Paralinguistic Recommendations for Affective Word Clouds

Tugba Kulahcioglu
Rutgers University
Piscataway, NJ, USA
tugba.kulahcioglu@rutgers.edu

Gerard de Melo
Rutgers University
Piscataway, NJ, USA
gdm@demelo.org

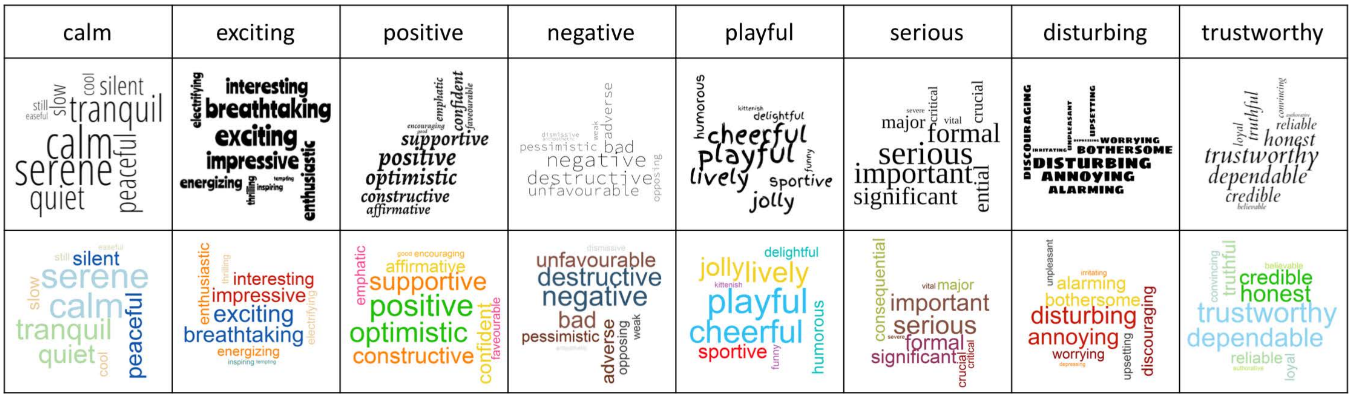

Figure 1: Word cloud examples using fonts (first row) and color palettes (second row) that are congruent with their message. Instead of randomly picking these paralinguistic signals as current tools do, in this study we determine their congruency with a set of eight affects, and propose a word cloud tool that helps users make congruent choices.

ABSTRACT

Word clouds are widely used for non-analytic purposes, such as introducing a topic to students, or creating a gift with personally meaningful text. Surveys show that users prefer tools that yield word clouds with a stronger emotional impact. Fonts and color palettes are powerful paralinguistic signals that may determine this impact, but, typically, the expectation is that they are chosen by the users. We present an affect-aware font and color palette selection methodology that aims to facilitate more informed choices. We induce associations of fonts with a set of eight affects, and evaluate the resulting data in a series of user studies both on individual words as well as in word clouds. Relying on a recent study to procure affective color palettes, we carry out a similar user study to understand the impact of color choices on word clouds. Our findings suggest that both fonts and color palettes are powerful tools contributing to the affect associated with a word cloud. The experiments further confirm that the novel datasets we propose are successful in enabling this. Based on this data, we implement a prototype that allows users to specify a desired affect and recommends congruent fonts and color palettes for the word cloud.

IUI '19, March 17–20, 2019, Marina del Rey, CA, USA

ACM ISBN 978-1-4503-6272-6/19/03…$15.00
https://doi.org/10.1145/3301275.3302327

CCS CONCEPTS

• **Human-centered computing → Information visualization**; **Visualization toolkits**.

KEYWORDS

affective interfaces, word clouds, typography, color palettes

ACM Reference Format:
Tugba Kulahcioglu and Gerard de Melo. 2019. Paralinguistic Recommendations for Affective Word Clouds. In *Proceedings of 24th International Conference on Intelligent User Interfaces (IUI '19).* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3301275.3302327

1 INTRODUCTION

When we see a social media posting, for instance, we are affected not just by the message itself but also by the way it is presented to us. Paralinguistic signals refer to auxiliary aspects beyond the essential units and structures of language [2]. They possess the potential to modulate it in diverse ways. For instance, a particular tone of voice may dramatically alter the underlying conceptual and pragmatic meaning conveyed by a speaker [32, 33]. While typically associated with verbal language, paralinguistic phenomena also manifest in written language [2, 25]. Psychological studies show that different typefaces may exhibit associations with divergent affective attributes [15, 36]. Accordingly, the congruency between the textual content and the selected font has shown to have a pronounced impact on the affective interpretation of the text [31, 36]. For instance, different typefaces may result in distinct ratings for the same textual content with respect to its perceived *excitingness*.

It has been observed that the response times of users decrease when fonts that are congruent with the message are used [12, 22]. Similar ties have also been observed for individual colors [4, 26, 47] and color palettes [3].

Word clouds appeal to people of all ages and occupations [43]. As analytic tools, word clouds have a number of notable drawbacks, particularly with regard to the relationship between size and salience [1, 10]. However, Viegas et al. show that the most common uses are in fact for non-analytic purposes, such as preparing a gift, introducing a topic to students, or just playful and creative exploration [43]. Viegas et al. focus on a specific type of word cloud, namely those generated by Wordle, finding that a key reason for the tool's greater success in comparison with alternative word cloud tools is the "emotional impact" it creates through the tight layout it applies, and the font and color palette options it provides [43]. This is exemplified by user comments from the study that emphasize the importance of these signals, e.g.:

> "Wordles have more emotional emphasis, colors, and layouts to enhance the meaning" [43]

> "Wordles are colorful, more visually interesting, more of an emotional response and connection with the viewer than the tag clouds." [43]

There are also use cases, such as the visualization of restaurant reviews or sentiment analysis results, where the primary goal of the cloud is to convey the sentiment of its input. In such cases, the affective perception of the cloud plays a critical role. Affect, thus, is a key factor to be considered in the design of word clouds. Typography and color are two important examples of paralinguistic choices that may influence the affects associated with a word cloud. Figure 1 demonstrates how choices of fonts and colors may evoke very different affective associations.

Designers routinely expend significant effort in making paralinguistic choices that accord with the message that visual material is intended to convey. Thus, a word cloud produced by a designer would typically incorporate colors and fonts that are congruent with the message of the cloud. On the contrary, this may not be the case for non-professional users, for a variety of reasons, including a lack of such awareness, or an unwillingness to put in the time and effort. Despite the established relationship between paralinguistic signals and affect, and the known impact of their congruency, word cloud tools thus far neglect to support the users in making semantically informed paralinguistic choices. Towards the aim of enabling such support, our contributions in this study are as follows:

(1) We propose computational methods to obtain affective relationships based on existing ones (Section 3).
(2) Through two user studies[1], in Sections 4 and 5, we show that congruent fonts and color palettes, respectively, better reflect word clouds with affective content, for a range of different affects.
(3) We find that fonts and color palettes have complementary strengths in conveying affects.
(4) Existing word cloud generators choose fonts and colors based solely on aesthetics or even randomly, neglecting their congruency with the intended affect. We instead show how they can be utilized to make paralinguistic recommendations based on user-specified affects.

[1]All studies in this paper have IRB approval.

2 RELATED WORK

We review related work on word clouds, fonts, and color.

2.1 Word Clouds

Word clouds have been of substantial interest to the academic community, especially with regard to the employed layout algorithms [9, 44]. Several studies [23, 34, 48] focus on semantic relationships of the words, e.g. placing semantically similar words closer to each other. There are also studies aiming to create comparison clouds by combining visualizations for multiple texts [7, 24].

