

(19)



(11)

**EP 3 058 564 B1**

(12)

**FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention de la délivrance du brevet:

**26.07.2023 Bulletin 2023/30**

(51) Classification Internationale des Brevets (IPC):

**G10K 15/12 (2006.01) H04S 1/00 (2006.01)**  
**H04S 3/00 (2006.01) H04S 7/00 (2006.01)**

(21) Numéro de dépôt: **14796814.3**

(52) Classification Coopérative des Brevets (CPC):

**G10K 15/12; H04S 7/306; H04S 2400/01;**  
**H04S 2420/01; H04S 2420/07**

(22) Date de dépôt: **14.10.2014**

(86) Numéro de dépôt international:

**PCT/FR2014/052617**

(87) Numéro de publication internationale:

**WO 2015/055946 (23.04.2015 Gazette 2015/16)**

(54) **SPATIALISATION SONORE AVEC EFFET DE SALLE, OPTIMISEE EN COMPLEXITE**

**KOMPLEXITÄTSOPTIMIERTE KLANGVERRÄUMLICHUNG MIT NACHHALL**

**SOUND SPATIALISATION WITH REVERBERATION, OPTIMISED IN TERMS OF COMPLEXITY**

(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**

(72) Inventeurs:

- **PALLONE, Grégory**  
**35000 Rennes (FR)**
- **EMERIT, Marc**  
**F-35000 Rennes (FR)**

(30) Priorité: **18.10.2013 FR 1360185**

(74) Mandataire: **Plasseraud IP**

**66, rue de la Chaussée d'Antin**  
**75440 Paris Cedex 09 (FR)**

(43) Date de publication de la demande:

**24.08.2016 Bulletin 2016/34**

(60) Demande divisionnaire:

**22211949.7 / 4 184 505**

(56) Documents cités:

**EP-A1- 2 840 811 US-A- 5 917 917**  
**US-A1- 2008 085 008 US-A1- 2011 170 721**  
**US-A1- 2012 201 389 US-B1- 7 835 535**

(73) Titulaire: **Orange**

**92130 Issy-les-Moulineaux (FR)**

**EP 3 058 564 B1**

Il est rappelé que: Dans un délai de neuf mois à compter de la publication de la mention de la délivrance du brevet européen au Bulletin européen des brevets, toute personne peut faire opposition à ce brevet auprès de l'Office européen des brevets, conformément au règlement d'exécution. L'opposition n'est réputée formée qu'après le paiement de la taxe d'opposition. (Art. 99(1) Convention sur le brevet européen).

## Description

[0001] La présente invention concerne une spatialisation sonore avec effet de salle.

[0002] L'invention trouve une application avantageuse mais non limitative à un traitement de signaux sonores respectivement issus de L canaux associés à des haut-parleurs virtuels (par exemple dans une représentation multicanale, ou encore dans une représentation ambiophonique, du son à restituer), pour une restitution spatialisée sur des haut-parleurs réels (par exemple deux oreillettes d'un casque en restitution binaurale, ou deux haut-parleurs séparés en restitution transaurale).

[0003] Par exemple, le signal de l'un de ces canaux peut être traité pour avoir une première contribution sur l'oreille gauche et une deuxième contribution sur l'oreille droite, en restitution binaurale, en appliquant en particulier une fonction de transfert avec effet de salle à chacune de ces contributions. L'application de ces fonctions de transfert à effet de salle contribue alors à offrir à l'auditeur un sentiment d'immersion lui permettant pratiquement de « situer dans l'espace » le haut-parleur virtuel associé à ce canal.

[0004] Dans une réalisation particulière, décrite notamment dans le document FR1357299, on applique une fonction de transfert avec effet de salle à chaque signal sonore d'un canal correspondant, dans le domaine temporel sous la forme de réponse impulsionnelle de type BRIR (pour « Binaural Room Impulse Response » ou « Réponse Impulsionnelle Binaurale de salle »). En particulier, dans ce document, incorporé ici par référence, on construit cette fonction de transfert BRIR comme la combinaison :

- d'une première fonction de transfert, propre à chaque signal, et
- d'une deuxième fonction de transfert, globale, commune à tous les signaux et caractérisant en particulier un champ diffus, dont la présence advient habituellement dans une salle après un certain temps, typiquement après les premières réflexions d'une onde sonore.

[0005] Une telle réalisation permet avantageusement d'appliquer un traitement commun à tous les signaux, ce qui correspond, dans une réalité physique, à un « mélange » des ondes acoustiques au fur et à mesure des réverbérations, donc au-delà d'une durée donnée (caractérisant un début de présence du champ diffus). Une telle réalisation permet alors de réduire la complexité des traitements de spatialisation avec effet de salle sur plusieurs canaux initiaux.

[0006] En outre, le document US 2011/170721 A1 divulgue un procédé de traitement d'au moins un signal d'entrée par un ensemble de filtres binauraux de telle sorte que les sorties peuvent être lues sur des écouteurs pour fournir une sensation d'écoute du son dans une pièce d'écoute via un ou plusieurs haut-parleurs virtuels.

[0007] Néanmoins, dans des modules de spatialisation intervenant en amont de la restitution, on cherche encore à réduire autant que possible la complexité des traitements de spatialisation. En effet, par exemple (mais non limitativement), les signaux des canaux sont reçus sous forme encodée, par un décodeur en compression. Ce décodeur envoie les signaux des canaux, une fois décodés, à un module de spatialisation pour une restitution sonore avec effet de salle, sur deux haut-parleurs. Il convient alors que cette étape de spatialisation (qui suit le décodage des signaux reçus) soit de complexité de traitement réduite pour ne pas retarder l'ensemble global des étapes de décodage et de spatialisation à la réception des signaux avant restitution.

[0008] La présente invention vient améliorer la situation.

[0009] L'invention propose à cet effet de réduire la complexité de l'application de la fonction de transfert à effet de salle, en réduisant en particulier cette complexité dans le domaine spectral. En effet, dans le domaine spectral, la convolution par une fonction de transfert devient une multiplication de composantes spectrales du signal d'une part, et d'un filtre représentant la fonction de transfert d'autre part (figure 1 commentée en détails plus loin).

[0010] L'invention part alors du constat avantageux selon lequel, après une propagation directe, une onde sonore tend à s'atténuer dans les hautes fréquences du fait des réflexions progressives sur des surfaces (murs typiquement, visage de l'auditeur, etc.) absorbant l'onde en particulier dans les hautes fréquences. En outre, l'air lui-même absorbe les composantes spectrales des fréquences les plus élevées du son pendant sa propagation. Ce phénomène est d'autant plus accru par exemple pour le champ sonore diffus, pour lequel il n'est pas nécessaire d'avoir une représentation fréquentielle pour des fréquences très élevées (par exemple supérieures à une fréquence dans une gamme de 5 à 15 kHz).

[0011] Ainsi, il est possible de réduire la complexité du traitement de l'application de la fonction de transfert avec effet de salle, dans le domaine spectral, en ne tenant simplement pas compte, pour effectuer les multiplications de composantes spectrales précitées, des composantes associées à des fréquences supérieures à une fréquence de coupure prédéterminée (par exemple supérieures à 5 à 15 kHz).

[0012] L'invention vise alors un procédé de spatialisation sonore, comportant l'application d'au moins une fonction de transfert à effet de salle à au moins un signal sonore, ladite application revenant à multiplier, dans le domaine spectral, des composantes spectrales du signal sonore par les composantes spectrales d'un filtre correspondant à la fonction de transfert précitée. Chaque composante spectrale du filtre comporte une évolution temporelle dans une représentation

temps-fréquence (comme détaillé plus loin en référence à la figure 3).

**[0013]** En particulier, ces composantes spectrales du filtre sont ignorées, pour les multiplications de composantes précitées, au-delà d'une fréquence seuil et après au moins un instant donné dans ladite représentation temps-fréquence. Ainsi, après cet instant donné, les composantes spectrales du filtre sont prises en compte jusqu'à une fréquence de coupure qui peut être choisie par exemple entre 5 et 15 kHz (en fonction de l'effet de salle à appliquer et/ou du signal à spatialiser, comme décrit ci-après). Au-delà de la fréquence de coupure, la multiplication n'est même pas effectuée, ce qui mathématiquement revient au même que de multiplier le signal par zéro.

**[0014]** Cet instant donné représente typiquement le moment où une onde sonore commence à subir des réverbérations (par réflexions successives, ou, plus tard encore, à partir d'une présence d'un champ sonore diffus). Ainsi, en termes généraux, dans une réalisation où la fonction de transfert tient compte de réverbérations dans l'effet de salle (avec prise en compte par exemple d'un champ sonore diffus), l'instant donné précité peut être choisi en fonction de telles réverbérations. Par exemple, l'instant donné précité peut être postérieur, dans l'effet de salle, à une propagation sonore directe avec des premières réflexions, et correspondre alors à un début de présence de champ sonore diffus.

**[0015]** En outre, on peut prévoir une réalisation dans laquelle la fréquence seuil précitée diminue en fonction du temps dans ladite représentation temps-fréquence. Par exemple, si le signal est échantillonné sur plusieurs blocs successifs, il peut être prévu, à titre d'exemple, de conserver les composantes spectrales présentes dans le signal, dans la multiplication des composantes, pour un premier bloc, puis de les ignorer au-delà d'une première fréquence seuil pour un deuxième bloc qui suit le premier bloc, puis de les ignorer au-delà d'une deuxième fréquence seuil pour un troisième bloc qui suit le deuxième bloc, etc., la deuxième fréquence seuil étant plus basse que la première.

**[0016]** Ainsi, en termes plus génériques, dans une réalisation où le signal est échantillonné sur plusieurs blocs successifs, les composantes spectrales du filtre peuvent être ignorées, pour la multiplication des composantes :

- au-delà d'une première fréquence seuil pour un bloc donné,
- puis, au-delà d'une deuxième fréquence seuil, pour un bloc qui suit le bloc donné,

la deuxième fréquence seuil étant plus basse que la première fréquence seuil. Le bloc donné précité peut inclure par exemple des échantillons situés temporellement à des instants qui correspondent à des moments où une onde sonore a subi une ou plusieurs réflexions, avec même un début de présence de champ sonore diffus. Le bloc qui suit ce bloc donné (immédiatement ou quelques blocs plus tard) peut inclure par exemple des échantillons situés temporellement après ou à partir d'un début de présence de champ sonore diffus.

