EXTENDING MORPHS IN AZEE USING POSE SPACE DEFORMATIONS

Paritosh Sharma, Michael Filhol

LISN, CNRS, Université Paris-Saclay, Orsay, France

ABSTRACT

Signing avatars have become increasingly important for sign language synthesis. However, to behave realistically, they must be able to replicate the coordinated activity of human hand movements and facial expressions. Most methods currently evaluate such motion using just kinematic techniques, which can limit the realism of the virtual characters. We propose a new methodology for creating a set of *morphs* in the AZee language to address this issue. We encapsulate a set of human movements and map their respective pose space deformations within our morphs. This allows us to capture the rigid as well as the non-rigid shape changes of the human anatomy and also addresses the stretching and contracting of the skin at its extremities. We create our pose space deformations based on the study of local avatar movements and a popular cognitive facial model for facial expressions. We integrate our set of morphs in our existing blender add-on implementation for AZee with a standard parameterized 3D avatar model, resulting in a fully articulated avatar that can produce more realistic movements with a faster real-time synthesis. The proposed methodology has the potential to enhance the realism of signing avatars and contributes to the development of a more intuitive toolkit for AZee linguists.

***Index Terms*—** Sign Language, Avatar, AZee, Morphs

1. INTRODUCTION

Procedural synthesis of sign language is a technique for creating the animation of a signing avatar based on a list of low-level motion constraints to be evaluated and synthesized on the avatar. The AZee model [1] allows us to synthesize a multi-track animation timeline specifying all parts of the utterance to render with the avatar [2]. This allows customisation and creation of new Sign Language content without requiring pre-animated data for the corresponding elements in the description.

One of the challenges in procedural synthesis is generating correct shape configurations inside the avatar's motion space. Using Inverse Kinematics(IK) or Forward Kinematics(FK) to calculate the joint rotations for shape configuration often leads to poses that either look unnatural or are incorrect since the description doesn't generalize to all avatars. Additionally, writing low-level descriptions to represent these shapes is often difficult for a linguist.

To address these issues, we propose a methodology for defining new sets of morphs that can be used to synthesize hand shapes and facial expressions, improving the synthesis of sign language gestures using the AZee model. Morph target animation, a computer animation technique that involves blending between different pre-defined shapes or *morph targets*, can be used to overcome the problems with just kinematic-based models as it only requires a single pre-defined morph to obtain intermediate poses. We aim to embed these morphs into our animation system, map them to their poses on our avatar, and add them to our existing Blender add-on for simpler, faster, and better shape synthesis.

2. RELATED WORK

We begin our discussion by first reviewing procedural Sign Language synthesis. Then we explore the use of morph target animation methods in computer graphics. Finally, we discuss the use of morph targets in signing avatars.

2.1. Procedural Sign Language Synthesis

Procedural Sign Language synthesis is an emerging area of research that aims to generate sign language animations from linguistic descriptions. The idea is to generate animations on the avatar directly from the description without the need for pre-recorded motion data. This has several applications, especially when the goal is to minimize the amount of pre-recorded motion data.

To do this, a set of constraints (provided by the linguistic description) act on the posture for a certain time [3]. The animation is generated by applying these constraints to the avatar's anatomy. To do this, constraints are evaluated using IK and FK techniques to generate postures. Finally, interpolations are applied to generate motion.

For example, we can define the shape of a closed index finger in AZee as follows,

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).

```
orient
  'dir
  !index
    ^side
   2
  'along
  oppvect
    dir
      !palm
        ^side
```

The above expression instructs the posture to close the index finger of *side* by orienting the index FK bone in the direction opposite to the palm.

Since the technique uses the skeleton space of the avatar, it often results in unnatural synthesized shapes, as shown in figure 1. Furthermore, with complex expressions, it is difficult for a linguist to write such code for a variety of hand shapes.

Fig. 1: Unnatural shape synthesis of the hand shapes using the kinematic model to generate hand shapes with AZee

2.2. Morph Target Animation in Signing Avatars

Morph target animation is a type of computer animation technique that involves changing the shape of a 3D model by blending between different pre-defined shapes or *morph targets*. Morph target animation is especially useful in character animation scenarios when the goal is to have more control over the movements because it allows us to key-frame the composite geometry of the mesh. A popular application of morph targets is facial animation [4].

