

19 RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
COURBEVOIE

11 N° de publication :  
(à n'utiliser que pour les  
commandes de reproduction)

3 037 760

21 N° d'enregistrement national : 15 55577

51 Int Cl<sup>8</sup> : H 04 N 21/854 (2016.01)

12

DEMANDE DE BREVET D'INVENTION

A1

22 Date de dépôt : 18.06.15.

30 Priorité :

43 Date de mise à la disposition du public de la  
demande : 23.12.16 Bulletin 16/51.

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) : ORANGE Société anonyme — FR.

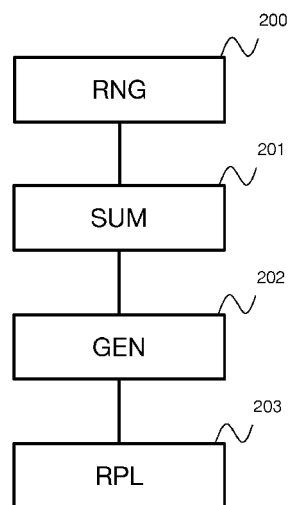
72 Inventeur(s) : MONCOMBLE GHISLAIN et RONDET  
PATRICK.

73 Titulaire(s) : ORANGE Société anonyme.

74 Mandataire(s) : ORANGE SA Société anonyme.

54 PROCÉDE ET DISPOSITIF DE SUBSTITUTION D'UNE PARTIE D'UNE SEQUENCE VIDEO.

57 L'invention concerne un procédé de substitution dans  
un contenu audiovisuel d'une partie du contenu définie par  
une plage temporelle, par une séquence audiovisuelle de  
substitution, le procédé étant caractérisé en ce que la sé-  
quence de substitution est représentative de la partie à rem-  
placer. L'invention concerne également un dispositif et un  
terminal mettant en oeuvre l'invention.



FR 3 037 760 - A1



## PROCÉDÉ ET DISPOSITIF DE SUBSTITUTION D'UNE PARTIE D'UNE SEQUENCE VIDEO

## DOMAINE TECHNIQUE

La présente invention se rapporte au domaine des services audiovisuels numériques et concerne en particulier un procédé de montage vidéo automatisé pour les formations et présentations en ligne.

## ART ANTÉRIEUR

Le domaine de la formation à distance subit d'importantes mutations depuis la démocratisation d'Internet. La formation à distance, ou enseignement à distance, désigne l'ensemble des dispositifs de formation à distance, qu'il s'agisse de cours par correspondance, de MOOC (Massive Open Online Course) ou de formations en ligne. Elle s'applique tant à la formation continue qu'à la formation initiale, tant de manière individuelle que collective. Ces formations utilisent de plus en plus les outils numériques et les moyens de l'Internet.

Si la formation à distance peut répondre à des situations d'éloignement géographique avec la formation souhaitée, elle trouve également une utilité dans des situations de handicap, de contrainte horaire ou financière. Ce type de formation concerne un public très large composé d'élèves, d'étudiants, de personnes en période d'inactivité ou de personnes en activité.

Le caractère flexible et ouvert de ces formations offre aux apprenants la possibilité de gérer de manière autonome le temps consacré à l'apprentissage, et les laisse libre d'entrer ou de sortir du dispositif. En outre, une formation est "ouverte" dans la mesure où elle est facilement accessible matériellement. Enfin, la formation "ouverte", "flexible", se veut accessible au plus grand nombre, et à toute personne souhaitant suivre une formation, contrairement à des enseignements inscrits dans un cursus académique normatif. Le tutorat, l'apprentissage et des séances de cours traditionnels en salle peuvent y être associés.

Les MOOC sont des cours en ligne ouverts à tous. Ce type d'enseignement rend accessible à tous les enseignements dispensés par les meilleures écoles et universités du monde, transformant ainsi profondément l'accès au savoir et à la connaissance dans la plupart des domaines.

Le succès des MOOC est lié à leur mode de fonctionnement très simple : les enseignants donnent des cours magistraux qui sont filmés avant d'être diffusés via le réseau Internet. Les cours sont le plus souvent découpés en tronçons de courte durée et associés à un forum de discussion permettant aux élèves d'obtenir des éclaircissements, d'approfondir un sujet ou de poser des questions. Lors de la

diffusion des cours en direct, ou lors de sessions de questions-réponses organisées à la suite d'un cours, les enseignants peuvent répondre en direct aux questions des étudiants. Cette interactivité entre étudiants et enseignants est une caractéristique forte des MOOC et participe beaucoup à leur succès.

- 5 Les MOOCS sont enregistrés et accessibles sur des sites dédiés, sur les sites des universités ou des écoles au travers d'un simple navigateur Internet. De cette façon, les cours sont consultables depuis n'importe où dans le monde et à tout moment.

10 L'efficacité d'un cours, qu'il soit présentiel ou qu'il s'agisse d'un MOOC, dépend pour beaucoup de la capacité de l'enseignant à capter l'attention de son auditoire. Dans le cas des MOOC, une attention particulière doit être portée sur la qualité de réalisation des vidéos ainsi que sur la fréquence et la qualité des échanges entre enseignants et élèves.

15 Bien que l'interactivité soit un élément important dans l'efficacité pédagogique d'un cours en ligne tel qu'un MOOC, des interventions multiples dans un cours peuvent s'avérer contre-productives. Le cours étant suivi par un large public, les interventions peuvent être nécessaires pour la bonne compréhension de certains mais s'avérer pénibles pour d'autres. Ainsi, il peut être avantageux lors de la mise en ligne de la vidéo d'un cours, de supprimer de la vidéo les parties correspondant à des interventions telles que des questions d'élèves et réponses de l'enseignant, tout en laissant ces séquences accessibles aux élèves qui désirent les consulter.

