NIELS OLE BERNSEN

MULTIMODALITY IN LANGUAGE AND SPEECH SYSTEMS - FROM THEORY TO DESIGN SUPPORT TOOL

1. INTRODUCTION

This paper presents an approach towards achieving fundamental understanding of unimodal and multimodal output and input representations with the ultimate purpose of supporting the design of usable unimodal and multimodal human-human-system interaction (HHSI). The phrase 'human-human-system interaction' is preferred to the more common 'human-computer interaction' (HCI) because the former would appear to provide a better model of our interaction with systems in the future, involving (i) more than one user, (ii) a complex networked system rather than a (desktop) 'computer' which in most applications may soon be a thing of the past, and (iii) a system which increasingly behaves as an equal to the human users (Bernsen, 2000). Whereas the enabling technologies for multimodal representation and exchange of information are growing rapidly, there is a lack of theoretical understanding of how to get from the requirements specification of some application of innovative interactive technology to a selection of the input/output modalities for the application which will optimise the usability and naturalness of interaction. Modality Theory is being developed to address this, as it turns out, complex and thorny problem starting from what appears to be a simple and intuitively evident assumption. It is that, as long as we are in the dark with respect to the nature of the elementary, or unimodal, modalities of which multimodal presentations must be composed, we do not really understand what multimodality is. To achieve at least part of the understanding needed, it appears, the following objectives should be pursued, defining the research agenda of Modality Theory (Bernsen, 1993):

(1) To establish an exhaustive taxonomy and systematic analysis of the unimodal modalities which go into the creation of multimodal *output* representations of information for HHSI.
(2) To establish an exhaustive taxonomy and systematic analysis of the unimodal modalities which go into the creation of multimodal *input* representations of information for HHSI. Together with Step (1) above, this will provide sound foundations for describing and analysing any particular system for interactive representation and exchange of information.

B. Granström, D. House and I. Karlsson (eds.), Multimodality in Language and Speech Systems, 93–148.

(3) To establish principles for how to legitimately combine different unimodal output modalities, input modalities, and input/output modalities for usable representation and exchange of information in HHSI.
(4) To develop a methodology for applying the results of Steps (1) – (3) above to the early design analysis of how to map from the requirements specification of some application to a usable selection of input/output modalities.
(5) To use results in building, possibly automated, practical interaction design support tools.

The research agenda of Modality Theory thus 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 constitute an optimal solution to the representation and exchange of that information. As we shall see and as has become obvious from the literature on the subject through the 1990s, this is a hard problem, for two reasons. Firstly, already at the level of theory there are a considerable number of unimodal modalities to consider whose combinatorics, therefore, is quite staggering. Secondly, when it comes to applying the theory in development practice, the context of use of a particular application must be taken thoroughly into account in terms of task, intended user group(s), work environment, relevant performance and learning parameters, human cognitive properties, etc. A particular modality is not simply good or bad at representing a certain type of information – its aptness for a particular application very much depends on the context. This adds to the combinatorics generated by the theory an open-ended space of possibilities for consideration by the developer, a space which, furthermore, despite decades of HCI/HHSI research remains poorly mastered, primarily because such is the nature of engineering as opposed to abstract theory.

Given the many different and confusing ways in which the terms ‘media’ and ‘modality’ are being used in the literature, it should be made clear from the outset what these terms mean in Modality Theory.

A *medium* is the physical realisation of some presentation of information at the interface between human and system. Media are closely related to the classical psychological notion of the human “sensory modalities”, i.e. vision, hearing, touch, smell, taste, and balance. Thus, the graphical medium is what humans or systems see, i.e. light, the acoustic medium is what humans or systems hear, i.e. sound, and the haptic medium is what humans or systems touch. Physically speaking, graphics comes close to being photon distributions, and acoustics comes close to being sound waves. In physical terms, haptics is obviously more complex than those two and no attempt will be made here to provide a physical description of haptics beyond stating that haptics involve touching. Media are symmetrical between human and system: a human hears (output) information expressed by a system in the acoustic medium, a system sees (input) information expressed by a human in the graphical medium (in front of a camera, for instance), etc. In the foreseeable future, information systems will mainly be using the three input/output media of graphics, acoustics and haptics. These are the media addressed by Modality Theory so far. To forestall a possible

misunderstanding, the medium of graphics includes both text and "graphics" in the sense of images, diagrams, graphs etc. (see below).

The term *modality* (or *representational modality* as distinct from the sensory modalities of psychology) simply means "mode or way of exchanging information between humans or between humans and machines in some medium". The reason why any approach to multimodality is bound to need both of the notions of media and modalities is that media only provide a very coarse-grained way of distinguishing between the many importantly different physically realised kinds of information which can be exchanged between humans and machines. For instance, a graphical output image and a typed Unix output expression are both output graphics, or 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, etc. It seems obvious, therefore, that we need a much more fine-grained breakdown among available representational modalities than what is offered by the distinction between different media. The notion of representational modalities just introduced is probably quite close to that intended by many authors. As early as ten years ago, Hovy & Arens (1990), observed that, e.g., tables, beeps, written and spoken natural language may all be termed 'modalities' in some sense.

Some additional terms are clarified briefly to avoid misunderstandings later on. *Input* means interactive information going from A to B and which has to be decoded by B. A and B may be either humans or systems. Typically in what follows, A will be a human and B will be a system. It is thus taken for granted that we all know a lot about what can take place in an interaction in which both A and B are humans, or in which several humans interact together as well as interacting with a system. *Output* means interactive information going from B (typically the machine) to A (typically a human). The term *interactive* emphasises that A and B exchange information deliberately or that they communicate. In this central sense of 'interaction', it is *not* interaction when, e.g., a surveillance camera tracks and records an intruder unbeknownst to that intruder. It should also be noted that Modality Theory is about (representational) modalities and not about the *devices* which machines and humans use when they exchange information, such as hands, joysticks, or sensors. The positive implication is that the world of modalities is far more stable than the world of devices and hence much more fit for stable theoretical treatment. The negative implication is that Modality Theory in itself does not address the – sometimes tricky – issues of device selection which may arise once it has been decided to use a particular set of input/output modalities for an application to be built. On a related note, the theory has nothing to say about how to do the detailed design (aesthetically or otherwise) of *good* output presentations of information using particular modalities. As the colourful field of animated interface agents illustrates at present, it is one thing to safely assume that these virtual creatures have strong potential for certain kinds of application but quite another to demonstrate that potential through successful design solutions. Finally, it should be pointed out that when we refer to the issue of which modalities to use for exchanging information of some kind, 'information' means information in the abstract, as in 'medical data entry information', information in a new interactive game to be developed, or

geographical information for the blind. Such descriptions are commonplace, and they leave more or less completely open the question of which modalities to use for the particular purpose at hand.

Modality Theory is, in fact, a century-old subject which easily antedates even the Babbage machine. People have interacted with information presentations on pyramids, in books or in magazines for a very long time. For instance, output modality analysis has a long tradition in the medium of (static) graphics. Outstanding examples are the results achieved on static graphic graphs (Bertin, 1983; Tufte, 1983, 1990). Given today's and tomorrow's input/output technologies, however, we need to address a much wider range of modalities and modality combinations. This is a truly collective endeavour. Modality Theory and the methodology for its practical application is an attempt to provide and illustrate a reasonably sound theoretical framework for integrating the thousands of existing and emerging individual contributions to our understanding of the proper use of modalities in interaction design and development.

This chapter addresses, at different levels of detail, all of the five points on the research agenda of Modality Theory described above, as follows. Section 2 presents the generation of the taxonomy for unimodal output modalities at several levels of abstraction. Section 3 proposes a draft standard representation format for modality analysis. Section 4 presents ongoing work on generating the taxonomy for input modalities. This part of the research agenda has proved to be hard and full of surprises. Section 5 presents our first full-scale application of the theory in its role as interaction design support. Finally, Section 6 concludes by discussing empirical and theoretical approaches for how to deal with the combinatorial explosion of modality combinations in multimodal systems. Due to space limitations, it has sometimes been necessary to refer to other publications for more detail.

For the obvious reason, the modality illustrations to be provided below are all presented in static graphics just like the present text itself. Current literature tends to focus on input/output modalities which are technically more difficult to produce, and which are less explored, than the static graphics modalities. It may be worthwhile to stress at this point, therefore, that all or most of the modality concept to be introduced below in fact do generalise to all possible modalities in the media of graphics, acoustics and haptics.

2. A TAXONOMY OF UNIMODAL OUTPUT

The taxonomy of unimodal output modalities to be presented is not the only one around although it appears to be the only one which has been generated from basic principles rather than being purely, or mainly, empirical in nature. In addition, its scope is as broad as that of any other attempt in the literature. A solid taxonomy based on decades of practical experience is Tufte's taxonomy of data graphics (Tufte, 1983). Twyman (1979) presents a taxonomy of static graphics representations (text, images, etc.). It is of wider scope than Tufte's taxonomy and, like the latter, based on long practical experience. Still in the static graphics domain, (Lohse et al., 1991) present a taxonomy which is based on experiments in which

they studied how subjects intuitively classify sets of static graphic representations. Of much broader scope, comparable to that Modality Theory, are the lists of modalities and modality combinations in (Benoit et al., 2000). These lists simply enumerate modalities found in a large sample of the literature on multimodality from the 1990s.

A taxonomy of representational modalities is a way of carving up the space of forms of representation of information based on the observation that different modalities have different properties which make them suitable for exchanging different types of information between humans and systems. Let us assume that modalities can be either unimodal or multimodal and that multimodal modalities are combinations of unimodal modalities, i.e. can be completely and uniquely defined in terms of unimodal modalities. These assumptions suggest that if we want to adopt a principled approach to the understanding and analysis of multimodal representations, we have to start by generating and analysing unimodal representations. Generation comes first, of course. So the crucial issue at this point is how to generate the unimodal modalities. 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 and is being used widely in the field, it appears that no stable scientific taxonomy was ever created in a purely empirical fashion from the bottom up. If, for instance, experimental subjects are asked to spontaneously cluster a more or less randomly selected set of analogue static graphic representations (Lohse et al., 1991), the subjects may classify according to different criteria, they may be unable to express the criteria they use, and in the individual subject the criteria that are being applied may be incoherent. An alternative to the purely empirical approach is to generate modalities from basic principles and then test through intuition, analysis, and experiment whether the generated modalities satisfy a number of general requirements. If not, the generative principles will have to be revised. Let us adopt the generative approach in what follows. We want to identify a set of unimodal *output* modalities which satisfies the following requirements:

(1) *completeness*, such that any piece of, possibly multimodal, output information in the media of graphics, acoustics and haptics can be exhaustively described as consisting of one or more unimodal modalities;
(2) *uniqueness*, such that any piece of output information in those media can be characterised in only one way in terms of unimodal modalities;
(3) *relevance*, such that the set captures the important differences between, e.g., beeps and spoken language from the point of view of output information representation; and
(4) *intuitiveness*, such that interaction developers recognise the set as corresponding to their intuitive notions of the modalities they need or might need. Given the practical aims of Modality Theory, it is of crucial importance to operate with intuitively easily accessible notions without sacrificing systematicity.

To satisfy requirements (a) and (b) in particular, the generative process itself must be completely transparent. The four requirements differ in status as regards empirical testing of the generated taxonomy. Thus (d), on intuitiveness, is the more immediately accessible to evaluation. But even with respect to (d) as well as for (a) through (c), the theory can and should be exposed to more systematic empirical testing of various kinds.

The space of unimodal output representations can be carved up at different levels of abstraction. We have seen that already above, in fact, because the three media of graphics, acoustics and haptics may be viewed as a very general way of structuring the space of unimodal output representations. What will be proposed in the following is a downwards extensible, hierarchical generative taxonomy of unimodal output modalities which at present has four levels, a *super level*, a *generic level*, an *atomic level* and a *sub-atomic level*. In terms of the generative steps to be made, the generic level comes first. Thus, the taxonomy is based on a limited set of well-understood generic unimodal modalities. In their turn, the generic modalities are generated from sets of basic properties. An earlier version of the taxonomy to follow is (Bernsen 1994).

2.1. Basic Properties

We generate the generic-level unimodal output modalities from a small set of *basic properties* which serve to robustly distinguish modalities from one another within the taxonomy. The properties are: *linguistic/non-linguistic, analogue/non-analogue, arbitrary/non-arbitrary* and *static–dynamic.* In addition, distinction is made between the physical *media of expression* of graphics, acoustics, and haptics, each of which are characterised by very different sets of perceptual qualities (visual, auditory and tactile, respectively). These media determine the scope of the taxonomy. It follows that the taxonomy does not cover, for instance, olfactory and gustatory output representations of information which would appear less relevant to current interaction design. Thus, the scope of the taxonomy is defined according to the relevance requirement (c) above.

By taking those basic properties as points of departure, unimodal output modality generation starts from what are arguably the most general and robust distinctions among the capabilities of physically realised representations for representing information to humans. The set of basic properties has 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 interaction design purposes. For instance, the same linguistic message may be represented in either the graphical, acoustic, or haptic medium but the choice of medium strongly influences the suitability of the representation for a given design purpose and is therefore considered a choice between different modalities. So, the first justification for the choice of basic properties is their profoundly different capabilities for representing information. The second justify-cation is that these basic properties appear to generate the right outcome, as we shall

see, i.e. to eventually generate the unimodal output modalities which fit the intuitions and the relevance requirements which developers already have.

The basic properties 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, such as speech and text, can, somehow, represent anything and one might therefore wonder why we need any other kind of modality for representing information in HHSI. The basic reason appears to be that linguistic representations lack the *specificity* which characterise analogue representations (Stenning & Oberlander, 1991; Bernsen, 1995). Instead, linguistic representations are *abstract* and *focused:* they focus, at some level of abstraction, on the subject-matter to be communicated without providing its specifics. The cost of linguistic abstraction and 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 will not know much about these specifics from understanding the expression 'my neighbour'. The presence of abstract focus and the 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 or Chinese. 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 representation purposes. For a detailed analysis of the implications of this complementarity for HHSI, see Bernsen (1995).

