# *Measuring the Aesthetics of Reading*

**Kevin Larson[†], Richard L. Hazlett[‡], Barbara S. Chaparro[§] & Rosalind W. Picard[¶]**

[†] *Microsoft Advanced Reading Technologies, One Microsoft Way, Redmond, WA 98052, USA*
Email: *kevlar@microsoft.com*

[‡] *Johns Hopkins University School of Medicine, 2045 York Road, Baltimore, MD 21093, USA*
Email: *rlhazlet@jhmi.edu*

[§] *Software Usability Research Laboratory, Wichita State University, Wichita, KS 67260-0034, USA*
Email: *barbara.chaparro@wichita.edu*

[¶] *MIT Media Laboratory, 20 Ames Street Cambridge, MA 02139, USA*
Email: *picard@media.mit.edu*

**Aesthetic considerations are as important as usability for human-computer interactions, but techniques for measuring aesthetics have been elusive. In this paper, we use the domain of reading to develop new measures of aesthetics. These measures could be applied to any domain. Reading is arguably the most ubiquitous task that people perform on computers. To date, reading research has focused on reader performance, which is typically measured by reading speed and comprehension. But many typographic improvements that make a more beautiful document show little to no measurable difference on traditional performance tasks. We conducted six studies that found two measures that successfully detect aesthetic differences: improved performance on creative cognitive tasks after text is optimized, and reduced activation in the corrugator muscle that is associated with frowning.**

**Keywords:** reading, typography, aesthetics, emotion, affective user interface.

## 1 Introduction

Reading has arguably the longest and richest history of any domain for scientifically considering the impact of technology on the user. From the 1920s to the 1950s, Miles Tinker [1963] and other researchers ran hundreds of user tests that examined the effects of different fonts and text layout variables, such as the amount of vertical space between each line of text (called leading). Their research focused on user performance, and reading speed was the favoured measure. They charted the effect of the manipulated variables on reading speed, looking for the point at which their participants could read the fastest. Their assumption was that faster reading speeds created a more optimal experience. Printers and publishers eagerly consumed this research.

In recent years, some of these variables have been reexamined as the technology and capabilities evolve with the advent of computers and computer screens. Dillon [1992] examined how to design textual information for an electronic environment. Boyarski et al. [1998] examined the effect of fonts that were designed for computer screens. Dyson & Kipping [1998] examined the effect of line length on computer screens. Larson et al. [2000] examined the effect of 3-D rotation on reading. Gugerty et al. [2004] demonstrated a reading performance advantage with the Microsoft ClearType display technology.

Some typographers argue that the focus for decades on traditional measures of user performance does a disservice to the design of text and documents. Bringhurst [2004] argues that the goal of good typography is to invite the reader into the text. The appropriate measure of invitingness is not reading speed or comprehension, but rather something that other measures have not been able to capture: aesthetics.

In this paper, we first describe two studies that inspired us to investigate aesthetics, and then describe a series of four more studies in which we experimented with two novel measures of aesthetics: creative cognitive tasks and facial muscle measurements. Before discussing the studies, we briefly explain some of the factors that are important to typographers for designing beautiful text.

## 2 The Art of Typography

Typographers are attuned to subtle features when they design and set type. One such feature is symmetry. It is surprisingly difficult to make and render symmetric type. Each stroke across a font needs to be of equal weight – if one vertical stem is heavier than the next, the relative darkness appears as a dark spot on the page. As shown in Figure 1, the white space *within* characters also needs to be balanced – for example, if the space under one arch of an m is narrower than the other arch or narrower than the arch of the n, this letter appears as a dark spot on the page. And the white space between characters needs to be equal to complete the desired symmetry.

To create the desired evenness across a page, typographers use a variety of techniques, including ligatures, kerning, small caps, and old-style numerals. Ligatures are special-combination characters that a type designer creates for a font when two characters clash with each other. The most common use of ligatures is for the letter pairs *fi* and *fl* because the top of the *f* looks uncomfortable with the dot on the *i* and the top of the *l* respectively. Kerning is the technique of adjusting the


**Figure 1:** The first paragraph shows uneven stroke weight and uneven spacing both within characters and between characters. The second paragraph shows only poor spacing between characters. The third paragraph shows symmetry both in characters and spacing.

