Animating virtual signers: the issue of gestural anonymization

Félix Bigand, E. Prigent, Annelies Braffort

Animating Virtual Signers : The Issue of Gestural Anonymization

Félix Bigand
LIMSI - CNRS
Orsay, France
felix.bigand@limsi.fr

Elise Prigent
LIMSI - CNRS
Orsay, France
elise.prigent@limsi.fr

Annelies Braffort
LIMSI - CNRS
Orsay, France
annelies.braffort@limsi.fr

ABSTRACT

This paper presents an ongoing PhD research project on visual perception and motion analysis applied to virtual signers (virtual agents used for Sign Language interaction).

Virtual signers (or signing avatars) play an important role in the accesibility of information in sign languages. They have been developed notably for their capability to anonymize shape and appearance of the content producer. While motion capture provides human-like, realistic and comprehensible signing animations, it also arises the question of anonymity. Human body movements contain important information about a person's identity, gender or emotional state. In the present work, we want to address the problem of gestural identity in the context of animated agents in French Sign Language. On the one hand, the ability to identify a person from signing motion is assessed through psychophysical experiments, using point-light displays. On the other hand, a computational framework is developed in order to investigate which features are critical for person identification and to control them over the virtual agent.

KEYWORDS

Sign Language, Virtual Signer, Motion Capture, Perception, Identity Recognition

ACM Reference Format:
Félix Bigand, Elise Prigent, and Annelies Braffort. 2019. Animating Virtual Signers : The Issue of Gestural Anonymization. In *ACM International Conference on Intelligent Virtual Agents (IVA '19), July 2–5, 2019, PARIS, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3308532.3329410

1 BACKGROUND AND MOTIVATION

Since 2005 [1], French public institutions must provide accessible content to users regardless of their disability, including deaf people using Sign Language (SL). For now, accessibility in SL is mainly achieved by pre-recorded videos. This cannot enable real-time interaction nor anonymity for the content provider.

More generally, the need for producing messages anonymously is an important demand of deaf people as many of them have problems communicating through written content. Indeed, most deaf people communicate in a sign language. The spoken language of their country (e.g. French) is only a second language and is almost impossible to be learnt without any auditory cues. As a consequence, video remains the main communication tool used by native signers. It arises the question of anonymity and concerns many domains, from the work of professional journalists to the social networks where deaf users leave comments. The development of virtual signers is thus expanding, in order to make written material on internet more accessible to deaf users.

Several studies at LIMSI in Automatic Sign Language Processing are part of the elaboration of virtual signers [7] [26] [3]. While this is not the only existing method [7], these virtual agents tend to be animated using motion capture (mocap) on real actors [8], as it provides highly realistic and comprehensible content. Movements of the 'mocapped' person are then mapped to the virtual agent. With such accurate systems, the avatar motion convey rich information that can make the actor identifiable, such as the voice can allow a given person to be identified.

Realism, agreeableness and comprehensibility of virtual signers have been assessed in several studies [13] [14] [1]. Other studies evaluated SL mocap data from a linguistic approach [5] [17] [18]. However, the problem of anonymization and gestural identity have not been addressed in the context of SL animation. The aim of this work is to better understand how this critical information is hidden in the complex motion stimulus, in order to provide a better control of the virtual signers. For that, we extend the gestural identity problem to the specific motion patterns of French Sign Language. First, we aim at finding critical features which differentiate gestural styles of human signers. Second, we apply it to motion generation systems in order to provide an animated agent model with controllable gestural style.

In the next section (section 2), we present the current developments of the project and the last section (section 3) presents what has already been done and what action is planned.

2 PROPOSED METHODOLOGY : COMBINING PERCEPTION STUDIES AND COMPUTING APPROACHES

Humans can extract important information from biological motion (i.e. movements produced by humans or animals), such as one's intentions, emotions, or identity. How such information can be retrieved from complex movements remains a challenging question for both computer vision and visual perception areas. It seems crucial to consider this complex matter for the generation of virtual agents movements. Providing realistic, human-like and anonymized animations would ensure higher acceptability and comprehensibility than actual signing avatars.

