# Consistency, uncertainty of eye path and visual appeal of webpages

## Abstract

We put forward the concept of Eye-path Consistency and conduct an eye-tracking experiment to study its correlation with the visual appeal of webpages. Several indexes are introduced to measure the EPC with three indexes (IF measuring local fixation consistency, EF measuring overall fixation consistency and the LCS value measuring the consistency among the AOI transition sequences) respectively showing significant correlations with visual appeal, which gives powerful support to the significant negative correlation between EPC and visual appeal as well as the Fluency Theory (Reber et al 2012). Further, by analyzing the eye-tracking data by time, we put forward plausible conjecture on human vision process.

## Keywords

Visual appeal, eye tracking, Consistency, fixation distribution, AOI, edit distance

## Introduction

Webpage viewers’ form their first impressions of the page in less than 50 ms. Such kind of first impressions can later influences the user trustiness of a website.

Viewing a webpage is essentially a procedure of image processing in human visual system. Theories have being proposed on the relationship between processing fluency and aesthetic. This paper we address this problem by introducing a new concept called Eye-Path Consistency (EPC), which is extracted from eye-tracking data collected from a group of people viewing webpages. The following are the main contributions we made:

* We put forward the hypothesis on the correlation between visual appeal and EPC.
* We define indexes to measure the EPC of a webpage and conduct experiments to proof these indexes significantly correlated with visual appeal
* We put forward the hypothesis on human early visual behavior base on the analysis result of the experiment data.

## Related Works

Several theories were raised to explain the inner mechanism of aesthetic response: the mere exposure theory(Zajonc 1968), arousal dynamics theory(Berlyne 1957,1960,1971), prototype theory(Rosch 1975), fluency theory(Reber 2012) and so on. They all respectively explain some of the observations in experiments while showing some limitations.

The development of Evolutionary Aesthetics(Voland et al. 1997) and Neuroaesthetics(Jacobs C 2003)(Cinzia D D 2009)( Celaconde C J 2011) provides us a chance to go deeper into the issue. And their methods of hypothesis and empirical analysis are better compatible with other domains of science. According to evolutionary aesthetics, 美感的知觉过程是由浅入深逐步唤醒的[Leder H, Belke B, Oeberst A, et al 2004],可分成三个阶段: The perceptional analysis, the implicit memory integration, and the deep cognition.

The perceptional analysis is a gradual-evolved visual instinct and perceptual knowledge mechanism (Voland et al. 1997), showing consistency among individuals. It is better explained by the fluency theory. The implicit memory integration is aroused slightly later than the perceptional analysis. It swiftly judges the familiarity of the surrounding environment through unconscious memory. The related theories are the mere exposure theory and the prototype theory. The deep cognition is a high level psychological behavior and shows great differences among individuals of various background and experience. It is therefore to a great extent responsible for the general thinking of the immeasurability of aesthetics.

According to Tuch et al 2012, both the perceptional analysis and the implicit memory integration are aroused within 50ms. Interestingly, researches show that aesthetics judgements can be made with only 50ms of explosure (Lindgaard et al. 2006). Experiments are conducted to better understand how perceptional analysis works. Michael S 2011 discussed visual attention distribution. Michailidou E 2008 and Tuch.A.N et al 2012 discussed complexity. Khalighy’s experiment 2015 showed the relationship between aesthetics and contrast which is measured by defining pureness and proportion through eye-tracking. The fluency theory which posits that people prefer visual displays to the extent that they are processed more easily (or fluently) is so far a good explanation to the results of those experiments though it has some self-consistency problems.

## Hypothesis

### Consistency, Uncertainty and visual appeal

Different people have different visual habits while browsing a stimulus for example a webpage. Some show more interests on images, while some others care more about verbal contents. Thus it is rather meaningless to measure the consistency between the eye paths of the viewers while deep reading. However since all human individuals share the same visual system, there could be some similarity among our visual behaviors in the very seconds of our exposure to a stimulus which is worthy researching.(这里需要关于人类视觉行为相似性方面的早先论文的支持，即从生理或神经层面支持人类眼睛行为的相似性，不受经验影响，可以是进化论方面的结果)

As the fluency theory arguing that ‘people prefer visual displays to the extent that they are processed more easily or fluently’ (Reber 2012; Reber et al. 1998, 2004), We believe that a webpage design with a good visual appeal should be able to conduct the viewer’s eye path in the initial seconds to reduce the uncertainty of the eye path so that the viewer could process the page more easily. To be short, the lower the uncertainty of the eye path is, the better is the visual appeal of the webpage.

Since the similarity of people’s visual system, uncertainty can to some extent be observed by measuring the consistency among the initial eye paths (eye paths in the initial seconds)of different viewers. The high consistency among people’s initial eye paths(EPC)indicates a low uncertainty of the eye path of each individual and further, indicates a good visual appeal of the webpage.

### Indexes measuring EPC

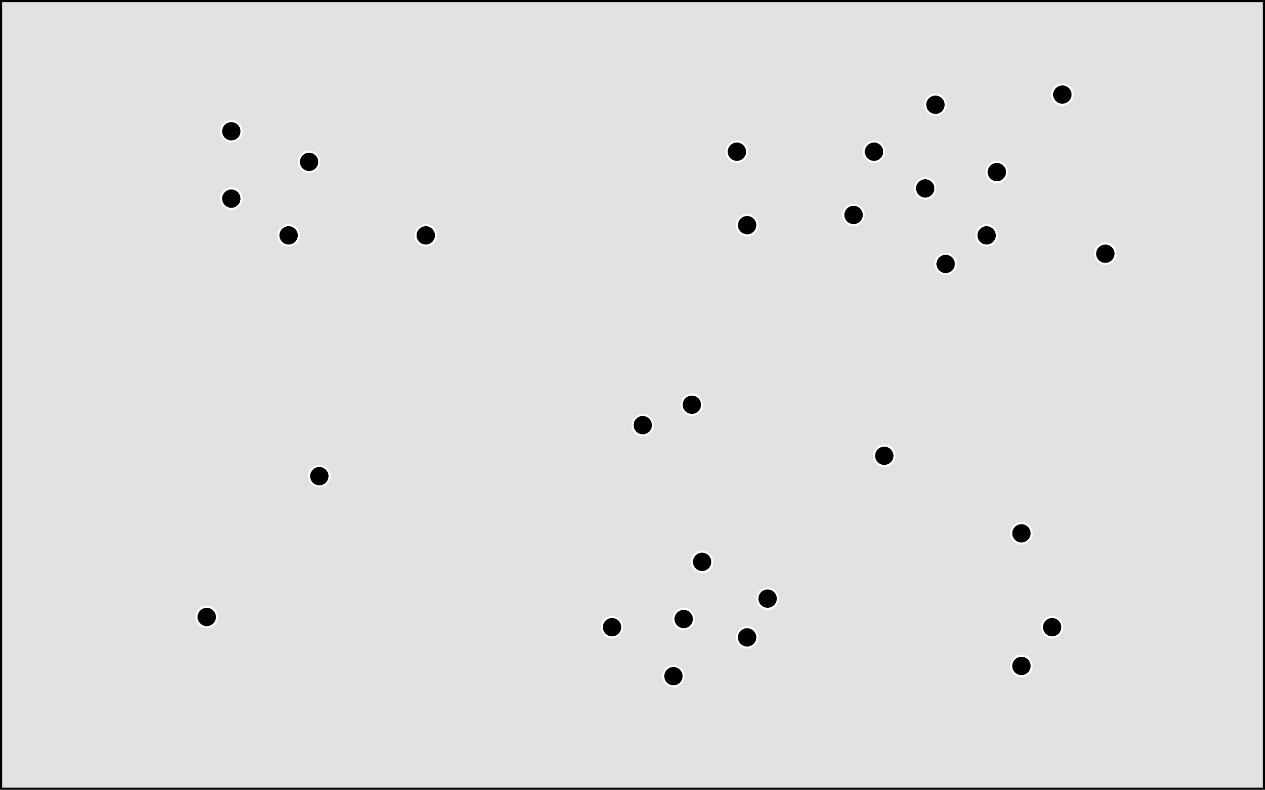
The EPC is a broad concept including time and space dimensions. In the following we introduce several indexes to measure different aspects of EPC. The indexes are sorted by the ways they are calculated.

#### Data format of eye trackers

All the indexes are calculated based on the data provided by the eye tracker. The eye tracker records the viewer’s eye path through a series of fixations. Each fixation has four parameters: Timestamp (starting time), fixationDuration, MappedFixationPointX and MappedFixatonPointY. The former two tell time information of the fixation and the latter two tell the location of the fixation. The fixation points are closely arrayed along time, sparing no time between each other. That is, the sum of the Timestamp and the fixationDuration of a fixation equals to the Timestamp of the next fixation.

#### Fixation-distribution indexes:

The fixation-distribution indexes are directly based on the fixation data of all viewers at a certain point of time, for example at 500ms after the webpage exposes, we introduce two indexes to measure the space-time range of the distribution of the fixations. The smaller is the range of the distribution, the higher is EPC.



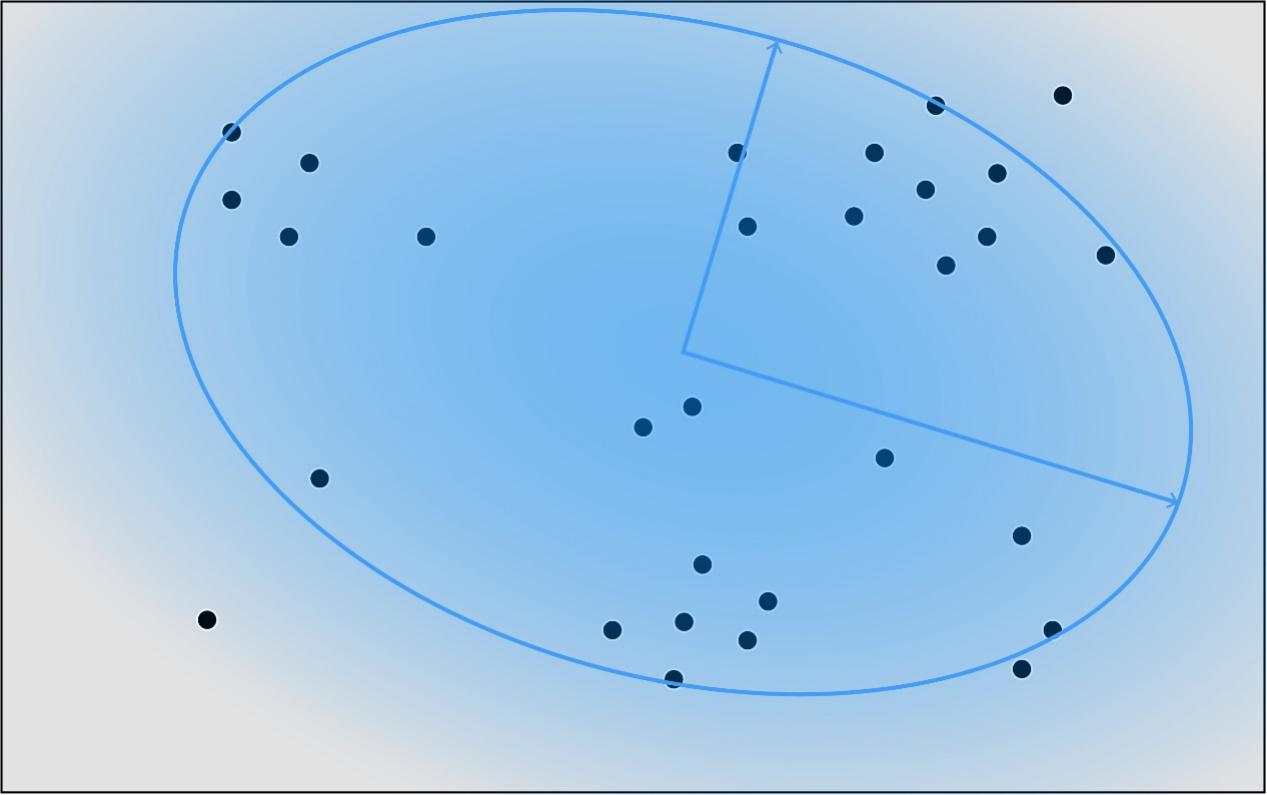
%the fixations of all viewers on a certain exposure time

**Measuring the range of overall distribution**

The covariance matrix of the locations of the fixations is calculated to measure the space range of overall distribution, based on which we can obtain the area of the ellipse representing the range of the overall distribution. An integral from t to t+deltat is introduced to include the range on time dimension.