**[0017]** Une telle réalisation permet par exemple de limiter des artéfacts éventuellement audibles de limitation de signal dans les hautes fréquences pour les réverbérations, cette réalisation étant accomplie progressivement sur plusieurs blocs. Elle permet aussi de considérer plusieurs formes de fonctions de transfert (notées ci-après  $B_{mean}^k(m)$ , m étant un indice de bloc) caractérisant un champ sonore diffus. En effet, il est possible par exemple d'appliquer une fonction de transfert  $B_{mean}^k$  à un bloc donné précité, et d'appliquer une fenêtre de coupure temporellement progressive (de type « fade out ») à cette fonction de transfert  $B_{mean}^k$ , pour le bloc qui suit, pour « terminer » la présence du champ sonore diffus.

**[0018]** Dans une forme de réalisation où le procédé est mis en oeuvre par un module de spatialisation sonore recevant une pluralité de signaux d'entrée et délivrant au moins deux signaux de sortie, pour délivrer chaque signal de sortie, on applique une fonction de transfert à effet de salle, à chaque signal d'entrée, chacun desdits signaux de sortie étant donné par application de la formule suivante:

$$O^k = \sum_{l=1}^L (I(l) *_{[0; \dots; f^k(l)]} A^k(l)) + \sum_{m=1}^M \left( z^{-iDDm} \cdot \sum_{l=1}^L \left( G(I(l)) \cdot \frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0; \dots; f^k(m)]} B_{mean}^k(m)$$

- $O^k$  étant un signal de sortie, et k étant l'indice relatif à un signal de sortie,
- $l \in [1; L]$ , étant l'indice relatif à un signal d'entrée parmi lesdits signaux d'entrée, L étant le nombre de signaux d'entrée, et  $I(l)$  étant un signal d'entrée parmi lesdits signaux d'entrée,
- $A^k(l)$  étant une fonction de transfert avec effet de salle spécifique à un signal d'entrée,
- $B_{mean}^k(m)$  étant une fonction de transfert globale, avec effet de salle, commune aux signaux d'entrée,

- $W^k(l)$  étant un poids de pondération choisi, et  $G(l(l))$ , un gain de compensation d'énergie prédéterminé,
- $z^{-iDDm}$  étant une application de délai, compté en nombre de blocs d'échantillons, correspondant à un écart temporel entre une émission sonore dans une salle correspondant à l'effet de salle, et un début de présence de champ diffus dans cette salle, l'indice  $m$  correspondant à un nombre de blocs d'échantillons de durée correspondant à ce délai,
- 5  $M$  étant le nombre total de blocs que dure une fonction de transfert dans une représentation temps-fréquence,
- le signe « . » désignant la multiplication,
- le signe «  $*$ [ $0^k(l)$ ] » désignant l'opérateur de convolution sur un nombre de fréquences limité et allant d'une plus basse fréquence à une fréquence maximale  $f^k(l)$  qui est fonction au moins du signal d'entrée d'indice  $l$ , et
- 10 - le signe «  $*$ [ $0^k(m)$ ] » désignant l'opérateur de convolution sur un nombre de fréquences limité et allant d'une plus basse fréquence à une fréquence  $f^k(m)$  qui est fonction du bloc d'échantillons d'indice  $m$ .

**[0019]** Cette réalisation sera décrite en détails plus loin en référence aux figures 2 et 5, notamment.

**[0020]** On peut aussi appliquer une limitation de calculs de multiplications au-delà d'une première fréquence seuil, dès le ou les premiers blocs d'échantillons, en fonction des caractéristiques du signal (par exemple sa fréquence d'échantillonnage, ou la fréquence la plus élevée représentée dans les composantes spectrales du signal) ou en fonction des caractéristiques de spatialisation appliquée (avec par exemple une limitation des composantes de hautes fréquences pour un trajet acoustique contra-latéral comme détaillé plus loin).

**[0021]** Dans ce cas, le signal issu des réverbérations (après réflexions ou dans le champ diffus) ne comporte pas, normalement, de composantes spectrales de fréquence plus élevée que le signal initial. Ainsi, la fréquence seuil précitée ne peut pas être plus grande que cette fréquence la plus élevée.

**[0022]** Ainsi, en termes plus génériques, dans une forme de réalisation, on obtient une information de composante spectrale de fréquence la plus élevée dans le signal sonore, et la fréquence seuil précitée est choisie comme étant le minimum parmi une fréquence seuil prédéterminée (par exemple entre 5 et 15 kHz) et ladite fréquence la plus élevée.

**[0023]** Typiquement, dans une réalisation où le signal sonore est issu d'un décodeur en compression, l'information de composante spectrale de fréquence la plus élevée peut être fournie par le décodeur.

**[0024]** De même, si la spatialisation est réalisée auprès d'un module capable de supporter différents formats de signaux, notamment en termes de fréquence d'échantillonnage de tels signaux, la fréquence la plus élevée, précitée, ne peut pas être supérieure à la moitié de la fréquence d'échantillonnage, et, ainsi, la fréquence seuil pour la mise en oeuvre de l'invention peut être choisie en outre en fonction de cette fréquence d'échantillonnage.

**[0025]** Dans une réalisation où le signal sonore est spatialisé sur au moins des premier et deuxième haut-parleurs virtuels, associés respectivement à un premier et un deuxième canal, on applique respectivement des première et deuxième fonctions de transfert avec effet de salle sur ces premier et deuxième canaux, comme expliqué ci-avant en introduction (par exemple en adaptant des signaux sur des canaux ambiophoniques pour passer à une restitution binaurale ou transaurale). En particulier, dans le cas où l'une des première et deuxième fonctions de transfert applique un effet de trajet acoustique ipsi-latéral, tandis que l'autre des première et deuxième fonctions de transfert applique un effet de trajet acoustique contra-latéral, on peut prévoir une élimination des composantes spectrales du signal sonore au-delà d'une fréquence d'écrantage donnée. Cette fréquence « d'écrantage » s'explique par le fait que pour un trajet contra-latéral entre un haut-parleur virtuel et une oreille considérée de l'auditeur, la tête de l'auditeur masque le trajet acoustique et absorbe les tonalités les plus aiguës de l'onde acoustiques (donc élimine les composantes spectrales associées aux fréquences les plus élevées de l'onde acoustique). Ainsi, la fréquence seuil précitée, pour la fonction de transfert appliquant un effet de trajet contra-latéral, peut être choisie comme un minimum parmi une fréquence seuil prédéterminée (par exemple choisie entre 5 et 15kHz) et cette fréquence d'écrantage. Cette réalisation est avantageuse pour être appliquée déjà pour le premier bloc d'échantillons. En revanche, elle n'exclut pas la possibilité d'augmenter à nouveau la fréquence seuil pour le bloc suivant pour simuler une première réflexion sur un mur situé en regard de l'oreille considérée, cette première réflexion étant reçue à cette oreille par un trajet ipsi-latéral.

**[0026]** On comprendra en tout état de cause que la fréquence de coupure peut être choisie commune à tous les signaux, dans une réalisation possible, après un instant donné qui correspond par exemple à la présence du champ diffus.

**[0027]** Ainsi, la réalisation décrite dans le document FR1357299 introduit plus haut peut être avantageuse dans le contexte de l'invention, et en particulier si chaque fonction de transfert appliquée à un signal comporte :

- une fonction de transfert spécifique à ce signal, additionnée à
- une fonction de transfert globale, commune à tous les signaux, et représentative d'une présence de champ diffus,

alors l'instant donné précité peut être commun à l'ensemble des signaux et correspondre par exemple à un début de présence de champ sonore diffus.

**[0028]** Dans une réalisation où les signaux comportent des blocs successifs d'échantillons, de mêmes tailles entre signaux, on prévoit au moins un instant donné pour la limitation de la prise en compte des composantes fréquentielles jusqu'à une fréquence de coupure, cet instant donné étant situé temporellement au début d'un bloc différent d'un premier

bloc dans une succession de blocs. Cet instant donné intervient donc après une propagation directe, et au moment de réflexions sonores ou de présence de champ diffus.

**[0029]** Cette réalisation sera détaillée plus loin en référence à la figure 5, illustrant aussi, dans un exemple de réalisation, un possible algorithme d'un programme informatique qu'exécuterait un processeur d'un module de spatialisation opérant le procédé au sens de l'invention. A ce titre, la présente invention vise aussi, de façon générale, un programme informatique comportant des instructions pour la mise en oeuvre du procédé ci-avant, lorsqu'elles sont exécutées par un processeur.

**[0030]** Cette divulgation présente également un exemple de mise en oeuvre d'un module de spatialisation sonore, comportant des moyens de calcul pour appliquer au moins une fonction de transfert à effet de salle à au moins un signal sonore d'entrée, ladite application revenant à multiplier, dans le domaine spectral, des composantes spectrales du signal sonore par les composantes spectrales d'un filtre correspondant à ladite fonction de transfert, chaque composante spectrale du filtre comportant une évolution temporelle dans une représentation temps-fréquence. En particulier, ces moyens de calcul sont configurés pour ignorer lesdites composantes spectrales du filtre pour lesdites multiplications de composantes, au-delà d'une fréquence seuil et après au moins un instant donné dans ladite représentation temps-fréquence.

**[0031]** Le module de spatialisation sonore recevant une pluralité de signaux d'entrée, délivre au moins deux signaux de sortie, les moyens de calcul étant configurés pour appliquer une fonction de transfert à effet de salle, à chaque signal d'entrée, chacun desdits signaux de sortie étant donné par application de la formule suivante:

$$O^k = \sum_{l=1}^L (I(l) *_{[0; \dots; f^k(l)]} A^k(l)) + \sum_{m=1}^M \left( z^{-iDDm} \cdot \sum_{l=1}^L \left( G(I(l)) \cdot \frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0; \dots; f^k(m)]} B_{mean}^k(m)$$

- $O^k$  étant un signal de sortie, et k étant l'indice relatif à un signal de sortie,
- $l \in [1; L]$ , étant l'indice relatif à un signal d'entrée parmi lesdits signaux d'entrée, L étant le nombre de signaux d'entrée, et  $I(l)$  étant un signal d'entrée parmi lesdits signaux d'entrée,
- $A^k(l)$  étant une fonction de transfert avec effet de salle spécifique à un signal d'entrée,
- $B_{mean}^k(m)$  étant une fonction de transfert globale, avec effet de salle, commune aux signaux d'entrée,
- $W^k(l)$  étant un poids de pondération choisi, et  $G(I(l))$ , un gain de compensation d'énergie prédéterminé,
- $z^{-iDDm}$  étant une application de délai, compté en nombre de blocs d'échantillons, correspondant à un écart temporel entre une émission sonore dans une salle correspondant à l'effet de salle, et un début de présence de champ diffus dans cette salle, l'indice m correspondant à un nombre de blocs d'échantillons de durée correspondant à ce délai, M étant le nombre total de blocs que dure une fonction de transfert dans une représentation temps-fréquence,
- le signe « . » désignant la multiplication,
- le signe «  $*_{[0; \dots; f^k(l)]}$  » désignant l'opérateur de convolution sur un nombre de fréquences limité et allant d'une plus basse fréquence à une fréquence maximale  $f^k(l)$  qui est fonction au moins du signal d'entrée d'indice l, et
- le signe «  $*_{[0; \dots; f^k(m)]}$  » désignant l'opérateur de convolution sur un nombre de fréquences limité et allant d'une plus basse fréquence à une fréquence  $f^k(m)$  qui est fonction du bloc d'échantillons d'indice m.

