基于局部最小二乘支持向量机的音频频带扩展方法

白海钏,鲍长春,刘 鑫
(北京工业大学电子信息与控制工程学院,北京 100124)

摘 要: 在网络传输过程中宽带音频会由于高频信息的缺失导致音频质量下降,因此,本文提出了一种基于局部最小二乘支持向量机的宽带向超宽带音频频带扩展方法.根据音频频域序列的非线性特性,本文采用相空间重构和局部最小二乘支持向量机对音频信号的高频频谱细节进行预测,并结合高斯混合模型对高频子带能量进行估计,最后经过高频频谱包络调整,所提方法能够有效地恢复7kHz~14kHz频率范围内的高频成分.主客观测试结果表明,该方法改善了宽带音频的听觉质量,其性能优于参考音频频带扩展方法.

关键词: 音频编码;频带扩展;高斯混合模型;局部最小二乘支持向量机
中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2016)09-2203-08
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2016.09.027

Audio Bandwidth Extension Method Based on Local Least Square Support Vector Machine

BAI Hai-chuan, BAO Chang-chun, LIU Xin
(*School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing* 100124, *China*)

Abstract: The auditory quality of wideband audio is generally degraded due to the lack of the high-frequency in network transmission, so this paper presents a kind of audio bandwidth extension method from wideband to super wideband based on local least square support vector machine. In the light of the nonlinearity of audio spectrum, the high-frequency fine spectrum of audio signals is predicted by using phase space reconstruction and local least square support vector machine. Combining with the estimation of high-frequency sub-band energy based on Gaussian mixture model, the proposed method can effectively recover the high-frequency components in the frequency range 7kHz ~ 14kHz through the envelope adjustment of high-frequency spectrum at last. Subjective and objective testing results indicate that the proposed method improves the auditory quality of wideband audio and outperforms the reference methods of audio bandwidth extension.

Key words: audio coding; bandwidth extension; Gaussian mixture model; local least square support vector machine

1 引言

在现阶段音频通信传输系统中,为了在保证音频主观质量的前提下同时提高信号传输效率,感知音频编码方法通常优先对音频信号的低频信息进行恢复.现有音频通信网络将传输宽带音频信号的有效带宽限制在50Hz~7kHz范围内,其采样率为16kHz.与32kHz采样、频带范围在50Hz~14kHz的超宽带音频相比,宽带音频在传输和存储过程中丢失了7kHz以上的高频部分,因此其重建音频的自然度和表现力有所欠缺[1].但是超宽带音频信号的传输会导致处理数据量的增加,并需要更为先进的网络设备.所以在不改变现有通信设备且不增加网络负担的前提下,本文采用频带扩展方法在接收端对重建的宽带音频人为地恢复所截去的高频成分,从而达到增强听觉质量,重现超宽带音频的目的[2].

目前音频编码标准中通常适当增加丢失频带的边信息,对宽带音频进行非盲目式频带扩展.这类方法需要在编码端计算音频时频能量,并根据高低频频谱之间的相关性来确定适当的频谱修补方法,最后将这两

收稿日期:2014-10-10;修回日期:2014-11-05;责任编辑:梅志强
基金项目:国家自然科学基金项目(No.61072089,No.61471014)

部分信息一并量化传输到解码端,从而近似重建高频成分,获得具有更强层次感和更加自然透明的主观音频质量.该方法的重建音质较好,然而却需要提供大量先验信息,会增加编解码端和网络设备的数据处理负担,实用性不佳.鉴于此,本文采用盲目式频带扩展方法,在不传输额外信息的前提下,实现超宽带音频信号的重现.

传统盲目式方法主要针对频谱包络和频谱细节两部分进行频带扩展.频谱包络估计的准确性直接影响重建音频的主观质量,目前频谱包络估计算法主要包括码书映射、高斯混合模型(Gaussian Mixture Model,GMM)[3,4]、隐马尔科夫模型[5]以及神经网络[6]等方法.而频谱细节扩展则源自音频信号"谐波+噪声"模型.其中,频谱翻折和频谱搬移方法将低频频谱细节直接翻折或搬移到高频成分[7],G.722.1C 音频编码器采用噪声填充来恢复高频子带中丢失的精细结构,而谐波频带扩展方法利用频谱拉伸将低频频谱扩展到高八度音来重建部分高频谐波[8],上述方法均未考虑高频频谱特征以及高低频频谱之间的相关性,因而会影响音频信号的层次感和自然度,尤其是重建音频在高低频衔接处发生的频谱偏移,会导致听觉感受不平滑或产生频谱失真现象.

以上频谱细节重建方法主要针对音调的高频谐波部分进行修复,而对于噪声信息的高频部分则保持其随机结构.但实际音乐信号的频谱特征比较复杂,在共振腔中声音发生的共振辐射会改变其高频谐波结构,因此上述频带扩展方法势必会有一定程度的预测偏差.而本课题组前期工作中验证了音频频谱具有一定的非线性特性,并将非线性预测引入了频谱细节扩展中,如文献[9]首先利用相空间重构将一维频域序列转换到多维相空间中,然后在此相空间中建立最近邻映射(Nearest Neighbor Mapping,NNM)模型来描述高低频频谱之间联系,最后根据低频相点的演变规律来对高频相点的运动轨迹进行预测,从而完成高频频谱细节的恢复.然而,实际音频频谱夹杂着某些类噪声成分,会影响 NNM 预测准确性,导致重建音频的主观听觉质量降低.据此,本文提出了一种基于局部最小二乘支持向量机(Least Squares Support Vector Machine,LS-SVM)的音频频带扩展方法,在对频域序列进行相空间重构的基础上,根据低频相矢量集合采用局部 LS-SVM 实现对高频相轨迹的非线性预测,从而完成高频频谱细节的逐点恢复.同时,该方法采用 GMM 来对高频子带能量和宽带音频时频特征之间的联合概率密度进行拟合,在最小均方误差准则下实现了对高频频谱包络信息的有效估计.最终,重建的高频成分和原始的宽带音频相结合,实现了宽带音频向超宽带音频的盲目式频带扩展.

