

方言转换系统中的音节切分算法研究

张 帅¹, 贾 珈², 杨大利¹, 徐明星², 蔡莲红²

(1. 北京信息科技大学 计算机学院, 北京 100101;

2. 清华大学 计算机科学与技术系, 北京 100084)

摘 要:方言转换系统实现了普通话到济南话、沈阳话和西安话的实时语音转换。北方方言之间的差异主要体现在声调上,声调是属于音节的,因此声调转换模式转换是以音节为单位实施的。主要研究了方言转换系统中关键技术:连续语流音节切分算法。提出了一种基于自动机的逐级音节切分算法,分为语段切分、音节切分自动机和切分点自动校正三部分。该算法在误差48ms时,正确率达到72.55%,并成功支持了方言转换中的基频模式转换。

关键词:方言转换;音节切分;自动机

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2009)07-0041-03

Speech Syllable Segmentation Algorithm Research in Dialect Conversion System

ZHANG Shuai¹, JIA Jia², YANG Da-li¹, XU Ming-xing², CAI Lian-hong²

(1. Computer School, Beijing Information Science and Technology University, Beijing 100101, China;

2. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Dialect conversion system fulfill the real-time speech conversion from Mandarin to Jinan, Shenyang and Xi'an. The difference between the northern dialect is mainly reflected in the tone, the tone is based on syllable, so the unit of tone conversion mode is syllable. In this paper, continuous syllable segmentation algorithm was researched. A multi-level syllable segmentation algorithm based on automaton including speech segmentation segment, syllable detection automaton and automatic cut-off point correction is proposed. The segment correct rate is achieved for 72.55% when tolerance is 48ms. This algorithm supports the pitch conversion in dialect conversion successfully.

Key words: dialect conversion; syllable segmentation; automaton

0 引言

中国地域辽阔,汉语包括多种方言。在饭店、银行、公共交通、旅游景点等场所,利用方言提供个性化的服务,会使方言使用者感受亲切,增加客户满意度;在娱乐领域,如网上聊天系统、地方广播和娱乐节目中使用方言,可以丰富人与人之间语音交流的方式,增加趣味性。方言转换技术可实现方言之间语音的转换。

分析了汉语方言的特点,特别研究了北方方言之间的差异。研究表明,其差异主要体现在声调的基频模式上。面向普通话到沈阳、济南和西安方言的语音转换,构建了方言转换系统,如图1所示。三种北方方言和普通话之间,除了声调差异以外,还具有特殊词汇的差异,比如济南话中的“我”读作“俺”,沈阳话中的

“聊天”读作“唠嗑”等。因此对于输入的普通话语音,经过基频模式的转换和特殊词汇替换,输出方言语音。

汉语是声调语言,声调是属于音节的。文中的基频模式转换以音节为单位实施。因此在方言转换中连续语流的音节切分占有至关重要的地位。文中重点介绍方言转换中的音节切分算法。

在音节切分研究中,主要有基于隐马尔可夫模型(Hidden Markov Model, HMM)的强制对齐算法^[1,2]和声学特征阈值法^[3~5]。除了两种主流的切分算法,文献[5]提出了基于分形维数的切分算法,文献[6,7]提出了基于小波变换的音节切分算法。HMM强制对齐算法,缺点是需要训练,计算量大,不能满足实时性要求。声学特征阈值法主要利用特征参数(例如能量、过零率、能频值、基频、倒谱系数)阈值分割音节,缺点是准确率较低。优点是与语言无关,计算量小,不用训练。

收稿日期:2008-10-28;修回日期:2009-03-06

作者简介:张 帅(1984-),男,河北邢台人,硕士研究生,研究方向为语音识别;杨大利,副教授,研究方向为语音识别、图像处理。

文中在音节切分自动机^[4]的基础上提出了基于自动机的逐级音节切分算法,包括语段切分、音节切分自动机和音节切分点自动校正三部分。语段切分减少了音节切分自动机的计算量,提高了切分的准确率。通过切分点自动校正机制,减少了切分点的删除错误。逐级音节切分算法提高了切分的准确率,并成功地支持了方言转换中的基频模式转换。

