****

**本科生毕业论文**



**学 院 计算机科学与技术**

**专 业 计算机科学与技术**

**年 级 2012级**

**姓 名 张洁**

**指导教师 王建荣**

**年 月 日**

****

**本科生毕业论文任务书**

题目：基于声管的语音合成

**学生姓名： 张洁**

**学院名称： 计算机科学与技术**

**专 业： 计算机科学与技术**

**学 号： 3012216058**

**指导教师： 王建荣**

**职 称： 副教授**

**一、原始依据（包括设计或论文的工作基础、研究条件、应用环境、工作目的等。）**

1. **题目背景：**

语音是人与人交流中最为重要的手段之一，也是实现和谐的人机交互的重要方法。语音合成是人机交互中的关键技术，它使得用户可以通过语音从计算机获取信息与反馈，从而使得人与计算机的交互变得更加和谐自然。

语音合成经过多年的发展，衔接合成是目前主要的高质量语音合成方法。尽管他已经接近自然演讲的水平了，其他的语音合成方法仍值得推崇。从长远来看，似乎最有前途的是发音语音合成。而基于声管的语音合成是基于发音机理的语音合成方法的重要组成部分。

**2.工作基础：**

一定的核磁共振（MRI）数据

已有的元音的声管模型

一定的开发基础

**3.研究条件：**

Windows系统PC机，VisualStudio2015，仿真系统等。

**4.应用环境**：

语音合成广泛的应用在多媒体技术，语音技术，声学技术甚至图像处理技术上。应用语音合成系统较为广泛的地方主要是一些已经形成一定规模的市场化组织，如股票、基金的查询，火车、飞机航班的查询，甚至是天气的查询等。

**5.工作目的：**

基于核磁共振（MRI）数据，完善现有的元音的声管模型，使模型能生成元音、辅音，并考虑声道的细节形态特征对声学特征的影响。

**二、参考文献**

[1] Peter Birkholz, Dietmar Jackèl, and Bernd J. Kröger ,“Simulation of Losses Due to Turbulence in the Time-Varying Vocal System”IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL.15, NO.4, MAY 2007

[2] Shinji Maeda,“A digital simulation method of the vocal-tract system” Centre National d'Etudes des Télécommunications, 22301 Lannion, France ,Speech Communication I (1982) 199-229

[3] P Birkholz, D Jackél - Rostocker “Noise Sources and Area Functions for the Synthesis of Fricative Consonants” Informatik Berichte, 2006 - vocaltractlab.de

[4] Parham Mokhtari , Hironori Takemoto, Tatsuya Kitamura “Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches” ATR Cognitive Information Science Laboratories, 2-2-2 Hikaridai, Seikacho, Kyoto 619-0288, Japan Received 20 February 2007; received in revised form 21 June 2007; accepted 9 August 2007

[5]M. H. Krane, “Aeroacoustic production of low-frequency unvoiced speech sounds,”*J. Acoust. Soc. Amer.*, vol. 118, no. 1, pp. 410–427,2005.

[6]D. J. Sinder, “Speech synthesis using an aeroacoustic fricative model,”Ph.D. dissertation, Rutgers Univ., New Brunswick, NJ, 1999.

[7]蔡明琦.融合发音机理的统计参数语音合成方法研究[D].中国科学技术大学,2015.

[8]郑鲁.情感语音发音机理的研究[D].山东师范大学,2011.

[9]王宇光.基于核磁共振成像的汉语普通话元音/i，（s）i，（sh）i/的研究[D].天津大学,2012.

[10]陈志杰,王韦皓.情感语音发音机理的研究[J].长春教育学院学报,2014,24:61-62.

[11]凌震华.基于统计声学建模的语音合成技术研究[D].中国科学技术大学,2008.

[12]卢恒.基于统计模型与发音错误检测的语音合成方法研究[D].中国科学技术大学,2011.

[13]黄南川,邓振杰,王嵬嵬,张皓健.语音合成技术的研究与发展[J].华北航天工业学院学报,2002,03:37-39.

[14]吕鹏.语音合成自然度的研究[D].河北科技大学,2010.

[15]吴振宇.个性特征的语音合成与模拟技术研究[D].华东师范大学,2011.

[16]郑骐.情感语音合成——研究现状与发展[J].宁波广播电视大学学报,2005,03:75-78.

[17]宋文龙.基于说话人自适应训练的统计参数语音合成的研究[D].西北师范大学,2013.

[18]宋阳.基于统计声学建模的单元挑选语音合成方法研究[D].中国科学技术大学,2014.

[19]陈静,李薇,崔忠伟,刘霞.语音合成技术的研究及其发展[J].中国科技信息,2007,14:280-281.

[20]赵学彬.声学、生理框架下的汉语基础元音发展研究[D].首都师范大学,2009.

[21]王磊.情感语音合成中的声源以及基频抖动的分析与建模[D].天津大学,2006.

[22]王永鑫.语音合成系统中关键技术的研究[D].清华大学,2013.

