

文章编号: 1004-4353(2018)03-0260-06

语音段中朝蒙单元音概率分布的对比分析

王桂荣, 金小峰*

(延边大学 工学院, 吉林 延吉 133002)

摘要: 为了从声学语音学角度比对朝鲜语与蒙古语的相似性,提出了一种利用相似单元音在连续语音段中出现的概率分布来对比分析语音相似程度的方法. 首先,采用双门限静音段检测方法将连续语音自动划分为一系列的语音段,然后采用模式识别方法检测和统计5对相似单元音在语音段首、段中和段尾的概率分布,最后通过对比分析概率分布比较了朝鲜语和蒙古语的语音相似性. 对比分析结果表明,朝鲜语与蒙古语语音之间存在差异性,这两种语音的声学特征不具有相似性. 本文研究方法和结果可为进一步研究朝鲜语与蒙古语的语音对比提供参考.

关键词: 语音信息处理; 声学语音学; 语音对比分析; 语音相似性度量; 静音段检测

中图分类号: TP391.42

文献标识码: A

Contrastive analysis of pure vowels probability distribution of Korean and Mongolian in speech segments

WANG Guirong, JIN Xiaofeng*

(College of Engineering, Yanbian University, Yanji 133002, China)

Abstract: In order to compare the similarity between Korean and Mongolian from the perspective of acoustic phonetics, a method for comparing and analyzing the degree of similarity of speech using a probability distribution of similar pure-vowels in continuous speech segments is proposed. First, the double-threshold silence segment detection method is used to automatically divide continuous speech into a series of speech segments. Then, the pattern recognition method is used to detect the probability distribution of 5 pairs of similar pure-vowels statistics in the head, middle and end segments of the segmented speech segment, and finally through comparison. Analyzing the probability distribution compares the similarity of Korean and Mongolian speech. The results of comparative analysis show that there are differences between Korean and Mongolian speech, indicating that the acoustic characteristics of these two speeches are not similar. The research methods and results in this paper can provide references for further study of the phonetic comparison between Korean and Mongolian.

Keywords: speech information processing; acoustic phonetics; contrastive analysis of phonetic; measurement of phonetic similarity; silence segment detection

0 引言

我国自古以来就是多民族聚居的国家,研究各民族语音对研究各个民族的历史具有重要的作用. 近年来,国内学者对少数民族语言的语音信息

处理进行了广泛研究,并取得了一定的研究成果. 例如:崔荣一等^[1]提出了一种基于基音频率特征的中国朝鲜族语言、韩国朝鲜语和朝鲜朝鲜语方言的自动辨识方法;朴明姬等^[2]提出了一种基于

朝鲜语数字串结构特点的识别方法;德格吉呼等^[3]对比分析了保安语与东乡语、蒙古语的元音声学参数,并研究了各语言元音的语音特点及其语音相关性;徐荣^[4]以元音为范畴,从发音方法和发音位置上对汉韩元音进行对比,分析了两种语音的共同点和不同点;许闰星^[5]通过实验语音学方法对比分析了蒙古语和朝鲜语单元音共振峰的值和韵律特征,并研究了朝鲜语与蒙古语发音相似音节的韵律特征.为了进一步分析朝鲜语和蒙古语之间的相似性,本文通过对朝鲜语和蒙古语中 5 对相似单元音 {l, ʁ}, {l, ɣ}, {l, ɯ}, {ɰ, ʁ/u}, {ɰ, ʁ/u}, 在连续语音段内出现的概率分布进行对比分析,以此研究两种语音的相似程度.

为了进一步提高对比的精准性,本文采用语音段检测方法将语音分割为一系列相对独立的连续语音段,并分别统计相似单元音在连续语音段的段首、段中和段尾 3 个部分出现的分布概率.由于目前缺乏语音和文本严格对齐的平行语料库,因此难以准确地提取语义上完整的句子.为此,本文根据连续语音中词、短语、句子以及段落之间存在的不同静音段时长,通过实验方法确定最佳的句子之间静音段时长,以此作为连续语音段分割的阈值.