Wordle [40] is a widely appreciated word cloud generation tool that aims to create more pleasing word clouds ("wordles") by adopting tighter layouts, e.g. allowing a tiny word to appear within a character from a larger word, with several font and color palette options. Through a user survey and analysis of resources on the Web, Viegas et al. [43] analyze the users and common use cases of wordles. The results show that it attracts people of all ages, most of whom are educators or students (29%), with no other occupation accounting for more than 6%. The main reason for their broad appeal is found to be the tool's power to create an emotional impact by means of fonts, colors, and layouts. Two other important reasons are their attention-grabbing nature and the organic nonlinear layout. The use cases of wordles vary from education (e.g., introducing a topic to students using Wordle) to gift-giving (e.g., creating a wordle from wedding wows), and even *guess the cloud* games[2]. Another insightful result of this analysis was that 88% of the users reported that they feel creative when using Wordle.

ManiWordle [17] enables custom manipulations in a wordle, based on changes with regard to fonts, colors, and the layout, both at the cloud and word level. For example, one may change the color of a selected word so as to emphasize it. Although this provides more control, users did not report feeling a strong difference in terms of creativity compared to regular wordles. WordlePlus [14] further extends ManiWordle to allow natural interaction on pen and touch-enabled tablets. EdWordle [45] facilitates multi-word editing, while preserving the neighborhood, i.e., keeping non-edited words close to their original locations. It applies a local re-wordle algorithm that re-arranges the words to close the gaps.

We are not aware of detailed studies on the use of fonts or colors in text visualization from a semantic congruency perspective. A related study [5] uses font attributes, such as underlining or small caps, to distinguish set membership in set visualizations, including for emotions. Wecker et al. [46] propose using font properties such as size and color to highlight the sentiment polarity of text passages. However, they do not consider typefaces or their semantic connotations.

2.2 Fonts

In the following, we consider related research on affective attributes of fonts, and studies revealing their impact.

[2]http://guessthewordle.weebly.com/

Affective Attributes of Fonts. Through a crowdsourced study, O'Donovan et al. [31] associate 200 diverse fonts with 37 semantic attributes (e.g., *happy*). Specifically, they request participants to pick one of two presented fonts for a given attribute, and then aggregate these choices to assign fonts a series of scores for each attribute. Another online survey [37] assesses the characteristics of 20 fonts with respect to 15 adjective-based scales (e.g. *stable–unstable*). Similarly, many font-focused websites[3] allow for tagging fonts with various attributes, some of which are semantic in nature. Further studies [41, 42] explored the relationship between visual font characteristics and taste attributes (sweet, sour, etc.) through user studies, concluding that round-shaped fonts are associated with a sweet taste.

Kulahcioglu and de Melo [20] extend the dataset from O'Donovan et al. [31] to a larger number of fonts using a k-Nearest Neighbors (k-NN) approach with deep Convolutional Neural Network (CNN) embeddings as a similarity metric between fonts. FontLex [19] is a lexicon providing word–font associations for arbitrary English words, automatically induced using affective associations of words and fonts, the latter based on word embedding similarities. In this study, we propose new computational methods that are based on affect relationships to obtain font–affect scores, instead of using word similarities.

Psychological Effects of Font Selection. We next survey studies that investigate the effect of font characteristics on perception.

Several Stroop-style studies have been conducted in this regard. Hazlett et al. [12] asked users to judge whether a displayed word is positive or negative, comparing 5 fonts and 25 words that are all strongly associated with positive or negative emotions. The results indicate that congruent typefaces yield faster responses. Lewis & Walker [22] ask users to press a left hand key if the words *slow* or *heavy* appear, versus a right hand one if *fast* or *light* appears. In a second experiment, they display related words (e.g., *fox*) instead of the original words (e.g., *fast*) to ensure that the user needs to grasp the meaning of the displayed word. In both experiments, they repeat the tasks with congruent and incongruent fonts, finding that the former significantly reduce the response time.

There are also studies that use survey-style approaches. Juni & Gross [15] present newspaper articles using two different fonts, finding that the same text may be perceived as more humorous or angry depending on the chosen font. Shaikh [36] collects ratings to understand the perception of the document and the perceived personality of the author. The results reveal that fonts with different congruencies heavily alter the perception of a document, while congruent and neutral fonts appear to evoke similar perceptions of an author's personality.

Another study by Shaikh et al. [38] implements a related test for the perception of emails. The results suggest that incongruent fonts may result in different perceptions of an email. A similar study on the perception of a company website [35] demonstrates that neutral and low congruency fonts can negatively affect a company's perception in terms of professionalism, believability, trust, and intent to act on the site.

[3]For instance, https://fontsinuse.com/, http://www.dafont.com, and http://www.1001fonts.com.

Table 1: Affective attributes associated with fonts

Index	Attribute Name	Index	Attribute Name
1	calm	5	playful
2	exciting	6	serious
3	positive	7	trustworthy
4	negative	8	disturbing

Marketing research has a strong interest on the impact of font choices in packaging design. Fligner [11] shows that *natural* looking fonts in a packaging design increase the perceived *healthfulness* of products, especially if the products' intrinsic (e.g., being fat-free) and extrinsic cues (e.g., being sold at Whole Foods Market) also support this perception. Childers & Jass [8] show that font choices have an impact on product perception for both high and low user engagement levels.

2.3 Color Palettes

The impact of colors has been studied extensively [4, 47]. Certain colors have been tied to specific forms of impact on cognitive tasks, although their effect may vary between tasks [28]. For instance, a choice of red has been found to be beneficial for detail-oriented tasks, but it does not have the same effect on creative tasks. Similar to fonts, colors have also been of special interest in marketing research [21, 39].

In contrast, the use of multiple colors together as a palette has not been studied to the same extent, as explained by Bartram et al. [3]. To address this, they recently conducted a study of color palette choices for a set of eight affects, which is also the source of the color palettes in our work. The study establishes several relationships between affects and perceptual color properties (hue, chroma, and lightness). As an example, the attributes *calm*, *playful*, and *positive* are found to be associated with the lightest colors. Among the core affects, *calm* is found to evoke the strongest color preferences. Other notable findings include that highly saturated light colors are not a good fit for *serious*, *trustworthy*, or *calm*, and that light colors, in general, are not very successful in conveying a *negative* sentiment.

3 AFFECTIVE FONTS

In this section, we present our method to computationally obtain font attribute associations using crowdsourced seed data, and we evaluate that method through a user study.

3.1 Method

Our goal is to compute font attribute vectors $\mathbf{v}_a \in \mathbb{R}^{|\mathbb{F}|}$ which, for a given affective attribute $a \in \mathbb{A}$, reflect the perception of fonts $f \in \mathbb{F}$ with respect to the affect. The set of affective attributes $\mathbb{A}$ is given in Table 1 and was chosen because it includes the core affects in the PAD emotional state model [27] and is used in a previous study on color [3], allowing us to compare the impact of fonts and color palettes. Given the indices i from Table 1, we denote each affective attribute as a_i and also use the notation v_i as a shorthand for v_{a_i}, e.g., v_1 as the vector for the affective attribute *calm*. Each dimension of a particular $\mathbf{v}_a$ reflects the congruency of attribute a with respect to a different font $f \in \mathbb{F}$.

