

## 声学仿真中的人工神经网络方法\*

张琼 石教英

(浙江大学 CAD&CG 国家重点实验室 杭州 310027)

**摘要** 虽然许多研究人员已认识到三维真实感声音在未来人机交互中的重要地位,但是三维真实感声音在计算机领域的真正实现仍有不少障碍有待克服。基于对声学及心理声学最新研究成果的调查和分析,本文设计并实现了一个基于神经网络方法的 HRTF(head-related transfer function)模型,用于三维真实感声音的生成,模型中的数据可通过网络学习进行重新设置,以满足多种场合的需要。并且,通过神经网络的非线性拟合能力,可以获取空间任意位置的 HRTF 数据。初步的实验表明了该方法的有效性和正确性。

**关键词** 人工神经网络,与头部相关的传递函数,三维真实感声音。

**中图法分类号** TP391

近年来,有关三维真实感声音的研究吸引了越来越多人的兴趣,许多多媒体和虚拟现实系统都将其作为系统的关键特性之一。在计算机领域,所谓的三维真实感声音,是相对于在人机接口中广泛使用的缺乏方向感和距离感的单声道声音而言的。<sup>[1]</sup>

虽然声学和心理声学的大量研究揭示,人的听觉空间定位感知特性受到许多因素的影响;但是一些学者的最新结果表明,人的双耳滤波效应(Pinnae Filtering)在听觉定位过程中扮演着极为重要的角色。<sup>[2,3]</sup>传统的 Duplex 理论企图仅依靠声音传达至聆听者双耳的时间差(ITD)和声级差(IID)来解释人类的听觉感知过程,但事实证明,该理论有着严重缺陷,比如它不能阐明与双耳间距相等的空间两点的定位问题。<sup>[4]</sup>

在声学领域,人们将声音从声源传至双耳耳鼓处的变换函数称之为与头部相关的传递函数 HRTF(head-related transfer function)。现在,该函数作为生成三维真实感声音的关键因素而广为人知。

一般说来,利用 HRTF 生成三维真实感声音有多种具体途径。<sup>[1]</sup>根据不同的应用场合,我们可选择合适的方法。

本文首先分析并指出了当前在利用 HRTF 生成三维真实感声音过程中存在的一些弊病和不足。然后基于一改进的神经网络模型,提出了两种用于产生三维真实感声音的具体方法。初步的实验结果表明,我们设计的神经网络模型及两种方法是非常有效的。

### 1 HRTF 的实现及存在的不足

所谓的 HRTF 可以看成是声音从声源传播至聆听者耳鼓处的变换函数。由于聆听者个体和声音传播过程中涉及因素的多样性,我们很难用一个统一的解析表达式来定义 HRTF。现阶段一般都是采用实验测量的方法来获取表征 HRTF 的部分脉冲响应数据,这其中包括两方面的内容:一部分为幅度响应;另一部分为相位响应。虽然在虚拟音频系统中应用 HRTF 被视为一项关键技术,但一般学者都认为,在人的听觉空间定位过程中,声波到达聆听者双耳的时间差要比单耳的 HRTF 相位响应发挥着更重要的作用。因此,许多研究人员在实现时通常将注意力集中在 HRTF 幅度响应上,并且使用声波(包括各种频率)到达双耳的平均时间差来取代 IIRTF 的相位响应。

一般认为<sup>[5,6]</sup>,决定 HRTF 的因素主要有 3 个:声波频率、声源位置、聆听者。我们特别注意到声波从声源传递至人耳耳鼓过程中的变化不仅是声源位置及声波频率的函数,而且不同的人其 HRTF 响应也不同。

不过,据我们调查所知,目前绝大多数的虚拟音频系统由于不可能面向特定的用户,所以在实现时一般不使用个人的 HRTF 测量数据,而是使用经过平均(Generalization)处理的某个人口集的 HRTF 数据。<sup>[7,8]</sup>这样一来虽然简化了 HRTF 的实现过程,但同时也应该看到,在不同的场合,使用平均的 HRTF 数据导致的聆听效果可能会有很大的差

