



March 20-24, 2023
San Francisco, CA

GPT-3 Powered Text to Lifelike Speech and Animation for NPCs

Dao Si

Director of AI Team @ JNG Studio, NUVERSE

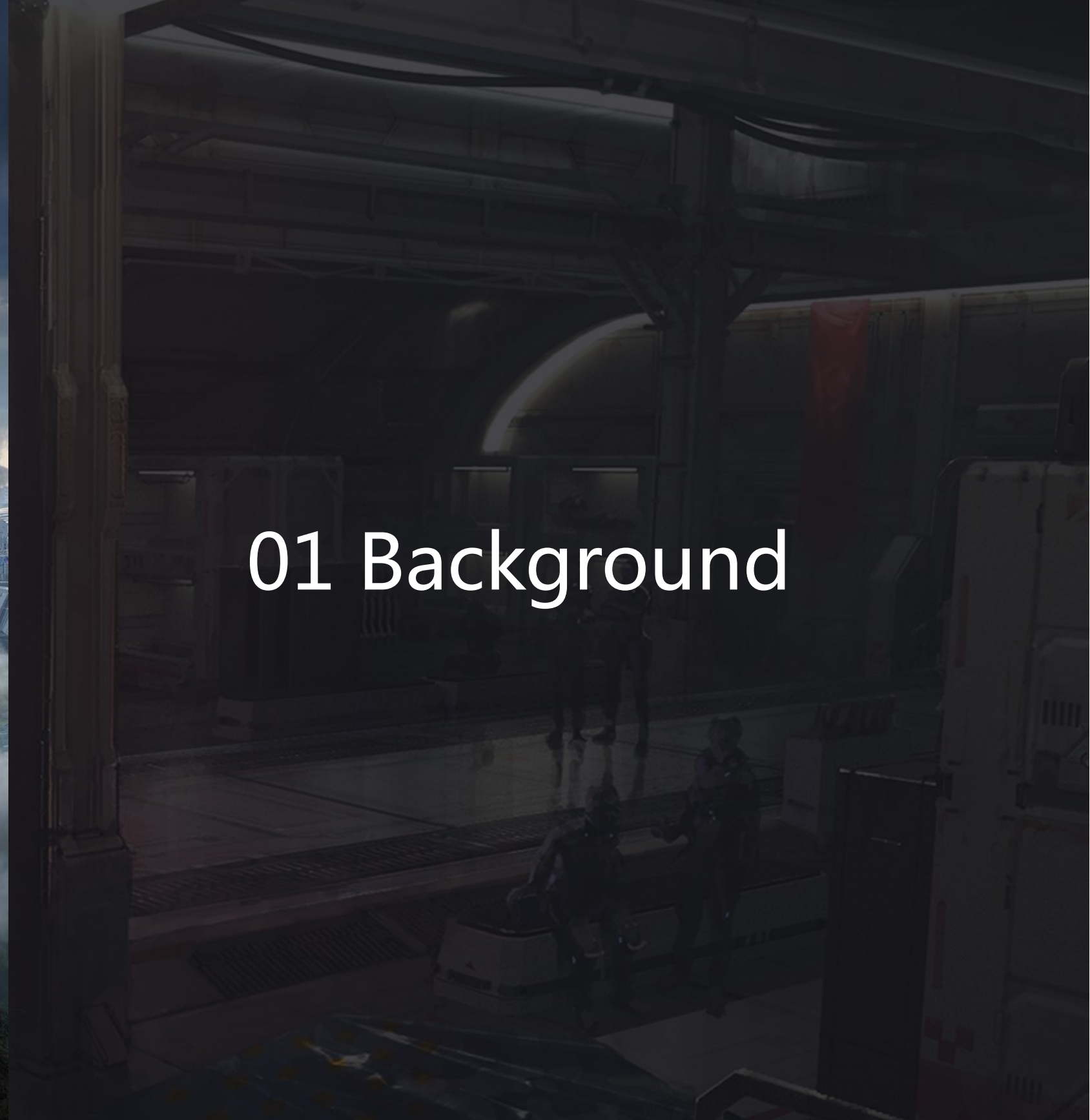


#GDC23

01 Background

02 HI System for NPCs

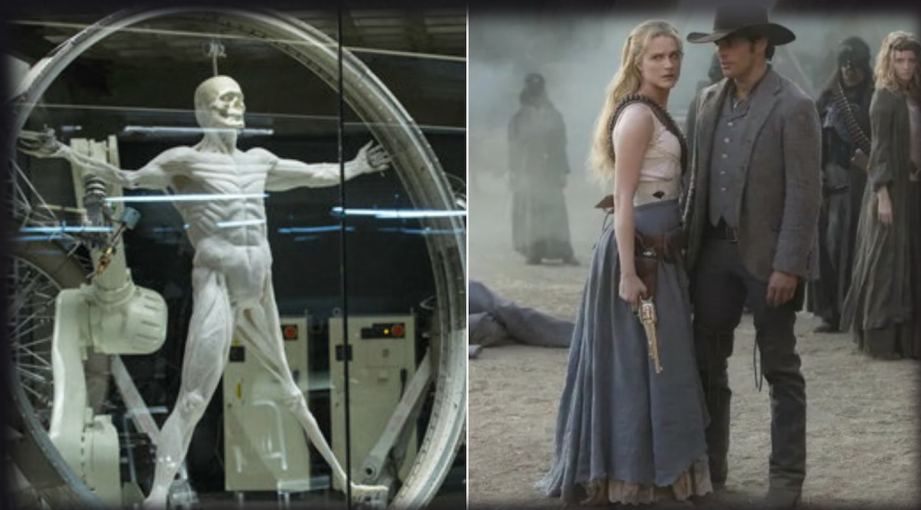
03 Application & Future



01 Background

Smarter NPCs make a more immersive experience

《West World》



《Detroit: Become Human》



《Free Guy》



The key elements of intelligent NPCs

- A variety of diverse personality appearances
- Natural lip expressions and body movements
- Emotionally rich authentic voices
- Personalized conversations and interesting character stories
- Human-like behavioral responses and plans

AI-powered Role-playing NPCs

3D Avatar

a beautiful blonde
with green eyes and pink lips



Speech&Animation

Text to Lifelike Speech and
Animation



NPC Behavior

Behavior generation based on
text description



Unscripted dialogue & Story generation

Personalization of dialogue and story
generation in line with character Settings

魔皇突然倒退了一步，伸手摸了一下身上的伤口，然后朝着四周的虚空一指，原本在地下的虚空魔法阵直接消失无踪，这里顿时恢复了活性。
“来了.....”魔皇喃喃的自语着，然后死死的盯着魔王的背影，不敢有丝毫的怠慢，在一个黑袍大汉的带领下，朝着人类的方向搜寻了过去。
几分钟之后，在前面的虚空魔法阵眼前一空，所有的人都露了出来，大家一个个的瞪大了眼睛。
“原来是魔皇大人。”听到魔皇的声音后，所有的人都感到惊讶，因为在他们眼中，这就是魔皇大人的魔导师，终于踏上了辉煌的魔法师之路，按理来说，能够拥有神力的魔导师是当之无愧的人类。
但是此刻魔皇的身影却消失在这片大海之中，似乎早就看穿了这一切，这让众人的心都悬了起来。
片刻之后，众人感觉到了空气中的清爽，所有人都猛的抬起头来，那巨大的眼睛里充满了向往，这样的表现让他们觉得此行不枉。
他们不|

