

G.729B 中话音激活检测算法的改进及仿真^①

胡 斌 黄 冰 罗保钦 王 苗 (桂林电子科技大学 信息与通信学院 广西 桂林 541004)

摘 要: VoIP 中 G.729B 的话音激活检测算法(VAD)通常只是在高信噪比下才会有良好的检测效果,而在低信噪比下或者有背景噪声的情况下检测效果却不佳。为此详细研究了自适应多速率语音编码(AMR)中的 VAD2 算法以及将其引入到 G.729B 中的可行性,并在此基础上对 VAD2 做了一些优化处理。通过 Matlab 仿真,发现优化后的 VAD2 无论是在高信噪比下还是在低信噪比下都有着相对良好的判决效果,同时也大大缩短了检测时间。实验结果表明:优化后的 VAD2 可用于 G.729B 算法中。

关键字: VoIP; G.729B; 话音激活检测; 信噪比; 自适应多速率语音编码

Simulation and Improvement of VAD in G.729B

HU Bin, HUANG Bing, LUO Bao-Qin, WANG Miao

(Information & Communication College, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: The Voice Activity Detector(VAD) of G.729B in VoIP can only work well in the high SNR, while in the low SNR or with a strong background noise it works poorly. To tackle this problem, the VAD2 mechanism of AMR is researched in detail and the possibility of introducing it into the G.729B is discussed. At the same time, some optimization work is also done. Through Matlab simulation, the optimized VAD2 is proved better by comparing it with the VAD algorithm of G.729B in both the high and low SNR, and its detection time is shortened greatly. So the optimized VAD2 can be used in G.729B.

Keywords: VoIP; G.729B; VAD; SNR; AMR

1 引言

VoIP 又称 IP 电话或者互联网电话,其中 VAD 技术是其三大关键技术之一^[1]。现实中,VoIP 常用到的 VAD 是基于线谱频率、全带宽信号能量、低频带信号能量和过零率的传统检测方法(G.729B 中的 VAD)^[2],这种方法在高信噪比情况下检测效果较好,而在低信噪比下或者存在非平稳背景噪声的情况下检测效果就会急剧下降。因此为了更好地保障重构后的语音质量,在 VoIP 中应该有一种更为良好的 VAD 判决机制,该机制即使在低信噪比下也应该能够进行较为准确的端点检测。为此详细研究了自适应多速率语音编码中的 VAD2 算法以及将其引入到 G.729B 中的可行性,并在此基础上对 VAD2 算法做了一些优化处理工作。实验结果证明:VAD2 算法不仅可以引入到 G.729B 中去,而且优化后的 VAD2 在检测性能和检测时间上均

要比原方法和 G.729B 中的 VAD 算法优越的多。

2 VAD2 判决的基本原理

VAD2 是自适应多速率语音编码中的一种话音激活检测方法,其基本判决原理如下^[3]:

2.1 声音 metric 的求解

首先将高通滤波后的语音信号的每一帧(20ms)分成两个子帧(10ms),即每个子帧的采样点为 80。然后对每一子帧进行预加重处理、加窗处理、离散傅立叶变换得到每一子帧的频域信号 $G(k)$ 。最后通过对 $G(k)$ 的迭代来估计每一子帧的频段能量 $E_{ch}(m,i)$,其迭代公式为:

$$E_{ch}(m,i) = \max \left\{ \begin{array}{l} E_{min}, \alpha_{ch}(m) E_{ch}(m-1,i) \\ + (1 - \alpha_{ch}(m)) \frac{1}{f_H - f_L + 1} \sum_{k=f_L(i)}^{f_H(i)} |G(k)|^2 \\ 0 \leq i < N_c \end{array} \right\} \quad (1)$$

① 基金项目:广西自然科学基金(桂科自 0640167)

收稿时间:2009-10-05;收到修改稿时间:2009-11-07

式中频段个数 N_c 为 16, $f_L(i)$ 和 $f_H(i)$ 分别取为 0 和 15, 最小允许频段能量 E_{min} 为 0.0625, $a_{ch}(m)$ 为频段能量平滑因子, $m=1$ 时 $a_{ch}(m)$ 为 0, 在后继子帧中 $a_{ch}(m)$ 取为 0.45^[4]。声音矩阵 $V(m)$ 的求解是通过对量化后的信道信噪比索引 $\sigma_q(i)$ 进行查表求和得到的, 其中 $\sigma_q(i)$ 被固定限制在 0 与 89 之间的一个整数。

