

一种基于匹配跟踪的谐波和独立谱线正弦模型实现方案

王晶 赵胜辉 匡镜明
(北京理工大学信息工程学院 北京 100081)

摘要 该文基于匹配跟踪的谐波和独立谱线正弦模型提出了一种用于参数音频编码的实现方案,输入音频信号的正弦成分由谐波联合独立谱线共同表示,分析合成过程采用50%叠接相加(OLA)消除帧间不连续,匹配跟踪(MP)算法在频域提取模型参数(幅度、频率和相位)大大降低运算复杂度。谐波谱线基频由谐波积谱法(SHS)获得,各次谐波频点确切位置借助MP迭代过程推出,并进行二次曲线拟合,对应的谐波幅度采用LPC谱包络近似。独立谱线的提取有效弥补了谐波提取不足。实验证明该文提出的正弦模型实现方案可以很好地表示出音频信号中的平稳成分,并对低比特率的参数音频编码有一定的借鉴性。

关键词 音频编码, 正弦模型, 谐波谱线, MP算法, 谐波积谱法

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2006)06-1016-05

An MP-based Harmonic and Individual Lines Sinusoidal Modeling in Parametric Audio Coding

Wang Jing Zhao Sheng-hui Kuang Jing-ming

(School of Information Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

Abstract This paper presents an MP-based harmonic and individual lines sinusoidal modeling used in parametric audio coding. The input audio signal can be represented by harmonic plus individual lines and the model parameters (amplitudes, frequencies and phases) are extracted by MP (Matching Pursuit) algorithm which can lower computational complexity. The fundamental frequency is obtained by subharmonic summation. Harmonic frequency points are decided by MP exactly and are curve-fitted by quadric polynomial. The corresponding harmonic amplitudes are simulated by LPC spectrum envelope. In the decoder, audio signal is reconstructed with IFFT and OLA. The experimental results show that this sinusoidal modeling can represent the stationary parts of the audio signal very well.

Key words Audio code, Sinusoidal model, Harmonic lines, Matching Pursuit (MP) algorithm, Subharmonic summation

1 引言

参数音频编码建立在信号源模型基础上,源模型假定输入音频信号可以分解成不同的成分,每一类分解均可由少量的模型参数表示。选择合适的信号模型和参数可以达到高质量的合成音频。目前研究的参数音频编码主要利用正弦模型、噪声模型和暂态模型对音频信号进行分析,提取相应的参数,分别表示音频信号中的平稳、类噪声和瞬时成分^[1]。正弦波参数提取是参数音频编码中极其关键和重要的部分,它的正确与否关系到整个编码器的效率和合成音频信号的质量。

正弦模型是参数音频编码系统的基础,用来表示音频信号中的平稳成分,最初仅是提取一定数量正弦分量,即所谓的独立谱线,编码端每帧传输正弦谱线的幅度、频率和相位信息,解码端对相邻帧连续的正弦谱线参数进行帧间内插,

发生变化的频率和幅度传输差值,最终产生每点的正弦波分量并在时域合成^[2]。独立谱线表示正弦模型有很好的灵活性,易提取参数,但传输比特随着提取谱线根数增加线性增长。大部分的音乐信号含有丰富的谐波成分,建立谐波模型并由少量的模型参数表示,可以大大降低编码比特。这样,正弦模型就扩展为谐波谱线和独立谱线共同表示,依次对音频信号中的平稳成分进行编码可获得更好的合成效果。

本文采用加窗叠接相加技术来消除帧间不连续,对输入音频信号首先进行50%的叠接分帧,每帧加Hamming窗作为每次的分析信号,对此用MP算法从频域提取正弦波参数。首先进行谐波谱线的提取,采用谐波积谱法估算^[3]分析帧基频和谐波根数取得很好的效果,利用MP迭代过程推出其余各次谐波频点,对应拟合后的谐波频点精确提取谐波谱线幅度,谐波频率点集合用二次曲线拟合系数表示,幅度用AR模型系数(即LPC系数和增益)代表,大大降低了所需参数的个数,利于量化编码;提取谐波后,分析帧中未提出的主要

