

## (12) BREVET D'INVENTION

(11) N° de publication : **MA 54537 B1** (51) Cl. internationale : **A61B 5/16; A61B 5/00**

(43) Date de publication :  
**30.08.2024**

---

(21) N° Dépôt :  
**54537**

(22) Date de Dépôt :  
**20.12.2019**

(30) Données de Priorité :  
**20.12.2018 CH 15712018**

(86) Données relatives à la demande internationale selon le PCT:  
**PCT/IB2019/061184 20.12.2019**

(71) Demandeur(s) :  
**CM Profiling Sàrl, Haldenweg 56 3074 Muri b. Bern (CH)**

(72) Inventeur(s) :  
**MATTEUCCI, Caroline ; BESSERT-NETTELBECK, Joanna**

(74) Mandataire :  
**IPPro**

(86) N° de dépôt auprès de l'organisme de validation : 19839416.5

---

(54) Titre : **SYSTÈME ET PROCÉDÉ DE LECTURE ET D'ANALYSE DU COMPORTEMENT  
COMPRENANT LES EXPRESSIONS VERBALES, CORPORELLES ET FACIALES AFIN  
DE DÉTERMINER LA CONGRUENCE D'UNE PERSONNE**

(57) Abrégé : La présente invention concerne un système de traitement de données pour déterminer la congruence ou l'incongruence entre l'expression corporelle et les paroles d'une personne, comprenant une machine d'auto-apprentissage, telle qu'un réseau neuronal, configurée pour recevoir comme entrée un ensemble de données comprenant : des données approuvées d'un ensemble de paroles analysées de personnes, lesdites données approuvées comprenant pour chaque parole analysée : \* un ensemble de séquences vidéo, comprenant des séquences audio et des séquences visuelles, chaque séquence audio correspondant à une séquence visuelle, et \* un indicateur de congruence approuvé pour chacune desdites séquences vidéo - ladite machine d'auto-apprentissage étant entraînée de sorte que le système de traitement de données est capable de fournir comme résultat un indicateur de congruence.

REVENDEICATIONS

1. Procédé de fourniture d'indicateurs de congruence ou d'incongruence entre le langage corporel et le discours d'une personne, mis en œuvre par un
- 5 système de traitement de données pour déterminer la congruence ou l'incongruence entre le langage corporel et le discours de la personne, le procédé comprenant les étapes ci-dessous consistant à :
- a/ fournir un dispositif d'enregistrement vidéo apte à enregistrer des images d'un sujet incluant le visage et au moins certaines parties du corps ;
- 10 b/ enregistrer une vidéo d'un discours de cette personne (130) à l'aide dudit dispositif d'enregistrement vidéo (126), ladite vidéo étant divisée en « n » séquences vidéo comprenant « n » séquences d'images (ou « n » séquences visuelles) et « n » séquences audio correspondantes ;
- c/ pour chaque séquence d'images, détecter au moins un repère visuel,  $V_c$ , et
- 15 attribuer au moins une note parmi « positive »,  $V_{c+}$ , « neutre »,  $V_{c0}$ , et « négative »,  $V_{c-}$ , pour chaque repère visuel,  $V_c$ ,
- d/ pour chaque séquence audio, détecter au moins un repère audio,  $A_c$ , et attribuer au moins une note parmi « positive »,  $A_{c+}$ , « neutre »,  $A_{c0}$ , et « négative »,  $A_{c-}$ , pour chaque repère audio,  $A_c$ ,
- 20 e/ pour chaque séquence vidéo, comparer la note dudit repère audio,  $A_c$ , à la note dudit repère visuel,  $V_c$ , et donner un indicateur de congruence qui est :
- un indicateur de congruence positif si les deux notes sont soit positives ( $V_{c+}$  et  $A_{c+}$ ), soit négatives ( $V_{c-}$  et  $A_{c-}$ ) ;
- un indicateur de congruence négatif si l'une des notes est positive et
- 25 l'autre est négative ( $V_{c+}$  et  $A_{c-}$ , ou  $V_{c-}$  et  $A_{c+}$ ) ; et
- un indicateur de congruence neutre si l'une des notes est neutre ( $V_{c0}$  ou  $A_{c0}$ ) ;
- dans lequel le procédé est caractérisé en ce qu'il comprend en outre, avant l'étape b), une étape préliminaire b0) d'établissement d'une ligne de base au cours de
- 30 laquelle les sous-étapes suivantes sont mises en œuvre, dans lesquelles :
- i) un film de référence est projeté à ladite personne (130), ledit film de référence comprenant « m » séquences de film de référence, au moins certaines des séquences de film de référence étant chargées en émotions ;

