一种新型基频变窗音频信号分析/合成系统

杨　诚,马永杰
(西北师范大学 物理电子工程学院,甘肃 兰州 730070)

摘　要：音频信号短时谱的基频随时间会发生变化,因此其谐波成分之间的间隔也会发生变化,在时域上信号随时间会发生或快或慢的变化,这导致短时谱分析所要求的时域和频域分辨率随时间是变化的。传统的固定分析窗由于其时频分辨率固定,无法同时满足上述要求,因而对短时分析造成偏差。本文基于正弦加噪声模型提出了一个分析窗宽受基频控制的自适应新型音频信号分析/合成系统方案,有效地提高了对信号实时分析的精度。并在此基础上,进一步对分析窗的使用、正弦成分的确定和追踪以及噪声成分的分离提出了新的算法和理论依据。本系统对实现音频信号的人为改造提供一套灵活高效的系统框架基础。

关键词：基频估计;正弦成分;噪声成分;自适应窗;频率追踪;音频分析/合成
中图分类号：TN912.3　　**文献标识码**：A

A new audio analysis/synthesis system of adapted window controlled by estimated fundamental frequency

YANG Cheng, MA Yong Jie
(Department of Physics and Electrical Engineering, Northwest Normal University, Lanzhou 730070, China)

Abstract: The fundamental frequency of short time spectral always varies with the real times, which results in the span between the harmonics varies, while the time signal shows fast or slow variations. The changes of resolution of time or frequency are required in the situation illustrated above. The conventional fixed windows can not satisfy the requirement of resolution both in time and frequency, as leads to the mistake in the analysis of short time. This paper proposes a new adapted audio analysis/synthesis system based on deterministic plus stochastic model in which the size of analysis window is controlled by estimated fundamental frequency in order to improve the effect of analysis for partials. New algorithms and theoretical methods are taken in the design of analysis window, partials determination and tracking and the department of residuals. This scheme offers a robust alternative as the flexible efficient fundamental frame for the music modifying.

Key words: fundamental frequency estimation; partials; residuals; adapted window; peaks matching; sound analysis/synthesis

根据基于正弦加噪声模型(deterministic plus stochastic model),音频信号 $s(t)$ 是由幅度 $A_r(t)$ 和频率 $\theta_r(t)$ 时变的正弦成分(partials)和噪声成分(residuals)$b(t)$ 构成的,即

$$s(t)=\sum_{r=1}^{R}A_r(t)\cos[\theta_r(t)]+b(t) \tag{1}$$

其中 R 为正弦成分的个数,它会随每一帧的不同而有所不同。由于正弦成分的幅度和频率随时间变化缓慢,在短时(几十毫秒)内可以认为不变,并且根据傅里叶级数它在该时段内应该存在一个基频 F_0,而正弦成分应为 F_0 的整数倍。于是加零相移窗 w_1 后经 STFT(短时傅里叶变换)得到:

$$\hat{s}(t)=\sum_{r=1}^{R}A_r\cos\theta_r+b(t) \tag{2}$$

其中 $\theta_r=\omega_r t+\theta_{r0}$,对(2)式进行傅里叶变换:

$$S(\omega)=F[\hat{s}(t)w_1(t)]=F\left\{\left[\sum_{r=1}^{R}A_r\cos\theta_r+e(t)\right]w_1(t)\right\}=$$

$$F\left[w_1(t)\sum_{r=1}^{R}A_r\cos(\omega_r t+\theta_{r0})\right]+F[w_1(t)b(t)]=$$

$$\pi A_r \sum_{r=1}^{R} e^{j\theta_{r0}} \left| W_1(\omega-\omega_r) \right| + \pi A_r \sum_{r=1}^{R} e^{-j\theta_{r0}} \left| W_1(\omega+\omega_r) \right| + \hat{B}(\omega) \quad (3)$$

短时谱 $S(\omega)$的确定部分是由被正弦成分在频域平移的窗谱$|w_1(t)|$叠加而成的，对应的相位为常数；而短时谱的随机部分 $S(\omega)$是由噪声成分的频谱 $B(\omega)$与窗谱 $W_1(\omega)$卷积的结果[1,2]。

