

【4】 Meng Jinsheng, Guan Dinghua, Acoustical method for remote sensing of seafloor sediment types, (将在《海洋学报》发表).

[5] Meng Jinshen, Guan Dinghua, Remote sensing of attenuation of sound in marine sediments, 11th International Congress on Acoustics. Paris. 1983.

语 言 通 讯 展 望

张 家 骥

(中国科学院 声学研究所)

语言是人类最重要的交际工具,随着人类社会的进步,交际活动也就更加频繁、更加重要。因而,语言作为交际工具的功能和作用也在不断扩大,这表现在近代语言通讯手段的不断发展和提高上。所谓交际主要是交换信息。语言——口头的和书面的,便是最直接、最方便、最有效的信息源,因为语言不但是交际工具,也是思维工具。

近代语言通讯技术不但延长了人类口、耳的作用距离,而且扩大了它们的功能。语言本来是人与人之间的交际工具;电子计算机的出现和计算技术的发展,人类的大脑和双手得到了进一步的延长。于是一种新的交际方式——人机对话出现了。第五代计算机,便是以具有自然语言输入输出为其特征的。今天语言通讯不但在空间距离上已经超出了人类居住的地球而伸向了宇宙空间,在功能上还扩展到了人与机器的“交往”之中。可以认为,语言通讯正处在一个发展的新阶段。因此,我们现在考察一下它的发展历史和展望未来的前景便很有必要了。

一、历史发展

1. 语言信号的波形原则

语言通讯的历史发展,可以追溯到相当遥远的过去。我们不妨从1874年电话机的发明,作为近代语言通讯的起点。在贝尔(A·G·Bell)发明电话机时,曾经建立了一个原则,那就是后来所谓的波形原则,即让线路上电流的变化相当于说话时空气密度的变化。这样一个原则几乎统治了电话通讯的设计达一个世纪之久。但是,由于语言信号占据很宽的频带和很大的动态范围,要想不失真地传递语言信号,不但技术上要求很高,经济上也是不现实的。于是,发展了一系列语言信号波形处理的方法。首先是把语言信号的频带限制在300—3400Hz之内,这称为“电话语言”。此外,还有在调制和解调技术基础上发展的多路通信。为了适应当时对语言通讯的要求,希望对通讯频带进一步压缩,以

提高通讯效率。1939年* 达得雷(Dudley)^[1]在纽约的国际博览会上公开地展出了可以把3000 Hz带宽的语言信号压缩在300Hz带宽之内的声码器(Vocoder)。他提出了一种对语言信号描述的新概念。他根据载波电话的原理,把人讲话时声带振动产生的声源,类比为载波信号,而把由口腔运动对声源的影响类比为调制信号。调制信号包含着有意义的信息,而载波只是负载这个信息,使之让人能听到,它本身并不重要。如果能将语言信号解体,分成载波源和调制波,并把它们分别加以传送,那将大大压缩所需频带宽度。因为,在讲话时发音器官的运动是比较缓慢的(低于10Hz左右)。载波源的频率较高,但并不是把载波信号全送出去,而只是传送周期振动(浊音)或噪音(清音)。这种把语言信号分解,只取其基本特征来加以传送的概念,开始打破了波形原则,孕育着日后语言

* 在1937年Dudley便已经做出了声码器,并在哈佛大学展出。

信号的参数模型的出现。但是,由于语言信号是一个复杂的瞬态过程,在当时还没有适当的技术手段对它进行仔细的分析研究。达得雷的通道声码器,在技术实现上也相当困难,造价也很昂贵。因此,这种新概念并未取得突破性的进展。