ii) pendant la projection du film, un enregistrement d'une vidéo de référence de la personne (130) est effectué ;

iii) la vidéo de référence est divisée en « m » séquences vidéo de référence, chaque séquence vidéo de référence correspondant à une séquence de film de référence dudit film ;

iv) pour chaque séquence vidéo de référence, est détecté au moins un repère visuel, Vc, d'une micro-expression qui est mémorisée dans une table de ligne de base de ladite personne (130).

10           2.       Procédé selon la revendication 1, dans lequel ledit repère visuel, Vc, est l'un des suivants : toutes les expressions faciales ou tous les repères de langage corporel, y compris un signe visuel d'inconfort, un signe visuel de confort ou un signe de pacification visuel.

15           3.       Procédé selon l'une quelconque des revendications 1 à 2, dans lequel ledit repère audio, Ac, est l'un des suivants : pour la voix : rythme (pause), vitesse (changement de vitesse), volume (élevé ou faible), hauteur tonale, ton (élevé ou faible) ; pour la voix émotionnelle (négative ou positive) ; style verbal : linguistique, interrogation, mot, nombre, changement de style verbal et un sentiment positif ou  
20 négatif exprimé dans la séquence audio, un signe audio d'inconfort, un signe audio de confort et un signe de pacification audio.

              4.       Procédé selon l'une quelconque des revendications 1 à 3, dans lequel est en outre fournie une table de référence avec la ou les correspondances de notes  
25 du repère visuel, Vc, et du repère audio, Ac.

              5.       Procédé selon l'une quelconque des revendications 1 à 4, dans lequel le discours de la personne (130) se déroule devant une autre personne considérée  
comme un intervieweur (132), de sorte que le discours constitue un entretien entre  
30 ladite personne, autrement dit l'interviewé (130), et un intervieweur (132), dans lequel le procédé comprend en outre les étapes ci-dessous consistant à :

f/ fournir un deuxième dispositif d'enregistrement vidéo (127) apte à enregistrer des images dudit intervieweur (132), y compris le visage et au moins certaines parties du corps ;

5 g/ enregistrer également une vidéo du discours de cet intervieweur (132), à l'aide dudit deuxième dispositif d'enregistrement vidéo (127), ladite vidéo étant divisée en « n » séquences vidéo comprenant « n » séquences d'images (ou « n » séquences visuelles) et « n » séquences audio correspondantes ;

10 h/ détecter au moins un repère visuel,  $V_c$ , de l'intervieweur (132) pour chaque séquence d'images et détecter au moins un repère audio,  $A_c$ , de l'intervieweur (132) pour chaque séquence audio ;

i/ pour chaque séquence vidéo, analyser la note du repère audio,  $A_c$ , et du repère visuel,  $V_c$ , de la personne constituant l'interviewé (130) par rapport au repère visuel,  $V_c$ , et au repère audio,  $A_c$ , de l'intervieweur (132), ce qui permet d'établir par conséquent un indicateur d'influence positif ou négatif,

15 moyennant quoi l'indicateur d'influence est positif lorsqu'il existe une influence détectée du repère visuel,  $V_c$ , et du repère audio,  $A_c$ , de l'intervieweur (132) sur la note du repère audio,  $A_c$ , et du repère visuel,  $V_c$ , de la personne constituant l'interviewé (130) ; et

20 où l'indicateur d'influence est négatif lorsqu'il n'existe pas d'influence détectée du repère visuel,  $V_c$ , et du repère audio,  $A_c$ , de l'intervieweur (132) sur la note du repère audio,  $A_c$ , et du repère visuel,  $V_c$ , de la personne constituant l'interviewé (130).

