

# 一种特征波提取速率自适应的波形内插语音编码方案

王晶 匡镜明 赵胜辉

(北京理工大学信息科学技术学院 北京 100081)

**摘要:** 该文提出了一种特征波形提取速率自适应于输入语音帧特性的波形内插编码方案。基于双加权长时预测增益最大原则并利用前向基音判决实现了较为可靠的基音周期估计算法,用基音周期、浊音度和波表面平坦度决定波形提取速率以及SEW(Slowly Evolving Waveform)和REW(Rapidly Evolving Waveform)的更新速率。实验证明,该文提出的波形内插(WI)编码算法相比固定波形提取速率的WI算法在平均码率和计算复杂度上均有一定程度的降低,且合成语音质量明显优于4.8kbps的CELP语音编码算法。

**关键词:** 语音处理; 波形内插; 特征波提取速率; 基音估计

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2007)07-1695-05

## A Characteristic Waveform Extraction Rate Adaptive WI Speech Coding Scheme

Wang Jing Kuang Jing-ming Zhao Sheng-hui

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

**Abstract:** This paper presents a Waveform Interpolation (WI) speech coder, which characteristic waveform extraction rate is adaptive to the feature of the input frame. Efficient pitch estimation algorithm is based on the principle of maximizing double weighted Long Time Prediction(LTP) gain and uses the forward pitch detection. The waveform extraction rate and the update rate of SEW(Slowly Evolving Waveform) and REW(Rapidly Evolving Waveform) are based on the three features: pitch cycle, voicing degree and stationary degree of waveform surface. Tests indicate that the proposed WI coding algorithm has lower average bit rate and computing complexity compared to the fixed-extraction-rate WI coder and obviously deliver better quality than FS1016 CELP at 4.8kbps.

**Key words:** Speech processing; Waveform Interpolation (WI); Characteristic waveform extraction rate; Pitch estimation

### 1 引言

目前在语音编码领域,4kbit/s以下低速率高质量语音编码算法仍然是研究热点,波形内插、正弦变换、多带激励以及混合激励线性预测等参数编码器在低速率时能产生较好的语音质量<sup>[1]</sup>。其中,瑞典皇家理工学院的Kleijin博士于1991年和1994年先后提出的原型波形内插(Prototype Waveform Interpolation, PWI)<sup>[2]</sup>和特征波形内插(Characteristic Waveform Interpolation, CWI)<sup>[3]</sup>语音编码算法及相关的演变算法极具潜力,受到普遍关注。

CWI算法对清浊音进行统一模型编码,较PWI算法的特征波形提取速率高,运用波形分解将特征波表面分为快渐变波(Rapidly Evolving Waveform, REW)和慢渐变波(Slowly Evolving Waveform, SEW),分别代表类噪声和准周期成分,利用人耳对它们不同的感性需要分别量化编码,增强编码效率<sup>[4]</sup>。但目前大部分CWI算法对于固定语音帧统一采用最大的特征波形提取速率和量化编码方案,对某些基音

周期较大及特征波表面平坦度较大的语音帧波形过采样,提取特征波越多意味着CWI算法的计算复杂度越大<sup>[5]</sup>,且从平均意义上增加了特征波表示的量化比特。本文设计了一种波形提取速率自适应于输入语音基音周期的CWI编码方案,采用双加权长时预测增益最大原则,并利用前向基音判决得到最佳基音估计值和浊音度信息,特征波形提取速率反比于基音周期,对于浊音度较小的帧采用最高提取速率,高、低提取速率下分别采用不同的REW和SEW更新速率;此外通过计算特征波序列间相关系数的均值和方差判断出二维波表面平坦度,对于平坦的语音帧不进行波形分解,SEW等同于原始特征波序列。通过实验验证,本文提出的特征波提取速率自适应的波形内插编码方案降低了计算复杂度,且语音帧的平均编码速率有较大程度下降,并获得了较好的编码质量。

