

词切分的韵律学线索^{*1)}

杨 玉 芳

中国科学院心理研究所,北京,100012

摘 要

本文是关于汉语普通话词切分的一个实验报告。用无意义重复句和合成句作语音材料,探讨音节时长分布、基频曲线、音节强度以及共振峰等韵律学线索在词切分中的作用。得到的主要结果是,在排除语义和音段层面信息的条件下,听者能够根据句子的韵律学特征进行句子结构的判断和词的切分;在对切分有效的韵律学线索中,最重要的是音节的时长分布模式,其次是基频曲线,音节的强度和共振峰变化对切分没有贡献。

关键词: 词的切分,韵律学线索,重复句

引 言

在阅读和会话过程中,为理解文章或话语所传达的信息内容,人不仅要识别其中的语言符号和音段,还要把它们组合成不同层次的语言学单位,如词、词组、短语、句子等,并确定各单位的作用和其间的关系。这个过程就是切分。在涉及自然语言的人工智能系统中,切分也是必不可少的环节。因此切分越来越多地受到语言学、心理语言学和计算语言学领域中研究者的注意。但到目前为止,关于切分所作的研究大部分都集中在书面语言,有关言语的相对较少。显然,言语的切分不同于书面语。我们需要加强对言语切分的独立研究。

词是能自由运用的最小意义单位。在不同的语言层次和单位中,词是最重要、使用最频繁的。可以说,词的切分是切分中最基本的问题。在语流中,特别是在句子或短语内部,词与词之间常常紧紧相连,甚至相互重叠。而听者可以毫不费力地听出一个个离散的词,其中一个重要原因是语流中存在有关词边界的声学语音学线索。有的语音学家认为,在每种语言中都有许多语音过程存在,其理由不在于语音本身或发音上的需要,而在于切分的需要。并推测,没有语音过程的语流如同没有标点和空格的文章,将使人类的通讯速度降低几个数量级^[1]。Church(1988)^[2]指出,言语信号富含各个表达层次上的语境冗余度,包括音位变体、音位配列规则、音节结构、重音区域、句法和语义等,可以用来帮助切分。例如,音位变体向来被人们看作是词识别中的统计噪声和引起混乱的根源。而 Church 认为,这些随音节环境产生的系统变化可以用来切分音节和重音区域,限制词的匹配过程。并建

* 国家自然科学基金资助项目。

1) 本文于 1992 年 2 月 15 日收到。

议,词的提取应分成两个问题,切分和匹配。切分主要利用那些随语境变化的特征,而匹配则使用不变的线索。

语流中存在有关词边界的声学语音学线索。对于某种特定的语言来说,这些线索是什么?听者如何利用它们达到切分的目的?已有一些研究者探讨了这些问题。Nakatani和Schaffer(1977)^[3]发现,在排除音段和语义信息的条件下,听者能够利用英语词的重音和节奏这类韵律学信息把话语切分成词。Cutler和Norris(1988)^[4]也证明,在英语连续语言识别中,听者常常把重读音节当作词特别是实词的开始,并在强音节处开始词义提取。许毅(1985)^[5]研究过汉语普通话的音联,认为在汉语普通话里不存在有语音边界的词。同时又认为,虽然没有词的语音边界,音联也不是与语法语义毫无关系,只是它们之间的关系并非一对一的简单关系。

如果汉语普通话的词界不像英语和其他语言那样容易确定,听者如何把语流切分成词,就是值得研究的问题了。在与普通话词切分有关的因素中,首先值得注意的是韵律学特征。近年来对韵律学特征的语音学研究发现,汉语普通话音节的时长除受到语音学和音系学这些内在因素影响外,还受到语言学因素的外在影响,包括语音结构特点、语法环境和底层语义要求等。普通话声调在语流中的变化也与语法结构有关。听者是否利用这些节奏和声调变化去判断语句的结构,进行正确的切分呢?

