图片包含 形状

描述已自动生成

**Analysis of Non-Functional Requirements Based on Eye Tracking**

**By Li Rongchen**

**Supervisor: Li Tong**

**Submitted to the Fan Gongxiu Honors College, BJUT**

**in Partial Fulfillment of Requirements for**

**the Degree of Bachelor of Engineering in Computer Science**

**Beijing University of Technology**

**Beijing, China**

**June, 2021**

**毕业设计（论文）诚信声明书**

本人郑重声明：在毕业设计（论文）工作中严格遵守学校有关规定，恪守学术规范；所提交的论文是我个人在导师指导下独立研究、撰写的成果，毕业设计（论文）中所引用他人的文字、研究成果，均已在毕业设计（论文）中加以说明；在本人的毕业生设计（论文）中未剽窃、抄袭他人的学术观点、思想和成果，未篡改实验数据。

本毕业设计（论文）和资料若有不实之处，本人愿承担一切相关责任。

签名：画里面的卡通动物

低可信度描述已自动生成 日期： 2022.06.12

**关于论文使用授权的说明**

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签名：画里面的卡通动物

低可信度描述已自动生成 导师签名：**卡通人物

低可信度描述已自动生成** 日期： 2022.06.12

# 摘要

质量需求，也被称为非功能需求，一般来说，更难被衡量表达，这使得它们在分析时更具挑战性。特别是，非功能需求往往是一个系统整体的属性，不能对单个组件进行验证。更加困难的是，它往往没有在正式的需求规格中明确提到，而在系统开发之初就错过了。因此，在随后的系统维护中，对这些功能的检测和修改会导致高成本。不断发展的系统在很大程度上依赖于用户反馈来评估非功能需求的满足情况。它分析的是明确的用户反馈，如App Store的评分和评论。然而，明确的用户反馈的收集具有挑战性，因此被认为是耗时的。许多用户不愿意提供这种明确的反馈。此外，对明确的用户反馈的分析也有缺陷，使得准确定义他们的需求成为挑战，导致评估需求被满足的程度的效率和准确性降低。一个潜在的解决方案是，从用户的使用过程中解析用户的隐性评论，这不需要给用户带来额外的负担。此外，用户的隐性反应直接来自于用户的客观反映，而不是他们的主观表达，这更可靠。

本论文研究了一种新的非功能需求分析方法，该方法通过研究眼动模式来自动收集用户的隐性反馈。具体来说，我们根据现有研究的认知和经验证据，建立了几种眼动指标和用户感知之间的语义联系，总结了它们与六种广泛使用的非功能需求的对应关系，并尝试将几种眼动模式正式化。最后，通过设计和实践实验验证了这些对应关系的正确性，并为这六个非功能需求提出了六个量化的评价标准。因此，该方法方法可以系统地评估六个非功能需求的满足情况，悄悄地揭开用户在正常使用软件或对其使用的软件进行用户测试时对软件的需求。此外，所提出的提供了不满足非功能需求的精确位置实现的细节，减少了对系统的修复和增强的成本。由于我们的模块化设计和 "基于眼球跟踪的需求评估框架 "的实现，这个工具现在相当有用。现有的应用程序可以迅速开始使用我们的方法，通过插入检查代码来评估他们的应用程序的非功能需求满足程度。

**关键词：非功能需求；眼动模式；实证评价；眼动跟踪**

**摘要**

所有的系统都有非功能需求，但因为对非功能性需求的研究相对功能性需求起步更晚，在系统开发的初期，非功能性需求更容易被忽视从而没有在正式的需求规格中被明确说明。当非功能性需求没有被满足时，会对用户的使用体验造成极大影响，并且这类问题有着极高的发现成本和修改成本。持续发展的系统因为其频繁迭代的特性，对非功能性需求的满足程度的评估高度依赖于用户反馈。到目前为止，大多数分析用户反馈的研究都是针对明确的用户反馈进行的，比如应用商店的评分和评论。然而，明确的用户反馈在收集上面临着困难，许多用户不愿意提供这种明确的反馈，因为这让他们付出了额外的精力和时间。另外，以明确的用户反馈作为对象的分析也存在弊端，用户很难准确地概括的需求，导致评估需求满足程度的准确程度和效率降低。而通过对用户的隐性反馈的分析可以很好地解决上述的问题。一方面，用户的隐性反馈来自于对用户使用过程的采集，不会需要对用户产生额外的负担。另一方面，用户的隐性反馈直接来自于用户的客观反映而非其主观表述，可靠性更高。但这类分析方法的难点在于如何收集用户的隐性反馈，因此相关的研究较少。

在本文中，我们对一种新的NFR分析方法进行了研究，该研究调查了眼动模式，目的是自动收集用户的隐性反馈。具体来说，我们根据现有研究的认知和经验证据，建立了若干眼动指标和用户感知之间的语义联系，总结了它们与六个被广泛使用的非功能需求之间的对应关系，并尝试将几种眼动模式形式化。最后通过设计并实践实验，验证了对应关系的正确性，并且提出了对于这六个非功能需求的六个量化的评价标准。因此，我们提出的方法可以系统地评估六个非功能需求的满足情况，在正常的用户使用过程或者用户测试过程中，悄悄地揭开用户对他们使用的软件的需求。此外，该方法提供了不满足非功能需求的精确位置实现的细节，并降低了系统的修复和增强的成本。得益于我们模块化地设计并实现的“基于眼动跟踪的需求评估框架”框架，该工具目前已具备相当的实用性。现有的应用程序可以通过插装检测代码的方式快速地开始使用我们的方法对其应用的非功能性需求满足程度进行评估。

**关键词： 非功能性需求；眼动模式；实态评估；眼动追踪**

**内容**

[摘要 III](#_Toc105867767)

[**第1章：简介** 1](#_Toc105867768)

[1.1 背景 1](#_Toc105867769)

[1.2 研究的重要意义 2](#_Toc105867770)

[1.3 目标 3](#_Toc105867771)

[1.4 论文纲要 4](#_Toc105867772)

[**第二章 :相关工作** 5](#_Toc105867773)

[2.1 基于用户明确反馈的分析 5](#_Toc105867774)

[2.2 基于隐性用户反馈的分析 5](#_Toc105867775)

[2.3 软件工程中的眼球追踪 6](#_Toc105867776)

[2.4 非功能要求的定义 6](#_Toc105867777)

[**第三章：基于眼动跟踪的需求评估框架** 8](#_Toc105867778)

[3.1 获得眼球追踪的指标 8](#_Toc105867779)

[3.1.1 第一层指标 9](#_Toc105867780)

[3.1.2 第二级指标 9](#_Toc105867781)

[3.1.3 第三层次的衡量标准 9](#_Toc105867782)

[3.1.4 第四级度量衡 10](#_Toc105867783)

[3.2 基于指标的分析 10](#_Toc105867784)

[3.2.1 准确度 11](#_Toc105867785)

[3.2.2 稳健性 12](#_Toc105867786)

[3.2.3 连贯性 13](#_Toc105867787)

[3.2.4 使用的便利性 14](#_Toc105867788)

[3.2.5 设备效率 15](#_Toc105867789)

[3.2.6 可及性 17](#_Toc105867790)

[3.3 基于眼动模式的分析 18](#_Toc105867791)

[**第四章：评估** 20](#_Toc105867792)

[4.1 参与者的选择 20](#_Toc105867793)

[4.2 实验材料和任务 20](#_Toc105867794)

[4.3 实验设计 29](#_Toc105867795)

[4.4 结果 30](#_Toc105867796)

[4.4.1 准确度 30](#_Toc105867797)

[4.4.2 稳健性 32](#_Toc105867798)

[4.4.3 连贯性 34](#_Toc105867799)

[4.4.4 使用的便利性 37](#_Toc105867800)

[4.4.5 设备效率 39](#_Toc105867801)

[4.4.6 可及性 42](#_Toc105867802)

[4.5 真实世界的测试 45](#_Toc105867803)

[**第五章：结论和未来工作** 47](#_Toc105867804)

[**书目** 49](#_Toc105867805)

[**鸣谢** 53](#_Toc105867806)

**第1章：简介**

## 1.1 背景

有效而准确地收集用户反馈对于理解用户需求是至关重要的，软件系统在此基础上不断发展。用户反馈包含功能需求（FR）和非功能需求（NFR），其中NFR的价值更高。粗略的说，FR描述了系统应该做什么，而NFR则限制了这些FR的实现方式 [1].NFRs被广泛认为在软件开发中起着举足轻重的作用。与FRs不同的是，NFRs通常有结构化的方法来捕获它们，而NFRs在总结和提取方面具有挑战性。因此，在软件开发过程中，NFR往往没有被很好地理解和充分考虑。 [2].

长期以来，软件工程的大部分注意力都集中在用于定义和交付必须执行的软件系统的功能符号和技术上。现实情况是，问题更多的是非功能性的，而不是以功能为导向的，例如，生产率低、处理速度慢、成本过高、质量低、客户不满意等 [3].在软件工程交付中，客户需要的是可用的应用程序，他们可以通过最少的培训和适当的用户界面设计立即操作，以减少错误。人们普遍认为，在开发和适当考虑这些要求时出现的遗漏或委托错误是最昂贵的，而且在信息系统完成后也很难纠正 [4].

近几十年来，将质量特征作为一个整体，而不仅仅是作为功能，一直是面向目标的需求工程领域的工作重点。特别是对待问题和解决方案的高度抽象的非功能性，已经被NFR框架广泛定义了 [3].最近的研究重点已经从定义和抽象NFR系统转移到寻找方法来评估NFR在实际生产用例中被满足的程度。

目前，从 "用户 "反馈中评估需求满意度的主流方法主要依赖于明确的用户反馈，如应用商店的分数和评论。研究人员对这种预测性信息进行分类和过滤，以从中获得有用的信息 [5–7].然而，这些方法依赖于积极的用户反馈，因此错过了那些不愿意或不擅长提供反馈的用户的需求。

此外，认知科学的研究表明，"参与者对其行为的感知并不总是与他们的基本过程和意图一致 [8]特别是当他们不得不脱离实际使用场景来进行反馈时。一些研究也开始关注隐性捕捉用户需求的方法 [9,10].这些研究使用基于条件随机场的欲望模型来诱导用户的NFRs。它们在使用前需要由相关领域的专家进行仔细的粒度调整，而且结果是由领域专家根据主观判断进行推理。这些研究使用基于条件随机场的欲望模型来诱导用户的NFRs。他们需要在使用前由相关领域的专家进行仔细的粒度调整，结果由领域专家根据主观判断进行推理。它们的分析是复杂而繁琐的，对于普通的工程师来说，需要很高的门槛才能从中获得有用的信息。此外，这些方法很难在系统之间迁移，特别是很难在实际应用中部署。此外，在其理论体系中补充和扩展它们的后续研究也面临更大的挑战。

## 1.2 研究的重要意义

为了应对这些挑战，我们提出了一个基于眼动跟踪的需求评估框架。

这篇论文在以下三个方面对艺术的现状做出了贡献。

* 首先，针对明确的用户反馈分析方法中的缺失和不准确，论文提出以非干扰性的方式定量评估NFR的满意度，用眼动技术隐性收集用户反馈。
* 其次，理论模型分为两个步骤。"将NFRs满意程度映射到用户行为 "和 "将用户行为映射到眼动指数"。这一部分的每个步骤都有一个清晰的推理过程，具有很强的可解释性，容易被他人理解。
* 第三，该模型易于应用，有非常明确的应用场景。通过简单的硬件要求（如图所示的红外摄像机）和简单的软件使用（模块化插件框架），该模型可以应用于软件开发的用户测试阶段，这样开发者就可以快速完成对用户的测试。 图 1)和简单的软件使用方法(模块化插件框架)，该模型可以应用于软件开发的用户测试阶段，这样开发者就可以快速完成可用性测试，减少该过程的时间和人力成本。该模型以非常细微的颗粒度分析了NFR在哪里以及为什么没有得到满足，这使得现有应用程序的开发者很容易维护和修改。

