A RESEARCH AGENDA FOR MODALITY THEORY

Niels Ole Bernsen, Centre for Cognitive Science, Roskilde University

Summary: The paper describes the comparatively new field of modality theory as trying to establish a principled scientific basis for solving the following information-mapping problem in human-computer interaction or usability engineering: Given any particular class of task domain information which needs to be exchanged between user and system during task performance, identify the set of input/output modalities which constitute an optimal solution to the representation and exchange of that information. A research agenda for modality theory is presented together with first steps towards its implementation.

1. Introduction

Contemporary designers of interactive human-computer interfaces have the opportunity of being able to use a rapidly increasing number of different, and sometimes alternative, input/output modalities for the expression and exchange of information between computer systems and their users. The interface designer's task can be described roughly as follows: (1) Identify the information to be exchanged between user and system from an analysis of the task domain of the artifact to be built; (2) know your inventory of input/output modalities; (3) perform an optimal match in terms of functionality, usability, naturalness, efficiency, etc. between the task domain information and the available input/output modalities; and (4) design, implement and test the artifact. Today's interface designers have become highly skilled at performing these steps (non-sequentially, of course) using graphical user interfaces (GUIs) in combination with keyboard and mouse. However, we are still far from having solid scientific theory that may explain and evaluate current design practices even in the limited area of GUI/task domain information-mapping. Interactive interfaces increasingly incorporate spoken and written natural language, sound, touch, gesture and so on, in addition to new forms of graphical expression. The term *Modality Theory* seems apt for characterising attempts to theoretically address the corresponding information-mapping problem in its general form, i.e.:

> *Given any particular class of task domain information which needs to be exchanged between user and system during task performance, identify the set of input/output modalities which constitute an optimal solution to the representation and exchange of that information.*

This mapping problem is a human-computer interaction or cognitive engineering problem rather than a software engineering problem. In the context of the ESPRIT

Basic Research project GRACE, the following research agenda for modality theory is being pursued:

1. To establish sound conceptual and taxonomic foundations for describing and analysing any particular type of unimodal or multimodal output representation relevant to human-computer interaction (HCI);
2. to create a conceptual framework for describing and analysing interactive computer interfaces;
3. to develop a practical methodology for applying the results of steps (1) and (2) above to the problem of information-mapping between work/task domains and human-computer interfaces in information systems design.

An ultimate objective is to use results in building computerised design tools for the support of interface usability engineering. Below follows a description of the three items on the proposed research agenda for modality theory with references to work in progress.

2. Concepts and Taxonomy of Output Modalities

The first item on the research agenda is to build a conceptually well-founded taxonomy of output modalities. There will be literally thousands of potentially useful modalities and combinations of modalities available to interface designers in the future. So the problem appears to be one of handling complexity on the basis of well-analysed elements. At least it seems difficult to envision any other viable approach. The hope is that once we have identified the elements and their basic properties we will be able to claim that, given any conceivable multimodal representation of information, we will have the apparatus required for analysing it. The question therefore becomes: what are the elements and what are their basic properties?

Table 1. The full set of permutations on the taxonomy. The 9 labelled columns represent the properties of analogueness and arbitrariness and their opposites, static and dynamic, and graphics, sound and touch. The 12 rows in dark shading are necessarily empty, except for the problem of touch. Rows a-f are empty because analogue representations cannot be arbitrary. Rows h-j, n-o and t-u are empty because of the dynamic character of sound and touch representations (see text, however). The table shows how the remaining 12 rows contain all the (21) pure generic modalities. / between two numbered modalities indicates that the difference between them is based on prototypes. Numbered modalities in boldface are linguistic modalities.

		an	-an	ar	-ar	stat	dyn	gra	sou	tou
a		x		x		x		x		
b		x		x		x			x	
c		x		x		x				x
d		x		x			x	x		
e		x		x			x		x	
f		x		x			x			x
g	5/6,**16**	x			x	x		x		
h		x			x	x			x	

i	12/19, **20**	x			x	x				x
j	8/9	x			x		x	x		
k	3/18,**15**	x			x		x		x	
l	12/19,**20**	x			x		x			x
m	7		x	x		x		x		
n			x	x		x			x	
o	13		x	x		x				x
p	10		x	x			x	x		
q	4		x	x			x		x	
r	13		x	x			x			x
s	**2**,11a		x		x	x		x		
t			x		x	x			x	
u	**14**		x		x	x				x
v	11b,**17**		x		x		x	x		
w	**1**		x		x		x		x	
x	**14**		x		x		x			x