2.2 语音门限值的求解

语音门限值 $V_{th}(m)$ 的求解是通过对长时峰值信噪比索引 $SNR_q(m)$ 进行查表得到的, 其中 $SNR_q(m)$ 被固定限制在 0 与 19 之间的一个整数, 而 $SNR_q(m)$ 的求解需要用到长时峰值信噪比 $SNR_p(m)$, 其迭代公式为:

$$SNR_p(m) = SNR = 56 - 10 \log_{10} \left(\sum_{i=0}^{N_c-1} E_n(m, i) \right) \quad m \leq 4 \quad (2)$$

$$SNR_p(m) = \begin{cases} 0.9 SNR_p(m-1) + 0.1 SNR \\ \quad (SNR > SNR_p(m-1)) \\ 0.998 SNR_p(m-1) + 0.002 SNR \\ \quad (0.625 SNR_p(m-1) < SNR \leq SNR_p(m-1)) \\ SNR_p(m-1) \\ \quad (otherwise) \end{cases} \quad m > 4 \quad (3)$$

(3)式中为瞬时帧信噪比, 其求解公式为:

$$SNR = 10 \log_{10} \left(\frac{1}{N_c} \sum_{i=0}^{N_c-1} 10^{\sigma_q(i)/10} \right) \quad (4)$$

2.3 语音偏差因子的求解

此外为了消除非平稳背景噪声和信号波动对 VAD 的影响, 我们还需要求解语音偏差因子 $\mu(m)$, 其求解公式为:

$$\mu(m) = \max \{ 12(\psi(m) - 0.65), 0 \} \quad (5)$$

式中 $\psi(m)$ 是变异因子。当 $m \leq 4$ 时, $\psi(m)$ 为零。当 $m > 4$ 时, $\psi(m)$ 为:

$$\psi(m) = \begin{cases} 0.99\psi(m-1) + 0.01 SNR^2 & (SNR(m) < 0) \\ \psi(m-1) & (otherwise) \end{cases} \quad (6)$$

2.4 频段噪声能量的更新和 VAD 判决

频段噪声能量 $E_n(m, i)$ 的更新是通过检测谱偏移量 $\Delta_E(m)$ 、频段能量总和 $E_{tot}(m)$ 、声音矩阵 $V(m)$ 、正弦波信号 $\sin wave_flag$ 、长时预测增益 β 等五个条件来综合判断的。即如果输出的更新标志 $update_flag$ 为 1, 则更新 $E_n(m, i)$, 否则不予更新^[4]。 $E_n(m, i)$ 的更新

公式为:

$$E_n(m, i) = \max \{ E_{init}, E_{ch}(m, i) \} \quad 0 < i < N_c \quad m \leq 4 \quad (7)$$

$$E_n(m+1, i) = \max \left\{ \begin{aligned} &E_{min}, \alpha_n E_n(m, i) + (1 - \alpha_n) E_{ch}(m, i) \\ &0 < i \leq N_c \quad m > 4 \end{aligned} \right\} \quad (8)$$

式中频段噪声能量的初始值 E_{init} 为 0.0625, 频段噪声平滑因子 α_n 取为 0.9。

VAD2 的判决是通过比较 $V(m)$ 、 $V_{th}(m)$ 与 $\mu(m)$ 之和的大小来进行的。即如果 $V(m) \geq V_{th}(m) + \mu(m)$, 则判决该子帧为语音帧, 否则判决为静音帧。一帧信号中只要有一个子帧被判决为语音帧, 则判决这一帧信号为语音帧^[3]。

2.5 VAD2 接入到 G.729B 中的可行性分析

从上述所有公式可以看出 VAD2 的判决是按照子帧的更新速度来进行的, 而且其划分子帧的参数是采样频率为 8000, 一子帧长度为 10ms。这正好与 G.729B 算法中划分一帧的参数是一致的, 也就是说用 VAD2 算法划分的子帧个数不仅与用 G.729B 算法划分的语音帧个数一样多, 而且还存在着——对应的关系。因此在 VoIP 环境中, 我们可以直接把 VAD2 对每一子帧的判决结果作为 G.729B 中 VAD 对每一帧的判决结果, 以此来区分语音帧和静音帧。即如果 VAD2 对某一子帧的判决结果为 1, 则用 G.729B 的编码方法来编码这一部分语音, 否则用 G.729B 的舒适噪声技术来编码这一部分语音^[5]。这就为在 G.729B 中利用 VAD2 算法来进行端点检测提供了前提条件。