图形用户界面

描述已自动生成

图 1项目所需的硬件和实际使用情况

## 1.3 目标

本论文将实现以下三个目标

* 首先，我们需要总结现有研究中用户眼动行为与相应眼动指标的映射关系。我们调查和研究了现有的眼动数据的指标，在此基础上建立了一套与评估NFRs密切相关的指标。
* 其次，我们需要确定从非功能需求到用户眼动行为的映射关系，这是一个未被研究的领域，为此我们在经验上以及调查上建立了六个典型的NFR和上述指标之间的语义联系。
* 最后，我们提出了一个评估计划。理论模型的有效性通过实际实验得到了验证。

## 1.4 论文纲要

在第1章中，我们将介绍本研究提出时的背景，并强调本研究的意义和需要完成的明确目标。在第二章中，我们将详细介绍需求分析领域类似研究的研究方法，并从过去的研究中给出本研究中使用的六个NFRs的定义。在第三章中，我们将正式介绍基于眼动跟踪的需求评估框架的理论模型的推导，然后是该框架的实际实施过程。首先给出了用户行为和眼动跟踪指标之间的关联，然后解释我们如何建立从NFR到用户行为，再到眼动跟踪指标的映射。在第四章中，我们提出了一个详细的实验计划来验证理论模型的正确性，然后列出了由实验结果决定的模型的参数。通过模拟真实应用的测试场景来验证模型的正确性。最后，在第五章中，我们对这项研究进行了总结和讨论，并对如何开展后续工作进行了展望。

**第二章 :相关工作**

获得用户的反馈可以分为两类：一类是通过用户的评论获得显性反馈，另一类是通过用户的行为获得隐性反馈。下面的小标题回顾了基于这种分类的最重要的相关工作，如下。

## 2.1 基于用户明确反馈的分析

评论是获得用户对应用程序的明确反馈的最简单的方法，其中包含许多与需求相关的信息，如错误报告、功能请求、用户体验、评级等。 [5].用户评论的特点是其数量巨大。最近的一项研究发现，移动应用程序每天收到约23条用户评论，而流行的应用程序，如Facebook，每天平均收到4275条用户评论 [11].目前有几种研究集中在从非结构化和非正式的用户评论中识别有价值的信息。

已经开发了各种分类技术来区分审查的类型 [5–7]其中大部分都集中在功能需求上。然而，用户评论包括功能和非功能需求，一些研究侧重于从用户评论中找到与质量特征相关的NFRs [12].从明确的用户反馈中分析系统的用户需求，面临一个不可避免的问题。也就是说，用户必须在产品使用之外主动表达他们的需求。这种明确的反馈通常需要额外的努力或时间，使开发者有可能错过那些不愿意或不擅长提供反馈的用户的看法。此外，用户的反馈可能是不真实的或扭曲的，因为它脱离了产品在现场的使用。

## 2.2 基于隐性用户反馈的分析

隐性评价用户需求的相对好处是，它不需要用户做任何事情，所捕获的一切都相对客观和有效。一些研究尝试了隐性的NFR评价，通过条件随机场定量地探索系统与用户的交互，以发现潜在的用户需求或要求 [10]；或通过监测环境背景和人类行为背景来推断人类需求 [9].然而，所有这些努力仍然是探索性的，并面临着诸如不容易部署和难以扩展等问题。因此，非常需要一种简单、直接和系统的方法来评估用户的隐性需求。

## 2.3 软件工程中的眼球追踪

眼动跟踪是指通过记录眼睛的注视数据来收集被试者公开的视觉注意力 [13,14].视觉注意力会触发理解和解决问题所需的认知过程，而认知过程会引导视觉注意力到特定的位置。因此，眼动仪对于研究被试在执行软件工程任务时的认知过程和努力是非常有用的。 [14].

眼球追踪器提供了一个客观的、实时的、定量的眼睛注视测量，无需有意识的过滤。它们帮助研究人员研究参与者无法表达的过程和意图。 [15].眼球追踪器根据参与者在任务中注意力集中的位置，提供了对参与者正在做什么和为什么做的额外见解。

眼球追踪器帮助研究人员确定：（1）为什么参与者在完成任务时有问题；（2）参与者期望在哪里找到某些元素；（3）元素是否分散注意力；（4）设计、布局或人工制品如何有效地引导参与者完成任务；（5）参与者的效率是否存在差异，基于他们的人口统计学或专业知识；以及（6）参与者是否关注细节或简要扫描刺激物。 [16,17].所有这些都是在任务执行过程中客观地完成的。

然而，正确使用眼动仪是至关重要的。进行眼动跟踪研究需要对精细的细节进行奉献，以确保数据被正确和准确地收集。必须对收集到的数据进行仔细分析，以便将参与者的固定与他们的认知过程和意图联系起来。 [18,19].特别是，目前还没有绝对的方法来了解参与者是否真的理解了他们所固定的刺激物的一部分。 [20].

## 2.4 非功能需求的定义

众所周知，非功能需求包含了软件设计的所有方面，比传统的功能需求更难定义、衡量、测试和跟踪 [21].考虑到在用户测试过程中需要提取与用户体验相关的非功能需求，我们在初始阶段没有选择只与软件维护和后续开发相关的非功能需求。经过对非功能需求的筛选和整理，我们最终选择了几个应用广泛、定义明确、与用户体验密切相关、具有明显特征的非功能需求 [22].下面列出了保留的要求，描述来自于公开的定义 [23].

**精确性**。服务提供正确结果或效果的精确程度。

**稳健性**。程度代表了服务正确行动的能力，即使一些输入参数丢失或不正确。

**易用性**。一个服务对用户来说毫不费力地操作和控制的程度。

**一致性**。代表性服务提供相同设计模式的程度。

**设备效率**。在使用一定数量的资源时，服务的迅速程度。

**可及性**。具有独特特征和能力的人能够在使用环境中使用一项服务以实现特定目标的程度。

**第三章：基于眼动跟踪的需求评估框架**

整个框架的结构显示在 图 2。我们的框架的输入是开发人员用关键的 "兴趣区"（AOI）标记的函数的子任务。在本节中，我们将首先描述我们如何获得眼球追踪的指标。然后，解释了我们如何使用眼动跟踪指标来评估子任务的非功能需求的满足程度。最后，讨论了一种方法，从指标中提取几种模式，并利用这些模式更精细、更全面地检测用户的隐形需求，评价软件的非功能需求的满足程度。

图示

中度可信度描述已自动生成

图 2基于眼睛追踪的需求评估框架的结构。

## 3.1 获得眼球追踪的指标

通过参考其他眼球追踪项目的研究方法 [24,25]，我们可以使用几个眼动跟踪的指标来描绘用户感知的不同方面。

高层次的指标是基于低层次的指标，它需要更复杂的处理，包括但不限于计算、去噪等。级别越高，对狭义语义和清晰特征的表述就越好。在我们的研究中，高层次的指标所占的权重就越高。

### 3.1.1 第一层指标

大多数经济实惠的眼球追踪设备可以收集的原始数据是屏幕上的眼睛注视点位置（X，Y）。一般来说，这一数据的收集依赖于计算机视觉技术来跟踪用户的头部和眼睛的旋转角度。

### 3.1.2 第二级指标

眼球追踪器可以通过使用事件检测算法来定位眼睛的指示着陆点，从而区分固定（Fixations）和移位（Saccades）。固定往往意味着用户正在对一个特定的物体产生意识，而移位则意味着用户的眼睛只是沿着位置快速移动。我们选择了基于离散阈值的方法，具有出色的准确性和鲁棒性 [26].

### 3.1.3 第三级指标

兴趣区（AOI）是在这个层面上引入的，作为标记刺激区和提取专门针对这些区域的指标的工具。我们在研究中使用软件中的组件作为AOI区域。

固定时间（FD）。凝视关键AOI的时间。以前的研究用这个指标来描述参与者的努力程度 [20].该指标的值与用户在AOI上投入的精力有关。

固定次数(FC)：关键区域的固定次数。它也可以描述参与者花在AOI上的精力或努力程度。然而，长时间的凝视可能只产生单一或少量的固定次数，这与固定时间不同。

固定率（FR）。固定在关键区域的总次数与其他区域的比率。固定率越低，说明搜索任务的效率越低：参与者花更多的精力去寻找相关区域 [27].较高的固定率可能表明需要更多的努力来完成任务。 [20].

首次定影时间（FFT）。在AOI中第一次定影的时间可以表明关键AOI的引人注意的程度。

### 3.1.4 第四级指标

四级指标中引入了 "扫描路径 "的概念，它是描述固定长度和时间的注视点或AOIs的序列。进行扫描路径之间的比较可以识别和分析''参与者解决任务的观看策略 [28].

扫描路径精度（SPA）。在被注视的AOIs中，关键AOIs的数量与其他AOIs的数量之比。该值越高，表明被试者对任务的理解程度越高 [29].

编辑距离（ED）。将扫描路径转换为开发者确定的最短路径的编辑成本是通过插入、替换和删除等基本操作来计算的。该值越小，参与者的行为就越符合开发者的期望。

注意力转换的数量（AS）。在单位时间内，在一连串的AOIs之间切换的总次数。切换次数越多，意味着参与者对正在进行的任务更不确定。

凸壳区（CA）。包含被试者所有固定点的最小的凸形集合。较小的表示凝视点之间的距离越近，被试者在寻找相关区域时投入的精力越少 [30].

## 3.2 基于指标的分析

众所周知，非功能需求包含了软件设计的所有方面，与传统的功能需求相比，非功能需求更难定义、衡量、测试和跟踪。经过对非功能需求的筛选和整理，我们最终选择了几个被广泛使用的非功能需求 [23]，与用户体验密切相关，并且有明确的特征。我们发现，一旦某个非功能需求没有被满足，用户的指标就会发生一些特征性的变化。我们根据上文A段提到的被广泛接受的响应用户的各个指标的行为特征，结合各个NFR不被满足时在人机交互中出现的用户特征，将眼动跟踪的指标与NFR对应。其中，有些指标在某个NFR不被满足时有很大的反应，而其他指标在这个NFR被满足或不被满足时没有明显的差别。到目前为止，我们所确定的NFR和变化的特征之间的对应关系如图所示 图 3.