方言转换系统结构如图 1 所示。

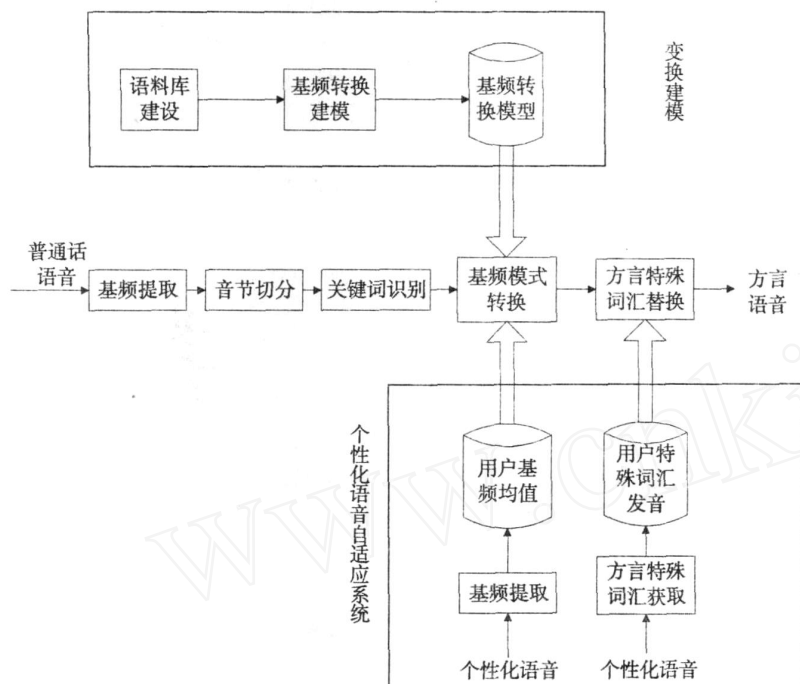


图 1 方言转换系统结构图

1 音节切分自动机

音节切分自动机首先利用能量、过零率来判断相邻帧之间有没有转折。如果能量和过零率发生突变,则将该帧定为转折点。没有发生突变的帧归并到同一个归并类似段中。

然后利用特征参数包括能量、过零率将归并段分为以下七类:静音、噪声、一类声母、二类声母(浊声母如 m, n, l 等以及零声母)、韵母、伪静音(声韵母过渡状态能量较低)和韵尾。七种不同的状态构成了音节切分自动机(如图 2 所示)。

由汉语的特点可知,当状态之间的转换满足表 1 中三种情况时,即可标注为音节切分点。

表 1 音节切分点的位置

	自动机前一状态	自动机当前状态
切分点 1	静音、噪声	声母、韵母
切分点 2	韵母	静音、噪声、声母
切分点 3	韵尾	声母、静音、噪声、韵母

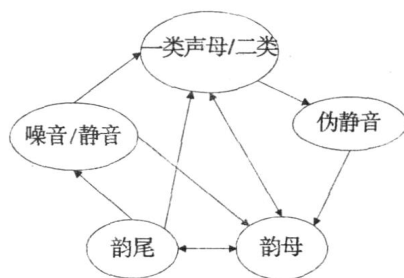


图 2 音节切分自动机

该算法能够有效地进行音节切分,切分效果较好,但是该算法也有一些不足。通过实验,发现了以下两个问题:

(1) 音节切分自动机对于静音的阈值选取是固定的。因此容易将能量较低的音节或者音节的尾部能量较低的部分,当作静音截断。

(2) 音节间能量和过零率没有明显突变的,切分自动机无法将其切开。

对于以上两种错误,都将直接导致基频模型的匹配错误,转换成错误的方言语音。

2 逐级音节切分算法原理

方言转换系统以音节为单位,通过修改基频模式来达到声调转换的效果。音节的错误切分会导致基频模型匹配错误,使转换出的方言语音和目标方言之间的相似度降低。文中在音节切分自动机算法^[4]的基础上,提出了基于自动机的逐级音节切分算法。改进了上节提出的音节自动机的两点不足。