基于上述分析，本文提出了基频变窗分析/合成系统新颖的设计方案和新的理论方法。同以往同类模型设计方案相比，本系统的特点在于：

(1)本系统使用实时变窗分析音频信号，能够同时提供高的时域分辨率和频域分辨率，从而较好地解决了两者之间一直存在的矛盾。X.Serra 曾经提出了这个设计思想[3]，但他并没有给出具体的实施细节。

(2)为确定正弦成分，同时使用了基频谐波估计和窗谱形状估计来确定最大可能的正弦成分，提高了估计的精度。

(3)对正弦成分合成，综合了相位信息。以此为基础提出了在时域分离噪声成分，并保留了它的相位信息。

(4)在频率追踪中引入频差和平滑度进行概率分析，提高了追踪的精度与速度。这与运算复杂的 HMM 方法[4,5]相比效率提高了许多。

(5)峰值检测、正弦成分与噪声成分分离、变窗调整和基频估计同时进行，提高了效能。

此外，本系统并没有使用传统的加合成方法(add synthesis)[6,7]，原因是它的运算量太大并且需要对幅度和相位进行插值。因此，决定使用高效的 OLA 方法进行合成[2]。

1 系统方案总体分析

整个系统由短时傅里叶分析(STFT)、基频估计(F_0 Estimator)、正弦成分处理(Partials Processing)和噪声成分处理(Residuals Processing)四部分构成。STFT 部分负责对信号加分析窗和短时变换，分析窗的宽度受反馈的基频周期实时调控。由于这里使用的是无相移窗，因此需要对样点进行解调(Demodulate)。解调后的信号进入正弦成分处理模块，首先快速粗略地检测出短时谱的峰值(peaks)，在粗选的峰值中进一步筛选出可能性最大的正弦成分进行人为修改，以加入某种效果。最后在频域对正弦成分进行再合成，经复数 IFFT 后变成时域信号，去掉分析窗 $w_1(t)$得到确定部分的信号 $s_d(t)$。从确定的正弦成分输出的信号经复数 IFFT 后得到未修改过的正弦成分合成的时域信号 $s'_d(t)$，它去减原加窗后的时域信号 $\hat{s}_d(t)$，获得分离出的噪声成分时域信号 $\hat{s}_s(t)$，去掉分析窗得到 $s_s(t)$，该信号与 $s_d(t)$相加得到最终的合成信号 $s'(t)$。

对基频估计是在时域采用简单高效的自相关函数实现的，本文使用高精度的累积均值归一化差分函数(CMNDF)[9]。音频信号基频是时变的，但可以认为短时不变，为提高测量的精度，提出使用自反馈变窗分析的方法，使 $w_2(t)$宽度保持在基频周期 T_0 的两倍左右。估计出的基频同时也要反馈给分析窗 $w_1(t)$，为精确分析正弦成分，$w_1(t)$的宽度保持在基频周期 T_0 的 4.5 倍左右。也就是说整个系统是工作在受基频控制自适应的状态中，系统框图如图 1 所示。

2 实现方法

2.1 窗问题的分析

2.1.1 STFT 使用的分析窗

为简化运算，希望使用一个没有相移的分析窗，这就要求窗使用关于原点对称的本地时间[10]。设输入音频信号 $s(t)$，分析窗 $w(t)$宽度为 M，短时分析的帧间隔为 I，且 $r=0,1,2,\cdots\cdots$，分析如下：

