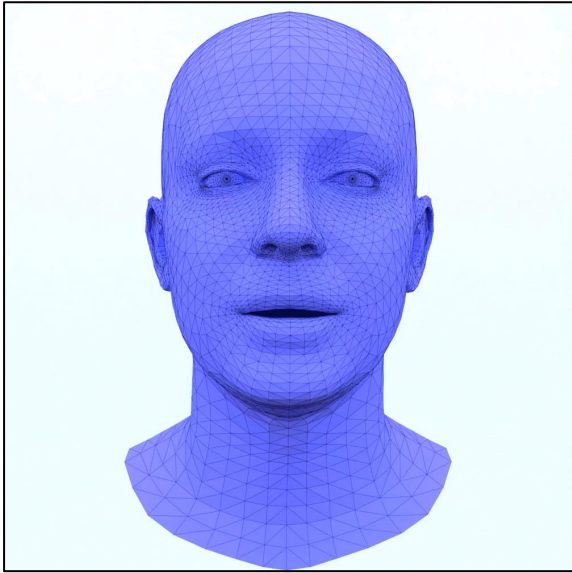
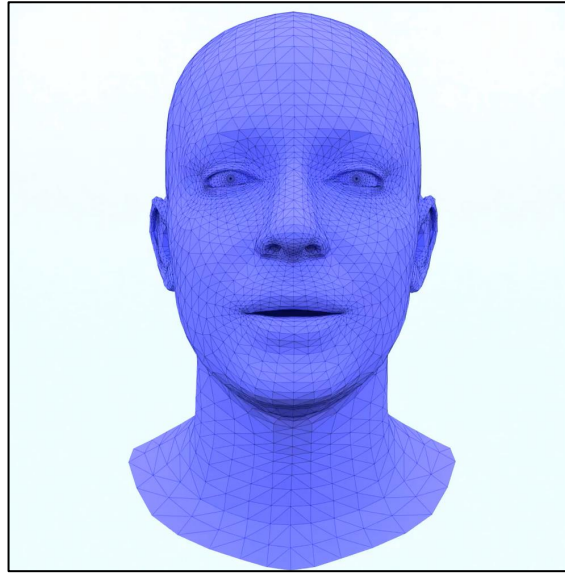


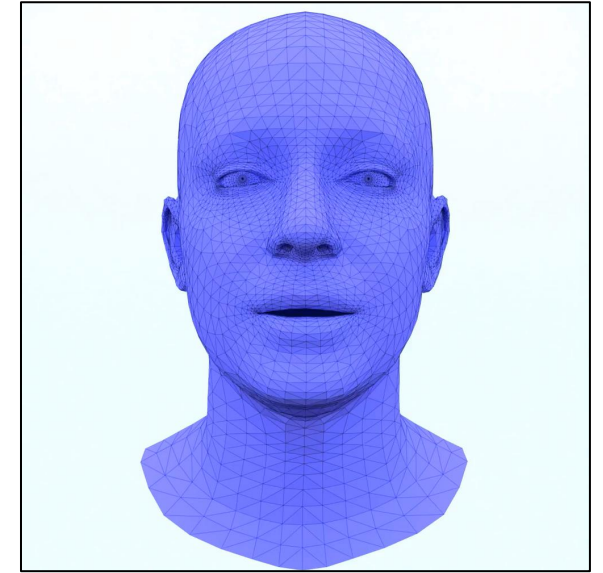
# Text Guided Generation of Head Avatars



Kissing



Sneering



Weeping

Christoph Höll  
Supervisor: Shivangi Aneja

---

Currently: expensive facial  
motion capturing

Problem: Creative restrictions

Solution:

Generate realistic 3D head &  
face animations from a textual  
input



Motion Capturing

<https://tenor.com/de/view/alita-behind-the-scenes-cgi-gif-13549391>

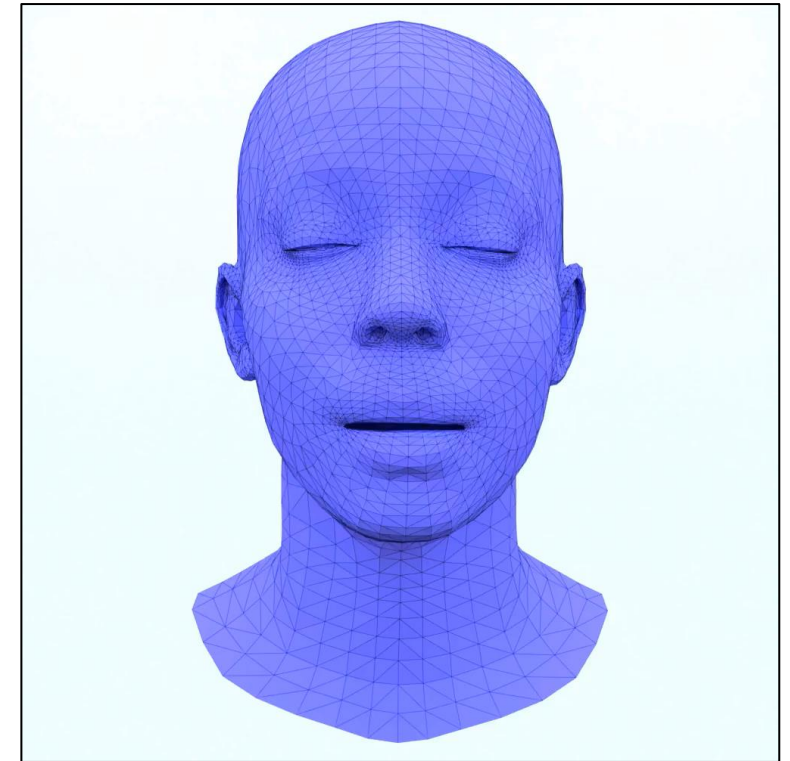
# Problem Statement

**Goal:** Generate realistic 3D head & face animations from a textual description

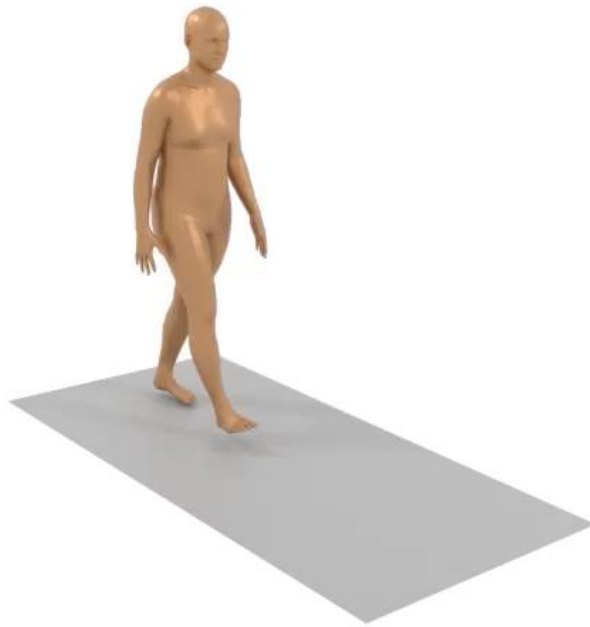
A Person is talking  
while blinking



Our Method



## Text-to-Body-Motion



HMDM

“A person walks forward, bends down to pick something up off the ground”



SMF

“He walks forward and then turns around fast and walks back”

## Audio-to-Facial-Motion



FaceFormer



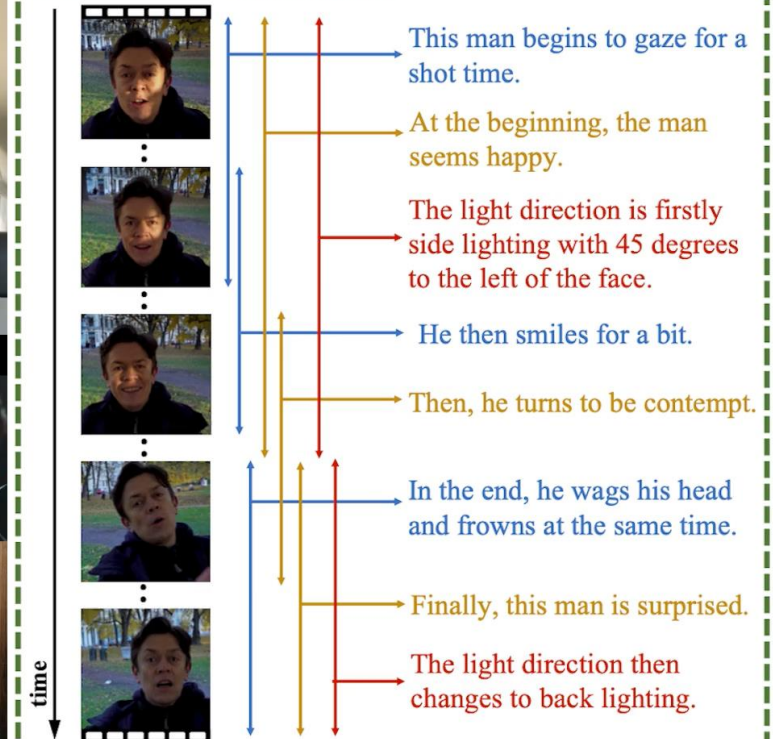
# Dataset Creation



<https://celebv-text.github.io/>

## Dynamic description:

(Action / Emotion / Light directions)



