(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

COURBEVOIE

(11) **N° de publication :** ***3 078 196***

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

(21) **N° d'enregistrement national :** **18 51311**

(51) Int Cl[8] : ***G 10 L 21/055*** *(2018.01), G 06 F 3/16*

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt :** 16.02.18.

(30) **Priorité :**

(43) **Date de mise à la disposition du public de la demande :** 23.08.19 Bulletin 19/34.

(56) **Liste des documents cités dans le rapport de recherche préliminaire :** *Se reporter à la fin du présent fascicule*

(60) **Références à d'autres documents nationaux apparentés :**

( ) **Demande(s) d'extension :**

(71) **Demandeur(s) :** *SWEESP Société par actions simplifiée* — FR.

(72) **Inventeur(s) :** DELAPORTE FABRICE et DIARD ANTHONY.

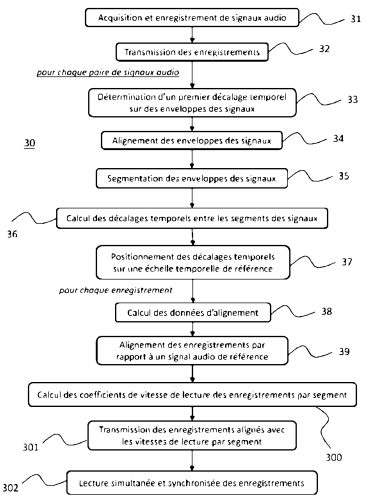
(73) **Titulaire(s) :** SWEESP Société par actions simplifiée.

(74) **Mandataire(s) :** IPAZ.

(54) PROCEDE ET DISPOSITIF DE SYNCHRONISATION D'ENREGISTREMENTS COMPRENANT UN SIGNAL AUDIO PROVENANT DE DIFFERENTS DISPOSITIFS D'ACQUISITION CO-LOCALISES.

(57) La présente invention concerne un procédé de synchronisation d'enregistrements comprenant un signal audio provenant de différents dispositifs d'acquisition (2, 3, 4, 5) co-localisés. Le procédé comprend les étapes :

- de détermination (33) d'un premier décalage temporel à partir d'une enveloppe des signaux audio, en segmentant lesdits signaux par détection de plages de silence ;
- de calcul (36) des décalages temporels entre les segments des signaux deux à deux ;
- d'alignement (39) des signaux selon un signal de référence ;
- de calcul (300) des coefficients de lecture des enregistrements.



FR 3 078 196 - A1

Procédé et dispositif de synchronisation d'enregistrements comprenant un signal audio provenant de différents dispositifs d'acquisition co-localisés

Domaine technique

L'invention se situe dans le domaine de la restitution d'enregistrements comprenant un signal audio. Plus particulièrement, l'invention peut s'appliquer à une synchronisation d'enregistrements musicaux associés à des enregistrements vidéo en vue de produire une vidéo musicale.

Etat de la technique – Description du problème technique

Avec l'avènement des téléphones multimédias, incluant des capteurs audio et vidéo d'une qualité sans cesse croissante, les capacités d'enregistrement du grand public ont largement progressé. Les réseaux sociaux permettent par ailleurs une publication et une diffusion d'enregistrements privés. Dans ce cadre, différentes études ont été menées pour synchroniser des contenus provenant de différents appareils capturant une même scène ou un même événement afin de produire un contenu agrégé à partir de différentes sources disponibles.

Un exemple d'application de la présente invention se trouve dans le domaine du spectacle puisque chaque personne du public est en capacité d'enregistrer et de filmer une même performance ou un même événement. Cependant, le public effectue en général une captation de son de mauvaise qualité. En effet, le microphone du téléphone peut être mal positionné ou soumis au bruit du public. Et surtout, les actuels microphones omnidirectionnels des téléphones ne permettent pas d'envisager un retraitement efficace à postériori des signaux enregistrés même par des moyens professionnels.

Un autre exemple peut être celui de musiciens jouant sur une même scène. Les musiciens peuvent chacun enregistrer le son de leur propre instrument par exemple. La problématique à résoudre dans ce cas est celle de la synchronisation à priori de leurs téléphones afin de pouvoir envisager un traitement ultérieur des enregistrements pour produire un contenu audio de

bonne qualité. À cette fin, les musiciens doivent s'équiper d'un matériel particulier qui nécessite des compétences techniques ainsi qu'un temps d'apprentissage et un investissement financier important.

Un prérequis à tout traitement audio et vidéo est de travailler sur un ensemble de fichiers présentant des caractéristiques temporelles communes permettant l'application de traitements et la restitution d'un contenu final cohérent. Ceci implique de déterminer un alignement ou un point de départ commun dans le temps et une vitesse de lecture commune. Or, pour des enregistreurs physiquement indépendants il n'est pas possible sur un plan pratique de déclencher un enregistrement exactement au même moment sur deux appareils ou plus. Il convient donc de repositionner les enregistrements sur une échelle de temps commune.

De plus, les fréquences d'échantillonnage, dans le domaine audio ou vidéo sont établies par des oscillateurs physiquement indépendants sur les différents enregistreurs. Il est donc à nouveau impossible de garantir l'égalité précise de ces fréquences d'échantillonnage. Les écarts entre les fréquences d'échantillonnage doivent donc être corrigés après l'enregistrement, par l'application d'une correction sur la vitesse de lecture des fichiers d'enregistrements audio et vidéo. On va donc chercher à maintenir au cours du temps la synchronisation initiale, c'est-à-dire à partir d'un point de départ temporel commun. Sans cette étape essentielle, après un temps variable de l'ordre de quelques dizaines de minutes de lecture du clip, un décalage perceptible par l'oreille humaine peut se produire. Une telle opération est donc indispensable pour la synchronisation d'enregistrements d'une durée supérieure à une dizaine de minutes.

Enfin, les dispositifs d'enregistrement à la disposition du plus grand nombre, et notamment les smartphones, présentent, du fait qu'ils ne sont pas uniquement dédiés à l'enregistrement, un risque d'erreur plus important lors d'un enregistrement. Par exemple, on peut constater des phénomènes dits de « buffer underrun » ou de « buffer overrun ». Un « buffer underrun » désigne une incapacité du dispositif d'échantillonnage du smartphone à fournir un signal dans les temps à un dispositif de stockage du smartphone, induisant une discontinuité, comme un blanc ou une répétition dans le signal

stocké. Un « buffer overrun » désigne une incapacité du matériel de stockage à écrire suffisamment rapidement le signal déjà échantillonné sur le dispositif de stockage, induisant une discontinuité qui se matérialise par une disparition d'une portion du signal réel, dans le signal stocké.

Ces problèmes de synchronisation ne se posent pas dans un studio d'enregistrement musical professionnel. En effet, dans le cas de prises de sons simultanées, un seul appareil, typiquement une console numérique multipiste, réalise l'enregistrement. Ainsi, il n'y a qu'un seul point de départ et qu'une seule fréquence d'horloge.

En revanche, dans le domaine de l'audio-visuel, un tel mode d'enregistrement n'existe pas. En effet, l'enregistrement est réalisé par des caméras qui effectuent directement l'échantillonnage. Une première technique bien connue pour synchroniser des enregistrements est celle du « clap » : un dispositif mécanique produit un son net et bref dont on peut aisément identifier l'instant dans un flux vidéo ou audio. Cet évènement commun à tous les enregistrements permet d'aligner les enregistrements provenant de différentes caméras. L'inconvénient de cette méthode est qu'elle nécessite une maîtrise complète du déroulement de l'enregistrement et se révèle donc inadaptée à des productions d'amateurs.

Alternativement, la synchronisation de plusieurs enregistrements peut aussi être réalisée par une production d'un « time code » ou marqueur temporel, avec une horloge maître sur un dispositif centralisé fournissant des signaux de synchronisation de manière cyclique et régulière. Ces informations sont transmises aux appareils par le biais de réseaux filaires, sans fil ou encore par des horloges de haute précision dites esclaves. Les horloges esclaves sont initialement synchronisées avec l'horloge maître et sont programmées pour ne pas présenter de décalage perceptible sur une durée nominale d'utilisation. Cette solution nécessite un matériel spécifique dont la mise en œuvre est complexe et couteuse et donc, par conséquent, réservée à des professionnels du domaine de l'audio-visuel.

Par ailleurs, différentes études ont été menées pour synchroniser des contenus capturés par différents appareils grand public dans le but de produire un contenu agrégé à partir des différentes sources disponibles. On peut citer l'exemple d'extraits de concerts filmés par plusieurs personnes dans le public. Plusieurs techniques ont été développées dans ce cadre pour répondre à ce besoin de synchronisation des enregistrements. Ces techniques sont basées sur une utilisation du signal audio comme moyen de synchronisation des vidéos. Ces techniques reposent fondamentalement sur deux méthodes d'analyse du signal sonore : une méthode dans le domaine temporel via des fonctions de corrélation et une méthode dans le domaine fréquentiel via le calcul d'empreintes fréquentielles caractéristiques. Ces techniques permettent un calcul d'un décalage temporel constant au cours du temps qui ne tient pas compte des variations de fréquences d'horloge au sein des différents appareils d'enregistrement. Ce type d'analyses est adapté pour une synchronisation et une restitution simultanée d'enregistrements courts ou disjoints mais ne permet pas de synchroniser du contenu d'une longueur de l'ordre de plusieurs dizaines de minutes devant être restitués de façon simultanée, tel un mixage audio d'un concert d'une heure par exemple. En effet, le décalage des horloges d'échantillonnage internes des différents appareils rend nécessaire l'application d'un coefficient correcteur sur la vitesse de lecture de chaque enregistrement.

Ensuite des erreurs potentielles de continuité dans le signal peuvent invalider ces synchronisations vu qu'elles calculent un décalage temporel constant qui ne peut donc tenir compte de la variation induite par l'erreur de continuité dans le signal. Ceci rend donc impossible une création d'un contenu agrégé cohérent. La correction de ce type d'erreur nécessite une intervention manuelle d'édition de l'enregistrement pour le corriger.

Dans le domaine de la synchronisation de signaux audio, on peut aussi noter l'existence de techniques visant à introduire depuis un système de traitement du signal des éléments caractéristiques dans le signal d'origine qui ne modifient pas pour autant la perception humaine de ces signaux, afin de détecter par la suite ces éléments lors d'un réenregistrement du signal. Ces techniques nécessitent encore que les enregistreurs soient par

avance connus et que l'on dispose d'un moyen de contrôle sur ces enregistreurs ce qui n'est pas le cas lors d'un enregistrement par plusieurs smartphones privés d'un même concert.

Résumé de l'invention

Un objectif de l'invention est notamment de palier aux inconvénients précités la technique.