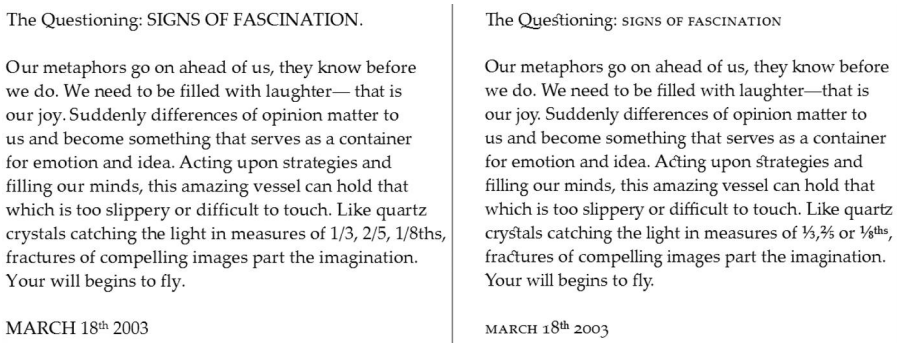



**Figure 2:** The paragraph on the left contains no OpenType features. The paragraph on the right demonstrates OpenType ligatures, kerning, small caps, old-style numerals, and subscript and superscript features.

default letter spacing to create more even spacing. Small caps and old-style numerals are necessary to reduce the large block that gets created on pages that use a large number with several digits or capital letters in a row, such as in an acronym, like HCI. Small caps are the shape of capital letters, but are comparable in size to lowercase letters. Like lowercase letters, old-style numerals have ascenders and descenders.

All of these techniques and dozens of others are now available to typographers for on-screen use with a technology called OpenType. Figure 2 shows an example of OpenType-adjusted text.

Most typographers believe that OpenType features benefit users, but there are many areas of controversy in the typography community about what is best for the user. For example, typographers have not determined if a serif font or sans serif is more appropriate for reading a textbook. Reading speed and comprehension are the primary scientific measures for testing typographic differences. In two performance studies, we discovered that reading speed and comprehension failed to detect differences between two kinds of typographic page manipulation, despite there being a large apparent difference in the two document types.

## 3 OpenType Reading Speed Study

In this study, we have participants read text with and without OpenType features.

**Participants:** Twenty college students (10 male, 10 female) with normal or corrected vision participated in the study and received compensation of $25. Participants were not told the purpose of the study until after its completion.

**Materials:** Text passages were chosen from college board practice examinations that included approximately 800 words (mean = 830.17, s.d. = 33.55). The documents with OpenType incorporated ligatures, kerning, small caps, old-style numerals, and subscript and superscript; the documents without OpenType did not. Text documents were randomly presented on Dell Inspiron 5100 laptops with a 15-inch display with 1400×1050 native resolution. Participants saw documents on two pages. They selected an arrow at the bottom of each page to go to the next page or the previous page. No scrolling was required. They read the documents at a distance of approximately 50cm while being automatically timed.

**Procedure:** In each condition, participants spent approximately 30 minutes reading three documents. We asked them to read each document at their own pace. After reading each document, participants answered eight comprehension questions about the document. They could look up the answers in the documents, but were advised that they had only five minutes to do so. After reading all three documents, the participants were administered the NASA Task Load Index to assess Mental Workload. Participants then took a short break and repeated the procedure with a different set of three documents for the other condition. After both conditions were completed, we showed the participants a sample page with images of the two conditions and asked them to state which layout they liked best. We counterbalanced the order of the two conditions and six passages across participants to ensure that each passage occurred equally in the two conditions.

### *3.1 Results*

**Reading performance:** We averaged the reading rate in words per minute across documents for each condition. Participants in the OpenType condition read at a rate of 194.73 words per minute (s.d. = 48.17); participants in the no OpenType condition read at a rate of 195.89 words per minute (s.d. = 55.64).

We then computed the comprehension scores as a sum score out of a total eight possible. Participants in the OpenType condition comprehended 4.67 questions on average (s.d. = 1.11); participants in the no OpenType condition comprehended 4.73 questions on average (s.d. = 1.02). A paired-samples *t*-test revealed no reliable differences across conditions for reading speed, $t(19) = 0.21$, $p = 0.84$, or comprehension, $t(19) = 0.18$, $p = 0.86$.

**Mental Workload (NASA Task Load Index):** Participants had a reliably higher Task Load Index (TLX) after reading the documents with no OpenType document condition (mean = 48.9) than after reading the OpenType documents (mean = 37.7), $t(19) = 2.79$, $p = 0.01$.

**Preference:** 55% of the participants preferred the OpenType format; 45% preferred the no OpenType format. A Wilcoxon Z test showed no reliable preference for either layout format, Z (N=20) = 0.447, $p = 0.66$.

Results from this study showed that OpenType formatting affected neither reading performance nor document preference. The participants indicated that they did not notice much difference between the two versions and had a hard time choosing which one they preferred. Some participants noticed the smaller caps and different style numbers in the documents with OpenType, but said they preferred the larger caps and numbers in the documents without OpenType. Despite the failure of traditional performance measures to distinguish the two aesthetic conditions, the NASA-TLX did capture a statistically significant difference.