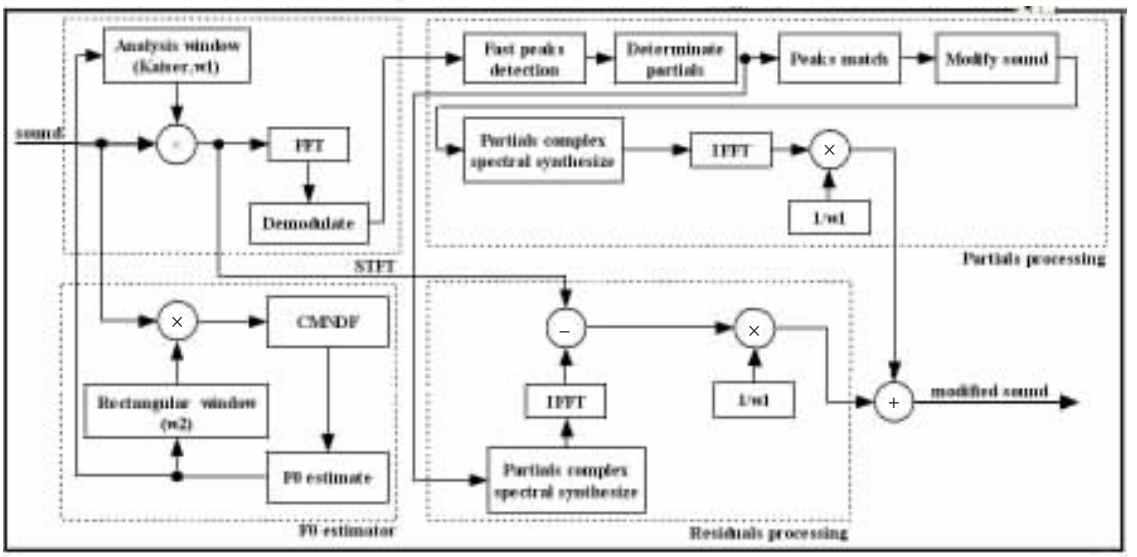

图 1 基频变窗分析/合成系统框图

$$S_r(k)=\sum_{n=\frac{M-1}{2}}^{\frac{M-1}{2}} w(n)s(n+rI)e^{-j\frac{2\pi}{M}kn} \quad (\text{设 } t=n+rI,\ n=t-rI)$$

$$=\sum_{t=\frac{M-1}{2}+rI}^{\frac{M-1}{2}+rI} w(t-rI)s(t)e^{-j\frac{2\pi}{M}k(t-rI)}$$

$$=\left[\sum_{t=\frac{M-1}{2}+rI}^{\frac{M-1}{2}+rI} s(t)w(t-rI)e^{-j\frac{2\pi}{M}k(t-rI)}e^{-j\frac{2\pi}{M}krI}\right]e^{j\frac{2\pi}{M}krI} \quad (4)$$

(4)式方括号中为经典理论中的 STFT。由于窗 $w(t)$ 具有对称性，可以得到(5)式。

$$s_r(k)=s(t)*w(t)e^{-j\frac{2\pi}{M}kt}e^{-j\frac{2\pi}{M}krI}e^{j\frac{2\pi}{M}krI}$$

$$=\left[s(t)*w(-t)e^{-j\frac{2\pi}{M}kt}e^{-j\frac{2\pi}{M}krI}\right]e^{j\frac{2\pi}{M}krI}$$

$$=\left[s(t)*w(t)e^{j\frac{2\pi}{M}kt}e^{-j\frac{2\pi}{M}krI}\right]e^{j\frac{2\pi}{M}krI}$$

$$=S'_r(k)e^{j\frac{2\pi}{M}krI} \quad (5)$$

$$S'_r(k)=S_r(k)e^{-j\frac{2\pi}{M}krI}$$

如果使用没有相移的分析窗，即看成信号在时域向左平移而时窗不动进行短时分析得到的 $S_r(k)$，它与时窗平移而信号不动的短时分析得到的 $S'_r(k)$存在差别。对于同一 k 而言，$S_r(k)$、$S'_r(k)$均可看成时域信号，但 $S'_r(k)$只是包络没有载波，而 $S_r(k)$是载波被调制过的信号。如果直接使用 $S_r(k)$将会导致短时谱样点变化过快，无法分析正弦成分的有效变化。因此，必须使用(5)式，也就是说，短时谱的每一个幅度样点要乘 $\exp\{-j2\pi/M(F_0)\cdot krI(F_0)\}$，其中 M、I 均为基频 F_0 的函数，受基频控制，从而将 $S_r(k)$的载波消除。称这种操作为解调(Demodulate)，见图 1。

