

(12) BREVET D'INVENTION

- (11) N° de publication : **MA 47830 B1** (51) Cl. internationale : **G06N 3/08; G06T 17/00; G06T 7/20**
- (43) Date de publication : **30.09.2021**

-
- (21) N° Dépôt : **47830**
- (22) Date de Dépôt : **27.12.2019**
- (71) Demandeur(s) : **Moroccan foundation for Advanced Science Innovation and Research (MAScIR), Rabat Design Center, Rue Mohamed Al Jazouli, Madinat Al Irfane, 10100 Rabat (MA)**
- (72) Inventeur(s) : **BOURJA OMAR ; NAGGAR OTHMANE ; ZENNAYI YAHYA ; AIT ABDELALI HAMD**
- (74) Mandataire : **AMMANI ABDELHAQ**

-
- (54) Titre : **Système et méthode pour la traduction et standardisation du langage des signes**
- (57) Abrégé : La présente invention présente un système et méthode basé sur des techniques de traitement de signale et d'intelligence artificielle pour l'objectif de faire une traduction interactif du langage des signes vers un langage naturel sélectionné (Français, Anglais, Arabe, ...), et inversement. D'une part ce système permet de récupérer une séquence vidéo provenant au moins d'une caméra filmant et d'une personne qui parle en langage des signes, puis de transcrire les signes vers une langue naturelle sous forme d'un texte ou un signal sonore ou autre, et d'autre part il permet de transcrire une source stockée sous forme d'un texte, audio ou vidéo, ou bien d'un signal sonore provenant d'un microphone vers un langage des signes sous forme d'un avatar affiché dans un écran.

Système et méthode pour la traduction et standardisation du langage des signes**Abrégé :**

La présente invention présente un système et méthode basé sur des techniques de traitement de
5 signale et d'intelligence artificielle pour l'objectif de faire une traduction interactif du langage
des signes vers un langage naturel sélectionné (Français, Anglais, Arabe, ...), et inversement.
D'une part ce système permet de récupérer une séquence vidéo provenant au moins d'une
caméra filmant et d'une personne qui parle en langage des signes, puis de transcrire les
signes vers une langue naturelle sous forme d'un texte ou un signal sonore ou autre, et
10 d'autre part il permet de transcrire une source stockée sous forme d'un texte, audio ou
vidéo, ou bien d'un signal sonore provenant d'un microphone vers un langage des signes
sous forme d'un avatar affiché dans un écran.

Système et méthode pour la traduction et standardisation du langage des signes

DOMAINE DE L'INVENTION

- 5 La présente invention concerne le domaine de la communication. Elle concerne en particulier un système et une méthode pour la traduction et la standardisation du langage des signes pour l'usage par les personnes sourdes.

ETAT DE L'ART

- 10 Dans le monde entier, les personnes sourdes communiquent souvent en langue des signes, avec des gestes des deux mains et des expressions faciales. Les langues des signes sont des langues naturelles à part entière dotées de leur propre grammaire et de leur lexique. Cependant, elles ne sont pas universelles bien qu'elles présentent des similitudes frappantes. Chaque pays a sa propre version modifiée qui contient de nouveaux signes pour son langage
- 15 familier. Même plus, pour le même mot ou concept, il peut correspondre à un signe différent dans différentes villes du même pays. Dans ce contexte, Le but de cette innovation est de modéliser un système interactif optimal basé sur les algorithmes d'intelligence artificielle (Figure 1), qui permettra aux personnes atteintes de surdit   d'avoir un acc  s au savoir (Apprentissage/Enseignement) et    l'int  gration sociale. Cela est par ailleurs un excellent
- 20 moyen de standardiser les langues des signes.

