

引用格式:黄镇坤,章小兵,朱俞清. 低信噪比环境下改进的新能零熵语音端点检测[J]. 微电子学与计算机,2020,37(6):19-23. [HUANG Z K,ZHANG X B,ZHU Y Q. A new improved energy-zero entropy speech endpoint detection with low signal-to-noise ratio[J]. Microelectronics & Computer,2020,37(6):19-23.]

低信噪比环境下改进的新能零熵语音端点检测

黄镇坤,章小兵,朱俞清

(安徽工业大学 电气与信息工程学院,安徽 马鞍山 24300)

摘 要: 语音端点检测是语音处理和识别至关重要的一环,针对传统端点检测方法在低信噪比情况下语音端点检测正确率低,抗噪能力差等问题,本文提出了一种改进的新能零熵特征参数语音端点检测方法,该方法通过对语音三个端点检测特征参数短时过零率,短时能量和基本谱熵分析研究并提出新的语音参数,即为短时能零熵值,最后采用双门限算法来进行端点检测,仿真实验表明,与传统的能零比端点检测法相比,该方法在不同低信噪比情况下有较高的端点检测准确性。

关键词: 端点检测;双门限算法;短时能零熵;低信噪比

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-7180(2020)06-0019-05

A new improved energy-zero entropy speech endpoint detection with low signal-to-noise ratio

HUANG Zhen-kun,ZHANG Xiao-bing,ZHU Yu-qing

(Anhui University of Technology, Ma Anshan 24300, China)

Abstract: Speech endpoint detection is an important part of speech processing and recognition. Traditional endpoint detection methods have low accuracy in speech endpoint detection and poor anti-noise ability under the condition of low signal-to-noise ratio. In this paper, an improved zero-entropy feature parameter speech endpoint detection algorithm is proposed. This method studies the short-time zero-crossing rate, short-time energy and basic spectral entropy of three speech endpoint detection feature parameters, and proposes a new speech parameter, namely, the short-time energy zero-entropy value. Finally, a two-threshold algorithm is adopted to carry out endpoint detection. Simulation results show that compared with the traditional zero-energy ratio endpoint detection method, this method has higher endpoint detection accuracy under different low SNR conditions.

Key words: endpoint detection; energy-zero rate ratio; short-term energy-zero entropy; Low signal-to-noise ratio

1 引言

语音端点检测是语音处理和识别的关键环节。语音端点检测的任务是从带噪声语音信号中分辨出语音段和非语音段,广泛的应用于语音增强、语音识别和语音编码等语音通讯领域。传统的语音端点检测方法例如谱熵法,能零比法,谱距离法,方差法等

在高信噪比环境下的端点检测仍然具有良好的检测效果。但随着噪声的增强,低信噪比环境下的传统端点检测方法的语音端点检测准确率则大幅度的下降,有些方法甚至失效。因此研究低信噪比环境下的语音端点检测方法具有重大的实际意义。

语音端点检测方法类别一般主要分为基于模型类方法和基于特征检测法这两大类。由于基于模型

类方法需要对语音以及噪音进行数学建模其算法比较复杂、嵌入式等平台移植困难、检测效果一般等原因一般采用基于特征检测的方法,即利用语音段与噪音段特性参数的不同来进行区分。

基于特征检测法又主要分为时域端点检测、频域端点检测、多特征结合检测、非线性特征检测。本文对涉及语音端点检测的三种特征参数,时域端点检测的短时过零率和短时能量,频域端点检测的基本谱熵进行研究分析。首先研究分析表明有话区域能量数值大,过零率数值小且谱熵值类似于过零率,在有话段内的谱熵值小于噪声段的谱熵值,其次时域与频域相结合的多特征的端点检测在低信噪比环境中具有更加良好的表现,从而提出新的能零熵语音特征参数来进行端点检测^[1]。经过仿真实验分析,验证了该算法在不同的低信噪比环境下具有较高的端点检测准确性。

