基于 DCT 分带谱熵与信号分解的高精度基音检测算法

罗亚飞,鲍长春
(北京工业大学电子信息与控制工程学院,北京 100022)

摘 要: 本文就低速率 WI 语音编码中的基音检测技术进行研究,针对基音检测在不同噪声与信噪比下容易发生清浊误判的问题,在基音检测前端引入基于 DCT 分带谱熵的语音检测算法划分语音段与非语音段;为了向基音检测算法提供更能准确反映基音周期实际变化的输入语音,基于谐波-噪声模型提出了一种改进的 DCT 域语音分解算法. 然后,根据变形的 MCAMDF(Modified Circular Average Magnitude Difference Function)与 NCCF(Normalized Cross-Correlation Function)的峰值共性,结合上述两项基音检测前端处理技术,提出了 MCAMDF-NCCF 基音检测组合算法. 为了满足不同环境下 WI 编码器对基音检测高精度的要求,在合成端更准确地恢复相位轨迹,本文又基于 MCAMDF-NCCF 算法提出了高精度 MCAMDF-NCCF-FRAC 基音检测算法以计算分数基音. 将算法应用与 2kb/s WI 编码器,主观 A/B 听力测试结果表明,本文提出的基音检测算法在低信噪比下明显抑制了基音加倍减半及清浊误判现象的发生,得到了优异的基音检测结果,合成语音质量完全满足低速率 WI 编码器对基音检测技术的要求.

关键词: 语音编码; 基音检测; 语音检测; 信号分解; 波形内插
中图分类号: TN912.3 **文献标识码**: A **文章编号**: 0372-2112 (2007) 01-0013-10

Super Resolution Pitch Detection Based on Band-Partitioning Spectral Entropy and Signal Decomposition in DCT Domain

LUO Ya-fei, BAO Chang-chun
(School of Electronic Information and Control Engineering Beijing University of Technology, Beijing 100022, China)

Abstract: In this paper, the research focuses on pitch detection techniques of the low-rate WI speech coding. As the pitch doubling and halving problems of pitch detection often occurred with varied noises and Signal to Noise Ratio (SNR), voice activity detection (VAD) algorithm based on DCT band-partitioning spectral entropy is employed in pre-processing to separate speech and non-speech segments. In order to provide an accurate-pitch-cycle speech for pith detection algorithm, an improved speech decomposition algorithm in DCT domain based on the Harmonic-Noise Model is presented. Then, using the same characteristic of maximum peaks of MCAMDF and NCCF and two pro-processing techniques mentioned above, a pitch detection algorithm in a combination both of two functions together named MCAMDF-NCCF is proposed. In order to satisfy the needs of the pitch accuracy of WI coder and synthesize phase track correctly, a super resolution pitch detection algorithm named MCAMDF-NCCF-FRAC based on MCAMDF-NCCF is also given to get fractional pitch. We applied these algorithms to WI coder, the results from the subjective A/B listening test indicated that both of these two algorithms have a great performance and heavily reduce pitch doubling and halving and voiced-unvoiced error in low SNR, the quality of the synthesized speech satisfies the accuracy of the pitch detection techniques of WI coder completely.

Key words: speech coding; pitch detection; voice activity detection; signal decomposition; waveform interpolation

1 引言

基音周期是发浊音时声带振动频率的倒数,它是语音时域特征里最重要的参数. 基音检测(Pitch Detection)是语音处理中非常重要的问题. 汉语是一种有调语音,基音的变换模式称为声调,它携带着重要的具有辨意作用的信息. 在低速率语音编码中,准确检测语音信号的基音周期非常关键,它直接影响到整个声码器系统的性能[1].

近年来人们已从时域、频域和时频混合域出发,针对不同情况提出了多种有效的基音检测算法. 在时域算法中,自相关函数算法通过比较原始信号与移位后信号间的类似性求取基音周期[2]. 由于基音较大时乘累加项数逐渐减少会造成自相关函数峰值降低,该算法在基音较大时检测不利. 短时平均幅度差函数算法通过寻找最小谷值点求取基音周期,但同样存

收稿日期:2006-05-08;修回日期:2006-08-21
基金项目:国家自然科学基金(No. 60372063);北京市自然科学基金(No. 4042009);北京市教委科技发展计划(No. KM200710005001)

在基音较大时检测不利的问题[3].在频域算法中,普通频域算法通过比较原始语音与重建语音的最小均方误差选择基频,提高了基音检测准确性,但同时也增加了算法复杂度[4].当语音信号的共振峰能量较高且出现位置和基频比较接近时,靠近共振峰的谐波提供了重建语音的主要能量,此时频域基音检测算法有可能将第一共振峰频率错判为基音频率.倒谱法[5]对纯净语音的基音提取精度较高,但算法比较复杂,受加性噪声影响较大.在时频混合域算法中,基于 MBE 模型的基音检测算法[6]首先在时域进行基音粗估计,然后在频域进行基音细搜索,精度很高但计算量较大.小波变换的基音检测算法[7]通过声门闭合瞬间语音信号相邻突变点的间隔确定基音周期,但计算量较大.

上述均为基音检测的传统算法,每种算法都在一定环境下有其优势.有的原理简单,计算量小;有的精度较高但计算复杂.针对以上问题,又有新的算法不断被提出,如变长的 AMDF 算法(Length-Varied AMDF,LVAMDF)[8]、循环的平均幅度差算法(CAMDF)[9,10]、利用平均幅度差函数倒数作用与自相关函数以突出基音周期处峰值的算法[11]等.

由于通常人们得到的语音都不同程度地受到各种噪声的污染,因此,基音检测技术的研究热点和难点已集中于处理低信噪比语音.比如,新近出现的利用时域信息的 APP 算法[12]、CAMDF 与频域算法的组合算法[13]、基于噪声白化过程的 LVAMDF 算法[14]、基于最大似然函数与谐波模型的算法[15]等都在低信噪比下得到了理想的基音检测结果.

总体而言,时域算法简单,计算量小,硬件易实现.由于基频离第一共振峰频率很近,噪声环境下会出现基音加倍减半.频域算法相对精确,代价是高的计算复杂度.同时,频域算法也同样易受共振峰影响,发生基音误判.时频混合域算法往往在时域进行基音初估,再在频域进行基音细搜索,纯净语音下通常具有很好的性能.算法精度与计算复杂度是一对矛盾,为了在两者之间寻求平衡,人们引入预处理、后处理措施以得到平滑渐变的基音轨迹.由于大多数基音检测算法针对纯净语音提出,信噪比较低时算法性能均有明显下降.信噪比越低,基音误判越严重,以致后处理措施也无能为力.

基于归一化互相关函数(NCCF)的基音检测算法[16]计算简单,高信噪比时性能出色,一直被本课题 WI 编码器所采用.为了得到噪声环境下高性能的基音检测算法,本文针对 NCCF 算法在不同噪声、信噪比下容易发生清浊误判的问题,提出了基于 DCT 分带谱熵的语音检测算法划分语音段与非语音段.为了向基音检测算法提供更能准确反映基频的输入语音,基于谐波-噪声模型提出了一种改进的 DCT 域语音分解算法.然后,根据变形的 MCAMDF 与 NCCF 的峰值共性,结合上述两项基音检测前端处理技术,提出了 MCAMDF-NCCF 基音检测组合算法.为了满足不同环境下低速率 WI 编码器对基音检测高精度的要求,在合成端更准确地恢复相位轨迹,本文又基于 MCAMDF-NCCF 算法提出了高精度 MCAMDF-NCCF-FRAC 基音检测算法.将算法应用于 2kb/s WI 编码器,主观 A/B 听力测试结果表明,本文提出的基音检测算法在低信噪比下明显抑制了基音加倍减半及清浊误判现象的发生,得到了优异的基音检测结果,合成语音质量完全满足低速率 WI 编码器对基音检测的要求.

