Intermediate Block Generation for Multi-Track Sign Language Synthesis

Paritosh Sharma, Michael Filhol

Intermediate Block Generation for Multi-Track Sign Language Synthesis

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

Figure 1: Example results for the AZee expression presented. the first two poses are snapshots of *TOPIC* and *INFO,* while the following represent the intermediate pose and the evaluated motion curves

ABSTRACT

Generating realistic Sign Language using signing avatars is a challenging task that typically involves synthesis using either procedural or pre-animated techniques like motion capture or artistic editing of signs. However, combining these two approaches is difficult. In this work, we propose a novel method for generating intermediate poses in a multi-track representation of a sign language discourse. The proposed method uses procedural generation with artistic techniques to prioritize certain aspects of the generated poses while sacrificing others to improve the overall consistency of the representation. The system is implemented as an add-on in Blender, an open-source 3D toolkit.

CCS CONCEPTS

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

KEYWORDS

sign language, animation, motion retargeting

ACM Reference Format:

Paritosh Sharma and Michael Filhol. 2023. Intermediate Block Generation for Multi-Track Sign Language Synthesis. In *The ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA '23), August 04–06, 2023, Los Angeles, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3606037.3606837

SCA '23, August 04–06, 2023, Los Angeles, CA, USA

ACM ISBN 979-8-4007-0268-6/23/08.
https://doi.org/10.1145/3606037.3606837

1 INTRODUCTION

The field of simulating realistic Sign Language using signing avatars has been a topic of growing interest in recent years. Procedural generation techniques are commonly used on a skeletal representation[Kipp et al. 2011] of the avatar to provide broad coverage of synthesized signs, while artistic and motion capture techniques are employed to achieve a more natural coverage of signs, albeit providing a smaller coverage. While both techniques have shown promising results individually, their combination requires careful tuning to produce accurate and realistic sign language generation.

Existing works for Sign Language synthesis use AZee[Filhol et al. 2014] - a model which facilitates writing expressions to represent Sign Language discourses and determine the forms to be articulated on a multi-track timeline which automatically maps directly to a non-linear editor[Sharma and Filhol 2022][Dauriac et al. 2022]. In this work, we propose a novel method to effectively generate intermediate blocks which contain interpolation information. To do this, we filter the most important components of the poses in the multi-track representation and generate intermediate poses which preserve these components while allowing for some deviation for naturalness. The algorithm is based on the use of artistically created motion templates. This allows us to generate motion curves that look more natural and accurately reflect the intended meaning of the respective track while also preserving the original constraints.

2 OVERVIEW

Filhol et al.[Filhol et al. 2017] proposed a fundamental guiding principle for animating AZee, according to which the coarser the basic animation blocks, the more natural the final animation. To understand this better, let's consider the following AZee expression representing signed utterance meaning "A cat is cute" or "Cats are cute".

```
:info-about
    'topic
    :cat
    'info
    :cute
```

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

2.1 Block Generation

We match each block with a motion template during generation. A block placed on the timeline can either be a *pre-animated block* with re-targeted animation data (for this example, let's assume "cat") or a *minimally constrained block* with several constraints to be satisfied on the avatar posture(example "cute").

Thus, figure 2 represents the corresponding block structure for the above expression. The blocks can be summarized as follows:

- **TOPIC**: Re-targeted sign for "cat".
- **INFO**: Contains more blocks with posture constraints for the sign "cute" such as HAND_PLACE_TRILL.INFO for placing the hand and trilling it on an axis and HAND_CONFIG.INFO for creating hand configuration.
- **HOLD1**: Constraints to hold the ending of the sign "cat".
- **HOLD2**: Constraints to hold the ending of the sign "cute".
- **BLINK**: Constraints for eye-blink.

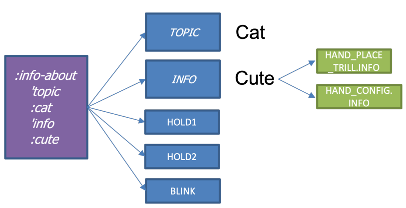

Figure 2: Blocks generated from the example AZee expression

A coarser animation block such as **TOPIC** will generate more natural animation than **INFO** which is a composite block made up of **HAND_PLACE_TRILL.INFO** and **HAND_CONFIG.INFO**.

2.2 Intermediate Block Generation

After baking these minimally constrained and pre-animated blocks on the timeline, intermediate blocks are generated between pairs of blocks by mapping the f-curves to each other, as shown in figure 3. The resultant f-curves are evaluated based on the motion template.

Figure 3: Intermediate blocks generated for blocks of *TOPIC* and *INFO*

2.3 Motion Template

A motion template informs how the f-curves in the intermediate blocks should be baked for pairs of blocks. Given a set of blocks $B = B_1, B_2, ..., B_n$ representing either poses or motions on a multi-track timeline, the motion template is used to generate f-curves inside the intermediate blocks that represent the interpolation between any two blocks B_i and B_j, where $1 \leq i < j \leq n$. This can be achieved using a function $f(B_i, B_j)$ that accepts two input blocks and generates an intermediate block between them. The function can be implemented using various interpolation techniques, such as linear, spline, etcetera. The specific choice of interpolation technique depends on the artist.

3 IMPLEMENTATION AND RESULTS

Our synthesis model is implemented in Blender as an add-on. We also have a library of re-targeted signs for synthesizing our pre-animated blocks as well as skeletal and facial[Ekman and Friesen 1978] morphs, which are used along with an IK solver for synthesizing the minimally constrained blocks.

We extend this synthesis model by adding motion templates and extend our algorithm by adding template checks for these motion templates. Figure 1 shows intermediate motion curves for the above example.

In future, we aim to improve our synthesis model by addressing the following areas:

- Improvements in the avatar model using SMPL [Loper et al. 2015] avatars and FLAME [Li et al. 2017] facial model.
- Creation of motion templates from motion capture.

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 Com- panies, LISN — University Paris-Saclay, and Gipsa-Lab — Grenoble Alpes University).

REFERENCES

Boris Dauriac, Annelies Braffort, and Elise Bertin-Lemée. 2022. Example-based Multilinear Sign Language Generation from a Hierarchical Representation. In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives*. 21–28.

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

Michael Filhol, Mohamed Hadjadj, and Annick Choisier. 2014. Non-manual features: the right to indifference. In *International Conference on Language Resources and Evaluation*. Reykjavik, Iceland. https://hal.archives-ouvertes.fr/hal-01849040

Michael Filhol, John Mcdonald, and Rosalee Wolfe. 2017. Synthesizing Sign Language by Connecting Linguistically Structured Descriptions to a Multi-track Animation System. 27–40. https://doi.org/10.1007/978-3-319-58703-5_3

Michael Kipp, Alexis Heloir, and Quan Nguyen. 2011. Sign language avatars: Animation and comprehensibility. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*. Springer, 113–126.

Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. 2017. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)* 36, 6 (2017), 194:1–194:17. https://doi.org/10.1145/3130800.3130813

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

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)*. Marseille, France. https://hal.archives-ouvertes.fr/hal-03721720