**[0032]** Ce module peut être intégré dans un dispositif de décodage en compression, ou plus généralement dans un système de restitution.

**[0033]** On a représenté sur la figure 6 un tel module de spatialisation SPAT, ainsi qu'un dispositif de décodage DECOD qui reçoit d'un réseau RES, dans l'exemple représenté, des signaux codés en compression  $I(l)$  (avec  $l = 1, \dots, L$ ) et les décode avant restitution en transmettant les signaux décodés  $I(l)$  (avec  $l = 1, \dots, L$ ) au module de spatialisation. Ce dernier comprend dans l'exemple représenté une interface d'entrée IN pour recevoir les signaux décodés, ainsi que des moyens de calculs tels qu'un processeur PROC et une mémoire de travail MEM coopérant avec les interfaces IN/OUT pour spatialiser les signaux  $I(l)$  et délivrer par l'interface de sortie OUT uniquement deux signaux  $O^d$  et  $O^g$  destinés à alimenter les oreillettes respectives d'un casque CAS.

**[0034]** D'autres caractéristiques et avantages de l'invention apparaîtront à l'examen de la description détaillée ci-après, et des dessins annexés sur lesquels :

- la figure 1 illustre une réalisation générale du procédé selon l'invention ;

- la figure 2 illustre un exemple d'application du procédé selon un mode de réalisation où les fonctions de transfert sont sous la forme d'une combinaison de deux fonctions de transfert dont l'une est appliquée avec un retard sur le signal à traiter ;
- la figure 3 représente un exemple de représentation temps-fréquence d'une fonction de transfert avec des fréquences de coupure (ou « fréquences seuils » précitées) variables notamment en fonction du temps ;
- la figure 4 illustre un organigramme correspondant à un algorithme général possible du programme informatique au sens de l'invention,
- la figure 5 représente une réalisation particulière issue du mode représenté sur la figure 2, mais sur plus de deux blocs temporels successifs, avec une évolution de la fonction de transfert  $B_{max}^k(m)$  représentant le champ diffus, en fonction des blocs  $m$  ;
- la figure 6 illustre un exemple de module de spatialisation au sens de l'invention ;
- la figure 7 illustre schématiquement les haut-parleurs virtuels et l'effet de salle pour appliquer une fonction de transfert appropriée, avec limitation des composantes fréquentielles de cette fonction de transfert jusqu'à une fréquence de coupure appropriée.

**[0035]** Avant de décrire la figure 1 et le principe général de l'invention, on se réfère à la figure 7 pour expliquer les phénomènes physiques qui sous-tendent la présente invention.

**[0036]** Une pluralité de haut-parleurs virtuels entourent, dans l'exemple représenté, la tête TE d'un auditeur. Chacun des haut-parleurs virtuels HPV est alimenté initialement par un signal  $l(l)$  avec  $l \in [1;L]$  par exemple précédemment décodé comme indiqué ci-avant en référence à la figure 6. La disposition des haut-parleurs virtuels peut concerner une représentation multicanale ou aussi ambiophonique des signaux  $l(l)$  à traiter pour les restituer ensemble de façon spatialisée avec un effet de salle sur un casque à oreillettes CAS (figure 6). A cet effet, on applique habituellement à chaque signal une fonction de transfert à effet de salle pour chaque signal d'oreillette à délivrer  $O^k$ , avec  $k = d$  (pour droite),  $g$  (pour gauche). Ainsi, en référence à la figure 7, on considère pour chaque haut-parleur virtuel HPV le trajet acoustique (ipsi-latéral TIL dans l'exemple représenté) du haut-parleur HPV à l'oreille gauche OG, ainsi que le trajet acoustique (contra-latéral TCL dans l'exemple représenté) du haut-parleur HPV à l'oreille droite OD, ainsi que des réflexions sur les murs MUR (trajet RIL), et finalement un champ diffus après plusieurs réflexions. A chaque réflexion, il est considéré que l'onde acoustique s'atténue dans les fréquences les plus élevées.

**[0037]** Ainsi, en référence à la figure 3 relative à une représentation temps-fréquence d'une fonction de transfert adaptée pour le haut-parleur virtuel HPV représenté sur la figure 7, il apparaît déjà que la tête de l'auditeur masque naturellement le trajet contra-latéral et les fréquences les plus élevées à considérer pour la fonction de transfert propre à l'oreille droite OD sont plus basses que celles à considérer pour la fonction de transfert propre à l'oreille gauche OG (qui est en regard du haut-parleur virtuel HPV selon un trajet ipsi-latéral). Ainsi, en considérant un premier bloc de temps de 0 à  $N-1$ , noté  $m=0$ , la fréquence maximale  $F_c^d(0)$  d'un filtre représentant la fonction de transfert propre à l'oreille droite peut être plus basse que la fréquence maximale  $F_c^g(0)$  d'un filtre représentant la fonction de transfert propre à l'oreille gauche. Un concepteur de tel filtre peut ainsi limiter les composantes de son filtre pour l'oreille droite jusqu'à la fréquence de coupure  $F_c^d(0)$  (correspondant à une fréquence d'écrantage de la tête) quand bien même le signal à traiter  $l(l)$  peut avoir des composantes spectrales supérieures et jusqu'à la fréquence  $F_c^g(0)$  au moins.

**[0038]** Ensuite, après des réflexions, l'onde acoustique tend à s'atténuer dans les fréquences élevées, ce qui est bien respecté par la représentation temps-fréquence de la fonction de transfert pour l'oreille gauche, comme pour l'oreille droite, pour les instants  $N$  à  $2N-1$ , correspondant au bloc suivant noté  $m=1$ . Ainsi, un concepteur de filtres représentant ces fonctions de transfert peut prévoir de limiter les composantes des filtres pour l'oreille droite jusqu'à la fréquence de coupure  $F_c^d(1)$  et pour l'oreille gauche jusqu'à la fréquence de coupure  $F_c^g(1)$ . Dans un mode de réalisation illustré en particulier sur la figure 5, on peut considérer que dans le bloc  $m=1$ , la fonction de transfert caractérise typiquement un champ diffus pour l'oreille droite comme pour l'oreille gauche, et, ainsi, il peut être établi (possiblement mais non limitativement) que  $F_c^d(1)=F_c^g(1)$ .

**[0039]** Ensuite, en présence de champ diffus avec atténuation globale du son (« fade out »), l'onde acoustique tend à s'atténuer davantage dans les fréquences élevées, ce qui est bien respecté encore par la représentation temps-fréquence de la fonction de transfert pour l'oreille gauche, comme pour l'oreille droite, sur la figure 3, pour les instants  $2N$  à  $3N-1$ , correspondant au bloc noté  $m=2$ . Ainsi, un concepteur de filtres représentant ces fonctions de transfert peut prévoir de limiter les composantes des filtres pour l'oreille droite jusqu'à la fréquence de coupure  $F_c^d(2)$  et pour l'oreille gauche jusqu'à la fréquence de coupure  $F_c^g(2)$ .

**[0040]** On relèvera que des blocs plus courts permettraient de faire varier plus finement la plus haute fréquence à considérer par exemple pour tenir compte d'une première réflexion RIL pour laquelle la fréquence la plus haute augmente pour l'oreille droite (traits pointillés autour de  $F_c^d(0)$  dans la figure 3) dans les premiers instants du bloc  $m=0$ .

**[0041]** Ainsi, on retiendra qu'il est possible de ne pas prendre en compte toutes les composantes spectrales d'un filtre représentant une fonction de transfert, en particulier au-delà d'une fréquence de coupure  $F_c$ . Dès lors, il est avantageux

de traiter l'application de la fonction de transfert, dans le domaine spectral. En effet, la convolution d'un signal  $I(f)$  par une fonction de transfert devient, dans le domaine spectral, une multiplication des composantes spectrales du signal  $I(f)$  par les composantes spectrales du filtre représentant la fonction de transfert dans le domaine spectral, et, en particulier, cette multiplication peut être opérée jusqu'à une fréquence de coupure seulement, qui est fonction d'un bloc donné, par

exemple, et du signal à traiter.

**[0042]** Ainsi, en référence à la figure 1,  $L$  signaux d'entrée  $I(1), I(2), \dots, I(L)$  sont transformés dans le domaine fréquentiel, respectivement aux étapes TF11, TF12, ..., TF1L. En variante, de tels signaux d'entrée peuvent déjà être disponibles sous forme fréquentielle (par exemple auprès du décodeur).

**[0043]** A l'étape BA11, une réponse impulsionnelle complète de spatialisation (typiquement de type BRIR pour « Binaural Room Impulse Response ») sous forme temporelle correspondant au signal  $I(1)$  du canal 1 est mise en mémoire. A l'étape TFA11, on transforme cette réponse impulsionnelle sous forme fréquentielle pour obtenir un filtre correspondant dans le domaine spectral. Dans un mode de réalisation avantageux, le filtre est stocké sous sa forme spectrale pour éviter de répéter le calcul de la transformée. On multiplie ensuite ce filtre au signal d'entrée sous forme fréquentielle du canal 1 (ce qui équivaut à une convolution dans le domaine temporel). On dispose donc du signal spatialisé pour le signal  $I(1)$  du canal 1.

**[0044]** On effectue les mêmes opérations pour les  $L-1$  autres canaux. On dispose ainsi au total de  $L$  canaux spatialisés. Ces canaux sont ensuite sommés pour obtenir un unique signal de sortie représentant les  $L$  canaux, et on repasse dans le domaine temporel à l'étape ITF11, pour délivrer l'un des signaux  $O^k$  (avec  $k=d,g$ ) alimentant une oreillette. Un traitement analogue est réalisé pour l'autre oreillette. Dans un mode de réalisation décrit en détails plus loin en référence aux figures 2 et 5, les  $L$  canaux spatialisés ne sont pas accessibles indépendamment avant sommation : l'unique signal de sortie est construit en sommant au fur et à mesure chaque canal spatialisé avec le signal de sortie précédent.

