

基于灰色 Verhulst 模型的音频频带扩展方法

白海钊, 鲍长春, 刘 鑫, 李红蕊

(北京工业大学电子信息与控制工程学院, 北京 100124)

摘 要: 受到传输带宽的限制, 宽带音频的自然度和听觉质量会有所下降, 因此, 本文提出了一种基于灰色 Verhulst 模型的宽带向超宽带音频频带扩展方法. 根据音频信号频谱包络序列的演变趋势, 采用灰色 Verhulst 模型对高频频谱包络进行估计, 并利用最近邻匹配方法对高频频谱细节进行预测, 最后经过高频频谱包络的调整, 所提方法能够有效地恢复 7~14kHz 频率范围内的高频成分. 主客观测试表明, 该方法改善了宽带音频的听觉质量, 并且优于传统的基于高斯混合模型的音频频带扩展方法.

关键词: 音频编码; 频带扩展; 灰色 Verhulst 模型; 最近邻匹配

中图分类号: TN912. 3 **文献标识码:** A **文章编号:** 0372-2112 (2014)08-1624-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.08.026

Audio Bandwidth Extension Method Based on Grey Verhulst Model

BAI Hai-chuan, BAO Chang-chun, LIU Xin, LI Hong-rui

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: The naturalness and auditory quality of the wideband audio is generally degraded due to the limitation of transmission bandwidth, so this paper presents a kind of audio bandwidth extension method from the wideband to super wideband based on grey Verhulst model. Grey Verhulst model was utilized for estimating the envelope of high-frequency spectrum, according to the evolution tendency of the spectral envelope series of audio signals. In addition, nearest-neighbor matching was utilized to predict the fine structure of high-frequency components. At last, the proposed method can effectively recover the high-frequency components in the frequency range 7-14 kHz through the envelope adjustment of high-frequency spectrum. Subjective and objective testing results indicate that the proposed method can improve the auditory quality of the wideband audio and outperforms the conventional method of audio bandwidth extension based on Gaussian mixture model.

Key words: audio coding; bandwidth extension; grey Verhulst model; nearest-neighbor matching

1 引言

在现阶段的音频通信系统中一般传输的是宽带信号, 其有效带宽为 50Hz~7kHz, 采样率提升到了 16kHz. 宽带音频具有良好的可懂度和优质的音频质量, 然而由于缺失 7kHz 以上的高频部分, 其重建音乐信号的透明度和表现力有所欠缺. 因此, 有学者^[1]提出了 32kHz 采样的超宽带音频的概念, 其频带范围在 50Hz~14kHz. 超宽带音频信号的重建声音具有较高的层次感和更好的自然度, 但是它的传输会导致处理数据量的增加. 所以可以采用频带扩展的方法, 在不影响网络传输和宽带信号信源编码的前提下, 在接收端对重建的宽带音频人为地增加一定的频率信息, 恢复其截去的高频带成分, 从而达到增强听觉质量、重现超宽带音频的目的^[2].

传统方法通常适当增加丢失频带的边信息, 对宽带

音频信号进行非盲目式频带扩展. 这种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, 无法满足移动通信网络传输的需求. 因此, 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.

盲目式频带扩展方法可以分为频谱包络扩展和频谱细节扩展两部分. 重建音频的主观质量主要由频谱包络估计的准确性来决定. 近年来, 许多研究者提出了不同的频谱包络估计算法. 其中, 线性外推法^[3]根据音频信号对数谱所呈现的线性递减特性利用外推方法来扩展高频频谱. 码书映射^[4]分别对高低频频谱包络参数建立并行码书, 利用高低频码书映射关系估计高频频谱包络成分. 而高斯混合模型^[5~8] (Gaussian Mixture Model, GMM) 和隐马尔科夫模型^[9,10] 则利用统计模型的方法, 通过计算宽带音频特征和高频频谱包络能量的联合概

率密度,在最小均方误差准则下对高频子带能量进行贝叶斯估计.该类方法具有良好的平滑近似特性,能够取得较好的重建音频质量.但是与原始音频信号相比较,其重建的高频频谱包络能量较低,同时高低频衔接部分的频谱包络连续性较差,因而重建的超宽带音频信号的明亮程度不足,音质不太自然.