1 = Spoken language.
2 = Written language.
3 = Real sound.
4 = Arbitrary sound.
5 = Diagram pictures.
6 = Non-diagram pictures.
7 = Arbitrary diagrams.
8 = Animated diagram pictures.
9 = Dynamic pictures.
10 = Animated arbitrary diagrams.
11a = Static graphs.
11b = Dynamic graphs.
12 = Real touch.
13 = Arbitrary touch.
14 = Touch language.
15 = Analogue spoken language (onomatopoetica).
16 = Analogue written language (hieroglyphs).
17 = Dynamic written natural language.
18 = Diagrammatic sound.
19 = Diagrammatic touch.
20 = Analogue touch language.

In a first cut at this problem, we have identified a limited number (21 in fact) of pure generic modalities drawn from three different media of expression. Table 1 shows the taxonomy of pure generic modalities drawn from (Bernsen 1993a). The duplication of touch modalities in the table reflects a so far unsolved problem about the static/dynamic character of touch. *Purity* means that we are dealing with unimodal in contrast to multimodal modalities in order to prepare the handling of multimodal complexity on the basis of a limited number of elements. *Genericity* means that each of these modalities have a number of different (unimodal) *types* subsumed under it. Spoken language, for instance, is one such generic modality which has well-known types such as spoken letters, words, numerals, other spoken language related sounds, text or lists. A *medium of expression* is a set of perceptual qualities and the corresponding sensory equipment needed for perceiving them, such as, e.g., the set of visual and graphical qualities/vision. What is interesting about this approach is that we seem to arrive at a reasonably exclusive and exhaustive taxonomy of pure generic modalities when we characterise each of our 21 pure generic modalities as either possessing or not possessing each of a small number of basic properties. The result is the following working definition of a pure generic modality: A pure generic modality is characterised by a specific *medium of expression* and what may be termed a *profile* constituted by its characteristics as selected from the following list of binary opposites: analogue/non-analogue, arbitrary/non-arbitrary, static/dynamic, linguistic/non-linguistic. This working definition may be extensible to modalities in general (John Lee, personal communication). The hypothesis is that we may have gone a long way towards completing Agenda Item 1 of modality theory when we have provided an in-depth analysis of those basic properties. Such an analysis would show, e.g., what the medium of graphics/vision can contribute to the expression of information. First steps towards basic property analysis have been taken in (Bernsen 1993b).

There is more work to do, however. We still have to descend to the level of (unimodal) generic modality *types* to analyse the information representation capabilities of types. As illustrated above, there are many different types subsumed under each generic modality and it might seem that the problem space is about to explode after all, which we want to avoid at all cost. Another interesting point, therefore, is that there seems to be a surprising amount of regularity among the one hundred or more individual types subsumed under the entire set of pure generic modalities. Lists, or icons, for instance, can be found under virtually every pure generic modality. I call such recurring phenomena *modality structures* (Bernsen 1993a). The hypothesis is that the types under each generic modality can be divided neatly into (1) one *atom* which is a typical or representative type-instance of the individual types subsumed under that generic modality, and (2) recurring modality structures. In other words, if we analyse the basic properties of pure generic modalities, the atomic types of pure generic modalities and the set of modality structures, we may have exhausted the first Agenda Item of modality theory. What remains is to be able to derive the properties of multimodal representations composed of these elements (May 1993a,b,c,d).

3. Analysing Interaction

Human-computer interfaces not only consist of individual output representations such as lines of text. They also consist of groupings and series of output representations such as entire graphical screens and series of such confronting the user during task performance (Hovy & Arens 1990). In addition, interactive interfaces also involve *input* modalities some of which, such as the mouse or (for the time being) gesture, only exist as input modalities and not also as output modalities. To implement Item 2 on the research agenda of modality theory, therefore, we need conceptual and taxonomic work on groupings and series of output representations, on input representations and on relevant aspects of interactive tasks performed on computers. Addressing these issues will form part of our work during the second year of GRACE.

4. The Information Mapping Problem

Agenda Items 1 and 2 of modality theory constitute contributions from basic science to HCI and in particular to the usability engineering of interactive human-computer interfaces. These contributions need to be packaged in such a way that they can be applied in and integrated into early interface design practice. A rough description of the interface designer's task was provided in the Introduction above. The question now is how the theoretical contributions of modality theory can be practically applied through mapping task domain information into optimal interface modalities. The main problem, once again, is one of handling complexity. Information systems and their interfaces are being designed for a multitude of widely different real life task domains and each artifact normally allows the user to perform many different tasks using it. There is no way that this complexity can be fully anticipated by theory. What we need is an operational method for performing information-mapping between task domain information and interface modalities. We have begun to specify such a method and test it on simple cases (Bernsen & Bertels 1993). The methodology proceeds in five steps.

