# 网页美感，眼动一致性和眼动实验数据的新可视化指标

>>>>>>>>>>颜色标注说明

红色标出的是有待补充或考证的

图片image 外部文件为jpg格式

图表graph 外部文件为fig格式

表格table 外部文件为excel

公式formula 暂无

<<<<<<<<<<<<<<

## - introduction

眼动检测技术在今天已经不算什么新鲜的话题了，在各类平面、网页乃至产品的设计中，它已经有了相当长一段时间的应用。很大一部分的眼动研究旨在评估图像上的目标区域对用户的注意力的吸引能力，从而确定一个设计的有效性，商业价值等。也有一部分的研究聚焦在行为学的研究上。

现今用于分析或是可视化眼动实验数据结果的指标有如下：

heatmap：一种评估画面上不同区域所受到眼动聚焦强度的二维图像；

【1-1热图例图】

bee swarm，一段沿时间播放的表现每个时刻点每个用户所注视的位置的视频；

【1-2 bee swarm例图】

gaze plot，一种把所有用户的眼动转移路径通过圆点和直线标注出来的复杂图像；

【1-3 gazeplot例图】

AOI cluster，一种对用户的眼动数据进行空间上的聚类，从而在二维图像上标注出若干个视觉重点区块的技术。

【1-4 AOI热图】

针对上述的可视化，尤其是AOI，又有相当多的指标来评估眼动的结果。以一个聚类得到的AOI为例，常见的指标有用户平均首次聚焦到AOI的时刻，在AOI上的平均停留时长，最长单次停留时长，首次停留时长，注视次数等，在此就不一一展开了。

不难发现，这些常见的眼动指标大都旨在通过不同的维度表现眼动的一种“强度”属性，可能是沿时间来展开，如首达时间、持续时间，可能是沿空间来展开如热图，AOI。我们的实验旨在研究一群用户的眼动行为，尤其是这些行为的一致性，与美感之间的联系。现有的可视化和指标在这方面，尤其是表现一致性方面，可能略显力不从心。因而本文中，我们将一定程度上基于现有的眼动指标，探讨优化和改进他们的方法，并提出一些新的指标，来考察他们与网页美感的联系，从而试图探寻人类眼动行为和美感之间的关联性。

## - related works

## - hypothesis：

一，猜想：一个公认好看的网页应该具有对视觉的较强的引导作用，引导被试以较小的眼动路径选择的代价来获取较多的页面信息量，从而在实验数据上，使得浏览者各自的浏览轨迹之间表现出更强的一致性（相对于较差的页面）。

\*需要注意的是，我们这里所讨论的一致性指的是个体之间的一致性而非单个个体多次浏览间的一致性。 后者由于存在印象残留等干扰因素，在本实验中不予探讨。

为了评估这种一致性，我们提出如下的眼动熵的概念。

二，概念： 眼动熵。现有的对于熵的定义包括热力学和信息论，概念上，熵表达一个系统的有序度、混乱度、或者反过来，纯度，一致度。

与此对应的，当我们把熵的概念应用在一个群体的被试者对同一个网页对象的观察所产生的一系列眼动行为上时，眼动熵的概念表达为这些被试者的眼动行为在时空上的一致性。越大的熵代表越弱的一致性，越小的熵代表越强的一致性。

在此，我们试图以基于概率散布的香侬信息熵为基础，来计算这种眼动的一致性。下文中我们将通过充分的实验数据分析，尝试定义能反映上述视觉引导作用的眼动熵的合理表达，分析其与美感的关系和其美学理论的合理性。

## - experiment：

一，网页的收集：由于眼动实验的数据规模难以做到很大，我们挑选已经具有先验评价的网页以增强代表性。我们对一个评选优秀网页设计的网站（\*）和一个评选丑陋网页设计的网站（www.websitesfromhell.com），从第一页开始，挑选基数序号的网页。在此过程中，我们滤除了包含具有非拉丁字母字符和具有高识别度的商标、符号或人脸的网站以减少非形式层面的干扰因素。如此直到获得来自上述两个网站的各20张网页。对这些网页我们统一在1280\*800的分辨率（\*一方面我们使用的眼动仪Tobii T50是1280\*1024的，一方面为了照顾到被试常用的横屏， 我们选择了这个分辨率作为实验分辨率）下截屏为图片（包含了使用的浏览器窗口），用作实验用的刺激物。这些刺激物（stimuli）如下。

