**清华大学硕士学位论文**

**基于动态难度调节（自主调节）的视唱游戏设计研究与实现**

院 系： 深圳国际研究生院

专 业： 电子信息（互动媒体设计与技术）

姓 名： 张嘉睿

学 号：2020214447

导 师： 倪士光

2023年 3 月 27 日

目 录

第1章 绪论 3

1.1 研究背景 3

1.2 研究内容……………………………………………………………………4

1.3 研究意义 5

第2章 相关研究 6

2.1 视唱概述 8

2.2 ICT视唱练耳的应用现状 8

2.3 现阶段ICT实现方案的缺陷 10

2.4 DDA研究现状 11

2.5 OLMs研究现状 16

2.6 曲目难度评估方法研究现状 16

第3章 基于视唱曲目生成难度评价 6

3.1 引言 8

3.2 曲目选择 8

3.3 探索视唱曲目难度评级方法 10

第4章 视唱游戏的动态难度构建 6

4.1 引言 8

4.2 游戏流程设计 8

4.3 水平分析模块 10

4.3.1 评分标准化 10

4.3.2 场内分析 10

4.3.3 场外分析 10

4.4 关卡匹配模块 11

4.4.1 搜索筛选 10

4.4.2 权重剔除 10

第5章 视唱游戏的交互设计与功能实现 6

5.1 引言 8

5.2游戏可视化设计 8

5.2.1 音高实时匹配可视化 10

5.2.2 反馈音效与特效 10

5.3关卡生成 8

5.4曲谱渲染 8

5.3 演唱音高提取 10

5.3.1 AMDF算法实现 10

5.3.2 演唱音高平滑处理 10

5.4曲目-演唱音高数据同步 11

5.5 曲目-演唱音高数据匹配度计算 16

5.6 演唱表现评价 16

第6章 用户实验评估 18

6.1 引言 18

6.2 用户实验一：评估动态难度视唱游戏的效率与效果 18

6.3 用户实验二：动态难度视唱游戏的可用性测试 19

6.4 本章小节 19

第7章 总结与展望 18

7.1 研究总结 18

7.2 研究局限与展望 18

参考文献 27

# 第1章 绪论

## 研究背景和意义

### 研究背景

音乐教育是落实“素质教育”的重要载体。十九大明确指出，要重视音乐教育的发展和审美能力的培养。音乐不仅是提高学生文化素质水平、审美水平的重要手段，也对儿童的心理和智力发展有积极的影响，据[1][2],音乐教育对认知功能的提升有规划、工作记忆、语言智商和抑制功能。而视唱练耳旨在培养学生对音高、音长、音色的感知能力。学生能够注意和区分简单的节奏和旋律模式，以及不同节奏之间的关系。通过系统地接触音乐，能够培养对美的感知、对优质音乐的接受和评价，为良好音乐品位的培养奠定基础。

由于大众传媒的发达，人们现在可以以相当低的成本获取音乐甚至创造音乐，MIDI、录音、声音处理的成本越来越低。另外，日常环境中，音乐充斥着几乎每一个角落，人们不可控制地在商场等环境中接收音乐，且在私人场景中使用，以创造特定的情绪状态，或缓和兴奋水平[3]。

而视唱训练在音乐学习中具有举足轻重的地位，是基础音乐理论教育中理论知识和实践技能最重要的技能之一。在中国，从义务教育到专业音乐学院，再到综合性大学音乐系，都有与视唱相关的课程。利用声音模仿声音的能力是人类早期出现的一种行为，也是语言习得中的一种关键行为(Kuhl，2010)。模仿声音的能力，特别是音高，是音乐制作(即唱歌)的重要组成部分。歌唱是一种无处不在的活动，服务于社会、文化和教育功能。

音乐教育面临的困难主要来源于几个不同的方面。首先是音乐课程设计本身的问题。音乐课程通常有非常明确和统一的书面目标，很少考虑个人的音乐经验；课程计划过于死板，几乎没有灵活的调整空间；对学生没有吸引力。

其次，师资问题。视唱练耳教学对教师的专业能力和教学能力都提出了较高的要求。但在我国，应试教育的风行导致了音乐教育的边缘化，师资质量难以得到保证。而在师资匮乏的乡村地区，推进音乐教育也是促进教育公平的重要部分。

2020年，在全球范围内COVID-19大流行的背景下，教育部发文表明，实施在线教育是推进教育公平的战略选择。而如何破解优质数字教育资源匮乏、教师信息技术应用能力不足的问题，成为本文研究的出发点。因此，我们提出该课题：基于动态难度调整的视唱练耳游戏研究与实现。研究聚焦如何推进基于DDA的个性化的游戏进程，有效提升学生的学习动机和学习效果。该应用部署于便携设备，支持学生的独立学习并提供及时反馈，以提供低成本、高质量的音乐学习体验。

### 研究意义

个性化教学是提高教学质量、培养创新型人才的关键，但在传统的教学环境下，由于教学资源和教师实力的限制，很难实现每个学生一对一的个性化教学[5]。随着网络教学的兴起，传统教学方法在时间和空间上的局限性得到了解决，基于网络的教学平台构建了虚拟化的教学情景和教学资源，使教学行为不再局限于课堂和课堂。学生可以通过无限时间和地点的在线学习平台自由学习，平台为私人教师等学习者提供各种服务，提供一对一的指导和帮助[4]。

使用声音模仿音调的能力，根据个体差异有所不同，相当一部分人口从来没有获得过该技能，不能匹配半音偏差之内的目标音调。因此，需要学习平台需要与学习者积极互动，了解他们的个体差异，通过模仿教师的教学风格为不同的学习者提供有针对性的教学服务，根据不同学习者的需要为他们提供个性化的学习路径和学习建议，通过行为分析和学习过程中的模拟来了解学习者的学习习惯和偏好，动态地引导学习者完成知识的掌握[5]。

综上所述，为音乐教育的游戏化设计做创新探索，尝试音乐教育在游戏理论与思想方向上的突破，找到学习和娱乐的结合点，并在此基础上挖掘学生的学习兴趣和提升学习效果，成为本研究的学术宗旨。

## 研究目标

在本研究中，我们主要围绕着提升视唱练耳的学习效果的目标，通过游戏化赋能音乐教育。我们将游戏领域常用于提升玩家心流的动态难度调节迁移到视唱练耳训练中，并通过游戏化的方式，增强用户的使用意愿。

