A TAXONOMY OF INPUT MODALITIES

Niels Ole Bernsen, Centre for Cognitive Science, Roskilde University

Summary: Advances in information technologies are producing a very large number of possible interface modality combinations which are potentially useful for the expression and exchange of information in human-computer interaction. However, a systematic basis for analysing arbitrary input/output modality types and their multimodal combinations as to their capabilities of information representation and exchange is still lacking. Such a basis would enable interface designers of multimodal systems to select a appropriate sets of interface modalities once the information representation and exchange requirements of an application have been specified. This exploratory paper presents a first generation of a taxonomy of unimodal input modalities within the framework of modality theory. The generation follows the principles which have been used in generation of a taxonomy and theory of output modalities. In order to take into account some major asymmetries between the domains of output and input modalities, it has been necessary to include a new medium of expression, i.e. kinaesthetics, in addition to the media of graphics, acoustics and haptics which appear jointly sufficient to chart the domain of output modalities. The paper discusses the next steps to be taken towards a taxonomy and theory of input modalities which can be used to support practical interface design.

Keywords: Interface modalities, input, multimodal systems, multimedia systems, virtual reality, taxonomy, usability engineering, modality theory.

1. INTRODUCTION

This exploratory paper presents a first step towards a principled understanding of the space of possible input modalities in human-computer interaction (HCI). Thousands of different combinations of input, output and input/output modalities for the representation and exchange of information between humans and machines are currently becoming available to designers of interfaces for human-computer interaction, from unimodal spoken language input to complete multimodal virtual reality interactive systems. However, whereas the enabling technologies for multimodal (including virtual reality) interaction are growing rapidly, there is a lack of theoretical understanding of the principles which should be observed in mapping information from some task domain onto sets of input/output modalities at the human-computer interface in a way which optimises the usability and naturalness of the interface, given the specific purposes of the artifact. To achieve at least part of this understanding, it appears, the following objectives should be pursued:

1. To establish a taxonomy and related concepts (i.e. a theory) of the unimodal modalities which go into the creation of multimodal output representations for HCI; this should enable
2. the analysis, based on sound theoretical foundations, of the information representation and exchange capabilities of any particular unimodal or multimodal output representation relevant to HCI;
3. to establish sound foundations for analysing input modalities and entire interactive computer interfaces as to their capabilities for the representation and exchange of information;
4. to develop a methodology for applying the results of steps (1) to (3) above to the analysis of the problems of information mapping between work/task domains and interactive human-computer interfaces in information systems design;
5. to use the results of steps (1) to (4) above in building, possibly automated, interface design support tools.

The objectives just mentioned form the research agenda of modality theory which addresses the following, general information mapping problem: *Given any particular set of information which needs to be exchanged between user and system during task performance in context, identify the* *input/output modalities which constitute an optimal solution to the representation and exchange of that information* (Bernsen 1993).

In the present paper, as in modality theory generally, the term 'modality' means "mode or way of representing information to humans or machines in a physically realised intersubjective form". Such representations are *external* representations as distinguished from the (non-intersubjective) *internal* representations of the human (or machine) cognitive system. Representational modalities, in this sense, should not be confused with the sensory modalities of psychology. The reason why representational modalities are external and intersubjective is that they are physically realised in one or more of the *media* of graphics, acoustics, haptics and kinaesthetics. Although the term 'modality' is being used in confusingly different ways in the literature on multimodal systems, the notion of representational modalities just introduced is probably close to that intended by many authors. Hovy and Arens (1990), for instance, observe that, e.g., tables, beeps, written and spoken natural language may all be termed ‘modalities’. Modality theory aims to achieve a systematic and comprehensive understanding of (information representation) modalities such as those exemplified by Hovy and Arens. To do so, the theory makes the following two basic assumptions: (1) that modalities can be unimodal or multimodal and (2) that multimodal modalities are combinations of unimodal modalities, i.e. can be uniquely defined in terms of unimodal modalities. These assumptions mean that if we want to adopt a principled approach to the analysis of multimodal input/output representations, we have to start by generating and analysing *unimodal* representations. When these have been generated and analysed, a new type of generation can begin, namely the generation-through-composition of all possible *multimodal* input/output combinations in the media covered by the taxonomy.

