Caractérisation d'unités gestuelles en vue d'une interaction humain-avatar

Ilaria Renna, Sébastien Delacroix, Fanny Catteau, Coralie Vincent, Dominique Boutet

Caractérisation d'unités gestuelles en vue d'une interaction humain-avatar

I. Renna[1] S. Delacroix[2] F. Catteau[1] C. Vincent[1] D. Boutet[1]

[1]Structures Formelles du Langage, UMR 7023 (CNRS / Université Paris 8)
[2]Laboratoire d'Analyse du Mouvement, Institut National de Podologie, Paris

ilaria.renna@gmail.com

Domaine principale de recherche: RFP
Papier soumis dans le cadre de la journée commune: NON

Résumé

Nous présentons ici une méthode de caractérisation d'unités gestuelles coverbales, en vue d'une exploitation dans une interaction humain-avatar. Nous avons enregistré 12 types de gestes avec un système de capture de mouvement. Nous avons utilisé les signaux de position obtenus afin d'en dégager des unités gestuelles à l'issue d'une segmentation de la partie significative. Pour soutenir notre analyse linguistique des gestes, nous présentons les hypothèses biomécaniques, notre méthode de segmentation, les hypothèses de caractérisation et les résultats obtenus.

Mots Clef

Unités gestuelles, segmentation de la partie significative d'un geste, caractérisation de geste.

Abstract

In this paper we present a method to characterize co-verbal gestures unities to be exploited in a human-avatar interaction. We recorded 12 different kinds of gesture with a motion capture system and exploited the obtained position signals to find gesture unities after a stroke segmentation. To prove a linguistic gestures analysis, we present the biomechanical assumptions, our segmentation method and its results as well as the characterization assumptions and their results.

Keywords

Gesture unities, stroke segmentation, gesture characterization.

1 Introduction

La caractérisation du sens des gestes est faite classiquement en fonction de descriptions égocentrées [17] appréhendant les éléments gestuels selon une description globale dans un repère du corps.

Nous voulons montrer que l'on peut caractériser le sens de différentes Unités Gestuelles (UG) sémantiquement proches à partir de formes distribuées sur le membre supérieur dans des repères multiples non égocentrés, centrés sur chacun des segments (main, avant-bras, bras), ce qui facilite une caractérisation automatique des gestes enregistrés en capture de mouvement (section 2).

Cette caractérisation servira de base pour alimenter un algorithme génétique qui anime un agent virtuel : les caractéristiques de plusieurs bases gestuelles seront hybridées pour donner lieu à des nouveaux gestes grâce auxquels un agent virtuel sera amené à interagir avec un acteur réel lors de performances théâtrales.

La segmentation de la partie significative des gestes (*stroke* [15], section 4) est un préalable nécessaire à la caractérisation : on ne peut pas caractériser le sens d'un geste sans savoir à quel moment il intervient. Pour cela, une segmentation automatique est présentée et testée par rapport aux vérités terrain constituées par la segmentation réalisée par deux annotateurs (section 4). Cette opération effectuée et validée, la caractérisation automatique repose sur un centrage par rapport à la variation du mouvement d'un degré de liberté (ddl) — la pronosupination (section 5). Naturellement, la caractérisation sémantique des gestes repose sur un modèle linguistique dont les grandes lignes sont exposées dans la section 3. Ce modèle, basé sur des constantes forme/sens, se décline en plusieurs niveaux dont chacun apporte une information sur la structuration du sens [2].

2 Présentation de la base de données

La base de données étudiée est composée de 91 gestes symboliques coverbaux et isolés réalisés selon un étiquetage sémantique contrôlé [1 et 2]. C'est l'une des 4 bases de données du projet CIGALE dont l'objectif est de créer une interaction avatar-humain.

Les gestes coverbaux présents couvrent l'ensemble des ddl du membre supérieur et sont sémantiquement autonomes (voir 3.1). Certains gestes sont réalisés sur l'ensemble du membre supérieur alors que d'autres peuvent n'être exécutés que sur les doigts, par exemple.

2.1 Capteurs et modèle biomécanique

Les gestes d'un acteur sont recueillis à l'aide d'un système de capture de mouvement 3D composé de 24 caméras numériques infrarouges (VICON, 120 fps), qui assure la fiabilité de la compréhension du geste et l'efficacité de la caractérisation des mouvements de l'avatar. Pour modéliser les segments corporels en trois dimensions, une liste des marqueurs cutanés est établie (*marker-set* de 90 points, Fig. 1).

Celle-ci référence les positions anatomiques à utiliser pour modéliser chaque segment comme un solide indéformable. Généralement, trois repères anatomiques non alignés suffisent à définir un segment. Dans notre modèle (Fig. 2), les segments tronc, bras, avant-bras et main ont été définis à partir des coordonnées spatiales

des mires, selon une méthode standardisée (détail dans la légende). Cette dernière permet la création des trois axes orthogonaux pour chaque système de coordonnées segmentaires ([23], [8]). Pour cela, les centres articulaires du poignet, du coude, de l'épaule ainsi que ceux de la région cervicale et lombaire sont calculés [8].

