## Analysis and results

## 3.1 Traditional statistical indexes

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。

【眼动注视个数的样本呈现】

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有微弱的推测性。

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数的表格

【table 6-4】fixation数的anova检验

【figure 6-5】fixation数的anova图

* + 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，页面上每个被试的平均眼动时长的样本如下

【duration个体平均值样本呈现】

而考察一个页面上发生的所有眼动注视的持续时长的平均值和方差与美感的相关性结果如下：

【duration的总体平均值和标准差与美感的关系】（均值：r:-0.15/F:0.3, 标准差：r:0.12/F:0.62）

都没有表现出较好的显著性。

* + 1. Khalighy2015的contrast

Khalighy2015中通过计算页面上的眼动注视的个数与duration的标准差的乘积来代表页面的对比度

Nf \* σf（Nf代表fixation的数量，sigmaf代表这些fixation的duration的标准差）

并发现对它的简单样本的美感评价存在不错的推测能力。然而在本实验中我们重复了他的计算，发现结果并不理想。该指标的相关系数为0.1855，anova置信系数为0.3736。这可能是因为在较复杂的刺激物下（如我们实验中所使用的网页），图像所带来的信息量较大，眼动的注视个数和duration的标准差已经无法分别代表khalighy文中提出的画面的复杂度和画面的元素比例的均衡度了。

* + 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】眼跳距离标准差Plumhoff and Schirillo 2009 的实验认为好的图片saccade length的标准差较大。

其均值与美感没有显著的相关性，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。

* + 1. AOI

AOI是Area of Insterest的简写，意为视觉注意区块。AOI往往可以通过人工手动标注的方式获得，一般用于研究设计中的指定的区域是否得到了足够的视觉关注。在这里为了保证一致性和公平性，我们采用对注视进行算法聚类得到的AOI，算法采用眼动实验软件Tobii Studio自带的AOI聚类算法。聚类得到的AOI例图如下。

【例图，AOI】

以下是每个页面聚类得到的AOI的个数，页面按照美感评分从小到大排序：

【AOI个数】

页面的AOI个数与美感的相关系数为-0.26，ANOVA显著性置信系数为0.1631。其本身对美感不具有推测性。

1.1.6 fixation number in aoi （mean and std）

## 3.2 Entropies in our glancing behaviors

为什么选择熵

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。下文中我们将通过实验数据分析，分析其与美感的关系和其美学理论的合理性。

香农公式：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。

Now we investigate the entropies in the data from fixation spatial distribution and scan path temple sequence two aspects

## 3.2.1 Entropy in gaze transition plot

从过程序列一致性的角度

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链假设的眼动注视序列熵的算法用于驾驶员视觉相关的一致性统计量。下面我们重复他的算法。

首先，为了统计转移，所有注视需要被聚类到更大的区块中，这样讨论的对象就从稀疏分布在画面上的注视点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。

通过AOI聚类，我们可以把一个页面下的每个被试的眼动注视数据转化为一系列的AOI跳转序列。以例图为例，部分序列如下：

【例图的部分AOI跳转序列】

Saccade 分为两种，aoi内部的和aoi之间的。这些跳转序列已经去除了AOI到自身的跳转。

* + 1. 转移序列的马尔科夫性假设

如果一个拥有有限个状态的离散的时间序列{x0, x1, ……, xt}满足P(xt+1 | xt)=P(xt+1|xt,xt-1,xt-2,……,x0)，则称该时间序列为马尔科夫链。对于眼动的AOI序列，可以通俗地解释为，下一次的跳转到哪个AOI仅仅与当前在哪个AOI上有关。

由于上述的性质，一个马尔科夫链的全部跳转概率信息可以通过一个一步转移概率矩阵来概括。在实验中，该矩阵通过统计该页面上发生的所有的一次AOI转移得到。如图是通过例图的眼动数据统计得到的一步转移概率矩阵：

【例图的markov转移概率矩阵】

矩阵中，pij代表从AOI i跳转到AOI j的概率。

* + 1. 相对信息熵

针对上述的markov矩阵【引文】给出了一阶转移熵的计算方法

【Σ(Σp(Y|X))p(X)/Hmax】

其中p(Y|X) = pxy表示从X出发转移到Y的条件概率。p(X)表示X的先验概率，即从X出发的概率有多大，通过统计页面上所有AOI序列中X的出现的频率可以获得。

Hmax是指当下AOI数量下的极大熵，极大熵在所有转移概率都相等且先验概率都相等时取得。通过除以极大熵，来归一化不同AOI数量下的熵，使他们相互之间具有可比性。

* + 1. 计算结果

我们取了0-1.5s和0-3s两个时间段来计算上述定义的一阶空间分布熵。都没有取得很理想的效果：

0-1.5s的熵的相关系数为0.0582，anova置信系数为0.864

0-3s的熵的相关系数为0.1585，anova置信系数为0.4741

这样的不显著性除了可能是因为AOI转移本身对美感不具有显著的推测性外，也可能是由下面两方面原因导致的：

* 马尔科夫假设的不适用：我们的眼动实验只考察被曝光后短时间内的被试的初始的眼动行为。这样的眼动过程是探索性的，会遵循一个潜在的顺序，因而马尔科夫链的这样一中宽平稳过程的跳转假设也许是不合适的。
* 样本量的限制：我们实验的样本量较小，无法获得充分多的AOI转移样本来更精确地估计一步转移矩阵。

