



基于 Householder 矩阵和 Butterworth 滤波器组的反馈延迟网络人工混响

摘要

针对现有人工混响方法中反馈延迟网络方法存在延迟线的个数选择受限和无法设定子带混响时间的不足,本文设计了一种基于 Householder 反馈矩阵和 Butterworth 滤波器组的反馈延迟网络人工混响器. Householder 反馈矩阵有助于最大化回声密度,延迟线个数的选择相对自由, Butterworth 滤波器组便于设定子带混响时间. 仿真实验中分别利用语谱图、语音质量感知评价和主观评价验证了本文方法的有效性.

关键词

人工混响; 反馈延迟网络; 反馈矩阵; 滤波器组

中图分类号 TU112; TN912.3

文献标志码 A

收稿日期 2019-03-22

资助项目 国家自然科学基金(12074192)

作者简介

吴礼福,男,博士,副教授,研究方向为音频信号处理. wulifu@nuist.edu.cn

1 南京信息工程大学 电子与信息工程学院, 南京, 210044

2 南京信息工程大学 江苏省大气环境与装备技术协同创新中心, 南京, 210044

0 引言

人工混响(artificial reverberation)是指人为地对原始声音信号进行处理,使其能够具有适当的混响效果^[1-2].在音乐、广播、电视和电影制作过程中,人工混响是不可缺少的部分.混响处理方法可以采用数字人工混响方法或模拟技术,其中数字人工混响方法是利用电声学知识以数字信号处理手段来模拟混响效果^[2].

数字人工混响方法主要有3种:一是反馈延迟网络(Feedback Delay Network, FDN),它将输入信号(干净无混响)延迟、滤波并根据参数化混响特性沿着多个路径反馈给前端,叠加后得到混响信号;二是卷积方法,它将输入信号与房间脉冲响应卷积得到混响信号;三是基于计算声学的方法,将输入信号模拟声能在几何模型中传播,从而得到混响信号^[3].反馈延迟网络方法在音乐技术领域使用较多.计算声学方法通常可用于声学设计和场景分析.而卷积方法在实时实现方面非常困难,脉冲响应的计算很耗时.本文主要研究内容是反馈延迟网络.

1971年,反馈延迟网络首先由 Gerzon 提出用于人工混响^[4],他指出单个反馈梳状滤波器质量很差,但是当交叉耦合时,使用几个反馈梳状滤波器效果会好很多.1996年, Jot 等^[5]提出了一种反馈延迟网络(FDN)方法来处理数字混响,将 FDN 方法发展到目前的应用水平. Jot 等^[5]提出的反馈延迟网络方法目前被认为是高质量人工混响的最佳选择之一,其中,正交反馈矩阵的选择是一个特别有趣的话题,它显著影响所获得的混响质量^[6-8].1998年, Piiril 等^[9]已经提出了如何使用2个具有稍微不同参数的 FDN 或其他修改的梳状滤波器结构来产生非指数衰减的混响响应.2011年, Sana 等^[10]通过结合频率相关的墙壁吸收、信源和接收器的方向性进一步扩展了 FDN 概念.

2010年, Smith 教授提出了反馈延迟网络的一种算法^[11],它是基于 Hadamard 反馈矩阵和 Butterworth 滤波器的,混响程度很深,但是延迟线个数的选择有限制.2017年,美国苹果公司发布一个名为 SoundS-oup 的应用程序,它是基于 Householder 反馈矩阵和单零点极点的滤波器,运算量小、处理时间短^[12],但是需要提供混响室的具体长宽高和墙壁的吸声系数,通过 Sabine 公式来计算混响时间 T_{60} ,无法精确

设定子带的混响时间.

为了克服以上 2 种方法的缺点,本文研究了一种新的反馈延迟网络方法,它基于 Householder 反馈矩阵和 Butterworth 滤波器组,延迟线的个数选择相对自由,可以精确设定子带混响时间.

1 反馈延迟网络

图 1 所示是用于人工混响的反馈延迟网络,它是一个使用 3 条延迟线的反馈延迟网络的例子. $x(n)$ 表示输入信号(干净无混响), $s_1(n), s_2(n), s_3(n)$ 是经过延迟线的信号, b_1, b_2, b_3 表示输入因子, c_1, c_2, c_3 表示输出因子, g_1, g_2, g_3 表示频带增益, q_{11} 等表示反馈矩阵的元素, $y(n)$ 则表示输出的混响信号.