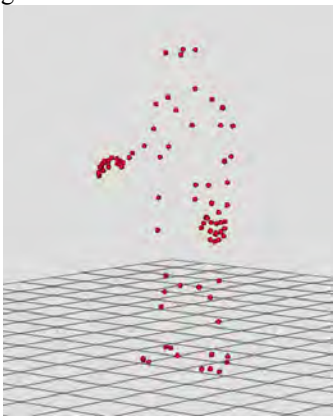

Figure 1. Visualisation du *marker-set*.

Afin de décrire les mouvements articulaires tridimensionnels de l'épaule, du coude et du poignet à chaque instant du geste, les systèmes de coordonnées de chaque articulation sont définis à partir des systèmes de cordonnées segmentaires adjacents à l'articulation. Pour cela, une séquence de rotations successives autour d'axes mobiles permettant d'obtenir la cinématique articulaire grâce aux angles d'Euler est utilisée [23]. La séquence mobile de rotation permet de définir le système de coordonnées articulaires en utilisant les axes de deux segments adjacents : un axe du segment proximal, un axe du segment distal et un axe flottant perpendiculaire aux deux précédents.

Grâce à ce modèle biomécanique, les différents mouvements articulaires du poignet, du coude et de l'épaule sont calculés. Ainsi, les mouvements de flexion palmaire/dorsale et d'adduction/abduction du poignet sont calculés. Ceux-ci correspondent aux mouvements de flexion/extension et d'adduction/abduction de la main tel que décrit dans les schémas d'actions (section 3). Les mouvements d'extension/flexion et de supination/pronation du coude correspondent respectivement aux mouvements d'extension/flexion de l'avant-bras et de supination/pronation de la main pour les schémas d'actions (voir 3.2). Enfin, les mouvements de rétropulsion/antépulsion, d'abduction/adduction et de rotation externe/interne de l'épaule sont mesurés. Ceux-ci correspondent respectivement à une extension/flexion, abduction/adduction et rotation extérieure/intérieure du bras pour les schémas d'actions.

3 Recadrage linguistique en vue d'une interaction humain-avatar

Les gestes coverbaux enregistrés correspondent à des emblèmes ou *quote gestures* ([19] et [13]), c'est-à-dire des gestes sémantiquement autonomes, à la signification indépendante du discours verbal associé.

Les 91 gestes se répartissent en une douzaine d'UG. Les significations sont les suivantes : rejeter, refuser, mépriser, déconsidérer, passer, accepter, considérer quelque chose, considérer quelqu'un, offrir, s'en fiche, s'engager, révérer. Ces étiquettes sémantiques ont été testées et validées auprès d'une population francophone dans un travail antérieur [2].

Chacune de ces UG répond à un schéma d'action singulier qui met en œuvre une partie ou l'ensemble des segments du membre supérieur.

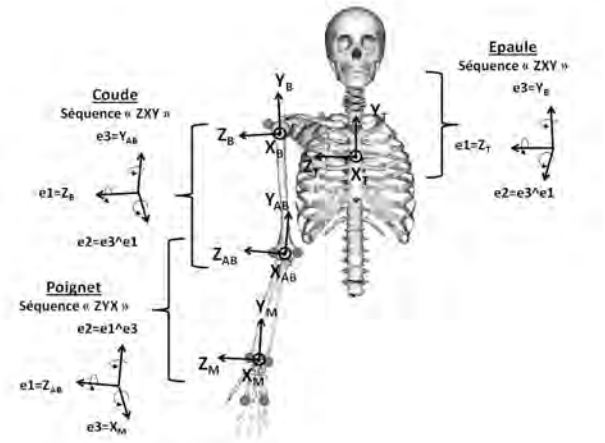

Figure 2. Modèle biomécanique. Pour la main, l'origine du système de cordonnées est situé au centre articulaire du poignet. L'axe Y est le vecteur unitaire reliant le centre des 2e et 5e têtes métacarpiennes à l'origine. L'axe X est le vecteur unitaire normal au plan contenant l'origine et les 2e et 5e têtes métacarpiennes. L'axe Z est le produit vectoriel des axes X et Y. Pour l'avant-bras, l'origine du système de cordonnées est situé au centre articulaire du coude. L'axe Y est le vecteur unitaire reliant le centre articulaire du poignet à l'origine. L'axe X est le vecteur unitaire normal au plan contenant l'origine et les processus styloïdes de l'ulna et du radius. L'axe Z est le produit vectoriel des axes X et Y. Pour le bras, l'origine du système de cordonnées est situé au centre articulaire de l'épaule. L'axe Y est le vecteur unitaire reliant le centre articulaire du coude à l'origine. L'axe X est le vecteur unitaire normal au plan contenant l'origine, l'épicondyle et l'épitrochlée. L'axe Z est le produit vectoriel des axes X et Y. Pour le tronc, l'origine du système de cordonnées est situé au centre articulaire cervical. L'axe Y est le vecteur unitaire reliant le centre articulaire lombaire à l'origine. L'axe Z est le vecteur unitaire normal au plan contenant l'origine, centre articulaire lombaire et l'espace supra sternal. L'axe X est le produit vectoriel des axes Y et Z.