1 基于静音检测的连续语音段分割方法

不同的研究目标其语音分割算法不同,而且对分割的精度也有不同的要求^[6];但是,无论何种级别的语音单位,其分割方法基本上都是通过对有声段和无声段(静音段)进行判别的基础上来处理的.语音检测是指从包含语音的一段信号中确定出语音的起始点和结束点位置.目前,语音检测方法主要包括以下 3 种方法.

1) 短时能量.因语音信号的能量随时间的变化较为明显,因此语音信号的能量分析可以用短时能量 $E(n)$ 和短时平均幅度来进行分析.短时能量反映的是语音振幅或能量随时间缓慢变化的特性,主要用于区分音素(如元音段和辅音段)以及无声和有声的分界.设经过加窗、分帧处理后的第 n 帧语音信号为 $y_n(m)$,则其短时能量 $E(n)$ ^[7] 的计算公式为:

$$E(n) = \sum_{m=0}^{M-1} y_n^2(m), 1 \leq m \leq M, 1 \leq n \leq N.$$

(1)

其中 M 为帧长, N 为总帧数.

根据式(1),首先计算语音信号 $x(m)$ 各个样点的平方值,然后通过一个冲激响应为 $h(m)$ 的滤波器,输出由短时能量构成的时间序列 $E(n)$:

$$E(n) = \sum_{n=-\infty}^{\infty} x^2(n)h(m-n) = x^2(m)h(m).$$

(2)

2) 短时过零率.短时过零率表示一帧语音信号的波形图中曲线穿过横轴(零电平或电平均值)的次数.由于它能够粗略地描述信号的频谱特性,因此利用短时过零率可判断无声段和有声段的起点和终点位置.在离散时间信号中,如果相邻的两个取样具有不同的代数符号,就会出现“过零”现象.音信号 $x(m)$ 的短时过零率的计算公式为^[7]:

$$Z_n = \frac{1}{2} \sum_{m=0}^{N-1} |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]|,$$

(3)

其中 $\operatorname{sgn}[\cdot]$ 为符号函数,其表示如下:

$$\operatorname{sgn}[x(m)] = \begin{cases} -1, & x(m) < 0; \\ 1, & x(m) \geq 0. \end{cases}$$

(4)

3) 双门限检测法.由于语音开始后总出现能量较大的浊音,此时单纯用短时能量或短时过零率不能准确地检查出语音信号,因此通常采用短时能量和短时过零率相结合的方法来确定语音信号的起始点和结束点位置,即双门限检测法^[8].双门限检测方法的具体实现步骤如下:

Step 1 根据式(1)和式(3)计算每一帧语音信号的短时能量和平均过零率.

Step 2 根据有声段短平均 $E(n)$ 确定一个高阈值参数 E_H 和一个低阈值参数 E_L ;根据无声情况下 Z_n 的均值设置一个阈值参数 Z_0 .

Step 3 根据 $E(n) > E_H$ 成立与否来确定某一帧语音是否为有声段,若成立标记此时的起点 \tilde{N}_1 和终点 \tilde{N}_2 .但此时的起点和终点不是精确的,需要进一步调整.

Step 4 由 \tilde{N}_1 向前找,当 $E(n)$ 减小至 E_L 时确定新的起点 N_1 ;类似地,由 \tilde{N}_2 可以确定新的终点 N_2 .可以肯定在 N_1 和 N_2 之间的信号仍属于

语音段。

Step 5 由短时过零率继续调整起点和终点。 N_1 向前搜索时,若 $Z_n > 3Z_0$, 则认为这些信号仍属于语音段;当出现 $Z_n \leq 3Z_0$ 时,则确定该信号为语音的精确起点;精确的终点采用相同的方法确定^[9];

Step 6 输出由精确的起点和终点所确定的有声段。

本文提出的基于静音检测的连续语音段分割算法的流程图如图 1 所示. 为了保证算法的起点和终点的准确性,以及分割出的每一段连续语音尽可能为一个完整的句子,需要设置合理的朝鲜语静音段时长阈值 S_K 和蒙古语静音段时长阈值 S_M . S_K 和 S_M 的值是通过反复调整阈值的方法来确定的. 如果阈值设置过高,分割结果可能多于一个完整的句子,反之分割结果可能是短语甚至是词. 当分割的静音段时长超过阈值 S_K 或 S_M 时,输出该语音段,否则记录当前的语音段并与上一次输出后不满足条件的语音段合并,直至超过时长阈值时输出该合并语音段。