In our study, the set $\mathbb{F}$ consists of 200 different fonts, taken from the aforementioned study by O'Donovan et al. [31], due to it being the largest of its sort available for download. Their data provides us with $|\mathbb{F}|$-dimensional vectors $\mathbf{x}_t$ for a set of 31 font traits $t \in \mathbb{T}$, providing the scores for that trait over all fonts in $\mathbb{F}$. The set $\mathbb{T}$ of 31 traits they consider includes visual ones such as *thin*, *angular*, but also more subjective ones. We rely on the following multi-pronged procedure to obtain the desired vectors $\mathbf{v}_a$ for affective attributes $a \in \mathbb{A}$ from this data.

Scores for "Calm", "Playful": Fortunately, the crowdsourced data already includes two of our eight considered affective attributes. Thus, we directly obtain $\mathbf{v}_1$ and $\mathbf{v}_5$ by selecting the relevant $\mathbf{x}_t$.

Scores for "Positive": We obtain scores for the attribute *positive* by clustering all emotional traits included in the data crowdsourced by O'Donovan et al. [31]. For this, we manually filter the set of font traits $t \in \mathbb{T}$ so as to retain only the ones of a strong emotional nature. Then we apply k-means clustering and obtain the following three clusters:

- C_1: *bad*, *boring*
- C_2: *happy*, *playful*, *attractive*
- C_3: *calm*, *charming*, *fresh*, *friendly*, *gentle*, *graceful*, *soft*, *warm*

The first observation about these clusters is that C_1 contains traits with a negative connotation, while the other two each contain positive traits. A detailed analysis reveals that C_2 includes high-arousal positive emotions, whereas C_3 includes low-arousal ones. We compute score vectors $\mathbf{x}_C$ for each positive cluster $C \in \mathbb{C} = \{C_2, C_3\}$ as

$$\mathbf{x}_C = \sum_{t \in C} \sigma(C, t)\, \mathbf{x}_t, \tag{1}$$

where the weight of each trait is calculated as

$$\sigma(C, t) = \frac{1}{|C| - 1} \frac{\sum\limits_{\substack{t' \in C \\ t \neq t'}} d(C, \mathbf{x}_{t'})}{\sum\limits_{t' \in C} d(C, \mathbf{x}_{t'})}. \tag{2}$$

with $d(C, \mathbf{x})$ corresponding to the distance between the cluster centroid and a trait-specific vector $\mathbf{x}$.

Using these two cluster centers, the scores for the target attribute *positive* is obtained as:

$$\mathbf{v}_3 = \frac{1}{|\mathbb{C}|} \sum_{C \in \mathbb{C}} \mathbf{x}_C \tag{3}$$

Scores for "Exciting", "Serious", and "Negative": For the attributes *exciting*, *serious*, and *negative*, we take advantage of the antonymy relationships and use scores for *calm*, *playful*, and *positive*, respectively, as:

$$\mathbf{v}_i = \mathbf{1} - \mathbf{v}_{\alpha(i)}, \tag{4}$$

where $\mathbf{1} \in \mathbb{R}^{|\mathbb{F}|}$ is a vector of ones, $i = 2, 4, 6$ are indices from Table 1, and $\alpha(i)$ denotes the respective antonym of affective attribute i. We follow this method based on the observation that a font that is least representative of an attribute is a candidate to best represent the opposite attribute. However, this method assumes that a font cannot be a good representative for each of the two opposing attributes, which may not always hold, given different contexts the fonts could be used in.

Scores for "Trustworthy" and "Disturbing": The PAD model [27] posits that complex emotions are composed of more basic ones. Following Bartram et al. [3], *trustworthy* can be defined as *positive + calm*. To obtain font-specific scores, we accordingly average the values for *positive* and *calm* as $\mathbf{v}_7 = \frac{1}{2}(\mathbf{v}_3 + \mathbf{v}_1)$. We performed an analogous derivation to obtain values for *disturbing* using *negative* and *exciting*, i.e. $\mathbf{v}_8 = \frac{1}{2}(\mathbf{v}_4 + \mathbf{v}_2)$.

3.2 Results

The top five congruent fonts for each affect in Table 1, as determined by the above process, are depicted in Figure 2. Top fonts for a particular affect seem to have similar visual characteristics, except for *exciting* and *negative*. To investigate this, we analyze Figure 3, which plots the distribution of fonts based on the *positive* and *exciting* attribute scores[4]. We find that the top fonts for *exciting* and *negative* attributes have a wide range of values in the other scale. For example, fonts with the highest scores for *exciting* exhibit a large degree of variance with regard to their scores for the attribute *positive*. The same applies for *negative*. More specifically, fonts with lowest scores for *positive* have a wide range of scores for *exciting*. This can explain why for these two attributes, fonts with the highest congruency exhibit different visual characteristics. The top fonts for other attributes, on the other hand, are found to reside in a smaller area in the chart, i.e., have a narrower range of scores in the other scale. As an example, fonts with the highest scores for *positive* are concentrated in a region with low scores for *exciting* (between 10 and 25). This analysis is further expanded and validated in Section 6, where we group the top *negative* and *positive* fonts based on their scores for *calm* and *exciting*.

Overall, the majority of fonts considered are deemed highly *positive* and *calm*. Figure 3 also provides the categories to which the fonts belong. *Handwriting* typefaces are designed to give the impression of being hand-rendered. The characters of *monospace* typefaces occupy equal horizontal space. *Serif* typefaces have small lines attached to the end of the strokes in their characters, whereas *sans-serif* ones lack such attached lines. *Display* typefaces do not share typical typographic properties other than a low degree of legibility when used for body text, so they are reserved mostly for headings and other kinds of display purposes.

When we analyze the score distributions in conjunction with their font categories, we find that *Serif* fonts appear to have higher *positive* scores compared to *sans serif* fonts. *Display* fonts are found to be *exciting*, which accords with their decorative nature. With significantly fewer instances in the dataset, *monospace* and *handwiriting* fonts are scattered along a wide range of different values.

3.3 Evaluation (User Study I)

We evaluate the above font score computations through a user study, in which we collect user preferences among congruent, incongruent, and neutral fonts for affective word clouds.

[4]These attributes correspond to the opposites of *negative* and *calm*. For instance, the fonts that correspond to the lower end of the *positive* scale are the fonts with high *negative* scores.

calm	exciting	positive	negative	playful	serious	disturbing	trustworthy
calm	exciting	positive	negative	playful	serious	disturbing	trustworthy
calm	EXCITING	positive	negative	playful	serious	DISTURBING	trustworthy
calm	exciting	positive	negative	playful	serious	disturbing	trustworthy
calm	exciting	positive	NEGATIVE	playful	serious	disturbing	trustworthy
calm	exciting	positive	negative	playful	serious	DISTURBING	trustworthy

Figure 2: Top five congruent fonts obtained for each of the eight affects used in this study.

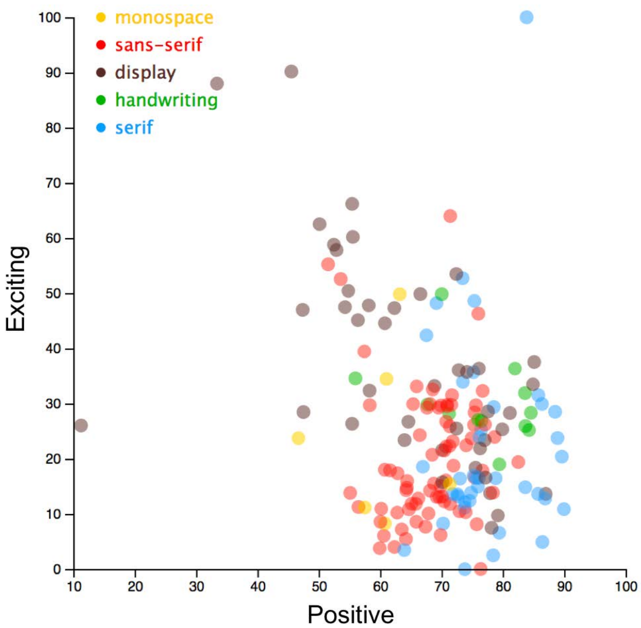

Figure 3: Scatter plot of the resulting font scores based on the *positive* and *exciting* attributes (scaled to [0, 100] range).