本文是关于汉语普通话词切分的一个实验研究报告。所探讨的问题是声调、节奏及更低一级的声学语音学性质,如音节的时长、强弱、基频曲线、共振峰变化等在词切分中的作用。

实验和结果

一、实验句的设计

作为初步研究,本实验采用由六个音节组成的简短陈述句。在这些句子中,词与节奏单元一致,即每个节奏单元只含一个词。句子的结构有三种:(1) $\underline{x\ x} \ \underline{x\ x} \ \underline{x\ x}$; (2) $\underline{x\ x} \ \underline{x\ x} \ \underline{x\ x}$; (3) $\underline{x\ x\ x} \ \underline{x\ x}$ 。为观察声调在切分中的作用,每种结构配有三种声调组合模式:(1)444444; (2)121423; (3)341421。表1为实验中所用的九句简短陈述句。

表1 实验句

句子结构	声 调 组 合 模 式		
	4 4 4 4 4 4	1 2 1 4 2 3	3 4 1 4 2 1
$\underline{x\ x} \ \underline{x\ x} \ \underline{x\ x}$	部队命令撤退	今年蔬菜稳产	储量超过台湾
$\underline{x\ x\ x} \ \underline{x\ x}$	到电去上课	说吉它的原理	你放心地回家
$\underline{x\ x\ x} \ \underline{x\ x}$	匡教授作报告	珊瑚礁在南海	打字机要维修

二、重复句的产生、分析和知觉切分

为排除语义和音段层面的影响,集中研究韵律学特征在词切分中的作用,使用了上述句子的重复句和合成句作实验材料。重复句的产生方法是,由一名男性发音人依次朗读实验句。在读完每句之后,用/da/代替句中的所有音节,按照原句的节奏、声调和语调重读。用/da/的原因是考虑测量音节时长的方便。

发音在隔声室中进行, 同时进行数字采样和存贮。采样频率为 10KHz, 采样前先对信号作功率放大和 4KHz 的低通滤波。

为考察重复句是否保留了实验句的韵律学特征, 用 ILS 软件包分别对实验句和重复句的数字样本文件作线性预测分析。然后就基频曲线、能量曲线和音节时长分布模式作成对对比。作为例子, 图 1 中给出句子“打字机要维修”和它的重复句的分析结果。通过分

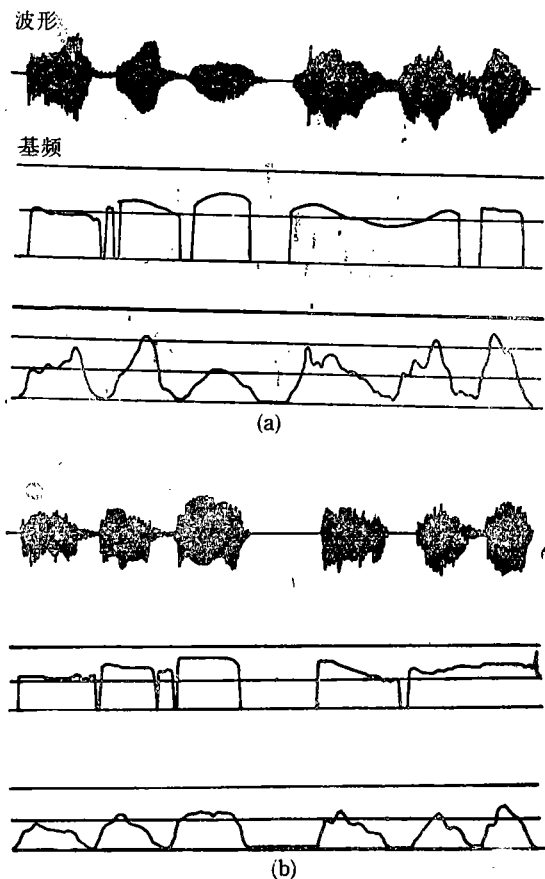


图 1 实验句“打字机要维修”

(a) 和它的重复句 (b) 的波形、基频曲线和能量曲线

析和对比可以看出, 在具有同样声调组合模式的三个句子中, 音节的时长分布和声调连接方式都因结构不同而不同。重复句保留了原句的主要韵律学特征。

在分析的基础上, 又用重复句作了知觉辨别实验。实验的目的有两个。一是进一步从知觉上说明重复句是否保留了原实验句的韵律学特征, 听者能否从重复句判断原句的结构和作词的切分; 二是为下一步的合成句实验打基础。