- Les personnes sourdes re  oivent tr  s peu d'aide    l'  ducation. La plupart des sourds n'ont pas la capacit   de lire ou d'  crire leur propre langue standard. Par cons  quent, leur seul moyen de communication est la langue des signes, qui n'est pas compr  hensible par les personnes normales et jusque-l  , non standardis  e. Cependant, les enseignants impliqu  s dans
- 25 l'  ducation des   l  ves sourds sont principalement des personnes qui entendent de fa  on normale, ils ma  trisent mal la langue des signes et manquent de personnels qualifi  s et d'outils d'assistance pour enseigner aux sourds    partir du texte   crit ou parl  .

Il existe plusieurs inventions et syst  mes pour le traitement des langages des signes.

Parmi ces syst  mes, on retrouve les brevets suivants :

- **WO2013/059894A1**, porte sur un système et une méthode de génération de contenu en langage des signes, qui est présenté au moyen d'un agent tridimensionnel virtuel.
- **EP3284019A1**, décrit un procédé pour une reconnaissance de langage des signes, consiste à recevoir au moyen d'un dispositif d'entrée, une entrée sur la base d'une pluralité de mouvements et de postures de main d'un utilisateur qui correspondent à une séquence de signes, à extraire une pluralité de caractéristiques à partir de l'entrée correspondant à la pluralité de mouvements et de postures de main, à identifier un début de la séquence de signes dans l'entrée sur la base d'un premier ensemble de caractéristiques dans la pluralité de caractéristiques et d'un premier modèle de Markov caché (HMM) stocké dans la mémoire et à identifier un premier signe dans l'entrée sur la base d'un second ensemble de caractéristiques dans la pluralité de caractéristiques et d'un second HMM stocké dans la mémoire. Ce procédé consiste également à générer une sortie correspondante au premier signe provenant de l'entrée.
- **WO2011107420A1**, décrit un système pour automatiser la traduction du langage parlé en langage des signes, il comprend les caractéristiques suivantes : une base de données, dans laquelle des données textuelles de mots et la syntaxe du langage parlé ainsi que des séquences de données vidéo avec les significations correspondantes en langage des signes sont stockées, et un ordinateur qui communique avec une base de données afin de traduire les données textuelles du langage parlé.
- **EP0586259B1**, décrit un système et un procédé de traduction de la langue des signes, permettant de reconnaître une langue de signes et de la traduire dans une langue parlée.

D'autres brevets partant sur l'amélioration des possibilités de communication des personnes sourdes en s'appuyant sur l'utilisation des approches probabilistes, des nouvelles technologies de l'information et de la communication (TIC).

Toutes ces solutions présente l'inconvénient d'être dépendant de la langue et limiter dans la capacité d'apprentissage pour faire face à la diversité des langues et des langages de signes qui en découle , et par la suite pouvoir servir dans n'importe quel environnement culturel et linguistique.

D'où l'intérêt de la présente invention qui par rapport à l'existant, elle présente les avantages suivant :

- Le système et la méthode de l'invention sont exclusivement basés sur les algorithmes d'intelligence artificielle (Deep learning, Machine learning, reinforcement learning, NLP), traitement du signal avec l'utilisation d'une ou plusieurs caméras/microphone/écran mobiles ou fixes.
- 5
- Le système et méthode de l'invention permettent d'extraire les informations à partir des fichiers audio, texte, vidéo et autres, de façon fiable, rapide (temps réel) et robuste.
 - Le système est muni d'un modèle de simulation du comportement humain sous forme d'un avatar virtuel 3D basé sur les algorithmes de « reinforcement learning ».
- 10
- Ce système comprend dans sa partie logicielle un module qui permet d'authentifier par la voix, empreinte digitale, mot de passe (chiffre, caractère), reconnaissance facial.

