

基于相似关联度神经网络的音频频带扩展

刘 鑫, 鲍长春

(北京工业大学电子信息与控制工程学院, 北京 100124)

摘 要: 宽带音频带宽的限制会降低其主观质量和自然度. 本文提出了一种基于相似关联度神经网络的宽带向超宽带音频频带扩展方法. 该方法将宽带音频的精细谱重构为多维相空间, 并建立相似关联度神经网络来恢复高频成分的精谱, 同时借助高斯混合模型估计高频谱包络, 并以 G.722.1 编码器为平台实现音频信号的带宽扩展. 测试结果表明, 本文方法扩展性能优于参考方法, 其主观质量接近于 G.722.1C 超宽带编码器.

关键词: 音频编码; 音频频带扩展; 相似关联度神经网络; 相空间重构; 高斯混合模型

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2015)04-0816-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2015.04.028

Audio Bandwidth Extension Method Using Similarity Correlation Degree-Based Neural Network

LIU Xin, BAO Chang-chun

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

Abstract: The bandwidth limitation of wideband audio degrades the subjective quality and the naturalness. In this paper, a bandwidth extension of audio signals from wideband to super-wideband was proposed by using a similarity correlation degree-based neural network. Firstly, the fine spectrum of wideband audio was converted to a multi-dimensional phase space. Then, a similarity correlation degree-based neural network was built up to reproduce the high-frequency fine spectrum. In addition, Gaussian mixture model was used to estimate the high-frequency spectral envelope. Finally, the bandwidth was extended to super-wideband by the proposed method in the ITU-T G.722.1 wideband codec. Evaluation results indicate that the proposed method is preferred over the reference methods and achieves a comparable subjective quality with the G.722.1C super-wideband codec.

Key words: audio coding; audio bandwidth extension; similarity correlation degree-based neural network; phase space reconstruction; Gaussian mixture model

1 引言

感知音频编码优先重建音频信号的低频 (Low-Frequency, LF) 成分以提升编码效率. 现有的通信网络通常传输 7kHz 带宽的宽带 (Wideband, WB) 音频. 相对于 14kHz 带宽的超宽带 (Super Wideband, SWB) 音频, 高频 (High-Frequency, HF) 成分的丢失必然影响 WB 音频的自然度和表现力. 为此, 频带扩展 (Bandwidth Extension, BWE) 可在 WB 解码端人为地恢复音频传输过程中截去的 HF 成分, 进而增强听觉质量、重现 SWB 音频^[1].

目前音频编码中多采用非盲目 BWE 方法^[2,3]. 这类方法将 HF 时频能量和频谱修补信息作为边信息传输到解码端, 并据此重建 HF 成分, 获得接近透明的音质^[4]. 然而多数通信系统不会预留 1 ~ 5 kbit/s 的码率^[5]

传输边信息, 这限制了非盲目方法的实用性. 盲目式扩展方法则不需传输附加信息, 只利用 LF 时频信息, 在解码端人为地恢复 HF 成分, 扩展音频带宽^[1,4]. 鉴于此, 本文着重针对盲目式 BWE 方法展开研究.

盲目式 BWE 可分为两步: 频谱包络扩展和精细谱扩展. 频谱包络扩展通常采用高斯混合模型 (Gaussian Mixture Model, GMM)^[6]、隐马尔科夫模型^[7] 以及神经网络^[8] 等方法. 而精细谱扩展则源自于“谐波 + 噪声”模型, 其中, 频谱翻折和频带搬移 (Spectral Translation, ST) 直接将 LF 谱折叠或搬移到 HF 中, 会在高低频衔接处造成频谱偏移^[4]. 为进一步提升音质, 有研究人员采用噪声填充^[3,9] 和频谱拉伸^[10] 来恢复 HF 谱细节, 并在音频编码器上获得了良好的主观质量.

上述精细谱估计方法均源自于“谐波 + 噪声”模

型,对于音调性信号着重修复 HF 谐波成分,对于噪声性信息则保持 HF 随机结构.但是实际器乐信号具有复杂的频谱特征,在共振腔中声音的共振辐射会改变 HF 谐波结构.前期我们从音频频谱的非线性特性着手,已将非线性预测引入到精细谱扩展中^[11,12].尽管这些方法能良好地重建出 HF 谱细节,但是扩展后音频仍存在可感知失真.

