

基于衰减因子的虚拟环绕系统听音区扩大方法

黄 强, 吴一波, 何 飞, 纪 震

(深圳市嵌入式系统设计重点实验室, 深圳大学计算机与软件学院, 广东深圳 518060)

摘 要: 面向音频多声道虚拟环绕处理的方法, 传统的串音抵消是将串音完全抵消掉, 听音区窄. 本文提出串音信号不必要完全消除, 剩余的信号会扩大最佳听音区域的范围并获得更好的音效质量. 新算法的主要特点是: 算法中加入衰减因子, 并给出了衰减因子和听音区的关系. 当衰减因子为 1 时, 新的算法完全具有传统算法的功能.

关键词: 虚拟环绕; 衰减因子; 串音消除; 扩大听音区

中图分类号: TN912 **文献标识码:** A **文章编号:** 0372-2112 (2012) 11-2342-04

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2012.11.032

A Novel Method to Enlarge Listening Area of Virtual Surround Sound Based on Attenuation Factor

HUANG Qiang, Wu Yi-bo, HE Fei, JI Zhen

(Shenzhen City Key Laboratory of Embedded System Design, College of Computer Science and Software Engineering, Shenzhen, Guangdong 518060, China)

Abstract: For audio multichannel virtual surround sound processing method. The crosstalk was offsetted completely in traditional method for crosstalk cancellation, which limits the optimal listening area. This paper puts forward the non-necessity of completely crosstalk elimination. The proposed method utilizes the remaining crosstalk signal to effectively enlarge the range of the optimal listening area and maintain the same sound quality. The primary features of the proposed algorithm are as follows: the algorithm is jointed with an attenuation factor. Moreover, the algorithm introduces the relationship between the attenuation factor and listening area. In addition, when the attenuation factor value is set to be unit, the new algorithm is fully functional identical to the traditional algorithm.

Key words: virtual surround; attenuation factor; crosstalk cancellation; enlarge listening area

1 引言

近年来, 虚拟听觉^[1]成为研究的热点问题, 虚拟环绕声^[2]的研究在人工听觉、机器人科学、虚拟现实等研究中都有着重要的应用. 虚拟环绕声系统的基本原理都是利用与头相关传递函数(HRTF, Head Related Transfer Function)^[3,4]对虚拟声源的方位进行虚拟, 当采用扬声器作为重放设备时, 会产生串音现象, 从而改变了重放的音色, 同时使得听音区域变窄. 为了改善虚拟声系统的听音效果, 提出了对应的解决方案, 包括扬声器的放置位置, 串音抵消网络^[5]的改进等.

2 串音消除扩大听音区的设计实现

3D 虚拟声系统^[6]通过在传递到听者耳道的声压中嵌入方位因素^[7], 可以生成来自于任意方向的声像, 使听者具有沉浸式感受. 利用音箱生成 3D 立体声效的一个主要问题是交叉串音(cross-talk)的存在, 使得其中任

意一个声道的声音都可以传递到听者另一侧的耳道中, 形成对期望声音传递的干扰, 严重影响重发效果. 因此普遍采用串音抵消网络以抵消重发时的交叉串音^[8]. 首先 O Kirkeby 证明了缩窄左、右扬声器之间的张角, 可提高声像的稳定性和扩大听音区域, 并提出了称为‘立体声偶极’的虚拟重放系统^[9], 将左、右扬声器布置在 ± 50 的位上. 但‘立体声偶极’需要对信号的低频部分做较大的提升, 信号处理较为困难.

2.1 典型的串音抵消系统结构分析

为了得到虚拟声源的信号处理模型^[10], 假设声源来自正对听者左耳的位置, 先利用一对麦克风记录下真实声音, 再通过音箱重发. 在声源发出一个声音信号之后, 先被左边的麦克风记录, 然后被右边的麦克风记录. 两个麦克风距离接近于一个头部宽度, 则时延大约为 $580\mu\text{s}$. 判断声源来自于左音箱位置. 因此包含其中的真实方位信息就受到破坏, 从而不能还原原始声像位置^[11], 影响虚拟立体声的逼真效果.

