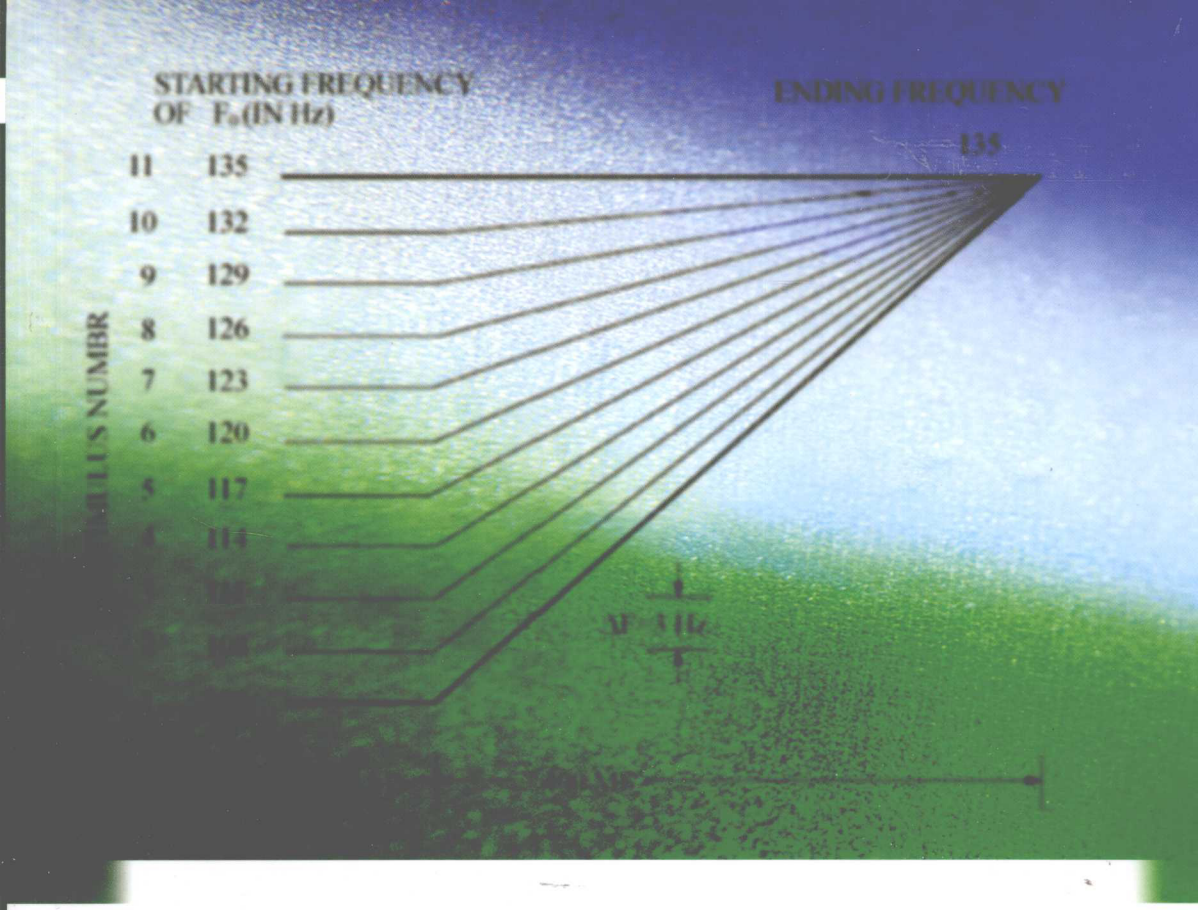

王士元语音学论文集

■王士元 著

世界图书出版公司

王士元语音学论文集

王士元 著

世界图书出版公司
北京·广州·上海·西安

图书在版编目（CIP）数据

王士元语音学论文集/王士元著. —北京：世界图书出版公司北京公司，2010.3
ISBN 978-7-5100-1772-8

I. ①王… II. ①王… III. ①语音学—文集—汉、英 IV. ①H01-53

中国版本图书馆 CIP 数据核字（2009）第 230807 号

王士元语音学论文集

著　　者：王士元
特约编辑：孔江平　汪　锋
责任编辑：梁沁宁　郭　力
封面设计：然则设计公司

出版发行：世界图书出版公司北京公司　http：//www. wpcbj. com. cn
地　　址：北京市朝内大街 137 号（邮编 100010，电话 010-64077922）
销　　售：各地新华书店及外文书店
印　　刷：北京高岭印刷有限公司