从信息论的角度和语谱图分析可以发现,音频信号的高低频频谱包络之间具有一定的相关性和近似的变化趋势^[3,11].据此,本文引入了灰色 Verhulst 模型(Grey Verhulst Model, GVM),通过建立灰色微分方程来描述音频频谱包络的变化趋势,并基于最小二乘法由音频信号的低频频谱包络序列预测高频频谱包络.同时,我们采用前期工作中提出的最近邻匹配(Nearest-Neighbor Matching, NNM)^[12]对高频频谱细节进行重建,最终实现了宽带向超宽带音频信号的频带扩展.

2 基于 GVM 的音频频带扩展方法

本文所提的频带扩展方法采用有效带宽为 7kHz 采样率为 16kHz 的宽带音频信号作为输入信号,其原理如图 1 所示.

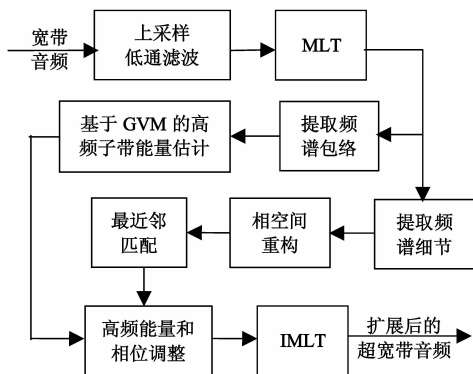


图1 所提音频频带扩展方法原理框图

输入音频信号通过基 2 上采样和低通滤波器后,得到的信号采样率为 32kHz,有效带宽为 7kHz.接下来,将该信号以 20ms 帧长进行划分,进行调制重叠变换(Modulated Lapped Transform, MLT),叠接长度为 20ms,得到 640 点的 MLT 系数.由于采用了低通滤波,7kHz 以上 MLT 参数的能量趋近于 0,所以低频部分相应的 MLT 系数可以表示为 $C_{mlt}(i), i=0 \cdots 279$.

盲目式频带扩展方法由频谱包络扩展和频谱细节扩展两部分构成.因此,输入信号经过 MLT 变换后,本文将对这两个部分分别进行处理.

2.1 基于 GVM 的频谱包络扩展

经过时频变换和低通滤波后,将 7kHz 以下的 280 个 MLT 参数按照每 20 个频点为一个子带进行划分,可以得到 14 个子带,并分别计算每个子带的均方根能量

$e_{rms}^{(0)}(n)$,即

$$e_{rms}^{(0)}(n) = \sqrt{\frac{1}{20} \sum_{i=0}^{19} C_{mlt}^2(20n + i)}, 0 \leq n < 14 \quad (1)$$

通过上式所得到的子带均方根能量,可以表示音频频谱的包络能量信息,因而非负序列 $E^{(0)} = \{e_{rms}^{(0)}(n) | n=0, \dots, 13\}$ 可以表示每一帧低频频谱包络信息.同理,7~14kHz 高频成分的频谱包络也可以采用子带均方根能量来表示,即 $E^{(0)} = \{e_{rms}^{(0)}(n) | n=14, \dots, 27\}$.

通过实验分析可以发现:大部分超宽带音频信号的低频频谱包络能量较大,随着频率的增加,子带包络能量逐步降低,最终趋于稳定.该变化过程与具有稳定状态的“S”型序列相近^[13],因此,本文采用 GVM 根据已知的低频频谱包络序列来估计高频子带能量.

所提方法的原理如图 2 所示.首先对原始频谱包络序列进行一次累加生成(Accumulated Generating Operation, AGO)和紧邻均值生成,目的是减少原始序列的随机性,然后通过数据序列的最小二乘拟合,确定频谱包络序列的模型参数^[14],最后建立 GVM 预测模型.

