

FOUNDATIONS OF MULTIMODAL REPRESENTATIONS

A Taxonomy of Representational Modalities

Niels Ole Bernsen, CCI, Risø National Laboratory and 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 principled basis for analysing arbitrary input/output modality types and combinations as to their capabilities of information representation and exchange is still lacking. The paper presents an generative approach to the analysis of output modality types and their combinations and takes some steps towards its implementation, departing from a taxonomy of generic unimodal modalities of representation. A small number of key properties appear sufficient for creating a taxonomy of generic output modalities which is both relatively simple, robust, intuitively plausible and reasonably complete. These (orthogonal) properties are: analogue and non-analogue representations; arbitrary and non-arbitrary representations; static and dynamic representations; linguistic and non-linguistic representations; different media of representation; and modality structure. The work presented is being located within the larger research agenda of modality theory.

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

1. INTRODUCTION

This paper presents a principled approach to the analysis of unimodal and multimodal output representations for usability engineering purposes. The work forms part of the ESPRIT Basic Research project GRACE which ultimately aims at providing a sound theoretical basis for usability engineering in the domain of multimodal representations. Whereas the enabling technologies for multimodal (including virtual reality) representation are growing rapidly, there is a lack of theoretical understanding of the principles which should be observed in mapping information from some task domain into presentations 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, listed in order of increasing complexity:

1. To establish a taxonomy and a set of related categorisations of the generic modalities which go into the creation of multimodal output representations for human-computer interaction (HCI); this should enable:
2. the establishment of sound foundations for describing and analysing any particular type of unimodal or multimodal output representation relevant to HCI;
3. to establish sound foundations for analysing input modalities and entire interactive computer interfaces;
4. to develop a methodology for applying the results of the steps above to the analysis of the problems of information-mapping and information-transformation between work/task domains and human-computer interfaces in information systems design;
5. to use, if possible, results of the work described in building design tools for the support of usability engineering.

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 1993a, 1994).

Throughout the work on taxonomy, categorisations and other conceptual apparatus the aim is to couch the analyses and their results in a clear, robust and useful terminology which can serve its purpose in usability engineering. At present, the terminology in the field or fields indicated above is confused and incoherent. The terminology proposed here is not claimed to be the ‘right’ one nor to be superior to any other. The aim is rather to propose a terminology which makes the important concepts and issues clear while not being overly complicated. A deeper point is that there may be also conceptual confusion in the way the field is currently addressed in the literature. It is hoped that the approach presented here will not increase conceptual confusion but rather help removing it.

The paper addresses the first and second objectives above in proposing a taxonomy of generic, unimodal representational *output* modalities (Sects. 2 and 3). The taxonomy builds on a small number of key properties which all express important dimensions of modality description and are analysed in Sect. 4. The analysis leads to characterisations of generic unimodal modalities and modalities in general (Sect. 5) and a discussion of the orthogonality and completeness of the taxonomy (Sect. 6). Sect. 7 discusses the issue of modality structure. The distinction between external and internal representations is crucial to the analysis of particular types of unimodal or multimodal representation and is discussed in Sect. 8. The concluding discussion (Sect. 9) argues that the taxonomy enables the specification of a principled approach to the analysis of any particular output modality type or combination of output modality types falling within the scope of this paper.

A key consideration in any attempt to address the issues dealt with here is the following. There are literally thousands of possible and potentially useful combinations of output modality types. Theory cannot and should not explicitly address all of these but should provide principles by which any given modality combination can be analysed when needed. *This calls for a generative approach from simple elements at the right levels of abstraction.* The simple elements themselves should be defined from a limited set of properties fundamental to the analysis of arbitrary representational modalities. So the core questions to be addressed in what follows are: What are the simple elements? What are the right levels of abstraction at which they should be identified? What are the fundamental properties defining the simple elements? Is the resulting generative taxonomy conceptually clear, robust and intuitively acceptable? A sufficient degree of intuitive acceptability is no doubt essential if the taxonomy is to be both useful to and used in actual interface design.

2. ON REPRESENTATION

The term *multimodal representation* designates combinations of two or more unimodal representational modalities for HCI purposes. Such representations are *external* to the human cognitive system. We are not here dealing with internal cognitive representations (see Sect. 8 below, however). External representations are considered as representations by the human cognitive system and are, as far as we are concerned, produced by data structures in computers and other items of information technology. The concept of an external representation implies a distinction between *what is represented* and its *representation*. In general, there is a one-to-many mapping relationship between what is represented and its possible representations. One and the same object, situation, event, process, set of data, procedural instruction, etc., can be represented in many different ways. Some of these representations may be better than others for given task goals, given also the information to be represented and the nature of human cognition. Moreover, in many cases representations do not allow a simple and universal decoding of what they represent but require additional knowledge for this to be possible with any degree of certainty or confidence. In general, there is a one-to-many mapping relationship between a representation and what it may represent. The additional knowledge which is needed for non-arbitrary interpretation is knowledge of the *mapping principles* between a representation and what is represented. Foreign spoken languages, unknown to us, are examples in point but so are many examples of graphical and other representations. The relationship between representations and what is to be represented is shown in Diagram 1. Diagram 2 in Sect. 8 below provides a less simplified representation.

What is to be represented<->mapping principles<->representations.

Diagram 1. Representation requires mapping principles. Diagram 1 is multimodal and is composed of typed static graphic language (natural language text and formal notation elements) and an explicit static graphic structure (the frame).

Someone, a designer, for instance, wants to represent something to, e.g., information technology users at the computer interface. The better the mapping principles between what is to be represented and its representation are known to the users in advance, the easier the communication to users will work and the less problems users will have in decoding the representations. The less fit there is between the mapping principles and users' knowledge, the more risk there is that users will misinterpret the representations or fail to understand them, and the more work there has to be done to somehow impart to users additional knowledge of the mapping principles involved. Unfortunately, however, mapping principles which fit the knowledge that users already have are neither necessary nor sufficient for securing optimal interface representations for given tasks.

3. A GENERATIVE TAXONOMY BASED ON GENERIC UNIMODAL OUTPUT MODALITIES

The term '(representational) modality' is being used in confusingly different ways in the literature. Let us proceed on the assumptions that (1), e.g., tables, beeps, written and spoken natural language may all be termed 'modalities' (Hovy and Arens 1990); (2) *representational modalities* should not be confused with the *sensory modalities* of psychology; (3) modalities can be unimodal or multimodal; (4) a taxonomy of modalities is a way of carving up the space of external representations of information based on the observation that different modalities have different properties which makes them suitable for representing different types of information. Given these assumptions, the taxonomy to be presented will produce a straightforward definition of modalities (Sect. 5).

The question to be addressed in this section is the following: what are the generic types of representational modalities in their unimodal or uncombined forms? For HCI purposes, combined representational forms are often more interesting because of the increase in expressive power that comes from combining different modalities. However, combined modalities or multimodal representations are combined from something, namely unimodal representations. I shall argue that if we want to adopt a principled approach to the analysis of multimodal representations, we have to start by analysing unimodal representations. The space of unimodal representations can be carved up at different levels of abstraction. A hierarchical generative taxonomy is proposed which has three levels, a *super level* of less principled significance, a basic *generic level* and an *atomic type level*. Starting with the super and generic levels ensures that the taxonomy is based on a limited set of generic unimodal modalities from which any given output modality or modality combination can be generated and analysed. Each generic modality has a number of actual or possible (and equally unimodal) *atomic modality types* subsumed under it which inherit its basic properties and have further discriminatory properties of their own.

