

# 基于眼动追踪实验的数字界面质量评价模型

陈晓皎<sup>1</sup>, 薛澄岐, 田静, 邵将

(东南大学机械工程学院,南京 211189)

**摘要:**本文针对现有界面评估数据主观性方面存在不足的问题,在眼动追踪技术的基础上,借鉴软件程序领域质量模型的构建思路,提出了基于眼动追踪技术的数字界面质量评估模型。模型以眼动追踪技术相关界面评估的指标作为质量评价模型的度量标准,从人、界面、交互、设计和情感的角度,把界面设计的基本元素、界面信息数量、界面布局、完成时间、操作准确率、认知资源投入等界面设计因素作为质量评价模型的质量要素,以实现界面功能的情感吸引性、设计复杂性、交互高效性、认知负荷性为质量实现目标,建立了全新的界面评估映射模型,提高了界面评估结果的可靠性。

关键词: 数字界面; 眼动追踪; 评价模型;

中图分类号: TP301 文献标志码: 文章编号:

**Abstract:** In view of the problem that the interface evaluation data is mostly subjectivity, we firstly proposed the digital interface quality evaluation model based on eye tracking technology, with the reference of software quality model. The standards of eye tracking technology index acted as the measure of quality evaluation model. From the perspectives of interface, interaction, design and emotion, we combined Interface design factors, such as the interface basic elements, interface information quantity, interface layout, accuracy and interface cognitive resources factors, into the quality evaluation model of quality elements .We established a new user interface evaluation mapping model to meet the functional requirements of interface emotional attraction, design complexity, interaction and cognitive load. We proposed a new idea for interface evaluation to avoid the interface of subjective evaluation

Key words: Digital Interface; Eye-tracking; Evaluation Model

数字界面作为信息化时代传递信息的重要载体, 用户通过数字界面获取信息进行决策判断。信息传递的精确、高效往往引起用户认知差异以及决策判断的失误,

---

作者简介: 陈晓皎 (1988-) 女, 博士生; 薛澄岐 (联系人), 男, 教授, ipd\_xcq@seu.edu.cn

基金项目: 面向大数据的信息可视化设计方法研究 (71471037), 数字界面视觉信息认知的脑机制研究 (71271053), 基于信息流和感知分层的大数据信息可视化设计方法研究 (KYLX15\_0061)

会导致操作事故。基于此问题的存在，对界面进行有效评估，实现界面的优化设计，降低界面不合理性概率，具有重要的作用。

目前国内外已经运用多种界面评估手段对数字界面进行有效评估，主要有主观评价法、绩效评估法、综合理论评估法、数学模型评估法和生理实验测评法。前四种评估方法存在主观判断因素，在样本采集数量不充分的情况下，具有一定的主观意识偏差，在数字界面评估中往往被当作辅助评价方法。生理实验评估方法是运用眼动跟踪技术和脑电技术的生理指标对界面开展的客观数据评估。特别是眼动跟踪技术对于复杂数字界面的布局的合理性有着较为客观的评估。自1950 年Fitts、Milton 和Jones开展了飞机驾驶舱界面设计的眼动评估研究以来<sup>[1]</sup>，2001 年Cowen 对网页界面设计的布局合理性进行了眼动评估研究<sup>[2]</sup>。2008 年Weinreich等人对企业网站和搜索引擎界面的用户关注热区进行了眼动研究<sup>[3]</sup>，2010 年东南大学薛澄岐、王海燕、刘青等人，开展了航天界面、地铁站台界面的眼动评估研究<sup>[4-5]</sup>。总的来说，这些研究着眼于利用个别生理指标对界面进行评估，不能将评估指标与界面功能元素指标联系起来，在评估全面性方面存在不足。本文将眼动追踪实验的技术指标作为度量标准，借鉴软件程序领域质量模型的构建思路，去分析影响数字界面质量的多种界面元素，以实现界面多重功能为目标去构建评价模型，本文构建的数字界面质量评估模型是对现有研究的补充和完善。

## 1 数字界面质量眼动评价模型构建

### 1.1 眼动追踪评价指标

生理实验测评法可以解决以往主观评价法、绩效评估法、综合理论评估法、数学模型评估法的主观因素制约和误差。眼动追踪技术作为目前应用最为广泛的界面评估方法，主要从剖析瞳孔直径，注视点热区，注视点次数，注视点序列等实验数据，对数字界面进行较为客观、科学的评估。以此，眼动追踪技术是本文建立数字界面质量评价模型的实验途径<sup>[6-7]</sup>。

