

腾讯云智能数智人 播报服务



腾讯云

【 版权声明 】

©2013–2026 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

文档目录

播报服务

视频生成服务 API 文档

视频生成服务 API 概述

音频制作接口

视频制作接口- PPT模式

视频制作接口-基础版

视频制作接口-进阶版

视频制作接口-视频免训练

视频制作接口-照片免训练

音视频制作进度查询接口

视频制作接口- PPT 模式进度查询接口

附录

附录一： 结果码字典

附录二： 回调请求体格式

接口对接常见问题

播报服务

视频生成服务 API 文档

视频生成服务 API 概述

最近更新时间：2024-11-13 14:38:22

本文档主要描述 腾讯云智能数智人 视频生成aPaas 平台的开放接口协议。

一、接口调用环境

正式环境：`https://gw.tv.s.qq.com`

二、权限申请及开通

视频生成 aPaas 接口使用权限可以通过 [资产领用](#) 页面领用，领用后即可开通15天试用权限。

三、接口调用方式

[数智人 aPaas 接口调用方式](#)

⚠ 注意：

如果通过腾讯云计费直接采购。appkey 将只能在选购的形象包内切换形象，不同形象包不通用。

四、接入流程示意图

资源查询逻辑示意图

个人资产管理API-查询主播



个人资产管理API-查询主播下的所有形象



个人资产管理API-查询音色列表



个人资产管理API-查询形象支持的动作列表



音频制作接口

最近更新时间：2025-12-10 14:56:32

接口说明

对输入文本进行试听，需要试听的音色可以通过接口 [查询声音资产信息](#) 查询。

调用协议

HTTPS + JSON

POST /v2/ivh/videomaker/broadcastservice/tts

Header Content-Type: application/json;charset=utf-8

请求参数

参数	类型	必须	说明
TimbreKey	string	否	音色 key，当 VirtualmanKey 为空时 TimbreKey 不能为空
VirtualmanKey	string	否	定义播报的角色、服装、姿态、分辨率等信息，参数为枚举值，当 TimbreKey 为空时 VirtualmanKey 不能为空，默认选取匹配该形象的第一个音色制作音频。
InputSsml	string	是	需要播报的文本内容，支持 SSML 标签，上限2万字（字数按 unicode 字符数计算）
Speed	float	是	语速（1.0为正常语速，范围[0.5-1.5]，值为0.5时播报语速最慢，值为1.5时播报语速最快）
AudioStorageS3Url	string	否	可传入含鉴权 S3 协议存储 URL，音频成品会上传至该 URL
SampleRate	int	否	采样率，支持24000（24k）和16000（16k），默认24000(24k)
Codec	string	否	音频格式，支持 mp3 和 wav，默认mp3
SentenceMaxWords	int	否	分句的字数上限，字数范围 [0,999]，传0或不传时默认值30
SentenceDisplayPunct	string	否	分句里需要展示的标点符号，其中，传入特殊字符"0"表示标点符号都不展示，特殊字符"1"（默认取值）表示标点符号全都展示，也可自定义

uation			义传入需要展示的标点符号
SentenceSplitPunctuation	string	否	需要断句的标点符号，默认取值。;?!……!?
Volume	int	否	音量大小，范围[0, 10]，对应音量大小。默认为0，代表正常音量，值越大音量越高。 说明： TimbreKey 在 male_1-20、female_1-23（即男声1-20、女声1-23）间的音色不支持音量调节。
EmotionCategory	string	否	控制合成音频的情感，仅支持多情感音色使用，可选值参考个人资产管理API 4.5音色列表接口。
EmotionIntensity	int	否	控制合成音频情感程度，取值范围为 [50,200]，只有 EmotionCategory 不为空时生效。
TimbreLanguage	string	否	音色语种，可选语种参考个人资产管理 API 分页查询音色列表 接口，多语种音色在合成时必须选择对应语种。

返回参数

参数	类型	必须	说明
TaskId	string	是	音频制作的任务 ID，携带 TaskId 访问<音视频制作进度查询接口>，可获得该音频的制作进度和下载地址。

请求示例

```

{
  "Header": {},
  "Payload": {
    "VirtualmanKey": "k3dda2d3b56f4ec89778f2e032c2e294",
    "InputSsml": "你好虚拟主播",
    "Speed": 1
  }
}

```

返回示例

成功

```
{
  "Header": {
    "Code": 0,
    "DialogID": "",
    "Message": "",
    "RequestID": "fde854eaa981c7f2f7285d1c7eca335b",
    "SessionID": "gzb7dec22117297528294581119"
  },
  "Payload": {
    "TaskId": "81883d47c6154edf8e276531f09227b6"
  }
}
```

失败

```
{
  "Header": {
    "RequestID": "d99cf740be48d7b0fb1eee4447b71781",
    "SessionID": "gz6240ef0317651792698007383",
    "DialogID": "",
    "Code": 100001,
    "Message": "InvalidParameter:请求参数错误: Speed value invalid,
ranges:0.5-1.5"
  }
}
```

视频制作接口- PPT模式

最近更新时间：2026-04-01 18:23:31

接口说明

基于 ppt 文件生成数智人讲解视频，当前一个账户仅允许同时提交一个 PPT 视频合成任务，任务最终会扣减相应数智人类型的播报小时包（制作失败不扣减小时包）。通过 [视频制作接口-PPT 模式进度查询接口](#) 最终返回成品视频。

调用协议

- HTTPS + JSON
- POST `/v2/ivh/mediaproduction/mediaproduction/videomakewithppt`
- Header `Content-Type: application/json;charset=utf-8`

请求参数

参数	类型	必须	说明
PPTFileUrl	string	是	PPT 文件 URL，支持格式：ppt、pptx、pdf。 大小要求：不超过100MB，页数要求：不超过50页。
GlobalConfig	object	否	定义全局默认参数，包含视频相关参数、音频相关参数等。参数优先级规则：PageConfig > GlobalConfig。即每页配置中显式指定的参数将覆盖全局默认值；未指定的参数自动继承全局配置。
GlobalConfig.AdaptationType	int	否	全局画面适配模式，当背景图和输出视频分辨率不一致时，提供两种适配模式：0，1；默认0。 以下是两种适配模式： <ul style="list-style-type: none">• 0：拉伸适配：强制将 PPT 内容拉伸至充满整个视频画面，可能导致文字、图片变形，适合对画面比例无严格要求的场景；• 1：等比缩放：按原始文件比例缩放内容，确保画面无变形，剩余区域以指定颜色背景填充
GlobalConfig.BackgroundColor	string	否	全局背景图填充颜色，当适配模式选择等比缩放时生效，默认值：#FFFFFF（白色）。 格式要求： <ul style="list-style-type: none">• 必须为合法的十六进制颜色代码

			<ul style="list-style-type: none"> 以#符号开头，后跟6位十六进制字符（0-9, A-F） 不区分大小写 不支持简写形式（如 #FFF）
GlobalConfig.ShowSubtitles	bool	否	<p>是否在视频内显示字幕（显示要求是文本驱动），默认不展示，展示会大大增加视频制作时间。</p> <div style="border: 1px solid #00aaff; padding: 5px;"> <p>说明： 需要注意当值为 true 时，如果某些音色（第三方音色）不支持字幕会导致任务失败。</p> </div>
GlobalConfig.LogoParams	Array of [LogoParam]	否	定义视频中 logo 相关参数。
GlobalConfig.SpeechParam	SpeechParam	否	定义全局音频的详细参数。
GlobalConfig.VideoParam	VideoParam	否	定义全局视频的详细参数。
PageConfigs	Array of [PageConfig]	是	<p>定义具体页 PPT 的参数，包含数字人形象、音频相关参数、视频相关参数。</p> <div style="border: 1px solid #00aaff; padding: 5px;"> <p>说明： PageConfigs 元素数量应不超过 PPT 页数，当 PageConfigs 元素数量为 N，并且少于 PPT 页数时，仅基于前 N 页 PPT 合成视频。</p> </div>

LogoParam

参数	类型	必须	说明
LogoFileUrl	stri	否	LOGO 图片文件下载路径，支持 jpg、png 格式

	ng		
PositionX	int	否	LOGO 图片左上角的 X 坐标（坐标范围和视频分辨率相关）
PositionY	int	否	LOGO 图片左上角的 Y 坐标（坐标范围和视频分辨率相关）
Scale	float	否	LOGO 图片缩放比例（1.0为原始图片大小，范围(0,15]）

PageConfig

参数	类型	必须	说明
VirtualmanKey	string	否	<p>数字人形象，包含了播报的角色、服装、姿态、分辨率等信息。要求每页配置的数智人形象为同一类型，并且分辨率相同。</p> <div style="border: 1px solid #ccc; padding: 10px; margin-top: 10px;"> <p>ⓘ 说明：</p> <ul style="list-style-type: none"> 当前仅支持1080P 横版（1920×1080）和 1080P 竖版（1080×1920）的形象。 如果不需要数智人出镜（只要 PPT 画面 + 语音旁白），该字段可为空。 可以通过 查询形象资产信息-查询主播下的所有形象接口查询。 </div>
InputAudioUrl	string	否	<p>驱动数智人的音频 URL。当 DriverType 为 OriginalVoice 时，该字段必填。</p> <p>音频格式要求：音频时长不超过60分钟（VirtualmanKey 为空时不超过30分钟），不少于0.5秒。</p> <p>支持格式：wav、mp3、wma、m4a、aac、ogg。</p> <p>建议文件大小：不大于100MB。</p>
DriverType	string	否	<p>驱动类型，为空时默认 Text。</p> <p>Text：文本驱动，要求 InputSsml 字段必填。</p> <p>OriginalVoice：原声音频驱动，要求 InputAudioUrl 字段必填。可通过 Speech.TimbreKey 指定音色，未填写时使用主播默认音色。</p>
InputSsml	string	否	<p>播报的文本内容，支持 SSML 标签，支持的标签类型参照 数智人 SSML 标记语言规范，标签写法参照示例，内容不能换行，符号需转义。上限2万字（字数按 unicode 字符数计算）。</p> <p>DriverType 为空、或 Text 时，该字段必填。</p>