Option 1	Option 2	Option 3	Option 4	Option 5
serious	serious	serious	serious	serious

Figure 4: A sample task from User Study I. The third and fifth images are generated using incongruent fonts, the second one uses a neutral font, and the first and fourth images use congruent fonts.

Hypothesis. Fonts with higher congruency scores for a given attribute are assessed as better representing that attribute than fonts with lower scores.

Participants and Method. We recruited 40 participants via Mechanical Turk, all from the United States, with at least 50 approved hits and an overall approval rating of 90% or more. Participants were paid $0.01 for each task. We rely on a within-subject design, and perform counterbalancing. The study involves 50 tasks for each participant, consisting of 6 tasks for each of the 8 affective

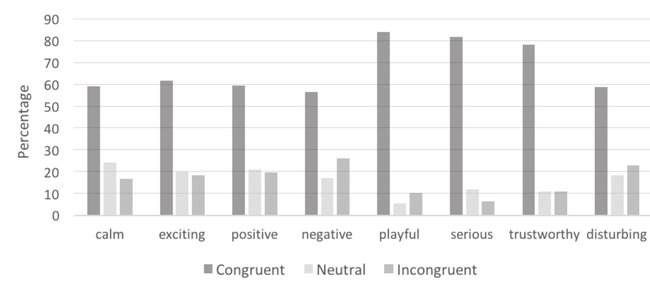

Figure 5: Results of User Study I, evaluating the font scores with respect to renderings of single words. The user preferences for congruent, neutral, and incongruent font choices are compared for each of the 8 affective attributes. With uniformly random selections, these, respectively, have a 40%, 20%, or 40% chance of being selected, since two each out of five options are congruent or incongruent, while one option out of five uses a neutral font.

attributes, and 2 additional validation tasks. Each task presents the name of the affective attribute using 5 different fonts, as shown in Figure 4. To allow for comparison with the color palette study by Bartram et al. [3], we select two congruent, two incongruent, and a neutral font. The congruent ones are selected randomly amongst the six most congruent fonts for the corresponding affect, while the two incongruent fonts are selected randomly among the least congruent six fonts. The fifth font is selected randomly among the three fonts that are in the middle of the ranked font list for each affect. The words are presented in random order within each task. The validation tasks include words written with 1 congruent font and 4 incongruent fonts. Participants were instructed to "*Pick the image that best represents the word*", with an additional detailed version given as "*Select the image that you think best reflects the meaning of the word shown in the images.*".

Results and Analysis. The results of our study are summarized in Figure 5. Across all attributes, fonts determined to be congruent are preferred by the participants, according with our hypothesis, while fonts determined to be incongruent are strongly dispreferred. We conducted chi-square goodness of fit tests of user preferences for each affect based on the three font category choices (congruent, neutral, and incongruent fonts). Table 2 provides results of these analyses, which are found to be statistically highly significant for

Table 2: Chi-square goodness of fit test results for the three user studies. Each affective attribute is represented by the corresponding index (*i*) as defined in Table 1. For all tests, expected values are specified as 0.4, 0.2, and 0.4 for the categories *congruent*, *neutral*, and *incongruent*, respectively, and significance level is set as 0.05. Results for each affective attribute in each user study are found to be highly statistically significant, as for each analysis $p < 0.001$.

	Study I		Study II		Study III	
i	$X^2(2)$	$p <$	$X^2(2)$	$p <$	$X^2(2)$	$p <$
1	56.79	0.001	54.14	0.001	121.84	0.001
2	56.33	0.001	16.80	0.001	44.96	0.001
3	48.10	0.001	43.78	0.001	43.13	0.001
4	29.03	0.001	51.17	0.001	89.54	0.001
5	195.07	0.001	70.15	0.001	115.02	0.001
6	180.03	0.001	142.06	0.001	43.09	0.001
7	149.29	0.001	86.68	0.001	20.79	0.001
8	38.94	0.001	126.69	0.001	76.75	0.001

each of the eight affects. The strongest statistical differences are observed for *playful*, *serious*, and *trustworthy*.

Given that the scores for *calm* and *playful* were obtained via crowd-sourcing, in our analysis, they may serve as ground truth benchmarks as to what range of scores we are to expect from high-quality human-provided ratings. Fonts rated strongly as *calm* appear to be less preferred than those for *playful*, possibly owing to the fact that even regular fonts may also have a tendency to be perceived as calm. Indeed, the median congruency score in our data for *calm* was 76.3%, while for *playful*, it was 34.5%, confirming that neutral fonts are more likely perceived as calm.

Both *exciting* and *serious* acquire similar results to the baselines, which suggests that our method of computing their scores as reversed opposites suffices to select fonts perceived as congruent. The clustering approach used for *positive*, from which in turn scores for *negative* are derived as well, appears to yield reasonable but not overly strong ratings. While this might stem from inaccuracies in the automatic clustering, it may also be the case that it is less trivial to convey positive and negative sentiment than to convey attributes such *playful* and *exciting*. The most peculiar finding is that fonts with high scores for the attribute *trustworthy* manifest stronger preferences than those for the attributes used to compute its values (namely *positive* and *calm*). Despite being computed analogously, the ratings for *disturbing* do not exceed those for *negative* and *exciting*.

4 AFFECTIVE FONTS IN WORD CLOUDS

Using the font scores obtained in the preceding sections, we seek to understand the impact of the affective nature of fonts on affective word clouds. In particular, we shall determine to what extent users may prefer fonts that accord with the content of a word cloud with regard to affect.

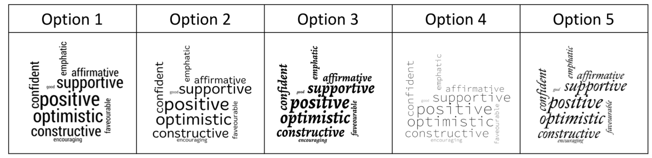

Figure 6: A sample task from User Study II. The third and fifth images use congruent fonts, the first uses a neutral font, and the second and fourth ones use incongruent fonts. The participants are asked to pick the image that best represents the words in the word cloud.

4.1 User Study II

Through a user study on Mechanical Turk, we evaluate the impact of fonts on word clouds with affective content.

Hypothesis. Word clouds using fonts determined to be congruent are assessed as being more representative of pertinent affect-evoking words than word clouds using fonts determined as neutral or incongruent.

Participants and Method. The participant and method information is as in Study I, except that the displayed renderings here include words clouds instead of single words. A sample task is given in Figure 6. For each affective attribute, we created word clouds of 10 words, one of which is the affective attribute name itself, coupled with 9 further semantically related words to avoid confounding effects potentially caused by irrelevant words. For the same reason as earlier, each task includes 5 word clouds with the same content and layout, just using a different font. As earlier, we randomly select two congruent fonts[5], two incongruent ones, and a neutral one, following the procedure for Study I.

4.2 Results and Analysis

The preference frequencies are visually presented in Figure 7, and chi-square goodness of fit test results are given in Table 2. Similar to Study I, results for each of the eight affective attributes are found to be highly statistically significant. The strongest statistical differences are observed for *serious* and *disturbing*.