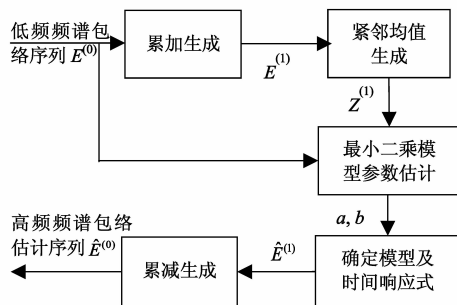


图2 基于GVM的高频子带能量估计原理框图

具体步骤如下:

首先,设原始宽带音频信号的频谱包络序列为:

$$E^{(0)} = \{e_{rms}^{(0)}(0), e_{rms}^{(0)}(1), \dots, e_{rms}^{(0)}(13)\} \quad (2)$$

其中, $e_{rms}^{(0)}(n) \geq 0, n=0, 1, \dots, 13, n$ 为低频频谱序列的子带索引号.

然后,根据原始频谱包络序列 $E^{(0)}$,可以得到其一次累加(AGO)序列 $E^{(1)}$,

$$E^{(1)} = \{e_{rms}^{(1)}(0), e_{rms}^{(1)}(1), \dots, e_{rms}^{(1)}(13)\} \quad (3)$$

其中,

$$e_{rms}^{(1)}(n) = \sum_{i=0}^n e_{rms}^{(0)}(i), n=0, \dots, 13 \quad (4)$$

进一步得到紧邻均值生成序列 $Z^{(1)}$,

$$Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(13)\} \quad (5)$$

其中,

$$z^{(1)}(n) = \frac{1}{2} e_{rms}^{(1)}(n) + \frac{1}{2} e_{rms}^{(1)}(n-1), n=1, \dots, 13 \quad (6)$$

接下来,根据 $E^{(0)}$ 和 $Z^{(1)}$,建立灰色微分方程来描述频

谱包络序列的变化趋势^[15],即

$$e_{rms}^{(0)}(n) + az^{(1)}(n) = b(z^{(1)}(n))^2 \quad (7)$$

其中,参数 a 和 b 分别称为发展系数和驱动系数, a 的大小和符号反映了高频频谱包络累加序列估计值 $\hat{E}^{(1)}$ 和原始序列估计值 $\hat{E}^{(0)}$ 的发展态势. 而 b 反映了高低频频谱包络序列变化的关系.

接下来借助最小二乘法原理求出式(7)中的参数 a 和 b .

令 $\mathbf{A} = (a, b)^T$, 且

$$\mathbf{B} = \begin{bmatrix} -z^{(1)}(1) & (z^{(1)}(1))^2 \\ -z^{(1)}(2) & (z^{(1)}(2))^2 \\ \vdots & \vdots \\ -z^{(1)}(13) & (z^{(1)}(13))^2 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} e_{rms}^{(0)}(1) \\ e_{rms}^{(0)}(2) \\ \vdots \\ e_{rms}^{(0)}(13) \end{bmatrix} \quad (8)$$

则灰色微分方程(7)的最小二乘估计参数列满足:

$$\mathbf{A} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} \quad (9)$$

在灰色系统理论中, GVM 的白化方程定义为^[15]

$$\frac{de_{rms}^{(1)}}{dt} + ae_{rms}^{(1)} = b(e_{rms}^{(1)})^2 \quad (10)$$

其时间响应函数可以表示为:

$$e_{rms}^{(1)}(t) = \frac{ae_{rms}^{(1)}(0)}{be_{rms}^{(1)}(0) + (a - be_{rms}^{(1)}(0))e^{at}} \quad (11)$$

将上式转换为离散形式^[15], 则式(7)的时间响应序列可以表示为:

$$\hat{e}_{rms}^{(1)}(n) = \frac{ae_{rms}^{(1)}(0)}{be_{rms}^{(1)}(0) + (a - be_{rms}^{(1)}(0))e^{a(n-1)}} \quad (12)$$

其中, $n = 14, \dots, 27$, n 为高频频谱序列的子带索引号.

