

一种基于小波的改进型带噪语音激活检测算法*

李强,黄冰,王苗

(桂林电子科技大学 信息与通信学院,广西 桂林 541004)

摘要: 提出了一种基于小波的改进型语音激活检测方法。在介绍小波变换、增强型变速率语音编码(EVRC)算法中的语音增强模块原理的基础上,着重讲述了利用小波对语音信号进行分解重构,利用语音增强模块中的判别方法对语音端点进行检测,依据判决更新对检测结果进行修正。通过对不同信噪比下的语音信号进行端点检测,给出了仿真结果及语音帧检测率,证明了该算法的有效性和可靠性。

关键词: 小波变换;语音增强;端点检测;低信噪比

中图分类号: TP301.6

文献标识码: B

A modified effective noisy voice activity detection based on wavelets

LI Qiang, HUANG Bing, WANG Miao

(Department of Computer Science and Technology, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In this paper, a modified voice activity detection based on wavelets is proposed. Firstly, it describes the principles of wavelets transform and the speech enhancement of EVRC. Then, the decomposition and reconstruction of speech signal used sym5 wavelet is emphasized and combined to speech enhancement model. According to the update mechanism of pre-judgments, an amendment to the results will be executed at the end of the algorithm. At last, speech signal is detected by this algorithm to get the result of simulation under different signal-to-noise ratios. And the result indicate the effectiveness of this algorithm.

Key words: wavelets transform; speech enhancement; voice activity detection; low signal-to-noise ratio

在人类的交流中,语音与非语音是同时存在且交替出现的。语音激活检测算法,就是对输入信号中语音/非语音加以区分,同时对语音的起止端给出准确判断的一类方法。该方法在语音信号处理领域中一直有着重要的意义。在语音编码领域中,准确的语音检测可以在一定程度上降低编码的平均比特率;而在语音识别系统中,检测结果的准确与否直接影响到最终的识别效果。

传统的端点检测方法采用短时能量、短时过零率等时域特征,它们在背景噪声较低的情况下,可以准确地找到语音的起止端。但随着信噪比的下降,其检测性能急剧恶化。

如果可以将低信噪比下语音信号的大体轮廓恢复出来,再对此近似信号进行语音端点检测,检测的准确率必然会得到一定的提升。因此,在心理声学模型的基础上,引入了小波分解重构来获取低信噪比下语音信号

的近似信号。

1 小波变换原理及相关小波介绍

小波变换是时间(空间)频率的局部化分析,它通过伸缩平移运算对信号(函数)逐步进行多尺度细化,最终达到高频处时间细分,低频处频率细分,能自动适应时频信号分析的要求,从而可聚焦到信号的任意细节^[1]。

对任意在实数域下满足平方可积条件的函数 $f(t)$,其连续小波变换定义如下:

$$(W_{\psi}f)(a,b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt = \langle f, \psi_{a,b} \rangle$$

其中, $\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$ 是一个窗口由尺度因子 a 和平移因子 b 共同决定的小波函数。

利用小波变换对信号进行分解、重构,需要将尺度因子 a 和平移因子 b 离散化,即

* 基金项目: 广西科学基金(0640167)

技术与方法 Technique and Method

$$\begin{cases} a = a_0^j \\ b = k a_0^j b_0 \end{cases}$$

将上式代入 $\psi_{a,k}(t)$ 可得到离散小波变换的母小波,其定义式为:

$$\psi_{j,k}(t) = a_0^{-\frac{j}{2}} \psi\left(\frac{t - k a_0^j b_0}{a_0^j}\right)$$

