doi:10.3969/j.issn.1000-436x.2013.08.016

基于非线性音频特征分类的频带扩展方法

张丽燕，鲍长春，刘鑫，张兴涛
（北京工业大学 电子信息与控制工程学院 语音与音频信号处理研究室，北京 100124）

摘 要：提出了一种基于非线性音频分类的频带扩展方法，即利用递归图和定量递归分析将音频信号的时间序列分成 4 类，并分别采用 4 种方法恢复高频频谱细节，最终利用高斯混合模型和基于软判决的码书映射调整频谱包络和能量增益。主客观测试表明，该方法优于传统的盲目式频带扩展方法，且应用到 ITU-T G.722.1 编解码器时，音频质量优于同码率下的 G.722.1C 编解码器。

关键词：音频编码；频带扩展；音频分类；递归图；定量递归分析
中图分类号：TN912.3 **文献标识码**：A **文章编号**：1000-436X(2013)08-0120-11

Bandwidth extension method based on nonlinear audio characteristics classification

ZHANG Li-yan, BAO Chang-chun, LIU Xin, ZHANG Xing-tao
(Speech and Audio Signal Processing Lab, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: A bandwidth extension method based on audio classification was proposed. Time series of audio signals were classified into four types based on recurrence plot and recurrence quantification analysis, and the fine spectrums were recovered by taking advantage of four methods respectively. In addition, the spectrum envelope and energy gain were adjusted by Gaussian mixture model and codebook mapping on the basis of soft decision respectively. Subjective and objective testing results indicate that the proposed method has good quality compared with conventional blind bandwidth extension methods, and the performance of ITU-T G.722.1 codec with the proposed algorithm is better than that of G.722.1C codec at the same bit rate.

Key words: audio coding; bandwidth extension; audio classification; recurrence plot; recurrence quantification analysis

1 引言

基于听觉感知理论的音频编码方法通常将有限的编码比特优先分配给低频带信息，这样能够有效地避免明显的听觉失真[1]。但是由于音频信号的频率带宽受到了限制，音频的明亮度和自然度都受到了很大影响。因此，在有限网络带宽和存储能力的条件下，解决编解码后音频质量不高的问题具有十分重要的现实意义。作为一种有效提高音质的音频增强方法，音频信号的频带扩展 (BWE, bandwidth extension)技术成为了现代音频编码领域的最新研究热点。它的原理是通过分析编码之前的原始音频信号的特点，在解码端对解码后的音频信息额外地增加一部分频率信息，恢复其丢失的高频信息，从而达到扩展信号的频率带宽、增强音频听觉质量的目的[2]。

频带扩展方法可分为“非盲目式”和“盲目式”

收稿日期：2012-05-22；**修回日期**：2013-01-29
基金项目：国家自然科学基金资助项目（60872027，61072089）；北京市教育委员会科技发展计划重点基金资助项目（KZ201110005005）；北京市自然科学基金资助项目（4082006）；北京市属高等学校人才强教计划基金资助项目；北京工业大学第九届研究生科技基金资助项目（ykj-2011-4910）
Foundation Items: The National Natural Science Foundation of China(60872027, 61072089); Beijing Natural Science Foundation Program and Scientific Research Key Program of Beijing Municipal Commission of Education(KZ201110005005); The Natural Science Foundation of Beijing (4082006); The Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality; The 9th Postgraduate Science Foundation of Beijing University of Technology (ykj-2011-4910)

2 种[3]。与“非盲目式”扩展方法需额外增加编码比特相比，“盲目式”音频频带扩展可以不用任何附加信息，仅依据音频频谱的统计特性和高低频的相关信息人为地扩展带宽，从而有效节省了编码高频信息所需要的比特数。因此，本文将着重研究“盲目式”扩展算法。

目前，针对音频信号宽带向超宽带“盲目式”频带扩展的研究工作尚处于起步阶段。音频信号频带扩展在传统语音信号频带扩展的基础上取得了很大进展。然而，国内外有关音频信号频带扩展算法的研究均是对该段信号采用统一的处理模型[4-6]，由于音频信号的复杂性，这些算法对于音频中的谐波信息和暂态信息预测效果并不明显。在对音频信号进行分帧处理时发现，每帧信号具有不同的时频特性。因此，可以考虑在对音频信号进行频带扩展之前，首先对音频信号的时间序列进行分类，然后对不同类别的音频信号采用不同的频带扩展方法。

传统的音频信号分类方法主要是基于内容将音频信号分成静音、纯净语音、音乐和带背景噪声的语音[7]。音频信号处理在这种分类方法的基础上进行，对于不同内容的音频信号分别采用不同的信号处理方式。然而，对于内容相同的信号（如所有的音乐信号），信号处理时仍然采用单一模型，这对于信号中的谐波信息和暂态信息的处理效果并无改进。因此，有必要对音频信号基于特征进行分类，然后针对不同的特征段采用不同的信号处理方法。

基于此，本文提出了一种基于非线性音频特征分类的频带扩展方法，即在对音频信号时间序列进行特征分类的基础上，针对每类信号分别采用不同的频谱细节扩展方式，并利用高斯混合模型估计高频子带能量，然后利用基于软判决的码书映射调整能量增益，最后构建出一套完整的宽带向超宽带音频盲目式频带扩展的算法框架。

2　基于递归图和定量递归分析的音频分类

音频信号的时间序列是极其复杂的，其结构变化较快，且无任何规律可言，这可以从图 1 中音频信号时间序列的波形图中看出。音频信号中即存在突然增大或减小的暂态结构（如图 1 中的 a 段信号），也存在类似周期信号的谐波结构（如图 1 中的 b 段信号），还存在与随机信号相似的噪声结构（如图 1 中的 c 段信号）。并且这些结构在频域中也有不同的表现形式，其对不同的频带扩展算法有不同的反应效果。因此，本文采用基于分类的方法对不同类型的音频信号采用不同的处理方式，以提高频带扩展的效果。

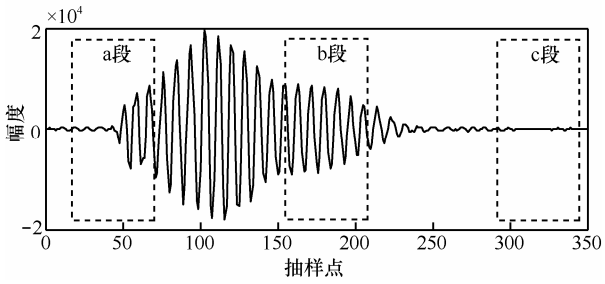


图 1　音频信号时间序列的波形图

相关研究发现，音频信号具有典型的非线性时域特征[8]。针对音频信号的短时非平稳的特性，本文引入非线性特征分析领域的递归图和定量递归分析技术对音频信号进行分类。

2.1　相空间重构

在实际问题中，多数情况下并不能直接观察到系统的状态点，而只能得到有关系统的一维或有限维的抽样时间序列，实际上，这些数据是原有状态函数在低于其相空间维数空间上投影的反映，往往只描述了系统不完全的信息。因此，只有一维的时间序列经过重构张开到三维或其以上的相空间，才能把时间序列中的多维动力学信息充分提取出来，相空间重构技术便应运而生。

利用时间延迟嵌入相空间重构方法[9]将一维的音频信号时间序列$\{x_n, n=1,2,\cdots,M\}$重构成 m 维的相空间 $\boldsymbol{Y}$ 为

$$\begin{aligned}\boldsymbol{Y}&=(\boldsymbol{y}_1,\boldsymbol{y}_2,\boldsymbol{y}_3,\cdots,\boldsymbol{y}_N)\\&=\begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{M-(m-1)\tau} \\ x_{1+\tau} & x_{2+\tau} & x_{3+\tau} & \cdots & x_{M-(m-2)\tau} \\ x_{1+2\tau} & x_{2+2\tau} & x_{3+2\tau} & \cdots & x_{M-(m-3)\tau} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1+(m-1)\tau} & x_{2+(m-1)\tau} & x_{3+(m-1)\tau} & \cdots & x_M \end{pmatrix}\end{aligned} \tag{1}$$

