

Supporting Content Design with an Eye Tracker: The Case of Weather-based Recommendations

Alejandro Catala, Jose M. Alonso, and Alberto Bugarin

Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS),
Universidade de Santiago de Compostela, Spain

{alejandros.catala, josemaria.alonso.moral, alberto.bugarin.diz}@usc.es

Abstract

Designing content output for weather-aware services based on domain experts can sometimes be arduous due to their limited availability and the amount and complexity of information considered in explaining their recommendations. As an initial step in our work towards generating recommendations that are acceptable and readable, our methodology involving an eye tracker attempts to simplify and capture more valuable data in early design stages. Our pilot study explored which information in weather-based recommendations seemed to be more useful to support users decision making. The results suggest that interactive content could be deployed based on the relevance of informational items and both graphical points of interest and legends could help in delivering content more efficiently.

1 Introduction

In the realm of context-aware services and interactive applications, Natural Language Generation (NLG) involving maps in combination with meteorological data is subject to active field research (Ramos-Soto et al., 2015). Automatically generating recommendations consisting of both text and figures can help users in making decisions while providing personalized services (Gkatzia et al., 2017). Furthermore, it is not just an issue of giving a suitable recommendation according to the user's context (Mocholi et al., 2012), but also to design content generators in such a way that the artificial intelligence associated to the service is better considered in terms of being explainable, accountable and intelligible (Abdul et al., 2018; Alonso et al., 2018).

The combination of such qualities means that we are facing a complex design problem that needs to deal with several issues before a successful algorithm can be implemented. In order to start addressing this issue, we propose to use an eye tracker with a double purpose: i) to set a priority and get a narrower focus on all the information elicited from meteorologists; and ii) to supply a method in order to gather from users more objective data that complement self-reporting questionnaires. In this way, we expect to enable better informed design decisions. Thus, this paper contributes a pilot empirical study exploring with the help of an eye tracker how stimuli containing recommendations with explanations supported by figures are processed by people and which elements can be more relevant for generating content in future designs.

2 Background on cognitive psychology and eye tracking

Eye trackers are devices capable of recording gaze or eye-movement data as users focus their visual attention. They have typically supported research concerned about reading patterns and content engagement (Liu, 2014), since visual attention triggers underlying cognitive processes. There are also studies regarding how the required tasks can influence people's eye movements (Kaakinen and Hyona, 2010). In addition, it is well-known from cognitive psychology that multimodal texts and underlying structures can enhance interaction with contents and their processing effort (Danielsson and Selander, 2016). For example, the study in (Holsanova et al., 2008) used an eye tracker to confirm that design principles such as spatial contiguity and attentional guidance that support both spatial navigation and semantic integration of concepts facilitate information processing in newspa-

per reading.

In order to analyze gaze data, there are several features and a range of metrics that eye-tracking tools can provide, as surveyed in (Sharafi et al., 2015). Among the raw data, eye fixations are especially useful, which refer to stabilization of the eye for a period of time (e.g. circa 200ms) and provide deeper understanding on where visual attention has been focused. Scanpaths are also interesting, which visualize chains of fixations. To the best of our knowledge, there is not much specific work using eye-tracking to explore weather-based stimuli besides the recent study by (Sivle and Uppstad, 2018). The authors explored how multimodal reading takes place and why readers move between representations, concluding that tables are more often used with respect to diagrams.

3 Study

3.1 Participants and equipment

Fifteen adult volunteers (mean: 23.13 years old, $sd=2.71$) participated in the empirical study. They were all postgraduate students or PhD candidates working on technical fields related to computer science. All except one stated to have prior knowledge of Galician geography.

The experimental setting was implemented using the EyeTribe Tracker¹ to track the eye gaze on a main screen where the stimuli were displayed (see Figure 1). Also, a device supported the subject's chin to prevent tracker calibration issues. A secondary bigger screen only active while answering questionnaires was set behind, placed at a distance so that both can be read without changing the pose, to not compromise the tracker calibration.

