QCon
全球软件开发大会

InfoQ 极客传媒

语音大模型：从级联到端到端

杨学锐

阶跃星辰语音负责人

2025.10

目录

InfoQ 极客传媒

北京

QCon 全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

北京

AiCon 全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

上海

QCon 全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月 — 5月 — 6月 — 8月 — 10月 — 12月

上海

AiCon 全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

深圳

AiCon 全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

北京

AiCon 全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOps
- 具身智能

极客邦科技 2025 年会议规划

促进软件开发及相关领域知识与创新的传播

参会咨询

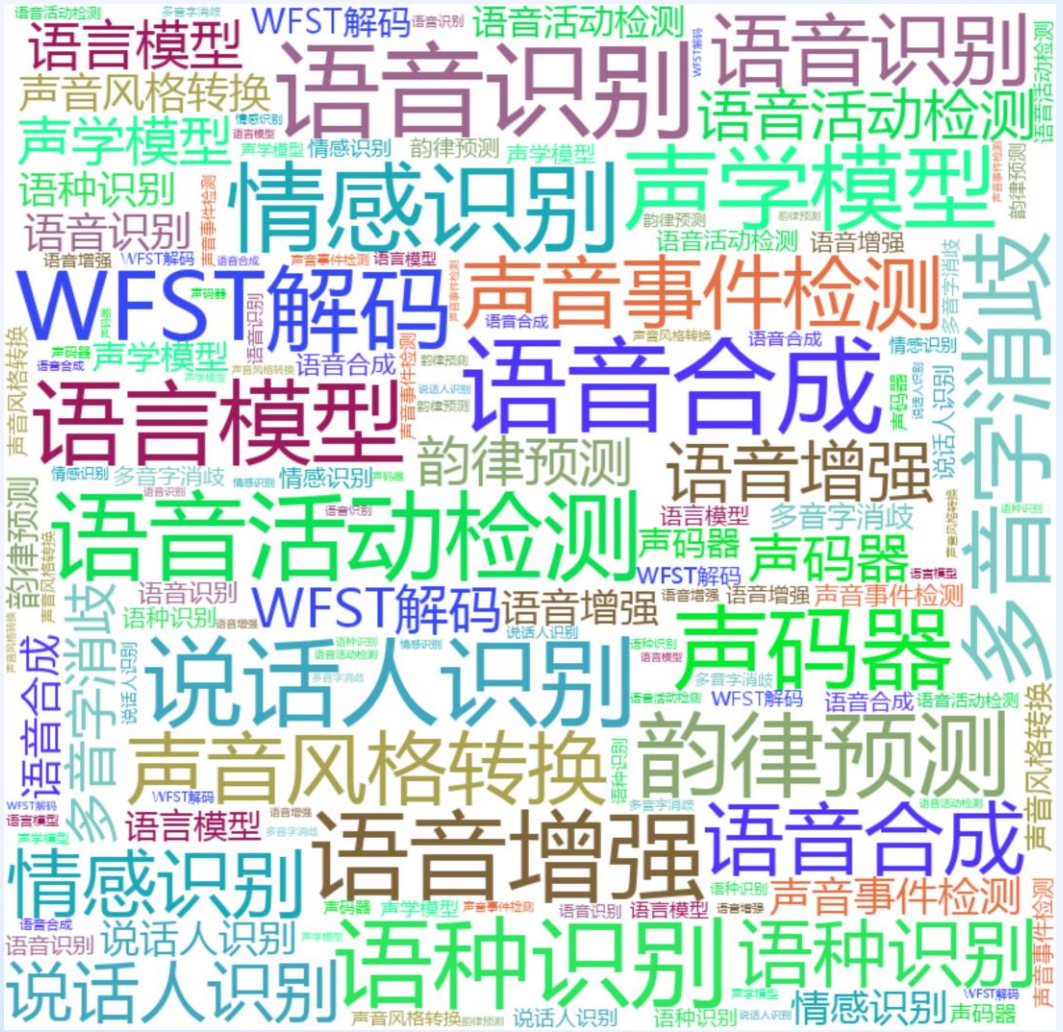
查看会议

01

LLM如何重塑语音技术

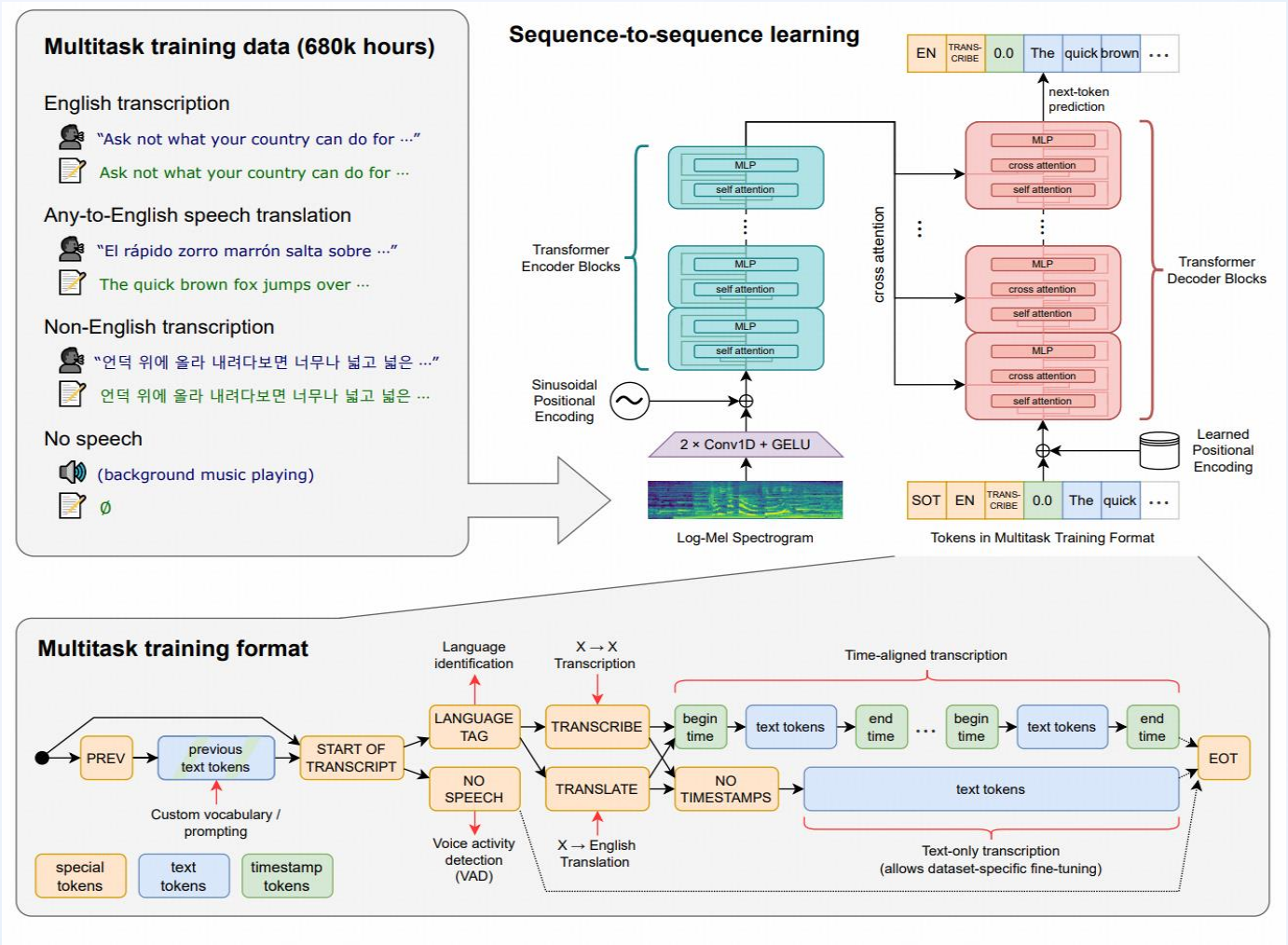
LLM如何重塑语音技术

- 传统语音技术
  - **流水线式架构：** 误差逐级传递，信息流失；
  - **理解：** 只能处理简单指令，无法进行多轮、多模态的深度推理；
  - **表达：** TTS声音机械，韵律模板化，没有真正的“人感”；



LLM如何重塑语音技术

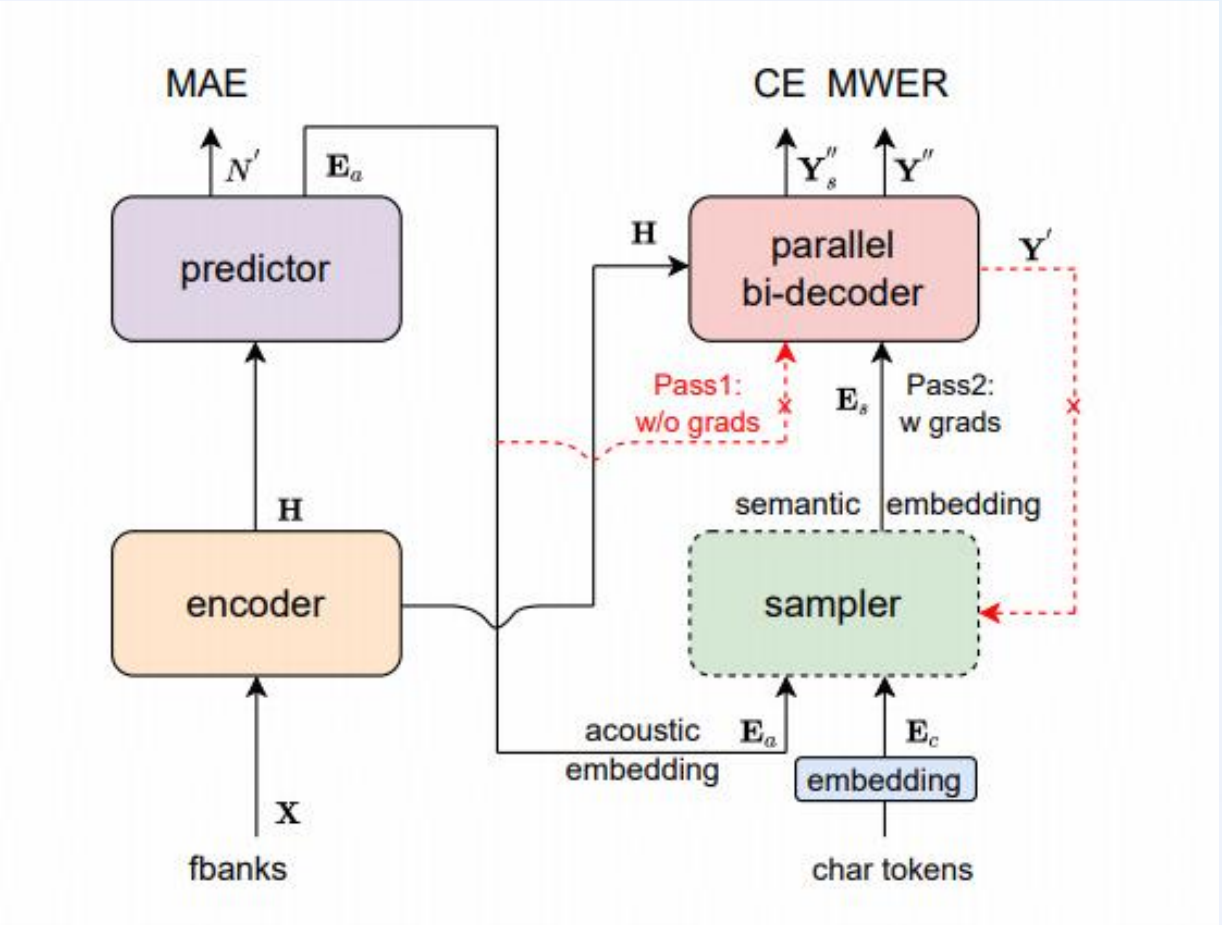
- ASR
  - Open-source
    - Whisper*
    - SenseVoice
    - FireredASR
  - Close-source/API
    - SeedASR
    - StepASR



*Radford A, Kim J W, Xu T, et al. Robust speech recognition via large-scale weak supervision[C]

LLM如何重塑语音技术

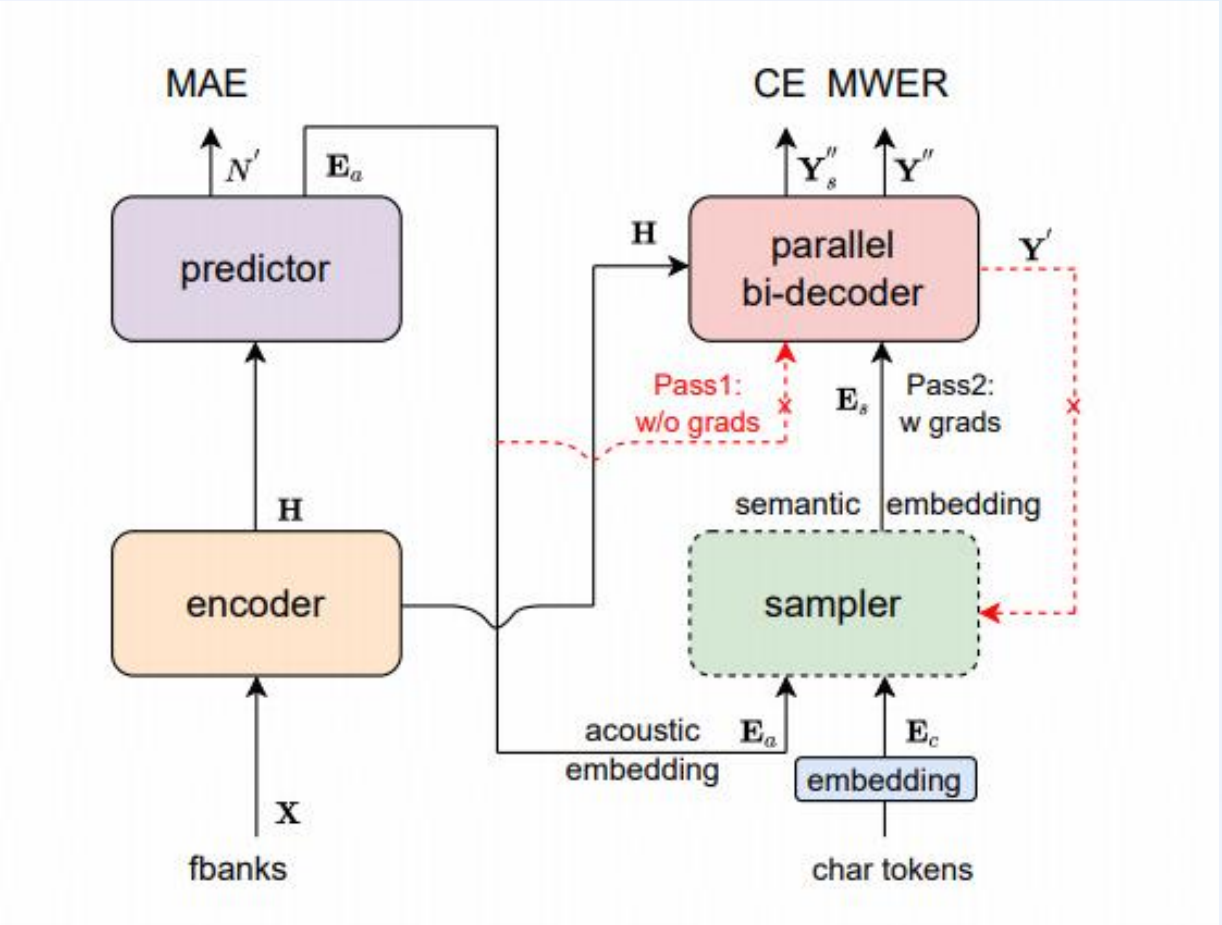
- ASR
  - Non-LLM
    - Paraformer*
  - Open-source
    - Whisper
    - SenseVoice
    - FireredASR
  - Close-source/API
    - SeedASR
    - StepASR



*Gao Z, Zhang S, McLoughlin I, et al. Paraformer: Fast and accurate parallel transformer for non-autoregressive end-to-end speech recognition[J].

