

带噪语流中浊音端点检测

张伟彬, 王海龙, 刘光军, 何 杰

(联芯科技有限公司, 上海 200233)

摘要: 浊音端点检测在语音处理中占有重要地位, 在语音编解码、语音识别、语音增强处理中都需要用到端点检测技术。常规的以短时能量、过零率等作为判决特征参数的方法无法在低信噪比的系统中满足应用需求。本文以信号的共振峰和基音周期检测为基础检测浊音端点, 算法首先提取语音信号的第一共振峰以及基音周期信息并以此为判决依据判断浊音的起点和终点。实验证明, 这种方法在噪声环境中相对于传统的基于能量检测的或 AMR_WB 标准中的端点检测算法有更高的正确率。

关键词: 端点检测; 共振峰; 基音周期; AMR_WB

中图分类号: T912.3

文献标识码: A

文章编码: 1000-3630(2009)-03-0284-03

DOI 编码: 10.3969/j.issn1000-3630.2009.03.019

Endpoint detection of noisy speech by the use of formant and pitch

ZHANG Wei-bin, WANG Hai-long, LIU Guang-jun, HE Jie

(Leadcore Technology Co., Ltd., Shanghai 200233, China)

Abstract: Voice activity detection, which is needed in speech codec, speech recognition and speech enhancement, is very important in speech processing. Conventional voice activity detection methods based on some simple features such as short term energy cannot meet the demand of application in noisy environment. Speech formants and pitches are used as the features to detect the voice activity in this paper. The information of the first formant and the pitch of speech signal are used to detect the starting points and end points of active voice. Experimental results show that this method can obtain higher accuracy than ordinary detection methods based on energy and the method proposed in AMR_WB standard.

Key words: endpoint detection; formant; pitch; AMR_WB

1 引言

浊音端点检测的主要任务是检测一段信号中包含浊音的部分, 精确的浊音端点检测能够在以下应用中发挥重要作用: 在语音编解码中降低编码速率; 在语音增强算法(用于噪声抑制, 回声消除等)中减少对有效语音的损害; 在语音识别系统中提高系统效率, 同时也提高识别率。

目前在噪声背景下已提出了很多种端点检测算法, 如子带能量映射^[1], 谱熵特征^[2], 自相关相似距离^[3]等, 这些方法在低信噪比的环境中效果比较差。近年来, 很多研究者试图将语音信号变换到另一个空间域, 利用有用信号和噪声在此空间域的不同特性来区分语音信号和背景噪声, 如小波变换域子带能量变化特征^[4]、独立分量分析特征提取^[5]、基

于 Hilbert-Huang 变换的特征提取^[6]等, 然而这些算法的计算量都比较大, 很难在实际应用中使用。

作者在研究中发现, 语音信号的共振峰相对于简单的能量检测具有更好的噪声鲁棒性, 加上基音检测, 在复杂噪声环境中具有更好的检测性能。而且假如在语音编码中使用, 只需要增加很少的计算量。

2 浊音端点检测方法

整个检测过程如图 1 所示, 主要包括 4 部分:

- (1) 预处理对语音信号进行滤波和分帧处理。
- (2) 共振峰分析模块用于分析语音信号的共振峰。
- (3) 基音检测模块用于检测语音信号的基音信息。
- (4) 分析判断模块用于根据共振峰模块和基音检测模块的输出判断当前帧是否是浊音帧。