We have adopted a pluridisciplinary approach, combining perception studies and computing methods. In a first step (section

[1] https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000809647&categorieLien=id

IVA '19, July 2–5, 2019, PARIS, France

ACM ISBN 978-1-4503-6672-4/19/07…$15.00
https://doi.org/10.1145/3308532.3329410

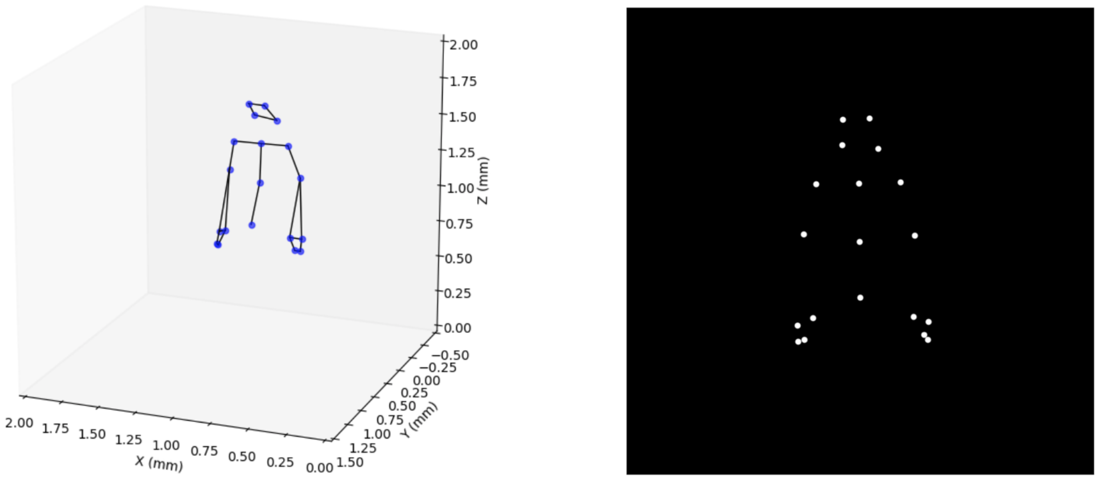

Figure 1: The 19 3D markers (left) provided by our mocap system, then transformed into point-light display (right).

2.1), we have to verify if the identity of a signer can be perceived only with motion cues. Then (section 2.2), we have to design methods that will allow us to extract the useful features using machine learning approaches.

2.1 Gestural identity perception

The studies of Johansson [11] [12] were the first to introduce the notion of point-light (PL) stimuli. This display separates information given by dynamic cues from characteristics such as shape or aspect of the person. This is why we have decided to use this method for our purpose. The key idea is to attach small point lights to the major joints of a person's body and film it in front of a dark background. Using this device, Johansson showed that humans could recognize a set of moving dots as a human walker. Point-light displays are widely used since then. Different studies demonstrated that they contain enough information to recognize familiar people from their gaits [6] [16] [10] [31]. It also inspired studies in sign language perception, evaluating comprehensibility with this reduced setup [24] [27].

Relying on passive reflective markers, mocap systems capture equivalent information as point-light displays, and mocap corpora can be used for this kind of studies. Our first analysis of signers motion has been done by computing PL stimuli from mocap data, in order to elaborate an identification task. The aim is to verify first that the identity of a signer can be transmitted through the movement only.

Identification task

In a perceptual experiment, participants are asked to identify signing actors shown as point-light displays. Four different signers describe images in French Sign Language. They are displayed as videos of white moving dots, on a dark background. Three of them are well known from the general public, the fourth is not. Each participant specifies which signer he/her knows before viewing the stimuli.

Stimuli

Point-light stimuli are generated from mocap data. Each mocap sensor representing major joints of the body is displayed, white on black. We use MOCAP1 [3], a 3D corpus of motion capture data on French Sign Language (LSF). The 3D body movements of eight LSF native signers have been recorded. From this collection, we selected four signers, two men and two women. A set of 23 reflective markers was attached to their body. We did not include movements of the signer's face. Actors were wearing suits to record markers of the shoulders, the elbows, the wrists, the hands and the chest. Four sensors were attached to a cap and recording the head movements.

