A SOFTWARE DEMONSTRATOR OF MODALITY THEORY

1. Introduction

For some years, the multimodal systems group at the Centre for Cognitive Science, Roskilde University, has been working on establishing and implementing the research agenda of modality theory. The research agenda for modality theory is the following (Bernsen 1993a):

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

Modality theory thus aims to establish the theoretical and methodological basis for addressing the information mapping problem in its general form, i.e.:

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

An ultimate objective is to use results in building computerised tools for the support of interface design.

We began work on the first part of this research agenda, i.e. the development of a taxonomy of output modalities in the media of graphics, acoustics and haptics. A (representational) *modality* is a way of representing information, e.g. at the human-computer interface. It was realised from early on that work progress might benefit from the support of a software tool in which we could represent large numbers of samples of output representations for the purposes of analysing their properties and testing possible taxonomy schemes. This lead to the development of Version 1 of the taxonomy workbench (May and Bernsen 1993, May and Tobin 1993), which was demonstrated at INTERCHI '93. Version 1 is a database tool designed to assist research by (a) setting up a common multimedia/multimodal database of example output representations, (b) assisting the description and classification of these examples according to different assumptions about the modalities involved, and (c) enabling thought experiments such as, e.g., the testing of different hypotheses about features of the modalities and their interrelations.

When a robust, intuitively plausible and principled taxonomy of output modalities (Bernsen 1994a,b,c, cf. below) had been established, the workbench in its current configuration had done its job of proving the usefulness of software support for modality theory development (Bernsen 1993b, May 1993a,b,c,d). This gave rise to the idea of re-designing the software tool with four objectives in mind: (1) to create a software demonstrator of the taxonomy of output modalities; (2) to use the demonstrator to further explore the functional properties of different output modalities in order to map out which information a particular modality is suited for representing; (3) to support exploration of the information mapping methodology (Bernsen and Bertels 1993, Verjans and Bernsen 1994, Verjans 1994, Bernsen and Verjans 1995); and (4) to move towards turning the demonstrator into a support tool for multimodal interface design.

The identification of functional properties of modalities is important to the achievement of the third part of the research agenda of modality theory (i.e. information mapping). We view representational modalities as having two broad kinds of property: declarative properties and functional properties. *Declarative properties* are the properties assigned to a particular modality in order to define or describe what it is. Thus, for instance, linguistic modalities share the property of being syntactic-semantic systems of meaning. The declarative properties of modalities explain how they represent information. *Functional properties* characterise which types of information a certain modality is good or bad, suitable or unsuitable at representing and sometimes also specify under which conditions this is the case. Arbitrary acoustics, for instance, may serve useful alert and alarm functions in low-acoustic environments but not in high-acoustic ditto. Successful information mapping must be informed by knowledge of functional properties. Whereas many functional properties can be analytically derived from the declarative properties of modalities, capture of other sets of functional properties requires an empirical, corpus-based approach in which different modality samples are analysed to identify their functional characteristics, often in conjunction with scenarios of use.

Based on the considerations just outlined, the original taxonomy workbench has now been completely redesigned. Its main objective still is to provide a computer-aided platform for analysing different modalities drawn from its large database in order to identify functional properties of unimodal modalities. The comprehensiveness of the functional properties identified will largely determine the applicability to interface design of modality theory.

The most important differences between Versions 1 and 2 of the system are the following:

1) *change of scope* from covering a variety of taxonomy ideas to concentrating on our current taxonomy of output modalities;
2) *change of focus* from a declarative taxonomy to a combined declarative/functional taxonomy (cf. above);
3) *upgraded functionality.* Version 2 incorporates revised versions of the analysis and classification facilities of Version 1 as well as a much improved search facility;
4) *two systems instead of one.* The system now has two distinct parts, namely a taxonomy workbench and a taxonomy demonstrator. Initially, the workbench and the demonstrator were both implemented on the OMNIS 7 platform (Bernsen, Lu and May 1994). The fact that frequent modification of hypermedia documents is a rather laborious endeavour in OMNIS 7, made us decide to port the taxonomy demonstrator into MOSAIC. This is currently being done.