其中，M 为时间序列$\{x_n\}$的长度；m 为相空间的嵌入维数；τ 为嵌入延迟时间；$N=M-(m-1)\tau$ 为相空间中相点的个数；$\boldsymbol{y}_i$ 为 m 维相空间中的一个相点，表示系统在时刻 i 的状态，这些有时间标记的向量序列$\{\boldsymbol{y}_i, i=1,2,3,\cdots,N\}$构成了系统的 m 维相空间轨迹。

2.2　递归图和定量递归分析

递归图(RP, recurrence plot)[10]是一种新的分析非线性时间序列的方法。一般来讲，三维及三维以下的相空间是可以通过图形描述的，三维以上的相空间，

必须通过投影到低维子空间（二维或三维）才能利用图形描述。而递归图便可以让在一个二维空间的递归表示图上对高维相空间($m \geqslant 3$)上的轨道进行观测研究。递归图是利用状态向量 $\boldsymbol{y}_i$ 以图形的方式来显示动力系统内部结构的变化规律，其数学表示式为

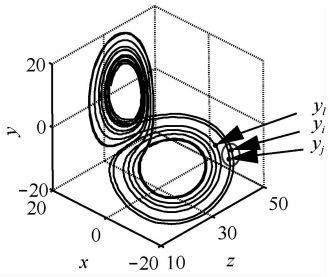
$$R_{i,j} = \Theta(\varepsilon_i - \|\boldsymbol{y}_i - \boldsymbol{y}_j\|), i, j = 1, 2, \cdots, N \tag{2}$$

其中，ε_i 是预先设定的临界距离，$\|\bullet\|$ 代表范数（如 1-范数，2-范数，∞-范数等），$\Theta(\bullet)$是 Heaviside 函数，其定义为

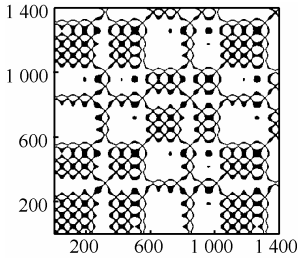
$$\Theta(z) = \begin{cases} 0, z < 0 \\ 1, z \geqslant 0 \end{cases} \tag{3}$$

由 Heaviside 函数的特性可知，$R_{i,j}$ 的值只可能为 0 或 1。当 $R_{i,j}$ 的值为 1 时，在递归图中坐标为(i, j)的位置上表示为一个黑点；相反，当 $R_{i,j}$ 的值为 0 时，递归图中坐标为(i, j)的位置上表示一个白点。

以 Lorentz 系统为例，图 2 给出了其三维相空间轨迹运行图及其对应的递归图。图 2 (a)中的圆圈表示以某个向量 $\boldsymbol{y}_i$ 为中心，临界距离 ε 为半径的邻域。圆圈内的 2 个黑点表示向量 $\boldsymbol{y}_i$ 和 $\boldsymbol{y}_j$，它们处于一个邻域内，在图 2(b)中坐标为(i, j)的位置上会显示一个黑点。图 2(a)中圆圈外的一点表示向量 $\boldsymbol{y}_l$，它不在邻域内，在图 2(b)中坐标为(i, l)的位置上会显示一个白点。



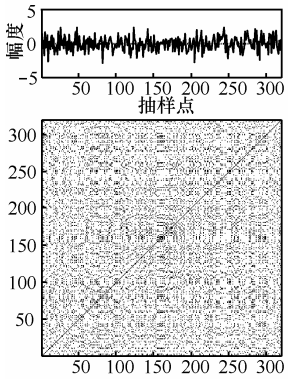
(a) Lorentz 系统的三维相空间轨迹图



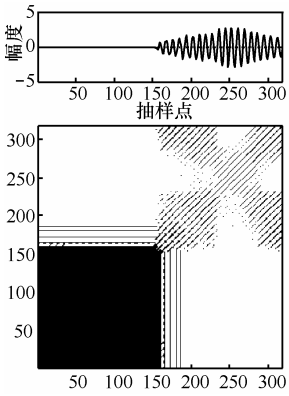
(b) Lorentz 系统的递归图

图 2　Lorentz 系统的三维相空间轨迹图及其对应的递归图

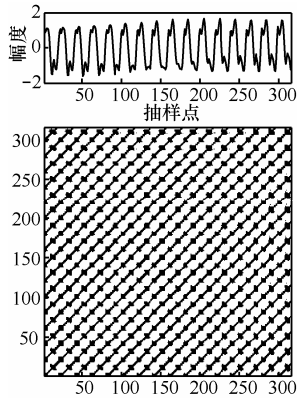
根据递归图的性质和音频信号的非线性特性，将音频信号时间序列分成 4 类：噪声型、暂态型、谐波型和混合型[11]。图 3 给出了 4 类音频信号时间



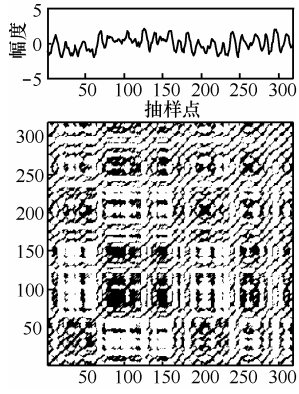
(a) 噪声型



(b) 暂态型



(c) 谐波型



(d) 混合型

图 3　音频信号的递归图

序列的递归图。噪声型音频信号的递归图如图 3 (a)所示，图中孤立散点分布均匀，很少出现较长的对角线段和竖直/水平线段，这揭示了信号的随机特性。暂态型音频信号的递归图如图 3 (b)所示，图中出现大片的、容易识别的白/黑色区域，这是由系统状态在短时间内变化较快而产生的。其中，黑色区域是稳定的前段信号彼此之间全部递归造成的，而白色区域是前半段信号与后半段冲击信号之间不递归造成的，因此，递归图的左上角和右下角几乎没有递归点。谐波型音频信号的递归图如图 3 (c)所示，图中出现类似对角线或棋盘结构的线条，其中，具有一定间距的平行对角线间的距离，揭示了状态演化的周期。而如图 3 (d)所示，混合型音频信号的递归图与噪声型相比孤立点不明显，与暂态型相比竖直/水平结构较少，与谐波型相比存在着较少的对角线，也就是说，它混合了噪声型、暂态型和谐波型 3 种类型的特性。

递归图方法虽然计算简单且容易实现，但它只能进行定性分析。为了利用递归图对音频信号进行分类，需要引入定量递归分析（RQA, recurrence quantification analysis)[12,13]。本文利用递归图中提取出的 7 个参数作为分类特征：递归度 R_{R}、确定度 R_{D}、对角线长度均值 L_{mean}、最长对角线长度 $L_{\max}$、层状度 R_L、竖直线段长度均值 V_{mean} 和最长竖直线段长度 $V_{\max}$。这 7 个参数分别从确定性、系统稳定性、相空间轨道的分离速率、系统状态的变化快慢等角度描述了音频信号时间序列的发展状态，下面将利用这些特征对音频信号进行分类。

2.3　音频分类

本文基于决策树的方法设计了一种层次化的音频信号时间序列分类器，它是把一个复杂的四分类问题转化为 3 个简单的分类问题，采用分级的形式逐步实现，其分类流程如图 4 所示。

