

基于扩展谱相减的RCAF基音周期检测算法

王玥 钱志鸿 张莹
(吉林大学通信工程学院 长春 130025)

摘要: 针对传统基音检测算法在信噪比低的情况下提取的基音周期错误率较高, 该文提出了一种基于RCAF (Reverse CAMDF Autocorrelation Function) 搜索试探平滑的基音轨迹提取方法。采用自适应判决准则的扩展谱相减进行语音增强, 在语音段实现了对噪声信号的估计。应用RCAF算法提取基音周期, 通过搜索试探平滑算法对提取出的基音周期进行平滑处理。该算法降低了误判率, 提高了提取精度。仿真结果表明, 该算法在-10dB信噪比情况下, 其性能优于传统的CAMDF和AWAC等方法。

关键词: 语音信号处理; 基音检测; 扩展谱相减; RCAF

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2009)05-1161-05

RCAF Pitch Detection Algorithm Based on Expanded Spectral Subtraction

Wang Yue Qian Zhi-hong Zhang Ying

(College of Communication Engineering, Jilin University, Changchun 130025, China)

Abstract: A new pitch detection of noisy speech signal for lower SNR is proposed in this paper, which is based on Reverse CAMDF Autocorrelation Function (RCAF) and searching tentative smooth measurement. The algorithm can estimate noise during speech presence, which employs the method of expanded spectral subtraction based on noise compensation structure. RCAF algorithm improves the robustness and precision of pitch detection. A number of experiments show that by RCAF method, higher efficiency and better detection accuracy can be obtained while the SNR is equal to -10dB. However, such performance can not be achieved by traditional methods, AMDF, CAMDF and AWAC under the same SNR.

Key words: Speech signal processing; Pitch detection; Expanded spectral subtraction; Reverse CAMDF Autocorrelation Function(RCAF)

1 引言

语音信号的基音提取存在许多困难, 因为语音信号特性受诸多因素的影响。语音信号是短时平稳信号, 很难准确判定基音开始与结束的位置; 在清音及浊音的过渡帧很难判断是周期还是非周期; 基音动态范围大, 在 1.25~40ms 之间变化; 基音容易受声道响应(共振峰频率)影响; 基音周期受说话人情绪影响, 受拼读方法影响等。

目前的基音估计方法都存在种种不完美之处, 倒谱法^[1]估计基音周期效果较好, 但是计算量大, 且在噪音干扰下其灵敏度明显下降。时域法最具代表性的是短时平均幅度差函数(AMDF Average Magnitude Difference Function)法^[2]和短时自相关法^[3]。AMDF算法在语音信号的幅度或频率变化比较快时, 基音检测精度明显下降。短时自相关法通常会导导致“倍频”或“半频”错误, 且考虑到语音信号的短时性, 其窗函数及其窗长对分析信号影响很大; 线性预测法^[4]需假设语音信号为全极点模型来计算线性预测误差, 对流音和鼻音标

记不明显。目前已经有的基音检测方法, 在信噪比较高时具有良好的性能, 而在低信噪比时性能均下降明显。因此, 为克服影响准确估计基音的种种因素, 本文提出了一种基于搜索试探平滑的RCAF(Reverse CAMDF Autocorrelation Function)带噪语音基音提取算法。这种方法利用噪声补偿结构, 使用自适应的判决准则, 在语音段也可以实现对噪声信号的估计。在基音提取时, 应用RCAF算法准确地提取基音周期。实验证明, 该方法在低信噪比的条件下也能达到较好的效果。

2 基于谱减法语音增强技术

2.1 谱减法

在提取基音之前, 首先对原始语音信号进行语音增强, 以期从带噪语音信号中获得尽可能纯净的语音信号, 减少噪音的干扰。谱减法的基本出发点是直接从带噪语音信号中减去噪声谱, 并利用带噪语音相位重建增强后的“纯净语音”。由于人耳对语音相位不敏感, 所以对语音的可懂度及质量起重要作用的是语音信号的短时谱幅度而不是相位。在谱减法中利用人耳的这种感知特性, 以原始含噪语音的相位代替估

