Evaluation of strategies for pedestrian guidance using commodity smartphones

Johan Mollevik
Department of Computing Science
Umeå University, Sweden
Email: mollevik@cs.umu.se

***Abstract*—This paper describes the user evaluation of two different strategies for guiding pedestrians. This work is based on a hands free, eyes free scenario explored by SpaceBook [1], using cheap Android phones. The first model described is a comparatively simple purely reactive model [2]. The second is using more involved techniques including prediction of future positions and scheduling [3]. The paper describes both models discussing identified problems with them. It then goes on to describe our evaluation procedure and the results of the evaluation.**

I. INTRODUCTION

Navigation application using GPS has been a reality on the market for some time now. However most such systems are aimed at car drivers and do not handle the situation of pedestrians all that well. Due to the fact that most pedestrians do not want to carry around yet another device there is a desire that such a system runs on their existing phone hardware. As pedestrians move on a much finer scale than cars, the maps used must be of higher resolution, meaning that small objects like bus stops, benches and pedestrian crossings can be important landmarks and should not be omitted. The need for large databases makes the system badly suited to install on a phone, prompting us to chose a server based solution. The EARS [4] system to our knowledge pioneered the approach of using only audio as both input and output, which we have elected to adhere to. The reasoning behind this approach is for the pedestrian to be hands free eyes free, a concept that the SpaceBook project [1] explored further. The combination of a server based solution and no multi-modality requiring special sensors makes this system well suitable to run on cloud services like Amazon EC2 [5]. The benefit of the cloud in this scenario is mostly one of elasticity as new instances can be started in the same rate new users log on and the destroyed when not in use. If this were to run from a dedicated data center the time it takes to expand a data center could delay adoption in case of a sudden interest because of inability to handle the required number of simultaneous users. Not to mention the cost of owning servers that are likely idle many hours of the day.

II. RELATED WORK

Quite a lot of work has been done in the broader field of computer aided navigation. Car GPS systems by companies like TomTom and Garmin and map based systems like Google Maps are widely available on the consumer market. Much work has also been put into developing systems to aid blind people, see [6] for an example of such work.

Looking more specifically at pedestrian navigation there have been recent work exploring which kinds of map data is needed to improve navigation for pedestrians specifically [7] and ways to obtain such data [8]. There have also been interest in the use of landmarks for guidance (for example [9]). In [10] the state of the art of integration of outdoor and indoor navigation guidance is explored. It is found that mixing indoor and outdoor data is still not solved, a major problem being lack of suitable data that covers both locales.

Regarding tourist navigation, which has been a focus of this work, we find both systems like “I Did It My Way” [11] and PocketNavigator [12] which elected to use haptic feedback as a way to communicate with the user as a way to free up their hands. The use of haptic displays, while unobtrusive, limits the systems to pure navigation instructions. The EARS systems [4] takes another approach by using a headset to serve as the interface with the user. This enables the EARS system to announce points of interest when they pass into view of the user. This approach is continued by the SpaceBook [1] project, which this work builds directly off. More specifically we are building on the system described in [13][14], evaluated in [2], and incorporating the ideas in [15] and evaluating them against the algorithm in the base system.

III. INTERACTION STRATEGIES

An interaction strategy in the sense used in this paper is a method of determining what words to say to the user based on sensor input from the mobile phone. As our work is a way-finding system the goal of the models we describe here are:

- Get the pedestrian to the target destination
- Send utterances to the phone enough in advance so that the pedestrian has time to react
- Guidance instructions should have precedence over reassuring utterances
- Avoid needless repetitions

In this work we have constrained ourselves to the use of speech only, with the idea that it will be safer and more enjoyable for the pedestrian to look around than to stare on their phone’s screen. The intent is that the user wears a headset and has their phone in their pocket.

Both models described guide the pedestrian by issuing turn instructions, instructions to keep moving, telling the pedestrian when moving the wrong way and describing landmarks passed by. Additionally both models read text from the database at the beginning and end of tours.

Both models work against the same infrastructure and database using the same application on the phone, only the code that determines what to say and when to speak differs. We call this code the interaction manager or IM for short.

The database is populated using OpenStreetmaps [16] data and can therefore be easily deployed in most locations without any complicated configuration. Both models, because of how this data is structured, use a street network representation based on connected line segments. This is problematic as it loses width information about roads, leading to the need to make assumptions. On the other hand, not using this type of data would mean that the system could not be deployed in as many locations.

