基于声源分离技术的虚拟低音增强算法

张三三，李四四，王　五

（上海大学 通信与信息工程学院，上海200072）

**摘要：**中文摘要200字左右，应包括目的、方法、结果和结论等要素。

**关键词：**（小五号宋体）关键词需3~8个；关键词一；关键词二；关键词三

**中图分类号：**　　　　　　　　　**文献标志码：**A

**Virtual Bass System Based on Rhythm Content** **Extracting**

ZHANG San-san, LI Si-si, WANG Wu

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

**Abstract:** Bass effect is an important criterion for audio system. However, the small loudspeakers in portable devices have poor low frequency responses. Conventional methods to enhance the bass effect using equalizers does not help significantly and may result in distortion or permanent damage to the loudspeakers. Recently, the virtual bass system (VBS) based on the psychoacoustic phenomenon called “missing fundamental” has been proposed, whereby human auditory system can perceive the fundamental frequency from its higher harmonics. Nonlinear devices (NLD) and phase vocoder (PV) are commonly used to generate harmonics in VBS. Yet, both approaches have their strength and weakness. In this paper, we proposed a virtual bass system by extracting the rhythm content to improve the bass effect as well as maintain the audio quality. Experiments show that the bass effect has been improved significantly with maintaining the good audio quality and the processing efficiency has increased too.

**Key words**: key

**1** 引言

随着移动多媒体设备变得更小、更薄、更轻，放置在这些设备中的扬声器也必须做到更小更薄。由于这些小型扬声器的成本和体积限制，使得它们不能够有很好的低频响应特性，但是低音效果又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。传统的处理低音效果不足的方法是利用均衡器，通过对低频段信号的增益控制来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的失真，甚至会对扬声器的振膜带来不可逆转的损坏。利用心理声学原理[1]的虚拟低音增强技术是一种增强小型扬声器低音重放品质的有效方法。

基音缺失(missing fundamental)现象是指人耳在基频缺失的情况下可以利用基频信号的谐波组合重建基频信号。虚拟低音增强技术按照谐波的产生方法可分为时域和频域两大类。时域法包括利用乘法器反馈产生谐波的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。在分析以上两种方法的优缺点后，一种混合虚拟低音算法被提出来[9]，这种算法结合了NLD和PV两种谐波生成算法，利用瞬态检测器（TCD）让谐波生成算法在NLD和PV之间切换，达到减少失真的目的。但是这种瞬态检测器对瞬态稳态信号的检测效果比较差，导致生成的低音仍然有失真，而且系统比较复杂，实时性较差，不适合于移动终端处理。例如，一段一分钟的音乐在PC机（Win 7, CPU: 3.4G, RAM: 4GB）上需要2.38分钟才可以处理完成，极大地限制了使用范围。

我们在分析现有的VBS算法后发现，传统的VBS算法只是将原始音频信号不经过处理，直接输入到谐波生成器中，而原始音频信号中是混有稳态信号和瞬态信号的，直接输入到谐波生成器肯定会带来一定程度的失真。而且，对传统VBS算法生成的低音效果进行主观测试后发现，虽然NLD算法会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。因为音乐中的低音大部分来源于架子鼓、贝斯等瞬态信号，而NLD算法对处理瞬态信号有优势，所以本文提出的算法就是在NLD算法基础上进行的。本文提出的算法利用声源分离算法将原始信号中的瞬态稳态信号分离，再利用瞬态信号和NLD算法生成虚拟低音信号，最后将虚拟低音信号与原始信号进行叠加，达到增强低音效果的目的。实验表明，本文提出的算法生成的低音效果饱满，失真小，而且处理速度也有很大提高。

基于以上讨论，本文提出了一种基于声源分离的虚拟低音增强算法，在保证低音效果的同时大幅降低了处理时间，提高了实时性。本文的安排如下，第一部分简要介绍了研究背景；第二部分简单介绍了一下虚拟低音算法的理论背景；第三部分详细讲解了所提出的虚拟低音增强算法；第四部分给出了实验和仿真结果；最后一部分进行总结。

**2 理论背景**

**2.1 基音缺失原理**

人耳的听觉系统对单音的主观感受是靠基频、振幅和频谱这三个因素来决定的，分别对应于心理声学的三个要素：音调、响度和音色。音调是这个单音信号给人的主观感觉，而一个单音的音色，是人脑对单音频谱（即各谐波成分比例）的主观感觉和判断。当几个纯音叠加时，比如100 200 300三个纯音同时存在，音调通常被识别为最低的100。也就是说当基频和它的谐波同时存在时，通常认为基频是这个音的音调。假设一个单音的频谱成分为f，2f，3f，4f，将基频f去掉后，音调仍被识别为f。这个被心理声学识别出来的f被称为“缺失的基音”，这种现象称为“基音缺失”。