## 论文主要工作

第一章：绪论。本章主要介绍了研究的选题背景和研究目标。

第二章：相关研究。本章主要对与本研究相关的选题背景展开阐述。

第三章：基于内容的音乐难度评价。我们首先对视唱曲目进行了筛选，并探索了一种评价曲目视唱难度的算法。这是因为动态难度的构建建立在对关卡的难度评估上。

第四章：视唱游戏的动态难度构建。这一部分具体介绍了如何实现一个能够匹配用户水平的可调节曲目推荐系统。

第五章：视唱游戏的交互设计与功能实现。这一部分详细介绍了如何使用Unity引擎开发可视化交互界面来提升游戏体验、以及后台的音高提取、数据对齐、匹配度计算，从而能快速地实现演唱表现评价。

第六章：用户实验评估。我们将用户分为三组：常规视唱软件应用组、自选曲目的视唱练耳用户组、动态难度调节的视唱练耳用户组。我们对这三组用户的完成率、视唱的水平变化进行评估。

第七章：总结与展望。

# 第二章 相关研究

## 2.1视唱训练的方法和目的

视唱是无法凭借学习者独立完成的，需要在练习的过程中得到反馈。

在老师的帮助下，视唱使学生能够识别和理解符号和音乐符号的意义，并将它们与它们所代表的声音和节奏联系起来[6]。传统的视唱教学方法通常是由教师演奏一种乐器，然后监督学生重复、模拟或通过手指敲击练习节奏[7]。传统教学一般以纸质手册为基础，学生据此学习理论，并根据课后练习开展练习。随着信息时代的来临，视唱练耳的视频首先弥补了纸质手册在感官体验上的缺陷。

视唱是音乐教学中一门技术性很强的基础性学科。视唱教学的内容既包括基础音乐理论的学习，又包括理论知识技能的应用，因而具有综合性学科的特点视唱教学的内涵是通过对音乐节奏、音高、音调感觉和音乐阅读能力的训练，培养学生对音乐的感知、理解、欣赏和创造能力，使学生掌握正确的音高、节奏和表达音乐情感的能力，从而更容易地理解音乐形象，感知音乐的风格、类型和主题，为今后的音乐学习打下坚实的基础。这意味着视唱教学不仅要提高学生的视唱能力，还要培养他们的音乐能力。研究表明，视唱音乐能力的培养是通过提高学生的节奏感、旋律感、强度感、和声感来实现的。

## 2.2音乐能力个体差异

### 2.2.1 声调模仿缺陷(VPID)

大多数人在模仿歌唱时都能准确地模仿音高，但少数人（约10-20%）的音高平均超出目标音高±100分（目标音高两侧的1个半音），被称为“音高不准的歌唱者”。音高匹配的难点在于，该任务不能像手势模仿一样，依赖于对目标产生的运动姿势的直接观察。因此，本质上与其说是模仿，不如说是一种通过调整姿态再现结果的训练。

根据Greenspon[i15]的一项研究表明，一部分音高匹配能力较弱的个体，由于音高感知能力弱，听觉意象构建能力也较差，这些人中的相当一部分，平均音高偏离范围超过了半音范围。听觉意象与大脑中构建心理图像的多模态关联相关，这些关联是感知/行为关联的基础。Hutchins[i16]的研究表明，还有一部分人在声音运动控制方面有较大的缺陷，也会影响音高匹配能力。Pfordresher[i17]的研究表明，部分音高不准的歌手，听觉-声音系统的逆模型运转不佳。该模型的工作过程是，将基于感知的目标音高与喉部声带的运动控制相关联。

## 2.3 音乐游戏应用现状

随着ICT(Information and Communication Technology)的发展,ICT对音乐理论教育的重要性和作用被证实[5]。

### 2.3.1 视唱练耳应用程序的游戏化尝试

Serdaroglu[6]等人基于《Troubadour》提出了一种训练音乐听写的游戏化应用，

游戏化元素为徽章和排行榜，提升他们的投入，从而提升他们的学习效果。另外，游戏设置了练习的四个基本难度级别，难度与音乐学院的学期相对应。而在每组训练中，随着游戏的进行，根据学生的表现，听写的关卡会解锁，解锁的关卡会按照先后顺序，逐步提升音符序列的复杂度，音符是根据预先收集的音程分布自动生成的，复杂度主要取决于音符序列的长度、间隔的大小和间隔出现的频率。但文章没有对难度调节提供数据和学术支持。

苹果App Store中发现的《Earpeggio》[7]是商店中涉及内容最全面的听力训练应用之一，包括音程比较、音程识别、和弦识别、和弦转位、和弦进行等。应用对这些训练模块进行的分类，以供学生自行选择。但应用没有课程组件，也没有难度的渐进，学生需要教师的帮助才能进步。该应用提供了统计数据接口供教师查看，老师可以根据学生的能力调整进度。尽管如此，题目的设计有大量重复的内容，对题目的难度评价也模糊缺乏理论依据。

Fletcher[8]等人提出了一种用于增强心理模型的策略：提示衰退，即先在训练中提供色彩的提示，通过色彩刺激学习者，再逐渐使提示衰退，帮助学习者建立视觉上的心理模型。该方式能够有效地提升听觉训练结果。

App Store中的新曲视唱这款应用的视唱训练具有简单的UI反馈，用户哼唱的音高以音符的方式被记录在曲谱内。偏差较小的音符将呈现绿色，偏差较大的呈现红色。玩家需要自主同难度等级的曲目数量非常大且同质化严重的情况下，并没有赋予玩家选择的意义，交互性较差。

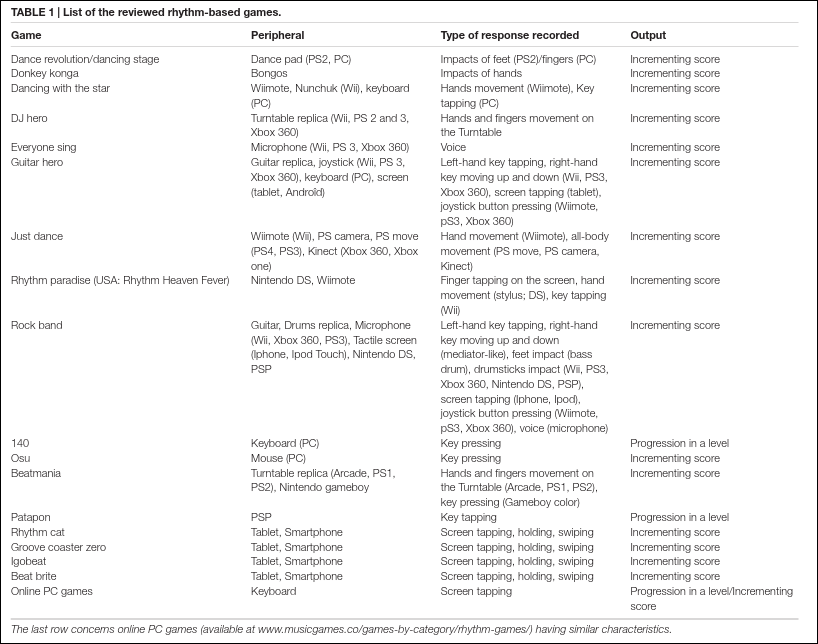
Ella采用了排行榜，视唱训练需要再通过前置任务的情况下，才能够解锁接下来的任务。具有简单的UI反馈，用户哼唱的音高以曲线的形式记录在曲谱上，并给予节奏和音高上的打分，但精确度不高。曲目没有对难度做设计。