【image 4 - 1】好的所有网页

【image 4 - 2】差的所有网页

二，被试的邀请：一共有30名被试（13男，17女）参与了眼动的实验，这些人全部都是来自上海交通大学的学生或老师。他们大都来自上海交通大学媒体与设计学院。他们来自中国的各地以及韩国。

三，实验环境：实验在上海交通大学的IXD交互设计实验室内进行。实验过程中室内没有除了实验操作者和被试者之外的人员以及噪声的干扰，房间的窗帘被拉上以避免不可控的光源和反光因素，实验用的眼动仪被布置在一面纯净的白墙前，以减少背景对用户注意力的分散。

四，实验设备：我们采用的实验设备为一台Tobii T50眼动仪，配合Tobii Studio 9软件使用。该眼动仪的分辨率为1280\*1024，如之前所描述，我们的实验页面采用1280\*800的尺寸。

五，眼动实验：对于每个被试，先进行的是眼动实验。所有被试都被保证是第一次浏览实验的网页，并被告知以平时放松地在网上闲逛的心态去浏览这些页面。为了减少疲劳的干扰，40张实验页面被随机（对于每个被试，这样的分组都是随机产生的）分为等量的两组，分前后两次进行实验，期间有半分钟的休息时间。每次，眼动仪会先自动播放3张Dummy页面以使被试进入实验状态，其后是以随机顺序出现的20张的实验页面，每张页面显示3秒钟，之后会有1秒钟的黑屏休息。这样每次实验大概进行1分半。

整个实验过程中，被试对鼠标是具有控制的，但由于所浏览的网页本质上是一张截图，滚动和交互事件是不被允许的。

眼动实验的更多细节：被试在开始试验前会先进行由眼动仪软件设计的一些矫正记录偏差的过程，主要是被要求先后看分布在画面不同位置的几个点。同时，被试的位置会被调整以使得眼睛距离屏幕60cm左右，这使得眼动仪的记录准确度更高。较短的曝光时间（3秒）足够使我们考察早期的下意识的眼动行为同时避免了过长的实验时间带来的疲劳。

六，评分实验：为了得到较为准确的评分同时减少评分过程对眼动的干扰，评分实验被安排在眼动实验之后分开进行。在评分实验中，被试得以再次浏览一遍刚才的40张实验网站，这次没有时长限制，被试需要对每张页面给出“好看”或是“不好看”的二值化评分。

七，奖励：每个被试在完成实验后会得到20元RMB的奖励。

## - raw data collection & format：

1，评分数据的产生：作为之后评估指标与美感关联性的一个基准，每个网页被赋予class和score两个值来评价其美感的得分：

评分(score)：令“不好看”为得0分，“好看”为得1分，得到如下的每个网页的评分。代表着每个网页被被试评为好看的比例。

【表5-1\_score\_category\_sorted】见excel

【图5-1】排序网页图片

类别class：对score值>=0.5的页面，我们将它归入“好看”网页类别，剩余的归入“不好看”网页的类别。归类的结果如下，比较实验开始前的先验分类结果，可以看到是完全一致的。也就是说，参与实验的40张网页是具有公认的代表性的“好看的”或“不好看”的网页。

2，眼动数据格式：眼动数据的收集原理在眼动仪和相关软件内部发生：眼动仪以一定的频率（对于我们的设备T50而言，这个频率是50Hz）采集用户聚焦的平面坐标，并通过插值得到每个时刻的眼动速率的值，并通过一定的阈值得到基于注视的眼动数据。

基于注视的眼动数据是一般能通过眼动仪直接获得的最原始的数据。它是由一系列的注视构成的，每个注视包含了四个参数：注视开始时间，注视持续时间，注视横坐标，注视纵坐标。通俗而言，记录了一个注视发生的时间和地点。下文中我们的一切分析都仅来自于这些基于注视的原始眼动数据。