We name it the **Ellipse function** (EF).



%the ellipse function measuring the range of overall distribution

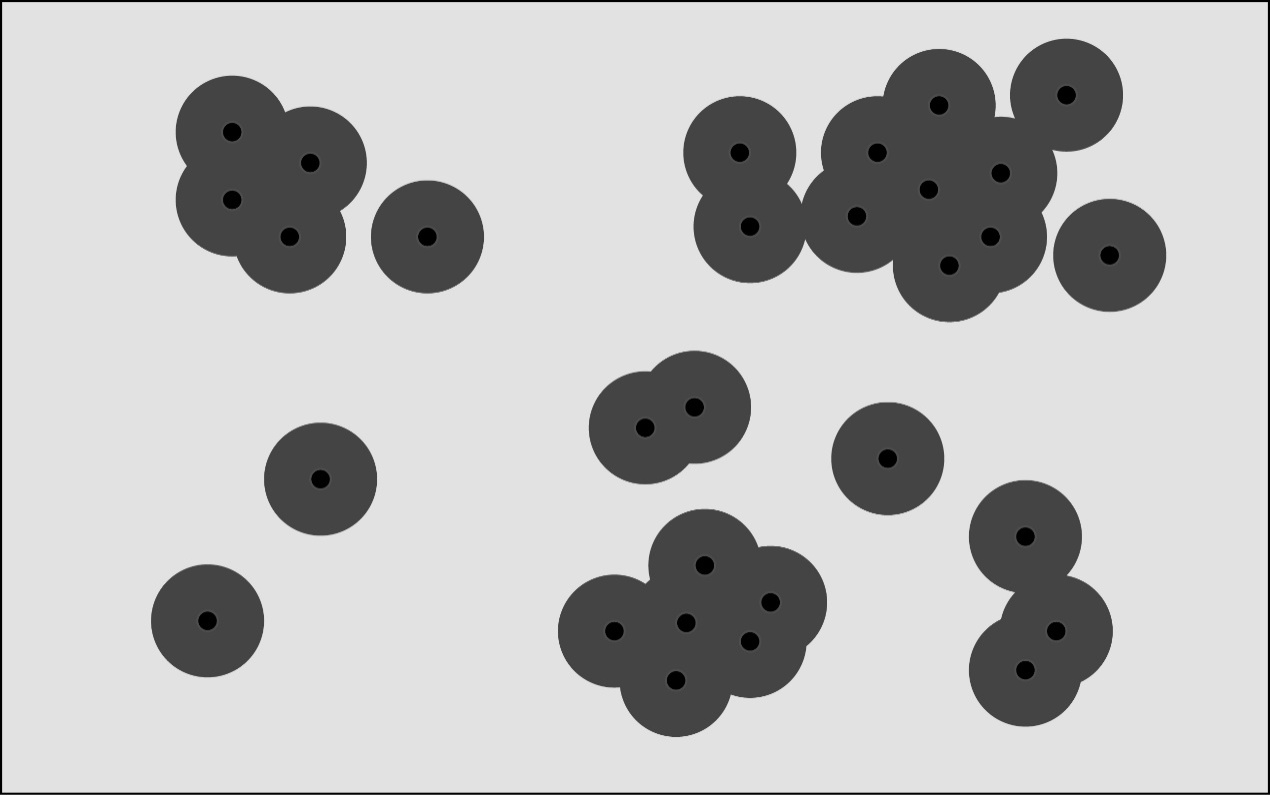
**Measuring the range of local distribution**

To measure how spatial locally concentrated the fixations distribute, we put a filled circle with a certain radius on each fixation and calculates the area covered by the circles. Those circles may overlap with each other, reducing the area calculated. Thus the smaller is the index, the locally closer are the fixations. Also, an integral is introduced to include the range on time dimension. Though the following [mathematical](javascript:void(0);) [expression](javascript:void(0);) may look a bit complicate, the concept is very simple and natural indeed.

 Mark %F(t) as the 2D point set of all the locations of the fixations at exposure time t



We name it the **Island function** as the overlapped circles look like a series of islands on the screen.

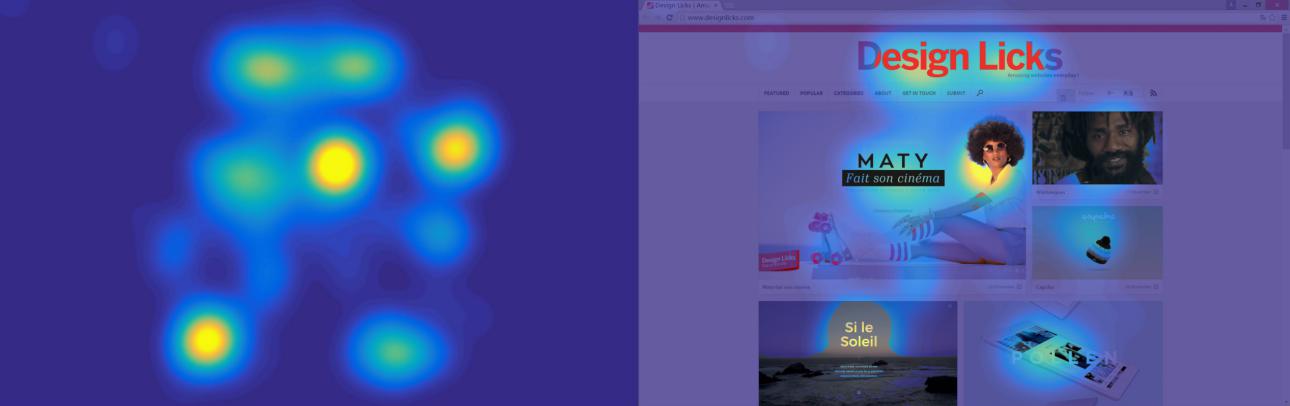


%the island function measuring the range of local distribution

#### Heat-map based indexes

**Heat map:**

The heat map shows the density of fixations on different locations by colors. Two EPC-related indexes are calculated based on the heat map: the peak value and the histogram. To ensure a standard measurement, the heat maps are normalized so that the sum of all values of the pixels equal to 1. The heat map can then be regard as a two-dimensional probability density function.

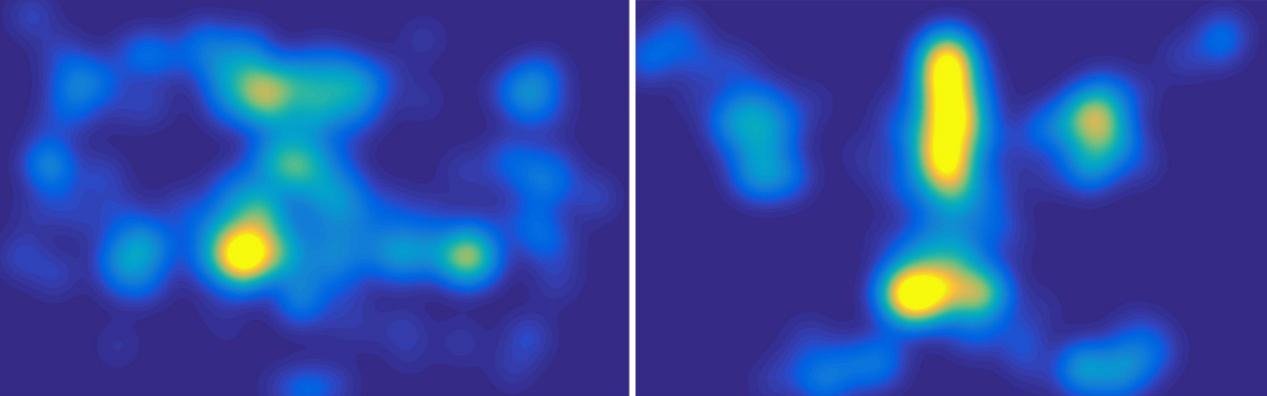


%doc: heatmap%instruction:The heat map of a webpage

**Peak value:**

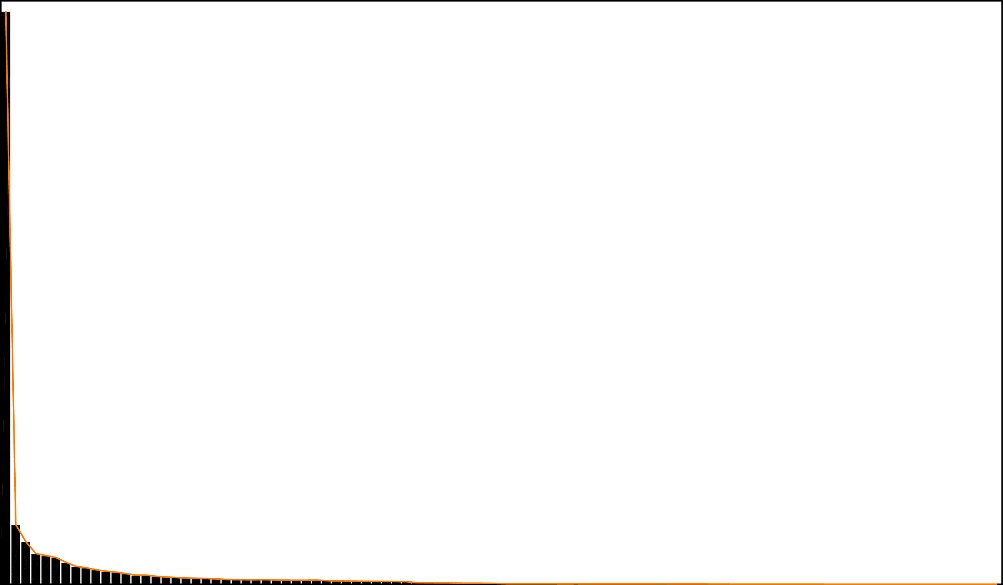
Peak value is the maximum value of the heat map produced by the fixations of all viewers. A high peak value shows how strong the strongest fixation concentration area of a webpage is. Though not directly related to EPC, it’s still taken into consideration as a comparison.

**Chaotic Area:**



%a heat map on the left has a larger Chaotic area than the right one

The nonzero low-stage area of a normalized heat map represents the proportion of chaotic fixation area. Theoretically, a less chaos indicates the high certainty and concentration of fixation locations. However experiments are required to find out which stage of heat map represents the heat map chaos best. Thus the heat map histogram is introduced to count the area proportion of each stage of value of a heat map and show them on a histogram, like the histogram of a picture used in software like Adobe Photoshop.