**[0045]** On effectue ces opérations pour chaque signal de sortie  $O^k$  à construire. Typiquement, s'il s'agit d'un restitution binaurale, ces étapes sont effectuées deux fois, une fois pour le signal de sortie destiné à alimenter une oreillette gauche d'un casque d'écoute et une fois pour le signal de sortie destiné à alimenter l'oreillette droite du casque. Ainsi, on obtient finalement deux signaux spatialisés  $O^d$  et  $O^g$  correspondant chacun à une oreille.

**[0046]** Les  $L$  signaux d'entrée peuvent correspondre typiquement aux  $L$  canaux d'un contenu audio multicanal censé alimenter des haut-parleurs (« virtuels »). Les  $L$  signaux d'entrée peuvent par exemple correspondre aux  $L$  signaux ambiophoniques d'un contenu audio en représentation ambiophonique.

**[0047]** En référence maintenant à la figure 2 illustrant une mise en oeuvre au sens de l'invention, on reprend le principe d'une spatialisation de  $L$  canaux comme présenté sur la figure 1. Néanmoins, la présentation de la figure 2 est simplifiée en ce que les  $L$  signaux d'entrée sont réunis en une seule voie  $I(f)$ . Ainsi,  $L$  signaux d'entrées  $I(1), I(2), \dots, I(L)$  sont transformés dans le domaine fréquentiel à l'étape S21. Comme indiqué précédemment, de tels signaux d'entrée peuvent en variante être déjà disponibles sous forme fréquentielle. A l'étape S22, une réponse impulsionnelle  $A^k(f)$  de spatialisation (typiquement de type BRIR) correspondant au signal  $I(f)$  du canal  $f$  est transformée dans le domaine spectral pour obtenir un filtre fréquentiel. Cette réponse impulsionnelle  $A^k(f)$  est incomplète dans la représentation de la figure 2 car elle correspond à un premier bloc temporel d'échantillons  $m=0$ . Comme indiqué précédemment, cette réponse impulsionnelle peut être déjà disponible sous forme fréquentielle. On multiplie ensuite les composantes de ce filtre au signal spectral du canal correspondant  $I(f)$ . Cette multiplication est paramétrée (comme indiqué ci-dessous en référence à la figure 4) pour que certaines composantes fréquentielles soient ignorées, au sens de l'invention. Typiquement, les composantes fréquentielles les plus élevées seront ignorées pour limiter la complexité des calculs. Sur les figures 2 et 5, on note alors la multiplication des composantes limitée à une fréquence de coupure par le signe :  $\chi$

**[0048]** Une fréquence de coupure  $f_{cA(f)}$  à partir de laquelle les composantes fréquentielles sont ignorées est définie (par exemple la fréquence maximale représentée dans le signal du canal  $I(f)$ , ou la moitié de sa fréquence d'échantillonnage). En outre, cette fréquence de coupure est spécifique à chaque filtre et pour chaque bloc (elle diminue par exemple pour les blocs  $m=1, m=2$ ). Comme les filtres sont ici spécifiques à chaque signal d'entrée et à chaque oreille, une fréquence de coupure est spécifique à un signal d'entrée, à une oreille (donc un signal de sortie) et à un bloc temporel.

**[0049]** On dispose alors du signal spatialisé pour le canal  $f$  pour ce premier bloc temporel. On effectue ces opérations pour tous les  $L$  canaux :  $f = 1, \dots, L$ . On dispose ainsi de  $L$  canaux spatialisés. Ces canaux sont ensuite sommés à l'étape S23 pour obtenir un unique signal représentant les  $L$  canaux sur le premier bloc temporel.

**[0050]** En pratique, la sommation est réalisée de façon particulière, car elle tient compte d'un retard sur les canaux pour caractériser les réverbérations (réflexions et champ diffus), comme détaillé ci-après. En effet, dans un mode de réalisation, les  $L$  canaux spatialisés ne sont pas accessibles indépendamment avant sommation : l'unique signal de sortie est construit en sommant au fur et à mesure chaque canal spatialisé avec le signal de sortie précédent. A cet effet, à l'étape DBD, on retarde les signaux d'entrée  $I(f)$  d'un certain délai donné par  $z^{-iDD.m}$  propre à chaque bloc  $m=1, \dots, M$ . On remarque que pour le premier bloc, le retard  $m$  est nul. Dans le cas d'une représentation fréquentielle, ce délai correspond généralement à la taille d'une trame de signal traité pour le premier bloc, et s'interprète comme le fait de prendre le bloc d'entrée précédent dans sa forme fréquentielle.

**[0051]** A l'étape S24, une réponse impulsionnelle  $B_m^k(f)$  incomplète de spatialisation (typiquement de type BRIR)

correspondant au signal  $I(l)$  du canal  $l$  est transformée dans le domaine spectral pour obtenir un filtre fréquentiel. Cette réponse impulsionnelle  $B_m^k(l)$  est incomplète car elle correspond à un second bloc temporel d'échantillons (puis à un troisième bloc et ainsi de suite, pour  $m=1, \dots, M$ ). Comme indiqué précédemment, cette réponse impulsionnelle peut en variante être déjà disponible sous forme fréquentielle. En appliquant le principe décrit dans le document FR1357299, il est possible de réduire la complexité du traitement en posant  $B_m^k(1) = \dots = B_m^k(l) = \dots = B_m^k(L) = B_{\text{mean}}^k(m)$  et de ne faire dépendre finalement cette fonction de transfert que du bloc  $m$  considéré (champ diffus principal, ou champ diffus secondaire avec atténuation « fade out ») et de l'oreille  $k$ . De même, le champ diffus ne dépend pas des canaux et il est possible de fixer la fréquence de coupure  $f_c$  comme étant identique pour chaque canal (mais qui peut diminuer encore d'un bloc au suivant, comme on l'a vu précédemment en référence à la figure 3). Cette réalisation est présentée sur la figure 5.

**[0052]** En référence à nouveau à la figure 2, on multiplie ensuite ce filtre  $B_m^k(l)$  au signal  $I(l)$  du canal  $l$ . Les fréquences de coupure sont différentes pour ce second bloc temporel. Comme présenté en référence à la figure 3, des mesures montrent que les hautes fréquences sont plus atténuées dans les blocs temporels éloignés (correspondants à des sons diffus et aux réverbérations multiples). Les fréquences de coupure pour ces blocs éloignés peuvent donc être plus faibles que pour les premiers blocs. Or, plus la fréquence de coupure est faible, plus le nombre d'opérations est limité. Ainsi, la complexité des calculs est avantageusement réduite.

**[0053]** On effectue les mêmes opérations pour les  $L$  canaux et on répète les opérations de multiplication du filtre sur les signaux spectraux progressivement retardés en sommant les contributions à l'étape S25 réitérée pour chaque retard  $m$  jusqu'à obtenir un unique signal représentant les  $L$  canaux sur l'ensemble  $M$  des blocs temporels  $m$  considérés. L'unique signal de sortie est construit en sommant au fur et à mesure chaque canal spatialisé avec le signal de sortie précédent comme on le verra maintenant en référence à la figure 4.

**[0054]** Enfin, on repasse dans le domaine temporel à l'étape S26 pour obtenir un signal de sortie destiné à alimenter l'une des oreillettes du casque.

**[0055]** En référence à la figure 4, on décrit maintenant un procédé de spatialisation pour un bloc temporel donné (par exemple pour le bloc représentant le champ sonore direct à valeurs dans l'intervalle temporel  $[0 ; N-1]$ ) et pour un signal correspondant, par exemple, à l'oreille droite. Bien sûr, le même procédé est appliqué pour le signal correspondant à l'oreille gauche. La distinction entre les deux oreilles est introduite par l'application de filtres spécifiques à chacune de ces oreilles.

**[0056]** A l'étape S40, le signal de sortie  $S$  est initialisé à 0. Ce signal de sortie s'exprime dans le domaine fréquentiel. Il possède une taille limitée d'une longueur supérieure à la fréquence de coupure  $f_c(l)$ . Par exemple, ce signal est défini sur  $[0; f_s(l)/2]$ ,  $f_s(l)$  étant la fréquence d'échantillonnage de ce signal  $I(l)$ . Une première variable de comptage  $l$  est également initialisée à 1. Cette première variable de comptage identifie un des signaux de canal  $I(1), I(2), \dots, I(l), \dots, I(L)$  sur le bloc temporel  $[0 ; N-1]$  pour l'oreille droite. A l'étape S41, une deuxième variable de comptage  $j$  est initialisée à 0. Cette seconde variable de comptage identifie une composante fréquentielle d'un signal  $I(l)$  sur le bloc temporel  $[0 ; N-1]$  pour l'oreille droite.

**[0057]** A l'étape S42, le coefficient  $c_{\text{BRIR}}(j ; l)$  est mis en mémoire. Ce coefficient correspond à la composante fréquentielle  $j$  du filtre  $\text{BRIR}(l)$  sur le bloc temporel  $[0 ; N-1]$  pour l'oreille droite. De même, le coefficient  $c_l(j ; l)$  est mis en mémoire. Ce coefficient correspond à la composante fréquentielle  $j$  du signal  $I(l)$  sur le bloc temporel  $[0 ; N-1]$  pour l'oreille droite. Ainsi, les coefficients  $c_{\text{BRIR}}(j ; l)$  et  $c_l(j ; l)$  correspondent à la même composante fréquentielle (identifiée par la variable  $j$ ) et pourront ainsi être ultérieurement multipliés terme à terme (étape S44).

**[0058]** Au test T47, on vérifie que la fréquence correspondant à la variable  $j$  est inférieure (par exemple strictement) à la fréquence de coupure  $f_c(l)$ . Cette fréquence de coupure correspond à la fréquence de coupure du signal  $I(l)$  pour le bloc temporel  $[0 ; N-1]$  pour l'oreille droite. Si la fréquence  $j$  est inférieure à la fréquence de coupure  $f_c(l)$ , on passe à l'étape S44.

**[0059]** A l'étape S44, on calcul une valeur  $\text{MULT}(j)$  correspondant à la multiplication des coefficients  $c_{\text{BRIR}}(j ; l)$  et  $c_l(j ; l)$ . Ces coefficients sont bien multipliés terme à terme car ils correspondent à la même composante fréquentielle  $j$  (pour un même canal, sur un même bloc et pour une même oreille).

**[0060]** A l'étape S45, on incrémente cette valeur  $\text{MULT}(j)$  au signal  $S$  à la position de la fréquence  $j$ .

**[0061]** Ainsi, on procède à une construction pas à pas d'un signal  $S$  qui comprend (à la fin de la boucle de longueur  $f_c(l)$ ) l'ensemble des composantes fréquentielles jusqu'à la fréquence de coupure  $f_c(t)$  (pour ce signal  $I(l)$  sur le bloc  $[0 ; N-1]$  et pour une oreille droite). Comme au début de la boucle de la figure 4, on a déjà toutes les composantes initialisées à 0, à la fin de la boucle, on a finalement rempli un buffer (initialement nul) jusqu'à la fréquence de coupure pour construire successivement le signal  $S$ . Ainsi, chaque multiplication  $\text{MULT}(j)$  de coefficients est ajoutée pas à pas au signal  $S$  en construction.