Work is in progress on all objectives on the research agenda of modality theory. After a series of multimodal systems design case studies (Bernsen and Bertels 1993, Verjans and Bernsen 1994, Bernsen and Verjans 1995), an information mapping methodology (objective 4) has been defined and a first test performed in an industrial multimodal systems design project (Bernsen et al. 1995). Version 2 of a hypermedia modality analysis and design support software (objective 5) has been developed to reveal the problems involved in automating the information mapping metodology (Lu and Bernsen 1995).

This paper deals exclusively with part of objective (3) of the research agenda of modality theory described above. The proposed approach is based on results obtained on output modalities (objectives 1 and 2). A taxonomy of all possible unimodal *output* modalities in the media of graphics, acoustics and haptics has been systematically generated, and each modality has been analysed to identify its capabilities and limitations of information representation in task context (Bernsen 1994a, 1994b,1995). The result is a "designer's toolbox" of unimodal output modalities for use in interface design. Section 2 provides a brief overview of the theory. Section 3 discusses how to apply to input modality generation the principles that were used in generating the taxonomy of output modalities. Section 4 presents the generation of unimodal modalities at the generic level of abstraction. Section 5 discusses the next steps which must be taken to arrive at a practically useful theory of input modalities.

2. THE TAXONOMY OF OUTPUT MODALITIES AND MODALITY THEORY

The taxonomy of output modalities generates a series of unimodal output representations or modalities from combinations of basic properties. The basic properties are the properties of being either *linguistic or non-linguistic, analogue or non-analogue, arbitrary or non-arbitrary, static or dynamic*, and of being physically realised as an external representation in one of the three media of *graphics, acoustics or haptics.* The basic assumption is that the presence or absence of each of these properties in a particular modality of representation, makes important differences to that representation's suitability for representing information in a particular task context. The issue of why the media of graphics, acoustics and haptics were chosen for the purpose of unimodal output modality generation, will be addressed in Section 3.

Mechanical combination of basic properties yields 48 unimodal modalities each of which is uniquely defined through the basic properties constituting it.

SUPER LEVEL CLASSES	GENERIC UNIMODAL LEVEL	NOTATION
I. Linguistic modalities	**1. Static analogue graphic language**	**<li,an,-ar,sta,gra>**
	2. Static analogue acoustic language Dynamic analogue acoustic language	**<li,an,-ar,sta/dyn,aco>**
	3. Static analogue haptic language Dynamic analogue haptic language	**<li,an,-ar,sta/dyn,hap>**
<li,-an,-ar>	**4. Dynamic analogue graphic language**	**<li,an,-ar,dyn,gra>**
	5. Static non-analogue graphic language	**<li,-an,-ar,sta,gra>**
	6. Static non-analogue acoustic language Dynamic non-analogue acoustic language	**<li,-an,-ar,sta/dyn,aco>**
	7. Static non-analogue haptic language Dynamic non-analogue haptic language	**<li,-an,-ar,sta/dyn,hap>**
	8. Dynamic non-analogue graphic language	**<li,-an,-ar,dyn,gra>**
II. Analogue modalities	**9. Static analogue graphics**	**<-li,an,-ar,sta,gra>**
	10. Static analogue acoustics Dynamic analogue acoustics	**<-li,an,-ar,sta/dyn,aco>**
	11. Static analogue haptics Dynamic analogue haptics	**<-li,an,-ar,sta/dyn,hap>**
<-li,an,-ar>	**12. Dynamic analogue graphics**	**<-li,an,-ar,dyn,gra>**
III. Arbitrary modalities	**13. Arbitrary static graphics**	**<-li,-an,ar,sta,gra>**
	14. Arbitrary static acoustics Dynamic arbitrary acoustics	**<-li,-an,ar,sta/dyn,aco>**
	15. Arbitrary static haptics Dynamic arbitrary haptics	**<-li,-an,ar,sta/dyn,hap>**
<-li,-an,ar>	**16. Dynamic arbitrary graphics**	**<-li,-an,ar,dyn,gra>**
IV. Explicit modality structures	**17. Static graphic structures**	**<-li,-an,-ar,sta,gra>**
	18. Static acoustic structures Dynamic acoustic structures	**<-li,-an,-ar,sta/dyn, aco>**
	19. Static haptic structures Dynamic haptic structures	**<-li,-an,-ar,sta/dyn,hap>**
<-li,-an,-ar>	**20. Dynamic graphic structures**	**<-li,-an,-ar,dyn,gra>**
SUPER LEVEL CLASSES	**GENERIC UNIMODAL LEVEL**	**NOTATION**