\* 本文研究得到国家 863 高科技项目基金和浙江省自然科学基金资助。作者张琼,1973 年生,博士生,主要研究领域为多媒体,虚拟现实,信号处理。石教英,1937 年生,教授,博士生导师,主要研究领域为计算机图形学,体系结构,多媒体及可视化。

本文通讯联系人:张琼,杭州 310027,浙江大学 CAD&CG 国家重点实验室

本文 1997-01-03 收到原稿,1997-03-02 收到修改稿

异,而且,对一些特殊的用户,比如听觉障碍者,我们可能希望使用某组特定的数据.由此可以认为,现在许多虚拟音频系统的一个主要缺陷就在于:它不能根据应用环境和使用对象的不同,动态地通过学习来获取 HRTF 参数,以适应多种场合的复杂需求.

另外,平均的 HRTF 数据仍然存在着数据量偏大、使用不方便等问题.而且,由于我们不可能通过测量的方法来获取空间所有位置的 HRTF 数据,所以,有些人试图在其中引入插值的方法.但如声学的研究成果所揭示的那样,HRTF 及其自变量之间存在着相当复杂的非线性关系,一般的线性插值途径所得结果并不理想.文献[9]中也指出插值方法存在着缺点.

为了克服以上不足,我们设计并实现了一改进的人工神经网络模型,应用于虚拟音频系统,取得了较好的结果.

## 2 改进的基于人工神经网络的 HRTF 模型

我们在三维真实感声音生成领域引入人工神经网络方法,其主要目的是实现:①自学习功能 这就是说,通过网络训练,我们能够重置系统的 HRTF 数据,以满足多种场合的要求.②非线性逼近功能 这就是说,利用人工神经网络的非线性逼近能力,在一些离散测量值的基础上,我们能够获取空间任意位置的 HRTF 数据.

另外,我们还试图尽可能地加速基于神经网络模型的系统的运行速度.

### 2.1 基本设想

一些研究者<sup>[10-11]</sup>指出,多层前馈网络可以作为通用的函数逼近工具来使用.很多人工神经网络模型,特别是多层前馈网络,能够方便地处理非常复杂的非线性关系.很自然地,我们设想在 HRTF 的实现过程中引入人工神经网络的方法,以图获得较大的改进.

多层前馈网络的另一个突出优点就是其执行速度相当快.考虑到将来我们可能用软件实时实现,该类网络也许是唯一实际的选择.

基于以上考虑,我们决定采用多层前馈网络为基本模型来实现初始的两个主要目标.

### 2.2 神经网络模型的设计与实现

为了设计并实现一多层前馈网络模型,并使其适于解决文中所提出的问题,在研究中,我们主要考虑了以下几点:

#### 2.2.1 神经元非线性变换函数的选择及实现

对于多层前馈网络,非线性变换函数用来决定神经元的输出.目前大多数网络模型采用 S 形的单调上升函数作为神经元的变换函数,我们则选用以下指数函数为神经网络模型的变换函数:

$$f(x) = 1/(1 + e^{-x}) \tag{1}$$

该函数的一个优点是导数形式非常简单.

$$f'(x) = f(x) * (1 - f(x)) \tag{2}$$

由于在网络运算时经常要用到变换函数的导数形式,所以选用以上简单的对数函数可以加速整个系统的执行.

实现时我们也并不严格按照(1)式的数学定义来进行,而是设计了一个“预置插值法”,以期改善网络的总体性能.