正弦成分用少量的独立谱线弥补,同样正弦波参数由MP算法获得。合成端,解码出所需的参数重新得到谐波和独立谱线的频率、幅度,考虑人耳对相位参数不敏感,采用随机相位,最后用IFFT叠接相加重构音频信号中的平稳成分。简要的谐波+独立谱线表示正弦模型示意图如图1所示:

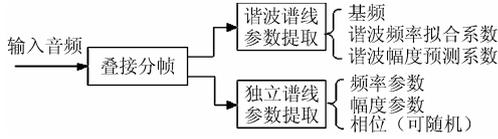


图1 谐波+独立谱线表示正弦模型

Fig.1 Harmonic + individual lines sinusoidal model

2 MP 算法提取正弦参数

本文提出的谐波和独立谱线表示正弦模型,模型参数亦即正弦波的幅度、频率、相位可以采用频域选峰或者是基于帧的分析合成方式获得。分析合成方法按照残差信号能量递减的顺序迭代提取正弦参数,这样能够保证收敛性,使得原始音频和合成音频之间的误差趋于0。MP匹配跟踪算法即是一种有效的分析合成方法^[4],当选定合适的字典元素,它可以给出压缩率很高的解决方案,又因为其可以采用FFT的快速实现,计算复杂性低,所以很适合用于正弦模型参数提取。

MP 算法是一种利用迭代从一个高度冗余的结构化字典中得到矢量的线性组合的信号分解方法,也就是说,该算法是从词典中选择一组原子,它们与待分析信号的结构最为匹配。MP 算法是一种贪婪算法,在每次递归中,该算法找到使下次余量能量最小的字典元素。找贪婪的最优字典元素首先将信号与字典元素作相关运算,然后从中找到最大的相关系数,例如字典元素选用复指数序列 $g_m = \left\{ g_m[n] = \frac{1}{N} \exp\left(j2\pi \frac{m}{M} n\right); n = 0, 1, \dots, N-1; m = 0, 1, \dots, M-1, \right.$ 字典子空间包括字典元素和其复共轭,相关系数以共轭对 α_k 和 α_k^* 出现,所以我们只需计算出第 k 次迭代过程中的最大自相关系数 $\alpha_k = a_k e^{j\theta_k}$ 。第 k 次递归后的残差信号为

$$\begin{aligned} r_{k+1}[n] &= r_k[n] - \alpha_k g_{m_k} - \alpha_k^* g_{m_k}^* \\ &= r_k[n] - \frac{2a_k}{N} \cos\left[2\pi \frac{m_k}{M} n + \theta_k\right] \end{aligned} \quad (1)$$

上式表明,每次递归在字典空间上得到的投影都是一个常数幅度、常数频率的余弦信号,利用叠接相加能够使得帧与帧之间平滑过渡。每个余弦的幅度和相位可以从相关系数 $\{\alpha_0, \alpha_1, \dots, \alpha_{k-1}\}$ 得到,频率可以从字典元素位置序列 $\{m_0, m_1, \dots, m_{k-1}\}$ 得到。MP 算法是按信号能量意义上的收敛,在每次迭代过程中都从原始信号中减去能量最大的分量,通过加权的方法就可以得到其他意义上的收敛。加权 MP 算法

可以用 DFT 来解释,对于第 k 次迭代,

$$\text{频率序号 } m_k = \max_m^{-1} \frac{|\langle g_m, r_k \rangle_w|}{\langle g_m, g_m \rangle_w} \quad (2)$$

$$\text{自相关系数 } \alpha_k = \frac{R_k^w[m_k/M]}{W[0/M]} = \frac{\alpha_k e^{j\theta_k}}{W[0/M]} \quad (3)$$