2.1.2 变窗的使用

由于音频信号每一帧的基频会缓慢地变化，从而每一帧的谐波成分之间的间隔将会有所变化，即对频率分辨率的要求是变化的。要很好地区分相邻的两个谐波成分，相邻两个被谐波移动的窗谱主瓣应该在第一个过零点处相遇，才可以清晰地分辨两个谐波成分，如图 2 所示。当 F_0 变小时，则要求窗谱变小才能清晰分辨两个谐波成分，即对频域分辨率的要求提高，同时时域窗会变宽；而当 F_0 变大时，说明高频成分增多，则要求较小的时域窗才能捕捉信号快速变化的部分，即对时域分辨率的要求提高。也就是说低频信号对频域分辨率要求较高，而高频信号对时域的分辨率要求较高。但这两种要求对时窗宽度的要求是相反的，对于一个固定宽度的时窗是不可能同时满足这两种要求的。因此，如果要对音频信号的频率成分进行准确有效地分析，必须根据 F_0 实时的调整窗的宽度，即使用变窗进行实时的分析。

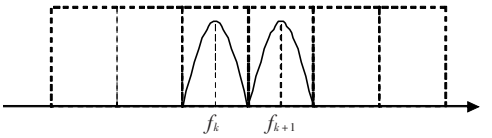

图 2 谐波成分的分辨

一般窗宽 M 取基频周期 T_0 的 4 倍以上，并且在进行 FFT 时为获得更大的频率分辨率要在时窗两边添零至长度为 N，以增加变换宽度。N 一般取 M 的两倍左右，但必须为 2 的整数次幂[8]。

$$F_0=f_{k+1}-f_k$$

$$B_f=k\frac{f_0}{N}=k\frac{f_0}{2M}$$

$$B_f\leqslant F_0$$

$$k\frac{f_0}{2M}\leqslant F_0$$

$$M\geqslant\frac{k}{2}\frac{f_0}{F_0}=\frac{k}{2}P \quad (6)$$

这里 B_f 为窗谱带宽，P 为离散基频周期点数，k 为窗谱主瓣宽度离散点数。当要求时窗宽度 M 为基频周期的 4.5 倍时，根据(6)式 $k=9$，窗谱主瓣宽度为奇数，峰值便可单取一点。

对于信号进行短时分析时我们使用性能好且灵活的 Kaiser 窗 $w_1(t)$，宽度要求是 T_0 的 4.5 倍；进行基频估计时使用矩形窗 $w_2(t)$，宽度要求是 T_0 的 2 倍，而基频估计的运算宽度是 T_0 的 4 倍。短时分析与基频估计对窗宽度的要求近似相等，即它们分析的可以针对同一段短时信号，因此可以使用相同的步长，即帧频相等。$w_1(t)$和 $w_2(t)$宽度均受 F_0 实时反馈控制，以决定最佳的分析尺度。

2.2 基频 F_0 估计

基频的估计分为时域方法[9,12]和频域方法[13-16]。这里对于基频的估计采用简便有效的时域方法，直接使用自相关函数(ACF)进行。不过，为了降低差错率使用 ACF 的衍生物——差分函数(DF)。DF 实际上是在比较信号 x_j 和其延迟 τ后的信号差别有多大，当 τ为信号周期时 DF 最小，如(7)式。

$$d_t(\tau)=\frac{1}{W}\sum_{j=t+1}^{t+W}(x_j-x_{j+\tau})^2 \quad (7)$$

$$d_t(\tau)=r_t(0)+r_{t+\tau}(0)-2r_t(\tau) \quad (8)$$

(8)式中 $r_t(\tau)$为信号的 ACF。$r_t(0)$、$r_{t+\tau}(0)$为信号功率，基本不变。当 τ 为信号周期时 ACF 出现最大值，而 DF 则为最小值。由于信号中噪声的存在，当 τ 不同时，$r_{t+\tau}(0)$会有波动，因此 DF 的最小值也会有波动。实验[19]表明使用 DF 检测 F_0 差错率降低了许多，但这个差错率还可以进一步减小。

文献[9]提出使用累计均值归一化差分函数 CMNDF

(Cumulative Mean Normalized Difference Function),即

$$d'_t(\tau)=\begin{cases}1 & \text{if } \tau=0\\ \dfrac{d_t(\tau)}{\dfrac{1}{\tau}\sum_{j=1}^{\tau}d_t(j)} & \text{otherwise}\end{cases} \quad (9)$$