逐级音节切分算法(如图 3 所示),分为语段切分、音节自动切分机、切分点自动校正三部分:

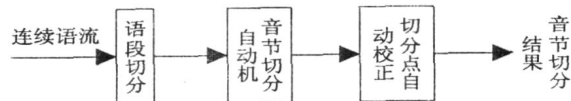


图 3 逐级音节切分算法

第一级语段切分,将连续语流划分为语音段和静音段。在阈值的选取上采用自适应阈值选取法^[3]。在开始录音时,会有短暂的间歇,说话人来不及说话,选取语音段的前 200ms,作为求取静音阈值的样本,实验表明,利用样本能量的均值加上 2 倍的标准差,作为静音的阈值效果较好。通过语段切分,减少了音节切分

自动机的计算量。动态选取静音的阈值,减少了音节切分错误率。

第二级采用2节介绍的音节切分自动机^[4]对语音进行切分,得到音节切分点。

第三级切分点自动校正。对音节切分自动机的错误切分点进行分析,发现以零声母开始的音节难以切分。原因是音节间能量和过零率没有明显的突变。通过观察发现(见图3),在一些语速并不是非常快的情况下,零声母与前一个音节之间,基频有短暂的断开,能量值处于低谷。具体修正原则为:如果基频曲线有断开,能量曲线低于阈值,则将其作为音节的结束和起始位置。

3 实验结果及分析

3.1 评测语料

评测语料库从本系统专用普通话语料库中选取。本次测评实验选取男女各100句,总音节数为4000多个。测试语速为正常语速。测试语句的采样率为16kHz,量化位数为16Bit,单声道。

3.2 音节切分评测算法

通过对音节切分评测算法的调研,发现文献中评测算法各异,很多算法并不能体现音节切分算法的性能。文中选用文献[8]提出的音节切分评测算法来对逐级音节切分算法进行评测。评测算法由切分正确率、删除错误率和插入错误率组成具体评测算法如下:

$$\text{切分正确率(Correct)} = \frac{C}{N} \quad (1)$$

$$\text{删除错误率(Del)} = \frac{D}{N} \quad (2)$$

$$\text{插入错误率(Insert)} = \frac{I}{N} \quad (3)$$

其中正确的切分点个数为 C 。音节切分点误差范围为 ms ,即一个自动切分点在手工切点的前后 ms ,则认为切点匹配正确。删除切分点个数为 D 。插入的切分点为 I 。 N 为手工标注的音节切分点总个数。

3.3 实验结果

对上述语料进行试验,当误差取如下几个典型值(16ms,32ms,48ms)时,实验结果如表2所示。实验数据表明,逐级音节切分算法总体上正确率高于音节切分自动机。实时性以及准确率达到了方言转换系统声调转换的要求。

对具体的实验数据分析也表明,音节切分自动机表现出来的问题也得到了有效的改进:

(1)如图5(a)所示,由于音节“西安”能量和过零率在音节间没有突然的转折,音节切分自动机未能切分。图5(b)经过对切分点进行校正得到了正确的切

分结果。

(2)如图5(c)所示,最后一个音节“出”能量较小。音节切分自动机由于选取固定的静音阈值导致结尾能量较低的音节被切掉。图5(d)通过逐级音节切分算法,最后的音节被正确保留。

表2 分级切分算法对比实验结果

偏差	切分算法	Correct	Del	插入率
16ms	切分自动机	58.18 %	41.82 %	26.64 %
	逐级切分	61.99 %	38.01 %	23.80 %
32ms	切分自动机	65.88 %	34.12 %	15.42 %
	逐级切分	68.53 %	31.47 %	17.26 %
48ms	切分自动机	70.37 %	29.63 %	14.13 %
	逐级切分	72.55 %	27.45 %	13.24 %