## Static description:

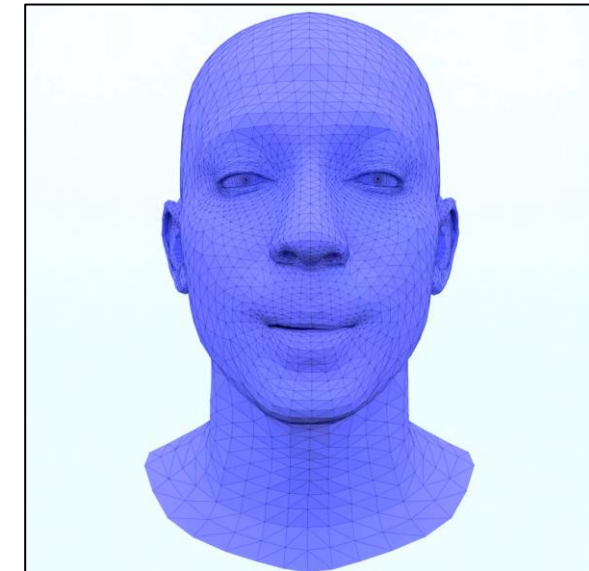
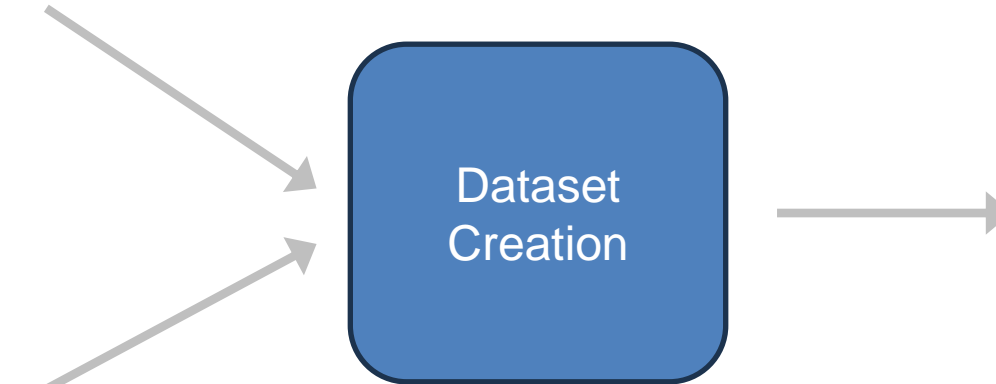
(General Appearance / Detailed Appearance / Light conditions)

The man has a short black hair with a receding hairline. He has arched eyebrows and bags under eyes. He gets a freckle on the left cheek.

The light intensity of this video is dark with a day light color temperature.



To begin with, the man talks  
for a short time







Turned



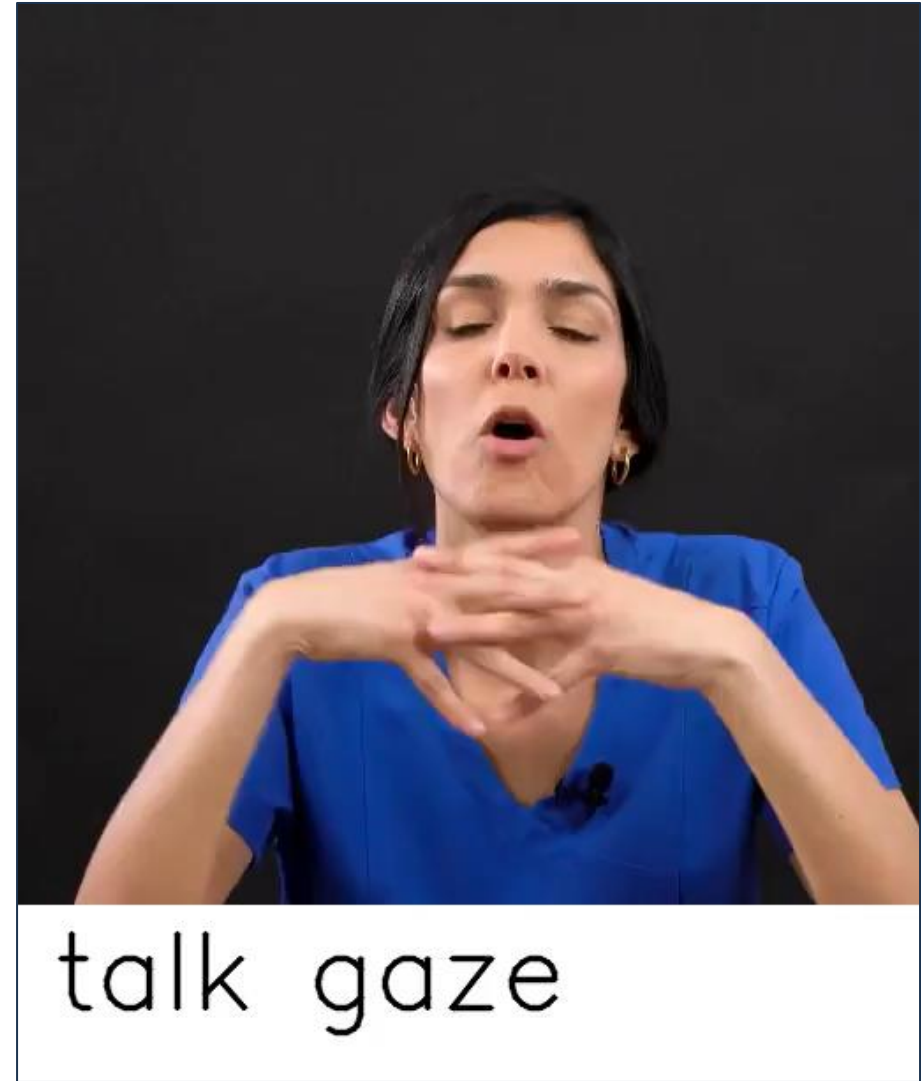
Bad Angle



Facial Occlusion

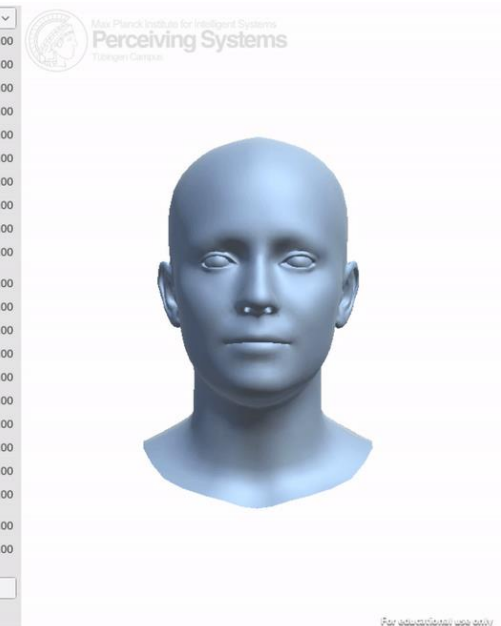
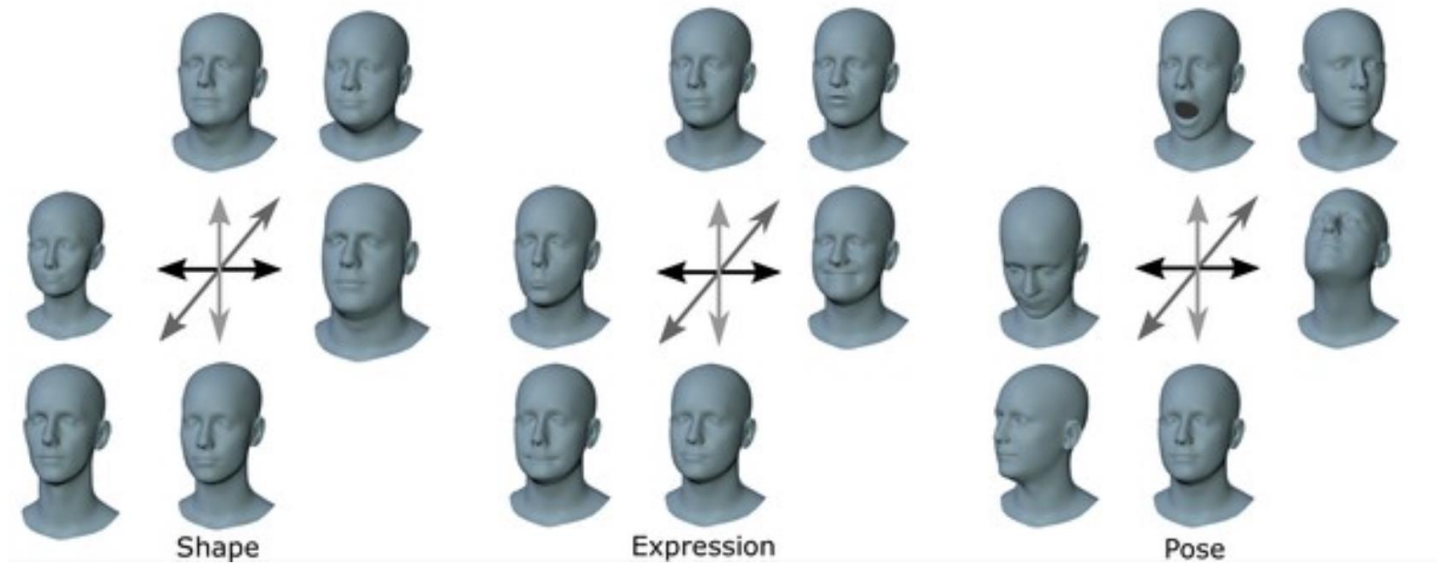
**Original:** To begin with, the woman **talks** **meanwhile gazing for a short time**, then she **frowns for a long time**, in the end she **blinks for a long time**

**New:** To begin with, the woman **talks and gazes for a short time**, then she **talks, gazes and frowns for a short time**, in the end she **talks, gazes and blinks for a long time**