α_k 表达式中的分子 $R_k^w[m_k/M]$ 即为第 k 个残差信号的 M 点 DFT。本文中选取的是针对正弦模型的结构化字典,内积加权矩阵 W 确定为一个窗(即 W 是一个对角矩阵,对角线上的元素为窗的系数),这样用 DFT(如果字典大小为 2 的幂,也可用 FFT)就能方便地实现相关运算,并且可以通过每次迭代时直接更新相关系数来减小计算复杂度^[2]。

迭代 k 次后,得到 k 个能量最大的正弦分量,其截止条件可以采用固定的迭代次数,也可以采用残差信号中包含的正弦分量低于某一门限。第 l 帧的重建合成音频信号:

$$\begin{aligned} \hat{x}^l[n] &= \sum_{k=0}^{K-1} \alpha_k g_{m_k}[n] + \alpha_k^* g_{m_k}^*[n] \\ &= \sum_{k=0}^{K-1} \frac{2a_k}{W[0/M]} \cos\left[2\pi \frac{m_k}{M} n + \theta_k\right] \end{aligned} \quad (4)$$

3 个正弦参数频率、幅度和相位可以存储为

$$\left\{ A_k = 2a_k / W[0/M], f_k = m_k / M, \theta_k = \theta_k \right\} \quad (5)$$

如图 2 所示,某帧音频信号用 MP 算法提取 10 根正弦谱线,峰值处标号表示提取的顺序。

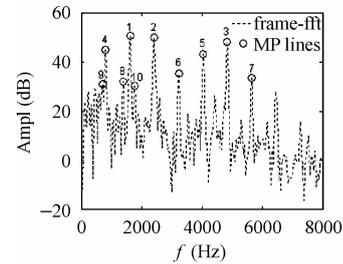


图2 MP 算法提取正弦谱线示意图

Fig.2 Illustration of extracting sinusoidal lines by MP algorithm

3 谐波谱线提取过程

谐波谱线提取主要包括:基频提取,MP算法提谐波谱线,二次曲线拟合谐波频点、获取各次谐波幅度以及幅度的 LPC 拟合。此过程简述如下:(1)用谐波积谱法计算基频 f_0 ,并得到谐波的根数 N_h ;(2)用 MP 算法提取谱线,与基频倍数误差在容许范围内的谱线就归为谐波,迭代过程中未被提出的少量谐波由基频的倍频表示;(3)对提出的谐波谱线频点进行二次曲线拟合,频率参数由基频 f_0 和拟合参数 S 表示;(4)由拟合参数重新得到频点,并借助 MP 算法在对应各次谐波频点上提取精确的谐波幅度;(5)谐波幅度值用 LPC 谱包络 (AR 模型参数) 来表示, LPC 系数及增益从本帧合成谐波中获得。相位信息可以不传输而在解码端用随机相位代替。

3.1 基频提取

基音检测是参数音频编码中一个非常重要的问题,其困难在于输入信号的不确定性和变化性,目前还无法准确而可靠地检测出所有声音信号的基频。基频提取可以在时域、频域及倒谱域进行,本文比较了经典的基于短时自相关、倒谱域峰值最大求基频以及频域谐波峰值累积三种方式,发现对于音频信号而言,谐波积谱法^[3]较其他方法更为有效。