Analogue representations, such as images and diagrams, represent through aspects of similarity between the representation and what it represents. These aspects can be many, as in holograms, or few, as in a standard data graphics pie graph (or pie chart). Note that the sense of 'analogue' in Modality Theory is only remotely related to that of 'analogue (vs. digital)'. Being complementary to linguistic modalities, analogue representations (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, video or hologram 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 important use of language is to *annotate* analogue representations, such as a 2D graphic map or a haptic compositional diagram; and one important use of analogue representation is to *illustrate* linguistic text. In annotation, analogue representation provides the specificity which is being commented on in language; 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 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 or else remain uninterpretable. Thus we stipulate things like "In this list, the boldfaced names are those who have already agreed to attend the meeting". In the case of non-arbitrary representations, such as when using the linguistic expressions of some natural language, introductory conventions are basically superfluous as the expressions already belong to an established system of meaning. It is not a problem for the taxonomy that representations, which used to be arbitrary, may gradually acquire common use and hence become non-arbitrary. Traffic signs may be a case in point.

Static representations and *dynamic representations* are mutually exclusive. However, the notion of static representation used in Modality Theory is not a purely physical one (what does not change or move relative to some frame of reference) nor is it a purely perceptual one (what does not appear to humans to change or move). 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. Dynamic representations are transient and do not afford freedom of perceptual inspection (Buxton 1983). According to this static/dynamic distinction, a representation is static even if it exhibits perceptible short-duration repetitive change. 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, are considered static rather than dynamic representations. The implication is that some *acoustic* representations are static. A lengthy video that plays indefinitely, on the other hand, would still be considered dynamic because it does not exhibit short-duration repetitive change. The reason for adopting this not-purely-physical and not-purely-perceptual definition of static representation is that, from a usability point of view, and that is what interaction designers have to take into account when selecting modalities for their applications, the primary distinction is between representations which offer freedom of perceptual inspection and representations which do not. Just imagine, for instance, that your standard Windows GUI main screen were as dynamic as a lively animated cartoon. In that case, the freedom of perceptual inspection afforded by static graphics would be lost with disastrous results both for the decision-making process that precedes much interaction and for the interaction itself. This particular way of drawing the static-dynamic distinction does not imply, of course, that a blinking graphic image icon has exactly the same usability properties as a

perceptually static one. Distinction between them is still needed and will have to be made internally to the treatment of static graphic modalities.

The *media* physically instantiate or embody representational modalities (se also Section 1). Through their respective physical instantiations, the various media are accessible through different sensory modalities, the graphic medium visually, the acoustic medium auditorily and the haptic medium tactilely. Different media have very different physical properties and are able to render very different sets of perceptual qualities. An important point which is sometimes ignored is that *all* of the perceptible physical properties characteristic of a particular medium, their respective scope of variation, and their relative cognitive impact are at our disposal when we use a certain representational modality in that medium. Standard typed natural language text, for instance, being graphical, can be manipulated graphically (boldfaced, italicised, coloured, rotated, highlighted, re-sized, textured, re-shaped, projected, zoomed-in-on etc.), and such manipulations can be used to carry information in context. Exactly the same holds for graphical images and other analogue graphical representations. This example shows that one should be careful when, or, indeed, preferably avoid, contrasting "text and graphics", because in the example just provided, the text *is* being graphically expressed. Text, or language more generally, need not be expressed graphically, however, but can be expressed acoustically (when read aloud) and haptically as well. Similarly, the reason why spoken language is so rich in information is that it exploits to the full the perceptible physical properties of the acoustic medium. We call these perceptible properties *information channels* and will return to them later (Section 3).

2.2. Generating the Generic Level

Given the basic properties presented in the previous section, the generation of the generic level of the taxonomy is purely mechanical, producing 48 (2x2x2x2x3) basic property combinations each of which represents a generic-level unimodal modality (Table 1). Each of the 48 generic unimodal modalities is completely, uniquely and transparently defined in terms of a particular combination of basic properties. Table 1 uses abbreviations to represent the basic properties. The meaning of these abbreviations should be immediately apparent. The term 'generic' indicates that unimodal modalities, as characterised at the generic level, are still too general to be used as a collection of unimodal modalities in an interaction designer's toolbox. The reason is that a number of important distinctions among different unimodal modalities cannot yet be made at the generic level (see Section 2.3).

All 48 unimodal modalities are perfectly possible forms of information representation. 48 unimodal output modalities at the generic level is a lot, especially since we are going to generate an even larger number when generating the atomic level of the taxonomy. However, closer analysis shows that it is possible to significantly reduce the number of generic modalities. The reductions to be performed are of two kinds. Both reductions are made with reference to the requirement of (current) relevance above. The first reduction is removal of modalities the use of which for interaction design purposes is in conflict with the purpose of Modality Theory. By

Table 1. The full set of 48 combinations of basic properties constituting possible unimodal output modalities at the generic level of the taxonomy. All modalities are possible ways of representing information.

	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*

contrast with the first reduction, the second reduction is not a removal of modalities but merely a fusion of some of them into larger categories. Both reductions are completely reversible, of course, simply by reinstating modalities from Table 1 which have been removed, or by re-separating modalities which were subjected to fusion. The reductions will be described in the following.

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 UNIMODAL LEVEL*	*NOTATION*
I. Linguistic modalities <li,-an,-ar>	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>
	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 sign graphic language	<li,-an,-ar,sta,gra>
	8. Static non-analogue sign acoustic language	<li,-an,-ar,sta,aco>
	9. Static non-analogue sign haptic language	<li,-an,-ar,sta,hap>
	10. Dynamic non-analogue sign graphic language	<li,-an,-ar,dyn,gra>
	11. Dynamic non-analogue sign acoustic language	<li,-an,-ar,dyn,aco>
	12. Dynamic non-analogue sign haptic language	<li,-an,-ar,dyn,hap>
II. Analogue modalities <-li,an,-ar>	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>
	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 <-li,-an,ar>	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>
	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 <-li,-an,-ar>	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>
	29. Dynamic acoustic structures	<-li,-an,-ar,dyn,aco>
	30. Dynamic haptic structures	<-li,-an,-ar,dyn,hap>
SUPER LEVEL	*GENERIC UNIMODAL LEVEL*	*NOTATION*

Some modalities in Table 1 are inconsistent with the *purpose* of the taxonomy. Modality Theory in general and the taxonomy of unimodal output modalities in particular, serve the clear and efficient presentation and exchange of information. Given this purpose, the *arbitrary use of non-arbitrary representations* constitutes a capital sin in the context of interaction design. What this involves is providing a representation which already has an established meaning, with an entirely different meaning. For instance, arbitrary use of established linguistic expressions in a static

graphic interface (Modality 13 in Table 1) should not occur in information systems output. To do so would be like wanting to achieve clear and efficient communication by letting 'yes' mean 'no' and vice versa. The result, as we know from children's games, is massive production of communication error and ultimate communication failure. Or if, for instance, a graphic interface designer lets iconic images of apples refer to ships on a graphic screen ocean map rather than using iconic images of ships for this purpose (assuming, among other things, that the ships do not carry apples), we have another case of using non-arbitrary representations arbitrarily. We also have a case of bad (i.e. confusing) interface design. 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. Modality Theory, on the other hand, aims to support designers in making the best use of representational modalities for the purpose of clear and efficient presentation and exchange of information, through building on the expressive strengths of each. The taxonomy, therefore, simply does not address cryptography. What about passwords? It would seem that, in general, passwords are not cryptographic representations. They are just meant to be kept secret, and that is something else. Similarly, it is perfectly acceptable to use numbers arbitrarily in the sense of, for instance, arbitrarily assigning different numbers to players on a team. This is not in conflict with the meaning of numbers. Problems only start to arise if, say, a player is being assigned the number 3 and everybody is being told that the player has, in fact, number 2.

We thus have to remove columns 1-6, 13-18 and 25-30 from the Table 1 matrix. The remaining 30 unimodal output modalities are presented in a more explicit form in Table 2 which names each modality and shows its notation. Table 2 also shows the super level of the taxonomy (see below).

The second reductive step in generating the generic level of the taxonomy is *purely pragmatic* or practical rather than theoretical. The reduction of the number of unimodal modalities from 30 to 20 (Table 3) has been done uniquely in order to simplify the work involved in using the taxonomy for practical purposes, cf. the intuitiveness and relevance requirements above. The resulting taxonomy becomes less scholastic, as it were, and more usable. Table 3 integrates the presentation and analysis of *static* acoustic modalities with the presentation and analysis of *dynamic* acoustic modalities, and integrates the presentation and analysis of *static* haptic modalities with the presentation and analysis of *dynamic* haptic modalities. No modality information is lost in the process, so the completeness requirement is not being violated.

The practical reasons are as follows. Static acoustics, such as acoustic alarm signals, constitute a relatively small and reasonably well-circumscribed fraction of acoustic representations in whatever acoustic modality. For practical purposes, the presentation and analysis of the static acoustic modalities may without loss of information be integrated with that of the corresponding dynamic acoustic modalities which constitute the main class of acoustic representations. Similarly, dynamic haptics, such as the invention of dynamic Braille text devices where users do not have to move their fingers because the device pad itself changes dynamically to display new signs, currently constitute a relatively small fraction of haptic

representations in whatever haptic modality. 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-instate the static/dynamic distinction in the haptic modalities part of the taxonomy.

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 UNIMODAL 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>
	19. Static haptic structures Dynamic haptic structures	<-li,-an,-ar,sta/dyn,hap>
<-li,-an,-ar>	20. Dynamic graphic structures	<-li,-an,-ar,dyn,gra>
SUPER LEVEL	*GENERIC UNIMODAL LEVEL*	*NOTATION*

Overall, at the generic and atomic levels combined, the proposed fusions reduce the number of modalities in the taxonomy by some 30 modalities. In a designer's toolbox, we want no more tools than we really need.

The 30 and 20 generic unimodal modalities of Tables 2 and 3, respectively, have been divided into four different classes at the super level, i.e. the linguistic, the analogue, the arbitrary, and the explicit structure modalities. The *super level* merely represents one convenient way of classifying the generic-level modalities among others, although, once laid down, it determines the overall surface architecture of the taxonomy. Other, equally useful, classifications are possible and can be used freely

in modality analysis and taxonomy use, such as classifications according to medium or according to the static/dynamic distinction. Furthermore, the chosen super level modalities may not be deemed to be modalities proper (yet) as they lack physical realisation. This is a purely terminological issue, however. A point of greater significance is that, at the generic level, four of the linguistic modalities (Modalities 1 through 4 in Table 3) use analogue *signs* and four use non-analogue signs (Modalities 5 through 8 in Table 3). Basically, however, these are all primarily *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 signs themselves 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 for native users to decode their analogue meanings.

2.3. Generating the Atomic Level

It may not be immediately obvious from Table 3 why the generic-level taxonomy cannot be used as a designer's toolbox of unimodal output modalities and does not meet the requirements of relevance and intuitiveness. This is partly due to the fact that some modalities are largely obsolete and hence irrelevant, 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, such as Modality 9, 'static analogue graphics'. The lack of intuitiveness is due to the relatively high level of 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*, because both are subsumed by 'static analogue graphics'. In interaction design, however, these two modalities are being used for very different purposes of information representation and exchange. In another example, static graphic written *text* is useful for rather different purposes than is static graphic written *notation*. Yet both are subsumed by 'static non-analogue sign graphic language' at the generic level. Since the generic level does not make explicit such important distinctions among modalities, it is even difficult to put the completeness of the taxonomy to the test.

To achieve the relevance and intuitiveness required, which in their turn are preconditions for testing the completeness of the taxonomy, we need to descend at least one level in the abstraction hierarchy defined by the taxonomy. This is done by adding further distinctions among basic properties, thereby generating the atomic level of the taxonomy as presented in the static graphic conceptual diagram in Figure 1. In Figure 1, many of the generic modalities have several (equally unimodal) atomic modalities subsumed under them which inherit their basic properties and have distinctive properties of their own.