SpeechParam	SpeechParam	否	定义当前页面的音频详细参数。
VideoParam	VideoParam	否	定义当前页面的视频详细参数。

SpeechParam

参数	类型	必须	说明
Speed	float	否	语速（1.0为正常语速，范围[0.5-1.5]，值为0.5时播报语速最慢，值为1.5时播报语速最快，DriverType 为音频驱动类型时，语速控制不生效）。
TimbreKey	string	否	音色 key，默认使用形象自有音色。
Volume	int	否	音量大小，范围[0, 10]，对应音量大小。默认为0，代表正常音量，值越大音量越高。 <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;"> <p>说明： TimbreKey 在 male_1 - 20、female_1 - 23（即男声1 - 20、女声1 - 23）间的音色不支持音量调节。</p> </div>
EmotionCategory	string	否	控制合成音频的情感，仅支持多情感音色使用，可选值参考个人资产管理 API 分页查询音色列表 接口。
EmotionIntensity	int	否	控制合成音频情感程度，取值范围为 [50,200]，只有 EmotionCategory 不为空时生效。
TimbreLanguage	string	否	音色语种，可选语种参考个人资产管理 API 分页查询音色列表 接口，多语种音色在合成时必须选择对应语种。

VideoParam

参数	类型	必须	说明
----	----	----	----

AnchorParam	AnchorParam	否	定义视频中主播相关参数
SmallSampleParam	SmallSampleParam	否	定义小样本相关特殊参数，当为非小样本形象时，此参数不生效。
SmartActionEnabled	bool	否	是否开启智能动作，默认不开启。生效条件： DriverType=Text、且 InputSsmI 中不携带动作标签。

VideoParam.AnchorParam

参数	类型	必须	说明
HorizontalPosition	float	否	定义主播横向位置（0为中间位置）。 <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;"> <p>说明： 具体计算方法请参见 主播位置计算方法。</p> </div>
VerticalPosition	float	否	定义主播纵向位置（0为中间位置）。 <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;"> <p>说明： 具体计算方法请参见 主播位置计算方法。</p> </div>
Scale	float	否	主播大小（1是默认大小，范围(0,20]）
Angle	int	否	主播角度（默认0度，范围[0,360]）目前仅3D 形象支持。
AnchorExtraParam	string	否	主播额外可配置参数，目前仅包含 3D 服装变色参数。不同主播可配置的参数不同，详情见接口 查询形象资产信息-查询主播下的所有形象 的 SupportAnchorExtraParam 参数。 需要将该参数组织为 JSON 字符串格式，请参考请求示例组织 JSON。

VideoParam.SmallSampleParam

参数	类型	必须	说明

MakeType	string	否	<p>定义小样本制作类型：</p> <ul style="list-style-type: none"> • Default：默认配置，会从随机起始帧开始制作，StartIdx 和 EndIdx 均不生效。 • Custom：指定视频片段，可以填写 StartIdx 和 EndIdx 来选择片段，默认生成的视频用该“指定的视频片段”进行正倒放循环。 • Circle：首尾帧归位，可以填写 StartIdx 来指定视频开始帧数(EndIdx 不生效)，注意：该类型只用于声音驱动。
StartIdx	int	否	<p>起始帧号，当 MakeType 为 Custom 或 Circle 时生效，当填写的时候，默认生成的视频从该位置开始；不填的时候，默认选取开始帧号，即第0帧。</p>
EndIdx	int	否	<p>结束帧号，当 MakeType 为 Custom 时生效，当填写的时候，默认结束的视频到该位置结束；不填的时候，默认选取该“选中的全部视频片段”的结束帧号。</p>

VideoParam.AnchorParam.AnchorExtraParam 示例

说明：

colorValue 转换规则：十六进制 Hex 色值，去掉 # 后转为十进制，例如：#FAFAFA =>16448250

```
{
  "shirtColor": {
    "colorValue": 10491928
  },
  "clotheColor": {
    "colorValue": 10491928
  },
  "shoeColor": {
    "colorValue": 10491928
  }
}
```

返回参数

参数	类型	必须	说明
TaskId	string	是	<p>视频制作的任务 ID，携带 TaskId 访问 PPT 合成视频进度查询接口，可获得制作进度和制作结果</p>

请求示例

```
{
  "Header": {},
  "Payload": {
    "PPTFileUrl": "https://virtualhuman-cos-prod-1251316161.cos.ap-
nanjing.myqcloud.com/12323/%E5%AE%9E%E4%B9%A0%E7%AD%94%E8%BE%A92.pptx",
    "GlobalConfig": {
      "AdaptationType": 0,
      "BackgroundColor": "",
      "SpeechParam": {
        "EmotionCategory": "",
        "EmotionIntensity": 0,
        "Speed": 0,
        "TimbreKey": "tts_female_yunxiaoxiao",
        "TimbreLanguage": "",
        "Volume": 0
      },
    },
    "VideoParam": {
      "AnchorParam": {
        "AnchorExtraParam": "",
        "Angle": 0,
        "HorizontalPosition": 0,
        "Scale": 0,
        "VerticalPosition": 0
      },
      "SmallSampleParam": {
        "EndIdx": 0,
        "MakeType": "",
        "StartIdx": 0
      },
      "SmartActionEnabled": true
    },
    "ShowSubtitles": true,
    "LogoParams": [
      {
        "LogoFileUrl": "https://virtualhuman-cos-prod-
1251316161.cos.ap-nanjing.myqcloud.com/123/123.png",
```

```
    "PositionX": 0,
    "PositionY": 0,
    "Scale": 0.1
  }
]
},
"PageConfigs": [
  {
    "DriverType": "",
    "InputAudioUrl": "",
    "InputSsml": "你好呀，我是数智人，这是我的第一次ppt视频",
    "SpeechParam": {
      "EmotionCategory": "",
      "EmotionIntensity": 0,
      "Speed": 1,
      "TimbreKey": "tts_female_yunxiaoxiao",
      "TimbreLanguage": "",
      "Volume": 0
    },
    "VideoParam": {
      "AnchorParam": {
        "AnchorExtraParam": "",
        "Angle": 0,
        "HorizontalPosition": 0,
        "Scale": 0,
        "VerticalPosition": 0
      },
      "SmallSampleParam": {
        "EndIdx": 0,
        "MakeType": "",
        "StartIdx": 0
      },
      "SmartActionEnabled": false
    },
    "VirtualmanKey": "2c3f4c11d1e54b7693156b060990544d"
  },
  {
    "DriverType": "",
    "InputAudioUrl": "",
    "InputSsml": "你好呀，我是数智人，这是我的第一次ppt视频",
```

```
"SpeechParam": {
  "EmotionCategory": "",
  "EmotionIntensity": 0,
  "Speed": 1,
  "TimbreKey": "tts_female_yunxiaoxiao",
  "TimbreLanguage": "",
  "Volume": 0
},
"VideoParam": {
  "AnchorParam": {
    "AnchorExtraParam": "",
    "Angle": 0,
    "HorizontalPosition": 0,
    "Scale": 0,
    "VerticalPosition": 0
  },
  "SmallSampleParam": {
    "EndIdx": 0,
    "MakeType": "",
    "StartIdx": 0
  },
  "SmartActionEnabled": false
},
"VirtualmanKey": ""
},
{
  "DriverType": "",
  "InputAudioUrl": "",
  "InputSsml": "你好呀，我是数智人，这是我的第一次ppt视频",
  "SpeechParam": {
    "EmotionCategory": "",
    "EmotionIntensity": 0,
    "Speed": 1,
    "TimbreKey": "tts_female_yunxiaoxiao",
    "TimbreLanguage": "",
    "Volume": 0
  },
  "VideoParam": {
    "AnchorParam": {
      "AnchorExtraParam": "",
```

```
        "Angle": 0,  
        "Scale": 0,  
        "VerticalPosition": 0  
    },  
    "SmallSampleParam": {  
        "EndIdx": 0,  
        "MakeType": "",  
        "StartIdx": 0  
    },  
    "SmartActionEnabled": false  
},  
"VirtualmanKey": "2c3f4c11d1e54b7693156b060990544d"  
}  
]  
}  
}
```

返回示例

成功

```
{  
  "Header": {  
    "RequestID": "5e2dd62fa7f3c1aa6030875637688706",  
    "SessionID": "gzfa9105ca17749284691612885",  
    "DialogID": "",  
    "Code": 0,  
    "Message": ""  
  },  
  "Payload": {  
    "TaskId": "098b482c71f643ad8f288df9f2961403"  
  }  
}
```

失败

```
{  
  "Header": {
```

```
"RequestID": "d99cf740be48d7b0fb1eee4447b71781",
"SessionID": "gz6240ef0317651792698007383",
"DialogID": "",
"Code": 100001,
"Message": "InvalidParameter:请求参数错误: Speed value invalid,
ranges:0.5-1.5"
}
}
```

视频制作接口-基础版

最近更新时间：2025-12-10 14:56:32

接口说明

使用 ssml 文本和数智人进行视频制作，通过 [音视频制作进度查询接口](#) 最终返回成品视频和字幕文件。

⚠ 注意：

不支持定义主播位置等未剪辑高级参数，如需使用需要切换到 [视频制作接口-进阶版](#) 接口。

调用协议

HTTPS + JSON

POST /v2/ivh/videomaker/broadcastservice/videomake

Header Content-Type: application/json;charset=utf-8

请求参数

参数	类型	必须	说明
Virtualman Key	string	是	定义播报的角色、服装、姿态、分辨率等信息，参数为枚举值。 说明： 可以通过 查询形象资产信息-查询主播下的所有形象 接口查询。
InputSsml	string	是	播报的文本内容，支持ssml标签，支持的标签类型参照 数智人 SSML 标记语言规范 ，标签写法参照示例，内容不能换行，符号需转义。上限2万字（字数按 unicode 字符数计算）。DriverType为空、或 Text 时，该字段必填 说明： 当 Format 为 TransparentMov 时，上限为2千字。
SpeechParam	object	是	定义音频的详细参数

