语音端点检测技术研究进展 *

韩立华[1], 王 博[2], 段淑凤[1]
(1. 石家庄铁道学院 计算机与信息工程分院, 石家庄 050043; 2. 国防科学技术大学 电子科学与工程学院, 长沙 410073)

摘 要: 总结了语音端点检测技术的基本原理、步骤及发展情况,介绍了当前主要语音端点检测算法的研究进展;并对各主要算法的检测性能进行了较详细的分析和比较。最后,总结了语音端点检测技术的发展特征,并展望了该技术的未来发展趋势。

关键词: 端点检测; 研究进展; 发展趋势
中图分类号: TN912.3; TP391 **文献标志码**: A **文章编号**: 1001-3695(2010)04-1220-07
doi:10.3969/j.issn.1001-3695.2010.04.005

Development of voice activity detection technology

HAN Li-hua[1], WANG Bo[2], DUAN Shu-feng[1]
(1. *Scool of Computing & Informatics, Shijiazhuang Railway Institute, Shijiazhuang* 050043, *China*; 2. *College of Electronic Science & Engineering, National University of Defense Technology, Changsha* 410073, *China*)

Abstract: This paper summarized the principle, the elementary steps and the development of VAD technology firstly; and then, introduced the mainly current VAD algorithms; furthermore, made the detailed analysis and comparison to the dominating VAD algorithms. Finally, summarized the characteristics of the development of VAD and discussed some viewpoints about the direction of the development.

Key words: voice activity detection; research development; direction of the development

0 引言

语音端点检测(voice activity detection, VAD)又称有声/无声检测(voiced/unvoiced detection)、语音边界检测(speech/word boundary detection)、语音终止点检测(speech endpoint detection)等,通常是指在复杂的噪声背景环境下的信号流中分辨出语音信号和非语音信号,并确定语音信号的起始点和终止点,为后续信号处理提供必要的支持。准确的语音端点检测对多通道传输系统、语音识别系统以及语音增强系统等都具有重要的现实意义。语音端点检测技术的发展不仅可以提高传输系统效率,而且能够提升识别系统精度,改善增强语音质量。

从 1959 年贝尔实验室最早提出语音端点检测开始,语音端点检测技术已经历了近五十年的发展,产生了数百种方法[1]。特别是近一二十年,由于大多数基于时域特性的端点检测方法在噪声条件下已不能实现可靠的检测,在语音识别、语音编码及语音增强等技术发展的推动下,又出现了很多种语音端点检测算法。它们主要是采用各种新的特征参数,以提高算法的抗噪声性能。这些新的特征参数主要有时频参数[2]、倒谱系数[3,4]、自相关相似距离[5]、信息熵[6~8]等。除此之外,还通过将信号的几种特征参数相结合[9]来检测端点。而对语音端点的判决方式也由原来的单一门限和双门限发展到基于模糊逻辑和模式分类的判决[10]。

到目前为止,成熟而应用最广的语音端点检测方法是国际电信联盟(International Telecommunication Union, ITU)提出的 G.729 标准[11]和欧洲电信标准化协会(European Telecommunications Standards Institute, ETSI)提出的应用于第三代移动通信系统的 AMR Option 1/Option 2 标准[12],这两个标准均采用多参数综合判决的检测方法,在较高信噪比(signal-to-noise ratio)条件下获得了很好的检测效果。

1 语音端点检测的基本原理

语音端点检测本质上是通过语音和噪声对于相同参数所表现出的不同特征来区分两者的,其基本流程如图 1 所示。其中预处理通常包括分帧和预滤波等。分帧是指将语音信号分段(称为语音帧,各帧通常是有交叠的),预滤波一般是指采用高通滤波器滤除低频噪声;参数提取是指选取可以反映语音和噪声差别的特征参数;端点判决是指采用一种判决准则(如门限判决或模式分类等)来区分语音帧与非语音帧;后处理是指对上述判决结果进行平滑滤波等处理,得到最终的语音端点判决结果。在语音端点检测的流程中,参数提取和端点判决是两个关键步骤。

参数提取是指选取能够反映语音和噪声差别的特征参数,是以语音和噪声的特性为基础。语音信号是一种典型的非平稳信号。但是,语音的形成过程是与发音器官的运动密切相关

收稿日期: 2009-10-09; **修回日期**: 2009-11-23 **基金项目**: 陕西省自然科学基金资助项目(2006F40)

作者简介:韩立华(1978-),女,河北辛集人,讲师,硕士,主要研究方向为多媒体技术应用、计算机软件应用、多媒体教学系统等(hanlihua@sjzri.edu.cn);王博(1981-),男,陕西商州人,博士研究生,主要研究方向为语音信号处理、多传感器管理调度、空间信息对抗技术;段淑凤(1980-),女,硕士,主要研究方向为多媒体技术、模式识别等.

