A Controlled User Study on Human Perception of Web Search Latency *

Ioannis Arapakis
Yahoo Labs Barcelona
arapakis@yahoo-inc.com

Xiao Bai
Yahoo Labs Barcelona
xbai@yahoo-inc.com

B. Barla Cambazoglu
Yahoo Labs Barcelona
barla@yahoo-inc.com

Abstract

We conduct two controlled user studies that reveal how response latency affects the user behavior in web search. The first study examines users' sensitivity to different levels of latency as well as their perception of response time. The second study demonstrates the effects of increasing response latency on the search experience and, more specifically, on user engagement and satisfaction. As a side contribution, we also look at potential bias due to search site branding.

1. Introduction

In practice, serving search results at the right speed is of vital importance to a commercial web search engine. Serving search results too slow or too fast both may result in certain financial consequences for the search engine. On the user side, the new generation of web users are impatient and have limited time. They expect subsecond response times from a search engine upon submission of their queries. High response latency is known to distract users and cause them to issue fewer queries than usual, decreasing users' engagement with the search engine in the long term [7]. This, in turn, can make a negative impact on the advertising revenue of the search engine. On the search engine side, commercial web search companies are known to make major investments in hardware infrastructures to cope with the growth of the Web as well as the growth of their user bases and query volumes, essentially trying to maintain their query response times at reasonable levels. These investments incur a financial burden on search engine companies and may even result in financial losses if the reduction attained in query response times due to these investments does not have any positive impact on the search experience of users.

In this paper, we aim to evaluate the potential impact of response latency on users' search behavior. To this end, we conduct a controlled user study, where we show users search results retrieved with varying response times from two different search engines (Yahoo and Google). The selected findings of our work are the following.

- The users of the fast search engine (Google) are more likely to notice the added delays than the users of the slow search engine (Yahoo).
- As long as the delay added to a response remains under 500ms, users cannot distinguish between a delayed response and a regular response with no added delay.
- When the introduced delay is larger than 1000ms, users are highly likely to notice the presence of delay.

2. User Sensitivity to Latency

Experimental design. The experiment used a repeated-measures design with two independent variables: search latency (with 12 levels in milliseconds: "0", "250", "500", "750", "1000", "1250", "1500", "1750", "2000", "2250", "2500", "2750") and search site speed (with two levels: "slow", "fast"). The search latency was controlled by using a client-side script that adjusted search latency by a desired amount of delay. The search site speed was controlled by using either a search engine with generally slow response rate (this is Yahoo, referred to as SE_{slow} in the paper) or a search engine with a generally fast response rate (this is Google, referred to as SE_{fast} in the paper). Although the two search engines were different, the returned search results were very similar due to the nature of queries used (see Procedure). The dependent variables were (i) sensitivity to search latency and (ii) prediction accuracy of search latency.

The scatter plot in Fig. 1 shows the response latency values observed for SE_{slow} and SE_{fast} upon submission of identical queries. We observe SE_{slow} to be somewhat slower than SE_{fast}. For almost any query, SE_{fast} has lower latency.

Apparatus. In our experiment, we used a desktop computer equipped with a $24''$ LCD monitor, keyboard, and mouse. In the background, we ran a custom-made JavaScript that controlled the search latency. The script was deployed using the Greasemonkey[1] extension in a Mozilla Firefox web browser. It captured a series of browser events (e.g., mouseover, click, or keypress) and logged the Unix timestamps for every query submitted and each search engine result page (SERP) rendered in response to a query.

* This work is part of a paper published in SIGIR'14. The remaining part on query log analysis was submitted as another TechPulse'14 paper.

[1] http://www.greasespot.net

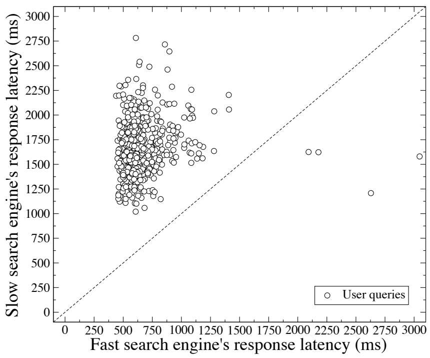

Figure 1. Response latency values attained by the fast and slow search engines for the same query.

