

拟公示算法机制机理内容

算法名称	趣闯生成式音频算法-2
算法基本原理	<p>该算法包括一系列语音要素的生成与编解码模型，可用于各类跨语种的语音生成与编辑任务，覆盖讲话、清唱等形式。其中，包括一组自回归模型，分别从语音中提取语义特征、音色特征，并实现语义、韵律、音色等语音要素的互相解耦；包括一组非自回归模型，根据参考语音、输入文本等内容，结合分别提取的语义、韵律、音色等特征，重建生成内容的韵律、音色等要素，并预测其声学特征；包括一个解码器，实现从声学特征到语音波形的重建。</p>
算法运行机制	<p>通过上传的一段参考语音，算法首先提取其中包括的音色、韵律等特征。</p> <p>根据不同的任务，提取生成文本信息，当进行语音翻译、语音编辑等任务时，利用相应语种的语音识别技术，由参考语音中提取相应的文本；当进行语音翻译时，利用翻译模型将提取文本翻译为目标语言对应的文本；当进行语音编辑时，修改提取的文本。当进行歌声合成时，输入文本标记当前生成的目标是清唱语音，若包含曲谱，算法按曲谱生成清唱的韵律特征；若输入文本不包含曲谱，算法根据文本目标内容自动生成韵律特征。</p> <p>然后，根据上一步得到的文本或输入的目标文本，提取其中的语义特征。</p> <p>将语义特征与音色、韵律特征等要素进行结合，预测生成内容对应的声学特征。</p>

	最后，由声学特征重建生成的语音波形。最后，由声学特征重建生成的语音波形。
算法应用场景	该算法可应用于视频制作、媒体与文化传播、直播、数字人应用等业务相关的语音合成、声音克隆、语音翻译、语音编辑、歌声合成等场景。
算法目的意图	服务于数字人应用提供商、视频制作平台、语音交互平台等机构，解决其在语音、歌声等相关内容的生成与编辑等方面的需求。
算法公示情况（选填）	