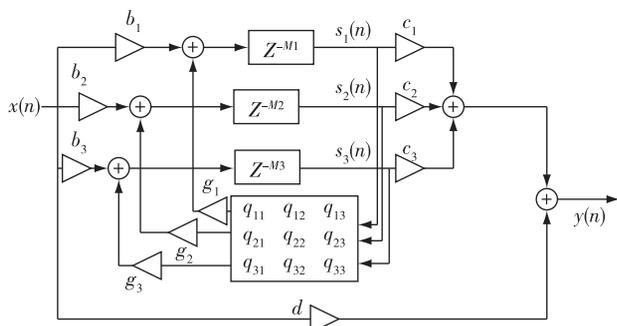


图 1 用于人工混响的反馈延迟网络

Fig. 1 Feedback delay network for artificial reverberation

对于 $N=3$, 图 1 所示 FDN 的关系式可以写成:

$$\begin{bmatrix} s_1(n + M_1) \\ s_2(n + M_2) \\ s_3(n + M_3) \end{bmatrix} = \begin{bmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \\ s_3(n) \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} x(n). \quad (1)$$

输出结果为

$$\begin{bmatrix} y_1(n) \\ y_2(n) \\ y_3(n) \end{bmatrix} = \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \\ s_3(n) \end{bmatrix} + d \begin{bmatrix} x(n) \\ x(n) \\ x(n) \end{bmatrix}. \quad (2)$$

推广到 $N=n$ 时,FDN 的关系式可以表示为

$$\begin{bmatrix} s_1(n + M_1) \\ s_2(n + M_1) \\ \vdots \\ s_n(n + M_n) \end{bmatrix} =$$

$$\begin{bmatrix} g_1 & 0 & \cdots & 0 \\ 0 & g_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_n \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_n(n) \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} x(n), \quad (3)$$

$$\begin{bmatrix} y_1(n) \\ y_2(n) \\ \vdots \\ y_n(n) \end{bmatrix} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_n(n) \end{bmatrix} + d \begin{bmatrix} x(n) \\ x(n) \\ \vdots \\ x(n) \end{bmatrix}. \quad (4)$$

或者,使用 z 变换,在频域中写为

$$S(z) = D(z) [GQS(z) + bX(z)], \quad (5)$$

$$Y(z) = c^T S(z) + dX(z), \quad (6)$$

其中 $G = \text{diag}(g_1, g_2, \dots, g_n)$ 为增益组成的对角矩阵, $Q = [q_{i,j}]_{N \times N}$ 为反馈矩阵, $b = [b_1, b_2, \dots, b_n]$ 为输入因子组成的列向量, $c = [c_1, c_2, \dots, c_n]$ 为输出因子组成的列向量, $D(z) = \text{diag}(z^{-M_1}, z^{-M_2}, \dots, z^{-M_n})$ 为延迟线组成的对角矩阵.

脉冲响应的后期混响部分理想情况下应该类似于指数衰减的随机噪声^[13].一旦在无损的脉冲响应中听到平滑的噪声,就可以在每个频带中获得期望的混响时间,而噪声的平滑性受 FDN 反馈矩阵以及延迟线长度的影响.

1.1 延迟线

平均延迟线的长度通常粗略地等于混响环境下的平均自由程.平均自由程的定义为声波在传播并衰减的过程中,经过每两个界面之间的平均距离.平均自由程 ρ 近似为

$$\rho = 4V/S, \quad (7)$$

其中 V 表示的是房间的体积, S 表示房间的表面积.如果将每条延迟线视为平均自由程延迟,则可将延迟平均值设为平均自由程:

$$\bar{M} = \frac{\rho}{cT} = \frac{1}{N} \sum_{i=1}^N M_i, \quad (8)$$

其中 c 表示声速, T 表示采样周期.给定 ρ 的值,选择一组具有预设的最小间距素数,其平均值尽可能接近所需值^[10].

