

数字内容生产线

声音制作录制指导

文档版本 01
发布日期 2024-12-11



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目录

1 真人声音录制.....	1
2 手机录制音频.....	4

1 真人声音录制

客户录制真人音频，上传至MetaStudio进行AI训练，即可得到和真人音色1:1复刻的声音模型。

声音模型可实现文本转语音，应用于数字人视频制作、直播、交互问答等场景中。不同版本录制规格如下：

- 基础版：整段录制生成一个WAV或MP3格式音频，总时长3~10分钟，建议5分钟。
- 进阶版：整段录制生成一个WAV或MP3格式音频，总时长10~30分钟，建议15分钟。
- 高品质：整段录制生成一个WAV或MP3格式音频，总时长1小时以上，建议1小时。

原声录制音频和其对应生成的音色播报示例，如客户声音制作案例所示。

录音准备

表 1-1 录音准备

声音类型	录制设备和软件	录音环境	录音文案
基础版	优先使用专业录音设备录制音频，推荐使用 Adobe Audition 软件进行录音。 如果不具备专业录音设备，可使用手机录制音频，详情请参见 手机录制音频 。	<ul style="list-style-type: none">● 录音环境安静，无回音、混响、噪声，避免出现汽车鸣笛、他人说话、走动等杂声。● 可使用“分贝仪”应用来测试录音环境的底噪，建议底噪低于0dB。● 同一批录音必须使用同一套录制设备和环境。	建议使用对应版本预置的文案样例：文案样例（基础版）、文案样例（进阶版）、文案样例（高品质）。 支持自定义文案，单句文案长短需和样例相当。 不建议临场即兴发挥录制，避免出现较多嗯、啊等语气词，影响连贯性。
进阶版			
高品质	专业录音棚+高保真MIC录制。		

开始录音

必须录制单人高质量、无噪音或背景音的人声音音，生成音频。支持使用苹果或者安卓手机录制视频，操作请参考[手机录制音频](#)。

具体的录音注意事项，如[表1-2](#)所示。

表 1-2 录音注意事项

录音事项	说明
话筒间距	调整与麦克风之间的距离，以一拳距离为宜。不宜离麦太近，防止喷麦或录入呼吸声。
录音内容	每句文案 起始数字编号无需阅读 。 示例：4. 它不仅拥有出色的功能，还具备卓越的性能，序号4无需阅读。
音频格式	推荐使用无损音质格式保存音频文件，如WAV、MP3格式。 原始录音数据，需为48kHz采样率、16bit编码和单声道。
录制风格	录制全程保持语境风格一致， 避免多种情绪混杂 。如直播场景时，录制风格需接近日常直播效果。
发音吐字	发音吐字清晰、准确，音量适中 。如果出现喷麦或嘶嘶声音，建议当前句子重新录制。
语速节奏	语速自然、平稳， 切忌过快过慢、忽快忽慢 。
音量适中	音量不能过小、过大、或忽大忽小，甚至削波爆音。峰值RMS在-9左右，无削波。
停顿断句	在 标点或适当断句处自然停顿 ，需轻声换气。 当录制一个长音频文件时，每句之间需要有2~3秒的停顿。
重音位置	重音位置要合理， 避免错误的重音 。
阅读发音	按顺序读，确保音字一致。避免 漏字多字、发音错误、阅读不流畅 等问题。如有这些问题，需 重新录制 。
内容要求	不能使用短音频复制拼接成长音频，会影响训练效果，无法通过审核。

录制提交

表 1-3 录音内容提交规范

声音类型	音频说明	音频命名
基础版	需 整段录制 所有语料生成一个长音频WAV或MP3格式文件，每句之间有2~3秒的停顿。WAV或MP3格式文件可直接上传MetaStudio控制台，无需压缩，无需提供语料txt文件。 推荐使用服务预置语料，也可自定义语料。系统会自动根据停顿做切割，自动识别文本内容。	音频文件命名无要求，可自定义，命名示例：Voice.wav。
进阶版		
高品质		

