

文章编号: 1002-0640(2004)06-0094-03

汉语语音识别中的一种音节分割方法

何 新, 王晓兰, 周献中

(南京理工大学, 江苏 南京 210094)

摘 要: 汉语语音识别研究中, 识别单元的选取是很重要的。随着大词汇量连续语音识别研究的发展, 越来越多汉语语音识别研究中选取次音节单位作为识别单元。结合汉语发音声学特性, 提出了音节的重叠音素分割策略, 并利用小波方法实现了音节的分割, 实验证明该方法分割准确可靠。

关键词: 音素; 语音识别; 小波变换; 突变检测

中图分类号: TN 912. 34

文献标识码: A

A Method for Syllable Segmentation in Mandarin Speech Recognition

HE Xin, WANG Xiao-lan, ZHOU Xian-zhong

(Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: It's very important to choose speech recognition units in Mandarin speech recognition researches. As the development of large vocabulary continuous speech recognition, sub-syllable is used as recognition units in many Mandarin speech recognition researches. In this paper, based on Mandarin acoustic characteristics, an overlapped segmenting scheme is presented and wavelets are used to segment phoneme. Experiments show that the method is available.

Key words: phoneme, speech recognition, wavelets transform, jump-detection

引 言

汉语普通话是以字为单位的。从声学角度看, 汉语中一个字就对应一个音节。因此, 音节是汉语语音识别中最自然的识别单元。目前, 随着汉语语音识别向大词汇量连续语音识别的方向不断发展, 音节作为识别单元已不再能满足要求。这是因为当词汇量增大时, 不可能要求在进行语音数据训练中每个音节出现很多次重复, 以得到可靠的结果。因此, 必须选取比音节更小的单元(音素)作为训练和识别的基本语音识别单元。基本的汉语语音识别单元一般应具备如下特点: 首先要具有良好的一致性, 即这种单元在不同情况下出现时应具有相似特性; 其次是可训练性, 即在训练数据中要有足够的重复出现次数。

那么如何合理地选取基本语音识别单元, 并建立相应的声学模型, 是汉语语音识别的关键。

汉语音节结构的显著特点是声-韵母结构, 因此将音节划分成声韵母音素作为识别单元符合汉语特点。将音节划分成声韵母音素后, 应对声母和韵母给予同等重视, 并对声母音素和韵母音素分别进行识别, 避免出现声母特征被韵母特征“淹没”的现象^[1]。同时, 应针对声母和韵母的不同特点, 分别采用不同的特征提取算法和声学模型将声母和韵母分开处理, 提高汉语语音识别系统的性能。此外, 这种基于音素的识别方法能使识别基元大大减少, 从而使运算量和贮存量减少, 而训练数据量却能相对增多。

基于音素识别单元进行汉语语音识别的基本问题之一是如何将音节分割成稳定性好的识别(音素)。由于汉语的特有结构使得其声韵母单元在声学信号上界限不够清晰, 所以, 要想通过音节的直接分割得到声韵母单元, 就存在一定的主观性, 导致分割的一致性比较差, 不适于实际系统采用。因此, 如何将声韵母单元提取出来是汉语语音识别系统训练和

收稿日期: 2004-01-03

作者简介: 何 新(1979-), 男, 山西太原人, 博士研究生, 主要从事语音信号处理;
王晓兰(1975-), 女, 博士研究生;
周献中, 男, 教授, 博士生导师。

识别的基础。针对这一问题, 本文提出了汉语音节的基础。针对这一问题, 本文提出了汉语音节的重叠音素分割策略; 同时利用小波方法能够检测信号突变的能力来进行音节分割; 最后通过实验室实验证明了使用小波方法进行音节分割的有效性和合理性。

