

关于对声音系统的室内修正和均衡

马赛厄斯·约翰逊, 狄拉克研究公司 (Mathias Johansson, Dirac Research AB)

本文将讨论若干有关均衡声音系统的滤波器设计问题。讨论的重点将放在基本原理上而不是在实验上, 我将聚焦几个常见的误解。我先简单介绍声均衡所需的几个基本概念, 例如有限冲激响应(FIR)和无限冲激响应(IIR)滤波器, 最小和线性相位; 和几个基本数学论据以及《狄拉克现场修正方法》的哲学背景。为了限制本文的长度我假定对在此所谈的话题有些基本了解。但我会避免在某些工程期刊中常出现的倾向, 即把一些不好的想法隐藏在看似复杂的方程式中。诚如开篇所言, 讨论的重点将放在不同均衡法的逻辑上, 而非实验。实验总会让人质疑其实验条件, 而本文的逻辑则可任由读者检验。正如概率论所教导我们的那样: 一个幸运的实验揭示不了其中的基本原理, 但基本原理却能够更多地推测出未来的实验结果。

附注: 本文的某些部分在纽约市第 123 届音频工程学会会议上已介绍过。会议讨论的议题是“有限冲激响应还是无限冲激响应? 那就是问题!”

有限冲激响应(FIR)和无限冲激响应(IIR)滤波器

一个有限冲激响应(FIR)滤波器的输出是 N 个最近的输入采样的加权平均。因此滤波器的阶数或长度表示对滤波器的记忆长度。无限冲激响应(IIR)滤波器也是如此, 只是它把以前的输出采样的加权平均加到第一个平均。当然权重可任意设置, 但如果对应于以前的输出的权重太大, 显而易见这将会导致输出呈指数增长。这时的滤波器被称为不稳定的。在经济学上, 常用一个更生动的术语称之为“爆炸的”。一个 IIR 滤波器会制约于定点运算处理器的精度而变得不稳定。现有不同的方法实现一个高阶 IIR 滤波器以减少这些问题所带来的风险。最常见的方法是把 IIR 滤波器改写成几个串联的二阶滤波器, 称之为双二阶滤波器。用这种方法, 数字误差的影响能够得以减少。由于 FIR 滤波器本身就是非递归的, 所以不会变得不稳定, 因此也更容易实现。

由于 IIR 滤波器固有的递归性, 其冲激响应是无限(长)的。因而称之为无限冲激响应滤波器。相比之下, FIR 滤波器(其阶数小于无穷大)则有有限长的冲激响应。因此 IIR 滤波器似乎更灵活。这在某些意义上是对的, 但 IIR 滤波器有其局限性。首先, 在实际当中搭建一个线性相位 IIR 滤波器是不可能的(更多关于相位响应的概念见下文), 因为一个线性相位系统必须有一个对称的冲激响应, 也就是说, 用主脉冲做参考, 响应的左边和右边的尾部是镜像对称的版本(准确地说, 响应可能也会是逆对称的, 即一个负镜像版本)。但对于一个因果 IIR 滤波器来说, 显然不是这样, 其右边(记忆所有过去的输入)有一无限长的尾巴而其左边被截去了(它不能预见将来)。但在实际当中, 如果我们想要一个线性相位滤波器我们也能用一个 IIR 滤波器得到一个很好的近似。

下面一个问题有其不同的特点。当人们按照最小化(滤波器响应和所需响应之间的)均方误差的准则设计一个优化的 IIR 滤波器时, 数学式会变得很“难看”。这是一个非线性问题, 找到其正确的响应在数学上是很难地。虽然“很难”并不意味着“不可能”。但这或多或少地为很多“懒汉”数学家们提供了远离 IIR 滤波器的借口(但这借口在用一个具有有限能力处理器来设计滤波器时是有道理的)。

那么为什么最简单的音频均衡器仍用 IIR 滤波器? 一个原因是滤波器往往通过手工配置, 并且有闭型公式用于构建几个简单的滤波器类型, 比如峰式/陷波滤波器, 低通/高通滤波器, 倾斜型滤波器, 等等。你能容易地实时调节 Q 值, 然后听其差别。不仅如此, IIR 滤波器与 FIR 滤波器相比还具有另一优点: 即在给定的低通滤波器的斜率下需要相对较少的系数。其原因就是 IIR 滤波器具有的递归性, 这就像让几个滤波器接连地为同一个目标工作。

FIR 滤波器更容易实现（能在定点结构上运行），能更灵活地进行数学操作，可以有任意相位响应，但对一给定的滤波器的陡度，FIR 滤波器则需更多系数。这一点对于低频范围的分辨率来说是非常重要的。直观地说，FIR 滤波器的长度与其处理的信号的波长有关。但只要可以采用具有足够长度的 FIR 滤波器，许多人就会更主张用 FIR 滤波器而不是在数值计算上更复杂的 IIR 滤波器。