20 La mise en ligne de tels cours nécessite souvent un montage vidéo préalable. Par exemple, lorsque plusieurs caméras sont mises en œuvre, l'intervention d'un technicien peut être nécessaire pour sélectionner les prises de vues. Ce montage ayant un coût, et les cours étant gratuits, il convient d'automatiser autant que possible cette étape de montage. Pour cela, différentes solutions techniques existent. Par exemple, il existe des techniques basées sur la détection d'activité vocale permettant de basculer d'une caméra à une autre selon la personne qui parle. D'autres solutions techniques permettent d'effectuer automatiquement un cadrage sur un visage ou encore de suivre une personne se déplaçant. De telles dispositions techniques permettent de réduire le coût de production de la vidéo en évitant la mise en place d'une régie ou la réalisation d'un montage avant la mise en ligne d'un cours.

30 Dans certains cas, les techniques existantes ne permettent pas d'automatiser complètement l'étape de montage. En particulier, l'intervention d'un technicien reste nécessaire pour repérer les parties du cours correspondant à des interventions et les extraire de la vidéo principale.

On connaît malgré tout dans le domaine de l'audiovisuel des techniques permettant par exemple de détecter le début et la fin d'une séquence publicitaire afin de la supprimer d'un contenu. La plage correspondant à la séquence peut alors être déterminée par la détection d'un changement de niveau sonore ou d'une image particulière dans le flux vidéo. Toutefois, ces techniques ne sont pas satisfaisantes lorsqu'elles sont utilisées dans le cadre d'un MOOC. En effet, les transitions de début et de fin d'une intervention lors d'un cours sont beaucoup plus vagues et ne peuvent être détectées par de telles techniques. D'autre part, dans le cadre d'un MOOC, il est important de garder une trace sur le moment de l'intervention et son contenu.

Ainsi, il existe un besoin pour une solution technique permettant de réaliser automatiquement un montage à partir d'une vidéo telle qu'une vidéo d'un cours en ligne de type MOOC, et en particulier de supprimer certains passages tout en conservant une information dans le contenu vidéo sur les passages supprimés.

#### RÉSUMÉ DE L'INVENTION

À cet effet, l'invention concerne un procédé de substitution dans un contenu audiovisuel d'une partie du contenu définie par une plage temporelle, par une séquence audiovisuelle de substitution, le procédé étant remarquable en ce que la séquence de substitution est représentative de la partie à remplacer.

De cette façon, il est possible d'extraire une partie d'un contenu vidéo, comme par exemple une partie correspondant à une intervention d'un élève lors d'un cours de type MOOC, et de remplacer cette partie par une autre séquence plus courte consistant par exemple en une transition entre la partie précédente la partie extraite et la partie suivant la partie extraite. Cette séquence de transition est telle qu'elle est représentative de la partie remplacée. Par exemple, la séquence de transition peut comporter un message composé de mots clefs relatifs à la partie remplacée, ledit message pouvant par exemple être vocalisé lors de ladite séquence de transition, ou encore être inclus dans au moins une image générée à cet effet, image affichée durant ladite transition. Ainsi, un utilisateur visionnant la vidéo est averti qu'une séquence a été extraite et dispose d'une information quant au contenu ayant été remplacé.

Selon une réalisation particulière, le procédé est tel que la séquence de substitution est générée à partir d'une analyse sémantique de la partie du contenu définie par la plage temporelle.

L'analyse sémantique de la séquence remplacée permet d'en extraire des concepts qui permettent de générer une séquence de substitution qui soit représentative de la séquence extraite.

Selon une réalisation particulière, le procédé est tel que les instants de début et de fin de la plage temporelle sont déterminés par la détection d'un premier et d'un second évènement particulier dans le flux audiovisuel.

- 5 La plage temporelle correspondant à la partie à remplacer est identifiée automatiquement par détection d'un évènement de début et d'un évènement de fin. De cette façon, l'intervention d'un technicien n'est pas nécessaire pour déterminer la ou les parties à remplacer dans le contenu vidéo. Une séquence vidéo correspondant par exemple à un cours en ligne peut ainsi être mise en ligne plus rapidement et à moindre coût.

10

Selon une réalisation particulière, le procédé est tel que la partie extraite est mémorisée et en ce que la séquence de remplacement comprend un lien interactif adapté pour déclencher la restitution de la partie extraite.

- 15 Lorsqu'une partie d'un contenu audiovisuel est remplacée, celle-ci est mémorisée de façon à pouvoir être consultée en cas de besoin. La séquence de remplacement peut alors comporter un lien vers cette partie. De cette façon, lorsqu'un utilisateur visionne le contenu, il peut facilement consulter la séquence remplacée s'il le souhaite. Par exemple, la séquence de remplacement peut comprendre un lien interactif de type lien hypertexte pointant vers l'emplacement de stockage de la partie remplacée.

20

Selon une réalisation particulière, le procédé est tel que le contenu audiovisuel résultant de la substitution comprend un index indiquant le début de la séquence de remplacement.

