Emotional Responses to Font Types and Sizes in Web Pages

Renata Germano Bianchi
renatagbianchi@gmail.com
DC, Universidade Federal de São Carlos
São Carlos, São Paulo, BR

Kamila Rios da Hora Rodrigues
kamila.rios@icmc.usp.br
ICMC, Universidade de São Paulo
São Carlos, São Paulo, BR

Vânia Paula de Almeida Neris
vania@dc.ufscar.br
DC, Universidade Federal de São Carlos
São Carlos, São Paulo, BR

ABSTRACT

Emotions play a critical role in decision-making and behavior. Several elements in the web design combined evoke specific emotions, although typography is usually related to preferences rather than the emotion evoked. For this reason, the aim of this paper is to investigate the emotional effects of font types and sizes to support design decisions. The main stages in the study were: preparing web pages; varying the font types and sizes; carrying out a case study with 60 volunteers; statistical data analysis with various evaluative instruments to comprise emotional components (cognitive appraisals, behavioral tendencies, motor expressions and subjective feelings). From the data collected with the subjects, Arial was more closely allied to positive emotions than Times New Roman. Font size 8pt caused unpleasant experiences, with the highest number of errors. The medium and large sizes, represented by 12pt and 16pt, were mainly related to pleasant and comforting experiences. There were some differences in the findings on font types from the literature and further studies are necessary.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Typography, Font sizes, Font types, Emotions, Design, Empirical study

1 INTRODUCTION

Emotions play a critical role in decision-making and behavior, which was formerly regarded as rational, such as deciding which chair to sit on or which sandwich to eat, just to cite some of those with low impact in someone's life [20, 33, 38]. Thus, they should be examined in people's daily activities, such as in their interaction

with the media (e.g., web pages, television etc.) to provide a better experience of use [29, 33].

Some authors state that it is possible to design user interfaces to evoke particular emotions in people [22]. According to Jiang et al. [21], a web page is basically formed of colors, images and text. The aim of this paper is to explore the last element more deeply, by examining the relationship between certain typographical features and emotions.

Typographical features are based on fonts and typesetting. The first includes color, background color, sizes and types, while the latter consists of plain, bold, italics, bold-italics, among other features [39]. In seeking to achieve their goals, there are many ways for designers to display online texts, such as changing these combinations [27]. It should be noted that each variation could evoke different emotions [21].

Most studies investigate people's preference for a particular type rather than the emotion evoked [2, 3, 6, 27]. Furthermore, few works were found that investigated the emotional effects of the font size [23, 39, 40]. This study explores these font elements in greater depth with the aim of filling the gap between font types and emotions and looking more closely at the effects of font sizes.

The desire to overcome the mentioned drawbacks and to support design decisions where there is an emotional intention about which font type and size to use to evoke emotions, has led this paper to link font features with the emotions defined in Scherer's Semantic Emotional Space [36]. This research study followed a sequence of stages: 1. preparing web pages and varying font elements; 2. carrying out case study with 60 volunteers; 3. conducting an analysis of the collected data, both individually and grouped according to the Hybrid Approach to the Assessment of Emotional Experience [43, 44]; and 4. critical data analysis.

The results with regard to the font type suggested that Arial was more positive than Times New Roman in the comments made by the subjects. There were no significant differences between these font types with regard to the reading time and number of errors. When all the instruments were merged to produce a final emotional experience, there were similar results among the font types. Concerning the font sizes, the volunteers reported 8pt as being hard to read, unpleasant and tiring, in both subjects' discourse and also converging all instruments. Further measures suggested that it had the highest number of errors. In contrast, 12pt and 16pt were mainly related to pleasant and comforting experiences.

2 THEORETICAL FRAMEWORK

Owing to the complexity of the subject, the literature has some difficulty in defining the term 'emotion', and this has led to a good deal of discussion among researchers [12, 33, 34, 36, 43]. This study has adopted Scherer's approach [35–37] as a means of covering a

broad spectrum of emotions within a dynamic model, with the purpose of assisting designers to understand and evaluate the effects of their decisions. The author claims that emotion is "the interface between an organism and its environment mediating between constantly changing situations and events and the individual's behavioral responses" [37], and thus involves preparing reactions of the organism to situations and events seen as important [35]. During these evaluations that are carried out by the emotion, five components undertake specialist functions: cognitive appraisals, physiological reactions, behavioral tendencies, motor expressions, and subjective feelings [37].

For this reason, the Semantic Emotional Space [36], model illustrated in Figure 1, includes emotions that cover the components cited. Four dimensions are referred to in this model: Pleasure (pleasant/positive - unpleasant/negative), Arousal (active/aroused - passive/calm), Control (high control/power - low control/power) and Goal Conduciveness (conducive - unfavorable) [36, 44]. The graph is divided into octants (i.e., in eight sets), the first (Octant 1) being the uppermost on the right, following in a clockwise direction so that each can be identified. The last one, Octant 8, is the uppermost on the left.

Figure 1: Structure of semantic emotional space [43].
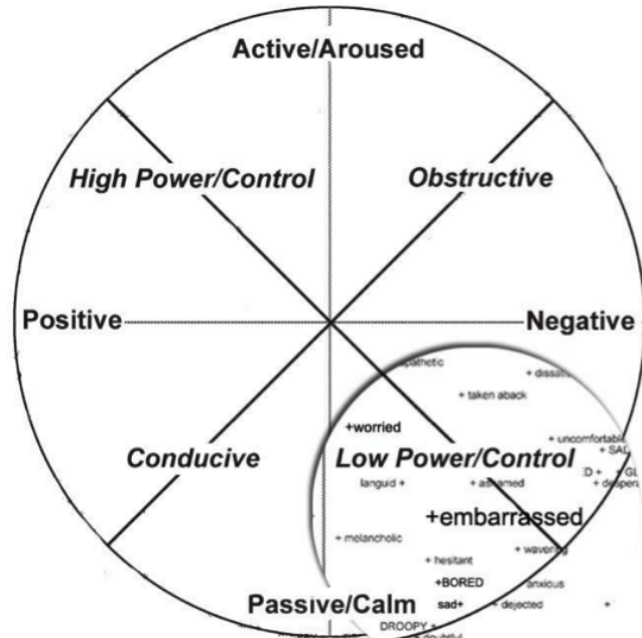
This paper adopted the Hybrid Approach [44] to make an appraisal of the emotional experience. It takes account of the viewpoints of both participants and specialists, since the former may not be able to express or choose their emotions or emotional state while, at the same time, the latter may be inaccurate or partial.