现假设对  $x > 0.0$  有  $xd = \text{int}(x)$ ; 对  $x < 0.0$  有  $nx = -x, nxd = \text{int}(nx)$ , 即  $xd$  与  $nxd$  分别表示浮点数  $x$  与  $nx$  的整数部分. 则对任意的  $x$ , 可以将(1)式表示成:

$$f(x) = \begin{cases} f(10.0) & x \geq 10.0 \\ f(xd) + f(x - xd) * (f(xd + 1) - f(xd)) & 0.0 \leq x < 10.0 \\ 1 - (f(nxd) + (nx - nxd) * (f(nxd + 1) - f(nxd))) & -10.0 < x < 0.0 \\ f(-10.0) = 1 - f(10.0) & x \leq -10.0 \end{cases} \tag{3}$$

根据以上插值公式,我们在网络初始化时可以预先建立一函数表,至于该表的大小则依赖于运行时可用内存的大小及所要求的插值精度.这样,在网络训练及执行过程中,根据  $x$  值可以很方便地确定相应的变换函数值(即神经元的输出).

很明显,通过以上方法,我们得以大大加速网络模型的执行速度.并且更重要的是,使用预置插值法后,对任意的  $x$  值,其对应的函数值不是无限接近极限值 0 或者 1,而是受到预置函数表的限制(在(3)式中取函数表的极限值为  $f(10.0)$  与  $f(-10.0)$ ).由(2)式可知,当  $x=1/2$  时,  $f(x)$  的导数取得最大值,这就是说在训练过程中神经元的连接权值此时获得最大变化.但是当连接权值的绝对值已变得极大时,如仍沿用严格的数学定义进行计算,则(2)式将趋于 0 值,相对应地神经元的连接权值此时变化很小.经验表明,仅通过网络自身学习,它很难在较短时间内纠正这一不良现象.所以说,采用预置插值法,不仅加速了网络训练和执行,而且还能改善训练过程和训练结果.

### 2.2.2 网络输入输出层的设计

多层前馈网络的输入、输出层的维数完全需要按照不同的问题进行具体的设计. 对于我们所要求的 HRTF 逼近问题, 如前所述, 可以将该函数表示为:

$$HRTF \begin{cases} \text{magnitude} = h_1(f, azimuth, elev) \\ \text{phase} = h_2(f, azimuth, elev) \end{cases} \quad (4)$$

其中  $f$  代表声波频率(定义域为可听声的声波频率范围);  $azim(0^\circ \leq azim < 360^\circ)$  为声源相对于聆听者的偏角;  $elev(-90^\circ \leq elev \leq 90^\circ)$  声源相对于聆听者的仰角. 这样  $azim=0^\circ, elev=0^\circ$  表示声源位于聆听者的正前方,  $azim=90^\circ, elev=0^\circ$  表示声源位于聆听者的正右方, 由此通过  $azim$  和  $elev$  这一对变量可以标识整个空间的任一位置.

很自然地, 我们想到网络输入层的维数应为 3, 即用 3 个神经元分别对声波频率、偏角、仰角进行编码; 输出层的维数为 2, 分别代表 HRTF 的幅度响应和相位响应值.

但是我们又注意到偏角的定义是从  $0^\circ \sim 360^\circ$ , 这样当该变量从  $360^\circ$  向  $0^\circ$  方向变化时, 就存在着某种意义上的不连续性. 这就是说, 如果我们仅用一个输入神经元来对该变量进行编码, 则两个表征位置基本相同的量  $1^\circ$  和  $359^\circ$ , 其输入的差别却是最大的. 对于这样的输入变量, 实践证明神经网络的学习效果很差.

解决上述问题一个较好的方法就是用 2~3 个输入神经元来对偏角变量进行编码, 经过实验我们发现,  $(\sin, \cos)$  函数对能够很好地表征偏角变量, 并且解决了原变量存在的不连续性问题.

$$\begin{cases} g_1(azim) = \sin(azim) \\ g_2(azim) = \cos(azim) \end{cases} \quad (5)$$

这是由于以上两函数当  $0^\circ < azim < 360^\circ$  时, 与自变量之间存在着一一对应的关系, 而且二者相结合可以完全传递“偏角”这一变量所包含的方位信息.