具体而言,如果用 (T_{11}, T_{21}) 、 (T_{12}, T_{22}) 分别代表左、右音箱到双耳的声学传递函数, \mathbf{H} 是由虚拟声源到理想输出的传输矩阵, D_L 、 D_R 是最终到达人耳的信号. 假设期望两耳接收到的信号为 x_L 、 x_R . 由图 1 可以看到, 在没有经过处理之前, 到达双耳的信号为:

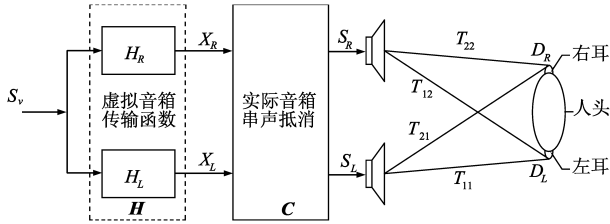


图1 基于扬声器的虚拟重发系统结构

$$\begin{aligned} \begin{pmatrix} D_L \\ D_R \end{pmatrix} &= \mathbf{T} \cdot \begin{pmatrix} S_L \\ S_R \end{pmatrix} = \mathbf{T} \cdot \mathbf{C} \cdot \begin{pmatrix} X_L \\ X_B \end{pmatrix} \\ &= \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} X_L \\ X_B \end{pmatrix} \end{aligned} \quad (1)$$

若使得实际信号与期望信号相同, 应有

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (2)$$

即 $\mathbf{C} = \mathbf{T}^{-1}$. 从而串音消除网络的实质就是构建逆滤波器^[12], 网络的设计也影响着最终的消除效果, 决定了听者能否接受到稳定的信号.

为避免对音箱传递函数矩阵 \mathbf{T} 求逆, 引入反馈式的消除网络(见图 2), 设反馈矩阵为 \mathbf{F} , 前向传输矩阵为 \mathbf{Q} , 则串音消除网络的输出为:

$$\begin{aligned} \mathbf{S} = \begin{pmatrix} S_L \\ S_R \end{pmatrix} &= \mathbf{F} \cdot \mathbf{S} + \mathbf{Q} \cdot \mathbf{X} \\ &= \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \begin{pmatrix} S_L \\ S_R \end{pmatrix} + \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} X_L \\ X_R \end{pmatrix} \end{aligned} \quad (3)$$

为简化设计, 令 \mathbf{Q} 为对角化矩阵

$$\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{T_{11}} & 0 \\ 0 & \frac{1}{T_{22}} \end{pmatrix} \quad (4)$$

则由上式可得到

$$\begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} = \begin{pmatrix} 0 & -\frac{T_{12}}{T_{11}} \\ -\frac{T_{21}}{T_{22}} & 0 \end{pmatrix} = \begin{pmatrix} 0 & f_1 \\ f_2 & 0 \end{pmatrix} \quad (5)$$

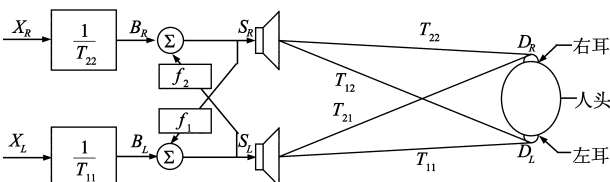


图2 串音消除的反馈法模型

由于音箱传递函数 T_{11} 、 T_{12} 、 T_{21} 、 T_{22} 都是最小相位系统, 所以 \mathbf{Q} 和 \mathbf{F} 也都是最小相位系统, 从而保证了整个系统的稳定性.

该方案结构简单, 但听者对位置具有依赖性, 如果听者头部在前方转动超过 10° , 就会使听者感到声像偏到头的内部, 从而降低了听音效果.

2.2 基于衰减因子的串音消除算法实现

本文针对听者头部运动的问题, 提出了如下的系统结构:

整个串音抵消网络的系统结构如图 3 所示. 包括左通道、右通道, 每个通道有一个独立的输入 W_L 、 W_R , 用于接受双耳信号, 每个通道由分枝节点到 $x(\mathbf{C})$ 和求和节点、校正、增益滤波器和交叉通道滤波器 G . 其中 S 、 A 为音箱到人头之间的头部相关传递函数(HRTF). 信号经过以上通道, 进入扬声器 S_L 、 S_R 到达听者. 其中 $x(\mathbf{C})$ 为引入衰减因子的交叉通道滤波器, 其传递函数为 $x(-\frac{A}{S})$, G 为校正、增益滤波器. 该算法包含两部分功能: 串音抵消和频谱均衡.