25 6. Procédé selon la revendication 5, dans lequel ledit indicateur d'influence détectée est utilisé pour fournir à l'intervieweur une série de formulations d'hypothèses sous la forme d'affirmations.

30 7. Procédé selon l'une quelconque des revendications 1 à 4, dans lequel le discours de la personne (130) se déroule devant une autre personne considérée comme un intervieweur (132), de sorte que le discours constitue un entretien entre ladite personne, autrement dit l'interviewé, (130), et un intervieweur (132), dans lequel le procédé comprend en outre les étapes ci-dessous consistant à :

f/ fournir un deuxième dispositif d'enregistrement vidéo (127) apte à enregistrer des images dudit intervieweur (132), y compris le visage et au moins certaines parties du corps ;

5 g/ enregistrer également une vidéo du discours de cet intervieweur (132) à l'aide dudit deuxième dispositif d'enregistrement vidéo (127), ladite vidéo étant divisée en « n » séquences vidéo comprenant « n » séquences d'images (ou « n » séquences visuelles) et « n » séquences audio correspondantes ;

10 h/ détecter au moins un repère visuel,  $V_c$ , de l'intervieweur (132) pour chaque séquence d'images, et détecter au moins un repère audio,  $A_c$ , de l'intervieweur (132) pour chaque séquence audio ;

i/ pour chaque séquence vidéo, analyser la note du repère audio,  $A_c$ , et du repère visuel,  $V_c$ , de la personne constituant l'intervieweur (132), par rapport au repère visuel,  $V_c$ , et au repère audio,  $A_c$ , de l'interviewé (130), ce qui permet d'établir par conséquent un indicateur d'influence positif ou négatif ;

15 moyennant quoi l'indicateur d'influence est positif lorsqu'il existe une influence détectée du repère visuel,  $V_c$ , et du repère audio,  $A_c$ , de l'interviewé (130) sur la note du repère audio,  $A_c$ , et du repère visuel,  $V_c$ , de la personne constituant l'intervieweur (132) ; et

20 où l'indicateur d'influence est négatif lorsqu'il n'existe pas d'influence détectée du repère visuel,  $V_c$ , et du repère audio,  $A_c$ , de l'interviewé (130) sur la note du repère audio,  $A_c$ , et du repère visuel,  $V_c$ , de la personne constituant l'intervieweur (132).

8. Procédé d'entraînement d'une machine d'auto-apprentissage, telle  
25 qu'un réseau de neurones, afin de déterminer la congruence ou l'incongruence entre le langage corporel et le discours d'une personne (130), le procédé comprenant les étapes ci-dessous consistant à :

a) fournir une machine d'auto-apprentissage, telle qu'un réseau de neurones, agencée de manière à recevoir, en tant qu'entrée, un ensemble de données  
30 d'entrée incluant :

des données approuvées d'une collection de discours analysés de personnes, lesdites données approuvées comprenant, pour chaque discours analysé :

\* un ensemble de séquences vidéo, comprenant des séquences audio et des séquences visuelles, chaque séquence audio correspondant à une séquence visuelle ; et

5 \* pour chaque séquence vidéo, un indicateur de congruence approuvé pour ladite séquence vidéo, ledit indicateur de congruence approuvé étant obtenu par la mise en œuvre du procédé de fourniture d'indicateurs de congruence ou d'incongruence selon l'une quelconque des revendications 1 à 7,

b) entraîner la machine d'auto-apprentissage avec ledit ensemble de données d'entrée.