Table 1. The 20 generic unimodal output modalities resulting from the reductive operations described in the text.

The list of 48 unimodal modalities is subsequently reduced based on considerations of (a) the purpose of modality theory, which is to support the optimisation of clear and unambiguous information representation, and (b) the purely pragmatic viewpoint that the theory should focus on unimodal modalities of significance to current human-computer interface design. The result is 20 unimodal modalities which, subject to the reductions described in (a) and (b) above, exhaustively cover possible output representations of information in the three *media* of graphics, acoustics or haptics (see Table 1). More details can be found in (Bernsen 1994a,b).

Exhaustive coverage, however, is not enough for the taxonomy to serve as a foundation of information mapping in the design of multimodal (output) interfaces. Table 1 shows the unimodal modalities as characterised hierarchically at two descending levels of abstraction, i.e. the *Super Level* and the *Generic Level*. Descent from the super level to the generic level happens by adding distinctions in terms of basic properties. As it turns out, the generic level is still too abstract to allow distinction between a sufficient number of unimodal output modalities important to interface design. In other words, modalities as described at the generic level still lack the properties of intuitive simplicity and practical usability which must be required of a "designer's toolbox" of unimodal representational modalities from which the large space of useful multimodal output representations may be generated. Resolution of this problem requires new distinctions to be made based on novel sets of basic properties. The result is an additional level of the taxonomy which is called the *Atomic Level* (see Figure 1). More details can be found in (Bernsen 1995).

The step from a *taxonomy* of output modalities generated from first principles to a *theory* of output modalities requires a thorough analysis of all the unimodal modalities in the taxonomy with a view to identifying those of their properties which are important to information mapping in practical interface design. Partial results of modality analysis can be found in Bernsen

and Lu (1995), Lu and Bernsen (1995) and Bernsen (1995). A comprehensive presentation is in preparation.

3. FROM OUTPUT TAXONOMY TO INPUT TAXONOMY

Several points need discussion before turning to the generation of a first taxonomy of input modalities. These points are addressed in the present section which deals with basic similarities and asymmetries between output and input modalities, the relationship between modality theory and input/output devices, and the very sense of the terms 'input' and 'output'.

Input modalities are forms of information representation communicated to the system (or computer) by its users. In principle, input modalities would seem to be as many and as diverse as output modalities. We have come a long way from the character-based interfaces using typed language notation and are beginning to see the practical use of input modalities such as spoken language, eye-tracking, camera-recorded gesture and 3D analogue bodily mo-