SpeechParam.Speed	float	是	语速（1.0为正常语速，范围[0.5-1.5]，值为0.5时播报语速最慢，值为1.5时播报语速最快，DriverType为音频驱动类型时，语速控制不生效）
SpeechParam.TimbreKey	string	否	音色key，默认使用形象自有音色
SpeechParam.Volume	int	否	音量大小，范围[0, 10]，对应音量大小。默认为0，代表正常音量，值越大音量越高。 <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;"> <p>说明： TimbreKey 在 male_1-20、female_1-23（即男声1-20、女声1-23）间的音色不支持音量调节。</p> </div>
SpeechParam.EmotionCategory	string	否	控制合成音频的情感，仅支持多情感音色使用，可选值参考个人资产管理 API 分页查询音色列表 接口。
SpeechParam.EmotionIntensity	int	否	控制合成音频情感程度，取值范围为 [50,200]，只有 EmotionCategory 不为空时生效。
SpeechParam.TimbreLanguage	string	否	音色语种，可选语种参考个人资产管理 API 分页查询音色列表 接口，多语种音色在合成时必须选择对应语种。
VideoParam	object	否	定义合成视频的详细参数。
VideoParam.Format	string	否	视频输出格式，默认 TransparentWebm TransparentWebm：带 Alpha 通道的 webm 格式视频 TransparentMov：带 Alpha 通道的 mov 格式视频 Mp4：绿幕 mp4 格式视频；非抠像形象为原始素材背景
CallbackUrl	string	否	当用户增加回调 URL 时，将把视频制作结果以固定格式发送 POST 请求到该 url 地址，固定格式见 附录二：回调请求体格式 ，需注意： 1、限制 CallbackUrl 长度小于1000 2、只发送一次请求，无论是哪种问题导致的请求失败，都不会再进行发送。
DriverType	string	否	驱动类型，默认Text 1. Text：文本驱动，要求 InputSsml 字段必填 2. OriginalVoice：原声音频驱动，要求 InputAudioUrl 字段必填

			3. ModulatedVoice: 变声音频驱动, 可通过 Speech.TimbreKey 指定音色, 未填写时使用主播默认音色
InputAudioUrl	string	否	<p>驱动数智人的音频 URL, 当 DriverType 为 OriginalVoice、ModulatedVoice 时, 该字段必填。</p> <p>音频格式要求:</p> <ol style="list-style-type: none"> 1. 当输出视频格式为: TransparentWebm、Mp4时, 音频时长不超过60分钟, 不少于0.5秒 2. 支持格式: wav、mp3、wma、m4a、aac、ogg 3. 建议文件大小: 不大于100M <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;"> <p>说明: 当 Format 为 TransparentMov 时, 时长不超过6分钟, 不少于0.5秒。</p> </div>
VideoStorageS3Url	string	否	可传入含鉴权 S3 协议存储 URL, 视频成品会上传至该 URL
SubtitleStorageS3Url	string	否	可传入含鉴权 S3 协议存储 URL, 字幕成品会上传至该 URL
ConcurrencyType	string	否	<p>视频制作任务使用的并发类型, 默认采用优先使用独享, 其次使用共享的策略</p> <ol style="list-style-type: none"> 1. Exclusive: 独享并发, 若无独享并发, 任务提交失败 2. Shared: 共享并发

返回参数

参数	类型	必须	说明
TaskId	string	是	视频制作的任务 ID, 携带 TaskId 访问 音视频制作进度查询接口 , 可获得制作进度和制作结果

请求示例

文本驱动

```

{
  "Header": {},
  "Payload": {
```

```
"VirtualmanKey": "k3dda2d3b56f4ec89778f2e032c2e294",
  "InputSsml": "你好，我是虚拟<phoneme alphabet=\"py\" ph=\"fu4\">主
</phoneme>播",
  "SpeechParam": {
    "Speed": 1.0
  },
  "VideoParam": {
    "Format": "TransparentWebm"
  },
  "DriverType": "Text"
}
```

音频驱动

```
{
  "Header": {},
  "Payload": {
    "VirtualmanKey": "k3dda2d3b56f4ec89778f2e032c2e294",
    "InputAudioUrl": "https://videomaker-
resources.ivh.qq.com/5da17352aa1e4cba9cada7bd20571295.mp3",
    "SpeechParam": {
      "Speed": 1.0
    },
    "VideoParam": {
      "Format": "TransparentWebm"
    },
    "DriverType": "OriginalVoice"
  }
}
```

返回示例

成功

```
{
  "Header": {
    "Code": 0,
    "DialogID": "",

```

```
    "Message": "",
    "RequestID": "fde854eaa981c7f2f7285d1c7eca335b",
    "SessionID": "gzb7dec22117297528294581119"
  },
  "Payload": {
    "TaskId": "81883d47c6154edf8e276531f09227b6"
  }
}
```

失败

```
{
  "Header": {
    "RequestID": "d99cf740be48d7b0fb1eee4447b71781",
    "SessionID": "gz6240ef0317651792698007383",
    "DialogID": "",
    "Code": 100001,
    "Message": "InvalidParameter:请求参数错误: Speed value invalid,
ranges:0.5-1.5"
  }
}
```

视频制作接口-进阶版

最近更新时间：2026-04-01 18:23:31

接口说明

使用 SSML 文本和数智人进行视频制作，通过 [音视频制作进度查询接口](#) 最终返回成品视频和字幕文件。在原有接口基础上、进阶版接口新增了部分资源参数，扩展了微剪辑能力，支持能力如下表：

LOGO	支持调整 LOGO 位置（自定义 X 轴 Y 轴）
	支持调整 LOGO 大小（缩放比例）
指定视频背景	在 MP4 视频文件中，可以指定背景图。
主播	横向位置调节
	主播大小调整
	主播角度调整
<div style="border: 1px solid #00aaff; padding: 5px;">⚠ 注意： 仅支持 3D 形象</div>	
片头/片尾视频	可指定远程地址添加视频片头片尾
内嵌字幕	打开、关闭内嵌字幕

调用协议

- HTTPS + JSON
- POST /v2/ivh/videomaker/broadcastservice/videomakeadvanced
- Header Content-Type: application/json;charset=utf-8

请求参数

参数	类型	必须	说明
VirtualmanKey	string	是	定义播报的角色、服装、姿态、分辨率等信息，参数为枚举值。 <div style="border: 1px solid #00aaff; padding: 5px;">ⓘ 说明：</div>

			<p>可以通过 查询形象资产信息-查询主播下的所有形象 接口查询。</p>
InputSsml	string	否	<p>播报的文本内容，支持 SSML 标签，支持的标签类型参照 数智人 SSML 标记语言规范，标签写法参照示例，内容不能换行，符号需转义。上限2万字（字数按 unicode 字符数计算）。DriverType 为空、或 Text 时，该字段必填</p> <p>说明： 当 Format 为 TransparentMov 时，上限为2千字。</p>
SpeechParam	object	是	定义音频的详细参数。
SpeechParam.Speed	float	是	语速（1.0为正常语速，范围[0.5-1.5]，值为0.5时播报语速最慢，值为1.5时播报语速最快，DriverType 为音频驱动类型时，语速控制不生效）。
SpeechParam.TimbreKey	string	否	音色 key，默认使用形象自有音色。
SpeechParam.Volume	int	否	<p>音量大小，范围[0, 10]，对应音量大小。默认为0，代表正常音量，值越大音量越高。</p> <p>说明： TimbreKey 在 male_1 - 20、female_1 - 23（即男声1 - 20、女声1 - 23）间的音色不支持音量调节。</p>
SpeechParam.EmotionCategory	string	否	控制合成音频的情感，仅支持多情感音色使用，可选值参考个人资产管理 API 分页查询音色列表 接口。
SpeechParam.EmotionIntensity	int	否	控制合成音频情感程度，取值范围为 [50,200]，只有 EmotionCategory 不为空时生效。
SpeechParam.TimbreLanguage	string	否	音色语种，可选语种参考个人资产管理 API 分页查询音色列表 接口，多语种音色在合成时必须选择对应语种。

VideoParam	object	否	定义合成视频的详细参数。
VideoParam.Format	string	否	视频输出格式，默认值：Mp4 <ul style="list-style-type: none"> • TransparentWebm: 带 Alpha 通道的 WebM 格式视频，支持部分微剪辑能力（主播参数支持）。 • TransparentMov: 带 Alpha 通道的 MOV 格式视频，支持部分微剪辑能力（主播参数支持）。 • Mp4: 支持微剪辑能力的 MP4 格式视频，抠像形象默认为绿幕背景，可通过 BackgroundFileUrl 参数设置背景；非抠像形象为原始素材背景。
VideoParam.BackgroundFileUrl	string	否	视频背景图片/视频下载路径，支持 JPG、PNG、MP4 格式，图片/视频分辨率需要与视频分辨率相同，不传默认为绿幕视频。文件大小限制500MB。
VideoParam.VideoHeadFileUrl	string	否	片头视频，支持 MP4 格式，分辨率需要与视频分辨率相同，文件大小限制500MB。
VideoParam.VideoTailFileUrl	string	否	片尾视频，支持 MP4 格式，分辨率需要与视频分辨率相同，文件大小限制500MB。
VideoParam.ShowSubtitles	boolean	否	是否在视频内显示字幕，默认不展示，展示会大大增加视频制作时间。
VideoParam.SubtitlesParam	object	否	定义视频中字幕如何显示的相关参数。
VideoParam.SubtitlesParam.MaxWords	int	否	字幕每页展示的字数上限，字数范围 [0,999]；默认值为0，默认展示规则：在分辨率80%宽度范围内展示，超出时分页。
VideoParam.SubtitlesParam.DisplayPunctuation	string	否	字幕需要展示的标点符号，其中，传入特殊字符"0"表示标点符号都不展示，特殊字符"1"（默认取值）表示标点符号全都展示，也可自定义传入需要展示的标点符号。
VideoParam.SubtitlesParam.SplitPunctuation	string	否	字幕需要分页的标点符号，默认取值：；？！………！？
VideoParam.LogoParams	Array of [LogoParam]	否	定义视频中 LOGO 相关参数。