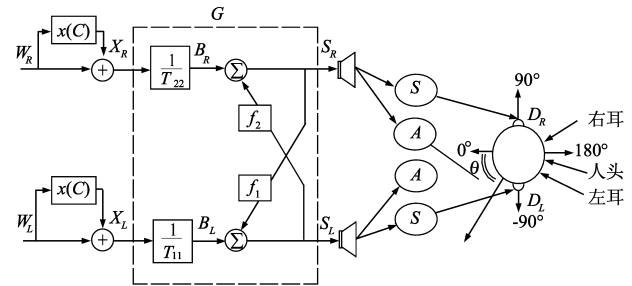


图3 基于衰减因子的虚拟环绕声系统框架图

假设虚拟的双声道信号为 W_L 、 W_R , 我们以 W_R 传到右耳为例. 信号在输入结点分为 2 个通道, 如下:

(1) 通过交叉通道滤波器 $x(\mathbf{C})$, 再经过校正、增益滤波器 G , 最后经过的头部相关传递函数 A 进入右耳. 此时, 信号通过这一通道的所经过传递函数为

$$R_{r1}(f) = x\left(-\frac{A}{S}\right) GA \quad (6)$$

(2) W_R 经过增益滤波器 G , 再通过头部相关递函数 S 进入右耳, 信号经过此通道的传递函数为:

$$R_{r2}(f) = GS \quad (7)$$

所以, 右声源到右耳的传递函数为:

$$R_r(f) = R_{r1}(f) + R_{r2}(f) = x\left(-\frac{A}{S}\right) GA + GS \quad (8)$$

为保证该传递函数对右声源信号 R 是“透明的”, 则 $R_r(f) = 1$, 那么 G 为:

$$G = \left(\frac{S}{S^2 - xA^2}\right) \quad (9)$$

同理, 右声源到左耳的传递函数为:

$$R_1(f) = x \left(-\frac{A}{S} \right) SG + GA$$

$$= GA(1-x) = \left(\frac{SA}{S^2 - xA^2} \right) (1-x) \quad (10)$$

即使低频段 S 和 A 近似相等的, 由于 x 的引入, 使得 $(S^2 - xA^2)$ 是有限的, 从而保证了滤波器的稳定性. 左声源到左耳、右耳的传递函数与右声源相同.

根据该方案中右声道到左耳的传递函数可知, 串音信号没有完全消除, 因此当听者的位置发生变化时, 人耳接收的串音信号不会发生明显的变化, 这样剩余一些串音信号扩大了听音区域.

3 测试结果与数据分析

3.1 加入衰减因子算法与传统串音消除算法比较

测试是采用音乐(德沃夏克第九交响曲片段, 长度 10s)作为信号. 由于左、右对称性, 只需要对左半水平面 $0^\circ \leq \theta \leq 180^\circ$ 的声像进行测试即可, 虚拟声像的间隔为 15° , 分别对应水平面内 0° 、 15° 、 30° …… 180° 等方向. HRTF 可以通过大量的个人 HRTF 的统计平均来获得, 也可以通过建模的方法由实验室得到. 本文所用的 HRTF 数据是麻省理工学院(MIT)媒体实验室提供的一组“一般 HRIR^[3]”数据, 然后通过傅立叶变换所得的结果. 并将每个位置虚拟声像的定位实验结果取平均值后作为最后的实验结果(实验值).

定位实验结果(实验)与传统的串音抵消的对比见表 1、表 2、和表 3.

可以看出, 对前半水平面, 在 $0^\circ \leq \theta \leq 45^\circ$ 的前方范围, 声像定位较为准确, 平均误差不超过 4° ; 在 $45^\circ \leq \theta \leq 90^\circ$ 以上方向平均定位误差逐渐加大, $\theta = 90^\circ$ 方位时达到最大(14°), 随着 θ 的增加, 平均定位误差和均方差逐渐缩小, 但是即使在 165° 方位处, 仍然存在一定的平均定位误差. 可以看到, 传统的串音消除在 $\theta = 0^\circ$ 为最佳的听音位置.

