

文章编号: 2095-2163(2020)03-0069-04

中图分类号: TP37

文献标志码: A

# 基于 MDCT 的线性带宽扩展方法

李思源<sup>1</sup>, 姜林<sup>2</sup>

(1 东华理工大学 信息工程学院, 南昌 330013; 2 湖南工商大学 计算机与信息工程学院, 长沙 410205)

**摘要:** 音频带宽扩展技术是音频编解码技术的重要组成部分, 研究高频信号的编码方法, 目的是大幅减少编码码率。针对传统频域率带宽扩展框架中直接高频信号在频率域编码导致高码率和高复杂度的问题, 本文提出基于 MDCT 变换的带宽扩展编码方法, 该方法基于源滤波器模型对高频信号进行分析合成。编码中对高频信号进行 LPC 参数提取, 并将高频激励信号进行 MDCT 变换, 提取 MDCT 子带能量系数, 然后对 LPC 和 MDCT 子带能量进行矢量编码并传送至解码端。解码时将低频激励作为高频激励信号的精细结构, 采用 MDCT 子带能量对精细结构进行调整, 最后采用 LPC 合成得到高频重建信号。实验结果表明, 本文方法相比经典的 SBR 方法, 主观音质相当, 编码码率下降了 56.36%, 整体时间复杂度下降明显。

**关键词:** 带宽扩展; 音频压缩; 线性预测; MDCT 变换

## A MDCT based linear bandwidth extension method

LI Siyuan<sup>1</sup>, JIANG Lin<sup>2</sup>

(1 College of Information Engineering, East China University of Technology, Nanchang 330013, China;

2 College of Computer and Information Engineering, Hunan University of Technology and Business, Changsha 410205, China)

**[Abstract]** The audio bandwidth extension technology is an important part of the audio codec technology. It studies the encoding method of audio compression, which can greatly reduce the coding rate. For the traditional frequency domain rate bandwidth extension framework, the time-frequency transform of the original audio signal after passing the filter directly leads to the problem of coding bit rate higher. The linear encoder based on the linear prediction is used in the traditional source filter bandwidth method. In combination with the MDCT transform, the residual signal obtained by the LPC analysis is firstly used as the fine structure of the high-frequency signal, and the fine structure is subjected to MDCT transform to obtain the fine-domain frequency domain signal, and then the frequency domain of the fine structure is achieved. The signal is encoded to obtain the final reconstructed high frequency signal. The experimental results show that the encoded sound quality is equivalent to the SBR algorithm, the coding rate is reduced by 56.36%, and the complexity is also reduced.

**[Key words]** bandwidth extension; audio compression; linear prediction; MDCT transform

## 0 引言

音频带宽扩展 (Bandwidth Extension, BWE) 技术是现代音频编解码技术的重要组成部分, 主要研究的是音频压缩编码的方法, 可以仅使用极低的码率实现高质量的高频信号重建<sup>[1]</sup>。音频带宽扩展利用了高低频间具有相关性这一物理特性, 可以利用低频信号实现高频信号的重建。根据重建高频信号时是否使用了高频参数音频带宽扩展可分为盲式和非盲式两种<sup>[2]</sup>, 本文仅针对非盲式带宽扩展 (传输少量高频参数) 进行讨论。

现有的音频带宽扩展算法根据针对的音频种类不同分为 2 类, 即: 时域带宽扩展和频率域带宽扩展。本文的研究是针对类音乐信号, 因此主要探讨频率域带宽扩展技术。频率域带宽扩展技术以 2002 年瑞典 Dietz 等人<sup>[3]</sup>首次提出的频带复制技术

(Spectral Band Replication, SBR) 为代表。SBR 技术被应用于 MPEG ACC<sup>[4]</sup> 编码标准中, 其原理是将低频信号从时域转换到频率域后, 再将其复制到高频得到用于重建高频信号的信号源。使用低频信号得到高频信号的重建信号是利用了高低频间具有相关性来实现的, 此方法在高低频间相关性强的时候效果较好, 当相关性变弱时, 高频重建信号的谐波将会产生严重失真。为了解决这一问题, 2009 年, Nagel 等人<sup>[5]</sup>提出采用频谱拉伸的方法进行谐波调制方法, 用以恢复高频谐波。此后, 又提出了谐波连续调制的方法, 进一步提高了谐波恢复的效果<sup>[6]</sup>。目前, 重建音质最好的 BWE 方法是 2013 年 MPEG USAC 标准提出的增强型的 SBR (简称 eSBR) 技术<sup>[7]</sup>, 但是其使用的伪正交镜像滤波器导致编码复杂度较高。本文使用了更加稳定的离散余弦变换

