



DEMANDE DE BREVET EUROPEEN

(43) Date de publication:
31.03.2021 Bulletin 2021/13

(51) Int Cl.:
G10L 25/51^(2013.01) G10L 25/03^(2013.01)

(21) Numéro de dépôt: **20193073.2**

(22) Date de dépôt: **27.08.2020**

(84) Etats contractants désignés:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Etats d'extension désignés:
BA ME
Etats de validation désignés:
KH MA MD TN

(30) Priorité: **27.09.2019 FR 1910678**

(71) Demandeur: **ORANGE**
75015 Paris (FR)

(72) Inventeurs:
• **LE RAZAVET, Danielle**
92326 CHATILLON CEDEX (FR)
• **PERON, Katell**
92326 CHATILLON CEDEX (FR)
• **PRIGENT, Dominique**
92326 CHATILLON CEDEX (FR)

(74) Mandataire: **Vidon Brevets & Stratégie**
16B, rue de Jouanet
BP 90333
35703 Rennes Cedex 7 (FR)

(54) **DISPOSITIF, SYSTÈME ET PROCÉDÉ D'IDENTIFICATION D'UNE SCÈNE À PARTIR D'UNE SÉQUENCE ORDONNÉE DE SONS CAPTÉS DANS UN ENVIRONNEMENT**

(57) L'invention concerne un dispositif d'identification, un procédé d'identification et un système d'identification d'une scène dans un environnement, ledit environnement comprenant au moins un moyen de captation de son (C1, C2, C3). Le dispositif d'identification (INTRP) est configuré pour identifier ladite scène à partir d'au moins deux sons captés dans ledit environnement, chacun desdits au moins deux sons étant respectivement associé à au moins une classe de sons, ladite scène étant identifiée en tenant compte de l'ordre chronologique dans lequel lesdits au moins deux sons ont été captés.

[Fig 1]

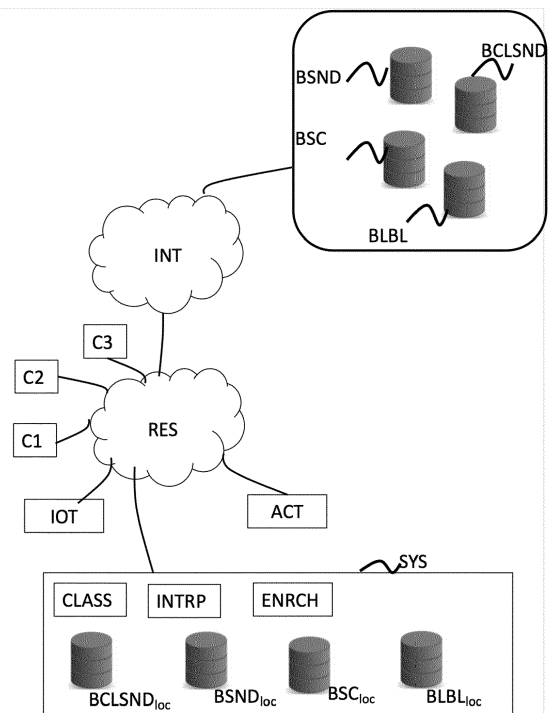


FIG. 1

Description

1. Domaine de l'invention

[0001] L'invention concerne un système d'identification d'une scène à partir de sons captés dans un environnement.

2. Art Antérieur

[0002] Des systèmes d'identifications de situations ou de cas d'usages peuvent être particulièrement intéressants pour un usage domestique ou professionnel, notamment dans le cas de situations détectées qui nécessitent des actions urgentes à opérer.

Par exemple, dans le cas d'une personne âgée maintenue à domicile, un système de surveillance pourrait identifier des situations nécessitant une intervention.

De tels systèmes peuvent aussi présenter un intérêt dans le cas de scènes sans caractère d'urgence, qui nécessitent de manière systématique un ensemble d'actions répétitives pour lesquelles l'automatisation de ces actions répétitives serait profitable à l'utilisateur (par exemple : verrouillage de la porte suite au départ d'un dernier occupant, mise en état de veille des radiateurs, ...).

De tels systèmes peuvent avoir un intérêt également pour des personnes avec handicap pour lesquelles le système peut être une aide.

De tels systèmes d'identification de situations, peuvent également avoir un intérêt dans un domaine domestique ou professionnel, par exemple dans le cas de systèmes de surveillance d'une entreprise ou d'un domicile lors de l'absence des personnes occupant l'entreprise ou le domicile, par exemple afin de prévenir une effraction, un incendie, un dégât des eaux, etc..., ou également dans le cas de systèmes proposant divers services aux usagers.

Aujourd'hui, il n'existe pas de solution industrielle de reconnaissance/ identification de situation, d'événement ou de cas d'usages dont le fonctionnement repose sur l'identification de plusieurs sons.

Les systèmes existants basés sur une reconnaissance de sons, comme celui de la compagnie « Audio Analytics », ne ciblent que l'identification d'un seul son parmi des sons ambiants captés. Un tel système n'identifie pas une situation associée au son identifié. L'interprétation du son est laissée à la responsabilité d'un tiers, libre de déterminer par exemple si un bris de glace identifié par l'équipement est dû à une intrusion ou à un accident domestique.

Les systèmes actuels d'identification de sons utilisent des bases de données de sons qui sont actuellement insuffisamment fournies et variées, à la fois en nombre de classes, mais aussi en nombre d'échantillons par classe. Ce nombre insuffisant d'échantillons ne rend pas compte de la variabilité des sons de la vie quotidienne et peut conduire à des identifications erronées.

Les techniques actuelles d'identification des sons et de leurs émetteurs se basent sur des comparaisons avec des modèles de classes de sons. Ces modèles sont construits à partir de bases de données souvent mal qualifiées. Ils sont alors susceptibles de générer des résultats approximatifs, voire des erreurs ou des contresens. Les Bases de Données de Sons disponibles et accessibles, gratuitement ou non (comme la base de données collaborative Freesound ou la base de données de la société Google « Google Audio Set ») sont très hétérogènes en termes de quantité et de qualité d'échantillons sonores.