LLM

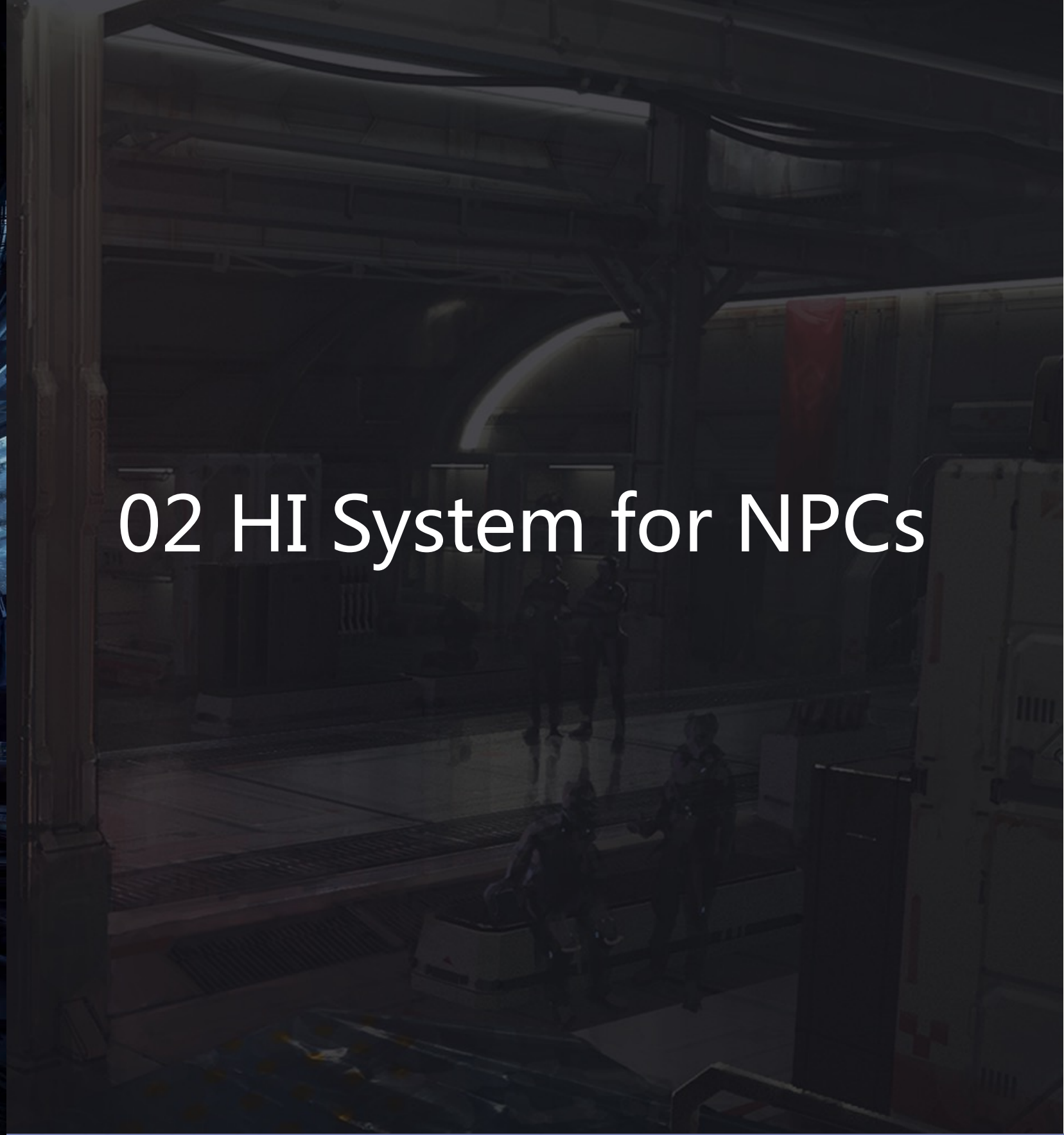
Background

AI-powered role-playing NPC
Lydia
Real-time Interactive

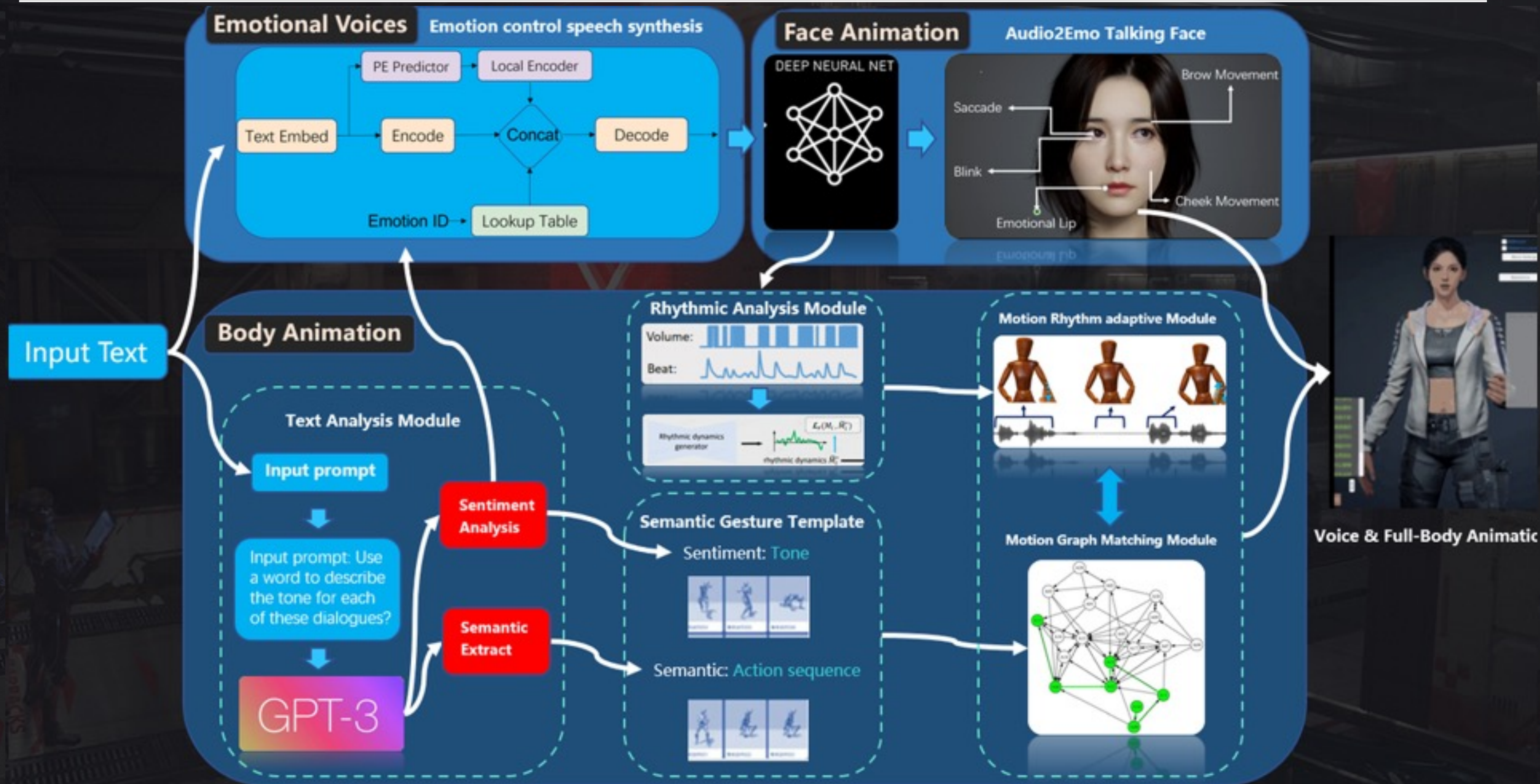




02 HI System for NPCs

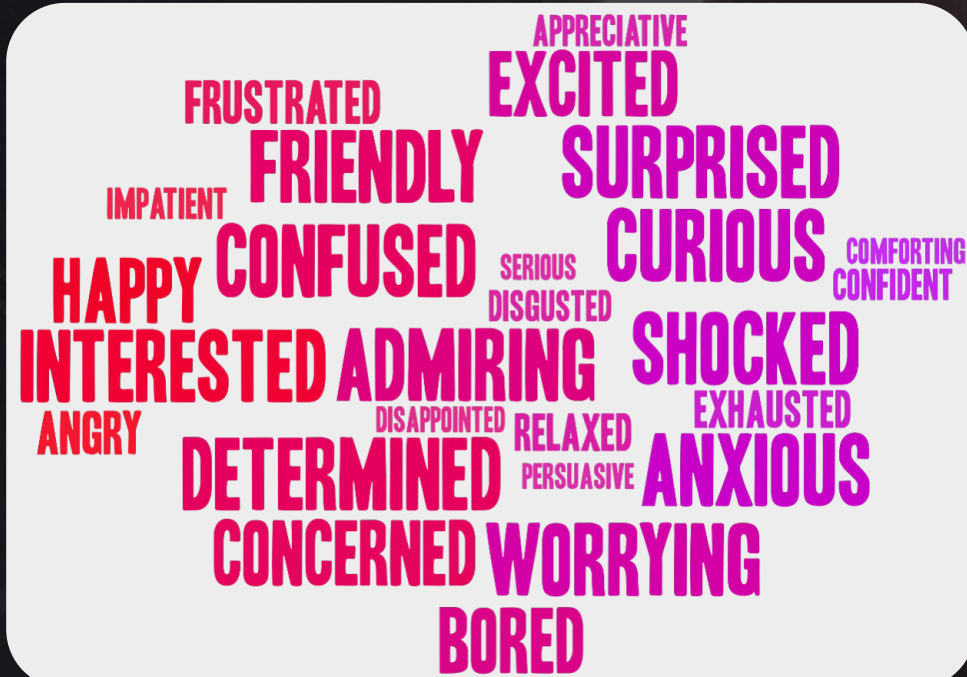


The framework of HI System for NPCs



Text Analysis Module

Emotional Labels (30+)



Positive Negative Ambiguous

hey [gender title _2], long time no see, I'm Yang, you must remember me, after all, w O. Friendly		friendly
The Duga Valley isn't very safe, but I'm holding on. How about you?	1. Cautious	cautious
I'm not bad either, it took some effort but I found a nice place and I set up a brand ne	2. Proud	proud
Asylum camp? what is that?	3. Curious	curious
My God, don't you know about the sanctuary camp yet? It's been a while since the A	4. Surprised	surprised
The members of the shelter camp will build the camp together to defend against the	5. Explaining	explaining
By the way, my companions and I also figured out how to build all kinds of advanced	6. Impressed	impressed
Sounds really good, maybe I can go to you?	7. Hopeful	hopeful
Well, I hate to say that, but we're full at the moment, sorry, but I suggest you find one	8. Apologetic	apologetic
Union law enforcement is like this sometimes, there are many stunned green soldier	9. Nervous	nervous
What about the baby at the scene?	0. Concerned	concerned
That kid has been placed and my friends are doing medical tests for him.	1. Reassuring	reassuring
Can you talk to me about the details? To be honest, I'm still in the fog.	2. Confused	confused
There was a murder, the unfortunate victim was a young girl, and a baby was found	3. Matter-of-fact	matter-of-fact
The victim was a 19-year-old girl named Kiki.	4. Sad	sad

Semantic Gesture Labels (200+)