In *Step 1* the goal is to identify the information to be exchanged by user and system during task performance in the selected application domain. Often, a central part of the information needed for solving an information-mapping problem is information on users' tasks (but sometimes part of the information-mapping problem can be solved even earlier in the design process). Since we cannot examine in depth each and every possible task involving the artifact to be designed, it is necessary to be selective as to the tasks considered. The ideal way to be selective is to identify a limited set of *representative* tasks or scenarios to be performed by using the intended artifact and carry out the information-mapping analysis on these. The problem is that no guaranteed method for generating an appropriate set of scenarios currently exists in HCI or usability engineering. One proposed method (Carroll et al. 1992) is to weak for our purposes and we are currently testing an alternative method (Klausen & Bernsen 1993). Let us simply assume that the best current methods or heuristics are being applied for the purpose of identifying representative tasks. The result of Step 1 is a small set of representative tasks which users should be able to carry out using the intended artifact. This result constitutes an operationalisation of the information-mapping task.

In *Step 2* the representative tasks are individually analysed in as much detail as possible in order to identify their goals and initial states, the activities and procedures involved, the task (work) environment, the intended users and their experience, etc. The analysis should primarily aim at revealing the input/output information representation needs of the tasks. However, there is no need to explicitly represent all the information gathered in this Step. Step 2 is closely related to Step 3:

In *Step 3* the information relevant to information-mapping and acquired through Step 2 is represented explicitly and succinctly, for instance using the DSD (Design Space Development) notation for representing developing design space structure (Bernsen 1993c,d). In principle, this representation should contain everything which is relevant to the input/output modality choices to be subsequently made. The conceptual apparatus and terminology used in the representation should be the same as the one described in Step 4 below. Step 3 serves the purpose of making explicit the requirements for interactive information to be satisfied by the interface to the intended artifact. Step 3 thus concludes the first main phase of the methodology.

In *Step 4* we consider and apply the theoretically developed framework for representing the components of interactive unimodal or multimodal interfaces, i.e., the results of research on Agenda Items 1 and 2 of modality theory. The framework should eventually contain the elements and properties needed for analysing any specific type of unimodal or multimodal input or output representation. We then perform the mapping of the results of Step 3 into the elements and properties of Step 4. The result will be sets of possible input/output modalities and modality combinations which might be capable of representing the information needed for the representative tasks. It is likely that the mapping will produce a number of alternative solutions which subsequently have to be judged and compared to identify the optimal set of input/output modalities for the representative tasks.

In Step 5 a 'higher level filtering' is performed in order to trade off potential solutions against one another given the results of Steps (1) to (4) above. The result of Step 5 is a solution to the task domain/interface mapping problem. In some cases, several solutions can be expected to come out of the trade-off process with identical evaluations from a usability engineering point of view.

References

Bernsen, N.O. (1993a). Foundations of multimodal representations: A taxonomy of pure modalities. To appear in *Interacting with Computers*.

Bernsen, N.O. (1993b). Specificity and focus. Two complementary aspects of analogue graphics and natural language. *Esprit Basic Research project GRACE Deliverable* 2.1.3, 1993.

Bernsen, N.O. (1993c). Structuring design spaces. *INTERCHI '93 Adjunct Proceedings*. pp. 211-12, 1993.

Bernsen, N.O. (1993d). The structure of the design space. In Byerley, P.F., Barnard, P.J. and May, J. (Eds.): *Computers, Communication and Usability: Design issues, research and methods for integrated services.* Amsterdam, North-Holland, 1993, 221-244.

Bernsen, N.O. & Bertels, A. (1993). A methodology for mapping information from task domains to interactive modalities. *Esprit Basic Research project GRACE Deliverable* 10.1.3, 1993.

Carroll, J. M., Kellogg, W. A. & Rosson, M. B. (1992). Getting around the task - artifact cycle: How to make claims and design by scenario. *ACM Transactions on Information Systems,* Vol. 10, No. 2, 181-212.

