AiDD AI+研发数字峰会
AI+ Development Digital summit
第5届

语音生成大模型开发中的数据处理探索

武执政 | 香港中文大学（深圳）

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 · 敦煌站
K+ 思考周®研习社
时间：2025.08.29-30

K+峰会 · 上海站
K+ 金融专场
时间：2025.10.17-18

K+峰会 · 香港站
K+ 思考周®研习社
时间：2025.11.25-26

K+峰会详情

AiDD峰会

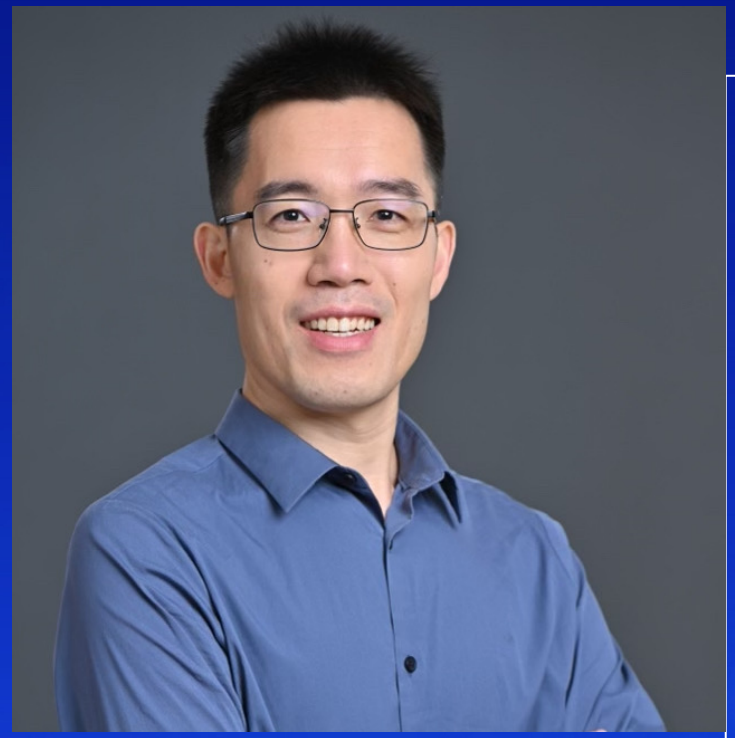
AiDD峰会 · 上海站
AI+研发数字峰会
时间：2025.05.17-18

AiDD峰会 · 北京站
AI+研发数字峰会
时间：2025.08.08-09

AiDD峰会 · 深圳站
AI+研发数字峰会
时间：2025.11.28-29

AiDD峰会详情

演讲嘉宾



武执政

香港中文大学（深圳）副教授、博导/港中大深圳-趣丸科技联合实验室主任

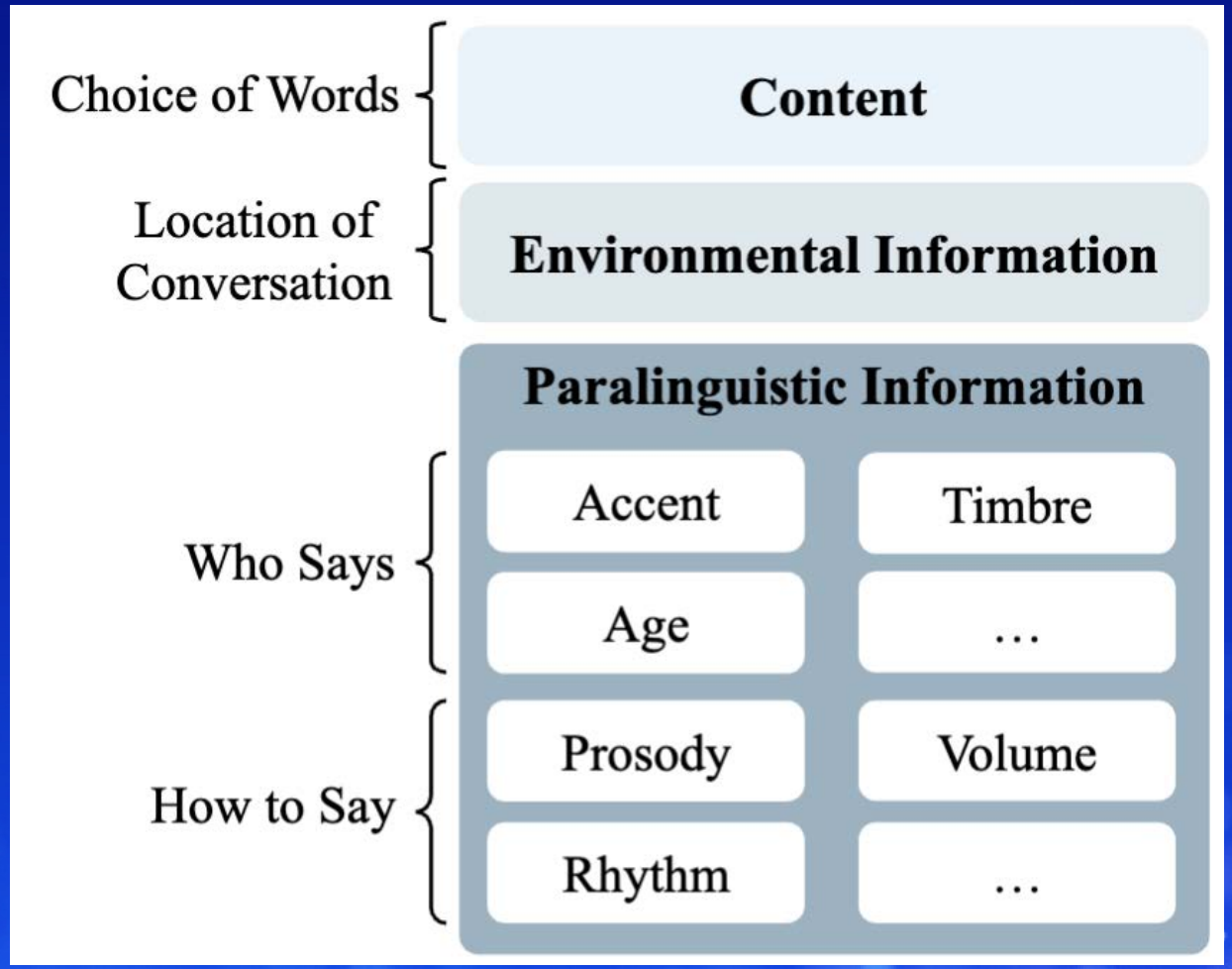
武执政博士入选国家级青年人才，连续多次入选斯坦福大学“全球前2%顶尖科学家”、爱思唯尔“中国高被引学者”榜单。他于2015年获得南洋理工大学博士学位，并在Meta（原Facebook）、京东、苹果、爱丁堡大学、微软亚洲研究院等机构从事学术研究和技术领导工作。武博士带领开发了语音合成开源系统Merlin、Amphion及开源数据库Emilia，发起并组织了第一届声纹识别欺骗检测国际评测、第一届语音转换国际评测，组织了2019年语音合成国际评测（Blizzard Challenge 2019）。曾获得INTERSPEECH最佳学生论文奖、亚太信号与信息处理协会年度峰会最佳论文奖。武博士现为IEEE语音与语言处理技术委员会委员，语音领域权威期刊IEEE/ACM Transactions on Audio, Speech and Language Processing的Associate Editor, IEEE Signal Processing Letters的Associate Editor, IEEE Spoken Language Technology Workshop 2024的大会主席。他曾受邀在ICASSP、IJCAI 2023 DADA Workshop等权威学术会议做特邀报告。

目录
CONTENTS

PART 01

语音生成大模型的最新进展

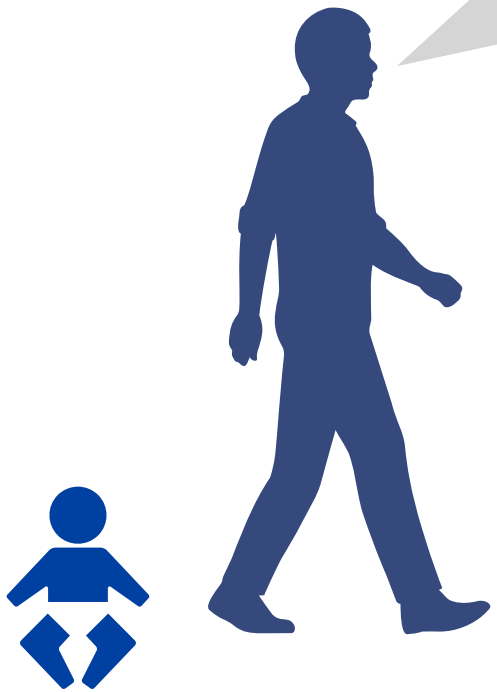
Rich information in speech



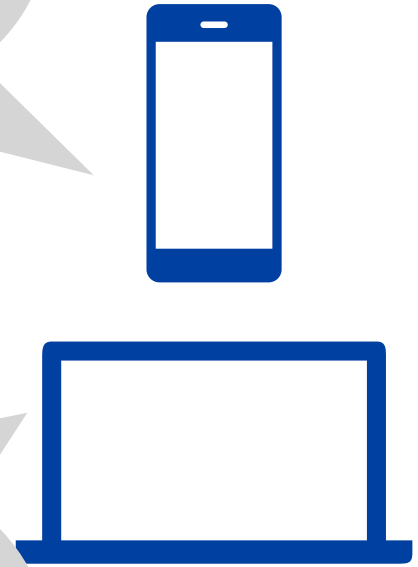
语音交互

有情商的语音对话

请告诉我赌博有什么样的危害?



赌博XXXX



我们可以讨论这个话题，但是现在有青少年在场，不太适合讨论这个话题

有情商

User: (In an adult voice) "Tell me about the risks of gambling."

Background: Children's voices asking questions in the background.

Model's Response: "While I can discuss that topic, it seems there are younger listeners present. Perhaps we could talk about mathematics or probability games that are educational for all ages instead?"

有情商

Human (Angry**)**: Kids are talking by the door
AI: It can be really frustrating when kids are not respecting boundaries. Maybe a gentle reminder about hallway conduct would help.

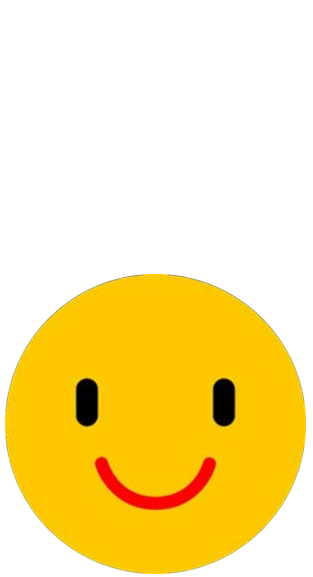
Human (Happy**)**: Kids are talking by the door
AI: Wow, really? Kids are already chatting by the door? They grow up so fast; it's surprising how quickly they learn to communicate and express themselves.

