

一种基于高阶循环累积量的改进型 VAD 算法

朱晓亮 杜旭 杨宗凯

(华中科技大学电子与信息工程系 武汉 430074)

摘 要 根据语音信号的短段循环平稳(CycloStationary, CS)特征, 该文提出了一种应用于复杂背景噪声条件下的基于高阶循环累积量的改进型 VAD(Voice Activity Detection)算法, 算法采用 MA(Moving Average)模型对语音信号建模, 并选择平均幅度差(Average Magnitude Difference Function, AMDF)的方法来估算循环频率以降低算法复杂度。经 VoIP(Voice over Internet Protocol)平台测试, 算法对高斯(白色或有色)噪声以及其它平稳噪声自适应能力强、检测性能突出, 且恢复后语音质量损失较小, 对于非对称噪声也具备可检测能力。

关键词 语音信号处理, 高阶循环累积量, 非连续传输, VoIP, VAD, G.729

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2006)06-1021-05

An Improved Higher-Order Cyclic Cumulant Based VAD Algorithm

Zhu Xiao-liang Du Xu Yang Zong-kai

(Dept. of Electronics & Information Eng., Huazhong University of Science & Technology, Wuhan 430074, China)

Abstract On the basis of cyclostationary feature of speech in short segment, an improved Voice Activity Detection (VAD) algorithm based on higher-order cyclic cumulant is proposed in this paper, which can be applied in complex noise background. In this algorithm, modeling speech signal under MA(Moving Average)model is adopted and estimating cyclic frequency by AMDF(Average Magnitude Difference Function) is chosen to decrease computing complexity as well. Extensive testing on VoIP platform reveals that the algorithm owns good adaptive classification capability in steady noise, such as Gauss (whited and colored) noise, while with a little loss in speech quality, even when it is applied in asymmetric noise conditions.

Key words Processing of speech signals, Higher-order cyclic cumulant, Discontinuous transmission, VoIP, VAD, G.729

1 引言

非连续传输技术是现代移动通信系统以及 VoIP 系统重要的组成部分。其主要包括语音活动性检测(VAD)和舒适音产生(Comfort Noise Generation, CNG)技术。其中, VAD 算法是非连续传输技术的关键, 它判断当前输入帧是否为语音帧。最终, 仅语音帧被发送。这在移动通信系统中带来的直接好处是降低移动台的发射功率、减少系统间干扰、增加系统的容量和保证通话质量。在 VoIP 系统中, 则可以抑制非语音帧传输, 从而节省带宽、降低网络负载、保证服务质量^[1]。VAD算法的瓶颈在于如何寻找复杂背景噪声下实时性和检测性能的最佳平衡。世界各地的研究人员已提出了众多 VAD算法, 其中包括了时域和频域处理方法^[2-5], 如短时能量检测^[2]、过零率检测^[2]、LPC参数分析^[2]、倒谱分析^[4]、共振峰检测^[5]等。这些方法较简洁, 但在复杂类型噪声(如有色高斯噪声、瑞利、指数噪声等)环境中将降低或失去检测效果。ITU-T的G.729 B VAD^[3,6]算法结合了线性频率、低频能量、过零率、全频段能量等特征, 但算法采用保守策略, 难以满

足一些特殊应用场合的性能要求, 例如在移动Voip网络中, 特别是存在需要有效切换收/发功能的半双工设备时算法表现不佳, 因此有必要寻找新的解决方案。高阶累积量对于高斯噪声(有色和白噪声)或其它平稳加性噪声等均具有良好的盲处理能力^[7-9], 因此可将其作为VAD检测特征。如Nemer^[10]在假设语音信号是短时平稳信号的前提下采用正弦信号模型建模, 提出了一种对LPC滤波结果进行高阶累积量统计的VAD算法, 但未考虑语音信号是典型的非平稳信号的事实且算法复杂。为此, 本文引入循环平稳的概念对高阶累积量统计方法进行优化。通常, 循环平稳信号的相关函数随事件按周期或多周期(各周期不能通约)规律变化, 其周期性采用循环频率来衡量, 所有的循环频率组成循环频率集。水文、气象数据以及心电图等人体信号均具备循环平稳性^[7]。语音信号可认为是经基音调制的一组随机序列^[11], 显然该序列调制后在一短段时间内具备周期平稳性。将循环平稳分析方法引入VAD算法有两点工作需要完善: (1)算法实时性的保证, 即算法复杂度的降低; (2)循环频率估计方法的选择。