Metaphorics

Iconics

Deictics

Emblems



Using GPT-3 API

Gather inputs

Requires text to classify and training examples of pre-labeled text

Input text

Example Text, Label

Example Text, Label

Example Text, Label

Example Text, Label

Narrow examples

Searches to select most relevant examples up to `max_examples` number

Input text

Example Text, Label

Example Text, Label

Example Text, Label

Example Text, Label



Fine-tuning

Additional training to become better at a certain task

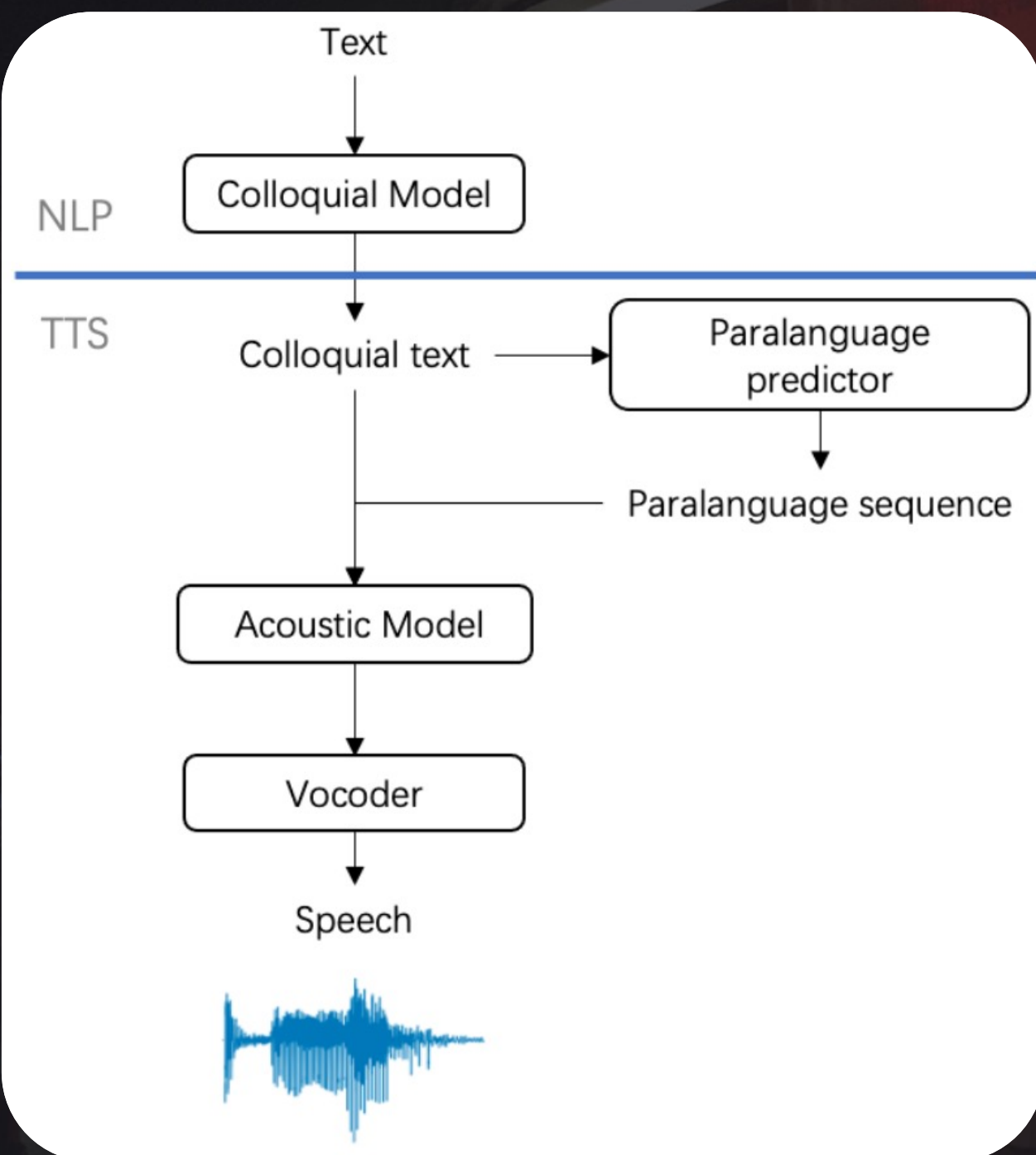
Task: Textual colloquialism

Task: Emotion Detection

Task: Semantic Label Extraction for Gesture

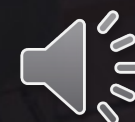
Text to Speech Module

Textual colloquialism & Paralanguage



Original Text	Colloquial Text
南方菜系偏爱蘸料，例如我第一次去上海才知道烧烤里的蔬菜也需要配蘸料 (Southern cuisine tends to favor dips and sauces. For example, it wasn't until I visited Shanghai for the first time that I realized vegetables in barbecue dishes also needed dipping sauce.)	嗯，南方菜系的话，超级偏爱用蘸料啊什么的，就比如说我第一次呢，第一次去上海的时候，才知道这个烧烤里的蔬菜也得配着蘸料 (Well, when it comes to the southern cuisine, I really love to use dipping sauces, such as when I first went to Shanghai and realized that even the grilled vegetables should be paired with a dipping sauce.)

Original

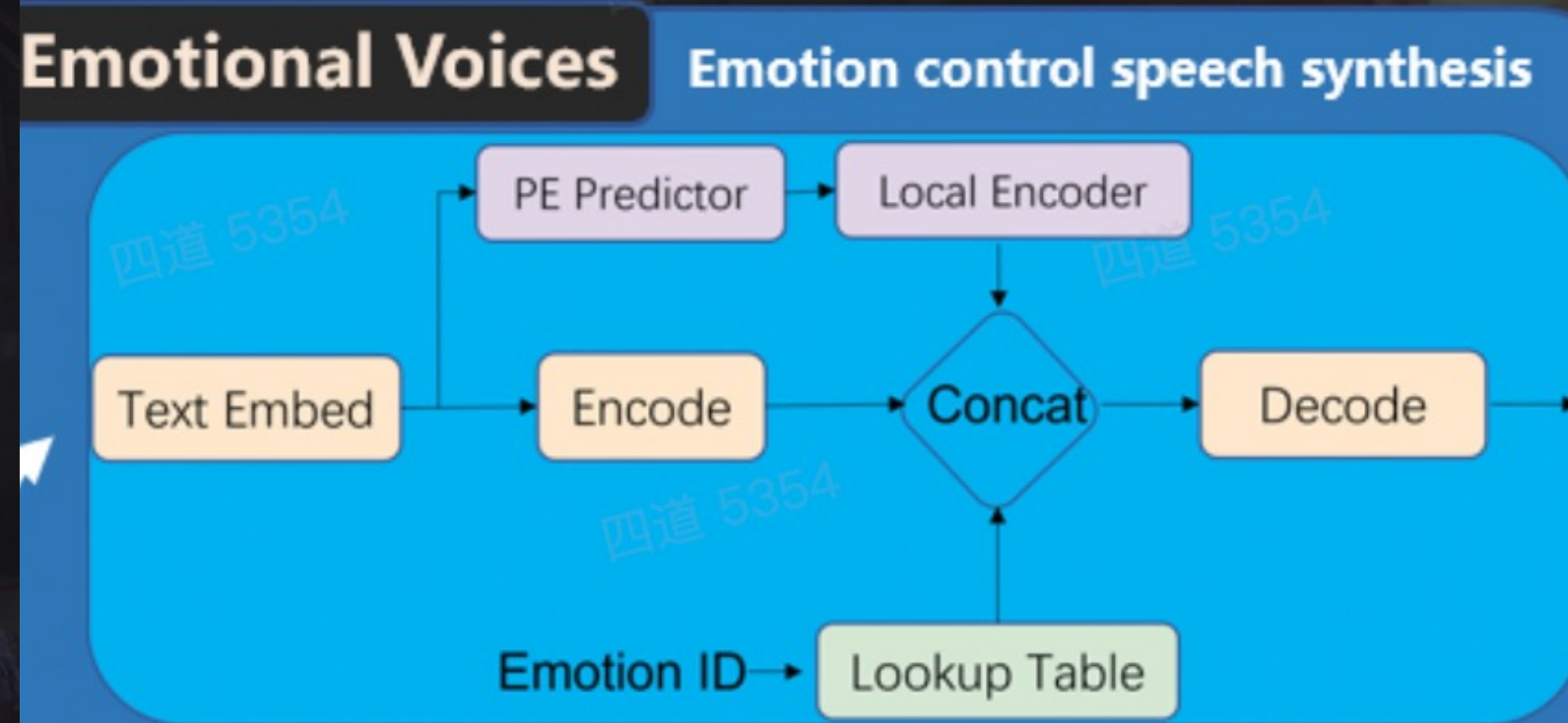


Optimization



Text to Speech Module

Emotion Tag From GPT-3



Happy



Disappoint



Surprise



Curious

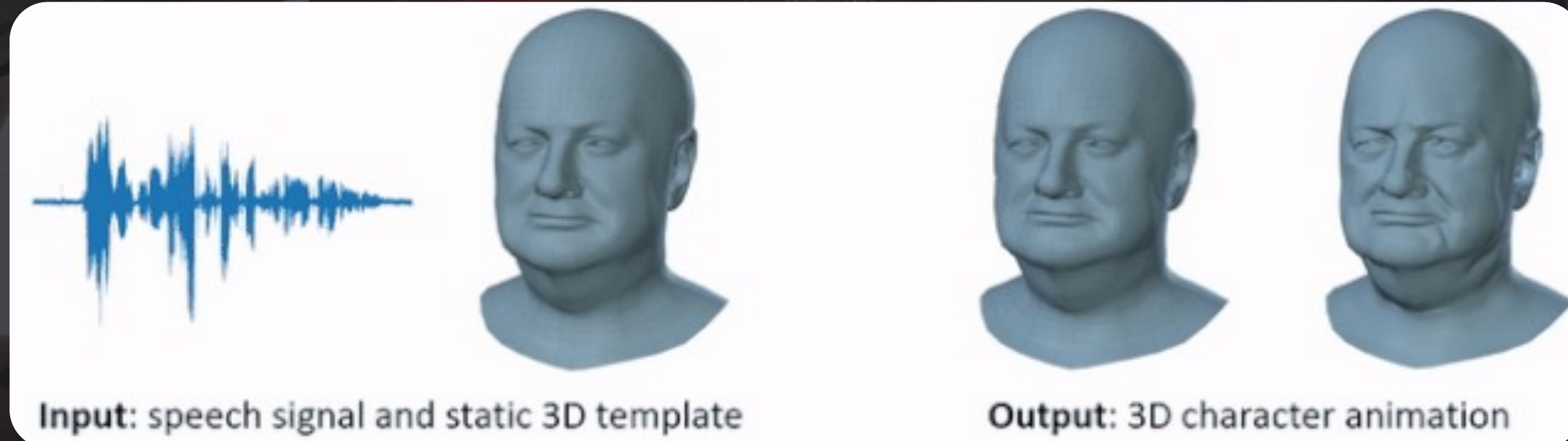


Praise



Facial Animation Generation Module

Problem Define :



