

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la
Propriété Intellectuelle
Bureau international



(43) Date de la publication internationale
4 septembre 2014 (04.09.2014)

WIPO | PCT

(10) Numéro de publication internationale
WO 2014/131984 A2

- (51) Classification internationale des brevets :
G10L 25/51 (2013.01)
- (21) Numéro de la demande internationale :
PCT/FR2014/050399
- (22) Date de dépôt international :
25 février 2014 (25.02.2014)
- (25) Langue de dépôt : français
- (26) Langue de publication : français
- (30) Données relatives à la priorité :
1351752 27 février 2013 (27.02.2013) FR
- (71) Déposant : INSTITUT MINES TELECOM [FR/FR];
TELECOM PARISTECH 46 rue Barrault, F-75634 Paris
Cedex 13 (FR).
- (72) Inventeurs : FENET, Sébastien; 4 rue de Provence, F-
94230 Cachan (FR). GRENIER, Yves; 26 rue de la Cha-
pelle, F-78114 Magny Les Hameaux (FR). RICHARD,
Gaël; 37 rue Chanzy, F-78220 Viroflay (FR).
- (74) Mandataires : HASSINE, Albert et al.; CABINET
PLASSERAUD, 52 rue de la Victoire, F-75440 Paris Ce-
dex 09 (FR).

(81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, RU, TJ, TM), européen (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Publiée :

— sans rapport de recherche internationale, sera republiée dès réception de ce rapport (règle 48.2.g)

(54) Title : GENERATION OF A SIGNATURE OF A MUSICAL AUDIO SIGNAL

(54) Titre : GENERATION D'UNE SIGNATURE D'UN SIGNAL AUDIO MUSICAL

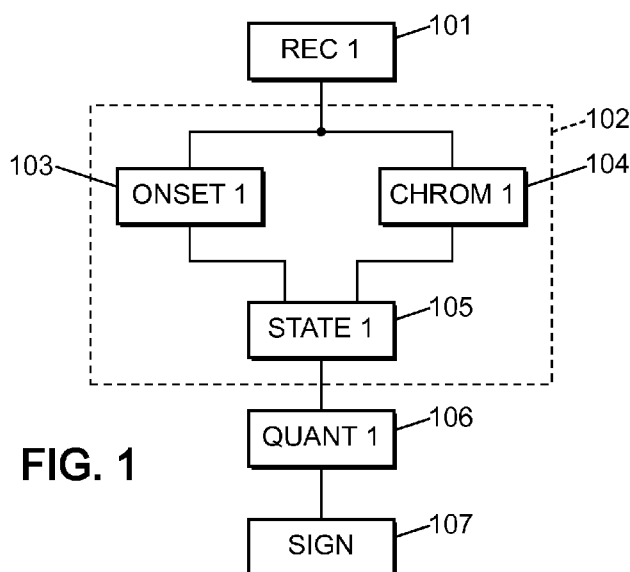


FIG. 1

(57) Abstract : The invention concerns a method for generating a signature of a musical audio signal of a given duration, the method comprising the following steps: - modelling (104) the musical audio signal to obtain, for each frequency band of a set of n frequency bands, a diagram representing the energy of the audio signal for the frequency band, on the basis of the time during said given duration; - determining (103) musical transition times t_k of the audio signal during the given duration; - associating (105) each musical transition time t_k with an item of local information comprising a vector of n values representative, respectively, of the energy of the audio signal in each of the n diagrams obtained between musical transition time t_k and a subsequent musical transition time t_{k+1} and/or a vector of n values representative, respectively, of the energy of the audio signal in each of the n diagrams obtained between musical transition time t_k and a preceding musical transition time t_{k-1} ; - determining (106), on the basis of the local information associated with each musical transition time t_k , a key associated with the musical transition time, the determined keys forming a first set of keys of the audio signal; - generating (107) a signature of the musical audio signal comprising pairs of keys from the first set of keys and associated musical transition times t_k .

(57) Abrégé :

[Suite sur la page suivante]

WO 2014/131984 A2

L'invention concerne un procédé de génération d'une signature d'un signal audio musical d'une durée donnée, le procédé comprenant les étapes suivantes : - modélisation (104) du signal audio musical pour obtenir, pour chaque bande de fréquences d'un ensemble de n bandes de fréquences, un diagramme représentant l'énergie du signal audio pour la bande de fréquences, en fonction du temps pendant ladite durée donnée; - détermination (103) d'instant t_k de transition musicale du signal audio pendant la durée donnée; - association (105) de chaque instant t_k de transition musicale avec une information locale comprenant un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale et/ou un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale; - détermination (106), en fonction de l'information locale associée à chaque instant t_k de transition musicale, d'une clé associée à l'instant de transition musicale, les clés déterminées formant un premier ensemble de clés du signal audio; - génération (107) d'une signature du signal audio musical comprenant des couples de clés du premier ensemble de clés et d'instant t_k de transition musicale associés.

Génération d'une signature d'un signal audio musical

La présente invention concerne la génération de signature pour l'identification de signaux audio.

5 Elle concerne plus précisément l'identification de signaux audio musicaux par extraction d'empreintes (communément appelée « audio-fingerprinting »).

L'identification de signaux audio a pour but de récupérer des métadonnées afin de les associer à un extrait audio inconnu a priori. Les terminaux mobiles tels que des téléphones mobiles de type Smartphone tendent à intégrer de telles applications d'identification. A cet effet, l'utilisateur capture un extrait audio au moyen du microphone de son Smartphone et l'application renvoie un ensemble de métadonnées telles que le titre d'une chanson, le nom d'un artiste, le nom d'un album, etc.

15 L'identification de signaux audio possède également des applications dans le contrôle automatique de droit d'auteur, dans la segmentation et l'annotation automatiques de flux multimédia, dans la détection de jingle, etc.

L'extraction d'empreinte est l'outil le plus répandu pour effectuer l'identification de signaux audio de requête lorsqu'aucune métadonnée n'est attachée au signal audio inconnu. Elle consiste à extraire, de signaux audio de référence, une représentation compacte (empreinte ou signature) qui est ensuite stockée dans une base de données. Lors de l'identification d'un signal audio de requête, sa signature est extraite. La meilleure correspondance avec la signature extraite est alors recherchée dans la base de données comprenant les signatures des signaux audio de référence.

On connaît de nombreux exemples de modèles d'empreintes (ou signatures) ainsi que de nombreuses méthodes de recherche de la meilleure correspondance.

En particulier, deux objectifs sont généralement mis en avant dans les méthodes connues :

- le signal audio capturé (de requête) doit être reconnu même s'il a subi une série de distorsions ;

- l'algorithme mis en œuvre doit pouvoir gérer une base de données contenant des quantités très importantes de signaux audio de référence.

5 Dans l'état de l'art, la robustesse à la distorsion est généralement définie comme étant la robustesse à différents post-traitements du signal audio. En effet, la problématique qui est généralement considérée est l'identification d'un signal qui a été tronqué, ré-égalisé, compressé en amplitude, étiré temporellement, décalé au niveau du pitch, et/ou enregistré dans un
10 environnement avec un mauvais microphone. Autrement dit, il s'agit de distorsions introduites dans le canal de transmission lors de la diffusion du signal audio.

Selon les méthodes, la robustesse couvre plus ou moins ces types de distorsions, mais ces distorsions correspondent toutes à des post-traitements appliqués à un signal de référence. Le signal audio de requête correspond alors
15 à l'un des signaux audio de référence auquel a été appliqué une série de distorsions, et l'identification consiste à retrouver le signal audio de référence. Une telle identification est appelée « identification quasi-exacte ».

En audio on définit deux signaux audio similaires comme étant deux
20 signaux audio qui présentent les mêmes caractéristiques harmoniques, rythmiques et mélodiques. Intuitivement, ce sont deux signaux audio qui seront considérés comme proches par un être humain.

Cependant, les méthodes par extraction d'empreintes traditionnelles ne traitent généralement pas du cas dans lequel deux signaux audio sont
25 similaires. Cela signifie notamment que si le système a appris un titre (alors stocké dans une base de données sous la forme d'un signal audio de référence) et qu'il est requis l'identification d'une reprise de ce titre (signal audio de requête), le système sera incapable d'indiquer la correspondance.

De manière plus gênante, un même titre, avec la même
30 instrumentation, qui est réenregistré par le même artiste, ne sera pas considéré comme correspondant au titre original. En effet, bien que, pour l'oreille humaine, les deux enregistrements semblent être les mêmes, les légères variations qui existent entre les deux versions (légers décalages rythmiques,

chant légèrement modifié, exécution instrumentale un peu différente) en font des signaux radicalement différents d'un point de vue informatique.