的,这种物理运动比起声音振动速度要缓慢得多,因此语音信号常常可假定为短时平稳的。语音可粗略分为清音和浊音两大类。浊音在时域上呈现出明显的周期性,在频域上出现共振峰,而且能量大部分集中在较低频段内。但清音段相对于很大一类噪声没有明显的时域和频域特征,类似于白噪声。在语音端点检测算法研究中,可利用浊音的周期性特征,而清音则难以与宽带噪声区分。

语音端点检测流程如图 1 所示。

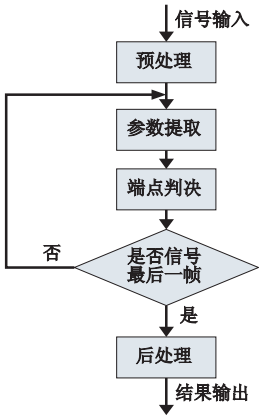


图1　语音端点检测流程图

噪声来源于实际的应用环境,因而其特性变化无穷。混入语音中的噪声可以是加性的,也可以是非加性的。考虑到加性噪声更普遍且易于分析问题,并且对于非加性噪声,有些可以通过一定变换转换为加性噪声,因此几乎所有研究语音端点检测技术的学者都是针对加性噪声展开研究的。

端点判决通常是通过门限判决的方式实现的,即设定一定的判决门限,当所提参数大于(小于)该门限时认为是语音,而小于(大于)该门限时认为是噪声。语音端点的判决方式最初是单一门限和双门限的形式,随后又逐步发展为基于模糊逻辑和模式分类的判决方式。

一种好的语音端点检测算法通常应具备以下四个特征:

a)语音端点判决的准确率高,尤其强调对于清音段端点的正确判决;

b)检测算法必须具有对绝大多数噪声的鲁棒性,抗干扰能力强;

c)判决准则应具有自适应性,而不是简单的固定门限判决;

d)检测算法应较简单,且运算量较小,易于硬件实现。

以上特征分别是从语音端点检测的准确性、稳定性、自适应性和运算量上对算法提出的要求。

2　语音端点检测技术发展过程

1959 年,贝尔实验室在其开发的电话传输系统中最早提出了语音端点检测技术,并应用于该系统的通信信道时间分配中。第一个系统而完整的语音端点检测算法是贝尔实验室的 Rabiner 等人[13]于 1975 年提出的基于短时能量和短时过零率的方法。该方法通过设定短时能量参数的高、低两个门限进行端点位置初判,然后设定短时过零率参数的检测门限最终确定语音段的端点。根据作者在无噪声环境和信噪比大于 30dB 的弱噪声背景环境下所进行的仿真实验,该方法在高 SNR 环境下具有很好的检测性能。之后许多学者在此基础上又提出了一些改进算法[14~18]。He 等人[15]详细比较了平方能量、绝对值能量和对数能量分别应用于端点检测的性能:绝对值能量是应用于端点检测效果最好的能量参数;平方能量在应用于端点检测时会丢失幅值较小的音节和清音;对数能量则会使语音信号低能量部分(这部分信号与微弱的噪声性质相似)增强,不利于分离孤立词识别系统中的音节,且运算量较前两个能量参数大。

1980 年,日本京都技术大学的 Kobayashi 等人[19]通过 FFT 提取语音频域信息,提出了一种基于语音频域信息的语音端点检测算法。之后,研究人员通过分析带噪语音信号的频域成分信息,提出了许多基于频域参数的端点检测算法,包括基于频谱变化量[20]的方法、基于子带能量[21]的方法、基于多带调制能量[22]的方法、基于谱相关参数[23]的方法以及基于语音基频信息[24]的方法。随后,中国台湾交通大学的 Lin 等人[25]提出了最小 Mel 尺度频带(minimum mel-scale frequency band, MiMSB)参数和增强时频(enhanced time-frequency, ETF)参数。其中,MiMSB 参数自适应地从 Mel 滤波器组中选择具有最小能量的频带;ETF 参数既包含时域信息,又包含频域信息。Lin 等人将这两种参数相结合,提出了一种新的基于时频域参数的检测方法——MiMSB-ETF 方法。通过对汽车噪声环境下语音信号进行仿真实验,结果表明,其检测性能较之 Brian Mak 等人所提的时域—频域参数检测方法有较大改善。此外,一些学者还提出了基于小波变换[26]、Walsh 谱能量分布[27]和 Hilbert-Huang 变换[28]的检测方法。

1991 年,美国匹兹堡大学的 Ghiselli-Crippa 等人[29]首次提出了一种基于人工神经网络的语音、非语音、静音区分算法。该算法采用前馈神经网络区分语音、非语音和静音,并通过一种快速收敛的训练算法(类牛顿误差最小化方法与 Hessian 矩阵正定近似的结合)确定网络权重,其检测性能较之统计决策方法有明显提高,且对于输入特征不需要进行繁琐的假设。随后,美国亚利桑那大学的 Qi 等人[30]将多层前馈网络(multi-layer feedforward network, MFN)应用于语音端点检测领域,于 1993 年提出了一种基于 MFN 的检测方法。该方法较之需要大量训练的最大似然方法[31]来说,只需很少的样本即可完成有效训练,且仍能达到较高的正确检测率。中国台湾交通大学的 Wu 等人[32]2000 和 2001 年先后提出了基于自组织模糊推理神经网络(self-organizing neural fuzzy inference network, SONFIN)和循环自组织模糊推理神经网络[33](RSONFIN)的检测方法。SONFIN 是模糊逻辑系统的一个普通连接模型,可以自动找到其最优结构和最优参数,且学习速度和模拟能力均强于普通的神经网络,而 RSONFIN 与 SONFIN 不同的是在三四层之间多一个反馈连接。作者采用时频(TF)参数和精确时频(RTF)参数作为网络输入,分别建立了四种检测方法,即 TF-SONFIN、TF-RSONFIN、RTF-SONFIN 和 RTF-RSONFIN,实验表明这四种方法检测性能的优劣顺序为 RTF-RSONFIN > TF-RSONFIN > RTF-SONFIN > TF-SONFIN。此外,研究人员还提出一些基于三态神经网络[34]、径向基函数(radial basis function, RBF)网络[35]、多层感知器(multi-layer perceptron, MLP)网络[36]和自适应线性神经元(adaptive linear neuron, ADALINE)网络[37]的方法。