EXPOSE DETAILLE DE L'INVENTION

L'invention se base en grande partie sur un système d'apprentissage permanent et interactif entre l'utilisateur, son environnement et la machine. Le processus d'interaction selon la

15 **figure 1** est basé sur constitution d'une base de données labellisée dans un contexte de machine / deep learning qui est le facteur clé de succès de l'apprentissage automatique. Cette étape est composée des éléments principaux suivants :

- Les fichiers texte, vidéo et audio pour la construction des modèles (101);
 - Le prétraitement des données (102);
- 20
- L'unité de stockage des données (103);
 - L'extraction des caractéristiques (104);
 - Les algorithmes d'apprentissage automatique (IA) (106);
 - L'entraînement des modèles (107).

Selon un mode de fonctionnement, l'utilisation du système tel que présenté à la **Figure 2** Le

25 système innovant passe par les tâches suivantes:

- **La reconnaissance des signes dans une vidéo**

Le but principal de cette tâche est de concevoir un ensemble d'outils basés sur l'intelligence artificielle pour transformer une vidéo qui contient un(e) ou plusieurs mots/phrases sous format de langue des signes vers une transcription textuelle d'un

30 langage naturel. Vu que la langue des signes n'est pas internationale, la construction

d'un dictionnaire complet propre à la langue naturelle donnée est une phase nécessaire.

La reconnaissance des signes dans une vidéo passe par plusieurs étapes, les plus importantes sont :

- 5 ● La segmentation de la vidéo en plusieurs plans identifiants les transitions entre les signes, cette phase nous permet d'identifier le début et la fin du mot et ou phrase.
- La reconnaissance de chaque signe ou plusieurs signes pour les transcrire en texte tout en prenant en considération les expressions de visage de l'acteur.
- 10 ● La mise en place des algorithmes de segmentation et de transcription seront dotés d'intelligence pour s'adapter à tous les environnements de prise de la vidéo (parmi les problèmes d'analyse de la vidéo, sont les conditions d'acquisition, dans ce cas une phase de prétraitement est obligatoire).

15 - **La traduction des langages des signes**

L'objectif principal de cette seconde tâche est fortement lié à la première au sens où il s'agit de partir sur des images segmentée grâce aux résultats des algorithmes de reconnaissance et de segmentation de la première tâche pour les retranscrire en langage naturel (Français, Anglais, Arabe, etc) en prenant soin d'assurer une

20 cohérence grammatical et morphosyntaxique correcte, intelligible et adaptée à la langue cible.

Les algorithmes de traduction devront passer par une phase d'apprentissage de structure de la langue des signes. La structure de la langue une fois identifiée, il sera possible de la construire via des algorithmes d'intelligence artificielle, traditionnels

25 (NLP) combinés à du Machine Learning / Deep Learning, un traducteur de langue des signes vers la langue naturel.

La réalisation de cette tâche passe par les étapes suivantes :

- Mise en place d'algorithmes d'apprentissage (Machine Learning / Deep Learning) sur les données construites avec des tests et itérations
- 30 jusqu'à la stabilisation d'un modèle exhaustif intégrant l'ensemble des signes et structures de la langue des signes.

- Mise en place du système de traduction sur la base des données et algorithmes entraînés avec analyse des rejets et réentraînement au besoin (boosting).

5 - **La traduction des langages naturels à la langue des signes**

Le but principal de cette tâche est de se corrélér à la précédente au sens où il s'agit de pouvoir générer à partir de la langue naturel parlée, une liste de signes, cela implique notamment que la génération de la suite de signes doit répondre à différents critères tels que la structure cible de la langue des signes. Cette complexité est à prendre en

10 considération car en fonction de la position du mot dans la phrase, des aspects morphosyntaxiques et des opinions exprimés, les signes équivalents et leurs dynamiques tendent à changer, ainsi que les expressions faciales associées.

