

# 一种强噪声环境下的语音端点检测算法

蔡萍

(闽江学院 物理学与电子信息工程系, 福建 福州 350108)

**摘要:** 语音识别中语音信号端点检测的准确度起着至关重要的作用, 传统的频带方差算法在强背景噪声环境下性能下降。提出一种改进算法, 在保证算法复杂度较低的前提下, 大大提高了检测的准确度。仿真证明, 该算法在低信噪比条件下也具有较鲁棒性。

**关键词:** 端点检测; 低信噪比; 频带方差

**中图分类号:** TP301.6

**文献标识码:** A

**文章编号:** 1671-5322(2015)03-0044-03

端点检测的作用是从连续采样的数字信号中确定语音段和噪声段的起止时刻<sup>[1]</sup>。语音识别的一个关键问题就是如何将语音信号精确地检测出来, 以减少运算量及提高效率<sup>[2]</sup>。在高信噪比条件下, 传统的端点检测算法, 如基于短时能量和过零率、能量谱方差、倒谱距离和熵的算法已经得到很好的运用; 在低信噪比条件下, 这些算法的性能急剧下降, 有的甚至无法工作<sup>[3]</sup>。如何在强背景噪声环境下准确地进行端点检测已成为一个棘手的问题。本文首先描述了传统的频带方差检测算法, 在此基础上进行改进, 提出一种新的语音段起止端点检测算法, 最后通过具体实验得出结论。

## 1 端点检测的基本步骤

语音信号端点检测的基本步骤为:

(1) 分帧。把语音信号分割为相互重叠的若干帧。由于语音信号具有短时平稳性, 即语音信号在 10~30 ms 的时间内可以看做是平稳过程, 所以在帧长的选取上也是基于这个原则。

(2) 提取特征向量。对每一帧语音, 计算其某一种特征向量。比如在能量判决法中, 每帧信号的能量或能量的对数被视为判决特征。

(3) 判决。根据(2)中提取的特征向量序列, 利用一定的判决规则, 判断该帧信号是语音段还是噪声段。

(4) 后处理。后处理过程的作用是对(3)得出的结果进行修正, 最终得出语音信号和噪声信

号的检测结果。在步骤(3)中, 可能会把人在正常发声过程中的自然停顿判为噪声, 这个停顿包括句子内的停顿, 甚至是词内的停顿, 这些错误判决要进行纠正。

## 2 频带方差及改进算法

频带方差, 指的是短时频带方差, 因为系统是个时变系统。它的实质就是计算一帧信号各个频带能量的方差。语音和噪声的频带方差相差很大。假设噪声是高斯白噪声, 之所以称其为白噪声, 是因为在噪声的频谱特性中, 各个频带的能量相差不大, 或者说各频带之间变化很平缓。而语音信号是有色的, 各频带之间变化较剧烈。根据这一特征, 可以很明显地区分语音和噪声<sup>[4]</sup>。

定义矢量

$$Y = \{y_1(\omega_1), y_2(\omega_2), \dots, y_N(\omega_N)\}$$

$\omega_1, \omega_2, \dots, \omega_n$  表示带通滤波器的中心频率。 $y_i(\omega_i)$  表示数字信号通过中心频率为  $\omega_i$  的滤波器的输出能量。数字信号最低频是 0, 最高频是  $\pi$ , 于是带通滤波器的带宽为  $\pi/N$ , 中心频率从 0 到  $\pi$  等步长递增。 $y_i(\omega_i)$  的计算方法: 计算一帧信号的 FFT (Fast Fourier Transform, 快速傅里叶变换), 然后把几个相邻频率分量组合而得<sup>[5]</sup>。

计算  $Y$  的平均值

$$E = \frac{1}{N} \sum_{i=1}^N y_i(\omega_i)$$

则

$$D = \frac{1}{N} \sum_{i=1}^N [y_i(\omega_i) - E]^2 \quad (1)$$

定义为频带方差。

通常可认为一段语音信号的前几帧是噪声帧。以  $n=5$  为例,取前 5 帧计算频带方差,计算 5 个频带方差的均值  $D_n$ 。取门限值  $G=2D_n$ ,超过这个门限值判断为噪声帧,否则为语音帧。

