

文章编号:1002-8684(2009)03-0059-04

EVB Phase Vocoder 算法与 MaxxBass 算法在 虚拟低音增强应用中的比较

·论文·

郭 杰, 莫福源, 蔡志博, 李晓东
(中国科学院 声学研究所, 北京 100190)

【摘 要】介绍了虚拟低音增强实现中的 MaxxBass 算法与 EVB Phase Vocoder 算法。通过对比分析了基于时频分析的虚拟低音算法相对于传统非线性算法在算法灵活性、低音效果还原和合成信号畸变控制等方面的优势。实验证明, EVB Phase Vocoder 算法合成的虚拟低音增强效果明显优于 MaxxBass 算法。

【关键词】虚拟低音; 虚拟音调; MaxxBass 算法; 相位声码器; EVB Phase Vocoder 算法

【中图分类号】 TN912

【文献标识码】 A

Comparison of Applications of EVB Phase Vocoder Algorithm and MaxxBass Algorithm in Virtual Bass Enhancement

GUO Jie, MO Fu-yuan, CAI Zhi-bo, LI Xiao-dong

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)

【Abstract】 The applications of MaxxBass algorithm and EVB Phase Vocoder algorithm in virtual bass enhancement are introduced. By comparison, the superiority of time-frequency processing approach over the traditional non-linear algorithms is presented in the aspects of flexibility, bass impression and distortion control. Objective experiments and subjective tests indicate that the EVB Phase Vocoder system achieves better results for bass enhancement than the MaxxBass system.

【Key words】 virtual bass; virtual pitch; MaxxBass algorithm; phase vocoder; EVB Phase Vocoder algorithm

1 引言

低频频带扩展一直是电声器件设计的热点问题; 低频重放能力也是衡量扬声器性能的重要指标之一。近年来, 随着 PC, 笔记本计算机, 智能手机, 掌上游戏机以及 MP3, MP4 等便携视音频设备的普及, 小型扬声器以其体积小、功耗低、成本低廉的优点得到广泛应用, 但其低频重放能力差也成了困扰用户以及产品设计者的典型问题。由于受到单元尺寸等物理条件的限制, 小型扬声器对低频声音信号还原能力很弱甚至没有。采用均衡电路直接增强信号的低频能量通常会导致扬声器效率降低、重放信号产生畸变, 严重时甚至可能损坏扬声器系统。虚拟低音技术则针对这一问题提供了另一种解决方法。

在心理声学实验中, 存在着一种被称为“虚拟音调”^[1]的实验现象: 对于 1 段和谐复音信号, 其基频成分决定了信号音调的高低; 然而, 通过某些手段将复音信号的基频部分除去后, 余下的各次谐波的叠加依然能

使人感受到相同的音调高低, 即人耳能在基频缺失情况下利用谐波组合重建信号音调高低。虚拟低音技术逆向应用了该现象。对于低于扬声器有效频率范围下限 f_L 的某个频率 f_0 , 人为构造谐波 $kf_0, (k+1)f_0, \dots, (k+N)f_0$, 其中 $kf_0 > f_L$, 叠加谐波信号可获得与 f_0 相同的音调感知, 增强低音。实验证明, 当 k 满足 $(k-1)f_0 < f_L$ 且 $kf_0 > f_L$, 同时, N 取 3 可获得较好的合成效果, 即选取频率大于 f_L 的连续 3 个最小阶次谐波。

现有算法中, 谐波的构造主要依靠非线性方法实现, 参考文献[2]中对此类方法作了详尽介绍。其中, Daniel Ben-Tzur 等人提出的 MaxxBass 算法^[3]最具代表性, 并发展成为成功的商用系统; Meir Shashoua 和 Daniel Glotter 给出了该算法的一个具体实现^[4]: 利用 1 个乘法回路产生谐波信号并以等响度曲线为基础引入声压级-响度扩展比率 (SPL-to-Phon Expansion Ratio) 以保证合成信号与原信号在响度音色等方面的一致性; 另一方面, 针对非线性方法固有的非线性畸变问

题,白明宪将数字信号处理中的时频分析算法引入虚拟低音系统,提出了基于相位声码器的虚拟低音增强算法^[3],该算法通过短时FFT分析信号低频成分,在此基础上计算各频率通道的瞬时频率并构造谐波相位。同时算法利用参数化的等响度曲线,精细计算了合成谐波每个频率通道能量与对应的基频通道能量比用来保证响度及音色还原。但是相位声码器音调变换技术利用瞬时频率合成谐波相位也会将该算法典型的“Phasiness”失真引入合成信号造成信号畸变。