VideoParam.SmartActionEnabled	bool	否	是否开启智能动作，默认不开启。生效条件：DriverType=Text、且 InputSsmI 中不携带动作标签。
VideoParam.ActionParams	Array of [ActionParam]	否	定义音频驱动模式下数字人主播的指定动作序列。每个动作将在对应时间点触发，使主播执行相应行为。 生效条件： 1. 所选动作（Code）必须已预置在该主播的可用动作库中，主播支持的动作可查询： 查询形象支持的动作列表 ； 2. 驱动模式（DriverType）需设置为 OriginalVoice； 3. 动作时间点（StartTime）必须严格递增且位于音频总时长范围内。
VideoParam.AnchorParam	object	否	定义视频中主播相关参数。
VideoParam.AnchorParam.HorizontalPosition	float	否	定义主播横向位置（0为中间位置）。 <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;">说明： 具体计算方法请参见 主播位置计算方法。</div>
VideoParam.AnchorParam.VerticalPosition	float	否	定义主播纵向位置（0为中间位置） <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;">说明： 具体计算方法请参见 主播位置计算方法。</div>
VideoParam.AnchorParam.Scale	float	否	主播大小（1是默认大小，范围(0,20]）
VideoParam.AnchorParam.Angle	int	否	主播角度（默认0度，范围[0,360]），目前仅 3D 形象支持。
VideoParam.AnchorParam.AnchorExtraParam	string	否	主播额外可配置参数，目前仅包含 3D 服装变色参数。不同主播可配置的参数不同，详情见接口 查询形象资产信息-查询主播下的所有形象 的 SupportAnchorExtraParam 参数。需要将该参数组织为 JSON 字符串格式，请参考请求示例组织 JSON。
VideoParam.SmallS	object	否	定义小样本相关特殊参数，当为非小样本形象时，此

ampleParam			参数不生效。
VideoParam.SmallSampleParam.MakeType	string	否	定义小样本制作类型： <ul style="list-style-type: none"> • Default: 默认配置，会从随机起始帧开始制作，StartIdx 和 EndIdx 均不生效。 • Custom: 指定视频片段，可以填写 StartIdx 和 EndIdx 来选择片段，默认生成的视频用该“指定的视频片段”进行正倒放循环。 • Circle: 首尾帧归位，可以填写 StartIdx 来指定视频开始帧数（EndIdx 不生效），注意：该类型只用于声音驱动。 • CustomOnlyStart: 通过 StartIdx 指定视频起始帧，EndIdx 不生效。
VideoParam.SmallSampleParam.StartIdx	int	否	起始帧号，当 MakeType 为 Custom 或 Circle 时生效，当填写的时候，默认生成的视频从该位置开始；不填的时候，默认选取开始帧号，即第0帧。
VideoParam.SmallSampleParam.EndIdx	int	否	结束帧号，当 MakeType 为 Custom 时生效，当填写的时候，默认结束的视频到该位置结束；不填的时候，默认选取该“选中的全部视频片段”的结束帧号。
CallbackUrl	string	否	当用户增加回调 URL 时，将把视频制作结果以固定格式发送 POST 请求到该 URL 地址，固定格式见 附录二：回调请求体格式 。 需注意： <ol style="list-style-type: none"> 1. 限制 CallbackUrl 长度小于1000。 2. 只发送一次请求，无论是哪种问题导致的请求失败，都不会再进行发送。
DriverType	string	否	驱动类型，默认 Text。 <ol style="list-style-type: none"> 1. Text: 文本驱动，要求 InputSsml 字段必填。 2. OriginalVoice: 原声音频驱动，要求 InputAudioUrl 字段必填。 3. ModulatedVoice: 变声音频驱动，可通过 Speech.TimbreKey 指定音色，未填写时使用主播默认音色。
InputAudioUrl	string	否	驱动数智人的音频 URL，当 DriverType 为 OriginalVoice、ModulatedVoice 时，该字段必填。音频格式要求： <ol style="list-style-type: none"> 1. 当输出视频格式为：TransparentWebm、Mp4 时，音频时长不超过60分钟，不少于0.5秒。

			<p>2. 支持格式: wav、mp3、wma、m4a、aac、ogg。</p> <p>3. 建议文件大小: 不大于100MB。</p> <div style="border: 1px solid #00aaff; padding: 5px; margin-top: 10px;"> <p>说明: 当 Format 为 TransparentMov 时, 时长不超过6分钟, 不少于0.5秒。</p> </div>
VideoStorageS3Url	string	否	可传入含鉴权 S3 协议存储 URL, 视频成品会上传至该 URL。
SubtitleStorageS3Url	string	否	可传入含鉴权 S3 协议存储 URL, 字幕成品会上传至该 URL。
ConcurrencyType	string	否	<p>视频制作任务使用的并发类型, 默认采用优先使用独享, 其次使用共享的策略</p> <p>1. Exclusive: 独享并发, 若无独享并发, 任务提交失败。</p> <p>2. Shared: 共享并发。</p>

LogoParam

参数	类型	必须	说明
LogoFileUrl	string	否	LOGO 图片文件下载路径, 支持 jpg、png 格式
PositionX	int	否	LOGO 图片左上角的 X 坐标 (坐标范围和视频分辨率相关)
PositionY	int	否	LOGO 图片左上角的 Y 坐标 (坐标范围和视频分辨率相关)
Scale	float	否	LOGO 图片缩放比例 (1.0为原始图片大小, 范围(0,15])

ActionParam

参数	类型	必须	说明
Code	string	否	动作 code, 必须属于相应主播。

StartTime	int	否	动作执行的开始时间，单位是 ms，必须在音频时长范围内，且需要严格递增。需要注意动作执行需要一定时长，如果动作之间执行存在冲突（时间冲突）可能会造成部分动作失效。
-----------	-----	---	---

VideoParam.AnchorParam.AnchorExtraParam 示例

说明：

colorValue 转换规则：十六进制 Hex 色值，去掉#后转为十进制，例如：#FAFAFA =>16448250

```
{
  "shirtColor": {
    "colorValue": 10491928
  },
  "clotheColor": {
    "colorValue": 10491928
  },
  "shoeColor": {
    "colorValue": 10491928
  }
}
```

返回参数

参数	类型	必须	说明
TaskId	string	是	视频制作的任务 ID，携带 TaskId 访问 音视频制作进度查询接口 ，可获得制作进度和制作结果

请求示例

文本驱动

```
{
  "Header": {},
  "Payload": {
    "VirtualmanKey": "k3dda2d3b56f4ec89778f2e032c2e294",
    "InputSsml": "你好，我是虚拟<phoneme alphabet=\"py\" ph=\"fu4\">主</phoneme>播",
    "SpeechParam": {
```

```
    "Speed": 1
  },
  "DriverType": "Text",
  "VideoParam": {
    "Format": "Mp4",
    "BackgroundFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/bg.png",
    "VideoHeadFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/video_head.mp4",
    "VideoTailFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/video_tail.mp4",
    "ShowSubtitles": true,
    "LogoParams": [
      {
        "LogoFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/logo.png",
        "PositionX": 1561,
        "PositionY": 751,
        "Scale": 1.0
      }
    ],
    "AnchorParam": {
      "HorizontalPosition": 0.5,
      "Scale": 1.0,
      "AnchorExtraParam": "{\"hair\": {\"boolValue\":true}, \"Jacket\": {\"boolValue\":true}, \"Shoes\": {\"boolValue\":true}, \"Skirt\": {\"boolValue\":true}, \"hairColor\": {\"colorValue\":12759296}, \"JacketColor\": {\"colorValue\":7976228}, \"ShoesColor\": {\"colorValue\":16777215}, \"SkirtColor\": {\"colorValue\":16777215}}}"
    }
  }
}
```

音频驱动

```
{
  "Header": {},
}
```

```
"Payload": {
  "VirtualmanKey": "k3dda2d3b56f4ec89778f2e032c2e294",
  "InputAudioUrl": "https://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/resource-test/shanjian/audio.wav",
  "DriverType": "OriginalVoice",
  "VideoParam": {
    "Format": "Mp4",
    "BackgroundFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/bg.png",
    "VideoHeadFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/video_head.mp4",
    "VideoTailFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/video_tail.mp4",
    "ShowSubtitles": true,
    "LogoParams": [
      {
        "LogoFileUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/logo.png",
        "PositionX": 1561,
        "PositionY": 751,
        "Scale": 1.0
      }
    ],
    "ActionParams": [
      {
        "Code": "action1",
        "StartTime": 5000
      },
      {
        "Code": "action2",
        "StartTime": 22000
      }
    ],
    "AnchorParam": {
      "HorizontalPosition": 0.5,
      "Scale": 1.0,
      "AnchorExtraParam": "{\"hair\": {\"boolValue\":true}, \"Jacket\": {\"boolValue\":true}, \"Shoes\": {\"boolValue\":true}, \"Skirt\": {\"boolValue\":true}, \"hairColor\":
```

```
{\"colorValue\":12759296},\"JacketColor\":  
{\"colorValue\":7976228},\"ShoesColor\":  
{\"colorValue\":16777215},\"SkirtColor\":{\"colorValue\":16777215}}"  
  }  
}  
}
```

返回示例

成功

```
{  
  "Header": {  
    "Code": 0,  
    "DialogID": "",  
    "Message": "",  
    "RequestID": "fde854eaa981c7f2f7285d1c7eca335b",  
    "SessionID": "gzb7dec22117297528294581119"  
  },  
  "Payload": {  
    "TaskId": "81883d47c6154edf8e276531f09227b6"  
  }  
}
```

失败

```
{  
  "Header": {  
    "RequestID": "d99cf740be48d7b0fb1eee4447b71781",  
    "SessionID": "gz6240ef0317651792698007383",  
    "DialogID": "",  
    "Code": 100001,  
    "Message": "InvalidParameter:请求参数错误: Speed value invalid,  
ranges:0.5-1.5"  
  }  
}
```

视频制作接口-视频免训练

最近更新时间：2026-04-01 14:19:50

接口说明

无需训练，即可基于一段真人视频素材，通过输入文本或音频，生成人物说话口型与输入内容相匹配的新视频。通过 [音视频制作进度查询接口](#) 最终返回成品视频。目前音视频资源只保留7天，请尽快下载。

调用协议

- HTTPS + JSON
- POST /v2/ivh/videomaker/broadcastservice/videomakenotrain
- Header Content-Type: application/json;charset=utf-8