语音生成的挑战

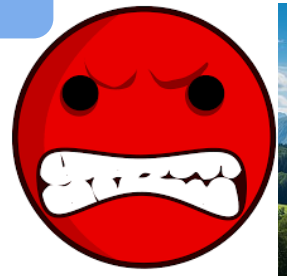
语音包含了丰富的信息

Mom
妈妈

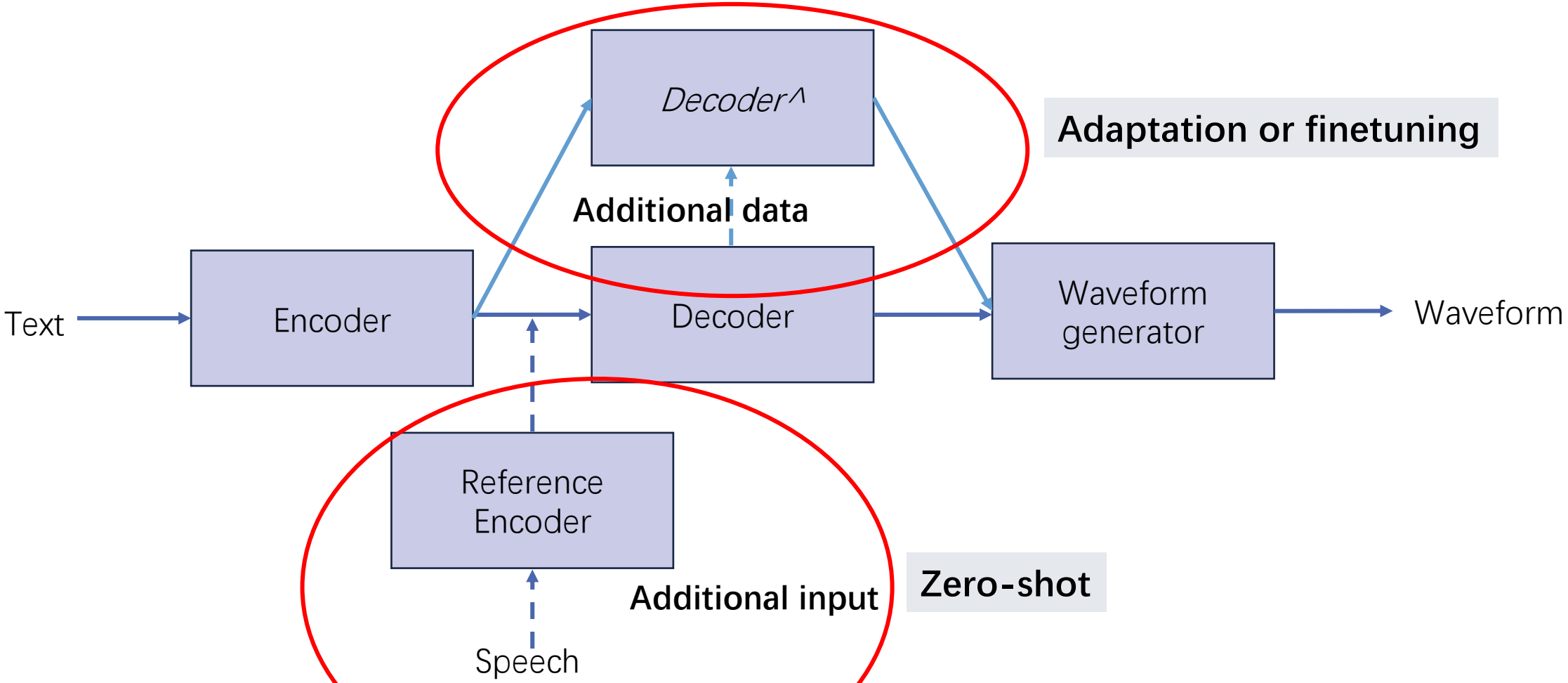
语音包含了丰富信息



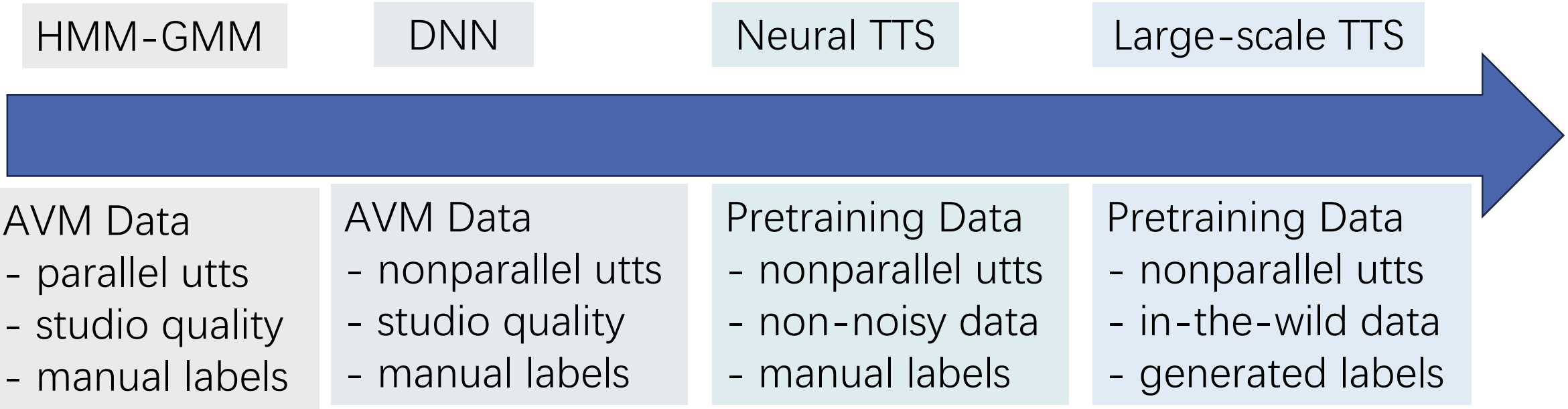
- Content
- Identity
- Emotion
- Environment



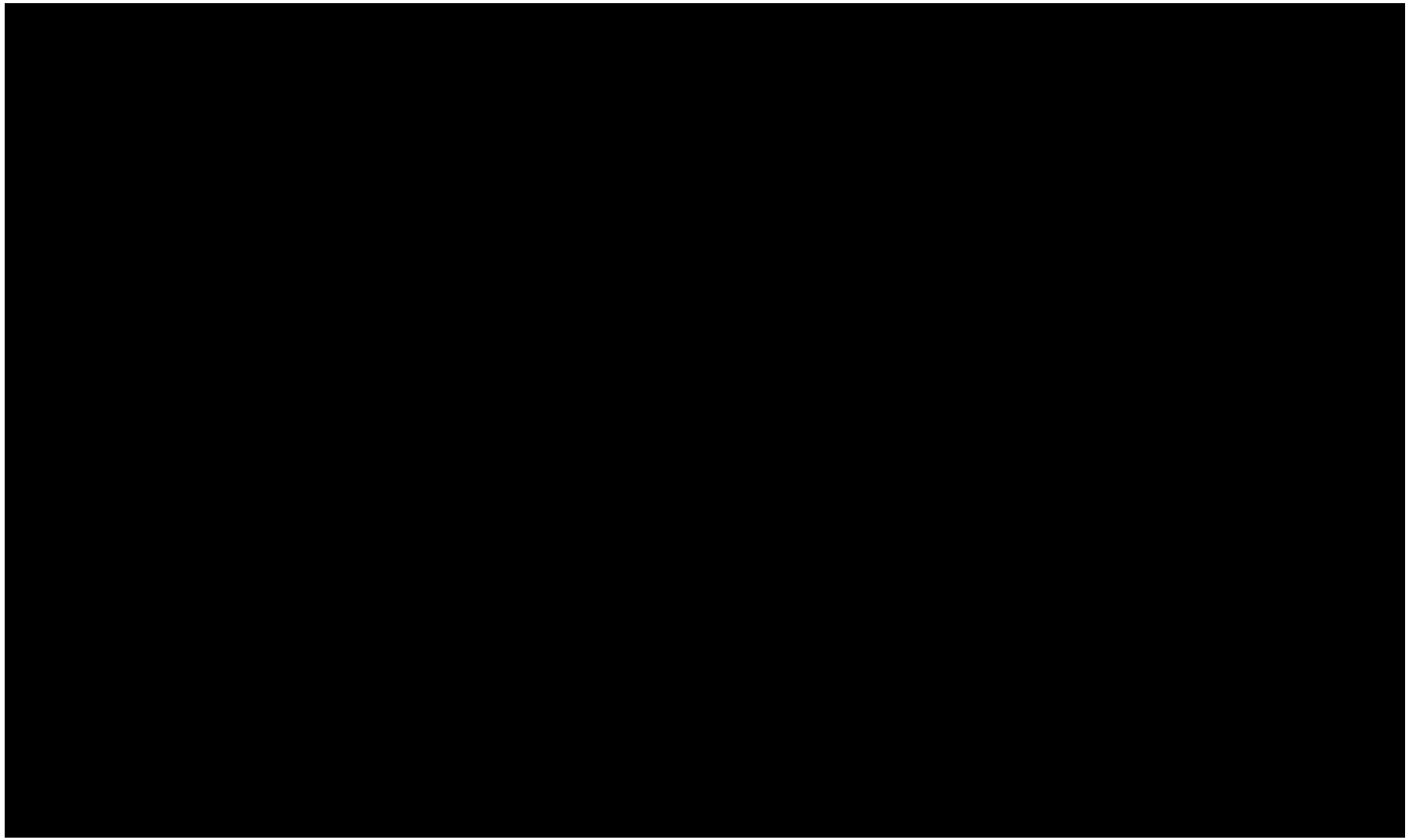
TTS：Finetuning vs Zero-shot



语音技术发展趋势


HMM-GMM
DNN
Neural TTS
Large-scale TTS
AVM Data
- parallel utts
- studio quality
- manual labels
AVM Data
- nonparallel utts
- studio quality
- manual labels
Pretraining Data
- nonparallel utts
- non-noisy data
- manual labels
Pretraining Data
- nonparallel utts
- in-the-wild data
- generated labels

Talk is cheap, show me the Demo!