开　　本：787×1092　1/16
印　　张：42
字　　数：742 千
版　　次：2010 年 3 月第 1 版　2010 年 3 月第 1 次印刷
书　　号：ISBN 978-7-5100-1772-8/H·1078
定　　价：98.00 元

编 者 导 言

王士元先生是世界著名的语言学家和语音学家，对语言学理论有重要贡献。他祖籍安徽，1933 年生于上海，1960 年于美国密歇根大学获得博士学位，研究领域主要是声学语音学。博士毕业后，王先生曾先后任职于 M. I. T. 电子研究实验中心和 I. B. M. 约克镇高科技研究中心。1963 年至 1965 年，他在美国俄亥俄州立大学创办了语言学及东亚语言文学系，并担任系主任；1965 年被聘为美国加州大学伯克利分校语言学系教授，并在该系任教 30 余年后退休。王士元先生现为香港中文大学电子工程学系伟伦研究教授，兼任语言学及现代语言学系、翻译学系和东亚研究中心教授。

王士元先生致力于语言学和语音学的研究和教学近 50 年，共发表论著 200 多篇。他的“词汇扩散理论”打破了新语法学派音变理论一统天下的局面，是至今华人对语言学理论最为重要的贡献之一。从开始执教至今，王先生指导过近 40 名博士生，其中多位已成为语言学界的知名教授。王士元先生早年致力于语音学和言语声学的研究，取得了丰硕的成果。1973 年，他回国讲学，在语音学和言语声学学界取得了很大的反响，对中国文革以后语言学和语音学研究和教学的恢复做出了重要的贡献。为了使我国语言学学者和研究生能更方便地阅读王士元先生的论文，我们收集和整理了王士元先生有关语音学和语言学方面的论文编成文集出版。王士元先生的研究涉及语言学和语音学的各个领域，相互之间都有密切的关系，因此很难进行明确的分类。在征求了王先生本人的意见后，我们在本文集中共收入王士元先生有关语音学和语言学研究的论文 37 篇，大致分为 6 个部分，它们分别是：1）语音学研究 14 篇；2）语音信息量研究 2 篇；3）音系学研究 3 篇；4）词汇扩散理论 11 篇；5）语言演变模型研究 2 篇；6）语音合成和识别研究 5 篇。王先生的论文原创性和理论性都很强，限于我们的水平，很难全面介绍，下面就这 6 个部分简要介绍一下，希望对读者的阅读能有所帮助。

一、语音学研究

王士元先生早期的研究多集中在语音学方面，其语音学的论文可以分为描写研究、语音感知研究（包括塞音感知、音调声调感知、句子边界感知）和语音学理论研究，下面就这三个方面做一些简要的介绍。