A. Reactive Model

The interaction manager we are using to demonstrate the reactive model is the ASAP Controller [2].

The main parts of this model are:

- A set of propositions (see figure 1), with associated conditions
- A set of speech acts
- A decision tree

- p1: receiving-tts
- p2: gps-adequate
- p3: at-goal
- p4: in-network
- p5: at-branching-point
- p6: aligned-with-edge
- p7: heading-accurate

Fig. 1. The ASAP models propositions

To figure out what to tell the pedestrian, start by checking the conditions and set the propositions to true or false. Then consult the decision tree in figure 2 until you reach a leaf.

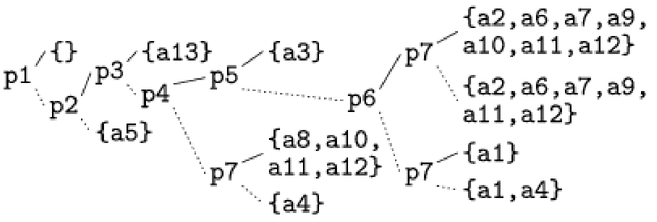

Fig. 2. The decision tree used by ASAP

At the leaves are a set of speech acts (see figure 3); among them select the least recently used and break ties by random selection.

This three step computation, (propositions, decision tree, utterances) is carried out again and again in a loop, along with recomputations of the route if the pedestrian strays off it or reaches the goal. The steps in a single iteration of this loop is shown in figure 4. The loop will run at a pace of one lap every second and to avoid nagging the user with a lot of utterances. It will not voice two utterances within ten seconds of each other unless they are information about an immediate turn or information about a reached goal. To avoid repeating those two it will not reissue them if they are the last thing it said (so no,“turn left”,“turn left”). Most of the time this works fine but it can fail it the route requires two turns in the same direction immediately after each other.

- a1: inform heading correction
- a2: continuation of walk
- a3: immediate turn operation
- a4: move to get a heading
- a5: wait for better GPS
- a6: inform continuation of segment for a span
- a7: inform of-future turn
- a8: state Euclidean distance to goal
- a9: state path distance to goal
- a10: give the heading to goal
- a11: state that progress is being made
- a12: describe landmarks as they pass
- a13: notify a goal has been reached

Fig. 3. The speech acts used by ASAP

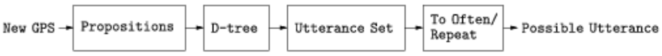

Fig. 4. Computational states of an iteration of the reactive model

B. Predictive Model

The interaction manager for the predictive model is based on the ideas in [3] of using prediction and scheduling to account for changing movement speeds and prioritization among different utterances.

A central part of the system is the use of a meaning representation language (MRL). Each utterance the system produces has an associated MRL sentence. This MRL encodes the pragmatic effect of the sentence and enables the system to know what it has told the user even when there are many ways to say the same thing. The MRL used is based on the MRL in [17].

The model is subdivided into what we call activities, which in the implementation are; route guidance, progress reports and descriptions of the surrounding. The first serves to actually get the pedestrian to the target position while the later two are used for reassurance.

The model works by creating a schedule of utterances each time new GPS information is received from the pedestrians phone.

To do this it iteratively first checks if any activity has anything to say to the pedestrian at the current position at this time. If so the possible utterances, including MRL representation, are stored in a set called the agenda. Then the next current position is predicted and used as the new current position. The number of times this loop runs is controlled by a setting. After the loop a schedule is computed from the agenda using a greedy scheduling algorithm and hard coded utilities associated with each utterance. Our implementation weights these by how far ahead they are in time. If anything should be said immediately it is sent to the phone to be voiced. After this the process restarts. A visualization of one such iteration is shown in figure 5 and a more procedural description of the algorithm can be seen in figure 6.