The application of this approach to the Assessment of Emotional Experience can be divided into three stages: the selection of measures, the generalization of the results, and incidence with octants [43, 44].

The first stage - the selection of measures - characterizes what instruments are needed to evaluate the five components of emotion cited previously. Table 1 illustrates this decision-making.

Table 1: Example of the Hybrid Approach – the first stage [43].

Emotional Component	Measure	Dimension evaluated
Cognitive appraisals	Thinking Aloud [41]	Control
Physiological reactions	Electrocardiogram	Arousal
Behavioral tendencies	Performance metrics and effectiveness	Conduciveness
Motor expressions	Facial Action Coding System (FACS) [11]	Pleasure
Subjective feelings	Self-Assessment Manikin (SAM) [10]	Pleasure, Arousal and Control
	Oral questionnaire	Conduciveness

The second stage is the generalization of the results for each measure in terms of positive, neutral or negative. The evaluator relates each positive or negative result to the respective hemisphere (four octants), which are assessed as positive or negative. However, although it can be assumed that the subjects did experience one or more emotions during the interaction, in case it is not possible to complete a final experience for the measure, the result can be regarded as neutral and not related to any dimension [43].

Next, in the third stage, the octants' incidence process is carried out. For each measure evaluated, the result is the octant with the highest rate [43], i.e. a statistic mode. In the end, a set of emotions represents the participant's emotional experience, which has been elicited from the interaction with computing systems.

3 RELATED WORKS

As well as the other user interface elements, typography has an influence on observer evaluations and decision-making [31]. They instantly affect emotions, satisfaction, preferences and the intention to use the media again [31]. Hence, it is of value to both designers and users to examine how typography (and the design itself) affects people's emotions, and thus to provide a better user experience.

According to Altaboli [2], the most common font types are Times New Roman (Serif font style) and Calibri (Sans-Serif), owing to their everyday use and for being default in word processing and web development software. In an attempt to measure the user's perceptions of visual aesthetics, the author displayed four designs of

the same webpage to a group of 30 people, and obtained 22 valid responses. The pages had an identical format (e.g., color and menu), but different font types in the title and body; there were only variations between Times New Roman, and Calibri and mixtures of types in Times New Roman and Calibri. After seeing a picture of each design, the participants were instructed to answer the instrument Visual Aesthetic of Website Inventory (VisAWI) [31]. Overall, they regarded the designs with Times New Roman as having a better visual aesthetics.

Another study sought to determine what font type (Times New Roman or Arial) and size (10pt or 12pt) are the most readable and preferred for reading on computers [6]. In terms of accuracy, there were no significant differences for each font type and size. However, with regard to the reading time, Arial with a font size of 10pt lasted a longer time than Arial and Times New Roman with 12pt. Despite these results, most of the participants showed preference for Arial. This preference was also noted in [27].

A group of researchers related font size to the following dimensions of emotions: Pleasure, Arousal and Control. They found that the emotional state Pleasure increased by up to 15pt, while Arousal and Control declined; and between 15pt and 18pt, Pleasure declined but Arousal and Control increased [39]. In more specific terms, the results of a more detailed study revealed a reduction in Pleasure and increase of Arousal between 9pt and 13pt and from 27pt, whereas between 14pt and 27pt, Pleasure increased and Arousal decreased [40].

The work of Koch [23] investigated whether viewing typefaces evokes emotions and whether all the subjects feel the same emotions. The participants in the case study were mainly designers (32 out of 42). They interacted with an online questionnaire in which the alphabet was shown in six different typefaces and, after each one, for twelve emotions they responded to the PrEmo™ instrument [17], as illustrated in Figure 2, to show whether they felt each or not. The typefaces investigated were: Helvetica Bold, Helvetica Ultra Light, Helvetica Bold Extended, Helvetica Condensed Bold, Helvetica Rounded Bold, and Glypha Medium. In general, the results showed that the participants did feel emotions and ones that could be compared. A note was that emotions reported were different among the typefaces. For example, Glypha Medium and Helvetica Light were closely related to satisfaction, while Helvetica Extended Bold was more linked to boredom [23].

Most previous studies investigate people's preferences with regard to font type, visual aesthetics, legibility and readability rather than the emotion evoked [2, 4, 6, 7, 19, 27, 31]. On the other hand, some research studies, such as those outlined above, are showing the first signs of support for designers who wish to examine the emotional impact of typography. Since few works were found about font size, this paper seeks to ascertain the findings of these studies and focus on the sizes that are generally used by designers in user interfaces. It can also be differentiated from previous studies on font types and sizes for a) gathering emotions employing more than one instrument established in the literature and b) embodying four emotional components - cognitive appraisals, behavioral tendencies, motor expressions, and subjective feelings. This leads to a better view of the emotions resulting from the interaction in question, in addition to addressing a broad spectrum of emotions.

Figure 2: The PrEmo instrument [23].

4 EXPERIMENTAL PROCEDURE

4.1 Hypotheses

The overall null hypothesis states that different font types and sizes in user interface design evoke the same emotions and/or sets of emotions. The alternative hypothesis is that the various ways these elements are shown have different emotional effects, i.e., those of the fonts Arial and Times New Roman are all different, as well as the 8pt, 12pt and 16pt sizes.

4.2 Participants

A total of 60 volunteers participated in the case study involving both font sizes and font types (i.e., all of them interacted with both stimuli). There were 43 men and 17 women, aged between 18 and 63, with an average age of 24.89. The educational level ranged from college students and graduates to PhD students.

4.3 Instantiation of the Hybrid Approach

The instruments selected to assess the volunteers' emotional state when faced with each emotional component are illustrated in Table 2.

Table 2: Instruments selected to evaluate the emotional experience of the participants.

Emotional Component	Measure	Dimension evaluated
Cognitive appraisals	Discourse of the Collective Subjects [26]	Control
Physiological reactions	-	-
Behavioral tendencies	Reading time and number of errors	Conduciveness
Motor expressions	Ten Emotion Heuristics [16]	Pleasure
Subjective feelings	SAM [10]	Pleasure, Arousal and Control

The Discourse of the Collective Subject [26] was the instrument selected to measure cognitive appraisals, and these were analyzed by one evaluator. A semi-structured interview was conducted after each stimulus. The participants were encouraged to comment on what the experience was like - whether it changed their emotional state or whether they had any general comments.