在描写方面，论文（1978）介绍了一种自主研发的基于光反射感应原理的仪器，通过这个仪器，可以测量到硬腭和舌面的距离，从而推算出舌面的动态参数。利用该仪器不仅可以研究元音的动态特性，也可以研究辅音的发音动作。论文（1987）采用了测量关闭时长的方法来研究塞音的分类。根据韩语紧塞音的研究结论可知，紧的塞音往往需要较长的关闭时长，因此研究塞音的关闭时长可以对塞音进行分类研究。研究表明，清塞音 p、t、k 的关闭时长较短。论文（1995）利用声学分析的方法，研究了吴语塞音的性质，并根据数据，从统计分析和理论的角度讨论了吴语塞音的特性，认为“传统的清浊分类法对吴语不适合，用松紧来描写口腔闭塞时间的长短则更为合适。吴语中三类塞音可以称为松塞音，不送气紧塞音和送气紧塞音。”在塞音感知方面，论文（1959）讨论了词语末尾塞音感知声学线索的相对意义，同时提出了从声学特征研究词语末尾塞音感知的基本方法。论文（1961）根据辅音的声学和感知性质，提出和讨论了辅音的“外在”和“内在”性质的基本概念，并通过语音声学分析和感知实验验证了辅音感知“外在”性质和“内在”性质的区别和不同的声学表现。论文（1961）讨论了英语词首塞音对立的性质，通常认为词首的浊塞音常常不浊，而非浊音总是送气，这表明送气与否可能是一对塞音对立的主要区别特征。利用磁带拼接的方法，在研究了 s 后塞音的感知后，完全证明了这一推论。

在音调和声调的感知方面，论文（1967）就普通话中“阳平 + 上声”和“上声 + 上声”的双音节调位的异同做了科学的听辨实验研究。研究表明，人们在辨别这两个双音节调位时几乎分辨不出差别，其百分比为 49.2% 对 54.2%。论文（1972）讨论了人类语言中基频这个必不可少的语音特征在语音物理生理层面和语言认知层面的关系和差别。该文从理论上提出问题，结合大量的语言实际，深入地讨论了基频在语言中的意义。论文（1976）是一篇很经典的语音学文章，这是因为此文第一次采用实验语音学和感知的方法研究了两个调位之间的感知机理，发现了汉语阳平和阴平之间调位的感知范畴，因此在语音学、音位学、

心理语音学和语言历史音变等领域都有理论上的指导价值。论文（1978）讨论了两种音调在心理上的倾向。一种心理倾向是在元音不相同的情况下，音调在心理感知上会有一定的倾向。另一种是在振幅不同时，音调在感知上会有一定的倾向。研究结果表明，音调的感知会受到元音和振幅的影响，从而在心理感知上产生差别。论文（1984）用合成语音和自然语音拼接组合的方法研究了北京话声调的感知问题，研究结果发现听音人对双音词中第一个音节调类的判断会受第二音节音高和音长的影响。阴平在一个较高的声调前会被听成阳平，如果阴平时长较长，则会被听成上声。在句子边界的感知方面，论文（1976）通过听辨实验，研究了听话者对句子边界的感知能力，根据此文的研究成果和启示，可以进一步研究音系边界信号和句法边界关系的可能性。

在普通语音学理论方面，论文（1974）从理论角度提出了语言学为什么要研究语音的问题。因为从人类语言起源的角度看，人类生理的进化，特别是发音器官的进化与人类语言的起源有着密切的必然关系，绝不是偶然的。因此，语言研究的实质是研究语音信号和概念及表义结构的关系。偏离了这个原则就会使语言学的研究走入歧途。因此该文在语言学研究的方法论方面具有宏观的指导意义。论文（2008）介绍了语音学研究从用于诗词格律及语言教学，到19世纪印欧语言学的发展过程中逐渐成为历史语言学的基本部分，同时介绍了随着声学及生理学的发展，语音学在20世纪对语音物理和听觉机制的研究，基于此人们有可能开始进行语音进化的研究。最后，论文展望了21世纪语音学在脑科学技术的带动下，利用核磁共振等科技，会将语音大脑机制的研究纳入语音学的研究体系。

二、语音信息量研究

在语音信息量的研究方面，王士元先生有两篇重要的论文。论文（1960）针对美国英语辅音出现频率的统计差异进行了研究，研究结果表明美国英语辅音的频率受文献风格、方言差异和样本数量的影响不大，其差异主要是来自不同的统计词表（电子字典）和版本。论文（1967）是一篇关于“音位功能信息量”研究的经典文章，论文首先讨论了音位功能负担的概念，该研究可以追溯到早期的布拉格学派时期，当时主要注重音位学的二元对立。王先生上世纪50年代的研究主要介绍了霍凯特（Hockett）的研究和格林博格（Greenberg）的研究。霍凯