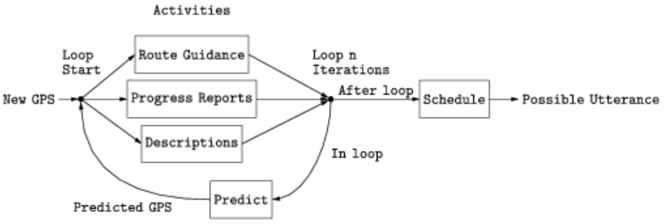

Fig. 5. Computational states of an iteration of the predictive model

1) set agenda to empty set
2) set state to the current DB state
3) let n be the number of seconds into the future to predict
4) let s be the number of seconds for each prediction step
5) loop for n/s iteration
    a) run all activities
    b) add the results from the activities to the agenda
    c) compute next predicted position using algorithm in figure 7
    d) store the prediction in the state
6) partition the agenda based on the MRL of the elements
7) select at most one element from each partition in such a way that no selected utterance overlap in time and the total utility is maximized (implemented using a greedy algorithm). This forms the schedule.
8) If there is an utterance in the schedule that is to be spoken at the current time, send it to the phone to be voiced

Fig. 6. The scheduling algorithm

The prediction model is decoupled from the scheduling model and can be changed for another one if that is desirable. The predictor used in the experiments in this paper works as follows. In the base case it is assumed that the pedestrian keeps their current direction and velocity, interpolated over the last few seconds. In the case that would move the pedestrian off the road or path they are walking on, the predictor makes a small adjustment to the movement angle so that the pedestrian gets closer to become aligned with the road's direction. If the pedestrian is moving quickly and perpendicularly to the road this will not force them on to it but allow them to cross. On the other hand if they are kind of moving along the road the predictor uses this to ensure that they keep doing so, accounting for road bends etc. Finally if a guidance utterance has been spoken or planned about the current position the predictor will assume that the pedestrian follows the guidance and for example turns onto another street. In all cases the speed is assumed to be the same as in the last few seconds. A more algorithmic description of this is seen in figure 7

The prediction checks the following conditions in order and then moves according to the first moving condition maintaining current speed

1) if a turn instruction has been issued for the intersection at the current position (uttered or planed):
   move in the direction of the street we are turning onto
2) if moving in the average direction of the last few time steps would move the pedestrian off the road:
   turn the pedestrian gently in the roads direction (note that if moving quickly and perpendicularly to the road this still make us move off the road by design)
3) otherwise:
   move in the average direction of the last few time steps

Fig. 7. The prediction algorithm

Using this predictor causes an important coupling between the prediction of the next position and the utterances that are planned and waiting in the agenda to be scheduled. This means that the predictor has to take the current agenda as an input. Not doing this makes the predictor dumb, in the sense that if the last position caused a turn instruction to be issued the predictor would not heed it until it was voiced, thus making incorrect predictions. But this coupling also brings complications, such as the need for the prediction and agenda creation to run interleaved, hindering parallelism.

IV. EVALUATION METHOD

To test our algorithms we conducted a user evaluation during the first quarter of 2016. We had 15 subjects test our system using a set of three tours each run once for each interaction manager. This had each subject going on six tours before returning. Upon return the subject was given a questionnaire to fill out and was getting paid 200 SEK for their trouble. The order of the tours was varied for each subject according to the following system.

Assign the possible combinations of 3 tours and 2 interaction managers to the variables $a - f$, where the combinations will be used in alphabetical order for a subject, such that.

- a and f will use different interaction managers
- a and b will use different interaction managers
- e and f will use different interaction managers
- the tours of $[a, b, c]$ will not be the same as $[d, e, f]$ ie no tour $1, 2, 3, 1, 2, 3$ or $1, 3, 2, 1, 3, 2$
- the same tour will not be walked twice in a row

The constraints are chosen to exclude tours that will give away the fact that each tour is taken twice as well as ensuring that the different interaction managers are alternating.

Each subject will use a different one of the possible permutations matching the above constraints.

The tours used are shown in figure 8 and are placed in a residential area close to Umeå University chosen because of a combination of good map coverage and a street network that is both compact and has many turns and non-trivial closest routes.

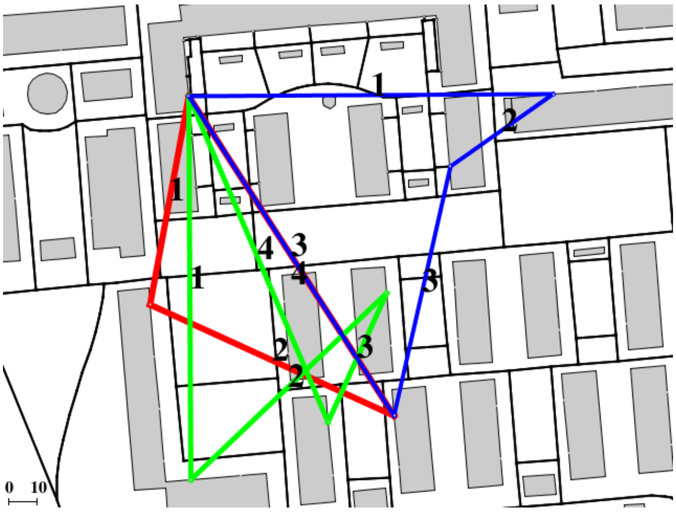

Fig. 8. The tours used in the evaluation

Each test was conducted as follows,

- Instruct subject to move outside and press the start button in the phone application
- Monitor subject and guide them manually using TTS if time to reach any goal exceeds 5 minutes
- Once subject has reached all goals have them return and fill in a questionnaire and get paid

