A Reference Model for Output Information in Intelligent Multimedia Presentation Systems

Niels Ole Bernsen[1]

Abstract. The paper addresses an issue that must be resolved in order to produce a scientifically sound and practically useful reference model for intelligent multimedia presentation systems (IMP systems), namely that of providing a systematic understanding of the types of output information to be presented by IMP systems. The term 'medium', though well-defined, is too coarse-grained for distinguishing between different types of output information. The paper introduces the notion of (representational) 'modalities' to enable sufficiently fine-grained distinctions to be made. For the term itself to be meaningful, 'multimodal' presentations must be composed of unimodal representations. In the approach presented, unimodal representations are defined from a small number of basic properties whose combinations specify the 'generic' level of a taxonomy of unimodal output modalities. To be scientifically sound as well as practically useful, the taxonomy must satisfy requirements of completeness, orthogonality, relevance and intuitiveness. The generic level of the taxonomy turns out to be too abstract to satisfy these requirements. By consequence, an 'atomic' and a 'sub-atomic' level are generated by analysis from the generic level, which satisfy the mentioned requirements. Based on the atomic and sub-atomic levels, all possible multimodal representations in the media of graphics, acoustics and haptics can now be generated by composition. The concluding discussion raises the issues of empirical validation of the taxonomy, its practical usefulness, and of expanding the approach to cover input modalities of information as well as user-system interactivity. [1]

Keywords. Intelligent multimedia presentation systems, reference model, multimodal systems, output information, modality theory.

1 INTRODUCTION

In recent years, networked intelligent multimedia presentation systems (IMP systems) have become a focal point in the development of advanced information technologies. Part of the background for this development is the technological advances that have been made in affordable computing power, communication bandwidth, external devices, sensors, mobile communication and novel forms of human-computer interfaces. Another part of the background is the fact that virtually everybody is becoming a computer user. IMP systems promise vastly increased intuitiveness of interaction between ordinary citizens and their computing systems. Future users of computing systems will be looking back upon the recent past as one of highly limited, primitive and cumbersome, desktop-bound systems providing "islands of computing" in a sea of unexplored opportunities. This paper addresses one of the many issues that must be resolved in order to produce a scientifically sound reference model for IMP systems, namely that of providing a systematic understanding of the types of output information (or output modalities) to be presented by IMP systems.

The work on output modalities to be described in this paper forms part of the research agenda of modality theory which addresses the following general 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 from the user's point of view constitute an optimal solution to the representation and exchange of that information*. The research agenda of modality theory requires that the following objectives be pursued:

1. To establish a taxonomy of the unimodal modalities which go into the creation of multimodal output representations of information for human-computer interaction (HCI). When coupled with concepts appropriate to modality analysis, this should enable the establishment of sound foundations for describing and analysing any particular type of unimodal or multimodal output representation relevant to HCI;
2. to establish a corresponding taxonomy and related analyses of the unimodal input modalities which go into the creation of multimodal input representations for HCI. This should enable the establishment of sound foundations for describing and analysing any particular type of unimodal or multimodal input representation relevant to HCI;
3. to establish a "grammar" for how to legitimately combine different unimodal output modalities, different unimodal input modalities, and different input and output modalities for the usable representation and exchange of information at the human-computer interface;
4. to develop a methodology for applying the results of the steps above to the analysis of the problems of information mapping between work/task domains and human-computer interfaces in information systems design;
5. to use results in building, possibly automated, practical interface design support tools.

The ultimate aim of modality theory is thus a practical one, namely to support the design of usable IMP systems and interfaces. This paper addresses objective (1) of the research agenda of modality theory. In output modality analysis we are primarily interested in knowing which information a particular unimodal modality or modality combination is suited or unsuited for representing in context. This kind of output modality analysis has long traditions, particularly in the medium of static graphics which antedates the computer. Outstanding examples are the results achieved on static graphic graphs [10,30,31]. In HCI, Hovy and Arens [15] called for a more general approach. Today, results in modality analysis are proliferating. However, most of these results

[1] Centre for Cognitive Science, Roskilde University, PO Box 260, 4000 Roskilde, Denmark, emails: nob@cog.ruc.dkl phone: +45 46 75 77 11 fax: +45 46 75 45 02

ECAI 96 workshop: *Towards a Standard Reference Model for Intelligent Multimedia Systems.*
Edited by G.P. Faconti and T. Rist.

concern individual unimodal modalities, such as speech input [18] or 3D graphics output [25], modality combinations, such as speech and writing [26], application areas, such as business applications [11,24], or user groups [27], without the complementary benefit of a systematic framework into which the results could feed. Such taxonomic work is still in its infancy [20,32].

In what follows, Section 2 defines the terms 'media' and 'modalities'. Section 3 states four requirements on an adequate theory of output modalities and presents the basic properties underlying the taxonomy. Section 4 describes the generation of the generic level of the taxonomy. Section 5 describes the generation of the atomic level of the taxonomy. Section 6 describes the selective generation of the sub-atomic level of the taxonomy. Section 7 describes how individual unimodal modalities are analysed in modality theory. Section 8 briefly discusses multimodal generation. And Section 9 concludes the paper by discussing empirical validation, providing evidence of the usefulness of the theory in design practice and pointing to ongoing work on input modalities and interactivity.

2 MEDIA AND REPRESENTATIONAL MODALITIES

In the present approach, a *medium* is the physical realisation of some presentation of information. In the foreseeable future, IMP systems will mainly be using three such media, i.e. graphics, acoustics and haptics. A *multimedia* system is one which outputs information in several media either simultaneously or sequentially. Obviously, the term 'medium' only provides a very coarse-grained way of distinguishing between different types of output information. For instance, a graphical image illustration and a piece of typed UNIX notation are both output graphics, and an alarm beep and a synthetic spoken language instruction are both output acoustics, even though those representations have very different properties which make them suited or unsuited, as the case may be, for different tasks, users, environments, communicative acts, or systems, or for optimising different performance parameters, learning parameters or cognitive properties. A more fine-grained approach to output information is therefore needed in addition to the distinction between media of expression. It is becoming common in the literature on IMP systems to refer to "multimodal" systems and interfaces [1,12,15,16,18, 22,23,26] but there is still a lack of consensus about what this term actually means or should be taken to mean in this context. The following proposal for a definition is simple and appears to agree with those parts of the literature which do not rely on a classical psychological notion modalities. A *modality* is a mode or way of representing information to humans or machines in a physically realised intersubjective form, such as in one of the media of graphics, acoustics and haptics. A modality is thus a *representational* modality and not a *sensory* modality as the term 'modality' has traditionally been used in cognitive psychology. Examples of representational modalities are tables, beeps, written and spoken natural language [15]. Given the sense of 'modality' just introduced, a *multimodal* (output) system or interface is one which outputs information as represented in several different modalities either simultaneously or sequentially.

3 REQUIREMENTS AND BASIC PROPERTIES

Clearly, it does not make operational sense to describe a piece of output information as being 'multimodal' unless we are able to decide what the *unimodal* constituents of the multimodal representation are. The crucial issue is how to identify those constituents. We want to identify a set of universally acceptable constituents of multimodal representations based on the observation that different modalities have different properties which makes them suitable for representing different types of information in context. How might this be done? Basically, two approaches are possible, one purely empirical, the other hypothetico-deductive, i.e. through empirical testing of a systematic theory or hypothesis. Note that both approaches are empirical ones, just in different ways. Although the purely empirical approach has a strong potential for providing relevant insights, it should be remembered that no stable scientific taxonomy was ever created in a purely empirical fashion from the bottom up. If we ask experimental subjects to cluster a more or less randomly selected set of, e.g., static graphic representations [20], the subjects may classify according to different criteria, be unable to express the criteria, and in the individual subject the criteria may be incoherent. The alternative to the purely empirical approach is to generate modalities from basic principles and then test through intuition and experiment whether the generated modalities satisfy a number of general requirements. If not, the basic principles will have to be revised. This is how generative grammar works in linguistics [13]. Ultimately, it is we, the native language speakers, who decide whether a proposed generative grammar actually generates all and only the syntactically correct sentences in some fragment of natural language. Note also that a generative grammar has different levels of generality, i.e. can generate sentences at different levels of syntactic detail from the top down. This analogy will be helpful in what follows.

To be of use in specifying a reference model for IMP systems, we want to identify a set of unimodal modalities which satisfies the following requirements:

(a) *completeness,* such that any piece of output information in the media of graphics, acoustics and haptics can be exhaustively described as consisting of one or more unimodal modalities;
(b) *orthogonality,* such that any piece of output information in those media can be characterised in only one way in terms of unimodal modalities;
(c) *relevance,* such that it captures the important differences between, e.g., beeps and spoken language from the point of view of output information representation; and
(d) *intuitiveness,* such that IMP systems and interface designers can recognise the set as corresponding to their intuitive notions of differences between modalities. Given the practical aims of modality theory, it is of crucial importance to operate with intuitively easily accessible notions without sacrificing theoretical systematicity.

These four requirements differ in status with respect to the empirical testing of the proposed reference model. Thus (d), on intuitiveness, is the more immediately accessible to evaluation, much in the same way as we judge whether a sentence in our native language is syntactically correct or not. The empirical issues will be addressed in Section 9.

