Following route directions: The role of landmark reference, intersection type and visual clutter

Keywords: route directions, visual clutter, intersection type, landmarks

1 Introduction

Technological advances in augmented reality (e.g., Google Glasses) enable context-aware pedestrian navigation systems to generate new types of content, such as in situ instructions that make use of both variable (“the moving car”) and stable (“the tall building”) information. However, we know little about how visual surroundings in pedestrian navigation instructions influence the turn-by-turn comprehension of such instructions. We address this issue by analyzing the effects of environment complexity and instruction type on the comprehension of route directions (RDs). Visual complexity is operationalized in terms of intersection structure and the richness of detail in the scene; and the type of instructions is varied (with / without landmarks).

RDs have two communicative goals: instructing the user on how to go from one location to another (via instructions for actions) and describing the environment (via referring expressions). Hence, RDs usually include an action prescription coupled with direction (“go left”), and can be enhanced with references to the visual aspects of the environment: path information (“first street”) and landmarks (“at the pharmacy”). This descriptive information has the role of achieving referential determinacy [1]. In this study, we analyze how much time participants require to determine what the correct next step in a route is based on a RD.

We hypothesize that interpreting RDs varies depending on the (geometrical) structure of the intersection [3]. In simple intersections the level of uncertainty is low, thus we expect instructions consisting of a minimum amount of information (e.g., reference to action, direction and path) to be sufficient for making a fast and a correct street choice. As the complexity of an intersection increases (e.g., turn right at a K-shaped intersection), people might need more time and make more mistakes in determining the correct turn.

In addition, we hypothesize that the amount of time needed for choosing the correct street is influenced by low level visual factors such as the amount of detail (visual clutter) in a scene. First, objects (street and landmarks) in cluttered scenes are expected to take more time to find. Second, speakers produce longer descriptions for targets in cluttered scenes, adding more references to salient objects [2], [5], that might facilitate the addressee’s task [4]. Thus, we expect participants to make less errors when receiving instructions with landmarks.
2 Method

2.1 Participants

78 Dutch-speaking students of Tilburg University (26 men, mean age 22.1 years) participated in exchange for partial course credits.

2.2 Materials

64 Dutch RDs were created as follows: a first set of 32 RDs were selected from a former data collection study so that (1) each instruction consists of an action verb coupled with direction and (2) has sufficient path information for making a correct choice. Based on the first set, the second set of RDs was created \((N = 32)\) by adding one landmark to each instruction. The scene set consisted of 32 images retrieved from Google Maps, used for the data collection. The set consists of two scene types: simple (‘T’– and ‘+’– shaped) and complex intersections (‘Y’– and ‘K’– shaped, as well as crossroads with 5 branches) (see Fig.1 and Fig.2). The level of visual clutter in these pictures was estimated using the Feature Congestion algorithm \([6]\) and validated using human ratings. The streets in the intersections were marked with 4 dots of different colours, that corresponded to 4 keyboard keys marked with the same colours. These dots marked the four choices that listeners could make based on the RD. A red arrow showed the position of the viewer and the direction he is facing.

2.3 Procedure

The task for participants was to read the RDs and choose as fast as possible the street that was indicated in the instructions. The experiment started with 4 warm-up trials followed by 32 randomized experimental trials.

Fig. 1. Example of simple / complex intersections in scenes with a low level of visual clutter
3 Results

Crossing the factors Clutter (low / high levels), Intersection type (simple / complex) and Instruction type (instructions with / without landmarks) resulted into a 2 x 2 x 2 design with Clutter and Intersection type as within participants factors and Instruction type as between participants factor. We measured reaction times and correctness scores. These were analyzed separately using logit mixed model analysis with Clutter, Intersection type and Instruction type as fixed factors; participants and item pictures as random factors; p-values were estimated via parametric bootstrapping. Only significant results are reported.

For reaction times, there was a significant effect of Intersection type ($\beta = 0.39, SE = 0.13, p < .05$). Participants decided faster which street to take in simple intersections ($M = 2408ms$), than in complex intersections ($M = 3413ms$). There was an effect of Instruction type ($\beta = 0.26, SE = 0.05, p < .001$). This effect might be due to the different length of the instructions (those with landmarks were longer). Subjects responded faster after reading instructions with no landmarks ($M = 2645ms$), than instructions with landmarks ($M = 3179ms$). There was a significant interaction between Clutter and Instruction type ($\beta = 0.029, SE = 0.028, p < .01$). Subjects who received instructions without landmarks, responded faster for scenes with low levels of clutter (no landmarks $M = 2392ms$), than for scenes with high level of clutter ($M = 2898ms$); the same pattern is observed for subjects who received instructions with landmarks (low clutter $M = 2863ms$; high clutter $M = 3495ms$).

As for correctness rates, participants choose the correct route in 88.73% of the cases. There was an effect of Intersection type ($\beta = -0.90, SE = 0.52, p < .05$). In simple intersections, participants made more often correct choices.
than in complex intersections (86%). There was a significant interaction between the Intersection type and Instruction type ($\beta = 1.55, SE = 0.77, p < .01$). While in simple intersections having or not landmarks did not influence the correctness of choice (91%), in complex intersections participants had more correct responses when receiving landmark information (with landmarks 89%; no landmarks 84%).

4 Conclusions

In this paper, we have investigated how scene complexity and landmark information influence the speed with which participants choose the street described in the instructions and the correctness of their choice. In the light of technological developments our results highlight that, for becoming context-aware, navigation systems should not only add landmarks to the instructions, but also adjust the type of landmarks included to the complexity of the environment.

References