During this process we are recording:

- the subject's position approximately every second
- when they are given a new goal and which goal
- when they reach each goal
- all the utterances the system makes and when they are made

as well as other information more internal to the system, for example which route is planned when and in the case of the predictive model where the user is predicted to be in the future. This stored information allows us a clear view of what the system did and why it did so allowing us to analyze the system's behaviour and hopefully improve it.

For the evaluation, the data we expected to be most useful are the times the user reached each goal, to determine how long they spent on each part of the trip. While we can use the position data to measure the distance traveled, the uncertainty of GPS measurements makes the metric uncertain. This can probably be improved by some kind of smoothing.

V. EVALUATION RESULTS

The evaluation was performed during the first quarter of 2016 in Umeå. As it was winter we had snow covering the ground during most experiments. The weather was varied but for most subjects quite good, meaning no rain or snowfall and no extremely cold temperatures. We ran 15 subjects through the evaluation and had them fill in a questionnaire. The results of that are presented below.

As can be seen in table I we had quite an even age distribution, we have the lowest interval as 16-25 as we had 16 as the minimum age to participate and thought that 10 year blocks was an appropriate resolution.

16-25	26-35	36-45	46-55	56-65
4	4	3	2	2

TABLE I
AGE DISTRIBUTION IN THE EVALUATION

Similarly in table II we can see an even gender distribution, (no one identified as anything other than the classical genders).

Male	Female
6	9

TABLE II
NUMBER OF MALES AND FEMALES IN THE EVALUATION

We did not manage to discern any groups of occupations among the subjects; they where too diverse, but regarding education level none had not attended high school and none had any degree higher than a master's degree. The distribution among high school and university education can be seen in table III.

High school	University
7	8

TABLE III
NUMBER OF SUBJECTS WITH HAD HIGH SCHOOL EDUCATION AND UNIVERSITY EDUCATION

	Reactive	Predictive
Time taken	09:00:48	08:29:32
$>$ 5min to goal	43	32
Quicker for user	4	11

TABLE IV
NUMERIC COMPARISON OF THE RESULTS OF THE REACTIVE AND PREDICTIVE MODEL

The results for each system are compared in table IV. Apart from the information in the table, in 8 of the cases of a subject taking more than 5 minutes to reach a goal the same goal took over 5 minutes using both methods. Comparing the times it took for the different methods we can see that the reactive method is 6% faster.

To test how statistically certain it is that the predictive system is faster than the reactive system we sum the time each subject has spent using each system, clamping those over 5 minutes to 5 minutes to avoid measuring the efficiency of the manual help they received, and perform a paired Wilcoxon signed rank test with the alternate hypothesis that the predictive system is faster than the reactive one. Computing this we get a p-value of 0.04163 and can reject the null hypothesis at the 5% level.

A. *User feedback*

In the questionnaire we gave to the users when they had been run through the tours we included a set of open ended questions. Translated from Swedish they where:

- How many systems do you think we tested?

- Which system did you prefer? Describe:
- Was the system something that could be of use to you?
- Was there anything about the system that was good?
- Was there anything about the system that was bad?

The first question about how many systems where tested received answers from 2-6 with several subjects indicating that they where not sure. As most people answered something other than two, and several of those who did indicated verbally that they where guessing, we conclude that the subjects did not manage to distinguish between the reactive and predictive system as a whole. The subjects did however identify aspects of those systems they liked and disliked; more on that in further questions.

As the subjects did not manage to identify the different systems the question about which they preferred ended up being answered by descriptions about what they liked about the system.

The answers to the question about whether the system could be of use to them the users were quite evenly divided among those saying ‘no’, ‘yes’ and ‘yes but only in a more refined form’. Some of the yes answers specified what they would use it for. Two subjects mentioned being guided in a new city, which the system was originally designed for. Another two mentioned using it to just get to an address. Also mentioned were; as exercise (following a tour on a walk presumably), finding new places to walk the dog and as one user put it ”IRL pacman”. Interestingly one user also mentioned that it could be of more use when bicycling, as the number of hands is more restricted in that case.