2. 语言信号的参数模型

经过多年的努力,贝尔电话公司发明了一种新装置。它可以把语言信号以三维(时间、频率、强度)的形式显示在屏幕上或者烧灼在电传纸上。这便是所谓的语图仪,有了这种仪器,便可以对语言信号进行大量的仔细研究,从此产生了一门新的学科——声学语音学。1947年波特(Potter)等^[2]出版了“可见语言”一书,系统地分析了英语的语图。1952年皮特森(Peterson)和巴内(Barney)^[3]又对大量发音人(包括男人、女人和儿童)所发的英语十个元音进行了系统的分析,并把这些元音标记在以第一共振峰 F_1 为横轴、以第二共振峰 F_2 为纵轴的平面上。这样,各元音的相对位置,便和语音学家的元音舌位图上的位置一致。

与此同时,语音分析的结果,不但直接促进了电话通讯的发展,而且推动了语言机器的诞生。1952年戴维斯(Davis)^[4]等利用 F_1 - F_2 作图案匹配,研制了口呼数字识别机(Audrey),其识别率最高可以达到97%—99%。奥尔森(Olson)^[5]也在五十年代研制语言打字机。一时间各种不同类型的声码器和自动语言识别机方案纷纷出现。一种颇为乐观的情绪鼓舞着语言通讯的研究人员。我们国内也在这个时候,在中国科学院电子学研究所、语言学研究所和南京大学,开展了语言声学的研究工作。电子学研究所进行了声码器和自动语言识别的研究,在语言学研究所的配合下,邮电学院开始研制语图仪。

但是,在工作中很快地遭遇到了两方面的困难。一方面对语言信号的认识,当时还既不系统也不够深入;另一方面在当时的电子技术所能提供的手段也不足以应付语言信号的复杂性。因此,虽然语言机器的研究取得了很大进

展,但没有取得决定性的成功。

正是在这个时期,方特(Fant, 1960)^[6]发表了他的著名论文“语言产生的声学理论”,为建立语言产生的模型奠定了基础。他首先用多段不同截面的管子相联来近似表示声道特性,采用传输线类比的方法进行分析。他以大量的实验数据来检验理论分析,取得开创性的成果。这样就把达得雷的语言产生的载波调制概念,上升到了一个新的理论高度。

在Fant开拓的新的方向上,不少人又做了许多理论和实验研究。特别是弗兰那干(Flanagan)在贝尔实验室的系统的研究,把方特的工作又向前推进了一步。1972年再版了他的有名的著作“语言分析、合成和感知”^[7]。这本书把语言通讯过程中,从说到听的主要环节以及分析合成技术,作了系统的论述。

以方特和弗兰那干的著作为代表的理论成果,不但给语言通讯研究提供了新的理论武器,也给语言通讯工程指明了新的努力目标,尽管在当时这个理论模型,由于一些不可避免的近似,还存在着相当的局限性。可是,一直到七十年代后期,它所指出的理论目标,在技术实现上还存在着技术和经济上的困难。

3. 语言信号处理的新发展

电子计算机是进行大量数据分析处理的有力工具。1962年弗兰那干首先把它用于共振峰声码器的合成器的模拟。十年后,1971年在第七届国际声学会议(布达佩斯)上,他作了计算机应用于语言声学的报告。与此同时一些新的算法提出来了。1963年包格特(Bogert)等^[8]提出了倒频谱(Cepstrum)的算法,1965年库利(Cooley)和屠开(Tukey)^[9]提出了快速傅氏变换(FFT)算法。这对语言分析提供了新的数字化计算的手段。这也为计算机深入语言通讯研究架起了桥梁。随后,1968年奥本海默(Oppenheim)^[10]把倒频谱算法用于语言信号处理,建立了同态分析技术。

在七十年代初发展的线性预测技术^[11],是基于语言产生模型而建立的算法的最好例子。尽管为了计算上的便利,而把声道传递函数用

全极点模型来近似,从而忽略或以变通的方法来引进零点的影响,这种算法在应用上取得了令人基本满意的结果。这些零点的影响,主要来自发鼻音时和发辅音时声道的耦合作用。发鼻音时,鼻通道与口腔并联;发辅音时,激励声源位于口腔内部某点,声源前后都有一部分声道作负载。这两种情况都有零点存在。全极点模型不能很好地表示这些发音过程。