2 语音端点检测特征参数

2.1 短时过零率

短时平均过零率为一帧语音中语音信号波形穿过横轴(零电频)的次数。一方面对于连续语音信号,时域波形经过时间轴则为过零;另一方面对于离散信号,相邻的取样值符号相反则成为过零^[2]。所以统计单位时间内抽取样本数值改变符号的次数就可以得到短时平均过零率。短时平均过零率可以用于辨别清音和浊音,清音处于高频段有着较高的过零率而浊音所处的低频段过零率则较低。

短时平均过零率定义:

$$Z_i = \frac{1}{2} \sum_{n=0}^{L-1} |\text{sgn}[y_i(n)] - \text{sgn}[y_i(n-1)]|$$

$$1 \leq i \leq f_n \quad (1)$$

式中, $\text{sgn}[x]$ 是符号函数,即

$$\text{sgn}[x] = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases} \quad (2)$$

因为直流分量的存在会影响短时平均过零率的正确估算,所以在实际计算短时平均过零率参数时,需要再语音信号处理前先消除直流分量。

2.2 短时能量

语音和噪音的区别可以从它们的能量上来进行表现短时能量表示,在噪音环境中语音段的能量是语音能量和噪音能量的和,所以语音段的能量是大于噪声段的。通过计算语音输入信号的短时能量可以对语音和噪音进行区分。在不含噪音或高信噪比情况下短时能量在区分语音段和噪音段有着较好的

表现,而在低信噪比环境下效果大幅度的下降。短时能量为一帧语音信号能量的大小,短时能量定义:

设语音波形时域信号为 $x(n)$, 加窗函数 $\omega(n)$ 分帧处理后得到的第 i 帧语音信号为 $y_i(n)$, 则 $y_i(n)$ 满足:

$$y_i(n) = \omega(n) * x((i-1) * inc + n)$$

$$1 \leq n \leq L, 1 \leq i \leq f_n \quad (3)$$

式中, $\omega(n)$ 为窗函数,一般取汉明窗; inc 为帧移长度; f_n 为分帧后的总帧数; $y_i(n)$ 是一帧数值, $i = 1, 2, \dots, f_n$; $n = 1, 2, \dots, L$; L 为帧长。

则计算第 i 帧语音信号 $y_i(n)$ 的短时能量公式可推导得出:

$$E_i = \sum_{n=0}^{L-1} y_i^2(n) \quad 1 \leq i \leq f_n \quad (4)$$

2.3 基本谱熵

熵这个字源于统计热力学,是一个重要的物理量,用来反映紊乱程度。熵在控制论、概率论、数论等领域都扮演着重要的角色。1948 年香农(C. E. Shannon)将熵的概念引入信息论,把熵作为一种随机事件的不确定性的度量^[3]。当一个系统越有序,其熵就越低。在语音信号端点检测中也广泛的用到熵的概念。Shen 等人则将信息熵成功引入语音端点检测并应用于语音编码领域之中^[4]。

在信息论中通过公式推导得出信息熵是关于概率分布的函数,将熵的概念引入语音端点检测,利用噪音和语音概率分布不同的特性将语音与噪音进行分离。谱熵算法利用输入语音信号的功率谱平坦度来计算语音和噪音的熵值,设定阈值来区分语音和噪音。由于噪音段功率谱有着比较平坦的分布,而语音段功率谱分布较为集中,在 500~3000 Hz 左右,因此语音段相比噪音段在频域内分布概率突变起伏比较明显。

一般设含噪语音信号时域波形为 $x(n)$, 加窗分帧处理后得到的第 i 语音信号为 $x_i(m)$, 经过傅里叶变换后,求得其中第 k 条谱线频率的分量 f_k 其能量谱为 $Y_i(k)$, 则采用公式(8)来定义每个平率分量的谱概率密度

$$p_i(k) = \frac{Y_i(k)}{\sum_{l=0}^{N/2} Y_i(l)} \quad (5)$$

式中, $p_i(k)$ 为第 i 帧第 k 个频率分量 f_k 对应的归一化谱概率密度; N 为 FFT 长度。

第 i 帧对应语音信号的短时谱熵定义:

$$H_i = - \sum_{k=0}^{N/2} p_i(k) \log p_i(k) \quad (6)$$

相比噪音而言,语音信号的频谱具有共振峰频谱的特点,它的归一化谱概率密目分布不均匀,使得语音的谱熵通常都要小于噪声的谱熵^[5].通常利用这种差异化的特性来从带噪语音中提取语音的端点.

