

①⑨ RÉPUBLIQUE FRANÇAISE
—
**INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE**
—
COURBEVOIE
—

①① N° de publication : **3 079 706**

(à n'utiliser que pour les
commandes de reproduction)

②① N° d'enregistrement national : **18 52774**

⑤① Int Cl⁸ : **H 04 N 21/20 (2018.01), H 04 L 29/02**

⑫

BREVET D'INVENTION

B1

⑤④ PROCÉDE ET SYSTÈME DE DIFFUSION D'UN FLUX AUDIO MULTICANAL A DES TERMINAUX DE SPECTATEURS ASSISTANT A UN ÉVÈNEMENT SPORTIF.

②② Date de dépôt : 29.03.18.

③⑦ Priorité :

④③ Date de mise à la disposition du public
de la demande : 04.10.19 Bulletin 19/40.

④⑤ Date de la mise à disposition du public du
brevet d'invention : 04.06.21 Bulletin 21/22.

⑤⑥ Liste des documents cités dans le rapport de
recherche :

Se reporter à la fin du présent fascicule

⑥⑦ Références à d'autres documents nationaux
apparentés :

○ Demande(s) d'extension :

⑦① Demandeur(s) : *INSTITUT MINES TELECOM
Etablissement public — FR.*

⑦② Inventeur(s) : BLOUET RAPHAEL et ESSID SLIM.

⑦③ Titulaire(s) : *INSTITUT MINES TELECOM
Etablissement public.*

⑦④ Mandataire(s) : PLASSERAUD IP.

FR 3 079 706 - B1



Procédé et système de diffusion d'un flux audio multicanal à des terminaux de spectateurs assistant à un évènement sportif

5 [001] L'invention concerne les procédés et les systèmes de diffusion permettant d'améliorer la perception sonore de spectateurs assistant à un évènement sportif.

[002] Plus précisément, l'invention porte sur des procédés et des systèmes de diffusion utilisant la réalité augmentée sonore, et permettant à un spectateur assistant à un évènement
10 sportif de percevoir des flux audio personnalisés issus de l'évènement.

[003] Le document WO 2007/115392 décrit un système permettant à des spectateurs d'assister à distance à plusieurs évènements sportifs se déroulant simultanément. Plus précisément, les spectateurs sont munis d'un terminal mobile électronique sur lequel ils
15 reçoivent plusieurs flux multimédias correspondant chacun à un des évènements sportifs retransmis en direct. Typiquement, les flux multimédias peuvent être des flux vidéo ou des flux audio. Le spectateur peut choisir à quels flux multimédia il souhaite avoir accès. Toutefois, un tel système ne permet pas une immersion complète du spectateur.

20 [004] La présente invention vient améliorer la situation.

[005] A cet effet, elle propose un procédé de diffusion d'un flux audio multicanal à des terminaux de spectateurs assistant à un évènement sportif,
le procédé comprenant, auprès d'un serveur :

25 - une acquisition, au moyen d'une pluralité de capteurs audio répartis sur un espace accueillant l'évènement sportif, d'une pluralité de flux audio bruts constituant une scène sonore comportant une ou plusieurs sources sonores,

- une analyse de la scène sonore comprenant :

30 ○ une isolation de chaque source sonore de la scène sonore et l'attribution d'un flux audio monocanal, propre à chaque source sonore,

 ○ pour chaque flux audio monocanal, une localisation dans l'espace en fonction d'une position dans l'espace accueillant l'évènement sportif de la source sonore associée au flux audio monocanal,

35 ○ la génération d'une table de description des flux audio monocanal, la table de description comportant un identifiant de chaque flux audio monocanal et, en

association avec chaque identifiant de flux audio monocanal, au moins une donnée de localisation,

5 - une transmission de la table de description avec un flux audio multicanal vers un ou plusieurs terminaux, le flux audio multicanal agrégeant les flux audio monocanal, et auprès d'un terminal:

10 ○ une utilisation de la table de description reçue et du flux audio multicanal pour appliquer un filtrage de spatialisation sonore des flux audio monocanal en fonction des localisations respectives de chaque flux audio monocanal, d'une part, et d'une position du spectateur dans ledit espace accueillant l'évènement sportif, d'autre part, en vue d'une restitution sonore spatialisée d'une partie au moins des flux audio monocanal sur le terminal..

15 [006] Grâce à ces dispositions, le spectateur est réellement immergé, de manière auditive, dans l'évènement sportif. En effet, la restitution sonore spatialisée permet de prendre en compte la position du spectateur afin que les flux audio diffusés correspondent exactement à ce qu'il est en train de vivre.

[007] Selon une réalisation, l'analyse de la scène sonore comprend en outre :

20 - pour chaque flux audio monocanal, une classification dans une classe parmi une pluralité de classes, en fonction d'un type de la source sonore associée au flux audio monocanal,

la table comportant en outre une donnée de classe de flux audio monocanal en association de chaque identifiant de flux audio monocanal,

25 et le procédé comprenant, auprès du terminal :

30 - une lecture de la table de description, pour identifier la classe de chaque flux audio monocanal,
 - l'application de gains respectifs aux flux audio monocanal pour la restitution sonore spatialisée, les gains étant choisis par un spectateur du terminal en fonction de la classe de chaque flux audio monocanal.

[008] De cette manière, le spectateur peut choisir quel type de flux audio il souhaite écouter. Son immersion dans l'évènement sportif n'en est qu'améliorée puisqu'il peut mettre de côté les flux audio dont la source sonore ne l'intéresse pas. Le spectateur vit une expérience
 35 personnalisée de l'évènement sportif.

[009] Selon une réalisation, le procédé comprend une détermination d'une orientation du spectateur par rapport à un référentiel choisi, le procédé comprenant en outre une application de gains respectifs aux flux audio monocanal spatialisés, les gains respectifs appliqués aux flux audio monocanal spatialisés évoluant en fonction de la position et/ou de l'orientation du spectateur par rapport au référentiel choisi.

[010] De cette manière, l'immersion du spectateur dans l'évènement sportif est encore augmentée. En effet, la détermination de son orientation lui permet d'effectuer un « zoom audio » sur la partie de l'espace accueillant l'évènement sportif qu'il est en train de viser.

[011] Selon une réalisation, le procédé comprend une division de l'espace accueillant l'évènement sportif en une pluralité de sous-espaces, au moins un capteur audio étant prévu dans chaque sous-espace, le procédé comportant la construction de la scène sonore par un mixage des flux bruts captés par les capteurs audio, le procédé comprenant, pour l'analyse de la scène sonore, une détermination des positions de sources sonores relativement à chacun desdits sous-espaces.

