# 基于声源分离技术的虚拟低音增强算法

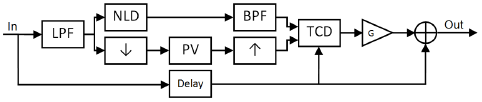
## 1 引言

随着移动多媒体便携设备变得更小、更薄、更轻，嵌入在这些设备中的扬声器也必须做到更小更薄，但由于小型扬声器的成本和体积限制，他们的低频响应特性不会做到很出色，但是低音效果是衡量一个多媒体设备听音感受的一个很重要的因素，所以如何解决这个矛盾是目前音效增强领域的一个研究热点问题。传统增强低音效果的方法是利用均衡器对低频段信号进行增益控制，以达到增加低频信号能量的目的。但是这种方法往往会带来糟糕的低音效果和声音失真，甚至会对扬声器的振膜造成不可逆转的损坏。

近年来，基于心理声学[1]的虚拟低音增强技术被很多研究者所热衷，它被认为是一种增强小型扬声器低音效果的有效方法。基音缺失是心理声学理论中的一种听觉错位现象，它是指人耳在声音基频缺失的情况下，利用基频信号的谐波组合来重建基频信号。虚拟低音增强技术中重要的一点就是如何产生谐波信号，目前常见的谐波生成方法可分为时域和频域两大类。时域法包括利用乘法器加反馈回路产生谐波的MaxxBass技术和基于非线性器件(Nonlinear Devices, NLD)的谐波生成技术。频域法主要是利用相位声码器进行频谱搬移来产生高次谐波。时域法的优点是算法简单，处理速度快有利于实时性，但是会产生非线性畸变和互调失真，而且时域法适合处理音频信号中的瞬态部分，如鼓声。频域法优点是失真小，谐波幅度控制精确，但是计算量和延迟较大，存在瞬态模糊和混响效应等缺陷，适合处理音频信号中的稳态部分，如乐器声。在分析了以上两种方法的优缺点后，一种混合虚拟低音增强算法被提出来[9]，这种算法结合了NLD和PV两种谐波生成算法，并使用了瞬态成分检测器（TCD）让系统在NLD和PV之间切换，来达到减少是真的目的。但是这种系统中的瞬态成分检测器的检测效果比较差，导致生成的低音信号仍然有失真，而且整个系统比较复杂，算法复杂度高实时性差，不适用于移动平台。

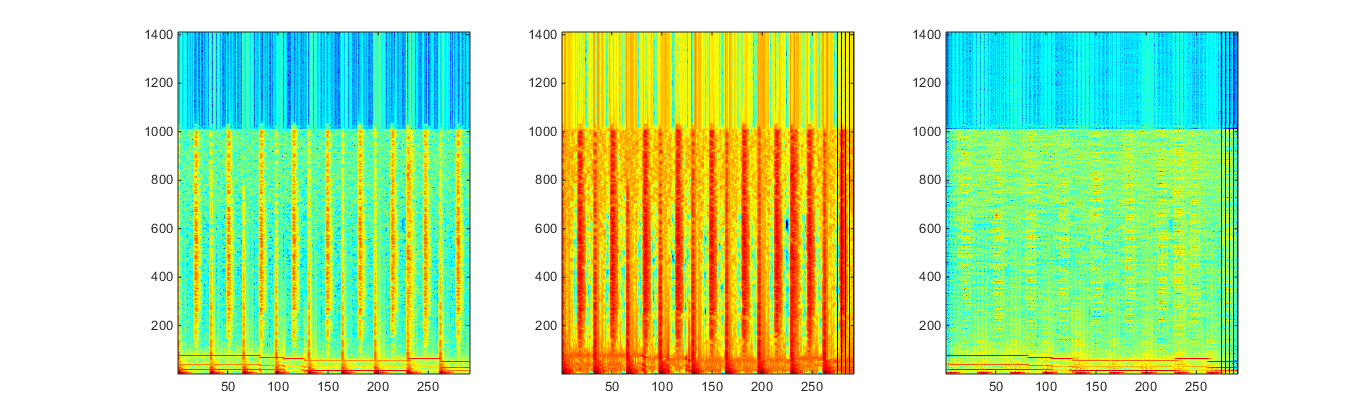
我们在分析现有的VBS算法后发现，传统的VBS算法是将未经处理的原始音频信号直接输入到谐波生成模块中，而原始音频信号是稳态和瞬态的混合信号，由于谐波生成算法对NLD和PV对瞬态和稳态信号的响应是不同的，所以将这种混合信号直接输入到谐波生成模块里会给生成的虚拟低音信号带来一定程度的失真。而且，我们对传统VBS算法生成信号的低音效果进行主观测试后发现，虽然NLD算法会给稳态信号带来一定的交叉调制失真，但是其生成的低音效果要远好于PV算法。这是因为一段音乐中的低音效果大部分来源于爵士鼓、贝斯等瞬态信号，而NLD算法更适合于处理瞬态信号，所以本文提出的算法的基本思想就是：利用声源分离算法对原始音频信号进行稳态、瞬态分离，再利用NLD算法对瞬态信号进行处理，最后将生成的虚拟低音信号与原始信号进行叠加，输出低音增强后的信号。实验结果表明，本文提出的虚拟低音生成算法的低音效果较传统方法更好，失真更小，并且处理速度也有很大提高，更适用于移动便携平台。