Physiological reactions were not included because of the high cost of the instruments, such as electroencephalogram (EEG), or monitors for skin temperature and heart rate.

The evaluation of behavioral tendencies was carried out on the basis of the measurement of the reading duration and number of errors. The subjects were asked to read the texts aloud, and this made it possible to take notes.

In this study, the motor expressions were evaluated with the aid of the Ten Emotion Heuristics [16], a set of guidelines to assess participants' emotions by analyzing their facial expressions and reactions during the interaction (e.g., eyebrow-raising, smiles, compression of the lips etc.). A negative emotional experience characterizes five or more of the different negative heuristics identified [16]. The application of the Ten Emotion Heuristics took place in two phases: the first was individually - each evaluator watched the videos with the interactions and noted each heuristic violated, while in the second phase the group of evaluators discussed their findings so that they could reach a consensus. The group was formed of an experienced evaluator (with more than 5 applications), a less experienced (2-5) and another with no experience, in accordance with levels defined by Norman [32].

The subjective feelings were measured with SAM [10], a nonverbal pictorial instrument illustrated in Figure 3. It evaluates Pleasure, Arousal and Control with ratings based on their degree of intensity. Manikins vary for each dimension: Pleasure ranges from 'smiling' to 'frowning'; Arousal from 'excited' to 'sleepy'; and Control from a large to a small manikin. Classifications between 1 and 4 correspond to a negative state (e.g., low pleasure), 5 represents a neutral response, and values from 6 to 9 are positive (e.g., high pleasure) [24].

Figure 3: The SAM instrument [10].
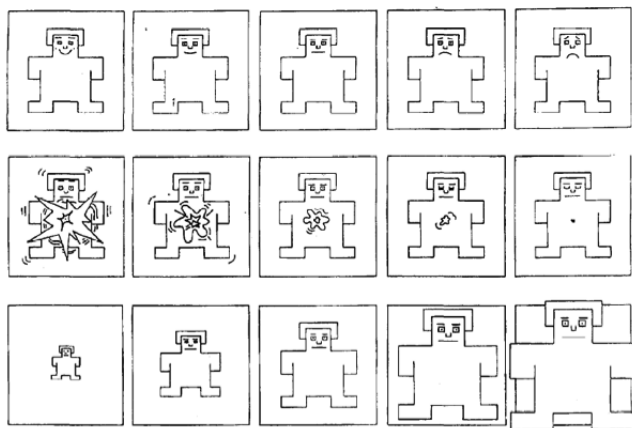
4.4 Stimuli

With regard to font styles and types, the literature mainly focuses on investigating Serif and Sans-Serif styles; the former is generally represented by the font type Times New Roman, while the latter by Arial or Calibri [3, 6, 31, 40]. This paper assumes that Arial and Calibri do not represent the style in its totality since each of them has particular features. Times New Roman and Arial are the types that are most often reported in the literature, but still have little relation to emotions, and for this reason were investigated (in the same font size – 12pt).

The selection of font sizes to analyze their emotional impact was based on a range of strong candidates for user interface design. In the newspaper web pages omitted for blind review, the contents are displayed between 10 and 12pt. Similarly, in general, the scientific studies employ the sizes 10pt, 12pt and 14pt [6, 7, 27, 30], and hence the medium size investigated was 12pt. A value to characterize a small size should be lower than 10pt, as this is what is generally used. For the proximity, 9pt was disregarded and 8pt adopted instead. Since this research is examining font sizes on web pages', sizes larger than 20pt were not included because they are generally used for titles (see online newspapers such as omitted for blind review). For this reason, the value determined to represent content in a large size was 16pt.

When the same text is read several times, it results in an automatic reading without the participants paying any attention to the text and to its typographical forms. This influences the outcome of the case study, in particular, in the results, which are hardly attenuated. Different topics were selected from newspaper articles to keep the participants interested: these were from the areas of Science and Health, Economics, Sport, World Affairs, and Technology - two per topic, making a total of ten. Each passage had approximately 130 words, within a range of 15%.

This study takes into account that the participants' emotions may be linked to the content and not to the font type and size. However, different subjects can arouse different interests and emotions among the various participants, which reduces the significance of this factor.

The papers selected were divided into two periods: font type (Phase 1) and font size (Phase 2). In Phase 1, the texts alternated between Times New Roman and Arial. In Phase 2, the participants always interacted with the same font type (Times New Roman or Arial) and the texts were displayed at random using 8pt, 12pt and 16pt; in both phases the font color was black. The experimental design ensured that all the themes were shown the same number of times in each reading.

Briefly, different newspaper articles were selected to investigate the emotional effects of Times New Roman and Arial, as well as the influence of the sizes 8pt, 12pt and 16pt in the reader's emotions, in both font types. After that, we prepared the webpages. A sample of the stimulus is illustrated in Figure 4.

4.5 Procedure

The experimental procedure complied with the protocols approved by the Ethics Committee and was as follows.

First of all, there was a brief description of the research objectives, as well as the tasks that would be carried out, risks involved

Figure 4: Stimulus sample.

Being focused, concentrating and meditating help to keep the brain active and avoid a loss of memory

The head holds so much information that the memory can fail when someone least expects it. The program invited the psychiatrist and consultant Daniel Barros and the neuroscientist Rogerio Panizziatti to talk about these brain failures.

Being focused, concentrating and meditating are ways to keep the brain working and avoid loss of memory. The mind has to be exercised in the same way as our body. It is also important to keep your senses sharpened. Good vision and hearing, for example, help us to retain information.

And why does memory fail? According to the psychiatrist Daniel Barros, our brain works in an economical way. At the time of remembering, it resorts to an information network and reconstructs what has occurred. We are also influenced by factors such as attention span, stress and the significance of particular facts.

and estimated time frame. At the same time, the volunteers were instructed to read and sign a participant consent form if they agreed, as well as to give their authorization to record the interaction. Next, the participants provided information about their age, gender and educational level.