L'identification de deux signaux audio similaires est appelée « identification approchée ».

5 De telles limitations sont principalement dues au fait que les méthodes par extraction d'empreintes traditionnelles utilisent des informations de bas niveau pour caractériser le signal. Ces informations, qui intègrent très peu de signification musicale, ne sont pas conservées d'une version à une autre d'un même titre.

10 Ainsi, les méthodes par extraction d'empreintes sont peu utilisées dans les domaines de la similarité musicale ou de la détection de reprises.

Par ailleurs, des approches permettant de mettre en correspondance deux versions différentes d'une même chanson ont été développées. Cependant, de telles approches sont incapables de traiter des bases de données de signaux de référence de grande taille. En effet, ces méthodes sont consommatrices en ressources et requièrent de comparer un par un les signaux audio de référence au signal audio de requête.

15 Il existe donc un besoin de définir une signature d'un signal audio qui permette à la fois l'identification quasi-exacte et l'identification approchée de signaux audio, tout en permettant de gérer des bases de données de grande taille.

La présente invention vient améliorer la situation.

25 Selon un premier aspect de l'invention, on décrit un procédé de génération d'une signature d'un signal audio musical d'une durée donnée, le procédé comprenant les étapes suivantes :

- modélisation du signal audio musical pour obtenir, pour chaque bande de fréquences d'un ensemble de n bandes de fréquences, un diagramme représentant l'énergie du signal audio pour la bande de fréquences, en fonction du temps pendant la durée donnée ;
- 30 - détermination d'instant t_k de transition musicale du signal audio pendant la durée donnée ;
- association de chaque instant t_k de transition musicale avec une information

- locale comprenant un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale et/ou un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale;
- 5 - détermination, en fonction de l'information locale associée à chaque instant t_k de transition musicale, d'une clé associée à l'instant de transition musicale, les clés déterminées formant un premier ensemble de clés du signal audio ;
- 10 - génération d'une signature du signal audio musical comprenant des couples , chaque couple comprenant une clé du premier ensemble de clés et un instant t_k de transition musicale associé.

Le procédé permet ainsi de générer une signature qui est représentative d'informations rythmiques et harmoniques du signal audio musical. Ainsi, les clés obtenues selon le procédé précité peuvent être communes à deux versions d'un même titre audio, et la signature peut être utilisée dans le contexte de l'identification approchée. De plus, la signature étant compacte, elle permet la comparaison efficace sur une durée importante de signal ce qui est notamment nécessaire dans le cadre de l'identification approchée.

15

20

On entend par signal audio musical un signal audio de requête (que l'on cherche à identifier) ou un signal audio de référence (pour stockage dans une base de données et pour comparaison ultérieure avec un signal audio de requête).

25 Selon une réalisation de l'invention, les clés du premier ensemble de clés peuvent être obtenues par quantification des vecteurs de n valeurs représentatives respectivement de l'énergie du signal dans chaque bande de fréquences de l'ensemble de n bandes de fréquences dans une période comprise entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale ou par quantification des vecteurs de n valeurs représentatives respectivement de l'énergie du signal dans chaque bande de fréquences de l'ensemble de n bandes de fréquences dans une période comprise entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de

30

transition musicale.

La quantification utilisée pour générer les clés permet de baisser la sélectivité opérée sur la base de ces clés. En quantifiant des vecteurs de valeurs d'énergie issus de deux signaux similaires mais différents, il est possible d'obtenir deux fois la même clé, et d'ainsi pouvoir identifier un signal de requête de manière approchée.

Selon un mode de réalisation de l'invention, chaque clé du premier ensemble de clés peut être un vecteur de n bits, chaque bit d'index i étant déterminé par comparaison entre la valeur d'énergie d'index i dans le vecteur de n valeurs à partir duquel la clé est déterminée, et la moyenne des valeurs du vecteur de n valeurs.

Ainsi, la quantification appliquée aux valeurs des vecteurs est maximale, ce qui permet d'abaisser la sélectivité des clés tout en conservant une signification d'un point de vue musical (utilisation d'un vecteur de n bits associé à un instant de transition musicale représentatif).

En complément, si la valeur d'énergie d'index i dans le vecteur de n valeurs à partir duquel la clé est déterminée est supérieure à la moyenne des valeurs dudit vecteur de n valeurs, alors le bit d'index i peut être égal à 1, le bit d'index i étant égal à 0 sinon.

Selon un mode de réalisation de l'invention, les instants t_k de transition musicale correspondent aux maxima locaux d'une fonction d'onset calculée à partir du signal audio musical.

On entend par fonction d'onset toute fonction dont le rôle est d'estimer les attaques de notes et les attaques percussives dans un signal audio. A titre d'exemple, une fonction de ce type est décrite dans la publication « *Accurate Tempo Estimation Based on Harmonic + Noise Decomposition* », de Alonso et Al, 2007.

Une telle réalisation permet d'assurer que les temps de transition musicale sont significatifs d'un point de vue rythmique.

Un deuxième aspect de l'invention concerne un procédé d'identification d'un signal musical, de requête, parmi des signaux musicaux, de référence, le signal de requête et les signaux de référence étant représentés par des signatures respectives obtenues par la mise en œuvre du procédé

selon le premier aspect de l'invention, chaque signature de signal de référence comprenant ainsi un second ensemble de clés auxquelles sont associés des instants respectifs τ_k de transition musicale, le procédé comprenant les étapes :

- 5 - pour chaque signature de signal de référence ayant au moins une clé du second ensemble en commun avec le premier ensemble de clés de la signature du signal de requête, constitution d'un troisième ensemble de clés comprenant la au moins une clé du second ensemble en commun avec le premier ensemble, chaque clé du troisième ensemble étant
- 10 associée à la fois à l'instant de transition musicale associé à la clé dans la signature du signal de référence, et à l'instant de transition musicale associé à la clé dans la signature du signal de requête ;
- détermination d'un score en fonction d'une correspondance entre chaque
- 15 paire d'instants de transition musicale associée à chaque clé du troisième ensemble ; et
- sélection d'au moins une signature de signal de référence ayant l'un des scores les plus élevés, en tant que signature d'un signal de référence candidat pour correspondre au signal de requête.

20 Ce mode de réalisation permet ainsi d'obtenir avec une complexité réduite un sous ensemble de signatures des signaux audio de référence qui sont susceptibles d'identifier de manière quasi exacte ou approchée le signal audio de requête.

25 En complément, on peut sélectionner un jeu de signatures de signaux de référence ayant des scores les plus élevés, et on peut appliquer un post-traitement :

- à des signaux de référence candidats, représentés par les signatures dudit
 - jeu, et
 - au signal de requête,
- 30 et une étape d'identification du signal de requête comme l'un des signaux de référence.

 Ainsi, il est possible d'appliquer un post-traitement, qui met en œuvre des calculs généralement complexes, à un ensemble réduit de signaux de référence candidats, ce qui permet de diminuer les ressources requises pour

identifier le signal de requête, en comparaison du même post-traitement appliqué à l'ensemble de la base de données.

En variante ou en complément, la détermination du score d'une signature de signal de référence comprend les étapes suivantes:

- 5 - détermination d'une fonction affine mettant en correspondance un plus grand nombre d'instant de transition musicale du signal de référence associés aux clés du troisième ensemble avec les instants de transition musicale du signal de requête associés aux clés respectives du troisième ensemble.

10 Le score attribué à la signature du signal de référence peut alors être égal au plus grand nombre précité.

Une telle réalisation permet de sélectionner avec une complexité réduite des signatures de signaux de référence ayant un score élevé, le score étant représentatif d'une similarité musicale entre les signaux audio.

15 En complément, la fonction affine peut être obtenue par application d'une transformée de Hough à l'ensemble de paires d'instant de transition musicale associées aux clés du troisième ensemble.

Selon un mode de réalisation, la détermination du score d'un signal audio de référence peut comprendre les étapes suivantes:

20 - construction d'un nuage de points, chaque point correspondant à une clé du troisième ensemble, l'abscisse du point étant l'instant de transition musicale du signal de référence associé à la clé du troisième ensemble et l'ordonnée du point correspondant à l'instant de transition musicale du signal de requête associé à la clé du troisième ensemble ;

25 - détermination d'une droite reliant un maximum de points du nuage de points construit ;

- détermination du score de la signature du signal de référence comme étant égal au nombre de points reliés par la droite déterminée dans le nuage de points.

