**面向数字人应用的人体建模和多模态驱动手势动作生成 张颖颖 指导教师 曾鸣 厦门大学**

2029320381452802095400

**本 科 毕 业 论 文 （设 计）**

**（主修专业）**

# 面向数字人应用的人体建模与

# 多模态驱动手势动作生成

**Human Body Modeling and Multimodal Driven Gesture Generation for Digital Human Applications**

姓 名：张颖颖

学 号：21620182203547

学　 院：信息学院

专 业：数字媒体技术

年 级：2019级

校内指导教师： 曾鸣 副教授

**二〇二三年 4月 10 日**

**厦门大学本科学位论文诚信承诺书**

本人呈交的学位论文是在导师指导下独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合相关法律规范及《厦门大学本科毕业论文（设计）规范》。

该学位论文为（ ）课题（组）的研究成果，获得（ ）课题（组）经费或实验室的资助，在（ ）实验室完成（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明）。

本人承诺辅修专业毕业论文（设计）（如有）的内容与主修专业不存在相同与相近情况。

学生声明（签名）：

年 月 日

致 谢

值此论文即将完成之时，请允许我像对我的毕业设计有过帮助的人们表示衷心的感谢！

首先，我要感谢我的导师曾鸣老师，从选题的确定，到实验的指导，大纲的拟定等等，曾鸣老师都在百忙之中给予非常重要的指导。我的毕设能够顺利完成，首先要感谢的就是曾鸣老师，对我的毕设提出很多专业指导，以及很有建设性的建议。

其次，我要感谢Vcg实验室的同学们的帮助，不论是问题的答疑解惑还是工程的部分技术支持，都离不开同学们的帮助与引导。

最后，我要感谢我的父母与朋友，感谢你们对我生活上的关爱与支持。

摘　要

（黑体小三号）

（--未改）随着信息的发展，出现了越来越多的非结构化信息。并且非结构化信息在政府和企业等的决策中扮演着重要的角色。如何将非结构化数据有效的管理起来，能够进行数据和知识挖掘，提取当中的隐含信息，提供一种形象的可视分析，为政府和企业决策提供支持成为当今亟待解决的主要问题。

（宋体小四号，1.5倍行间距，两端对齐）

本文以北京市科委的指数统计文档为研究对象，主要任务是针对以北京市科委的指数统计文档为代表的非结构化信息的抽取和企业指标信息的可视分析。主要工作包括三个方面：第一，设计了一套以北京市科委的指数统计文档编写规范为标准的确实可行的信息抽取算法；第二，针对抽取出来的指标信息，借助于Dundas可视化工具进行可视分析；第三，完成了一个满足客户需求的企业信息库管理系统。

论文从项目背景出发，介绍了系统开发的背景和研究价值。然后，详细介绍了企业指标信息智能处理的可行性和算法设计，以及企业指标信息可视分析的原理及其实现。再次，论文详细阐述了系统的需求，具体介绍了企业信息库管理系统的设计及其实现，最后论文针对企业信息库管理系统进行了分析和评价，并指明了下一步的改进计划。

（宋体小四号）

（黑体小四号）

关键词：非结构化信息；信息可视化；可视分析

**Abstract**

（未改）With the development of information, there has been an increasing number of unstructured information. And it plays an important role in decision of government and enterprise, etc. How to manage the unstructured information efficiently, mine the data and knowledge, extract the implicit information, provide a visual image analysis, and then support the government and enterprise's decision have become the main issues to be settled urgently.

（Times New Roman小四号，1.5倍行间距，两端对齐）

（Times New Roman小三号加粗）

In this question for discussion, we mainly have a research in indicator of enterprise documents from the Beijing Science and Technology Commission and try to obtain the indicators of the unstructured information, and then provide a visual image analysis. It includes three aspects: First, to design a set of practical information extraction algorithm; second, through the use of the Dundas Chart toolbox, providing visual analysis; third, completed Enterprise Information Management System which meet customers requirement.

The beginning of the dissertation introduced the background of the project, introduced the background of the system and research value. Second, detailing information extraction algorithms and principles of Information Visualization. Third, the dissertation elaborated the system's requirement, specifically introduced the system design and implementation. Finally, some possible improvements and future works were presented.

（Times New Roman小四号加粗）

（Times New Roman小四号）

**Key words:** Unstructured Information; Information Visualization; Visual Analysis

目 录

[第一章 绪论 1](#_Toc230575357)

（小四号黑体）

（四号黑体）

（黑体小三号）

[1.1 研究背景 1](#_Toc230575358)

[1.2 研究现状 1](#_Toc230575359)

[第二章 系统相关技术概述 1](#_Toc230575360)

[2.1 非结构化信息处理 1](#_Toc230575361)

（小四号宋体）

[2.1.1 非结构化信息管理概述 1](#_Toc230575362)

[2.1.2信息抽取技术](#_Toc230575363) 5

[2.2 信息可视化 6](#_Toc230575364)

[2.3 其它系统技术介绍 7](#_Toc230575365)

[2.3.1 ASP.NET简介 7](#_Toc230575366)

[2.3.2 ASP.NET AJAX简介 8](#_Toc230575367)

[2.3.3 ASP.NET Ajax Control Toolkit组件 9](#_Toc230575368)

[2.3.4 Dundas Chart工具箱简介 13](#_Toc230575369)

[2.4 本章小结 15](#_Toc230575375)

[第三章 非结构化信息处理和可视分析 17](#_Toc230575376)

[3.1 企业指标信息统计分析设计方案 17](#_Toc230575377)

[3.2 企业指标信息的智能处理 19](#_Toc230575378)

[3.2.1企业指标信息文档的结构分析 20](#_Toc230575379)

[3.2.2指标信息的提取算法设计 22](#_Toc230575380)

[3.2.3指标值的提取算法设计 24](#_Toc230575381)

[3.3 信息可视化的设计方案 26](#_Toc230575382)

[3.3.1信息可视分析过程模型 28](#_Toc230575383)

[3.3.2基于Dundas的信息可视分析设计 30](#_Toc230575384)

[3.4 本章小结 31](#_Toc230575385)

[第四章 企业信息库管理系统的实现 32](#_Toc230575386)

[4.1 系统概述及功能 32](#_Toc230575387)

[4.1.1开发背景与系统目标 32](#_Toc230575388)

[4.1.2 系统功能和模块划分 34](#_Toc230575389)

[4.2 系统的框架设计 36](#_Toc230575390)

[4.3 指数统计模块的实现 38](#_Toc230575391)

[4.3.1统计分析模块的实现 38](#_Toc230575392)

[4.3.2问卷管理模块的实现 39](#_Toc230575393)

[4.4 文档资源库模块的实现 40](#_Toc230575394)

[4.5 系统维护模块的实现 42](#_Toc230575395)

[4.5.1用户管理子模块的实现 43](#_Toc230575396)

[4.5.2角色管理子模块的实现 44](#_Toc230575397)

[4.5.3文档类型定义子模块的实现 46](#_Toc230575398)

[4.5.4数据库备份&还原的实现 48](#_Toc230575399)

[4.7 本章小结 49](#_Toc230575400)

[第五章 系统测试及运行结果 50](#_Toc230575401)

[5.1 系统测试 50](#_Toc230575402)

[5.2 运行结果 50](#_Toc230575403)

[5.2.1统计分析模块的运行结果 50](#_Toc230575404)

[5.2.2问卷管理模块的运行结果 51](#_Toc230575405)

[5.2.3文档资源库模块的运行结果 52](#_Toc230575406)

[5.2.4用户管理子模块的运行结果 53](#_Toc230575407)

[5.2.5角色管理子模块的运行结果 54](#_Toc230575408)

[5.2.6文档类型定义子模块的运行结果 55](#_Toc230575409)

[5.2.7数据库备份及还原的运行结果 56](#_Toc230575410)

[5.2.8 改善用户体验的工作 57](#_Toc230575411)

[5.3 本章小结 58](#_Toc230575412)

[第六章 总结与展望 51](#_Toc230575413)

[6.1 总结 51](#_Toc230575414)

[6.2 工作展望 1](#_Toc230575415)

[参考文献 63](#_Toc230575416)

**Contents**

（Times New Roman四号加粗）

（Times New Roman小三号加粗）

[Chapter 1 Preface](#_Toc230076798) 1

[1.1 Introduction 1](#_Toc230076799)

[1.2 The Structure of This Dissertation 2](#_Toc230076800)

[Chapter 2 System Related Technologies Outline 4](#_Toc230076801)

[2.1 Unstructured Information Management 4](#_Toc230076806)

（Times New Roman小四号加粗）

[2.1.1 Introduction of Unstructured Information 4](#_Toc230076807)

[2.1.2 Information Extraction 5](#_Toc230076808)

[2.2 Information Visualization 7](#_Toc230076809)

[2.3 Other Related Technologies Introduce 8](#_Toc230076802)

[2.3.1 Introduction of ASP.NET 8](#_Toc230076803)

[2.3.2 Introduction of ASP.NET AJAX 9](#_Toc230076804)

（Times New Roman小四号）

[2.3.3 ASP.NET Ajax Control Toolkit 11](#_Toc230076805)