信号 $x(n)$ 的谐波积谱为

$$P_n(e^{j\omega}) = \prod_{r=1}^k |X_n(e^{jr\omega})| \quad (6)$$

采用对数谱的形式为

$$P_n(e^{j\omega}) = \sum_{r=1}^k \log |X_n(e^{jr\omega})| \quad (7)$$

$X_n(e^{jr\omega})$ 的频谱结构是 $X_n(e^{j\omega})$ 在频域压缩 r 倍的结果,它的 r 次谐波的峰值位置总是与 $X_n(e^{j\omega})$ 的基波频率对齐,如图3所示列出了 $r=1, 2, 3$ 各项对数谱示意图。因此,无论原来的频谱中基波成分是否具有最高的峰,谐波积谱 $P_n(e^{j\omega})$ 都会在基频处出现最高峰,即使所分析的声音信号截去了一定的基频分量,它仍可在基频处获得最高的峰值。这是一种谐波累积的效果,可以有效地避免基音频率估值落到倍频或分频上的可能性。这种技术特别能对抗加性噪声,因为噪声频谱中没有相干的结构,尤其在基频处有干扰的音乐段比之时域自相关法和倒谱最大法更有优势。我们可以从第4节的实验结果图6(b)和图7(b)看出本文采用谐波积谱法估计基频效果令人满意。

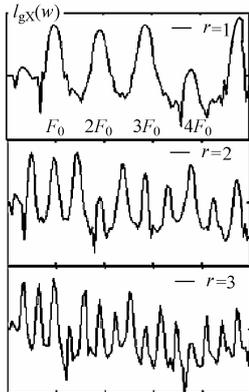


图3 对数谐波积谱中各项压缩频谱示意图(F_0 代表基频)

Fig.3 Compressed spectrums of log-SHS

(F_0 is the fundamental frequency)

3.2 基于 MP 算法提谐波谱线

由于 FFT 精度有限造成基频提取误差,其余各次谐波频率如果直接用估计出的基频的倍频表示,倍数越高误差越大。为了更准确地决定各次谐波频点位置,本文在频域上借助 MP 算法迭代过程推出各次谐波频率,并由二次多项式对

频点曲线进行拟合。MP 算法每次提取出能量上较大的正弦

波,而谐波能量一般比较明显,在迭代过程中容易发现大部分谐波谱线。根据前一步求出的谐波根数设置 MP 算法的迭代次数为 1.5 倍的谐波根数($N_i = 1.5 \times N_h$)。每次迭代中检验提出的谱线是否是谐波的倍频,考虑容差范围,我们认为如果谱线频率与基频的倍频差异不超过两倍的频率分辨率(频率分辨率由 FFT 的点数决定),则此谱线归为谐波。每次提取出谐波频率后,都对基频值 f_0^i 进行加权平均更新。设第 i 次提取的第 h 根谐波频率为 f_h^i ,则下一次迭代需要的基频为 $f_0^{i+1} = 0.5f_0^i + 0.5f_h^i/h$ (h 是谐波 f_h^i 相对于 f_0^i 基频的倍数)。设置一个谐波位置标志向量 HarmFlag,提出谐波的位置置 1,则剩余在迭代过程中没有提出的少量谐波谱线频点由基频的倍频直接表示。基于 MP 算法提谐波谱线的流程图如图4所示。

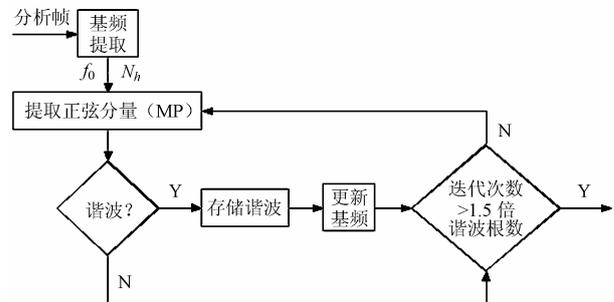


图4 基于 MP 算法提谐波谱线流程图

Fig.4 Extract harmonic lines based on MP algorithm

3.3 频点拟合及各次谐波幅度的获得

由于 FFT 分辨率有限,上一步提出谐波频率近似是基频的整数倍,为了获得更高的精确度,我们用二次多项式对频点曲线进行拟合,公式如下:

$$f_h[i] = f_0 \cdot i \cdot (1 + S \cdot i), \quad i = 1, \dots, N_h \quad (8)$$

$f_h[i]$ 为各次谐波频率, f_0 为基频, S 为拟合系数, N_h 是谐波根数。 f_0 和 $f_0 \cdot S$ 分别对应二次多项式 $y = ax + bx^2$ 的系数 a 和 b , 如此各次谐波频率可由 f_0 和 S 这两个参数获得。

