Using Thinking-aloud Data to Characterize Cognitive Tasks in Hypertext

Charlotte van Hooijdonk
Tilburg University
C.M.J.vanHooijdonk@uvt.nl

Alfons Maes
Tilburg University
Maes@uvt.nl

Nicole Ummelen
Tilburg University
Ummelen@uvt.nl

Abstract

In this paper, we explore the usefulness of thinking-aloud protocols as a means of characterizing cognitive tasks involved in executing digital information tasks. In particular, we used protocol analysis to detect levels of cognitive tasks in which hypertext users are involved, and to investigate the way in which spatial descriptions are used to conceptualize users' tasks.

Keywords: *hypertext tasks, cognitive tasks, spatial representation*

Cognitive Tasks in Hypertext Use

When searching, reading, studying, or browsing information in a hypertext, users perform a large number of tasks, which can be evaluated as more or less efficient and effective. However complex hypertext use may be, technology allows us to meticulously register the user's behavior in terms of rational quantitative (log) measures like searching time, number of visited or unique nodes, link patterns, number of correct answers, etc.[1-3] Just as psychologists agree that reading time is indicative of cognitive effort, hypertext researchers assume logical relationships between specific log measures and the quality of task execution. For example, a longer task time, an intensive use of the back button, or a low proportion of unique pages visited are considered to signal poor hypertext performance, while a high proportion of unique nodes or long distance clicks are assumed to signal successful performance. However plausible these assessments may be, they do not explain the user's behavior in terms of the cognitive ramifications they are involved in. Thus, the preferential use of the back button may be the result of different cognitive strategies, operating on different levels of task execution. It may reflect the preference of users to carefully navigate hierarchically by going back the same route through the information structure. It may also signal the users' fear to lose track or their inability to set and realize their information goals efficiently. But it may also result from a low level strategy, which is based on the perceptual saliency of this button, as it takes an invariable and preferential position in the upper left of an otherwise fairly variable screen. In other words, quantitative measures cannot be mapped unambiguously to cognitive tasks or strategies, nor do they relate unequivocally to effectiveness measures like information recall. [4]

In this article, we first discuss theoretical and methodological barriers in defining and categorizing cognitive tasks involved in hypertext use. Furthermore, we describe the research method and preliminary results of an exploratory study in which thinking-aloud protocols were used to determine and characterize cognitive tasks involved in hypertext use.

Detecting and Categorizing Cognitive Tasks

Determining the range and interaction of cognitive tasks in hypertext use poses theoretical as well as methodological problems. Theoretically, there is no generally accepted overall cognitive theory of hypertext use from which types of cognitive tasks as well as their interaction can be deduced. [5] What we have is a large number of proposals in which different types of cognitive tasks of hypertext uses are modeled, starting from a particular (ergonomic, learning or usability) perspective. For example, Shneiderman's Object-Action Interface Model suggests different types of actions, associated with either the task or the interface, as can be seen in Figure 1: the left-hand trees represent to the relevant (semantic) content and (pragmatic) intentions, the trees on the right refer to the conceptualization and manipulation of the interface, respectively. [6]

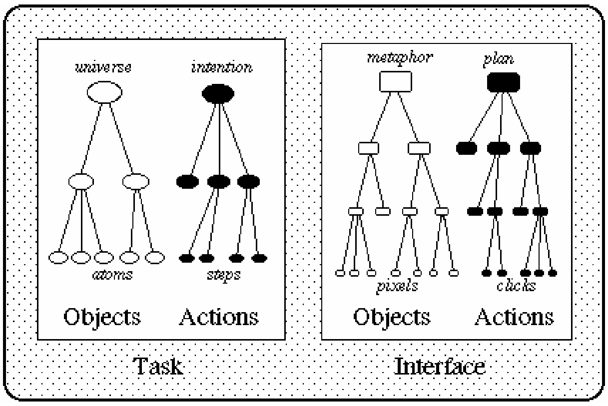

Figure 1. Object-Action Interface Model (Shneiderman, 1998: p.62)

