

文章编号:1009 - 671X(2001)03 - 0013 - 02

一种基于信息熵的语音端点检测方法

陈四根, 和应民

(哈尔滨工程大学 电子工程系, 黑龙江 哈尔滨 150001)

摘 要: 根据语音信号的波形特征, 利用熵函数的性质, 构造了一种特殊的熵函数, 通过判断此函数值的大小, 确定是语音还是无声状态。实验表明, 此方法计算简单而且具有很高的准确性。

关 键 词: 语音信号处理; 端点检测; 信息熵

中图分类号: TN912.3 **文献标识码:** A

A Scheme of Speech Endpoint Detection Based on Information Entropy

CHEN Si-gen, HE Ying-min

(Dept. of Electronic Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: A new scheme of speech endpoint detection based on information entropy was presented in this paper. We exploited the properties of the speech and the information entropy function, and constructed an entropy function adapted to speech signal processing. Experiments showed that the new scheme could obtain high scores on speech endpoint detection, and its complexity was very low.

Key words: speech signal processing; information entropy; speech end-point detection

0 引 言

在语音信号处理中, 语音端点检测是语音编码与语音识别的一个重要组成部分, 端点检测准确性对减少语音编码的比特数与正确地进行语音识别起到十分重要的作用。

在以往的研究过程中, 人们已经提出了许多端点检测方法, 其中比较经典的有语音短时能量、短时平均幅度等。但在应用过程中, 其效果都难尽人意, 这是因为语音信号中的清音与无声状态很相似。本文在研究了语音信号的波形特征后, 利用熵函数的性质, 提出了一种熵函数的判决方法, 为端点检测提供了相当满意的效果。

1 语音信号的形成及性质

在研究了发声器官和语音的产生过程之后, 通常把语音信号分成清音与浊音, 但从信号处理的角度出发, 语音信号产生统一为一个模型^[1], 即语音信号为激励源通过激励声道模型产生, $S(Z) = E(Z) \times H(Z)$, 发声模型如图 1 所示。

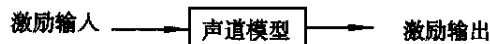


图 1 发声模型

其中 $S(Z)$, $E(Z)$, $H(Z) = 1/A(Z)$ 分别为语音信号、激励源和声道模型的 Z 变换。清浊音的主要区别是其激励源不同。对于浊音而言, 气流通过声门致使声带周期性的闭合, 形成周期性的脉冲气流, 这种周期性的脉冲气流就是浊音的激励源。气流激励声道产生浊音。因此, 浊音信号具有非常明显的周期性, 同时由于人的发声器官的限制, 这一周期又不会太小, 一般大于 2 ms, 而且, 相对于清音而言, 浊音信号幅度比较大。清音产生时声门完全闭合, 声道不受周期脉冲气流激励而是利用口腔内存的空气释放出来而发声。该气流通过一个狭长的通道时, 在口腔中形成湍流, 因而清音信号没有明显的周期性, 幅度小, 呈现出随机噪声的特点。清浊音对比如图 2 所示。

2 熵函数的构造

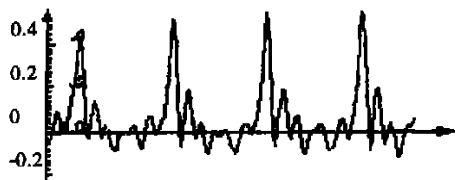
对于一个有限离散概率场的不确定性可由随

收稿日期: 2000 - 12 - 08

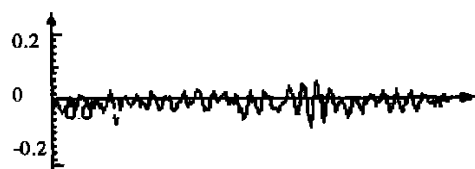
作者简介: 陈四根(1973 -), 男, 江西高安人, 哈尔滨工程大学电子工程系硕士研究生, 主要研究方向: 数字信号处理、语音信号处理与识别。