30 En complément ou en variante, le post-traitement est appliqué successivement pour chaque signal de référence correspondant à une signature sélectionnée, et au signal de requête modifié en fonction de la fonction affine déterminée pour l'obtention du score de la signature du signal de référence.

Ainsi, ce mode de réalisation permet de tirer profit d'informations issues de l'étape d'obtention d'une fonction affine (pouvant être la droite dans le nuage de points précédemment introduit), dans le cadre du post-traitement, améliorant ainsi l'efficacité du procédé d'identification.

5 En complément, le signal de requête est modifié par mise à l'échelle du signal de requête en fonction d'une pente de la fonction affine et par recadrage temporel du signal de requête en fonction d'une ordonnée à l'origine de la fonction affine.

10 Un tel mode de réalisation permet d'assurer que le signal audio de référence et le signal audio de requête modifié sont comparés avec un même étirement temporel et partant du même instant de départ.

15 Un troisième aspect de l'invention concerne un programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé selon le premier aspect de l'invention, lorsque ce programme est exécuté par un processeur.

 Un quatrième aspect de l'invention concerne un programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé selon le deuxième aspect de l'invention, lorsque ce programme est exécuté par un processeur.

20 Un cinquième aspect de l'invention concerne un dispositif de génération d'une signature d'un signal audio musical d'une durée donnée, le dispositif comprenant :

25 - une unité de modélisation du signal audio musical pour obtenir, pour chaque bande de fréquences d'un ensemble de n bandes de fréquences, un diagramme représentant l'énergie du signal audio pour la bande de fréquences, en fonction du temps pendant la durée donnée ;

 - une première unité de détermination d'instant t_k de transition musicale du signal audio pendant la durée donnée ;

30 - une unité d'association de chaque instant t_k de transition musicale avec une information locale comprenant un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale et/ou un vecteur de n valeurs représentatives

respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale;

- une deuxième unité de détermination, en fonction de l'information locale associée à chaque instant t_k de transition musicale, d'une clé associée à l'instant de transition musicale, les clés déterminées formant un premier ensemble de clés du signal audio ;
- une unité de génération d'une signature du signal audio musical comprenant des couples, chaque couple comprenant une clé du premier ensemble de clés et un instant t_k de transition musicale associé.

Un sixième aspect de l'invention concerne un dispositif d'identification d'un signal musical, de requête, parmi des signaux musicaux, de référence, le signal de requête et les signaux de référence étant représentés par des signatures respectives obtenues par la mise en œuvre du procédé selon le premier aspect de l'invention, chaque signature de signal de référence comprenant ainsi un second ensemble de clés auxquelles sont associés des instants respectifs τ_k de transition musicale, le dispositif comprenant :

- une unité de constitution, pour chaque signature de signal de référence ayant au moins une clé du second ensemble en commun avec le premier ensemble de clés de la signature du signal de requête, d'un troisième ensemble de clés comprenant la au moins une clé du second ensemble en commun avec le premier ensemble, chaque clé du troisième ensemble étant associée à la fois à l'instant de transition musicale associé à la clé dans la signature du signal de référence, et à l'instant de transition musicale associé à la clé dans la signature du signal de requête ;
- une unité de détermination d'un score en fonction d'une correspondance entre chaque paire d'instant de transition musicale associée à une clé du troisième ensemble ; et
- une unité de sélection d'au moins une signature de signal de référence ayant l'un des scores les plus élevés, en tant que signature d'un signal de référence candidat pour identifier le signal de requête.

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

- 5 - la figure 1 illustre un procédé de génération d'une signature d'un signal audio musical d'une durée donnée selon l'invention ;
- la figure 2 présente une superposition d'une fonction d'onset calculée à partir d'un signal audio musical et d'un chromagramme déterminé à partir du même signal audio musical, selon l'invention ;
- 10 - la figure 3 illustre un procédé d'identification d'un signal audio musical, de requête, parmi des signaux audio musicaux, de référence, le signal audio de requête et les signaux audio de référence étant représentés par des signatures respectives obtenues par la mise en œuvre du procédé selon la figure 1 ;
- 15 - la figure 4 est un diagramme temps/temps représentant un nuage de points selon un mode de réalisation de l'invention ;
- la figure 5 illustre un dispositif de génération d'une signature d'un signal audio musical d'une durée donnée, selon une réalisation de l'invention ;
- 20 - la figure 6 présente un dispositif d'identification 600 d'un signal audio musical, de requête, parmi des signaux audio musicaux, de référence, le signal audio de requête et les signaux de référence étant représentés par des signatures respectives obtenues par la mise en œuvre du procédé selon la figure 1, selon l'invention.

25 La figure 1 illustre un procédé de génération d'une signature d'un signal audio musical d'une durée donnée selon l'invention.

30 A une étape 101, un signal audio musical est reçu par un dispositif de génération de signature selon l'invention, qui sera illustré ultérieurement en référence à la figure 5. Un tel signal audio musical peut correspondre à un signal audio pour lequel l'utilisateur aimerait accéder à des métadonnées, telles que le nom de l'artiste ou le titre. Un tel signal audio est appelé signal audio de requête. Le signal audio musical peut également correspondre à un signal audio de référence dont la signature est destinée à être stockée dans une base de données en vue d'une comparaison ultérieure avec une signature de signal

audio de requête (comme détaillé dans ce qui suit en référence à la figure 3) en vue d'une identification du signal audio de requête.

5 Comme expliqué précédemment, l'invention vise à permettre non seulement l'identification quasi-exacte (à savoir la recherche efficace d'un extrait audio dans une base importante avec robustesse aux distorsions) mais également l'identification approchée (à savoir le rapprochement d'un extrait audio avec un signal audio de référence qui lui est similaire au sein d'une base importante).

10 A partir du signal audio musical reçu, une étape de détermination d'états du signal audio musical est effectuée en référence à l'étape 102.

L'étape 102 comprend, à une étape 103, la détermination d'instant t_k de transition musicale du signal audio musical, pendant la durée donnée du signal audio musical.

15 En parallèle de l'étape 103, à une étape 104, l'étape de détermination d'états comprend la modélisation du signal audio musical pour obtenir, pour chaque bande de fréquences d'un ensemble de n bandes de fréquences, un diagramme représentant l'énergie du signal audio de requête pour la bande de fréquences en question, en fonction du temps pendant la durée donnée précitée.

20 Les étapes 103 et 104 seront mieux comprises en référence à la figure 2, qui présente une superposition d'une fonction d'onset 201 calculée à partir du signal audio musical et d'un chromagramme 205 déterminé à partir du signal audio musical. A noter que la détermination du chromagramme 205 et de la fonction d'onset 201 peuvent requérir tous les deux l'application d'une transformée de Fourier au signal audio de requête, avec des fenêtrages différents. Afin d'éviter tout décalage temporel entre la fonction d'onset 201 et le chromagramme 205, l'invention peut prévoir de synchroniser les transformées de Fourier.

30 La présente invention peut par exemple prévoir de détecter les instants t_k de transition musicale comme correspondant aux maxima locaux 202 de la fonction d'onset.

Le chromagramme 205 est constitué des n diagrammes 206.i, i variant de 1 à n , illustrant l'énergie du signal audio musical pour chaque bande

de fréquences. Plus l'énergie est importante, plus le chromagramme est clair.

Le procédé selon l'invention comprend ensuite une étape 105 de définition d'états du signal audio musical. Un état s_k est défini par l'association d'un instant t_k avec une information locale i_k :

$$5 \quad s_k = (t_k, i_k).$$

L'information locale i_k peut être donnée par les caractéristiques suivantes :

$$i_k = (c_{k,l}, c_{k,r})$$

dans lequel $c_{k,l}$ est un vecteur chromatique moyen à gauche de t_k et $c_{k,r}$ est un vecteur chromatique moyen à droite de t_k .

Le vecteur $c_{k,r}$ est un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio musical dans chacun des n diagrammes 206 obtenus dans la période 203 entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale. Le vecteur $c_{k,l}$ est un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio de requête dans chacun des n diagrammes 206.i obtenus dans la période 204 entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale. Par exemple, la composante d'index i du vecteur $c_{k,r}$ (notée $c_{k,r,i}$ ci-après) correspond à la moyenne de l'énergie de la bande de fréquences représentée par le diagramme 206.i sur la période 203 et la composante d'index i du vecteur $c_{k,l}$ (notée $c_{k,l,i}$ ci-après) correspond à la moyenne de l'énergie de la bande de fréquence représentée par le diagramme 206.i sur la période 204.