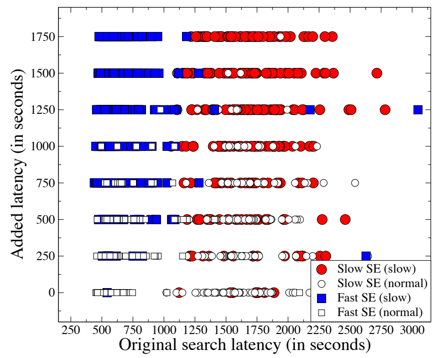

Figure 2. Perceived latency versus the original latency (each point represents a query).

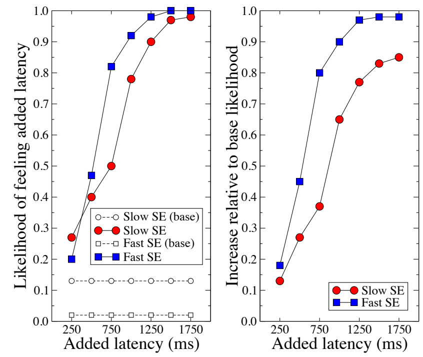

Figure 3. The likelihood of participants to feel increasing values of added latency.

Questionnaires. At the beginning of the study, the participants were asked to fill in an entry questionnaire, which gathered background and demographic information, as well as information about their previous experience with online search. A set of scales was developed specifically for our study (e.g., easy/difficult, relaxing/stressful, and satisfying/frustrating) based on users' response to the statement "Using a search site is generally...".

Participants. There were 12 participants (female=6, male=6) aged from 24 to 41 and free from any obvious physical or sensory impairment. The participants were of mixed ethnicity (Catalan, Chinese, Italian, German, Greek, Korean, Persian), came from a variety of educational backgrounds (41.6% had an MSc degree and 58.3% had a PhD degree), and were all proficient with the English language (8% intermediate level, 75% advanced level, 17% native speakers). They were primarily pursuing further studies while working (54.3%) although there were a number of students (33.3%) and full-time employees (16.6%). Participants reported using a search site at home or work very often ($M = 6.58, SE = .79$). In addition, they indicated that they find online searching a very easy ($M = 6.00, SE = 1.53$) and somewhat satisfying ($M = 5.50, SE = 1.16$) task.

Procedure. The user study was carried out in a laboratory setting and followed a think-aloud protocol. At the beginning of each session, the participants were informed about the conditions of the experiment and were asked to complete a demographics questionnaire. Each participant then performed two tasks. Both tasks involved submitting a fixed number of randomly selected navigational queries, i.e., queries that seek a single website or web page of a single entity (the web domain list was created using the web analytics provided by Alexa[2]). We limited the study to navigational queries because they impose a smaller cognitive load to the searcher (compared to other types of queries), promote a convergence in the search intent across all users, and do not require native-level knowledge of the English language. Therefore, by mitigating the effort of query formulation, our participants were able to assess the latency effect better.

The first task asked the participants to report to the experimenter their subjective impression of the search site's response latency after each submitted query, i.e., whether they felt that the response was "slow" or "normal". In this task, the search latency was increased by a fixed amount that ranged from 0 to 1750ms, using a step of 250ms. Each latency value (0ms, 250ms, ..., 1750ms) was introduced five times and in a random order, in combination with 40 randomly selected navigational queries. The provided navigational queries were submitted to the search site the same way they would be submitted in a realistic search scenario, i.e., through typing and clicking.

The second task required the participants to provide an estimation of the search latency in milliseconds for each submitted query. Participants were instructed to consider as search latency the time from the query submission until the rendering of SERP. The search latency was set to a fixed value that ranged from 500ms to 2750ms, using a step of 250ms. Similar to the previous task, each latency value was introduced five times in a random order, in combination with 50 navigational queries. To familiarize themselves with the default behavior of the search site and establish a measure of comparison, the participants were asked to submit a set of training queries before each task. Finally, to control for order effects, the task assignment was randomized.