计语音的相位。它的优点是比较简单,只需要进行正反傅里叶变换,而且实时实现较容易。但是标准的谱减法需要利用语音激活检测对无语音段进行判断,进而估计背景噪声的特性,这样在语音段就无法对噪声进行更新,影响了背景噪声估计的准确性,同时语音激活检测的准确性也对背景噪声的估计有影响。因此本文采用改进型的维纳滤波器作为提取基音前语音增强的工具。

2.2 利用改进型维纳滤波器估计噪声

改进的谱相减法采用噪声补偿结构,对带噪语音中的输入噪声进行估计,然后利用谱相减的方法去除带噪语音中的噪声^[5],得到语音信号的估计值,再利用输出的信号来重新估计维纳滤波器的传递函数,形成了一个反向结构,因此可以看作是循环的维纳滤波器,其内部结构如图1所示,通过维纳滤波器 H_n 得到的输出为噪声功率谱估计,通过虚线部分后,得到噪声幅度谱估计的更新,最后系统输出为信号幅度谱的平滑估计。对于改进的维纳滤波器而言,不需要进行语音激活检测,在语音段和非语音段都可以对背景噪声进行更新,因此对背景噪声的估计相对于传统方法更加准确。

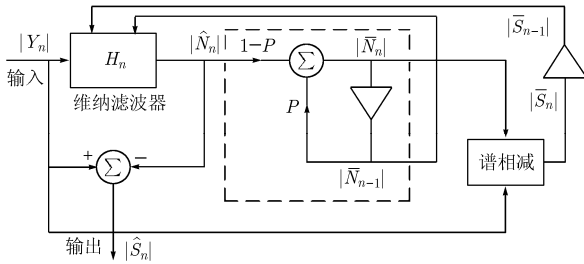


图1 扩展谱相减法的内部框图

改进的扩展谱相减的实现步骤如下:

(1) 设输入语音前 N_{noise} 帧为纯噪声信号,可以利用其幅度的统计平均得到噪声功率谱估计,其递推公式为

$$|\bar{N}'_n(e^{j\omega})|^2 = \frac{(n-1)|\bar{N}'_{n-1}(e^{j\omega})|^2 + |Y_n(e^{j\omega})|^2}{n} \quad (1)$$

其中 $n=1, \dots, N_{\text{noise}}$, $|\bar{N}'_n(e^{j\omega})|^2$ 为背景噪声的第 n 次统计估值,其初始值 $|\bar{N}'_0(e^{j\omega})|^2 = 0$ 。 $|Y_n(e^{j\omega})|^2$ 为第 n 帧带噪语音信号的功率谱。得到背景噪声幅度谱的初始平滑估计值:

$$|\bar{N}_0(e^{j\omega})| = |\bar{N}'_{N_{\text{noise}}}(e^{j\omega})| \quad (2)$$

(2) 利用幅度谱相减法计算语音信号的幅度谱初始平滑估计 $|\bar{S}_0(e^{j\omega})|$:

$$|\bar{S}_0(e^{j\omega})| = ||Y_{N_{\text{noise}}}(e^{j\omega})| - |\bar{N}_0(e^{j\omega})|| \quad (3)$$

其中 $|Y_{N_{\text{noise}}}(e^{j\omega})|$ 为第 N_{noise} 帧带噪语音信号的幅度谱。

(3) 初始化帧变量 $n=1$ 。

(4) 计算第 n 帧的维纳滤波器的传递函数:

$$H_n(e^{j\omega}) = \left(\frac{|\bar{N}_{n-1}(e^{j\omega})|^2}{|\bar{N}_{n-1}(e^{j\omega})|^2 + |\bar{S}_{n-1}(e^{j\omega})|^2} \right)^{1/2} \quad (4)$$