To meet (a)-(d), we first need a basis for generating unimodal modalities. We start by defining a first set of unimodal modalities from a small set of *basic properties* which serve to robustly distinguish modalities from one another. The properties are: *linguistic/non-linguistic*, *analogue/nonanalogue*, *arbitrary/non-arbitrary* and *static–dynamic*. The analogy with generative grammar is helpful when addressing the question of how the choice of basic properties can be justified. Generative grammar starts with the most prominent features of sentences, as in the rule "S (sentence) -> NP (noun phrase) VP (verb phrase)", proceeding with increasingly detailed distinctions between the sentence parts. Modality theory starts with what are arguably the most basic distinctions between the capabilities of physically realised representations for representing information to humans. The set of basic properties have been chosen such that it is evident that their presence in, or absence from, a particular representation of information makes significant differences to the usability of that representation for some specific human-computer interface design purpose.

The (non-negatively defined) basic properties used in the generation may be briefly defined as follows, linguistic and analogue representations being defined in contrast to one another:

Linguistic representations are based on existing syntactic-semantic-pragmatic systems of meaning. Linguistic representations can, somehow, represent anything and one might therefore wonder why we need any other kind of modality for representing information in IMP systems. The basic reason appears to be that linguistic representations lack the *specificity* which characterise analogue representations [6,29]. Instead, linguistic representations are *focused*: they focus, at some level of abstraction, on the subject-matter to be communicated without providing its specifics. The cost of abstract 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 in the way he looks, sounds and feels to distinguish him from any other person in the history of the universe, but you won't know much about these 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, graphic, acoustic or haptic, or whether the linguistic signs used are themselves non-analogue as in the present text, or analogue as in iconographic sign systems such as hieroglyphs. Linguistic representation therefore 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, sound, feel, smell, taste or unfold, whereas other types of information can hardly be rendered at all using analogue representations, such as abstract concepts, states of affairs and relationships or the contents of non-descriptive speech acts. The complementarity between linguistic and analogue representation explains why their combination is so excellent for many representational purposes. A detailed analysis of the implications of this complementarity for HCI is presented in [6].

Analogue representations represent through aspects of similarity between the representation and what it represents. These aspects can be many or few. Being complementary to linguistic modalities, analogue representations (which are sometimes called 'iconic' or 'isomorphic' representations) have the virtue of specificity but lack abstract focus, whether they be static or dynamic, graphic, acoustic or haptic. Specificity and lack of focus and, hence, lack of interpretational scope, generate the characteristic, limited expressive power of analogue representations. Thus, a photograph, haptic image, sound track or video representing my neighbour would provide the reader with large amounts of specific information about how he looks and sounds, which might only be conveyed linguistically with great difficulty, if at all. As already noted, the complementarity between linguistic and analogue representation explains why their (multimodal) combination is eminently suited for many representational purposes. Thus, one basic use of language is to *annotate* analogue representations, such as a 2D graphic map or a haptic compositional diagram; and one basic use of analogue representation is to *illustrate* linguistic text. In annotation, analogue representation provides the specificity; in illustration, language provides the generalities and abstractions which cannot be provided through analogue representation.

The distinction between *non-arbitrary* and *arbitrary representations* 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 must be accompanied by appropriate representational conventions at the time of its introduction. In the former case, such as when using the linguistic expressions of some natural language known to the interlocutor, introductory conventions are unnecessary as the expressions already belong to an established system of meaning. It is not a problem for the taxonomy that representations which were originally intended as being arbitrary, may gradually acquire common use and hence become non-arbitrary. Traffic signs may be a case in point.

Rather trivially, *static representations* are *non-dynamic representations* and dynamic representations are non-static representations. However, modality theory does not have a purely physical notion of static representation. Rather, static representations are such which offer the user *freedom of perceptual inspection*. This means that static representations may be decoded by users in any order desired and as long as desired. According to this static/dynamic distinction, a representation is static also when it exhibits short-duration repetitive change. Thus, for instance, an acoustic alarm signal which sounds repeatedly until someone switches it off, or a graphic icon which keeps blinking until someone takes action to change its state, is considered static rather than dynamic. The implication is that some acoustic representations are static. A movie video that plays indefinitely, on the other hand, would still be considered dynamic. The reason for adopting this not-purely-physical definition of static representation is that, from a usability point of view, and that is what interface designers have to take into account when selecting modalities for their applications, the main distinction is between representations which offer freedom of perceptual inspection and representations which do not. Just imagine, for instance, that your standard GUI Macintosh or Windows main screen had been as dynamic as a lively movie. In that case, the freedom of perceptual inspection afforded by static graphics would have been lost with disastrous results both for the decision-making process which precedes most interaction and for the interaction itself. Finally, the static-dynamic distinction adopted does not imply, of course, that a

blinking graphic image icon has exactly the same usability properties as a physically static one. Distinction between them is still needed and will have to be made internally to the treatment of static graphic modalities.

So the first justification for the choice of basic properties is their profoundly different capabilities of representing information. A second justification for the choice of basic features is that they *serve to generate the right outcome*, i.e. to generate the output modalities which fit the intuitions that designers already have, just like in generative grammar where it can be difficult to judge a single generative rule by itself, the rule being judged, rather, from its generative contribution.

To the above basic properties we add distinction between the physical media of expression of graphics, acoustics and haptics. These media determine the *scope* of the taxonomy. Thus the taxonomy will not cover, for instance, olfactory and gustatory output representations of information or robot gesture which would all appear largely irrelevant to current IMP design. The *media* physically instantiate modalities of information representation. Through their respective physical instantiations, each medium is accessible through different sensory modalities, the graphic medium visually, the acoustic medium auditorily and the haptic medium tactilely. This means that different media, such as graphics, acoustics and haptics, have very different physical properties and are able to render very 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 certain representational modality in designing an interface. Standard typed natural language, for instance, being graphical but non-analogue, 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 mainly non-analogue, can be manipulated acoustically (changed in pitch, volume, rhythm and so on) and the results used to carry meaning in context as we do when we speak. The term 'medium' (of expression), therefore, is much closer to the psychological notion of sensory modalities than is the term '(representational) modality'.

When, in designing a human-computer interface, we choose a certain (unimodal) output modality to represent information, this modality inherits a specific medium of expression which it shares with a number of other unimodal modalities. This makes it possible to use the concept of information channels for the analysis of types and instances of representational modalities and modality combinations. An *information channel* is a perceptual aspect, that is, an aspect accessible through human perception, of some medium, which can be used to carry information in context. If, for instance, differently numbered but otherwise identical iconic ships are being used to express positions of ships on a screen map, then different colourings of the ships can be used to express additional information about them. Colour, therefore, is an example of an information channel [15]. Information channels characteristic of a certain medium of expression are illustrated in Section 7.

4 GENERIC-LEVEL UNIMODAL MODALITIES

Exhaustive combination of the basic properties presented in Section 3 mechanically produces the 48 (= 2x2x2x2x3) basic property combinations or unimodal modalities shown in Table 1. We call the level of abstraction at which modalities are presented in Table 1 the *generic level* of the taxonomy of unimodal modalities [3].

Table 1. The full set of 48 combinations of basic properties constituting the possible modalities at the generic level of the taxonomy. All modalities provide possible ways of representing information but only 30 of them are useful for IMP purposes. These are named in Table 2.

	li	**-li**	**an**	**-an**	**ar**	**-ar**	**sta**	**dyn**	**gra**	**aco**	**hap**
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**	**hap**

Whereas each of the generated 48 unimodal output modalities is perfectly acceptable as a mode of information representation, not all of them are acceptable for the purpose of IMP systems and interface design [4]. For instance, the arbitrary use of established linguistic expressions in a static graphic interface (e.g. modality 13 in Table 1) should not occur in IMP systems output. To do so would be like playing the children's' game of letting 'yes' mean

'no' and vice versa. As we all know, the result is massive production of communication error and ultimate communication failure. Formally, what is involved is providing a representation which already has an established meaning, with an entirely different meaning. This style of information representation is certainly meaningful and sometimes even useful, as in classical cryptography which makes use of the expressive strength of particular tokens belonging to some representational modality in order to mislead. The taxonomy, on the other hand, aims to support designers in making the best use of representational modalities through building on the expressive strengths of each, which implies the avoidance of predictable communication error. In terms of the requirements (a)-(d) above, the arbitrary use of non-arbitrary modalities conflicts with the requirement of relevance. Removing the modalities which represent the arbitrary use of non-arbitrary modalities produces the pruned set of generic level unimodal output modalities shown in Table 2. In Table 2, the remaining modalities have been named and represented in abbreviated notation.

In addition, Table 2 subsumes the generic-level modalities under the *super level* of the taxonomy. The 30 generic unimodal modalities of Table 2 have been divided into 4 different classes at the super level, i.e. the linguistic, the analogue, the arbitrary and the explicit structures. The super level merely represents one convenient way of classifying the generic-level modalities. Other, equally valid, classifications are possible, for instance in terms of the static-dynamic distinction or in terms of the distinction between media. The super level, therefore, has no deeper theoretical significance although, once laid down, it determines the overall architecture of the taxonomy.