3 短时能零熵

上述逐一对于短时过零率,短时能量以及基本谱熵的特征参数进行了分析.通过实验仿真和前人研究表明,在无噪音环境和高信噪比的情况下,通过这三种特征参数单独来进行端点检测表现出良好的准确性^[6].但是在低信噪比的情况下端点检测的准确率大大下降.

为了提高在低信噪比环境下的端点检测准确率,本文提出一种结合短时过零率、短时能量、基本谱熵这三种语音端点检测特征参数的新能零熵端点检测算法.新能零熵端点检测算法即保留了时域检测法中短时过零率与短时能量在检测清音浊音方面的优点,又结合频域检测法中通过对语音和噪音谱熵值的计算来显示它们频域分布概率特性的不同从而区分语音与噪音的能力.

算法其定义如下:

$$EZH_i = (E_i - E_0) / [(Z_i - Z_0)(H_i - H_0) + b]^{\alpha} \quad (7)$$

式中, E_i 、 Z_i 和 H_i 分别代表第 i 帧短时能量,短时过零率和基本谱熵. E_0 、 Z_0 和 H_0 表示前十帧的对数能量均值,短时过零率均值和基本谱熵均值. b 为一个较小的常数,防止 $(Z_i - Z_0)(H_i - H_0)$ 为 0 是时出现溢出情况.经过多次实验 α 取值为 0.15.

能零熵特征值提取的实现流程图如下:

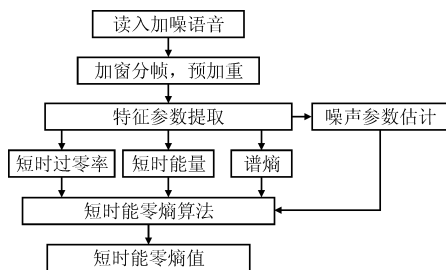


图1 能零熵特征值提取实现流程图

短时能零熵结合时域与频域的语音特征,能够发挥更好的增大噪音与纯净语音差异的效果^[7],可以更好的区分有话语段与噪音段,在低信噪比的环境下有良好的端点检测准确性.

4 双门限算法

双门限法最初是基于短时平均能量和短时平均

过零率而提出的,其原理是汉语的韵母中有元音,能量较大,所以可以从短时平均能量中找到韵母,而声母是辅音,拥有较高的频率,所以其短时过零率较大,利用上述这两个特点找出声母和韵母,等于找出完整的汉语音节.

由于噪音的存在,单凭短时能量或者短时过零率进行判断可靠度低,所以一般采用两者结合的双门限端点检测算法来进行语音端点检测

语音一般分为无声段,过渡段,语音段.短时平均能量主要是测量每一帧的信号能量,由于语音段的短时能量远大于噪音段的,因此根据背景帧设置合理的阈值对每一帧语音信号的短时能量进行对比,最终将语音与噪音进行分离.双门极限判决的端点检测算法首先设置两个能量阈值:高能量阈值 T_1 和低能量阈值 T_2 .当能量参数高于低能量阈值 T_2 时则代表进入语音的过渡段,语音过渡段一般为辅音即汉语中的声母;如果能量参数低于低能量阈值 T_2 则表示为无声段;能量参数高于高能量参数 T_1 则为进入语音段,当在语音段时能量参数又下降至低能量阈值 T_2 以下时则表示语音段结束^[8].其次,利用清音时具有较高的过零率,浊音是具有较低的过零率来且清音的短时过零率明显高于浊音和无声段的特性来很好的区分清音段和无声段,所以通过短时过零率来对判决结果进行补充确认.