1 重叠音素分割策略

1.1 问题的提出

汉语音节特有的 C+V 结构见图 1。

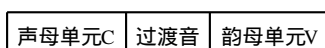


图1 汉语音节的构成

1。使得汉语音节的

声学单元组合具有一定的规律性。在这种结构中, 汉语音节的过渡音体现了音节的一种过渡性质。即在过渡阶段的起始段保留了较多的声母特性而韵母特性较少; 在过渡音末段则保留的声母特性较少而包含较多的韵母特性。正是由于这种声母与韵母的混合, 存在过渡音, 从而使音节中的声母和韵母之间的界线不是很清晰。所以想绝对地将声母和韵母分割开是不太可能的, 因此, 本文采用了音节的重叠音素分割策略。在实际的汉语语音识别系统中, 只希望分割得到的单元能保持良好的一致性即可, 而不需要其在声学模型和语言模型上严格一致, 所以, 将声韵母单元完全绝对分割是没有必要的。

目前, 常用的声韵母音素分割方法主要是利用分段 K-means 算法^[2]训练模型。该模型用单音节进行整体训练, 待模型收敛后人为地取其若干个状态为声母单元模型参数, 剩余状态为韵母单元的模型参数。另外一种方法是以固定的比例进行音节划分的, 例如文献^[3,4]取音节的前 1/3 作为声母单元。此外, 还有通过动态规划/多段矢量量化方法^[5], 通过状态划分来分割音节的。上述这几种分割方法实现比较简单, 但基本上都是基于采用声韵母“绝对”分割策略的基础上的, 这并不符合汉语发音的实际情况, 而且都存在分割上的随意性。

1.2 重叠音素分割策略

从声音的产生机理来看, 可以将语音的产生简化为喉、声道和嘴三部分的作用。其中, 在喉的发音中, 声门起到了重要的作用。在发浊音(韵母)时, 是通过声门的张闭运动产生周期性的脉冲气流进入声道, 而清音(声母)的产生是在声门关闭的情况下, 声道将口腔存有的空气释放, 从而形成发声, 这就是所谓有声门关闭(Glottal Closure, GC)事件。而从波形上看, 声母部分的变化很快, 没有明显的周期特征; 到了过渡音段, 逐渐呈现出周期性, 而到了韵母部分, 则呈现出明显的周期特性, 波形显得稳定而有规律; 最后能量逐渐减少, 但依然保持周期性的特征。因此, 任何一个汉语音节(零声母除外), 其声

门运动周期性呈现无序-基本有规律-有规律。所以只要能够检测到语音信号中的波形突变, 就可以获得声门运动的周期规律, 从而可以将音节分割成音素。

针对这一特点, 本文提出了用重叠音素分割方法实现音节分割, 以避免声韵母绝对化分割的缺陷, 使之更符合汉语发音的实际情况。

所谓音素重叠分割, 是指分割后的声母单元和韵母单元之间存在部分重叠, 即在声母单元的结尾和韵母单元的开始含有相同的过渡音。

设某一个汉字的帧长为 N , 其对应的语音信号为 $X = \{x_1, x_2, \dots, x_N\}$, 其中 x_i 表示第 i 个采样点。那么在此序列中取一点 x_c , 使得 $\{x_1, x_2, \dots, x_c\}$ 对应于声母部分(包含声母+部分过渡音, 或声母+全部过渡音, 或声母+全部过渡音+小部分韵母), 然后在 x_1 与 x_c 之间再取一点 x_v , 使得 $\{x_v, x_{v+1}, \dots, x_c, \dots, x_N\}$ 对应于韵母部分(包含绝大部分韵母, 或部分过渡音+韵母, 全部过渡音+韵母)。这样, 语音信号 X 分成了 A 段(即 $\{x_1, x_2, \dots, x_c\}$) 和 B 段(即 $\{x_v, x_{v+1}, \dots, x_c, \dots, x_N\}$)。为了简便起见, A 段和 B 段分别称为声母音素单元和韵母音素单元(如图 2 所示)。因为在 A 段和 B 段之间重叠了 $\{x_v, \dots, x_c\}$, 因而这种分割方法就称为“重叠音素分割”, 其中 x_c 即为语音分割点。在识别时, 就是先分别识别声母音素单元和韵母音素单元, 然后再将两部分的识别结果组合起来得到最终的识别结果。这样的分割方法就不存在“绝对”分割的问题了。