1) 瞳孔直径。瞳孔直径作为用户视觉信息加工时的负荷指标之一人在进行视觉信息加工时，瞳孔直径的变化幅度与进行信息加工的认知努力程度密切相关。当认知负荷较大时，瞳孔直径会扩大，反之认知负荷变小时，瞳孔直径也会相应变小。

2) 注视热点。直观反应受试者在哪个区域的停留的时间长短，注视点的空间分布，并用颜色标注出注视时间在此区域的时间长短。此评价指标能较为直观的反应受试者对界面信息的视觉兴趣的关注区域。界面中重要信息的突出显示，是否存在信息干扰设计可直观关联此指标。

3)注视点数目。与视觉搜索的绩效成反比，注视点数目越多表明视觉搜索的效率越低，注视点数目越少表明搜索的效率越高。受试者对界面的信息获取时，以及对信息获取后的决策判断的效率都与此指标相关。

4)注视点序列。跟踪受试者眼球的转动，依次标记出注视点在注视兴趣之间的位置转换、停留时间，通过移动顺序和路线描绘。此指标用来度量界面布局的有效性以及层级信息的合理性。

## 1.2 质量评价模型构建

现有的研究一般是从宏观用户需求、产品属性、产品质量特性度量来构建整个质量评价框架<sup>[8-9]</sup>。将该构建逻辑引入到数字界面的评估体系中，以眼动实验评价指标作为底层度量标准，以界面元素特性作为中层质量子要素，以界面功能特性作为顶层质量要素。数字界面质量的眼动评价模型构建思路如图1所示：

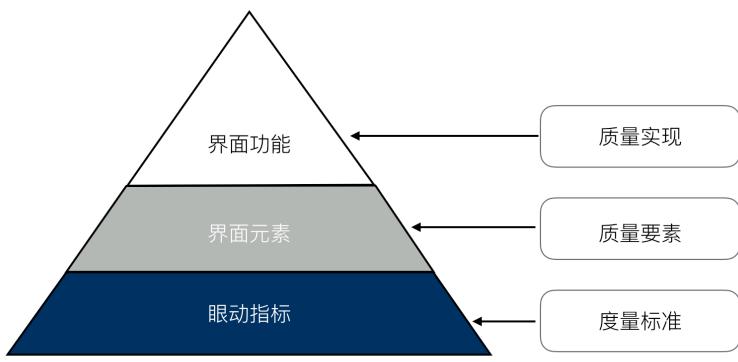


图1 数字界面质量的眼动评价模型构建思路

根据眼动实验评价指标分析，将瞳孔直径、注视点热区、注视点数目、注视点序列、平均注视时间作为底层度量标准；提炼界面设计的基本元素图形、控件、字体和色彩，界面信息的数量和界面布局，界面的完成时间和操作正确率，认知资源投入量作为界面的质量要素，把情感吸引性、设计复杂性、交互高效性和认知负荷作为界面质量实现的评测目标，建立数字界面质量的眼动评价模型，如图2所示。

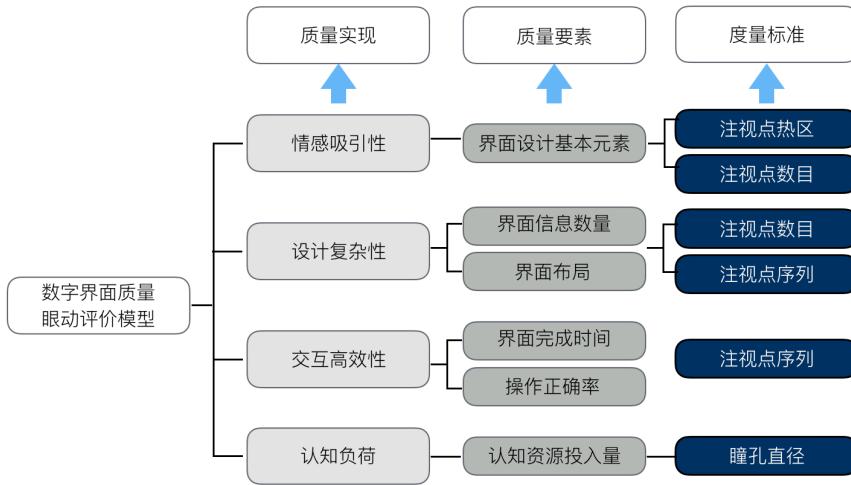


图2 数字界面质量眼动评价模型

1) 情感吸引性。本文从数字界面对用户的吸引性来反映用户对于界面的情感因素，这是从界面功能实现的情感吸引性出发，把界面的图形、控件、字体、色彩等界面设计的基本元素作为评价模型的质量要素。

