

从符号学看标点符号在语音文本中的技术实现

刘健刚 董静

(东南大学 外国语学院, 江苏 南京 210096)

摘要: 符号学研究已成为一种国际性的研究热点, 但是真正运用到计算机在对语音文本识别实验中较为鲜见。符号学在计算机语音识别实现中的研究, 成为新的观照。语言是一种符号, 标点是语言符号系统中的个体符号, 有其特有的能指和所指。长期以来我国对标点的研究大多基于文字符号的诠释, 缺乏声音符号的诠释。文章通过对声音符号的诠释, 选择语音文本中的语音信号特征参数, 识别出静音段和语音段, 从而达到识别语音文本中的标点符号。

关键词: 符号学; 标点; 语音文本

中图分类号: H017

文献标识码: A

1. 引言

美籍语言学家罗曼·雅各布森曾经说过: “一个信息都是由符号构成的”^{[1]129}。符号是意义活动(表达意义与理解意义)的必须而且独一无二的工具, 不用符号无法表达任何意义; 反过来, 任何意义必须用符号才能表达, 因此, 符号学即是研究意义活动的学说。

在西方, 古希腊哲学的集大成者亚里士多德曾探讨了语言符号问题。他在《诗学》和《修辞学》中提出了区分有意义符号和无意义符号的主张, 他实际上创造了科学语言的基础^{[2]13}。在东方的中国, 早在东周时期便开始对汉民族独特的语言符号系统——汉语和汉字进行了研究, 并诞生了《说文解字》这部解释古汉语文字的不朽之作。《说文解字》中蕴藏着丰厚的符号学思想, 把汉字作为一个符号系统来理解和阐释是《说文解字》中体现的语言文字思想的核心^{[3]3}。

第二次世界大战后创建的控制论、信息论等具有跨学科倾向的新学科, 不仅直接包括与符号学通讯部分相关的内容, 而且成为一些当代符号学思想的一般理论基础之一^{[3]4}。

2007年, 广东省首次实施普通高校英语口语网考(粤招【2006】22号); 2008年12月20日, 我国53所高校实施四、六级英语网考试点考试; 自2010年起, 江苏省的中考也开始试行英语口语网考。英语网考加入了英语口语测试部分, 成功地将英语口语考试引入我国无纸质考试的语言考试的实践之中。口语机考中语音识别实现的可能性犹如一石击水, 在教育界和科技界引起激起阵阵涟漪。人们发现, 新世纪伊始, 托福就着手探索网考(internet-based test, 简称iBT), 并从2005年8月起, 在全球试用、启用iBT-TOEFL考试^{[1]8}。渐渐人们进一步发现, 早在1997年7月, IBM就向世人推出了ViaVoice连续语音识别系统的中文版本识别平台, 表明该系统能够使人们把一篇文章和想说的话“朗读出来”(还要说出标点符号!)^{[4]26}。

总体来说, 我国计算机语音科研人员已经开始了对于符号中的“个别符号”——标点进

项目来源: 教育部人文社会科学研究一般项目“英语口语机考评分系统的建模与应用”(10YJA740061)资助; 东南大学教学改革项目“大学英语四、六级网考模拟网站的建模与分析应用”(2010-54)资助; 江苏省高等教育教改研究立项重点项目“基于英语口语网络测试平台建设的教改模式研究”(2011JSJG453)资助。

行了开发研究,但是作者拙见认为,目前的技术实现仅为“逆向”实现技术完成,“巧夺天工”。在技术实现手法上,到目前为止,基本上还停留在“字和词处理”阶段(如孤立词识别,连续识别等),对于句子的研究尤其是复句的计算机语处理相对较少^{[5]1}。即使如此,对句子识别(让文章的话“朗读起来”)也仅仅主要停留在TTS(Text-to-Speech)的研究上,通过语音合成技术对话语句子进行分词标注,然后自动处理适当的标点符号。但是,在对STT(Speech-To-Text)标点符号识别的研究上尚存空白地段。在语音识别过程中,如何在STT实现智能分词过程后自动处理生成标点符号呢?如果目前STT技术不能进行标点符号的实现,我们能否跳出常规技术思路和方法,直接使语音文本生成标点符号呢?为此,我们有必要从符号学的角度来探讨在语音识别过程中要解决的关键问题:标点符号识别的实现方法与实现途径。