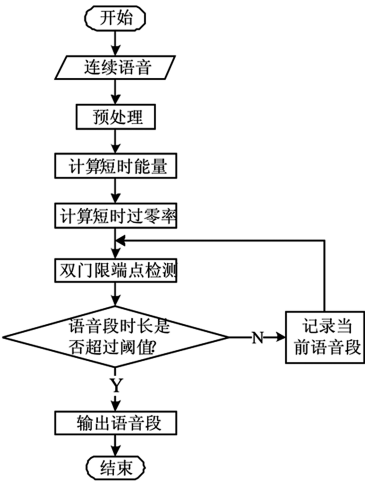


图 1 基于静音段检测的语音分割流程图

2 相似单元音概率分布的对比分析

基于 HMM 模型^[10] 的语音音素自动识别方法因具有较高的识别准确度和较强的可扩展性,目前被语音研究领域广泛使用. 由于本文要在连续语音中识别这些单元音出现的位置及其概率,因此采用模式识别的方法先对朝鲜语和蒙古语的相似单元音语音进行模型训练,以此得到每一对

朝鲜语和蒙古语相似单元音的 HMM 模型,然后采用识别的方法对分割的连续语音段进行检测,以此得到这些模型出现的位置和概率,从而进一步统计分析它们的分布概率. 朝鲜语和蒙古语的相似单元音概率分布对比分析过程如图 2 所示。

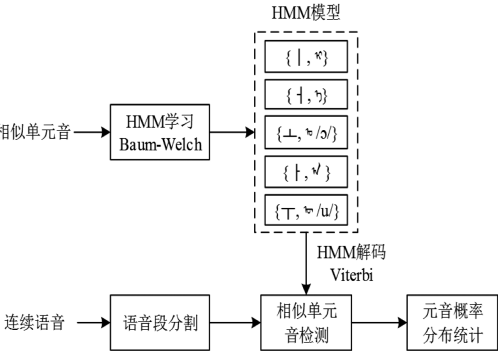


图 2 相似单元音概率分布对比的过程

相似单元音概率分布算法的具体步骤如下:

Step 1 对相似单元音进行 HMM 模型训练.

Step 2 基于静音检测的连续语音段分割算法从输入的朝鲜语和蒙古语连续语音中分割出连续语音段,将每个连续语音段的前 30% 定义为段首,后 10% 定义为段尾,中间 60% 定义为段中. 提取这 3 个部分的 MFCC 特征参数;

Step 3 利用 Viterbi 算法识别 Step 1 训练得到的 HMM 模型在 3 个划分段中出现的位置和概率,生成 5 对相似单元音的概率分布结果.

按照上述算法分别统计段首、段中和段尾等不同位置中相似单元音出现的位置和频数,并对对比分析朝鲜语和蒙古语相似单元音的概率分布情况。

3 实验及结果分析

首先确定连续语音段分割的最佳阈值,然后进行相似单元音分布概率的统计与分析. 所使用的朝鲜语和蒙古语语料为中国广播网“民族之声”、“全国新闻联播”和“新闻 30 分”等栏目的共计 180 min 语音(朝鲜语和蒙古语语音各 90 min). 使用 Audacity 软件对语料进行格式转换和重采样,采样率为 16 kHz,精度为 16 bit,采用 wav 文件格式存储。