The results are in general consistent with those from the first study, as both show strong support for congruency with *serious* and *trustworthy*. Although the differences are less pronounced than earlier, across all attributes, congruent fonts were chosen notably more frequently than incongruent ones, and incongruent fonts were chosen substantially less frequently than chance would predict, i.e., $\frac{2}{5} = 40\%$. It is observed that the congruent font options are more frequently preferred for complex affects compared to core affects.

For *calm* and *positive*, neutral fonts were also chosen in many cases. As explained for Study I, and shown in Figure 3, larger numbers of fonts might appear somewhat calm or positive, and hence fonts in the middle of the ranked lists, which we assumend as *neutral*, may be more congruent.

Interestingly, *disturbing* received higher scores in this experiment compared to Study I. This may result from the layout of the

[5]Figure 1 provides samples of word clouds with congruent fonts from this user study.

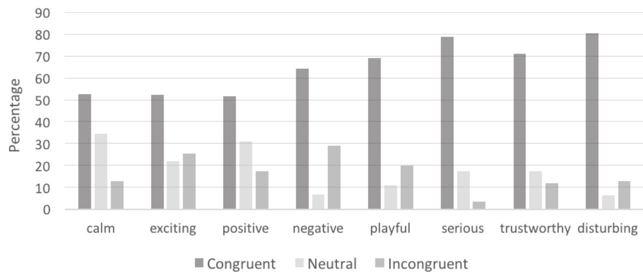

Figure 7: Results of User Study II, providing percentages of user preferences for congruent, incongruent, and neutral font choices in the word clouds. With uniformly random selections, the expected values for the congruent and incongruent options are 40%, while for neutral it is 20%.

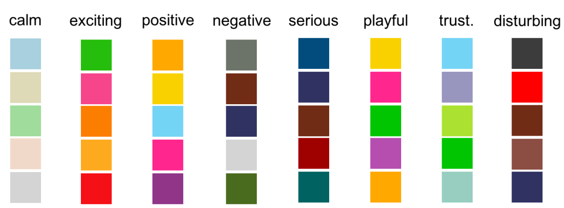

Figure 8: Congruent color palette samples from Bartram et al. [3].

word cloud, which mixes horizontal and vertical orientations, different font sizes, and different alignments, and, hence, in itself may already embody an appearance congruent with the notion of being *disturbing*. This circumstance may also explain the comparably lower scores for congruence with *calm* in comparison with Study I. We conclude that, in addition to the layout, font congruency as well merits significant consideration when designing word clouds.

5 AFFECTIVE COLOR PALETTES IN WORD CLOUDS

We wish to understand the impact of affective color palettes on affective word clouds, specifically, to what extent users prefer color palettes that match the content of the word cloud affectively. To achieve this, we rely on the data from Bartram et al. [3] and carry out a user study using these palettes to create affective word clouds.

5.1 Data

The Bartram et al. study considers the same affect categories as this study, and their methodology to obtain the color palettes is as follows. First, 8,608 images tagged with one of the eight affective categories are analyzed to find the most common colors for each category. Then, with the support from a visualization color expert, a set of 41 colors is determined, which combines the most representative colors for all eight affects.

Using this set of colors, a user study is carried out requesting participants to design color palettes for one of the affect categories

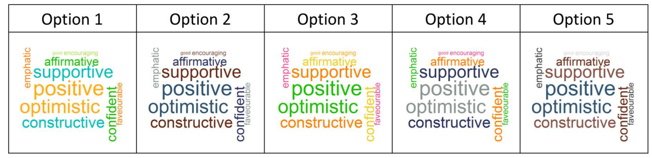

Figure 9: A sample task from User Study III. The second and fifth images are generated using incongruent color palettes, the fourth one uses a mixed color palette, and the first and the third images use congruent color palettes. The users are asked to pick the image that best represents the words in the word cloud.

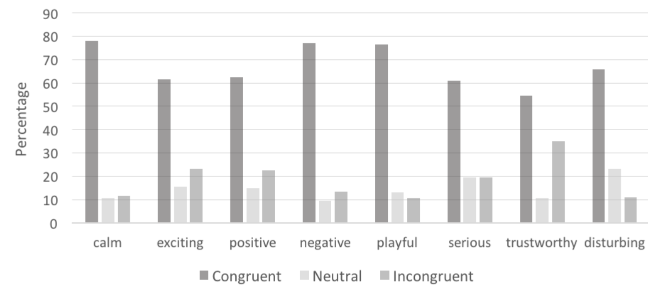

Figure 10: Results of User Study III, providing percentages of user preferences for congruent, incongruent, and neutral color palette choices in the word clouds. For uniformly random selections, the respective expected values for the congruent, incongruent, and neutral options are 40%, 40%, and 20%.

to be used in either a bar chart or a map visualization. The frequencies from this user study reveal the preferred colors for each affect, which are then used to create weights to come up with a palette weight concept. A weight for a palette is determined using the frequencies of the colors used in the palette. Finally, another user study is carried out to verify the results of the previous one, by generating palettes of different weights, and asking users to pick the best one. The results suggest that these weights indeed can be a good predictor of user preferences towards the color palettes.

We use color palettes from the second user study in their paper, which reveals the most preferred colors for each affect. Specifically, we rely on the palettes in Figure 7 from their paper [3] to determine the colors. Figure 8 in our paper shows a congruent color palette sample for each of the eight affect categories. The color palettes for complex affects appear to follow a pattern based on their underlying core affects. As an example, the colors for *disturbing* seem to be a combination of colors for *negative* and *exciting*.

5.2 User Study III

Through a user study carried out on Mechanical Turk, we evaluate the impact of color palettes on word clouds with affective content.

Hypothesis. Word clouds using colors determined to be congruent are assessed as being more representative of pertinent affect-evoking words than word clouds using mixed or incongruent color palettes.

Participants and Method. The participant and method information is as in Study I and II, except that in this study, we use different color palettes in each option of the tasks, while keeping other signals, namely font and layout, the same. A sample task is given in Figure 9. Two of the color palettes used in a task are *congruent*, two of them *incongruent*, and one is *neutral*. The procedure to create the color palettes is as follows. We use Bartram et al.'s Figure 7 [3] to obtain a list of *congruent* colors, referred to as B_i, for each attribute i as defined in Table 1. To generate a *congruent* color palette[6] for attribute i, we randomly pick five colors from B_i. To generate an *incongruent* color palette for attribute i, we randomly pick five colors from B_j, where j is the index of the opposite attribute[7]. To generate a *neutral* palette, we randomly pick two colors from B_i, two other colors from B_j, and one more color from $B_i \cap B_j$.

5.3 Results and Analysis

As summarized in Figure 10, and in Table 2, the results are consistent with our hypothesis: Participants exhibit a preference for congruent color palettes. Similar to the results of our previous two user studies, results for each of the eight affective attributes are found to be highly statistically significant. This is particularly pronounced for *calm*, *negative*, and *playful* for User Study III, while the results appear comparably less strong for *exciting*, *serious*, *positive*, and *trustworthy*. These findings mirror those from Bartram et al. [3], in which *trustworthy* and *serious* were not strongly associated with specific colors, whereas the palettes for *calm* and *playful* had highly weighted colors, reflecting a strong preference for their respective affects. One exception is *exciting*, which had highly weighted colors in its palette, but did not result in comparably high preference levels in our study.