La présente invention propose à cette fin un procédé de synchronisation d'au moins deux enregistrements. Chaque enregistrement comprend un signal audio. Les enregistrements proviennent de différents dispositifs d'acquisition co-localisés. Ledit procédé comprend au moins les étapes suivantes :

a) acquisition et enregistrements de signaux audio par les dispositifs d'acquisition ;
b) transmission des enregistrements par les dispositifs d'acquisition à un serveur de calcul ;
c) pour chaque paire d'enregistrements, les traitements suivants sont réalisés par le serveur de calcul :
    - détermination d'un premier décalage temporel entre les deux signaux audio à partir d'une enveloppe de chaque signal audio ;
    - alignement des enveloppes des signaux audio en fonction du premier décalage temporel ;
    - segmentation de chaque enveloppe de signal audio par détection de plages de silence sur lesdites enveloppes de chaque signal audio ;
    - calcul des décalages temporels entre les segments correspondant de chaque signal audio ;
    - définition d'une échelle temporelle de référence et positionnement de l'ensemble des décalages temporels entre chaque segment de chaque signal audio sur l'échelle temporelle de référence ;

d) calcul de données d'alignement à partir d'un positionnement de l'ensemble des décalages temporels sur l'échelle temporelle de référence, pour chaque enregistrement, par le serveur de calcul ;

e) alignement des enregistrements par rapport à un signal audio de référence par le serveur de calcul à partir des données d'alignement ;

f) calcul de coefficients de vitesse de lecture pour chaque segment de chaque enregistrement par le serveur de calcul ;

g) transmission des enregistrements alignés, avec leurs coefficients de vitesse de lecture, à un dispositif de restitution d'enregistrements ;

h) lecture simultanée et synchronisée de l'ensemble des enregistrements par un dispositif de restitution d'enregistrements.

Une plage de silence peut être définie par une énergie du signal comprise entre une valeur minimum et une valeur maximum prédéfinies.

L'enveloppe peut être une enveloppe d'énergie du signal audio.

Alternativement l'enveloppe peut être une moyenne mobile calculée sur le signal audio ré-échantillonné.

Les enveloppes peuvent être normalisées et alignées en réalisant une corrélation croisée pour déterminer le premier décalage temporel.

Ledit procédé peut comprendre en outre une étape de filtrage des plages de silence détectées.

Ledit procédé peut également comprendre une étape d'élimination de valeurs aberrantes parmi l'ensemble des décalages temporels calculés entre chaque segment de chacun des deux signaux audio.

Le calcul des données d'alignement peut comprendre les étapes suivantes :

- construction d'une fonction d'interpolation des décalages temporels d'un signal par rapport aux décalages temporels d'un autre signal ;
- modélisation d'un ajustement des décalages par exemple par un système de ressort linéaire ;
- construction d'un système d'équations linéaires pour calculer les valeurs de décalages ajustées constituant les données d'alignements.

Ledit procédé peut s'appliquer à des enregistrements comprenant en outre un signal vidéo.

Ledit procédé peut comprendre des étapes de mixage et de montage des enregistrements par le dispositif de restitution d'enregistrements avant la lecture desdits enregistrements.

La présente invention concerne également un dispositif de synchronisation d'enregistrements, comprenant au moins :

- des dispositifs d'acquisition de signaux audio co-localisés réalisant notamment les opérations d'acquérir, d'enregistrer et de transmettre des signaux audio et vidéo ;
- un serveur de calcul mettant notamment en œuvre les étapes c), d), e), f), du procédé selon l'invention, ledit serveur de calcul comprenant à cette fin au moins un processeur de calcul, une mémoire, des moyens de réception et de transmission de signaux ;
- un dispositif de restitution d'enregistrements comprenant par exemple des moyens de mixage, de montage et de lecture des enregistrements.

Les dispositifs d'acquisition peuvent être disposées chacun à proximité d'une parmi des sources sonores.

Les dispositifs d'acquisition de signaux audio peuvent être adaptés à une acquisition de signaux vidéo.

L'invention concerne également un produit programme d'ordinateur comprenant des instructions qui, lorsque le programme est exécuté par au moins un processeur du serveur de calcul, conduisent celui-ci à mettre en œuvre les étapes c), d), e), f) du procédé selon l'invention.

L'invention se rapporte aussi à un support d'enregistrement lisible par ordinateur comprenant des instructions qui, lorsqu'elles sont exécutées par au moins un processeur d'un serveur de calcul, conduisent celui-ci à mettre en œuvre les étapes c), d), e), f) du procédé selon l'invention.

Avantageusement, l'invention permet une synchronisation continue c'est-à-dire sur la durée totale d'un ensemble d'enregistrements, chacun d'une durée arbitraire, réalisés dans un même lieu, c'est-à-dire co-localisés, et réalisés à un même moment.

Description des figures

D'autres avantages et caractéristiques de l'invention apparaîtront à l'examen de la description détaillée de plusieurs modes de réalisation nullement limitatifs, et des dessins annexés, sur lesquels :

- La figure 1 représente un exemple de dispositifs d'enregistrements ;
- La figure 2 représente un exemple d'une mise en œuvre d'un dispositif selon l'invention ;
- La figure 3 représente plusieurs étapes d'un procédé de synchronisation selon l'invention ;
- La figure 4a représente un premier signal audio en entrée du procédé de synchronisation selon l'invention ;
- La figure 4b représente une enveloppe du premier signal audio ;
- La figure 5a représente une enveloppe d'un deuxième signal audio en entrée du procédé de synchronisation selon l'invention ;
- La figure 5b représente l'enveloppe du premier signal audio ;
- La figure 5c représente une fonction résultat d'une corrélation entre les enveloppes des premier et deuxième signaux audio ;
- La figure 6a représente l'enveloppe du deuxième signal audio tronquée ;
- La figure 6b représente l'enveloppe du premier signal audio tronquée ;
- La figure 6c représente une combinaison des enveloppes du premier et deuxième signal audio ;
- La figure 7 représente un premier spectrogramme d'un intervalle de l'enveloppe d'un troisième signal audio ;
- La figure 8a représente un deuxième spectrogramme d'un intervalle de l'enveloppe d'un quatrième signal audio ;
- La figure 8b représente un troisième spectrogramme d'un intervalle d'une enveloppe d'un cinquième signal ;
- La figure 8c représente un exemple d'une fonction de corrélation spectrale des parties réelles des deuxième et troisième spectrogrammes ;

- La figure 9 représente un exemple de partitionnement des valeurs de décalage temporel.

Description détaillée

La description de la présente invention utilise pour l'exemple une production de clips vidéo musicaux agrégeant de multiples captures sonores et visuelles, c'est-à-dire provenant de différents enregistreurs. L'invention peut avantageusement être utilisée pour réaliser des captations via différents téléphones multimédia d'évènements comme des présentations, des conférences, des colloques ou encore pour de la captation de spectacles non musicaux tels des pièces de théâtre. D'autres exemples d'applications peuvent être envisagés sans sortir du cadre de l'invention. Notamment, et de manière générale, l'invention peut s'appliquer à une synchronisation de signaux numériques quelconques dès lors qu'ils comprennent un signal audio.

La figure 1 représente un événement 1, comme un concert par exemple ou toute autre représentation ou spectacle. L'évènement 1 est caractérisé par au moins une source sonore 6, 7, 8, 9 émettant un signal audio. Le dispositif selon l'invention comprend des dispositifs d'acquisition de signaux audio indépendants 2, 3, 4, 5. Les dispositifs d'acquisition de signaux audio 2, 3, 4, 5 comprennent au moins un micro. Les dispositifs d'acquisition de signaux audio 2, 3, 4, 5 peuvent acquérir soit uniquement des signaux audio, soit des signaux audio et vidéo. Par exemple un microphone 4 peut enregistrer principalement une première source sonore 8 de l'événement 1. Alternativement il est possible d'utiliser d'autres types d'enregistreurs audio comme des magnétophones, dictaphones. Par exemple les dispositifs d'acquisition de signaux audio peuvent être des téléphones multimédias 2, 3, 5 ou encore des caméras avec une prise de son. Avantageusement chaque dispositif d'acquisition de signaux audio peut être orienté de façon à capter principalement une source sonore en particulier. Par exemple un premier téléphone multimédia 2 peut enregistrer principalement une deuxième source sonore 7 en étant situé à proximité de ladite deuxième

source sonore 7. Un deuxième téléphone multimédia 3 peut enregistrer principalement une troisième source sonore 9, un troisième téléphone multimédia 5 peut enregistrer principalement une quatrième source sonore 6. Ceci est particulièrement avantageux pour enregistrer un concert réunissant plusieurs instruments et par exemple un ou plusieurs chanteurs. Ainsi en disposant d'enregistrements plus particulièrement centrés sur certaines sources, l'ensemble sera rendu de manière très qualitative. Alternativement il est possible que les enregistreurs soient disposés à des endroits aléatoires par rapport aux sources. Il est également possible d'envisager que les enregistreurs puissent être en mouvement par rapport aux sources sonores.

La figure 2 représente un exemple d'un dispositif de synchronisation 20 selon l'invention. Le dispositif de synchronisation 20 peut par exemple comprendre les dispositifs d'acquisition de signaux audio 2, 3, 4, 5 tels que représentés sur la figure 1. Les dispositifs d'acquisition, ou enregistreurs, de signaux audio 2, 3, 4, 5 après enregistrement de l'évènement 1 peuvent transmettre leurs enregistrements par liaison filaire ou non filaire. La transmission peut se faire en passant par exemple par un réseau de télécommunication et plus spécifiquement par l'intermédiaire d'un réseau relié à un système de stockage distant 26 en exploitant les technologies de l'informatique en nuage, ou « cloud computing » en langage anglo-saxon. Ces enregistrements sont ensuite récupérés par un serveur de calcul 21 distant des enregistreurs fonctionnant sur un ou plusieurs ordinateurs eux-mêmes reliés au réseau externe et au « cloud » 26 et apte à être connectés aux dispositifs d'acquisition de signaux audio 2, 3, 4, 5. Le serveur distant 21 peut comprendre un ou plusieurs processeurs 22, 23, 24 aptes à mettre en œuvre différentes étapes d'un procédé de synchronisation selon l'invention. Le serveur distant comprend en outre une ou plusieurs unités de mémoire de stockage 25. Les unités de mémoire de stockage ou mémoire permettent le stockage des données d'enregistrement ainsi que les instructions du programme informatique exécutant des étapes du procédé selon l'invention et artefact résultat du procédé selon l'invention. Le serveur distant 21 peut également transmettre des informations à destination du « cloud » 26 ou

par liaison filaire pour diffusion d'un clip vidéo résultant de la synchronisation des enregistrements selon l'invention. Le serveur distant 21 peut également transmettre le résultat du procédé selon l'invention à des moyens de lecture et de diffusion audio et éventuellement vidéo. Par exemple le résultat du procédé selon l'invention peut être transmis aux téléphones multimédias 2, 3, 5 pour lecture de l'enregistrement final, après traitement par le procédé selon l'invention. Le serveur distant 21 peut également transmettre l'enregistrement synchronisé à des moyens de mixage et de montage. Avantageusement, lesdits moyens de mixage et de montage peuvent s'exécuter sur le serveur distant. Ainsi, le serveur distant peut réaliser les opérations d'alignement, de synchronisation et d'analyse des plages audio de l'ensemble des enregistrements puis la sélection des pistes audio pour le mixage audio et enfin la sélection des pistes vidéo et l'inclusion de l'audio mixé dans le flux vidéo puis le montage final de la vidéo.