**三、设计（研究）内容和要求**

**1.研究内容：**

基于声管的语音合成是基于发音机理的语音合成方法的重要组成部分。本研究将基于核磁共振（MRI）数据，完善现有的元音的声管模型，使模型能生成元音、辅音，并考虑声道的细节形态特征对声学特征的影响。本课题的仿真实验将用OpengGL和VC++实现。

**2.主要指标和技术参数：**

(1) 数据应当准确 ,可靠

(2) 环境要提前搭好,并保持稳定；

(3) 资料要充分，结构要完整，论述要清晰；

(4) 重要数据及引用他人成果要表明出处；

   (5) 符合我校毕业论文书写规范；

(6) 字数不少于15,000字；

(7) 按时完成毕业论文各阶段工作，不突击、不抄袭；

(8) 每周主动向指导教师汇报工作进度，探讨研究内容；

指导教师（签字）

年 月 日

审题小组组长（签字）

年 月 日

天津大学本科生毕业论文开题报告

|  |  |  |  |
| --- | --- | --- | --- |
| 课题名称 | 基于声管的语音合成 | | |
| 学院名称 | 计算机科学与技术 | 专业名称 | 计算机科学与技术 |
| 学生姓名 | 张洁 | 指导教师 | 王建荣 |
| **一**、课题来源及意义：  语音合成技术(Text To Speech)主要是以计算机为媒介，将文本、Word 等各种文字形式的信息，利用语音学的韵律规则合成语音信号并向外部设备进行输出的一种语言处理技术。它的目的就是为了使人们能够更加容易，清晰的听明白文本的内容，赋予计算机一个“嘴巴”，使它通过“说”的方式“背诵”给人们听。这种利用语音学将文本内容转变成语音的方式称之为 TTS 技术即语音合成技术。目前语音合成技术已经成为各国竞相研究的一个主要内容，它广泛的应用在多媒体技术，语音技术，声学技术甚至图像处理技术上。  语音合成主要分为两种方法，即参数合成法和波形拼接法。较早前语音合成主要是针对人的发音方式，对发声器官参数进行模拟的参数合成法。进入上个世纪 90年代，波形拼接技术已经在英、法、日、德等国家得到了广泛的研制和应用。这种技术较之前的共振峰技术和 LPC 技术在自然度方面已经有了很大的提高，同时利用PSOLA 算法制造的合成器在应用前景上更大，设计结构也既简单又易于实现。  语音合成经过多年的发展，衔接合成是目前主要的高质量语音合成方法。尽管他已经接近自然演讲的水平了，其他的语音合成方法仍值得推崇。从长远来看，似乎最有前途的是发音语音合成。而基于声管的语音合成是基于发音机理的语音合成方法的重要组成部分。声道在人体内部，不容易被观测到，这给声道研究带来了一定的技术困难。早期，Edholm 等曾用往尸体的鼻腔灌膜的方法来获得鼻腔形状。Fant 和 Ladefoge也曾用这种方法测量发音人的口腔形状，但这种方法是有局限性的。后来随着科技的进步，X光摄像、核磁成像等技术都被发明出来，于是对声道的测量也就进入了新的研究阶段。  现在，世界各大国家都在集中力量进行语音合成，这一技术要求很高的领域，尽管我们国家现在只是在通信方面对语音合成投入了比较多的研究精力，但是相信在不久的将来，随着国家队语音行业的重视程度不断提高，人们对语音自然度要求也越来越苛刻，我国将会有更多的企业和部门从事语音合成行业的研究。这样我国将会在语音合成的各个方面与国外的先进技术长期处于竞争的局面，并会在这种良性的竞争中有所收获。  本研究将基于核磁共振数据，完善现有的元音的声管模型，使模型能生成元音、辅音，提高语音合成的清晰度和自然度并考虑声道的细节形态特征对声学特征的影响。同时也将采用高空间分辨率的MRI数据去讨论汉语普通话的元音、辅音，具有很高的可靠性，而且利用高分辨率的三维声道数和高精度的三维数值算法可以很精准的分析声道内部结构的声学特性。分析不同发音的特点将有助于理解语音识别中对于一些不容易辨识的发音犯错误的原理，对语音合成有重大的意义。 | | | |