6 AFFECT-AWARE WORD CLOUD PROTOTYPE

In this section, we present our prototype tool that incorporates affects into the word cloud generation process by enabling the user to specify the intended affects such that congruent font and color palette options can be recommended.

6.1 Implementation Details

6.1.1 User Input. The tool takes as input a text, as well as the desired affective preferences. In the above user studies, we used a set of words that are synonyms with the affect word to prevent any confounding impact on the evaluation. However, on real-world data, multiple affects may be evoked within a single text. So, one challenge in font selection is to cope with such cases.

To overcome this, we introduce restrictions on possible combinations of the affects as follows. We group the affects in two groups: core affects (*calm*, *exciting*, *positive*, *negative*) and complex affects (*playful*, *serious*, *trustworthy*, *disturbing*). Using these groups, we only allow the following selections:

(1) A single complex affect.
(2) A single core affect
(3) Two core affects, except for opposite pairs (*positive* and *negative*, *calm* and *exciting*)

6.1.2 Font Selection. In this part, we describe how we handle font selections for the different affect choices listed above. When a complex affect is chosen, we use the fonts that have the highest scores for that complex affect. However, the case for a single core affect selection is not that straightforward. As an example, when only *positive* is selected, this means that the intended affect does not contain *calm* or *exciting*. As seen in Figure 3, a *positive* font could be *calm* or *exciting*, so simply using fonts with high *positive* values may not be an optimal solution. To handle such cases, when only one core affect is selected, we filter fonts so as to only retain those that exhibit neutral behavior in the unspecified affect direction. Figure 11 shows some examples for the affect *positive*, which includes cases where neither *calm* nor *exciting* is selected along with it, or one of them is selected. Figure 12 gives similar examples for *negative*. Analyzing these figures in conjunction with Figure 2 reveals that the fonts that are congruent with multiple affects tend to combine characteristics of fonts for each of the participating affects. We also see that while having diverse visual characteristics in Figure 2, the top fonts for *negative* tend to be more similar when grouped by their scores for the unspecified affect dimension, as shown in Figure 12. The same is also true for *exciting*.

The final scenario is that of two core affects being selected. In such cases, for each font, we ensure that the values for each of the selected affects is above a certain threshold (e.g., 50%). If they are,

Neutral & Positive	Calm & Positive	Exciting & Positive
positive	calm+positive	exciting+positive
POSITIVE	calm+positive	exciting+positive
positive	calm+positive	exciting+positive

Figure 11: Samples of fonts with high scores for *positive* alone, and for its combinations with other core affects *calm* and *exciting*.

Neutral & Negative	Calm & Negative	Exciting & Negative
negative	calm+negative	EXCITING+NEGATIVE
negative	calm+negative	exciting+negative
negative	calm+negative	EXCITING+NEGATIVE

Figure 12: Samples of fonts with high scores for *negative* alone, and for its combinations with other core affects *calm* and *exciting*. The visual characteristics of the fonts are similar in a combination, whereas they are somewhat more diverse between different combinations.

[6] Figure 1 provides samples of word clouds with congruent color palettes from this user study.
[7] We used *exciting* vs. *calm*, *negative* vs. *positive*, *serious* vs. *playful*, and *disturbing* vs. *trustworthy* as opposite pairs. The color palette for *exciting*, e.g., provides incongruent colors for *calm*.

Figure 13: Word clouds visualizing two restaurant reviews. The colorful cloud on the left visualizes a five-star review using a *positive* font and color palette, whereas the two star review on the right uses *negative* ones.

then we use the average of these scores to determine the congruency of the font for this affect combination. Otherwise, we use the minimum of the affect values. We have adopted this approach to prevent unsuccessful matches caused by very high values of certain affects dominating a low value of another.

The actual font that is recommended by the tool is selected randomly among the top candidates using the above logic.

6.1.3 Color Selection. Similar to the font selection methodology described above, in this part we explain how we handle color selections for the different affect choices. If only one affect is chosen, irrespective of whether it is a core affect or a complex one, we simply apply the corresponding color palette. If multiple core affects are selected, we use the corresponding complex affect, which is based on the relationships of emotions. Specifically, we use *trustworthy* for *calm* and *positive*, *serious* for *calm* and *negative*, *disturbing* for *exciting* and *negative*, and *playful* for *exciting* and *positive*. Reviewing Figure 7 from Bartram et al. [3] reveals that the color palettes for complex emotions are indeed very close to the combination of color palettes of their underlying core affects.

6.1.4 Visualization. We preprocess the input to remove punctuation etc. as well as to obtain lemmatized versions of the words. To generate the word cloud, we rely on an external library[8] that is developed using the D3 framework[9]. The library allows for specifying preferred angle options for the words, but in our prototype we render all words horizontally. It also allows for specifying a scale with which the sizes of the words are determined based on their distributions from the input. We enable users to control this parameter so that they are able to cater to inputs with different distribution characteristics. Based on the affect choices described earlier, we generate a word cloud with the automatically recommended font and color palette choices. Users have the option of browsing through further recommendations, or they can simply select any preferred fonts or colors.

6.2 Output Samples

We showcase outputs of our prototype tool using data from several domains, while demonstrating the importance of affective visualizations for these domains.

6.2.1 Visualizing Restaurant Reviews. Word clouds are commonly used to visualize user-written reviews, such as of restaurants. A

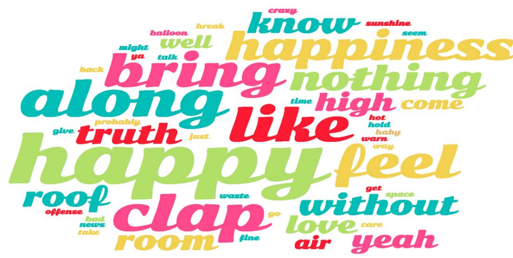

Figure 14: Word cloud visualizing the song *Happy* by *Pharrell Williams*, illustrating the potential of congruent font and color choices.

Google web search using keywords *”restaurant review” and ”word cloud”* yields around eighteen thousand results, including academic papers [16]. Since a central goal of reviews is to convey the sentiment of a customer with respect to products or experiences, the affective perception of the word cloud is significant. Figure 13 portrays two word clouds generated by our tool. Based on the sentiment in the reviews, the input affects are specified as *positive* and *negative*, respectively, resulting in corresponding affective visualizations. These clouds also reflect the results of our font selection method when there is only one specified affect, i.e., the input is neither *calm* nor *exciting*.

6.2.2 Visualizing Songs. Word clouds are frequently used to visualize songs [6]. The web query *song and ”word cloud”* yields 2.1 million search results. There have also been games, such as *guess the song* ones, using word clouds as their medium. Figure 14 provides a visualization of the song *Happy* by *Pharrell Williams*. As the name suggests, it is a *positive* song and visualized accordingly. The word cloud seeks to evoke an uplifting feeling as the original song does.

For song lyrics visualization, paralinguistic signals may be particularly important due to the additional affective signals provided by the music. The words from the lyrics sometimes cannot directly be used to elicit these type of signals. Fonts and color palettes, thus, may be used to evoke an appropriate set of affects. This case also applies to any other form of input carrying additional affective signals beyond those in the text.

6.2.3 Visualizing Movies. Movies are another frequent type of word cloud content. A web search using keywords *movie and ”word cloud”* yields 2.7 million results, which suggests substantial interest in such visualizations. Naturally associated with emotions, movies benefit from congruent paralinguistic signals in their visualizations. Figure 15 shows two movie word clouds created by our tool[10]. The animation/comedy movie on the left uses *playful* options, whereas the horror movie uses *exciting* and *negative* ones. It may even be possible to guess the genres of these movies just in light of the font and color choices.

