

Avatar signeurs : que peuvent-ils nous apprendre ?

Caroline Larboulette, Sylvie Gibet

Université Bretagne Sud, IRISA, UMR 6074, F-56000, Vannes, France

Contexte. Le projet *Avatar Signeurs* de l'équipe EXPRESSION de l'IRISA a pour but de combiner la technologie d'un avatar signeur avec les mécanismes linguistiques des langues des signes afin de réduire le fossé de communication qui existe entre les sourds ou mal-entendants et les entendants, et ceci au moyen d'un système de traduction automatique du français vers la LSF et l'animation expressive de l'avatar (<http://lsf.irisa.fr/>).

Problème. Quels peuvent être le rôle et l'utilité d'un avatar signeur ? Quel type d'interaction observe-t-on en fonction du contexte : réalité virtuelle, réalité augmentée ou visualisation traditionnelle sur un écran ?

Méthode. Pour animer l'avatar, nous utilisons de la capture du mouvement afin d'obtenir des animations et des interactions plus naturelles. Les nouveaux énoncés sont construits par de la synthèse concaténative multi-canal.

Résultats. Nous proposons des pistes de réflexion.

Conséquences. Ces réflexions pourront guider le design d'applications interactives utilisant un avatar signeur pour apprendre et manipuler les concepts de la LSF mais aussi pour favoriser l'apprentissage en général transmis en LSF.

Lien avec l'énaction. L'énaction se situe dans le cadre de l'apport de connaissances linguistiques ayant pour vecteur de communication les gestes dans des situations interactives entre un humain et son environnement.

Mots clés : avatar signeur, Langue des Signes Française (LSF), réalité virtuelle, réalité augmentée, interaction, apprentissage, énaction.

1 Introduction

Dans cette proposition, nous souhaitons nous intéresser au rôle et à l'utilité que peut avoir un *avatar signeur*, c'est-à-dire un avatar 3D communiquant en Langue des Signes Française (LSF), en interaction avec un humain, dans un processus d'énaction.

Les interactions entre l'humain et l'avatar, et donc les possibilités d'apprentissage, dépendent du rôle joué par chacun. Les situations suivantes, dans lesquelles l'humain occupe des rôles diversifiés peuvent être envisagées : l'humain en tant qu'*observateur* (il regarde l'avatar) ; l'humain en tant qu'*acteur* (ses mouvements sont reproduits par un

avatar) ; l'humain en tant que *créateur* (l'interaction entre l'humain et l'avatar conditionne les actions de l'avatar). À cela, nous pouvons ajouter différents contextes d'utilisation. L'avatar peut être visualisé sur un écran traditionnel, en immersion dans un système de **réalité virtuelle** (RV), ou bien en superposition au monde réel en **réalité augmentée** (RA).

Le type d'interaction et le contexte définissent alors un cadre qui permet certains types d'apprentissages grâce à l'avatar signeur. Par exemple, on peut imaginer un système en immersion où l'humain est observateur et l'avatar le professeur, afin d'apprendre la Langue des Signes Française (voir section 2). Dans une application différente, l'humain peut endosser le rôle de créateur dans un système de réalité augmentée. Ainsi, à travers de l'interaction gestuelle, il peut poser des questions à un avatar sur un lieu dans lequel il se trouve ou sur un objet de son environnement afin d'acquérir des connaissances sur ce lieu/objet (voir section 3). Quelles sont les possibilités offertes par un avatar signeur ? Quel type d'environnement est adapté à quel type d'application ? Quel type d'interaction est alors proposée ? Ces questions constituent des défis à part entière qui pourront être développés et discutés dans le cadre du colloque.

