(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
14 juin 2012 (14.06.2012)

WIPO | PCT

(10) Numéro de publication internationale
WO 2012/076807 A1

(51) Classification internationale des brevets :
G10L 13/08 (2006.01) *G10L 15/00* (2006.01)
G10L 13/02 (2006.01) *G10L 15/02* (2006.01)

(21) Numéro de la demande internationale :
PCT/FR2011/052882

(22) Date de dépôt international :
6 décembre 2011 (06.12.2011)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
1060158 7 décembre 2010 (07.12.2010) FR

(71) Déposant *(pour tous les États désignés sauf US)* : **FRANCE TELECOM** [FR/FR]; 6, place d'Alleray, F-75015 Paris (FR).

(72) Inventeurs; et
(75) Inventeurs/Déposants *(pour US seulement)* : **BAGSHAW, Paul** [GB/FR]; 12, allée Paul Sérusier, F-35235 Thorigne Fouillard (FR). **HINARD, Edouard** [FR/FR]; 7, rue Compagnie Roger Barbé, F-22300 Lannion (FR).

(74) Mandataire : **FRANCE TELECOM R&D/PIV/BREVETS**; MILET Isabelle, 38-40, rue du Général Leclerc, F-92794 Issy Moulineaux Cedex 9 (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, 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, IS, JP, KE, KG, KM, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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,

[Suite sur la page suivante]

(54) Title : METHOD AND SYSTEM FOR VOCALISING A TEXT

(54) Titre : PROCEDE ET SYSTEME DE VOCALISATION D'UN TEXTE

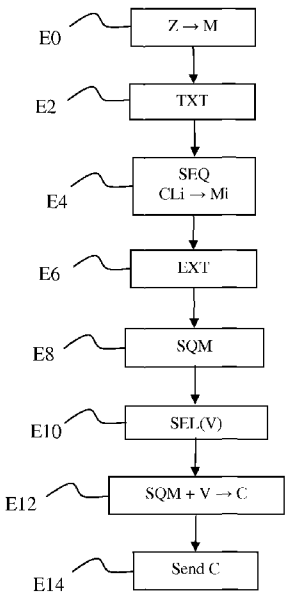


Figure 2

(57) Abstract : The invention relates to a method for vocalising a text, including detecting, in the text, at least one text extract that fulfils a criterion for identifying undesirable words, and producing audio content, resulting from the vocalisation of said text, in which the detected extract is not audible. According to the invention, the method is adapted such that a text extract fulfils said identification criterion when said extract has at least one predefined linguistic characteristic and usage context that are identical to those of a group of at least one undesirable word of a predefined set of groups of undesirable words. The invention also relates to a system implementing the vocalisation method.

(57) Abrégé : L'invention se rapporte à un procédé de vocalisation d'un texte comprenant la détection dans le texte d'au moins un extrait de texte vérifiant un critère d'identification de mot indésirable et la production d'un contenu audio, résultant d'une vocalisation dudit texte et dans lequel l'extrait détecté n'est pas audible. Selon l'invention, le procédé est adapté de façon à ce qu'un extrait de texte vérifie ledit critère d'identification s'il possède au moins une caractéristique linguistique prédéfinie et un contexte d'usage identiques à ceux d'un groupe d'au moins un mot indésirable d'un ensemble prédéfini de groupes de mots indésirables. L'invention se rapporte également à un système mettant en œuvre le procédé de vocalisation.

WO 2012/076807 A1

UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Déclarations en vertu de la règle 4.17 :

— *relative à la qualité d'inventeur (règle 4.17.iv))*

Publiée :

— *avec rapport de recherche internationale (Art. 21(3))*

Procédé et système de vocalisation d'un texte

L'invention se rapporte au domaine de la synthèse vocale ("Text-to-Speech synthesis" en anglais).

De façon connue, un système automatique de synthèse vocale transforme un texte écrit en un contenu audio intelligible correspondant.

Un procédé de synthèse vocale à partir d'un texte mis en œuvre dans un tel système se décompose classiquement en deux étapes : une étape de traitement linguistique et une étape de production d'un contenu audio. Le traitement linguistique comprend l'analyse du texte à synthétiser et la génération d'une suite de représentations symboliques des sons correspondant au texte. Lors de l'étape de production, un contenu audio correspondant au texte est construit à partir de la suite de représentations symboliques obtenue lors de l'étape de traitement linguistique. De manière à ce que le contenu audio ainsi produit simule la voix d'un locuteur, l'étape de production du contenu audio prend en compte un modèle de voix, propre à ce locuteur, qui est construit à partir de phrases prédéfinies prononcées par ce locuteur et enregistré dans une base de données lors d'une étape préalable.

De nombreuses applications sont basées sur de tels procédés de synthèse vocale. Parmi celle-ci, des applications proposent à un utilisateur l'envoi d'un contenu audio obtenu à partir d'un texte fourni par cet utilisateur et d'un modèle de voix d'une personne prédéterminée, par exemple une personnalité connue. Par exemple, l'application permet à un utilisateur de saisir un texte de son choix à l'aide d'une interface graphique et de sélectionner une personne parmi un ensemble de personnes prédéfini. Le contenu audio construit en fonction du texte et d'un modèle associé à la personne sélectionnée est ensuite transmis à un ami de l'utilisateur. Cet ami, lors de l'écoute du contenu audio reçu entend le texte de l'utilisateur prononcé avec la voix de la personne choisie.

Le texte saisi peut contenir des mots et des phrases offensantes pour une personne écoutant le contenu audio, ou attaquant l'image publique d'une personne dont la voix est exploitée par le système ou encore qui détériore l'image du fournisseur de service.

Ces mots ou phrases (par exemple des mots obscènes) sont considérés comme indésirables.

Des méthodes existantes permettent de détecter et de supprimer les mots indésirables du contenu audio restitué.

Ces méthodes sont basées sur une comparaison des mots du texte avec une liste de mots indésirables enregistrée dans une mémoire du système. Lorsqu'un mot du texte figure dans la liste, il est remplacé par un autre mot ou supprimé du texte à vocaliser.

La comparaison est réalisée avant le traitement linguistique du texte nécessaire à la vocalisation.