笔者在这一算法的基础上改进了相位合成方法,同时加入峰值匹配计算并简化能量增益计算过程,提出了一种改进的基于相位声码器的虚拟低音增强算法 EVB Phase Vocoder。实验证明,EVB Phase Vocoder 算法相比 Maxxbass 算法在低音听感和畸变控制方面有明显提高。

2 Maxxbass 算法

2.1 谐波产生

图1给出了1个简单的乘法器谐波产生电路,设输入信号 $x(t)$ 为频率为 f_0 的纯音信号,即

$$x(t) = \sin(2\pi f_0 t) \quad (1)$$

设 g_1, g_2, g_3, g_4 为线路增益; x_2, x_3, x_4 分别为乘法器输出信号,于是有

$$x_2 = \frac{g_1 g_2}{2} [1 - \cos(2 \times 2\pi f_0 t)] \quad (2)$$

依次计算 x_3, x_4 , 系统的最终输出 $y(t)$ 为

$$y(t) = h_0 + \sum_{i=1}^2 \{h_{2i-1} \sin[(2i-1) \times 2\pi f_0 t] + h_{2i} \cos(2i \times 2\pi f_0 t)\} \quad (3)$$

式中, $h_0 = \frac{g_1 g_2}{2} [1 + \frac{1}{4} g_3 g_4]$, $h_1 = g_1 [1 + \frac{3}{4} g_2 g_3]$, $h_2 = -\frac{g_1 g_2}{2} [1 + g_3 g_4]$, $h_3 = -\frac{g_1 g_2 g_3}{4}$, $h_4 = \frac{g_1 g_2 g_3 g_4}{8}$ 。由式(3)可知, $y(t)$ 包含了输入信号的2,3,4次谐波,滤去直流和基频信号即可得到虚拟低音信号。上述结果是在输入为纯音信号的前提下推得的,对于复音信号输入,不同频率间的互调干扰会给输入带来一定程度的非线性畸变。

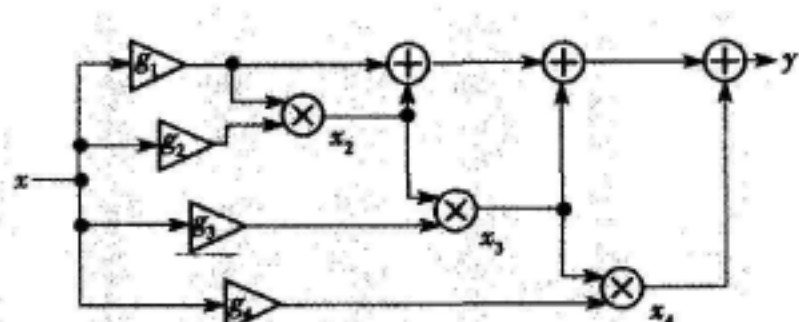


图1 乘法器谐波产生电路

2.2 谐波幅度控制

虚拟音调理论只保证了谐波叠加所得的虚拟低音信号与原低频信号的音调保持一致,要得到理想的低音效果还原还需要考虑合成信号响度与音色同原信号的一致性。图2为等响度曲线。