LLM如何重塑语音技术

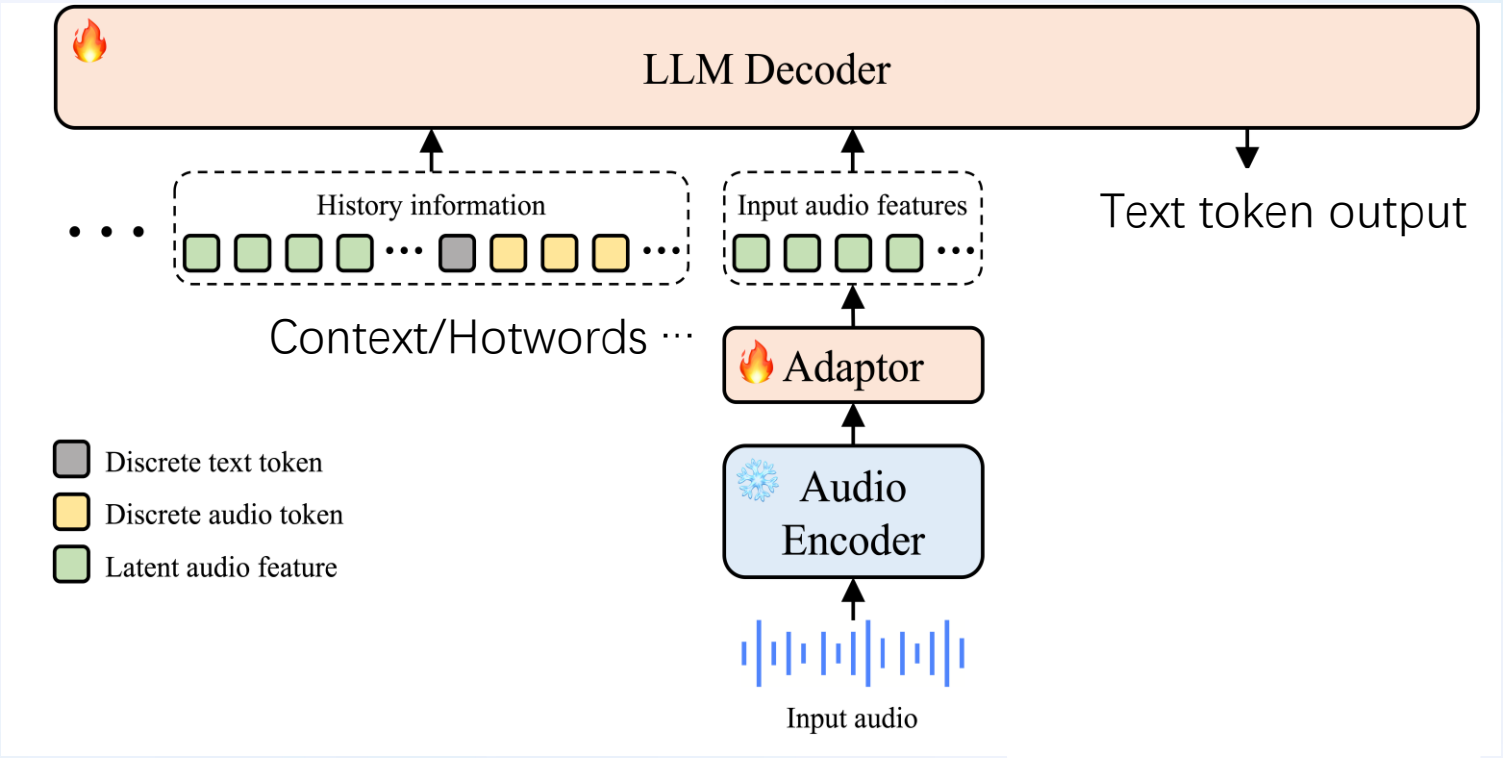
- ASR
  - Non-LLM
    - Paraformer*
  - Open-source
    - Whisper
    - SenseVoice
    - FireredASR
  - Close-source/API
    - SeedASR
    - StepASR



*Gao Z, Zhang S, McLoughlin I, et al. Paraformer: Fast and accurate parallel transformer for non-autoregressive end-to-end speech recognition[J].

LLM如何重塑语音技术

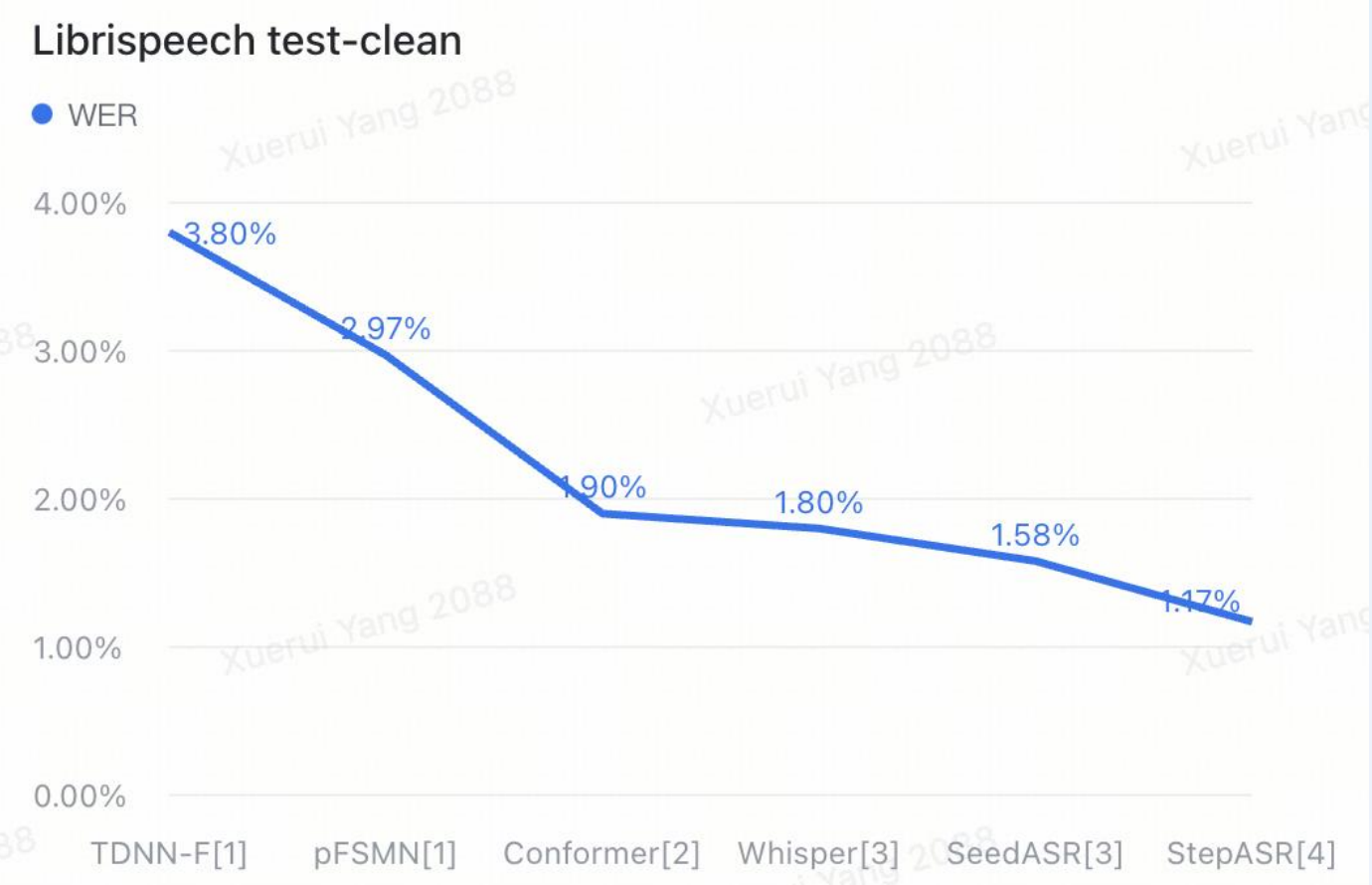
- ASR
  - Open-source
    - Whisper
    - SenseVoice
    - FireredASR
  - Close-source/API
    - SeedASR
    - StepASR*



*https://platform.stepfun.com/docs/api-reference/audio/transcriptions

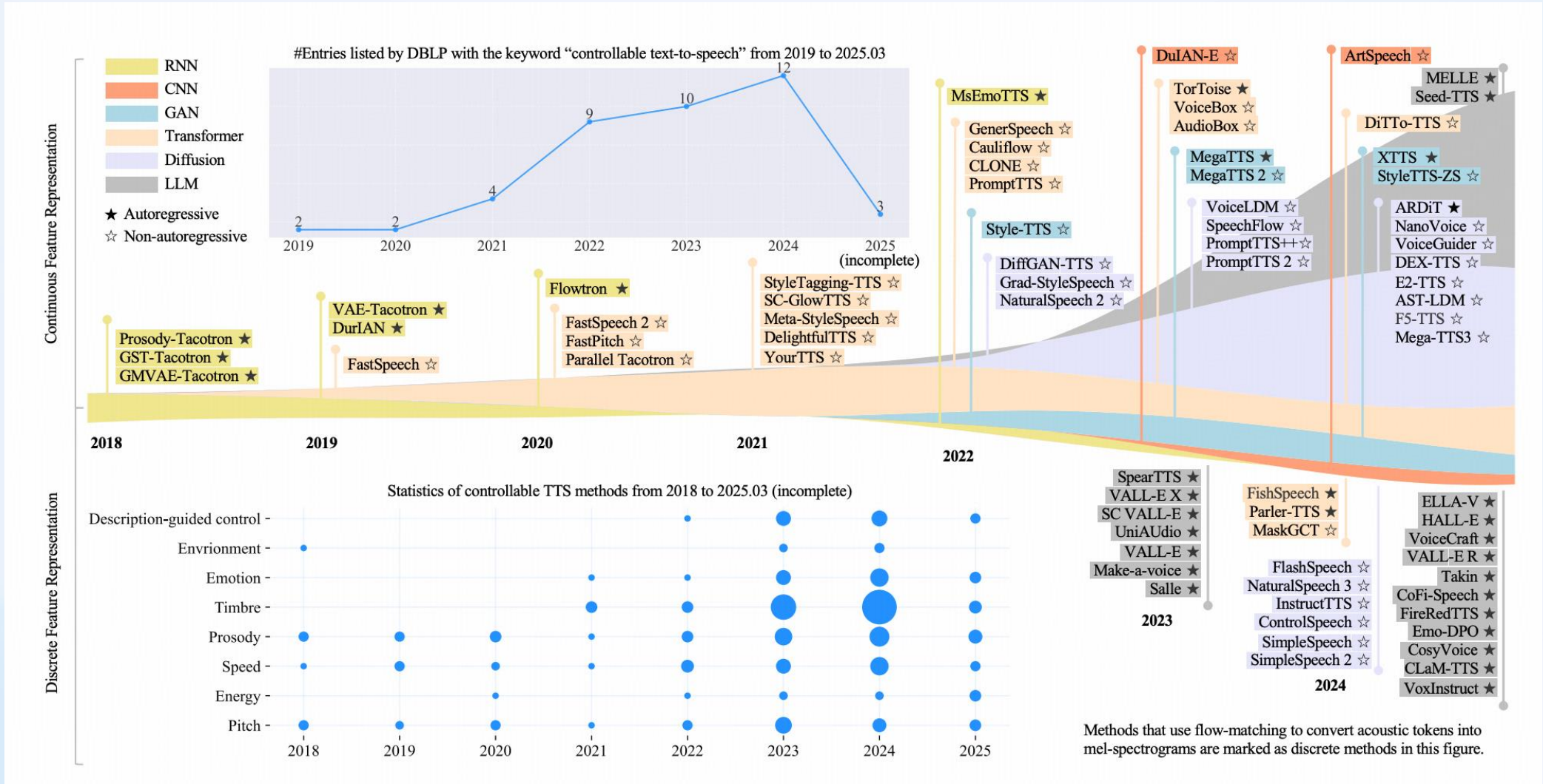
LLM如何重塑语音技术

- ASR
  - Open-source
    - Whisper
    - SenseVoice-L
    - FireredASR
  - Close-source/API
    - SeedASR
    - StepASR



[1] Yang X, Li J, Zhou X. A novel pyramidal-FSMN architecture with lattice-free MMI for speech recognition[J]. arXiv preprint arXiv:1810.11352, 2018.
[2] Gulati A, Qin J, Chiu C C, et al. Conformer: Convolution-augmented transformer for speech recognition[J]. arXiv preprint arXiv:2005.08100, 2020.
[3] Bai Y, Chen J, Chen J, et al. Seed-asr: Understanding diverse speech and contexts with llm-based speech recognition[J]. arXiv preprint arXiv:2407.046
[4] Wu B, Yan C, Hu C, et al. Step-audio 2 technical report[J]. arXiv preprint arXiv:2507.16632, 2025.

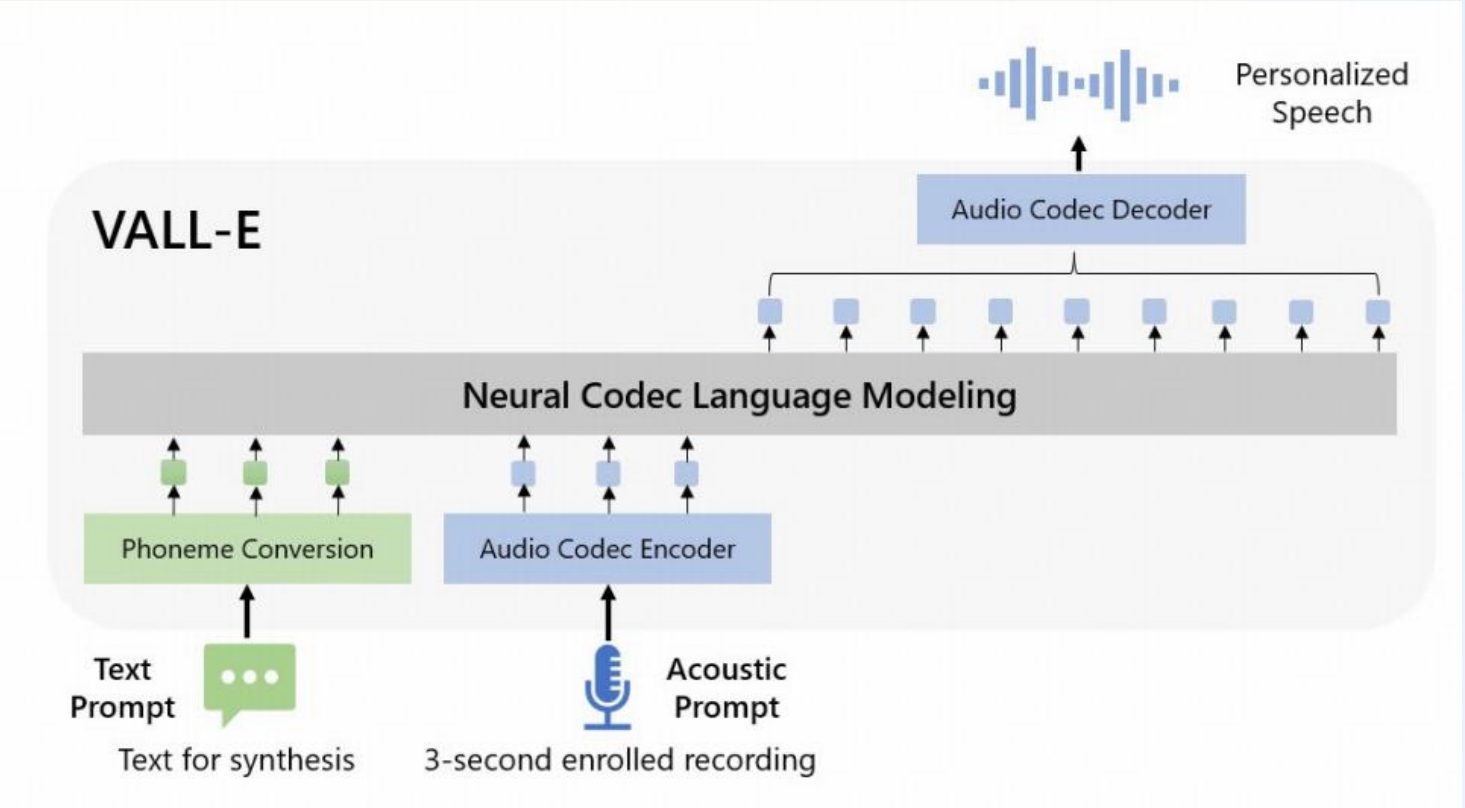
LLM如何重塑语音技术



*Xie T, Rong Y, Zhang P, et al. Towards controllable speech synthesis in the era of large language models: A survey[J]. arXiv e-prints, 2024: arXiv: 2412.06602.

LLM如何重塑语音技术

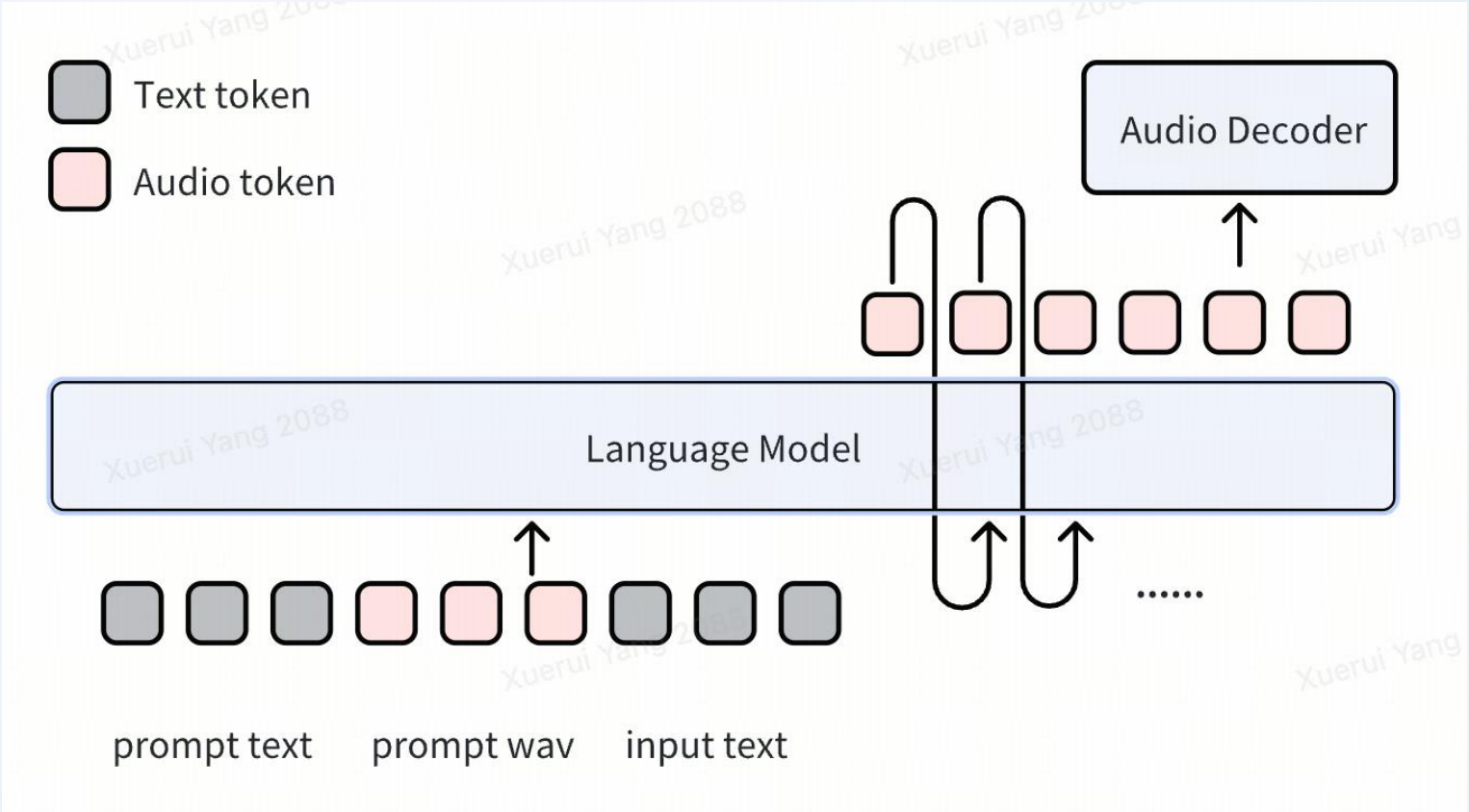
- TTS
  - NAR
    - FastSpeech
    - NatualSpeech
    - E2-TTS
  - AR（LLM）
    - VALLE*
    - CosyVoice
    - Minimax-Speech
    - StepTTS
    - DiTAR
    - VibeVoice



Zhang, Ziqiang, et al. "Speak foreign languages with your own voice: Cross-lingual neural codec language modeling." *arXiv preprint arXiv:2303.03926* (2023).