We used the Ogama software (Voßkühler et al., 2008) (version 5.0.5754) to assemble the stimuli and manage the gaze data recording.

3.2 Stimuli design

The empirical study consisted of 6 trials, which were randomized to prevent order effects. The stimulus in each trial included a recommendation about the suitability to carry out activities on the Beach, Surfing or activities on the Mountains. Typically, a stimulus included a textual description on the upper side of the screen. The text was in Spanish, the native language of the participants. A literal transcription of a sample text on the Beach topic into English is as follows:

¹<https://github.com/EyeTribe/documentation>

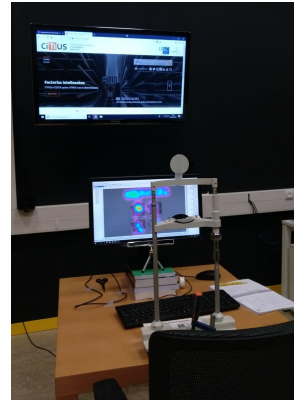


Figure 1: Experimental setting.

“Today will be a perfect day to enjoy the beaches of A Mariña luguesa, like in As Catedrais or Arealonga, since the temperatures will be very pleasant and the skies will remain clear all day. Likewise, it is also recommended to attend the fluvial beaches of the interior of Galicia.

The reason for the good weather prevailing on the Cantabrian coast and inland of Galicia is due to the move of the Anticyclone from the Azores to the east. Such a synoptic situation will cause both territories to be left out of the mists and low cloud cover that will do affect the Galician Atlantic coast.”

The text consisted of both a recommendation R (the first paragraph) and an explanation of the weather forecast E (the second paragraph). The order of these parts can lead to two possible arrangements (i.e., $\langle R, E \rangle$ or $\langle E, R \rangle$). The stimuli also came with a set of maps supporting the explanation (see Figure 2-Left): weather forecast, UV index, max temperature, and sea state maps. The fourth map was replaced by a storm warning map in the mountain recommendations. The stimuli were designed by a meteorologist with experience in generating weather reports, taking into account that the target user is the general public.

3.3 Procedure

Before the experimental session, each participant was provided with an informed consent form explaining the research context and empirical tasks, and agreed to participate voluntarily. Then, the experimenter proceeded to make the adjustments to the chair as needed, and a calibration procedure was carried out in order to initialize the eye tracker and ensure gaze data recording. Some informational screens were displayed regarding ge-

ographical information to get acquainted with the type of maps and related locations. Then the user performed the trials following instructions on the main screen, just switching to the secondary display when requested to answer questionnaires. For each trial, there was a first screen presenting the textual description as a stimulus. The task was to read the text, in order to gather typical reading patterns, warm up, and be sure that tracking worked correctly. A second screen presented the same text plus the figures supporting the textual description. The task for the user was to inspect the recommendation to assess to which extent the visual information provided matched the textual description. The stimuli were self-paced, and participants kept their hand on the space bar all the time, which had to be pushed to move forward. Switching between displays was handled by the experimenter, turning them off and on as needed. Instruction screens were set between stimuli in order to ensure that gaze recording was separated accordingly. Once the 6 trials were finished, the participant answered the demographics questionnaire.

3.4 Results

3.4.1 Gaze data

We carried out a qualitative analysis by replaying the fixations, scanpaths and calculating the attention maps as fixation count heatmaps with fixations weighted by duration as provided by the Ogama software. While fixations just give point clouds where users looked at on screen, the attention maps can be used to identify regions of special attention in specific stimuli, filtering noise and enhancing visual analysis. Longer fixations, and therefore attention, have several implications. In reading tasks, longer fixations are typically over words that took longer processing time (Rauzy and Blache, 2012; Sharafi et al., 2015), either because the word was more difficult to understand or just because it was considered a relevant and important term to remember. In matching tasks, fixations and attention maps provide insight into which spots can be more informational and relevant to support the textual description. This allows us to decide which information should be kept as it is, highlighted or discarded.