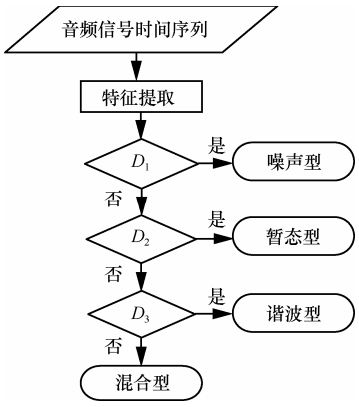


图 4　音频信号时间序列分类流程

特征提取时，选取特征向量{R_{R}, R_{D}, L_{mean}, $L_{\max}$, R_{L}, V_{mean}, $V_{\max}$}来区分 4 类音频信号，图中节点的决策规则 D_1、D_2 和 D_3 定义为

$$
\begin{aligned}
D_1&:\begin{cases} L_{\mathrm{mean}} < T_{\mathrm{Lmean}} \\ V_{\mathrm{mean}} < T_{\mathrm{Vmean}} \\ L_{\max} < T_{\mathrm{Lmax1}} \\ V_{\max} < T_{\mathrm{Vmax}} \end{cases} \\
D_2&:\begin{cases} R_{\mathrm{D}} > T_{\mathrm{RD1}} \\ R_{\mathrm{L}} > T_{\mathrm{RL1}} \\ V_{\mathrm{mean}} > T_{\mathrm{Vmean}} \\ V_{\max} > T_{\mathrm{Vmax}} \end{cases} \\
D_3&:\begin{cases} R_{\mathrm{D}} > T_{\mathrm{RD2}} \\ R_{\mathrm{L}} > T_{\mathrm{RL2}} \\ L_{\max} > T_{\mathrm{Lmax2}} \end{cases}
\end{aligned}
\tag{4}
$$

式(14)中的分类阈值 T 的确定采用如下方法[7]：随机选取音频类型 A 和 B 各 n 段，设 T 为 2 个音频类型对于某个特征的分类阈值，提取特征值，并计算均值 m_A、m_B 和方差 Δ_A、Δ_B，如图 5 所示。

设 $m_A>m_B$，如果 $m_A-\Delta_A<m_B+\Delta_B$，那么该特征不适合作为 A 和 B 的区别性特征；如果 $m_A-\Delta_A\geqslant m_B+\Delta_B$ 则有

$$
\frac{\Delta_A}{\Delta_B}=\frac{(m_A-\Delta_A)-T}{T-(m_B+\Delta_B)} \tag{5}
$$

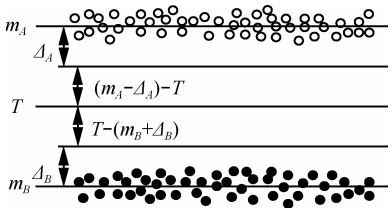


图 5　特征集的阈值确定

由式（5）便可计算出区分 A 和 B 的分类阈值

$$
T=\frac{m_A\cdot\Delta_B+m_B\cdot\Delta_A}{\Delta_A+\Delta_B} \tag{6}
$$

采用上述方法，根据式(6)分别求取阈值，这样便可以设置分类器中定量递归参数的阈值。基于阈值的分类方法是所有方法中算法最简单、计算量最小的方法，它能够做到对音频信号的实时分类。

2.4　分类性能测试

本文选取了 MPEG 数据库中鼓乐、小提琴、口琴、交响乐和流行歌曲等音频片段以及 SQAM 数据库中响棒、边鼓、三角铁和锣等相关打击乐共 20 min 的数据对分类方法进行测试，并与人工标注的标准

类别进行比较。表 1 给出了分类结果，可以看出，4 类信号的准确率均在 82%以上，误判率较低。

表 1　　音频分类方法的分类准确率

概率类型	噪声型	暂态型	谐波型	混合型
噪声型的概率	87.51%	0.08%	1.98%	7.08%
暂态型的概率	2.73%	83.81%	2.03%	3.48%
谐波型的概率	2.25%	10.76%	90.21%	7.01%
混合型的概率	7.51%	5.35%	5.78%	82.43%

3　基于音频分类的频带扩展

本文提出了一种基于非线性音频特征分类的频带扩展算法，其原理框图如图 6 所示。对于输入的 16 kHz 抽样、7 kHz 有效带宽的宽带音频，首先基于递归图和定量递归分析技术在时域对时间序列进行分类，共分为 4 类：噪声型、暂态型、谐波型和混合型；之后对每类时间序列进行抽样率转换，以转换到 32 kHz 抽样；然后进行时频变换，将时间序列变换到频域，以便进行高频部分的重建；变换到频域的 4 类信号在去除子带包络之后将分别采用与之相适应的方法进行频谱细节扩展；同时提取时频特征，利用高斯混合模型估计高频子带能量；最后进行融合并调整能量增益，时频反变换输出抽样率为 32 kHz，带宽为 14 kHz 的超宽带音频信号。

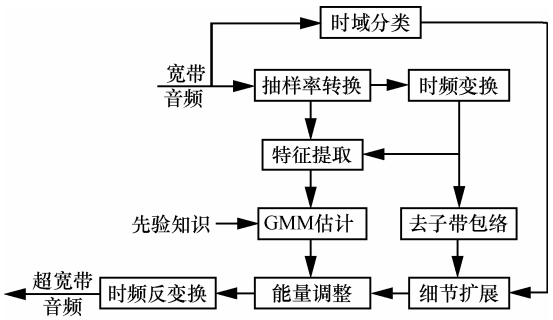


图 6　音频频带扩展的原理

3.1　抽样率转换

本文提出的音频频带扩展算法的输入信号是 16 kHz 抽样、7 kHz 有效带宽的宽带音频信号。为了达到将带宽扩展为 14 kHz 的要求，首先需要进行抽样率转换。本文采用内插的方式进行抽样率转换，宽带音频 $s(n)$经过零值内插变为 $x'(n)$

$$x'(n)=\begin{cases}s(n/2), n=0,\pm2,\pm4,\cdots\\ 0, \quad \text{其他}\end{cases} \tag{7}$$

经过简单插值后的 32 kHz 抽样信号 $x'(n)$，其频谱在高频成分中会出现低频频谱的镜像，即原始低频频谱信息会沿着 8 kHz 谱线镜像搬移到 8 kHz 以上的高频频谱中。为了保证抽样率转换后音频频谱信息不发生改变，插值后的信号 $x'(n)$还需要进一步通过一个截止频率为 7 kHz 的低通滤波器，以去除 7 kHz 以上频谱的多余镜像，并保持插值上抽样后音频信号的平滑性。滤波后的 32 kHz 抽样信号记为 $x(n)$，其有效带宽为 7 kHz。

3.2　时频变换

时频变换将时间序列 $x(n)$转换到频域，以获取信号的频域信息。由于基于时域混叠抵消思想的调制叠接变换(MLT, modulated lapped transform)可以有效地抑制帧间块效应，精确地进行信号重构，同时具有与离散余弦变换相似的能量集中特性。因此，本文采用 MLT 对音频信号进行时频分析。抽样率为 32 kHz 的时间序列 $x(n)$在处理过程中以帧为单位，每帧时长为 20 ms（即 640 个样点），设 MLT 的叠接时长也是 20 ms。将当前帧和前一帧信号共 1 280 个样点的信号进行 MLT 变换，变换产生 640 个 MLT 频谱系数 $f_{\text{mlt}}(i)$，$0\leqslant i<640$。

$$f_{\text{mlt}}(i)=\sum_{n=0}^{1\,279}\sqrt{\frac{2}{640}}\sin\left(\frac{\pi}{1\,280}(n+0.5)\right)\cdot \cos\left(\frac{\pi}{640}(n-319.5)(i+0.5)\right)x(n) \tag{8}$$