本文在第 2 节介绍基于 DCT 分带谱熵的语音检测算法,第 3 节基于谐波-噪声模型介绍改进的 DCT 域语音分解算法,第 4 节结合第 2、3 节的两项基音检测前端处理技术,介绍 MCAMDF-NCCF 基音检测组合算法及其性能,第 5 节介绍高精度 MCAMDF-NCCF-FRAC 基音检测算法,第 6 节给出本文建议的基音检测算法应用于低速率 WI 编码器的性能评价,最后是本文的结论.

2 基于 DCT 分带谱熵的语音检测算法

众所周知,若浊音被误判,会丢失重要的语音信息;若清音被误判,会增加基音检测算法的负载,影响合成语音质量.为了在不同噪声、信噪比下最大程度消除清浊误判,明确区分语音段与非语音段,本文将从语音检测角度研究该问题.

2.1 语音检测概况

准确进行语音检测在语音信号处理领域一直具有重要意义,在语音识别前端可提高识别的准确率;用于语音增强系统可进行准确的噪声模型估计;在语音编码领域可降低编码速率及编码器负载.语音检测可分为完全的显性、完全的隐性、混合的语音检测三类[17].其中,显性语音检测比较适合基音检测前端处理,本文主要对此类语音检测进行研究.

语音检测技术主要依靠选取更有效的特征以区分语音段与非语音段.基于时域能量与过零率特征[18]的算法在处理纯净语音时效果明显,但在噪声环境下却无法达到要求.在时频特征[19]、高阶统计量特征[20]、基于帧的 Teager 能量特征[21]、谱熵特征[22~24]中,由于语音与噪声在频谱上有很大区别,用频谱的熵在噪声环境下进行语音检测有其独特的优势.下面对基于谱熵特征的语音检测算法进行介绍,并提出一种 DCT 分带谱熵语音检测算法.

2.2 基于谱熵的语音检测技术

熵的概念在语音编码领域已经被广泛使用.从熵的角度来看,语音信号具有较小的熵,噪音信号具有较大的熵.

Jialin Shen 于 1998 年首先提出运用熵进行语音检测[22].该算法先分析语音信号各频率成分的能量以求取概率密度函数,然后根据求得的各频率点概率密度估计谱熵.

为使不同噪声的谱熵轨迹比较接近而且平坦,同时语音谱熵与噪音谱熵又能有效被区分,Chuan Jia 于 2002 年提出一种改进的谱熵语音检测算法[24].该算法将输入语音的频率限制在一定范围内,引入正数 K 修正概率密度函数并得到了很好的效果.

针对语音幅度谱易受噪声谱污染,并导致语音检测算法质量下降的问题,Bing-Fei Wu 在 2005 年提出,对带噪语音的多带分析能够抑制噪声环境下语音幅度谱易受噪声污染的问题[23].该算法引入加权窗、自适应谱熵门限,采用归一化最小子带能量确定有效子带数求取谱熵,进行语音检测.

2.3 基于 DCT 分带谱熵的语音检测算法

上述基于谱熵的语音检测算法需对语音信号做 FFT 变换以反映时域信号相关性,而 FFT 变换在去除信号相关性上不

是最佳变换. KL 变换是最佳正交变换但是没有快速算法. DCT 变换的性能仅次于 KL 变换,并且以 2 为基底点数的 DCT 拥有快速算法,所有的数据变换全在实数域进行. 本文给出一种基于 DCT 分带谱熵的语音检测算法,算法如下:

(1) 对每帧语音信号计算 M 点 DCT 系数 s_dct_i, $i = 0, 1, \cdots, M-1$, $M = 512$.

(2) 将每帧语音信号分为 N_b 个子带,每个子带点数 $K = M/N_b$. 这里 $N_b = 32$, $K = 16$. 因此每一 DCT 子带的频率权重 sub_dct_j 为:

$$sub\_dct_j = \sum_{i=j*K}^{j*K+K-1} |s\_dct_i|, \quad j = 0, 1, \cdots, N_b - 1 \qquad (1)$$

(3) 计算归一化子带频率权重 p_dct_i:

$$p\_dct_i = \frac{sub\_dct_i}{\sum_{i=0}^{N_b-1} sub\_dct_i} \qquad (2)$$

由于语音信号的主要能量集中在 250～3500Hz,因此不属于该频率范围的 DCT 子带频率权重 sub_dct_j 置 0. 另外,为了消除一些集中在特殊频率的噪声,限定 $p\_dct_i < 0.9$,即采用如下约束关系:

$$sub\_dct_j = 0, \begin{cases} f < 250\text{Hz} \text{ 或 } f > 3500\text{Hz} \\ p\_dct_i < 0.9 \end{cases} \qquad (3)$$

(4) 由于归一化子带频率权重 p_dct_i 反映了 DCT 系数在频率点 i 处的出现概率,语音的 DCT 谱熵为:

$$H = -\sum_{i=0}^{N_b-1} p\_dct_i \log p\_dct_i \qquad (4)$$

因为负谱熵的变化轨迹和短时能量轨迹相似,便于观察分析,本算法使用 DCT 负谱熵进行语音检测.

2.4 实验分析

本文挑选内容为"沉舟侧畔千帆过"的语音,与 NOISEX-92 噪音语音数据库中 12 种常见噪声串接在一起处理,分别为 clap(鼓掌声)、siren(警笛声)、explode(爆炸声)、bell、horn(号角声)、f16、white、volve(汽车噪声)、pink(粉红噪声)、gun 噪声、babble(鸡尾酒会上人群的谈话声)、engine(机械声). 所有语音均用 Cool Edit Pro2.0 处理为 8kHz 采样,16 比特量化的单声道语音,实验结果如图 1. 分别基于短时能量(图 1(*c*))、短时过零率(图 1(*d*))、原始 DCT 谱熵(图 1(*e*))、32 分带并限制语音范围在 1000-3400Hz DCT 谱熵(图 1(*f*))、32 分带并限制语音范围在 250-3500Hz DCT 谱熵(图 1(*g*))的语音检测算法处理. 图 1(*b*)为输入语音语谱图.

从图 1(*c*)看出,由于噪声能量很高,简单的基于短时能量的语音检测算法已经失效,语音的能量轨迹完全被噪声淹没. 因为日常生活中噪声大小变化无常,基于能量特征的算法不可取. 从图 1(*d*)看出,不同噪声的过零率变化很大,基于过零率的特征对噪声和语音没有很好的区分度,因此不适合与其他特征结合使用. 从 volvo 噪音的语谱图可以看出其主要能量集中在 200Hz 以内,因此与图(*e*)相比,图 1(*f*)、(*g*)基于频率限制的 DCT 谱熵检测算法显出优势. 由于 gun 噪声的主要能量集中在 600Hz 以内,1000-3400Hz 频率限制算法起到很好的效果,但其在处理 babble 噪声和 engine 噪声时几乎失效,原因是这两种噪声的能量在 500Hz 以上仍然比较强,尤其是 engine 噪声,在 1500Hz、2000Hz、2300Hz 能量都很强. horn 噪声与 siren 噪声由于具有很强的周期性,谱熵区分度不高.

但总体而言,基于频带限制的 DCT 分带谱熵语音检测算法在大多数噪声背景下,能够得到比较平坦的噪声谱熵轨迹,语音与噪声有较好的区分度. 由于基音检测前端处理是为了区分语音段与非语音段,因此,基于 DCT 分带谱熵的语音检测算法结合门限控制可以得到很好的效果.

