Designing context aware instructions

Baltaretu, A.A.; Krahmer, E.J.; Maes, Alfons

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2014

Designing context aware instructions: Perceptual salience and task demands in the selection of natural landmarks

Adriana Baltaretu,
Alfons Maes,
Emiel Krahmer,
Tilburg center for Cognition and Communication (TICC), Tilburg University, The Netherlands
{a.a.baltaretu, maes, e.j.krahmer}@tilburguniversity.edu

Abstract. The current study examines a small but important ingredient of adequate route directions (RDs): the selection of landmarks ("go left at the red building"). The starting point is a collection of RDs elicited in visual environments systematically differing in two characteristics which are easy to be detected by navigation software (path complexity and visual clutter). In this paper we focus on the set of landmarks produced by respondents in this experiment and analyze to what extent they can be predicted on the basis of their perceptual salience (as a function of size and pixel salience). We conclude that perceptual salience influences landmark choice only to a limited extent, other factors related to the specific navigation task play an important role, such as the location and stability of landmark objects.

Keywords: landmarks; perceptual salience; task knowledge; route directions; instructions.

Route directions (RDs) instruct travellers how to get, turn-by-turn, from one place to another. RDs consist of instructions (how to navigate) and landmarks (descriptive information about the environment). Landmarks are core components of RDs (Allen, 2000), they function as points of reference and help identifying directions, locations or anchor actions in space (e.g., "go left at the red building"). Despite the fact that the role of landmarks in RDs is well documented, we know little of how different physical and cognitive properties of objects influence the selection of natural landmarks. Currently, many commercial systems that generate instructions are not context aware and rarely make reference to landmarks. Knowledge about landmark selection is a prerequisite for the design of humanlike navigational aids and a crucial aspect especially in view of technological developments (like Google Glass) which can feed navigation systems with detailed online visual information about the environment ("Go right just after the parked red car").

The most important characteristic of a landmark is its distinctiveness in the environment. Based on a widely used taxonomy (Sorrows & Hirtle, 1999), one can distinguish different types of distinctiveness: perceptual salience (How distinctive is an object in terms of size, colour, shape etc.), functional or structural relevance (How relevant is the object to the navigation task?) and familiarity (How familiar or important is the object?). In this study, we investigate the influence of perceptual salience and functional relevance on the selection of landmarks in RDs elicited in unfamiliar environments (thus, excluding the role of familiarity). We asked respondents to produce RDs based on visual input (Google StreetView scenes) systematically differing in path complexity and visual clutter. The results show that the higher the level of clutter and the more complex an intersection is, the longer the RDs and the more landmarks people include in the RDs. For the purpose of this study, we re-analyse all landmarks and question to what extent participants are sensitive to perceptual saliency of landmark objects (in terms of object size and pixel salience) and to their relevance to the navigation task.

Method

Materials

A pool of 200 scenes varying in clutter was created by taking snapshots with a pedestrian perspective of intersections in Google StreetView (Google Maps). Out of these, a final set of 36 items differing in terms of intersection complexity (simple intersections: T and + -shaped; complex intersections: Y and

K -shaped) and visual clutter (low levels of clutter: 1.93 – 2.73 feature congestion; high levels of clutter: 3.70 – 5.20 feature congestion, see Rosenholtz, Li, & Nakano, 2007) were chosen. In these stimuli, we counterbalanced intersection complexity and visual clutter. In each scene, subjects view yellow lines marking the route and the direction to be followed and are asked to describe an action into a given direction: “go left”, “go straight” and “go right”.

Participants and Procedure

The collection of landmarks (N = 268 references to 98 different landmarks in 1764 RDs) was produced by 43 English native speakers (20 males, mean age 41 years). They were presented with a scenario specifying that we are developing software that can generate real time/live pedestrian route descriptions by using live video input from Google Glasses. The respondents' task was to provide adequate instructions, making use of all the visual features provided by the visual scenes. Participants saw one picture at a time and filled in the RDs in the input field provided under the picture.

Data coding

On each picture, we labelled all objects chosen as landmarks with bounding boxes (see Figure 1) and counted how often these objects were chosen as landmarks. For each bounding box we estimated the perceptual salience by measuring: (i) pixel salience using Erdem & Erdem's (2013) saliency model, that estimates the saliency of each region of the scene relative to the rest of the picture and outputs a salience map where each pixel has a value between 0 (not salient) and 1 (maximum saliency); (ii) area: we calculated the size proportion of the bounding box compared to the whole picture. Both estimations measure low-level visual features, not informed by higher order navigation task features.