LLM如何重塑语音技术

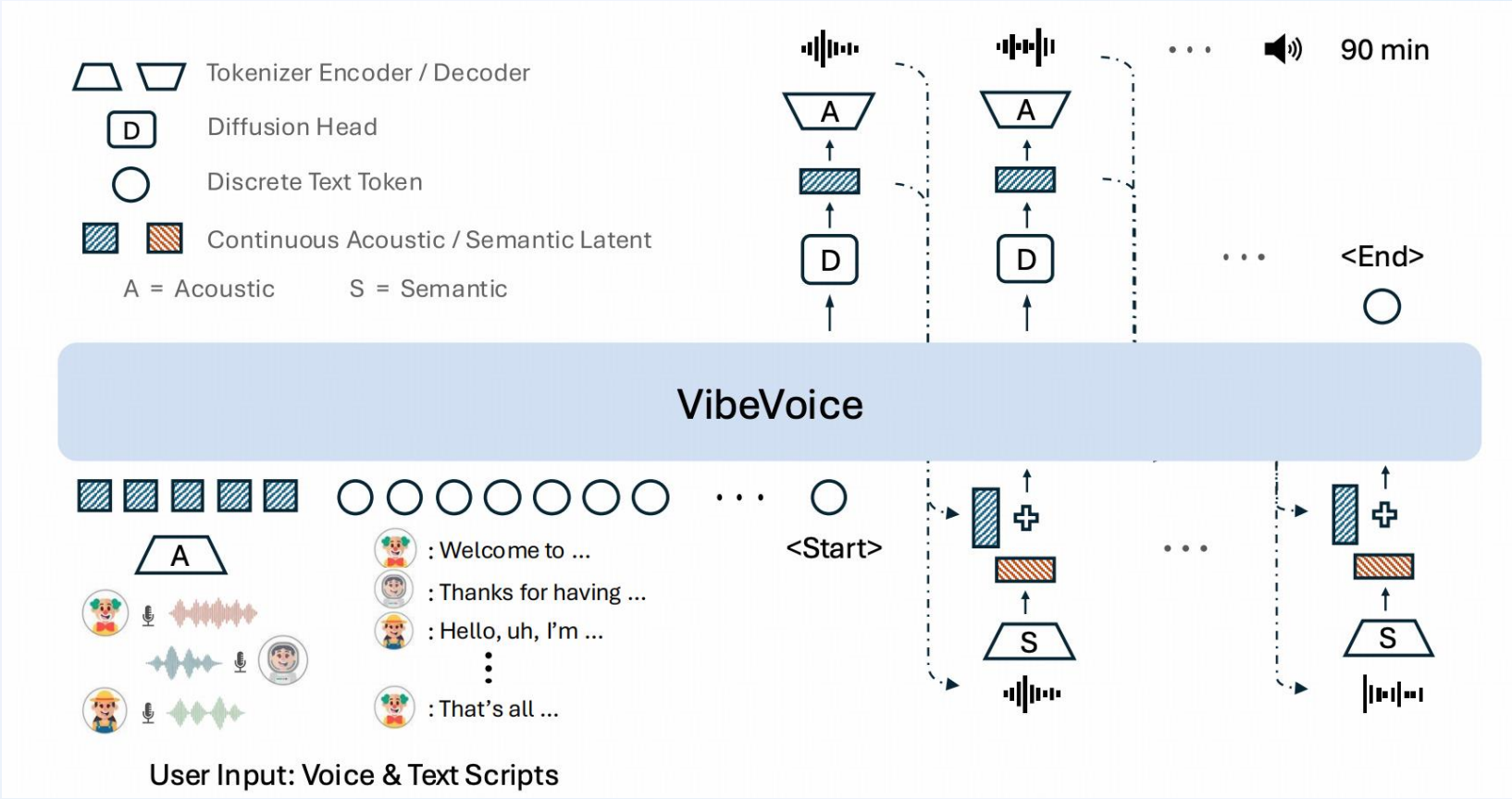
- TTS
  - NAR
    - FastSpeech
    - NatualSpeech
    - E2-TTS
  - AR（LLM）
    - VALLE
    - CosyVoice
    - Minimax-Speech
    - StepTTS*
    - DiTAR
    - VibeVoice



*https://platform.stepfun.com/docs/api-reference/audio/create_audio

LLM如何重塑语音技术

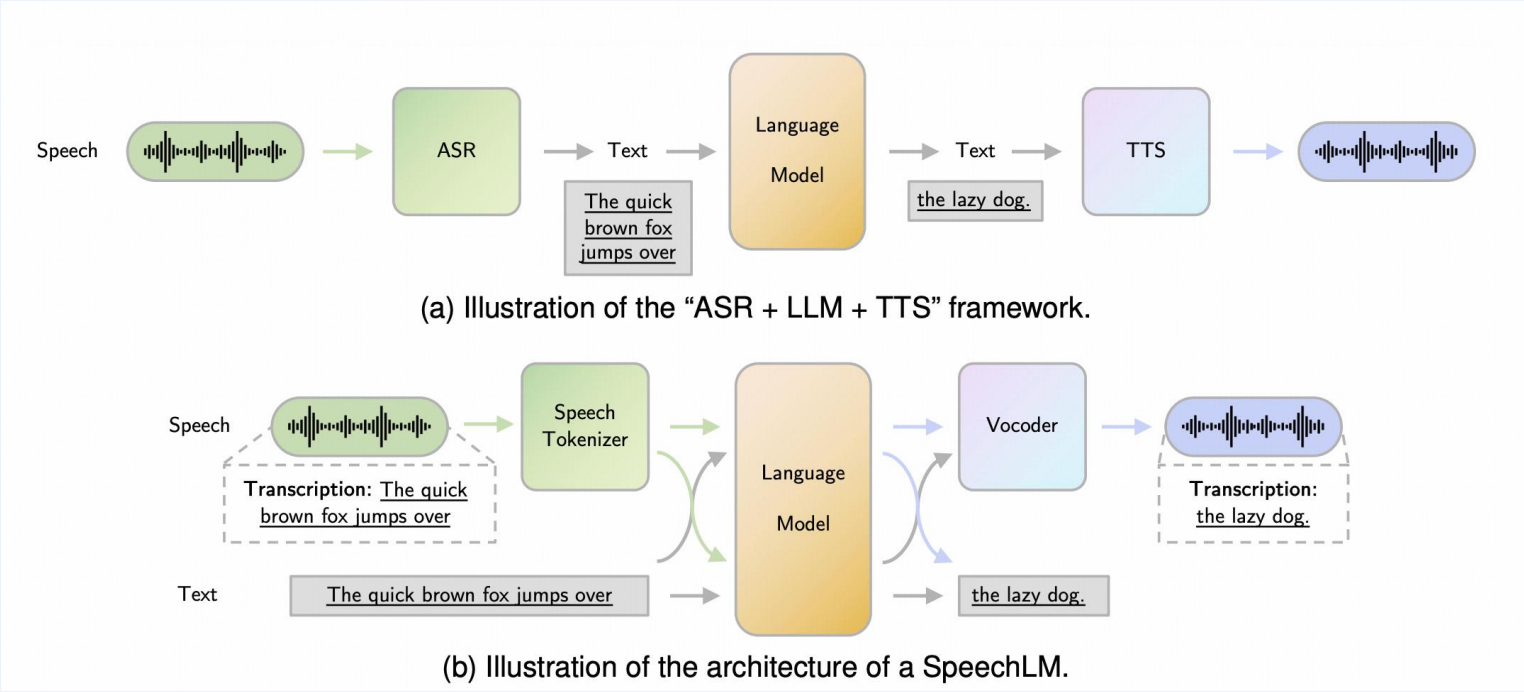
- TTS
  - NAR
    - FastSpeech
    - NatualSpeech
    - E2-TTS
  - AR（LLM）
    - VALLE
    - CosyVoice
    - Minimax-Speech
    - StepTTS
    - DiTAR
    - VibeVoice*



*Peng Z, Yu J, Wang W, et al. Vibevoice technical report[J]. arXiv preprint arXiv:2508.19205, 2025.

LLM如何重塑语音技术

- 对话/语音交互

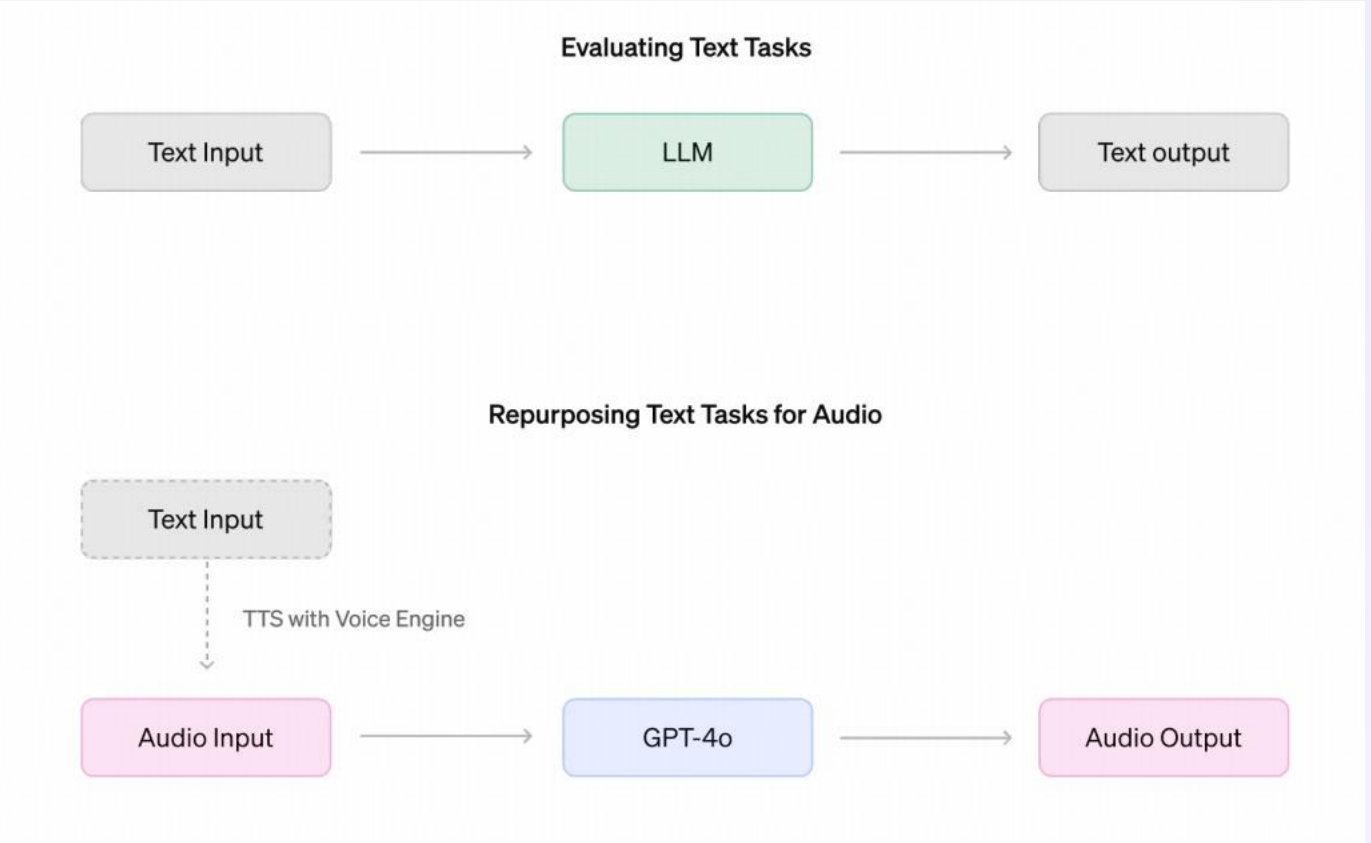


- 是否有一种端到端架构，能够实现理解与生成一体化？
  - YES

*Cui W, Yu D, Jiao X, et al. Recent advances in speech language models: A survey[J]. arXiv preprint arXiv:2410.03751, 2024.

LLM如何重塑语音技术

- GPT-4o



*Hurst A, Lerer A, Goucher A P, et al. Gpt-4o system card[J]. arXiv preprint arXiv:2410.21276, 2024.