综合以上分析, 我们设计的神经网络模型其输入输出层可用图 1 表示.

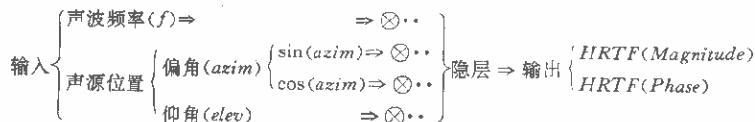


图 1 神经网络模型输入输出层示意图

### 2.2.3 网络训练时导师的选择

在进行网络训练时, 有必要对导师(即训练数据)进行选择. 我们的选择依据主要是人类的听觉感知模型. 根据 Oldfield 与 Parker<sup>[14]</sup> 对人的声音定位能力及定位精度的研究结果, 人类对耳平面上的声源进行定位的能力最强(即仰角  $= 0^\circ$ ); 随着仰角的增大, 声音定位能力要受到一定程度的影响.

这就是说, 随着声源仰角的增大, 人类对声源偏角及仰角的定位能力都随之下降; 当声源处于同一仰角平面时, 人类对前方声源的定位精度则要比后方高.

基于以上人类听觉感知模型, 导师选择策略应该是, 以高分辨率、高密度的训练样本对应人类声音定位能力强的空间区域, 而以低分辨率、低密度的训练样本对应人类定位能力弱的区域. 这样就将人类的听觉定位的感知模型与神经网络模型统一起来.

比如在耳平面上, 由于在该平面上人类的听觉定位能力最强, 相应地, 网络的训练样本可取得最密; 并且后半边的样本又可以比前半边取得稀疏一些.

### 2.2.4 数据的准备

由于训练所用的 HRTF 数据是在频域上的, 因此需要将测量获取的时域上的 HRTF 脉冲响应序列变换到频域上来. 另外, 因为在神经网络模型中采用了指数函数作为神经元的变换函数, 而该函数的值域为  $(0, 1)$ , 所以接着还必须对频域上的 HRTF 数据进行归一化处理, 使其满足网络的训练要求.

下面给出了数据准备的具体过程.

#### 算法 1. 神经网络模型训练数据的准备

输入: 对应于某声源位置的某对象的 HRTF 脉冲响应序列

输出: 相应的已经过归一化处理的 HRTF 频域数据(包括幅度响应  $M_i$  和相位响应  $P_i$ )

开始: (1) 傅里叶变换(FFT)

将时域数据通过 FFT 变换到频域上来(包括幅度响应数据  $M_i$  和相位响应数据  $P_i$ ).

(2) 幅度响应数据的归一化处理

设  $M_{\max} = \text{MAX}(M_i)$ ,  $M_{\min} = \text{MIN}(M_i)$ ,  $X_2$  趋近于 1,  $X_1$  趋近于 0

则  $M_i = (M_i - M_{\min}) * (X_2 - X_1) / (M_{\max} - M_{\min}) + X_1$

(3) 相位响应数据的归一化处理

设  $P_{\max} = \text{MAX}(P_i)$ ,  $P_{\min} = \text{MIN}(P_i)$ ,  $X_2$  趋近于 1,  $X_1$  趋近于 0

则  $P_i = (P_i - P_{\min}) * (X_2 - X_1) / (P_{\max} - P_{\min}) + X_1$

结束

### 2.2.5 网络隐层数目及隐层神经元数目的确定

现在已经发现,对于大多数实际问题,在多层前馈网络中一般都不需要使用 2 层或 2 层以上的中间层,业已证明,对于任一在闭区间内连续的函数都可以用 1 个隐层的多层前馈网络来逼近。<sup>[12]</sup>由于我们对 HRTF 的各种性质还不是很了解,合适的隐层数目只能通过实验来确定。