This paper presents the taxonomy demonstrator. The focus is on demonstrating the scope and depth of a central part of modality theory, i.e. the theory of output modalities. Due to the limited space available, only a synoptic view of the theory can be provided. The usefulness of the demonstrator in furthering theory development will also become apparent, we hope. An outline of the taxonomy demonstrator and its underlying theory of output modalities is presented in Sect. 2. Sects. 3, 4, 5 and 6 focus on various dimensions of modality theory. Sect. 7 concludes and discusses future work.

2. Outline of the Taxonomy Demonstrator

The taxonomy of unimodal output modalities has been generated (Bernsen 1994b,c) from a set of (declarative) basic properties (see Fig. 1). *Unimodal modalities* are representational modalities which, when combined together, constitute multimodal representations but which are not themselves multimodal. Unimodal modalities at the *super level* are defined by being either analogue or non-analogue, arbitrary or non-arbitrary, and either linguistic or non-linguistic. At the *generic level*, unimodal modalities are characterised, in addition, by being either static or dynamic, as well as being physically realised in one of the three media of graphics, acoustics and haptics. Additional basic properties are needed to distinguish between modalities at the *atomic level.* For instance, the basic properties of text, discourse, label/keyword, notation, gesture, writing and speech are used to distinguish between different atomic linguistic modalities. In one part of the taxonomy, i.e. analogue graphics, a *sub-atomic level* has been added at which even more fine-grained distinctions are needed for the taxonomy to properly serve its purpose. In this way, individual unimodal modalities are defined through their declarative *profile* as constituted by a set of basic properties.

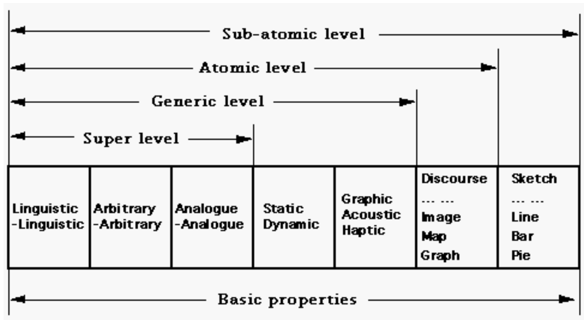


Figure 1. The underlying principles of the taxonomy of unimodal output modalities.