2. 符号学与标点的定位与研究

符号是意义活动(表达意义与理解意义)的必须而且独一无二的工具,不用符号无法表达任何意义;反过来,任何意义必须用符号才能表达,因此,符号学即是研究意义活动的学说^{[6]169}。研究符号学,首先要对符号概念本身有一个正确认识。现象学派认为,符号的功能作用就是它的物质化,符号是指认识主体的模式行为、直观形象方面的信号,即符号就是信号或特征;逻辑—心理学派则认为,符号是指意念的或功能的结构,这种结构对于其物质方面是漠不关心的,也就是说,符号的功能就是用事物内容和意思内容来充实它。显然,上述两个学派对符号的理解均有偏颇。与之相比,双重意义学派的观点较为科学,它把符号的物质性和思想性有机地统一起来,得到了大多数学者的认可^{[3]1}。

语言是客观存在的符号体系,包括符号和符号的组合规则。用语言说话(言语)即产生话语。不掌握形成话语的语言符号及其组合,无异于无本之木^{[7]219}。长期以来,人们对语言符号的研究大多停留在文字符号的狭义范围内:“文字符号可以称作是语言符号的符号,它是记录语言的符号,是语言的书面形式,是在语言产生后经过非常漫长的阶段才发明的,它的存在只是为了记录前者,表现前者^{[8]37}。语言的物质实体是声音。语言符号即声音符号,可称语音符号。语音符号的记录是文字,称为文字符号。语音符号组成口头语言。文字符号组成语言的书面形式,即语言的书面记录。从语言发展的源头上说,语音符号作为语言符号的本体,应处于研究的首位。形象比喻说,一个人可以不识文字,但却能言语,我们可以说这个人已掌握了语言。再证:母语的习得就是先口语后书面语。

标点符号同文字符号一样都是书面语的组成部分,与文字符号差不多同样古老。可是古往今来,人们对标点符号的学习、研究与应用远不如对文字符号那样重视。“标点”在我国古代有时又称“句读”和“圈点”。根据《现代汉语词典》的解释,所谓“句读”,古时称文词停顿的地方叫句读(dòu)。连称句读时,句是语意完整的一小段,读是句中语意未完,语气可停的更小的段落。所谓“圈点”,就是在书上加圆圈或点,作为句读的记号,或用来标出认为值得注意的语句。所谓“标点符号”,就是用来标明句读、语气和专名的书写符号。很明显,以上描述都是基于文本符号的诠释。在留声机发明以前,人们只能通过文字的形式来记录发生的事件。随着科技的发展,声音记录越来越显示出其记录事件的重要性,也就出现了“文字文本”和“声音文本”的符号形式。那么声音文本的诠释应该如何解释呢?

人类的话语中不仅包含了文字符号信息,还包含了语音符号信息,更包含了人们的感情和情绪等信息。表示停顿是标点符号的基本性质和作用。因此,计算机语音科研人员通常的做法就是基于“语音停顿也被认为是有声语言的标点符号”^{[9]542}这一特性以及寂声来表

示标点信息(见图1)。关于语音停顿,现在能被普遍认可的说法是:“说话时语音上的间歇”;“朗读语流中声音的中断”;它是有声语言的标点符号^{[10] 30-31}。