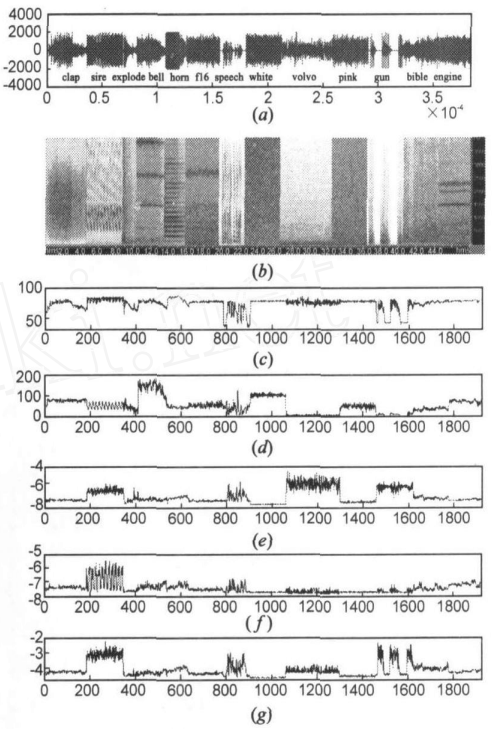

图 1　噪声串界语音的语音检测特征分析

本算法首先计算前 N 帧噪声语音的谱熵 $dct\_entropy_i$, $i = 0, 1, \cdots, N-1$, $N = 10$ 或 20. 然后计算前 N 帧噪音 dct_en-

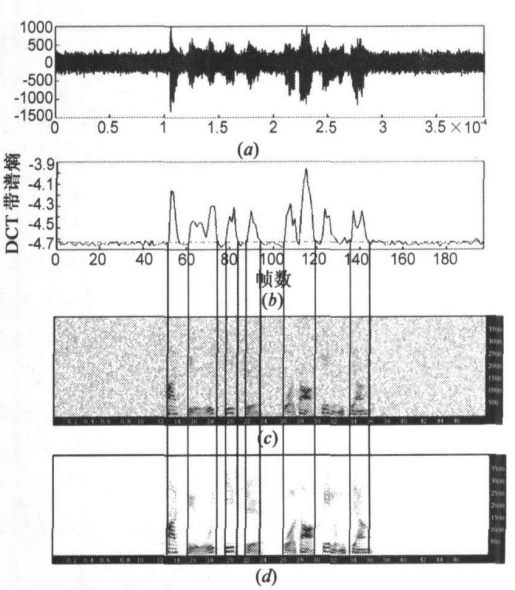

图 2　*SNR*=0dB 的语音检测结果及与纯净语音的语谱边界比较

$tropy_i$ 的均值 μ 与标准差 σ,控制门限 THR 取为:

$$THR = \mu + c \cdot \sigma \tag{5}$$

不同信噪比下的大量实验结果表明,当经验因子 $c = 1.5$ 时实验结果最理想. 其中,不同信噪比语音用 Matlab 计算生成. 图 2 为 $SNR = 0$dB 的语音检测结果. 其中,图 2(a) 为加噪语音(白噪声),图 2(b) 为采用本文算法检测语音段的情况,点划线为控制门限 THR. 图 2(c)、(d) 分别为 $SNR = 0$dB 的语音与纯净语音的语谱图. 将图 2(b) 的检测结果与图 2(c)、(d) 的语谱边界比较可得,本文算法在低信噪比条件下能有效区分语音段与非语音段,完全可以用于基音检测前端处理.

3 改进的 DCT 域语音分解算法

由于正确检测浊音段基音,保证基音轨迹的平滑渐变非常重要,因此,为基音检测核心算法提供更能准确反映基音周期实际变化的输入语音,将对合成语音质量产生重要影响.

本节将从信号分解角度出发,基于谐波-噪声模型提出一种改进的 DCT 域语音分解算法.

3.1 基于"谐波-噪声"模型的语音分解技术概述

在信号处理领域,人们提出了多种信号分解算法,如小波分解、奇异值分解、DCT 分解等.

为了尽量恢复低信噪比语音的周期性特征,基于谐波-噪声模型,Yegnanarayana 于 1995 年最先提出一种"周期-非周期成分"语音分解算法(简称 YAD 算法)[25]. 由于每个频率点上周期-非周期成分均有贡献,该算法将线性预测残差信号看作语音信号激励源的近似,在频域用"谐波-噪声比"(harmonic-Noise Ratio,HNR)粗略估计周期与非周期成分的频率点,并将确定的一系列谐波与噪声频率区域记为 F_p 和 F_r. 假设原始语音所有频率点的子集 F_r 为非周期成分的初始估计,则主要的问题便是在粗估计的周期成分位置生成原始浊音段的非周期成分. 为了达到此目的,首先将周期成分区域置 0,非周期成分为实际频率点值,然后用一种"频域-时域"的迭代算法从周期成分中估计非周期成分,最后用残差信号减去最后一次迭代得到的非周期成分得到周期成分. 详细的算法原理推倒见原文献[26].

当然,相似地也可以从噪声区重建周期性成分或既从谐波区重建非周期成分又从噪声区重建周期成分. 实验结果表明,这几种处理算法所得到的结果差异很小,因此 Yegnanarayana 建议使用谐波区下重建非周期成分的算法[26].

Ahn 和 Holmes 对 YAD 算法做改进,得到了更好的效果[27]. 之后,Nazih Abu-Shikhah 基于 YAD 算法,于 2000 年提出一种基于 DCT 域自适应门限的语音信号分解算法(简称 DCT-HN 算法)[28],得到了比上述两种算法更好的结果. 主要改进如下:

(1) 在 DCT 域分解语音,避免了在 DFT 域中复杂的复数运算. 基于 2.3 节的讨论,DCT 性能仅次于 KL 变换,并且具有快速算法,可以比 DFT 最大限度的去除信号相关性.

(2) 每帧用自适应门限估计非周期成分. YAD 算法将全部帧的非周期水平用固定门限判别不严格.

(3) 该算法既可用于语音信号,也可用于残差信号. 而 YAD 算法是在残差域给出的.

DCT-NN 与 YAD 算法原理相似,实验结果表明,二者在处理 SNR 大于 10dB 的语音时效果均很好,而在 SNR 小于 5dB 以下时 DCT-HN 算法更好.

由于 DCT-HN 算法迭代烦琐,见原文献[28],本文给出了一种改进的 *DCT-HN* 语音分解算法,在满足基音检测要求的同时降低了计算复杂度,并在一定程度上提高了输入语音的信噪比.

3.2 改进的 DCT 域语音分解算法

由于 DCT-HN 算法有两级迭代:Max iteration 处需计算 Max iteration 对 DCT-IDCT 算子,MSE < 0.001 处需根据迭代出的时域非周期成分计算对应周期成分的 DCT 系数,以求得新的自适应门限,因此完成两级迭代共需 2 · Max iteration + 1 次 DCT 运算. 若直接用于基音检测前端处理,计算复杂度是不容忽视的问题.

针对此问题,本文给出一种改进的 DCT-HN 语音分解算法,以求在不同信噪比下为基音检测核心算法提供更能准确反映基频的输入语音. 算法如下:

(1) 对每帧语音 $s(n)$ 补零做 M 点 DCT,记为 $s\_dct(k)$,帧长 $N = 200$,$M = 512$.

(2) 按下式计算该帧门限 Thr:

$$y(k) = \log|s\_dct(k)|, \quad k = 0,1,\cdots,M-1 \tag{6}$$

$$Thr = \mu_y + cc \cdot \sigma_y \tag{7}$$

cc 为经验因子,μ_y、σ_y 为 $y(k)$ 的均值、标准差.

(3) 依门限 Thr 确定初始的 DCT 域周期成分 $h^{(0)}(k)$ 与非周期成分 $n^{(0)}(k)$,$k = 0,1,\cdots,M-1$.