Results (first task). Fig. 2 shows the distribution of cases where the participants felt that the response was slow or normal. Based on that plot, Fig. 3 shows the likelihood that the participants will feel the added delay in response time. In case of SE_{fast}, when there was no added delay, the participants could almost always feel the absence of delay (with $1 - 0.02 = 0.98$ probability). In case of SE_{slow}, however, their accuracy was considerably lower ($1 - 0.13 = 0.87$ probability), potentially due to the high variation in response time of SE_{slow}. In general, participants could distinguish slow response with much higher likelihood when they were

[2] `http://www.alexa.com/topsites`

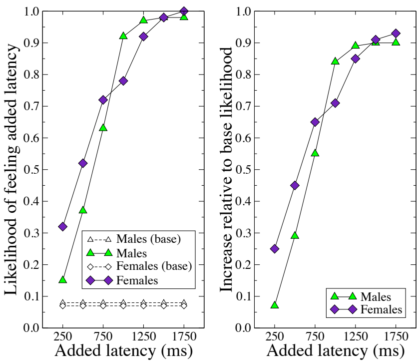

Figure 4. Impact of gender on the likelihood of feeling added latency.

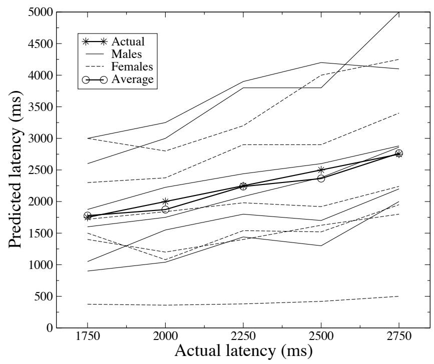

Figure 5. Latency predictions in case of SE_{slow}.

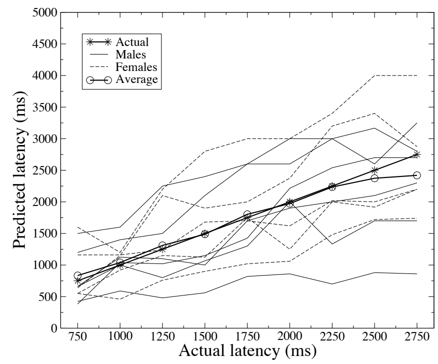

Figure 6. Latency predictions in case of SE_{fast}.

using SE_{fast}. For example, when the added latency was 750ms, the likelihood of participants to feel the added latency was not different than random in case of SE_{slow}, but they were able to notice the added latency with much higher likelihood (around 0.82 probability) in case of SE_{fast}. For both search engines, added delays under 500ms were not easily noticeable by participants (not better than random prediction) while added delays above 1000ms could be noticed with very high likelihood. Fig. 4 displays similar data, but this time comparing male and female participants. According to the figure, female participants are observed to be better in noticing small increases in response time than male participants. But, there is no significant difference between males and females when the added delays are large.

Results (second task). In Figs. 5 and 6, we show the predicted versus actual latency values for individual participants using SE_{slow} and SE_{fast}, respectively. The results reveal considerable differences in the way individuals perceive the latency. In case of SE_{slow}, about half of the participants consistently overestimated the latency while the other half consistently underestimated it. The prediction quality of participants have higher deviation in case of SE_{fast} than in case of SE_{slow}. Interestingly, the average of all participants' predictions are very close to the actual values in both cases.

3. Impact of Latency on Search Experience

The objective of this study is to investigate the effects of response latency on the search experience and, in particular, on user engagement and satisfaction. Two psychometric scales were used to capture hedonic and cognitive aspects of the user experience: the User Engagement Scale (USE) and IBM's Computer System Usability Questionnaire (CSUQ). In addition to the psychometric scales, participants were asked to evaluate the performance and speed of the search site, as well as report the experienced frustration after each task. We speculate that, as the search latency increases, the search experience will become less engaging (i.e., low scores on all psychometric scales) and the perceived usability of the search site will be negatively impacted.

Experimental design. The experiment had a two-way, mixed design. The related measures independent variable was the search latency (with four levels in milliseconds: "0", "750", "1250", "1750"). The unrelated measures independent variables was the search site speed (with two levels: "slow", "fast"). Search latency was controlled through a client-side script that adjusted the latency by a desired amount of delay. The choice of latency values was informed by the findings from the first study (see Section 2). The search site speed was controlled by using either a search site with a generally slow response rate (SE_{slow}) or a search site with a generally fast response rate (SE_{fast}). Despite the two search sites coming from different brands, the returned results were almost identical due to the nature of the search queries used (see Procedure). The dependent variables were (i) experienced positive and negative affect, (ii) level of focused attention, (iii) perceived system usability, and (iv) subjective beliefs about search site performance.