相对应的离散小波变换的定义如下:

$$(W_{\psi}f)(j, k) = \int_{-\infty}^{\infty} f(t) \overline{\psi_{j,k}(t)} dt$$

对信号进行离散小波变换,其效果相当于使用了一组正交镜像滤波器(QMF)对信号进行高通滤波和低通滤波。对信号进行三层小波分解,其原理图如图1所示。

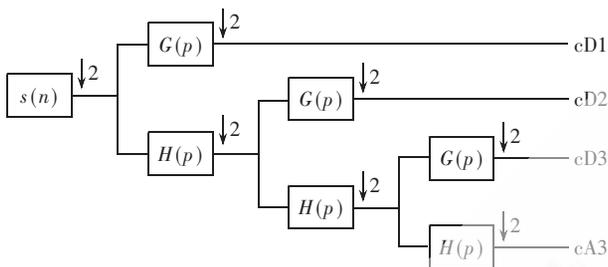


图1 对信号 $s(n)$ 进行三层小波分解

在图1中,输出的 cD1、cD2、cD3、cA3 分别代表第1层细节系数、第2层细节系数、第3层细节系数、第3层近似系数。

sym5 小波是 Symlets 小波函数族中的一个小波,它是由 Daubechies 提出的近似对称的小波函数,是对 db 函数的一种改进,它具有正交性、双正交性、紧支撑性,可以精确实现重构,能够更好地对谐波进行分析。函数 symn 中的 n 代表了这个小波函数的消失矩。消失矩越大,它的支撑长度就越大,对应的滤波器也就越平坦,这意味着小波函数的振荡会很强。小波消失矩的大小,决定了小波逼近光滑信号的能力,即越大的消失矩将使高频系数越小,小波分解后的能量也将越集中。但是,紧支撑性也是小波函数好坏与否的一个关键因素,而它的变化趋势与消失矩恰恰相反,即大的消失矩对应的小波紧支撑性差,小的消失矩对应的小波紧支撑性好。所以,实际中会根据需要选择合适的小波函数,以期达到实验的最佳效果^[2]。

2 EVRC 语音增强模块中语音帧的判别方法

在增强型变速率语音编码(EVRC)算法中^[3],语音增强模块包含在预处理部分内,其目的是为了减小噪声对有用信号的影响,尽可能地提高含噪语音的质量。而语音增强的首要工作就是对输入信号帧语音的判别,其处理过程如图2所示。

具体实现步骤^[4-5]如下:

(1) 预处理

对输入语音信号进行分帧处理(帧长为160,帧移为

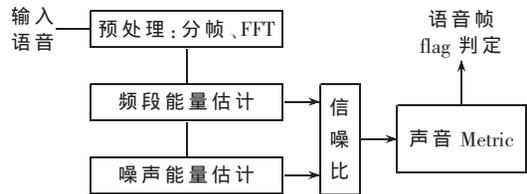


图2 语音增强模块中语音判别子模块处理框图

80),并对每帧信号进行离散傅立叶变换(DFT),变换长度 $M=256$ 。

(2) 频段能量估计

将频带划分为16个互不重叠的频段,然后计算第 m 帧第 i 个频段内的能量,其计算式可表示为:

$$E_{ch}(m, i) = \max\{E_{min}, a_{ch}(m)E_{ch}(m-1, i) +$$

$$(1 - a_{ch}(m)) \frac{1}{f_u(i) - f_l(i) + 1} \sum_{k=f_l(i)}^{f_u(i)} |G(k)|^2\} \quad 1 \leq i \leq N_c$$

其中 $E_{min}=0.0625$ 是允许的最小频段能量, $a_{ch}(m)$ 是信道能量平滑系数, $N_c=16$ 为频段总数。 $m=1$ 时, $a_{ch}(m)=0$, 其余各帧 $a_{ch}(m)=0.45$ 。第 i 帧的下限频率 $f_l(i)$ 和上限频率 $f_u(i)$ 如表1所示。

表1 各频段对应的频谱分量起止端

段号	0	1	2	3	4	5	6	7
f_l	2	4	6	8	10	12	14	17
f_u	3	5	7	9	11	13	16	19
段号	8	9	10	11	12	13	14	15
f_l	20	23	27	31	36	42	49	56
f_u	22	26	30	35	41	48	55	63