总而言之，在一个很好的近似范围内，我们可用 IIR 滤波器和 FIR 滤波器达到同样的目的。选择那一种滤波器更多地与其实现方式有关，并在某种程度上与其应用有关。于是为了对一组声传输函数进行适当的冲激响应校正，则需要一个较长的滤波器，而用 IIR 滤波器似乎得不到更多益处。从另一方面来说，如果我们处理器的能力仅够一个低阶滤波器（比如 10 个双二阶滤波器或 20-40 个抽头的 FIR 滤波器），那么冲激响应校正就不可能进行。此时更明智之举是把注意力放在幅度响应上和使用最小相位双二阶滤波器。

最后应该指出，如果冲激响应校正要求一个非因果部分，那么 IIR 模型就不能较好地模拟了。而 IIR 滤波器对于设计得很好的均衡滤波器的因果部分来说可以作为一个好模型，这意味着最有效率的混合相位均衡器的实现方法之一是 IIR 和 FIR 滤波器的结合。

关于相位的概念

一个滤波器的相位响应表示一个纯正弦波的输入在输出时其周期位移了多少。因此它是对频率变量延迟的一种表示方法。

对群延迟的误解

相位响应对频率的导数叫群延迟。它有时被误认为是在某一给定频率的延迟。事实上群延迟只是测量以那个频率为中心的窄带信号的包络（即形状）。把它解释成一个物理延迟是没有意义的。为了得到在某一频率的准确延迟，我们只要给相位响应一新的刻度，把角度量变成秒（如果我们更喜欢，也可以用米）。

例如：36 度相移指的是 1/10 周期的延迟。如果所说的频率是 1000 赫兹，一个周期就是 1 毫秒，那么 1/10 周期的延迟就是 0.1 毫秒。

回到相位来

对于一个线性相位响应系统，其相位响应作为频率的函数（在线性频率刻度时）是一直线。这意味着该系统在所有的频率上引入了一固定延迟。但如上所说，一个线性相位滤波器总有一个对称（或反对称）的冲激响应。这意味着如果这样一个滤波器的幅度响应不平坦，那么它将会在时域中引入前振荡和后振荡。其危险性有时会被在该领域的工作人员中被夸大。在实际中，如果一线性滤波器的幅度响应是平滑的（一均衡滤波器应该如此），其前振荡则可忽略不计。

现在，我们也必须讨论一下最小相位系统的概念。这是一个非常特殊的系统，它在所有的具有给定幅度响应的系统中具有最小能量延迟（在复平面中，一个最小相位系统的所有零极点都在单位圆内）。但必须强调，这绝不是指其冲激响应是所有的具有给定幅度响应的系统中是最短的，而是指它是最早的。这个差别是很大的。“它是早的”对于需要保证延时最小的应用来说是一个有用的特征。对于声音质量而言，没有证据显示最小相位均衡器听起来比线性相位均衡器更好。事实上，线性相位滤波器的冲激响应一般比最小相位滤波器的短。“最小相位”这个术语容易误导人，而如果在音频界采用“最小能量延迟”这个术语的话就会好很多。确切的定义如下：考虑一个最小相位冲激响应和一个其他任意的具有相同幅度响应的线性系统，比较在任意时刻 t 上两个系统输出的累积能量，最小相位系统的（从时间零到 t 的）累积能量总比其他系统的大。这意味着其响应必须开始得早，否则我们总能建立另一系统，其在时间 t 接

近零时比最小相位滤波器传递更多的能量。注意：最小相位系统，只是延时小于其他滤波器，而并不意味着零延时。

根据众所周知的复分析中的定律阐述，任何线性时不变系统可以因式分解成两个系统：一最小相位系统和一全通系统。如果我们设计一个能完全可逆的最小相位系统的均衡器，就可以使得幅度响应变得平坦（这或许不是我们的最终目的，但我们可以暂且这样假设，因为在此情况下这不会导致任何差别）。这时冲激响应发生了什么事情呢？有时我也会听到能干的人说：我们至少没有把它弄差，因为我们关心最小相位部分（最小相位系统的逆转本身还是一个最小相位系统。因为对任何一个幅度响应只有一个最小相位部分，所以我们已经自动地考虑了它）。这个叙述的谬误是显然的：全通系统的冲激响应可能比整个系统的冲激响应差很多。如果你拿掉最小相位因子你可能会去掉在全通因子中抑制严重振荡的特征。我们所选择的因式分解方式完全是任意的，没有任何物理解释。我们只是选择了一个不同的分解方式。

我们需要做的是看一看总的系统响应。如果我们因式分解它，我们最好是保证所用的因式分解具有物理声学或心理声学意义，并能根据其总响应评估最终结果。一个有很好物理动机的常用模型就是把室内不同位置的传递函数模拟成有一共同极点集和一个只有零点组成的位置变量传递函数。到现在为止，一直都还不错。下一步就是说：行，我们要进行鲁棒校正，因此只需要反转极点。误差就在这里产生了：即使它们（共同极点）对所有位置都的确相同，但抛开与位置相关的零点来看，它们将是没有意义的。如果你只是去掉在低频区域靠近单位圆的那些极点，这并不意味着你已经改善总体响应。首先，那些极点影响的并不只是它们碰巧最靠近的单位圆区域。它们也有助于减小在远离那个区域的频率的增益。类似地，靠近单位圆高频范围的那些与位置相关的零点会提升低频范围的能量。