(4) 在 $s\_dct(k)$ 中将周期成分 $h^{(0)}(k)$ 位置的 DCT 系数置 0 后,计算 IDCT 求得非周期成分的估计 $n(m)$,$m = 0,1,\cdots,M-1$.

(5) 令估计的非周期成分 $n(m) = 0$,$m = N, N+1, \cdots, M-1$(时域限制)并从新计算其 DCT.

(6) 将初始非周期成分 $n^{(0)}(k)$ 中不为 0 的频率点代替步骤(5)所得到的非周期成分 DCT 系数(频率限制),再做 IDCT 求得本次非周期成分估计.

(7) 依步骤(5)、(6)进行迭代,计算上一次与本次估计的非周期成分 DCT 系数的均方误差(MSE). 如果 MSE 小于预设门限,则从原始语音 $s(n)$ 中减去本次估计的非周期性成分 $n(n)$,得到周期性成分 $h(n)$.

注意:由于输入语音的大小影响 MSE 门限的设定,可用迭代次数作为终止算法的手段. 图 3 为改进的 DCT-HN 语音分解算法流程图.

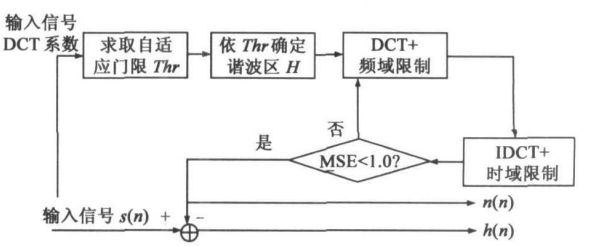

图 3 改进的 DCT-HN 语音分解算法流程图

该算法中 DCT 门限 *Thr* 经验因子 *cc* 的选取对分解结果影响很大. DCT-HN 算法建议 *cc* = 0.25. 由于本文基于汉语语音,经大量实验分析, *cc* = 0.25 时不能得到理想的分解效果.

图 4(*a*)为任取的一帧纯净浊音语音. 将纯净语音加高斯白噪声生成 *SNR* = 0dB 语音,经改进的 DCT-HN 算法分解,得到图 4(*b*)、(*c*)、(*d*)、(*e*)分别为 *cc* = 0.25、0.5、0.75、1.0 时分解出的周期成分.

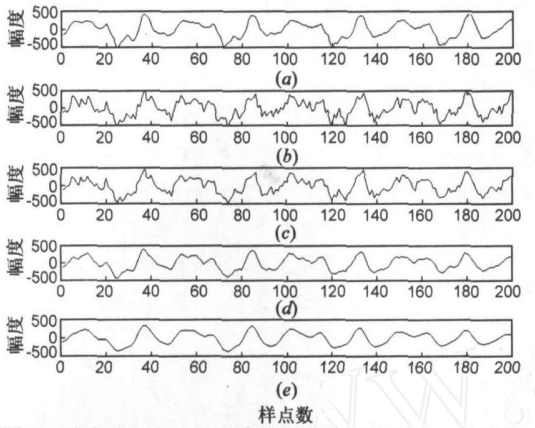

图 4　改进的 DCT-HN 语音分解算法经验因子 *cc* 的确定

可以看出, *cc* = 0.25、0.5 时分解效果不理想, *cc* = 1.0 周期性过强. 通过观察更多帧周期成分的分解结果发现, *cc* = 1.0 时过周期现象很普遍,甚至在非语音段也产生了周期性结构. 若将此类语音用于基音检测,会在非语音段对 2.3 节建议的谱熵判决产生影响. 只有 *cc* = 0.75 时产生的周期成分与纯净语音最接近. 结合数百帧的分析,本算法取 *cc* = 0.75.

3.3　实验分析

为了比较 DCT-HN 算法改进前后的计算复杂度,选取内容为"大家"*SNR* = 0dB 的语音,共 40 帧. 将两种算法的最大迭代次数定为 10 次,DCT-HN 算法的另一级循环强制最多循环 5 次,求取两种算法分解出的周期成分与纯净原始语音间的 MSE. 表 1 记录了不同的 MSE 阶段,两种算法的总迭代次数 ITER 与计算时间 TIME. 可以看出,改进的 DCT-HN 算法在计算速度上比 DCT-HN 算法有明显优势.

表 1　DCT-HN 算法改进前后的计算复杂度比较

	改进的 DCT-HN 算法			DCT-HN 算法		
MSE	15000	10000	5000	15000	10000	5000
ITER(次)	72	195	349	620	880	1260
TIME(秒)	14.8	30.2	49.9	87.4	125.5	177.5

为了比较 DCT-HN 算法改进前后的分解准确度,仍然使用内容为"大家"的纯净语音加高斯白噪声生成 *SNR* = 10dB、5dB、0dB、-5dB 的带噪语音. 两种算法的迭代次数定为 10 次, DCT-HN 算法的另一级循环强制循环 5 次. 图 5 显示了各种信噪比下,两种算法分解出的周期成分与纯净语音之间平均每帧 MSE 的变化趋势. 可以看出,DCT-HN 算法改进前后在各种信噪比下准确率相当,随着信噪比的降低,改进的 DCT-HN 算法的准确率还稍占优势.

为了进行有效的基音检测前端处理,语音分解算法的复杂度要尽量小. 由于作为前端处理的语音分解算法不要求将周期、非周期成分最佳的分解出来,因此,多次迭代没有必要也不可取.

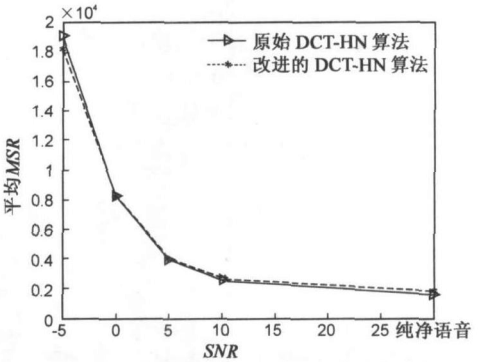

图 5　DCT-HN 算法改进前后的准确率比较

图 6 给出了 *SNR* = 0dB 的一帧浊音语音在迭代次数分别为 0、1、2、5 时分解出的非周期成分的频谱对比. 其中图 6(*a*)为 0 次迭代时的非周期频谱,频谱是 0 的部分为原始周期成分频谱位置. 将图 6(*b*)与图 6(*a*)比较可以看出,随着迭代次数的增加,原始周期成分区域中的非周期成分逐渐显现出来,并且,两次以内的迭代已经产生了很好的分解效果.

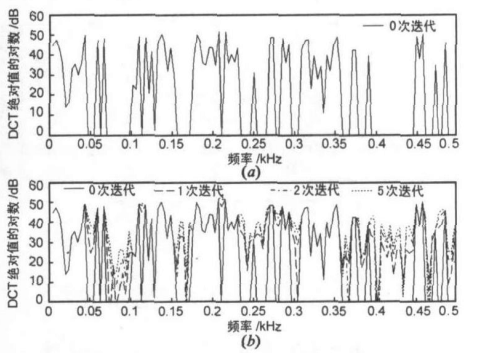

图 6　迭代算法在 DCT 域重建非周期频谱成分

图 7 给出了图 6 中各迭代次数下分解出的周期成分比较. 图 7(*a*)为 *SNR* = 0dB 的输入语音,图 7(*b*)～(*e*)为各次迭代分解出的周期成分. 可以看出,由于改进算法设定的门限可

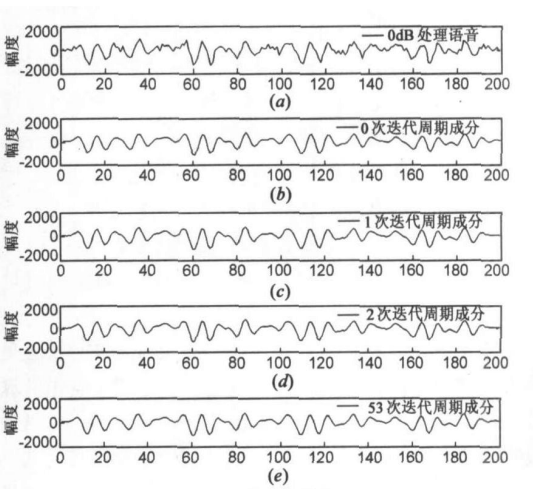

图 7　迭代算法在 DCT 域重建周期成分

靠有效,使得该算法收敛速度很快,两次以内的迭代就能达到很好的效果. 考虑到基音检测的计算复杂度,本文控制迭代次数小于 2.