视唱应用的难度设计普遍存在缺失或缺陷。唯一有课程编排的Troubadour，其音符序列生成的难度依据，无论是学术的还是统计数据的，都没有在文章中给出。早期有研究对不同音程实例的难度进行了排序，但仅是从宏观人群的正确率进行统计。然而，个体在音乐听力任务的子任务能力水平存在差异[9]，具体如音高、节奏、音色的感知，音乐工作记忆水平、对不同音符排列模式的熟悉程度，均存在不同程度的差异。单一的难度评价体系对个体的适用性较差。

游戏化方案依赖外部动机。现有的游戏化实现方式主要以排行榜、徽章体现。然而，赢得游戏的目标是一种外部动机，与用户实际接受的任务是分离的，而由于任务的视觉呈现方式和训练模式基本与纸质训练一致，只提供了基本的反馈功能，任务本身的内在动机几乎没有被加强。据研究[10]，与任务脱离联系的外部激励因素存在一些负面影响，比如用户无法专注于激励因素之外的内容。而让用户专注于任务的反思对学习来说是必要的，因此需要将正反馈与任务本身更紧密地结合。

忽视了学生在安排学习内容上的自主性。上述研究大多提供了教师监督学生和安排学习任务的功能，但没有一个应用提供了学生自我管理的方法。研究表明，在游戏任务执行时，自主选择任务的用户内在动机水平更高[10]。

### 2.3.2 音乐游戏研究概况

节奏技能是人类普遍具有的天赋之一，大多数人能够跟随音乐的节奏并随之律动。这些能力在认知方面具有重要意义，因为它们与语言和记忆等重要的运动和认知功能紧密相关。当节奏技能受到大脑损伤或神经发育障碍的挑战时，游戏作为一种交互式康复工具被广泛使用。

尽管上述一些游戏为训练节奏技能提供了良好的基础，但它们的主要缺点是在记录与节拍相关的运动时，时间精度相当差。因此，这些游戏提供的输出度量不足以分离性能的节奏特征（例如，运动性能的可变性，与节拍的同步精度等）。此外，在这些游戏中都没有操纵音乐刺激的节奏复杂性。难度仅通过游戏中所需响应的数量（例如，玩家必须对其做出反应的视觉标记数量）来操纵，这不是一个节奏特征。例如，使用具有不同程度节拍突出性的音乐可以在游戏中引入基于节奏的难度级别。这样做的优点是可以在整个游戏过程中逐渐呈现复杂度增加的节奏，从而可能导致跟踪技能的改善。

### 2.3.3 音乐游戏的PCG生成技术

节奏游戏的beatmap（所需动作及其时间）通常由人类设计师手工制作。Salsabilla[i19]提出了一种使用对象池管理绘制对象的方法，梁宇冰[i18]提出了一种“模糊标签”方法，采用C-BLSTM从著名的节奏游戏“OSU”的音乐文件中生成游戏内容。Destra[i20]利用机器学习方法和SVM进行节拍和旋律检测，步骤包括音符检测和音符放置。

## 2.4 DDA研究现状

动态难度调整是一种自适应地改变游戏属性以使游戏更容易或更难的技术，这种技术帮助开发者在游戏中实现良好平衡和优化玩家体验。为了保持玩家的专注度和参与度，游戏的挑战应取决于玩家的技能水平[19][20]。因此，游戏的难度必须得到控制，即对玩家来说不能太难也不能太简单。所有的DDA技术都有一个共同的步骤，即测量玩家在游戏中某一时刻面临的难度。

很多游戏都应用了这一系统。Valve公司在研发的《半条命》中植入了游戏的监控程序，收集玩家的表现，从而预估玩家未来的状态。并最终根据这些状态改变敌人的属性和行为。另外，Su Xue[18]等人研究了一种用于增强用户粘性的基于玩家状态转移有向图的DDA算法，Thai-Nghe N[19][20]等人提出了张量分解模型来建模学生的水平变化，他们将学生的水平建模为一个三维推荐系统问题（学生，任务，表现记录的时间），以预测他们在未来技能事件中的表现。Spronck P[21] 提出了一种规则库抽取的权重裁剪方法，选择规则的概率取决于分配给规则的权重。玩家的行为将会改变各种规则的权重比例。Mourato F[51]提出了一种基于遗传算法的DDA关卡自动生成，该方法先评估玩家表现，在多次种群迭代后生成玩家的最终代理，再基于这种表现输出关卡。

## 2.5 OLMs研究现状

许多现有的DDA教育游戏模型，往往基于学习者的表现，生成为每种结构特殊优化的推荐系统。这种系统自驱的适应性系统已经在视唱练耳游戏中实现。

人们正在积极研究开发引导学习者适应性特征的潜在模型，努力与学习者建立联系，提高系统透明度。由此研发的OLMs方法[23] ，通过记录学习者数据生成用户特征，并以此为凭据提出最佳学习路径[24],面向用户生成可调节的难度表盘，这些模型的使用已经导致学生在各个领域和系统中的学习得到改善[25]。

Lynnette[26]是一个使用 OLMs 来支持学习和自我调节的著名例子，该游戏为学习代数的应用，游戏中呈现了一个界面，显示学生在每个问题类型上的表现，同时也提供了关于问题选择的即时反馈; 学生在选择他们需要改进的问题类型时会收到正面反馈，而在选择他们已经掌握的问题类型时会收到负面反馈。Long和Aleven[27]的一项随机实验结果显示，与系统完全控制问题选择时相比，当学生被允许选择要做的问题类型并收到关于他们选择的反馈时，他们的学习效果更好。

## 2.6音乐内容处理

### 2.6.1曲目难度评估方法研究现状

在音乐教学中掌握乐谱的难易程度，不仅有助于学生根据自己的水平选择合适的乐谱作为学习材料，而且有助于提高教学评价的公正性。虽然没有明确的标准来分类乐谱，但是在乐谱网站、音乐等级考试和教科书的布局上有乐谱难度的分类[32]。例如，乐谱下载网站根据大致的音高,音阶,和声和节奏将分数分为10个分级[33]。