A set of morph targets is required to synthesize facial animations [5] and was used to further extend the JASigning system [6]. Similarly, the EMBR system [7], the Paula animation system [8] [9], and the SIGNCom system [10] use morphs to synthesize facial animations as well.

Along with facial animation, morphing can also be used as a Pose Space Deformation(PSD) [11]. A PSD is a hybrid method that combines skeleton space deformations(SSD) with morphing and employs scattered data interpolation to compute non-linear skin corrections in pose space. This gives us a kinematic model that also has poses which can be pre-sculpted by artists.

A similar approach to control deformable material using Dynamic Morph Targets derived from these PSDs was used by [12].

Thus, defining pose space for specific skeleton joints allows us to use morph targets for not just the geometry of the mesh but also the skeletal representation of the avatar.

3. METHODOLOGY

The AZee language tentatively defines morph constraints as a morph target with some weight. Hence, implementing a set of motion space using morphs is appealing as it simplifies the work of the linguist. However, AZee doesn't have definitions of the motion space, which could be covered with these morphs. The syntax of a morph expression can be defined as follows:

```
morph
    'morph_id
    weight[0, 1]
```

where $morph\_id$ is the name of the morph with which the language is extended, and $weight[0, 1]$ represents the amount of weight applied to the given morph.

Additionally, in previous works, a *morph* was referred to as a non-skeletal articulatory constraint. Here, we redefine morph as a constraint which can constrain both the skeleton and the mesh. The target motion space of a morph can be mapped to a single parameter, its weight. This makes morph a local constraint on the avatar.

3.1. Skeletal morphs

In this section, we discuss the parts of morph motion space which depend on the skeleton.

3.1.1. Adduction and Abduction

An *Adduction* refers to the movement of a limb or other part towards the mid-line of the body. On the contrary, an *Abduction* is a limb's movement away from the body's mid-line. Figure 2 shows the adduction and abduction of the palm.

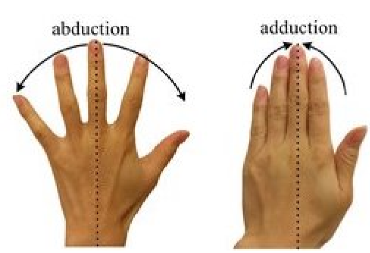

Fig. 2: Adduction and Abduction of the palm [13]

Although adduction and abduction generalise to various joints, we are concerned only by the scenarios representing a local movement.

3.1.2. Extension and Flexion

Extension refers to an extension or bending movement of a joint that increases the angle between two bones. *Hyper-extension* refers to an extension beyond its normal range of motion, typically in a backward direction. Lastly, *Flexion* is the opposite of Extension and refers to a bending movement of a joint that decreases the angle between two bones. Figure 3 shows the index finger's extension, hyper-extension, and flexion.

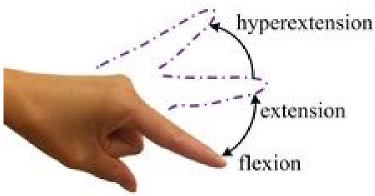

Fig. 3: Hyper-extension, Extension and Flexion of the index finger [13]

Just like adduction and abduction, we are only concerned by scenarios with local hyper-extension and flexion of the joints.

3.2. Non-Skeletal morphs

Apart from the skeletal morphs, we also want to use morphs for detailing facial characteristics. For this, we use a subset of the FACS model [14] for our facial morphs. We chose the model for its simplicity and as a baseline and aim to have a bigger set in the future.

We choose not to include the action units which correspond to global dependencies. These include Action Units(AUs) 51 to 60, which correspond to head movements AUs 61 to 69, which correspond to eye movements.

Fig. 4: AU 61(eye turn left) not implemented as morph since we already have *look* and *lookat* constraints in AZee

Fig. 5: Facial expression in rule *:inter-subjectivity* by combining AU18 and AU22

morph_id	Movement
I_closed	Hyperextension and Flexion of index fingers
M_closed	Hyperextension and Flexion of middle fingers
R_closed	Hyperextension and Flexion of ring fingers
L_closed	Hyperextension and Flexion of little fingers
T_closed	Hyperextension and Flexion of thumbs
palm_extended	Adduction and Abduction of the palms

Table 1: First Set of AZee Morphs

4. RESULTS

Based on the above methodology, we get a new list of morphs and the respective skeletal movements as shown in table 1, which can be used to extend the low-level constraints of AZee. We use this list to redefine the AZee vocabulary of morph IDs. Which can be used to create poses which can be synced independently.