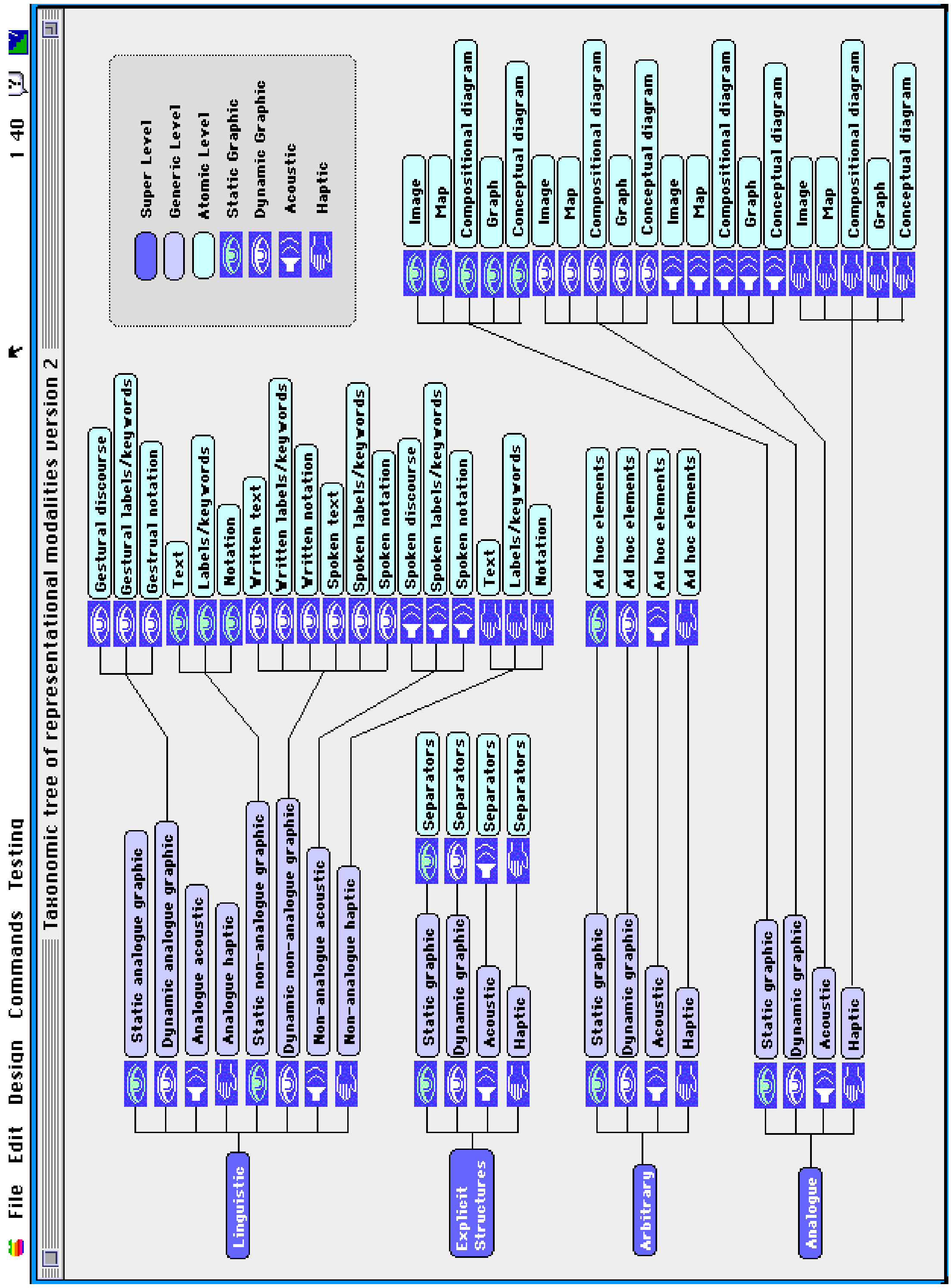

Figure 1. The super level, generic level and atomic level of the taxonomy of unimodal output modalities.

vement for virtual reality applications. However, one notes at once that several of these examples of input modalities do not have equivalents in the output domain. Spoken language can be used both as input and output. But whereas we may control the computer through eye-tracking, the computer does not look at us, through cameras or otherwise, in any similar way, or at least does not do so in current applications. And when it comes to gesture or bodily movement, it seems clear that current computers do not communicate information to us in these ways. Robots are normally built for purposes quite different from that of gestural or bodily communication with humans. And if we want the machine to communicate to us through gesture, it is much easier to implement such communication in graphics rather than through equipping the machine with limbs. For these reasons, output modality theory does not (yet) cover the medium of computer bodily gesture but covers only the media of graphics, acoustics and haptics in which all known forms of output representation would appear to take place. The conclusion is that *there is at best only partial overlap between unimodal input and output modalities,* for the simple reason that an entire medium for the physical realisation of information representation, i.e. bodily movement, is absent from the output domain but present in the input domain.