A noter que l'information locale i_k peut correspondre uniquement au vecteur chromatique moyen $c_{k,r}$ à droite de t_k ou au vecteur chromatique moyen $c_{k,l}$ à gauche de t_k .

L'avantage de la définition d'états présentée ci-dessus est que le signal audio de requête est représenté sous une forme compacte et représentative d'informations rythmiques et harmoniques du signal audio musical. Ainsi, la présente invention permet des applications d'identification quasi-exacte mais également d'identification approchée d'un signal audio de requête.

Le procédé selon l'invention comprend ensuite une étape 106 de

détermination, en fonction de l'information locale associée à chaque instant t_k de transition musicale, d'une clé associée à l'instant t_k de transition musicale, les clés déterminées formant un premier ensemble de clés du signal audio musical.

5 Il convient de générer un premier ensemble de clés qui soient invariantes aux variations du signal audio musical et suffisamment discriminantes. L'invariance aux variations doit couvrir les distorsions classiques en post-traitement (égalisation, décalage de pitch, etc) mais également les enregistrements d'un même titre dans des conditions différentes
10 (correspondance de signaux audio similaires).

Selon l'invention, une clé peut être déterminée pour chaque état s_k déterminé du signal audio musical, et peut correspondre à une version quantifiée de l'un des vecteurs $c_{k,l}$ et $c_{k,r}$. Dans ce qui suit, on prend l'exemple d'une version quantifiée du vecteur $c_{k,r}$. Chaque composante $c_{k,r,i}$ du vecteur $c_{k,r}$
15 peut par exemple être quantifiée à l'aide d'un seul bit $\tilde{c}_{k,r,i}$, de la manière suivante :

$$\tilde{c}_{k,r,i} = \begin{cases} 1 & \text{si } c_{k,r,i} > \bar{c}_{k,r} \\ 0 & \text{sinon} \end{cases}$$

20 avec $\bar{c}_{k,r}$ correspondant à la moyenne des $c_{k,r,i}$ pour i variant entre 1 et n .

Ainsi, la quantification du vecteur $c_{k,r}$ est déterministe. L'invention peut également prévoir, afin de limiter le risque de modifier une clé à ce stade du procédé, de générer un jeu probabiliste de vecteurs $\tilde{c}_{k,r}$ binaires, associés à
25 leurs probabilités respectives.

Chaque vecteur binaire $\tilde{c}_{k,r}$ constitue une clé du signal audio musical, et est stocké en association avec le temps t_k de transition musicale.

Le premier ensemble de clés ainsi générées et les instants t_k respectivement associés à ces clés forment alors une signature du signal audio musical, à une étape 107.
30

Une telle signature selon l'invention contient des informations rythmiques et harmoniques sur le signal audio musical, permettant ainsi

l'identification approchée aussi bien que l'identification quasi-exacte. Un tel procédé de génération de signature est appliqué à la fois au signal audio de requête, mais également aux signaux audio de référence en vue du stockage de leurs signatures respectives dans une base de données.

5 Lorsque le procédé décrit ci-avant est appliqué à un signal audio de référence, on prévoit en outre la définition d'une fonction d'index. On appelle fonction d'index, une fonction qui prend une clé en entrée et qui retourne un jeu de valeurs.

10 L'intérêt d'une telle opération est qu'elle est potentiellement réalisable avec une complexité $O(1)$. Dans notre cas, les valeurs sont des pointeurs sur les signaux audio de référence de la base de données et les clés sont des caractéristiques audio qui caractérisent ces signaux audio de référence.

15 Ainsi, lorsque le procédé de génération de signature a été appliqué à tous les signaux audio de référence, les clés de toutes les références sont extraites et stockées dans l'index. Ainsi, les valeurs d'une clé correspondent aux identifiants des signaux audio de référence ainsi que les temps t_k de transition musicale correspondant à la clé dans ces signaux audio de référence.

20 La figure 3 illustre un procédé d'identification d'un signal musical, de requête, parmi des signaux musicaux, de référence, le signal de requête et les signaux de référence étant représentés par des signatures respectives obtenues par la mise en œuvre du procédé selon la figure 1.

 A une étape 301, un signal audio de requête est reçu par un dispositif d'identification qui sera présenté ultérieurement en référence à la figure 6.

25 A une étape 302, une signature du signal audio de requête est obtenue, par application du procédé selon la figure 1 au signal audio de requête. La signature comprend ainsi un premier ensemble de clés K associées respectivement à des instants t_k de transition musicale (à noter que l'indice k des instants de transition musicale, est indépendant du symbole K représentant une clé).

30

 A une étape 303, on accède à une base de données comprenant les signatures des signaux audio de référence, en vue de constituer, à l'aide de la fonction d'index, et pour chaque signature de signal de référence ayant au

moins une clé du second ensemble en commun avec le premier ensemble de clés de la signature du signal de requête, un troisième ensemble de clés comprenant les clés du second ensemble en commun avec le premier ensemble. A noter qu'une clé K du troisième ensemble peut être associée à
5 plusieurs instants de transition musicale dans le premier ensemble et/ou à plusieurs instants de transition musicale dans le second ensemble.

Le procédé comprend ensuite une étape 304 de détermination d'un score en fonction d'une correspondance entre chaque paire d'instant de transition musicale associée à une clé du troisième ensemble. La détermination
10 du score est effectuée pour chaque signature de signal de référence comprenant au moins une clé en commun avec le premier ensemble de clés de la signature du signal audio de requête. La détermination du score peut par exemple consister en la détermination d'une fonction affine mettant en correspondance un plus grand nombre d'instant de transition musicale du
15 signal de référence associés aux clés du troisième ensemble avec les instants de transition musicale du signal de requête associés aux mêmes clés du troisième ensemble, le score attribué à la signature du signal de référence étant égal au plus grand nombre précité.

Un exemple détaillé de détermination du score d'une signature de signal de référence, à partir du troisième ensemble de clés commune à cette signature de signal de référence et à la signature du signal de requête, est illustré en référence à la figure 4.
20

La figure 4 est un diagramme temps/temps représentant un nuage de points 403, chaque point correspondant à une clé du troisième ensemble, l'abscisse 401 du point étant l'instant de transition musicale du signal de référence associé à la clé du troisième ensemble et l'ordonnée 402 du point correspondant à l'instant de transition musicale du signal de requête associé à la clé du troisième ensemble. On construit ainsi un nuage de points par signature de signal de référence qui partage au moins une clé en commun avec
25 la signature du signal de requête.
30

A partir de ce nuage de points, on détermine une droite 404 reliant un maximum de points du nuage de points construit. La détermination de la droite 404 peut être effectuée par application d'un algorithme de Hough à l'ensemble

de paires d'instants de transition musicale associées aux clés du troisième ensemble.

Le score de la signature du signal de référence pour lequel le nuage de points a été construit peut alors être égal au nombre de points reliés par la droite 404 déterminée dans le nuage de points. Sur la figure 4, on remarque qu'une droite 404 comprenant un grand nombre de points est obtenue. Le fait que de nombreux points soient situés en dehors de la droite s'explique par le fait qu'une clé du premier ensemble peut être présente plusieurs fois dans le deuxième ensemble. Ceci traduit la répétition d'une clé dans le titre représenté par le signal audio de référence. Il faut en outre garder à l'esprit que, de par la quantification très forte appliquée pour déterminer les clés, ces dernières ne sont pas très discriminantes (ce qui est volontaire, et utile pour permettre l'identification approchée du signal audio de requête).

A une étape 305, on sélectionne au moins une signature de signal de référence ayant l'un des scores les plus élevés, en tant que signature d'un signal de référence candidat pour correspondre au signal de requête. Par exemple, l'invention peut prévoir de sélectionner M signatures de signaux de référence candidats, M pouvant être prédéterminé ou pouvant dépendre du nombre de clés du premier ensemble de clés de la signature du signal de requête, par exemple. Aucune restriction n'est attachée à la valeur de M.

A une étape 306, on sélectionne un jeu de signatures de signaux de référence ayant des scores les plus élevés, et on applique un post-traitement :

- à des signaux de référence candidats, représentés par les signatures du jeu, et
- au signal de requête,

en vue d'identifier le signal de requête comme l'un des signaux de référence. Le jeu de signatures peut par exemple comprendre les M signatures précitées.

