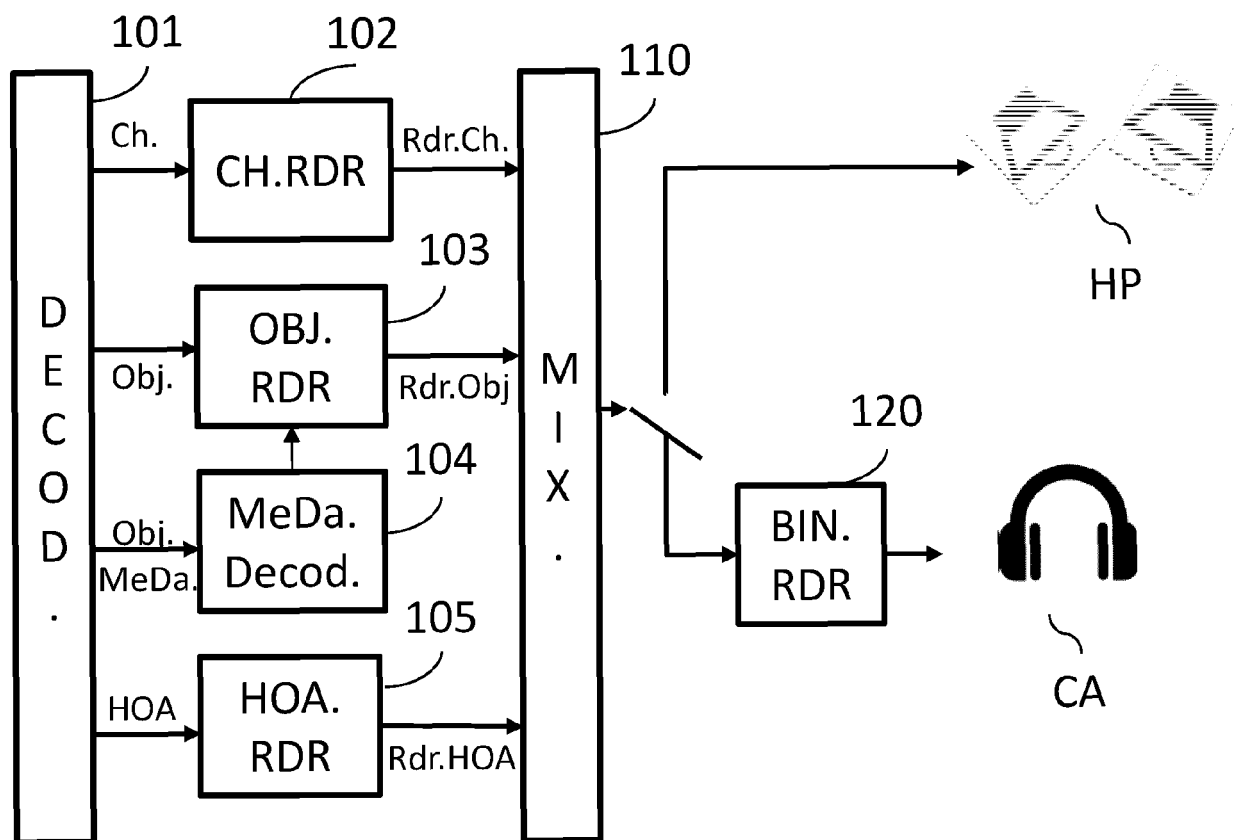
1. Procédé de traitement d'un signal monophonique audio dans un décodeur audio 3D comportant une étape de traitement de binauralisation des signaux décodés destinés à être restitué spatialement par un casque audio, **caractérisé en ce que,** à la détection (E200), dans un flux de données représentatif du signal monophonique, d'une indication de non-traitement de binauralisation associée à une information de position spatiale de restitution, le signal monophonique décodé est dirigé (O-E200) directement vers un module de mixage comportant un moteur de rendu stéréophonique prenant en compte l'information de position pour construire deux voies de restitution (E220) traitées directement par une étape de mixage direct (E230) sommant ces deux voies avec un signal binauralisé issu du traitement de binauralisation, pour être restitué (E240) sur le