## *3.2 Page Setting Reading Speed Study*

In this study, we have participants read online text with and without enhanced page setting. Enhanced page setting included the optimal use of headers, indentation, figure placement, and quote blocking.

**Participants:** Twenty college students (10 male, 10 female) with normal or corrected vision participated in the study and received compensation of $25.

**Materials:** Figure 3 shows an example of the poor page setting and optimized page setting conditions used in this study. Optimized page setting includes good headers, indentation, figure placement, and quote blocking. Text passages and presentation were the same as the ones that we used in the OpenType reading speed study.

**Procedure:** The procedures were the same as the ones in the OpenType reading speed study.

## *3.3 Results*

**Reading performance:** We averaged the reading rate in words per minute for documents for each condition. Participants in the optimized page setting condition read at a rate of 185.60 words per minute (s.d. = 47.22); participants in the poor page setting condition read at a rate of 183.38 words per minute (s.d. = 51.36). Comprehension scores were computed as a sum score out of

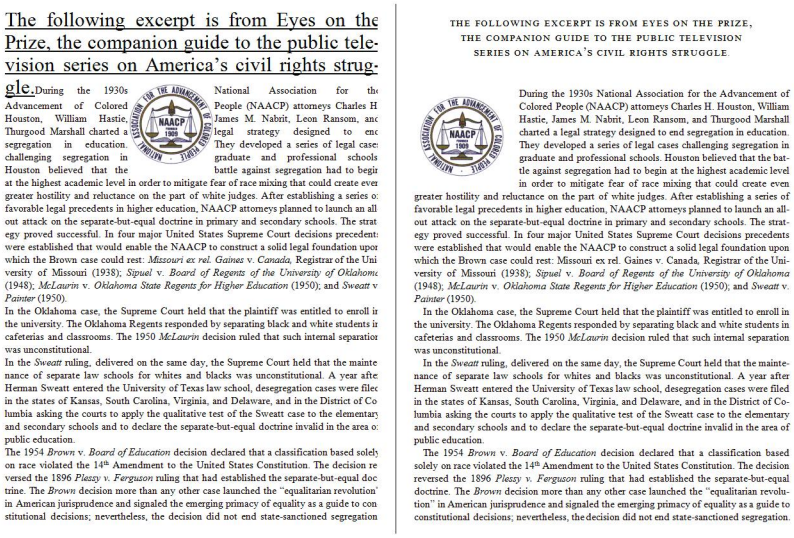



**Figure 3:** The page on the left shows poor page setting. The page on the right shows optimized page setting, including optimized image placement, good headers, and well-marked paragraphs.

---

a total of eight possible. Participants in the optimized page setting condition correctly answered 5.32 comprehension questions on average (s.d. = 0.99). In contrast, participants in the poor page setting condition correctly answered 5.08 questions on average (s.d. = 1.31). A paired-samples *t*-test revealed no reliable differences across conditions for reading speed, $t(19) = 0.41$, $p = 0.69$, or comprehension, $t(19) = 0.96$, $p = 0.35$.

**Mental Workload (NASA-TLX):** Participants showed no reliable differences in the Task Load Index between reading the poor page setting documents (mean = 56.02) and the optimized page setting documents (mean = 50.33), $t(19) = 1.41$, $p = 0.17$.

**Preference:** 90% of the participants preferred the optimized page setting format. In contrast, 10% preferred the poor page setting format. A Wilcoxon Z test showed a reliable preference for the optimized page setting layout, Z (N=20) = 3.58, $p = 0.002$.

Results from this study showed that the quality of the page setting format influenced neither reading performance nor comprehension, despite the fact that participants often had to read around a photograph in the poor page setting documents. All but two participants chose the optimized page setting layout as their favourite. One participant said the positioning of the pictures in the middle of a line was good because it 'broke up the long lines and made it seem shorter.' Both participants, however, said the optimized page setting layout looked better when it was shown alongside the poor page setting at the end of the study.

## 4 New Approaches for Measuring Aesthetics [Instead of Performance]

These two studies demonstrate that some improvements in document typographic quality do not reliably improve user reading speed and comprehension performance, and may not even be explicitly noticed by users. Current measurement tests, which are based on performance and on subjective reporting of workload, are not sophisticated enough to test for improvements that most typographers agree on. How can we test more controversial typographic features?

Research on aesthetics is becoming more common in human-computer interactions as the field shifts its focus from creating usable products to creating desirable products. Sykes & Brown [2003] demonstrated that it is possible to measure arousal during video game play by examining pressure on the gamepad buttons. Reijneveld et al. [2003] used a non-verbal subjective rating of emotions to detect differences while individuals sat in different office chairs. Ward & Marsden [2003] measured the physiological indicators of galvanic skin response, heart rate, and blood volume pressure while individuals performed information searching tasks. Branco et al. [2005] measured facial muscle differences while participants performed word processing tasks.