【表5-2一个被试对一个页面的眼动注视数据举例】

|  |  |  |  |
| --- | --- | --- | --- |
| start time | duration | x | y |
| 24224 | 700 | 483 | 341 |
| 24924 | 400 | 173 | 356 |
| 25324 | 350 | 58 | 351 |
| 25673 | 200 | 696 | 462 |
| 25873 | 150 | 738 | 483 |
| 26023 | 217 | 928 | 442 |
| 26240 | 300 | 561 | 239 |
| 26539 | 683 | 638 | 163 |

## - analysis

0 指标的评价手段

在开始前，我们首先需要约定初步评估一个计算指标与美感的相关性强弱的手段。这里我们采用分别基于网页美感评分和网页美感分类的两种方式来评估：

* 1. pearson系数：通过计算代表每个网页美感评分的score与指标的线性相关系数，可以快速获得一个指标是否与美感存在一定的线性相关性。

计算公式如下：

【公式】

pearson系数的绝对值越接近1代表指标与美感的关联越显著，反之越接近0则越不显著。

0.2 ANOVA显著性：通过对网页美感分类产生的好坏两个组，采用ANOVA方差分析来检验两个类别之间的差异的显著性。得到的F分布显著性置信系数代表在多大概率上可以拒绝两个组别之间存在着显著差异。

1.1简单的眼动指标

自然地，我们首先讨论可以直接快速地提取自原始眼动数据的一些度量，考察他们与美感的相关性。此处讨论到的这类指标包括眼动注视的个数，每次眼动注视的持续时长以及眼动扫掠的速率。

* + 1. 眼动注视的个数

数据表明，在同样的三秒钟的曝光时间内，不同被试对不同页面发生的眼动注视的个数之间具有较大的差异。下图展现了每个页面的眼动注视个数的分布盒图，这些页面是从左至右按照美感评分从小到大排列的。

【眼动注视的样本呈现】

图中，眼动注视大都分布在4-16之间，被试间方差较大。以每个页面所获得的眼动均值作为统计量，考察其与美感的相关系数以及方差验证，线性相关系数为0.33，anova显著性为0.0429。也即好的页面往往有更大可能获得更多的被试注视个数，但眼动注视个数对美感没有显著的推测性和关联性。

* + 1. 每个眼动注视的持续时长

每个眼动注视的持续时长同样表现出了较大的个体差异，按页面美感评分从小到大排序，每个被试的平均眼动持续时长、最大眼动持续时长和最小眼动持续时长的分布盒图分别如下

【duration平均样本呈现】

【duration最大样本呈现】

【duration最小样本呈现】

对上述三个指标作平均所获得的三组统计量都与美感缺乏显著的关联性。他们的相关系数和anova显著性分别为平均眼动持续时长的平均值（，），最大眼动持续时长的平均值（，），最小眼动持续时长的平均值（，），。

* + 1. 眼动速率

统计每次眼动注视跳转在画面上跨过的直线距离和，再除以实验时长可以得到估计的眼动平均速率。眼动速率在个体间同样有较大的差异，其每个页面的分布盒图如下：

【speed的样本呈现】

其均值与美感同样没有显著的相关性，相关系数为-0.25，anova置信系数为0.058。

通过上述的分析，可以认为眼动的注视的个数、注视时长以及速率这些简单指标的统计量对美感没有直接的推测能力。同时，它们受个人的浏览习惯影响较大，表现出了较大的个体差异性。

接下来，为了讨论更为复杂的注视的空间信息，我们需要引入信息熵的概念。信息熵是描述一个概率空间的信息混乱度的统计量，由香农与1948年首次提出。对于具有概率空间Ω = {x1, ……, xn}和概率分布P(X)的离散的随机变量X。信息熵H（希腊字母eta）定义为H(X) = -ΣP(xi)log2(xi)。H(X)越大，代表X的分布的混乱程度越高，信息量越大，反之H(X)越小则X的分布越是有序，信息量越小。

