

基于小波变换的语音段起止端点检测算法^{*}

董力, 陈宏钦, 马争鸣

(中山大学电子与通信工程系, 广东 广州510275)

摘要: 提出一种基于小波变换的语音段起止端点检测算法。传统的语音段起止端点检测算法大都是在时域上根据能量累积的大小判别语音段和噪声段, 这些算法只是适用于高信噪比的情况。对于低信噪比的情况, 这些算法往往要借助平均过零率等辅助特征。这样做不但增加了算法的复杂度, 而且也未必奏效。单音节或浊辅音汉字的平均过零率与噪声的平均过零率大致相当。根据小波变换的特性, 针对主要由白色噪声组成的噪声背景, 提出一种新的语音段起止端点检测算法。这种算法根据白色噪声在小波变换域各个子带的平均能量变化平缓的特点判别语音段和噪声段。实验结果表明, 算法即使在低信噪比的情况下也能正确判别语音段和噪声段。

关键词: 小波变换; 语音处理; 白色噪声

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0529-6579(2005)03-0116-03

语音段起止端点检测是语音分析、语音合成和语音识别中的一个必要环节。传统的语音段起止端点检测算法大都是在时域上基于能量累积的大小来判别语音段和噪声段。这些算法通常假设语音的能量总是大于噪声的能量^[1]。但是, 在实际应用中, 语音的能量随着音素的变化而变化。例如, 当语音段的开始和末尾都是弱摩擦音(如“四”字音)时, 其短时能量就比较小。以鼻音结尾的语音段, 其末尾的短时能量也比较小。它们都容易与噪声混淆。

本文根据小波变换的特性, 针对主要由白色噪声组成的噪声背景, 提出一种新的语音段起止端点检测算法。这种算法根据白色噪声在小波变换域各个子带的平均能量变化平缓的特点判别语音段和噪声段。本文提供实验结果表明, 本文提出的算法即使在低信噪比的情况下也能正确判别语音段和噪声段。

1 基于小波变换的语音段起止端点检测算法

传统的语音段起止端点检测算法都属于时域算法, 利用语音和噪声的时域特征检测语音段的起止端点。其实, 在许多情况下, 语音和噪声的频谱特性差异很大, 可以利用语音和噪声的频域特征检测语音段的起止端点。例如, 在一般情况下, 背景噪声可以认为是高斯白噪声, 其频谱是“白色”的, 而语音信号的频谱则肯定是有色的。利用频谱的“颜色”之分, 可以很明显地区分噪声和含噪语音。

一个信号 $x(m)$ 经过多次小波变换后分解为 N 个不同频段的子带信号:

$$x_1(m), x_2(m), \dots, x_N(m)$$

这些信号是时域信号, 对它们做FFT可以得到它们的DFT:

$$X_1(k), X_2(k), \dots, X_N(k)$$

如果 $x(m)$ 是白色信号, 则这些DFT都是同一个常数 C 。根据能量守恒, 有

$$\sum_{m=1}^{N_i} x_i^2(m) = \sum_{k=1}^{N_i} X_i^2(k) = N_i C^2 \Rightarrow$$
$$E_i = \frac{1}{N_i} \sum_{m=1}^{N_i} x_i^2(m) = C^2$$

这里 N_i 是第 i 个子带信号 $x(m)$ 的长度, $i=1, 2, \dots, N$ 。我们称 $E_1, E_2, \dots, E_N$ 为信号 $x(m)$ 的小波子带平均能量序列。进一步, 我们可以计算这个能量序列的均值和方差:

$$E = \frac{1}{N} \sum_{i=1}^N E_i \quad (1)$$

$$D = \frac{1}{N} \sum_{i=1}^N |E_i - E|^2 \quad (2)$$

显然, 如果信号 $x(m)$ 主要由白色噪声组成, 那么, 由于噪声的“白”谱性质, 这个能量序列应该近似为一个常数序列, 其方差近似为零。如果信号 $x(m)$ 主要由语音组成, 那么, 这个能量序列是一个数值变化的序列, 其方差不会近似为零。如果信号 $x(m)$ 主要由白色噪声和语音组成, 那么, 不论白色噪声能量有多大, 由于语音的存在, 这个能量序列依然是一个数值变化的序列, 其方差不会近似为零。因此, 我们可以根据方差的大小来区分