实验中以随机顺序回放九句重复句, 每句重复九次。要求听者判断每句所代表的句子的结构。实验由框架程序控制, 包括提供刺激和接收听者的反应。刺激是通过耳机传给听者的。听者如果认为听到的重复句属于第一种结构, 则按键盘上的数字键“1”; 如果属于第 2 种结构, 则按键“2”; 如果属于第三种, 则按键“3”。刺激出现一秒钟后, 如果听者

不作反应,刺激将重复出现。直到听者作出合法的反应后,另一个刺激才出现。实验在隔声室中进行,背景噪声不大于 40dB。参加实验的听者共 12 名,年龄在 20 至 40 岁之间,说纯正的普通话,无听力问题和耳疾。

在正式实验前,对每个听者进行了训练。训练中先成对地回放实验句和重复句,并指出句子的三种结构。然后随机地回放重复句,让听者练习判断句子结构。图 2 是 12 名听者对九个重复句判断的平均正确率。

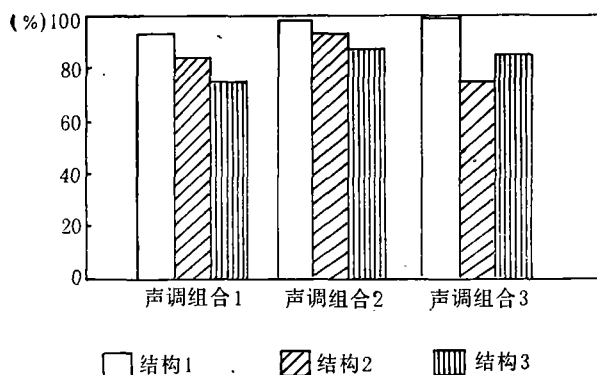


图 2 重复句结构判断正确率

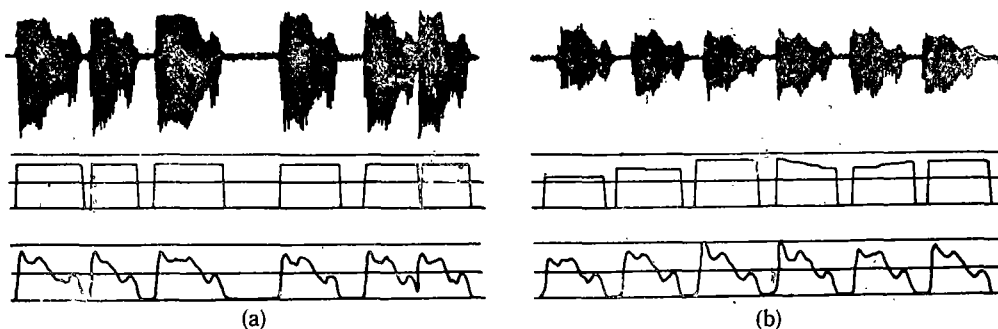


图 8 “打字机要维修”的合成句的波形,基频曲线和能量曲线

(a)保留音节时长分布模式的合成句(条件一);(b)保留基频曲线的合成句(条件三)

用重复句所作的辨别实验结果说明,在没有语义和音段信息的条件下,听者能够根据其中的超音段信息相当准确地判断结构,平均正确率达 81% 以上。二维因素方差分析结果说明,三种结构和三种声调组合的识别率之间没有显著差异($F = 1.712, P = 0.104$)。

三、合成句和词切分知觉线索

听者判断重复句的结构,利用了哪些韵律学特征呢?为回答这一问题,又分别用重复句的音节时长分布模式、基频曲线、能量曲线和共振峰参数合成一系列/da/组成的句子。每句仅保留原重复句的一种或两种韵律学线索。共有六种实验条件。在条件一中保留音节时长分布模式,使所有音节的基频为常数(170Hz),音节能量相等,辅音元音相接处共