Our goal is to develop a measure that is sensitive to improvements in aesthetics by extending two earlier methodologies. Our first measure is based on a body of research that shows that a good mood improves people's performance during creative cognitive tasks [Isen 1993].

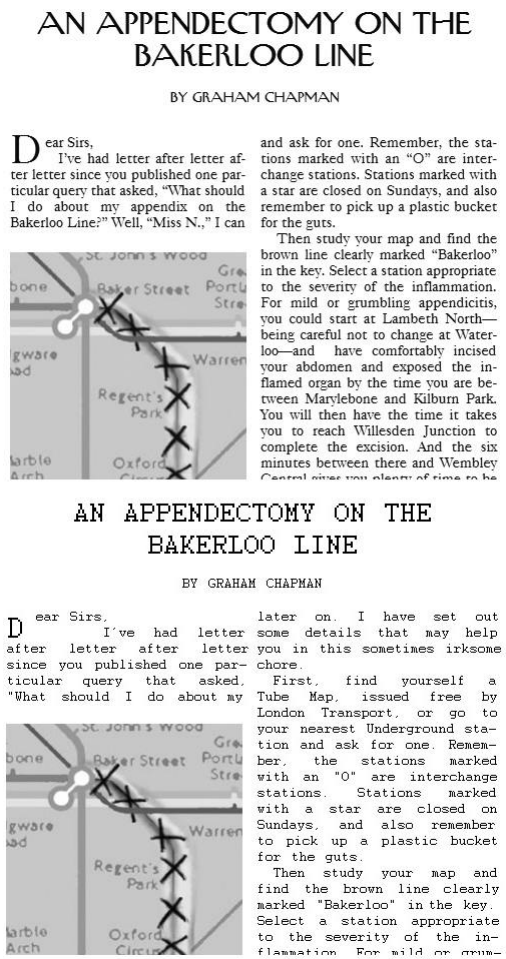
Alice Isen et al. [1987] have demonstrated that participants who are put in a good mood before performing certain cognitive tasks perform better than participants who are not put in a good mood first. Participants can be placed in a good mood by receiving a small gift, such as a candy bar, or by watching five minutes of a humorous video. After being induced into a good mood, participants performed better on creative cognitive tasks, such as the candle task [Duncker 1945] and remote associates task [Mednick 1962], compared to participants who were not put in a good mood.

If a candy bar or humorous video can induce a good mood, can good typography induce a similar kind of good mood? We expect that after reading documents that contain good typography, participants will perform better on the tasks that Isen used to measure creative cognition.

The next two studies use creative cognitive tasks as the dependent measure. Half of the participants read with good typography and half read with poor typography. Our hypothesis is that the participants in the good typography condition will perform better on cognitive tasks than the participants in the poor typography condition. If our hypothesis is true, it suggests that good typography may elevate mood.

### *4.1 ePeriodicals Candle Task Cognitive Measure Study*

Participants in this study read text that either had high-quality typography or poor-quality typography. All participants used special software on a Tablet PC that enabled them to read a full issue of *The New Yorker* magazine. The candle task was used as the measure of aesthetics.



AN APPENDECTOMY ON THE BAKERLOO LINE

BY GRAHAM CHAPMAN

Dear Sirs,
I've had letter after letter after letter since you published one particular query that asked, "What should I do about my appendix on the Bakerloo Line?" Well, "Miss N.," I can and ask for one. Remember, the stations marked with an "O" are interchange stations. Stations marked with a star are closed on Sundays, and also remember to pick up a plastic bucket for the guts.

Then study your map and find the brown line clearly marked "Bakerloo" in the key. Select a station appropriate to the severity of the inflammation. For mild or grumbling appendicitis, you could start at Lambeth North—being careful not to change at Waterloo—and have comfortably incised your abdomen and exposed the inflamed organ by the time you are between Marylebone and Kilburn Park. You will then have the time it takes you to reach Willesden Junction to complete the excision. And the six minutes between there and Wembley



AN APPENDECTOMY ON THE BAKERLOO LINE

BY GRAHAM CHAPMAN

Dear Sirs,
I've had letter after letter after letter since you published one particular query that asked, "What should I do about my later on. I have set out some details that may help you in this sometimes irksome chore.

First, find yourself a Tube Map, issued free by London Transport, or go to your nearest Underground station and ask for one. Remember, the stations marked with an "O" are interchange stations. Stations marked with a star are closed on Sundays, and also remember to pick up a plastic bucket for the guts.