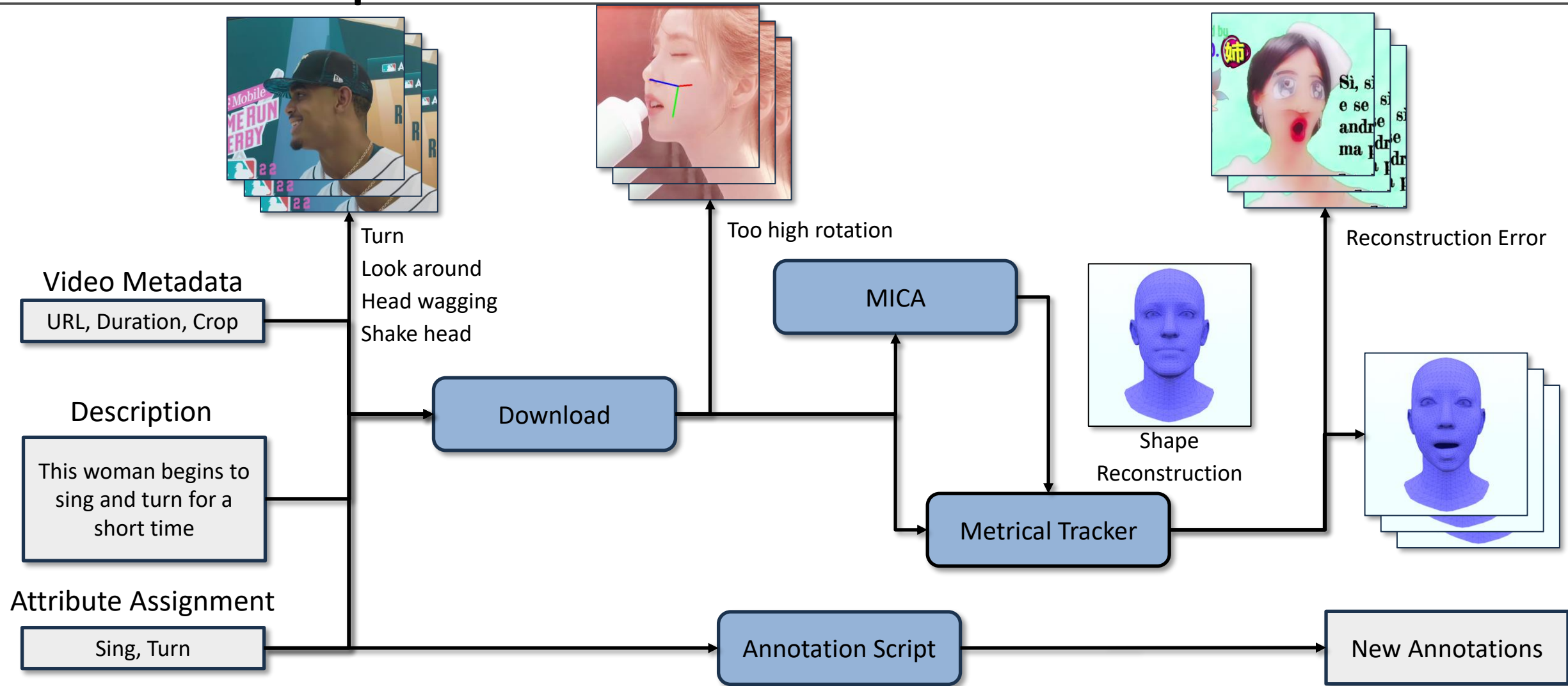


# Data Representation

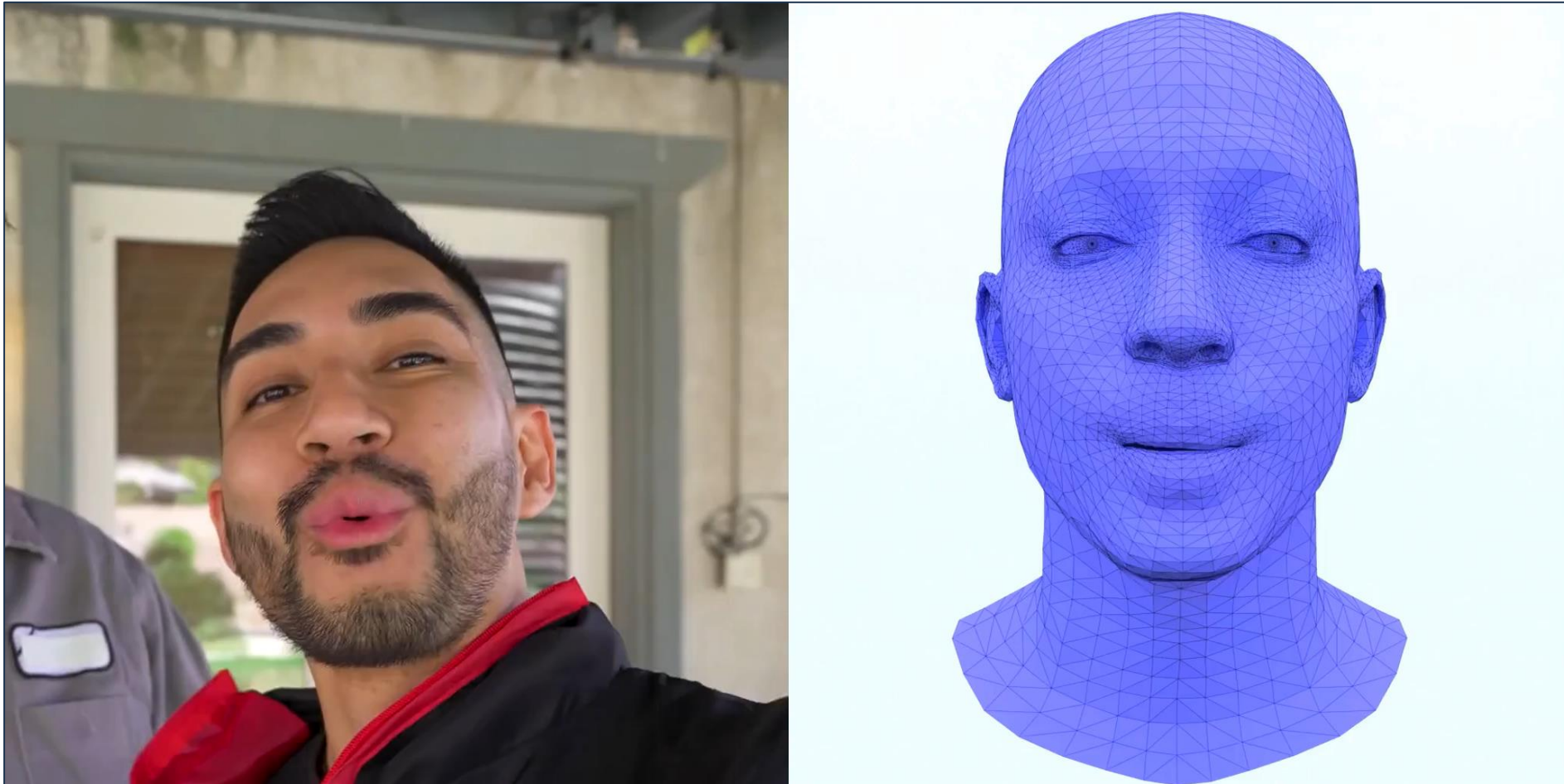


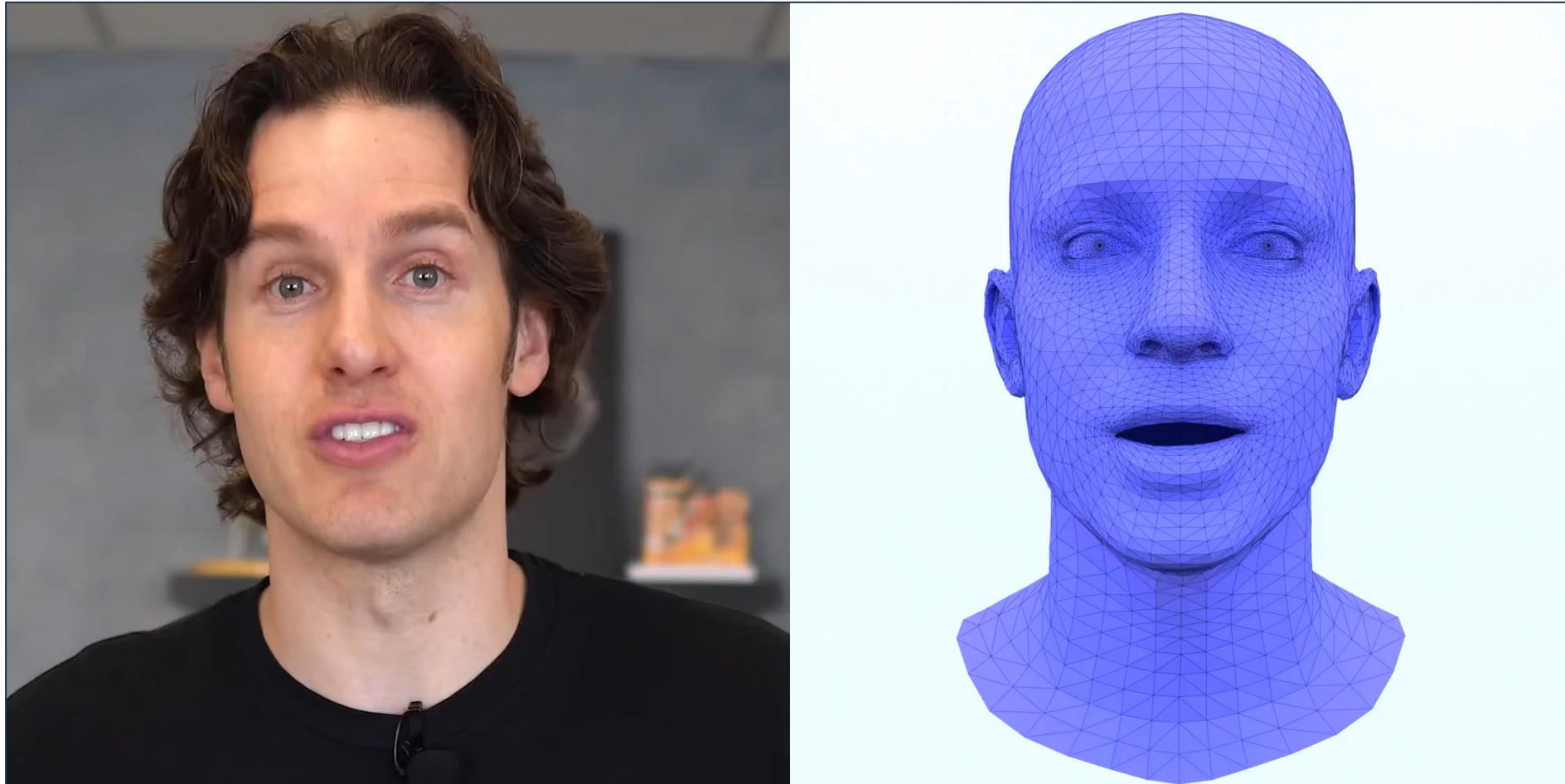
<https://flame.is.tue.mpg.de/>