La description sous forme de schémas d'action repose sur la mise en mouvement de différents ddl des segments du membre supérieur dans un ordre précis. Le mouvement est transféré en fonction des moments d'inertie qui sont attachés à chaque ddl et en fonction du mouvement conjoint (involontaire) de l'axe longitudinal (Rotation extérieure/intérieure ou Pronation/supination) associé à toute articulation à deux ddl ([5], [16] et [4]). Ainsi, pour l'UG « refuser » par exemple (Schéma 1), le schéma d'action déploie le geste de la main vers l'avant-bras.

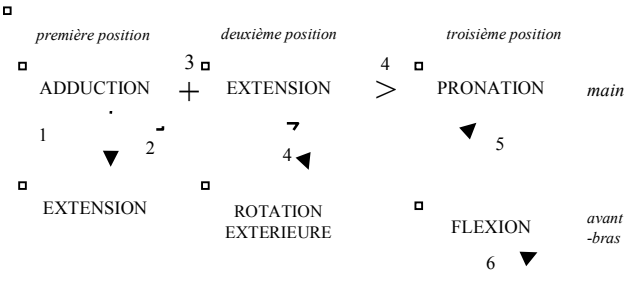

Schéma 1. Schéma d'action de l'UG « refuser ».

3.1 Flux de propagation du mouvement

Dans le schéma d'action, la position du pôle de l'adduction (mouvement vers l'auriculaire dans le plan de la paume) donne le sens de propagation du mouvement. Si l'adduction figure en première ou en deuxième position, le flux de propagation du mouvement est distal-proximal ; il va globalement de la main vers l'avant-bras. Au contraire, si l'adduction est en troisième position, alors elle est un mouvement conséquent des deux premiers et, à ce titre, ne présente pas une amplitude importante. Le geste est initié sur l'avant-bras et se propage alors vers la main selon un flux distal-proximal. On définit ainsi deux types d'UG. Les 8 premières UG de la liste ci-dessus sont structurées sur la main, tandis que les 4 dernières (offrir, s'en fiche, s'engager et révérer) le sont sur le bras. Lors d'une étude précédente, des UG identiques présentées en vidéo ont montré un taux de reconnaissance significatif selon une méthode des juges [2].

3.2 Schéma de l'enchaînement des mouvements sur la main

L'enchaînement des mouvements sur la main suit une structuration telle que le mouvement ou la position des deux premiers ddl entraînent le mouvement involontaire du troisième ddl. Ce troisième mouvement est dû soit à une contrainte biomécanique liée à un mouvement autour de l'axe longitudinal (pronation/supination), soit à un enchaînement entraîné par le moment d'inertie. Dans les deux cas, les pôles du mouvement en troisième position sont parfaitement déterminables et répondent à un ordre d'enchaînement des deux mouvements précédents tel que leur ordre impacte le pôle du troisième mouvement. Ainsi, la suite ADD.EXTEN entraîne un mouvement involontaire de PRONATION, tandis que l'ordre opposé, EXTEN.ADD implique un mouvement de SUPINATION ([1] et [2]).

3.3 Regroupement des UG par famille de sens

Le repérage de l'ordre des pôles mis en mouvement relève tout autant de l'amplitude du mouvement, de la succession temporelle de l'apparition du mouvement, de la position initiale et de l'accélération sans qu'il soit aisé d'en hiérarchiser les critères qui varient même de manière intra-individuelle. En revanche, il est possible de regrouper les UG par champ sémantique sur une base formelle (Schéma 2).

Dans un premier temps, il s'agit de déterminer le flux de propagation du mouvement : soit le geste part de la main et le mouvement remonte sur l'avant-bras, soit il part du bras et diffuse vers la main (Main et Bras dans le schéma). Pour la branche de la main (Schéma 2, à gauche), la position de la pronosupination initiale au geste est soit marquée, soit non marquée. Au niveau suivant, on examine le mouvement de la pronosupination par rapport à la position initiale. On obtient ainsi 8 schémas d'action manuels. Pour la branche du bras (Schéma 2, à droite), on examine la position ou le mouvement de l'ADD/ABD du bras. Au niveau suivant, le mouvement de la pronosupination permet de distinguer les 4 UG organisées sur le bras.

Chacune de ces UG a un label sémantique. Un premier niveau de regroupement hyperonymique compose 4 ensembles sémantiques : i/ Positionnement par rapport aux choses, ii/ Considération ou jugement, iii/ Implication et iv/ Intérêt. Ce niveau sémantique correspond au 2e niveau de disjonction formelle dans le schéma. On peut également procéder à un regroupement sémantique en deux parts correspondant à la première disjonction formelle (Main ou Bras) : Positionnement par rapport au monde *versus* Positionnement par rapport à la relation.

Ainsi, à différents niveaux de différenciation formelle correspond un étiquetage sémantique spécifique.