至于隐层神经元数目的选择则是一个非常经典而棘手的问题,也有一些学者<sup>[13]</sup>给出了具体的解析式,但是由于隐层神经元数目与具体的问题、输入输出单元的多少都有直接的关系,所以目前还没有具有普遍意义的推导公式,最佳隐层神经元数目的选择仍是一个有待解决的课题。

针对具体问题,我们采取了以下较费时但却操作方便,并且相当有效的选择策略:一开始,仅用单隐层且很小规模的神经元数目进行训练,根据学习结果,适当地增加隐层神经元数目,然后重新训练并测试;重复以上过程直至训练误差已可接受或者学习结果没有得到较显著地改善,如果是第 2 种情况,即说明用单隐层的网络(其隐层神经元数目已挺大)似乎不能很好地解决问题,则我们将增加一隐层进行训练,并试图减少隐层神经元的总数目。

### 2.2.6 网络初始值的选取及网络的学习策略

在解决实际问题过程中,神经网络的训练往往要花费相当长的时间,而且由于整个网络模型是非线性的,所以网络初始值的选取与学习是否能到达局部最小和是否能收敛的关系很大。

在选择与设计网络初始化及学习策略时,有两个基本点需要考虑:①网络的训练速度,通常的目标当然是在尽可能短的时间内获得较好的学习结果;②还需考虑在训练过程中如何避免陷入局部最小解的问题。

对于网络的基本训练算法,我们倾向于采用 Conjugate gradient 算法<sup>[15]</sup>,实践经验表明该算法的收敛速度要明显优于传统的一些训练算法,比如 Gradient descent 算法,另外我们还用模拟退火(Simulated Annealing)策略来进行网络连接权的初始化及训练过程中防止陷入局部最小解的工作。

图 2 结合网络隐层数和隐层神经元数目的选择,给出了神经网络模型的整个训练流程。

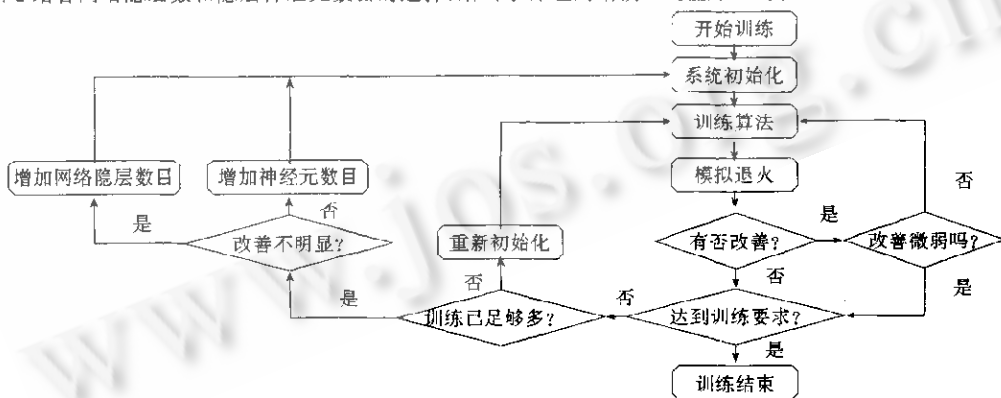


图2 神经网络模型训练流程图

## 3 基于神经网络模型的用于生成三维真实感声音的方法

根据不同的使用环境和要求,基于上述神经网络模型,我们设计了两种用于生成三维真实感声音的具体方法。

应该注意,下面所述方法都假设处理对象为 Windows 环境支持的标准的 WAV 声波文件,而且均认为初始的声波文件是单声道的,采样精度为 8 位或者 16 位,通过算法处理,根据声源相对于聆听者的位置(包括偏角和仰角),我们可以得到一个 16 位采样精度,并包含了空间信息的立体声 WAV 文件。

### 3.1 对于简单声音或谐波

如果需要处理的声波文件只包含很少几个频率的声波信号,或者甚至是一纯粹的单音或谐波信号,那么我们可以

用如文献[16]中所述的简化方式来进行处理.区别仅在于这里的算法采用了神经网络模型生成的 HRTF 数据进行计算,而不是原来所用的拟合公式;并且,现在这种简化方法已适用于整个三维空间,而不仅限于耳平面.