Hovy, E. & Arens, Y. (1990). When is a picture worth a thousand words? Allocation of modalities in multimedia communication. Paper presented at the *AAAI Symposium on Human-Computer Interfaces,* Stanford.

Klausen, T. & Bernsen, N.O. (1993). CO-SITUE: Towards a methodology for constructing scenarios. In Hollnagel, E. and Lind, M. (Eds.): *Proceedings of the Fourth European Meeting on Cognitive Science Approaches to Process Control (CSAPC '93): Designing for Simplicity*. Copenhagen, August 1993, 1-16.

May, M. (1993a). Representations and homomorphisms. A taxonomy of representations for HCI, part 1. *Esprit Basic Research project GRACE Deliverable* 2.1.5, 1993.

May, M. (1993b). Levels of representations and mappings. A taxonomy of representations for HCI, part 2. *Esprit Basic Research project GRACE Deliverable* 2.1.6, 1993.

May, M. (1993c). From semantic types to multimodal presentations: Case-studies. A taxonomy of representations for HCI, part 3. *Esprit Basic Research project GRACE Deliverable* 2.1.7, 1993.

May, M. (1993d). Taxonomy and levels of mappings in the construction of multimodal interfaces. In Cox, R., Petre, M., Brna, P. and Lee, J., Eds. *Proceedings of the Workshop on Graphical Representations, Reasoning and Communication.* World Conference on Artificial Intelligence in Education, Edinburgh, August 1993.

Acknowledgement: The work described in this paper was carried out under Esprit Basic Research project 6296 GRACE whose support is gratefully acknowledged.

List of publications in the series "Working Papers in Cognitive Science"

1991

91-1 *Kognitionsforskning ved AUC*. Per Hasle.
91-2 *Johnson-Laird's Theory of Mental Models*. Niels Ole Bernsen.
91-3 *Potential Flashbulbs: Memories of Ordinary News as the Baseline*. Steen F. Larsen.
91-4 *"Icons Make me Click" - The Problem of Semantics*. Ib Ulbæk.
91-5 *Mental Models in Human-Computer Interaction*. Niels Ole Bernsen.
91-6 *Two Games in Town. Systematicity in Distributed Connectionist Systems*. Niels Ole Bernsen and Ib Ulbæk.
91-7 *Bevidsthedsfilosofi og kognitionsforskning*. Ib Ulbæk.

1992

92-1 *Matching Information and Interface Modalities. An Example Study*. Niels Ole Bernsen.
92-2 *Systematicity, Thought and Attention in a Distributed Connectionist System*. Niels Ole Bernsen and Ib Ulbæk.
92-3 *Communication Failure and Mental Models*. Brynja Svane and Niels Ole Bernsen.
92-4 *Kognitiv semantik og diagrammatisk tænken: fra intuition til formalisering*. Ph.D. Thesis. Michael May.
92-5 *Design of a Spoken Language Dialogue System. A Study of the Initial Specification Phase*. Niels Ole Bernsen.
92-6 *Learning the time constants of a Dynamic Task*. Bernd Brehmer, Leif Løvborg, and Anders Winman.
92-7 *Cost Information in a Dynamic Decision Task*. Bernd Brehmer, Leif Løvborg, and Anders Winman.
92-8 *New Facilities in NEWFIRE*. Leif Løvborg and Bernd Brehmer.
92-9 *C.S. Peirce on the Possibility of a Temporal Logic*. Peter Øhrstrøm.
92-10 *Hypertid*. Peter Bøgh Andersen og Peter Øhrstrøm.

1993

93-1 *A logical approach to handling uncertainty*. Morten Elvang-Gøransson, Paul Krause, John Fox.
93-3 *Den kognitive revolution*. Niels Ole Bernsen.
93-4 *Systematicity in the vision to language chain*. Niels Ole Bernsen.
93-5 *The structure of the design space*. Niels Ole Bernsen.
93-6 *Errors in a team context.* James Reason and Niels Ole Bernsen.
93-7 *Talking to a Wizard: report from the design of a natural speech understanding system*. Tove Klausen.
93-8 *Eksempler på samspillet mellem simultane og sekventielle kognitive processer*. John Paulin Hansen.

93-9 *Foundations of Multimodal Representations. A Taxonomy of Pure Modalities*. Niels Ole Bernsen.

93-10 *A methodology for mapping information from task domains to interactive modalities*. Niels Ole Bernsen.

93-11 *A research agenda for modality theory.* Niels Ole Bernsen.

93-12 *Types of user probems in design.* Niels Ole Bernsen.