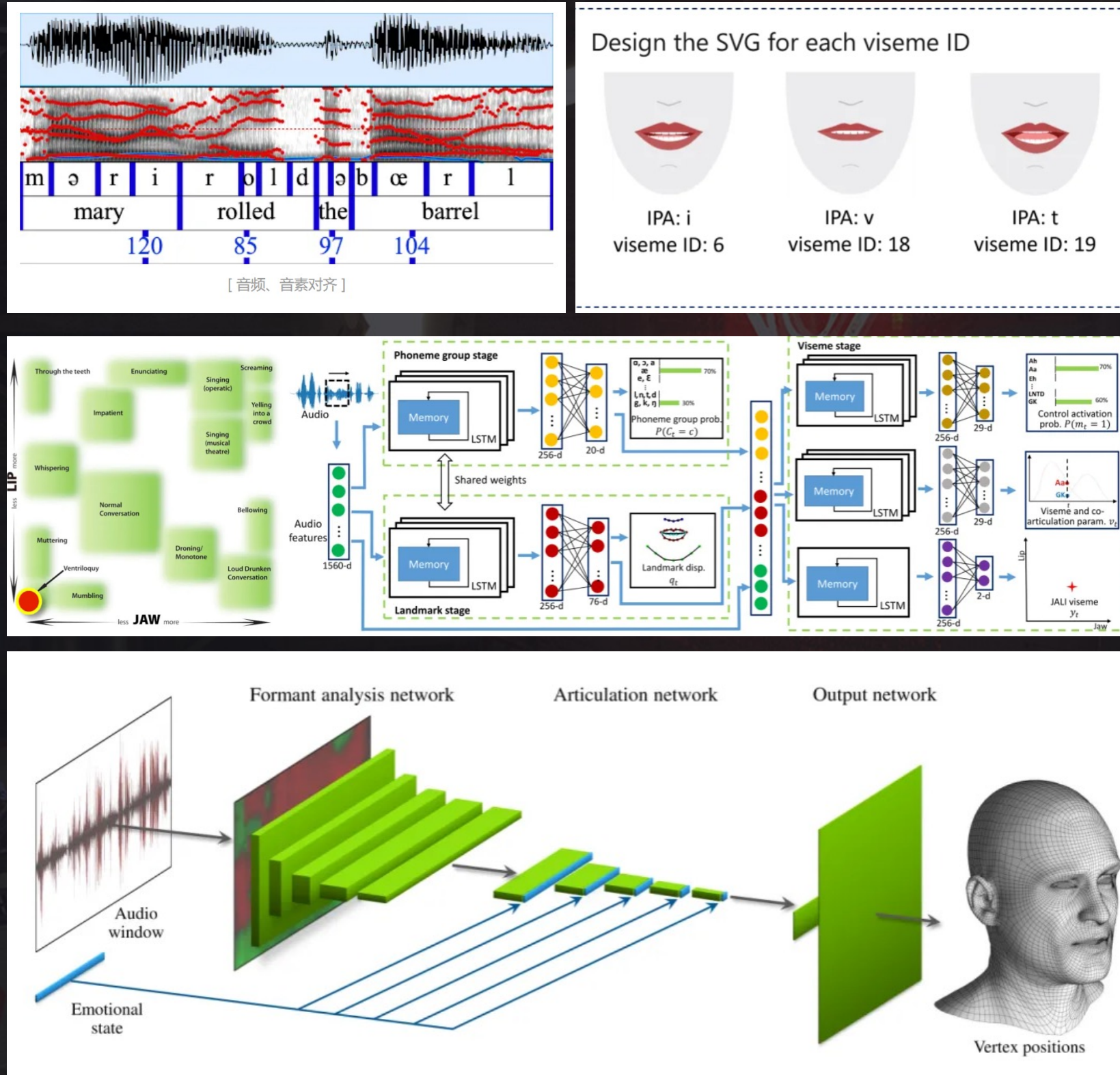
Challenge :

Little Data : The available data is scarce, and there is a significant shortage of accessible, synchronized datasets integrating speech and 3D models.

Authenticity : Many In many jobs today, facial expressions are often mechanistically programmed, leading to individuals with wooden and rigid faces lacking genuine emotional feedback.

Generalization: Different individuals have their unique ways of speaking, and even the same parameters applied to facial models of people with different appearances can result in diverse outcomes.

Facial Animation Generation Module



Research history:

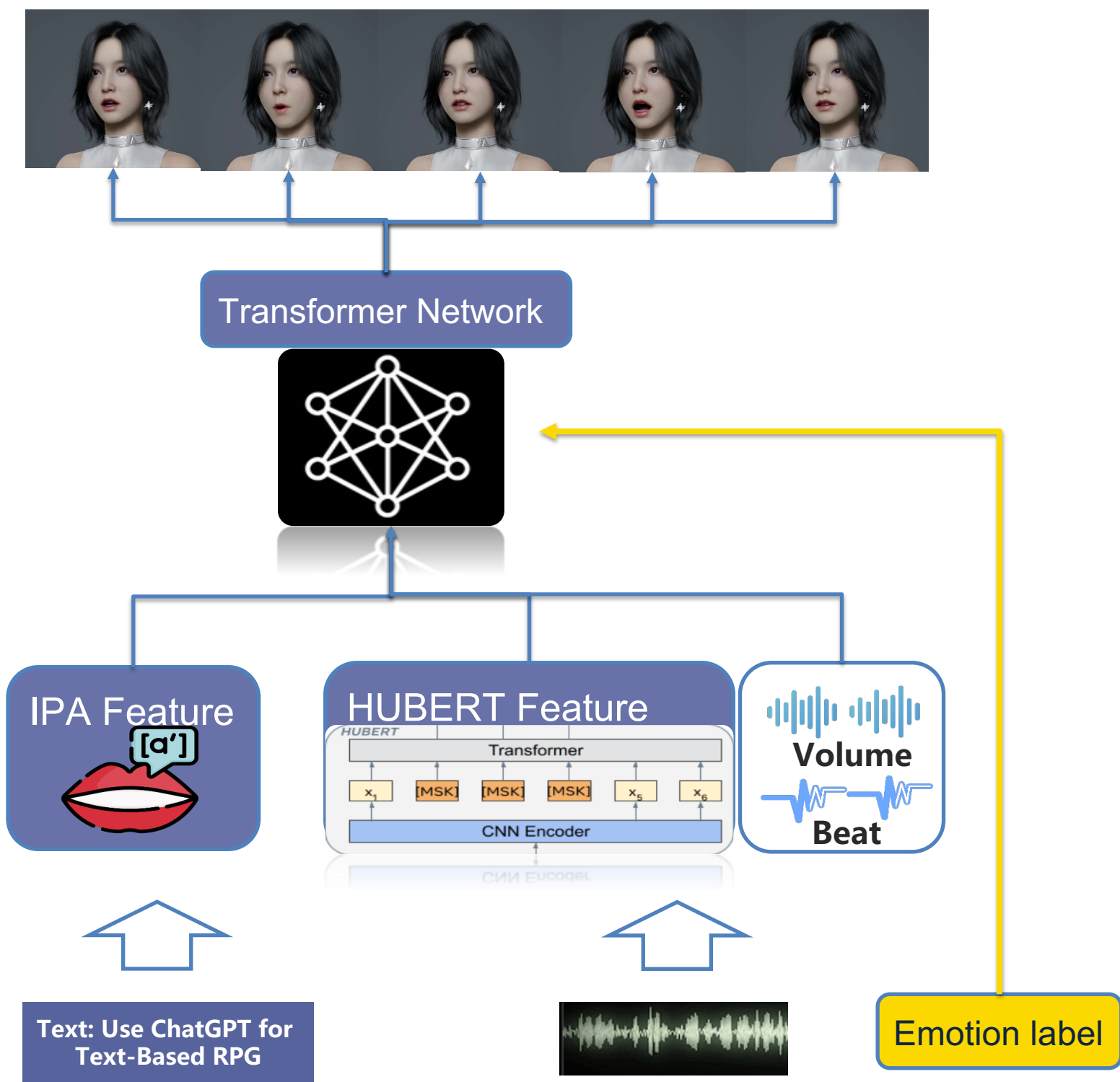
Method based on phoneme visual element mapping. The details are not natural, hard to adjust, dependent on ASR and art resources;

Methods based on LIP&JAW. On one hand, solving the phenomenon of collaborative pronunciation mainly depends on several rules defined by humans, which is neither complete nor systematic. On the other hand, the "style" of speaking requires a lot of artistic manual adjustment, which is not conducive to the highly automated workflow needed for large-scale dialogues.

An end-to-end approach based on deep learning. But high-precision large-scale data acquisition is also a pain point for this solution.