Ces méthodes ne sont toutefois pas suffisamment efficaces. En effet, l'utilisateur, auteur du texte, peut délibérément modifier l'orthographe d'un mot indésirable de façon à ce que le mot mal orthographié ne soit pas dans la liste des mots interdits. Dans ce cas, le mot indésirable ne sera pas supprimé du texte à vocaliser. Ceci est notamment possible car de nombreux langages naturels possèdent plusieurs formes écrites différentes (graphèmes) correspondant à une même perception d'un son lorsqu'elles sont prononcées (phonèmes).

Il existe donc un besoin pour améliorer l'efficacité du filtrage des mots ou séquences de mots indésirables.

L'invention vient améliorer la situation.

A cet effet, l'invention se rapporte à un procédé de vocalisation d'un texte comprenant:

- une étape de détection dans ledit texte d'au moins un extrait de texte vérifiant un critère d'identification de mot indésirable;

- une étape de production d'un contenu audio, résultant d'une vocalisation dudit texte et dans lequel ledit au moins un extrait détecté n'est pas audible;

caractérisé en ce qu'un extrait de texte vérifie ledit critère d'identification s'il possède au moins une caractéristique linguistique prédéfinie et un contexte d'usage dans ledit texte identiques à ceux définis pour un groupe d'au moins un mot indésirable d'un ensemble prédéfini d'au moins un groupe d'au moins un mot indésirable.

L'utilisation de caractéristiques linguistiques permet d'identifier dans le texte des extraits de texte, c'est-à-dire des mots, des expressions ou des phrases, suspectés d'être indésirables. L'analyse du contexte dans lequel ces extraits suspects sont employés permet de rejeter uniquement les extraits indésirables.

Selon une caractéristique particulière du procédé de vocalisation d'un texte, la caractéristique linguistique prédéfinie est une représentation phonétique ou le lemme.

L'utilisation de la représentation phonétique permet de s'affranchir du découpage en mots effectué par l'utilisateur et de l'orthographe utilisée par l'utilisateur et ainsi de détecter des extraits pour lequel l'utilisateur a volontairement modifié le découpage des mots et/ou l'orthographe.

L'utilisation du lemme permet de détecter toutes les formes d'un mot indésirable tout en gardant une taille raisonnable de l'ensemble prédéfini des mots indésirables.

Selon une caractéristique particulière du procédé de vocalisation d'un texte, le groupe comprend plusieurs mots indésirables et le procédé comporte une étape de comparaison mot par mot de caractéristiques linguistiques du groupe et de l'extrait.

L'attribution de caractéristiques par mot lors de la définition d'un groupe de mots de l'ensemble de groupe de mots indésirables permet de regrouper plusieurs expressions indésirables dans un même groupe de mots indésirables et d'éviter un volume important de l'ensemble prédéfini.

La comparaison mot par mot permet de détecter toutes les formes du groupe à partir de l'enregistrement d'un seul groupe.

Par exemple, le choix de la caractéristique lemme pour le premier mot et de la représentation phonétique pour le deuxième mot permet de détecter toutes les formes conjuguées ou déclinées du premier mot lorsqu'il est suivi du deuxième mot.

Selon une caractéristique particulière du procédé de vocalisation d'un texte, le contexte d'usage d'un extrait est déterminé à partir de la syntaxe d'une phrase contenant ledit extrait et/ou du rôle sémantique dudit extrait.

L'analyse de la syntaxe de la phrase permet par exemple de déterminer la fonction grammaticale d'un mot dans une phrase contenant un extrait considéré et ainsi de déterminer si l'extrait est indésirable ou non indésirable en fonction de la fonction grammaticale de ce mot.

Par exemple, un extrait est censuré s'il est précédé d'un verbe et accepté sinon.

Le rôle sémantique d'un mot permet de différencier deux mots ayant la même orthographe tout en ayant un sens différent. Le rôle sémantique est ici utilisé pour décider si l'extrait considéré est ou non indésirable.

Par exemple, le mot anglais "black" peut être conservé s'il est utilisé en tant que couleur mais doit être rejeté s'il est utilisé pour qualifier péjorativement une personne.

Selon une caractéristique particulière du procédé de vocalisation d'un texte, la vérification du contexte d'usage est effectuée par comparaison d'un groupe de mots contenant ledit extrait avec une liste d'au moins une expression autorisée et/ou une liste d'au moins une expression interdite.

Un mot ou groupe de mots qui est indésirable employé seul ou dans certaines expressions peut être employé correctement dans d'autres expressions. Par exemple, le mot français "cul" doit être détecté comme indésirable s'il est employé seul et être accepté s'il fait partie de l'expression française "cul de sac".

Selon une caractéristique particulière du procédé de vocalisation d'un texte, la comparaison du groupe et de la liste est basée sur la représentation phonétique.

Selon une autre caractéristique particulière du procédé de vocalisation d'un texte, utilisée seule ou en combinaison avec la caractéristique précédente, le rôle sémantique est utilisé pour la comparaison.

L'utilisation de caractéristiques linguistiques pour effectuer la comparaison d'un groupe de mots contenant un extrait suspect et un groupe de mots d'une liste permet de déterminer si l'extrait est ou non indésirable.

La représentation phonétique permet une comparaison simple et rapide.

Selon un mode de réalisation particulier du procédé de vocalisation d'un texte, la vocalisation du texte comporte une étape de traitement linguistique lors de laquelle des

caractéristiques linguistiques sont générées et, lors de l'étape de détection, le critère d'identification est vérifié à partir d'au moins une caractéristique linguistique générée.

Les caractéristiques linguistiques déterminées lors d'une phase de traitement linguistique visant à transformer le texte en une séquence de phonèmes sont également utilisées pour détecter les mots et extraits indésirables. Ainsi le processus de détection est peu consommateur de temps et peut être effectué en temps quasi réel lors d'une vocalisation d'un texte.

Selon un mode de réalisation particulier du procédé de vocalisation d'un texte, la vocalisation comporte une étape d'obtention d'une séquence de phonèmes pour ledit texte, le procédé comporte une étape de modification de ladite séquence de phonèmes lors de laquelle au moins un phonème d'un extrait détecté est remplacé par au moins un phonème de remplacement et le contenu audio est produit en fonction de la séquence modifiée.

La modification de la séquence de phonèmes permet de rendre inaudibles les extraits indésirables.

Selon une caractéristique particulière de ce mode de réalisation, lors de l'étape de modification, l'ordre des phonèmes d'un extrait détecté est modifié.