Things users thought where good about the system was very varied. The things several pople liked where;

- the direction to goal strategy using clock directions employed by the reactive model
- the distance to goal utterances both models shared
- the mentioning of visible objects (mostly addresses of buildings in the test area) as the user passed them that the prediction model uses
- that the system seemed to know were they where which was found reassuring.

Other than that no trends could be discerned in the answers.

Regarding things the subjects did not like about the system two comments stand out. Most obvious with five subjects mentioning it is the fact that the system gets stuck in a loop. This loop consists of saying turn around followed by a replan of the route. The user not aware of the new route turns around, resulting in another turn around utterance, and repeat. Second most obvious with four subjects mentioning it where that the system was to late with instructions. Less common objections where bad distance measurements, too fast instructions, frequent loss of GPS fix and contradictory instructions.

A consideration when interpreting survey results is the fact that comments in the survey can for the most part not be attributed to either model; this is a drawback of not telling them in advance when they were using each system which was done to prevent bias for one system or the other.

B. Problems encountered

The two major problems encountered by subjects was bad weather and problems with the, among the interaction managers shared, route planner.

The weather issues are relatively straightforward. The first of two weather related issues was that during some tests there was heavy cloud cover and during these tests the GPS precision suffered, leading to tired and frustrated subjects. Sadly no record of cloud cover was being performed as part of this evaluation.

The other weather related issue is that some of the paths where getting increasingly snowed over, seen in figure 9, as the tests progressed. This continued to the point that subjects had no idea that there was a path there, while the systems stubbornly kept trying to guide them over those paths. For some subjects this was a major issue while others handled it by either ignoring some instructions or walking through the snow.

Fig. 9. Snowed over road encountered in the evaluation

The shared route planner has several issues.

The first is a plain bug, we are storing the street network as a directed graph with every edge mirrored by an edge in the other direction (for easier computations in several cases). The route planner fails to recognise this when fetching the closest edge and instead of getting the mirrored pair it takes one of them at random (dependent on database insertion order). It then plans the route from the end node of this edge which in many cases results in non optimal and most importantly confusing routes for the user. This issue was unknown before the evaluation started and we elected not to fix it during the evaluation to make the data from all the subjects comparable.

The second issue, as identified by the subjects in the questionnaire, is the fact that both systems can get ’stuck’ in a loop. This consists of telling the user they are moving the wrong way and then replanning the route so that the direction the user now is moving is wrong, and so on. This is caused by the fact that we are running the route planner independently of the interaction manager. If the interaction manager was deciding when to call the route planner it could give the user time to react after issuing a turn around instruction before replanning the route.

The third issue is the fact that the route planner always tries to find the shortest way even when it would force the user to turn around. A better strategy is probably to check if there is a way that is not much longer where a turn around utterance can be avoided, and only if that does not exists select a route that forces the user to turn around.

Additionally there were some minor issues, among them the fact that both interaction managers are quite slow at detecting when the user is moving in the wrong direction if they are still on the path. This was made worse by the fact that the route planner can plan a new path that is the same as the old path and one additional edge, covering where the user has moved of the path, thus keeping them on the path. There were also issues getting a certain orientation for users who walk slowly, (this one) due to GPS noise.

VI. CONCLUSION

We show in this paper that our algorithm based on predicting user positions and scheduling utterances in advance gives a more efficient guidance instruction than our baseline system in [2], both by total time taken for subjects and by the number of times the subjects hit the time limit and had to be helped.

VII. DATA CONTRIBUTION

To facilitate validation of these results the data generated during the evaluation are available as an PostgreSQL 9.5 dump on the following address http://janus-system.org/jan2016eval/janus_eval_january_2016.sql. Feel free to use this for any research purposes, and if doing so it would be appreciated if I get to hear how it was used. Our software is not yet published but if anyone would like a copy feel free to email us.

VIII. DISCUSSION AND FUTURE WORK

By constructing this work as a comparison with an older system we constrained ourself to using similar strategies as that system for the parts of the system not being compared. As work continues to build on this we expect to incorporate the results of the user study to improve the system and also integrate guidance based on landmarks. We also expect to modify the route planning that has shown to be an issue in this evaluation. Both to fix the obvious problems as well as making the planning more fuzzy and conforming to the users expected path rather than planning strictly based on the shortest path. A further avenue of study could also be different strategies for prediction, both simpler and more complicated, as well as evaluating whether it gives better instruction to replan continuously as we do in this paper or if a strategy of replanning only when the user deviates from the plan is more efficient.