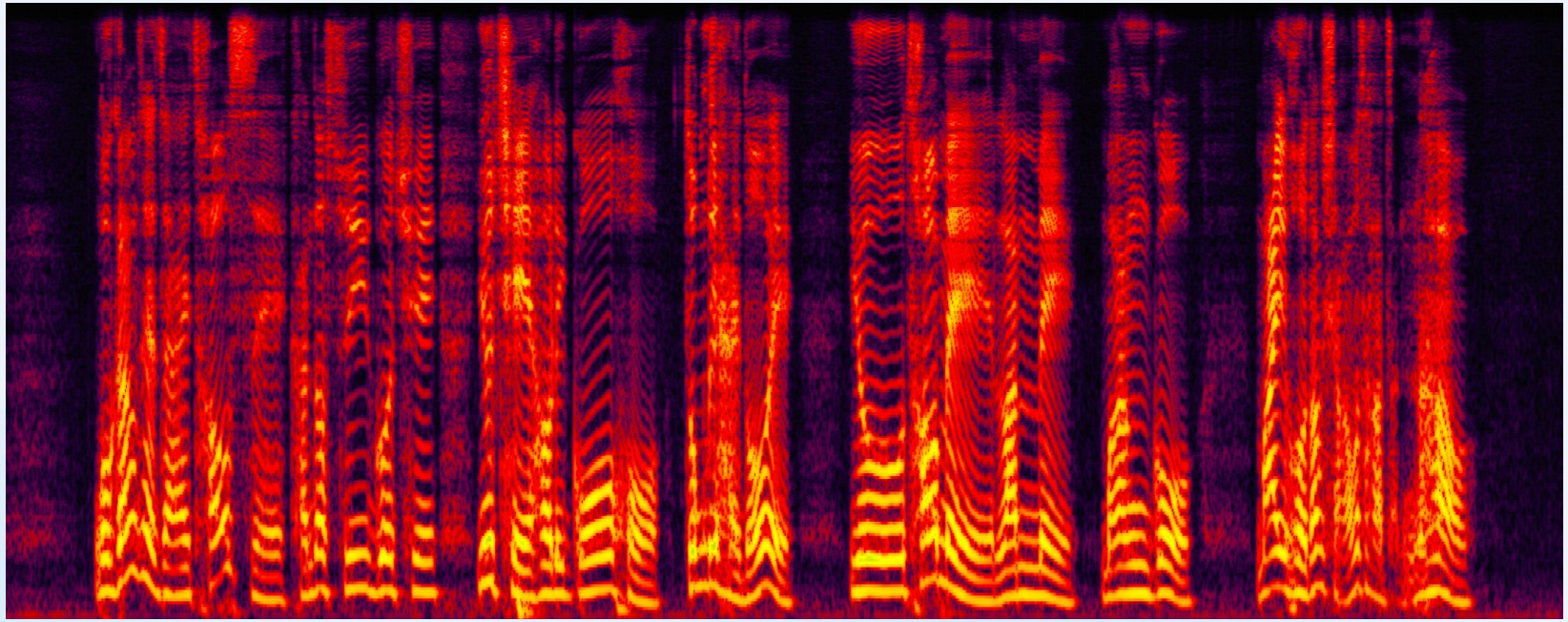
02

如何构建端到端语音模型

表征

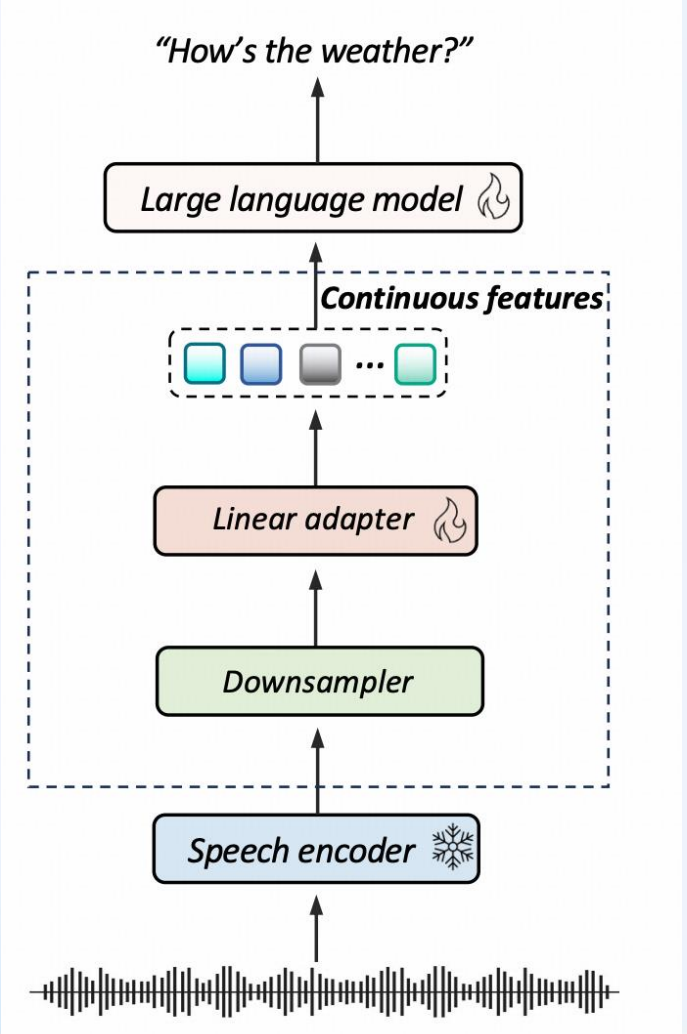
如何在大模型中表征语音与音频信号

- Mel-Spectrogram
  - 模数信号转换->分帧加窗->时频转换->梅尔滤波器组->取对数



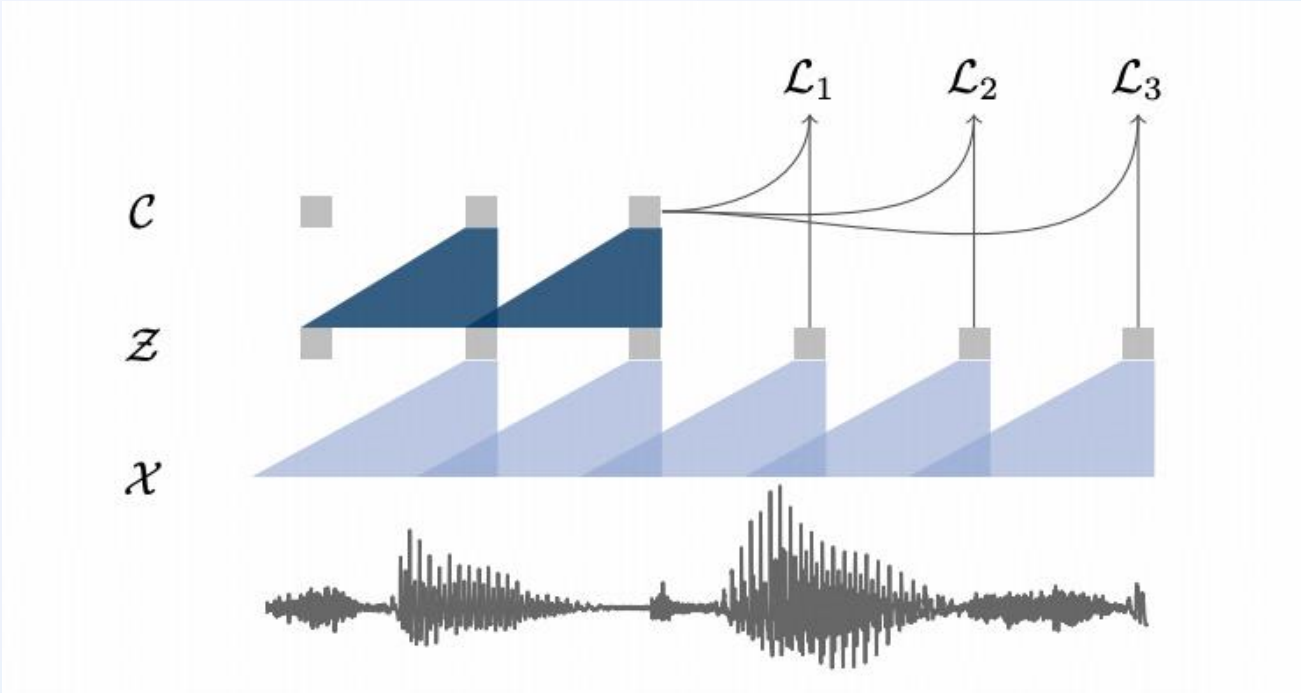
如何在大模型中表征语音与音频信号

- **Continuous Features**
  - Wav2Vec
  - HuBERT
  - WavLM
  - Whisper Encoder



如何在大模型中表征语音与音频信号

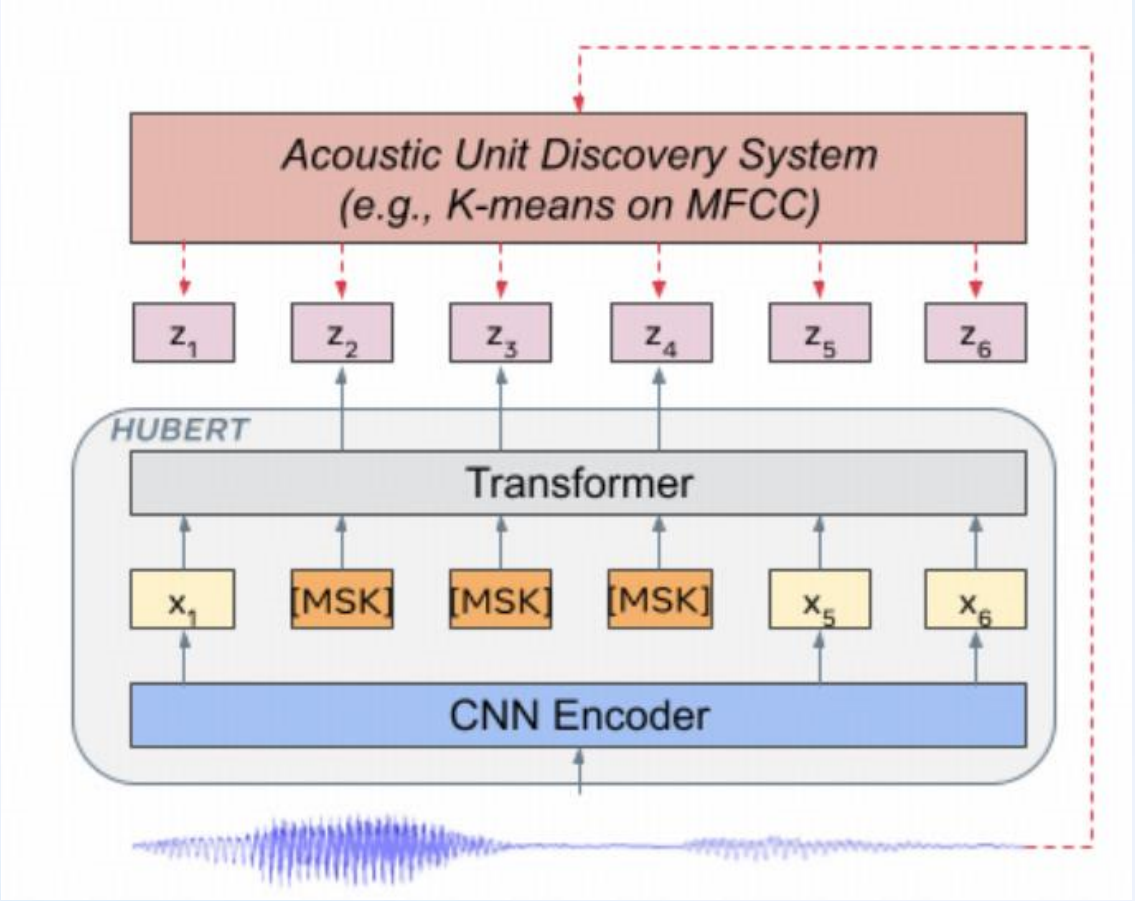
- **Continuous Features**
  - Wav2Vec*
  - HuBERT
  - WavLM
  - Whisper Encoder



*Baevski A, Zhou Y, Mohamed A, et al. wav2vec 2.0: A framework for self-supervised learning of speech representations[J]. Advances in neural information processing systems, 2020, 33: 12449-12460.

如何在大模型中表征语音与音频信号

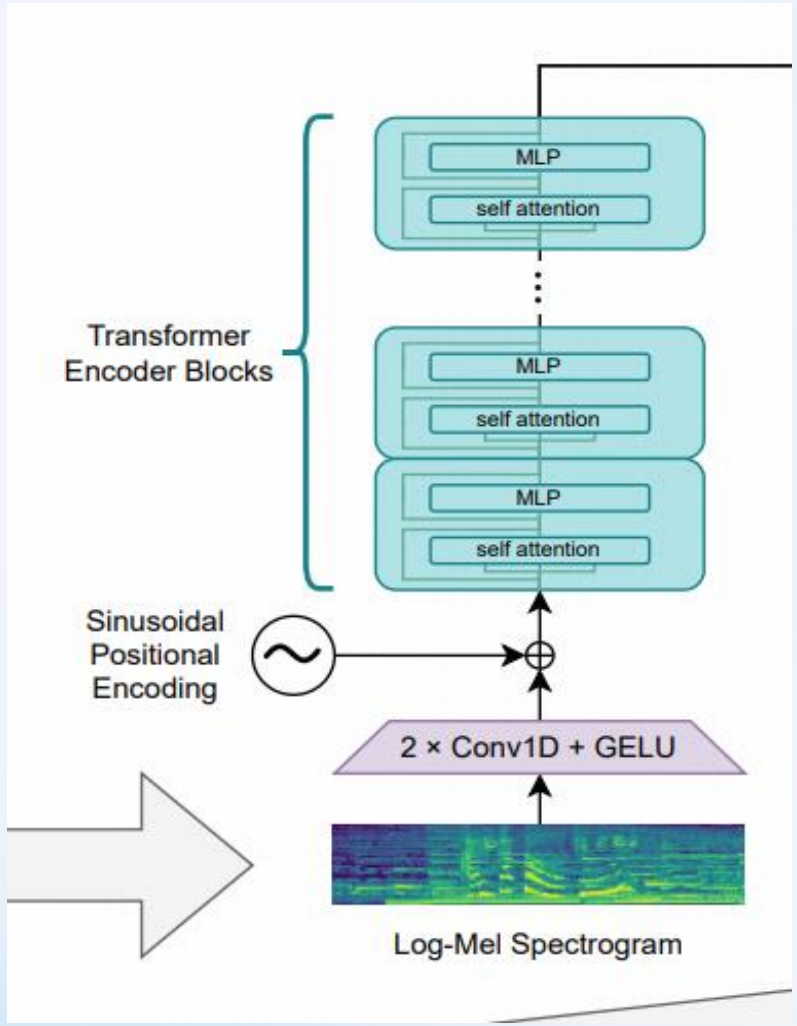
- **Continuous Features**
  - Wav2Vec
  - HuBERT*
  - WavLM
  - Whisper Encoder



*Hsu W N, Bolte B, Tsai Y H H, et al. Hubert: Self-supervised speech representation learning by masked prediction of hidden units[J]. IEEE/ACM transactions on audio, speech, and language processing, 2021, 29: 3451-3460.

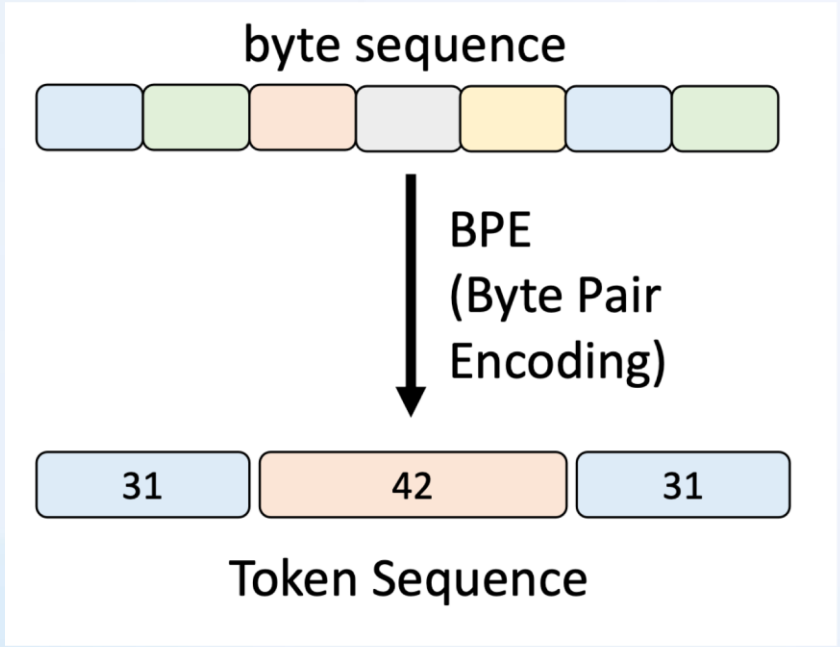
如何在大模型中表征语音与音频信号

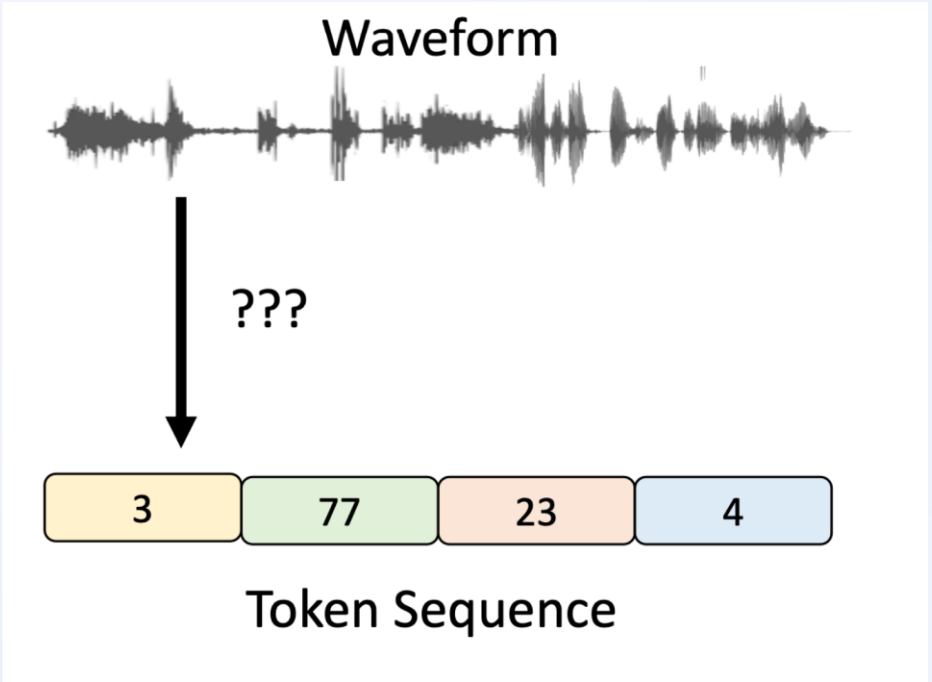
- **Continuous Features**
  - Wav2Vec
  - HuBERT
  - WavLM
  - Whisper Encoder*