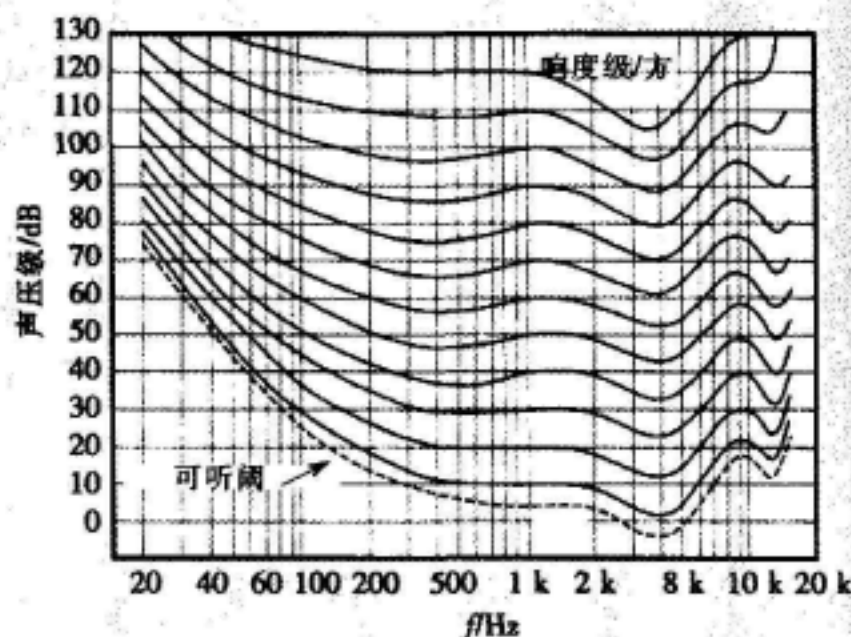


图2 等响度曲线

由图2可以看出,在不同的频率点,相同的声压级变化引起的响度变化是不同的。响度变化量与声压级变化量的比值 R 随频率的变化改变。例如在40 Hz频率点上,声压级下降18 dB,响度大约下降60方;而在80 Hz处,降低同样的声压级,响度约下降38方。 R 被定义为声压级-响度扩展比率(SPL-to-Phon Expansion Ratio)。在响度20~80方,频率20~700 Hz范围内, R 可近似表示为^[4]

$$R(f) = \frac{1.0}{\ln(f) \cdot 0.241 - 0.579} \quad (4)$$

保持谐波信号响度对声压级的动态特性与基频处一致对有效还原虚拟低音信号的响度与音色特征十分重要,设^[4]

$$RR(f, n) = \frac{R(f)}{R(n \cdot f)} = 1 + \ln(f) \cdot 0.241 \cdot R(f) \quad (5)$$

式中, n 为谐波阶次。谐波信号能量与基频信号能量应满足关系^[4]

$$E_n = RR(f, n) \cdot E_f + K \quad (6)$$

式中, E_n, E_f 分别为倍频点和基频点的信号能量,以分贝为单位; K 为增益常量。将式(6)转化为线性坐标的能量关系得

$$X_n^2 = 10 \lg \frac{K}{10} \cdot (X_f^2)^{RR(f, n)} = g \cdot (X_f^2)^{RR(f, n)} \quad (7)$$

式中, X_n, X_f 分别为倍频点和基频点的信号幅度。

2.3 MaxxBass 算法实现

MaxxBass 算法利用图3所示的乘法回路生成无

限阶次的谐波信号, 并利用衰减增益 g 消除高次谐波的影响。高通滤波器 OUT-HPF 和 FB-HPF 用来减小直流干扰和降低系统的互调畸变。这里 OUT-HPF 采用 4 阶椭圆高通滤波器, 截止频率为 $f_2=150$ Hz, FB-HPF 采用 2 阶巴特沃思高通滤波器, 截止频率为 $f_1=40$ Hz。

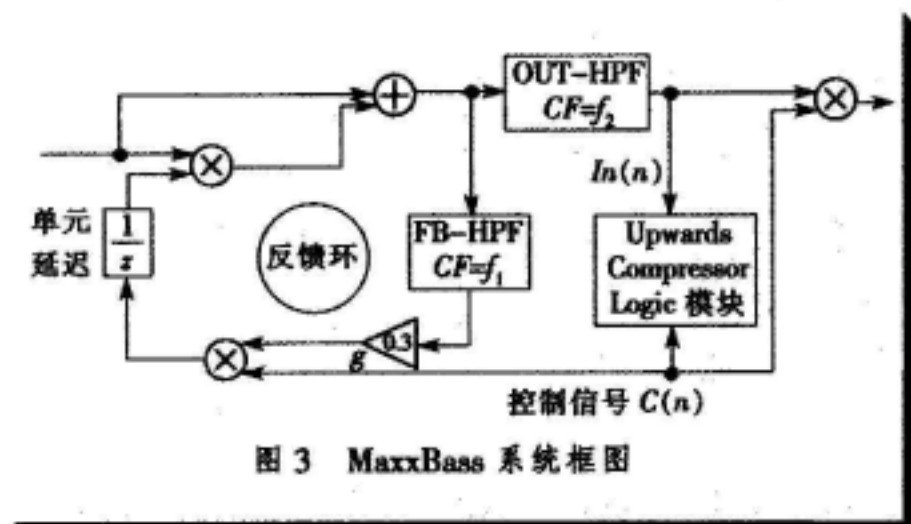


图3 MaxxBass 系统框图

Upwards Compressor Logic 模块通过生成控制信号 $C(n)$ 实现谐波幅度控制, 以保证响度对声压级在各次谐波处的动态平衡。对于输入 $In(n)$, 该模块产生的相应的控制信号为^[9]

$$C(n) = \frac{b_0 + b_1 \cdot In^2(n) + b_2 \cdot In^4(n)}{a_0 + a_1 \cdot In^2(n) + a_2 \cdot In^4(n)} \quad (8)$$

