



HAL
open science

Retrieving Human Traits from Gesture in Sign Language: The Example of Gestural Identity

Félix Bigand, E. Prigent, Annelies Braffort

► **To cite this version:**

Félix Bigand, E. Prigent, Annelies Braffort. Retrieving Human Traits from Gesture in Sign Language: The Example of Gestural Identity. International Conference on Movement and Computing, Oct 2019, Tempe, United States. pp.1-4, 10.1145/3347122.3359607 . hal-02400930

HAL Id: hal-02400930

<https://hal.science/hal-02400930>

Submitted on 16 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Retrieving Human Traits From Gesture In Sign Language : The Example Of Gestural Identity

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

Virtual signers (or signing avatars) play an important role in the accessibility of information in sign languages. They have been developed notably for their capability to anonymize the signer's appearance and to enable dynamic or interactive scenarios. Recording real movements thanks to motion capture provides human-like, realistic and comprehensible signing animations. However, such accurate systems may also convey extralinguistic information such as 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 (LSF). On the one hand, person identification from signing motion is assessed through psychophysical experiments, using point-light displays. On the other hand, a computational framework is developed for the analysis of LSF motion in order to investigate which features are critical for identification. For some applications, determining these movement parameters will enable controlling the gestural human traits of virtual signers.

CCS CONCEPTS

- **Human-centered computing** → **Visualization**; *Accessibility*;
- **Applied computing** → *Education*.

KEYWORDS

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

ACM Reference Format:

Félix Bigand, Elise Prigent, and Annelies Braffort. 2019. Retrieving Human Traits From Gesture In Sign Language : The Example Of Gestural Identity. In *Proceedings of International Conference on Movement and Computing (MOCO'19), Doctoral Consortium*. ACM, New York, NY, USA, 4 pages. <https://doi.org/https://doi.org/10.1145/3347122.3359607>

1 BACKGROUND AND MOTIVATION

The use of virtual signers brings many advantages: it is possible to modify the animation content dynamically, and to adapt the appearance of the avatar (age, gender, clothes, etc.) according to the target population [13]. Several studies are part of the elaboration of

virtual signers [7] [28] [13] [1] [4] [19] which tend to be animated using motion capture (mocap) on real actors [8], as it provides highly realistic and comprehensible content. With such accurate systems, not only the avatar motion provides the message but it also conveys rich information about the 'mocapped' person. As an example, it can make the actor identifiable, such as the voice can allow a given person to be identified.

Humans can extract important information from biological motion, such as one's intentions, emotions, or identity. How such information can be retrieved from complex movements remains a challenging question for both psychology and computer vision areas. Extraction of human traits from a complex signal is a fundamental problem that has been addressed in different domains. In the auditory field, studies investigated the perception of extralinguistic cues in speech [3] [27] and notably the recognition of a particular speaker [15]. Similar approaches have been used in the visual domain to study the categorization and identification of human faces [34] [23]. Studies of Johansson [11] [12] addressed the question for human motion, introducing 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. 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] [33].

As mocap systems capture equivalent information as point-light displays, such information might be perceived when mapping mocap recordings to a signing avatar. However, this problem has not been addressed in SL motion. Studies evaluated SL mocap data from a linguistic approach [5] [17] [18] but no one studied the extralinguistic cues conveyed by such signals. In this study, we want to address the question of gestural identity in LSF motion. The aim of this work is to find critical features which differentiate the gestural identity of human signers in order to provide a better control of the animated signers. This multidisciplinary research yields contributions to psychology, computational science and LSF animation with both fundamental and applied perspectives. On the one hand, it helps us better understand how critical information can be extracted from the complex SL motion stimulus. On the other hand, it enables controlling social and human characteristics of virtual signers regarding motion. For example, manipulating gestural identity would allow to produce signing animation "in the style of" someone or to anonymize its own signing. The need for producing messages anonymously is an important demand of deaf people as for many of them the use of writing is not obvious. Anonymizing the gesture of replayed motion can add substantial contribution to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MOCO'19, October 10-12, 2019, Tempe, AZ, USA

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7654-9/19/10.

<https://doi.org/https://doi.org/10.1145/3347122.3359607>

the existing virtual signing systems which only anonymize shape of the agent.

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

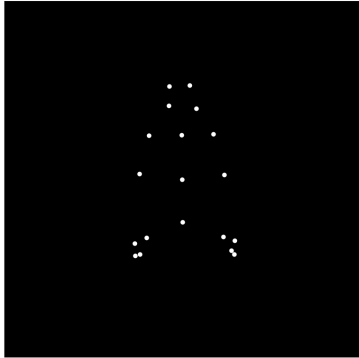


Figure 1: The 19 markers shown as point-light display.