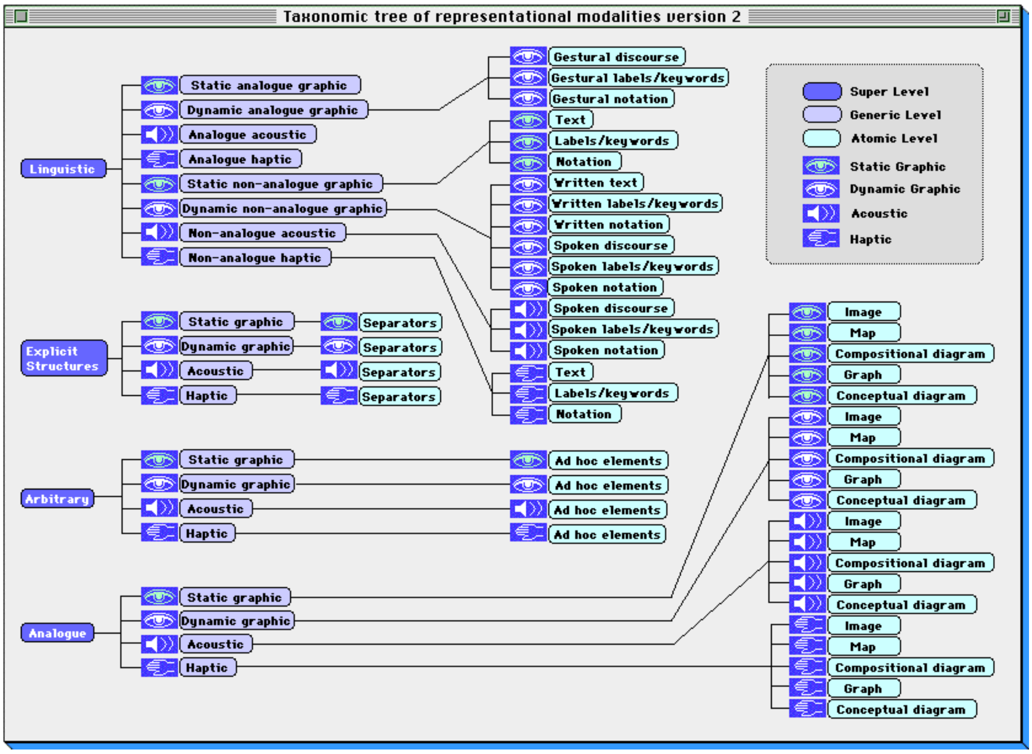


Figure 2. The taxonomy tree (OMNIS 7 implementation).

The hierarchical structure of the taxonomy is presented in Fig. 2 which shows the main screen of the taxonomy demonstrator. The taxonomy tree has 70 nodes, i.e. 4 at the super level, 20 at the generic level and 46 at the atomic level. In the tree structure, colour, being one of the information channels of graphics, is used to carry differential information. The super, generic and atomic levels are differentiated by their background colours, i.e. blue, blue/grey and light green, respectively. Different media are marked by different analogue icons, i.e. graphics by an eye, acoustics by a loudspeaker and haptics by a hand. Static and dynamic graphics are differentiated through the foreground colours of their icons, i.e. green and white, respectively. Property inheritance links are shown as lines connecting different unimodal modalities from left to right. Via these links, properties are inherited from the super level down to the atomic and sub-atomic levels, the latter of which is not shown in Fig. 2. The layout of the tree is mainly determined by spatial constraints. At the top right-hand corner, an explicit structure in a darker shade of grey contains the legend of the taxonomy tree.

The theory demonstrator consists of a series of hypertext/hypermedia documents which are of two categories, modality documents and lexicon documents. Accessed through the taxonomy tree and having a common document structure, *modality documents* define, explain and illustrate the unimodal modalities. Lexicon documents, accessed through modality documents and having free-style document structure, define, explain and illustrate additional key concepts of the theory of output modalities. The taxonomy tree provides the main structuring principle for these documents. The taxonomy tree is a graphic representation of the structure of the taxonomy of unimodal output modalities in a directly manipulatable form. Mouse-clicking on any node of the tree provides access to the relevant modality document.

The set of nodes in the taxonomy tree provides a pragmatically simplified version of the taxonomy of unimodal output modalities. For instance, the tree structure does not represent the theoretically valid distinctions between *static and dynamic acoustics* and between *static and dynamic haptics.* Instead, these distinctions are being applied internally in the documents on acoustics and haptics. Similarly, as a matter of theoretical principle some nodes in the tree do in fact have daughter nodes although these are not shown. For instance, the generic level modality *static analogue graphic language,* i.e. static graphic language using analogue signs, does have a set of atomic-level daughter nodes for representing hieroglyphic (or iconographic) writing in the modality types *text, labels/-keywords* and *notation.* However, this information has been incorporated into the presentation of *static non-analogue graphic language,* i.e. static graphic language using non-analogue signs such as those which the reader is currently reading. The reason for these purely pragmatic reductions which have been made without loosing any important information, are (i) to reduce the number and nature of unimodal atoms to those which are expected to be important to interface design; and (ii) to avoid proliferation of - sometimes even useful - atoms in the acoustic and haptic media. More atoms can always be added within the theoretical boundaries of modality theory, when substantial information on them becomes important to HCI. In the absence of the pragmatic reductions just described, the number of nodes in the taxonomy tree (super, generic and atomic levels) would have been increased by 30 from 70 to 100 (Bernsen 1994c).