MLT 变换得到的 640 个频谱系数表示的是 0~16 kHz 的频谱信息，由于输入的时间序列 $x(n)$只有 7 kHz 的有效带宽，因此时频变换后的频谱系数中每帧只有前 280 点有效，其余 360 点幅值均为 0。

3.3　去子带包络

借鉴语音信号频带扩展算法的思想，音频信号的频带扩展可分为高频频谱包络扩展和频谱细节扩展 2 部分。其中，频谱包络可以利用均方根能量(F_{rms})进行估计。将 280 点的 MLT 频谱系数 $f_{\text{mlt}}(i)$划分为 14 个子带，每个子带包含 20 个频点，计算各个子带的均方根能量

$$F_{\text{rms}}(r)=\sqrt{\frac{1}{20}\sum_{n=0}^{19}f_{\text{mlt}}(20r+n)f_{\text{mlt}}(20r+n)}\ ,0\leqslant r<14 \tag{9}$$

去除子带包络后的MLT系数 $f'_{\text{mlt}}(i)$ 表征信号的频谱细节信息。

$$f'_{\text{mlt}}(i)=\frac{f_{\text{mlt}}(i)}{F_{\text{rms}}(r)},0\leqslant i<280,r=\left\lfloor\frac{i}{20}\right\rfloor \tag{10}$$

3.4　频谱细节扩展

高频频谱细节恢复的准确性直接影响着频带扩展后音频的音色。传统的频谱细节扩展通常采用频谱搬移、频谱折叠、综合多带激励和非线性失真等方法中的一种进行扩展，这种单一的处理方法没有充分考虑到音频的差异性，在很大程度上掩盖了不同音频高频成分的本质规律，使得扩展后的音频存在较大的频谱失真，从而直接影响了信号的音色。本节将采用 4 种方法分别对噪声型、暂态型、谐波型和混合型音频信号的频谱细节进行扩展，以提高频带扩展后的音频听觉质量。

3.4.1　噪声型

图 7 是一帧超宽带噪声型音频信号（抽样率 32 kHz，带宽 16 kHz，时长 20 ms）的时域波形图、MLT 系数谱和去除子带包络的 MLT 系数谱。从图中可以看出，信号去除子带包络的频谱细节的高低频差别很小，幅度变化也很小，因此可以考虑采用频谱折叠法[14]对噪声型信号的频谱细节进行扩展。在语音信号的频带扩展中，频谱折叠法是最常用的频谱细节扩展方法，它是将低频的频谱折叠到高频部分，特别适用于高低频细节差异较小的信号。其实现方法也很简单，直接将去除子带包络的频谱细节上抽样即可。

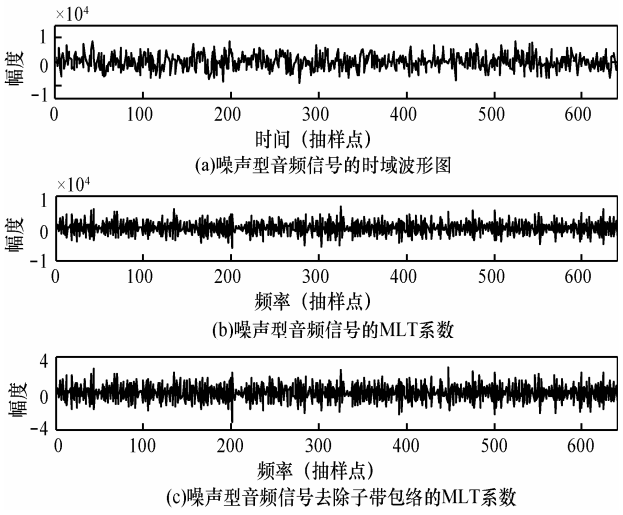


图 7　噪声型超宽带音频信号的时域波形图和 MLT 系数谱

3.4.2　暂态型

图 8 给出了一帧抽样率为 32 kHz，带宽为 16 kHz，时长为 20 ms 的典型的暂态型超宽带音频信号的时域波形图、MLT 系数谱及其去除子带包络的 MLT 系数谱。暂态型信号的时域波形中存在很多幅度突然变得很大或突然降得很小的部分，形成类似冲激信号的波形。从图中的时域波形图可以看出，波形在 300 点附近忽然产生一个冲击，幅度值在短时间内增幅较大，继而是类周期的结构，并且其去除子带包络的 MLT 系数的高低频细节部分差别很小，因此也可以考虑采用频谱折叠的方法来实现暂态型信号的细节扩展。然而暂态型信号的频谱并不像噪声型信号那样分布均匀，其低频 0~2 kHz 的部分仍然可能存在能量较高的部分，因此在实际操作中，一般将 3.5~7 kHz 带宽的部分向高频折叠，折叠 2 次即可完成频谱细节的扩展。

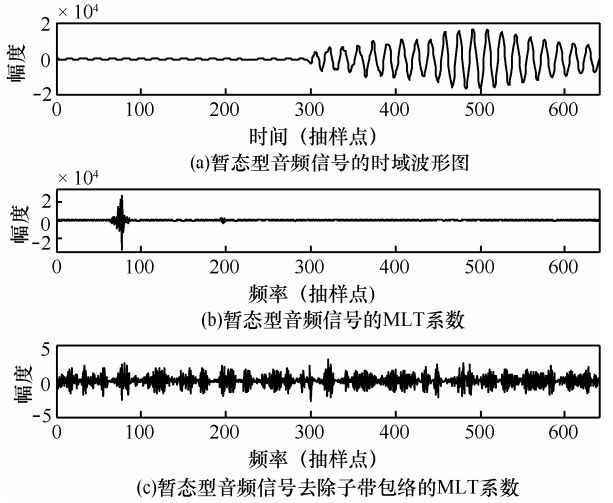


图 8　暂态型超宽带音频信号的时域波形图和 MLT 系数谱

3.4.3　谐波型

图 9 给出了一段典型的谐波型超宽带音频信号（口琴乐）的语谱图，图中明显可以看出，信号的高低频部分除了能量稍有差异外，频谱形状差别不大，存在很高的相似性。对于谐波型信号，不能采用之前的频谱折叠法进行频谱细节的扩展，因为频谱折叠会将低频段明显的谐波结构复制到高频段，从而从听觉感知上引入噪声。另外，谐波型信号的高低频谐波宽度并不一定完全一致，若利用频谱折叠法，则可能使高频段的谐波发生偏移，从而引入误差。因此考虑利用非线性动力学的相关知识进行分析。

实际上，音频信号的频谱序列同样具有非线性特征[4]，同样可以利用非线性动力学的相空间重构、非线性预测等技术进行分析。由于谐波型信号典型的谐波特性，本文将利用基于最近邻匹配原则的非线性局部预测技术对频谱细节进行扩展。其基本思想就是根据已有的低频频谱细节信息在重构相空