2 基于局部 LS-SVM 的音频频带扩展

本文所提频带扩展方法原理如图 1 所示.该方法采用有效带宽 7kHz 采样率 16kHz 的宽带音频信号作为其输入信号,它通过上采样和低通滤波器后,可获得同样 7kHz 有效带宽而采样率为 32kHz 的滤波信号 $x(n)$.首先,将 $x(n)$ 按照 20ms 分帧,并选取调制重叠变换(Modulated Lapped Transform,MLT)方法对音频信号进行时频分析,得到音频信号的频域序列.MLT 的时间叠接窗长为 40ms,所以在时频分析时需将上一帧和本帧数据共 1280 个采样点一起进行 MLT 变换,得到 640 个频谱参数 $C_{\mathrm{mlt}}(i)$,$i=0,1,\cdots,639$ 来表示 0~16kHz 的频谱信息.MLT 变换公式如下:

$$C_{\mathrm{mlt}}(i)=\sum_{r=0}^{1279}\sqrt{\frac{2}{640}}\sin\left(\frac{\pi}{1280}(r+0.5)\right)\times\cos\left(\frac{\pi}{640}(r-319.5)(i+0.5)\right)x(r)\quad(1)$$

由于输入的时域信号和滤波信号有效带宽均为 7kHz,因此得到的 640 个 $C_{\mathrm{mlt}}(i)$ 频谱参数中仅有前 280 点有实际幅度值,其他参数幅度值为 0.然后,将这 280 个 $C_{\mathrm{mlt}}(i)$ 参数进行子带划分,每个子带由 20 个频点构成,可得到 14 个子带.接下来,分别计算每个子带均方根能量 $e_{\mathrm{rms}}(n)$,$n=0,1,\cdots,13$ 来表示音频低频频谱包络信息,如下式所示

$$e_{\mathrm{rms}}(n)=\sqrt{\frac{1}{20}\sum_{r=0}^{19}C_{\mathrm{mlt}}(20n+r)C_{\mathrm{mlt}}(20n+r)}\quad(2)$$

在频谱包络估计模块中,本文将采用传统 GMM 方法根据 7kHz 以下的低频能量信息来估计 7kHz~14kHz 的高频子带均方根能量,从而实现高频频谱包络估计.

根据上面得到的子带均方根能量,本文采用归一化的 MLT 频谱参数 $C_{\mathrm{norm}}(i)$ 来表示频谱细节信息,即

$$C_{\mathrm{norm}}(i)=\frac{C_{\mathrm{mlt}}(i)}{e_{\mathrm{rms}}(n)},0\leqslant i<280,n=\lfloor i/20\rfloor\quad(3)$$

根据音频频谱序列的非线性特性,本文采用相空间重构将一维频谱细节序列转换到高维相空间中,并利用局部 LS-SVM 方法恢复高频频谱细节.最后,通过

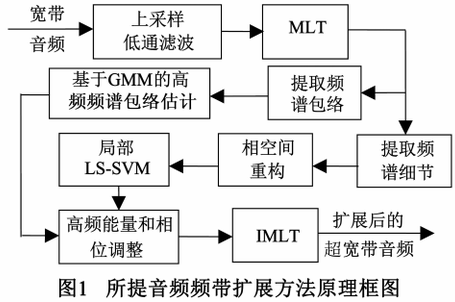


图1 所提音频频带扩展方法原理框图

谱包络调整,恢复高频频谱信息,并结合原始低频成分,借助 MLT 反变换(Inverse Modulated Lapped Transform, IMLT)得到有效带宽为 14kHz 采样率为 32kHz 的超宽带音频信号,实现完整的频带扩展. 具体步骤将在下文中进行详细阐述.

2.1　相空间重构

本课题组在前期工作中对音频信号的非线性特性进行了研究,借助相空间重构描述了音频信号频域相点的运动轨迹,并进一步利用基于最大李雅普诺夫指数的非线性分析方法验证了音频信号的频域序列具有非线性特性[10]. 在实际的音频频带扩展过程中,实验得到的一维音频频域序列无法直接反映音频频谱的非线性关系. 根据非线性动力学理论[11~14],本文将该一维序列通过延迟重建方法重构出与原始音频动力学系统拓扑等价的多维相空间,充分展示出音频频域系统所蕴含的非线性特性,并在重构的相空间中建立非线性数学模型,实现对高频频谱细节信息的非线性预测.

将一维音频频域序列表示为$\{C_{\text{norm}}(i)\}$,$i=0,1,\cdots,279$,通过相空间重构可以得到 m 维相点 $\boldsymbol{S}(i)$:

$$\boldsymbol{S}(i)=\{C_{\text{norm}}(i),C_{\text{norm}}(i+\tau),C_{\text{norm}}(i+2\tau),\cdots,C_{\text{norm}}(i+(m-1)\tau)\},i=0,\cdots,279-(m-1)\tau \tag{4}$$

其中,τ 和 m 是相空间重构方法中两个重要参数,分别代表延迟时间和嵌入维数. 根据所得到的相点,可以进一步得到音频频域序列重构的相空间,它可以看成是由 $279-(m-1)\tau$ 个相点所构成的相点集合,即:

$$\boldsymbol{S}=\begin{bmatrix} C_{\text{norm}}(0) & C_{\text{norm}}(1) & \cdots & C_{\text{norm}}(278-(m-1)\tau) & C_{\text{norm}}(279-(m-1)\tau) \\ C_{\text{norm}}(0+\tau) & C_{\text{norm}}(1+\tau) & \cdots & C_{\text{norm}}(278-(m-2)\tau) & C_{\text{norm}}(279-(m-2)\tau) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{\text{norm}}(0+(m-2)\tau) & C_{\text{norm}}(1+(m-2)\tau) & \cdots & C_{\text{norm}}(278-\tau) & C_{\text{norm}}(279-\tau) \\ C_{\text{norm}}(0+(m-1)\tau) & C_{\text{norm}}(1+(m-1)\tau) & \cdots & C_{\text{norm}}(278) & C_{\text{norm}}(279) \end{bmatrix} \tag{5}$$

在音频频域序列的相空间重构中,延迟时间和嵌入维数的确定至关重要,这两个参数的大小直接决定了所重构的相空间是否与原始非线性音频系统拓扑等价. 下面将对这两个参数的选取方法进行详细介绍.