1993 年,英国斯旺西大学的 Haigh 等人[3]将加权欧氏距离引入倒频谱域,定义了倒谱距离,并首次提出了一种基于倒

谱距离的语音端点检测算法。仿真实验表明,该方法可以实现 SNR 最低为 -2dB 时的平稳高斯白噪声环境下带噪语音端点检测。随后,又出现了一些改进型方案。Bou-Ghazale 等人[38]将倒谱距离参数与短时能量参数相结合,建立二维的判决准则,提出了一种基于短时能量和倒谱距离的端点检测算法,并对该方法和短时能量方法应用于语音识别系统的性能在不同噪声背景下进行了比较,结果显示该方法较之短时能量方法对于不同噪声背景下语音识别的正确率提高均超过了 10%。王博等人[39]在比较上述基于倒谱距离端点检测算法的基础上,根据不同 SNR 环境下端点检测所需最佳门限的统计结果,引入短时 SNR 估计,并由统计方法拟合短时 SNR 估值和检测所需最佳门限的关系曲线,提出了一种基于短时 SNR 估计及其与判决门限关联的倒谱距离端点检测算法。该方法在低 SNR 色噪声背景下仍能有效检测语音端点,且检测错误率相比文献[3]中的基本倒谱距离检测方法减少 5% 左右。

1997 年,美国的 McClellan 等人[40]将仙农熵引入端点检测领域,由信号余弦变换的频谱分布引出谱熵定义,提出了一种基于谱熵的端点检测算法,并成功应用于语音编码领域。之后,Huang 等人[9]深入研究了谱熵参数和短时能量参数应用于语音端点检测的优缺点,然后用短时能量对谱熵参数加权,提出了一种基于加权谱熵的检测算法。解放军信息工程大学的徐望等人[41]在研究带噪语音信号协方差矩阵特征分解的基础上,定义了一种新的信息熵函数,即特征空间能量熵,并将其应用于语音端点检测。通过对平稳高斯白噪声、战斗机座舱噪声和高速公路噪声环境下带噪语音信号进行仿真,并提取加权错误测度与谱熵算法进行比较,结果表明该算法的检测错误率较之谱熵算法减少 10% 以上,明显优于谱熵算法,但高维矩阵特征值的求解也带来了运算量的严重负担。此外,清华大学的李晔等人[42]还提出了一种基于模糊加权谱熵的端点检测算法;王博等人[43]还对常用的几种基于谱熵的端点检测算法性能进行了较为详细的比较分析,进一步验证了文献[41]方法的优势。

1998 年,中国香港的 Cheung 等人[44]提出了一种基于 2-D Markov 模型的端点检测方法。该方法定义了语音有无两个状态及状态转移概率,通过 Markov 模型的状态转移来判决语音的端点。同年,上海交通大学的朱杰等人[45]把语音识别中常用的 HMM 方法直接用于语音信号的端点检测,提出了一种基于 HMM 模型的端点检测算法。他们将被检测信号分为两部分:背景和废料(语音处理中习惯上把有用或无用的发音统称为废料)。废料就是上述两部分的分界处。在训练阶段,分别得出背景噪声和废料的模型参数;在测试阶段,用 Viterbi 解码方法在训练模型基础上对被检测语音进行分解,确定语音的哪些帧与背景噪声匹配,哪些帧与废料匹配,从而得到端点的位置。在上述研究的基础上,Couvreur 等人[46]还提出了一种基于小波收缩密度估计的 HMM 语音端点检测方法,Gazor 等人[47]提出了一种基于 Laplacian 模型的方法。

2000 年,土耳其巴斯肯特大学的 Tanyer 等人[48]将几何方法引入检测门限的确定,提出了几何自适应门限的检测方法。该方法采用幅值概率分布函数自适应更新能量门限,且不局限于在噪声段进行更新,对噪声具有较强的鲁棒性,尤其对于非平稳噪声。Tanyer 等人还对短时能量、短时过零率和最小平方周期估计(least-square periodicity estimator, LSPE)参数的检测性能进行了研究,并将这三个参数与几何自适应门限相结合,提出了一种多参数检测方法,通过对交通噪声、走廊噪声、饭店噪声及喷泉噪声环境下带噪语音信号的仿真实验结果分析,表明该多参数方法可以实现 SNR 大于 -5 dB 条件下带噪语音信号的有效端点检测。2004 年,意大利卡塔尼亚大学的 Beritelli 等人[49]提出了一种实时估计带噪语音信号当前 SNR 的方法,并由此 SNR 估计值定义一 Sigma 函数,提出一种基于 Sigma 函数的端点检测方法,该方法较之第三代移动通信系统的 AMR VAD Option 1 的检测性能有 10% 左右的提高。

近年来,一些学者还将分形技术[50]和混沌理论[51]引入端点检测,分别提出了相应检测方法。2005 年,清华大学的刘鹏等人[52]将视觉信息与音频信息相结合,提出了基于双模式的语音端点检测方法。仿真实验表明,其检测性能明显优于仅采用音频信息的情况。2007 年,Gorriz 等人[53]提出了一种基于联合高斯分布和似然比检验的端点检测算法及其实时实现方法;Tahmasbi 等人[54]提出基于 GARCH 滤波、Gamma 分布和自适应门限函数的端点检测算法。2008 年,Shin 等人[55]提出了基于修正最大后验准则的端点检测算法;Masakiyo Fujimoto 等人[56]提出基于多特征和信号决策自适应综合的语音端点检测算法;潘欣裕等人[57]将 Hilbert-Huang 变换中的经验模态分解(EMD)引入端点检测,并提出了基于 EMD 拟合特征的语音端点检测新方法。由此可见,现今学者们对语音端点检测的研究越来越倾向于多种技术或测度的融合。