其中 $|\bar{N}_{n-1}(e^{j\omega})|^2$ 是噪声功率谱的平滑估计, $|\bar{S}_{n-1}(e^{j\omega})|^2$ 是

信号功率谱的估计值,可以看出,当前帧的维纳滤波器的传递函数是利用上一帧的信息得到的。

(5) 对带噪语音信号幅度谱进行滤波,得到当前帧背景噪声幅度谱的估计值 $|\hat{N}_n(e^{j\omega})|$:

$$|\hat{N}_n(e^{j\omega})| = H_n(e^{j\omega}) |Y_n(e^{j\omega})| \quad (5)$$

(6) 计算谱减后得到信号的幅度谱估计值 $|\hat{S}_n(e^{j\omega})|$:

$$|\hat{S}_n(e^{j\omega})| = ||Y_n(e^{j\omega})| - |\hat{N}_n(e^{j\omega})|| \quad (6)$$

(7) 利用当前帧噪声幅度谱估计值 $|\hat{N}_n(e^{j\omega})|$ 对背景噪声的平滑估计值 $|\bar{N}_n(e^{j\omega})|$ 进行更新:

$$|\bar{N}_n(e^{j\omega})| = p |\bar{N}_{n-1}(e^{j\omega})| + (1-p) |\hat{N}_n(e^{j\omega})| \quad (7)$$

其中 p 为比例因子,如果语音信号变化的速率可以和噪声信号变化的速率很好的分开,那么就可以对 p 进行合理的设置,使谱相减得到较好的效果。这时 $|\hat{N}_n(e^{j\omega})|$ 和 $|\bar{N}_{n-1}(e^{j\omega})|$ 变化较慢,而 $|\bar{S}_n(e^{j\omega})|$ 变化较快。

(8) 计算信号幅度谱的平滑估计 $|\bar{S}_n(e^{j\omega})|$:

$$|\bar{S}_n(e^{j\omega})| = ||Y_n(e^{j\omega})| - |\bar{N}_n(e^{j\omega})|| \quad (8)$$

(9) 帧变量 n 递增;如果 $n > N$, 则结束,得到谱相减后的语音幅度谱的估计值 $|\hat{S}_n(e^{j\omega})|$ 作为输出;否则转(4)继续执行。

应用上述改进的维纳滤波器进行扩展谱相减,不需要语音激活检测,能够方便准确地对噪声谱进行估计,从而更好地对带噪语音信号进行增强。

3 RCAF 与搜索试探平滑算法

为了能够准确地提取出带噪语音信号的基音周期,首先对语音信号使用扩展谱相减去除噪声,再对降噪后的语音信号 $\hat{s}(n)$ 根据RCAF算法提取基音周期,最后对提取的错误点用搜索试探的平滑算法进行平滑,将平滑后的基音周期作为提取结果输出。具体的实现框图如图2。

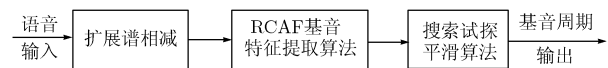


图2 基于搜索试探平滑算法的RCAF带噪语音基音提取方法框图

3.1 RCAF 算法

设 $s(n)$ 为纯净语音信号, $w(n)$ 为噪声信号, $y(n)$ 为带噪语音信号,则有 $y(n) = s(n) + w(n)$ 。 $s(n)$ 的自相关函数 $R_{ss}(\tau)$ 为

$$R_{ss}(\tau) = \frac{1}{N} \sum_{t=0}^{N-1-|\tau|} s(n)s(n+\tau) \quad (9)$$

其中 N 是语音分析帧的长度, τ 是延长度。自相关函数 $R_{ss}(\tau)$ 在 $s(n)$ 的基音周期 T 整数倍处呈现出峰值特性, $R_{ss}(\tau)$ 的第2最大峰值 ($\tau = T$) 相对于第1最大峰值 ($\tau = 0$) 的距离通常用来估计基音周期,这就是基于自相关法的基音提取。与自相关函数类似,平均幅度差函数 AMDF 函数同样具有基音周期性,具体函数定义为

$$\xi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} |s(n) - s(n + |\tau|)| \quad (10)$$

