Sign Language Synthesis using Pose Priors

Paritosh Sharma, Michael Filhol

Sign Language Synthesis using Pose Priors

Paritosh Sharma
paritosh.sharma@universite-paris-saclay.fr
CNRS, Université Paris Saclay
Orsay, France

Michael Filhol
michael.filhol@cnrs.fr
CNRS, Université Paris Saclay
Orsay, France

ABSTRACT

The challenge of simulating realistic Sign Language using avatars lies in achieving accurate human-like postures for effective communication. Unlike artistic or motion capture techniques, linguist-driven procedural generation methods are widely employed, relying on skeletal representations to synthesize a broad range of signs. However, determining appropriate joint limits for these avatars is intricate due to inter-joint and intra-joint dependencies, as well as variations in biomechanical properties. In this context, our work addresses this problem by introducing a pose corrector, enhancing an established Sign Language synthesis technique. Focused on rectifying extreme joint rotations, our approach incorporates a pre-trained poser based on existing work, integrated with a 21-joint character model. The correction process involves applying linguist-defined constraints using AZee language and subsequent pose corrections, showcasing promising advancements in obtaining more natural sign gestures.

CCS CONCEPTS

• **Procedural animation**; • **Motion processing**;

KEYWORDS

Motion Synthesis, Avatar, Sign Language, Procedural Animation

ACM Reference Format:
Paritosh Sharma and Michael Filhol. 2024. Sign Language Synthesis using Pose Priors. In *9th International Conference on Movement and Computing (MOCO '24), May 30–June 02, 2024, Utrecht, Netherlands.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3658852.3659080

1 INTRODUCTION

The field of simulating realistic Sign Language using signing avatars has been a topic of growing interest in recent years. In contrast to artistic and motion capture techniques, linguist-driven procedural generation techniques[7] are widely used on a skeletal representation of an avatar to provide a broad coverage of synthesized signs. Since these avatars imitate human signers, natural human-like posture generation is paramount to not only the functionality of the avatar but also for the comprehensibility of the Sign Language discourse.

MOCO '24, May 30–June 02, 2024, Utrecht, Netherlands

ACM ISBN 979-8-4007-0994-4/24/05
https://doi.org/10.1145/3658852.3659080

Modeling a human body accurately for Sign Language highly depends on the postures required to achieve a discourse. One common concern faced by most procedural animation based signing avatars is the awareness of the range of human motion due to physical joint limits. If the human joint limits are modeled too conservatively, the avatar might never attain the desired pose. On the other hand, overly relaxed human joint limits might lead to synthesis of impossible poses for an anatomical human body. Thus, setting appropriate human joint limits is challenging because not only that different joints have different ranges of angles, bio-mechanics literature [6] [17] [9] suggests that the range of angle varies depending on the positions of other joints (inter-joint dependency) or other degrees-of-freedom in the same joint (intra-joint dependency). For example, the range of motion for the middle finger depends on the configuration of the other fingers.

In this work, we enhance an established Sign Language synthesis technique by introducing a pose corrector based on existing data-driven methodologies within the field. Our corrector targets extreme joint rotations, aiming to rectify and improve the synthesis of poses in the context of Sign Language discourse. This innovation contributes to a more natural and fluid representation of sign gestures, ultimately enhancing the overall quality of Sign Language synthesis.

2 BACKGROUND

Our background research commences with an examination of joint extremities within current kinematic approaches employed in Sign Language synthesis. Subsequently, we delve into related studies within the broader domain of human motion synthesis. Finally, we conclude our investigation by reviewing recent research focused on pose-conditioned joint limits derived from human body priors.

2.1 Kinematics in Sign Language synthesis

The challenge of determining a character pose satisfying specified constraints is a well-explored problem, often characterized by its undetermined nature, allowing multiple poses to meet the given constraints [2, 18]. This can result in the generation of diverse and sometimes unnatural joint configurations for the same set of constraints. In Sign Language synthesis, existing systems often grapple with this challenge by restricting the avatar through limited Inverse Kinematics (IK) chains, such as employing two for hands, one for the spine, etc. [12]. However, this approach imposes a significant burden on animators (or linguists in this case) who must specify an excessive number of constraints. For example, movement of the left arm to an unreachable point for the hand ik chain will require a separate movement specification of the spine IK, allowing the avatar to bend and reach the point.

Recent research has pioneered techniques in sign language pose correction, marking a significant advancement in the field. These

techniques aim to confine animated characters within the space of natural poses based on biomechanics and kinesiology, representing the first work of its kind in sign language pose correction. One notable approach involves measuring the contribution of each individual IK chain separately to restrict the character's movements [11]. Another strategy utilizes a relaxation algorithm for specific joints to achieve more natural poses [8]. Despite these advancements, describing a broad range of body poses remains challenging, especially for dynamic styles lacking straightforward biomechanical interpretations. For instance, signers may intentionally stretch their arms beyond the typical comfort range or manipulate an IK chain beyond its contributions to convey meaning, such as illustrating the physical difficulty of a task.