(3) 噪声估计及更新

噪声的估计按下式进行:

$$E_n(m-1, i) = \max\{E_{min}, a_n E_n(m, i) + (1 - a_n) E_{ch}(m, i)\} \quad 1 \leq i \leq N_c$$

式中 $E_{min}=0.0625$, $a_n=0.99$ 是频段噪声平滑因子。

对前4帧频段噪声能量估计应初始化为频段能量估计:

$$E_n(m, i) = \max\{E_{ini}, E_{ch}(m, i)\} \quad 1 \leq m \leq 4, 1 \leq i \leq N_c$$

式中 $E_{ini}=16$ 是允许的最小频段噪声初始化能量。

(4) 频段信噪比估计

利用下式计算信噪比的量化值:

$$\sigma_q(m, i) = \max\{0, \min\{89, \text{round}\{10 \lg \left| \frac{E_{ch}(m, i)}{E_n(m, i)} \right| / 0.375\}\}\} \quad 1 \leq i \leq N_c$$

其中, σ_q 的值被限制在0和89之间。

(5) 声音 Metric 计算

第 m 帧的声音 Metric 值可由下式得到:

$$v(m) = \sum_{i=1}^{N_c} V(\sigma_q(i))$$

其中 $V(k)(k=\sigma_q(i))$ 是长度为90的声音 Metric 表 V 中的第 k 个元素。

技术与方法 Technique and Method

$$V = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 \\ 3 & 3 & 3 & 4 & 4 & 4 & 5 & 5 & 5 & 6 & 6 & 7 & 7 \\ 7 & 8 & 8 & 9 & 9 & 10 & 10 & 11 & 12 & 12 & 13 & 13 & 14 \\ 15 & 15 & 16 & 17 & 17 & 18 & 19 & 20 & 20 & 21 & 22 & 23 & 24 \\ 24 & 25 & 26 & 27 & 28 & 28 & 29 & 30 & 31 & 32 & 33 & 34 & 35 \\ 36 & 37 & 37 & 38 & 39 & 40 & 41 & 42 & 43 & 44 & 45 & 46 & 47 \\ 48 & 49 & 50 & 50 & 50 & 50 & 50 & 50 & 50 & 50 & 50 & 50 & 50 \end{bmatrix}$$

(6)端点检测

根据声音 Metric 设定的门限值,若 $v(m) > 34$ 时,语音帧标志 $flag$ 置 1; 否则,认定该帧不是语音帧, $flag$ 置 0。

3 基于小波的改进型语音激活检测方法

EVRC 语音增强模块的话音判别方法对于高信噪比的语音信号可以给出较为准确的判决,但当信噪比下降时其性能也逐渐降低。出现的问题主要是由于噪声能量的不断上升,某一连续的语音段被分裂成若干部分检测出来,造成连续语音的断裂。

语音信号作为非平稳信号,其频率会随着时间的变化而不断改变。其中变化较慢的部分,对应着信号的低频成分,反映了信号的大体轮廓;而变化较快的高频成分,代表了信号的细节^[6]。基于此,可以利用小波先将低信噪比下加噪信号中的语音部分近似地重构,然后利用 EVRC 语音增强模块中判别方法对语音帧进行判决。这样可以在一定程度上降低了语音出现断裂。算法流程图如图 3 所示。



图 3 基于小波—VAD2 语音激活检测算法流程

首先,利用 sym5 小波对输入的语音信号进行三层小波分解。其次,利用第三层分解得到的近似系数对原始信号进行单支重构,得到原始信号的近似波形。两者的对比效果可由图 4 看出。其中上面的信号是信噪比为 0 dB 的染噪波形,下面是经过单支重构后的近似波形。依据 EVRC 语音增强模块的话音判别原理,对由小波重构的近似信号进行语音帧