图示

描述已自动生成

图 3NFR和眼动指标之间的关系

### 3.2.1 准确度

准确度不达标表示该服务提供的结果不准确，因此不容易让用户获得返回结果的正确含义。也就是说，用户需要比较界面上的元素，以收集更多的信息来帮助他们理解。用户产生搜索行为来收集整个界面上的元素的信息。搜索行为导致定点分散在其他AOI上，这就降低了定点率和扫描路径的准确性。因此，用户搜索行为的不可预测性也导致了编辑距离的显著增加，而频繁的AOI切换也增加了注意力切换的次数。因此，简而言之，与标准行为相比，用户的行为在不一致时表现出以下特点。

* 固定率低
* 扫描路径精度低
* 长距离编辑
* 高数量的注意力转换

然而，由于用户可以毫不费力地找到相关的AOI，因此首次固定时间并没有改变。另外，在搜索中，由于用户被混乱的返回结果所迷惑，对大方向很清楚，所以凸壳区域很小。所以用户会围绕关键AOI进行搜索，范围不会很大。此外，一旦找到关键的AOI，用户就不再面对了，所以他的定点次数和定点时间都和平时一样。

* 在相关AOI上的首次固定时间
* 固定数
* 固定时间
* 凸形船体面积

总之，只有固定率、扫描路径准确性、编辑距离和注意力切换的数量发生了变化。对于编辑距离和注意力切换次数，较高的数值表明对这个NFR的不满意程度较高；对于固定率和扫描路径准确性这两个参数，较低的数值表明对这个NFR的不满意程度较高。我们用以下公式来评估对准确性的不满意程度。分配给固定率的系数相对较小，因为它是一个较低层次的指标。相比之下，分配给扫描路径准确性、编辑距离和注意力切换次数的系数较高，因为它们是更高层次的指标。

### 3.2.2 稳健性

当准确度不达标时，表明当输入系统的参数缺失或不正确时，系统不能正常工作，也就是说，用户的行为导致了他没有预期的结果。然而，用户往往很容易推断出系统异常的原因，并会在下一次尝试中加以改进。重试行为使用户再次操作相关的AOI，与标准行为相比，其定影次数和定影时间都增加了一倍甚至更多，由于用户没有在非相关的AOI上投入那么多精力，所以定影率也随之提高。所以，与标准行为相比，用户的行为在不符合一致性的情况下会表现出以下特点。

* 高固定数
* 高固定时间

然而，在输入错误的参数之前，对相关AOI的关注和使用基本上是正确的。用户不难发现什么是完成任务的相关区域，这将导致对相关AOI的首次固定时间与标准行为完全相同。重试行为发生在第一次正确尝试之后，这意味着编辑距离没有明显增加。重试行为集中在相关的AOI上，不会在非相关的AOI和相关的AOI之间切换，这将导致固定率、注意力转换次数和凸壳面积与标准行为没有明显区别。因此，以下指标将与标准行为完全一致，不会改变。

* 固定率
* 在相关AOI上的首次固定时间
* 编辑距离
* 注意力开关的数量
* 凸形船体面积

综上所述，只有固定次数和固定时间发生了变化。对于固定次数和固定时间这两个参数，较低的数值表示对这个NFR的不满意程度较高。我们用以下公式来评估对稳健性的不满意程度。因为定格次数和定格时间属于同一级别的眼球追踪指标，所以不需要增加用于划分不同级别眼球追踪指标的权重。

### 3.2.3 连贯性

当不符合一致性时，表明系统没有提供相同的设计模式，也就是说，具有相同表意文字的元素的行为是不同的，或者具有不同表意文字的元素具有相同的外观或行为模式。这可能会让用户感到困惑，因为他们很难在第一时间发现他们需要的元素。无法在第一时间找到的元素最明显的特征是，由于用户首先被其他元素所吸引，所以对相关元素的第一次固定很晚。此外，当用户在尝试失败后被迫找到正确的感兴趣的元素时，会表现出搜索特征。搜索的行为会导致定影分散在其他的AOIs上，从而导致定影率和扫描路径的准确性降低。用户搜索行为的不可预测性也导致了编辑距离的显著增加；在AOIs中的频繁切换增加了注意力切换的次数。一般来说，当不符合标准行为时，与标准行为相比，用户的行为会表现出以下特点。

* 固定率低
* 在相关AOI上的首次固定时间较晚
* 扫描路径精度低
* 长距离编辑
* 高数量的注意力转换

但在搜索过程中，由于用户只是被混乱的设计模式所迷惑，对大方向很清楚，所以凸壳的面积很小。因此，用户会围绕关键AOI进行搜索，范围也不会很大。此外，一旦找到关键的AOI，用户在操作上就不会再面临困难。所以他的定影次数和定影时间都和平时一样。因此，以下指标不会有任何变化，将与标准行为保持一致。

* 固定数
* 固定时间
* 凸形船体面积

总而言之，固定率、首次固定时间、扫描路径准确度、编辑距离和注意力切换的数量都取决于满足程度的高低。满意度越差，所遵循的固定率、首次固定时间和扫描路径准确性就越低，所遵循的编辑距离和注意力切换次数就越高。因此，我们使用下面的公式来评估这个NFR的满意度。注意，我们为不同级别的指标设置了不同的权重，级别越高，权重越高。

### 3.2.4 使用的便利性

如果易用性不理想，服务对用户来说就很难操作和控制。太多的操作难度会导致更多的努力，使用户持怀疑态度并进行多次尝试。太多的努力和用户的反复尝试会导致在关键的AOIs上花费异常多的时间和注意力，所以固定时间很长。然后，任务执行起来非常困难，用户可能会怀疑是否是正确的操作，从而引发怀疑和寻找周围的其他元素，这就带来了低的扫描路径准确性和高的注意力切换次数。由于搜索行为是与操作穿插进行的，所以固定次数会增加。所以，一般来说，当不符合标准行为时，与标准行为相比，用户的行为会表现出以下特点。

* 高固定数
* 高固定时间
* 扫描路径精度低
* 高数量的注意力转换

然而，由于关键AOIs内部和外部的固定数都很大，固定率并没有明显提高，而且搜索区域是在关键AOIs周围。因此，"凸壳 "区域很小。此外，所有的搜索行为都是在用户找到难以操作的相关元素后才出现的。用户被正确地引导到相关元素上开始操作，所以第一次固定是正常的早期，而且编辑距离也很短。

* 固定率
* 在相关AOI上的首次固定时间
* 编辑距离
* 凸形船体面积

只有定格次数、定格持续时间、扫描路径准确度和注意力切换次数有变化。对于定格持续时间和注意力切换次数这两个参数，较低的数值表示对该NFR的满意度较高；对于扫描路径准确性，较高的数值表示对该NFR的满意度较高。我们用下面的公式来评估对易用性的满意程度。分配给固定时间的系数相对较小，因为它是一个较低级别的指标。相比之下，分配给扫描路径准确性和注意力切换次数的系数较高，因为它们是更高层次的指标。

### 3.2.5 设备效率

当设备效率不达标时，表明服务在使用一定数量的资源时反应缓慢或延迟，意味着用户必须等待对某些内容的响应。用户产生了一个等待行为。等待一般分为两步。在第一步中，用户首先试图直接观察相关AOI中的内容，以便第一时间捕捉其内容的变化，这将导致用户在AOI中的固定次数非常多，固定时间也非常长。一旦等待时间超过一定限度，用户会因为无聊而进行第二种等待行为，即 "开始自己找东西"，即在界面中漫无目的地搜索。这反映在眼球追踪的数据指标上，就是扫描路径的准确性明显下降，注意力切换的次数明显增加。此外，由于搜索是无目的的，凸壳的面积也会增加。所以，一般来说，与标准行为相比，当不一致时，用户的行为表现出以下特点。

* 高固定数
* 高固定时间
* 扫描路径精度低
* 高数量的注意力转换
* 较大的凸形船体面积

然而，由于第一和第二等待操作分别增加了相关AOI和不相关AOI的固定次数，固定率趋于平缓，即与标准行为没有太大区别。此外，由于设备的低效率导致的响应延迟并不影响用户对页面元素的感知，所以相关AOI上的第一次固定时间也与标准行为没有明显的区别。最终，等待行为都发生在用户动作完成之后，所以编辑距离也没有什么变化。以下指标将与标准行为完全匹配，不会发生变化。

* 固定率
* 在相关AOI上的首次固定时间
* 编辑距离

综上所述，只有固定次数、固定时间、扫描路径准确性、注意力转换和凸壳面积发生了变化。对于固定次数、固定时间、注意力切换和凸壳面积，数值越高表示对这个NFR越不满意；对于扫描路径准确性，数值越低表示对这个NFR越不满意。我们用以下公式来评估对设备效率的不满意程度。分配给定影次数和定影时间的系数相对较小，因为它是一个较低层次的指标。相比之下，分配给扫描路径准确性、注意力开关和凸壳面积的系数较高，因为它们是更高层次的指标。

### 3.2.6 可访问性

当Accessibility不满足时，表明用户不具备访问该内容的相关技能，也就是说，用户需要大量的搜索、理解和实验才能逐渐找到任务的解决方案。然而，由于页面上的非相关AOI往往多于相关AOI，而且用户对目的地没有清晰的认识，所以最终的固定率仍然很低。对相关AOIs的第一次固定时间往往很晚。此外，广泛和随机的尝试和搜索也会导致非常低的扫描路径准确性，非常长的编辑距离，特别多的注意力切换，以及一个凸形的船体区域。所以，一般来说，与标准行为相比，在不一致的情况下，用户行为表现出以下特点。

高固定数

* 固定时间长
* 固定率低
* 在相关AOI上的首次固定时间较晚
* 扫描路径精度低
* 长距离编辑
* 高数量的注意力转换
* 凸形船体面积大

综上所述，固定次数、固定时间、固定率、在相关AOI上的首次固定时间、扫描路径准确性、编辑距离、注意开关数量和凸壳面积都有变化。对于定影次数、定影时间、在相关AOI上的首次定影时间、编辑距离、注意力切换次数和凸壳面积，数值越高，说明对NFR的不满意程度越高。对于固定率和扫描路径准确性这两个参数，数值越低，说明对NFR的不满意程度越高。我们用以下公式来评价无障碍设施的不满意程度。分配给四个参数定影次数、定影时间、定影率和首次定影时间的系数相对较小，因为它们是较低层次的指标。相比之下，编辑距离、注意力切换次数和凸壳面积这四个参数的系数较高，因为它们是较高层次的指标。

## 3.3 基于眼动模式的分析

上面的分析表明，许多固定的模式在多个NFR中反复出现。我们开展了进一步的工作，从眼动指标中抽象出这些模式，以帮助我们更好地评估非功能需求。我们使用可变长度的滑动窗口的想法，在用户完成任务的时间轴上以更精细的粒度分析上述的多个指标。

从用户感知的角度来看，当每个NFR不被满足时，用户的行为总是由多种模式组成的 [31].我们以两个模式为例，一个是 "搜索"，一个是 "执行任务"，我们将在下文中详细描述。这两个模式为更多模式的提出奠定了基础，以同样的方式提供对NFR评估的更细化的分析。

**搜索。**

通过分析经验，我们把低的固定率、低的扫描路径准确性和高的注意力切换次数作为一种模式。当它们在一个时间窗口中组合出现时，就代表了用户所使用的搜索行为，因为在搜索时，固定的着陆点必然会在多个AOI之间徘徊。

有多种NFR，当用户不满意时，他们在任务执行过程中表现出搜索行为。例如，当易用性不满足时，用户会在任务执行过程中穿插搜索行为，以排除因自己的行为而导致任务困难的可能性。当一致性不满意时，用户在一开始就表现出搜索行为来寻找关键的AOI；当设备效率不满意时，用户由于等待的无聊而开始搜索整个页面的变化，等等。

在不同的情况下，搜索模式出现在不同的位置，在每个NFR不被满足的情况下，这意味着每个NFR的不同。例如，在一致性的情况下，搜索模式是核心行为特征。使用捕获的数据，我们可以进一步分析哪些元素与搜索对象的关键AOI设计模式相冲突，从而使用户感到困惑。然而，在设备效率的情况下，当NFR不被满足时，搜索只是一个副产品，所以我们不需要关心具体的对象。

图表

描述已自动生成

图 4几个图案的原始眼球追踪数据，黄色荧光区域代表关键的AOI区域，点代表固定。固定时间越长，点越大。

**表演。**

当定格率极高，定格时间极长，扫描路径精度极高，并且在一段时间的时间窗口内，凸壳面积极小，我们可以断定，用户正在完成子任务的目标。无论是否存在NFR不满足，执行任务模式总会出现，因为对于每个子任务，完成是通往下一个子任务的必经之路。