The new basic properties and distinctions to be introduced in order to generate the atomic level are specific to the super and generic level fragments of the taxonomy to which they belong. Where do these properties and distinctions come from? The generation of the atomic level follows the same principles as that of the

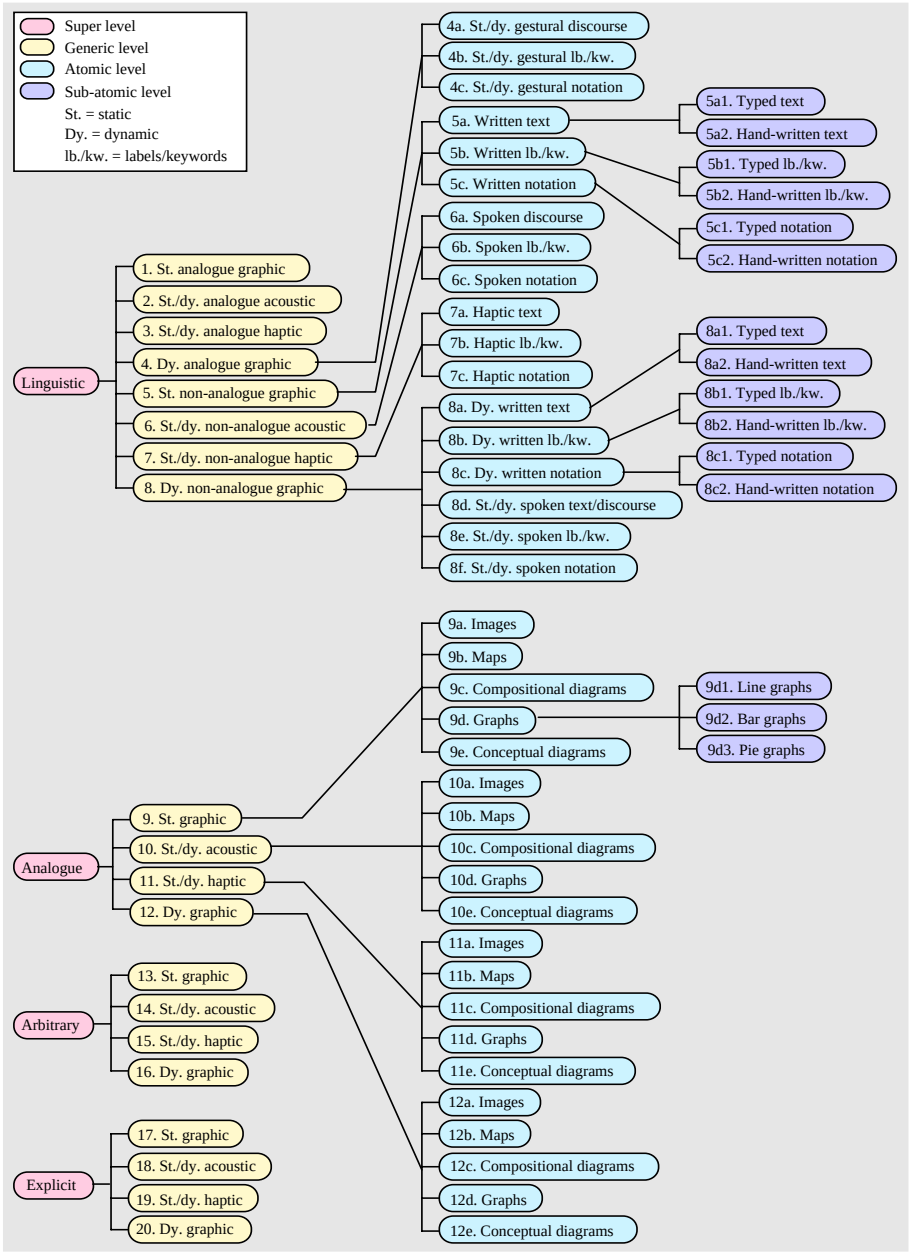

Figure 1. The taxonomy of unimodal output modalities. The four levels are, from left to right: super level, generic level, atomic level and sub-atomic level.

generic level. The new distinctions have been selected such as to support the generation of importantly different unimodal output modalities, which satisfy the intuitiveness requirement. In addition, pragmatic reductions have been made in order not to proliferate atomic modalities beyond those necessary in current interaction design, thus addressing the relevance requirement. In what follows, a justification will be presented for each super level segment generation of atomic modalities, starting with the linguistic modalities.

2.3.1. Linguistic Atomic Modalities

Two *types* of distinction go into the generation of the atomic level linguistic modalities. The first type of distinction is between (a) text and discourse, and (b) text/discourse, labels/keywords and notation. As to (a), it is a well-known fact that, grammatically and in many other respects, written language 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 compared to the context in which the author wrote the message. By contrast, spoken language has evolved to serve *situated* communication, the partners in the communication sharing location, situation and time. Hybrid situations of linguistic communication made possible by technology, such as telephone conversation, on-line e-mail dialogue or www chat, generate 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 and chat, temporal independence is missing and some situation sharing may be present. Generalising these observations, situated linguistic communication is termed *discourse* and situation independent linguistic communication is termed *text* (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 the original forms of text exchange, such as mail letters or books, than do on-line e-mail and chat dialogue because normal e-mail communication is independent of partners' place, situation and time.

The distinction (b) between text/discourse, labels/keywords and notation is straightforward and important. *Text* and *discourse* have unrestricted expressiveness within the basic limitations of linguistic expressiveness in general (Section 2.1). Discourse and text modalities, however, tend to be too lengthy for use in brief expressions of focused information in menu lines, graph annotations, conceptual diagrams, command expressions etc. across media. *Labels* or, in another, equivalent, term, *keywords* are well suited and widely used for this purpose. Their drawback is their inevitable ambiguity which, at best, may be somewhat reduced by the context in which they appear, such as the context of other menu-line keywords. Whereas text, discourse and labels/keywords are well suited for representing information to any user who understands the language used, *notation*, such as first-order logic or xml, 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 interaction design across media and static-dynamic modalities.

Table 4. The atomic level unimodal linguistic modalities are shown in boldface in the right-hand column.

GENERIC UNIMODAL LEVEL	ATOMIC UNIMODAL 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**

The second type of distinction involved in generating the atomic level is empirical in some restricted sense. 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 implies the possibility that Modality Theory might so far be missing some important type(s) of linguistic communication. However, testing made so far suggests that Table 4 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 (even if done by a gesturing robot). Static gestural language is included in 4a-4c (see below). 5a-5c cover the original 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 below). 6a-6c cover the original form of discourse, i.e. *spoken language.* 7a-7c includes static and dynamic haptic language, such as Braille. The atomic modalities in Section 8 of Table 4 illustrate 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 versions of 5a-5c, such as scrolling text. However, Section 8 also includes graphically represented (non-acoustic) spoken language as produced, for instance, by a talking head or face, including read-aloud text and spoken discourse, labels/keywords and notation (8d-8f).

The pragmatic reductions of the linguistic atomic modalities are straightforward. As argued in Section 2.2, 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 textual languages, such as Chinese, and whether these are static or dynamic, graphic or haptic (Sections 1, 3 and 4 of Table 4), may therefore be fusioned with their non-analogue, non-iconographic counterparts without effects on interaction design. (The "glyphs" which have been invented for expressing multi-dimensional data points in graph space are rather forms of static graphic arbitrary modalities (Joslyn et al., 1995, Section 2.3.3)). Analogue speech sounds (onomatopoietica and others), 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, such as the 'V' sign and many others, (Section 1 of Table 4), has been fusioned with dynamic gestural language (Section 4). Finally, the static graphic spoken language atomic modalities, such as a "frozen" talking head (Section 5), have been fusioned with their dynamic counterparts (Section 8). The result of this comprehensive set of fusions is shown as six triplets of atomic linguistic modalities in Figure 1. These modalities are shown in boldface in the right-hand column of Table 4. The strong claim of Modality Theory is that these modalities are all that interaction designers need in order to have a complete, unique, relevant and intuitive set of unimodal linguistic output modalities at the atomic level of abstraction. If more linguistic modalities are needed, they must either be generated from those at the atomic level and hence belong to some lower level of abstraction, such as the sub-atomic level, or they will re-appear by backtracking on some reduction (fusion) performed to obtain the modalities which presently constitute the linguistic atomic level.

The next section (2.3.2) will discuss *prototypical structure* and *continuity of representation,* phenomena which are prominent in the analysis of analogue representations and which need to be understood in order to avoid confusion in handling borderline issues of demarcation among different modalities. It should be noted that these phenomena are also present in the linguistic domain, so that, for instance, the issue over whether a certain representation is a collection of labels/keywords or is a notation may have to be decided by recourse to prototypical instances of labels/keywords and notation. In fact, prototypicality is a basic characteristic of conceptual structures. It follows that any theory of modalities will have to deal with the phenomenon.

2.3.2. Analogue Atomic Modalities

The analogue atomic modalities (Table 5) are generated without any pragmatic modality fusion or reduction. The generation is based on the concept of 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 entire domain of analogue representation, whether static or dynamic, graphic, acoustic or haptic. What is needed, just like in the linguistic domain, is a justification of the distinctions, which have been introduced to generate the atomic level of analogue representation of information. Why are these distinctions the right ones with which to carve up the vast and complex space 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 both intuitively distinct and relevant to interaction design. At least two more questions need to be addressed, however. The first is whether the five concepts at issue exhaust the space of analogue atomic representation, cf. the completeness requirement in Section 2. The second question is how these concepts are defined so as to avoid overlaps and confusion when applying them to concrete instances in design practice, i.e. how distinct and mutually exclusive are these concepts in practice? Let us begin with the second question.

Table 5. The atomic level unimodal analogue modalities.

GENERIC UNIMODAL LEVEL	ATOMIC UNIMODAL 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

The exclusiveness (uniqueness) issue is particularly difficult in the analogue domain. The problem about exclusiveness in the analogue domain is that representations belonging to one modality, such as images (e.g. Figure 11), can often be manipulated to become as close as desired to representations belonging to several

other modalities, such as compositional diagrams (Figure 2). Such *continuity of representation* is a well-known characteristic of many ordinary concepts and has been explored in prototype theory (Rosch, 1978). The point is that classical definitions using jointly necessary and sufficient conditions for specifying when an instance (or token) belongs to some category does 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 general implication is that the concepts of atomic and other modalities of Modality Theory cannot be fully intuitive in the sense of completely corresponding to our standard concepts. Any theory in the field would have to recognise this fact because such is the nature of the concepts, which we carry around in our heads. Ultimately, however, this is a desirable effect of theory. 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, landscape or otherwise (e.g. Figure 11). However, the static graphic images modality also includes 3D or 1D images, and these differ from those prototypes. In other words, Modality Theory can only meet the relevance and completeness requirements through some amount of *analytic generalisation*, which, in its turn, challenges the intuitiveness requirement somewhat. We shall see how the concept characterisations in the analogue domain work using abbreviated versions of the (unpublished) concept characterisations of Modality Theory which often run several pages per concept, excluding illustrations.

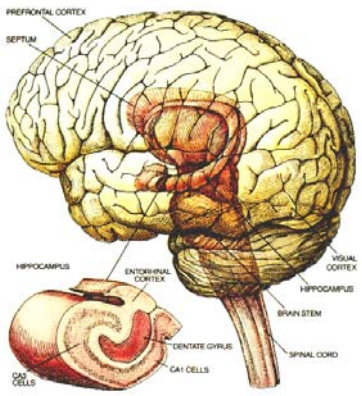

Figure 2. A prototypical compositional diagram: an annotated (hence bimodal) static 2D graphic representation of part of the structure of the brain.

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. This characterisation of diagrams will be expanded through characterisations of the various types of diagram below.

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, as in ideal virtual reality output representations or standard input from real-life scenes, images allow realistic (quasi-) perception of the rich specific properties of objects, processes and events, which cannot easily be represented linguistically (Section 2.1). Images vary from high-dimensionality, maximally specific images to images whose specificity has been highly reduced for some purpose ('sketches'). 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 (e.g. too high frequency, too low frequency), or normally hidden beneath some exterior, so that the objects, processes or events cannot themselves be perceived by humans. Images may also represent objects in a medium different from its 'normal' physical medium as when, for instance, acoustic information is being represented graphically (e.g. sonar images). Because images, considered on their own as unimodal representations, 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. Figure 11 shows a prototypical image, i.e. a high-specificity 2D static graphic colour photograph of a person.

It may be observed from the above characterisation of images that images are being contrasted to their closest neighbour in analogue modality space, i.e. compositional diagrams (see below). Furthermore, it is pointed out in the image characterisation that images have limited value as stand-alone unimodal representations because of their lack of focus. For most interaction design purposes, images need linguistic annotation, which explains the intended point contributed by the image, so that the combined representation becomes bimodal. As a convenient, albeit coarse and relative generalisation which should be handled with care, unimodal modalities may be 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 and discourse modalities, for instance, are among the most independent unimodal modalities there are. Graphs, on the other hand, tend to be powerless in expressing information unless accompanied by other modalities. This issue will recur several times in what follows but a full discussion goes beyond the scope of this chapter. It still needs to be kept in mind that no unimodal modality has unlimited expressive power.

Compositional diagrams, such as an exploded representation of a wheelbarrow, 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, i.a., structure and function. Combinations of analogue representation and linguistic annotation in compositional diagrams vary from highly labelled diagrams containing rather abstract (i.e. reduced-specificity or schematic) 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 they often use focusing mechanisms, saliency enhancement and dimensionality reduction (Bernsen, 1995). These selectivity mechanisms are used in order to optimise the compositional diagram for representing certain types of information rather than others. Figure 2 shows a rather high-specificity but otherwise prototypical compositional diagram, i.e. an annotated static 2D graphic representation of the structure of the brain. Figure 2 is a bimodal representation consisting of text labels and image elements.

Even more than images, compositional diagrams depend on linguistic annotation to do their representational job. Note also how compositional diagrams are being contrasted with their closest neighbours in analogue representation space, i.e. images and conceptual diagrams.

Maps are, in fact, a species of compositional diagrams, which are 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 present 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 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 (non-map) compositional diagrams. Figure 3 shows a non-prototypical map, i.e. a unimodal, highly specificity-reduced static 2D graphic representation of the Copenhagen subway system. Only the topology and the relative positions of lines and stations have been preserved. The unimodality of the map in Figure 3 makes it uninterpretable for all but those possessing quite specific background knowledge, which enables them to supply the information, which is missing in the representation.

Maps are thus a species of compositional diagrams, sharing most of the properties of these as described above. Maps have been included in the taxonomy because they are 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 diagram’. A taxonomy of unimodal analogue modalities that ignores this fact is likely to be less relevant and intuitive than a taxonomy, which respects the fact while preserving analytic transparency.

Graphs represent quantitative or qualitative information through the use of 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. Their analogue character makes graphs well suited for facilitating users' identification of global data properties through making comparisons, perceiving data profiles, spotting trends among the data, perceiving tempo

Figure 3. A non-prototypical map: a unimodal, highly specificity-reduced static 2D graphic representation of the subway system of Copenhagen.

ral developments in the data, and/or discovering new relationships among data, and hence supports the analysis of, and the reasoning about, data 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, consistent with the analogue representation, for their interpretation. Thus, graphs are in practice (at least) bimodal modalities. Graphic graphs frequently incorporate graph space grids and other explicit structures (see below), which makes them trimodal modalities. The graph notion is quite robust and does not require contrasting with other analogue modalities - it has no close neighbours in analogue representation space. The huge diversity of graph representations requires a sub-atomic expansion of at least the static graphics graph node of the taxonomy (Section 2.5).

Conceptual diagrams use various analogue representational elements to represent the analytical decomposition of an abstract entity such as an organisation, a family, a theory or classification, or a conceptual structure or model. Thus, conceptual diagrams 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 the analogue elements in conceptual diagrams is to make the diagram's abstract subject-matter more easily accessible through spatial structure and layout. The abstract subject-matter of conceptual diagrams requires that the information they represent be carried, to a very important extent, by the linguistic modalities involved. Figure 1 shows a prototypical (multimodal) conceptual diagram.