这里需要特别提一下的是,在七十年代初期,尽管语言产生模型在理论上给语言信号处理和语言机器的设计提供了有力的工具,但是,在具体实现时还是存在着许多困难。这主要表现在技术手段还不够,以及经费消耗常很大。自动语言识别研究工作,由于这种影响,在七十年代初期曾出现过一次低潮。有人甚至悲观地认为,实现语言识别是和登月一样的困难。但是也有一部分人,积极努力从事实用化的研究,以期转变人们对自动语言识别的认识,争取得到更大的支持。日本铁路系统部门便开始研制自动订票系统。

微电子学工艺和计算技术的发展,再一次给予语言信号处理和语言通讯工程以巨大的推动。首先是微处理器芯片,继而是专用信号处理芯片,不但使各种语言机器的构造大大减化,而且成本也大大降低了。供儿童学习语音拼读的简易自动语言识别装置出现了。它可以小到计算器大小,便宜到几十美元。一些简易的语言合成器也纷纷问世,并应用在许多方面。

二、当前状况

自从可以显示三维动态频谱的语图仪出现以后,在有效的技术手段的支持下,出现了语音产生的声学理论,建立了语言产生的模型。至今在理论方面,除了考虑了口腔侧壁阻抗的影响,声门阻抗的影响和声门波波形的影响之外,并没有根本性的进展。与此相反,在这一时期之内,在技术方面的发展,包括一些语言信号处理的新算法及其实现手段,却是突飞猛进的。现在语言声学理论所提供的模型和方法,用当前

技术手段都可以轻而易举地实现了。可是,语言通讯工程中的许多具体问题还没有很好地解决。比如,即使是高级的语言合成装置,它的音质也还不能完全令人满意;十分复杂的自动语言识别设备,也还不能对连续语言和不同的发音人,作出满意的响应。这些都反映出现有模型和理论的局限性。也表现出我们对语言信号的认识还相当肤浅。

按照共振峰的定义,它是声道的固有频率。严格地说,只有当激励声源停发同时声门关闭时,声道内的自由衰减振荡,才反映了它的特性。但是,我们总是在声源连续激励状态下,来测量共振峰的,当声带振动时,声门就有打开的时候,这样,喉阻抗便不是无限大,这就与理论上的闭管模型有所不同。实验已经证明,声带波的波形对浊音的音质影响相当大。我们的实验还表明,当发音人以不同的总声级讲话时,不但基频发生了变化,而且共振峰也会变化。所以,我们又必须考虑到语音产生的动态过程。在发音的动态过程当中,还要有协同发音(*Coarticulation*)和超音段特征(*Suprasegmental*)的影响。这些影响都是在现有的理论模型和技术实现当中所没能考虑的。因此,要想合成出高自然度的合成语言,还需要做很多努力。

当前,正在发展着的一种高级语言合成技术——文语转换系统(*Text-to-speech system*)^[12],便是用一系列生成音系学规则,来控制语言合成器,以产生连续语言。由于在规则当中,可以考虑语言的动态性质和超音段特性,所以它可以产生较为自然的语言输出。它的输入方式是灵活的,既可以把文字材料由键盘输入,也可以帮助聋哑人和有发音障碍的人说话,为盲人朗读书报,给计算机配上语言输出。同时,它也是进行语言分析合成研究的有力工具。

在语言自动识别当中,我们还会遇到,什么是最佳的语言自动识别的单位,是音素还是音节?语音有没有常量,即不变因素,如果有它在语音的那一部分?也就是说,不同的发音人,在讲同一句话时,他们的共性在那里?他们的个性又在那里?对这样一些问题,我们研究得还

很不够。因此,目前对专人的语言识别系统可以得到较高的识别率,而适用于多数人的,特别是对连续语言的识别系统,识别率都还很低。一些专家认为,对于未来的市场来说,语言自动识别可能潜力较大;但就近期发展来看,语言合成设备将会更快地扩大应用。