ACKNOWLEDGMENTS

This work is a continuation of the work at Umeå University during the SpaceBook project by the research group lead by Michael Minock. Thanks is also extended to Mattias Åsander for doing a large part of the work of implementing the technical infrastructure [13].

REFERENCES

[1] M. Minock, J. Boye, S. Clark, M. Fredriksson, H. Geffner, O. Lemon, W. Mackaness, and B. Webber, "Summary of spacebook project results," SpaceBook Project, Public Deliverable, 2014.

[2] M. Minock, J. Mollevik, and M. Asander, "Does tts-based pedestrian navigation work?" Technical Report UMINF-14.07, Umeå University, 2014.(http://www8. cs. umu. se/research/uminf/index. cgi, Tech. Rep., 2014.

[3] M. Minock and J. Mollevik, "Prediction and scheduling in navigation systems," in *GeoHCI 2013-Geographic Human Computer Interaction Workshop at CHI 2013, Paris, France, April 27-28, 2013.*, 2013, pp. 30–32.

[4] P. J. Bartie and W. A. Mackaness, "Development of a speech-based augmented reality system to support exploration of cityscape," *Transactions in GIS*, vol. 10, no. 1, pp. 63–86, 2006.

[5] "Amazon ec2 - virtual server hosting," 2016, [Online; accessed 20-April-2016]. [Online]. Available: \url{https://aws.amazon.com/ec2/}

[6] H. Makino, I. Ishii, and M. Nakashizuka, "Development of navigation system for the blind using gps and mobile phone combination," in *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, vol. 2, 1996, pp. 506–507.

[7] C. Gaisbauer and A. U. Frank, "Wayfinding model for pedestrian navigation," in *AGILE 2008 Conference-Taking Geo-information Science One Step Further, University of Girona, Spain*, 2008.

[8] H. A. Karimi and P. Kasemsuppakorn, "Pedestrian network map generation approaches and recommendation," *International Journal of Geographical Information Science*, vol. 27, no. 5, pp. 947–962, 2013.

[9] A. Basiri, P. Amirian, A. Winstanley, S. Marsh, T. Moore, and G. Gales, "Seamless pedestrian positioning and navigation using landmarks," *Journal of Navigation*, vol. 69, no. 01, pp. 24–40, 2016.

[10] A. Vanclooster, N. Van de Weghe, and P. De Maeyer, "Integrating indoor and outdoor spaces for pedestrian navigation guidance: A review," *Transactions in GIS*, 2016.

[11] S. Robinson, M. Jones, P. Eslambolchilar, R. Murray-Smith, and M. Lindborg, "I did it my way: moving away from the tyranny of turn-by-turn pedestrian navigation," in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, 2010, pp. 341–344.

[12] M. Pielot, B. Poppinga, and S. Boll, "Pocketnavigator: vibro-tactile waypoint navigation for everyday mobile devices," in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, 2010, pp. 423–426.

[13] M. Minock, J. Mollevik, and M. Åsander, "Toward an active database platform for guiding urban pedestrians," Technical Report UMINF-12.18, Umeå University, 2012.(http://www8. cs. umu. se/research/uminf/index. cgi, Tech. Rep., 2012.

[14] M. Minock, J. Mollevik, M. Åsander, and M. Karlsson, "A test-bed for text-to-speech-based pedestrian navigation systems," in *Natural Language Processing and Information Systems*. Springer, 2013, pp. 396–399.

[15] M. Minock and J. Mollevik, "Prediction and scheduling in navigation systems," in *GeoHCI 2013-Geographic Human Computer Interaction Workshop at CHI 2013, Paris, France, April 27-28, 2013.*, 2013, pp. 30–32.

[16] M. M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, Oct. 2008. [Online]. Available: http://dx.doi.org/10.1109/MPRV.2008.80

[17] A. Vlachos, S. Clark, T. Dalmas, R. Hill, S. Janarthanam, O. Lemon, M. Minock, and B. Webber, "D4.1.2 final request analysis component," SpaceBook Project, http://spacebook-project.eu/pubs/D4.1.2.pdf, Public Deliverable, August 2013.