* 收稿日期: 2004-07-19

基金项目: 广东省科技计划资助项目(2004B10101031); 中山大学重点建设高水平大学专项资金资助项目

作者简介: 董力(1970年生), 女, 研究生; 通讯联系人: 马争鸣; E-mail: mzmzm@pub.guangzhou.gd.cn

噪声帧和含噪的语音帧，从而检测语音段的起止端点。

在语音段起止端点检测的实际应用中，如果 x (m) 是需要检测的输入信号，则我们令这里 w (m) 是取样窗。然后，我们对 x_n (m) 进行上述的基于小波变换的端点检测，判断 x_n (m) 是噪声还是含噪语音。

2 实验结果

图 1 表示对含噪语音信号处理的结果，第一个波形是语音波形（横坐标是离散信号的点数，纵坐标是幅值大小，下同）；第二个波形是噪声波形；第三个波形是语音加噪声（含噪语音）波形；第四个波形是语音的小波子带平均能量序列（横坐标是子带的数目，纵坐标是能量大小，下同）；第五个波形是噪声的小波子带平均能量序列；第六个波形是含噪语音的小波子带平均能量序列。通过计算，语音能量序列的均值是 0.832 27，方差是 1.571 01，方差较大，说明语音是“有色”信号；噪声能量序列的均值是 0.037 66，方差是 0.000 36，均值较小，说明信噪比较高，方差也较小，说明噪声是白色噪声；含噪语音能量序列的均值是 0.825 03，方差是 1.695 23。语音信号的能量为 11.03，噪声信号的能量为 1.50，信噪比为 $S/N=7.34$ 。

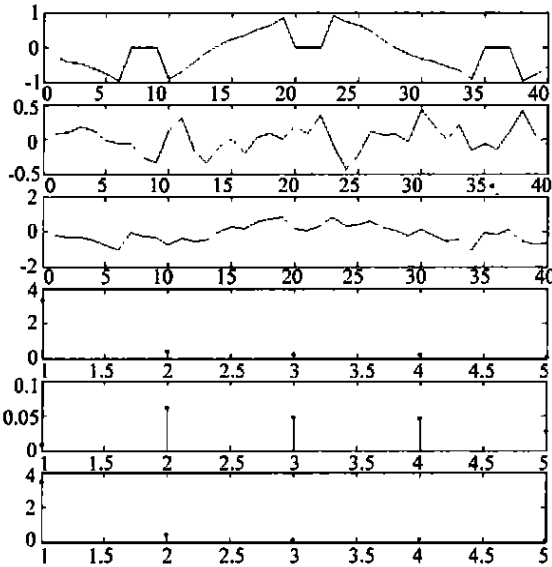


图 1 含噪语音信号（信噪比为： $S/N=7.34$ ）

Fig. 1 Noised voice signals (SNR: $S/N=7.34$)

图 2 表示对语音和噪声信号处理的结果。图中第一个波形是语音波形；第二个波形是噪声波形；第三个波形是语音加噪声（含噪语音）波形；第四个波形是语音的小波子带平均能量序列；第五个波形是噪声的小波子带平均能量序列；第六个波形是含噪语音的小波子带平均能量序列。通过计算，语

音的能量序列的均值是 0.832 27，方差是 1.571 01；噪声的能量序列的均值是 0.271 66，方差是 0.005 90。含噪语音能量序列的均值是 0.987 90，方差是 1.731 01。语音信号的能量为 11.03，噪声的能量为 9.38；信噪比为 $S/N=1.17$ 。

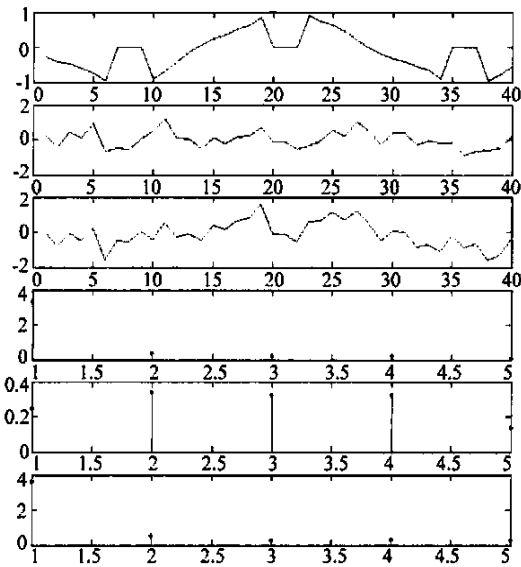


图 2 含噪语音信号（信噪比： $S/N=1.17$ ）

Fig. 2 Noised voice signals (SNR: $S/N=1.17$)