La figure 3 représente plusieurs étapes du procédé de synchronisation 30 selon l'invention mis en œuvre par le dispositif 20 de synchronisation selon l'invention.

Une première étape 31 du procédé selon l'invention est une étape d'acquisition de données par les dispositifs d'acquisition de signaux audio et vidéo 2, 3, 4, 5. Les données acquises sont ensuite traitées sous la forme de signaux audio et vidéo. Les données acquises peuvent être enregistrées par exemple dans une mémoire du dispositif d'acquisition de signaux audio et vidéo 2, 3, 4, 5. Ces données sont par la suite nommées enregistrements.

Les enregistrements sont transmis au cours d'une deuxième étape 32 au serveur distant 21 par l'intermédiaire d'un réseau internet par exemple. Avantageusement, les enregistrements peuvent être collectés via un espace de stockage distant 26.

Le serveur distant 21 réalise ensuite différents traitements sur ces enregistrements. Les enregistrements sont, dans un premier temps synchronisés en utilisant les informations audio contenues dans chaque enregistrement.

Tout d'abord les étapes suivantes sont réalisées pour chaque paire d'enregistrements possible parmi l'ensemble des enregistrements. Il est

procédé, dans un premier temps, à un premier alignement préliminaire peu précis des enregistrements. L'alignement préliminaire comprend notamment une troisième étape 33 de détermination d'un premier décalage temporel sur des enveloppes des signaux audio et une quatrième étape 34 d'alignement des enveloppes des signaux audio deux à deux.

Pour chaque signal audio S_i, S_j de chaque enregistrement et pour chaque paire de signaux audio d'enregistrements (S_i, S_j), la troisième étape 33 de détermination d'un premier décalage temporel comprend tout d'abord une étape de calcul d'une enveloppe du signal audio S_i, S_j. Si l'on définit N comme le nombre total d'enregistrements à traiter alors $i = 1,..,N$ et $j = 1,..,N$. Les paires (S_i, S_j) de signaux d'enregistrements sont constituées telles que $i \neq j$. Une enveloppe d'un signal peut être définie comme un signal descriptif du signal initial en ce qu'il en conserve les caractéristiques principales ou tendances. Réaliser les calculs sur une enveloppe de signal plutôt que sur le signal lui-même permet de réduire le volume des calculs réalisés.

Une première méthode de calcul d'une enveloppe $E_i(\tau)$ peut être un calcul d'une enveloppe d'énergie du signal, τ étant une variable temporelle :

$$E_i(\tau) = \sum_{k=0}^{F_S^{-1}} S_i(k + \tau \cdot F_S)^2 \quad (1000)$$

avec F_S une fréquence d'échantillonnage du signal S_i.

On réalise ainsi un sous-échantillonnage, ou downsampling en langage anglo-saxon, afin d'obtenir un signal de travail dans lequel n'apparaissent que les variations les plus importantes, par exemple. Cette première méthode de calcul est très rapide et donne de bons résultats par exemple pour réaliser un alignement de pistes comprenant des enregistrements d'un concert de musique amplifiée de type rock, pour lesquels les débuts et fins de morceaux sont bien marqués et de fait, facilement identifiables.

Une deuxième méthode de calcul possible d'une enveloppe d'un signal peut être une utilisation d'une moyenne mobile calculée sur une version ré-échantillonnée du signal. Ensuite, la moyenne mobile est ré-échantillonnée. Un avantage de cette méthode est l'augmentation de la précision des calculs d'alignement des enveloppes des signaux à venir pour gagner par la suite

en rapidité de calcul notamment pour les calculs de corrélation des signaux proprement dit. En effet, cette deuxième méthode permet d'augmenter le facteur de « downsampling » et donc d'accélérer le calcul à venir de corrélation des signaux tout en maintenant une bonne précision.

De plus, la deuxième méthode donne de bons résultats pour des captations de signaux audio dans lesquels les frontières entre silence et information sont moins marquées.

Soit

$$S'_i(\tau) = resample_1(S_i, q_1) \quad (1001)$$

avec la fonction $resample_1()$ étant une fonction réduisant l'échantillonnage d'un signal S d'un premier facteur d'échantillonnage ou de « downsampling » q_1. Cette fonction préserve avantageusement les caractéristiques principales du signal S en utilisant un filtre passe bas afin de limiter l'effet de repliement du spectre, ou « aliasing ». Par exemple il est possible d'utiliser un filtre de Chebyshev d'ordre n.

On peut choisir q_1 tel que $50 \leq q_1 \leq 200$ et par exemple q_1 de l'ordre de la centaine et n de l'ordre de la dizaine. Il s'agit de trouver un compromis entre la précision de l'alignement et le temps de calcul.

Ensuite, on calcule une moyenne mobile quadratique sur le signal ré-échantillonné :

$$\overline{S'_i}(\tau) = \sqrt{\frac{1}{N}\sum_{k=0}^{N-1}(S'_i(\tau + k) \times W(\omega, rc, k))^2} \quad (1002)$$

dans laquelle $W(\omega, rc, k)$ est une fonction de fenêtrage variant entre 0 et 1, de largeur ω, et de temps de montée caractéristique rc. On peut choisir ω entre 0.01s et 0.5s, avec par exemple une valeur de l'ordre du dixième de seconde. On peut choisir rc entre 0.1ms et 10ms avec par exemple une valeur de l'ordre de 5ms.

On peut ensuite à nouveau ré-échantillonner ce signal $\overline{S'_i}$ pour obtenir une enveloppe du signal s'exprimant ainsi :

$$E_i(\tau') = resample_2(\overline{S'_i}, q_2) \quad (1003)$$

avec q_2 un deuxième facteur d'échantillonnage, ou de « downsampling » qui peut être choisi tel que $50 \leq q_2 \leq 200$, avantageusement $q_2 = q_1$ et $q_2 q_1$ est de l'ordre de 10000. On peut choisir à titre d'exemple q_2 de

l'ordre de la centaine. La fonction $resample_2$ est une autre fonction d'échantillonnage réduisant le signal $\overline{S'_t}$ d'un facteur de ré-échantillonnage q_2.

Après le calcul de la moyenne mobile quadratique (1002) les informations fréquentielles disparaissent du signal et l'utilisation d'un filtre passe bas pour l'anti-aliasing n'est donc pas nécessaire.

D'autres méthodes de sous-échantillonnage peuvent également être utilisées.

Une quatrième étape 34 du procédé de synchronisation 30 selon l'invention est une étape d'alignement des enveloppes des signaux deux à deux.

Les niveaux des signaux enregistrés sont fortement dépendants des conditions d'enregistrement du signal ainsi que des caractéristiques de la source sonore. Par exemple, la distance entre un microphone et la source sonore est l'un des facteurs déterminants dans la puissance reçue par un transducteur sonore et par conséquent dans sa traduction en un signal numérique. La disparité dans les niveaux sonores reçus doit tout d'abord être réduite avant de pouvoir effectuer une comparaison des signaux entre eux. À cette fin, une normalisation des deux enveloppes est réalisée permettant ainsi de comparer des signaux avec des amplitudes similaires. Un objectif de cette normalisation est de permettre de distinguer des phases dites de silence, des phases d'émission de la source. Par définition une plage de silence représente les moments pendant laquelle la source que l'on souhaite enregistrer n'émet pas. Par contre pendant ces plages de silence, il n'est pas garanti qu'il n'y ait pas de signal, c'est-à-dire que le niveau de puissance du signal enregistré soit nul. Le signal contient alors uniquement les sons issus de l'environnement de la source sonore dont il convient de quantifier le niveau en définissant deux paramètres S_{min}, un niveau de puissance minimum, et S_{max}, un niveau de puissance maximum. Lorsque le niveau d'énergie du signal est détecté entre S_{min} et S_{max} alors on considère que l'on est en présence d'une plage de silence de la source. Par la suite on utilise simplement l'expression « plage de silence ».

Afin de détecter les plages de silence on réalise alors un partitionnement des niveaux du signal en deux sous-ensembles par la méthode dite des K-moyennes. On choisit ensuite la partition P_μ dont le barycentre μ a la valeur la plus faible. Pour cette partition on calcule une nouvelle enveloppe $E'_i(\tau')$ tel que :

$$\forall\, \tau' \in P_\mu \quad E'_i(\tau') = K.E_i(\tau')/\mu \qquad (1004)$$

avec K une constante d'ajustement de la normalisation. Par exemple on peut utiliser $K = 4/30$.

Cette opération (1004) permet de recentrer les points de l'enveloppe faisant partie de périodes de silence, autour d'une valeur commune.

Les figures 4a et 4b donnent un exemple de l'application des précédentes étapes du procédé selon l'invention. La figure 4a représente un premier signal audio 40 en entrée du procédé selon l'invention. Le niveau du premier signal 40 évolue entre -0.002 et $+0.002$ sur une échelle de temps échantillonnée. La figure 4b représente une première enveloppe 41 du premier signal 40 représenté sur la figure 4a. La première enveloppe 41 obtenue par moyenne mobile normalisée est comprise entre 0 et 1 sur la même échelle de temps que celle de la figure 4a. Sur la figure 4b, on peut par exemple déterminer des périodes de silence entre $S_{min} = 0$ et $S_{max} = 0{,}2$.

Il est ainsi possible de calculer une corrélation croisée entre deux vecteurs comprenant chacun une enveloppe du signal S_i, S_j telle que précédemment calculée. À cette fin le procédé selon l'invention réalise une recherche du maximum de corrélation $\tau'_{ij\,max}$ tel que $\forall\tau, \Gamma_{ij}(\tau') \leq \Gamma_{ij}(\tau'_{ij\,max})$ (1005)

Les figures 5a, 5b, 5c illustrent la détection du maximum de corrélation entre deux enveloppes 50, 51 de deux signaux audio, respectivement représentées sur les figures 5a et 5b. La figure 5c représente sous la forme d'une courbe 52 le résultat de la corrélation entre les deux enveloppes 50, 51. Le maximum de corrélation est visible sous la forme d'un pic 53 sur la courbe 52. Pour exemple une deuxième enveloppe 50 d'un deuxième signal est représentée sur la figure 5a. La figure 5b représente quant à elle la première enveloppe 51 du premier signal 40 avec un niveau de zoom différent de la première enveloppe 41 représentée sur la figure 4b.

Le maximum de corrélation 53 permet de déterminer à quel instant les deux enveloppes des signaux sont les plus similaires. L'instant du maximum de corrélation 53 signifie qu'en première approximation les première et deuxième enveloppes 51, 50 coïncident. Ainsi il est possible d'en déduire que les deux signaux correspondant aux première et deuxième enveloppes 51, 50 représentent des sources sonores ayant émis approximativement durant les mêmes périodes.