本文工作基于以上两点展开, 提出以语音段的基音周期作为循环频率的估计, 并采用平均幅度差方法估计循环频率, 定义了以各子窗口高阶循环累积量实部的累加和作为检

测量, 设计了改进型 VAD 算法, 最后将其与 G.729 B VAD 做详细的比较分析。

2 高阶循环累积量理论分析

2.1 高阶循环累积量的定义^[7,8]

定义 1 随机变量 x 具有概率密度函数 $f(x)$, 定义特征函数 $\phi(\omega)$ 是密度 $f(x)$ 的 Fourier 变换:

$$\phi(\omega) = \int_{-\infty}^{\infty} f(x) e^{j\omega x} dx \quad (1)$$

$\phi(\omega)$ 又称为矩生成函数或第一特征函数。 $\phi(\omega)$ 在原点的 k 阶导数称为 x 的 k 阶矩 m_k ; 设累积量生成函数 $\psi(\omega) = \ln \phi(\omega)$, 其在原点的 k 阶导数称为 x 的 k 阶累积量 c_k 。

定义 2 对幅度受限的随机信号序列 $x(t)$ 进行非线性变换 $g[x(t)]$, 若可从中产生有限强度的加性正弦波, 且序列本身典型的并不包含有限强度的加性正弦波分量, 则 $x(t)$ 称作一个循环平稳序列。产生的正弦波的频率称为循环频率 ω , 所有的循环频率组成循环集 $\Omega = \{\omega_i\}$, 则 $g[x(t)]$ 的时变数学期望 $\hat{E}[g(t)]$ 可用 Fourier 级数表示:

$$\hat{E}[g(t)] = \sum_{\omega} G(\omega) e^{j\omega t} \quad (2)$$

式中 Fourier 系数 $G(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} g(t) e^{-j\omega t} dt$ 。

定义 3 定义序列 $x(t)$ 的 k 阶滞后积

$$L_{kx}(t; \tau_1, \dots, \tau_{k-1}) = \left\{ \prod_{j=1}^k x(t + \tau_j) \right\} \quad (3)$$

k 阶时变矩函数 $m_{kx}(t; \tau_1, \dots, \tau_{k-1})$ 定义为 k 阶滞后积的时变数学期望, 则

$$m_{kx}(t; \tau_1, \dots, \tau_{k-1}) = \hat{E}\{L_{kx}(t; \tau_1, \dots, \tau_{k-1})\} \quad (4)$$

定义 4 对于固定的 $\tau_1, \dots, \tau_{k-1}$, 若矩函数 $m_{kx}(t; \tau_1, \dots, \tau_{k-1})$ 存在一个 Fourier 级数展开, 则

$$m_{kx}(t; \tau_1, \dots, \tau_{k-1}) = \sum_{\omega \in \Omega'} M_{kx}(\omega) e^{j\omega t} \quad (5)$$

$$M_{kx}(\omega) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^{N-1} m_{kx}(t; \tau_1, \dots, \tau_{k-1}) e^{-j\omega t} \quad (6)$$

其中循环频率集 $\Omega = \{\omega: M_{kx}(\omega) \neq 0, 0 \leq \omega \leq 2\pi\}$, 且 Ω 中的元素个数是可数的。Fourier 系数 $M_{kx}(\omega)$ 称作在循环频率的 k 阶循环矩。

定义 5 对于固定的 $\tau_1, \dots, \tau_{k-1}$, 若时变累积量 $c_{kx}(t; \tau_1, \dots, \tau_{k-1})$ 存在一个 Fourier 级数展开, 则

$$c_{kx}(t; \tau_1, \dots, \tau_{k-1}) = \sum_{\omega \in \Omega'} C_{kx}(\omega) e^{j\omega t} \quad (7)$$

$$C_{kx}(\omega) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^{N-1} c_{kx}(t; \tau_1, \dots, \tau_{k-1}) e^{-j\omega t} \quad (8)$$

