PhD in Computer Graphics / Animation

Editing 3D Keyframed Animation via Gesture & Vocal Inputs

Place : Campus de l'Ecole Polytechnique, Palaiseau

Employer : Ecole Polytechnique (*Taking place at LIX – Laboratoire d'Informatique de l'Ecole Polytechnique)*

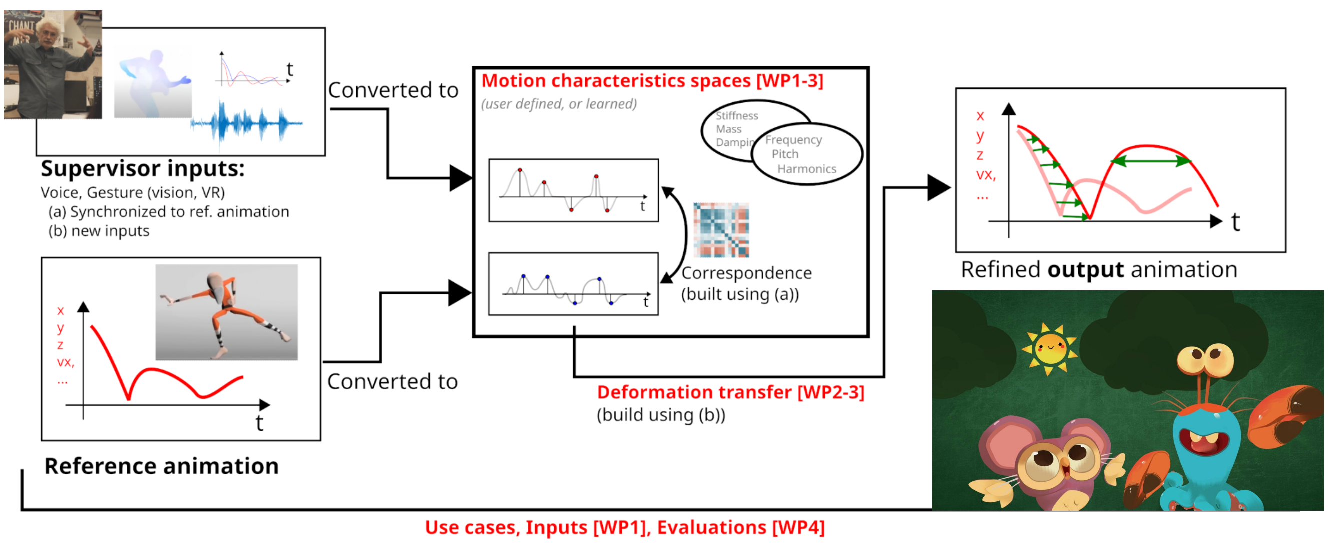
PhD start: October 2024 or before (flexible)

Supervisors: Damien Rohmer (damien.rohmer@polytechnique.edu), LIX

With possible co-supervision with Vicky Kalogeiton (LIX), Pascal Guehl (LIX), Marc Christie (IRIT).

Work in collaboration with the Animation Studio: Dada ! Animation in Paris - www.dada-animation.com

URL: https://www.lix.polytechnique.fr/vista/job/index.html



Expressing a modification of a virtual character via gesture and sound.

Context:

Over the past years, the demand for high-quality 3D virtual content has been constantly increasing, and will further grow with the generalization of distant interaction in our society, the rise of digital doubles in evermore immersive environments, and the emergence of AI-driven content generation techniques such as stable diffusion or motion diffusion. One of the main bottlenecks faced by Computer Graphics (CG) industries in this context is the animation of non-realistic (*cartoony*) 3D shapes and characters: despite the recent emergence of AI-fueled motion generation techniques, a minute of animation usually requires between 6 (for basic TV level) to 60 days (for high-end cinema shows) of work by several skilled animators.

Quite noticeably, humans have a remarkable and natural ability to mimic and **communicate motions at a high level** through elaborate body gestures and/or vocal expressions. This ability is intuitive; for instance, a choreographer, like a real-time conductor with his musicians, is able to direct dancers through gestures and sounds. Interestingly, such expressive forms of inputs have never been fully and simultaneously exploited to conduct existing animated content, yet they display a considerable potential in improving the naturalness of the interaction between the user and the machine, especially when considering complex spatial and temporal editing tasks. The key scientific question stated in this proposal is therefore whether it is possible to computationally interpret these multimodal high-level signals used by animators, which are ambiguous by nature, but yield precise and timely intentions, and convert them into changes on existing animations via modifications of rythms, amplitudes, accelerations, variability; thus offering an innovative and natural way of authoring animated contents.

Objective and Methodology:

The fundamental idea of this PhD is propose an **Intuitive editing for dynamic physics-inspired phenomenon**. To this end, we propose to (i) express simultaneous multimodal inputs as high-level animation principles into a motion characteristics space, (ii) exploit spatial and temporal characteristics of the input signals to edit existing animations using learning techniques inspired by style transfer, (iii) combine the style transfer techniques with authoring constraints such as physics-based and (iv) co-design interactive tools with creative artists to exploit them in industrial pipelines.

The methodology followed in this PhD will first consider the fitting and motion transfer via physical parameter analogy to an animated shape undergoing a global deformation, which will be extended to hierarchical skeletal-based character with a specific focus on local deformation propagation along the hierarchy as well as time synchronicity between the subparts. Finally, we will study the case of the more general natural-like phenomenon with a larger number of degrees of freedom in relying on the expressivity of multimodal inputs.