5 仿真实验与分析

实验原始语音为纯净不含噪语音,语音内容为“蓝天,白云,碧绿的大海”.噪音选自 Noisex-92 噪音数据库中的白噪音,样本采用频率为 8 kHz, 16 bit 进行量化.通过函数 wavread() 进行调用. wav 文件

实验一通过添加 Noisex-92 噪音数据库中的‘f16. wav’飞机噪音以及‘pink. wav’粉红色噪音分别生成 0 dB 的含噪语音,通过文中所提出的新的能零熵语音端点检测算法进行 MATLAB 语音端点检测仿真并与传统的能零比语音端点检测仿真结果进行对比.信噪比 0 dB 时两种不同噪音环境下的新能零熵算法端点检测和传统能零比算法端点检测仿真结果如图 2—5.

实验一由仿真结果图可以表明本文所提出的新能零熵语音端点检测算法在信噪比为 0 dB 的飞机噪音和粉红色噪音情况下都具有较高的语音端点检测准确率,虽然新能零熵语音端点检测并未检测出“蓝天,白云,碧绿的大海”中的“海”字,但是相较于

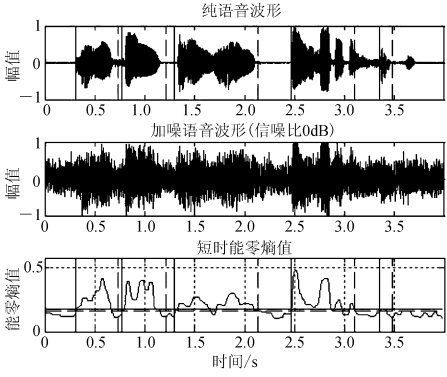


图2 SNR=0 dB 飞机噪音新能零熵端点检测

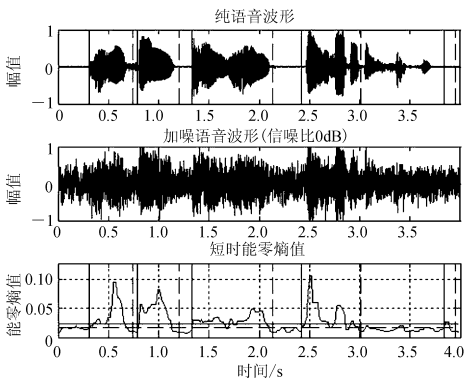


图5 SNG=0 dB 粉红色噪音能零比端点检测

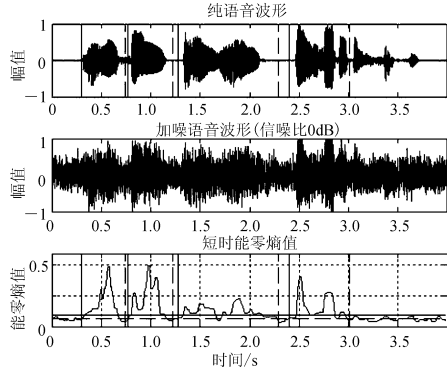


图3 SNG=0 dB 飞机噪音能零比端点检测

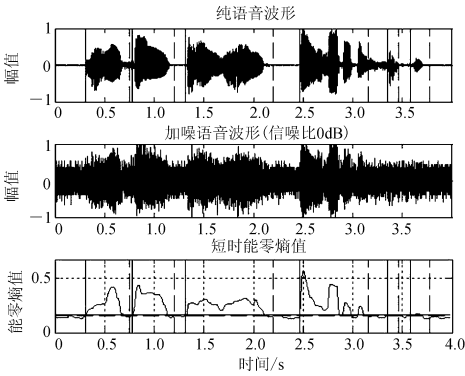


图6 SNR=0 dB 白噪声新能零熵端点检测

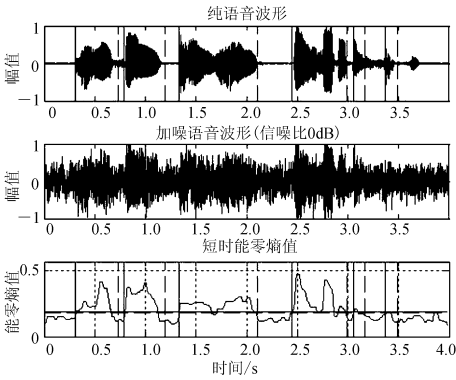


图4 SNR=0 dB 粉红色噪音新能零熵端点检测