我们通过讨论眼动注视的“0阶”空间分布熵（即眼动注视的空间位置分布的一致性，不考虑眼动注视跳转）和1阶空间分布熵（也即眼动跳转产生的长度为二的序列的一致性），考察眼动注视位置的一致性与美感的关系。

* 1. 0阶空间分布熵——眼动注意熵

1.2.1定义：

考虑一个代表眼动注视位置的二维随机变量Z = (X, Y)，则眼动试验画面中的全体像素点构成它的概率空间（即每个像素点都作为眼动注视的一种可能的选择）。对于通过实验数据估计的该空间上的概率分布P(X, Y)，Z的信息熵H(Z) = Σ(-P(xi, yi)\*log(P(xi, yi))) (for all (x, y) s.t. P(x, y) != 0)

理论上，上述0阶空间分布的信息熵反映被试之间的眼动行为在空间分布上的一致性或是集聚性：当所有注视集中在一个像素上的时候它取得最小的指，当注视分散在不同空间位置的时候它取得最大的值。

1.2.2 空间概率分布P(X, Y)的估计：

概率分布P（X，Y）满足ΣP(Xi, Yi) = 1。对它的估计是该指标计算的关键。由于下面我们讨论几种获得P的估计的方法，并逐一评估基于他们的熵与美感的关联性。由于此处的熵旨在反映被试的眼动之间的一致性，故一个页面所接受到的所有注视都应被纳入计算。

为了有一个直观的大体的认识，我们用如下的实验网页页面来帮助可视化指标：

【image 6-1】举例的页面

1.2.2.1热图

一个能够直接想到的，也是现有的能够直观反映注视在画面上的空间分布的可视化方式是热图。热图是将整个眼动记录时间内的所有被试的眼动注视（fixation）以其注视时长为强度，以其注视坐标为位置叠加的一系列高斯核组成的图像。具体地，对于一个选定的标准差σ，一个持续时长为t，坐标为(x0, y0)的注视会叠加的高斯核的表达式为-1/2/r/r\*e^(-((x-x0)^2+(y-y0)^2)/2/r/r)。如此叠加一个页面所受到的所有注视之后，对得到的热图进行归一化处理，就获得了可以用于计算眼动熵的概率分布P。

【image6-2】归一化的热图举例，取不同σ（σ = 5， 20， 100）

高斯核叠加的意义在于利用较少的眼动数据来估计注视分布的概率，故其标准差的选取（即σ）应该是具有理论合理性的，应该在一定程度上反映由人类视觉中心（即黄斑）的聚焦范围、眼动注视过程中的微动范围以及眼动仪记录的系统误差综合造成的注视范围概率分布。为了获得一个合适的标准差，我们对从1px到133px的所有标准差下得到的基于热图的眼动熵都做了计算。评估他们的pearson系数和分类准确率，得到如下的图像。

【figure 6-1】传统热图的眼动熵随标准差变化

可以发现在整个标准差计算区间里，基于传统热图的眼动熵都没有表现出任何与美感的关联性。

1.2.2.2去权热图

由于空间分布熵只包含空间维度的信息，我们尝试对传统的热图做一些改进，在计算热图时去除由眼动注视的持续时长所造成的权重，给予每个注视相同的权重。 这样就得到了我们称之为去权热图的概率分布估计。

【image6-3】去权热图举例，取不同标准差（σ = 5, 20, 100）

类似的，对于基于去权热图的眼动熵进行关于高斯核标准差r的遍历计算，并评估他们的pearson系数和分类准确率，得到如下的图像。

【figure 6-2】去权热图的眼动熵随标准差变化

虽然只是改变了注视的权重比例，但结果相比传统的热图却产生了很大的差别。在从13px-60px的广泛标准差区间里，眼动熵与美感的相关系数的都小于-0.5，表现出了与美感的显著地负相关性。