从以上两个例子中可学到的是：即使某一因式分解有物理意义或是便于数学考虑，但如果把每一个因子看成一个孤立的系统且认为只要能改进那个因子就会改善整个系统，那么这将是危险的愚蠢行为。但是正如基本逻辑教给我们的：一个假的前提意味着任何一个命题，即假的或真的。因此你也许能给出这样一些例子，它们有时听起来很好但基本原理有纰漏，但这并不意味着，有一个更好的基本原理你不能做得更好。

傅里叶变换和频率的概念

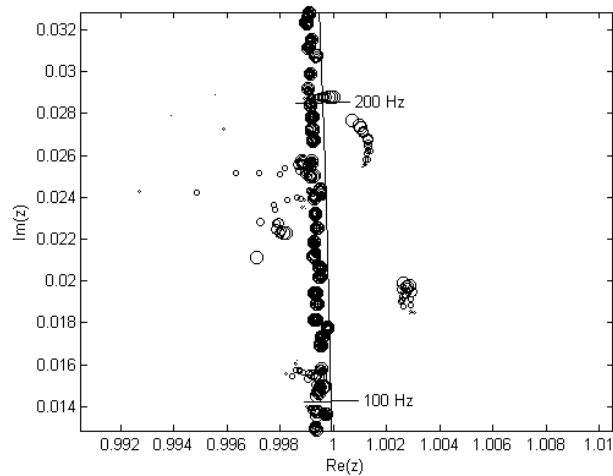
正如丹尼斯·盖伯（Dennis Gabor）于 1945 年所注意到的，一个信号的傅里叶的表达与人对频率的感觉有明显的不同。傅里叶变换是对一个时间序列在一个无限时间窗口上进行积分。这会带来一些有趣的结果。例如，一个变化的频率就是一个矛盾措辞，按定义来说这是不可能的事情，因为时间已经排除在方程式以外。因此汽笛就有一个看似很有趣的傅里叶频谱，但这不容易表明其“本质”。如果我们根据盖伯框架，或短时傅里叶变换，或其他时频基底来分析时间序列，我们就能容易区分在某一时刻的某一频率，在连续地变化音调。这种时频表示法也能更好地描述人类的听觉。对于人来说，频率是与时间相关的。傅里叶表示法和知觉印象之间有明显差别。

这对声均衡有某些重要的含义。当我们读到幅度响应的估计值时，我们用一个非常简单的方法估计被感觉的频谱，完全不考虑时间概念。比如，取一最小相位冲激响应然后将其反转。前者在时刻 0 开始，然后随某一时间常数衰减，直至其消失。而后者有大的前振荡但没有后振荡。所有对人们感觉瞬态的研究表明，前回声和后回声听起来完全不同，但它们的幅度响应却是相同的。当然从这一点也同样可看到，如果我们说一件事，然后以相反的方向回放此事。这两个样本都有相同的傅里叶幅度响应。这表明相位响应或者冲击响应的确影响对声音的感觉，甚至对非立体声源也是如此。显然存在一个我们对相位有多敏感的阈值。但有关这方面的文献（从 20 世纪 30 年代起，已有很多）已得出结论，这个阈值或积分时间常数是自适应的并随我们所听的东西的不同而不同。我们所能肯定的是，我们的确可以听出绝对相位，但频率越高我们就越不敏感（波传播物理学指出，用声学的方法在高频时只能发射很少的相关信息，原因是高频声传输函数的混沌特性）。这意味着一个好的均衡器也应考虑相位而不仅仅幅度。