间中研究其运动轨迹的演变方式，并基于最近邻原则建立非线性预测模型，从而依据模型估计高频频谱细节信息。

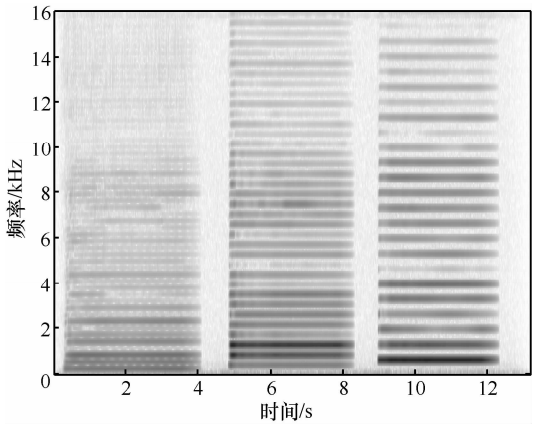


图 9 谐波型超宽带音频信号的语谱图

频谱细节扩展的目的，就是要通过 0~7 kHz 的去除子带包络的 MLT 系数 $f'_{\text{mlt}}(i)$,i=1,⋯,M 预测出 7~14 kHz 的 MLT 系数 $f'_{\text{mlt}}(i)$, i= M+1,⋯,2M。首先对序列 $f'_{\text{mlt}}(i)$, i=1,⋯,M 进行相空间重构

$$\left(\boldsymbol{y}_1,\boldsymbol{y}_2,\boldsymbol{y}_3,\cdots\boldsymbol{y}_{M-(m-1)\tau}\right)=\begin{pmatrix} f'_{\text{mlt}}(1) & f'_{\text{mlt}}(2) & \cdots & f'_{\text{mlt}}(M-(m-1)\tau) \\ f'_{\text{mlt}}(1+\tau) & f'_{\text{mlt}}(2+\tau) & \cdots & f'_{\text{mlt}}(M-(m-2)\tau) \\ \vdots & \vdots & \vdots & \vdots \\ f'_{\text{mlt}}(1+(m-1)\tau) & f'_{\text{mlt}}(2+(m-1)\tau) & \cdots & f'_{\text{mlt}}(M) \end{pmatrix} \tag{11}$$

显然，首先需要预测 $f'_{\text{mlt}}(M+1)$，也就是相空间中下一个向量点（待估相点）的最后一维，如式(12)所示，而这个向量点的前 m–1 维都是已知的。

$$\left(\boldsymbol{y}_{M-(m-1)\tau+1},\boldsymbol{y}_{M-(m-1)\tau+2},\cdots\right)=\begin{pmatrix} f'_{\text{mlt}}(M-(m-1)\tau+1) & f'_{\text{mlt}}(M-(m-1)\tau+2) & \cdots \\ f'_{\text{mlt}}(M-(m-2)\tau+1) & f'_{\text{mlt}}(M-(m-2)\tau+2) & \cdots \\ \vdots & \vdots & \vdots \\ f'_{\text{mlt}}(M+1) & f'_{\text{mlt}}(M+2) & \cdots \end{pmatrix} \tag{12}$$

最近邻匹配法的基本思想是：如果 $\boldsymbol{y}_i$ 是 $\boldsymbol{y}_j$ 的最近邻点，那么 $\boldsymbol{y}_{i+1}$ 也是 $\boldsymbol{y}_{j+1}$ 的最近邻点。因此问题转化为求取待估相点的前一个相点的最近邻点，这样，这个最近邻点的下一个相点的最后一维即为所求。但是相空间轨迹变化剧烈，为了提高预测精度，可以选择多个最近邻点，然后加权平均，以减少计算误差。预测的具体步骤如下。

1) 选取第一个待估相点 $\boldsymbol{y}_{M-(m-1)\tau+1}$ 的前一个相点 $\boldsymbol{y}_{M-(m-1)\tau}$ 为中心相点。

2) 在整个相空间中寻找中心相点的 K 个最近邻点，并求出中心相点与所有相点之间的最大距离 $d_{\max}$ 和最小距离 $d_{\min}$，假设 K 个最近邻点的下一个相点的最后一维为 $f'_{\text{mlt}}(j)$, j=1,⋯,K。

3) 利用加权局部平均预测方法进行预测

$$f'_{\text{mlt}}=\frac{\sum_{j=1}^{K} f'_{\text{mlt}}(j)\dfrac{d_{\max}-d_j}{d_{\max}-d_{\min}}}{\sum_{j=1}^{K}\dfrac{d_{\max}-d_j}{d_{\max}-d_{\min}}} \tag{13}$$

4) 选取下一个待估相点 $\boldsymbol{y}_{M-(m-1)\tau+2}$ 的前一个相点 $\boldsymbol{y}_{M-(m-1)\tau+1}$ 为中心相点，重复步骤 2)和 3)，求出预测值。依次类推，逐点恢复，直到预测出 $f'_{\text{mlt}}(2M)$ 为止，从而得到 7~14 kHz 的高频频谱细节信息。

通过基于最近邻匹配的非线性局部预测算法，低频成分对应的相点能够快速预测出高频成分对应的相点，从而恢复出了高频频谱的细节信息。

3.4.4 混合型

本文所指的混合型信号是指没有典型的噪声性、暂态性或谐波性的音频信号，它一般是上述 3 种特性的混合产物，没有明显的特性偏向性。因此，对于混合型信号，本文将采用频谱拉伸的方法进行频谱细节扩展。

宽带向超宽带音频频带扩展需要在 7 kHz以上重新构建一个 7 kHz的频段，使信号带宽提升到 14 kHz。为了保持音频的频谱细节变化不要太大，可以对$f'_{\text{mlt}}(i)$, i=140,⋯,279 进行频谱拉伸，从而估计出高频段的频谱细节$f'_{\text{mlt}}(i)$, i=280,⋯,559。在计算过程中，频谱拉伸可以通过插零的方式进行。

$$f'_{\text{mlt}}(i)=\begin{cases} f'_{\text{mlt}}(i), & i=2k \\ 0, & i=2k+1 \end{cases}(k=140,\cdots,279) \tag{14}$$

然而，这种简单的插零方法会引入一定的听觉噪声。原始低频谐波谷中噪声成分占据的频率范围会被人为地拓展开，导致人耳可以感知到原始被谐波峰掩蔽的部分噪声。因此，考虑将幅度小于某一门限T_0的频点的MLT系数值置零，从而降低谐波谷处失真对听觉感知的影响

$$f'_{\text{mlt}}(i)=\begin{cases} f'_{\text{mlt}}(i), f'_{\text{mlt}}(i)<T_0 \\ 0, \text{其他} \end{cases}(i=280,\cdots,559) \tag{15}$$

3.5 特征提取

本文利用高斯混合模型(GMM, Gaussian mix-

ture models)对高频子带能量进行估计。该方法通过计算训练数据中提取的低频时频特征和高频子带能量的联合概率密度，实现对高低频信息相关性的描述。一般情况下，训练过程中提取的宽带特征参数的性能直接影响着超宽带子带能量估计的准确性。本文分别从传统听觉感知[1]和MPEG-7音色描述[15]2 个方面对提取的宽带时频特征进行描述。

对于基于听觉感知的时频特征的选取，低频特征与高频频谱包络信息之间应该具有一定的相关性，从而保证根据低频特征就可以对高频进行准确估计。因此将选择计算复杂度较低的过零率、梯度指数、子带通量和子带能量均方根。而对于MPEG-7音色描述的特征选择，由于音色主要由频谱分布特性决定，因此将选用音频谱重心、音频扩展度和音频谱平坦度 3 种描述音频音色的特征，从而在一定程度上增强参数估计的有效性。如表 2 所示，7 个音频描述特征构成了 26 维的特征参数来刻画宽带音频的时频信息。

表 2　　宽带音频的时频特征参数

特征类型	特征名称	特征维数
听觉感知特征	过零率	1
	梯度指数	1
	子带通量	1
	子带能量均方根	14
音色描述特征	音频谱重心	1
	音频扩展度	1
	音频谱平坦度	7

3.6　GMM 估计

基于 GMM 的高频子带能量估计可以分为训练和估计 2 个阶段。在训练阶段，GMM 首先对宽带音频的时频特征和超宽带音频的高频子带能量的联合概率密度进行计算，并将 2 类特征进行结合组成 GMM 样本集，然后利用 EM 算法计算 GMM 各个分量的相关参数。最终，利用训练得到的最大似然估计参数实现对高低频特征参数联合概率密度的估计。其原理如图 10 所示。