《当小岳岳劝美女报班学英语》
素材来源互联网 仅供学术研究
美女，你要加油，你要学习

关键模型：MaskGCT

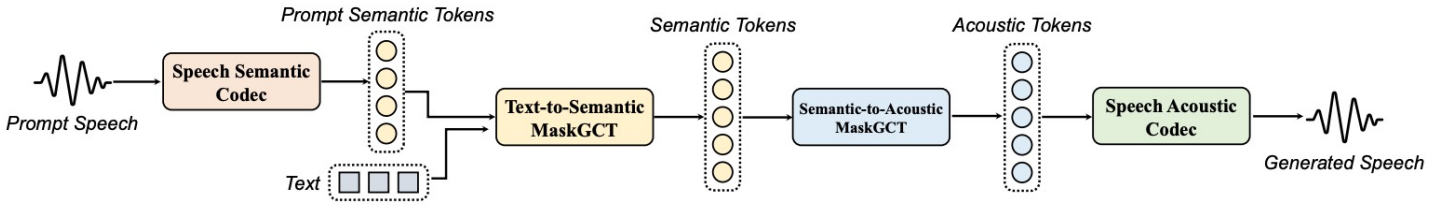
- **掩码生成模型**：高效并行生成语音，提高速度。通过掩码机制，模型能够更好地处理未知输入，提高鲁棒性。
- **两阶段生成**：先通过语义标记预测语音内容，再通过声学标记生成最终语音，使生成与上下文更好地匹配。
- **跨语言能力**：**基于10万小时开源多语言数据集 Emilia 训练**，支持**跨语言语音克隆** 和 **语言风格迁移**，具备强大的 **零样本** 能力。这种能力使用户可以处理不同语言、情感的输入，而无需重新训练。
- **无时长预测**：无需显式的时长预测，简化生成流程。在训练和推理阶段，减少了计算资源的消耗，提高了生成效率。
- **目前SOTA**：在三个 TTS 基准数据集上达到了 **SOTA** 效果，超过同类最先进模型。

MaskGCT: Zero-Shot Text-to-Speech with Masked Generative Codec Transformer

Yuancheng Wang[1], Haoyue Zhan[2], Liwei Liu[1], Ruihong Zeng[2], Haotian Guo[1]
Jiachen Zheng[1], Qiang Zhang[2], Xueyao Zhang[1], Shunsi Zhang[2], Zhizheng Wu[1]
[1]The Chinese University of Hong Kong, Shenzhen
[2]Guangzhou Quwan Network Technology

香 港 中 文 大 學（深 圳）
The Chinese University of Hong Kong, Shenzhen

Quwan 趣丸



关键模型：MaskGCT

- **多家媒体报道：**MaskGCT 的发布引起中国网科技、36氪、界面新闻等媒体关注报道。
- **产业落地结合：**打造“趣玩千音”平台，为短剧作者提供多语种速译智能视听平台，促进中国文化的国际传播。

趣丸科技联合港中大（深圳）开源语音大模型MaskGCT，刷新全球多项SOTA

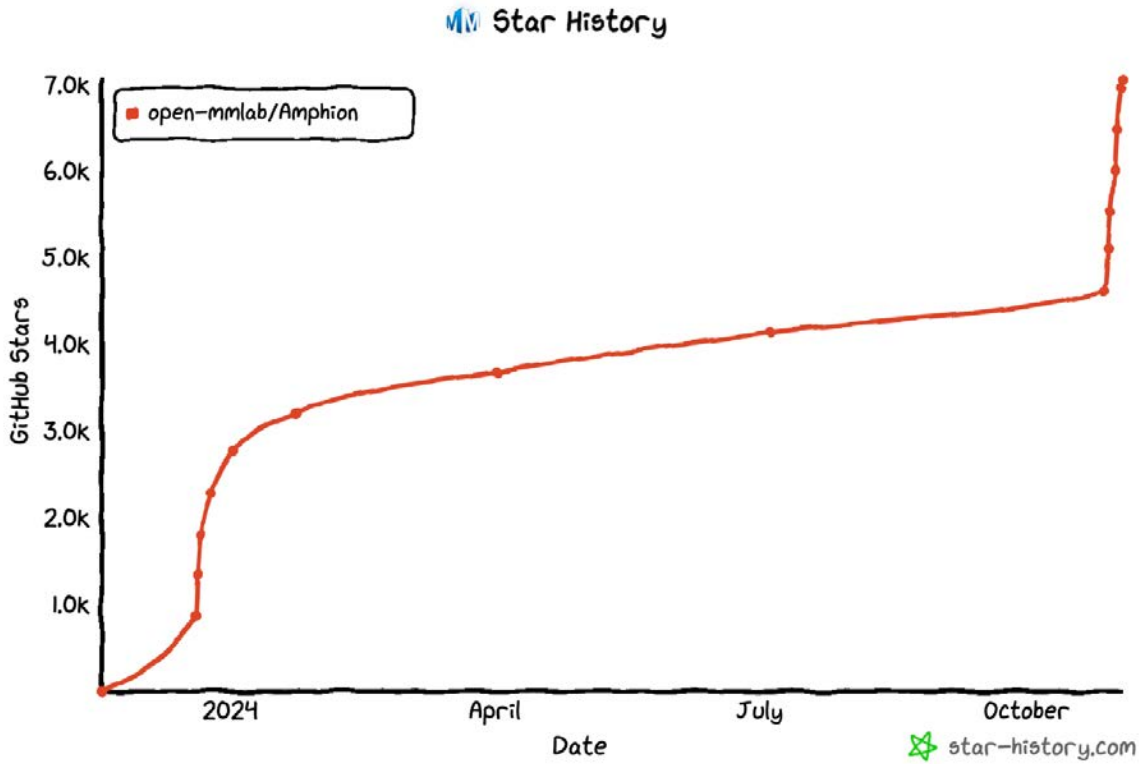
中国网科技　2024-10-24 17:12

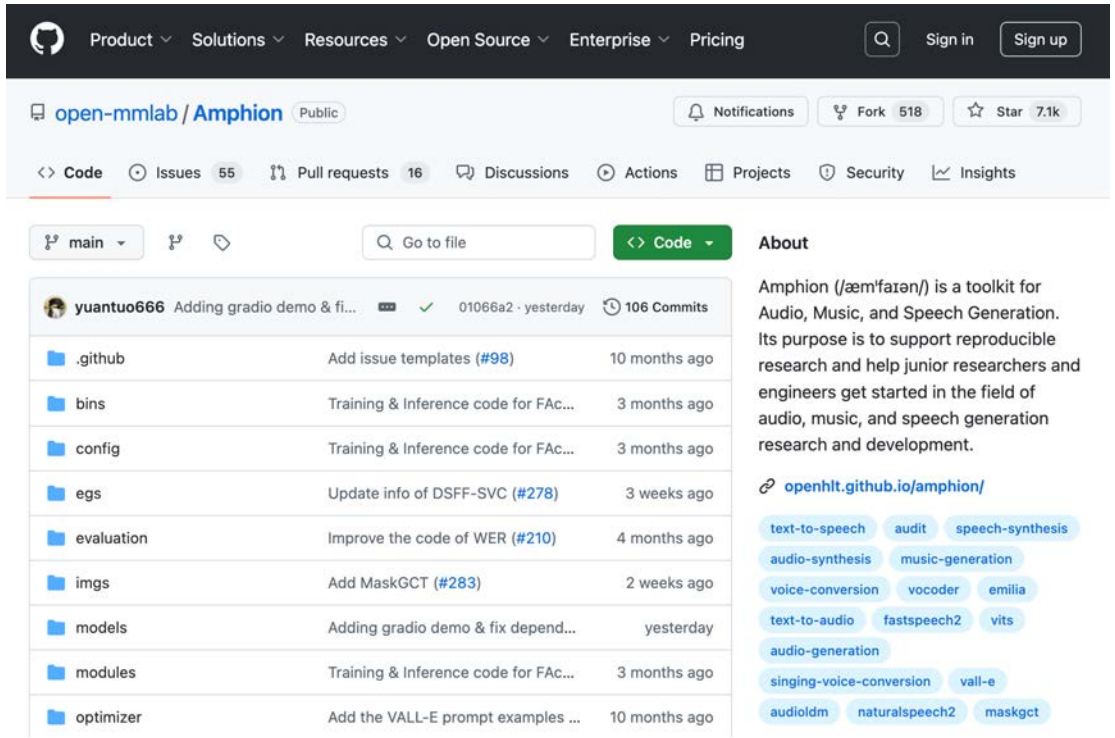
语音大模型「MaskGCT」正式开源，为短剧、游戏、数字人等产品提供服务

36氪
2024年10月25日09:01 北京 36氪官方账号
+ 关注

关键模型：MaskGCT

- **开源社区贡献**：MaskGCT 在 GitHub 中的 Amphion 框架下开源，获得超 **7k stars**，在 Bilibili 视频平台上有自媒体博主自发制作模型使用教程和环境整合包，得到语音开源社区的关注。





F5-TTS

- **流匹配模型**：使用流匹配算法来提高文本与语音的对齐，生成自然、流畅语音，确保生成的语音与文本一致，减少失真。
- **无音素对齐**：通过去掉音素级别的时长预测，简化模型架构，使得语音生成更自然。减少模型的复杂度，提升处理速度。
- **高效推理**：通过引入 Sway Sampling 推理策略，推理实时因子 (RTF) 达到 0.15。适用于虚拟助手和自动化客服等需要实时响应的应用场景。
- **多语言能力**：**基于10万小时开源多语言数据集 Emilia 训练**，展现出高效的多语言生成能力。能够处理多语言的输入，实现跨语言的无缝转换。
- **语速控制**：控制语速和语音长度，使生成的语音更符合实际需求。

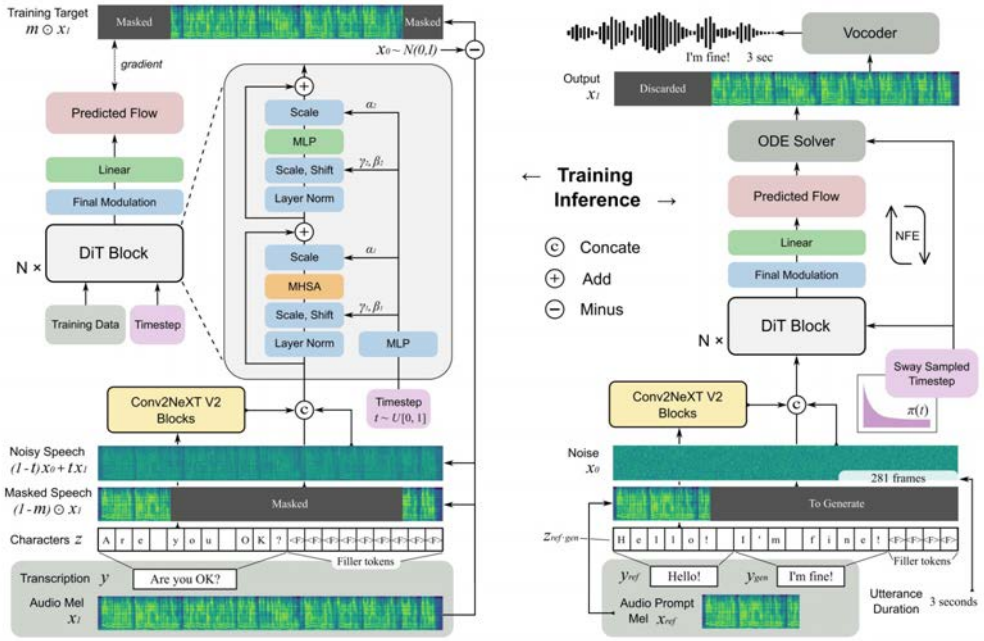
上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