Facial Animation Generation Module

Emotional 3D Facial Animation

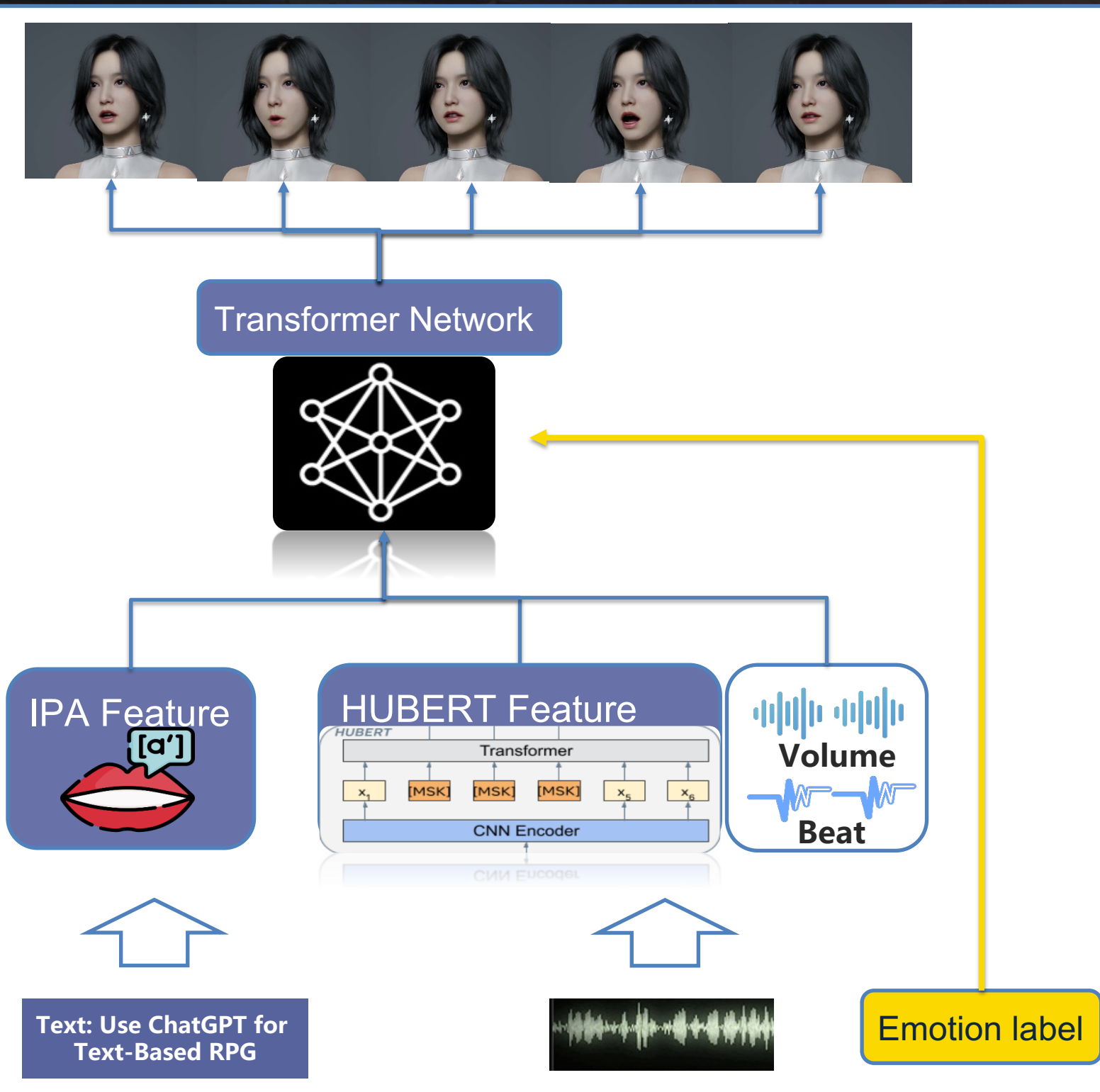


Mandarin

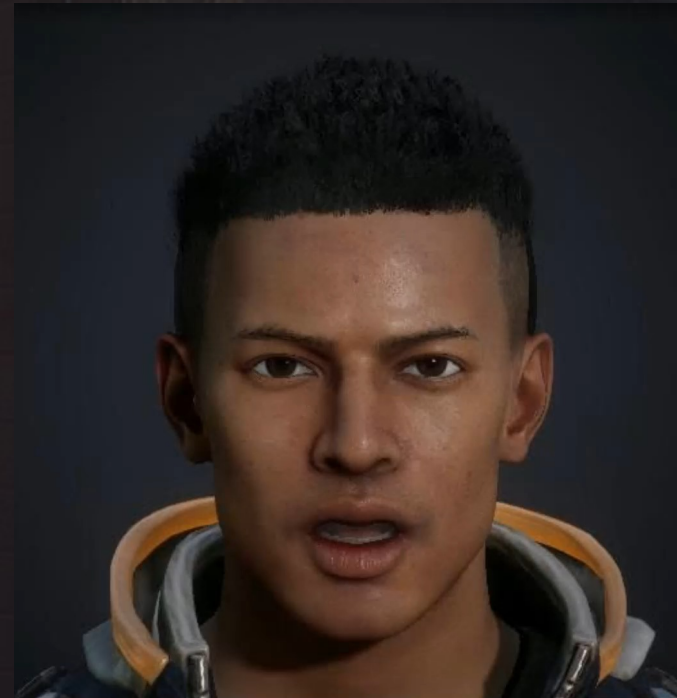


Facial Animation Generation Module

Emotional 3D Facial Animation



Neutral



Doubt



Happy

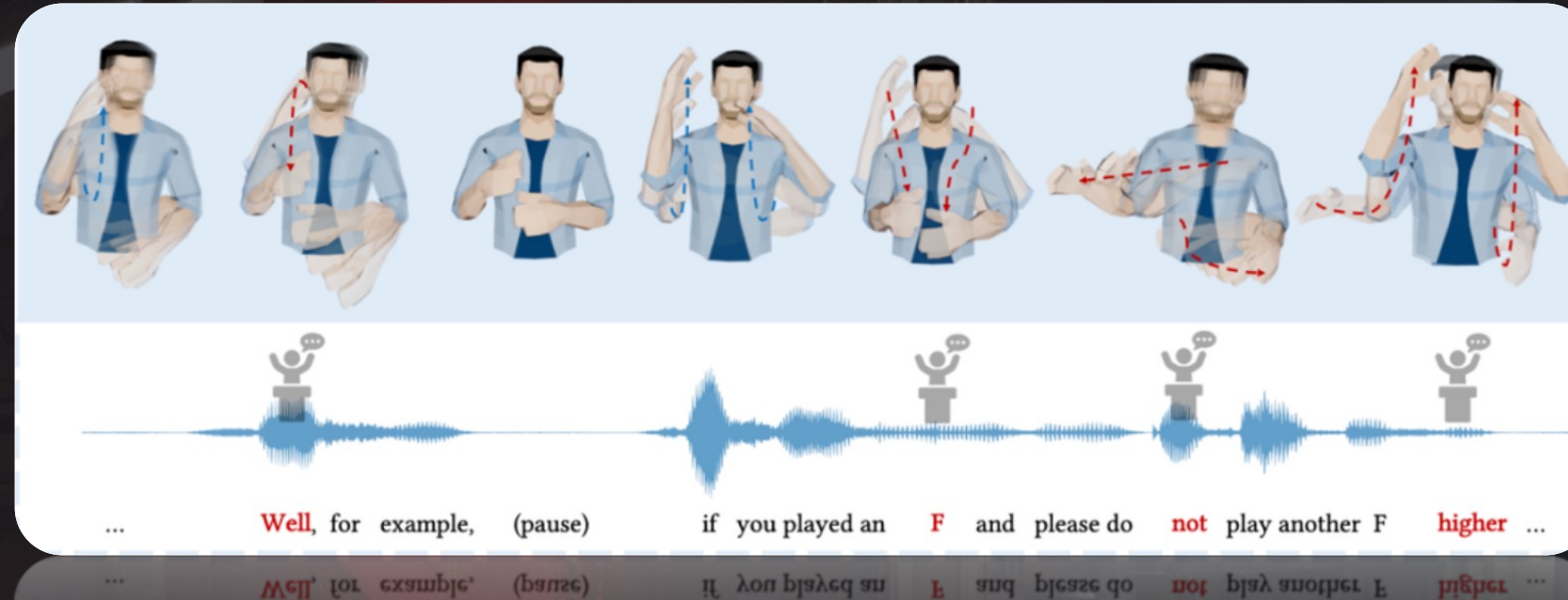


Sad



Body Motion Generation Module

Problem Define :



Challenge :

