# 一种头相关传输函数的双维度稀疏压缩方法[\*](#项目信息)

张　斌1，方　勇1，刘华平1，王红梅2

(1.上海大学 通信与信息工程学院，上海 200444；2.浙江天格信息技术有限公司，浙江金华 321000)

**摘 要**：头相关传输函数（HRTF）描述了人耳空间听觉特征，但其巨大的数据量是影响空间信息实时重现的主要障碍。目前的压缩方法主要是基于主成分分析（PCA）在方向维度的压缩。本文分析了方向维度压缩信号的稀疏性，提出在方向维度压缩后，再利用压缩感知进行时间维度的压缩方法。在重构误差仅增加0.2%的前提下，数据压缩率较PCA方法可提升11%左右。

**关键词**：头相关传输函数；数据降维；主成分分析；压缩感知

**【中图分类号】**TN912.1 【**文献标识码】**A

**A Two Dimensions Compression Method for Head-related Transfer Functions**

ZHANG Bin1, FANG Yong1, LIU Hua-ping1, WANG Hong-mei2

（1. School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China ;

2. Zhejiang Tiange Information Technology Co., Ltd, Jinhua Zhejiang 321000, China）

**Abstract:** Head-related transfer function (HRTF) is used to represent the characteristics of human spatial hearing. However, its huge amount of data is the main impediment to real time reconstruction of spatial information. The principal component analysis (PCA) is used to compresse HRTF in direction dimension. In this paper, the sparsity of the compressed signal in direction dimension is analysized, and a novel compression method is proposed by utilizing compressive sensing theory to recompression in time dimension for the HRTF, which has been compressed in direction dimension. With reconstruction error only increased by 0.2%, the proposed algorithem can promote the compression ratio by 11% compareing with the PCA method.

**Key words:** HRTF; Dimensionality reduction; Principal component analysis; Compressive sensing

**1**引言

[[1]](#footnote-1) 人的听觉系统可以感受到来自四面八方的声音信号，但是一般的声音重放系统并不能够将声音的方位信息重现出来。三维音频（3D audio）技术是根据人耳对声音信号的感知特点，使用信号处理方法对声源到两耳之间的传递过程进行模拟，从而重现出声音的空间信息。头相关传输函数（HRTF，Head Related Transfer Function）是人耳空间听觉特性的一种表示方法。HRTF以冲激响应函数的形式，描述了不同方位的声音在耳朵、头部、躯干等身体部位的反射、折射和衍射的综合作用，它包含了双耳时间差（ITD）、双耳声级差（ILD）、高频谱因素等声源定位信息，所以将声音信号与一定方向的HRTF进行卷积处理，就可以认为对声音信号加入了方向信息[[1](#文献1),[2](#文献2)]。

为了产生任意方位的虚拟声像或模拟连续运动的声像，需要存储和调用大量的HRTF数据，由于HRTF巨大的数据量，虚拟声技术的应用受到了限制。一个重要的方法便是对高维的HRTF数据进行压缩。针对减小HRTF的数据存储量的研究已有很多，其大致可以分为三种。第一种是对确定方向的HRTF进行精确的声学建模[[3](#文献3),[4](#文献4)]，这种方法只需很小的存储量，但是由于人体结构的复杂性，其定位效果还有待检验。第二种方法是对测量得到的HRTF进行零极点建模，就是用低阶零极点模型来代表HRTF[[5](#文献5)]。第三种是利用数学分解的方法把HRTF用一些列的基函数的权重组合来表示。作为第三种方法的典型例子，Kistler和Wightman运用主成分分析（PCA）方法在对数尺度上对HRTF的幅值进行了分析，并将其表示成若干基函数的线性组合[[6](#文献6)]。这种方法充分考虑了HRTF数据之间的相似性，所以第三种方法的数据压缩性能是这三种方法中最好的，但是处理后的数据量仍然很大。目前的PCA降维处理是在方向维度上进行压缩的，而我们在进行HRTF

数据分析时发现，HRTF数据在时间维度上是稀疏的，是可以在时间维度进行压缩处理的。本文

提出了一种从方向维度和时间维度两个方面对HRTF数据进行压缩处理方法，通过对PCA压缩

信号的稀疏化，利用压缩感知方法在时间维度上进一步压缩。各项实验结果表明，比以往的单独用PCA进行降维的方法，本文提出的降维方法可以在重构误差仅增大0.2%的前提下使数据压缩率提高了11%左右。