我们挑出其中相关系数表现较优的标准差等于30px的去权热图的熵值进行进一步的分析：

【table 6-1】标准差等于30px的去权热图的眼动熵数据

【table 6-2】标准差等于30px的去权热图的anova表

其盒图和二维散点图如下：

【figure 6-3】sigma=30px的去权热图的anova盒图

【figure 6-4】sigma=30px的去权热图的散点图

1.2.2.3对分布估计算法的小结

上述的计算似乎表明与美感相关的眼动空间分布的一致性是一个较为简洁的概念，其依赖的概率分布与单个眼动注视的时长（duration）没有直接关系。结合1.1中的分析，这可能是因为duration在个体之间的较大差异造成的。

在本文的后续讨论中，我们把基于去权热图的0阶眼动注视空间分布熵称为视觉注意熵（Visual Attention Entropy），并默认取去权热图的高斯标准差为30px（当然由于在很大的标准差范围内去权热图与美感都表现出显著的关联性，取30px左右的其他值也是可以的）。

* 1. 一阶空间分布熵——一阶转移熵

0阶的空间分布熵只讨论注视之间的独立的空间分布的一致性。下面我们讨论眼动注视之间的相互转移的确定性。最早在【Gilland\_FINALDissertationREV1】中提出了基于markov链的眼动注视序列熵的算法。下面我们在我们试验中重复他的算法。

首先，为了统计转移，所有注视需要被聚类到更大的区块中，然后讨论就转化为了眼动注视在这些区块间的转移。在本实验中我们采用AOI聚类算法来实现这种对注视的区块分类。

* + 1. AOI
    2. 转移序列的马尔科夫性假设
    3. 相对信息熵
    4. 计算结果

2 对VAE的进一步讨论

2.1 眼动熵的有效性

上述实验结果初步表现了基于去权热图的眼动熵与美感的关联性，为了验证这种关联性是否可靠有效以及眼动熵是否能够代表眼动的一致性，我们从注视总数，被试人数两个维度展开探讨。

2.1.1注视总数

对于一个页面而言，如果所有的注视都是随机落到画面的随机位置的，那么显然越多的注视个数更有可能导致较高的信息熵。 是否实际上是因为好看的网页相比难看的网页在同样的实验时间内获得了更少的注视从而使得其眼动熵较小呢？实验数据的分析结果表明，恰恰相反！

统计得到的注视个数与网页美感表现出微弱的正相关性，代表着好看的网页在总体上反而有更大可能获得更多的注视。

【table 6-3】每个页面的fixation总数的表格

【table 6-4】fixation总数的anova检验

【figure 6-5】总数的anova图

【figure 6-6】总数的散点图

另一方面，视觉注视总数与眼动熵的相关系数仅为0.039，这意味着页面的视觉注意个数与其眼动熵的大小没有关联性。

2.1.2被试人数

为了进一步验证眼动熵的与美感之间的关联，同时探究眼动熵的稳定性与被试人数的关系，我们从对被试的人数进行限制。从随机抽取2人、随机抽取3人一直到抽取全部30个被试，并分别计算基于抽出的被试的数据的眼动熵。这里眼动熵的计算中，对传统热图和去权热图的标准差我们都取xxpx。

【figure 6-7】传统热图和去权热图的眼动熵分别随人数的变化

可以看出，相比于传统热图的飘忽不定，去权热图与美感的相关系数随着被试人数的增多逐步趋于稳定并逐步趋于-0.5至-0.6之间。

本质上在眼动熵的计算中，被试的眼动数据提供一个对页面眼动分布的估计，样本量越大这样的估计的误差越小。因而眼动熵与美感的关联性是需要足够多的被试来表现的。

2.2 眼动熵随时间的发展

在一个时间维度上，眼动熵的值是如何发展变化的，其与美感的关联是怎样随时间变化的？计算多长时间的眼动熵能取得与美感的更好的关联性。这部分，我们讨论沿时间发展的眼动熵的一些属性，从眼动熵自身随时间的变化规律和不同时刻的眼动熵与美感的相关性两个方面展开讨论：