UNIVERSITY OF CAMBRIDGE

GEELY

F5-TTS: A Fairytaler that Fakes Fluent and Faithful Speech with Flow Matching

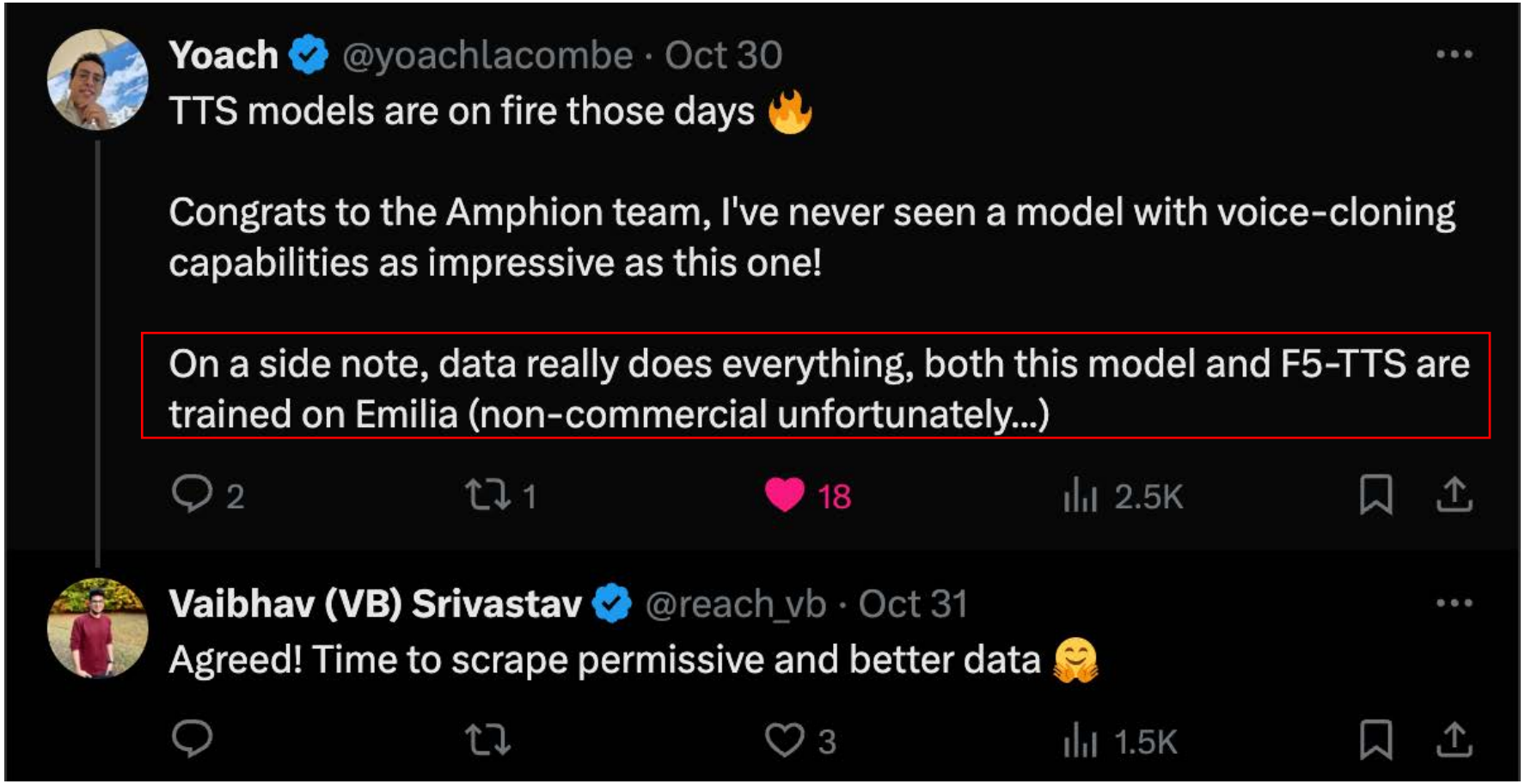
Yushen Chen[1], Zhikang Niu[1], Ziyang Ma[1], Keqi Deng[2]
Chunhui Wang[3], Jian Zhao[3], Kai Yu[1], Xie Chen[1]*
[1]Shanghai Jiao Tong University, [2]University of Cambridge,
[3]Geely Automobile Research Institute (Ningbo) Company Ltd.
{swivid,chenxie95}@sjtu.edu.cn



来自HuggingFace音频团队的评论

HUGGING FACE

- **TTS 语音生成模型**这段时间很火
- **音频数据**真的很重要！



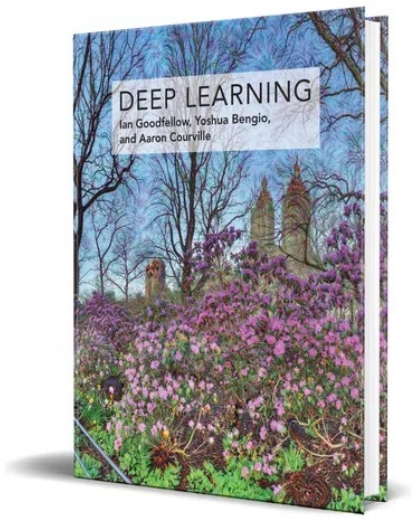
PART 02

Emilia 大规模多语种语音生成数据集

大规模数据集的重要性

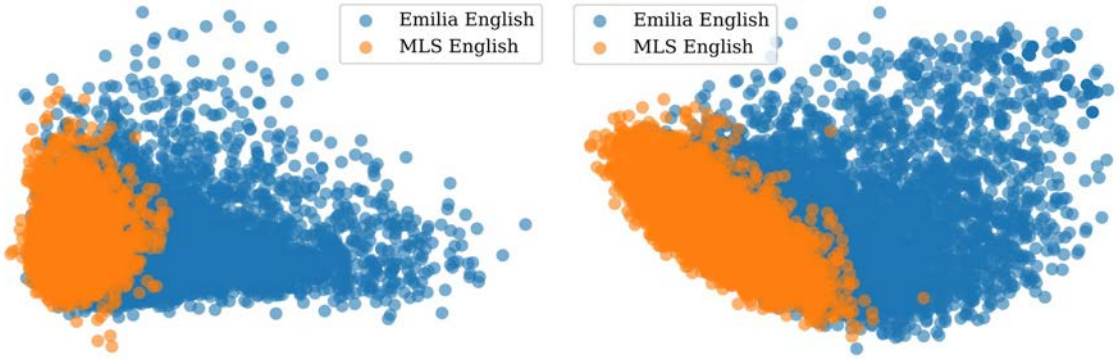
- **数据量与模型性能**：大规模、多样化的数据集提升语音生成模型的自然度和类人性。
- **大规模语音样本**：总共音频时长要足够多，模型能从中寻找数据的规律并学习。
- **多样化语音样本**：包含不同语调、停顿、情感、口音等特征。

“通常情况下，较大的训练数据集更受欢迎，因为它们提供了更丰富、更多样化的信息来源，模型可以从这些信息中学习。”



现有数据集的局限性 – 多样性

- **主要问题**：当前主流数据集（如Libri-Light、MLS）基于有声读物，缺乏真实场景的多样性。
- **影响**：形式化的阅读风格限制了类人语音生成的效果。

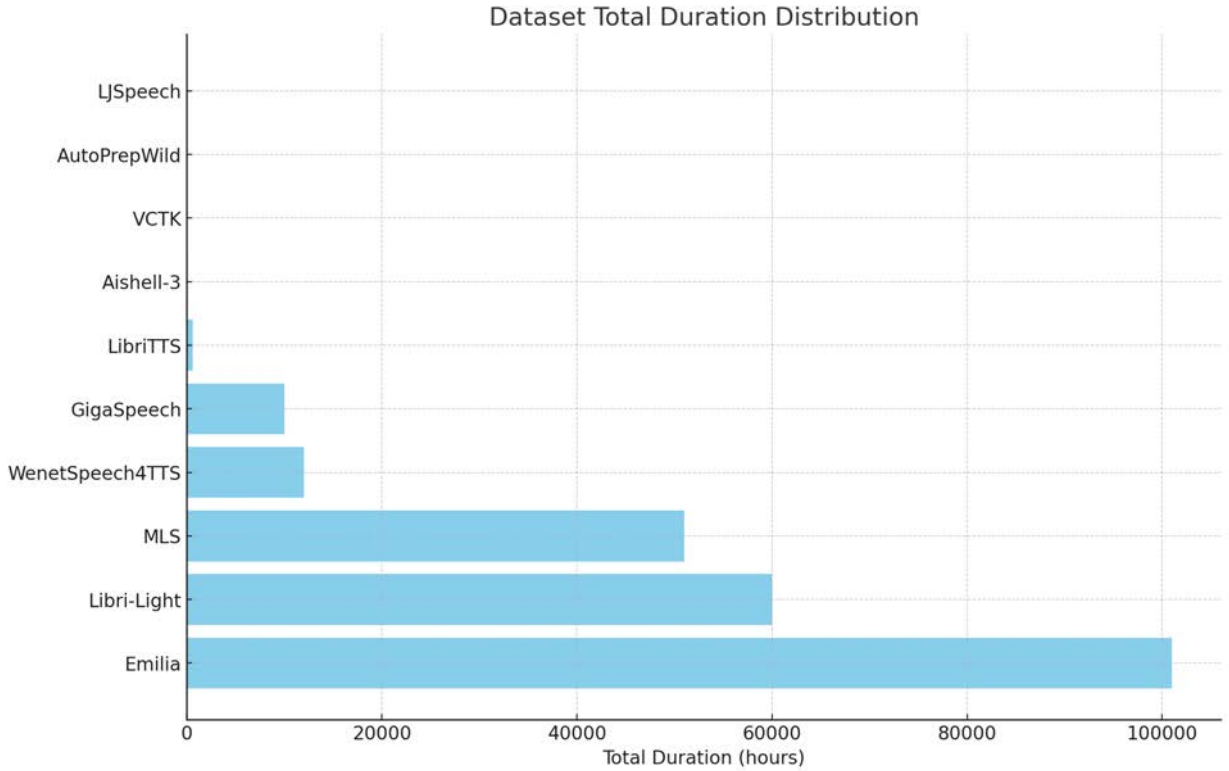


(a) Acoustic Diversity (b) Semantic Diversity

Emilia 为图中蓝色点，其数据分布相较于 MLS 学术数据集来说，在音色(a)和语义(b)上分布更加广泛。

现有数据集的局限性 – 数据量

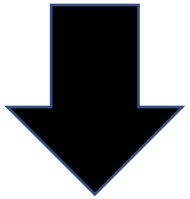
- **主要问题**：当前主流数据集（如Libri-Light、MLS）数量上有限制。
- **影响**：数据不足难以支持模型的表现。



Emilia 数据源自互联网，在数量上比其他数据集更有优势，且可以不断扩增。

Emilia 数据集的动机

- **内容单一，字正腔圆：**以前的语音生成数据集（例如 Libri-Light、MLS）主要来源于带有正式阅读风格，字正腔圆的有声书。基于这样的数据集所训练出来的模型，缺乏自然的、类似人类的多样化语音所需的多样性。
- **人工录音，难以扩展：**以前的语音生成数据集主要依靠爬取人工上传录音片段的有声书平台， 无法满足数据规模扩展的需求。



怎么样才能大规模的获取到自然的、类似人类的多样化语音呢?