3 语音端点检测算法分类

语音端点检测算法是各种技术的大融合,到目前为止还没有统一的分类方法。一般可以按照应用的范围分类,也可以按照所使用的特征参数或判决准则分类。按照所采用的特征参数或判决准则的不同,本文将语音端点检测算法分为七类,分别是:

a)时域参数方法

主要是指基于短时能量和过零率[13~18]、短时自相关[58]及一些其他时域参数(如对数能量、绝对值能量、最小均方参数[59]等)的方法。此外,基于几何自适应门限[48]和基于 Sigma 函数[49]的方法也归入这一类。

b)变换域参数方法

包括基于频域参数[19~24]、时频域参数[2,25,60]及小波域参数[26]的方法。此外,还包括一些基于 Walsh 谱能量分布和 Hilbert-Huang 变换的检测方法。

c)距离和失真测度方法

包括基于 LPC 距离[31]、倒谱距离[3,38,39]、Kullback- Leibler 距离[61]及长时谱差异[62]的方法等。

d)信息论方法

主要是指基于熵函数[6,9,40~43]的检测算法和基于信号编码理论的检测算法[63]。

e)人工神经网络方法

包括基于前馈网络[29,30]、自组织和循环自组织模糊推理神经网络、径向基函数网络、多层感知器网络、自适应线性神经元网络等的方法。

f)统计模型和模式分类方法

主要是指基于 HMM 模型[44~46]、Laplacian 模型[47]和 Bayesian 模型[64]的方法。此外,还有一些基于多统计模型[65]、

似然检验[66](likelihood ratio test, LRT)、模式识别[67]、模糊逻辑[68]及高阶累计量[69]的方法。

g)其他方法

除了前述六类方法之外,近年来研究人员还提出了一些基于其他参数的方法,主要包括基于分形技术、混沌理论及双模式[52](将视觉信息和音频信息相结合)的方法。

4　语音端点检测算法性能比较

针对语音端点检测算法的性能评价问题,学者们提出了很多方法,其中绝大多数是客观评价方法。由于主观评价受人的因素影响很大,本文只关注客观评价结果。客观评价方法可分为直接评价和间接评价。直接评价是指直接对端点检测结果提取一定的性能参数进行分析,而间接评价是指将端点检测作为语音识别、语音编码、语音增强等系统的一部分,通过识别、编码、增强等结果提取参数进行分析。直接评价方法很多,普遍认同的较好的评价方法是由 Beritelli 等人[68]提出的四种错误参数,此外还有正确检测比例、加权错误测度[41]等。间接评价一般是通过语音识别正确率、语音编码效率、增强语音可懂度等进行评价的。

对各类端点检测算法性能的比较见表 1。

由于受实验方法和数据的影响,相同的方法在不同的文献中也会得到不同的实验结果。因而本文在进行算法性能比较时作如下假定:

a)由于实验方法、实验数据以及实验环境的不同,不同文献中的定量结果之间不具有可比性,本文只给出检测性能的定性比较;

b)对同一类方法在不同文献中的实验结果,选择背景噪声种类多的情况,因为这样所得到的结果更具有说服力;

c)对同一类方法在同一文献中不同背景噪声下的实验结果取平均值,进行比较分析。

笔者通过对第一、三、四类中主要语音端点检测算法在平稳高斯白噪声、F16 战斗机噪声和 M109 坦克噪声环境下进行仿真,并提取四种错误参数[69]和加权错误测度[41],仿真结果如图 2 所示。由图 2 可以得到如下结论:总体来讲,基于倒谱距离的检测算法性能优于基于谱熵函数的算法(特征空间能量熵算法除外),基于短时能量的算法性能最差;对于较低 SNR 环境下的检测,特征空间能量熵算法的性能最好,而对于较高 SNR 环境下的检测,基于 SNR 估计的倒谱距离算法与基于短时能量和倒谱距离的检测算法的性能最好。

5　语音端点检测技术发展的特点和趋势

从语音端点检测技术的发展历程及各种方法的性能看,语音端点检测技术的发展具有如下特点:

a)语音端点检测技术研究取得了长足的进展,特别是近十几年来,学者们针对噪声环境下语音端点检测所做的大量工作。语音端点检测算法的性能已经由最初的仅适用于较高 SNR 平稳高斯白噪声环境发展到可适用于 −5 dB 这样低 SNR 部分非平稳噪声环境下的检测。

b)从语音端点检测算法所采用的判决准则来看,语音端点检测技术的发展大致经历了两个阶段:门限判决阶段和模式分类阶段。两个阶段并不是完全分开的,而是互相混叠的,新方法往往是建立在传统方法的基础上的。