From this time on, the interaction was recorded so that the Ten Emotion Heuristics could be employed and notes taken of the reading time and number of errors. The subjects read five texts aloud, two in Phase 1 and three in Phase 2; these were delivered one at a time and in a random order at each stage, as described previously.

After each reading, they responded to the SAM test for gathering data on emotional states and were encouraged to comment on the font types and sizes.

4.6 Data Analysis and Results

Each collected measure was assessed individually and the findings were also summarized in accordance with the guidelines of Stages Two and Three of the Hybrid Approach [43, 44] (i.e., generalization of results and incidence with octants), to obtain the final sets of emotions evoked.

By employing this method, after the measures (i.e., instruments) had been selected and the data collection, the results of each measure had to be generalized in terms of positive, neutral and negative [43, 44].

The generalization of the SAM data for Times New Roman is illustrated in Table 3, which describes the data of some participants. It should be noted that owing to constraints on the length of this article, it has only been possible to show the results of 10 out of the 60 participants in the case study (which were chosen at random). The signs "+", "0" and "-" mean positive, neutral and negative, respectively.

The same procedure was carried out for other measures of each font type and size – Discourse of the Collective Subject, number of errors and Ten Emotion Heuristics. Since the reading times were very close together, it was not possible to create a parameter to

Table 3: Generalization of SAM data of ten participants for Times New Roman.

Participant	Pleasure	Arousal	Control
1	+	+	+
2	+	+	+
3	+	+	-
4	+	-	0
5	+	+	-
6	+	+	+
7	+	0	+
8	+	+	+
9	+	0	+
10	0	-	+

determine whether they were positive or negative, and hence they were disregarded at this stage.

After this, the octants' incidence process took place. The Semantic Emotional Space contains eight sets for each dimension of emotion: four positive octants and four negative. In this stage, results obtained for each domain are related to their respective octants, as illustrated in Table 4. Neutral results are not included.

Table 4: Positive and negative octants of each dimension.

| Dimension	Positive Octants	Negative Octants
	5, 6, 7, 8	1, 2, 3, 4
Arousal	1, 2, 7, 8	3, 4, 5, 6
Control	6, 7, 8, 1	2, 3, 4, 5
Conduciveness	4, 5, 6, 7	1, 2, 3, 8

Figure 5 illustrates the incidence stage for Participant 6, in Phase 1 of the case study, which involves reading Times New Roman. It also shows the resulting emotional experience of this subject, i.e., the emotions contained in Octant 7 of the Semantic Emotional Space represent the set of emotions felt by this person. This underlines the fact that the participants may respond to their emotions in one way in their speeches while, when filling forms (e.g., SAM), they report the exact opposite. This became apparent when several different instruments were used to assess their emotional experience, as proposed by the Hybrid Approach.

The following subsections examine the data analysis and results obtained from both the statistical methods and the application of the Hybrid Approach in Phase 1 (font type) and Phase 2 (font size).

4.6.1 Font type.
Discourse of the Collective Subject
Words such as "normal" and "small" were often used in the volunteers' comments after they read the first text in Times New Roman. In contrast, Arial, the second font type, was described as "better". Some examples of sentences expressed by the participants are as follows: "*This font is pleasant, quite normal*" (first reading); "*I had some difficulty in viewing it, the letters were too small for me*" (first

Figure 5: Incidence of octants for Participant 6 in Phase 1.
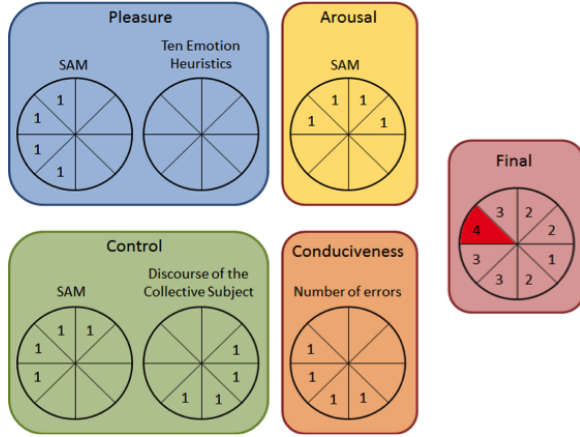
reading); "*This one is better to read than the first*" (second reading); "*I think the letters were bigger the second time, so it was more pleasant*" (second reading).

In cases where Arial was the first font type, the participants also often described it as "normal". However, after reading the second text in Times New Roman, they thought that this font type was worse than the one before. The following are some typical examples of the participant's comments: "*This is what I usually use, it was normal*" (first reading); "*It seems normal to me*" (first reading); "*I had some more difficulty*" (second reading); "*It looks like this font size is smaller*" (second reading).

Reading time and number of errors
Arial and Times New Roman had similar reading times. Table 5 summarizes the average values for each font type.

Table 5: Average reading time, in minutes, for each font type.

Newspaper article theme	Arial (min)	Times New Roman (min)
Science and Health	01:03	01:02
Economy	01:12	01:16
Sports	00:51	00:58
World	00:59	00:59
Technology	00:48	00:49

The number of errors was calculated by the nonparametric-paired Wilcoxon test [42], adopting a global confidence level of 95% ($\alpha = 0.05$). There was no significant difference between Arial and Times New Roman, the p-value ranged from 0.06 to 0.85 (p-value > 0.5).

Ten Emotion Heuristics
The measure adopted for motor expressions requires at least five different heuristics violated to classify a positive or negative emotional experience [16]. Only one participant had a final emotional experience after reading in Arial font type, by achieving a negative result.

The other subjects did not violate five heuristics, since it was not possible to classify their emotional experience.

SAM
The Wilcoxon paired test [42] was also used to analyze the SAM results; this is an evaluative measure of subjective feelings. A global confidence level of 95% ($\alpha = 0.05$) was adopted. There was a significant difference between font types in Control: Arial had a higher rate than Times New Roman, that is, the participants stated that there was more control when the text was in Arial. On the other hand, the method did not indicate significant differences in the areas of Pleasure and Arousal.

The Hybrid Approach
Table 6 illustrates the results obtained in the incidence rate of octants for ten participants in Phase 1, when Times New Roman was the first font type read. The gray cells highlight the highest frequencies of the respective octant, i.e., the resulting set of emotions.

Table 6: Individual incidence rate of ten participants when Times New Roman was the first reading.