2.2.1眼动熵的大小与曝光时间的关系

对于一个时刻t（秒），我们计算0-t（秒）的眼动熵的值。则对每一张页面，可以得到一条眼动熵关于时间的曲线，把所有页面曲线在一起展示，如下图：

【figure6-8】xxpx时眼动熵的值随时间的发展

可以看到，整体上，网页无论好看与否，其眼动熵是随时间不断增大的。这一定程度上表明，早期的眼动行为是偏向生理性的，本能性的，从而在个体间表现出一定的一致性。而随着时间的发展，个体间的认知、阅读习惯等差异逐渐表现出来，从而使得眼动熵愈来愈高。

2.2.2眼动熵与网页美感的关联性随时间的变化

0-t秒的眼动熵与美感评分的相关系数关于时间t的曲线如下

【figure 6-9】xxpx时沿时间眼动熵与美感的关联性

可以看到眼动熵与美感的显著关联性最早在1000ms前后表现出来，并随着时间的继续发展有一定波动，但一直保持在-0.5以下的较强的负相关性。也就是说至少需要1s左右的时间人的眼动行为才能对不同美感水平的页面表现出较为显著的“区别对待”。

由于实验时长所限，难以考察到3s以后的眼动熵与美感关联的情况，但可以猜测，随着个体之间的浏览习惯差异逐步表现，这种关联性会随时间慢慢变得不显著。

2.3先验条件的讨论

上述基于眼动熵的讨论和实验似乎表明：被试的眼动注视在画面上的分布越是局部集中，则网页越有可能是好看的网页。但这个论述是很容易举出反例的：一个纯白色的画面上的一个黑点很有可能吸引很集中的视觉重点从而获得很小的眼动熵，但却不太可能会是最好看的页面。

【6-4不合理的页面】

在我们的实验中，由于样本量较小，各网页之间的信息量差异不大，故使得单纯的眼动熵也表现出了不错的与美感的相关性。要进一步探究美感与眼动一致性的关系，我们需要纳入考量一些先验信息量相关的因素。这里，我们采用平均个体眼动熵来考量一个页面的信息量。

与总体的眼动熵不同，计算发现平均个体眼动熵与页面注视总数具有0.59的高相关度，越多的注视个数导致越高的个体眼动熵，这与随机状态下的眼动熵与注视总数的关系接近。这表明在我们的实验时间内，个体的眼动注视是倾向于散布在画面上的，而非倾向于集中在几个区块上。从而短时间内的个人的眼动熵能在一定程度上表现他在页面上所浏览的范围并估计他获取的信息量大小。较大的眼动熵表示被试者浏览了较为分散的页面空间，获取了较大的页面信息量。

【公式】

以下是每一个页面的所有被试的个体眼动熵的均值和标准差范围和页面得分排名的数据图。

【6-10 页面的所有个体眼动熵的方差】

以下是计算得到的每个页面的个体平均眼动熵的盒图和散点图

【6-11 个体平均眼动熵的盒图】

【6-12 个体平均眼动熵的散点图】

可以看到在本实验中，个体平均眼动熵与美感页并没有很强的联系。这是符合预期的——信息量本身与美感没有直接关系。而针对我们提出的猜想：一个好看的页面应该引导被试以较小的眼动路径选择的代价来获取较多的页面信息量。我们把个体平均眼动熵作为先验条件，来考察总体眼动熵/平均个体眼动熵这一指标与美感的关联性，其数据，盒图和散点图如下。

【6-5 眼动熵比的数据】

【6-13 眼动熵比的盒图】

【6-14 眼动熵比的散点图】

可以看到，相比只用总体眼动熵，眼动熵比指标表现出了明显更强的与美感的相关性，并在理论上更完备，与我们的预期猜想一致。同时，眼动熵比作为一个比值，其取值在1附近，更方便推广到不同时长、被试人数和画面分辨率的实验情形下去。

## - discussion

一、半径意味着什么？与黄斑有没有关系

二、美感与眼动的关联与信息论的关系

三、基于沿时间的眼动熵对人类眼动行为的一些推测。

## - further discussion

眼动实验的新可视化指标

基于眼动熵的对现有眼动数据可视化的探索，如去权热图，布尔饼图和相关眼动熵数据用来配合传统热图表现空间眼动的一致性；

【image 8-1】去权热图的可视化界面

## - conclusion