We now take a closer look at the different levels of the taxonomy, the media of graphics, acoustics and haptics, the distinction between static and dynamic modalities, and the modality and lexicon documents.

3. Levels of Representation

The levels of the taxonomy correspond to considering the world of representational modalities at different levels of abstraction. From the super level down to the atomic and sub-atomic levels, distinction is made between an increasing number of modalities as more and more basic properties are being introduced. The levels allow the study of unimodal modalities to be pursued in an orderly and step-by-step manner. They make it possible, for instance, to analyse a limited number of basic properties at a time. Starting from the super level, declarative as well as functional characteristics are passed on to the descendant modalities at the levels below. The characteristics of each basic property are analysed at the level at which it has been introduced. A core assumption is that the basic properties used to define a modality are all central to that modality's ability to represent information. For each basic property, its presence or absence in a modality has important effects on that modality's abilities to represent various kinds of information.

3.1 The Super Level

The super level represents the highest level of abstraction in the taxonomy. One step up from the super level is the root, i.e., all possible representational modalities in the media of graphics, acoustics and haptics. The modalities at the super level are defined by combinations of three sets of basic properties, such that these modalities are each either linguistic or non-linguistic, arbitrary or non-arbitrary, and either analogue or non-analogue. This yields four possible combinations of basic properties which each define a super level unimodal modality (Bernsen 1994b,c), namely, the linguistic, explicit structure, arbitrary and analogue modalities, respectively (see Fig. 2). Linguistic modalities, for instance, are defined from the following set of basic properties: linguistic, non-arbitrary and non-analogue. Linguistic modalities are non-arbitrary as they are based on an already existing system of meaning; they are non-analogue because language does not have a primarily analogue relationship to what it represents (Bernsen 1994b,d). In a simple notation, the profile of linguistic modalities is <li,-ar,-an>. The profiles for explicit structure, arbitrary and analogue modalities are expressed as <-li,-ar,-an>, <-li,ar,-an> and <-li,-ar,an>, respectively. **The super level does not have any deep theoretical significance as it simply reflects one among several possible, all incomplete, classifications of the generic modalities.**

3.2 The Generic Level

From the super level to the generic level, the taxonomy is further differentiated through distinctions between different media and states of representation. The *medium* determines how a modality is being instantiated physically while the *state of representation* specifies whether or not a modality affords freedom of perceptual inspection. The taxonomy covers the media of graphics, acoustics and haptics and distinguishes between static and dynamic states of representation. The former is elaborated in Sect. 4, the latter in Sect. 5 below. Taking into account the pragmatic reductions noted in Sect. 2 above, the taxonomy includes twenty generic level unimodal modalities as presented in Fig. 2. There are eight linguistic generic unimodal modalities, i.e., static graphic language using analogue signs, dynamic graphic language using analogue signs, acoustic language using analogue signs, haptic language using analogue signs, static non-analogue graphic language, dynamic non-analogue graphic language, non-analogue acoustic language, non-analogue haptic language. There are four generic unimodal modalities under each of the super level categories explicit structure, arbitrary and analogue modalities, respectively. The profiles of, e.g., the four analogue generic unimodal modalities are <-li,-ar,an,st,gr>, <-li,-ar,an,dyn,gr>, <-li,-ar,an,st/dyn,ac> and <-li,-ar,an,st/dyn,ha>, respectively. This shows how the first three basic properties in each profile have been inherited from the super level. The reader may have noticed the seemingly contradictory notion of 'analogue' in first four linguistic generic level modalities. These refer to languages using analogue *signs* rather than languages which are, as such, analogue. The latter would be self-contradictory according to the theory. The point is that languages are primarily syntactic-semantic systems of meaning and, as such, essentially non-analogue modalities. Only secondarily, through the use of analogue signs in some languages (such as hieroglyphs), may a language possess an analogue aspect.