In presenting the analogue atomic modalities, we have so far concentrated on the question of exclusiveness raised in the beginning of the present section. Let us now address the second question raised, i.e. the question of completeness. The present taxonomy assumes four categories of analogue representation: images, compositional diagrams (including maps), graphs and conceptual diagrams. In an empirical study, Lohse et al. (1991) found that subjects tended to robustly categorise a variety of analogue 2D static graphic representations into categories, which they termed 'network charts', 'diagrams', 'maps', 'icons', and 'graphs/tables'. A free-hand drawing of their findings made by the present author is shown in Figure 4. It should be noted that the pile of representations, which the subjects had to classify, did not include any static graphic images. As shown in Figure 4, 'network charts' correspond to conceptual diagrams in Modality Theory, 'diagrams' correspond to compositional diagrams, 'maps' to maps, and 'graphs' to graphs. The terminology in the field has not been standardised, so there is nothing unexpected about these variations in terminology. As no images were presented to the subjects by Lohse et al. (1991), we can ignore images in what follows. Disregarding for the moment 'icons' and 'tables' which have not been discussed above, the correspondence between the result of Lohse et al. and the present taxonomy is very close indeed, at least in the domain of static graphics representation. However, Figure 4 also includes the well-known representation types 'icons' and ‘tables’, suggesting that the taxonomy above is not complete or exhaustive. So, what is the status of icons and tables in Modality Theory? The answer is that the theory provides what is arguably some much needed analytic generalisations concerning icons and tables.

Tables, although clearly distinct from any of the atomic modalities considered above, do not constitute a separate representational modality. Rather, tables are a convenient way of structuring information represented in most graphic or haptic modalities. Tables are a particular type of *modality structure* rather than a modality. Tables are often bimodal, as in prototypical 2D static graphic tables, which combine typed language with explicit structures (cf. Tables 1 through 5 above). Note, however, that explicit structures are not necessary constituents of tables. Simple tables can be elegantly presented without any use of explicit structures at all, just by appropriate spatial distribution of the tabular information. The fact that the subjects in Lohse et al. (1991) combined graphs and tables into one category is probably due to the fact that graph information can often be represented in tables, just like tabular information is often used to generate graphs. However, the fact that the abstract information contents of a table is sometimes equivalent to the information contents of a graph does not address the question of when it is preferable to represent that information in a table and when it is preferable to represent the information in a graph. Depending on the nature of the information and the purpose of the representation, graphs may be preferable to tables or vice versa (Tufte, 1983). Graphs and tables are, therefore, different forms of information representation. Moreover, tables

may contain much else besides standard spreadsheet quantitative information, such as text, labels/keywords, notation, static or dynamic images, or graphs. The latter tables, such as a table showing a variety of dynamic graphics talking heads, do not have any graphs corresponding to them. In conclusion, tables can be uniquely defined in terms of unimodal modalities, i.e. as a particular, and quite general, way of structuring them, and tables are clearly distinct from graphs. *Lists* are another type of modality structure, which is different from, but related to, tables.

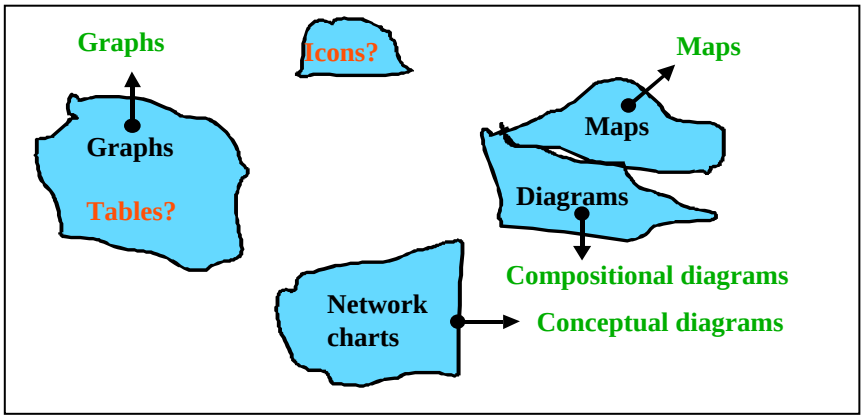

Figure 4. Subjects' classifications of analogue representations after Lohse et al. (1991). Arrows point to the corresponding terms used in Modality Theory. Question marks indicate phenomena, which are not modalities but something else.

Like lists and tables, *icons* do not constitute a separate modality. Rather, icons are "generalised labels/keywords" and the generalisation reaches far beyond 'icons' in the prototypical sense of static 2D graphic representations. Just 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 token of any modality, it would appear, 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 quite 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 exhaustive or complete. In other words, like tables and lists, icons can be uniquely defined in terms of unimodal modalities. Still, as Lohse found, icons are different from other modalities. The above analysis shows that the difference can be expressed in terms of the particular role, which a modality assumes when being used as an icon.

In conclusion, the correspondence between the taxonomy of analogue 2D static graphic modalities and the empirical results of Lohse et al. (1991) is now complete. Until an entirely different taxonomy of the space of analogue representation comes forward, and that has not happened yet, the present taxonomy would appear to be partially empirically confirmed. The results of Lohse et al. (1991) have confirmed

the assumption that the four concepts of images, compositional diagrams (including maps), graphs, and conceptual diagrams exhaust the space of analogue atomic representation. It should still be kept in mind, however, that completeness does not imply exclusiveness. We have seen that classical-style definitions of analogue and other modalities are hardly possible. This implies that borderline cases will inevitably occur. For instance, is a static graphic image one part of which is labelled by a single typed label only, an image or a compositional diagram? But if classical-style definitions are impossible, *any* taxonomy of analogue modalities will be subject to the existence of continuity of representation and of borderline cases, which are difficult to categorise without ambiguity. 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 downward 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 2.5).

2.3.3. 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 (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 in the use, which is being made of them. Arbitrary modalities are therefore by definition non-linguistic and non-analogue. As argued in Section 2.1, it is against the purpose of the taxonomy that non-arbitrary modalities be used arbitrarily. This imposes rather severe restrictions on which representations may be used arbitrarily. Nonetheless, arbitrary modalities can be very useful for representing information and we use them all the time. Information channels, in particular, are often useful for assuming arbitrary roles (Section 3). In general, any information channel in any medium can be arbitrarily assigned a specific meaning in context. This operation is widely used for expressing information in compositional diagrams, maps, graphs and conceptual diagrams and is illustrated in Figures 5 and 10 below. In another example, arbitrary modalities can be used to express acoustic 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 UNIMODAL LEVEL	ATOMIC UNIMODAL 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

2.3.4. Explicit Structure Atomic Modalities

As in the case of the arbitrary atomic modalities, no reason has been found to introduce new distinctions in order to generate a larger set of explicit structure modalities at the atomic level than was already present at the generic level (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 is 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 interaction 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 at a certain point (Figure 12).

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

GENERIC UNIMODAL LEVEL	ATOMIC UNIMODAL 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

2.4. The Generative Power of the Taxonomy

The hypothesis, which has been confirmed up to this point in the development of Modality Theory and which is inherent 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, uniqueness, relevance and intuitiveness stated above. Any multimodal output representation can be exhaustively characterised as consisting of a combination of atomic-level modalities.

Assuming that the atomic level of the taxonomy of unimodal output modalities has been generated successfully, 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, rather, the principles that were applied in generating the atomic level and the new distinctions introduced in the process. However, what has been generated goes far beyond the generative apparatus so far described. 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 particular atomic modalities generated

are the results of this multiplication process. 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 the generative power of the taxonomy is to focus on atomic modalities which are yet to become used in interaction design; which hold unexploited potential for useful information representation; which have not yet been discovered as representational modalities; or which are so "exotic" as to appear difficult to even exemplify for the time being. As several critics have noted, the very existence of just one such "exotic" unimodal modality goes against the requirement of relevance above. Before pragmatically removing them, however, which is easy as explained in the case of other modalities above, they should be scrutinised for potential relevance. In what follows, some of the generated unimodal output modalities, which fit the above descriptions are briefly commented upon.

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 come across. 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 (cf. Figure 9). 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 as Morse-code specialists. Such specialists might find uses for lengthy gestural and spoken notation as well. If (acoustic) spoken notation is being expressed as synthetic speech, the specialists might need support from graphic spoken notation in order to properly decode the information expressed.

In the analogue atomic modalities domain, *acoustic images* are becoming popular, for instance 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 to remain largely unexplored, except as redundant representations accompanying, e.g., static graphics bar graphs shown on TV. *Acoustic maps* appear to have some potential for representing spatial layout. *Acoustic compositional diagrams* offer interesting possibilities. Think, for instance, of a system for supporting the training of car repair trainees. Acoustic 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* may appear not to have any clear application potential. Yet it is possible to map, for instance, the different inheritance levels of a static graphic inheritance hierarchy into different keys. Primarily for reasons of technology and cost, output *dynamic analogue haptics* appears to be mostly unexplored territory, whether in the form of images, maps, compositional

diagrams, graphs or conceptual diagrams. Yet their potential for special user populations would appear considerable. *Dynamic analogue graphics* are extremely familiar to us but still has great unused potential. Immersive virtual reality must combine dynamic, perceptually rich analogue graphics, acoustics and haptics.

Arbitrary static graphics, *acoustics* and *haptics* are widely used already. It is less obvious how much we shall need their dynamic counterparts in future applications. A ringing telephone, of course, produces arbitrary dynamic acoustics, and a vibrating mobile phone produces arbitrary dynamic haptics. Beyond such saliency-based applications, however, it is less 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, for encircling linguistic or analogue graphic information of current interest during multimodal graphics and spoken language presentations. *Static* and *dynamic haptic explicit structures* have unexplored potential for the usual (technology and cost) reasons. In fact, it is only in the singular case of *acoustic explicit structures* that 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. Still, there might be useful functions out there for acoustic explicit structures.

It may be concluded that virtually all of the unimodal atomic output modalities in the taxonomy hold a claim to belong to a designer's toolbox of output modalities.

2.5. Generating the Sub-Atomic Level of the Taxonomy

Explicit completeness at any level of the taxonomy is still limited by level of abstraction and hence by the number of basic propertie which has been introduced to generate that level. The generic level is as complete as the atomic level, but the former is less intuitive and relevant than the latter. One virtue of the taxonomy is its unlimited downward extensibility. That is, once the need has become apparent for distinguishing between different unimodal modalities subsumed by an already existing modality, further basic properties can be sought which might help generate the needed distinctions. In the domains of arbitrary and explicit structure modalities, this possibility remains unused already at the generic level (Sections 2.3.3 and 2.3.4). Below the atomic level, however, such as at the sub-atomic level of the taxonomy, there is still much representational diversity to be identified and used when required by theory and/or technology. In what follows, the purpose is merely to illustrate possibilities. The reader is invited to generate some of the other sub-atomic parts of the taxonomy, which already beckon to be generated. 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* (cf. Figure 1). This extension may not be terribly important to output modality choice in interaction design except for reasons of aesthetic design, which go beyond Modality

Theory. The extension is important, however, to the developer's choice of *input* modalities.

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

ATOMIC UNIMODAL LEVEL	SUB-ATOMIC UNIMODAL 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 shows what is still a hypothetical application of the principle of extensibility in the domain of static graphic graphs (cf. Figure 1). This example is much more complex than the one in Table 8. Static graphic graphs are extremely useful for representing quantitative information. The domain has been the subject of particularly intensive research for decades with the result that the atomic modality 'static graphic graphs' has become much too coarse-grained a notion for handling the large variety of information representations that exist. In fact, static graphic graph theory is one of the earliest examples of systematic work on a methodology for mapping from requirements specification to modality choice (e.g. Bertin, 1983; Tufte, 1983, 1990; Lockwood, 1969; Holmes, 1984). Even in this limited "corner" of the domain addressed by Modality Theory, however, there is still no consensus on taxonomy of static graphic graphs.

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

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

Our current hypothesis is that distinguishing between three basic types of graphs is sufficient for analysing the capabilities and limitations for information representation of all possible static graphic graphs. To avoid confusion it seems necessary to distinguish, in addition, between standard graphs and enhanced graphs. The *standard graphs* are: line graphs, bar graphs and pie graphs. This may appear at once both trivial and controversial. Trivial, because these graph types, as ordinarily understood, are the most common among all static graphic graph modalities. Controversial, both in so far as Tufte (1983) argues that nobody needs pie graphs and because different authors tend to provide different, and always considerably

longer, lists of graph types. Typically, however, these lists are based on examples rather than careful definition. It would seem that the proposed graph types have not been sufficiently analysed as to their information representation capabilities. In addition, little attempt has been made to achieve reasonably generalised concepts of each type claimed to be basic to information representation. Finally, as is the norm rather than the exception in the modalities field, data graph terminology remains confusing and without any clear convergence towards a standard.

A standard *line graph* ('fever graph', 'curve chart') represents data points, whether empirical or generated, in a 1D, 2D or 2 1/2D/3D graph space, which is normally defined by co-ordinate axes. The term 'line graph' is somewhat misleading in referring to this class of graphs. The term derives from the fact that prototypical line graphs are 2D graphs in which the data points have been connected or computed over, the result being expressed in one or more lines which show how the changes in one quantity (dependent variable) are related to changes in another quantity (independent variable). However, discrete, non-continuous data point patterns need not be connected for a graph to be a line graph ('scatter diagrams/plots/graphs', 'dot charts'). In 1D no connecting lines are possible, whereas data point connections in 2 1/2D/3D are most often done using curved surfaces (Figure 5). Lines in 2D line graphs may be replaced by curved surfaces ('surface charts'). Cyclical data may be represented in circular line graphs using a circular co-ordinate system. Line graphs are good at representing large data sets with variability, showing flow, profiles, trends, history and projections, and are good for time-related data.