[012] De cette manière, cette étape permet d'effectuer la restitution sonore spatialisée avec un nombre réduit d'étapes, de calculs et d'échanges de données. En effet, il suffit de déterminer de manière grossière la localisation des sources sonores.

[013] Selon une réalisation, l'isolation de sources sonores est effectuée par séparation de sources.

[014] Selon une réalisation, l'isolation de sources sonore est effectuée par masquage temps-fréquence.

[015] Selon une réalisation, la classification des flux audio monocanal est menée par apprentissage, les flux audio monocanal étant classifiés par une technique de réseaux de neurones profonds.

[016] Ainsi il est possible de mettre en œuvre le procédé en temps réel.

[017] Selon une réalisation, le procédé comprend en outre, auprès du serveur :

- pour chaque phase de jeu, une détermination d'un type de la phase de jeu, parmi une pluralité de types de phases de jeu,
- une sélection, dans une bibliothèque de flux audio informatifs, d'au moins un flux audio informatif en fonction du type de la phase de jeu déterminé, et
- une transmission du flux audio informatif au terminal en vue de sa restitution sonore.

[018] Ainsi, cela permet au spectateur d'être immergé de manière encore plus importante dans l'évènement sportif. Cela permet également au spectateur de ne pas rater de phases de jeu cruciales puisque le système permet d'avertir le spectateur de la phase de jeu en cours.

[019] Selon une réalisation, la détermination du type de phase de jeu et éventuellement la localisation sur le terrain, associée à cette phase de jeu, peuvent être menées initialement par apprentissage sur des images acquises par capture vidéo de phases de jeu par au moins une caméra, l'apprentissage étant effectué au moyen d'une technique de réseaux de neurones profonds.

[020] De cette manière, la détermination et éventuellement la localisation du type de phase de jeu peuvent être faites en temps réel, de manière automatique.

[021] La présente invention vise aussi un système de diffusion d'un flux audio multicanal à des terminaux de spectateurs assistant à un évènement sportif comprenant :

- un module d'acquisition, comprenant une pluralité de capteurs audio répartis sur un espace accueillant l'évènement sportif, d'une pluralité de flux audio bruts constituant une scène sonore comportant une ou plusieurs sources sonores,
- un serveur comprenant un module informatisé d'analyse de la scène sonore configuré pour :
 - isoler chaque source sonore de la scène sonore et attribuer un flux audio monocanal propre à chaque source sonore,
 - pour chaque flux audio monocanal, localiser la source sonore associée au flux monocanal dans l'espace accueillant l'évènement sportif,
 - générer une table de description des flux audio monocanal, la table de description comportant un identifiant de chaque flux audio monocanal et, en association avec chaque identifiant de flux audio monocanal, au moins une donnée de localisation,

- un module de transmission configuré pour transmettre la table de description avec un flux audio multicanal vers un ou plusieurs terminaux, le flux audio multicanal agrégeant les flux audio monocal,
- 5 - un terminal configuré pour :
- utiliser la table de description reçue et le flux audio multicanal pour appliquer un filtrage de spatialisation sonore des flux audio monocal en fonction des localisations respectives de chaque flux audio monocal, d'une part, et d'une position du spectateur dans ledit espace accueillant l'évènement sportif, d'autre
- 10 part, en vue d'une restitution sonore spatialisée d'une partie au moins des flux audio monocal sur le terminal.

[022] Selon une réalisation, un capteur d'orientation est apte à déterminer l'orientation du spectateur dans l'espace, le capteur d'orientation étant choisi parmi une centrale inertielle

15 et/ou un accéléromètre.

[023] D'autres avantages et caractéristiques de l'invention apparaîtront à la lecture de la description détaillée ci-après d'exemples de réalisation de l'invention, et à l'examen des dessins annexés sur lesquels :

20

[024] La figure 1 est une vue schématique illustrant le fonctionnement d'un système de diffusion d'un flux audio multicanal selon l'invention,

[025] La figure 2 est un ordiogramme des étapes principales d'un procédé de diffusion d'un flux audio multicanal à des terminaux de spectateurs assistant à un évènement selon un exemple de réalisation selon l'invention,

25

[026] La figure 3 représente de manière schématique le système selon l'invention.

30 [027] Par « évènement sportif », on entend tout évènement sportif réunissant un grand nombre de spectateur à un même endroit pour assister à une performance sportive. Par exemple, le système peut être utilisé dans le cadre d'un match de football, de basketball, de baseball, de rugby, de hockey, de tennis, etc. Cette liste n'est pas limitative.

35 [028] Comme illustré sur les figures 1 et 3, le système SYST comprend un module d'acquisition MAC. Le module d'acquisition MAC comprend une pluralité de capteur audio

M1 - MN, par exemple des microphones. Les capteurs audio M1 – MN sont répartis sur l'espace accueillant l'évènement sportif. L'aspect de l'espace accueillant l'évènement sportif peut grandement varier en fonction du type d'évènement sportif. Notamment, l'espace accueillant l'évènement sportif peut être un stade avec des gradins, un circuit de course, etc.

[029] Pour chaque type d'espace accueillant l'évènement sportif (ou « espace » dans la suite de la description), les capteurs audio M1 – MN sont répartis de manière à ce que chaque flux audio bruts $F_{B1} - F_{BN}$ issu de l'évènement sportif soit capté. Les capteurs audio M1 – MN peuvent alors être répartis de manière homogène autour de l'espace accueillant l'évènement sportif.

[030] On entend par « flux audio bruts » les flux audio qui sont enregistrés par les capteurs audio M1 – MN et qui ne sont pas encore traités. Les flux audio bruts peuvent comprendre des clameurs de la foule assistant à l'évènement sportif, des bruits issus du jeu des sportifs, par exemple un coup dans une balle, un cri d'effort etc. Les flux audio bruts peuvent également comprendre des paroles échangées par les sportifs. Les flux audio bruts peuvent également comprendre les paroles échangées entre les spectateurs. Les flux audio bruts forment une scène sonore comportant une ou plusieurs sources sonores.

[031] Dans l'exemple de la figure 1, des flux audio bruts $F_{B1} - F_{BN}$ et des sources sonores S1 - SN sont représentés. Le flux audio brut F_{B1} , associé à la source sonore S1, correspond par exemple à un coup donné par un joueur dans un ballon. Le flux audio brut F_{B2} , associé à la source sonore S2, correspond à un coup de sifflet et le flux audio F_{BN} , associé à la source sonore SN, correspond à des paroles échangées entre des spectateurs. Les flux audio $F_{B1} - F_{BN}$ sont captés par les capteurs audio M1 – MN du module d'acquisition MAC.