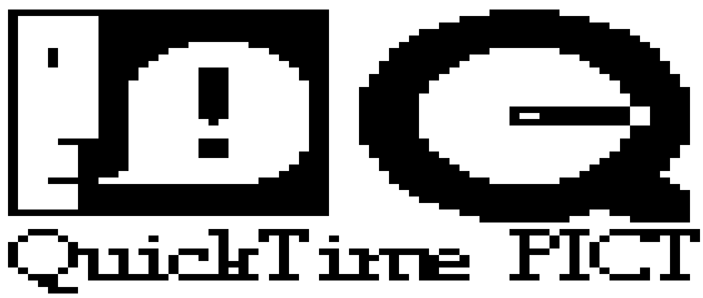
3.3 The Atomic Level

Generally speaking, the atomic level is the lowest level in the taxonomy so far. The unimodal modalities at this level are intended to be used as 'basic building blocks' by interface designers. Basic properties introduced at the atomic level, unlike those introduced at the generic and super levels, do not manifest themselves across all branches of the taxonomy. Rather they are specific to the descendants of a certain super level modality. In the linguistic family, the basic properties are: gestural, written, spoken, text, discourse, label/keyword and notation. The basic properties of image, map, compositional diagram, graph and conceptual diagram are introduced in the analogue branch of the taxonomy, whereas the basic properties separator and ad hoc element are the grandchildren of the explicit structure and arbitrary modalities, respectively. Atomic modalities inherit both declarative and functional properties from their parental super and generic level modalities. The profile of, e.g., static graphic conceptual diagrams is <-li,-ar,an,st,gr,con.dia>.

3.4 The Sub-Atomic Level

Unimodal modalities at the sub-atomic level are not visible on the main screen of the theory demonstrator but must be accessed via their parental atomic level documents. As said above, the unimodal modalities at the atomic level are normally the lowest-level modalities in the taxonomy. However, in some parts of the taxonomy, notably the analogue modalities, information representations have been so richly developed that more fine-grained distinctions are necessary for the sakes of both sophistication of the theory and its potential practical use in interface design. One such analogue atomic modality is the static graphic graph which **in our analysis** has three sub-atomic descendants: the line graph, bar graph and pie graph, respectively. **They show functionally related quantities, independent quantities and percentages of wholes, respectively.** In the corresponding modality documents, these notions have been appropriately generalised. For instance, bar graphs are in fact graphs in which any geometrical shape in 1D, 2D, or 3D can be used whose length, area, or volume represents the quantities in question. It follows that Figs. 3 (a) and (b) are as much instances of bar graphs as is Fig. 9 further below. In Fig. 3 (a), area is being used for showing size in square miles of states relative to the size of their overseas empires whereas in Fig. 3 (b) volume is used to show coal and oil-gas reserves in various regions.

(a)



(b)

Figure 3. Variety of bar graphs (Lockwood 1969).

4. Media of Representation

The medium of a representational modality is the physical substrate in which it is realised and perceived. Among all possible media, the media of graphics, acoustics and haptics are considered the most relevant for interface design purposes. The medium of a certain unimodal modality is of fundamental importance to that modality's suitability for representing information in a specific interface design context. The fact that different media have different information channels at their disposal similarly has important implications for their usability for different design purposes. Each output medium is uniquely characterised by a set of information channels (Hovy and Arens 1990). An *information channel* is a humanly perceivable aspect of a medium which may be used to carry information.

4.1 The Graphic Modalities

Throughout human civilisation, the medium of graphics or the visual medium has been a central vehicle for the representation of information. It is therefore not surprising that certain graphic modalities, such as static typed language, graphs and images, have been more extensively studied than any type of acoustic or haptic representation. Yet, a full appreciation of all graphic modalities still remains to be made. Graphics have at least the following information channels: shape, size (length, width, height), texture, resolution, contrast, value (grey scales), **colour (brightness, hue and saturation),** position, orientation, viewing perspective, spatial arrangement, short-duration repetitive change of properties, non-repetitive change of properties, movement, displacement (relative to the observer), and temporal order.