式中 Fourier 系数 $M_{kx}(\omega)$ 称作在循环频率的 k 阶循环累积量。可认为 $M_{kx}(\omega)$ 是 $x(t)$ 的经循环频率加权的广义相关函数。

2.2 高阶循环累积量估计

式(5)和式(7)中的循环频率集 Ω 和 Ω' 不一定相同, 如

$x(t)$ 是零均值信号, 则 $k=2,3$ 时两者相同, 不过, 在大多数情况通常仅考虑两者相同的情况^[7]。若将两者中的元素分别采用 β, α 表示, 得到循环矩和循环累积量之间的转换公式。如式(9)所示, k 阶循环累积量可以用较低阶的循环矩函数表示为^[7,8]

$$C_{kx}(\alpha) = \sum_{\cup_{p=1}^q I_p = I} \left[(-1)^{q-1} (q-1)! \sum_{\beta_1 + \dots + \beta_q = \alpha} \prod_{p=1}^q M_{n_p, x}(\beta_p) \right] \quad (9)$$

式中 $\sum_{\cup_{p=1}^q I_p = I}$ 表示在指数符集 $I = \{0, 2, \dots, k-1\}$ 的所有无交连的非空集合分割 ($1 \leq q \leq k$) 内的求和; 而 $n_p = |I_p|$ 表示分割 I_p 中的元素个数。 $x(t)$ 的 k 阶样本循环矩估计值^[8]:

$$\hat{M}_{kx}(\omega) = \frac{1}{N} \sum_{t=0}^{N-1} x(t) x(t + \tau_1) \cdots x(t + \tau_{k-1}) e^{-j\omega t} \quad (10)$$

当样本数 $N \rightarrow \infty$ 时, 该估计值为其真实值的强一致估计。若假定循环阶数为 3, 简化处理式(9)得到

$$\hat{C}_{3x}(\omega) = \hat{M}_{3x}(\omega) \quad (11)$$

2.3 循环频率估计

估算循环累积量的关键是有用地估计循环频率, 语音信号的循环频率可用基音周期表示, 为了降低算法复杂度, 本文采用平均幅度差函数(AMDF)方法^[11]获得。假定语音序列为 $x(n)$, 窗口函数为 $w(n)$, 定义短时平均幅度差函数:

$$F_n(\tau) = \sum_{m=0}^{N-1-\tau} |x_n(m) - x_n(m + \tau)| \quad (12)$$

其中 $x_n(m) = w(n)x(n+m)$; $x_n(m + \tau) = w(n)x(n+m + \tau)$ 。显然, 如果 $x(n)$ 在窗口取值范围内具有周期性, 设周期对应的采样点数为 N_p , 则 $F_n(\tau)$ 在 $\tau = N_p, 2N_p, \dots$ 时将出现零值。实际的语音信号是一个准周期信号, 经计算机模拟, $F_n(\tau)$ 在浊音语音的基音周期上出现极小值, 而在清音语音时没有明显的极小值。对于不同的 τ 值, 计算 $F_n(\tau)$ 的差值是不同的, 随着 τ 值的增加, 差值的个数将减少。为了克服此缺点, 可令 $x_n(m) = w(n)x(n+m)$, 而 $x_n(m + \tau) = w'(n)x(n+m + \tau)$, 其中

$$w'(n) = \begin{cases} 1, & 0 \leq n \leq N + \tau - 1, \\ 0, & \text{其他} \end{cases} \quad (13)$$

则 $F_n(\tau)$ 修正为

$$F_n(\tau) = \sum_{m=0}^{N-1} |x_n(m) - x_n(m + \tau)| \quad (14)$$

在此基础上, 为避免由于信号强弱带来的误差, 对 AMDF 进行归一化处理: 定义改进的平均幅度差 $D(\tau)$ 如下:

$$D(\tau) = \frac{\sum_{m=0}^{\tau_{\max}-1} |x_n(m) - x_n(m + \tau)|}{2\tau \sum_{m=0}^{\tau_{\max}-1} |x_n(m)|}, \quad \tau = \tau_{\min}, \tau_{\min+1}, \dots, \tau_{\max} \quad (15)$$