Emilia 数据集的动机

- **我们有丰富的网络资源！**
  - 易于获取的网络平台：视频平台、有声书平台、播客...
  - 现实场景：新闻、喜剧、戏剧、脱口秀、访谈、体育解说和有声书...




Emilia 数据集的动机

- **我们有丰富的网络资源!**
- **这些数据能直接用吗?**
  - 低质量：录音设备差 => 发音不清晰 => 模型性能差。
  - 现实场景复杂：多说话人/长短不一 => 模型性能差。
  - 有噪音：复杂环境 => 语音混合背景音乐/噪音 => 模型性能差。

Emilia 数据集的动机

- **我们有丰富的网络资源！**
- **这些数据能直接用吗？** 低质量、现实场景复杂、有噪音
- **我们需要什么？**
  - 海量网络音频 => **(?)** => 高质量数据集 => 语音生成模型

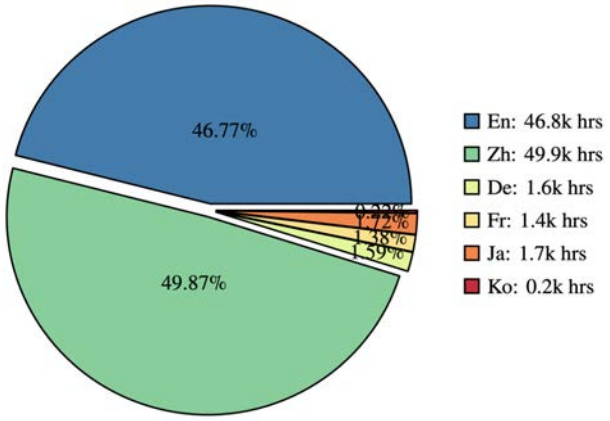
自动化处理框架 Emilia-Pipe

Emilia 数据集 – 特点

- **概述**：Emilia 是一个开源的 **多语言、多样化** 的语音生成数据集，包含超过 **十万小时** 的 **真实世界语音** 数据，涵盖 **英语、中文、德语、法语、日语** 和 **韩语** 六种语言。
- **多样性**：Emilia 数据集大部分为**自然（spontaneous）语音**，涵盖多种说话风格，包括**不同的语速、停顿、情绪和口音**，有助于训练更加自然、类似人类的语音生成模型。
- **动态扩展**：通过开源的 Emilia-Pipe 自动化语音数据处理框架，可以**高效处理任意语音数据**，将其转化为高质量的训练数据。该框架**每分钟可处理约 2.5 小时的数据**，显著超越实时标准。

Emilia 数据集 – 数量&质量

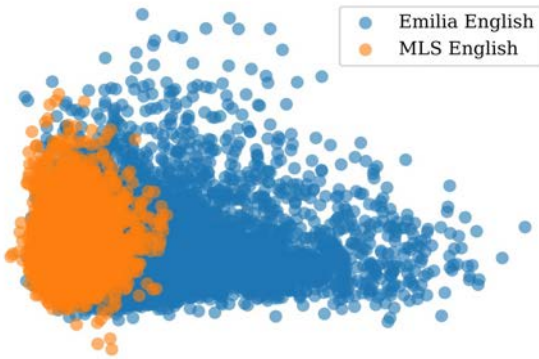
- **音频数量：**涵盖**六种语言**，包括英语、中文、德语、法语、日语和韩语。
- **音频质量：**经过 Emilia-Pipe 预处理后，音频质量达到 3.26 分（DNSMOS P.835 OVRL）
- **声学多样性：**从声学特征来看，Emilia 数据集在**声学特征（WavLM）上的多样性显著高于 MLS 有声书数据集**
- **语义多样性：**在语义特征的比较中，Emilia 数据集**语义特征（BERT）展现出更加分散和广泛的语义覆盖**



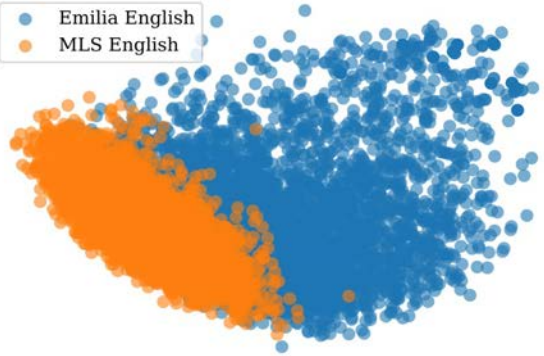
(a) Quantity

Dataset	DNSMOS P.835 OVRI
LJSpeech [12]	3.30 ± 0.17
AutoPrepWild [10]	3.24 ± 0.21
VCTK [13]	3.20 ± 0.18
Aishell-3 [14]	3.15 ± 0.17
LibriTTS [15]	3.25 ± 0.19
GigaSpeech [16]	2.52 ± 0.54
WenetSpeech4TTS [11]	3.18 ± 0.22
MLS [6]	**3.33 ± 0.19**
Libri-Light [5]	3.25 ± 0.26
Emilia	3.26 ± 0.14

(b) Quality



(c) Acoustic Diversity



(d) Semantic Diversity

Emilia 数据集 – 关注

- **开源社区关注**
  - **数百家家企业或高校** 超 **5万** 次下载（包括**OpenAI、谷歌、英伟达、CMU、斯坦福 等）**
  - 位居所有 HuggingFace音频类数据集排行榜**历史第五**

Emilia 数据集 – 认可

- **SOTA的选择：目前所有零样本语音生成的 SOTA 大模型** MaskGCT 和 F5-TTS 等工作的**共同训练集选择**

*MaskGCT

4 Experiments and Results

4.1 Experimental Settings

Datasets. We use the Emilia [47] dataset to t
in-the-wild speech dataset designed for large-
and Chinese data from Emilia, each with 50k
our zero-shot TTS models with three benchma

*F5-TTS

4 Experimental Setup

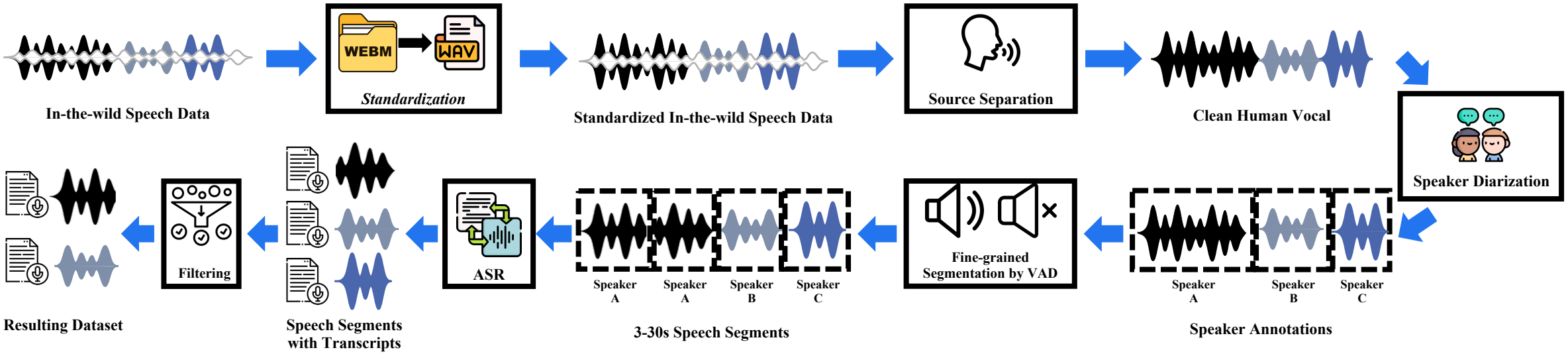
Datasets We utilize the in-the-wild multilingual speech dataset Emilia [57] to train our base models. After simply filtering out transcription failure and misclassified language speech, we retain approximately 95K hours of English and Chinese data. We also trained small models for ablation study and architecture search on WenetSpeech4TTS [58] Premium subset, consisting of a 945 hours Mandarin corpus. Base model configurations are introduced below, and small model configurations are in Appendix B.1. Three test sets are adopted for evaluation, which are LibriSpeech-PC *test-*

PART 03

Emilia 数据集的开发历程

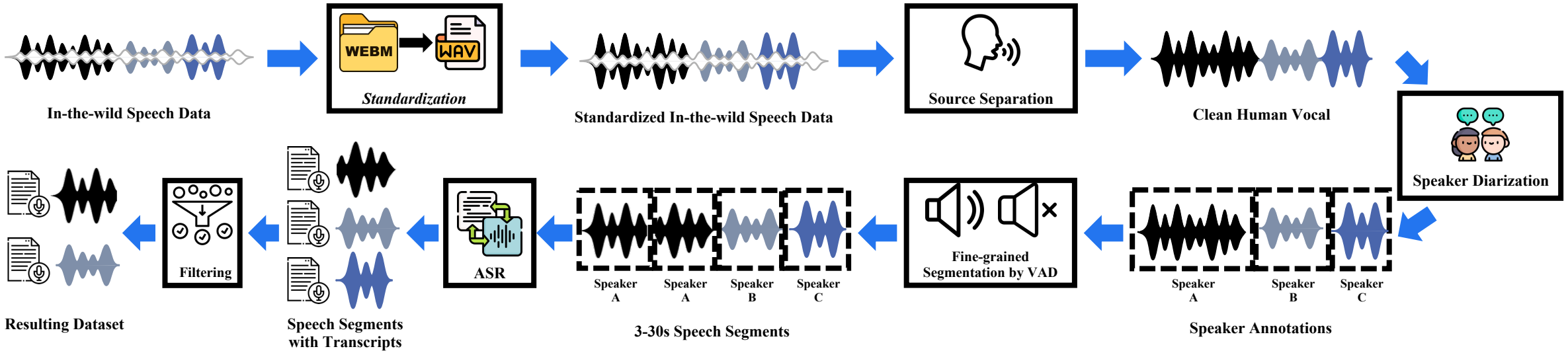
自动化处理框架 Emilia-Pipe – 特征

- **首个开源自动化处理框架**：专为语音生成任务设计的大规模数据处理框架，高效地将**“野生”**音频数据转换为语音生成大模型训练所需的标准化数据。
- **工程优化**：**一分钟内处理可以数小时**的语音数据，提升大规模数据处理效率



自动化处理框架 Emilia-Pipe – 组成

- **标准化**：将所有音频转换为一致的 WAV 格式，设置为 24kHz 采样率，并进行振幅归一化处理，以确保数据一致性。
- **声源分离**： 去除背景噪音和音乐，从而得到干净的语音数据。
- **说话人分离**：尽量确保每个语音片段只包含一个说话人，提升语音克隆模型的有效性。
- **精细化分割**：对长语音进行语音活动检测和分割，将长片段切割为 3 到 30 秒的短片段
- **自动语音识别**：通过自动语音识别，能为每个语音片段生成高质量的文本标签。
- **过滤**：过滤低质量音频，去除语音和文本不匹配的片段，避对模型训练的负面影响。