$\xi(\tau)$ 在基音周期处呈现谷值特性, 而自相关函数 ACF (AutoCorrelation Function) 则在对应处 ($\tau = \rho T$) 呈现峰值特性, 文献[6]采用 AMDF 倒数加权的 ACF 方法, 即自相关加权平均幅度差方法 AWAC (AMDF Weighted Auto Correlation) 提取基音。AWAC 方法是通过 AMDF 的谷值来增强 ACF 的峰值, 然而传统的 AMDF 由于其自身的谷值随着时间的增加呈递减趋势, 这样在取倒数后会对自相关函数的较小峰值起加强作用, 影响真正峰值的搜索。同时在低信噪比下, $\xi(\tau)$ 的谷值可能出现在 $T < \tau \leq 2T$ 处, 而循环平均幅度差函数 CAMDF (Circular AMDF)^[7] 则只在基音周期处出现谷值。且 CAMDF 的谷值随着时间的推移呈现递增趋势, 可以弥补 AMDF 的谷值变化的不足, CAMDF 函数可以定义为^[8]

$$\xi'(\tau) = \sum_{n=0}^{N-1-|\tau|} |s_w(\text{mod}(n + \tau, N)) - s_w(n)| \quad (11)$$

其中 $\text{mod}(n + \tau, N)$ 是取模操作, 将 $n + \tau$ 对 N 取模, $s_w(n)$ 是一段加窗的语音信号。

由于 CAMDF 在基音周期整数倍处存在谷值, 并且谷值呈递减趋势, 因此本文提出一种基于 CAMDF 的倒数加权自相关来进行基音周期估计, 即 RCAF (Reverse CAMDF Autocorrelation Function) 算法, 其计算公式为

$$\chi(\tau) = \frac{\frac{1}{N} \sum_{t=0}^{N-1-|\tau|} s(n)s(n + \tau)}{\frac{1}{N} \sum_{n=0}^{N-1-|\tau|} |s(n) - s(n + |\tau|)| + k} = \frac{R_{ss}(\tau)}{\xi'(\tau) + k} \quad (12)$$

其中 $R_{ss}(\tau)$ 是语音信号的自相关函数, $\xi'(\tau)$ 为语音信号的循环幅度差函数, k 为一常数 ($k > 0$), 当式(11)中 $\tau = 0$ 时,

$$\xi'(0) = 0 \quad (13)$$

此时式(12)中分母需要加入常数 k , 以避免分母为 0。

采用式(12)提取基音, 会使基音周期处的谷值点检测更容易。因为水平的波动趋势, 更容易确立一致的谷值点显著性检验标准, 简化了基音周期的检测过程。由于基音周期位置上谷值点的函数值依次递增, 利用求峰值的方法可以一次定位到估计的基音周期位置, 而不需要复杂的基音检测逻辑, 与此同时, 还可避免因不适当的基音判决逻辑而造成基音周期误判的情况。此外, RCAF 函数对清音帧与浊音帧有着明显的区别特征。对于类似白噪声的清音, RCAF 也呈现类噪声特性, 信号的噪声特性越强, 该函数的波动就越剧烈, 局部最小点的分布就越密集, 函数值的变化范围也越来越小, 提供了新的类似于过零率的清浊判决依据^[9]。RCAF 函数还具有较好的抗噪性能, 也不易受谐波和共振峰的影响。在清浊音混合的情况下, RCAF 函数也能够估计出浊音部分的基音周期。

