

Artificial Audio Intelligence

产品介绍

Product Introduction



Tencent
Cloud

Copyright Notice

©2013-2017 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Documentation Legal Notice	2
产品介绍.....	4
Product Intro.....	4
Basic Concepts.....	6
Speech Recognition.....	8
Text To Speech	10
Voiceprint Recognition.....	12
Customer Service Robot.....	14

产品介绍

Product Intro

Artificial Audio Intelligence meets various voice processing demands such as speech recognition, text to speech, and voiceprint recognition. Artificial Audio Intelligence has a powerful customized service for vertical industry, which is designed to create a highly professional and efficient voice brain to provide a full range intelligent speech solutions for enterprises.

Product Features

Speech Recognition

Artificial Audio Intelligence supports multiple features including voice-to-text, searching speech offset value according to keyword list, emotion recognition, mute monitoring. It also provides customized service for acoustics and language models.

Text to Speech

Artificial Audio Intelligence can convert text into high-quality voice. It offers multiple options for tones and speeds, and supports Chinese, English or a mixture of both.

Voiceprint Recognition

It provides one-on-one speech recognition to verify speaker's identity. In financial industry, it can be used for verifying user's identity to secure his/her assets.

Cloud + Embedded

Cloud service can be scaled as needed and provides persistent storage capacity. Embedded engine can meet the requirement for Artificial Audio Intelligence in scenarios where network is unavailable. The combination of "Cloud + Embedded" can improve user experience by achieving real-time speech interaction of intelligent hardware and smart home.

Application Scenarios

Voice Quality Testing

The combination of various capabilities such as voice-to-text, emotion recognition, mute monitoring and keyword search can increase the efficiency of voice compliance check. The marketing information hidden in corpus information can also improve the productivity of enterprises.

Voice Search

It is a speech recognition engine customized for vertical industry, which provides an easy way of human-machine interaction in an environment with or without network connection to improve product experience.

Intelligent Customer Service

Locate users' problems through machine learning to eliminate the need for multi-level voice menu navigation. Robot can help you save over 50% of the labor and minimize the operating cost.

Identity Verification

Voiceprint recognition technology has a 99% accuracy on identity verification, which ensures the security for many fields in a reliable manner.

Basic Concepts

Artificial Audio Intelligence involves some basic concepts related to COS and speech recognition. To help you understand the follow-up contents, we explain these concepts as follows:

1. APPID

Tencent Cloud application ID.

How to obtain: Register and log in to the [Tencent Cloud Console](#), and then an APPID is generated immediately.

2. ProjectId

Tencent Cloud project ID. Default is 0.

How to obtain: Log in to the [Tencent Cloud Console](#), and select "Project Management" under "User Center" to view the default project with the ProjectId of 0. Click "New" to create a project and obtain a new ProjectId.

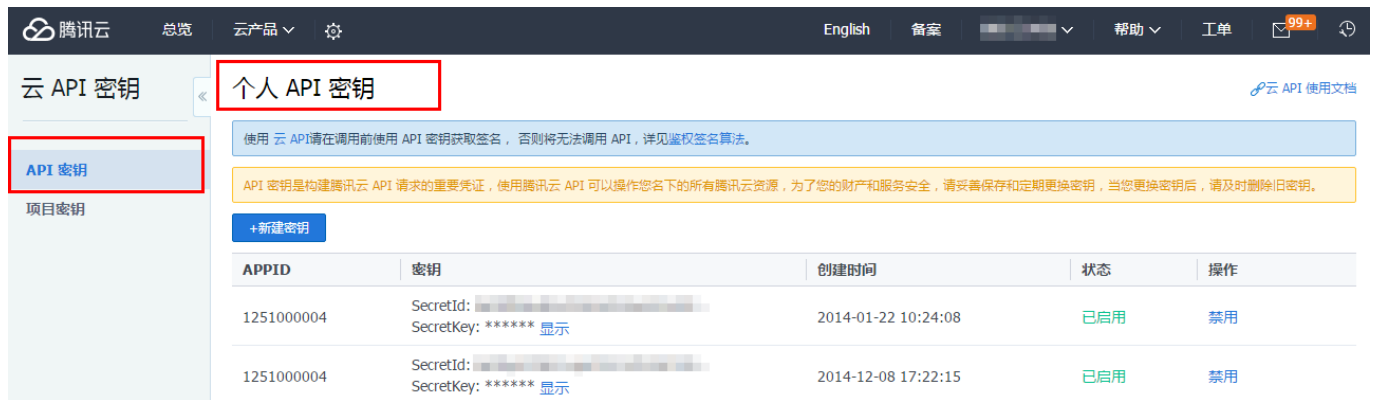
3. SecretId and SecretKey

SecretId and SecretKey are Tencent Cloud security credentials. SecretId is used to identify an API caller, and SecretKey is used to encrypt the signature string and verify its key on the server. Users must keep their SecretKeys private to avoid disclosure.