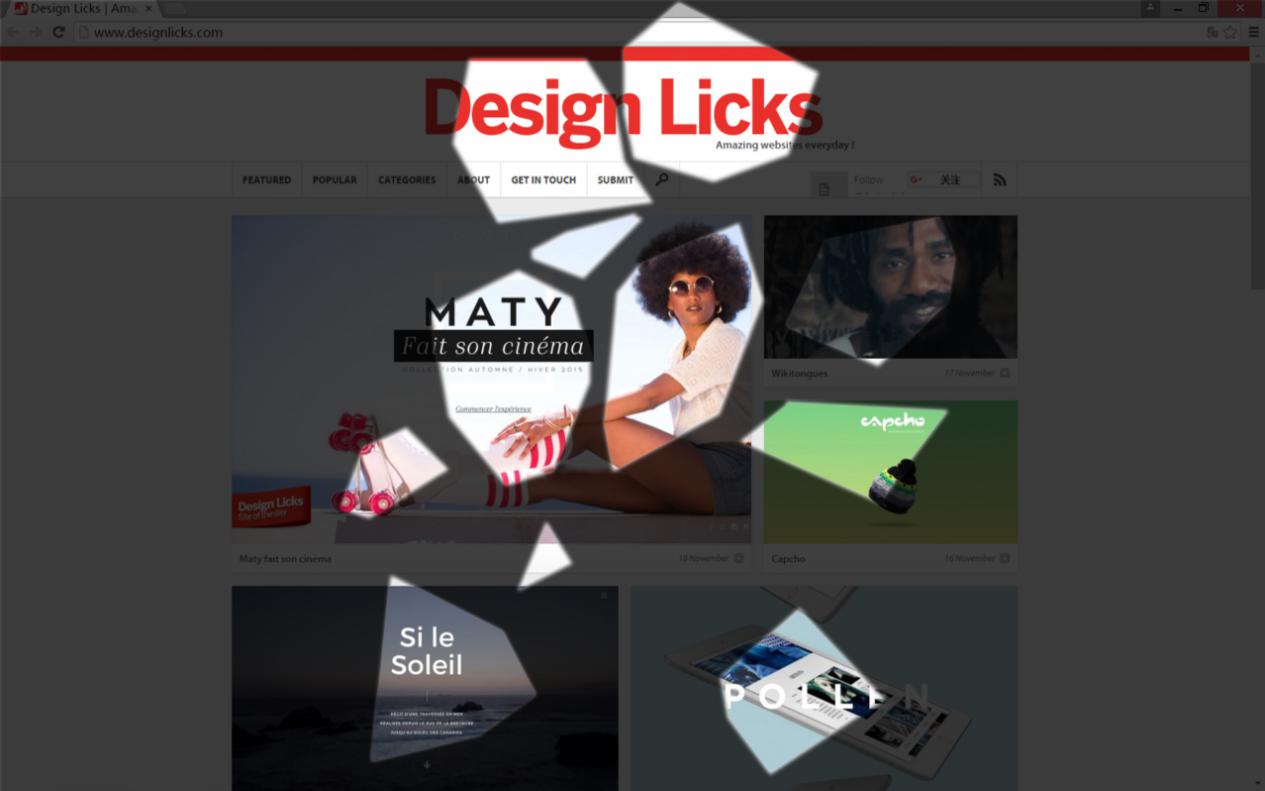
%the histogram of a heat map

By searching every stage of value on the histogram, we may be able to ascertain the best correlated stage to represent heat map chaos.

#### AOI Based Indexes

AOI, area of interest, marks area that is likely to concentrate the viewers’ fixations on a stimulus. Webpages may have different amounts of AOIs. In researches focusing on graphic design or post design, AOIs are often marked manually to study whether the expectant area gets enough focus. However, as a base on which the indexes of the visual behavior are calculated, it is more indiscriminately to mark the AOIs by algorithmic ways. Ways like dividing the stimuli equally into grids can easily generate a series of AOIs for all webpages without any eye tracking tests.

In this study, we adopt the cluster algorithm provided by Tobii studio to generate the AOIs. The algorithm is based on the collected eye path data of all viewers, it clusters the spatially neighboring fixations together into a series of convex polygons(凸多边形), each representing an AOI. Although the AOIs can probably be wrongly clustered, for example clustering the neighboring title area and text area into one AOI or clustering a complete image area into two AOIs, it’s still a best way to generate AOIs with an indiscriminate scale. Later we will discuss the indexes’ robustness to these potential mistakes.



%the polygons form AOIs clustered by Tobii studio

**AOI sequence**

Based on the AOIs, a series of fixations of an eye path of a stimulus can be translate into an AOI sequence by the following rules

* Those fixations who locate in none of the AOIs are simply deleted.
* Shifting from an AOI to itself is allowed.
* Each AOI is marked by a unique capitalized letter and the sequences are presented by strings consisting of those letters.

%%

With the sequences preserving only the shifting between AOIs, indexes can be defined to measure the consistency among those sequences which is also the consistency of the eye paths to some extent.

**Edit distances**

The edit distance is a way of quantifying how dissimilar two [strings](https://en.wikipedia.org/wiki/String_(computing)) are to one another by counting the minimum number of operations required to transform one string into the other.(出处) It’s been widely applied in natural language processing and the quantification of the similarity of DNA sequences. With different operations allowed, edit distances can be defined differently, here we introduce the Levenshtein Distance and the LCS index.

* Levenshtein Distance

The Levenshtein Distance is the most common used edit distance which allows three different operations: removal, insertion and substitution. The smaller is the Levenshtein Distance, the more similar are the two sequences. For example, %%%

* LCS index

LCS index or LCS function, where LCS is short for Longest Common Subsequences, is widely used in bioinformatics and revision control systems. It finds the longest common subsequence of two sequences (which is often not unique) and counts its length to be its value. Thus the bigger the index is, the more similar are the two sequences, opposite to the Levenshtein Distance. For example, %%

As the LCS index doesn’t fit the mathematical definition of distance, it can’t be named as LCS distance. However it is intrinsically related to the edit distance when only insertion and deletion are allowed and has the following property.

%%%

The aforementioned indexes can both be calculated with a time complexity equaling to the product of the lengths of the two sequences by dynamic programming.

Since the indexes only compare the similarity between two sequences, statistical [magnitude](javascript:void(0);)s will be needed to count the values of all the combinations of sequences to measure the consistency among all the eye paths. This will be discussed later in the RESULT section.

**Scan path entropy**

The scan path entropy was defined by Ignace Hooge 2013 to measure the consistency of the AOI sequence to a certain AOI. The original AOI sequence is first varied into a first-arriving sequence. For example%%.Then for a certain chosen AOI, the probability of every different kind of first-arriving sequence types are calculated based on which the Shannon entropy is generated.

Hooge used the entropy to quantify the consistency of the scan path to a certain AOI. However we find it hard to find a suitable statistical [magnitude](javascript:void(0);) to count the entropies of all the AOIs on a webpage to represent the consistency of the intact eye paths. Moreover, the Shannon entropy is so strict to the sequences that different sequences are equally treated despite the similarity, ending up in a quite low calculability.

**AOI transition probability matrix and F value**

The concept of the AOI transition matrix was first mentioned by Holmqvist et al 2011. It’s an N×N matrix (N=the amount of the AOIs), mathematical expressed as %%{aij}n where the %%aij element of the matrix represent for the probability of a saccade from AOI I to AOI j.The sum of the probability of one AOI to all AOIs (including itself) is either 1 or 0(only occurs when the AOI only appears at the end of a sequence), very much like the Markov transition probability matrix.

%%%sigma aik=1 or 0

However, because whether the AOI transition process is a Markov chain is still unknown, we cannot name the matrix after Markov.

Hooge et al 2013 introduced an index F to measure the certainty of the AOI transition based on the matrix. Here we altered the original definition a bit, replacing the Anet matrix with the absolute value of the Anet matrix. %%% A high F value indicates a firm transition direction for every AOIs and the strong certainty of AOI transition.

The F value is then define as followed:

%%%%%%

We summarized the aforementioned indexes in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **method** | | **index name** | | **measure** |
| Fixation Distribution | | ellipse function (EF) | | overall space-time range of fixations |
| island function (IF) | | local space-time range of fixations |
| Heat map | | peak value | | strength of the strongest fixation concentration area |
| Chaotic area | | proportion of chaotic fixation area |
| AOI cluster | AOI sequence | edit Distance | LCS index | consistency of AOI sequences |
| Levenshtein Distance |
| Scan path entropy | | consistency of the sequences towards a certain AOI |
| Transition Matrix | F value | | the certainty of transition between AOIs |

## Experiment

### Eye tracker

The eye tracker we used is Tobii T60.It uses a pair of cameras to capture the user's continuous eye movement at a fixed frequency of 60Hz, collecting a series of disperse points. The associated algorithm then fits the points with quadratic curves and turns the data into a series of fixations.

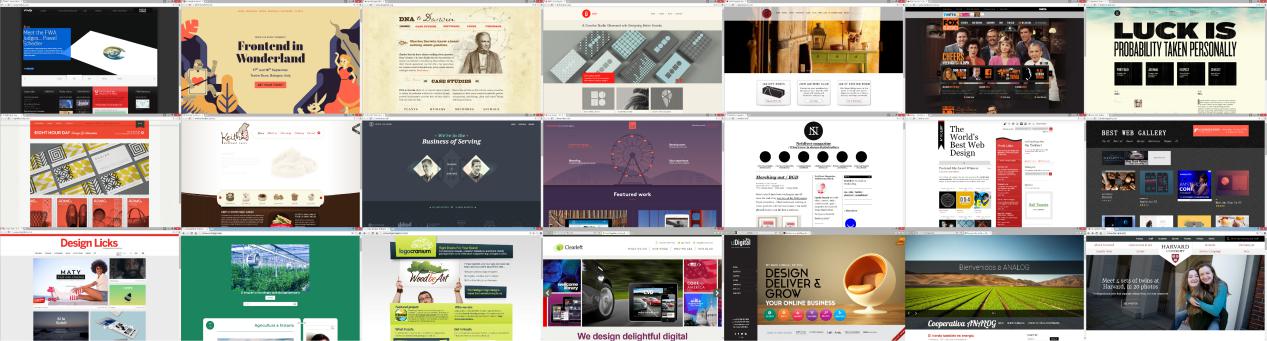
To control the deflection, a calibration is done before each test by requiring the participants to focus on several points on the screen. However, errors are still inevitable due to slight calibration deflections and occasional behaviors of the viewers like blinking. Later we will discuss the influences of these errors on the following introduced indexes.

### Environment

The room for the experiment has an area of about 20\*15m2 using only artificial lights. The eye tracker is placed on a table in front of a clean white wall. No significant noise can be heard during the test.

### Stimuli

The webpages are rated by 3 web design experts in advance. Only the pages with consistent ratings are selected as the stimuli of the experiment to ensure they are representative enough as good or bad visual pages. It is also made sure that no commonly familiar logo, mark or names can be found on the webpages. Finally, 21 good-rated pages and 21 bad-rated pages are chosen as the stimuli of the experiment. The webpages are shown to the viewers in the form of screenshot images with a resolution ratio of 1280×800 pixels allowing no scrolling.



### Participants

11 male and 19 female participants with age ranging from 18-25 participate in the test. 27 are students of the school of Media and Design of SJTU. Each participant gets 20RMB as a reward.

### Procedure

The experiment includes two tests: the eye tracking test and the rating test.

* In the **eye tracking test** the participants are asked to view all the 42 webpages 3 seconds each in a normal website browsing condition having control of the mouse. Between every two pages a one second black screen is set to reset the viewers’ fixations. A short break is set in the middle of the test to reduce the influence of fatigue.
* Shortly after the eye tracking test, the participants are asked to do the **rating test** where they rate the visual appeal of the just-browsed websites with0 for negative and 1 for positive. The pages are exposed to the viewers while rating so that they can make rational judgments.

## Result

### Visual Appeal Ratings:

The proportion of positive ratings is considered as the score of a webpage. With the expected good webpages having an average score of 0.85 ranging from 0.58-1 and the expected bad ones having an average score of 0.045 ranging from 0-0.19, the rating task result is not so far from expectation: the good and bad groups are then clearly separated. One notable phenomenon is that 1/3 (7/21) of the bad pages get a score of 0 but only 1/7(3/21)of the good pages gets a score of 1 which gives a proof to the theory that it's easier to reach consensus on bad visual appeal webpages than on good ones.&&引用

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| score | 1 | 1 | 1 | 0.97 | 0.97 | 0.97 | 0.97 | 0.94 | 0.94 | 0.94 | 0.87 | 0.87 | 0.84 | 0.81 | 0.77 | 0.71 | 0.7 | 0.68 | 0.65 | 0.61 | 0.58 |
| class | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| score | 0.19 | 0.13 | 0.1 | 0.1 | 0.07 | 0.06 | 0.06 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| class | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

%The rating score and class table, with class1 refers to the good group and class 2 refers to the bad group

### Indexes significance and correlation

All of the indexes mentioned in HYPOTHESIS are calculated except for the scan path entropy due to its bad calculability. Many problems occur while its calculation including the lack of suitable statistical [magnitude](javascript:void(0);), the lack of suitable treating to AOIs that haven’t been viewed by all the viewers. However the concept of the scan path entropy is noteworthy in further studies.

