



中华人民共和国国家标准

GB/T 34145—2017

中文语音合成互联网服务接口规范

Specification of programming interface for Chinese
speech synthesis internet service

2017-07-31 发布

2018-02-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会

发布

中 华 人 民 共 和 国
国 家 标 准
中文语音合成互联网服务接口规范
GB/T 34145—2017

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲2号(100029)
北京市西城区三里河北街16号(100045)

网址: www.spc.org.cn

服务热线: 400-168-0010

2017年8月第一版

*

书号: 155066·1-57053

版权专有 侵权必究

目 次

前言	I
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 语音合成服务系统能力范围	2
4.1 概述	2
4.2 基本功能	2
4.3 扩展功能	3
5 输入数据	3
5.1 合成文本	3
5.2 合成参数	3
6 输出数据	8
6.1 音频数据格式要求	8
6.2 合成音频描述信息	8
7 服务接口	12
7.1 初始化组件	12
7.2 创建语音合成会话	12
7.3 设置语音合成会话参数	12
7.4 写入合成文本	13
7.5 暂停语音合成会话	13
7.6 恢复语音合成会话	13
7.7 停止语音合成	13
7.8 获取合成音频	13
7.9 获取合成音频描述信息	13
7.10 结束语音合成会话	13
7.11 逆初始化组件	13
8 接口返回值要求	14
附录 A (资料性附录) 合成音频描述信息消息示例	15
A.1 不含标注信息的消息	15
A.2 含标注信息的消息	15
图 1 语音合成一般流程示意图	2

表 1	语音合成参数	4
表 2	合成音频数据格式	8
表 3	合成音频描述信息参数	8

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位:科大讯飞股份有限公司、中国电信集团公司、中国电子技术标准化研究院、北京慧听科技有限公司、苏州思必驰信息科技有限公司、阿里云计算有限公司、哈尔滨工业大学。

本标准主要起草人:李洪亮、严峻、于继栋、朱恩德、刘坤、张陈、胡修文、杨震、赵菁华、双志伟、周伟达。

中文语音合成互联网服务接口规范

1 范围

本标准规定了中文语音合成服务系统在互联网环境下提供服务的能力范围、输入数据、输出数据、服务接口、接口返回值要求等。

本标准适用于基于互联网的中文语音合成服务系统接口的设计、开发和应用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB 13000 信息技术 通用多八位编码字符集(UCS)(GB 13000—2010,ISO/IEC 10646:2003, IDT)

GB 18030 信息技术 中文编码字符集

GB/T 21024—2007 中文语音合成系统通用技术规范

汉语拼音方案(中华人民共和国第一届全国人民代表大会第五次会议决议通过)

RFC 2045 多用途互联网邮件扩展 第1部分:互联网消息体格式[Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies]

RFC 2586 多用途互联网邮件扩展内容类型(The Audio/L16 MIME content type)

RFC 5646 语言标识标签(Tags for Identifying Languages)

SSML 1.1 语音合成置标语言[Speech Synthesis Markup Language (SSML) Version 1.1]

国际语音学会手册:国际音标使用指南(Handbook of the International Phonetic Association: A Guide to the Use of International Phonetic Alphabet)

3 术语和定义

下列术语和定义适用于本文件。

3.1

语音合成 **speech synthesis**

将给定的文本转换成与之对应的语音的过程。

3.2

语音合成置标语言 **speech synthesis markup language;SSML**

一种丰富的、基于XML的语音合成置标语言,为的是在不同的平台之上提供标准的方式来控制语音合成的发音、音量、语速等。

3.3

中文语音合成置标语言 **Chinese speech synthesis markup language;CSSML**

一套基于XML的、用于中文语音合成的置标语言。

注:CSSML可被用于中文语音合成中的资源共享、模块互换,以及分布式应用中的中间数据格式,或者独立系统的输入控制。