Figure 4. A multimodal representation composed of a dynamic graphic image and a dynamic acoustic image.

4.2 The Acoustic Modalities

Until recently, acoustic modalities have had a rather limited role in HCI. This contrast sharply with human-human interaction, in which acoustic modalities are of central importance. However, this disparity is gradually being removed by advances in technology. In many cases, acoustic modalities may provide an alternative or supplement to graphic modalities. For instance, acoustic images strongly augment the (virtual) realism of events, processes or situations rendered in dynamic graphic images. Fig. 4, using dynamic graphic and acoustic images, shows the demolition of a building. Furthermore, acoustic modalities generally have two primary advantages over graphic and haptic modalities: (1) they allow users to simultaneously monitor and identify sources of information in all possible directions, not just in the direction of the gaze or body surface; (2) they enable users to distinguish, monitor, and switch attention among simultaneous sources of sound or among sounds with disparate sound parameters. However, the potential of acoustic modalities is still largely waiting to be explored. Dynamic acoustic language (speech) is likely to become a powerful rival of typed graphical user interfaces, and the power of acoustic graphs for the exploration of high-dimensional data is going to be heard.

Acoustics have at least the information channels loudness, pitch, timbre, rhythm, duration, temporal order and source location. Acoustic linguistics has the following additional information channels: voice quality, stress, intonation, dialect, accent, personality.

4.3 The Haptic Modalities

Haptics are currently an impoverished medium in interface design, partly because of technological limitations and partly because, under normal circumstances, haptic representations are inferior to graphic representations in terms of information density and speed of acquisition. This situation is likely to change as the technology for haptic representation improves. Haptic representations comprise much more than representations for the visually handicapped and have important potential roles in representing information to normal users. They increase the (virtual) realism of representations and may convey information through bypassing overloaded visual and acoustic sensors.

Haptics have at least the following information channels: shape, size (length, width, height), texture (surface quality), position, orientation, temperature, pressure, voltage, spatial arrangement, movement, repetitive change of properties, non-repetitive change of properties, and temporal order.

Figure 5. Scanned image of a unimodal haptic compositional diagram with bimodal legend **using texture and haptic keywords, respectively. The diagram shows the geological composition of the Earth. Differently coloured surfaces have different texture.**

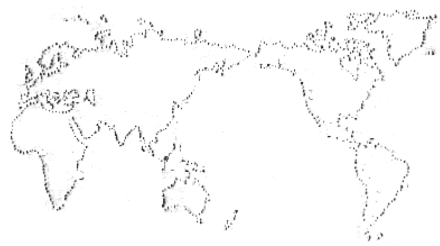


Figure 6. An unimodal haptic map of the world.

The taxonomy demonstrator includes scanned images of real haptic representations rather than the haptic representations themselves. Two haptic representations are presented here. Fig. 5 is taken from the modality document on haptic compositional diagrams and shows the internal structure of the earth. Fig. 6 is from the haptic map modality document.

5. States of Representation

Any unimodal modality is either static or dynamic. The distinction between static and dynamic modalities is, however, defined in terms of whether or not *freedom of perceptual inspection* is possible of the information represented by the modality rather than in terms of, say, the absence or presence of perceptible physical change. Freedom of perceptual inspection means that the user is allowed time to inspect, in random order and as long as desired, the information presented. Freedom of perceptual inspection is compatible with some amount of perceptible change as long as the change is repetitive **and the cycle of repetitions is of relatively short duration** as in, e.g., acoustic alarm signals or blinking graphics (cf. Fig. 7). As remarked in Sect. 2