La modification de l'ordre des phonèmes permet de changer la prononciation d'un extrait indésirable tout en conservant la durée du contenu. Le contenu correspondant aux extraits indésirables est incompréhensible mais reste harmonieux car prononcé de façon similaire au contenu non censuré.

Selon encore un autre mode de réalisation particulier du procédé d'un texte, utilisé seul ou en combinaison avec un des deux modes de réalisation précédents, un signal sonore prédéfini masque le contenu audio représentant au moins un extrait détecté.

Le masquage du signal sonore rend inaudible les extraits détectés indésirables.

Selon un mode de réalisation particulier, le procédé de vocalisation d'un texte comporte en outre une étape préalable de mémorisation d'un groupe d'au moins un mot indésirable dudit ensemble en association avec des données définissant au moins une caractéristique linguistique associée et un contexte d'usage dans lequel ledit groupe est indésirable.

L'ensemble de mots indésirables est construit en fonction de l'application, de l'évolution d'un langage, de la langue puis mémorisé. Chaque groupe de mots de l'ensemble permet de détecter des extraits indésirables dans un texte à vocaliser.

Les données définissant la ou les caractéristiques linguistiques associées à un groupe de mots permettent de détecter un extrait suspecté d'être indésirable.

Les données définissant le contexte d'usage permettent de déterminer si l'extrait suspecté doit être accepté ou censuré.

Selon un mode de réalisation particulier, le procédé comporte en outre une étape de sélection d'un modèle de voix prédéterminé et le contenu audio est en outre produit en fonction dudit modèle de voix sélectionné.

Le procédé de vocalisation selon l'invention permet ainsi de respecter l'image de la personne associée au modèle de voix.

L'invention se rapporte également à un système de vocalisation d'un texte comprenant

- des moyens de détection dans ledit texte d'au moins un extrait de texte vérifiant un critère d'identification de mot indésirable;

- des moyens de production d'un contenu audio, résultant d'une vocalisation dudit texte et dans lequel ledit au moins un extrait détecté n'est pas audible;

caractérisé en ce que les moyens de détection d'au moins un extrait de texte vérifiant ledit critère d'identification sont agencés pour déterminer si un extrait de texte possède au moins une caractéristique linguistique prédéfinie et un contexte d'usage identiques à ceux d'un groupe d'au moins un mot indésirable d'un ensemble prédéfini d'au moins un groupe d'au moins un mot indésirable.

L'invention se rapporte enfin à un produit programme d'ordinateur comprenant des instructions pour mettre en œuvre les étapes du procédé de vocalisation tel que décrit précédemment, lorsqu'il est chargé et exécuté par un processeur.

D'autres particularités et avantages de la présente invention apparaitront dans la description suivante de modes de réalisation donnés à titre d'exemples non limitatifs, en référence aux dessins annexés, dans lesquels :

- la figure 1 illustre un système de vocalisation selon un mode de réalisation de l'invention,

- la figure 2 est un organigramme illustrant les différentes étapes d'un procédé de vocalisation mis en œuvre dans un système de vocalisation selon un mode de réalisation,

- la figure 3 est un tableau illustrant un ensemble de groupes de mots indésirables enregistrés lors d'une étape préalable de mémorisation et mis en œuvre dans une étape de détection d'un procédé de vocalisation selon un premier mode de réalisation,

- la figure 4 est un tableau illustrant un ensemble de groupes de mots indésirables enregistrés lors d'une étape préalable de mémorisation et mis en œuvre dans une étape de détection d'un procédé de vocalisation selon un deuxième mode de réalisation.

Un mode de réalisation particulier de l'invention va maintenant être décrit en référence aux figures 1 et 2.

En référence à la **figure 1**, un système SYS comprend un serveur de vocalisation SV, un premier terminal T1 d'un premier utilisateur A et un deuxième terminal T2 d'un deuxième utilisateur B aptes à communiquer au travers d'un réseau de communication R.

Le premier terminal T1 est, par exemple, un ordinateur de type PC (pour "Personal Computer"), un PDA (pour "Personal Digital Assistant") ou un téléphone mobile.

Le deuxième terminal T2 est, par exemple, un téléphone fixe, un téléphone mobile, un PDA.

Le serveur de vocalisation SV possède notamment une unité de traitement 100 équipée d'un microprocesseur, une mémoire vive de type RAM 102, une ou plusieurs mémoires mortes de type ROM ou EEPROM 104 dans laquelle sont enregistrés des programmes pouvant être exécutés par le microprocesseur.

Le serveur de vocalisation SV possède également un module de réception REC, un module de traitement linguistique MTL, un dispositif de détection DET, un module de sélection d'un modèle de voix SEL, un module de production d'un contenu audio PRO et un module d'émission ENV.

Le serveur de vocalisation SV peut comporter de manière classique et non exhaustive les éléments suivants: un écran, un clavier, un microphone, un haut-parleur, un moyen de stockage…

L'unité de traitement est pilotée par un programme informatique PG afin de mettre en œuvre notamment le procédé de vocalisation selon un mode de réalisation de l'invention décrit ultérieurement en référence à la figure 2.

Le programme informatique PG comporte des instructions de code pour mettre en œuvre les étapes du procédé de vocalisation d'un texte et notamment les étapes :

- de détection dans le texte d'au moins un extrait de texte vérifiant un critère d'identification de mot indésirable, un extrait de texte vérifiant le critère d'identification s'il possède au moins une caractéristique linguistique prédéfinie et un contexte d'usage dans ledit texte identiques à ceux d'un groupe d'au moins un mot indésirable d'un ensemble prédéfini d'au moins un groupe d'au moins un mot indésirable.
- de production d'un contenu audio, résultant d'une vocalisation du texte et dans lequel ledit au moins un extrait détecté n'est pas audible.

Le programme informatique PG peut également mettre en œuvre une étape d'enregistrement d'un ensemble prédéfini de groupes de mots indésirables et de données définissant pour chaque groupe au moins une caractéristique linguistique associée et un contexte d'usage dans lequel le groupe considéré est indésirable. Cet ensemble est nommé par la suite "ensemble de référence".

Le module de réception REC est apte à recevoir des données via le réseau R, par exemple un texte à vocaliser transmis par le terminal T1 du premier utilisateur A.