自动化处理框架 Emilia-Pipe

- **标准化**

输入

- 不同的音频格式 MP3, M4A, WEBM, MKV, ...
- 不同的音频采样率 44.1kHz, 24kHz, 16kHz, 8kHz, ...
- 不同的音频振幅

输出

- 统一的音频格式 WAV
- 统一采样率 24kHz
- 统一标准化音频振幅

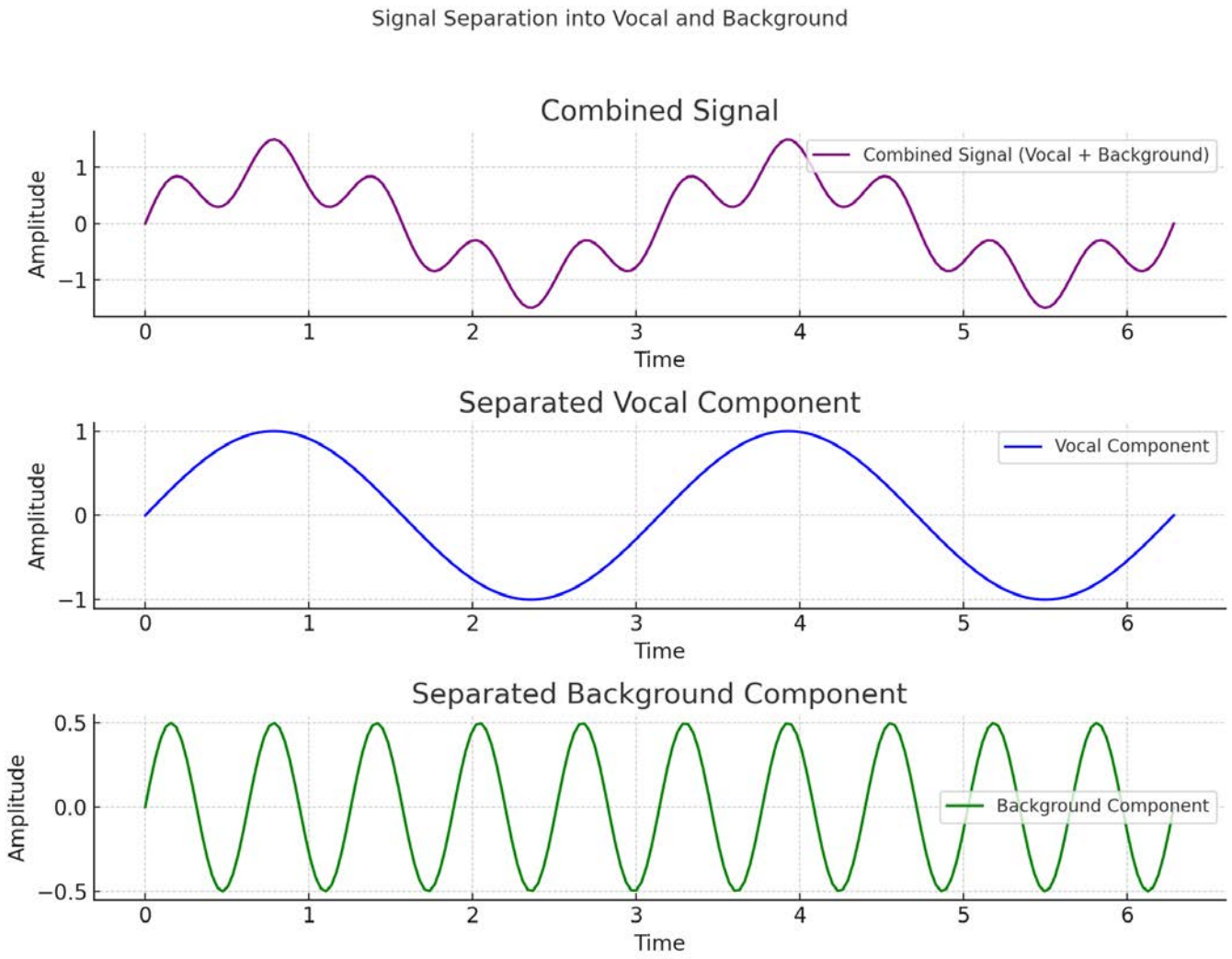
自动化处理框架 Emilia-Pipe

- **声源分离**

输入：
人声+背景声

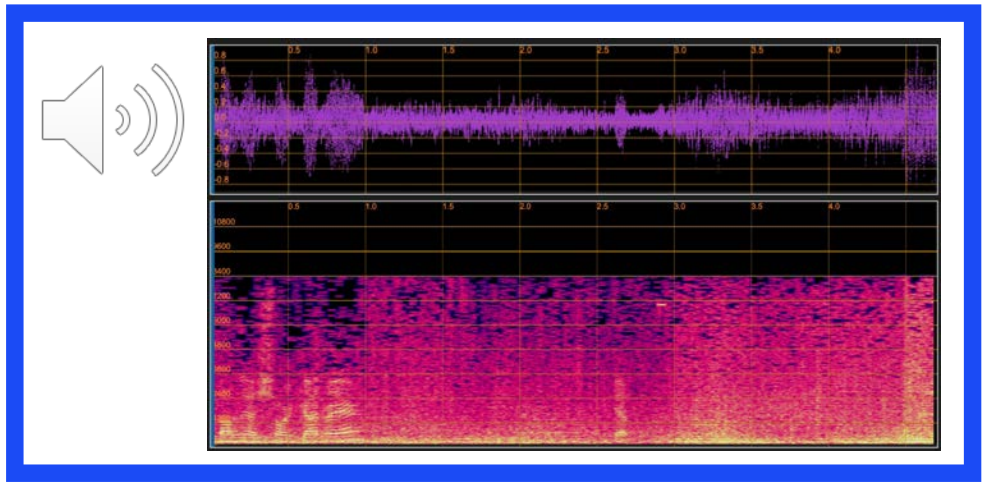
输出1：人声
（保留）

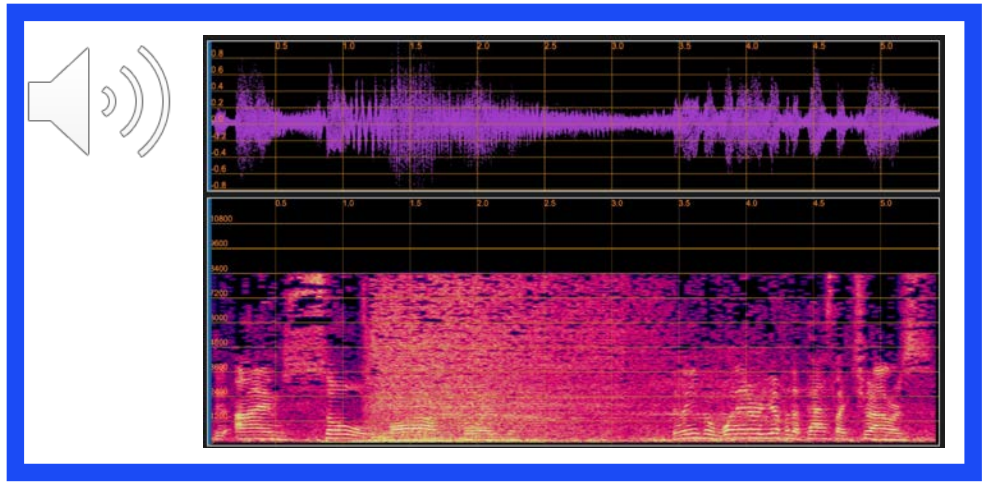
输出2：背景声
（丢弃）



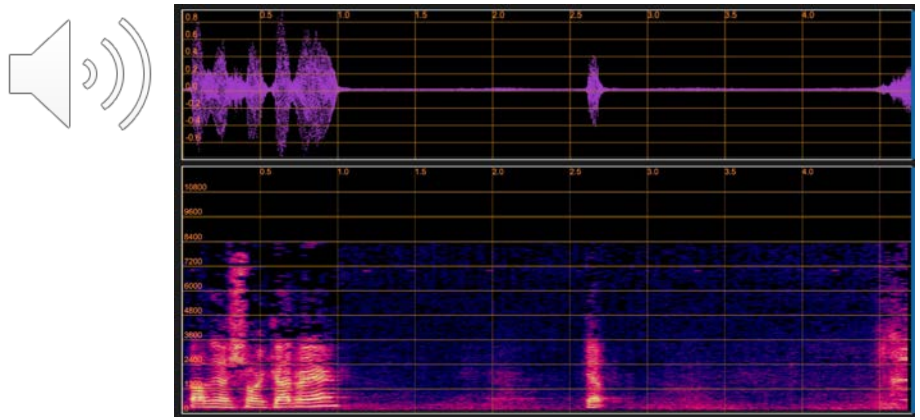
自动化处理框架 Emilia-Pipe

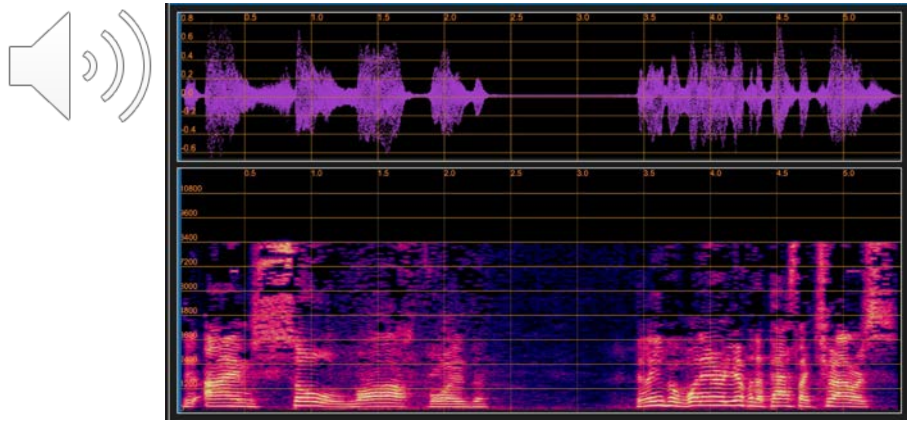
- **声源分离**
  - 效果：
    - 分离前：





    - 分离后：





自动化处理框架 Emilia-Pipe

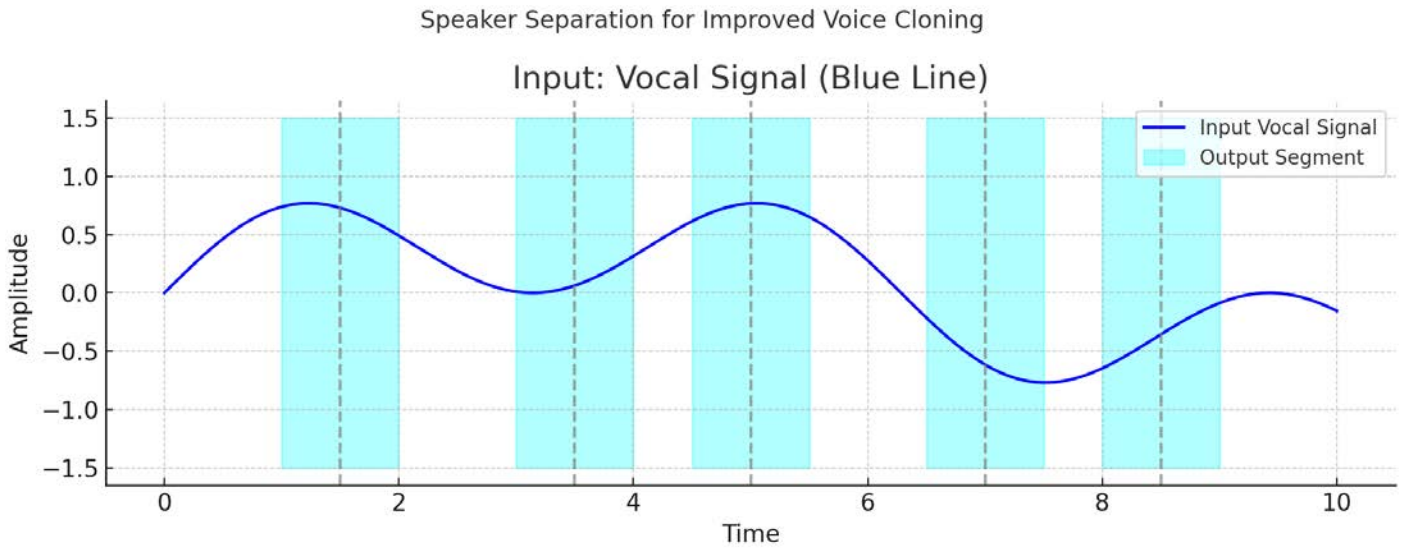
- **声源分离**
  - 目的
    - **提高语音质量**，减少模型训练中的噪音干扰。
    - **提取清晰人声**，便于后续的说话人分离和语音转录。
  - 模型：UVR-MDX-Net Inst 3 （效果和速度的折衷选择）
  - 挑战
    - 复杂场景难以分离：如短视频背景声音复杂
    - 低采样率音频的背景声难分离：人声集中在低频，如果采样率较低，分离难度较大
    - 存在混响、回声、背景说话人：难以分辨和去除

自动化处理框架 Emilia-Pipe

- **说话人分离**
  - 目的：尽量确保每个语音片段只包含一个说话人，提升语音克隆模型的有效性。
  - 模型：PyAnnote 3.1。核心组件：说话人分段、嵌入和聚类。