[032] Les flux audio $F_{B1} - F_{BN}$ sont diffusés dans l'espace. Ainsi, chaque flux audio $F_{B1} - F_{BN}$ est capté de manière plus ou moins intense par chaque capteur audio M1 – MN. Par exemple, le capteur audio M1 capte principalement le flux audio F_{B1} , mais il capte également le flux audio F_{B2} . De même, le capteur audio M5 capte de manière très précise le flux audio brut F_{BN} , un peu moins le flux audio F_{B1} et encore moins le flux audio F_{B2} . La scène sonore SS est construite par un mixage de tous les flux audio bruts F_{B1} , F_{B2} et F_{BN} . La scène sonore SS comporte plusieurs sources sonores S1 – SN.

[033] Le système SYST peut comprendre au moins deux réseaux de capteurs audio M1 – MN. Le premier réseau de capteurs audio permet par exemple de capter le fond sonore de l'évènement sportif, typiquement la clameur de la foule assistant à l'évènement sportif. Le premier réseau de capteurs audio peut également permettre de capter les paroles échangées entre les spectateurs. Ainsi, les capteurs audio du premier réseau sont de préférence disposés près des spectateurs, et de préférence directement sur les spectateurs.

[034] Le deuxième réseau de capteurs audio M1 – MN peut également permettre de capter le fond sonore de l'évènement sportif ou la clameur de la foule. Le deuxième réseau peut également capter les flux audio bruts associés à l'évènement sportif. Cela peut comprendre les bruits issus du jeu, tel qu'un coup dans une balle, ou encore les paroles échangées entre les joueurs.

[035] Il est possible de définir au moins trois types de sources sonores. Le premier type correspond à l'ambiance de l'évènement sportif. Le premier type comprend notamment la clameur de la foule assistant à l'évènement sportif. Le deuxième type correspond aux évènements de l'évènement sportif. Il comprend par exemple les coups dans une balle, des coups de sifflet, etc. Le troisième type correspond aux paroles échangées entre les spectateurs de l'évènement sportif.

[036] La pluralité de flux audio bruts captés par chacun des capteurs audio M1 – MN est envoyée vers un local d'analyse LOC. La transmission est représentée par la flèche « IN » sur la figure 1. Le local d'analyse comprend au moins un serveur SERV auprès duquel un filtrage des flux audio bruts $F_{B1} - F_{BN}$ est appliqué. Notamment, un filtrage par séparation de sources est effectué. La séparation de sources permet d'isoler les différents évènements sonores de jeu à partir de la scène sonore enregistrée par les capteurs audio. Les flux filtrés sont transmis à un terminal TERM en vue d'une restitution spatialisée d'une partie au moins des flux audio sur le terminal TERM. Le terminal TERM est par exemple un terminal mobile de type téléphone intelligent, tablette électronique ou ordinateur.

[037] Des dispositifs de diffusion DIFF sont également prévus pour diffuser les flux audio spatialisés. Les dispositifs de diffusion DIFF peuvent être des enceintes personnelles ou des casques audio que chaque spectateur a à sa disposition.

[038] Le système SYST selon l'invention permet notamment de faire une restitution sonore personnalisée. Par exemple, la restitution sonore des sons relatifs à l'évènement sportif est faite en fonction de la localisation de chaque source sonore $S1 - SN$ ainsi que de la position du spectateur, utilisant le terminal TERM, dans l'espace accueillant l'évènement sportif. La position du spectateur peut être déterminée par un module de positionnement MP, apte à localiser le spectateur par géolocalisation par satellite (« GPS »), par triangulation ou par communication en champ proche (« Near field communication » ou « NFC » en anglais). Un spectateur entendra de manière plus audible un flux audio dont la source sonore associée est proche de lui. En reprenant l'exemple de la figure 1, un spectateur situé à gauche du terrain entendra de manière plus prononcée le flux audio F_{B1} associé à la source $S1$ que le flux audio $F2$ associé à la source sonore $S2$.

[039] Dans un exemple de réalisation, le spectateur a la possibilité d'entrer des préférences de restitution de la scène sonore via le terminal TERM. Par exemple, le spectateur peut choisir d'entendre des flux audio associés à un certain type de source sonore. Par exemple, le spectateur peut choisir de n'entendre que les paroles échangées entre les spectateurs, ou que les coups de sifflet de l'arbitre. Le spectateur peut également choisir d'entendre de manière plus ou moins intense tous les flux audio, en fonction du type de source sonore à laquelle ils sont associés.

[040] Selon un autre exemple de réalisation, le spectateur entend de manière plus intense certains flux audio en fonction de son orientation dans l'espace accueillant l'évènement sportif. Par exemple, le système comprend module d'orientation MO comprenant un capteur d'orientation pour déterminer dans quelle direction regarde le spectateur. Le spectateur peut alors être équipé d'un accessoire, par exemple un casque ou une vignette, supportant le capteur d'orientation. Le capteur d'orientation peut être un accéléromètre ou une centrale inertielle. Il est alors possible de déterminer vers quelle source sonore $S1 - SN$ le spectateur est tourné pour amplifier le volume de diffusion des flux audio associés à cette source sonore. L'expérience du spectateur en est améliorée puisque cela permet d'être complètement immergé dans l'évènement sportif auquel il assiste.

[041] La figure 2 illustre de manière plus précise les principales étapes du procédé mis en œuvre par le système SYST.

[042] Le module d'acquisition MAC capte tous les flux audio bruts $F_{B1} - F_{BN}$ au moyen des capteurs audio $M1 - MN$. Les flux audio bruts $F_{B1} - F_{BN}$ forment une scène sonore SS comprenant au moins une source sonore $S1 - SN$.

5

[043] Les étapes E1 à E5 sont effectuées auprès du serveur SERV. Le serveur SERV comprend un module informatisé d'analyse MIA permettant d'isoler chaque source sonore $S1 - SN$ de la scène sonore SS . La scène sonore SS est transmise au module informatisé d'analyse MIA. A l'étape E1, le module informatisé d'analyse MIA isole chaque source sonore $S1 - SN$ de la scène sonore SS . L'isolation des sources sonores $S1 - SN$ peut être faite par une méthode de séparation de sources par formation de voies, tel qu'illustré à l'étape E1A. En variante, l'étape d'isolation des sources sonores $S1 - SN$ est faite par un masquage temps-fréquence, tel qu'illustré à l'étape E1B. Le masquage temps-fréquence permet notamment de faire une séparation de sources lorsque deux évènements se sont produits simultanément dans une même zone de l'espace.