Le module d'émission ENV est apte à transmettre des données, par exemple un contenu audio, à destination du deuxième terminal T2 de deuxième utilisateur B, via le réseau de télécommunications R.

Un mode de réalisation du procédé de vocalisation mis en œuvre dans le système SYS est maintenant décrit en référence à la **figure 2**.

Lors d'une étape préalable E0, un ensemble de référence Z d'un ou plusieurs groupes de mots indésirables est mémorisé dans une mémoire M, par exemple une mémoire morte 104, du serveur de vocalisation SV. Dans cette mémoire, sont mémorisées des données définissant, pour chaque groupe de l'ensemble de référence Z, au moins une caractéristique linguistique associée et un contexte d'usage dans lequel le groupe considéré est indésirable.

Ces données comprennent par exemple, concernant la définition de la ou des caractéristiques linguistiques:

- une représentation phonétique du groupe; et/ou
- une représentation phonétique d'un mot du groupe; et/ou
- le lemme de chaque mot du groupe; et/ou
- le lemme d'un mot du groupe; et/ou
- une étiquette grammaticale d'un mot du groupe.

Une étiquette grammaticale comporte une fonction grammaticale du mot dans une phrase du texte et un ou plusieurs attributs associés.

Une fonction grammaticale est par exemple, un nom, un verbe, un adjectif, un article…

Un attribut permet de préciser la fonction grammaticale.

Un nom peut par exemple être associé à un attribut de genre : masculin, féminin ou indéfini et à un attribut précisant si le nom est employé au singulier ou au pluriel.

Un attribut associé à un verbe est par exemple un attribut indiquant le temps auquel le verbe est conjugué.

De façon connue, la représentation phonétique d'un mot est constituée d'une suite symbolique de phones et de phonèmes.

Ces données enregistrées dans la mémoire M comprennent en outre, concernant la définition du contexte d'usage dans lequel le groupe est indésirable, une liste d'au moins une expression autorisée et/ou une liste d'au moins une expression interdite. Par exemple si le mot français "cul" est indésirable lorsqu'il est utilisé seul mais autorisé s'il est utilisé dans l'expression française "cul de sac". Dans ce cas, la liste d'expressions interdites comprendra l'expression "cul" et la liste d'expression autorisée comprendra l'expression "cul de sac". Selon ce qui est le plus simple pour définir les mots ou groupes indésirables, une seule de ces deux listes peut être utilisée ou bien les deux.

En outre, chaque expression d'une de ces listes d'expressions peut être représentée dans la mémoire M sous sa forme phonétique, de manière à pouvoir détecter la présence dans le texte d'une telle expression malgré d'éventuelles fautes d'orthographe dans le texte à vocaliser.

Une expression d'une de ces listes peut également être définie en fonction de caractéristiques linguistiques telle qu'une étiquette grammaticale d'un mot ou un rôle sémantique d'un mot de l'expression.

Par exemple, la liste des mots autorisés peut comprendre le mot français "jet" défini par son orthographe et un rôle sémantique "aviation". Dans ce cas, le mot français "jet" ayant un rôle sémantique "jeter" n'est pas dans la liste des mots autorisés.

A titre d'alternative, l'ensemble de référence Z est enregistré dans un serveur distant (non représenté) accessible par le serveur de vocalisation SV à travers le réseau R.

Lors d'une étape E2, le module de réception REC reçoit un texte à vocaliser TXT en provenance du premier terminal T1 de l'utilisateur A.

A titre d'alternative, le texte à vocaliser TXT est lu dans une mémoire du serveur de vocalisation SV.

Lors d'une étape suivante E4, le module de traitement linguistique MTL analyse le texte à vocaliser TXT.

Plus précisément, le module de traitement linguistique MTL détermine les mots M1, M2, M3…du texte TXT. De façon classique, ce découpage en mot prend en compte la ponctuation du texte. Un signe de ponctuation, par exemple le signe ".", est considéré comme un mot par le module de traitement linguistique MTL.

Pour chaque mot Mi du texte, le module de traitement linguistique MTL détermine ensuite au moins une caractéristique linguistique CLi.

Les caractéristiques linguistiques déterminées pour un mot ou un extrait du texte sont la phonétique, le lemme, une étiquette grammaticale, un rôle sémantique…

Egalement lors de l'étape E4, le module de traitement linguistique MTL détermine une suite de phonèmes associée à chaque mot en fonction des caractéristiques linguistiques déterminées.

Puis, le module de traitement linguistique MTL détermine une séquence SEQ de phonèmes représentant le texte TXT. La séquence de phonèmes SEQ est obtenue par la concaténation des suites de phonèmes de mots du texte TXT.

L'étape E4 est suivie d'une étape de détection E6 lors de laquelle au moins un extrait de texte indésirable EXT est détecté dans le texte TXT en fonction de l'ensemble de référence Z mémorisé lors de l'étape E0.

Un extrait de texte EXT est soit un mot du texte TXT, soit un ensemble de mots consécutifs du texte TXT.

Lors de l'étape de détection E6, un extrait de texte TXT est détecté s'il vérifie un critère d'identification de mot indésirable. L'extrait TXT vérifie un critère d'identification de mot indésirable s'il possède au moins une caractéristique linguistique prédéfinie et un contexte d'usage identiques à ceux d'un groupe de l'ensemble de référence Z.

Le module de détection DET compare les propriétés d'un extrait du texte à vocaliser avec celles des groupes de mot de l'ensemble de référence.

Plus précisément, le module de détection DET compare les caractéristiques linguistiques utilisées pour définir un groupe de l'ensemble de référence Z avec les caractéristiques linguistiques déterminées pour un mot ou un extrait du texte lors de l'étape de traitement linguistique E4.

Puis, il compare le contexte d'usage utilisé pour définir un groupe de l'ensemble de référence Z avec le contexte d'usage associé au mot ou à l'extrait considéré.

Ainsi, par simple comparaison de propriétés, on peut vérifier si un critère d'identification de mot indésirable est vérifié ou non.

Puis lors d'une étape E8, le module de modification MOD modifie la séquence SEQ obtenue lors de l'étape E4 de traitement linguistique en une séquence modifiée SQM.

Dans le mode de réalisation décrit, la séquence modifiée SQM correspond à la séquence initiale SEQ dans laquelle les phonèmes représentant les extraits TXT indésirables détectés sont remplacés par une suite de phonèmes prédéterminée.