鉴于此,本文提出了一种基于相似关联度神经网络(Similarity Correlation Degree-based Neural Network, SCD-NN)的音频 BWE 方法.该方法在重建相空间中建立 SCD-NN 模型.根据相矢量和神经元中心矢量的几何相似程度,逐点拟合 HF 精细谱,并借助 GMM 实现 HF 谱包络扩展.最终,重建的 HF 成分和原始 WB 音频相结

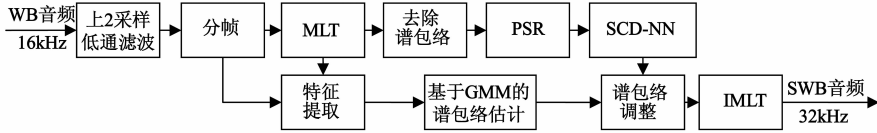


图1 基于SCD-NN的BWE方法原理框图

$$E_{ms}(r) = \sqrt{\frac{1}{20} \sum_{n=0}^{19} C_{mlt}(20r+n) C_{mlt}(20r+n)} \quad (1)$$

式(1)中, $r=0, \dots, 13$ 表示子带序号. 音频频谱的精细结构可采用能量归一化 MLT 系数 $C_{norm_mlt}(i)$ 表示如下,

$$C_{norm_mlt}(i) = \frac{C_{mlt}(i)}{E_{ms}(r)}, 0 \leq i < 280, r = \left\lfloor \frac{i}{20} \right\rfloor \quad (2)$$

接下来,利用相空间重构(Phase Space Reconstruction, PSR)法^[13,14]将一维精细谱序列转换到相空间中,并建立非线性模型来恢复 HF 精细结构.下面将逐步对本文方法进行介绍.

2.1 相空间重构

前期研究结果^[11,12]表明,音频信号相邻频点间存在着非线性相关性.根据 PSR 原理^[13,15],原始音频中蕴含的全部动力学特性可在多维相空间中展现出来.因此,本文在相空间中建立非线性模型来描述归一化 MLT 系数间的非线性关系,如下式所示,

$$\begin{aligned} C_{norm_mlt}(i) &= F[s(i-1)] \\ &= F[C_{norm_mlt}(i-1), C_{norm_mlt}(i-1-\Delta i), \\ &\quad \dots, C_{norm_mlt}(i-1-(m-1)\Delta i)] \end{aligned} \quad (3)$$

其中, $F[\cdot]$ 表示非线性函数, i 为归一化 MLT 参数的频谱序号.根据延迟重建法^[13],相矢量 $s(i)$ 可表示为

$$\begin{aligned} s(i) &= [C_{norm_mlt}(i), C_{norm_mlt}(i-\Delta i), \\ &\quad \dots, C_{norm_mlt}(i-(m-1)\Delta i)] \end{aligned} \quad (4)$$

其中,嵌入延迟 Δi 和嵌入维数 m 称作嵌入参数,并分别采用自相关法^[13]和虚假近邻点法^[16]来确定.这些重

合,形成 SWB 信号.此外,本文将所提方法应用到 ITU-T G.722.1 WB 编码器中,并对其主客观性能进行比较.

2 基于相似关联度神经网络的音频频带扩展方法

本文所提方法的基本原理如图 1 所示.输入 WB 音频的采样率为 16kHz,带宽 7kHz.经过上二采样和低通滤波后,获得的信号按照 20ms 帧长 50% 叠接的汉明窗分帧,并做调制叠接变换(Modulated Lapped Transform, MLT)^[9].7kHz 以下 MLT 系数 $C_{mlt}(i)$, $i=0 \sim 279$ 将均匀划分为 14 个子带,并利用子带均方根能量 $E_{ms}(r)$ 来描述谱包络,如下式所示,

构相矢量 $s(i)$ 可构成多维相空间 S ,

$$S = \{s(i)\}, i = (m-1)\Delta i, (m-1)\Delta i + 1, \dots, N-1 \quad (5)$$

其中, $N-1=279$ 对应 WB 音频截止频率.不同音频信号的相轨迹分析结果显示,高低频相矢量均聚集在相空间的固定区域中,并且呈现出类椭球空间结构.一旦确定嵌入参数,近邻相点会遵循相似的演变轨迹.受此启发,本文选择 SCD-NN 来描述相矢量间的演变关系,并恢复 HF 精细结构.