**[0062]** A l'étape S46, on incrémente la variable  $j$  et on reprend à l'étape S42. Si la variable  $j$  est supérieure (par exemple ou égale) à la fréquence de coupure  $f_c(l)$ , on passe au test T48. Ainsi, on a rempli le signal  $S$  sur l'intervalle  $[0 ; f_c(l)]$ .

**[0063]** Comme précisé ci-avant, ce signal peut être défini sur un intervalle plus grand que  $[0 ; f_c(l)]$  (par exemple  $[0 ; f_s(l)/2]$ ). De plus, ce signal avait été initialisé à 0 sur la totalité de son intervalle de définition. Dès lors, il est nul sur le reste de l'intervalle qui n'a pas été rempli (par exemple  $[f_c(l) ; f_s(l)/2]$ ). La complexité est donc ici améliorée car des

étapes de remplissage du signal S n'ont pas été effectuées, ce qui réduit le nombre de calculs nécessaires.

**[0064]** Au test T48, on vérifie que la variable de comptage / correspondant au signal  $I(l)$  du canal / est inférieure (par exemple strictement) au nombre L de canaux. Si la variable / est inférieure ou égale à L, on incrémente la variable / à l'étape S49 et on reprend le procédé à l'étape S41. Si la variable / est supérieure à L, le signal S correspondant au signal spatialisé pour le bloc temporel [0 ; N-1] pour l'oreille droite est disponible à l'étape S50.

**[0065]** Ce signal S correspondant au bloc temporel [0 ; N-1] est ensuite sommé aux autres signaux générés de manière similaire pour d'autres blocs temporels [N ; 2N-1], [2N ; 3N-1], etc., (et pour lesquels un retard approprié a été appliqué conformément à l'étape DBD ci-avant de la figure 2 par exemple).

**[0066]** Typiquement, pour construire le bloc [N ; 2N-1], on applique dans le domaine fréquentiel un filtre correspondant à une fonction de transfert commune à tous les signaux d'entrée  $I(l)$ , représentant le champ diffus, avec une fréquence de coupure  $f_c$  dans la multiplication des composantes spectrales qui correspond au minimum entre :

- une fréquence maximale de champ diffus  $F_c$  (diffus) comme illustrée sur la figure 3 décrite ci-avant (choisie par exemple entre 10 à 15 kHz pour le bloc  $m=1$  et entre 5 à 10 kHz pour le bloc  $m=2$ ), et
- la fréquence maximale  $f_{max}$  représentée dans chaque signal d'entrée (par exemple sa fréquence d'échantillonnage ou la fréquence maximale dont la composante spectrale n'est pas nulle, cette valeur étant habituellement donnée par un décodeur en compression).

**[0067]** Il est à noter que la multiplication fréquentielle en s'arrêtant à une fréquence de coupure donnée (ce qui revient mathématiquement à multiplier par 0 au-delà) n'est pas triviale pour l'homme de l'art. En effet, dans un contexte de filtrage d'un signal audio, ce type de filtre passe-bas très violent apporte généralement des artefacts audibles (dits « d'aliasing »), dus à des phénomènes d'écho ou de pré-écho provenant du repliement temporel généré par la convolution circulaire, qu'il est généralement désirable d'éviter. Or, dans le contexte de l'invention, ce filtre passe-bas n'est pas appliqué au signal audio, mais au filtre BRIR (qui est lui-même convolué au signal audio) qui est déjà composé de multiples réflexions ; les artefacts produits seront donc, au pire, perçus comme des réflexions supplémentaires du filtre BRIR d'origine, et en pratique rarement perceptibles. Il est toutefois possible d'atténuer néanmoins ces artefacts en modifiant légèrement les fréquences du filtre précédant la fréquence de coupure (par exemple par atténuation douce par application d'une demi-fenêtre de Hanning (type fade out)).

**[0068]** De manière générale, en référence à la figure 4, on remarquera que deux opérations sont effectuées dans une même instance de boucle (typiquement un coup d'horloge) : la multiplication MULT(k) et son ajout au signal S de sortie. Ceci permet notamment d'implémenter ce procédé sur des processeurs qui ont la possibilité d'effectuer plusieurs opérations sur une même instance de boucle (typiquement un coup d'horloge) et de réduire ainsi le temps nécessaire aux calculs.

**[0069]** On a illustré sur la figure 5 une forme algorithmique complète du traitement, conforme à la formule donnant un signal de sortie  $O^k$  présentée ci-avant :

$$O^k = \sum_{l=1}^L (I(l) *_{[0; \dots; f^k(l)]} A^k(l)) + \sum_{m=1}^M \left( z^{-iDDm} \cdot \sum_{l=1}^L \left( G(I(l)) \cdot \frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0; \dots; f^k(m)]} B_{mean}^k(m)$$

**[0070]** Comme indiqué ci-avant les pondérations  $W^k(l)$  et les gains  $G(I(l))$  peuvent être fixés à 1. On n'a pas représenté sur la figure 5 les gains  $G(I(l))$  car il convient de lire cette figure comme une intégration des gains aux poids  $1/W^k(l)$ . D'ailleurs, à la conception des filtres, ces deux paramètres sont déterminés, fixés et multipliés l'un à l'autre une fois pour toutes.

## Revendications

1. Procédé de spatialisation sonore, comportant l'application d'au moins une fonction de transfert à effet de salle à au moins un signal sonore, ladite application revenant à multiplier, dans le domaine spectral, des composantes spectrales du signal sonore par les composantes spectrales d'un filtre correspondant à ladite fonction de transfert, chaque composante spectrale du filtre comportant une évolution temporelle dans une représentation temps-fréquence, procédé dans lequel lesdites composantes spectrales du filtre sont ignorées, pour lesdites multiplications de com-

posantes, au-delà d'une fréquence seuil et après au moins un instant donné dans ladite représentation temps-fréquence et dans lequel, pour une mise en oeuvre par un module de spatialisation sonore recevant une pluralité de signaux d'entrée et délivrant au moins deux signaux de sortie, pour délivrer chaque signal de sortie, on applique une fonction de transfert à effet de salle, à chaque signal d'entrée, le procédé étant **caractérisé en ce que** chacun desdits signaux de sortie est donné par application de la formule suivante :

$$O^k = \sum_{l=1}^L (I(l) *_{[0;\dots;f^k(l)]} A^k(l)) + \sum_{m=1}^M \left( z^{-iDDm} \cdot \sum_{l=1}^L \left( G(I(l)) \cdot \frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0;\dots;f^k(m)]} B_{mean}^k(m)$$

- $O^k$  étant un signal de sortie, et  $k$  étant l'indice relatif à un signal de sortie,
- $l \in [1; L]$ , étant l'indice relatif à un signal d'entrée parmi lesdits signaux d'entrée,  $L$  étant le nombre de signaux d'entrée, et  $I(l)$  étant un signal d'entrée parmi lesdits signaux d'entrée,
- $A^k(l)$  étant une fonction de transfert avec effet de salle spécifique à un signal d'entrée,
- $B_{mean}^k(m)$  étant une fonction de transfert globale, avec effet de salle, commune aux signaux d'entrée,
- $W^k(l)$  étant un poids de pondération choisi, et  $C(I(l))$ , un gain de compensation d'énergie prédéterminé,
- $z^{-iDDm}$  étant une application de délai, compté en nombre de blocs d'échantillons, correspondant à un écart temporel entre une émission sonore dans une salle correspondant à l'effet de salle, et un début de présence de champ diffus dans cette salle, l'indice  $m$  correspondant à un nombre de blocs d'échantillons de durée correspondant à ce délai,  $M$  étant le nombre total de blocs que dure une fonction de transfert dans une représentation temps-fréquence,
- le signe « . » désignant la multiplication,
- le signe «  $*_{[0;\dots;f^k(l)]}$  » désignant l'opérateur de convolution sur un nombre de fréquences limité et allant d'une plus basse fréquence à une fréquence maximale  $f^k(l)$  qui est fonction au moins du signal d'entrée d'indice  $l$ , et
- le signe «  $*_{[0;\dots;f^k(m)]}$  » désignant l'opérateur de convolution sur un nombre de fréquences limité et allant d'une plus basse fréquence à une fréquence  $f^k(m)$  qui est fonction du bloc d'échantillons d'indice  $m$ .

2. Procédé selon la revendication 1, dans lequel la fréquence seuil diminue en fonction du temps dans ladite représentation temps-fréquence.
3. Procédé selon l'une des revendications précédentes, dans lequel on obtient une information de composante spectrale de fréquence la plus élevée dans le signal sonore, et dans lequel ladite fréquence seuil est un minimum parmi une fréquence seuil prédéterminée et ladite fréquence la plus élevée.
4. Procédé selon la revendication 3, dans lequel le signal sonore est issu d'un décodeur en compression et l'information de composante spectrale de fréquence la plus élevée est fournie par le décodeur.
5. Procédé selon l'une des revendications 3 et 4, dans lequel le signal sonore est échantillonné à une fréquence d'échantillonnage donnée, ladite fréquence seuil étant choisie en fonction de ladite fréquence d'échantillonnage.
6. Procédé selon l'une des revendications précédentes, dans lequel le signal sonore est spatialisé sur au moins un premier et un deuxième haut-parleurs virtuels associés respectivement à un premier et un deuxième canal, et on applique respectivement des première et deuxième fonctions de transfert avec effet de salle sur lesdits premier et deuxième canaux,

l'une des première et deuxième fonctions de transfert appliquant un effet de trajet acoustique ipsi-latéral, et l'autre des première et deuxième fonctions de transfert appliquant un effet de trajet acoustique contra-latéral avec une élimination des composantes spectrales du signal sonore au-delà d'une fréquence d'écranage donnée, et dans lequel ladite fréquence seuil pour la fonction de transfert appliquant un effet de trajet contra-latéral est un minimum parmi une fréquence seuil prédéterminée et ladite fréquence d'écranage.

7. Procédé selon la revendication 1, dans lequel les signaux comportent des blocs successifs d'échantillons, de mêmes tailles entre signaux, et dans lequel ledit au moins un instant donné est situé temporellement au début d'un bloc distinct d'un premier bloc dans une succession de blocs.

8. Programme informatique, comportant des instructions pour la mise en oeuvre du procédé selon l'une des revendications précédentes, lorsqu'elles sont exécutées par un processeur.