Dillon represents hypertext use in terms of his TIMS framework, consisting of four interactive elements as they are displayed in Figure 2: task, information space, manipulation skills, and standard reading processor. These components reflect the cognitive, perceptual, and psychomotor aspects of information usage. [7]

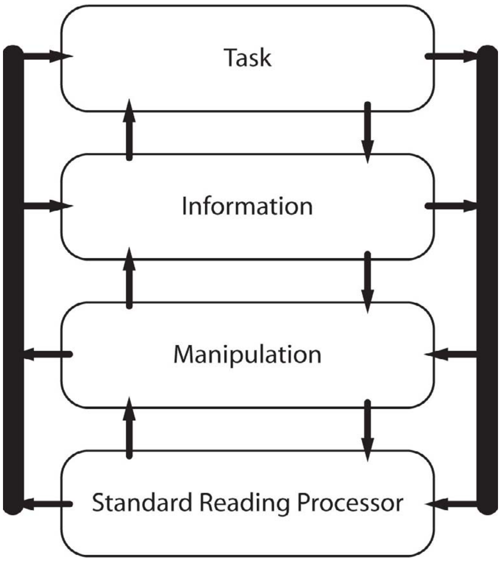

Figure 2. TIMS framework (Dillon, 1996, p.111)

Likewise, Rouet and Tricot (1996), and Tricot, Puigserver, Berdullo and Diallo (1998) model information usage tasks in a three layer framework, presented in Figure 3. [4, 8] The three levels are the rational task model (i.e., the most efficient 'rational' representation of the information task), the cognitive task model (i.e., the users' representation of the goal, the means and the environment) and the cognitive activity level (resulting from the application of a cognitive task in a particular situation). Each of these levels is defined in terms of goals, means, and the relevant environment.

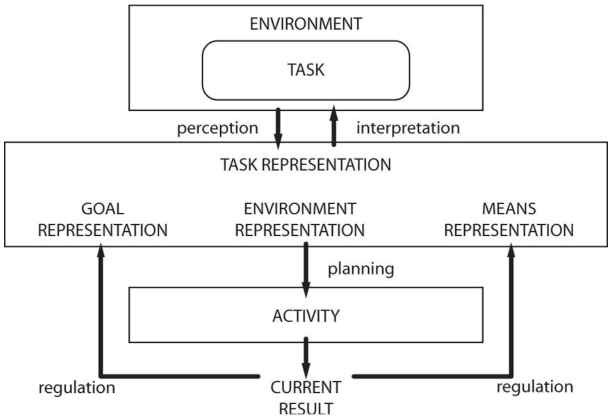

Figure 3. Task and activity model (Tricot, Puigserver, Berdullo and Diallo, 1999, p.24)

In many other studies, hypertext tasks are not so much modeled, but merely listed. For example, Kim and Hirtle (1995) mention three types of tasks involved in hypertext use: Navigational tasks (planning and executing routes through the network), informational tasks (reading and understanding the content presented in the nodes and their relationships for summary and analysis) and management tasks (coordinating the informational and navigational tasks). [9]

Although the theoretical ambition of these models and typologies and the way they present the interaction between types of cognitive tasks differ considerably, they largely agree on the basic categories of cognitive tasks involved. In our exploratory study, we intend to cover these categories by using the well-known tripartition *pragmatic*, *semantic* and *perceptual/syntactic*. Hypertext users can be seen as being engaged basically in three types of cognitive tasks, and three levels of mental representation:

1. The *pragmatic level* of planning, setting (sub)goals, selecting, and monitoring task execution
2. The *semantic level* of building a coherent representation of the information content and the conceptual information structure
3. The *perceptual level* of representing the actual execution of the task in terms of time, space, and supporting actions (e.g. screen management, hand/eye coordination, and mouse/keyboard management)

This tripartition is not claimed to be better than the components mentioned in the models above, but we assume that the tripartition allows for a more general characterization of tasks, which is not dependent on specific hypertext tasks or environments. Furthermore, we hope that the three linguistic notions will be appropriate to determine cognitive tasks in terms of linguistic characteristics of thinking-aloud data.