In the reading tasks, we captured the reading patterns, which led to scanpaths line per line from left to right. It took on average about 29 seconds (sd=7.99), resulting in 91.28 (sd=22.97) fixations

and 3.18 fixations per second on average. The attention maps show that more processing effort focused on the general forecast description *E* rather than in the recommendation *R* itself regardless of the arrangement. We must also be aware that explanations were usually longer than the proper recommendation. When analyzing the words lying in the spots, the most prominent ones are related to weather events (e.g., showers, wind, or very significant waves) and geographical locations (e.g., Patos beach, A Madalena beach, or province of Pontevedra). Also, we noticed that some single words (e.g., *synoptic*) were signaled, which are uncommon terms often used by meteorologists.

In the matching tasks, each trial took 24.42 seconds on average (sd=9.71), with a mean number of fixations about 70.28 (sd=26.67) and 2.9 fixations per second. Regarding the gaze data, weather events and geographical locations are again prominent in the text (e.g., light showers, high temperature, inland region, or beach of Carnota). Regarding the figures, the most prominent spots are over the weather forecast map (e.g., related to specific areas mentioned in the text such as *A Mariña luguesa* in Figure 2-Center), the max temperature map and the maps' legends. When the gaze focused longer on weather graphic symbols, they were about weather events such as showers rather than good weather conditions. The sea and the storm warning maps had some relevant role in the surfing and mountain trials respectively as depicted in Figure 2-Right. Overall, the dwell times on the defined Areas of Interests (AOIs) confirm the relevance of maps for users (see Figure 3).

3.4.2 Questionnaires

We gathered additional information through questionnaires provided after each trial. We used a 7-point Likert scale for assessing questions regarding *Coherence text-graphics* (m=6.26, sd=1.13), *Readability* (m=6.23, sd=1.02), and *Understandability* (m=6.37, sd=0.98). Some open questions to gather the most and less relevant items according to participants were included. Table 1 reports the frequencies of topics in the content analysis. Overall, these self-reported remarks were consistent with the observation from the gaze data.

4 Discussion and future work

We have explored eye-tracking as a complementary method to the self-reporting questionnaires that are typically used in similar research.

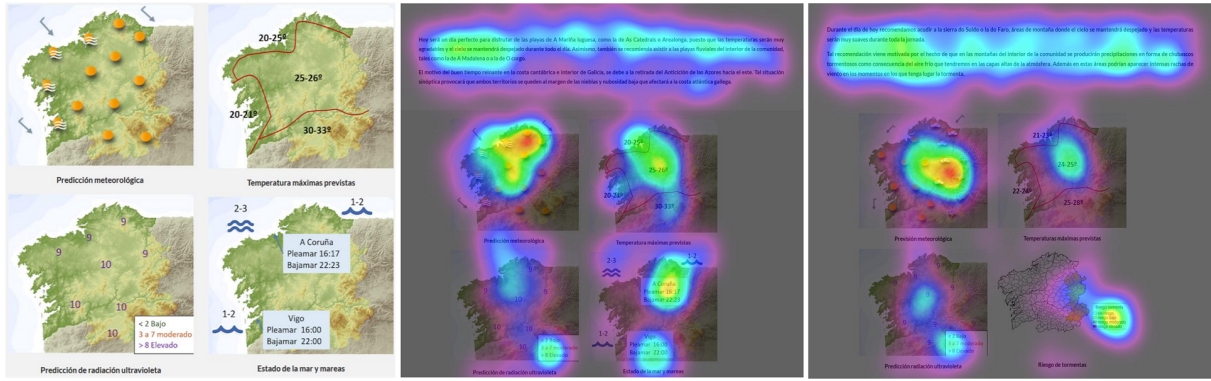


Figure 2: Sample maps (Left). Attention maps for a beach (Center) and a mountain stimulus (Right), calculated as fixation count height maps with fixations weighted by duration and using the following colour scale normalization: purple (10%), blue (25%), turquoise (40%), green (65%), yellow (75%), orange (93%), red (100%). The stimuli size was 1920x1080, the kernel size was Ogama's default 201.