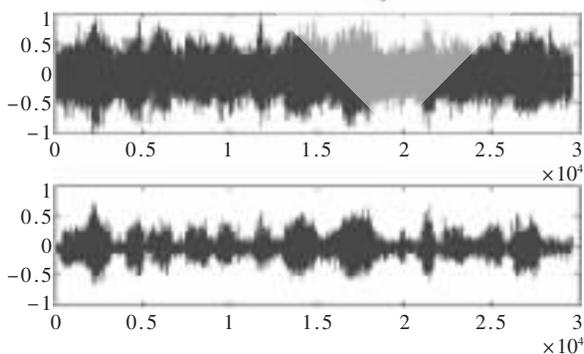
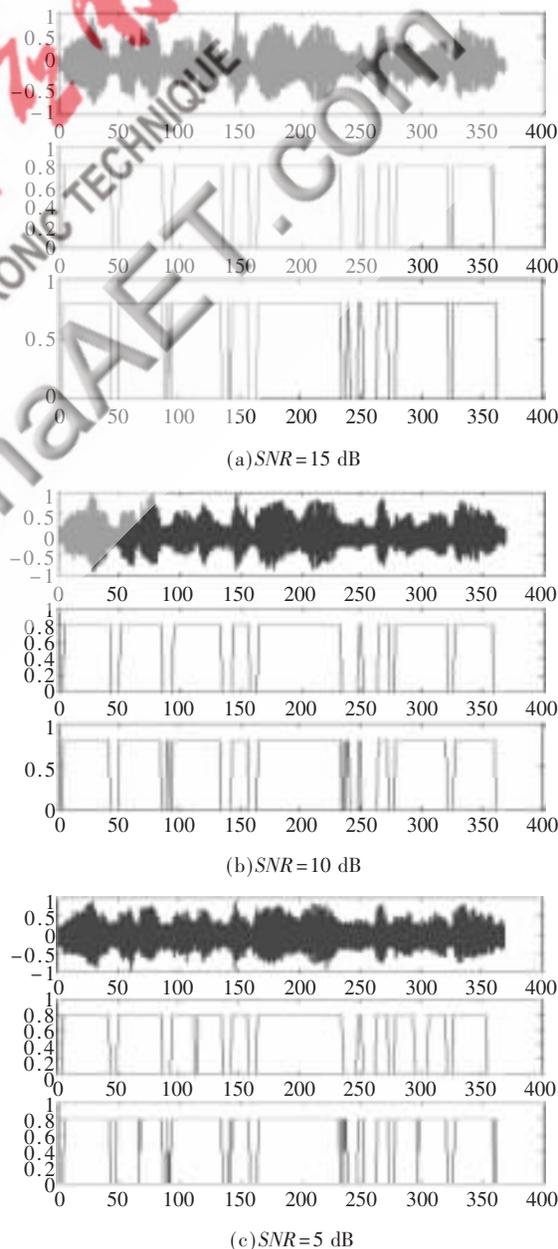


图 4 SNR=0dB 的加噪文件经过小波分解重构后的波形

预判决。这里用到的阈值是硬阈值,为 34。最后,对预判结果进行更新。更新算法的具体过程是,搜索所有帧的语音预判决结果,其中少于 3 帧的非语音段,根据前后帧的判决结果,选择是否将该段的语音标志位 $flag$ 置 1;而对少于 4 帧的语音段,也需根据前后帧的判决结果,选择是否将语音标志位 $flag$ 置 0。

4 仿真结果及分析

首先,利用 Windows 自带的录音软件录制语音,语音文本是“语音信号处理的一个重要应用是语音编码”;其次,利用 matlab 软件的 awgn 函数对录制的语音文件进行加噪处理,其 SNR 值分别为 15 dB、10 dB、5 dB、0 dB、-5 dB;最后,对生成的不同信噪比的加噪语音文件分别利用小波—心理声学模型和心理声学模型算法对端点进行检测。其不同 SNR 值对应的检测结果如图 5 所示。



