**Applied Speech and Audio Processing 第一章读书报告**

**姓名：湛颖**

**1.介绍**

声音和语音处理系统在大部分发展中国家的人们的生活当中的地位是稳定上升的。在hifi音乐系统中，不管是无线电广播还是便携式音乐播放器，音频处理对于消费者的娱乐来说都是很重要的。数字音频技术现在在使用CD播放器，互联网电台，mp3播放器和i POD的音频传输方面占统治地位。在电视和电影工作室，以及现场演出的混音台，数字音频处理也是占优势地位的，音乐和音效在电脑游戏当中也很重要。

随着蜂窝通信的发展（尤其是欧洲GSM标准），语音处理在世界范围内的发展是一个上升的趋势。如今，GSM几乎是无处不在的，甚至在世界上最贫穷的地区，也有广泛的采用度。

当然，数字语音信号使用的年限已经不短，比如卫星通讯连接，但如今，传统电话网络（named POTS for ‘Plain Old Telephone Services’）也在向数字化屈服。最后一公里使用几百米的铜双绞线接入消费者的家中，这样的数字技术工作方式是从来不会出现在工程师的脑海中的，光纤通信以太网和无线连接在多年来也是一直被抵制的。DSL (digital subscriber line – normally asymmetric so it is faster in one direction than the other, hence ADSL) ，即使是这种模拟的双绞线，也可以可观地高速传达数字信号。a d s l，对于互联网视频传输或者是像skype这样的数字语音传输都是足够快的。

**1.1数字音频**

目前数字处理是解决音频和语言问题的选择方式：事实上，新的音频软件或系统都是占优势地位的。这种从模拟到数字的改变大多发生在过去十年，直到现在已经变成了一种默认的模式。

这样的趋势大部分发展到了语音音频和听力相关的研究中，许多相关的研究现在都是基于数字的。数字技术的一大好处是它几乎不依赖设备：一个数字处理平台上的技术，可以一直到另外一个上。一个标准在开发平台上会变的易用以及易测试，开发平台的标准可能完全不同，低功耗，小容量，或高功耗，低价格等。

从易用的层面开发，在PC上运行的matlab在这个领域的工作中广泛使用。它是专为处理数字信号而设计的，尤其是长串的音频采样。其内建的函数，允许大部分可能的常用的操作控制，也可以录音和回放，其可视化和绘图工具也很棒。本文作者常在m a c o sx和linux平台使用matlab完成它的音频工作。

尽管目前matlab并没有提供语音音频或听觉工具箱，信号处理工具箱包含了大多需要的函数，Department of Electrical and Electronic Engineering, Imperial College, London的voicebox开源工具箱也提供了许多有用的函数。也可以使用开源的octave开发环境执行本书中的大部分音频和语音进程，但需要对本书中的某些案例作出小的变更。Octave比matlab小众一些，它缺乏高级的绘图和侦错功能，但是其她的方面很相像。

**1.2收集和转换声音**

系统所有内容都是关于声音的。不是声音被机械产生，就是声音被人或机械听到。在纯物理层面，声音是由于分子振动穿越空气的纵波（或其她媒质中的横波）。在空气中，声音是作为高低压力变化传播的，高低变化的速度决定了频率。气压变化程度决定了幅度。

麦克风通过感受薄膜偏移捕获到的声波，成比例的转换成电压或电流。作为结果的电信号，通常是由模数转换器提供的的已编码的数字信号序列。最常见的格式，脉冲编码调制，将在5.1.1部分介绍。

如果相同的编码数据序列流经一个兼容的数模转换器，通过放大器送往扬声器，便可以产生声音。这样提供给扬声器的电压信号，都是每一个瞬间由电脑提供的采样值经过数模转换后比例放大所得来的。扬声器上的电压会导致圆锥体向内或向外位移，这样的位移时刻压缩附近的空气，从而产生声波。

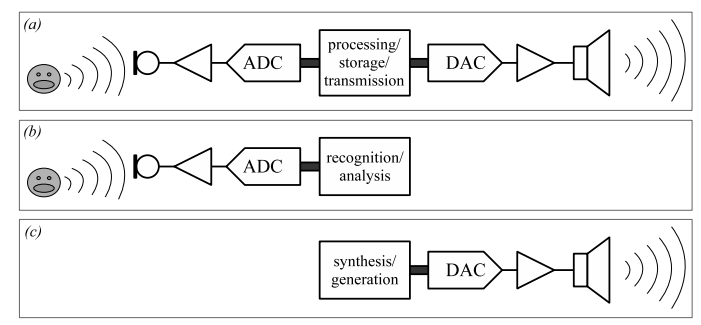


Figure 1.1 Block diagram of three classes of digital audio system showing (a) a complete digital audio processing system comprising (from left to right) an input microphone, amplifier, ADC, digital system, DAC, amplifier and loudspeaker. Variations also exist for systems recognising audio or speech (b), and systems synthesising audio (c).