2.1.1　延迟时间的选择

对于无限长、无噪声、无误差的理想观测序列,延迟时间可以任意选取. 然而,在实际应用中,所选取的序列不可避免地会受到背景噪声和计算误差的影响,且长度也有一定限制,因而需要人为选取延迟时间. 选取参数 τ 的基本原则是音频频域序列重构相空间中每个相点的任意两个相邻元素之间具有独立性但又不完全不相关. 该原则不仅保证了相点中每个元素可以作为构成相空间的独立坐标,同时保证了重构相空间能够呈现出原始音频系统的非线性频谱特征. 如果延迟时间过大,相点元素之间接近完全独立,过低的相关性不能充分描述音频频谱细节信息;反之,相点元素之间相关性过强,相轨迹过于集中,同样无法呈现出音频信号的频谱特性,反而在一定程度上增加了计算复杂度. 鉴于此,为了有效地反映音频频域序列相轨迹的真实演变规律,本文采用自相关方法对音频频域序列重构相空间的延迟时间进行适当选取[12].

自相关法在保证原始音频系统信息不会过多丢失的基础上,适当去除相点中相邻元素之间的线性相关性,它采用频域序列$\{C_{\text{norm}}(i)\}$在时间间隔 τ 下的归一化自相关函数来进行计算,如式(6)所示.

$$R(\tau)=\frac{\sum_{i=0}^{279-\tau}C_{\text{norm}}(i)C_{\text{norm}}(i+\tau)}{\sqrt{\sum_{i=0}^{279-\tau}C_{\text{norm}}{}^{2}(i)\sum_{i=0}^{279-\tau}C_{\text{norm}}{}^{2}(i+\tau)}} \tag{6}$$

一般情况下,$C_{\text{norm}}(i)$ 和 $C_{\text{norm}}(i+\tau)$ 之间的相关性会随着时间间隔的增大而逐渐降低. 而实际经验表明,当自相关函数下降为初始值 $R(0)$ 的 $(1-1/\mathrm{e})$ 倍,或第一次降低为零值,或下降到第一个极小值处时,可以得到最佳的延迟时间 τ.

2.1.2　嵌入维数的选择

根据 Takens 嵌入定理[14],当嵌入维数 $m\geqslant 2D+1$(其中 D 表示真实的空间维数)时,可以在重构的相空间内将原始动力学系统相轨迹的几何结构完全恢复出来. 该条件是保证非线性音频系统能够在重构相空间中完全展开的必要不充分条件,当选取嵌入维数过大时,能够保证相点轨迹完全打开,但是会增加轨迹预测和控制的计算量并且会放大噪声对系统性能的影响. 然而,当选取的嵌入维数过小时,相空间中相点的运动轨迹可能会发生交叠,无法借助该重构相空间分析原始音频系统的非线性特性. 因此,本文采用虚假近邻点法来适当选取嵌入维数,在保证相轨迹完全展开的基础上,进一步降低计算量和噪声的影响[10].

对于 d 维空间,每个音频频域序列的相点 $\boldsymbol{S}(i)$ 都有其对应的最近邻点 $\boldsymbol{S}(i')$,使两点间距离 $R_d(i)$ 最小:

$$R_d(i)=\|\boldsymbol{S}(i)-\boldsymbol{S}(i')\|_2^{(d)}=\min_{s=1,\cdots,N,s\neq i}\|\boldsymbol{S}(i)-\boldsymbol{S}(s)\|_2^{(d)} \tag{7}$$

其中,N 代表相空间中相点的个数,i'代表 $\boldsymbol{S}(i)$ 最近邻点所对应的序号.

在适当的嵌入空间中,原本相邻的近邻点在高维空间中可能不再是近邻点. 这类近邻点可以定义为虚假最近邻点. 随着嵌入维数的增加,相空间中的虚假最

近邻点会逐渐消失,当它的比例不再随嵌入维数的增加而变化时,则可以确定最优的嵌入维数.

虚假近邻点法的步骤如下:

步骤 1 据式(7),确定每个相点 $\boldsymbol{S}(i)$ 所对应的初始最近邻点 $\boldsymbol{S}'(i)$;

步骤 2 空间维数由 d 增加到 $d+1$,重新计算相点 $\boldsymbol{S}(i)$ 与其最近邻点 $\boldsymbol{S}'(i)$ 的距离,记为 $R_{d+1}(i)$;

步骤 3 满足下式中的条件,可以认为相点 $\boldsymbol{S}(i)$ 对应的最近邻点 $\boldsymbol{S}'(i)$ 为其虚假近邻点;

$$\left|\frac{R_d(i)-R_{d+1}(i)}{R_d(i)}\right|>R_{\mathrm{T}} \tag{8}$$

根据实际经验,阈值 R_{T} 选择 10%[12].

步骤 4 计算虚假近邻点占全部相点的比例;

步骤 5 判断上述比例是否小于一定的阈值或不再随着维数 d 的增加而减小,则认定相点轨迹已完全展开,结束循环并确定嵌入维数 $m=d$,否则重新从步骤 2 开始进行循环计算,直到满足截止条件.

本文通过上述方法,可以确定延迟时间和嵌入维数,并根据由这两个参数所构成的每个相点,根据式(5)对音频信号的频域序列进行相空间重构.

2.2 基于局部 LS-SVM 的高频频谱细节恢复

经过音频频域序列的相空间重构后,本文将采用局部 LS-SVM 方法来对高频相点的轨迹进行预测.

归一化 MLT 系数间的非线性函数关系可表示为,

$$\begin{aligned}C_{\mathrm{norm}}(i+1)&=F[S(i)]\\&=F[C_{\mathrm{norm}}(i),C_{\mathrm{norm}}(i+\tau),C_{\mathrm{norm}}(i+2\tau),\\&\quad\cdots,C_{\mathrm{norm}}(i+(m-1)\tau)]\end{aligned} \tag{9}$$

其中,i 代表归一化 MLT 系数的频谱序号.

$F[\cdot]$ 是一个非线性函数,它表示音频频域序列前一个相点与当前相点中最后一维 MLT 频域参数值之间的非线性关系,下面将采用支持向量回归机的方法对该非线性函数进行求取.