### 2 CWI 波形内插编码算法

CWI 波形内插编码算法将语音信号表示为一个二维特征波表面  $u(t, \phi)$  沿相位轴  $\phi$  显示按一定时间间隔提取的特征

波序列 CWs, 沿时间轴  $t$  显示特征波的渐变。解码端通过内插 CWs 获得中间的特征波, 由内插后基音轨迹得到相位轨迹, 用于从二维特征表面得到一维的合成语音信号。CWs 分解为对应于准周期信号成分的 SEWs 和对应于类噪声成分的 REWs, 根据人耳感知特性分别进行降采样量化编码。本文首先设计了一种码率为 3.8kbps, 采用最大波形提取速率(400Hz)的 CWI 实现方案, 在此基础上引入特征波形提取速率自适应方法。输入语音格式为 8kHz 采样, 16bit PCM 量化, 分析帧长为 20ms, 每帧提取 8 个特征波形。

## 2.1 编解码框架建立

首先在语音域上逐帧进行线性预测(Linear Prediction, LP)分析和基音检测。线性预测系数转化为线谱频率(Linear Spectrum Frequency, LSF)参数进行量化和内插。在残差域根据提取点的基音周期提取特征波形, 浊音帧提取的是基音周期波(Pitch Cycle Waveform, PCW), 清音帧没有特定的意义, 所以统称为特征波形(CW), 并用离散傅里叶级数(DTFS)表征:

$$s(n, \phi) = \sum_{k=1}^{\lfloor P(n)/2 \rfloor} [A_k(n) \cos(k\phi) + B_k(n) \sin(k\phi)], \quad 0 \leq \phi(\cdot) < 2\pi, \quad (1)$$

然后, 对提取的特征波形进行相位对齐和功率归一化处理, 使其能够精确地描述语音信号的渐变过程。可由下面的 DTFS 域互相关最大准则来进行相位对齐操作, 得到精确的归一化时间移位信息(假定特征波维数已经调整到相同)。

$$\tau = \arg \max_{0 \leq \tau' < 2\pi} \sum_{k=1}^M \left\{ [A_k(n_0)A_k(n_1) + B_k(n_0)B_k(n_1)] \cos(k\tau') + [B_k(n_0)A_k(n_1) - B_k(n_1)A_k(n_0)] \sin(k\tau') \right\} \quad (2)$$

其中,  $A, B$  是 CWs 的 DTFS 系数,  $M = \text{floor}(P/2)$  是 DTFS 的维数,  $n_0$  和  $n_1$  分别是前后两个相邻提取点,  $\tau$  是归一化时移, 用来调整 CWs 的 DTFS 系数以达到特征波对齐。特征波相位对齐过程互相关计算量大, 其结果是随着提取特征波个数的增加, 运算更为复杂。

随后, 用一个截止频率为 25Hz 的 17 阶 FIR 低通滤波器将归一化的特征波分解成 SEW 和 REW。此过程等效于对 DTFS 系数序列进行低通滤波得到 SEW 的 DTFS 系数, 然后将原特征波 DTFS 序列与 SEW 的相减得到 REW 的 DTFS 系数序列。

$$\left. \begin{aligned} \tilde{A}_k(n) &= \sum_{i=-8}^8 A_k(n - iL_{sf})H_{lp}(i), \\ \tilde{B}_k(n) &= \sum_{i=-8}^8 B_k(n - iL_{sf})H_{lp}(i), \end{aligned} \right\} k = 1, 2, \dots, \lfloor P(n)/2 \rfloor \quad (3)$$

其中,  $L_{sf}$  是特征波提取间隔, 由帧长和提取速率决定,  $H_{lp}$  是 FIR 低通滤波器冲击响应(汉明窗截断的 sinc 函数)。为了获得高效的编码并考虑到滤波器的非理想性, 分别对 SEW 和 REW 降低抽样率到 100Hz 和 200Hz 进行量化。特征波功率在对数域进行低通滤波降抽样量化。

解量化和恢复抽样率的 SEW 和 REW 相加即为重建的特征波形, 与解码后的功率相乘得到去归一化的特征波形。为保证特征波内插的有效性, 需要利用编码中的对齐方法对去归一化的特征波形进行重新对齐。解码后的基音周期用于内插每一点的基音周期并推得相位轨迹。恢复得到的每个提取点的特征波形进行线性内插得到每点的特征波形, 利用相位轨迹恢复一维残差信号。最后, 用重建的残差信号激励线性预测合成滤波器得到重建语音。相邻提取点上的 CW 通过线性内插得到中间时刻的 CWs, 基音周期以同样的方式内插。

$$\left. \begin{aligned} u(t, \phi) &= (1 - \alpha(t))u(t_m, \phi) + \alpha(t)u(t_{m+1}, \phi) \\ p(t) &= (1 - \alpha(t))p(t_m) + \alpha(t)p(t_{m+1}) \end{aligned} \right\} \quad (4)$$