Each generic unimodal modality is characterised by a small set of *basic features* which serve to robustly distinguish modalities from one another within the taxonomy. Each of these features have profound implications for a certain modality's capacity of representing information. The features are: *linguistic/non-linguistic*, *analogue/non-analogue*, *arbitrary/non-arbitrary*, *static/dynamic*. In addition, we distinguish between the *media of expression* of graphics, sound and touch which are each characterised by very different sets of perceptual qualities (visual, auditory and tactile, respectively). These media de

	modality	li	-li	an	-an	ar	-ar	sta	dyn	gra	sou	tou
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	1	x		x			x	x		x		
8		x		x			x	x			x	
9		x		x			x	x				x
10	2	x		x			x		x	x		
11	3	x		x			x		x		x	
12	4	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	5	x			x		x	x		x		
20		x			x		x	x			x	
21		x			x		x	x				x
22	6	x			x		x		x	x		
23	7	x			x		x		x		x	
24	8	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	9/10/11		x	x			x	x		x		
32			x	x			x	x			x	
33			x	x			x	x				x
34	12/13/14		x	x			x		x	x		
35	15/16/17		x	x			x		x		x	
36	18/19/20		x	x			x		x			x
37	21		x		x	x		x		x		
38			x		x	x		x			x	
39			x		x	x		x				x
40	22		x		x	x			x	x		
41	23		x		x	x			x		x	
42	24		x		x	x			x			x
43	25		x		x		x	x		x		
44			x		x		x	x			x	
45			x		x		x	x				x
46	26		x		x		x		x	x		
47	27		x		x		x		x		x	
48	28		x		x		x		x			x
	modality	li	-li	an	-an	ar	-ar	sta	dyn	gra	sou	tou

Table 1. The full set of (48) combinations in the taxonomy. The 28 rows in dark shading are empty, in some cases for several reasons: 1-6 and 25-30 because analogue representations cannot sensibly be used arbitrarily; 2-3, 8-9, 14-15, 20-21, 26-27, 32-33, 38-39 and 44-45 because sound and touch are dynamic; 1-6 and 13-18 because language is non-arbitrary. / between two modalities indicates that the difference between them is based either on prototypes or on the issue of data abstraction. Numbers in the Modalities column refer to the generic unimodal modalities named in Table 2.

termine the scope of the taxonomy. For instance, the same linguistic information may be represented in either the graphical, sound or touch medium but the choice of medium strongly influences the suitability of the representation for a given design purpose and is therefore considered a choice between different modalities.

A matrix of generic unimodal modalities distinguished according to the basic features above contains 48 feature combinations (2x2x2x2x3). These are presented in Table 1.

After removal of potential modalities which are not possible for one reason or another, we obtain the taxonomy presented in Table 2. 28 feature combinations were ruled out leaving 20 combinations comprising 28 generic unimodal modalities divided into 4 different classes at the super level, i.e. the primarily linguistic, the primarily analogue, the arbitrary and the explicit structures. The reasons, sometimes double, for ruling out combinations are straightforward. 12

modality	li	-li	an	-an	ar	-ar	sta	dyn	gra	sou	tou
1. Static analogue graphic language	x		x			x	x		x		
2. Dynamic analogue graphic language	x		x			x		x	x		
3. Analogue spoken language	x		x			x		x		x	
4. Analogue touch language	x		x			x		x			x
5. Static non-analogue graphic language	x			x		x	x		x		
6. Dynamic non-analogue graphic language	x			x		x		x	x		
7. Non-analogue spoken language	x			x		x		x		x	
8. Non-analogue touch language	x			x		x		x			x
9. Diagrammatic pictures		x	x			x	x		x		
10. Non-diagrammatic pictures											
11. Static graphs											
12. Animated diagram pictures		x	x			x		x	x		
13. Dynamic pictures											
14. Dynamic graphs											
15. Real sound		x	x			x		x		x	
16. Diagrammatic sound											
17. Sound graphs											
18. Real touch		x	x			x		x			x
19. Diagrammatic touch											
20. Touch graphs											
21. Arbitrary static diagrams		x		x	x		x		x		
22. Animated arbitrary diagrams		x		x	x			x	x		
23. Arbitrary sound		x		x	x			x		x	
24. Arbitrary touch		x		x	x			x			x
25. Static graphics structures		x		x		x	x		x		
26. Dynamic graphics structures		x		x		x		x	x		
27. Sound structures		x		x		x		x		x	
28. Touch structures		x		x		x		x			x
modality	li	-li	an	-an	ar	-ar	sta	dyn	gra	sou	tou

Table 2. A taxonomy of generic unimodal modalities. Except for the rows containing modalities 9-20, each row exclusively represents one single generic unimodal modality. Four classes of modalities are separated by boldface lines, providing the super level of the taxonomy: primarily linguistic, primarily analogue, arbitrary and explicit structures. The table itself is a multimodal combination of modalities 5 and 25. combinations are ruled out because analogue representations should not be used arbitrarily. It makes little sense, for instance, to use static diagrammatic graphical representations of bananas to represent, e.g., cars. 16 combinations are ruled out because sound and touch are dynamic, not static, media; and 12 combinations are ruled out because language is non-arbitrary.

The basic features used to create the taxonomy will be explained in Sect. 4 below. Let us make some observations on Table 2. Except for the rows containing modalities 9 to 20, each row contains one single generic unimodal modality. To distinguish between the four triplets of analogue modalities 9 to 20, two more distinctions are needed. As these distinctions are not made at the generic level represented by Table

2, they will have to be represented at the lower, atomic level. One is between *real-world* representations and *diagrammatic* representations. Both are analogue but, prototypically, diagrammatic representations manipulate the representation of what is represented in various ways (e.g., abstracting from irrelevant detail or reducing dimensionality) whereas real-world representations do this to a lesser extent. Given current manipulation possibilities, this distinction (between 9/10, 12/13, 15/16 and 18/19, respectively) seems to have to be prototype-based. That is, the distinction has to be based solely on sets of intuitively central examples (or prototypes) of each category. There does not seem to be any strongly principled way of distinguishing between external representations that are ‘really’ like what they represent and external representations that are less like what they represent because of leaving out or having changed some aspects of what they represent (cf. Twyman 1979). In other words, there seems to be a *continuum of representation* between completely faithful external representations and more or less abstract, schematic or otherwise manipulated representations. For instance, an ordinary, non-manipulated, photograph is a prototypical real-world representation whereas diagrams of house layouts, engine parts or traffic accidents are prototypical analogue diagrams. Similarly, there are prototypical dynamic real-world representations such as videos and there are prototypical animated diagrammatic representations such as ‘virtual reality’ computer games using sound and graphics, and many scientific visualisations. Nobody would call a photograph a ‘diagram’ but if one wants to appeal to principles in order to justify such a distinction, it seems that prototypically determined categories are the best one can hope for.