Participant	Octants							
	1	2	3	4	5	6	7	8
1	2	1	0	1	2	3	4	3
2	2	1	0	0	1	2	3	3
3	2	3	2	1	2	1	2	3
4	0	0	1	2	3	3	2	1
5	1	2	1	2	3	2	3	2
6	2	2	1	2	3	3	4	3
7	1	0	0	0	1	2	2	2
8	3	2	1	0	1	2	3	4
9	1	0	0	1	2	3	3	2
10	2	0	1	1	1	3	2	2

Results from the participants that interacted with Times New Roman in the second reading, still in Phase 1, are illustrated in Table 7.

Table 7: Individual incidence rate of ten participants when Times New Roman was the second reading.

Participant	Octants							
	1	2	3	4	5	6	7	8
1	1	1	1	2	3	3	3	2
2	2	1	0	1	2	3	4	3
3	0	1	1	1	2	1	1	1
4	2	1	0	1	2	3	4	3
5	1	1	1	2	2	2	2	1
6	1	1	0	1	2	2	3	2
7	2	1	0	0	1	2	3	3
8	1	1	1	2	2	2	2	1
9	2	1	0	1	2	3	4	3
10	3	2	1	2	1	2	3	2

Octant 7 had the highest incidence rate with regard to all the participants in Phase 1, when the Times New Roman was read either

first or later. This means that most of the volunteers experienced positive Pleasure, Arousal, Control, and Conduciveness.

Table 8 illustrates how many times each octant was the resulting set of emotions for readings in Times New Roman.

Table 8: Overall incidence rate for Times New Roman.

Incidence rate for Times New Roman	Octants							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
First reading	1	3	3	0	5	12	16	10
Second reading	3	0	1	5	6	10	17	2

The findings for Arial were similar to those of Times New Roman. When all the participants are taken into account, Octant 7 had the highest incidence rate. Table 9 illustrates the number of times each octant was the final emotional experience.

Table 9: Overall incidence rate for Arial.

Incidence rate for Arial	Octants							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
First reading	3	0	1	3	4	9	21	9
Second reading	4	1	2	8	6	11	18	3

4.6.2 Font size.
Discourse of the Collective Subject
Most of the participants stated that size 8pt was unpleasant, "bad", tiring and hard to read. Some of the sentences used by the participants were as follows: "*Is it possible to make this font size bigger? It's hard to read*"; "*I thought this size bad and tiring, it was harder to read*"; "*This font is really small, because of the reading it was tiring*".

Words such as "good", "comforting" and "pleasant" were those that were most associated with 12pt. The volunteers described this size in phrases such as "*This font size was more comforting*"; "*This is a good size to use*"; "*This font size made me more satisfied with my reading*".

Most of the participants described 16pt as good, pleasant and comforting. However, it was interesting to note that a few people (approximately ten) also considered the last to be disquieting and tiring because it was so large. A few other comments were: "*The bigger the better*"; "*This was a good size*"; "*It wasn't more difficult, but it was less comforting*".

Reading time and number of errors
There were a few differences in the reading times between the font sizes. Table 10 illustrates average values for each.

The non-parametric statistical Friedman test [18] was used to investigate the number of errors among the font sizes, since there are three categories (8pt, 12pt and 16pt). A global confidence level of 95% (α = 0.05) was adopted. On the basis of the analysis, significant differences were found, with the p-value = 0.048 (p-value < 0.05). The results showed that 8pt led to a total number of 17 errors, while 12pt had 14 errors and 16pt, 6 errors.

Ten Emotion Heuristics
Three participants had a negative emotional experience in the 8pt

Table 10: Average reading time, in minutes, for each font size.

Newspaper article theme	8pt (min)	12pt (min)	16pt (min)
Science and Health	00:58	01:02	00:59
Economy	00:59	00:58	00:55
Sports	00:47	00:48	00:49
World	01:00	00:58	00:59
Technology	01:02	01:02	01:01

size. Other subjects did not violate five heuristics, and so it was not possible to classify their emotional experience.

SAM
The results of the SAM questionnaire were also evaluated by running Friedman test [18], with a global confidence level of 95% (α = 0.05). The ratings may range from one to nine, as cited in the Instantiation of the Hybrid Approach.

When all the texts were in Arial, the average ratings of the significant differences (p-value < 0.05) were at the levels illustrated in Table 11.

Table 11: Arial significant differences between sizes.

Font size	Economics - Pleasure	Sport - Control
8pt	6.17	5.67
12pt	7.83	7.67
16pt	5.17	7.33

The significant findings (p-value < 0.05) in Times New Roman are illustrated in Table 12.

Table 12: Times New Roman significant differences between sizes.

Font size	Economy		Sports	World	
	Pleasure	Control	Pleasure	Arousal	Control
8pt	5	4	5	3.83	4.33
12pt	6.33	7.17	6.17	4	5.83
16pt	6.67	7.33	7.67	5.5	7.3

Previous tables of significant differences between the font sizes indicate that 8pt had lower rates than all the other sizes. The readings in Arial had a higher rate when 12pt was used. On the other hand, when the texts were in Times New Roman, the highest rates were related to 16pt.

The Hybrid Approach
Table 13 summarizes the individual incidence rate of some of the participants in Phase 2 while reading texts using 8pt in Times New Roman.

An overview of the incidence rate is illustrated in Table 14 for 8pt, 12pt and 16pt sizes in Times New Roman.

In short, reading Times New Roman using 8pt evoked conductive emotions, but negative Pleasure, Arousal and Control (Octant 4).

In opposition to this, sizes 12pt and 16pt achieved Octant 7, which addresses positive values of Pleasure, Arousal and Control, and also includes conductive emotions.

The overall ratings in Arial for each size are listed in Table 15.

Table 13: Individual incidence rate of ten participants for 8pt in Times New Roman.

Participant	Octants							
	1	2	3	4	5	6	7	8
1	3	3	2	2	1	1	2	2
2	2	4	3	4	3	1	2	1
3	0	2	3	4	4	2	1	0
4	2	4	4	3	2	0	0	1
5	1	2	2	2	1	0	0	0
6	3	1	0	1	2	4	5	4
7	0	1	2	3	3	2	1	0
8	2	2	3	2	3	3	2	3
9	1	1	1	2	2	2	2	1
10	0	2	2	3	4	2	2	1

Table 14: Overall incidence rate of 8pt, 12pt and 16pt in Times New Roman.