- 25 Lorsqu'une séquence est remplacée dans un contenu audiovisuel, son emplacement est indexé au niveau du contenu audiovisuel à partir duquel la séquence est extraite. De cette façon, un utilisateur peut accéder à une liste des séquences remplacées dans le contenu. L'index peut également servir à inscrire une marque correspondant aux parties remplacées sur une ligne de temps affichant de manière classique l'avancement de la lecture. Un utilisateur peut ainsi facilement repérer les plages temporelles remplacées dans le contenu et y accéder en y déplaçant un curseur de lecture.

Selon un mode de réalisation particulier, le procédé est tel qu'au moins un des premier et second évènements particuliers est identifié par reconnaissance d'au moins un mot clef dans une composante audio du contenu audiovisuel.

Le procédé permet de sélectionner automatiquement une séquence à remplacer en se basant sur un ou plusieurs mots clefs détectés dans une composante audio du contenu. Par exemple, le début d'une intervention de type question/réponse entre un élève et un enseignant dans un cours de type MOOC peut être détecté lorsque l'enseignant prononce, une expression du type « vous pouvez poser votre question ». Une telle expression peut être détectée par des techniques classiques de reconnaissance vocale. De même, la fin d'une intervention peut être détectée par la reconnaissance d'une expression du type « reprenons le cours ». L'enseignant peut ainsi contrôler le début et la fin des séquences à remplacer en prononçant au moment opportun des mots ou expressions appropriées.

Selon un mode de réalisation particulier, le procédé est tel qu'au moins un des premier et second évènements particuliers est identifié par un changement de locuteur dans une composante audio du contenu audiovisuel.

En détectant un changement de locuteur, le procédé peut détecter le début d'une séquence à remplacer. Par exemple, dans un cours de type MOOC, le procédé peut déterminer qu'un élève prend la parole par des techniques d'analyses vocales connues, afin de déterminer le début d'une plage temporelle correspondant à une séquence à remplacer. Le procédé permet ainsi de déterminer automatiquement le début et/ou la fin d'une plage temporelle à remplacer.

Selon un mode de réalisation particulier, le procédé est tel qu'au moins un des premier et second évènements particuliers est identifié par un changement de source audio dans une composante audio du contenu audiovisuel.

De cette façon, le procédé permet de détecter le début ou la fin d'une plage temporelle selon la source audio utilisée. Par exemple, dans un amphithéâtre dont les bancs sont équipés de microphones individuels, le procédé peut déterminer qu'un individu prend la parole lorsque qu'un autre microphone que celui de l'enseignant est activé. Le procédé permet ainsi de déterminer automatiquement le début et/ou la fin d'une plage temporelle à remplacer. Ce mode de réalisation

est particulièrement opportun lorsqu'au moins une partie du montage vidéo est réalisée lors du tournage initial du document audiovisuel.

5 Selon un mode de réalisation particulier, le procédé est tel qu'au moins un des premier et second évènements particuliers est identifié par analyse d'image dans une composante vidéo du contenu audiovisuel.

10 Le début et/ou la fin de la plage temporelle correspondant à une partie du contenu à remplacer est déterminée par une analyse des images de la composante vidéo du contenu audiovisuel. Par exemple, l'invention peut mettre en œuvre un algorithme de reconnaissance de geste afin de détecter par exemple un élève levant la main dans une assemblée d'étudiant afin de demander la parole, ou encore une reconnaissance de visage du locuteur pour distinguer le visage du présentateur des autres visages dont les élèves. Le procédé permet ainsi de détecter automatiquement le début d'une partie de contenu à remplacer .

15 Selon une réalisation particulière, le procédé est tel que la séquence de substitution est représentative de la séquence extraite en ce qu'elle comprend un résumé de la partie extraite.

20 La séquence de substitution est composée suite à une analyse sémantique de la partie à remplacer permettant la génération d'un résumé. Par exemple, la composante audio du contenu est soumise à un algorithme de reconnaissance de la parole afin de déterminer des mots clefs représentatifs du contenu substitué. Ces mots clefs peuvent être intégrés à la séquence de substitution de manière à donner des informations synthétiques sur la partie remplacée. Le résumé peut également consister en une reprise sous forme textuelle d'une question posée par un élève. Le résumé peut être intégré dans la séquence vidéo de remplacement, sous la forme par exemple d'un message vocalisé lors de ladite séquence de transition, ou d'un texte inclus dans au moins une image de la séquence générée  
25 à cet effet. De cette façon, un étudiant visionnant un cours en ligne peut immédiatement avoir une idée de l'intérêt que peut constituer une intervention avant de décider ou non de la visionner.

L'invention se rapporte également à un dispositif de substitution dans un contenu audiovisuel, d'une partie du contenu définie par une plage temporelle, par une séquence audiovisuelle de substitution,  
30 le dispositif comportant une unité de traitement adaptée pour générer une séquence de substitution représentative de la partie dudit contenu.

Dans un mode de réalisation particulier, le dispositif comprend :

- Un module de détection des instants de début et de fin de la plage temporelle par détection d'un premier et d'un second évènement particulier dans le flux audiovisuel,
- 5 - Un module d'extraction de la partie du contenu audiovisuel comprise entre le début et la fin de la plage temporelle,
- Un module d'analyse sémantique de la partie extraite et générer une séquence de substitution à partir du résultat de l'analyse,
- Un module d'insertion de la séquence de substitution à la place de la partie extraite.

10 L'invention concerne aussi un terminal comportant un dispositif de substitution tel que décrit ci-dessus.

L'invention concerne également un programme d'ordinateur comportant les instructions pour l'exécution du procédé de substitution, lorsque le programme est exécuté par un processeur.