2.2 基于相似关联度的神经网络预测模型

本文结合最近邻映射(Nearest Neighbor Mapping, NNM)准则^[12]和 SCD-NN 来预测 HF 精细谱,其原理如图 2 所示.首先,对 LF 精细结构的幅度谱 $|C_{norm_mlt}(i)|$, $i=0, \dots, 279$ 进行 PSR,并令其重构相矢量 $s_p(i)$, $i=(m-1)\Delta i, (m-1)\Delta i + 1, \dots, 279$ 作为预测模型的输入信号,然后根据最小二乘法训练模型参数,构建 SCD-NN,最终利用 NNM 逐点预测 HF 相点,扩展 HF 精细谱,具

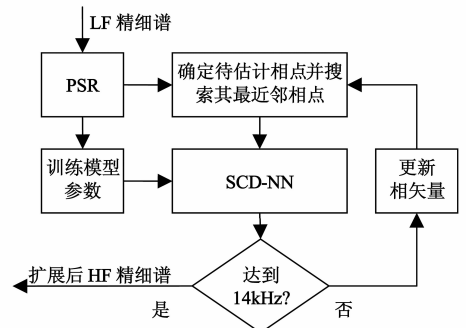


图2 基于SCD-NN的精细谱预测原理框图

体算法如图 2 所示.

2.2.1 基于相似关联度的神经网络

(1) 网络结构

本文采用前向神经网络来描述 $s_p(i)$ 及 $|C_{\text{norm_mlt}}(i+1)|$ 间的非线性关系, 其结构包括一个隐含层和一个线性输出层. 令 $s_p(i)$ 和 $|C_{\text{norm_mlt}}(i+1)|$ 分别为预测模型的输入矢量和输出值, 则预测模型可描述为

$$\begin{aligned} | \hat{C}_{\text{norm_mlt}}(i+1) | &= F[s_p(i)] \\ &= \sum_{k=0}^{K-1} w_k \varphi_k(s_p(i)) \end{aligned} \quad (6)$$

其中, $| \hat{C}_{\text{norm_mlt}}(i+1) |$ 为归一化 MLT 参数的估计值, w_k 是隐含层第 k 个神经元与输出层之间的连接权值, $K=16$ 为隐含层神经元个数, φ_k 为非线性激活函数.

(2) 相似关联度函数

本文引入了 SCD 函数^[17]作为前向网络的非线性激活函数.

首先, 分别对输入矢量 $s_p(i)$ 和隐含层第 k 个神经元的中心矢量 $\mathbf{c}_k = [c_k(0), c_k(1), \dots, c_k(m-1)]$ 进行始点零化, 则零化矢量 $s_p^0(i)$ 和 \mathbf{c}_k^0 可表示为

$$s_p^0(i) = [s_p^0(i, 0), s_p^0(i, 1), \dots, s_p^0(i, m-1)] \quad (7)$$

$$\mathbf{c}_k^0 = [c_k^0(0), c_k^0(1), \dots, c_k^0(m-1)] \quad (8)$$

其中,

$$s_p^0(i, j) = |C_{\text{norm_mlt}}(i - j\Delta i)| - |C_{\text{norm_mlt}}(i)| \quad (9)$$

$$c_k^0(j) = c_k(j) - c_k(0) \quad (10)$$

接下来, 计算相对距离,

$$\begin{aligned} d(s_p(i), \mathbf{c}_k) &= \left| \sum_{j=1}^{m-2} (s_p^0(i, j) - c_k^0(j)) \right. \\ &\quad \left. + \frac{1}{2}(s_p^0(i, m-1) - c_k^0(m-1)) \right| \end{aligned} \quad (11)$$

最终, 定义 $s_p(i)$ 和 \mathbf{c}_k 之间的相似关联度为,

$$\varphi_k(s_p(i)) = \frac{1}{1 + \alpha [d(s_p(i), \mathbf{c}_k)]^\beta} \quad (12)$$

其中, α, β 分别控制 SCD 函数的宽度和形状, 其数值可根据实际应用性能人为确定.