We also get another set of facial morphs based on FACS, which can be combined for facial expressions such as in *:inter-subjectivity*(figure 5).

5. IMPLEMENTATION

We implement our set of morphs as shape keys in blender [15]. We also use FACSHuman [16] to extract the relevant set of FACS shape keys for our MakeHuman [17] based blender avatar. To use this shape key dataset, we map the AZee morph definitions with our defined shape key names in a *skeleton.morphmap* which is initialized with our avatar posture.

We extend our AZee animator add-on in Blender for

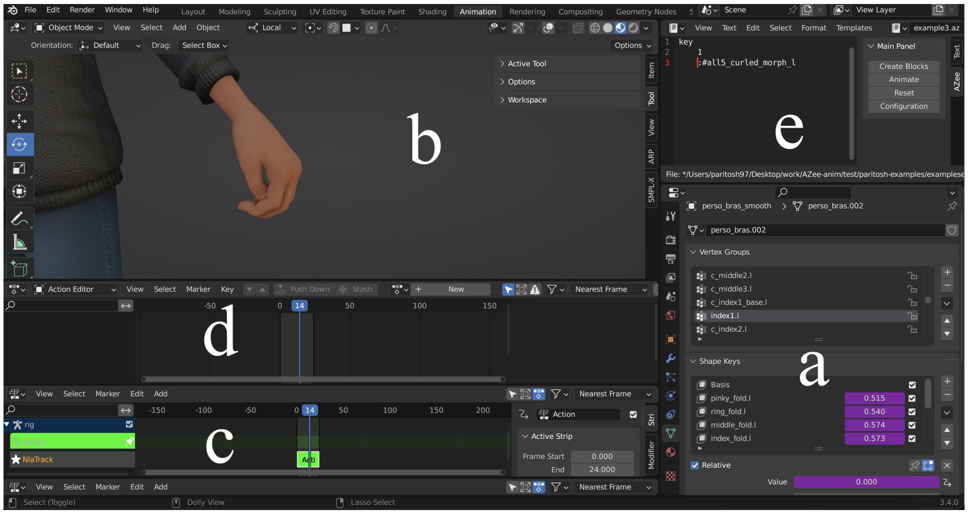

Fig. 6: Blender interface. (a) Shape Key properties (b) 3D Viewport (c) Non-linear Editor (d) Action Editor (e) AZee editor

morph support. Figure 6 shows the Blender interface configured with the new shape keys. Its main components include:

(a) Shape Key properties
Modify, add and debug shape keys

(b) Viewport
Shows the 3D scene with the avatar.

(c) Non-linear Editor
To place all the animated blocks from the utterance.

(d) Action Editor
Allows us to modify and visualize the generated actions as well as the pre-recorded animations.

(e) AZee editor
An editor to write AZee expressions and change additional settings.

5.1. Outcome

Our blender add-on can be used with AZee morphs. We observe significant improvements in the synthesis of hand shapes compared to the previous approach, which can be seen in figure 7. Figure 5 shows the facial expression for the rule *:inter-subjectivity*, which was synthesized using the morphs linked to the FACSHuman extracted shape keys.

6. CONCLUSION AND FUTURE WORK

We presented a methodology to extend the low-level of AZee with a new set of morphs. This allowed us to map low-dimensional pose space deformations to AZee morphs and produces better shapes, is easier to pose for the linguist and is faster at run-time since it is based on pre-recorded animation data.

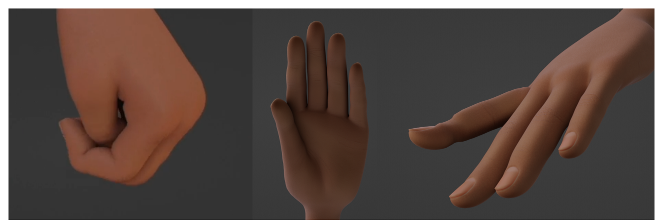

Fig. 7: Improvements in the shape synthesis compared to the kinematic approach in figure 1

Our system still has some limitations which we want to address in the future:

Larger Use Cases For now, we use morphs for defining hand shapes and facial expressions only. I would be interesting to study more use cases such as for spine-extension, head movements, etcetra.