特认为，功能负担的重要性在于它对描写音韵系统有重要的价值，从而使我们可以有一个尺度来认识语言信息、语言冗余度和言语识别。格林博格认为，功能负担以通用的方式反映了一组音位或一组对立特征各成员之间的对有区别意义信号的贡献。王先生上世纪60年代的论文则主要介绍了赫厄希斯瓦尔德（Hoenigswald）关于功能负担和音变的研究，他认为，功能负担和语言的音变有关，并提出了一个假说，即在一种语言里，如果一种对立用得很少，它的消失对系统造成的危害要小于功能负担大的对立的消失。京·罗伯特（King R. D.）将音变和功能负担一同进行研究，并着重研究了音位功能和语音音变的关系，发现在日耳曼语中，功能负担和历史音变的关系不大。

在前人研究的基础上，王士元先生首次实现了功能负担的计算并指出了计量功能负担的困难，而且给出了解决这些困难的方法。首先，他讨论了音位系统中常见的三种分布、霍凯特与格林博格的测量方法，以及这些方法和香农（Shannon）的通信理论与各种语言学概念的关系。其次，在这些背景知识的基础上，王士元先生讨论了功能负担计量必须满足的五个条件。最后，他系统地发展了四种计量功能负担的方法。另外，王士元先生还指出："音变受到许多其他因素的严重影响，如音位之间语音的相似度和语言的接触等，但正如许多历史语言学家相信的那样，如果功能负担在音变中确实起作用的话，那么用量化的解释至少可以从一个方面阐明音变这一难题。"关于音位系统的分布，王士元教授指出："对任意一个语音序列，有三种相互关联的分布，即相似性分布、交叉性分布和互补性分布。当语音序列中每一个音位的排列都共有一组相同的环境时，它就处于相似性分布中；当音位的排列共有一些相同的环境，但不是所有的环境时，该序列处于交叉性分布中；当音位排列没有任何共有环境时，该序列处于互补性分布中。"王士元教授的研究为后来功能负担的研究建立了一个理论上的基本框架。实际上现代语音识别技术中常用的双音子和三音子的概念就起源于音位功能负担量的研究。

三、音系学研究

在音系学研究方面，我们选了王士元先生的三篇重要的论文。论文（1967）全面研究和讨论了声调的音系学特征。论文首先对声调的特征和音段的特征进行了梳理和定义，并在此基础上逐步介绍了音系学特征

的表述方式，提出了一套声调的基本特征，同时介绍了冗余度规则、语音解释、标记规则和闽语的声调循环。论文（1968）给出了元辅音对相似性的听辨结果，研究结果发现听辨者对相似的判断是非常一致的，同时也进行了音系特征对判断的研究。研究表明，有些特征对立并不影响相似度的判断，而另一些特征却对相似性的影响很大。论文（1968）对分离组别音系规则缩写所表示变量的使用进行了审查，提出是否有需要一个以上变量的规则组，回答是肯定的。证据包含了一个英语元音大转移的新公式，其中用到了一个不对称特征，它能区别四个元音舌位的高低。最后根据这一规则，讨论了规则化和音系演变的理论问题。

四、词汇扩散理论

王士元先生通过对大量语音演变材料的深入研究，特别是通过建立汉语方言数据库（DOC），提出了著名的“词汇扩散理论”。为了能较好地反映词汇扩散从概念的提出到理论体系的建立，我们将所选的论文大致分为概念的提出、实证研究和理论建立三个方面，希望能展示出王先生建立这一理论的整个过程。