根据音阶、和声、演奏方法等因素，将英国音乐演奏评分系统分为 8 个等级，为评价音乐演奏和理解提供了一个评分系统。就美国NYSSMA而言，根据音阶和八度音阶、音阶、演奏速度等，分为6个等级标准，这些标准是根据音乐的难度主观地分类出来的。视唱和听评分系统(SSMA)的教学大纲也根据乐谱的音高和节奏的复杂程度，将教学内容从容易到难度进行分类[34]目前对钢琴乐谱难度等级识别的研究主要是基于现有教科书或权威机构给出的具有明确难度等级标签的乐谱，并利用线性回归支持向量机等方法建立难度识别模型，试图利用该模型预测大量未标记乐谱的难度等级，以实现难度等级对乐谱进行自动分类[35]。旋律复杂性已经被研究了很长时间，它已经被用作研究的主要指标，如唤醒艺术刺激的潜力和享乐价值(偏好)音乐流行因素的分析[36]。这些研究说明了旋律的复杂性作为一种诊断工具的潜力，以确定何时出现旋律偶然性，或者旋律的哪些方面(音高组合、层次结构、顺序等)对于分析旋律结构(相似度评级、旋律曲目等)更重要[37][38][39][40][41]。

Eugene Narmour首先提出旋律复杂性的概念，他认为内隐实现模型在描述原始民歌的旋律复杂性方面表现良好，但缺乏对听众音乐和文化背景的考虑。为此，Eerola等人提出了一种基于期望的复杂性预测模型(EBM) ，他们发现音调 (由节奏位置和音高持续时间修正)、距离方向和距离变化的区间原则、节奏原则切片和节奏变化在预测听众的复杂性判断方面具有显著的能力。除了特征定义之外，模型构造还有算法选择。Cilibrasi等人质疑音乐聚类算法的有效性，并使用标准化的信息距离作为广义相似性度量来构建模型。在最近的一项研究中，Eerola等人使用信息理论和期望违背作为构建旋律复杂性的模型，并在七个数据集上验证了它，在这些数据集中，旋律复杂性的总体静态评分已经被收集和执行得很好。音乐评分的传统步骤是首先提取测试者的歌唱音乐，从歌唱输入信号中提取旋律特征，然后将提取的旋律特征与模板旋律进行相似性比较，根据它们之间的差异来衡量音乐表现。

### 2.6.2 旋律特征分析

音乐特征提取是针对用户进行曲目推荐的依据之一，主要是通过解析音乐结构、模仿人类分析感知音乐的方式进行的。机器学习被广泛地应用于音乐的分类，主要分为两个阶段：1.将音频转换为一组代表性的特征。2.根据这些特征构建模型。

2.3.2.1基于神经网络的音频特征提取

Nikki[i2]研究了一种使用CNN神经网络的方法，将歌曲时间片段转换为频谱图象，并输出到不同的音乐流派。

2.3.2.2基于向量机的频谱特征提取

Matsui[i1]等人研究了一种新颖的音乐特征提取方法，使用SIFT算法提取梯度特征。这种特征可以有效地捕捉对数频率域中的局部动态信息。他们随机选择样本并将其表示为2D谱图像。然后，他们为每个图像提取SIFT关键点，使用具有线性核的SVM作为分类器，他们通过与基于GMM和MFCC特征向量的简单方法进行比较来检查其方法的时间和频率独立性。Sri[i3]提出了一种能够结合频域、时域、倒谱域和调制频域多种域的SVM分类器，能够有效提升自动分类与人工分类的一致性。

2.3.2.3使用决策树的分类方法

Hssina[i4]不断迭代树的结构，在每一次迭代中选择最具有区分性的节点，并保留与该路径一致的数据子集，通过这样的方法提高了决策树在分类音乐时的速度和表达。这种算法强依赖于信息度量，也就是熵。

2.3.2.4基于符号序列的分类方法

除了使用音频进行音乐信息提取，有些分类方法直接使用符号格式存储的音乐，常见的音乐符号格式有MIDI格式、MusicXML格式以及abc格式等，这些格式在理论上可以互相转换。这些文件格式主要记录了与标准化演奏的信息，如音符和休止的时延，以及演奏时的表情、力度。还会额外存储曲目的节奏型、调号、声部等。一般使用这些符号序列进行分类，需要对这些信息进行建模。例如Corrêa[i5]等人使用隐马尔可夫模型进行旋律分类与预测，Wang J C[i6]等人提出了一种将音乐序列与情感关联的模型。Niu N[i7]开发了一种基于门控递归单元网络和多特征提取的情感识别模型，在效能上超过了LSTM。

以上模型的应用范围具有较大的局限性，目前的识别方向主要集中于情感识别与歌曲推荐检索，且需要大量的数据库进行训练，否则可能出现对小概率样本的偏向问题。

### 2.6.2歌唱信息提取方法

#### 2.6.2.1音高提取方法

音高检测的目的是提取人声中的基音周期，并将其转换为单音音高序列。通常会利用语音信号短时平稳性这一特点，先将语音信号分帧，然后对每一帧进行处理。主要有两种基频提取方式：基于时域或者基于频域的方法。时域法的输入是声音的波形，它通过寻找波形的最小正周期来实现。而频域法则是先对信号进行傅里叶变换，得到幅度谱（忽略相位谱）。在基频的整数倍处，频谱上会出现尖峰。频域法的基本原理就是求出这些尖峰频率的最大公约数。

在大部分情况下，并不是所有的音频信号都含有基频，所以提取基频之前，需要先判别该帧是处于清音还是浊音。一般来说，使用逐帧的方式提取基音，最常出现的问题是半倍频错误，从表现上来看，得到的基频是真实情况的倍数或半数，这是由于检测到了其他泛音导致的。为了解决这个问题，通常会再对结果进行二次处理，可能是通过与符号序列进行匹配，或是加以特殊的平滑处理。

一般来说，基频提取分为三种：基于规则的信号处理方法、基于机器学习的方法、基于心理声学的方法。

时域基频提取方法分为三种，第一种是自相关函数（ACF）方法，通过将位移信号与原始信号相乘，然后在一帧内求和，得出能够表示该点信号周期位移的峰值，来找到波形的最小正周期。第二种是平均幅度差函数法[i12]，通过减去信号及其移位版本来衡量它们之间相似性，通过将原始信号与位移信号相减，以求得绝对值，之后再积分。第三种也是最常见的方法是YIN，这种方法建立在ACF的差函数思想基础上，定义一种累积均值归一化差函数。它通过计算差分函数来估计音高，并使用累积平均值来规范化差分函数（cumulative mean normalized difference function, CMNDF）。最后，使用阈值来确定可能的音高位置，并使用二次插值来提高音高估计的精度。