The second distinction is between diagrammatic and real-world representations, on the one hand, and graphs (11, 14, 17 and 20, respectively) on the other. Graphs manipulate the representation in specific ways (see Sect. 4.2 below).

To put some meat on the bones of the taxonomy, Table 3 provides some familiar atomic types (if any) instantiating each of the generic unimodal modalities. These types are candidate types at the atomic level of the taxonomy. It is crucial to note that, although Table 3 contains nothing but well-known candidate atomic types, these types should not be confused with their corresponding, and someti-

Modality	Well-known atomic types
1. Static analogue graphic language	Hieroglyphs. Rarely used.
2. Dynamic analogue graphic language	Gestural language. Dynamic hieroglyphs would appear anachronistic.
3. Analogue spoken language	Part of everyday spoken language.
4. Analogue touch language	Apparently none.
5. Static non-analogue graphic language	Letters, words, numerals, other written language signs, text, logograms (e.g. arrows), special-purpose notations (e.g., programming languages, formal logic, music). Sequential, list and tabular orderings.
6. Dynamic non-analogue graphic language	Dynamizations of (5). Graphically viewed spoken language discourse.
7. Non-analogue spoken language	Spoken letters, words, numerals, other spoken language signs, discourse. List orderings.
8. Non-analogue touch language	Touch letters, numerals, words, other touch language related signs, text, list and table orderings. Example: Braille.
9. Static diagrammatic pictures	Pure diagrams, maps, cartoons. Sequential, list and tabular orderings.
10. Static non-diagrammatic real-world pictures	Photographs, naturalistic drawings, holograms. Sequential, list and tabular orderings.
11. Static graphs	1D, 2D or 3D graph space containing geometrical forms or other elements. Pure charts (dot charts, bar charts, pie charts, etc.). Sequential, list and tabular orderings. Non-iconic use difficult due to the absence of linguistic annotation.
12. Animated diagrammatic pictures	Pure animated diagrams. Pure standard animations.
13. Dynamic real-world pictures	Pure movies, videos, realistic animations.
14. Dynamic graphs	Pure graphs (see 11) evolving in graph space. Non-iconic use difficult due to the absence of linguistic annotation.

15. Real-world sound	Sound 'pictures'. Sound signals. List orderings.
16. Diagrammatic sound	Sound 'pictures'. Sound signals. Many possibilities, synthetic or manipulated, exist. Music? List orderings.
17. Sound graphs	E.g. Geiger counters.
18. Real-world touch	Single touch representations, touch sequences.
19. Diagrammatic touch	Apparently none, but many possibilities exist.
20. Touch graphs	1D, 2D or 3D graph space containing geometrical forms. Pure charts (dot charts, bar charts, pie charts, etc.).
21. Arbitrary static diagrammatic forms	Diagrams consisting of geometrical elements. Sequential, list and tabular orderings. Non-iconic use difficult due to the absence of linguistic annotation.
22. Animated arbitrary diagrammatic forms	Diagrams consisting of geometrical elements. Non-iconic use difficult due to the absence of linguistic annotation.
23. Arbitrary sound	Sound signal icons.
24. Arbitrary touch	Touch signals of different sorts.
25. Static graphic structures	Form fields, frames, table grids, line separations, trees, windows, bars. Sequential, list and tabular orderings. Non-iconic use difficult due to the absence of linguistic annotation.
26. Dynamic graphic structures	Dynamic frames, windows, scroll bars. Non-iconic use difficult due to the absence of linguistic annotation.
27. Sound structures	Apparently none.
28. Touch structures	Form fields, frames, grids, line separations, trees.

Table 3. Well-known atomic types (if any) of each of the generic unimodal modalities.

mes better known, *multimodal* counterparts. To avoid confusion, the types in Table 3 have sometimes been qualified as 'pure', i.e. unimodal. For instance, in contrast to most ordinary maps a pure map does not contain written language naming towns, rivers and other locations as this would make the map a multimodal (bi-modal) representation.

4. THE PROPERTIES OF LINGUISTIC, ANALOGUE, ARBITRARY AND STATIC/DYNAMIC REPRESENTATIONS, AND MEDIA

It appears from the combinations of basic features belonging to each generic modality in Tables 1 and 2 that the generic modalities are clustered and interrelated in various ways. Exposing such clusterings and interrelationships helps demonstrate the origins and nature of various classifications different from the taxonomy itself, some of which are common in the literature or in everyday use. Even more importantly, the features involved in creating different orthogonal classifications of generic modalities are crucial to the analysis of any particular unimodal or multimodal representation. These features are briefly discussed in this section. A detailed analysis is outside the scope of this paper.

4.1 Linguistic and Non-Linguistic Representational Modalities

Linguistic representations can, somehow, represent anything and one might therefore wonder why we need any other kind of modality for representing information in HCI. The basic reason seems to be that linguistic representations lack the *specificity* which characterise analogue representations (Stenning and Oberlander 1991, Bernsen 1993b, cf. 4.2 below). Instead, linguistic representations are *focused*: they focus on the subject-matter to be communicated without providing its specifics. The price of linguistic focusing is to leave open an *interpretational scope* as to the nature of the specific properties of what is being represented. My neighbour, for instance, is a specific person who may have enough specific properties to distinguish him from any other person in the history of the universe, but you won't know much about his specifics from understanding the expression 'my neighbour'. The presence of focus and lack of specificity jointly generate the characteristic, limited expressive power of linguistic representations, whether these be static or dynamic, graphical, auditory or tactile, or whether the linguistic signs used are themselves non-analogue (as in the present text) or analogue. Linguistic representation is, in an important sense, complementary to analogue representation. Many types of

information can only with great difficulty, if at all, be rendered linguistically, such as how things, situations or events exactly look or unfold, whereas other types of information can hardly be rendered at all using analogue representations, such as abstract states of affairs and relationships or non-descriptive speech acts. The complementarity between linguistic and analogue representation explains why their combination is optimal for many representational purposes. A detailed analysis of the implications of this complementarity for HCI is presented in Bernsen (1993b).

The taxonomy contains 8 different generic linguistic modalities constituting its linguistic super level:

1. Static analogue graphic language.
2. Dynamic analogue graphic language.
3. Analogue spoken language.
4. Analogue touch language.
5. Static non-analogue graphic language.
6. Dynamic non-analogue graphic language.
7. Non-analogue spoken language.
8. Non-analogue touch language.

The first 4 of these modalities are analogue in addition to being linguistic. They should be characterised as being *primarily* linguistic and only secondarily as being analogue representations because the integration of analogue signs into a linguistic system subjects the signs to a set of rules which make them far surpass the icons (signs) themselves in expressive power.