Quel que soit le contexte choisi, l'avatar doit effectuer des mouvements réalistes afin de provoquer une interaction naturelle avec l'humain et donc d'enclencher le processus d'énaction. Pour cela, nous utilisons de la capture du mouvement pour animer nos avatars. Toute la difficulté de l'utilisation de la capture de mouvement réside dans la création de nouveaux énoncés qui n'ont pas été capturés. En effet, limiter la capture du mouvement à du rejeu n'aurait que peu d'intérêt par rapport à la vidéo. La possibilité de créer de nouvelles phrases à la demande dans un système dynamique offre des perspectives beaucoup plus intéressantes, notamment en terme d'enrichissement du corpus initial ou au niveau de l'interaction [DLT11]. La solution que nous proposons est la synthèse concaténative de mouvements [GCDLN11] qui permet d'éditer des composants de la langue des signes à différents niveaux. En plus de la production d'animations de l'avatar qui peuvent être utilisées dans des applications diverses, il est à noter que la capture et l'édition du mouvement offrent d'autres avantages, en particulier la sauvegarde de la langue signée (voir section 4).

Plusieurs situations d'énaction dans le cadre d'applications concrètes sont explorées dans les sections suivantes.

2 Apprentissage de la LSF

Comme toutes les langues, la Langue des Signes Française (LSF) s'apprend grâce à l'interaction des individus avec d'autres individus pratiquant la même langue. Lorsque cette interaction n'est pas possible, pour les langues orales, les autres individus peuvent être remplacés par des bandes son enregistrées. La LSF étant une langue visuo-gestuelle, l'équivalence dans ce cas est la vidéo d'une personne signant. Nous proposons de remplacer la vidéo par un système programmable capable de simuler et d'animer un avatar signeur.

On peut noter que l'avatar est une alternative bien plus riche et complète que la vidéo. Tout comme la vidéo, l'avatar signeur peut rejouer les mouvements à l'infini là où un humain serait fatigué, jouer le mouvement plus ou moins rapidement ou encore se focaliser sur une partie du corps. Outre l'anonymat garanti au signeur, l'avatar permet d'avoir une vision 3D des gestes (contrairement à la vidéo qui est 2D) et permet donc de mieux visualiser et mesurer les composants des signes, leur emplacement dans l'espace de signation, ou encore de changer de point de vue en tournant autour ou bien en zoomant

sur l'avatar. Il est possible, entre autres, de choisir entre regarder l'avatar signeur et *être* l'avatar signeur afin de regarder ses propres mains à travers lui, et par la même occasion, apprendre à signer de son propre point de vue. La compréhension des gestes est alors facilitée car l'apprenant n'a plus à mentalement faire l'effort d'inverser les gestes visualisés, une option très utile pour vérifier que les signes ont été appris correctement et de s'auto-corriger. L'avatar favorise donc l'apprentissage dans des conditions proches de celles de la vie réelle, voire même dans de meilleures conditions. Ainsi l'apprentissage de la LSF est rendu plus efficace et plus accessible à tous.

De plus, bien que l'avatar soit utilisable grâce à une visualisation 2D sur un écran classique, l'idéal, pour profiter pleinement de la 3D, est une visualisation en environnement immersif, de réalité virtuelle ou de réalité mixte (pour continuer à voir ses propres mains). Les questions soulevées par l'usage de la réalité virtuelle ou augmentée sont les suivantes : quelles sont les possibilités offertes par les nouvelles technologies ? Quelles en sont les limitations ?

Enfin, l'usage du numérique permet de modéliser l'avatar, mais également l'espace de signation, donnant ainsi la possibilité d'augmenter la représentation des énoncés en LSF par la visualisation dynamique de données représentant les entités du discours. Là également plusieurs questions émergent : comment cette visualisation peut-elle améliorer l'apprentissage de la LSF ? Doit-elle intervenir de manière simultanée à la production des signes et dans quel espace de représentation ? Lorsqu'il s'établit un dialogue entre l'utilisateur et l'avatar, comment évoluent respectivement les espaces de signation des signeurs ?

