

本文是音频处理的朋友icoolmedia (QQ:314138065) 的投稿。对音频处理有兴趣的朋友可以通过下面的方式与他交流：

作者：icoolmedia

QQ：314138065

音视频算法讨论QQ群：374737122

原文公式较多，因此直接贴上图片。

一、混响时间的计算与预测

所谓混响就是声音的直达声与反射声很紧凑的重合在一起时人耳所听到的声音，这个效果在语音的后期处理时特别有用。能产生混响最常见的场景就是房间内，尤其是空旷的房间中

混响有直达声，早期反射和后期反射声组成。其中直达声是声源信号不经过任何障碍物直接到达人耳的那部分、早期反射声由一次或者几次反射的声音信号组成、后期反射声由随后更多次的反射声音信号组成。混响效果的空间感主要由早期反射声决定。

在一个房间中，声音衰减所消耗的时间是房间的吸声系数和声波走过的距离的函数。声波在其传输并衰减的过程中，经过每两个界面之间的平均距离就是众所周知的平均自由程，它可以通过如下公式计算得到：

$$MFP = \frac{4V}{S} \quad (1.1)$$

这里：

MFP 平均自由程，单位为米

V 房间的体积，单位为立方米

S 房间的总表面积，单位为平方米

将式（1.1）除以声速，可以得到两次反射的时间间隔为：

$$\tau = \frac{4V}{Sc} \quad (1.2)$$

假如每次反射时，都有 a 部分能量被墙壁吸收，这时 a 为墙壁的吸声系数，那么则有 $(1-a)$ 部分的声能被反射回来，并作用于下一次反射，且每一次都有 a 部分声能被吸收，因此，经过 n 次反射后，回来的声能为：

$$E_n = E_0(1-a)^n \quad (1.3)$$

由于 a 为小于 1 的系数， $(1-a)$ 也小于 1，因此式（1.3）表示声能按指数规律衰减。为了求出声能衰减到指定比例所需要的时间，就要先求出在指定时间间隔内声波发生反射的次数，而这个可以通过将这个时间间隔除以两次反射的时间间隔得到，基于式（1.2）得到：

$$n = \frac{t}{\left(\frac{4V}{Sc}\right)} = t \left(\frac{Sc}{4V}\right) \quad (1.4)$$

这里 t 为混响的总时间（单位为秒），将式（1.4）代入式（1.3）中，得到经过时间 t 后剩余的声能为：

$$E_t = E_0(1-a)^{t \left(\frac{Sc}{4V}\right)} \quad (1.5)$$

因此，经过一定的时间 t 后，声能的衰减比例为：

$$\frac{E_t}{E_0} = (1-a)^{t \left(\frac{Sc}{4V}\right)} \quad (1.6)$$

对式（1.6）两边取以 $(1-a)$ 为底的对数得：

$$\log_{(1-a)}\left(\frac{E_t}{E_0}\right) = t\left(\frac{Sc}{4V}\right) \quad (1.7)$$

因此，达到一定衰减比例所需要的时间为：

$$t = \left(\frac{4V}{Sc}\right) \log_{(1-a)}\left(\frac{E_t}{E_0}\right) \quad (1.8)$$

将以 (1-a) 为底的对数转换为自然对数（使用高中学过的换底公式）得：

$$t = \left(\frac{4V}{Sc}\right) \frac{\ln\left(\frac{E_t}{E_0}\right)}{\ln(1-a)} \quad (1.9)$$

式 (1.9) 给出了声能衰减比例与所需时间的关系，可用于计算所需的时间，即混响时间。我们可以选择无限多种比例进行计算，但最常用的是声能衰减到 60dB 的声能比，即： 10^{-6} 。将此值代入式 (1.9)，得到声能衰减到 60dB 所需的时间，即混响时间 T_{60} 为：

$$T_{60} = \left(\frac{4V}{Sc}\right) \left(\frac{\ln(10^{-6})}{\ln(1-a)}\right) = \left(\frac{V}{S \ln(1-a)}\right) \frac{4 \times (-13.82)}{344} = \frac{-0.161V}{S \ln(1-a)} \quad (1.10)$$

式 (1.10) 被称为艾润公式，式中负号与自然对数计算结果的负号相抵消，得到一个正的混响时间值。在实际使用中，也可能会遇到计算其它衰减比例的混响时间，其计算方式与式 (1.10) 相同，不同的是常数部分。

上述公式的推导基于统计学，因此式 (1.10) 的使用必须具备以后前提条件：

- (1) 声波以相同的机率从各个方向到达墙面的各个位置，即声场为扩散声场。这个条件是为了利用房间平均吸声系数的概念。
- (2) 平均自由程概念成立。这个也是利用平均吸声系数的条件。它也意味着房间的形状不能太极端，如：很长的隧道就不适合。然而，对于大多数房间，平均自由程的概念是成立的

另外，还有一个较简单的混响时间计算公式，称为赛宾 (Sabine) 公式，是以其推导者 Wallace Clement Sabine 的名字命名的。也经常使用。赛宾公式最初是通过考虑房间的平均声能损失、解简单的微分方程推导出来的，但它也可以从艾润公式推导出来，从而更清楚的看到赛宾公式的条件。艾润公式表示如下：

$$T_{60} = \frac{-0.161V}{S \ln(1-a)} \quad (1.11)$$

上式的困难在于需要计算 (1-a) 的自然对数。但是自然对数可以用以下无限级数表示：

$$T_{60} = \frac{-0.161V}{S \left(-a - \frac{a^2}{2} - \frac{a^3}{3} - \dots - \frac{a^n}{n} - \dots - \frac{a^\infty}{\infty} \right)} \quad (1.12)$$