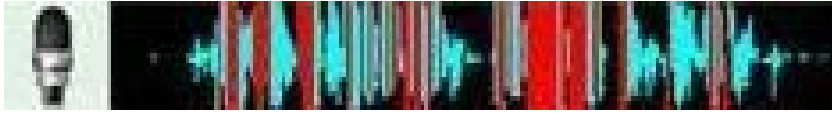


图1 声音文本中语声段和寂声段的关系示意图

3. 语言符号的三个重要特性

20世纪前期,“索绪尔式语言学”在叶尔慕斯列夫、马丁奈、班维尼斯特等大家手中迅速成熟,符号学在1960年代起步发展时,语言学为之提供了清晰的理论框架。语言符号学有三大原则:语言符号的任意性(arbitrariness)、语言符号的线性(linearity)和语言符号的分节性(articulation)^{[11]1}。

索绪尔将语言的线性特征表述为: The signifier, being auditory, is unfolded solely in time from which it gets the following characteristics (1) it represents a span, and (2) the span is measurable in a single dimension, it is a line (Saussure, 1986). 很明显,在以文字符号文本为主导研究的环境下,但是语言的线性特征经常被语言学家们所忽视,因为线性特征在文字符号文本中的表述比较累赘和复杂。相比之下,听觉能指可自由支配的只有时间上的一条线。这些要素是一个接一个得呈现,并形成一条链条(图2)。

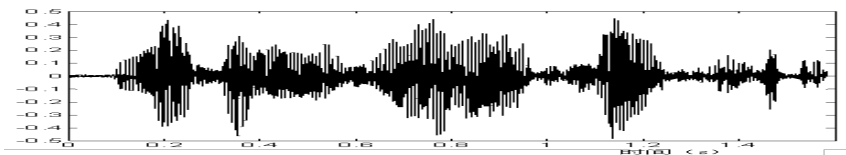


图2 “Nothing was really accomplished” 时间波形线性特征示意图

郭鸿教授根据以上两个特性,提出了语言的分节特性(articulation)^{[11]2}。Daniel Chandler 在 Semiotics for Beginners 中写道:“具有‘双重分节’(double articulation)的符号编码有两个结构层:高一层次的叫‘第一分节层’,低层次的叫‘第二分节层’。在第一分节层上,符号系统是由最小的意义单位组成的(如语言中的词素或单词)。……在第二分节层面上,符号分成最小的功能单位。它们本身没有意义(例如口语中的音素和笔语中的词素)。例如。在语言中, /h/, /p/ 和 /t/ 是第二分节层的,它们的功能是区别语言的第一分节层的单词 bin, pin 和 tin……。英语只有40个到50个第二分节成分,但这些有限的成分能产生几十万个语言符号”。这种语言的分节特性可以用语谱图淋漓之至地示意出来(图3)。

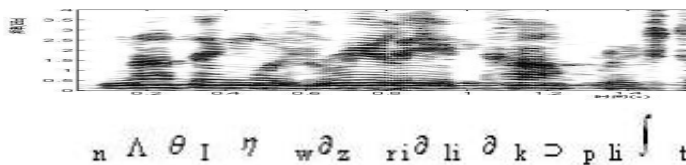


图3 “Nothing was really accomplished” 语谱分节特性示意图

4. 标点符号计算机智能分析和处理的习得

现代语言学的开创者索绪尔(Saussure)指出,能指即音响形象(sound—image)。音响形象并非单纯的声音,而是心中所理解并形成的固定的语音形象。音响形象之成为语音,是因为代表或意指一定意义,即其所指。它是一定的语言群体约定来代表某一意义的特定声音。声音没有意义代表的约定,就不成其为语音。也正是有了意义代表的约定,语音也才能在心中固定,形成一定的形象。语音所指和其能指音响形象之间的关系,是约定的音义关系。口头语言的听说能力的培养,所需的第二要素,就是这种音义关系的牢固建立^{[7]219}。