制作声音模型

准备好音频文件后，就可以上传至MetaStudio控制台，进行声音训练。详细操作如下所示：

- 创建声音制作任务（自研模型）
- 查看声音

声音模型制作耗时，如下所示：

- 基础版：约1~3个工作日。
- 进阶版：约1~3个工作日。
- 高品质：约5个工作日。

自定义声音应用方式，如下所示：

- 自定义声音生成后，会自动展示在MetaStudio控制台声音列表中，可用于分身数字人视频制作、视频直播或智能交互等场景中。
- 通过MetaStudio的API调用自定义声音。

2 手机录制音频

请根据手机类型，执行相应操作：

- [苹果手机操作](#)
- [安卓手机](#)

苹果手机操作

苹果手机推荐下载软件“简易录音机（Easy Voice Recorder）”进行录音，如[图2-1](#)所示。

图 2-1 简易录音机



音频录制说明，如[表2-1](#)所示。

表 2-1 音频录制说明

音频录制	说明
采样率	<p>推荐使用48kHz采样率录制音频。</p> <p>苹果手机设置简易录音机录音参数的操作如下所示：</p> <ol style="list-style-type: none"> 1. 打开简易录音机软件，切换至设置“Settings”页面。 2. 选择“Format”，进入“Format”页面。 设置如下选项： <ul style="list-style-type: none"> ● 文件格式：选择“.wav (PCM)”或MP3格式。 ● 采样率：选择“48 kHz (high quality)”。 3. 返回至录音页面，开始录制。
采样位	<p>推荐使用48bit采样位录制音频。</p> <p>苹果手机的简易录音机的采集位默认为16bit，无需修改。</p>
声道	<p>推荐使用单声道录制音频。</p> <p>苹果手机的简易录音机默认为单声道录制，无需修改。</p>
音频格式	<p>推荐使用无损音质格式保存音频，如WAV、MP3格式。</p>
音频命名	<p>所有语料录制生成一个长音频WAV或MP3文件，命名示例：Voice.wav。</p>
音频导出	<p>操作如下所示：</p> <ol style="list-style-type: none"> 1. 打开简易录音机软件，单击底部的“播放”菜单，进入“播放”页面。 2. 单击界面右上方的“编辑”，进入音频文件编辑页面。 3. 选中所有的音频文件，单击“分享”，可选择隔空投送或其他第三方应用导出音频文件至PC中。 一个长音频文件无需压缩，可直接上传至MetaStudio控制台，进行声音训练。

安卓手机

安卓手机推荐下载软件“超级录音机”进行录音，如图2-2所示。

图 2-2 超级录音机



音频录制说明，如表2-2所示。

表 2-2 音频录制说明

音频录制	说明
采样率	推荐使用48kHz采样率录制音频。 安卓手机的超级录音机的采样率默认为高清 44100Hz，无需修改。
采样位	推荐使用48bit采样位录制音频。 安卓手机的超级录音机的采集位默认为16bit，无需修改。
声道	推荐使用单声道录制音频。 安卓手机的超级录音机默认为单声道录制，无需修改。
音频格式	推荐使用无损音质格式保存音频，如WAV、MP3格式。 安卓手机的超级录音机的录音参数设置方式，如下所示： 1. 打开超级录音机软件。 2. 为了能正常录音和保存文件，需要先授权给软件录音和存储权限，即允许“超级录音机”访问媒体。 3. 进入软件设置页面。 设置如下选项： <ul style="list-style-type: none">● 录制设置 > 录音音质：选择“高清”。● 录制设置 > 录音保存格式：选择“WAV”或“MP3”。 4. 退出设置页面，进入录制页面开始录制。
音频命名	所有语料录制生成一个长音频WAV或MP3文件，命名示例：Voice.wav。
音频导出	操作如下所示： 1. 打开超级录音机软件，进入音频文件列表页面。 2. 通过长按音频文件的方式，选中所有文件，单击分享图标，可选择隔空投送或其他第三方应用导出音频文件至PC中。 一个长音频文件无需压缩，可直接上传至MetaStudio控制台，进行声音训练。