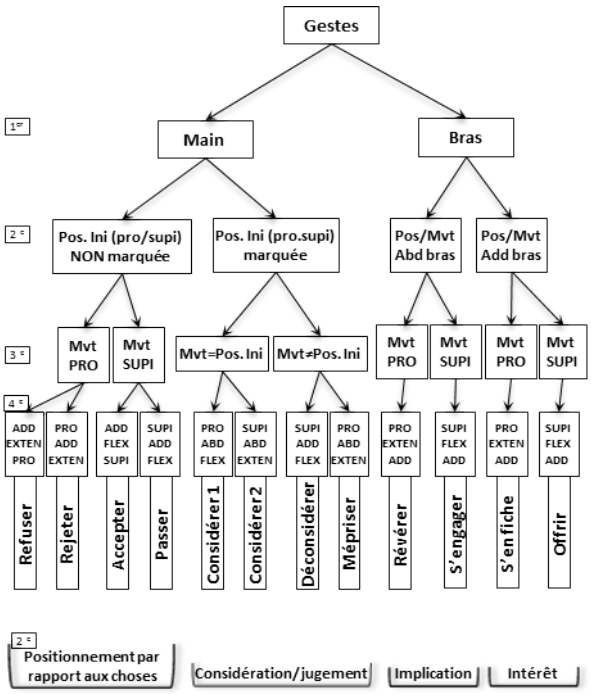

Schéma 2. Présentation formelle des gestes en vue de leur caractérisation sémantique.

4 Segmentation des signaux gestuels

Dans notre base de données, chaque signal gestuel est constitué d'un enchaînement pose en T, geste, pose en T. Une segmentation automatique est nécessaire afin d'extraire le geste.

La séquence généralement admise est celle correspondant à une suite de 4 phases ([17] et [6]) :

1- Position de repos
2- La préparation (pré-*stroke*)
3- Le cœur (*stroke*)
4- La rétraction (post-*stroke*)

Cette séquence décrit la structure du geste.

La difficulté réside dans l'impossibilité de trouver un critère objectif automatique pour extraire le *stroke,* la partie sémantiquement significative. Cette opération est complexe même pour un humain et reste incertaine [21].

Dans notre cas, nous effectuons la segmentation sur la base de propriétés morpho-cinématiques (*morpho-kinetics* selon Kendon [14]). En effet, la préparation du mouvement consiste en un mouvement balistique qui amène le(s) bras vers le cœur du mouvement [3]. Cette balistique consiste en une accélération puis une décélération à l'approche de la pose finale, puis une

accélération et une décélération symétriques aux premières pour revenir à la position de repos.
Les poses en T sont également caractérisées par des mouvements d'accélération et de décélération.
En conséquence on a décidé d'extraire le *stroke* de chaque geste en considérant la valeur absolue de la dérivée en Y des positions de l'index (considéré dans tous les cas comme le membre du corps qui bouge le plus) : les minima de ce signal représentent le passage entre décélération et accélération. Pour la segmentation automatique on considère donc que le *stroke* est la partie comprise entre la phase minimale qui précède le deuxième maximum (caractéristique du début du *stroke*) et la phase minimale qui suit l'avant-dernier maximum (fin du *stroke*) (Fig. 3). Pour éviter de prendre en compte des phase maximales et minimales dues au bruit (petits mouvements d'ajustement ou de préparation) un seuil est fixé à 0.5 car nous considérons qu'un mouvement sémantiquement valide présente un pic supérieur à cette valeur.

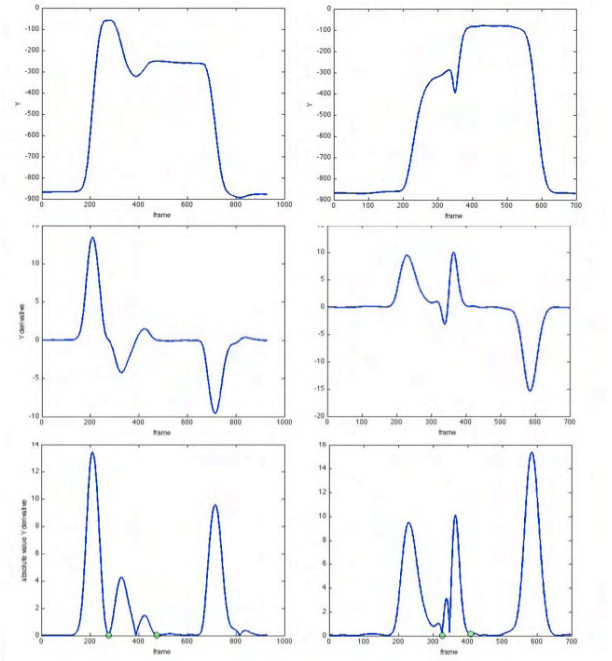

Figure 3. A gauche : signaux du geste « révérer » ; à droite : signaux d' « accepter ». De haut en bas : positions de l'index en Y, vitesse (dérivée) et valeur absolue de la vitesse avec individuation automatique du *stroke* comprise entre deux points verts.