2.2 Human Body Prior

Capturing the range of motion for an avatar poses a considerable challenge, as existing methods often necessitate the estimation of numerous specific parameters through repeated experiments. This complexity makes it challenging to apply these methods as general joint-limit constraints [9]. Consequently, there has been notable research focusing on the representation of joint angles and the integration of forward kinematics within neural networks. Akhter and Black [1] propose a pose-dependent model of joint angle limits, demonstrating effective generalization while avoiding the generation of physically impossible poses. Building upon this idea, they introduce a heuristic score function, denoted as "prior," over body pose that penalizes impossible configurations while permitting feasible ones [13], leveraging a variational auto-encoder for this purpose. This heuristic score function acts as a pre-established criterion based on certain features, guiding the understanding of the distribution of human joint postures. This implies that the joint posture of the human body is not arbitrary but follows a specific distribution. Through pre-training a model to understand this posture distribution, we can identify the most plausible posture among those that satisfy the Inverse Kinematics (IK) requirements. This is achieved by utilizing a latent space to represent poses and learning the distribution density function of poses. Post-training and distribution of action sets, a reasonable pose can be derived. This pose serves as the IK solution by determining the posture with the highest probability within the latent space under given constraints.

Built on this work, we use this pose prior to restrict the poses generated from our synthesis model.

3 OVERVIEW

The primary idea of our work involves leveraging a pre-trained poser based on work in [13]. This pre-trained poser is then integrated with our existing constraint resolution algorithm to *correct* any unnatural joint angles, ensuring that the generated poses adhere to realistic and feasible configurations.

3.1 Character Model

We employ a 21-joint character model, excluding finger joints, and enhance it by incorporating a restricted set of body sites [12] to facilitate Inverse Kinematics (IK) operations. These sites are affixed to the mesh through parent-child constraints, contributing to the overall flexibility and effectiveness of the IK operations within the character model.

3.2 Pose Synthesis

The pose synthesis can be divided into 2 stages - constraining the avatar using constraints defined by a linguist in the AZee language, and then applying pose corrections to those evaluations to fix joint limits.

3.2.1 Constraining the Avatar using AZee. AZee [7] allows us to write parameterised signed forms for semantic functions. A sign language utterance is encoded in the form of a hierarchy of applied production rules instead of a sequence. Given a description, it produces a timeline specifying all parts of the utterance to render with the avatar, thereby addressing the issues of non-manual features synchronisation, sign concurrency, and timing. Furthermore, AZee's timeline specifications also carry interpolation information and are essential for synthesising the utterance. To illustrate this concept, let's examine the following AZee expression from a pre-existing corpus [3] where the signer places "Iraq" on the eight signing space(to reuse later on for some reference).

```
:about-ref
    'pt
    ^Rssp
    'info
    :Irak
```

Evaluating this expression with the AZee interpreter generates a recursive representation of blocks to be animated [15].

Each of these blocks contain some constraints for the avatar. Figure 2 represents the corresponding block structure for the above expression. These blocks can be summarized as follows:

- **SIG**: The constraints for the sign of the country *Iraq*. This is done by constraining the side of the avatar's right palm on the right side of the head **??**
- **GAZE**: Constraints for gazing on the right signing space with a slight movement of the bust.

Thus, for time frame f, a set of ordered constraints $c = c_1, c_2, ..., c_n$ inside a set of parallel blocks $B = B_1, B_2, ..., B_n$ generate a posture for the avatar.

To synthesize a pose for time t, our algorithm optimises the loss for each of these constraints. For inverse kinematics constraints, this loss is the distance of the current body site position from its desired position while for forward kinematics constraint, it is the angular magnitude of the actual bone orientation and the desired bone orientation.

3.2.2 Pose Corrector. After obtaining our calculated pose for each time frame from the evaluated AZee constraints, we send each of these poses to the pose predictor module to apply the joint limit corrections. Our combined optimisation algorithm is shown in algorithm 1.