如何在大模型中表征语音与音频信号

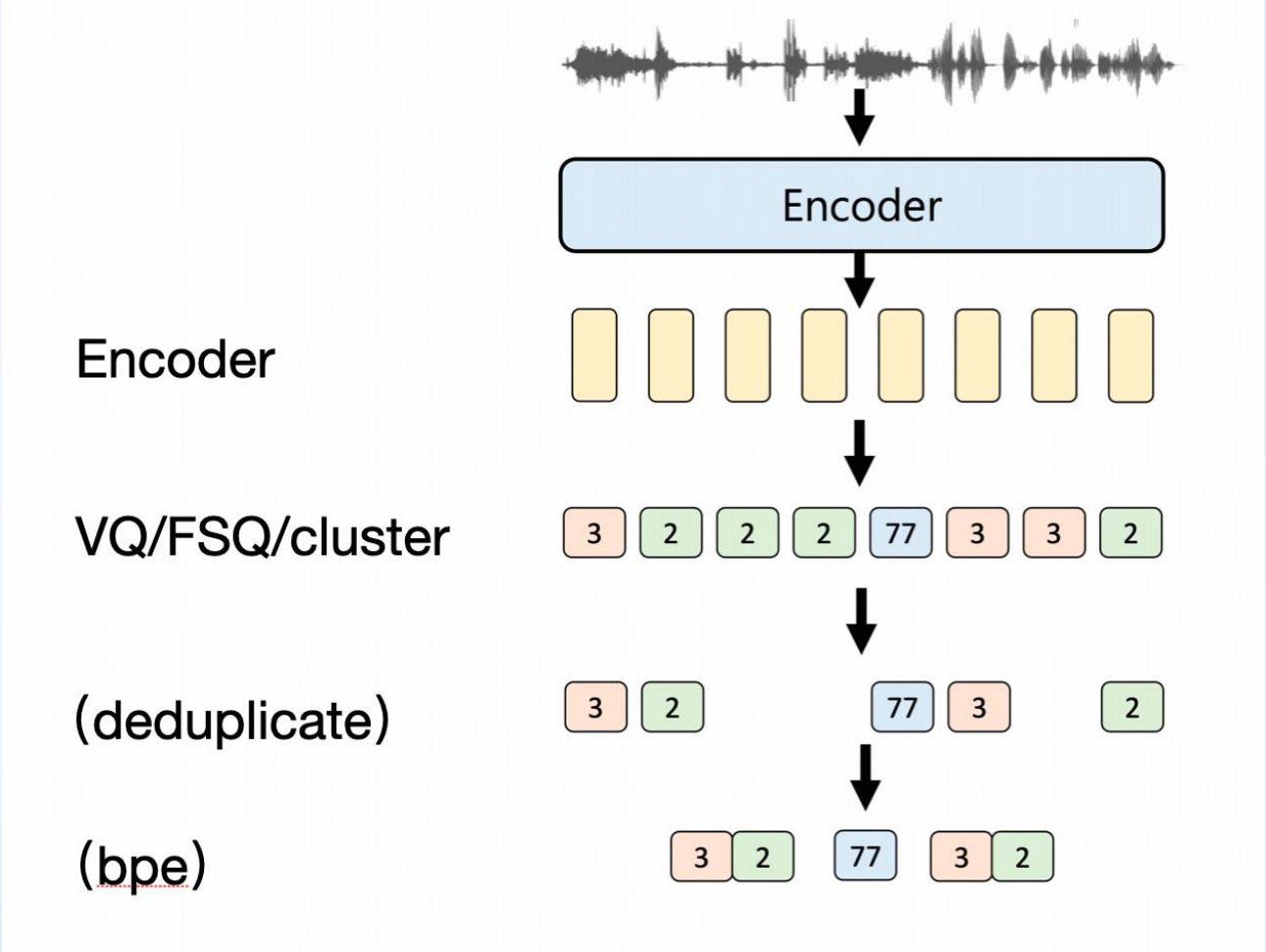
- LLM text token





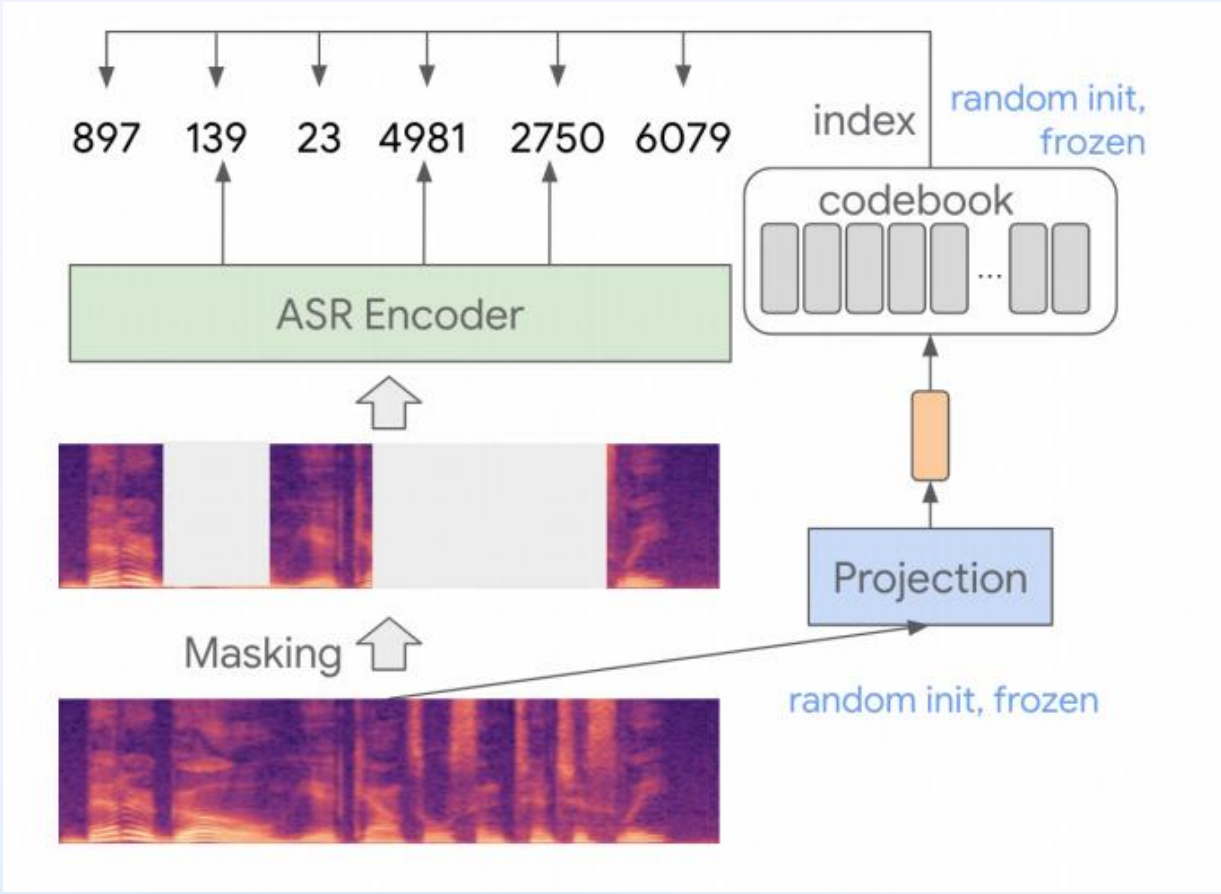
如何在大模型中表征语音与音频信号

- **Discrete Token**
  - Semantic Token
  - Acoustic Token
  - Unified Token
  - Other（Pitch/Style）



如何在大模型中表征语音与音频信号

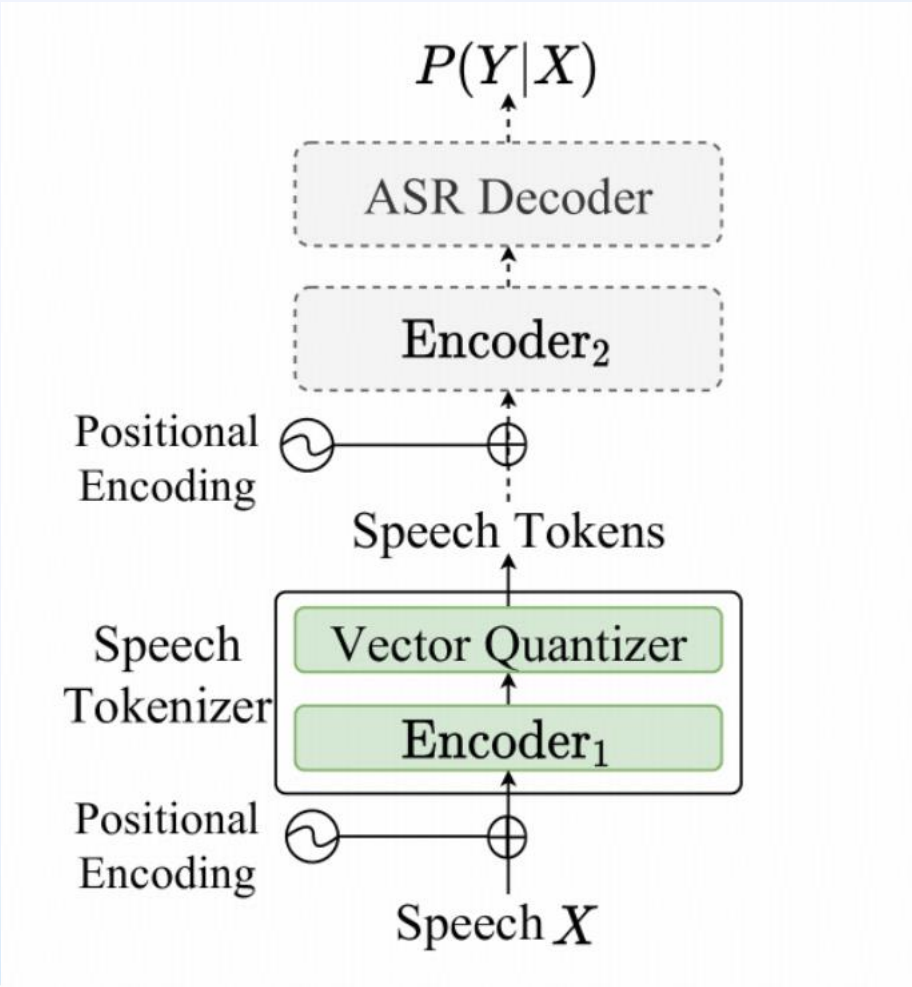
- **Semantic Token**
  - Wav2Vec
  - HuBERT
  - WavLM
  - BEST-RQ*
  - S3Tokenizer



*Chiu C C, Qin J, Zhang Y, et al. Self-supervised learning with random-projection quantizer for speech recognition[C]//International Conference on Machine Learning. PMLR, 2022: 3915-3924.

如何在大模型中表征语音与音频信号

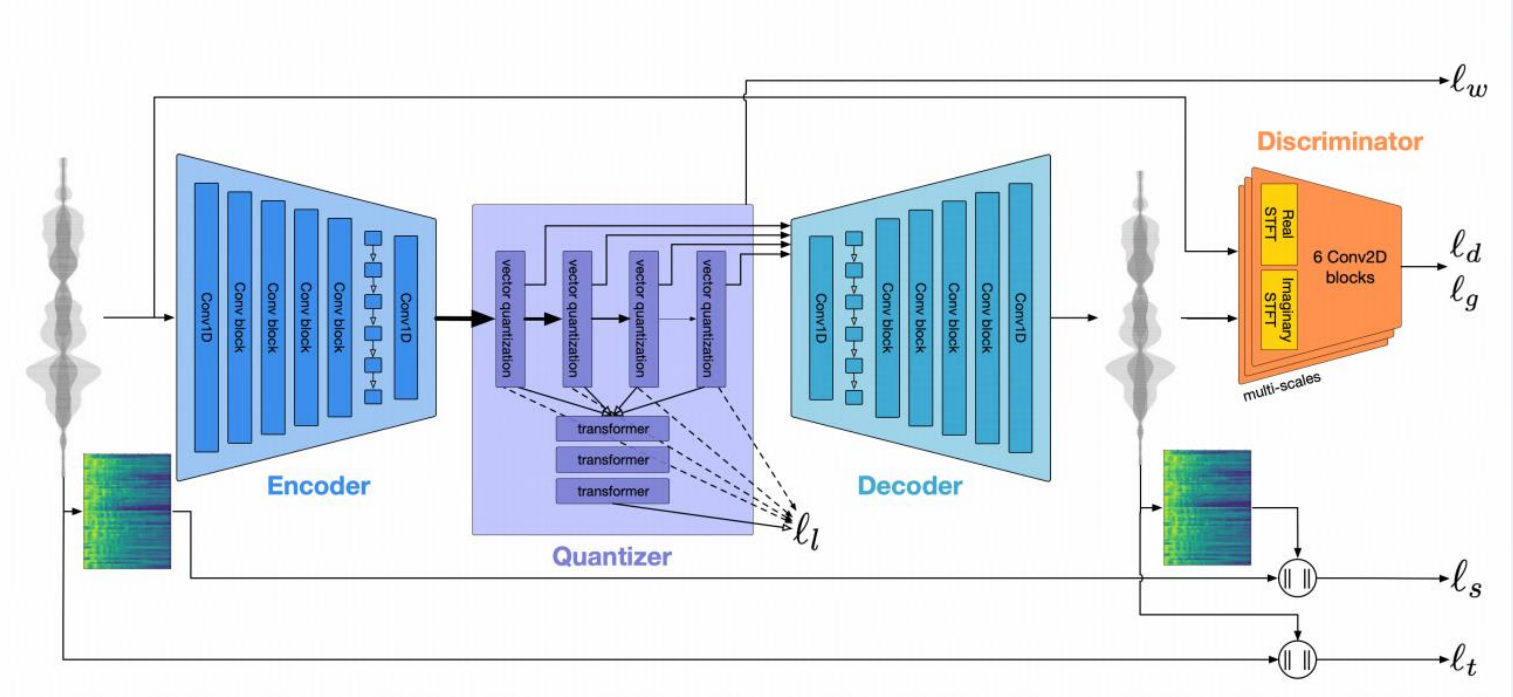
- **Semantic Token**
  - Wav2Vec
  - HuBERT
  - WavLM
  - BEST-RQ
  - S3Tokenizer*



*Du Z, Chen Q, Zhang S, et al. Cosyvoice: A scalable multilingual zero-shot text-to-speech synthesizer based on supervised semantic tokens[J]. arXiv preprint arXiv:2407.05407, 2024.

如何在大模型中表征语音与音频信号

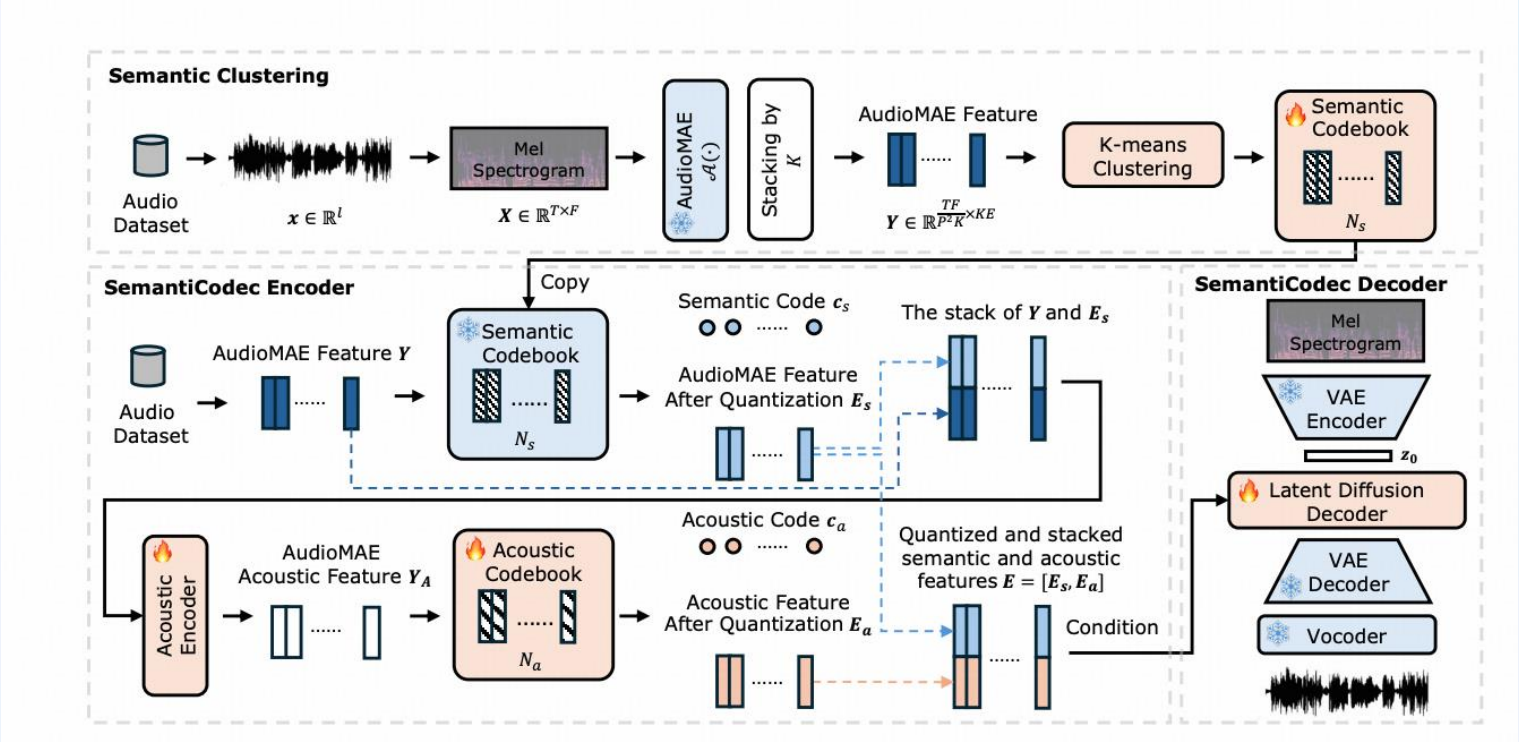
- **Acoustic Token**
  - Soundstream
  - Encodec*
  - DAC
  - FACodec
  - WavTokenizer



*Défossez A, Copet J, Synnaeve G, et al. High fidelity neural audio compression[J]. arXiv preprint arXiv:2210.13438, 2022.

如何在大模型中表征语音与音频信号

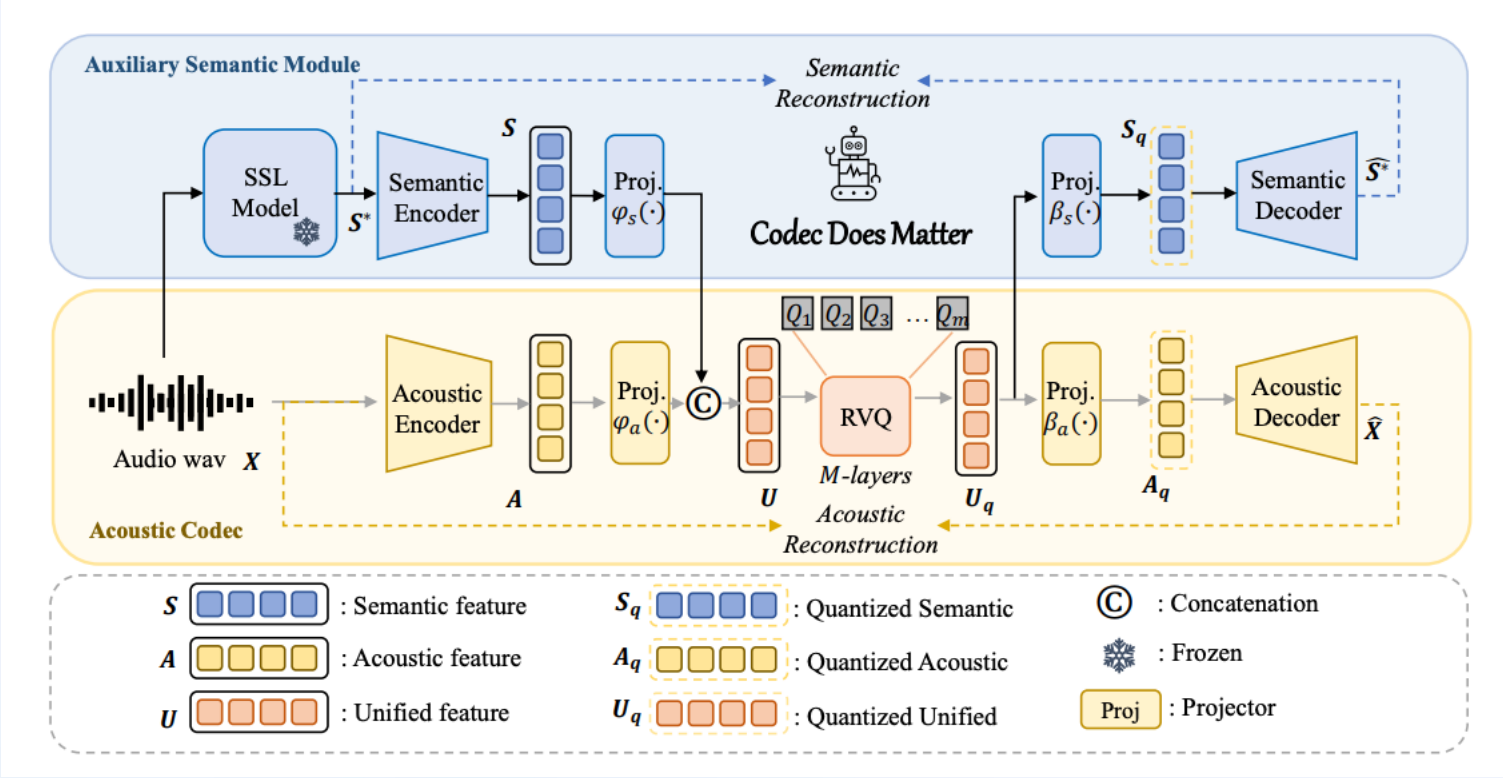
- **Unified Token**
  - SpeechTokenizer
  - SemantiCodec*
  - X-Codec
  - XY-Tokenizer
  - UniCodec
  - Mimo-Tokenizer



*Liu H, Xu X, Yuan Y, et al. Semanticodec: An ultra low bitrate semantic audio codec for general sound[J]. IEEE Journal of Selected Topics in Signal Processing, 2024.

