

高噪声环境下的语音活动检测技术

胡大雷 刘珩

(中国农业大学 信息与电气工程学院,北京 100083)

摘要 以 MOTOROLA 公司生产的 DSP56F801 芯片为核心搭建系统硬件电路,对语音信号进行处理。采用能够根据背景噪声变化而自适应调节门限的算法对语音活动进行精确检测,实现整机系统在高噪声环境下的正常通讯。实际检测结果表明:在 100、110、115 dB 的强背景噪声环境下,系统识别率分别为 100%、96%和 94%,丢字漏字率分别为 2%、5%和 10%;在 100~110、110~115 dB 的背景噪声环境,系统稳定性分别为 100%和 98%。采用此种算法,增强了系统的抗干扰能力,在高噪声和噪声不太稳定的环境下,实现了对语音活动的精确检测。

关键词 语音活动检测;短时能量;过零率;短时自相关;自适应门限

中图分类号 TN 912.3

文章编号 1007-4333(2005)03-0065-04

文献标识码 A

Voice detection technology (VDT) in a background of high noise

Hu Dalei, Liu Heng

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract In order to identify the voice activity accurately in a background of 100 - 115 dB high noise, voice signals were interpreted by using a circuit based on the DSP56F801 chip produced by the Motorola. The voice detection algorithm used in the circuit could adjust threshold automatically according to the change of noise signals, which ensured a normal communication of the whole system. The test results showed that the detection rate reached 100%, 96% and 94% corresponding the noise levels of 100, 110 and 115 dB respectively; and the words loss rate were 2%, 5% and 10% respectively. The system stability reached 100% and 98% respectively in high noise background of 100 - 110 dB and 110 - 115 dB. The system ability on anti-jamming was improved and a precision detection of VD was implemented in a high noise background or to unstable noise signals by the developed algorithm.

Key words VAD (voice activity detection); short-time energy; zero-cross ratio; short-time auto-correlation; adaptive threshold

在实际场合中,语音经常与各种噪声混合在一起,因此,语音活动检测的抗噪声性能越来越受到人们的重视,在移动通信、野战通信和声控系统中尤为突出。传统的语音端点检测方法都是针对实验室安静环境的,20世纪90年代末,随着自动语音识别(ASR)技术在车载电话通信、关键字识别、人员身份确认、连续语音识别等方面的应用逐渐广泛,人们才开始研究噪声下的语音端点检测。如基于自相关相似距离的语音端点检测算法^[1],基于倒谱特征的端点检测方法^[2]等;但是这些方法都只是针对白噪声的,而且计算方法复杂,难以实现。动态追踪能量包络算法^[3-5]复杂度小,在高噪声下的语音检测效果

有所提高,但在 100 dB 以上高噪声环境下的检测效果仍较差,且算法的精确度也不够高,误判的几率仍很大。语音激活检测(voice activity detection, VAD)算法主要依据语音与噪声的不同特性进行语音和噪声的判决,能够从连续采样得到的数字信号中检测出语音信号段和噪声信号段,其准确的语音端点检测提高了系统识别率和处理效率。该算法具有能量、过零率、零能比、时域和频域等基音检测功能,在其早期发展阶段,由于没有先进的高速数字处理芯片,这些功能主要靠仿真器件来实现。随着数字信号处理器(DSP)技术的发展,过去仅限于军队使用的价格昂贵的 DSP 价格成几何级数下降,使得民用

收稿日期:2004-11-25

作者简介:胡大雷,硕士研究生;刘珩,副教授,主要从事高噪声环境下语音通讯研究。

产品也可以采用其作为技术支撑。故本研究采用 VAD 算法,以 DSP 作为硬件平台对 100~115 dB 高噪声环境下的语音检测系统进行了研究。

1 语音检测算法

VAD 采用的语音特征参数^[6]:

1) 短时能量

$$E = \sum_{n=1}^N x^2(n)$$

式中: E 表示 1 帧信号(可能为语音信号,也可能为语音和噪声的叠加信号)的能量值; N 为帧长; $x(n)$ 为数字化后的语音幅度,单位根据不同数字化方法而定。语音可分为清音、浊音两大类。清音相对类似于白噪声和强度不太大的背景噪声;浊音的短时能量要比噪声的高很多,可以用来区分噪声。