# Dataset Pipeline

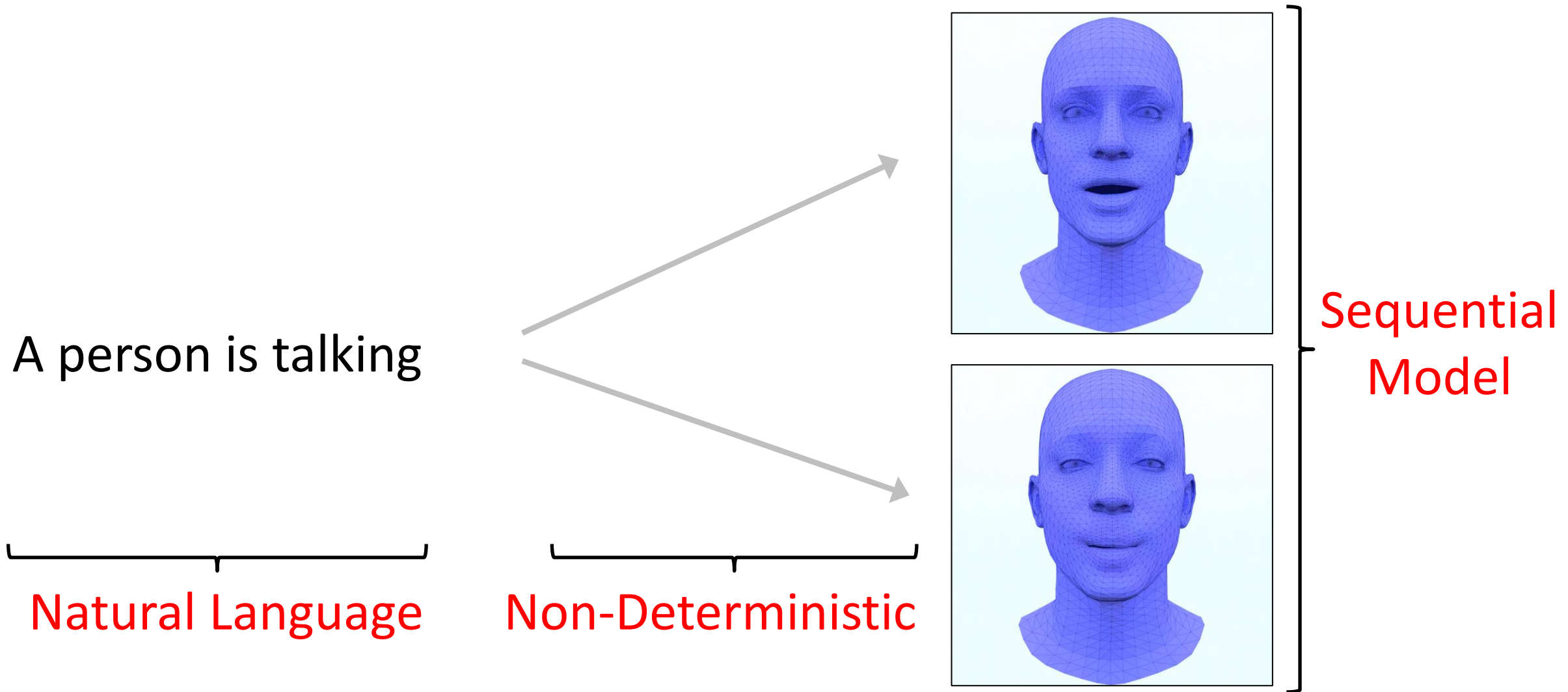


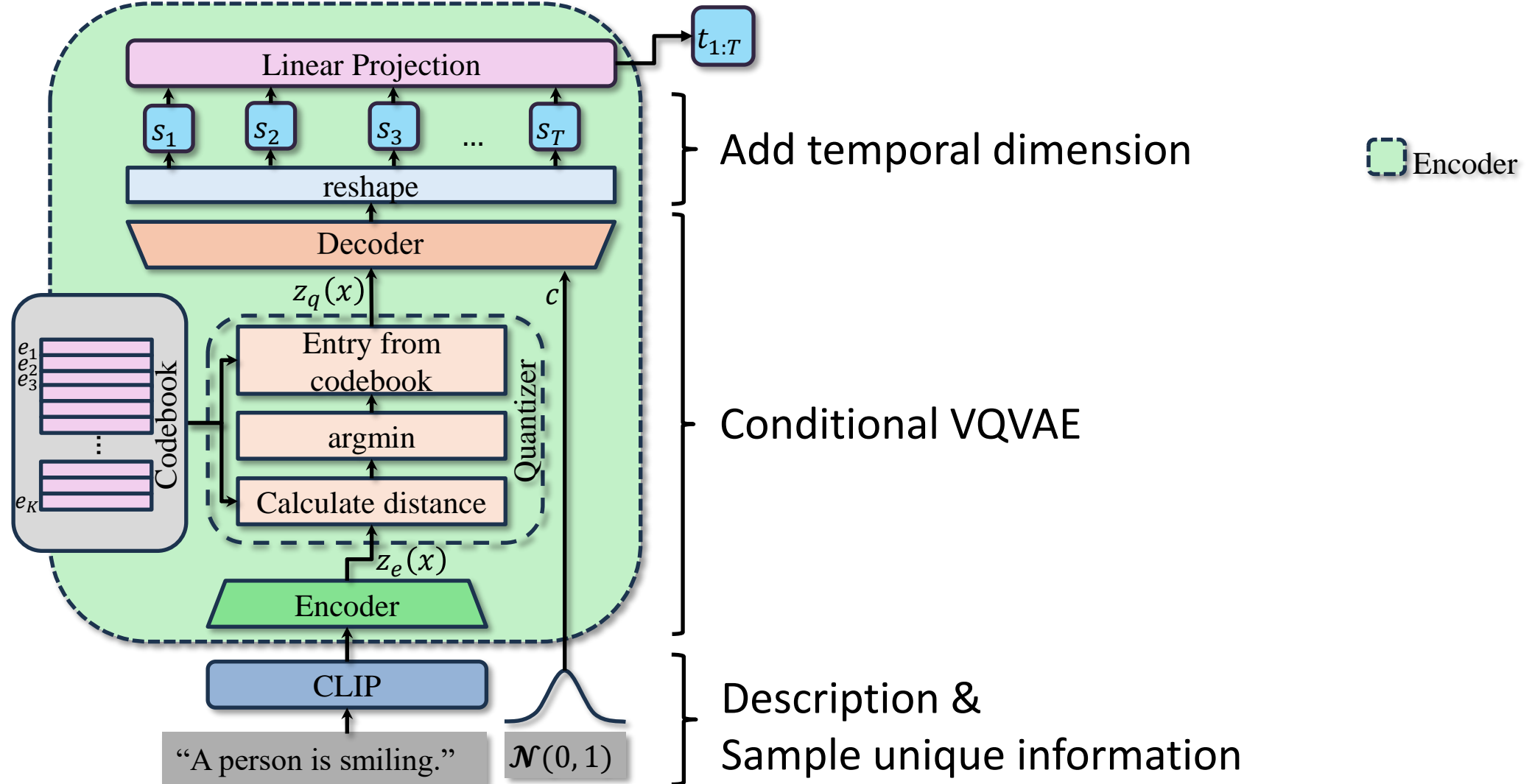
# Dataset Examples

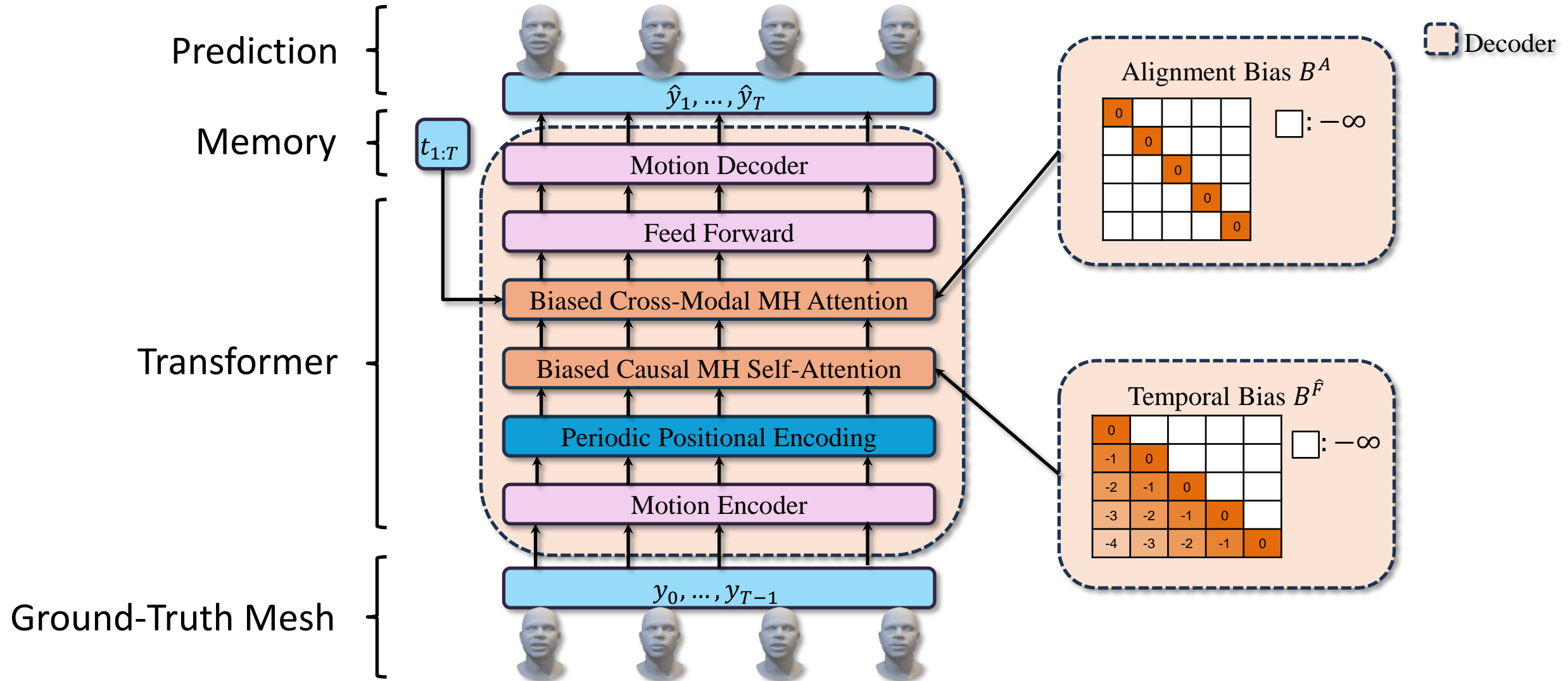