3 VAD2 算法的优化

由公式(1)可以看出, 当 N_c 划分为 16 时, 这相当于将整个频谱分量划分为 16 个互不重叠的临界带, 各个频段的带宽和人耳的巴克临界带宽大致相近, 然而这些频段并没有覆盖全部的频谱分量^[4]。因而在求解语音能量公式 $\sum_{k=f_L(i)}^{f_H(i)} |G(k)|^2$ 中只是求解了人耳听觉范围内的语音能量而漏掉了人耳以外的一部分语音能量, 而且当频谱分量划分为 16 时, 计算量也相当的复杂。结合实际情况, 我们应该求解全部频谱内的语音能量, 并以此来重新估计频段能量。其方法就是用频域变换过后的 $G(m)$ 来直接求解语音能量, 然后用此语音能量来重新估计 $E_{ch}(m)$ 。这样做不仅可以节省一部分计算量, 而且重新估计得到的 $E_{ch}(m)$ 还包含了所有频段内的语音能量, 这将有利于增强 VAD 判决的正

确性,此外实际情况也比较符合。另外需要对所有和 N_c 有关的数学表达式都要进行修正,如频段能量 $E_{ch}(m,i)$ 的数学表达式则需要修正为:

$$E_{ch}(m) = \max \left\{ E_{min}, \alpha_{ch}(m) E_{ch}(m-1) + (1 - \alpha_{ch}(m)) \cdot |G(m)|^2 \right\} \quad (9)$$

原 VAD2 算法中含有大量的对 $E_n(m,i)$ 的更新判断程序,考虑到现实中的语音总是会受到噪声的干扰,从而会导致原始语音信号中夹带很多正弦波信号分量,因此在 VAD 判决中,对 *sinewave_flag* 的判决结果在大多数情况下都会为 1,从而导致对 *update_flag* 的判决为 1^[6],详细判断过程见参考文献[6]。所以在 Matlab 仿真中我们就直接设置 *update_flag* 为 1,这样我们就可以去掉一些不必要的更新判决程序,从而可以减小代码的执行时间。虽然同时导致多更新估计了一部分频段噪声能量,但是多更新的那一部分频段噪声能量是很少的,其对 VAD2 的影响也是微乎其微的,因而直接设置 *update_flag* 为 1 是可以接受的。

另外原算法的窗函数为矩形窗函数。我们知道加窗的目的是为了使主瓣更尖锐,旁瓣更低,防止或者减少主瓣的频率泄漏到旁瓣上。而矩形窗的主瓣宽度比同样窗口长度的汉宁窗、海明窗的主瓣宽度小,旁瓣能量却比后两者大。因此在本实验中,窗函数选的是性能较好的汉宁窗。

4 实验结果及分析

为了评价优化后的 VAD2 算法的检测效果,我们分别进行了在高信噪比下和带背景噪声的低信噪比下的实验。本实验条件为:语音采样频率为 8000;一帧的采样点为 160,分为两个子帧;窗函数为 Hanning 窗,长度为 80;FFT 变换长度为 128;高信噪比下的语音文件为 185 帧的语音信号,低信噪比下的语音文件为 125 帧的语音信号。其中图 1、图 2、图 3 为高信噪比下的检测结果;图 4、图 5、图 6 为低信噪比下的检测结果;表 1 为三种算法各自的语音漏检率和误检率;表 2 为三种算法分别运行 6 次所需要的运算时间。

通过比较可以看出,在高信噪比下,优化后的 VAD2、VAD2、G.729B 的 VAD 对语音文件的判决结果基本一致。在语音漏检率方面,甚至 G.729B 的 VAD 算法还要更好一点。

在低信噪比下,G.729B 的 VAD 判决结果全为 1,语音误检率很大,而优化后的 VAD2、VAD2 在语音误检方面就要比其好的多。这是因为 G.729B 的 VAD 检测过程可分为两步。其第一步为:VAD 模块从输入的语音信号中提取 4 个特征参数,这 4 个特征参数分别是线谱频率、全带宽信号能量、低频带信号能量和过零率,然后分别计算这些特征参数与其长时平均值的差值,最后根据这些特征参数的差值做出初步静音判决。第二步:对初步静音判决的结果进行平滑,使平滑后的判决结果更加地符合语音信号中存在的长时静音现象,同时也使得语音帧和静音帧编解码器之间的切换更为自然些。但是从第一步中我们可以看出,在 G.729B 的 VAD 判决中设置的语音门限值(平均值)是固定不变的,而第二步只是一个平滑过程,其对最终的 VAD 判决结果影响不大。而 VAD2 是按照每一子帧的更新速度来自适应地更新语音门限值的,也就是说 VAD2 的语音门限值是时变的。因而很显然,对于那些低信噪比下的语音文件,由于其过零率和背景噪声能量往往比较大,所以当用 G.729B 的 VAD 算法时就会很容易产生语音误检现象。而 VAD2 由于采用了量化查表的方法来得到声音 *metric* 和语音门限值,然后再进行 VAD 判决,所以用该方法就不存在误检率很大的问题。另外在语音漏检率方面,优化后的 VAD2 要略好于原算法,但是对于那些完全淹没在噪声中的语音部分,无论是优化后的 VAD2 还是原 VAD2 算法则都无法识别出来,这种现象是和人耳的屏蔽效应是相一致的。