式中, $b_0, b_1, b_2, a_0, a_1, a_2$ 为 Upwards Compressor Logic 模块决定的多项式系数。这一方法只能满足对式(6)的近似实现, 系统中 $RR(f, n)$ 被近似为以阶次 n 为变量与频率 f 无关的函数 $\tilde{RR}(f, n)$, 将相应的参数代入式(8), 计算得^[9]

$$\begin{cases} RR(f, 2) \approx \tilde{RR}(2) = 1.34 \\ RR(f, 3) \approx \tilde{RR}(3) = 1.72 \\ RR(f, 4) \approx \tilde{RR}(4) = 2.01 \end{cases} \quad (9)$$

系统输入为 $[f_1, f_2]$ 范围内的基频信号, 输出为虚拟低音信号。将输出信号与原信号频率在 f_2 以上的高频成分叠加即可得到虚拟低音增强信号。

3 EVB Phase Vocoder 算法

3.1 Phase Vocoder 与 Pitch Shifting

Phase Vocoder 可通过时频分析方法实现信号的音调变换 (pitch shifting)^[10], 信号的短时加窗傅里叶变换可以写为

$$X(n, k) = \sum_{m=-\infty}^{+\infty} x(m)h(n-m)e^{-j2\pi mk/N} = |X(n, k)|e^{j\varphi(n, k)} \quad (10)$$

在此基础上, 信号可以写为 1 组幅度和相位随时

间变化的正弦信号的叠加

$$\begin{aligned} X(n) &= \frac{X(n, 0) + X(n, N/2)}{N} + \sum_{k=1}^{N/2-1} \frac{2}{N} |X(n, k)| \cdot \\ &\quad \cos[\varphi(n, k) + \frac{2\pi kn}{N}] = \\ &= \frac{X(n, 0) + X(n, N/2)}{N} + \sum_{k=1}^{N/2-1} \frac{2}{N} |X(n, k)| \cdot \\ &\quad \cos[\varphi(n, k) + \Omega_k n] = \\ &= A_0(n) + \sum_{k=1}^{N/2-1} A(n, k) \cdot \cos[\theta(n, k)] \end{aligned} \quad (11)$$

改变各频率通道的幅度与相位参数即可实现对信号的相应变换。对于音调变换, 设帧移为 R_s , 由相邻 2 帧的相位值可求得频率通道 k 的瞬时频率为

$$d\theta(k) = \Delta\theta(k)/R_s \quad (12)$$

式中, $\Delta\theta(k)$ 为解卷绕后的相位差, 将所求得的瞬时频率乘以音调变换系数 α 累加, 便可得到音调变换后的信号相位

$$\psi(n+1, k) = \psi(n, k) + \alpha \cdot d\theta(k) \quad (13)$$

将各正弦分量叠加即可得到音调变换后的信号

$$y(n) = A_0(n) + \sum_{k=1}^{N/2-1} A(n, k) \cos[\psi(n, k)] \quad (14)$$

设 $[f_1, f_2]$ 为需要增强的低频范围, f_1 为计算方便取 40 Hz, f_2 取扬声器有效频率范围下限。若 $x(n)$ 只包含 $[f_1, f_2]$ 范围内的低频成分, 对于不同频率点 f , 设整数 l_f 满足 $(l_f-1)f < f_2$ 且 $l_f f > f_2$, 分别取 $\alpha = l_f, l_f+1, l_f+2$, 即可实现由基频信号构造 l_f, l_f+1, l_f+2 次谐波, 将谐波叠加便得到虚拟低音信号。对于不同频率点 f , 整数 l_f 满足 $(l_f-1)f < f_2$ 且 $l_f f > f_2$ 。