请求参数

参数	类型	必须	说明
RefVideoUrl	string	是	<p>模板视频 URL，该字段必填。</p> <ol style="list-style-type: none">支持封装格式：mp4、mov、avi <div style="border: 1px solid #00a0e3; padding: 10px; margin: 10px 0;"><p>⚠ 注意： 视频本身要求编码格式为 H.264 格式，若不是，可通过下述方法转换。请参见 编码格式转换。</p></div> <ol style="list-style-type: none">文件大小：不超过5G。视频分辨率：单边像素要求在360 - 4096之间。视频时长：时长不超过20分钟，不少于1秒。请保证文件的下载速度，否则会下载失败或导致视频制作实时率下降。 <p>对于输入视频素材的规范性要求：</p> <ol style="list-style-type: none">人脸画面：要求真人出镜（如果是卡通人物，需要人物五官和真人比例相近），画面中的人脸说话时，建议正对镜头，水平转动不超过45度，俯仰不超过15度；人脸尽量不遮挡，面部光线稳定。说话音频：对音频无限制。 <p>对于输出视频的说明：</p> <ol style="list-style-type: none">输出格式：mp4输出分辨率：最大支持输出单边4096，即 4K 视频。<ol style="list-style-type: none">2.1 当 ConcurrencyType 为 Shared 时，支持最大输出 2560*1440。当输入视频分辨率<=2560*1440时，输出分辨

			<p>率和输入视频一致。</p> <p>2.2 当 ConcurrencyType 为 Exclusive 时，支持最大输出 4096*4096。当输入视频分辨率<=4096*4096时，输出分辨率和输入视频一致。（4K 视频是 2K 视频合成实时率的2倍）</p>
DriverType	string	是	<p>驱动类型，该字段必填。</p> <ol style="list-style-type: none"> 1. Text: 文本驱动，要求 InputSsml 字段必填。 2. OriginalVoice: 原声音频驱动，要求 InputAudioUrl 字段必填。
IdentityWrittenUrl	string	否	PDF 格式授权书，格式为 PDF，文件大小小于10M。
IdentityVideoUrl	string	否	视频格式授权书，格式为 mp4，文件大小小于5G。
InputAudioUrl	string	否	<p>驱动数智人的音频 URL，当 DriverType 为 OriginalVoice 时，该字段必填。</p> <ol style="list-style-type: none"> 1. 支持格式：wav、mp3、wma、m4a、aac、ogg 2. 时长不超过20分钟，不少于1秒。 3. 大小要求：不超过100MB 4. 请保证文件的下载速度，否则会下载失败或导致视频制作实时率下降。
InputSsml	string	否	<p>播报文本内容，支持 SSML 标签，支持的标签类型参照 数智人 SSML 标记语言规范，标签写法参照示例，内容不能换行，符号需转义。上限2000字，不少于4个字（字数按 unicode 字符数计算）。DriverType 为空或 Text 时，该字段必填。</p>
SpeechParam	object	否	定义音频的详细参数。DriverType 为 Text 时，该字段必填。
SpeechParam.Speed	float	否	语速（1.0为正常语速，范围[0.5-1.5]，值为0.5时播报语速最慢，值为1.5时播报语速最快，DriverType 为音频驱动类型时，语速控制不生效），DriverType 为 Text 时，该字段必填。
SpeechParam.TimbreKey	string	否	音色 key，DriverType 为 Text 时，该字段必填。
SpeechParam.Volume	int	否	音量大小，范围[0, 10]，对应音量大小。默认为0，代表正常音量，值越大音量越高。
			<p> 说明：</p>

			<p>TimbreKey 在 male_1-20、female_1-23（即男声1-20、女声1-23）间的音色不支持音量调节。</p>
SpeechParam.EmotionCategory	string	否	控制合成音频的情感，仅支持多情感音色使用，可选值参考个人资产管理 API 分页查询音色列表 接口。
SpeechParam.EmotionIntensity	int	否	控制合成音频情感程度，取值范围为 [50,200]，只有 EmotionCategory 不为空时生效。
SpeechParam.TimbreLanguage	string	否	音色语种，可选语种参考个人资产管理 API 分页查询音色列表 接口，多语种音色在合成时必须选择对应语种。
ConcurrencyType	string	否	<p>视频制作任务使用的资源类型。</p> <ol style="list-style-type: none"> 1. Exclusive：使用并发调用，不扣除小时包，需要购买并发，如果没有购买，任务提交失败。 2. Shared：调用会扣除小时包，需要购买小时包，如果没有购买，任务提交失败。 3. 不填：如果购买了并发或者并发和小时包都购买则为“Exclusive”，如果没有购买并发但购买了小时包则为“Shared”，如果都没有购买，任务提交失败。
VideoLoop	int	否	<p>当音频时长大于视频时长时，生成的视频会对齐音频时长。以下是两种对齐方式： 0：反向拼接、1：正向拼接。默认 0。</p>
VideoParametersConsistent	int	否	<p>视频输出标准（码率与帧率）对齐选项，默认0。</p> <ul style="list-style-type: none"> ● 0：不强制对齐输入视频格式。 ● 1：对齐输入视频格式。 <p>参数说明：</p> <p>一、选择“0”时</p> <ol style="list-style-type: none"> 1. 码率： <ul style="list-style-type: none"> ○ 默认采用 CRF=17 的编码方式，编码器会根据视频内容复杂度动态调整输出码率，确保质量稳定。 2. 帧率： <ul style="list-style-type: none"> ○ 默认输出帧率为 25fps。 <p>二、选择“1”时</p> <ol style="list-style-type: none"> 1. 码率处理规则： <ul style="list-style-type: none"> ○ 输入码率 < 1000kbps → 输出码率调整为 1000kbps。 ○ 输入码率 > 9000kbps → 输出码率调整为 9000kbps。

			<ul style="list-style-type: none"> ○ 输入码率在 [1000kbps, 9000kbps] 范围内 → 输出码率与输入一致。 <p>2. 帧率处理规则:</p> <ul style="list-style-type: none"> ○ 输入帧率 < 15fps → 输出帧率调整为 15fps。 ○ 输入帧率 > 60fps → 输出帧率调整为 60fps。 ○ 输入帧率在 [15fps, 60fps] 范围内 → 输出帧率与输入一致。
CallbackUrl	string	否	<p>当用户增加回调 URL 时，将把视频制作结果以固定格式发送 POST 请求到该 URL 地址，固定格式见 附录二：回调请求体格式，需注意：</p> <ol style="list-style-type: none"> 1. 限制 CallbackUrl 长度小于1000。 2. 只发送一次请求，无论是哪种问题导致的请求失败，都不会再进行发送。
VideoStorageS3Url	string	否	可传入含鉴权 S3 协议存储 URL，视频成品会上传至该 URL。
VideoParam	object	否	定义合成视频的详细参数。
VideoParam.RefPhotoUrl	string	否	<p>当“人脸 ID 追踪开关”开启时，该字段生效。</p> <p>用户上传的人脸参考图 URL。</p> <p>输入视频中包含多张人脸时，VideoMakeNoTrain 视频生成 API 仅能选择一张人脸作为目标替换其口型，实现口型与音频相匹配的效果。此参数用来指定以哪个人物作为目标。</p> <p>若不输入人脸参考图，默认将选择视频中第一个有人脸的画面中，人脸占比最大的人物为目标。</p> <p>图像文件要求：</p> <ol style="list-style-type: none"> 1. 文件大小：文件 ≤ 10MB 2. 图像大小：单边分辨率要求在 [192~4096] 3. 格式支持 jpg、jpeg、png、bmp、webp 4. 内容：需包含一张清晰的人物正脸，且为视频中出现的人物。
VideoParam.RefPhotoUrls	string[]	否	<p>当“人脸 ID 追踪开关”开启时，该字段生效。</p> <p>用户上传的多张人脸参考图 URL 数组，用于辅助定位目标人脸。</p> <p>适用场景：</p> <p>当输入视频中包含多张人脸，或单张参考图不足以唯一确定目标人脸时（如动漫场景中人物脸型相似度较高），可通过提供多张不同角度、表情或光照条件下的人脸图像（如人物三视图），提升人脸识别的准确性与鲁棒性。系统将综合多张图像提取的人脸特征，与视频中的人脸进行匹配，最终选定目标人脸进行口型替换。</p> <p>与 RefPhotoUrl 的关系：</p> <p>若同时传入 RefPhotoUrl 与 RefPhotoUrls，以 RefPhotoUrls 为准，RefPhotoUrl 将被忽略。若两者均未传入，默认将选择视频中第一个有人脸的画面中，人脸占比最大的人物为目标。</p>

			<p>图像文件要求（每张图片）：</p> <ol style="list-style-type: none"> 1. 文件大小：单张文件 ≤ 10MB 2. 图像大小：单边分辨率要求在 [192, 4096] 3. 格式支持 jpg、jpeg、png、bmp、webp 4. 内容：需包含清晰的人物正脸或不同角度的人脸，且均为视频中出现的人物。 <p>使用建议： 建议提供 2~5 张不同姿态的图片以获得更优匹配效果，最多支持同时上传 10 张图片。搭配 FaceMatchMode 设为 Strict（严格匹配模式）可进一步提升定位精度。</p>
VideoParam.DisabledDetect	int	否	<p>人脸 ID 追踪开关。默认0。</p> <p>0：开启人脸 ID 追踪。开启后，可以通过人脸 ID 参数（VideoParam.RefPhotoUrl）指定人脸 ID 进行驱动（如果不指定就以识别到的第一个有效人脸 ID 为主）。</p> <p>1：关闭人脸 ID 追踪。关闭后，人脸 ID 参数（VideoParam.RefPhotoUrl）无效，识别出人脸 ID 均会驱动（如同一帧画面中出现多张人脸，最大的人脸会作为有效人脸 ID）。</p>
VideoParam.DisableOcclusionDetect	int	否	<p>遮挡检测开关。默认0。</p> <p>0：开启遮挡检测。开启后，当嘴部有遮挡时，不会驱动口型。</p> <p>1：关闭遮挡检测。关闭后，当嘴部有遮挡时，也会驱动口型。</p>
VideoParam.EnableFakeTeeth	int	否	<p>假牙处理开关。默认0。</p> <p>0：关闭假牙处理。关闭后，最终合成的视频里的牙齿，会学习原视频里的牙齿特征。</p> <p>1：开启假牙处理。开启后，最终合成的视频里的牙齿会重新生成，不会参考原视频的牙齿特征。</p>
VideoParam.EnableMusicMode	int	否	<p>唱歌口型模式（适用于输入音频是歌曲的场景）开关。默认 0。</p> <p>0：关闭唱歌口型模式，使用口播模式。</p> <p>1：开启唱歌口型模式。</p>
VideoParam.MakeType	string	否	<p>视频定制类型。</p> <ul style="list-style-type: none"> • Default：默认配置，视频片段默认会从原视频第0秒开始，StartTime 和 EndTime 均不生效 • Custom：指定视频片段，可以填写 StartTime 和 EndTime 来选择片段（需要大于1s），默认生成的视频用该“指定的视频片段”进行正倒放循环。 • Circle：首尾帧归位，可以填写 StartTime 来指定视频开始时间（EndTime 不生效），并且此时 VideoLoop 参数不生效。