For the edit Distance indexes, we calculated three kind of statistical magnitudes named as: mean, max and mle. We will introduce these magnitudes later.

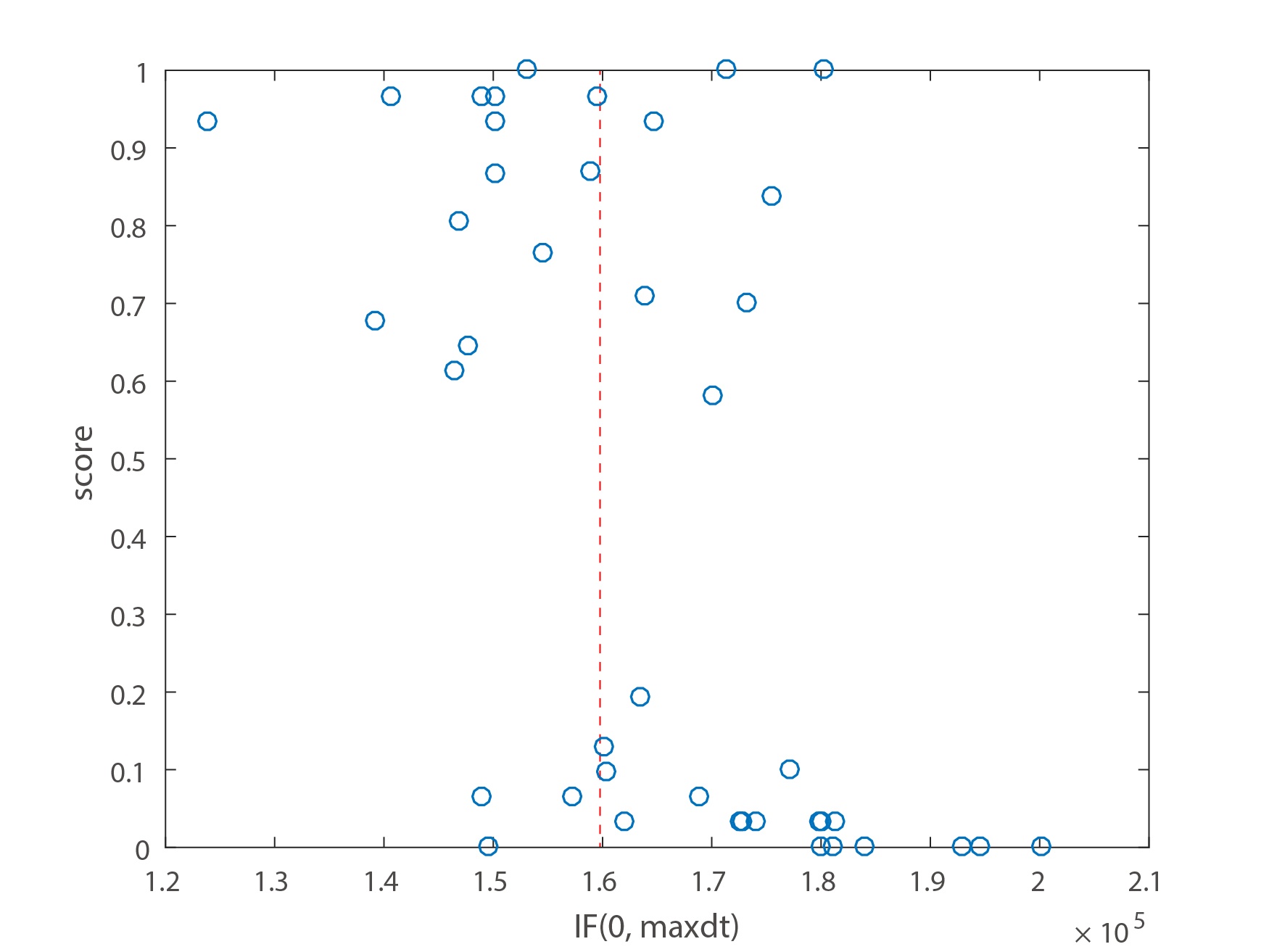
In the following table, we list the ANOVA p value (the probability to reject that the indexes of the two classes are not equal, also known as the significance) and the correlation coefficient with score for each index.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| measure | index name | | | ANOVA p | corr |
| overall space-time range of fixations | ellipse function (EF) | | | 0.0247098 | -0.35659 |
| **local space-time range of fixations** | **island function (IF)** | | | **0.0003153** | **-0.54352** |
| strength of the strongest fixation concentration area | peak value | | | 0.1573581 | 0.208867 |
| **proportion of chaotic fixation area** | **Chaotic area** | | | **5.89E-05** | **-0.59356** |
| **consistency of AOI sequences** | **edit Distance** | **LCS index** | **mean** | **0.0011874** | **0.495173** |
| **max** | **0.0015412** | **0.48892** |
| mle | 0.0072211 | 0.416874 |
| Levenshtein Distance | mean | 0.0577614 | -0.29853 |
| max | 0.025849 | -0.32921 |
| mle | 0.0577614 | -0.29853 |
| the certainty of transition between AOIs | F value | | | 0.0616347 | -0.27529 |

Under the Confidence Coefficient of 0.005, three indexes show significantly correlated with score. The IF measuring the local space-time range of fixations, the heat map chaotic areas and the LCS index measuring the consistency of AOI sequences. In the following we will analyze these indexes in detail and try to explain why these indexes work while the others don’t.

### Island Function

Both the EF and IF calculated in the above table take the maximum %deltat value, representing the time-average values of the indexes. When radius=55px, the IF value reaches its best correlation. The high significance of IF and the rather low significance of EF tell us that local space-time fixation distribution is more visual-appeal correlated.



%The IF-score figure of all the 42 webpages, the red line gives a best classification with an error rate of (7+3)/42

* **Robustness**

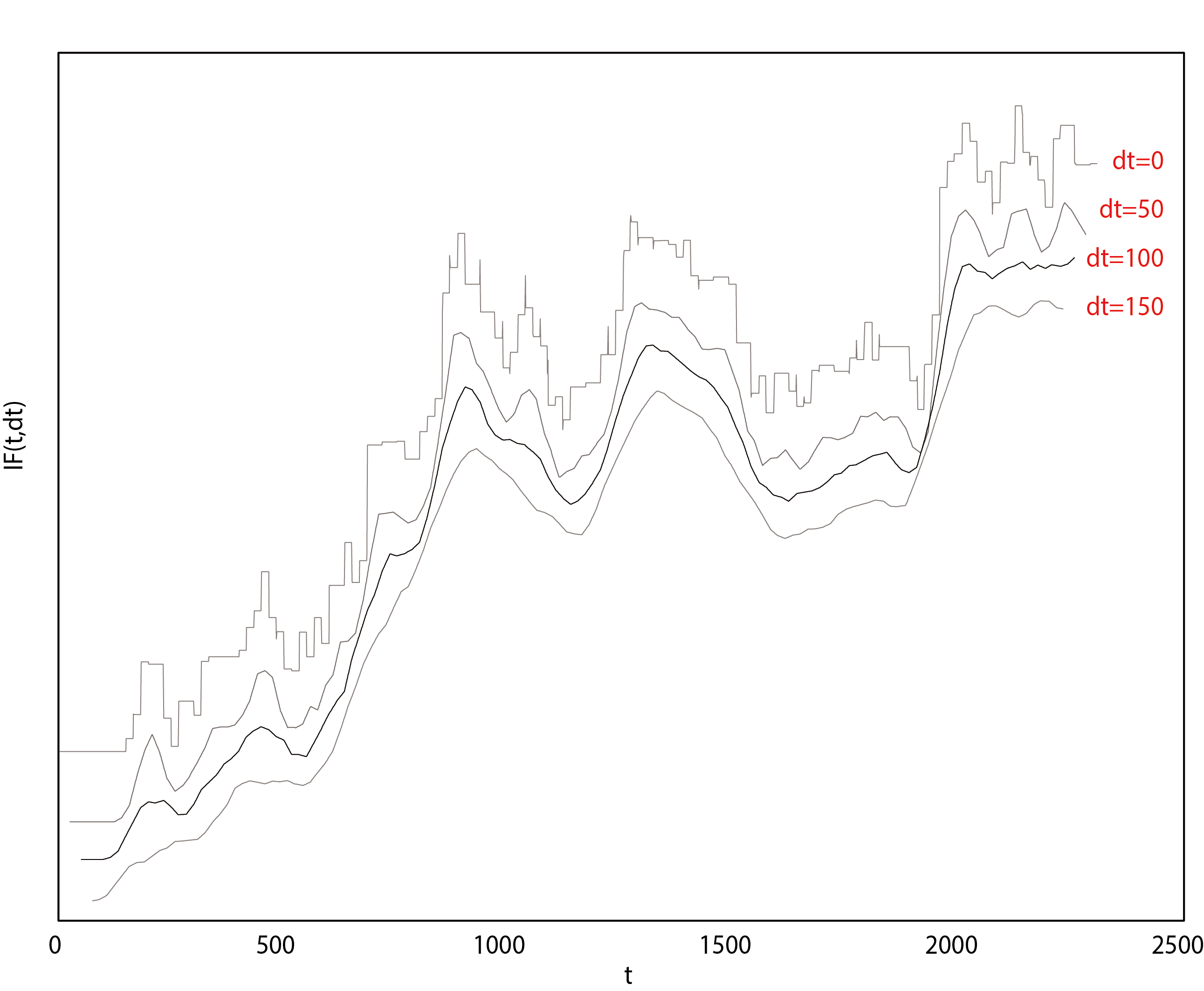
A robust check is conducted to ensure the significance of the IF index is reliable. The original fixation data are dithered on both space and time dimensions to obtain a new fixation data, simulating a low precision eye-track recording situation. The IF value is recalculated based on the new data:

|  |  |  |
| --- | --- | --- |
| index | ANOVA p | corr |
| IF | 0.000315 | -0.54352 |
| IF dither | 0.000438 | -0.53082 |

The result prove that even under a quite low precision recording situation, IF still shows a significant negative correlation with score.

* **Continuity by time**

Due to a rather average distribution of the timestamp of all the fixations, IF shows some continuity along time. The following graph shows the IF-t curves of a webpage under different deltat.