最后, 根据 $\hat{E}^{(1)}$ 利用一次累减可以还原出超宽带音频信号高频频谱包络的原始序列估计值 $\hat{E}^{(0)}$, 即有:

$$\hat{e}_{rms}^{(0)}(n) = \hat{e}_{rms}^{(1)}(n) - \hat{e}_{rms}^{(1)}(n-1) \quad (13)$$

通过上述过程可以对每一帧高频部分的各个子带能量 $\mathbf{E}^{(0)} = \{e_{rms}^{(0)}(n) | n = 14, \dots, 27\}$ 逐个进行预测. 最终, 结合原始的低频频谱包络序列, 该方法能够较好地重建超宽带音频信号的频谱包络信息.

2.2 基于最近邻匹配的频谱细节预测

根据每个子带的均方根能量, 将 7kHz 以下的 MLT 系数 $C_{mlt}(i)$, $i = 0, \dots, 279$ 进行归一化, 即本文采用归一化的 MLT 系数 $C_{norm}(i)$, $i = 0, \dots, 279$ 来表示 7kHz 以下的低频频谱细节信息, 如式(14)所示:

$$C_{norm}(i) = \frac{C_{mlt}(i)}{e_{rms}^{(0)}(n)}, 0 \leq i < 280, n = \left\lfloor \frac{i}{20} \right\rfloor \quad (14)$$

基于音频信号的非线性特性, 可以根据 $C_{norm}(i)$ 来构成低频相点集合, 从而进行相空间重构^[16]. 本文频谱细节预测部分引入了最近邻匹配法, 利用低频相点的

变化轨迹对高频相点进行预测, 从而逐点恢复高频频谱细节信息^[12].

首先进行相空间重构, 则 m 维的相点 $S(i)$, $i = 0, \dots, 279 - (m-1)\tau$, 可以表示为:

$$S(i) = \{C_{norm}(i), C_{norm}(i + \tau), \dots, C_{norm}(i + (m-1)\tau)\} \quad (15)$$

其中, τ 和 m 分别表示当前帧的音频频谱序列的延迟时间和嵌入维数, 分别利用自相关法和虚假近邻点法对两个参数进行计算^[16].

经过相空间重构后, 采用最近邻匹配方法来预测高频相点的运动轨迹.

(1) 逐一计算新相点 $S_N(i)$, $i = 280 - (m-1)\tau$ 与逐帧更新的低频相点集合 $S = \{S(k)\}$, $k = 0, \dots, 279 - (m-1)\tau$ 中各相点的内积 $\langle S_N(i), S(k) \rangle$;

(2) 选择其中内积模最大的相点 $S(k_{max})$, 作为 $S_N(i)$ 的最近邻点, k_{max} 可以表示为

$$k_{max} = \arg \max_{k=0, \dots, 279 - (m-1)\tau} \{|\langle S_N(i), S(k) \rangle|\} \quad (16)$$

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

本文根据 GVM 恢复的高频频谱包络来对高频频谱细节的能量进行调整, 并最终重建所截去的高频部分. 高频部分相应的 MLT 系数可以表示为 $C_{mlt}(i)$, $i = 280, \dots, 559$, 其估计值如式(17)所示:

$$\hat{C}_{mlt}(i) = \hat{C}_{norm}(i) \times \hat{e}_{rms}^{(0)}(n), \quad n = \left\lfloor \frac{i}{20} \right\rfloor, \quad i = 280, \dots, 559 \quad (17)$$

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基于 GMM 的频谱包络估计方法^[12]进行了评测比较, 其中参考算法同样采用最近邻匹配的方式重建高频频谱细节.

本文选用 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. 在进行测试前, 所有音频数据的信号能量均被调整为

-26dB.