- **Génération d'avatar 3D représentant les signes**

L'objectif principal de cette tâche se concentre sur la conception et l'animation d'un

15 avatar virtuel 3D capable de générer des gestes de la Langue des Signes (LS). Les modèles de génération de LS doivent être plus naturels. Pour ce faire, il faut comprendre comment des individus sourds perçoivent la LS en fonction de la quantité d'informations disponible. Étudier les paramètres importants pour que la LS générée automatiquement soit compréhensible, et dans ce but s'inscrit le besoin d'étudier le

20 mouvement de LS afin de fournir des modèles de génération de mouvements plus naturels et plus compréhensibles. Aussi, consiste à étudier et vérifier les lois classiques du contrôle dans les deux types de mouvements « signes » et « transitions » en prenant en considération les différentes types de transition (transition intersigne et transition intra-signes), ainsi qu'identifier des différences entre la cinématique des

25 mouvements des signes et des transitions. Proposition de création d'un corpus de données 3D pour LS pour la génération d'animation de l'avatar virtuel pour une utilisation future sur différents domaines d'application. La tâche doit prendre en considération les déplacements des mains qui semblent porteurs d'une partie non négligeable de l'information linguistique.

30 La **Figure 3** présente le mode de fonctionnement du système sur un ordinateur (202), une tablette (208) ou un téléphone intelligent (Smartphone). Il est composé des éléments suivants:

- Une ou plusieurs caméras (203);

5. Méthode selon la revendication 1 **caractérisée en ce que** l'avatar 3D est capable d'apprendre via un système d'entraînement basé sur les algorithmes de l'intelligence artificielle notamment l'apprentissage par renforcement et de parler plusieurs langages des signes de manière interactif avec plusieurs utilisateurs selon leurs besoin (apprentissage/communication).

Revendications modifiées :

1. Méthode pour la traduction et standardisation du langage des signes comprenant les étapes suivantes :
 - L'acquisition des signes sous un format vidéo, séquence des images et autres (101), permettant de constituer une base de données labélisée grâce à des moyens (102), (103) et (104), qui contient les expressions suivantes : gesticulation, facile, verbale et textuel.
 - L'entraînement et générations des modèles grâce à des algorithmes de Machine Learning et de Deep Learning (105) et (106) afin de construire les dictionnaires de langues des signes (107).
 - L'application des algorithmes de traitement du langage naturel (NLP et autres), de traitement du signal (audios, vidéos, images, etc) et d'intelligence artificielle analysant les suites de signes (textuel, verbale, visuel, ...) pour détecter l'unité linguistiques générées par l'utilisateur.
 - L'application des algorithmes de l'apprentissage par renforcement grâce à des algorithmes choisis parmi "Natural Policy Gradient, Q-learning", "Deep Q-learning", "Deep Deterministic Policy Gradient DDPG", "Optimize Natural Policy Gradient with Conjugate Gradient TRPO" pour animer un avatar intelligent 3D (207) de façon interactive à travers la reconnaissance en temps réel des signes dans une vidéo ou séquence d'images au via une caméra.
2. Méthode selon la revendication 1 **caractérisée en ce que** pour la reconnaissance des signes dans une vidéo ou une séquence des images passe par plusieurs étapes :
 - L'acquisition des signes qui s'adaptent à tous les environnements de prise de la vidéo et d'audio.
 - La segmentation de la vidéo ou l'audio, en plusieurs plans identifiants les transitions entre les signes ou mots, cette phase nous permet d'identifier le début et la fin du mot et ou phrase ;
 - La reconnaissance de chaque signe ou plusieurs signes pour les transcrire en (texte, vocale, un autre signe) tout en prenant en considération les expressions de visage de l'utilisateurs, les mouvements de la main ou une expression vocale ;
3. Méthode selon la revendication 1 **caractérisée en ce que** le modèle de simulation du comportement humain (avatar 3D intelligent) a une capacité de communiqué avec plusieurs utilisateurs simultanément à traverser un algorithme de détection multiple des signes.
4. Méthode selon la revendication 1 **caractérisée en ce que** pour animé l'avatar 3D, le système utilisé les algorithmes d'apprentissage par renforcement (Natural Policy Gradient, Q-learning, Deep Q-learning, Deep Deterministic Policy Gradient DDPG, Optimize Natural Policy Gradient with Conjugate Gradient TRPO, etc).

