

一种组合参数的语音信号清/浊音判决方法

郭英, 李雪娇, 李宏伟
(军工程大学 电讯工程学院, 陕西 西安 710077)

摘要:提出一种基于组合参量的语音信号清/浊判决方法,减小了清/浊单信号特征概率分布的重叠区域,程序设计简练。经计算机仿真表明,与单一参量判决方法相比,其错判率小,抗噪性好。

关键词:语音信号;清/浊单;判决

中图分类号:TN911.7 **文献标识码:**A **文章编号:**1009-3516(2002)04-0019-03

在语音信号处理中,对输入信号进行清音/浊音(U/V)的判别时是很重要的。其判决的复杂和准确度对后续的语音处理有很大影响。常用的U/V判决法多采用语音信号的短时能量、短时平均过零率及线性预测编码(LPC)参数等作为判别的参量。由于单一参量值的动态范围在各分类之间通常会发生叠接,只用一种特征来作为语音信号的清/浊判决依据,容易发生错判。所以要取得高识别率,需要采用参量组合,如新近制定的G.729B^[1]标准中对语音的U/V判别就是采用4种参数(能量、过零率、线谱频率及低通能量)的组合进行的。但组合参量判决法会增加判别复杂度。文中借鉴G.729B标准的作法,提出了一种基于三参数(能量、过零率、归一化自相关函数值)组合的语音信号清/浊判决方法,并同时得到浊音的基音周期。

1 三参量组合判决算法

语音信号能量随时间有相当大的变化,一般来说清音段的能量比浊音段的小得多。所以,语音信号的能量是判决清/浊的一个重要参数。计算能量 E_0 的公式如下:

$$E_0 = \sum_{n=0}^{N-1} S_{\omega}^2(n) \quad (1)$$

其中 $S_{\omega}(n)$ 是语音信号 $s(n)$ 的加窗信号。窗函数一般可用矩形窗或汉明窗, N 为窗长。在纯净语音环境下,采样率为8 kHz,帧长为30 ms(每帧包含240个样点)时经大量统计发现浊音的帧能量大于某固定值,而清音的帧能量则低于此数。

语音的过零率 Z_0 是表征一帧语音中语音信号波形穿过横轴(零电平)的次数。它可以用相邻两个取样改变符号的次数来计算,其公式如下:

$$Z_0 = \frac{1}{2} \left\{ \sum_{n=0}^{N-1} |\operatorname{sgn}[S_{\omega}(n)] - \operatorname{sgn}[S_{\omega}(n-1)]| \right\} \quad (2)$$

$\operatorname{sgn}(n)$ 为符号函数。清音的过零率较高,浊音的过零率较低,当采样率为8 kHz,帧长为20 ms(每帧包含160个样点)时清音的平均过零率约为70,浊音约为20。

语音的短时自相关函数 $R_{\omega}(l)$ 对于清/浊音判决来说也是一个重要参数,计算公式为

$$R_{\omega}(l) = \sum_{n=0}^{N-1-l} S_{\omega}(n) S_{\omega}(n+l) \quad (3)$$

若语音信号是周期性信号,则语音的短时自相关函数 $R_{\omega}(l)$ 也是周期性信号,且其周期与信号周期相同。由这一特性可以通过计算语音的短时自相关函数来判断语音的清/浊性(在语音模型中清音信号视为随机信号,浊音信号则被视为周期信号),并同时可判决出浊音信号的基音周期。

基于这三种参数在清/浊音情况下的条件概率密度函数交叠小,本文提出一种基于三参数(能量、过零

率、归一化自相关函数峰值)组合的两级判决算法。其具体作法是:

1) 将语音信号 $s(n)$ 进行分帧, 30 ms (每帧包含 240 个样点) 为一帧, 加矩形窗, 并作预处理 (包括预加重及低通滤波) 得到每帧信号 $S_w(n)$ 。