|  |
| --- |
| 二、国内外发展状况：  Baer(1991)首次用 MRI 一起测量了人体声道的三维形状，实验中采集了两位男性实验对象持续发 10 个元音时候的的轴面、冠状面和平行面的三维 MRI  图片，同时记录了声音信号。Baer 的研究中详细讨论声道的面积函数、声道长度、平行面的宽度与侧面的宽度的关系、平行面的宽度与面积的关系等。  Baer 的研究具有很重要意义，为后面的许多研究着利用 MRI 仪器研究声道的特征奠定了基础。但是，由于当时设备的所限，Baer 的图片比较模糊，存在较大的测量误差。  利用 MRI 测量声道的研究也有涉及到不同的性别、不同年龄段的实验者。Yang 和 Kasuya（1994）第一次利用 MRI 测量不同年龄段的实验对象的声道，分别是小孩、成年女性和成年男性。Yang 和 Kasuya（1996）利用 MRI 测量了三男三女持续发日语五个元音的声道形状，分析不同人之间的前后腔体，认为后  /前腔的体积（back/front cavity volume）和声道的长度可以作为人与人之间的个性差异的一个重要特征。MRI 测量声道数据是一件耗时耗力的工作，数据采集比较困难。即便如此，也有人对大量的实验对象的声道数据进行了分析。Vorperian等（2005）对人在成长过程中声道的变化进行研究，实验涉及到 63个儿童和 12个成年，声道的区域（口腔/后腔）和声道方向的比值（横向对纵向）确定了声道的生长模式。  分支（鼻腔、梨状窝、会厌沟等）是声道的重要组成部分，利用 MRI对声道分支的研究一直是一个研究热点。Dang 和 Honda（1994）对不同的实验对象发鼻音/m/和/n/的鼻腔和声道形状进行了 MRI 测量，并重构了鼻腔的三维形状，提出了准确的计算模型。Dang 和 Honda（1997）进而用机械模型、生体实验以及计算模型等方法首次严谨地证明了梨状窝在人的语音生成及个性化中的作用。  Takemoto 和 Honda 等（2006）利用一个成年的男性的实验对象发日语五个元音的三维声道数据研究喉腔在声道共振方面的作用，证明了喉腔生成第四共振峰。此外，Takemoto（2010）年还利用时域有限差分法讨论喉腔、声道的其他分支（梨状窝、会厌沟、齿间空间）等对于声道共鸣的作用。  MRI还被大量用于很多语言的研究，例如英语，德语，日语，法语等等。Story（1996）年对英语的元音的三维声道形状和面积函数进行了系统、全面的测量和分析。Takemoto(2006)利用一个实验对象发日语五个元音的动态三维和静态三  维声道数据进行对比，发现静态情况舌头等受重力的影响发生微小的偏差。  关于汉语普通话元音的声道的形态学和声学方面的研究非常少。鲍怀翘（1983）利用 X 射线测量的一个女性实验对象在中位矢面上的数据，使用声道数据对汉语的九个元音进行了一个系统的分析，但是由于 X 射线测量的声道非  常模糊，是二维平面上的数据，测量出的声道的面积函数具有很大的误差，另外  由于当时鲍采用的很简单的传输线模型方法（没有考虑声道壁的振动等因素）所  以计算出来的共振峰与测量出的共振峰具有很大的误差，超出 30%。Wang（2008）年，建立了一套系统的汉语普通话的 MRI 数据库，涉及 6 男 2 女，汉语普通话元音、辅音、鼻音，动态二维、静态三维、动态三维数据。然后，对其中的一个女性实验对象的元音的进行了声道面积函数的提取，测量和计算的共振峰的对比，误差在 10%以下。  综上所述，前人对声道的研究具有一下几点：  （1）MRI 的研究包括英语、日语、法语、德语、中文等等。但是对于汉语  普通话的元音的声道的声学的研究极少。  （2）针对声道的部分的研究涉及喉腔及声道的分支，即鼻腔、梨状窝、齿  间空间等。  （3）MRI 实验对象以男性实验对象为主，也包括少数的女性和儿童实验对象。  （4）绝大多数考察元音系统的特性研究都使用一到两个实验对象的数据。  三、研究目标：  基于核磁共振（MRI）数据，完善现有的元音的声管模型，使模型能生成元音、辅音，并考虑声道的细节形态特征对声学特征的影响。  四、研究内容：  发不同性质的声音时，声道的情况是不同的，大致可以将这些情况分为两大类：   1. 发元音的情况：这时声道中的口腔为稳定的某种形状的谐振腔。由声门来的准周期脉冲波激励声道而产生响应。所有的单元音，复元音及复鼻音的元音部分都属于这种情况。 2. 发辅音的情况：此时又可以分为塞音、擦音、鼻音等情况。发塞音时，声道的某部分构成阻碍，使声道完全封闭，由声门来的激励波在此处形成高压湍流，然后突然开放，发出声音。发擦音时，声道的某部分构成为完全封闭的阻碍，使激励在此处形成高速湍流，与该处摩擦而发出声音。而发鼻音时，软腭下垂，鼻腔参加谐振响应。   对于声道的数学模型，有两种观点：一种是将声道看作是由多个不同的截面积的声管串联而成的系统，称为声管模型；另一种是将声道视为一个谐振腔，共振峰就是这个腔体的谐振频率，从这个角度出发来描述声道的模型，即为共振峰模型。本课题采用的是第一种，即声管模型。  本课题需要在现有的元音的声管模型的基础上，完善模型，使模型能够生产元音，辅音，同时要考虑声道的细节形态特征对声学特征的影响，以此来提高语音合成的清晰度、自然度和连续性。为此，需要研究以下内容：   1. 利用核磁共振技术采集中国人发汉语普通话元音辅音时的三维声道数据，然后利用一维和三维的声学分析方法对汉语元音，辅音形态学和声学进行全面，详细分析，重点介绍不同元音、辅音之间的差异和不同实验对象之间的差异级这些差异引起的声学差异。 2. 构建模型，采用新方法实现发音合成器，预防声伪像。   五、研究方法：  针对以上研究内容，本课题采用的研究方法包括：   1. 明确具体需求：通过查询相关文献，或观看相关视频，了解基于声管的语音合成方面的研究状况，结合本课题要求，明确自己的任务。 2. 收集数据：采集MRI数据，并对数据进行一定的预处理。 3. 构建模型：分析数据，采用一些比较好的方法来构建模型，使模型尽可能的完善。   六、进度安排：  2016年1月4日-2016年1月24日 阅读相关文献，撰写开题报告  2016年1月25日-2016年3月1日 阅读相关文献，查阅相关资料。  2016年3月2日-2016年3月12日 收集数据  2016年3月13日-2016年3月18日 搭配仿真系统所需的环境  2016年3月19日-2016年4月18日 构建模型  2016年4月19日-2016年5月1日 期中教学检查，修改相关内容；做基本测试，分析使用结果；  2016年5月2日-2016年6月1日 完成毕业论文的撰写  2016年6月2日-2016年7月3日 准备并完成答辩  七、实验方案的可行性分析：  语音合成技术是信息处理领域的一个重要分支，实现计算机文本文件语音合成，就是让计算机开口说话，这也是人工智能的一个重要研究方向。语音合成是一门跨学科的技术，涉及到自然语言理解、语音学、信号处理、心理学、声学、多媒体技术等多个领域，是当今世界强国竞相研究的热门技术之一。语音合成技术必将在更为广泛的范围内得到推广和应用。从长远来看，似乎最有前途的是发音语音合成。它不受任何基本的限制，并且超出纯文本-语音合成的应用程序，如发音驱动的面部动画和视觉支持第二语言学习和语音障碍的治疗。此外，语音教育和研究可以受益于发音语音合成。发音语音合成会有很大的发展和应用空间。  由此，此课题是可行的。  八、已具备的实验条件：  Windows系统PC机，OpengGL和VC++，仿真系统等。  九、主要参考文献：  [1] Peter Birkholz, Dietmar Jackèl, and Bernd J. Kröger ,“Simulation of Losses Due to Turbulence in the Time-Varying Vocal System” IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 15, NO. 4, MAY 2007  [2] Shinji Maeda,“A digital simulation method of the vocal-tract system” Centre National d'Etudes des Télécommunications, 22301 Lannion, France ,Speech Communication I (1982) 199-229  [3] P Birkholz, D Jackél - Rostocker “Noise Sources and Area Functions for the Synthesis of Fricative Consonants” Informatik Berichte, 2006 - vocaltractlab.de  [4] Parham Mokhtari , Hironori Takemoto, Tatsuya Kitamura “Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches” ATR Cognitive Information Science Laboratories, 2-2-2 Hikaridai, Seikacho, Kyoto 619-0288, Japan Received 20 February 2007; received in revised form 21 June 2007; accepted 9 August 2007  [5]M. H. Krane, “Aeroacoustic production of low-frequency unvoiced speech sounds,”*J. Acoust. Soc. Amer.*, vol. 118, no. 1, pp. 410–427,2005.  [6]D. J. Sinder, “Speech synthesis using an aeroacoustic fricative model,”Ph.D. dissertation, Rutgers Univ., New Brunswick, NJ, 1999.  [7]蔡明琦.融合发音机理的统计参数语音合成方法研究[D].中国科学技术大学,2015.  [8]郑鲁.情感语音发音机理的研究[D].山东师范大学,2011.  [9]王宇光.基于核磁共振成像的汉语普通话元音/i，（s）i，（sh）i/的研究[D].天津大学,2012.  [10]陈志杰,王韦皓.情感语音发音机理的研究[J].长春教育学院学报,2014,24:61-62.  [11]凌震华.基于统计声学建模的语音合成技术研究[D].中国科学技术大学,2008.  [12]卢恒.基于统计模型与发音错误检测的语音合成方法研究[D].中国科学技术大学,2011.  [13]黄南川,邓振杰,王嵬嵬,张皓健.语音合成技术的研究与发展[J].华北航天工业学院学报,2002,03:37-39.  [14]吕鹏.语音合成自然度的研究[D].河北科技大学,2010.  [15]吴振宇.个性特征的语音合成与模拟技术研究[D].华东师范大学,2011.  [16]郑骐.情感语音合成——研究现状与发展[J].宁波广播电视大学学报,2005,03:75-78.  [17]宋文龙.基于说话人自适应训练的统计参数语音合成的研究[D].西北师范大学,2013.  [18]宋阳.基于统计声学建模的单元挑选语音合成方法研究[D].中国科学技术大学,2014.  [19]陈静,李薇,崔忠伟,刘霞.语音合成技术的研究及其发展[J].中国科技信息,2007,14:280-281.  [20]赵学彬.声学、生理框架下的汉语基础元音发展研究[D].首都师范大学,2009.  [21]王磊.情感语音合成中的声源以及基频抖动的分析与建模[D].天津大学,2006.  [22]王永鑫.语音合成系统中关键技术的研究[D].清华大学,2013. |
| 选题是否合适： 是□ 否□  课题能否实现： 能□ 不能□  指导教师（签字）  年 月 日 |
| 选题是否合适： 是□ 否□  课题能否实现： 能□ 不能□  审题小组组长（签字）  年 月 日 |