4.1 Evaluation de la segmentation

Pour évaluer les méthodes de segmentation automatique, il est nécessaire de comparer les performances de la segmentation réalisée automatiquement avec celle de juges humains (codeurs). En général, les méthodes pour effectuer cette comparaison utilisent la segmentation d'un seul codeur comme référence [11]. Une telle référence ne devrait pas être considérée comme unique vérité terrain, étant donné que l'accord inter-annotateurs est souvent assez faible [12]. De plus, pour s'assurer qu'une segmentation automatique n'est pas biaisée par rapport aux choix d'un codeur, elle devrait être comparée directement au travail de plusieurs codeurs [10]. Pour évaluer, d'une part, l'accord inter-annotateurs, et d'autre part, la segmentation automatique, nous avons décidé d'adopter deux méthodes : l'*Accurate Temporal Segmentation Rate* (ATSR) (taux de segmentation temporelle précis) [20] et le F-score [22].
L'ATSR est une mesure basée sur le temps, qui permet d'évaluer la performance en termes de précision de détection des début et fin de *stroke* pour chaque geste, alors que le F-score donne plus d'informations sur la précision et permet l'individuation de la typologie d'erreur.
Trois comparaisons sont effectuées avec ces deux méthodes :

1. la segmentation automatique est comparée au premier annotateur, considéré comme vérité terrain (cas 1) ;
2. la segmentation automatique est comparée au second annotateur considéré comme vérité terrain (cas 2) ;
3. les deux annotateurs sont comparés (cas 3).

Pour chaque geste considéré, l'ATSR a été calculé de la manière suivante : l'*Absolute Temporal Segmentation Error* (ATSE) (erreur de segmentation temporelle absolue) est évaluée en additionnant l'erreur temporelle absolue entre la vérité terrain et le résultat de l'algorithme pour les événements de début et de fin, le tout divisé par la durée totale de l'occurrence du *stroke* mesurée à partir de la vérité terrain, comme formalisé dans l'équation 1.
Une fois les ATSE obtenues, les mesures d'ATSR sont calculées en soustrayant l'ATSE moyenne à 1, de manière à obtenir le taux de précision comme montré dans l'équation 2. Une segmentation parfaitement précise produit un ATSR de 1.

$$ATSE = \frac{\left|\text{Start}_{GT} - \text{Start}_{Alg}\right| + \left|\text{Stop}_{GT} - \text{Stop}_{Alg}\right|}{\text{Stop}_{GT} - \text{Start}_{GT}} \qquad (1)$$

$$ATSR = 1 - \frac{1}{n}\sum_{i=1}^{n} ATSE(i) \qquad (2)$$

L'équation 1 compte les différences qui se produisent image par image, donc une erreur est prise en compte même quand les annotations diffèrent de seulement quelques images.
Pour limiter cet effet, il est possible de fixer un seuil de tolérance α de manière à ce que

si ATSE(i) < α, alors ATSE(i) = 0. (3)

Comme, en général, un *stroke* dure environ 100 images, nous fixons α =0.2. Ceci correspond à une différence globale de α *100=20 images (ce qui signifie environ 0.17s à la fréquence d'acquisition de 120i/s) ce qui est un choix adéquat comparé à la durée de la vérité terrain, considérant que, en moyenne, il est facile d'avoir 10 images de décalage pour chaque début et fin.
Nous obtenons : pour le cas 1, ATSR=0.6038 ; pour le cas 2, ATSR=0.5857 ; pour le cas 3, ATSR=0.8707.
Ce genre de méthode manque néanmoins d'exhaustivité car il ne fournit pas le type d'erreur.
Nous pouvons, en fait, avoir 5 types d'erreur (Fig. 4).
De fait, il est important de savoir si la segmentation automatique est erronée, mais préserve le *stroke* (Fig. 4, erreur 2) ou si elle le coupe (tous les autres cas).

De manière à évaluer la qualité de notre segmentation et l'accord inter-annotateurs, considérons la précision (*p*) et le rappel (*r*) ([9] et [18]) : la précision est la fraction de détections qui sont des vrais positifs plutôt que des faux positifs (équation 4), alors que le rappel est la fraction de vrais positifs qui sont détectés plutôt que manqués (équation 5). En termes probabilistes, la précision est la probabilité que la détection soit valide, et le rappel est la probabilité que les données de vérité terrain soient détectées.

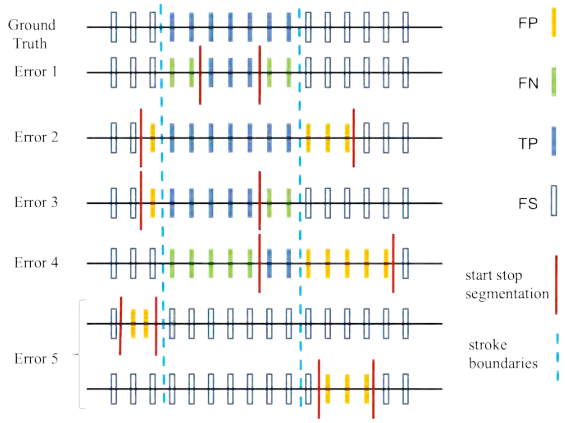

Figure 4. Différentes erreurs de segmentation possibles. FP = Faux Positif, FN = Faux Négatif, TP = Vrai Positif et FS = Hors Segmentation.

$$p = \frac{TP}{TP + FP} \quad (4) \qquad r = \frac{TP}{TP + FN} \quad (5)$$

La précision et le rappel peuvent être combinés dans le F-score de cette manière : $F_\beta = (1+\beta^2) * \frac{p * r}{\beta^2 p + r}$ (6)

Quand le paramètre β=1, le F-score est dit équilibré et s'écrit F_1 : $F_1 = 2 * \frac{p * r}{p + r}$ (7)

Le score F_1 peut être vu comme une moyenne pondérée de la précision et du rappel. F_1 est compris entre 0 (moins bonne valeur) et 1 (meilleure valeur).