频域提取基频的方法主要是Camacho提出的SWIPE[i13]，核心思想是，设计一个显著度函数，表达每一个候选频率是基频的可能性。显著度函数取得最大值的频率f就被当作基频候选值。Thomas[i14]在SIFT算法基础上提出了一种能够在嘈杂环境下实现基频提取的方法，使用残差信号纯化谐波，能够显著地去除噪声和共振。

基于机器学习的方法往往从数据入手，比如 C Kim[i11]采用一种深度卷积神经网络（CREPE），可以直接在时域波形上运行，具有良好的抗噪性能。但该算法的训练数据并非人工标注获得，而是源自程序自动合成，它根据预先确定的音高、音量和音色来生成数据，以此来保证它们的标准答案100%准确。但这可能会导致CREPE在应用到更广泛的真实基频数据时，无法复现在合成数据中展现的良好效果。

#### 2.6.2.2 节奏提取方法

声音的包络通常包括起始、衰减、延音及释音四个部分。而歌曲的节奏检测源于歌声的起始检测。

超过半音距离的音高的变化代表出现了音符转换，这种思想带来了一类方法，通过观察基频轨迹的陡升和陡降检测起始节点。Behnam[i8]提出了一种通过pYIN算法获取基频的轨迹变化，并使用相邻帧之间的音高差值，找到音符衔接位置，识别唱名的起始结束位置的方法。

还有一种方法是使用合理的探测函数，检测音频信号的能量峰值。建模并优化探测函数可以提升该类方法的准确度和效率。常用的深度神经网络检测，例如林敏等[i9]提出了一种使用起始函数拾取基频中峰值的方法，使用激活值描述在该点起始出现的概率，并用峰值拾取算法来验证确切的起始项。关于提取音符相关时间信息的方法，Sara[i10]提出了一种人工参与的多参数筛选峰值的算法，通过不断地修正可以达到良好的识别效果。Holzapfel[i11]使用群延迟、幅度、音高评估乐器起音的三维方法。

以上获取音符起始和终止节点的方法，在应用层面上都尚未完全实现鲁棒性，此外，由于它们往往应用于音乐自动记谱等功能，这些功能脱离于既有的乐谱模板而存在，因此常常由于预设和约束的缺失，导致偶发的错误。在我们的应用场景中，节奏识别是有对照的模板存在的，因此更需要关注如何检测用户输出的节奏和乐谱信息的一致性，在一定程度上可以简化到搜索匹配音高的范畴。

# 第3章 视唱练习动态难度构建

## 3.1 引言

本章节主要介绍了本文针对研究目标1：如何通过选择合适的视唱练耳曲目，评估它们的难度，并设计合理的动态难度调节系统，为用户提供全面和自适应的视唱训练体验。我们希望通过这一研究，能够为视唱练耳教学提供更多的帮助和支持。

为了实现这一目标，我们设计了一个动态难度调节系统，它能够根据用户的水平和表现，动态调整游戏关卡的难度。这一系统的核心是一个评估模块，它能够根据用户在游戏中的表现，评估用户的水平，并根据评估结果，动态调整游戏关卡的难度。

通过选择合适的视唱练耳曲目，评估它们的难度，并设计合理的动态难度调节系统，为用户提供全面和自适应的视唱训练体验。选择视唱练耳的曲目对于确保视唱训练材料的教学意义至关重要。这些曲目必须精心挑选，涵盖广泛的节奏型、音乐主题、风格，以便用户能够获得全面而具有合理性的训练体验。此外，还必须评估这些曲目的难度，以便为用户设计合理的动态难度调节系统。该系统可以根据用户的水平动态调整游戏关卡的难度，提供个性化和自适应的训练体验。

在游戏动态难度的研究中，我们发现，评估关卡难度与用户水平匹配程度是动态难度调节的核心。视唱游戏的关卡难度由曲目的复杂度组成，可以通过各种音乐特征（如旋律、节奏、速度和熵）来衡量。用户水平的测量也需要一个科学合理的方法，可以通过评估用户在游戏中的表现（如节奏一致性、旋律一致性、声音稳定性）来实现。

## 3.2 视唱的曲目选择

本工作需要较多的曲目作为训练材料，主要的目的是尽可能满足不同水平用户的需求，因此需要选定一个全面的曲目数据集。

最后，在视唱曲目的选择范围方面，我们采用了现国内音乐院校常用的法国《视唱教程》系列和上海音乐学院编写的《单声部视唱》上下册，他们都是久经考验的典型教材。这两个系列由于使用时间长，范围广而为学生和教师熟知，权威性很高。尤其是法国《视唱教程》，它被《新格罗夫音乐与音乐家辞典》评价为在视唱练耳学科发展过程中最重要的视唱练耳教程之一。

我们选择了其中的一些曲目，它们覆盖了全面的节奏型，包括八分音符、前十六后八、前八后十六、十六分音符、切分音、前附点、后附点、三连音等。这样可以让用户尽可能在使用过程中，接触到各种不同类型的节拍，从而更好地掌握节奏。

在调式的选择上，我们考虑到过高的读谱难度会降低目标用户的使用意愿，因此我们的训练曲目全部采用C大调和a小调，即没有升降号。这样可以让学生更容易地读谱，并且能够更快地掌握基础知识。