图 8 为改进的 DCT-HN 算法对纯净语音的分解示例. 图 8 (*a*)、(*b*)、(*c*)分别为原始语音、分解出的周期成分、非周期成分. 其中,周期成分与原始语音的周期结构近似,能量较大(语谱颜色较深);非周期成分能量较小(语谱颜色较浅),几乎没有周期性.

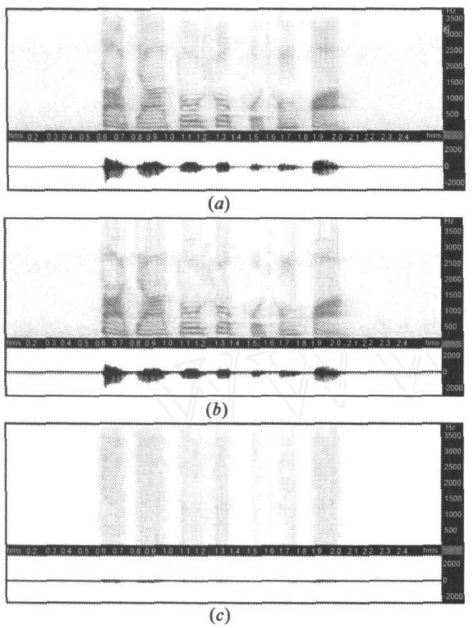

图 8 改进的 DCT-HN 算法进行周期-非周期成分分解示例

4 MCAMDF-NCCF 基音检测组合算法

NCCF 基音检测算法性能出色,纯净语音环境下完全满足低速率语音编码的要求[16] (NCCF 算法是本文基础,请读者查阅文献原文,不在赘述). 但在低信噪比下,NCCF 算法的去均值、800Hz 低通滤波、数值滤波的预处理性能明显下降,固定的判决门限无法发挥有效作用,以致出现明显的清浊误判与基音检测错误,严重影响合成语音质量.

第 2、3 节提出的基于 DCT 分带谱熵的语音检测算法、改进的 DCT-HN 语音分解算法在相当程度上解决了低信噪比下语音的清浊误判与预处理问题,能够为基音检测核心算法提供准确反映基频信息的输入语音. 对于基音检测的核心算法,考虑到计算复杂度,本文基于时域算法进行研究.

传统的 AMDF 算法随着基音延迟的增大,峰值幅度逐渐下降,这使谷值点的检测变得困难. LVAMDF 算法通过可变长度,试图跟踪帧内基音周期波形的变化,并采用归一化处理,但其基音估计值有时与真实值偏离较大. CAMDF 算法采用循环方式使用帧内样点,求和项数均相同,克服了 AMDF 算法因求和项数减少所造成的函数幅度逐渐下降的缺点. CAMDF 函数值围绕均值线水平波动,而且峰值基本保持不变. 由于 CAMDF 函数特性上的优势,用于基音检测明显减少了基音加倍减半误判. 但是,CAMDF 与 LVAMDF 算法一样,基音检测范围只有帧长的一半,若想检测正常范围内的基音需增大帧. MCAMDF 定义为:

$$\Psi(\tau)=\sum_{n=0}^{\beta}|s(\mathrm{mod}(n+\tau,N+\tau_{\max}))-s(n)|,\quad \tau=0,1,\cdots,\beta \tag{8}$$

该算法将计算序列增加到 $N+\tau_{\max}$项,其中 $\tau_{\max}$为限定的最大基音延迟,N 为帧长. 由于 MCAMDF 保持了 CAMDF 的对称特性,易知 MCAMDF 关于 $\tau_s=(\beta+1)/2$ 对称. 由于 $N>\tau_s>N/2$ 且 $\tau_s>\tau_{\max}$,MCAMDF 克服了 CAMDF 不能检测 $N/2$ 以上基音周期的缺点.

观察到 MCAMDF 定义式用到了本帧以外的数据,与 NCCF 算法的处理手段近似,考虑到 NCCF 在基音周期整数倍位置出现峰值,同时 MCAMDF 出现谷值,本文基于 C. Shahnaz 对 MCAMDF 定义式所做的变形[13],将 MCAMDF 的谷值变为峰值,与 NCCF 的峰值做乘积生成基音检测核函数(MCAMDF-NCCF). 变形的 MCAMDF 定义为:

$$\chi(\tau)=\frac{\zeta_{\max}\cdot N}{N-\Omega_{\max}}-\Psi(\tau),\ \tau=0,1,\cdots,\tau_s \tag{9}$$

$\zeta_{\max}$为 $0<\tau\le\tau_s$ 时 $\Psi(\tau)$的最大值,$\Omega_{\max}$为 $\zeta_{\max}$的下标,$\Psi(\tau)$的定义见公式(8).

为了使用 DCT 分带谱熵语音检测算法,需从前 10-20 帧噪声中提取谱熵参数(即 0.25～0.5 秒)及噪声平均能量,这在通常环境下是合理可行的. 若信噪比较高,考虑到计算复杂度,直接进行 DCT 分带谱熵分析,然后用 MCAMDF-NCCF 模块进行基音检测;若信噪比较低,则先对带噪语音进行改进的 DCT-HN 语音分解,增强信噪比及信号周期性,迭代次数设为 1 次. 之后,将得到的周期成分送入 MCAMDF-NCCF 模块完成基音检测.

根据 Rabiner 的文章[29],浊音与清音的基音检测性能指标定义如下:

(1) 总错误帧数. 定义为浊音帧内,比真实基音周期偏差大于等于 1ms 的基音个数. 该错误主要由于基音加倍减半和对共振峰抑制不够造成. 总错误率定义为浊音帧内,总错误数帧与总帧数的比值.

(2) 估计较准确的基音误差. 定义为在浊音帧内,比真实基音周期偏差小于 1ms 的基音与真实基音差的均值 μ 与标准差 σ. 普遍认为此时估计的基音周期是准确的.

(3) 清浊判决错误率. 定义为清浊误判总帧数与处理总帧数的比值.

本文选取内容为"大家都说普通话,奥林匹克运动会,他去无锡市,我到黑龙江"的实验句子,共 559 帧,其中浊音 142 帧,分别加入高斯白噪声生成 SNR = 20dB、10dB、5dB、0dB、-5dB语音. 对 NCCF 算法、离散小波变换与 NCCF 的组合算法(DWT-NCCF)[30]、本文建议的 MCAMDF-NCCF 组合算法在不同信噪比下给出如下性能比对结果.

从表 2 可以看出,三种算法在检测浊音基音时效果均很好,而 MCAMDF-NCCF 算法在低信噪比环境下($SNR\le$0dB 时)检测更加准确.

表 3 给出了不同信噪比下浊音段真实基音与估计较准确

基音的差的均值与标准差,单位为样点.可以看出,三种基音检测算法在检测浊音基音时偏差很小,说明检测准确度都相当高.