De plus, elles sont dépourvues de systèmes performants de recherche ou de sélection, car les échantillons audio sont insuffisamment documentés et qualifiés. Lors de la recherche d'un échantillon, c'est après une série de tests auditifs manuels d'un grand nombre d'échantillons sonores repérés sur la base d'un ou 2 critères simples : émetteur, état (chat, chien, cafetière...) que la sélection d'un son ad-hoc peut être envisagée.

[0003] Toutes ces difficultés entraînent des incertitudes sur les classes de sons reconnues et diminuent sensiblement la performance d'un système d'identification d'une situation qui serait basé sur l'identification d'un son capté. Un tel système d'intelligence ambiante peut en être rendu inopérant, non adéquat (comme prévenir les gendarmes alors qu'on a simplement cassé un verre), voire dangereux.

[0004] Les systèmes d'analyse computationnelle des scènes sonores relatives à des activités (comme faire la cuisine), sont encore à l'état de recherche. Ils reposent sur l'analyse d'un corpus de sources non identifiées de sons récurrents, qui ne permettra donc pas à terme de mieux qualifier les classes de sons de référence pour entraîner les modèles. Aujourd'hui, grâce à des techniques d'apprentissage machine (ou machine-learning en anglais), ces procédés permettent de catégoriser des contextes habituels et répétitifs, mais ils sont mal adaptés à l'analyse d'événements sonores exceptionnels.

3. Exposé de l'invention

[0005] L'invention vient améliorer l'état de la technique. Elle concerne à cet effet un dispositif d'identification d'une scène dans un environnement, ledit environnement comprenant au moins un moyen de captation de sons. Le dispositif d'identification est configuré pour identifier ladite scène à partir d'au moins deux sons captés dans ledit environnement, chacun desdits au moins deux sons étant respectivement associé à au moins une classe de sons, ladite scène étant identifiée en tenant compte de l'ordre chronologique dans lequel lesdits au moins deux sons ont été captés.

L'invention propose ainsi un dispositif d'identification de scène à partir de sons captés dans un environnement. Avantageusement, un tel dispositif est basé sur une succession chronologique des sons captés et classés de sorte à discriminer des scènes lorsqu'un même son capté

peut correspondre à plusieurs scènes possibles.

En effet, un système d'identification de scène qui serait basé sur l'identification d'un unique son capté dans l'environnement serait peu fiable car dans certains cas, un son capté peut correspondre à plusieurs interprétations possibles, donc plusieurs situations ou scènes identifiées possibles. En effet, lorsqu'une scène n'est caractérisée que par un unique son, plusieurs scènes différentes peuvent correspondre à une même empreinte acoustique. Par exemple, un son de verre brisé peut être associé à une scène d'intrusion ou à un accident domestique, ces deux scènes correspondent à deux situations distinctes qui sont de nature à générer des réponses appropriées différentes. De plus, le dispositif d'identification selon l'invention permet de réduire l'incertitude sur l'identification de la source du son. En effet, certains sons peuvent avoir des empreintes acoustiques voisines qui sont difficiles à distinguer : par exemple un son d'aspirateur et un son de ventilateur, or ces sons ne sont respectivement pas révélateurs de la même situation. La prise en compte de plusieurs sons et de l'ordre chronologique dans lequel ces sons sont captés permet de fiabiliser les résultats du dispositif d'identification de scène. En effet, l'interprétation d'une scène est améliorée par la prise en compte de plusieurs sons captés lorsque cette scène se produit, ainsi que de l'ordre chronologique dans lequel ces sons se produisent.

[0006] Selon un mode particulier de réalisation de l'invention, la scène est identifiée parmi un groupe de scènes prédéfinies, chaque scène prédéfinie étant associée à un nombre prédéterminé de sons marqueurs, lesdits sons marqueurs d'une scène prédéfinie étant ordonnés de manière chronologique.

[0007] Selon un autre mode particulier de réalisation de l'invention, le dispositif est outre configuré pour recevoir au moins une donnée complémentaire fournie par un objet connecté dudit environnement et associer un label à une classe de sons d'un son capté ou à ladite scène identifiée. Selon ce mode particulier de l'invention, des objets connectés placés dans l'environnement dans lequel les sons sont captés transmettent au dispositif d'identification des données complémentaires.

De telles données complémentaires peuvent par exemple être une information de localisation du son capté, une information temporelle (heure, jour/nuit), une température, une information de type service : par exemple une information domotique indiquant qu'une lumière est allumée, une fenêtre est ouverte, une information météo fournis par un serveur....

Selon ce mode particulier de l'invention, des labels sont prédéfinis en fonction du type et de la valeur des données complémentaires susceptibles d'être reçues. Par exemple, des labels de type : jour/nuit sont définis pour des données complémentaires correspondant un horaire, des labels de type : chaud/froid/tempéré sont définis pour des données complémentaires correspondant à des valeurs de températures, des labels représentatifs de la localisation peuvent être définis pour des données com-

plémentaires correspondant à la localisation du son capté. Dans certains cas, les données complémentaires peuvent également correspondre directement à un label, par exemple un objet connecté peut transmettre un label de localisation qui lui a été préalablement renseigné...

Par la suite, un label peut également être appelé qualificatif.

Selon ce mode particulier de réalisation de l'invention, les données complémentaires permettent de qualifier (i.e. décrire sémantiquement) une classe de sons ou une scène identifiée. Par exemple, pour un son capté correspondant à de l'eau qui coule, une information de localisation du son capté permettra de qualifier la classe de son à l'aide d'un label associé à la localisation (par exemple : douche, cuisine, etc...).

[0008] Selon un autre mode particulier de réalisation de l'invention, le dispositif est outre configuré pour, lorsqu'un son capté est associé à plusieurs classes de sons possibles, déterminer une classe de sons dudit son capté à l'aide au moins une donnée complémentaire reçue. Selon ce mode particulier de réalisation de l'invention, les données complémentaires permettent de discriminer des sons ayant des empreintes acoustiques proches. Par exemple, pour un son capté correspondant à de l'eau qui coule, une information de localisation du son capté permettra de discriminer si le son doit être associé à une classe de son de type douche ou à une classe de son de type pluie.