上图所示的处理过程，代表了任何数字音频处理系统的主要步骤。音频，即本例中自由空气里传播的语音，被麦克风转换成电信号，然后放大以及可能会有的滤波，然后被ADC转换成数字信号。进入数字领域后，这些信号可以被很多方式处理，发送或者储存，或者被matlab试验。相反的过程，会将这些信号转换成声音。

上图中的处理、储存和传输系统都可能是串联或者并联的，各种情况下标准不一。光纤和无线传输也很受欢迎。

这个基本系统的变化如上图使用了其分析或合成声音的部分。立体声系统需要两个麦和扬声器，有些甚至更多。每个简单的放大器，图上的adc和dac模块也会隐藏某些系统中的复杂性——如模拟滤波、自动增益控制等，作为基础放大器的增强功能。

ADC个DAC也有自己的特点：采样率、技术、信噪比、动态范围等，通常由他们输出的位数决定。

**1.3采样**

一段音频采样序列首先需要考虑的是连续采样点间的时间间隔，通常它们被设计成均匀的。这些间隔的频率即采样率。在上图它应该由一个周期时钟信号发出送给adc和dac，需注意两个采样率没有必要完全相同，因为数字处理本身可以更改采样率。根据著名的奈奎斯特准则，最高频率达到采样率的一半时便能被精确表达。

ADC送出的采样本身是16位深的固定点，高端系统可能将这个标准提高到20甚至24bit。在计算机中处理它们可能使用固定或浮点数表示。合理的一般化准则是，在16位输入输出的系统需要20位定点分辨率来操作和处理。

缺乏其他因素时，基本准则是n位相同的数字音频信号最好有以下动态范围（the ratio of the biggest amplitude that can be represented in the system to the smallest one）：

电话音质的语音，分辨率在相应设备上低至8-12位。GSM标准的手机上14位为标准。电话音质也叫电话相同音质，对于声音通信是可用的，但不能算是高音质。出于这个原因现代声音通信系统倾向于在实践中再增加8位的分辨率。

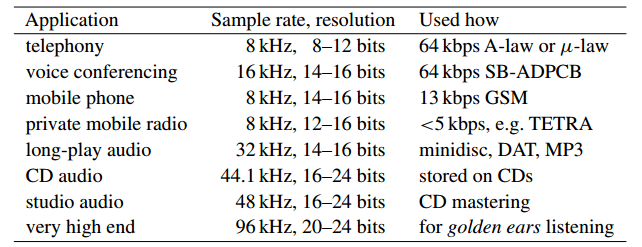
采样率从电话音质的7.2khz到8khz、cd音质的44.1khz不等。长时间播放的数字音频系统偶尔选择32khz，高音质系统选择48khz，近年也流行增加到96khz。除了对于某些金耳朵（注释1）听众，是否96khz的采样率对于通常听不见高于18khz的人耳有用是可争辩的。但这样的系统也许对宠物更友好：据说狗的听力高达44khz，猫则高达80khz。

**注释1：喜欢阀放大器的顽固音频爱好者花费多年薪水购买扬声器，这些人宣称自己能听见20khz，所以他们是“金耳朵”。**

【信息框1.1音频保真度】

需要注意的是整个转换过程是不确定的：你听到的是声波冲击耳膜，而电脑获得的声音由空气直接送达，其间可能经过一些障碍，如被麦克风拾音、激励震动膜，被转换成电信号、放大，接着被采样。放大器会产生噪音、失真，并不是完全的线性元件。麦克风则是各种意义上的糟糕。模数转换同样饱经非线性、噪音、失真和电压采样过程的精确度导致的量化误差。所有上述原因可能是你的采样结果与希望的真实录音不同，或者裸耳回放当中信号发生的细微变化是否明显将难以预测。

Table 1.1. Sampling characteristics of common applications.



常见场合的采样率及采样精度上表已列出。

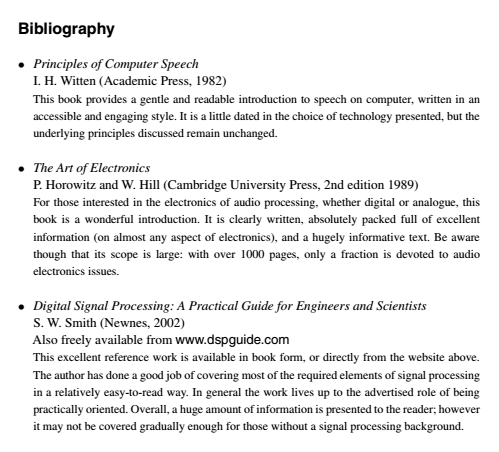
**1.4 总结**

大多关于转换和传播的过程的技术细节已经超出本书的范畴，相关内容请移步参考文献。通常来说，声音处理专家不需要了解声音的采集和回放过程就能完成自己工作。因此，我们讨论的范畴主要在信号的处理/保存/传输，以及图1.1上的信号辨别/分析以及合成/生成，而忽略模拟部分的技术细节。

对人类来说声音有一些特征：如时域的持续时间、节拍、启动时间和衰减时间等，也有频域的特征如声调（tone）和音高，以及定义不甚明确的音质、音色、音调（tonality）等等，通常声波承载了信息：如火警警报音、狮子的咆哮、婴儿的哭泣、雷声隆隆或国歌等。

然而正如我们所见，ADC采样（大多经过脉冲编码）所得的声音，仅仅是通过瞬时采样得到的幅度值来表达声音的容器。本书接下来将试图建立这些数字和声音的联系，以及试图理解、解释这些声音。

**参考文献**



**小结**

第一章没有代码，主要提及的内容是数字音频系统整体概况，兼谈少量之后会重点讨论的内容。