2.2.1 基于 LS-SVM 的非线性预测

根据支持向量回归机理论,相点 $\boldsymbol{S}(i)=\{C_{\mathrm{norm}}(i)\}$,$i=0,1,\cdots,279-(m-1)\tau$ 的估计函数可设为:

$$F(\boldsymbol{S}(i))=(\boldsymbol{w}_{\mathrm{s}}{}^{\mathrm{T}}\varphi(\boldsymbol{S}(i)))+b_{\mathrm{s}} \tag{10}$$

其中,$\varphi(\boldsymbol{S}(i))$ 代表进行非线性映射的核函数,它可以将输入相点 $\boldsymbol{S}(i)$ 由低维空间映射到高维空间,$\boldsymbol{w}_{\mathrm{s}}$ 和 b_{s} 分别代表权值和偏置量.

SVM 源于线性可分情况下的最优分类面,算法中要求该分类面能够将两类样本点无错误的分开,而且要使其分类空隙最大. 当样本满足线性可分时,其最优分类面为函数 $\varphi(\boldsymbol{w}_s)=\|\boldsymbol{w}_s\|^2/2$ 中最小的分类面,当样本线性不可分时,SVM 通过引入非负松弛变量,以求在错误最小的情况下将样本分离. 类似分类问题,用于回归的 SVM 算法综合考虑函数复杂度和拟合误差,引入了非负松弛变量 $\xi_i{}^*$ 和 ξ_i,并通过对目标函数进行最小化可以得到其最优化问题[15],如下式所示:

$$\min\frac{1}{2}\|\boldsymbol{w}_{\mathrm{s}}\|^2+C\sum_{i=1}^{T_n}(\xi_i+\xi_i{}^*) \tag{11}$$

约束条件为:

$$F(\boldsymbol{S}(i))-\boldsymbol{w}_{\mathrm{s}}{}^{\mathrm{T}}\varphi(\boldsymbol{S}(i))\leqslant\xi_i{}^*+\varepsilon,i=1,\cdots,T_n \tag{12}$$

$$\boldsymbol{w}_{\mathrm{s}}{}^{\mathrm{T}}\varphi(\boldsymbol{S}(i))-F(\boldsymbol{S}(i))\leqslant\xi_i+\varepsilon,i=1,\cdots,T_n \tag{13}$$

$$\xi_i{}^*,\xi_i\geq0,i=1,\cdots,T_n \tag{14}$$

其中,ε 代表不敏感损失参数,用于控制拟合精度,C 代表惩罚因子,用于控制对错分数据点的惩罚程度,$T_n=279-(m-1)\tau$ 代表训练样本的个数,即相点的个数.

以上的最优化问题可以通过标准的二次规划算法得到,并可以通过其对偶问题来进行求解. 然而,标准支持向量回归机的训练复杂度高,且样本数据越大,求解二次规划问题越复杂,所以本文考虑采用 LS-SVM 来对该非线性函数进行估计.

LS-SVM 与标准 SVM 不同之处在于它用训练误差的平方代替了松弛变量,并用等式约束代替了不等式约束[13],回归型 LS-SVM 同样利用高维特征空间里的线性函数来对样本集合进行拟合,其优化问题为:

$$\min\frac{1}{2}\sum_{i=1}^{T_n}\boldsymbol{w}_i^2+\frac{1}{2}\gamma\sum_{i=1}^{T_n}e_i^2,\gamma>0 \tag{15}$$

约束条件为:

$$F(\boldsymbol{S}(i))=\boldsymbol{w}_{\mathrm{s}}{}^{\mathrm{T}}\varphi(\boldsymbol{S}(i))+b_{\mathrm{s}}+e_i,i=1,2,\cdots,T_n \tag{16}$$

其中 γ 代表正则化参数,b_{s} 为偏置,$\boldsymbol{e}=[e_1,e_2,\cdots,e_{Tn}]^{\mathrm{T}}$.

为了求解上述最优化问题,建立 Lagrange 函数为

$$\begin{aligned}L(\boldsymbol{w}_{\mathrm{s}},b_{\mathrm{s}},\boldsymbol{e},\alpha)&=\frac{1}{2}\|\boldsymbol{w}_{\mathrm{s}}\|^2+\frac{1}{2}\gamma\sum_{i=1}^{l}e_i^2\\&\quad-\sum_{i=1}^{l}\alpha_i(\boldsymbol{w}_{\mathrm{s}}^{\mathrm{T}}\varphi(\boldsymbol{S}(i))+b_{\mathrm{s}}+e_i\\&\quad-F(\boldsymbol{S}(i)))\end{aligned} \tag{17}$$

式(17)中,α_i 代表 Lagrange 乘子. 函数对 $\boldsymbol{w}_{\mathrm{s}}$、$b_{\mathrm{s}}$、$\boldsymbol{e}$、$\alpha$ 求偏微分,得到上式的最优条件:

$$\frac{\partial L}{\partial\boldsymbol{w}_{\mathrm{s}}}=0\rightarrow\boldsymbol{w}_{\mathrm{s}}=\sum_{i=1}^{T_n}\alpha_i\varphi(\boldsymbol{S}(i)) \tag{18}$$

$$\frac{\partial L}{\partial b_{\mathrm{s}}}=0\rightarrow\sum_{i=1}^{T_n}\alpha_i=0 \tag{19}$$

$$\frac{\partial L}{\partial e_i}=0\rightarrow\gamma e_i=\alpha_i \tag{20}$$

$$\frac{\partial L}{\partial\alpha_i}=0\rightarrow\boldsymbol{w}_{\mathrm{s}}{}^{\mathrm{T}}\varphi(\boldsymbol{S}(i))+b_{\mathrm{s}}+e_i-F(\boldsymbol{S}(i))=0 \tag{21}$$

消去式中的 $\boldsymbol{w}_{\mathrm{s}}$ 和 $\boldsymbol{e}$,得到线性方程组:

$$\begin{bmatrix}0&\boldsymbol{1}^{\mathrm{T}}\\\boldsymbol{1}&\boldsymbol{\Omega}+\gamma^{-1}\boldsymbol{I}\end{bmatrix}\begin{bmatrix}b_{\mathrm{s}}\\\boldsymbol{\alpha}\end{bmatrix}=\begin{bmatrix}0\\F(\boldsymbol{S}(i))\end{bmatrix} \tag{22}$$