Ainsi on peut obtenir un premier décalage temporel δ'_{ij} avec une précision correspondant à la fréquence d'échantillonnage Fs et au produit des premier et deuxième coefficients de « downsampling » tel que :

$$\delta'_{ij} = (\tau'_{ij\,max} - max(|E_i|, |E_j|) \quad (1006)$$

avec $\delta_{ij} = q_1 . q_2 . \delta'_{ij}$ (1007)

La précision obtenue, de l'ordre de la seconde, n'est pas encore suffisante pour déclencher des traitements par exemple de mixage sur les signaux audio. Ce premier alignement permet néanmoins d'effectuer ultérieurement un deuxième alignement plus précis avec des données pré-alignées ce qui diminue de manière importante le temps de calcul nécessaire pour l'alignement précis.

Les méthodes de détermination d'enveloppes du signal ainsi que la méthode de normalisation précédemment décrites, sont données uniquement à titre d'exemple. D'autres méthodes peuvent également être utilisées dans le cadre de l'invention sans en remettre en cause le principe.

Une cinquième étape 35 du procédé selon l'invention est une étape de segmentation de chacune des enveloppes des signaux précédemment calculées. L'étape de segmentation emploie une méthode de détection des plages de silence dans l'enveloppe de chaque signal.

Les enveloppes des deux signaux sont combinées en un vecteur de même dimension afin de rechercher les périodes de silence relatives des deux signaux.

Par exemple il est possible d'utiliser une fonction donnant un maximum comme fonction de combinaison des enveloppes telle que :

$$\tau \rightarrow E_{ij}(\tau) = max(E_j(\tau + \delta'_{ij}), E_i(\tau)) \quad (1008)$$

sur un intervalle pour lequel les deux signaux sont définis.

Ainsi à partir du vecteur combinaison des deux enveloppes, sont obtenues les intervalles où le signal évolue entre les deux seuils S_{min} et S_{max}. Ces intervalles représentent des périodes au cours desquelles aucune des sources sonores que l'on cherche à capter ne produit de son significatif. En d'autres termes, ces intervalles reflètent simultanément l'environnement sonore ambient.

Alternativement, il est possible d'utiliser de la même manière une fonction donnant le minimum comme fonction de combinaison des enveloppes sur chaque intervalle pour lequel les deux signaux sont définis. Ainsi entre deux seuils S_{min} et S_{max} on peut détecter des périodes pendant lesquelles au moins une source n'émet pas. Ainsi les périodes de silence seront relatives à l'au moins une source n'émettant pas.

L'ensemble A de ces intervalles I est déterminé de la manière suivante :

$$A = \{I, \forall \tau \in I, s_{min} \leq E_{ij}(\tau) \leq s_{max}\} \quad (1009)$$

Par exemple, on peut utiliser s_{min} proche de 0 et s_{max} de l'ordre de 1.10^{-1}.

Les figures 6a, 6b et 6c illustrent les traitements réalisés pour la cinquième étape 35. La figure 6a représente la deuxième enveloppe 60 du deuxième signal audio et la figure 6b représente la première enveloppe 61 du premier signal 40. Les première et deuxième enveloppes 61, 60 telles que représentées sur les figures 6a et 6b sont tronquées en début et fin, par rapport aux première et deuxième enveloppes 51, 50 représentées sur les figures 5a et 5b. La troncature des première et deuxième enveloppes 51, 50 résulte de l'alignement initial réalisé au cours de la quatrième étape 34. L'objectif de cette troncature est de pouvoir travailler sur des enveloppes de même longueur, sur la base de l'alignement initial afin de pouvoir calculer une troisième enveloppe combinée 62 telle que représentée sur la figure 6C. C'est sur cette troisième enveloppe combinée 62 que les plages de silence sont recherchées.

La figure 6C représente donc la troisième enveloppe 62 résultant de la combinaison des première et deuxième enveloppes 61, 60 des premier et deuxième signaux. Sur la figure 6c la valeur s_{max} est représentée par la ligne

horizontale 63. On peut donc identifier l'ensemble S des intervalles I 64, 65, 66, 67, 68, 69 pour lesquels la troisième enveloppe 62 reste sous le seuil $s_{max} = 1.10^{-1}$ par exemple. Dans l'exemple représenté sur la figure 6c, on obtient six intervalles 64, 65, 66, 67, 68, 69 de silence.

L'ensemble des intervalles ou plages de silence A peut ensuite être filtré afin d'en extraire les plages de silence permettant un calcul d'alignement précis.

Pour filtrer les intervalles, il est possible d'utiliser par exemple les règles suivantes :

- en déterminant un écart minimal entre deux plages de silence consécutives, lorsque deux des plages de silence I se suivent avec un écart temporel assez faible, on peut réaliser une concaténation desdites plages de silence. Par exemple : si $I = [a;\ b] \in A$ et $J = [c\ ;\ d] \in A$, avec $b < c$, alors si $\mathrm{c} - \mathrm{b} \leq \varepsilon$, avec ε une tolérance de proximité fixée, alors on peut remplacer I et J par $IJ = [a\ ;\ d]$ dans A. On peut par exemple utiliser $\varepsilon = 1s$ rapporté au temps échantillonné de l'enveloppe.
- toute plage de silence I_k dont la durée est inférieure à une durée minimale paramétrable est exclue de l'ensemble des plages de silence A. Par exemple $\forall\, I \in A$, si $|I| \leq d_{min}$ avec d_{min} la durée minimale d'une plage de silence, alors I_k est exclu de A. Il est par exemple possible d'utiliser $d_{min} = 6s$ ramené au temps échantillonné de l'enveloppe.
- il est également possible de paramétrer une durée maximale d'une plage de silence afin de ne pas avoir un nombre de calculs trop important, ce qui pourrait advenir en présence de plages de silence trop longues. Par exemple, en prenant $I = [a;\ b] \in S$, si $|I| \geq d_{max}$, où d_{max} est la durée maximale d'une plage de silence, alors la plage peut être exclue de S. Par exemple on peut poser $d_{max} = 120s$ ramené au temps échantillonné de l'enveloppe.

Ensuite, on réalise un changement de référentiel temporel pour passer de la première échelle de temps τ à une deuxième échelle de temps t, correspondant au temps réel. Par exemple on peut remplacer $I = [a\ ;\ b\ ]$ par

$$I_i = [u_i = q_1.q_2.a\ ;\ v_i = q_1.q_2.b] \quad (1010)$$

$$I_j = \left[u_j = q_1.q_2.a + \delta_{ij};\ v_j = q_1.q_2.b + \delta_{ij}\right] \quad (1011)$$

Chaque signal est ainsi divisé en plusieurs segments dont des plages de silence relatif.

Ensuite on procède à un alignement plus précis des signaux deux à deux. À cette fin, une sixième étape 36 du procédé selon l'invention est une étape de calcul de décalages temporel entre les segments des signaux deux à deux. L'alignement plus précis est réalisé sur chacun des intervalles I_k du signal combinaison des enveloppes des signaux, déterminés au cours de la cinquième étape 35. Pour chaque intervalle I_k du signal combinaison des enveloppes, on sélectionne les intervalles $I_{i,k}$ et $I_{j,k}$ correspondants dans chacun des signaux réels S_i, S_j.

Un premier spectrogramme 70 de $I_{i,k}$, tel que représenté sur la figure 7, est ensuite calculé de la manière suivante : le signal réel S_i sur l'intervalle de silence $I_{i,k}$ est découpé en segments de longueur n, les segments consécutifs se recouvrant. Ensuite on applique sur ces segments une fonction de fenêtrage permettant de transformer le signal dans le domaine fréquentiel par une transformée de Fourier. On obtient ainsi le spectrogramme :

$$SP_{i,k} = specgram(I_{i,k}, n, F_S, window, overlap) \quad (1012)$$

dans lequel $window$ est la taille de la fenêtre temporelle de la fonction de fenêtrage, $overlap$ est la taille du recouvrement entre les segments, n est la résolution de la transformée de Fourier, et $specgram$ est une fonction de calcul du spectrogramme.

Par exemple, on peut utiliser les gammes de valeurs suivantes : n est une puissance de deux pouvant varier de 2^4 à 2^{15}, $window$ peut être comprise entre 500 et 5000, et $overlap$ est fonction de $window$ et peut par exemple varier entre 50% et 98% de $window$.

Ensuite, on étudie la partie réelle des signaux transformés : $R_{i,k} = \Re(SP_{i,k})$ et $R_{j,k} = \Re(SP_{j,k})$.

Une corrélation croisée est ensuite réalisée sur les deux signaux pour chaque bande de fréquence f du spectrogramme. Ensuite en sommant les

corrélations croisées, on obtient une corrélation spectrale des signaux $xcorr_{ij,k}$:

$$xcorr_{ij,k} = \sum_f xcorr(R_{i,k}(f), R_{j,k}(f)) \qquad (1013)$$

Le maximum de la fonction de corrélation croisée indique un instant $\tau_{ij,k\,max}$ où deux signaux présentent une similitude maximale, permettant ainsi de déterminer un alignement temporel précis et local à l'intervalle de silence relatif.

Les figures 8a, 8b et 8c représentent un exemple des traitements réalisés à partir des spectrogrammes. La figure 8a représente la partie réelle d'un deuxième spectrogramme $R_{i,k} = \Re(SP_{i,k})$ et la figure 8b représente la partie réelle d'un troisième spectrogramme $R_{j,k} = \Re(SP_{j,k})$. La figure 8c représente la fonction de corrélation spectrale $xcorr_{ij,k}$ des parties réelles des deuxième et troisième spectrogrammes $R_{i,k}$ et $R_{j,k}$. Ainsi on peut observer un maximum de corrélation 80 au centre de la fonction de corrélation spectrale $xcorr_{ij,k}$.

La transformée de Fourier réduisant naturellement la résolution du signal, il est ainsi possible de localiser le maximum de corrélation croisée en réalisant une interpolation polynomiale quadratique.

On obtient ainsi sur chaque intervalle I_k une valeur de décalage $\delta_{ij,k}$ telle que :

$$\delta_{ij,k} = \delta_{ij} + \tau_{ij,k\,max} - |I_{i,k}| \quad (1014)$$

Chaque valeur de décalage peut être positive ou négative.

Ces calculs sont réalisés sur l'ensemble des intervalles I_k de silence. Ces calculs permettent de déterminer un ensemble de valeurs de décalages temporels, relatifs au premier décalage préliminaire calculé au cours de la troisième étape 33 du procédé selon l'invention.

Une septième étape 37 est une étape de positionnement des valeurs des décalages temporels obtenus au cours de la sixième étape 36, sur une échelle temporelle commune dite de référence.