现在用一个例子来更清楚说明这一点。考虑在一房间里有一个扬声器。A 先生测量了某一收听容积中的冲激响应，丧气地发现在所有的空间位置上其幅度响应在某一低频处（比如说是 300 赫兹）有一很宽的衰落。他先校准一峰值滤波器，然后去弥补该幅度响应的衰落。这被以后的测量所证实是对的。此时 B 先生进来了。B 先生是一位音乐家，他来听被均衡的系统。他说：“这听起来很糟糕。你们对这个系统做了什么？！听起来很奇怪！”A 先生有点紧张不安起来，因为 B 先生是一位重要的顾客，于是 A 先生赶紧给他的一个朋友 C 先生。C 先生回答说：“啊，当然是的。衰落确实是因为反射造成的。你永远不要去提升任何衰落。因为他们一般是由反射造成的。”于是 A 先生拆掉了他的均衡器，让 B 先生再来听。但 B 先生仍然不满意，说：“好一些。但还不够好。这声音有一些虚。”这时 D 女士进入交谈。她一直静静地坐在角落里听着。她说：“A 先生错了，因为他忘记了时域。他只是看一看傅里叶变换的幅度然后解释其与我们的频率概念极为相关，便认为他能提升这个区域而得到更好的声音。问题在于他用了最小相位滤波器，因此早些时间在那个频率增加了能量。但如果我们只看直达波，在频响中就没有洞需要填满。如果我们看任何时刻的一个窄时窗，这个洞根本就不存在。”B 先生皱皱眉头地说：“C 先生说地对，我拿它没办法。但如果是这么回事，为什么在我的录音中仍然听到一个声音奇怪双簧管？”D 女士认真地看着他并说：“C 先生也不对。这问题的确是因为时域性质引起的，也就是由是反射引起这个问题。这只能用时域的方法来校正。如果我们设计一个减小反射的滤波器，你最终就会得到一个有趣的结果，即这个洞会消失、双簧管听起来会更自然。”“但是，”D 女士补充道，“不要把这个例子作为证明，说明你总能用此方法校正这些衰落！只是在当前这种情况下是可能的，因为所有的位置都有同样的问题。”

当论及一个好的声学幅度响应应该是什么样子的的时候，不加批判地认为傅里叶变换的幅度是对我们的感觉（并因此忽视时间）可以准确描述显然是种误解。其答案取决于被测房间的时域特性。一个在大房间听起来好的幅度响应可能在汽车内听起来很糟，因为它们的混响比完全不同。

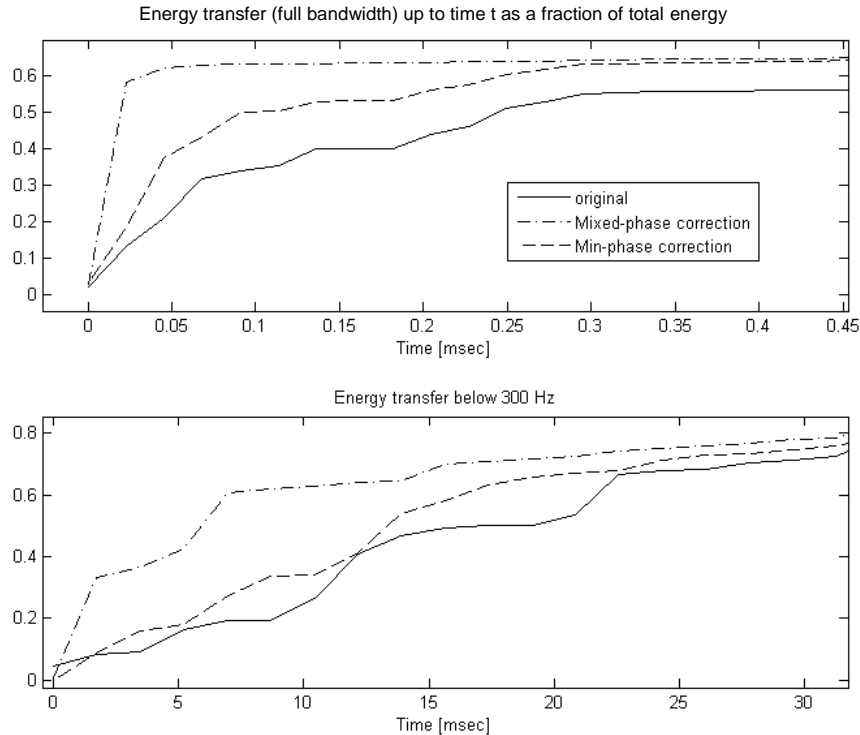
提升衰落的基本问题是：通常用一个最小相位均衡器来实现。因此，这在一个错误的时间注入能量。这对一个线性相位滤波器或任何一个其他的没有考虑整个系统脉冲响应是什么样的滤波器来说同样也是对的。教训是：不要把傅里叶变换和被感觉的频率响应混起来。被感觉的频率响应是与时间有关的。因此设计一个好的均衡器需要进行时频联合分析。除此以外，随空间变化的问题必须认真地考虑。在某一测量点上一个零点可能正好在单位圆内；但移动麦克风，它就可能会跑到在单位圆外去。在这种情况下，观察其平均效果（一个零点在单位圆上）然后优化那个响应可能会很糟糕。一个适合平均响应的校正非常不同于在总体上适合任何被测响应的校正。位于单位圆外但在近似同样地点（不管麦克风的位置如何）的一些零点是鲁棒可逆的，而其它的就不是这样（再请注意：这并不意味我们考虑一个与响应的其他部分相隔离的零点）。下面放大的零点图显示了这些特点。测量是在一个好的收听室的不同位置用很好的 Genelec 扬声器进行地。图中所画的零点采用了不同的尺寸，用来对应不同的测量位置，其目的是为了便于看这些零点怎样随空间的变化而移动。