当一致性或易用性不被满足时，用户只能有两种行为模式，即搜索和执行。然而，区别在于模式的顺序：一致性不被满足时的用户行为是 "搜索"-"执行"，而易用性不被满足时的用户行为是 "执行"-"搜索"-"执行"。

**第四章：评估**

目前，我们仍然只能对所有的指标模型进行粗略的定性分析，只能猜测每个指标的高低。为了获得详尽的、系统的、定量的模型，我们需要设计实验，在真实的用户测试中收集与每个NFR对应的眼动数据。

## 4.1 参与者的选择

对照实验的参与者将从具有一定计算机使用基础知识的人中招募。实事求是地说，为了保证参与者的数量能得出相对合理的结论，我们计划招募20名计算机科学与技术专业的学生参与实验。

## 4.2 实验材料和任务

我们的测试材料的组织结构如下图所示 图 7中所示。对于硬件设备，我们使用Tobii Eye Tracker 5游戏眼球追踪器，如图4和图5所示。使用基于流媒体引擎开发的驱动程序，我们从硬件设备上获得了原始的眼球追踪数据，并将其提交给后端应用程序。我们设计了测试中的软件来模拟用户直接与之互动的真实系统。一些代码被输入到被测程序中，将子任务的实时执行和组件的位置发送到后端应用程序。



图 5Tobii眼球追踪器5游戏眼球追踪器设备



图 6眼球追踪器使用的实际实验场景

后台应用程序分为两部分，其中HTTP服务器接受并记录来自被测软件和眼动跟踪组件的数据。分析系统对记录的眼动数据、子任务执行情况和组件坐标位置进行分析，包括但不限于从原始眼动数据中提取高级指标、眼动模式等。最后，分析系统将报告每个子任务的NFR满足情况。

图示

描述已自动生成

图 7实验材料组织

被测软件通过后台程序与数据采集平台进行通信，后台程序记录硬件设备获取的原始眼球追踪数据和前端子任务的执行情况以及组件的实时位置，从而为后期分析提供依据。

我们设计了一个基于网络应用的模拟大学图书馆系统作为被测软件。该软件本身具有信息查询和座位预订等功能，所有年龄段的预览图显示在 图 8到 图15.我们开发了在其上实时插入故意不良设计的能力。我们计划将该应用程序作为评估上述六个典型NFR的目标软件。我们将让参与者使用该软件，并着重于评估图书馆的座位预订功能。

我们将整个过程分为五个子任务，即 "导航到预订功能页面"、"选择学习室地点"、"选择日期和时间"、"选择座位 "和 "提交预订信息"。具体的子任务以及我们如何在这些子任务上插入有意的不良设计，如图所示 表 1下图。当参与者完成每个子任务时，我们将使用上述的眼动数据收集系统来收集他们的眼动数据。

表 1实验分组以及故意不良设计、非功能需求和子任务之间的对应关系

|  |  |  |  |
| --- | --- | --- | --- |
| 测试组 | 副任务 | 经过测试的NFR | 有意的设计缺陷 |
| 测试组1 | 导航至预订功能页面 | 无障碍设施 | 将导航主页的语言设置为俄语 |
| 选择日期和时间 | 使用的便利性 | 选择时间的确定范围很小，需要高度准确地击中按钮的中心。 |
| 选择座位 | 稳健性 | 不选择或选择一个以上的座位，系统将崩溃，迫使用户重新选择 |
| 测试组2 | 选择学习室场地 | 准确度 | 点击主图书馆按钮将选择图书馆1；点击图书馆2将只选择主图书馆 |
| 选择日期和时间 | 一致性 | 开始时间和结束时间的选择提供了不同的设计模式 |
| 提交预订信息 | 设备效率 | 点击提交按钮后，页面进入加载状态，人为地降低了响应速度。 |
| 控制组 | 没有故意的设计缺陷 | | |

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

图 8不插入故意不良设计的主页预览

子任务 "导航到预订功能页面 "发生在用户第一次进入系统时出现的主页上，如上图所示。用户需要认识和理解侧边栏和顶部导航栏，并点击正确的栏目，进入座位 "预订 "功能页面。这样，完成当前的 "导航到订票功能页面"，就完成了当前的 "导航到订票功能页面 "的子任务，开始下一个子任务。

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

图 9在插入了有意为之的不良设计后的主页预览图

我们在这个方案中插入的非功能需求是可及性。它被定义为 "具有独特特征和能力的人可以使用一项服务的程度，以便在使用环境中实现特定的目标"。特意的反向设计为：针对 "没有俄语阅读和理解能力 "的参与者。故意用俄语取代指南，以防止他们使用服务和实现当前子任务的任务目标。在参与者的使用过程中，出现了改善当前页面的 "无障碍性 "的需求。

图形用户界面, 应用程序

描述已自动生成

图 10选择学习室场地页面的预览

子任务 "选择学习室场地 "的场景如上图所示，它发生在用户完成子任务 "导航到预订功能页面 "之后。进入 "预订 "功能页面后，出现了第一个页面。用户需要通过识别和理解页面中央的引导来选择和点击他所选择的场地。用户选择的场地会有精确的阴影作为反馈，表示该场地已被选中。然后，用户需要点击 "下一步 "按钮，完成当前的 "选择学习室场地 "子任务，开始下一个子任务。

我们在这个场景中插入的非功能需求是准确性，定义为 "服务提供正确的精确程度"。在当前的场景中，不是用户点击的场地被选中，而是用户点击的下一个场地被选中。在参与者的使用过程中，出现了改善当前页面的 "准确性 "的需求。

日历

描述已自动生成

图 11选择日期和时间页面的预览，没有插入故意的坏设计

子任务 "选择日期和时间 "的情景，如上图所示。它发生在用户完成了 "导航到预订功能页面"，并完成了 "导航到预订功能页面 "和 "选择学习室场地 "子任务之后，这是进入 "预订 "功能页面后出现的第二个页面。用户需要认识和理解页面中央的功能组件的作用，在左边的日历页面中选择日期，在右边的两个时间选择列表中选择开始和结束时间。两个时间选择列表都使用相同的设计模式，并且对日期和时间都有广泛的点击判断能力。然后用户点击 "下一步 "按钮，完成当前的 "选择日期和时间 "子任务，开始下一个子任务。

我们在这个方案中插入的非功能需求是 "易用性 "和 "一致性"。我们首先描述了我们对 "易用性 "的实验设计，它被定义为 "一个服务对用户来说容易操作和控制的程度"。我们特意做了一个反向设计：为了使服务难以操作和控制，在当前场景下，我们特意缩小了日期和时间选择控件的判断范围，将判断范围改为原来的十分之一。在目前的情况下，用户必须谨慎行事才能完成任务。在参与者的使用过程中，出现了改善当前页面的 "易用性 "的需求。

日历

描述已自动生成

图 12选择日期和时间页面的预览，为了保持一致性，故意插入了不好的设计

接下来，我们描述一下我们对一致性的实验设计，如上图所示，对于一致性的定义，即 "代表服务提供相同设计模式的程度"。在目前的情况下，两个含义相同的时间选择控件使用了两种不同的设计模式：一个扩展菜单和一个下拉菜单。让参与者在使用过程中产生改善当前页面的 "一致性 "的需求。

图形用户界面, 文本, 应用程序

描述已自动生成

图 13提交预订信息页面的预览

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

图 14提交预订信息页面加载时的预览

子任务 "提交预订信息 "的场景如上图所示，发生在用户完成 "导航到预订功能页面"、"选择学习室场地 "和 "选择日期和时间 "子任务之后。这是进入 "预订 "功能页面后出现的第三个页面。用户需要点击 "下一步 "按钮，确认并提交前两个子任务中选择的预订信息，完成当前的 "提交预订信息"。提交后，系统将在数据库中搜索当前可用的座位，并在下一页显示。在系统搜索完成之前，加载图标将在页面上循环显示，不会进入下一个页面和下一个子任务。我们在这个场景中插入的非功能需求是 "设备效率"。它被定义为 "服务在使用一定数量的资源时的迅速程度"。特意反向设计为：降低服务返回结果的及时性，即在当前场景下，让后端至少需要10秒才能完成搜索，这样，参与者在使用当前页面的过程中，就会出现提高设备效率的需求。

图片包含 图形用户界面

描述已自动生成

图 选择座位页的预览，没有插入故意的坏设计

"选择座位 "子任务的场景如上图所示，发生在完成 "导航到预订功能页面"、"选择自习室场地"、"选择日期和时间"、"提交预订信息 "子任务之后，这是进入 "预订 "功能页面后出现的第四页，也是整个实验中的最后一页。用户需要在页面上找到自己需要的座位号，点击相应的图标，最后点击下一步按钮，完成该子任务。

图形用户界面, 应用程序

描述已自动生成

图 选择座位页的预览，为了稳健起见，插入了有意的坏设计

我们在这个场景中插入的非功能需求是鲁棒性。它被定义为 "代表服务在某些输入参数缺失或不正确的情况下仍能正常行动的程度"。刻意的反向设计：使服务在输入参数缺失或不正确的情况下容易崩溃，不能正常行动，即在目前的场景中，一旦用户选择了一个以上的座位或没有座位，服务就会 "崩溃"，用户必须重新加载页面并重试。这使得参与者有可能使用该服务，这样就出现了改进当前页面的鲁棒性的需求。

## 4.3 实验设计

我们将参与者分为两组，每组10人。基于上述自行开发的插入不良设计的测试软件，我们在不同的子任务中设计了不同的NFRs的测试任务，涵盖了我们提到的六种典型的NFRs。在对实验组的测试中，我们将为参与者开启故意的不良设计，以触发他们的NFR。这些设计是故意错开的，这样就不会互相干扰，并给参与者时间从之前由困惑和不满引起的不良设计中恢复过来。而在这两组中，故意打开不良设计的子任务将被作为实验组数据使用。相反，没有开启不良设计的子任务将被用作对照组数据，这样分配是为了最大限度地利用有限的参与者数量。控制组的数据将被用作完成每个子任务的基线数据。我们将能够把对照组与实验组进行比较，以验证我们上述理论模型是否正确。在实验过程中，我们会要求参与者使用 "大声思考 "的方法来验证故意的不良设计确实诱发了他们相应的NFR。

## 4.4 结果

我们比较了收集到的实验数据和对照数据之间的差异，分析它们之间的差异是否与我们提出的指标模型的特征一致。最后，结合这些数据，我们希望通过将每个指标特征的判断从高或低的定性提升到具体的尺度来改进我们的模型。通过这种方式，模型允许在判断上更加精确。实验结果如下。

