

【自动化技术】

一种对单兵通讯模块中短时能量 语音端点检测的改进方法

吕卫强

(中国兵器工业第五八研究所 军品部, 四川 绵阳 621000)

摘要:现有的单兵通讯模块中的实时语音端点检测为短时能量加过零率的检测方法,在实际使用中发现该方法在低信噪比的环境下的识别率较低,针对该问题,在原来检测方法的基础上提出一种音频特征值交叉熵,根据三者阈值的综合判断来确定当前音频信号是否是语音信号。

关键词:语音端点检测;短时能量;过零率;交叉熵

中图分类号: TN914

文献标识码: A

文章编号: 1006-0707(2011)03-0059-02

语音端点检测算法也称语音、噪声分类算法。该方法在语音信号处理领域有着重要意义,单兵通讯模块中的语音端点检测是该方法的重要应用。当前单兵通讯模块中使用的是短时能量加过零率2个特征值来判断音频信号是否是语音信号,在使用中该算法在高信噪比的环境中可以获得较好的正确率,但在低信噪比系统中算法的性能会大幅度降低。为了解决这一问题,本文提出了一个针对原算法的改进算法,在原音频特征值的基础上加入了交叉熵这个特征值。该特征值在低信噪比环境下的语音信号的检测率较高,故而可极大地提升整个算法的性能。

1 算法描述

1.1 短时能量加过零率的计算

语音端点检测依靠分析每帧输入信号的特征值是否符合语音信号的特征来确认输入信号是否是语音信号。短时能量和过零率就是2个典型的语音信号的特征值,根据它们可以有效检测输入信号是否是语音信号。

对于输入的语音信号 $x(n)$,其中 n 为采样点,首先进行分帧操作。将语音分为 30 ms 一帧,计为帧长。相邻 2 帧起始点之间相隔 10 ms,计为帧移。相邻 2 帧之间有 20 ms 的重叠。对于不同的采样频率,帧长和帧移对应的采样点数不同,本文采用 8 kHz 的采样频率,30 ms 对应的帧长为 240 个采样点,计为 N 。10 ms 的帧移对应 80 个采样点,计为 M 。则对于第 i 帧,第 n 个采样点,它与原始语音信号之间的关系为:

$$X_i(n) = X[(i-1)M + N]$$

第 i 帧语音信号的短时能量可用下述公式计算:

$$e(i) = \sum_{n=1}^N |X_i(n)|$$

将分帧信号的短时能量计算出来后,设定一个门限值,超过这个门限值的信号被认为是语音信号。这样即可实现语音端点检测。但是仅凭这一个特征判断信号是语音信号

不是十分可靠。人的语音分清音和浊音 2 种:浊音为声带发出,对应语音信号的幅度高,周期特性明显,采用短时能量可以有效检测;清音则不会有声带的震动,仅凭空气在口腔中的摩擦、冲击或爆破发声,其短时能量一般较小,仅用短时能量法容易漏检。通过分析清音信号的波形发现,在清音阶段,由于口腔空气摩擦的效果,导致波形在幅度上的变化比较剧烈,故引入一个新的特征来描述这种在一帧信号中波形穿越零电平的次数变化剧烈程度的参数,称之为过零率,用如下公式描述:

$$ZCR(i) = \sum_{n=1}^{N-1} |X_i(n) - X_i(n-1)|$$

为避免静音阶段的随机噪声产生过高的过零率,设定一个门限值,当前后两个采样值符号不同,并且差值大于这个门限时,才将过零率的数值加一。

1.2 交叉熵的计算

针对上述检测的音频帧,在计算上述 2 个特征值的同时计算该音频帧的交叉熵特征值。计算方法如下:

1.2.1 子带能量计算

首先对该音频帧作 256 点的快速 Fourier 变换,将 0 ~ 4 000 Hz 的全频段分为 8 个频段,每 16 个点分为一个频带,计算每一个频带的能量:

$$S_i = \sum_{k=16, i=15}^{16i} R_{i,k}^2 \quad (1)$$

其中 R_k 是对应子带的相应得 Fourier 变换的第 K 个幅度值。相应子带在总带中占有的比例为

$$p_i = s_i / \sum_{j=1}^8 S_j \quad (2)$$

1.2.2 交叉熵的计算

设随即变量 X 可能的取值为 $\{a_1, a_2, \dots, a_k\}$,且 X 的概率分布情况与 H_1, H_2 有关。在假设 H_1 下, X 的概率分布为

$$\begin{pmatrix} X \\ P_1(X) \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \dots & a_k \\ p_1(a_1) & p_1(a_2) & \dots & p_1(a_k) \end{pmatrix} \quad (3)$$

在假设 H_2 下, X 的概率分布为

收稿日期:2011-01-13

作者简介:吕卫强(1975—),男,硕士,工程师,主要从事计算机软件开发研究。

$$\begin{pmatrix} X \\ P_2(X) \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \cdots & a_k \\ p_2(a_1) & p_2(a_2) & \cdots & p_2(a_k) \end{pmatrix} \quad (4)$$

假设 H_1 和 H_2 成立的概率分别为 $P(H_1)$ 和 $P(H_2)$, 根据全概率公式和条件概率公式有

$$p(H_1 | a_k) = \frac{p(H_1)p_1(a_k)}{p(H_1)p_1(a_k) + p(H_2)p_2(a_k)} \quad (5)$$

$$p(H_2 | a_k) = \frac{p(H_2)p_2(a_k)}{p(H_1)p_1(a_k) + p(H_2)p_2(a_k)} \quad (6)$$