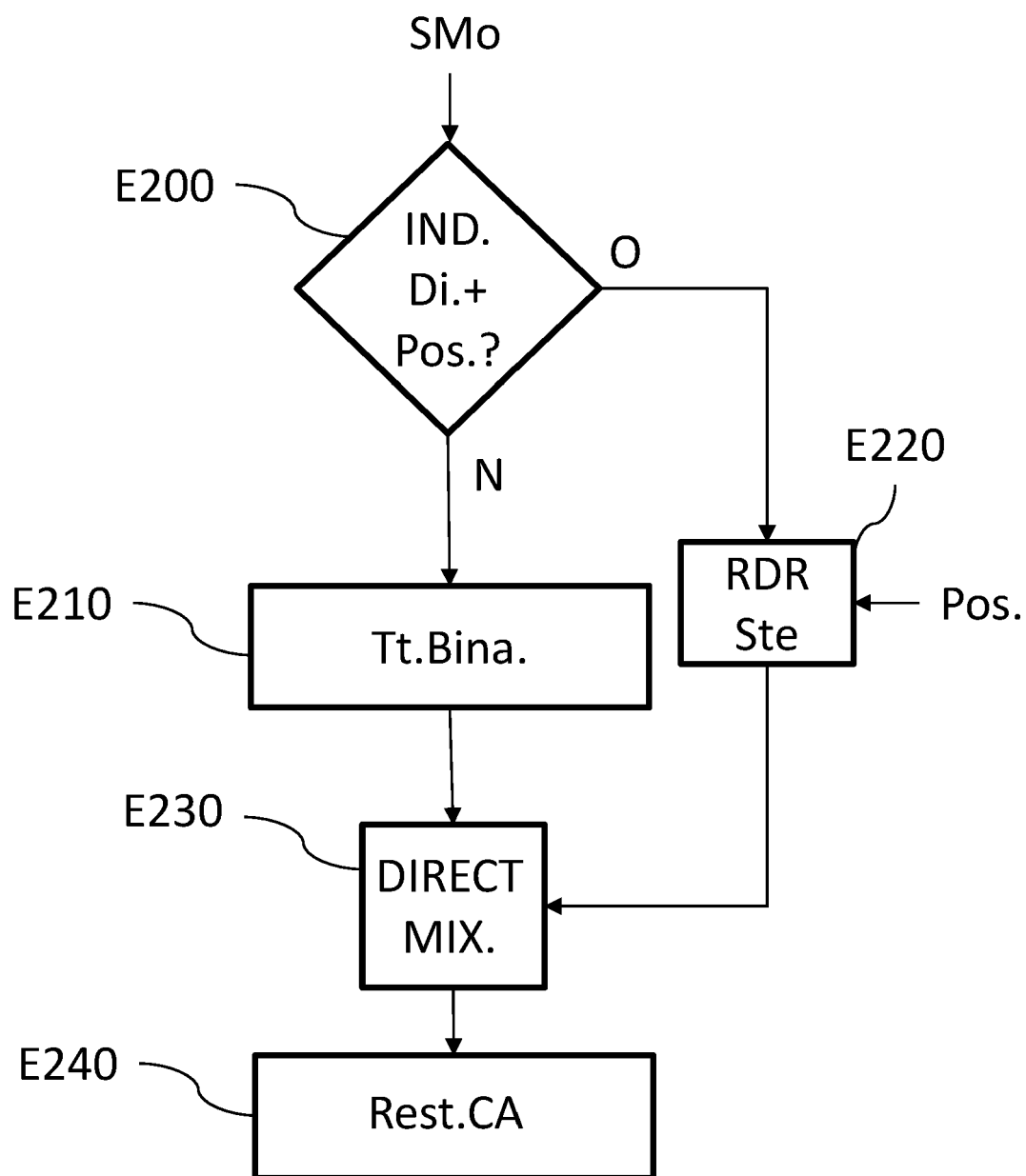
casque audio.