We have adopted a pluridisciplinary approach, combining perception studies and computing methods. In a first step, we develop evaluation methods for person identification based on motion cues. Then (section 2.2), we apply machine learning techniques in order to classify signer’s identity based on motion and to extract critical features.

2.1 Gestural identity perception

We already mentioned the different studies investigating the perception of biological motion using PL stimuli. Relying on passive reflective markers, mocap systems capture equivalent information as point lights and mocap corpora can be used for this kind of studies. The first evaluation methods we develop is to compute PL stimuli from mocap data and elaborate an identification task. The aim is to assess whether the identity of a signer can be transmitted through the movement only. This is ongoing work, data will be collected in the coming months. Our three hypothesis are H1. Signers identify themselves (>chance), H2. Participants identify known signers (>chance), H3. The higher the familiarity, the higher the scores. Next sections present how the experiment was built.

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. It has been ensured that signers described different images so that participants can’t differentiate them based on the signs they chose to describe the same object. Three of them are well known from the general public, so that participants have good chance to recognize them

without prior training. Participants are shown the identities of the signers prior to the experiment and they specify to what extent they know each one.

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 [4], 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 PL 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

In conjunction with this study, we aim at providing a computational framework for the analysis of LSF motion and the extraction of social and human attributes. We apply machine learning techniques in a way to extract human features from mocap data (e.g. identity, age, emotions...). Based on the same stimuli (MOCAP1), we run a classifier using a similar approach to the one developed for face, voice and gait recognition [15][24][33]. Then, these investigations will be applied to motion synthesis and allow to control specific human features in motion. For synthesis, a new mocap corpus is needed in which we constrain signers to perform the same utterances (vs free descriptions in MOCAP1). The design of this new dedicated LSF mocap corpus will be carried out in the coming months.

Related work and general framework

The key hypothesis is that mostly the dynamic part of signing motion accounts for person identification, as it has been shown for walking [33]. We present here the general framework developed in biological motion studies, which reduces the motion data to a low-dimensional space and then runs classification methods to differentiate the data regarding a specific attribute. This framework provides a way to investigate critical features with no a priori assumption.

Prior studies assessed candidate features by manipulating motion stimuli and evaluating the effect on recognition through experiments [14] [33] [2] [20]. 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 [31] [32] [21] [22] [26]. 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 [25] [24]. O’Toole et al. demonstrate that identifiable faces can be reconstructed from only a subset of eigenvectors of a covariance matrix of faces. Similar studies used different computational techniques but used the same data-driven approach [30] [29] [9]. We extend this approach to LSF motion.

Ongoing work and preliminary results

The first step is the analysis of LSF mocap data. The aim is to extract low-dimensional features and then discriminate the categories. In this study, our goal is to differentiate signers motion regarding their gestural identity. Ongoing work is based on the same corpus as the one used for perception tests, MOCAP1. The dataset consists of 24 extracts (10 seconds each) of the utterances given by 4 signers. We present here the first analysis we ran but comparison with other techniques is ongoing too.

The motion data is represented as the 3D coordinates of 19 virtual markers (Figure 1) relative to the pelvis marker. This data can undergo two steps of normalization : it can be size-normalized or shape-normalized. In the first case, each signer’s data has the same size but relative positions of the articulations still differ, keeping intact shape. In the latter case, all signers have the same average posture. In that condition, the only information remaining is purely kinematic. We present preliminary results about the shape-normalized data.

We perform PCA in the first place to reduce the dimensionality of the original data, and decide to keep the first k principal components that represent 95% of the cumulative percentage of information. This criterion is empirical and changes whether the goal is classification or reconstruction. As shown in Figure 2, the first 2 components capture some differentiation between gestural identity of the 4 signers. We also note that the different exemplars of each signer are consistently distributed across the 2 axis. The first component enables differentiating signers 4 and 8, while the second component enables differentiating signers 2 and 6. This is confirmed by the projection of the data over the 2 axis (Figure 3). Next components (>2) don’t seem to enable any differentiation.

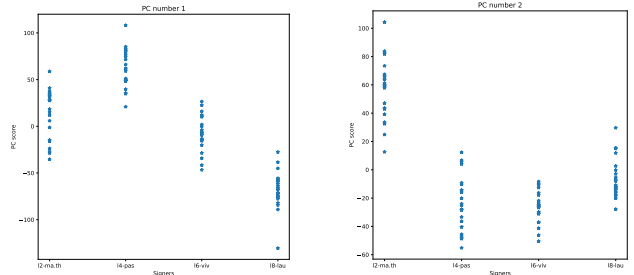


Figure 2: Scores of the first two principal components

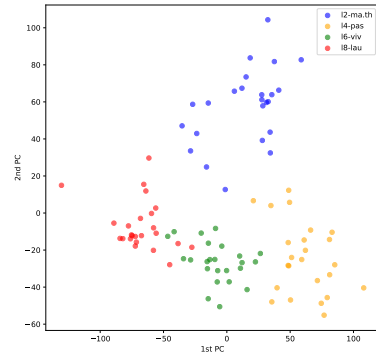


Figure 3: Projection of the dataset over the first 2 PCs.