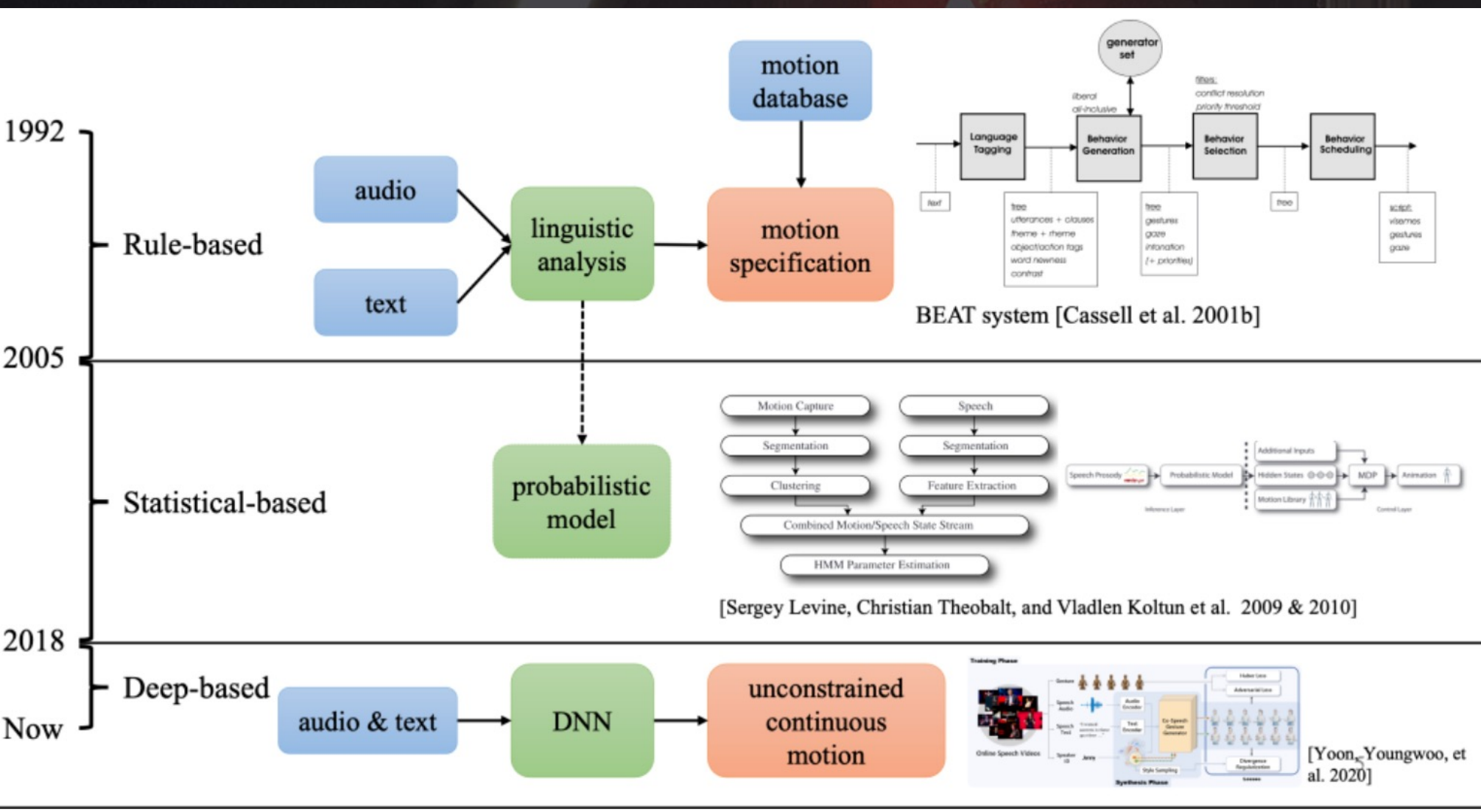
Complex rhythmic structure. The prosodic structure of spontaneous gestures in human speech is complex and has not been strictly studied in linguistics.

Strong semantic correlation. Gestures are often used to explain the meaning of speech (such as marking a number with your hand), but the term "semantics" covers a wide range of things.

Polysemy. We can often use different gestures to reasonably express the same sentence.

Body Motion Generation Module

Research history:

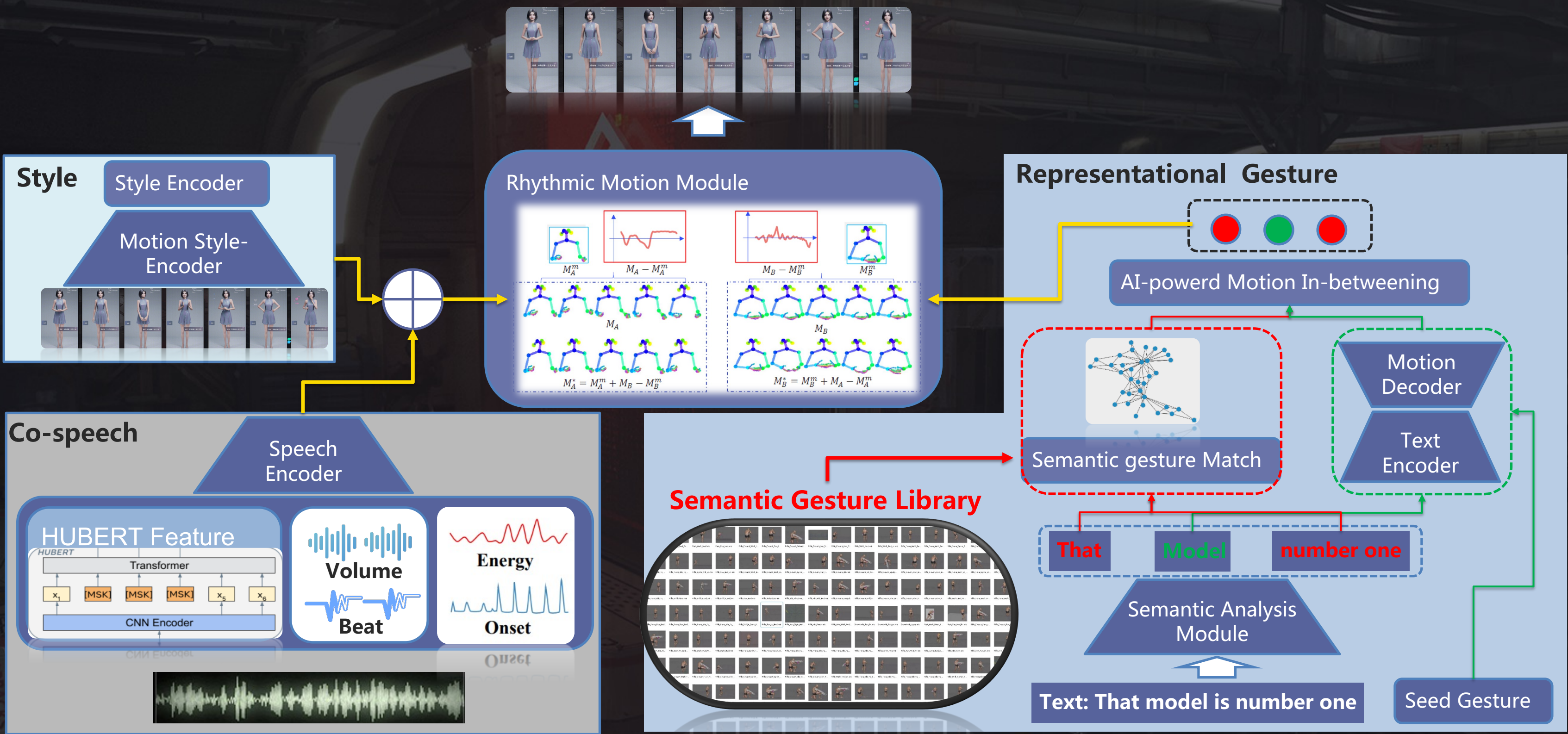


From 1992 to 2005, a rule-based approach was adopted. Explicit mapping rules bring interpretability, which is conducive to processing semantic gestures, but limited rules are difficult to cover all corner cases, and the threshold of expert knowledge is high.

From 2005 to 2015, the method was based on statistical model. Manual high cost local construction of gesture library;