根据 Schroeder^[1] 的观点,FDN 的延迟线的长度通常选择为互为质数的 \hat{M}_i , 所需延迟线长度 M_i 按

照升序排列,

$$M_1 < M_2 < \dots < M_N,$$

那么延迟线的长度 \hat{M}_i 就可以使用自然顺序的素数

p_i 来近似,素数延迟线长度 \hat{M}_i 的计算公式如下:

$$\hat{M}_i = p_i m_i, \quad (9)$$

其中 m_i 的计算方法如下:

$$m_i = \text{round}\left[\frac{\log M_i}{\log p_i}\right] = \text{floor}\left[0.5 + \frac{\log M_i}{\log p_i}\right], \quad (10)$$

其中 M_i 为所需延迟线的长度, p_i 为使用自然顺序的素数. $\text{round}(\cdot)$ 表示对数值进行四舍五入运算, $\text{floor}(\cdot)$ 表示对数值进行向下取整.

1.2 Butterworth 滤波器组

当房间的具体几何模型和墙壁的吸声系数未知时,就无法使用 Sabine 公式计算出所需的混响时间.多频带 FDN 延迟滤波器可以克服这个困难,它可以单独设置混响时间,混响时间应该至少在 3 个频段内独立可调^[14].相对于一阶延迟滤波器,更多地使用多频带 FDN 延迟滤波器,通常可以使用滤波器组来实现多频带延迟滤波器.例如,每条延迟线的输出被分成 $K(K \geq 3)$ 个频带,那么长度为 M_i 的延迟线的第 k 个频带的增益为

$$G = 10^{\frac{-3M_i}{n_{60}(\omega_k)}} = e^{\frac{-3\ln(10)M_i}{n_{60}(\omega_k)}}, \quad (11)$$

其中 $n_{60}(\omega_k) = t_{60}(\omega_k)/T$, $t_{60}(\omega_k)$ 表示频率 ω_k 的混响时间.

本文在反馈延迟网络中选用的是 Butterworth 滤波器组. Butterworth 滤波器是一种通带频率响应曲线很平坦的滤波器,是使用低通和高通 Butterworth 滤波器来实现所需特性的滤波器组.也就是说,整个频谱在最高的交叉频率被分割,在下一个交叉频率将低通区域再分割成 2 个频带.本文设置 S 个交叉频率,整个频带就被分成 $S+1$ 个频带,那么就需要 $2S$ 个 Butterworth 滤波器组成滤波器组.例如,设置 2 个交叉频率,那么整个频带就被分成 3 个频带,就需要 4 个滤波器组成滤波器组.

图 2 所示的是 $N=3$, 基于 Butterworth 滤波器组的 FDN. $x(n)$ 为输入信号(干净无混响), filter1 ~ filter4 为 Butterworth 滤波器组, $y(n)$ 为输出信号.

1.3 Householder 反馈矩阵

N 阶反馈矩阵 \mathbf{Q}_N 满足 $\mathbf{Q}_N = \mathbf{I}_N - \frac{2}{N} \mathbf{u}_N \mathbf{u}_N^T$, 其中 \mathbf{I}_N 为 $N \times N$ 的单位矩阵, $\mathbf{u}_N^T = [1, 1, \dots, 1]$. 采用 Householder 矩阵, 延迟线个数的选择相对比较自由, 可以

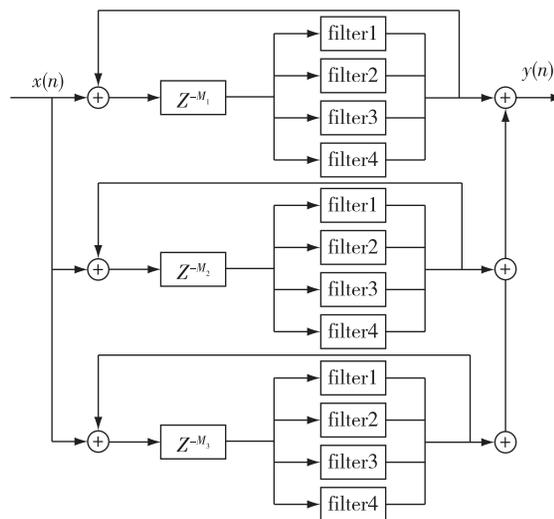


图 2 基于 Butterworth 滤波器组的 FDN

Fig. 2 FDN based on Butterworth filter bank

解决 Hadamard 反馈矩阵的缺点.