实际应用中, 用瞬时频率构造谐波相位, 当 α 取值大于 2 时, 容易产生由相位畸变引起的信号失真。针对这一问题, EVB Phase Vocoder 算法直接选取时频分析后相频曲线上低频通道倍频点处的相位作为谐波相位, 从而保证了合成相位的连续性。这一做法也等同于按照信号低频部分能量对相应倍频处的能量进行提升以达到虚拟低音增强的目的。

3.2 峰值匹配

实验表明, 虚拟音调的形成并不要求参与合成的高次谐波频率与基频频率满足严格的整数倍关系。在倍频点约 5% 的频率范围均可看作是能够进行虚拟音调合成的谐波频率。例如 740 Hz, 750 Hz 及 760 Hz 都可被视作 150 Hz 频点的 5 次谐波。相比而言, 人耳对信号峰值的位置分布更加敏感, 倘若在信号幅度谱上与信号低频峰值对应的倍频点为 1 谷点, 增强该点的

能量有可能破坏合成信号与原信号峰值位置的一致性而产生可感知的畸变,因而有必要在合成前加入峰值匹配。在 $[f_1, f_2]$ 范围内搜索峰值所在位置 f_p ,以 $l \cdot f_p$ 为中心在 $\pm 5\%$ 的范围内搜索幅度最大值所在的位置 $l \cdot f_p + \text{offset}$, offset 为峰值偏移量。对于基频通道 f ,取 $l \cdot f + \text{offset}$ 处的相位为对应的 l 次谐波相位,同样的方法可求得 $l+1, l+2$ 次谐波相位。峰值匹配保证了在倍频点附近的峰值处进行能量提升。

3.3 谐波能量控制

在对每一帧信号进行FFT时频分析后,依照式(7)对 $[f_1, f_2]$ 范围内各频率通道的幅度进行调整,以实现谐波信号的响度-声压级动态特性与基频信号保持一致。相对MaxxBass算法,这里可实现对每个频率点的精确调整。

4 实验对比

为验证算法的有效性,笔者针对MaxxBass和EVB Phase Vocoder进行了客观频谱分析与主观测听对比。

4.1 频谱对照

图4和图5分别给出了23 ms长的音乐信号经MaxxBass算法和EVB Phase Vocoder算法处理前后的频谱对比。为便于观察,图4与图5中,横坐标取频率的对数坐标,纵坐标取幅度归一化值的20倍对数值。从图4与图5中可看出,虽然对倍频处能量补偿方法不同,MaxxBass与EVB Phase Vocoder算法的共同点是算法处理后的信号150 Hz以下的低频部分被大幅度衰减,而150~900 Hz倍频部分得到相应的提升。由此可看出虚拟低音增强算法得到的低音听感是通过提升倍频能量获得的,而非直接对低频部分进行增强,因此可应用于低频重放能力差的小型扬声器。同时,由图4中可以看出,MaxxBass算法处理后的信号高频部分的能量也被提升,这是由于算法引入的非线性畸变引起的。

4.2 主观测听对比

虚拟低音算法是以心理声学模型为基础的,所以有必要对算法结果进行主观测听实验从而验证算法的有效性。

实验选取了8段低频成分丰富的信号作为测试样本,其中包括鼓声、交响乐、低音贝司、直升机螺旋桨声等。11名年龄在20~30岁,听力正常的人员参加了测试,测试扬声器的低频截止频率为150 Hz。测试者以未经处理的原始文件为参考分别对MaxxBass及EVB Phase Vocoder算法处理后的文件进行主观评分。评分