Recordings were done with an Optitrack S250e, equipped with 10 cameras with a spatial resolution of 0.7 Mpixels and a temporal resolution of 250 Hz. From the 23 markers, we derived 19 virtual markers which optimally describe the major joints of the body in PL displays. Figure 1 shows the 19 markers in a 3D view and point-light display. All the stimuli are displayed in front view.

Design and procedure

The participants take part in the experiment through an online-survey. Native signers are more likely to identify the actors in the stimuli, however the experiment is also open to non-signers who could interact with signing people. The language level and the degree of interaction with signers are evaluated in the first place.

Before the test sessions, the four signing actors are introduced to the participants with a short video of them using LSF. Unlike the test stimuli, this introducing video includes the person appearance, clothing etc. The participants specify whom of the four signers they know. After one training trial, the test session consists of 20 trials. A trial begins with a 10-sec stimulus display, followed by the presentation of 5 buttons. The task is to identify the signer. Four buttons are illustrated by the four signers (screenshot from the introducing video) and another button enables them to answer that they did not recognize the signer. Each signer is presented five times, with five different contents (each photo is presented only once and by only one signer).

2.2 Gestural identity and machine learning

Figure 2: A virtual signer developed at LIMSI [4].

In conjunction with this study, we aim at providing a computational framework for the control of signing agent gestural style. The algorithm is developed for analysis and synthesis of the signers motion. Using machine learning techniques, we will be in a position to extract critical features for identification and to control them over the agent.

Proposed methodology based on motion studies

The key hypothesis is that mostly the dynamic part of signing motion accounts for person identification, as it has been shown for walking [31]. We present here one possible approach developed in biological motion studies, which reduces the motion data to a low-dimensional space. The principal components of this reduced representation are then used to feed a linear classifier. This framework provides a way to investigate critical features for person identification, with no a priori assumption.

Studies assessed candidate features by manipulating motion stimuli and evaluating the effect on recognition through experiments [15] [31] [2] [19]. A causal link was then established between the candidate features and the task. Other approaches address the question as a pattern recognition problem, with no a priori hypothesis [29] [30] [20] [21] [25]. Statistical analysis (PCA, Fourier-based decomposition) of the motion data extracts components which are critical for the task (identification, gender classification..). They might be less interpretable than classic features such as velocity, quantity of movement etc. This approach takes its inspiration from eigenfaces, in the face recognition domain [22] [23]. O'Toole et al. demonstrate that identifiable faces can be reconstructed from only a subset of eigenvectors of a covariance matrix of faces. In this study, we propose to extend the idea of eigenpostures and eigenwalkers [29] to the signers.

Data processing

The proposed methodology consists of working at the sign level, in order to limitate the linguistic complexity that could be a source of variation between signers, allowing the person identification. Isolated signs will be analysed thanks to a mocap system, beginning with periodic signs such as "pédaler" (to cycle) or "marcher" (to walk). Signing data can be seen as a time series of postures described by the 3D coordinates of 19 markers (Figure 1). Each posture refers to a 57-dimensional vector :

$$p = (x_1, y_1, z_1, ..., z_{19}) \tag{1}$$

This representation can be reduced to an average posture and eigenpostures for walking data [29] [30], we will investigate its extent to our signing data. A 57xN (e.g. N=5,000 postures) matrix **S** representing one signer phrase, can then be reduced to a 57xd matrix **F** (d << N), depending on the number d of components we use. This value is set so that the components cover the overall variance at the most.

Linear classifier for person identification

The second step of this computational framework is to train a linear classifier to identify signers with the extracted features. A linear discriminant function is computed by regressing the signers labels on their features. The weights of the linear regression **w** are optimized (minimizing the least-square error) following this relation :

$$w^T F = c \tag{2}$$

c is the class vector indicating the label to which signer i belongs ($c_i = 1$ if signer 1). This discriminant function **w** thus optimally separates the feature space according to signers gestural identity.

