
科大讯飞股份有限公司
IFLYTEK CO.,LTD.

语音评测参数设置、试题格式、
结果说明文档

目 录

1. 参数设置.....	1
2. 试题格式.....	1
2.1.汉语评测试题.....	1
2.2.英语评测试题.....	2
3. 评测结果.....	2
3.1.精简结果.....	2
3.2.详细结果.....	3
3.2.1.汉语评测.....	3
3.2.2.英语评测.....	5

1. 参数设置

在开始评测之前，需要设置评测语种、题型等参数，这些参数如下：

表 1 评测可设置参数

参数	说明	是否必需
language	评测语种，可选值： <code>en_us</code> （英语）、 <code>zh_cn</code> （汉语）。	是
category	评测题型，可选值： <code>read_syllable</code> （单字朗读，汉语专有）、 <code>read_word</code> （词语朗读）、 <code>read_sentence</code> （句子朗读）。	是
text_encoding	上传的试题编码格式，可选值： <code>gb2312</code> 、 <code>utf-8</code> 。当进行汉语评测时，必须设置成 <code>utf-8</code> ，建议所有试题都使用 <code>utf-8</code> 编码。	是
vad_bos	前端点超时，默认 5000ms。	否
vad_eos	后端点超时，默认 1800ms。	否
speech_timeout	录音超时，当录音达到时限将自动触发 vad 停止录音，默认-1（无超时）。	否
result_level	评测结果等级，可选值： <code>plain</code> （仅英语评测支持）、 <code>complete</code> ，默认为 <code>complete</code> 。	否

2. 试题格式

评测试题需要符合一定的格式，汉语试题和英语试题有所不同，同语种的不同题型也有差异。

2.1. 汉语评测试题

表 2 汉语评测试题格式

评测题型	格式	示例
read_syllable	支持两种试卷格式： 1) 拼音标注试卷。以 <code><customizer: interphonic></code> 开头，字单独一行，紧接着的一行为字的拼音（用拼音后加数字代表声调，1-4 分别代表一到四声，5 表示轻声）； 2) 纯文本试卷，每个字间用逗号隔开； 3) 最大长度 180 字节，超长则会报 10109 错误。	<code><customizer: interphonic></code> 丰 <code>feng1</code> 呈 <code>cheng2</code> 政 <code>zheng4</code> 或者 丰，呈，政

read_word	支持两种试卷格式： 1) 拼音标注试卷。与字的拼音标注试卷一样，只不过将字换成词语； 2) 纯文本试卷，每个词间用逗号隔开； 3) 最大长度 180 字节，超长则会报 10109 错误。	<customizer: interphonic> 宁可 ning4 ke3 非难 fei1 nan4 灾难 zai1 nan4 或者 宁可，非难，灾难
read_sentence	普通中文文本，最大长度为 180 字节，超长则会报 10109 错误。	这是中文语句评测示例。

2.2. 英语评测试题

表 3 英语评测试题格式

评测类型	格式	示例
read_word	一个单词占一行，首行用[word]标记，最大长度 180 字节，超长则会报 10109 错误。	[word] apple banana orange
read_sentence	普通英文文本，最大长度为 180 字节，超长则会报 10109 错误。	This is an example of sentence test.

3. 评测结果

SDK 返回 XML 格式的评测结果，可以分成两种：一种是十分精简的结果；另一种是详细结果。

3.1. 精简结果

在英语评测中，将 result_level 设置成 plain 将返回精简结果（注：汉语评测返回的都是详细结果），格式如下：

```
<?xml version="1.0" ?>  
<FinalResult>
```

```
<ret value="0"/>
<total_score value="4.496416"/>
</FinalResult>
```