			<ul style="list-style-type: none"> CustomOnlyStart: 通过 StartTime 指定视频起始时间, EndTime 不生效, 此时 VideoLoop 参数不生效, 默认为反向拼接。
VideoParam.MouthMode	string	否	<p>口型模式, 控制视频中人物的口型幅度, 默认值为 Soft。</p> <p>Soft: 柔和口型 (默认)。口型效果柔和自然。</p> <p>Exaggerated: 夸张口型。口型幅度更大, 适用于欧美风格表达或需要强化口型表现力的视频场景。</p>
VideoParam.StartTime	float	否	<p>起始时间, 单位秒 (有效数字为小数点后3位), 当 MakeType 为 Custom 或 Circle 时生效, 当填写的时候, 默认生成的视频从该位置开始; 不填的时候, 默认选取视频开始时间。</p>
VideoParam.EndTime	float	否	<p>结束时间, 单位秒 (有效数字为小数点后3位), 当 MakeType 为 Custom 时生效, 当填写的时候, 默认结束的视频到该位置结束; 不填的时候, 默认选取该“选中的全部视频片段”的结束时间</p>
VideoParam.DisableIntervals	Array of [DisableInterval]	否	<p>定义视频中不驱动时间片段列表 (目前最大支持5个时间片段), 列表需按照时间顺序递增, 否则任务提交失败。</p> <p>例如: [[1,2],[3,4],[5,6]] 表明原视频的1s~2s、3s~4s、5s~6s不做口型驱动。</p>
VideoParam.FaceMatchMode	string	否	<p>人脸匹配模式, 当"人脸 ID 追踪开关"开启且传入 RefPhotoUrl 或 RefPhotoUrls 时生效。默认值为 Loose。</p> <p>Strict: 严格匹配模式。要求参考照片与视频中人脸高度相似, 适用于需要精确锁定特定人物的场景。推荐在以下情况使用: 视频中存在多个相似人脸 (如动漫角色)、或需要配合多张参考图 (RefPhotoUrls) 精准定位目标人脸时。</p> <p>Loose: 宽松匹配模式 (默认)。允许参考照片与视频中人脸存在一定差异, 适用于参考照片角度、光线与视频不一致的情况, 容错性更高。</p>

DisableInterval

参数	类型	必须	说明
StartTime	float	否	不驱动片段起始时间, 单位秒, (有效数字为小数点后3位)
EndTime	float	否	不驱动片段结束时间, 单位秒, (有效数字为小数点后3位)

返回参数

参数	类型	必须	说明
TaskId	string	是	视频制作的任务 ID，携带 TaskId 访问 音视频制作进度查询接口 ，可获得制作进度和制作结果

请求示例

文本驱动

```
{
  "Header": {},
  "Payload": {
    "RefVideoUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/ref_video.mp4",
    "DriverType": "Text",
    "InputSsml": "你好，我是虚拟<phoneme alphabet=\"py\" ph=\"fu4\">主</phoneme>播",
    "SpeechParam": {
      "TimbreKey": "female_1",
      "Volume": 1,
      "Speed": 1.0
    }
  }
}
```

音频驱动

```
{
  "Header": {},
  "Payload": {
    "RefVideoUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/ref_video.mp4",
    "DriverType": "OriginalVoice",
    "InputAudioUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/audio.mp3"
  }
}
```

返回示例

成功

```
{
  "Header": {
    "Code": 0,
    "DialogID": "",
    "Message": "",
    "RequestID": "fde854eaa981c7f2f7285d1c7eca335b",
    "SessionID": "gzb7dec22117297528294581119"
  },
  "Payload": {
    "TaskId": "81883d47c6154edf8e276531f09227b6"
  }
}
```

失败

```
{
  "Header": {
    "RequestID": "d99cf740be48d7b0fb1eee4447b71781",
    "SessionID": "gz6240ef0317651792698007383",
    "DialogID": "",
    "Code": 100001,
    "Message": "InvalidParameter:请求参数错误: Speed value invalid, ranges:0.5-1.5"
  }
}
```

编码格式转换小教程

该教程专用于将视频转成 H.264 编码格式，主要原理为：通过 FFmpeg 实现转换。

第一步：安装 FFmpeg

- Windows 系统：

从 FFmpeg 官网下载 Windows 版本的压缩包。

- macOS 系统：

可以使用 Homebrew 进行安装，命令为 `brew install ffmpeg`。

- Linux 系统:

- Ubuntu/Debian 系统使用命令 `sudo apt-get install ffmpeg`。
- Fedora 系统使用命令 `sudo dnf install ffmpeg`。

第二步：转换

1. 准备好两个路径：转换前视频路径和转换后视频路径，然后在“终端”中输入以下指令。

```
ffmpeg -i 转换前视频路径 -c:v libx264 -preset medium -crf 23 -c:a aac -  
b:a 128k 转换后视频路径
```

例如：

- 转换前视频路径为：“/Users/xxxx/Downloads/picture/video/视频名.mov”
- 转换后视频路径为：“/Users/xxxx/Downloads/picture/video/转换后视频名.mp4”

2. 则终端输入指令为：

```
ffmpeg -i "/Users/xxxx/Downloads/picture/video/视频名.mov" -c:v libx264  
-preset medium -crf 23 -c:a aac -b:a 128k  
"/Users/xxxx/Downloads/picture/video/转换后视频名.mp4"
```

⚠ 注意：

Windows 系统需要通过 `cd` 切换目录到压缩包 .exe 所在的文件夹后，再执行指令。

第三步：验证

验证是否为 H.264：在上一步指定路径下获取已转换的视频后，可借助系统自带播放器或第三方 VLC 等软件检查。

视频制作接口-照片免训练

最近更新时间：2025-12-16 17:14:52

接口说明

无需训练，即可基于一张形象照片（真人/卡通人物/卡通动物皆可），通过输入文本或音频，生成人物说话口型与输入内容相匹配的新视频。

通过 [音视频制作进度查询接口](#) 最终返回成品视频。目前音视频资源只保留7天，请尽快下载。

调用协议

- HTTPS + JSON
- POST `/v2/ivh/videomaker/broadcastservice/phototovideonotrain`
- Header `Content-Type: application/json;charset=utf-8`

请求参数

参数	类型	必须	说明
RefPhotoUrl	string	是	请传入包含人像/卡通/动物面部的图片URL地址，并注意： <ul style="list-style-type: none">• 格式：支持 JPG、PNG、BMP、WEBP• 大小：不超过 6MB• 尺寸建议：单边在 192px ~ 4096px 之间，宽高比建议在 1:2 ~ 2:1• 画面建议：请使用面部清晰、无严重遮挡的图片，避免使用无面部主体的图片（如风景、物体或文字截图）
DriverType	string	是	驱动类型，该字段必填。 <ul style="list-style-type: none">• Text：文本驱动，要求 InputSsml 字段必填。• OriginalVoice：原声音频驱动，要求 InputAudioUrl 字段必填。
InputAudioUrl	string	否	驱动数智人的音频 URL，当 DriverType 为 OriginalVoice 时，该字段必填。格式要求如下： <ul style="list-style-type: none">• 时长范围：2秒-60秒• 支持格式：wav、mp3、wma、m4a、aac、ogg• 文件大小：≤10M
InputSsml	string	否	播报文本内容，支持 SSML 标签，支持的标签类型参照 数智人 SSML 标记语言规范 ，标签写法参照示例，内容不能换行，符号需转义。上限300字（文本驱动底层会转换为音频，如果时长超过60秒时任

			务会制作失败)，不少于4个字（字数按 Unicode 字符数计算）。DriverType 为空、或 Text 时，该字段必填。
SpeechParam	object	否	定义音频的详细参数。DriverType 为 Text 时，该字段必填。
SpeechParam.Speed	float	否	语速（1.0为正常语速，范围[0.5-1.5]，值为0.5时播报语速最慢，值为1.5时播报语速最快，DriverType 为音频驱动类型时，语速控制不生效），DriverType 为 Text 时，该字段必填。
SpeechParam.TimbreKey	string	否	音色 Key，DriverType 为 Text 时，该字段必填。
SpeechParam.Volume	int	否	音量大小，范围[0, 10]，对应音量大小。默认为0，代表正常音量，值越大音量越高。 说明： TimbreKey 在 male_1-20、female_1-23（即男声1-20、女声1-23）间的音色不支持音量调节。
SpeechParam.EmotionCategory	string	否	控制合成音频的情感，仅支持多情感音色使用，可选值参考个人资产管理 API 分页查询音色列表 接口。
SpeechParam.EmotionIntensity	int	否	控制合成音频情感程度，取值范围为 [50,200]，只有 EmotionCategory 不为空时生效。
SpeechParam.TimbreLanguage	string	否	音色语种，可选语种参考个人资产管理 API 分页查询音色列表 接口，多语种音色在合成时必须选择对应语种。
CallbackUrl	string	否	当用户增加回调 URL 时，将把视频制作结果以固定格式发送 POST 请求到该 URL 地址，固定格式见 附录二：回调请求体格式 ，需注意： 1. 限制 CallbackUrl 长度小于1000。 2. 只发送一次请求，无论是哪种问题导致的请求失败，都不会再进行发送。
VideoStorageS3Url	string	否	可传入含鉴权 S3 协议存储 URL，视频成品会上传至该 URL。
VideoParam	object	否	定义输出视频的相关参数，不填时取各字段默认值。

VideoParam.Resolution	int	否	<p>输出视频的分辨率，可选。</p> <ul style="list-style-type: none"> 0: 720P (默认值): 高效省时之选，每秒视频渲染预计需要20秒。 1: 1080P: 画面质感更优，渲染时间约为720P的2倍以上。
VideoParam.Prompt	string	否	<p>人物表现提示词，用于定义数字人的行为与动作。最大长度不超过5000字符（按 Unicode 字符数计算）。</p> <p>注意：当此参数留空时，系统将自动使用默认提示词：“画面中的人物正在对着镜头讲话，偶尔做些手势匹配说话的内容。镜头保持固定。”</p>