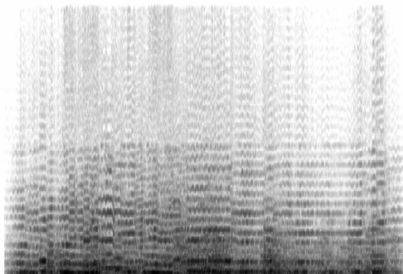
3.1 高频频谱包络估计参考算法

本文选择传统基于 GMM 的频谱包络估计方法作为参考算法来进行性能测试. 该方法可以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 \mathbf{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 \mathbf{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^[5].

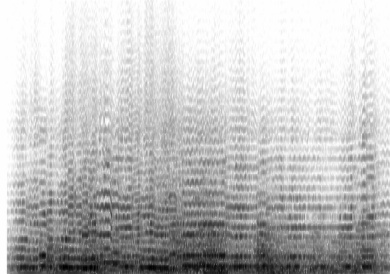
高低频特征的联合概率密度 $p(\mathbf{x}, \mathbf{y} | \boldsymbol{\lambda})$ 可以表示为:

$$p(\mathbf{x}, \mathbf{y} | \boldsymbol{\lambda}) = \sum_{i=1}^M w_i p(\mathbf{x}, \mathbf{y} | \mathbf{m}_i, \mathbf{C}_i) \quad (18)$$

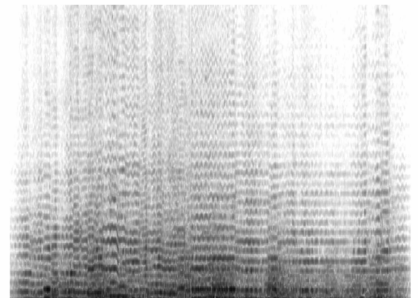
其中, $M=64$ 为高斯分量的个数, $p_g(\mathbf{x}, \mathbf{y} | \mathbf{m}_i, \mathbf{C}_i)$ 为第 i 个高斯分量的联合概率密度, w_i 、 \mathbf{m}_i 和 \mathbf{C}_i 分别是第 i 个高斯分量的权值、均值矢量和方差矩阵, 这三个参数联合起来称作高斯混合模型参数 $\boldsymbol{\lambda}$, 可以采用期望最大算法(expectation maximization, EM)来计算^[5].



(a) GMM 扩展的音频信号语谱图



(b) 所提方法扩展的音频信号语谱图



(c) 原始超宽带音频信号语谱图

图3 交响乐信号的重建超宽带音频语谱图比较

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包络, 高低频成分过渡自然. 与原始超宽带音频对比, GMM 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以上缺点, 由 GVM 扩展的频谱包络与原始音频更为接近, 并且提高了重建信号的音频质量.

3.3 客观质量测试

在客观音频质量测试阶段, 本文采用对数域谱失真(Logarithmic Spectral Distortion, LSD)测度^[17]以及音频质量感知评价法(Perceptual Evaluation of Audio Quality, PEAQ)^[18], 与基于 GMM 进行频带扩展的参考方法进行评测.

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:

在实际应用中, 可以根据宽带音频中提取的时频特征 \mathbf{x} 实现对高频子带能量 \mathbf{y} 的最小均方误差估计, 估计函数如式(19)所示

$$\begin{aligned} \hat{\mathbf{y}} &= E[\mathbf{y} | \mathbf{x}] = \sum_{i=1}^M p(c_i | \mathbf{x}) E[\mathbf{y} | c_i, \mathbf{x}] \\ &= \sum_{i=1}^M p(c_i | \mathbf{x}) [\mathbf{m}_i^y + \mathbf{C}_i^{yx} (\mathbf{C}_i^{xx})^{-1} (\mathbf{x} - \mathbf{m}_i^x)] \end{aligned} \quad (19)$$

其中, c_i 表示第 i 个高斯分量, $p(c_i | \mathbf{x})$ 表示 \mathbf{x} 对应的第 i 个高斯分量的后验概率.