图中横轴- $\text{Re}(z)$ 和纵轴- $\text{Im}(z)$ 分别表示 z 的实部和虚部。

就在 200 Hz 以上我们可以看到一个零点从最小相位向混合相位移动。其平均值正好在单位圆内，于是粗心地进行最小相位逆转可能会在此导致许多严重的振荡。在 150 和 200Hz 之间我们可以看到一个零点在单位圆外，它随位置变化地很多。因为它靠近单位圆并且在移动，所以鲁棒校正似乎不会改变这一特性。但是在 100 和 150Hz 之间位于单位圆外的一零点是有可能被去除而不会在任何位置引起任何问题（到单位圆的距离相当大并随位置的变化小）。

基于这些测量，设计了两种不同的均衡器。下图显示对于三个系统（原系统、用最小相位校正的系统和混合相位校正的系统）的脉冲响应，该图描述了被转换的总累积能量随时间的变化。该图是根据在不同收听位置进行新的测量而不是根据用于设计滤波器的原来测量数据来画的。应该强调的是，最小相位校正基本上是采用与混合相位逆转相同的鲁棒原则。其中上图显示的是全频带的结果，下图显示的是响应到 300Hz 的结果。这些基本上是能量阶跃响应。因此一个更好的系统就会有一个更快的上升时间，理想的情况是立刻达到 1。很清楚，即便是在一个本身具有优质扬声器的高级收听室，采用细致的混合相位设计仍能得到实质性的性能改善。最小相位滤波器虽不及混合相位设计，但它能改善时域性能，所以显然也能胜任。在一个大的、设计得很好的收听室中，其冲激响应近似最小相位更加合适；这对前面的收听室同样如此。因此我们只要用一些普通的最小相位滤波器就能得到改善。不幸地是这不能用于像汽车厢这样较难处理的环境。最后应提到的是：使用混合相位滤波器所引起的前震荡，其最高幅度应低于冲激响应中峰值的 60dB。（再请注意：其他收听位置用真实的测量数据。）因此，我们能放心地得出结论：混合相位逆转是用的，它能改善音响系统的性能且不会损失鲁棒性。



到时间 t 的能量转换（全频带）在总能量中所占分数。图中横轴为时间（毫秒）和纵轴为总累积能量被转换率（被转换的能量在总能量中所占分数）。

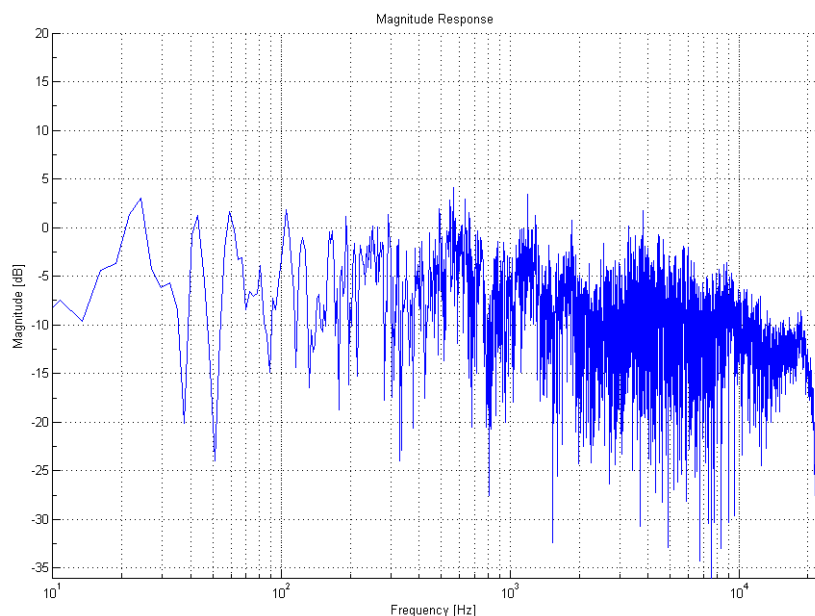
我们已经阐释：与最小相位滤波器相比，混合相位均衡器在增加鲁棒性的同时能改善原先就已较好的状况。即使是在这个具体例子中被感觉的差异对于一个一般的听众来说不是很大，这个例子的重要性在于它反映了这样一个事实：混合相位均衡器比最小相位均衡器更鲁棒（因为最小相位滤波器没有考虑整个脉冲响应）。同时这也显示脉冲响应的衰减时间能通过避开最小相位响应来减小。当然其唯一的缺点是存在延时。但如果这不是问题，那么我推荐使用混合相位逆转法，毕竟它是一个较为安全的方法。（作为一个旁注：在以上的例子所用的方法中，一个参数控制前振荡的量，其与后振荡的量相关。因此，如果收听测试显示人们对于直到-30dB 的前振荡不敏感，那么我们就能够达到一个更好的系统响应。）