5 **Patentansprüche**

1. Klangverräumlichungsverfahren, aufweisend das Anwenden wenigstens einer Übertragungsfunktion mit Raumeffekt auf wenigstens ein Klangsignal, wobei das Anwenden dem Multiplizieren, im Spektralbereich, der Spektralkomponenten des Klangsignals mit den Spektralkomponenten eines Filters entspricht, der der Übertragungsfunktion entspricht, wobei jede Spektralkomponente des Filters eine zeitliche Entwicklung in einer Zeit-Frequenz-Darstellung aufweist, wobei bei dem Verfahren die Spektralkomponenten des Filters für die Multiplikationen von Komponenten über einer Schwellenfrequenz und nach wenigstens einem bestimmten Zeitpunkt in der Zeit-Frequenz-Darstellung ignoriert werden, und wobei für eine Umsetzung durch ein Klangverräumlichungsmodul, das eine Mehrzahl von Eingangssignalen empfängt und wenigstens zwei Ausgangssignale ausgibt, zur Ausgabe jedes Ausgangssignals eine Übertragungsfunktion mit Raumeffekt auf jedes Eingangssignal angewandt wird, wobei das Verfahren **dadurch gekennzeichnet ist, dass** jedes der Ausgangssignale durch Anwendung der folgenden Formel geliefert wird:

$$O^k = \sum_{l=1}^L \left( I(l) *_{[0; \dots; f^k(l)]} A^k(l) \right) + \sum_{m=1}^M \left( z^{-iDDm} \cdot \sum_{l=1}^L \left( G(I(l)) \cdot \frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0; \dots; f^k(m)]} B_{mean}^k(m)$$

- wobei  $O^k$  ein Ausgangssignal und  $k$  der Index in Bezug auf ein Ausgangssignal ist,
- wobei  $l \in [1; L]$  der Index in Bezug auf ein Eingangssignal von den Eingangssignalen ist,  $L$  die Anzahl von Eingangssignalen ist und  $I(l)$  ein Eingangssignal von den Eingangssignalen ist,
- wobei  $A^k(l)$  eine Übertragungsfunktion mit Raumeffekt ist, die spezifisch für ein Eingangssignal ist,
- wobei  $B_{mean}^k$  eine übergreifende Übertragungsfunktion mit Raumeffekt ist, die für die Eingangssignale gemeinsam ist,
- wobei  $W^k(l)$  ein gewähltes Gewichtungsgewicht und  $G(I(l))$ , eine vorbestimmte Verstärkung zum Energieausgleich ist,
- wobei  $z^{-iDDm}$  eine Anwendung einer Verzögerung ist, die als Anzahl von Abtastwertblöcken gezählt wird, die einer Zeitabweichung zwischen einer Schallaussendung in einem Raum, die dem Raumeffekt entspricht, und einem Beginn des Vorhandenseins eines Diffusfelds in diesem Raum entspricht, wobei der Index  $m$  einer Anzahl von Abtastwertblöcken mit einer Dauer entspricht, die dieser Verzögerung entspricht, wobei  $M$  die Gesamtzahl von Blöcken ist, die eine Übertragungsfunktion in einer Zeit-Frequenz-Darstellung dauert,
- wobei das Zeichen "  $\cdot$  " die Multiplikation bezeichnet,
- wobei das Zeichen " $*_{[0; \dots; f^k(l)]}$ " den Faltungsoperator auf einer begrenzten Anzahl von Frequenzen bezeichnet und von einer tiefsten Frequenz zu einer Maximalfrequenz  $f^k(l)$  reicht, die wenigstens vom Eingangssignal mit dem Index  $l$  abhängig ist, und
- wobei das Zeichen " $*_{[0; \dots; f^k(m)]}$ " den Faltungsoperator auf einer begrenzten Anzahl von Frequenzen bezeichnet und von einer tiefsten Frequenz zu einer Frequenz  $f^k(m)$  reicht, die vom Abtastwertblock mit dem Index  $m$  abhängig ist.

2. Verfahren nach Anspruch 1, wobei die Schwellenfrequenz in der Zeit-Frequenz-Darstellung in Abhängigkeit von der Zeit abnimmt.

3. Verfahren nach einem der vorhergehenden Ansprüche, wobei eine Information über eine Spektralkomponente mit höchster Frequenz im Klangsignal erhalten wird und wobei die Schwellenfrequenz ein Minimum aus der vorbestimmten Schwellenfrequenz und der höchsten Frequenz ist.

4. Verfahren nach Anspruch 3, wobei das Klangsignal aus einem Kompressionsdecoder kommt und die Information über die Spektralkomponente mit höchster Frequenz vom Decoder bereitgestellt wird.
5. Verfahren nach einem der Ansprüche 3 und 4, wobei das Klangsignal bei einer bestimmten Frequenz abgetastet wird, wobei die Schwellenfrequenz je nach der Abtastfrequenz gewählt wird.
6. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Klangsignal über wenigstens einen ersten und einen zweiten virtuellen Lautsprecher verräumlicht wird, die einem ersten bzw. zweiten Kanal zugeordnet sind, und eine erste bzw. zweite Übertragungsfunktion mit Raumeffekt auf den ersten und zweiten Kanal angewandt wird,
- wobei die eine der ersten und zweiten Übertragungsfunktion einen Effekt einer ipsilateralen akustischen Strecke anwendet und die andere der ersten und zweiten Übertragungsfunktion einen Effekt einer kontralateralen akustischen Strecke mit einer Eliminierung der Spektralkomponenten des Klangsignals über einer bestimmten Abschirmfrequenz anwendet,
- und wobei die Schwellenfrequenz für die Übertragungsfunktion, die einen Effekt einer kontralateralen Strecke anwendet, ein Minimum aus einer vorbestimmten Schwellenfrequenz und der Abschirmfrequenz ist.
7. Verfahren nach Anspruch 1, wobei die Signale aufeinanderfolgende Abtastwertblöcke gleicher Größen zwischen Signalen aufweisen und wobei sich der wenigstens eine bestimmte Zeitpunkt zeitlich am Anfang eines Blocks befindet, der von einem ersten Block in einer Aufeinanderfolge von Blöcken gesondert ist.
8. Computerprogramm, umfassend Anweisungen für die Durchführung des Verfahrens nach einem der vorhergehenden Ansprüche, wenn sie von einem Prozessor ausgeführt werden.

### Claims

1. Method for achieving audio spatialization, comprising the application of at least one room-effect transfer function to at least one audio signal, said application amounting to multiplying, in the spectral domain, spectral components of the audio signal by spectral components of a filter corresponding to said transfer function, each spectral component of the filter comprising a variation as a function of time in a time-frequency representation, in which method said spectral components of the filter are ignored, in said multiplications of components, beyond a threshold frequency and after at least one given time in said time-frequency representation and in which method, in an implementation by a module for achieving audio spatialization that receives a plurality of input signals and that delivers at least two output signals, to deliver each output signal, a room-effect transfer function is applied to each input signal, the method being **characterized in that** each of said output signals is given by application of the following formula:

$$O^k = \sum_{l=1}^L \left( I(l) *_{[0; \dots; f^k(l)]} A^k(l) \right) + \sum_{m=1}^M \left( z^{-iDDm} \cdot \sum_{l=1}^L \left( G(I(l)) \cdot \frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0; \dots; f^k(m)]} B_{mean}^k(m)$$

- $O^k$  being an output signal, and  $k$  being the index relative to an output signal,
- $l \in [1; L]$ , being the index relative to one input signal among said input signals,  $L$  being the number of input signals, and  $I(l)$  being one input signal among said input signals,
- $A^k(l)$  being a transfer function with room effect specific to one input signal,
- $B_{mean}^k(m)$  being an overall transfer function, with room effect, common to the input signals,
- $W^k(l)$  being a selected weighting weight, and  $G(I(l))$  being a predetermined energy compensation gain,
- $z^{-iDDm}$  being an application of a delay, counted in number of blocks of samples, corresponding to a time

difference between an audio emission in a room corresponding to the room effect and an onset of presence of a diffuse field in this room, the index  $m$  corresponding to a number of blocks of samples of duration corresponding to this delay,  $M$  being the total number of blocks that a transfer function lasts in a time-frequency representation,

- the sign "." designating multiplication,

- the sign "[0;...;f<sup>k</sup>(l)]" designating convolution over a limited number of frequencies ranging from a lowest frequency to a maximum frequency  $f^k(l)$  that is dependent at least on the input signal of index  $l$ , and

- the sign "[0;...;f<sup>k</sup>(m)]" designating convolution over a limited number of frequencies ranging from a lowest frequency to a frequency  $f^k(m)$  that is dependent on the block of samples of index  $m$ .

2. Method according to Claim 1, wherein the threshold frequency decreases as a function of time in said time-frequency representation.

3. Method according to one of the preceding claims, wherein information is obtained on the spectral component of highest frequency in the audio signal, and wherein said threshold frequency is a minimum among a predetermined threshold frequency and said highest frequency.

4. Method according to Claim 3, wherein the audio signal is generated by a compression decoder and the information on the spectral component of highest frequency is delivered by the decoder.

5. Method according to one of Claims 3 and 4, wherein the audio signal is sampled at a given sampling frequency, said threshold frequency being chosen depending on said sampling frequency.

6. Method according to one of the preceding claims, wherein the audio signal is spatialized via at least a first virtual loudspeaker and a second virtual loudspeaker that are associated with a first channel and a second channel, respectively, and first and second transfer functions with room effect are applied to said first and second channels, respectively,

one of the first and second transfer functions applying an ipsi-lateral acoustic pathway effect, and the other of the first and second transfer functions applying a contra-lateral acoustic pathway effect with removal of spectral components above a given screening frequency from the audio signal,

and wherein said threshold frequency for the transfer function applying a contra-lateral pathway effect is a minimum among a predetermined threshold frequency and said screening frequency.

7. Method according to Claim 1, wherein the signals comprise successive blocks of samples, of same sizes between signals, and wherein said at least one given time is located temporally at the start of a block other than a first block in a succession of blocks.

8. Computer program, containing instructions for implementing the method according to one of the preceding claims, when they are executed by a processor.