其中 $\tau_{\min}$ 和 $\tau_{\max}$ 表示由基音频率决定语音段移动的最小和最

大数值。正常情况下,男性、女性的基音频率一般为 50~250Hz 和 100~500Hz,综合考虑,通常考察的语音信号的基音频率为 150~300Hz。若采样频率为 8kHz,则语音的基音周期为 27~54 个采样周期,因此可取 $\tau_{\min}=25$, $\tau_{\max}=55$ 。若每帧语音 80 个样点,则对应循环频率 $\omega \in [0.625\pi, 1.375\pi]\text{rad/s}$ 。当语音段在这个范围移动恰为该基音周期或周期的整数倍时, $D(\tau)$ 出现较小值。 $D(\tau)$ 值反映了语音信号的循环周期性,周期性越强,其值越小;反之,则越大。因此,单音的 $D(\tau)$ 值最小,语音的 $D(\tau)$ 值其次,噪声的 $D(\tau)$ 值最大。取 $D_0 < D(\tau) < D_{\text{th}}$ 时 τ 值,根据 $\omega = (2\pi/N) \cdot \tau$ 则可得到循环频率集 Ω 中的各元素。 D_0 为单音输入 $D(\tau)$ 上限,设定该门限可有效避免单频噪声干扰, D_{th} 为语音 $D(\tau)$ 上限,各门限取值见实验分析。

2.4 高阶循环累积量性质

高阶循环累积量具备以下主要性质^[7,8]:

性质 1 若随机变量 $\{x_i\}$ 与随机变量 $\{y_i\}$ 独立, λ_1, λ_2 是常数且 $s = \lambda_1 x + \lambda_2 y$ 则

$$C_{ks}(\omega) = \lambda_1 C_{kx}(\omega) + \lambda_2 C_{ky}(\omega) \quad (16)$$

性质 2 若 N 个随机变量 $\{x_i\}$ ($i=1, \dots, N$) 的一个子集同其他部分独立, 则

$$C_{kx}(\omega) = 0 \quad (17)$$

可以证明,高阶循环累积量理论上可以完全抑制任何平稳(高斯或非高斯)噪声和非平稳的高斯噪声,不论高斯噪声是平稳的还是非平稳的,甚至是与语音信号具有相同的循环频率的循环平稳的^[8]。下面通过语音信号模型从理论上验证基于循环累积量的信号可分离性。

如图 1 所示。

$$s(t) = x(t) + v(t), \quad t = 0, 1, \dots, N \quad (18)$$

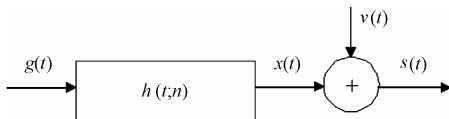


图 1 语音信号模型

Fig. 1 Model of speech signals

式中 $v(t)$ 是具有未知的任意功率谱密度,与 $x(t)$ 独立的平稳加性噪声或者高斯噪声。由循环累积量性质得到 $k \geq 3$ 时,

$$C_{ks}(\omega) = C_{kx}(\omega) \quad (19)$$

可见该模型下计算 $s(t)$ 的循环累积量等同于计算 $x(t)$ 的循环累积量。通常纯语音信号 $x(t)$ 服从时变滑动平均(MA)模型:

$$x(t) = \sum_{n=0}^q h(t; n)g(t-n) \quad (20)$$

式中 $h(t; n)$ 为时变滑动平均(MA)参数, $g(t)$ 是由声道特性决定的时变函数。当发送清音时, $g(t)$ 类似于白噪声;当发送浊音时, $g(t)$ 是一准周期函数,同时为独立同分布过程,并且具有有界的 k 阶循环累积量 $\rho_{kg} \neq 0$, 此时得到时变累积量^[8]

$$c_{kx}(t; \tau) = \rho_{kg} \sum_{n=0}^{\infty} h(t; n)h(t + \tau_1; n + \tau_1) \cdots h(t + \tau_{k-1}; n + \tau_{k-1}) \quad (21)$$