2) 短时过零率

$$Z_{CR} = \frac{1}{2} \sum_{n=1}^{N-1} |\text{sign}[x(n)] - \text{sign}[x(n-1)]|$$

式中: Z_{CR} 表示 1 帧信号传过零电平的次数,其中

$$\text{sign}[x(n)] = \begin{cases} 1 & x(n) > 0 \\ -1 & x(n) < 0 \end{cases}$$

为符号函数。清音的特点是短时能量比较小,甚至接近背景噪声,用短时能量比较难分辨,但过零率很高,可以此作为判断的一个依据。

3) 短时自相关

$$R(m) = \sum_{n=0}^{N-1-m} x(n)x(n+m)$$

式中: m 为抽样周期个数。噪声的自相关除了 $m=0$ 时都很小,而语音信号的相关性很高,除了主峰外还有较高的副峰,可以根据这一特点来区分两者。这里不直接用自相关,而是采取求主副峰比的办法。人声音的频率为 300~3 000 Hz,在 8 000 点/s 的采样率下,当 m 为 25~108 时至少会有 1 个副峰,取最近的副峰来求主副峰比。主副峰比很大时认为是噪声,较小时是语音。

2 系统结构及功能

2.1 系统构成

系统构成见图 1。若系统启动一段时间没有语音输入,这段时间内的输入则作为背景噪声处理。系统启动时即开始数据采集,将从麦克风输入的信号转换成 16 位数字信号,为此在内存中开辟了一个缓存区。将 1 帧语音数据存入缓存区后,再采集

1 帧语音数据存入其中,程序据此求出门限。当检测到语音端点到来后,控制器发出一个选通控制信号控制可编程放大器,使语音输出。在此过程中继续检测语音的终点,检测到终点后,由控制器发出衰减控制信号,停止语音的传送。

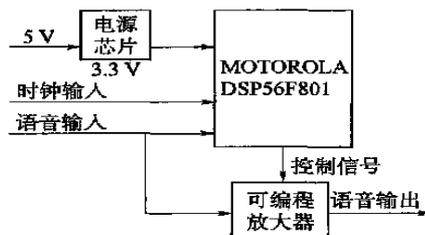


图 1 语音检测系统构成

Fig. 1 Structure of VD system

2.2 器件选择

系统采用 MOTOROLA 公司生产的 DSP56F801 芯片作为数字信号处理芯片。该芯片的特点是体积小、传输速度快、运算能力强且功耗低、价格便宜,适用于便携式产品的开发应用。其基本特性如下:

1) 关键部分采用双哈佛结构,支持并行处理。DSP56F801 内部有 7 条总线,而通常微处理器的 CPU 中只有地址总线、数据总线和控制总线。增加的内部总线是为了实现多指令的并行处理,提高处理速度。

2) 在 80 MHz 时钟频率下运算速度可达到每 s 40 兆条指令。

3) 支持 DSP 和 MCU 2 种功能风格的指令系统。

4) JTEG/ONCE 程序调试接口,允许在系统设计过程中进行调试,并可对软件进行实时调试。电源芯片采用 LD1117,把 5 V 电压转化为 3.3 V 以驱动 DSP 工作。

2.3 软件算法

首先依据能量进行判别:

1) 自适应电平均衡。目的在于使受干扰信号以背景噪声为基准进行能量的归一化(均衡)。

2) 确定背景噪声能量门限。

3) 起点判别。用均衡后的语音能量与噪声能量门限比较进行判别。

4) 终点判别。方法同 3)。

图 2 示出软件算法流程,这里采用的噪声能量门限是自适应的。在实际系统中,噪声能量门限不

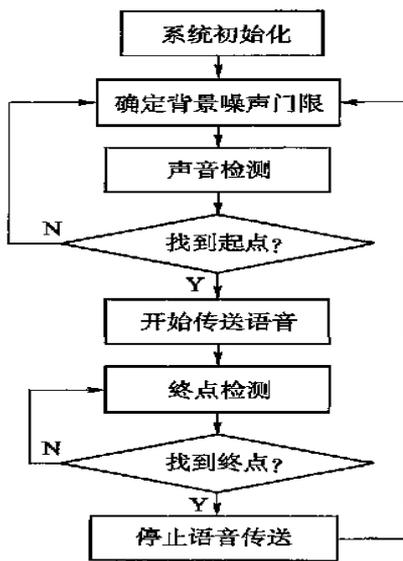
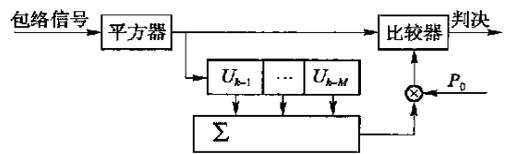


图 2 软件算法流程

Fig. 2 Flowchart of software arithmetic

是固定不变的,而是随背景噪声电平的变化而起伏变化的;因此可以在信号捕获过程中通过对噪声功率的最大似然估计来改变判决噪声能量门限大小,实现自适应门限^[7],使能量门限随背景噪声的强度变化而发生近似线性的变化,保持门限值始终比噪声高 3~6 dB。这样,当噪声发生波动时,门限值也相应变动。当语音出现时,由于其能量比噪声高 3~6 dB,从而系统检测到语音活动。图 3 示出自适



输入包络信号为窄带复高斯白噪声时 U_k 为中心 χ^2 分布,输入包络信号为噪声和语音的混和信号时 U_k 为具有二阶自由度的非中心 χ^2 分布; P_0 为固定门限

图 3 自适应门限的信号相关检测框图

Fig. 3 Signal-correlation detection chart of self-adaptive gate-limit

应门限的信号相关检测框图。

本文中采用的判决门限与噪声功率有关,自适应门限的设置方法要求对信号中的噪声功率进行估计。最大似然估计 (ML) 算法是信号估计中一种最佳方法,噪声功率的最大似然估计可以表示为:

$$P_{ML}^2 = \frac{1}{2 M_0} \prod_{l=1}^{M_0} U_{k-l}$$

式中: M_0 为功率估计所采用的包络数。估计噪声功率的过程可称为自适应过程 (AD), 自适应判决门限

$$P_{AD} = P_0 \cdot P_{ML}^2$$

随着噪声的波动,门限值也相应变化(图 4)。

采用自适应门限的优点是可以动态地实现语音启动检测功能,使系统能够在不同噪声环境下工作,

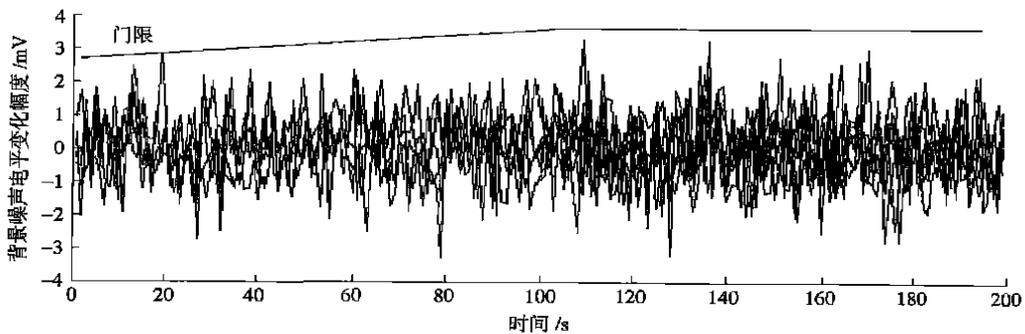


图 4 自适应门限曲线

Fig. 4 A self-adaptive gate-limit curve

提高了系统的可靠性和实用性。

3 系统性能测试

为了测试系统对语音启动反应的可靠性和稳定性,分别在 100、110 和 115 dB 的背景噪声环境下对系统进行了评测。方法是将 VAD 功能模块放入通话器前端电路中,对其进行性能测试和分析。评测

指标如下:

1) 整个句子的可懂度。测试目的是检查在噪声环境下整机系统对语音的跟随能力和抗噪声、抗干扰能力。采用标准的词表测试,检测在整机系统中受话者对发话者发出的整个句子的理解程度。经统计,在 100、110 和 115 dB 背景噪声环境下,句子的可懂度分别为 100%、96%和 94%。