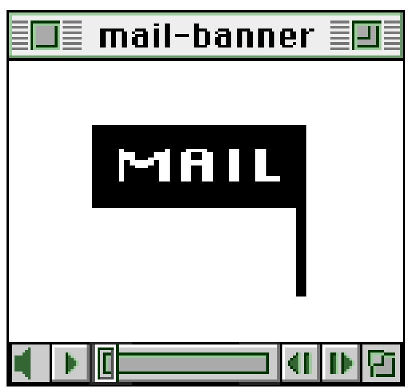


Figure 7. The blinking mail icon in this window is a static multimodal representation consisting of a static graphic image and a static graphic typed label (keyword).

above, the taxonomy tree does not incorporate distinctions between static and dynamic modalities in the haptic and acoustic media. These distinctions are made in the underlying modality documents, however.

6. Document Structure

The theory demonstrator currently includes 134 hypermedia documents comprising some 150 Kb of (non-illustrated) text. Structured in terms of the taxonomy tree (Fig. 2), these documents are of two types, modality documents and lexicon documents.

6.1 Modality Documents

Modality documents define, explain, analyse and illustrate the unimodal modalities from the point of view of interface design support. These documents share the same document structure which includes the following entries:

- Profile
- Inherited declarative and functional properties
- Specific declarative and functional properties
- Information mapping rules
- Combinatorial analysis
- Relevant operations
- Identified types-of

Each modality document is illustrated by some 5-10 illustrations selected such as to show both prototypical examples, important non-prototypical and marginal cases, interesting multimodal combinations, etc. What follows is a walkthrough of the modality document structure exemplified by illustrations from various modality documents.

1) *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 in kind from other modalities at the same level. This was illustrated in Sect. 3.1 above.

2) *Inherited declarative and functional properties.* These are the properties, basic or otherwise, which the modality inherits from higher levels of the taxonomy. Except for the super level modalities, all modalities inherit an important part of their properties from higher levels. Thus, generic level modalities inherit the categorical and functional properties of their parent node at the super level, atomic modalities inherit the properties of their parent nodes at the super and generic levels, etc. To keep individual modality documents short, these properties must be retrieved through the hypertext links. The following example shows the list of links to inherited properties in the gestural notation modality document (hypertext links are underlined):

- linguistic modalities
- dynamic modalities
- graphic modalities
- notation

Dynamic graphics have the following information channels: (a) those of static graphics: shape, size (length, width, height), texture, resolution, contrast, value (grey scales), colour, brightness, hue, saturation, position, orientation, viewing perspective, spatial arrangement, short-duration repetitive change of properties; (b) in addition to those of static graphics: non-repetitive change of properties, movement, displacement (relative to the observer), and temporal order.

The dimensionality of dynamic graphics is: 1-D, 2-D and 3-D spatial, time.

Gestural notation thus inherits the properties of the linguistic, 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 policy (Sect. 2), the gestural notation document presents both static and dynamic gestural notation. Fig. 8 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 may share 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 and aspects of 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, requires to be introduced in some non-arbitrary modality, such as some linguistic modality or other. This is demonstrated in Fig. 9 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 language labels/keywords in the graph legend. Without this linguistic annotation, nobody would be able to interpret the graph. The graph compares waste recycling of aluminium, glass and paper in the years 1970 and 1991 in the USA.

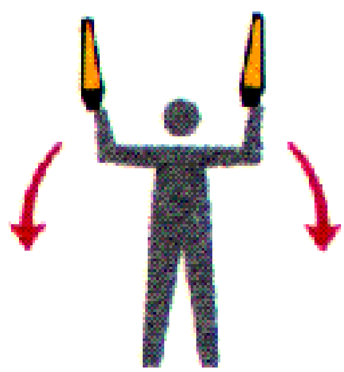


Figure 8. Static gestural notation: a marshalling signal which means 'move ahead' (Tufte 1990).

4) *Information mapping rules* are similar in many respects to production rules. They express aspects of information which a particular unimodal modality is good at, or unsuited for, representing and sometimes under which conditions this is the case. Information mapping rules are crucial to, and their use for interface design support is being investigated as part of, the development of the information mapping methodology (cf. Sect. 1 above).