由于第一个最小值经常发生误判，干脆把 $\tau=0$ 提到 1，并且通过归一化放大最小值，增加对最小值判断的精度。(9)式的分母中求出除 0 点以外 τ 中 $d_t(\tau)$的均值，用这个均值去除和它大小有可比性的 $d_t(\tau)$来确定第一个最小值。

$$\sum_{j=t+1}^{t+W}2(x_t^2+x_{t+T}^2)=\sum_{j=t+1}^{t+W}(x_t+x_{t+T})^2+\sum_{j=t+1}^{t+W}(x_t-x_{t+T})^2$$

$$\frac{1}{2W}\sum_{j=t+1}^{t+W}(x_t^2+x_{t+T}^2)=\frac{1}{4W}\sum_{j=t+1}^{t+W}(x_t+x_{t+T})^2+\frac{1}{4W}\sum_{j=t+1}^{t+W}(x_t-x_{t+T})^2$$

$$\frac{1}{2W}\sum_{j=t+1}^{t+W}(x_t^2+x_{t+T}^2)=\frac{1}{4W}\sum_{j=t+1}^{t+W}(x_t+x_{t+T})^2+\frac{1}{4}d_t(\tau) \quad (10)$$

为判定最小值需要给定一个最小值考虑范围，即要设置一个阈值。(10)式左侧为信号总功率 P_a，右侧第一项信号功率 P_s，第 2 项为噪声功率 P_n，$d'_t(\tau)$分式的分母近似为 $2P_a$。

$$d_t(\tau)=4(P_a-P_s)$$

$$d'_t(\tau)=\frac{4(P_a-P_s)}{2P_a}=2\left(1-\frac{P_s}{P_a}\right)\leqslant 0.1 \quad (11)$$

由(11)，仅考虑时间上第一个 $d'_t(\tau)$小于 0.1 的 $\tau=T_0$。

基频估计使用矩形窗 $w_2(t)$，根据经验规则[9]其窗宽应为所估计基频周期 T_0 的两倍左右。F_0 的估计上限不超过采样频率的四分之一(11 kHz)，下限为 10 Hz。T_0 的范围约为 0.1~100 ms，对窗宽的要求不固定。如果对较低的基频使用低于基频周期的窗，或对较高的基频使用远大于基频周期的窗，都会造成较高的差错率。因此为精确的估计基频，本文提出使用实时调整的适度宽度的窗 $w_2(t)$。$w_2(t)$宽度的调整是通过基频的自反馈控制的，如图 1 所示。

2.3 正弦成分处理

2.3.1 快速峰值检测

快速峰值检测需要快速、粗略的检出峰值，简单地使用一个 FIFO 解决这个问题，如图 3 所示。数据流进入 FIFO，对 a、b、c 进行比较，如果满足 $a<b$ 且 $b>c$，则记录 b 的位置，作为初选峰值 $Peak0$。

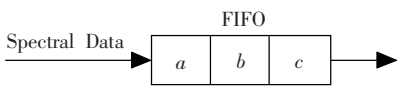

图 3 快速检峰

2.3.2 正弦成分确定

正弦成分的确定主要依据两条准则：基频谐波和形状相似。如果某个峰是正弦成分应该在基频谐波附近，并且它的形状应与窗谱形状相近[2,17]，尽管有小幅度的信号噪声扰动。具体步骤如下：

(1)根据估计出的 F_0 计算 $kF_0\in[10\text{ Hz},22\text{ kHz}]$，$k=1,2,3,\cdots$，在 $kF_0\pm\varepsilon_1$ (ε_1 为频点波动幅度) 范围内与 $Peaks0$ 比对，筛选出 $Peaks1$[14]。