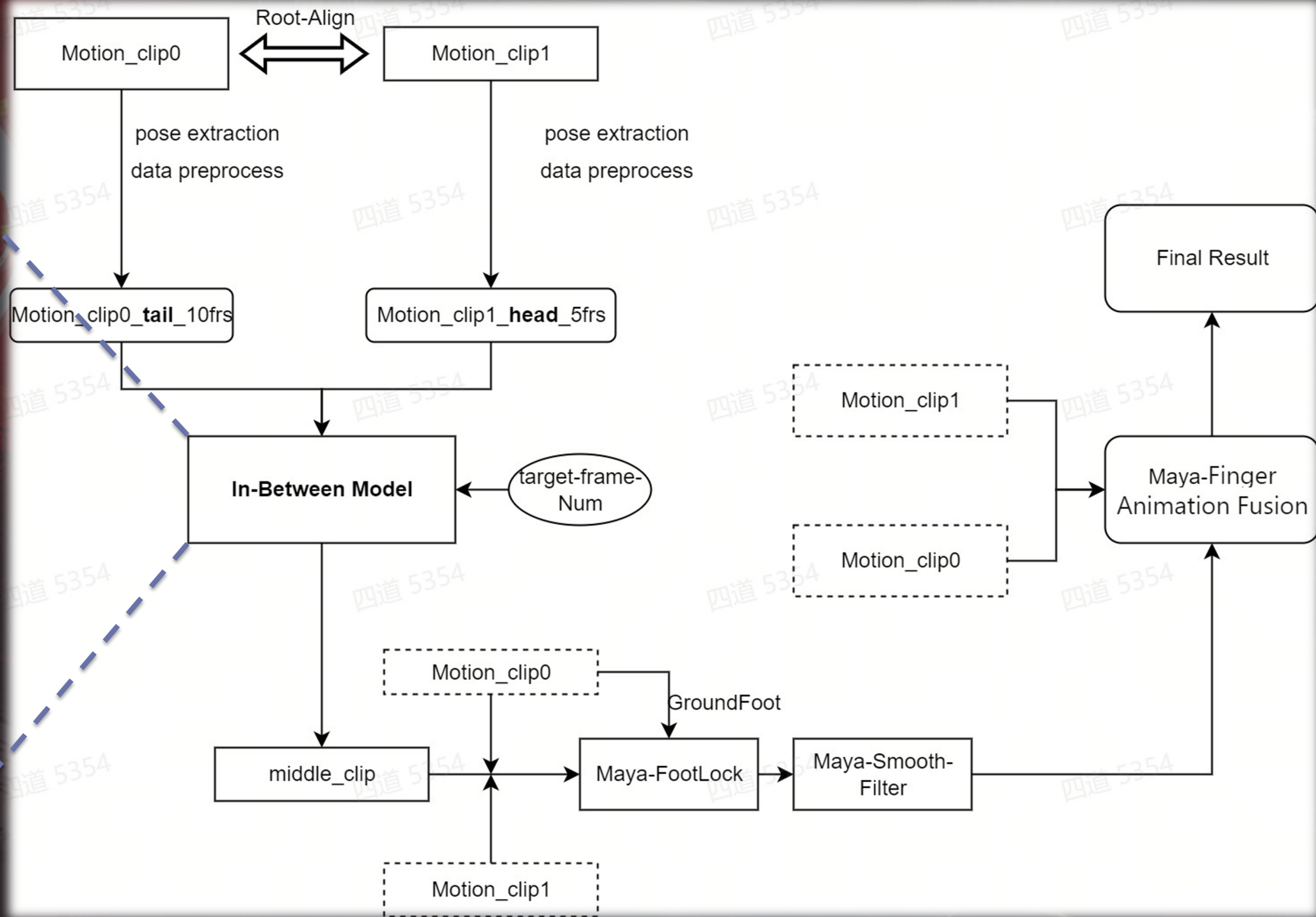
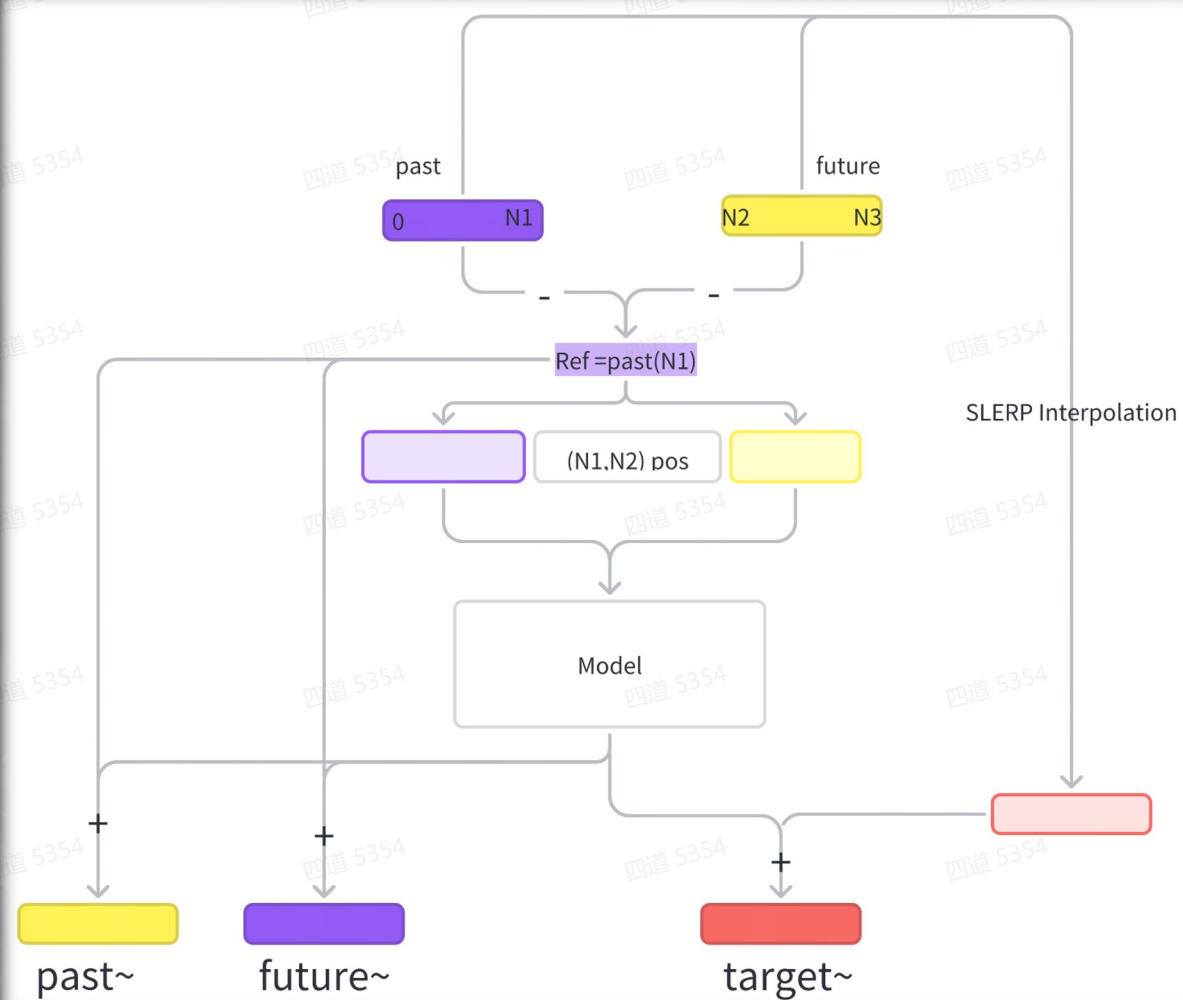
From 2018 to now, an end-to-end approach based on deep learning. However, the unexplainability and uncontrollability brought by the "black box" of neural network bring challenges to how to ensure accurate generation of rhythmic gestures and semantic gestures.

Body Motion Generation Module



AI-powered Motion In-betweening Module

In-betweening Model



AI-powerd Motion In-betweening Module

MOS: mean opinion score
5 points for full grading

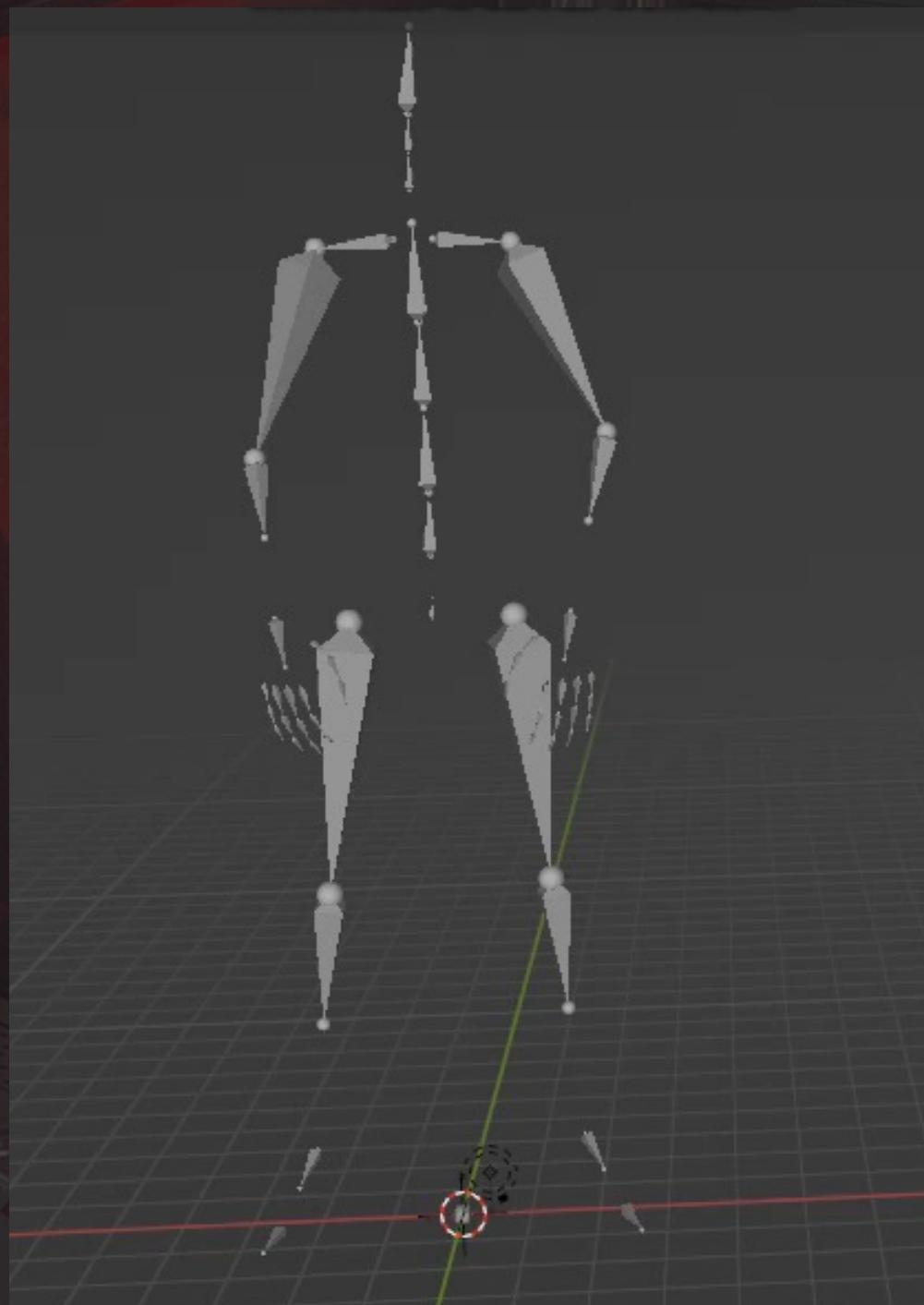
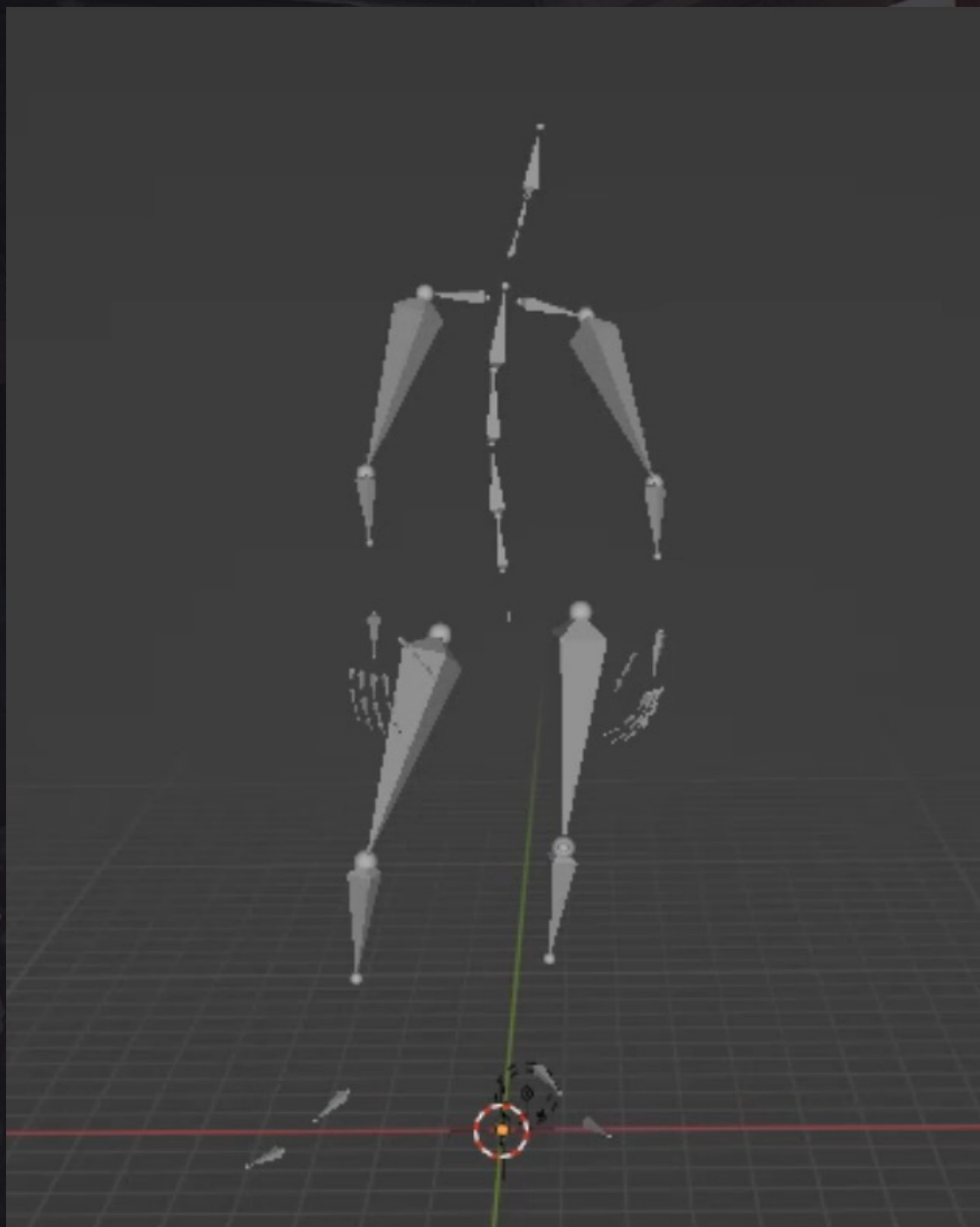
A:UE; B:NN Motionblend	Fluency	Stability
NN MotionBlend Significantly better than UE	MotionBlend (3.96) > UE (2.67)	MotionBlend (4.13) > UE (3.67)