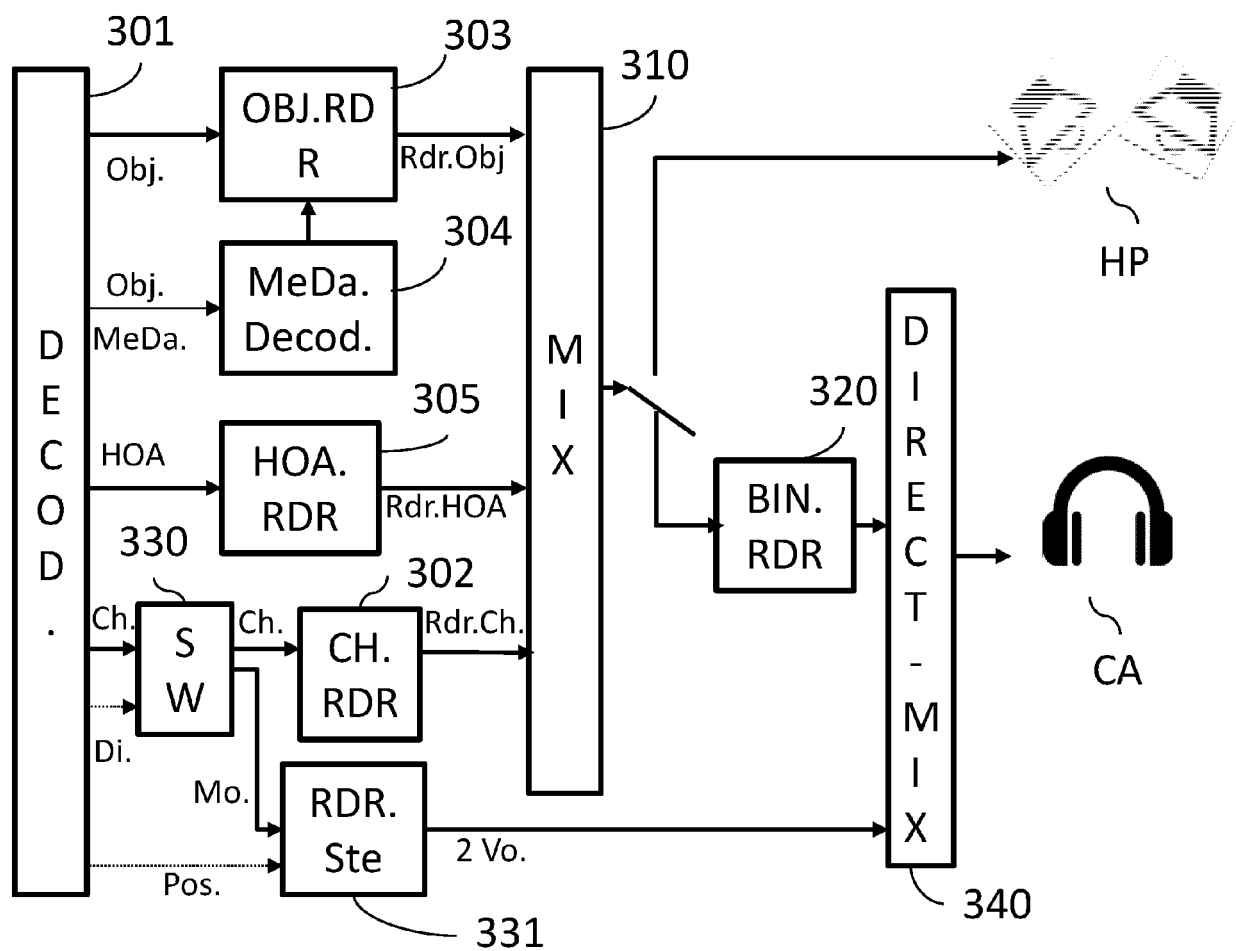
2. Procédé selon la revendication 1, dans lequel l'information de position spatiale de restitution est une donnée binaire indiquant une seule voie du casque audio de restitution. 5
3. Procédé selon la revendication 2, dans lequel seule la voie de restitution correspondant à la voie indiquée par la donnée binaire est sommée à la voie correspondante du signal binauralisé à l'étape de mixage direct, l'autre voie de restitution étant de valeur nulle. 10
4. Procédé selon la revendication 1, dans lequel le signal monophonique est un signal de type canal dirigé vers le moteur de rendu stéréophonique, avec l'information de position spatiale de restitution. 15
5. Procédé selon la revendication 4, dans lequel l'information de position spatiale de restitution est une donnée de différence interaurale de niveau sonore (ILD). 20
6. Procédé selon la revendication 1, dans lequel le signal monophonique est un signal de type objet associé à un ensemble de paramètres de restitution comprenant l'indication de non-binauralisation et l'information de position de restitution, le signal étant dirigé vers le module de mixage avec l'information de position de restitution. 25
7. Procédé selon la revendication 6, dans lequel l'information de position spatiale de restitution est une donnée d'angle azimut. 30
8. Dispositif de traitement d'un signal monophonique audio comportant un module de traitement de binauralisation de signaux décodés destinés à être restitués spatialement par un casque audio, **caractérisé en ce qu'il comporte :** 35
 - un module de détection (330 ;430) apte à détecter, dans un flux de données représentatif du signal monophonique, une indication de non-traitement de binauralisation associée à une information de position spatiale de restitution ; 40
 - un module de redirection (330, 430), dans le cas d'une détection positive par le module de détection, apte à diriger le signal monophonique décodé directement vers un module de mixage; 45
 - un module de mixage (340 ; 440) comportant un moteur de rendu stéréophonique (331 ; 431) apte à prendre en compte l'information de position pour construire deux voies de restitution, le module de mixage étant apte à traiter directement les deux voies de restitution en les sommant avec un signal binauralisé issu du module de traitement de binauralisation (320 ; 420), 50