Methodologically, there are different ways of discovering the nature and the interaction of cognitive tasks involved in hypertext use. As we mentioned above, direct (i.e., unintrusively obtained) online data, such as log files, enable us to reconstruct the user's behaviour in quantitative terms, and they are mostly used to qualify the efficiency of task execution. These data allow for interpretations of behaviour in terms of different types of cognitive tasks as well, as our example of using the back button demonstrated. However, there is a fairly large interpretation range, which partly results from the lack of a predictive theoretical model, and partly from the lack of empirical validation of the relationship between rational measures and cognitive tasks.

Cognitive tasks can be deduced from direct offline data, such as effectiveness measures, like (content or structure) recall or transfer tasks. For example, differences in the quality of structure map drawings or in content recall can suggest differences in semantic or structural mental representations of the hypertext environment.

Indirect, offline data, such as post test evaluation questions can be interpreted in terms of cognitive tasks as well. For example, differences in perceived disorientation can be interpreted in terms of differences in mental representations of the hypertext environment. [10]

In this study, we will concentrate on indirect online data, i.e., thinking-aloud protocols. The primary goal is to explore to what extent cognitive tasks and task levels can be deduced from thinking-aloud protocols, how task levels and task types can be concluded based on linguistic characteristics of thinking-aloud data. We realize that thinking-aloud protocols do not directly tap the users' cognitive tasks, or provide us with indisputable evidence for the psychological validity of tasks or levels, but we assume that the users' verbalizations can reveal the nature and diversity of cognitive tasks of hypertext users, as well as the relative distribution of attention over types and levels of cognitive tasks.

Apart from that, we are interested in one specific aspect of users' verbalizations, i.e., the way in which cognitive tasks are conceptualized in terms of spatial descriptions. That way, we hope to use this method to assess the importance of space in hypertext use.

Space in Hypertext

Space is a powerful means of conceptualizing digital tasks, witness the use of notions like "hyperspace", "browsing", and "visiting a web site navigation". This is not surprising given the fact that space is an overall powerful metaphor to mould many different conceptual domains, as is clear from research on spatial cognition. [11, 12] The premise of the spatial metaphor in hypertext is that using hypertext environments has similar psychological features as navigating in a physical environment. [13, 14]

Much research has been done in the field of geography and psychology on how users represent spatial information in physical space. When people are operating in a spatial environment, they need to know the location and the attributes of objects in that particular environment. [15] The location information of an object consists of the concepts of distance and direction. Attributes of an environmental object can be descriptive and evaluative. Shum (1990) tried to apply these spatial notions to hypertext. He attributed distance to hypertext nodes, and proposed to quantify link distance in absolute or relative terms. Each hypertext node has a certain distance, which - according to Shum - can be quantified in absolute terms as the number of nodes a user has to visit, or in relative terms as the cognitive distance. Cognitive distance is related to aspects like system response time, ease of returning to the previous node, the number of link traversals/mouse clicks involved. Although he tried to conceptually apply space to hypertext, he did not investigate whether or not users make use of spatial concepts to mentally represent hypertext tasks. [16] Boechler (2001) meticulously analyzed the physical and psychological correspondences between physical and digital space. She concludes that it is hard to investigate the psychological value and validity of the spatial metaphor in using hypertext.

Although thinking-aloud protocols do not provide us with conclusive evidence on the mental effect of spatial representations, they can shed more light on the proportion of spatial descriptions, the linguistic characteristics and the nature and levels of the cognitive tasks, which are conceptualized in spatial terms. This may indirectly assess the importance of spatial conceptualizations in using hypertext.

Research Method

Material

Twenty verbal protocols were collected from three usability experiments. In all these experiments users were asked to perform search tasks in a hypertext, and were asked to think aloud while executing these tasks.