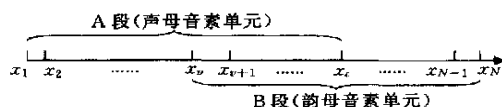


图2 音节中重叠音素示意图

此外, 由于在音节 X 的离散采样序列 $\{x_1, x_2, \dots, x_N\}$ 里声母单元和韵母单元是部分重叠的, 其重叠的部分为子序列 $\{x_v, x_{v+1}, \dots, x_c\}$, 其中 x_v 为包含全部韵母的子序列的起始点。从而有如下的定义。定义表征声母单元和韵母单元重叠程度的参数为重叠距离 ψ 。满足 $\psi = c - v$ 。

其中, c 和 v 分别代表 x_c 和 x_v 的下标号, ψ 体现了重叠序列的长度。

在重叠音素分割中最关键问题就是如何确定音素分割点 x_c 的位置。在识别过程中, 首先将 x_c 定在语音信号的周期性平稳段中的某一点, 再通过实验调整以获得重叠距离 ψ , 最后通过前面所定义的 c, v 和 ψ 的关系来确定点 x_v , 从而完成对一个语音信号的重叠音素分割。下面将主要介绍利用小波方法来确定 x_c 。

2 小波方法的音素分割

小波分析对信号突变的检测能力正是本文音素分割算法的理论基础。首先简单介绍有关小波突变检测的原理。

2.1 信号突变的小波检测原理^[6-10]

小波变换可以用于信号突变的检测, 其基本原理概述如下。

选择光滑函数 $\varphi(x)$ 的一阶导数作为小波函数

$\varphi(x) = \frac{d\varphi(x)}{dx}$, 则信号 $f(x)$ 的小波变换式可写为:

$$w_s f(x) = f(x) * \varphi(x) = f(x) * (s \frac{d\varphi(x)}{dx}) = s \frac{d}{dx} (f(x) * \varphi(x))$$

其中, $\varphi(x) = \frac{1}{s} \varphi(\frac{x}{s})$, $\varphi_s(x) = \frac{1}{s} \varphi(\frac{x}{s})$

$W_s f(x)$ 正比于 $f(x)$ 经 $\varphi(x)$ 平滑后关于时间 x 的一阶导数, 所以小波变换模量 $|W_s f(x)|$ 的局部极大值对应于 $f(x)$ 及其平滑量的尖锐变化点; 模量 $|W_s f(x)|$ 的局部极小值则对应于 $f(x)$ 及其平滑量的平缓变化点。因此, 对于 $f(x)$ 的一个阶跃来说, 小波变换沿 x 方向的局部极大值模量将沿伸缩系数 s 方向传播, 从而在比例—时间的相空间 (s, x) 中形成一条近似垂直于阶跃变化处且跨越若干伸缩尺度的局部极大值曲线。这就是小波对突变信号的检测原理。

2.2 基于小波方法重叠音素分割算法

在 1.2 节, 从识别的角度提出了重叠音素分割策略。这里首先确定出分割点 x_c 应满足的条件:

(1) x_c 应位于声门突变的稳定周期段;

(2) x_c 距离语音起始点应有足够的长度, 通常取 $60\text{ms} \sim 70\text{ms}$, 以保证浊辅音的长度。

由此可见, 要确定 x_c , 首先就是要先确定出声门突变的稳定周期段。由于受声门张闭的影响, 在周期开始时由脉冲气流激励会产生一个主要突变点, 其后还会产生一些次要突变点。本文就是利用小波的突变检测能力, 通过逐层小波分解, 去除次要突变点而检测主要突变点, 从而找出 GC 事件的准周期性来进行音节分割。

基于小波方法和重叠音素分割算法主要包括以下三步骤(如图 3 所示): 小波分解和预处理、声门关闭事件的多尺度确认、重叠音素分割点的确定。

其中, 在小波分解和预处理阶段主要是对不同层次上的小波分解信号的正极大值进行筛选, 求出相应的局部极大值, 去掉次要极大值, 从而找出不同层次的主要正极大值点位置, 即声门突变点; 再通过多尺度确认将得到的主要正极大值点进行确认, 分别计