En variante, les données complémentaires peuvent être utilisées pour affiner une classe de sons en créant de nouvelles classes de sons plus précises à partir de la classe de sons initiale. Par exemple, pour un son capté qui a été associé à une classe de sons correspondant à de l'eau qui coule, une information de localisation du son capté permettra de qualifier le son capté à l'aide d'un label associé à la localisation (par exemple : douche, cuisine, etc...). Une nouvelle classe de sons de type eau qui coule dans une pièce de type douche/cuisine peut être créée. Cette nouvelle classe de sons sera alors plus précise que la classe de sons initiale « eau qui coule ». Elle permettra une analyse plus fine lors des prochaines identifications de scène.

[0009] Selon un autre mode particulier de réalisation de l'invention, le dispositif est configuré en outre pour déclencher au moins une action à exécuter suite à l'identification de ladite scène.

[0010] Selon un autre mode particulier de réalisation de l'invention, le dispositif est configuré en outre pour transmettre à un dispositif d'enrichissement au moins une partie des données suivantes :

- une information indiquant la scène identifiée, et au moins deux classes de sons et un ordre chronologique associés à la scène identifiée,
- au moins une partie des fichiers audio correspondant aux sons captés associés respectivement à une classe de sons,
- le cas échéant au moins une classe de sons asso-

ciée à un label.

[0011] L'invention concerne également un système d'identification d'une scène dans un environnement, ledit environnement comprenant au moins un moyen de captation de son, ledit système comprend :

- un dispositif de classification configuré pour :
 - recevoir des sons captés dans ledit environnement,
 - déterminer pour chaque son reçu, au moins une classe de sons,
- un dispositif d'identification selon l'un quelconque des modes particuliers de réalisation décrits ci-dessus.

Selon un mode particulier de réalisation de l'invention, le système d'identification comprenant en outre un dispositif d'enrichissement configuré pour mettre à jour au moins une base de données avec au moins une partie des données transmises par le dispositif d'identification. Selon ce mode particulier de l'invention, le système selon l'invention permet d'enrichir des bases de données existantes, ainsi que les relations liant des éléments de ces bases de données entre eux, par exemple :

- une base de données de sons à l'aide d'au moins une partie des fichiers audio correspondant aux sons captés,
- une base de données de qualificatifs à l'aide des labels obtenus par les données complémentaires par exemple.
- les relations entre des fichiers audio, des classes de sons et des labels (qualificatifs) complémentaires provenant de données de capteurs ou de services.

[0012] L'invention concerne également un procédé d'identification d'une scène dans un environnement, ledit environnement comprenant au moins un moyen de captation de son, ledit procédé d'identification comprend l'identification de ladite scène à partir d'au moins deux sons captés dans ledit environnement, chacun desdits au moins deux sons étant respectivement associé à au moins une classe de sons, ladite scène étant identifiée en tenant compte de l'ordre chronologique dans lequel lesdits au moins deux sons ont été captés.

Selon un mode particulier de réalisation de l'invention, le procédé d'identification comprend en outre la mise à jour, d'au moins une base de données, à l'aide d'au moins une partie des données suivantes :

- une information indiquant la scène identifiée, et au moins deux classes de sons et un ordre chronologique associés à la scène identifiée,
- au moins une partie des fichiers audio correspondant aux sons captés associés respectivement à une

classe de sons,

- le cas échéant au moins une classe de sons associée à un label.

[0013] L'invention concerne également un programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé ci-dessus selon l'un quelconque des modes particuliers de réalisation décrits précédemment, lorsque ledit programme est exécuté par un processeur. Le procédé peut être mis en œuvre de diverses manières, notamment sous forme câblée ou sous forme logicielle. Ce programme peut utiliser n'importe quel langage de programmation, et être sous la forme de code source, code objet, ou de code intermédiaire entre code source et code objet, tel que dans une forme partiellement compilée, ou dans n'importe quelle autre forme souhaitable. L'invention vise aussi un support d'enregistrement ou support d'informations lisible par un ordinateur, et comportant des instructions d'un programme d'ordinateur tel que mentionné ci-dessus. Les supports d'enregistrement mentionnés ci-avant peuvent être n'importe quelle entité ou dispositif capable de stocker le programme. Par exemple, le support peut comporter un moyen de stockage, tel qu'une ROM, par exemple un CD ROM ou une ROM de circuit microélectronique, ou encore un moyen d'enregistrement magnétique, par exemple un disque dur. D'autre part, les supports d'enregistrement peuvent correspondre à un support transmissible tel qu'un signal électrique ou optique, qui peut être acheminé via un câble électrique ou optique, par radio ou par d'autres moyens. Les programmes selon l'invention peuvent être en particulier téléchargés sur un réseau de type Internet. Alternativement, les supports d'enregistrement peuvent correspondre à un circuit intégré dans lequel le programme est incorporé, le circuit étant adapté pour exécuter ou pour être utilisé dans l'exécution du procédé en question.

4. Liste des figures

[0014] D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description suivante de modes de réalisation particuliers, donnés à titre de simples exemples illustratifs et non limitatifs, et des dessins annexés, parmi lesquels :

[Fig 1] La figure 1 illustre un exemple d'environnement de mise en œuvre de l'invention selon un mode particulier de réalisation de l'invention,

[Fig 2] La figure 2 illustre des étapes du procédé d'identification d'une scène dans un environnement, selon un mode particulier de réalisation de l'invention,

[Fig 3] La figure 3 illustre schématiquement un dispositif d'identification d'une scène dans un environnement, selon un mode particulier de réalisation de l'invention,

[Fig 4] La figure 4 illustre schématiquement un dis-

positif d'identification d'une scène dans un environnement, selon un autre mode particulier de réalisation de l'invention,

[Fig 5] La figure 5 illustre schématiquement un dispositif d'identification d'une scène dans un environnement, selon un autre mode particulier de réalisation de l'invention.

5. Description d'un mode de réalisation de l'invention

[0015] L'invention propose par l'identification successive de sons captés dans un environnement la détermination d'un cas d'usage qui leur est associé.

Par cas d'usage, on entend ici un ensemble constitué d'un contexte et d'un événement. Le contexte est défini par des éléments de l'environnement, comme la localisation, les acteurs en jeu, le moment présent (jour/nuit), etc....