The first five verbal protocols originated from a usability research at a Dutch web site of the European Committee (EU)[1]. Five participants between 20 and 25 years old took part in this experiment. The second five verbal protocols originated from a usability research at a Dutch medical website (MED)[2]. Five participants between 18 and 25 years old contributed in this experiment. The other ten protocols were derived from an experiment at the official web site dedicated to the Dutch writer Harry Mulisch (MUL)[3]. Ten subjects between 20 and 25 years old participated in this investigation.

The three websites used to elicit the protocols differed substantially in terms of perceptual and spatial look (see Figure 4). The EU and MED websites were traditional in their information structure. They used textual hyperlinks, and had several standard search facilities, like a sitemap and a search function. The MUL website is highly unconventional. The website is designed as a journey through the mind of the Dutch writer Mulisch and is supposed to be a metaphor for the way he thinks. [17] This is symbolized on the home page by a dark castle with various entrances, corridors and windows. Each of these features represents one of the themes of the writer's work. Users can explore these themes by navigating through the relevant parts of the castle. Standard hyperlinks are absent. Instead visitors have to manipulate virtual objects to navigate. [18] That way, the website contained a strong spatial metaphor. This variation in website type enabled us to investigate whether the spatial character of a website is reflected in the thinking-aloud protocols.

Coding Procedure and System

The verbal protocols were coded with the program MAXQDA[4]. Of each verbal protocol 50 utterances were segmented in clauses and coded by one analyst (the first author). Altogether, the corpus consisted of 1170 coded segments in total. To increase the reliability of the analysis, a second analyst independently coded parts of the corpus.

Each utterance was coded on the following variables:

Type of Action We distinguished two global types of actions: active and reactive actions. The variable is comparable to Norman's difference between executions and evaluations. [19] Also, it refers to different information types in instructive discourse (goal/action vs result information).
1. *Active actions* are verbalization of tasks originating from deliberate users' goals and intentions *(I am going back to the homepage)*.
2. *Reactive utterances* start from an observation in the task context (mainly on the computer screen), which asks for a reaction of the user (*I cannot click on this item)* or an evaluation *(This is not what I am looking for)*.

Levels of Cognitive Task As mentioned above, we defined three levels of cognitive task, a syntactic/perceptual level, a semantic and a pragmatic level. Users have to perceive the different information modalities on the screen. They have to decide on their navigation path, and they have to understand the information, not only in semantic terms, but also pragmatically, in view of their tasks and goals.

[1] http://europa.eu.int/comm/index_nl.htm
[2] http://www.life4u.nl
[3] http://www.mulisch.nl
[4] http://www.maxqda.com/