15 L'invention concerne aussi un support d'informations lisible par un processeur sur lequel est enregistré un programme d'ordinateur comprenant des instructions pour l'exécution des étapes du procédé de substitution.

Les différents modes ou caractéristiques de réalisation précités peuvent être ajoutés indépendamment ou en combinaison les uns avec les autres, aux étapes du procédé substitution.

Les terminaux, dispositifs, programmes et supports d'information présentent au moins des avantages analogues à ceux conférés par le procédé de substitution décrit ci-dessus.

20

#### LISTE DES FIGURES

D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description suivante d'un mode de réalisation particulier, donné à titre de simple exemple illustratif et non limitatif, et des dessins annexés, parmi lesquels :

- 25 - Les **figures 1a, 1b et 1c** représentent de manière schématique une séquence vidéo dans laquelle est une partie remplacée selon un mode particulier de réalisation.
- La **figure 2** illustre les différentes étapes du procédé de substitution selon un mode de réalisation particulier de l'invention.
- La **figure 3** représente de manière simplifiée un dispositif de substitution selon une  
30 réalisation particulière de l'invention.



## DESCRIPTION DÉTAILLÉE

L'invention va maintenant être décrite en prenant l'exemple d'un cours dispensé par un enseignant devant une assemblée d'étudiant, le cours étant filmé afin d'être diffusé en ligne sous forme de MOOC. L'invention peut toutefois trouver une application dans d'autres contextes. Par exemple, 5 l'invention peut être utilisée afin de remplacer une séquence publicitaire détectée dans un programme audiovisuel par un écran d'information énumérant sous forme de liste les différents spots publicitaires contenus dans la séquence originale, l'écran d'information pouvant en outre contenir des liens vers les différents spots, ou encore dans une vidéo de présentation de produit.

La **figure 1a** représente une séquence vidéo 100 enregistrée lors d'un cours de type MOOC dispensé 10 par un enseignant devant une assemblée d'élèves. Dans cet exemple, la vidéo comprend une composante vidéo 101 et une composante audio 102.

Lors d'une première étape 200 représentée sur la **figure 2**, le contenu audiovisuel est analysé de façon à détecter un premier et un second événements particuliers. Ces premiers et seconds événements particuliers constituent le début et la fin d'une plage temporelle correspondant à une 15 partie du contenu à remplacer. Dans cet exemple, l'analyse vise à déterminer une plage temporelle correspondant à une intervention d'un élève durant le cours. Cette analyse peut être mise en œuvre selon différentes techniques.

Selon un premier mode de réalisation particulier, la composante audio est analysée de façon à repérer un ou plusieurs mots clefs signalant le début ou la fin d'une intervention. Pour cela, le 20 procédé peut utiliser un algorithme de reconnaissance vocale adapté. De manière classique, l'algorithme de reconnaissance vocale peut être entraîné au préalable pour reconnaître des mots clefs particuliers. Par exemple, l'algorithme peut être entraîné à partir d'un corpus d'utilisateurs prononçant différents mots clefs. Dans un deuxième temps, un filtre adapté pour filtrer les fréquences vocales et supprimer les bruits de fond peut être appliqué à la composante audio de la 25 séquence. Enfin, la composante audio est présentée en entrée de l'algorithme afin de détecter des mots clefs signalant le début et la fin d'une intervention. De tels mots clefs peuvent par exemple être prononcés par l'enseignant lorsqu'il donne la parole à un élève et lorsqu'il reprend le fil de son cours. Il est ainsi possible de délimiter une partie du contenu correspondant à une interruption du cours.

Selon une réalisation particulière, la composante audio est analysée de façon à repérer un 30 changement de locuteur. Une telle disposition permet de détecter par exemple un instant où un individu autre que l'enseignant prend la parole. Pour cela, un algorithme classique adapté à la reconnaissance du timbre de la voix peut être utilisé. Cet algorithme peut être entraîné au préalable

pour reconnaître la voix de l'enseignant. Ainsi, lorsque la composante audio de la séquence est présentée en entrée de l'algorithme, celui-ci peut détecter un instant correspondant à la prise de parole par un individu autre que l'enseignant. Une telle disposition permet la détection d'un instant de début d'une interruption de cours.

5 Selon un mode de réalisation particulier, la composante audio de la séquence est analysée de façon à détecter un changement de source audio caractéristique d'un changement de locuteur. Par exemple, lorsque l'aménagement d'un amphithéâtre prévoit un microphone pour l'enseignant et au moins un microphone destiné aux étudiants, le procédé interprète un changement de microphone comme un changement de locuteur permettant de signaler le début et/ou la fin d'une intervention. Pour cela, le  
10 procédé peut mettre en œuvre un algorithme adapté pour reconnaître dans le signal audio une signature caractéristique d'un microphone particulier. Pour cela, l'algorithme peut subir au préalable un apprentissage à partir de signaux audio capturés à partir de différents microphones afin de déterminer une signature caractéristique de chacun des microphones. Par exemple, la signature peut comprendre la largeur de la bande de fréquence capturée par un microphone, un gain ou un niveau  
15 de bruit propre au microphone. Il est ainsi possible, suite à l'analyse de la composante audio par un tel algorithme, de déterminer un instant dans la séquence correspondant à l'activation ou la désactivation d'un microphone particulier. Une telle disposition permet la détection d'un instant de début ou de fin d'une interruption du cours.