3.2 搜索试探平滑算法

在提取基音的过程中, 无论哪种方法提取的基音轨迹与真实的基音轨迹都不可能完全吻合。其错误点主要包括倍频

点、半频点和随机错误点。因此在基音提取之后需要对基音曲线作平滑处理。

传统的平滑方法, 如线性平滑和中值平滑^[10]能够得到较好的结果, 但不能处理连续的基音频率有随机错误点的情况。而一般的线性插值, 很容易使出现连续随机错误点的部分形成直线, 影响最后的识别效果。本文采用一种新的平滑方法, 用搜索试探的方法来得到基音轮廓。实验表明这种方法对汉语发音获得的结果比传统方法有显著提高, 而且运算简单, 速度快。另一方面, 该方法不用做语音切分来区分清音和浊音, 而在搜索过程中得到的最终基音轮廓的边界就是浊音部分。

设 f_1, f_2, \dots, f_N 代表连续 N 帧用 RCAF 方法得到的基音频率, f'_i 为第 i 帧平滑后得到的基音频率^[10]。当处理第 i 帧基音频率时, 首先处理提取过程中出现的半频和倍频错误:

$$f'_i = \begin{cases} f_i/2, & |f_i/2 - f_{i-1}| < c_1 \\ 2f_i, & |2f_i - f_{i-1}| < c_1 \end{cases} \quad (14)$$

消除半频倍频错误后, 相邻基音如果还有较大幅度变化, 较大幅度变化的点则称之为野点, 针对野点平滑方案如下:

$$\left. \begin{aligned} f'_i &= 2f_{i-1} - f_{i-2}, & |f_i - f_{i-1}| > c_1 \text{ 且 } |f_{i+1} - f_{i-1}| > c_2 \\ f'_i &= (f_{i-1} + f_{i+1})/2, & |f_i - f_{i-1}| > c_1 \text{ 且 } |f_{i+1} - f_{i-1}| \leq c_2 \end{aligned} \right\} \quad (15)$$

式(14), 式(15)中, c_1, c_2 为经验阈值。根据采样频率和人发音的频率范围, c_1 通常为 10, c_2 通常为 20。两个阈值和上面规则的作用就是限制相邻帧之间的基音频率变化不超过 c_1 , 而相隔两帧的基音频率变化不超过 c_2 , 从而达到对基音平滑的效果。

4 仿真实验

为了验证基于平滑的 RCAF 带噪语音基音检测法的有效性, 在不同信噪比下, 对输入信号分别进行 AMDF, CAMDF, AWAC 和 RCAF 四种方法的处理, 并对其结果进行比较, 如图 3 所示。实验数据采样率为 8kHz, 量化等级为 16bit, 帧长 52ms, 帧移 26ms。语音内容为男声“数字信号处理”短句。图中分别给出了语音信号在 -5dB, -10dB 和 -15dB 信噪比下本文提出方法与其他基音提取算法所得基音周期结果。仿真图中的 x 轴为基音采样点数, y 轴为基音频率(Hz)。

从图3的4种基音提取方法比较实验可以看出, AMDF 方法随着滞后时间的增加, 峰值幅度逐渐下降。这使得谷值点检测以及谷值点的清晰度检查比较困难, 加倍误判、减半误判的概率比较高。可以看出 SNR 在 -5dB 时 AMDF 方法几乎已经失效, 但 CAMDF, AWAC 和 RCAF 都可以准确提取基音轮廓。当信噪比大于等于 -10dB 时, CAMDF 和 AWAC 的带噪语音基音检测算法出现了提取错误的情况, 均不能准确地

提取出基音周期，但是基于平滑的RCAF带噪语音基音检测法能够准确的提取出基音周期。当SNR恶劣到-15dB时，所有方法都不能有效的进行基音提取。由此可见，与传统的AMDF，CAMDF和AWAC算法相比较，基于平滑的RCAF带噪语音基音检测法能在更低的信噪比较准确地提取基音轮廓，算法的抗噪性能更强。计算量方面，CAMDF只需要加，减和取绝对值运算，运算量约是AMDF的两倍^[8]，而RCAF与AWAC方法都要得到CAMDF函数值与AMDF函数值，再进行一次加法和除法运算，因此RCAF运算量约是AWAC的两倍。为了减少数据访问次数，基音估计算法中的幅度差函数及其均值的计算，局部最小点的检测与统计都可以在一个数据遍历中完成。