Une telle étape de post-traitement peut mettre en œuvre des calculs complexes, sans pour autant mobiliser des ressources trop importantes, dans la mesure où la base de données de signaux de référence a été réduite à M signaux de référence candidats.

Un exemple de post-traitement est décrit ci-après. Toutefois,

l'invention n'est aucunement restreinte à cet exemple, et toute méthode connue pourra être appliquée afin de sélectionner l'un des signaux audio de référence candidats comme identifiant le signal audio de requête. L'état de l'art propose de nombreuses techniques à cet effet.

5 A ce stade du procédé, il convient de comparer les M signaux audio de référence sélectionnés avec le signal audio de requête.

A titre d'exemple, il est proposé ci-après une approche de programmation dynamique qui repose sur la modélisation des signaux audio par états telle que précédemment décrite.

10 Tout d'abord, pour chaque signature de signal de référence possédant des clés en commun avec le premier ensemble de clés, le signal audio de requête peut être modifié en fonction de la fonction affine déterminée pour l'obtention du score de la signature du signal de référence.

En effet, la droite 404 définit une fonction affine caractérisée par une pente référencée 405 sur la figure 4 et notée a , et par une ordonnée à l'origine référencée 406 et notée b . La droite 404 signifie qu'une clé extraite du signal audio de requête à un instant t , correspondra à une clé dans le signal audio de référence à l'instant $a(t-b)$.

15 L'inverse de a représente alors un étirement temporel du signal audio de requête tandis que b représente l'instant dans le signal audio de référence auquel commence le signal audio de requête. Ainsi, le signal audio de requête peut être modifié par mise à l'échelle du signal de requête en fonction de la pente a et par recadrage temporel du signal audio de requête en fonction de l'ordonnée à l'origine b .

20 Le signal audio de requête ainsi modifié est représenté par la séquence d'états $\{s_1 \dots s_i \dots s_m\}$, m étant le nombre d'instant de transition musicale déterminé précédemment dans l'intervalle de temps correspondant à la durée du signal audio de requête. Pour représenter les états des signaux de référence candidats, on utilise dans ce qui suit la notation $\{\sigma_1 \dots \sigma_j \dots \sigma_p\}$, p étant le nombre d'instant de transition musicale dans le signal de référence candidat considéré, p étant dépendant du signal de référence candidat.

30 Afin de dynamiquement aligner deux séquences d'états, il convient de définir trois types de pénalités :

$f^d(s_i)$: pénalité assignée pour la suppression de l'état s_i ;

$f^i(s_i)$: pénalité assignée à l'insertion de l'état s_i ;

$f^s(s_i, \sigma_j)$: pénalité assignée au remplacement de l'état s_i par l'état σ_j .

A titre d'exemple, les fonctions f^d et f^i peuvent toutes deux prendre
5 une valeur constante égale à 0.3.

Sachant que $s_i = (t_i, (c_{i,l}, c_{i,r}))$ et $\sigma_j = (\tau_j, (\gamma_{j,l}, \gamma_{j,r}))$, en notant τ_j l'instant de transition musicale d'indice j dans le signal de référence, $\gamma_{j,l}$ le vecteur chromatique moyen à gauche de τ_j , et $\gamma_{j,r}$ le vecteur chromatique moyen à droite de τ_j , on définit f^s par :

$$10 \quad f^s(s_i, \sigma_j) = \cos(c_{i,l}, \gamma_{j,l}) \cdot \cos(c_{i,r}, \gamma_{j,r}) \cdot e^{-\frac{|t_i - \tau_j|}{c}}$$

Le premier terme en cosinus pénalise la ressemblance du vecteur chromatique moyen à gauche de t_i avec le vecteur chromatique moyen à gauche de τ_j . Le second terme en cosinus pénalise la ressemblance du vecteur chromatique moyen à droite de t_i avec le vecteur chromatique moyen à droite de τ_j . Le terme exponentiel pénalise l'erreur temporelle entre
15 l'occurrence de l'état s_i dans le signal de requête modifié, et l'occurrence de l'état σ_j dans le signal de référence.

Une programmation dynamique consiste à remplir de manière itérative une matrice D . Pour chaque couple $(i, j) \in \{1..m\} \times \{1..p\}$, $D(i, j)$ contient le score de l'alignement de la sous-séquence $\{s_1...s_i\}$ avec la sous séquence $\{\sigma_1... \sigma_j\}$. D est calculée de la manière suivante :

$$20 \quad D(i, j) = \max \left\{ \begin{array}{l} D(i, j-1) \cdot f^d(\sigma_j) \\ D(i-1, j-1) \cdot f^s(s_i, \sigma_j) \\ D(i-1, j) \cdot f^i(s_i) \end{array} \right\}$$

Le score de l'alignement de $\{s_1...s_m\}$ avec $\{\sigma_1... \sigma_p\}$ est finalement donné par $D(m, p)$.

A une étape 307, à l'issue du post-traitement, le signal audio de requête est identifié de manière quasi-exacte ou approchée comme

correspondant au signal audio de référence ayant le score $D(m,p)$ le plus élevé. Une telle identification peut par ailleurs être conditionnée par une étape de comparaison avec un seuil prédéterminé : si le score $D(m,p)$ le plus élevé est inférieur à ce seuil, aucune identification n'est obtenue pour le signal audio de

5 requête.

La figure 5 présente un dispositif 500 de génération d'une signature d'un signal audio musical d'une durée donnée.

A cet effet, le dispositif 500 comprend une unité de réception 501

10 d'un signal audio musical. Comme précédemment décrit, le signal audio musical peut être un signal audio de requête ou de référence.

Le dispositif 500 comprend en outre une unité de modélisation 502 du signal audio musical pour obtenir, pour chaque bande de fréquences d'un ensemble de n bandes de fréquences, un diagramme représentant l'énergie du signal audio pour la bande de fréquences, en fonction du temps pendant la

15 durée donnée. Les diagrammes obtenus peuvent être regroupés sous la forme du chromagramme 205 illustré sur la figure 2. En particulier, l'unité de modélisation 502 est apte à mettre en œuvre l'étape 104 du procédé illustré en figure 1.

Le dispositif 500 comprend une première unité de détermination 502 d'instant t_k de transition musicale du signal audio pendant la durée donnée, qui est apte à mettre en œuvre l'étape 103 du procédé illustré en figure 1.

20

Afin de synchroniser les opérations effectuées par les unités 502 et 503, le dispositif 500 comprend une unité de synchronisation 504 apte à mettre

25 en œuvre l'étape 104 du procédé illustré en figure 1.

Le dispositif 500 comprend une unité d'association 505 de chaque instant t_k de transition musicale avec une information locale comprenant un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale et/ou un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans

30 chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale. L'unité d'association 505 est apte à

mettre en œuvre l'étape 105 du procédé illustré en figure 1.

Le dispositif 500 comprend en outre une deuxième unité de détermination 506, en fonction de l'information locale associée à chaque instant t_k de transition musicale, d'une clé associée à l'instant de transition musicale, les clés déterminées formant un premier ensemble de clés du signal audio. L'unité de détermination 506 est apte à mettre en œuvre l'étape 106 du procédé illustré en figure 1.

Le dispositif 500 comprend une unité de génération 507 d'une signature du signal audio musical comprenant des couples de clés du premier ensemble de clés et d'instant t_k de transition musicale associés, qui est apte à mettre en œuvre l'étape 107 du procédé illustré en figure 1. L'unité de génération 507 peut en outre être apte à transmettre le signal audio musical ainsi que la signature générée à une entité distante.

Par exemple, lorsque le signal audio musical est un signal audio de requête, il peut être transmis avec la signature générée à un serveur distant en charge d'identifier le signal audio de requête, ou plus généralement à un dispositif selon la figure 6. Dans ce cas le dispositif 500 peut être inclus dans un terminal utilisateur tel qu'un terminal mobile de télécommunications, un ordinateur portable ou fixe, etc.

En variante, lorsque le signal audio musical est un signal audio de référence, il peut être transmis avec la signature générée à une base de données de référence apte à stocker le signal audio de référence en association avec la signature générée.