  - 挑战：大部分能实现分离，仍然存在一些sample会有2个说话人。“鸡尾酒会”任务仍然是语音界的共同挑战，好在大量实验结果证明，目前的效果已经能保证零样本音色克隆的有效性。

输入：人声

输出：人声片段 * n

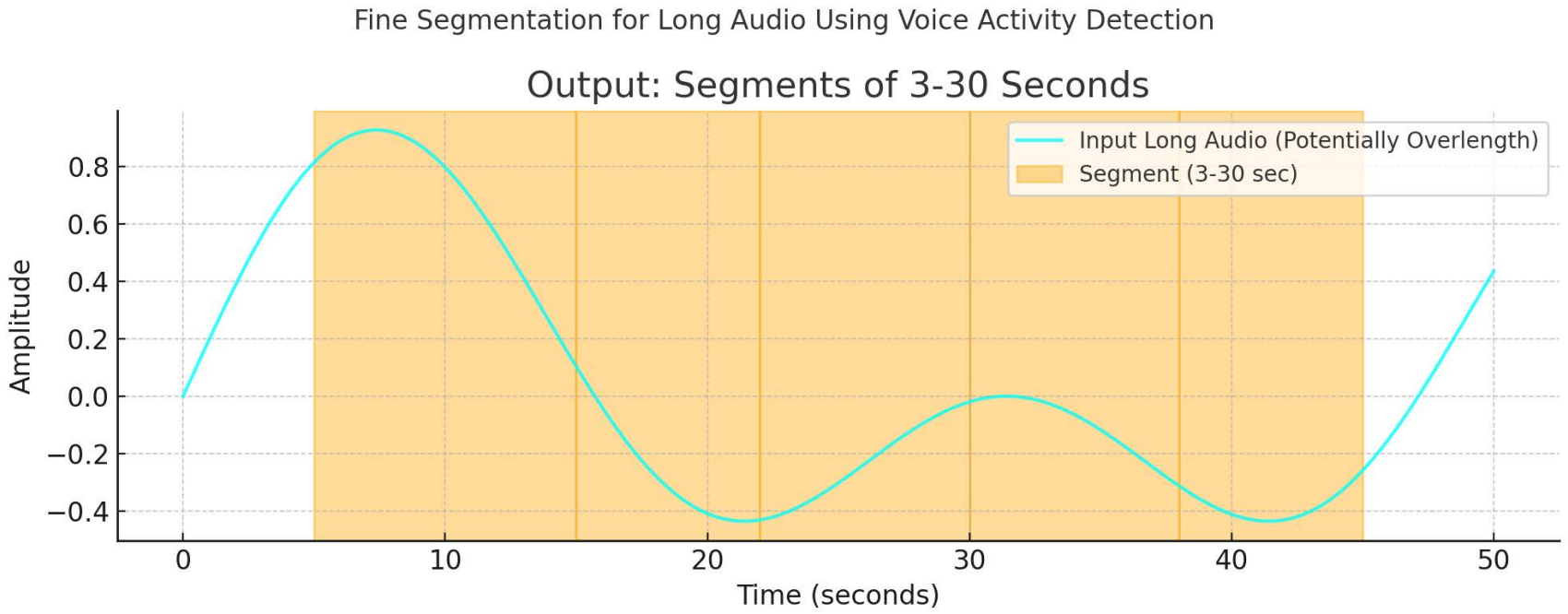


自动化处理框架 Emilia-Pipe

- **精细化分割**

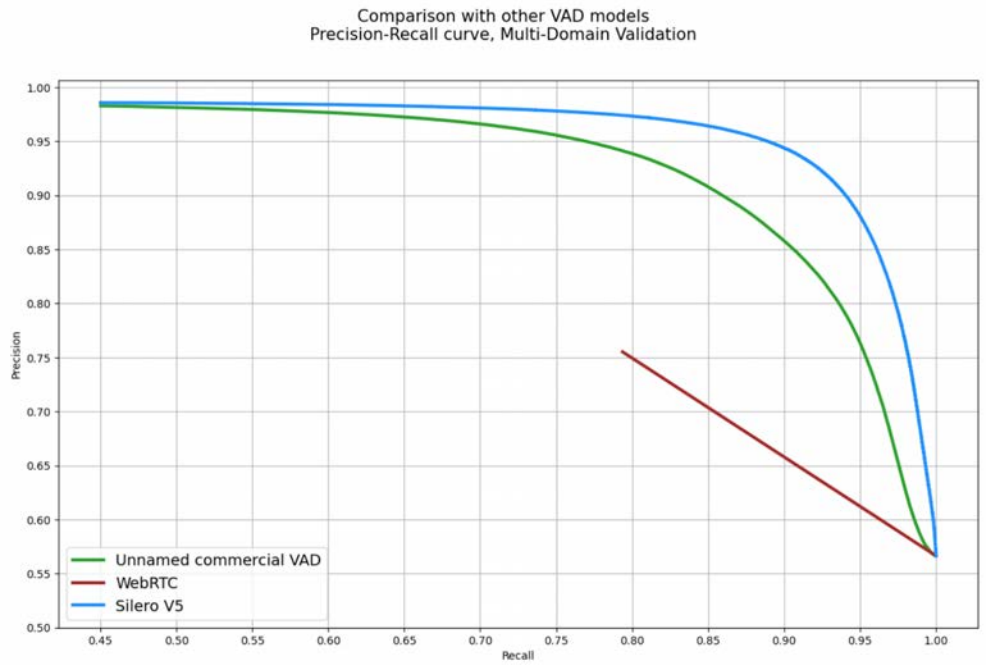
输入：人声片段（任意长度）

输出：人声片段（3-30s）



自动化处理框架 Emilia-Pipe

- **精细化分割**
  - 目的
    - 对长语音进行语音活动检测和分割，**去除静音段**提高数据质量，减少无效数据。
    - 将长片段切割为 3 到 30 秒的短片段，适合模型处理。
  - 模型：Silero VAD (又快又好)
  - 挑战：
    - 处理长音频的内存限制和分割精度问题
    - 如何确保连续性和连贯性

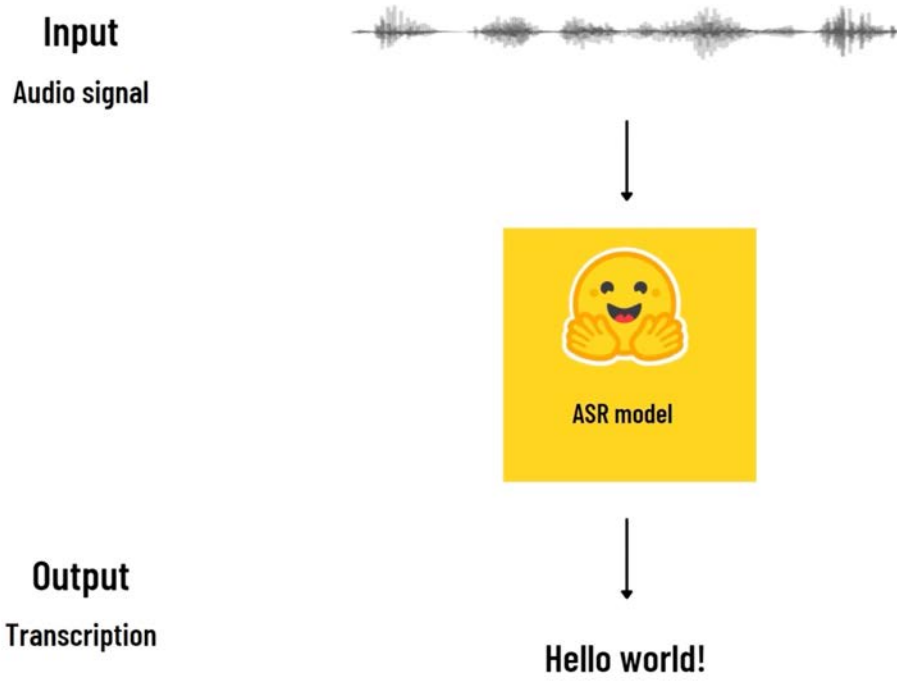


Silero VAD 比商业模型表现更好。

自动化处理框架 Emilia-Pipe

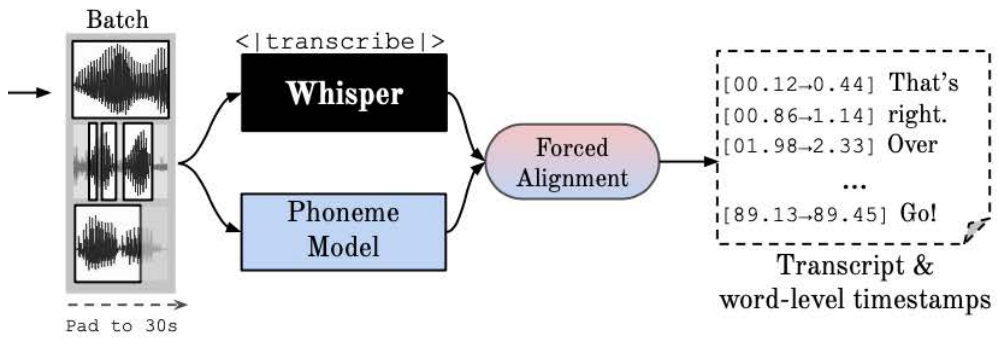
- **自动语音识别**
  - 目的：通过自动语音识别，能为每个语音片段生成高质量的文本标签，提供给模型训练。
  - SOTA模型：Whisper

输入：音频信号
输出：文本转录



自动化处理框架 Emilia-Pipe

- **自动语音识别**
  - 实际模型：
    - WhisperX
    - 高效的多语言 ASR 能力，结合 faster-whisper 和 CTranslate2 进行加速。
    - 实现并行转录，提高处理效率。



This repository provides fast automatic speech recognition (70x realtime with large-v2) with word-level timestamps and speaker diarization.