摘 要

基于声管的语音合成是基于发音机理的语音合成方法的重要组成部分。本研究将基于核磁共振（MRI）数据，采用时域模拟方法，用传输线电路TLM来模拟声道，并加入了噪声源模型。模型中，控制声波生成和传播的声波方程通过应用一定的规则转化为离散变量，并在基于一个更现实的对流体动压变化的分布式考虑及基础上进行改进，同时考虑声道的分支将三个不同稀疏矩阵运用数学方法合并成单一矩阵，以此来完善现有的元音的声管模型，使模型能更准确的生成元音、辅音，并且可以成功避免声伪像。

关键词：发音语音合成；声道；时域模拟方法；噪声源；

**ABSTRACT**

Articulatory synthesis based on acoustic pipe is an important part of the articulatory synthesis method which is based on the articulatory mechanism. This research will be based on magnetic resonance imaging (MRI) data, time domain simulation method, using electrical transmission-line(TLM) circuit to simulate the vocal tract and joined the noise sources model. The acoustic equations that govern the generation and the propagation of acoustic waves inside the model were transformed into the discrete variable representation by applying certain rules. We improve the model basing on a more realistic distributed consideration of fluid dynamic pressure changes. At the same time, considering vocal tract branch, we will merge three different sparse matrixes into a single matrix by using mathematical methods. All above is to improve the existing vowel sound tube model and make the model generating vowels and consonants more accurately and avoiding artifacts successfully.

**Key words**： Articulatory synthesis; Vocal tract; Time domain simulation;

Noise Sources

目□□录

*（三号、黑体、居中、目录两字空两格、与正文空一行）*

□□第一章□□*（空两格）*☆☆☆*（小三号、宋体）*…………×

□□□1.1□☆☆☆☆*（空一格、四号宋体）*………………………×

□□□1.2□☆☆☆☆…………………………………………………×

□□□1.3□☆☆☆☆…………………………………………………×

………………

□□第四章□□*（空两格）*☆☆☆*（小三号、宋体）…*………×

□□□4.1□☆☆☆☆…………………………………………………×

□□□4.2□☆☆☆☆…………………………………………………×

………………

□□参考文献*（小三号、宋体）……………………*……………×

□□附录*（小三号、宋体）…………………………*……………×

□□外文资料*（小三号、宋体）*

□□中文译文*（小三号、宋体）*

□□致谢*（小三号、宋体）*

*目录单独排页码, 阿拉伯数字（宋体小五号）连续编排，居中书写。*

第一章 绪论

1.1 研究背景

语音是人与人交流中最为重要的手段之一，也是实现和谐的人机交互的重要方法。语音合成是人机交互中的关键技术，它使得用户可以通过语音从计算机获取信息与反馈，从而使得人与计算机的交互变得更加和谐自然。