L'échelle de temps de référence est choisie arbitrairement sur le signal S_i. L'ensemble des décalages temporels calculés sont positionnés au centre des intervalles $I_{i,k}$. On obtient un premier ensemble de décalages du

premier signal S_i par rapport au deuxième signal S_j : $\Delta_{ij} = \left\{(\frac{u_{i,k} + v_{i,k}}{2};\ \delta_{ij,k})\right\}$ et de la même manière, un deuxième ensemble de décalages du deuxième signal S_j par rapport au premier signal S_i.

La méthode de la corrélation croisée décrite ici est l'une parmi d'autres méthodes qui peuvent être utilisées pour déterminer un instant de similitude maximale entre deux signaux. Tout autre méthode notamment la corrélation de signaux réels ou filtrés ou encore le calcul d'empreintes acoustiques ou fréquentielles peuvent également être utilisées.

Les calculs de corrélation sur signaux étant par nature imprécis, il arrive que localement un résultat soit erroné. Par exemple un des décalages peut avoir une valeur dite aberrante. Afin de ne pas introduire d'erreur dans la synchronisation, une étape de filtrage est mise en œuvre après que tous les calculs de corrélations sur les plages de silence piste à piste aient été réalisés.

Pour cette étape de filtrage, on utilise une modélisation des décalages entre deux segments par une approche linéaire, ou régression linéaire, avec possibilité de discontinuité. Les discontinuités modélisent les éventuelles coupures du signal et les linéarités représentent le fait que sur les échelles de temps considérées, c'est-à-dire de l'ordre de l'heure, les rapports de fréquence d'échantillonnage sont globalement constants. En d'autres termes, pour deux signaux S_i et S_j un changement de référentiel temporel peut être donné par l'expression :

$$\forall t, t_j = k.t_i + \lambda(t_i) \quad (1015)$$

dans laquelle $\lambda(t)$ est une fonction constante par morceaux sur les intervalles de temps où les signaux ne présentent pas de discontinuité.

Pour détecter les points aberrants parmi les valeurs de décalage, on étudie les rapports r, tels que :

$$\forall\ (\delta_m\ , \delta_n) \in P_2(\Delta_{ij})\ r_{mn} = \frac{\delta_n - \delta_m}{t_n - t_m} \quad (1016)$$

où t_m et t_n sont les centres des intervalles I_m et I_n dans le référentiel temporel du signal de référence S_i.

Ensuite, on utilise un algorithme de partitionnement pour regrouper les rapports r_{mn} dans des partitions de rapports similaires et ainsi en écarter les valeurs aberrantes.

Un exemple d'algorithme couramment utilisé est l'algorithme DBSCAN pour Density-Based Spatial Clustering of Applications with Noise proposé par Martin Ester, et al. Il est également possible d'utiliser une régression linéaire.

L'algorithme DBSCAN peut par exemple être paramétré de la manière suivante :

- un nombre de points minimum $minpts$ peut être de l'ordre de la cardinalité de Δ_{ij} : $minpts = \frac{Card(\Delta_{ij})}{2} + 1$;
- une distance minimale d_{min} entre les points peut être de l'ordre de $100/3600.10^{-3}$, ce qui correspond à un ratio r_0 de 100 millisecondes par heure ; une valeur de 50 à 100 millisecondes peut également convenir.

Une fois le partitionnement réalisé, on élimine de l'ensemble Δ_{ij} l'ensemble des mesures m pour lesquelles aucune partition ne contient de rapport r_{mn}. Autrement dit, toute mesure m pour laquelle au moins un des rapports r_{mn} est inclus dans une partition est incluse dans l'ensemble Δ_{ij}, toutes les autres mesures en sont exclues.

Avantageusement l'utilisation de l'étape de filtrage utilisant la modélisation des décalages t_j telle que définie par l'équation (1015) permet de gérer les problématiques de « buffer overrun » et « underrun ». Alternativement, tout algorithme de filtrage utilisant le modèle décrit dans l'équation (1015) permet également de résoudre ces problématiques.

La figure 9 montre un exemple de filtrage d'un point aberrant en utilisant un algorithme de partitionnement de données. On obtient une première partition 90 comprenant une majorité de points représentant des valeurs de décalage et une deuxième partition 91 contenant une minorité de points représentant des valeurs de décalage. Sur la figure 9, la première partition 90 comprend une douzaine de points alors que la deuxième partition 91 ne contient qu'un seul point. On peut en déduire que la valeur de décalage représentée par le point de la deuxième partition 91 est une valeur

aberrante et quelle peut être supprimée de l'ensemble des valeurs de décalage.

Alternativement, la régression linéaire fonctionne également directement sur les décalages.

Le filtrage peut être réalisé de plusieurs autres manières différentes bien connues de l'état de la technique. Par exemple, il est possible de définir simplement un intervalle de valeurs acceptables, en rejetant des décalages supérieurs à une seconde par exemple. Un autre exemple est de calculer un écart type, ou sigma, des mesures, et d'exclure les mesures s'écartant trop de cet écart-type, par exemple plus de 500ms de distance avec l'écart-type.

Une fois l'ensemble de ces calculs réalisés pour l'ensemble des paires possibles de signaux, une huitième étape 38 est une étape de calcul de données d'alignement pour chaque signal par rapport à l'un des signaux pris comme signal de référence. À cette fin, l'ensemble des décalages précédemment calculés sont combinés entre eux pour calculer les données d'alignement de chaque signal par rapport au signal de référence.

Pour chaque signal S_i, on considère l'ensemble des mesures de décalage relativement aux autres signaux $\{S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_N\}$. On considère donc : $\mathrm{U}_{j=1, j\neq i}^{N} \Delta_{ij}$.

Pour chaque ensemble de mesures relatives Δ_{ij} on construit une fonction d'interpolation linéaire $\delta_{ij}(t)$ permettant de calculer le décalage du deuxième signal S_j par rapport au premier signal S_i, en tout point de S_j. Cette fonction d'interpolation s'appuie sur l'ensemble des mesures de décalages Δ_{ij}.

On peut nommer T_i l'ensemble des mesures de décalage entre le premier signal S_i et les autres signaux $\{S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_N\}$.

En tout point t de T_i, on modélise un ajustement des décalages temporels interpolés par la fonction $\delta_{ij}(t)$ en utilisant un système de ressorts de raideur $k = 1$ à l'équilibre, de longueur au repos $\delta_{ij}(t)$, et de longueur à l'équilibre $x_{ij}(t)$. Chaque ressort du système exerce donc une force $F_{ij}(t)$ telle que $F_{ij}(t) = x_{ij}(t) - \delta_{ij}(t)$. Pour calculer un décalage corrigé, on recherche l'ensemble des longueurs des ressorts à l'équilibre soit :

$$\sum_{j=1,j\neq i}^{N} F_{ij}(t) = 0 \qquad (1017)$$

On a donc

$$\sum_{j=1,j\neq i}^{N} x_{ij}(t) = \sum_{j=1,j\neq i}^{N} \delta_{ij}(t) \qquad (1018)$$

D'autre part, à l'équilibre, les valeurs orientées de décalage vérifient les relations :

$$\forall\, k \in [1\,;\, N] \setminus \{i,j\},\ x_{ij}(t) + x_{ik}(t) + x_{kj}(t) = 0 \qquad (1019)$$

Les équations (1018) et (1019) permettent de construire une matrice carrée A_i des coefficients d'un système linéaire de dimension $\frac{N\times(N-1)}{2}$.

Il s'agit alors de résoudre en tout point $t \in T_i$ un système d'équations linéaires de la forme :

$$A_i X_i(t) = b_i(t) \qquad (1020)$$

Les valeurs de décalage ajustées en t sont les valeurs de $X_i(t)$ déterminées par la résolution du système d'équations (1020) pour le signal S_i.

Ensuite, pour permettre des traitements audio, il faut choisir un signal de référence S_r, de façon arbitraire dans l'ensemble des signaux, afin d'exprimer sur une échelle de temps commune tous les autres signaux.

On pose $t' = t + \delta_{kr}(t)$ où $\delta_{kr}(t)$ est une fonction d'interpolation construite avec les solutions $X_k(t)$ permettant de connaître le décalage entre le signal de référence S_r et un autre signal S_k en tout point de S_k.

Ce traitement est réalisé pour tous les signaux et permet de construire les données d'alignement des signaux.

Une neuvième étape 39 est une étape d'alignement des enregistrements par un alignement des signaux audio par rapport au signal audio de référence S_r.

Pour les signaux S_r et S_k l'alignement consiste à faire coïncider à l'instant t les valeurs $S_r(t)$ et $S_k(t')$, et ce, pour tous les signaux audios.

Une dixième étape 300 est une étape de calcul des coefficients de lecture des enregistrements par segment. Par exemple, il est possible de calculer un coefficient de vitesse de lecture d'un segment $[a\,;\, b]$ du signal S_i par la relation :

$$p = 1 + \frac{\delta_{ir}(b) - \delta_{ir}(a)}{b - a} \qquad (1021)$$

dans laquelle a et b sont deux valeurs consécutives de l'ensemble des mesures de décalage T_i.

Ainsi, les enregistrements sont synchronisés en utilisant les informations audio contenues dans chaque enregistrement. Les informations de synchronisation peuvent être utilisées pour synchroniser le contenu audio et/ou vidéo de l'enregistrement.

Une onzième étape 301 peut être une étape éventuelle de transmission des enregistrements alignés avec leurs vitesses de lecture par segment, à un dispositif de lecture, ou de restitution, d'enregistrements. Alternativement, les enregistrements alignés et leurs vitesses de lecture peuvent être transmis à un dispositif de mixage et de montage audio et vidéo.

Une douzième étape 302 peut être une étape de lecture simultanée et synchronisée de l'ensemble des enregistrements.

Avantageusement, le procédé 30 selon l'invention permet un véritable enregistrement multipiste. Chaque dispositif d'enregistrement peut a priori se concentrer sur une captation efficace d'une unique source sonore, par exemple par le biais d'un microphone unidirectionnel. De tels microphones présentent l'avantage de ne capter que très peu de l'environnement sonore de la source à enregistrer ce qui permet un mixage de meilleure qualité par la suite. C'est pourquoi deux pistes audio différentes ne vont présenter que très peu de similarités. L'une des spécificités de la méthode présentée ici est qu'elle recherche spécifiquement les zones où l'information est présente mais faible : la source sonore n'émet pas mais on capte du bruit en provenance de l'environnement, à l'exclusion des autres zones qui ne peuvent être corrélées.

Les différents modes de réalisation de la présente invention comprennent diverses étapes. Ces étapes peuvent être mises en œuvre par des instructions d'une machine exécutable au moyen d'un microprocesseur par exemple.

Alternativement, ces étapes peuvent être réalisées par des circuits intégrés spécifiques comprenant une logique câblée pour exécuter les étapes, ou par toute combinaison de composants programmables et composants personnalisés.