真实的立体声再现

立体声再现是基于这样一个事实，即人类来自不同声源的声音的相关性具有敏感性。即使在一个房间里有两个真实的声源，我们在某些时候能感觉一个声源位于两个真实声源中间。如果这两个声源播放同样的声音并且与我们相隔同样的距离，一个在左、一个在右，我们会感觉一个虚拟的声源在这两个真实声源的中间。当这两个声源不是完全相关，其空间印象就会被改变。如果一个声源的声音被延迟足够多，那么我们甚至会听不到那个声音；所有声音会被感觉是来自较早的声源。这是有大量文献显示的优先效应。优先效应的一个有趣方面是，感觉不仅取决延时而且取决两个声源的空间间隔。比如说，如果两个声源位于一条直线上，听者在该直线的延伸线上，这样对于两个声源的收听角度是相同的，所听到的声音是“有色的”。如果使用同样的延时但两扬声器从水平方向分开，就没有有色化的感觉，我们听到的只是第一个声源。但是，如果我们放一个全向麦克风在听者的位置，以上两个实验设置会给出两个完全相同的录音结果（假设房间有良好的衰减性能），两个录音听起来都很糟。第二声源与第一相干。取两个录音的傅里叶变换的幅值我们会发现两者完全相同，都呈梳状；许多深的波谷遍布频谱。我们会再次发现傅里叶变换不是一个表达我们听觉的好方法。其主要原因是：基本的傅里叶变换不会区分不同的入射角。

现在我们考虑在收听室中一个传统的立体声音响的设置。如果两个声源有不同的传递函数，那么它们就会改变左右声道信号的相关性，于是到耳朵的信号也就不再会以录音时的方式而相关。如果左右声道信号相差特别大，它们就有可能完全解相关。在这种情况下，即使是左右声道信号相等，我们也能听出两个截然不同的声源，而不是预想的位于它们中间的一个虚拟声源。例如，隔音场/电影棚的声场会是漫反射的，如果我们对左右喇叭各自的传输函数进行一个合理的冲激响应校正，那么我们就从它们的录音中重建原来的相关特性并且使得隔音场/电影棚中的声场会变得清晰和相干。遗憾的是，在上述使用高阶滤波器的房间校正系统中，相反的情形有时也会发生。这些校正系统无意中使声场解相关，而不是改善左右声道信号的相关性。

原因是一个高阶滤波器有一个长冲激响应，该响应需要合理地控制。如果我们基于一个（对在不同收听位置测量的原来的传递函数进行）过分简单化的平均模型搭建一个逆转滤波器，我们最终会校正与感知不一致（在那个模型尾部）的这些问题。考虑下面的图，该图是基于一个性能较差、但较大的收听室，使用了非常好的扬声器，在 12 个不同的位置（对应于一个沙发的两个不同的座位）的测量。图示的是一个傅里叶幅度响应，它是基于简单的“延迟和平均”响应，即在傅里叶变换之前，所有的脉冲响应按时间对齐，然后进行平均（也即波束成形）。

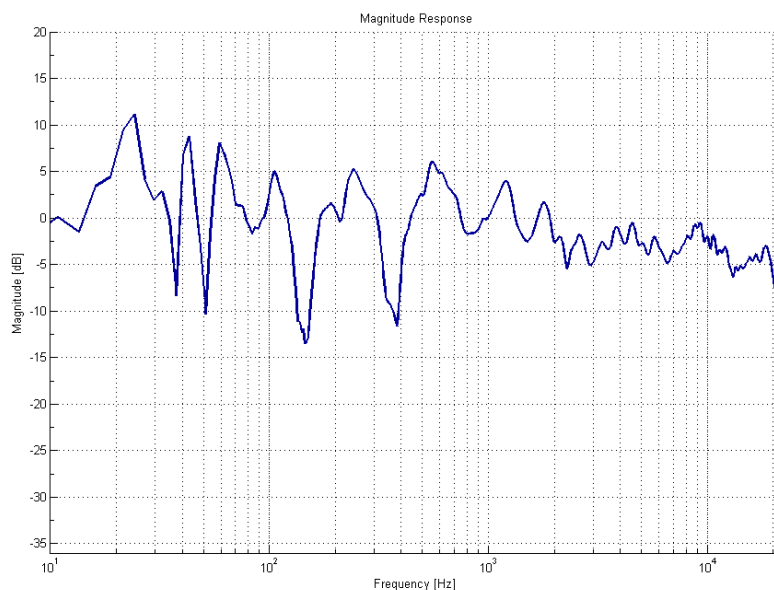


幅值响应。图中横轴-Frequency (Hz)和纵轴-Magnitude (dB)分别表示频率和幅度。

注意纹波的大小，特别是在高频区域（如果我们用分倍频程平滑，我们就不会看到这些纹波（如果我们不细心），但它们仍然存在，而且会影响最终的滤波器）。纹波的问题在于，它不是我们所能感觉的频率响应的特性，而只是傅里叶频率响应的特性。傅里叶响应既没考虑空间角度也没考虑时间。纹波是听不见的，因为它只是在房间中后期反射的特性（能被认为是“后期的”当然是与频率相关的）。假设我们现在搭建一个逆转傅里叶变换的均衡器。不管它是不是最小相位，它在高频区域仍然有相似的变化性，这在时域中表现为冲激响应中的附加的高频垃圾（随时间传播开）。如果我们现在对左右喇叭使用不同的均衡器，我们最终会有两个截然不同的传输函数，它们会产生任意的空间效应，通常是弱的解相关的形式。幼稚的平均法致使我们去补偿（a）在任何单一的收听位置不存在的东西和（b）在心理声学上不相关的东西。非常值得注意的是：校正某个听不见的东西可能会导致某个东西可听见。