3.7　能量调整

现有的音频频带扩展算法都是逐帧进行计算的，而本文同时采用了逐帧分类技术对 4 类音频类型分别进行扩展，因此，在 2 类不同帧的衔接处，可能会出现能量不能平滑过渡，从而影响听觉感知质量的情况。因此，有必要在频谱能量估计和频谱细节扩展后端加入能量增益调整模块，从而降低衔接噪声的影响，有效提高人耳听觉质量。

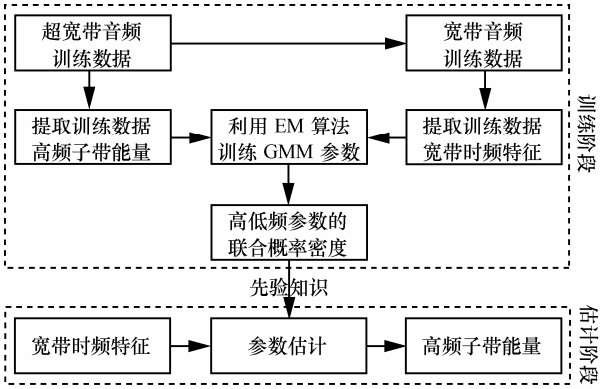


图 10　基于高斯混合模型的高频子带能量估计的原理

本文采用基于软判决[16]的码本映射算法调整能量增益，其算法原理如图 11 所示。算法包含宽带特征和增益 2 个码本，它们之间是一一对应的。

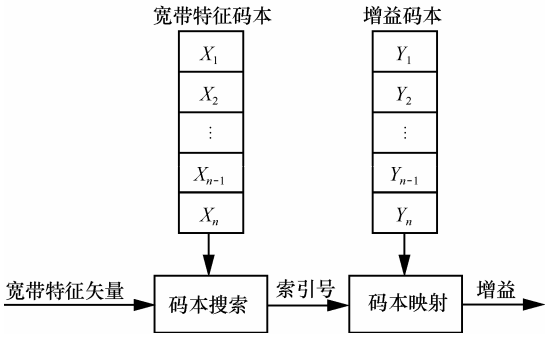


图 11　码本映射算法的原理

宽带特征码本中存储的特征向量与 3.5 节中一致，而增益码本中存储的增益为

$$g=\lg\left(\frac{\sum_{i=280}^{559}f_{\mathrm{mlt}}^{2}(i)}{\sum_{i=0}^{279}f_{\mathrm{mlt}}^{2}(i)}\right) \tag{16}$$

训练数据通过LBG算法进行码本训练，即可得到宽带特征码本及其相应的增益码本。对于输入的每一帧音频频域序列，首先计算其宽带特征向量$\boldsymbol{X}$，之后在宽带特征码本中寻找与其最相似（欧式距离最小）的N个码字$\boldsymbol{X}_1, \boldsymbol{X}_2,\cdots,\boldsymbol{X}_N$，记相应的距离分别为$d_1, d_2,\cdots,d_N$，并且这$N$个码字对应的增益分别为$\boldsymbol{Y}_1, \boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_N$，然后基于软判决的方

式对这N个增益进行加权，从而估计出高频能量增益因子

$$\hat{g}=\sum_{i=1}^{N}\frac{\boldsymbol{Y}_i}{d_i}\Big/\sum_{i=1}^{N}\frac{1}{d_i} \tag{17}$$

3.8　时频反变换

综合上述内容，将频谱能量估计和频谱细节扩展进行结合，并进行能量增益调整后得到扩展后的 7~14 kHz高频频谱信息$f_{\mathrm{mlt}}(i)$，$280\leqslant i<560$。结合 0~7 kHz的原始低频频谱，形成完整的 14 kHz有效带宽的超宽带音频频谱$f_{\mathrm{mlt}}(i)$，$0\leqslant m<560$。最终利用调制叠接反变换(IMLT, inverse modulated lapped transform)，将扩展后的MLT系数转换为 640 个样点的时间序列，从而得到抽样率为 32 kHz，有效带宽为 14 kHz的重建音频时间序列。

3.9　频带扩展在编解码器中的应用

为了测试本文提出的频带扩展算法在实际音频通信系统中的性能，本文采用ITU-T G.722.1 宽带音频编解码器作为测试对象，利用本文提出的频带扩展算法将G.722.1 解码得到的宽带音频进行频带展宽，并将扩展后的超宽带音频与G.722.1C编解码器重建的超宽带音频进行主客观质量测试。

本文提出的频带扩展算法在G.722.1 宽带音频解码器中应用的原理如图 12 所示。首先，算法从宽带音频的编码码流中解码得到子带包络值、分类方式控制字和MLT系数，并利用子带包络值重建低频子带能量信息，在此基础上利用高斯混合模型估计出高频子带能量信息。然后根据分类方式控制字表示的相应量化编码参数来解码MLT系数，从而获得低频成分的频谱细节信息，结合低频子带能量时频反变换到时域，并在时域进行分类，分类后保留类别序号。接下来，分别利用 4 种扩展算法重建出高频频谱细节信息。最终，联合高低频成分的频谱能量和细节信息进行高频能量增益调整，并在高抽样率下采用时频反变换重建出超宽带音频信号。

4　实验比较与评测结果

为了验证方法的有效性，本文将分别从客观质量测试、语谱图分析和主观偏爱测试 3 个角度对所提方法进行评价比较。

为了提高方法的普适性，训练数据和测试数据是完全独立的。参数训练阶段，GMM中的训练数据分别来自于SQAM数据库中与其特性相似的 4 类音频，高斯基函数的个数均为 64；码书映射中的参数训练数据全部来自于SQAM数据库，码本大小为 128。音频测试阶段所用的音频数据全部来自于MPEG标准音频测试数据库，共计 19 段不同风格的音频数据，包括流行音乐、电子乐、人声演唱以及交响乐等。在进行训练和测试之前，所有音频数据的信号能量都归一化到–26dB。

4.1　客观质量测试

本文采用的客观质量测试是基于FFT功率谱的对数域频谱失真测度(LSD, log spectral distance)方法[17]，公式如下

$$d_{\mathrm{LSD}}(i)=\sqrt{\frac{1}{N_{\mathrm{h}}-N_{\mathrm{l}}+1}\sum_{n=N_{\mathrm{l}}}^{N_{\mathrm{h}}}\left[10\lg\frac{P_i(n)}{P_i'(n)}\right]^2} \tag{18}$$

其中，$d_{\mathrm{LSD}}(i)$为第i 帧的对数域谱失真；$P_i(n)$为原始超宽带音频第i帧的功率谱值；$P_i'(n)$为频带扩展后音频第i帧的功率谱值；N_{l}=280，代表高频起始频率；N_{h}=560，代表高频截止频率。