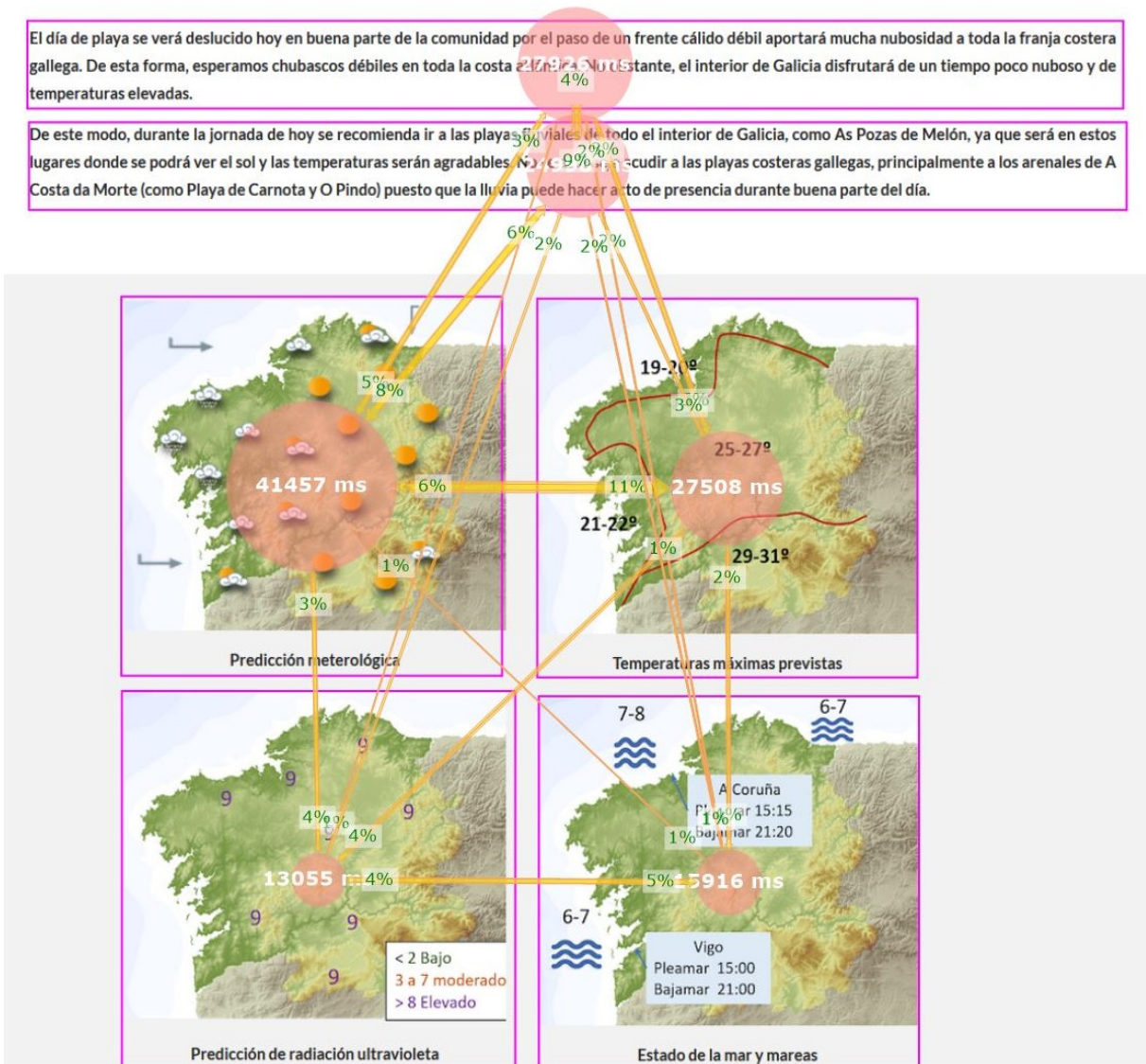


Figure 3: Complete fixation time on Areas of Interests (AOIs) and relative transition paths for a Beach stimulus. It provides a magnitude for the overall time spent in each AOI.