2) 设计复杂性。从界面设计的复杂性出发，将信息数量和界面布局也归纳为实现评估界面模型的质量要素。设计复杂性的计算公式<sup>[10]</sup>如式1所示：

$$C = -N \sum_{i=1}^n P_i \log P_i \quad (1)$$

式中：  $C$  表示复杂性，  $N$  表示所有组件的数目、  $n$  表示组件的种类、  $P_i$  表示第  $i$  类组件出现的概率。由公式1可以看出复杂性与组件数目，组件种类和  $P_i$  相关，  $N$  越大，  $n$  越大，  $P_i$  越大， 对象设计越复杂。

3) 交互高效性。人机界面交互是指人和界面之间进行信息沟通的一个双向过程。人机界面通过交互反映界面设计的高效性，完成时间和正确率作为表征反映。完成时间代表用户交互操作完成所需的时间，交互的高效性通过正确率数据作为反映，因此，界面的交互高效性总结为完成时间和正确率两个质量要素。完成时间可以利用完成任务的眼动数据指标所需平均注视时间来测量，正确率是用眼动数据指标的正确率来测量。

4) 认知负荷。在用户与界面产生交互的过程中，各种认知加工活动，都需要认知资源的投入，认知资源的投入比例与界面设计给用户的认知负荷是息息相关的。在本文中采用眼动追踪的生理测量方法来开展认知负荷的测量。眼动数据指标中的瞳孔直径与认知负荷有直接表象联系。当认知负荷较大时，瞳孔直径会扩大，反之认知负荷变小时，瞳孔直径也会相应变小。

## 2 基于界面质量模型的眼动实验

### 2.1 实验材料

选取F-18 战斗机显控界面作为实验素材样本，主要是考虑到此座舱显控界面的数据完整性可作为界面评估的典型<sup>[11]</sup>。本次实验选取的实验图片尺寸为 1024\*768 px，格式为 BMP 的图片，位深度为 16，以符合眼动仪屏幕的显示需求。

本次将对 F-18 战斗机显控界面的两个设计方案开展眼动实验，按照数字界面质量眼动评价模型中，选取度量标准一眼动追踪实验的评价指标瞳孔直径、注视点热区、注视点数目、注视点序列，意在测试数字界面质量眼动评价模型的质量要素，评估数字界面质量眼动评价模型的评测目标的有效性。

### 2.1 实验准备

本次实验中，选择被试为东南大学和南京大学的本科和研究生，共15人。其中主要考虑飞行员大多为男性，因此被试选取男生较多，男生10人，女生5人。拥有本科学历1人，硕士学历10人，博士学历4人，年龄在23-27岁之间。被试人群无色弱或色盲现象。实验设备为德国SMI公司的iView X系列Hi-Speed型高精度眼动仪，包括iView PC 测试计算机、Stimulus PC 图像显示计算机和测试托架。

### 2.2 实验流程

本次实验的任务包括观察任务和检查任务两个方面。观察任务是让被试按照自己对界面的理解去观察界面，设定界面呈现时间2000ms，中间空屏呈现1000ms。观察任务的实验流程如图3所示：

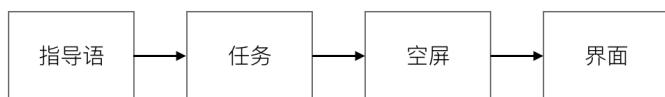


图 3 观测任务实验流程图

检查任务是让被试检查界面的引擎界面是否处于正常状态，并出声回答是否正常，按任意键结束。检查任务实现流程如图4所示。



图 4 检查任务实验流程图

## 3 眼动实验数据分析

### 3.1 注视点数目

根据数字界面质量眼动评价模型的质量实现层级之情感吸引性，统计用户的注视点热区和注视点数目，划分界面的兴趣关注区域AOI。界面1兴趣关注区域AOI的划分如图5所示，界面2兴趣关注区域AOI的划分如图6所示：