SCD 函数具有良好的光滑性, 其数值介于 0 和 1 之间, 并且仅与 $s_p(i)$ 和 \mathbf{c}_k 的几何相似程度有关^[17]. 若两者几何上越相似, 其 SCD 越大.

(3) 模型参数训练

确定网络结构和激活函数后, 下面对 SCD-NN 模型参数进行混合学习.

首先, 采用 k -均值聚类确定中心矢量 \mathbf{c}_k . 同样选择 SCD 函数作为空间划分准则, 对于第 k 个胞腔 Φ_k , 用 α_k 来控制激活函数宽度. 根据实际结果, 令 $\beta = 2$, α_k 可由下式确定,

$$\alpha_k = \frac{2}{\sqrt{\frac{1}{N_k} \sum_{i \in \Phi_k} d^2(s_p(i), \mathbf{c}_k)}} \quad (13)$$

其中, N_k 为第 k 个胞腔中相矢量个数.

接下来, 计算连接权值 w_k . 令 $s_p(i)$, $i = (m-1)\Delta i, (m-1)\Delta i + 1, \dots, 278$, 作为网络训练的输入矢量, 相应的 $|C_{\text{norm_mlt}}(i+1)|$ 为输出信号, 那么非线性激活函数的输出矩阵表示为,

$$\begin{aligned} \boldsymbol{\varphi} &= [\varphi_{i,j}] \\ &= \begin{bmatrix} \varphi^{(m-1)\Delta i, 0} & \varphi^{(m-1)\Delta i, 1} & \cdots & \varphi^{(m-1)\Delta i, K-1} \\ \varphi^{(m-1)\Delta i + 1, 0} & \varphi^{(m-1)\Delta i + 1, 1} & \cdots & \varphi^{(m-1)\Delta i + 1, K-1} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi^{278, 0} & \varphi^{278, 1} & \cdots & \varphi^{278, K-1} \end{bmatrix} \end{aligned} \quad (14)$$

隐含层与输出层间的连接权矢量表示为

$$\mathbf{w} = [w_0, w_1, \dots, w_{K-1}]^T \quad (15)$$

经过 SCD-NN 的输出矢量为,

$$\hat{\mathbf{C}} = \boldsymbol{\varphi} \mathbf{w} \quad (16)$$

如果给定真实输出矢量

$$\mathbf{C} = [C_{\text{norm_mlt}}((m-1)\Delta i + 1), \dots, C_{\text{norm_mlt}}(279)] \quad (17)$$

则拟合误差 ϵ 可表示为

$$\epsilon = \|\mathbf{C} - \hat{\mathbf{C}}\| = \|\mathbf{C} - \boldsymbol{\varphi} \mathbf{w}\| \quad (18)$$

为了使 ϵ 最小化, 采用最小二乘法估计连接权值矢量 \mathbf{w} ,

$$\mathbf{w} = (\boldsymbol{\varphi}^T \boldsymbol{\varphi})^{-1} \boldsymbol{\varphi}^T \mathbf{C} \quad (19)$$

根据以上步骤, 可建立 SCD-NN, 并结合 NNM, 预测高频谱细节.

2.2.2 高频谱细节的预测

在相空间中, $s_p(i)$ 的最近邻点可表示为 $s_p(i_{NN})$, i_{NN} 为近邻点对应的标号. 根据 NNM 原则^[12,13], $s_p(i)$ 和 $s_p(i_{NN})$ 通常遵循相似的演变轨迹, 则可利用 SCD-NN 近似 $s_p(i_{NN})$ 附近的演变规律. 只要两者 SCD 较小, 可根据 $s_p(i_{NN})$ 来估计 $|C_{\text{norm_mlt}}(i+1)|$, 如下式所示,

$$\begin{aligned} | \hat{C}_{\text{norm_mlt}}(i+1) | &= F[s_p(i)] = \sum_{k=0}^{K-1} w_k \varphi_k(s_p(i)) \\ &\approx \sum_{k=0}^{K-1} w_k \varphi_k(s_p(i_{NN})) \\ &= | \hat{C}_{\text{norm_mlt}}(i_{NN} + 1) | \end{aligned} \quad (20)$$

接下来, 根据 $| \hat{C}_{\text{norm_mlt}}(i+1) |$ 重构下一个待估计相矢量, 并重新做近邻点搜索和非线性预测, 直到逐点恢复出 7~14kHz 的 HF 频谱精细结构.