Selon un mode de réalisation particulier, au moins un évènement marquant le début ou la fin d'une  
20 plage temporelle correspondant à une partie de la séquence à remplacer est déterminé par une analyse de l'image. Par exemple, l'analyse d'image peut consister à rechercher un changement de plan, un changement de caméra ou encore un geste particulier ou un visage particulier dans la séquence. Pour cela, le procédé peut mettre en œuvre un algorithme de caractérisation des mouvements afin de détecter un geste particulier effectué par exemple par un élève ou un  
25 enseignant participant à un cours de type MOOC. Par exemple, l'algorithme de caractérisation d'un mouvement peut déterminer qu'un élève demande la parole en levant la main dans l'assemblée. Pour cela, une séquence vidéo faisant apparaître l'assistance lors d'un cours peut être présentée en entrée de l'algorithme de façon à ce qu'une analyse des images soit réalisée. À l'issue de cette analyse, l'algorithme détermine différentes plages temporelles correspondant à des interruptions du  
30 cours.

Selon un mode particulier de réalisation, les différentes méthodes d'analyse décrites ci-dessus peuvent être combinées de façon à utiliser par exemple une analyse de la vidéo pour détecter le

début ou la fin d'une plage temporelle correspondant à une séquence à remplacer, et une analyse de la composante audio pour en détecter la fin.

Suite à cette analyse, un premier évènement 103 et un second évènement 104 peuvent être détectés dans la séquence 100. Ces deux évènements permettent de définir une plage temporelle située entre 5 l'évènement 103 et 104 correspondant à une partie de la séquence à remplacer.

La **Figure 1b** illustre la séquence 100 de laquelle a été extraite une partie 105 correspondant à la plage temporelle définie par les évènements 103 et 104. La partie 105 correspond par exemple à l'intervention d'un élève pendant un cours suivie de d'une réponse de l'enseignant. Cette partie est analysée à l'étape 201 illustrée sur la **figure 1**. Cette analyse est par exemple effectuée par un 10 algorithme de reconnaissance vocale travaillant sur la composante audio de la partie 105 après extraction de la séquence et application d'un filtre de débruitage par exemple. L'algorithme mis en œuvre est adapté pour produire une liste de mots clefs représentatifs de l'intervention de l'élève. Pour cela, l'étape comprend par exemple la mise en œuvre d'un algorithme de conversion de la parole en texte (« speech to text » en anglais, ou TTS) de façon à obtenir le texte de l'intervention. À 15 partir de ce texte, un second algorithme peut par exemple produire un résumé de l'intervention en reprenant par exemple des phrases clefs, la question de l'élève et/ou la réponse apportée par l'enseignant. De tels algorithmes sont connus de l'homme du métier et ne sont pas abordés ici plus en détails.

À partir de cette analyse, une séquence de remplacement 106 représentative de la séquence 105 est 20 générée à l'étape 202. Par exemple, une séquence représentative peut être générée par incrustation vidéo de mots clefs identifiés à l'étape 201 ou d'un résumé. La durée de la séquence générée peut être différente de la durée de la séquence à remplacer. Par exemple, la séquence générée peut être une courte séquence comportant un message indiquant qu'une intervention a eu lieu.

Selon une réalisation particulière, un logiciel de « text mining » est appliqué au texte de 25 l'intervention obtenu à partir d'une reconnaissance vocale, afin de définir une pertinence de l'interruption. De tels algorithmes sont connus de l'homme du métier et ne sont pas abordés ici en détails.

À partir de cette analyse, une séquence de remplacement 106 peut être générée à l'étape 202 en 30 remplacement de la séquence 105, mais sans contenu représentatif de ladite séquence 105 si le contenu n'a pas été jugé pertinent. Par exemple, la séquence 106 est alors une simple transition de plan.

Selon un mode particulier de réalisation, une série de séquences de remplacement 106 peuvent être générées et positionnées successivement à l'étape 202 en remplacement de la séquence 105, dans le cas où une série de contenus représentatifs distincts de ladite séquence 105 ont été identifiés. Par exemple, la séquence 106 générée comprend une succession de messages indiquant chacun qu'une intervention a eu lieu.

En variante, la séquence 106 ne comprend qu'un seul message de remplacement, ledit message indiquant que plusieurs interventions distinctes ont eu lieu.

Selon une réalisation particulière, la séquence de remplacement comporte au moins un lien interactif 107 permettant d'accéder à la séquence remplacée, comme par exemple un lien hypertexte. Dans le cas où plusieurs parties sont remplacées un seule séquence, la séquence peut comporter plusieurs liens interactifs permettant d'accéder aux différentes parties remplacées. De tels liens hypertextes, lorsqu'ils sont sélectionnés par un utilisateur, peuvent déclencher la lecture d'une partie remplacée mémorisée dans un espace de stockage. Ainsi, un utilisateur peut visionner le cours sans qu'il soit interrompu par des interventions et accéder à la séquence remplacée en cas de besoin par une simple interaction sur la séquence de remplacement ou à partir de l'index.

La **figure 1c** représente la séquence vidéo initiale 100 dans laquelle la partie 105 a été remplacée par la séquence générée 106. Ce remplacement est réalisé à l'étape 203 du procédé de substitution illustrée sur la **figure 2**. Le remplacement peut être effectué en utilisant des techniques classiques d'édition vidéo. Selon une réalisation particulière, le contenu audiovisuel résultant de la substitution comprend un index indiquant le début de la séquence de remplacement. Par exemple, des informations d'horodatage relatives à différentes parties remplacées dans une séquence vidéo peuvent être répertoriées dans un index de façon à ce qu'un utilisateur puisse immédiatement avoir accès à l'une des séquences remplacées. Par exemple, lorsque la vidéo est visionnée à partir d'un lecteur multimédia adapté, les différentes entrées de l'index peuvent apparaître sous la forme d'indices visuels intégrés dans une barre de progression de la lecture.