pour être restitué sur le casque audio.

9. Dispositif selon la revendication 8, dans lequel le signal monophonique est un signal de type canal et dans lequel le moteur de rendu stéréophonique est intégré à un moteur de rendu canal construisant par ailleurs des voies de restitution pour des signaux à plusieurs canaux. 5
10. Dispositif selon la revendication 8, dans lequel le signal monophonique est un signal de type objet et dans lequel le moteur de rendu stéréophonique est intégré à un moteur de rendu objet construisant par ailleurs des voies de restitution pour des signaux monophoniques associées à des ensembles de paramètres de restitution. 10
11. Décodeur audio comportant un dispositif de traitement selon l'une des revendications 8 à 10. 15
12. Programme informatique comportant des instructions de code pour la mise en oeuvre des étapes du procédé de traitement selon l'une des revendications 1 à 7, lorsque ces instructions sont exécutées par un processeur. 20
13. Support de stockage, lisible par un processeur, mémorisant un programme informatique comportant des instructions pour l'exécution du procédé de traitement selon l'une des revendications 1 à 7. 25

**FIG. 1 (Etat de l'art)**

**FIG. 2**

**FIG. 3**

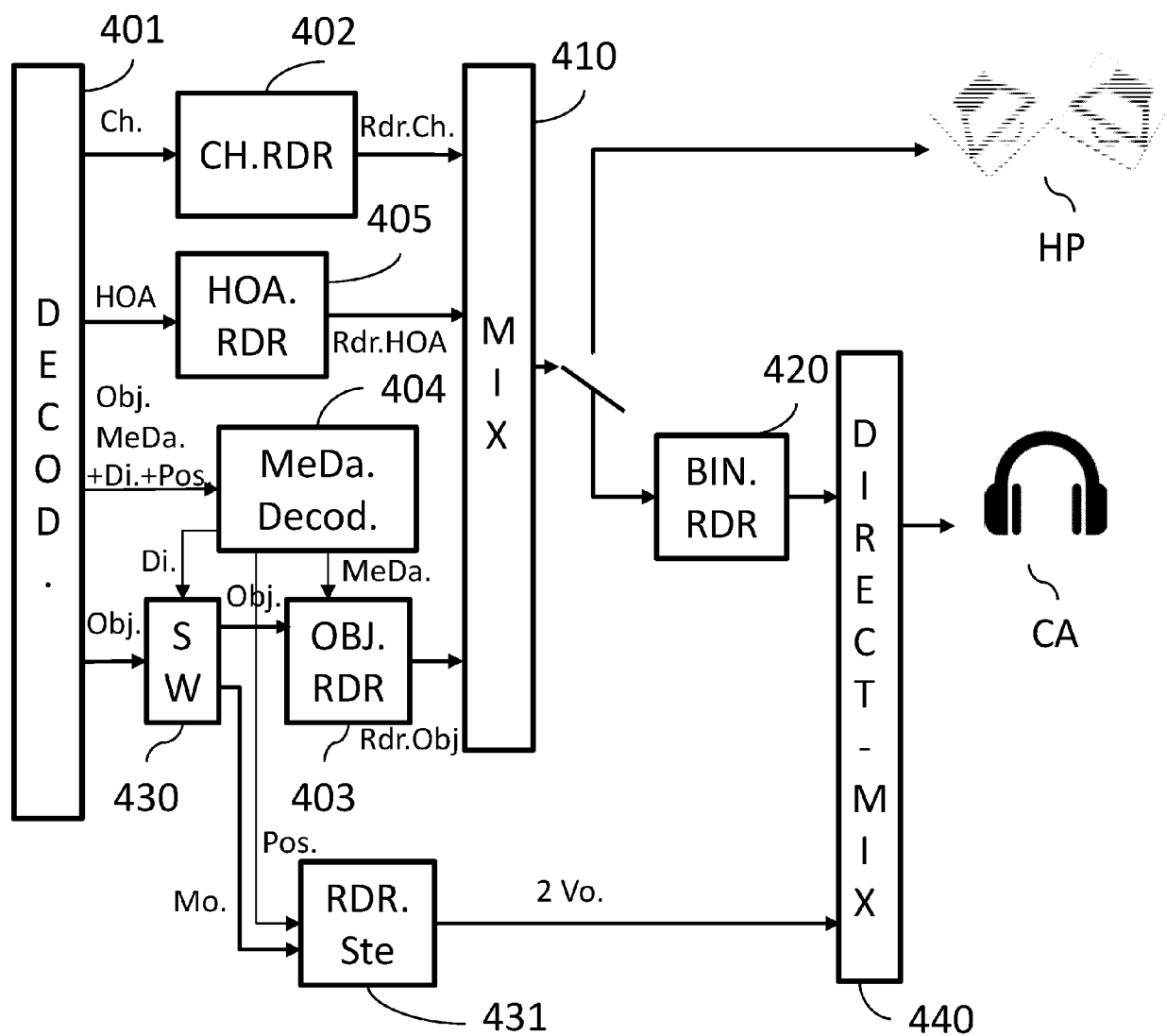


FIG. 4

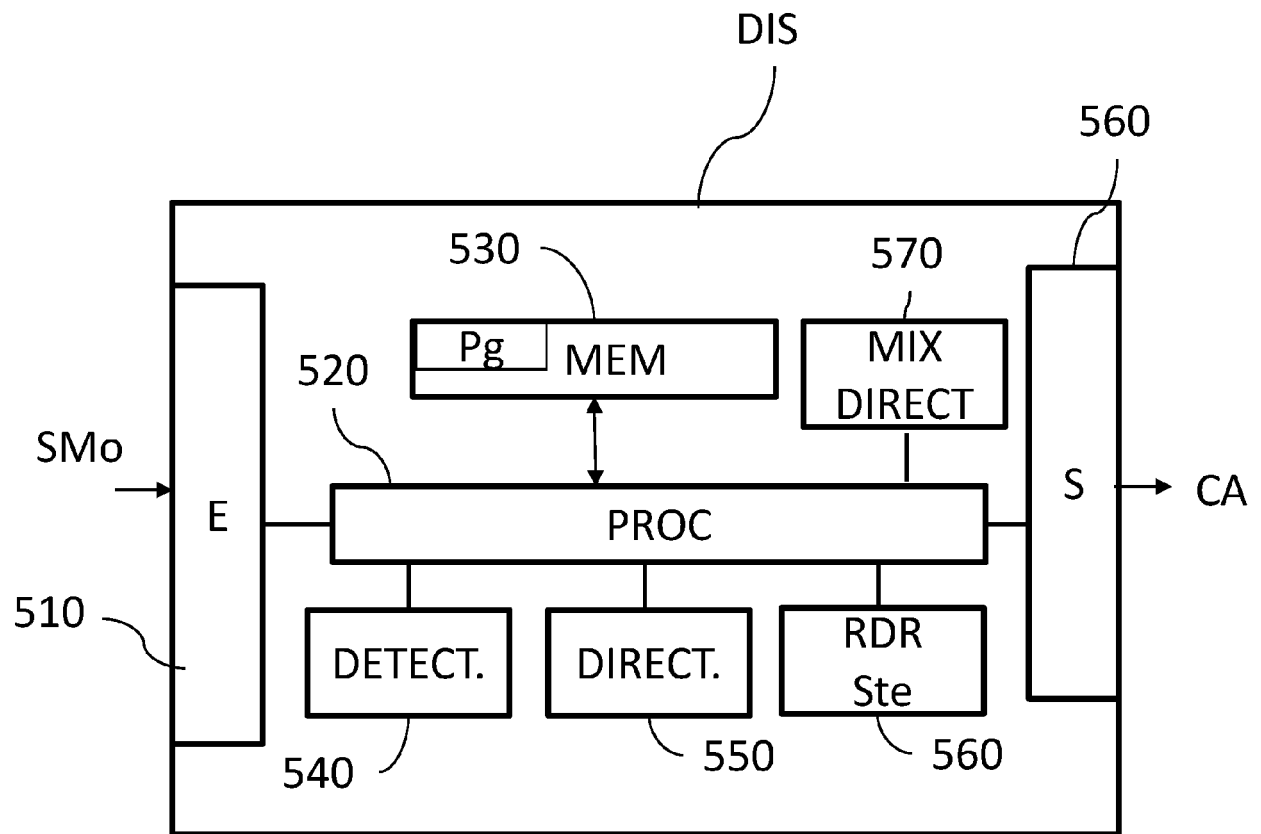


FIG. 5



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande

EP 22 19 7901

5

10

15

20

25

30

35

40

45

50

55

1

EPO FORM 1503 03.82 (P04C02)

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
X	US 2016/266865 A1 (TSINGOS NICOLAS R [US] ET AL) 15 septembre 2016 (2016-09-15)	1, 4-13	INV.
A	* alinéa [0035] - alinéa [0044]; figures 1-5 *	2, 3	H04S7/00 G10L19/008
A	US 2016/300577 A1 (FERSCH CHRISTOF [DE] ET AL) 13 octobre 2016 (2016-10-13) * alinéa [0022] - alinéa [0030]; figure 1; tableau 1 *	2, 3	
			DOMAINES TECHNIQUES RECHERCHES (IPC)
			H04S G10L
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche		Date d'achèvement de la recherche	Examineur
Munich		9 janvier 2023	Joder, Cyril
CATEGORIE DES DOCUMENTS CITES		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire			

**ANNEXE AU RAPPORT DE RECHERCHE EUROPEENNE
RELATIF A LA DEMANDE DE BREVET EUROPEEN NO.**

EP 22 19 7901

5 La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche européenne visé ci-dessus.
Lesdits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

09-01-2023