Appealing to the intuitiveness requirement above, a further reduction in the number of generic unimodal modalities can be made. Acoustic modalities are mostly dynamic. Static acoustics, such as acoustic alarm signals, constitute a relatively small and reasonably well-circumscribed fraction of acoustic representations. Furthermore, given the present state-of-the-art in output devices, haptic modalities are mostly static. The dynamic haptics fraction may not be well circumscribed, however, and may be expected to grow dramatically with the growth of haptic output technologies. When this happens, we may simply re-introduce the static/dynamic distinction in the haptic modalities part of the taxonomy. These considerations allow a *pragmatic fusion* of the static and dynamic acoustic modalities and the static and dynamic haptic modalities (Table 3). Table 3 represents the final version of the generic level of the taxonomy. Pragmatic fusion implies no loss of information, i.e. does not sacrifice completeness, is completely reversible at any time and reduces the overall complexity of the taxonomy. The latter is important given the intuitiveness requirement. The taxonomy becomes less scholastic, as it were, and more usable.

A final remark on the generic level modalities in Tables 2 and 3 is the following. Four of the linguistic modalities use analogue *signs* and four use non-analogue signs. Basically, however, they are all linguistic, and hence *non-analogue* representations because the integration of analogue signs into a syntactic-semantic-pragmatic system of meaning subjects the signs to sets of rules which make them far surpass the analogue signsthemselves in expressive power. This may be the reason why all known, non-extinct iconographic languages have seen their stock of analogue signs decay to the point where it became difficult to decode their analogue meanings.

Table 2. 30 generic unimodal modalities result from removing from Table 1 the arbitrary use of non-arbitrary modalities of representation. The left-hand column shows the super level of the taxonomy. Modality theory notation has been added in the right-hand column.

SUPER LEVEL	GENERIC 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 haptic language	<li,an,-ar,sta,hap>
<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 haptic language	<li,an,-ar,dyn,hap>
	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 haptic language	<li,-an,-ar,sta,hap>
	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 haptic language	<li,-an,-ar,dyn,hap>
II. Analogue modalities	13. Static analogue graphics	<-li,an,-ar,sta,gra>
	14. Static analogue acoustics	<-li,an,-ar,sta,aco>
	15. Static analogue haptics	<-li,an,-ar,sta,hap>
<-li,an,-ar>	16. Dynamic analogue graphics	<-li,an,-ar,dyn,gra>
	17. Dynamic analogue acoustics	<-li,an,-ar,dyn,aco>
	18. Dynamic analogue haptics	<-li,an,-ar,dyn,hap>
III. Arbitrary modalities	19. Arbitrary static graphics	<-li,-an,ar,sta,gra>
	20. Arbitrary static acoustics	<-li,-an,ar,sta,aco>
	21. Arbitrary static haptics	<-li,-an,ar,sta,hap>
<-li,-an,ar>	22. Dynamic arbitrary graphics	<-li,-an,ar,dyn,gra>
	23. Dynamic arbitrary acoustics	<-li,-an,ar,dyn,aco>
	24. Dynamic arbitrary haptics	<-li,-an,ar,dyn,hap>
IV. Explicit modality structures	25. Static graphic structures	<-li,-an,-ar,sta,gra>
	26. Static acoustic structures	<-li,-an,-ar,sta,aco>
	27. Static haptic structures	<-li,-an,-ar,sta,hap>
	28. Dynamic graphic structures	<-li,-an,-ar,dyn,gra>
<-li,-an,-ar>	29. Dynamic acoustic structures	<-li,-an,-ar,dyn,aco>
	30. Dynamic haptic structures	<-li,-an,-ar,dyn,hap>
SUPER LEVEL	**GENERIC LEVEL**	**NOTATION**

5 ATOMIC-LEVEL UNIMODAL MODALITIES

The generic-level taxonomy does not fully meet the requirements on relevance and intuitiveness (Section 3). This is partly due to the fact that some of its modalities are largely obsolete, such as the hieroglyphs subsumed by modality 1 in Table 3. Much more

important, however, is the lack of intuitiveness of several of the modalities in Table 3, such as modality 9 'static analogue graphics', which is due to the relatively high level of

Table 3. The 20 generic unimodal modalities resulting from pragmatic fusion of the static and dynamic acoustic modalities and the static and dynamic haptic modalities in Table 2.

SUPER LEVEL	GENERIC LEVEL	NOTATION
I. Linguistic modalities	1. Static analogue sign graphic language	<li,an,-ar,sta,gra>
	2. Static analogue sign acoustic language Dynamic analogue sign acoustic language	<li,an,-ar,sta/dyn, aco>
	3. Static analogue sign haptic language Dynamic analogue sign haptic language	<li,an,-ar,sta/dyn, hap>
<li,-an,-ar>	4. Dynamic analogue sign graphic language	<li,an,-ar,dyn,gra>
	5. Static non-analogue sign graphic language	<li,-an,-ar,sta,gra>
	6. Static non-analogue sign acoustic language Dynamic non-analogue sign acoustic language	<li,-an,-ar,sta/dyn, aco>
	7. Static non-analogue sign haptic language Dynamic non-analogue sign haptic language	<li,-an,-ar,sta/dyn, hap>
	8. Dynamic non-analogue sign 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>
<-li,an,-ar>	11. Static analogue haptics Dynamic analogue haptics	<-li,an,-ar,sta/dyn, hap>
	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>
<-li,-an,ar>	15. Arbitrary static haptics Dynamic arbitrary haptics	<-li,-an,ar,sta/dyn, hap>
	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>
<-li,-an,-ar>	19. Static haptic structures Dynamic haptic structures	<-li,-an,-ar,sta/dyn, hap>
	20. Dynamic graphic structures	<-li,-an,-ar,dyn,gra>
SUPER LEVEL	**GENERIC LEVEL**	**NOTATION**

abstraction at which modalities are being characterised at the generic level. At the generic level, for instance, analogue static graphic *images* cannot be distinguished from analogue static graphic *graphs,* but to an interface designer these two modalities are being used for rather different information representation purposes. In another example, static graphic written *text* is useful for rather different purposes than is static graphic written *notation.* It should be remembered that the generic level is simply a pruned result of combining a small number of basic properties (Section 4). To achieve the intuitiveness required, we need to descend at least one level in the abstraction hierarchy of the taxonomy. This is done by adding further basic property distinc-

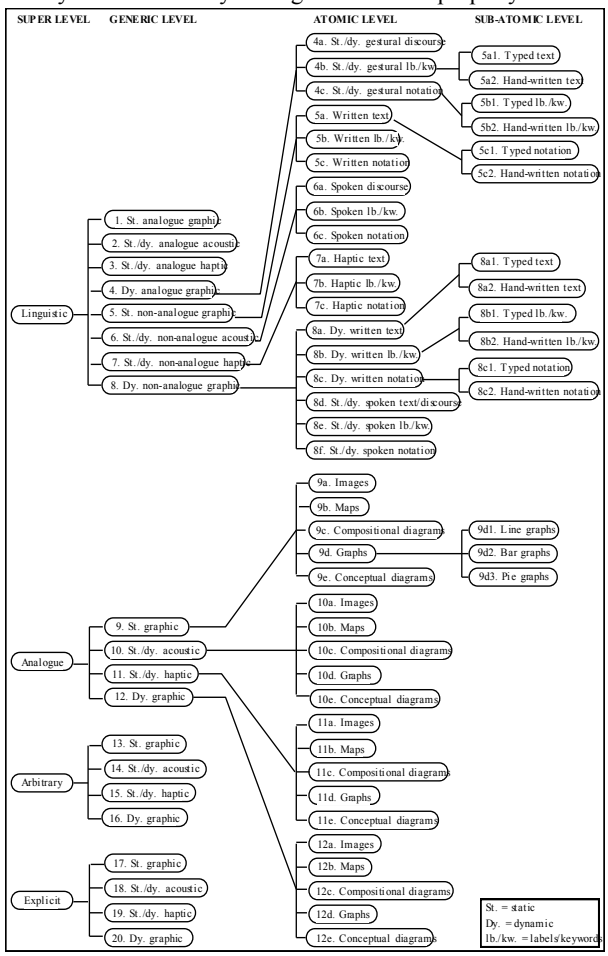

Figure 1. The taxonomy of unimodal output modalities. The four levels are, from left to right: super level, generic level, atomic level and sub-atomic level.

tions - just like when further distinctions are being added among sentence parts in generative grammar - thereby generating the atomic level of the taxonomy as presented in the static graphic conceptual diagram in Figure 1. The sub-atomic level in Figure 1 will be described in Section 6.

Given the diversification among modalities achieved at the super and generic levels, the novel basic properties that have been introduced to generate the atomic level are specific to the super and generic level fragments of the taxonomy to which they belong. Thus, the atomic level of the linguistic fragment of the taxonomy has been generated from the basic properties of *text, discourse, labels/keywords* and *notation* (Table 4). The atomic level of the analogue fragment of the taxonomy has been generated from the basic properties of *diagram, image, map, compositional diagram, graph* and *conceptual diagram* (Table 5). With respect to the atomic level of the arbitrary and explicit structure fragments of the taxonomy, no further basic properties were needed, with the result that the atomic level remains identical to the generic level for these fragments (Tables 6 and 7).