4.2 Analogue and Non-Analogue Representational Modalities

The distinction between analogue and non-analogue (external) representations is quite important as well as being intuitively obvious in most cases. Being complementary to linguistic representations, analogue representations (which are sometimes called ‘iconic’ or ‘isomorphic’ representations) have the virtue of specificity but lack focus, whether they be static or dynamic, graphical, auditory or tactile. Specificity and lack of focus and, hence, lack of interpretational scope, generate the characteristic, limited expressive power of analogue representations. As already noted, this complementarity explains why (multimodal) combinations of linguistic and analogue representations are eminently suited to many representational purposes. Thus, one basic use of language is to *annotate* analogue representations (e.g., a map, a diagram or a dynamic measurement representation for control room use), and one basic use of analogue representation is to *illustrate* linguistic discourse (Bernsen 1993b). The specificity of analogue representation is related to the fact that analogue representations have ‘shape’ or dimensionality, i.e. are encoded relative to a system of dimensions such as, e.g., 2-D space in the case of 2-D static graphics (Haugeland 1991, Slack, Fedder and Oberlander 1993). *Graphs* constitute a particular form of analogue representation in that they represent data in a graph space according to one or more selected dimensions of interest. In graphs, in contrast to real-world representations and diagrams, any ‘pictorial’ similarity to the represented subject-matter has disappeared but since dimensionality is still represented, graphs remain analogue representations.

Whereas analogue representations represent through providing some of the specifics of what is represented, non-analogue representations represent through conventional pairing between representation and what is represented. As long as we focus only on external representations (including touch) and do not consider the nature of internal cognitive representations, this distinction is clear in most cases. In practice, however, the distinction sometimes can be difficult to draw primarily because of the existence of *levels of abstraction* in analogue representation, whether the representation be a sound, a piece of graphics such as an ordinary diagram (of machine parts, say) or a tactile/kinaesthetic one. A highly abstract analogue representation may have so little dimensionality in common with what it represents that it may be close to acting as a non-analogue representation. The less common dimensionality there is between what is represented and its representation, the more we may have to rely on additional knowledge of the *representational conventions* used in order to decode particular representations. In the

limit, where we find, e.g., (most) natural language and arbitrarily chosen icons and diagrams, we have to rely exclusively on representational conventions. As noted above, many graphs come close to this limit.

Another problem in applying the analogue/non-analogue distinction is that it is sometimes unclear how *real* are the states of affairs to be represented in analogue representations. The equator, for instance, is nearly always represented on the relevant maps, but what does this representation correspond to? An arbitrary triangular icon, on the other hand, recognisably resembles many triangular shapes to be found in nature, so is it really arbitrary after all or is it a highly abstract analogue representation? These two examples may be distinguished according to the criterion that the equator on the map does represent a fixed topological property of the globe whereas the triangular icon really is intended as being arbitrary - it might just as well have been a circle or something else again. The represented 'reality' of analogue representation, therefore, is certainly more comprehensive than the tangible world of spatio-temporal objects, processes and events. A conceptual graph, for instance, does have a topology (and hence dimensionality) but it may be questioned whether the topology is an analogue representation of conceptual relations on the ground that such relations are not themselves topological. However, it is not evident at this point that the topology criterion just used will be able to resolve all problems about the analogue versus non-analogue character of particular external representations. We may have to accept the existence of an undecidable area between analogue graphic diagrams and non-analogue (arbitrary) graphic diagrams which are often alternatively called 'abstract' or 'conceptual' diagrams. The sound and touch domains might pose similar decidability problems.

The taxonomy ignores the difference between analogue external representations which have a 'real original' which they more or less faithfully represent and analogue external representations which in some sense might have had a real original but just happen not to have one, for instance because the real entity is about to be built as a result of ongoing work on CAD analogue screen representations. Such distinctions belong to the level of analysis of atomic types (see Sect. 9 below).

Categorising the 28 generic modalities according to the analogue/non-analogue distinction generates a classification of external representations which is orthogonal to the taxonomy of generic unimodal modalities. The class of *primarily* analogue external representations comprises the following modalities which constitute the (primarily) analogue super level of the taxonomy:

9. Static diagrammatic pictures.
10. Static non-diagrammatic real-world representations or pictures.
11. Static graphs.
12. Animated diagrammatic pictures.
13. Dynamic real-world representations or pictures.
14. Dynamic graphs.
15. Real-world sound representations.
16. Diagrammatic sound representations.
17. Sound graphs.
18. Real-world touch representations.
19. Diagrammatic touch representations.
20. Touch graphs.

The *secondarily* analogue modalities of the taxonomy are the linguistic modalities 1-4 which employ analogue signs (cf. 4.1 above).

It is important to stress again that we are only dealing with external representations. The analogue/non-analogue distinction behaves quite differently when we consider internal cognitive representations (see Sect. 8 below).

4.3 Arbitrary and Non-Arbitrary Representational Modalities

The distinction between non-arbitrary and arbitrary generic unimodal modalities marks the difference between external representations which, in order to perform their representational function, rely on an already existing system of meaning and representations which do not. In the latter case, the representation has to be accompanied by appropriate representational conventions at the time of its introduction. The reason why this distinction tends to be overlooked is that, in a number of cases, it coincides with the distinction between non-analogue and analogue representations. Thus the following generic modalities are both non-analogue and arbitrary:

21. Arbitrary static diagrams.
22. Animated arbitrary diagrams.
23. Arbitrary sound.
24. Arbitrary touch.

However, the following modalities are at the same time non-analogue and non-arbitrary:

5. Static non-analogue graphic language.
6. Dynamic non-analogue graphic language.
7. Non-analogue spoken language.
8. Non-analogue touch language.
25. Static graphic structures.
26. Dynamic graphic structures.
27. Sound structures.
28. Touch structures.

That standard spoken, graphic and touch language are non-analogue and non-arbitrary is straightforward. The same is true of the class of unimodal *structures*, such as empty forms or tables in static graphics. The existence of non-analogue and non-arbitrary modalities means that as many as 24 out of the 28 generic unimodal modalities exploit already existing systems of meaning. The only generic unimodal modalities which do not do so are the expressly arbitrary graphic diagrammatic forms, sound representations and touch representations. It seems obvious that, *ceteris paribus*, exploiting already existing systems of meaning is an advantage in usability engineering as in the external representation of information in general. Unfortunately, this can be done in many different ways for a given design or other representational purpose, not all of which are appropriate.

The separation performed between the analogue/non-analogue distinction, on the one hand, and the arbitrary/non-arbitrary distinction, on the other, does seem quite important. It shows why, e.g., natural language can compete successfully with analogue graphics for many interface representational purposes. Despite being non-analogue (or primarily non-analogue) considered as a form of external representation, natural language does build on an already existing system of meaning. And the separation between the analogue/non-analogue and arbitrary/non-arbitrary distinctions demonstrates that explanations of why, e.g., natural language modalities are in some cases inferior, and in others superior to analogue graphical and other modalities cannot simply be provided through appeal to the analogue/non-analogue distinction. One has to look deeper than that, namely into the distinction between specificity and focus. Furthermore, the distinction between arbitrary and non-arbitrary representational modalities is the one to consider when analysing the basic differences between representations which are, and representations which are not, based on already existing systems of meaning. It is not a problem for the taxonomy that representations which were originally intended as being arbitrary, gradually may acquire common use and hence become non-arbitrary. Traffic signs may be a case in point.

Music is a difficult case. It does seem to be based, in some sense, on an already existing system of meaning. One possibility might be to re-categorise music as being non-analogue, non-arbitrary and dynamic just like spoken language, and then attempt to identify the differences between musical 'meaning' and linguistic 'meaning'. However, this problem would seem marginal to HCI.