La figure 6 illustre un dispositif d'identification 600 d'un signal musical, de requête, parmi des signaux musicaux, de référence, le signal de requête et les signaux de référence étant représentés par des signatures respectives obtenues par la mise en œuvre du procédé selon la figure 1. A cet effet, le dispositif 600 peut être relié au dispositif 500 et peut ainsi recevoir par une unité de réception 601 un signal audio de requête accompagné d'une signature. Dans ce cas, une unité 602 extrait la signature du signal audio de requête pour communication à une unité de constitution 603, décrite dans ce qui suit. En variante, l'unité de réception 601 peut recevoir un signal audio de requête sans signature, la signature étant générée par une unité 602. L'unité

602 peut ainsi correspondre au dispositif de génération de signature 500 et peut transmettre la signature générée à l'unité de constitution 603.

En outre, le dispositif 600 peut être relié à une base de données 604 stockant des signatures respectivement associées à des signaux audio de référence. Dans la base de données 604, chaque signature de signal de référence comprend ainsi un second ensemble de clés auxquelles sont associés des instants respectifs τ_k de transition musicale. Selon l'invention, la base de données 604 peut être intégrée dans le dispositif 600.

Le dispositif 600 comprend l'unité de constitution 603, qui accède à la base de données 604 et qui, pour chaque signature de signal de référence ayant au moins une clé du second ensemble en commun avec le premier ensemble de clés de la signature du signal de requête, constitue un troisième ensemble de clés comprenant les clés du second ensemble en commun avec le premier ensemble, chaque clé du troisième ensemble étant associée à la fois à l'instant de transition musicale associé à la clé dans la signature du signal de référence et à l'instant de transition musicale associé à la clé dans la signature du signal de requête. L'unité de constitution 603 est apte à mettre en œuvre l'étape 303 du procédé selon la figure 3.

Le dispositif 600 comprend en outre une unité de détermination 605 d'un score en fonction d'une correspondance entre chaque paire d'instant de transition musicale associée à une clé du troisième ensemble, qui est apte à mettre en œuvre l'étape 304 du procédé selon la figure 3.

Le dispositif 600 comprend une unité de sélection 606 d'au moins une signature de signal de référence ayant l'un des scores les plus élevés, en tant que signature d'un signal de référence candidat pour identifier le signal de requête. L'unité de sélection 606 est apte à mettre en œuvre l'étape 305 du procédé selon la figure 3.

Le dispositif comprend en outre une unité de post-traitement 607 apte à mettre en œuvre l'étape 306 du procédé selon la figure 3, afin de déterminer parmi un ensemble de signaux de référence candidats, le signal audio de référence identifiant le signal audio de requête. L'unité 607 peut en outre ne retourner aucun résultat lorsque le score le plus élevé d'un signal audio de référence est inférieur au seuil introduit précédemment.

Le dispositif 600 comprend une unité de transmission 608 pour retourner le signal audio de référence identifiant le signal audio de requête, à un terminal utilisateur ou à une unité d'affichage du dispositif 600, non représentée sur la figure 6.

5 L'invention prévoit également le cas où le dispositif 600 ne comprend pas l'unité 306. Dans ce cas, l'unité de transmission 608 transmet les signaux de requêtes candidats à une entité distante apte à comprendre l'unité 306, ce qui est particulièrement avantageux dans les cas où le dispositif 600 possède des capacités de calcul réduites.

10

A noter que le dispositif 500 selon la figure 5 et le dispositif 600 selon la figure 6 peuvent être regroupés au sein d'un même dispositif global (lorsque le dispositif 500 remplace l'unité 602 du dispositif 600). Le dispositif global peut alors être utilisé dans un terminal utilisateur ou dans un serveur distant accessible depuis un terminal utilisateur. Dans ce dernier cas, le terminal

15 utilisateur transmet au serveur distant un signal de requête.

La présente invention prévoit également un programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé illustré en figure

20 1, lorsque ce programme est exécuté par un processeur, et un programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé illustré en figure 3, lorsque ce programme est exécuté par un processeur.

25

REVENDICATIONS

1. Procédé de génération d'une signature d'un signal audio musical d'une durée donnée, le procédé comprenant les étapes suivantes :

- 5 - modélisation (104) dudit signal audio musical pour obtenir, pour chaque bande de fréquences d'un ensemble de n bandes de fréquences, un diagramme représentant l'énergie du signal audio pour ladite bande de fréquences, en fonction du temps pendant ladite durée donnée ;
- détermination (103) d'instants t_k de transition musicale du signal audio
10 pendant la durée donnée ;
- association (105) de chaque instant t_k de transition musicale avec une information locale comprenant un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de
15 transition musicale et/ou un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale;
- détermination (106), en fonction de l'information locale associée à chaque
20 instant t_k de transition musicale, d'une clé associée audit instant de transition musicale, les clés déterminées formant un premier ensemble de clés du signal audio ;
- génération (107) d'une signature du signal audio musical comprenant des couples, chaque couple comprenant une clé du premier ensemble de clés et un
25 'instant t_k de transition musicale associé.

2. Procédé selon la revendication 1, dans lequel les clés du premier ensemble de clés sont obtenues par quantification des vecteurs de n valeurs représentatives respectivement de l'énergie du signal dans chaque bande de
30 fréquences de l'ensemble de n bandes de fréquences dans une période comprise entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale ou par quantification des vecteurs de n valeurs représentatives respectivement de l'énergie du signal dans chaque bande de

fréquences de l'ensemble de n bandes de fréquences dans une période comprise entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale.

5 3. Procédé selon la revendication 2, dans lequel chaque clé du premier ensemble de clés est un vecteur de n bits, chaque bit d'index i étant déterminé par comparaison entre la valeur d'énergie d'index i dans le vecteur de n valeurs à partir duquel ladite clé est déterminée, et la moyenne des valeurs dudit vecteur de n valeurs.

10

4. Procédé selon la revendication 3, dans lequel si la valeur d'énergie d'index i dans le vecteur de n valeurs à partir duquel la clé est déterminée est supérieure à la moyenne des valeurs dudit vecteur de n valeurs, alors le bit d'index i est égal à 1, le bit d'index i étant égal à 0 sinon.

15

5. Procédé selon l'une des revendications précédentes, dans lequel les instants t_k de transition musicale correspondent aux maxima locaux d'une fonction d'onset calculée à partir du signal audio musical.

20

6. Procédé d'identification d'un signal musical, de requête, parmi des signaux musicaux, de référence, le signal de requête et les signaux de référence étant représentés par des signatures respectives obtenues par la mise en œuvre du procédé selon l'une des revendications 1 à 5,

25

chaque signature de signal de référence comprenant ainsi un second ensemble de clés auxquelles sont associés des instants respectifs τ_k de transition musicale, ledit procédé comprenant les étapes :

30

- pour chaque signature de signal de référence ayant au moins une clé du second ensemble en commun avec le premier ensemble de clés de la signature du signal de requête, constitution (303) d'un troisième ensemble de clés comprenant la au moins une clé du second ensemble en commun avec le premier ensemble, chaque clé du troisième ensemble étant associée à la fois à l'instant de transition musicale associé à la clé dans la signature du signal de référence, et à l'instant de

transition musicale associé à la clé dans la signature du signal de requête ;

- détermination (304) d'un score en fonction d'une correspondance entre chaque paire d'instantants de transition musicale associée à une clé du troisième ensemble ; et
- sélection (305) d'au moins une signature de signal de référence ayant l'un des scores les plus élevés, en tant que signature d'un signal de référence candidat pour correspondre au signal de requête.

7. Procédé selon la revendication 6, dans lequel on sélectionne un jeu de signatures de signaux de référence ayant des scores les plus élevés, et on applique un post-traitement (306) :

- à des signaux de référence candidats, représentés par les signatures dudit jeu, et
 - au signal de requête,
- et comprenant une étape d'identification du signal de requête comme l'un des signaux de référence.

8. Procédé selon l'une des revendications 6 et 7, dans lequel la détermination (304) du score d'une signature de signal de référence comprend les étapes suivantes:

- détermination d'une fonction affine mettant en correspondance un plus grand nombre d'instantants de transition musicale du signal de référence associés aux clés du troisième ensemble avec les instantants de transition musicale du signal de requête associés aux clés respectives du troisième ensemble ;
- dans lequel le score attribué à la signature du signal de référence est égal audit plus grand nombre.

9. Procédé selon la revendication 8, dans lequel la fonction affine est obtenue par application d'une transformée de Hough à l'ensemble de paires d'instantants de transition musicale associés aux clés du troisième ensemble.