语音合成是人机语音交互的主要技术之一，语音合成技术是通过一些机械的、电子的方法产生人造语音的技术，是实现人机语音通信，建立一个有说话能力的口语系统所必需的关键技术，伴随着电子计算机运算和存储能力的迅猛提升，语音合成有了它新的定义——利用电子计算机和一些专业装置模拟人制造语音的技术。语音合成近些年在技术和应用方面都取得了长足的进展，被应用到不同领域[1]。

按照人类言语功能的不同层次，语音合成可分成三个层次，它们是：

1、从文字到语音的合成（Text-To-Speech, TTS）；

2、从概念到语音的合成（Concept-To-Speech, CTS）；

3、从意向到语音的合成（Intention-To-Speech, ITS）；

现阶段，语音合成的研究还是集中到文字到语音的合成这一阶段，也就是

TTS合成系统。

TTS主要以计算机为媒介，将文本、Word 等各种文字形式的信息，利用语音学的韵律规则合成语音信号并向外部设备进行输出的一种语言处理技术。它的目的就是为了使人们能够更加容易，清晰的听明白文本的内容，赋予计算机一个“嘴巴”，使它通过“说”的方式“背诵”给人们听。目前语音合成技术已经成为各国竞相研究的一个主要内容，它广泛的应用在多媒体技术，语音技术，声学技术甚至图像处理技术上。

语音合成主要分为两种方法，即参数合成法和波形拼接法。较早前语音合成主要是针对人的发音方式，对发声器官参数进行模拟的参数合成法。进入上个世纪 90年代，波形拼接技术已经在英、法、日、德等国家得到了广泛的研制和应用。这种技术较之前的共振峰技术和 LPC 技术在自然度方面已经有了很大的提高，同时利用PSOLA 算法制造的合成器在应用前景上更大，设计结构也既简单又易于实现。

语音合成经过多年的发展，衔接合成是目前主要的高质量语音合成方法。从长远来看，似乎最有前途的是发音语音合成。而基于声管的语音合成是基于发音机理的语音合成方法的重要组成部分。声道在人体内部，不容易被观测到，这给声道研究带来了一定的技术困难。早期，Edholm 等曾用往尸体的鼻腔灌膜的方法来获得鼻腔形状。Fant 和 Ladefoge也曾用这种方法测量发音人的口腔形状，但这种方法是有局限性的。后来随着科技的进步，X 光摄像、核磁成像等技术都被发明出来，于是对声道的测量也就进入了新的研究阶段。

本研究将基于核磁共振数据，完善现有的元音的声管模型，使模型能生成元音、辅音，提高语音合成的清晰度和自然度并考虑声道的细节形态特征对声学特征的影响。

1.2 发展现状

20世纪是计算机技术迅猛发展的时代，数字化是这个时代的主题，它的发展使得电子技术简单化了，语音合成的发展也证实了这一点。无论是激励声源还是滤波器都可以用数字化方法去实现，从整个系统来看，除了模数和数模转换器以外，就不再需要特殊的硬件设备。另外，随着个人计算机的成熟和推广应用，语音合成研究所需的基本条件有了很大的不同，研究场合不再局限于大的、著名的一些实验室，语音合成的研究开始普及化了，使得更多的人有机会研究语音合成。语音合成技术迎来了迅速的发展，我们进入了计算机语音合成的新时代。此外，计算机存储容量不断增大，运算和检索速度不断提高，大规模真人语音的存储、检索和拼接成为可以实施的工程方案，产生了新的语音合成理论，引起了波形拼接合成、隐马尔可夫参数合成技术的兴起，出现了基于统计参数的语音合成方法。

随着语音学的不断发展，文语转换技术(TTS)已经从最初的满足于声音的可懂性及连贯性逐渐演变为追求语音的高自然度的特性。现在的语音自然度的研究虽然已经成为语音学家们越来越关注的话题，但是其自然度所达到的程度仍然不能满足人们的需求，因此，目前在语音合成的自然度方面各生产厂商仍然在不断地创新。可见，语音合成在未来将会有更大的商业市场和机遇。

衔接合成是目前主要的高质量语音合成方法。尽管他已经接近自然演讲的水平了，其他的语音合成方法仍值得推崇。从长远来看，似乎最有前途的是发音语音合成。它不受任何基本的限制，并且超出纯文本-语音合成的应用程序，例如，发音驱动的面部动画和视觉支持第二语言学习和语言障碍的治疗。此外，语音教育和研究可以受益于发音语音合成。

语音合成作为信息处理领域的一个重要分支，涉及到自然语言理解、语音学、信号处理、心理学、声学、多媒体技术等多个领域，是当今世界强国竞相研究的热门技术之一。语音合成技术必将在更广泛的范围内得到推广和应用。而且发音语音合成会有很大的发展和应用空间。

现在，世界各大国家都在集中力量进行语音合成，这一技术要求很高的领域，尽管我们国家现在只是在通信方面对语音合成投入了比较多的研究精力，但是相信在不久的将来，随着国家队语音行业的重视程度不断提高，人们对语音自然度要求也越来越苛刻，我国将会有更多的企业和部门从事语音合成行业的研究。这样我国将会在语音合成的各个方面与国外的先进技术长期处于竞争的局面，并会在这种良性的竞争中有所收获。

1.3 本文研究内容和组织结构

1.3.1 研究内容

发不同性质的声音时，声道的情况是不同的，大致可以将这些情况分为两大类：

(1)发元音的情况：这时声道中的口腔为稳定的某种形状的谐振腔。由声门来的准周期脉冲波激励声道而产生响应。所有的单元音，复元音及复鼻音的元音部分都属于这种情况。

(2)发辅音的情况：此时又可以分为塞音、擦音、鼻音等情况。发塞音时，声道的某部分构成阻碍，使声道完全封闭，由声门来的激励波在此处形成高压湍流，然后突然开放，发出声音。发擦音时，声道的某部分构成为完全封闭的阻碍，使激励在此处形成高速湍流，与该处摩擦而发出声音。而发鼻音时，软腭下垂，鼻腔参加谐振响应。