结果中包含一个返回值 ret（0 表示评测正常）和总分 total_score（分值 0-5 分，满分为 5 分）。

3.2. 详细结果

当 result_level 设置为 complete 或不设置时，返回详细的 XML 评测结果，结果中包含的主要标签如下：

表 4 评测 xml 结果标签

标签	说明
rec_paper	<rec_paper>和</rec_paper>中间的内容即为评测结果，是开发者需要关心的部分。
read_syllable	表明这是单字朗读评测（汉语评测专有）结果。
read_word	表明这是单词朗读评测结果。
read_sentence	表明这是句子朗读评测结果。
sentence	句子。
word	单词，句子的组成部分。
syllable	音节，单词发音的组成部分。对于汉语，一个音节对应一个字的发音。
phone	音素，基本发音元素，音节的组成部分。

read_syllable、read_word 和 read_sentence 标签中的属性记录了本次评测的各项得分、整体时长等信息。sentence、word、syllable 和 phone 标签记录的是云端引擎对评测语音的分析结果，如哪些是静音，静音时长多少，哪些是有效发音，每个字、词的音节发音是否正确等等。

3.2.1. 汉语评测

以下为汉语评测结果的属性说明（注：一帧长度为 10ms，汉语评测得分为 5 分制）：

表 5 汉语评测结果属性说明

评测题型	层级	属性	含义
read_syllable	read_syllable	total_score	总分
		beg_pos	始/末位置（单位：帧）
		/end_pos	
		content	试卷内容
		time_len	朗读时长（单位：帧）
	sentence	time_len	朗读时长（单位：帧）

		beg_pos/end_pos	始末位置（单位：帧）
		content	试卷内容
	word	beg_pos/end_pos	始末位置（单位：帧）
		symbol	拼音：数字代表声调，5 和 5 以上表示轻声
		content	试卷内容
		time_len	时长（单位：帧）
	syll	beg_pos /end_pos	始/末位置（单位：帧）
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		symbol	拼音：数字代表声调，5 和 5 以上表示轻声
		content	试卷内容（sil、silv 表示静音，fil 表示噪音）
		time_len	时长（单位：帧）
	phone	beg_pos/ end_pos	始末位置（单位：帧）
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		content	试卷内容（sil、silv 表示静音，fil 表示噪音）
		time_len	时长（单位：帧）
read_word	read_word	total_score	总分
		beg_pos/ end_pos	始/末位置（单位：帧）
		content	试卷内容
		time_len	时长（单位：帧）
	sentence	time_len	时长（单位：帧）
		beg_pos/ end_pos	始末位置（单位：帧）
		content	试卷内容
	word	beg_pos/ end_pos	始末位置（单位：帧）
		symbol	拼音：数字代表声调，5 和 5 以上表示轻声
		content	试卷内容
		time_len	时长（单位：帧）
	syll	beg_pos/ end_pos	始/末位置（单位：帧）
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		symbol	拼音：数字代表声调，5 和 5 以上表示轻声
		content	试卷内容（sil、silv 表示静音，fil 表示噪音）
		time_len	时长（单位：帧）
	phone	beg_pos/ end_pos	始末位置（单位：帧）
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		content	试卷内容（sil、silv 表示静音，fil 表示噪音）
		time_len	时长（单位：帧）
read_sentence	read_sentence	total_score	总分
		beg_pos/ end_pos	始/末位置（单位：帧）

		content	试卷内容
		time_len	时长（单位：帧）
	sentence	total_score	总分
		beg_pos/ end_pos	始末位置（单位：帧）
		content	试卷内容
		time_len	时长（单位：帧）
	word	beg_pos/ end_pos	始末位置（单位：帧）
		symbol	拼音：数字代表声调，5 和 5 以上表示轻声
		content	试卷内容
		time_len	时长（单位：帧）
	syll	beg_pos/ end_pos	始/末位置（单位：帧）
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		symbol	拼音：数字代表声调，5 和 5 以上表示轻声
		content	试卷内容（sil、silv 表示静音，fil 表示噪音）
		time_len	时长（单位：帧）
	phone	beg_pos/ end_pos	始末位置（单位：帧）
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		content	试卷内容（sil、silv 表示静音，fil 表示噪音）
		time_len	时长（单位：帧）

dp_message 属性说明如下：

表 6 dp_message 属性说明

取值	含义	说明
0	正常	引擎认为该单元读了，但不一定朗读正确
16	漏读	该单元没有读
32	增读	该单元是多读的文本内的内容
64	回读	该单元是重复读的相邻文本的内容
128	替换	该单元读成文本内其他的内容