[2.3.4 Dundas Chart Toolkit 12](#_Toc230076811)

[2.4 Summary 14](#_Toc230076812)

[Chapter 3 Unstructured Information Management](#_Toc230076813) 16

[3.1 The Design Philosophy of Enterprise Indicators](#_Toc230076814) 16

[3.2 The Design Philosophy of Enterprise Indicators Extraction 1](#_Toc230076815)7

[3.2.1 The Statistics Document’s Structure Analysis](#_Toc230076816) 17

[3.2.2 The Statistics Information Extraction Algorithm](#_Toc230076817) 24

[3.2.3 The Value of Statistic Extraction Algorithm](#_Toc230076818) 33

[3.3 The Design Philosophy of Information Visualization](#_Toc230076819) 35

[3.3.1 Information Visualization Model](#_Toc230076820) 35

[3.3.2 Information Visualization Base on Dundas Chart Toolkit](#_Toc230076821) 36

[3.4 Summary 38](#_Toc230076822)

[Chapter 4 Implementation of Enterprise Infromation Management](#_Toc230076823) 39

[4.1 System Profiler and Function](#_Toc230076824) 39

[4.1.1 Development Background and Overall Objective](#_Toc230076825) 39

[4.1.2 Functional Requirements and Module Division](#_Toc230076826) 40

[4.2 System Architecture](#_Toc230076827) 42

[4.3 Indicators of Statistics Module Design](#_Toc230076828) 43

[4.3.1 Statistical Analysis Module Design](#_Toc230076829) 43

[4.3.2 Questionnaire Management Module Design](#_Toc230076830) 45

[4.4 Document Management Module Design](#_Toc230076831) 46

[4.5 System Maintenance Module Design 48](#_Toc230076832)

[4.5.1 User Management Sub-module Design 48](#_Toc230076833)

[4.5.2 Role Management Sub-module Design](#_Toc230076834) 49

[4.5.3 Document Attribute Management Sub-module Design](#_Toc230076835) 49

[4.5.4 Database Backup and Restore](#_Toc230076836) 49

[4.7 Summary](#_Toc230076837) 49

[Chapter 5 System Testing and the Running Results](#_Toc230076838) 50

[5.1 System Test](#_Toc230076839) 50

[5.2 Running Results](#_Toc230076840) 50

[5.2.1 Statistical Analysis Module Running Results](#_Toc230076841) 50

[5.2.2 Questionnaire Management Module Running Results](#_Toc230076842)  52

[5.2.3 Document Management Module Running Results](#_Toc230076843) 54

[5.2.4 User Management Module Running Results](#_Toc230076844) 55

[5.2.5 Role Management Module Running Results 56](#_Toc230076845)

[5.2.6 Document Attribute Management Module Running Results](#_Toc230076846)  57

[5.2.7 Database Backup and Restore Module Running Results](#_Toc230076847) 58

[5.2.8 Improve the System-experience](#_Toc230076848) 59

[5.3 Summary](#_Toc230076849) 60

[Chapter 6 Conclusions and Future works](#_Toc230076850) 61

[6.1 Conclusions](#_Toc230076851) 61

[6.2 Future works](#_Toc230076852) 62

[References](#_Toc230076853) 63

Acknowledgements 64

# 第一章 绪论

（奇数页页眉为当前章名，宋体、小五号；从第一章开始后增加页眉，之前（即论文前置部分）不要有页眉。每一章都从奇数页开始，封订的时候在右边）

（黑体四号）

（黑体小三号加粗）

## 1.1研究背景

当今数字虚拟人技术在游戏、电影和虚拟现实等领域得到了广泛应用。数字虚拟人的高逼真度和互动性，对于提高用户体验和创造更真实的虚拟世界至关重要。然而，在科学研究领域，构建真实的数字虚拟人仍需要进行漫长的探索和研究。在《2023数字人产业发展白皮书》中提到，数字人可以划分为三种，分别是“内容/IP型”、“功能服务型”和“虚拟分身型”，其中这三种数字人的核心竞争力分别包括“形象的艺术性，IP的打造”，“只能交互能力”和“沉浸化、实时化，体验感”，因此现如今的数字人如果能拥有独特的形象，有真实自然的交互，那么会给这个数字人带来更多的核心竞争力提升。同时，白皮书还提到，当前数字人拥有以下三方面特征，分别是“人的外观”、“人的行为”和“人的思想”。为了掌握这三方面特征，我们可以实现一个数字人系统，能够自定义人的外观，同时能自动生成人的行为来表达人的思想，那么这个数字人就拥有更完善的能力与个性。基于这一研究背景，本研究旨在探究数字虚拟人的人体建模和多模态驱动手势动作生成方法，以提高数字虚拟人的逼真度和表现力。

## 1.2研究现状

在人体建模方面，传统的建模方法需要花费大量的时间和精力，而且难以满足个性化需求。近年来，许多游戏开发商都开发了个性化的捏脸系统，并不断迭代提高其自由度，旨在让玩家能够创造属于自己的人物角色，增加游戏的自由度和可玩性。基于这一研究现状，我们的数字人系统也实现了个性化的人体建模模块。

另一方面，除了具备人的形象，数字人还应该具备人的动作行为，才能更加完整。国内外各大厂商都在探索更加多样化和个性化的数字人手势和动作表达方式，以提高用户体验。针对数字人的手势和动作，目前有三种主要的解决方案：

（1）采用动捕技术，即在特定情境下的行为进行动作捕捉，并生成对应手势和动作。

（2）另外，一些对话系统直接复用之前已有的一些动作模组来表达意图。然而动作复用带来的问题很明显，例如在游戏中，由于NPC手势和动作的不断重复，会导致审美疲劳的问题。

（3）使用人工智能实现动作生成，使数字虚拟人呈现出多样、真实、自然的动作和手势。多模态驱动是一种基于人工智能技术的动作生成方法，它可以利用语音、姿势、面部表情等多种信息来生成自然流畅的动作和手势，从而增强数字虚拟人的表现力和互动性。这也是我们数字人系统实现的一个模块，旨在探究多模态自动手势动作生成方案。

## 1.3研究目的和意义

我们期望数字人具备以下能力：

1、自定义形象能力：能够自定义外观、拥有特殊相貌和体型的数字人。

2、表达能力：能够通过人体姿态、手势和面部表情进行情感表达，以及呈现人类动作。

为了让数字人拥有自定义外观、表情和动作，需要利用人体建模技术。此外，我们希望有一种动作生成技术，可以通过语音和情感等多种模态输入来驱动数字人自然地表现出相应的手势和动作，以显著提升数字人的表达能力，从而降低虚拟主播、游戏NPC等多种对话和直播场景的开发和管理成本。此技术还有望为增强现实、虚拟现实和智能机器人等多个领域带来革命性变化。因此，进一步研究和探索该技术的应用前景具有重要意义。

## 1.4 论文结构

本文设计并实现了一个数字人系统， 主要包含人体建模以及多模态驱动手势动作生成模块。论文结构如下：

第一章：绪论。这章介绍了本文的研究背景和研究现状，阐述了本文的研究目的和意义以及本文的论文结构安排。

第二章：人体建模系统的设计与实现。这章中我们介绍了基于Blendshape和基于骨骼驱动的人体建模的原理和不同之处，并基于这些原理实现了自定义人脸建模，表情建模以及体型建模并应用于我们的数字人。

第三章：多模态驱动手势动作生成模块的设计与实现。这章中我们介绍了基于多模态驱动生成手势与动作的系统，同时介绍本系统采用的 CaMN 网络中的5个编码器和2个解码器及其损失函数，基于这个系统实现了多模态驱动手势动作生成并落地相关技术。

第四章：本章介绍了数字人的动作系统原理及其在Unity引擎中的应用。我们结合多模态驱动手势动作生成模块，解决了数字人在Unity引擎中应用此模块遇到的问题。

第五章：本章介绍了数字人在Unity引擎中的综合应用。我们将人体建模模块，多模态驱动手势模块和动作模块相结合，展示了它们之间共同作用后更完备的数字人系统，并对这个综合的系统进行测试。此外，我们还会提及需要完成这个综合数字人系统所需要的其他部分的技术。

第六章：本章介绍了

# 第二章 人体建模系统的设计与实现

## 2.1 引言

这部分介绍了一个基于Blendshape和骨骼驱动的人体建模系统的设计和实现方案。该系统可以对人体的面部、表情和体型进行建模，以实现人体动作合成。第一部分介绍基于Blendshape和骨骼驱动的两种不同的人体建模方法，并对它们进行了比较和分析。第二部分给出实验结果，即根据上述原理方法，介绍在Unity中实现的人脸建模、表情建模和体型建模功能。

## 2.2 基于Blendshape与骨骼驱动的人体建模

### 2.2.1基于Blendshape的人体建模方法

1、Blendshape简介

Blendshape的原理是在两个网格（Mesh）之间做插值运算，从一个网格到另一个网格的融合变形。例如左嘴角从放松到上扬的一个变换，那么左嘴角放松可以作为基形状，左嘴角上扬可以作为变形目标，它们之间就可以实现融合变形。通过制作大量的Blendshape应用在数字人身上就可以实现人体建模。