L'événement est singulier, occasionnel et fugace. L'événement marque une transition ou une rupture dans une situation vécue. Par exemple, dans une situation où une personne s'active dans une cuisine et réalise des tâches de préparation de repas, un événement peut correspondre au moment où cette personne s'entaille la main avec un couteau. Selon cet exemple, un cas d'usage est alors défini par le contexte comprenant la personne présente, la cuisine, et par l'événement d'accident de coupe.

Un autre exemple de cas d'usage est par exemple une scène de départ du domicile par son occupant. Selon cet exemple, le contexte comprend l'occupant du domicile, la localisation (entrée du domicile), des éléments avec lesquels l'occupant est susceptible d'interagir lors de ce cas d'usage (placard, clés, chaussures, vêtements, ..), et l'événement est le départ du domicile.

L'invention vient identifier de tels cas d'usages définis par un contexte et un événement qui se déroulent dans un environnement. De tels cas d'usages sont caractérisés par une suite chronologique de sons générés par le mouvement et les interactions entre les éléments/personnes de l'environnement lorsque le cas d'usage se produit. Il peut s'agir de sons spécifiques au contexte ou à l'événement du cas d'usage. C'est l'identification successive de ces sons et selon l'ordre chronologique dans lequel ils sont captés que le cas d'usage pourra être déterminé.

Par la suite, on utilisera indifféremment les termes de situation, cas d'usage ou scène.

[0016] On décrit ci-après la figure 1 qui illustre un exemple d'environnement de mise en œuvre de l'invention selon un mode particulier de réalisation de l'invention, en relation avec la figure 2 illustrant le procédé d'identification d'une scène.

L'environnement illustré en figure 1 comprend notamment un système SYS de collecte et d'analyse de sons captés dans l'environnement via un ensemble de moyens de captations de sons.

Un réseau de moyens de captation de sons est localisé dans l'environnement. De tels moyens de captation de

sons (C1, C2, C3) sont par exemple des microphones intégrés dans différents équipements situés dans l'environnement. Par exemple, dans le cas où l'environnement correspond à un domicile, il peut s'agir de microphones intégrés dans des terminaux mobiles lorsque l'utilisateur propriétaire du terminal est au domicile, de microphones intégrés dans des terminaux de type ordinateur, tablettes, etc... et de microphones intégrés dans tout type d'objets connectés tels que radio connectée, télévision connectée, assistant personnel, des terminaux intégrant des systèmes microphoniques dédiés à la reconnaissance de sons, etc...

On décrit ici le procédé selon l'invention à l'aide de trois microphones. Toutefois, le procédé selon l'invention peut également être mis en œuvre avec un seul microphone. De manière générale, le réseau des moyens de captation de sons peut comprendre tous types de microphones intégrés dans des équipements informatiques ou multimédia déjà placés dans l'environnement ou spécialement placés pour la reconnaissance sonore. Le système selon l'invention peut utiliser des microphones déjà localisés dans l'environnement pour d'autres usages. Il n'est ainsi pas toujours nécessaire de placer spécifiquement dans l'environnement des microphones.

Dans le mode particulier de réalisation décrit ici, l'environnement comprend également des objets connectés IOT, par exemple un assistant personnel, une TV connectée, ou une tablette, équipement domotique, ...

[0017] Le système SYS de collecte et d'analyse de sons communique avec les moyens de captations et éventuellement les objets connectés IOT via un réseau local RES, par exemple un réseau WiFi d'une passerelle domestique (non représentée).

L'invention n'est pas limitée à ce type de modes de communication. D'autres modes de communications sont également possibles. Par exemple, le système SYS de collecte et d'analyse de sons peut communiquer avec les moyens de captations et/ou les objets connectés IOT en Bluetooth ou via un réseau filaire.

Selon une variante, le réseau local RES est connecté à un réseau de données plus large INT, par exemple l'Internet via la passerelle domestique.

Selon l'invention, le système SYS de collecte et d'analyse de sons identifie à partir des sons captés dans l'environnement une scène ou un cas d'usage.

Dans le mode particulier de réalisation décrit ici, le système SYS de collecte et d'analyse de sons comprend notamment :

- un module de classification CLASS,
- un module d'interprétation INTRP,
- une base de données de fichiers audio BSND_{loc},
- une base de données de classes de sons BCLSN-D_{loc},
- une base de données de labels BLBL_{loc},
- une base de données de cas d'usage BSC_{loc}.

Le module de classification CLASS reçoit (étape E20)

des flux audio en provenance des moyens de captations. Pour cela, une application spécifique peut être installée dans les équipements de l'environnement intégrant des microphones pour que ces équipements transmettent le flux audio du son qu'ils captent. Une telle transmission peut être réalisée en continu, ou à intervalle réguliers, ou sur détection d'un son ayant une certaine amplitude. Suite à la réception d'un flux audio, le module de classification CLASS analyse le flux audio reçu pour déterminer (étape E21) la ou les classes de sons correspondant au son reçu via un ou plusieurs modèles de prédiction issus d'apprentissage automatique. Les sons de la base de données de sons sont mis en correspondance avec des classes de sons mémorisées dans la base de données de classes de sons $BCLSD_{loc}$. Le module de classification détermine la ou les classes de sons correspondant au son reçu en sélectionnant la ou les classes de sons associées à un son de la base de données de sons proche du son reçu. Le module de classification fournit ainsi en sortie au moins une classe CL_i de sons associée au son reçu avec un taux de probabilité P_i .

Les classes de sons retenues pour un son analysé correspondent à un seuil de probabilité acceptable préalablement déterminé. Autrement dit, on ne retient que les classes de sons pour lesquelles le taux de probabilité que le son reçu corresponde à un son associé à la classe de son est supérieur à un seuil prédéterminé.

Les classes de sons et leur probabilité associée sont ensuite transmises au module d'interprétation INTRP pour qu'il identifie la scène en train de se dérouler. Pour cela, le module d'interprétation s'appuie sur un ensemble de cas d'usage stockés dans la base de données de cas d'usage BSC_{loc} .