The generation of the atomic level follows the same principles as that of the generic level. The new distinctions introduced have been selected such as to support the generation of importantly different representational modalities which satisfy the intuitiveness requirement described in Section 3. In addition, pragmatic reductions have been performed in order not to proliferate atomic modalities beyond those necessary in practical interface design, thus satisfying the relevance requirement from Section 3. In what follows, justifications will be presented for each super level segment of the generation of atomic modalities, starting with the linguistic modalities.

5.1 Linguistic atomic modalities

Two types of distinction have gone into the generation of the atomic level linguistic modalities. The first type of distinction includes distinction between (a) text and discourse and (b) text or discourse, labels/keywords and notation. As to (a), it is a well-known fact that, grammatically speaking, written and spontaneous spoken language behave rather differently. This is due, we hypothesise, to the deeper fact that written language has evolved to serve the purpose of *situation independent* linguistic communication. The recipient of the communication would normally be in a different place, situation and time when decoding the written message. By contrast, spoken language has evolved to serve *situated* communication, the partners in the communication sharing location, situation and time. Hybrid uses of spoken and written language, such as telephone conversation or on-line e-mail dialogue are partially awkward forms of communication. In telephone conversation, the shared location is missing completely and the shared situation is missing more or less. In on-line e-mail dialogue, temporal independence is missing and some situation-sharing may be present. Situated linguistic communication has been termed *discourse* and situation-independent linguistic communication has been termed *text* (cf. Table 4). Videophone communication comes closer to discourse than does telephone communication because videophones establish more of a shared situation than telephones do. Normal e-mail communication comes closer to text exchange than on-line e-mail dialogue because normal e-mail communication is independent of partners' place, situation and time.

The distinction (b) between text or discourse, labels/keywords and notation is straightforward and important. *Text* and *discourse* have unrestricted expressiveness within the basic limitations to linguistic expressiveness in general (cf. Section 3). Discourse and text, however, tend to be too lengthy for being suited to the brief expression of focused information in menu lines, graph annotations, conceptual diagrams etc. across media. *Labels* or *keywords* are well-suited and widely used for this purpose. Their drawback is their inevitable ambiguity which, at best, may be reduced by the context in which they appear. Whereas text, discourse and keywords are well-suited for representing information to any user who understands the language used, *notation* is for specialist users and always suffers from limited expressiveness compared to text and discourse. Text, discourse, labels/keywords and notation thus have importantly different but well-defined roles in interface design across media and the static-dynamic distinction.

The second type of distinction involved in generating the atomic level is empirical in some restricted sense of the term. That is, once the above distinctions have been made, it becomes an empirical matter to determine which important types of atomic linguistic modalities there are. This again means that modality theory might so far have missed out on some important type of linguistic communication. However, Table 4 probably presents all the important ones. In fact, the search restrictions imposed by the taxonomy does seem to enable close-to-exhaustive search in this case. When output by current machines, *gestural language* (4a-4c) is (mostly) dynamic and always graphic. Gesturing robots are not addressed by the output modality taxonomy. Static gestural language is included in 4a-4c (see below). 5a-5c covers the most basic form of textual language, i.e. *static graphic written language*. The distinction between typed and hand-written static graphic written language belongs to the sub-atomic level (see Section 6). 6a-6c include the most basic form of discourse, i.e. *spoken language*. 7a-7c include static and dynamic haptic language, such as Braille. Section 8 of Table 4 illustrates the empirical nature of atomic level generation. One might have thought that dynamic (non-analogue sign) graphic language simply includes 8a-8c, i.e. the dynamic version of section 5, such as scrolling text. It turns out, however, that section 8 also includes graphically represented spoken language (speaking faces), whether this be read-aloud text, discourse, labels/keywords or notation (8e-8f). The latter modalities have gained favour recently as supporting the disambiguation of otherwise not easily understood synthetic speech [23].

The pragmatic reductions of the linguistic atomic modalities are straightforward. As argued in Section 4, the fact that some written language uses analogue signs is ultimately insignificant compared to the fact that written language is a syntactic-semantic-pragmatic system of meaning. Written hieroglyphs and other iconographic expressions, whether static or dynamic, graphic or haptic (sections 1, 3 and 4 of Table 4), may therefore be reduced to their non-analogue, non-iconographic counterparts without effects on interface design. The "glyphs" which have been invented for expressing multi-dimensional data points in graph space are rather forms of arbitrary static graphic modalities ([17], see below). Analogue speech sounds, by contrast (section 2 of Table 4), constitute a genuine sub-class of speech. As such, they have been pragmatically included in section 6 of Table 4. Static gestural language (section 1 of Table 4) has been fusioned with dynamic gestural language (section 4). Finally, the static graphic spoken language atoms (section 5) have been pragmatically fusioned with their dynamic counterparts (section 8). The result of this comprehensive set of reductions is shown as six triples of atomic linguistic modalities in Figure 1 above. As already remarked, we have all the prerequisites for creating more atomic modalities than those of Table 4, but the point in doing so is not clear when our purpose is a usable theory for interface design support.

5.2 Analogue atomic modalities

The analogue atomic modalities (Table 5) have been generated without any pragmatic modality fusion. The generation is based on the concept of a *diagram* and the distinction between (a) *images,* (b) *maps,* (c) *compositional diagrams,* (d) *graphs* and (e) *conceptual diagrams*. Diagrams subsume maps (b), compositional diagrams (c) and conceptual diagrams (e). The distinction between (a), (b), (c), (d) and (e) has been applied across the domain of

analogue representation, whether static or dynamic, graphic, acoustic or haptic. How can these distinctions be justi-

Table 4. The atomic level unimodal linguistic modalities with pragmatic fusions shown.

GENERIC LEVEL	ATOMIC LEVEL
1. Static analogue sign graphic language	Static gesture included in 4 a-c. Static text, labels/keywords, notation included in 5 a-c.
2. Static analogue sign acoustic language Dynamic analogue sign acoustic language	Included in 6 a-c.
3. Static analogue sign haptic language Dynamic analogue sign haptic language	Included in 7 a-c.
4. Dynamic analogue sign graphic language	Dynamic text, labels/keywords, notation included in 8 a-c. 4a. Static/dynamic gestural discourse 4b. Static/dynamic gestural labels/keywords 4c. Static/dynamic gestural notation
5. Static non-analogue sign graphic language	Static graphic spoken text, discourse, labels/ keywords, notation included in 8d-f. 5a. Static graphic written text 5b. Static graphic written labels/keywords 5c. Static graphic written notation
6. Static non-analogue sign acoustic language Dynamic non-analogue sign acoustic language	6a. Static/dynamic spoken discourse 6b. Static/dynamic spoken labels/keywords 6c. Static/dynamic spoken notation
7. Static non-analogue sign haptic language Dynamic non-analogue sign haptic language	7a. Static/dynamic haptic text 7b. Static/dynamic haptic labels/keywords 7c. Static/dynamic haptic notation
8. Dynamic non-analogue sign graphic language	8a. Dynamic graphic written text 8b. Dynamic graphic written labels/keywords 8c. Dynamic graphic written notation 8d. Static/dynamic graphic spoken text or discourse 8e. Static/dynamic graphic spoken labels/keywords 8f. Static/dynamic graphic spoken notation

fied as being the right ones for carving up the vast and complex domain of analogue representation at the atomic level? For a start, it may probably be acknowledged that the concepts of images, maps, compositional diagrams, graphs and conceptual diagrams are intuitively distinct and meaningful, and, as such, fulfil the intuitiveness and relevance requirements from Section 3. However, three more questions need to be addressed. The first is whether the space of analogue atomic representation should be carved up in an entirely different way (orthogonality). The second is whether the five concepts at issue exhaust the space of analogue atomic representation (completeness). The third question, which is also to do with orthogonality, is how these concepts are defined so as to avoid overlaps and confusion when they are being applied to concrete instances in design practice, i.e. how distinct and mutually exclusive are these concepts in practice? Let us begin with the third question.

The exclusiveness issue is particularly difficult in the analogue domain. The problem with exclusiveness in the analogue domain is that representations belonging to one category, such as images, can often be manipulated to become as close as desired to representations belonging to several other categories, such as compositional diagrams. This *continuity of representation* is a well-known characteristic of many ordinary concepts and has been empirically explored in prototype theory [28]. The point is that classical definitions using jointly necessary and sufficient conditions for specifying when an instance belongs to some category, do not work well in the analogue domain. Instead, concept definitions have to rely on a combination of reference to prototypical instances (or paradigm cases) of a category combined with characterising descriptions that include pointers to contrasts between different categories. An important implication is that the concepts of atomic modalities of modality theory cannot be fully intuitive in the sense of completely corresponding to our standard concepts. For instance, one of our present prototypical concepts of a static graphic image is the concept of a well-resembling 2D photograph of a person or landscape. However, static graphic images are also 3D or 1D, and these differ from those prototypes. In other words, modality theory can only meet the completeness requirement of Section 3 through some amount of analytic generalisation. We shall see how the concept characterisations in the analogue domain work using abbreviated versions of the concept characterisations of modality theory which often run several pages per concept, excluding illustrations.