2) 按式(1)和(2)计算每帧能量 E_0 和过零率 Z_0 , 计算两参数组合的零能积 $EZM = E_0 \times Z_0$ 及零能比 $EZR = E_0/Z_0$, 判断零能积与零能比的值, 若 $EZM \leq$ 阈值 T_1 或 $EZR \leq$ 阈值 T_2 , 则可将该帧视为清音, 此为第一级判断。其中, 阈值 T_1 和 T_2 是分别在纯净语音环境下, 对不同男女声的稳定浊音帧和稳定清音帧计算其零能积与零能比的值 (分帧方法与本算法相同), 经过大量统计而得, 本算法中将阈值 T_1 定为 5 000, 阈值 T_2 定为 25。

3) 若 EZM 和 EZR 不满足 2) 中的条件, 进行第二级判断 (采用中央平滑自相关函数法)。首先对输入的语音信号与 1) 作相同的处理 (但不进行预加重), 再对所得每帧信号作补长处理, 将其 240 点数据延至 300 点, 延长方法可以参考文献^[2]。自相关函数作归一化处理将其峰值与阈值 T_3 相比, 若峰值不大于阈值 T_3 , 则把其视为清音, 基音周期可以认为为零。阈值 T_3 的选取与 2) 中阈值一样, 即在同一条件下对自相关函数峰值统计而来。本算法中定为 0.26^[3], 其计算流程如图 1 所示:

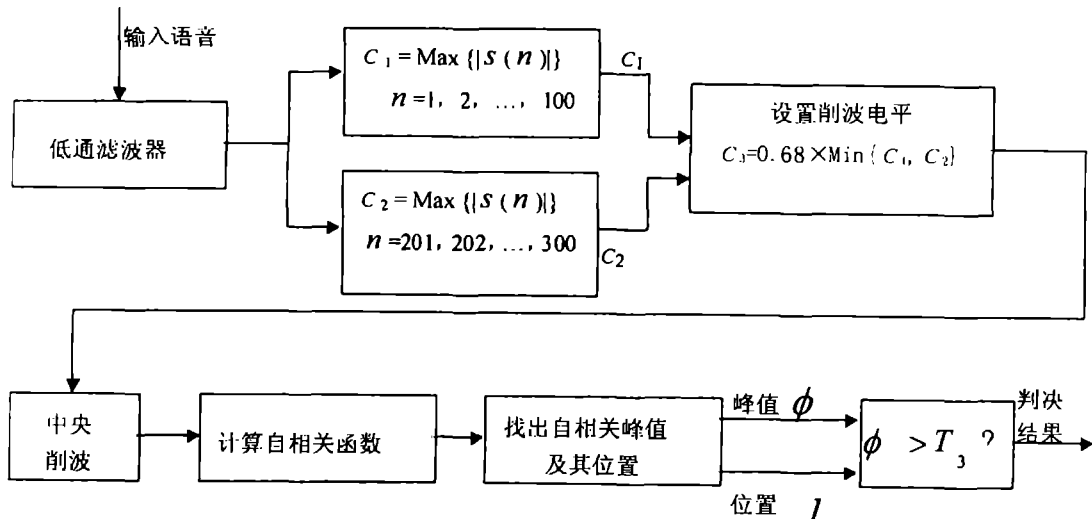


图 1 计算流程图

3 实验结果

对该算法进行了 Matlab 仿真, 由于对参数进行组合, 综合考虑了各个特征参数在清音和浊音时的不同特性以及语音信号本身所固有的模糊性, 它提高了清/浊判决的准确性且误判率低。图 2 是女声“电院”的话音的清/浊判决结果。

将本文所述算法与单参量(过零率)的清/浊判决算法作比较, 统计结果如表 1:

表 1 实验结果统计表

判决算法	汉语男声	汉语女声	判决结果	总错判率
单参量判决算法 (过零率判决)	清音 1 378 帧 浊音 3 201 帧	清音 2 993 帧 浊音 1 586 帧	判错 1 254 帧 判错 1 389 帧	26.857%
组合参量判 决算法	清音 1 378 帧 浊音 3 201 帧	清音 2 993 帧 浊音 1 586 帧	判错 497 帧 判错 526 帧	10.395%

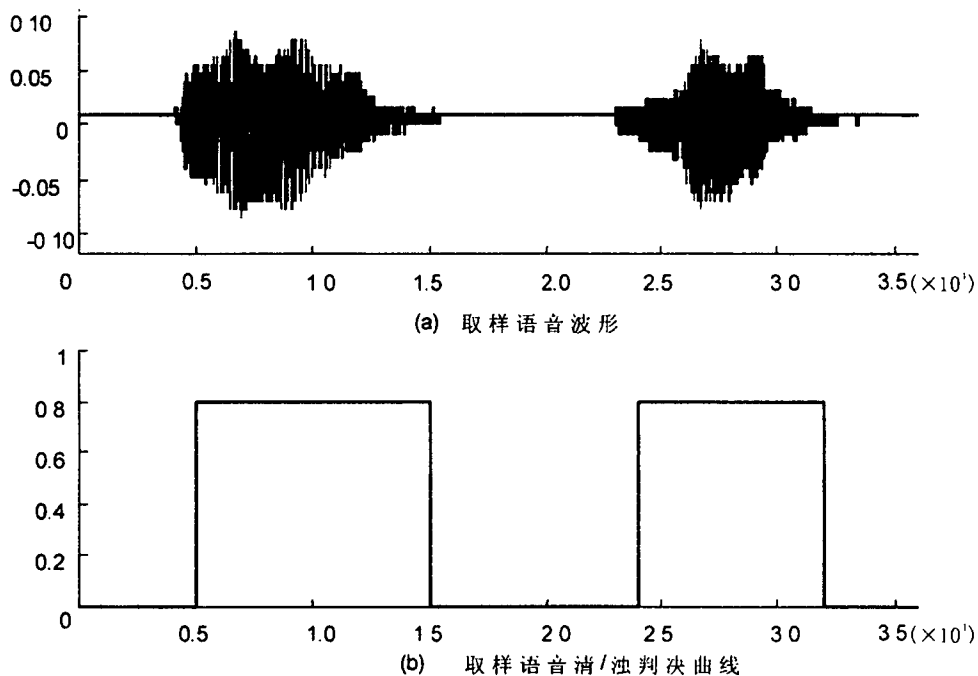


图2 女声“电院”清/浊判决结果

4 结论

算法所用参量短时过零率和短时能量的计算量少,加上自相关函数三参量组合后使得清音/浊音特征概率分布重叠区域减小,实验结果表明算法切实可行,且算法复杂度低、误判概率小。

参考文献:

- [1] ITU - T G - Series Recommendation G. 729B. 8kb/s Conjugate Structure Algebraic - Code - Excited Linear - Prediction (CS - ACELP) [S].
- [2] 易克初,田 斌,付 强. 语音信号处理[M]. 北京:国防工业出版社,2000.
- [3] 陈新富. 高音质 bps 声码器的研究与实现[D]. 西安:空军工程大学电讯工程学院,1998.
- [4] 段艳丽,郑 荣. 语音时变频谱分析的 FFT 技术[J]. 西安:空军工程大学学报(自然科学版),2000,1(2):26 - 28.

(编辑:门向生)

A Simple Method of Unvoiced/Voiced Classification of Speech Signal

GUO Ying, LI Xue - jiao, LI Hong - wei

(The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710077, China)

Abstract: This paper presents a simple method of unvoiced/voiced classification based on combinatorial parameters, which decreases the overlap area of the characteristic probability density of the unvoiced/voiced speech signals. Simulations on computer show that the method is more efficient and robust than that of classification based on single parameter, and the design of the program is simple and concise.

Key words: speech signal; unvoiced/voiced; classification