其中,$F(\boldsymbol{S}(i))=[F(\boldsymbol{S}(1)),\cdots,F(\boldsymbol{S}(T_n))]^{\mathrm{T}}$,$\boldsymbol{1}=$

$[1,1,\cdots,1]^{\mathrm{T}}$,$\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_l]^{\mathrm{T}}$,$\boldsymbol{\Omega}$ 是一个 $n\times n$ 的对称矩阵,$\boldsymbol{\Omega}_{ij}=\varphi(\boldsymbol{S}(i))^{\mathrm{T}}\varphi(\boldsymbol{S}(j))=K(\boldsymbol{S}(i),\boldsymbol{S}(j))$,$i,j=1,2,\cdots,T_n$. 令 $\boldsymbol{A}=\boldsymbol{\Omega}+\gamma^{-1}\boldsymbol{I}$,可得

$$b_{\mathrm{s}}=\frac{\mathbf{1}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{y}}{\mathbf{1}^{\mathrm{T}}\boldsymbol{A}^{-1}\mathbf{1}},\boldsymbol{\alpha}=\boldsymbol{A}^{-1}(\boldsymbol{y}-b_{\mathrm{s}}\mathbf{1}) \tag{23}$$

对上述方程组进行求解,可得到估计函数,进而可以基于 LS-SVM 获得高频频谱细节的估计值,如式(24)所示.

$$\begin{aligned}C_{\mathrm{norm}}(i+1)&=F(\boldsymbol{S}(i))=\boldsymbol{w}_{\mathrm{s}}^{\mathrm{T}}\varphi(\boldsymbol{S}(i))+b_{\mathrm{s}}\\&=\sum_{j=1}^{T_n}\alpha_j K(\boldsymbol{S}(i),\boldsymbol{S}(j))+b_{\mathrm{s}}\end{aligned} \tag{24}$$

在式(24)中,本文采用径向基核函数 $K(\boldsymbol{S}(i),\boldsymbol{S}(j))$ 替代内积计算 $\varphi(\boldsymbol{S}(i))^{\mathrm{T}}\varphi(\boldsymbol{S}(j))$,灵活处理了高维运算的问题:

$$K(\boldsymbol{S}(i),\boldsymbol{S}(j))=\exp\left(-\frac{\|\boldsymbol{S}(i)-\boldsymbol{S}(j)\|^2}{\sigma^2}\right) \tag{25}$$

根据上述回归 LS-SVM 的求解过程可知,该算法将标准支持向量机的二次规划的求解问题转化为运用最小二乘法解线性方程组的问题,所以 LS-SVM 的最大优势在于计算简便、明显提升了训练速率.

2.2.2　高频频谱细节恢复

根据上文推导得到的高频频谱细节预测公式,本文将对基于相空间重构和 LS-SVM 的高频频谱细节恢复方法进行详细介绍,其原理如图 2 所示. 所提方法输入信号为宽带音频的频谱细节信息,采用归一化的 MLT 系数来表示.

首先,对音频频域序列进行相空间重构,得到由归一化 MLT 频谱参数所构成的相点集合$\{\boldsymbol{S}(i),i=0,\cdots,279-(m-1)\tau\}$,作为 LS-SVM 预测模型的训练数据.

然后,依据所得低频相矢量集合,本文采用最小二乘法来训练模型参数,并构建基于 LS-SVM 的预测模型. 这里采用径向基核函数将输入相矢量通过非线性变换映射到高维空间,并在该空间内求取最优线性拟合函数,从而得到能够表示未知 MLT 系数与已知相矢量之间非线性函数关系的预测公式.

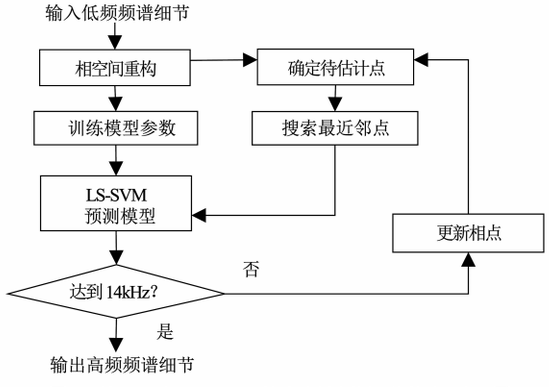


图2　基于局部LS-SVM的高频频谱细节恢复原理图

接下来,由当前已预测的高频相点来确定下一个估计点,并在低频相矢量集合中对该点的最近邻点进行搜索. 根据 NNM 原则,在相空间中近邻点之间通常遵循相近的演变轨迹,因此可以利用最近邻点附近的演变规律对待预测点进行近似估计.

最后,将搜索到的最近邻点作为基于 LS-SVM 非线性预测模型的输入,并利用其预测值作为高频相点归一化 MLT 频谱参数的估计. 更新相点并重复执行近邻点搜索和非线性预测,直到最终逐点恢复出 7kHz ~ 14kHz 频带范围内的高频频谱细节信息.

由于上述方法借助 NNM 原则仅在局部相空间中采用 LS-SVM 对相点进行了预测,因而本文称其为基于局部 LS-SVM 的高频频谱细节恢复方法.

2.3　基于 GMM 的高频频谱包络估计

本文选取传统 GMM 方法来估计高频子带能量,从而恢复高频频谱包络信息. 该方法可分为线下训练和参数估计两个阶段. 在线下训练阶段,利用具有 M 个高斯分量的 GMM 来近似拟合宽带时频特征 $\boldsymbol{F}_x$(包括过零率、梯度指数、子带均方根能量、子带通量、音频谱重心、音频扩展度以及音频谱平坦度)和高频子带能量 $\boldsymbol{F}_y$ 的联合概率密度,并以其作为高低频特征间的先验知识来指导高频频谱包络的贝叶斯估计[4].