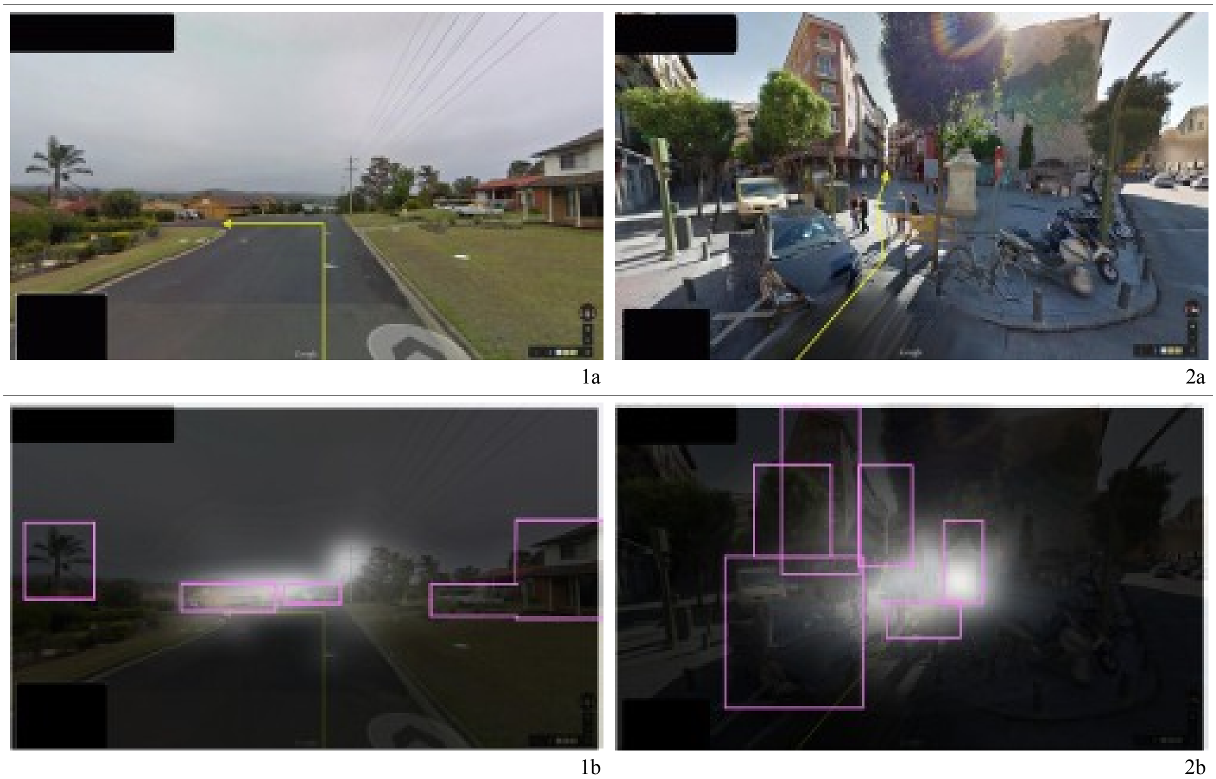

Figure 1. Examples of stimuli: low clutter / simple intersection (1a); high clutter / complex intersection (2a); salience maps below. The white areas are most pixel salient areas of the scene; landmarks marked with boxes. Salience maps might be unclear in black and white print.

Results and discussion

First, we analysed the effect of perceptual salience on the choice of landmarks. The frequency with which objects are chosen as landmarks seems to be influenced by Pixel Salience ($\beta = .226$, $p < .05$) and Object Size ($\beta = .223$, $p < .05$). However, these factors are limited in explaining variance in the choice participants made, ($R2 = .062$, $F(1, 92) = 3.00$, $p < .05$). The results might be due to the small sample of references. We also checked whether the chosen landmarks fall within the most salient scene regions defined by the salience maps. Out of the 98 landmarks, 46.9 % are located in the most salient regions. Despite the fact that these objects have the highest pixel saliency and as a result they are good candidates to be frequently mentioned, they cumulate only 147 references (54.9 % references). Thus, a system only informed by perceptual salience would only partially predict the use of landmarks by humans. Other, task related factors also play a role in selecting relevant landmarks.

One of these task related factors is the location of objects with respect to the route. We know from navigation studies that humans more often include landmarks at decision points (e.g. intersections) than along the route. In our data, most of the landmarks (N = 210 references to 88 different landmarks) indeed are objects placed nearby the intersections where a decision had to be made. In simple intersections there are more landmarks placed just before the intersection (64%). In complex intersections, the landmarks' position not only marks the point where a reorientation action is required, but also disambiguates the correct direction. Most of the landmarks are placed just before the street that needs to be taken (48%), while 30.8% mark the correct direction ("go left *towards* the white house") and are placed imediatelly after the intersection, on the street on which the pedestrian should continue walking.

Lastly, in order for objects to provide reliable navigational information, it is generally believed that they need to be stable in the environment. Given that we presented subjects with static scenes, stable objects (such as buildings) are frequently chosen: 77.6% of the landmarks are stable objects (including the trees mentioned and the shop names which have a relative temporal stability). Out of the stable landmarks, almost half fall outside the most salient regions of the saliency maps (48.7%). Approximately a quarter (23%) of the chosen objects are unstable landmarks (cars and pedestrians). It might be the case that the unstable objects are named because they fall within the most salient scene regions; but, only 31.8% of these unstable landmarks are pixel - salient and, in general, their area is quite small compared to other landmarks such as buildings. This suggest that in situ route directions present particularities that are not currently captured by route direction production experiments.

In sum, we analysed a corpus collected in a larger study and focused on factors that influence language production in visual scenes. Our data suggests that bottom-up perceptual salience is not the only factor that influence landmark selection. Contextual information relevant to the task given (such as the location and the stability of the landmarks) plays an important role, guiding the selection of objects in visual scenes.

References

Allen, G. L. (2000). Principles and practices for communicating route knowledge. *Applied Cognitive Psychology*, 14(4), 333-359.

Erdem, E., & Erdem, A. (2013). Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, 13(4), 11.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), 17.

Sorrows, M. E., & Hirtle, S. C. (1999). The nature of landmarks for real and electronic spaces. In *Spatial information theory*. (pp. 37-50). Springer Berlin Heidelberg.