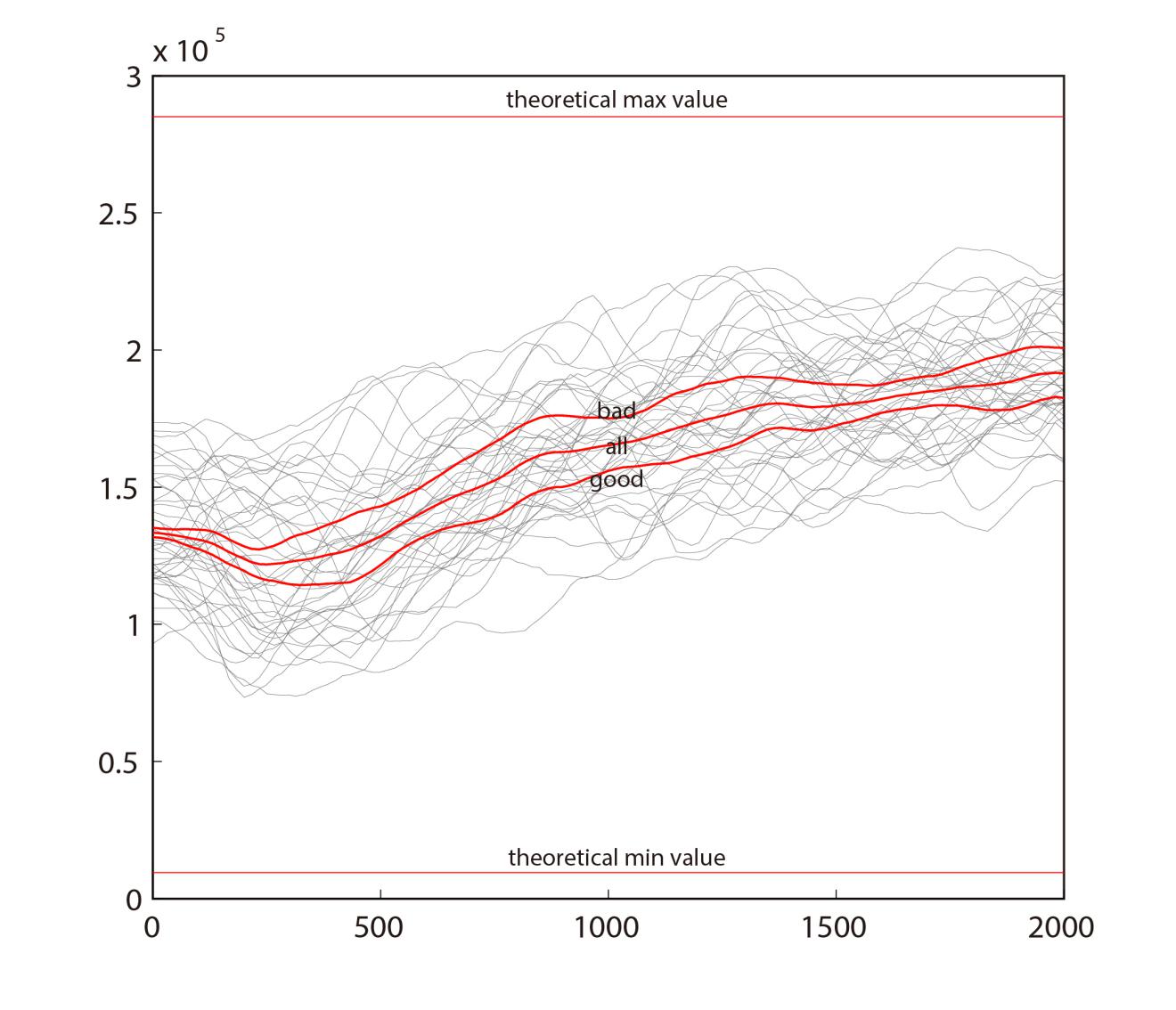
%the IF-t curves at different dt. The curves have been moved vertically for a clearer visual appearance.

In the following discussion, we fix %deltat at 100ms to ensure enough including of time dimension at the same time preserve the variety at different stages of exposure time.

* **Trend by time**

The trend of the island function shows how the local distribution consistency varies by time. The average curves of the IF-t curves of all the webpages and the good and bad classes are respectively calculated for comparison.

The following graph shows these three average curves together.



%the average curve of all the IF curves, the good class and the bad class. The light grey shows all the curves. The two red horizontal lines are the theoretical maximum value and the theoretical minimum value. The axis is adjusted to 0-2000ms as a few individuals start their first fixation quite late causing a rather short total fixation time. The 2000ms time ensures the involvement of every individuals.

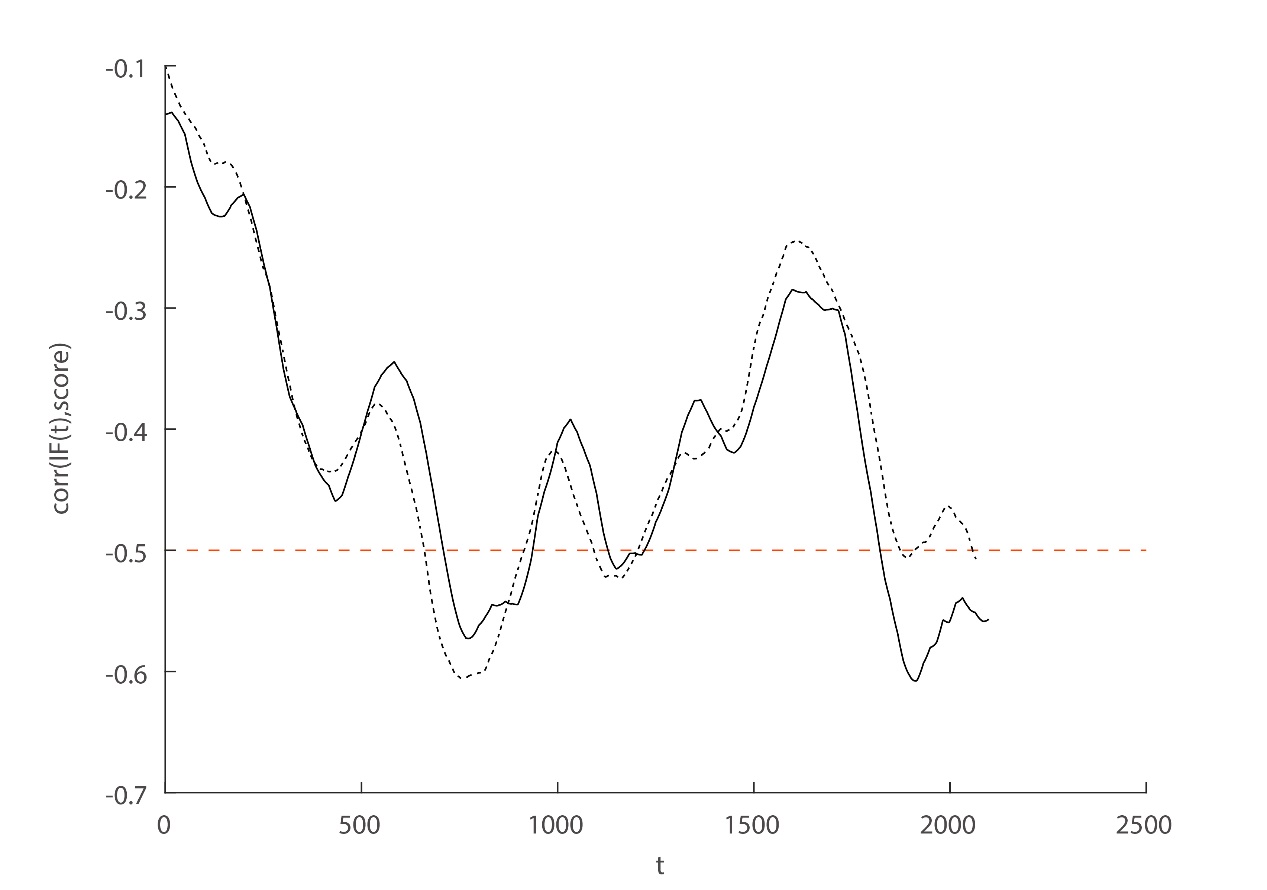
It’s obvious that the IF values increases by time, which means the consistency of local fixation distribution decrease by time. With the exposure time increases, the difference of people’s individual visual habits gradually works, decentralizing the fixations of different viewers.

* **Correlation by time**

To study the correlation between IF and score on different stage of time a correlation curve is calculated as follow



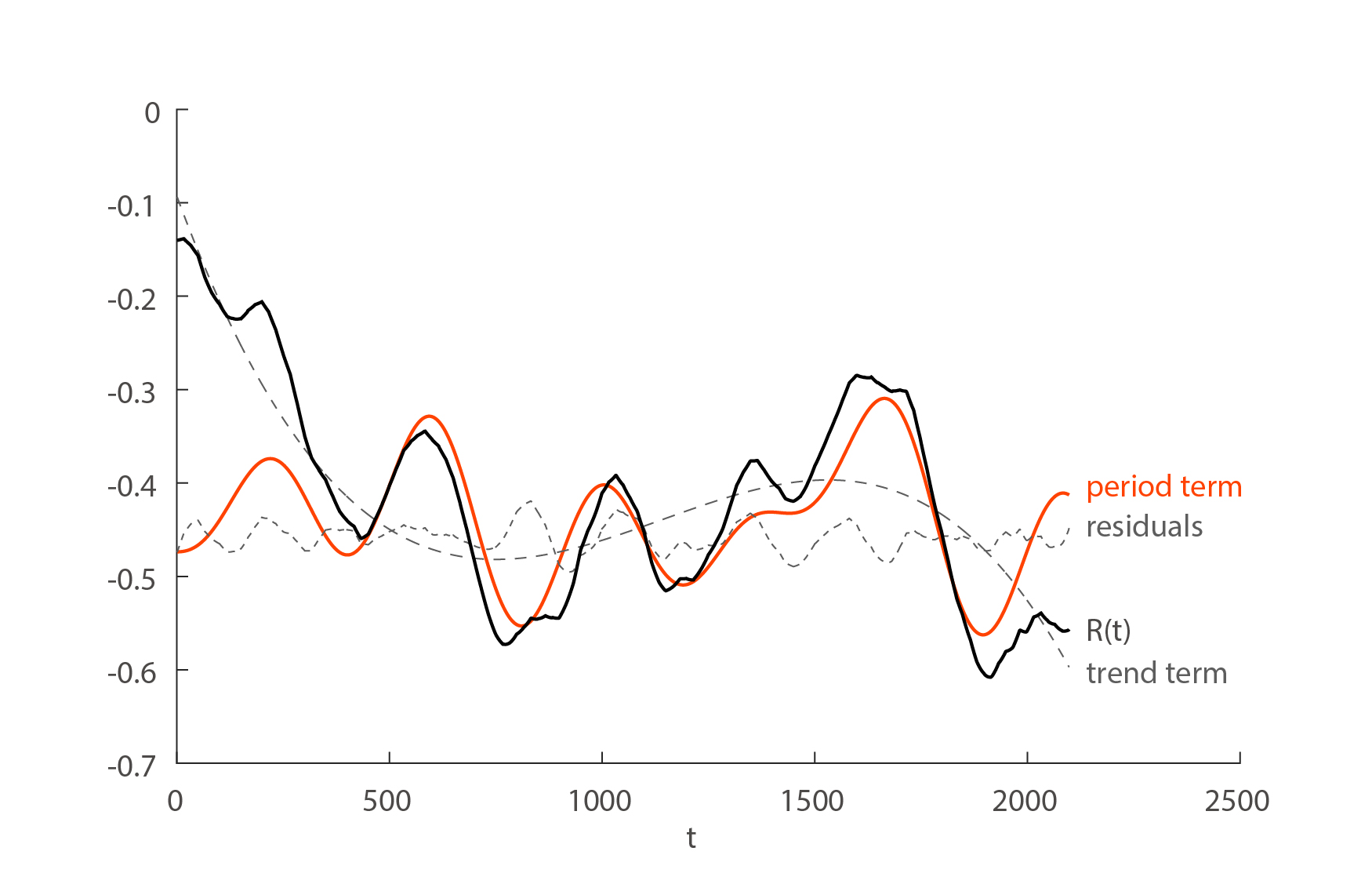
The R-t curve is drawn on the following graph. Another R-t curve based on the dithered data is drawn together to ensure the shape of the curve is not formed coincidentally.



%the R-t curves, the solid curve is based on the original data and the dotted curve is based on the dithered data

It can be observed from the graph that the IF index shows no remarkable correlation in the initial about 400ms. One noticeable feature of the curve is the periodic oscillation of the curve: successively reaching its local extreme value at 434ms, 584ms, 764ms, 1032ms, 1138ms, 1368ms, 1449ms, 1622ms and 1912ms, with coefficients of -0.4593, -0.3445, -0.5725, -0.3919, -0.5151, -0.3764, -0.4193, -0.2861, -0.6078. To better confirm and study this feature of oscillation, we use fitting tools to decompose the curve into trend term, periodic term and noise term. (Fourier spectrum analysis also works but has a lower precision on period than the fitting method) The trend term fits the original curve with a 3rd-ordered polynomial; the periodic term fits the residual of the trend term with the sum of 3 sine functions; the noise term is the residual of the periodic term.

The original correlation curve and the three terms are as follow.



