

言语知觉研究*

杨玉芳

(中国科学院心理研究所 北京 100012)

1996年1月19日收到

摘要 言语知觉是心理学的一个领域,其发展与语音学、语音工程和人工智能等许多学科有关.本文简要介绍了在认知心理学和其它相关学科的推动下,言语知觉发展的主线、当前的主要问题和研究现状.

关键词 知觉常性,词知觉加工,语句知觉加工

Research on speech perception

Yang Yufang

(Institute of Psychology, Chinese Academy of Sciences, Beijing 100012)

Abstract Speech perception, as a research field of Psychology, relates its development to many research fields, such as Phonetics, Speech Engineering and Artificial Intelligence. This article gives a brief description of the development trend, questions in highlight and the present status of research in the field of speech perception.

Key words Perceptual constancy, Perceptual processing of word, Perceptual processing of sentence

言语知觉,作为心理学的一个领域,已有四十余年发展历史.在发展的早期,言语知觉研究的主要对象是音位,目的是寻找最基本的音位知觉线索,希望这些线索一旦确定下来,就可以解决言语识别的根本问题.随着心理学,特别是认知心理学的发展,同时也在其它相关学科,特别是言语通讯和语音工程发展的推动下,言语知觉研究的视野从音位扩展到词、语句,甚至话语.在每一个更高的语言学层次上,都有许多低层次所没有的语言现象和知觉问题值得研究.从心理学基本过程的角度看,言语知觉从只研究刺激与反应之间的关系,即寻找某一语言单位或语言现象的知觉线

应用声学

索,到研究知觉和认知加工过程,探讨人作为一个信息加工系统,为达到识别和理解的目的,对言语信号进行的知觉加工,及在各加工阶段上信号的编码和表征形式.不仅研究人在外界信号驱动下“自下而上”的加工过程,还研究“自上而下”的加工过程,即探讨存贮在长时记忆中的知识和经验对感知觉加工的影响和制约.从通讯的环境和条件看,言语知觉从研究理想通讯条件下的言语知觉加工,发展到研究有信息缺失和噪声干扰情况下人的语言加工能力.言语知觉研究的发展趋向,可以说是研究

* 国家自然科学基金资助项目

自然通讯条件下人作为一个复杂的信息加工系统处理言语信号的能力和过程。

现在,言语知觉已经成为认知心理学的一个重要组成部分,同语言学、人工智能、语音工程等相关学科的关系也越来越密切。

下面仅从几个方面简述一下言语知觉研究的一些新进展。

1 言语信号的知觉常性

言语知觉研究的基本问题,如同视觉等其它知觉研究领域一样,是信号的心理表征及表征的知觉常性问题。

众所周知,言语信号中缺少声学语音学不变量。要确定独立于语境的与知觉到的语言单位唯一匹配的音段和特征是极其困难的。由于协同发音,与一个音位相关的声学信号在不同语境中有很大的差异。语流中一个声段常常含有几个邻近音位的信息。这给语音识别带来极大的困难。另一方面,面对这样一种信号,人的言语知觉过程却是自主地、近于实时地进行。人是怎样处理变异识别语音的?对这样一个问题,研究者们正在从不同角度和层次上进行探索,寻求答案。根据关于语音内部表征的假设,大致可以分成两种研究路线或方法。

其中一种,有人称之为抽象主义路线或分析方法。这类研究工作接受一种共同的假设,认为在知觉加工的某些阶段上,言语的内部表征是离散的音段、特征系列或描象模板,尽管在表征的具体描述上研究者之间持有不同的看法。

为了解言语在外周的初级感觉编码,人们一方面用动物作一系列重要的生理研究,以考察听神经对简单言语信号的反应活动,确定与言语声中重要的声学性质或特征直接对应的听神经纤维放电模式的性质;另一方面研究以心理物理学为基础的言语加工模型。在建立言语内部表征时,将临界频带、掩蔽向上扩散等心理物理学事实考虑在内。

Sawusch^[1]提出了以不变的听觉特征为基础的言语知觉模型。在这个模型中,有一系列

变换和表征。言语信号首先在外周听觉系统转换成内部时频表征,提取出一系列基本特征,如幅度包络、波形的周期性或非周期性性质、基频和频谱峰值的幅度和频率等。再从这些局部特征中提取一系列不受讲话者、语速和语境影响的听觉特征,如谱峰频率差和短时谱的统计动差等。最后是把这些听觉特征与记忆存储中的目标进行匹配。