此外, 对于 MLT 参数的符号信息, 可将 LF 符号直接复制到 HF 中, 从而改善了其时域连续性. 最终 HF 谱细节成分可表示为,

$$\hat{C}_{\text{norm_mlt}}(i) = \text{sign}(\hat{C}_{\text{mlt}}(i - 280)) |\hat{C}_{\text{norm_mlt}}(i)| \quad (21)$$

其中, $i = 280, \dots, 559$.

2.3 基于 GMM 的高频谱包络估计

在 HF 谱包络方面,本文采用基于 GMM 的估计方法^[18],该方法首先从 32kHz 采样的 WB 音频信号中提取出时频特征矢量 F_X (包括过零率、梯度指数、子带均方根能量、子带通量、音频谱中心、音频谱扩展度、子带谱平坦度),并采用 7~14kHz 范围 14 个子带的均方根能量来表示 HF 谱包络信息 F_Y . 接下来,利用 $M = 128$ 个高斯分量的 GMM 来拟合高低频矢量的联合概率密度,且每个高斯分量均采用全协方差矩阵. 在最小均方误差准则下, HF 子带能量的估计值表示为,

$$\hat{F}_Y = E[F_Y | F_X] = \sum_{i=0}^{M-1} p(i | F_X) E[F_Y | i, F_X] \quad (22)$$

其中, $p(i | F_X)$ 表示 F_X 属于第 i 个高斯分量的后验概率, $E[F_Y | i, F_X]$ 表示给定 F_X 的条件下第 i 个高斯分量对 F_Y 的期望. 最终,利用 \hat{F}_Y 来调整 HF 谱包络,并结合原始 WB 频谱,利用调制叠接逆变换 (Inverse Modulated Lapped Transform, IMLT) 重建出 SWB 音频.

2.4 所提扩展方法在音频解码器中的应用

本文利用所提方法来扩展 ITU-T G.722.1 重建 WB 音频的有效带宽,其原理如图 3 所示. 首先,根据接收到的码流分别解码 WB 音频的谱包络和频谱细节,并利用 320 点 IMLT 生成 WB 时域信号. 接下来,借助 GMM 估计 HF 子带均方根能量. 同时,利用 SCD-NN 扩展频谱细节. 最终,结合 LF 成分采用 640 点 IMLT 重建出 SWB 音频.

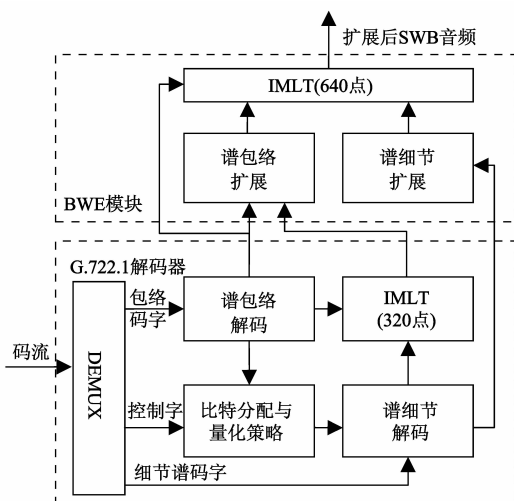


图3 所提方法在G.722.1解码器中应用的原理框图

3 实验比较和测评结果

本节将从对数谱失真和主观偏爱测试来对所提方法和参考方法进行评测。

3.1 音频数据与参考算法

GMM 线下训练数据来自于全美音乐颁奖典礼 SWB 无损音频,时长 2 小时,包括对话、音乐、演唱以及背景音等. 对该信号进行下采样得到 WB 音频,并经时间对齐形成并行数据库. 此外,本文选择了 MPEG 音频数据库中的 18 段纯净音频做测试数据,每段长度为 10~20s. 这些音频分为简单音频、复杂音频和人声演唱三种类型. 其中,简单音频最多包含三种乐器同时演奏,而复杂音频中包含三种以上乐器. 在测试之前,信号能量需归一化到 -26dBov.