表 1　语音端点检测算法性能比较

算法类别	优点	缺点	检测性能综合评价
时域参数方法	1)原理简单; 2)运算量小,便于实时实现	1)仅适应于平稳噪声的检测,对于不同噪声来说算法的鲁棒性差; 2)绝大多数方法只能检测 SNR > 10 dB 的带噪语音,只有个别多参数方法可以检测 SNR 约 0 dB 的带噪语音; 3)几乎不能区分清音和噪声	较差
变换域参数方法	1)原理较简单(基于小波的检测方法除外); 2)大多数方法运算量较小,便于实时实现; 3)可以适应部分非平稳噪声的检测	1)大多数情况下只能实现 SNR > 0 dB 条件下带噪语音信号的检测,对于低 SNR 的情况方法失效; 2)当噪声与语音信号具有类似的变换域特征时,方法失效; 3)对于清音的检测效果一般	一般
距离和失真测度方法	1)原理较简单; 2)特征参数受语音信号时域、频域参数变化影响小,鲁棒性较好; 3)运算量较小,便于实时实现; 4)能够实现平稳噪声环境下较低 SNR 条件下带噪语音信号的检测; 5)可以适应部分非平稳噪声的检测	1)对于大多数非平稳噪声的检测仅限于较高 SNR 的情况; 2)当噪声与语音信号的相应特征参数差别不大时,方法失效; 3)对于清音的检测效果一般	好
熵函数和信号编码方法	1)原理较简单; 2)特征参数受语音信号时域、频域参数变化影响小,鲁棒性较好; 3)大多数算法运算量较小,便于实时实现; 4)能够实现低 SNR 平稳噪声环境下带噪语音信号的检测和部分非平稳噪声环境下的检测	1)当噪声与语音信号的时域或频域分布类似时,方法失效; 2)对于清音的检测效果较差	较好
人工神经网络方法	1)能够实现低 SNR 平稳噪声环境下带噪语音信号的检测; 2)选取适当参数,通过预先训练,也可以完成对非平稳噪声环境下带噪语音信号的检测; 3)对于清音的检测效果较好	1)原理复杂; 2)运算量大,实时实现受限; 3)需要预先对算法参数进行训练,受先验信息限制	较好
统计模型和模式分类方法	1)能够实现低 SNR 平稳噪声环境下带噪语音信号的检测; 2)选取适当统计模型或模式分类方法,也可以完成对非平稳噪声环境下带噪语音信号的检测; 3)对于清音的检测效果较好	1)原理较复杂; 2)运算量较大,受具体应用环境限制,不一定能够实时实现; 3)对于不同噪声环境下的检测,可能需要不同的统计模型	好
其他方法	或是基于新理论的应用,或是独辟蹊径的新想法,大多数都存在运算量大或是适用于特殊噪声的缺点,算法本身还需进一步的深入研究		

c)新的特征参数的提出和新的判决准则的建立是推动语音端点检测算法演进的两个重要因素,也是语音端点检测技术研究取得突破的关键。

d)基于模式分类的判决方法在语音端点检测技术的发展过程中具有重要的作用。建立在统计模型基础上的模式分类方法相比门限判决方法对噪声具有很强的鲁棒性,因而近年来效果较好的检测方法几乎都是采用基于模式分类的判决方法。

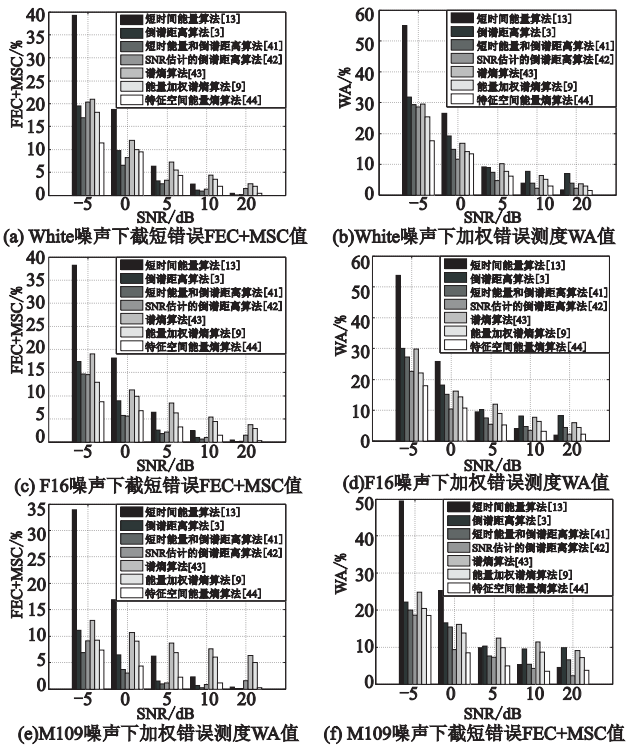
当前,语音端点检测技术还远滞后于通信技术发展的脚步,在此领域还有很多问题需要研究:

a)对于强干扰非平稳噪声和快速变化的噪声环境,如何找到更好的端点检测方法将是进一步研究的主要方向。

b)提取人耳听觉特性可以更加有效地区分语音和噪声,从而更加准确地检测语音端点。因而,融入人耳听觉特征的语音端点检测算法也将是未来主要研究的方向。

c)预先未知噪声统计信息条件下的语音端点检测算法已经出现,但仍处于萌芽阶段。虽然预先未知噪声统计信息条件下的端点检测是未来语音端点检测技术的发展方向,但在理论方法和技术参数等方面还有待进一步突破。

d)提取非音频信息辅助区分语音和噪声开辟了语音端点检测技术的一个新的思路。文献[52]已对该方法作了初步探索,但更加深入的研究尚需进一步展开。



(a) White噪声下截短错误FEC+MSC值　(b)White噪声下加权错误测度WA值

(c) F16噪声下截短错误FEC+MSC值　(d)F16噪声下加权错误测度WA值

(e)M109噪声下加权错误测度WA值　(f) M109噪声下截短错误FEC+MSC值

图2　仿真结果

6　结束语

本文回顾了语音端点检测技术的研究发展情况,介绍了语音端点检测技术的分类和性能,同时探讨了今后语音端点检测算法的研究发展方向。随着通信技术的不断进步,对语音端点检测技术提出了更高的要求,语音端点检测技术将得到进一步的发展。

参考文献:

[1] BULLINGTON K, FRASER J M. Engineering aspects of TASI[R]. 1959: 353-364.

