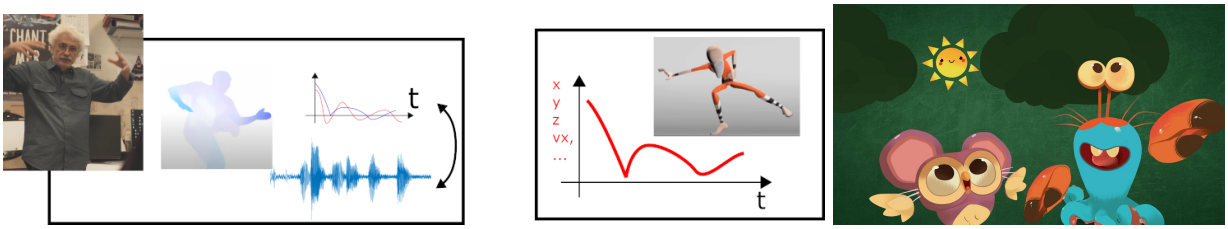
Master Internship in Computer Graphics / Animation

Editing 3D Keyframed Animation via Vocal Inputs



Expressing a modification of a virtual character via gesture and sound. Right: Example of character produced by Dada Animation

Place : Campus de l'Ecole Polytechnique, Palaiseau

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

Internship start: March 2024 (flexible)

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

Supervisors:

- Damien Rohmer (damien.rohmer@polytechnique.edu)
- Pascal Guehl (pascal.guehl@polytechnique.edu)

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

Expected continuation in PhD*.*

Context of the Project:

Creating and tuning 3D animation composed of virtual characters and/or animated elements is a major task of production studios. 3D animators typically use standard tools such as Maya or Blender to specify the so-called *animation curves*. These curves are built by interpolating positional and orientation coordinates of shape controllers called "rig" (e.g. joints of an animation skeleton), and passing through specific values called key-frames.

Once an animation is created, a process of review is performed by the artist itself or its supervisor, to improve iteratively this animation. This review is often performed with our human expressivity, namely our gestures and voices, that needs to be interpreted by the animator as modifications to be applied to the animation curves. This manual edition of these curves allows very precise modifications but remains time-consuming, therefore not allowing fast interactive iterations.

The ANR project AnimationConductor, in which this internship takes place, aims at creating new methodologies and tools to ease the semi-automatic edition and interaction of these animation curves via

expressive human inputs such as acquired gestures (video & VR handles), and recorded vocal sounds able to naturally convey a high-level notion of timing, frequency and magnitude. This project is a collaboration between the two research teams VISTA at LIX/Ecole Polytechnique and VIRTUS IRISA/Univ. Rennes/Inria, and the Animation Studio: Dada! Animation.

Different approaches have been proposed in research to synthesize animation from human inputs, starting from classical Motion Capture (MOCAP) providing a 1-to-1 mapping between the sensor and the character joint, using more general gestures mapping [Arora19, Garcia19], or via drawn sketches [Gay15, Ciccone17, Dvoroznak20]. These works, however focused on synthesizing from scratch a motion that matches as best as possible a given trajectory, but did not rely on an existing motion in order to suggest an appropriate space-time deformation. A few inspirational works attempt to have more general mappings between the human body and a character [Dontcheva03, Fender15], but they focused mostly on character posing rather than on animation authoring. Beyond gestures, sound inputs have also been studied in relation to character animation via natural language for lip synchronization [Zhou18], or to music for dancing characters [Kim03, Li21]. The use of general sound features for a more general animation authoring remains an uncovered research area.

Internship Objective and Proposed Methodology:

Within this internship taking place at LIX, we propose to create a preliminary methodology to edit an animation using vocal inputs expressed typically as onomatopoeic sounds. We will consider the following three inputs: (i) An initial pre-existing set of animation curves, (ii) a synchronized vocal reference paired with these animation curves, (iii) a new vocal input with different timing and sound variation. The scientific objective is to compute a set of deformed animation curves considering this new vocal input.

An initial work was proposed by Nivaggioli et al. 2019 to extract onomatopoeic sound patterns (short peak-like sound, and long-lasting sounds) and their characteristics (time interval, pitch, magnitude) from a voice recording in order to generate an animation. However, this work was limited to pre-scripted mapping between sound and animation. We propose to extend this methodology to automatically compute the correspondence between detected sound events and animation features. A first step will consist in exploring the different ways to analyze and extract space-time features from (a) the animation curve (Peaks and discontinuities, cycles, noise parameter, acceleration properties) [Dvoroznak20, Ciccone17], and (b) the sound input (Spectral analysis, wavelets, Hibert-transform, etc.) [Dong21] separately. Then in a second step, the features maximizing their correlation in the two tracks will be matched. Finally, these characteristics will be interpreted as scaling in space, and/or translation in time in order to deform the curve based on a new vocal signal.

The project will start by analyzing such mapping for simple examples – typically deformable bouncing spheres and oscillating cylinder-like primitives – before trying to generalize it to more complex shapes seen as a hierarchical merging of such simple shapes.

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 for Animation Studio Production. Note that 3D art knowledge or use of 3D modeler is not required, but it can be a plus.
- Interest in pursuing this topic in PhD (the funding is already acquired).

How to Apply:

Send an email to Damien Rohmer (damien.rohmer@polytechnique) and Pascal Guehl (guehl@lix.polytechnique.fr) 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.

[**Dvoroznak20**] M. Dvoroznak, D. Sykora, C. Curtis, B. Curless, O. Sorkine-Hornung, D. Salesin. Monster Mash: A Single-View Approach to Casual 3D Modeling and Animation. SIGGRAPH Asia 2020.

[**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.

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

[**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.

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