original

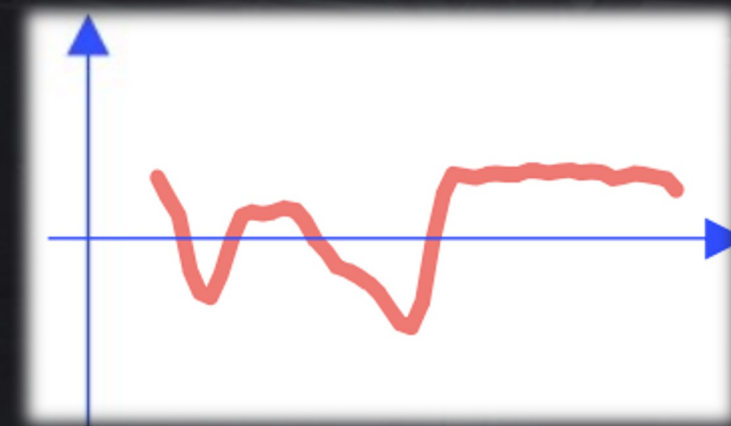


After optimization





(1) Rhythmic dynamics



(2) Rhythmic motion loss

$$\mathcal{L}_r = \left\| \widetilde{M}_i^* - (M_i - M_i^m) \right\|$$

Final Result





03 Application & Future



Game storyline Editor

Timeline

view

215.3

Dialogue_1_100181_TL (Timeline)

0 60 120 180 240 300 360 420 480 540 600 660 720 780 840 900 960 1020 1080 1140 1200 1260 1320 1380 1440 1500 1560

None (Signal Rec) Common

Dialogue Flag Track_2 Common

Camera (Cinema) Common

10002

NPC_10002 Common

Sound Action Track_5 Common

Caption Event Track_6 Common

NPC_10002 Common

NPC_10002 Common

10002 (t)

None (Transfc) Common

None (Transfc) Common

None (Transfc) Common

axis_double_L_3_normal

axis_double_L_3_nor...

axis_double_R_3_normal

axis_single_L_6_normal

axis_single_L_5_normal

axis_double_R...

axis_double_L_4_should...

axis_double_L_5_s...

JNAnimatorControlClip_0

JNAnimatorControlC...

JNAnimatorControlClip_2

JNAnimatorControlClip_3

JNAnimatorControlClip...

JNAnimatorCont...

SoundActionClip_0

SoundActionClip...

SoundActionClip_2

SoundActionClip_3

SoundActionClip_4

SoundAction

Exactly! He knows you're the only one who can ...

How was your n...

Calm down, soldier, <color=#ff7833>Who do you think

You were part of the <...>

Yeah, you've been ...

I've been looking for one pe

口型: Exactly! He knows you're the only one who...

口型: How was y...

口型: Calm down, soldier, <color=#f...

口型: You were part of ...

口型: Yeah, you've b...

口型: I've been looking for on

10002LookAt:0_0

10002L...

10002LookAt:0_2

10002LookAt:0_3

10002LookAt:0_4

10002LookAt:0_5

10002Looka...

10002LookAt:0_7

JNAnimatorControlClip_0

JNAnimatorControlC...

JNAnimatorControlClip_2

JNAnimatorControlClip_3

JNAnimatorControlCl...

JNAnimatorControlClip...

口型: 刘姐, 还没来得及自我介绍, 我叫陈雅, 你...

口型: 你为什么...

口型: 陈雅姐! 你参与了<color=#ff7833>...

口型: 我知道! 一定还有没...

口型: 你昨晚太久了.....

口型: 我一直在寻找弟弟的下落, 你...

10002LookAt:0_0

10002L...

10002LookAt:0_2

10002LookAt:0_3

10002LookAt:0_4

10002LookAt:0_5

10002Looka...

10002LookAt:0_7

对白类型

对白ID

生成语言

读取已经生成的Prefab

参与环境

场景

环境

刷新灯光效果

男主角

开启景深

角色信息

站位类型

固定站位

NPC列表

npc_10002

槽位0

删除

选中

贴地

avatar

槽位1

删除

选中

贴地

0 Default Or Null

添加

对白信息

精简模式

动作

情绪

角色ID

对白组ID

时长

语音

台词

Exactly! He knows you're the only one who can help him, and he's embarrassed to ask himself.

Hey, <color=#B2CFFF>[playerName]</color>!

How was your nap?

I woke from... cryostasis... My serial number is...

Calm down, soldier. <color=#ff7833>Who do you think woke you</color>?

You were part of the <color=#ff7833>Kindling Project</color> and...

...<color=#ff7833>Kindling</color>... Project?

更新选项

Manual Refresh Modules

7 Items

模块名称

Camera Shot

模块名称

Animation

模块名称

Audio

模块名称

Caption

模块名称

Mouth And Emotion

模块名称

Look At

模块名称

Move

辅助工具

设置固定镜头

刷新选项轨道

选中Timeline组件

J1对白资源检查

加载当前对白

批量导出动画轨

自动生成NPC资源

Look At开始时间轴 (本影片剪辑剪辑和身体动画)

生成对白

刷新对白

保存

提交

Application





下一段

Pause

下一句

The call of data by the system backend



Hourly request volume of TTS



Hourly request volume of full-body animation

- Approximately 68 hours animation and speech a week
- Over 100 NPCS, multiple games
- It takes less than 800ms by one T4 GPU to generate speech and compute full body animations for interactions that last between 10 and 20 seconds.

Future Work : More Natural Performance



AI-powered Role-playing NPCs

3D Avatar

a beautiful blonde
with green eyes and pink lips



Speech&Animation

Text to Lifelike Speech and
Animation



NPC Behavior

Behavior generation based on
text description



Unscripted dialogue & Story generation

魔皇突然倒退了一步，伸手摸了一下身上的伤口，然后朝着四周的虚空一指，原本在地下的虚空魔法阵直接消失无踪，这里顿时恢复了活性。
“来了……”魔皇喃喃的自语着，然后死死的盯着魔王的背影，不敢有丝毫的怠慢，在一个黑袍大汉的带领下，朝着人类的方向搜寻了过去。
几分钟之后，在前面的虚空魔法阵眼前一空，所有的人都露了出来，大家一个个的瞪大了眼睛。
“原来是魔皇大人。”听到魔皇的声音后，所有的人都感到惊讶，因为在他们眼中，这就是魔皇大人的魔导师，终于踏上了辉煌的魔法师之路，按理来说，能够拥有神力的魔导师是当之无愧的人类。
但是此刻魔皇的身影却消失在这片大海之中，似乎早就看穿了这一切，这让众人的心都悬了起来。
片刻之后，众人感觉到了空气中的清爽，所有人都猛的抬起头来，那巨大的眼睛里充满了向往，这样的表现让他们觉得此行不枉。
他们不|

LLM



Thank You