为了进一步验证算法通用性，采用北京大学视觉与听觉信号处理国家重点实验室录制的PKU-SRSC^[11]语音数据库进行大量实验。从数据库中选择其中20个说话人，10个男生

M1-M10，10个女生F1- F10，截取纯净说话语音15s，噪声采用NOISEX-92数据库中的高斯白噪声，汽车噪声以及babble噪声。将PKU-SRSC库中纯净语音混入不同响度的噪声，使得输入信噪比为0dB，-5dB和-10dB。

算法的评估方法基于Rabiner^[12]公式

$$e(n) = F_t(n) - F_e(n) \tag{16}$$

其中 $F_t(n)$ 是真正的基音频率， $F_e(n)$ 是所提取的基音频率， $e(n)$ 为提取误差。若 $|e(n)| \geq 10\text{Hz}$ ，称为粗略基音误差GPE(Gross Pitch Error)，否则称为精细基音误差FPE(Fine Pitch Error)。对于GPE只计算百分比，对FPE则给出频率误差平均值。为了测试所提出的RCAF方法针对传统3种方法的基音检测精度，依次提取20个说话人的基音，表1和表2分别给出了3种典型噪声在不同信噪比下各种算法提取基音的测试结果。

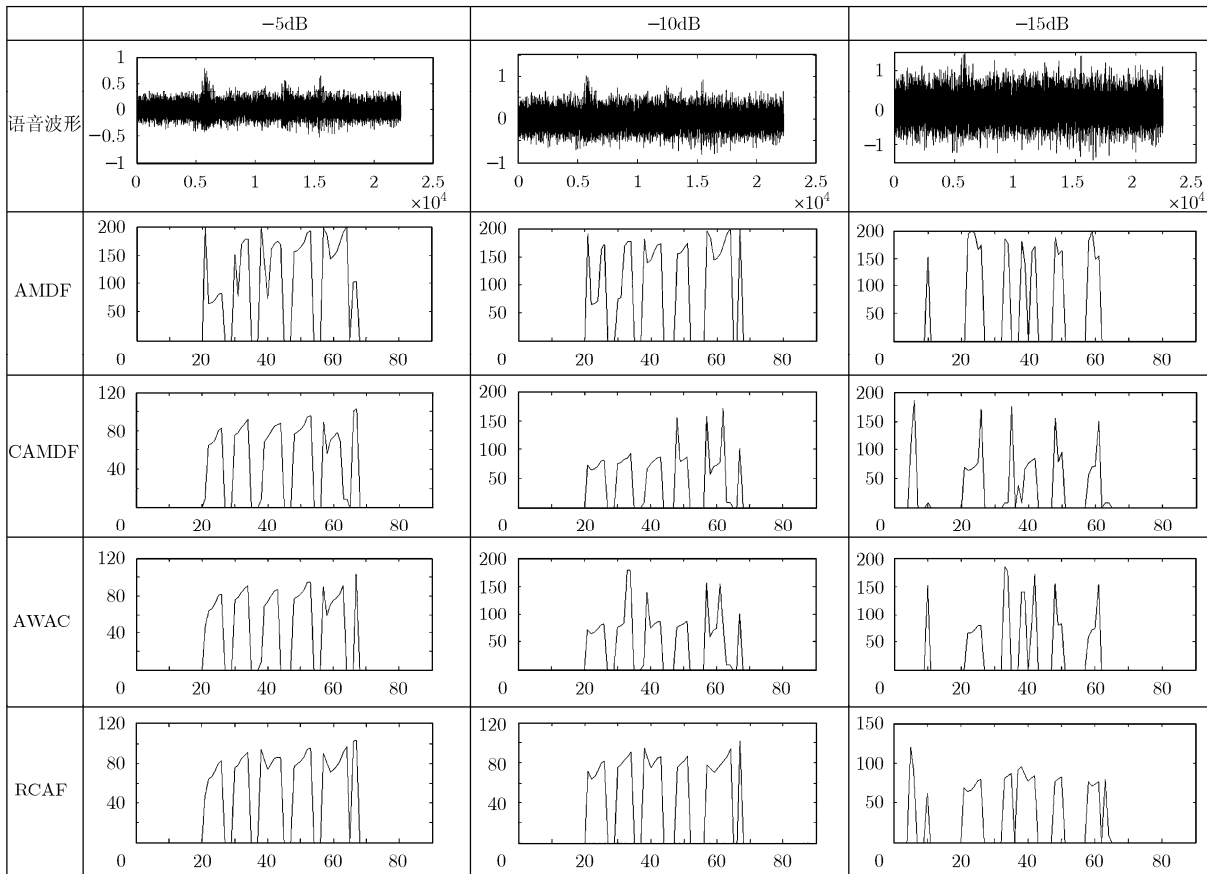


图3 4种基音提取方法比较

表1 粗略基音误差GPE(%)

| 算法 | White 噪声 | | | Babble 噪声 | | | Volvo 噪声 | | |
|-------|----------|------|-------|-----------|------|-------|----------|------|-------|
| | 0dB | -5dB | -10dB | 0dB | -5dB | -10dB | 0dB | -5dB | -10dB |
| AMDF | 5.4 | 28.2 | 55.0 | 9.1 | 42.1 | 60.7 | 8.4 | 36.7 | 58.1 |
| CAMDF | 2.6 | 13.7 | 31.5 | 6.2 | 25.5 | 46.1 | 3.9 | 15.3 | 37.5 |