如何在大模型中表征语音与音频信号

- **Unified Token**
  - SpeechTokenizer
  - SemantiCodec
  - X-Codec*
  - XY-Tokenizer
  - UniCodec
  - Mimo-Tokenizer



*Ye Z, Sun P, Lei J, et al. Codec does matter: Exploring the semantic shortcoming of codec for audio language model[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(24): 25697-25705.

如何在大模型中表征语音与音频信号

- 对比
  - 输入侧
  - 输出侧：离散表征生成更稳定，但需要diffusion补充细节；连续表征预测目标过于平滑。

SSL model	*Token type*	ASR (WER↓)			PR (PER↓)	ST (BLEU↑)	KS (ACC↑)	IC (ACC↑)	ER (ACC↑)
		LS (clean|other)	GigaSpeech (test)	CHiME4 (test)	LS-100 (clean)	GigaST (En-Zh|En-De)	SC (test|val)	SLURP (test)	IEMOCAP (test)
Qwen 1.5-0.5B									
HuBERT	Discrete	4.56 / 9.79	19.40	13.35	9.69	22.75 / 20.14	93.70 / 93.85	57.04	38.65
WavLM	Discrete	4.72 / 10.45	16.34	12.94	9.64	24.62 / 21.22	92.87 / 92.45	59.96	37.98
HuBERT	Continuous	4.91 / 6.43	17.45	8.62	12.84	26.63 / 25.42	95.38 / 95.70	76.84	56.72
WavLM	Continuous	2.92 / 4.61	13.96	8.68	12.62	29.44 / 28.12	97.76 / 97.36	81.35	59.45
Llama 3.1-8B									
HuBERT	Discrete	2.56 / 6.49	12.86	10.56	7.85	26.64 / 25.13	96.75 / 96.69	63.44	39.84
WavLM	Discrete	2.96 / 7.48	13.35	9.13	**7.02**	28.62 / 26.87	97.92 / 98.17	66.96	36.12
HuBERT	Continuous	1.76 / 4.58	9.04	5.72	9.83	32.02 / 34.42	**99.74 / 98.59**	**86.84**	64.54
WavLM	Continuous	**1.65 / 4.22**	**8.86**	**5.43**	10.44	**35.17 / 37.20**	97.34 / 98.26	85.57	**65.87**

*Wang D, Li J, Cui M, et al. Speech discrete tokens or continuous features? a comparative analysis for spoken language understanding in speechllms[J]. arXiv preprint arXiv:2508.17863, 2025.

02

如何构建端到端语音模型

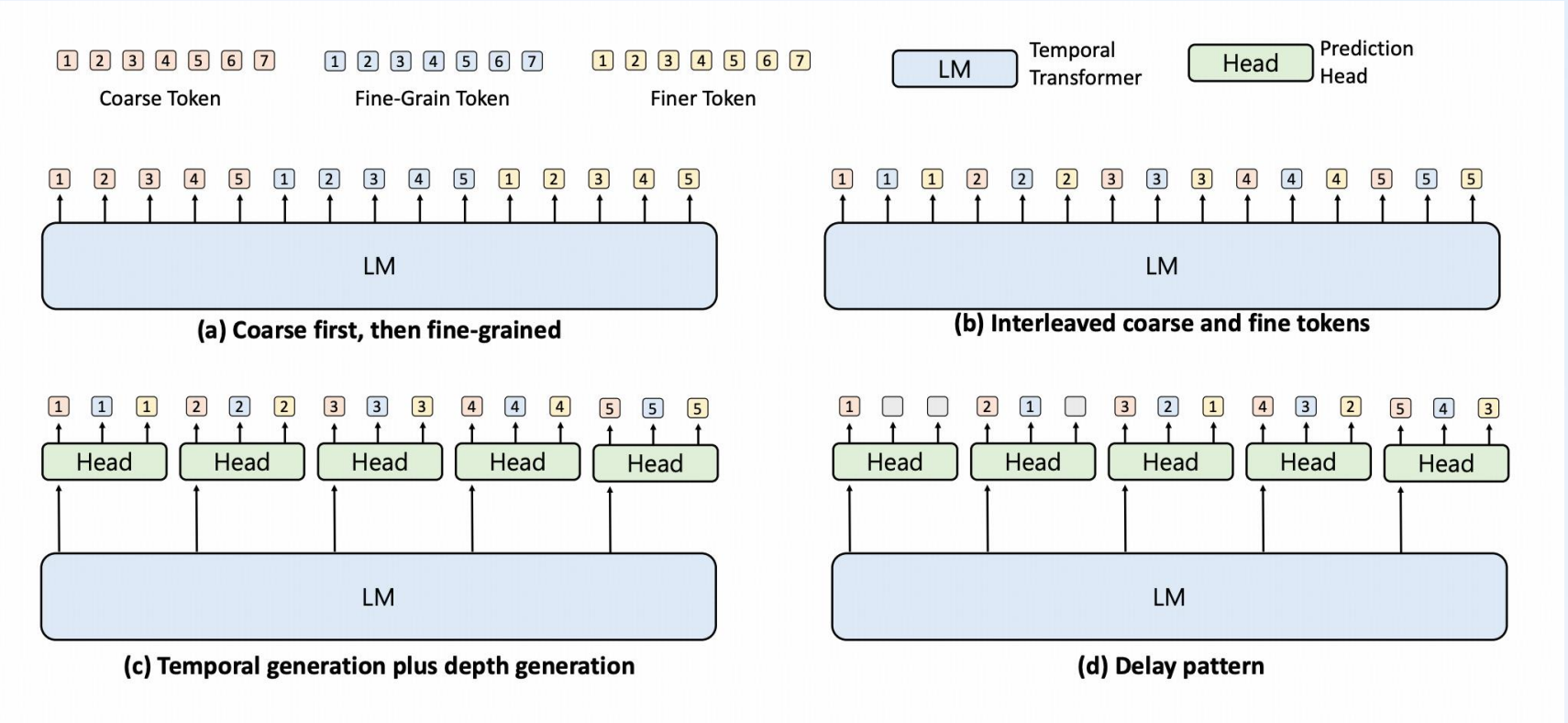
架构

模型架构

- 理解侧
  - Continuous Features
  - Discrete Token（Single Codebook）
  - Discrete Token（Multi Codebook）

模型架构

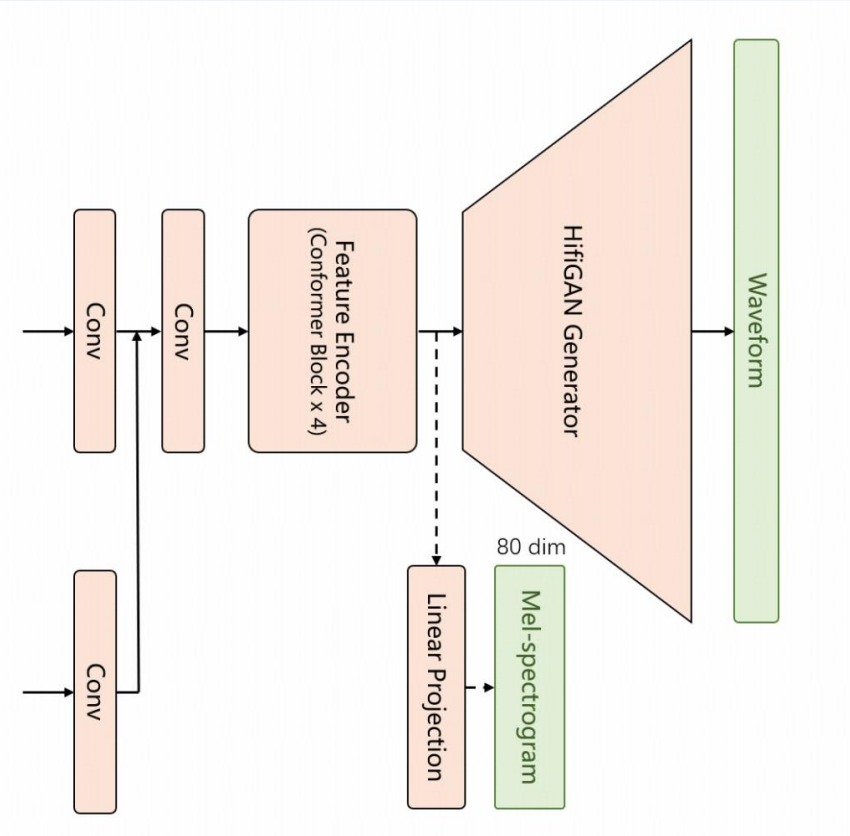
- **Token生成策略**
  - Naive
  - Multihead
  - Interleaved

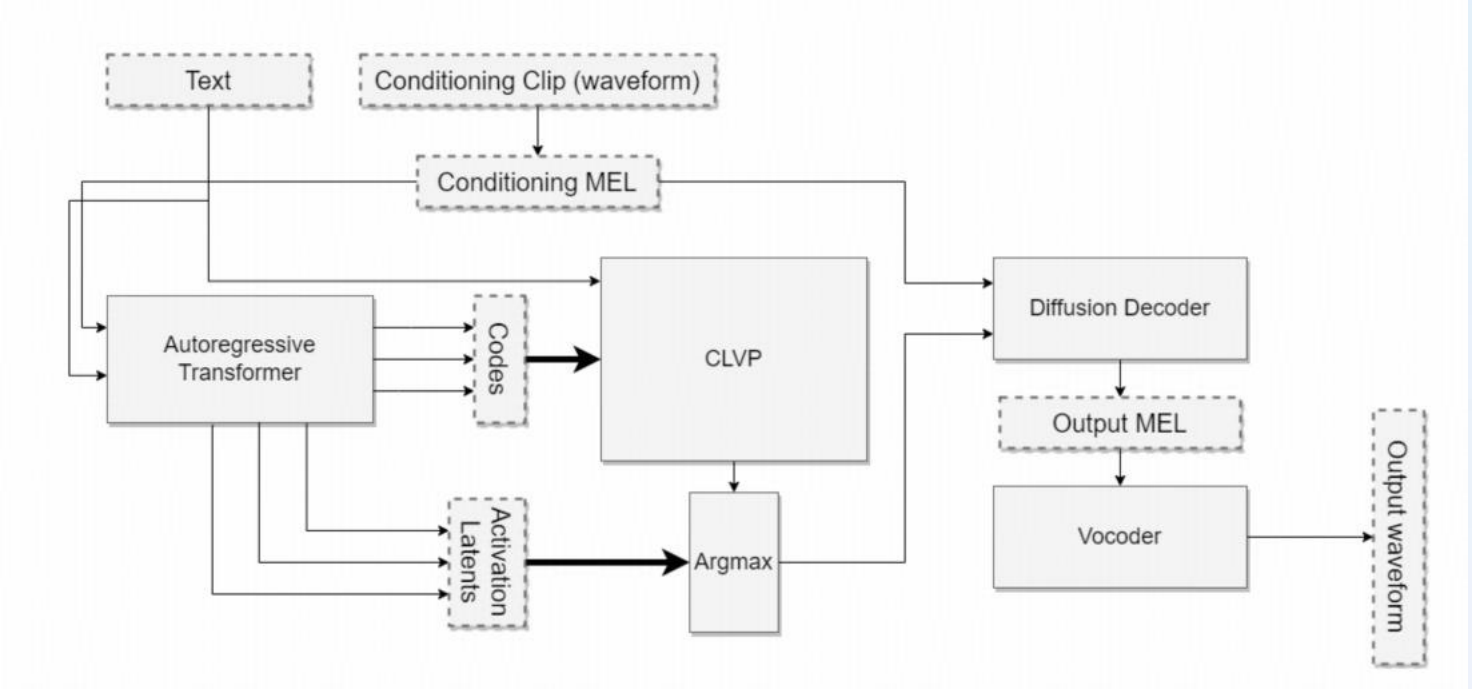


*Arora S, Chang K W, Chien C M, et al. On the landscape of spoken language models: A comprehensive survey[J]. arXiv preprint arXiv:2504.08528, 2025.

模型架构

- Token2Wav/Audio Detokenizer/Vocoder/Speech Decoder

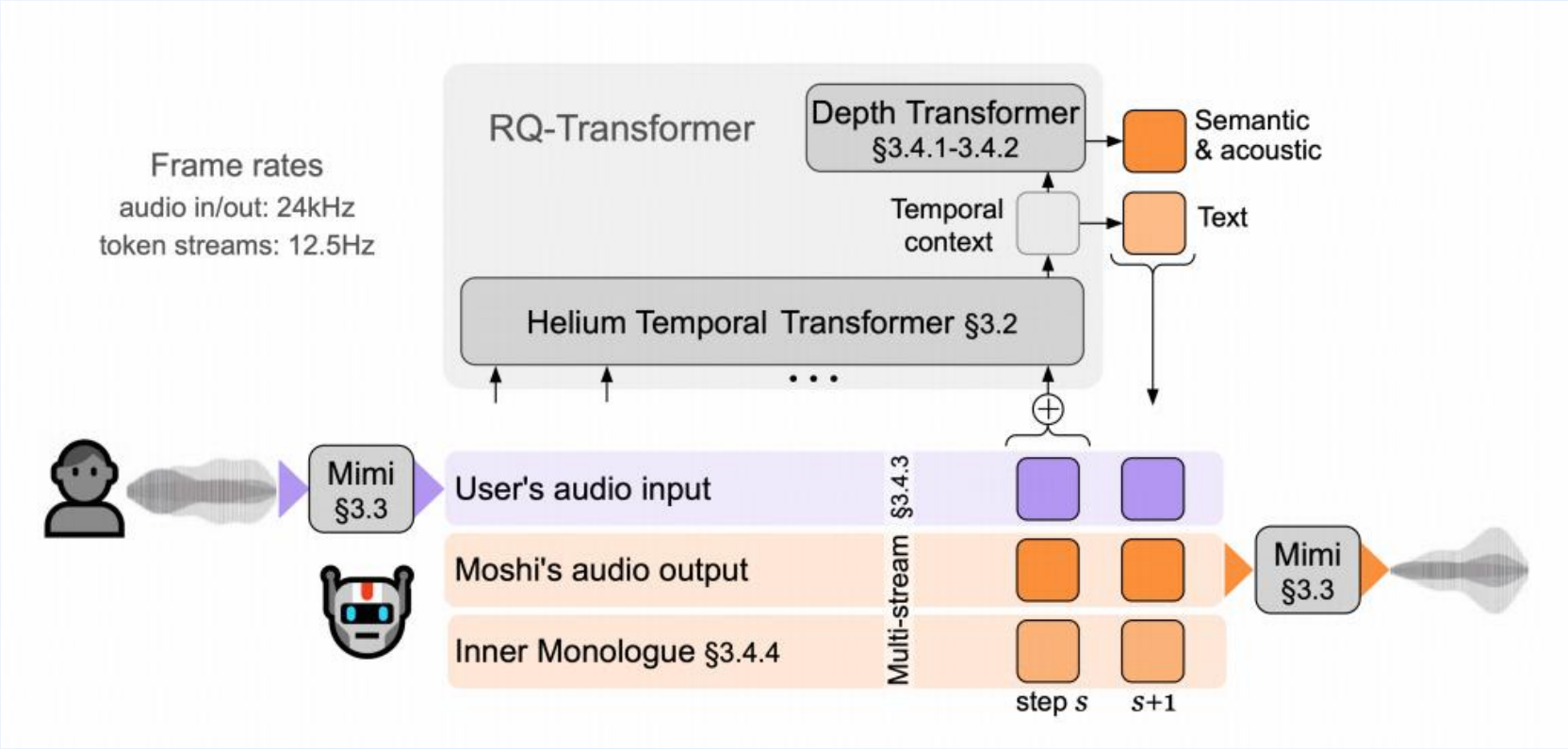




*Du C, Guo Y, Chen X, et al. VQTTS: High-fidelity text-to-speech synthesis with self-supervised VQ acoustic feature[J]. arXiv preprint arXiv:2204.00768, 2022.
Betker J. Better speech synthesis through scaling[J]. arXiv preprint arXiv:2305.07243, 2023.