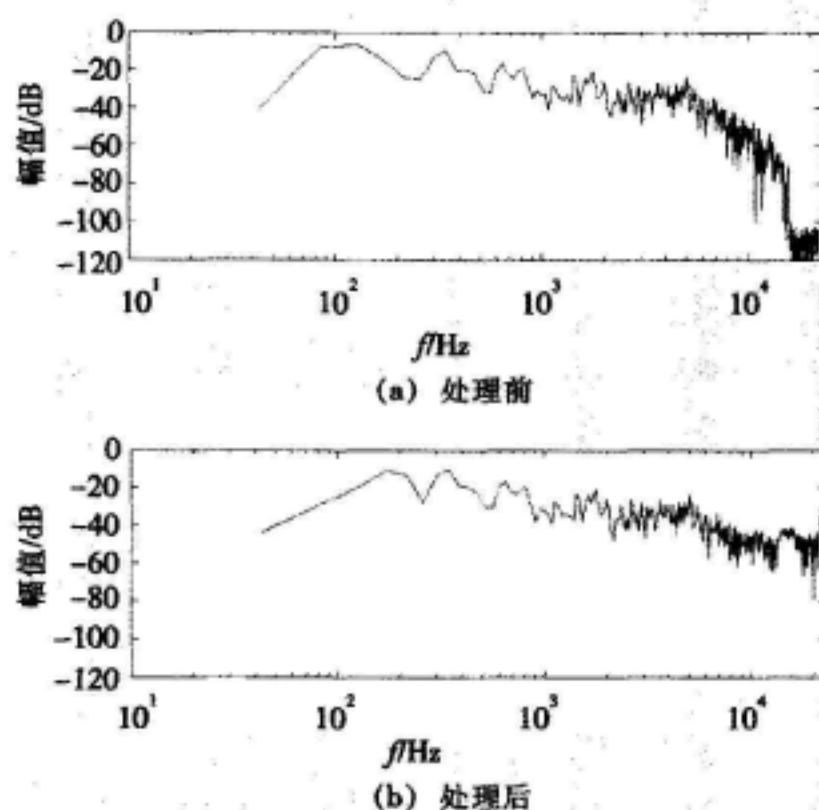


图4 23 ms长的音乐信号经MaxxBass算法处理前后的频谱

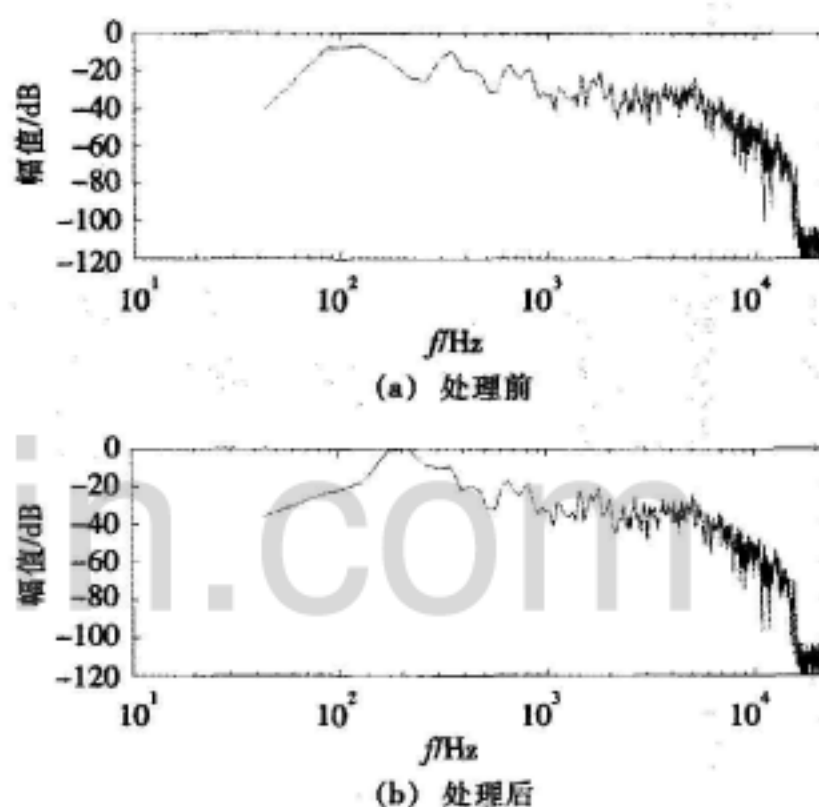


图5 23 ms长的音乐信号经EVB Phase Vocoder算法处理前后的频谱

内容包括2个方面:低音效果评分和音质评分。评分范围为1~5分,对于低音效果,1~5分分别代表低音效果“无”、“不明显”、“明显”、“很明显”、“非常好”;对于音质,1~5分分别代表音质“不可忍受”、“差”、“明显下降”、“稍有下降”、“无明显下降”。对11人的结果取平均值,结果见表1。

从表1可看出,EVB Phase Vocoder算法相对于MaxxBass算法,低音效果和音质两方面表现均有明显提高,其中对鼓声表现最好,同时,引入虚拟低音增强

(下转第67页)

表 1 实验结果

α	准确率/%			
	Top 1	Top 3	Top 5	Top 10
0	25	40	44	58
0.25	36	55	63	70
0.50	36	59	70	87
0.75	27	48	58	70
1.00	11	15	16	20