La **figure 3** illustre un dispositif 300 mettant en œuvre le procédé de substitution, selon un mode particulier de réalisation de l'invention. Le dispositif comprend un espace de stockage 301, par exemple une mémoire MEM, une unité de traitement 303 équipée par exemple d'un processeur PROC. L'unité de traitement peut être pilotée par un programme 302, par exemple un programme d'ordinateur PGR, mettant en œuvre le procédé de substitution tel que décrit dans l'invention en référence à la **figure 2**, et notamment les étapes de détermination des instants de début et de fin d'une plage temporelle par détection d'un premier et d'un second évènement particulier dans le flux audiovisuel, d'extraction de la partie du contenu audiovisuel comprise entre le début et la fin de la

plage temporelle, d'analyse sémantique de la partie extraite et de génération d'une séquence de substitution à partir du résultat de l'analyse, et d'insertion de la séquence de substitution à la place de la partie extraite.

5 À l'initialisation, les instructions du programme d'ordinateur 302 sont par exemple chargées dans une mémoire RAM (Random Access Memory en anglais) avant d'être exécutées par le processeur de l'unité de traitement 303. Le processeur de l'unité de traitement 303 met en œuvre les étapes du procédé de substitution selon les instructions du programme d'ordinateur 302.

10 Pour cela, le dispositif comprend, outre la mémoire 301, une unité de traitement adaptée pour déterminer des instants de début et de fin d'une plage temporelle par détection d'un premier et d'un second évènement particulier dans le flux audiovisuel. Un tel traitement peut être réalisé par un module de détection 304 (RNG) mettant en œuvre par exemple un algorithme de caractérisation de gestuelle et/ou un algorithme de reconnaissance de la parole. Le dispositif comprend également une unité de traitement adaptée pour extraire la partie du contenu audiovisuel comprise entre le début et la fin de la plage temporelle identifiée par le module de détection 304. L'extraction peut être  
15 réalisée par un module d'extraction 305 (EXT) mettant en œuvre un algorithme de traitement vidéo capable d'identifier les trames vidéo et les échantillons audio compris dans la plage temporelle afin de les extraire de la séquence. Le dispositif comporte aussi une unité de traitement apte à réaliser une analyse sémantique de la partie extraite, comme par exemple un module 306 (SUM) mettant en œuvre un algorithme adapté pour créer un résumé à partir d'une composante audio d'une vidéo. Le  
20 dispositif comporte également un module de génération 307 (GEN) d'une séquence de substitution à partir du résultat de l'analyse sémantique, le module 307 pouvant correspondre par exemple à une unité de traitement vidéo adaptée pour réaliser l'incrustation d'un résumé ou de mots clefs dans une vidéo. Enfin, le dispositif comprend un module d'insertion 308 (INS) de la séquence de substitution à la place de la partie extraite dans la séquence vidéo.

25 Selon un mode particulier de réalisation, le dispositif peut être intégrée dans un équipement de traitement vidéo ou un terminal tel qu'un ordinateur.

## REVENDICATIONS

1. Procédé de substitution dans un contenu audiovisuel d'une partie du contenu définie par une plage temporelle, par une séquence audiovisuelle de substitution, le procédé étant caractérisé en ce que la séquence de substitution est représentative de la partie substituée dudit contenu.  
5
2. Procédé selon la revendication 1 caractérisé en ce que la séquence de substitution est générée à partir d'une analyse sémantique de la partie du contenu définie par la plage temporelle.
- 10 3. Procédé selon l'une des revendications précédentes caractérisé en ce que la partie substituée est mémorisée et en ce que la séquence de substitution comprend un lien interactif adapté pour déclencher la restitution de la partie substituée.
- 15 4. Procédé selon l'une des revendications précédentes caractérisé en ce que le contenu audiovisuel résultant de la substitution comprend un index indiquant le début de la séquence de substitution.
- 20 5. Procédé selon l'une des revendications précédentes caractérisé en ce que les instants de début et de fin de la plage temporelle sont déterminés par la détection d'un premier et d'un second évènement particulier dans le flux audiovisuel.
- 25 6. Procédé selon la revendication 5 caractérisé en ce que en ce que au moins un des premiers et second évènements particuliers est identifié par reconnaissance d'au moins un mot clef dans une composante audio du contenu audiovisuel.
- 30 7. Procédé selon l'une des revendications 5 à 6 caractérisé en ce qu'au moins un des premier et second évènements particuliers est identifié par un changement de locuteur dans une composante audio du contenu audiovisuel.
8. Procédé selon l'une des revendications 5 à 7 caractérisé en ce qu'au moins un des premier et second évènements particuliers est identifié par un changement de source audio dans une composante audio du contenu audiovisuel.

9. Procédé selon l'une des revendications 5 à 8 caractérisé en ce que au moins un des premier et second évènements particuliers est identifié par analyse d'image dans une composante vidéo du contenu audiovisuel.
- 5 10. Procédé selon l'une quelconque des revendications précédentes caractérisé en ce que la séquence de substitution est représentative de la séquence substituée en ce qu'elle comprend un résumé de la partie substituée.
- 10 11. Dispositif de substitution dans un contenu audiovisuel, d'une partie du contenu définie par une plage temporelle, par une séquence audiovisuelle de substitution, le dispositif comportant une unité de traitement adaptée pour générer une séquence de substitution représentative de la partie substituée dudit contenu.
- 15 12. Terminal caractérisé en ce qu'il comporte un dispositif de substitution selon la revendication 11.
13. Programme d'ordinateur comportant les instructions pour l'exécution du procédé de substitution selon l'une quelconque des revendications 1 à 10, lorsque le programme est exécuté par un processeur.
- 20 14. Support d'informations lisible par un processeur sur lequel est enregistré un programme d'ordinateur comprenant des instructions pour l'exécution des étapes du procédé de substitution selon l'une quelconque des revendications 1 à 10.