若 $h(t; n)$ 是一个严格周期 T_0 的函数,则可证 $c_{kx}(t; \tau)$ 是周期函数;若 $h(t; n)$ 是一个 MA(q) 过程,则可证 $c_{kx}(t; \tau)$ 是准周期函数,因此必定存在 $C_{kx}(\omega)$ ^[8]。可见高阶循环累积量的方法可作为语音检测的依据。

3 算法设计

以上分析均假设 $x(n)$ 为零均值随机信号,在应用中 $x(n)$ 常为非零均值的时间序列,此时可通过减去均值估计变成零均值,即采用 $x'(n)$ 替代 $x(n)$ ^[8]:

$$x'(n) = \{x(n) - E[x(n)]\} \quad (22)$$

在循环频率估计的基础上,构造了一种计算区域高阶循环累积量的函数:

$$\hat{M}_{3x}(\omega) = \frac{1}{N'} \sum_{t=0}^{N'-1} L_{3x}(t; \tau_1, \tau_2) e^{-j\omega t} \quad (23)$$

式中滞后积 $L_{3x}(t; \tau_1, \tau_2) = x'(3t)x'(3t+1)x'(3t+2)$; $t=1, \dots, N'$; N' 为 $N/3$ 的下取整,表示规整后的子窗口总数;滞后 $\tau_1=1$, $\tau_2=1$ 。取定义检测变量

$$\text{VAD}_x = \lg \left[\sum_{\omega \in \Omega} \left| \text{Re}(\hat{M}_{3x}(\omega)) \right| \right] \quad (24)$$

可认为 VAD_x 是一种改进的循环频率加权的广义相关函数,它表征了语音信号的循环平稳性特征。将各子窗口高阶循环累积量实部的累加和作为检测量的方法可极大降低算法复杂度,在三阶应用时降低约 67%。为验证该变量的有效性,选择一段原始质量等级为 5,包含清音、浊音和非语音段(加噪后为纯噪声段)的中文语音添加高斯噪声(信噪比 10dB),每帧语音为 80 个样点,统计 100 帧。根据循环频率估计得到 11 个频点,因此该语音源具有多循环周期性。统计各帧 VAD_x 的谱线见图 2。

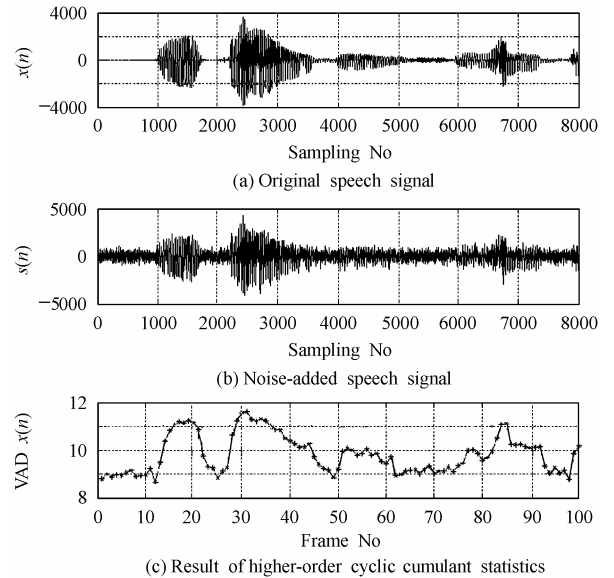


图 2 高阶循环累积量统计实例

Fig. 2 Higher-order cyclic cumulant statistics example

如图2所示,该谱线基本按语音信号的包络变化,其中噪声帧的 VAD_x 值较小,因此采用局部阈值化的方法来实现VAD算法。基于以上分析设计VAD算法流程图见图3。此外,为避免误判语音中的间隔和咳嗽等突发性干扰,设定前、后方保护函数。规定当且仅当连续多帧信号检测无语音时才发出无语音判决,即前方保护;当且仅当连续多帧信号检测有语音时才发出有语音判决,即后方保护。前方、后方保护帧数可依据现场实验调整,此处暂取1。为避免前、后向保护带来的截止效应,解码语音恢复时结合CNG技术。对图2加噪语音段的检测结果见图4,其中横轴为测试帧数,纵轴为检测标志位结果,“1”表示语音。