高低频特征的联合概率密度 $p(\boldsymbol{F}_x,\boldsymbol{F}_y\mid\lambda)$ 可表示为:

$$p(\boldsymbol{F}_x,\boldsymbol{F}_y\mid\boldsymbol{\lambda})=\sum_{i=1}^{M}w_i p_{\mathrm{g}}(\boldsymbol{F}_x,\boldsymbol{F}_y,\boldsymbol{m}_i,\boldsymbol{C}_i) \tag{26}$$

其中,$M=64$ 为高斯分量的个数,$p_{\mathrm{g}}(\boldsymbol{F}_x,\boldsymbol{F}_y,\boldsymbol{m}_i,\boldsymbol{C}_i)$ 为第 i 个高斯分量的联合概率密度,w_i、$\boldsymbol{m}_i$和 $\boldsymbol{C}_i$分别是第 i 个高斯分量的权值、均值矢量和方差矩阵,这三个参数联合起来称作高斯混合模型参数 $\boldsymbol{\lambda}$,可以采用期望最大算法(Expectation Maximization,EM)来计算.

在实际应用中,可以根据宽带音频中提取的时频特征 $\boldsymbol{F}_x$实现对高频子带能量 $\boldsymbol{F}_y$的最小均方误差估计,估计函数如下式所示

$$\begin{aligned}\hat{y}&=\mathrm{E}[\boldsymbol{F}_y\mid\boldsymbol{F}_x]=\sum_{i=1}^{M}p(c_i\mid\boldsymbol{F}_x)\mathrm{E}[\boldsymbol{F}_y\mid c_i,\boldsymbol{F}_x]\\&=\sum_{i=1}^{M}p(c_i\mid\boldsymbol{F}_x)[\boldsymbol{m}_i^{y}+\boldsymbol{C}_i^{yx}(\boldsymbol{C}_i^{xx})^{-1}(\boldsymbol{F}_x-\boldsymbol{m}_i^{x})]\end{aligned} \tag{27}$$

其中,c_i表示第 i 个高斯分量;$p(c_i\mid\boldsymbol{F}_x)$表示 $\boldsymbol{F}_x$对应的第 i 个高斯分量的后验概率;$\mathrm{E}[\boldsymbol{F}_y\mid c_i,\boldsymbol{F}_x]$代表高频子带能量的条件期望;$\boldsymbol{m}_i^{y}$ 表示第 i 个高斯分量中高频子带能量 $\boldsymbol{F}_y$的均值矢量;$\boldsymbol{m}_i^{x}$ 表示第 i 个高斯分量中宽带时频特征 $\boldsymbol{F}_x$的均值矢量;$\boldsymbol{C}_i^{yx}$表示第 i 个高斯分量

的互相关矩阵；$\boldsymbol{C}_i^{xx}$表示第 i 个高斯分量的自相关矩阵.

最后，本文算法将 GMM 和局部 LS-SVM 方法相结合，能够实现对高频成分的有效重建. 另外，结合原始的低频频谱信息，通过 MLT 逆变换将扩展后的频谱由频域转换到时域，最终完成完整的超宽带音频信号的频带扩展方法.

3 实验比较和评测结果

为了验证所提频带扩展方法的有效性，本文在对语谱图进行分析的基础上，从主客观质量评测角度对所提方法与频谱搬移(Spectral Translation，ST)[7]和 NNM[9]两种频谱细节恢复参考算法进行了评测比较，其中参考算法同样采用 GMM 重建高频频谱包络.

本文选用 Moving Picture Experts Group(MPEG)编码质量测试音频数据库中的音频信号作为测试数据，包括小提琴、交响乐、流行音乐等 10 段音频片段，每段音频的长度为 10～20s. 原始音频数据为 32kHz 采样的 16 位 PCM 超宽带音频，并将其下采样到 16kHz，作为频带扩展方法的输入信号. 其中，GMM 所需宽带和超宽带训练数据源自全美音乐颁奖典礼转录的无损音频，其长度约为 2 小时，包含流行音乐、人声演唱和背景音效等. 在进行测试前，所有音频数据的信号能量均被调整为 -26dB.

3.1 高频频谱细节恢复参考算法

本文选择 ST 和 NNM 作为频谱细节重建的参考算法来进行性能测试. 其中，ST 为盲目式频带扩展中常用的频谱细节重建方法，它直接将低频频谱细节成分复制到高频频带中，进而实现对高频频带的有效扩展. 而 NNM 则同样采用相空间重构方法，根据低频相点的变化轨迹采用 NNM 方法对高频相点进行预测，从而逐点恢复高频频谱细节信息[9]，具体步骤如下：

步骤 1 低频相点集合 $\boldsymbol{S}=\{\boldsymbol{S}(k)\}$，$k=0,\cdots,279-(m-1)\tau$ 进行逐帧更新；

步骤 2 逐一计算新相点 $\boldsymbol{S}_N(i)$，$i=280-(m-1)\tau$ 与低频相点集合中各相点的内积$\langle \boldsymbol{S}_N(i),\boldsymbol{S}(k)\rangle$；

步骤 3 选择其中内积模最大的相点 $\boldsymbol{S}(k_{\max})$，作为 $\boldsymbol{S}_N(i)$的最近邻点，$k_{\max}$可表示为

$$k_{\max}=\underset{k=0,\cdots,279-(m-1)\tau}{\arg\max}\{|\langle \boldsymbol{S}_N(i),\boldsymbol{S}(k)\rangle|\} \tag{28}$$

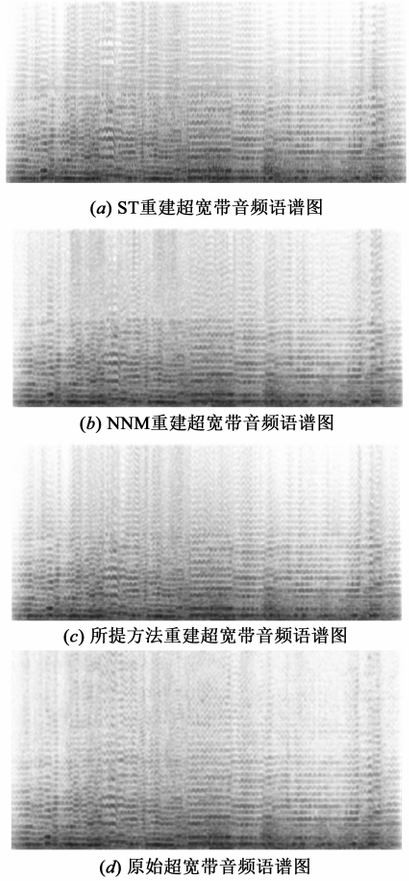
步骤 4 $\boldsymbol{S}(k_{\max})$的最高位元素作为新相点中最后一维 $C_{\text{norm}}(i+(m-1)\tau)$的预测值.