Plus précisément, chaque phonème d'un extrait détecté EXT est remplacé par un phonème de remplacement.

Un phonème de remplacement est, par exemple, un phonème correspondant à un son inaudible, à un silence, à un son de fréquence prédéterminée ("beep"), à un cri d'animal ou encore un phonème choisi aléatoirement.

A titre d'alternative, la séquence modifiée SQM correspond à la séquence initiale SEQ dans laquelle les phonèmes représentant les extraits TXT indésirables détectés ont été enlevés.

Encore, à titre d'alternative, la séquence modifiée SQM correspond à la séquence initiale SEQ dans laquelle l'ordre des phonèmes représentant un ou plusieurs extraits TXT indésirables détectés est modifié.

Lors d'une étape E10, le module de sélection SEL sélectionne un modèle de voix V parmi un ensemble de modèles de voix.

Dans le mode de réalisation décrit, un identifiant d'une personnalité, par exemple, un nom, est sélectionné par le premier utilisateur A parmi un ensemble d'identifiants de personnes au moyen d'une interface graphique du premier terminal T1 et l'identifiant sélectionné est transmis au module de sélection SEL du serveur de vocalisation SV au travers du module de réception REC. Le module de sélection SEL sélectionne ensuite le modèle de voix V en fonction de l'identifiant sélectionné reçu.

Puis, lors d'une étape E12, le module de production PRO détermine un contenu audio C en fonction de la séquence de phonèmes modifiée SQM et du modèle de voix sélectionné V.

Lors d'une étape E14, le module d'envoi ENV transmet le contenu audio C obtenu au deuxième terminal T2 du deuxième utilisateur B.

L'utilisateur B peut écouter le contenu audio C reçu par le deuxième terminal T2.

A titre d'alternative, le contenu audio C est déterminé en fonction de la séquence de phonèmes initiale SEQ et du modèle de voix sélectionné V et d'un signal prédéterminé S. Plus précisément, le signal S masque le contenu audio correspondant aux phonèmes des extraits indésirables détectés. Le masquage par le signal S des mots ou groupes de mots détectés comme étant indésirables rend inaudible ces mots ou groupes.

Un premier mode de réalisation particulier des étapes de traitement linguistique E4 et de détection E6 va maintenant être décrit en référence à **figure 3**.

Un ensemble de référence Z1 a été enregistré lors de l'étape préalable de mémorisation E0 et le texte à vocaliser est le texte : "MT1 MT2 MT3 MT4" dans lequel MT1, MT2, MT3 et MT4 sont des mots saisis par l'utilisateur A.

Comme illustré figure 3, l'ensemble de référence Z1 comprend 3 groupes de mots : GR1, GR2 et GR3.

Le groupe de mots GR1 est défini par un ensemble de caractéristiques linguistiques CL1 et un contexte d'usage C1. L'ensemble de caractéristiques CL1 est par exemple constitué d'une seule caractéristique : la représentation phonétique du groupe de mots GR1 est égale à une valeur prédéterminée PH1. Le contexte d'usage C1 est par exemple : "le mot suivant le groupe de mots GR1 est différent d'un nom".

Le groupe de mots GR2 est défini par un ensemble de caractéristiques linguistiques CL2 et un contexte d'usage C2. L'ensemble de caractéristiques CL2 est par exemple constitué de deux caractéristiques : la représentation phonétique du premier mot du groupe de mots GR2 est égale à une valeur prédéterminée PH2 et l'étiquette grammaticale du deuxième et dernier mot du groupe de mots GR2 est égale à une valeur prédéterminée PH3. Le contexte d'usage C2 est par exemple : "indifférent". Un contexte d'usage du type "indifférent" signifie qu'un mot ou groupe de mots vérifie ce contexte quel que soit le contexte.

La valeur prédéterminée PH2 est ici égale à la concaténation d'une valeur prédéterminée P2 et d'une valeur prédéterminée P3.

Le groupe de mots GR3 est défini par un ensemble de caractéristiques linguistiques CL3 et un contexte d'usage C3. L'ensemble de caractéristiques CL3 est par exemple constitué d'une seule caractéristique : le lemme du groupe de mots GR3 est égal à une valeur prédéterminée L1. Le

contexte d'usage C3 est par exemple : "l'étiquette grammaticale du groupe de mots GR3 est différent d'un verbe".

Lors de l'étape E4, le module de traitement linguistique MTL détermine que :

- le mot MT1 a pour phonétique P1, pour lemme L1 et pour étiquette grammaticale G1=verbe,
- le mot MT2 a pour phonétique P2, pour lemme L2 et pour étiquette grammaticale G2,
- le mot MT3 a pour phonétique P3, pour lemme L3 et pour étiquette grammaticale G3,
- le mot MT4 a pour phonétique P4, pour lemme L4 et pour étiquette grammaticale G4 = "nom, masculin, singulier".

Lors de l'étape E6, le module de détection DET détecte que l'extrait MT2-MT3 composé des mots MT2 et MT3 possède les mêmes caractéristiques linguistiques et le même contexte d'usage que GR1. En effet, PH1=P2+P3 et le mot MT4 suivant le mot MT3 est un nom. L'extrait MT2-MT3 est en conséquence considéré comme indésirable.

Le contexte d'usage C1 est défini en fonction d'une caractéristique linguistique, qui est ici une étiquette grammaticale. Lors de l'étape de détection, l'étiquette grammaticale est appliquée à un mot du texte qui n'est pas un mot d'un extrait suspect, pour déterminer si le contexte d'usage de l'extrait suspect et du groupe GR1 sont identiques.

Le mot MT1 possède la caractéristique linguistique CL3 du groupe GR3. Cependant, le mot MT1 ne possède pas le contexte d'usage C3 car l'étiquette grammaticale G1 est un verbe. Aussi, le mot MT1 n'est pas considéré comme un mot indésirable.

Un deuxième mode de réalisation particulier des étapes de traitement linguistique E4 et de détection E6 va maintenant être décrit en référence à la **figure 4**.

Dans ce mode de réalisation, un ensemble de référence Z2 a été enregistré lors de l'étape préalable de mémorisation E0 et le texte à vocaliser est le texte : "MT1 MT2 MT3 MT4", identique au texte du mode de réalisation précédent.