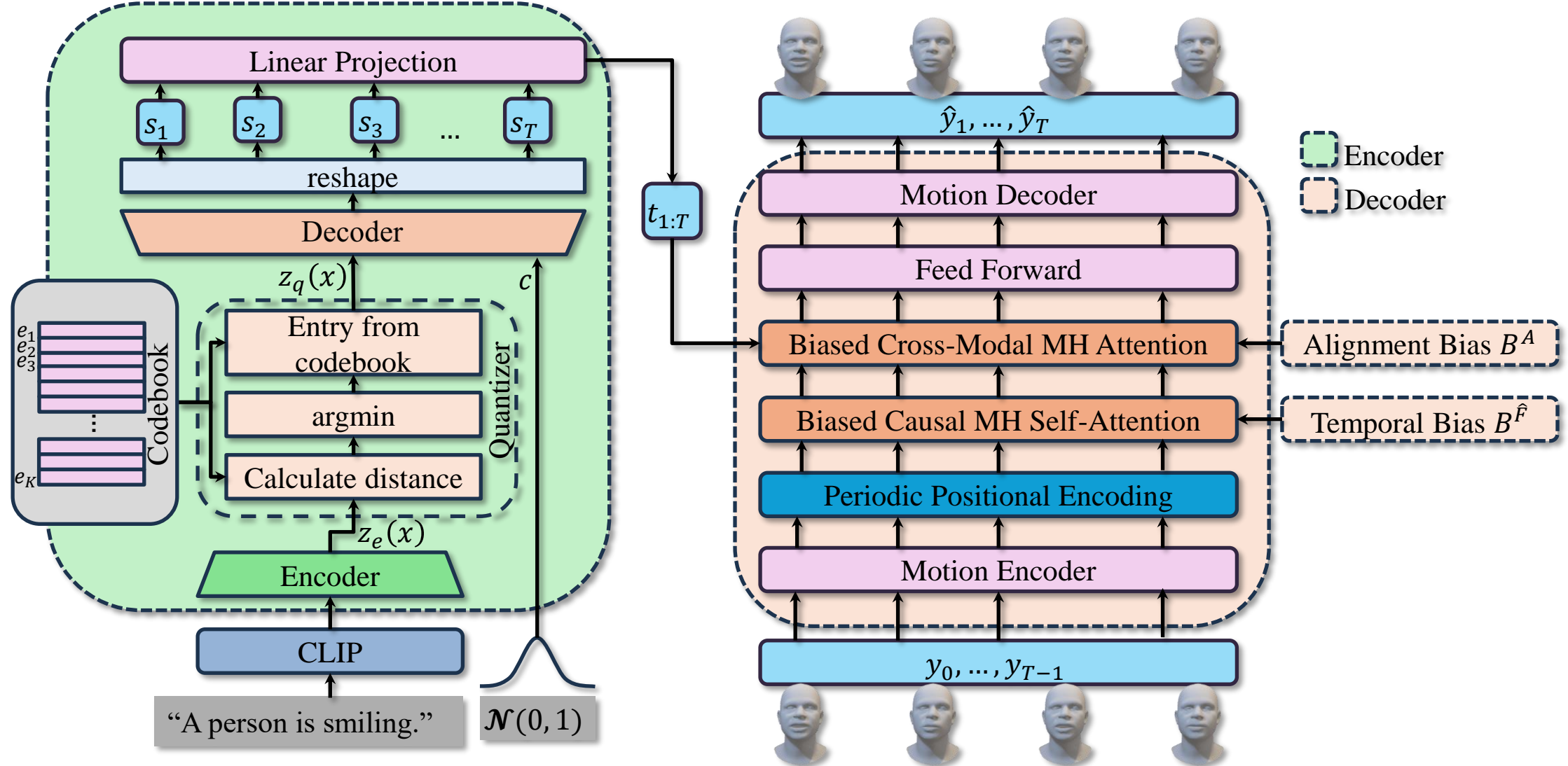








# Method



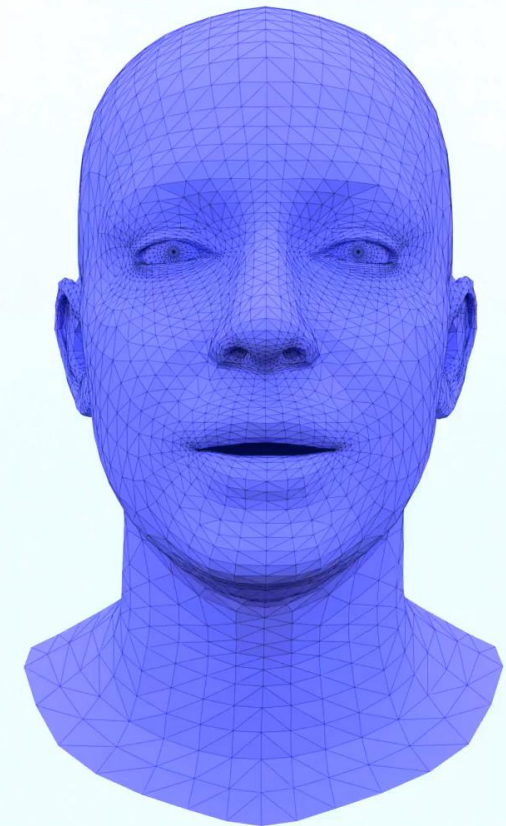
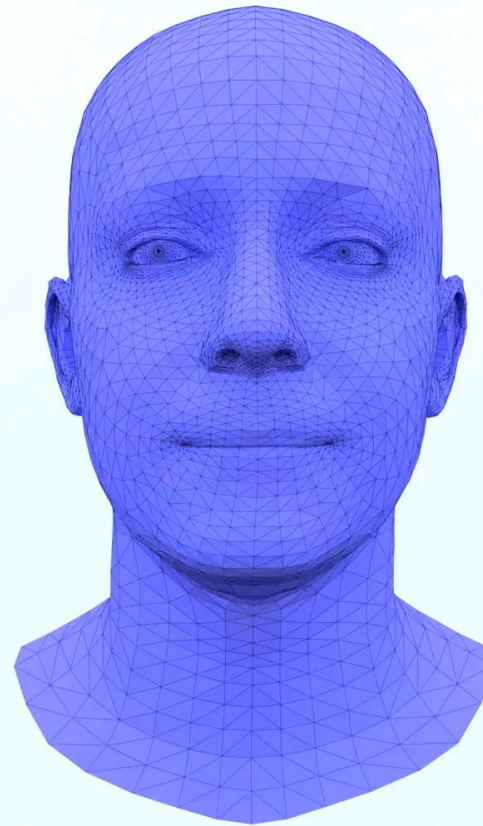
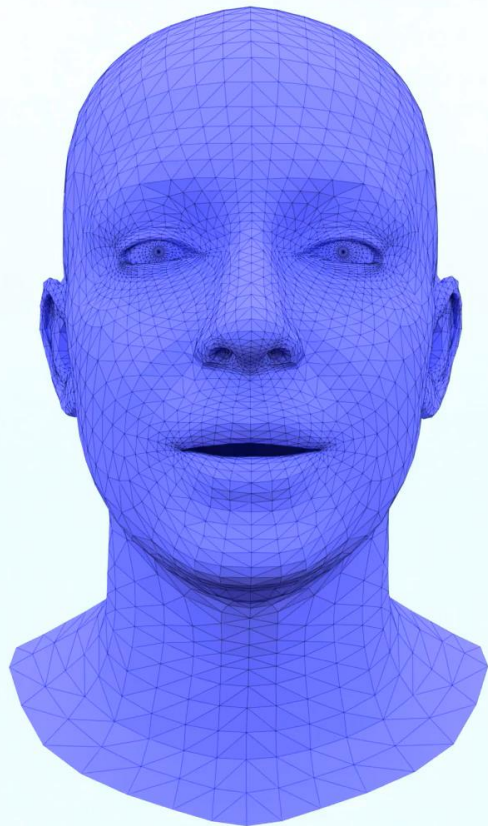