3 FUTURE WORK

In this PhD project, we adapted MOCAP1 corpus to the identification purpose and we proposed lines of research for motion analysis and synthesis, using machine learning techniques. At this stage :

- We designed a perceptual experiment to assess the identification of signers shown as point-light displays. It will provide preliminary results and will help to design a second experiment including computational manipulations on specific motion cues. It will also enable us to evaluate the need for the creation of a new mocap corpus, specific to identification.
- We are currently developing algorithmic methods for motion analysis. Features are extracted from the mocap data with a data-driven approach, and then feed a linear classifier for person identification.
- We evaluated methods for motion synthesis in order to animate the virtual signer and control the extracted features, as presented in the following.

First part of the future work consists of running the perceptual experiment, collecting and analysing data. Second, we will refine the proposed methods for motion decomposition and aim at developing a framework that is invertible (i.e. that enables visualization of the extracted features). A simple Fourier-based decomposition of the time-series postures is one possible solution. If the motion data is trained for person identification based on its Fourier components, we are in a position to manipulate the components and then come back to the time-series representation. Extension of this approach to our dataset of periodic signs is ongoing. For unperiodic sign movements, more complex techniques are needed. We would refer to the notion of spatio-temporal correspondences and morphable models [9]. The idea is to define spatio-temporal differences between prototypes and a reference pattern. New motion can then be generated from a linear combination of these correspondences. Another way to circumvent the problem of temporal description is to normalize the number of frames for each sign. Tilmanne & Dutoit apply PCA to time-normalized data and demonstrate that they can re-generate walking motion in a specific style thanks to Gaussian Modeling [28]. Controlling the gestural style of our virtual signer thanks to Gaussian Mixture Models is another approach that we are evaluating as part of this doctoral project.

4 INTEREST OF THE DOCTORAL CONSORTIUM TO THIS RESEARCH

This consortium would enable us to get valuable feedback on the chosen approach and methods. The problem we address involves the development of algorithms for gesture analysis and synthesis, but also methods to evaluate the virtual agent. We would be honored to interact with the other participants on this matters. In addition, the Doctoral Consortium would offer us the opportunity to learn from other researchers about person identification in other domains, such as speech or facial expressions.

REFERENCES

[1] Nicoletta Adamo-Villani, Ronnie Wilbur, Petra Eccarius, and Laverne Abe-Harris. 2009. Effects of character geometric model on perception of sign language animation. In *Visualisation, 2009. VIZ'09. Second International Conference in.* IEEE, 72–75.

[2] Catharine D Barclay, James E Cutting, and Lynn T Kozlowski. 1978. Temporal and spatial factors in gait perception that influence gender recognition. *Perception & psychophysics* 23, 2 (1978), 145–152.

[3] Mohamed-el-Fatah Benchiheub, Bastien Berret, and Annelies Braffort. 2016. Collecting and Analysing a Motion-Capture Corpus of French Sign Language. In *Workshop on the Representation and Processing of Sign Languages*. Portoroz, Slovenia. https://hal.archives-ouvertes.fr/hal-01633625

[4] Annelies Braffort, Michael Filhol, Maxime Delorme, Laurence Bolot, Annick Choisier, and Cyril Verrecchia. 2016. KAZOO: A Sign Language Generation Platform Based on Production Rules. *Universal Access in the Information Society* 15, 4 (2016), 541–550.

[5] Fanny Catteau, Marion Blondel, Coralie Vincent, Patrice Guyot, and Dominique Boutet. 2016. Variation prosodique et traduction poétique (LSF/français): Que devient la prosodie lorsquâĂŹelle change de canal?. In *Journées dâĂŹÉtude sur la Parole*, Vol. 1. 750–758.

[6] James E Cutting and Lynn T Kozlowski. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society* 9, 5 (1977), 353–356.

[7] Michael Filhol and John Mcdonald. 2018. Extending the AZee-Paula shortcuts to enable natural proform synthesis. In *Workshop on the Representation and Processing of Sign Languages*.

[8] Sylvie Gibet. 2018. Building French Sign Language Motion Capture Corpora for Signing Avatars. In *Workshop on the Representation and Processing of Sign Languages: Involving the Language Community, LREC 2018*.

[9] Martin A Giese and Tomaso Poggio. 2000. Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision* 38, 1 (2000), 59–73.