测试中, G.722.1 和 G.722.1C 在 24kb/s 码率下重建的音频信号分别作为 WB 和 SWB 音频参考项. 所提方法和参考方法分别对 G.722.1 重建音频进行扩展,这些音频都将作为 BWE 测试数据,如下表所示.

表 1 测试数据的处理方式

测试项	处理方法
G.722.1	G.722.1 在 24kb/s 下重建的 WB 音频
G.722.1C	G.722.1C 在 24kb/s 下重建的 SWB 音频
ST	频谱搬移方法扩展后的 SWB 音频
NNM	NNM 方法扩展后的 SWB 音频
SCD-NN	本文所提方法扩展后的 SWB 音频

G.722.1C 编码器在 24kb/s 码率时,会利用 2kb/s 的码流来传输 HF 子带能量,并利用噪声填充重建 HF 谱细节信息. 为了与本文方法比较, ST^[4]与 NNM^[12]都采用基于 GMM 的 HF 谱包络估计器,构成参考 BWE 方法.

3.2 对数谱失真

本文采用对数谱失真 (Log Spectral Distortion, LSD)^[19]来比较所提出方法与参考项的客观性能,其定义如下,

$$d_{\text{LSD}}(i) = \sqrt{\frac{1}{N_{\text{high}} - N_{\text{low}} + 1} \sum_{n=N_{\text{low}}}^{N_{\text{high}}} [10 \log_{10} \frac{P_i(n)}{\hat{P}_i(n)}]^2} \quad (23)$$

其中, $d_{\text{LSD}}(i)$ 为第 i 帧的 LSD, P_i 和 \hat{P}_i 分别为原始 SWB 音频和处理后音频的功率谱, N_{high} 和 N_{low} 分别对应 HF 成分的上下限频率. 测试前,所有数据都需要和原始 SWB 音频进行时域对齐,并重采样到 32kHz. 最终,将整段数据的平均 LSD 值作为其客观质量测度.

所有处理后音频数据的 LSD 如图 4 所示. 经 BWE 后,复杂音频的 LSD 大于另两种音频,其主要原因是多种乐器同时演奏增大了 HF 成分的动态范围,影响了谱包络估计的准确性. 而简单音频的 HF 成分衰减剧烈,降低了其频谱失真.

从图 4 中还可以看出,由于 G.722.1C 量化传输 HF 谱包络,其重建音频的 LSD 最低. 而 ST 方法将 LF 谱细

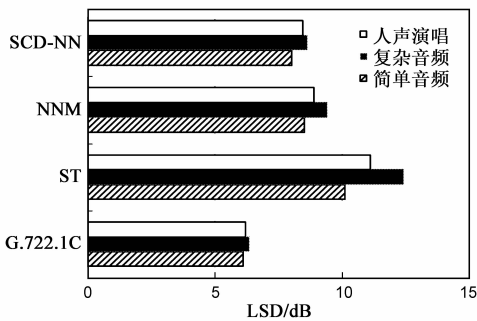


图4 处理后音频数据的LSD比较

节直接搬移到 HF 中,其 LSD 为 11dB 左右. NNM 方法的 LSD 在 9dB 左右,而本文方法比 NNM 有 0.5dB 左右的提升.

3.3 主观偏爱测试

此外,本文针对 G.722.1C、NNM 与所提方法重建音频逐对进行主观偏爱测试.测试过程中邀请了 20 名年龄在 22~28 岁的被测者来选择两种被测项中偏爱的一种,或者选择无偏爱.主观测试安排在静音室中,并选择了 MPEG 数据库中的五句作为测试数据.为了保证公平性,测试音频随机排列,且被测者在做出判断前可重复监听所有数据.

表 2 主观偏爱测试结果

	偏爱前者	偏爱后者	无偏爱
所提方法与 NNM 比较(百分比)	28.3%	21.5%	50.2%
所提方法与 G.722.1C 比较(百分比)	22.5%	24.4%	53.1%
G.722.1C 与 NNM 比较(百分比)	29.5%	22.6%	47.9%

本次主观测试分为三组:所提方法与 NNM 比较、所提方法与 G.722.1C 比较、G.722.1C 与 NNM 比较.最终结果如表 2 所示,可以看出本文所提方法扩展后的音频主观质量和 G.722.1C 编码器质量十分接近,而 NNM 方法则略逊色于其他两种处理方法.