振峰变化也相同。在条件二中保留了重复句的音节强度关系,而使各音节的时长、基频和共振峰变化都相同。在条件三中保留了重复句的基频变化特征。在把重复句的基频参数移到合成句前,对各音节基频曲线作了线性压缩或拉伸,使其具有相同的时长。这种时阈上的伸缩变化不会影响基频曲线的起始频率值、走向和终止频率值,但会影响曲线的变化率。条件四中保留了重复句中各/da/音节辅音元音过渡处共振峰的变化,而使合成句中所有音节时长、强度和基频值都相同。图 3(a)和(b)分别是保留了图 1(b)中重复句的音节时长分布模式(条件一)和基频曲线(条件三)的合成句的波形、基频曲线和能量曲线。本实验的工作假设是,通过听者在听辨实验中根据合成句作结构判断的正确率可以确定各种韵律学线索在词切分中的贡献。

实验中还使用了另外两种实验条件。条件五保留了重复句的时长和基频特征;条件六保留了时长和强度分布特征。条件五包含了主要的声调信息,条件六则包含了主要的节奏信息。通过这两个实验条件要考察的是声调和节奏在切分中的作用,及时长与基频、时长与强度之间的相互作用。

上述六种条件下的合成句都是通过与 IBM PC/AT 微机联系的 LSI 并联共振峰语音合成器进行合成的。

实验中将六种条件下的合成句分别放音,形成六个实验模块。在各个模块中,每个合成句重复九次,以随机顺序排列,刺激间隔为 1 秒。要求听者判断合成句的句子结构,并通过键盘作出反应。12 名听者全都参加过重复句实验。六个实验模块以随机顺序排列,完成每个模块约需二十分钟。完成一个模块后,听者至少要休息两分钟。整个实验约需两个半小时。表 2 是 12 名听者在六种实验条件下对合成句结构判断的平均正确率。为便于比较,表中还列出了对重复句判断的正确率。

表 2 合成句结构判断正确率(%)

声调	结构	条 件						
		1. 时长	2. 强度	3. 基频	4. 共振峰	5. 时长+基频	6. 时长+强度	7. 重复句
1	1	100.0	33.3	87.0	20.4	99.1	98.1	93.5
	2	84.3	29.4	58.3	40.7	85.2	75.0	84.3
	3	64.8	51.9	68.5	46.3	86.1	72.2	75.9
2	1	97.1	38.9	57.4	41.7	96.3	96.3	99.1
	2	96.3	45.4	29.6	29.6	89.8	96.3	95.4
	3	88.0	25.0	54.6	32.4	72.2	73.1	86.1
3	1	91.7	52.8	49.1	36.1	85.2	80.6	99.1
	2	82.4	31.5	36.1	35.2	89.8	69.4	75.0
	3	86.1	27.8	47.2	30.6	75.9	79.6	85.2
平 均		89.9	36.2	54.4	34.8	86.6	82.3	88.2

从表 2 可以看出,在六种实验条件中,条件一、五和六的正确率与重复句相比十分接近,都高达 80% 以上。条件三的正确率为 54%,明显低于上述三种条件,但大大高于概率水平。条件二和四正确率最低,分别为 36.2% 和 34.8%,略高于概率水平。

对实验数据作三维因素方差分析[条件(7)×结构(3)×声调组合模式(3)]。分析结果

是,实验条件作用显著($F = 114.220, P = 0.000$),声调组合模式作用不显著($F = 2.543, P = 0.079$),结构作用显著($F = 9.438, P = 0.000$),这些因素的交互作用均不显著($P > 0.172$)。

Pos hoc 检验表明,条件一、五、六和重复句之间差异不显著($P > 0.177$);条件三与其它五种条件的正确率差异均显著($P < 0.001$);条件二和四之间差异不显著($P = 0.746$),但它们和其它几种条件之间的差异均显著($P < 0.001$)。结构一和结构二、三之间差异显著($P < 0.01$),结构二、三之间差异不显著($F = 0.002, P = 0.919$)。

讨 论

本实验得到的主要结果是:(1)在排除语义和音段层面信息的条件下,听者能够根据句子的韵律学特征进行句子结构判断和词切分。(2)在对切分有效的韵律学线索中,最主要的是音节时长分布模式,其次是基频曲线。而音节的共振峰和强弱则对切分没有贡献。(3)时长与基频、时长与强度之间没有相互作用。下面就此作几点讨论。