口语的一大特点就是不带有标点符号,没有明确句子的边界^{[12]2}。因此,在口语标点符号确认过程中,在一个无标点长句的各个词语之间,有停顿的可能性大小可以用该处出现标点的可能性大小来估计。国家自然科学基金资助项目(60572159、60573184和60473139)研究中曾对各个标点处出现的语音无声段(寂声段)的长度进行试验和验证,表明各标点处的语音无声段平均时长都在600ms以上,它们明显大于语料所有音节间语音无声段的平均时长,因而说明了书面语中的标点符号能够表示停顿的性质——无声(寂声)时间长于有声时间(见表1)。

表1 标点处的语音无声(寂声)段性质

标点符号特性	句末标点				句内标点			部分标点	
标点符号展示	.	?	!	,	,	;	:	---	—
无声(寂声)段时长范围	1315- 2294	513- 2800	433- 3311	106- 1481	165- 2277	566- 2764	404- 2803	637- 2371	303- 1008
无声(寂声)段平均时长	1794	1088	1132	628	738	1066	882	1205	711

5. 英语寂声段和语声段在语音文本中的判别方法

对语音文本的寂声段和语声段进行判别需要对语音信号的一些重要特征参数进行选择。这些参数主要是语音信号的统计特性,如幅度、能量、准周期性、过零率和频域特性等。本文根据赵力教授撰写论文中的方法选择语音文本中的语音信号特征参数^{[13]148-150}。选取短时平均幅值 M 和幅值的方差 V 、短时平均过零率 Z 和过零率的方差 U_k 作为识别特征参数。但是,实验结果告诉我们,仅仅利用短时平均幅值、平均幅值的方差、短时平均过零率和过零率的方差4个参数中的任意一个都不能很好地完全寂声段和语声段的识别,必须同时利用4个参数来综合特性然后进行寂声段和语声段的识别,这样实验的效果才佳。

5.1. 实验对象

实验中使用的数据是由10名在读工科研究生(6男,4女),每人对210句子各发音一遍得到的2100个句子,其中,1000个句子作为学习用数据,1100个句子作为识别用数据。句子的平均朗读速率为8.2字/秒,数据采样频率为12KHZ,窗长21.33ms(256点),窗移10ms。

5.1. 实验对象

5.2. 实验条件

本实验主要根据英语四、六级网考的实际情况,在语句上附加噪声,同时在语句两端也各附接一定长度的噪声,然后,从中检测出语声段。噪声是由计算机产生的随机噪声,SNR分别为20db、10db、5db、0db。语句两端附接噪音的长度为1秒。

5.2. 实验条件

5.3 实验过程与结果

本实验首先对于各类别的学习数据求出其相关矩阵,然后求出相关矩阵的特征值和特征向量,再通过主元素分析来决定隐射矩阵,最后把隐射量用特征矢量的大小正规化后,按下式与阈值 θ 比较,大于阈值 θ 的判定为语声段,小于阈值 θ 的判定则为寂声段:

5.3 实验过程与结果

本实验首先对于各类别的学习数据求出其相关矩阵,然后求出相关矩阵的特征值和特征向量,再通过主元素分析来决定隐射矩阵,最后把隐射量用特征矢量的大小正规化后,按下式与阈值 θ 比较,大于阈值 θ 的判定为语声段,小于阈值 θ 的判定则为寂声段:

$\frac{X'BX}{X'X} > \theta$ 。实验结果如表 2 所示（示意图见图 4）。

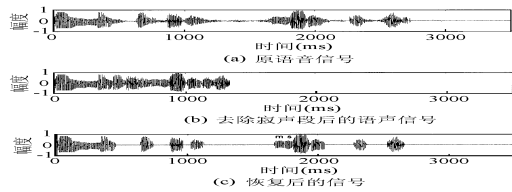


图 4 寂声段/语声段分离效果示意图

表 2: 寂声/语声段判别实验结果

SNR	20db	10db	5db	0db
始端	100	100	98.7	97.7
终端	100	100	98.3	96.7
平均	100	100	98.5	97.2