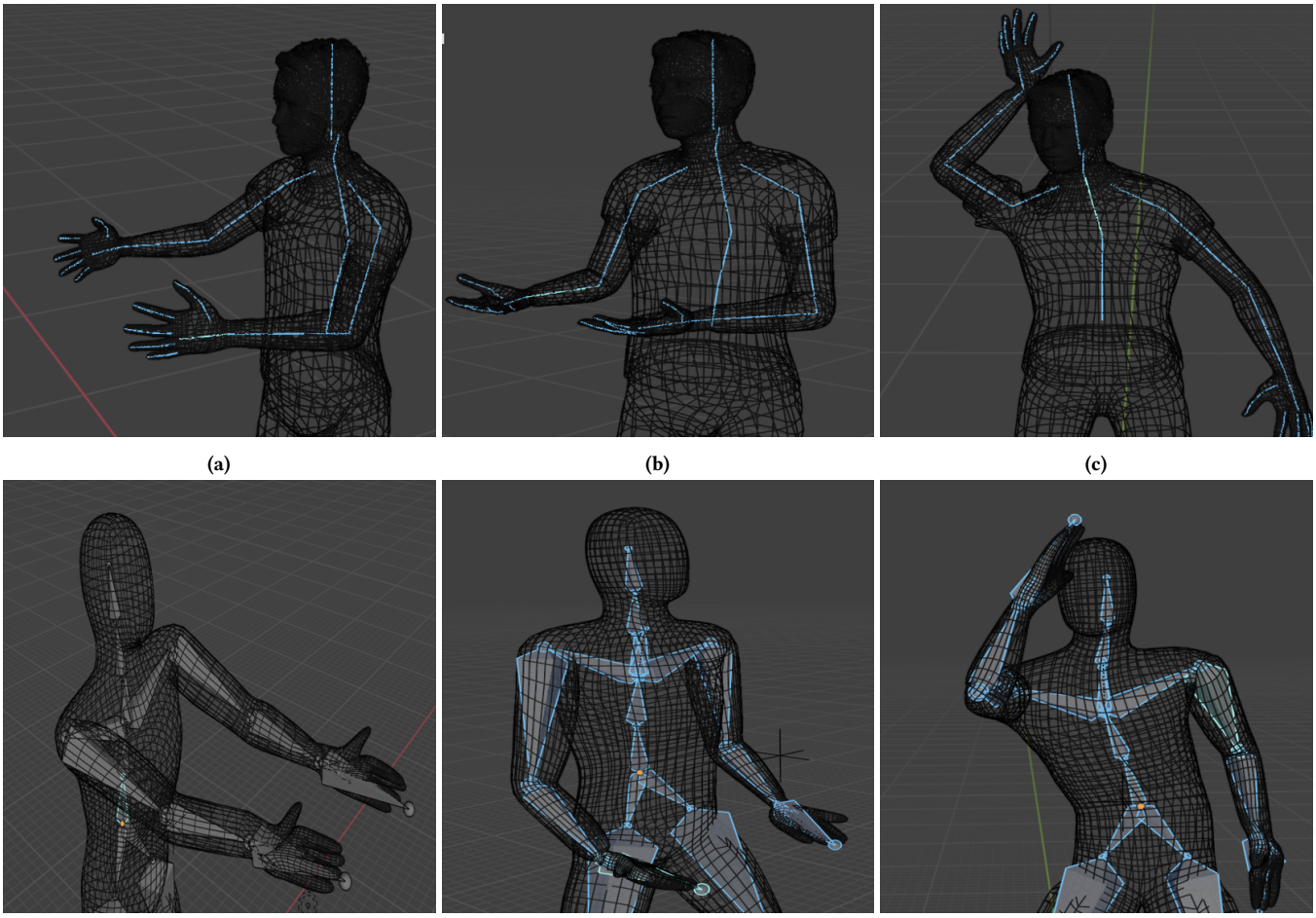

Figure 1: Results for the AZee expressions (from left to right): *armoire, maintenant, about-ref(R̂ssp, Irak)*

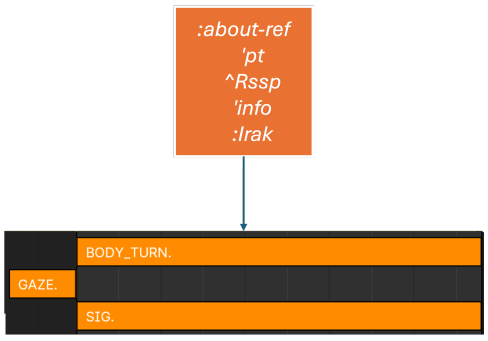

Figure 2: Block breakdown for the example expression

4 DISCUSSION AND FUTURE WORK

Figure 3: Problems with elbow placements for the sign *tree*(left) compared to a valid synthesis(right)

We implemented our animator module inside Blender[4]. We can observe in figure 1 more natural poses and less joint breaking when compared to the previous approach especially for spine and palms.