Apparatus. The study used the setup in Section 2.

Questionnaires. We used two types of questionnaires. The first questionnaire (entry) was introduced at the beginning of the study and gathered background and demographic information, as well as information about previous experience with online search. The second questionnaire (main) was administered at post-task and included the USE and CSUQ scales. The questions were all forced-choice type and appeared in a random sequence to mitigate potential bias due to the ordering effect. The UES is multi-dimensional; its items pertain to positive and negative affect, perceived usability of the system, as well as users' felt involvement and focused attention during the task. Affect refers to the emotion mechanisms that influence our everyday interactions and can act as the primary motivation for sustaining our engagement [6] during information processing tasks or computer-mediated activities. Focused attention refers to the feeling of energised focus and total involvement, often accompanied by loss of awareness of the outside world and distortions in the subjective perception of time. The CSUQ [4] is a multi-dimensional user satisfaction questionnaire. Out

Table 1. I-PANAS-SF

Positive Affect items	Negative Affect items
active	afraid
alert	ashamed
attentive	hostile
determined	nervous
inspired	upset

Table 2. Focused attention scale

1. I forgot about my immediate surroundings while performing this search task.
2. I was so involved in my search task that I ignored everything around me.
3. I lost myself in this search experience.
4. I was so involved in my search task that I lost track of time.
5. I blocked out things around me when I was completing the search task.
6. When I was performing this search task, I lost track of this world around me.
7. The time I spent performing the search task just slipped away.
8. I was absorbed in my search task.
9. During this search task experience I let myself go.

of the four items it consists, we considered only the scores from the responses to system usefulness (SYSUSE). Taken together, the UES and CSUQ probe users' perceptions of the pragmatic and hedonic qualities of their search interactions, as well as their perceptions of the search engine and of themselves using a technology, all of which are considered key facets of the user experience [3]. More in specific, the questionnaires inquired about the following aspects:

I-PANAS-SF. The international Positive and Negative Affect Schedule (PANAS) Short Form was used to measure the affect before and after each task (Table 1). I-PANAS-SF is a validated test for measuring affect changes. It includes ten items measuring positive (PAS) and negative (NAS) affect. Participants were asked to respond on a 7-point Likert scale (very slightly or not at all; a little; moderately; quite a bit; extremely) their agreement to the statement: "You feel this way right now, that is, at the present moment", for each item. Although I-PANAS-SF may not be as efficient and accurate for capturing temporal micro-resolutions of emotional responses, there are examples of studies from the domain of Library & Information Science [6] where PANAS has been successfully applied for measuring searchers' affect between search tasks. Considering that the duration of our search tasks is comparable to those in the aforementioned studies, we believe that our experimental approach to measuring emotion was reasonably accurate.

Focused attention. A 9-item focused attention subscale, part of a larger scale for measuring user engagement [6], was adapted to the context of the search tasks. The focused attention subscale has been used in past work [5] to evaluate users' perceptions of time passing and their degree of awareness about what took place outside of their interaction with the given task. Given the context of our work, focused attention was a more meaningful dimension, at least compared to other subscales of engagement (e.g., aesthetics, novelty) that were not relevant enough or were addressed by the other questionnaires employed in our study (USE, CSUQ, i-PANAS-SF). To measure focused attention, the participants were instructed to report on a 7-point Likert scale (strongly agree; disagree; neither agree nor disagree; agree; strongly agree) their agreement to each item shown in Table 2.

System usability. The CSUQ [4] was developed by IBM for measuring the perceived usability of systems in the context of realistic scenarios. A 7-point Likert scale of agreement (strongly agree; strongly disagree) that quantifies system usefulness is used for each of the 8 statements in the SYSUSE subscale. Two examples statements are "I can complete my work quickly using this search site" and "I can efficiently complete my work using this search site".