通过上述过程，音频频谱细节序列被不断更新，直到达到 14kHz 的截止频率，从而完成对高频频谱细节部分的逐点恢复.

3.2 语谱图分析

以交响乐信号为例，用不同频带扩展方法重建音频信号的语谱图和原始音频信号的语谱图如图 3 所示.

由语谱图表明所提方法有效地恢复了音频信号的高频频谱细节，高低频成分过渡自然. 而 ST 方法的重建音频在高低频结合处存在明显的频谱偏移，同时低频的强谐波成分复制到高频后，也会影响其主观听觉质量. 与原始超宽带音频对比，NNM 扩展的超宽带音频高频能量过于平滑，且高低频衔接处连续性较差，因而会不可避免地造成重建音频信号听觉质量的下降. 而本文所提方法克服了上述缺点，由 LS-SVM 扩展的频谱细节信息与原始音频更为接近，且进一步提高了重建信号的音频质量.



(*a*) ST重建超宽带音频语谱图

(*b*) NNM重建超宽带音频语谱图

(*c*) 所提方法重建超宽带音频语谱图

(*d*) 原始超宽带音频语谱图

图3 交响乐信号的重建超宽带音频语谱图比较

3.3 客观质量测试

在客观音频质量测试阶段，本文采用对数谱失真(Logarithmic Spectral Distortion，LSD)测度[16]以及音频质量感知评价法(Perceptual Evaluation of Audio Quality，PEAQ)[17]，对所提方法进行评测.

3.3.1 对数域谱失真测度

对数谱失真测度 d_{LSD}被广泛应用在客观质量评测中，其定义为：

$$d_{\mathrm{LSD}}=\frac{1}{M}\sum_{i=0}^{M-1}\sqrt{\frac{1}{N_{\mathrm{h}}-N_{\mathrm{l}}+1}\sum_{n=N_{\mathrm{l}}}^{N_{\mathrm{h}}}\left[10\log_{10}\frac{p_i(n)}{\hat{p}_i(n)}\right]^2} \tag{29}$$

式中,n 为功率谱的频率索引值,M 为音频信号总帧数,N_{l}为高频起始频率,对应于 7kHz 频点,N_{h}为高频截止频率,对应于 14kHz 频点,$p_i(n)$ 和 $\hat{p}_i(n)$ 分别表示第 i 帧原始音频功率值和重建音频功率值. 最后,将每帧音频信号的 d_{LSD}进行平均,作为最终的 LSD 测度.

应用 LSD 方法进行客观测试时,音频帧长为 20ms 并采用汉明窗进行处理,相邻帧间进行 50% 叠接. 图4 是两种方法之间的谱失真比较. 谱失真测试结果表明,所提频带扩展方法重建音频频谱失真测度的平均值为 7.287dB,NNM 参考方法重建音频谱失真为 9.503dB,而 ST 方法重建音频谱失真为 11.73dB. 所提频带扩展方法重建音频谱失真较参考方法有了明显降低,客观质量有明显改进. 此外,根据测试结果可以发现,小提琴、吉他等音频信号中高频成分相对暗淡,采用所提方法重建的超宽带音频与 NNM 算法效果基本相当. 而弦乐、摇滚、大提琴、贝斯和电子乐等乐曲高频能量较高,采用所提方法重建的高频成分更接近原始音频,重建效果明显优于 NNM 方法. 对于鼓乐信号,其原始音频高低频成分频谱细节差异较大,采用本文提出的盲目式扩展方法重建高频频谱的频谱细节仍存在一定的失真.

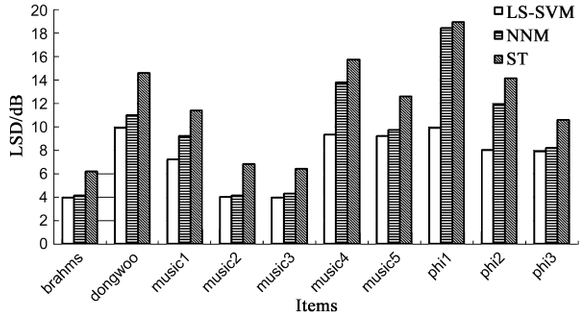


图4　LSD评测比较

3.3.2　音频质量感知评价

本文采用 PEAQ 进一步对所提方法与参考方法重建的超宽带音频信号进行客观测试. PEAQ 和主观测试统计结果具有良好的相似性,是一种重要的音频客观质量评价方法,其得分称作客观差异等级(Object Difference Grade,ODG). ODG 得分的范围是 -4(无法忍受)~0(无失真). 当 ODG 增加 0.1 时,表明合成音频产生了显著改善. 测试前,需要将所有音频数据上采样到 48kHz. ODG 得分情况如图 5 所示.

通过图 5 中的客观音频质量测试结果表明:所提方法重建音频 ODG 得分为 -2.951,较 NNM 方法有 0.162 的提高,较 ST 方法有 0.38 以上的提升. 在测试数据中,本文方法对交响乐、鼓乐、贝斯、弦乐等信号进行扩展后,其 ODG 得分较参考方法有较为明显地提高. 而对于鼓乐信号,其高频成分与低频成分的频谱细节差异明显,采用本文方法和参考方法重建鼓乐信号的 ODG 得分均低于 -3.4. 总体而言,本文所提方法的客观听觉质量优于参考音频频带扩展方法.

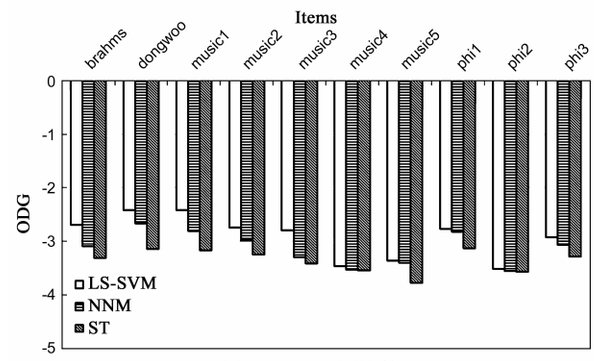


图5　ODG评测比较

3.4　主观质量测试

在主观音频质量测试阶段,本文采用 A/B 测试对所提方法和 NNM 方法所得到的音频信号进行质量评测,测试中邀请12 名测试者来进行主观测试. 为了保证公平性,测试数据以随机顺序进行排列,要求测试者从两组测试数据中选择较偏爱的一组,或者选择两者几乎无差异. 测试结果如表 1 所示:可知本文所提频带扩展方法得到的超宽带音频信号主观听觉质量同样要优于参考方法重建的音频质量.