[10] Alissa Jacobs, Jeannine Pinto, and Maggie Shiffrar. 2004. Experience, context, and the visual perception of human movement. *Journal of Experimental Psychology: Human Perception and Performance* 30, 5 (2004), 822.

[11] Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics* 14, 2 (1973), 201–211.

[12] Gunnar Johansson. 1976. Spatio-temporal differentiation and integration in visual motion perception. *Psychological research* 38, 4 (1976), 379–393.

[13] Michael Kipp, Alexis Heloir, and Quan Nguyen. 2011. Sign language avatars: Animation and comprehensibility. In *International Workshop on Intelligent Virtual Agents*. Springer, 113–126.

[14] Michael Kipp, Quan Nguyen, Alexis Heloir, and Silke Matthes. 2011. Assessing the deaf user perspective on sign language avatars. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 107–114.

[15] Lynn T Kozlowski and James E Cutting. 1977. Recognizing the sex of a walker from a dynamic point-light display. *Perception & psychophysics* 21, 6 (1977), 575–580.

[16] Fani Loula, Sapna Prasad, Kent Harber, and Maggie Shiffrar. 2005. Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance* 31, 1 (2005), 210.

[17] Evguenia Malaia, John Borneman, and Ronnie B Wilbur. 2008. Analysis of ASL motion capture data towards identification of verb type. In *Proceedings of the 2008 conference on semantics in text processing*. Association for Computational Linguistics, 155–164.

[18] Evie Malaia, Ronnie B Wilbur, and Marina Milković. 2013. Kinematic parameters of signed verbs. *Journal of Speech, Language, and Hearing Research* 56, 5 (2013), 1677–1688.

[19] George Mather and Linda Murdoch. 1994. Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 258, 1353 (1994), 273–279.

[20] Johannes Michalak, Nikolaus F Troje, Julia Fischer, Patrick Vollmar, Thomas Heidenreich, and Dietmar Schulte. 2009. Embodiment of sadness and depressionâĂŤgait patterns associated with dysphoric mood. *Psychosomatic medicine* 71, 5 (2009), 580–587.

[21] Lars Omlor and Martin A Giese. 2007. Extraction of spatio-temporal primitives of emotional body expressions. *Neurocomputing* 70, 10-12 (2007), 1938–1942.

[22] ALICE J O'TOOLE, KENNETH DEFFENBACHER, Hervé Abdi, and JAMES C BARTLETT. 1991. Simulating the âĂŸother-race effectâĂŹas a problem in perceptual learning. *Connection Science* 3, 2 (1991), 163–178.

[23] Alice J OâĂŹToole, Hervé Abdi, Kenneth A Deffenbacher, and Dominique Valentin. 1993. Low-dimensional representation of faces in higher dimensions of the face space. *JOSA A* 10, 3 (1993), 405–411.

[24] Howard Poizner, Ursula Bellugi, and Venita Lutes-Driscoll. 1981. Perception of American sign language in dynamic point-light displays. *Journal of experimental psychology: Human perception and performance* 7, 2 (1981), 430.

[25] Claire L Roether, Lars Omlor, and Martin A Giese. 2009. Features in the recognition of emotions from dynamic bodily expression. In *Dynamics of Visual Motion Processing*. Springer, 313–340.

[26] Jérémie Segouat and Annelies Braffort. 2009. Toward the study of sign language coarticulation: methodology proposal. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*. IEEE, 369–374.

[27] Vivien C Tartter and Kenneth C Knowlton. 1981. Perception of sign language from an array of 27 moving spots. *Nature* 289, 5799 (1981), 676.

[28] Joëlle Tilmanne and Thierry Dutoit. 2010. Expressive gait synthesis using PCA and Gaussian modeling. In *International Conference on Motion in Games*. Springer, 363–374.

[29] Nikolaus F Troje. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of vision* 2, 5 (2002), 2–2.

[30] Nikolaus F Troje. 2002. The little difference: Fourier based synthesis of gender-specific biological motion. *Dynamic perception* (2002), 115–120.

[31] Nikolaus F Troje, Cord Westhoff, and Mikhail Lavrov. 2005. Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics* 67, 4 (2005), 667–675.