2) 句子的丢字漏字情况。测试的主要目的是检查在噪声环境下整机系统中清音辅音字的丢字漏字情况。清音辅音能量很低,很难冲破噪声门限,所以噪声与其混淆,造成句子开头清音辅音字的丢字现象以及在句子中间因清音辅音字丢失造成的漏字现象,从而影响整机系统的性能。采用标准的词表测试,检测在整机系统中受话者对发话者发出的整个句子中丢字漏字情况的判断。经统计,在 100、110 和 115 dB 的背景噪声环境下,句子的丢字漏字率分别为 2%、5% 和 10%。

3) 系统的可靠性。测试目的是检查在噪声波动或者有冲击噪声干扰的情况下整机的跟随能力。采用标准的词表测试,噪声区间分别为 100~110 dB 和 110~115 dB。先使噪声以 3 dB/(5 s) 的速度从区间底值开始增大,达到区间顶值后,再以相同的速度减小到区间底值。在这个过程中,检测是否有噪声从系统通过。另外,分别在 100~110 dB 和 110~115 dB 区间内,随机加入一些幅度较大的短时噪声,检测是否有噪声从系统通过。经统计,在 100~110 dB 和 110~115 dB 区间内,系统的稳定性分别为 100% 和 98%。

可见,随着背景噪声等级的提高,整机系统的识别率和稳定性有所降低,丢字漏字率有所上升。如何在背景噪声提高和噪声波动速度较大的情况下保证整机系统的优良效果,尚需进一步研究。

4 有待深入研究和解决的问题

1) 在包含 VAD 模块的整机系统中,除电路和算法部分能够起到克服背景噪声的作用外,送话器性能也具有十分重要的作用,因此,应选用抗噪声送话器。这种送话器利用声平衡原理,可以把送入系统的噪声降低 12 dB 左右,并且具有很好的稳定性。这样就可以避免普通送话器因外界环境影响发生抖动而影响系统性能的情况发生。

2) 解决清音、辅音丢字漏字问题。一方面要做大量试验来调节参数使系统达到最优化,一方面在算法方面进一步深入研究。

由于清音和辅音的能量以及过零率与背景噪声

很接近,所以采用普通算法对这个问题的解决效果不是很理想,这就需要寻找并试验其他更加实用的算法,比如如何增强语音,通过寻找语音的其他特征来区别清音、辅音和背景噪声。

目前,从噪声中准确判别清音、辅音的常用方法有:1) 基于语音信号频谱坦度的判别法。这是一种基于信息论,在频域进行判别的方法,需进行滤波计算,运算量很大,所以实时性较差。2) 语音波形图像模板匹配法。基本原理是首先利用语音信号的波形图,导出清音、辅音的标准向量模板,然后将输入信号模型转换成向量图,与标准模板进行比较。此种算法复杂度较低,有利于实时应用。3) 通过汉语塞音选择性特征自动萃取的小波变换方法。

在下一步的工作中将参考以上方法,寻求更好的解决途径,对清音和辅音的检测进行深入研究,结合实际操作环境,来实现高噪声环境下对清音和辅音的准确判别。

参 考 文 献

- [1] 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(9): 1097-1099
- [2] 胡光锐,韦晓东. 基于倒谱特征的带状语音端点检测[J]. 电子学报,2000(10):95-97
- [3] Mark M, Birger K. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics [J]. IEEE Transactions on Speech and Audio Processing,2002(2):109-129
- [4] Juraj K, Gregor R. Speech signal detection in a noisy environment using neural networks and cepstral matrices [J]. Journal of Electrical Engineering,2004(5):131-137
- [5] Javier R, Jose S, Carmen B. Efficient voice activity detection algorithms using long-term speech information [J]. Speech Communication, 2004(2):271-287
- [6] Kim Nam Soo, Chang Joor Hyuk. Voice activity detection based on complex laplacian model [J]. Electronics Letters,2003(7):632-633
- [7] 何振亚. 自适应信号处理[M]. 北京:科学出版社,2002. 60-75