Incidence rate	Octants							
	1	2	3	4	5	6	7	8
8pt	1	6	5	13	12	5	8	1
12pt	3	1	1	2	6	11	14	9
16pt	3	3	2	1	4	4	18	9

Table 15: Overall incidence rate of 8pt, 12pt and 16pt in Arial.

Incidence rate	Octants							
	1	2	3	4	5	6	7	8
8pt	3	3	3	4	3	3	17	9
12pt	1	2	1	2	4	11	19	20
16pt	1	1	1	2	5	9	20	6

In contrast with the results for Times New Roman, texts using 8pt in Arial led to an experience including emotions with positive Pleasure, Arousal, Control, and Conduciveness (Octant 7). Sizes 12pt and 16pt had the same final set of emotions, as well as Octant 7.

5 DISCUSSION

Previous studies of font types have mainly addressed issues related to visual aesthetics, readability and preference, although Kock [23] demonstrated that typefaces evoke different emotions in subjects. It was found that there was a gap in the literature with regard to the emotional effects of font types. This paper has carried out an extensive study on specific types - Arial and Times New Roman, while avoiding font styles such as Serif or Sans Serif, since each

font type has its own different and particular features, even in the same style.

The findings from measures related to font types suggested that Arial was more closely linked to positive emotions in the subject's discourse, whereas motor expressions caused a negative emotional experience. The Ten Emotion Heuristics did not achieve a final experience for most of the participants, which may be due to the short period of interaction. The reading time and number of errors were quite similar between Times New Roman and Arial, with no significant differences. On the basis of the SAM data, it was noted that the participants reported greater control from the reading of texts in Arial.

The Hybrid Approach [43, 44], that is adopted in this paper covering four out of five components of emotions, obtained the same results for both Times New Roman and Arial. The final emotional experience contained positive Pleasure, Arousal, Control, and Conduciveness.

With regard to the measures employed for the analysis of font sizes, the participants' comments suggest that the smallest, 8pt, was unpleasant, tiring and hard to read. The medium and large sizes were mainly associated with pleasant and comforting experiences. As was the case with the font types, the reading times were close among the font sizes. The largest number of errors was encountered in the readings of texts in 8pt, followed by 12pt, the lowest rate being 16pt. The 8pt size caused a negative emotional experience in the evaluation of motor expressions for three people, while the other sizes did not achieve a final emotional experience for this component. When the texts were in Arial, the size with highest Pleasure was 12pt, followed by 8pt and 16pt, respectively. Times New Roman had different findings in this area: the highest rate of Pleasure was in 16pt, and afterwards in 12pt and 8pt. The highest rate for Control in Arial was in 12pt and 16pt for Times New Roman, while the lowest for both was in 8pt. Arousal had significant values in Times New Roman, and had the highest rate for 16pt and lowest for 8pt.

One reason for the different results obtained for the font sizes among the Times New Roman and Arial may have been their x-height. This is a characteristic of typefaces which represents the letter "x" height in lower case [9]. Most of the participants believed that Arial in 12pt was bigger than Times New Roman in the same size; this may have been because the first has a larger x-height.

When Xavier's approach was adopted for font sizes, the resulting emotional experiences were as follows. The smallest size, 8pt, shown in Times New Roman achieved Octant 4 – negative Pleasure, Arousal and Control, but still conducive. The final emotions for the other sizes in Times New Roman, as well as all the sizes in Arial, were related to Octant 7 (positive Pleasure, Arousal, Control, and Conduciveness).

The literature on font sizes has linked ranges of sizes to dimensions of emotions, from 9pt up to 35pt. As mentioned previously, this paper is based on the assumption that sizes larger than 20pt are mainly used in titles, and are thus beyond the scope of this study. Apart from this, there is a divergence in the findings of previous works, in which 12pt had a lower Pleasure rate than 16pt, as well as the fact that the smallest size reported in [41] 9pt reached higher levels of Arousal and Dominance than 12pt and 16pt. However, this paper employed Friedman test [18] to analyze statistical and

significant differences, while the works referred to in the literature were based on mean values for each dimension.

In general, the null hypothesis was accepted for font types but rejected for font sizes. As described previously, Times New Roman and Arial had the same final emotional experience, while there were few differences among the 8pt, 12pt and 16pt font sizes.

6 CONCLUSION AND FUTURE WORK

The literature states that various combinations of font types and sizes evoke different emotions in users interacting with computing systems. This means that they influence a subject's decision-making, preferences and attitudes [31].

The main purpose of this paper was to explore the emotional effects of font types and sizes in greater depth. A number of different evaluative instruments and measures, for assessing emotional components, were combined to reduce the risk of obtaining inaccurate or distorted results. The data were analyzed individually, together with the statistical methods, and in groups, with the aid of the Hybrid Approach [43, 44] to assess emotional experiences.

This paper carried out a study with 60 participants with no previous formal knowledge of typography and their links to emotions (they were not design students, for example). In addition, four components of emotions were measured for each subject – cognitive appraisals, behavioral tendencies, motor expressions, and subjective feelings. In summary, among the font types, Arial tends to have positive emotions and as a result, is recommended for general use. Our findings showed significant differences between the font sizes, and this suggests that designers should treat the question of choosing their texts and font sizes with caution. Table 16 illustrates recommended use.

Table 16: Recommendation for font sizes.

Font size	Recommendation
8pt	No – generally related to negative emotions
12pt	Yes – usually associated with positive emotions
16pt	Caution necessary – related to both positive and negative emotions

Due to the wide use of CSS and to make alignments simpler, websites use px, em and rem as size units [5, 8]. It is noteworthy, however, that at the time of this research, the literature indicated the use of pt and that is why the study with this unit was included in this work. Usually, web browsers exhibit text in different units – for example, Google Chrome site inspector shows text size in px, while Mozilla Firefox in rem. According to browser use stats, Chrome was the most used between Jul 2020 and Jul 2021, thus examples of the recommended use were collected on it and will be presented next [14].

Taking into account the proportion between these different size units [15] , we have the conversion shown on Table 17.

An example of recommended use of the findings is presented in Figure 6, that shows a BBC news in 16px (i.e., 12pt) [25].