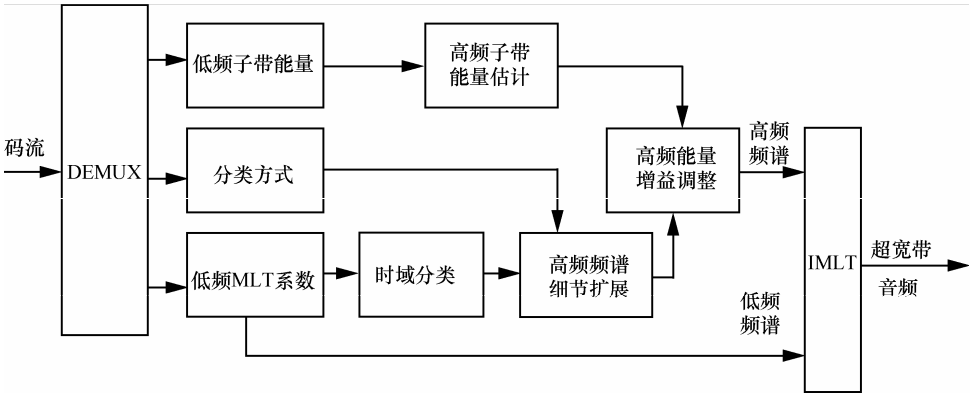


图 12　频带扩展在 G.722.1 宽带解码器中应用的原理

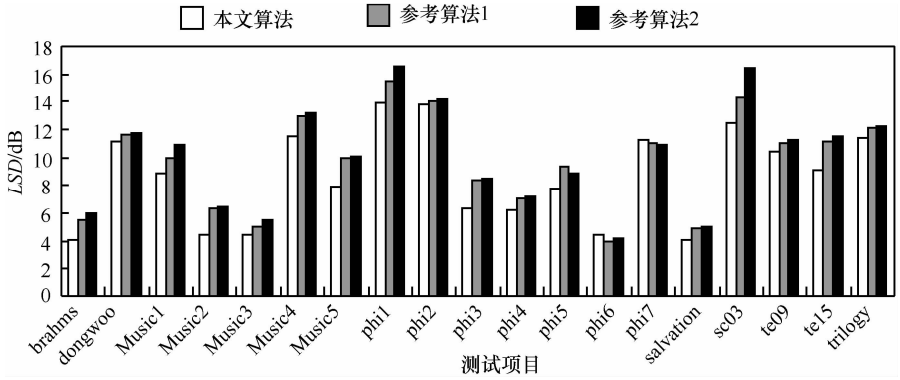


图 13　平均对数域频谱失真结果

客观质量测试将本文方法与 2 种参考音频频带扩展算法进行评价比较。其中参考算法 1 为 3.4.3 节中介绍的针对谐波型音频进行扩展的非线性局部预测算法，参考算法 2 为基于最近邻匹配的非线性音频频带扩展算法[18]。2 种参考算法均为未分类算法，即利用 2 种算法对所有类型的音频进行扩展，并与本文提出的分类方法进行比较，可以看出分类对频带扩展的影响以及影响大小。

不同算法扩展后的超宽带音频经过时域对齐，逐帧计算对数域谱失真，最后取整段音频的平均对数域谱失真，结果如图 13 所示。可以看出，参考算法 1 的平均失真程度低于参考算法 2，反映了本文提出的非线性局部预测算法的有效性。同时，本文所提方法的平均失真程度低于参考算法 1，这反映了音频分类的有效性。因此，从总体上而言，本文所提方法的频带扩展效果优于 2 种参考算法。

4.2　语谱图分析

语谱图分析过程中，G.722.1 音频编码器输入的是 16 kHz抽样、8 kHz带宽的 16 bitPCM信号，编码码率为 24 kbit/s。在解码端，解码的宽带音频经过频带扩展处理以后得到 32 kHz抽样、14 kHz有效带宽的超宽带信号。作为参考算法，G.722.1C 超宽带音频编码器的输入输出信号均为 32 kHz抽样、16 kHz带宽的 16 bit音频信号，编码码率也为 24 kbit/s。

图 14~图 16 分别给出了原始超宽带音频信号的语谱图，G.722.1C超宽带音频编解码重建得到的超宽带音频信号语谱图和G.722.1 编码器结合频带扩展技术得到的超宽带音频信号语谱图。比较 3 幅图可知，由于G.722.1C超宽带音频编解码器中高频部分采用噪声填充技术，因此图 15 中的语谱图高频部分严重失真，它完全改变了音频的音色特征。而图 16 中G.722.1 宽带音频编解码器结合基于非线性音频特征分类的频带扩展方法，其重建的音频信号的高频部分得到良好修复，尽管存在一定的频谱失真，但是音频信号的高频谐波结构在很大程度上得到了保留，更接近于原始音频。

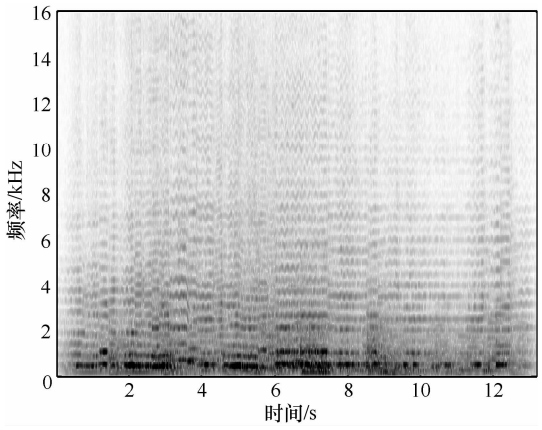


图 14　原始超宽带音频信号的语谱图

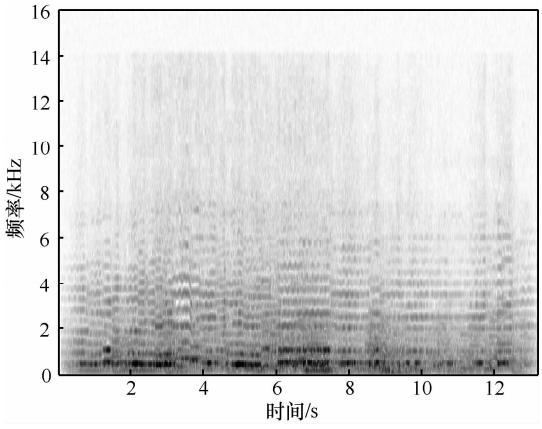


图 15　G.722.1C 编解码的音频语谱图

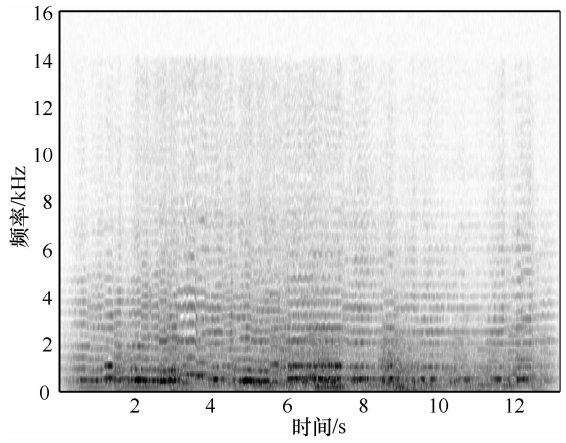


图 16　G.722.1 编解码并用本文方法频带扩展的音频语谱图

4.3　主观偏爱测试

本文同样采用A/B主观偏爱听力测试对 2 种音频重建方法进行质量评价。测试过程中邀请 12 名测听人员进行测试，这些测听人员年龄分布在 20~29 岁，没有任何听力缺陷。测试采用MPEG 数据库 19 段音频中的 8 段作为测试数据。为了保证公平性，测试数据以随机顺序进行排列。主观偏爱测试要求测听人员从 2 种测试数据中选择较偏爱的一种，或者选择两者几乎无差异。实验结果如表 3 所示。可以看到宽带音频编解码器结合频带扩展的方法重建的超宽带音频的主观听觉质量同样要优于单纯超宽带音频编码方法重建的音频。

表 3　2 种方法重建超宽带音频的主观测试结果

方法	主观偏爱比例
G.722.1+频带扩展	37.9%
无偏爱测试	28.3%
G.722.1C	33.8%

5　结束语

本文提出了一种基于音频分类的频带扩展方法。该方法引入递归图和定量递归分析技术，将音频信号时间序列分成噪声型、暂态型、谐波型和混合型 4 类，并分别利用 4 种方法对信号高频频谱细节进行预测。同时利用高斯混合模型实现了高频频谱能量的有效估计。最终，以宽带音频编解码器 ITU-T G.722.1 为平台实现了由宽带向超宽带音频信号的盲目式频带扩展。主客观测试结果表明，本文提出的方法与传统盲目式音频频带扩展算法相比，在重建音频感知质量上有明显的提高。同时，在实际宽带音频编解码系统应用中，经过该方法扩展的超宽带音频质量优于同码率下直接编解码重建的超宽带音频。

参考文献：

[1] LARSEN E, AARTS R M. Audio Bandwidth Extension-Application of Psychoacoustics, Signal Processing and Loudspeaker Design[M]. UK: John Wiley & Sons Ltd, 2004.