**2**HRTF的主成分矩阵的稀疏化

在HRTF的研究中，习惯采用以下的坐标系统：定义头中心为三维坐标的原点，声源与原点的距离为，声源的空间方向由，，其中为水平面，为正上方，（）代表水平面正右方向，（）代表水平面正左方向，即得在自由场的情况下，HRTF的频域定义为





其中，，分别是声源在倾听者左、右耳处产生的复数声压，它的测量点可以是从耳道入口到鼓膜的任意截面；是头部移开后原点位置处的复数声压，是声波的频率；由于HRTF还和生理特征（如头的大小和形状）有关，所以用来表示个性化参数。HRTF的时域表述称为头相关脉冲响应（Head-Related Impulse Response，HRIR），它与头相关传输函数互为傅里叶变换对。

目前比较常用的HRTF数据库是MIT媒体实验室的数据库[[7](#文献7)]，它是使用Kemar假人头模型进行的测量，测量结果是时域的HRIR，采用44.1KHz的采样率，测量的水平方位角范围在到，每隔测量一次，俯仰角范围在到，以为间隔，一共测量了710个离散位置，每个位置进行多次测量然后取平均值。每个角度的HRIR数据实际上是一段长度为512个采样点的PCM文件。我们以水平面上的HRIR数据为例，通过以上介绍可以得出，水平面上的左右耳HRIR数据共有个，如果不对其进行压缩而直接应用的话，其大数据量会影响信号处理的时延，而声音信号处理对时延是比较敏感的，所以需要对HRIR数据进行压缩处理。

从式（1）和式（2）可知HRIR数据是方向和时间的函数，由于HRIR数据是每隔一个比较小的角度测量一次，所以相邻或者距离比较近的方向位置上的HRIR数据之间并不是不相关的，存在着一定的相关性，也就是说数据是有冗余的，因此目前使用的PCA分析就是将这些冗余去掉，以少量数据的组合来表示原始数据，达到降维的目的。

如果不考虑HRTF的个性化特性，只讨论远场HRTF，则可以略去变量和；为简单起见，空间方向仅用一个变量来表示（实际上是和的函数），因而可以将HRTF理解为方向变量和频率的连续函数。考虑到对左、右耳HRTF的讨论与前类似，所以略去表示左、右耳的下标，则HRTF可简记为。类似的，时域HRIR可简记为。

文献[8]指出，HRTF的PCA处理可以在时间域进行，也可以在复频域进行，两者完全等价。因此本文是在时间域也就是对HRIR进行PCA处理。以MIT媒体实验室的HRTF数据库为例，只取水平面的数据，即所对应的数据集。时域HRIR表示为



其中，为水平方位角的个数，为每个HRIR对应的样本数。对HRIR进行PCA的步骤可以描述为：（1）先求出去均值后的HRIR，；（2）求的自协方差矩阵；（3）求出的特征值和对应的特征向量，并按照从大到小的顺序排列；（4）取个主成分及其对应的特征向量；由正交变换矩阵系数用权向量表示：



重构数据可以表示为：



当选择的主成分个数时，式（5）可以完全恢复出原始的HRIR波形；当选取前个（）主成分时，式（5）得到的是的近似。我们只需要存储主成分矩阵和权向量矩阵就可以恢复出原始HRIR数据，以主成分数为例，经过压缩后的数据量为22980（），压缩率为，从而从方向维度上实现了数据的压缩。

PCA降维后得到的主成分矩阵是个长度为的列向量所组成的，以为例，主成分矩阵的每一列都是一个主成分分量，图1所示为第一到第五个主成分（PC1~PC5）的时域波形，可以看出，其在大部分的时间上是没有幅值或者幅值很小，只有部分时间是有值的，所以他是一个稀疏信号，可以从时间维度上在对主成分矩阵进行压缩处理。



图1 PC1~PC5的时域波形

与时域信号在频域中会表现出不同的形态一样，信号的稀疏度在不同变换域也会有很大差别，如果在某个变换域的稀疏度大的话那么压缩效果会更好。我们将主成分矩阵变换到常见的离散傅里叶（DFT）域，离散余弦（DCT）域和离散小波（DWT）域，观察变换后的到的系数，实验表明主成分分量在时域和DWT域上具有明显的稀疏性。