由上式计算出拟合后的各次谐波频点,再次借助 MP 迭代过程在对应位置上提取谐波幅度参数,这样重新单独提取谐波幅度使得频率和幅度更精确地对应起来,使合成音频谱线从谱上更逼近原始音频。

3.4 LPC 谱包络拟合谐波幅度

可以根据已知谐波频率在谐波幅度包络上找到相应的幅度值,整个幅度包络即对应合成谐波信号的幅度谱。AR 模型能很好地进行谱预测,而且可以方便解线性方程来得到 AR 模型参数。这个过程相当于对合成谐波信号用自相关法求解 LPC 系数和增益,并用 p 阶的全极点滤波器频响 $H(\omega)$ 的幅度谱来拟合谐波信号的幅度谱包络 $S(\omega)$, 如下式:

$$S(e^{j\omega}) = |H(e^{j\omega})| = \left| \frac{G}{A(e^{j\omega k})} \right| = \left| \frac{G}{1 - \sum_{k=1}^p a_k e^{-j\omega k}} \right| \quad (9)$$

式中 G 为 LPC 增益因子, p 为预测阶数, a 为预测系数。

具体过程是在编码端,提取完谐波参数后,用 IFFT 重构成谐波分量,此信号通过 LPC 分析获得 LPC 系数和增益因子,解码端通过式(9)生成谐波谱幅度包络,并在对应频点上取得幅度信息。如图 5 所示一段谐波帧原始的幅度谱和 LPC 谱包络拟合后的结果,预测阶数设为 16。可以看出在低频段以及波峰处拟合较好,而高频段和波谷处稍差,但这种差异从听觉上不会对合成音频造成很大的影响,而 LPC 系数转换为 LSP 系数进行矢量量化所需的编码比特将大大降低。

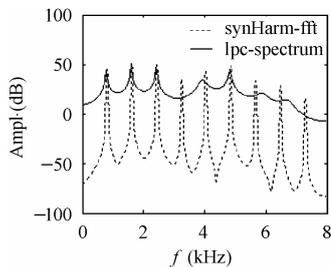


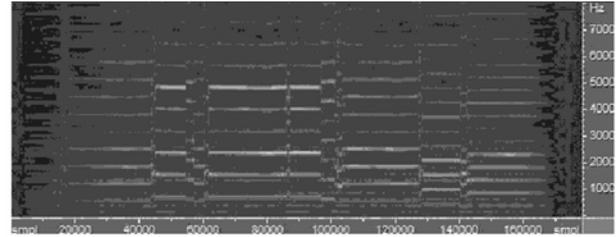
图5 谐波帧幅度(虚线),
LPC 谱包络拟合(实线)

Fig.5 Amplitude of harmonic frame (dotted),
LPC spectrum envelope (solid)

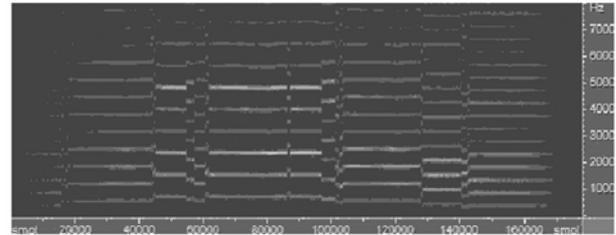
4 实验结果及分析

本次实验中,选取的输入音频信号类型有独奏曲、协奏曲、合奏曲和声乐等几种,利用谐波联合独立谱线的正弦模型表示音频信号中的平稳成分,其中协奏乐和合奏乐在低频段主奏乐器的谐波成分受旁奏乐器干扰较大,表示相对困难,因此图示主要以这两类音乐段为例来说明实验结果。此外,非正式的主观测听实验对这几类包含丰富正弦谱线的音乐段中平稳成分合成结果进行测试表明合成音质清晰,可很好地表现出原信号主特征,令人满意。