A first possibility is to explore explicit correspondence between input signal and animation curve. We plan to build our analysis on top of Nivaggioli et al.'s work [2019]. In particular, we will study the link between the low frequency of the voice signal envelope obtained, for instance, from its Hilbert transform as well as its variation along time with physical parameters of an elastic model. For more complex models, the frequency of their oscillations may depend on their geometry. Relying on a modal decomposition of eigenmodes, also known as "natural vibration" [Kry2009], provides an interesting tradeoff between physical-related spatial and temporal frequencies while offering a low dimensionality compatible with user control. A second objective will focus on the time synchronization between hierarchical elements whose motions are set one after the other, before tackling the problem of multi-scale natural phenomena such as explosion of fluid-based models.

A second possibility is to design a motion-characteristics latent space that is able to encode the spatial and temporal characteristics of the multimodal inputs. For each given input motion, we will propose to compare the temporal and spatial correlations within the classical features (rhythm, amplitude acceleration, amplitude) in order to compute (i) the amount of motion that needs to be added to the animation curves and (ii) the timing of these motions. Previous techniques have relied on CCA (Canonical Correlation Analysis) to discover spatial mappings, yet are limited to linear transforms [Dontcheva03]. We propose to rely on a multi-stage approach. The first stage will rely on a differential correlation model to measure and understand the temporally aligned spatial differences between input characteristics and the animation curves at different levels of the animation hierarchy. In a second stage, we will rely on flow networks, recently used with success for improving temporal aspects in style transfer tasks [Huang22] to transfer the motion characteristics. This latent motion characteristics space will be designed in two iterations, first by only considering single input modalities (gestures/voice/videos), and second by considering combined input modalities. Cross-attention networks have been successful in accounting for different modalities [Wei20] and will be exploited

Requirements:

- Master level student, or last year of Engineering School, with good Computer Science and Applied Math background.
- Followed class or performed projects in Computer Graphics and/or 3D geometry.
    - Specific knowledge in Computer Animation is a plus
    - Specific knowledge in Signal Processing and Audio Processing is a plus
- Good practical skills in programming typically in Python and/or C++. Being able to autonomously develop code interfaced if needed Blender or Unity plugins, and parsing 3D animated assets.
- Interest in Animation Studio Production. Note that 3D art knowledge or use of 3D modeler is not required, but it can be a plus.

How to Apply:

Send an email to Damien Rohmer (damien.rohmer@polytechnique) with the following elements:

- Your CV
- The school transcript obtained so far in master or engineering school.
- Please mention explicitly in your email:
    - The reason for your interest in the subject
    - The class and/or projects you have done in Computer Graphics
    - The name of the teachers/supervisors you had in Computer Graphics

References

[**Arora19**] R. Arora, R. Habib Kazi, D. Kaufman, W. Li, K. Singh. MagicalHands: Mid-Air Hand Gestures for Animating in VR. UIST 2019.

[**Ciccone17**] L. Ciccone, M. Guay, M. Nitti, R. Sumner. Authoring Motion Cycles. SCA 2017.

[**Dong21**] R. Dong, Y. Lin, Q. Chang, J. Zhong, D. Chai, S. Ikuno. Motion Feature Extraction and Stylization for Character Animation using Hilbert-Huang Transform. ACM ICEA, 2021.

[**Dontcheva03**] M. Dontcheva, G. Yngve, Z. Popovic. Layered Acting For Character Animation. ACM SIGGRAPH 2003.

[**Fender15**] J. Chen, S. Izadi, A. Ditzgibbon. KinEtre: Animating the World with the Human Body. UIST 2012.

[**Garcia19**] M. Garcia, R. Ronfard, **M.-P. Cani.** Spatial Motion Doodles: Sketching Animation in VR Using Hand Gestures and Laban Motion Analysis. MIG 2019.

[**Guay15**] Guay, M., Ronfard, R., Gleicher, M., & **Cani, M. P.** (2015). Space-time sketching of character animation. ACM Transactions on Graphics (TOG), 34(4), 1-10.

[**Huang22**] Huang, Y., Mo, H., Liang, X., & Gao, C. (2022, July). Unpaired Motion Style Transfer with Motion-Oriented Projection Flow Network. IEEE ICME 2022.

[**Kim03**] T.-H. Kim, S. Park, S. Shin. Rhythmic-motion synthesis based on motion-beat analysis. ACM SIGGRAPH 2003.

[**Kry09**] P. Kry, L. Revéret, F. Faure, M.-P. Cani. Modal locomotion: Animating virtual characters with natural vibrations. EUROGRAPHICS 2009.

[**Li21**] R. Li, S. Yang, D. A. Ross, A. Kanazawa. AI Choreographer: Music Conditioned 3D Dance Generation with AIST++. ICCV 2021.

[**Nivaggioli19**] Adrien Nivaggioli, Damien Rohmer. Animation Synthesis Triggered by Vocal Mimics. Motion, Interaction and Games, 2019.

[**Wei20**] Wei, X., Zhang, T., Li, Y., Zhang, Y., & Wu, F. (2020). Multi-modality cross attention network for image and sentence matching. CVPR (pp. 10941-10950).

[**Zhi23**] Y. Zhi, X. Cun, X. Chen, X. Shen, W. Guo, S. Huang, S. Gao. LivelySpeaker: Towards Semantic-Aware Co-Speech Gesture Generation.

[**Zhou18**] Y. Zhou, Z. Xu, C. Landreth, E. Kalogerakis, S. Maji, K. Singh. VisemeNet: Audio-Driven Animator-Centric Speech Animation. ACM SIGGRAPH 2018.