# Baseline Comparison

HMDM

SMF

Ours



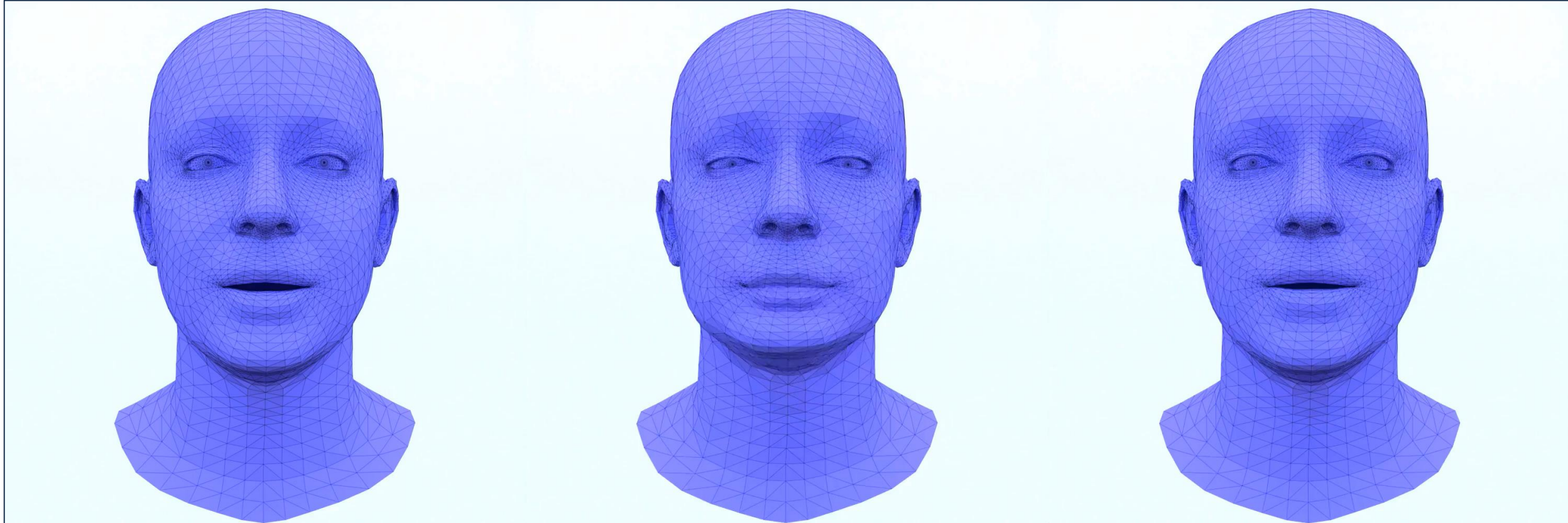
A person is kissing

# Baseline Comparison

HMDM

SMF

Ours



A person is talking

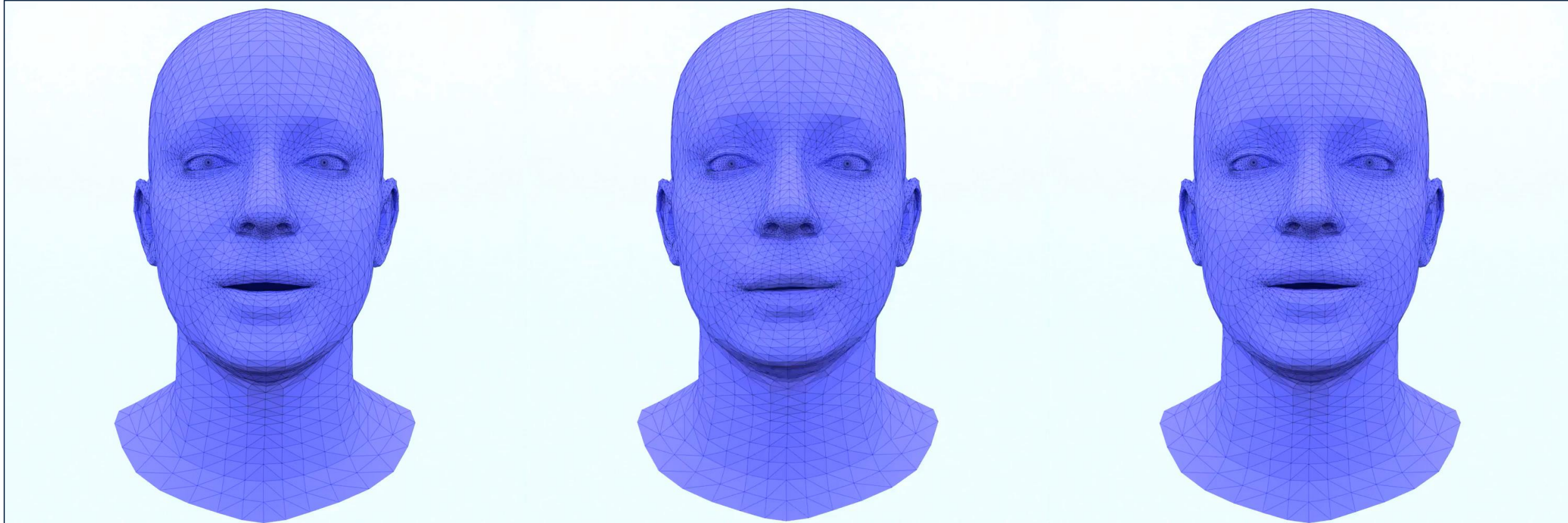


# Baseline Comparison

HMDM

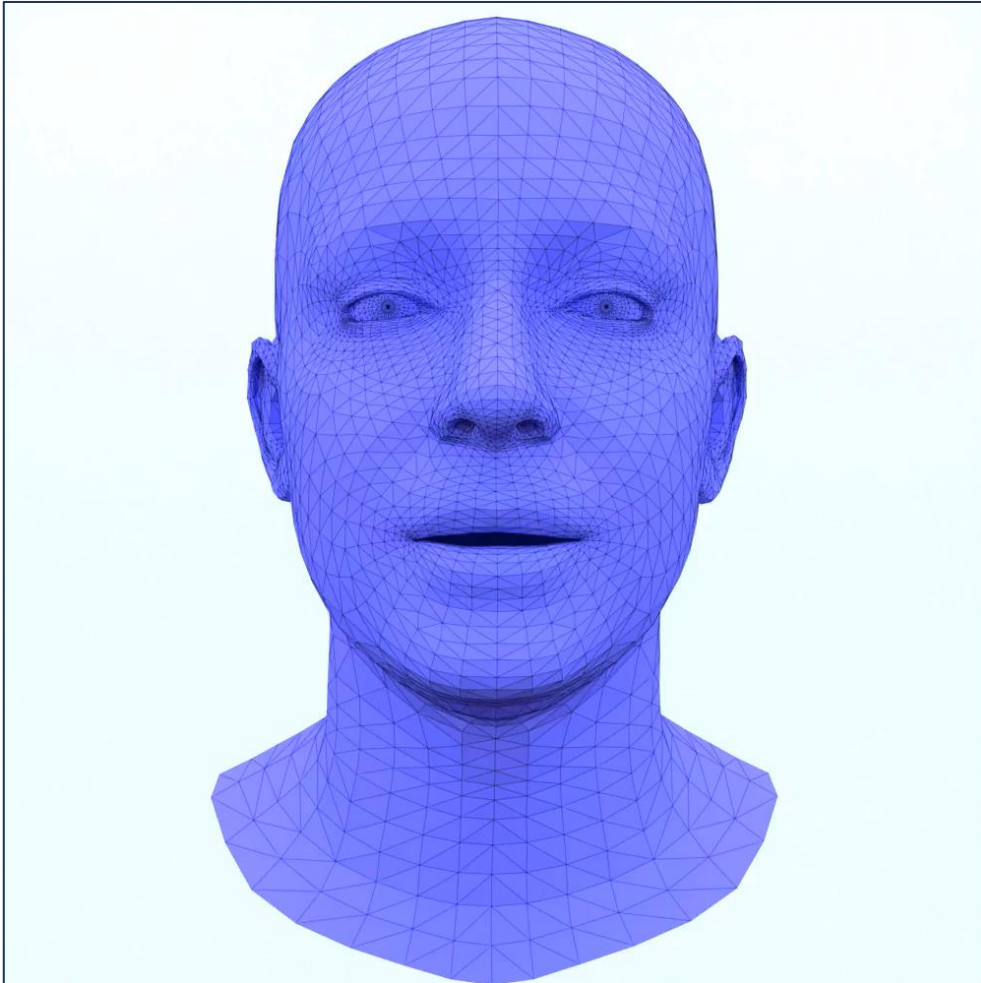
SMF

Ours

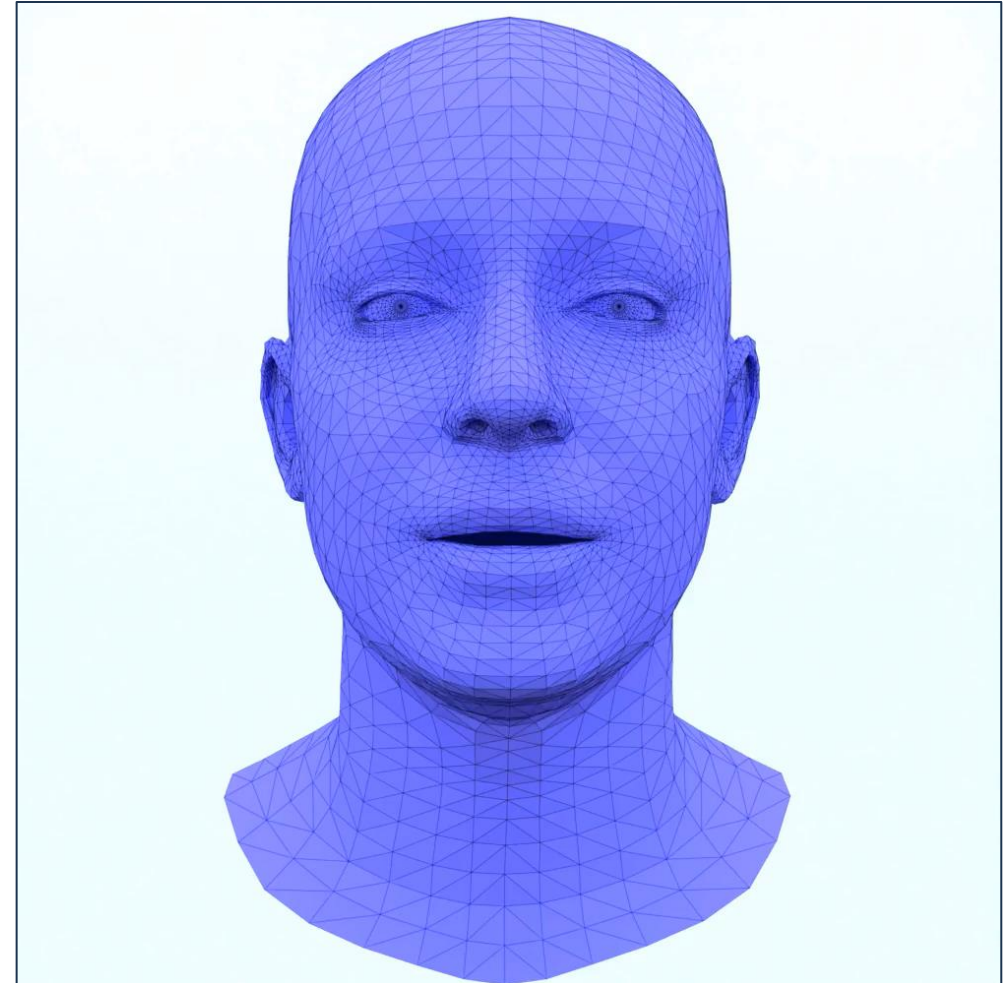


A person is yawning

# Qualitative Analysis



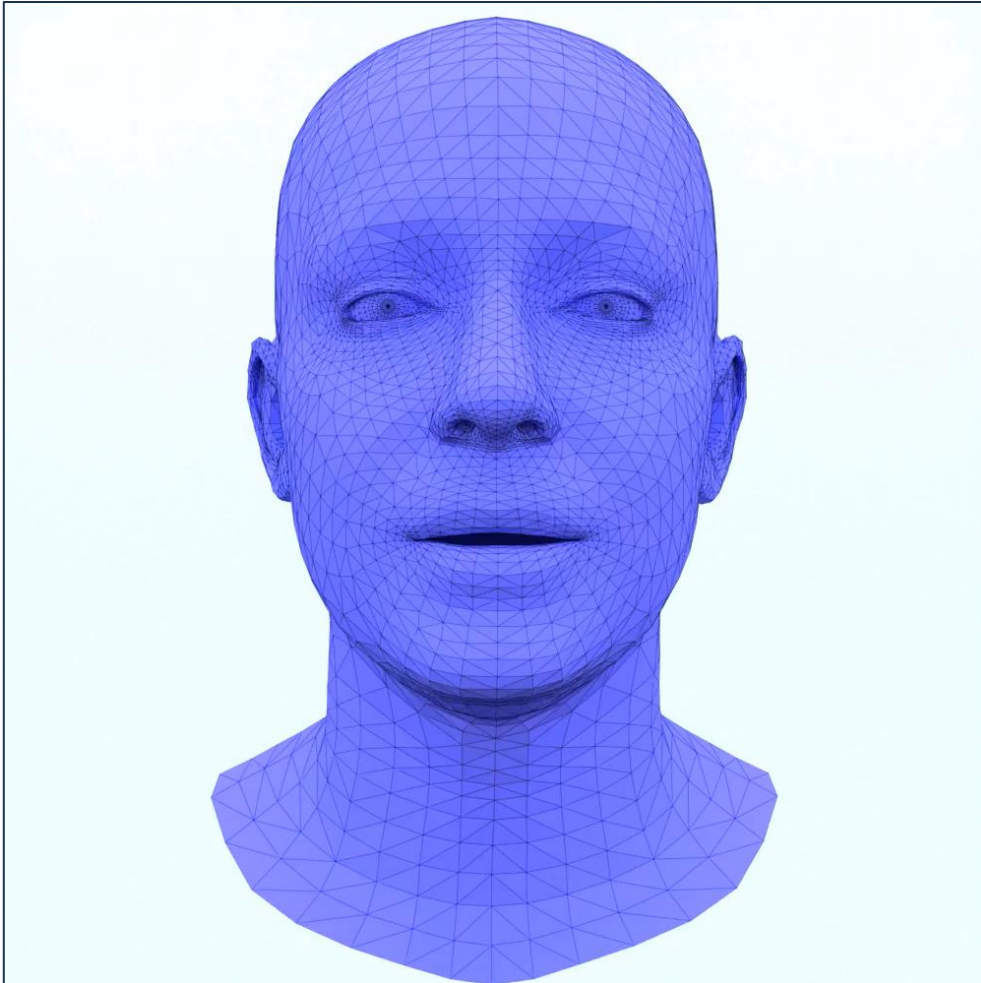
A person is crying



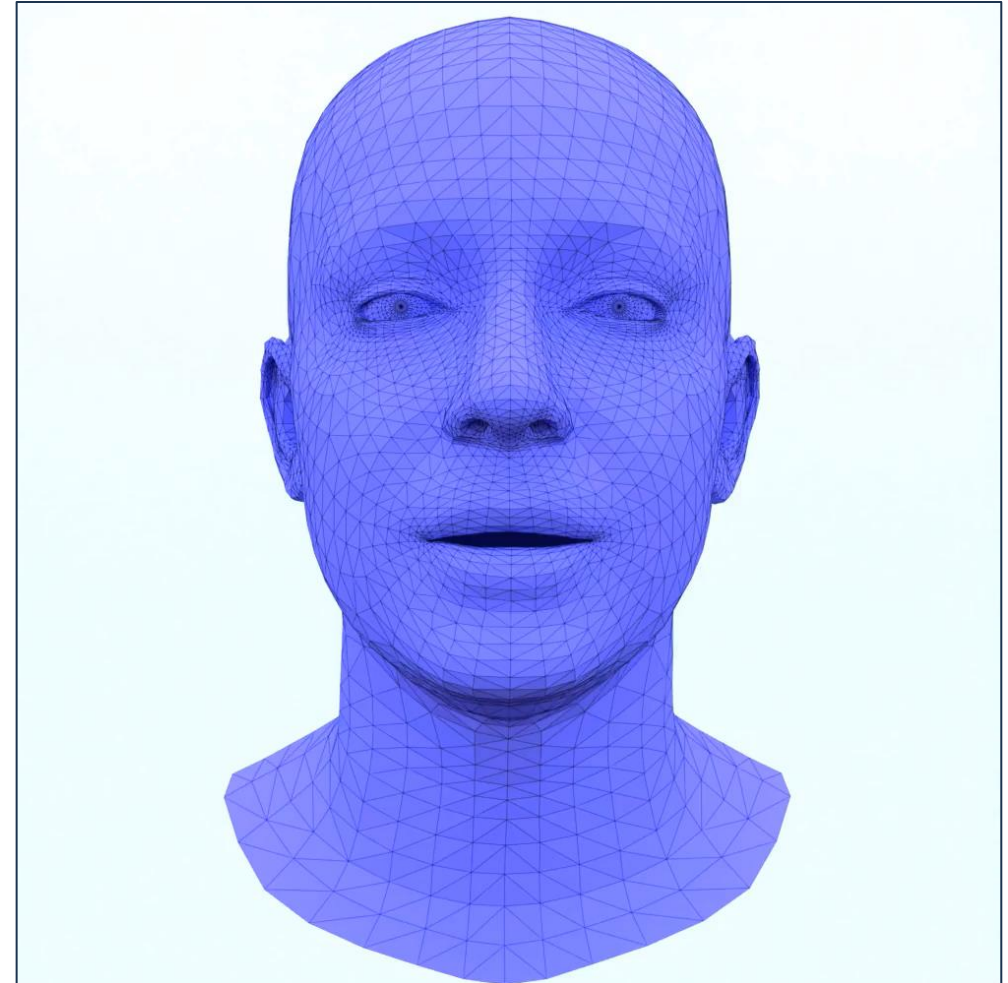
A person is weeping



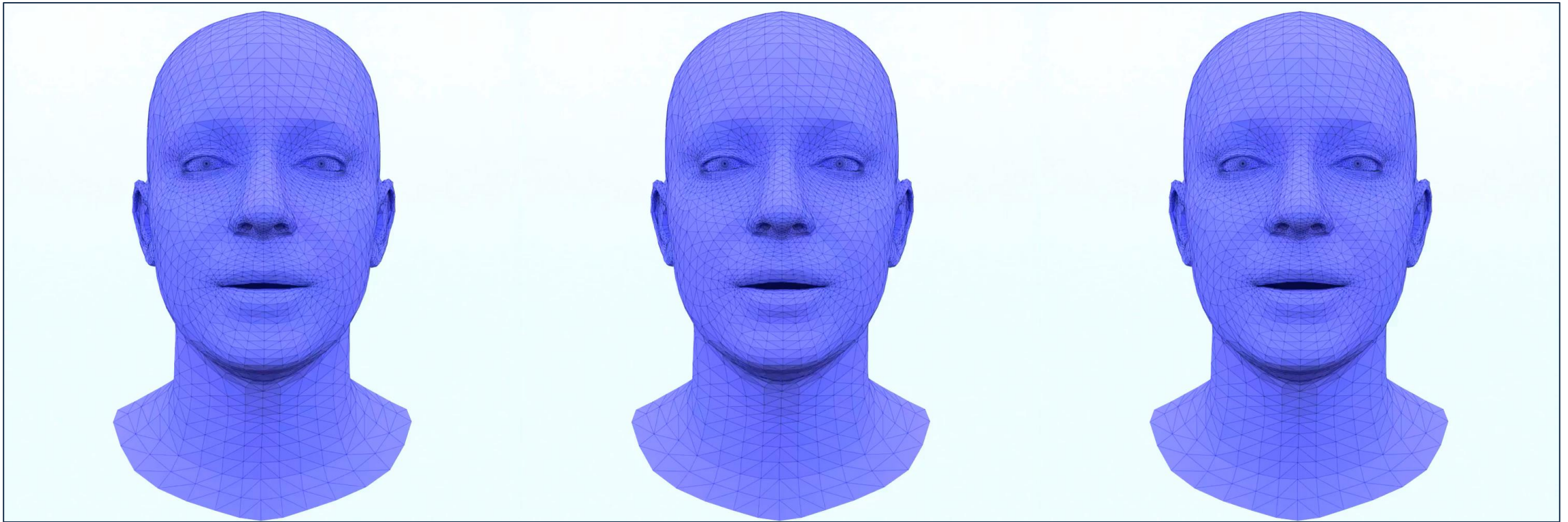
# Qualitative Analysis



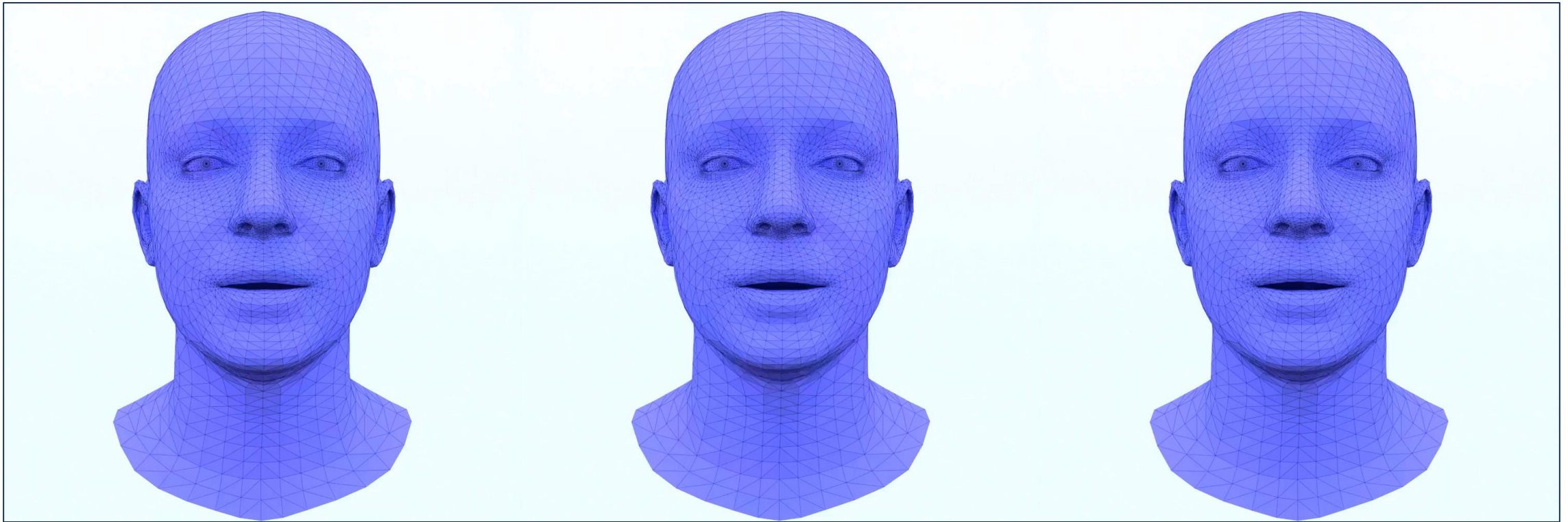
A person is sneering



A person is kissing



A person is drinking



A person is yawning



**Bold:** best result  
Underlined: second best

Method	MAE ↓	RMSE ↓	Diversity →
Real (Dataset)	—	—	$10.8744^{\pm 0.483}$