在概念的提出方面，论文（1969）讨论了在音系演变过程中，在语音学层面可能是突变的，而在词汇层面可能是渐变的。当音变在词汇中扩散时，可能不是所有的语素都能达到适合演变的要求，这样，在音变竞争下，就会有一些残留。论文就音系演变理论中的一些基本问题进行了讨论，首次提出语言历史语音演变的“词汇扩散理论”。论文（1970）讨论了音系演变的一些细节。首先，文章简要介绍了导致词汇扩散的理论思考；其次，给出了支持词汇扩散的证据——汉语双峰方言；最后，给出了对音系演变轮廓的思考。

在实证研究方面，论文（1971）以潮州话声调的变化为例，研究和探讨了词汇扩散的过程。文章认为，尽管音系演变过程中伴随着词汇的突变，但并没有被系统地证实，而且有多种大量的论据和证据都表明词汇是渐变的。在词汇突变确实被证明之前，历史音韵学研究必须以词汇扩散是演变的基本机制为假设前提，并以潮州汉语声调演变的词汇扩散为实例证明了这一点。论文（1975）讨论了音变的两个基本问题，一是音变是怎样进行的，二是为什么音系变化过程被假设成一种形式并遵循一个特定的模式。为了回答音变是怎样进行的，文章提出了词汇扩散的概念，并用汉语、英语和瑞典语以及其他大量证据来证明。为了回

答第二个问题，论文认为音变的决定条件是说话人天生具有的音系和直觉的制约性。论文调查了大量汉语方言韵尾的损耗，发现印欧语系语言有同样的状况，并认为跨语言的普遍模式和演变进程可以用于追寻“语音演变的动因”，感知和心理学的证据都支持这一点。论文（1982）首先介绍了生物进化的基本理论以及生物学家和语言学家关于生物进化和语言进化关系的研究成果，并在此基础上，从变异和选择两个方面进一步讨论了语言演变的基本情况，指出“词汇扩散”反映了“变异和选择”的相互作用，选择的强制性主要是来自人体器官在认知、发音和感知方面的制约。论文（1987）首先介绍了在汉语声调的研究中，一般都认为汉语声调的发展过程主要是“分裂”，并以此为前提用语言学笔记的方式提出了另一种汉语声调演变的途径，认为汉语声调在中古汉语的发展过程可能是“合并”，并给出了证据。论文最后指出这些声调的合并很可能在词汇上是逐渐扩散的。论文（1970）介绍了电子词典是开发计算机在音系学研究中潜力的重要组成部分，它不仅可以用来重构汉语方言的音系历史，而且还可以从中了解通常的音系演变方式、演变之间的关系和演变所导致的共时系统。最后介绍了计算机快速和精准处理大量语料的能力，并给出了使用这种研究方法的实例。论文的特殊意义在于利用语音事实验证了词汇扩散的假设。论文（1987）介绍了利用汉语方言数据库对汉语声调从中古音到现代汉语各方言的演变规律，并给出了一系列具体对应规律表格。这一研究在方法上充分利用了计算机数据库容量大、计算快速和精准的优势，开创了历史语言学研究的先河。

在理论建立方面，论文（1969）是一篇理论性和综合性都很强的文章，文章详细介绍了不同历史时期语言变化研究的各个学派和基本概念，还介绍了新语法学派关于语言演变的基本概念和理论，即“词汇是突变的，而语音是渐变的”。在介绍这些情况的基础上，对以往关于语言演变的基本理论和观点提出了质疑。论文用语言演变的事实证明了以往关于语言演变理论中的谬误和不足之处，最后深入讨论了语言演变的基本规律，并提出“语音突变和词汇渐变”的基本理论，即“词汇扩散理论”。论文（1993）通过提供内部演变和接触演变相互作用的证据讨论了音变相互作用的情况。为了奠定讨论的基础，论文回顾了新语法学派的理论与词汇扩散理论冲突的基本问题。音变研究一直都被认为是新语法学派的专利，然而词汇扩散理论的提出冲击了其地位和理论基础。争论基于这样一种语言学理论的观点，即每一种语法成分都是相关