- La barre de recherche (206);
- Les fichiers audio, vidéo, texte (205);
- Un avatar 3D (207);
- La barre des langues Arabe, Anglais, Français,etc (204);
- Les utilisateurs du système (201).

5

Selon l'invention (figure 1) le système contient plusieurs méthodes dédiées à l'acquisition des signes sous un format vidéos, séquence des images et autres (101), permettant de constituer une Data-Set pour l'entraînement des modèles (107) et la constitution des dictionnaires.

10

Le système est muni des méthodes et des modèles de traitement du langage naturel (NLP et autres), de traitement du signal (audios, vidéos, images, etc) et d'intelligence artificiel analysant les suites de signes (gustelle, verbale, visuel, ...) pour détecter l'unité linguistiques générées.

15

Le système permet la traduction de la langue naturel écrite /parlée en langue des signes, pour les besoins communication et pédagogiques (apprentissage).

20

Selon un autre aspect de l'invention et pour animé l'avatar 3D, le système utilisé les algorithmes d'apprentissage par renforcement (Natural Policy Gradient, Q-learning, Deep Q-learning, Deep Deterministic Policy Gradient DDPG, Optimize Natural Policy Gradient with Conjugate Gradient TRPO, etc).

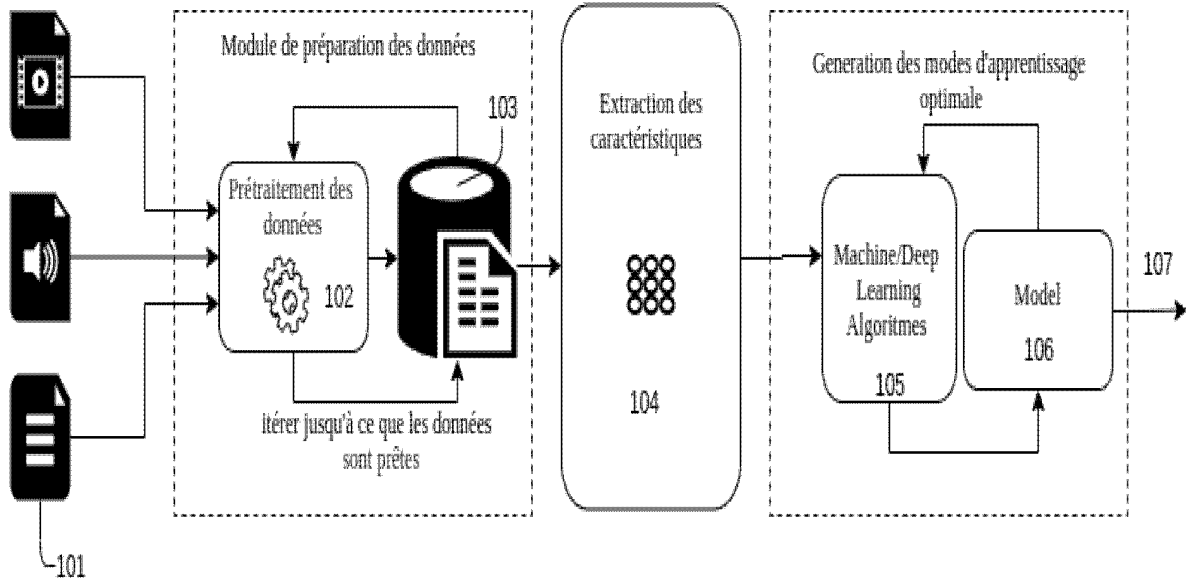