15

[044] L'étape E1 ayant permis d'isoler chaque source sonore $S1 - SN$, un flux audio monocanal $F_{mono1} - F_{monoN}$ propre est peut être attribué à chaque source sonore $S1 - SN$ à l'étape E2.

20

[045] A l'étape E3, chaque source sonore $S1 - SN$ est localisée dans l'espace accueillant l'évènement sportif. Ainsi, chaque flux audio monocanal $F_{mono1} - F_{monoN}$ peut être localisé en fonction de la localisation de la source sonore $S1 - SN$ qui lui est propre. Notamment, l'espace accueillant l'évènement sportif peut être divisé en une pluralité de sous-espaces. Par exemple, l'espace accueillant l'évènement sportif peut être quadrillé, bien que l'espace puisse être divisé selon une autre configuration. Par exemple, chaque sous-espace peut correspondre à une zone propre de jeu, par exemple le centre du terrain, chaque zone de but, les zones de penalty, etc. Il est prévu au moins un capteur audio $M1 - MN$ par sous-espace. La localisation des sources sonores $S1 - SN$ est réalisée relativement à chaque sous-espace. Plus précisément, la localisation de chaque source sonore $S1 - SN$ correspond au sous-espace dans lequel elle se situe. La localisation de chaque flux audio monocanal $F_{mono1} - F_{monoN}$ et celle de la source sonore $S1 - SN$ qui lui est associée.

30

[046] Le module informatisé d'analyse MIA génère une table de description TAB à l'étape E5. Chaque flux audio monocanal $F_{mono1} - F_{monoN}$ est défini par un identifiant ID. Chaque

35

identifiant ID est associé à une donnée de localisation LOC du flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$, déterminé à l'étape E3.

5 [047] Selon une variante de réalisation, le procédé selon l'invention comprend en outre une étape préalable E4 de classification des flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$. Cette classification est faite en fonction du type de la source sonore $S1 - SN$ associée au flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$. Comme décrit plus haut, il y a au moins trois types de sources sonores $S1 - SN$. Le premier type correspond à l'ambiance de l'évènement sportif et
 10 comprend notamment la clameur de la foule. Le deuxième type correspond aux évènements de l'évènement sportif et comprend notamment les coups des joueurs et les sifflets de l'arbitre. Le troisième type correspond aux paroles échangées entre spectateurs. Chaque type de source sonore peut être associé à au moins une classe. Plusieurs classes peuvent correspondre à un même type de source sonore. Par exemple, une première classe
 15 correspondant aux coups de pied dans une balle et une deuxième classe comprenant les coups de sifflet de l'arbitre peuvent être associées au deuxième type de source sonore.

[048] Selon une réalisation, la classification des flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$ peut être effectuée au moyen d'une méthode de séparation de sources supervisée. La classification
 20 est menée par apprentissage, au moyen d'une technique de réseaux de neurones profonds. Cela a l'avantage de permettre une mise en œuvre en temps réel du procédé. En effet, l'apprentissage permet de reconnaître, dès la transmission vers le serveur SERV des flux audio captés par le module d'acquisition MAC, le type de la source sonore à laquelle les flux audio sont associés.

25 [049] Selon cette variante de réalisation, une classe TYP est en outre attribuée à chaque identifiant ID de flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$ compris dans la table de description TAB.

30 [050] La table de description TAB est ensuite transmise vers le terminal TERM à l'étape E6 par un module de transmission MT. La transmission comprend également un flux audio multicanal F_{multi} comprenant l'ensemble des flux audio monocanal.

[051] Selon une variante de réalisation, le système SYST ne comprend pas de module de
 35 transmission MT. Les flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$ sont compris dans une bibliothèque BIB et sont synthétisés en local en vue de leur restitution sonore spatialisée.

[052] Les étapes E1 à E6 sont réalisées en temps réel auprès du serveur SERV.

5 [053] Les étapes E7 à E11 sont effectuées auprès du terminal TERM. Le terminal TERM a reçu la table de description TAB ainsi que le flux audio multicanal F_{multi} agrégeant l'ensemble des flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$.

[054] Le terminal TERM peut utiliser la table de description pour appliquer au moins un
10 filtrage de spatialisation sonore. Le terminal TERM reçoit ainsi à l'étape E7 des données de positionnement du spectateur utilisant le terminal TERM. Les données de positionnement sont obtenues par le module de positionnement MP et/ou par le module d'orientation MO. Les données de positionnement comprennent au moins la position géographique du spectateur dans le référentiel de l'espace accueillant l'évènement sportif.

15 [055] Les données de positionnement peuvent en outre comprendre une donnée relative à l'orientation du spectateur par rapport à un référentiel fixe de l'espace, par exemple le centre du terrain. L'orientation du spectateur correspond au sous-espace vers lequel il est tourné ; L'utilisation de données d'orientation du spectateur permet, comme vu plus haut, de faire un
20 « zoom audio » sur la partie de l'espace visée par le spectateur.

[056] Les données de positionnement sont acquises en temps réel et en continu par le module de positionnement MP.

25 [057] A l'étape E8, un filtrage est appliqué aux flux audio monocanal $F_{\text{mono}1} - F_{\text{mono}N}$ en fonction des données de positionnement du spectateur et de la localisation des flux audio monocanal compris dans la table de description TAB. Une pluralité de flux audio monocanal spatialisés $F_{\text{spatia}1} - F_{\text{spatia}N}$ est ainsi obtenue.

30 [058] Toutefois, le spectateur a la possibilité de mettre en avant certains flux audio par rapport à d'autres. Il a donc la possibilité de choisir s'il souhaite faire un zoom audio en fonction de son orientation dans l'espace, ou quelles classes de flux audio à diffuser. Ainsi, à l'étape E9, le terminal TERM reçoit des instructions de préférence du spectateur. Le spectateur peut par exemple entrer ses instructions via une interface du terminal TERM. Par exemple, une
35 application dédiée peut être prévue pour permettre un échange entre le spectateur et le terminal TERM.

[059] A l'étape E10, le terminal TERM applique des gains respectifs à au moins une partie des flux audio monocanal spatialisés $F_{\text{spatia}1} - F_{\text{spatia}N}$.

5

[060] Selon la variante de réalisation dans laquelle l'orientation du spectateur est prise en compte, la table de description TAB est lue. Un gain respectif est appliqué aux flux audio monocanal spatialisés $F_{\text{spatia}1} - F_{\text{spatia}N}$ dont les localisations associées dans la table de description TAB correspondent à la direction dans laquelle le spectateur est orienté. Plus
10 précisément, un gain plus important est appliqué aux flux audio monocanal spatialisés $F_{\text{spatia}1} - F_{\text{spatia}N}$ dont la source sonore est localisée dans le même sous-espace que celui visé par le spectateur.