[2] JUNQUA J C, MAK B, REAVES B. A robust algorithm for word boundary detection in the presence of noise[J]. IEEE Trans on Speech and Audio Processing, 1994, 2(3): 406-412.

[3] HAIGH J A, MASON J S. Robust voice activity detection using cepstral features[C]//Proc of IEEE Region 10 Conference TENCON. 1993: 321-324.

[4] 胡光锐, 韦晓东. 基于倒谱特征的带噪语音端点检测[J]. 电子学报, 2000, 39(10): 95-97.

[5] 陈斐利, 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.

[6] ABDALLAH I, MONTRESOR S, BAUDRY M. Robust speech/non-speech detection in adverse conditions using an entropy based estimator[C]//Proc of the 13th International Conference on Digital Signal Processing. 1997: 757-760.

[7] SHEN Jia-lin, HUNG J W, LEE L S. Robust entropy-based endpoint detection for speech recognition in noisy environments[C]//Proc of ICSLP. 1998: 232-235.

[8] 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.

[9] HUANG Liang-sheng, YANG C H. A novel approach to robust speech endpoint detection in car environments[C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing-Proceedings. 2000: 1751-1754.

[10] WU Ya-dong, LI Yan. Robust speech/non-speech detection in adverse conditions using the fuzzy polarity correlation method[C]//Proc of IEEE International Conference on Systems, Man, and Cybernetics. 2000: 2935-2939.

[11] ANNEX B. ITU-T Rec. G.729, A silence compression scheme for G.729 optimized for terminals conforming to ITU-T V.70[S]. 1996.

[12] ETSI TS 126 073 v4.1.0(2001), Universal mobile telecomm systems (UMTS); mandatory speech codec speech processing functions, AMR speech codec; voice activity detector(VAD) (3GPP TS 26.094 version 4.0.0 release 4)[S]. 2001.

[13] RABINER L R, SAMBUR M R. An algorithm for determining the endpoints of isolated utterances[J]. Bell System Technical Journal, 1975, 54(2): 297-315.

[14] TABOADA J, FEIJOO S, BALSA R, *et al*. Explicit estimation of speech boundaries[J]. IEEE Proceedings Science Measurement and Technology, 1994, 141(3): 153-159.

[15] HE Qiang, ZHANG You-wei. On prefiltering and endpoint detection of speech signal[C]//Proc of ICSP-98. 1998: 749-752.

[16] LI Qi, ZHENG Jin-song, TSAI A, *et al*. Robust endpoint detection and energy normalization for real-time speech and speaker recognition [J]. IEEE Trans on Speech and Audio Processing, 2002, 10(3): 146-157.

[17] HO K, YANG T Y, PARK K J, *et al*. Robust voice activity detection algorithm for estimating noise spectrum[J]. IEEE Electronics Letters, 2002, 36(2): 180-181.

[18] MARZINZIK M, KOLLMEIER B. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics[J]. IEEE Trans on Speech and Audio Processing, 2002, 10(2): 109-118.

[19] KOBAYASHI Y, NIIMI Y. Word boundary detection by pitch con-

tours in an artificial language[C]//Proc of IEEE ICASSP'80. 1980: 900-903.

[20] OSHIKIRI M, AKAMINE M. A 2.4-kbps variable-bit- rate ADP-CELP speech coder[J]. Electronics and Communications in Japan, 2000,83(7):32-41.

[21] VAHATALO A, JOHANSSON I. Voice activity detection for GSM adaptive multi-rate codec[C]//Proc of IEEE Workshop on Speech Coding for Telecommunications Proceedings. 1999:55-57.

[22] EVANGELOPOULOS G, MARAGOS P. Multiband modulation energy tracking for noisy speech detection[J]. IEEE Trans on Audio, Speech, and Language Processing,2006,14(6):2024-2038.

[23] JEBARA S B. Coherence-based voice activity detector[J]. IEEE Electronics Letters,2002,38(22):1393-1395.

[24] RAMANA R G V, SRICHAND J. Word boundary detection using pitch variations[C]//Proc of the 4th International Conference on Spoken Language Proceedings. 1996:813-816.

[25] LIN C T, LIN J Y, WU G D. A robust word boundary detection algorithm for variable noise-level environment in cars[J]. IEEE Trans on Intelligent Transportation Systems,2002,3(1):89-101.

[26] CHEN Shi-huang, WANG J F. A wavelet-based voice activity detection algorithm in noisy environments[C]//Proc of the 9th IEEE International Conference on Electronics, Circuits and Systems. 2002: 995-998.

[27] HUANG J, TSENG B D. A walsh transform based endpoint detection of isolated utterances[C]//Proc of Asilomar Conference on Signal, Systems & Computers. 1991:335-338.

[28] WANG Wu, LI Xue-yao, ZHANG Ru-do. Speech detection based on Hilbert-Huang transform[C]//Proc of the 1st International Multi-Symposiums on Computer and Computational Sciences. 2006: 290-293.

[29] GHISELLI-CRIPPA T, El-JAROUDI A. A fast neural net training algorithm and its application to voiced-unvoiced-silence classification of speech[C]//Proc of IEEE International Conference on Acoustics, Speech, Signal Processing. 1991:441-444.