因为 $a < 1$ ，所以上述级数总是收敛的。而且，当 $a < 0.3$ 时，级数值与 a 值的误差小于 5.7%，这时，式 (6.18) 可以近似用下式近似：

$$T_{60} = \frac{-0.161V}{S(-a)} = \frac{0.161V}{aS} \quad (1.13)$$

式 (1.13) 就是赛宾公式, 是最早出现的混响时间定量计算公式, 也常被使用。

二. 混响算法建模与实现

现在我们已经知道了如何计算混响的时间, 下面就来看下如何对混响进行建模并实现。要对房间响应进行建模, 我们假设房间内有一个向各个方向辐射的声源, 其响应将由声波走过的路径、房间的吸声系数 (取决于光滑程度等) 来决定、其中房间的吸声系数决定了混响的总体时间, 通常混响时间是指在封闭环境中, 声能衰减到 60dB 时所需要的时间。

由于每次反射都会因为墙壁、障碍物等吸收一部分能量, 所以声音信号的能量呈指数衰减。因此混响信号可以看成由直达声与许多逐步衰减、不断延迟的回声信号叠加而成, 很自然的想到可以利用等比数列求和来进行模拟混响:

$$y(n) = x(n) + ax(n-D) + a^2x(n-2D) + \dots \quad (2.1)$$

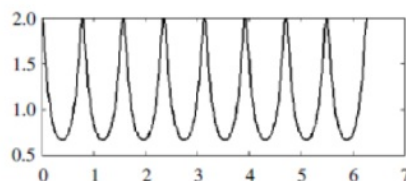
这里 $x(n)$ 为原始声音信号、 $y(n)$ 为混响信号、 a 为衰减系数、 D 为延迟时间。写成传递函数为:

$$H(z) = 1 + az^{-D} + a^2z^{-2D} + \dots \quad (2.2)$$

由等比数列求和公式, 传递函数可以转换为:

$$H(z) = \frac{1}{1 - az^{-D}} \quad (2.3)$$

熟悉信号处理的朋友可能会比较眼熟, 这不正是梳状滤波器吗? , 一点没错! 正是 IIR 梳状滤波器, 其典型的频谱特性为:



如果写成差分方程, 可表示为:

$$y(n) = ay(n-D) + x(n) \quad (2.4)$$

然而, 正如上图中所示, 梳状滤波器频谱曲线不平坦, 呈现明显的梳状效应, 从而对不同的频率成分幅度产生波动, 导致有金属声染色效应, 听起来不够自然, 另外, 单纯使用梳状滤波器, 其回声密度还是不够多。

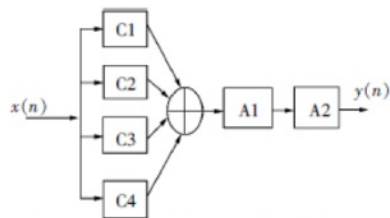
那么继续进行思考, 什么样的滤波器既有信号的延时效应, 频谱又比较平坦呢, 很多朋友应该想到了: 全通滤波器。是的, 接下来使用全通滤波器对混响模型进行改进, 它比梳状滤波器多了一个前向反馈支路, 一阶全通传递函数可以表示为:

$$H(z) = \frac{z^{-m} - g}{1 - gz^{-m}} \quad (2.5)$$

其差分方程为:

$$y(n) = -gx(n) + x(n-m) + gy(n-m) \quad (2.6)$$

最后, 让我们来看 Schroeder 发明的他的著名的混响产生模型: 4 个并联的梳状滤波器与 2 个串连的全通滤波器, 如下图所示, 在 Schroeder 混响模型中, 梳状滤波器提供了混响效果中延时较长的回声, 全通滤波器提供了较短延时的回声, 从而增加了反射回声的密度, 同时又避免了由于梳状滤波频谱不平坦造成的金属染色效应。



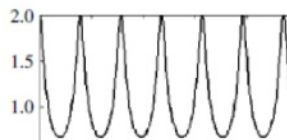
然而，从实际效果来看，如果完全按照全通滤波器来实现，回声密度还是有点不够理想，改进方法是：不完全使用全通滤波器的实现，人为的增大反馈，使回声的能量增大。这样同样的衰减因子情况下，混响持续的时间就会变长，出来的混响效果会更理想。

人为的增大反馈，带来的唯一问题是。当个别的语音能量本身比较大时，有时候会带来溢出问题，不过也不用过于担心，通过增大空气湿度的衰减系数就可以很容易调整过来，整体来讲实现效果是比较好的，这部分内容已经在群里进行过好几种方案的讨论，大家都认为实践中人为增加回声反馈效果会更好。

关于各个滤波器混响时间参数哪些配合起来听感比较自然，其实也是不用去一一进行研究的，我发现好几个混响的开源实现中，都使用相同的参数配置（忘了最初的资料是哪里了，如果有找到的望不吝赐教），据说这样听感比较自然。

我提取了这些参数，略做了一点修改和整理，写了个测试 Demo（TestReverb）。如果有感兴趣的朋友可以到[音视频算法讨论 QQ 群（374737122）](#)里下载并交流。欢迎就延时网络法和立体声混响实现进行讨论。

梳状滤波器，其典型的频谱特性为：



文章标签：[音频](#) [声学](#) [混响](#)

个人分类：[音频编码](#)

此PDF由spyyg生成,请尊重原作者版权!!!

我的邮箱:liushidc@163.com