[061] Les données de positionnement étant acquises en temps réel et en continu, les gains
15 appliqués peuvent évoluer en fonction de ces données. Un changement d'orientation ou de position géographique du spectateur entraîne une modification de la valeur des gains appliqués aux différents flux audio monocanal spatialisés. La modification de la valeur des gains appliqués est faite en temps réel.

[062] Selon la deuxième variante de réalisation dans laquelle la classe des flux audio est prise
20 en compte, la table de description TAB est lue afin de déterminer quelle classe est associée à quel flux audio monocanal spatialisé $F_{\text{spatia}1} - F_{\text{spatia}N}$. Un gain respectif est appliqué aux flux audio monocanal spatialisés $F_{\text{spatia}1} - F_{\text{spatia}N}$ dont la classe correspond à celle choisie par le spectateur. Le spectateur peut modifier à tout moment ses instructions de préférence, de sorte
25 qu'une application de gains différente est réalisée. La modification des valeurs des gains respectifs est effectuée en temps réel.

[063] A l'étape E11, les flux audio monocanal spatialisés $F_{\text{spatia}1} - F_{\text{spatia}N}$ sont envoyés vers un module de diffusion de type casque audio ou enceinte personnelle. Le spectateur peut alors
30 profiter de l'immersion auditive fournie par le système selon l'invention.

[064] Selon une variante de réalisation, il est également prévu d'intégrer aux flux audio diffusés aux spectateurs des informations relatives au déroulement de l'évènement sportif. Selon cette variante, le serveur SERV peut en outre comprendre une bibliothèque BIB
35 comprenant une pluralité de flux audio informatifs $F_{\text{info}1} - F_{\text{info}N}$ associés chacun à une phase de jeu. Les flux audio informatifs $F_{\text{info}1} - F_{\text{info}N}$ peuvent être diffusés au spectateur

automatiquement. Le type de la phase de jeu en cours est déterminé parmi une pluralité de types de phases de jeu. Les types de phases de jeu comprennent par exemple un but, un penalty, une faute, etc. Selon le type de phase de jeu déterminé, au moins un flux audio informatif $F_{\text{info}1} - F_{\text{info}N}$ est sélectionné dans la bibliothèque BIB en vue de sa restitution sonore. Le flux audio informatif $F_{\text{info}1} - F_{\text{info}N}$ peut être automatiquement diffusé.

[065] En variante, les flux audio informatifs $F_{\text{info}1} - F_{\text{info}N}$ ne sont diffusés que si le spectateur l'a choisi. Les flux audio informatifs $F_{\text{info}1} - F_{\text{info}N}$ permettent notamment d'ajouter une information auditive sur ce que le spectateur visualise. Par exemple, dans une phase de but, le flux audio auditif diffusé indique au spectateur qu'il y a but. Cela permet d'augmenter encore l'immersion du spectateur dans l'évènement sportif auquel il assiste. Cela permet également de ne manquer aucune phase importante de jeu.

[066] Selon un exemple, la détermination du type de la phase de jeu en cours est réalisé au moyen d'un dispositif de capture DC comprenant au moins une caméra vidéo, tel qu'illustré sur la figure 3. Le dispositif de capture DC est disposé de manière à pouvoir suivre tous les phases de l'évènement sportif. Les images capturées par le dispositif de capture peuvent être transmises au module informatisé d'analyse qui est apte à déterminer, en temps réel, un type de la phase de jeu que le dispositif de capture DC est en train de capturer. Les types de phases de jeu peuvent être déterminés au moyen d'une méthode d'apprentissage par réseau de neurones profonds.

Revendications

1. Procédé de diffusion d'un flux audio multicanal (F_{multi}) à des terminaux (TERM) de spectateurs assistant à un évènement sportif,
- le procédé comprenant, auprès d'un serveur (SERV) :
- une acquisition, au moyen d'une pluralité de capteurs audio ($M1 - MN$) répartis sur un espace accueillant l'évènement sportif, d'une pluralité de flux audio bruts ($F_{B1} - F_{BN}$) constituant une scène sonore (SS) comportant une ou plusieurs sources sonores ($S1 - SN$),
 - une analyse de la scène sonore (SS) comprenant :
 - une isolation de chaque source sonore ($S1 - SN$) de la scène sonore (SS) et l'attribution d'un flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), propre à chaque source sonore ($S1 - SN$),
 - pour chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), une localisation dans l'espace en fonction d'une position dans l'espace accueillant l'évènement sportif de la source sonore ($S1 - SN$) associée au flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),
 - la génération d'une table de description (TAB) des flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), la table de description (TAB) comportant un identifiant (ID) de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) et, en association avec chaque identifiant (ID) de flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), au moins une donnée de localisation,
 - une transmission de la table de description (TAB) avec un flux audio multicanal (F_{multi}) vers un ou plusieurs terminaux (TERM), le flux audio multicanal (F_{multi}) agrégeant les flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),
- et auprès d'un terminal (TERM):
- une utilisation de la table de description (TAB) reçue et du flux audio multicanal (F_{multi}) pour appliquer un filtrage de spatialisation sonore des flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) en fonction des localisations respectives de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), d'une part, et d'une position du spectateur dans ledit espace accueillant l'évènement sportif, d'autre part, en vue d'une restitution sonore spatialisée d'une partie au moins des flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) sur le terminal (TERM),
- et dans lequel l'analyse de la scène sonore (SS) comprend en outre :
- pour chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), une classification dans une classe parmi une pluralité de classes, en fonction d'un type de la source sonore ($S1 - SN$) associée au flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),

la table de description (TAB) comportant en outre une donnée de classe de flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) en association de chaque identifiant (ID) de flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),

5 et le procédé comprenant, auprès du terminal (TERM) :

- une lecture de la table de description (TAB), pour identifier la classe de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),

- l'application de gains respectifs aux flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) pour la restitution sonore spatialisée, les gains étant choisis par un spectateur du terminal (TERM) en fonction de la classe de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$).

10

2. Procédé selon la revendication 1, comprenant une détermination d'une orientation du spectateur par rapport à un référentiel choisi,

le procédé comprenant en outre une application de gains respectifs aux flux audio monocanal spatialisés ($F_{\text{spatia}1} - F_{\text{spatia}N}$), les gains respectifs appliqués aux flux audio monocanal spatialisés ($F_{\text{spatia}1} - F_{\text{spatia}N}$) évoluant en fonction de la position et/ou de l'orientation du spectateur par rapport au référentiel choisi.