在上述实验结果中，判别是逐帧进行的，并且端点检测误差在 10 帧之内即认为判断正确。从表 2 数据看寂声段/语声段判别效果是较好的，平均最低值达 97.2%；从图 4 可以看到寂声/语声段分离效果明显。

6. 结论

无论是英语语音教学还是英语语音评测，语声段的正确识别都是非常重要的环节。本文引用赵力教授的英语寂静/语声段判别方法，通过实验证明具有算法简单，误检率低等特点，是一种在网考实施过程中有效的、有一定实际推广应用价值的方法。同时在计算机语音识别技术实现中诠释了符号学的三个重要特征：任意性(arbitrariness)、线性(linearity)和分节性(articulation)，为我国大规模网考的技术实现提供了理论指导和具体解决方案。

参考文献

- [1] 周心慧. 中国古代戏曲版画考[M]. 北京: 学苑出版社, 1998.
- [2] 肖峰. 从哲学看符号[M]. 北京: 中国人民大学出版社, 1989.
- [3] 王铭玉. 从符号学看语言符号学[J]. 解放军外国语学院学报, 2004. 1 (27).
- [4] 一清/文. 语音识别: 以应用为王[N]. 互联网周刊, 1999-6-21(26).
- [5] 俞小娟. 面向中文信息处理的汉语复句中书读短语的自动识别研究[D]. 华中师范大学硕士学位论文, 2008. 6. 《中国学位论文全文数据库》
- [6] 赵毅衡. 符号学文化研究: 现状与未来趋势[J]. 西南民族大学学报(人文社科版), 2009. 12.
- [7] 谭芷芳. 外语教学的符号学视野[J]. 西南民族学院学报(哲学社会科学版), 2001. 7(22).
- [8] 刘青. 科技术语的符号学诠释[J]. 科技术语学研究, 2002. 4(2).
- [9] 钱揖丽, 荀恩东. 基于标点信息和统计语言模型的语音停顿预测[J]. 模式识别与人工智能, 2008. 8(21卷第四期).
- [10] 叶竹钧. 朗读中的停顿探析[J]. 语文教学通讯, 1995 (7).
- [11] 郭鸿. 索绪尔语言符号学与皮尔斯符号学两大理论系统的要点——兼论对语言符号任意性的质疑和对索绪尔的挑战[J]. 外语研究, 2004(4).
- [12] 李幸. 汉语句法分析方法研究[D]. 中国科学院研究生院硕士学位论文, 2005. 5. 《中国学位论文全文数据库》.
- [13] 赵力等. 基于子空间分析的语音信号寂声语声段识别方法[A]. 信号处理(第十届全国信号处理学术年会(CCSP-2001)论文集), .2001. 11. 01.

Technical Realization of Punctuation from the Semiotic Perspective

Liu Jian-gang, Dong Jing

(School of Foreign Languages, Southeast University, Nanjing 210096, China)

Abstract: Study on semiotics has a hot topic around the world, but it hesitates to be applied into computer-field, which would be a wonder if it has access to the speech recognition. Language is usually described as speech signal, which involves punctuation in sound with silence and speech. All these silence and speech sounds have characteristics in significant and signifié. The study on punctuation in China has been focused on word-based text, lacking in precise explanation of sound punctuation, which leading to our study on recognition in sound-based text. In this paper, we discuss the definition and characteristics of punctuation in script and sound with parameter analysis of wavelet transform, and therefore defines the critical nature of punctuation: stop in silence, trying to realize the recognition of punctuation in sound-text.

Keywords: semiotics; pronunciation; sound-text

作者简介: 第一作者刘健刚(1959--), 男, 东南大学外国语学院副教授; 主要研究方向为实验语言学、语料库语言学、CALL和iBLL。电邮地址: jhonliunj@163.com
第二作者董静(1975--), 女, 东南大学外国语学院讲师; 主要研究方向: 应用语言学、CALL和iBLL等。