To the extent that there actually is overlap between output and input modalities, this overlap may be exploited through transfer of results from output modality theory to input modality theory. Such overlap does seem to obtain in the media of expression of graphics and acoustics. When, for instance, the machine observes its environment, including the user, through devices such as a camera and a microphone, it perceives the world in the graphic and acoustic medium, respectively. The user is graphics to the machine's camera and related software programs, and acoustics to the machine's microphone and related software programs. In haptics, on the other hand, there is no overlap between output and input modalities. Haptic output does seem to hold important potential which for technological reasons has remained to a large extent unexplored (Bernsen 1995). It seems unclear, on the other hand, whether haptic input will ever be of significant interest to interface design. This notwithstanding, machines decoding Braille text through their limbs or receiving electrical impulse alarm signals from their users are fascinating to think of.

Given that overlap between output and input modalities obtains in the media of graphics and acoustics, it can be concluded that *the types of information representation in the media of graphics and acoustics*, as charted in the taxonomy of output modalities and analysed in modality analysis, *remain the same for input and output.* This is guaranteed by the exhaustiveness of the taxonomy of unimodal output representations at the levels of abstraction at which it operates. This further implies that *the distinction between unimodal and multimodal representations is valid for input as well as for output, at least in the media of graphics and acoustics,* haptics having been excluded from consideration above. Conversely, input and output modalities do not overlap in the domain of bodily movement which is important to input but not to output interface design, nor in the domain of haptics, which is important to output but not to input interface design.

Let us note an important *caveat* to the reasoning on graphics and acoustics above. It is that machines do not always use sensory devices which have any direct correspondence in the human perceptual apparatus. It is true that the machine's camera(s) roughly corresponds to the eyes of humans, that the machine's microphone(s) corresponds to the ears of humans, and that appropriate programming can make machines partly correspond to humans equipped with vision and hearing. However, in systems design we are free to use whatever input/output devices may serve useful interface purposes. In other words, *the taxonomy of input modalities should be clearly distinguished from a typology of input devices,* just like the taxonomy of output modalities is distinct from a typology of output devices. We are addressing forms of human-to-machine input representation of information irrespective of how these forms, or modalities, map onto the devices used for inputting them. A mouse, for instance, is not an interface modality but an input device which is similar in functionality to several other input devices, such as trackerballs and some joysticks.

It is crucial to note that *a taxonomy and theory of input modalities,* just like the taxonomy and theory of output modalities, *deals with exchange of information between user and system.* User and system communicate, or conduct a dialogue, during such exchanges. It is not the case that each of them merely obtains information on the other, as when a human diagnoses or repairs a fault in the machine or when the machine obtains information on a person through a surveillance camera, a hidden microphone or through other sensors which may even be tactile ones. In other words, input, in the sense with which modality theory is concerned, is information which is being interactively communicated from user to machine, just as output is information interactively conveyed from machine to user. Normally, when a user interacts with a machine in the sense of human-computer interaction, the user is engaged in interactively performing some task, a task being a more or less ordered series of human activities towards the achievement of a certain goal. The machine has been designed to participate in solving the task. So, we are dealing with task-oriented exchange of information in which the task is being shared between human(s) and machine(s). The reason why this is important is that it allows us to exclude from consideration all non-interactive, non-task-oriented relationships between humans and machines, even if these relationships involve the transfer of information from human to machine or from machine to human. When a user does not provide the machine with deliberate, task-oriented input, then, whatever information the machine may sense, perceive, register or otherwise obtain about the user, this information is not input in the sense of human-computer interaction or the taxonomy of input modalities in particular.