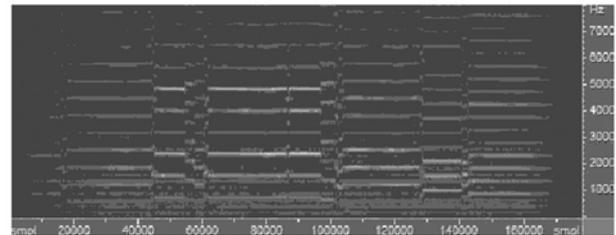
实验条件:原始音频带宽 $BW=8\text{kHz}$,采样频率 $f_s=16\text{kHz}$,帧长 $N=512$,步长 $L=256(50\%$ 叠接),MP字典大小 $M=1024$,加权矩阵 W 为 1024 点的 Hamming 窗。谐波谱线提取的模型参数有基频(用对数谐波积谱法获得)、频率拟合系数及用于谐波幅度包络表示的 16 阶 LPC 预测系数(转换为 LSP 参数便于矢量量化)和预测增益。独立谱线提取中 MP 算法迭代次数 K 设定为 15,即另外再提 15 根独立谱线,传输谱线的频率、幅度参数,相位由瞬时频率积分得到。用 MATLAB 实现仿真结果,图 6 和图 7 所示分别为一段协奏乐和合奏乐的(a)原始音频、(b)谐波合成信号及(c)谐波+独立谱线合成信号的语谱图(cooledit 软件中显示结果)。



(a) 原始音频



(b) 谐波合成信号



(c) 谐波+独立谱线合成信号

图6 协奏乐语谱图

Fig.6 Spectgram of concerto

结果分析如下:

(1)音频信号的分析与合成采用了帧间叠接相加(OLA)的方法,可以有效避免帧间不连续现象,重建效果比较理想。

(2)谐波谱线模型参数量化编码所需比特不受音频信号谐波根数的影响,可较好地适用于固定比特的参数音频编码。模型参数用二次曲线拟合系数和 AR 模型参数表示在不损害音质的基础上提高了音频信号压缩效率。

(3)对数谐波积谱法估计谐波基频收到很好的效果,如图 6(b),图 7(b)所示原始音频中较明显的谐波成分均准确提出。对于多基频的音乐信号(多个乐器发声),谐波积谱法可以检测到反映音乐特征的主要谐波组基频,而其它正弦成分由独立谱线提取过程弥补。如果提取一组谐波后的剩余信号仍然可以区分出很强的谐波组,则可以利用独立的谐波提取过程进行二次提取,只是牺牲一些编码复杂度。此外可以考虑借心理声学模型针对感觉上有意义的正弦谱线进行提取,被掩蔽的正弦成分不编码或分配较少比特。

(4)借助 MP 算法对各次谐波频率进行确切位置提取,采用固定根数的截止门限,只需提取少量独立谱线即可弥补谐波谱线表示的不足,根据 MP 算法特性,独立谱线大部分集中在对人耳听觉影响较大的低频段。快速的匹配跟踪算法相比传统的对一帧信号的 STFT 用频域选峰的方法提取正弦谱线更有效且降低了编码复杂度。