5 结论

在深入研究已有音乐检索算法的基础上,提出了有背景音乐的基音检测算法,改进了 DTW 算法使之支持高度动态调整和可变长搜索。实验结果表明,笔者提出的有背景音乐的基音检测算法能准确提取得到音乐的特征,改进后的 DTW 算法能更好地衡量哼唱输入与音乐库中的音乐的近似程度。由于音乐库较大,笔者使用的算法复杂度较高,所以搜索的时间较长,需要进一步改进。

参考文献

- [1] LI Tao, MITSUNORI Ogihara. Toward intelligent music information retrieval[J]. IEEE Trans. on Multimedia, 2006, 8(3): 564-574.
- [2] JIANG Dan-ning, PICHENY M, QIN Yong. Voice-melody transcription under a speech recognition framework[C]//

Proceedings of 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing.[S.l.]: IEEE Press, 2007, 4: 617-620.

- [3] GAO Liang, WU Ya-dong. A system for melody extraction from various humming inputs [C]// Proceedings of 2006 IEEE International Symposium on Signal Processing and Information Technology.[S.l.]: IEEE Press, 2006: 680-684.
- [4] WU Ya-dong, LI Yang, LIU Bao-long. A new method for approximate melody matching [C]// Proceedings of 2003 International Conference on Machine Learning and Cybernetics. Xi'an: IEEE Press, 2003, 5: 2687-2691.
- [5] 蔡莲红, 黄德智, 蔡锐. 现代语音技术基础与应用[M]. 北京: 清华大学出版社, 2003.
- [6] ZHU Yong-wei, KANKANHALLI M S. Robust and efficient pitch tracking for query-by-humming [C]// Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia.[S.l.]: IEEE Press, 2003: 1586-1590.

作者简介

薛振武, 硕士研究生, 主要研究方向为音频信号处理、嵌入式系统设计和实现;
胡剑凌, 导师, 主要研究方向为音频信号处理、嵌入式系统设计和实现、高清数字电视、DSP 应用等。

[责任编辑] 侯莉

[收稿日期] 2008-12-16

(上接第 62 页)

算法都会对信号的音质产生一定的影响。

表 1 主观测听评分结果

曲目	MaxxBass		EVB Phase Vocoder	
	低音效果 评分	音质评分	低音效果 评分	音质评分
01 TheMass Excerpt1	2.2	3.2	3.0	4.1
02 TheMass Excerpt2	2.1	3.0	3.3	4.1
03 TheMass Excerpt2	2.3	2.7	3.5	4.4
04 Drum	2.0	1.9	4.3	4.4
05 Zorbass	2.0	1.5	4.0	3.2
06 Radetzky march	1.9	4.3	2.3	4.3
07 Rock n' Roll	2.5	3.5	3.0	4.3
08 Propeller	2.5	2.3	3.5	3.2

5 结论

通过实验对比可以看出, 将 Phase Vocoder 时频分析方法应用于虚拟低音增强技术, 有效解决了 MaxxBass 算法中合成信号存在非线性畸变的问题, 同时实现了对合成谐波参数的精确控制。EVB Phase Vocoder 方法在低音效果还原, 信号畸变控制等方面均优于 MaxxBass 方法。

参考文献

- [1] ZWICKER E, FASTL H. Psychoacoustics facts and models [M]. 2nd ed. Berlin: Springer verlag, 1999.
- [2] LARSON E, AARTS R M. Audio bandwidth extension[M]. Hoboken: John Wiley & Sons, 2004.
- [3] DANIEL B-T, MARTIN C. The effect of maxxbass psychoacoustic bass enhancement system on loudspeaker design[C]// AES Convention 106.[S.l.]: AES, 1999: 4892.
- [4] SHASHOUA M, GLOTTER D. Method and system for enhancing quality of sound signal: US, 5930373[P]. 1999.
- [5] MINGSIAN R B, LIN Wan-chi. Synthesis and implementation of virtual bass system with a phase-vocoder approach[J]. JAES, 2006, 54(11): 1077-1091.
- [6] ZOLER U. DAFX digital audio effects[M]. Hoboken: John Wiley & Sons, 2002.

作者简介

郭杰, 硕士研究生, 主要研究方向为语音与音乐信号处理;
李晓东, 研究员, 主要研究方向为声学及声信号处理。

[责任编辑] 潘浩然

[收稿日期] 2008-12-24