	li	-li	an	-an	ar	-ar	sta	dyn	gra	aco	kin
1	x		x		x		x		x		
2	x		x		x		x			x	
3	x		x		x		x				x
4	x		x		x			x	x		
5	x		x		x			x		x	
6	x		x		x			x			x
7	x		x			x	x		x		
8	x		x			x	x			x	
9	x		x			x	x				x
10	x		x			x		x	x		
11	x		x			x		x		x	
12	x		x			x		x			x
13	x			x	x		x		x		
14	x			x	x		x			x	
15	x			x	x		x				x
16	x			x	x			x	x		
17	x			x	x			x		x	
18	x			x	x			x			x
19	x			x		x	x		x		
20	x			x		x	x			x	
21	x			x		x	x				x
22	x			x		x		x	x		
23	x			x		x		x		x	
24	x			x		x		x			x
25		x	x		x		x		x		
26		x	x		x		x			x	
27		x	x		x		x				x
28		x	x		x			x	x		
29		x	x		x			x		x	
30		x	x		x			x			x
31		x	x			x	x		x		
32		x	x			x	x			x	
33		x	x			x	x				x
34		x	x			x		x	x		
35		x	x			x		x		x	
36		x	x			x		x			x
37		x		x	x		x		x		

38		x		x	x		x			x	
39		x		x	x		x				x
40		x		x	x			x	x		
41		x		x	x			x		x	
42		x		x	x			x			x
43		x		x		x	x		x		
44		x		x		x	x			x	
45		x		x		x	x				x
46		x		x		x		x	x		
47		x		x		x		x		x	
48		x		x		x		x			x
	li	-li	an	-an	ar	-ar	sta	dyn	gra	aco	kin

Table 2. The full set of 48 combinations of basic properties constituting the possible modalities at the generic level of the input taxonomy. All modalities provide possible ways of representing information.

4. A FIRST TAXONOMY OF INPUT MODALITIES

We have seen that input and output modalities are identical as far as the media of graphics and acoustics are concerned. I propose to call the medium of bodily movement the *kinaesthetic* medium. Thus, when humans interact with machines, their behaviour is regarded by the machine as representing information in one or more of the media of graphics, acoustics and kinaesthetics. Apart from replacing output haptics by input kinaesthetics, we may use the same assumptions about basic properties as in output modality theory. Thus, interactive, human-machine, unimodal input modalities can be linguistic and non-linguistic, analogue and non-analogue, arbitrary and non-arbitrary, static or dynamic, and graphic, acoustic or kinaesthetic. This produces 2x2x2x2x3=48 basic property combinations, as shown in Table 2.

	li	-li	an	-an	ar	-ar	sta	dyn	gra	aco	kin
1	x		x			x	x		x		
2	x		x			x	x			x	
3	x		x			x	x				x
4	x		x			x		x	x		
5	x		x			x		x		x	
6	x		x			x		x			x
7	x			x		x	x		x		
8	x			x		x	x			x	
9	x			x		x	x				x
10	x			x		x		x	x		
11	x			x		x		x		x	
12	x			x		x		x			x
13		x	x			x	x		x		
14		x	x			x	x			x	
15		x	x			x	x				x
16		x	x			x		x	x		
17		x	x			x		x		x	
18		x	x			x		x			x
19		x		x	x		x		x		
20		x		x	x		x			x	
21		x		x	x		x				x
22		x		x	x			x	x		
23		x		x	x			x		x	
24		x		x	x			x			x
25		x		x		x	x		x		
26		x		x		x	x			x	

27		x		x		x	x				x
28		x		x		x		x	x		
29		x		x		x		x		x	
30		x		x		x		x			x
	li	**-li**	**an**	**-an**	**ar**	**-ar**	**sta**	**dyn**	**gra**	**aco**	**kin**

Table 3. The 30 combinations of basic properties which remain after removal from Table 2 of all combinations in which non-arbitrary expressions of information are being used arbitrarily.