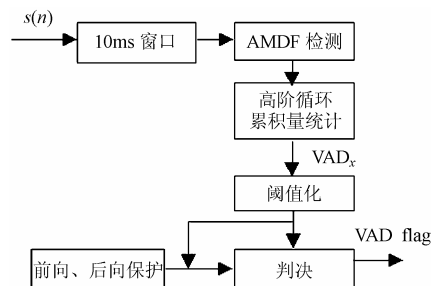


图3 改进型VAD算法框架
Fig.3 Block diagram of improved VAD algorithm

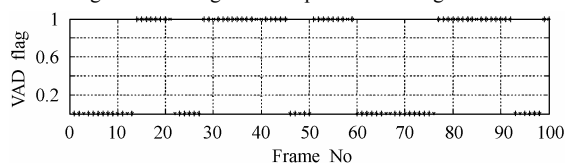


图4 语音检测结果
Fig.4 Result of VAD

4 实验分析

本文提供的VAD算法已经应用于我们的VoIP测试平台。该平台基于定点DSP(TMS320VC5509),采用144MHz时钟。语音数据采集采用12位A/D芯片,采样率8000Hz,每帧语音数据80个点,代表10ms语音段。DSP与采样芯片接口采用DMA乒乓方式工作来保证数据的实时性。首先,完成平均幅度差的门限确定工作。语音的数据采用现场采集作为数据源,单音信号采用300Hz正弦信号源输入,通过Matlab生成噪声及噪声与语音混合信号,然后在CCS2.0上搭建的测试环境使用。取每段的 $D(\tau)$ 序列的最小值 $D_{\min}(\tau)$ 作为统计语音段的表示值,最终得到以下实验结果如图所示(横轴是仿真帧数,纵轴是最小平均幅度差值(采用Q14定标,即该数值为表示值/ 2^{14}))。

从图5~图8 4幅图中可以看到,单频信号的平均幅度差值最小,在139左右,几乎为零,非零原因是A/D采集误差引起;其次是纯语音信号,其值大致在1151~6015之间变化;语音加噪的平均幅度差值在3481~11000之间变化;随机噪声在8319~10930之间变化。一般情况下,在硬件仿真中得到的平均幅度差都比软件仿真中大,可以根据实际情况进行适当地调整。本系统中采用的语音信号平均幅度差值范

围为2128~16384,去定标值范围0.130~1,其上下限分别对应 D_0 和 D_{th} 。通过软、硬件仿真和VoIP平台现场测试,该门限对于中文男、女语音均具备较好的自适应能力。