对于声道的数学模型，有两种观点：一种是将声道看作是由多个不同的截面积的声管串联而成的系统，称为声管模型；另一种是将声道视为一个谐振腔，共振峰就是这个腔体的谐振频率，从这个角度出发来描述声道的模型，即为共振峰模型。本课题采用的是第一种，即声管模型。

本文将基于核磁共振（MRI）数据，采用时域模拟方法，用传输线电路TLM来模拟声道，并加入了噪声源模型。模型中，控制声波生成和传播的声波方程通过应用一定的规则转化为离散变量，并在基于一个更现实的对流体动压变化的分布式考虑及基础上进行改进，同时考虑声道的分支将三个不同稀疏矩阵运用数学方法合并成单一矩阵，以此来完善现有的元音的声管模型，使模型能更准确的生成元音、辅音，并且可以成功避免声伪像。

1.3.2 组织结构

本论文共包含五个章节，具体组织结构如下。

第一章：绪论。主要描述研究背景，发展现状以及本文研究内容和组织结构。

第二章：相关理论基础。主要介绍一些本文需要的相关知识。

第三章：模型构建。主要描述整个声道模型构建的各个部分。 。。。。。

第四章：结果与分析。 。。。。。。

第五章：总结与展望。对论文工作进行总结以及未来工作展望。

第二章 相关理论基础

2.1 RIFF/WAV

2.1.1 简介

RIFF全称为子源互换文件格式(ResourcesInterchange FileFormat)，RIFF文件是Windows环境下大部分多媒体文件遵循的一种文件结构，RIFF文件所包含的数据类型由该文件的扩展名来标识，能以RIFF文件存储的数据包括：

·音频视频交错格式数据(.AVI)

·波形格式数据(.WAV)

·位图格式数据(.RDI)

·MIDI格式数据(.RMI)

·调色板格式(.PAL)

·多媒体电影(.RMN)

·动画光标(.ANI)

·其他RIFF文件(.BND)

2.1.2 CHUNK

CHUNK 是组成RIFF文件的基本单元，它的结构如下：

structchunk{

u32id;/\*块标志\*/

u32size;/\*块大小\*/

u8dat[size];/\*块内容\*/

};

·id 由四个ASCII字符组成，用以识别块中所包含的数据。如：‘RIFF’,‘LIST’,‘fmt’，‘data’,‘WAV’,‘AVI’等等，由于这种结构最初是由Microsoft和BM为PC机所定义，RIFF文件是按照little-endian字节顺序写入的。

·size(块大小)是存储在data域中数据的长度，id与size域的大小则不包括在该值内。

·dat(块内容)中所包含的数据是以字(WORD)为单位排列的，如果该数据结构长度是奇数，则在最后添加一个空(NULL)字节。

其中有且仅有两种块:‘RIFF’和‘LIST’块可以包含其他块，而其他块仅能含有数据。

2.1.3 WAV格式

WAVE文件作为作为多媒体中使用的波文件格式之一，它是以RIFF格式为标准的。每个WAVE文件的头四个字节便是”RIFF”。Wav文件由文件头和数据体两部分组成。文件头又分为RIFF/WAV文件标识段和声音数据格式说明段。

WAVE文件是由若干个Chunk组成的。按照在文件中出现的位置包括：RIFF WAVE Chunk，Format Chunk，Fact Chunk(可选),Data Chunk。具体见下图：

-------------------------------------------------

RIFF WAVE Chunk

ID = ‘RIFF’

RiffType= ‘WAVE’

------------------------------------------------- ------------------------------------------

Format Chunk | | 所占字节数 | 具体内容|

ID = ‘fmt’ -------------------------------------------

--------------------------------------------------- | ID | 4 Bytes | ‘RIFF’ |

Fact Chunk(optional) ------------------------------------------

ID = ‘fact’ | Size | 4 Bytes | |

----------------------------------------------------- ------------------------------------------

Data Chunk | Type | 4 Bytes | ‘WAVE’ |

ID = ‘data’ ------------------------------------------

----------------------------------------------------

图2-1 Wav 格式包含Chunk示例 图2-2 RIFF WAVE Chunk

图2-1中除了Fact Chunk 外，其他三个Chunk 是必须的。每个Chunk 有各自的ID，位于Chunk 最开始位置，作为标示，而且均为四个字节。并且紧跟在ID后面的是Chunk大小(去除ID和Size所占的字节数后剩下的其他字节数目)，四个字节表示，低字节表示数值低位，高字节表示数值高位。Format Chunk以fmt为标示，一般情况下 Size为16，如果是18的话，最后有附加信息。Fact Chunk是可选字段，一般当wav文件由某些软件转化而成，则包含该Chunk。Data Chunk是真正保存wav数据的地方，以‘data’作为该Chunk的标示；然后是数据的大小；紧接着就是wav数据。

图2-2以‘RIFF’为标示，然后紧跟着为size字段，该size是整个wave文件大小减去ID和Size所占用的字节数，即 FileLen-8=Size。然后是Type字段，为‘WAVE’，表示是wav文件。

2.2 一些相关物理知识

2.2.1 伯努利定理

在一个流体系统，比如：气流，水流中，流速越快，流体产生的压力就越小。

由不可压，理想流体（不可压缩，不计粘性即粘度为0的流体）沿流管作定常流动时的伯努利定理知，流动速度增加，流体的静压将减少；反之，流动速度减小，流体的静压将增加。但流体的静压和动压之后，称为总压保持不变。动压为 。