Algorithm 1 Combined Optimization Algorithm

```
for frame in frames do
    switch_cursor_to_frame(f)
    for parallel_block in self.parallel_blocks do
        constraints.add(parallel_block.constraints)
    end for
    for constraint in constraints do
        constraint.apply(frame)
    end for
    model.pose_embedding
    model.global_trans
    optimizer = . . .
    for epoch in range(max_epochs) do
        optimizer.zero_grad()
        . . .
        optimizer.step()
        if loss.item() < threshold then
            break
        end if
    end for
    posture.keyframe(frame)
end for
```

For certain scenarios, the pose corrector over-corrects the pose by extensively, thus changing the meaning of the sign(example in figure 3). This happens mostly because the poser has a bayesian bias and a potential solution to this could be studying the continuity of the pose corrector with respect to signing spaces and subsequently improving the learnt pose prior based on neural distance fields [16] or diffusion [10].

In future, we can apply similar data-driven corrections for the hand model[14] and for facial synthesis as well(currently implemented using FACS[5]).

ACKNOWLEDGMENTS

This work has been funded by the Bpifrance investment "Structuring Projects for Competitiveness" (PSPC), as part of the Serveur Gestuel project (IVès et 4Dviews Companies, LISN — University Paris-Saclay, and Gipsa-Lab — Grenoble Alpes University).

REFERENCES

[1] Ijaz Akhter and Michael J Black. 2015. Pose-conditioned joint angle limits for 3D human pose reconstruction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1446–1455.

[2] Bobby Bodenheimer, Chuck Rose, Seth Rosenthal, and John Pella. 1997. The process of motion capture: Dealing with the data. In *Computer Animation and Simulation'97: Proceedings of the Eurographics Workshop in Budapest, Hungary, September 2–3, 1997*. Springer, 3–18.

[3] Camille Challant and Michael Filhol. 2022. A First Corpus of AZee Discourse Expressions. In *Language Resources and Evaluation Conference (Proceedings of the 13th Language Resources and Evaluation Conference)*. Marseille, France. https://hal.archives-ouvertes.fr/hal-03714951

[4] Blender Online Community. 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam. http://www.blender.org

[5] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior* (1978).

[6] AE Engin and S-M Chen. 1986. Statistical data base for the biomechanical properties of the human shoulder complex—I: Kinematics of the shoulder complex. (1986).

[7] Michael Filhol, Mohamed Hadjadj, and Annick Choisier. 2014. Non-manual features: the right to indifference. In *International Conference on Language Resources and Evaluation*.

[8] Michael Filhol and John McDonald. 2020. The synthesis of complex shape deployments in sign language. In *Proceedings of the 9th workshop on the Representation and Processing of Sign Languages*.

[9] H Hatze. 1997. A three-dimensional multivariate model of passive human joint torques and articular boundaries. *Clinical Biomechanics* 12, 2 (1997), 128–135.

[10] Junzhe Lu, Jing Lin, Hongkun Dou, Yulun Zhang, Yue Deng, and Haoqian Wang. 2023. DPoser: Diffusion Model as Robust 3D Human Pose Prior. *arXiv preprint arXiv:2312.05541* (2023).

[11] John McDonald, Rosalee Wolfe, Jerry Schnepp, Julie Hochgesang, Diana Gorman Jamrozik, Marie Stumbo, Larwan Berke, Melissa Bialek, and Farah Thomas. 2016. An automated technique for real-time production of lifelike animations of American Sign Language. *Universal Access in the Information Society* 15 (2016), 551–566.

[12] Fabrizio Nunnari, Michael Filhol, and Alexis Heloir. 2018. Animating azee descriptions using off-the-shelf ik solvers. In *Workshop on the Representation and Processing of Sign Languages*.

[13] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. 2019. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10975–10985.

[14] Javier Romero, Dimitrios Tzionas, and Michael J Black. 2022. Embodied hands: Modeling and capturing hands and bodies together. *arXiv preprint arXiv:2201.02610* (2022).

[15] Paritosh Sharma and Michael Filhol. 2022. Multi-Track Bottom-Up Synthesis from Non-Flattened AZee Scores. In *7th Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual & the Textual Challenges and Perspectives (SLTAT 7)*.

[16] Garvita Tiwari, Dimitrije Antić, Jan Eric Lenssen, Nikolaos Sarafianos, Tony Tung, and Gerard Pons-Moll. 2022. Pose-ndf: Modeling human pose manifolds with neural distance fields. In *European Conference on Computer Vision*. Springer, 572–589.

[17] Xuguang Wang, Michel Maurin, Frédéric Mazet, Nilo De Castro Maia, Karine Voinot, Jean Pierre Verriest, and Michel Fayet. 1998. Three-dimensional modelling of the motion range of axial rotation of the upper arm. *Journal of biomechanics* 31, 10 (1998), 899–908.

[18] Chris Welman. 1993. Inverse kinematics and geometric constraints for articulated figure manipulation. (1993).