返回参数

参数	类型	必须	说明
TaskId	string	是	视频制作的任务 ID，携带 TaskId 访问 音视频制作进度查询接口 ，可获得制作进度和制作结果

请求示例

文本驱动

```

{
  "Header": {},
  "Payload": {
    "RefPhotoUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-nanjing.myqcloud.com/ref_photo.jpg",
    "DriverType": "Text",
    "InputSsml": "你好，我是虚拟<phoneme alphabet=\"py\" ph=\"fu4\">主</phoneme>播",
    "SpeechParam": {
      "TimbreKey": "female_1",
      "Volume": 1,
      "Speed": 1.0
    }
  }
}

```

音频驱动

```
{
  "Header": {},
  "Payload": {
    "RefPhotoUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-
nanjing.myqcloud.com/ref_photo.jpg",
    "DriverType": "OriginalVoice",
    "InputAudioUrl": "http://virtualhuman-cos-test-1251316161.cos.ap-
nanjing.myqcloud.com/audio.mp3"
  }
}
```

返回示例

成功

```
{
  "Header": {
    "Code": 0,
    "DialogID": "",
    "Message": "",
    "RequestID": "fde854eaa981c7f2f7285d1c7eca335b",
    "SessionID": "gzb7dec22117297528294581119"
  },
  "Payload": {
    "TaskId": "81883d47c6154edf8e276531f09227b6"
  }
}
```

失败

```
{
  "Header": {
    "RequestID": "d99cf740be48d7b0fb1eee4447b71781",
    "SessionID": "gz6240ef0317651792698007383",
    "DialogID": "",
    "Code": 100001,
    "Message": "InvalidParameter:请求参数错误: Speed value invalid,
ranges:0.5-1.5"
  }
}
```

```
}
```

音视频制作进度查询接口

最近更新时间：2025-12-10 14:56:32

接口说明

通过 TaskId 来查询任务的制作进度和结果，当返回值里的 Progress 字段值为100时，即可通过 MediaUrl 获取最终音视频的下载地址，如果是视频制作也会返回 SubtitlesUrl 字段获取最终 SRT 字幕地址。

⚠ 注意：

制作完成的视频资源默认存储在数智人内部账号的 COS 下，只保留7天，可以通过在制作接口中的 VideoStorageS3Url 字段指定第三方存储 URL，这样制作完的视频会上传到指定地址。

调用协议

HTTPS + JSON

POST /v2/ivh/videomaker/broadcastservice/getprogress

Header Content-Type: application/json;charset=utf-8

请求参数

参数	类型	必须	说明
TaskId	String	是	任务 ID，在 音频制作接口 、 视频制作接口-基础版 、 视频制作接口-进阶版 接口中返回。

返回参数

参数	类型	必须	说明
Progress	int	是	制作进度，-1~100，-1代表生成失败，100代表生成成功（预留字段，目前不具备参考意义）
MediaUrl	String	是	音视频结果地址
SubtitlesUrl	String	否	当制作视频时，返回视频对应的 SRT 字幕地址

Status	String	是	制作状态 <ul style="list-style-type: none"> "COMMIT": 已提交需要排队 "MAKING": 制作中 "SUCCESS": 制作成功 "FAIL": 制作失败
ArrayCount	int	是	Status 为 "COMMIT" 状态时在该任务之前排队的任务数量
FailCode	int	否	制作失败返回的错误码, 对应附录一中的“音视频制作结果错误码”
FailMessage	String	否	制作失败返回的失败原因, 便于排查问题
TextTimestampResult	Array of [Sentence]	否	该字段返回 TTS 试听任务的文本时间戳信息
Duration	int	是	视频时长, 单位 ms

Sentence

参数	类型	必须	说明
Sentence	String	是	拆句得出的句子
Words	Array of [Word]	是	句子中每个字的信息

Word

参数	类型	必须	说明
Word	String	是	句子中的一个字
StartTimestamp	long	是	该字的时间起点, 该数值/10000为 ms, 相当于单位为 0.1us
EndTimestamp	long	是	该字的时间尾点, 该数值/10000为 ms, 相当于单位为 0.1us

请求示例

```
{
  "Header": {},
}
```

```
"Payload": {
  "TaskId": "81883d47c6154edf8e276531f09227b6"
}
```

返回示例

成功

```
{
  "Header": {
    "Code": 0,
    "Message": "",
    "RequestID": "fde854eaa981c7f2f7285d1c7eca335b"
    "SessionID": "gzb7dec22117297528294581119"
  },
  "Payload": {
    "Progress": 100,
    "MediaUrl": "https://videomaker-
resources.ivh.qq.com/81883d47c6154edf8e276531f09227b6.mp3",
    "SubtitlesUrl": "",
    "ArrayCount": 0,
    "FailMessage": "",
    "Duration": 11810,
    "FailCode": 0,
    "TextTimestampResult": [
      {
        "Sentence": "你好虚拟主播。",
        "Words": [
          {
            "Word": "你",
            "EndTimestamp": 6100000,
            "StartTimestamp": 4500000
          },
          {
            "Word": "好",
            "EndTimestamp": 8200000,
            "StartTimestamp": 6100000
          }
        ]
      }
    ]
  }
}
```

```
        "Word": "虚",
        "EndTimestamp": 9700000,
        "StartTimestamp": 8200000
    },
    {
        "Word": "拟",
        "EndTimestamp": 11100000,
        "StartTimestamp": 9700000
    },
    {
        "Word": "主",
        "EndTimestamp": 12900000,
        "StartTimestamp": 11100000
    },
    {
        "Word": "播。",
        "EndTimestamp": 16000000,
        "StartTimestamp": 12900000
    }
]
}
]
```

失败

```
{
  "Header": {
    "RequestID": "74aabb5a4852942c300699eba39e8cb6",
    "SessionID": "gz6240ef0317651913998371486",
    "DialogID": "",
    "Code": 110006,
    "Message": "InternalServerError:记录不存在"
  }
}
```

视频制作接口- PPT 模式进度查询接口

最近更新时间：2026-04-01 18:23:31

接口说明

通过 TaskId 来查询 PPT 合成视频任务的制作进度和结果，当返回值里的 Progress 字段值为100时，即可通过 MediaUrl 获取最终视频的下载地址。

⚠ 注意：

制作完成的视频资源默认存储在数智人内部账号的 COS 下，只保留7天。

调用协议

HTTPS + JSON

POST /v2/ivh/mediaproduction/mediaproduction/getprogress

Header Content-Type: application/json;charset=utf-8

请求参数

参数	类型	必须	说明
TaskId	String	是	任务 ID，在 视频制作接口- PPT 模式 接口中返回。

返回参数

参数	类型	必须	说明
Progress	int	是	制作进度，-1~100，-1代表生成失败，100代表生成成功（预留字段，目前不具备参考意义）
MediaUrl	String	是	视频结果地址
Status	String	是	制作状态 <ul style="list-style-type: none">"COMMIT": 已提交需要排队"MAKING": 制作中"SUCCESS": 制作成功"FAIL": 制作失败
FailCode	int	否	制作失败返回的错误码， 对应附录一中的“音视频制作结果错误码”

FailMessage	String	否	制作失败返回的失败原因，便于排查问题
Duration	int	是	视频时长，单位：ms

请求示例

```
{
  "Header": {},
  "Payload": {
    "TaskId": "81883d47c6154edf8e276531f09227b6"
  }
}
```

返回示例

成功

```
{
  "Header": {
    "Code": 0,
    "Message": "",
    "RequestID": "fde854eaa981c7f2f7285d1c7eca335b",
    "SessionID": "gzb7dec22117297528294581119"
  },
  "Payload": {
    "Progress": 100,
    "MediaUrl": "https://videomaker-resources.ivh.qq.com/81883d47c6154edf8e276531f09227b6.mp3",
    "FailMessage": "",
    "Duration": 11810,
    "FailCode": 0
  }
}
```

失败

```
{
  "Header": {
    "RequestID": "74aabb5a4852942c300699eba39e8cb6",
  }
}
```

```
"SessionID": "gz6240ef0317651913998371486",  
"DialogID": "",  
"Code": 110006,  
"Message": "InternalServerError:记录不存在"  
}  
}
```

附录

附录一：结果码字典

最近更新时间：2025-05-14 17:43:22

一、接口错误码

代码	含义	措施
0	成功	无
-101000	系统内部错误	重试或联系管理员定位错误
-101001	算力不足	重试或联系管理员定位错误
-101002	下载图片失败	确认提供的图片 url 是否可正常访问、下载
-101003	请求S3失败	重试或联系管理员定位错误
100001/100003	请求参数错误（包括参数格式、类型等错误，缺少参数错误，必传参数没填）	对照 API 协议核查参数格式、类型
100002	参数取值错误	对照 API 协议核查参数列表
100005	未授权操作	检查对请求的会话是否有操作权限
100008	超过配额限制	会话超过并发限制
100009	资源不存在	当前权限范围是否越权
110005	记录已经存在	过快的重复请求
110006	记录不存在	请检查请求的会话是否存在
110007	virtualmankey 已过期	检查 key 所属主账号的有效期
100013	未找到有效的小时包或小时包余量不足	申请或加购对应小时包
100015	数智人形象已经过期	申请形象延期
100016	未找到形象资产	确认使用的 VirtualmanKey 是否有效
801000	调用远程服务异常	重试或联系管理员定位错误
801002	调用远程服务，返回状态码异常	重试或联系管理员定位错误