Then study your map and find the brown line clearly marked "Bakerloo" in the key. Select a station appropriate to the severity of the inflammation. For mild or grum-



**Figure 4:** The New Yorker with optimized (top) and poor (bottom) typography.

---

**Participants:** Twenty participants (10 male, 10 female) with normal or corrected vision participated in this study and received a compensation of Microsoft software. All participants classified themselves as occasional readers of The New Yorker. We used a between-participants design, in which we assigned 10 participants to the poor typography condition and the other 10 participants to the optimized typography condition. The data from one participant in the poor typography condition was discarded because she was familiar with the candle task.

**Materials:** ePeriodicals are electronic versions of print magazines. Each page is designed to fit perfectly on a 768×1024 native resolution Tablet PC in portrait

orientation without scrolling. People turn pages (page up and page down) by pressing a tablet hardware button. People can also use a tablet pen to navigate directly to the table of contents or any article. In this study, we used the content from the January 5th, 2004 edition of *The New Yorker*, including the text, images, and advertisements that were used in that edition of the print magazine. The page layout differed from that of the print magazine to accommodate the size of the Tablet PC screen.

We created two versions of the ePeriodical. As shown in Figure 4, the optimized typography version used the New Yorker font with ClearType and good hyphenation and justification. The poor typography version used the bitmap version of the Courier font and included an extra two points of space between each word. Although the text looked terrible, users had no trouble reading the text – and the content was exactly the same in the two conditions.

**Procedure:** We gave each participant a brief tutorial about the ePeriodical user interface. The tutorial explained how to use the tablet hardware buttons to move to the next page or the previous page, and how to navigate to the table of contents and to any article. Participants could choose to read anything they wanted to read from this ePeriodical. We asked the participants to read for 20 minutes.

After the reading session, we gave the participants the candle task. In this mood-detecting task, we gave participants a box full of tacks, a candle, a match, and a corkboard, which was affixed to a wall. Their task was to attach the candle to the corkboard in such a way that the wax wouldn't drip all over the place when lit. They had 10 minutes to solve the task. We considered the task to be correctly solved if a participant emptied the tacks from the box, tacked the box to the corkboard, and placed the candle inside the box. All other solutions were considered incorrect.

## *4.2 Results*

Candle task. We found that 4 of 10 participants correctly solved the candle task in the optimized typography condition. In contrast, 0 of 9 participants correctly solved the task in the poor typography condition. This is a reliable difference, $\chi^2(1) = 2.47$, $p = 0.04$. We propose that the typographic difference between poor and optimized typography drove the difference in the creative cognitive task. This finding is very similar to Isen's finding that a small gift or humorous video improved performance in creative cognitive tasks. This result indicates that typographic differences can also impact mood.

## *4.3 ePeriodicals Remote Associates Cognitive Measure Study*

To examine whether we could replicate the above results by changing the creative cognitive task that is used, we replaced the candle task with the remote associates task, which is another cognitive task that Isen has shown to be influenced by positive mood. One advantage of the remote associates task compared to the candle task is that it can be used in within-participant studies, though this study uses a between-participant design.

In the remote associates task, we gave participants three words, such as *water*, *skate*, and *cream*, and then asked them to think of a word that created a common compound with each of the three words. In this example, the correct answer is *ice*. Participants saw the three items on a computer screen for up to 15 seconds, and then pressed a keyboard button as soon as they knew the answer. After the participants pressed the button or 15 seconds had elapsed, we prompted them to type their response. We collected the reaction time and the accuracy data from this test.

All other details of the study are identical to the first study. The same experimenter gave 20 new participants the same instructions about how to use an ePeriodical.

### *4.4 Results*

**Remote associates test:** We found that the participants who viewed the optimized typography succeeded at 52% of the trials, at an average speed of 6395ms. Participants who viewed the poor typography succeeded at 48% of the trials, at an average speed of 6715ms. Although the means are in the predicted direction, neither the accuracy, $t(18) = 1.15$, $p = 0.27$, nor the speed differences, $t(18) = 0.42$, $p = 0.68$, are statistically reliable.

In summary, we conducted two studies using cognitive measures, which produced findings that suggest that more aesthetic typography can create greater positive feelings. When we induced a positive mood in the participants, they performed better on the candle task. However, the remote associates task did not show a statistically reliable difference.

## 5 Facial Muscle Activation

Another approach to measuring the emotion of readers is to measure changes in facial expressions by using facial electromyography (EMG). By placing tiny sensors over certain facial muscles, one can measure the minute changes in the muscles' electrical activity, which reflects changes in muscle tension. Facial EMG studies have found that activity of the corrugator muscle, which lowers the eyebrow and is involved in producing frowns, varies inversely with the emotional valence of the presented stimuli and reports of mood state; and activity of the zygomatic muscle, which controls smiling, is positively associated with positive emotional stimuli and positive mood state [Cacioppo et al. 1992; Dimberg 1990]. These effects have been found in passive viewing situations with various protocols and media, such as photos, videos, words, sounds, and imagery [Hazlett & Hazlett 1999; Larsen et al. 2003].