### 4.4.1 准确度

实验组和对照组在 "选择空间 "子任务中的数据对比计算如下。

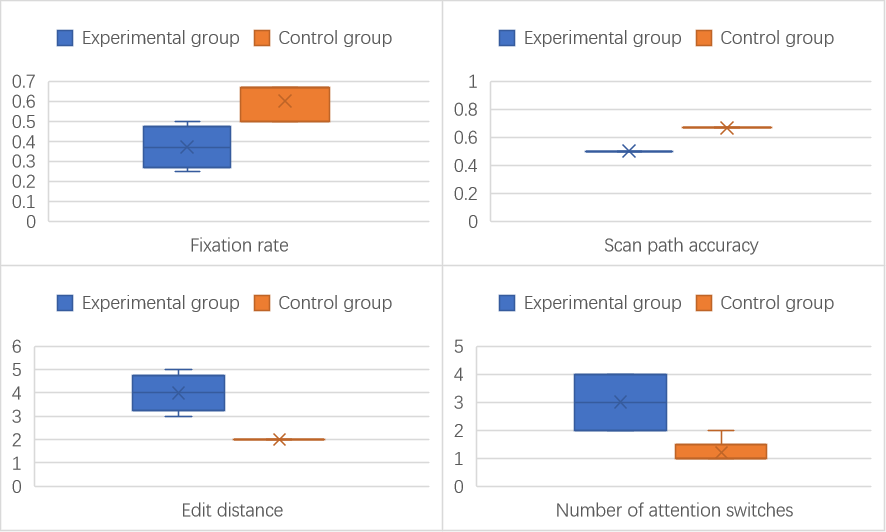


图 17在 "选择空间 "子任务中，实验组和对照组之间有显著差异的四个眼动跟踪数据的比较

根据上述 图 17，可以看出3.2节中的分析。实验组的 "固定率 "都低于对照组，这可以作为一个特点。而 "扫描路径准确率 "由于AOI区域数量较少，且结果相对单一，但仍能明显看出实验组平均低于对照组的特点。实验组的 "编辑距离 "均高于对照组，这可以作为一个特点。实验组的 "注意转换 "高于对照组，这可以作为一个特征。



图 四个眼球追踪数据的比较，在 "选择空间 "子任务中，实验组和对照组之间没有显著差异

图18显示，如3.2节所分析。实验组的 "固定次数 "包含了对照组的所有数据，这意味着它们之间没有明显的差异。实验组的 "固定时间 "和 "第一次固定在AOI上的时间 "的中位数和平均值都与对照组非常相似，说明它们之间没有明显的差异。对照组的 "凸壳面积 "包含了实验组的所有数据，说明它们之间没有显著差异。上述表现表明，当 "准确度 "满足和不满足时，这些指标的表现没有变化，所以不能作为识别 "准确度 "满足程度的一个特征。

综上所述，当 "准确度 "不满足时，其特点是固定数低，扫描路径准确度低，编辑距离高，注意力转换高。因此，将不满足度的评价公式中的各参数带入真值，以找到

对照组和实验组的 "准确度 "不满意分数如下所示。

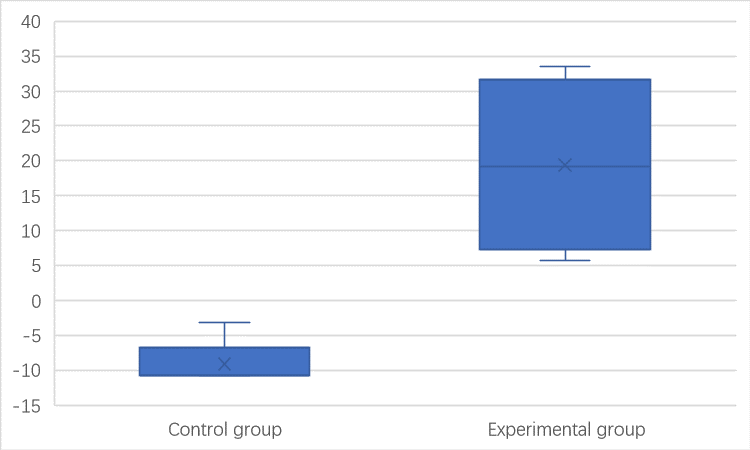


图 19实验组和对照组在 "选择空间 "子任务中的 "准确度 "不满意分数的比较

这表明，"准确度 "不满足的公式程度可以用0为界来评价准确度是否满足。

### 4.4.2 稳健性

实验组和对照组在 "选择座位 "子任务中的数据比较如下。

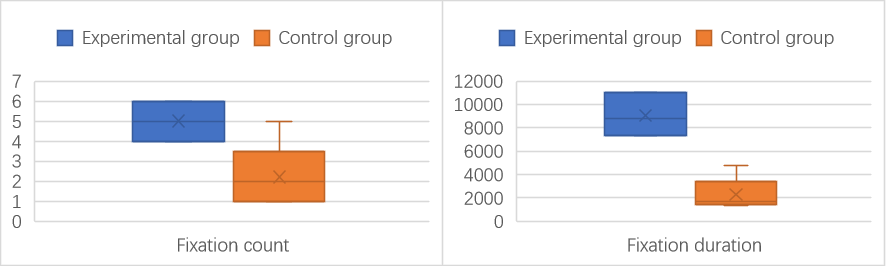


图 20在 "选择座位 "子任务中，实验组和对照组的两个眼球追踪数据的比较有显著差异

在 图 20上，可以看到3.2节的分析。实验组的 "固定次数 "绝大部分高于对照组，这可以作为一个特征。实验组的 "固定时间 "比对照组高，这可以作为一个特征。

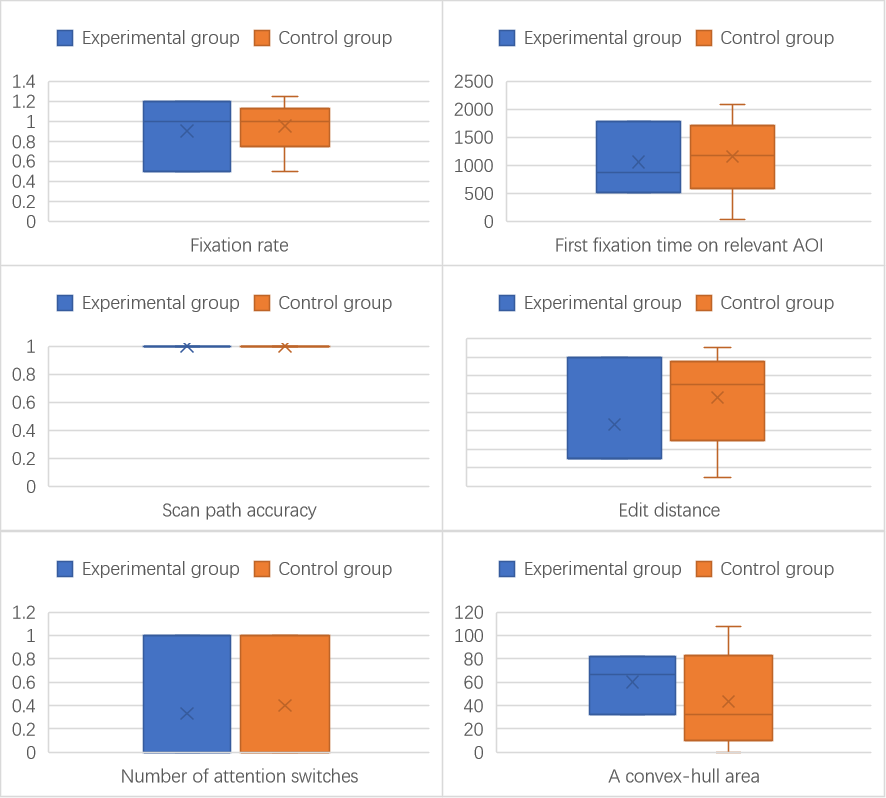


图 21六个眼球追踪数据的比较，在 "选择座位 "子任务中，实验组和对照组之间没有显著差异。

图 21显示，正如第3.2节所分析的那样。对照组的 "固定率"、"在相关AOI上的首次固定时间"、"扫描路径准确度"、"编辑距离"、"注意力切换次数 "和 "凸壳面积 "包含了实验组的所有数据，这说明它们之间没有明显的差异。上述表现表明，当 "鲁棒性 "满足和不满足时，这些指标的表现没有变化，所以不能作为识别 "鲁棒性 "满足程度的一个特征。

综上所述，当 "稳健性 "不满足时，其特点是固定数和持续时间高。因此，不满足度的评价公式中的每个参数都要带入真实值来寻找。

对照组和实验组对 "稳健性 "不满意的评分如下。

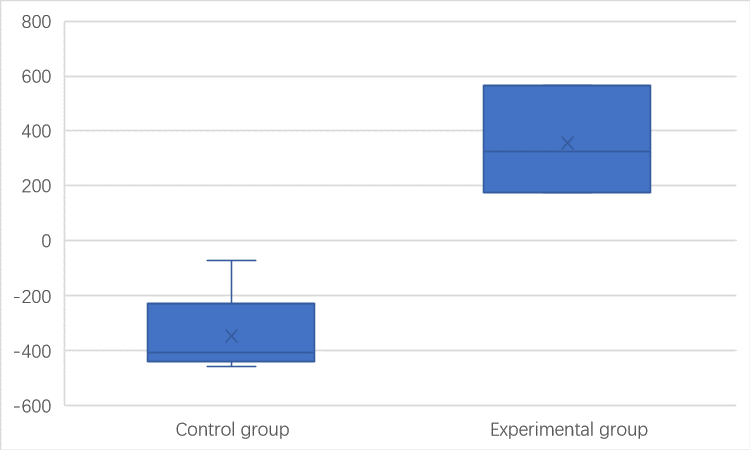


图 22实验组和对照组在 "选择座位 "子任务中的 "稳健性 "不满意分数的比较

这表明，"Robustness "不满足公式的程度可以用0为界来评价Robustness是否满足。

### 4.4.3 连贯性

实验组和对照组在 "选择日期和时间 "子任务中的数据比较如下。

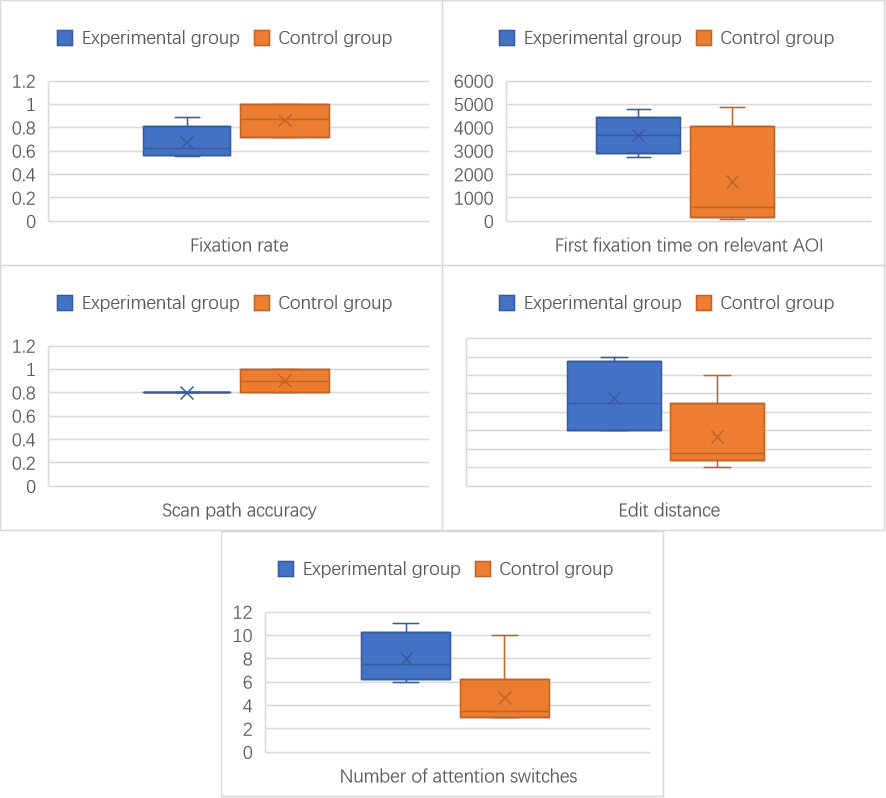


图 23在 "选择日期和时间 "子任务中，实验组和对照组之间有显著差异的五个眼动跟踪数据的比较

画面中的 图 23以上可以看作是3.2节中的分析。实验组的 "固定率 "的平均值和中位数都低于对照组，这可以作为一个特征。实验组的 "首次固定在相关AOI上的时间 "的平均值和中位数均高于对照组，这可以作为一个特征。实验组的 "扫描路径准确度 "绝大部分低于对照组，可作为一个特征。实验组的 "编辑距离 "的平均值和中位数都高于对照组。实验组的 "注意力切换次数 "绝大多数都比对照组高，这可以作为一个特征来使用。