10

9. Procédé selon la revendication 8, dans lequel ledit ensemble de données approuvé résulte :

- d'enregistrements vidéo d'un grand nombre de discours de différentes personnes ;

15

- de la division de chaque discours en « n » séquences vidéo comprenant « n » séquences d'images ou « n » séquences visuelles et « n » séquences audio correspondantes ; et

20

- de l'annotation de chaque séquence vidéo par des spécialistes du profilage, c'est-à-dire par des profileurs qui évaluent manuellement les séquences vidéo avec au moins un repère audio,  $A_c$ , et au moins un repère visuel,  $V_c$ , chaque repère se voyant attribuer, par le spécialiste, au moins une note parmi « positive » (+), « neutre » (0) ou « négative » (-).

25

10. Système destiné à fournir des indicateurs de congruence ou d'incongruence entre le langage corporel et le discours d'une personne, le système comprenant :

une machine d'auto-apprentissage configurée de manière à recevoir une entrée comprenant des ensembles de séquences audio du discours d'une personne, et des séquences d'images de cette personne pendant ledit discours ;

30

un dispositif d'enregistrement vidéo configuré de manière à enregistrer des images d'un sujet, y compris le visage et au moins certaines parties du corps ;

dans lequel le dispositif d'enregistrement vidéo est configuré de manière à enregistrer une vidéo d'un discours de cette personne (130), ladite vidéo étant

divisée en « n » séquences vidéo comprenant « n » séquences d'images (ou « n » séquences visuelles) et « n » séquences audio correspondantes ;

un détecteur de repères visuels (108) configuré de manière à, pour chaque séquence d'images, analyser ladite séquence vidéo et détecter au moins un repère visuel,  $V_c$ , dans lequel au moins une note parmi « positive »,  $V_{c+}$ , « neutre »,  $V_{c0}$ , et « négative »,  $V_{c-}$ , est attribuée pour chaque repère visuel,  $V_c$  ;

un détecteur de repères audio (114) configuré de manière à, pour chaque séquence audio, analyser ladite séquence audio et détecter au moins un repère audio,  $A_c$ , dans lequel au moins une note parmi « positive »,  $A_{c+}$ , « neutre »,  $A_{c0}$ , et « négative »,  $A_{c-}$ , est attribuée pour chaque repère audio,  $A_c$ ,

dans lequel le système est configuré de manière à, pour chaque séquence vidéo, comparer la note dudit repère audio,  $A_c$ , à la note dudit repère visuel  $V_c$ , et donner un indicateur de congruence qui est :

un indicateur de congruence positif si les deux notes sont soit positives ( $V_{c+}$  et  $A_{c+}$ ), soit négatives ( $V_{c-}$  et  $A_{c-}$ ) ;

un indicateur de congruence négatif si l'une des notes est positive et l'autre est négative ( $V_{c+}$  et  $A_{c-}$ , ou  $V_{c-}$  et  $A_{c+}$ ) ; et

un indicateur de congruence neutre si l'une des notes est neutre ( $V_{c0}$  ou  $A_{c0}$ ) ;

caractérisé en ce que le système est configuré de manière à établir une ligne de base, comprenant les étapes ci-dessous consistant à :

i) projeter un film de référence à ladite personne (130), ledit film de référence comprenant « m » séquences de film de référence, au moins certaines des séquences de film de référence étant chargées en émotions ;

ii) pendant la projection du film, enregistrer une vidéo de référence de la personne (130) ;

iii) diviser la vidéo de référence en « m » séquences vidéo de référence, chaque séquence vidéo de référence correspondant à une séquence de film de référence dudit film ;

iv) pour chaque séquence vidéo de référence, détecter au moins un repère visuel,  $V_c$ , d'une micro-expression qui est mémorisée dans une table de ligne de base de ladite personne (130).