4.4 Static and Dynamic Representational Modalities

This is another important distinction because the differences between static and dynamic external representations have profound implications for their usability in specific task domain contexts. What is dynamic changes over time. In the domain of non-linguistic graphics, this distinction marks the obvious differences between:

Static generic modalities

9. Static diagrammatic pictures.
10. Static non-diagrammatic real-world representations or pictures.
11. Static graphs.
21. Static arbitrary diagrammatic forms.

and

Dynamic generic modalities

12. Animated diagrammatic pictures.
13. Dynamic real-world representations or pictures.
14. Dynamic graphs.
22. Animated arbitrary diagrammatic forms.

Written language is sometimes presented dynamically (cf. generic modality 6). The sound medium is inherently dynamic. The medium of touch appears to be inherently dynamic because of its close relationship with kinaesthesia. However, it is quite possible that finer distinctions will ultimately have to be made in this latter domain. Just as the distinction between specificity and focus seems to provide an ultimate explanation of the major differences between the expressive power of analogue and linguistic representational modalities, we have been looking for an ultimate explanation of the difference between static and dynamic representations which may inform a detailed view on when to use static or dynamic representations for given HCI task purposes. We are currently focusing on the notion of *freedom of perceptual inspection*. Thus, perceptually accessible static representations, such as most computer screens after startup, allow freedom of perceptual inspection whereas dynamic representations do not. This approach seems to work in the media of graphics and sound. The price to be paid for defining the static/dynamic distinction in terms of freedom of perceptual inspection is that recurrent patterns of dynamic change such as, e.g., a blinking cursor, should be classified as static representations of information. This price may be worth paying.

4.5 Representational Modalities in Different Media

A fourth classification, orthogonal to the taxonomy of unimodal modalities, is the one between different media of representation. Whereas the term ‘modality’ currently is being used in many different ways in the literature, there seems to be considerable consensus on the sense of the term ‘output medium of expression’. An output medium is a physical vehicle for the expression of information or for realising external representations. Different media, such as graphics, sound and touch have very different physical properties and are able to render very different sets of perceptual qualities. The term ‘medium’ (of expression), therefore, is much closer to the psychological notion of sensory modalities than is the term ‘(representational) modality’. The 28 generic modalities of the taxonomy are expressed in three different *media*, namely:

Visual and graphical qualities/vision

1. Static analogue graphic language.
2. Dynamic analogue graphic language.
5. Static non-analogue graphic language.
6. Dynamic non-analogue graphic language.
9. Static diagrammatic pictures.
10. Static non-diagrammatic real-world representations or pictures.
11. Static graphs.
12. Animated diagrammatic pictures.

- 13. Dynamic real-world representations or pictures.
- 14. Dynamic graphs.
- 21. Arbitrary static diagrams.
- 22. Animated arbitrary diagrams.
- 25. Static graphic structures.
- 26. Dynamic graphic structures.

Sound qualities/audition

- 3. Analogue spoken language.
- 7. Non-analogue spoken language.
- 15. Real-world sound representations.
- 16. Diagrammatic sound representations.
- 17. Sound graphs.
- 23. Arbitrary sound.
- 27. Sound structures.

Tactile and kinaesthetic qualities/touch

- 4. Analogue touch language.
- 8. Non-analogue touch language.
- 18. Real-world touch representations.
- 19. Diagrammatic touch representations.
- 20. Touch graphs.
- 24. Arbitrary touch.
- 28. Touch structures.

The relationship of modality types to the same or different media of expression is important to the external representation of information in usability engineering and elsewhere for the following reason. Different media imply quite different sets of perceptual qualities. These qualities, their respective scope of variation and their relative cognitive impact are at our disposal when we use a given representational modality in designing an interface. Standard written natural language, for instance, being graphical although not pictorial, can be manipulated graphically (coloured, rotated, highlighted, re-sized, textured, re-shaped, projected and so on), and such manipulations can be used to carry meaning in context. Spoken natural language, although basically non-analogue, can be manipulated auditorily (changed in pitch, volume, rhythm and so on) and the results used to carry meaning in context as we do when we speak.

If, in other words, we choose a given (unimodal) modality for the representation of information, this modality inherits a specific medium of expression whose different generic modalities of representation share a number of perceptual qualities which can be manipulated for representational purposes. This makes it possible to use the concept of *information channels* for the analysis of types and instances of representational modalities and modality combinations. A channel of information is a perceptual aspect of some medium which can be used to carry information. If, for instance, differently numbered but otherwise identical iconic ships are being used to express positions of ships on a screen map, then different colouring of the ships can be used to express additional information about them. Colour, therefore, is an example of an information channel (Hovy and Arens 1990, Bernsen 1993c) as is the shading used in Table 2 above to indicate infelicitous or impossible combinations of basic properties.

Evidently, there are other media of expression than the three media considered in this paper and the taxonomy may eventually have to be expanded to include them. So far, (output) media of expression such as physical machine gesture, smell and taste are outside the scope of the taxonomy. Much closer to current technological output possibilities, however, is the inclusion of graphically expressed linguistic information in the forms of lip movements, gestural language and facial expression (cf. generic modalities 2 and 6). Extension of the taxonomic work to cover input modalities will obviously have to include additional media such as keyboard and mouse.

4.6 A Common Sense Classification

As argued in the Introduction, a useful taxonomy of generic output modalities not only must be established in a principled and transparent way from combinations of analysed basic properties and at useful levels of abstraction. To be found useful it also has to correspond, to some significant extent, to our common sense intuitions about different categories of external representation. The taxonomy presented would seem to meet this latter requirement. The 28 modalities can be divided into the categories:

a. Language (natural or otherwise)

1. Static analogue graphic language.
2. Dynamic analogue graphic language.
3. Analogue spoken language.
4. Analogue touch language.
5. Static non-analogue graphic language.
6. Dynamic non-analogue graphic language.
7. Non-analogue spoken language.
8. Non-analogue touch language.

b. Pictures of something (in the ordinary sense)

9. Static diagrammatic pictures.
10. Static non-diagrammatic real-world representations or pictures.
11. Static graphs.
12. Animated diagrammatic pictures.

c. Non-visual 'pictures' or analogue representations

15. Real-world sound representations.
16. Diagrammatic sound representations.
18. Real-world touch representations.
19. Diagrammatic touch representations.

d. Graphs (i.e. analogue representations of a special kind)

11. Static graphs.
14. Dynamic graphs.
17. Sound graphs.
20. Touch graphs.

e. Representations which need a conventionally assigned meaning in order to represent something

21. Arbitrary static diagrams.
22. Animated arbitrary diagrams.
23. Arbitrary sound.
24. Arbitrary touch.

f. Explicitly rendered structurings of information

25. Static graphic structures.
26. Dynamic graphic structures.
27. Sound structures.
28. Touch structures.

The categories (a), (b), (e) and (f) are familiar in their own right. Category (c) corresponds to category (b), only covering different media. Language does not seem to have a label for category (c) but makes it tempting to use the term 'picture' analogously in characterising (c). Common sense may not have a position on graphs which were characterised above as a particular form of analogue representations. What is special about graphs is that, on the one hand, they are useless as external representations without an accompanying explanation of their mapping principles. This is also true of arbitrary unimodal modalities. On the other hand, graphs represent structured data which in their turn represent the world, and graphs do have structural similarities with the data they represent (Bernsen 1993c). Graphs are therefore in a very specific sense 'in between' analogue and non-analogue representations without this fact acting as a threat to the clarity of the distinction.