| AWAC | 3.2 | 15.8 | 34.1 | 6.8 | 22.7 | 44.2 | 4.3 | 17.8 | 37.3 |
| RCAF | 2.9 | 11.3 | 18.6 | 5.4 | 16.4 | 33.3 | 2.4 | 11.9 | 24.7 |

表2 精细基音误差FPE(Hz)

| 算法 | White 噪声 | | | Babble 噪声 | | | Volvo 噪声 | | |
|-------|----------|------|-------|-----------|------|-------|----------|------|-------|
| | 0dB | -5dB | -10dB | 0dB | -5dB | -10dB | 0dB | -5dB | -10dB |
| AMDF | 3.93 | 4.62 | 7.15 | 4.24 | 5.23 | 7.67 | 4.16 | 4.85 | 7.41 |
| CAMDF | 3.71 | 3.94 | 6.54 | 3.50 | 4.27 | 6.35 | 3.74 | 4.21 | 7.12 |
| AWAC | 3.54 | 3.85 | 5.47 | 3.73 | 4.38 | 6.43 | 3.35 | 3.64 | 6.77 |
| RCAF | 3.38 | 3.82 | 5.02 | 3.67 | 4.14 | 6.51 | 3.54 | 3.76 | 6.64 |

从表1中可以看出,3种不同噪声在信噪比为0dB情况下,AMDF,CAMDF,AWAC和RCAF算法所得到的基音误差都在10%的范围内,提取精度相差不大。但在-5dB和-10dB信噪比时,随着信噪比的下降背景噪声加强,算法逐渐出现误差。其中AMDF方法已经严重失效,而CAMDF,AWAC和RCAF3种方法中,RCAF方法具有最好的效果,白噪声情况下得到的GPE在20%以内,Babble鸡尾酒会噪声下误差最大。在表2的精细误差数据中也可以看到同样的情况。RCAF方法的优点在于CAMDF取倒数以后其递减的峰值不会加重自相关函数波形,而且能够减少由共振峰和噪声所引起的突出点对搜索峰值的影响,从而精确地提取基音周期。大量实验说明RCAF方法相对于传统算法具有更强的抗噪声性能。