15

3. Procédé selon l'une des revendications précédentes, comprenant une division de l'espace accueillant l'évènement sportif en une pluralité de sous-espaces, au moins un capteur audio (M1 – MN) étant prévu dans chaque sous-espace,

20

le procédé comportant la construction de la scène sonore (SS) par un mixage des flux audio bruts ($F_B1 - F_BN$) captés par les capteurs audio (M1 – MN),

le procédé comprenant, pour l'analyse de la scène sonore (SS), une détermination des positions de sources sonores (S1 – SN) relativement à chacun desdits sous-espaces.

25

4. Procédé selon l'une des revendications 1 à 3, dans lequel l'isolation de sources sonores (S1 – SN) est effectuée par séparation de sources.

30 5. Procédé selon l'une des revendications 1 à 3, dans lequel l'isolation de sources sonore (S1 – SN) est effectuée par masquage temps-fréquence.

6. Procédé selon l'une des revendications 1 à 5, dans lequel la classification des flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) est menée par apprentissage, les flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) étant classifiés par une technique de réseaux de neurones profonds.

35

7. Procédé selon l'une des revendications 1 à 6, comprenant en outre, auprès du serveur (SERV) :

- pour chaque phase de jeu, une détermination d'un type de la phase de jeu, parmi une pluralité de types de phases de jeu,
- une sélection, dans une bibliothèque (BIB) de flux audio informatifs ($F_{\text{info}1} - F_{\text{info}N}$), d'au moins un flux audio informatif ($F_{\text{info}1} - F_{\text{info}N}$) en fonction du type de la phase de jeu déterminé, et
- une transmission du flux audio informatif ($F_{\text{info}1} - F_{\text{info}N}$) au terminal (TERM) en vue de sa restitution sonore.

8. Procédé selon la revendication 7, dans lequel la détermination du type de phase de jeu est menée par apprentissage sur des images acquises par capture vidéo de phases de jeu par au moins une caméra, l'apprentissage étant effectué au moyen d'une technique de réseaux de neurones profonds.

9. Système de diffusion (SYST) d'un flux audio multicanal (F_{multi}) à des terminaux (TERM) de spectateurs assistant à un évènement sportif comprenant :

- un module d'acquisition (MAC), comprenant une pluralité de capteurs audio ($M1 - MN$) répartis sur un espace accueillant l'évènement sportif, d'une pluralité de flux audio bruts ($F_{B1} - F_{BN}$) constituant une scène sonore (SS) comportant une ou plusieurs sources sonores ($S1 - SN$),
- un serveur (SERV) comprenant un module informatisé d'analyse (MIA) de la scène sonore (SS) configuré pour :
 - isoler chaque source sonore ($S1 - SN$) de la scène sonore (SS) et attribuer un flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) propre à chaque source sonore ($S1 - SN$),
 - pour chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), localiser la source sonore ($S1 - SN$) associée au flux monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) dans l'espace accueillant l'évènement sportif,
 - classifier chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) dans une classe parmi une pluralité de classes, en fonction d'un type de la source sonore ($S1 - SN$) associée au flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),
 - générer une table de description (TAB) des flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), la table de description (TAB) comportant un identifiant (ID) de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) et, en association avec chaque identifiant (ID) de flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), au moins une

donnée de localisation et une donnée de classe de flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),

- un module de transmission (MT) configuré pour transmettre la table de description (TAB) avec un flux audio multicanal (F_{multi}) vers un ou plusieurs terminaux (TERM), le flux audio multicanal (F_{multi}) agrégeant les flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$),

- un terminal (TERM) configuré pour :

- utiliser la table de description (TAB) reçue et le flux audio multicanal (F_{multi}) pour appliquer un filtrage de spatialisation sonore des flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) en fonction des localisations respectives de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), d'une part, et d'une position du spectateur dans ledit espace accueillant l'évènement sportif, d'autre part, en vue d'une restitution sonore spatialisée d'une partie au moins des flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) sur le terminal,
- lire la table de description (TAB) pour identifier la classe de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$), et
- appliquer des gains respectifs aux flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$) pour la restitution sonore spatialisée, les gains étant choisis par un spectateur du terminal (TERM) en fonction de la classe de chaque flux audio monocanal ($F_{\text{mono}1} - F_{\text{mono}N}$).

10. Système selon la revendication 9, dans lequel un capteur d'orientation est apte à déterminer l'orientation du spectateur dans l'espace, le capteur d'orientation étant choisi parmi une centrale inertielle et/ou un accéléromètre.