Blendshape作为数字人物建模和面部动画的技术，它基于一系列预定义的面部表情，每个表情由一组顶点的位移量来表示。通过对这些表情进行线性组合，可以生成几乎无限数量的面部表情，使得面部动画更加真实、自然。

Blendshape的计算公式可以表示为：

(公式2-1)

其中， 表示任意面部表情， 表示第 i 个表情基（Blendshape）， 表示自然表情（neutral expression）， 表示第 i 个表情基的权重，n 表示表情基的数量。该公式将每个表情基与其相应的权重相乘，然后将结果相加，得到最终的面部表情。

2、基于Blendshape的人体建模方法

在基于Blendshape的人体建模中，可以使用基于顶点的Blendshape方法。

在具体实现时，可以通过使用PCA（Principal Component Analysis）算法对不同的Blendshape（如面部表情）进行降维处理，然后通过权重调整来控制变形。此外，还可以通过使用人工神经网络和深度学习算法来实现更加准确和逼真的建模表现。

Blendshape的权重即从初始基形状到目标变形（即另一个拓扑结构一样，但顶点的transform不同的网格）的程度。

### 2.2.2 基于骨骼驱动的人体建模方法

在捏人系统中，使用Blendshape需要大量的工作量。因此，我们探索了基于骨骼驱动的方法，该方法在制作量和性能消耗方面相对较低。

骨骼驱动的捏人方法的核心原理是改变部分骨骼的变换矩阵。例如，在捏脸系统中，我们需要控制面部的关键部位，比如眼、耳、口、鼻、下巴等部位。但是实际上生物学上的控制人脸变形的结构更加复杂。我们需要自己做真实世界中人脸结构与虚拟的重建世界中骨骼的映射，达到较为真实的效果。

这个模块的实现我们使用了骨骼变换和线性蒙皮（Linear Blending Skinning，LBS）的原理：

1、骨骼变换

每个骨骼都可以通过矩阵变换（如旋转、平移、缩放）来控制模型的姿态。右手坐标系下骨骼变换可以使用以下公式表示：

(公式2-1)

其中，是平移矩阵，是旋转矩阵， 是缩放矩阵，是骨骼的总变换矩阵。

2、线性蒙皮

线性蒙皮即LBS，它是由一系列骨骼驱动网格顶点的移动，特定骨骼对附近特定顶点会分配一个权重。而顶点和骨骼的关系，在三维建模软件中如果要手动制作，我们会通过对骨骼刷权重来赋值，即对应这个原理反过来操作：我们选中一根骨骼，对这根骨骼运动会带动的顶点刷上不同的权重，代表骨骼运动对这个顶点的影响轻重。将生成的顶点权重图与骨骼一起计算出最终的顶点位置。当然，也可以使用SSDR来实现从动作中分解出蒙皮信息。

在LBS算法中我们可以利用骨骼当前位置相对于基准位置的变换矩阵和顶点相对于骨骼的权值来计算受骨骼影响下顶点的位置。

具体公式如下：

其中，顶点i的最终位置为 ,是第j个骨骼对于第i个顶点的权重，bonesN是骨骼个数。是骨骼j的平移矩阵，即公式2-1中的，骨骼j的旋转矩阵，即公式2-1中的，骨骼j的旋转矩阵，即公式2-1中的。是i这个顶点初始位置的坐标。

### 2.2.3 Blendshape与骨骼驱动方案对比

Blendshape能够很精准地控制表情或者人脸的建模，且在引擎中方便控制。Blendshape是直接控制了Mesh顶点，因此和骨骼动画完全不冲突，但是一般一个模型可能有几十最多几百根骨骼但是顶点数量却可能有不止几千个，因此注定了Blendshape计算量肯定比骨骼控制计算量大一些。我们举个例子，如果游戏中有100个不同的表情，并且每个表情都需要记录1234个顶点的位置信息，那么总共我们需要记录的位置数量就是100 \* 1234=123400个。不过，如果我们可以将这些表情分解成由123个骨骼驱动的蒙皮动画，那么我们只需要存储一张顶点权重图和100 \* 123=12300个位置，就可以实现这些表情了。可以说，基于骨骼驱动方法对比Blendshape既能够有效地减少存储空间的使用，而且还可以提高动画播放的效率。

但骨骼驱动也有明显的缺点，就是权重的分配有过多的限制，人体建模的细节很难像基于Mesh顶点的Blendshape那样收放自如。Blendshape更像所见即所得，制作的变形目标是什么，这个部位就固定是这么融合计算的， 而骨骼的蒙皮有着Rotation、Position和Scale三种变形矩阵，要达到某种表情效果时，需要对其做大量测试与计算。

## 2.3 实验结果与分析

### 2.3.1 人脸建模功能实现

这里我们分别展示基于Blendshape和基于骨骼驱动的人脸建模功能的实现。

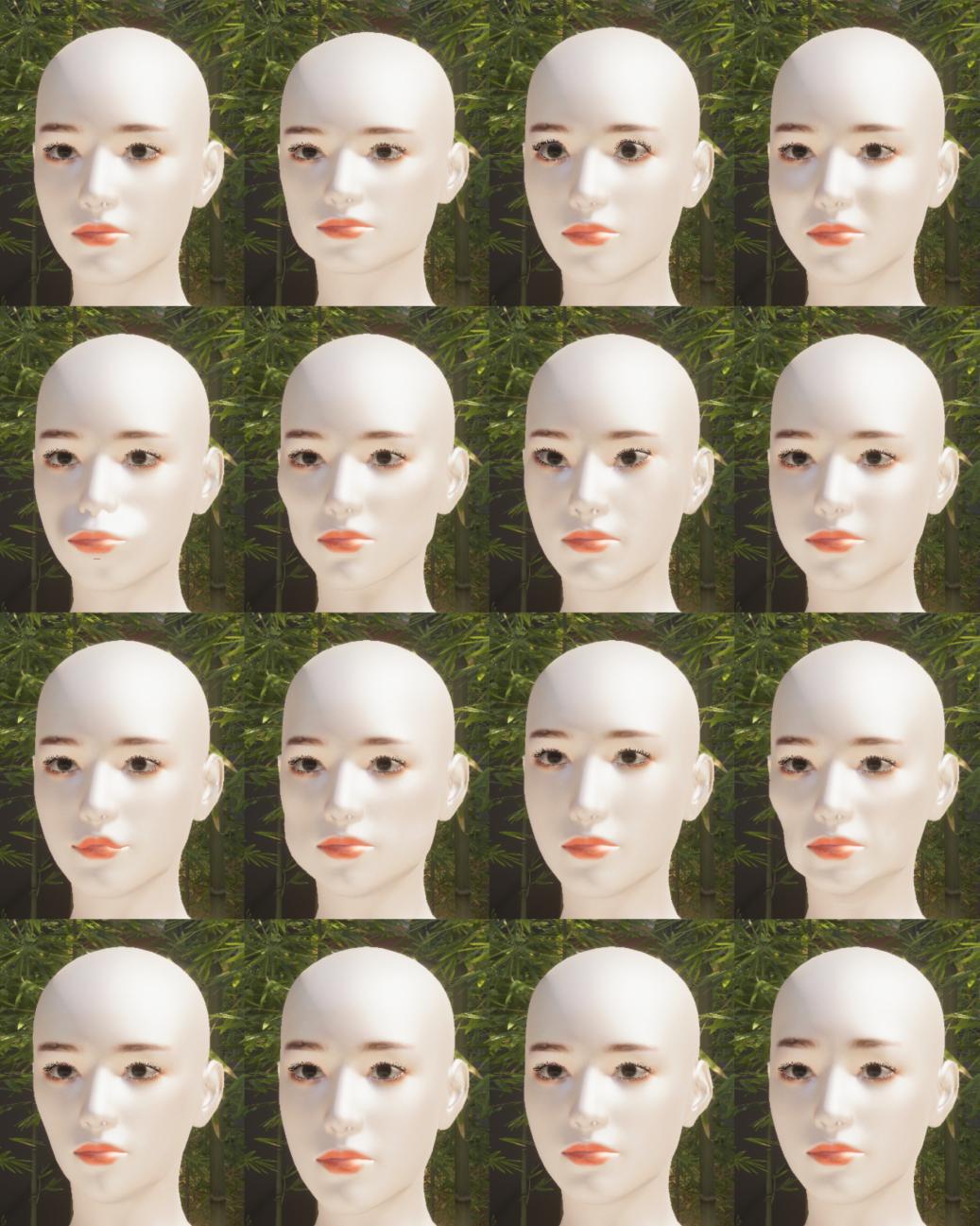
1、基于Blendshape：

我们一共用51个Blendshape作为基来控制人脸建模。包括以下几个部位：

脸型，眼型，嘴型，脖型，眉型，鼻型，耳型，脸颊，下巴。涵盖了人脸的几乎所有部位，每一个部位我们提供多个Blendshape可供用户捏脸，基本上可以实现用户自定义出自己想要的人脸。