其中  $\alpha(t)$  是从  $\alpha(t_m) = 0$  到  $\alpha(t_{m+1}) = 1$  单调递增的函数。内插得到的二维特征表面通过相位轨迹由下式恢复一维残差信号:

$$e(t) = u(t, \phi(t)) = u \left( t_m, \phi(t_m) + \int_{t_m}^t \frac{2\pi}{p(t')} dt' \right) \quad (5)$$

## 2.2 基音周期估计

本文在语音域实现了一种低复杂度的双加权长时预测增益基音估计算法, 利用前向和后向基音预测信息有效防止基音加倍与减半现象, 使基音轨迹尽量平滑, 确保 CWI 语音编码算法获得可靠的特征波提取速率和尽可能高的重建音质。

每个输入语音帧的基音延时计算通过两个估计窗分别进行, 第 1 个估计窗的范围是整个当前帧信号, 第 2 个估计窗的范围是当前帧的后一半和未来帧的前一半。首先对待分析语音信号进行预处理<sup>[6]</sup>, 包括去均值, 800Hz 低通滤波(减少高频共振峰和高频噪声影响)和数值低通滤波(去除第 1 和第 2 共振峰对基音估计的影响)。对于每个预处理后的分析帧基音周期估计是基于加权长时预测增益最大准则, 加重较低基音延时处和前一帧的基音预估位置处的预测增益值<sup>[7]</sup>。加权预测增益表达式如下:

$$\beta_w = \beta(d)w_1(d)w_2(d) \quad (6)$$

$$\beta(d) = \max \left\{ 0, \min \left[ \frac{\sum_{i=0}^{L-d-1} s(i)s(i+d)}{\sqrt{\sum_{j=0}^{L-d-1} s^2(j) \sum_{k=0}^{L-d-1} s(k+d)}}, 1.0 \right] \right\}$$

$$P_{\min} \leq d \leq P_{\max}$$

其中,  $d$  是整数基音延时, 搜索范围为  $[P_{\min}, P_{\max}]$  (本文  $P_{\min} = 20$ ,  $P_{\max} = 120$ ),  $s$  是预处理后的原始语音波形段,  $L$  是分析帧长。加权函数  $w(d) = w_1(d)w_2(d)$  由两部分组成, 其中  $w_1(d)$  用来加重较低基音延时处的相关值以去除几倍于基音延时位置处的预测增益值造成的不利影响。另外一个加权函数  $w_2(d)$  用作前一帧基音预估值邻近加权, 尽可能保持相邻基音估计值的连续性。通过搜索加权基音预测增益最大值决定

分析帧最佳基音延时,然后重新根据原始语音计算得到最佳基音预测增益,作为输入语音段浊音度的度量。

最后,根据每帧两个预测窗分别得到的最佳基音延时参数( $d_0, d_1$ )和长时预测增益( $\beta_0, \beta_1$ ),依据图1所示逻辑准则得到本帧最终的最佳基音周期 $d_{opt}$ 和最佳长时预测增益 $\beta_{opt}$ 。

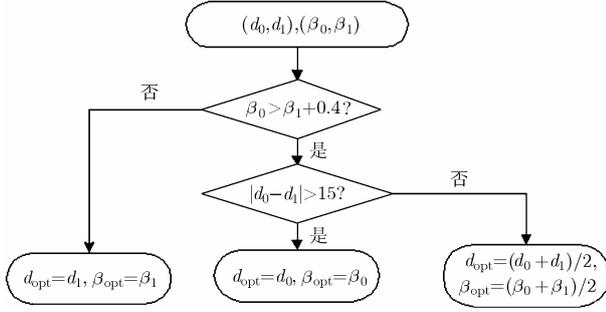


图1 基音判决逻辑准则

### 3 特征波提取速率自适应

根据Kleijin的结论<sup>[4]</sup>,理论上为了防止沿时间轴 $t$ 的混叠,至少每隔一个基音周期提取一个CW,若基音周期为 $p(t)$ ,则 $u(t, \phi)$ 沿时间轴 $t$ 的渐变带宽最多为 $1/2p(t)$ 。现有的WI编码框架的提取速率为满足最小基音周期的要求,对所有输入语音帧用固定的最高提取速率 $1/p(t)$ ,对于不同基音周期的情况,采用最高提取速率大多情况下相当于特征波过采样,而波形提取和对齐过程计算量较大,其结果是随着提取特征波个数的增加,运算更为复杂,且统一的SEW和REW降采样量化比特分配增加了平均编码比特。实际中,CWI的波形提取速率可以自适应于分析语音段的基音周期,亦即每帧的特征波提取速率根据基音周期和浊音度情况自适应地改变,亦即特征波提取速率反比于基音周期:

$$\text{if } \beta_{opt} < 0.4, R_{extr} = R_{max}; \text{ else } R_{extr} = [f_s / P_i] \quad (7)$$

其中 $R_{extr}$ 是用赫兹表示的波形提取速率, $f_s$ 是用赫兹表示的输入语音信号采样率, $P_i$ 是用样点数表示的第 $i$ 帧的基音周期。

平均提取速率降低可以大大减小CWI波形提取和波形对齐等模块的计算复杂度,在REW和SEW的量化编码方面,当CW提取速率降低时SEW和REW的抽样率也可随之降低,从而使平均编码速率有所降低,而合成语音质量不变。为了避免清音段或清浊过渡段提取速率不够,当浊音程度较小时则增加提取速率。特征波形临界抽样理论上需要用限带内插的方式进行重建,但实际上人耳对线性内插的失真并不敏感。波形分解用到了17阶非因果线性相位FIR低通滤波器,滤波时需要过去一帧和未来一帧的特征波序列,为了保持滤波阶数不变从而保证低通滤波性能,将不同提取速率下的特征波序列重采样到最高提取速率,再进行SEW和REW的降采样量化。实践证明这个过程可以简单用临近波形替代

法实现。

提取速率用每帧提取的特征波个数 $K_{extr}$ 表示,最大提取波形个数为 $K_{max}$ ,每个特征波的DTFS系数序列和功率归一化因子序列统一用 $Z$ 表示,则重采样后的序列 $Z'$ 公式如下:

$$Z'(k) = \begin{cases} Z(i), & k = \lfloor K_{max} \times i / K_{extr} \rfloor, \quad 1 \leq i \leq K_{extr} \\ Z(k+1), & k \text{ 为其它值}, \quad 1 \leq k \leq K_{max} \end{cases} \quad (8)$$

本文采用非均匀的特征波提取间隔,每20ms分析帧提取2~8个特征波,相应特征波采样率为100~400Hz,提取速率由基音周期决定,不需要传边信息。波形提取、对齐和归一化后进行重采样,使特征波及功率归一化因子的更新速率恢复至400Hz后再进行低通滤波。

考虑到低通滤波器的非理想性及人耳的感知特性,设置200Hz以上提取速率对应SEW和REW的更新速率分别为100Hz和200Hz。通过实验验证,对于200Hz以下提取速率,SEW和REW的采样率可以分别降为50Hz和100Hz而对合成音质无影响。同时判断最佳基音预测增益,如果小于门限值则采用 $R_{max}$ 特征波提取速率。解码端根据解量化的基音周期和浊音度标志得到当前帧的特征波提取速率,将解量化后的SEW和REW序列分别通过线性内插和临近替代法上采样到实际的提取速率 $R_{extr}$ ,两部分相加得到原始提取速率下的特征波序列。

此外,通过实验观察分析,语音帧的二维特征波表面存在很多平坦度较高的情况,可以用原始特征波序列作为SEW成分而无需进行波形分解。本文在特征波序列重采样后利用相关系数序列的均值和方差联合判断二维特征波表面的平坦度。每帧提取的特征波序列用DTFS系数表示为 $CW_n(k) = [A_n(k) \ B_n(k)]$ , $n$ 为每帧提取的特征波序号,相邻特征波DTFS系数的维数用截短或补零的方法调整到相同,则相邻特征波的相关系数定义为

$$c_n = \frac{\left| \sum_k CW_{n-1}(k) \times CW_n(k)^T \right|}{\sqrt{\sum_k CW_{n-1}(k) \times CW_{n-1}(k)^T \sum_k CW_n(k) \times CW_n(k)^T}} \quad (9)$$