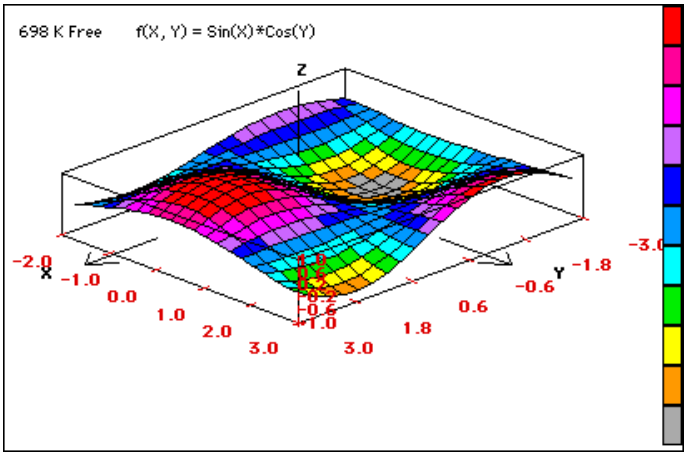

Figure 5. 2 1/2D static graphic surface function line graph using arbitrary colour coding for distinguishing hills and basins of the function.

A prototypical *bar graph* ('column graph') represents a small number of separate quantities lying within a comparative range in a 2D graph space. The term 'bar graph' is somewhat misleading in referring to this class of graphs. In a prototypical bar graph, horizontally or vertically aligned 2D bars are being used to represent

quantities through their length. However, the bars may be replaced by other geometrical shapes in 1D, 2D, 2 1/2D or 3D whose length, area or volume represents the quantities in question (Figure 6). In ‘circle graphs’, for instance, it is not the length of the bar but the area of the circle, which represents the information. In 2D ‘histograms’ (‘step charts’), the bars touch or have been replaced by stepwise curves and there is usually no spacing between the columns. Bars may be non-aligned (‘float’ or ‘slide’), diagonal, radiate from the centre of a circle or from the circumference towards the centre, be stacked in a (‘population’) pyramid, shown at an angle or as receding towards the horizon, folded to encompass out-of-range quantities, etc. Bar graphs are good at enabling comparison, particularly when horizontally or vertically aligned, and the spotting of relationships among relatively small numbers of individual quantities. If the represented data are of a lower dimensionality than the information-carrying dimensions of the (generalised) ‘bars’ used to represent them, there is a high risk of generating misinformation. Similarly, humans are bad at correctly perceiving proportional relationships between areas or volumes. Work on 3D virtual reality graphs has identified interesting problems in using 3D graphic graphs, such as occlusion and perspective (Mullet & Schiano, 1995).

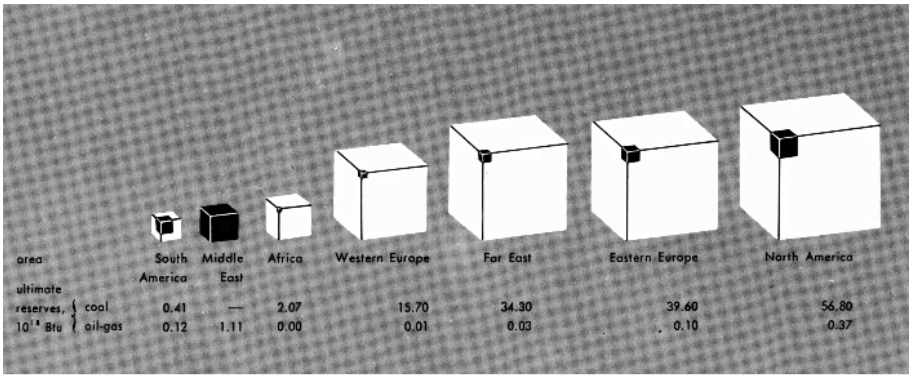

Figure 6. An example of a non-prototypical bar graph, which compares coal and oil resources in various parts of the world.

A standard *pie graph* (‘divided circle graph’, ‘sector graph’) represents a whole as decomposed into a relatively small number of quantitative constituents. The term ‘pie graph’ is somewhat misleading in referring to this class of graphs. In a prototypical pie graph, 2D or 2 1/2D circles or pies are being used to represent quantities through their percentage-wise partitioning. However, the circles or pies may be replaced by other regular geometrical shapes including, e.g., lines, squares, rectangles, triangles, ellipses, boxes, spheres, cylinders, cones, etc. Pie graphs are good at representing small numbers of parts of some whole with a view to comparing them. However, humans are bad at correctly perceiving proportional relationships between areas or volumes.

The above characterisations of sub-atomic-level static graphic graphs illustrate the need to refer to prototypes, the need to conceptually generalise quite strongly beyond these, the approach to defining graphs from the types of quantitative information which each type is best suited to represent, and the "reduction" of graph types proposed in the literature into a small number of graph modalities. If the proposed generation works, it can be rather straightforwardly generalised to include sub-atomic *dynamic* graphic graphs and static and dynamic *haptic* graphs. The latter sub-atomic generations are not shown in Figure 1. It is an interesting question whether all *acoustic* graphs must be line graph modalities or whether useful applications can be found for acoustic bar graphs and pie graphs.

Enhanced graphs are multimodal representations which not only, as in all interpretable graphs, in bimodal fashion combine an analogue standard graph with linguistic annotation, but include one or several additional (typically) analogue modalities such as images or maps. Enhanced graphs go beyond the present discussion of unimodal modalities. They are mentioned here because of the importance of one of their forms, i.e. the 'data map' or 'thematic map' in which a reduced-specificity map assumes the role of graph space. Enhanced graphs are widely used to represent graph information to "graph illiterates", such as children. An enhanced bar graph is shown in Figure 10.

2.6. Beyond Literal Meaning. Metaphor and Metonymy

So far, we have focused only on the *literal meaning* of information representations. Thus, in interaction design, an image of an apple would be meant by the designer to refer to an apple or to apples, as the case may be; an acoustic image of people convening for a meeting would be meant to refer to people convening for a meeting; etc. However, using representational modalities in their literal meaning with the intention of their being understood as such during information representation and exchange is only one, albeit fundamental, form of information representation and hence of the use of representational modalities. Sometimes it may be preferable to use *non-literal meaning* instead, or in addition, i.e. to use modalities intending them to be understood in a way, which is different from their literal meaning. *Metaphoric use* of modalities is probably the best known kind of non-literal use in interaction design so far, such as in the static graphic desktop metaphor. What metaphors do is to bring a host of meaning and knowledge from a known source domain, such as the ordinary desktop, to bear on the user's understanding of the target domain, such as the computer screen. The trick is that the user knows the source domain already, so that simple and brief reference to that domain often suffice to marshal all of that understanding for the comprehension or reception of something new and unfamiliar. Metaphors are not the only kind of potentially useful non-literal meaning, however. We will consider only two kinds of non-literal meaning, i.e. metaphor and metonomy. In *metonomy*, a complex subject-matter is being referred to through presenting some simple part of it.

In general, Modality Theory views non-literal meaning as being derived from literal meaning through subtraction of a smaller or greater amount of the literal

connotations of an expression of information in some modality. Thus, the apple of MacIntosh Computers does not proclaim that you should buy, or use, an apple (Figure 7a). The claim is rather that you should get yourself something, which, like the apple shown, is related to many good things, such as knowledge, love, natural beauty and health. With respect to literalness, use of an apple to refer to computers is pretty far-fetched, coming close, but not quite there, to an arbitrary use of the representation of an apple. The metonymical chair used to refer to the program Director (Figure 7b) comes somewhat closer to literal meaning as the chair is one which is often associated with directing films.

Figure 7. Metonymic (a) and metaphoric (b) representations.

In other words, beneath the surface of literal meaning, there is a deep in which any non-arbitrary expression of information in any modality can be used to potentially good effect in interaction design by making the expression convey non-literal information. At the bottom of this deep we find the arbitrary modalities, which have no (intended) relationship of meaning to that to which they refer. The idea is depicted in Figure 8. How to exploit the idea to good effect in interaction design is something, which, as usual, goes beyond Modality Theory.

In Section 2.3.2 we found that modalities can do more than just represent information. Modalities can be organised into *modality structures*, such as lists and tables, and modalities can assume *modality roles*, such as when being used as icons. In this section, we have seen that modalities can have *non-literal uses* in addition to, or in replacement of, their literal uses. These phenomena are not exclusive. It is perfectly possible, for instance, to make a table of metaphorical icons. Figure 8 is a case in point. The figure uses an explicit structure, abbreviated text, and image icons to metaphorically illustrate conceptual points made in the text.

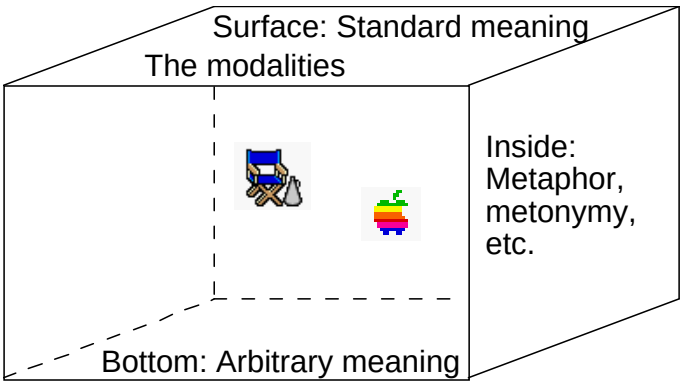

Figure 8. Non-literal uses of modalities can be located in the conceptual space between literal meaning and arbitrary meaning. The figure itself is a metaphor.

3. A FORMAT FOR REPRESENTING MODALITIES

Considered in isolation, the taxonomy of unimodal output modalities simply is a hierarchical analysis of the space of representational modalities in the media of graphics, acoustics and haptics. Gradually, as it were, 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, (c) all individual unimodal modalities have been analysed as regards their properties and their capabilities and limitations for representing different types of information in context, and (d) other phenomena related to modality use, such as tables and metaphors, have found their proper place in the model. Points (a), (b), (c) and (d) have all been addressed above to some extent. Some time ago, we analysed all the unimodal modalities presented above and implemented them in a software demonstrator (Bernsen & Lu, 1995). We represented the analysis of each modality in a common format using a *modality document* template. Subsequent work on speech functionality (Section 5) has introduced some modifications, which have been included in the template below. The - still pending - completion of the taxonomy of unimodal input modalities may produce further revisions of the proposed common format for modality representation. This means that the following modality document template is a draft only.

Modality documents define, explain, analyse, and illustrate unimodal modalities from the point of view of interaction design support. The shared document structure includes the following entries:

(1) Modality profile
(2) Inherited declarative and functional properties
(3) Specific declarative and functional properties
(4) Combinatorial analysis

(5) Relevant operations
(6) Identified types-of
(7) Illustrations

What follows is a walkthrough of the modality document structure exemplified by illustrations from various modality documents.

(1) *Modality profile.* A notation is used to express the profile of the modality, i.e. the combination of basic properties which defines the modality as being distinct from other modalities at the same level of abstraction (cf. Tables 2 and 3).

(2) *Inherited declarative and functional properties.* These are the properties, basic or otherwise, which the modality inherits from its parent nodes in the taxonomy hierarchy. *Declarative properties* characterise the unimodal modality in a way, which is independent of its use. *Functional properties* characterise the unimodal modality as to which aspects of information it is good or bad at representing in context. An example is the claim that high-specificity (a large amount of detail in as many information channels as possible), high-image resolution, high-dimensionality (2 1/2D or 3D better that 2D) graphic images are useful for facilitating the visual identification of objects, processes, and events. An illustration of this functional property, 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 (Bernsen, 1995). Use of static graphic images, such as the one shown in Figure 11, 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, and irrespective of whether these are embodied in graphics, acoustics or haptics. The issue of functional properties of modalities is an extremely complex one, however, as we shall see (Section 5). For this reason, it is not entirely clear at present, which, if any, functional information should be included in individual modality documents. Except for the super level modalities, all unimodal modalities inherit important parts of their properties from higher levels in the taxonomy. A generic-level modality inherits the declarative and functional properties of its parent node at the super level, an atomic modality inherits the properties of its parent nodes at the super and generic levels, etc. To keep individual modality documents short, those properties should be retrievable through hypertext links. The following example shows the list of links to inherited properties in the atomic-level *gestural notation* modality document (Table 4) as well as the information channel and dimensionality information provided in the document.

- 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 1D, 2D and 3D spatial, time.

The important analytical concept of an information channel may be briefly explained as follows. When, in designing human-human-system interaction, we choose a certain (unimodal) output modality to represent information, this modality inherits a specific medium of expression, such as acoustics, which it shares with a number of other unimodal modalities. An *information channel* is a perceptual aspect (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. A list of the information channels, which are characteristic of a particular unimodal modality makes explicit the full inventory of means for information representation available to the developer using that modality.

Returning to the example above, gestural notation inherits the properties of the linguistic, static, dynamic, graphic, and notational modalities. As 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 (or fusion) strategy (Section 2.2), the gestural notation document presents both static and dynamic gestural notation. Figure 9 shows an example of static gestural notation.

(3) *Specific declarative and functional properties.* 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.

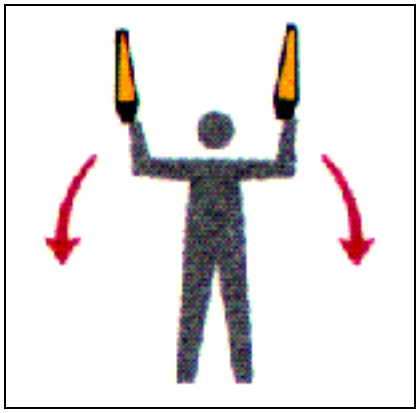

Figure 9. Static gestural notation: a marshalling signal which means 'move ahead'.

This is illustrated in Figure 10 in which ad hoc use of the graphic information channel colour (blue for the left-hand bar and green for the right-hand bar in a pair) has been defined in static graphic typed labels/keywords in the graph legend in the top right-hand corner. Without this linguistic annotation, it would not be possible to interpret the bar graph shown. The graph compares waste recycling of aluminium, glass and paper in the years 1970 and 1991 in the USA. An interesting point about the bar graph in Figure 10 is that the choice of the colours blue and green is not entirely arbitrary. In fact, the green colour used for 1991 manages to metaphorically suggest environmentalist progress. Metaphoric and other non-literal uses of modalities are discussed in Section 2.6.