[30] QI Ying-yong, HUNT B R. Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier[J]. IEEE Trans on Speech and Audio Processing,1993,1(2):250-255.

[31] RABINER L R, SAMBUR M R. Application of an LPC distance measure to the voiced-unvoiced-silence detection problem[J]. IEEE Trans on Acoustics, Speech, and Signal Processing,1977,25(4):338-343.

[32] WU G D, LIN C T. Word boundary detection with mel- scale frequency bank in noisy environment[J]. IEEE Trans on Speech and Audio Processing,2000,8(5):541-554.

[33] WU G D, LIN C T. A recurrent neural fuzzy network for word boundary detection in variable noise-level environments[J]. IEEE Trans on Systems, Man, and Cybernetics-Part B: Cybernetics,2001, 31(1):84-97.

[34] BERITELLI F, CASALE S, SERRANO S. Adaptive V/UV speech detection based on acoustic noise estimation and classification[J]. IEEE Electronics Letters,2007,43(4):249-251.

[35] HOYT JD, WECHSLER H. Detection of human speech using hybrid recognition models[C]//Proc of the 12th IAPR International Conference on Pattern Recognition. 1994:330-333.

[36] HUSSAIN A, SAMAD S A, FAH L B. Endpoint detection of speech signal using neural network[C]//Proc of TENCON 2000 Proceedings. 2000: 271-274.

[37] 胡瑞敏, 薛东辉, 姚天任,等. 神经网络方法及其在语音识别中的应用[J]. 高技术通讯, 1995,5(6): 11-15.

[38] BOU-GHAZALE S E, ASSALEH K. A robust endpoint detection of speech for noisy environments with application to automatic speech recognition[C]//Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2002:3808-3811.

[39] 王博, 郭英, 段艳丽,等. 基于倒谱特征的语音端点检测算法研究[C]//Proc of CCSP-2005. 2005: 212-215.

[40] McCLELLAN S, GIBSON J D. Variable-rate CELP based on subband flatness[J]. IEEE Trans on Speech and Audio Processing, 1997,5(2):120-130.

[41] 徐望, 丁琦, 王炳锡. 一种基于特征空间能量熵的语音信号端点检测算法[J]. 通信学报, 2003,24(11):125-132.

[42] 李晔, 张仁智, 崔慧娟,等. 低信噪比下基于谱熵的语音端点检测算法[J]. 清华大学学报:自然科学版, 2005,45(10):1397-1400.

[43] 王博, 郭英, 韩立峰. 基于熵函数的语音端点检测算法研究[J]. 信号处理,2009,25(3):368-373.

[44] CHEUNG M T, LEA C T. CCI improvement by voice activity detection & power-control in a cellular system[C]//Proc of the 48th IEEE Vehicular Technology Conference. 1998:1229-1233.

[45] 朱杰, 韦晓东. 噪声环境中基于 HMM 模型的语音信号端点检测算法[J]. 上海交通大学学报, 1998,32(10):14-16.

[46] COUVREUR L, COUVREUR C. Wavelet-based non-parametric HMM's: theory and applications[C]//Proc of IEEE Conference. 2000:604-607.

[47] GAZOR S, ZHANG W. A soft voice activity detector based on a Laplacian-Gaussian model[J]. IEEE Trans on Speech and Audio Processing,2003,11(5):498-505.

[48] TANYER S G, OZER H. Voice activity detection in nonstationary noise[J]. IEEE Trans on Speech and Audio Processing,2000,8(4):478-482.

[49] BERITELLI F, CASALE S, SERRANO S. A low-complexity speech-pause detection algorithm for communication in noisy environments[J]. European Trans on Telecommunications,2004,15(1):33-38.

[50] YANG Su, LI Zong-ge, CHEN Yan-qiu. A fractal based voice activity detector for Internet telephone[C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2003: 808-811.

[51] 林嘉宇, 王跃科, 黄芝平,等. 一种新的基于混沌的语音、噪声判别方法[J]. 通信学报, 2001,22(2):123-128.

[52] 刘鹏, 王作英. 多模式语音端点检测[J]. 清华大学学报:自然科学版,2005,45(7):896-899.

[53] GORRIZ J M, RAMIREZ J, PUNTONET C G. Effective jointly PDF-based voice activity detector for real-time applications[J]. Electronics Letters, 2007,43(4):1-2.

[54] TAHMASBI R, RRZAEI S. A soft voice activity detection using GARCH filter and variance gamma distribution[J]. IEEE Trans on

Audio, Speech and Language Processing, 2007, 15(4): 1129-1134.

[55] SHIN J W, KWON H J, JIN S H, *et al*. Voice activity detection based on conditional MAP criterion[J]. IEEE Signal Processing Letters, 2008, 15: 257-260.

[56] FUJIMOTO M, ISHIZUKA K, NAKATANI T. A voice activity detection based on the adaptive integration of multiple speech features and a signal decision scheme[C]//Proc of ICASSP 2008. 2008: 4441-4444.

[57] 潘欣裕, 赵鹤鸣, 陈雪勤, 等. 基于 EMD 拟合特征的耳语音端点检测[J]. 电子与信息学报, 2008, 30(2): 362-366.

[58] WU Ya-dong, LI Yan. Robust speech/non-speech detection in adverse conditions using the fuzzy polarity correlation method[C]//Proc of IEEE International Conference on Systems, Man and Cybernetics. 2000: 2935-2939.