1/3

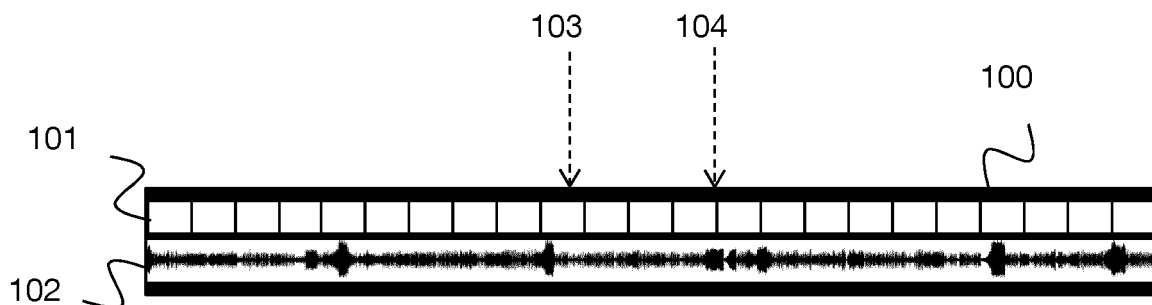


Fig. 1a

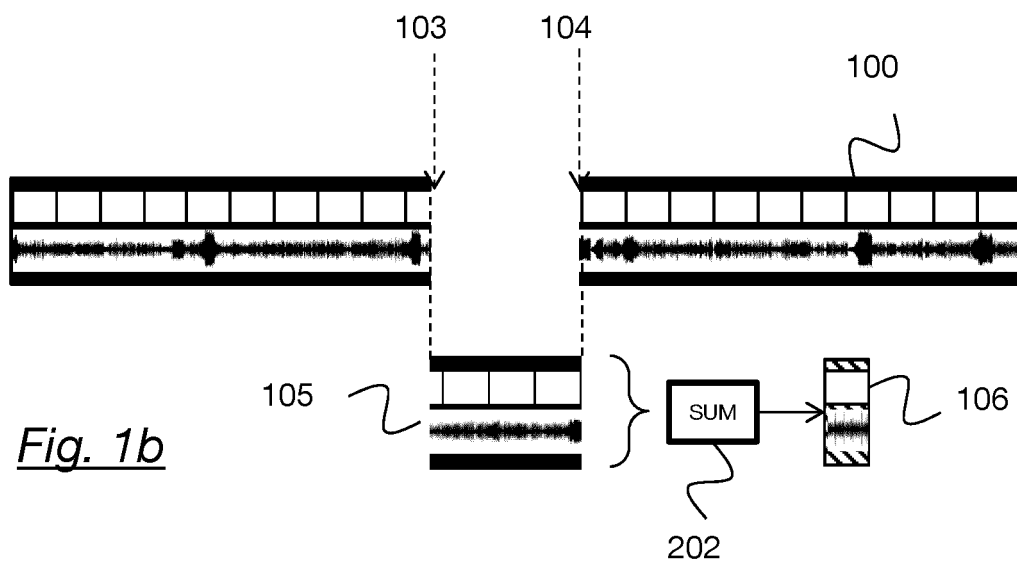


Fig. 1b

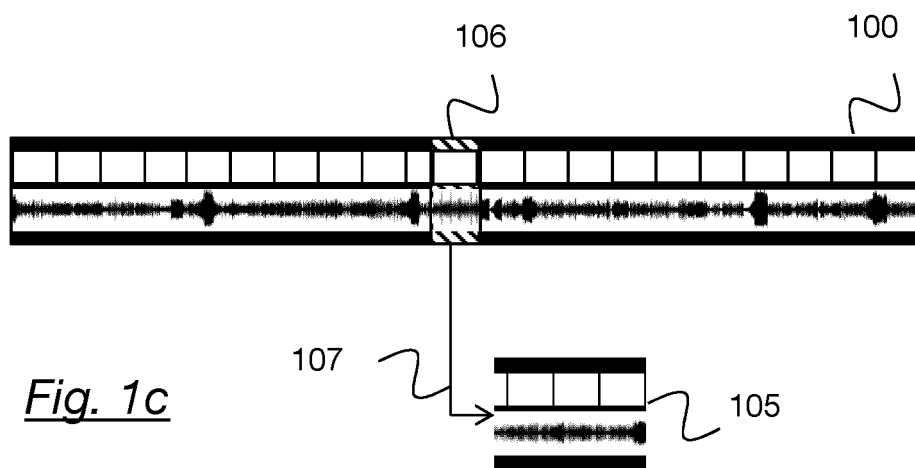


Fig. 1c



2/3

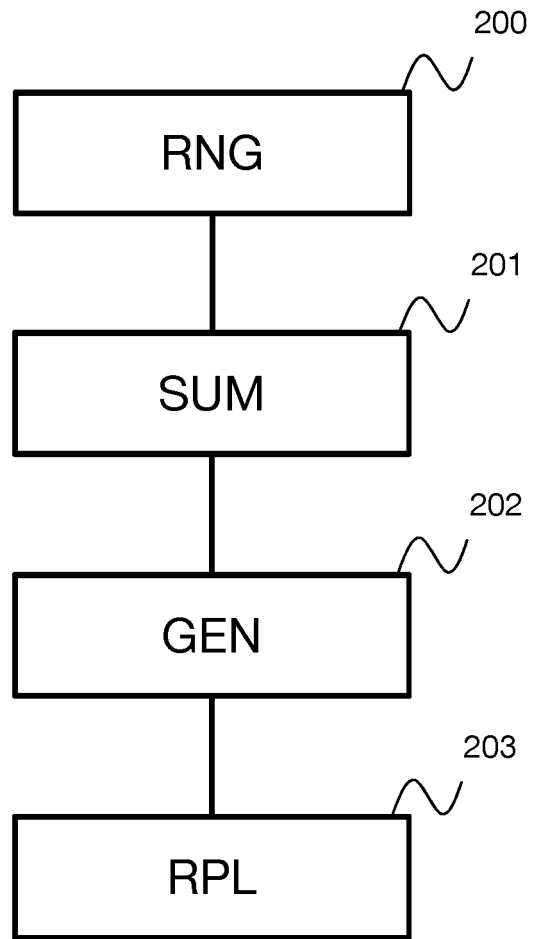


Fig. 2

3/3

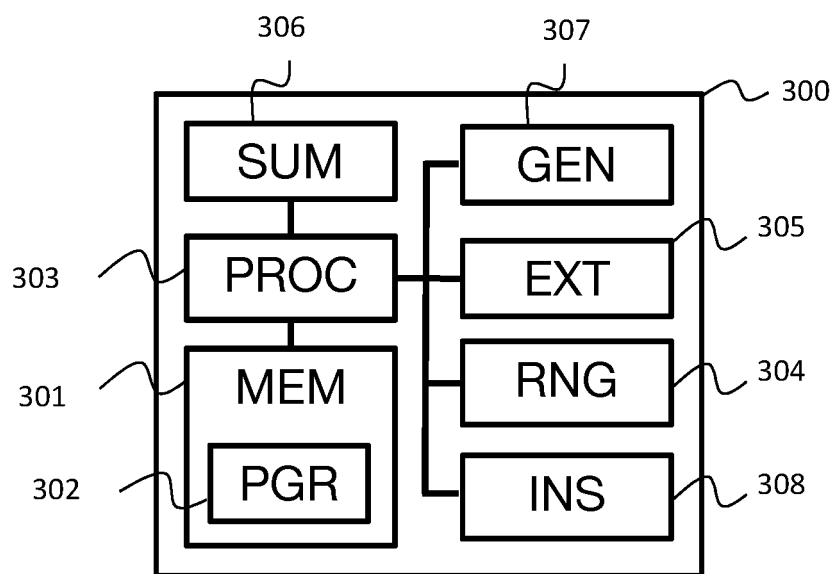


Fig. 3



**RAPPORT DE RECHERCHE  
PRÉLIMINAIRE**

N° d'enregistrement  
national

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

FA 813293  
FR 1555577

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		
X	CN 104 065 977 A (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING) 24 septembre 2014 (2014-09-24) * alinéa [0064] * * figures 1-2 * & EP 2 953 132 A1 (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO LTD [CN]) 9 décembre 2015 (2015-12-09) * alinéa [0064] * * figures 1-2 *	1-14	H04N21/854
X	----- US 2014/278308 A1 (LIU NATHAN [US] ET AL) 18 septembre 2014 (2014-09-18) * alinéa [0047] * * figures 1-30 *	1-14	
X	----- EP 2 575 341 A1 (FRANCE TELECOM [FR]) 3 avril 2013 (2013-04-03) * alinéa [0058] * * figures 1-7c *	1-14	
X	----- US 2014/310342 A1 (GOUESBET PIERRE-ERWANN [FR] ET AL) 16 octobre 2014 (2014-10-16) * alinéa [0024] * * figures 1-4b *	1-14	DOMAINES TECHNIQUES RECHERCHÉS (IPC) H04N