Custom statements. In addition to the USE and CSUQ-SYSUSE scales, we gathered information about the search sites' performance. We used a 7-point Likert scale of agreement for the following positive statements: (i) "This search site was fast in responding to my queries", (ii) "This search site helped me to accomplish my task in a reasonable amount of time", and (iii) "I feel satisfied with the retrieved results". Moreover, we asked our participants to indicate on a 7-point Likert scale how frustrating each search task was.

Demographics. This study gathered the same demographics as those discussed in Section 2.

Participants. There were 20 participants (female=10, male=10) aged from 18 to 41 and free from any obvious physical or sensory impairment. The participants were of mixed ethnicity (Dutch, English, Farsi, French, German, Greek, Italian, Korean, Persian, Spanish, Turkish, Urdu), came from a variety of educational backgrounds (10% had a BSc degree, 50% had an MSc degree and 40% had a PhD degree), and were all proficient with the English language (10% intermediate level, 70% advanced level, 20% native speakers). They were primarily pursuing further studies while working (40%) although there were a number of students (35%) and full-time employees (25%). Participants reported using a search site at home or work very often ($M = 6.85, SE = 0.36$). In addition, they indicated that they find online searching an easy ($M = 5.75, SE = 1.91$) and somewhat satisfying ($M = 5.30, SE = 0.86$) task.

Procedure. The user study was carried out in a laboratory setting. At the beginning of each session, the participants were informed about the conditions of the experiment and were asked to complete a demographics questionnaire. Each participant had to perform four search tasks (one for each latency value). The tasks were presented in the context of a short cover story, which asked the participants to evaluate the performance of four different backend search systems. All tasks involved submitting out of a list of 200 web domains as many navigational queries as possible, within ten minutes. Participants were presented with two web browser windows: the first window displayed the search site while the second window displayed the the questionnaire. For each navigational query, participants were instructed to locate the associated URL among the first ten results of the SERP and copy-paste it in the corresponding box of the questionnaire.

Table 3. Descriptive statistics (Mean, SD) for reported UE and CSUQ-SYSUSE scales

	SE_{slow} latency				SE_{fast} latency			
	0ms	750ms	1250ms	1750ms	0ms	750ms	1250ms	1750ms
postPAS	16.20 ± 9.04	14.50 ± 7.59	15.50 ± 7.21	15.20 ± 7.47	20.50 ± 7.82	19.00 ± 9.01	20.80 ± 9.48	19.30 ± 8.23
postNAS	7.00 ± 3.80	6.80 ± 2.70	7.60 ± 3.27	6.90 ± 3.28	6.80 ± 2.44	7.40 ± 3.03	7.40 ± 2.72	7.20 ± 2.49
postPAS-prePAS	-3.10 ± 8.49	-4.80 ± 6.46	-3.80 ± 6.34	-4.10 ± 7.11	2.50 ± 5.95	1.00 ± 6.13	2.80 ± 6.01	1.30 ± 6.29
postNAS-preNAS	0.30 ± 2.31	0.10 ± 1.10	0.90 ± 1.79	0.20 ± 2.30	-0.40 ± 2.46	0.20 ± 2.53	0.20 ± 2.74	0.00 ± 1.33
Frustration	3.20 ± 2.20	3.10 ± 2.02	2.90 ± 2.02	3.30 ± 2.21	2.80 ± 1.40	3.00 ± 1.63	3.50 ± 1.08	2.60 ± 0.84
FA	22.80 ± 9.37	22.90 ± 8.29	19.90 ± 9.26	22.20 ± 10.38	27.90 ± 13.20	26.60 ± 10.41	23.90 ± 9.23	29.50 ± 9.85
SYSUS	32.80 ± 6.73	28.90 ± 5.40	29.80 ± 7.63	27.90 ± 6.89	35.20 ± 5.35	31.30 ± 8.25	29.80 ± 8.34	33.20 ± 8.22

A set of training queries was used at pre-task to allow participants to familiarize themselves with the "default" behavior of the search site and the search task. To provide further motivation and engage the participants with the task, they were informed that a prize would be awarded to the person who will submit the most URLs in total. To control the order effects, the task assignment was randomized. Finally, the participants were randomly allocated to two groups, ensuring an even number of female and male participants per group.