最终, 参考算法将 GMM 法和最近邻匹配法相结合, 能够实现对高频成分的有效重建. 在性能测试过程中, GMM 方法所需的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.

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.

$$d_{LSD} = \frac{1}{M} \sum_{i=0}^{M-1} \sqrt{\frac{1}{N_h - N_l + 1} \sum_{n=N_l}^{N_h} [10 \log_{10} \frac{p_i(n)}{\hat{p}_i(n)}]^2} \quad (20)$$

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.

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 7.991dB, GMM 方法重建音频谱失真为 9.503dB, 所提频带扩展方法重建音频谱失真较 GMM 方法提高了 1.5dB 左右, 客观质量有明显的改进. 测试数据中, 除了鼓乐、吉他、小提琴与参考方法效果相当, 其他数据包括贝斯、弦乐、摇滚、电子乐和大提琴等乐曲, 重建效果明显

优于 GMM 参考方法.

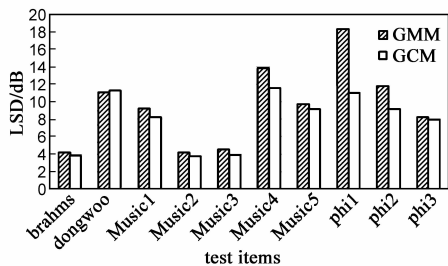


图4 LSD评测比较

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 GMM 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.

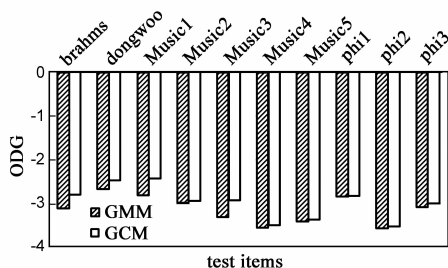


图5 ODG评测比较

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 -2.960 ,较参考方法有 0.146 的提高. 另外,在 95% 置信水平下,所提方法重建音频 ODG 得分的置信区间为 $[-3.228, -2.708]$,参考算法的置信区间为 $[-3.327, -2.901]$. 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 得分较参考方法有所提高. 该结果显示本文所提方法的客观听觉质量优于传统的基于 GMM 进行频带扩展的方法.

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 测试对两种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基于 GMM 重建的音频质量.

表 1 两种方法的主观 A/B 测试结果比较

	GMM	GVM	无偏爱
主观偏爱比例	26%	39%	35%

3.5 计算量分析

为了进一步测试所提出算法的计算量,本文对程序实现中每秒需要的乘法次数进行了统计,并与基于 GMM 的参考算法进行了比较. 由于两个算法采用相同的时频变换和频谱细节扩展方法,因此,所统计的计算量只包含频谱包络估计模块. 对于 GMM 方法,每秒需要进行 4.36 百万次乘法,而本文提出的 GVM 方法每秒则需要 2.93 百万次乘法. 此外,本文所提方法没有利用任何线下训练先验知识,因而无需额外占用存储空间. 而基于 GMM 的方法则需要 195kByte 的空间来存储各个高斯分量的权值、均值和方差矩阵. 由此可见,所提算法在计算量和存储空间上都要低于基于 GMM 的参考方法.

4 结束语

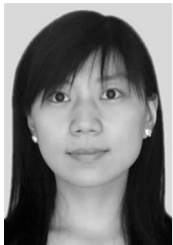
本文提出了一种基于 GVM 的宽带向超宽带音频信号频带扩展方法. 该方法根据音频信号的低频频谱包络序列,采用 GVM 对高频部分频谱包络信息进行预测;同时采用最近邻匹配方法来重建所截去的高频频谱细节;最后经过高频能量调整,保证了重建超宽带音频信号的较高质量. 语谱图分析和主客观评测结果表明:本文方法能够有效地扩展宽带音频的带宽,其主客观性能优于传统的基于 GMM 的音频频带扩展方法.