Householder 反馈矩阵 \mathbf{Q}_N 的另一个很好的特性就是, 当 $N \neq 2$ 时, 矩阵中的所有数都是非零的, 这就意味着每条延迟线都会反馈给其他延迟线, 从而有助于尽可能地最大化回声密度. 例如, 当 $N=4$ 时, Householder 反馈矩阵为

$$\mathbf{Q}_4 = \frac{1}{2} \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}. \quad (12)$$

由于 $N=4$ 的 Householder 反馈矩阵的平衡性, Jot 等^[14] 在此基础上提出了一种 $N=16$ 的反馈矩阵嵌入 FDN 内:

$$\mathbf{Q}_{16} = \begin{bmatrix} \mathbf{Q}_4 & -\mathbf{Q}_4 & -\mathbf{Q}_4 & -\mathbf{Q}_4 \\ -\mathbf{Q}_4 & \mathbf{Q}_4 & -\mathbf{Q}_4 & -\mathbf{Q}_4 \\ -\mathbf{Q}_4 & -\mathbf{Q}_4 & -\mathbf{Q}_4 & -\mathbf{Q}_4 \\ -\mathbf{Q}_4 & -\mathbf{Q}_4 & -\mathbf{Q}_4 & \mathbf{Q}_4 \end{bmatrix}. \quad (13)$$

本文也采用式(13)的反馈矩阵.

2 实验及分析

在实验仿真中, 将一段采样率为 8 000 Hz 的干净语音依次通过延迟线、滤波器组, 产生所需要的混响时间, 再通过反馈矩阵, 最后得到混响信号. 表 1 为仿真中的基本参数设置. 其中房间(学校体育馆)的长为 48 m, 宽为 19 m, 高为 18 m, 声源位置的三维坐标为 $[18, 11, 12]$, 传声器位置的三维坐标为 $[18, 8, 12]$.

Smith^[11]的 FDN 的方法使用 16 条延迟线,设定 3 个子带的混响时间,SoundSoup 使用 16 条延迟线,一个总的混响时间,而本文的方法是用 18 条延迟线(延迟线不必是 2 的整数次幂,选择自由),设定 3 个子带的混响时间.

表 1 仿真实验中的基本参数设置

Table 1 Basic parameter settings in the simulation experiment

方法	参数设置
Smith ^[11] 的 FDN	延迟线 16 条;输入因子 b 为 1,输出因子 c 为 1;交叉频率 3 150 和 315 Hz;混响时间 0.5、1.3 和 2.2 s;素数延迟线 M_i 最大值为 2 809,最小值为 125
SoundSoup	延迟线 16 条;输入因子 b 为 1,输出因子 c 为 1;房间的吸声系数 α 取 0.53;素数延迟线 M_i 最大值为 2 809,最小值为 125
本文的 FDN	延迟线 18 条;输入因子 b 为 1,输出因子 c 为 1;交叉频率 3 150 和 315 Hz;混响时间为 0.5、1.3 和 2.2 s;素数延迟线 M_i 最大值为 2 809,最小值为 125

图 3a—3d 分别为真实房间产生的混响信号、Smith^[11]方法处理后的信号、SoundSoup 处理后的信号以及本文方法处理后的信号的语谱图.对比图 3a、图 3b 和图 3c 可以看出 Smith^[11]方法、SoundSoup 处理过的信号能量强的频率衰减过程变得不清楚.