机变量的熵来定义^[2]. 设 X 是取有限个值的随机变量, $i = 1, 2, \dots, n$; 其中, n 为有限值, 则 X 的熵定义为

$$H(X) = - \sum_{i=1}^n P_i \log P_i; \text{且 } P_i = 0 \text{ 时, } P_i \log P_i = 0.$$



(a) 浊音波形



(b) 清音波形

图 2 清浊音对比图

熵 H 代表了 X 的信息量, 且熵函数具有这样的性质, 当 n 个概率相等时, 取最大值, 也就是 X 的概率分布越模糊, 越难以判断, 熵值越大。

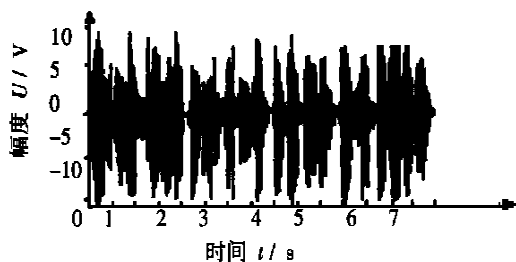


图 3 语音信号

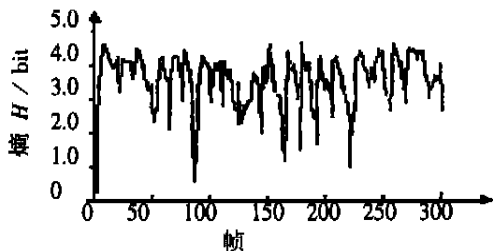


图 4 语音熵反映的特征

下面来构造用于清浊判决的熵函数。

假设语音 $S(n)$ 的帧长为 N , 在一帧语音中最大与最小幅度分别为 M , $-M$, 分别统计出其值等于属于这一区域的每一幅度值的比例作为出现这一值的概率。 $S_i \in (-M, M)$, n_i 为 $S(n) =$

S_i 的个数, 其概率为 $P_i = n_i / N$; 显然 $P_i = 1$, 故定义此帧语音的熵为

$$H = - \sum_{i=-M}^M P_i \log P_i$$

由于语音信号的幅度相对于背景噪声而言其幅度的动态范围大, 因此直观地讲, 可以认为随机信号在 $(-M, M)$ 中的随机事件多, 故平均信息量大, 也就是熵值大。而无声状态的幅度小, 分布相对集中, 因而熵值小。图 3, 图 4 分别是语音信号和熵值图示, 实验语音采样频率为 8 kHz, 处理帧长为 $N = 200$ 采样点。注意图 4 中纵坐标为熵值, 在峰值较低处才是无声帧。

3 结 论

针对语音信号自身的特征, 本文讨论了清浊音各自的发声原理及特征, 在研究了信息熵函数特点的基础上, 提出了一种利用熵函数的语音信号端点检测方法, 极大地提高了检测的准确性, 经过多次实验, 发现其准确率在 98% 以上。这种方法简单而且准确率高, 对于语音编码与语音识别具有十分重要的意义。

参 考 文 献

- [1] 杨行峻, 迟惠生. 语音信号数字处理[M]. 北京: 电子工业出版社, 1995.
- [2] 贾世楼. 信息理论基础[M]. 哈尔滨: 哈尔滨工业大学出版社, 1986.
- [3] 杨震, 毕厚华. ATM 网中语音编码的新方案[J]. 通信学报, 2000, 21(5): 23 - 29.

科技信息

英发明自动螺旋形运动吸尘器

英国 Dyson 电器公司生产的 DC06 全自动吸尘器装备有 3 台计算机, 据说它装备的感应器比登陆火星的机器人还多。今年它将在美国上市, 价格为 3500 美元。DC06 工作时先从墙边开始。然后就自动沿着房间四周做螺旋形运动。直到把整个房间打扫干净。它可以自动绕开障碍物, 甚至在楼梯边缘也能自动停下来, 避免摔跤。