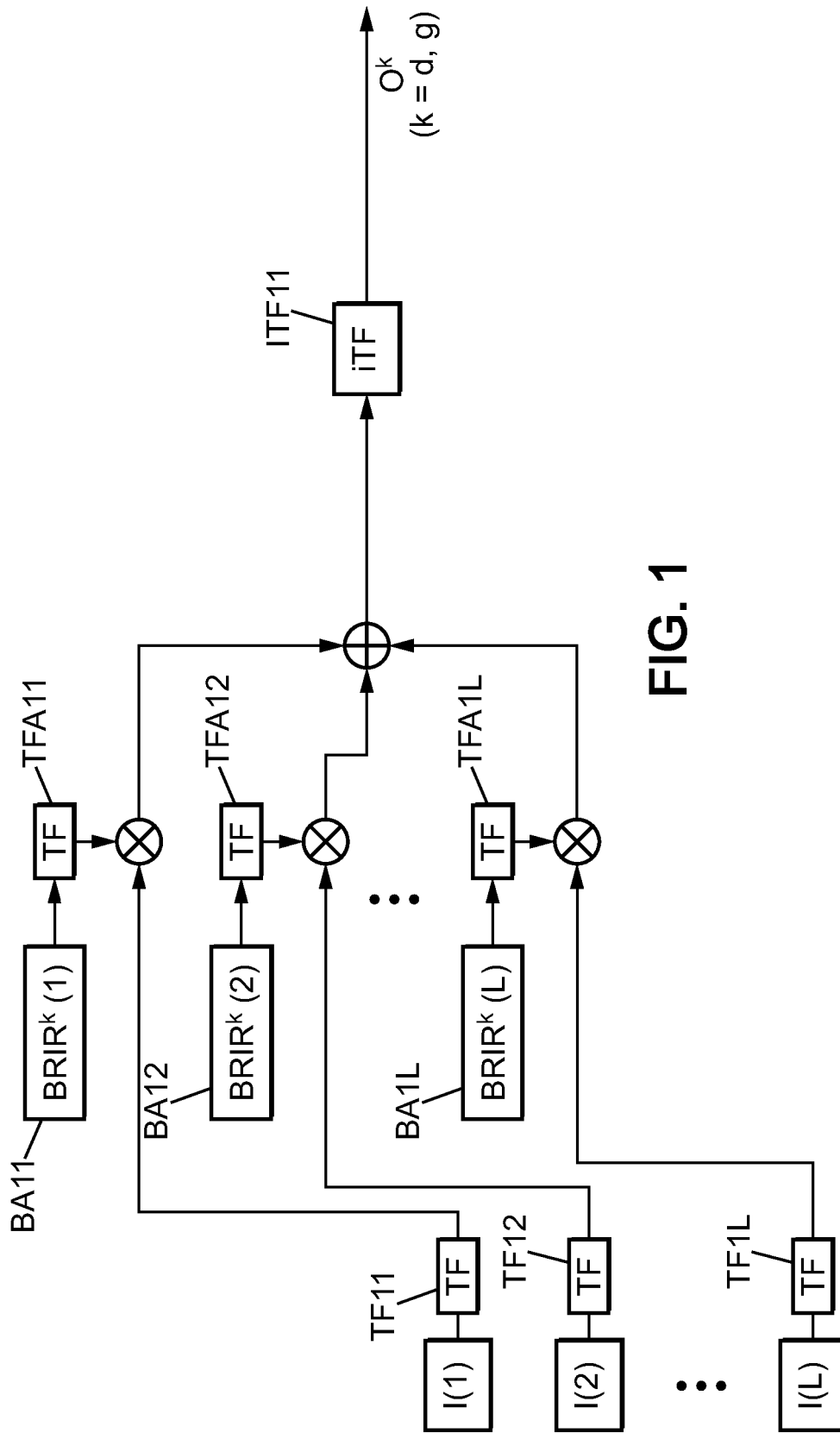


FIG. 1

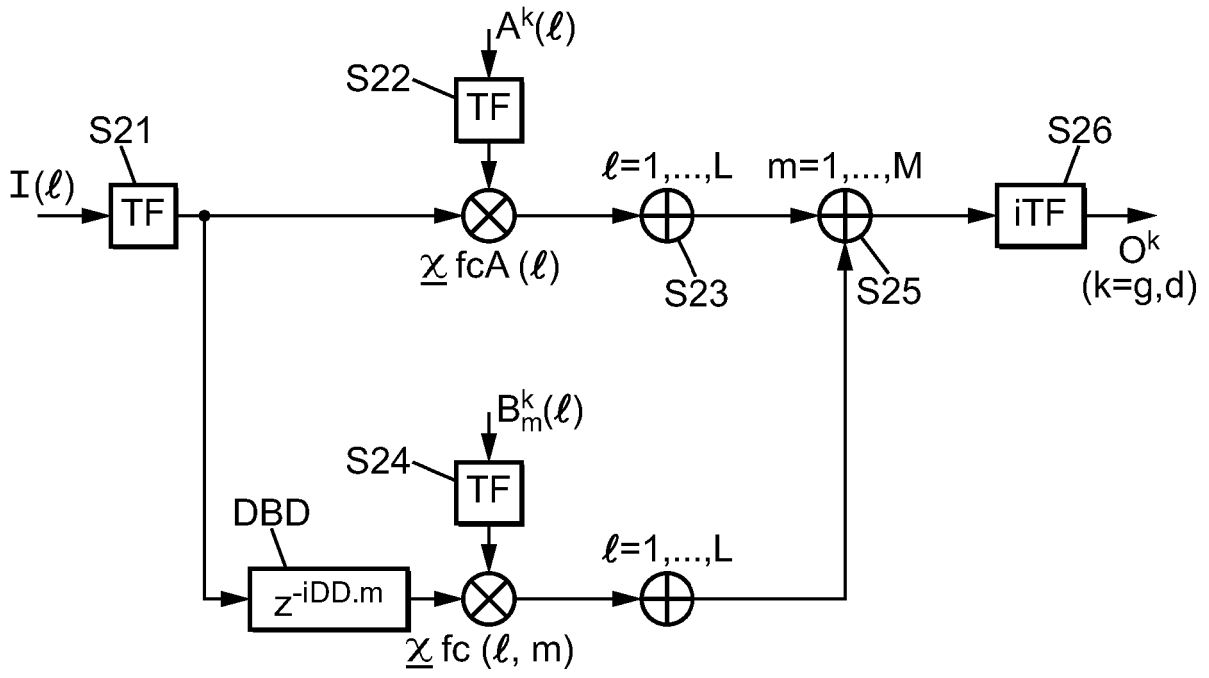


FIG. 2

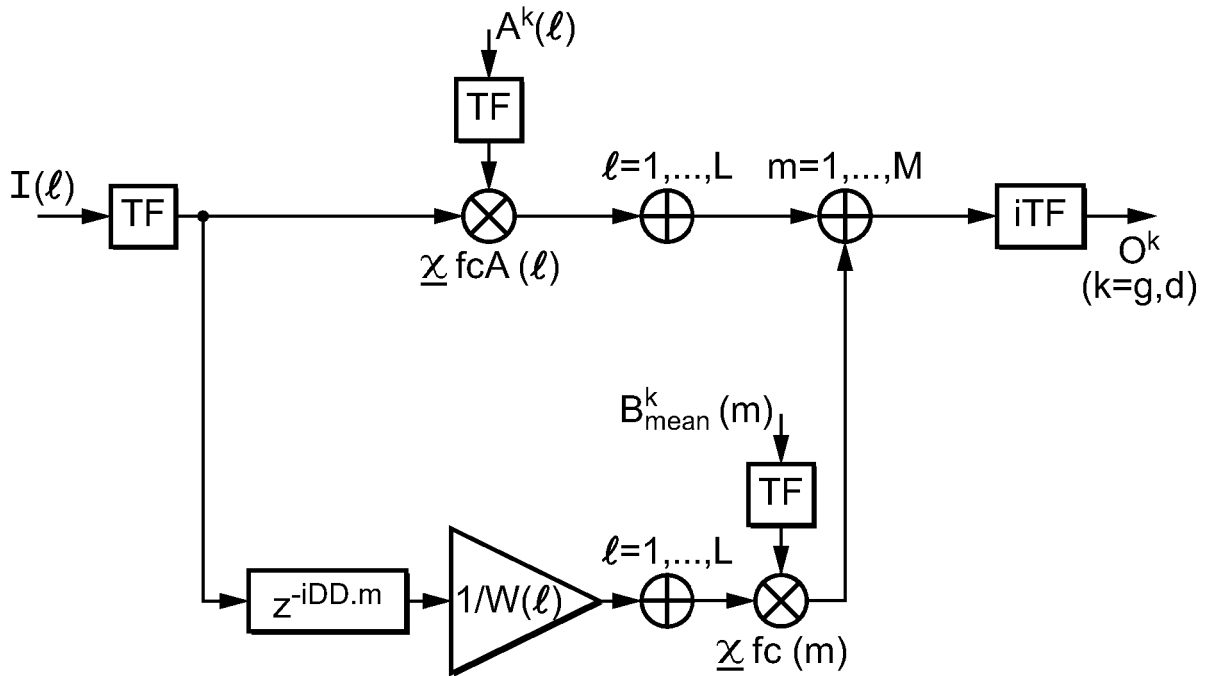


FIG. 5

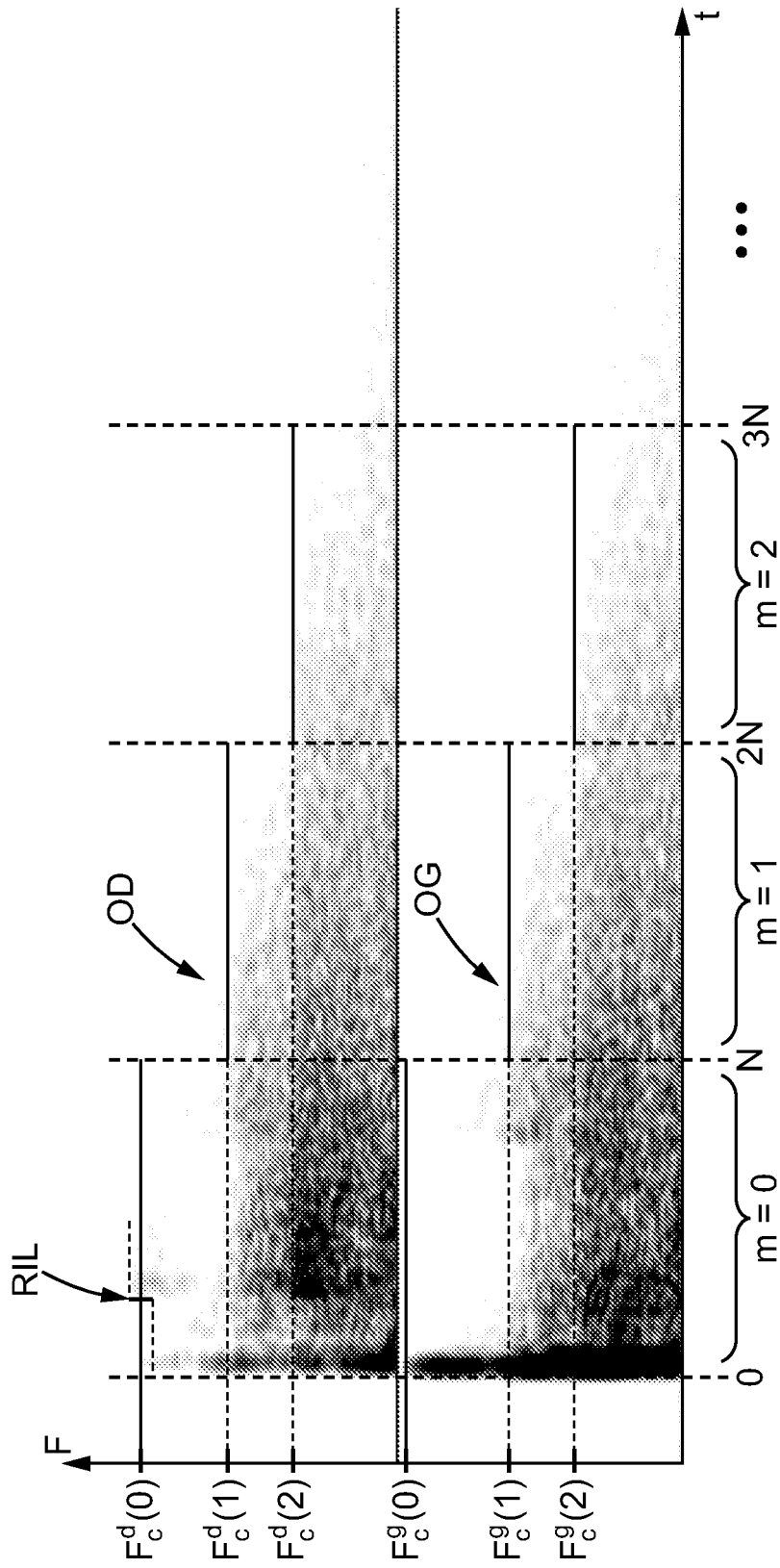


FIG. 3

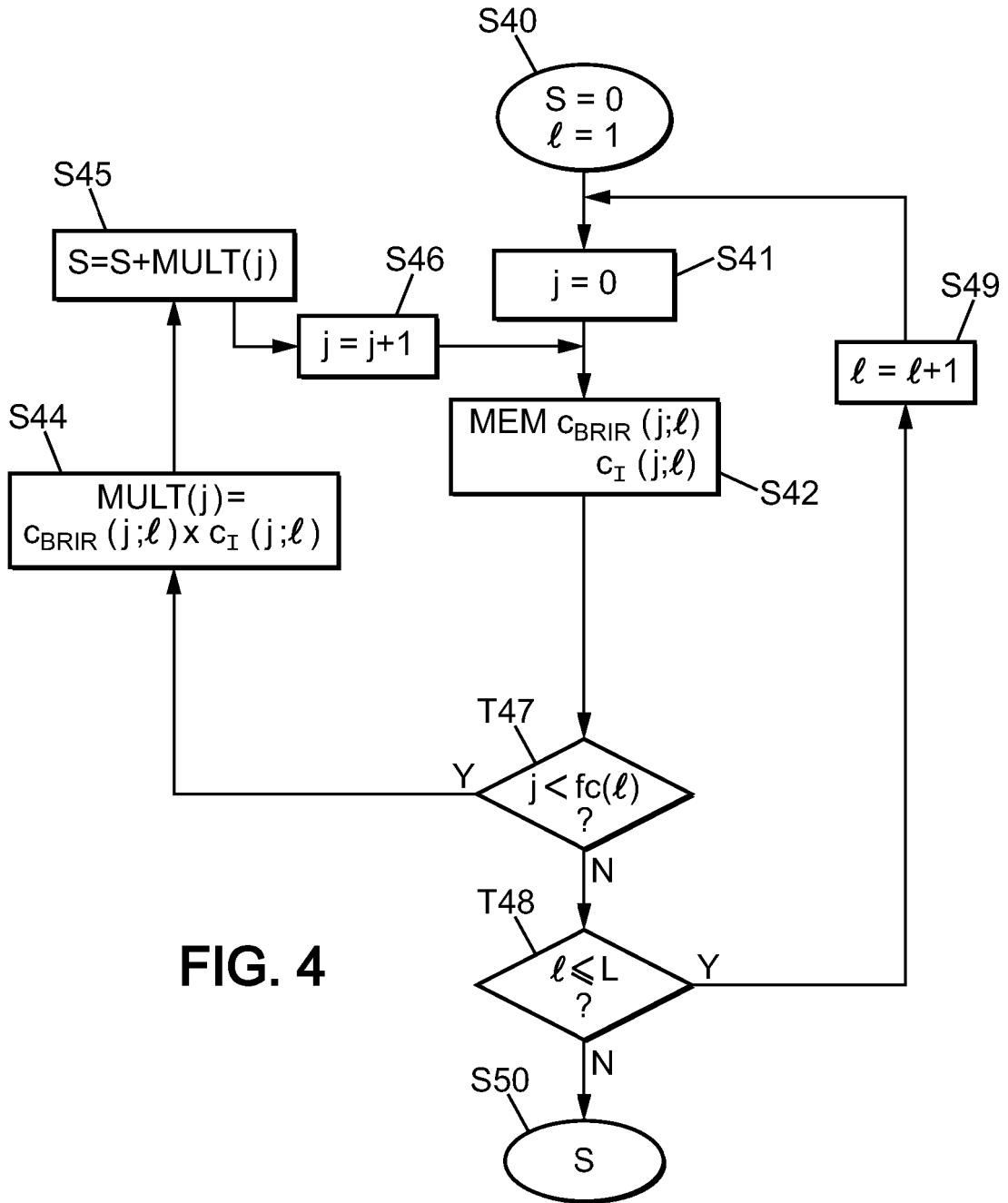