A	----- FR 2 811 111 A1 (QUESTION D IMAGE [FR]) 4 janvier 2002 (2002-01-04) * figures 1-2b *	1-14	
		Date d'achèvement de la recherche	Examineur
		5 avril 2016	Votini, Stefano
CATÉGORIE DES DOCUMENTS CITÉS		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	
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			

1

EPO FORM 1503 12.99 (P04C14)

**ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE  
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 1555577 FA 813293**

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 05-04-2016

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
CN 104065977	A	24-09-2014	CN 104065977 A	24-09-2014
			EP 2953132 A1	09-12-2015
			JP 2015231237 A	21-12-2015
			US 2015358653 A1	10-12-2015
-----				
US 2014278308	A1	18-09-2014	AUCUN	
-----				
EP 2575341	A1	03-04-2013	EP 2575341 A1	03-04-2013
			ES 2502476 T3	03-10-2014
			FR 2980940 A1	05-04-2013
			US 2013083907 A1	04-04-2013
-----				
US 2014310342	A1	16-10-2014	CN 101953135 A	19-01-2011
			FR 2925800 A1	26-06-2009
			US 2011016170 A1	20-01-2011
			US 2014310342 A1	16-10-2014
			US 2015172346 A1	18-06-2015
			WO 2009083459 A2	09-07-2009
-----				
FR 2811111	A1	04-01-2002	AUCUN	
-----				