Figures 7 and 8 show websites from different areas (professional network and entertainment), such as LinkedIn [28] and Instagram [1], using 14px (i.e., 11pt). Given the proximity to 12pt, this size is also recommended.

Table 17: Conversion px to pt.

px	pt
12px	9pt
13px	10pt
14px	11pt
15px	11pt
16px	12pt
17px	13pt
18px	14pt
19px	14pt
20px	15pt
21px	16pt
22px	17pt

A bad design decision on font sizes is illustrated by Figure 9 [13], which presents text in 12px (i.e., 9pt) – close to 8pt, it may cause negative emotions on users and it is not recommended by our findings.

Further studies are needed on font types and on the effects of texts, topics and typographical features. Future work should address the question of whether typography (when compared with different topics) has as much impact as the content.

Figure 6: Recommended: use of medium font size (12pt).
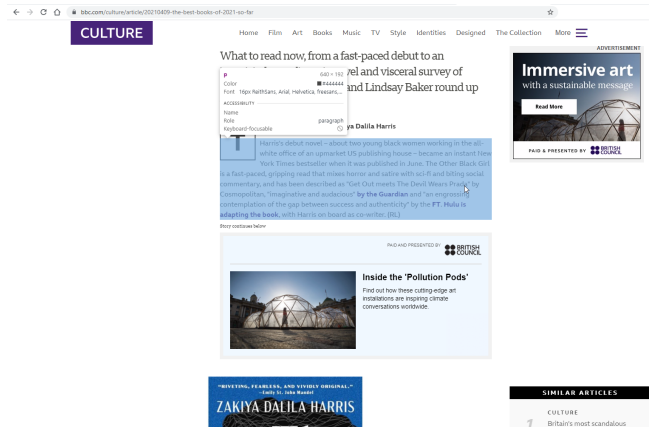
7 ACKNOWLEDGMENTS

REFERENCES

[1] 9GAG. 2021. *Instagram: 9GAG.* Retrieved August 17, 2021 from https://www.instagram.com/p/CSgL-NsnNPl
[2] Ahamed Altaboli. 2013. Investigating the effects of font styles on perceived visual aesthetics of website interface design. In *International Conference on Human-Computer Interaction.* Springer, 549–554.
[3] Ahamed Altaboli. 2014. Further Investigation of the Effects of Font Styles on Perceived Visual Aesthetics of Website Interface Design. In *International Conference of Design, User Experience, and Usability.* Springer, 199–207.
[4] Aries Arditi and Jianna Cho. 2005. Serifs and font legibility. *Vision research* 45, 23 (2005), 2926–2933.
[5] Joe Attardi. 2020. *Modern CSS.* Springer.

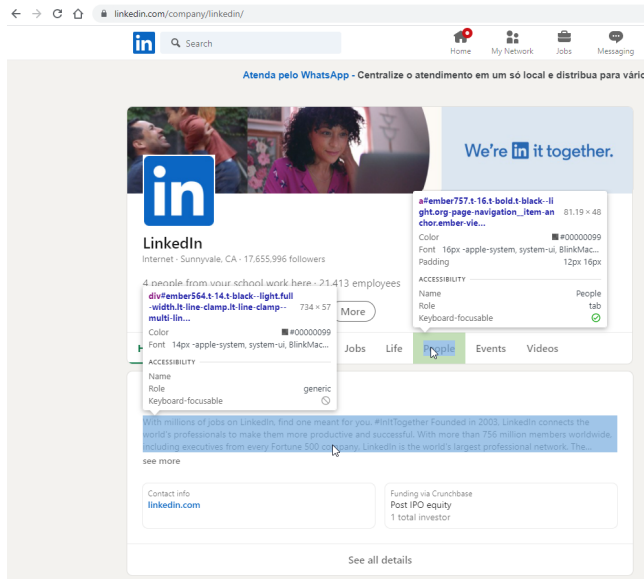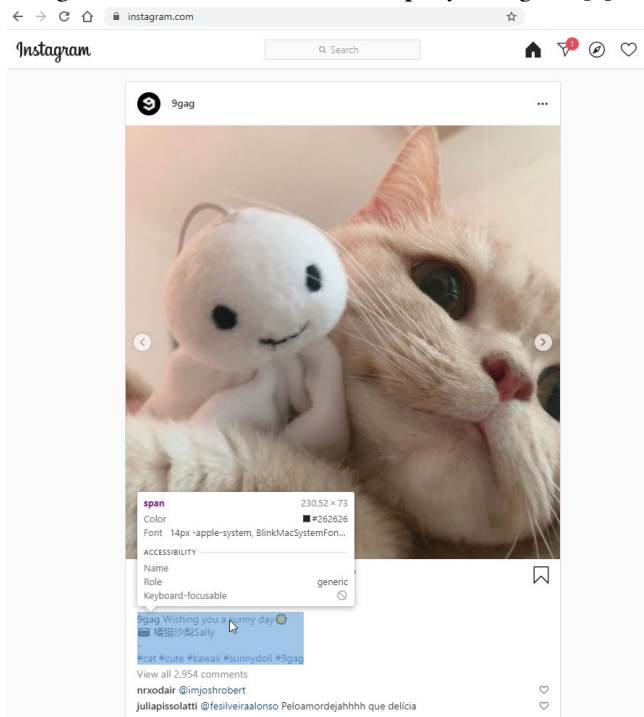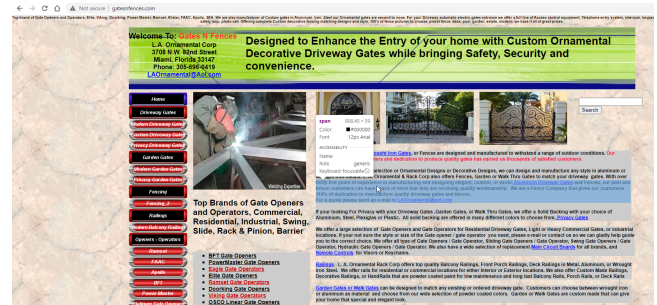**Figure 7: Recommended: use of 11pt by LinkedIn [28].**

Figure 8: Recommended: use of 11pt by Instagram [1].

Figure 9: Not recommended: use of small font size (9pt) [13].

annual conference, Vol. 2.

