

JALI-Driven Expressive Facial Animation and Multilingual Speech in Cyberpunk 2077

Pif Edwards
JALI Research

Chris Landreth
JALI Research

Mateusz Popławski
CD PROJEKT RED

Robert Malinowski
CD PROJEKT RED

Sarah Watling
JALI Research

Eugene Fiume
JALI Research

Karan Singh
JALI Research

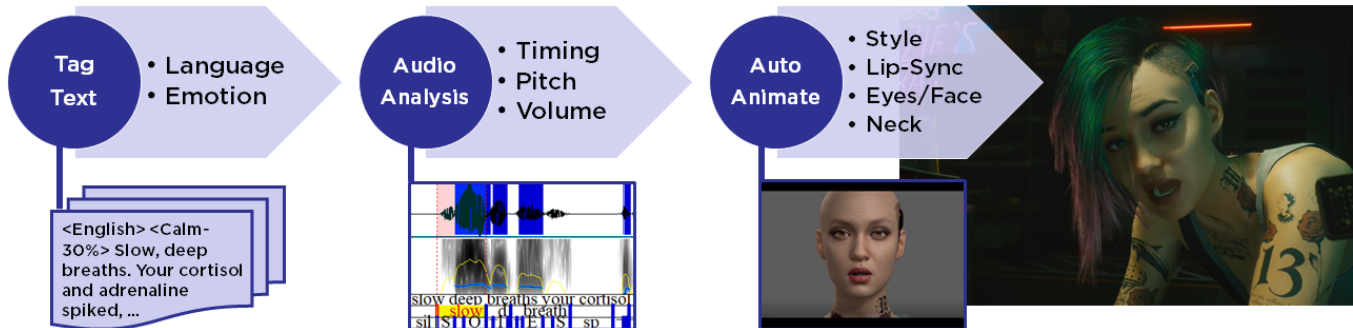


Figure 1: Cyberpunk 2077 workflow: An audio performance and tagged animation transcript is combined with hand-crafted animation to produce expressive, multi-lingual speech at an unprecedented scale.

ABSTRACT

Cyberpunk 2077 is a highly anticipated massive open-world video game, with a complex, branching narrative. This talk details new research and innovative workflow contributions, developed by JALI, toward the generation of an unprecedented number of hours of realistic, expressive speech animation in ten languages, often with multiple languages interleaved within individual sentences. The speech animation workflow is largely automatic but remains under animator control, using a combination of audio and tagged text transcripts. We use insights from anatomy, perception, and the psycho-linguistic literature to develop independent and combined language models that drive procedural animation of the mouth and paralingual (speech supportive non-verbal expression) motion of the neck, brows and eyes. Directorial tags in the speech transcript further enable the integration of performance capture driven facial emotion. The entire workflow is animator-centric, allowing efficient key-frame customization and editing of the resulting facial animation on any typical FACS-like face rig. The talk will focus equally on technical contributions and its integration and creative use within the animation pipeline of the highly anticipated AAA game title: Cyberpunk 2077.

ACM Reference Format:

Pif Edwards, Chris Landreth, Mateusz Popławski, Robert Malinowski, Sarah Watling, Eugene Fiume, and Karan Singh. 2020. JALI-Driven Expressive Facial Animation and Multilingual Speech in Cyberpunk 2077. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '20 Talks)*, August 17, 2020. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3388767.3407339>

1 INTRODUCTION AND RELATED WORK

Animating an engaging anthropomorphic face is the most challenging aspect of character animation. Humans are highly adapted to recognize and understand faces. Imperfect emulation of the subtlest facial nuance on an animated character can alarmingly drop the character into the *Uncanny Valley*, an emotional abyss where the audience loses trust and empathy with the character.

Facial animation today is produced in one of three ways.

- *Professionally key-framed animation* is high quality but laborious, often hindered by a language barrier for animators in a multi-lingual setting.
- *Performance Capture* is high quality but not editable (use or re-capture), challenging for occluded parts like the tongue, and limited by access to professional capture setups.
- *Procedural models* are often plagued by a cartoony one-to-one phoneme-viseme mapping, and poor co-articulation and paralinguals.

High-end film and game studios have traditionally relied on their most talented face actors and polyglot animators for this critical and labor-intensive task, neither of which have scaled beyond a few hundred lines of spoken content. In contrast, Cyberpunk 2077's ambitious vision of a complex narrative, driven by tens of thousands of

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGGRAPH '20 Talks, August 17, 2020, Virtual Event, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7971-7/20/08.

<https://doi.org/10.1145/3388767.3407339>