3.1 连续语音段最佳分割阈值的选择

采用上述基于静音段的连续语音段分割方法,先对朝鲜语和蒙古语语料(各 90 min)进行静

音段时长的检测,并对时长为 20 ms 到 559 ms 的静音段进行划分,划分间隔为 10 ms,统计结果如图 3 所示。

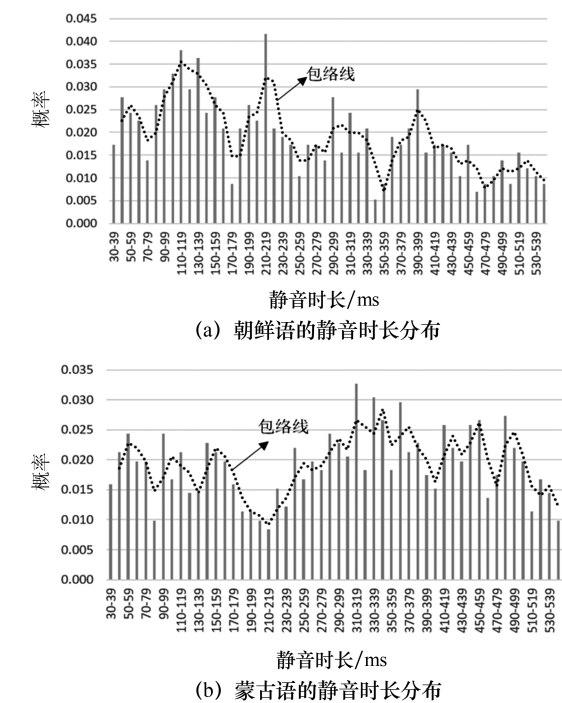


图 3 朝鲜语和蒙古语的静音时长分布

假设连续语音中静音段是由词与词、短语与短语、句子与句子、段落与段落等 4 种类型的停顿引起,则这 4 种类型静音段在图 3 中的包络线会各自呈现出准高斯分布特点,尤其是专业播音员朗读的新闻稿,其特点会更加明显。将各高斯分布之间的谷值作为朝鲜语和蒙古语的静音段阈值 S_K 、 S_M 的候选值,然后通过静音段划分实验确定最佳阈值。通过对比图 3(a) 和图 3(b) 发现,朝鲜语和蒙古语各类型静音段时长阈值不存在严格的对应关系。

利用候选的静音段阈值 S_K 和 S_M 对朝鲜语和蒙古语连续语音进行分割实验,结果见表 1。表 1 中,准确率是用自动化分的段数除以人工划分的段数后乘以 100% 得到的。显然,自动划分的段数不应大于人工划分的段数,因此在表 1 中将准确率为 4.6% 以下的标记为“—”。

从表 1 可以看出,朝鲜语在静音段阈值 $S_K = 176$ ms 时,准确率达到最大(92.8%);蒙古语在静音段阈值 $S_M = 219$ ms 时,准确率达到最大(91.2%)。

表 1 连续语音段分割实验结果

朝鲜语				蒙古语			
分割阈值 S_K	自动划分段数	人工划分段数	准确率/%	分割阈值 S_M	自动划分段数	人工划分段数	准确率/%
78	6 616	5 035	—	85	6 418	4 010	—
176	4 673		92.8	135	4 590		—
256	1 663		33.0	219	3 658		91.2
342	895		17.8	400	809		20.2
465	506		10.0	480	584		14.6
542	233		4.6	519	396		9.9

3.2 相似单元音分布概率的统计分析

采用上述相似元音概率分布对比的分析方法分别对朝鲜语和蒙古语的 5 对相似单元音的概率分布情况进行统计,结果如表 2 所示。

依据表 2 绘制的朝鲜语和蒙古语各相似单元音在段首、段中和段尾的分布概率如图 4 所示。

通过对比表 2 和图 4 发现:

1)朝鲜语和蒙古语中的单元音/a/的概率分布最为相似,其他依次为/e/、/i/、/o/和/u/。

2)段首概率分布中,朝鲜语单元音/o/的概率最高(25.5%),蒙古语单元音/i/的概率最高

表 2 相似单元音在连续语音段中的概率分布

单元音	语种	概率分布/%		
		段首	段中	段尾
/a/	朝鲜语	18.5	49.0	32.5
	蒙古语	16.5	52.8	30.7
/e/	朝鲜语	20.3	60.6	19.1
	蒙古语	13.6	63.4	23.0
/i/	朝鲜语	15.0	53.5	31.5
	蒙古语	23.9	61.0	15.1
/o/	朝鲜语	25.5	52.6	21.9
	蒙古语	14.3	61.5	24.2
/u/	朝鲜语	20.4	50.1	29.5
	蒙古语	12.0	35.5	52.5