对传统的频带方差算法进行以下几点改进,既减小了计算量,又提高了检测的准确度。

(1)将平方运算转化为绝对值运算。

(2)在频域上取相邻 5 个点的值求平均进行平滑。与传统的相邻 3 点平滑相比,语音段的检测参数没有太大的变化;而在噪声段,假定噪声是白噪声,本身能量谱的起伏就不大,加上 5 点平滑后起伏就更小了,从而拉开了语音帧和噪声帧的差距,使门限的设定更加简单。

(3)仅取前 50 个点的能量谱进行计算。假设对每帧进行 256 点的离散傅里叶变换。语音信号的能量主要集中在低频部分。从谱特征上看,也是低频部分的震荡幅度大,高频部分的震荡幅度小。仅取前半部分参与计算,可大大简化计算量。

(4)能量谱间隔 5 个点取差值,即

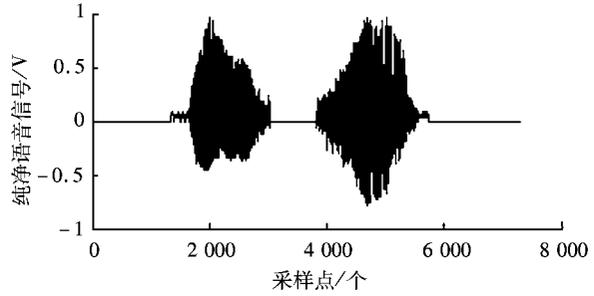
$$D = \frac{1}{45} \sum_{i=6}^{50} |y_i(\omega_i) - y_{i-5}(\omega_{i-5})| \quad (2)$$

这一改进措施同样是为了拉开语音帧与噪声帧的检测参数。对噪声帧来说,相邻 2 点的差与相邻 5 点的差差别不大,而语音帧差别极大。

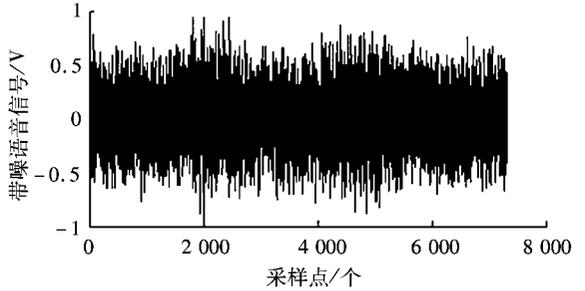
### 3 仿真结果与分析

在 MATLAB 环境下进行仿真实验。声音信号样本由男、女生发声,在实验室安静环境下以声卡采集,纯净语音信号采样率为 8 kHz,量化位数为 16bit,以 wav 格式存储。噪声采用平稳高斯白噪声。纯净语音和高斯白噪声以 -5 dB 的信噪比进行混合,形成带噪语音。

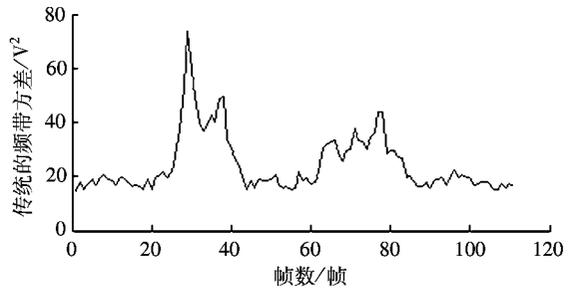
图 1a 是归一化后的纯净语音信号,图 1b 是 -5dB 的带噪语音信号,图 1c 与图 1d 是根据传统算法即公式(1)计算得出的频带方差以及根据该曲线设定门限后得到的检测结果,图 1e 与图 1f 是根据改进算法即公式(2)计算出的参数和判定结果。从图 1 可以看出,用传统算法计算的检测参数在语音段和噪声段的差别不明显,而改进算法语音段的检测参数大大高于噪声段,因此后者对判决门限的灵敏度低。