La présente invention peut également être fournie sous forme d'un produit programme d'ordinateur qui peut comprendre un support mémoire informatique non-transitoire contenant des instructions exécutables sur une machine informatique, ces instructions pouvant être utilisées pour programmer un ordinateur (ou tout autre dispositif électronique) pour exécuter les procédés.

REVENDICATIONS

1. Procédé de synchronisation (30) d'au moins deux enregistrements, chaque enregistrement comprenant un signal audio, lesdits enregistrements provenant de différents dispositifs d'acquisition (2, 3, 4, 5) co-localisés, ledit procédé étant caractérisé en ce qu'il comprend au moins les étapes suivantes :

a) acquisition et enregistrements (31) de signaux audio par les dispositifs d'acquisition ;

b) transmission (32) des enregistrements par les dispositifs d'acquisition à un serveur de calcul ;

c) pour chaque paire d'enregistrements, les traitements suivants sont réalisés par le serveur de calcul :

- détermination (33) d'un premier décalage temporel entre les deux signaux audio à partir d'une enveloppe de chaque signal audio ;
- alignement (34) des enveloppes des signaux audio en fonction du premier décalage temporel ;
- segmentation (35) de chaque enveloppe de signal audio par détection de plages de silence sur lesdites enveloppes de chaque signal audio ;
- calcul (36) des décalages temporels entre les segments correspondant de chaque signal audio ;
- définition d'une échelle temporelle de référence et positionnement (37) de l'ensemble des décalages temporels entre chaque segment de chaque signal audio sur l'échelle temporelle de référence ;

d) calcul de données d'alignement (38) à partir d'un positionnement de l'ensemble des décalages temporels sur l'échelle temporelle de référence, pour chaque enregistrement, par le serveur de calcul ;

e) alignement (39) des enregistrements par rapport à un signal audio de référence par le serveur de calcul à partir des données d'alignement ;

f) calcul (300) de coefficients de vitesse de lecture pour chaque segment de chaque enregistrement par le serveur de calcul ;

g) transmission (301) des enregistrements alignés, avec leurs coefficients de vitesse de lecture, à un dispositif de restitution d'enregistrements ;

h) lecture simultanée et synchronisée (302) de l'ensemble des enregistrements par un dispositif de restitution d'enregistrements.

2. Procédé selon la revendication 1, caractérisé en ce qu'une plage de silence est définie par une énergie du signal comprise entre une valeur minimum et une valeur maximum prédéfinies.

3. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que l'enveloppe est une enveloppe d'énergie du signal audio.

4. Procédé selon l'une quelconque des revendications 1 et 2, caractérisé en ce que l'enveloppe est une moyenne mobile calculée sur le signal audio réechantillonné.

5. Procédé selon l'une quelconque des revendications 3 et 4, caractérisé en ce que les enveloppes sont normalisées et alignées en réalisant une corrélation croisée pour déterminer le premier décalage temporel.

6. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il comprend une étape de filtrage des plages de silence détectées.

7. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il comprend une étape d'élimination de valeurs aberrantes parmi l'ensemble des décalages temporels calculés entre chaque segment de chacun des deux signaux audio.

8. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que le calcul des données d'alignement comprend les étapes suivantes :

- construction d'une fonction d'interpolation des décalages temporels d'un signal par rapport aux décalages temporels d'un autre signal ;
- modélisation d'un ajustement des décalages par un système de ressort linéaire ;
- construction d'un système d'équations linéaires pour calculer les valeurs de décalages ajustées constituant les données d'alignements.

9. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il s'applique à des enregistrements comprenant en outre un signal vidéo.

10. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il comprend des étapes de mixage et de montage des enregistrements par le dispositif de restitution d'enregistrements avant la lecture desdits enregistrements.

11. Dispositif de synchronisation d'enregistrements, caractérisé en ce qu'il comprend :

- des dispositifs d'acquisition (2, 3, 4, 5) de signaux audio co-localisés réalisant les opérations d'acquérir, d'enregistrer et de transmettre des signaux audio et vidéo ;
- un serveur de calcul (20) mettant en œuvre les étapes c), d), e), f), du procédé selon la revendication 1, ledit serveur de calcul comprenant à cette fin au moins un processeur de calcul (22, 23, 24), une mémoire (25), des moyens de réception et de transmission de signaux ;
- un dispositif de restitution d'enregistrements comprenant des moyens de mixage, de montage et de lecture des enregistrements.

12. Dispositif selon la revendication 11, caractérisé en ce que les dispositifs d'acquisition sont disposées chacun à proximité d'une parmi des sources sonores (6, 7, 8, 9).

13. Dispositif selon l'une quelconque des revendications 11 et 12, caractérisé en ce que les dispositifs d'acquisition (2, 3, 5) de signaux audio sont adaptés à une acquisition de signaux vidéo.

14. Produit programme d'ordinateur comprenant des instructions qui, lorsque le programme est exécuté par au moins un processeur (22, 23, 24) du serveur de calcul (20), conduisent celui-ci à mettre en œuvre les étapes c), d), e), f) du procédé selon la revendication 1.

15. Support d'enregistrement lisible par ordinateur comprenant des instructions qui, lorsqu'elles sont exécutées par au moins un processeur d'un serveur de calcul, conduisent celui-ci à mettre en œuvre les étapes c), d), e), f) du procédé selon la revendication 1.