2.2.2 雷诺数

一种可用来表征流体流动情况的无量纲数。

 :流体流速 ，：密度， ：粘性洗漱， d：一特征长度

流体力学中表征粘性影响的相似性准则数。

雷诺数较小时，粘滞力对流场的影响大于惯性力，流场中流速的扰动会因粘滞力而衰减，流体流动稳定，为层流。反之，若雷诺数较大时，惯性力对流场的影响大于粘滞力，流体流动较不稳定，流速的微小变化容易发展，增强，形成紊乱。

粘性流体的求解不仅和边界条件有关，而且也和雷诺数有关，若雷诺数很小，则粘性力是主要因素，压力项主要和粘性力项平衡；若雷诺数很大，粘性力项则成为次要因素，压力项主要和惯性项平衡。

2.2.3 巴特沃斯滤波器

巴特沃斯滤波器是一种通频带的频率响应曲线很平坦的信号处理滤波器，也叫最大平坦滤波器。

巴特沃斯滤波器的特点是通频带内的频率响应曲线最大限度平坦，没有起伏，而在阻频带则逐渐下降为零。在振幅的对数对角频率的波特图上，从某一边界角频率开始，振幅随着角频率的增加而逐步减少，趋向负无穷大。

一阶巴特沃斯滤波器的衰减率为每倍频6分贝，每十倍频20分贝。二阶巴特沃斯滤波器的衰减率为每倍频12分贝、三阶巴特沃斯滤波器的衰减率为每倍频18分贝、如此类推。巴特沃斯滤波器的振幅对角频率单调下降，并且也是唯一的无论阶数，振幅对角频率曲线都保持同样的形状的滤波器。只不过滤波器阶数越高，在阻频带振幅衰减速度越快。其他滤波器高阶的振幅对角频率图和低阶数的振幅对角频率有不同的形状。

可扩展

2.2.4 电路相关

刚度：材料或者结构在受力时抵抗弹性变形的能力。

电导：电阻的倒数，单位：西门子。Grad

电纳：电抗的倒数，单位：西门子。Srad

电抗：交流电通过电感或者电容压降时，电压与电流之比，虚数表示，单位：欧姆。

2.3 语音学相关知识

2.3.1 声道

声道是很多动物和及人类都有的一个腔室，从声源（哺乳动物是喉头，鸟类则是鸣管）产生的声音经由此处滤出。在哺乳动物中，声道包括喉腔、咽头、口腔和鼻腔，且在一些非人类的哺乳动物中，有些亦包括气囊。

2.3.2 声带

声带是位于喉部的两瓣左右对称的膜状解剖结构，主要功能是振动以及发声，声带肌肉受迷走神经的控制，可以调整声带的张力，以改变振动频率。

在呼吸时，声带张开，允许肺部与外界的空气交换；在憋气时，声带关闭。在说话、唱歌等动作时，声带通过与空气的相互作用产生振动。声带振动产生的声波是语音中浊音的声源，气流通过气管到了喉部，因为声门附近的气道比较窄，所以气流通过声门的速度加快，根据伯努利定理，气流加快处压力会降低，加上两侧声带粘膜极为松软，于是两片声带就向中央靠近，合拢。紧接着当声门闭合之后，来自肺部的气流无法通过，累积的气体压力降再度将声门推开，如此周而复始便产生振动。

2.3.3声门

声门是两瓣声带之间的开口。

声门是肺部压出的空气通过声带的出口。发声过程中牵涉到声门的音素称为声门音。声门的大小受到声带的控制，不同的声门大小将导致不同的语音音色，声门闭合时间与气流呼出时间协调一致时，才可能出现自然的噪音。发声时，声门闭合呈现“1”形。闭合状态下 ，呼出气流通过声门，使声带产生振动。

在言语发声中，以声带的振动作为声音源的音称为浊音。

2.3.4发音部位

发音部位在语音学上指的是辅音发音时，口腔或者咽腔中受到阻碍的位置。

2.3.5 发音方法

发音方法指的是发音时，喉头、口腔、鼻腔节制气流的方式和状态，包括发音时构成阻碍和克服阻碍的方式，气流强弱的情况及声带是否振动等几个方面。

2.3.6 协同发音

协同发音是指发音时在声道中的两个或者偶遇多个不同的部位形成阻碍，这两个阻碍可能同样是完全阻塞（如协同发音的塞音[kp]），或其中一个的阻碍程度比较轻，（如圆唇化的软腭塞音[kw]）Catford称前一种情形为‘等同协同发音’，后一种情形为‘非等同协同发音’。

2.3.7 辅音

辅音又叫子音，是气流在口腔或咽头受到阻碍而形成的音。

发音时气流受到发音器官的各种阻碍，声带不一定振动，不够清晰响亮的音素叫辅音，气流从肺里出来时不一定振动声带，通过口腔时受到一定的阻碍，这种主要依靠阻碍发出的音叫做辅音。

辅音是相对元音而言的，而元音是发音时从肺部呼出的气流通过共鸣器作用的口腔，阻力极小并无摩擦声音的语音。辅音是构成音节的重要组成部分，是区别于元音而称呼的。

* 辅音按照发音部位可以分为：

唇音、双唇音、唇齿音、舌尖音、齿音、卷舌音、齿龈音、齿龈后音、舌面音、硬颚音、唇硬颚音、软腭音、唇软腭音、小舌音、舌根音、咽音、会厌音、喉音。

* 辅音按照发音方法可以分为：

1. 鼻音：发音时，气流的口腔通路闭塞，软腭下垂，带音的气流从鼻腔流出。
2. 塞音(爆音)：也叫爆发音或者破裂音，发音时气流完全阻塞，然后突然放开，让气流出来而形成声音。
3. 擦音：也称“摩擦音”。发音时，气流通路没有完全闭塞，但很狭窄，气流是从窄缝中挤出，因摩擦而成音。
4. 塞擦音：成阻时气流通路先闭塞，而后转为窄缝状态。发音开始时和塞音一样，收尾时和擦音一样，所以叫塞擦音。
5. 清浊音：辅音发音时声带的振动模式，发浊辅音时，声带有充分振动。发清辅音时，声带完全不振动。
6. 边音：发音时，用舌头挡着口腔中央部分的气流通路，使气流从舌头两边流出。
7. 近音（无摩擦通音）
8. 闪音（弹音）
9. 颤音