Un cas d'usage est défini sous la forme de N sons marqueurs, avec N un entier positif supérieur ou égal à 2. Les cas d'usage ont été définis au préalable de manière expérimentale et montés à l'aide d'une succession de sons caractérisant chaque étape de la scène. Par exemple, dans le cas d'une scène de départ du domicile, la succession de sons suivante a été montée : son du placard qui s'ouvre, son d'enfilage de manteau, son de placard qui se ferme, son de pas, son de porte qui s'ouvre, son de porte qui se ferme, son de fermeture à clé. Chaque montage de scène a été soumis à des personnes malvoyantes pour déterminer la pertinence des sons-étapes choisies et déterminer des sons marqueurs permettant d'identifier la scène.

L'expérimentation a permis d'identifier qu'un nombre de 3 sons marqueurs est suffisant pour identifier une scène et d'identifier, pour chaque scène, les sons marqueurs qui la caractérise, parmi les sons de la succession de sons montée lors de l'expérimentation.

Dans le mode particulier de réalisation de l'invention décrit ici, on considère donc $N=3$. D'autres valeurs sont toutefois possibles. Le nombre de sons marqueurs peut dépendre de la complexité de la scène à identifier. Dans d'autres variantes, seuls 2 sons marqueurs peuvent être utilisés, ou bien des sons marqueurs supplémentaires

($N > 3$) peuvent être ajoutés afin de préciser une scène ou distinguer des scènes trop proches acoustiquement. Le nombre de sons marqueurs utilisé pour identifier une scène peut également varier en fonction de la scène à identifier. Par exemple, certaines scènes pourront être définies par 2 sons marqueurs, d'autres scènes par 3 sons marqueurs, etc... Dans cette variante, le nombre de sons marqueurs n'est pas fixe.

[0018] La base de données de cas d'usage BSC_{loc} a ensuite été peuplée avec les scènes définies, chaque scène étant caractérisée par 3 sons marqueurs selon un ordre chronologique.

Selon un mode particulier de réalisation de l'invention, les scènes définies dans la base de données de cas d'usage BSC_{loc} peuvent provenir d'une base de données de cas d'usage plus large BSC, par exemple préalablement définie par un fournisseur de service selon l'expérimentation décrite ci-dessus ou tout autre méthode. Les scènes mémorisées dans la base de données de cas d'usage BSC_{loc} peuvent avoir été préalablement sélectionnées par l'utilisateur, par exemple lors d'une phase d'initialisation. Cette variante permet d'adapter les cas d'usage possibles à identifier pour un utilisateur en fonction de ses habitudes ou de son environnement.

Afin d'identifier une scène en cours, le module d'interprétation INTRP s'appuie donc sur une succession de sons reçus et analysés par le module de classification CLASS. Pour chaque son reçu par le module de classification CLASS, celui-ci transmet au module d'interprétation INTRP au moins une classe associée au son reçu et une probabilité associée.

Le module d'interprétation compare (étape E22) la succession de classes de sons reconnues par le module de classification, dans l'ordre chronologique de captation des sons correspondants, avec les sons-marqueurs caractérisant chaque scène de la base de données de cas d'usage BSC_{loc} .

Selon un mode particulier de réalisation de l'invention, le module d'interprétation INTRP prend aussi en compte des données complémentaires transmises (étape E23) au module d'interprétation INTRP par des objets connectés (IOT) placés dans l'environnement. De telles données complémentaires peuvent par exemple être une information de localisation du son capté, une information temporelle (heure, jour/nuit), une température, une information de type service : par exemple une information domotique indiquant qu'une lumière est allumée, une fenêtre est ouverte, une information météo fournis par un serveur....

Selon le mode particulier de l'invention décrit ici, des labels ou qualificatifs sont prédéfinis et stockés dans la base de données de labels LBL_{loc} . Ces labels dépendent du type et de la valeur des données complémentaires susceptibles d'être reçues. Par exemple, des labels de type : jour/nuit sont définis pour des données complémentaires correspondant un horaire, des labels de type : chaud/froid/tempéré sont définis pour des données complémentaires correspondant à des valeurs de tem-

pératures, des labels représentatifs de la localisation peuvent être définis pour des données complémentaires correspondant à la localisation du son capté.

Dans certains cas, les données complémentaires peuvent également correspondre directement à un label, par exemple, lorsque le son reçu par le module de classification a été transmis par un objet connecté, l'objet connecté peut transmettre avec le flux audio, un label de localisation correspondant à son emplacement...

Les données complémentaires permettent de qualifier (i.e. décrire sémantiquement) une classe de sons ou une scène identifiée. Par exemple, pour un son capté correspondant à de l'eau qui coule, une information de localisation du son capté permettra de qualifier la classe de son à l'aide d'un label associé à la localisation (par exemple: douche, cuisine, etc...). Selon cet exemple, le module d'interprétation INTRP peut alors qualifier la classe de sons associée à un son reçu.

Selon un autre exemple, pour un son capté associé à deux classes de sons qui sont proches acoustiquement, donc avec des taux de probabilités assez proches, une information de localisation du son capté permettra d'affiner la classe de sons la plus probable. Par exemple, un label associé à la localisation permettra de distinguer un son d'une classe de sons correspondant à de l'eau qui coule d'un robinet d'une classe de sons correspondant à de la pluie.

En sortie, le module d'interprétation fournit la scène identifiée et un taux de probabilité associé. En effet, comme pour l'identification d'une classe de sons correspondant à un son capté, l'identification d'une scène est faite par comparaison des sons captés avec des sons-marqueurs caractérisant un cas d'usage. Les sons captés ne sont pas identiques aux sons-marqueurs, car les sons marqueurs peuvent avoir été générés par d'autres éléments que ceux de l'environnement. De plus, le bruit ambiant de l'environnement peut également impacter l'analyse des sons.

Le module d'interprétation fournit également en sortie pour chaque classe de sons identifiées par le module de classification, des données complémentaires comme la scène identifiée, les données fournies par les objets connectés, les fichiers des sons captés.