我们这只展示了Blendshape一个方向，但实际上我们也实现了逆向。即嘴可以突出，也可以凹进去，只需通过调整Blendshape权重即可完成，而unity的UI上用户只需要通过拉滚动条来调整正向或者是逆向操作。

图2-2展示了用于人脸建模的部分Blendshape，包括头型，眼型，嘴型，脸型，耳型等。第一个子图为初始的人脸建模基准。



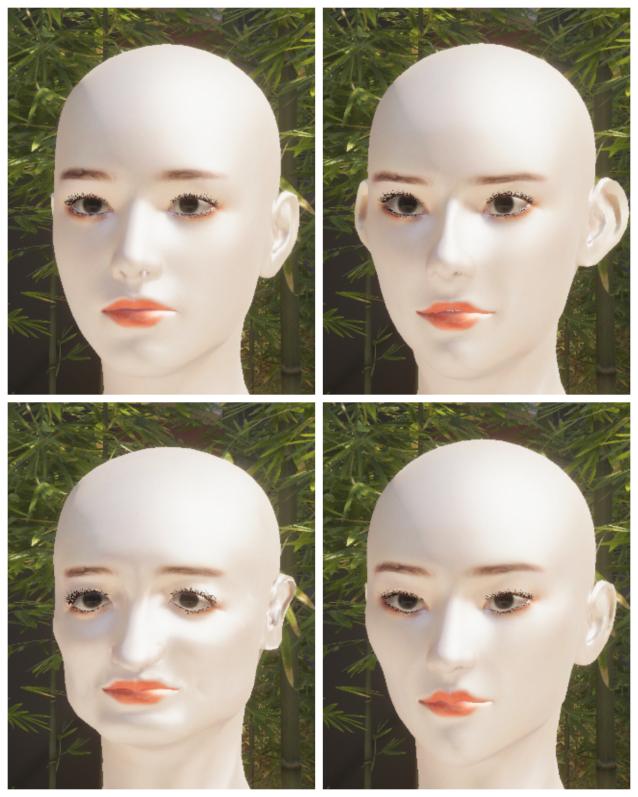
**图2-2 用于人脸建模的部分Blendshape**

用户可以通过我们在unity上设置的UI滚动条来自定义设置人脸，组合上述所说的51个Blendshape进行人脸建模。例如在图2-3中用我们实现的数字人系统进行了捏脸。左边是用户可以操作的UI滑动条，在捏脸部分一共有51个提供给用户编辑。



**图2-3 捏脸效果与对应UI**

图2-4是一些捏脸后的图与原图对比。第一个子图是Blendshape基准模型。

****

**图2-4 捏脸效果**

2、基于骨骼驱动：

我们的数字人一共有251根骨骼来控制整个人体建模，其中159根骨骼用于人脸包括表情的建模，可见人脸骨骼的复杂之处。这159根骨骼同样能够覆盖到Blendshape所覆盖的几乎所有部位。

图2-X我们将251根骨骼显示在屏幕上，点击某个骨骼右边会出现这根骨骼的缩放和位置滚动条（因为旋转值一般不用于人脸建模）。拉动滚动条，数字人的骨骼对应的蒙皮顶点会随之改变。可以看到，直接对骨骼驱动，效果并不如Blendshape好。我们的改进方案是，给用户呈现如Blendshape所呈现的效果，但是这背后可能需要多根骨骼作用，我们将预先进行判断和定义。或者通过SSDR(Smooth Skinning Decomposition With Rigid Bones)技术将各种形变转化为骨骼驱动的蒙皮。



**图2-1 使用骨骼建模实现的捏脸效果**

### 2.3.2 表情建模功能实现

这里我们分别展示基于Blendshape和基于骨骼驱动的表情建模功能实现。

1. 基于Blendshape：

首先，我们此处展示一下我们的初始的Blendshape表情基，即标准放松姿态模型，这个模型上没有任何表情：

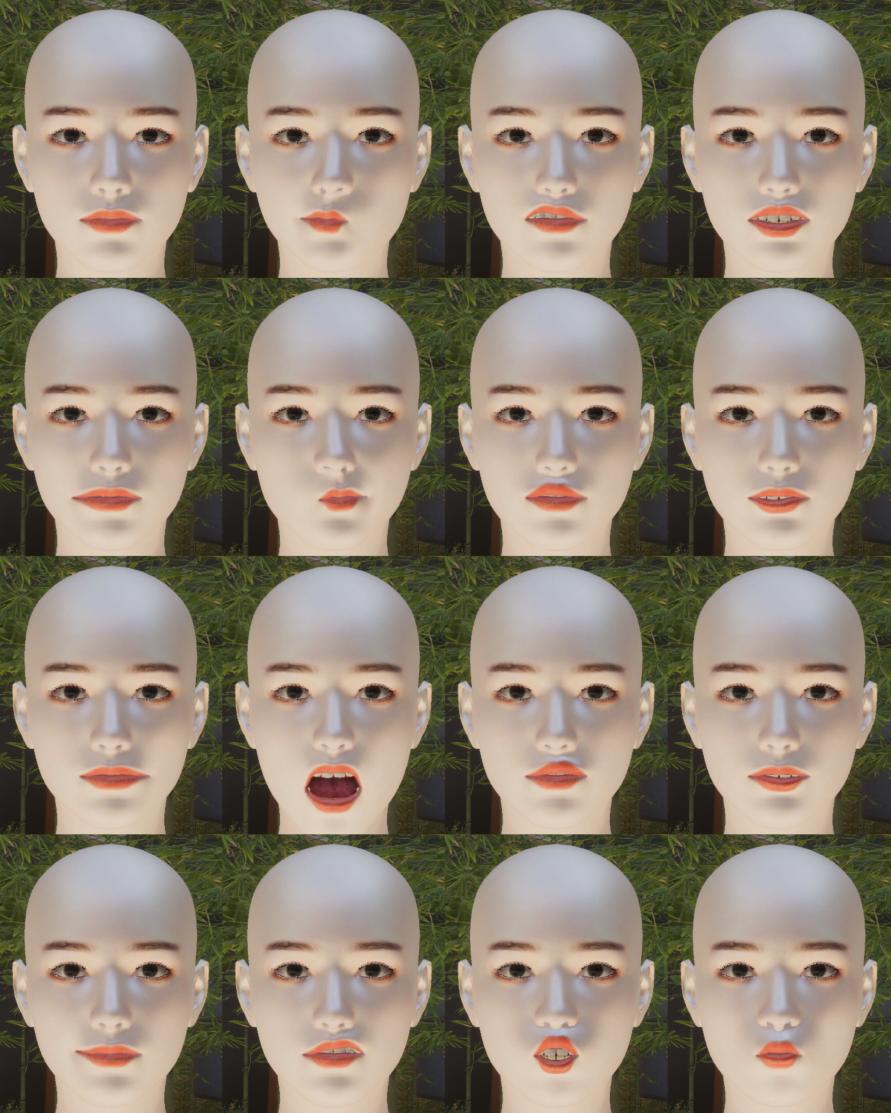


**图2-1 初始的Blendshape表情基**

我们一共使用64个Blendshape作为基来控制表情建模。包括嘴角的各种牵动，嘴的开合，鼻子的抽动，脸颊的鼓起与吸气，下巴的移动，眉毛里外的移动等等，基本能覆盖所有的脸部移动，也基本能够做出所有表情。

需要注意的是，我们的基本Blendshape只需制作出表情变化的最小单位即可。例如，微笑是一个调用众多肌肉的表情，可以由眼睛，眉毛，鼻子，脸颊，嘴等部分构成，再比如嘴的运动，又可以分为上唇、下唇、嘴角、牙齿、舌头等的运动。而上唇，又包含了上嘴唇中部、左部和右部。其中每一个部位的运动方式又是多样的，我们只需制作其中最基本的Blendshape即可。

图2-2为基于Blendshape的表情建模可调整的单个Blendshape的的部分嘴型展示。其中图2-2中的第一个子图是表情基。有一些Blendshape的变化非常细节，但是在真实世界中，人们的各个表情也是由一些细微的肌肉拉伸变化实现的。图2-2中我们可以看到，很多的基本Blendshape是需要成对制作的，比如左嘴角上扬和右嘴角上扬。



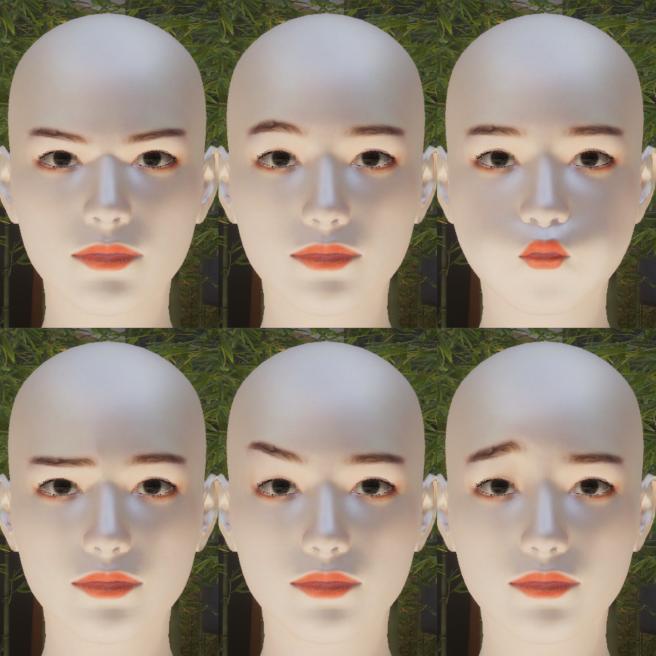
**图2-2 表情建模的部分嘴部Blendshape效果**

图2-3为部分眼睛部位的Blendshape，下图中我们用一个Blendshape来表示成对的，例如闭左眼和闭右眼中的闭左眼。第一个子图为初始的Blendshape表情基，用于对比。