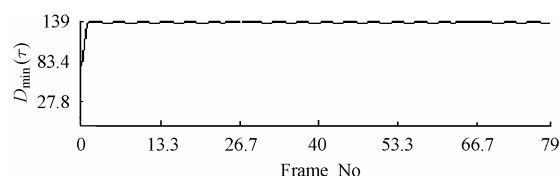


图5 单频信号的平均幅度差值
Fig.5 AMDF value of monochromatic signal

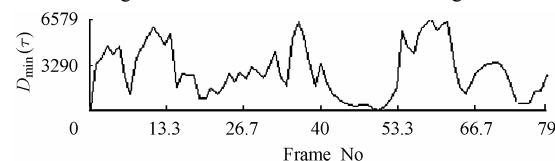


图6 语音信号的平均幅度差值
Fig.6 AMDF value of speech signal

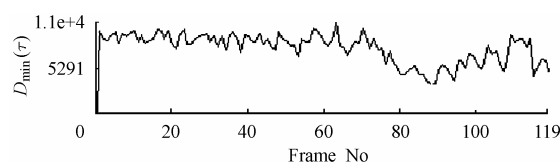


图7 语音加噪的平均幅度差值图
Fig.7 AMDF value of noise-added speech

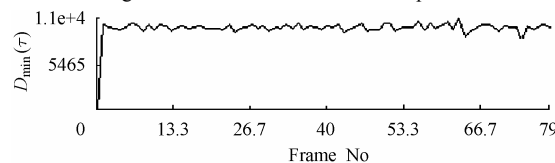


图8 随机噪声的平均幅度差值
Fig.8 AMDF value of random noise

其次,为验证算法的有效性,在CCS2.0环境下进行改进型VAD算法实验。作为例子,实验中采用原始语音的质量等级为5的汉语女声独白语音源并添加有色Gauss噪声、指数噪声等。信噪比分别是10, 5, 0。前方、后方保护帧数均设为5帧(50ms),为了进一步比较最后的实验结果,引入文献[3]定义的定量测试参数并参照G.729BVAD算法对同一语音源的检测结果来评估该算法的综合性能。在为了更直观表达,第1项采用处理时延来代替,这些参数包括:

处理时延(Processing delay): 算法应用在TMS320VC5509上的处理时间。代表算法的实时性指标。为了满足实时性要求,DSP处理时延必须小于10ms,更小的时延表示单位时间内将有更多通道的语音被处理和传输。

压缩率(Percentage compression, 简称Compression): 判断为无效的语音帧与总帧数的比值。

错误率(Percentage misdetection, 简称Misdetection): 语音帧被判作无效帧以及无效帧被判作语音帧的总和与语音帧总和的比值。

主观评价语音质量: 即MOS(Mean Opinion Score)因子,采用5级表示,其中1表示最差,5表示最好。

性能完好的 VAD 算法必须满足低时延、高压缩率、低误码率和好的语音质量等特点。下面在表 1 和图 9 至图 12 中给出测试比较结果。

表 1 处理时延
Tab.1 Processing delay

算法	CPU 时钟数	平均时延 (ms)
G.729B VAD	298806	2.074
New VAD	449891	3.122

测试结果表明, 算法时延均值为 3.122ms, 远小于 10ms 的语音帧处理限定时间范围, 因此具备双路语音实时并行处理的潜力。在高斯噪声环境下, 算法在信噪比 10dB, 5dB, 0dB 条件下压缩比与 G.729B VAD 相比分别提高 30%, 29%, 30%, 尤其是当低信噪比条件下依旧获得较高的压缩比; 而误检率则分别改善 9%, 7%, 9%; 语音质量方面对高斯噪声改善效果明显(见图 12)。对非对称噪声(以指数噪声为例)的检测效果较差, 压缩比分别达到 43%, 40%, 11%, 误检率则分别为 9%, 9%, 22%, 在语音质量方面则在高信噪比条件下改善效果明显, 对低信噪比改善效果不明显, 但同等条件下, G.729B VAD 算法基本失去检测效果, 其在图 12 中标识的信噪比为 0dB 时的语音质量是未检测的带噪声语音的质量, 5dB 时仍未有改善。总之, 本算法在高斯噪声和平稳噪声条件下具备较好的检测性能, 在非平稳噪声下也具备一定的检测效果。