算并比较不同尺度下的声门突变时间间隔, 如果相同就确认发生了 GC 事件, 否则就舍弃, 从而得到声门突变周期和 GC 事件发生的位置; 最后通过相应的周期长度参数计算确定出分割点 x_c (具体流程如图 4 所示)。

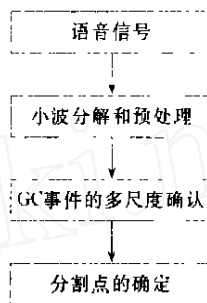


图3 重叠音素分割算法步骤

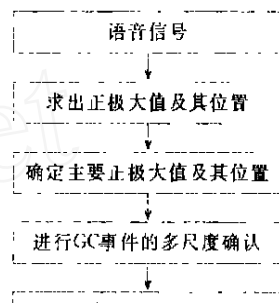


图4 分割算法流程图

4 实验

本实验主要是验证基于小波的重叠音素分割方法的有效性, 采用的小波是 db1 小波(即 Haar 小波), 这种小波的主要特点是正则性比较好, 可以更好地检测信号的突变。采用的语音信号采样频率为 11 025Hz, 16 位数字采样, 特定人在实验室环境下发音, 字表为汉语普通话全音节单字共 405 个, 每个单字发音 10 次, 共有样本 4 050 个。分别对这些样本利用基于小波的重叠音素分割法进行音节分割。观察声韵母单元的分割情况, 并进行人工判断分割正确与否, 分割标准是要求分割点应位于声门突变的稳定周期段。实验的结果如图 5 所示, 其中图 5 中所示的为音节“ba”的处理过程及结果。图 6 所示的为音节“pa”的处理过程及结果。

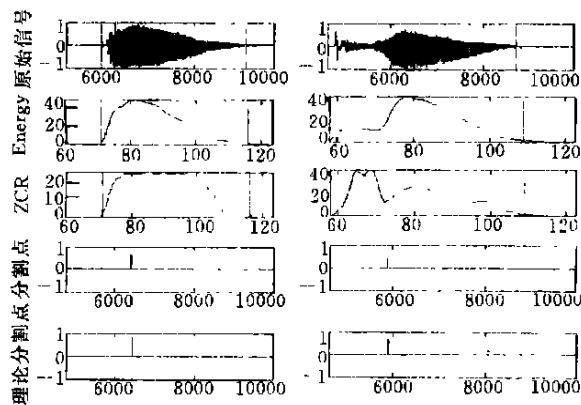


图5 音节“ba”的处理过程及结果

图6 音节“pa”的处理过程及结果

利用实验对语音样本进行基于小波的重叠音素分割, 结果表明运用小波方法进行音节分割是有效的。此外, 利用同样的语音样本还利用分段 K-means 算法进行了声韵母分割, 发现分段 K-means 算法不是对于所有的音节都合适的。

(下转第 99 页)

$$\begin{aligned}
 x_i &= \frac{1}{2} \cdot (x_i + x_{i+1}) + v_{x1} \\
 y_i &= \frac{1}{2} \cdot (y_i + y_{i+1}) + v_{y1} \\
 x_{i+1} &= \frac{1}{2} \cdot (x_i + x_{i+1}) - v_{x1} \\
 y_{i+1} &= \frac{1}{2} \cdot (y_i + y_{i+1}) - v_{y1}
 \end{aligned} \quad (10)$$

滚转角为:

$$\theta = \tan^{-1} \frac{z_{i+1} - z_i}{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}} \quad (11)$$

z_i 为与点 (x_i, y_i) 对应的高程, z_{i+1} 为与点 (x_{i+1}, y_{i+1}) 对应的高程。

该网格内的坡度为: $\alpha = \sqrt{\psi_i^2 + \theta^2}$

俯仰角、滚转角和坡度可用来判断作战单元在该段是否能通行, 通行速度等。

点 (x_i, y_i, z_i) , $(x_{i+1}, y_{i+1}, z_{i+1})$ 之间的距离为:

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2} \quad (12)$$

A (l_1, b_1) 到 B (l_k, b_k) 之间的行军距离为:

$$d = \sum_{i=1}^{MN} d_i \quad (13)$$

直线距离为:

$$d = \sqrt{(x_k - x_1)^2 + (y_k - y_1)^2 + (z_k - z_1)^2} \quad (14)$$

$$\text{平均坡度为: } \alpha = \frac{\sum_{i=1}^{MN} |\alpha_i|}{MN} \quad (15)$$

2.5 通视性计算^[3]

计算任意点 (l_i, b_i) 的高程 z_i 。

计算直线 AB (图 1) 在点 (l_i, b_i) 处的高程:

$$z_i = z_1 + \frac{\sqrt{(l_i - l_1)^2 + (b_i - b_1)^2}}{\sqrt{l_k - l_1)^2 + (b_k - b_1)^2}} \cdot (z_k - z_1) \quad (16)$$

z_1, z_k 为起点和终点的高程。

如果 $z_i > z_i$, 该点可以通视, 否则不能通视,

$z_i - z_i$ 为遮蔽部分的高度。

3 结 论

数字化部队作战过程中, 根据情况的变化, 要求指挥员快速地确定各作战单元的运动航迹, 航迹生成是数字化部队各级指挥控制系统重要组成部分。本文研究的地面机动目标通讯环境的可视化建模, 可以快速、直观地生成逼真的地面运动航迹, 已应用于实际的指挥控制系统。

参考文献:

- [1] 李增路, 李 真 飞机航迹自动生成技术[J]. 火力与指挥控制, 2002, 27(5): 50-52
- [2] 石 峰, 等 联合作战仿真应用中的想定系统框架[J]. 系统仿真学报, 2003(2).
- [3] 张野鹏, 等 作战模拟基础[M]. 北京: 解放军出版社, 1994

(上接第 96 页)

比如, 由于在 V iterbi 路径中选择的状态个数不同, 会出现无法取到合适的过渡音长度, 这不利于识别。

5 结 论

本文从识别的目的出发, 在汉语发声规律的基础上, 利用重叠音素分割策略合理地解决了汉语音节的划分问题。重叠音素分割策略不仅具有一般分割算法的优点, 而且能解决汉语发音中的过渡音归属问题。选择基于重叠音素分割策略的声学单元, 既符合人的听觉规律, 又避免了冗余信息对声母识别的负作用, 提高了声韵母划分的合理性。同时还利用小波方法作为音素分割工具, 并通过实验证明其可以准确地实现汉语声韵母重叠音素分割。

参考文献:

- [1] 杜利民, 侯自强 汉语语音识别中研究面临的一些科学问题 电子学报[J]. 1995, 23(10): 110-116
- [2] Raviner L R, et al A Segmental K Means Training Procedure for Connected Word Recognition Based

on Whole Word Reference Pattern [J]. AT&T Tech J. 1986, 165: 21-31.

- [3] 高慧琴 用 1/3 音节作匹配基元研究汉语单字音识别(Ⅰ)[J]. 电声技术, 1988(6): 13-16
- [4] 高慧琴 用 1/3 音节作匹配基元研究汉语单字音识别(Ⅱ)[J]. 电声技术 1989(1): 1-5
- [5] 张延平, 陈锡先, 蔡长年 一种新的全汉语单字音识别算法[J]. 信号处理 1992, 8(3): 143-151.
- [6] Mallat S, Zhong S Characterization of Signals from Multiscale Edges[J]. IEEE Trans, on PAM I 1992, PAM I-14(7): 710-732
- [7] Mallat S, Hwang M L. Singularity Detection and Processing with Wavelets[J]. IEEE Trans On IT. 1992, 32(2): 617-643
- [8] Mallat S Zero-Crossing of a Wavelet Transform [J]. IEEE Trans, on IT. 1991, 37(4): 1019-1033
- [9] 杨福生 小波变换的工程分析与应用[M]. 第 1 版 北京: 科学出版社, 2000
- [10] Hao Yu, Zhou Xiaoyan A New Feature in Speech Recognition Based on Wavelet Transform [J]. Proceedings of ICSP2000 2000: 1526-1529