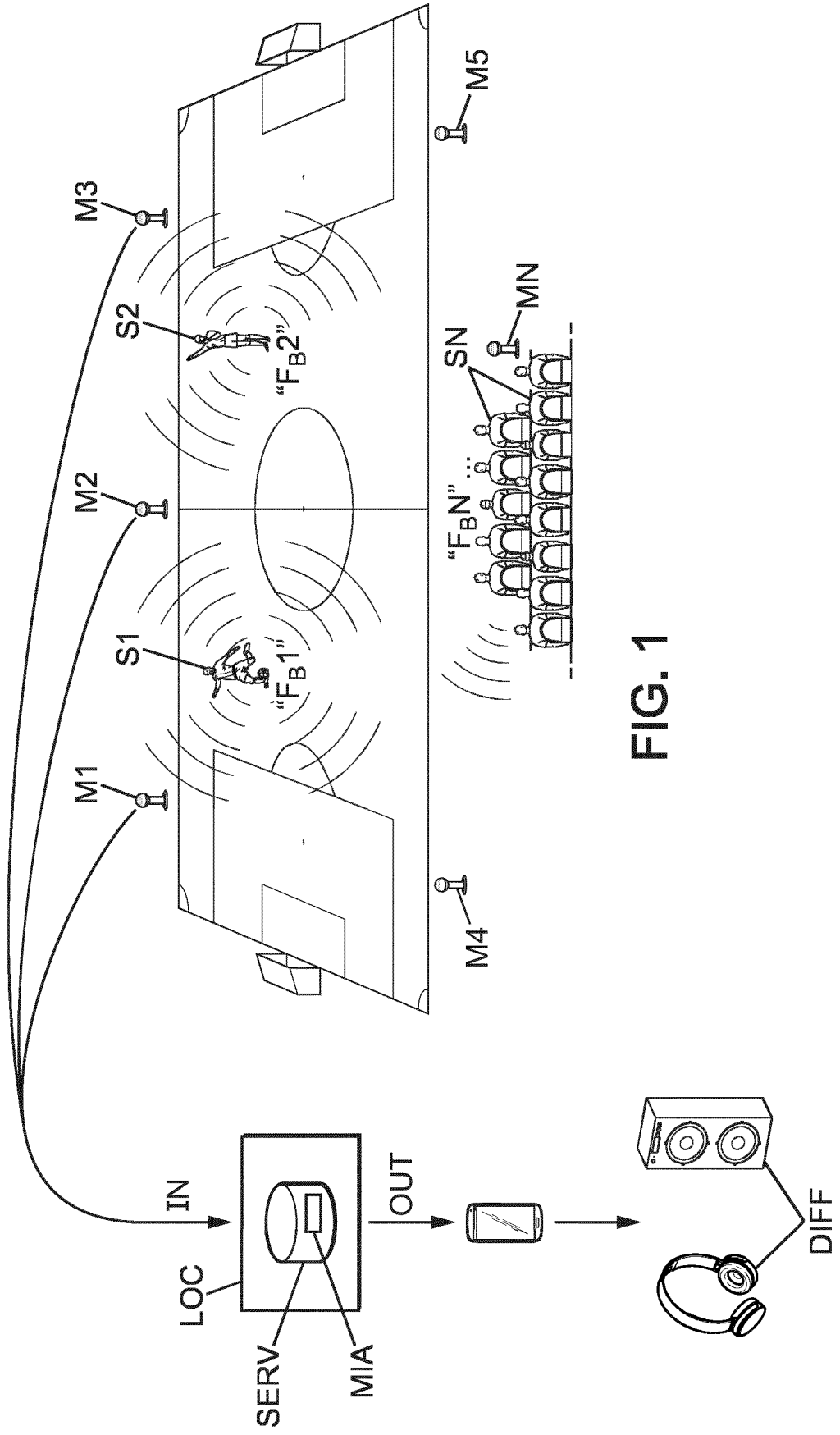
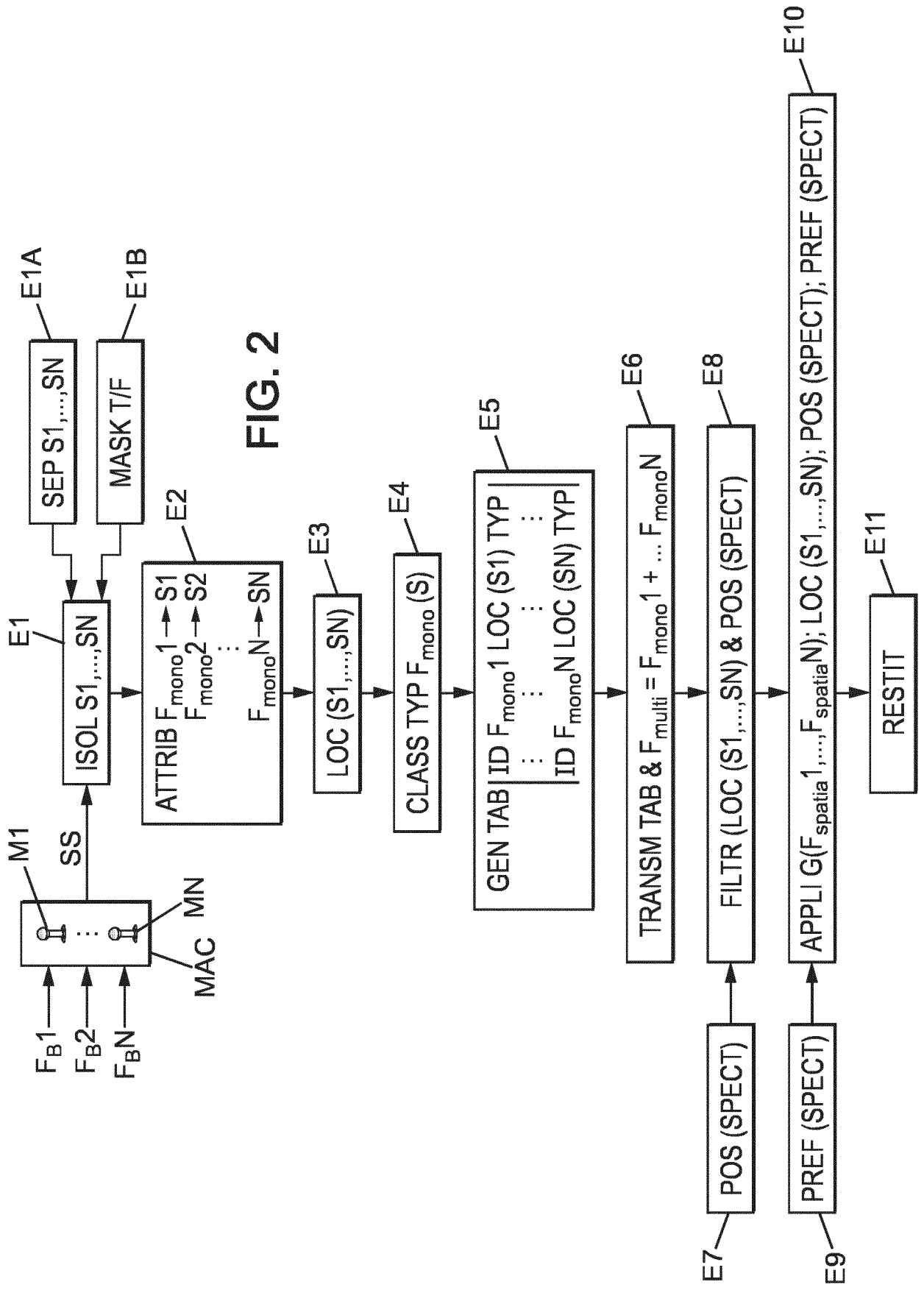