Table 1: Content analysis: the most ‘+’ and less ‘-’ relevant items (number of occurrences in brackets).

	Beach	Surfing	Mountain
+	weather forecast map (7), max temperature map (4), UV index map (2), sea state map (2)	sea state map (7)	weather forecast map (8), storm warning map (4), max temperature map (3)
-	complex descriptions including either technical terms or place names (5), UV index (3)	max temperature map (2), UV index map (2), place names (3)	place names and technical terms, storm warning map if there is no risk

Involving domain experts to provide well-founded descriptions and explanations is a challenge. They provide much information to be fully precise, and therefore prioritizing or simplifying the pieces of information is not straightforward.

Following a traditional approach would require several design cycles to elicit information from meteorologists, who are not always available, whereas testing with users is costly even for small samples. Thus, our approach attempts to speed up the process at an early stage of development by starting with a more exploratory scenario that allowed us to get multiple observations at once in order to back up future design decisions. This motivated that our request to the meteorologist for designing the stimuli included some practical constraints such as text no longer than a short paragraph and no more than four maps fitting a single screen for a web service application. In this way, the expert still had some room to create a report and we are not discarding informational items beforehand without a good reason. Moreover, having a setting with a PC desktop screen was deliberately chosen because we can focus on the content without any interference imposed by interactions (e.g., navigating between smaller screens in a mobile user interface), and the design space is better understood by both the domain expert and users.

The results confirmed that the domain expert who designed the stimuli used more source information than users demand and can naturally process, as suggested by underused maps and user comments regarding complex descriptions and technical terms. Thus, one design principle is to provide *simplified on-screen information*. Choosing a limited set of information sources would help to reduce complexity and cognitive load. For example, by providing only the two most relevant maps as reported in the results and by giving the option to interactively explore more complex and extended information. Salient gaze spots for

text were on referring expressions, such as proper nouns of places, and weather events. This is an expected result as these words are actually the key information being conveyed, in line with (Rauzy and Blache, 2012).

When talking about specific places (e.g., the name of a beach or a peak), the maps should also include landmarks to facilitate its interpretation, mitigate any gap in the user’s geographical background knowledge and simplify text. Furthermore, when text is the only possible output, because maps are not available or another modality is being used (e.g., speech), the *specific place should be accompanied of a more general location*. For example, a recommendation referring to the beach called “Patos” could be improved by expanding the information in the text with a more general location well-known by users such as *Ría de Vigo*. We can also focus on the recommendation *R*, and then consider the general *explanation in a follow-up interaction*. This is still important to provide more intelligibly context-aware applications (Lim and Dey, 2013). Expanded explanations under request could include a more technical view indeed. We must *use the legends properly* as can be a very powerful resource, with users looking at them systematically. Although using heatmaps can be a very useful tool, they must be handled with care to prevent misinterpretations (Bojko, 2009). Our study used fixation count height maps with a correction to take into account the length of fixations. However, we must be aware that they just represent average fixation behaviour, and as any averaged computation it can be subject to bias due to very different fixation behaviours or longer exposures to the stimuli. Accordingly, more advance and robust computations to complement and counteract such limitations should be considered whenever possible.

We can conclude that an eye-tracker provides additional objective and valuable data which are

complementary, but quite in agreement to those derived from questionnaires. As future work, we aim to develop a data-to-text module ready to automatically produce multimodal recommendations. Content design will be initially guided by the conclusions derived from this study. Furthermore, we will analyze how other different structures (that can be explored interactively) may affect the explainability and intelligibility of a weather-aware service.