A *diagram* may be briefly defined as an analytic analogue representation. A diagram provides an analytic account of its subject-matter rather than an account of its mere appearance.

An *image* is an analogue representational modality which imitates or records the external form of real or virtual objects, processes and events by representing their physical appearance rather than serving analytical or decompositional purposes such as those served by compositional diagrams. In the limit, images allow realistic quasi-perception of the rich specific properties of objects, processes and events, which cannot easily be represented linguistically (cf. Section 3). Images vary from high-dimensionality, maximally specific images to images whose specificity has been highly reduced (‘sketches’) for some purpose. Depending on the medium, images may represent non-perceivable objects, processes and events, whether these be too small, too big, too remote, too slow, too fast, beyond the human sensory repertoire or normally hidden beneath some exterior, to be perceived by humans. Images may also represent objects in a medium different from its 'normal' physical medium, e.g. by representing acoustic information graphically. Because images, on their own, represent unfocused, association-rich 'stories', linguistic annotation is often needed to add focus and explanatory contents to the information they provide. In addition, many types of image, such as medical X-ray images, microscope images or many types of sound pattern, require considerable skill for their interpretation.

We observe from this definition that images are being contrasted to their closest neighbour in analogue modality space, i.e. compositional diagrams (see below). Furthermore, we note that images have limited value as stand-alone unimodal representations. For many interface design purposes, images need

Table 5. The atomic level unimodal analogue modalities.

GENERIC LEVEL	ATOMIC LEVEL

9. Static analogue graphics	9a. Static graphic images 9b. Static graphic maps 9c. Static graphic compositional diagrams 9d. Static graphic graphs 9e. Static graphic conceptual diagrams
10. Static analogue acoustics Dynamic analogue acoustics	10a. Static/dynamic acoustic images 10b. Static/dynamic acoustic maps 10c. Static/dynamic acoustic compositional diagrams 10d. Static/dynamic acoustic graphs 10e. Static/dynamic acoustic conceptual diagrams
11. Static analogue haptics Dynamic analogue haptics	11a. Static/dynamic haptic images 11b. Static/dynamic haptic maps 11c. Static/dynamic haptic compositional diagrams 11d. Static/dynamic haptic graphs 11e. Static/dynamic haptic conceptual diagrams
12. Dynamic analogue graphics	12a. Dynamic graphic images 12b. Dynamic graphic maps 12c. Dynamic graphic compositional diagrams 12d. Dynamic graphic graphs 12e. Dynamic graphic conceptual diagrams

linguistic annotation with the result that the combined representation becomes bimodal. More generally, unimodal modalities may be roughly distinguished into *independent unimodal modalities* which can do substantial representational work on their own, and *dependent unimodal modalities* which need other modalities if they are to serve any, or most, representational purposes. Text, discourse and image modalities, for instance, are among the most independent unimodal modalities there are.

Compositional diagrams are 'analytical images', i.e. they are analogue representations which represent, using image elements, the structure or decomposition of objects, processes or events. The decomposition is standardly linguistically labelled. Compositional diagrams focus on selective part-whole decomposition into structure and function. The combination of analogue representation and linguistic annotation in compositional diagrams may vary from highly labelled diagrams containing rather abstract (i.e. reduced-specificity) analogue elements to highly image-like diagrams containing a modest amount of labelling. Highly labelled and abstract compositional diagrams, or compositional diagrams combining the representation of concrete and abstract subject-matter, may occasionally be difficult to distinguish from conceptual diagrams (see below). To serve their analytic purpose, compositional diagrams standardly involve important reductions of specificity, and often use focusing mechanisms, saliency enhancement and dimensionality reduction [6]. These selection mechanisms are used in order to optimise the compositional diagram for representing certain types of information rather than others.

We see that, much more than images, compositional diagrams depend on linguistic annotation to do their representational job. And we note again how compositional diagrams are being contrasted to their closest neighbours in analogue representation space, i.e. images and conceptual diagrams.

Maps are a species of compositional diagrams, defined by their domain of representation. Maps provide geometric information about real or virtual physical objects and focus on the relational structure of objects and events, in order to provide locational information about parts relative to one another and to the whole. A prototypical map is a reduced-scale, reduced-specificity 2D graphic representation of part of the surface of the Earth, showing selected, linguistically labelled features such as rivers, mountains, roads and cities, and having been designed to enable travellers to find the right route between geographical locations. Maps may otherwise represent spatial layout of any kind, being on occasion difficult to distinguish from images and (other) compositional diagrams.

Maps are thus a species of compositional diagrams and share most of the properties of these as described above. Maps have been included in the taxonomy because of being quite common and application-specific and because of the robustness of the map concept. We seem to think in terms of maps rather than in terms of a-certain-sub-species-of compositional diagrams. A taxonomy of unimodal analogue modalities which ignores this fact is likely to be less useful than a taxonomy which respects the fact while preserving, at the same time, analytic transparency.

Graphs represent quantitative or qualitative information through the use of abstract analogue means which standardly bear no recognisable similarity to the subject-matter or domain of the representation. The quantitative information is statistical information or numerical data which may either be gathered empirically or generated from theories, models or functions. Use of analogue representation makes graphs well-suited for facilitating users' identification of global data properties through making comparisons, perceiving data profiles, spotting trends among the data, perceiving temporal developments in the data and/or discovering new relationships among data, and hence supports the analysis of, and the reasoning about, quantitative information. Whilst quantitative data can in principle be represented linguistically and are often presented in tables (see below), the focused and non-specific character of linguistic representation makes this form of representation ill-suited to facilitate the interpretation of global data properties. Given their primarily abstract analogue nature, graphs virtually always require clear and detailed linguistic annotation for their interpretation, consistent with the analogue representation. Graphs are thus in practice at least bimodal modalities. Graphic graphs frequently incorporate graph space grids and other explicit structures, which makes them trimodal modalities. The huge diversity of graph representations requires a sub-atomic expansion of at least some of the graph nodes of the taxonomy (see Section 6). Graphs clearly exemplify the dependent unimodal modalities which strictly need linguistic annotation in order to be intelligible. The graph notion is quite robust and does not require contrasting with other analogue modalities - it has no close neighbours.

Conceptual diagrams use various analogue representational elements in representing the analytical decomposition of an abstract entity such as an organisation, a family, a theory or classification, or a conceptual structure or model. Conceptual diagrams thus enhance the linguistic representation of abstract entities through analogue means which facilitate the perception of structure and relationships. Conceptual diagrams constitute an abstract counterpart to compositional diagrams. The abstract, not primarily spatio-temporal representational purpose and the decompositional purpose of conceptual diagrams jointly mean that conceptual diagrams require ample linguistic annotation and hence are at least bimodal. The role of analogue elements in conceptual diagrams is to make the diagram's abstract subject-matter more easily accessible through spatial and/or temporal structure and

layout. The abstract subject-matter of conceptual diagrams requires that the information they represent is to a very important extent being carried by the linguistic modalities involved. Figure 1 in Section 5 shows a prototypical (bimodal) conceptual diagram.

Like graphs, conceptual diagrams are dependent unimodal modalities which always require linguistic annotation in order to enable proper decoding.

In presenting the analogue atomic modalities, we have so far concentrated on the question of exclusiveness raised in the beginning of the present section. Two further questions were raised. One was whether the space of analogue atomic representation might, or even should, be carved up in an entirely different way. Modality theory assumes four categories of analogue representation: images, compositional diagrams (including maps), graphs and conceptual diagrams. In an empirical study, Lohse et al. [20] (analysed in Bernsen [2]), found that subjects tended to robustly categorise a variety of analogue 2D static graphic representations into the categories 'network charts', 'diagrams', 'maps', 'icons', and 'graphs/tables'. 'Network charts' correspond to the conceptual diagrams of modality theory, 'diagrams' to compositional diagrams, 'maps' to maps and 'graphs' to graphs. As no images were presented to the subjects in the study of Lohse et al., we can ignore images in what follows. Apart from 'icons' and 'tables', the correspondence between the result of Lohse et al. and the present taxonomy is very close indeed. What is the status of icons and tables in modality theory?

In modality theory, *tables,* although clearly distinct from any of the atomic modalities considered above, are not viewed as constituting a separate modality of representation but as a convenient way of spatially structuring information as represented in most graphic or haptic modalities. Tables, like *lists,* are thus *modality structures* rather than modalities. They are often bimodal, as in prototypical 2D static graphic tables which combine typed language with explicit structures, such as the tables in the present paper. That the subjects in Lohse et al. [20] combined graphs and tables into one category is probably due to the fact that graph information can often, if not always, be represented in tables, and vice versa. However, this fact is of no help to an interface designer whose task it is to optimise the representation of information in context. Depending on the nature of that information, graphs may be preferable to tables, or vice versa [30].