Here we only focus on the periodic term whose analytic formula is f(x) =

a1\*sin(b1\*x+c1) + a2\*sin(b2\*x+c2) + a3\*sin(b3\*x+c3) +a4\*sin(b4\*x+c4)

with a1 =0.0498 b1 =0.01236

a2 =0.0533 b2 =0.01719

a3 =0.0498 b3 =0.005689

a4 =0.02096 b4 =0.008394

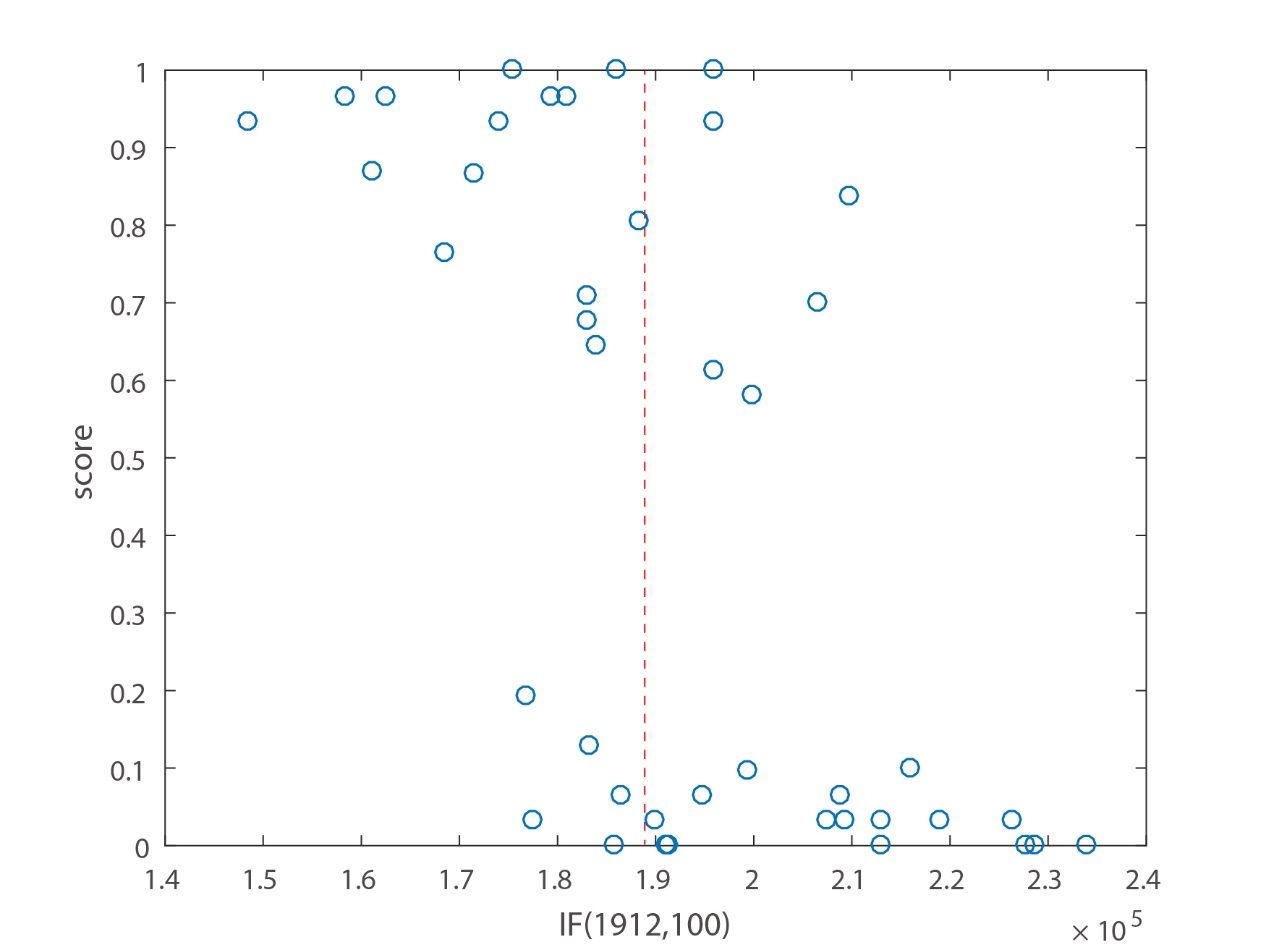
The three strongest periods are then 365.5ms, 508.3ms and 1104.4ms among which the 365.5ms period is the strongest and best observed. The strong oscillation feature shows that while viewing webpages with different visual appeal, people switch back and forth between a more indiscriminate condition and a more discriminate condition with a stable period time at about 365.5ms. The longer periods are weaker and exist fewer times for the limitation of the exposure time. The causing of the oscillation is worthy discovering and will be discussed in DISCUSSION.

* **Best-correlated situation**

According to the curve, the best correlation value is IF(1912,100),

|  |  |  |
| --- | --- | --- |
| index | ANOVA p | corr |
| IF(1912,100) | 0.00023 | -0.6048 |

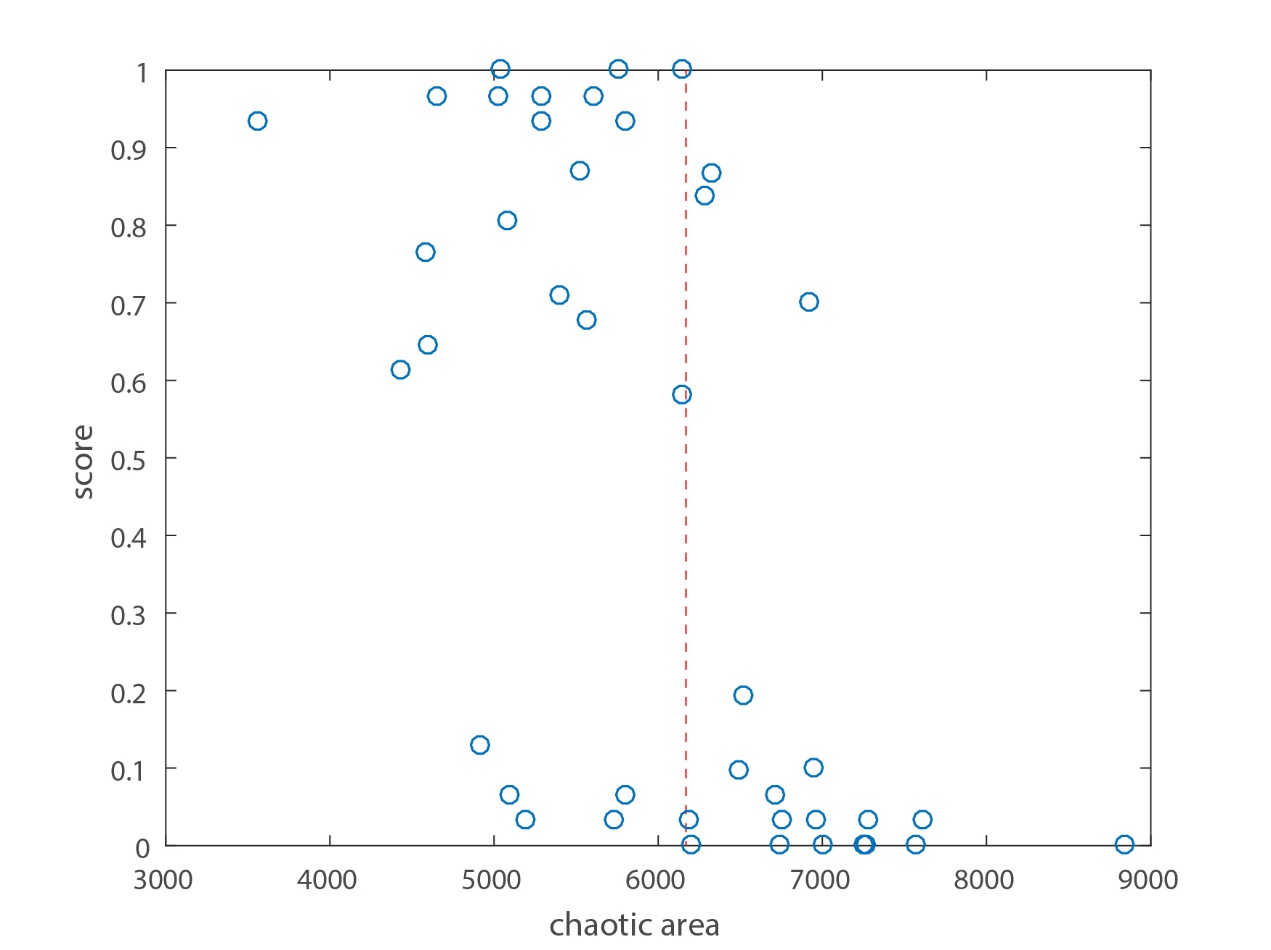
the figure with score is as followed.



%The IF(1912,100)-score figure of all the 42 webpages, the red line gives a best classification with an error rate of (6+5)/42

### Heat map chaotic areas:

It’s not surprising that the heat map peak value fails to show any remarkable relevance as visual appeal is not simply about drawing people’s attention to a few single points. In other words, if high peak value should indicate a high visual appeal, a single black dot on a pure background should no doubt be the best visual appeal stimulus.



%The chaotic area -score figure of all the 42 webpages, the red line gives a best classification with an error rate of (3+5)/42

The chaotic area which is calculated by count the nonzero low-stage area of a normalized heat map, however, shows a quite significant negative correlation with the score. The index is calculated by counting the area of the points whose value in the interval (0, 0.00047) of a normalized heat map. To be more mathematical:



The essence of the index is similar to IF, measuring how locally concentrated are the fixations, which is an essential dimension of the EPC. The strong negative correlation of the chaotic area proves that the less the fixation chaos is, the better the visual appeal of a webpage is.

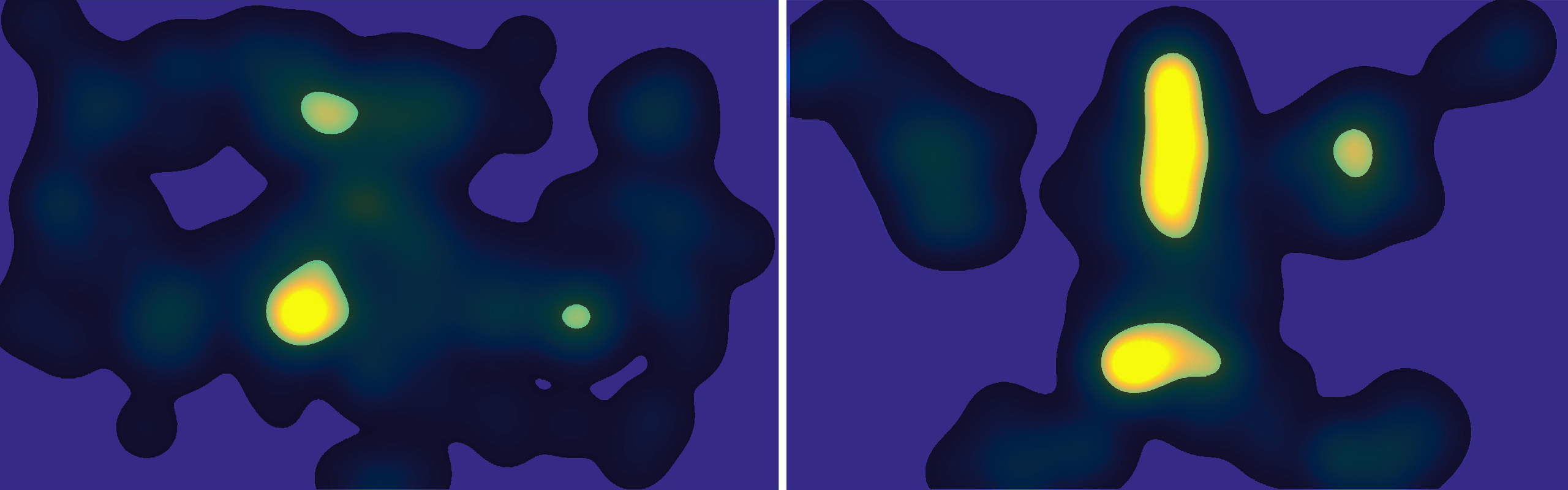
* **Robustness:**

The index is also calculated based on the dithered data to check its robustness.

|  |  |  |
| --- | --- | --- |
| index | ANOVA p | corr |
| Chaotic area | 5.89E-05 | -0.5936 |
| Chaotic area dither | 4.87E-05 | -0.5985 |

The result prove that even under a quite low precision recording situation, The index still shows a significant negative correlation with score.