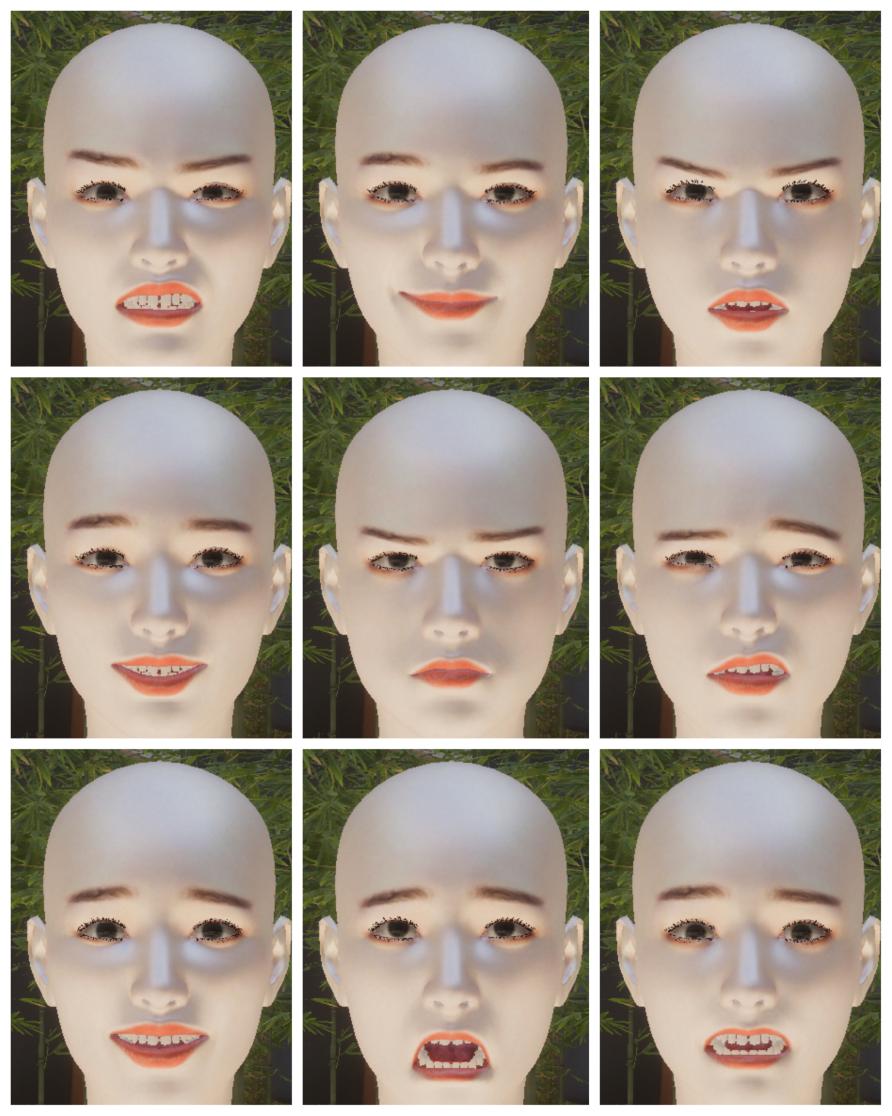
**图2-3 表情建模的部分眼睛Blendshape效果**

图2-3展示其他部位的部分Blendshape，包括眉毛，脸颊等。

****

**图2-3 表情建模的部分其他部位Blendshape效果**

以下图2-4中我们使用这些基本Blendshape来做组合，实现一些具体的表情。其中每一个都是由众多的基本Blendshape组合而成。



**图2-4 使用基本Blendshape组合成不同表情**

以上数字人表情都是由众多基本Blendshape组成，举个例子：例如图2-4中的子图1，就包含了：上唇左侧上下，上唇右侧上下，上唇进出，不张开牙齿的张嘴唇等等很多与嘴相关的Blendshape，还有眼睑上侧下侧、脸颊左侧右侧、左右侧眉毛的内外侧等等一共是29个Blendshape的融合。

2、基于骨骼驱动：

这部分用到的骨骼与上一部分人脸建模功能实现用到的骨骼一致，不过多介绍了。

图2-X使用骨骼驱动的表情建模，图2-X为愤怒表情，unity的UI与人脸建模的骨骼驱动部分一样。



**图2-4 使用基本骨骼驱动合成愤怒表情**

### 2.3.3 体型建模功能实现

这里我们分别展示基于Blendshape和基于骨骼驱动的体型建模功能的实现。

1、基于Blendshape：

我们通过Blendshape来分别实现自定义人体体型的肌肉含量，即健美程度、人体梨型身材程度和肥胖或瘦弱程度等等。

图2-4表示了控制梨形身材程度的Blendshape在不同权重下对人体体型的建模结果。



**图2-4 控制梨形身材程度的Blendshape**

图2-4表示了控制肌肉含量的Blendshape在不同权重下对人体体型的建模结果。



**图2-4 控制肌肉含量的Blendshape**

还有一些是控制腹肌、或皱纹等皮肤质量的细节， 这里就不展示了。人体建模Blendshape之间本身也能叠加使用，人的体型建模外加上面的人脸建模，我们可以得到类似图2-X的拥有特殊外貌和体型的独一无二的数字人。



**图2-4 人体建模和人脸建模的结合**

2、基于骨骼驱动：

关于体型的身体单个部位建模，我们更多的是通过骨骼来驱动。我们的数字人一共有251根骨骼来控制整个人体建模，其中92根骨骼用于除去人脸骨骼的人体建模。具体部位包括各个关节，如手部关节，脚部关节，腿部关节，胸部关节，肩部关节等等，基本涵盖了所有可能的人体部位。

图2-4中，我们基于骨骼进行体型建模，分别是小腿，肚子和胸部。可以看到，骨骼驱动可以很方便调整人体的各个部位。不过骨骼驱动对于细节的处理仍然很难比得上Blendshape，比如对肌肉、皮肤皱纹的模拟等。



**图2-4 基于骨骼的体型建模**

### 2.3.3 unity中的实现细节

1、基于Blendshape：

在unity中，我们需要同时控制人体、眉毛和人物各个服装的Blendshape。例如我们需要保证在类似闭眼睁眼等场景下人脸和眉毛的Blendshape会一起更改，或者更改体型时人物服装也会随着体型更改。

在unity的UI呈现上，我们需要获取所有的skinnedMeshRenderer下的sharedMesh中之前制作了的各个Blendshape的名称，将其blendshap权重分别对应到unity的滑动条权重上，实现表情建模的功能。

需要注意的是，不同数字人会对应制作不同数量的Blendshape，我们需动态获取数字人的Blendshape大小，并动态显示在UI上。

2、基于骨骼驱动：

在unity中，我们动态获取数字人所有的骨骼，包括面部骨骼，动态生成带有所有所需骨骼名称的button，对每个button启用监听事件，然后将其transform的属性对应到UI的滑动条上，这样就可以根据自己的需要使用骨骼来进行人体重建。

但是我们可能会遇到穿模问题，我们的方案是制定好每根骨骼的最小最大的波动范围，而不是仍由其自由发挥，制定好规范以后一般情况下就不会穿模了。还有需要注意的是，unity中的Transform是引用类型，当需要对其进行赋值、修改等操作时需要注意。

# 第三章 多模态驱动手势动作生成模块的设计与实现

## 3.1 引言

在上一章中，数字人拥有了自定义人的外形，也有了一定的表情表达能力。但是想进一步让数字人在“人的行为”方面有更好的表现，我们自然想到了其动作上的表现。数字人常常会以虚拟主播、虚拟偶像、品牌代言人的形象出现，这些形象不仅需要有特定的外貌，也需要大量处于对话或演说的场景之中。当演说时，其身体特别是手势的动作，也常常是富有特点，带有情感的动作。而这些动作的动画资源获取方式费时费力，只能找到千篇一律的手势动作，或者花费大量时间制作适配于这个情景的动作，又或者使用开销较大的动捕。具体来说，传统的手势动作生成方法通常是基于关键帧的方式，这样不仅需要大量的人工工作，还难以实现细腻的手势运动。又或者，采用动捕的方式需要付出高昂的人力成本和设备成本。