6.2.4 Visualizing Trustworthy *Content.* Another type of word clouds we wish to specifically analyze here are *trustworthy* ones. Word

[8]https://www.jasondavies.com/wordcloud/
[9]https://d3js.org/
[10]The input texts are movie summaries obtained from IMDB.

Figure 15: Word clouds visualizing two movies. The colorful cloud on the left visualizes the animation movie *Smurfs: The Lost Village (2017)*, using a *playful* font and color palette. The word cloud on the right is for the movie *Scream (1996)*, and, reflecting the movie's genre (horror), it is visualized using the affects *exciting* and *negative*.

Figure 16: Word cloud for the United Nations using paralinguistic signals assessed as *trustworthy*.

clouds are frequently used to describe the values of a corporation. Searches in web sites providing professional design products, such as ShutterStock[11], return hundreds of word clouds specifically designed to reflect core values of corporations. A common goal of these corporations is to be perceived as *trustworthy*, which definitely needs to be incorporated into their visualizations. Figure 16 provides a sample output of our tool using text[12] related to the *United Nations*, visualized using the affective attribute *trustworthy*.

7 DISCUSSION

We discuss our results, their implications for users and use cases, and other potential research directions.

7.1 Affective Strengths of Fonts and Color Palettes

Across all experiments and attributes, congruent font and color palette choices were preferred by a plurality or majority, while incongruent choices were dispreferred by a majority of responses. The findings in the first two user studies have shed light on the relationship between affective responses and fonts. For attributes such as *serious* and *trustworthy*, this relationship is found to be particularly strong. Interpreting these results together with the third user study, we observe that different signals exhibit different strengths in terms of their affective impact. Based on our experiments, color palettes prove particularly powerful for *calm*, *negative*, and *playful*. Thus, a *serious* or *trustworthy* perception appears easier to evoke with fonts, whereas a *calm* or *negative* appearance can arise from appropriate color palettes. For the remaining attributes, to achieve a more pronounced effect, a combination of both fonts and colors may be a compelling option.

7.2 Supporting Creativity

In a survey [43], 88% of users reported feeling creative when using Wordle, due to the use of font and color palette options that can be explored. 81% of users reported they were trying it for fun. Hence, providing a medium in which users can feel creative ought to be an important aim of word cloud tools. We believe that allowing users to try applying different affects, or to try different options for the same affect, would substantially increase the creative potential to be explored by them, especially if they are given the chance to explore font and color palette options that are congruent with the affect(s) they specified. Nonetheless, a user study is needed to verify this hypothesis, since the users' sense of creativity is known to be hard to predict [45].

7.3 Sentiment Analysis

Our output samples show that affective-aware choices of fonts are crucial for data from several domains. The same is more generally true of word clouds for sentiment analysis. There are many online resources, presentations, and academic papers [18, 29] that make use of word clouds for sentiment analysis, showcasing affective words that are used in the text. Most of them use a different word cloud for each emotion, or they use a comparison cloud to compare their intensities. In both cases, paralinguistic signals could

[11] https://www.shutterstock.com
[12] The text is obtained from the Wikipedia page for the United Nations.

prove helpful to facilitate the perception of different emotions in sentiment visualizations.

7.4 Other Paralinguistic Signals and Visualizations

We have explored the affective usage of fonts and color palettes in word clouds. However, there are further affective signals to be explored in word clouds, as well as other text visualization methods to be investigated with respect to their incorporation of paralinguistic signals. For word clouds, one other such signal is the *layout*. We suspect that this could prove powerful, especially for attributes such as *disturbing* and *playful*. Again, this needs to be verified through a user study as well. In our experiments, in contrast, we keep the layout identical between different word cloud options to reduce the potential for confounding effects.

There are also several other visualizations that could benefit from an affect-aware approach. One line of such visualizations, as mentioned earlier, are visualization tools used for sentiment analysis. However, the list is not limited to this task. A tree map visualization, for instance, already conveys affect, since it relies on color palettes, and hence should perhaps be made affect-aware. In fact, any visualization that makes use of fonts or color palettes ought to allow for making such choices more carefully and deliberately, considering the affective nature of these signals.

7.5 Other Semantic Connections

Our study could be considered as a starting point towards exploring the connections between the semantics of the input and word clouds, or more generally, any visualization technique using the considered signals. More specific associations exist between fonts and semantic attributes [31], and these could be drawn upon to create word clouds that are an even a better thematic fit to the input. An example is using fonts found to be *technical* for *technical* content. Other more specific connections also exist between color and words, such as invoking the color *red* for the word *strawberry*, or *blue* for a word cloud relating to the *Smurfs*.

7.6 Regional and Cultural Differences

Previous studies reveal that regional and cultural differences affect color choices [13, 30]. A potential research direction is to advance the word cloud tool to consider such differences based on user demographics. This might entail investigating whether such differences exist for font choices as well, especially with regard to the affective connections. Currently, in our tool, we provide users the opportunity to change the font and color choices without any restrictions, so that arbitrary personal preferences can be accommodated.

8 CONCLUSION

We have invoked a set of techniques to obtain font congruency values for several affective attributes based on a crowdsourced seed set. The results of our studies establish that such semi-automatically acquired font scores accord with human assessments of congruence, similar to previous studies that relied on human-chosen fonts. This has led us to explore the use of such paralinguistic signals in computational applications and tools, which we have considered in particular to increase the affective power of word clouds.

Our findings reveal that both fonts and color palettes are potent signals in creating affective word clouds. Moreover, their respective strengths turn out to be complementary. Hence, we can conclude that fonts could be used as an additional dimension in visualizations to intentionally encode affect, and not only designers but also developers of computational tools need to account for the possibilities afforded by font and color associations with affective attributes.

9 ACKNOWLEDGEMENTS

We thank Alex Borgida, Nil Tugce Kulahcioglu, and Min-Jeong Yang for helpful comments and discussions. Gerard de Melo's research is funded in part by ARO grant no. W911NF-17-C-0098 (DARPA SocialSim program).

REFERENCES

[1] Eric Alexander, Chih-Ching Chang, Mariana Shimabukuro, Steven Franconeri, Christopher Collins, and Michael Gleicher. 2018. Perceptual biases in font size as a data encoding. *IEEE transactions on visualization and computer graphics* 24, 8 (2018), 2397–2410.

[2] Christina Alm-Arvius. 1998. *Introduction to semantics.* Studentlitteratur, Lund.

[3] Lyn Bartram, Abhisekh Patra, and Maureen Stone. 2017. Affective Color in Visualization. In *Proceedings of CHI 2017.* ACM, 1364–1374.

[4] Faber Birren. 2016. *Color psychology and color therapy; a factual study of the influence of color on human life.* Pickle Partners Publishing.

[5] Richard Brath and Ebad Banissi. 2016. Typographic Sets: Labelled Set Elements with Font Attributes. In *Proceedings of the SetVR 2016 Workshop at Diagrams.* 29–43.

[6] Michael Burch, Tobias Fluck, Julian Freund, Thomas Walzer, Uwe Kloos, and Daniel Weiskopf. 2016. Lyrics Word Clouds. In *Information Visualisation (IV), 2016 20th International Conference.* IEEE, 51–56.

[7] Michael Burch, Steffen Lohmann, Fabian Beck, Nils Rodriguez, Lorenzo Di Silvestro, and Daniel Weiskopf. 2014. RadCloud: Visualizing multiple texts with merged word clouds. In *Information Visualisation (IV), 2014 18th International Conference on.* IEEE, 108–113.