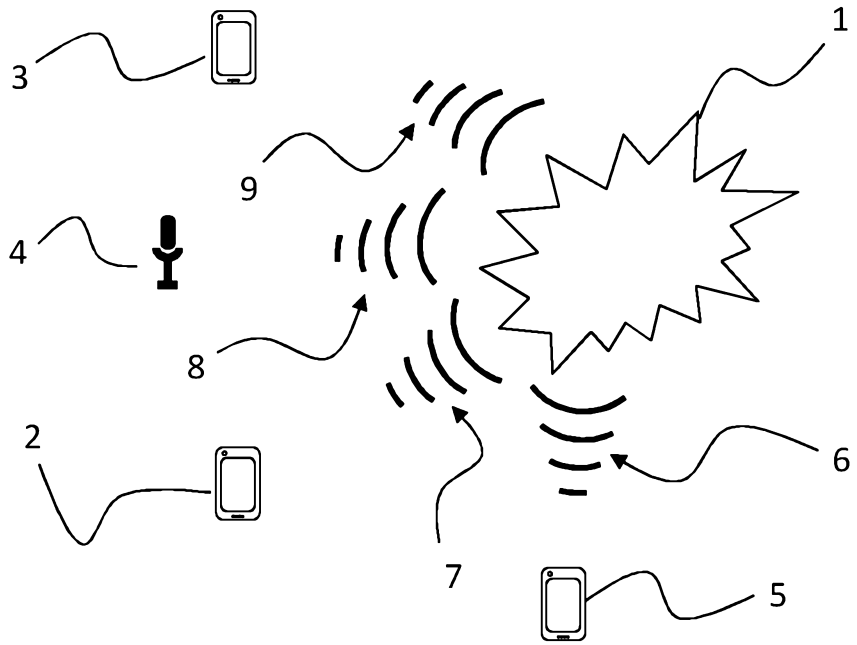


FIG 1

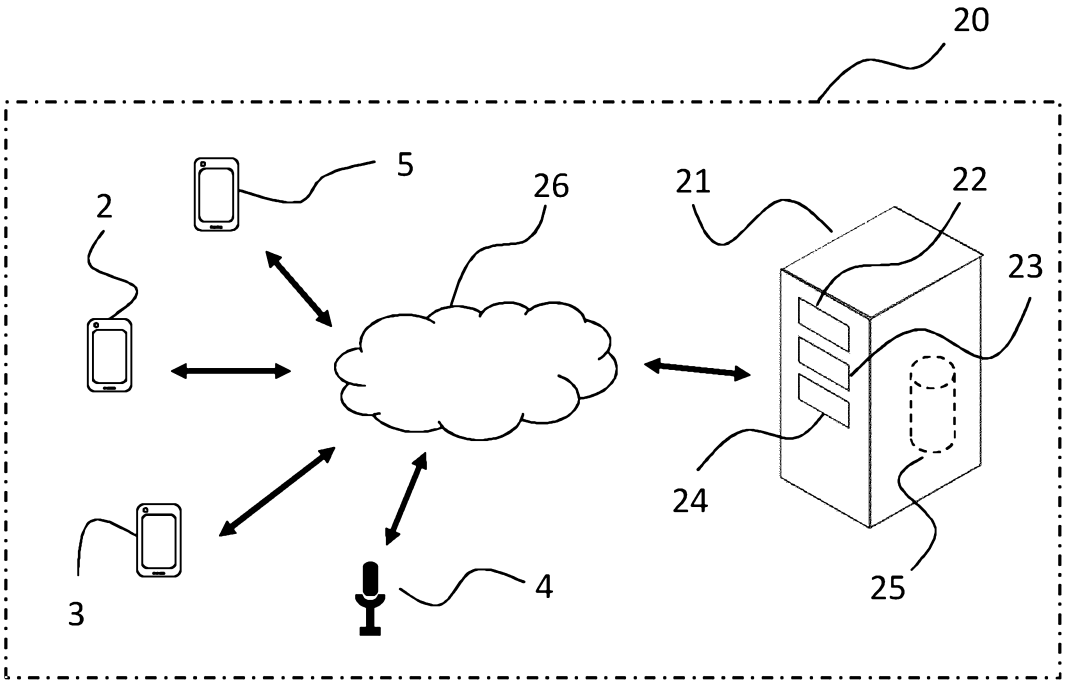


FIG 2

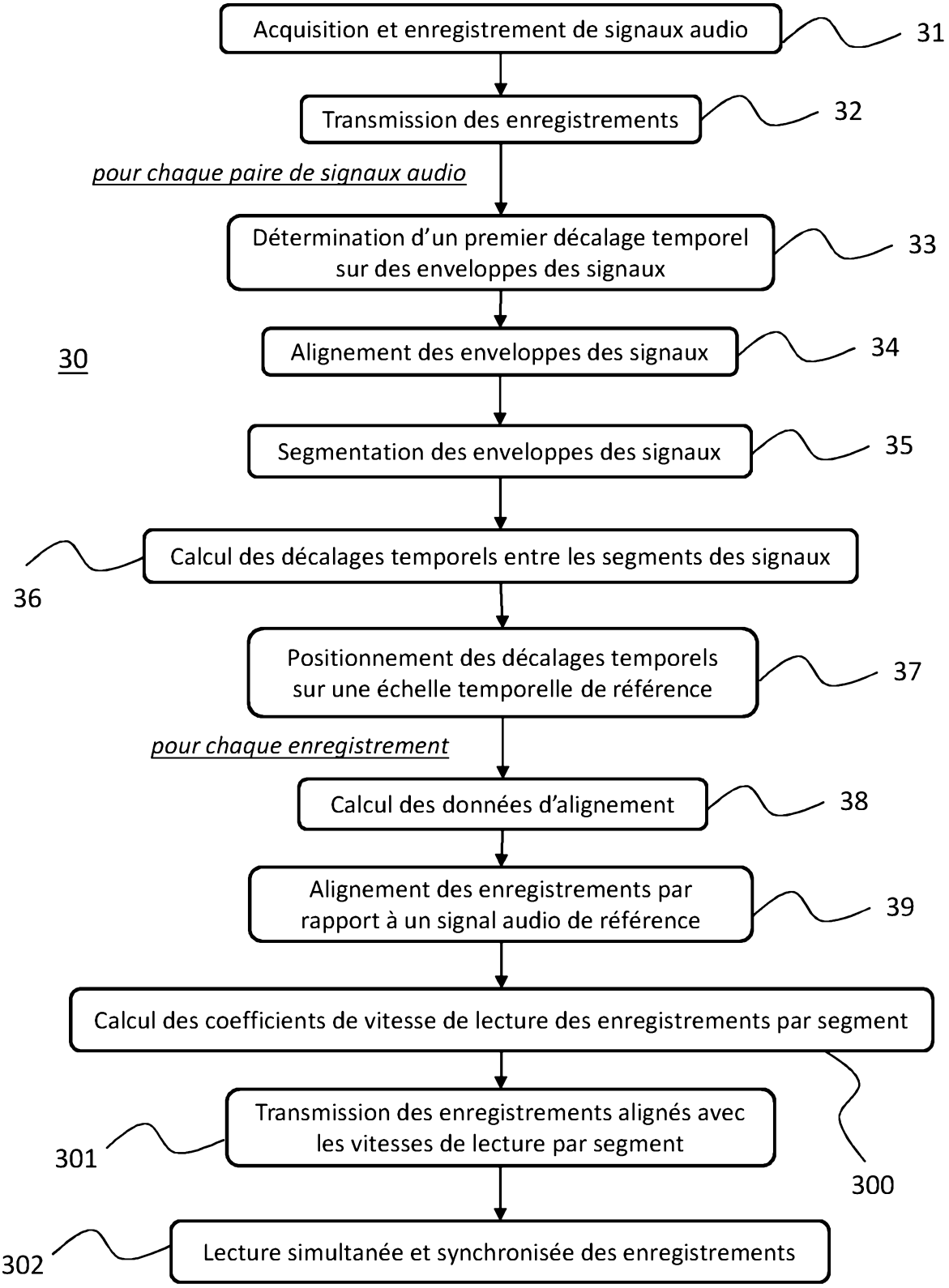

Acquisition et enregistrement de signaux audio
31
Transmission des enregistrements
32
pour chaque paire de signaux audio
Détermination d'un premier décalage temporel sur des enveloppes des signaux
33
30
Alignement des enveloppes des signaux
34
Segmentation des enveloppes des signaux
35
Calcul des décalages temporels entre les segments des signaux
36
Positionnement des décalages temporels sur une échelle temporelle de référence
37
pour chaque enregistrement
Calcul des données d'alignement
38
Alignement des enregistrements par rapport à un signal audio de référence
39
Calcul des coefficients de vitesse de lecture des enregistrements par segment
Transmission des enregistrements alignés avec les vitesses de lecture par segment
301
300
Lecture simultanée et synchronisée des enregistrements
302