的，对语言学成分依赖的观点意味着音系和文法词汇的相互影响。为了质疑新语法学派的理论，论文提出了一个更广泛的词汇扩散的版本，认为音变和借词的冲突是可以解决的，相互影响可以存在于由于语言接触而产生的两个系统中。同时还提出了双向词汇扩散的现象和声调层级的现象和概念。文章在最后给出了大量相关文献，是一篇词汇扩散理论和新语法学派理论争论的经典论文。论文（1996）以英语第三人称单数加 s 为例，利用词汇扩散理论讨论了其变化的各个方面，首先讨论了词到词的扩散，然后讨论了说话人到说话人的扩散，最后讨论了地点到地点的扩散。研究结果表明，词汇的扩散开始时速度较慢，到中期速度加快，到了后期，扩散的速度又开始减慢，这种效应如同滚雪球，呈现出一条 S 形曲线，因此被称为“词汇扩散”的“滚雪球效应”。

五、语音演变数学模型研究

在本世纪初，由于生物遗传算法研究的快速进展，语言进化模型的研究逐渐又被大家所重视，王士元先生和他的语言工程实验室也对语言的进化模型进行了研究。我们选了两篇相关论文，希望这两篇论文能使大家了解语言进化数字模型的研究情况。论文（2003）介绍了一个基于遗传算法的语音系统优化模型。论文采用此模型来研究元音和声调的结构，其方法是用一个假设的能够控制语音系统结构的特定标准来预测最佳的元音和声调系统。以往的研究往往只采用一个标准，当采用两项标准时，这两项标准通常会并入一个纯量函数。研究结果表明，尽管预测系统和被观察系统与实际得到的系统在一致性上还不是很好，但这一研究方向非常有前途。论文（2004）介绍了不同学科对语言研究的新进展，使语言进化的研究重新为人们所重视，由于计算模型领域成果丰硕，利用数学模型和仿真对语言进化各个方面的研究都有了新的进展，论文最后介绍和讨论了几种语言演变和突现的计算研究成果。

六、语音识别研究

王士元先生有深厚的理工科背景，早在上世纪 50 年代他就将声学语音学的知识用于语音合成的研究，并取得了很好的效果。论文（1958）从语音学、音位学和言语声学的角度讨论了音位与语音表层单位的关系，并基于语音学和音位的理论探讨了利用波形拼接进行语音合

成的理论基础和可能性，同时讨论了利用语图进行拼接的技术，给出了一个合成系统的框架。最后，计算了一个波形拼接合成系统理论上的信息量，讨论了实现该系统的可能性，并提出了“二半音段”（dyad）的概念。论文（1958）讨论了美国英语中一个发音者所必需的语音基本单位应该有43个，但考虑到所有的语音环境，如不同的语音组合、语调、重音等，同时考虑到波形拼接的频谱特性，如浊音的谐波、共振峰和振幅包络等，整个合成系统应有8500个语音片断，最终给出了用这个系统合成的句子样本。

在本世纪初，王先生领导的语言工程实验室在怎样将有用的语音学和语言学信息用于语音识别进行了尝试，并取得了很好的效果，这里选的3篇论文都是基于语音信息的语音识别系统的研究。论文（2000）介绍了一种基于时长信息的汉语数字识别系统。在通常的语音识别系统中，汉语语音的时长一般不会被利用在识别系统中，这项研究将语音的时长信息用于汉语数字识别的解码算法，从而大大地降低了词的错误率。论文（2004）介绍了一种基于语言韵律知识的言语语音识别系统。对于汉语来说，韵律特征是十分重要的，但怎样利用这些特征一直是语音识别中的一个待解决的问题。该系统首先评价了韵律信息的可靠性，并在此基础上，将可靠的韵律信息用语音的识别，建立起一个汉语韵律识别模型。研究表明，韵律信息的可靠性越高，识别率也越高，从而提高了整个语音识别系统的性能。论文（2005）介绍了一种基于“支持向量机”的汉语粤语声调识别系统。粤语有很复杂的声调系统，包含了大量的语言信息，因此充分利用这些信息能够大大提高识别系统的性能。研究表明，使用这种方法，粤语声调的识别率可以达到71.5%，从而能大大提高整个语音识别系统的性能。