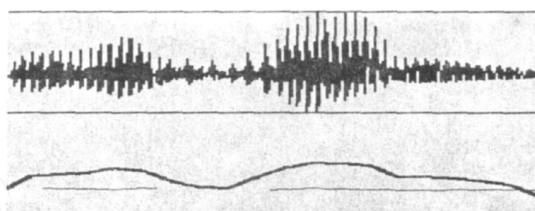


图4 “西安”语音图

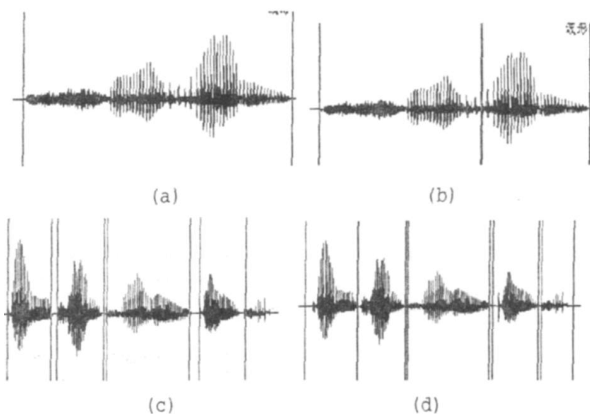


图5 音节切分自动机和逐级音节切分算法的切分效果对比

图5中,(a)为音节切分自动机切分结果,语音内容为“西安”,没能将两个音节切开;(b)为逐级切分算法切分示意图;(c)为音节切分自动机切分结果,末尾音节当作静音切除;(d)为逐级音节切分结果,正确保留结尾音节信息。

4 结论及展望

文中提出的逐级音节切分算法,提高了音节切分的准确率。成功的应用到方言转换系统中,为声调转换提供了音节信息。声调转换以音节为单位,通过基频模式修改达到声调转换的目的。

(下转第47页)

许的提高,尤其在节点运动比较剧烈存在相同节点间多次连接的情况下。当相同节点间之建立一个数据流时,改进后的协议与原协议性能相当。

参考文献:

- [1] Perkins C, Royer E, Das S. Ad Hoc On Demand Distance Vector (AODV) Routing[S]. RFC 3561, 2003.
- [2] 臧婉瑜,于 勤. 按需式 ad hoc 移动网络路由协议的研究进展[J]. 计算机学报, 2002, 25(10): 1009 - 1017.
- [3] Royer E M, Toh C K. A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Network [J]. IEEE Personal Comm., 1999(4): 46 - 55.
- [4] 何昆鹏,李腊元. Ad Hoc 网络中按需路由协议的仿真与性能分析[J]. 计算机技术与发展, 2008, 18(3): 81 - 84.
- [5] 王新生,张 昕. MANET 环境下 AODV 协议的研究和改进[J]. 微机发展(现名:计算机技术与发展), 2005, 15(12): 139 - 141.

- [6] 陈晓曙,李 霞. 一种高效的 Ad Hoc 网络 AODV 改进路由协议[J]. 东南大学学报:自然科学版, 2002, 33(2): 127 - 130.
- [7] 袁 柱. Ad Hoc 中 AODV 路由协议的改进[J]. 重庆工学院学报, 2007, 21(3): 77 - 79.
- [8] Marina M K, Das S R. Ad hoc On - demand Multipath Distance Vector Routing[C]// Proceeding of 9th IEEE International Conference on Network Protocols (ICNP). CA: Riverside, 2001.
- [9] 郑相全. 无线自组网技术实用教程[M]. 北京:清华大学出版社, 2004.
- [10] Belding - Royer E M, Perkins C E. Evolution and future directions of the ad hoc on - demand distance - vector routing protocol[J]. Ad Hoc Networks, 2003(1): 125 - 150.
- [11] 谢希仁. 计算机网络[M]. 北京:电子工业出版社, 2003.
- [12] 徐雷鸣,庞 博,赵 耀. NS2 与网络模拟[M]. 北京:人民邮电出版社, 2003.