减少语音变异,解决言语信号知觉常性问题的另一方面工作是关于听觉规范化问题的研究。由于讲话者的声道长度存在差别,发音时的姿势和在不同语境下使用的发音方法也不相同,口语中使用的多数语音学区别性特征的声学语音学表现,在讲话者之间存在重大差别。通过知觉补偿和规范化过程,听者能够轻易地忽略和减小这种差异。了解这种知觉补偿的确切机理并把它应用于语音识别系统,是正在研究的问题。

人们认为,言语知觉加工不仅仅取决于输入的言语信号,人的语言学知识和语境对加工过程也起着重要的限定和制约作用,尽管人们目前对其确切的作用机制还看法不一。张家骏等^[2]在研究汉语语音知觉特征时发现,语音诸区别特征在知觉上的重要性对不同的语言而言是不同的,而且与音节结构有关。在汉语中,发音方式不同、但具有相同音节结构的声母j、q、x在知觉上极易混淆。这一事实说明有关音节结构的语言学知识对言语加工的影响。

关于言语信号不变量和知觉常性问题的另一条研究路线是与前面所述的抽象或分析方法相对的一种“情景”或非分析方法。Pisoni^[3]和他的同事所进行的关于发音人和语速变化对言语知觉的作用和知觉学习的研究说明,有关发音人的嗓音和语境信息存贮在长时记忆中并影响着言语知觉过程。这一发现对传统的抽象方法的假设提出了问题,说明抽象主义方法将言语的语言学性质和指示性性质加以分离是不正确的。抽象主义方法强调输入信号的符号编码,与之相对的情景记忆对言语的指示性信息(发音人性别、方言、语速)进行编码,并保存

着听者在识别过程中对发言人嗓音作的知觉操作或程序性知识, 这些内容以内隐记忆的形式无意识地影响着听者的言语知觉活动. 因此, 按照非分析认知理论, 抽象主义方法中当作噪声和干扰的变异实际上为言语知觉提供了有规律可循的、有用的信息.

2 词的知觉加工

口语词识别和词意提取研究在近二十年中得到人们的重视, 成为独立于视觉词识别和孤立音位识别的一个部分.

词是最小的自由意义单位, 也是最重要、使用最频繁的一个语言学层次. 当把目光移到词时, 会遇到许多在处理音位时没有的现象. 关于词的知觉加工有许多重要的实验事实和现象, 使人们有可能去思考和构建口语词加工的心理模型. 这些事实和现象包括词加工中的音位恢复、高频词优先、词汇密度效应, 以及对词实时加工研究中揭示的现象等.

研究者们假设, 词加工如同音位加工一样, 首先要从接受语音输入开始. 词加工的基本阶段包括: 感觉输入与词的初始接合、激活(与输入匹配的词汇状态的改变)、选择(从激活的词集中选择目标词)、识别(选择阶段的终点)和词汇信息提取. 词加工还受到语境的制约和限制. 按照来源, 语境的影响可以分成结构性和非结构性影响两种. 结构性影响指由元素组成更高级单位的方式的限制, 这种影响涉及系统各加工层次间的相互制约. 如词对音位加工的影响, 句法对词加工的影响等属于这一类. 非结构性影响指由于词之间的语义和语用联系产生的词之间的相互影响, 这种影响发生在系统的同一加工水平上. 按照作用方式, 语境影响可以分成自主和相互作用两种.

已有一些口语词加工的心理模型对词加工各个阶段的内容和语境影响的方式加以描述和说明. 这些模型在所考虑的实验事实、所侧重的加工阶段以及各种知识源的相互作用方式等问题上都不相同.

W. Marslen-Wilson^[4]等提出的 Cohort 模应用声学

型, 其核心部分处理的是自音位识别完成到词识别的过程. 模型的实验基础是用 Shadowing 方法对语流中词实时加工过程所作的观察和研究结果. 模型认为, 词的识别过程由两个阶段组成, 第一个阶段, 输入词起始部分的声学语音学信息, 激活了记忆中所有以同样信息开始的词, 这是一个自主加工或数据驱动阶段. 这个词集一旦被激活, 所有可能的信息源都对选择有影响. 随着输入信息的增加, 与之匹配的词汇减少, 直到确定目标词. 这是一个相互作用的加工阶段. 这个模型仍在扩展和完善中, 以便能更好地说明词知觉加工的基本事实.

Elman 和 McClelland^[5]提出的 Trace 模型, 是将认知的平行分布加工理论应用于口语词识别而产生的. 它是由称为“节点”的加工单元构成的系统. “节点”可以代表特征、音位或词. 节点的激活水平指示输入与节点所代表的单元的一致程度. 节点之间相互联结, 当一个节点到达激活阈限时, 会增加不同层次上节点的激活程度, 同时降低相同层次上节点的激活程度. 这个模型说明一个高度相互作用的系统是如何概念化的. 该模型有两个值得注意的特征, 一个是把协同发音和语境变异当作有用的信息源加以利用. 另一个是这个模型没有明显的对于输入语音信号的切分步骤, 切分是识别过程本身的结果.