Selon un mode particulier de réalisation de l'invention, lorsqu'une scène a été identifiée, le module d'interprétation INTRP transmet (étape 24) l'identification de la scène à un système d'actionneurs ACT connecté au système SYS via le réseau local RES ou bien via le réseau de données INT lorsque le système d'actionneurs n'est pas localisé dans l'environnement. Le système d'actionneurs permet d'agir en conséquence en fonction de la scène identifiée, en exécutant les actions associées à la scène. Par exemple, il peut s'agir de déclencher une alarme lors de l'identification d'une effraction, ou bien de prévenir un service d'urgence lors de l'identification d'un accident, ou bien tout simplement de brancher l'alarme lors de l'identification d'un départ du domicile....

Selon un mode particulier de réalisation de l'invention,

le système SYS de collecte et d'analyse de sons comprend également un module d'enrichissement ENRCH. Le module d'enrichissement ENRCH met à jour (étape 25) les bases de données de sons BSND_{loc}, de classes de sons BCLSND_{loc}, de cas d'usage BSC_{loc} et de labels BLBL_{loc} à l'aide des informations fournies en sortie par le module d'interprétation (INTRP). L'enrichisseur permet ainsi d'enrichir les bases de données à l'aide des fichiers sons des sons captés, permettant d'améliorer les analyses de sons ultérieures réalisées par le module de classification et d'améliorer l'identification d'une scène, en augmentant le nombre de sons associés à une classe de sons. L'enrichisseur permet également d'enrichir les bases de données à l'aides des labels obtenus, par exemple en associant un son capté mémorisé dans la base de données de sons BSND_{loc} le label obtenu pour ce son et mémorisé dans la base de données de label. Le module d'enrichissement permet d'enrichir de manière dynamique les données nécessaires à l'apprentissage du système SYS pour améliorer la performance de ce système.

[0019] Dans l'exemple décrit ici, les bases de données de sons BSND_{loc}, de classes de sons BCLSND_{loc}, de cas d'usage BSC_{loc} et de labels BLBL_{loc} sont locales. Elles sont par exemple stockées en mémoire du module de classification ou du module d'interprétation, ou dans une mémoire connectée à ces modules.

Dans d'autres modes particuliers de réalisation de l'invention, les bases de données de sons BSND_{loc}, de classes de sons BCLSND_{loc}, de cas d'usage BSC_{loc} et de labels BLBL_{loc} peuvent être distantes. Le système SYS de collecte et d'analyse des sons accède à ces bases de données, par exemple via le réseau de données INT.

[0020] Les bases de données de sons BSND_{loc}, de classes de sons BCLSND_{loc}, de cas d'usage BSC_{loc} et de labels BLBL_{loc} peuvent comprendre tout ou partie de bases de données distantes plus larges BSND, BCLSND, BSC et BLBL, par exemple des bases de données existantes ou fournies par un fournisseur de service.

Ces bases de de données distantes peuvent servir à initialiser les bases de données locales du système SYS et être mises à jour à l'aide des informations collectées par le système SYS lors de l'identification d'une scène. Ainsi, le système SYS de collecte et d'analyse des sons permet d'enrichir les bases de données de sons, de classes de sons, de cas d'usages et de labels pour d'autres utilisateurs.

[0021] Selon le mode particulier de réalisation décrit ci-dessus, les modules de classification, d'interprétation et d'enrichissement ont été décrits comme des entités séparées. Toutefois, tout ou partie de ces modules peut être intégrées dans un ou plusieurs dispositifs comme on le verra ci-dessous en relation avec les figures 3, 4 et 5.

[0022] La figure 3 illustre schématiquement un dispositif DISP d'identification d'une scène dans un environnement, selon un mode particulier de réalisation de l'invention. Selon un mode particulier de réalisation de l'in-

vention, le dispositif DISP a l'architecture classique d'un ordinateur, et comprend notamment une mémoire MEM, une unité de traitement UT, équipée par exemple d'un processeur PROC, et pilotée par le programme d'ordinateur PG stocké en mémoire MEM. Le programme d'ordinateur PG comprend des instructions pour mettre en œuvre les étapes du procédé d'identification d'une scène tel que décrit précédemment, lorsque le programme est exécuté par le processeur PROC.

A l'initialisation, les instructions de code du programme d'ordinateur PG sont par exemple chargées dans une mémoire avant d'être exécutées par le processeur PROC. Le processeur PROC de l'unité de traitement UT met notamment en œuvre les étapes du procédé d'identification d'une scène selon l'un quelconque de modes particuliers de réalisation décrits en relation avec la figure 2, selon les instructions du programme d'ordinateur PG. Le dispositif DISP est configuré pour identifier une scène à partir d'au moins deux sons captés dans ledit environnement, chacun desdits au moins deux sons étant respectivement associé à au moins une classe de sons, ladite scène étant identifiée en tenant compte de l'ordre chronologique dans lequel lesdits au moins deux sons ont été captés. Par exemple, le dispositif DIP correspond au module d'interprétation décrit en relation avec la figure 1.

Selon un mode particulier de réalisation de l'invention, le dispositif DISP comprend une mémoire BDDLOC comprenant une base de données de sons, une base de données de classes de sons, une base de données de cas d'usage et une base de données de labels.

Le dispositif DISP est configuré pour communiquer avec un module de classification configuré pour analyser des sons reçus et transmettre une ou des classes de sons associées à un son reçu, et éventuellement avec un module d'enrichissement configuré pour enrichir des bases de données telle que base de données de sons, base de données de classes de sons, base de données de cas d'usage et base de données de labels. Selon un mode particulier de réalisation de l'invention, le dispositif DISP est également configuré pour recevoir au moins une donnée complémentaire fournie par un objet connecté de l'environnement et associer un label à une classe de sons d'un son capté ou à ladite scène identifiée.

[0023] La figure 4 illustre schématiquement un dispositif d'identification DISP d'une scène dans un environnement, selon un autre mode particulier de réalisation de l'invention. Selon cet autre mode particulier de réalisation de l'invention, le dispositif DISP comprend les mêmes éléments que le dispositif décrit en relation avec la figure 3. Le dispositif DISP comprend en outre un module de classification CLASS configuré pour analyser des sons reçus et transmettre une ou des classes de sons associées à un son reçu et un module de communication COM2 adapté pour recevoir des sons captés par des moyens de captation de l'environnement.