Increases in zygomatic EMG and decreases in corrugator EMG tend to indicate a positive emotional response. The orbicularis oculi muscle EMG controls eye smiling and is also positively associated with positive mood state. Some research has suggested that because this muscle is less subject to voluntary control than the zygomatic muscle, it may more truly reflect actual felt emotion [Cacioppo et al. 1992]. In task performance research, activity of the corrugator muscle provides a sensitive index of the degree of exerted mental effort [Waterink & van Boxtel 1994] and increases with the perception of goal obstacles [Pope 1994]. Several studies

**Figure 5:** Participant with facial EMG sensors.

have demonstrated that the corrugator EMG can reflect the computer user's tension and frustration [Scheirer et al. 1999; Hazlett 2003; Branco et al. 2005].

## *5.1 Page Setting Facial Muscle Activation Study*

In this study, we compared the activation of facial muscles while participants read articles with optimized page setting and poor page setting. During the experiment, continuous emotional response data were collected from the corrugator (frown) muscle EMG and the zygomatic (smile) EMG.

**Participants:** 25 participants (13 male, 12 female) with normal or corrected vision participated in this study and received compensation of $90.

**Materials:** The page setting documents were the same as the documents that we used in the page setting reading speed study. We presented documents on a Sony Vaio laptop with a 15-inch display running 1400×1050 resolution. We continuously measured facial EMG by placing Rochester miniature Silver/Silver Chloride surface electrodes over the left zygomaticus major and corrugator supercilii muscles. We followed recommended guidelines for preparing the skin and placing the electrodes [Tassinary et al. 1989]. The raw EMG signals were amplified and filtered by using two Psylab (Contact Precision Instruments) bioamplifiers and processing system. EMG detection bandpass was set at 30Hz–500Hz, and the analogue EMG signal was digitized at 1000Hz.

**Procedure:** This study used a repeated measures design. We asked subjects to read six two-page articles: three in poor type and three in optimized type. The poor and optimized conditions were alternated for each participant, and the articles were counterbalanced between participants. After we attached the EMG sensors and a quiet baseline period had passed, the participants began reading the series of six articles. During the reading, the facial EMG measures were collected continuously.

## 5.2 Results

**Reading speed and comprehension:** Like the page setting reading speed study, there were no reliable differences in either reading speed or comprehension.

**Corrugator EMG:** We rectified the EMG values and calculated the 100ms data points. Because we were interested in the more subtle background effects of the typography, we filtered out extreme values of the EMG data series that were greater than 1.5 standard deviations (determined by pilot testing) above the series mean. These extreme EMG values were due to movement or other extraneous factors. Two participants' EMG data were not valid due to technical problems, which left us with 23 valid data series in the EMG statistical analyses. The corrugator mean value for the poor page setting documents was 11.59 microvolts; the mean value for the optimized page setting documents was 11.08 microvolts, which resulted in a mean difference of 0.51 microvolts (s.d. = 1.10). The corrugator muscle had reliably greater EMG when participants read the poor page setting documents, $F(1,22) = 4.90$, $p = 0.04$. This data showed that people used their frown muscles more while reading the text with the poor typography.

**Zygomatic EMG:** The zygomatic mean value for the poor page setting documents was 16.34 microvolts; the mean value for the optimized page setting documents was 15.89 microvolts, which resulted in a mean difference of 0.45 microvolts (s.d. = 4.83). This difference was not reliable, $F(1,22) = 0.22$, $p = 0.65$.

## 5.3 OpenType Facial Muscle Activation Study

This study expanded on the results of the first facial muscle activation study, which showed corrugator differences for the page setting documents. This study investigated EMG differences when participants read the documents with and without OpenType features. The difference between documents with vs. without OpenType is more subtle than the difference between page setting documents. In the OpenType reading speed study, we did not find a difference between the OpenType conditions, either in reading speed and comprehension or in document preference. In contrast, in the page setting reading speed study, we found a difference in document preference for the page setting documents.

In this study, we measured the orbicularis oculi (eye smile) muscle instead of the zygomatic muscle (both are associated with positive emotions) because we failed to find any zygomatic muscle differences in the page setting facial muscle activation study.

**Participants:** 25 participants (13 male, 12 female) with normal or corrected vision participated in this study and received compensation of $90.

**Materials:** The OpenType documents were the same documents as the ones we used in the OpenType reading speed study. The equipment and settings were the same as the ones we used in the page setting facial muscle activation study.

**Procedure:** The procedures were the same as the procedures we used in the page setting facial muscle activation study, except that we placed the EMG sensors on each participant's corrugator and orbicularis oculi muscles.