图 5 界面 1 兴趣关注区域 AOI 的划分

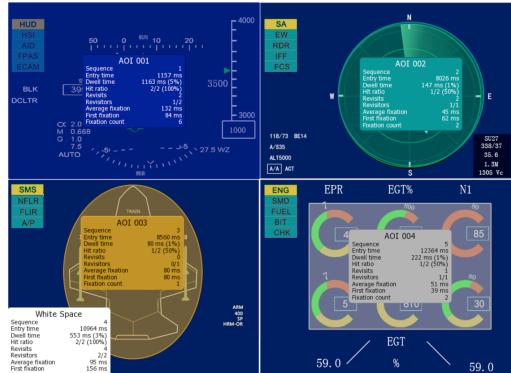


图 6 界面 2 兴趣关注区域 AOI 的划分

15个被试查看界面1、界面2时在兴趣关注区域AOI的注视点数目的分析结果如图7所示。黑色折线表示界面1的兴趣关注区域的注视点数目，红色折线表示界面2的兴趣关注区域的注视点数目。

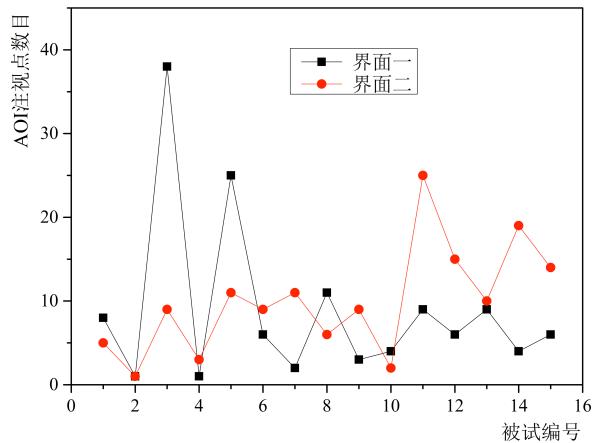


图 7 AOI 区域注视点数目对比

两组数据的统计指标如表1所示。通过对该组数据的统计分析，15个被试在界面1的兴趣关注区AOI的注视点数目平均值为8.67，标准差是10.01，15个被试在界面2的兴趣关注区的注视点数目平均值为9.87，标准差是6.35，如表所示。根据数字界面质量眼动评价模型，在兴趣关注区域的注视点数目多则吸引性好，因此界面2在情感吸引性角度优于界面1。

表 1 兴趣关注区域 AOI 的注视点数目分析

主界面方案	样本容量(人)	均值(个)	标准差
1	15	8.666667	10.0119
2	15	9.866667	6.356849

## 5.2 注视点序列

15名被试在实验过程中对界面1、界面2进行操作，通过实验分析，被试的眼动数据可以呈现，注视点序列是否按照界面既定任务的操作流程的轨迹，同时可以考察被试是否理解实验任务，按照实验任务的指导语进行实验任务。因为被试人数较多，本文不可能一一呈现所有被试的注视点序列图，经对比分析后，挑选了被试6的实验数据作为例证分析。被试6在界面1与界面2的注视点序列如图8与图9所示：

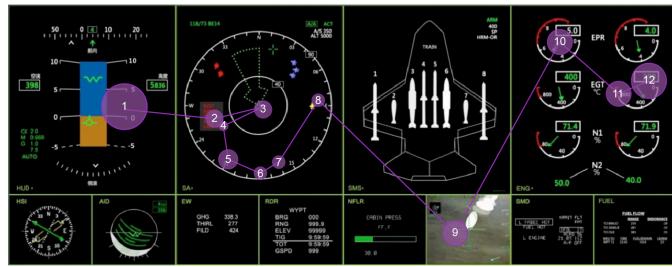


图 8 被试 6 在界面 1 的注视点序列

图 9 被试 6 在界面 2 的注视点序列

如图所示，被试6在界面1中的注视点序列路径比界面2的注视点序列复杂，有迂回路径，且注视点数目多于界面2。因此，根据数字界面质量眼动评价模型，界面的注视点序列清晰明确，且注视点数目较少，则界面布局较好，因此界面2在设计复杂性角度优于界面1，界面2的交互高效性优于界面1。