Comme illustré sur la figure 4, l'ensemble Z2 comprend un premier sous ensemble de référence Z21 et un deuxième sous ensemble de référence Z22.

Le premier sous ensemble Z21 comprend ici 3 groupes de mots : GR4, GR5 et GR6.

Le groupe de mots GR4 est défini par un ensemble de caractéristiques linguistiques CL4. L'ensemble de caractéristiques CL4 est par exemple constitué d'une seule caractéristique : la phonétique du groupe de mots GR4 est égale à la valeur prédéterminée PH1.

Le groupe de mots GR5 est défini par un ensemble de caractéristiques linguistiques CL5. L'ensemble de caractéristiques CL5 est par exemple constitué de deux caractéristiques : la phonétique du premier mot du groupe de mots GR2 est égale à la valeur prédéterminée PH2 et l'étiquette grammaticale du deuxième et dernier mot du groupe de mots GR2 est un "nom".

Le groupe de mots GR6 est défini par un ensemble de caractéristiques linguistiques CL6. L'ensemble de caractéristiques CL6 est par exemple constitué d'une caractéristique : le lemme du groupe de mots GR3 est égal à la valeur prédéterminée L1.

Le deuxième sous ensemble Z22 comprend ici 2 groupes de mots : GA1 et GA2. Chaque groupe de mots du deuxième sous ensemble Z22 représente ici un ensemble de mots autorisés.

Le deuxième sous ensemble Z22 représente un exemple de liste de mots autorisés.

Par exemple, le groupe GA1 comprend les mots ou groupes de mots dont la phonétique est PH4 et le groupe GA2 comprend les mots ou groupes de mots dont la phonétique est PH1 suivi d'un mot dont l'étiquette grammaticale est un nom.

Dans ce mode de réalisation, l'étape de traitement linguistique E4 est similaire à celle décrite en relation avec le mode de réalisation précédent. Les caractéristiques linguistiques définies pour chaque mot du texte à vocaliser TXT sont celles obtenues dans le mode de réalisation précédent décrit en relation avec la figure 3.

Lors de l'étape de détection E6, le module de détection DET détecte que l'extrait MT2-MT3 composé des mots MT2 et MT3 possède les caractéristiques linguistiques de GR4. L'extrait MT2-MT3 est ainsi considéré comme étant potentiellement indésirable.

Puis le module de détection DET recherche si l'extrait MT2-MT3 correspond à un mot ou groupe de mots du deuxième sous ensemble de référence Z22. Si l'extrait MT2-MT3 correspond à un groupe de mots du deuxième sous ensemble de référence Z22, par exemple au groupe de mots GA2, l'extrait MT2-MT3 est considéré comme possédant le même contexte d'usage que GA2 et est considéré comme indésirable.

Le mot MT4 étant un nom, l'extrait "MT2-MT3-MT4" est un groupe de mots du groupe GA2. Aussi, l'extrait "MT2-MT3-MT4" est un extrait autorisé. L'extrait "MT2-MT3" étant utilisé dans un groupe de mots autorisé, la condition d'usage de cet extrait qui est ici de ne pas appartenir à un groupe du deuxième sous ensemble Z22 de référence n'est pas vérifiée et l'extrait "MT2-MT3" n'est pas considéré comme indésirable.

REVENDICATIONS

1. Procédé de vocalisation d'un texte (TXT) comprenant :
- une étape de détection (E6) dans ledit texte d'au moins un extrait de texte (EXT) vérifiant un critère d'identification de mot indésirable;
- une étape de production (E12) d'un contenu audio (C), résultant d'une vocalisation dudit texte et dans lequel ledit au moins un extrait détecté n'est pas audible;
caractérisé en ce qu'un extrait de texte vérifie ledit critère d'identification s'il possède au moins une caractéristique linguistique prédéfinie et un contexte d'usage dans ledit texte identiques à ceux définis pour un groupe d'au moins un mot indésirable d'un ensemble prédéfini d'au moins un groupe d'au moins un mot indésirable.

2. Procédé de vocalisation d'un texte selon la revendication 1 dans lequel la caractéristique linguistique prédéfinie est une représentation phonétique ou le lemme.

3. Procédé de vocalisation d'un texte selon la revendication 1 dans lequel le groupe comprend plusieurs mots indésirables et le procédé comporte une étape de comparaison mot par mot de caractéristiques linguistiques du groupe et de l'extrait.

4. Procédé de vocalisation d'un texte selon la revendication 1 dans lequel le contexte d'usage d'un extrait est déterminé à partir de la syntaxe d'une phrase contenant ledit extrait et/ou du rôle sémantique dudit extrait.

5. Procédé de vocalisation d'un texte selon la revendication 1 dans lequel la vérification du contexte d'usage est effectuée par comparaison d'un groupe de mots contenant ledit extrait avec une liste d'au moins une expression autorisée et/ou une liste d'au moins une expression interdite.

6. Procédé de vocalisation d'un texte selon la revendication 5 dans lequel la comparaison du groupe et de la liste est basée sur la représentation phonétique.

7. Procédé de vocalisation d'un texte selon la revendication 1 dans lequel le rôle sémantique est utilisé pour la comparaison.

8. Procédé de vocalisation d'un texte selon la revendication 1 caractérisé en ce que la vocalisation du texte comporte une étape de traitement linguistique lors de laquelle des

caractéristiques linguistiques sont générées et en ce que, lors de l'étape de détection, le critère d'identification est vérifié à partir des caractéristiques linguistiques générées.

9. Procédé de vocalisation d'un texte selon la revendication 1 dans lequel la vocalisation comportant une étape d'obtention d'une séquence de phonèmes pour ledit texte, le procédé comporte une étape de modification de ladite séquence de phonèmes lors de laquelle au moins un phonème d'un extrait détecté est remplacé par au moins un phonème de remplacement, et dans lequel le contenu audio est produit en fonction de la séquence modifiée.

10. Procédé de vocalisation d'un texte selon la revendication 9 dans lequel lors de l'étape de modification, l'ordre des phonèmes d'un extrait détecté est modifié.

11. Procédé de vocalisation d'un texte selon la revendication 1 dans lequel un signal sonore prédéfini masque le contenu audio représentant au moins un extrait détecté.