Forster^[6]的搜索模型是一个严格意义的自主加工系统, 由三个语言处理器组成: 词处理器, 句法处理器和信息处理器. 在加工的第一个阶段, 信息从外周知觉系统送到词加工器, 对外周提取文件中词汇进行定位. 对该文件的搜索顺序是按照词的使用频率排列. 词汇在外周文件中的位置确定后, 再确定它在主要词库中的位置. 这一位置信息传递给句法加工器, 再传给信息加工器, 建立目标信息的概念结构. 这几个加工器的信息都可以直接传送给通用加工器. 它综合语言加工器的输出信息和一般概念知识后作出决策.

人们曾经认为, 词切分是词识别的自然和必然结果, 即识别出语流中的一个词, 其终点

也是后一个词的开始。对英语词汇的统计分析结果证明,这不能解决切分的全部问题。更重要的是,不符合人类听者识别口语词的实际情况。研究证明,口语中的一些词,特别是虚词的识别,常常是后续的几个词识别之后完成的。

自然语言不同于人工语言的方面之一是,存在语音过程和语音变化,这些变化的主要作用是给出各种语言单位的边界信息。利用这些信息对语流进行切分是人的语言能力和认知能力的一部分,也是保证人类快速通讯的必要条件。对于一种特定的语言,知觉相关的词边界线索是什么?人在产生和识别言语时,是怎样表达和知觉词边界的?已有一些研究者对这些问题进行过探讨,并将研究成果成功地应用于口语词切分系统的开发和研制。

研究证明,语流中词边界的表达和知觉主要是通过语句的韵律学线索和配列规则。对英语、荷兰语和汉语所作的研究表明^[7,8],音节时长分布模式以及辅音在音节中所占的时长比例是最重要的词边界线索。对于汉语而言,词之间的连读变调模式是独立于音节时长分布模式之外的一个有用的线索^[9]。研究结果还说明,词边界知觉,同音段层面上语言单位的识别一样,是受听者的语言学知识制约的。英语中有70%的单词重音在第一个音节上,口语中常用词以重读位音节开始的更多。对以英语为母语的听者的词切分错误所作的分析表明,错误多数出现在以非重读音节开始的词中。

3 语句的知觉加工

句子是由词构成的。在解决了词的识别问题之后,还不能解决语句知觉和理解的全部问题。为了表达语义和语用内容,讲话者除了选择恰当的词并将其按照一定的顺序排列外,还要根据句法和语境要求生成句子的韵律结构、语调和重音分布模式,系统地改变音段的时长和基频等超音段特征。语句的韵律特征是表达语句类型、句法结构和语用信息的最重要的手段。听者要了解语句的全部内容,必须在识别

词的同时,对句子的韵律特征进行知觉加工。听者利用哪些线索感知语句的韵律特征,又如何将其运用于语句信息加工,是言语知觉领域近年来研究的比较多的方面。

语句加工的重要问题之一是,听者如何从韵律中发现语句的句法结构。韵律是把言语切割成尺度不同的成分的组织框架,在一定程度上反映着句法内容。建立话语等级结构的韵律学手段有韵律结构和重音系统。韵律结构主要与语句的时间组织有关。Shen^[10]研究了汉语的时间结构和句法结构的关系,证实听者在分析歧义句时使用边界前延时和停顿等韵律学线索,并认为在评价不同韵律特征对于口语句子内部结构的相对贡献时,时间维将是最重要的。杨玉芳^[11]研究了汉语语句句法成分边界的声学语音学表现,以及语句韵律结构和重音分布模式的知觉表征,并以此为基础初步探讨了韵律结构和句法结构之间的关系。重音系统通过标记成分的中心(the head)建立话语的等级结构。中心把成分中的其它元素加以组织和归拢,并与其它成分的中心相呼应,形成彼此间的句法和语义联系。语句重音分布取决于信息分布方式,它的形成与语调有关。沈炯的语调模型^[12]认为,汉语的语调是由音节声调的高音点和低音点构成的极限线表现的。这些极限线有明显的错位或突变。高音线的错位与语势重音有关,低音线的错位与节奏重音有关。