王士元先生在语音方面的研究涉及学科广泛，利用数据材料众多，理论性强，因此，上面的介绍肯定会有不少错漏之处，希望读者批评指正。同时也希望广大读者以读原文为主，真正了解王士元先生语音学研究的方法论和理论精髓，推动中国的语言学和语音学的理论发展。

孔江平

2009年9月9日

于香港沙湾径23号港大职员公寓

目　录

Segmentation Techniques in Speech Synthesis[①]

Gordon E. Peterson[1], *William S-Y. Wang*[2] and *Eva Sivertsen*[3]
[1,2]Speech Research Laboratory, University of Michigan, Ann Arbor, Michigan
[3]University of Oslo, Oslo, Norway

A basic method of speech synthesis is described in which discrete segments of recorded utterances are joined together to produce continuous speech. The segments are characterized as (a) containing parts of two phones with their mutual influence in the middle of the segment, and (b) beginning and ending at the phonetically most stable position of each phone. All segments containing the same articulatory sequence have been defined as a dyad. The method of synthesis described includes not only articulatory phones, but also intonation, stress, and duration. A large number of segments is required and various techniques of obtaining the segments for speech synthesis are discussed. The method is limited to a specific dialect, and practically it is limited to a single speaker.

For many centuries scholars have demonstrated a somewhat casual curiosity about the mechanical synthesis of speech. With modern advances in communication technology, however, the possibility of important applications of the artificial creation of speech has been recognized. Among these are the transmission of standardized speech over channels of low information capacity and the transformation of orthography to speech. The former of these two examples could form the basis for telegraph speech and the latter the basis for print reading for the blind or for the production of a speech output in mechanical language translation.

Of primary interest, however, is the basic information about the structure of speech to be derived from efforts to create it artificially. In this case the ultimate objective is to synthesize utterances which are indistinguishable from normal human speech production. It is obvious that a great deal of knowledge about the spectral and also the transient or dynamic aspects of speech would

---

① Originally published in *The Journal of the Acoustical Society of America*, Vol. 30, No. 8, August, 1958.

This research was supported by the Horace H. Rackham School of Graduate Studies of the University of Michigan, and more recently by the Information Systems Branch of the Office of Naval Research under contract Nonr 1224(22), NR 049-122.

be necessary in order to achieve such a synthesis.

In the past, various mechanical devices have been constructed for the synthesis of isolated speech sounds, and some attempts have been made with instruments to create actual continuous speech. More recently, however, the convenience of an electral analog to the human vocal tract has been recognized, and several electronic synthesizers have been developed.

Methods of speech synthesis

There appear to be two basic approaches to the automatic synthesis of speech. One approach involves the use of dynamic electronic or other instrumentation to generate the speech according to a sequence of control signals. These control signals set up specific parameter values which depend upon the language to be synthesized, and the transitional effects between these values are provided by the dynamics of the system. The second general approach involves the use of discrete segments which are connected together to produce the speech. These segments may be of actual human utterances or they may be discrete segments which are artificially generated to form basic units for the synthesis. Both approaches may employ discrete inputs. In the former approach the dynamic properties of the vocal tract are an integral part of the analog, whereas in the latter approach segments are stored which contain the appropriate dynamic properties. This latter approach is that discussed in the present paper.