[8] Bert Bos. 2021. *CSS: em, px, pt, cm, in.* Retrieved August 17, 2021 from https://www.w3.org/Style/Examples/007/units.en.html

[9] Mark Boulton. 2009. Designing for the Web. *Penarth: Mark Boulton Design Ltd* (2009).

[10] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.

[11] Jeffrey F Cohn, Zara Ambadar, and Paul Ekman. 2007. Observer-based measurement of facial expression with the Facial Action Coding System. *The handbook of emotion elicitation and assessment* 1, 3 (2007), 203–221.

[12] Ricardo Colomo-Palacios, Cristina Casado-Lumbreras, Pedro Soto-Acosta, and Ángel García-Crespo. 2011. Using the affect grid to measure emotions in software requirements engineering. (2011).

[13] L. A. Ornamental Rack Corp. 2008. *Gates N Fences.* Retrieved August 17, 2021 from http://www.gatesnfences.com/

[14] Stat Counter. 2021. *Browser Market Share Worldwide.* Retrieved August 17, 2021 from https://gs.statcounter.com/browser-market-share

[15] Brian Cray. [n.d.]. *PXtoEM.com: PX to EM conversion made simple.* Retrieved August 17, 2021 from http://pxtoem.com/

[16] Eva De Lera and Muriel Garreta-Domingo. 2007. Ten Emotion Heuristics: Guidelines for assessing the user's affective dimension easily and cost-effectively. In *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21.* 1–4.

[17] Pieter MA Desmet and Paul Hekkert. 2002. The basis of product emotions. *Pleasure with products, beyond usability* (2002), 60–68.

[18] Milton Friedman. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association* 32, 200 (1937), 675–701.

[19] Richard H Hall and Patrick Hanna. 2004. The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour & information technology* 23, 3 (2004), 183–195.

[20] Eva Hudlicka. 2003. To feel or not to feel: The role of affect in human–computer interaction. *International journal of human-computer studies* 59, 1-2 (2003), 1–32.

[21] Nan Jiang, Xinyu Feng, Hui Liu, and Juan Liu. 2008. Emotional design of web page. In *2008 9th International Conference on Computer-Aided Industrial Design and Conceptual Design.* IEEE, 91–95.

[22] Jinwoo Kim and Jae Yun Moon. 1998. Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces. *Interacting with computers* 10, 1 (1998), 1–29.

[23] Beth Elynn Koch. 2011. *Human emotion response to typographic design.* Ph.D. Dissertation. University of Minnesota.

[24] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. 1997. International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention* 1 (1997), 39–58.

[25] Rebecca Laurence and Lindsay Baker. 2021. *The best books of 2021 so far.* Retrieved August 17, 2021 from https://www.bbc.com/culture/article/20210409-the-best-books-of-2021-so-far

[26] Fernando Lefevre and Ana Maria Cavalcanti Lefevre. 2014. Discourse of the collective subject: social representations and communication interventions. *Texto & Contexto-Enfermagem* 23, 2 (2014), 502–507.

[27] Jonathan Ling and Paul Van Schaik. 2006. The influence of font type and line length on visual search and information retrieval in web pages. *International journal of human-computer studies* 64, 5 (2006), 395–404.

[28] LinkedIn. 2021. *LinkedIn: Overview.* Retrieved August 17, 2021 from https://www.linkedin.com/company/linkedin/

[29] S Livingstone. 2003. The changing nature of audiences: from the mass audience to the interactive media user [online]. London: LSE Research Online. *Disponível:*

[6] Michael L Bernard, Barbara S Chaparro, Melissa M Mills, and Charles G Halcomb. 2003. Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. *International Journal of Human-Computer Studies* 59, 6 (2003), 823–835.

[7] D Beymer, D Russell, and P Orton. 2008. An eye tracking study of how font size and type influence online reading. People and computers XXII: culture, creativity, interaction: proceedings of HCI 2008. In *the 22nd British HCI Group*

http://eprints. lse. ac. uk/archive/00000417 [Consultado: agosto de 2013] (2003).

[30] Jo Mackiewicz and Rachel Moeller. 2004. Why people perceive typefaces to have different personalities. In *International Professional Communication Conference, 2004. IPCC 2004. Proceedings*. IEEE, 304–313.

[31] Morten Moshagen and Meinald Thielsch. 2013. A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology* 32, 12 (2013), 1305–1311.

[32] Donald A Norman. 2002. Emotion & design: attractive things work better. *interactions* 9, 4 (2002), 36–42.

[33] Donald A Norman. 2004. *Emotional design: Why we love (or hate) everyday things*. Basic Civitas Books.

[34] Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89, 4 (2001), 344–350.

[35] Klaus R Scherer. 2001. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research* 92, 120 (2001), 57.

[36] Klaus R Scherer. 2005. What are emotions? And how can they be measured? *Social science information* 44, 4 (2005), 695–729.

[37] Klaus R Scherer et al. 1984. On the nature and function of emotion: A component process approach. *Approaches to emotion* 2293, 317 (1984), 31.

[38] Mark Solms. 1997. Descartes' Error: Emotion, Reason, And The Human Brain.: By Antonio Damasio. New York: Putnam, 1994, xix+ 312 pp., $24.95. *Journal of the American Psychoanalytic Association* 45 (1997), 959–964.

[39] Dimitrios Tsonos, Kalliopi Ikospentaki, and Georgios Kouroupetrolgou. 2008. Towards modeling of Readers' Emotional State response for the automated annotation of documents. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 3253–3260.

[40] Dimitrios Tsonos and Georgios Kouroupetroglou. 2008. *A methodology for the extraction of reader's emotional state triggered from text typography*. Citeseer.

[41] MW Van Someren, YF Barnard, and JAC Sandberg. 1994. The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress* (1994).

[42] Frank Wilcoxon. 1992. Individual comparisons by ranking methods. In *Breakthroughs in statistics*. Springer, 196–202.

[43] Rogério Aparecido Campanari Xavier and Vânia Paula de Almeida Neris. 2012. A Hybrid Evaluation Approach for the Emotional State of Information Systems Users.. In *ICEIS (3)*. 45–53.

[44] Rogério Aparecido Campanari Xavier and Vânia Paula de Almeida Neris. 2014. Measuring the emotional experience of users through a hybrid semantic approach. In *Proceedings of the 13th Brazilian symposium on human factors in computing systems*. 226–235.