### *5.4 Results*

**Reading speed and comprehension:** As with the OpenType reading speed study, there were no reliable differences in either reading speed or comprehension.

**Corrugator EMG:** One participant's EMG data was not valid due to technical problems, which left us with 24 valid data series in the EMG statistical analyses. The corrugator mean value for the no OpenType documents was 13.80 microvolts; the mean value for the OpenType documents was 12.56 microvolts, which resulted in a mean difference of 1.23 microvolts (s.d. = 2.83). The corrugator muscle had reliably greater EMG when participants read the no OpenType documents, $F(1,23) = 4.56$, $p = 0.04$.

**Orbicularis oculi EMG:** The orbicularis oculi mean value for the no OpenType documents was 10.35 microvolts; the mean value for the OpenType documents was 9.45 microvolts, which resulted in a mean difference of 0.90 microvolts (s.d. = 3.82). This difference was not reliable, $F(1,23) = 1.34$, $p = 0.25$.

The two facial muscle activation studies have demonstrated that we can detect subtle typographic differences by measuring the corrugator muscle. The corrugator muscle was more active when participants read the poor typography documents. The corrugator EMG decreases with increases in positive emotional state, and increases with increases in negative emotional state, tension, and mental effort. It is not clear from these studies which of these factors was associated with the type differences that we noted. The fact that we did not find differences with either of the positive emotion EMG measures – the zygomatic or orbicularis oculi facial muscles – does not help clarify the associated factors, because these measures are known to have a higher activation threshold [Larsen et al. 2003], and we are measuring very subtle effects.

## 6 Conclusions

The goal of this project was to develop new measures of subtle aesthetics differences for any domain. We have described two new measures and examined their application to detecting subtle typographic differences. These measures detected reliable differences, where classic measures of reading speed and comprehension failed to detect any differences. The first measure was based on Isen's findings that mood inducers, such as receiving a small gift or watching a humorous video, caused people to perform better on creative cognitive tasks, such as the candle task and remote associates task. In the first ePeriodical cognitive measure study, we manipulated typographic quality and demonstrated the same performance benefit that Isen found, which indicates that optimized typography has similar mood-improving powers as a small gift or humorous video does. In the second ePeriodical cognitive measure study, we did not find the same advantage in the remote associates

task that Isen found, but we are still optimistic about exploring other creative cognitive tasks.

The second measure involved measuring the differences in facial muscle activation. In the facial muscle activation studies, EMG on the corrugator muscle successfully detected differences on both the page setting documents and the OpenType documents. The corrugator activation was greater for the poor typography condition in both studies, which indicates that participants experienced a greater level of frustration, disapproval, tension, or mental effort while reading documents that contain poor typography.

Our expectation is that these measures are generally useful for detecting subtle aesthetic differences in any field, and are not particular to typographic differences. Moving forward, we plan to use these measures to detect differences where we don't have a predetermined expectation for the outcome.

## 7 Future Work

Future research should incorporate these practical measures into a theoretically sound model of emotion. Right now, we cannot say if the good typography conditions induce a more positive emotion in the reader, or whether poor typography conditions require more tension or effort to read. Also, our six studies tapped into the overall positive and negative dimensions of emotion only. Future work might fruitfully investigate more specific emotional characteristics of typography.

We want to see aesthetic and emotional research become as commonplace as usability research and other performance measures. We prefer to use less intrusive physiological measures than an electrode attached to the skin. For example, the sensors could be hidden in chairs [Anttonen & Surakka 2005], or thermal imaging could be used to detect corrugator differences from a distance [Puri et al. 2005].

## Acknowledgements

Special thanks to Michael Duggan for designing the page setting and OpenType stimuli, to Tracy Premo who ran the ePeriodical studies, and to Radoslav Nickolov and Steven Kim for helping create the ePeriodicals that contained optimized and poor page settings.

## References

Anttonen, J. & Surakka, V. [2005], Affect and Intimacy: Emotions and Heart Rate while Sitting on a Chair, *in* G. van der Veer & C. Gale (eds.), *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*, ACM Press, pp.491–9.

Boyarski, D., Neuwirth, C., Forlizzi, J. & Regli, S. H. [1998], A Study of Fonts Designed for Screen Display, *in* C.-M. Karat, A. Lund, J. Coutaz & J. Karat (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'98)*, ACM Press, pp.87–94.

Branco, P., Firth, P., Encarnacao, L. M. & Bonato, P. [2005], Faces of Emotion in Human–Computer Interaction, *in* G. van der Veer & C. Gale (eds.), *CHI'05 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.1236–9.

Bringhurst, R. [2004], *The Elements of Typographic Style*, Hartley and Marks Publishers.