(上接第 40 页)

- [2] 吴 蒙,季丽娜,王 莹. 无线异构网络的关键安全技术[J]. 中兴通讯技术, 2008, 14(3): 32 - 37.
- [3] 张 泉. 复杂卫星组网任务中卫星需求量解析模型研究[J]. 航空科学技术, 2006(6): 26 - 28.
- [4] Li Jia - qing, Zheng Feng, Zuo Zheng - jun. New Ad hoc reliable multicast transport protocol[J]. Application Research of Computers, 2008, 25(2): 581 - 583.
- [5] 汪小燕,王 浩,董 燕,等. 引入 ad hoc 方式的 LEO 卫星终端双模组网模型及性能分析[J]. 通信学报, 2006, 27(8): 139 - 143.
- [6] Zang Wan - Yu, Yu Meng, Xie Li, et al. A Survey of On - demand Routing Protocols for Ad hoc Mobile Networks[J]. Chinese Journal of Computers, 2002, 25(10): 1009 - 1017.

- [7] 刘发军,李冠宇,张 俊,等. 基于角色的分布式事务处理模型设计[J]. 微机发展(现更名:计算机技术与发展), 2005, 15(6): 25 - 27.
- [8] 关向凯. GSM 和 WCDMA 双模组网应注意的问题[J]. 通讯世界, 2002, 8(10): 98 - 99.
- [9] 宋晓婵,刘连忠. 异构操作系统环境的统一用户管理的研究[J]. 计算机技术与发展, 2008, 18(4): 144 - 146.
- [10] Hung Hui - Nien. Modeling of Ad - hoc and Infrastructure Dual Mode Mobile Networks[J]. International Journal of Automation and computing, 2005, 2(1): 75 - 84.
- [11] 韩 猛,钟广军. 一种适用于网络环境的三维网格压缩算法[J]. 计算机技术与发展, 2007, 17(12): 8 - 11.

(上接第 43 页)

由于语音信号的复杂和多变性,逐级音节切分的效果还不够理想。要想达到十分精确的切分,还需要从语音信号的本质去分析,进而完善切分算法。

参考文献:

- [1] Lee Chin - Hui, Soong F K, Juang Bing - Hwang. A segment model based approach to speech recognition[C] International Conference on Acoustics, Speech, and Signal Processing (ICASSP - 88). New York, NY, USA:[s. n.], 1988.
- [2] Toledano D T, Gomez L A H, Grande L V. Automatic phonetic segmentation[J]. IEEE Transactions on Speech and Audio Processing, 2004, 11: 617 - 621.
- [3] 张继勇,郑 方,杜 术,等. 连续汉语语音识别中基于归并的音节切分自动机[J]. 软件学报, 1999, 10: 1212 -

1216.

- [4] 林 帆,徐明星. 一种改进的基于时域参数的语音切分算法[J]. 计算机科学, 2006, 33(4): 164 - 170.
- [5] 罗世谦,冯子亮,张 恒. 一种基于能量聚类分析的句子语音端点检测法[J]. 计算机技术与发展, 2008, 18(4): 13 - 15.
- [6] 王 帆,郑 方,吴文虎. 基于多尺度分形维数的汉语语音声韵切分[J]. 清华大学学报, 2002, 42(1): 68 - 71.
- [7] 李永光,李雪耀. 基于小波变换的自动声/韵切分的研究[J]. 哈尔滨工程大学学报, 1998, 19(4): 75 - 81.
- [8] Villing R, Timoney J, Ward T, et al. Automatic Blind Syllable Segmentation for Continuous Speech[C] Irish Signals and Systems Conference. Belfast, Ireland:[s. n.], 2004.