参考文献

- [1] ITU-T G. 722.1 Annex C, Low complexity coding at 24 and 32kb/s for hands-free operation in systems with low frame loss Annex C 14kHz mode at 24,32 and 48kb/s[S].
- [2] Vary P, Martin R. Digital Speech Transmission-Enhancement, Coding and Error Concealment[M]. UK: John Wiley & Sons Ltd, 2006.
- [3] Chig-Min Liu, Wen-Chieh Lee, Han-Wen Hsu. High frequency reconstruction for band-limited audio signals[A]. Proceeding of 6th Conference on Digital Audio Effects[C]. London, UK: IEEE Press, 2003. 1-6.
- [4] Soon I Y, Yeo C K. Bandwidth extension of narrowband speech using soft-decision vector quantization[A]. Proceeding of 5th International Conference on Information Communications and Signal Processing[C]. Bangkok, Thailand: IEEE Press, 2005. 734-738.
- [5] Liu Xin, Bao Changchun, Jia Maoshen, Sha Yongtao. A harmonic bandwidth extension based on Gaussian mixture model [A]. Proceeding of 10th International Conference on Signal

- Processing (ICSP2010) [C]. Beijing: IEEE Press, 2010. 474 – 477.
- [6] Pulakka H, Remes U, Yrttiaho S, et al. Bandwidth extension of telephone speech to low frequencies using sinusoidal synthesis and a Gaussian mixture model [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20 (8): 2219 – 2231.
- [7] Nilsson M, Gustafsson H, Andersen S V, Kleijn W B. Gaussian mixture model based mutual information estimation between frequency bands in speech [A]. Processing of the IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Orlando, USA: IEEE Press, 2002. I – 525 – 528.
- [8] Nour-Eldin A H, Kabal P. Memory-based approximation of the Gaussian mixture model framework for bandwidth extension of narrowband speech [A]. Processing of Interspeech 2011 [C]. Florence, Italy: ISCA Press, 2011. 1185 – 1188.
- [9] Yagli C, Turan T T, Erzin E. Artificial bandwidth extension of spectral envelope along a Viterbi path [J]. Speech Communication, 2013, 55(1): 111 – 118.
- [10] Song G B, Martynovich P. A study of HMM-based bandwidth extension of speech signals [J]. Signal Processing, 2009, 89 (10): 2036 – 2044.
- [11] Iser B, Schmidt G, Minker W. Bandwidth Extension of Speech Signals [M]. USA: Springer-Verlag, 2008.
- [12] Liu Xin, Bao Changchun, Jia Maoshen, Sha Yongtao. Nonlinear bandwidth extension based on nearest-neighbor matching [A]. Processing of the Second APSIPA (Asia-Pacific Signal and Information Processing Association) Annual Summit and Conference (ASC) [C]. Biopolis, Singapore: APSIPA Press, 2010. 169 – 172.
- [13] Wang Zhengxin, Dang Yaoguo, Liu Sifeng. Unbiased grey Verhulst model and its application [J]. Systems Engineering-Theory & Practice, 2009, 29(10): 138 – 144.
- [14] Kayacan E, Ulutas B, Kaynak O. Gray system theory-based models in time series prediction [J]. Expert System with Application, 2010, 37: 1784 – 1789.
- [15] Liu Sifeng, Lin Yi. Grey Information-Theory and Practical Applications [M]. USA: Springer-Verlag, 2006.
- [16] Kantz H, Schreiber T. Nonlinear Time Series Analysis [M]. Britain: Cambridge University press, 2004. 42 – 51.
- [17] Pulakka H, Laaksonen L, Vainio M, Pohjalainen J, Alku P. 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.
- [18] ITU-R Rec BS. 1387-1, Method for objective measurements of perceived audio quality, 2001 [S].

作者简介



白海钊 女, 1986 年生于河北邯郸. 北京工业大学硕士研究生, 研究方向为音频信号处理.
E-mail: baihaichuan@emails.bjut.edu.cn



鲍长春 (通信作者) 男, 1965 年生于内蒙古赤峰. 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