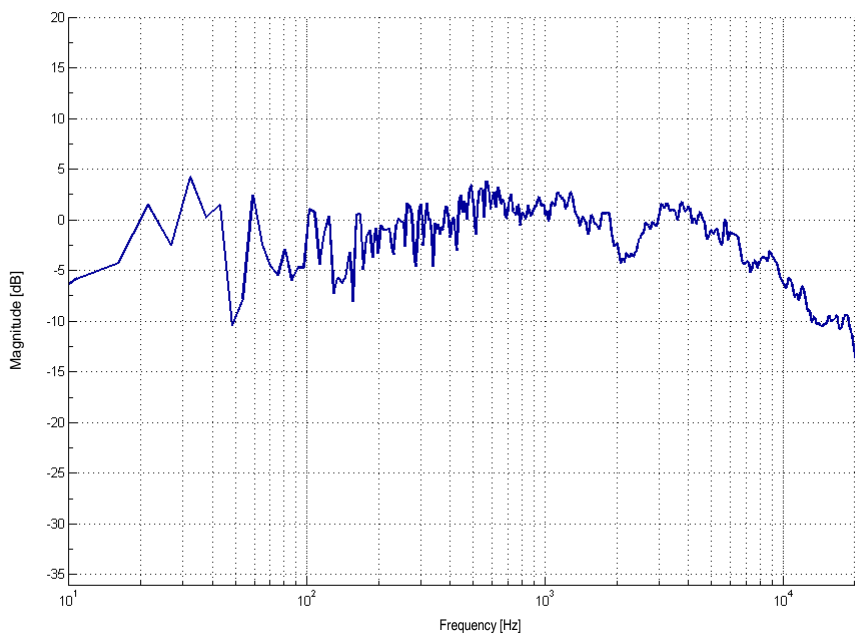
基于我们用来制作上图的同样的 12 个测量，我们用不同的类型的时频分析来进行一些对与心理声学更相关的表示方法的估计。下面两张图显示两种不同的表示方法；两者都在频域

中，但在心理声学上与两个不同的目的相关。第一种表示方法采用了一种处理，它设法效仿人对混响室内的瞬态声音的感觉。我们可以看到高频纹波消失了但有一些相当强的衰落点出现在1kHz一下。出于某种简化，我们已经去除了那些与直达波不同方向的后期反射。



幅值响应。图中横轴-Frequency (Hz)和纵轴-Magnitude (dB)分别表示频率和幅度。

但此图并没有适当地描述出我们在听某一演奏（比如弦乐四重奏）时所感觉的东西。后期反射仍然加进了声染色效应，它影响着我们对稳态和声信号的听觉感受。因为这个原因，一个图能给我们对稳态频域声染色效应的理解有一个合理的估计，该图看上去就像如下描述的。



幅值响应。图中横轴-Frequency (Hz)和纵轴-Magnitude (dB)分别表示频率和幅度。

我认为：以上二图都有意义地描述我们在收听室中实际感觉到的。前者描述我们怎样感觉瞬态声，后者描述我们怎样感觉稳态和声信号。也就是说，我们即不应该基于前者进行幅度响应校正，也不应该基于后者进行冲激响应校正。

在用混和相位校正时所得的一些重大改进在于立体声和多声道的再现。在左右声道传递函数不同时，空间信息就失真了。因为房间响应是非最小相位的，我们需要混合相位逆转从而重建正确空间图像。合适的混合相位逆转与最小相位逆转之间的差别在糟糕的收听空间中会变得更大；比如说在汽车中，其左右声道的传输函数就有些不同。应该注意，声音质量的差别通常很大。虽然差别小的时候一定存在（比如说当主要的房间响应的变化可以用最小相位滤波器来校正时），但通常的差别都很可观。然而，如前所警示的，进行一个合适的冲激响应校正是不容易的，因为人们很容易以一个解相关滤波器（和/或前回声）而了之。但这是可以被发觉的：在收听测试中，空间图像变得模糊（与清晰相对），声音被感觉是来自两个喇叭之间的某个区域，这是相位校正变坏的清楚迹象；画一个高分辨率的滤波器频率响应，滤波器在高频区域不应该有任何快速变化，这是一个设法校正冲激响应中后期尾部特性的清楚迹象。幅度响应中的快速变化表示相位响应中有类似的快速变化；一个非常不太可能的事是，对房间响应的相位校正应该在高频应该是有用的（因为根据波的传播规律，在某一物理位置的校正一定会在离该位置几个厘米远处就变坏）。