[0024] La figure 5 illustre schématiquement un dispositif DISP d'identification d'une scène dans un environ-

nement, selon un autre mode particulier de réalisation de l'invention. Selon cet autre mode particulier de réalisation de l'invention, le dispositif DISP comprend les mêmes éléments que le dispositif décrit en relation avec la figure 4. Le dispositif DISP comprend en outre un module d'enrichissement ENRCH configuré pour enrichir des bases de données telle que base de données de sons, base de données de classes de sons, base de données de cas d'usage et base de données de labels.

Revendications

1. Dispositif d'identification d'une scène dans un environnement, ledit environnement comprenant au moins un moyen de captation de son (C1, C2, C3), ledit dispositif d'identification (DISP) est configuré pour identifier ladite scène à partir d'au moins deux sons captés dans ledit environnement, chacun desdits au moins deux sons étant respectivement associé à au moins une classe de sons, ladite scène étant identifiée en tenant compte de l'ordre chronologique dans lequel lesdits au moins deux sons ont été captés.
2. Dispositif d'identification d'une scène selon la revendication 1, dans lequel ladite scène est identifiée parmi un groupe de scènes prédéfinies, chaque scène prédéfinie étant associée à un nombre prédéterminé de sons marqueurs, lesdits sons marqueurs d'une scène prédéfinie étant ordonnés de manière chronologique.
3. Dispositif d'identification d'une scène selon l'une quelconque des revendications 1 ou 2, configuré en outre pour recevoir au moins une donnée complémentaire fournie par un objet connecté dudit environnement et associer un label à une classe de sons d'un son capté ou à ladite scène identifiée.
4. Dispositif d'identification d'une scène selon la revendication 3, configuré en outre pour, lorsqu'un son capté est associé à plusieurs classes de sons possibles, déterminer une classe de sons dudit son capté à l'aide de ladite au moins une donnée complémentaire reçue.
5. Dispositif d'identification d'une scène selon l'une quelconque des revendications 1 à 4, configuré en outre pour déclencher au moins une action à exécuter suite à l'identification de ladite scène.
6. Dispositif d'identification d'une scène selon l'une quelconque des revendications 1 à 5, configuré en outre pour transmettre à un dispositif d'enrichissement au moins une partie des données suivantes :

- une information indiquant la scène identifiée,

- et au moins deux classes de sons et un ordre chronologique associés à la scène identifiée,
 - au moins une partie des fichiers audio correspondant aux sons captés associés respectivement à une classe de sons, 5
 - au moins une classe de sons associée à un label.
7. Système d'identification d'une scène dans un environnement, ledit environnement comprenant au moins un moyen de captation de son, ledit système comprend : 10
- un dispositif de classification configuré pour recevoir des sons captés dans ledit environnement, et déterminer pour chaque son reçu, au moins une classe de sons, 15
 - un dispositif d'identification selon l'une quelconque des revendications 1 à 5. 20
8. Système d'identification d'une scène selon la revendication 7, dans lequel le dispositif d'identification est selon la revendication 6, le système d'identification comprenant en outre un dispositif d'enrichissement configuré pour mettre à jour au moins une base de données avec au moins une partie des données transmises par le dispositif d'identification. 25
9. Procédé d'identification d'une scène dans un environnement, ledit environnement comprenant au moins un moyen de captation de son, ledit procédé d'identification comprend l'identification de ladite scène à partir d'au moins deux sons captés dans ledit environnement, chacun desdits au moins deux sons étant respectivement associé à au moins une classe de sons, ladite scène étant identifiée en tenant compte de l'ordre chronologique dans lequel lesdits au moins deux sons ont été captés. 30 35
10. Procédé d'identification d'une scène selon la revendication 9, comprenant en outre la mise à jour, d'au moins une base de données, à l'aide d'au moins une partie des données suivantes : 40
- une information indiquant la scène identifiée, et au moins deux classes de sons et un ordre chronologique associés à la scène identifiée, 45
 - au moins une partie des fichiers audio correspondant aux sons captés associés respectivement à une classe de sons, 50
 - au moins une classe de sons associée à un label.
11. Programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé d'identification selon l'une quelconque des revendications 9 ou 10, lorsque le programme est exécuté par un processeur. 55

[Fig 1]