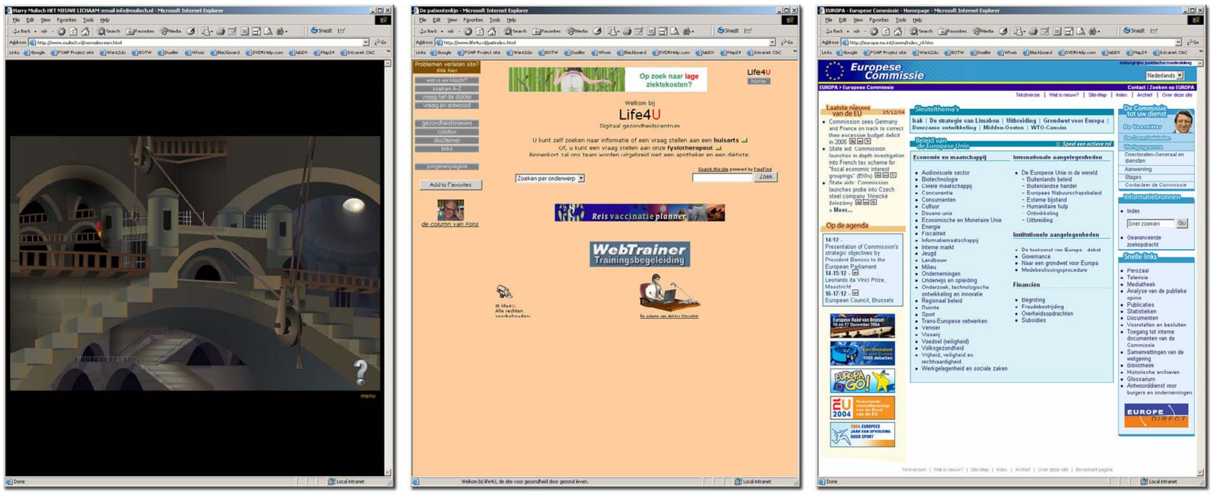

Figure 4. The three web sites (MUL, MED, EU)

Utterances on the *syntactic level* concern the perception of elements in the environment (*I see three hyperlinks*), the coordination of actions with mouse and keyboard (*where is the mouse ointer*?), assumptions on technical aspects of the hypertext (*is this element clickable?*), the verbalization of technical actions (*I type in the word 'movement*).

2. Utterances on a s*emantic level* relate to the comprehension (problems) of elements on the screen (*This is an interview about the books he likes*), semantic inferences made during reading and interpreting (*This is about the nations who are united in the European Union*), or conceptualizations about the global or local structure of the website *(This is a chronological list of the available numbers)*.
3. Utterances on a *pragmatic level* concern higher and lower order goal orientation and expressions of users' strategies (*I am going to search on seats)*, reflections on the specific task (*What is the name of the book I am looking for?*), and reflections and evaluations of screen results in terms of the task (*I think I have found the answer*).

Spatial conceptualisation of tasks Utterances can be spatial or non-spatial. Verbalizations are *spatial* if they take the form of a spatial proposition (*I'm going back to the homepage)*. Spatial expressions can refer to the virtual space of the hypertext information or to locations beyond the screen (*I have the feeling that I am close*, or *I do not know where I am)*.

Results
The first author coded a total of 1170 utterances. Table 1 shows the results of this analysis. The results show that the spatial variable is less problematic than the other two analytical variables. The linguistic characterization of the values (spatial vs. non spatial) makes it relatively easy to classify an utterance as spatial or non-spatial. It is less easy to assess the action type and the task level. About one out of four utterances cannot be coded.

Table 1. Overall results of the type of action, level of cognitive task, and conceptualisation of the task (Scores are percentages; N = 1107)

Type of action	Active actions	28.6
	Reactive actions	38.7
	Non-coded	32.7
Level of cognitive task	Syntactic/perceptual level	18.6
	Semantic level	32.9
	Pragmatic level	20.6
	Non-coded	27.9
Conceptual isation of the task	Spatial	11.7
	Non-spatial	86.3
	Non-coded	2.0

We were also interested in the relation between the spatial conceptualisation of tasks and the levels of cognitive task. Table 2 shows the results of this analysis.

Table 2. Spatial conceptualisation of tasks related to the levels of cognitive task (Scores are percentages; N = 797)

	Syntactic level	Semantic level	Pragmatic level
Spatial conceptualisation	5.4	9.3	1.3
Non-Spatial conceptualisation	20.5	36.4	27.2

Spatial conceptualizations occurred mostly at the semantic and syntactic level. Examples of these cases are: *I am going back* and *I visited Time*[5] *before*. An example of a spatial conceptualisation on a pragmatic level was: *This is the web site were I should be*. Furthermore, most of the spatial conceptualizations refer to a backward rather than a forward movement of the user (*I am going back to the homepage now).*

Given the qualitative nature of this analysis, and the precarious status of the variables, it is important to increase the reliability of the analysis. Therefore, a second analyst coded part of the corpus. A total of 264 utterances were coded by the two analysts. The results indicated that both analysts highly corresponded in judging the utterances as *active or reactive* (inter-rater reliability score Cohen's κ = .80) or *spatial or non-spatial* (Cohen's κ = .67). However, there was low correspondence in coding the utterances as *syntactic/perceptual, semantic, or pragmatic* (inter-rater reliability score Cohen's κ = .47)[6].