[59] TUCKER R. Voice activity detection using a periodicity measure[J]. IEE Proceedings-1, 1992, 139(4): 377-380.

[60] MAK B, JUNQUA J C, REAVES B. A robust speech/non-speech detection algorithm using time and frequency-based features[C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 1992: 269-272.

[61] RAMIREZ J, SEGURA J C, BENITEZ C, *et al*. A new kullback-leibler VAD for speech recognition in noise[J]. IEEE Signal Processing Letters, 2004, 11(2): 266-269.

[62] RAMIREZ J, SEGURA J C, BENITEZ C, *et al*. Efficient voice activity detection algorithms ising long-term speech information[J]. Speech Communication, 2004, 42(3-4): 271-287.

[63] LIU C H, HUANG C C. Voice activity detector based on CAPDM architecture[J]. Electronics Letters, 2001, 37(1): 68-69.

[64] ZHANG Jian-ping, WARD W, PELLOM B. Phone based voice activity detection using online Bayesian adaptation with conjugate normal distributions[C]//Proc of International Conference on Acoustics, Speech and Signal Processing. 2002: 321-324.

[65] CHANG J H, KIM N S, MITRA S K. Voice activity detection based on multiple statistical models[J]. IEEE Trans on Signal Processing, 2006, 54(6): 1965-1976.

[66] CHO Y D, KONDOZ A. Analysis and improvement of a statistical model-based voice activity detector[J]. IEEE Signal Processing Letters, 2001, 8(10): 276-278.

[67] ATAL B S, RABINER L R. A pattern recognition approach to voice-dunvoiced-silence classification with applications to speech recognition[J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1976, 24(3): 201-212.

[68] BERITELLI F, CASALE S, CAVALLARO A. A robust voice activity detector for wireless communications using soft computing[J]. IEEE Journal on Selected Areas in Communications, 1998, 16(9): 1818-1829.

[69] LI Ke, SWAMY M N S, AHMAD M O. An improved voice activity detection using higher order statistics[J]. IEEE Trans on Speech and Audio Processing, 2005, 13(5): 965-974.

(上接第 1211 页)

[29] SONG Shan-shan, HUANG Kai, ZHOU Run-fang. Trusted P2P transactions with fuzzy reputation aggregation[J]. Internet Computing, 2005, 9(6): 24-34.

[30] RICHARDSON M, AGRAWA R, DOMINGOS P. Trust management for the semantic Web[C]//Proc of the 2nd International Semantic Web Conference. Berlin: Springer-Verlag, 2003: 351-368.

[31] DESPOTOVIC Z, ABERER K. Maximum likelihood estimation of peers' performance in P2P networks[C]//Proc of the 2nd Workshop on the Economics of Peer-to-Peer Systems. Cambridge: Harvard University, 2004.

[32] GOLBECK J, HENDLER J. Accuracy of metrics for inferring trust reputation in semantic Web-based social networks[C]//Proc of International Conference on Knowledge Engineering and Knowledge Management. Berlin: Springer-Verlag, 2004: 116 -131.

[33] SABATER J, SIERRA C. REGRET: a reputation model for gregarious societies[C]//Proc of the 4th Workshop on Deception Fraud and Trust in Agent Societies. 2001: 61-70.

[34] 常俊胜, 王怀民, 尹刚. DyTrust: 一种 P2P 系统中基于时间的动态信任模型[J]. 计算机学报, 2006, 29(8): 1301-1307.

[35] 郭磊涛, 杨寿保, 王菁, 等. P2P 网络中基于矢量空间的分布式信任模型[J]. 计算机研究与发展, 2006, 43(9): 1564-1570.

[36] 李景涛, 荆一楠, 肖晓春, 等. 基于相似度加权推荐的 P2P 环境下的信任模型[J]. 软件学报, 2007, 18(1): 157-167.

[37] ZIEGLER C, GOLBECK J. Investigating interactions of trust and interest similarity[J]. Decision Support Systems, 2007, 43(2): 460-475.

[38] 刘艳玲, 白宝兴, 王大东, 等. 应用关系集合的 P2P 网络信任模型[J]. 吉林大学学报: 信息科学版, 2009, 29(2): 210-214.

[39] SONG Wei-hua, PHOHA V V. Neural network-based reputation model in a distributed system[C]//Proc of IEEE CEC. San Diego: [s. n.], 2004: 321-324.

[40] BURAGOHAIN C, AGRAWAL D, SURI S. A game theoretic framework for incentives in P2P systems[C]//Proc of the 3rd International Conference on Peer-to-Peer Computing. Los Alamitos: IEEE Press, 2003: 48-56.

[41] GOLBECK J. Generating predictive movie recommendations from trust in social networks[C]//Proc of the 4th International Conference on Trust Management. Berlin: Springer-Verlag, 2006: 93-104.

[42] Bibserv[EB/OL]. http://www.bibserv.org/.

[43] KAUTZ H, SELMAN B, SHAH M. ReferralWeb: combining social networks and collaborative filtering[J]. Communications of the ACM, 1997, 40(3): 63-66.

[44] SINGH A, LIU Ling. TrustMe: anonymousmanagement of trust elationships in decentralized P2P systems[C]//Proc of IEEE International Conference on P2P Computing. Sweden: [s. n.], 2003: 142-149.

[45] MITRA A, UDUPA R, MAHESWARAN M. A secure trust and incentive management framework for public resource based computing utilities[C]//Proc of IEEE International Symposium on Cluster Computing and the Grid. Cardiff, UK: [s. n.], 2005: 267-274.