Figure 1

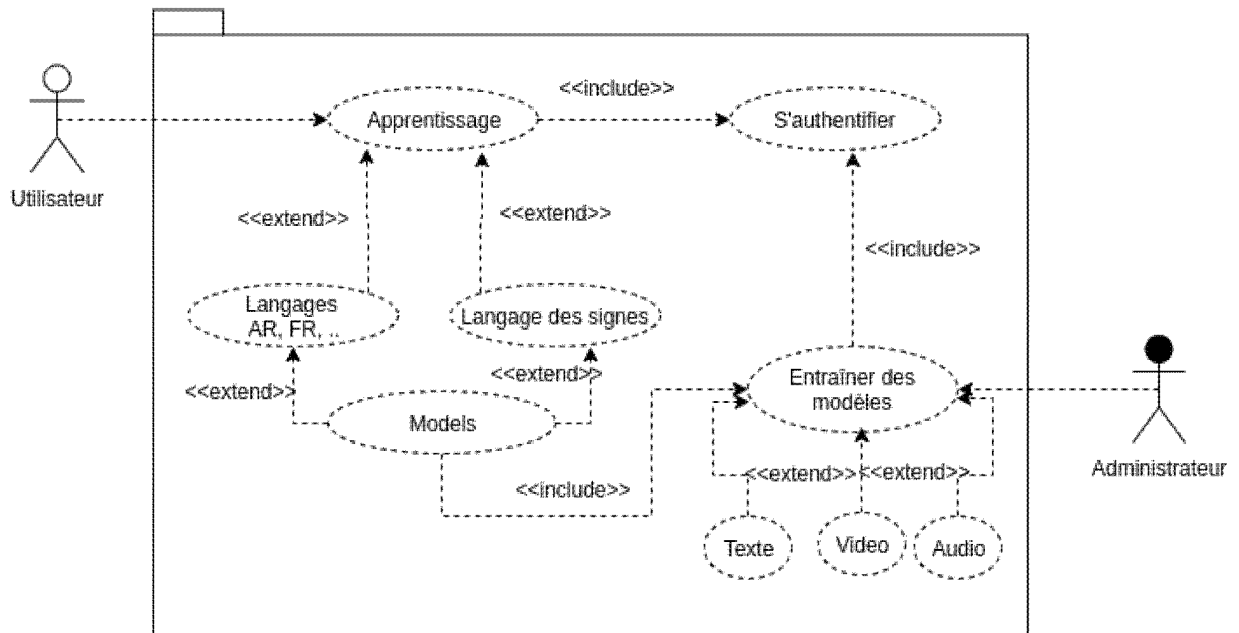


Figure 2

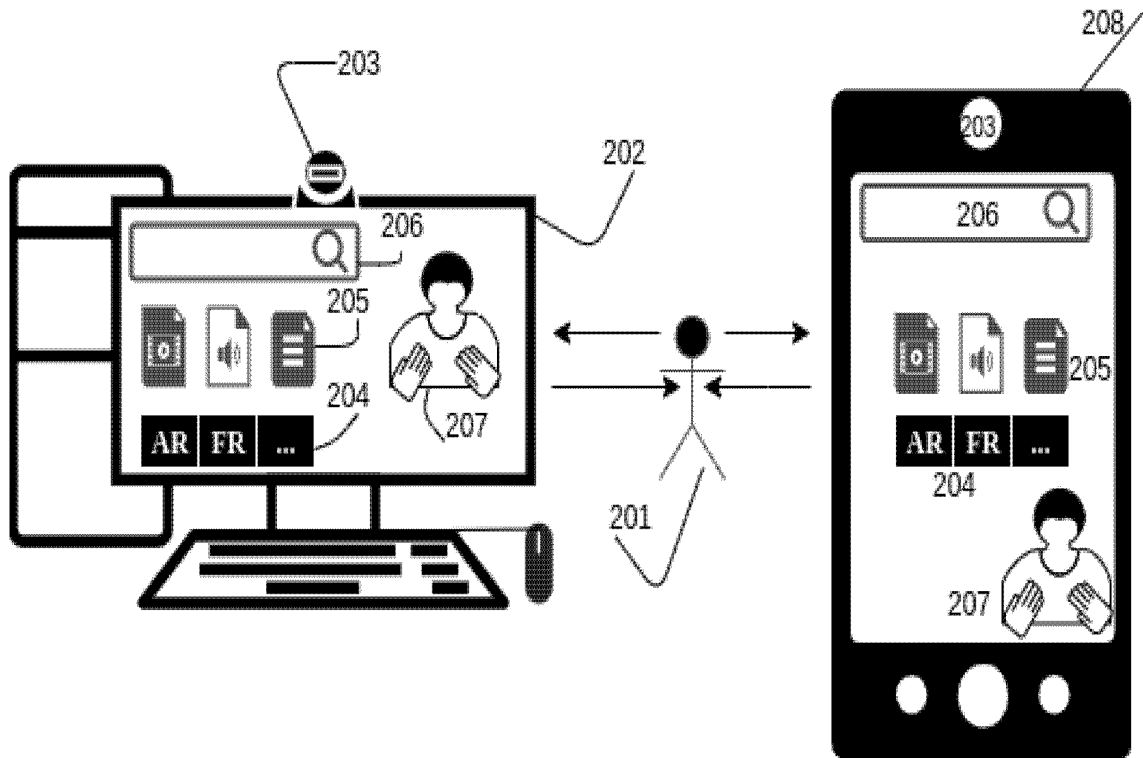


Figure 3

RAPPORT DE RECHERCHE DEFINITIF AVEC OPINION SUR LA BREVETABILITE

Établi conformément à l'article 43.2 de la loi 17-97 relative à la protection de la propriété industrielle telle que modifiée et complétée par la loi 23-13

Renseignements relatifs à la demande	
N° de la demande : 47830	Date de dépôt : 27/12/2019
Déposant : Moroccan foundation for Advanced Science Innovation and Research (MAScIR)	
Intitulé de l'invention : Système et méthode pour la traduction et standardisation du langage des signes	
Classement de l'objet de la demande : CIB : G06F40/40, G06F40/55, G06F40/56, G06N 3/08, G06T 7/20, G06T 17/00 CPC : G09B21/009, G06K9/00355	
Le présent rapport contient des indications relatives aux éléments suivants : Partie 1 : Considérations générales <input checked="" type="checkbox"/> Cadre 1 : Base du présent rapport <input type="checkbox"/> Cadre 2 : Priorité Partie 2 : Opinion sur la brevetabilité <input type="checkbox"/> Cadre 3 : Remarques de clarté <input type="checkbox"/> Cadre 4 : Observations à propos de revendications modifiées qui s'étendent au-delà du contenu de la demande telle qu'initialement déposée <input type="checkbox"/> Cadre 5 : Défaut d'unité d'invention <input type="checkbox"/> Cadre 6 : Observations à propos de certaines revendications exclues de la brevetabilité <input checked="" type="checkbox"/> Cadre 7 : Déclaration motivée quant à la Nouveauté, l'Activité Inventive et l'Application Industrielle	
Examineur: Ilham Oubiyi Téléphone: (+212) 5 22 58 64 14	Date d'établissement du rapport : 26/08/2021 

Partie 1 : Considérations générales**Cadre 1 : base du présent rapport**

Les pièces suivantes servent de base à l'établissement du présent rapport :

- Demande telle qu'initialement déposée
- Demande modifiée suite à la notification du rapport de recherche préliminaire :
- Revendications
5
- Observations à l'appui des revendications maintenues
- Observations des tiers suite à la publication de la demande
- Réponses du déposant aux observations des tiers
- Nouveaux documents constituant des antériorités :
- Suite à la recherche complémentaire (Couvrant les documents de l'état de la technique qui n'étaient pas disponibles à la date de la recherche préliminaire)
 - Suite à la recherche additionnelle (couvrant les éléments n'ayant pas fait l'objet de la recherche préliminaire)
- Observations à l'encontre de la décision de rejet