Les résultats obtenus sont résumés dans le tableau 1.

	p	*r*	F_1
Cas 1	0.7430	0.9216	0.8227
Cas 2	0.7358	0.9151	0.8157
Cas 3	0.9077	0.9053	0.9065

Tableau 1. Résultats obtenus pour les trois cas d'étude.

En général, des valeurs élevées de F_1 sont obtenues ; *r* est plus élevé que *p* dans la comparaison avec la segmentation automatique, ce qui signifie que l'algorithme renvoie la plupart des résultats pertinents, alors que *p* est plus élevé pour l'accord inter-annotateurs : ce qui est obtenu le plus, ce sont les accords pertinents.

Les résultats concernant les types d'erreur sont présentés dans le tableau 2.

	Cas 1	Cas 2	Cas 3
Erreur 1	4	5	17
Erreur 2	53	49	15
Erreur 3	24	16	54
Erreur 4	10	21	5
Erreur 5	0	0	0

Tableau 2. Erreurs mesurées dans les cas d'étude.

Il est important de souligner que dans les cas 1 et 2, l'erreur 2 se produit le plus : ceci signifie que la méthode de segmentation préserve le *stroke*. De plus, l'erreur la plus basse est la coupe du *stroke* : nous pouvons affirmer que la méthode de segmentation utilisée est robuste pour analyser le type des gestes présentés car elle permet d'avoir des F-score élevés avec une coupure du *stroke* très basse.

Pour l'accord inter-annotateurs, nous soulignons que, quand ils se trompent, c'est principalement parce que l'un d'eux coupe le *stroke* (erreur 1) ou parce que l'un anticipe l'autre (erreur 3).

Cette méthode de segmentation pour simple et robuste qu'elle soit demande néanmoins à être étendue et évaluée pour des situations différentes.

5 Caractérisation des composantes des schémas d'action

Afin de caractériser les schémas d'action pour chacun des gestes coverbaux enregistrés, les signaux segmentés sont transformés en données cinématiques selon la modélisation biomécanique (section 2.1). Les mouvements des différents ddl de chaque articulation du membre supérieur droit (épaule, coude et poignet) sont pris en compte et normalisées temporellement sur 101 points [7].

Pour la caractérisation, on part de la constatation que dans une communication humain-humain, pour n'importe quel type de geste (réalisé sur l'ensemble du membre supérieur ou juste sur un de ses segments), les mouvements des ddl de pronosupination sont les plus visibles. Il est donc décidé, dans un premier temps, de focaliser l'analyse sur la partie des signaux alignée temporellement avec la zone de pronosupination contenant la plus grande variation. Les paramètres biomécaniques tels que les positions initiales et finales de chaque ddl ainsi que leur amplitude maximale sont pris en considération (Fig.5).

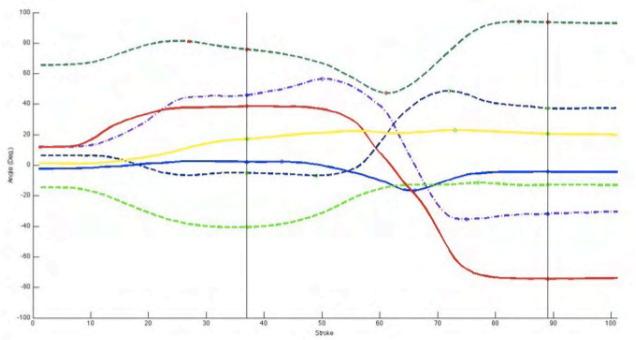

Figure 5. Signaux des différents segments du membre supérieur. Pour la main : en rouge la supination/pronation, en bleu ciel l'abduction/adduction et en violet pointillé la flexion/extension. Pour l'avant-bras en vert foncé pointillé l'extension/flexion. Pour le bras : en bleu foncé pointillé la rotation interne/externe, en vert pointillé l'abduction/adduction et en jaune l'extension/flexion. Les lignes verticales indiquent la plus grande variation de pronosupination.

La caractérisation a été effectuée sur les 33 gestes (des 91 captés) mobilisant des mouvements de tous les segments du membre supérieur. Pour cela, les étapes considérées de l'arbre de décision présenté au schéma 2 vont i/du premier nœud (détermination du flux manuel