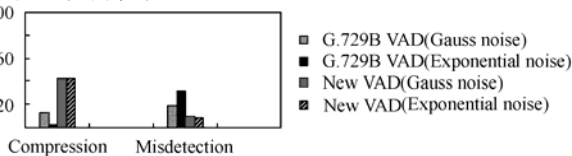


图 9 信噪比 10dB 检测结果比较

Fig. 9 Comparison of VAD for SNR = 10dB

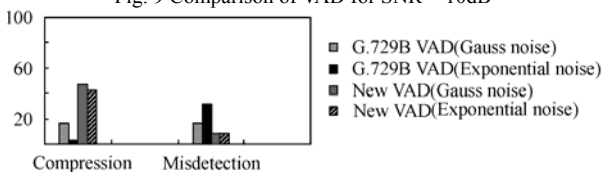


图 10 信噪比 5dB 检测结果比较

Fig. 10 Comparison of VAD for SNR = 5dB

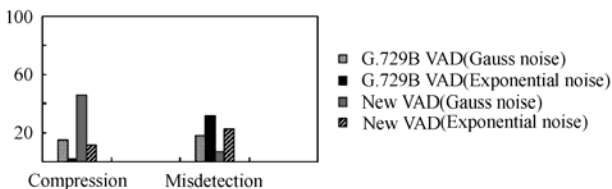


图 11 信噪比 0dB 检测结果比较

Fig. 11 Comparison of VAD for SNR = 0dB

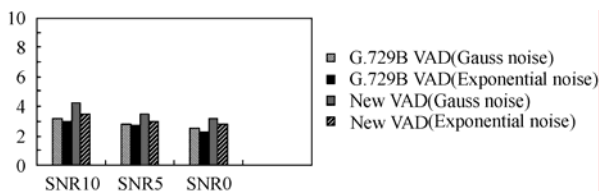


图 12 不同信噪比(分贝)的语音质量比较

Fig.12 Comparison of MOS for different SNR(dB)

5 结束语

应用循环累积量分析的语音检测算法着眼于语音信号的循环平稳特征, 算法的难点在于循环频率的估计和降低算法的复杂度。本文提出以语音段的基音周期作为循环频率的思想, 并采用修正的平均幅度差方法估计循环频率, 同时为降低算法复杂度对循环累积量的估计方法进行简化, 前向、后项保护策略可有效消除错误判决, 也有利于在保守与激进检测策略中寻找平衡。实验结果表明, 本文中提出的 VAD 算法较好地兼顾了实时性和性能要求, 采用的循环频率估计方法简捷, 算法实现了对不同背景平稳加性噪声和高斯噪声的自适应检测, 对于其他非平稳噪声也具备一定的检测能力。在 DSP 平台上实现该算法简洁、可靠, 综合性能优于 G.729 B VAD, 是一种性能优良的替代算法。在移动 VoIP 系统中, 该 VAD 算法可以为半双工设备提供可靠的发送、接收选择指示。该算法的思想也可以推广至语音识别等领域。

参考文献

- [1] Beritelli F, Casale S, Ruggeri G. New speech processing issues in IP telephony. Proc. Of IEEE ICCT'2000, Beijing, China, 2000: 652-656.
- [2] Beritelli F, Casale S, Ruggeri G. Performance evaluation and comparison of ITU-T/ETSI voice activity detectors. Proc. of IEEE ICASSP'2001, SaltLake City, USA, 2001, (3): 1425-1428.
- [3] Prasad R, Sangwan A, Jamadagni H S, Chiranth M C, Sah R, Gaurav V. Comparison of voice activity detection algorithms for VoIP. Proc. of IEEE ISCC'2002, Taormina, Italy, 2002: 530-535.
- [4] Haighand J A, Mason J S. Robust voice activity detection using cepstral features. Proc. of IEEE TENCON'1993, Beijing, China, 1993: 321-324.
- [5] Hoyt J D, Wechsler H. Detection of human speech in structured noise. Proc. of IEEE ICASSP'1994, Adelaide, Australia, 1994: 237-240.
- [6] Rec.G.729 Annex B. A silence compression scheme for G.729 optimized for terminal conforming, Tech. Rep. ITU-T, Oct. 1996.
- [7] 张贤达. 时间序列分析——高阶统计量方法. 北京: 清华大学出版社, 1996: 356-380.
- [8] 张贤达, 保铮. 非平稳信号分析和处理. 北京: 国防工业出版社, 2001: 324-358.
- [9] Mendel J M. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. Proc. IEEE, 1991, 79(3): 278-305.
- [10] Nemer E, Goubran R, Mahmoud S. Robust voice activity detection using higher-order statistics in the LPC residual domain. Speech and Audio Processing, 2001, 9 (3): 217-231.
- [11] 姚天任. 数字语音处理. 武汉: 华中理工大学出版社, 1994: 14-24.

朱晓亮: 男, 1975 年生, 博士生, 研究方向为 VoIP、流媒体等。
杜旭: 男, 1969 年生, 副教授, 研究方向为流媒体、现代通信网理论。
杨宗凯: 男, 1964 年生, 教授, 博士生导师, 研究方向包括现代网络理论及其应用等。