## 5.3 瞳孔直径分析

15个被试在完成界面1和界面2检查任务时的瞳孔直径平均值如图10所示：

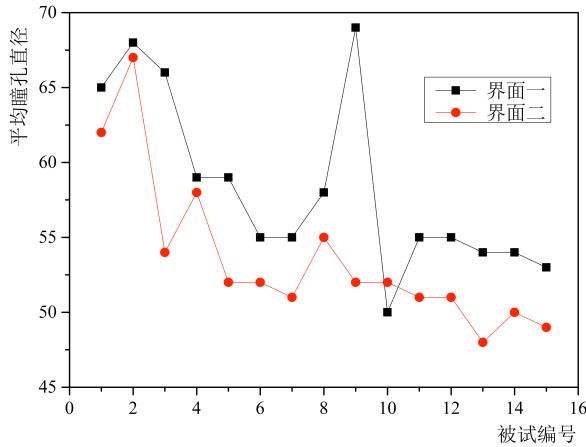


图 10 被试瞳孔平均直径分布图

图中黑色折线表示15个被试检查界面1时的瞳孔直径平均值为58.15，标准差为5.25；红色折线表示15个被试检查界面2时的瞳孔直径平均值为54.42，标准差为4.87。具体数值如表2所示：

表 2 瞳孔直径数据分析

主界面方案	样本容量(人)	均值(个)	标准差
1	15	58.15377	5.256566
2	15	54.42476	4.872989

根据数字界面质量眼动评价模型，完成界面任务时，被试的瞳孔直径较大其对于界面的认知负荷则越大。因此，根据数字界面质量眼动评价模型的质量实现层级之认知负荷性，界面1需要用户投入的认知资源量大于界面2所投入的认知资源量，界面2的认知负荷优于界面1。

#### 4 结论

本文将在现有眼动追踪技术的基础上构建了数字界面质量评价模型，从眼动实验的数据里提取了注视点数目、注视点序列、注视点热区、瞳孔直径等技术指标，建立评价模型的度量标准；从用户认知资源投入量、界面元素、界面布局、界面完成时间和操作正确率的角度建立模型的质量要素；在度量标准与界质量要素的映射关系基础之上，建立评价模型顶层的质量实现目标，作为数字界面质量评价模型的评估界面评测目标。并通过基于任务的战斗机显控界面实验，最终实现对战斗机显控界面客观综合的评估，以此来验证此数字界面质量评价模型的有效性。

## 5 参考文献

- [1] Fitts P., Jones R., Milton J. Eye movements of aircraft pilots during instrument-landing approaches [J]. Aeronautical Engineering Review, 1950(9). 24-29.
- [2] Laura Cowen. An eye movement analysis of web-page usability [D]. UK, Lancaster: Lancaster University, 2001.
- [3] Weinreich, H., Obendorf, H., Herder, E., Mayer, M. Not Quite the Average: An Empirical Study of Web Use [J]. ACM Transactions, 2008, 2(1). 22-25.
- [4] 刘青, 薛澄岐. 基于眼动追踪技术的界面可用性评估[J]. 东南大学学报(自然科学版), 2010, 40(2). 332-334.
- [5] 牛亚峰, 薛澄岐, 王海燕, 李晶. 复杂系统数字界面中认知负荷的脑机制研究[J]. 工业工程与管理, 2012, 17(6). 72-75.
- [6] 胡凤培, 韩建立, 葛列众. 眼部跟踪和可用性测试研究综述[J]. 人类工效学, 2005, 11(2). 52-54.
- [7] 康卫勇, 袁修干, 柳忠起, 刘伟. 飞机座舱视觉显示界面脑力负荷综合评价方法 [J]. 航天医学与医学工程, 2008, 21(2). 104-107.
- [8] 孙洋, 袁玉宇. ISO/IEC 25010 质量模型标准现状[J]. 信息技术与标准化, 2008 (11). 33-36.
- [9] 杨俊, 景疆. 软件质量度量技术浅谈[J]. 信息技术, 2005, 29(11). 150-152.
- [10] Comber T, Maltby J R. Screen complexity and user design preferences in windows applications[J]. 1994.
- [11] 卞婷, 薛澄岐, 仇岑. 涡轮喷气发动机状态显示生态界面设计[J]. 机械设计与制造, 2010. 9-11.