模型架构

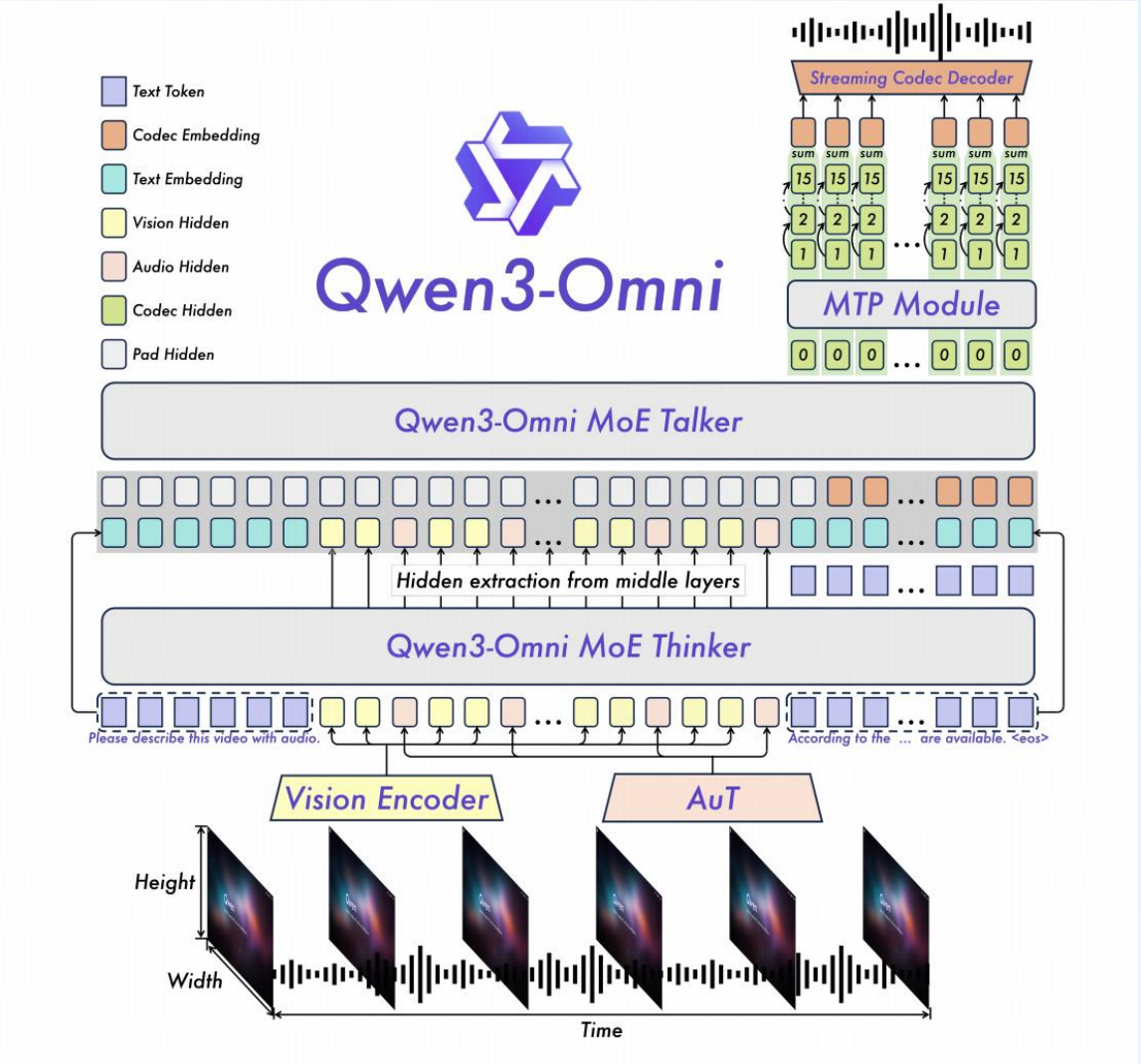
- **Large Audio Language Model**
  - GPT-4o-Audio
  - Moshi*
  - GLM-4-Voice
  - Qwen2.5/3-Omni
  - Kimi-Audio
  - Mimo-Audio
  - Step-Audio1/2



*Défossez A, Mazaré L, Orsini M, et al. Moshi: a speech-text foundation model for real-time dialogue[J]. arXiv preprint arXiv:2410.00037, 2024.

模型架构

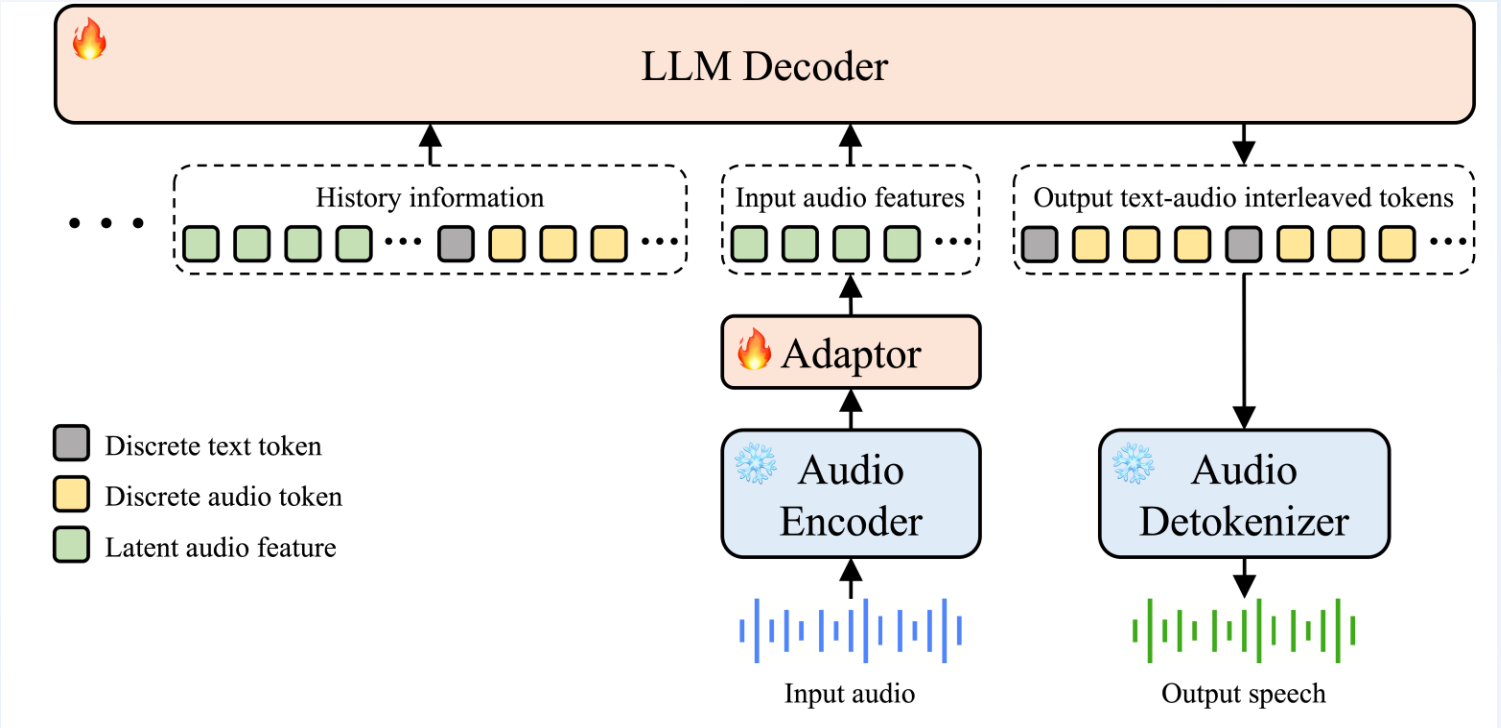
- **Large Audio Language Model**
  - GPT-4o-Audio
  - Moshi
  - GLM-4-Voice
  - Qwen2.5/3-Omni*
  - Kimi-Audio
  - Mimo-Audio
  - Step-Audio1/2



*Xu J, Guo Z, Hu H, et al. Qwen3-omni technical report[J]. arXiv preprint arXiv:2509.17765, 2025.

模型架构

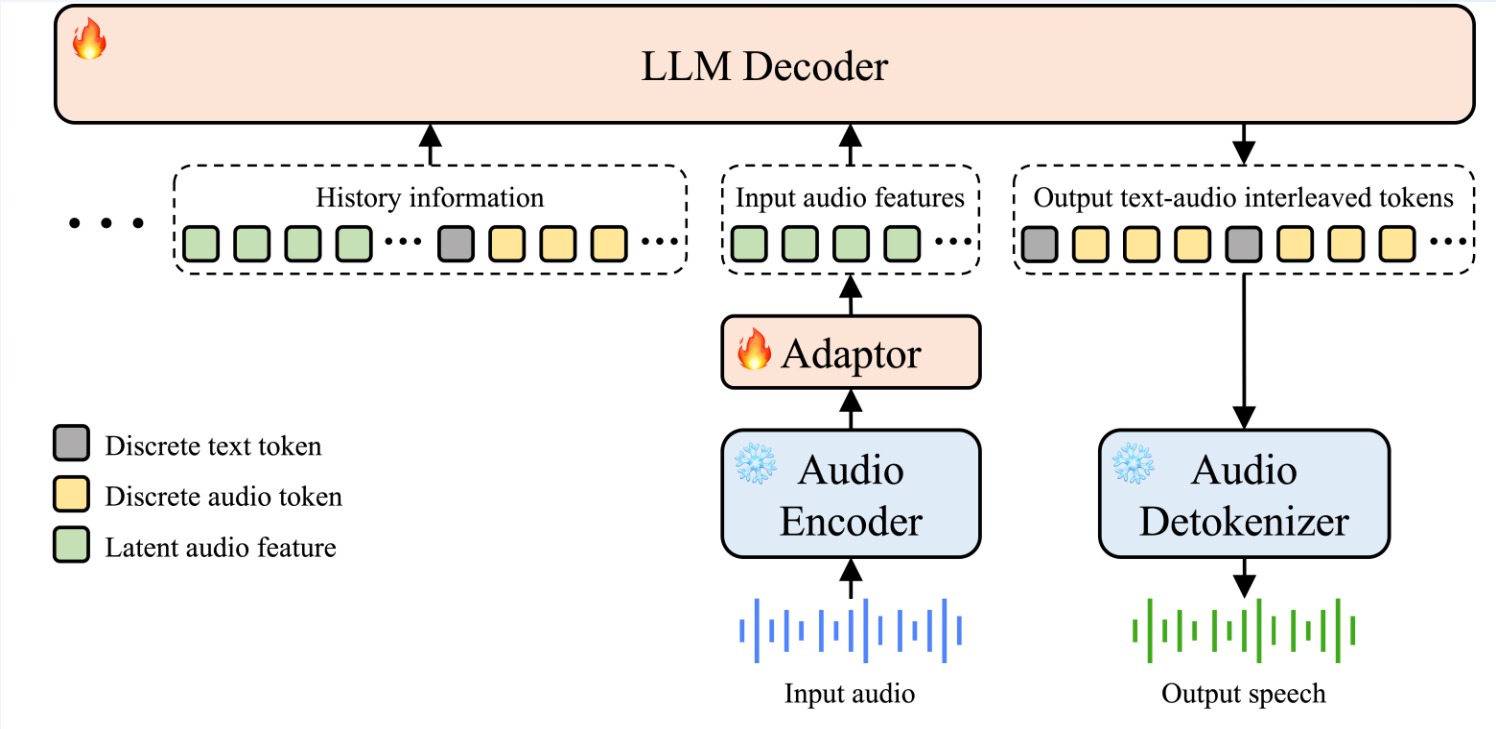
- **Large Audio Language Model**
  - GPT-4o-Audio
  - Moshi
  - GLM-4-Voice
  - Qwen2.5/3-Omni
  - Kimi-Audio
  - Mimo-Audio
  - Step-Audio1/2*



*Wu B, Yan C, Hu C, et al. Step-audio 2 technical report[J]. arXiv preprint arXiv:2507.16632, 2025.

模型架构

- Step-Audio2
- 平衡理解与生成
  - Audio Encoder连续表征输入+离散token输出
- 输入： 12.5Hz
- 输出： 25Hz



02

如何构建端到端语音模型

训推

模型训练——以Step-Audio2为例

- **Continue Pretrain**
  - 基于纯文本 LLM
  - Stage 1：纯 ASR 对齐语音特征和文本特征空间（100B）
  - Stage 2：扩码表，TTS + 语音问答任务建模语音 token（128B）
    - 平衡 128B 文本数据
  - Stage 3：正式预训练阶段，大量较低质量数据（400B）
    - ASR，TTS，语音翻译，语音对话，文本语音混排续写
    - 平衡 400B 文本数据
  - Stage 4：Midtrain，少量较高质量、领域化数据（100B）
    - ASR，TTS，语音翻译，语音对话，文本语音混排续写，副语言信息理解
    - 平衡 100B 文本数据

模型训练——以Step-Audio2为例

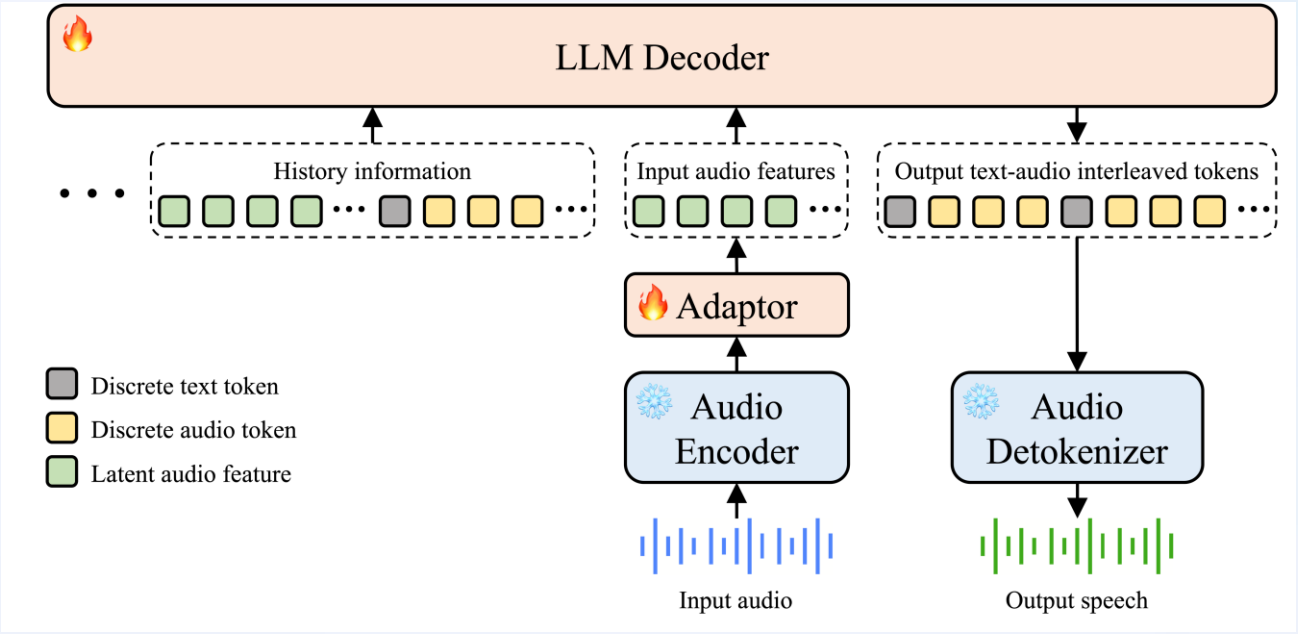
- **Posttrain**
- SFT阶段引入 Human-Assistant 对话结构，精标数据轻量 SFT 方案（2B）
  - **语音识别**：精标ASR数据以及开源高质量数据集；
  - **语音理解**：AudioSet, AudioCaps 等；
  - **语音翻译**：CoVoST 2 等
  - **工具调用**：网页搜索，音频检索；
  - **端到端语音对话**：合成播客级对话数据；

模型训练——以Step-Audio2为例

- **Posttrain-RL**
- 深度推理冷启动：合成语音理解、副语言信息理解深度推理数据
- 优化深度推理，PPO 控制长度与偏好，GRPO 强化结果
  - **PPO:** 长度 binary reward，1 if 0 < length < 200 else 0
    - 实测大概对应 3~5s 的思考时间
  - **GRPO:** group size 8, temperature 1
- 亮点能力：情感深度推理
  - 面向心理咨询、情感安抚等高情商场景

模型推理——以Step-Audio2为例

- 交错Token处理
  - 1:4
- 多轮上下文
  - Audio+Text Context



02

如何构建端到端语音模型

任务

下游任务——以Step-Audio2为例

- 语音理解
  - 语音识别
  - 音频与副语言理解

Category	Test set	Doubao LLM ASR	GPT-4o Transcribe	Kimi-Audio	Qwen-Omni	Step-Audio 2
English	Common Voice	9.20	9.30	7.83	8.33	**5.95**
	FLEURS English	7.22	**2.71**	4.47	5.05	3.03
	LibriSpeech clean	2.92	1.75	1.49	2.93	**1.17**
	LibriSpeech other	5.32	4.23	2.91	5.07	**2.42**
	Average	6.17	4.50	4.18	5.35	**3.14**
Chinese	AISHELL	0.98	3.52	**0.64**	1.17	0.63
	AISHELL-2	3.10	4.26	2.67	2.40	**2.10**
	FLEURS Chinese	2.92	**2.62**	2.91	7.01	2.68
	KeSpeech phase1	6.48	26.80	5.11	6.45	**3.63**
	WenetSpeech meeting	4.90	31.40	5.21	6.61	**4.75**
	WenetSpeech net	**4.46**	15.71	5.93	5.24	4.67
	Average	3.81	14.05	3.75	4.81	**3.08**
Multilingual	FLEURS Arabian	N/A	**11.72**	N/A	25.13	14.22
	Common Voice yue	9.20	11.10	38.90	**7.89**	7.90
	FLEURS Japanese	N/A	**3.27**	N/A	10.49	3.18
In-house	Anhui accent	**8.83**	50.55	22.17	18.73	10.61
	Guangdong accent	4.99	7.83	**3.76**	4.03	3.81
	Guangxi accent	3.37	7.09	4.29	**3.35**	4.11
	Shanxi accent	20.26	55.03	34.71	25.95	**12.44**
	Sichuan dialect	**3.01**	32.85	5.26	5.61	4.35
	Shanghai dialect	47.49	89.58	82.90	58.74	**17.77**
	Average	14.66	40.49	25.52	19.40	**8.85**

下游任务——以Step-Audio2为例

- 语音理解
  - 语音识别
  - 音频与副语言理解
    - VAD
    - 多说话人
    - 情感理解
    - 副语言理解
    - 性别识别
    - 年龄识别
    - 音乐理解

Sys prompt
你是一位经验丰富的音频分析专家，擅长对各种语音音频进行深入细致的分析。你的任务不仅仅是将音频内容准确转写为文字，还要对说话人的声音特征（如性别、年龄、情绪状态）、背景声音、环境信息以及可能涉及的事件进行全面描述。请以专业、客观的视角，详细、准确地完成每一次分析和转写。

这是第一个说话人
说话的时间范围是0.0s到18.65s。
语音的内容是：刚刚刷抖音让我刷到一句特别难过的话。我觉得好难过啊。他说，我可以明目张胆的想你，但是我不能明目张胆的去找你。我可以明目张胆的喜欢你，但是不能明目张胆的拥有你。
这是一位年轻女性，年龄大约在15到25岁之间。她的声音略显稚嫩，带有一些鼻音，语速偏慢，清晰度较高。她以倾诉和独白的方式表达自己的情绪，语气中充满了悲伤和失落感。她的声音低沉而富有情感，语调中透着压抑和无奈，似乎在向某人倾诉内心的痛苦。在背景中，有抒情流行歌曲的音乐声，音量适中，营造出一种录音室或个人录音的环境。她的表达让人感受到一种脆弱感和引人同情的印象。

下游任务——以Step-Audio2为例

- 语音合成
  - 前端处理
  - 可控生成
  - 音频编辑

以多音字处理为例：

- <|THINK_START|>用户的输入包含拼音,分别为zen3, me5,hui4,you3,zhe4,zhong3,shi4,zhe4, pian4, hai3,yu4, shi2, zai4, tai4, shen2, mi4, le5
- 根据拼音、声调和上下文的语义理解,将这句话的拼音转成汉字,结果为:怎么会有这种事，这片海域实在太神秘了!
- 在说这句话时,要注意使用用户标注的拼音和音调作为对应汉字的发音<ITHINK_END|>

下游任务——以Step-Audio2为例

- 语音合成
  - 前端处理
  - 可控生成
  - 音频编辑

我… [Sigh]…我现在脑子里一团乱，[Uhm]真的不知道下一步该怎么走了……

下游任务——以Step-Audio2为例

- 语音合成
  - 前端处理
  - 可控生成
  - 音频编辑

原始音频

编辑为撒娇风格

下游任务——以Step-Audio2为例

- 对话
  - 情感对话
  - 工具调用

Table 5: Comparison between Step-Audio 2 and Qwen3-32B on StepEval-Audio-Toolcall. †Qwen3-32B is evaluated with text inputs. ‡Date and time tools have no parameter.