FIG 3

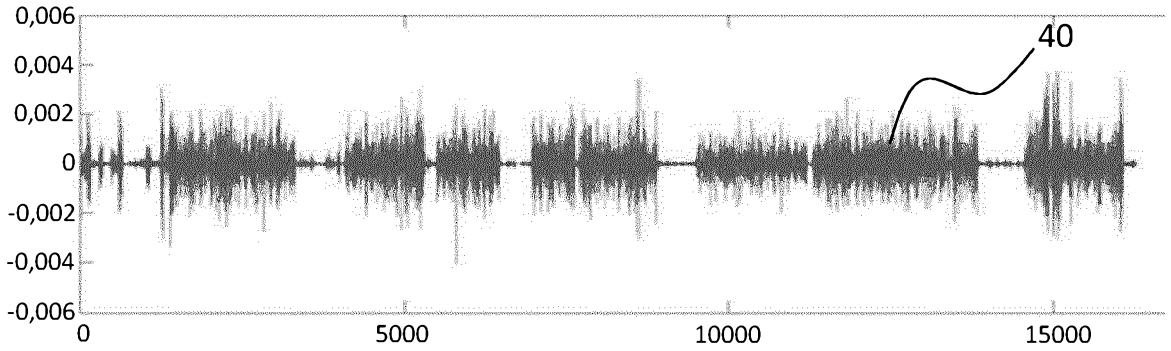


FIG 4a

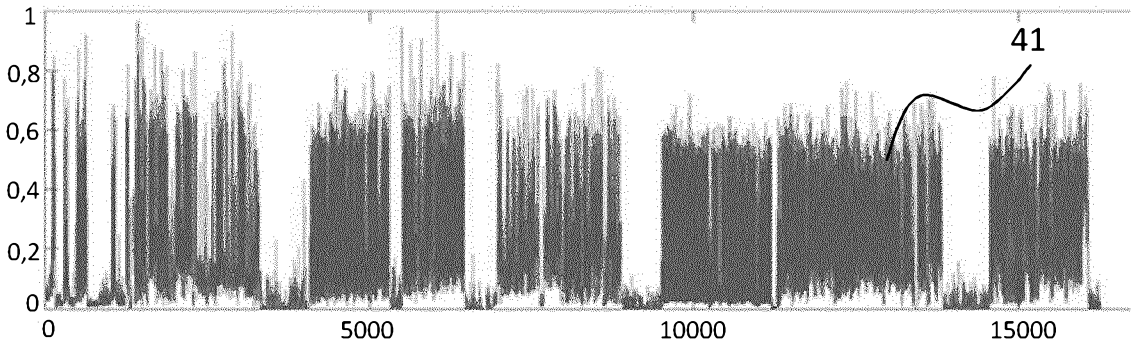


FIG 4b

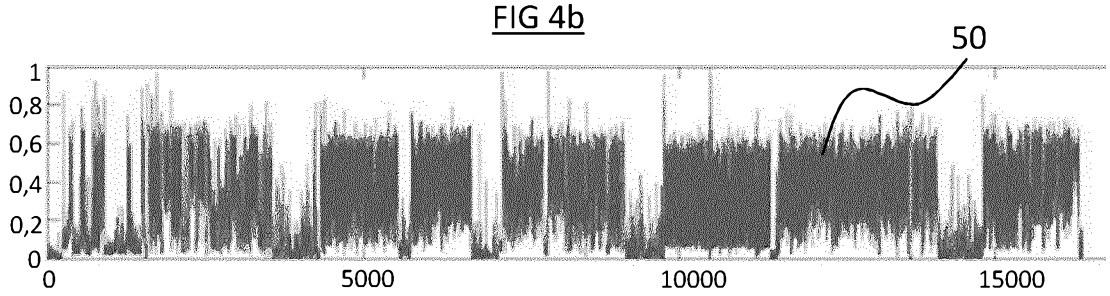


FIG 5a

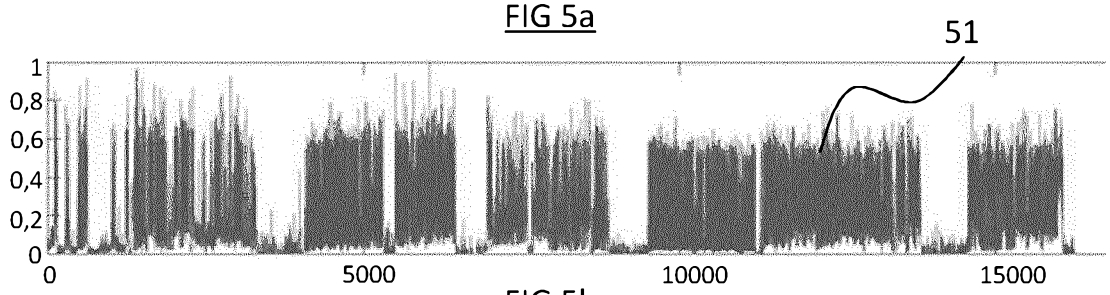


FIG 5b

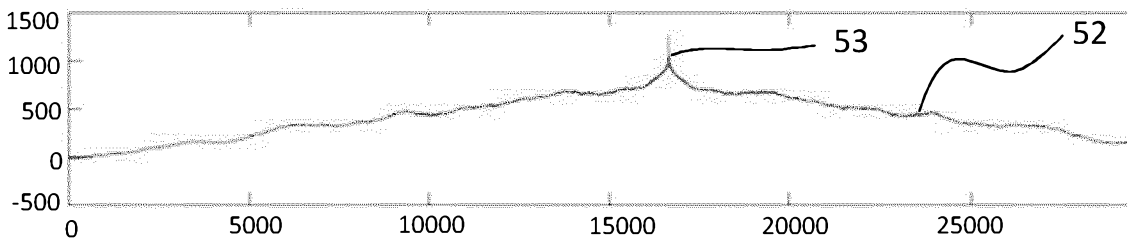


FIG 5c

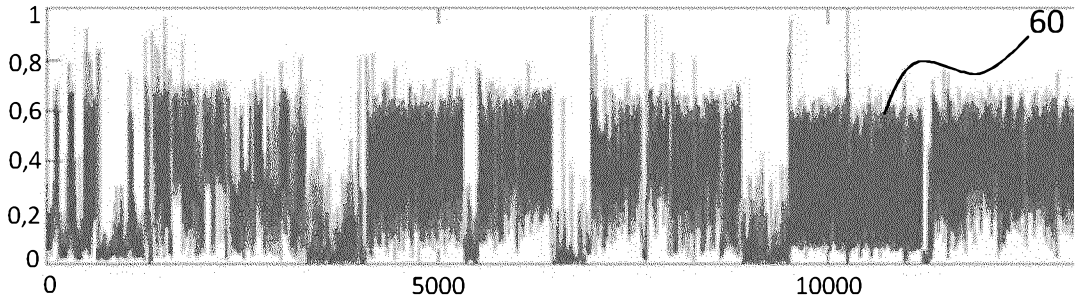


FIG 6a

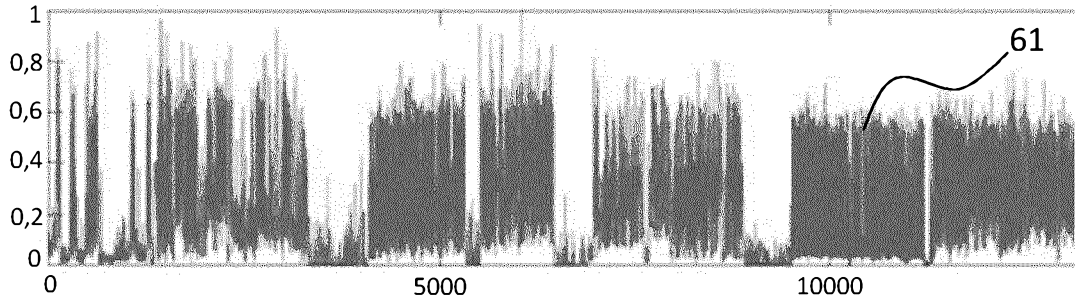


FIG 6b

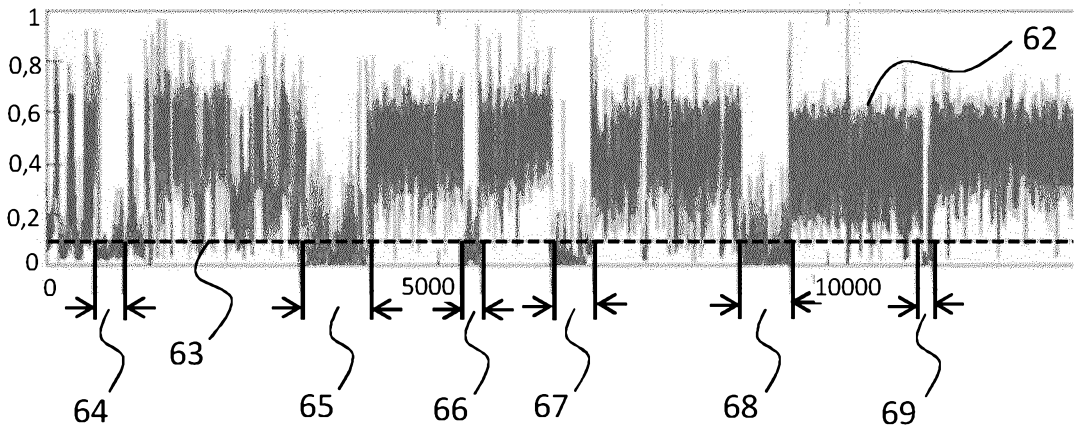


FIG 6c

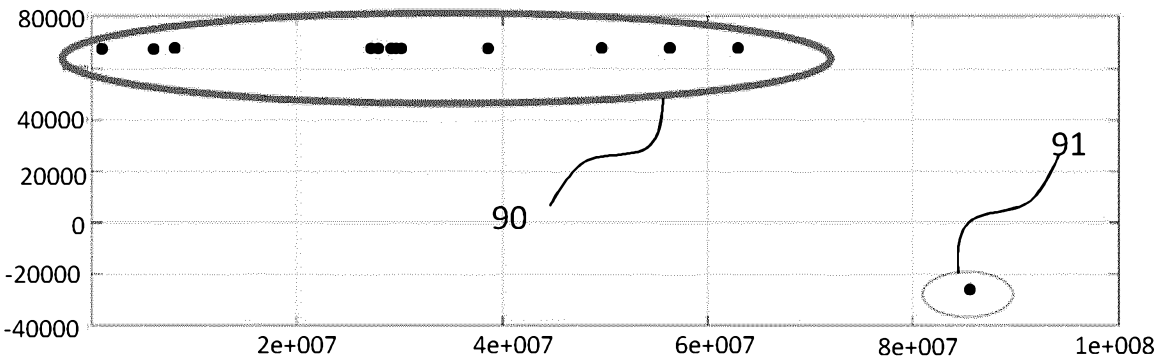


FIG 9

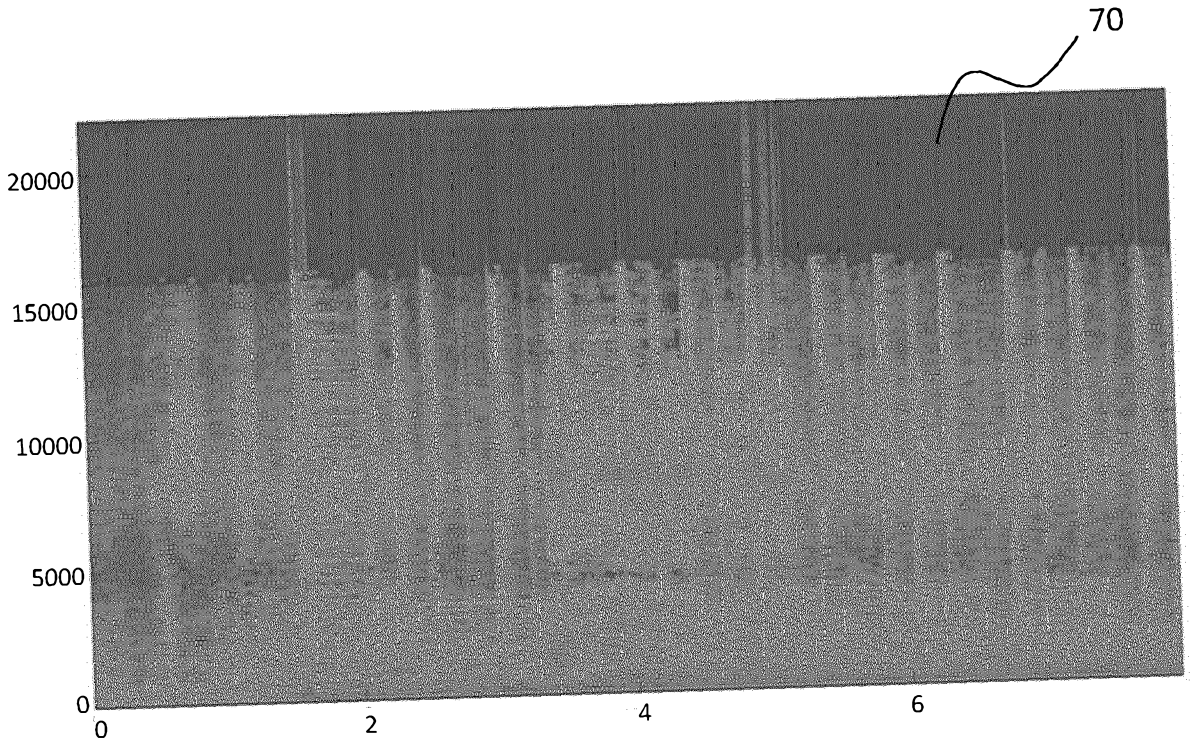


FIG 7

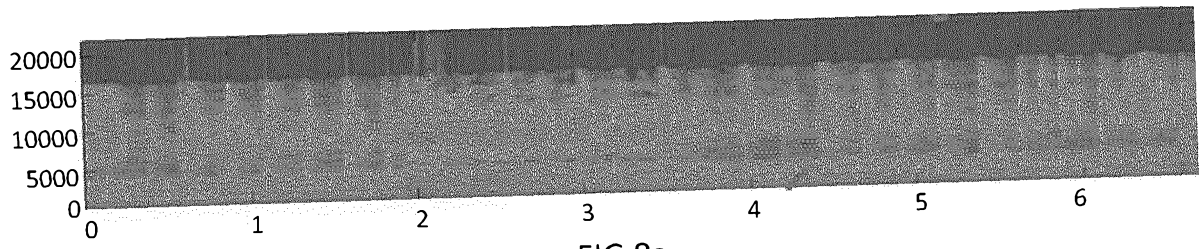


FIG 8a

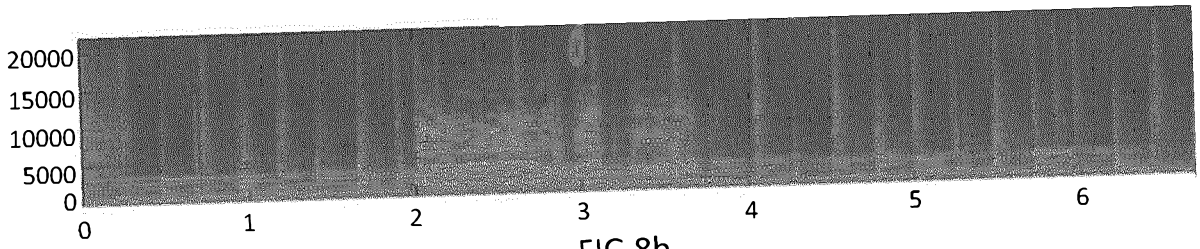


FIG 8b

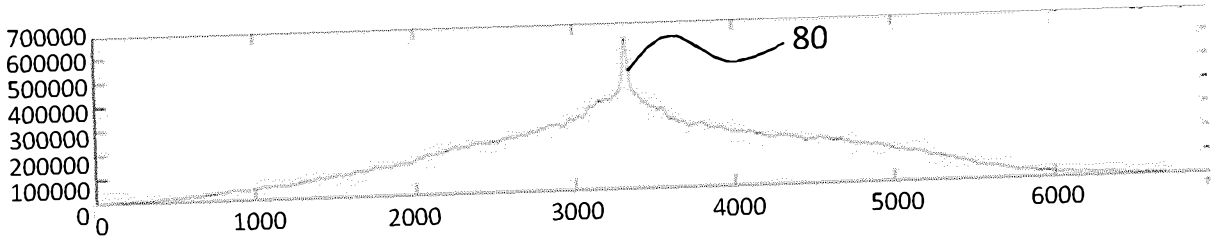


FIG 8c

RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

RAPPORT DE RECHERCHE PRÉLIMINAIRE

établi sur la base des dernières revendications déposées avant le commencement de la recherche

N° d'enregistrement national

FA 852522
FR 1851311

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
A	US 9 235 552 B1 (ECK DOUGLAS [US] ET AL) 12 janvier 2016 (2016-01-12) * colonne 7; figure 1 * * colonne 5, alinéa 46 - alinéa 55; revendication 13; figure 2 * -----	10	G10L21/055 G06F3/16
A	KIM MINJE ET AL: "Efficient neighborhood-based topic modeling for collaborative audio enhancement on massive crowdsourced recordings", 2016 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), IEEE, 20 mars 2016 (2016-03-20), pages 41-45, XP032900559, DOI: 10.1109/ICASSP.2016.7471633 [extrait le 2016-05-18] * abrégé * -----	1-15	
A	US 2012/257875 A1 (SHARPE BRUCE [CA] ET AL) 11 octobre 2012 (2012-10-11) * figures 1,3 * * alinéa [0073] - alinéa [0096]; figure 6 * -----	1-15	

DOMAINES TECHNIQUES RECHERCHÉS (IPC)

G11B
H04N
G10H
G06F

Date d'achèvement de la recherche	Examinateur
11 octobre 2018	Glasser, Jean-Marc

CATÉGORIE DES DOCUMENTS CITÉS

X : particulièrement pertinent à lui seul
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie
A : arrière-plan technologique
O : divulgation non-écrite
P : document intercalaire

T : théorie ou principe à la base de l'invention
E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.
D : cité dans la demande
L : cité pour d'autres raisons

& : membre de la même famille, document correspondant

EPO FORM 1503 12.99 (P04C14)

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 1851311 FA 852522

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.
Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du 11-10-2018
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 9235552 B1	12-01-2016	AUCUN	
US 2012257875 A1	11-10-2012	US 8205148 B1	19-06-2012
		US 2012257875 A1	11-10-2012

EPO FORM P0465