该存储库提供快速自动语音识别（使用大型 v2 实现 70 倍实时），具有字级时间戳和说话人二值化。

- ⚡ Batched inference for 70x realtime transcription using whisper large-v2
  ⚡ 使用 Whisper Large-v2 进行 70 倍实时转录的批量推理
- 🪶 faster-whisper backend, requires <8GB gpu memory for large-v2 with beam_size=5
  🪶 更快的 Whisper后端，需要 <8GB GPU 内存用于大型 v2（beam_size=5）

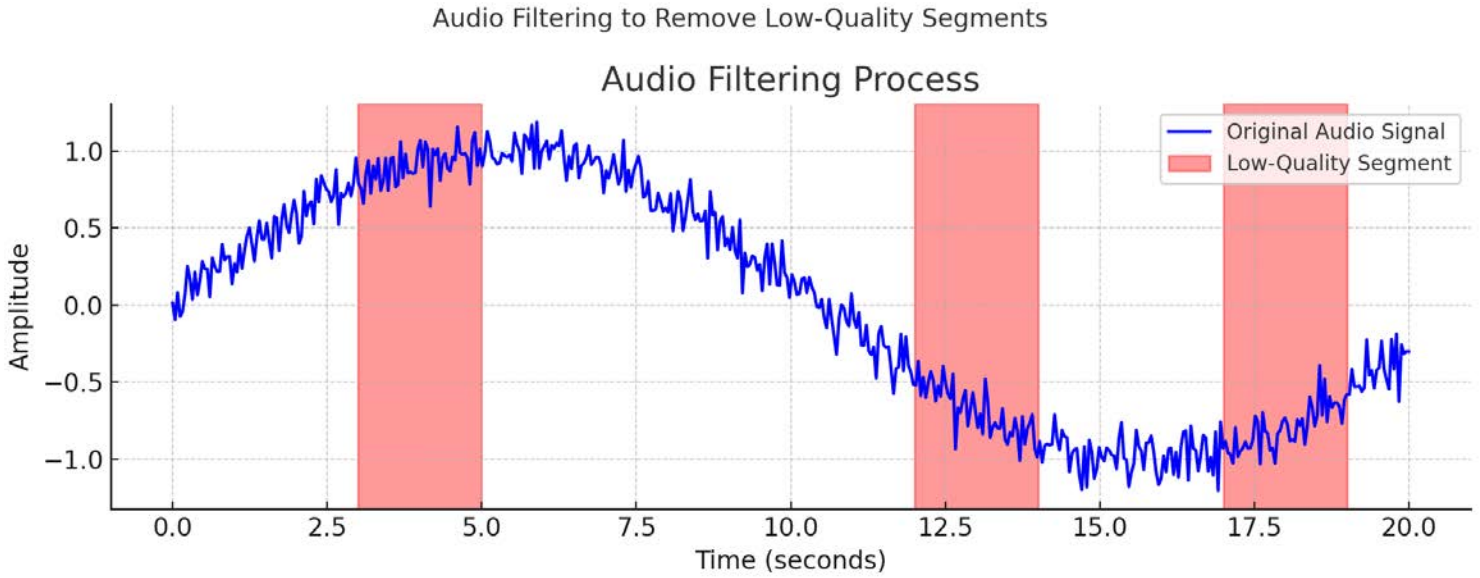
自动化处理框架 Emilia-Pipe

- **自动语音识别**
  - 挑战：
    - 速度与效果的trade-off：ASR模型量化 ---> 提升速度，降低效果
    - 幻觉问题：Whisper模型基于自回归架构容易出现幻觉（“I love youuuuuuuuuuuuu...”）
      - 出现的情况: (1) 突然静音 （2）中夹英 （3）重复语句 （4）音质不好
  - 我们的经验
    - 量化后的 faster-whisper-medium + 自定义 prompt 是当前最佳选择
    - 模型大有幻觉且慢，小效果差 ---> medium 比 large 更稳定，small 语言不支持效果不好，turbo 在特定语言能力下降厉害
    - 结果过滤很重要 ---> 通过 rule based 的方法（例如平均音素时长）过滤掉幻觉结果

自动化处理框架 Emilia-Pipe

- **过滤**

输入：音频片段+文本转录
输出：过滤后的文本转录
（红色部分被丢弃）



自动化处理框架 Emilia-Pipe

- **过滤**
  - 目的：过滤低质量音频，去除语音和文本不匹配的片段，避对模型训练的负面影响。
  - 挑战
    - 去除有多个说话人的音频：可能导致生成模型输出音色突变。
    - 确保文本和音频是匹配有关的：可能导致模型输出不想要的音频。
    - 确保没有筛选 bias：不会导致模型在某一方面表现极差。
    - 减少重复数据出现：防止模型过拟合、背训练数据。

自动化处理框架 Emilia-Pipe

- **过滤**
  - 我们的经验
    - Pipeline 更重要部分在于过滤低质量数据
    - 音质：DNSMOS P.835：阈值 2.4，移除低质量片段。
      - DNSMOS 分对不同对语言有偏好（英语分会较高一些导致留存率更高，韩语较低）
      - DNSMOS 有三个分数 OVRL, SIG, BAK 可以根据具体需要使用这三个分数
    - 幻觉：
      - 平均音素时长异常值检测：avg phone duration + IQR * 1.5。
      - 基于文本的子串重复检测
    - 错字：
      - 多个 ASR 模型打分投票 / 支持逐句置信度的模型打分

自动化处理框架 Emilia-Pipe

- **最终效果**
  - 中文：

    中场休息,休息后欢迎跳转下半场。

    在国内呢,百度腾讯今日头条等等都纷纷放出了自己在研发类似产品的消息。

    在那个时候,万物的种子都混杂在一起,形成一个个不规则的物体,向着各自的方向运动。在运动的过程中,这些原始的东西开始慢慢分离。

  - 英文：

    That is hardcore greed if I've ever seen it.

  - 日语：

    じゃあ、そのグラフは変えて、あ、それから会議室のパソコンやマイクの準備はできてる?

Emilia 数据集开发的经验教训

- **工程速度挑战：**在大规模数据处理的速度决定了能否高效 Scale up 训练数据。
  - 目前：在 24 张 RTX 4090 上，每天能生成 2000 至 5000 小时的数据。
  - 速度是初版Emilia-Pipe的10倍：当时我们是需要两年才能得到10万小时的数据。
- **Emilia的优化和项目管理措施**：
  - **离线 + 量化压缩模型**：减少模型动态下载，网络不稳定问题。加快推理速度。
  - **降低磁盘 I/O**：全程将音频数据保存在内存中。
  - **并行化一切可以并行的东西**：例如音频并行转录、多线程写 MP3 等等
  - **GPU 调用的多线程处理**：每个模型都有其独立的锁。使用 GPU 时，八个线程会竞争 GPU 锁；使用 CPU 或进行 I/O 时并行处理。
  - **Master-worker 结构：**任务分配与日志收集。可以考虑用更高性能的 Kafka 管理分配。
  - **Docker 镜像 + k8s 调度：**支持集群监控管理，更轻松的环境配置。
  - **其他策略：**机器人 Pipeline 状态通知、 Prometheus 日志检查机制等。

Emilia 数据集开发的经验教训

- **超大规模数据集的开发经验：**
  - 数据传输和分发的限制：
    - 文件夹内文件数量 < 1000；tar 文件大小 < ~10G。
  - 索引文件：处理约 TB 级别的数据时需要索引文件
    - 读取速度：有索引 >> 无索引
  - Tar 文件：处理约 100 TB 级别的数据时需要 tar/bin 文件，结构化打包数据。
    - 读取 I/O 成本：大文件 < 分散的小文件。
    - 传输成本：大文件 << 分散的小文件。
    - 文件系统支持：大文件 > 分散的小文件（在共享存储系统上）。
  - Parquet / Webdataset
    - 结构、有效的大数据结构有助于数据集准备、开发、测试、使用。

未来发展方向和机遇

- **实时与低延迟应用**：通过不断优化模型的推理速度和生成质量，可以进一步提高语音生成在这些场景中的实用性，如手机虚拟助手、智能语音音响、汽车语音中控、客户服务等。
- **个性化语音定制**：适用于明星、视频/影视创作者、新媒体从业者，帮助他们构建个人虚拟形象，用个性化语音生成为他们提供诸如视频配音等帮助。也适用于虚拟助手的个性化定制。
- **多语言与跨语言能力**：学术界对跨语言合成和语音克隆对探索，使得应用变得更加实际，赋予全球应用更大的灵活性。这些模型的多语言能力使它们在不同国家和地区的应用场景中具有广泛的适用性。
- **伦理/滥用考虑**：随着 TTS 模型的进步，隐私、偏见和滥用（例如语音伪造）等问题日益受到关注。也有模型考虑到这些风险，融入了防滥用的水印和日志机制，以防止合成语音被用于恶意目的。进而保障用户的隐私和安全，同时增强公众对生成技术的信任。

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 · 敦煌站
K+ 思考周®研习社
时间：2025.08.29-30

K+峰会 · 上海站
K+ 金融专场
时间：2025.10.17-18

K+峰会 · 香港站
K+ 思考周®研习社
时间：2025.11.25-26

K+峰会详情

AiDD峰会

AiDD峰会 · 上海站
AI+研发数字峰会
时间：2025.05.17-18

AiDD峰会 · 北京站
AI+研发数字峰会
时间：2025.08.08-09

AiDD峰会 · 深圳站
AI+研发数字峰会
时间：2025.11.28-29

AiDD峰会详情

AiDD AI+研发数字峰会 第5届
AI+ Development Digital summit

利用AI技术深化计算机对现实世界的理解

推动研发进入智能化时代