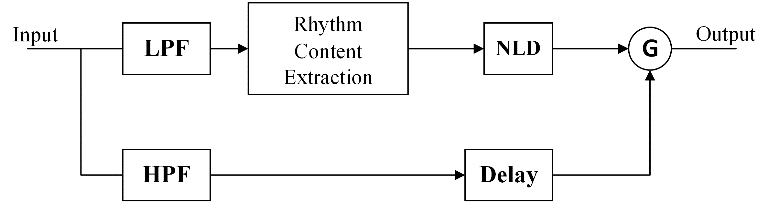
利用“基音缺失”现象，对输入到扬声器的信号进行处理，减少基音，增加谐波成分，利用扬声器在谐波频段的重放能力较强的优势，可以从主观听觉上改善扬声器的低频表现力，这就是虚拟低音技术的基本原理。

**2.2 现有VBS算法**

在基音缺失原理的基础上，目前的VBS算法主要是利用非线性器件（NLD）[2-5]和相位声码器（PV）[8]来产生低音增强所需要的谐波，分别是从时域上和频域上对信号进行处理。MaxxBass[6]算法是最早投入市场的虚拟低音增强技术，它采用了基于非线性电路的乘法器元件生成低频成分的高次谐波。虽然这种方法结构简单、处理速度较快，但是存在着较为明显的交叉调制（IM）失真，在处理乐器声等稳态信号时表现较差，影响音频音质，所以NLD方法比较适合于处理鼓声等瞬态信号。为解决此问题，台湾学者Mingsian R.Bai提出一种基于相位声码器的虚拟低音算法，这种算法采用数字信号处理时频分析中的变调方法生成谐波，采用参数化的等响度曲线精细控制生成谐波的能量。相对于非线性器件算法，PV算法消除了非线性畸变，在等响度控制方面也较为精确。但是，PV算法在时域上需要足够大的分析窗口来满足频域分辨率，这将会导致算法复杂度增大，实时性受影响，而且在处理鼓声等瞬态信号时会产生相位失真，比较适合处理稳态声音信号。

在分析以上两种方法的优缺点后，一种混合虚拟低音算法被提出来[9]，这种算法结合了NLD和PV两种谐波生成算法，利用瞬态检测器（TCD）让谐波生成算法在NLD和PV之间切换，达到减少失真的目的，但是这种瞬态检测器对瞬态稳态信号的检测效果比较差，导致生成的低音仍然有失真，而且系统比较复杂，实时性较差，不适合于移动终端处理。例如，一段一分钟的音乐在PC机（Win 7, CPU: 3.4G, RAM: 4GB）上需要2.38分钟才可以处理完成，极大地限制了使用范围。

**3 系统详解**

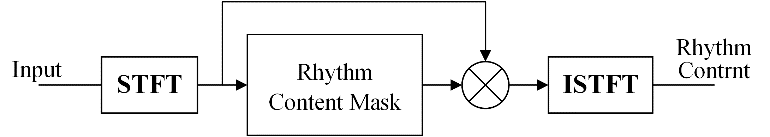


图一 系统框图

本文所提出的系统框图如图所示，首先将原始音频信号分别通过低通滤波和高通滤波处理，将低通信号通过节奏信息提取算法得到音频信号的节奏信号，通过NLD生成低音的谐波信号，再通过幅度控制后与高频滤波信号经延迟后的信号进行叠加，输出虚拟低音增强信号。

**3.1 低通滤波器和高通滤波器**

由于虚拟低音增强只需要对扬声器截止频率一下的低频信号进行谐波生成处理，所以输入的音频信号需要先通过低通滤波器滤除高频信号，本算法选用了截止频率为120Hz的四阶椭圆滤波器。经高频滤波器处理后的高频信号需要经过延迟处理后，与虚拟低音信号进行叠加。



图二 节奏信息提取算法

**3.2 节奏信息提取算法**

由于

**3.3 非线性器件**

根据文献[4]中对不同非线性器件在谐波生成中的效果的分析，反正切平方根（Arc Tangent Square Root， ATSR）所生成的虚拟低音效果最好，所以本算法选用ATSR作为谐波生成器。

**3.4 谐波幅度控制**

谐波信号生成以后，必须经过适当的幅度控制来达到合适的音调和响度。传统的PV算法中将等响度曲线参数化来实现谐波幅度控制的，这种方法非常耗时。MaxxBass算法中介绍了一种响度分析算法来控制每个谐波的幅度，这个响度分析器是基于响度扩展比率（SPL-to-PHON）。这种方法与PV算法中的幅度控制方法有相同的效果而且更高效。

基于响度扩展比率是频率f的函数，可由式给出：



第l阶谐波的响度扩展比率如下式所示：



基频和谐波的能量之比满足下式：



由此我们可以推得：



上式中是l阶谐波的幅度，是基频f的幅度。

文献[10]指出，在原始音频上添加虚拟低音效果后会影响音乐的品质，例如由9式生成的谐波信号。通过心理听觉实验，文献[15]研究了谐波的相对强度对虚拟低音音色的影响。实验表明，当各谐波成分的相对幅度按照的指数衰减曲线（x是谐波次数）进行幅度控制时，所达到的虚拟低音音色最佳。基于这个研究，本文所用的幅度控制策略是先将最低次谐波的幅度计算出来，然后根据下式计算其他各次谐波的幅度