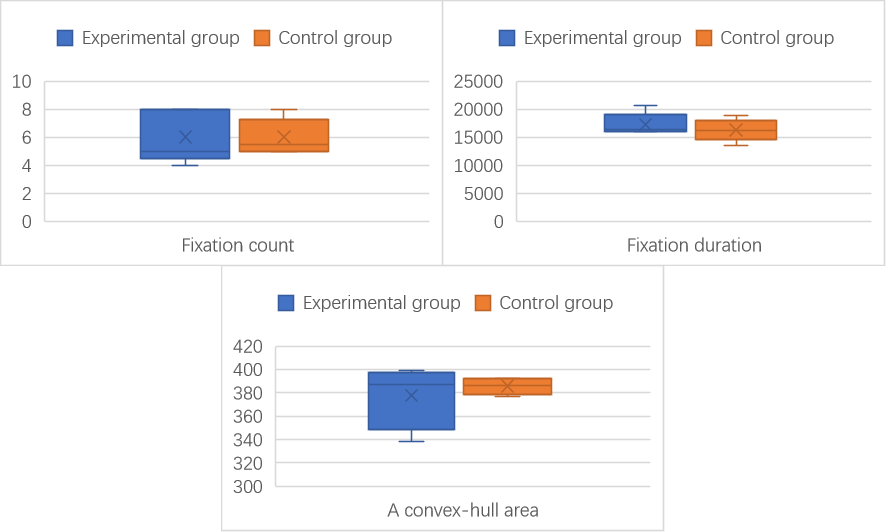


图 24在 "选择日期和时间 "子任务中，实验组和对照组的三个眼球追踪数据的比较没有明显差异

图 24显示，如3.2节所分析的。实验组的 "固定次数 "和 "固定时间 "的中位数和平均值都与对照组非常相似，说明它们之间没有明显的差异。实验组的 "a凸壳面积 "包含了对照组的所有数据，说明它们之间没有明显的差异。上述表现表明，当 "一致性 "满足和不满足时，这些指标的表现没有变化，所以不能作为识别 "一致性 "满足程度的一个特征。

综上所述：当 "一致性 "不满足时，其特点是固定率低，在AOI上的首次固定时间高，扫描路径准确性低，编辑距离高，注意力转换次数多。因此，将评价方程中各参数的不满足程度带到真实值中去寻找。

对照组和实验组的 "一致性 "不满意度得分如下所示。

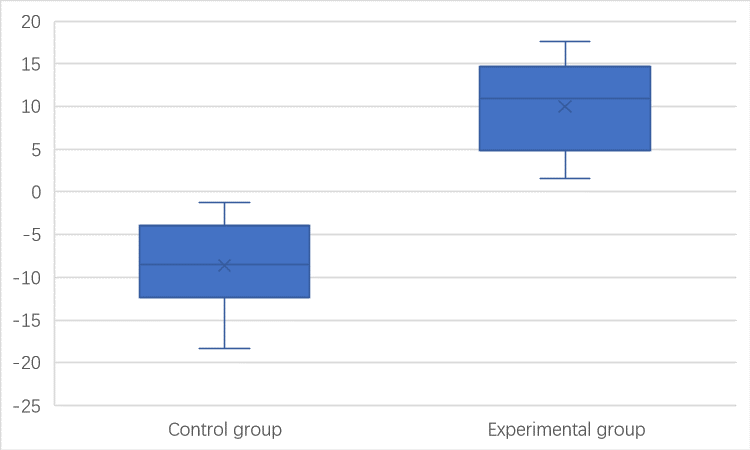


图 25实验组和对照组在 "选择日期和时间 "子任务中的 "一致性 "不满意分数的比较

可以看出，"一致性 "不满足度公式中的度可以用0为界来评价一致性是否满足。

### 4.4.4 使用的便利性

实验组和对照组在 "选择日期和时间 "子任务中的数据比较如下。

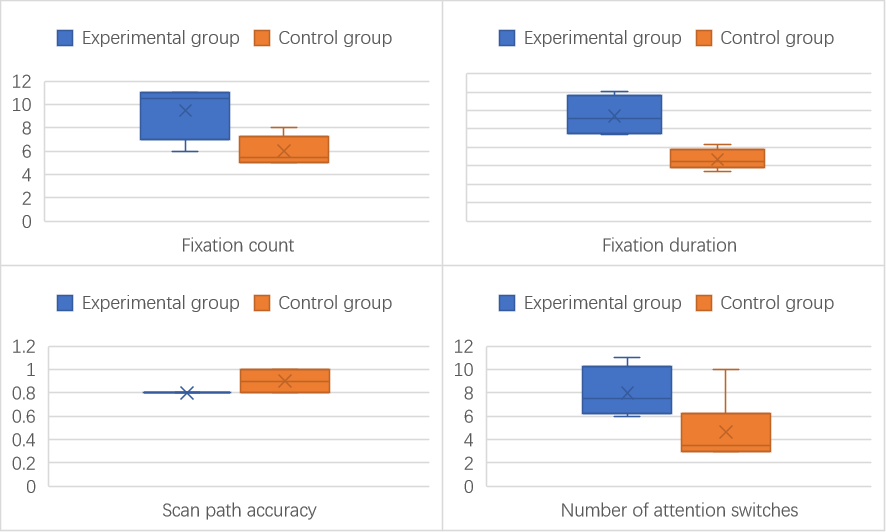
****

图 26在 "选择日期和时间 "子任务中，实验组和对照组之间有显著差异的四个眼动跟踪数据的比较

根据上述 图 26，可以看出3.2节中的分析。实验组的 "固定次数 "绝大部分高于对照组的平均值，这可以作为一个特征。实验组的 "固定时间 "明显高于对照组的平均值，这可以作为一个明显的特征。由于AOI区域的数量较少，"扫描路径准确度 "相对单一，但仍可明显看出实验组的平均扫描准确度低于对照组的路径，可作为一个特征。实验组的 "注意开关数 "明显大于对照组的平均值，可以作为一个特征。

****

图 27四个眼动数据的比较，在 "选择日期和时间 "子任务中，实验组和对照组之间没有显著差异

图 27显示，如3.2节所分析的。实验组的 "固定率 "与对照组的中位数和平均值都非常接近，说明它们之间没有明显的差异。对照组的 "在相关AOI上的第一次固定时间 "和 "一个凸壳区域 "包含了实验组的所有数据，说明它们之间没有显著差异。实验组的 "编辑距离 "与对照组的中位数非常接近，说明它们之间没有显著差异。上述表现表明，当 "易用性 "满足和不满足时，这些指标的表现没有变化，所以不能作为识别 "易用性 "满足程度的特征。

综上所述：当 "易用性 "不满足时，其特点是固定次数多，固定时间长，扫描路径准确度低，注意力转换高。因此，不满足程度的评价公式中的每个参数都要带入真实值来寻找。

对照组和实验组的 "使用便利性 "不满意度得分如下。

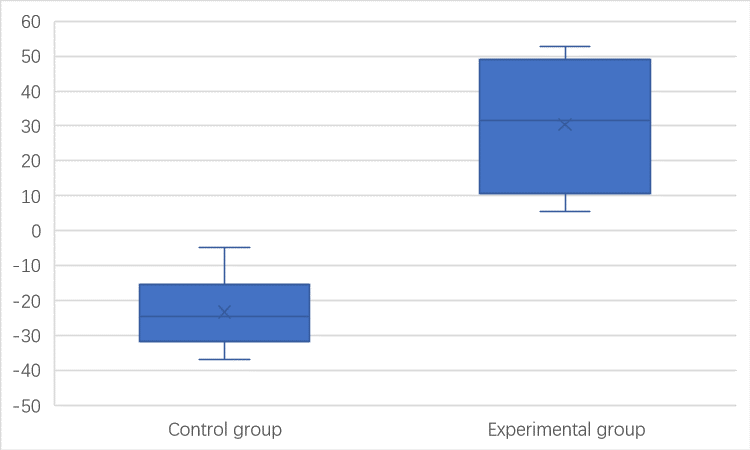


图 28实验组和对照组在 "选择日期和时间 "子任务中 "易用性 "不满意分数的比较

可以看出，"易用性 "不满意公式的程度可以用0来评价易用性是否得到满足。

### 4.4.5 设备效率

在 "提交预订信息 "子任务中，实验组和对照组的数据比较如下：。

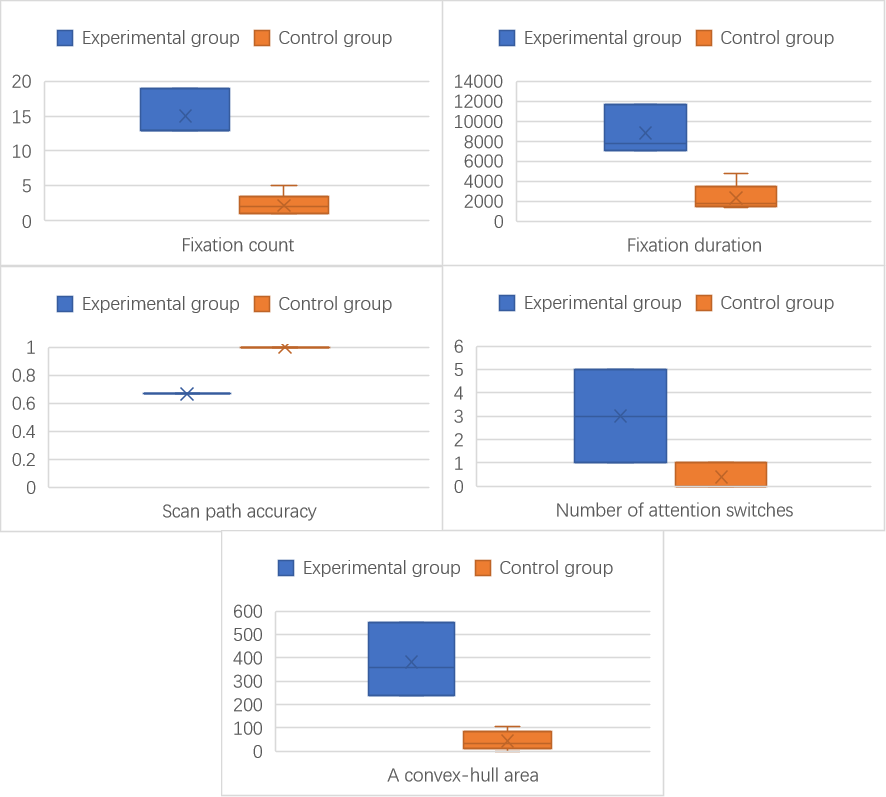


图 29在 "提交预订信息 "子任务中，实验组和对照组之间有显著差异的五个眼动跟踪数据的比较

上述的 图 29可以看作是3.2节中的分析。实验组的固定次数比对照组高，这可以作为一个特征。实验组的固定时间比对照组高，这可以作为一个特征。由于AOI区域的数量较少，扫描路径准确度相对单一，但仍能明显看出实验组的平均扫描准确度低于对照组，这可作为一个特征。实验组的注意开关绝大部分都高于对照组，这可以作为一个特征。实验组的凸壳比对照组高，这可以作为一个特征。