## 2 声源分离算法



在混合虚拟低音增强算法被Adam.J Hill第一次提出来时，他并没有利用声源分离算法对原始音频信号进行处理，而是设计了一个瞬态成分检测器（TCD），对NLD和PV生成的谐波信号进行检测，根据检测到的音频信号的成分不同，分配不同的权重进行叠加，最终的效果就是当某部分检测到的稳态成分比较多时，那么该部分所分配的叠加权重就比较大。但是通过该算法的系统框图可以看出，原始音频信号通过低通滤波器处理后，瞬态和稳态成分依然混合在一起，再通过NLD和PV处理后，两个谐波生成器产生的干扰信号依旧没有减少，只是在最后叠加之前进行了权重分配。这样所造成的结果就是干扰信号依旧存在，并且原始音频中稳态和瞬态成分的比重和叠加权重并不是简单的线性关系，所以合成出的低音效果并不是很好。

声源分离算法被广泛应用于音乐信息检索（MIR）中，包括和弦估计、主旋律提取和节奏模式识别等。其中有一类算法可以实现音频信号稳态成分和瞬态成分的分离，这正符合我们对虚拟低音增强算法的改进需求。稳态瞬态分离算法大致可以分为三种类型，第一种是利用信号盲分离算法来实现，比如ICA和NMF。第二种是利用节奏特征匹配来检测瞬态信号，例如MFCC和LPCC等音频特征。这两种算法的分离准确度还是很高的，但是缺点就是计算量大，实时性差。第三种是数据集合分解算法，也就是本系统用到的分离算法，这种算法把音频信号的时频谱看作一个数据集合，利用稳态信号和瞬态信号在时频谱中表现出来的不同特点，把信号分离问题转化为一个简单的最优化问题，并利用拉格朗日乘数法对最优化问题进行求解，最终得到了一种分离效果好并且计算量小，实时性好的稳态瞬态分离算法。



如图所示是一段人工合成的音乐信号的时频谱图，横轴代表时间，纵轴代表频率，颜色深度代表此点的数值大小。左图是单独的鼓声信号，右图是单独的乐音信号，中间是两者合成后的信号。从左图和右图可以看出，瞬态信号（鼓声）和稳态信号（乐音）在时频谱中的表现是不同的，瞬态信号由于其脉冲特性，在时频谱图中的频率轴（纵轴）方向表现为连续，而稳态信号由于其频率稳定性，在时频谱图中的时间轴（横轴）表现为连续。我们将混合信号的时频谱（STFT变换）记为，幅度谱记为，和分别代表STFT的时间窗口数和单个时间窗内的频率点数；稳态信号的幅度谱记为，瞬态信号的幅度谱记为。对于上述的稳态信号和瞬态信号在时频谱中的特点，我们可以用下面的公式来表示：

## 3 系统分析

## 4 仿真实验

## 5 结论