于是我们迫切需要有种方法，来让数字人能自适应地生成各种场景下和不同对话中的手势动作。因此，本文使用了一种基于CaMN网络与BEAT数据集的多模态驱动手势动作生成方法，通过学习多模态数据集中的语义和情感信息，自动地生成流畅、自然的手势动作，从而提高数字人的表现力和交互性。

本研究的手势动作生成模块是基于多模态数据的，其中涉及到了语音、语言和手势等多个模态。为了获取充足的手势数据并提升生成效果，我们参考了论文《BEAT: A Large-Scale Semantic and Emotional Multi-Modal Dataset for Conversational Gestures Synthesis》及其开源代码，使用了这个论文提出的BEAT数据集作为网络输入来驱动生成我们数字人的手势动作最终实现了一个具有较高生成质量的多模态驱动手势动作生成模块。

在本章中我们会介绍网络总体架构，以及各个编码器与解码器结构，还有损失函数，来探究手势动作是如何从多模态数据中生成的。最后我们会展示在unity中应用在我们的数字人身上的效果。

## 3.2 基于CaMN网络与BEAT数据集的多模态驱动手势动作生成

CaMN网络是一种多模态级联网络，可以将多个输入模态的信息融合在一起，提高模型的表现力。BEAT数据集是一个大规模的语义和情感多模态数据集，包含了视频、音频、动作等多种信息。本文使用BEAT数据集中的动作，语音，情感标签等信息作为网络的输入，使用bvh格式作为输出，并在unity中应用于我们的数字人身上。

### 3.2.1 网络结构

本文的手势动作生成模块基于编码器-解码器结构，采用了级联网络，包含文本编码器，演讲者ID编码器，情感编码器，音频编码器，面部表情编码器这五个编码器，解码器包含身体解码器和手势解码器。

### 3.2.2 编码器

1、文本编码器

文本编码器的输入是来自于将演讲者的语音通过自动语音识别器(ASR)转为文本，并进行了人工的文本校对，最终形成的字符串列表。

第一步是将文本数据中的词转换为词嵌入集，具体实现是使用预训练的FastText模型，将每个单词映射为一个低维度的词向量表示。这一步的目的是减少特征维度，简化模型复杂度，提高训练效率。

第二步是对文本序列进行编码，具体实现是使用一个自定义的编码器，它由一个8层的TCN网络和跳过连接组成。通过 TCN 模块进行时间维度的卷积操作，为了提取当前帧的特征，TCN通过从前17帧和后17帧的信息中进行分析和提取。之后使用跳过连接将信息从较浅的层直接传递到较深的层，有助于缓解梯度消失的问题。网络中一共包含4个TemporalBlock块，每块一共有2个卷积层，组成8层TCN网络。此外，还包含了一些relu激活函数和 dropout 操作。公式如下：

公式3-中的代表文本编码器，或者代表的是i的前f或者后f帧的包含文本信息的向量。代表文本编码器最终输出的文本特征向量。

2、演讲者ID编码器

演讲者ID编码器的输入是一个one-hot向量，代表演讲者的ID。

由于演讲者ID在同一个视频的各个帧不会改变，因此仅使用当前帧的演讲者ID计算其潜在特征。

这个模型采用了嵌入层（Embedding）作为演讲者ID编码器，线性层进行线性变换以及LeakyReLU 作为激活函数。

其中，嵌入层将每个演讲者ID映射为一个向量，然后经过线性层和激活函数层进行处理。最终将其与其他输入数据进行拼接，参与后续的级联网络，以增强神经网络的表达能力。

这个编码器定义了一个能够将演讲者ID编码转换为固定长度向量表示的神经网络模型，为后续对话手势合成任务提供了演讲者ID的上下文信息，从而可以更好地反映对话中不同演讲者之间的个性以及情感和语境。

公式3-中的代表演讲者ID编码器，是代表演讲者ID信息的one-hot向量。代表演讲者ID编码器最终输出的演讲者ID特征向量。

3、情感编码器

情感编码器的输入是一个情感标签的one-hot编码向量。

由四层TCN组成，每层TCN的基本结构都由卷积层、批量归一化层 (Batch Normalization, BN)和LeakyReLU作为的激活函数组成。

每一层TCN的BN层会进行归一化操作，LeakyReLU作为激活函数进行非线性变换，卷积层输出情感特征向量。这个过程可以提取出输入情感的高级特征，提取情感随着时间的变化。

公式3-中的代表情感编码器，是第i帧的前f或者后f帧的包含代表情感信息的向量。代表第i帧情感编码器最终输出的情感特征向量。

4、音频编码器

音频文件首先会被预处理程序下采样为16KHZ，帧速率则为15FPS。

音频编码器不仅接收预处理后的音频特征，还会将文本特征，演讲者ID特征和情感特征与这个音频特征拼接并处理，形成级联网络结构，最终得到更合适的音频特征。

音频编码器包含了一系列卷积层，第一个卷积层使用了较大的卷积核和大的膨胀率（即卷积核中像素之间的距离）来捕捉长期的时间依赖关系，以提取较宽的频带信息。接下来的卷积层逐渐缩小卷积核和膨胀率，以提取更局部的特征。最终，经过一系列的卷积操作，将音频数据转换为时域上的特征表示。总共是由12层TCN和2层MLP组成。

公式3-中的代音频编码器，是第i帧的前f或者后f帧的包含代表音频信息的向量。代表第i帧音频编码器最终输出的音频特征向量。、、 是前面提到的文本、情感、演讲者ID编码器的特征向量，在音频编码器中将他们与音频信息拼接一起来提取音频特征。

5、面部表情编码器

面部表情的表示为Iphone12Pro标准的52 维面部 blendsshape 权重。

面部表情编码器与音频编码器类似，会串联所有之前的特征，已提取更合适的面部表情特征。

面部表情编码器是一个包含四个基本块（BasicBlock）的序列模型。这些基本块具有不同的输入维度、输出维度、卷积核大小、步幅、扩张率等参数。在实现中，第一个基本块具有更大的扩张率和下采样，而其他三个基本块具有较小的扩张率和下采样。这些基本块包含两个卷积层和一个可选的下采样层，这些层通过残差连接连接在一起，以使网络能够更好地学习特征。面部表情编码器总共包含8层TCN，以及2层MLP。

最后，以上所有特征会拼接起来与预处理的种子姿势序列进行拼接，再通过LSTM模型进行处理，得到最终的输出姿势序列。

公式3-中的代面部表情编码器，是第i帧的前f或者后f帧的包含代表面部表情信息的向量。代表第i帧面部表情编码器最终输出的面部表情特征向量。、、、是前面提到的其他各个编码器的特征向量，在面部表情编码器中将他们与面部信息拼接一起来提取面部表情特征。

### 3.2.3 解码器

解码器分为身体解码器和手势解码器，而这二者是分开的，这是因为需要身体动作的输入来进行更好的生成手势。

其中音频、演讲者ID、情感和文本特征会进行融合，融合后的特征会输入到音频和面部表情编码器中进行特征提取，提取出的特征会再次进行融合，最后融合后的特征和前一时刻的姿势特征一起输入到LSTM中进行训练，得到手势动作的预测结果。即将五种模态的特征和之前的手势动作（即种子姿势）结合起来，来合成潜在手势动作。

身体解码器和手势编码器都是使用LSTM结构进行潜在特征提取，并使用2层MLP进行动作重建。

身体的特征向量公式如下：

手势的特征向量公式如下：

公式3-表示的是提取完特征之后，使用MLP进行手势和动作的重建。

### 3.2.4 损失函数

我们的损失函数主要有2个，分别是手势动作重建损失函数以及对抗损失函数。

1、手势动作重建损失函数

(公式3-4)

(公式3-5)

(公式3-6)

公式3-4是生成的身体动作与源身体动作在空间上的距离，公式3-5是生成的手势动作与源手势动作在空间上的距离，公式3-6是总的手势动作重建损失函数，决定手势损失的权重。

我们使用L1损失来衡量生成的手势动作与源手势动作之间的空间差异程度，以此来鼓励网络生成与源手势动作尽可能相似的手势动作。

2、对抗损失函数

(公式3-4)

公式3-4中的D表示生成对抗模型中的判别器，此公式表示手势和动作的对抗损失函数。

该对抗损失是通过训练一个判别器模型，来鼓励生成的手势动作越来越难以被判别器区分出来，从而提高前面的级联网络构成的生成器模型的生成能力和生成动作的真实感。

3、总损失函数

(公式3-4)