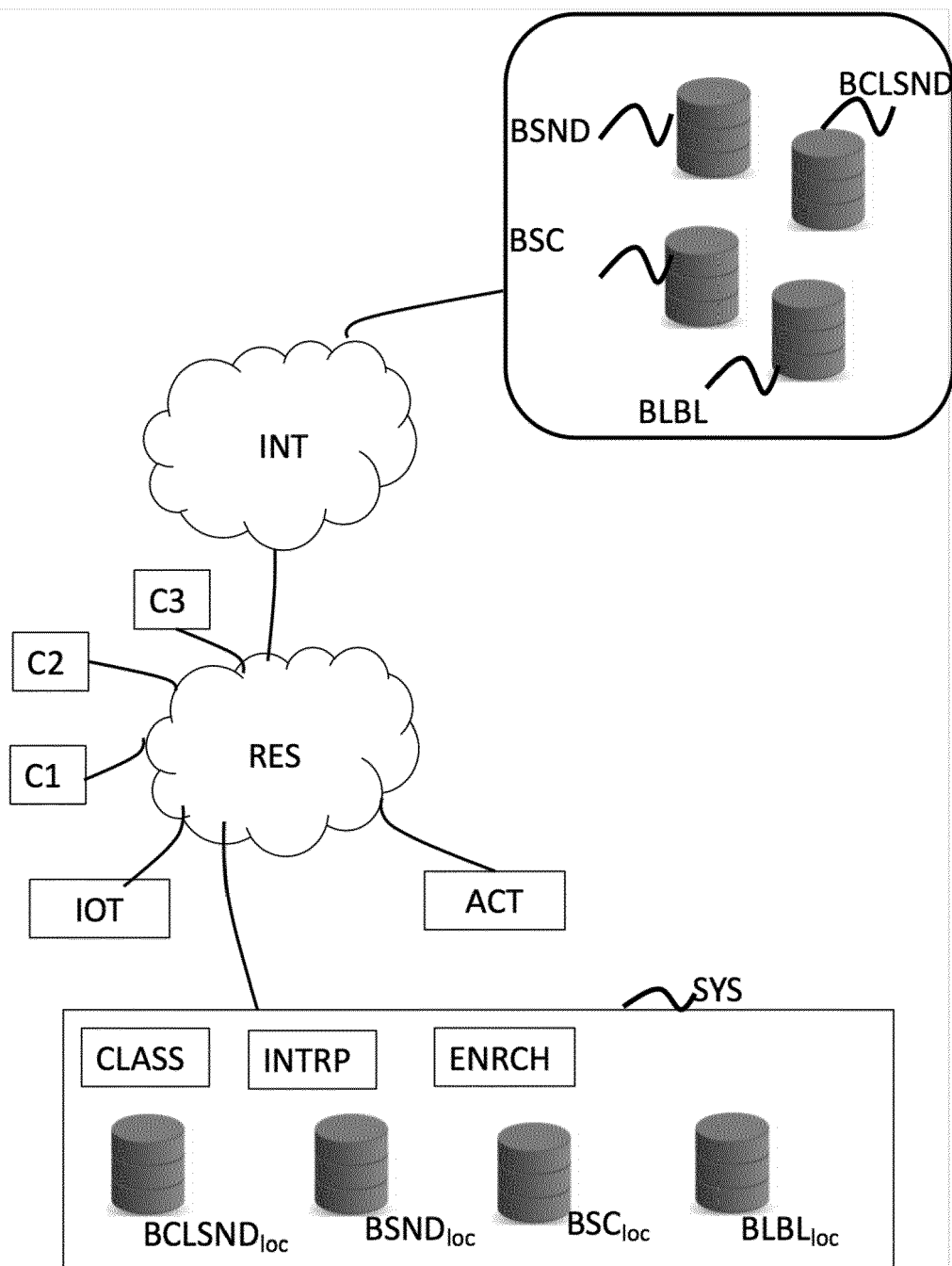


FIG. 1

[Fig 2]

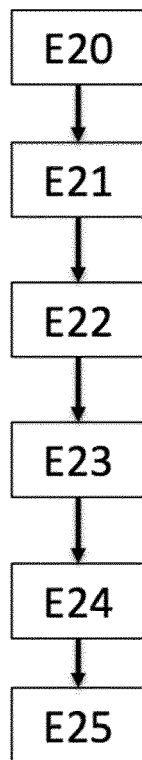


FIG. 2

[Fig 3]

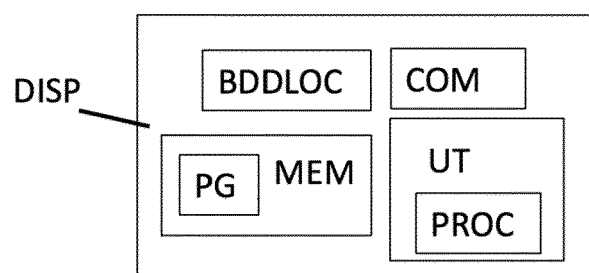


FIG. 3

[Fig 4]

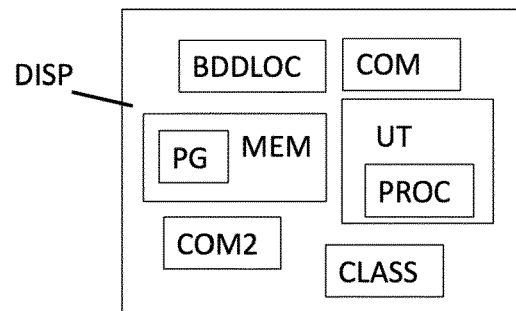


FIG. 4

[Fig 5]

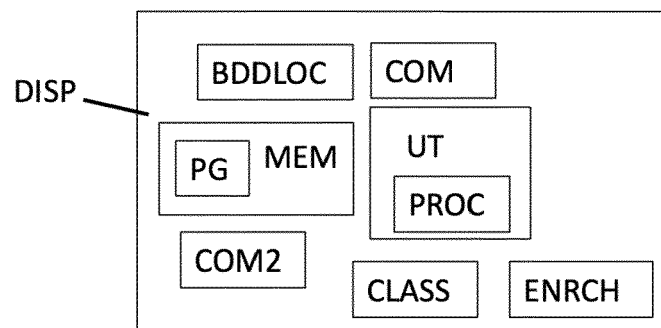


FIG. 5



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande

EP 20 19 3073

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
X	Brian Clarkson ET AL: "Auditory Context Awareness via Wearable Computing", Proceedings of 1998 Workshop on Perceptual User Interfaces, 1 janvier 1998 (1998-01-01), XP055677044, Extrait de l'Internet: URL:https://pdfs.semanticscholar.org/7e58/b584e1422d1be3a315dcdeef8016d5cb43f4.pdf [extrait le 2020-03-17]	1,2,5-11	INV. G10L25/51 ADD. G10L25/03
Y	* figure 3 * * section 3. The Current System * * section 1. Introduction, premier et second alinéas *	3,4	
X	CHAKRABARTY DEBMALYA ET AL: "Exploring the role of temporal dynamics in acoustic scene classification", 2015 IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS (WASPAA), IEEE, 18 octobre 2015 (2015-10-18), pages 1-5, XP032817953, DOI: 10.1109/WASPAA.2015.7336898 * abrégé *	1,9,11	
Y	BARCHIESI DANIELE ET AL: "Acoustic Scene Classification: Classifying environments from the sounds they produce", IEEE SIGNAL PROCESSING MAGAZINE, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 32, no. 3, 1 mai 2015 (2015-05-01), pages 16-34, XP011577488, ISSN: 1053-5888, DOI: 10.1109/MSP.2014.2326181 [extrait le 2015-04-02] * page 17, colonne droite, alinéa second *	3,4	
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche Munich		Date d'achèvement de la recherche 21 janvier 2021	Examineur Chétry, Nicolas
CATEGORIE DES DOCUMENTS CITES X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

EPO FORM 1503 03.82 (P04C02)