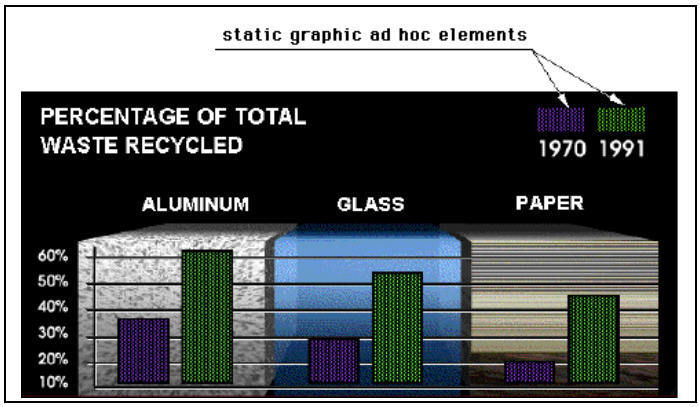

Figure 10. Dependence on linguistic modalities of an information channel used ad hoc.

(4) *Combinatorial analysis* expresses 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". This is illustrated in Figure 12. The figure represents a Macintosh window as a layered series of unimodal explicit static graphic structures. 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" or "chemistry". Like modality functionality analysis, combinatorial analysis faces a high-complexity problem space. We cannot claim as yet to have demonstrated a workable solution to the problem of how to do systematic combinatorial analysis (see Section 6).

Figure 11. A unimodal static graphic image of high specificity.

(5) *Relevant operations* are operations, which can be applied to the unimodal output modality described in a particular template. 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 typically is to bring out more clearly particular aspects of the information to be presented. *Dimensionality reduction,* as in reducing common road maps from 3D to 2D without loss of key information; *specificity reduction,* as in replacing an image with a sketch; *saliency enhancement,* as in selective colouring; and *zooming* are some of the operations applicable to analogue graphic modalities. Similarly, **boldfacing**, *italicising*, underlining and re-sizing are common operations in graphic typed languages.

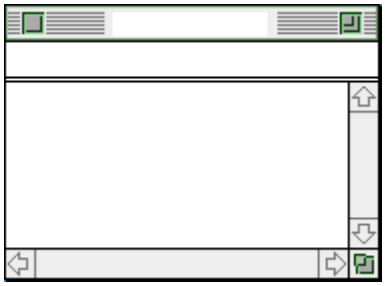

Figure 12. Nested unimodal explicit static graphic structures: the Macintosh window.

(6) *Identified types-of.* These are simply the sub-types of a unimodal modality, which are to be found one level down in the taxonomy hierarchy. In this way, the modality template presents an overview of the daughter modalities of the unimodal modality, which is currently being inspected. For instance, dynamic non-analogue sign graphic language (generic level) has six atomic types:

(1) Dynamic graphic written text
(2) Dynamic graphic written labels/keywords
(3) Dynamic graphic written notation
(4) Graphic spoken text or discourse
(5) Graphic spoken labels/keywords
(6) Graphic spoken notation

(7) *Illustrations.* Each modality document is illustrated by some 5-10 illustrations selected such as to show both prototypical examples, important non-prototypical or marginal cases, interesting multimodal combinations, non-literal uses, etc.

In addition to the systematic representation of all unimodal modalities in a common format, Modality Theory comes with a number of supporting theoretical concepts some of which have been mentioned already, such as 'specificity', 'interpretational scope', ‘modality role’, 'saliency', 'information channel', or 'dimensionality'. Theoretical concepts are explained and illustrated in *lexicon documents.* There are currently about 70 such documents (or concepts). Due to the heterogeneous nature of their topics, no rigid document structure has been imposed on lexicon documents.

4. TOWARDS A TAXONOMY OF UNIMODAL INPUT

The present section briefly describes ongoing work on a taxonomy of unimodal input modalities. We are almost, but not quite, there with a taxonomy of unimodal input modalities corresponding to the taxonomy of unimodal output modalities presented above. It may be illuminating to present and discuss some reasons why it

has proved somewhat hard to arrive at a taxonomy of input modalities because, naively, one might have assumed that this would be merely a matter of mirroring the taxonomy of unimodal output modalities. Well, it is not.

It is interesting to note that the state of the art in input taxonomies is probably even less developed than the state of the art in output taxonomies. With respect to the latter, we have seen (Sections 1 and 2) that most existing taxonomies were confined to the static graphic modalities, reflecting the fact that static graphics antedate even the pyramids. The exception is the purely empirical output modality lists, which have been produced quite recently based on the multitude of studies of individual modality combinations during the 1990s.

Input, on the other hand, is in one sense a historical novelty because we did not have "real" input until the computer came along. In another sense, this picture is a rather superficial one, as we shall see. Not surprisingly, the first input taxonomies were proposed in the 1980s in an attempt to produce a scientific basis for addressing design choices among the growing number of different haptic input devices, such as the mouse, the joystick, the button, the pen, etc. The leading question, thus, was *not* one of distinguishing among modalities but of distinguishing among devices according to what a particular device was good or bad for. Some examples are Lenorovitz et al. (1984), who based their taxonomy of haptic input devices on *user goals*, such as 'create', 'indicate', 'eliminate', 'manipulate' and 'activate'. Foley et al. (1984) based their taxonomy of haptic input devices on basic input *interaction tasks*, such as 'select', 'position', 'orient', 'path', 'quantify' and 'text', and *control tasks*, such as 'stretch', 'sketch', 'manipulate' and 'shape'. Similarly, Greenstein & Arnaut (1988) distinguished between input tasks, such as 'point', 'draw', 'trace' and 'track'. Other researchers in this line of research are, e.g., Buxton (1983) and Mackinlay et al. (1990). Their common goal was to create a systematic basis of rules or guidelines of the form: "if the task (or user goal) is Tx (or UGy), then use haptic device HDn". Good systematic work was done towards these aims. Nevertheless, this line of research seems to have disappeared in the early 1990s. The reasons why this happened are interesting and important. The researchers gradually realized that they were building on extremely unstable foundations. The fact is that we still do not have any stable taxonomies of user goals (as it happens, psychologists already gave up on that in the 1940s) or user tasks. Moreover, we do not seem likely to have such taxonomies in the foreseeable future, as will be argued in Section 5. But if those taxonomies are unattainable for practical purposes, prospects are dim to ever reach any kind of systematicity or closure on sets of rules or guidelines of the form described above. Another fact of the matter is that new devices get invented at increasing speed, leaving any device taxonomy gasping for breath to catch up and revise its user goals or task type foundations post hoc. This is a problem for any attempt at creating a science-based device taxonomy.

Other than the haptic input taxonomies just described, few attempts appear to have been made, and none appear to have been made to gain a systematic grasp of input *modalities.* The best we have are lists of input devices, modalities and tasks based on the literature from the 1990s (e.g. Benoit et al., 2000), and focused sets of contributions towards understanding the pros and cons of particular modalities and multimodal combinations but lacking in attempts at systematic theoretical

comprehension. A good example is the volume of articles on speech functionality edited by Baber & Noyes (1993).

Input modalities are forms of representation of information from the user to the machine. When we began work on a taxonomy for unimodal input modalities, a natural first question was whether it might be possible to simply re-use the output taxonomy or whether the output taxonomy would have to be modified to account for input. Closer analysis revealed a number of – real or purported - asymmetries between output and input modalities. The ones identified so far are discussed in the following.

4.1. Asymmetries between Output and Input

4.1.1. Perceptual asymmetry

Obviously, interactive *output* must be perceptible to humans in HHSI. This is not necessarily true for input where it is sufficient that the machine is able to perceive what arrives at its sensors. Thus, some input media aspects are not (fully) perceptible to humans, such as radar, infrared, ultrasound, magnetic fields, skin conductivity, etc. An implication is that our list of output media *information channels* is likely to have to be augmented when dealing with input.

4.1.2. Media asymmetry

It is an obvious fact that haptics is, at least so far, much more prominent in input than in output. As far as output is concerned, current systems can output information in any output modality as long as we have plenty of time to create the output at design-time, and as long as we have built the output devices we need. In input, haptics dominate. We key in text, point, track and click with the mouse, write and draw with the pen, manipulate 3D objects to modify graphical output representations, move in force-feedback output space, etc. Correspondingly, as we saw above, most input theory is about haptic input, focusing on the haptic control and/or creation process. There are at least two reasons why this is the case. The first reason is that humans are very good at using their hands, which is reflected in the fact that most of the control devices, which have been invented before and after the industrial revolution are haptic devices. The second reason is that acoustic (including speech) recognition and understanding by machine has been difficult to achieve and that visual recognition and understanding by machine is more difficult still. Workable speech recognition systems have been around for about two decades only, and workable spoken dialogue systems have been around for about a decade only (Bernsen et al., 1998). It is not surprising, therefore, that haptic devices continue to proliferate.

Based on the above observations, the following claim would appear plausible: we should forget about (most of) *device theory* and focus on Modality Theory instead. As was pointed out earlier, haptic device theory became extinct because scientific soundness was missing. There are no formally sound taxonomies of user goals or tasks in HHSI. Modality theory has stronger formal properties, at least as concerns output. Secondly, device theory would seem destined to become much too

simplistic for the non-haptic input media. All we need as regards input devices for acoustics and graphics are microphones, cameras, infrared sensors, and a few more. Even factoring in more detailed distinctions among technologies, such as between direction-sensitive microphones, microphone arrays and the like, device theory is not likely to attain the level of detail which is needed to support interaction design. Thirdly, input devices are generally easy to build and modify, which is also why it is virtually impossible to keep track on developments. In fact, there is something deeply puzzling about the haptic input device theory endeavor, something which we have termed the *door knob problem*. Who ever saw the ultimate textbook on door-knobs, or on knives? Mankind has been using door-knobs for centuries and knives for much longer, yet the literature on door-knobs or knives, the taxonomies of them, the proposals for scientific foundations for judging the appropriateness or inappropriateness of various types of knives and door-knobs for particular purposes, etc., is non-existent. The simple reason appears to be that if a particular ”type” of knife or door-knob does not work for a particular purpose, we just design and manufacture a better one. The same holds, or will soon hold, it is suggested, for haptic input devices. So, who needs device theory? Or, re-phrased from the point of view of Modality Theory, let us start with the information to be exchanged in HHSI and then find, or make, the input devices, which can meet the requirements.

4.1.3. Is input ”more essentially interactive” than output?

When investigating possible asymmetries between output and input, we also came across more questionable purported asymmetries than the above, such as the following.

If we focus on what people *do* with systems during HHSI, it is tempting to conclude that we mostly just *receive* ready-made output but that we often *create* things with the input we produce, such as text, drawings, 3D graphics objects, soundscapes, a flight ticket reservation made through spoken input, etc. We are less accustomed to the point of view that machines also do something to, or with, humans during interaction. Thus, the output taxonomy presented above has focused on ready-made results, such as the images or the music to be presented as system output, rather than on the process of interactively creating those output modalities. However, when considering input, it is much harder to ignore the *input process* through which we create those (output) representations. In other words, *input is input into an output space*, which gets modified as a result of the input. With today’s WYSIWYG and direct manipulation interfaces, we have become accustomed to getting ample immediate feedback on the results of our input actions.

It is less easy to accept that *output is output into an input space* (i.e. the user’s cognitive system), which gets modified as a result of the output. Nevertheless, this is how the *developer* always had to conceive of output creation, i.e. that the output is there to modify the cognitive state and behavior of the humans interacting with the system. To be meaningful in context, both input and output require knowledge of the interlocutor’s state and how it is likely to change as a result of the information, which is being sent across. An example in point is spoken language dialogue systems output design during which the developer must constantly refer to how the

user is likely to understand the system's output (Bernsen et al., 1998). With increasingly advanced (or more intelligent) machines able to do flexible and adaptive interaction on-line, this burden will gradually shift from the developer to the system itself. In other words, so far, users need feedback more than machines do; so far, users tend to work into an output environment more than machines tend to work into an input environment on-line; so far, machines mostly produce "finished things" on-line; and so far, machines are tools for humans more than humans are "tools" for machines. But this only goes to say that, so far, machines are (mostly) dumber than humans in their perception and understanding of their interlocutors during HHSI. And this will gradually change.

4.1.4. Dynamic/static modalities asymmetry

An important difference between output and input seems to be that the dynamic/static distinction does not apply to machines. It does not matter to the machine whether the input it receives is static or dynamic in the sense of these terms used in Modality Theory (Section 2.1). The reason is that machines are endowed with "photographic memory", as a result of which they tend to capture all the information contained in the input, within, of course, the capabilities of their input sensors. Once captured, the machine does not profit from, or need, having exactly the same input repeated. Instead, they use their processing power to internally exploit the information already received. Humans are different. They do profit from repetition of complex information, such as a static graphics screenful of information, just as they profit from the constant availability of complex information whilst they make up their minds on what to do next. This is why the static modalities are so important to humans in many cases, and why humans have difficulties receiving complex information dynamically. Nothing prevents us from building machine, which share with humans the property of selective focusing within a complex information space, machines which quickly forget most of the complex information they are being exposed to and which have to refresh their memory from time to time by re-perceiving the information with a changing perceptual and cognitive focus. So far, however, there seems to be little point in building such machines. Removal from the unimodal input taxonomy of the static/dynamic distinction significantly simplifies input taxonomy generation. For input, the basic matrix in Table 1 reduces to 24 initial modalities (Table 10).

4.1.5. Conclusion

In conclusion, when generating the taxonomy of unimodal input modalities, we need to watch out for input which is perceptible to machines but not perceptible to humans. We will probably have to expand the haptic section of the input taxonomy compared to that in the output taxonomy given the importance of haptics in human-to-machine interaction. The problem, however, in expanding on input haptics is that this should not be done through reference to scientifically unsound taxonomies of user goals, tasks or otherwise. Similarly, we should continue to resist the temptation to involve (input) device properties in the taxonomy even if those properties are sometimes important to device selection. Rather, if none of the devices available are

appropriate, we build a new and better (faster, more precise, etc.) one. Finally, we should omit the static/dynamic distinction in the input taxonomy.