[2] VARY P, MARTIN R. Digital Speech Transmission-Enhancement, Coding and Error Concealment[M]. UK: John Wiley & Sons Ltd, 2006.

[3] MARTIN R, HEUTE U, ANTWEILER C. Advances in Digital Speech Transmission[M]. UK: John Wiley & Sons Ltd, 2008.

[4] SHA Y T, BAO C C, JIA M S. high frequency reconstruction of audio signal based on chaotic prediction theory[A]. ICASSP2010[C]. Dallas, USA, 2010. 381-384.

[5] LIU X, BAO C C, ZHANG L Y. Nonlinear bandwidth extension of audio signals based on hidden markov model[A]. ISSPIT2011[C]. Bilbao, Spain, 2011. 144-149.

[6] LIU H J, BAO C C, LIU X. Audio bandwidth extension based on RBF neural network[A]. ISSPIT2011[C]. Bilbao, Spain, 2011. 150-154.

[7] 崔玉强. 基于内容的音频分类方法研究[D]. 武汉: 华中科技大学, 2007.
CUI Y Q. Research on Content-Based Audio Classification[D]. Wuhan: Huazhong University of Science and Technology, 2007.

[8] 闫润强. 语音信号动力学特性递归分析[D]. 上海: 上海交通大学, 2006.
YAN R Q. Recurrence Analysis of Dynamical Characteristics for Speech Signals[D]. Shanghai: Shanghai Jiao Tong University, 2006.

[9] 刘秉正, 彭建华. 非线性动力学[M]. 北京:高等教育出版社, 2004.
LIU B Z, PENG J H. Nonlinear Dynamics[M]. Beijing: Higher Education Press, 2004.

[10] ECKMANN J P, KAMPHORST S O, RUELLE D. Recurrence plots of dynamical systems[J]. Europhys Lett, 1987, 4(9):973-977.

[11] ZHANG L Y, BAO C C, LIU X. Audio classification algorithm based on nonlinear characteristics analysis[A]. APSIPA ASC 2011[C]. Xi'an, China, 2011.

[12] ZBILUT J P, WEBBER C L. Embeddings and delays as derived from quantification of recurrence plots[J]. Phys Lett A, 1992, 171(3/4): 199-203.

[13] MARWAN N. Encounters with Neighbours-current Developments of Concepts based on Recurrence Plots and their Applications[D]. Germany: University of Potsdam, 2003.

[14] 窦庚欣. 4kbit/s 快速 DP-CELP 语音编码与频带扩展技术研究[D]. 北京: 北京工业大学, 2006.
DOU G X. Research on 4kbit/s Fast DP-CELP Speech Coding and Bandwidth Extension[D]. Beijing: Beijing University of Technology, 2006.

[15] ISO/IEC 15938-4: Information Technology-Multimedia Content Description Interface - Part 4: Audio[S]. 2001.

[16] 张丽燕, 刘鑫, 鲍长春. 基于软判决矢量量化的语音频带扩展[A]. 中国电子学会第十六届青年学术年会论文集[C]. 2010: 307-313.
ZHANG L Y, LIU X, BAO C C. Bandwidth extension of speech based on soft-decision vector quantization[A]. CIE-YC 2010[C]. 2010. 307-313.

（下转第 139 页）

IEEE Transactions on Information Theory, 2000, 46(4): 1204-1216.

[4] KATTI S, RAHUL H, HU W, *et al*. XORs in the air: practical wireless network coding[J]. IEEE/ACM Transactions on Networking, 2008, 16(3): 497-510.

[5] GUPTA P, KUMAR P R. The capacity of wireless networks[J]. IEEE Transactions on Information Theory, 2000, 46(2): 388-404.

[6] WAN P J. Multiflows in multihop wireless networks[A]. ACM MOBIHOC[C]. New Orleans, USA, 2009. 85-94.

[7] CHAPORKAR P, PROUTIERE A. Adaptive network coding and scheduling for maximizing throughput in wireless networks[A]. ACM MOBICOM[C]. Montreal, Canada, 2007. 135-146.

[8] LIU J, GOECKEL D, TOWSLEY D. Bounds on the gain of network coding and broadcasting in wireless networks[A]. IEEE INFOCOM[C]. Anchorage, USA, 2007. 724-732.

[9] SENGUPTA S, RAYANCHU S, BANERJEE S. An analysis of wireless network coding for unicast sessions: the case for coding-aware routing[A]. IEEE INFOCOM[C]. Anchorage, USA, 2007. 1028-1036.

[10] LE J, LUI J, CHIU D M. How many packets can we encode? - an analysis of practical wireless network coding[A]. IEEE INFOCOM[C]. Phoenix, USA, 2008. 371-375.

[11] GAREY M R, JOHNSON D S. Computers and Intractability: a Guide to the Theory of NP Completeness[M]. New York, USA: W H Freeman and Company, 1979.

[12] AHUJA R K, MAGNANTI T L, ORLIN J B. Network flows: Theory, Algorithms, and Applications[M]. New Jersey, USA: Prentice Hall, 1993.

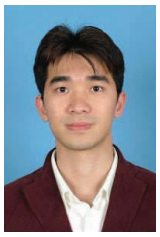
[13] IBM ILOG CPLEX optimizer[EB/OL]. http://www-01.ibm.com/software/integration/optimization/cplex-optimizer. 2012.

作者简介：

周进怡（1979-），男，湖南娄底人，清华大学博士生，主要研究方向为网络体系结构、无线网络编码等。

夏树涛（1972-），男，辽宁大连人，博士，清华大学教授、博士生导师，主要研究方向为编码理论与应用等。

江勇（1975-），男，重庆人，博士，清华大学副教授，主要研究方向为计算机网络体系结构和下一代互联网技术等。



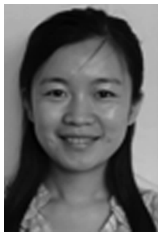
郑海涛（1978-），男，广东湛江人，博士，清华大学副教授，主要研究方向为网络科学、语义网及信息检索等。

（上接第 130 页）

[17] PULAKKA H, LAAKSONEN L, VAINIO M. 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] LIU X, BAO C C, JIA M S. Nonlinear bandwidth extension based on nearest-neighbor matching[A]. APSIPA ASC 2010[C]. Singapore, 2010. 169-172.

作者简介：



张丽燕（1983-），女，山东烟台人，北京工业大学硕士生，主要研究方向为音频频带扩展。

鲍长春（1965-），男，蒙古族，内蒙古赤峰人，博士，北京工业大学教授、博士生导师，主要研究方向为话音与音频信号处理。

刘鑫（1986-），男，北京人，北京工业大学博士生，主要研究方向为音频频带扩展。

张兴涛（1986-），女，河北保定人，北京工业大学硕士生，主要研究方向为音频频带扩展。