Among the input modalities defined through combinations of basic properties in Table 2, we need to remove 18 modalities that contradict the purpose of the taxonomy. The purpose of the taxonomy is to support the optimisation of clear and unambiguous information representation. We therefore need to remove all modalities which are defined by the arbitrary use of non- arbitrary expressions. We do not want, for instance, to design interfaces

SUPER LEVEL CLASSES	GENERIC UNIMODAL LEVEL	NOTATION
I. Linguistic modalities	**1. Static analogue sign graphic language**	**<li,an,-ar,sta,gra>**
	2. Static analogue sign acoustic language	**<li,an,-ar,sta,aco>**
	3. Static analogue sign kinaesthetic language	**<li,an,-ar,sta,kin>**
<li,-an,-ar>	**4. Dynamic analogue sign graphic language**	**<li,an,-ar,dyn,gra>**
	5. Dynamic analogue sign acoustic language	**<li,an,-ar,dyn,aco>**
	6. Dynamic analogue sign kinaesthetic language	**<li,an,-ar,dyn, kin>**
	7. Static non-analogue graphic language	**<li,-an,-ar,sta,gra>**
	8. Static non-analogue acoustic language	**<li,-an,-ar,sta,aco>**
	9. Static non-analogue kinaesthetic language	**<li,-an,-ar,sta, kin>**
	10. Dynamic non-analogue graphic language	**<li,-an,-ar,dyn,gra>**
	11. Dynamic non-analogue acoustic language	**<li,-an,-ar,dyn,aco>**
	12. Dynamic non-analogue kinaesthetic language	**<li,-an,-ar,dyn, kin>**
II. Analogue modalities	**13. Static analogue graphics**	**<-li,an,-ar,sta,gra>**
	14. Static analogue acoustics	**<-li,an,-ar,sta,aco>**
	15. Static analogue kinaesthetics	**<-li,an,-ar,sta, kin>**
<-li,an,-ar>	**16. Dynamic analogue graphics**	**<-li,an,-ar,dyn,gra>**
	17. Dynamic analogue acoustics	**<-li,an,-ar,dyn,aco>**
	18. Dynamic analogue kinaesthetics	**<-li,an,-ar,dyn, kin>**
III. Arbitrary modalities	**19. Arbitrary static graphics**	**<-li,-an,ar,sta,gra>**
	20. Arbitrary static acoustics	**<-li,-an,ar,sta,aco>**
	21. Arbitrary static kinaesthetics	**<-li,-an,ar,sta, kin>**
<-li,-an,ar>	**22. Dynamic arbitrary graphics**	**<-li,-an,ar,dyn,gra>**
	23. Dynamic arbitrary acoustics	**<-li,-an,ar,dyn,aco>**
	24. Dynamic arbitrary kinaesthetics	**<-li,-an,ar,dyn, kin>**
IV. Explicit modality structures	**25. Static graphic structures**	**<-li,-an,-ar,sta,gra>**
	26. Static acoustic structures	**<-li,-an,-ar,sta,aco>**
	27. Static kinaesthetic structures	**<-li,-an,-ar,sta, kin>**
<-li,-an,-ar>	**28. Dynamic graphic structures**	**<-li,-an,-ar,dyn,gra>**

	29. Dynamic acoustic structures	**<-li,-an,-ar,dyn,aco>**
	30. Dynamic kinaesthetic structures	**<-li,-an,-ar,dyn, kin>**
SUPER LEVEL CLASSES	**GENERIC UNIMODAL LEVEL**	**NOTATION**

Table 4. At the "generic" level of abstraction, modality theory generates 30 unimodal input modalities in the media of graphics, acoustics and kinaesthetics.

which require users to enter the term 'apple' when communicating the meaning of the term 'ship' to the machine, or show the computer a picture of a ship when communication the meaning of a picture of an apple. The arbitrary use of non-arbitrary expressions of information does have its use, such as for cryptographic purposes, but such use goes against the purpose of human-computer interface design. We must therefore remove from Table 2 rows 1-6, because linguistic *and* analogue representations should not be used arbitrarily, rows 13-18, because linguistic representations should not be used arbitrarily, and rows 25-30, because analogue representations should not be used arbitrarily. The (re-numbered) result of this reduction is shown in Table 3.

In Table 4, the modalities represented in Table 3 have been named and expressed in the notation of modality theory.