(2)根据时窗 w_1 的窗谱 W_1 的主瓣上的 9 点构建窗谱矢量 $\boldsymbol{W}_1$，从 Peaks1 每一个峰中对称选出 9 点构成峰矢量 $\boldsymbol{P}_1$。利用 $\boldsymbol{W}_1$、$\boldsymbol{P}_1$ 对窗谱和测量峰的形状进行比较[13]，确定它们的相似程度 η，计算如式(12)。为相似程度 η 设定一个阈值 ε_2，对 $Peaks1$ 进行第二次筛选得到 $Peaks2$，即通过该系统的方法所确认的正弦成分，并对正弦成分由低频到高频进行编号(1,2,3,……)。对于正弦成分的确认运算量主要集中在这步，并且先前的两次筛选已经把参与运算的 $Peaks$ 的数量缩至最小，以降低计算的消耗。

$$\eta=\frac{\overrightarrow{W_1}\cdot\overrightarrow{P_1}}{2|\overrightarrow{W_1}||\overrightarrow{P_1}|} \quad (12)$$

2.3.3 峰值匹配

峰值匹配要完成对各个正弦成分的匹配追踪任务，以便今后对其各个成分进行连续地修改，实施各种音效技术。本文是在麻省理工学院林肯实验室 Mcaulay 和 Quatieri 在文献[7][8]中所使用的算法的启发下，引入频率通道 FC(Frequency Channel)和迁移概率 P_T 的概念，对原有算法进行有效地改进，提高了追踪的精度。并且与基于 HMM 算法的方法相比较，运算大为简化。

首先，要根据相邻两帧的基频 F_0 求出一个平均基频 AF_0，根据 AF_0 确定 FC，如图 4 所示。在频率通道的搜索范围内对下一帧的正弦成分进行追踪并传递编号。

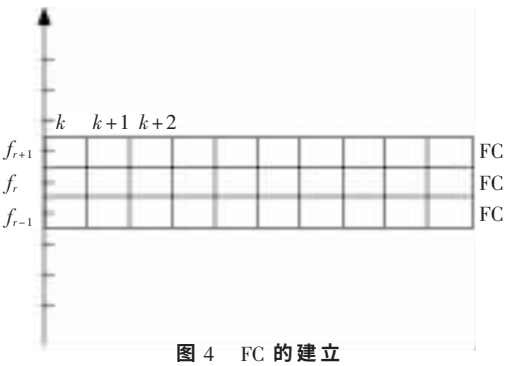

图 4 FC 的建立

以相邻两帧内的 5 个小格为搜索范围，按从上至下、从左至右的顺序进行搜索和编号传递。搜索下一帧同一正弦成分依据的准则是：同一正弦成分在相邻两帧内频差 Δf 缓慢变化，幅度差 ΔA 不会变化很大，即幅度保持平滑[13]。因此建立一个迁移概率 P_T，综合上述准则，来描述正弦成分迁移的可能性，见式(13)。

$$P_T=e^{-\Delta} \quad (13)$$

$$\Delta=\alpha\frac{\Delta f}{f}+\beta\frac{\Delta A}{A}$$

这里，Δ为幅度和频率的总偏差。f和A为本帧的频率和幅度，参考本帧的频率和幅度，可以有效地描述偏差的程度。例如，同样相差10 Hz，在20 Hz附近为50%，在20 kHz附近为0.05%。α、β为试验测定的频率和幅度偏差的权重。$\alpha+\beta=1$，且$\alpha<\beta$，即我们更倾向于幅度平滑。分四种情况考虑频率追踪，如图5所示。

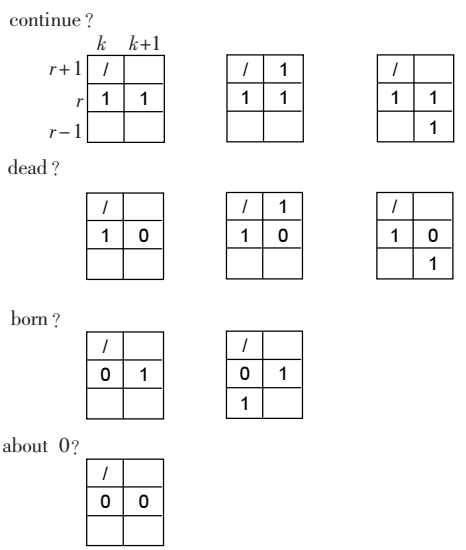

图5 频率线追踪

(1)考虑“连续”(continue)情况，第k帧第r个正弦成分存在峰值，如果第$k+1$帧第r个正弦成分也存在峰值是否连续，即它们是否为同一正弦成分。这取决于k+1帧的$r+1$、$r-1$的正弦成分情况，如果$r+1$、$r-1$的正弦成分均存在峰值，那么就需要判断第k帧第r个正弦成分向三个方向迁移的概率P_T，取概率最大的一个方向迁移，并传递频率编号。