下面结合 8kHz 采样的带噪语音信号的处理过程详细介绍各个模块。

2.1 预处理模块

预处理模块把语音信号进行滤波处理, 滤除

收稿日期: 2008-05-01; 修回日期: 2008-09-05

作者简介: 张伟彬(1983-), 男, 广东人, 硕士研究生, 研究方向为语音信号处理。

通讯作者: 张伟彬, E-mail: zhangweibin01@sina.com.cn

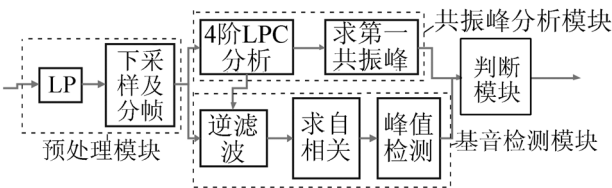


图 1 算法流程
Fig.1 The flow of the algorithm

1700Hz 以上的部分并进行 2 倍下采样,然后将信号分成帧,每帧长度 80 个采样点,帧与帧之间重叠 50%,并对帧信号进行加窗处理,此处使用海明窗。

2.2 共振峰分析模块

利用 Levinson-Durbin 算法计算 4 阶线性预测模型的滤波器系数:

$$H(z)=1/A(z)=\frac{1}{1+\sum_{k=1}^p a_{pk}z^{-k}}, p=4 \tag{1}$$

由于声道的效应,语音的谐振(共振峰)特别明显,而噪声信号一般不具有这种谐振特性。语音信号的第一个谐振峰一般出现在 300Hz~1100Hz 的范围内,共振峰分析模块寻找传输函数(式(1))在 300Hz~1100Hz 内的极大增益 $P(n)$,并将极大增益 $P(n)$ 输出到分析判断模块,若在 300Hz~1100Hz 区间找不到极大值,则输出 0。实验中发现,对于噪声,特别是高斯白噪声,用上面的方法计算出来的极大值几乎不随噪声的强度发生变化,表 1 列举了对不同强度的高斯白噪声应用上述的共振峰分析方法得到的最大增益值,噪声长度是 4s,0dB 时信号幅度为 32767。

表 1 不同强度高斯白噪声的第一共振峰的最大增益值
Table 1 The max gain of the first formant in Gaussian white noise of different strength

信号强度/dB	-18	-15	-14	-10	-6	-3
最大值/dB	7.2	7.4	7.5	8.1	8.5	9.2

2.3 基音检测模块

浊音是由周期性声门脉冲激励声道而产生的。因为基频变化缓慢,因此在窄带语谱图上变为一系列均匀平行的亮线。加窗语音信号在基音及谐波频率处均存在能量峰值。由于大多数噪声信号不具有这种谐波特性,所以可以在输入信号中检测谐波特性以检测浊音信号。

基音最低频率可达 80Hz,最高可达 500Hz,但基音频率大多处在 100Hz~200Hz。语音信号包含十分丰富的谐波分量,由于声道的谐振效应,基波分量往往不是最强的分量。语音信号的第一共振峰通常出现在 300Hz~1100Hz,于是第 2~8 次谐波分量往往比基波分量还强,使得语音信号的波形变复

杂,给基音检测带来困难。为了减少声道谐振特性对基音检测的影响,低通滤波处理以后的语音信号需经过逆滤波处理,逆滤波的目的是使得输入信号的谱平坦化^[7]。使用下面的步骤寻找语音信号的基音周期。

首先求经过逆滤波处理以后的信号 $S_f(n)$ 的自相关:

$$R_k=\sum_{n=0}^{80} S_f(n)S_f(n-k),k=9,10,11,...,71,72,73 \tag{2}$$

在三个不同区间的极大值为: $i=3: 9,10,...,17$
 $i=2: 18,19,...,35$ $i=1: 36,37,...,73$

把找到的极大值和对应的延迟标记为 (M_i,t_i) , $i=1,2,3$ 。接下来在寻找实际基音周期的过程中需避免把基音周期的倍数当作实际的基音周期,采用下面的算法决定实际的基音周期:

$$\begin{aligned} T &= t_1 \\ M_T &= M_1 \\ \text{if}(M_2 > \alpha M_T) \\ M_T &= M_2 \\ T &= t_2 \\ \text{end} \\ \text{if}(M_3 > \alpha M_T) \\ M_T &= M_3 \\ T &= t_3 \\ \text{end} \end{aligned} \tag{3}$$