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 2016266865 A1	15-09-2016	CN 105684467 A	15-06-2016
		CN 108712711 A	26-10-2018
		CN 109040946 A	18-12-2018
		CN 109068263 A	21-12-2018
		CN 113630711 A	09-11-2021
		EP 3063955 A1	07-09-2016
		EP 3672285 A1	24-06-2020
		ES 2755349 T3	22-04-2020
		US 2016266865 A1	15-09-2016
		US 2018210695 A1	26-07-2018
		US 2019205085 A1	04-07-2019
		US 2020065055 A1	27-02-2020
		US 2021132894 A1	06-05-2021
		US 2022269471 A1	25-08-2022
US 2016300577 A1	13-10-2016	WO 2015066062 A1	07-05-2015
		CN 106162500 A	23-11-2016
		CN 111586533 A	25-08-2020
		US 2016300577 A1	13-10-2016

Pour tout renseignement concernant cette annexe : voir Journal Officiel de l'Office européen des brevets, No.12/82

RÉFÉRENCES CITÉES DANS LA DESCRIPTION

Cette liste de références citées par le demandeur vise uniquement à aider le lecteur et ne fait pas partie du document de brevet européen. Même si le plus grand soin a été accordé à sa conception, des erreurs ou des omissions ne peuvent être exclues et l'OEB décline toute responsabilité à cet égard.

Littérature non-brevet citée dans la description

- **VILLE PULKKI.** Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *J. Audio Eng. Soc.*, Juin 1997, vol. 45 (6 [0090])