How to obtain:

- 1) Log in to the [Tencent Cloud Console](#), click "Cloud Products", and select "Cloud API Key" under "Monitoring and Management" to enter the cloud API key management page.
- 2) On the cloud API key management page, click "New" to create a pair of SecretId/SecretKey. Each account can have two pairs of SecretId/SecretKey at most.





The screenshot shows the Tencent Cloud console interface. The top navigation bar includes the Tencent Cloud logo, '总览' (Overview), '云产品' (Cloud Products), and 'English'. The sidebar on the left has '云 API 密钥' (Cloud API Key) and 'API 密钥' (API Key) selected. The main content area is titled '个人 API 密钥' (Personal API Key) and contains a table of API keys.

APPID	密钥	创建时间	状态	操作
1251000004	SecretId: [redacted] SecretKey: ***** 显示	2014-01-22 10:24:08	已启用	禁用
1251000004	SecretId: [redacted] SecretKey: ***** 显示	2014-12-08 17:22:15	已启用	禁用

4. Service Type

Artificial Audio Intelligence provides four types of services, that is, Automatic Speech Recognition (ASR), Text to Speech (TTS), Voiceprint Recognize (VPR), and Customer Service Robot (CSR).

5. Subservice Type

The ASR, TTS, VPR, CSR services of Artificial Audio Intelligence are subdivided into several subservices, such as offline speech recognition and streaming speech recognition under ASR.

6. Template

A template consists of a set of parameters for voice processing. The name of a template must be unique within Project. The required parameters include APPID, ProjectId, service type, subservice type, as well as other parameters required in the subservice. You can create, modify and delete templates in the console. You can also specify templates to deal with requests submitted to Artificial Audio Intelligence.

7. Bucket

Bucket is the first-level directory under COS. Each Bucket has its own access domains (public, private, and accelerated domains), access permissions (public and private), origin-pull settings, hotlink protection settings, and other attribute configuration items.

Speech Recognition

Featuring the most advanced learning algorithm in the industry, Automatic Speech Recognition (ASR) provides excellent capabilities of voice-to-text, keyword search, mute detection, speed detection, and emotion recognition. It completely meets the voice processing needs in scenarios of telephone recording quality control, real-time voice input, LVB subtitles and review.

Advantages

Powerful recognition capability: Based on Tencent social data platform, ASR accumulates more than 100 thousand hours of annotation voice data, possesses massive language-layer data, and provides competitive recognition accuracy.

A complete set of solutions: ASR adopts industry-leading speech recognition modeling methods including LSTM, BLSTM, Deep CNN, LFMMI, CTC revaluation and other acoustic modeling techniques.

Wide application: ASR has been fully verified in major internal and external business lines, including zhenai.com, zhihu.com, Beijing Intellectual Property Court, WeChat and QQ. It covers a wide range of industries, serves 100 million DAUs, and applies to a rich variety of business scenarios.

Speech to Text

Speech-to-text feature can convert speech into text. Speech input device can be a phone, mobile phone, gooseneck microphone, or microphone array. In addition, you can specify a generic or specific field engine based on your business needs. It is widely used in scenarios such as transcription of telephone records, input of electronic medical records, synchronization of subtitles for conferences, and adding subtitles in LVBS.

Keyword Search

Keyword search feature can help you search the business-focused words from your voice files. You can submit a maximum of 200 keywords at a time. With this feature, you can locate keywords in a piece of speech, and determine their time offset from the beginning of the speech, and also obtain confidence level of the result. It is widely used in telephone recording, LVB, and APP audio quality check.

Speed Detection

Speed detection feature can detects the speed of speech, and gives the number of words spoken in a minute. Compared to manual detection, it is more accurate, efficient with lower cost. This feature is widely used in scenarios such as telephone speed monitoring, speed monitoring in the simultaneous interpretation scenario, speed evaluation for television and broadcast announcers.

Text To Speech

Text to Speech (TTS) is designed to generate speech based on a known text, to create a closed loop for human-machine interaction. TTS provides a variety of tone options and allows you to customize volume and speed. It provides enterprise customers with custom vocabulary and personalized speaker services to make pronunciation more natural, professional and suitable to scenarios. TTS is widely used in voice navigation, audiobooks, reading with standard pronunciation, automatic news broadcast and other scenarios.

Advantages

1. Leading algorithm

The TTS system developed by Tencent based on Hidden Markov Model (HMM) adopts the advanced deep learning synthesis algorithm in the industry, featuring fast synthesis speed as well as natural and fluent synthesized speech. It makes machines and applications "speak" easily, and achieves human-machine interaction in a more smooth and efficient manner.