[8] Terry L Childers and Jeffrey Jass. 2002. All dressed up with something to say: Effects of typeface semantic associations on brand perceptions and consumer memory. *Journal of Consumer Psychology* 12, 2 (2002), 93–106.

[9] Weiwei Cui, Yingcai Wu, Shixia Liu, Furu Wei, Michelle X Zhou, and Huamin Qu. 2010. Context preserving dynamic word cloud visualization. In *2010 IEEE Pacific Visualization Symposium (PacificVis).* IEEE, 121–128.

[10] Cristian Felix, Steven Franconeri, and Enrico Bertini. 2018. Taking word clouds apart: An empirical investigation of the design space for keyword summaries. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 657–666.

[11] Alysha Fligner. 2013. *The effect of packaging typeface on product perception and evaluation.* Ph.D. Dissertation. The Ohio State University.

[12] Richard L Hazlett, Kevin Larson, A Dawn Shaikh, and Barbara S Chaparo. 2013. Two studies on how a typeface congruent with content can enhance onscreen communication. *Information Design Journal* 20, 3 (2013).

[13] Johannes Itten. 1970. *The elements of color.* John Wiley & Sons.

[14] Jaemin Jo, Bongshin Lee, and Jinwook Seo. 2015. WordlePlus: expanding wordle's use through natural interaction and animation. *IEEE computer graphics and applications* 35, 6 (2015), 20–28.

[15] Samuel Juni and Julie S Gross. 2008. Emotional and persuasive perception of fonts. *Perceptual and motor skills* 106, 1 (2008), 35–42.

[16] David Kamerer. 2014. Understanding the Yelp review filter: An exploratory study. *First Monday* 19, 9 (2014).

[17] Kyle Koh, Bongshin Lee, Bohyoung Kim, and Jinwook Seo. 2010. Maniwordle: Providing flexible control over wordle. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1190–1197.

[18] Kostiantyn Kucher and Andreas Kerren. 2014. Text visualization browser: A visual survey of text visualization techniques. *Poster Abstracts of IEEE VIS* 2014 (2014).

[19] Tugba Kulahcioglu and Gerard De Melo. 2018. FontLex: A Typographical Lexicon based on Affective Associations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).*

[20] Tugba Kulahcioglu and Gerard de Melo. 2018. Predicting Semantic Signatures of Fonts. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC).* IEEE, 115–122.

[21] Lauren I Labrecque and George R Milne. 2012. Exciting red and competent blue: the importance of color in marketing. *Journal of the Academy of Marketing Science* 40, 5 (2012), 711–727.

[22] Clive Lewis and Peter Walker. 1989. Typographic influences on reading. *British Journal of Psychology* 80, 2 (1989), 241–257.

[23] Dongying Li and Xiaolu Zhou. 2017. "Leave Your Footprints in My Words"—-A Georeferenced Word-Cloud Approach. *Environment and Planning A* 49, 3 (2017), 489–492.

[24] Steffen Lohmann, Florian Heimerl, Fabian Bopp, Michael Burch, and Thomas Ertl. 2015. Concentri cloud: Word cloud visualization for multiple text documents. In *Information Visualisation (iV), 2015 19th International Conference on*. IEEE, 114–120.

[25] Andrea Webb Luangrath, Joann Peck, and Victor A Barger. 2017. Textual paralanguage and its implications for marketing communications. *Journal of Consumer Psychology* 27, 1 (2017), 98–107.

[26] Colin M MacLeod. 1991. Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin* 109, 2 (1991), 163.

[27] Albert Mehrabian and James A Russell. 1974. *An approach to environmental psychology.* the MIT Press.

[28] Ravi Mehta and Rui Juliet Zhu. 2009. Blue or red? Exploring the effect of color on cognitive task performances. *Science* 323, 5918 (2009), 1226–1229.

[29] Saif M Mohammad. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems* 53, 4 (2012), 730–741.

[30] Jantawan Noiwan and Anthony F Norcio. 2006. Cultural differences on attention and perceived usability: Investigating color combinations of animated graphics. *International journal of Human-computer studies* 64, 2 (2006), 103–122.

[31] Peter O'Donovan, Janis Libeks, Aseem Agarwala, and Aaron Hertzmann. 2014. Exploratory font selection using crowdsourced attributes. *ACM Trans. Graph.* 33, 4 (2014), 92:1–92:9.

[32] Klaus R Scherer. 2003. Vocal communication of emotion: A review of research paradigms. *Speech communication* 40, 1-2 (2003), 227–256.

[33] Klaus R Scherer, Harvey London, and Jared J Wolf. 1973. The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality* 7, 1 (1973), 31–44.

[34] Johann Schrammel, Michael Leitner, and Manfred Tscheligi. 2009. Semantically structured tag clouds: an empirical evaluation of clustered presentation approaches. In *Proceedings of CHI 2009*. ACM, 2037–2040.

[35] A Dawn Shaikh. 2007. The Effect of Website Typeface Appropriateness on the Perception of a Company's Ethos. *Usability News* 9, 2 (2007).

[36] Audrey Dawn Shaikh. 2007. *Psychology of onscreen type: Investigations regarding typeface personality, appropriateness, and impact on document perception.* Ph.D. Dissertation.

[37] A Dawn Shaikh, Barbara S Chaparro, and Doug Fox. 2006. Perception of fonts: Perceived personality traits and uses. *Usability News* 8, 1 (2006), 1–6.

[38] A Dawn Shaikh, Doug Fox, and Barbara S Chaparro. 2007. The effect of typeface on the perception of email. *Usability News* 9, 1 (2007), 1–7.

[39] Satyendra Singh. 2006. Impact of color on marketing. *Management decision* 44, 6 (2006), 783–789.

[40] Julie Steele and Noah Iliinsky. 2010. *Beautiful visualization: looking at data through the eyes of experts.* " O'Reilly Media, Inc.".

[41] Carlos Velasco, Alejandro Salgado-Montejo, Fernando Marmolejo-Ramos, and Charles Spence. 2014. Predictive packaging design: Tasting shapes, typefaces, names, and sounds. *Food Quality and Preference* 34 (2014), 88 – 95.

[42] Carlos Velasco, Andy T. Woods, Sarah Hyndman, and Charles Spence. 2015. The Taste of Typeface. *i-Perception* 6, 4 (2015).

[43] Fernanda B Viegas, Martin Wattenberg, and Jonathan Feinberg. 2009. Participatory visualization with wordle. *IEEE transactions on visualization and computer graphics* 15, 6 (2009).

[44] Ji Wang, Kyle D Dent, and Chris L North. 2013. Fisheye word cloud for temporal sentiment exploration. In *Proceedings of CHI 2013 Extended Abstracts*. ACM, 1767–1772.

[45] Yunhai Wang, Xiaowei Chu, Chen Bao, Lifeng Zhu, Oliver Deussen, Baoquan Chen, and Michael Sedlmair. 2018. EdWordle: Consistency-preserving Word Cloud Editing. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 647–656.

[46] Alan J Wecker, Joel Lanir, Osnat Mokryn, Einat Minkov, and Tsvi Kuflik. 2014. Semantize: visualizing the sentiment of individual document. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. ACM, 385–386.

[47] TW Whitfield and TJ Whiltshire. 1990. Color psychology: a critical review. *Genetic, social, and general psychology monographs* (1990).

[48] Jin Xu, Yubo Tao, and Hai Lin. 2016. Semantic word cloud generation based on word embeddings. In *Pacific Visualization Symposium (PacificVis), 2016 IEEE*. IEEE, 239–243.