3.2.2. 英语评测

以下为英语评测结果的属性说明（注：一帧长度为 10ms，英语评测得分为 5 分制）：

表 7 英语评测结果属性说明

评测类型	层级	属性	含义
read_word	read_word	beg_pos	多个单词开始边界时间
		content	多个单词内容

		end_pos	多个单词结束边界时间
		except_info	异常信息，0 表示无异常，常见的异常代码见表 8 常见 except_info 列表
		is_rejected	是否被拒绝：true，false。为 true 表明引擎检测到乱读。注：只有当评测试题长度达到 5 个单词时才会进行乱读检测，否则该属性的取值总为 false
		total_score	多个单词总分的平均分
	sentence	beg_pos	句子开始边界时间
		content	句子内容
		end_pos	句子结束边界时间
		index	句子索引
	word	beg_pos	单词开始边界时间
		content	单词内容
		end_pos	单词结束边界时间
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		global_index	单词在全篇章索引
		index	单词在句子索引
		total_score	单词总分
	syll	beg_pos	音节开始边界时间
		content	音节内容（sil、silv 表示静音，fil 表示噪音）
		end_pos	音节结束边界时间
		serr_msg	音节错误信息
		syll_accent	音节重读标记
	phone	beg_pos	音素开始边界时间
		content	音素内容（sil、silv 表示静音，fil 表示噪音）
		end_pos	音素结束边界时间
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
read_sentence	read_chapter	beg_pos	篇章开始时间
		content	篇章内容
		end_pos	篇章结束时间
		total_score	篇章总分
		except_info	异常信息，0 表示无异常
		is_rejected	是否被拒绝：true，false。为 true 表明引擎检测到乱读。注：只有当评测试题长度达到 5 个单词时才会进行乱读检测，否则该属性的取值总为 false
		word_count	篇章全部单词数量
	sentence	beg_pos	句子开始边界时间
		content	句子内容
		end_pos	句子结束边界时间

		index	句子索引
		total_score	全部分数，结构体（隐藏）
		word_count	句子全部单词数量
	word	beg_pos	单词开始边界时间
		content	单词内容
		end_pos	单词结束边界时间
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）
		global_index	单词在全篇章索引
		index	单词在句子索引
		total_score	单词总分
	syll	beg_pos	音节开始边界时间
		content	音节内容（sil、silv 表示静音，fil 表示噪音）
		end_pos	音节结束边界时间
	phone	beg_pos	音素开始边界时间
		content	音素内容（sil、silv 表示静音，fil 表示噪音）
		end_pos	音素结束边界时间
		dp_message	增漏读信息：0（正常），16（漏读），32（增读），64（回读），128（替换）

常见的 except_info 代码有：

表 8 常见 except_info 列表

except_info	说明
0	无异常
28673	该评测语音为无语音或音量太小
28676	该评测语音为乱说类型
28680	该评测语音为信噪比太低
28690	该评测语音出现截幅

XML 结果中音素和音节的音标标记（即 syll 和 phone 标签中的 content 属性）并不是标准音标，以下是它们与标准音标的对应关系：

表 9 音标对照

音标标记	标准音标	音标标记	标准音标	音标标记	标准音标
aa	ɑ:	uw	u:	f	f
oo	ɔ	ch	tʃ	g	g
ae	æ	dh	ð	p	p
ah	ʌ	hh	h	r	r
ao	ɔ:	jh	dʒ	s	s
aw	aʊ	ng	ŋ	t	t
aw	ə	sh	ʃ	v	v
ay	aɪ	th	θ	w	w

eh	e	zh	ʒ	z	z
er	ə:	y	j	ar	eə
ey	eɪ	d	d	ir	iə
ih	ɪ	k	k	ur	ʊə
iy	i:	l	l	tr	tr
ow	əʊ	m	m	dr	dr
oy	ɔɪ	n	n	ts	ts
uh	ʊ	b	b	dz	dz