用 $c\_mean$ 和 $c\_std$ 分别表示相关系数序列的均值和方差,通过对参数进行统计实验和合成音质的试听,设定若 $c\_mean > 0.85$ 且 $c\_std < 0.3$ ,则可认为波表面平坦而不进行波形分解,将原始特征波序列作为SEW,且SEW和功率增益因子的更新速率都降为50Hz。此种情况下,编码比特会有很大降低,而合成音质保持不变。在特征波提取速率自适应的CWI编码框架下,3种编码速率和REW/SEW/功率增益因子的更新速率对应关系如表1所示。

如图2和图3所示编码端和解码端特征波形提取速率自适应的CWI算法框图。

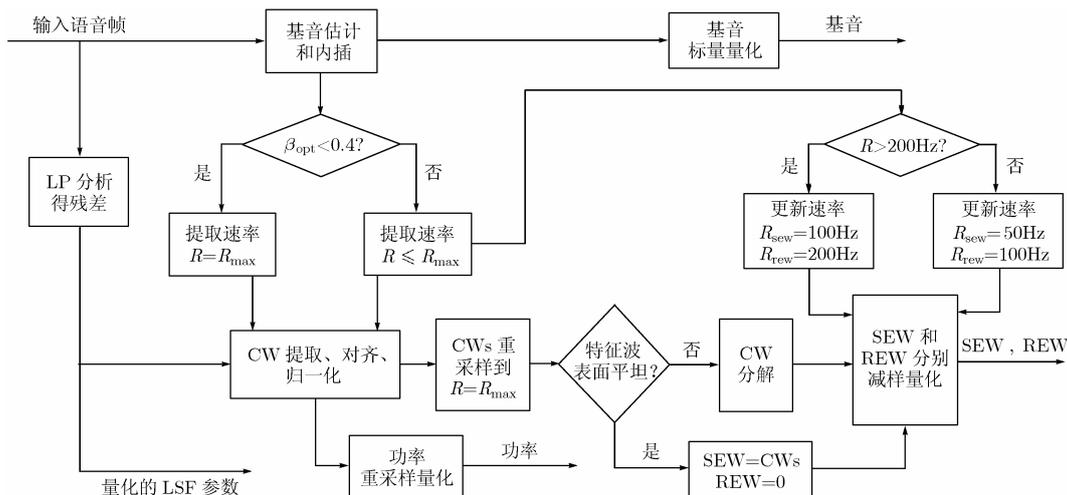


图 2 特征波提取速率自适应的 CWI 编码端算法框图

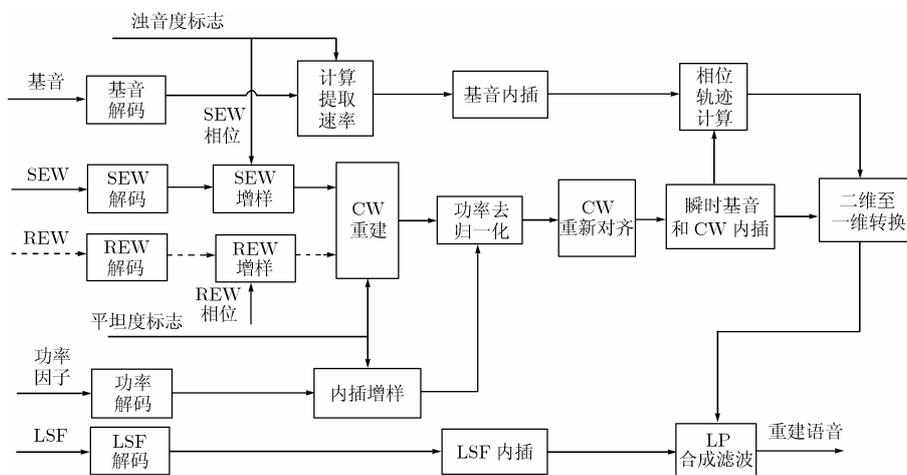


图 3 特征波提取速率自适应的 CWI 解码端算法框图

#### 4 编码参数量化

编码器需要量化的参数有 LP 系数, 基音周期、功率因子、SEW/REW 的幅度谱及浊音度和波表面平坦度标志。10 阶 LP 系数转化为线谱频率 LSF 参数用 20bit 的一阶预测分裂矢量量化<sup>[1]</sup>。CW 功率因子采用减样对数域一阶差分量化, 在解码端通过内插方法将量化的 CW 功率增样到原始更新速率。波形分解后的慢渐变波 SEW 和快渐变波 REW 分别表达了语音信号中的准周期成分和噪声成分, 它们的感性特性非常不同, SEW 幅度谱量化特点是高精度量化和低速率更新, REW 则相反。当  $R_{\text{extr}} = R_{\text{max}} = 400\text{Hz}$  时, SEW 和 REW 的更新速率分别是 100Hz 和 200Hz, 而当  $R_{\text{extr}} \leq 200\text{Hz}$  时, 更新速率则采用 50Hz 和 100Hz 的方案。根据平坦度标志, 若本帧的特征波表面平坦则  $\text{SEW} = \text{CWs}$ ,  $\text{REW} = 0$ , SEW 的采样率降为 50Hz 进行量化编码。由于每个提取时刻的基音周期不同, 每个特征波离散傅里叶级数的维数也不同, 对 REW 和 SEW 的幅度谱进行矢量量化时采用变维矢量量化方法 VDVQ<sup>[8]</sup>, 将不同维数幅度谱矢量调整到相同的 60 维进行量化。人耳对慢变波幅度谱比较敏感, 要对它进行精确的量化, 采用变维分裂矢量量化, 将 60 维幅度谱分裂成 3