**基金项目:** 国家自然科学基金 (61762005); 湖南省自然科学基金 (2019JJ40144); 湖南省教育厅科学研究项目 (18A304)。

**作者简介:** 李思源 (1995-), 女, 硕士研究生, 主要研究方向: 音频编码与处理; 姜林 (1977-), 男, 博士, 副教授, 主要研究方向: 语音信号处理、人工智能。

收稿日期: 2019-10-25

(Modified Discrete Cosine Transform, MDCT)作为时频变换方法。

## 1 基于 MDCT 的音频带宽扩展框架

传统的频域率带宽扩展技术针对类音乐信号的高频重建效果较好,一般输入信号后通过滤波器直接获取原始音频信号进行时频变换得到高频信号的边信息,本文使用基于线性预测(Linear Prediction, LP)的核心编码器与频域率核心编码器相结合的方法进行音频信号的带宽扩展降低了编码码率。

由于改进的离散余弦变换(Modified Discrete Cosine Transform, MDCT)具有稳定的量化性能,最新的 AVS2-P3 频域率编码的核心技术就是设计在 MDCT 的基础上<sup>[8]</sup>,因此本文频域率核心编码器使用的时频变换方法为 MDCT 变换,其正变换的数学定义可表示为<sup>[9]</sup>:

$$X_{i,k} = 2 \cdot \sum_{n=0}^{N-1} z_{i,n} \cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right), \quad 0 \leq k \leq N/2, \quad (1)$$

MDCT 反变换的数学定义可表示为<sup>[9]</sup>:

$$x_{i,n} = \frac{2}{N} \cdot \sum_{k=0}^{N/2-1} X_{i,k} \cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right), \quad 0 \leq k \leq N/2, \quad (2)$$

其中,  $X_{i,k}$  表示 MDCT 频谱系数;  $x_{i,n}$  表示时域信号值;  $z_{i,n}$  表示加窗时域输入序列;  $n$  表示采用点序号;  $k$  表示频谱系数序号;  $i$  表示变换块序列;  $N$  表示变换窗窗长;  $n_0$  为  $(N/2 + 1)/2$ 。

### 1.1 编码器

编码框架如图 1 所示,首先将音频信号通过高通滤波器得到高频(High Frequency, HF),然后将高频信号进行 LP 分析得到线性预测系数(Linear Prediction Coefficient, LPC)和低频激励信号  $e_h$ , LPC 系数是由式(3)通过 Levinson - Durbin 算法进行求解得到<sup>[10]</sup>,即:

$$A(z, n) = 1 - \sum_{k=1}^p a_k(n) z^{-k}, \quad (3)$$

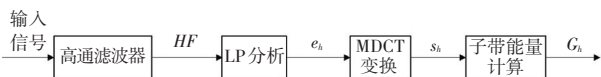


图 1 基于 MDCT 的 BWE 编码框架

Fig. 1 The encoding flowchart of BWE based on MDCT

高频激励信号  $e_h$  是原始信号  $s(n)$  与预测信号  $\tilde{s}(n)$  的误差,也叫残差信号(Residual Signal),由式(4)计算得到:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k), \quad (4)$$

LPC 系数可以用来表示信号频谱包络,但这只能表示高频粗糙包络,很难生成高质量高频信号。麻省理工学院 Larsen 等人<sup>[1]</sup>的研究表明,除了高频粗糙包络,高频精细结构对于音频主观质量也具有极其重要的影响。研究中用激励信号  $e_h$  表示高频信号的精细结构,为了获得高频精细结构的信息,通过对  $e_h$  进行 MDCT 变换得到其频域表达信号  $S_h$ ,同时计算频率域内的子带能量  $G$ ,将其传输到解码端。具体公式如下:

$$G = \{G_1, G_2, \dots, G_i, \dots, G_M\}, \quad G_i = \sum_{k=1}^N E_h^2(i, k). \quad (5)$$

其中,  $M$  为子带个数,  $N$  为子带长度,本文采用均匀划分子带<sup>[11]</sup>。

### 1.2 解码器

解码框架如图 2 所示,对解码后的低频信号 LP 分析 LPC 系数和低频激励信号  $e_l$ ,并对  $e_l$  进行 MDCT 变换,得到频域表达信号  $\tilde{S}_l$ ,接着采用线性映射将低频频域表达信号  $\tilde{S}_l$  复制到高频位置,得到高频频域信号  $\tilde{S}_{nh}$ ,再使用编码端传送的子带能量  $\tilde{G}$  进行调整,得到重建的高频频率信号  $\tilde{S}_{nah}$ ,最终通过 MDCT 反变换得到最终的高频重建时频信号  $\tilde{s}_{nah}$ ,研究中还可以采用时域能量平滑对最终的高频重建时频信号进一步提高最终信号的自然度。能量调整计算见式(6):

$$\tilde{S}_{nah} = \tilde{S}_{nh}(i, n) \cdot \frac{\tilde{G}(i)}{\tilde{G}_l(i)}, \quad 0 \leq n \leq N-1. \quad (6)$$