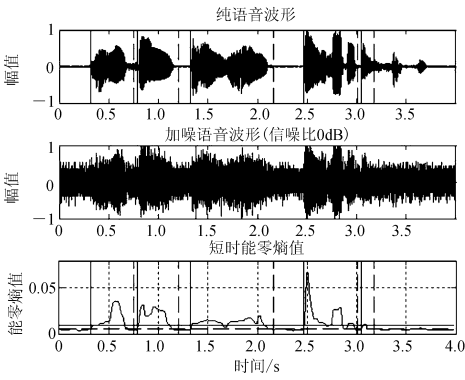


图7 SNR=0 dB 白噪声能零比端点检测

传统的能零比端点检测的准确率低和出现误判的缺点,新能零熵端点检测在低信噪比的不同噪音环境下仍然具有一定的优势。

实验二采用新的能零熵端点检测算法,通过添加噪音库中的白噪音,生成 0 dB 和 -5 dB 两种不同低信噪比情况下的合成含噪语音信号进行 MATLAB 仿真。图 6—9 为传统的能零比与短时能零熵的端点检测法效果对比。

实验二同样也对信噪比为 5 dB 和 10 dB 的情况进行了能零比和能零熵端点检测效果对比,实验表明

当信噪比为 10 dB 时,能零比端点检测仍然具有较好的准确性,但是随着信噪比的降低,能零比的端点检测准确性大大降低,甚至会出现大量的误判断和漏判。而新的能零熵端点检测算法在低信噪比情况下依然表现出很好的鲁棒性和较高端点检测正确性^[9]。

对图 6 和图 7 进行对比观察,可以很明显可以发现,当在信噪比为 0 dB 时,能零比端点检测算法的端点检测效果并不理想,虽然判断有话语段区间基本准确,但是出现很明显漏判,“大海”二字没有被检测出来。

对图 8 和图 9 进行对比观察,在信噪比为 -5

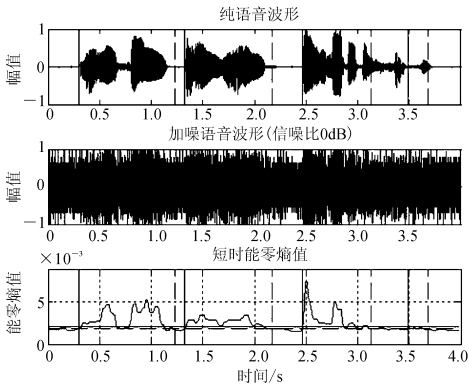


图 8 SNR=-5 dB 白噪声新能零熵端点检测

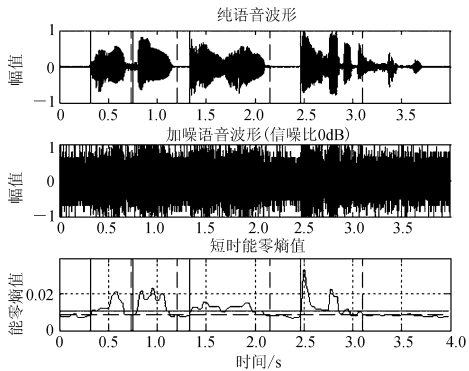


图 9 SNR=-5 dB 白噪声能零比端点检测

dB 情况下,噪音基本已经盖过语音段,能零比端点检测效果如图 9,不仅对“碧绿的大海”进行了误判,还对“大海”二字漏判,而新能零熵相较传统能零比端点检测准确率上有明显的提高。