个子矢量 15, 15, 30, 分别对应 0~1kHz, 1~2kHz, 2~4kHz 3 个子频带; 相位谱用固定相位代替(来自于高基音周期的男性讲话者浊音语音)。在清音或清浊过渡段, 固定相位谱可能会造成合成语音过强的周期性, 所以本文根据基音预测增益来决定 SEW 的相位谱采用随机相位或是固定相位, 当  $\beta_{\text{opt}} < 0.4$  时则采用随机相位。人耳对 REW 的幅度谱细节不敏感, 只需用少量比特对 REW 的幅度谱进行矢量量化; REW 的相位谱携带的感性信息不多, 可以用随机相位代替。

表 1 为本文的 CWI 编码器比特分配方案和 3 种编码码率设置。

#### 5 实验结果和分析

本文选取的 16 段测试语音来自于 NTT-AT 标准汉语语音库, 输入信号采样率调整为 8kHz, 每段 8s(400 帧), 4 男 4 女发音。按照 3.8kbps 的 CWI 和特征波提取自适应的 CWI 算法分别对输入语音进行编解码, 统计所有帧的平均码率, 并考察各个输入语音段的编解码仿真耗时作为计算复杂度比较, 最后通过主观 R-A/B 测试本文提出的特征波提取自适应的波形内插编码算法合成音质。

表1 编码参数比特分配

编码参数	bit/更新点	更新速率(Hz)	bit/帧	传输速率(bit/秒)
LSFs	20	50	20	1000
基音	7	50	7	350
功率	4	100/ 100/ 50	8/ 8/ 4	400/ 400/ 200
SEW 幅度	14	100/ 50/ 50	28/ 14/ 14	1400/ 700/ 700
REW 幅度	3	200/ 100/ 0	12/ 6/ 0	600/ 300/ 0
浊音度判决	1	50	1	50
平坦度判决	1	50	1	50
总比特			77/ 57/ 40	3850/ 2850/ 2350

### 5.1 平均码率和计算复杂度统计

测试语音帧长为 20ms(帧速率 50Hz), 3.8kbps的CWI算法的特征波提取速率固定为 $R_{\max}=400\text{Hz}$ , REW和SEW的更新速率固定为 200Hz和 100Hz; 特征波提取速率自适应情况下, CW的提取速率和SEW/REW的更新速率由基音周期、预测增益及波表面平坦度共同决定, 每帧的编码速率在 3.85kbps、2.85kbps和 2.35kbps间切换。在MATLAB环境下进行算法仿真, 统计 3 种码率 $r_1, r_2, r_3$ 各占的帧数 $n_1, n_2, n_3$ , 通过式 $r = \frac{n_1 \times r_1 + n_2 \times r_2 + n_3 \times r_3}{n_1 + n_2 + n_3}$ 得到平均编码

速率 $r$ (不计静默帧), 并统计各个语音段分别在两种编码情况下的算法仿真时间来衡量计算复杂度。如此对 16 段测试语音进行统计得到波形提取速率自适应方案平均码率为 3.03kbps, 相对于 3.8kbps 方案码率有 20.3%的下降, 而非正式主观测听表明合成音质没有下降。