The list of categories (a)-(f) seems close to our standard intuitions about the domain of investigation. In particular, categories (a), (b), (e) and (f) are familiar, (c) is easily understood as an extension of (b) into two media different from that of vision and visual qualities and the category of graphs is easily perceived to be *sui generis*. Thus it appears possible to generate from basic principles (properties) modalities which are familiar at the super and generic levels.

5. WHAT IS A MODALITY?

Having presented a taxonomy of generic unimodal modalities and some orthogonal classifications of these above, the following operational definition of a generic unimodal modality comes out straightforwardly. A generic unimodal 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. To extend this definition to the atomic level, we simply open the list of profile properties to include further properties. This has already been done to some extent above in order to distinguish between analogue real-world representations, diagrams and graphs. A general definition of the term ‘modality’ including both unimodal and multimodal representations from the super level downwards, thus becomes the following: A modality has at least one medium of expression together with a profile which includes at least one property from the following list of binary opposites: analogue/non-analogue, arbitrary/non-arbitrary, static/ dynamic, linguistic/non-linguistic. A physical medium is necessary for any modality to act as an external representation. The profile may be quite general, as when we analyse the properties of all analogue modalities, or all dynamic modalities, or it may be quite specific as when we analyse the properties of 2-D static graphic pie charts annotated in static typed natural language.

6. ORTHOGONALITY AND COMPLETENESS OF THE TAXONOMY

Questions pertaining to the *orthogonality* of the taxonomy of generic unimodal modalities have raised at various occasions above. The crux is the extent to which the five property distinctions basic to the taxonomy are orthogonal. This is the case to a considerable extent. There are, however, some necessary interdependencies between those properties. Sound and touch are dynamic, not static. Linguistic representations are primarily non-analogue but may be secondarily analogue. Conversely, analogue representations are primarily non-linguistic but may be linguistic in a secondary sense. Linguistic, analogue and structural representations are non-arbitrary. As to the orthogonality of the generic modalities generated from the basic properties, it has been ensured from the way these modalities were generated. Orthogonality in this sense is compromised, but not severely, in at least two types of cases. One is that the distinction between analogue and non-analogue modalities in the same medium and either being both static or both dynamic, is sometimes difficult to draw (cf. Sect. 4.2). Another is that two important distinctions among analogue modalities require additional properties and therefore should be drawn at the next level down (i.e. the atomic level). One of these distinctions refers to prototypes, the other refers to the peculiarities of graphs as an important type of analogue modalities (cf. Sect. 4.2).

The question whether the taxonomy is *complete* is the question whether there might exist other generic unimodal modalities than those listed in Tables 1 and 2 above. In one sense this is evidently the case since we have ignored media of expression such as machine gesture, taste and smell and their corresponding perceptual qualities. Let us consider here a second sense of the question, namely whether there are or might be other generic unimodal modalities in the three media we are considering. The answer is that this is only possible to a very limited extent. Why this is so becomes apparent from Table 4.

According to Table 4, all the cells that were empty in Table 2 are empty by definition. In most cases, Table 4 reflects such trivial facts as that sound is not graphics and the like. Some might want to claim that this is a basic fact of nature rather than a matter of definition, but even then the corresponding cell would be empty. This result is not surprising on the assumption that Table 2 was generated through

performing only legitimate ‘pruning’ of Table 1 which in its turn contained all possible combinations of the properties basic to the taxonomy. The question of completeness, therefore, comes down to the question whether some of the ‘definitions’ underlying the taxonomy are inherently problematic. The only example we have been able to find is whether the medium of touch is necessarily dynamic or, in other words, whether static touch modalities should be added to the taxonomy. One intuition, for what it is worth, is that touch should be generally categorised as being dynamic because of the intimate connection be-

modality	li	-li	an	-an	ar	-ar	sta	dyn	gra	sou	tou
1. Static analogue graphic language	x	-d	x	-d	-d	x	x	-d	x	-d	-d
2. Dynamic analogue graphic language	x	-d	x	-d	-d	x	-d	x	x	-d	-d
3. Analogue spoken language	x	-d	x	-d	-d	x	-d	x	-d	x	-d
4. Analogue touch language	x	-d	x	-d	-d	x	-d	x	-d	-d	x
5. Static non-analogue graphic language	x	-d	-d	x	-d	x	x	-d	x	-d	-d
6. Dynamic non-analogue graphic language	x	-d	-d	x	-d	x	-d	x	x	-d	-d
7. Non-analogue spoken language	x	-d	-d	x	-d	x	-d	x	-d	x	-d
8. Non-analogue touch language	x	-d	-d	x	-d	x	-d	x	-d	-d	x
9. Diagrammatic pictures	-d	x	x	-d	-d	x	x	-d	x	-d	-d
10. Non-diagrammatic pictures											
11. Static graphs											
12. Animated diagram pictures	-d	x	x	-d	-d	x	-d	x	x	-d	-d
13. Dynamic pictures											
14. Dynamic graphs											
15. Real sound	-d	x	x	-d	-d	x	-d	x	-d	x	-d
16. Diagrammatic sound											
17. Sound graphs											
18. Real touch	-d	x	x	-d	-d	x	-d	x	-d	-d	x
19. Diagrammatic touch											
20. Touch graphs											
21. Arbitrary static diagrams	-d	x	-d	x	x	-d	x	-d	x	-d	-d
22. Animated arbitrary diagrams	-d	x	-d	x	x	-d	-d	x	x	-d	-d
23. Arbitrary sound	-d	x	-d	x	x	-d	-d	x	-d	x	-d
24. Arbitrary touch	-d	x	-d	x	x	-d	-d	x	-d	-d	x
25. Static graphics structures	-d	x	-d	x	-d	x	x	-d	x	-d	-d
26. Dynamic graphics structures	-d	x	-d	x	-d	x	-d	x	x	-d	-d
27. Sound structures	-d	x	-d	x	-d	x	-d	x	-d	x	-d
28. Touch structures	-d	x	-d	x	-d	x	-d	x	-d	-d	x
modality	li	-li	an	-an	ar	-ar	sta	dyn	gra	sou	tou

Table 4. This table shows that there are strict limits to the existence of generic unimodal modalities in addition to the ones already identified. The label **-d** indicates that it is a matter of definition that a cell is empty. Cells that were empty in Table 2 are uniformly empty by definition.

tween touch sensations and movement of the body surface of the person doing the touching. However, we can of course receive many different touch sensations without movement, such as electrical current, heat, passive contact with objects, etc. I must confess to not having a clear answer to propose on this question, nor to have any clear idea of its significance. Otherwise, there seems to be little opportunity for creating additional unimodal modalities within the media addressed. So if the information provided in Table 4 is reasonably correct, the taxonomy of unimodal modalities is close to being complete given the limited number of media it addresses. This means that, at the level of descriptive generality adopted, we have a reasonably robust taxonomy of the generic unimodal modalities of external representation which, either in their unimodal forms or in combination with other modalities, go into the building of human-computer interfaces to constitute multimodal and virtual reality representations.