总的来说,近年来语言通讯在技术方面的发展速度,远远超过了理论方面的进展。许多在六十年代、七十年代望而生畏的技术难题和经济负担,今天都可以轻而易举地解决了。当前的情况是,在微电子学和计算技术武装了的语言通讯工程师们,为解决语言通讯中的各种难题,向语音学家、声学专家、心理学家们提出,我们还要做什么?怎样做?而这些理论家们却正面临理论枯竭的状态和高大的知识壁垒。

三、未来前景

在语言通讯领域中,技术与理论水平之间,出现了两次不平衡。在六十年代和七十年代,理论水平可以说超前于技术发展。一些理论概念和技术设计,在具体实现时,不是遇到技术困难便是经济代价太高。因而,一些语言机器,诸如声码器、语言自动识别装置、自动应答系统……,不是不能进入市场便是得不到推广。七十年代末,八十年代初,形势发生了逆转。超大规模集成电路,大容量的存贮,高功能的专用处理器,以及低廉的价格,使得语言机器获得了新的生命力。人工智能也已成为更具有吸引力的新方向,于是第五代计算机便应运而生,无疑地这将给语言通讯的发展以新的推动力。

在进一步要做的工作中,首先便是应当对语言信号的动态特性加以深入的分析,从而使语言产生的模型更加完善。当然,为了能更彻底地利用生成音系学规则,那么对现有的区别特征系统^[43]也应当进一步加以完善,特别是在语音的区别特征及其声学体现上的研究,还值得进一步深入。现在的文语转换系统,进行语言合成的基本单位还是音素(位)和词素,只

是用区别特征,作为书写语音规则,描写语音变化的一种手段。我们还不能用区别特征,直接来进行声学上的语言分析、合成和自动识别,这是与生成音系学的理论体系相矛盾的。

任何一种语言,语音和单词都是处在自然的语流当中的。因此,有些语言学的特征,是在一定长度的语音的序列当中体现的。这就是超音段特征或韵律特征(Prosody)^[44]。比如,声调是汉语的重要特征之一。它便是在音节上体现出来的,汉语普通话有四种声调,用来区别不同的意义;又比如,语调是在单句中体现出来的。升调表示疑问,降调表示陈述。着重点或语意中心,常常是在单词上体现出来的。需要强调的地方重读,则音调升高、音强加重也可能伴有音长加长。重音也是在单词上体现出来的,在许多语言中起着区别意义的作用。超音段特征的研究还非常不够,但它是非常重要的。在语法分析当中有歧义的句子,只要一唸(说)出来就明确了。所以,这也是应当进一步深入研究的一个领域。

除此之外,对于语言信号来说,它还负载很多非语言学信息。诸如发音人的性别、年龄、情感、态度、情绪、个人特征等。甚至某人的职业也不难从谈话中猜到。比如同样一句话:“你去吧!”在语音体现上可以表示出几种不同的情况:1.表示祈求,小孩子缠着大人去做某事;2.表示恫吓,大人不希望小孩子去某地,意思是,你去吧,后果将对你不利;3.表示态度,你去吧,反正我不去;4.表示猜想,你去吧,是不是。这种信息在文字上就必须加长才能传递。而口语则只需改变某些参数随时间的变化方式就可以了。说起来又经济又生动。这在语言通讯中,几乎是一个未开发的领域。

显然,语言交际是包括说和听两个过程的。

* 区别特征的概念,是在三十年代由布拉格学派的语言学家楚别茨柯依(Trubetzkoy)提出来的。在五十年代初,由雅柯布森(Jakobson)、方特和哈雷(Halle)加以系统化,并且以声学分析的具体事实来支持这种体系。其后,又有乔姆斯基(Chomsky)和哈雷提出了他们的体系。至今,几种体系并存不能统一。