## 3.3探索视唱的曲目难度评级方法

曲目的认知难度评级，由旋律复杂度、节奏复杂度、速度和熵四个维度组成。

# 第4章 视唱游戏的开放式学习模式探索

## 4.1 引言

## 4.2 游戏流程设计

## 4.3水平分析模块

### 4.3.1评分标准化

### 4.3.2场内分析

### 4.3.3场外分析

## 4.4关卡匹配模块

### 4.4.1搜索筛选

### 4.4.2权重剔除

# 第5章 视唱游戏的交互设计与功能实现

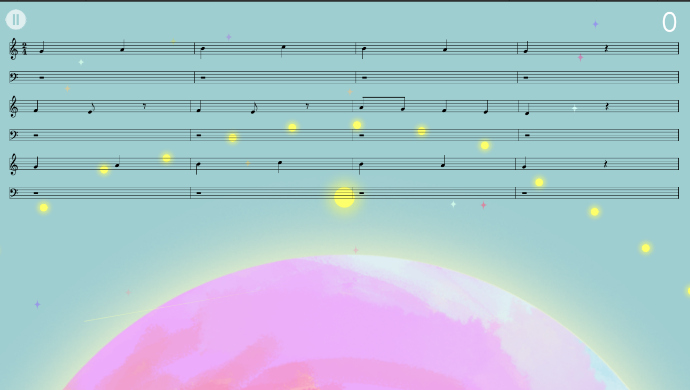
## 5.1游戏可视化设计

基于多感官学习理论，游戏中对用户输入的视觉反馈，可显著提升沉浸式学习的效果。而多样的可视化方法可能产生学习效果上的差异。

### 5.1.1 音高实时匹配可视化

音高匹配，是认知心理学中的一类研究对象，即人类通过发声模仿指定的声调，这是视唱任务的重要组成部分，具有复杂的心理认知过程。根据已编码的音高序列准确发出音调依赖于听觉-声音系统的抽象逆模型，该模型允许歌唱者根据目标感知结果来计划喉部的运动姿势。研究表明，在这一过程中，听觉意向能够激活大脑的一些听觉和运动区域，以及次级听觉皮层和辅助运动区。这些区域的活动表明，大脑在执行不同的听觉意向任务，其中就包括操纵旋律的听觉图像。

根据MMIA模型，喉部运动和特定声音结果之间的关系模型，是构建于此个体过去的知觉事件。个人的一生中，统计学习使个体能够采集音高和喉部运动的关联经验，从而扩展形成抽象的模式。而多模态图像的示意图映射，允许个体将这些关联推广到新的情况。具有更高音乐经验的声乐演唱者更可能使用音高想象进行声乐音高模仿，因此，音乐图像与音高的映射关系，对演唱水平有一定的影响，本研究将视唱训练游戏化的理论依据得到证实。

音高决定了玩家在空中的位置，而曲谱内的音高将在序列化后，转换为空中星辰的高度。随着球体的转动，曲目的音高信息不断地被转换为星辰序列，玩家在这个过程中，需要通过控制发声的音高，使自身尽可能准确地接触每一个星辰。

### 5.1.1 反馈特效与音效

## 5.2 关卡生成

## 5.3 曲谱渲染

## 5.4演唱音高提取

## 5.4.1AMDF算法实现

使用AMDF(短时平均幅度差函数)算法，可以有效地提取演唱者输入音频的基础频率，设语音信号的时间序列为,加窗分帧处理后得到的第帧语音信号为,其中下标表示第帧，设每帧帧长为。的短时平均幅度差函数定义为：

对于周期性的浊音语音，Di(k)呈现出了与浊音语音周期相一致的周期特性，利用这种特性，可以估计出浊音语音的基音周期。

此外，半倍频错误也是现有算法共有的问题。

## 5.4.2音高平滑处理

在基于AMDF算法的基本提取中，将音频信号分为帧，得到的基音序列也是基于帧数刷新的。假设每位学习者的目标是准确地唱出乐谱，那么存在的野点不会持续很长时间，而且它们之前和之后应该有一个平滑的音高序列。事实上，通过分析大量视唱分帧数据的音高序列中野点的持续时间，在两个平滑信号的中间，发现大量野点出现。这主要是由于大部分演唱者由于声带气流的颤动，往往会产生颤音，该技法也能提升演唱的表现力。因此，这种音高的合理偏离应尽量在数据处理中被校正，不应该按照原始数据来进行动态规整。

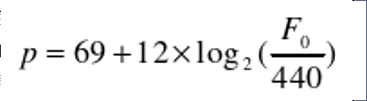
因此，我们可以将音调序列平滑如下。

(1) 将基音序列中具有相等和相近基音值的帧作为一个子序列，并计算帧的数量。

(2)对每个基音序列的帧进行迭代，找到帧数在1到2之间且其前后帧大于2的野点。

(3)将发现的野点的音高值设置为其前后序列音高的均值。

(4) 将中值平滑应用于优化的基音序列。

 通过此方法得到的频率可以直接用公式转换为MIDI音高，该度量是国际通用的音高衡量标准。

## 5.5 曲目-演唱音高数据同步

该步骤将游戏过程中的音节与曲目信息同步匹配。视唱技能评估的主要功能是通过比对两组数据，得出用户的旋律和节奏的表现评估。在本研究中，比对的输入数据分为三组：1)以演唱时间为自变量，所对应的曲谱内的音高；2)以演唱时间为自变量，所对应的用户演唱的音高。3) 以演唱时间为自变量，所对应用于节奏识别的音头。

首先是演唱时间窗口的获取。在Unity引擎中，有一个生命周期函数是按固定的帧率调用的，根据设置步长固定间隔的执行。例如：设置50帧，那么不管实际帧中一帧执行了多少秒，该生命周期函数固定每20ms调用一次。而每帧所对应的时间戳可以自由获取。我们的游戏帧率在50帧以上，远大于检测节奏准确性所需要的精度，这为我们记录均匀且足够细分的时间序列提供了基础。

获取曲谱内音高的第一步是序列化MusicXML文件，读取其中的音符、休止符节点的音高、附点、时值信息，按照先后顺序存储到对象池中，根据音符时值占总时值的比例，计算出每个音符或休止符的持续时间，从而获取音符的起始时间戳和终止时间戳。经过这样的文件处理，我们可以直接通过游戏当前时间获取到对应的音高信息。

获取演唱音高的第一步是获得当前时间的用户演唱音高，并使用滑动平均滤波法进行平滑处理。

## 5.6 曲目-演唱音高数据匹配度计算

DTW(Dynamic Time Warping)动态时间规整算法，是一种计算两个时间序列数据相似度的一种动态规划算法。

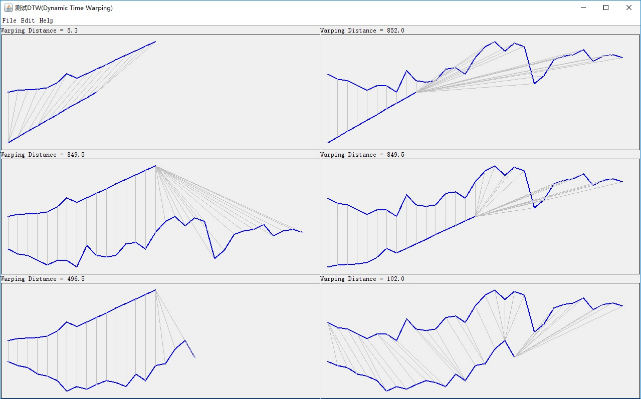
在实际任务中，用于计算相似性的两条数据在时间维度上是大致相似的，但很难得知每个时刻的对应关系。但此算法可以有效地解决整个难点。