Like lists and tables, icons are not viewed as constituting a separate modality. Rather, icons represent an extreme generalisation of the notion of labels/keywords. This generalisation reaches far beyond 'icons' in the standard sense of static 2D graphic representations. Like a label or keyword, an *icon* is a singular representation or expression, which normally has one intended meaning only, and which is subject to ambiguity of interpretation. *Any* modality token, it appears, can be used as an icon, even a piece of text. Being an icon is, rather, a specific *modality role* which can be assumed by any modality token. It would therefore be misleading to consider icons as a separate kind of modality. This means that icons are covered by the taxonomy to the extent that the taxonomy is complete.

In conclusion, the correspondence between the present taxonomy and the empirical results of Lohse et al. [20], is remarkable. Until someone comes up with an entirely different taxonomy of the space of analogue representation, the present taxonomy would appear to be at least empirically confirmed as to its orthogonality and relevance as well as being intuitively plausible.

The second question raised above was whether the four concepts of images, compositional diagrams (including maps), graphs and conceptual diagrams exhaust the space of analogue atomic representation. The results of Lohse et al. [20] confirm this assumption (cf. above). It should be kept in mind, however, that exhaustiveness does not imply exclusiveness. We have seen that classical-style definitions of analogue modalities are hardly possible. This implies that borderline cases will inevitably occur. But if classical-style definitions are impossible, *any* taxonomy of analogue modalities will be subject to the existence of borderline cases which are difficult to categorise unambiguously. What matters is that the number of borderline cases is relatively small and that it is possible to clearly state on which borderline between which specific analogue atomic modalities a particular borderline case lies. Finally, the downwards extensibility of the atomic level of the taxonomy means that there is still a richness of different sub-atomic modalities to be discovered. As it stands, the taxonomy only addresses this richness in a few cases (see Section 6).

5.3 Arbitrary atomic modalities

The arbitrary unimodal atomic modalities are simple to deal with because, so far, at least, no reason has been found to introduce new distinctions in order to generate the atomic level (see Table 6). *Arbitrary modalities* express information through having been defined ad hoc at their introduction. This means that arbitrary modalities do not rely on an already existing system of meaning. Arbitrary modalities are therefore non-linguistic and non-analogue by definition. As argued in Section 4, it is against the purpose of the taxonomy that non-arbitrary modalities be used arbitrarily. This imposes severe restrictions on which representations may be used arbitrarily. Nonetheless, arbitrary modalities can be quite useful for the representation of information. In general, any information channel in any medium can be arbitrarily assigned a specific meaning in context. This operation is widely used in the expression of information in compositional diagrams, maps, graphs and conceptual diagrams. In another example, arbitrary modalities are often used to express alarms in cases where the only important point about the alarm is its relative saliency in context.

Table 6. The atomic level unimodal arbitrary modalities are identical to those at the generic level.

GENERIC LEVEL	ATOMIC LEVEL
13. Arbitrary static graphics	See generic level
14. Arbitrary static acoustics Dynamic arbitrary acoustics	See generic level
15. Arbitrary static haptics Dynamic arbitrary haptics	See generic level
16. Dynamic arbitrary graphics	See generic level

5.4 Explicit structure atomic modalities

As in the case of arbitrary atomic modalities, no reason has so far been found to introduce new distinctions in order to generate the explicit structure modalities at the atomic level (see Table 7). *Explicit structure modalities* express information in the limited but important sense of explicitly marking separations between modality tokens. Explicit structure modalities rely on an already existing system of meaning and are therefore non-arbitrary. This is

because the purpose of explicit markings are immediately perceived. Explicit structure modalities are non-linguistic and non-analogue. Despite the modest amount of information conveyed by an explicit structure, these structures play important roles in interface design. One such role is to mark distinction between different groupings of information in graphics and haptics. This role antedates the computer. Another, computer-related role is to mark functional differences between different parts of a graphic or haptic representation. Static graphic windows, for instance, are based on arbitrary structures which inform the user about the different consequences of interacting with different parts of the screen.

Table 7. The atomic level unimodal explicit structure modalities are identical to those at the generic level.

GENERIC LEVEL	ATOMIC LEVEL
17. Static graphic structures	See generic level
18. Static acoustic structures Dynamic acoustic structures	See generic level
19. Static haptic structures Dynamic haptic structures	See generic level
20. Dynamic graphic structures	See generic level

The claim, or hypothesis, with respect to the atomic level of the taxonomy of unimodal output modalities, is a rather strong one. It is that the atomic level fulfils the requirements of completeness, orthogonality, relevance and intuitiveness stated in Section 3 above. Any multimodal output representation can be exhaustively characterised as consisting of a combination of atomic-level modalities (se Section 9).

Assuming that the atomic level of the taxonomy of unimodal modalities has been successfully generated, an interesting implication follows. Space has not allowed the definition of each individual atomic modality presented in Tables 4 through 7 above. What have been described are the principles that were applied in generating the atomic level and the novel distinctions introduced in the generation. However, what has been generated surpasses the apparatus described above. This is because the distinctions introduced in generating the atomic level get "multiplied" by the static/dynamic distinction and the distinction between different media of expression. The specific atomic modalities are the results of this multiplication. Each atomic modality is distinct from any other and has a wealth of properties. Some of these are inherited from the modality's parent nodes at higher levels of abstraction in the taxonomy. Other properties specifically belong to the atomic modality itself and serve to distinguish it from its atomic-level neighbours. One way to briefly illustrate this generative power of the taxonomy is to focus on atomic modalities which are yet to become used in interface design; which have not yet received a separate identification as representational modalities; or which are so "exotic" as to appear difficult to exemplify for the time being.

Like any other atomic modality, *gestural notation* is a possible form of information representation. Except for use in brief messages, examples of gestural notation may be hard to find. The reason probably is that notation, given its non-naturalness as compared to natural language, normally requires freedom of perceptual inspection to be properly decoded. Like *spoken language notation*, gestural notation would normally be dynamic and hence does not allow freedom of perceptual inspection. This leads to the prediction that, except for brief messages in dynamic notation, *static* gestural and spoken notation would be the more usable varieties. For the same reasons, there would seem to be little purpose in using lengthy dynamic written notation, except for specialists capable of decoding such notation on-line. Such specialists might find uses for lengthy gestural and spoken notation as well. If the (acoustic) spoken notation is expressed as synthetic speech, the specialists might need support from graphic spoken notation (i.e. from a speaking face on the screen) in order to properly decode the information expressed.

In the analogue atomic modalities domain, *acoustic images* are becoming popular, e.g. in the 'earcon' modality role. *Acoustic graph-like images* have important potential for representing information in many domains other than, e.g., those of the clicking Geiger counter or the pinging sonar. The potential of *acoustic graphs* proper would seem be largely unexplored. *Acoustic maps* appear to have some potential in representing spatial layout. *Acoustic compositional diagrams* are interesting. Think, for instance, of a system for the training of novice car repair persons. Sound diagnosis plays an important role in the work of skilled car repairers. The training system might take apart the relevant diagnostic noises into their components, explain the causes of the component sounds and finally put these together again in training-and-test cycles. *Acoustic conceptual diagrams* are a fascinating subject but their application potential is unclear. For technological reasons, output *dynamic analogue haptics* appears to be mostly unexplored territory, whether in the form of images, maps, compositional diagrams, graphs or conceptual diagrams. *Dynamic analogue graphics* is extremely familiar to us but still has great unused potential. Full virtual reality will need to combine dynamic, perceptually rich analogue graphics, acoustics and haptics.

Arbitrary static graphics, acoustics and *haptics* are widely used already. It is much less obvious how much we shall need their dynamic counterparts in future applications. A ringing telephone, of course, produces arbitrary dynamic acoustics. Beyond such saliency-based applications, however, it is not entirely clear which information representation purposes might be served by the dynamic arbitrary atomic modalities.

Finally, in the explicit structure domain, *static graphic explicit structures* are as commonplace as static graphics itself. *Dynamic graphic explicit structures* are in use as focusing mechanisms, for instance, which encircle linguistic or analogue graphic information of current interest during multimodal graphics/spoken language presentations. *Static* and *dynamic haptic explicit structures* have unexplored potential for the usual (technological) reasons. As for *acoustic explicit structures*, we have had problems coming up with valid examples. It is common, for instance, in spoken language dialogue applications to use beeps to indicate that the system is ready to listen to user input. However, as these beeps do not rely on an already existing system of meaning, they rather exemplify the use of arbitrary dynamic acoustics.

6 SUB-ATOMIC-LEVEL UNIMODAL MODALITIES

Exhaustiveness at any level of the taxonomy is still limited by level of abstraction and hence by the number of basic properties which have been introduced to generate that level. One virtue of the taxonomy is its unlimited downwards extensibility. That is, once the need has become apparent to distinguish between dif-

ferent unimodal modalities subsumed by an already existing modality, further basic properties can be sought that might help generate the needed distinctions. Given the above strong claims on behalf of the atomic level, such needs are currently most likely to be felt at this level. Table 8 shows how the principle of extensibility has been applied to static and dynamic graphic written text through the simple distinction between *typing* and *hand-writing*. Table 9 shows what is still a hypothetical application of the principle in the domain of static graphic graphs. Static graphic graphs are extremely useful for representing quantitative information. The domain has been the subject of particularly intensive research for decades [10,14,19,30,31] with the result that the atomic modality 'static graphic graphs' has become much too coarse-grained a notion to handle the large variety of information representations that exist. However, there is still no consensus on a taxonomy of static graphic graphs. Given the experimental nature of Table 9 and the complexity of the issues involved, the matter will be left for later presentations of modality theory.