Results. We present the findings based on 80 search tasks, carried out by 20 participants. For our analysis we used several related and unrelated measures tests, like the Mann-Whitney and Wilcoxon Signed-Rank test for pair-wise comparisons, and Friedman's ANOVA for three or more conditions. Participants response to the 5-item PAS, 5-item NAS, 9-item focused attention, and 8-item CSUQ-SYSUSE scales were summed to obtain the final scores. Results are reported at a statistical significance level of .05. To take an appropriate control of Type I errors in multiple pair-wise comparisons we applied the Bonferroni correction.

Experienced affect. Table 3 (top) shows the mean scores for the positive (postPAS) and negative (postNAS) affect scale at post-task, as well as the difference Δs between the scores reported at pre- and post-task for SE_{slow} and SE_{fast}. The results indicate a decrease in positive affect for both search sites as we introduce larger latency values. The inverse effect is observed for negative affect, which increases as higher latency values are used, but this effect is more consistent in the case of SE_{fast}. None of the differences identified above were statistically significant. However, when comparing the reported postPAS and postNAS scores between SE_{slow} ($Mdn = 16.50$) and SE_{fast} ($Mdn = 21.00$) and across all latency values, the Mann-Whitney test indicated a statistically significant difference for postPAS, $U = 550.50, p < .05, r = -.31$. This small to medium effect observed for PAS between the two search sites suggests a positive bias towards SE_{fast}, despite participants having experienced the same range of added latencies. Table 3 also displays the mean scores for reported level of frustration. There were no differences among the latency values, nor between the two search sites.

Focused attention. Table 3 (middle) displays the mean scores for focused attention (FA). For the participants of SE_{slow}, the variation of the scores across the latency values does not indicate any visible trend. For the participants of SE_{fast}, we observe a decrease in small- and medium-size latencies. However, there were no significant differences between the latency values. When comparing the reported focused attention between the participants of SE_{slow} ($Mdn = 21.00$) and SE_{fast} ($Mdn = 26.00$), and across all latency values, the Mann-Whitney test indicated a significant difference, $U = 568.50, p < .05, r = -.27$. This implies a small to medium effect for the focused attention observed between the two search sites. It also suggests that the participants of SE_{fast} felt more deeply involved with the search task, despite having experienced the same range of added latencies.

System usability. Table 3 (bottom) displays the mean CSUQ-SYSUSE scores per latency value and per search site. For both search sites we observe a noticeable increase in the reported usability scores. More in specific, for SE_{slow}, there was a statistically significant difference in the perceived usability of the search site depending on which amount of added latency was introduced, $\chi^2(3) = 11.00, p < .05$. Post-hoc analysis with Wilcoxon Signed-Rank test indicated a statistically significant difference in the perceived usability, as reported scores were significantly higher for latency value of "0" ($Mdn = 31.00$) compared to "1750" ($Mdn = 28.50$), $Z = -2.66, p < .008, r = -0.42$. This represents a large effect in the levels of perceived usability when search latency was increased by 1750ms. No significant differences were observed for SE_{fast}, suggesting that the participants were more tolerant towards the delays experienced for that search site despite the large latency values introduced to their search interactions. Additionally, the reported scores for perceived usability differed significantly between the participants of SE_{slow} ($Mdn = 30.00$) and SE_{fast} ($Mdn = 35.00$), $U = 596.00, p < .05, r = -.22$. Finally, none of the differences identified in the number of submitted queries per latency value were significant.

Search experience. We evaluated the search experience promoted by the two search sites by asking our participants to report their agreement to a set of custom statements. With respect to statement (i), the Friedman's ANOVA test indicated for SE_{slow} a significant difference in the perceived search site speed, depending on which latency value was added. Wilcoxon tests were used to follow up this finding but no significant differences were observed for any of the pair-wise comparisons. Furthermore, the reported perceived

Table 4. Correlations of subjective beliefs on search site performance and reported UE and CSUQ-SYSUSE scales