最后的建议是，混合相位校正是绝对有用的，并且在我的心目中是一条改进声源重现的正确的前进方向，但是它比去阅读过去二十年关于冲激响应校正的研究要困难得多，这值得人们思考。

有些反射好，有些不好

在这个领域似乎有一个共识，某些早期反射实际上有助于使演讲更容易理解。但是，也有大量文献显示，在通常的收听室中，主要脉冲在 5-10ms 内的反射会高于主声源移动或传播的声级（即使在听一个单独的主声源时）。从前面和后面（ $\pm 40^\circ$ 以内）的反射感觉起来是不利于声音质量的，然而来自侧面的反射（在合理的声级以内）常常会改善被感觉的声音质量。

这可以从信息论的观点来理解。来自前面的反射一般很难与主声源区分开，原因在于我们的耳朵的位置。一个单独的前方反射的传递函数在左耳和右耳几乎是一样的，就像主声源的响应。的确，这就是为什么一个滤波器能够鲁棒地校正这些反射的原因。它们在相当大的收听空间中是恒定的。这一点与来自侧面的反射形成对比。这些（侧面的）反射会随位置变化很大，原因是入射角。在左耳和右耳之间总存在很大差别，因此侧面反射能带来分集增益。分集是信息论的概念，它使可分辨通信信道的数量量化。分集越高，信息传递的香农容量就越高。这在移动无线通信系统中被用到；在该系统中，独立的直接路径反射实际上被用来增加无线通信系统的比特率。有一个令人感兴趣的推论，它反映出前方和后方反射对人听觉的不利影响。因为这些信道不能跟主声源区分开（两者几乎是常数），所以这些信道就会造成自相干扰。在容量的表达式中，这直接解释成较低的信噪比和较低的信息传递容量。

因为在此所谈的影响会降低声源质量，它相对位置来说是近似不变的，所以这意味着，这些也是一个非因果混合相位滤波器能够校正的影响。类似地，携带信息的反射是不可能被改善的，就是因为它们携带信息的原因是它们随位置而变化。如果它们在两只耳朵总是一样，就没有办法让我们从反射中分离出声源，所以声音的逼真度就会降低。这的确是一个令人感兴趣的事实，是因为渐进而可以预料的，人的听觉系统是有能力利用房间提供给我们的附加信息。同时我们都会直观地知道，这种推理取决于这样的假设，即反射实际上携带可分辨的和附加的信息。当然这只是一个相当特别的情况，显然大多数的反射只会降低声音的质量。

总结和建议

FIR 和 IIR 滤波器有几乎相同的可能性。在许多情况下，如果调节正确，最小相位均衡能够达到相当不错的效果。但是为了提高空间清晰度和达到真正好的声音质量，必须对整个系统

响应仔细评估。为达到最佳性能，混合相位均衡是必需的。在不同的收听环境中（比如汽车）用混合相位逆转法能达到很明显的不同效果。但是进行混合相位逆转，会引入延时，所需的滤波器的长度一般会相当长。在延时必须最小或滤波器阶数受限（比如在 44.1kHz 的采样频率小于 100 阶）的应用中，我的建议是采用合理设计的最小相位 IIR 均衡器。即使如此，人们也必须认真，设法达到针对位置的鲁棒设计。

如果你的 DSP 预算允许你使用高价滤波器，我的建议是采用仔细的混合相位设计。一个设计得好的混合相位滤波器能提供快于最小相位或线性相位滤波器所能达到的系统响应。注意：“仔细”一词是很重要的。一些设计高阶滤波器（不管是最小相位或是混合相位）的算法总有困难的时候，例如试图摆脱在脉冲响应尾部加进一个高频纹波。这对立体声的感觉有负面影响（即虚拟声源传播）。如果算法设法做的太多，就会引入前回声。简言之，混合相位设计从性能和鲁棒方面来说有更好的潜力，但此潜力更难达到（但是对于超低音箱声道，最小相位逆转一般就够了，因为房间与超低音箱相结合的传递函数在这些低频范围可以很好地用一个最小相位系统来模拟）。

一个混合相位设计可以用一个 FIR 或一个 IIR 滤波器来实现，但最有效的实现方法一般是两者的结合。

如上所述，一个合适的混合相位滤波器设计类似于去掉扬声器附近的反射面。此外，它在喇叭本身的设计中使线性缺陷最小化。因此数字混合相位滤波器在设计音响系统和房间时，对通常的电声性考虑是一个强有力的、划算的互补。

致谢

本文仅代表作者本人的观点。但它确实是多年来在狄拉克研究公司内部讨论中所得到的见解的积累。它包含有几个人贡献。拉尔斯-约翰·布然马克（Lars-Johan Brännmark）提供了本文中所用的实验数据。我与他及狄拉克其他人的讨论极大地影响和丰富了我对该问题的理解。