图3.6 DTW算法示意图

上述DTW算法是针对所有时间序列的相似度计算而设计的，在本文的视唱评分中，可以结合数据的特殊性来改进基音序列的匹配与一般的语音信号匹配相比，视唱评分中基音序列的匹配过程具有以下两个特点。

(1) 视唱是指根据乐谱中的旋律唱出每个音符，乐谱中每个音符的持续时间远远长于基本检测中一帧的持续时间，通常呈乘法关系。这意味着从演唱音频中提取出来的音高序列由许多段组成，其音高值与序列相同。意味着当使用

DTW时, 基音序列相似度计算算法，可以通过子序列进行匹配，以指数形式减少算法的计算时间。 (2) 这里有三个方向用于选择原始 DTW 中的下一个路径点, 垂直选择表明序列 x 中的一个点可以对应于 y 中的多个连续点，反之亦然。在视唱评分中，假设视唱音频的音高序列为 x，模板的音高序列为 y，如果 y 中的多个连续点与 x 中的一个点对齐，则表示演唱者遗漏了一些音符，在这种情况下，y 中的这些点原来是错误的位置，在DTW 的路径选择中可以忽略，最后可以与相邻的点对应。即，在最小累积距离路径选择中，可以跳过 y 中的点，这种改进也可以缩短计算时间。 一个用户多次唱同一首歌，在个人音符和整体节奏上有所不同。因此，基音序列不能严格按时间匹配到模板，需要对基音序列进行时移、缩放等动态正则化操作。考虑到这一客观要求，本研究使用前面部分的DTW算法来实现模板匹配。

## 5.7 演唱表现评价

### 5.7.1 修正时值错误的延续性

目前的研究已经证明，DTW能够通过匹配乐谱MIDI音高序列和演唱音高序列之间的非线性关系，可以用于度量歌声和曲谱的相似程度。

但虽然度量的算法本身已经经过有效性的检验，但演唱者的某个音符的个别时值错误可能会影响后续的节奏匹配。虽然这一问题由于视觉反馈的及时性，已被弱化，但在多个连音重复出现的时候，依然会影响对音高偏差的反映准确率。因此，我们使用

### 5.7.2 音高和节奏准确度的主观感测度量

## 5.8 数据处理流程

图3.7 数据处理流程

由于本游戏区别于大部分商业游戏的高频交互和基于时间的线性关卡体验，更偏向静态任务，我们采取

先生成任务与任务评分，再预测玩家表现的方式。在生成任务阶段，我们将随机生成音高任务和节奏任务。我们将利用算法生成大量的音高任务和节奏任务。通过标准测试和专家评审得出被试者的音乐感知水平，将被试者的感知水平和在任务中的表现，输入RNN深度神经网络进行训练，得出任务在多个维度上的难度评分。利用玩家个体在其中部分任务中的技能水平，结合任务的难度，预测其在任务中的测试表现，从而计算其在不同任务中的准确率预测值。

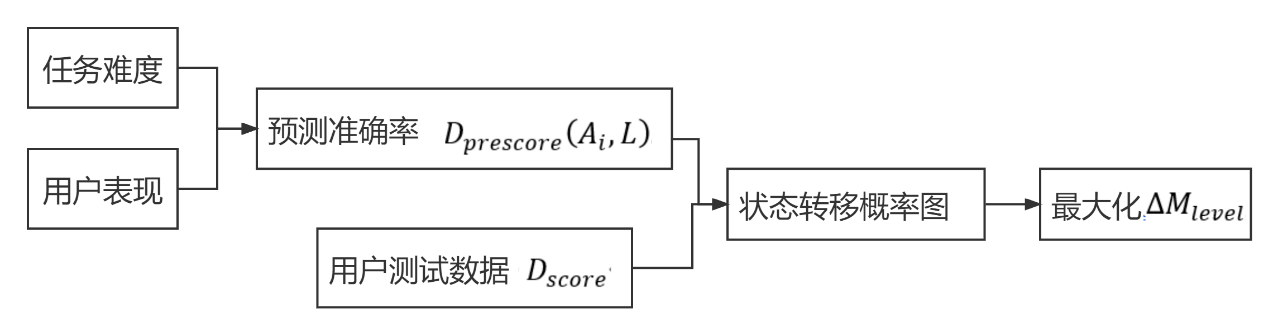
获取准确率预测值后，我们采用贝叶斯网络绘制玩家的状态转移概率图，通过动态规划方法求证最优的,以最大化玩家的水平提升值。依据MET音乐能力标准测试[49]结合玩家表现计算。

图3.8 求证最优水平提升值的数据流向图

3.9 自主调节难度的界面

上文提到，大部分的建模和适应过程都是难度调节系统在幕后进行的，学生们自己无法接触到。据进一步文献调查发现，跨不同类型系统的研究揭示了开放用户评估模型的好处，从而提高了系统透明度和用户参与度[50]。因此，在本研究中，我们希望尝试一种更加透明的适应性形式，除了提供游戏难度推荐之外，通过一个个性化的界面向学生显示相关的评估信息，并给予学习者自主调节的空间，如图3.9所示。