在信噪比为-5~10 dB 环境下加入白噪声,并对能零比与新能零熵端点检测正确率数据进行收集对比。端点检测正确率(R)定义为^[10]

$$R = \frac{\text{语音总帧数} - \text{错误帧数}}{\text{语音总帧数}} \times 100\% \tag{8}$$

$$\text{错误帧数} = \text{语音误判噪音帧数} + \text{噪音误判语音帧数} \tag{9}$$

四种不同信噪比情况下的传统能零比算法和改进的新能零熵算法的语音端点检测准确率见表 1

表 1 不同信噪比端点检测准确率(%)		
信噪比(SNG)	能零比	能零熵
10	91.2	93.1
5	82.6	87.5
0	73.5	81.6
-5	59.7	73.3

随着信噪比的下降,能零比在低信噪比情况下

的端点检测准确率雪崩式下降,而能零熵表现出较平稳的端点检测准确率,鲁棒性更好,且相较能零比具有较高的端点检测准确率如图 10

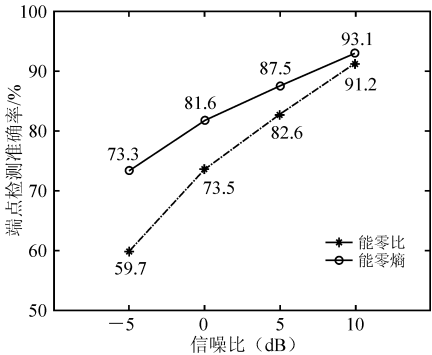


图 10 能零比和能零熵不同信噪比端点准确率

6 结束语

低信噪比环境下的端点检测是语音识别处理的难点.为了得出改进的后的新能零熵语音端点检测算法,本文首先通过对语音端点检测的三种不同特征参数进行分析研究,再将三种特征参数进行提取并结合.最终所得到新能零熵端点检测算法在经过一系列的仿真实验之后并对比了传统的能零比语音端点检测算法之后发现,相较传统的能零比端点检测算法在低信噪比情况下改进的新能零熵算法具有更好的稳定性和鲁棒性等优点,具有较高的端点检测准确率,且对进一步研究低信噪比情况下的语音端点检测提供了参考。

参考文献:

- [1] 韩志艳,王旭,王健. 基于短时能零熵和鉴别信息的语音端点检测[J]. 东北大学学报(自然科学版), 2009, 30(12):1690-1693.
- [2] 尹岩岩,殷业,罗汉文. 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.
- [3] 宋知用. MATLAB 在语音信号分析与合成中的应用[M]. 北京:北京航空航天大学出版社,2017:75-80.
- [4] HUNG J W, SHEN J L, LEE L S. Improved robustness for speech recognition under noisy conditions using correlated parallel model combination [C]// IEEE International Conference on Acoustics, IEEE, 1998:553-556
- [5] 赵东雪. 低信噪比下语音识别系统的研究[D]. 浙江:浙江工业大学, 2014.
- [6] 董胡. 基于先验信噪比和能零熵的语音端点检测算法[J]. 计算机技术与发展, 2017, 27(7):72-75.

表 1 本文 ADC 性能与其他设计比较

性能参数	文献[10]	文献[11]	本文
工艺/nm	180	40	180
电源电压/V	1.8	0.7	0.7
分辨率/bit	12	12	12
采样率/KS/s	200	200	100
SNDR/dB	71.55	69.1	72.8
有效位数/bit	11.59	11.19	11.79
功耗/ μ W	336	0.51	0.95
FOM/fJ/conv	545	1.1	2.68
文献来源	ISCAS2016	VLSI2017	本文

6 结束语

基于 TSMC 180 nm 工艺,设计了一款 12 bit 二进制缩放重组电容加权 SAR ADC,与传统冗余校准方法相比,本设计在不添加额外冗余电容的情况下,实现了校准的功能,并且未对输入信号的摆幅产生影响,提高了面积利用率和性能. 仿真结果显示,在 0.7 V 电源电压下,当采样率为 100 KS/s,输入频率为 46.6 KHz 时,SAR ADC 的 SNDR 为 72.8 dB,有效位数为 11.79 bit,功耗仅 0.95 μ W,FOM 值为 2.68 fJ/conv.