7. MODALITY STRUCTURE

The term ‘icon’ is often being used to designate singular 2-D static graphic representations, be they analogue representations or non-analogue written words or letters. However, during work on the taxonomy it became apparent that icons can be created from *any* generic modality. The term ‘icon’, therefore, does not, strictly speaking, designate a modality. Rather, icons are defined by their singularity and their representational function. An icon is chosen to symbolise, all by itself, something in a particular context, possibly as part of a larger set of icons. And given their singularity, icons are almost always semantically ambiguous as to what they symbolise. Context may significantly help in disambiguating an icon but its ambiguous character is independent of whether the icon is analogue or not, arbitrary or not, static or dynamic, linguistic or non-linguistic, and is also medium-independent (see below). Its ambiguity is primarily due to its singularity. If icons are neither generic modalities nor constitute an atomic type subsumed under one particular generic modality, what are they? It is proposed to view icons as a particular type of *modality structure*.

Let us reconsider the ‘well-known types’ of each of the generic unimodal modalities in Table 3 above. It is clearly important to the understanding of the expressive potential of modalities to analyse in depth the different atomic types subsumed under the 28 generic modalities. While this is outside the scope of this paper, I want to mention an interesting observation suggested by the omnipresence of icons in the taxonomy. Their omnipresence suggests that modality structures may exist across all the distinctions which have so far been recognised as basic to the taxonomy. Intuitively, one might perhaps have assumed that even if the set of generic unimodal modalities is both finite, quite limited and reasonably orthogonal and complete, the number of atomic modalities subsumed under it would form a huge and poorly structured class whose individual members would have to be analysed independently of each other. The case of icons suggests that this might not be so. Following this lead, it turns out that several other structural types are found across the taxonomy. *Lists*, for instance, can be made up of words, text, pictures, animations, touch qualities and so on. Analogue *diagrams* can be created across the media. *Topological maps* can be created not only in graphics but also in sound and touch. Lifting the ‘well-known types’-column from Table 3 above and applying this idea, we obtain a more structured view of unimodal modality types (see Table 5).

One interesting point about Table 5 is that the domain of each generic unimodal modality has now been split into two different subsets of types. The first subset contains the *atomic representations* characteristic of the modality. Each of these will have to be analysed in its own right. The second subset contains the *modality structures* or structuring principles which can be applied to the atoms of a particular generic modality. Although not exhausted in Table 5, these modality structures appear limited in number and each cut across several different generic modalities. A second point of interest is that the list of atomic types no longer contains only well-known types. Some less well-known types have been added through using the structuring principles generatively. I shall return to these two points in the conclusion.

1. Hieroglyphs. Rarely used. Sequences, lists, tables, icons.
2. Gestural language. Dynamic hieroglyphs would appear anachronistic. Sequences, lists, icons.
3. Part of everyday spoken language. Sequences, lists, icons.
4. Apparently none. Sequences, lists, tables, icons.
5. Letters, words, numerals, other written language signs, text, logograms (e.g. arrows), special-purpose notations (e.g., programming languages, formal logic, music). Sequences, lists, tables, icons.
6. Dynamizations of (5). Graphically viewed spoken language discourse. Sequences, lists, icons.
7. Spoken letters, words, numerals, other spoken language signs, discourse. Sequences, lists, icons.
8. Touch letters, numerals, words, other touch language related signs, text, list. Example: Braille. Sequences, lists, tables, icons.
9. Pure diagrams, maps , cartoons. Sequences, lists, tables, icons.
10. Photographs, naturalistic drawings, holograms. Maps, sequences, lists, tables, icons.

11. 1D, 2D or 3D graph space containing geometrical forms or other elements. Pure charts (dot charts, bar charts, pie charts, etc.). Non-iconic use difficult due to the absence of linguistic annotation. Sequences, lists, tables, icons.
12. Pure animated diagrams . Pure standard animations. Maps, sequences, lists, tables, icons.
13. Pure movies, videos, realistic animations. Maps, sequences, lists, tables, icons.
14. Pure graphs (see 11) evolving in graph space. Non-iconic use difficult due to the absence of linguistic annotation. Sequences, lists, tables, icons.
15. Sound 'pictures'. Sound signals. Sound diagrams, maps, sequences, lists, icons.
16. Sound 'pictures'. Sound signals. Many possibilities, synthetic or manipulated, exist. Music? Sound diagrams, maps, sequences, lists, icons.
17. E.g. Geiger counters. Sequences, lists, icons.
18. Single touch representations, touch sequences. Touch diagrams, maps, sequences, lists, tables, icons.
19. Apparently none, but many possibilities exist. Touch diagrams, maps, sequences, lists, tables, icons.
20. 1D, 2D or 3D graph space containing geometrical forms. Pure charts (dot charts, bar charts, pie charts, etc.). Sequences, lists, tables, icons.
21. Diagrams consisting of geometrical elements. Non-iconic use difficult due to the absence of linguistic annotation. Sequences, lists, tables, icons.
22. Diagrams consisting of geometrical elements. Non-iconic use difficult due to the absence of linguistic annotation. Sequences, lists, tables, icons.
23. Sound signal icons. Sequences, lists, icons.
24. Touch signals of different sorts. Sequences, lists, tables, icons.
25. Form fields, frames, table grids, line separations, trees, windows, bars. Non-iconic use difficult due to the absence of linguistic annotation. Sequences, lists, tables, icons.
26. Dynamic frames, windows, scroll bars. Non-iconic use difficult due to the absence of linguistic annotation. Sequences, lists, tables, icons.
27. Apparently none. Sequences, lists, icons.
28. Form fields, frames, grids, line separations, trees. Sequences, lists, tables, icons.

Table 5. Atomic types and modality structures.

8. EXTERNAL AND INTERNAL REPRESENTATIONS

We have been considering only external representations in this paper. If, on the other hand, we attempt to go inside the cognitive system to consider the nature of *internal* representations, then the analogue/non-analogue distinction may not have much relevance any more. The reason is apparent from the distinction between arbitrary and non-arbitrary external representations in Sect. 4.3 above. It turned out that there are more categories of external representations which exploit already existing systems of meaning than there are analogue external representations. A different way of expressing this is to point out that internal representations, in order to serve their purpose of representing the world, may themselves to a large extent be analogue representations (in some sense) which build on material that has ultimately been derived from perceptual input to the human cognitive system. This may also be true of representations to which the words of natural language have been conventionally attached (see Bernsen 1993b). Internal representations, therefore, constitute a domain of research very different from the domain of external representations addressed above. It would perhaps have been easier if we did not have to enter this domain when analysing the foundations of modalities and multimodal representation. This cannot be avoided, however, since we cannot avoid issues concerning what, e.g., natural language is good at representing, or what graphics is good at representing, what various combinations of natural language and graphics are good at representing, or indeed what many different combinations of modalities are good or bad at representing. It is not possible to provide relevant explanations of such issues without discussing the nature of the internal representations linked to, e.g., natural language and graphics. The reason is both simple and compelling:

Considered purely as an external representation, a written natural language sequence on a screen would merely clutter up the screen without contributing anything else. What makes the sequence potentially useful for the representation of information is the fact that users *understand* the language used, i.e., that

they have access to the system of meaning on which this particular language is based. This means that they are able to form appropriate internal representations of what some written word sequence represents. These internal representations are not identical to the external (written) representations which cause or evoke them. If they were, then another process of interpretation would have to take place, and so on *ad infinitum*. So if we want to, e.g., optimally combine natural language and graphics for representing something on a screen, we cannot avoid considering the nature of the internal representations which are likely to be evoked in users by the natural language we consider using and by the graphics we might use. Nor can we avoid considering the cognitive processes operating on internal representations, the various cognitive limitations on these processes, effects of users' background knowledge on their understanding of mapping principles as well as on their interpretation of the specific types of external representation used, and so on. The same is true of many other modality combinations.