# 第7章 总结与展望

## 6.1本文工作总结

## 6.2未来工作展望

参考文献

1. Habibi A, Damasio A, Ilari B, et al. Music training and child development: a review of recent findings from a longitudinal study[J]. Annals of the New York Academy of Sciences, 2018, 1423(1): 73-81.
2. Jaschke A C, Honing H, Scherder E J A. Longitudinal analysis of music education on executive functions in primary school children[J]. Frontiers in neuroscience, 2018, 12: 103.
3. North A C, Hargreaves D J, O'Neill S A. The importance of music to adolescents[J]. British journal of educational psychology, 2000, 70(2): 255-272.
4. Carrillo C, Baguley M, Vilar M. The influence of professional identity on teaching practice: Experiences of four music educators[J]. International Journal of Music Education, 2015, 33(4): 451-462.
5. Eyles A M. Teachers' perspectives about implementing ICT in music education[J]. Australian Journal of Teacher Education, 2018, 43(5): 110-131.
6. Debevc M, Weiss J, Šorgo A, et al. Solfeggio learning and the influence of a mobile application based on visual, auditory and tactile modalities[J]. British Journal of Educational Technology, 2020, 51(1): 177-193.
7. Serdaroglu E. Ear Training Made Easy: Using IOS Based Applications to Assist Ear Training in Children[J]. European Journal of Medicine and Natural Sciences, 2018, 2(1): 61-68.
8. Hwang Y T, Chu C N. The Design of Music Ear Training System in Building Mental Model with Image Stimulus Fading Strategy[C]//International Conference on Learning and Collaboration Technologies. Springer, Cham, 2018: 127-135.
9. Kragness H E, Swaminathan S, Cirelli L K, et al. Individual differences in musical ability are stable over time in childhood[J]. Developmental Science, 2021, 24(4): e13081.
10. Ryan R M, Deci E L. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions[J]. Contemporary Educational Psychology, 2020, 61: 101860.
11. A. Stenberg and I. Cross, “White spaces, music notation and the facilitation of sight-reading,” Scientific Reports, vol. 9, no. 1, pp. 1–12, 2019.
12. E. Huovinen, “Early attraction in temporally controlled sight reading of music,” Journal of Eye Movement Research, vol. 11, no. 2, pp. 1–30, 2018.
13. J. C. Smith, “Hidden in plain sight: a music therapist and music educator in A public school district,” International Journal of Music Education, vol. 36, no. 2, pp.
14. H. Mitchell and B.. Roger, “'e moot audition: preparing music performers as expert listeners,” Research Studies in Music Education, vol. 39, no. 2, pp. 195–208, 2017.
15. Pathania A, Leiker A M, Euler M, et al. Challenge, motivation, and effort: Neural and behavioral correlates of self-control of difficulty during practice[J]. Biological psychology, 2019, 141: 52-63.
16. Tan C H, Tan K C, Tay A. Dynamic game difficulty scaling using adaptive behavior-based AI[J]. IEEE Transactions on Computational Intelligence and AI in Games, 2011, 3(4): 289-301.
17. Csikszentmihalyi M, Csikzentmihaly M. Flow: The psychology of optimal experience[M]. New York: Harper & Row, 1990.
18. Xue S, Wu M, Kolen J, et al. Dynamic difficulty adjustment for maximized engagement in digital games[C]//Proceedings of the 26th International Conference on World Wide Web Companion. 2017: 465-471.
19. Thai-Nghe N, Horváth T, Schmidt-Thieme L. Factorization models for forecasting student performance[C]//Educational Data Mining 2011. 2010.
20. Zook A E, Riedl M O. A temporal data-driven player model for dynamic difficulty adjustment[C]//Eighth Artificial Intelligence and Interactive Digital Entertainment Conference. 2012.
21. Spronck P, Sprinkhuizen-Kuyper I, Postma E. Difficulty scaling of game AI[C]//Proceedings of the 5th International Conference on Intelligent Games and Simulation (GAME-on 2004). 2004: 33-37.
22. Bull, S. (2020). There are open learner models about! IEEE Transactions on Learning Technologies., 13, 425– 448.
23. Dascalu, M.-I., Bodea, C.-N., Mihailescu, M. N., Tanase, E. A., & Ordoñez de Pablos, P. (2016). Educational recommender systems and their application in lifelong learning. Behaviour & Information Technology, 35(4), 290–297.
24. Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. IEEE Transactions on Learning Technologies, 10(4), 405–418.
25. Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. Handbook of Research on Learning and Instruction, 522–560.
26. Long, Y., & Aleven, V. (2016). Mastery-oriented shared student/system control over problem selection in a linear equation tutor. International Conference on intelligent tutoring systems, 90–100.
27. [14] L. Herrero and N. Carriedo, “'e role of cognitive flexibility and inhibition in complex dynamic tasks: the case of sight reading music,” in Current Psychology, pp. 1–13, Springer, Berlin, Germany, 2020.
28. [15]T. Herbert, “Public military music and the promotion of patriotism in the British provinces, c. 1780-c. 1850,” Nine-teenth-Century Music Review, vol. 17, no. 3, pp. 427–444, 2020.
29. [16] A. J. Bovin, “'e effects of frequent use of a web-based sightreading software on eighth graders’ music notational literacy,” Journal of Music, Technology and Education, vol. 11, no. 2, pp. 131–147, 2018. 131-147,2018.
30. [17] Ç. Ozen, “'e analysis and comparison of pre-service music teachers’ attitudes towards the piano lesson and their sightreading skills in terms of certain variables,” Journal of Education and Practice, vol. 8, no. 29, pp. 162–167, 2017.
31. [18] M. E. D¨uzbastılar, “'e comparasion of music teacher candidates’ instrument test grades and instrument sight reading grades in talent exam with individual instrument exam grades during four-years education period,” Journal of New Results in Science, vol. 17, no. 1, pp. 315–323, 2020.
32. [19] A. Adamyan, “Sight-reading as an important factor in the professional growth of future music teacher,” Revista Vortex, vol. 8, no. 2, pp. 1–16, 2020.
33. [20] Y. Liu, J. Wang, J. Li et al., “Zero-bias deep learning for accurate identification of Internet of things (IoT) devices,” IEEE Internet of 5ings Journal, vol. 84, p. 2627, 2634.
34. [21] S. Niu, Y. Liu, J. Wang et al., “A decade survey of transfer learning (2010–2020),” IEEE Transactions on Artificial In-telligence, vol. 1, no. 2, pp. 151–166, 2020.
35. [22] J. Yang, J. Wen, B. Jiang et al., “Blockchain-based sharing and tamper-proof framework of big data networking,” IEEE Network, vol. 34, no. 4, pp. 62–67, 2020.
36. [23] ] J. Yang, M. Xi, B. Jiang et al., “FADN: fully connected attitude detection network based on industrial video,” IEEE Trans-actions on Industrial Informatics, vol. 17, no. 3, pp. 2011–2020, 2021.
37. Stephenson M, Renz J. Agent-based adaptive level generation for dynamic difficulty adjustment in angry birds[J]. arXiv preprint arXiv:1902.02518, 2019.
38. Plass J L , Homer B D , Pawar S , et al. The effect of adaptive difficulty adjustment on the effectiveness of a game to develop executive function skills for learners of different ages[J]. Cognitive Development, 2019, 49:56-67.
39. Sampayo-Vargas S, Cope C J, He Z, et al. The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game[J]. Computers & Education, 2013, 69: 452-462.
40. Mikkel, Wallentin, and, et al. The Musical Ear Test, a new reliable test for measuring musical competence[J]. Learning & Individual Differences, 2010.

[50 ] Sarkar A, Cooper S. Meet your match rating: providing skill information and choice in player-versus-level matchmaking[C]//Proceedings of the 13th international Conference on the foundations of digital games. 2018: 1-8.

[51] Mourato F, dos Santos M P, Birra F. Automatic level generation for platform videogames using genetic algorithms[C]//Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology. 2011: 1-8.