Cacioppo, J. T., Bush, L. K. & Tassinary, L. G. [1992], Microexpressive Facial Actions as a Function of Affective Stimuli: Replication and Extension, *Psychological Science* **18**(5), 515–26.

Dillon, A. [1992], Reading from Paper vs. Screens: A Critical Review of the Empirical Literature, *Ergonomics* **35**(5), 1297–326.

Dimberg, U. [1990], Facial Electromyography and Emotional Reactions, *Psychophysiology* **27**(5), 481–94.

Duncker, K. [1945], On Problem Solving, *Psychological Monographs* **58**(270). The American Psychological Association.

Dyson, M. C. & Kipping, G. J. [1998], The Effects of Line Length and Method of Movement on Patterns of Reading from Screen, *Visible Language* **32**(2), 150–81.

Gugerty, L., Tyrrell, R. A., Aten, T. R. & Edmonds, K. A. [2004], The Effects of Sub-pixel Addressing on Users' Performance, *ACM Transactions on Applied Perception* **1**(2), 81–101.

Hazlett, R. L. [2003], Measurement of User Frustration: A Biologic Approach, *in* G. Cockton, P. Korhonen, E. Bergman, S. Björk, P. Collings, A. Dey, S. Draper, J. Gulliksen, T. Keinonen, J. Lazar, A. Lund, R. Molich, K. Nakakoji, L. Nigay, R. Oliveira Prates, J. Rieman & C. Snyder (eds.), *CHI'03 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.734–5.

Hazlett, R. L. & Hazlett, S. Y. [1999], Emotional Response to Television Commercials: Facial EMG vs. Self-report, *Journal of Advertising Research* **39**(2), 7–23.

Isen, A. M. [1993], Positive Affect and Decision Making, *in* M. Lewis & J. M. Haviland (eds.), *Handbook of Emotions*, The Guilford Press, pp.261–77.

Isen, A. M., Daubman, K. A. & Nowicki, G. P. [1987], Positive Affect Facilitates Creative Problem Solving, *Journal of Personality and Social Psychology* **52**(6), 1122–31.

Larsen, J. T., Norris, C. J. & Cacioppo, J. T. [2003], Effects of Positive and Negative Affect on Electromyographic Activity over Zygomaticus Major and Corrugator Supercilii, *Psychophysiology* **40**(5), 776–85.

Larson, K., van Dantzich, M., Czerwinski, M. & Robertson, G. [2000], Text in 3-D: Some Legibility Results, *in* M. Tremaine (ed.), *CHI'00 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.145–6.

Mednick, S. A. [1962], The Associative Basis of the Creative Process, *Psychological Review* **69**(3), 220–32.

Pope, L. K. & Smith, C. A. [1994], On the Distinct Meanings of Smiles and Frowns, *Cognition and Emotion* **8**(1), 65–72.

Puri, C., Olson, L., Pavlidis, I., Levine, J. & Starren, J. [2005], StressCam: non-contact measurement of users? emotional states through thermal imaging, *in* G. van der Veer & C. Gale (eds.), *CHI'05 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.1725–8.

Reijneveld, K., de Looze, M., Krause, F. & Desmet, P. [2003], Understanding What to Design through Empathy and Emotion: Measuring the Emotions Elicited by Office Chairs, *in Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces*, ACM Press, pp.6–10.

Scheirer, J., Fernandez, R. & Picard, R. W. [1999], Expression Glasses: A Wearable Device for Facial Expression Recognition, *in* M. E. Atwood (ed.), *CHI'99 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.262–3.

Sykes, J. & Brown, S. [2003], Affective Gaming: Measuring Emotion through the Gamepad, *in* G. Cockton, P. Korhonen, E. Bergman, S. Björk, P. Collings, A. Dey, S. Draper, J. Gulliksen, T. Keinonen, J. Lazar, A. Lund, R. Molich, K. Nakakoji, L. Nigay, R. Oliveira Prates, J. Rieman & C. Snyder (eds.), *CHI'03 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.732–3.

Tassinary, L. G., Cacioppo, J. T. & Geen, T. R. [1989], A Psychometric Study of Surface Electrode Placements for Facial Electromyographic Recording: I. The Brow and Cheek Muscle Regions, *Psychophysiology* **26**(1), 1–16.

Tinker, M. A. [1963], *Legibility of Print*, Iowa State University Press.

Ward, R. D. & Marsden, P. H. [2003], Physiological Responses to Different Web Page Designs, *International Journal of Human–Computer Studies* **59**(1–2), 199–212.

Waterink, W. & van Boxtel, A. [1994], Facial and Jaw-elevator EMG Activity in Relation to Changes in Performance Level During a Sustained Information Task, *Biological Psychology* **37**(3), 183–98.

# *Author Index*

# *Keyword Index*