Partie 2 : Opinion sur la brevetabilité**Cadre 7 : Déclaration motivée quant à la Nouveauté, l'Activité Inventive et l'Application Industrielle**

Nouveauté	Revendications 1-5	Oui
	Revendications aucune	Non
Activité inventive	Revendications 1-5	Oui
	Revendications aucune	Non
Application Industrielle	Revendications 1-5	Oui
	Revendications aucune	Non

Il est fait référence aux documents suivants:

D1 : US20190251343 A1

1. Nouveauté

Aucun des documents cités dans le rapport de recherche préliminaire ne divulgue l'ensemble des caractéristiques techniques énoncées dans les revendications 1-5. Par conséquent, l'objet des revendications 1-5 est nouveau au sens de l'art. 26 de la loi 17-97 telle que modifiée et complétée par la loi 23-13.

2. Activité inventive

Le document D1 (les références entre parenthèses s'appliquent à ce document) qui est considéré comme l'état de la technique le plus proche de l'objet de la revendication 1, divulgue Un système et son procédé pour la traduction automatique et la standardisation du langage des signes (§[0005]) doté d'une intelligence d'apprentissage permanent et interactif (§[0039], §[0070], fig. 1) adapté pour différentes langues (§[0035]), dernière phrase du §[0033]), entre l'utilisateur, son environnement et la machine en utilisant (voir figures, §[0034]):

- Des moyens pour l'acquisition de l'information à traduire sous forme audio et/ou vidéo (revendication 1, fig. 1) et une bibliothèque de gestes de langage visuel de base ;
- Une interface utilisateur (§[0039]) ;
- L'application des algorithmes de traitement du langage naturel ;
- L'application des algorithmes d'apprentissage automatique permanent et interactif entre l'utilisateur, son environnement et la machine (§[0055]) ;
- Une intelligence embaquée sous forme d'algorithmes pour la transformation des fichiers audio et vidéo en langages des signes (§[0055]) ;
- Un entraînement par renforcement de réseaux neuronaux peut être utilisé pour améliorer les performances et augmenter la flexibilité et l'adaptabilité des modes de réalisation (§[0055]).

Par conséquent, l'objet de la revendication 1 diffère donc de ce système connu en ce qu'il contient :

- L'entraînement et générations des modèles grâce à des algorithmes de Machine Learning et de Deep Learning afin de construire les dictionnaires de langues des signes.
- L'application des algorithmes de l'apprentissage par renforcement grâce à des algorithmes choisis pour animer un avatar 3D de façon interactive à travers la reconnaissance en temps réel des signes dans une vidéo ou séquence d'images au via une caméra.

Le problème objectif que la présente invention se propose de résoudre peut donc être considéré comme celui de donner plus d'autonomie aux systèmes de communication par le langage des signes.

La solution à ce problème proposé dans la revendication indépendante 1 de la présente demande est considérée comme impliquant une activité inventive. En effet, l'homme du métier ne serait pas parvenu d'une manière évidente à reproduire l'invention revendiquée en partant de D1. Aussi, aucun enseignement n'a été trouvé dans le reste de l'état de la technique disponible qui aurait incité la personne du métier, en partant du document D1, à atteindre le résultat recherché. Par conséquent, l'objet de la revendication 1 implique une activité inventive au sens de l'article 28 de la loi 17-97 telle que modifiée et complétée par la loi 23-13.

Les revendications 2-5 dépendent de la revendication 1 dont l'objet est considéré inventif, comme indiqué auparavant, et elles satisfont donc également, en tant que telles, aux exigences de l'article 28 de la loi 17-97 telle que modifiée et complétée par la loi 23-13 concernant l'activité inventive.

3. Application industrielle

L'objet de la présente invention est susceptible d'application industrielle au sens de l'article 29 de la loi 17-97 telle que modifiée et complétée par la loi 23-13, parce qu'il présente une utilité déterminée, probante et crédible.