The generative result presented in Table 4 will be analysed in future work. As in the generation of output modalities, it may be expected that the additional generation of an "atomic" level of unimodal input modalities will be needed in order to arrive at a level of abstraction which is intuitively satisfactory in practical design (Bernsen 1995). One interesting challenge to the analysis of the result presented in Table 4, will be the analysis of the kinaesthetic modalities. Some of these are already evident from the table, such as typed input through use of devices such as the keyboard, which belongs to dynamic non-analogue kinaesthetic language (row 12). Another important point of interest will be the analysis of the input modalities which have counterparts in the output domain, i.e. the graphic and acoustic modalities. It may be expected that their relative importance as input modalities will be at times significantly different from their relative importance as output modalities. Finally, preliminary analysis of Table 4 suggests that pragmatic reduction (cf. Section 2) of the number of modalities will be a natural and feasible way to obtain a reduced set of atomic modalities.

5. CONCLUSION

This paper has presented a first generation of a taxonomy of input modalities expressed in the conceptual framework of modality theory. Asymmetries between output and input information, which are ultimately based in the different capabilities of humans and machines, have led to the introduction of a new medium of expression, i.e. that of kinaesthetics. To arrive at a principled taxonomy of input modalities, it has been necessary to perform a complete dissociation between input devices and input modalities. It has also been necessary to define the term 'input' more precisely than is customarily the case in HCI.

The next steps in the exploration of input modalities from the point of view of modality theory, will follow the steps taken in output modality analysis. Firstly, a large number of examples of input modalities will be preliminarily analysed. Secondly, based on this analysis, an atomic level of abstraction will be generated from the generic level of abstraction in the taxonomy. Thirdly, a proper modality analysis will be made of each unimodal input modality represented in the taxonomy. The results of modality analysis will be represented and illustrated in the taxonomy workbench (Lu and Bernsen 1995). Assuming that the work just described proceeds as planned, modality theory will eventually provide a principled foundation for addressing the problem of human-machine interaction in terms of multimodal combinations of unimodal input/output modalities.

REFERENCES

Bernsen, N.O.: A research agenda for modality theory. 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, 43-46.

Bernsen, N.O. (1994b): A revised generation of the taxonomy of output modalities. *Esprit Basic Research project AMODEUS Working Paper* RP5-TM-WP11, 1994.

Bernsen, N. O (1995) A toolbox of output modalities. Representing output information in multimodal interfaces. *Esprit Basic Research Project AMODEUS-2 Working Paper* RP5-TM-WP21.

Bernsen, N.O. (1994a): Bernsen, N.O.: Foundations of multimodal representations. A taxonomy of representational modalities. *Interacting with Computers* Vol. 6 No. 4, 1994, 347-71.

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

Bernsen, N.O., Jensager, F. and Lu, S.: Designing interfaces by information mapping. An industrial case study. *Esprit Basic Research project AMODEUS-2 Working Paper* RP5-TM-WP18 1995.

Bernsen, N.O. and Lu, S.: A software demonstrator of modality theory. In Bastide, R. and Palanque, P. (Eds.): *Proceedings of DSV-IS'95: Second Eurographics Workshop on Design, Specification and Verification of Interactive Systems.* Springer-Verlag 1995 (to appear).

Bernsen, N.O. and Verjans, S. (1995) Designing interfaces by information mapping. A case study of the CERD design. *Esprit Basic Research Project AMODEUS-2 Working Paper* RP5-TM-WP13.

Hovy, E. and Arens, Y.: 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 1990.

Lu, S. and Bernsen, N.O.: The taxonomy workbench. A multimedia database system for analysing representational modalities. In *Proceedings of the 4th International Conference on Interface to Real and Virtual Worlds,* Montpellier, June 1995 (to appear).

Verjans, S. and Bernsen, N.O.: PaTerm: A case study in information-mapping. *Esprit Basic Research project AMODEUS II Working Paper* RP5-TM-WP6, 1994.

Acknowledgements: The work described in this paper was carried out under Esprit Basic Research project 7040 AMODEUS-2 and under a grant from the Danish Government's Informatics Programme. The support is gratefully acknowledged. I want to thank the participants in the discussions which led to the result reported in this paper: Laila Dybkjær, Shijian Lu, Judith Ramsay and Steven Verjans.