表 4 给出了三种基音检测算法在不同信噪比下的清浊误判数,可以看出,MCAMDF-NCCF 算法在低信噪比环境下(*SNR* ≤0dB 时)优势明显,大大消除了清浊误判现象.

表 2　不同信噪比下的浊音段总错误帧数

算法 \ *SNR*(dB)	∞	20	10	5	0	-5
NCCF	0	1	2	6	10	21
DWT-NCCF	1	1	2	5	10	18
MCAMDF-NCCF	0	1	2	5	8	12

表 3　不同信噪比下浊音段真实基音与估计较准确基音的差的均值 μ 与标准差 σ(单位:样点)

算法 \ *SNR*(dB)	∞	20	10	5	0	-5
NCCF	0	1.4	1.2	1.24	1.35	1.13
	0	0.89	0.52	1.12	1.04	0.39
DWT-NCCF	0	1.86	1.19	1.39	1.34	1.15
	0	1.57	0.51	1.38	1.03	0.50
MCAMDF-NCCF	1.04	1.10	1.12	1.11	1.54	1.33
	0.20	0.47	0.38	0.38	1.16	0.72

表 4　不同信噪比下的清浊误判数

算法 \ *SNR*(dB)	∞	20	10	5	0	-5
NCCF	0	2	3	8	55	104
DWT-NCCF	1	2	3	8	41	79
MCAMDF-NCCF	1	2	3	7	9	19

图 9、10 分别给出了三种基音检测算法在不同信噪比下浊音段的总错误率及整段语音的清浊误判率比较.可以看出,MCAMDF-NCCF 算法在信噪比较低时仍能保证很高的准确率.由于使用了有效的基音检测前端处理技术,整段处理语音的清浊误判率大大下降,明显抑制了基音加倍减半的发生.

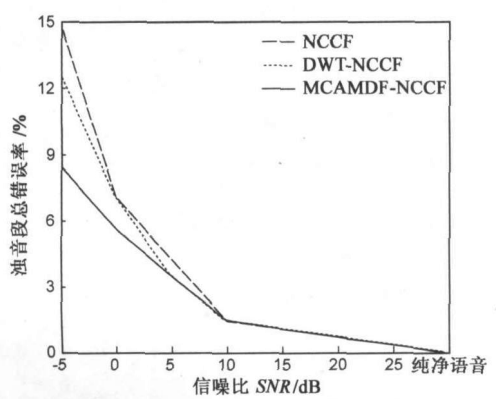

图 9　不同信噪比下的浊音段总错误率比较

图 11 给出了 *SNR* = -5dB 时三种算法得到的基音轨迹.可以看出,在低信噪比下采用 NCCF 与 DWT-NCCF 算法不仅在非语音段产生了频繁的清浊误判,而且在语音段产生了较多的基音加倍减半错误.而采用 MCAMDF-NCCF 算法所产生的基音轨迹较前两种算法平滑得多,清浊误判发生率很低,并且保持了较高的准确率.

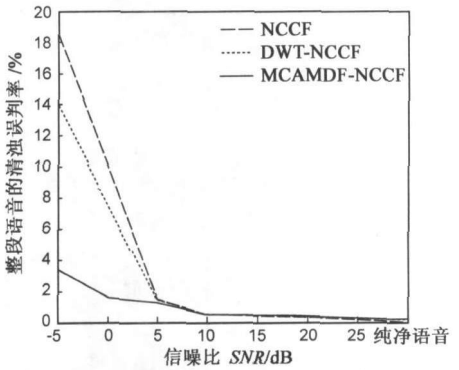

图 10　不同信噪比下的清浊误判率比较

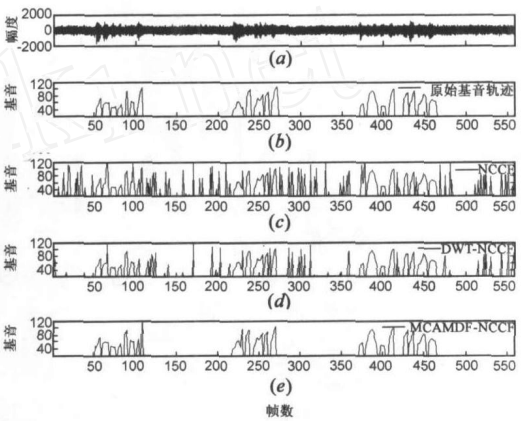

图 11　*SNR*= -5dB 时三种基音检测算法的基音轨迹

5　高精度 MCAMDF-NCCF-FRAC 基音检测算法

基音检测求取的基音周期通常为整数.为了满足不同情况下语音编码器对基音检测精度的要求,有时需要检测分数基音周期.

由于每帧语音可能包含多个基音周期波形,这样会导致求出的基音周期为一帧基音周期的平均值,即通常的整数基音周期.同时,采样率精度会限制基音周期的提取精度,即一个采样点以内的基音变化是无法被表示的.综合以上原因,Yoav Medan、Eyal Yair、Dan Chazan 三人于 1991 年提出了分数基音周期的提取技术[31].该技术被混合激励线性预测(Mixed-Excitation Linear Prediction,MELP)声码器所采用,分数基音被应用于整个编解码过程当中,在 2.4kb/s 速率得到了高质量的合成语音.

本文将分数基音检测技术与第 4 节提出的 MCAMDF-NCCF 基音检测算法相结合,提出了一种高精度 MCAMDF-NCCF-FRAC 基音检测算法.

分数基音提取基于如下假设,即浊音语音的基音周期缓慢渐变.假设整数基音周期为 T,利用线性插值技术与正交投影定理可以得到真实基音与整数基音的偏移量 β 及分数基音 $(T+\beta)$ 对应的互相关函数值 $\rho_{(T+\beta)}$,定义式见公式(10～12).

$$\beta=\frac{C_T(0,T+1)\,C_T(T,T)-C_T(0,T)\,C_T(T,T+1)}{C_T(0,T+1)\,[C_T(T,T)-C_T(T,T+1)]+C_T(0,T)\,[C_T(T+1,T+1)-C_T(T,T+1)]} \tag{10}$$

$$\rho_{(T+\beta)}=\frac{(1-\beta)\,C_T(0,T)+\beta C_T(0,T+1)}{\{C_T(0,0)\,[(1-\beta)^2\,C_T(T,T)+2\beta(1-\beta)\,C_T(T,T+1)+\beta^2\,C_T(T,T+1)]\}^{1/2}} \tag{11}$$

其中

$$C_T(i,j)=\sum_{n=0}^{N-1}s_{n+i}s_{n+j} \tag{12}$$

s_i 为输入语音.

高精度 MCAMDF-NCCF-FRAC 基音检测算法如下:首先用 MCAMDF-NCCF 算法求取整数基音.因为该算法可有效抑制基音加倍减半,所以直接将此整数基音用于分数基音提取.为了判定归一化互相关函数最大值落在区间 $(T-1,T)$ 还是区间 $(T,T+1)$,首先计算 $C_T(0,T-1)$ 与 $C_T(0,T+1)$.若 $C_T(0,T+1)>C_T(0,T-1)$,则归一化互相关函数最大值落在 $(T,T+1)$,可直接用公式(10)、(11)计算分数基音 β 与归一化互相关函数最大值 $\rho_{(T+\beta)}$;若 $C_T(0,T-1)>C_T(0,T+1)$,则归一化互相关函数最大值落在 $(T-1,T)$,此时需用 $T-1$ 替换公式(10)、(11)中的 T 后再计算 β 与 $\rho_{(T+\beta)}$.