参考文献:

[1] QU W Y, ZHAO J Q, ZHANG Z F, et al. Low-energy switching method based on asymmetric binary search algorithm for SAR ADCs[J]. Journal of Circuits Systems and Computers, 2019, 29(6): 2050087-1-2050 087-10.

[2] 汪正锋. 用于植入式医疗设备的超低功耗 SAR ADC 设计[D]. 成都: 电子科技大学, 2015.

[3] ZHANG H, WANG X H, ZHANG L, et al. A 10-bit 120-MS/s SAR ADC in 90nm CMOS with redundancy compensation[C]// IEEE MTT-S International Wireless Symposium. Chengdu, China, 2018: 1-3.

[4] 倪亚波, 张创, 徐世六, 等. 一种带二进制校正的 10 位 100MS/s SAR ADC[J]. 微电子学, 2016, 46(2):

145-149.

[5] LIU C C. Design of high-speed energy-efficient successive-approximation analog-to-digital converters [D]. Tainan: National Cheng Kung University, 2010.

[6] LIU C C, KUO C H, LIN Y Z. A 10 bit 320 MS/s Low-Cost SAR ADC for IEEE 802. 11ac Applications in 20 nm CMOS[J]. IEEE Journal of Solid-State Circuits, 2015, 50(11): 2645-2654.

[7] ZHU Z M, LIANG Y H. A 0.6-V 38-nW 9.4-ENOB 20-kS/s SAR ADC in 0.18- μ m CMOS for Medical Implant Devices[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2015, 62(9): 2167-2176.

[8] 吴青龙. 8 位 20KSPS 0.4V 超低功耗 SAR ADC 的研究[D]. 西安: 西安电子科技大学, 2015.

[9] 赵金强, 梅年松, 张钊锋, 等. 一种超低功耗 SAR A/D 转换器[J]. 微电子学, 2019, 49(1): 1-6.

[10] RABUSKE T, FERNANDES J. A 12-bit SAR ADC with background self-calibration based on a MOSCAP-DAC with dynamic body-biasing[C]// IEEE International Symposium on Circuits and Systems (ISCAS). Montreal, QC, Canada, 2016: 1482-1485.

[11] HU Y S, LIN K Y, CHEN H S. A 510nW 12-bit 200kS/s SAR-assisted SAR ADC using a re-switching technique[C]// Symposium on VLSI Circuits. Kyoto, Japan, 2017: C238-C239.

作者简介:

曲维越 男,(1994-),硕士研究生. 研究方向为混合信号集成电路.

张钊锋 男,(1972-),博士生导师,研究员. 研究方向为物联网核心芯片、柔性快速开发平台和车联网关键技术等.

梅年松(通讯作者) 男,(1978-),博士,副研究员. 研究方向为射频 CMOS 集成电路系统设计、智能传感器和能量采集等.

E-mail: meins@sari. ac. cn

(上接第 23 页)

[7] 董胡, 钱盛友. 改进的能量谱熵端点检测算法[J]. 测控技术, 2016, 35(6): 26-29.

[8] 王满洪, 张二华, 王明合. 基于双门限算法的端点检测改进研究[J]. 计算机与数字工程, 2017, 45(11): 2223-2228.

[9] WU B F, WANG K C. Robust endpoint detection algorithm based on the adaptive band-partitioningspectral entropy in adverse environments[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 762-775.

[10] 陈莹莹, 毕春艳, 龙建忠. 低信噪比下的语音端点检测算法研究[J]. 电视技术, 2018, 42(6): 9-12.

作者简介:

黄镇坤 男,(1992-),硕士研究生. 研究方向为语音识别及算法研究. E-mail: 543904796@qq. com

章小兵 男,(1972-),博士,教授. 研究方向为测控技术与语音识别.

朱俞清 男,(1996-),硕士研究生. 研究方向为语音识别及算法研究.