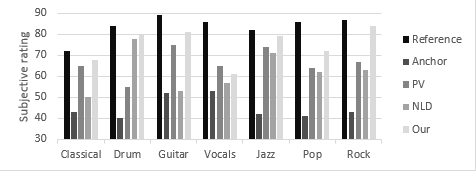


由于这种方法只需通过MaxxBass方法计算出最低次谐波的幅度，其他各次谐波通过式10计算出来，这样会减少系统的复杂度，提高性能。

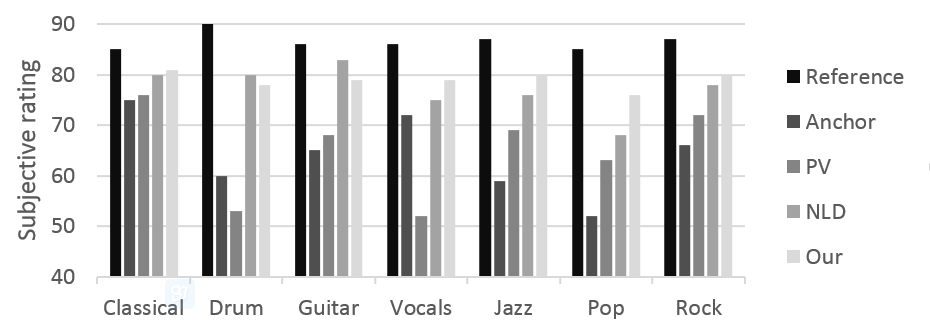
**4 实验仿真**

“多激励隐参考”（Multiple Stimuli with Hidden Reference and Anchor, MUSHRA）主观测试法[17]是2003年ITU组织发布的一种对音频质量进行判断的主观评价标准。本文使用这种方法对所提出的虚拟低音系统与传统的虚拟低音系统进行对比，并对实验结果进行详细说明。

作为一种音效增强系统，该系统应该对各种音乐类型都有一致的增强效果，所以我们选取了7种不同音乐类型的WAV文件并都截取了30秒作为参考音频（Reference）。经过120Hz高通滤波处理的音频文件作为“锚”（Anchor），测试音频也就是激励（Stimuli）是经过NLD、PV和本文所提算法处理的7种不同类型的音频文件，分别是古典、鼓声、吉他、人声、爵士、流行和摇滚。7个经过简单听力训练并事先不了解具体音频处理过程的被试者，分别在安静的房间里独自对7种不同类型的音频文件进行主观评价，并根据音频的音色、失真度和低音效果综合进行打分，分别有Excellent、Good、Fair、Poor和Bad五种级别，以百分制分别计为80~100分、60~80分、40~60分、20~40分和0~20分，最后统计每种音乐类型的平均得分。实验是在PC平台进行，具体配置为Win 7, CPU: 3.4GHz, RAM: 4GB，测试用耳机为Bose OE2帖耳式耳机，外置声卡为REM FIREFACE UC USB高清音频声卡。



1. Average subjective ratings for bass effect.



1. Average subjective ratings for audio quality.

实验结果如图所示，从图一可以看出，本文所提系统在大多数音乐类型的低音表现上要好于传统的PV和NLD算法，而在人声音乐类型上，传统的PV算法的低音效果要好于本文所提系统，这是因为人声类型的音乐包含丰富的稳态信号，瞬态信号相对匮乏，所以比较适合于PV算法处理。图二所示为处理后的音频的音质对比图，从图中可以看出，作为“锚”的音频信号得分最低。这是因为高通滤波会滤除一些有用信号，这样不但使得音频的低音效果几乎全部没有，而且也会使音质有损失。从图一图二可以发现，传统的PV算法在处理不同音乐类型的音频时，所得到的低音效果和音质都是得分最低的，这是由于传统的PV算法在产生谐波的时候同时会带来相位畸变和混响效应，这些都会时音频信号的音色和音调发生变化，而音色和音调又是人耳判断音频质量的重要因素；同时也可以看出，本文所提算法在低音效果和音频质量两个方面表现都十分优异，但是与真实低音的感觉还有一定差距。

不同虚拟低音算法的处理时间如下表所示，由于本文所提系统中使用了处理时间短的HPSS算法，所以整个系统的处理时间大大缩短。一段60秒的音频片段在PC(Win 7, CPU: 3.4GHz, RAM: 4GB)上的处理时间只有36s，相比于传统VBS算法有了50%左右的提高，更适合于移动终端的虚拟低音增强。

1. computation time of different VBS.

|  |  |  |  |
| --- | --- | --- | --- |
| Approach | PV | NLD | Our |
| Time (s) | 96 | 57 | 36 |

**5 总结**

本文提出了一种新型的虚拟低音增强算法，在分析了原始音频信号的频谱信息后，通过HPSS算法将瞬态稳态信号分离，然后将瞬态信号经过NLD生成虚拟低音信号，最后将信号合并起来。根据主观音频评价结果可以看出，本文所提算法在低音增强效果、音频质量保真度和处理速度等方面要优于传统虚拟低音增强算法。