801004	操作频繁，请稍后重试	避免同一时间提交多个任务
801005	文本安全校验未通过	请检查 ssml 是否有敏感信息

二、音视频制作结果错误码

代码	含义	措施
0	成功	无
801000	系统内部错误	重试或联系管理员定位错误
801001	数据库操作错误	重试或联系管理员定位错误
801002	调用远程服务错误	重试或联系管理员定位错误
801003	任务制作超过重试次数	重试或联系管理员定位错误
801004	任务制作过程发生崩溃	重试或联系管理员定位错误
801006	操作S3失败	重试或联系管理员定位错误
801007	任务执行超时	重试或联系管理员定位错误
801008	下载文件失败	检查提供的文件资源是否可正常访问、下载
801009	参数错误	检查 FailMessage 中提示的参数是否符合接口要求
801010	获取 url 资源失败	检查提供的 url 资源是否可正常访问、下载
801012	任务参数解析失败	重试或联系管理员定位错误
801014	下载字幕文件失败/下载生成视频失败	重试或联系管理员定位错误
801015	上传文件到指定的S3失败	检查提供的 VideoStorageS3Url、SubtitleStorageS3Url 是否可用
801401	音频制作，解析任务参数失败	重试或联系管理员定位错误
801402	音频制作，创建本地文件失败	重试或联系管理员定位错误
801403	音频制作，接收音频数据失败	重试或联系管理员定位错误
801404	音频制作，接收音频数据，发生业务异常	重试或联系管理员定位错误

801405	音频制作，mp3 文件转换失败或合成背景音乐失败	重试或联系管理员定位错误
801501	视频制作，解析参数错误	重试或联系管理员定位错误
801503	视频制作，调用远程服务失败	重试或联系管理员定位错误
801504	视频制作，创建拼接文件失败	重试或联系管理员定位错误
801506	视频制作，制作 logo/字幕或者拼接视频失败	重试或联系管理员定位错误
801507	视频制作，传入图片等资源分辨率转换失败	重试或联系管理员定位错误
801508	视频制作，获取小样本指定视频片段帧数失败	重试或联系管理员定位错误
801509	视频制作，检查小样本输入帧数不符合规定	检查 StartIdx、EndIdx 是否符合接口定义
801510	视频制作，音频驱动文件时长校验失败	检查音频驱动文件、重试或联系管理员定位错误
802000	免训练视频制作，照片审核失败	检查上传的照片内容是否涉及敏感信息
802001	免训练视频制作，照片人脸检测失败	检查上传的照片中是否仅有单一人脸
802002	免训练视频制作，预处理失败	重试或联系管理员定位错误
802003	免训练视频制作，SDK 处理错误	请检查输入视频或照片中是否存在人脸丢失情况
802004	免训练视频制作，回调错误	重试或联系管理员定位错误

附录二：回调请求体格式

最近更新时间：2024-11-13 14:38:22

参数	类型	必须	说明
Progress	int	是	制作进度，-1代表生成失败，100代表生成成功（预留字段，目前不具备参考意义）
MediaUrl	string	是	音视频结果地址
Subtitles Url	string	否	当制作视频时，返回视频对应的SRT字幕地址
Status	string	是	制作状态 "SUCCESS": 制作成功 "FAIL": 制作失败
FailMessage	string	否	制作失败返回的失败原因，便于排查问题
TaskId	string	是	请求视频制作返回的 TaskId

请求示例

```
{
  "Payload": {
    "Progress": 100,
    "MediaUrl": "https://videomaker-
resources.ivh.qq.com/broadcast/Basic/video/25215772ab424ad38858f6b692426
989.mp4",
    "SubtitlesUrl": "https://videomaker-
resources.ivh.qq.com/broadcast/Basic/subtitle/25215772ab424ad38858f6b692
426989.srt",
    "Status": "SUCCESS",
    "FailMessage": "",
    "TaskId": "25215772ab424ad38858f6b692426989"
  }
}
```

```
}
```

接口对接常见问题

最近更新时间：2025-07-11 18:41:12

调用接口时出现鉴权不通过？

1. 确认签名时参数按照字典序排列了。
2. 确认时间戳参数为当前时间。
3. 参考以下示例参数，按照自己的签名步骤计算一下，如果计算结果与示例一致则表示签名计算步骤无误。

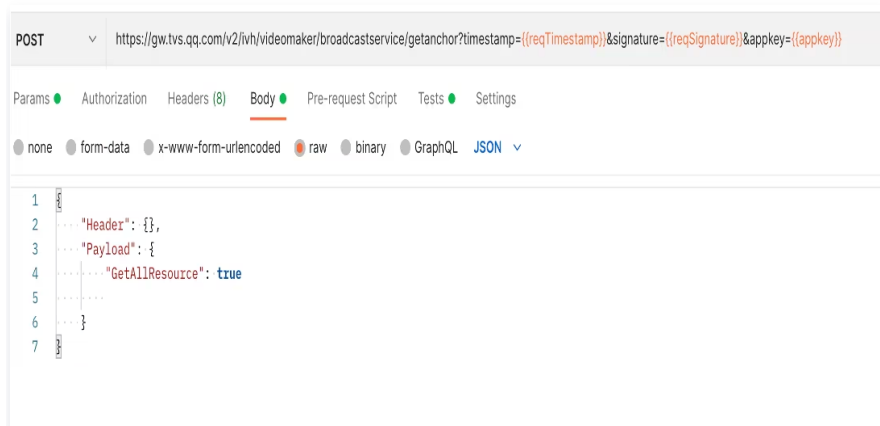
- 参数：

- appkey = e38267c0e86411ebb02aed82acb0ed99
- accesstoken = f68f2d10ae9e4604b76fb05cf46bccec
- timestamp=1646636485

- 签名结果：BfWuaC9kmaicCggXc693uK%2BsZQ8qe88O4HVQNTdwZuo%3D

调用接口出现参数错误？

1. 请确认请求的 body 是否加了 header 参数，不加的话会报错。如右图所示。
2. 参数的类型要严格按照接口文档中填充。



查询出的形象数量不对或查询不到形象？

- 如果您开通的是小样本形象，则需要参考 [分页查询小样本形象列表接口](#) 进行查询。
- 如果您开通的是 2D 精品形象，则需要参考 [客户资源查询主播接口](#) 进行查询。

ⓘ 说明：

官网下单播报小时包后，开通的形象既有 2D 小样本，又有 2D 精品，需要参考两套接口分别查询相关形象。

是否支持生成透明背景的视频？

支持，可以生成 WebM 格式的带透明通道的视频。

为什么 WebM 透明背景的视频播放出来是绿幕的？

说明客户使用的播报器不支持 WebM 格式，可以尝试 Chrome 中进行播放，如果出来的视频是黑底的，说明视频本身没问题。

形象位置参数调整后不生效？

1. 确认调用的接口是基础版还是进阶版，只有进阶版接口支持。
2. 确认视频输出格式是否为 GreenScreenMp4，设置为该值时不支持调整位置。

主播位置计算方法？

您可以参考下图进行计算。



参数1: HorizontalPosition: 主播横向位置 (默认值是0, 负数为向左调整, 正数为向右调整)
 定义: 屏幕中轴线距形象中轴线距离 / 整个屏幕宽度;
 公式: $HorizontalPosition = (x + anchorWidth / 2 - containerWidth / 2) / containerWidth$

参数2: VerticalPosition: 主播纵向位置 (默认值是0, 负数为向上调整, 正数为向下调整)
 定义: 形象底部距屏幕底部距离 / 两倍形象高度
 公式: $VerticalPosition = (anchorHeight + y - wrapperHeight) / 2 / anchorHeight$

参数3: Scale: 主播大小 (默认值是: 1; 数值越小, 主播越小)
 缩放规则: 左下顶点保持不动, 按照形象宽度、形象高度等比例缩放。

相关参数的定义:

- ① anchorHeight: 即为叠加了缩放系数的形象高度, 即“原始图片高度” * originzoom * Scale
- ② anchorWidth: 即为叠加了缩放系数的形象宽度, 即“原始图片宽度” * originzoom * Scale
- ③ 原始图片高度、宽度: 即《个人资产管理apaas接口协议》→4.1分页查询小样本形象列表接口中, PoselImage的高度和宽度。
- ④ originzoom: 即默认的缩放系数。为了在不同的分辨率下人物默认有最佳展示效果, 赋予了一个默认的缩放系数。可以在《个人资产管理apaas接口协议》→4.1分页查询小样本形象列表接口 查询。

如何计算数智人初始大小？

数智人查询接口中会返回 OriginZoom (原始缩放系数) 和 PoselImage (数智人姿态图片), OriginZoom 乘以 PoselImage 的宽高即为数智人初始大小。

文案中插入动作不生效？

动作前后需要预留一定的文字，每个动作对预留文字都不一样，建议前后预留10个字以上。

数智人是否支持调整语速，韵律，多音字？

可以，参考接口文档中的 SSML 标签 [附录二：SSML 规范](#) 部分。

指定了外部的 S3 存储，数智人视频未上传上去？

请检查 S3 的权限设置是否允许外部写入，或者参数中的 S3 链接是否带入了鉴权信息。

为什么没有字幕文件？

只有文本驱动才有字幕文件，声音驱动没有。

超过配额限制：提交音频制作数量超过限制：1？

开通权限时默认只开通1路音频制作并发，即同时只能有一个音频任务处于制作中，必须等上一个制作完成后才能提交下一个任务。可以通过购买音频并发解决。

超过配额限制：提交视频制作数量超过限制：n？

1. 针对每种形象类型，每个用户默认支持同时处理5个视频制作任务，由主子账号共享。这5个任务使用的并发资源所有客户共享，可能会出现任务排队的情况。
2. 可增购播报独享并发，例如购买了1路播报独享并发，则一共支持同时处理6个视频制作任务。增购的并发由该用户专享。

解决方案：

1. 用户增加任务排队策略。
2. 购买独享并发。

视频制作接口-视频免训练中的 VideoParametersConsistent 参数，选强制对齐和不强制对齐，对生成时间有什么影响吗，具体影响分别是什么？

- 这里指输出视频帧率码率跟输入视频是否进行对齐。
- 码率对齐的影响10%左右；帧率影响会大一些，例如25帧到60帧，大概+30%。

视频制作接口-视频免训练中的 VideoLoop 参数，反向拼接和正向拼接的含义是什么，是否有示例？

视频的正反向拼接，主要在音频时长大于视频时长时会用到。

假设输入音频2秒，视频1秒。

- 正向拼接：2个1秒视频直接拼接，可能会出现跳帧的情况，第一秒最后一帧和第二秒第一帧大概率不一致。
- 反向拼接：1秒视频先正放再倒放。

视频制作接口-视频免训练生成的形象有哪些音色可选择？

- 请参考 [音色列表](#) 中音色类型为“精品音色”的音色清单。
- 您可以在 [音色试听](#) 进行试听。