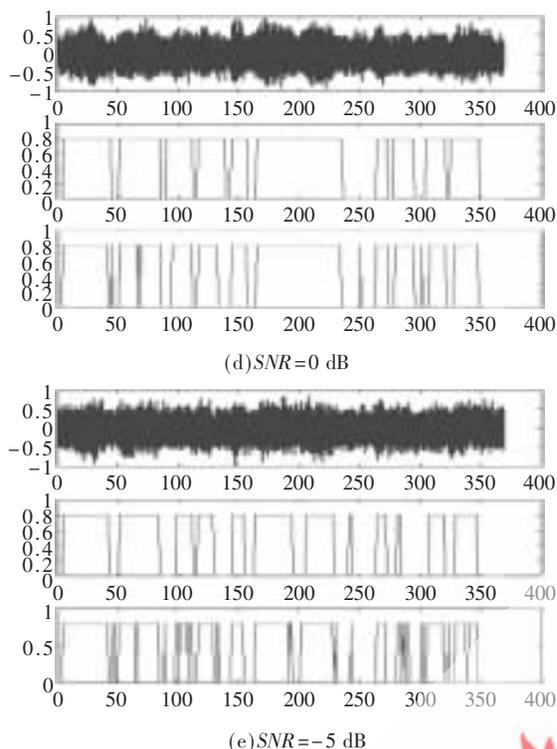


图5 不同信噪比下 endpoint 检测的结果对比图

首先,从图5的5个图中可以发现,EVRC语音增强模块的语音判别出的某些语音起止段有错误,例如(a)中第235帧判断的语音段,其实是前端语音的尾端,但却被算法隔断。出现问题的主要原因是由于判决门限采用的是硬阈值。

其次,由图5中(a)、(b)、(c)、(d)给出的结果可以看出,虽然信噪比不断下降,但改进型算法对端点的检测性能依然稳定。从图(e)给出的检测结果可以判断,此信噪比下算法的检测性能已经开始恶化,比如出现在250帧左侧的语音段,在其他4种信噪比下未被检测到,所以该语音段属于误判。

最后,图5(d)中的检测结果相对(a)、(b)、(c),差别在于250帧附近的语音段未被检测出来,其原因主要是该语音段包含的语音文本为“是”字,而录音者由于个人习

惯,读这个字时发音未到位。

综上所述,可以认定基于小波的改进型语音激活检测算法在不同信噪比下的检测性能较为稳定,尤其是在信噪比较低的情况下。其检测准确率如表2所示。

表2 语音帧检测准确率

SNR	15 dB	10 dB	5 dB	0 dB	-5 dB
基于小波的改进型算法/%	96.6	95.4	94.1	92.1	64.7
EVRC语音增强模块的判别算法/%	95.1	93.7	90.3	84.9	52.3

基于基于小波的改进型语音激活检测算法,可以动态地跟踪噪声的变化来确定语音的起止位,尤其是在低信噪比下依然可以保持较高的检测率,是一种较为可靠的端点检测方法。

参考文献

- [1] 丁宣浩.小波分析基本理论[M].北京:中国科学技术出版社,2007.
- [2] 葛哲学,沙威.小波分析理论与MATLAB R2007实现[M].北京:电子工业出版社,2007.
- [3] TIA/EIA/IS-127.Enhanced variable rate codec,speech service option 3 for wideband spread spectrum digital systems[S].January,1997.
- [4] 王炳锡,王洪.变速率语音编码[M].西安:西安电子科技大学出版社,2004.
- [5] 姚艳红,宋刚.基于EVRC的带噪汉语语音端点检测[J].计算机工程与应用,2007,43(35).
- [6] 乔飞,杨小军,马岸英.基于小波分析的多普勒弱信号检测方法[J].探测与控制学报,2008,30(1).

(收稿日期:2009-11-19)

作者简介:

李强,男,1982年生,硕士研究生,主要研究方向:语音信号处理的研究。

黄冰,男,1965年生,教授,主要研究方向:光通信和语音信号处理的研究。

王苗,男,1987年生,硕士研究生,主要研究方向:语音信号处理的研究。