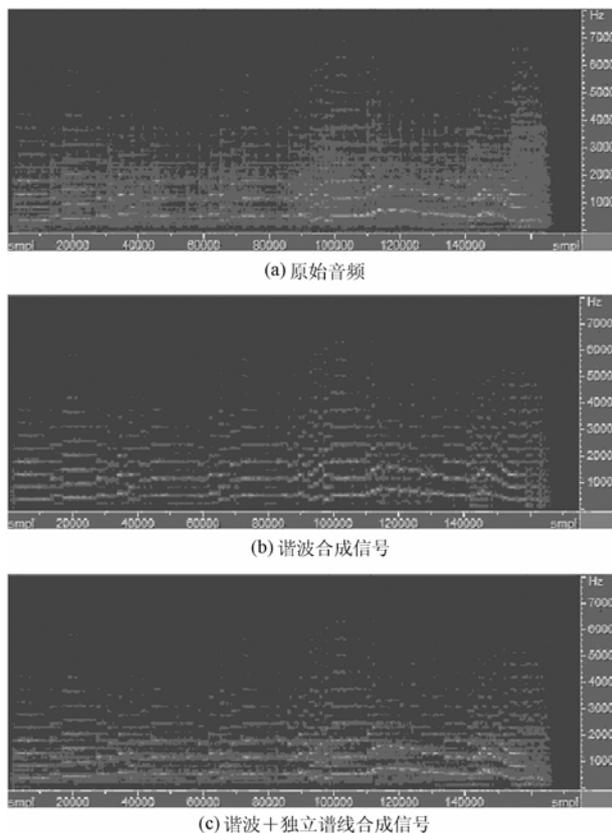


图7 合奏乐语谱图

Fig.7 Spectrogram of instrumental ensemble

(5)从图6(c),图7(c)所示合成信号可看出,实验中的正弦模型很好地表示出原始音频中的平稳成分,即正弦成分,非主观测听也获得良好的效果。进一步,在谐波和独立谱线表示平稳成分的基础上再配合噪声模型表示类噪声成分,暂态模型表示瞬时成分(如敲击声)就可以合成出完整的音频信号,而采用目前较流行的频带扩展(BWE)技术则可获得更低的编码比特。

(6)本文的后续工作中对正弦提取后剩余的噪声成分用基于感知的巴克带频谱包络重构噪声模型进行表示,此外用基于能量比值的办法有效地检测出音乐信号中可能出现的暂态信号这类瞬时成分,并对其在DCT(离散余弦变换)域提取正弦参数编码传输,实践证明扩展的正弦(谐波+独立谱线)+噪声+暂态的模型对音乐信号编码效果显著,值得进一步探索研究。

5 结束语

本文提出的谐波+独立谱线正弦模型表示方法相比仅用独立谱线表示音频信号中的正弦成分可以节省编码比特,谐波和独立谱线相应的模型参数借助MP算法获得并在频域进行分析合成,提高了编码效率。采用叠接相加(OLA)的方法可以消除帧间不连续现象,谐波积谱法很好地对音频信号基频进行估计,各次谐波频点经由MP迭代过程准确推出,LPC谱包络有效地对谐波幅度谱进行拟合,此种方案实现正弦模型可以很好地表示出原始音频中的平稳成分,且对低比特率的参数音频编码有一定的借鉴性。

参考文献

- [1] Serra X, Smith J O. Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *International Computer Music Journal*, 1990, 14(4): 14-24.
- [2] George E B, Smith M J T. Analysis-by-synthesis/overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones. *J. Audio Eng. Soc.*, 1992, 40(6): 497-515.
- [3] Hermes D. Measurement of pitch by subharmonic summation. *J. Acoust. Soc. Am*, 1988, 83(1): 257-264.
- [4] Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 1993, 41(12): 3397-3415.
- [5] 周宏, 陈健. 基于正弦波的参数音频编码算法. *通信技术*, 2002, (2): 4-6.
- [6] 陶峻, 陈健, 曾文珺. 参数音频编码中的正弦波参数提取. *数据采集与处理*, 2003, 18(2): 161-164.
- [7] 张文耀, 许刚, 王裕国. 基于匹配跟踪的感知梯度正弦建模方法. *软件学报*, 2003, 14(3): 467-472.

王晶: 女, 1980年生, 博士生, 从事语音、音频编码的研究。

赵胜辉: 男, 1970年生, 副教授, 从事移动通信及语音、音频信号处理的研究和教学工作。

匡镜明: 男, 1943年生, 教授, 博士生导师, 从事数字通信及数字信号处理的研究和教学工作。