ou brachial) ii/ au quatrième (séparation des gestes par la pronosupination).
Pour le premier nœud, il s'agit de déterminer si la propagation du mouvement part du bras (flux proximal-distal) ou de la main (distal-proximal). Pour cela, on a calculé : 1/ l'instant où pour la position initiale apparaissent le min. et le max. de chaque ddl à l'intérieur du *stroke* découpé automatiquement ; 2/ la différence temporelle entre le min. et le max. des ddl d'un segment à l'autre (bras [offrir, s'en fiche, s'engager et révérer], avant-bras et main [pour tous]). Dans ce dernier calcul, le choix de la valeur min. ou max. pour tel ou tel ddl correspond à la position initiale et donc, *a priori*, à l'opposé du pôle en mouvement repéré au cours du *stroke*. Si le mouvement de la main est une EXTEN (valeur positive), alors la position initiale correspond à un minimum (flexion, valeur négative). Ainsi par exemple, pour la ligne supérieure du schéma 1 qui illustre les pôles en mouvement du geste « refuser » : ADD.EXTEN>PRO, on choisit comme positions initiales la valeur max. de l'ADD/ABD, la valeur min. de la FLEX/EXTEN et la valeur min. de la SUPI/PRO.
On a fixé un seuil minimal de 10 images, correspondant à un volant de 2 images vidéo à 25i/s, pour la différence temporelle permettant de connaître le flux.
Sur 33 gestes testés qui recouvrent les 12 UG présentées dans la section 3 (chaque UG a été réalisée entre 2 et 3 fois), la détermination du flux par cette méthode valide 87,88% de ce qui était attendu. Parmi les 4 cas non validés, 3 sont en dessous du seuil des 10 images et ne répondent donc à aucun flux déterminable et un seul cas (une réalisation de « révérer ») montre un flux inverse de ce qui était attendu.
A l'autre bout de l'arbre de décision (schéma 2), la quatrième étape de la caractérisation — celle des pôles en mouvement pour déterminer le schéma d'action — a été faite selon une méthode avec deux types de données (voir a/ et b/ ci-dessous).
Dans un premier temps, les calculs concernent la moyenne des deux ou trois réalisations par UG (33 gestes en tout) et portent donc sur 12 UG moyennées dont on connaît *a priori* les pôles en mouvement ainsi que les étiquettes sémantiques. Dans un second temps, on détermine l'amplitude maximale pour chaque ddl, a/ soit à l'intérieur du bornage de la pronosupination tel qu'il est présenté dans la figure 5 ; b/ soit plus largement, à partir du *stroke*, en calculant la différence entre la position finale et la position initiale de chaque ddl. On obtient ainsi les pôles en mouvement pour l'ensemble des ddl qui caractérisent l'UG, c'est-à-dire 60 ddl pour la somme des 12 UG.
Les résultats avec le premier type de données (a/ dans le bornage de pronosupination) donnent un taux de reconnaissance de 76,67%. L'autre option (b/ dans le *stroke* avec la différence de position finale et initiale) voit un taux de caractérisation bien meilleur : 90%. Sur les 60 ddl, seuls 6 pôles attendus voient leur pôle opposé apparaître. Dans les deux options, sur une moyenne de 6 ddl mesurées par UG, le pôle le plus sujet à erreur est l'ABD/ADD de la main (a/ 36% des erreurs, b/ 67%) ; il s'agit du pôle présentant la plus petite amplitude (25° et 35°).
Les étapes intermédiaires — 2 et 3 — de caractérisation (voir schéma 2) consistent à déterminer :
- pour l'étape 2, le marquage des positions initiales de la pronosupination et le mouvement d'ABD *vs* ADD du bras ;
- pour l'étape 3, la position initiale et le mouvement de la pronosupination identique *vs* opposé et le pôle du mouvement entre PRO et SUPI.
La caractérisation du mouvement ABD *vs* ADD du bras et PRO *vs* SUPI est effectuée sans aucun problème. En revanche le marquage des positions initiales de la pronosupination (étape 2) ne donne pas les résultats escomptés. Seule la différence d'amplitude de la rotation intérieure/extérieure entre le début et la fin du *stroke* est significative dans ce cas. Pour un intervalle de confiance à 95%, il n'y a pas de zones de recouvrement entre "rejeter/refuser", d'une part, et "mépriser" d'autre part. Pour le trio "passer/accepter/déconsidérer", ce non recouvrement est également vérifié. Ainsi, il convient de modifier le critère de marquage ddl (PRO/SUPI) de l'étape 2 en un différentiel d'amplitude de rotation extérieure/intérieure plus ou moins marqué.
Pour l'étape 3, l'identité ou l'opposition entre la position initiale et le mouvement de pronosupination est un bon critère puisque pour un intervalle de confiance à 95%, il n'y a pas de zones de recouvrement entre "rejeter/refuser/mépriser", d'un côté, et "considérer quelque chose", de l'autre. Il en va de même entre "passer/accepter/déconsidérer", d'une part, et "considérer quelqu'un", d'autre part.
En résumé, les seules étapes qui ne donnent pas entière satisfaction sont donc la 1re (un seul cas d'inversion pour « révérer ») et la 4e (90% des pôles attendus). Les étapes intermédiaires sont fiables à 100%.