Low Coverage We want to improve our facial animation system to have a larger coverage like the FLAME model [18] or Paula [8].

Naturalness Our morph interpolations look robotic. One way to improve this could be to modify the bezier handles of the underlying f-curves between morph-based poses based on the morph.

Universality We introduce an additional step i.e. creation and mapping of shape keys which could get cumbersome when re-targeting because the same shape keys may not work on different types of avatars. A potential solution to this could be to explore other models for avatars like the SMPL model [19].

7. REFERENCES

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

[2] Paritosh Sharma and Michael Filhol, "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)*, Marseille, France, June 2022.

[3] Alexis Heloir and Michael Kipp, "Embr–a realtime animation engine for interactive embodied agents," in *Intelligent Virtual Agents: 9th International Conference, IVA 2009 Amsterdam, The Netherlands, September 14-16, 2009 Proceedings 9*. Springer, 2009, pp. 393–404.

[4] Zhigang Deng and Junyong Noh, "Computer facial animation: A survey," *Data-driven 3D facial animation*, pp. 1–28, 2008.

[5] Vince Jennings, Ralph Elliott, Richard Kennaway, and John Glauert, "Requirements for a signing avatar," in *Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Philippe Dreuw, Eleni Efthimiou, Thomas Hanke, Trevor Johnston, Gregorio Martínez Ruiz, and Adam Schembri, Eds., Valletta, Malta, May 2010, pp. 133–136, European Language Resources Association (ELRA).

[6] Ralph Elliott, F. Bueno, Richard Kennaway, and John Glauert, "Towards the integration of synthetic sl animation with avatars into corpus annotation tools," 01 2010.

[7] Alexis Heloir and Michael Kipp, "Real-time animation of interactive agents: Specification and realization," *Applied Artificial Intelligence*, vol. 24, no. 6, pp. 510–529, 2010.

[8] John Mcdonald, Rosalee Wolfe, Julie Hochgesang, Diana Jamrozik, Marie Stumbo, Larwan Berke, Melissa Bialek, and Farah Thomas, "An automated technique for real-time production of lifelike animations of american sign language," *Universal Access in the Information Society*, vol. 15, 05 2015.

[9] Ronan Johnson, "Towards enhanced visual clarity of sign language avatars through recreation of fine facial detail," *Machine Translation*, vol. 35, no. 3, pp. 431–445, sep 2021.

[10] Sylvie Gibet, Nicolas Courty, Kyle Duarte, and Thibaut Le Naour, "The signcom system for data-driven animation of interactive virtual signers: Methodology and evaluation," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 1, no. 1, pp. 1–23, 2011.

[11] J. P. Lewis, Matt Cordner, and Nickson Fong, "Pose space deformation: A unified approach to shape interpolation and skeleton-driven deformation," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, USA, 2000, SIGGRAPH '00, p. 165–172, ACM Press/Addison-Wesley Publishing Co.

[12] Nico Galoppo, Miguel A. Otaduy, William Moss, Jason Sewall, Sean Curtis, and Ming C. Lin, "Controlling deformable material with dynamic morph targets," in *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games*, New York, NY, USA, 2009, I3D '09, p. 39–47, Association for Computing Machinery.

[13] Lefan Wang, T. Meydan, and Paul Williams, "A two-axis goniometric sensor for tracking finger motion," *Sensors*, vol. 17, 04 2017.

[14] Paul Ekman and Wallace V Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[15] Blender Online Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[16] Michaël Gilbert, Samuel Demarchi, and Isabel Urdapilleta, "Facshuman, a software program for creating experimental material by modeling 3d facial expressions," *Behavior Research Methods*, vol. 53, no. 5, pp. 2252–2272, 2021.

[17] MakeHuman Online Community, *MakeHuman*, MakeHuman Community, 2023.

[18] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero, "Learning a model of facial shape and expression from 4D scans," *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, vol. 36, no. 6, pp. 194:1–194:17, 2017.

[19] Sergey Prokudin, Michael J Black, and Javier Romero, "Smplpix: Neural avatars from 3d human models," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1810–1819.