从图 4 中的仿真耗时对比可以看出, 引入特征波提取速率自适应后 WI 算法计算复杂度有明显降低, 因波形的平均提取速率小于最大波形提取速率, 使得使 CW 的提取、表示、对齐等模块的计算量下降很多, 且在特征波表面平坦情况下无需进行波形分解过程。

从图 4 中的仿真耗时对比可以看出, 引入特征波提取速率自适应后 WI 算法计算复杂度有明显降低, 因波形的平均提取速率小于最大波形提取速率, 使得使 CW 的提取、表示、对齐等模块的计算量下降很多, 且在特征波表面平坦情况下无需进行波形分解过程。

### 5.2 主观 R-A/B 测试

选取 10 名男女各半的听音人组成的测试小组通过试听比较本文的平均码率为 3.03kbps 的 WI 编码算法(B)和 FS1016 的 4.8kbps 的 CELP 编码标准(A)的合成音质。R 是原始语句仅作为参考, A 和 B 的放音顺序随机。听音人对 A, B 的音质做出偏爱选择, 对 10 名听音人的测听结果进行统计平均, 得到主观评价结果如表 2 所示。

从表 2 可以看出, 对于女声语音、男声语音和所有测试语音, 偏爱本文提出的语音帧平均码率为 3.03kbps 的 CWI 编码算法的人数要比偏爱 4.8kbps 的 CELP 语音编码算法的人数分别高出 23.4%, 4.7%, 14.1%, 表明本文的特征波提取速率自适应的 CWI 语音合成质量女声明显优于 4.8kbps 的 CELP 算法, 而男声的重建语音质量则略胜于 4.8kbps 的 CELP。

表2 主观 R-A/B 测试结果

	偏爱 A	偏爱 B	无偏爱
女声语音	25.0%	48.4%	26.6%
男声语音	32.8%	37.5%	29.7%
所有语音	28.9%	43.0%	28.1%

### 6 结束语

本文在实现 3.8kbps 的固定提取速率特征波形内插编码算法基础上, 提出了特征波形提取速率自适应的实现方案, 并设计了有效的基音估计算法联合浊音度决定每帧提取的特征波个数, 平均提取速率远低于最大提取速率, 通过相关系数的均值和方差判断特征波平面平坦度, 可使 SEW 和 REW 的更新速率降为 50Hz 和 0Hz。分别设置 3 种 SEW/REW 的更新速率, 整个算法的平均编码比特和计算复杂度均有不同程度的下降。主观测听表明, 本文提出的语音帧平均码率为 3.03kbps 的 WI 编码算法合成音质明显优于 4.8kbps 的 CELP 算法。此外, 可以利用时间弯折(time-warping)方法将时变基频信号转化为恒定基频信号从而实现临界基音周期波形提取, 获得自然的特征波形提取速率, 在 SEW 和 REW 的参数的量化方案上还可以做进一步改善。

### 参考文献

- [1] 鲍长春. 低比特率数字语音编码基础[M]. 北京:北京工业大学出版社, 2001: 233-234.
- [2] Kleijin W B. Continuous representation in linear predictive coding[A]. IEEE ICASSP'91[C]. Toronto, 1991, 201-204.
- [3] Kleijin W B and Haagen J. Transformation and decomposition of the speech signal for coding[J]. *IEEE Signal Processing Letters*, 1994, 1(9): 136-139.
- [4] Kleijin W B. A speech coder based on decomposition of characteristic waveforms[A]. IEEE ICASSP'95[C]. Detroit, 1995: 508-511.
- [5] Kleijing W B, Shoham Y, and Sen D, *et al.* A low-complexity waveform interpolation coder[A]. IEEE ICASSP'96[C]. Atlanta, 1996, 212-215.
- [6] 鲍长春, 樊昌信. 基于归一化互相关函数的基音检测算法[J]. 通信学报, 1998, 19(10): 27-31.
- [7] 3GPP TS 26.190. AMR Wideband speech codec; Transcoding functions[S], 2001, Release 5.
- [8] Das A, Rao A V, and Gersho A. Variable-dimension vector quantization[J]. *IEEE Signal Processing Letters*, 1996, 3(7): 200-202.

王晶: 女, 1980年生, 博士生, 目前主要从事语音、音频编码的研究。

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。

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。