10. Procédé selon l'une des revendications 6 à 9, dans lequel la détermination

(304) du score d'un signal audio de référence comprend les étapes suivantes:

- construction d'un nuage de points (403), chaque point correspondant à une clé du troisième ensemble, l'abscisse (401) du point étant l'instant de transition musicale du signal de référence associé à la clé du troisième ensemble et l'ordonnée (402) du point correspondant à l'instant de transition musicale du signal de requête associé à la clé du troisième ensemble ;
- détermination d'une droite (404) reliant un maximum de points du nuage de points construit ;
- détermination du score de la signature du signal de référence comme étant égal au nombre de points reliés par ladite droite déterminée dans le nuage de points.

11. Procédé selon la revendication 8 ou 9, dans lequel le post-traitement (306) est appliqué successivement pour chaque signal de référence correspondant à une signature sélectionnée, et au signal de requête modifié en fonction de la fonction affine déterminée pour l'obtention du score de la signature du signal de référence.

12. Procédé selon la revendication 11, dans lequel le signal de requête est modifié par mise à l'échelle dudit signal de requête en fonction d'une pente (405) de la fonction affine et par recadrage temporel dudit signal de requête en fonction d'une ordonnée à l'origine (406) de la fonction affine.

13. Programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé selon l'une des revendications 1 à 5, lorsque ce programme est exécuté par un processeur.

14. Programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé selon l'une des revendications 6 à 12, lorsque ce programme est exécuté par un processeur.

15. Dispositif de génération d'une signature d'un signal audio musical d'une durée donnée, le dispositif (500) comprenant :

- une unité de modélisation (503) dudit signal audio musical pour obtenir, pour chaque bande de fréquences d'un ensemble de n bandes de fréquences, un diagramme représentant l'énergie du signal audio pour ladite bande de fréquences, en fonction du temps pendant ladite durée donnée ;
- 5 - une première unité de détermination (502) d'instants t_k de transition musicale du signal audio pendant la durée donnée ;
 - une unité d'association (505) de chaque instant t_k de transition musicale avec une information locale comprenant un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes
 - 10 obtenus entre l'instant t_k de transition musicale et un instant suivant t_{k+1} de transition musicale et/ou un vecteur de n valeurs représentatives respectivement de l'énergie du signal audio dans chacun des n diagrammes obtenus entre l'instant t_k de transition musicale et un instant précédent t_{k-1} de transition musicale;
- 15 - une deuxième unité de détermination (506), en fonction de l'information locale associée à chaque instant t_k de transition musicale, d'une clé associée audit instant de transition musicale, les clés déterminées formant un premier ensemble de clés du signal audio ;
 - une unité de génération (507) d'une signature du signal audio musical
 - 20 comprenant des couples, chaque couple comprenant une clé du premier ensemble de clés et un instant t_k de transition musicale associé.

16. Dispositif d'identification d'un signal musical, de requête, parmi des signaux musicaux, de référence, le signal de requête et les signaux de référence étant

25 représentés par des signatures respectives obtenues par la mise en œuvre du procédé selon l'une des revendications 1 à 5,

chaque signature de signal de référence comprenant ainsi un second ensemble de clés auxquelles sont associés des instants respectifs τ_k de transition musicale, ledit dispositif (600) comprenant :

- 30 - une unité de constitution (603), pour chaque signature de signal de référence ayant au moins une clé du second ensemble en commun avec le premier ensemble de clés de la signature du signal de requête, d'un troisième ensemble de clés comprenant la au moins une clé du second

ensemble en commun avec le premier ensemble, chaque clé du troisième ensemble étant associée à la fois à l'instant de transition musicale associé à la clé dans la signature du signal de référence et à l'instant de transition musicale associé à la clé dans la signature du signal de requête ;

5

- une unité de détermination (604) d'un score en fonction d'une correspondance entre chaque paire d'instant de transition musicale associée à une clé du troisième ensemble ; et
- une unité de sélection (605) d'au moins une signature de signal de référence ayant l'un des scores les plus élevés, en tant que signature d'un signal de référence candidat pour identifier le signal de requête.