在以往的研究与切分有关的韵律学特征的文献中,多半是研究重音的作用。本实验借鉴 Nakatani 和 Schaffer 的研究方法,用无意义重复句和合成句作材料研究切分的韵律学线索。这里先将二者的结果作一对比。英语是一种重音语言,重音在词中起着辨义作用。而且,有百分之七十左右的英语单词的重音在第一个音节上。重音对于从语流中进行词的切分起着十分重要的作用。在重音不足以提供词切分线索时,节奏有助于切分。在英语中,音高对切分没有贡献。汉语是一种声调语言,汉语单词没有固定重音位置,重音不起辨义作用。但在语流中与轻重有关的一个重要因素——音节的时长却受到语音单元结构和句法环境的制约。有关的研究结果说明,在语流中,音节多半作为更大语音单元的一部分活动,其时长往往取决于那些单元的结构特点。在这些单元中,音节时长有比较稳定的分布模式。其中最稳定又最具构词能力的是双音节词。在语流中,双音节词的前音节时长通常大于后音节时长。三音节和四音节词内部时长分布与其组合结构有关。反过来,我们是否可以,音节的时长分布模式又反映了语音单元结构和句法关系,听者可以利用这一关系,根据音节时长进行结构判断和词的切分。另一方面,汉语普通话有四声,每个声调有自己的区别性特征和固定的调型。在语流中,由于语音单元结构等的要求和协同发音作用,声调会改变音域、起始或终止频率值。这种变化是有规律可循的。语音学研究结果说明,普通话的语调由单字词和二字词作为基本单元。在语句中这些单元相互连接时,调值是升高还是降低,调域是扩展还是压缩,取决于语音结构和句法环境。基于这个原因,听者能在排除时长等因素的情况下,根据基频曲线进行结构判断和词的切分。如同 Church 所说的那样,利用随语境变化的那些声调特征作切分,利用不变的特征作识别和匹配。

本实验和 Nakatani 的实验结果相一致的地方是,都没有看到频谱变化和强度在切分中的作用。但从其它的有关研究看,特别是考虑到 Remez^[6]等从知觉组织角度所作的一些研究,应该说,这两个因素在切分中的作用值得进一步研究。

本实验仅仅是对词切分的韵律学线索问题作的初步研究,使用的语音材料十分有限,在设计的语句中,词与节奏单元相一致,而且仅有一名发音人。在更自然更复杂的语音材料中,这些结论是否还正确,需要作进一步的探讨。

参 考 文 献

- [1] Kaye, J. (1989), *Phonology: A Cognitive View*, Lawrence Erlbaum Associate, Publishers, p. 49.
- [2] Church, K. W., (1987) Phonological parsing and lexical retrieval, *Cognition*, 25, p.53—59.
- [3] Nakatani, L. H. and Schaffer, J. A. (1978) Hearing "words" without words, Prosodic cues for word perception, *Journal of the Acoustical Society of America*, 63(1), p. 234.
- [4] Cutler, A. and Norris, D.(1988) The role of strong syllables in segmentation for lexical access, *Journal of Experimental Psychology: Human Perception & Performance*, 14(1), p. 113—121.
- [5] 许毅(1985). 汉语普通语音联, 中国语文.
- [6] Reméz, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D., (1981) Speech perception without traditional speech cues, *Sciences*, 212, p.947—950.

PROSODIC CUES FOR LEXICAL PARSING

Yang Yufang

Institute of Psychology, Academia Sinica, Beijing

Abstract

This is an experiment report on lexical parsing of Chinese utterances, Using as stimuli reiterant and synthesized sentences of six-syllable Chinese short utterances, the roles in lexical parsing were investigated of such prosodic cues as distribution patterns of syllable durations, fundamental frequency and intensity curves of utterances, and forment transitions of syllables. It was found that listeners could identify with high accuracy sentence structures and parse words in short utterances without any segment and semantic information, and that among the prosodic cues possibly usefull for lexical parsing the most important were distribution patterns of syllable durations and fundamental frequency curves of utterances.

Key words: lexical parsing, prosodic cues, reiterant sentences