### 3.2 对于复杂声音或者在真实感要求更高的环境下

在一些场合,比如说 C<sup>3</sup>T 系统,有可能对声音的空间真实感提出更高的要求,因为在这种环境下感知上的错误会导致非常严重的后果.再加上现阶段我们仍没有较好的方法来准确地抽取相对复杂声波信号的基频<sup>[13]</sup>,这就使得以基频抽取为前提的简化方法失去了普遍适用的基础.同时测试结果也表明简化方法对复杂声波的处理效果并不是很理想,因此,基于文中提出的神经网络模型,下面给出另一种方法用于生成三维真实感声音,以求获得真实感方面更好的聆听效果.

#### 算法 2. 三维真实感声音的生成

输入: WAV 文件格式的单声道文件; 由偏角和仰角确定的声源位置

输出: 用于驱动耳机的立体声 WAV 文件

开始: (1)输出文件 WAV 头的准备

根据输出要求,生成输出文件的 WAV 头.

(2)获得相应位置的 HRTF 数据

根据输入的声源位置(包括偏角和仰角信息),训练过的神经网络模型可以生成一系列的 HRTF 数据(包括幅度响应和相位响应).

(3)逆傅里叶变换(IFFT)

由于在第(2)步获得的数据是频域上的,通过 IFFT,我们可以获得时域上相应空间位置的 HRTF 脉冲响应序列.

(4)卷积运算

以输入 WAV 文件数据和 HRTF 脉冲响应序列为对象,作一卷积运算.

(5)将计算结果写入输出文件并关闭所有打开文件

结束

## 4 测试与验证

我们设计了一个分成两个阶段实施的测试方案,首先对神经网络模型的运行性能进行评估,然后再对基于该神经网络模型生成的三维真实感声音进行声学及心理声学方面的验证.

### 4.1 网络性能评估

几乎所有的人工神经网络模型都采用最小平方差作为其基本的性能测试手段,同样,该测试策略也适于对非线性关系逼近类问题进行评估.

根据文中的训练流程,我们利用现有的 HRTF 数据进行了大量测试.表 1 给出了一测试实例,该实例所用训练数据集共为 86 组,代表  $azim$  从  $0^\circ \sim 15^\circ$ ,  $elev$  为  $0^\circ$  的 HRTF 频域数据.

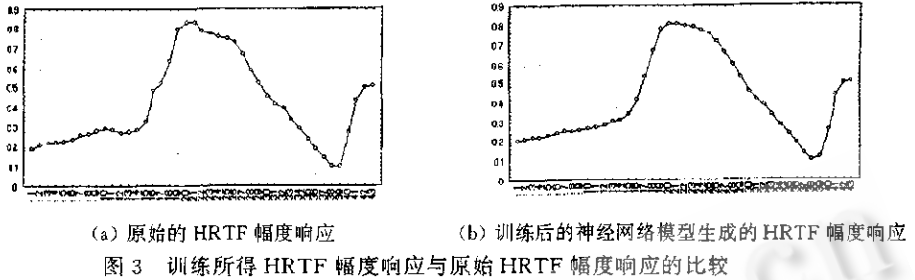
表 1 神经网络模型训练与测试实例

隐层神经元数目	训练误差	测试误差	测试误差平方根
15	0.150 708	0.001 993 2	0.044 645
19	0.121 943	0.001 330 1	0.036 471
23	0.105 548	0.001 209 2	0.034 774
27	0.090 468	0.001 208 3	0.034 751
19,8	0.028 875	0.000 335 0	0.018 028
17,10	0.034 47	0.000 604 7	0.024 591
15,8	0.028 964	0.000 376	0.018 373
17,6	0.034 139	0.000 622 3	0.024 916