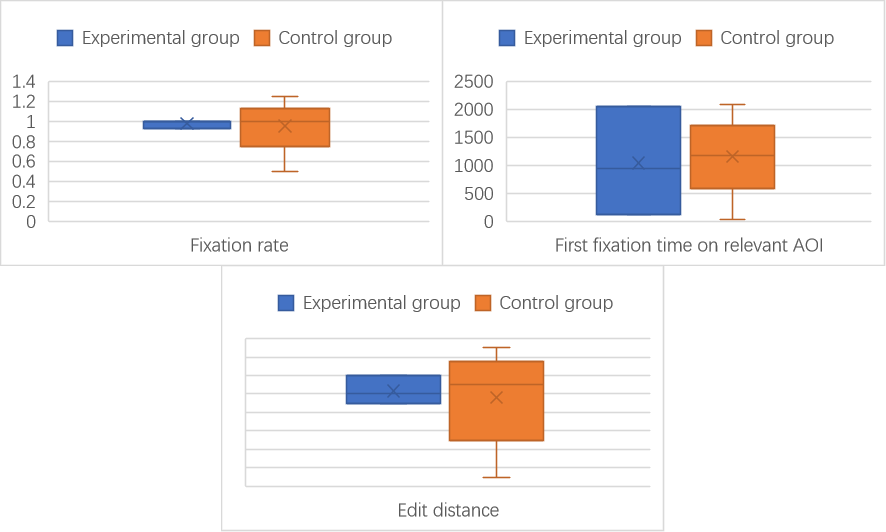


图 30在 "提交预订信息 "子任务中，实验组和对照组的三个眼动数据的比较没有显著差异

图 30显示，正如第3.2节所分析的那样。固定率"、"在相关AOI上的首次固定时间"、"编辑距离 "这三个指标都显示，对照组的数据完全包含了实验组的所有数据，说明对照组和实验组之间没有明显差异。上述表现表明，在 "设备效率 "满足和不满足时，这些指标的表现没有变化，因此不能作为识别 "设备效率 "满足程度的一个特征。

综上所述：当 "设备效率 "不满足时，其特点是固定次数多，固定时间长，扫描路径精度低，注意力转换多，凸壳面积大。因此，将各参数的不满足程度的评价公式带到真实值中去寻找。

对照组和实验组对 "设备效率 "不满意的评分，如下图所示。

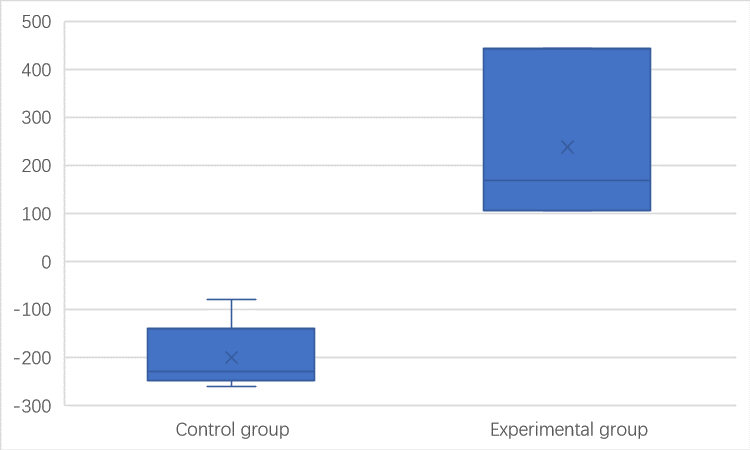


图 31实验组和对照组在 "提交预订信息 "子任务中 "设备效率 "不满意分数的比较

可以看出，不满足的公式度可以以0为界来评价是否满足 "设备效率"。

### 4.4.6 可及性

实验组和对照组在 "选择座位 "子任务中的数据比较如下。

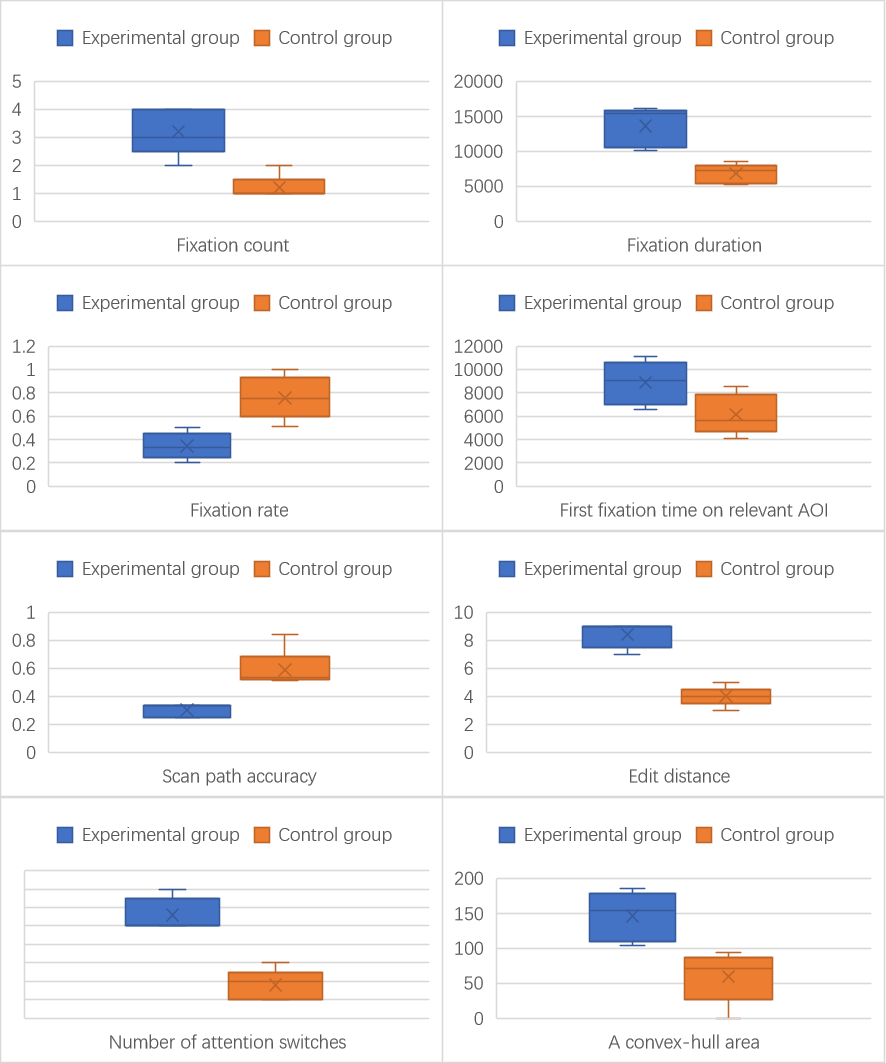


图 32在 "选择座位 "子任务中，实验组和对照组之间有显著差异的八个眼动跟踪数据的比较

中的 图 32，可以看出3.2节中的分析。实验组的固定次数比对照组高，这可以作为一个特征。实验组的固定时间高于对照组的固定时间，这可以作为一个特征。实验组的固定率绝大部分低于对照组，这可以作为特征。实验组在AOI上第一次定影的平均数和中位数都高于对照组，这可以作为一个特征。实验组的扫描路径精度低于对照组，可以作为一个特征。实验组的编辑距离高于对照组，这可以作为一个特征。实验组的注意力转换比对照组高，这可以作为一个特征。实验组的凸壳比对照组高，可以作为一个特征。

综上所述：当 "可及性 "不满足时，其特点是固定次数高、固定时间高、固定率低、在AOI上的首次固定时间高、扫描路径精度低、编辑距离高、高；因此，将不满足程度的评价公式中的各参数带到真实值中去寻找。

对照组和实验组的 "可及性 "不满意分数如下所示。

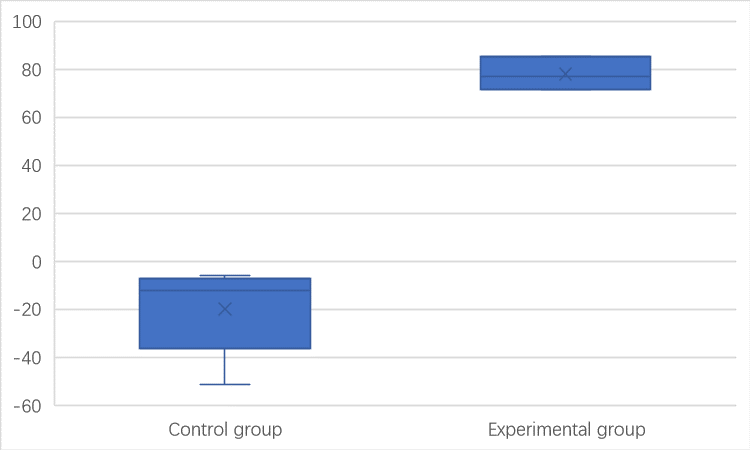


图 33实验组和对照组在 "选择座位 "子任务中的 "可及性 "不满意分数的比较

这表明，不满足的公式程度可以用0为界来评价 "可及性 "是否满足。

## 4.5 真实世界的测试

为了验证整个基于眼动跟踪的需求评估框架的可用性，上述已经被参数化的非功能需求评估公式也被包括在内。 可以实现后端分析系统。为此，我们邀请了另外两名参与者，要求他们重新进行第一组和第二组的实验。不同的是，分析系统将不只是记录眼动参数，而是直接给他们每个子任务中的非功能需求的满足度，结果如下。

表 2真实世界的测试结果

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **副任务** | **准确度得分** | **稳健性得分** | **一致性得分** | **易用性得分** | **设备效率得分** | **可访问性得分** |
| **参与者1** | 导航至预订功能页面 | -5.45 | -344.11 | -8.16 | -20.30 | -187.19 | 85.93 |
| 选择学习室场地 | -11.91 | -387.69 | -8.76 | -19.93 | -232.94 | -33.91 |
| 选择日期和时间 | -7.73 | -272.34 | -11.68 | 25.46 | -241.89 | -21.70 |
| 提交预订信息 | -10.11 | -222.22 | -11.16 | -23.98 | -191.57 | -16.25 |
| 选择座位 | -13.39 | 569.24 | -12.98 | -30.88 | -227.78 | -35.45 |
| **参与者2** | 导航至预订功能页面 | -10.31 | -307.95 | -5.33 | -24.81 | -195.94 | -16.53 |
| 选择学习室场地 | 17.93 | -293.57 | -9.88 | -31.43 | -242.98 | -20.35 |
| 选择日期和时间 | -12.34 | -257.58 | 8.94 | -17.94 | -226.52 | -35.32 |
| 提交预订信息 | -9.18 | -239.60 | -12.35 | -21.59 | 365.21 | -32.22 |
| 选择座位 | -6.25 | -339.58 | -8.21 | -22.42 | -199.99 | -15.52 |

如表中 所示 表 2，在开启了非功能需求的故意不良设计的子任务中，基于Eye-tracking的需求评估框架可以对未满足的非功能需求产生正确的输出，而其他非功能需求则不会被误报。可以看出，基于眼动跟踪的需求评估框架可以在实时环境中检测非功能需求的满足程度。

**第五章：结论和未来工作**

寻找从用户那里提取有价值的非功能需求信息的方法是最重要的，也是研究人员仍然感兴趣的。非功能属性是整个系统的属性，无法对单个组件进行验证。在随后的系统维护中，对这些功能的检测和修改会导致高成本。目前还没有明确和全面的解决方案来评估非功能需求的满足情况。