我们将第一个主成分PC1分别变换到DFT域、DWT域和DCT域，可以看到将其变换到DWT域时（图2（b）），变换系数的能量主要集中在少数系数上，绝大部分系数的幅值为零或者接近零，表现出良好的稀疏性；原始时域波形也具有一定的稀疏性，但不如DWT域明显；而DFT域和DCT域的变换系数稀疏性更差。主成分矩阵经DWT稀疏变换后可表示为，其中，为某个正交变换矩阵，是原信号在变换域的系数向量。

**3** 时间维度上HRTF主成分矩阵

的压缩

针对PCA处理后的主成分矩阵的稀疏表示，我们采用压缩感知理论(Compressed Sensing)[[9](#文献9)]在时间维度进行压缩处理。

对于稀疏信号，在某一测量矩阵（）的作用下产生一组线性测量值，



式（6）可看做是原信号在下的线性投影，由于，所以测量值可以表示为：





图2 不同变换域PC1的稀疏度，  
（a）PC1的时域波形（b）PC1的DWT域波形（c）PC1的DFT域波形（d）PC1的DCT域波形。

当矩阵满足约束等距性（RIP）时，可以用以下优化的方法来重构式（7）



在实际中常采用范数来代替范数，而且理论上证明当满足一定条件时，和优化问题同解。



式（7）可以通过LP（Linear Programming）求解。

测量矩阵使用高斯随机矩阵，因为高斯随机矩阵可以以很大的概率满足RIP性质，这一性质决定了其与大多数固定正交基构成的变换矩阵不相关。重构算法采用块稀疏贝叶斯学习（Block sparse Bayesian learning，BSBL）算法[[10](#文献10)]。因为从PC1分量的时域特别是DWT域系数可以看出，幅值较大的系数可以看成是成块出现的，某些区块内的系数是稀疏的，某些区块内的系数是不稀疏的，所以满足信号的块稀疏特性，采用针对块稀疏信号的重构算法会有较好的性能。

所以PCA和压缩感知结合的HRTF压缩方法步骤如下：

① 对HRIR数据进行PCA处理，得到主成分矩阵和方向权向量矩阵。

② 将主成分矩阵进行离散小波变换，得到。

③ 将高斯随机矩阵分别与相乘，得到测量值。

④ 用重构算法分别对测量值进行重构，并对其进行离散小波反变换，重构出的主成分矩阵记为。

⑤ 利用重构出原始的HRIR数据。

在对主成分矩阵在时间上利用压缩感知进行压缩的过程中，会有或多或少的数据损失，所以重构的数据会有一定的误差。由上所述的压缩方法可知，原始数据是通过两次重构后得出的，每次的重构都会有一定的重构误差，最终的重构误差并不是压缩感知重构误差和PCA重构误差的简单叠加，而是一个非线性过程。由后面的实验可以得出，当压缩感知重构误差小于某个阈值时，其对整体重构误差的影响很小，但是一旦压缩感知重构误差超过某个阈值，其对整体重构误差的影响就很大。

**4** 实验结果及分析

为了判断不同位置的HRIR的重构效果，引入均方根误差（RMSE）



其中为第次试验的重构信号，为原始信号，是实验的次数。

定义压缩比为



本实验主要研究的是压缩感知的重构误差对整体重构误差的影响，压缩感知的重构误差与稀疏变换矩阵，压缩比（测量矩阵的维数）和重构算法有关。所以本实验就从这三个方面来分析重构误差影响。稀疏变换矩阵采用DWT变换和时域对比；压缩比的变化直接影响着压缩感知重构误差的大小，压缩比越小代表着经过压缩感知方法处理后的数据量越小，同时压缩感知的重构误差也就会越大，所以压缩比对整体重构误差的影响就代表着压缩感知重构误差对整体重构误差的影响；重构算法为OMP算法、CoSaMP算法、BP算法和BSBL-BO算法的对比。实验选取的信号是HRIR水平面上的数据进行PCA处理后得到的主成分矩阵的第一个主成分向量PC1，信号长度。

当压缩比和重构算法一定时，对比时域信号和经过DWT域稀疏变换后的信号的重构误差。由图3所示，蓝色曲线是时域信号的整体重构误差和压缩比的变化图，红色曲线是经过DWT变换后的整体误差和压缩比的变化图，可以看出经过DWT稀疏变换后的重构误差要比没有经过变换的重构误差低；还可以看出，BSBL-BO算法相比其他三种算法对稀疏变换的敏感程度要低，这说明了这种算法具有很好的鲁棒性。