Table 10. The full set of 24 combinations of basic properties constituting the possible unimodal input modalities at the generic level of the taxonomy. All modalities provide possible ways of representing information.

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

4.2. Towards a taxonomy of unimodal input modalities

Given the argument above, the first step of generating a complete matrix of unimodal input modalities at the generic level is easy (Table 10). Removing the modalities which represent the arbitrary use of information tokens which already have an established meaning is simple as well (Table 11). We are currently experimenting with the generation of the atomic level of unimodal input modalities. The way this is being done is to use questionnaires in the form of tables showing an atomic-level breakdown of the generic level into intuitive and relevant atomic unimodal input modalities. Subjects are asked to fill in each cell of the table with one or more concrete examples in which the atomic input modality in point is being produced interactively into a particular output space and using specific input devices. This process generates issues of classification and consistency with the atomic output taxonomy, which are then being analyzed and discussed. For illustration, some examples from the questionnaires follow.

Table 11. 15 generic unimodal input modalities result from removing from Table 10 the arbitrary use of non-arbitrary modalities of representation.

	li	*-li*	*an*	*-an*	*ar*	*-ar*	*gra*	*aco*	*hap*
1	x		x			x	x		
2	x		x			x		x	
3	x		x			x			x
4	x			x		x	x		
5	x			x		x		x	
6	x			x		x			x
7		x	x			x	x		
8		x	x			x		x	
9		x	x			x			x
10		x		x	x		x		
11		x		x	x			x	
12		x		x	x				x
13		x		x		x	x		
14		x		x		x		x	
15		x		x		x			x
	li	*-li*	*an*	*-an*	*ar*	*-ar*	*gra*	*aco*	*hap*

(1) Linguistic graphics: Written notation: A mathematician in front of a blackboard filled with formulae tells the system: “Please digitise and store!”
(2) Analogue graphics: Maps: A person shows the system a map and asks for the nicest itinerary to a particular landmark.
(3) Analogue acoustics: Graphs: Singing people compete to produce overtones in Italian arias. The system is the judge.
(4) Analogue haptics: Manipulation/action: A person explores a VR cityscape on a bicycle.

5. APPLYING MODALITY THEORY TO SPEECH FUNCTIONALITY

Following the research agenda of Modality Theory, we should at this point address the issue of how to combine unimodal output modalities, unimodal input modalities, and unimodal input/output modalities into usable multimodal representations. However, as the taxonomy of unimodal input modalities is not quite ready yet, this issue will be postponed to the final section of the present chapter. This section briefly addresses the two final issues on the research agenda of Modality Theory, i.e. (4) to develop a methodology for applying Modality Theory to early design analysis of how to map from the requirements specification of some application to a usable selection of input/output modalities, and (5) to use results in building, possibly automated, practical interaction design support tools. These two issues have been the subjects of substantial work for several years already, resulting in a number of system design case studies, comprehensive studies of speech functionality, and a

design support tool. Before describing the progress made, it may be useful to review some false starts on methodology, which, as was discovered in the process, we were not the only ones to make.

5.1. How (not) to Map from Requirements Specification to Modality Selection

As a result of a series of multimodal systems design case studies, a consolidated methodology for mapping from requirements specification to modality selection was specified (Bernsen & Verjans, 1997). The basic notion of the methodology was that of *information mapping rules.* The modality document template (cf. Section 3) would list, per unimodal output or input modality, the information mapping rules applying to that modality. The rules would then be used, manually at first but eventually automatically by a rule-based system, to generate advice on which modalities to use in designing interaction for a particular application. The final product would be an 'interface sketch' which identifies the best input and output modalities to use, the devices to use, and the interactive functionalities which the artefact to be developed would need. All the sketch would need to become a full design specification was a more detailed task analysis together with the application of a particular standard or aesthetic design concept for controlling the detailed specification of the interface. As it turned out, the inherent complexity of the problem space of selecting among thousands of potential modality combinations subject to multiple constraints imposed by the context of use of the artefact to be developed, proved too great for a rule-based approach. We were envisioning getting bogged down by thousands of rules expressed in a basically unsound and non-transparent state of the art conceptual apparatus for describing tasks, user groups, user goals, and many other parameters as well. We also discovered that other researchers, such as the group in Namur, Belgium (e.g. Bodart et al., 1995), had spent years working on a small (static data graphics) fraction of the modalities covered by Modality Theory, only to arrive at the same adverse conclusion. In other words, we had to abandon methodology A below.

A. Rule-based mapping onto modality selection:

(1) Initial requirements specification ->
(2) rules specifying what a particular modality can be used for ->
(3) interface sketch and device selection ->
(4) detailed task analysis and interface specification.

In retrospect, this methodology shares the flaws of the haptic input taxonomy work reviewed in Section 4, i.e. methodology B below.

B. Rule-based mapping onto device selection:

(1) Requirements specification ->
(2) task/goal taxonomy + rules ->
(3) mapping onto device selection.

After these lessons in infeasibility, it was clear that any solution to the mapping problem would have to be based on a significant reduction in complexity. We also knew that some of the possible complexity reductions would not be sufficient. Thus, merely reducing complexity to a sub-section of unimodal modality space would not help. We would still be bogged down by scientifically unsound, as well as unmanageable, rules. So, the rule-based approach common to Methodologies A and B above had to go. Still, reducing complexity to a sub-section of unimodal modality space might at least help us experiment with new methodologies without having to take on all unimodal modalities and their thousands of multimodal combinations from the outset. We decided to consider only speech output and speech input together with multimodal representations, which include speech. During the study of the literature on speech functionality, i.e. on the issue of when it is (not) advisable to use speech output and/or speech input in interaction design, it became clear that there is a crucial distinction between the rules involved in the rule-based approaches A and B above, and principles which simply state declarative properties (cf. Section 3) of unimodal modalities. A rule would state something like:

> *If the task is Tl and the user group is UGm and the goal is to optimise efficiency of interaction, then use modality (or modality combination) Mn.*

The problem, as we have seen, is the scientifically messy notions of task types, user group types and performance parameter types (such as efficiency). Moreover, there is not even consensus about the number and nature of the *types* of relevant parameters. On the other hand, the modalities themselves and their declarative characteristics *is*, potentially at least, a scientifically sound part of a possible methodology. An example of a declarative characteristic of a modality is:

> *Speech is omnidirectional.*

We call such characteristics *modality properties.* The methodology (C below), then, is based on modality properties rather than rules.

C. Modality property-based mapping onto modality selection:

(1) Requirement specification ->
(2) modality properties + natural intelligence ->
(3) advice/insight with respect to modality choice.

Clearly, methodology C is considerably more modest than methodologies A and B above. In particular, C is non-automated and relies on human intelligence for deriving clues for modality choice decisions from modality properties. All we can do, it would seem, to help train and sharpen the human intelligence doing the derivations, is to provide a wide range of concrete interaction design examples, each of which has been analysed with respect to the modality choice claims or decisions made. This we set out to do for the speech functionality problem.

5.2. Speech Functionality

In two studies, we have analysed in depth 273 claims about speech functionality found in the literature on the subject from 1993 onwards (Bernsen, 1997; Bernsen & Dybkjær, 1999a). The claims were selected in an objective fashion in order to prevent exclusion of examples, which might prove difficult to handle through Methodology C (see Section 5.1). Thus, the first set of 120 claims studied was the entire set of claims to be found in Baber & Noyes (1993). The second set of claims was identified from the literature on speech functionality since 1993 by a researcher other than the one who would be analysing the set. The exercise has provided a solid empirical foundation for judging just how complex the speech functionality problem is in terms of the number of instantiated parameters involved (the modality complexity we knew already, more or less). The complexity of the problem of accounting for the functionality of speech in interaction design is apparent from the semi-formal expression from (Bernsen & Dybkjær, 1999b) in Figure 13. Parameters are in boldface.

[Combined speech input/output, speech output, or speech input modalities M1, M2 and/or M3 etc.] or
[speech modality M1, M2 and/or M3 etc. in combination with non-speech modalities NSM1, NSM2 and/or NSM3 etc.] are [useful or not useful] for:
[**generic task** GT and/or **speech act type** SA and/or **user group** UG and/or **interaction mode** IM and/or **work environment** WE and/or **generic system** GS and/or **performance parameter** PP and/or **learning parameter** LP and/or **cognitive property** CP] and/or are:
[preferable or non-preferable] to:
[alternative modalities AM1, AM2 and/or AM3 etc.] and/or are:
[useful on conditions] C1, C2 and/or C3 etc.

Figure 13. Minimal complexity of the speech functionality problem.

Note that each of the boldfaced parameters can be instantiated in many different ways. For instance, Bernsen (1997) found 38 different instantiations of the parameter *performance parameter* in the 120 claims analysed. It is a sobering thought that any systematic approach to modality choice support must face this complex parameter space in addition to the complexity of the space of unimodal input/output modalities and their combinations.

Prior to the claims analysis, each of the 273 claims from the literature was rendered in a standard semi-formal notation in order to facilitate analysis. Each rendering-cum-analysis had to be independently approved by a second researcher. An example of a rendering-cum-analysis is Claim (or data point) 48 in Figure 14, quoted from (Bernsen & Dybkjær, 1999b). The claim representation first quotes the original expression of the claim followed by a literature reference. The claim is then expressed in a standard format referring to the modalities and parameters involved.

The statement "Justified by MP5" refers to one of the modality properties used in evaluating the claims (Table 12). The claims evaluation is the centre-piece of the work reported here. For each claim, search was made among the hundreds of modality properties of Modality Theory to identify those properties, which could either justify, support (but not fully justify), or correct the claim. In the case of Claim 48, as shown, one modality property was found which justifies the claim. The Claims type "Rsc" refers to claims, which recommend the use of speech in combination with other modalities. The claims evaluation in Figure 14 is followed by an (optional) note on the claim and its evaluation. Finally, the claim itself is evaluated as being true. This evaluation of the truth-value of claims is important, because the potential of modality properties for claims analysis is basically to be judged by the percentage of true claims which Modality Theory, through reference to modality properties, is able to justify or support, and the percentage of false or questionable claims which the theory is able to correct. It would be bad news for the approach if there were many true, questionable, or false claims on which the theory had nothing to say.

48. Interfaces involving spoken ... input could be particularly effective for interacting with dynamic map systems, largely because these technologies support the mobility [walking, driving etc.] that is required by users during navigational tasks. [14, 95]
Data point 48. **Generic task** [mobile interaction with dynamic maps, e.g. whilst walking or driving]: a speech input interface component could be **performance parameter** [particularly effective].
Justified by MP5: "Acoustic input/output modalities do not require limb (including haptic) or visual activity." Claims type: **Rsc.**
NOTE: The careful wording of the claim "Interfaces involving spoken ... input". It is not being claimed that speech could suffice for the task, only that speech might be a useful interface ingredient. Otherwise, the claim would be susceptible to criticism from, e.g., MP1. Note also that the so-called "dynamic maps" are static graphic maps, which are interactively dynamic.
True.

Figure 14. Evaluation of a claim about speech functionality.

What we found, however, was that Modality Theory was able to justify, support, or correct 97% of the claims in the first study of 120 claims (Bernsen, 1997), and 94% of the claims in the second study of 153 claims (Bernsen & Dybkjær 1999a). In other words, assuming, as argued in those two studies, the representativity of the analysed claims with respect to all possible claims about speech functionality, modality properties – i.e. the clean, declarative messages of Modality Theory – are highly relevant to judging speech functionality in early design and development.

A final important question is: how many modality properties were needed to achieve the high percentages reported in the preceding paragraph? In fact, the first

Table 12. Modality properties found relevant to speech functionality evaluation.

No	Modality	MODALITY PROPERTY
MP1	Linguistic input/output	Linguistic input/output modalities have interpretational scope, which makes them eminently suited for conveying abstract information. They are therefore unsuited for conveying high-specificity information including detailed information on spatial manipulation and location.
MP2	Linguistic input/output	Linguistic input/output modalities, being unsuited for specifying detailed information on spatial manipulation, lack an adequate vocabulary for describing the manipulations.
MP3	Arbitrary input/output	Arbitrary input/output modalities impose a learning overhead which increases with the number of arbitrary items to be learned.
MP4	Acoustic input/output	Acoustic input/output modalities are omnidirectional.
MP5	Acoustic input/output	Acoustic input/output modalities do not require limb (including haptic) or visual activity.
MP6	Acoustic output	Acoustic output modalities can be used to achieve saliency in low-acoustic environments. They degrade in proportion to competing noise levels.
MP7	Static graphics/haptics input/output	Static graphic/haptic input/output modalities allow the simultaneous representation of large amounts of information for free visual/tactile inspection and subsequent interaction.
MP8	Dynamic input/output	Dynamic input/output modalities, being temporal (serial and transient), do not offer the cognitive advantages (wrt. attention and memory) of freedom of perceptual inspection.
MP9	Dynamic acoustic output	Dynamic acoustic output modalities can be made interactively static (but only small-piece-by-small-piece).
MP10	Speech input/output	Speech input/output modalities, being temporal (serial and transient) and non-spatial, should be presented sequentially rather than in parallel.
MP11	Speech input/output	Speech input/output modalities in native or known languages have very high saliency.
MP12	Speech output	Speech output modalities may complement graphic displays for ease of visual inspection.
MP13	Synthetic speech output	Synthetic speech output modalities, being less intelligible than natural speech output, increase cognitive processing load.
MP14	Non-spontaneous speech input	Non-spontaneous speech input modalities (isolated words, connected words) are unnatural and add cognitive processing load.
MP15	Discourse input/output	Discourse input/output modalities have strong rhetorical potential.
MP16	Discourse input/output	Discourse input/output modalities are situation-dependent.
MP17	Spontaneous spoken labels/keywords and discourse input/output	Spontaneous spoken labels/keywords and discourse input/output modalities are natural for humans in the sense that they are learnt from early on (by most people and in a particular tongue and, possibly, accent). (Note that spontaneous keywords and discourse must be distinguished from designer-designed keywords and discourse which are not necessarily natural to the actual users.)
MP18	Notational input/output	Notational input/output modalities impose a learning overhead which increases with the number of items to be learned.