在算法的运行时间上,无论是在高信噪比下还是在低信噪比下,优化后的 VAD2 的运行时间均小于原 VAD2 和 G.729B 中 VAD 的运行时间。

下图分别为本次实验的结果和实验数据。

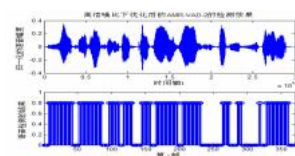


图 1 高信噪比下优化后的 VAD2 检测的结果

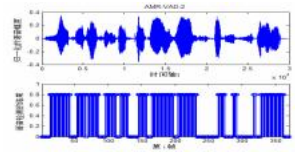


图 2 高信噪比下原 VAD2 检测的结果

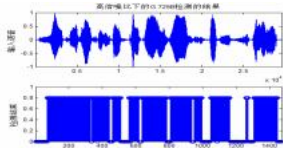


图 3 高信噪比下 G.729B 的 VAD 检测的结果

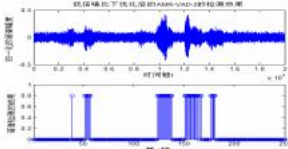


图 4 低信噪比下优化后的 VAD2 检测的结果

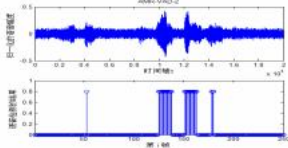


图 5 低信噪比下原 VAD2 检测的结果

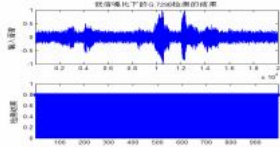


图 6 低信噪比下 G.729B 的 VAD 检测的结果

表 1 三种 VAD 检测结果的漏检率和误检率

		优化后的 VAD2	VAD2	G.729B 的 VAD
高信 噪比	漏检率	1.2%	1.2%	1.0%
	误检率	0%	0%	0%
低信 噪比	漏检率	4.9%	6.8%	0%
	误检率	0%	0%	62%

表 2 三种 VAD 算法各自运行 6 次所需要的运算

	高信噪比下		低信噪比下	
G.729B 的 VAD	0.9910s	0.9810s	0.6010s	0.5810s
	0.9410s	0.8610s	0.5310s	0.7510s
	0.8810s	0.8210s	0.6610s	0.6710s
VAD2	0.5810s	0.5100s	0.5310s	0.5100s
	0.5400s	0.5610s	0.5510s	0.4900s
	0.5500s	0.5810s	0.5110s	0.5100s
优化后的 VAD2	0.4810s	0.3800s	0.4310s	0.4000s
	0.4100s	0.3910s	0.3510s	0.3600s
	0.3700s	0.4710s	0.4210s	0.3300s

5 结论

本文主要研究了 VAD2 算法及其在 G.729B 中应用的可行性,并在此基础上对 VAD2 进行了一些优化处理。通过 Matlab 仿真,结果证明:经过优化处理后的 VAD2 算法无论是在语音检测的性能上还是在算法的运行时间上均要比 G.729B 算法的 VAD 优越,因而其可用在 G.729B 算法中。

参考文献

1 Colline D.舒华英,李勇等译. VoIP 语音技术与应用. first ed., 北京:人民邮电出版社, 2000. 1-18.

2 刘丙娥,张为宁,毕文青. G.729B 编解码算法在 DM642 上的实现与优化.自动化表, 2007,28(10):48-49.

3 3GPP TS 26.094:AMR Speech Codec; Voice Activity Detector, 2009.

4 王炳锡,王洪.变速率语音编码.西安:西安电子科技大学出版社, 2004. 161-167.

5 张拥军. VoIP 中语音压缩编解码器的研究与实现. 安徽电子信息职业技术学院学报, 2007,6(31):102-103.

6 3GPP TS 26.073: Adaptive Multi-Rate (AMR); ANSI C source code, 2009. © 中国科学院软件研究所 <http://www.c-s-a.org.cn>