上面 α 是一经验值,在实验中取 0.84。

2.4 分析判断模块

对于共振峰处理模块的输出,假设前 10 帧是噪声帧,并计算其平均值得到噪声强度 E_{bck} ,当共振峰检测模块的输出增益值 $P(n)$ 满足式(4)时:

$$P(n)/E_{bck} > TH \tag{4}$$

其中 TH 是预先设定的阈值, n 表示帧序号;并且基音检测模块的输出 $T(n)$ 满足式(5)时,才表示当前帧是浊音帧,此时分析判断模块输出 1,否则输出 0。

$$|T(n)-T(n-1)| < 2 \tag{5}$$

噪声强度在非浊音帧的时候采用式(6)更新:

$$E_{bck} = \beta E_{bck} + (1-\beta)P(n) \tag{6}$$

其中 β 是一遗忘因子。

3 实验结果

浊音端点检测算法在不同的噪声条件下进行测试,并与传统的能量检测和 AMR_WB[8]标准中的检测方法进行比较。实验中所用的阈值参数 $TH=2.2$, $\beta=0.95$ 。图 2 给出一段含噪语音的浊音端点检测过

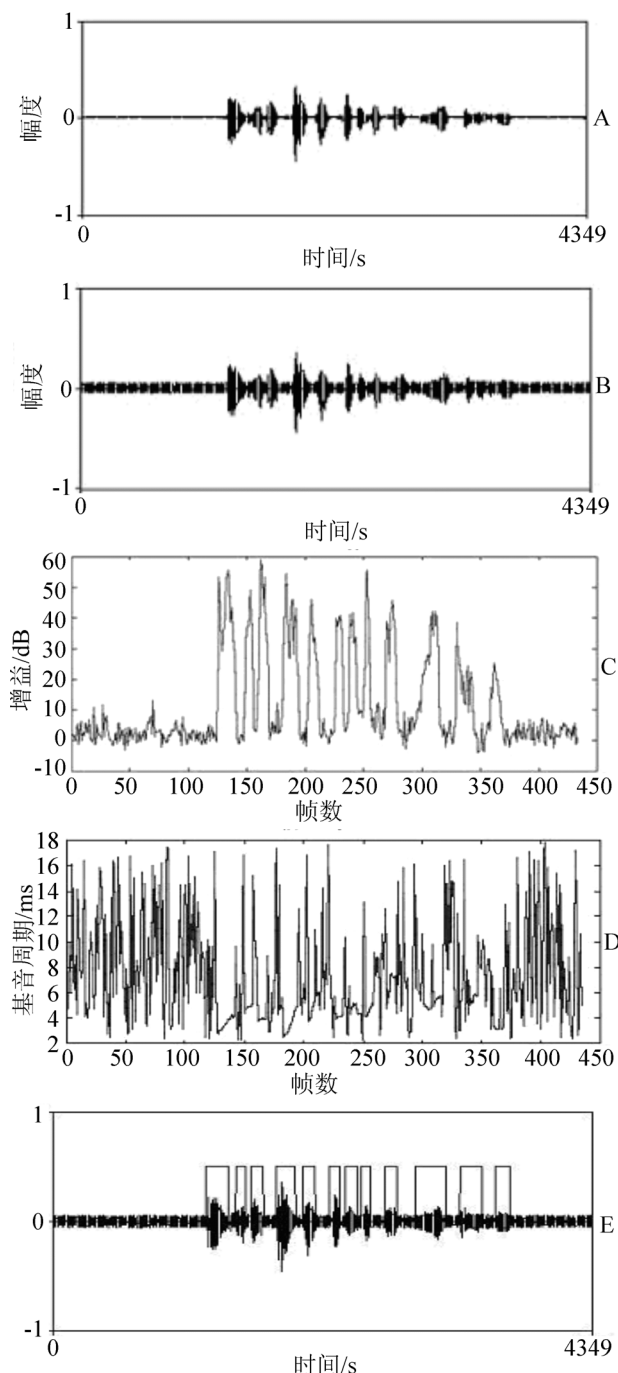


图2 一段4秒左右含噪语音的端点检测过程。

Fig.2 The process of end-point detection for a noisy speech of about 4s.