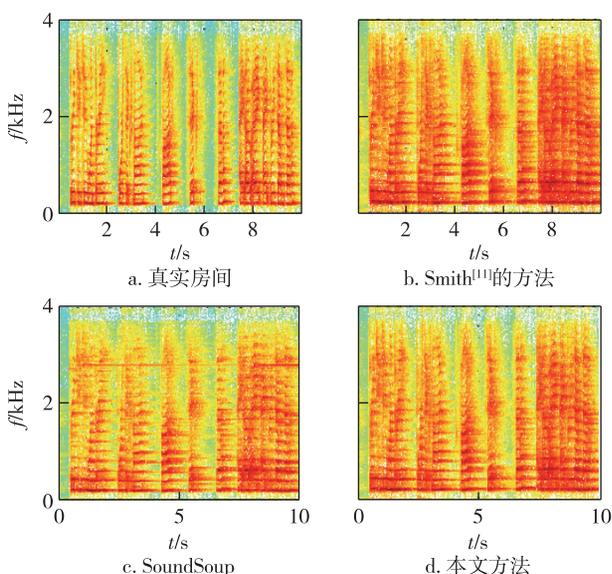


图 3 不同方法的语谱图对比

Fig. 3 Spectrogram comparison of different methods

为了进一步评价加混响的效果,采用语音质量感知评价^[15](Perceptual Evaluation of Speech Quality, PESQ)对混响效果进行评价,ITU-T(国际电信联盟电信标准化部)的相关资料已经证明:PESQ

能够精确地给出编码失真、传输丢失、环境噪声和时间扭曲的预测值.PESQ 得分的高低可以用来评价信号的好坏,通常情况下, PESQ 的得分在 1.0~4.5 之间.

本文选用了 10 条测试语音和 10 个房间脉冲,分别通过 3 种方法处理共得到 300 个混响信号.对比混响信号和真实房间产生的信号,得到 3 种方法 PESQ 得分的平均值.图 4 所示的是 3 种处理方法得到的混响信号 PESQ 得分情况.Smith^[11]的 FDN 方法得到的混响信号 PESQ 的平均值为 2.48, SoundSoup 方法得到的混响信号 PESQ 的平均值为 2.36,而经过本文方法处理得到的混响信号 PESQ 的平均值为 2.55.图 4 表明经过本文方法处理过的信号的 PESQ 得分较高,比 Smith^[11]方法提高了 0.07,比 SoundSoup 方法提高了 0.19.

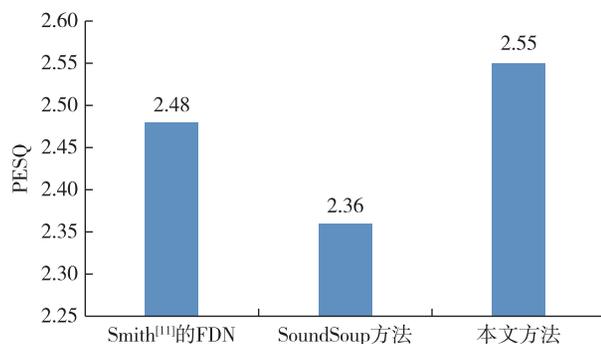


图 4 3 种方法处理后信号的 PESQ 得分

Fig. 4 PESQ scores of signals processed by the three methods

听者的主观感觉是判断混响感的重要评价标准^[16].因此,本文还采用听音实验来评价 3 种不同的人工混响信号.测试中音频文件采样率为 8 kHz,单声道,分别经过 3 种方法进行处理得到的混响信号.实验中选择 10 名听众,均为在校研究生,听力正常,对处理后的混响信号和真实房间产生的信号进行试听,选出 3 种方法中最佳、最接近真实房间产生的混响信号.10 名听众选出来的 100 条中语音中, Smith^[11]的 FDN 方法处理的语音有 26 条,占 26%; SoundSoup 方法处理的语音有 2 条,占 2%;本文方法处理的语音有 72 条,占 72%.大部分听者选择了本文方法,表明了在这 3 种方法中,本文方法能产生最佳且最接近真实房间的混响信号.

表 2 将 3 种方法的组成结构和优缺点进行了比较,可以看出本文的 FDN 方法更便于进行参数的设置.

表 2 3 种方法对比结果

Table 2 Comparison of three methods

	结构	优点	缺点
Smith ^[11] 的FDN	Hadamard 反馈矩阵、Butterworth 滤波器组	1) 可以设置子带的混响时间; 2) 不需要给出房间的长宽高、墙壁的吸声系数	1) 延迟线的选择有局限,只能是 2 的整数次幂; 2) 运算量相对大一点,处理时间长
SounSoup	Householder 反馈矩阵、单零点极点滤波器	1) 延迟线的个数不必是 2 的整数次幂; 2) 运算量相对小一点,处理时间短	不能设置子带的混响时间
本文的 FDN	Householder 反馈矩阵、Butterworth 滤波器组	1) 延迟线的个数不必是 2 的整数次幂; 2) 可以设置子带混响时间; 3) 不需要给出房间的长宽高及墙壁的吸声系数	运算量相对大一点,处理时间长

3 结论

本文在 Smith^[11] FDN 方法和苹果公司推出的 SoundSoup 基础上,提出了一种基于 Householder 反馈矩阵和 Butterworth 滤波器组的人工混响方法.语谱图、语音质量感知评价和主观评价结果表明,本文方法能产生比其他 2 种方法更加接近真实房间的混响信号,证明了本文方法的有效性.