Discussion

The joint analysis of thinking-aloud data in terms of cognitive tasks shows a large degree of agreement with regard to conceptualization of the task and type of action. However, analysts disagree to a large extent on the level of cognitive tasks. In a post hoc analysis, a number of these disagreements can be explained by systematic differences in the way in which utterances were interpreted in terms of cognitive tasks.

[5] *Time* is a particular part of the Mulisch web site.

[6] Another indication for the problematic status of this variable is the high number of cases in which the first analyst doubted about the coding (about 20% of all cases).

Semantic Versus Syntactic Tasks

In a number of cases, the analysts disagreed on the semantic vs. syntactic level of the task verbalized in the segments. Utterances were coded as semantic if they somehow gave evidence of a conceptualization of the content or contained a consideration on the structure of the semantic environment. Syntactic/perceptual utterances verbalized actions in terms of situational and visible interface objects.

One obvious linguistic signal for discriminating between these two interpretations is the use of nominal vs. pronominal expressions to refer to the relevant objects and actions. Thus, in an example like *I am going back to the Publication Magazine of the European Union*, the user clearly refers to the information segment *The Publication Magazine* by semantically conceptualizing it within the relevant information space, thus demonstrating semantic knowledge about the information environment. On the other hand, in an example like *Let see if I can read this*, the user simply uses a deictic referential strategy to refer to the perceptual environment of the screen to verbalize the next task.

This heuristic, however, is not fully waterproof. Nominal reference does not always signal semantic conceptualization, but it can also be the result of a simple naming or reading act right from the screen, in which case it is more like a syntactic action. Consider the following example: *I am clicking now on education*. The nominalization *education* most probably indicates a perceptual reading act: the user reads the link education from the screen. In that case, the nominal is the result of a naming act of a visible object, rather than a semantic conceptualization.

An analogous type of disagreement can be seen in cases like *I am reading the headings* or *I am reading the content of this page*. These cases can be seen as semantic because these actions clearly imply the processing and comprehension of information. However, the examples can be read as syntactic as well, if they are regarded as expressing the technical, low-level activity of reading from the screen. The relevance of these cases is reflected in Dillon's TIMS framework by the so-called standard reading processor. This component deals with the user perceiving the document and performing the activity of reading [

Syntactic Versus Pragmatic Tasks

In a number of cases, the analysts disagreed on the syntactic versus pragmatic character of the actions verbalized in the segments. Utterances were coded as pragmatic if they gave evidence of higher and lower order goal orientation and expressions of users' strategies. Syntactic utterances concern the perception of elements in the environment, and verbalization of technical actions. However, there were cases in which the difference between these two codes was not clear. Look at the following example: *I am going to use the search engine*. From a syntactic perspective this utterance could refer to a generic, 'low-level' syntactic strategy. But from a pragmatic perspective it can also be the result of a 'higher-level' intentional strategy, which implies the translation of the user's task in terms of the interface.

The difference between these two interpretations is accounted for in Shneiderman's Object-Action Interface Model, especially in the difference between the task oriented and interface oriented pragmatic actions. [6] In the above-mentioned case, the pragmatic task clearly refers to the interface level.

Conclusion

The results of the analysis of the verbal protocols showed that it is possible to define types of actions and spatial utterances, probably because they are fairly generic or to be defined in specific linguistic terms. However, the analysis of cognitive tasks shows that cognitive tasks involved in using hypertext do not show themselves that easy and unequivocally in thinking-aloud protocols. Yet, the protocols gradually show interesting parallels with the models and task types as they are presented in the hypertext literature.

References

[1] D. R. Danielson, "Web navigation and the behavioral effects of constantly visible site maps," in *Interacting with computers*, vol. 14, 2002, pp. 601-618.

[2] M. Otter and H. Johnson, "Lost in hyperspace: metrics and mental models," *Interacting with Computers*, vol. 13, pp. 1-40, 2000.

[3] P. A. Smith, "Towards a practical measure of hypertext usability," *Interacting with Computers*, vol. 8, pp. 365-381, 1996.