HMDM (adapted)	<b><math>0.6999^{\pm 0.0000}</math></b>	<b><math>0.9006^{\pm 0.0000}</math></b>	$0.0000^{\pm 0.0000}$
SMF (adapted)	$0.9272^{\pm 0.0052}$	<u><math>1.1773^{\pm 0.0060}</math></u>	<b><math>9.2109^{\pm 0.0620}</math></b>
Ours	<u><math>0.9239^{\pm 0.0067}</math></u>	$1.2013^{\pm 0.0083}$	<u><math>7.5996^{\pm 0.1755}</math></u>



**Bold:** best result

Method	MAE ↓	RMSE ↓	Diversity →
Real (Dataset)	—	—	10.8744 $\pm$ 0483
Autoregressive	<b>0.7003<math>\pm</math>.0000</b>	<b>0.9008<math>\pm</math>.0000</b>	0.0431 $\pm$ .0004
Style	0.9066 $\pm$ .0047	1.1673 $\pm$ .0054	6.7660 $\pm$ .1065
Ours (Full)	0.9239 $\pm$ .0067	1.2013 $\pm$ .0083	<b>7.5996<math>\pm</math>.1755</b>

**Bold:** best result

$D_{emb}$	$N_{emb}$	MAE ↓	RMSE ↓	Diversity →
Real (Dataset)		—	—	$10.8744^{\pm.0483}$
256	256	$0.8872^{\pm.0046}$	$1.1479^{\pm.0051}$	$7.1254^{\pm.1026}$