FIG. 1



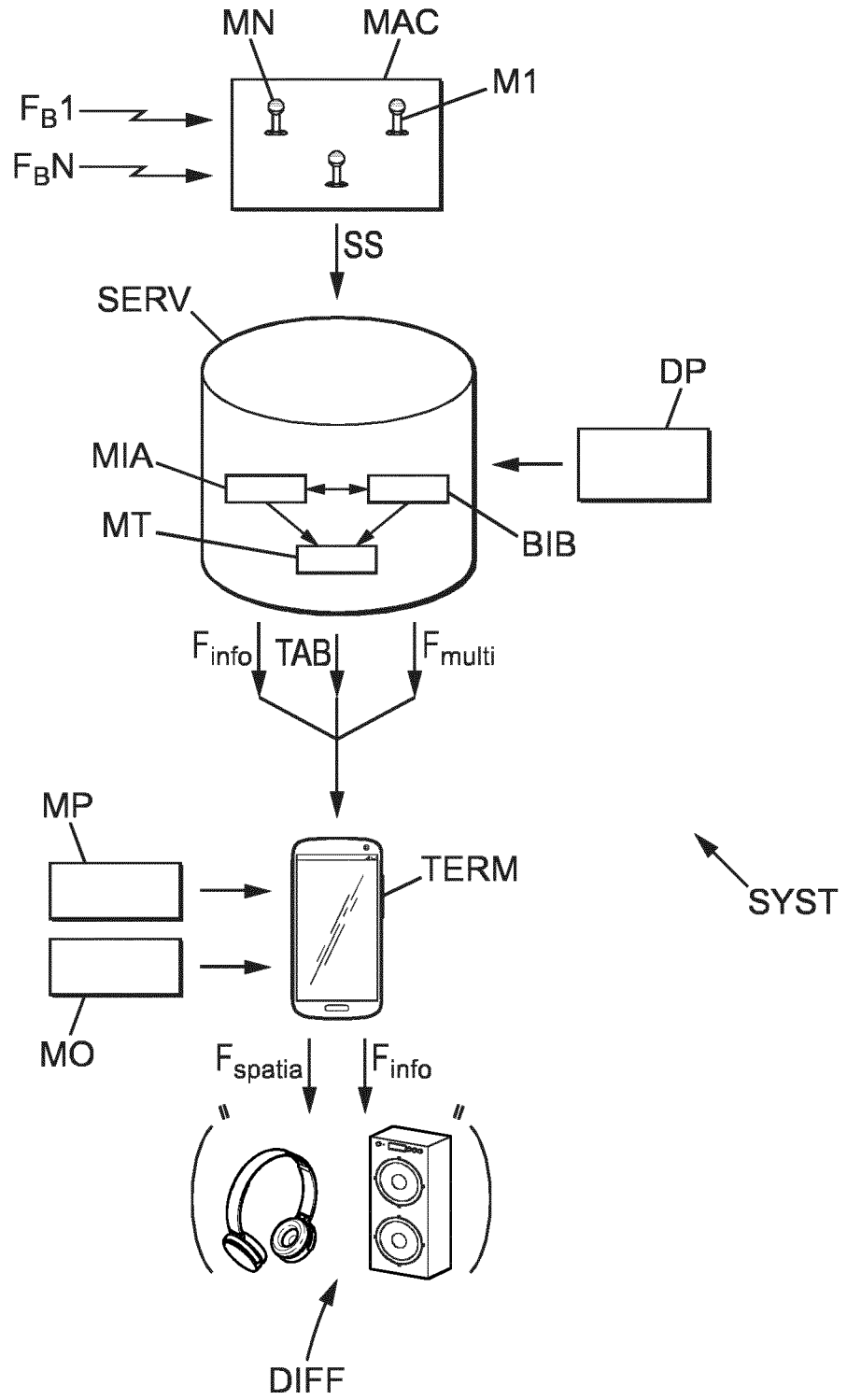


FIG. 3

RAPPORT DE RECHERCHE

articles L.612-14, L.612-53 à 69 du code de la propriété intellectuelle

OBJET DU RAPPORT DE RECHERCHE

L'I.N.P.I. annexe à chaque brevet un "RAPPORT DE RECHERCHE" citant les éléments de l'état de la technique qui peuvent être pris en considération pour apprécier la brevetabilité de l'invention, au sens des articles L. 611-11 (nouveau) et L. 611-14 (activité inventive) du code de la propriété intellectuelle. Ce rapport porte sur les revendications du brevet qui définissent l'objet de l'invention et délimitent l'étendue de la protection.

Après délivrance, l'I.N.P.I. peut, à la requête de toute personne intéressée, formuler un "AVIS DOCUMENTAIRE" sur la base des documents cités dans ce rapport de recherche et de tout autre document que le requérant souhaite voir prendre en considération.

CONDITIONS D'ETABLISSEMENT DU PRESENT RAPPORT DE RECHERCHE

Le demandeur a présenté des observations en réponse au rapport de recherche préliminaire.

Le demandeur a maintenu les revendications.

Le demandeur a modifié les revendications.

Le demandeur a modifié la description pour en éliminer les éléments qui n'étaient plus en concordance avec les nouvelles revendications.

Les tiers ont présenté des observations après publication du rapport de recherche préliminaire.

Un rapport de recherche préliminaire complémentaire a été établi.

DOCUMENTS CITES DANS LE PRESENT RAPPORT DE RECHERCHE

La répartition des documents entre les rubriques 1, 2 et 3 tient compte, le cas échéant, des revendications déposées en dernier lieu et/ou des observations présentées.

Les documents énumérés à la rubrique 1 ci-après sont susceptibles d'être pris en considération pour apprécier la brevetabilité de l'invention.

Les documents énumérés à la rubrique 2 ci-après illustrent l'arrière-plan technologique général.

Les documents énumérés à la rubrique 3 ci-après ont été cités en cours de procédure, mais leur pertinence dépend de la validité des priorités revendiquées.

Aucun document n'a été cité en cours de procédure.

1. ELEMENTS DE L'ETAT DE LA TECHNIQUE SUSCEPTIBLES D'ETRE PRIS EN CONSIDERATION POUR APPRECIER LA BREVETABILITE DE L'INVENTION

US 5 714 997 A (ANDERSON DAVID P [US])
3 février 1998 (1998-02-03)

WO 2017/182714 A1 (NOKIA TECHNOLOGIES OY
[FI]) 26 octobre 2017 (2017-10-26)

US 9 832 587 B1 (PARK HYUN JIN [US] ET AL)
28 novembre 2017 (2017-11-28)

US 2014/355769 A1 (PETERS NILS GÜNTHER
[US] ET AL) 4 décembre 2014 (2014-12-04)

US 2015/195641 A1 (DI CENSO DAVIDE [US] ET
AL) 9 juillet 2015 (2015-07-09)

US 2015/373474 A1 (KRAFT NOAH [US] ET AL)
24 décembre 2015 (2015-12-24)

2. ELEMENTS DE L'ETAT DE LA TECHNIQUE ILLUSTRANT L'ARRIERE-PLAN TECHNOLOGIQUE GENERAL

NEANT

3. ELEMENTS DE L'ETAT DE LA TECHNIQUE DONT LA PERTINENCE DEPEND DE LA VALIDITE DES PRIORITES

NEANT