4 结论

本文提出了一种宽带向超宽带音频盲目式频带扩展方法.该方法在重建相空间中建立了一个基于相似关联度神经网络的预测模型,并根据低频成分来恢复高频的频谱细节,同时借助高斯混合模型估计高频谱包络.最终,在 G.722.1 宽带音频编码器上,所提方法有效地扩展了重建音频的带宽.测试结果表明,本文方法恢复了重建音频的高频成分,主客观性能均优于参考频带扩展方法.此外,所提方法提升了 G.722.1 宽带编码器的性能,其重建音频接近于 G.722.1C 超宽带编码器的主观质量.

参考文献

[1] P Vary, R Martin. Digital Speech Transmission—Enhancement,

Coding and Error Concealment[M]. UK: John Wiley & Sons Ltd, 2006.

- [2] P Ekstrand. Bandwidth extension of audio signals by spectral band replication[A]. Proceedings of 1st IEEE Benelux Workshop on Model Based Processing and Coding of Audio[C]. New Paltz, NY, USA: IEEE, 2002. 53–58.
- [3] ITU-T Rec G.722.1, Coding at 24 and 32 kbit/s for Hands-free Operation in Systems with Low Frame Loss[S].
- [4] E Larsen, R M Aarts. Audio Bandwidth Extension -Application of Psychoacoustics, Signal Processing and Loudspeaker Design [M]. UK: John Wiley & Sons Ltd, 2004.
- [5] F Tobias, G Schuller. Spectral band replication tool for very low delay audio coding applications [A]. Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics[C]. New Paltz, NY, USA: IEEE, 2007. 199–202.
- [6] A H Nour-Eldin, P Kabal. Memory-based approximation of the Gaussian mixture model framework for bandwidth extension of narrowband speech [A]. Proceedings of Interspeech[C]. Florence, Italy: ISCA, 2011. 1185–1188.
- [7] P Jax, P Vary. Wideband extension of telephone speech using a hidden Markov model [A]. Proceedings of IEEE Workshop on Speech Coding [C]. Delavan, WI, USA: IEEE, 2000. 133–135.
- [8] H Pulakka, P Alku. Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband mel spectrum [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(7): 2170–2183.
- [9] 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].
- [10] F Nagel, S Disch. A harmonic bandwidth extension method for audio codecs [A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C]. Taipei, Taiwan: IEEE, 2009. 145–148.
- [11] Y Sha, C Bao, M Jia, X Liu. High frequency reconstruction of audio signal based on chaotic prediction theory [A]. Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing[C]. Dallas, Texas, USA: IEEE, 2010. 381–384.
- [12] X Liu, C Bao, M Jia, Y Sha. Nonlinear bandwidth extension based on nearest-neighbor matching [A]. Proceedings of the Second APSIPA Annual Summit and Conference[C]. Biopolis, Singapore: APSIPA, 2010. 169–172.
- [13] K Holger. Nonlinear Time Series Analysis, 2nd edition [M]. New York: Cambridge University Press, 2004.
- [14] M Small. Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance [M]. Singapore: World Scientific, 2005.
- [15] N H Packard, J P Crutchfield, J D Farmer, R S Shaw. Geome-

try from a time series[J]. Physical Review Letters, 1980, 45 (9):712 – 716.

[16] C Rhodes, M Morar. The false nearest neighbors algorithm: an overview[J]. Computers & Chemical Engineering, 1997, 21: S1149 – S1154.

[17] Liu Si-feng, Lin Yi. Grey Information—Theory and Practical Applications[M]. USA: Springer-Verlag, 2006.

[18] X Liu, C Bao, M Jia, Y Sha. A harmonic bandwidth extension based on Gaussian mixture model[A]. Proceedings of the 10th International Conference on Signal Processing [C]. Beijing, China: IEEE, 2010. 474 – 477.

[19] H Pulakka, L Laaksonen, M Vainio, J Pohjalainen, P Alku. 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.

作者简介



刘 鑫 男, 1986 年 9 月出生于北京. 现为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



鲍长春 男, 1965 年 6 月出生于内蒙古赤峰市, 博士, 教授, 博士生导师, 主要研究方向为语音与音频信号处理.

E-mail: chchbao@bjut.edu.cn