5 结论

传统的基音检测算法多数只能在信噪比较为理想的情况下进行准确的基音提取,但实际应用中多为噪声污染信号,本文针对此现象,提出了一种基于RCAF搜索试探平滑的基音轨迹提取方法,采用噪声补偿结构,使用自适应的判决准则,使得在语音段也可以实现对噪声信号的估计。在CAMDF算法基础上,使用RCAF算法准确地提取带噪语音的基音周期,并通过搜索试探平滑算法对提取出的基音周期进行平滑处理,纠正个别错误点。对比实验表明,基于RCAF搜索试探平滑的基音轨迹提取算法明显地降低了误判率,提高了基音估计精度。该算法已经在实际嵌入式DSP说话人识别系统中得到了应用,取得了良好的效果。

参 考 文 献

- [1] Zhao X F, O'Shaughnessy D, and Nguyen M Q. A processing method for pitch smoothing based on autocorrelation and cepstral F0 detection approaches. International Symposium on Signals, Systems and Electronics, Montreal, 2007: 59-62.
- [2] Deshmukh O, Espy-Wilson C Y, Salomon A, and Singh J. Use of temporal information: Detection of periodicity, aperiodicity, and pitch in speech. *IEEE Transaction on Speech and Audio Processing*, 2005, 13(5): 776-786.
- [3] Amado R G and Filho J V. Pitch detection algorithms based on zero-cross rate and autocorrelation function for musical notes. International Conference on Audio, Language and Image Processing, Shanghai, 2008: 449-454.
- [4] Ykhlef F, Mhania G, and Mesaoud B. Pitch modification of speech signal using source filter model by linear prediction for prosodic transformations. Proceedings of the Future Generation Communication and Networking, Jeju-Island, 2007: 414-419.
- [5] Shahnaz C, Zhu W P, and Ahmad M O. A pitch detection method for speech signals with low signal-to-noise ratio signals, systems and electronics. International Symposium on Signals, Systems and Electronics, Montreal, 2007: 399-402.
- [6] Shimamura T and Kobavashi H. Weighted autocorrelation for pitch extraction of noisy speech. *IEEE Transaction on Speech and Audio Processing*, 2001, 9(7): 727-730.
- [7] Shahnaz C, Zhu W P, and Ahmad M O. Robust pitch estimation at very low SNR exploiting time and frequency domain cues. International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, 2005: 389-392.
- [8] Zhang W Y, Xu G, and Wang Y G. Circular AMDF and pitch estimation based on it. *Acta Electronica Sinica*, 2003, 31(6): 886-890.
- [9] Zhu X Y, Wang Y, and Liu J. An approach of fundamental smoothing for Chinese tone recognition. *Journal of Chinese Information Processing*, 2000, 15(2): 45-50.
- [10] Zhao Z D, Hu X M, and Tian J F. An effective pitch detection method for speech signals with low signal-to-noise ratio. International Conference on Machine Learning and Cybernetics, Kunming, 2008, Vol.5: 2775-2778.
- [11] Wu Xi-hong. A Chinese speech database for speaker recognition. <http://nlpr-web.ia.ac.cn/english/irds/chinese/sino-biometrics-pdf/wuxihong.pdf>, 2002.
- [12] Rabiner L R, Cheng M J, and Rosenberg A E, et al. A comparative performance study of several pitch detection algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1976, 24(5): 399-417.

王 玥: 男, 1980年生, 博士生, 研究方向为语音信号处理和DSP应用。

钱志鸿: 男, 1957年生, 教授, 博士生导师, 主要从事领域为无线网络通信系统的信号分析和处理、通信系统微弱信号检测理论与应用等。

张 营: 女, 1984年生, 硕士, 研究方向为语音信号处理。