在说的方面已经有了比较成熟的语言产生模型,而在听的方面则还没有建立起公认的模式来。实际上区别特征理论,是应当在说和听统一的基础上建立起来。因为人们的交际过程,是说和听的统一过程。在发展语言产生模型的同时,还必须十分注意语言感知模型的研究,并且把它们统一起来。这将是一项艰巨的任务,我们不妨把这样的模型称为语言的信息模型。

根据现有的语言产生的理论模型和语言信号处理技术,所建造的各种语言机器,在人机对话中所表现的语言能力,充其量也不过是一个小学生的水平。要想进一步提高语言机器的水平,就要对它进行训练。对机器进行训练,就首先要求我们对语言的产生和感知过程加以形式化,也就是要写出一系列规则来,这便是语言信息模型的内容。要想完成这些工作,必须靠声学家、语言学家、语音学家、心理学家等的通力合作才能胜任,这便是我们所面对的未来。

参 考 文 献

- [1] H. Dudley, *J. Acoust. Soc. Am.*, 6(1939), 169--177; *Bell Labs. Record* 17(1939) 122—126.
- [2] R. K. Potter, P. A. Kopp and H. C. Green, *Visi-*

ble Speech, New York, D. Van Nostrand Co., 1947.

- [3] G. E. Peterson and H. L. Barney, *J. Acoust. Soc. Am.*, 24(1952), 175—184.
- [4] K. H. Davis et al, *J. Acoust. Soc. Am.*, 24(1952), 637—642.
- [5] H. F. Olson, *Speech Mechine Considerations*, Proc. Stockholm Speech Comm. Seminar, R. I. T., 1962.
- [6] G. Fant, *Acoustic Theory of Speech Production*, Mouton and Co. 1960.
- [7] J. L. Flanagan, *Speech Analysis Synthesis and Perception*, 2nd Edition, Springer-Verlag, New York, 1972.
- [8] B. Bogert et al, The Quefrency Analysis of Time Series for Echoes, Proc. Symp. on *Time Series Analysis*, M. Rosenblatt Ed Ch. 15, Wiley New York, 1963.
- [9] J. W. Cooley, J. W. Tukey, *Math. of Comp.*, 19-90 (1965), 297—301.
- [10] A. V. Oppenheim and R. W. Schaffer, *IEEE Trans., on Audio and Electraoustics*, AU-16-2 (1968), 221—226
- [11] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.
- [12] R. Carlson B. Granström and S. Hunhicut, A Multi-language Text-to-speech Module QPSR 4/1981, 18—28.
- [13] G. Fant, The Nature of Distinctive Features, in "To Honor Roman Jakobson: Essays on the Occasion of His Sevevtieth Birthday" Mouton, The Hague, 1967.
- [14] I. Lehiste, *Suprasegmentals*, M. I. T. Press, Cam., Mass., 1970.

(上接35页)

在混响场测量高频声功率时[2],例如测量10—20kHz的声源功率,就往往需要相应频率范围的混响时间资料。而常用的扬声器发声方式,由于高频的声功率输出不够而难以达到要求。本声源则高频特性较好,用它来测量高频混响时间就十分方便。

高声强冲击波换能器灵敏度的校正,目前一般采用点爆炸源的方法[3]。由于本火花脉冲声源所发的声级高,重复率又稳定,它的波形与空中爆炸的冲击波相似(同为N形波,见图4)。因此利用它可作为高声强冲击波换能器近

场的校正声源,比之用点爆炸源的方法简便得多。

本工作得到王季卿副教授和周志孝工程师的关心与支持,方启文等同志对声学特性的测量给予了很大的帮助,作者深表感谢。

参 考 文 献

- [1] Howard. G. Latham, *Applied Acoustics*, 12-4(1979), 253.
- [2] George. C. Maling, Determination of Sound power-The last ten years, Proc. Inter-noise 82, P. 12.
- [3] 邵志常 裴佐霖, 同济大学学报(声学版), 1(1979), 123.