256	64	$0.9613^{\pm.0052}$	$1.2324^{\pm.0064}$	$6.8904^{\pm.1132}$
128	128	$0.8861^{\pm.0040}$	$1.1376^{\pm.0048}$	$6.8693^{\pm.1042}$
512	128	<b><math>0.8830^{\pm.0053}</math></b>	<b><math>1.1338^{\pm.0064}</math></b>	$6.4074^{\pm.1032}$
Ours (Full)		$0.9239^{\pm.0067}$	$1.2013^{\pm.0083}$	<b><math>7.5996^{\pm.1755}</math></b>

## **Dataset Quality**

- Sometimes problems with the tracking accuracy
- Unbalanced distribution of expressions
- Textual descriptions varies very little

## **Model Diversity**

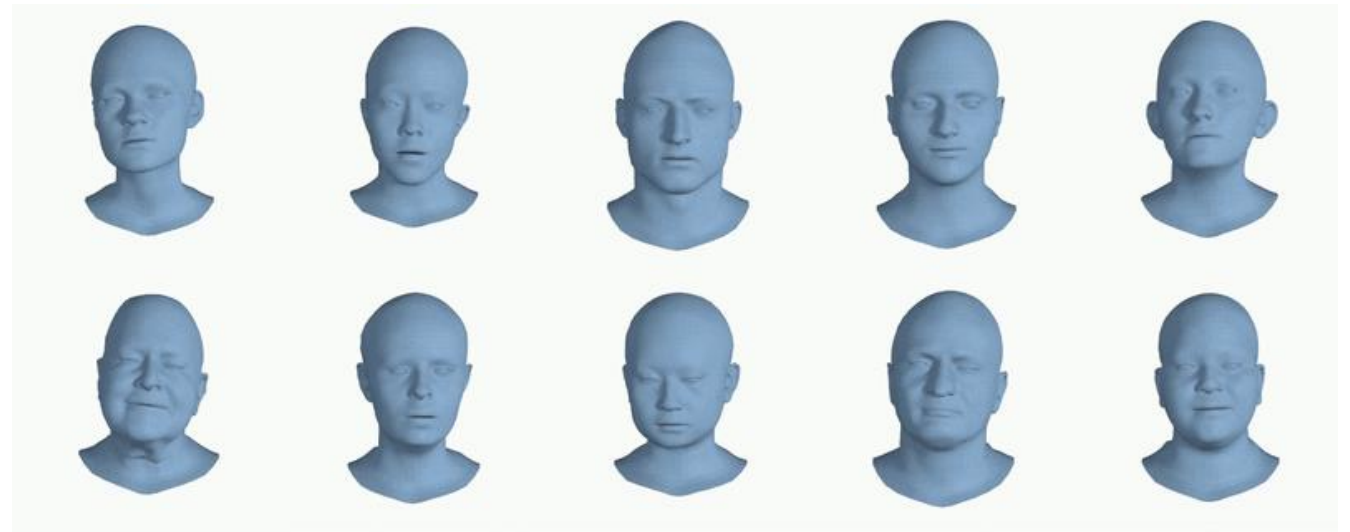
- Model could reach the desired expression, but remained mostly static thereafter

## Dataset:

- Explore FaMoS dataset
- Create new Dataset with better descriptions

## Model:

- Explore SMF Diffusion Approach



FaMoS Dataset

<https://tempeh.is.tue.mpg.de/>