Quantitative information about the dynamics of the human vocal tract is at

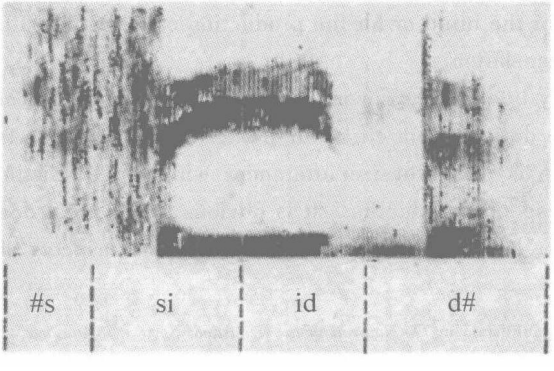

Fig. 1 Broad band sound spectrogram showing the segmentation of the word *seed*.

present very limited. The essential parameters involved in the dynamics of the mechanism have not yet been well defined. One characteristic in particular which has been considerably neglected thus far in speech synthesis is the servo property of the system.

The obvious complexity of the human system suggests that, even if its parameters were well understood, the construction of synthesizers which adequately represent its dynamic properties would be difficult. This suggests the possible convenience of synthesizing speech from recorded segments of actual human utterances, in which the parameters and dynamics of the system are already normally represented. A possible alternative would be to employ segments constructed artificially, but which were individually adjusted and carefully modified until each approximated the properties of normal speech.

The first major study of the synthesis of speech from segments of human utterances was made by Harris.[①] He discussed the theoretical and practical importance of this approach to speech synthesis, and developed instrumentation for programming sequences of recorded segments.

Dyads

The present study was based upon two major assumptions, resulting chiefly from experience with the analysis of speech: (1) that speech is quantized and the intelligibility of speech is carried by the more sustained or target positions of the vowels, consonants, and other phonetic features; and (2) if synthesized speech is to sound natural, the normal dynamics of speech production must be maintained.

On the basis of these two hypotheses it was concluded that the segment junctions should be made in the more sustained portions of the speech signal, and that the transitional sections should form the segmental units for the synthesis. Thus in the method described here the segments consist of adjoining portions of two phonetic units. To construct the segment [si], for example, the [s] and the [i] would be sectioned at about the middle of their durations. The segment to be employed in speech synthesis would thus be formed by the latter portion of [s] attached to the first portion of [i]; accordingly, the dynamics of the articulation of [si] would form a single recorded segment.

① C. M. Harris, *J. Acoust. Soc. Am.* 25, 962-969(1953).

The method is illustrated in Fig. 1 which shows a broad band sound spectrogram (to 3500 cps) for the word *seed*. The vertical lines indicate the positions at which different segments would be joined together if the word were to be constructed artificially.

No language, however, can be reduced to simple unidimensional sequences of phonetic or phonemic elements. As a minimum, consideration must be given to (a) articulatory sequences, (b) tone and intonation, (c) duration, and (d) stress. Thus several different versions of the segment [si] would be required depending upon the conditions of allophonic variation, tone or intonation, duration, and stress which it proved necessary to represent in the synthesis. The emphasis which must be given to any one of the above aspects or features, of course, will vary with the particular language involved. In some languages one aspect may be very highly correlated with another aspect. Such correlations and interrelationships will normally simplify the total segment catalog required.

Throughout the remainder of this paper we shall define *articulation* to include all supra-laryngeal phenomena in the vocal tract. The term *prosody* includes tone and intonation and stress. Duration and linguistic juncture may involve both articulation and prosody.

We define a *segment* as an articulatory sequence pair (including the case in which silence is either the first or second element of the sequence) with a specific prosody. A *dyad* is defined as the set of all segments involving a single articulatory sequence pair and all conditions of prosody associated with that sequence.

Languages differ greatly in their structural patterns of articulation and prosody. Not only is a particular system of synthesis restricted to a given language and a given dialect, but in order to have the segments abut properly it is necessary to derive them from a single speaker. Thus the method is restricted to standardized speech as discussed by Harris.

The articulatory sequences contained in the dyads are not phonemes. The articulatory target sequences may be considered allophones of the phonemes of the idiolect involved (as will be shown later); more than one allophone of a phoneme may be necessary in order to produce normal speech, but all allophones of a phoneme may not be necessary.