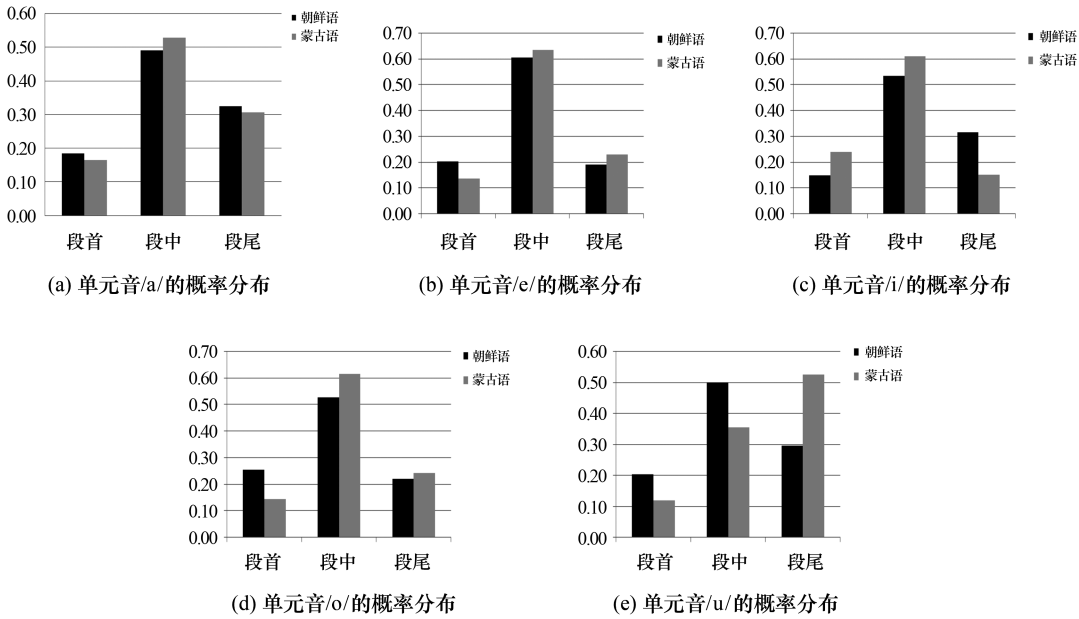


图 4 相似单元音在不同位置出现的概率分布

(23.9%),这表明朝鲜语句子起始若干个音节中以/o/构成的音节多,而蒙古语句子起始若干个音节中以/i/构成的音节多.

3)段尾概率分布中,朝鲜语中 3 个单元音的概率都较高(/a/ 32.5%、/i/ 31.5%和/u/ 29.5%),表明朝鲜语句子末尾的若干音节中以这 3 个单元音构成的音节较多.而蒙古语中单元音/u/的概率最高(52.5%),表明蒙古语句子末尾的若干音节中以/u/构成的音节较多.

对整个语料中的相似单元音的总体分布概率进行统计,结果如表 3 所示.

表 3 相似单元音的总体分布概率

单元音	分布概率/%	
	朝鲜语	蒙古语
/a/	25.6	27.0
/e/	19.2	21.3
/i/	25.4	13.8
/o/	17.5	12.8
/u/	12.3	25.1

根据表 3 数据绘制的朝鲜语和蒙古语 5 对相似单元音的总体概率对比曲线如图 5 所示.结合表 3 和图 5 可以看出,朝鲜语和蒙古语在整个语料中,/a/和/e/出现的概率基本相同,其他 3 个单元音表现出差异性,尤其是/i/和/u/的概率差异最为显著.

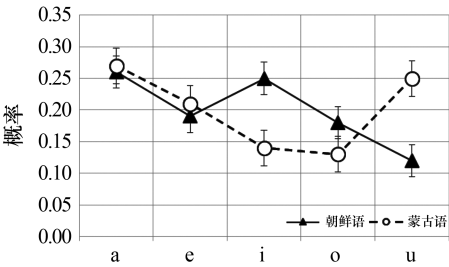


图 5 相似单元音的总体分布概率

4 结论

本文将朝鲜语和蒙古语中的 5 对相似单元音作为研究对象,提出了一种基于概率分布统计和对比的朝鲜语和蒙古语语音相似度分析方法.通过相似单元音在连续语音段段首、段中和段尾出现的概率对比分析发现,朝鲜语和蒙古语中的单元音/a/的概率分布最为相似,其次为/e/、/i/和/o/,而/u/的概率分布差异较大;在整个语料中对相似单元音的总体分布概率统计分析发现,/a/和/e/出现的概率基本相同,其他 3 个单元音表现出差异性,尤其是/i/和/u/的概率差异最为显著.以上结果表明,朝鲜语和蒙古语的语音不具有相似性.本文研究方法和结果可为进一步研究朝鲜语与蒙古语的语音对比提供参考.本文在研究中只对朝鲜语和蒙古语的单元音系统进行了相似性对比分析,在今后的工作中将考虑辅音和复合元