3 Visite culturelle et touristique en LSF

Dans cette section, nous prenons l'exemple de l'humain qui a un rôle de créateur dans un environnement en réalité augmentée. Là encore, pour le type d'applications visées, nous pouvons aisément faire un parallèle avec le monde de l'audio. Lors de la visite d'un musée, il est maintenant monnaie courante d'utiliser des *audio-guides* plutôt qu'un guide, afin de procéder à la visite d'un site culturel de façon personnalisée. Imaginons alors l'équivalence en LSF : la *vidéo-guide*. Le visiteur du site touristique peut s'équiper de lunettes de réalité augmentée, elles-mêmes équipées de caméras et d'un système de capture de la direction du regard. Lorsque l'utilisateur regarde un monument ou bien un objet précis, un avatar signeur peut alors endosser le rôle du guide et l'instruire sur le monument ou l'objet qu'il est en train de regarder. Un processus d'interaction entre l'humain, l'environnement et l'avatar signeur se met alors en place. Les bénéfices d'un tel dispositif sont évidents au regard du manque cruel de guides capables de pratiquer la LSF ou comme alternative à la vidéo sur tablette pour les mêmes raisons que celles évoquées dans la section précédente. Dans ce cadre, on peut s'interroger d'une part sur l'existence d'autres avantages à utiliser un avatar signeur, et d'autre part sur les autres technologies qui pourraient être envisagées.

4 Sauvegarde du patrimoine culturel

L'enregistrement des signes et des phrases signées en LSF dans une base de données indexée est un moyen d'*écrire* la langue signée, et d'y accéder automatiquement. De la même façon qu'il est primordial d'enregistrer des sons pour les langues orales, l'enregistrement du mouvement sous la forme de données numériques permet de garder en mémoire les informations en LSF. Le format utilisé pour les avatars est généralement

donné par les positions (respectivement les angles) successives de postures du squelette. C'est un moyen efficace et peu onéreux de stocker et de transférer l'information, comparativement à l'information vidéo. En ce qui concerne l'accès, il est également possible de récupérer les morceaux de mouvements sur tout ou partie du corps à partir de la spécification d'un groupe articulations ou de zones corporelles (par exemple les mains ou les yeux), d'une succession temporelle de postures spécifiques (par exemple le mouvement de la main entre deux instants), de signes ou de phrases en LSF, les informations indexées étant préalablement annotées manuellement ou automatiquement [NLG17, NRLG18].

Ces enregistrements constituent la base pour la création (i) d'un dictionnaire du français vers la LSF – qui existe déjà très partiellement sous forme de vidéos de signes isolés, comme sur le site Elix [ELI] par exemple –, mais aussi (ii) de la LSF vers le français, étant donné que l'approche par indexation multi-canal permet de retrouver un signe (ou un groupe de signes) en spécifiant par exemple la configuration manuelle recherchée, l'emplacement dans l'espace de signation ou le type de mouvement. Le dictionnaire LSF vers français est actuellement inexistant et plus difficilement réalisable à partir de vidéos.

5 Conclusion

Nous souhaiterions étudier les possibilités qu'offre l'avatar signeur au sein de mécanismes d'énaction dans des contextes interactifs avec différents degrés d'immersion ou de réalisme.

Références

- [DLT11] Pierre De Loor and Jacques Tisseau. Réalité virtuelle et énaction. *Journal de l'Association Française de Réalité Virtuelle*, 10 :3 pages, 2011.
- [ELI] ELIX. Elix : le dictionnaire vivant en langue des signes. <https://www.elix-lsf.fr/>.
- [GCDLN11] Sylvie Gibet, Nicolas Courty, Kyle Duarte, and Thibaut Le Naour. The sign-com system for data-driven animation of interactive virtual signers : Methodology and evaluation. In *Transactions on Interactive Intelligent Systems*, 2011.
- [NLG17] Lucie Naert, Caroline Larboulette, and Sylvie Gibet. Annotation automatique des configurations manuelles de la langue des signes française à partir de données capturées. In *Proceedings of JFIG*, oct 2017.
- [NRLG18] Lucie Naert, Clément Reverdy, Caroline Larboulette, and Sylvie Gibet. Per channel automatic annotation of sign language motion capture data. In *Proceedings of LREC*, 2018.