2. High stability

With a massive user base, the service is proven to be stable and efficient, and better matches business scenarios. It provides service to such products as WeChat, map.qq.com, fm.qq.com, tianqi.2345.com with 100 million DAUs.

Product Features

It supports synthesis in Chinese, English, bilingual and other languages. Properties such as volume and speed can be configured based on business needs, and more options of human voices are available. The rich emotional corpora and the speech synthesis engine that applies advanced machine learning algorithm in industry make the synthesized speech more natural and close to human reading level.

Common Scenarios

Intelligent Customer Service Robot

Intelligent customer service robot automatically recognizes customer's phonetics and semantics, and automatically responds to customers' questions with voice through TTS, to

create a closed loop for human-machine interaction.

Electronic Audio Books

With audiobooks, people can enjoy learning while driving, walking or at a moment when it is not convenient to read. Manual synthesis of audiobooks takes time and effort, and the accuracy cannot be ensured. TTS makes it easier to create audiobooks. With emotion synthesis technology, voice becomes more natural and pleasant. Custom speaker feature supports personalized voices to meet various business needs.

Smart Education

In combination with the smart education system, TTS can achieve standard reading and guided reading of English/Chinese phonemes, words, phrases and texts. In addition to the public basic education-based classes, TTS is widely used in extra-class education and training institutions and teaching auxiliary software.

Voiceprint Recognition

Voiceprint Recognition (VPR) performs identification based on the speaker's acoustic characteristics, and can be used for speaker recognition and validation. Accent and language do not affect the identification.

Advantages

a. High stability

With a massive-level user base, the service proved to be stable and reliable, and can expand along with the business. It has been applied in WeChat voice lock, WeBank remote account opening, and other scenarios, serving more than 100 million users.

b. High security

Voiceprint characteristics and speech are verified at the same time, which provides a two-factor verification for enhanced security and reliability. Voiceprint verification with random numbers or text can determine whether the playback unit is a counterfeit recording.

Number-based Password Recognition

You can read out instead of inputting the "digital password" automatically assigned by the system to complete authentication. This feature eliminates the problem of forgotten or confused passwords and provides fast and secure authentication in real time.

Text-based Password Recognition

During registration, you need to read out a specified text and keep the text password in mind. The system verifies the specified text and voiceprint characteristics at the same time for authentication. You need to keep the text password well and confidential. Besides, you can either use a text password automatically assigned by the system or customize one as needed, which is very flexible and convenient.

Speaker Identification

Before speaker identification, you need to provide a target speaker set that includes all users' voice characteristics sequences. It is a multiple choice question to identify the speaker by recognizing his/her voice from the speaker set.

Customer Service Robot

Customer Service Robot (CSR) is developed based on Tencent's leading natural language processing capabilities and huge enterprise knowledge base. It can respond to customer issues in multiple channels at any time and supports interaction with internal business systems to help enterprises lower labor costs and provide intelligent customer services.

Auto Robot Reply

CSR provides auto reply when the customer service is offline. The context-based multi-session capability supports auto reply on business information problems and guidance on business operation. API and SDK access methods are provided to achieve seamless connection to internal business systems, leading to effective integration of enterprise data and more intelligent services.

Online Manual Customer Service

The key business problems or those that cannot be replied by CSR are seamlessly conveyed to manual customer service in order to balance service efficiency and customer satisfaction. human customer service representatives are automatically allocated based on the business system, work load, service history and other rules.

Data Operation Analysis

It can provide real-time statistics and analysis of session volume, service quality, and key words for robots and customer service, as well as comprehensive analysis of customer satisfaction, emotion, hot issues to gain insight into customer service in real time.

Application Scenarios

Financial Industry

CSR can help enterprises effectively save more than 50% phone customer service staff. The multi-channel access capability can help quickly build all-media customer service. CSR can reduce costs while improving business service system.

E-commerce

The human plus robot customer service system allows e-commerce customers to easily connect to WeChat, APP, web pages and other service channels to reduce labor costs. The history retention mechanism can be used to effectively track user issues and enhance service experience.

Travel

CSR accurately covers the high frequency user consulting, and can access internal business systems to provide self-service interactive operation services, such as refund, ticket change, and query. CSR can greatly simplify the service process, provides quick and easy operation methods with uniform service standards, and effectively improves business efficiency.

Government Affairs

CSR keeps abreast of hot issues concerned by the masses and automatically replies to the known problems. It can also intelligently retrieve unsolved problems for government staff to research and discuss, and gives a standard reply via the system to maintain quick and accurate contact with the masses and enhances the service experience of government organs and enterprises.