Table 8. The sub-atomic level unimodal graphic written language modalities.

ATOMIC LEVEL	SUB-ATOMIC LEVEL
5a. Static graphic written text	5a1. Static graphic typed text 5a2. Static graphic hand-written text
5b. Static graphic written labels/keywords	5b1. Static graphic typed labels/keywords 5b2. Static graphic hand-written labels/keywords
5c. Static graphic written notation	5c1. Static graphic typed notation 5c2. Static graphic hand-written notation
8a. Dynamic graphic written text	8a1. Dynamic graphic typed text 8a2. Dynamic graphic hand-written text
8b. Dynamic graphic written labels/keywords	8b1. Dynamic graphic typed labels/keywords 8b2. Dynamic graphic hand-written labels/keywords
8c. Dynamic graphic written notation	8c1. Dynamic graphic typed notation 8c2. Dynamic graphic hand-written notation

Table 9. The sub-atomic level unimodal static graphic graph modalities.

ATOMIC LEVEL	SUB-ATOMIC LEVEL
9d. Static graphic graphs	9d1. Line graphs 9d2. Bar graphs 9d3. Pie graphs

7 MODALITY ANALYSIS

Considered in isolation, the taxonomy of unimodal output modalities is primarily just that, a principled hierarchical analysis of the space of representational modalities in the media of graphics, acoustics and haptics. The taxonomy turns into modality *theory* proper when (a) its generative principles are being accounted for in more detail, (b) its basic properties have been analysed in depth, and (c) individual unimodal modalities have been analysed as to their properties and capabilities and limitations of representing different types of information in context. We have analysed all the unimodal modalities presented above and implemented them in a hypertext/hypermedia software demonstrator [7] which is currently being ported to the WWW. The analysis of each modality is represented using a modality document template. Modality documents define, explain, analyse and illustrate the unimodal modalities from the point of view of IMP systems and interface design support. The shared document structure includes the following entries:

- *Modality profile, information channels and dimensionality.* The modality profile is expressed in the notation introduced in Table 2. Information channels and dimensionality, such as 1D or time, are properties of the medium in which a particular modality is being expressed.
- *Inherited declarative and functional properties.* Each modality inherits part of its properties from its parent nodes in the taxonomy. To keep individual modality documents short, these properties must be retrieved through hypertext links. Declarative properties describe the modality independently of its use. Functional properties state which types of information the modality is good or bad at representing in context. The following example shows the list of links to inherited properties in the atomic-level gestural notation modality document (Table 4). Hypertext links are underlined:
  - linguistic modalities
  - static modalities
  - dynamic modalities
  - graphic modalities
  - notation
  - Static graphics have the following information channels: shape, size (length, width, height), texture, resolution, contrast, value (grey scales), colour (including brightness, hue and saturation), position, orientation, viewing perspective, spatial arrangement, short-duration repetitive change of properties.
  - Dynamic graphics have the following information channels in addition to those of static graphics: non-short-duration repetitive change of properties, movement, displacement (relative to the observer), and temporal order.
  - The dimensionality of dynamic graphics is 1-D, 2-D and 3-D spatial, time.

  Gestural notation thus inherits the properties of the linguistic, static, dynamic, graphic and notational modalities. Since the information channel and dimensionality information is important to have close-at-hand, it is repeated in the document rather than having to be retrieved through hypertext links. Because of the pragmatic node reduction strategy (Section 5), the gestural notation document presents both static and dynamic gestural notation.
- *Specific declarative and functional properties.* Each modality, being a combination of basic properties, has properties of its own in addition to those it has inherited. These are the properties which characterise the modality as being specifically different from its sister modalities with which it shares a common ancestry. For instance, in the arbitrary modality document (super level), the entry on 'Specific declarative and functional properties' includes the point that "Arbitrary modalities express information through having been defined ad hoc at their introduction." This implies that information represented in arbitrary modalities, whether graphic, acoustic or haptic, in order to be properly decoded by users, must be introduced in some non-arbitrary modality, such as some linguistic modality or other.

- *Information mapping rules.* These rules represent functional analyses of each modality and express which types of information that modality is suited or unsuited for representing. The rules are similar in many respects to production rules. We have been exploring for some time a methodology for applying the rules to the design of IMPs [5,8,9]. One of the information mapping rules in the static graphic image document is:

  > Facilitate the visual identification of objects, processes, or events <->
  > Consider including high-specificity static graphic images in as high dimensionality and resolution as possible.

  This rule effectively states that static graphic images are good tools for supporting the identification of objects, and that identification is further enhanced through high specificity (a large amount of detail in as many information channels as possible), high dimensionality (2 1/2D or 3D better that 2D), and high image resolution. The rule is read from left to right as an if-then rule. Read from right to left, the rule says that "Modality X is good at representing Y". An illustration of this rule, and hence of one of the advantages of the static graphic image modality, is the use of photographs in criminal investigation. It is virtually impossible to linguistically express what a person looks like in such a way that the person may be uniquely identified from the linguistic description [6]. Use of static graphic images can make this an effortless undertaking. Indeed, a picture can sometimes be worth more than a thousand words. Or, rather, this proverbial classic not only applies to pictures but to analogue representations in general, irrespective of whether they are embodied in graphics, acoustics or haptics.
- *Combinatorial analysis.* These analyses express which other unimodal modalities a particular modality may or may not be combined with to compose multimodal representations. For instance, in the modality document on explicit static graphic structures, the combinatorial analysis states that "explicit static graphic structures combine well with any static or dynamic graphic modality, whether linguistic, analogue or arbitrary". Combinatorial analysis is highly important to the discovery of patterns of compatibility and incompatibility between unimodal modalities. Such patterns would begin to constitute a (unimodal) modality combination "grammar" (see Section 8).
- *Relevant operations.* Each modality can be subjected to a number of operations, such as, in analogue graphics and haptics, dimensionality reduction. Normal road maps, for instance, reduce the topology from 3D to 2D. An operation may be defined as a meaningful addition, reduction, or other change of information channels or dimensionality in a representation instantiating some modality. The purpose of an operation normally is to bring out more clearly particular aspects of the information to be presented. Other examples in the domain of analogue graphic modalities are *specificity reduction,* as in replacing an image with a sketch; *saliency enhancement,* as in selective colouring; and *zooming*. Similarly, **boldfacing**, *italicizing* and underlining are common operations in graphic typed languages [6].
- *Illustrations.* A very important part of the demonstrator is to illustrate each modality using annotated prototypical and less prototypical examples. These illustrations serve to demonstrate the points made elsewhere in a particular modality document.

In addition to the modality documents, a *modality lexicon* introduces the technical terms applied during modality analysis, such as 'salience' or 'information channel'. There are currently about 70 such documents (or concepts). Due to the heterogeneous nature of their topics, no rigid document structure has been enforced on lexicon documents.

8 MULTIMODAL GENERATION

The generation of the taxonomy of unimodal output modalities has been outlined above. Once the taxonomy and theory is in place, an entirely new type of generation becomes possible. This type of generation is not, as in taxonomy generation, an analytic or decompositional process of adding ever finer distinctions, but is a synthetic process of composition in which multimodal representations are being produced from unimodal representations. This opens the prospect of establishing a "chemistry" or "grammar" of modality theory, in which complex multimodal representations are being composed from their atomic and sub-atomic constituents according to principles derived from the combinatorial analysis of modalities (Section 7). In this process, and only limited by the levels of abstraction of the taxonomy itself, the taxonomy allows generation of all possible multimodal output modalities in the media of graphics, acoustics and haptics. Simple computation shows that the atomic and sub-atomic modalities described in this paper can be combined into multimodal expressions of information in thousands of different ways. The problem, therefore, is to create a principled basis for multimodal generation, which allows the generation of all and only those multimodal representations which are useful to IMP systems and interface design. We are currently investigating a "filtering" mechanism based on the study of all possible *pairs* of unimodal output modalities (all bimodal modalities). Multimodal representation can of course be n-modal. Now suppose that we are considering a multimodal representation in which n = 10 and that the filtering mechanism has identified a highly questionable bimodal combination [a,b]. If [a,b] occurs in the n-modal multimodal representation under consideration, chances are that this representation will fail as a design solution. However, as several have pointed out, the complementary strengths of information representation of different modalities might conceivably falsify this general hypothesis.

9 CONCLUSION

The empirical status of the taxonomy presented above merits further comment. We have seen how the intuitiveness and relevance requirements (Section 3) have been used in generating the taxonomy. However, even granted the intuitiveness of the taxonomy as it stands, there might be representations out there which turn out to prove recalcitrant to classification and reveal new dimensions of relevance whilst preserving intuitiveness. This is where further work of the type presented in [20] might prove interesting. Furthermore, a novel type of work is needed to test for completeness and orthogonality (Section 3). This work will have to analyse large samples of multimodal material to test whether their tokens can be exhaustively described as consisting of one or

more unimodal modalities and can be thus described in only one way. This is ongoing work.