注意:一般情况 $\beta\in[0,1)$.但有时因估计的整数基音 T 与真实基音 T 偏差超过一个样点,β 会落在区间 $[0,1)$ 以外.此时,若 $\beta\geq1$,整数基音加 1;若 $\beta<0$,整数基音减 1.然后用公式(10)、(11)及更新后的整数基音从新计算 β 与 $\rho_{(T+\beta)}$.图 12 为高精度 MCAMDF-NCCF-FRAC 基音检测算法流程图.

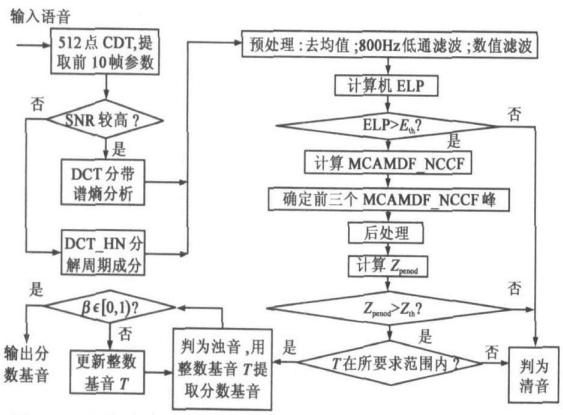

图 12 高精度 MCAMDF-NCCF-FRAC 基音检测算法流程图

选取 $SNR=10$dB 的实验语音,对本文建议的 MCAMDF-NCCF 算法及 MCAMDF-NCCF-FRAC 算法进行实验分析,图 13

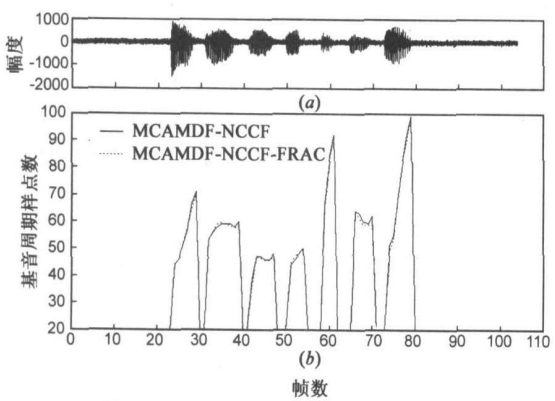

图 13 两种基音检测算法的基音轨迹

(b)中虚线、实线分别为两种算法检测的基音轨迹.可以看出,分数基音与整数基音轨迹非常接近,并在浊音段保持平滑渐变.

图 14 为上述两种算法应用于 WI 模型的结果.图 14(a)为任取的一帧合成语音,图 14(b)为该合成语音 0～2000Hz 频谱,图 14(c)为该合成语音的相位轨迹.图中实线、虚线分别表示应用 MCAMDF-NCCF、MCAMDF-NCCF-FRAC 基音检测算法后的结果.

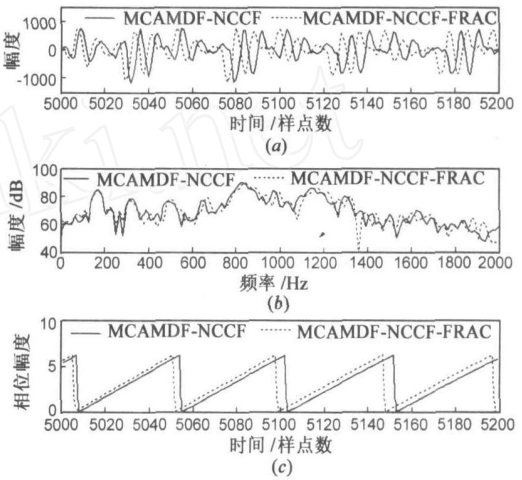

图 14 两种基音检测算法应用于 WI 模型的结果对比

从图 14 可以看出,两种基音检测算法所得到的基音之间微小分数偏差 β 的影响,在合成语音上表现为时间轴上微小的延迟,偏差 β 导致 WI 分析器在特征波形(Characteristic Waveform,CW)的提取与对齐时产生的结果略有不同,如图 14(a);在合成语音的频谱上表现为谐波峰值点处的微小偏差,基频时偏差不明显,随着谐波次数的增加,该偏差逐渐增大,如图 14(b);在合成语音所对应的相位上表现为相位轨迹的微小改变,这也充分证明了 WI 合成器用各帧基音线性内插出帧内每个样点的瞬时基音进行积分运算后对相位轨迹的影响.

6 低速率 WI 编码器中的性能评价

将本文提出的基音检测算法应用与 2kb/s WI 语音编码器,帧长 25ms(帧速率 40Hz),每帧提取 10 个 CW,所有参数的量化每帧执行一次.改进后的 WI 编码器与原始 2kb/s WI 编

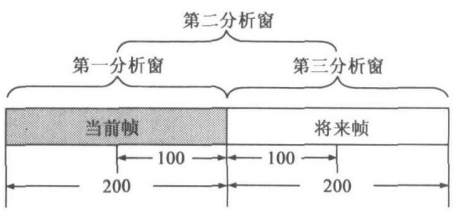

图 15 基音检测分析窗

码器[30]只有基音检测模块不同. 原始 2kb/s WI 编码器采用 DWT-NCCF 基音检测算法(其检测性能见第 4 节末). 为了有效防止基音加倍减半,对每帧语音都在三个混叠的分析窗[30]内分别进行基音检测,最后选择一个最优基音作为最终结果. 三个分析窗如图 15.

为得到最终基音 P_{opt},令三个分析窗的最优基音分别为 P_1、P_2、P_3,则当前帧的最终基音 P_{opt} 为:

$t_1 = (float)\ P_2/P_1;\ t_2 = (float)\ P_2/P_3;$

$t_3 = (float)\ P_1/P_2;\ t_4 = (float)\ P_3/P_2;$

if $((t_1 > 1.8\ \&\&\ t_2 > 1.8) \| (t_3 > 1.8\ \&\&\ t_4 > 1.8))\ P_{opt} = (P_1 + P_2)/2$; else $P_{opt} = P_2$.

本文选取男女各 8 句汉语语音实验句子,组织 10 名测试者采用主观 A/B 听力测试方式比较改进前后 WI 编码器的合成语音质量. 不同信噪比下的统计结果见表 5,表中数据表示喜欢某种 WI 方案合成语音质量的人数占总测试人数的比例,无偏爱表示测试者认为两种 WI 方案的合成语音质量相差不大,背景噪声为高斯白噪声.

从表 5 看出,WI 编码器改进前后性能接近,基音检测都很准确. 从表 6、7 看出,改进后的 WI 编码器质量占优. 这说明本文提出的基音检测算法在低信噪比下具有优势. 此外还应指出,由于 *SNR* 很低(比如 - 5dB)时背景噪声刺耳,在相当程度上影响测试者对改进后 WI 编码器提升性能的辨别. 但从第 4 节图 11(*d*)、(*e*)关于 DWT-NCCF 与 MCAMDF-NCCF 基音检测算法在 *SNR* = - 5dB 时的性能对比上看,MCAMDF-NCCF 所产生的浊音段基音轨迹比 DWT-NCCF 更趋于平滑和准确,清浊误判明显减少,对合成语音质量作用明显.