从言语知觉研究文献中可以清楚地看到,语句加工过程中的音段知觉不同于孤立音节和词的情况。言语信号中包含了大量冗余信息。听者对话语中的音段不会而且也不必给予同样的注意。听者会寻找作为知觉决策基础的可靠信息,“可靠岛”(Islands of Reliability)去提取各种知识源。这样的信息包括重读音节、词的开头和结尾部分等。实验还证明,在一些知觉分析水平上可能存在选择延迟。就是说,当输入信息不充分而需要等待另外的信息时,决策可能会延时。有人研究了作语音决策的扫描窗口,证明在句子中,其后续六个音节(约2秒)以内的语境会对音位知觉加工产生影响。

参 考 文 献

- [1] Sawusch I R. Invariant auditory attributes and model of speech perception. In Proceedings of X I ICPhS, 1992, vol. 1, 63—67.
- [2] Zhang Jialu, Lu Shinan, Qi Shiqian. *Journal of Chinese Linguistics*, 1982, **10**, 189—205.
- [3] Pisoni D B. *Speech Communication*, 1993, **13**, 109—125.
- [4] Marslen-Wilson W. *Cognition*, 1987, **25**, 71—102.
- [5] McClelland J L, Elman J L. *Cognitive Psychology*, 1986, **18**, 1—86.
- [6] Forster K I. Levels of processing and the structure of the language processor, in Cooper W. E., and Walker B. C. T (eds), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, NJ: Erbaum, 27—85.
- [7] Quene H. *Journal of Acoustical Society of America*, 1993, **71**(4), 996—1007.
- [8] Cutler A, Norris D. *Journal of Experimental psychology: Human Perception and Performance*, 1988, **14**, 113—121.
- [9] 杨玉芳, 孙健. *心理学报*, 1994, **1**, 8—13.
- [10] Shen X S. *Journal of the International Phonetic Association*, 1993, **22**(1—2), 35—43.
- [11] 杨玉芳. *声学学报*, 1997, **3**.
- [12] 沈炯. *语文研究*, 1992, **4**, 16—24.

中国声学学会 1997 年青年学术会议在哈尔滨召开

中国声学学会 1997 年青年学术会议于元月 10 日 13 日在冰城哈尔滨召开, 来自全国各地的七十余名代表, 克服了经费紧张、旅途不便等许多困难, 到会交流讨论了 42 篇学术论文。

本次会议由中国声学学会主办, 黑龙江省自然科学基金委员会与省科协协办, 中国声学学会青年工作委员会、哈尔滨工程大学与黑龙江声学学会共同承办的。中国科学院声学研究所田静研究员任会议主席, 西北工业大学孙超副教授、哈尔滨工程大学杨德森副教授任会议副主席, 青年工作委员会主任委员、西北工业大学马远良教授主持了会议组织委员会的工作。中国声学学会副理事长、南京大学吴文虬教授和哈尔滨工程大学杨士莪教授等八位著名学者应邀到会指导, 并组成优秀论文评选委员会。会议由哈尔滨工程大学出版社出版了文集, 收入论文 86 篇, 内容包括物理声学、水声学、超声学、量子声学、声学物理效应、噪声及控制、语言声学、语音通讯、生物声学、声学测量、信号处理与分析、声学换能器等重要的声学学科方面的新进展, 文集主编为杨德森副教授。

顾亚平、李剑、印建华、张向东与何元安等五位青年声学学者应邀到会, 分别就自适应信号处理、超声检测、声电荷输运器件、汉语人机语言对话、水下结构振动与声及其控制等方面的研究进展作综述报

告。会议还评选出四篇优秀论文, 分别为哈尔滨工程大学水声工程系关浩同志的“高分辨率多波束图像声纳的仿真研究”、上海船舶电子设备研究所杜选民同志的“多波束条带测深系统中正交信号的获取技术”、中国科学院声学研究所杨军同志的“非线性颤振的有源声控制”和四川压电与声光技术研究所杨晓东同志的“SAW 11 位巴克码低旁瓣滤波器”。会议为他们颁发了优秀论文证书和评语, 并将推荐和资助他们参加 1997 年 11 月在香港召开的第六届西太平洋国际声学会议。

专家指出, 本次会议论文的整体水平很高, 多数选题先进, 反映出与会代表及他们背后的老一代声学家的辛勤工作, 得到了许多有价值的研究和应用成果。

会议还利用休息时间, 安排代表参观了国际著名的冰灯和雪雕展览, 令代表们领略了冰城的迷人风光和哈尔滨人的热情好客。与会代表普遍反映, 会议安排紧凑合理, 学术气氛浓厚, 收获很大, 不虚此行。

在闭幕式上, 会议主席田静向本次会议的积极倡导和支持者杨士莪教授及其领导的哈尔滨工程大学水声工程系赠送了花篮, 表示诚挚的谢意。

(中科院声学所 田 静)