MetaStudio

Voice Recording Guide

Issue   01
Date   2024-12-18

Huawei Cloud Computing Technologies Co., Ltd.

Contents

1 Recording a Human Audio

You can upload a human audio recording to MetaStudio for AI training to obtain a voice model that reproduces the human timbre at 1:1.

The voice model can be used for text-to-speech conversion and applied to scenarios such as virtual avatar video production, livestreaming, and intelligent interaction.

For voice modeling, record and generate an entire WAV or MP3 audio file of 10 to 30 minutes (recommended: 15 minutes).

Preparing for Recording

Table 1-1 Recording preparations

Recording Device and Software	Recording Environment	Recording Script
Professional recording devices (recommended: **Adobe Audition**) are preferred for audio recording. If professional recording devices are not available, you can use your mobile phone for recording. See **Recording an Audio on a Mobile Phone**.	• Try to record the audio in a quiet environment without echo, reverberation, and noises caused by car horns, talking, or walking. • You can use the decibel (dB) meter app to test the background noise in the recording environment. **The background noise should be lower than 0 dB.** • The recording device and environment cannot be changed for the same recording task.	You are advised to use Script Examples (Advanced Edition). You can also customize the script. The length of one phrase must be the same as that in the example. Improvised recording is not recommended as there may be too many fillers that compromise the speech coherence.

Starting Recording

The recorded audio must be high-quality, free of noises and background sounds, and of the same person. You can use an iPhone or Android mobile phone to record videos. See **Recording an Audio on a Mobile Phone**.

Table 1-2 describes the precautions for recording.

Table 1-2 Recording precautions

Item	Description
Distance from the microphone	Adjust the distance from the microphone. The one-punch distance is appropriate. To avoid pop sound effects or recording the breath sound, do not be too close to the microphone.
Recording content	The **starting number** of each piece of script **does not need to be read**. For example, for the script "4. It features a multitude of functions and superior performance", **4** does not need to be read.

Item	Description
Audio format	Save the audio file in a lossless format, such as WAV and MP3. The recording data should not be encoded (sample rate of 48 kHz, sample bit of 16 bits, and mono).
Speech style	Keep the speech style consistent throughout the recording to **avoid excessive emotions**.
Pronunciation	**Pronunciation should be clear and accurate, and the volume should be moderate.** If there is undesired sound, record the phrase again.
Speed and rhythm	The speed of speech should be natural and stable. **Do not be too fast or too slow**.
Moderate volume	The volume cannot be too low or too high, or fluctuate. Clipping noise is not allowed.
Pause	**Pause naturally** and breathe softly **upon punctuations and appropriate positions**. There must be a pause of 2–3 seconds between phrases for a long audio file.
Accent position	Find the correct accent position to **avoid wrong accent**.
Reading pronunciation	Read in order, ensure the phonetic consistency (avoid **missing or adding words**), and avoid **wrong pronunciation**. If there is a misreading or the **reading is not smooth**, **record the whole phrase again**.
Content	Merging several audio files into one audio file for training will fail the review.

Submitting an Audio File

Record all phrases in **one single** WAV or MP3 audio file, with a pause of two to three seconds between each phrase. You can upload the WAV or MP3 file to the MetaStudio console without compressing it or providing a TXT script file. The preset script is recommended. You can also customize the script. The text is automatically split based on pauses and identified.

You can customize the audio file name, for example, Voice.wav.

Creating a Voice Model

After the audio file is available, you can upload it to the MetaStudio console for voice training by following:

- Creating a Voice Modeling Task
- Checking a Voice

The task takes about seven working days.

Application scenarios of a customized voice:

- After a customized voice is generated, it is automatically displayed in the voice list on the MetaStudio console. This voice can be used for virtual avatar video production, livestreaming, or intelligent interaction.
- A customized voice can be called using the APIs of MetaStudio.

2 Recording an Audio on a Mobile Phone

Perform operations based on the mobile phone type:

- **Operations on iPhone**
- **Operations on Android**

Operations on iPhone

Easy Voice Recorder is recommended for recording on iPhone.

See **Table 2-1**.

Table 2-1 Audio recording requirements

Audio Recording	Description
Sampling rate	48 kHz is recommended for audio recording. Configure the recording parameters of Easy Voice Recorder: 1. Open Easy Voice Recorder and switch to the **Settings** page. 2. Select **Format**. The **Format** page is displayed. Configure the following parameters: • **FILE FORMAT**: Select **.wav (PCM)** or **MP3**. • **SAMPLE RATE**: Select **48 kHz (high quality)**. 3. The recording page is displayed and the recording starts.
Sampling bit	48 bits is recommended for audio recording. The default sampling bit of Easy Voice Recorder is 16 bits. You do not need to change it.
Audio channel	Mono is recommended for audio recording. Mono is set by default for Easy Voice Recorder. You do not need to change it.
Audio format	Save the audio file in a lossless format, such as WAV and MP3.

Audio Recording	Description
Audio naming	All script pieces are recorded in a WAV or MP3 file (file name example: Voice.wav).
Audio export	Procedure: 1. Open Easy Voice Recorder and click **Listen** at the bottom. 2. On the page displayed, click **Edit** in the upper right corner. The page for editing audio files is displayed. 3. Select all audio files and click **Share**. You can transfer the audio files to the PC or use other third-party applications to export the audio files to the PC. You can upload a long audio file to the MetaStudio console for voice training without compressing it.

Operations on Android

Super Recorder is recommended for recording on Android.

See **Table 2-2**.

Table 2-2 Audio recording requirements

Audio Recording	Description
Sampling rate	48 kHz is recommended for audio recording. The default sample rate is 44,100 Hz (HD). You do not need to change it.
Sampling bit	48 bits is recommended for audio recording. The default sampling bit of Super Recorder is 16 bits. You do not need to change it.
Audio channel	Mono is recommended for audio recording. Mono is set by default for Super Recorder. You do not need to change it.

Audio Recording	Description
Audio format	Save the audio file in a lossless format, such as WAV and MP3. Configure the recording parameters of Super Recorder: 1. Open Super Recorder. 2. To record and save files properly, you need to grant the recording and storage permissions to Super Recorder. 3. The **Record Options** area is displayed. Configure the following parameters: • **Recording quality**: Select **High**. • **Recording format**: Select **WAV** or **MP3**. 4. Exit the settings page and go to the recording page to start recording.
Audio naming	All script pieces are recorded in a WAV or MP3 file (file name example: Voice.wav).
Audio export	Procedure: 1. Open Super Recorder and go to the audio file list page. 2. Touch and hold an audio file, select all files, and touch the icon of sharing to transfer the audio files to the PC or use other third-party applications to export the audio files to the PC. You can upload a long audio file to the MetaStudio console for voice training without compressing it.