上表第 1 列给出了网络所用的隐层神经元数目,如用逗号分开则表示网络有两个隐层,训练误差一列给出了网络的最佳训练结果,用整个训练集平方差的 100 倍表示;下一列给出网络采用测试集(包含了与训练集不同的 100 组相关 HRTF 频域数据)后的平方差;最后一列给出了上一列误差,即测试集平方误差的平方根数据.

根据实验结果,我们发现用两个隐层的网络进行训练似乎比单隐层的更有效.另外,测试集误差与训练集误差相比并没有显著地增长也说明我们的训练策略及数据表示是有效的.

图 3(a), (b) 分别显示了初始的 HRTF 幅度响应变化情况及训练后所得的相应幅度响应变化情况(两图都代表当  $azim=0^\circ$ ,  $elev=0^\circ$  时某对象左耳的频域 HRTF 幅度响应数据).



我们应该注意图 3 中的幅度变化值都已进行了归一化处理。

#### 4.2 声学及心理声学的验证

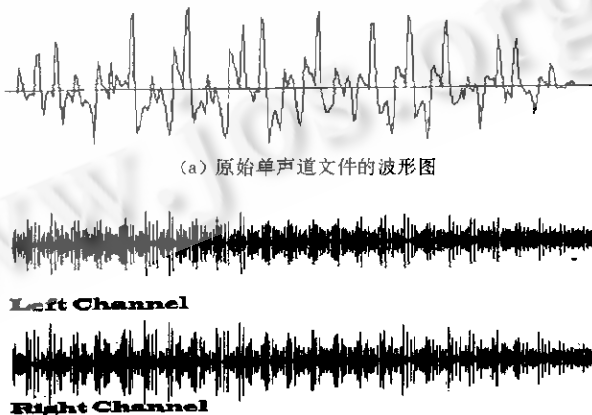
由于我们已获取了部分 HRTF 测量数据,所以声学验证只需面向由训练结束后的网络产生的中间数据(即未包括在原始数据中,当然更未包括在网络训练数据中的新产生数据)进行,这就相应地减轻了声学验证的工作量。

为了测试经过训练的神经网络模型是否能够正确生成所需的 HRTF 中间数据,我们进行了许多实验工作.然后根据上文提出的生成三维真实感声音的具体算法,利用神经网络产生的 HRTF 数据对原始的单声道 WAV 文件进行处理以获得包含空间信息的声波文件.结果表明效果是相当不错的。

测试者被分成两组,第 1 组为比较测试组,测试者首先听到两组边界声音信号(所谓边界声音信号,是指用来生成该段声音的对应 HRTF 数据是网络的两组训练数据,比如仰角=0°,偏角=5°;仰角=0°,偏角=15°的两组 HRTF 数据).然后我们利用训练后的网络生成的一组或几组中间 HRTF 数据(比如对应仰角=0°,偏角=10°的一组中间 HRTF 数据),按照上文中给出的三维真实感声音生成算法对同样的声音信号进行处理,然后由测试者聆听.初步的实验结果表明,在耳平面上时,当两段边界声音对应偏角差值大于 8°~10°,测试者对中间声音信号一般均有认同感;竖直方向上,只有当仰角差值大于 15°时,测试者才能获得比较明显的中间感.并且随着仰角的不断增大,测试者所需边界声音的偏角差值与仰角差值也需相应增大。

第 2 组测试者只需聆听由中间 HRTF 数据生成的一段或多段声音信号,然后要求他们说出对虚拟声源仰角和偏角的判断.结果表明测试者对由原始 HRTF 数据和新生成 HRTF 数据所生成的声音信号的定位精度没有本质差别。

图 4 给出了一个单声道波形文件用网络生成数据处理之后所得立体声 WAV 文件的对应波形。



(b) 生成的立体声文件的对应波形( $azim=10^\circ, elev=0^\circ$ )

图 4 波形文件实例

### 5 结 论

通过本文,我们在三维真实感声音生成领域引入了人工神经网络的方法,而且基于一种改进的神经网络模型,设计并实现了两种用于生成三维真实感声音的具体算法.初步的实验和调查表明该神经网络模型从根本上来说是

效的.