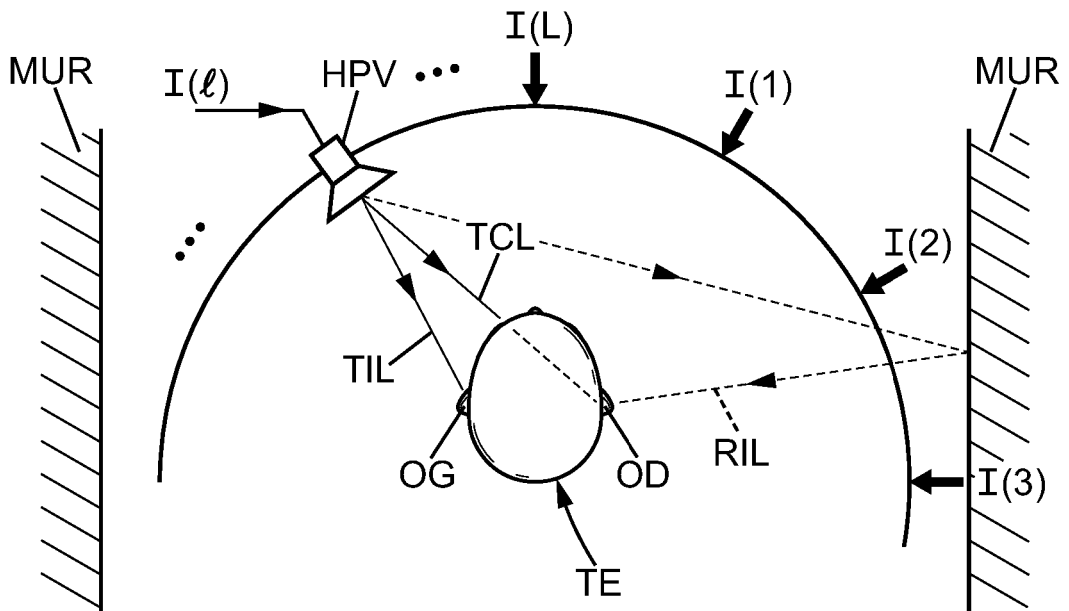
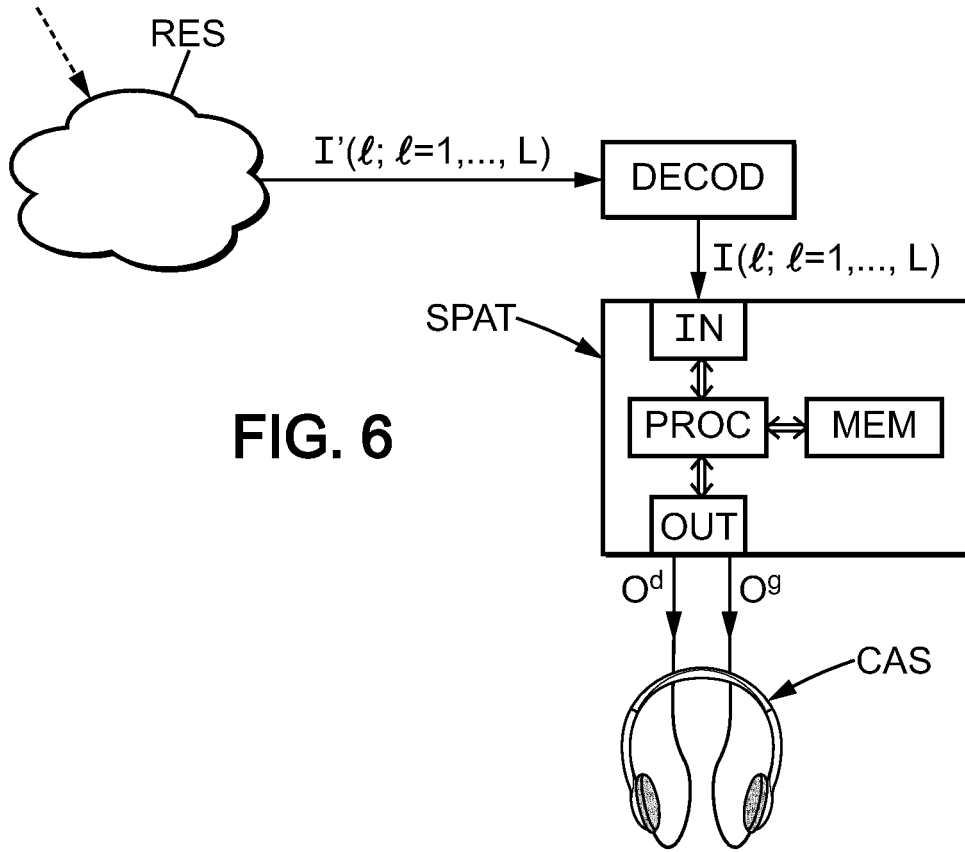


FIG. 4



**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.*

**Documents brevets cités dans la description**

- FR 1357299 [0004] [0027] [0051]
- US 2011170721 A1 [0006]