## 3.2.2 Entropy in gaze/attention map

Heat maps (also referred to as attention maps) are often used to visualize the eye movement data in this evaluation process.

heatmap entropy中反映过程次序的信息被剔除了

相当于眼动注视的“0阶”markove熵，吧fixation看做一个个独立的事件。

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。

我们首先给出如下的基于gaze map的眼动熵的定义：

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gazemap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值，当注视分散在不同空间位置的时候它取得最大的值。

该指标中的p是眼动注视的二维离散空间的概率分布, 满足sum(p) = 1，p(x, y)是眼动注视像素点(x, y)的可能性，是后验的，待估计确定的。

概率分布P（X，Y）满足ΣP(Xi, Yi) = 1。

传统热图：一个能够直觉想到的，也是现有的能够直观反映注视在画面上分布的可视化方式是热图：Heat maps (also referred to as attention maps) 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)。 如此叠加所有被试的每一个的注视之后对得到的热图进行归一化处理，就可以用于计算眼动熵的概率分布p。

【image6-2】传统热图举例，（σ =30px）

【image6-3】去权热图举例，（σ =30px）

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，故采用适当大小的高斯分布应该是具有一定的理论合理性的，它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，1-2度夹角，在屏幕上大约2厘米直径、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。这里我们暂取（σ =30px）

讨论Fixational eye movements

不幸的是，基于duration的传统热图，其眼动熵对美感的推测能力并不理想，他们的pearson系数和分类准确率只有

二）去权热图duration free heat map：对传统的热图做一些改进，只考量每个注视的坐标位置，并给每予个注视相同的权重而不再根据他们的持续时长来决定权重。 这样就得到了我们称之为去权热图的概率分布估计。最好加上和传统热图的对比

除掉duration数据后，结果发生了根本性的改变。

标准差等于30px的去权热图的眼动熵数据如下

【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=30的去权热图的anova盒图

【figure 6-4】sigma=30的去权热图的散点图

上述的计算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。在计算一致性的过程中加入duration造成的权重反而会使得眼动熵几乎毫无效果。

穿插对Duration 为何无效的讨论

目前可以确定的是position的熵是最为有效，其本质是注意的区域分布，没有过程的信息，我们称其为vae，区别于ve

## 3.3 further analysis of The efficacy of vae

在后续更深入的分析中，我们发现，vae具有很好的稳定性，对于不同的sigma、采样时间和人数

1. 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。可以发现在整个r值的区间里，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。这是令人有点吃惊的。

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。

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，表现出了与美感的负相关性。

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。

【figure 6-1】传统热图的眼动熵随标准差变化 及 去权热图的眼动熵随标准差变化

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随机抽取，从2人、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30px。

【figure 6-7】传统热图和去权热图眼动熵分别随人数的变化与美感关系

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。

3）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：

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。

1. 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。

上述实验结果表现了基于去权热图的眼动熵对美感预测的稳定性。

## 4 Conditional vae

Vae受到画面信息及其分布的影响，

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。

也许你注意到，不同的页面内容和兴趣点的多少，可以影响vae

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。那么 是否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？实际的情形恰恰是相反， 在3.1中我们看到统计得到的注视个数与网页美感表现出微弱的正相关性0.33，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。

在3.3中页面的AOI个数与美感的相关系数为-0.26，ANOVA显著性置信系数为0.1631。

总体上好看的网页，视觉更加活跃，但是关注更为集中。

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

这里，我们引入平均个体眼动熵来考量一个页面的信息量。Base vae（Be）

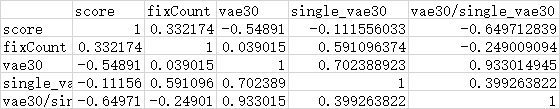
与总体的眼动熵不同，计算发现be与页面注视总数具有0.59的高相关度，和aoi有xxxxx ，越多的注视个数和aoi数量 导致越高的个体眼动熵，

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。

Base vae 和美感呈现了较为微弱的 -0.11的相关性， 可以认为be中和了fixation数量和aoi数量两者，与美感的关系更加趋于中性。一种合理的猜测，过高和过低的be都不是太好的网页。

可以认为Base vae 对美感没有推测能力，但是对vae影响很大， 正相关0.7 ，大的be 会导致大的vae，所以我们引入了conditional vae的概念，为了便于不同复杂度页面之间的比较。

在我们的实验中，由于样本量较小，各网页之间的be差异不大，故使得单纯的vae也表现出了不错的与美感的相关性。



对于内容相同的页面, vae本身也许就够了, base vae 本质是noise free 的 vae 的估计. 意味着所有人的视觉注意都一致，是特定画面视觉传达信息必不可少的注意力资源. 和画面中有意义模式的数量有关, 理想的设计中，base vae 与美感成正比，与vae成反比

【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】

（将这些图与总体眼动熵6.1 图放在一起）

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。