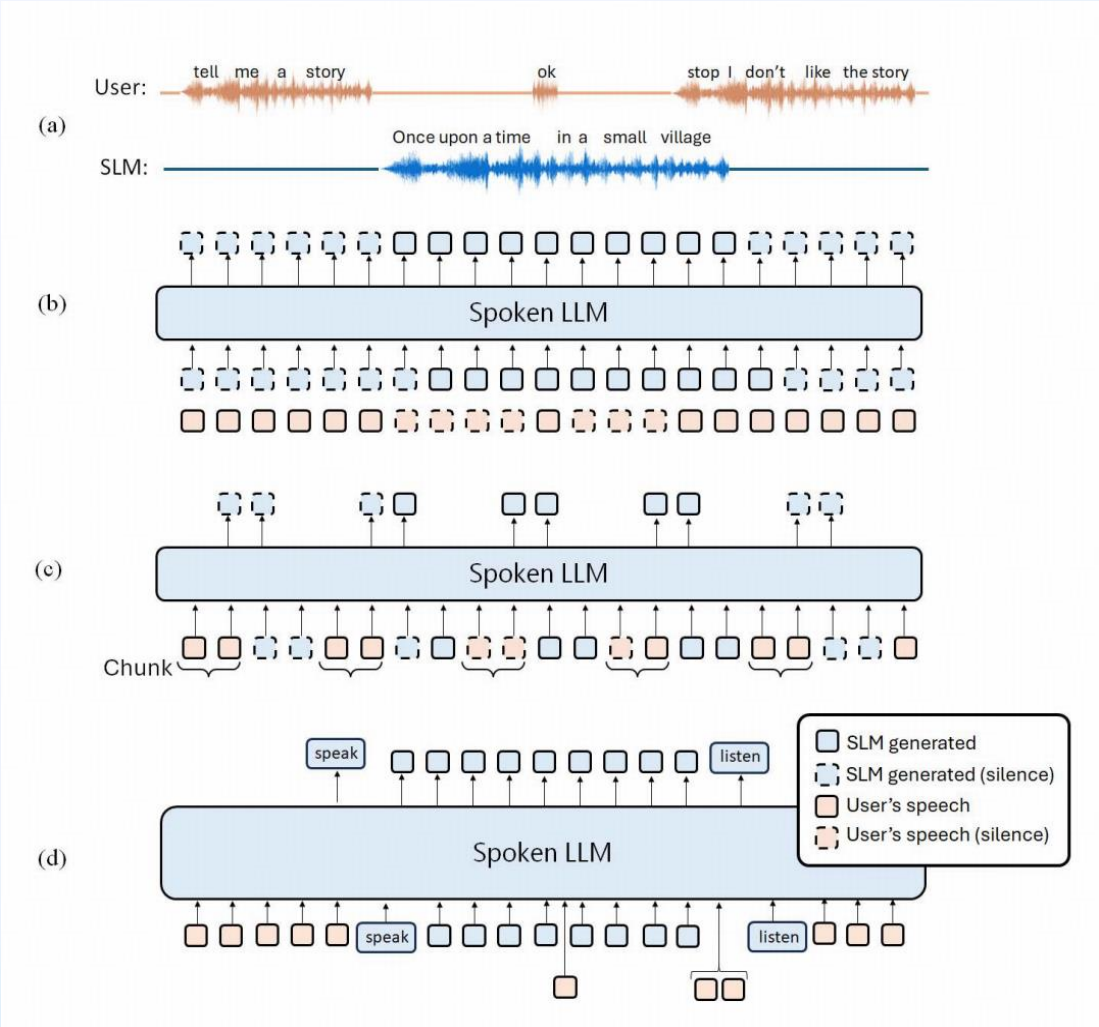
Model	Objective	Metric	Audio search	Date & Time‡	Weather	Web search
Qwen3-32B†	Trigger	Precision / Recall	67.5 / 98.5	98.4 / 100.0	90.1 / 100.0	86.8 / 98.5
	Type	Accuracy	100.0	100.0	98.5	98.5
	Parameter	Accuracy	100.0	N/A	100.0	100.0
Step-Audio 2	Trigger	Precision / Recall	86.8 / 99.5	96.9 / 98.4	92.2 / 100.0	88.4 / 95.5
	Type	Accuracy	100.0	100.0	90.5	98.4
	Parameter	Accuracy	100.0	N/A	100.0	100.0

下游任务——以Step-Audio2为例

- **对话**
  - 情感对话
  - 工具调用
- 基于**工具调用**中的**音频检索**
  - Human: 切换成一个清朝格格的音色
  - Assistant: audio_search(query=清朝格格)
  - Input: <清朝格格 audio prompt>
  - Assistant: <参考 Input 生成的语音回复>
- 音色库
  - 10w 量级的 audio – 描述对
  - 不含名人

下游任务

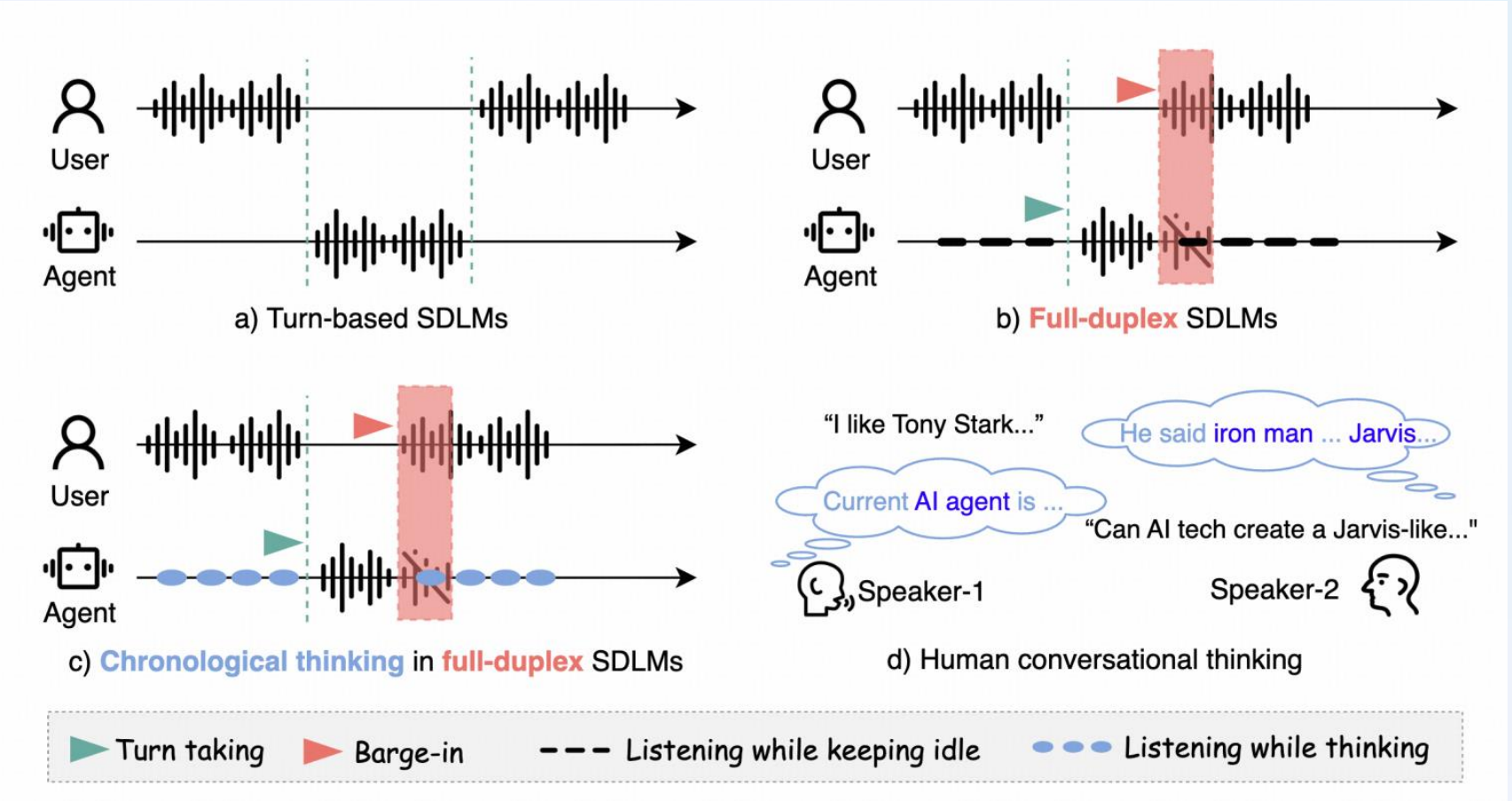
- 全双工对话
  - VAD-Based
  - Multichannel
  - Interleaved
  - Chunk-wise



*Arora S, Chang K W, Chien C M, et al. On the landscape of spoken language models: A comprehensive survey[J]. arXiv preprint arXiv:2504.08528, 2025.

下游任务

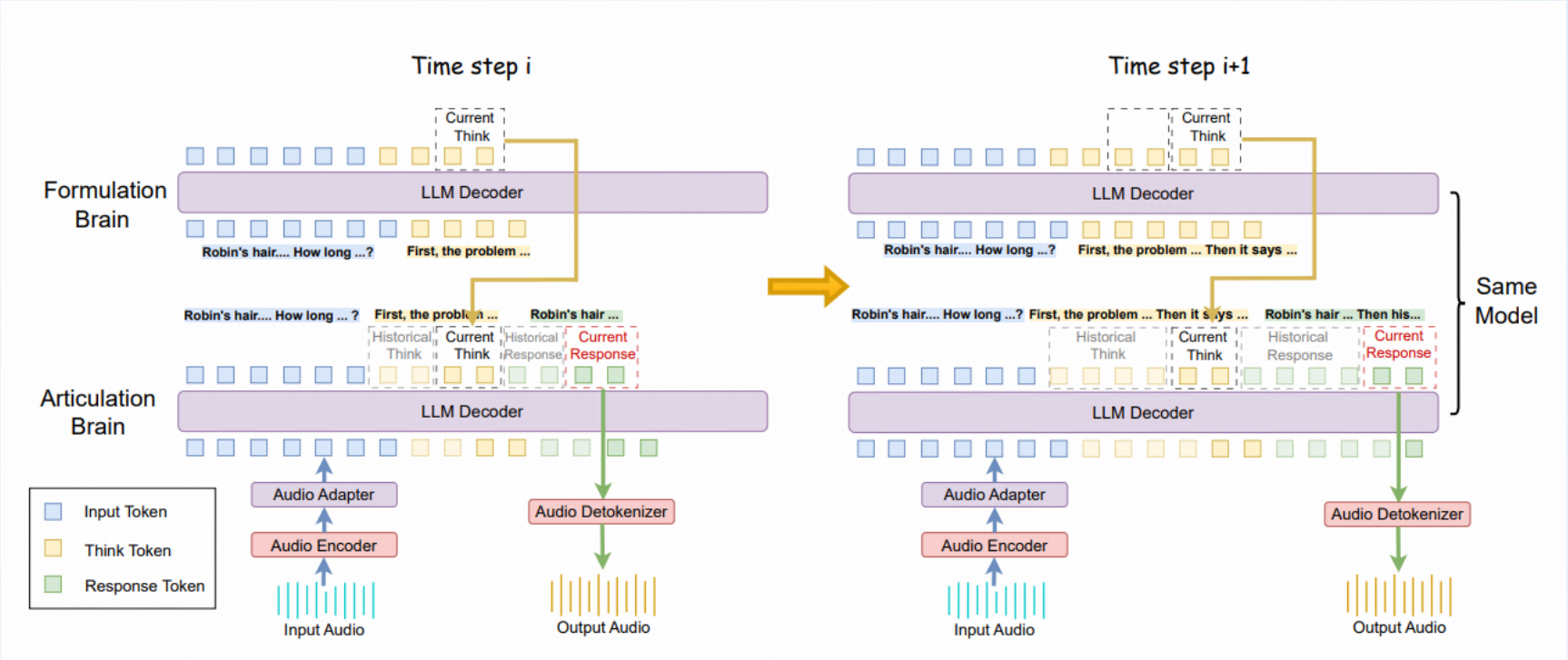
- 全双工推理
  - STITCH
  - SHANKS
  - CT*
  - Step-MPS



*Wu D, Zhang H, Chen C, et al. Chronological Thinking in Full-Duplex Spoken Dialogue Language Models[J]. arXiv preprint arXiv:2510.05150, 2025.

下游任务

- 全双工推理
  - STITCH
  - SHANKS
  - CT
  - Step-MPS*



*Wu D, Zhang H, Chen J, et al. Mind-Paced Speaking: A Dual-Brain Approach to Real-Time Reasoning in Spoken Language Models[J]. arXiv preprint arXiv:2510.09592, 2025.

03

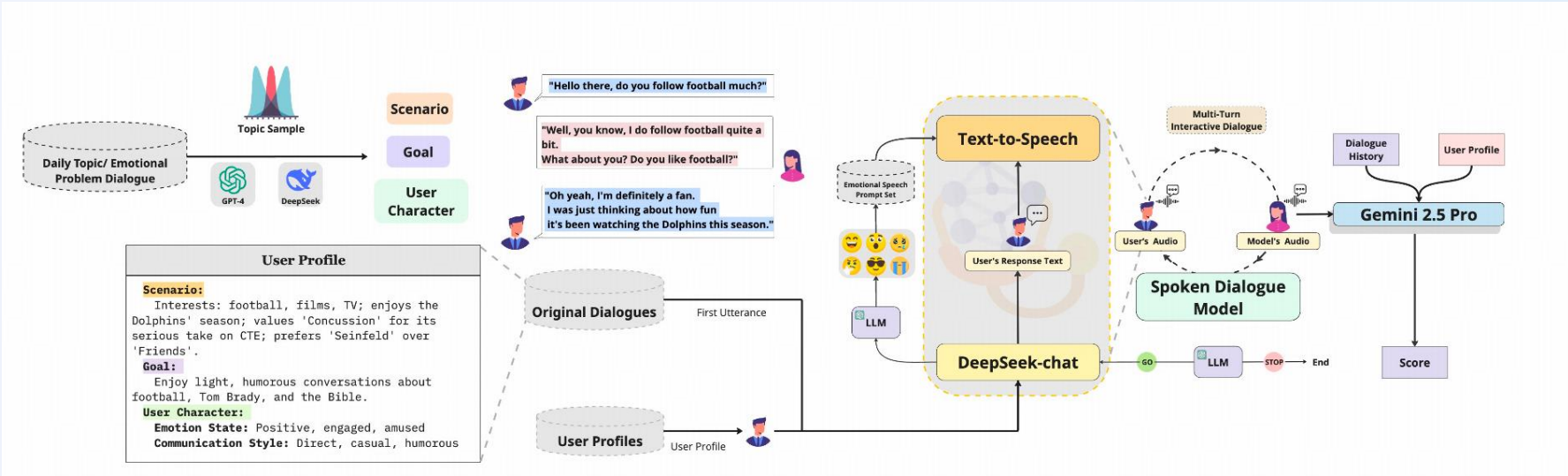
模型评估：什么是好模型

模型评估

- 这很重要!
- **单点能力**
  - 语音识别
  - 语音翻译
  - 情感和副语言理解
  - 音频理解
  - 语音合成
  - 语音克隆
  - 音频编辑
- **对话能力**
  - 知识性
  - 创作能力
  - 工具调用
  - 推理与规划
  - 指令遵循
  - 全双工
  - 多轮一致性
  - 情感与共情

模型评估

- 这很重要!
- 语音对话BMK
  - VoiceBench
  - AIR-Bench
  - ADU-Bench
  - SD-Eval
  - C3Benchmark
  - URO-Bench
  - VoxDialogue
  - MULTI-BENCH*



InfoQ 极客传媒

北京

QCon 全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

北京

AiCon 全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

上海

QCon 全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月 — 5月 — 6月 — 8月 — 10月 — 12月

上海

AiCon 全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

深圳

AiCon 全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

北京

AiCon 全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOps
- 具身智能

极客邦科技 2025 年会议规划

促进软件开发及相关领域知识与创新的传播

参会咨询

查看会议

QCon
全球软件开发大会

InfoQ 极客传媒

THANKS

大模型正在重新定义软件

Large Language Model Is Redefining The Software