音等音素及其韵律特征的对比,以便更加全面地对比朝鲜语和蒙古语的语音相似性.

参考文献:

[1] 刘双君,金小峰,崔荣一. 基于基频的朝鲜语方言辨识方法的研究[J]. 中文信息学报,2017,31(2):55-60.

[2] 朴明姬,芦世丹,崔荣一. 多位朝鲜语数字语音识别方法的研究[J]. 延边大学学报(自然科学版),2011,37(4):337-341.

[3] 德格吉呼,巴图格日勒,金玲. 保安语、东乡语、蒙古语单音节短元音声学对比分析[J]. 丝绸之路,2015,297(8):54-55.

[4] 徐荣. 汉韩元音对比与韩语语音教学[J]. 高教学刊,2017,66(17):81-83.

[5] 许闰星. 朝鲜语-蒙古语单元音和韵律特征的实验语音学对比研究[D]. 延吉:延边大学,2013:42-68.

[6] Juneja A, Espy-Wilson C. Speech segmentation using probabilistic phonetic feature hierarchy and support vector machines [C]//International Joint Conference on Neural Networks. Killarney, Ireland, 2003:675-679.

[7] 吕卫强,黄荔. 基于短时能量加过零率的实时语音端点检测方法[J]. 兵工自动化,2009,28(9):69-70.

[8] 薛胜尧. 基于改进型双门限语音端点检测算法的研究[J]. 电子设计工程,2015,23(4):78-81.

[9] Mporas I, Ganchev T, Fakotakis N. Speech segmentation using regression fusion of boundary predictions[J]. Computer Speech & Language, 2010, 24(2):273-288.

[10] Ranjard L, Reed B S, Landers T J, et al. MatlabHTK: a simple interface for bioacoustic analyses using hidden Markov models[J]. Methods in Ecology & Evolution, 2017,8(5):8-12.

—————
(上接第 249 页)

参考文献:

[1] Fang Xinjian, Liu Xiangwei, Fang Xianwen. An effective constraint-aware optimization method for university course timetabling problem[J]. Bio Technology: An Indian Journal, 2014,10(9):4001-4006.

[2] Chen Hefeng, Wu Naiqi, Zhou Mengchu. A novel method for deadlock prevention of AMS by using resource-oriented Petri nets[J]. Information Sciences, 2016,1:178-189.

[3] Ye Z, Wang S, Zhao T. IMA dynamic reconfiguration modeling and resource criticality analysis based on Petri net [C]//ICRSE 2017: the 2nd International Conference on Reliability Systems Engineering. Beijing: IEEE, 2017:1-6.

[4] 侯春龙,齐新战,卫翔. 基于 Petri 网建模的互斥问题优化方案[J]. 系统仿真技术,2012(3):238-243.

[5] 秦奋涛. 一种资源分配方案及其 Petri 网模型[J]. 河北北方学院学报(自然科学版),2005(1):24-25.

[6] 罗冬梅. 进程互斥算法的 Petri Net 实现[J]. 青岛大学学报(自然科学版),2015,28(4):77-79.

[7] Smirnov S, Weidlich M, Mendling J. Business process model abstraction based on behavioral profiles[C]//International Conference on Service-Oriented Computing. Berlin, Heidelberg: Springer, 2010:1-16.

[8] 方贤文. Petri 网行为轮廓理论及其应用[M]. 上海:上海交通大学出版社,2017:1-16.

[9] 吴哲辉. Petri 网理论[M]. 北京:机械工业出版社,2006:1-28.

[10] 汤小丹. 计算机操作系统[M]. 西安:西安电子科技大学出版社,2014:47-60.