与分别代表二损失函数的预定义权重。

其中也能一定程度决定手势动作重建的权重，其值越小，表示语义相关性可能越低，与情感的关联会更小，同时对抗损失更大，更不容易被察觉是生成动作而非真人动作。

这两种损失函数的作用相互协作，共同用于优化网络的训练过程，提高生成手势动作的质量和逼真度。

## 3.3 实验结果与分析

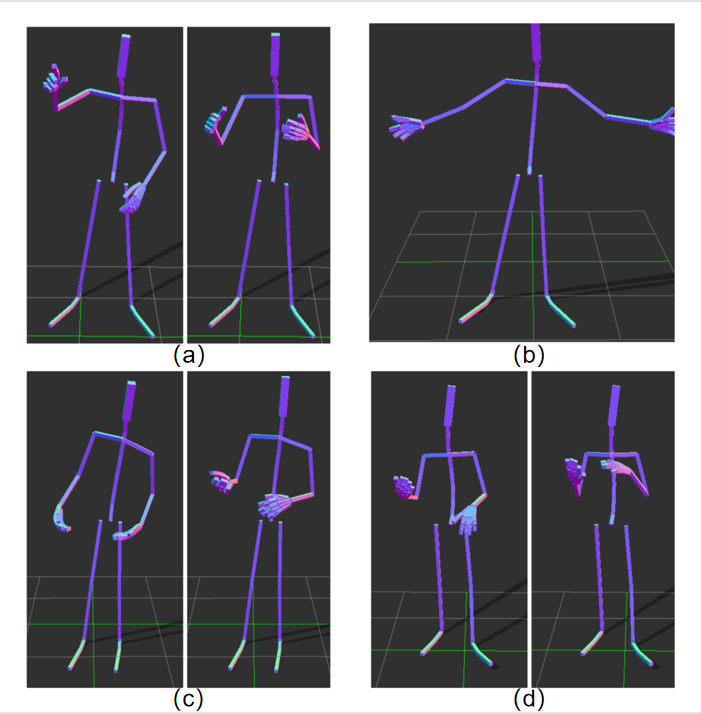
我们在unity中应用了这个模块，点击播放对应测试用例的同时，unity会开始播放对应的测试音频，且数字人会自然地做出相应手势动作。这个手势动作会有对应的情感和节奏变化。

同一段语音，演讲者不同，生成的手势动作也会有不同的风格。

经过调研，大部分人已经无法分别这里的手势动作是由神经网络生成的，而认为是由动捕或者什么别的方式获取的手势动作，可见此时生成出来的手势和动作已经非常自然和真实。

此处我们用到的测试输入数据为BEAT数据集提供的部分数据，包括文本信息，情感标签信息、演讲者信息、音频信息和面部表情信息。

输出数据为我们使用CaMN神经网络自动生成的手势和动作的动画文件，格式为bvh，我们放入bvh解析器中读取其动作。图3- 表示的是输入数据中演讲者ID不同，输出生成不同动作的截图。图中共有a、b、c、d分别是输入4个不同的演讲者ID，输出生成的不同动作风格。可以看到，不同演讲者有着各不相同的演讲风格，比如a、b这2位演讲者为男性，生成站姿就与下面了b、c演讲者不同，同时b演讲者生成的动作幅度较a、c、d演讲者较大，d演讲者重心放在右腿，而a、b、c演讲者重心更多放在左腿。



**图3- 输入不同演讲者ID生成不同风格的个性化手势动作**

当然，我们还需要将多模态驱动手势动作应用到我们的数字人身上，以下是在unity中的效果，具体应用于数字人模型后出现的问题与解决方案在第四章中会提到。



**图3- 应用于unity数字人的多模态手势动作驱动生成**

# 第四章 动作模块的设计与实现

## 4.1 引言

当我们将多模态数据中生成的手势动作应用到数字人身上时，需要考虑数字人执行动作时所需的各种细节，因此，我们需要设计一种有效的动作模块来实现这些细节。

在动作模块中，我们通常会使用逆向运动学（IK）和蒙皮技术来模拟数字人的动作。逆向运动学是一种能够根据目标末端效应器的位置计算出关节角度的技术，它可以帮助我们模拟数字人的运动。蒙皮技术是一种将数字人的表面网格与骨架进行绑定的技术，它可以让数字人的外观与骨架相匹配，并且可以让数字人的运动更加自然。

动作模块是数字人建模中非常重要的一个部分，它可以帮助我们实现数字人的运动行为，并且可以让数字人看起来更加自然和真实。

## 4.2 数字人动作模块设计

### 4.2.1 基于LBS的动画系统设计

本节介绍基于线性蒙皮(LBS)的动画系统。该系统能够处理数字人的动作数据，通过权重与骨骼相结合，来实现数字人的动画效果。

1. 线性蒙皮

我们在第一章的第二小节中已经提到过LBS的原理及其公式等，LBS不仅仅是基于骨骼驱动的人体建模的原理，同时也是动作系统的原理。LBS会根据骨骼变换矩阵，得到人体皮肤等的顶点变换，实现相应动画效果。

2、动画骨架

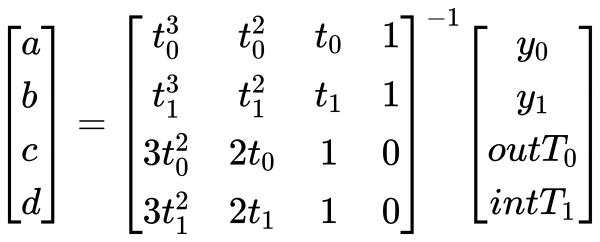
在本系统中，使用LBS来进行动画蒙皮，同时我们尝试使用两套骨骼进行动画骨架控制，一套是unity统一的humanoid骨架，一套是为数字人定制的专门骨架。通过将每个骨骼与相应的权重结合，我们可以计算出每个顶点受到骨骼影响的程度。这些权重以及骨骼的位置将用于计算动画蒙皮的结果。

3、动画插值

为了实现更加流畅的动画效果，我们使用了动画插值技术。在运行时，动画系统会根据用户的输入动态地混合多个动画。这种方法能够实现更加自然的过渡，从而提高动画的质量和真实感。

unity动画的关键帧插值采用三次多项式插值，在unity中，可调整的参数我们设为inT(inTangent，左侧的斜率)，outT(outTangent，右侧的斜率)和时间t。同时也对应公式4-的参数，y是t时刻动画的插值结果。

代入、时刻的y及其导数，我们可以得到a、b、c、d的矩阵表示，见公式4-。



实验结果表明，unity的动画系统在数字人动画的制作中具有良好的表现和应用前景。

### 4.2.2 基于IK的角色动画校正

1、反向动力学

反向动力学（IK，Inverse Kinematics）的原理是根据角色的目标点位置来自动计算每个关节的旋转和位移，从而实现更加精细、自然的动作效果。即是子节点带动父节点的运动，相比较之下的正向动力学就是从父节点到子节点的运动。

基于IK的角色动画校正可以帮助我们更加方便地对角色的动作进行控制和校正。

对于关节树来说，当已知末端的坐标时，需要逆向求解每个关节的坐标，这就是逆向动力学的核心思想。

常见的逆向动力学解决方案包括CCD（循环坐标下降）和CAA（圆形对齐算法）等。

在Unity中，IK可以设置的部位包括头部、左右手、左右脚、身体、左右手肘和左右膝盖这10个关节。

例如，在一个人体动画中，我们可以利用脚部的IK来控制角色的脚部位置，从而实现更加真实、自然的效果。

## 4.3 数字人动作模块实现与优化

### 4.3.1 数字人动作模块实现

1、骨骼系统

我们首先遇到的问题就是使用的模型骨骼与传统的软件和unity引擎中的骨骼系统有所不同，因此并不能让我们的数字人模型骨骼无缝衔接Unity的骨骼系统。

我们使用了两种可行的解决方案：

（1）全部使用和模型骨骼一致的动画，此时动画资产和角色模型资产的Rig全部设置为Generic，并且动画和人物的骨骼需要保持一致。

（2）使用其他常见动画素材库中的骨骼动画资源（如Mixamo），此时Mixamo下载的动画资产和我们的数字人和动画资产全部都要适配于Unity的骨骼系统（也就是把Rig调整为Humanoid），也就是需要进行重定向的操作。

2、动画资产

（1）自己手动制作关键帧动画。

（2）使用网上动画资源。

（3）根据需求自动生成动作。即我们上一章实现的多模态驱动手势动作生成。

### 4.3.1 数字人动作模块优化

1、动画资产错误解决方案

大部分动画资产应用在我们的数字人身上，都会有一些抖动，突变和骨骼不适配问题，此时都可以通过对动画资产进行精修的方式来适当地解决该问题，解决方案有很多，比较常见的是修改序列帧动画，最好是在Blender或Maya这样的专门建模软件当中进行，当然也可以直接在unity中做调整。