本论文提出了可以有效区分非功能需求的眼球运动指标，并使用几个非功能需求的例子来证明系统的工作原理。此外，本论文提出的方法主要是提供一种直接观察用户的方式，可以用来描述和告知非功能需求的研究。这篇论文提出了一个综合的、结构化的方法的起点。论文研究了一种新的非功能需求分析方法，它探索了眼球运动模式，以自动收集用户的隐性反馈。论文根据现有研究的认知和经验证据，建立了眼动和用户感知之间的语义联系。它进一步将几种眼动模式形式化，并提供了它们与非功能需求之间的对应关系。此外，它还根据实验数据总结和量化了评价标准的六个标准非功能要求。

建议的方法提供了不满足非功能要求的精确位置实现的细节，并降低了系统的修复和增强的成本。

尽管本论文提出了有效的方法，但由于COVID-19疫情的影响，本研究招募的实验参与者数量不多。作为第一项将眼动跟踪与需求分析相结合的工作，本研究获得的结果足以作为后续工作的灵感，但由于样本较小，很难将结果作为一般化的结论，直接用于广泛的实际工作中。在未来的工作中，我们希望招募更多的参与者，以便结论可以得到更有力的验证，并且可以分析更多的非功能需求。

在未来的工作中，我们将给出方法来确定眼动跟踪指标的每个维度的定量分析，以兼容更多的非功能需求分类。我们还打算在开源社区验证所提出的技术，结合真实的项目与检测系统和实验来确定潜在的监督问题和对我们方法的改进。这篇论文是一个综合和结构化方法的起点。长期的愿景是建立一个系统的、高度符合要求的、可定制的评级系统，结合适当的开发技术，可以促进中小开发者的发展，并为更杰出的开发者提供新的灵感。

此外，NFRs可能是相互关联的。这也意味着，在现实世界的应用中，它们可能会相互影响，以至于它们不能从用户那里得到单独的反馈。这是一个客观存在的问题，有很多研究都关注NFRs之间的相互作用，但这不在我们的研究范围之内。在这项工作中，我们试图根据NFRs的定义，构建最明显和最简单的需求场景，将指标和NFRs独立联系起来。在我们后续的工作中，我们将尝试考虑NFRs之间的关系，并通过结合相关主题的研究成果来完善映射关系。

对于更多的应用场景。这个项目使用的硬件设备，本质上只是一个带红外补光灯的相机传感器，以获得所有需要的参数。这意味着该框架可以部署在大量的设备上（带有Windows Hello人脸认证的PC，带有Face ID的苹果移动设备和带有人脸认证的安卓移动设备）。在充分遵守开放性和自愿性原则，并适当处理隐私法规的情况下，我们可以期待一种能从众多远程未知用户隐性获得反馈的方法。这将为软件系统的用户需求反馈提供一种新的可能性。

**参考资料**

[1] SOMMERVILLE I, SAWYER P. Requirements Engineering: A Good Practice Guide[M]. 1st 版. USA: John Wiley & Sons, Inc., 1997.

[2] MAIRIZA D, ZOWGHI D, NURMULIANI N. An investigation into the notion of non-functional requirements[C]//Proceedings of the 2010 ACM Symposium on Applied Computing. New York, NY, USA: Association for Computing Machinery, 2010: 311-317[2022-05-05]. https://doi.org/10.1145/1774088.1774153. DOI:10.1145/1774088.1774153.

[3] CHUNG L, LEITE J C S do P. On Non-Functional Requirements in Software Engineering[C]. 2009.

[4] MYLOPOULOS J, CHUNG L, NIXON B A. Representing and using nonfunctional requirements: a process-oriented approach[J]. IEEE Transactions on Software Engineering, 1992, 18: 483-497.

[5] MAALEJ W, KURTANOVIĆ Z, NABIL H, 等. On the automatic classification of app reviews[J]. Requirements Engineering, 2016, 21(3): 311-331. DOI:10.1007/s00766-016-0251-9.

[6] GUZMAN E, MAALEJ W. How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews[C]//2014 IEEE 22nd International Requirements Engineering Conference (RE). 2014: 153-162. DOI:10.1109/RE.2014.6912257.

[7] PANICHELLA S, DI SORBO A, GUZMAN E, 等. How can i improve my app? Classifying user reviews for software maintenance and evolution[C]//2015 IEEE International Conference on Software Maintenance and Evolution (ICSME). 2015: 281-290. DOI:10.1109/ICSM.2015.7332474.

[8] BINKLEY D, DAVIS M H, LAWRIE D, 等. The impact of identifier style on effort and comprehension[J]. Empirical Software Engineering, 2013, 18: 219-276.

[9] SUN P, YANG J, MING H, 等. A Multi-layered Desires Based Framework to Detect Users’ Evolving Non-functional Requirements[C]//2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC): 卷 01. 2018: 28-37. DOI:10.1109/COMPSAC.2018.00013.

[10] XIE H, YANG J, CHANG C K, 等. A statistical analysis approach to predict user’s changing requirements for software service evolution[J]. Journal of Systems and Software, 2017, 132: 147-164. DOI:10.1016/j.jss.2017.06.071.

[11] PAGANO D, MAALEJ W. User feedback in the appstore: An empirical study[C]//2013 21st IEEE International Requirements Engineering Conference (RE). 2013: 125-134. DOI:10.1109/RE.2013.6636712.

[12] WANG T, LIANG P, LU M. What Aspects Do Non-Functional Requirements in App User Reviews Describe? An Exploratory and Comparative Study[C]//2018 25th Asia-Pacific Software Engineering Conference (APSEC). 2018: 494-503. DOI:10.1109/APSEC.2018.00064.

[13] RAYNER K. Eye movements in reading and information processing[J]. Psychological Bulletin, 1978, 85(3): 618-660. DOI:10.1037/0033-2909.85.3.618.

[14] ANDREW T. D. Eye Tracking Methodology[M]. Springer Cham, 2017[2022-06-01]. https://link.springer.com/book/10.1007/978-3-319-57883-5.

[15] Eyetracking: Is It Worth It? :: UXmatters[EB]. [2022-06-01]. https://www.uxmatters.com/mt/archives/2009/10/eyetracking-is-it-worth-it.php/.

[16] A Survey on the Usage of Eye-Tracking in Computer Programming | ACM Computing Surveys[EB]. [2022-06-01]. https://dl.acm.org/doi/abs/10.1145/3145904.

[17] SHARAFI Z, SOH Z, GUÉHÉNEUC Y G. A systematic literature review on the usage of eye-tracking in software engineering[J]. Information and Software Technology, 2015, 67: 79-107. DOI:10.1016/j.infsof.2015.06.008.

[18] JACOB R J K, KARN K S. Commentary on Section 4 - Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises[M]//HYÖNÄ J, RADACH R, DEUBEL H. The Mind’s Eye. Amsterdam: North-Holland, 2003: 573-605[2022-06-01]. https://www.sciencedirect.com/science/article/pii/B9780444510204500311. DOI:10.1016/B978-044451020-4/50031-1.

[19] KARN K S, ELLIS S, JULIANO C. The hunt for usability: tracking eye movements[C]//CHI ’99 Extended Abstracts on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery, 1999: 173[2022-05-31]. https://doi.org/10.1145/632716.632823. DOI:10.1145/632716.632823.

[20] SHARAFI Z, SHARIF B, GUÉHÉNEUC Y G, 等. A practical guide on conducting eye tracking studies in software engineering[J]. Empirical Software Engineering, 2020, 25: 3128-3174.

[21] GLINZ M. On Non-Functional Requirements[C]//15th IEEE International Requirements Engineering Conference (RE 2007). 2007: 21-26. DOI:10.1109/RE.2007.45.

[22] DOERR J, KERKOW D, KOENIG T, 等. Non-functional requirements in industry - three case studies adopting an experience-based NFR method[C]//13th IEEE International Conference on Requirements Engineering (RE’05). 2005: 373-382. DOI:10.1109/RE.2005.47.

[23] CASTRO C, FANTINATO M. Dictionary of Non-functional Requirements of Business Process and Web Services[M]. Sao Paulo, SP. Brazil.: University of São Paulo, School of Arts, Sciences and Humanities, 2018.

[24] WANG C C, HUNG J C, HUANG C H, 等. Advertising Visual Attention to Facebook Social Network: Evidence from Eye Movements[C]//2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI). 2018: 68-73. DOI:10.1109/IIAI-AAI.2018.00023.

[25] SHARAFI Z, SHAFFER T, SHARIF B, 等. Eye-Tracking Metrics in Software Engineering[C]//2015 Asia-Pacific Software Engineering Conference (APSEC). 2015: 96-103. DOI:10.1109/APSEC.2015.53.

[26] SALVUCCI D D, GOLDBERG J H. Identifying fixations and saccades in eye-tracking protocols[C]//Proceedings of the 2000 symposium on Eye tracking research & applications. New York, NY, USA: Association for Computing Machinery, 2000: 71-78[2022-06-08]. https://doi.org/10.1145/355017.355028. DOI:10.1145/355017.355028.

[27] POOLE A, BALL L J, PHILLIPS P. In Search of Salience: A Response-time and Eye-movement Analysis of Bookmark Recognition[C]//FINCHER S, MARKOPOULOS P, MOORE D, 等. People and Computers XVIII — Design for Life. London: Springer, 2005: 363-378. DOI:10.1007/1-84628-062-1\_23.

[28] BUSJAHN T, BEDNARIK R, BEGEL A, 等. Eye Movements in Code Reading: Relaxing the Linear Order[C]//2015 IEEE 23rd International Conference on Program Comprehension. 2015: 255-265. DOI:10.1109/ICPC.2015.36.

[29] PETRUSEL R, MENDLING J. Eye-Tracking the Factors of Process Model Comprehension Tasks[C]//SALINESI C, NORRIE M C, PASTOR Ó. Advanced Information Systems Engineering. Berlin, Heidelberg: Springer, 2013: 224-239. DOI:10.1007/978-3-642-38709-8\_15.

[30] GOLDBERG J H, KOTVAL X P. Computer interface evaluation using eye movements: methods and constructs[J]. International Journal of Industrial Ergonomics, 1999, 24(6): 631-645. DOI:10.1016/S0169-8141(98)00068-7.

[31] BHATTARAI R, PHOTHISONOTHAI M. Eye-Tracking Based Visualizations and Metrics Analysis for Individual Eye Movement Patterns[C]//2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE). 2019: 381-384. DOI:10.1109/JCSSE.2019.8864156.

**鸣谢**

在本论文的整个写作过程中，我得到了大量的支持和帮助。

我首先要感谢我的导师李彤，他非常有远见，在本项目的研究方向上发现了相关领域的一个空白。在一年多的研究过程中，他用他丰富的专业知识打消了我所有的疑虑，并在明确研究问题和方法方面给予我宝贵的帮助。你富有洞察力的反馈促使我改进了我的思路，使我的工作更上一层楼。

接下来我最想感谢的人是我们的辅导员戴璐。她以热情的工作态度，致力于解决我们在学校生活中的所有问题。她的存在使我们能够毫无顾虑地专注于我们的课程和项目。她不仅是我们的学业指导和组织者，也是我们的生活伴侣，愿意理解和解决我们的任何想法。我无法想象没有她的大学生活，那一定比我们现在经历的要艰难好几倍。

最后，我想感谢我亲切的同学们。作为一个在生活中经常粗心大意的人，你们总是毫不吝啬地提供信息和帮助，与你们交流总是让我知道自己的位置。我很荣幸能与你们这些优秀的人一起工作、学习、生活和分享四年的时光。

我还要感谢范公修荣誉学院在整个大学期间对我的支持。学院通过专门的课程和大量的项目培养了我的研究能力。更重要的是，学院给了我们一个广阔的平台，让我们完全自由地选择自己感兴趣的方向，并一直慷慨地资助我们的研究。在学院的教育为我建立了信心，使我在以后继续从事研究工作。