## 柯南变声器的算法原理解析

链接：https://mp.weixin.qq.com/s/McWHS0esRRhIiBlB4CAGXA

### 第一章

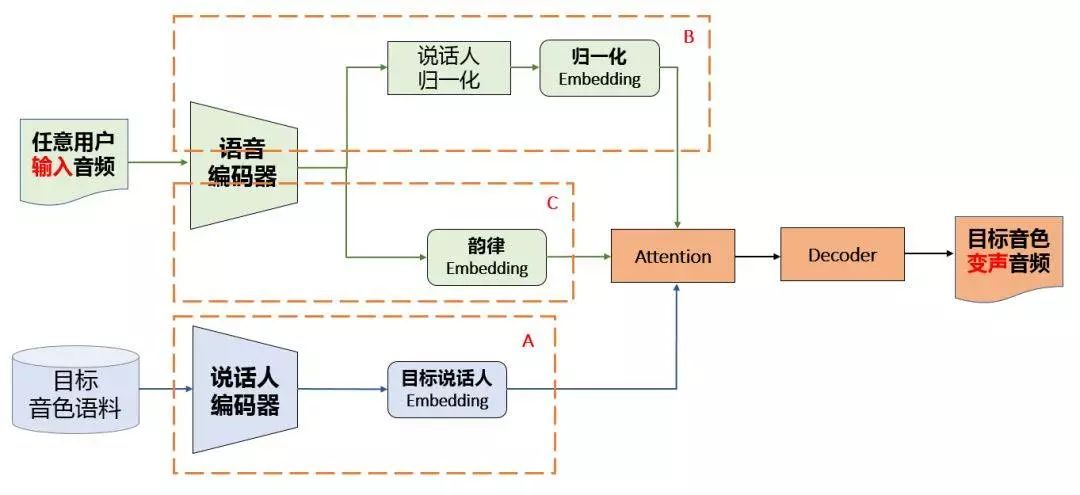
小时候特别喜欢看柯南，里面有个变声蝴蝶结，让我印象深刻。变声蝴蝶结的作用就是可以把输入的语言转化成其它人的声音。举个例子，如果我有这个语音转换功能，下次录歌的时候就可以用周杰伦的声音录制（唱功还是自己的）。



如果要实现变声功能，这里面包含哪些技术点呢，首先要提取出语音中的3种表征：

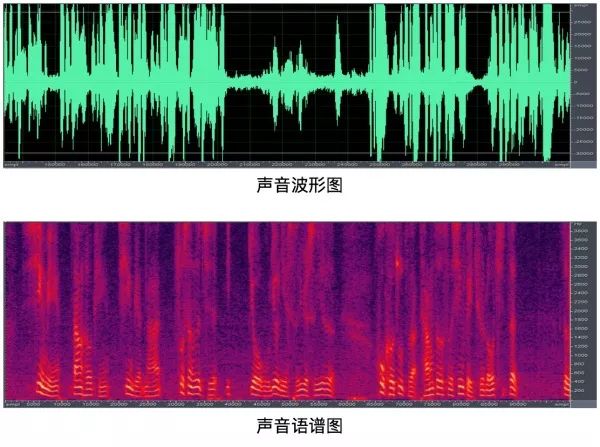
* 内容表征：输入语音的具体内容
* 声纹表征：可以理解为人的声色，每个人都是不同的
* 说话风格表征：说话的语气、强弱、快慢等风格相关

通过上述分析，如何实现变声功能就比较好理解了。首先把语音分解为内容表征、声纹表征、说话风格表征3种特征，然后把声纹表征替换成其他人的声纹特征即可实现。

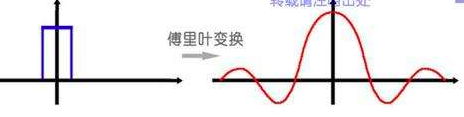


### 第二章

那么声纹识别如何实现就成了变声功能的关键。先看下声音信号是什么样的：



大家应该在各种k歌工具中都多少了解些声音信号的样子，基本上声音信号就是一堆脉冲。脉冲是一个时序信号数据，很难处理。需要对声音时序信号进行傅里叶转换，才能转换成频谱图（学过通信的同学应该比较好理解，通信专业的我当年碰到傅里叶这3个字就抖），转换效果如下图：



傅里叶转换后的图其实就成了一个与时间无关的图，所以就可以通过深度学习CNN算法去学习规律，对于深度学习有了解的同学应该知道CNN可以快速学习图像中的一些细节差别，其实根本原理跟人脸识别或者文字识别等场景都是一致的，于是就可以基于这种方法实现对于语音信号的声纹的分析和识别。

### 第三章

上文大概讲了下变声这个功能的大体理论以及声纹识别的算法原理，接下来讲一个适用场景吧，早在14年我在某外资手机巨头做手机软件开发的时候就想到一个场景，通过声纹识别解决骚扰电话。

当时想的方法是这样的：首先通过声纹识别把用户所有有关联的人的声纹pattern都识别出来，存储到手机内。一旦有个电话进来，先不着急接通，而是要求拨电话的人说一句话，这时候马上识别他的声纹与手机中存储的声纹pattern作比对。如果拨电话的人的声纹是一个新声纹，则判定为骚扰电话，拒绝接通~

现在想想，那时候这种通过声纹识别拦截骚扰电话的想法还是挺超前的。好啦，今天就简单分享到这里，谢谢大家，希望能对大家有所帮助~