6 Conclusion

Dans ce travail, nous avons présenté une méthode de caractérisation de 12 unités gestuelles sémantiquement proches à partir de formes distribuées sur le membre supérieur. Pour cela une capture du mouvement a été effectuée ainsi qu'une méthode de segmentation automatique. Les tests faits à partir du protocole de segmentation montrent sa robustesse dans l'individuation du *stroke* nécessaire pour la caractérisation gestuelle.
Les méthodes simples de caractérisation répondent pour l'instant à l'exigence d'associer, à chaque étape, un étiquetage sémantique à la caractérisation formelle. C'est le cas, puisque les UG qui partagent les mêmes pôles se différencient uniquement par l'ordre d'apparition dans le schéma d'action. Or, nous n'avons pas encore fait cette caractérisation pour 4 d'entre eux. Pour cette étude, nous ne pouvons discriminer, d'une part, « refuser » de « rejeter » et, d'autre part, « accepter » de « passer ». Par contre ces deux groupes sont étiquetables : d'un côté, un *positionnement négatif par rapport aux choses* et de l'autre, le même type de positionnement *positif* cette fois. Ainsi, tous les gestes ont un étiquetage sémantique associé avec une granularité variable. Il reste à éprouver ces méthodes de caractérisation pour des réalisations ne mettant en mouvement qu'un seul segment comme la main.

Bibliographie

[1] Boutet, D., Une morphologie de la gestualité : structuration articulaire. *Cahiers de linguistique analogique,* n°5, Abell, pp. 80–115, décembre 2008.

[2] Boutet, D., Structuration physiologique de la gestuelle : modèle et tests. *Lidil* 42, 77–96, 2010.

[3] Chellali R., Renna I., Bernier E., Détection et Reconnaissance des Gestes Emblématiques, *Interaction Homme-Machine pour l'Apprentissage Humain -IHMA –RFIA*, 2012.

[4] Cheng, P. L., Simulation of Codman's paradox reveals a general law of motion, *Journal of Biomechanics*, 39(7), 1201–1207, 2006.

[5] Codman, E. A. *The shoulder: rupture of the supraspinatus tendon and other lesions in or about the subacromial bursa*. RE Kreiger, 1934.

[6] Corradini, A. et Cohen, P. R, Speech-gesture Interface for Handfree Painting on a Virtual Paper suing Partial Neural Networks as Gesture Recognizer, *Proceedings IJCNN'02, HI*, 2293–2298, 2002.

[7] Desroches, G., Dumas, R., Pradon, D., Vaslin, P., Lepoutre, F.-X., Chèze, L., Upper limb joint dynamics during manual wheelchair propulsion. *Clinical Biomechanics* 25, pp. 299–306, 2010.

[8] Dumas, R., Chèze, L., Verriest, J.-P., Adjustments to McConville et al. and Young et al. body segment inertial parameters. *Journal of Biomechanics* 40(7), 543–553, 2007.

[9] Fawcett. T., An introduction to ROC analysis. *Pattern Recogn. Lett.* 27(8), pp. 861–874, 2006.

[10] Fournier, C., Inkpen, D., Segmentation Similarity and Agreement, *NAACL HLT '12 Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2012.

[11] Gonzalez Preciado M., Computer Vision Methods for Unconstrained Gesture Recognition in the Context of Sign Language Annotation, PhD thesis, Toulouse, 2012.

[12] Marti A. Hearst. 1997. TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages. *Computational Linguistics* 23(1), 33–64. MIT Press, Cambridge, MA, USA.

[13] Kendon, A., How Gestures Can Become like Words. In *Cross-Cultural Perspective in Nonverbal Communication*, 131–141. C.J. Hogrefe, Toronto & Lewiston, N.Y.: Fernando Poyatos, 1988.

[14] Kendon, A. An agenda for gesture studies. *Semiotic Review of Books* 7(3), 8–12, 1996.

[15] Kita, Sotaro, Ingeborg van Gijn, et Harry van der Hulst. « Gesture and Sign Language in Human-Computer Interaction ».. *Lecture Notes in Computer Science*. 1371:23–35, Springer, Berlin / Heidelberg, 1998.

[16] MacConaill, M. A. « The Movements of Bones and Joints ». *Journal of Bone & Joint Surgery, British Volume* 30-B(2), 322–326, 5 janvier 1948.

[17] McNeill, D., *Hand and Mind : What Gestures Reveal about Thought*. University of Chicago Press, Chicago & London, 1992.

[18] Olson, D. L. et Delen, D.. *Advanced Data Mining Techniques* (1st ed.). Springer, 2008.

[19] Payrató, L., A pragmatic view on autonomous gestures: A first repertoire of Catalan emblems, *Journal of Pragmatics* 20(3), 193–216, 1993.

[20] Ruffieux, S., Lalanne, D., Mugellini, E., ChAirGest: a challenge for multimodal mid-air gesture recognition for close HCI. *ICMI,* 483–488, 2013.

[21] M. Sigalas, H. Baltzakis, P. E. Trahanias. Gesture recognition based on arm tracking for human-robot interaction. *IROS*, 5424–5429, 2010.

[22] Van Rijsbergen, C. J., Information Retrieval (2nd ed.). Butterworth, 1979.

[23] Wu, G., van der Helm, F. C., Veeger, H. E., Makhsous, M., Van Roy, P., Anglin, C., Nagels, J., Karduna, A. R., McQuade, K., Wang, X., Werner, F. W., Buchholz, B., ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion--Part II: shoulder, elbow, wrist and hand. *Journal of Biomechanics* 38(5), 981–992, 2005.