压缩比和重构算法对整体重构误差的影响可以从图4得出。从图中可以看出，随着压缩比的增大，也就是随着压缩感知重构误差的减小，整体重构误差呈减小的趋势，最终趋近于只用PCA方法的重构误差。当主元数为20时，只用PCA方法的重构误差为0.072，当压缩比在0.6~0.9时，BSBL-BO算法的整体重构误差最小，其值在0.079~0.072之间，与PCA方法的重构误差相差无几。 但是当压缩比小于0.4时，四种方法的整



图3 稀疏变换域对整体重构误差的影响，  
（a）OMP算法的RMSE（b）CaSaMP算法的RMSE（c）BP算法的RMSE（d）BSBL-BO算法的RMSE。

体重构误差都在0.1以上，比PCA方法的重构误差增大了38%以上，这是由于PCA重构时对压缩感知重构误差有放大作用，这种放大作用就导致了压缩比不能无限的小，当压缩比小到一定程度时（即压缩感知重构误差大到一定程度时），PCA重构出的信号就会有很大的误差，导致数据不可用。

当压缩比一定时可以看出，四种算法中BSBL-BO算法的RMSE曲线最低，重构误差最小，这也验证了PC1信号的块稀疏性，所以针对块稀疏信号重构的BSBL-BO算法的效果才最好。对比图4的四条曲线可以发现，BSBL-BO算法的重构误差也是最低的，并且和不经过稀疏变换的RMSE曲线基本一致，这说明BSBL-BO算法可以很好的克服信号稀疏度的影响。



图4 压缩比和重构算法对整体重构误差的影响

经过以上的实验可以得出当压缩比为0.6时，整体重构误差为0.079，此时的压缩比和整体重构误差有着很好的平衡。这种情况下所压缩出的数据量为，压缩率为，比PCA算法压缩率提高了11%左右。

**5** 结论

为了克服合成三维音频时HRTF数据量过大的缺点，本文从方向和时间两个方面，将新兴的压缩感知理论与经典的主成分分析方法相结合，用于HRTF数据的压缩降维。仿真实验表明：和以往的单独用主成分分析法对HRTF数据进行降维相比，该方法可以在重构误差只增加0.2%的前提下，将HRTF数据压缩率提高了11%左右，使其更好的应用于移动便携平台上。

**参考文献：**

1. Spors S, Wierstorf H, Raake A, et al. Spatial sound with loudspeakers and its perception: A review of the current state[J]. Proceedings of the IEEE, 2013, 101(9): 1920-1938.
2. Algazi V R, Duda R O. Headphone-based spatial sound[J]. Signal Processing Magazine, IEEE, 2011, 28(1): 33-42.
3. Algazi V R, Duda R O, Duraiswami R, et al. Approximating the head-related transfer function using simple geometric models of the head and torso[J]. The Journal of the Acoustical Society of America, 2002, 112(5): 2053-2064.
4. Xiao T, Liu Q H. Finite difference computation of head-related transfer function for human hearing[J]. The Journal of the Acoustical Society of America, 2003, 113(5): 2434-2441.
5. Kulkarni A, Colburn H S. Infinite-impulse-response models of the head-related transfer function[J]. The Journal of the Acoustical Society of America, 2004, 115(4): 1714-1728.
6. Kistler D J, Wightman F L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction[J]. The Journal of the Acoustical Society of America, 1992, 91(3): 1637-1647.
7. Gardner B, Martin K. HRTF measurements of a KEMAR dummy-head microphone[J]. Massachusetts Institute of Technology, 1994, 280(280): 1-7.
8. Rao D, Xie B. The equivalence between principal components analysis of head-related transfer functions in the time domain and frequency domain[C]//Image and Signal Processing (CISP), 2010 3rd International Congress on. IEEE, 2010, 8: 3895-3898.
9. Donoho D L. Compressed sensing[J]. Information Theory, IEEE Transactions on, 2006, 52(4): 1289-1306.
10. Zhang Z, Jung T P, Makeig S, et al. Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning[J]. Biomedical Engineering, IEEE Transactions on, 2013, 60(2): 300-309.

1. \*金华市科技计划重点项目(2013-1-007) [↑](#footnote-ref-1)