MP 19	Analogue graphics input/output	Analogue graphics input/output modalities lack interpretational scope, which makes them eminently suited for conveying high-specificity information. They are therefore unsuited for conveying abstract information.
MP 20	Haptic manipulation selection input	Direct manipulation selection input into graphic output space can be lengthy if the user is dealing with deep hierarchies, extended series of links, or the setting of a large number of parameters.
MP 21	Haptic deixis (pointing) input	Haptic deictic input gesture is eminently suited for spatial manipulation and indication of spatial location. It is not suited for conveying abstract information.
MP 22	Linguistic text and discourse input/output	Linguistic text and discourse input/output modalities have very high expressiveness.
MP 23	Images input/output	Images have specificity and are eminently suited for representing high-specificity information on spatio-temporal objects and situations. They are therefore unsuited for conveying abstract information.
MP 24	Text input/output	Text input/output modalities are basically situation-independent.
MP 25	Speech input/output	Speech input/output modalities, being physically realised in the acoustic medium, possess a broad range of acoustic information channels for the natural expression of information.

study needed only 18 modality properties whilst the second study needed seven additional modality properties. It may thus be concluded that a relatively small number of modality properties constitute an extremely powerful resource for evaluating most speech functionality claims or assumptions in early design and development. The modality properties used in the two studies are listed in Table 12.

5.3. The SMALTO Tool

The results presented in Section 5.2 convinced us that it might be worthwhile to develop a design support tool for supporting early design decisions on whether or not to use speech-only or speech in multimodal combinations for particular applications. The tool is called SMALTO and can be accessed at http://disc.nis.sdu.dk/smalto/. Basically, what SMALTO does is to enable hypertext navigation among hundreds of evaluated claims made in the literature as well as among the modality properties, which bear on those claims. The benefits to be derived from using the tool are to become familiar with the specific modality thinking which bears on the design task at hand in case claims are found which are immediately relevant to that design task, and to become increasingly familiar with the general modality thinking which can be done straight from an understanding of the modality properties themselves (Luz & Bernsen, 1999).

6. MULTIMODALITY

Getting a theoretical handle on multimodality would constitute a major result of Modality Theory. As this is work in progress, we are not yet able to present any

well-tested approach, which could be claimed to be superior to, or a valuable complement to, the best current approach.

6.1. The Best Current Approach

The best current approach to the issue of multimodality as described in this chapter is an empirical one. The approach consists in, quite simply, analysing and publicising "good compounds", i.e. good modality combinations with added observations on the parameter instantiations which have been studied or which are being hypothesised about (cf. Section 5.2). As has become clear from the argument in this chapter, no modality combination is good for every conceivable purpose and it is difficult, to say the least, to precisely circumscribe the circumstances in which a particular modality combination should be preferred to others for the representation and exchange of information. Still, it is useful to build systematic overviews of modality combinations, which have proved useful for a broad range of specified purposes. Some such combinations have already been described above, such as bimodal static graphics including, e.g., linguistically labelled graphs and static graphics images, which illustrate static graphic text. There are very many other "good compounds", such as speech combined with static or dynamic graphics output, speech and pen input, speech and analogue haptics output for the blind, etc. At least, when using a modality combination which has been certified as a good one under particular circumstances, developers will know that they are not venturing into completely unexplored territory but can make the best use of what is already known about their chosen modality combination. The problems about this approach are, first, that it tends to be a rather conservative one, dwelling upon modality combinations which have been used so often already that some kind of incomplete generalisation about their usefulness has become possible. Thus, the approach lacks the predictive, creative and systematic powers of a firm theoretical grasp of the space of possible modality combinations. Secondly, given the complex parameter space addressed by any claim about the usefulness of a particular modality combination, most of those generalisations are likely to be scientifically unsound ones, and increasingly so the more general they are.

6.2. Modality Theory–Based Approaches

How might Modality Theory do better than the best current approach (Section 6.1)? The theory is superior to that empirical approach in that Modality Theory allows a complete *generation* of all possible input, output, and input/output modality combinations at any level, such as the atomic level, and cross-level as well. However, whilst complete generation is possible in a way that is sufficient for all practical interaction design purposes, the combinatorial explosion involved makes it practically impossible to systematically *investigate* all the generated modality combinations. For instance, if we wanted to investigate all possible n-modal atomic input/output modality combinations where n=10, the number of combinations to be investigated would run into millions. Still, there do seem to be interesting

opportunities for exploring this generative/analytic approach by carving out *combinatorial segments* from the taxonomy for systematic analysis, such as a speech-cum-other-modalities segment, or an input-manipulation-cum-other-modalities segment. These exercises could be further facilitated by tentatively clustering families of similar modalities and treating these as a single modality whose interrelations with other modalities are being investigated. An example could be to treat all analogue static graphics modalities as a single modality given the fact that these modalities have a series of important properties in common. It is perfectly legitimate to ask questions, such as "How does this particular family of tri-modal combinations combine with other modalities?"

An alternative to the generative approach just described could be to *scale up the SMALTO tool* to address all possible modality combinations. The problem, however, is that this would be likely to produce lists of hundreds of relevant modality properties, creating a space of information too complex for practical use. Part of the usefulness of SMALTO lies in the fact that SMALTO operates with such a small number of modality properties that it is humanly possible to quickly achieve a certain familiarity with all of them, including the broad implications for interaction design of each of them. It might be preferable, therefore, to use the SMALTO approach in a slightly different way, i.e. by producing modality properties for limited segments of combinatorial input/output modality space according to current needs just like SMALTO itself does.

We are currently working on a third approach which is to *turn Modality Theory as a whole into a hypertext/hypermedia tool* using a common format for modality representation similar to the format described in Section 3 but with an added entry for modality properties. By definition, the tool would include all identified modality properties. The challenge is to make the tool useful for interaction designers who are not, and do not want to become, experts in the theory, for instance by including a comprehensive examples database. In itself, this tool would not constitute a full scientific handle on multimodality in the sense of a systematic approach to multimodal combinations. However, building the Modality Theory tool does seem to constitute a necessary next step, which would also facilitate achievement of the ultimate goal of mastering the huge space of multimodal combinations.

Finally, a fourth approach is to *analyse the "good compounds"* (Section 6.1) in terms of modality properties in order to explore whether any interesting scientific generalisations might appear.

6.3. Conclusion

This chapter has outlined the current state of progress on the research agenda of Modality Theory. So far, Modality Theory has taken a different route from most recent work on modalities, which has focused on exploring, on an ad hoc basis, useful modality and/or device combinations based on an emerging conceptual apparatus, including concepts such as modality 'complementarity' and 'redundancy'. Modality Theory, on the other hand, has focused on generating fundamental concepts and taxonomies of unimodal output and input modalities subject to the

requirements of completeness, uniqueness, relevance, and intuitiveness; on exploring methodologies for applying the emerging theory in design and development practice; and on developing demonstrator tools in support of modality choice decision making in early design of human-human-system interaction. In the process, a good grasp has been achieved of the extreme complexity of the problem of modality functionality. To complete the research agenda of Modality Theory, we need a well-tested taxonomy of unimodal input modalities, a Modality Theory hypertext/hypermedia tool, and exploration of additional ways in which the theory can be of help in achieving a systematic, creative, and predictive understanding of input/output modality combinations, including those which have not yet been widely used, if at all. These themes are topics of current research, which the reader is kindly invited to join.

7. REFERENCES

Baber, C. & J. Noyes (Eds.). *Interactive Speech Technology*. London: Taylor & Francis, 1993.

Benoit, C., J.C. Martin, C. Pelachaud, L. Schomaker & B. Suhm. "Audio-Visual and Multimodal Speech Systems." In: D. Gibbon (Ed.), *Handbook of Standards and Resources for Spoken Language Systems* - Supplement Volume. Kluwer, 2000.

Bernsen, N.O. "A research agenda for modality theory." In: Cox, R., Petre, M., Brna, P., and Lee, J. (Eds.), *Proceedings of the Workshop on Graphical Representations, Reasoning and Communication.* World Conference on Artificial Intelligence in Education. Edinburgh, 1993: 43-46.

Bernsen, N.O. "Foundations of multimodal representations. A taxonomy of representational modalities." *Interacting with Computers* 6.4 347-71, 1994.

Bernsen, N.O. "Why are analogue graphics and natural language both needed in HCI?" In: Paterno, F. (Ed.), *Design, Specification and Verification of Interactive Systems. Proceedings of the Eurographics Workshop*, Carrara, Italy, 165-179. *Focus on Computer Graphics.* Springer Verlag, 1995: 235-51, 1994.

Bernsen, N.O. "Towards a tool for predicting speech functionality." *Speech Communication* 23: 181-210, 1997.

Bernsen, N.O. "Natural human-human-system interaction." In: Earnshaw, R., R Guedj, A. van Dam & J. Vince (Eds.). *Frontiers of Human-Centred Computing, On-Line Communities and Virtual Environments.* Berlin: Springer Verlag, 2000.

Bernsen, N.O. & L. Dybkjær. "Working Paper on Speech Functionality." *Esprit Long-Term Research Project DISC Year 2 Deliverable* D2.10. University of Southern Denmark. See www.disc2.dk, 1999a.

Bernsen, N.O. & L. Dybkjær. "A theory of speech in multimodal systems." In: Dalsgaard, P., C.-H. Lee, P. Heisterkamp & R. Cole (Eds.). *Proceedings of the ESCA Workshop on Interactive Dialogue in Multi-Modal Systems*, Irsee, Germany. Bonn: European Speech Communication Association: 105-108, 1999b.

Bernsen, N.O., H. Dybkjær & L. Dybkjær. *Designing Interactive Speech Systems. From First Ideas to User Testing.* Springer Verlag, 1998.

Bernsen, N.O. & S. Lu. "A software demonstrator of modality theory." In: Bastide, R. & P. Palanque (Eds.). *Proceedings of DSV-IS'95: Second Eurographics Workshop on Design, Specification and Verification of Interactive Systems.* Springer Verlag, 242-61, 1995.

Bernsen, N.O. & S. Verjans. "From task domain to human-computer interface. Exploring an information mapping methodology." In: John Lee (Ed). *Intelligence and Multimodality in Multimedia Interfaces.* Menlo Park, CA: AAAI PressURL: http://www.aaai.org/Press/Books/Lee/lee.html, 1997.

Bertin, J. *Semiology of Graphics. Diagrams. Networks. Maps.* Trans. by J. Berg. Madison : The University of Wisconsin Press, 1983.

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

Buxton, W. “Lexical and pragmatic considerations of input structures.” *Computer Graphics* 17,1: 31-37, 1983.

Foley, J.D., V.L., Wallace & P. Chan. “The Human Factors of Graphic Interaction Techniques.” *IEEE Computer Graphics and Application* 4.11: 13-48, 1984.

Greenstein, J.S. & L.Y. Arnaut. “Input devices.” In: M. Helander (Ed.). *Handbook of Human-Computer Interaction,* Amsterdam: North-Holland, 495-519, 1988.

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

Hovy, E. & 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.

Joslyn, C., C. Lewis & B. Domik. “Designing glyphs to exploit patterns in multidimensional data sets.” *CHI’95 Conference Companion,* 198-199, 1995.

Lenorovitz, D.R., M.D. Phillips, R.S. Ardrey & G.V. Kloster. “A taxonomic approach to characterizing human-computer interaction.” In: G. Salvendy (Ed.). *Human-Computer Interaction.* Amsterdam: Elsevier Science Publishers, 111-116, 1984.

Lockwood, A. *Diagram. A visual survey of graphs, maps, charts and diagrams for the graphic designer*. London: Studio Vista, 1969.

Lohse, G., N. Walker, K. Biolsi & H. Rueter. “Classifying graphical information.” *Behaviour and Information Technology* 10, 5419-36, 1991.

Luz, S. & Bernsen, N.O. “Interactive advice on the use of speech in multimodal systems design with SMALTO.” In: Ostermann, J., K.J. Ray Liu, J.Aa. Sørensen, E. Deprettere & W.B. Kleijn (Eds.). *Proceedings of the Third IEEE Workshop on Multimedia Signal Processing,* Elsinore, Denmark. IEEE, Piscataway, NJ: 489-494, 1999.

Mackinlay, J., S.K. Card & G.G. Robertson. “A semantic analysis of the design space of input devices.” *Human-Computer Interaction* 5: 145-90, 1990.

Mullet, K. & D.J. Schiano. “3D or not 3D: ‘More is better’ or ‘Less is more’?” *CHI’95 Conference Companion,* 174-175, 1995.

Rosch, E. “Principles of categorization.” In: Rosch, E. & B.B. Lloyd (Eds.). *Cognition and Categorization.* Hillsdale, NJ: Erlbaum, 1978.

SMALTO: http://disc.nis.sdu.dk/smalto/

Stenning, K. & J. Oberlander. “Reasoning with words, pictures and calculi: Computation versus justification.” In: Barwise, J., J.M. Gawron, G. Plotkin & S. Tutiya (Eds.). *Situation Theory and Its Applications.* Stanford, CA: CSLI, Vol. 2: 607-62, 1991.

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

Tufte, E.R. *Envisioning information.* Cheshire, CT: Graphics Press, 1990.

Twyman, M. “A schema for the study of graphic language.” In: Kolers, P., M. Wrolstad & H. Bouna (Eds.). *Processing of Visual Language* Vol. 1. New York: Plenum Press, 1979.

8. AFFILIATION

Prof. Niels Ole Bernsen
Director of the Natural Interactive Systems Laboratory at the University of Southern Denmark
Main Campus: Odense University
Science Park 10
5230 Odense M, Denmark
Tel. (+45) 65 50 35 44 (direct)
Tel. (+45) 65 50 10 00 (switchboard)
Fax (+45) 63 15 72 24
email: nob@nis.sdu.dk
URL http://www.nis.sdu.dk