Beliefs	postPAS	postNAS	Focused attention	CSUQ-SYSUS	custom-1	custom-2	custom-3
SE_{slow} will respond fast to my queries	$.455^{**}$	.041	$.702^{**}$	.267	.177	.177	.082
SE_{slow} will provide relevant results	.262	$-.083$	$.720^{**}$	$.411^{**}$	.278	.263	.232
SE_{fast} will respond fast to my queries	$-.051^{**}$	.245	$.341^{*}$	$.591^{**}$	$.330^{*}$	$.443^{**}$	$.624^{**}$
SE_{fast} will provide relevant results	$-.272$	.133	$-.133$	$.378^{*}$	.212	.259	$.390^{*}$

*. Correlation is significant at the .05 level (2-tailed). **. Correlation is significant at the .01 level (2-tailed).

search site speed by participants of SE_{slow} did not differ significantly from that of participants of SE_{fast}, which is an interesting finding considering the notable difference in the search sites' performance. In regards to statement (ii), participants' belief that the search site helped them accomplish their task more quickly changed significantly over the latency values ($\chi^3 = 10.80, p < .05$). This effect was observed only for SE_{slow}. Post hoc tests revealed a significant difference between the latency values "0" and "1750", $Z = -2.63, p < .008, r = -.83$. Finally, for statement (iii), none of the differences identified in the reported scores were statistically significant across the latencies and search sites.

These results help us understand that the subjective search experience may be influenced by branding, as well as users' preconceptions about the search site performance. For example, a search site perceived as "fast" or "efficient" may still result in engaging search interactions despite occasional poor performance. This suggests that a successful marketing approach could go a long way to improve the reputation of a product and positively bias the end-users.

Correlation analysis of all factors. Finally, we report the results of a correlation analysis performed across all search experience factors discussed above, including participants' prior beliefs of the search site performance. The importance of this analysis is to understand better the influence of subjective beliefs on the hedonic and cognitive aspects on the search experience. Table 4 shows all interactions between UE and SYSUS factors, and subjective beliefs. We observe that in the case of SE_{slow}, positive bias in regards to the search site speed results in higher positive affect and focused attention, whereas strong belief that the search site will provide relevant results is positively correlated with perceived usability. On the other hand, for SE_{fast}, we observe that participants' positive expectations regarding to the search site speed is negatively correlated with positive affect and positively correlated with focused attention and perceived usability. Moreover, this favorable bias is also positively correlated with expectations that the given search site will respond fast to the queries, will be helpful in accomplishing the task in a reasonable amount of time, and will provide satisfactory results. Despite our relatively small sample, these findings suggest that search engine bias cannot be ruled out and users tend to interpret ambiguous evidence as supporting their existing beliefs. Hence, these tendencies to overestimate, or underestimate, system performance biases their interpretations of search interactions and invokes negative behaviors that may result in search site abandonment.

4. Conclusions

We investigated the impact of increasing response latency on user behavior in web search by conducting a controlled user study. The study revealed that up to a point (500ms) added response time delays are not noticeable by the users. However, after a certain threshold (1000ms), the users could feel the added delay with very high likelihood. We believe that the subjective nature of perceived latency creates an opportunity for search engines. Search results can be served to each user at custom latencies depending on the estimated behavioral impact on the user. For example, if no negative impact is estimated on the user experience, search results may be served at high latencies by computing them using less resources. Serving results at right latencies may bring further financial benefits to search engines in the form of decreased hardware investments and reduced energy consumption. All of this, of course, requires devising certain forecasting mechanisms for accurate prediction of user-perceived response latency as well as the impact on user experience.

References

[1] L. Azzopardi. The economics in interactive information retrieval. In *Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, pages 15–24, 2011.

[2] L. Azzopardi, D. Kelly, and K. Brennan. How query cost affects search behavior. In *Proc. 36th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, pages 23–32, 2013.

[3] M. Hassenzahl. Funology. In M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright, editors, *The Thing and I: Understanding the Relationship Between User and Product*, pages 31–42. Kluwer Academic Publishers, 2004.

[4] J. R. Lewis. Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78, 1995.

[5] L. McCay-Peet, M. Lalmas, and V. Navalpakkam. On saliency, affect and focused attention. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 541–550, New York, NY, USA, May 2012. ACM.

[6] H. L. O'Brien and E. G. Toms. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69, January 2010.

[7] E. Schurman and J. Brutlag. Performance related changes and their user impact. In *Velocity – Web Performance and Operations Conf.*, 2009.