其中: $p_1(a_k) = p(a_k/H_1)$, $p_2(a_k) = p(a_k/H_2)$ 。

由式(5)、(6)可得

$$l_b \frac{p_2(a_k)}{p_1(a_k)} = l_b \frac{p(H_2 | a_k)}{p(H_1 | a_k)} - l_b \frac{p(H_2)}{p(H_1)} \quad (7)$$

对数似然比 $l_b \frac{p_2(a_k)}{p_1(a_k)}$ 在假设 H_2 的期望就称为交叉熵,

记为 $I(p_2, p_1; X)$, 即

$$I(p_2, p_1; X) = \sum_{k=1}^K p_2(a_k) l_b \frac{p_2(a_k)}{p_1(a_k)} \quad (8)$$

1.3 算法流程

基于能量加过零率的端点检测算法都先假设开始 n 帧为噪声帧, 从而进行噪声能量的估计, 在后续的判断过程中, 若当前帧的能量高于噪声能量超过一个阈值, 便判断为语音帧, 否则, 判断为噪声帧, 并进行噪声能量的更新。这种端点检测算法在噪声突变的情况下会造成误判, 并且误判会一直持续下去。基于这种情况, 算法作了改进, 噪声能量不是仅仅在基于能量判断为噪声的帧中进行, 而是在所有的帧中都进行更新, 只不过更新的因子会根据交叉熵信息进行自适应调整, 这样可以很好地跟踪背景噪声的变化。

算法首先通过开始的 n 帧噪声估计噪声各子带能量的分布, 因为算法针对噪声环境建立, 考虑到安静环境下算法的实用性, 若噪声估计帧的能量小于某个阈值则在每个信号帧中(包括噪声帧)中叠加少量的白噪声, 从而保证算法的可行。

算法中, 交叉熵的计算采用子带能量分布代替概率分布, 用来衡量当前帧频谱分布于噪声频谱分布的相似程度。 H_1 假设当前帧为噪声帧, H_2 假设当前帧为语音帧。交叉熵用来表示当前帧偏离噪声的程度。式(9)中的 $p_1(a_k)$ 表示噪声帧第 k 个子带的能量在总能量中的比例, 即

$$p_1(a_k) = \frac{N_k}{\sum_{i=1}^8 N_i} \quad (9)$$

式(10)中的表示语音帧第 k 个子带的能量在总能量中的比例, 即

$$p_2(a_k) = \frac{S_k}{\sum_{i=1}^8 S_i} \quad (10)$$

其中: N_i 代表噪声帧中第 i 个子带的能量; S_i 代表语音帧中第 i 个子带的能量。将式(9)、(10)代入式(8), 可计算出当前帧与噪声帧的鉴别信息, 表示当前帧偏离噪声帧的程度。

n 帧过后, 每接受一帧, 首先求该帧各子带能量的分布以及总能量, 若总能量超过噪声帧的能量达到一定的域值, 则将当前帧判断为语音帧, 同时计算该帧信号与噪声帧信号的交叉熵, 交叉熵反映了该帧信号类似噪声的程度, 因此根据交叉熵可以进行噪声能量估计的更新。

2 试验结果

基于能量加过零率的语音端点检测算法和本算法在白噪声下的语音检测正确率试验结果对比如表 1 所示。

表 1 试验结果对比

信噪比/ dB	白噪声下的语音检测正确率/%	
	能量加过零率的算法	本算法
30	98.48	98.56
15	89.12	96.41
0	67.49	89.28

根据表 1 可以看出基于能量加过零率的语音端点检测算法在高信噪比环境下效果良好, 但随着信噪比的降低, 算法检测的正确率大幅下降。但在加入交叉熵特征值后, 语音检测的正确率在低信噪比的环境下依旧能够保持稳健。

3 结束语

采用能量加过零率特征来进行端点检测, 必须能够自适应地调整门限域值, 否则在噪声能量急剧变化的环境中误识率会很大, 而交叉熵作为衡量信号与噪声相似度的特征, 单独用来进行端点检测时会因为信号帧与噪声帧的交叉熵特征相交区域较大, 失识率会较大(二者在单独运用时, 各有其不足, 联合运用时, 可以优势互补)。能量加过零率特征是日前用来进行端点检测的特征中最为普遍也最好的, 但是为了能够较快地跟踪背景噪声的变化, 门限需要不断更新。在传统的端点检测算法中, 只在噪声帧进行更新, 如果噪声能量突然升高或者降低, 则这种方法的检测结果会一直错下去。交叉熵具有与噪声幅度变化无关的特性, 利用交叉熵, 可以根据当前帧与噪声帧的相似度, 在所有的帧中进行门限能量更新, 从而及时地跟踪噪声能量的变化, 即使在其发生剧烈变化时, 仍能很快地适应。利用能量进行端点检测时, 个别噪声帧会被判断为语音帧, 利用交叉熵, 即使该帧能量超过门限, 也可以根据其噪声的相似程度进行综合判断, 从而可以在一定程度上减少这种错误。使用 3 个特征值综合判断在各种信噪比下可以有效提升检测准确率。

参考文献:

- [1] 何强, 何英. matlab 扩展编程[M]. 北京: 清华大学出版社, 2002.
- [2] 李荣祥, 马春庭, 程翔, 等. 一种新的小波阈值法在语音信号处理中的应用[J]. 四川兵工学报, 2008, 29(3): 45-47.
- [3] 肖沙里, 高家利. 日盲紫外光语音通信系统的设计与实现[J]. 重庆理工大学学报: 自然科学版, 2010, 24(1): 51-54.
- [4] 张卫宁. DSP 原理与应用教程[M]. 北京: 科学出版社, 2008.
- [5] 陈怀深. 数字信号处理与教程 - MATLAB 释义与实现[M]. 北京: 电子工业出版社, 2008.

(责任编辑 刘 舸)