注：实验均采用 Matlab 小波工具箱，所采用的小波为 DB7，对信号进行 4 次小波分解，共产生 5 个子带。

3 结 论

(1) 本文提出的方法的主要思路是利用频域特征来区分语音和噪声。任何一种子带滤波的方法都可以利用这个思路达到区分语音和噪声的目的^[2]。但是，小波分解恐怕是最为快捷的子带滤波方法，很多现成的工具（如 Matlab 的小波工具箱）可供使用^[3-5]。另外，利用本文提出的方法确定信号为含噪语音信号后，可以接着进行基于小波分解的去噪处理，从而形成一套完全基于小波变换的语音识别预处理系统。

(2) 本文提供的实验表明，即使在很低的信噪比下（如），采用本文提出的小波子带平均能量序列的方差做为判别标准，可以轻易地判别出信号是否含有语音成分。如果采用通常的短时能量做为判别标准，在这样低的信噪比下，是绝对不可能判别出信号是否含有语音成分的。这时必然要采用诸如平均过零率等辅助判别标准。但是，这样做不但增加了算法的复杂性，而且，由于这些辅助判别标准往往都具有各自的局限性，实际应用中常常会得不偿失。例如，浊音的平均过零率与噪声的平均过零率大致相当。如果语音段是以单音节汉字或浊辅音

汉字（它们的时域特征均为浊音）开头或结尾，则平均过零率判别也无能为力。

参考文献：

[1] 李祖鹏, 佩阳. 一种语音段起止端点检测新方法[J]. 电讯技术, 2000(3): 68—71.
[2] 刘庆升, 徐霄鹏, 黄文浩. 一种语音端点检测方法的探

究[J]. 计算机工程, 2003, 29(3): 27—31.
[3] 董小刚, 秦喜文. 信号消噪的小波处理方法及 Matlab 实现[J]. 长春工业大学学报, 2003, 24(2): 13—15.
[4] 胡昌明. 基于 Matlab 的系统分析与设计: 小波分析[M]. 西安: 西安电子科技大学出版社, 1999.
[5] 陈怀琛, 吴大正, 高西全, Matlab 及在电子信息课程中的应用[M], (第二版). 北京: 电子工业出版社, 2003.

A Wavelet Based Algorithm for Detecting the Beginning and End Points of Voice Segments

DONG Li, CHEN Hong qin, MA Zheng ming

(Department of Electronics and Communication Engineering, Sun Yat sen University, Guangzhou 510275, China)

Abstract: A wavelet based algorithm for detecting the beginning and end points of voice segments was proposed. The common used algorithms are mostly based on their energy differences in time domain and therefore only suitable for the high SNR situation. For the low SNR, the common used algorithms have to be strengthened by making use of some additional features such as average across zero rates. This improvement will increase the algorithm complexity and still be possible to fail in some cases such as those Chinese characters with sonant consonants. A new wavelet based algorithm has been proposed for the situation with white noise environment. White noise environment can remain unchanged across various subbands of wavelet transform domain and therefore can be detected and removed from voice signal. The experimental results show that this new algorithm can efficiently distinguish the beginning and end points of voice segments even in the low SNR environments.

Key words: wavelet transform; voice signal process; white noise

(上接第 115 页)

An Improvement to Brands E-cash Scheme

LI Jin, LIN Qun, WANG Yan ming

(School of Mathematics, Sun Yat sen University, Guangzhou 510275, China)

Abstract: It is suggested that the bank just directly use different private and public key pairs to issue different denominations cash in previous e-cash schemes. An efficient multi denomination off-line e-cash scheme is presented, in which the bank only needs one private key other than n different private keys when issuing n kinds denominations e-cash. The new scheme is more efficient than previous multi denomination schemes.

Key words: e-cash; multi denomination; restrictive blind signature