As repeatedly said above, the ultimate aim of modality theory development is to provide practical support for IMP systems and interface design. The generation of novel unimodal modalities was discussed in Section 5. Ongoing work on information mapping was briefly mentioned in Section 7. A case study in preparation may illustrate the potential usefulness of modality theory. The case study addresses the thorny issue of the functionality of speech input and/or output based on 120 different claims made in the literature. These claims are extremely different along many dimensions, some appealing to properties of the work environment, others to properties of the task, yet others to certain performance parameters or cognitive properties for which the application should be optimised. Their sources of evidence vary from intuition through usability observation to laboratory experiments, and the claims differ widely in generality. In fact, their only commonality is that they all address modality theory issues. Preliminary results show that a mere 18 properties of unimodal modalities, such as "acoustic modalities are omnidirectional", suffice to justify 83,5% of the (109) claims which were not false or unclear, support 14% of these claims and correct 7,5% of the total number of claims. This suggests that proper understanding, during early design, of a limited set of modality properties might enable designers to largely (but not fully) dispense with the complex, patchy, imperfect and difficult-to-obtain knowledge embodied in the 120 claims collected from the literature.

Just like the downwards extension of the taxonomy of unimodal output modalities, modality analysis and multimodal generation are open-ended and collaborative endeavours. Large amounts of results are currently being produced across the world on the information representation capabilities of individual unimodal modalities and their multimodal combinations. Novel modalities are being investigated and cases prepared for making additional downwards extensions of the taxonomy. If valid and practically useful, the candidate reference model for output information representation in IMP systems presented above might act as a common frame of reference for this work.

A key advantage of the coming generation of IMP systems is their augmented user-system *interactivity* as compared with current systems. Whereas the approach to output modalities presented above may serve the design of system *presentations* of information, it has nothing to say about interactivity and input modalities. Interactivity is a more complex problem than (output) representation, both from an information, a device and a software engineering perspective. Output modality theory invites an approach to interactivity via the addition of a theory of input modalities. Interactivity is viewed as sequences of input/output information exchanges in which user information is being input into the output domain of the system as represented through its output modalities. A mouse-click, for instance, would represent the user's input of information into some part of a graphic output representation. We have begun explorative developments of a theory of input modalities. Not surprisingly, given the complexity of the problem, we found that significantly less work has been done on input modalities (e.g. [21]) compared to the results that are available on output [33].

ACKNOWLEDGEMENTS

Thanks are due to Laila Dybkjær who developed Figure 1 and provided valuable comments and criticisms along the way. Thanks are also due to the four anonymous reviewers who made many constructive points which, I hope, have now been addressed in the revised version of the paper. The work reported was done on grants from the Esprit Basic Research projects GRACE and AMODEUS-2 and from the Danish Natural Sciences Research Council whose support is gratefully acknowledged.

REFERENCES

[1] N. Bellalem, and L. Romary, Reference Interpretation in a Multimodal Environment Combining Speech and Gesture, *Proceedings of the First International Workshop on Intelligence and Multimodality in Multimedia Interfaces*, Edinburgh, Scotland, 1995.

[2] N. O. Bernsen, Matching Information and Interface Modalities. An Example Study, *Working Papers in Cognitive Science* WPCS-92-1, Centre for Cognitive Science, Roskilde University, 1992.

[3] N. O. Bernsen, Foundations of Multimodal Representations: A Taxonomy of Representational Modalities, *Interacting with Computers* Vol. 6, 4, 347-71, 1994.

[4] N. O. Bernsen, A Revised Generation of the Taxonomy of Output Modalities, *Esprit Basic Research project AMODEUS-2 Working Paper* RP5-TM-WP11. *CCI Working Papers in Cognitive Science and HCI,* WPCS-94-7, Centre for Cognitive Science, Roskilde University, 1994.

[5] N. O. Bernsen, Information Mapping in Practice. Rule-Based Multimodal Interface Design, *Proceedings of the First International Workshop on Intelligence and Multimodality in Multimedia Interfaces*, Edinburgh, Scotland, 1995.

[6] N. O. Bernsen, Why are Analogue Graphics and Natural Language both Needed in HCI? In F. Paterno (Ed.), *Interactive Systems: Design, Specification, and Verification,* Focus on Computer Graphics, Springer Verlag, 235-51, 1995.

[7] N. O. Bernsen, and S. Lu, A Software Demonstrator of Modality Theory. In Bastide, R. and Palanque, P. (Eds.), *Design, Specification and Verification of Interactive Systems '95,* Springer-Verlag, 242-61, 1995.

[8] N. O. Bernsen, and S. Verjans, Information Mapping. Knowledge-Based Support for User Interface Design, *Proceedings of the CHI '95 Workshop on Knowledge-Based Support for the User Interface Design Process,* Denver, Colorado 1995.

[9] N. O. Bernsen, and S. Verjans, From Task Domain to Human-Computer Interface. Exploring an Information Mapping Methodology, *AAAI Press* 1996 (to appear).

[10] J. Bertin, *Semiology of Graphics. Diagrams. Networks. Maps,* Madison, University of Wisconsin Press, 1983.

[11] F. Bodart, A. M. Hennebert, J.-M. Leheureux, I. Provot, G. Zucchinetti, and J. Vanderdonckt, Key Activities for a Development Methodology of Interactive Applications. In Benyon, D. and Palanque, P. (Eds.), *Critical Issues in User Interface Systems Engineering,* Springer-Verlag, 1995.

[12] N. Carbonell, (Ed.), Multimodal Human-Computer Interaction, *Proceedings of the ERCIM Workshop on Multimodal Human-Computer Interaction,* Nancy, France, 1994.

[13] N. Chomsky, *Syntactic Structures,* Haag, Mouton, 1957.

[14] N. Holmes, *Designer's Guide to Creating Charts and Diagrams,* New York, Watson-Guptill Publications, 1984.

[15] E. Hovy, and Y. Arens, 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.

[16] P. Johnson, (Ed.), Intelligent Multi-Media Multi-Modal Systems, *Proceedings of the AAAI Spring Symposium,* Stanford, 1994.

[17] C. Joslyn, C. Lewis, and B. Domik, Designing Glyphs to Exploit Patterns in Multidimensional Datasets, *CHI'95 Conference Companion,* 198-199, 1995.

[18] P. Lefebvre, G. Duncan, and F. Poirier, Speaking with Computers: A Multimodal Approach, *Proceedings of EUROSPEECH '93,* Berlin, 1665-68, 1993.

[19] A. Lockwood, *Diagram. A Visual Survey of Graphs, Maps, Charts and Diagrams for the Graphic Designer,* London, Studio Vista, 1969.

[20] G. Lohse, N. Walker, K. Biolsi, and H. Rueter, Classifying Graphical Information, *Behaviour and Information Technology* 10, 5, 419-36, 1991.

[21] J. Mackinlay, S. Card, and G. Robertson, A Semantic Analysis of the Design Space of Input Devices, *Human-Computer Interaction,* 5, 2-3, 145-90, 1990.

[22] J.-C. Martin, and D. Béroule, Multimodal Interfaces Based on Types and Goals of Cooperation Between Modalities, *Proceedings of the First International Workshop on Intelligence and Multimodality in Multimedia Interfaces*, Edinburgh, Scotland, 1995.

[23] D. W. Massaro, and M. M. Cohen, Auditory/Visual Speech in Multimodal Human Interfaces, *Proceedings of ICSLP '94,* 531-534, 1994.

[24] M. T. Maybury, (Ed.), *Intelligent Multimedia Interfaces,* Cambridge, MA, MIT Press, 1993.

[25] K. Mullet, and D. J. Schiano, 3D Or Not 3D: "More is Better" Or "Less is More"? *CHI'95 Conference Companion,* 174-175, 1995.

[26] C. M. Neuwirth, R. Chandhok, D. Charney, P. Wojahn, and L. Kim, Distributed Collaborative Writing: A Comparison of Spoken and Written Modalities for Reviewing and Revising Documents, *Proceedings of CHI'94,* 51-57, 1994.

[27] V. Z. Ogozalek, A Comparison of the Use of Text and Multimedia Interfaces to Provide Information to the Elderly, *Proceedings of CHI'94,* 65-71, 1994.

[28] E. Rosch, Principles of Categorization. In Rosch, E. and Lloyd, B. B. (Eds.), *Cognition and Categorization,* Hillsdale, NJ, Erlbaum, 1978.

[29] K. Stenning, and J. Oberlander, 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, Vol. 2, 607-21, 1991.

[30] E. R. Tufte, *The Visual Display of Quantitative Information,* Cheshire, CT, Graphics Press, 1983.

[31] E. R. Tufte, *Envisioning Information,* Cheshire, CT, Graphics Press, 1990.

[32] M. Twyman, 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.

[33] S. Verjans, State-of-the-Art for Input Modalities, *Esprit Basic Research Project AMODEUS-2 Working Paper* RP5-TM-WP20, 1995.