参考文献

References

[1] Schroeder M R. Natural-sounding artificial reverberation [J]. J Audio Eng Soc, 1962, 10(3): 219-233

[2] 周婧洁,陈建平,吉晓东.几种数字混响器模型的分析与比较[J].南通大学学报(自然科学版),2010,9(2):23-28
ZHOU Jingjie, CHEN Jianping, JI Xiaodong. Comparison and analysis of several digital reverberator models [J]. Journal of Nantong University (Natural Science Edition), 2010, 9(2): 23-28

[3] Valimaki V, Parker J D, Savioja L, et al. Fifty years of ar-

tificial reverberation [J]. IEEE Transactions on Audio Speech & Language Processing, 2012, 20(5): 1421-1448

[4] Gerzon M A. Synthetic stereo reverberation parts I [J]. Studio Sound, 1971, 13: 632-635

[5] Jot J M, Chaigne A. Method and system for artificial spatialisation of digital audio signals; US, US 5491754 A [P]. 1996

[6] Rocchesso D, Smith J O. Circulant and elliptic feedback delay networks for artificial reverberation [J]. IEEE Transactions on Speech & Audio Processing, 1997, 5(1): 51-63

[7] Rocchesso D. Maximally diffusive yet efficient feedback delay networks for artificial reverberation [J]. IEEE Signal Processing Letters, 2002, 4(9): 252-255

[8] Menzer F, Faller C. Unitary matrix design for diffuse Jot reverberators [C] // Convention of the Audio Engineering Society, 2010

[9] Piiril E, Lokki T, Vlimki V. Digital signal processing techniques for non-exponentially decaying reverberation [C] // Proc Digital Audio Effects Conference (DAFX'98), 1998

[10] Sena E D, Hacıhabiboğlu H, Cvetkovic Z. Scattering delay network: an interactive reverberator for computer games [C] // AES 41st International Conference, London, UK, 2011

[11] Smith J O. Physical audio signal processing: for virtual musical instruments and digital audio effects [M]. W3K Publishing, 2010

[12] Bullen R. Simulation of room reverberation using a feedback delay network [J]. Acoustics Australia, 2015, 43(1): 1-4

[13] Moorer J A. About this reverberation business [J]. Computer Music Journal, 1979, 3(2): 13-28

[14] Jot J M, Chaigne A. Digital delay networks for designing artificial reverberators [C] // Convention of the Audio Engineering Society, 1991

[15] Hu Y, Loizou P C. Evaluation of objective quality measures for speech enhancement [J]. IEEE Transactions on Audio Speech & Language Processing, 2007, 16(1): 229-238

[16] 孟子厚,戴璐.混响的感知与设计 [M].北京:中国建筑工业出版社,2018:57-88
MENG Zihou, DAI Wei. Perception and design of reverberation [M]. Beijing: China Building Industry Press, 2018: 57-88

A feedback delay network reverberator based on Householder matrix and Butterworth filter bank

WU Lifu^{1,2} TAO Mingming¹ SHEN Hao¹ GUO Yecai¹

1 School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044

2 Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract In order to improve the limitation of the number of delay lines in the existing feedback delay network method and the inability to determine the sub-band reverberation time, this paper introduces a feedback delay network for artificial reverberation based on the Householder feedback matrix and Butterworth filter bank. The Householder feedback matrix can maximize the echo density, the choice of delay lines number is relatively free. The Butterworth filter bank is convenient for specifying the sub-band reverberation time. In the simulation experiment, the validity of the proposed method was verified by spectrogram, perceptual evaluation of speech quality, and subjective evaluation.

Key words artificial reverberation; feedback delay network; feedback matrix; filterbank