(2)当第k帧第r个正弦成分存在峰值，而第$k+1$帧第r个正弦成分不存在峰值，这时就需要判断该频率线是否“死亡”(dead)。如果第$k+1$帧第$r-1$、$r+1$正弦成分均存在，那么就需要求出它们的P_T。当两者的P_T都小于阈值ε_d时，可以确定上一帧正弦成分已经dead。

(3)频率线还存在“诞生”(born)的情况，如果第k帧第r个正弦成分不存在峰值，第$k+1$帧第r个正弦成分存在峰值，此时尚不能确定是否一个新的正弦成分真的born。因为有可能第$k+1$帧第r个峰值是由第k帧第$r-1$个峰值迁移过来的，只有这个迁移概率小于阈值ε_b时才能肯定频率线的born，并给新频率赋新的编号。

(4)最后一种情况是第k帧第r个正弦成分不存在峰值，第$k+1$帧第r个正弦成分也不存在峰值，此时频率线维持dead，不会有其他情况发生。

2.3.4 正弦成分(Partials)再合成

经过加工改造后的正弦成分平移窗谱$|W_1(\omega)|$，相对应的相位为常数，在频域进行正弦成分频谱的复数叠加合成确定部分的频谱$s_d(\omega)$，经复数IFFT后在时域乘$1/w_1(t)$去掉窗，在时域等幅拼合，相当于使用矩形窗OLA。假如去掉窗后使用三角窗重叠50%叠加[2]，在本系统中存在两个问题：首先，两帧重叠部分的正弦成分由于频差会产生调制，从而使幅度扭曲[18]。考虑到人耳对幅度的敏感性要强于相位，因此直接使用矩形窗重叠加(OLA)，以保证幅度不变。尽管相邻两帧的正弦成分频率存在不连续，但差别不大，因此对人耳影响不大。另一个问题是，本系统使用的是变窗，窗宽随基频实时变化，这样使三角窗在保持重叠部分幅度不变的情况下进行重叠加遇到严重的困难。

$$s_d(\omega)=\sum_{r=1}^{R} e^{j\theta} r_0 |W_1(\omega-\omega_r)| \tag{14}$$

在式(14)中，由于负频率部分和正频率部分是对称的，因此只考虑正频率部分。并且对于窗谱$|W_1(\omega)|$进行复数叠加时，为简化计算只考虑它的主瓣。

再合成后，正弦成分处理完成，输出确定部分的时域信号$s_d(t)$。

2.4 噪声成分处理

合成后缺少噪声成分的音频信号缺乏真实感，现代音频信号处理已将对噪声成分的认识提高到一个新的程度[3,11,19,20]。本文提出使用时域方法对噪声成分进行处理，并且保留噪声成分的幅度和相位信息。首先对未加工改造的正弦成分在频域进行复数叠加，然后进行复数IFFT，得到时域$s'_d(t)$。该时域信号直接去减原始加窗时域信号$\hat{s}(t)$，由此得到噪声成分的时域加窗信号$\hat{s}_s(t)$。需要注意的是，信号在时域使用的是标量形式，两个信号在时域可以直接进行运算，但如果在频域直接进行幅频特性运算或相频特性运算是不行的，因为信号在频域是一种矢量形式。

由上得到的噪声成分保留原来的幅度和相位信息，并且其中还含有窗成分，因此需要乘上$1/w_1$去掉窗成分，最终得到噪声成分的时域形式$s_s(t)$，如(15)式，

$$\hat{s}_s(t)=\hat{s}(t)-s'_d(t)$$

$$s_s(t)=\frac{\hat{s}_s(t)}{w_1} \tag{15}$$

确定部分的时域形式$s_d(t)$和随机部分的时域形式$s_s(t)$在时域相加便得到最终所需要的在合成信号$s'(t)$，如式(16)。

$$s'(t)=s_d(t)+s_s(t) \tag{16}$$

本文设计的系统提供了一种对音频信号的正弦成分进行有效修改的分析/合成方案。根据基频变化的自适应窗能够有效地提高音频信号分析的准确性，从而使对谐波成分的有效修改成为可能。对于复杂多变的频率线追踪问题，本文提出根据迁移概率进行判决，提高了准确性与灵活性。在正弦成分再合成中保留了相位信息，抑制了声音的扭曲。将正弦成分与噪声成分在时域进行标量分离，完整的保留了噪声成分部分的相位和幅