Chaotic areas can be marked on heat maps for a visualization.

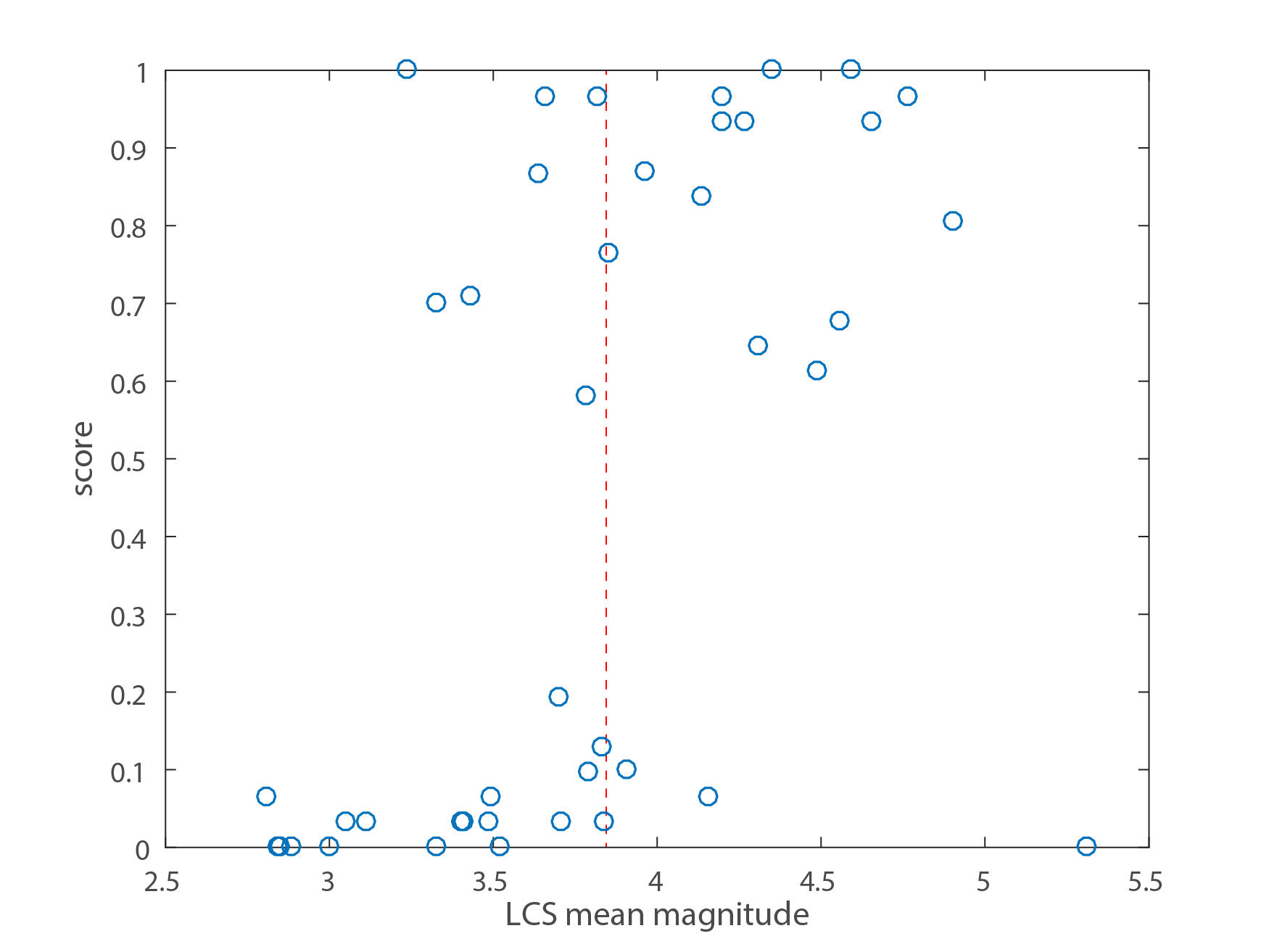


%the dark areas are the chaotic areas. The left heat map has an obviously larger chaotic area than the right one.

#### LCS value:

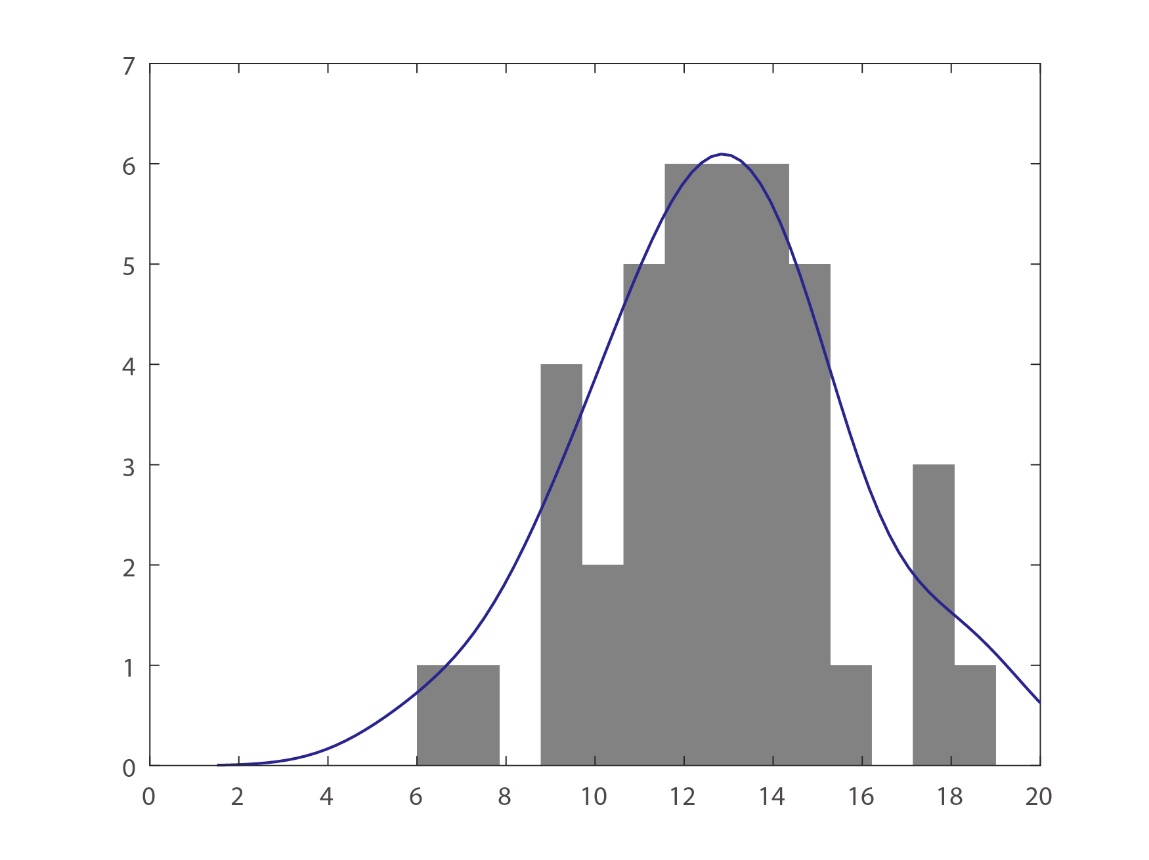
The significance of LCS value but the Levenshtein Distance indicates that substitution operation should cost more than removal and insertion operations. The following introduces the details of the calculation.

%The LCS-score figure of all the 42 webpages, the red line gives a best classification with an error rate of (7+2)/42



* **AOI cluster results**

The histogram of the amount of AOIs of the webpages.



|  |  |  |
| --- | --- | --- |
| index | ANOVA p | corr |
| AOI amount | 0.18593 | -0.2564 |

The amount of the AOIs shows no significant correlation to the score.

* **Statistical magnitudes**

For both of the two indexes related to edit distance, three statistical magnitudes are adopted to count the all 30\*29/2 pairs of values:

Mean: calculates the average of all 435 values.

MLE: MLE refers to Maximum Likelihood Estimation. Since the indexes calculated both have discrete probability spaces, MLE refs to the value which with the highest probability.

Max: For each AOI sequence, counts the maximum of all the 29 values related. Average those 30 values as the ‘Max’ magnitude

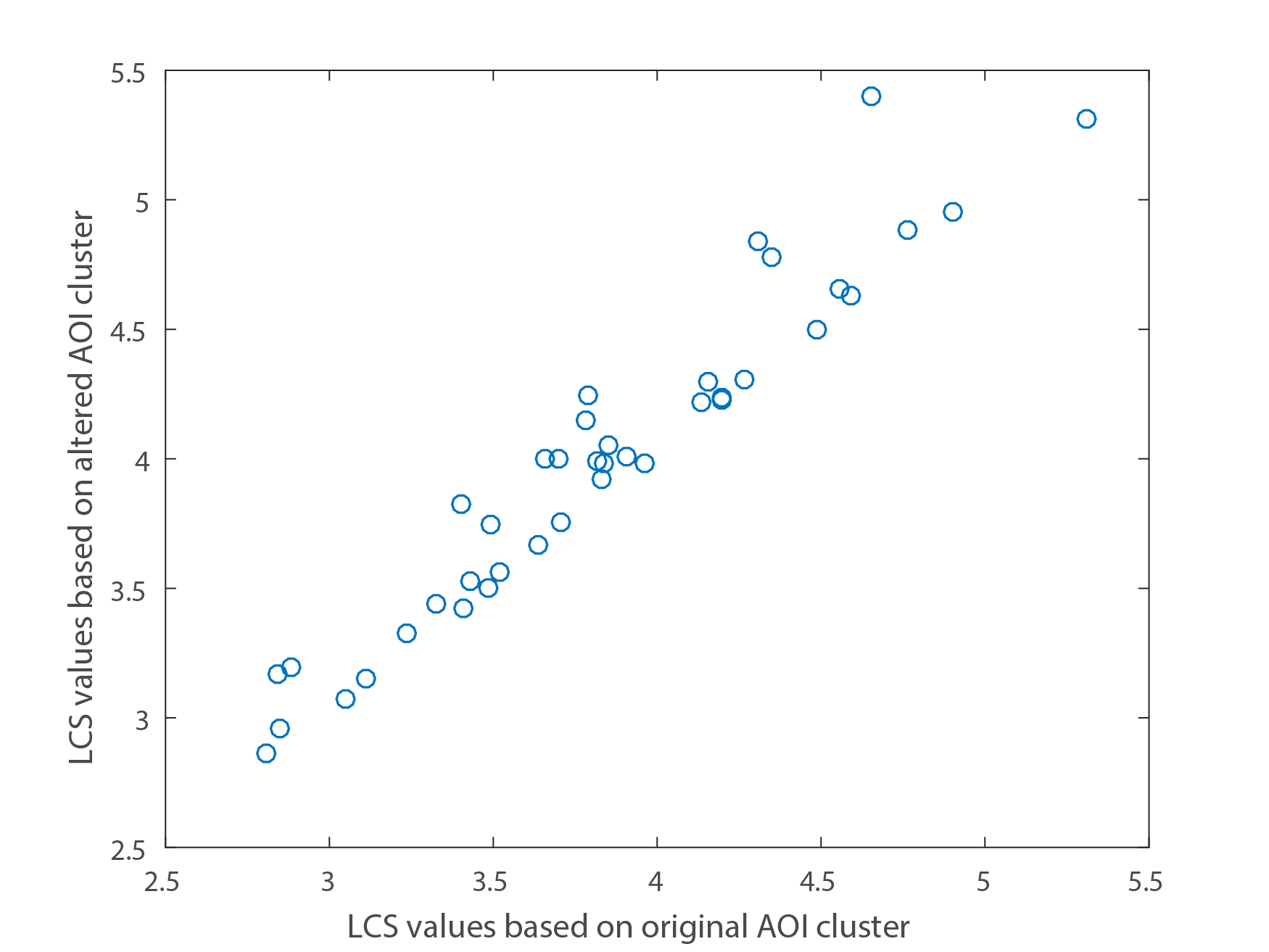
The experiment result shows that mean and max magnitudes are better score-correlated than the mle magnitude. In the following discussion, we select mean as the standard magnitude for LCS index.