12. Procédé de vocalisation d'un texte selon la revendication 1 caractérisé en ce que le procédé comporte en outre une étape préalable (E0) de mémorisation d'un groupe d'au moins un mot indésirable dudit ensemble en association avec des données définissant au moins une caractéristique linguistique associée et un contexte d'usage dans lequel ledit groupe est indésirable.

13. Procédé de vocalisation selon la revendication 1 caractérisé en ce qu'il comporte en outre une étape de sélection (E10) d'une voix prédéterminée et en ce que le contenu audio est en outre produit en fonction de ladite voix sélectionnée.

14. Système (SYS) de vocalisation d'un texte comprenant :
- des moyens de détection (DET) dans ledit texte d'au moins un extrait de texte vérifiant un critère d'identification de mot indésirable;
- des moyens de production (PRO) d'un contenu audio, résultant d'une vocalisation dudit texte et dans lequel ledit au moins un extrait détecté n'est pas audible;
caractérisé en ce que les moyens de détection (DET) d'au moins un extrait de texte vérifiant ledit critère d'identification sont agencés pour déterminer si un extrait de texte possède au moins une caractéristique linguistique prédéfinie et un contexte d'usage identiques à ceux d'un groupe d'au moins un mot indésirable d'un ensemble prédéfini d'au moins un groupe d'au moins un mot indésirable.

15. Produit programme d'ordinateur comprenant des instructions pour mettre en œuvre les étapes du procédé de vocalisation d'un texte selon la revendication 1, lorsqu'il est chargé et exécuté par un processeur.