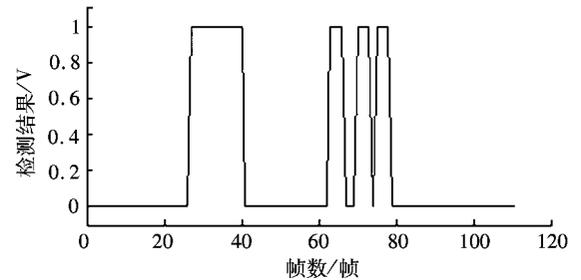
a 纯净语音信号



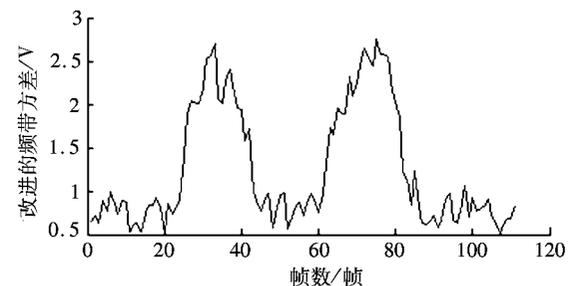
b 带噪语音信号



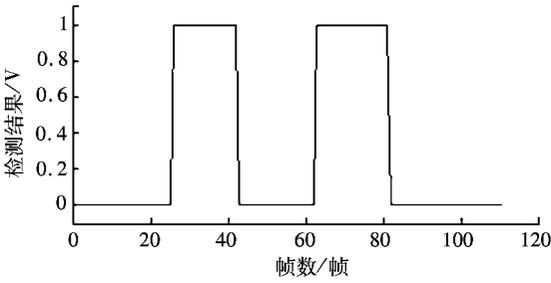
c 传统算法计算得的频带方差



d 传统频带方差检测结果



e 改进算法的频带参数



f 改进算法的检测结果

图 1 传统算法与改进算法的仿真结果

Fig.1 Simulation result of traditional algorithm and improved algorithm

从实验结果可以看出,传统的频带方差检测方法在低信噪比的条件下性能不佳,一个发音音节内判决出既有语音段又有噪声段的结果;改进后的算法具有较强的鲁棒性。这是因为在低信噪比条件下,噪声的能量较强而语音的能量较弱,传统的频带方差检测算法在语音段和噪声段的检测参数差距不大,因而对门限选择的要求极高,否则会出现误判的结果;而改进后的算法一是增加了平滑,使噪声段的检测参数明显下降,二是用相邻 5 点的能量谱进行差分,提高了语音段的检测参数,使检测结果泾渭分明,提高了算法的准确度。

参考文献:

[1] 席大林,李如玮,陈海龙. 基于自相关最大值和过门限率的语音端点检测[J]. 语音技术,2010,34(4):53-57.  
 [2] 董胡. 一种改进的基于倒谱距离端点检测算法研究[J]. 计算机与数字工程,2013,41(7):1 078-1 080.  
 [3] 尹岩岩,殷业,罗汉文,等. 基于短时能零熵的端点检测方法[J]. 计算机仿真,2012,29(11):408-411.  
 [4] 王忠禹. 现代谱估计[M]. 南京:东南大学出版社,1990:35-36.  
 [5] 李祖鹏,姚佩阳. 一种语音段起止端点检测新方法[J]. 电讯技术,2000,20(3):68-70.

## Speech Endpoint Detection Algorithm Based on Strong Noise Environment

CAI Ping

(Department of Physics & Electronic Information Engineering, Minjiang University, Fuzhou Fujian 350108, China)

**Abstract:** The accuracy of speech recognition in speech signal endpoint detection plays an important role. Traditional frequency band variance algorithm gives a bad performance in strong noise environment. In ensuring a lower computational complexity conditions, Put forward an improved algorithm, which can greatly improve the accuracy of detection. The simulation results show that the algorithm has strong robustness under low SNR condition.

**Keywords:** endpoint detection; low SNR; frequency band variance

(责任编辑:李华云)