表 5　主观 A/B 测试结果

纯净语音	偏爱原始 WI	偏爱改进 WI	无偏爱
男性语音	33.8%	33.8%	32.5%
女性语音	28.8%	27.5%	43.8%
所有语音	31.3%	30.7%	38.2%

表 6　主观 A/B 测试结果

SNR = 0dB	偏爱原始 WI	偏爱改进 WI	无偏爱
男性语音	26.3%	35.0%	38.8%
女性语音	31.3%	42.5%	26.3%
所有语音	28.8%	38.8%	32.6%

表 7　主观 A/B 测试结果

SNR = - 5dB	偏爱原始 WI	偏爱改进 WI	无偏爱
男性语音	27.5%	38.8%	33.8%
女性语音	25.0%	35.0%	40.0%
所有语音	26.3%	36.9%	36.9%

7　结论

本文针对基音检测算法在不同噪声、信噪比下容易发生清浊误判的问题,提出基于 DCT 分带谱熵的语音检测算法划分语音段与非语音段;为了向基音检测核心算法提供更能准确反映基音周期实际变化的输入语音,基于谐波-噪声模型提出一种改进的 DCT 域语音分解算法. 然后,根据变形的 MCAMDF 与 NCCF 的峰值共性,结合上述两项基音检测前端处理技术,提出了 MCAMDF-NCCF 基音检测组合算法. 为了满足不同环境下 WI 编码器对基音检测高精度的要求,在合成端更准确地恢复相位轨迹,本文又基于 MCAMDF-NCCF 算法提出了高精度 MCAMDF-NCCF-FRAC 基音检测算法. 将本文建议的算法应用于 2kb/s WI 编码器,主观 A/B 听力测试结果表明,本文提出的基音检测算法在低信噪比下明显抑制了基音加倍减半及清浊误判的发生,得到了优异的基音检测结果,合成语音质量完全满足低速率 WI 编码器对基音检测技术的要求.

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.

[2] L R Rabiner. On the use of autocorrelation analysis for pitch detection[J]. IEEE Transactions on Acoustics Speech and Signal Processing,1977,25(1):24 - 33.

[3] M J Ross, H L Shaffer, A Cohen, R Freudberg, H J Manley. Average magnitude difference function pitch extractor[J]. IEEE Transactions on Acoustics Speech Signal Processing. 1974, 22(5):353 - 362.

[4] R J McAulay, T F Quatieri. Pitch estimation and voicing detection based on a sinusoidal speech model[A]. ICASSP'90[C]. Albuquerque, NM, USA:1990. 1:249 - 252.

[5] A M Noll. Cepstrum pitch determination[J]. The Journal of the Acoustical Society of America, 1967, 41(2):293 - 309.

[6] D W Griffin, J S Lim. Multi-band excitation vocoder[J]. IEEE Transactions on ASSP, 1988, 36(8):1223 - 1235.

[7] S Kadambe, G F Boudreaux-Bartels. Application on the wavelet transform for pitch detection of speech signals[J]. IEEE Transactions on Information Theory, 1992, 38(2):917 - 924.

[8] 顾良,刘润生. 利用声调判别提高汉语数码语音识别性能[J]. 清华大学学报,1998,38(9):36 - 39.

[9] 张文耀,许刚,王裕国. 循环 AMDF 及其语音基音周期估计算法[J]. 电子学报,2003,31(6):886 - 890.
ZHANG Wen-yao, XU Gang, WANG Yu-guo. Circular AMDF and pitch estimation based on it[J]. Acta Electronica Sinica, 2003, 31(6):886 - 890. (in Chinese)

[10] W Zhang, G Xu, Y Wang. Pitch estimation based on circular AMDF[A]. ICASSP'2002[C]. Orlando, Florida, USA:2002. 341 - 344.

[11] T Shimamura, H Kobayashi. Weighted autocorrelation for pitch extraction of noisy speech[J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(7):727 - 730.

[12] Om Deshmukh, Carol Y Espy-Wilson, Ariel Salomon, and Jawahar Singh. Use of temporal information: detection of periodicity, aperiodicity, and pitch in speech[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5):776 - 786.

[13] C Shahnaz, W-P Zhu, M O Ahmad. Robust pitch estimation at very low SNR exploiting time and frequency domain cues [A]. ICASSP'2005[C]. Philadelphia, PA, USA: 2005. 1: 389 - 392.

[14] C Shahnaz, W-P Zhu, M O Ahmad. A robust pitch estimation approach for colored noise-corrupted speech[A]. IEEE International Symposium on Circuits and System[C]. Kobe, Japan: 2005. 4: 3143 - 3146.

[15] Joseph Tabrikian, Shlomo Dubnov, Yulya Dickalov. Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model[J]. IEEE Transations on Speech and Audio Processing, 2004, 12(1): 76 - 87.

[16] 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.
Bao Changchun, Fan Changxin. Pitch detection algorithm based on normalized cross-correlation function[J]. Journal of China Institute of Communications, 1998, 19(10): 27 - 31. (in Chinese)

[17] 田野. 噪声环境下语音检测的稳健性问题[D]. 北京: 清华大学工学博士学位论文, 2003.

[18] L F Lamel, L R Rabiner, A E Rosenberg. An improved endpoint detector for isolated word recognition[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1981, 29(4): 777 - 785.

[19] J C Junqua, B Reaves. A robust algorithm for word boundary detection in the presence of noise[J]. IEEE Transactions on Speech and Audio Processing, 1994, 2(3): 406 - 412.

[20] E Nemer, R Goubran, S Mahmoud. Robust voice activity detection using higher-order statistics in the LPC residual domain [J]. IEEE Transactions on Speech and Audio Processing, 2001, 9(3): 217 - 231.

[21] G S Ying, C D Mitchell, L H Jamieson. Endpoint detection of isolated utterances based on a modified teager energy measurement[A]. ICASSP'1993[C]. Minneapolis, MN, USA: 1993. 2: 732 - 735.

[22] J L Shen, J W Hung, L S Lee. Robust entropy-based endpoint detection for speech recognition in noisy environments[A]. ICSLP'1998[C]. Sydney, Australia: 1998. 12 - 16.

[23] Bing-Fei Wu, Kun-Ching Wang. Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 762 - 775.

[24] Chuan Jia, Bo Xu. An improved entropy-based endpoint detection algorithm[A]. ICSLP'2002[C]. Denver: 2002. 285 - 288.

[25] C d'Alessandro, B Yegnanarayana, V Darsinos. Decomposition of the speech signal into deterministic and stochastic components[A]. ICASSP'1995[C]. IEEE, 1995. 1: 760 - 763.

[26] B Yegnanarayana, Christophe d'Alessandro and Vassilis Darsinos. An iterative algorithm for decomposition of speech signals into periodic and aperiodic components[J]. IEEE Transactions on Speech and Audio Processing, 1998, 6(1): 1 - 10.

[27] Raphael Ahn, W Harvey Holmes. An improved harmonic-plus-noise decomposition method and its application in pitch determination[A]. IEEE Workshop on Speech Coding for Telecommunications[C]. Pocono Manor, PA, USA. 1997: 41 - 42.

[28] Nazih Abu-Shikhah, Mohammed Deriche. Speech decomposition using the discrete cosine transform and adaptive thresholding[A]. Proceedings of IEEE TENCON 2000[C]. Kuala Lumpur, Malaysia: IEEE, 2000, 1: 49 - 52.

[29] L R Rabiner, M J Cheng, A E Rosenberg. A comparative performance study of several pitch detection algorithms[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, 24(5): 399 - 418.

[30] 李靓. 高质量的 2kb/s 波形内插语音编码算法研究[D]. 北京: 北京工业大学工学博士学位论文, 2005. 14 - 114.

[31] Yoav Medan, Eyal Yair, Dan Chazan. Super resolution pitch determination of speech signals[J]. IEEE Transactions on Signal Processing, 1991, 39(1): 40 - 48.

作者简介:

罗亚飞 男, 1980 年 9 月出生于河北省邯郸市. 2006 年于北京工业大学电控学院硕士毕业. 研究方向为低速率语音编码.
E-mail: luoyafei@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, 国际语音通信学会(ISCA)会员, 中国电子学会理事, 信号处理学会委员, 《通信学报》与《信号处理学报》编委. 主要研究领域为语音信号处理与编码.
E-mail: chchbao@bjut.edu.cn