10

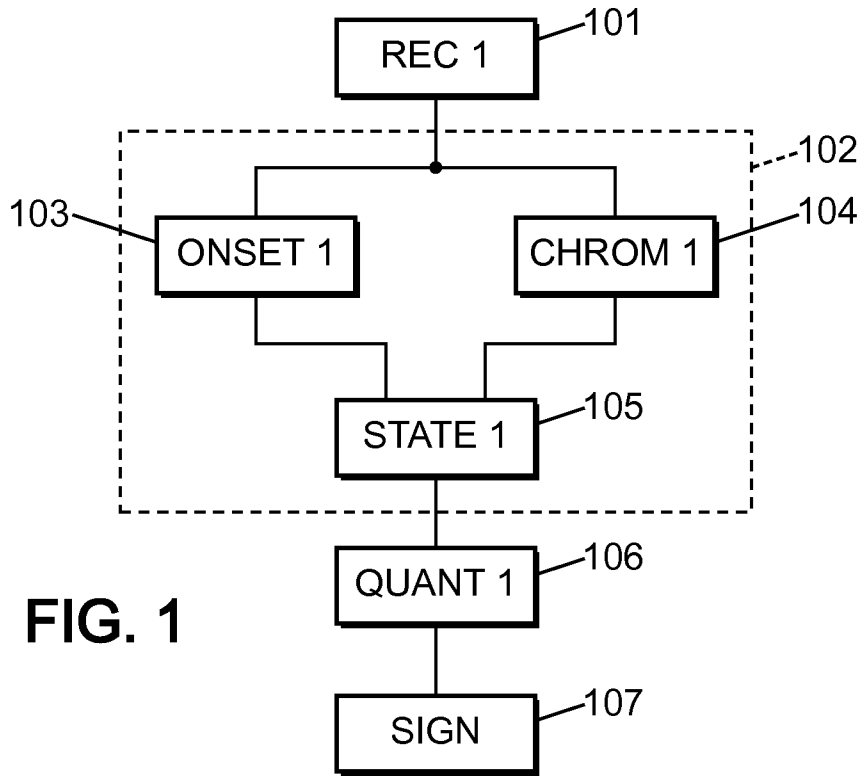


FIG. 1

FIG. 2

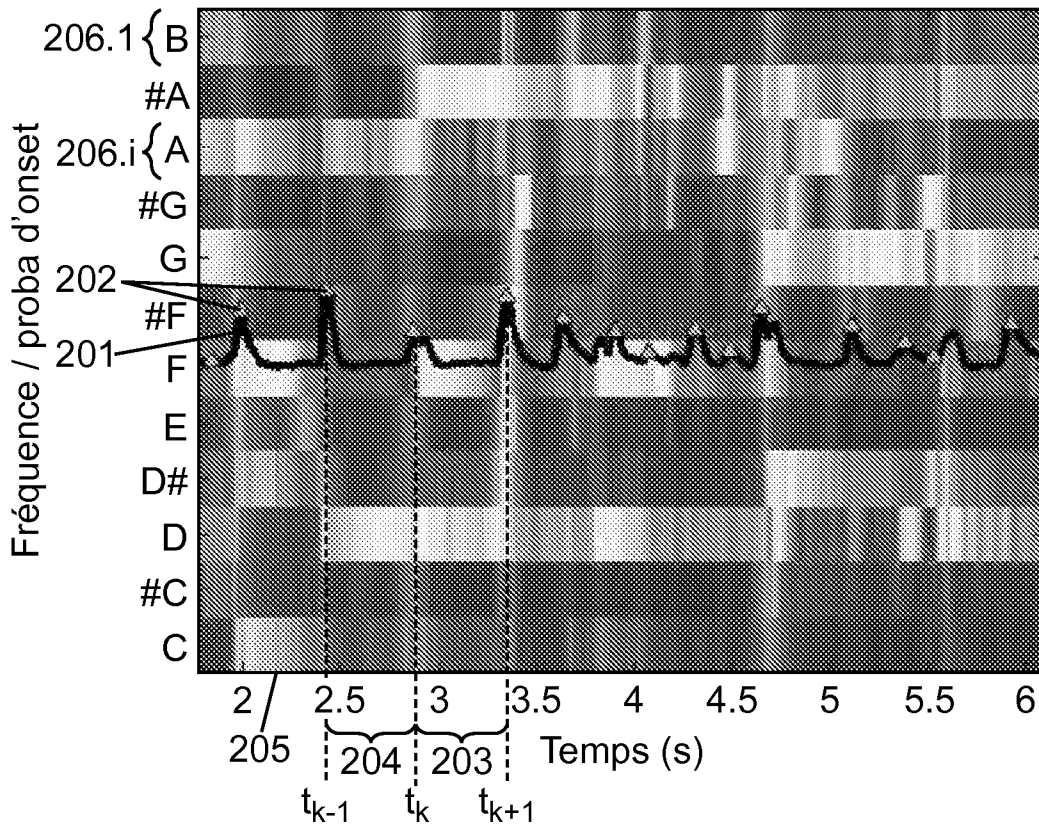


FIG. 3

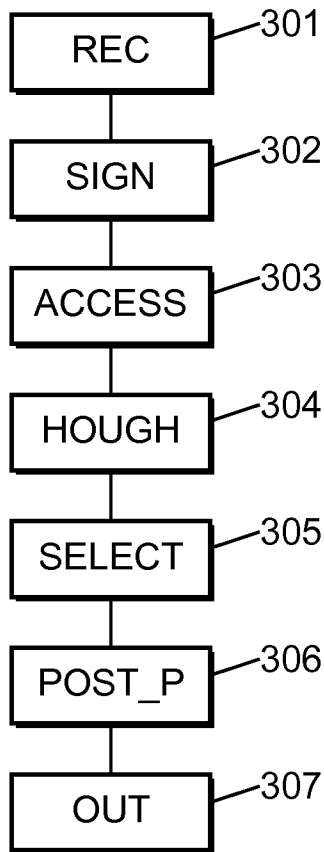
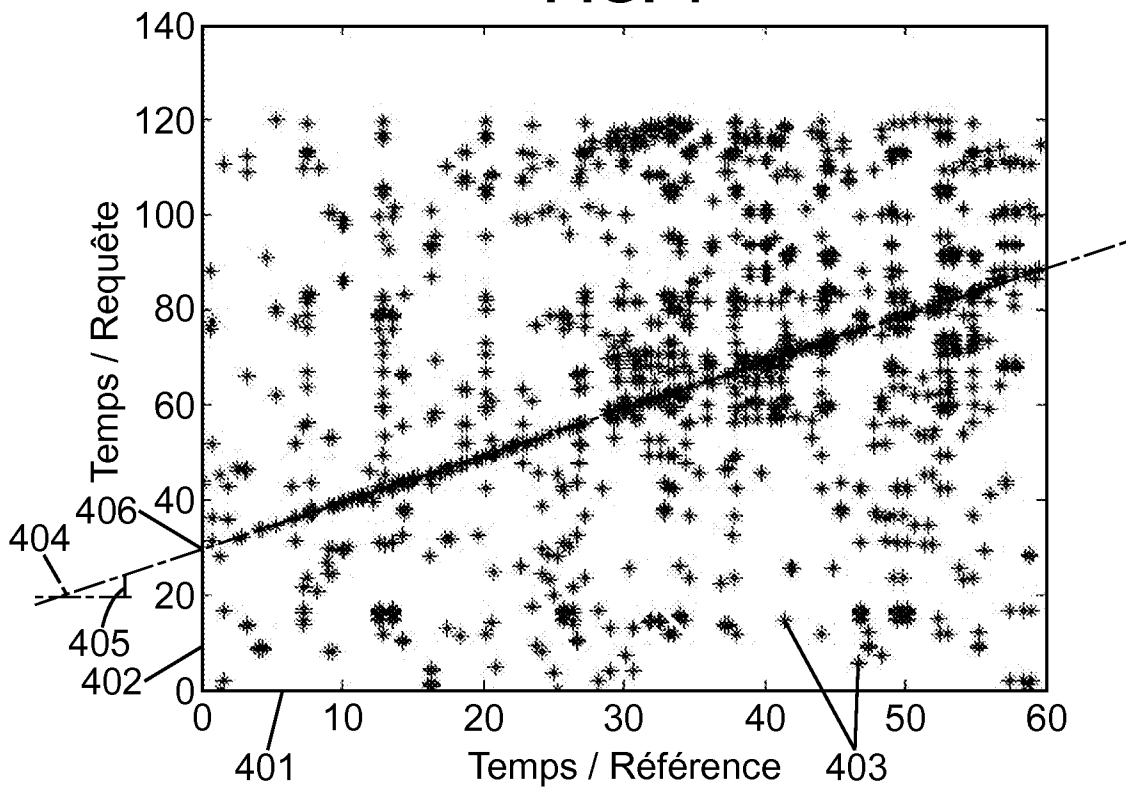


FIG. 4



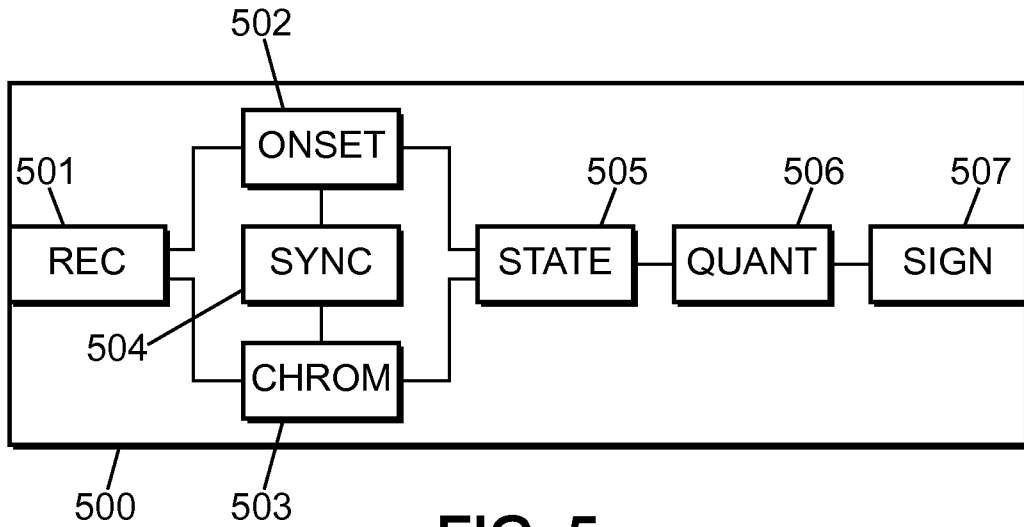


FIG. 5

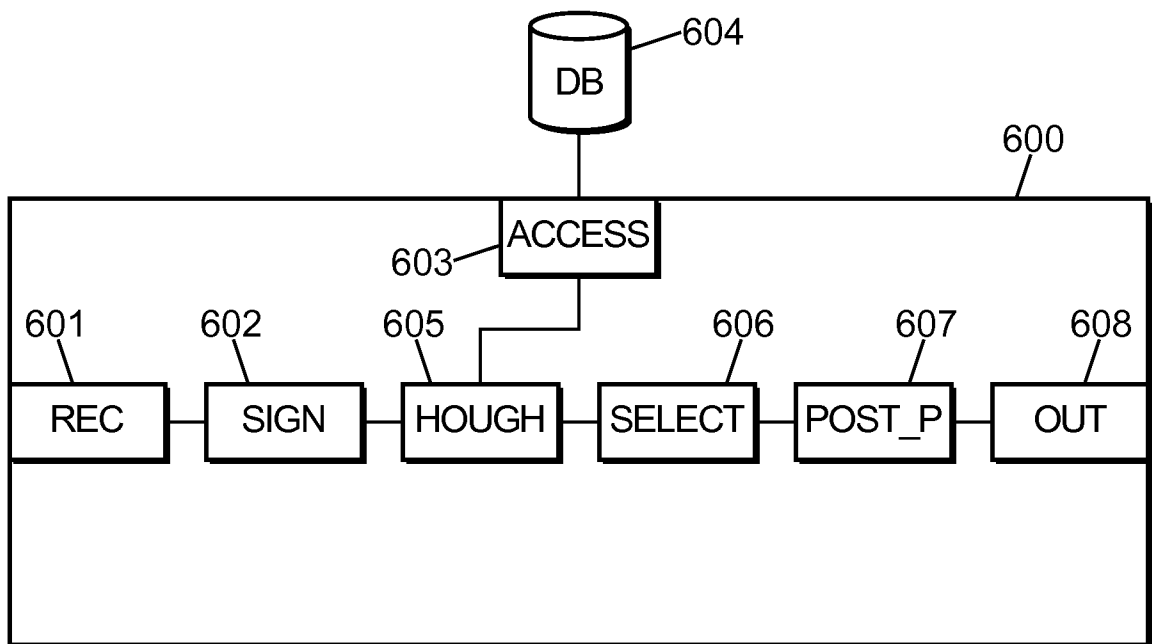


FIG. 6