其中,  $\tilde{G}(i)$  为高频解码的子带能量参数;  $\tilde{G}_l(i)$  为低频子带能量;  $\tilde{S}_{nh}(i, n)$  表示通过线性映射得到的高频频率信号;  $i$  为子带序号;  $N$  为子带长度。

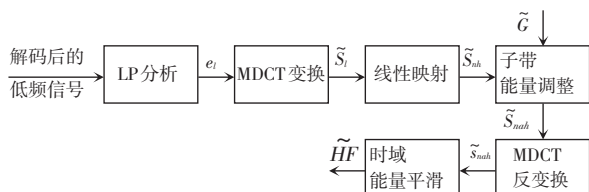


图 2 基于 MDCT 的 BWE 解码框架

Fig. 2 The decoding flowchart of BWE based on MDCT

## 2 实验设置

### 2.1 主观测试

主观音质测试阶段, 本文采用 A/B 测试对所提出的方法和 SBR 方法所得到的音频信号进行音质评测<sup>[12]</sup>, 测试材料选择的是音质测试中常用的 MPEG 的语音库中的 12 个测试序列, 包含语音、音乐、乐器等音频信号见表 1。

表 1 测试材料

Tab. 1 Test material for subjective testing

编号	名称	类型	描述	时长/s
1	es01	语音	歌唱语音	10
2	es02		德语语音	8
3	es03		英语语音	7
4	sc01	音乐	喇叭独奏	10
5	sc02		管弦乐	12
6	sc03		流行音乐	11
7	si01	乐器	大键琴	7
8	si02		响板	7
9	si03		律管	27
10	sm01		风笛	11
11	sm02		钟琴	13
12	sm03		琴弦	13

听音测试选择的是 8 名受过听音测试训练的研究生(4 男 4 女, 20 ~ 28 岁)<sup>[13]</sup>。测试过程中, 2 组测试数据随机各播放 8 次, A、B 各 4 次。测试结果见表 2。

表 2 主观听音测试结果

Tab. 2 Subjective listening results

结果	偏好/%			
	语音	音乐	乐器	平均
A(SBR)	34.5	36.5	33.8	34.9
B(本文)	36.4	34.3	33.1	34.6
无偏好	29.1	29.2	33.1	30.5

从主观听音测试结果可以看出, 本文算法相比 SBR, 编码音质相对接近, 从信号类型来看, 本文算法在语音信号的编码音质也有略微提升, 这是因为研究中使用了激励信号作为高频信号的精细结构对高频信号进行重建。

### 2.2 时间复杂度测试

时间复杂度测试阶段, 使用简单的算法运行时间测试。算法测试材料为一段长度为 368 s 的波形文件, 测试环境相同。测试运行的时间见表 3。

表 3 运行时间比较

Tab. 3 Run time comparison

	SBR	本文
编码端	12.32	8.54
解码端	9.41	7.56
总用时	21.73	16.10

从实验结果可以看出, 相比于 SBR 算法本文少了 5.63 s, 时间复杂度优势明显。

### 2.3 编码码率

编码端的参数中, 本文 BWE 模型下每帧(1 024 样点), 提取 LPC 阶数为 12, 取 16 个子带能量(均匀划分)。所有参数采用矢量量化, 量化比特率及码率见表 4。已知 SBR 的码率为 2.2 kbit/s, 本文码率下降 56.36%。

表 4 编码参数量化及码率

Tab. 4 Coding parameter quantization and code rate

	频率域模型(1 024 样点/帧)	
	高频 LP 系数	子带能量
参数个数	12	16
量化比特数	13	64
码率/(kbit · s <sup>-1</sup> )	0.96	0.96

## 3 结束语

本文针对现有的频域滤波器带宽扩展框架中使用伪正交镜像滤波器导致编码复杂度较高的问题和使用原始信号获取高频边信息的编码码率较高的问题, 提出了一种基于 MDCT 的音频带宽扩展方法, 该方法首先将通过滤波的原始信号使用 LP 分析得到的激励信号作为高频信号的精细结构, 对其精细结构进行 MDCT 变换得到精细结构的频率域信号, 再对精细结构的频域信号进行编码得到最终的重建高频信号。实验结果表明, 本文提出的方法能够有效对音频进行带宽扩展, 相比于 SBR 算法, 编码码率下降 56.36%, 复杂度也有所降低, 且编码音质相当。但本方法有一个很明显的问题, 这也是传统非盲带宽扩展的共同问题, 本方法只能在高低频信号间相关性较强的时候结果更好, 当相关性变弱, 音频音质将会急剧下降, 因此, 如何解决这一问题也是未来工作中将要研究的一个重要方向。

### 参考文献

[1] LARSEN E R, AARTS R M. Audio bandwidth extension: Application of psychoacoustics, signal processing and loudspeaker design[M]. USA: John Wiley & Sons Inc., 2004.  
 [2] ISER B, SCHMIDT G, MINKER W. Bandwidth extension of speech signal [M]. New York: Springer-Verlag New York Inc., 2010. (下转第 75 页)