表 1　两种方法的主观 A/B 测试结果比较

	NNM	LS-SVM	无偏爱
主观偏爱比例	28%	39%	33%

3.5　算法复杂度分析

本文借助 WMOPS(Weighted Million Operations Per Second)来统计所提方法的算法复杂度. 测试数据同样源自于 MPEG 音频数据库. 表 2 中分别给出了 MLT 变换、频谱包络估计、相空间重构以及基于最小二乘支持向量机的频谱细节恢复四个模块的复杂度数值. 总体上,整体算法的复杂度在 18WMOPS 左右.

表 2　所提算法各个模块复杂度统计结果

	算法复杂度(WMOPS)
MLT 变换	3.42
频谱包络估计	5.24
相空间重构	3.14
频谱细节恢复	6.27

4　结束语

本文提出了一种基于局部 LS-SVM 的宽带音频向

超宽带音频的盲目式频带扩展方法. 该方法根据音频信号的非线性特性,在相空间重构的基础上,采用局部 LS-SVM 对高频部分频谱细节信息进行预测;同时采用 GMM 来对高频频谱包络进行估计;最后将上述扩展后的两部分进行整合,并结合原始低频信息,有效实现了超宽带音频信号的重现. 语谱图分析和主客观评测结果表明:本文所提方法能够有效扩展宽带音频的带宽,其主客观性能优于传统的基于 NNM 的音频频带扩展方法.

参考文献

[1] ITU-T G. 722. 1 Annex C, Low Complexity Coding at 24 and 32 kb/s for Hands-free Operation in Systems with Low Frame Loss Annex C 14kHz Mode at 24, 32 and 48 kb/s [S]. 2005.

[2] Peter Vary, Rainer Martin. Digital Speech Transmission-Enhancement. Coding and Error Concealment [M]. UK: John Wiley & Sons Ltd, 2006.

[3] 张勇,胡瑞敏. 基于高斯混合模型的语音带宽扩展算法的研究[J]. 声学学报,2009,35(5):471 – 480.
Zhang Yong, Hu Ruimin. Speech wideband extension based on Gaussian mixture model [J]. Acta Acustica, 2009, 35(5):471 – 480. (in Chinese)

[4] Liu Xin, Bao Chang-chun. A harmonic bandwidth extension based on Gaussian mixture model [A]. 10th International Conference on Signal Processing [C]. Beijing: IEEE, 2010. 474 – 477.

[5] Liu Xin, Bao Chang-chun. Nonlinear bandwidth extension of audio signals based on hidden Markov model [A]. IEEE International Symposium on Signal Processing and Information Technology [C]. Bilbao, Spain: IEEE, 2011. 144 – 149.

[6] Liu Hao-jie, Bao Chang-chun. Audio bandwidth extension based on RBF neural network [A]. IEEE International Symposium on Signal Processing and Information Technology [C]. Bilbao, Spain: IEEE, 2011. 150 – 154.

[7] Erik Larsen, Ronald M Aarts. Audio Bandwidth Extension-application of Psychoacoustics. Signal Processing and Loudspeaker Design [M]. UK: John Wiley & Sons Ltd, 2004.

[8] Frederik Nagel, Sascha Disch. A harmonic bandwidth extension method for audio codecs [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Taiwan: IEEE, 2009. 145 – 148.

[9] Liu Xin, Bao Chang-chun. Nonlinear bandwidth extension based on nearest-neighbor matching [A]. Asia-Pacific Signal and Information Processing Association [C]. Singapore: APSIPA, 2010. 169 – 172.

[10] 刘鑫. 宽带音频的非线性频带展宽技术[D]. 北京:北京工业大学电控学院,2011.
Liu Xin. Nonlinear Bandwidth Extending for Wideband Audio [D]. Beijing: Beijing University of Technology, 2011. (in Chinese)

[11] 王海燕,卢山. 非线性时间序列分析及其应用[M]. 北京:科技出版社,2006. 10 – 11,12 – 16,102 – 103.

[12] 刘秉正,彭建华. 非线性动力学[M]. 北京:高等教育出版社,2004. 396 – 398,400 – 414,441 – 449.

[13] 韩敏. 混沌时间序列预测理论与方法[M]. 北京:中国水利水电出版社,2007. 155 – 172.

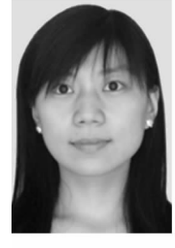
[14] Holger Kantz, Thomas Schreiber. Nonlinear Time Series Analysis [M]. Britain: Cambridge University Press, 2004. 42 – 51.

[15] 张燕平,张铃. 机器学习理论与算法[M]. 北京:科学出版社,2012.

[16] Pulakka H, Laaksonen L. Evaluation of an artificial speech bandwidth extension method in three languages [J]. IEEE Transactions on Audio, Speech and Language Processing, 2008, 16(6):1124 – 1137.

[17] ITU-R BS. 1387-1, Method for Objective Measurements of Perceived Audio Quality [S]. 2001.

作者简介



白海钏 女,1986 年出生,河北邯郸人,北京工业大学硕士研究生. 主要研究方向为音频信号处理.
E-mail:baihaichuan@ emails. bjut. edu. cn

鲍长春 男,1965 年出生,内蒙古赤峰人,博士,北京工业大学教授、博士生导师,IEEE 高级会员,国际语音通信学会(ISCA)会员,亚太信号与信息处理学会(APSIPA)会员,中国电子学会理事,中国声学学会理事,信号处理专业委员会委员. 主要研究方向为语音与音频信号处理.
E-mail:chchbao@ bjut. edu. cn

刘 鑫(通信作者) 男,1986 年出生,北京人,北京工业大学博士研究生. 主要研究方向为语音与音频信号处理.
E-mail:liuxin0930@ emails. bjut. edu. cn