Acknowledgements

This work has been partially funded by the Spanish Ministry of Economy and Competitiveness under the project grants BAI4SOW (TIN2014-56633-C3-1-R), BIGBISC (TIN2017-84796-C2-1-R), and the Ramón y Cajal Research grant RYC-2016-19802, by the European Regional Development Fund (ERDF/FEDER program) and Consellería de Cultura, Educación e Ordenación Universitaria (acc.2016-2019, ED431G/08), and by the CiTIUS Summer scholarship program. Special thanks go to Damian Caballero for his invaluable help and expertise on weather recommendations and forecasting. Thanks to all volunteers who participated in the study.

References

- Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. [Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda](#). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 582:1–582:18, New York, NY, USA. ACM.
- Jose M. Alonso, C. Castiello, and C. Mencar. 2018. [A bibliometric analysis of the explainable artificial intelligence research field](#). In *17th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 3–15. Springer.
- Agnieszka (Aga) Bojko. 2009. [Informative or misleading? heatmaps deconstructed](#). In *Human-Computer Interaction. New Trends*, pages 30–39, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Kristina Danielsson and Staffan Selander. 2016. [Reading multimodal texts for learning : a model for cultivating multimodal literacy](#). *Designs for Learning*, 8(1):25–36.
- Dimitra Gkatzia, Oliver Lemon, and Verena Rieser. 2017. [Data-to-text generation improves decision-making under uncertainty](#). *IEEE Computational Intelligence Magazine*, 12(3):10–17.
- Jana Holsanova, Nils Holmberg, and Kenneth Holmqvist. 2008. [Reading information graphics: The role of spatial contiguity and dual attentional guidance](#). *Applied Cognitive Psychology*, 23(9):1215–1226.
- Johanna K. Kaakinen and Jukka Hyona. 2010. [Task effects on eye movements during reading](#). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6):1561–1566.
- Brian Y. Lim and Anind K. Dey. 2013. [Evaluating intelligibility usage and usefulness in a context-aware application](#). In *Human-Computer Interaction. Towards Intelligent and Implicit Interaction*, pages 92–101, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Pei-Lin Liu. 2014. [Using eye tracking to understand learners' reading process through the concept-mapping learning strategy](#). *Computers & Education*, 78:237 – 249.
- Jose A. Mocholi, Javier Jaen, Kamil Krynicki, Alejandro Catala, Artzai Picón, and Alejandro Cadenas. 2012. [Learning semantically-annotated routes for context-aware recommendations on map navigation systems](#). *Applied Soft Computing*, 12(9):3088 – 3098.
- Alejandro Ramos-Soto, Alberto Jose Bugarín, Senen Barro, and Juan Taboada. 2015. [Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data](#). *IEEE Transactions on Fuzzy Systems*, 23(1):44–57.
- Stéphane Rauzy and Philippe Blache. 2012. [Robustness and processing difficulty models. A pilot study for eye-tracking data on the French Treebank](#). In *Workshop on Eye-tracking and Natural Language Processing at The 24th International Conference on Computational Linguistics (COLING)*, Proceedings of the First Workshop on Eye-tracking and Natural Language Processing, pages 1–15, Mumbai, India.
- Zohreh Sharafi, Timothy Shaffer, Bonita Sharif, and Yann-Gaël Guéhéneuc. 2015. [Eye-tracking metrics in software engineering](#). In *2015 Asia-Pacific Software Engineering Conference (APSEC)*, pages 96–103.
- Anders D Sivle and Per H Uppstad. 2018. [Reasons for relating representations when reading digital multimodal science information](#). *Visual Communication*, 17(3):313–336.
- Adrian Voßkühler, Volkhard Nordmeier, Lars Kuchinke, and Arthur M. Jacobs. 2008. [Ogama \(open gaze and mouse analyzer\): Open-source software designed to analyze eye and mouse movements in slideshow study designs](#). *Behavior Research Methods*, 40(4):1150–1162.