表 2 和表 3 分别给出了给出了衰减因子 $x = 1$ 与衰减因子 $x = 0.5$ 的声像定位的平均差与均方差.

表 1 传统的串音消除声像定位的平均误差与均方差

方位($^\circ$)	平均误差($^\circ$)	均方差($^\circ$)
0	0	0
15	1	1
30	1	2
45	4	4
60	5	4
75	10	4
90	14	6
105	17	9
120	11	8
135	8	5
150	4	4
165	2	2
180	1	1

由表 2 可知, 但衰减因子为 1 时, 声像定位的测试结果和传统的串音消除方法非常接近, 效果基本相同. 由表 3 可知, 声像的平均定位误差较传统的串音消除方法总体误差要小, 将最佳听音点平均到了 180° 范围, 克服了传统方法虚拟系统听音区^[13]窄的缺陷, 能够较好地

3.2 最佳的听音位置的尺寸和衰减因子 x 的关系

对原始的 5.1 通路信号处理后得到单声道的数据信息, 进行虚拟处理, 以获得用于声像定位实验的重发信号, 其详细步骤如下:

(1) 将待处理的音频信号^[4]按照图 3 的结构图处理, 可得到虚拟重发信号.

(2) 选取适当的原始信号, 生成所需要的某个虚拟扬声器信号, 保存为双通路 WAVE 文件.

实验结果如图 4:

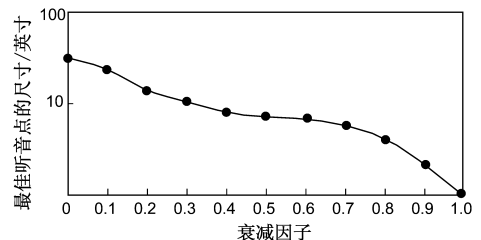


图 4 衰减因子 x 和最佳听音点尺寸的关系

由图 4 可知, 当 x 在 0.2 和 0.8 变化时, 最佳的听音区域的范围变化缓慢, 小于 10 英寸, 当 x 小于 0.2 时, 最佳的听音区域的范围变化很快.

x 为 1 时, 只有一个特殊的最佳听音点, 超出这个位置, 重放的音频效果就会降低.

表 2 衰减因子为 1 的声像定位平均误差与均方差

方位($^\circ$)	平均误差($^\circ$)	均方差($^\circ$)
0	0	0
15	1	1
30	1	3
45	4	4
60	4	4
75	9	4
90	14	6
105	16	7
120	11	5
135	8	4
150	5	4
165	2	2
180	1	1

表 3 衰减因子为 0.5 的声像定位在水平面内 360° 的声像平均误差与均方差

方位($^\circ$)	平均误差($^\circ$)	均方差($^\circ$)
0	1	0
15	1	1
30	1	2
45	2	2
60	3	2
75	2	1
90	3	2
105	2	1
120	1	2
135	2	3
150	1	1
165	2	2
180	1	1

x 为 0.95 时,最佳的听音区的尺寸达几个英寸,听者可以向前、向后以及垂直的运动或者旋转头部。

x 为 0.5 时,最佳听音范围是 10 英寸(25cm),听者移动椅子的位置仍然可以获得最佳的听音效果。

4 结论与未来展望

声音信号通过扬声器重放和通过耳机重放是不同的。扬声器与耳机的区别在于,耳机的声音是直接传到两耳,而扬声器与人耳之间有一段距离,根据声音传播原理,扬声器会产生耳间串音。针对这些问题,本文提出了一种新的基于衰减因子的串音消除算法。新算法具有以下优点:

(1)提供了听者头部处于最佳的听音位置的双耳信号的重放处理方法。

(2)交叉通道引入了一个衰减,使得双耳信号中有剩余的串音信号,当双耳信号重放时,允许处于最佳的听音位置的听者头部运动,不会感到双耳信号的明显改变。

(3)新的算法当衰减因子为 1 时,完全具有传统算法的功能。

参考文献

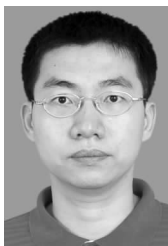
- [1] LIN W, FULIANG Y, ZH E C. An "out of head" sound field enhancement system for headphone[A]. International Conference on Neural Networks and Signal Processing[C]. Nanjing, China: IEEE press, 2008. 517 - 521.
- [2] D R Bngault. 3-D Sound for Virtual Reality and Multimedia [M]. Cambridge: Morgan Kaufmann Pub, 1994.
- [3] Kyriakakis C. Fundamental and technological limitations of immersive audio systems[J]. Proceedings of the IEEE, 1998, 86(5): 941 - 951.
- [4] Kim Y, Deille O, Nelson PA. Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners[J]. Journal of Sound and Vibration, 2006, 297(2): 251 - 266.
- [5] Bai MR, Lee C. Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction[J]. The Journal of the Acoustical Society of America, 2006, 120(4): 1976 - 1989.
- [6] 张超鹏. 双声道三维虚拟声系统的研究与实现[D]. 南京: 东南大研究生院, 2009.
ZHANG Chao-peng. Research and Implementation of Dual Channel 3D Virtual Sound System [D]. Nanjing: Graduate School of Southeast University, 2009. (in Chinese)
- [7] Jean J Labrosse. MicroC/OS-II The Real-Time Kernel [M]. CMP Books, 2002.
- [8] 贾懋 ■, 鲍长春. 一种符合 ITU-T 指标要求的嵌入式立体语音音频编码方法[J]. 电子学报, 2009, 37(10): 2291 -

2297.

JIA Mao-shen, BAO Chang-chun. A embedded stereo speech and audio coding method meeting the requirements of ITU-T terms of reference[J]. Acta Electronica Sinica, 2009, 37(10): 2291 - 2297. (in Chinese)

- [9] O Kirkeby. The stereo dipole a virtual source imaging system using two closely spaced loudspeakers[J]. Journal of the Audio Engineering Society, 1998, 46(5): 387 - 395.
- [10] 张一彬, 周杰, 边肇祺, 张大鹏. 一种新的基于分类的音频流分割方法[J]. 电子学报, 2006, 34(4): 612 - 617.
ZHANG Yi-bin, ZHOU Jie, BIAN Zhao-qi, ZHANG Da-peng. A novel classification-based audio segmentation algorithm[J]. Acta Electronica Sinica, 2006, 34(4): 612 - 617. (in Chinese)
- [11] 王鑫, 唐舒岩. 数字声频多声道环绕声[M]. 北京: 人民邮电出版社, 2008.
WANG Xin, TANG Shu-yan. Digital Audio Multi-Channel Surround Sound [M]. Beijing: POST & TELECOM PRESS, 2008. (in Chinese)
- [12] 谭良, 吴波, 刘震, 周明天. 一种基于混沌和小波变换的大容量音频信息隐藏算法[J]. 电子学报, 2010, 38(8): 1812 - 1818.
TAN Liang, WU Bo, LIU Zhen, ZHOU Ming-tian. An audio information hiding algorithm with high-capacity which based on chaotic and wavelet transform[J]. Acta Electronica Sinica, 2010, 38(8): 1812 - 1818. (in Chinese)
- [13] 谢波菽. 对 5 通路 3/2 环绕声系统缺陷的分析应用声学[J]. 应用声学, 1997, 16(5): 1 - 7.
Xie Bosun. Analysis on defects of the 5 channel 3/2 surround sound system[J]. Applied Acoustics, 1997, 16(5): 1 - 7. (in Chinese)

作者简介



黄 强 男, 教授. 1977 年 1 月出生于广东湛江. 2004 年在英国利物浦大学获工学博士学位. 现为深圳大学计算机与软件学院软件工程系主任, 主要从事嵌入式系统、数字信号处理和数字图像处理等方面的研究工作.

E-mail: jameshq@szu.edu.cn



吴一波 男, 1985 年 9 月生, 湖南娄底人, 硕士研究生, 主要研究方向为嵌入式系统, 数字信号处理.

E-mail: abowu@qq.com