* 英语中28个辅音音素

(1)清辅音：[p]、[t]、[k]、[f]、[s]、[]、[]、[h]、[t]、[tr]、[ts]。

(2)浊辅音：[b]、[g]、[v]、[z]、[ʒ]、[e]、[d ʒ]、[dr]、[dz]。

(3)鼻音(浊辅音)：[m]、[n]、[]。

(4)舌侧音(浊辅音)：[l]、[r]。

(5)半元音(浊辅音)[w]、[j]。

第三章 模型构建

3.1 声学模型

本课题中采用声道系统的声学模型包括咽管，口腔，以及鼻腔。

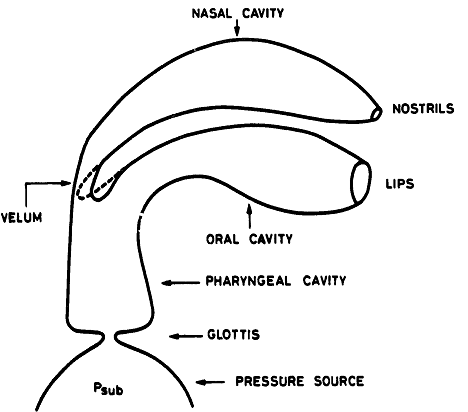


图3-1 声道系统模型

在该模型中，最内层的咽管通过代表声门孔的狭窄的收缩部分连接到压力源，这个模型中忽略了气管导管(在声门与压力源即肺之间)。当频率低于4KHZ时，声道内的声学波可以被看作是平面波，这是被大家所广泛接受的。

有不同的技术来模拟离散管中声波的传播模型。最常用的技术是基于波数字滤波器,或者基于传输线电路模型的直接数值模拟，或者是基于时域-频域的混合仿真系统模拟声道。每种方法都有其特有的优点和缺点，我们的声学模拟是基于直接数值模拟的传输线电路模型。TLM很容易解释和描述，因为它是直接基于声学和电子系统进行的类比。此外，它没有限制每个管部分的长度。

我们用相邻的不同长度的圆柱管部分来代替声道管。每个管部分由图XX所示的T型网络构成，这样整个管道就可以等价于传输线电路。在发辅音时声门由面积为0.3,长度为7的管来表示。声门到嘴唇之间的面积函数由40个等价的管来表示。

3.2 传输线电路模型

1.1.1□☆☆☆*（四号、黑体、顶格）*

□□☆☆☆☆☆☆☆☆☆正文*（小四号、宋体） 宋体五号，居中，位于表上*

表1-3□□☆☆☆

|  |  |  |
| --- | --- | --- |
|  | **×××** | **×××** |
| ××× | ××× （*宋体五号，垂直居中*） | ××× |
| *×××* | ××× | ××× |

*（表与正文空一行）*

□□☆☆☆☆☆☆☆☆☆☆☆☆☆

*（下一章另起一页）*

第二章□□☆☆☆☆☆*（居中、小三号、黑体）*

2.1□☆☆☆*（四号、黑体、顶格）*

2.1.1□☆☆☆*（四号、黑体、顶格）*

□□☆☆☆☆☆☆☆☆☆正文*（小四号、宋体、空两格）*

总经理

经营科

后勤中心

销售科

质量检测科

设备科

副总经理

秘书室

办公室

图2-5□□**×××**组织结构图

*宋体五号居中，位于图下*

*图名与下文空一行*

*正文开始标注页码：*

*位置：页面底端（页脚）；*

*对齐方式：居中*

*正文单独排页码, 阿拉伯数字（宋体小五号）连续编排，居中书写。*

参考文献（*四号、黑体、顶格*）

[1]凌震华. 基于统计声学建模的语音合成技术研究[D]. 合肥: 中国科学技术大学, 2008.

[1]Liu G J，Yang P Y，Zhang W *et al*．Research on separation of minor elements from coal during combustion[J]．Journal of China University of Ming and technology，2000，6(1)：62—66．

[2]岑可法，倪明江，落仲泱等．循环流化床锅炉理论、设计与运行[M]．北京：中国电力出版社，1998．10—12．

[3]Li C M, Anbulagan. Heuristics based on unit propagation for satisfiability problems[A]. In: Proceedings of the International Joint Conference on Artificial Intelligence [C]. Nagova, Japan ： 1997. 232 — 238.

[4]李 宁．煤高温燃烧过程中脱硫反应机理的研究及其工业性的应用[D]．杭州：浙江大学，2000．

[5]ISO 5660-1. Fire Tests-Reaction to Fire-Rate of Heat Release from Building Products[S]. 1993.

[6]Hansen P A. Fires in Types, Heat Release and Response of Vehicles[R]. Norway: Norwegian Fire Research Laboratory, 1995.

[7]Kureha, Kagaku, Kabushiki Kaisha. Method of treatment for stoma-peripheral inflammation diseases[P]. US: 5556622, 1996-09-17.

[8]OASIS. Universal description , discovery and integration[EB/OL]. http://www .uddi . org, 2001-11-14.

[9]Anderson G W. A Burning Rate Model for Charring Materials[Z]. National Institute of Standards and Technology: Gaitherburg, 1997.

*（以上，如果需要两行的，第二行文字要位于序号的后边，与第一行文字对齐。中文的用五号宋体，外文的用五号Times New Roman字体。）*

外文资料

中文译文

*中文译文单独排页码, 阿拉伯数字（宋体小五号）连续编排，居中书写。*

致 谢