Then, we train a linear classifier on a reduced number of first components. By running a multinomial logistic regression, we obtained the following confusion matrix for the first 2 components (Figure 4). Recognition rates for signers 2, 4, 6 and 8 are (respectively) 0.96, 0.92, 0.88 and 0.88. These results are consistent with Figure 3 showing that the highest confusion occurs between signer 6 and 8.

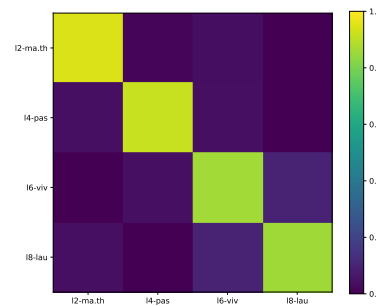


Figure 4: Confusion matrix computed from the logistic regression between PC scores and signer’s identity.

According to these results, component 1 plays an important role in differentiating the gestural identity of signer 2 and 6, and component 2 for signers 4 and 8. These are preliminary results, we propose to compare different classification techniques (linear vs non-linear), different inputs (adding velocity, acceleration, jerk)

in the future. Linear Discriminant Analysis might do a better job at discriminating the gestural identity. Neural networks such as LSTMs¹ might perform better at capturing the dynamic properties of the data. Still, the computational framework we build needs to provide the intermediate "identity features", as one application of this work is to generate new controlled motion. Our mocap corpus is still small but these preliminary results suggest that differentiation is possible between gestural identities, also depending on the neutrality of the signer's gesture.

3 FUTURE WORK

At this stage, we designed a perceptual experiment to assess the identification of signers shown as point-light displays. Data collection will be done in the coming months. We developed algorithmic methods for LSF motion analysis. Shape-normalized data is reduced using PCA, and then feeds a linear classifier for person identification. First components seems to enable differentiation between some signers, but not all of them.

Future work is to refine the computational framework for analysis by comparing different classification techniques. Moreover, a new mocap corpus needs to be recorded in order to overcome limitations of both the perception test and the computational work. Rather than freely describing images, signers will be asked to perform the same utterances so that the variability between signers is exclusively related to gestural identity. It will also integrate motion recordings of isolated signs which will allow the generation of new controlled motion. In prior work, the generation of stylistic or gender-specific manipulated motion was done almost exclusively on walking patterns. A walking pattern is quite simple compared to a complex signing utterance. Extending these synthesis approaches to isolated signs will be a first step to investigate the "identity-control" of signing animations.

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] Pascal Belin, Patricia EG Bestelmeyer, Marianne Latinus, and Rebecca Watson. 2011. Understanding voice perception. *British Journal of Psychology* 102, 4 (2011), 711–725.

[4] Mohamed-el-Fatah Benchiheb, 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>

[5] Fanny Catteau, Marion Blondel, Coralie Vincent, Patrice Guyot, and Dominique Boutet. 2016. Variation prosodique et traduction poétique (LSF/français): Que devient la prosodie lorsqu'ÄZelle change de canal?. In *Journées d'ÄZtude 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 *Intelligent Virtual Agents*. Springer, 113–126.

[14] 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.

[15] Roland Kuhn, Patrick Nguyen, Jean-Claude Junqua, Lloyd Goldwasser, Nancy Niedzielski, Steven Fincke, Ken Field, and Matteo Contolini. 1998. Eigenvoices for speaker adaptation. In *Fifth International Conference on Spoken Language Processing*.

[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] Vonjiniaina Domohina Malala, Elise Prigent, Annelies Braffort, and Bastien Berret. 2018. Which picture? A Methodology for the Evaluation of Sign Language Animation Understandability.

[20] 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.

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

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

[23] Alice J O'Toole, 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] Alice J O'Toole, 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.

[25] Alice J O'Toole, Kenneth Deffenbacher, Hervé Abdi, and James C Bartlett. 1991. Simulating the 'other-race effect' as a problem in perceptual learning. *Connection Science* 3, 2 (1991), 163–178.

[26] 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.

[27] Stefan R Schweinberger, Hideki Kawahara, Adrian P Simpson, Verena G Skuk, and Romi Zäske. 2014. Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science* 5, 1 (2014), 15–25.

[28] 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.

[29] Joëlle Tilmanne, Nicolas d'Alessandro, Maria Astrinaki, and Thierry Ravet. 2014. Exploration of a stylistic motion space through realtime synthesis. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, Vol. 2. IEEE, 803–809.

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

[31] 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.

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

[33] 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.

[34] Matthew A Turk and Alex P Pentland. 1991. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 586–591.

¹Long short-term memory