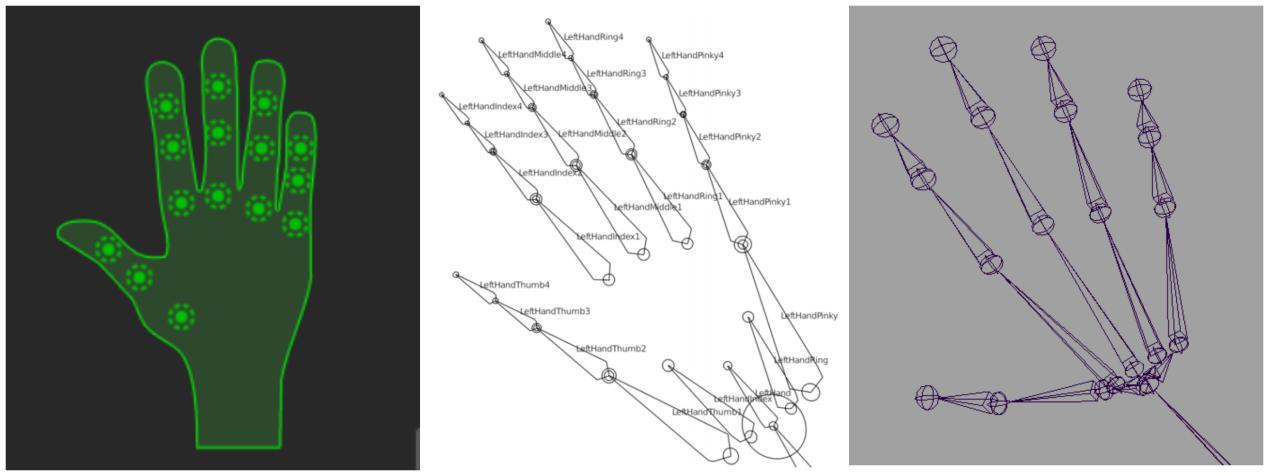
2、穿模问题解决方案

IK校正动画: 此时我们可以使用IK来校正动画，我们上面提到了基于IK的角色动画校正的原理，当应用于unity中时，需要注意的是，我们应用到unity中时，动画类型必须是Humanoid，且在动画Layer的设置里面，需要勾选“IK Pass”选项，然后在C#代码中设置反向动力学目标的转换权重，让对应关节更靠近IK Goal，可以调整关节位置的不自然，以及一部分穿模问题。

限制角度：在unity中，肌肉(Muscles)可以控制骨骼的运动范围，我们对数字人Avater的肌肉选项卡进行调整，比如将其旋转角限制的更小，防止穿模，同时，也能让角色动作更加自然。

3、部分骨骼不对应问题

在将多模态驱动生成手势动作的动画结果应用到数字人身上时，我们使用的是将我们的数字人和生成结果动画的Rig都调整为统一的Humanoid的方案，即进行骨骼重定向以适配unity的骨骼系统。在骨骼重定向过程中，会出现很多骨骼不对应的现象，这是由于骨骼不匹配的原因，而我们需要将这些不匹配的骨骼进行手动的大量的比对，以寻找最合适的相对应的骨骼。



**图3- unity humanoid、多模态驱动生成的动画、数字人三者的左手骨骼对比**

从图3- 我们可以看到unity humanoid、多模态驱动生成的动画、数字人三者的左手骨骼有很大区别，比如多模态驱动生成的动画的每根手指关节有4个，但是其他两个骨骼只有3个。多模态驱动生成的动画骨骼中手指有四个关节是为了在动作捕捉时更准确地捕捉手部姿势，而数字人骨骼和unity humanoid骨骼是为了模拟真实世界中，符合人类手指只有三个关节的事实。且多模态驱动生成的动画骨骼例如左手食指和拇指会有一个额外的父骨骼，而其他两套骨骼都没有。

最终产生的骨骼不匹配效果我们可以在unity中进行手动重定向，完成较合理的映射。

## 4.3 实验结果与分析

### 4.2.2 多模态驱动生成的手势动作的unity应用



**图3- 多模态驱动生成的手势动作的unity应用**

右边的UI上有多个测试用例按键，当然实际上可以有更多，这里只做展示测试。用户点击不同的测试用例会驱动不同的多模态数据去生成手势动作，并最终展示在我们的数字人身上，同时我们让unity同时播放多模态数据中的语音，就可以直观看到不同语音对应不同的手势动作。

### 4.2.3 数字人动作模块优化效果分析

1、IK矫正

从图3- 可以看到，同一帧动画中，IK矫正前的动画应用于我们数字人穿模严重，IK矫正后，纠正了穿模问题，且动作也很自然。



**图3- 同一帧IK矫正前后对比**

# 第五章 数字人系统综合与测试

## 5.1 引言

## 5.2 数字人综合系统的设计与实现

### 5.2.1 人体建模与手势驱动动作结合

由于我们的数字人系统本身就是应该集合各个模块，并且各个模块之间可以兼容和共存，因此我们也实现了动作模块和人体建模模块的结合。例如，数字人在多模态驱动生成的动作下，可以随时去调整数字人的人脸建模，表情建模体型建模，让数字人更具有个性化。当然更合理的做法是用户先用我们的系统自定义一个角色，即按照自己喜好捏人，让数字人拥有特殊的体型，相貌，然后这个数字人就可以通过多模态驱动手势动作生成，实现有独特个性特征和相貌的数字人，在不同场景下，通过屏幕与人自然交流，不管是对话，新闻播报，还是游戏中的聊天都可以有广泛应用。

### 5.2.2 其他部分

我们的数字人工程由于是一个完整的工程，因此也不仅仅只有人体建模模块，多模态驱动手势模块和动作模块，还有代表人物完整性的服装模块和头发模块等，本文我们并不展开细说这些模块，而是说一说在人体建模模块，多模态驱动手势模块和动作模块下，为了数字人更合理的表现，这些其他模块出现的问题和解决方案。

1、服装适配问题

服装方面我们采用了骨骼点替换技术，将服装的模型骨骼点替换为我们数字人模型的骨骼点，这样能够使模型运动时，服装也能贴着人体运动。当需要换装时，我们只需在对应文件夹下放入服装资源，即使骨骼并不完全适配也并不需要手动的给服装制作新的骨骼，而是直接自动识别人体骨骼并贴合人体运动。且在做例如骨骼驱动人体建模任务的时候，也不会穿模，而是随骨骼适配人体。

2、头发计算量问题

为了数字人的美观，我们其有更美观的头发。我们尝试过面片式与发丝式的头发，发丝式即头发的每一根发丝都被建模出来，视觉效果上确实更真实，也更方便物理模拟，但是计算量过于庞大。为了追求数字人的实时性，我们就使用了手机游戏端常见的面片式头发。 面片式即头发由多个面片组成，面片的材质包含透明通道，部分材质贴图是绘制了发丝。我们采用面片式，在unity中定义适配头发渲染的材质。此时出现的问题是会出现错误的半透明效果，解决方案是将深度写入打开来做正常的透明度混合。

## 5.2 实验结果与分析

# 第六章 总结与展望

## 6.1总结

首先，论文介绍基于Blendshape和基于骨骼驱动的人体建模系统的设计与实现，分别介绍了两者的原理，并对它们进行优劣对比。然后基于这些原理，介绍了我们在unity中实现的人脸、表情和体型建模功能。

其次，本文还介绍了数字人的基于多模态驱动手势动作生成模块的设计与实现。我们参考相应论文进行实现。我们介绍了CaMN网络的整体架构，包括各个编码器和解码器的结构，以及损失函数的设置。通过这些内容，我们将探讨如何从多模态数据中生成手势动作。最后，我们展示了该网络在Unity中应用于数字人所取得的效果。

我们在将多模态驱动生成的手势动作生应用到我们的数字人身上时，还要经过许多步骤，因此我们也引出了动作模块，介绍了数字人要实现动作模块所需要学习的原理，以及在unity中的应用方案。

最后，我们将各个模块整合在一起，让我们的数字人能够在拥有特殊的自定义外观的同时，可以个性化地驱动手势和动作生成，生成有独特外观，独一无二手势和动作的行为的数字人。

## 6.2 工作展望

尽管已经实现了上述功能，我们的数字人系统还存在很多不足之处，接下来将继续对数字人系统进行进一步的改进和优化：

1、多模态数据集获取和处理起来很麻烦，我们希望后续能通过优化网络，使得能从更少的信息中提取出更多有效的结果。

2、多模态驱动手势动作应用在我们的数字人模型上时，出现了因数字人和动画文件的骨骼不匹配带来动作并没有那么自然的问题，因此我们可以更加细致地使用重定向，确保数字人骨骼与驱动生成的手势动画骨骼足够接近。如果重定向不能解决，可以对我们的数字人手动添加或删除骨骼与重新蒙皮，使其与生成的动画文件骨骼相匹配。

3、代码的优化。

# 参考文献

[1] 杨福生，高上凯. 生物医学信息处理[M]. 北京：高等教育出版社，1988.

[2] 陈川波. 基于半结构化文本信息抽取的简历识别系统[D]. 北京邮电大学，2008.

（宋体五号）

[3] 张德政，张萍萍. 非结构化信息管理[J]. 微计算机信息，2006，22（3）：218-239.

[4］谢希德. 创造学习的新思路[N].人民日报，1998－12－25（10）

[5] 龙狼. Oracle数据库9i DataGuard的安装与维护. <http://www.soidc.net/articles/1213781343264/20080610/1215945423692_1.html>.

[6] <http://support2.dundas.com/Default.aspx?section=138>.

[7] Ian Foster, Carl Kesselman. The Grid: Blueprint for a New Computing Infrastructure[M]. Morgan Kaufmann Publishers, 1998.

[8] Ian Foster, C.Kesselman, J.Nick, et al. Grid Services for Distributed Systems Integration[J]. IEEE Computer, 2002, 35(6):135-160.

# 附 录

（黑体小三号）

……………………