度信息，增强了再合成信号声音听觉感知的真实感。对于噪声成分的人为改造方法，尚有待于进一步的研究。

参考文献

[1] SERRA X, SMITH J. Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition.Computer Music Journal, 1990,14(4): 12-24.

[2] RODET X, DEPALLE P. Spectral Envelopes and Inverse FFT Synthesis. AES 1992, San Francisco.

[3] SERRA X. Musical Sound Modeling with Sinusoids plus Noise.Musical Signal Processing, 1997:1-25.

[4] PARIS S, JAUFFRET C.Frequency Line Tracking Using HMM-based Schemes. IEEE Transactions on Aerospace and Electronic Systems, 2003,39(2):439-449.

[5] DEPALLE P, GARCFA G,RODET X. Tracking of Partials for Additive Sound Synthesis Using Hidden Markov Models. Acoustics, Speech, and Signal Processing, ICASSP-93, IEEE, 1993:225-228.

[6] SMITH J, SERRA X. PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based On a Sinusoidal Representation. ICMC-87.

[7] MEAULAY R J, QUATIERI T F.Speech Analysis/Synthesis Based on a Sinusoidal Representation. IEEE transactions on Acoustics, Speech, and Signal Processing, 1986,34(4):744-754.

[8] QUATIERI T F, MCAULAY R J. Audio Signal Processing Based on Sinusoidal Analysis/Synthesis. Kluwer International Series in Engineering and Computer …, Spring 1998:343-416.

[9] CHEVEIGNE A D, KAWAHARA H. YIN, a Fundamental Frequency Estimator For Speech and Music J. Acoust. Soc. Am. 111(4),April 2002:1917-1930.

[10] GOOWIN M M. The STFT, Sinusoidal Models, and Speech Modification. Springer Handbook of Speech Processing Benesty, Sondhi, Huang(Eds.),2008:229-258.

[11] SERRA X. A System for Sound Analysis/Transformation/Synthesis Based on A Deterministic Plus Stochastic Decomposition[D]. Stanford University, 1989.

[12] ROBEL A. Fundamental Freqency Estimation.Summer 2006 lecture on analysis, modeling and transformation of audio signals.

[13] DOVAL B, RODET X. Fundamental Frequency Estimation and Tracking Using Maximum Likelihood Harmonic Matching and HMMs. Acoustics, Speech, and Signal Processing, ICASSP-93, IEEE, 1993:221-224.

[14] MACHER R C, BEAUCHAMP J W. Fundamental Frequency Estimation of Musical Signals Using a Tow-Way Mismatch Procedure.Journal of the Acoustical Society of America, 1993:2254-2263.

[15] KLAPURI A P. Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness. IEEE Transactions on Speech and Audio Processing, 2003,11(6): 804-816.

[16] KLAPURI A P. Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes. Proc. ISMIR, 2006.

[17] DEPALLE P, HELIE T. Extraction of Spectral Peak Parameters Using a Short-Time Fourier Transform Modeling and No Sidelobe Windows. Applications of Signal Processing to Audio and Acoustics, 1997,IEEE.

[18] GOODWIN M,RODET X. Efficient Fourier Synthesis of Nonstationary Sinusoids. Proceedings of the International Computer Music Conference,1994.

[19] RODET X.Musical Sound Signal Analysis/Synthesis: Sinusoidal+ Residual and Elementary Waveform Models. Applied Signal Processing,1998.

[20] GOODWIN M. Residual Modeling in Music Analysis-Synthesis.Acoustics, Speech, and Signal Processing, ICASSP-96. IEEE. 1996:1005-1008.

（收稿日期：2009-03-12）