We must therefore reconsider the simple abstract diagram of Sect. 2 above which only dealt with states of affairs to be represented, external representations of these and the mapping principles from states of affairs to representations:

What is to be represented <-> mapping principles <-> representations.

The real situation in, e.g., interface design is somewhat more complex if internal representations are taken into account (see Diagram 2). It is, therefore, no wonder that many things can go wrong in interface design. The artifact designers may not have adequate ideas of what is to be represented at the human-computer interface. The mapping principles may be unknown or only partially known to the users who therefore misunderstand or fail to understand the external representations used by the designers unless provided with additional information through manuals, training, exploration, etc. The designers may have used inappropriate unimodal representational modalities or multimodal combinations of these for the specific representational purpose at hand, in which case the inappropriateness may have many different sources: the modalities chosen may be inappropriate for the information to be represented, users' cognitive architectures may be unable to cope with the information as represented although the mapping principles are known to the users, and so on.

States of affairs to be represented <-> designers' ideas of what is to be represented <-> mapping principles <-> external representations of the states of affairs at the interface <-> users' internal representations of the states of affairs represented.

Diagram 2: The complexity involved in trying to externally represent states of affairs to others.

9. CONCLUDING DISCUSSION

We are, clearly, still far from having provided the full foundations for analysing multimodal and virtual reality representations for the purpose of supporting usability engineering. This was already made clear from the outset in this paper where five consecutive steps of increasing complexity were described as being necessary to the creation of such foundations (Sect. 1). We haven't even scratched the surface of steps three and four which dealt with input modalities, interaction and task domain information/interface mapping, respectively (preliminary work on step 4 are reported in Bernsen and Bertels 1993, Bernsen 1994). Let us here merely consider step two and how the results of this paper might facilitate approaching the complexity involved:

- to establish sound foundations for describing and analysing any particular type of unimodal or multimodal output representation relevant to HCI;

If the taxonomy of generic unimodal output modalities is anything to go by, there is a huge number of existing and possible combinations of atomic modality types. A tiny fraction of these are well-known and

have been rather extensively analysed in the literature, such as (static) analogue graphic/written natural language maps, (static) graphic/written natural language analogue or abstract diagrams, or (static) graphic/written natural language graphs (Twyman 1979, Tufte 1983,1990). However, there are literally thousands of possible output modality combinations. For instance, no one is currently able to exclude the unfamiliar prospect that some combination of written natural language tables and datagloves might some day achieve prototypical status and a name in language because of having become popular for the performance of some prominent task category; or, some combination of atomic types belonging to most of our 28 generic unimodal modalities might soon become involved in advanced virtual reality representations of, say, flight decks. There is no clear sense at this point in undertaking a detailed analysis of each and every such actual or possible combination of atomic modalities. Even ignoring the scale of such an undertaking, it would not be feasible without sound foundations. Rather, the only viable solution seems to be to establish foundations which enable a principled scientific analysis of *any given* output modality combination *once* it is considered for analysis.

This is where the taxonomy and principles presented above might prove useful. The taxonomy actually reduces the problem into the following sub-problems:

(1) Provide a deep analysis of the binary opposites used in the taxonomy, i.e., analogue/non-analogue, arbitrary/non-arbitrary, static/dynamic and linguistic/non-linguistic representations, as well as of the expressive potential of each of the three media including analysis of their respective information channels.

(2) Identify and analyse the characteristics of the atomic types under each generic unimodal modality starting from their characterisation through the taxonomy. It should be noted that the set of 'well-known types' of the generic unimodal modalities listed in Table 3 above does not yet constitute a complete set and has not been derived in a principled way.

(3) Analyse the modality structures which cut across the boundaries imposed by the different categorisations of the taxonomy, i.e., the 28 generic modalities, the binary opposites and the media.

Implementing this programme as we are currently attempting to do is still no minor task. However, the task is limited and of well-defined scope. Parts of (1) and (2) have been addressed above and (3) would seem eminently feasible. Furthermore, the approach described is principled rather than *ad hoc*. Last but not least, this approach seems to have the potential needed for enabling the analysis of any given generic modality type or combination of modality types as external representations of information.

REFERENCES

- Bernsen, N.O. (1993a): 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. (1993b): Specificity and focus. Two complementary aspects of analogue graphics and natural language. *Esprit Basic Research project GRACE Deliverable 2.1.3*, 1993 (submitted).
- Bernsen, N.O. (1993c): Matching information and interface modalities. An example study. *Esprit Basic Research project GRACE Deliverable 2.1.1*, 1993.
- 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. (1994): Modality Theory: Supporting multimodal interface design. To appear in *Proceedings from the ERCIM Workshop on Multimodal Human-Computer Interaction*, Nancy, November 1993.
- Haugeland, J.: Representational genera. In Ramsey, W., Stich, S.P. and Rumelhart, D.E. (Eds.): *Philosophy and Connectionist Theory*. Hillsdale, NJ, Erlbaum, 1991.
- 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.
- Slack, J., Fedder, L. and Oberlander, J.: Animation: A representation framework. *Esprit Basic Research project GRACE Deliverable 4.1.1*, 1993.
- Stenning, K. and Oberlander, J.: Reasoning with words, pictures and calculi: Computation versus justification. In Barwise, J., Gawron, J.M., Plotkin, G. and Tutiya, S. (Eds.): *Situation Theory and Its Applications*. Stanford, CA, CSLI, 1991, Vol. 2, 607-621.
- Tufte, E.R.: *The Visual Display of Quantitative Information*. Cheshire, CT, Graphics Press, 1983.
- Tufte, E.: *Envisioning Information*. Cheshire, CT, Graphics Press, 1990.
- Twyman, M.: A schema for the study of graphic language. In Kolers, P., Wrolstad, M. and Bouna, H. (Eds.): *Processing of Visual Language* Vol. 1. New York: Plenum Press 1979.

Acknowledgements: The work described in this paper was carried out under Esprit Basic Research project 6296 GRACE whose support is gratefully acknowledged. Discussions with Michael May have been of great help during the development of ideas. Kenneth Holmquist has provided valuable comments on an earlier draft. Many thanks are also due to the two anonymous reviewers who commented on the submitted paper.