[4] A. Tricot, E. Puigserver, D. Berdugo, and M. Diallo, "The validity of rational criteria for the interpretation of user-hypertext interaction," *Interacting with Computers*, vol. 12, pp. 23-36, 1998.

[5] C. Chen and R. Rada, "Interacting with Hypertext: A Meta-Analysis of Experimental Studies," *Human-Computer Interaction*, vol. 11, pp. 125-156, 1996.

[6] B. Shneiderman, *Designing the user interface. Strategies for effective human-computer interaction*, 3 ed. Reading, Mass.: Addison-Wesley, 1998.

[7] A. Dillon, *Designing usable electronic text: ergonomic aspects of human information usage*. London: Taylor & Francis, 1994.

[8] J. F. Rouet and A. Tricot, "Task and Activity Models in Hypertext Usage," in *Cognitive Aspects of Electronic Text Processing*, vol. 58, *Advances in Discourse Processes*, H. Van Oostendorp and S. De Mul, Eds. Norwood, New Jersey: Ablex Publishing Corporation, 1996, pp. 239-264.

[9] H. Kim and S. C. Hirtle, "Spatial metaphors and disorientation in hypertext browsing," *Behaviour & Information Technology*, vol. 14, pp. 239-250, 1995.

[10] J. S. Ahuja and J. Webster, "Perceived disorientation : an examination of a new measure to assess web design effectiveness," in *Interacting with computers*, vol. 14, 2001, pp. 15-29.

[11] B. Tversky, "Spatial schemas in depictions," in *Spatial schemas and abstract thought*, M. Gattis, Ed. Massachusetts: The MIT Press, 2001, pp. 79-112.

[12] M. Raubal, M. J. Egenhofer, D. Pfoser, and N. Tryfona, "Structuring Space with Image Schemata: Wayfinding in Airports as a Case Study," presented at COSIT, 1997.

[13] P. M. Boechler, "How spatial is hyperspace? Interacting with hypertext documents: cognitive

processes and concepts," *CyberPsychology & Behavior*, vol. 4, pp. 23-46, 2001.

[14] J. S. Farris, K. S. Jones, and P. D. Elgin, "Users' schemata of hypermedia: what is so "spatial" about a website?," *Interacting with computers*, vol. 14, pp. 487-502, 2002.

[15] R. M. Downs and D. Stea, "Cognitive maps and spatial behavior: process and products," in *Image and Environment*, R. M. Downs and D. Stea, Eds. Chicago: Aldine Publishing, 1973, pp. 8-27.

[16] S. Shum, "Real and virtual spaces: mapping from spatial cognition to hypertext," *Hypermedia*, vol. 2, pp. 133-158, 1990.

[17] H. den Hartog Jager, "Dolen door des schrijvers geest," in *NRC Handelsblad*, 2000.

[18] E. Krahmer and N. Ummelen, "Thinking about thinking aloud: A comparison of two verbal protocols for usability testing," *IEEE*, pp. 1-17, 2004.

[19] D. A. Norman, *The Design of everyday things*. London: The MIT Press, 1998.

About the Authors

Charlotte van Hooijdonk is a PhD student at Tilburg University, the Netherlands. A part of her PhD thesis deals with spatial cognition in digital environments. Another part of her thesis, deals with information presentation and evaluation of multimodal information in the medical domain.

Alfons Maes holds a chair at Tilburg University, the Netherlands, and is head of the Communication and Cognition section. His research interests include multimodal and digital communication, and document design. He has published articles on discourse reference, document design, instructive discourse. He is member of IEEE.

Nicole Ummelen (Ph.D. 1997) is currently affiliated with the Communication and Cognition section of the Faculty of Arts at Tilburg University. Her research focuses on information selection processes, especially in instructional documents and digital environments. She is a member of IEEE-PCS.

Acknowledgments

This research was partly conducted within the IMIX project Interactive Multimodal Output Generation (Imogen), sponsored by The Netherlands Organization of Scientific Research (NWO).