* **Robustness to AOI cluster algorithm**

As is mentioned, the AOI cluster algorithm may have AOIs wrongly clustered. Mainly three types of mistakes may occur: clustering a minority area as an AOI; clustering the neighboring areas of different categories into one AOI (for example cluster a text area and an image area into one AOI); and clustering a complete area into two AOIs.

For the three types of possible mistakes, only the 2rd type is simulated for a robustness test. Because for the minority situation, its minority makes sure that few people have fixed on these AOIs, thus their existence can hardly cause variations to the LCS values. For the ‘one to two’ situation, as the algorithm treats the AOIs indiscriminately, the way of dividing one big area into two is with a reason, and is a more objective way of AOI marking, uncovering the essential details of the viewers’ eye path information.

The second situation however is often caused by close fixation locations, and may cause variations to LCS values. Here for every webpage we randomly select two neighboring AOIs, mark them with a same number, and calculate a new group of LCS values. The results are as followed.



% The original-AOI-cluster-based values vs the altered-AOI-cluster-based values.

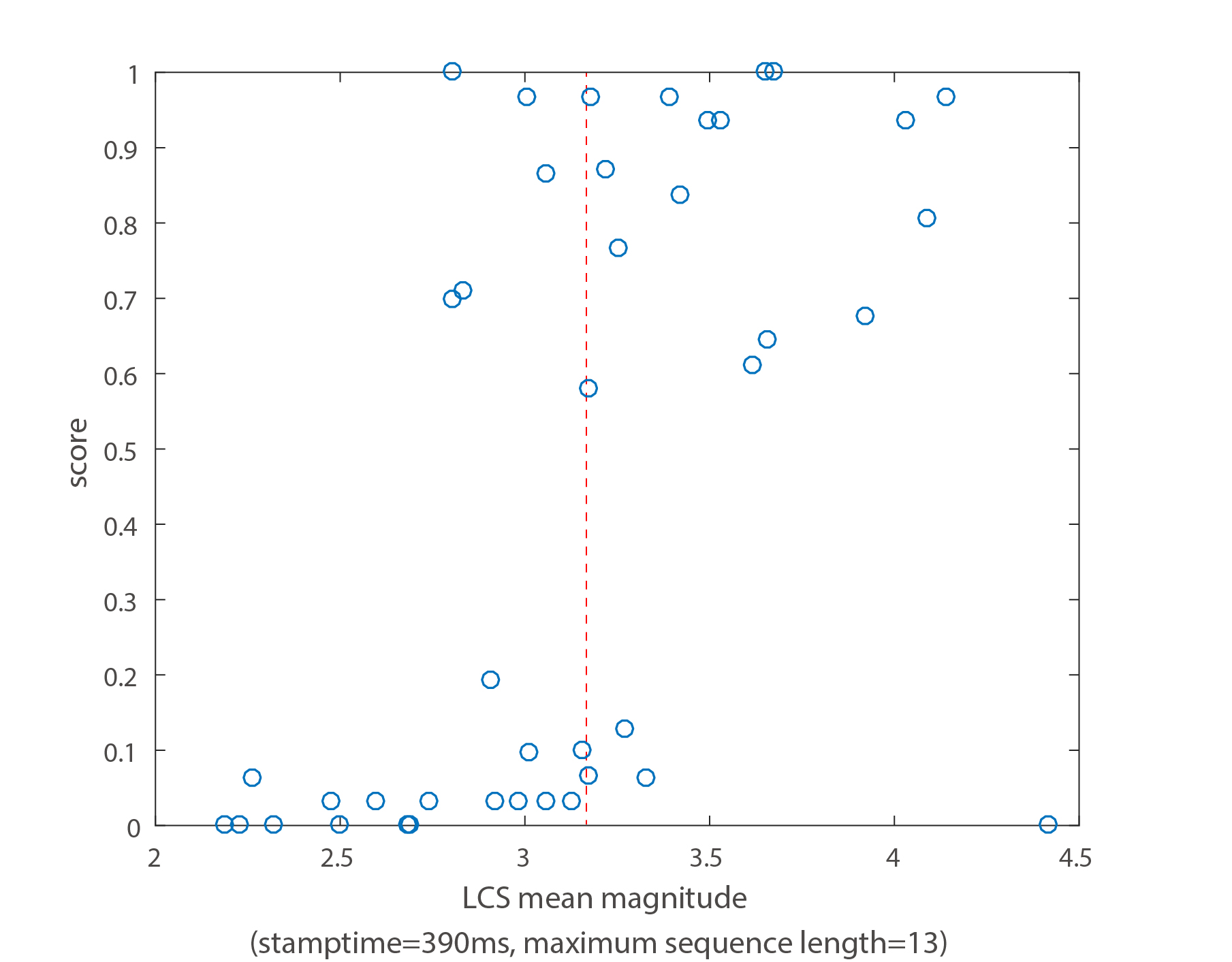
|  |  |  |
| --- | --- | --- |
| Index | ANOVA p | corr |
| LCS mean based on original AOI cluster | 0.0012 | 0.4952 |
| LCS mean based on altered AOI cluster | 0.0013 | 0.4906 |

The robustness the LCS value as well as its correlation can be verified.

* **Best correlation situation**

To find out which range of the AOI sequences shows the best correlation, we search the whole range of stamptimes and maximum sequence lengths. And reaches the best correlation at the stamptime of 390ms with a maximum sequence length of 13. The following table gives the ANOVA p and coefficient of the mean magnitude of the best LCS value. The dithered-data-based result is also shown to check the robustness.

|  |  |  |
| --- | --- | --- |
| index | ANOVA p | corr |
| LCS (stamptime=390ms, max sequence length=13) | 0.00049 | 0.5241 |
| LCS based on altered AOI cluster | 0.00043 | 0.5380 |



%The best correlated LCS-score figure of all the 42 webpages, the red line gives a best classification with an error rate of (5+3)/42

Similar to the correlation-by-time result of the IF index, the initial about 400ms again shows no remarkable correlation with the score. The cause of this will be discussed in DISCUSSION.

### Mutual correlations and Classification error rate

For the three indexes that show significant correlations with the scores, their mutual correlation coefficients are calculated (here we use the full-exposure-time values instead of the best-correlated-situation values to avoid specific cases):

|  |  |  |  |
| --- | --- | --- | --- |
|  | IF(0,maxdt) | Chaotic area | Full-sequence LCS |
| IF(0,maxdt) | 1 | 0.848 | -0.716 |
| Chaotic area | 0.848 | 1 | -0.618 |
| Full-sequence LCS | -0.716 | -0.618 | 1 |

The IF and chaotic area are more closely correlated as their essence are both the measuring of the consistency of fixation distributions, while the LCS index concerns the consistency of AOI transitions.

The classification error table of the logistic regression produced by the three indexes.

|  |  |  |  |
| --- | --- | --- | --- |
|  | prediction | good | bad |
| class |  |
| good | | 18 | 3 |
| bad | | 4 | 17 |

The classification rate is just a little better than the respective ones, indicating a high overlapping among the indexes.

## Discussion:

### Chaos in the initial 400ms

The correlation curves of the IF and the best-correlated LCS both show that the eye path property in the initial 400ms is quite different from later. The result strongly suggests that people’s eye treat the bad and good webpages indiscriminately in the initial about a half second.

The result may seem inconsistent with the 50ms theory. However the judgement on visual appeal discussed in Lingaard’s study is not behavioral. And it’s not affirmed yet how the 50ms works physically: Is it a duration enabling people to snapshot the webpage and make judgements later based on the impression or is it a duration in which people’s cortex or conscience could give different react towards bad pages from good ones.

What can be sure in our study is that it is not until 400ms can our visual behavior tell the possible good or bad visual appeal of the webpages through EPC indexes. The initial 400ms then could be regarded as a kind of chaos or probe scanning.

### Correlation oscillation and visual behavior

The immanent cause of the periodic oscillation of the correlation curve of IF is worthy discussion. The oscillation shows that people are switching back and forth between a more indiscriminate condition to a more discriminate condition with an almost fixed period of about 365.5ms.

The cause of the oscillation could probably be the human visual behavior of planning and focusing. Human vision are composed by the foveal vision which has high resolution but a tiny visual field and the peripheral vision which has a large visual field but quite low resolution. While planning, we use peripheral vision to evaluate the surroundings and decide where to focus next. While focusing, we use foveal vision to read the information on the focus.

Comparing to the focusing process, the planning process may have a more uncertain focus location. Thus with planning and focusing process going alternately, the correlation curve of IF shows an oscillation feature.

For viewing a webpage, the easier the planning process, the lower is the consistency, the higher is the EPC, and the better is the visual appeal of the webpage.

#### EPC and Visual appeal

EPC, as a concept, has only theoretical correlation with visual appeal. With indexes measuring EPC in different aspects, we are able to study specifically what dimensions of EPC are visual appeal correlated and discover further principles for webpage design. Although the visual behavior of an individual is subjective, the EPC indexes values of a webpage are rather objective as they reflect collective behavior of human vision.

The nonsignificance of EF value indicates that visual appeal is not strongly related to the overall space-time distribution of the fixations; the low correlations of the heat map peak value implies that a single strong fixation concentration won’t definitely result in a good webpage design; the result of the F value indicates that visual appeal is not simply about the certainty of AOI transitions; And the non-significance of the Levenshtein Distance implies that for AOI sequences, substitutions cost more than the insertions and deletions.

For the significant indexes, the negative correlations of IF and chaotic area tell that a good webpage is able to catch the viewers’ attention sequentially to a series of explicit focuses. While the LCS index shows that a positive visual appeal a webpage is able to guide the viewers’ eye paths in a way not to a specific path, but to a common path. Just like people passing through a grass lawn, they may walk in different paths, but there is a common path on which they all pass through.

## Conclusions:

Through the experiment and the analysis of the result, we proved the several specific aspects of EPC, eye path consistency, being significantly correlated with visual appeal. They are, the IF (island function) measuring the local space-time consistency of fixations, the heat map chaotic area measuring the proportion of chaotic fixation area and the LCS value (Longest Common Subsequence) which measures the consistency of the AOI sequences.

We give powerful support to the hypothesis that the less uncertainty a visual process has, the better is the visual appeal, also to the Fluency Theory which argued that the easier the visual process is, the better is the visual appeal. According to our research, a good webpage design should have the following features: (值得商榷)

* Being able to lead the viewers’ attention sequentially to a series of explicit focuses in the initial exposure time (in this study, about 2100ms)
* Being able to guide the viewers’ eye path to a common path (passing throw a series of AOIs sequentially) in the initial seconds.

We also discovered the probably visual processing behavior of human beings:

* In the initial about 400ms of the exposure time our eyes are doing a rather random or chaotic probe scanning as a preliminary planning.
* From about 400ms of the exposure time our eyes start switching back and forth between planning mode and focusing mode alternately to scan the whole webpage. This is also the range of time when webpages with good visual appeal can be distinguished from the bad ones through EPC indexes.
* The EPC decreases during the exposure, with individual differences of visual habits increase.

The result of the study may have applications on visual theories, design principles and qualification of the visual appeal of webpages. Still, further research is required to thoroughly study the relations between human visual behavior and visual appeal.

## References:

Complexity

One delicate aspect of the LCS index is that it counts in the complexity of the webpages to some extent. A webpage whose complexity is quite low will generate short AOI sequences (causing by long fixations and noisy fixations which belong to no AOIs), based on which the LCS index is very probably to be high. Though not emphasis in this study, we believe that a sufficient complexity is a premise to visual appeal.