程: A 表示干净的语音; B 表示含噪语音; C 表示共振峰分析模块的输出结果; D 表示基音检测模块的输出结果; E 表示整个端点检测的输出。表2给出在不同的噪声源和信噪比情况下端点检测的正确率。检测方法为: 1-能量检测; 2-AMR_WB; 3-本文方法。实验语音信号在安静环境下录制, 采用8kHz 采样, 16 比特量化, 录音材料包含5 个男性声音和5 个女性声音, 并与不同的强度的噪声源相混合。对每个语音样本通过手工标号以区分实际语音与背景噪声, 以此作为测试端点检测正确率的标准。

表2 不同噪声源以及不同信噪比情况下不同检测算法的正确率。

Table 2 The accuracy of different algorithm in the case of different SNR and different noise

检测方法	正确率				
	15dB 白噪声	10dB 白噪声	5dB 白噪声	0dB 白噪声	-5dB 白噪声
1	0.92	0.84	0.75	0.63	0.42
2	0.96	0.94	0.9	0.84	0.75
3	0.95	0.95	0.93	0.90	0.86

4 结 语

本文利用浊音信号的声道谐振特性以及浊音激励源的伪周期特性, 基于共振峰检测和基音周期检测提出了一种浊音端点检测算法, 实验证明, 该算法在噪声环境中的鲁棒性好, 这一特性使其能够满足实际应用的需要, 在低信噪比的环境中能够获得比较高的浊音检测正确率。

参 考 文 献

- [1] 田野, 王作英. 基于子带能量线性映射的噪声中端点检测算法[J]. 清华大学学报(自然科学版), 2002, 42(7): 953-956.
TIAN Ye, WANG Zuoying, Robust word boundary detection through linear mapping of the sub-band energy in noisy environments[J]. Journal of Tsinghua University(Science and Technology), 2002, 42(7): 953-956.
- [2] 王让定, 柴佩琪. 一个基于谱熵的语音端点检测改进方法[J]. 信息与控制, 2004, 33(1): 77-81.
WANG Randing, CHAI Peiqi. An improved speech endpoint detection method based on spectral entropy[J], Information and Control, 2004, 33(1): 77-81.
- [3] 陈裴利, 朱杰. 一种新的基于自相关相似距离的语音信号端点检测方法[J]. 上海交通大学学报, 1999, 33(9): 1097-1099.
CHENG Peili, ZHU Jie. A new speech endpoint detection method based on Dis-similarity of auto-correlation[J]. Journal of Shanghai Jiaotong University, 1999, 33(9): 1097-1099.
- [4] 董力, 陈宏钦, 马争鸣. 基于小波变换的语音段起止端点检测算法[J]. 中山大学学报: 自然科学版, 2005, 44(3): 116-118.
DONG Li, CHEN Hongming, MA Zengming. A wavelet-based algorithm for detecting the beginning and end points of voice segments[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2005, 44(3): 116-118.
- [5] 何清波, 孔凡让, 王建平, 等. 基于独立分量分析特征提取的带噪信号端点检测[J]. 数据采集与处理, 2007, 22(1): 25-30.
HE Qingbo, KONG Fanrang, WANG Jianping, et al. Endpoint detection of noisy signal based on feature extraction using independent component analysis[J]. Journal of Data Acquisition and Processing, 2007, 22(1): 25-30.
- [6] WANG, LI Xueyao, ZHANG Rubo. Speech detection based on Hilbert-Huang transform[J]. Computer and Computational Science, June, 2006, (1): 290-293.
- [7] Rabiner L.R., Schafer R.W. Digital processing of speech signals[M]. Prentice-Hall, Inc., 1978.
- [8] 3GPP TS 26.194.AMR wideband speech codec. Voice Activity Detection (VAD)[S].