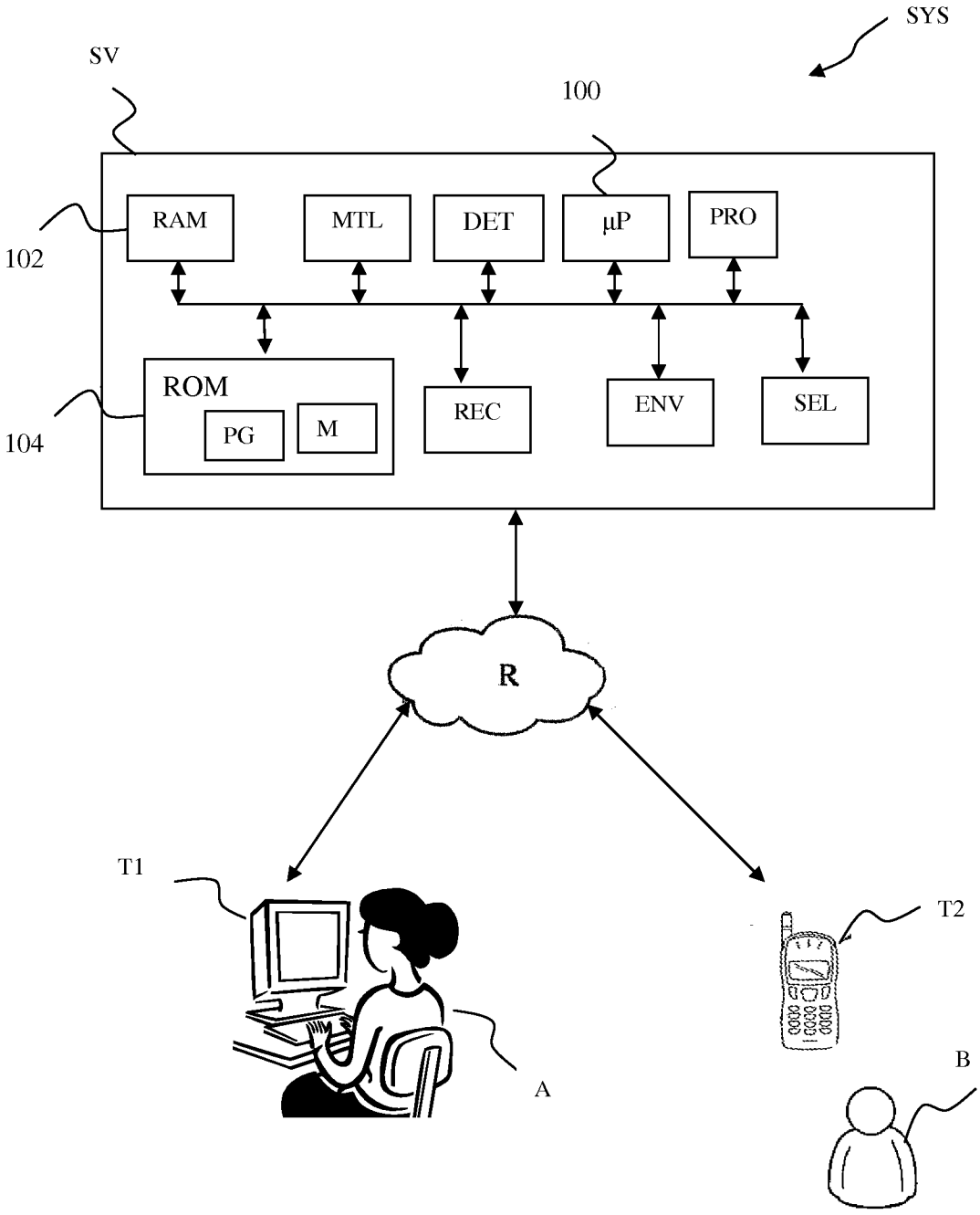

SYS
SV
100
RAM
MTL
DET
µP
PRO
102
ROM
PG
M
104
REC
ENV
SEL
R
T1
T2
A
B

Figure 1

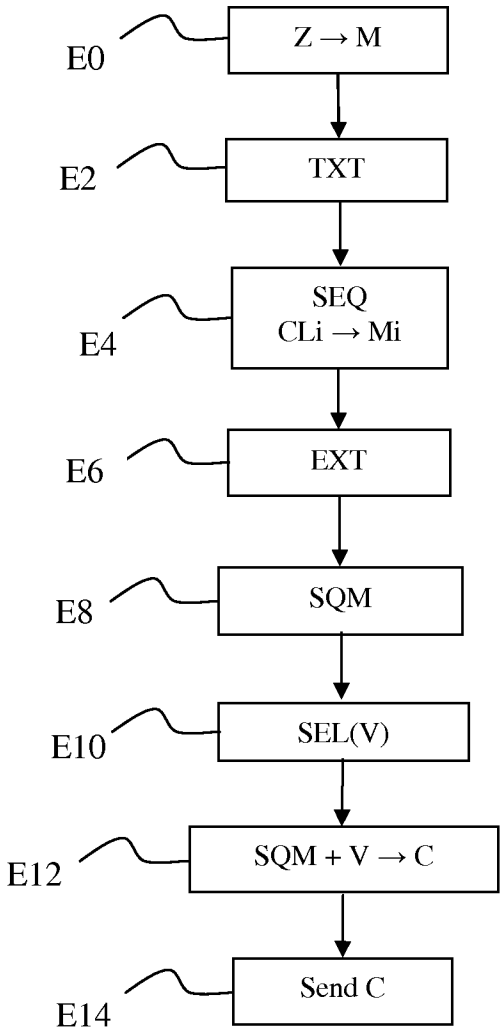

E0
Z → M
E2
TXT
E4
SEQ
CLi → Mi
E6
EXT
E8
SQM
E10
SEL(V)
E12
SQM + V → C
E14
Send C

Figure 2

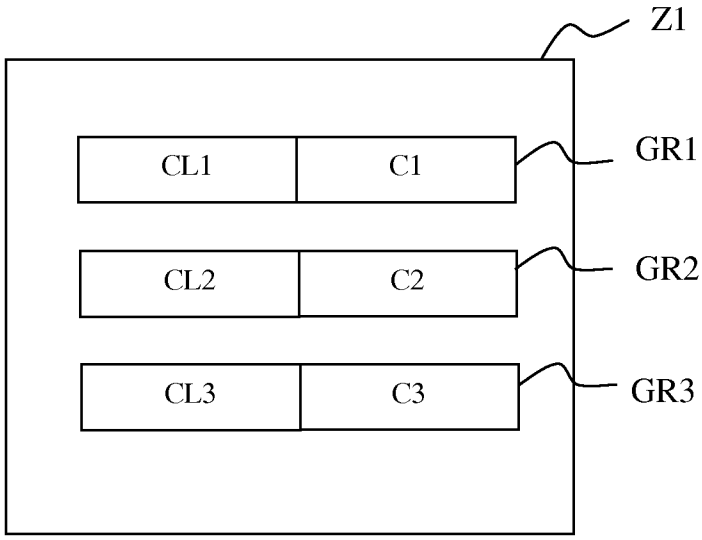


Figure 3

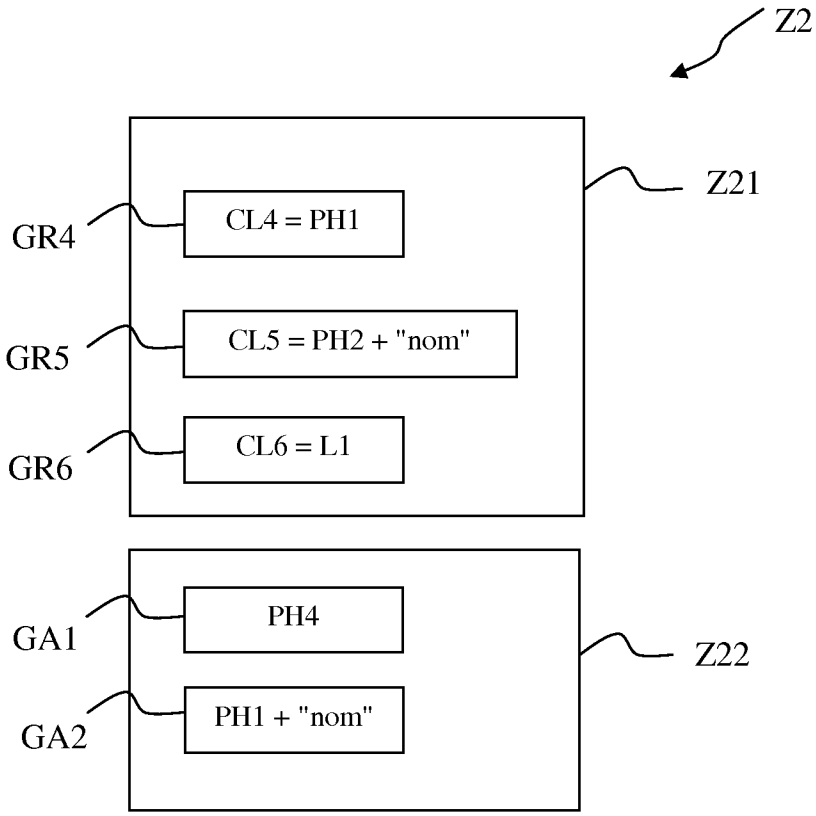


Figure 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/FR2011/052882

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L13/08 G10L13/02
ADD. G10L15/00 G10L15/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 2009/124272 A1 (WHITE MARC [US] ET AL) 14 May 2009 (2009-05-14) abstract; figures 5,13 paragraphs [0126], [0176]	1,12,14, 15 2-9,11, 13
X Y	US 2006/095262 A1 (DANIELI DAMON V [US]) 4 May 2006 (2006-05-04) abstract; figures 5,5a,7,9	1,12,14, 15 2-9,11, 13

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 February 2012	05/03/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Quélavoine, Régis

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/FR2011/052882

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009124272 A1	14-05-2009	NONE	
US 2006095262 A1	04-05-2006	NONE	

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°
PCT/FR2011/052882

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
INV. G10L13/08 G10L13/02
ADD. G10L15/00 G10L15/02

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)
EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X Y	US 2009/124272 A1 (WHITE MARC [US] ET AL) 14 mai 2009 (2009-05-14) abrégé; figures 5,13 alinéas [0126], [0176]	1,12,14, 15 2-9,11, 13

X Y	US 2006/095262 A1 (DANIELI DAMON V [US]) 4 mai 2006 (2006-05-04) abrégé; figures 5,5a,7,9	1,12,14, 15 2-9,11, 13

☐ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée	Date d'expédition du présent rapport de recherche internationale
28 février 2012	05/03/2012

Nom et adresse postale de l'administration chargée de la recherche internationale	Fonctionnaire autorisé
Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Quélavoine, Régis

Formulaire PCT/ISA/210 (deuxième feuille) (avril 2005)

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale n°
PCT/FR2011/052882

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 2009124272 A1	14-05-2009	AUCUN	
US 2006095262 A1	04-05-2006	AUCUN	