### 参考文献

- 1 Zhang Qiong, Zhao Yu, Xiang Hui *et al.* The research on the methods for generating 3-D realistic sound in HCL. *Journal of the Computer-Aided Design and Computer Graphics*, 1996,8(Special Issue on Computer Graphics):68~73
- 2 Blauert J. *Spatial hearing: the psychophysics of human sound localization*. Cambridge, MA: MIT Press, 1983
- 3 Wenzel E M, Arruda M *et al.* Localization using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America*, 1993,94:111~123
- 4 Mills W. Auditory localization. In: *Foundations of Modern Auditory Theory*. New York: Academic Press, 1972
- 5 Mehrgardt S, Mellert V. Transformation characteristics of the external human ear. *Journal of the Acoustical Society of America*, 1977,61:1567~1573
- 6 Wightman W L, Kistler D J. Headphone simulation of freefield listening. *Journal of the Acoustical Society of America*, 1989, 85:858~878
- 7 Martens W L. Principal components analysis and resynthesis of spectral cues to perceived direction. In: *Proceedings of the 1987 International Computer Conference*. San Francisco, 1987
- 8 Kistler D J. A model of head-related-transfer functions based on principal components analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America*, 1992,91:1637~1647
- 9 Moore F R. *Elements of Computer Music*. NJ, Prentice-Hall, 1990
- 10 Hornik K M, Stinchcombe M *et al.* Multilayer feedward networks are universal approximators. *Neural Networks*, 1989,2:359~366
- 11 Blum, Edward *et al.* Approximation theory and feedward networks. *Neural Networks*, 1991,4:511~515
- 12 Hecht-Nielsen R. Theory of the back propagation neural network. In: *Proceedings of the International Joint Conference on Neural Network*, 1989,1:593~603
- 13 Mirchandani G, Cao W. On hidden nodes for neural nets. *IEEE Transactions on CAS*, 1989,36:661~664
- 14 Oldfield S R, Parker S P A. Acuity of sound localisation: a topography of auditory space. I. normal hearing conditions. *Perception*, 1984,13:581~600
- 15 Press, William H *et al.* *Numerical recipes in C*. New York: Cambridge University Press, 1988
- 16 Zhao Yu, Zhang Qiong, Xiang Hui *et al.* A Simplified model for generating 3-D realistic sound in the multimedia and virtual reality system. *Journal of Computer Science and Technology*, 1996,11(5):461~470

## An Artificial Neural Network Method in Acoustical Simulation

ZHANG Qiong SHI Jiao-ying

(State Key Laboratory of CAD&CG Zhejiang University Hangzhou 310027)

**Abstract** Although most scholars have realized the importance of 3-D realistic sound embedded in future human-computer interaction, there are still many obstacles to overcome. After intensive investigation of the current research situations, an improved HRTF (head-related transfer function) model on the basis of neural network method was introduced to generate 3-D realistic sound. The HRTFs data employed can be reset by network training to satisfy a variety of requirements. Furthermore, with the utilization of neural network's nonlinearity approximation ability, all the HRTFs data (including magnitude characteristics and phase characteristics) at any incident angles can be obtained based on discrete sample values. Experiment and verification demonstrate the effectiveness and correctness of this scheme.

**Key words** Artificial neural network, HRTF(head-related transfer function), 3-D realistic sound.

**Class number** TP391