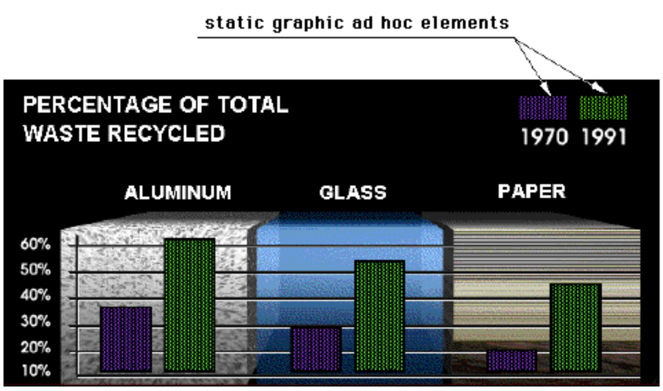


Figure 9. Dependence on linguistic modalities of an information channel used ad hoc.

One of the information mapping rules in the static graphic image document is:

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

This rule effectively states that static graphic images are good tools for identifying, e.g., persons sought by the police, and that identification is further supported by high specificity (a large amount of detail in as many information channels as possible), high dimensionality (2 1/2D or 3D better that 2D), and high image resolution.

The rule is read from left to right as an if-then rule. From right to left, the rule says that "Modality X is good at representing Y". An illustration of this rule, and hence of one of the advantages of the static graphic image modality, is the use of photographs in criminal investigation. It is virtually impossible to linguistically describe what a person looks like in such a way that the person may be uniquely identified from the linguistic description (Bernsen 1994d). Use of static graphic images, such as the one shown in Fig. 10, makes this an effortless undertaking. Indeed, a picture can sometimes be worth more than a thousand words. Or, rather, this proverbial classic not only applies to pictures but to analogue representations in general, irrespective of whether they are embodied in graphics, acoustics or haptics.

Figure 10. A unimodal static graphic image of high specificity.

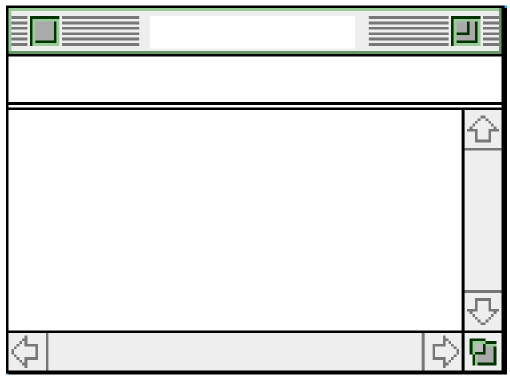


Figure 11. Nested unimodal explicit static graphic structures: the Macintosh window.

5) *Combinatorial analysis.* This form of analysis addresses compatibilities and incompatibilities between the modality presented in a particular modality document and other unimodal modalities. For instance, in the modality document on explicit static graphic structures, it is stated under 'combinatorial analysis' that "explicit static graphic structures combine well with any static or dynamic graphic modalities, whether linguistic, analogue or arbitrary". This may be illustrated by Figs. 7 and 11. In Fig. 11, a Macintosh window is represented as a layered series of unimodal explicit static graphic structures. In Fig. 7, these unimodal explicit static graphic structures form part of the multimodal representation. Speaking more generally, 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 'syntax'.

6) *Relevant operations.* These are operations which may be applied to the current unimodal modality. 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 is always to bring out more clearly particular aspects of the information to be presented. Dimensionality reduction, as in reducing common road maps from 3-D to 2-D without any loss of relevant 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**, *italicizing* and underlining are common operations in graphic typed languages (see Bernsen 1994d).

7) *Identified types-of.* These are the specific types of a unimodal modality, which are found one level down in the taxonomy hierarchy. For instance, dynamic non-analogue graphic language (generic level) has six atomic types:

- Dynamic written text
- Dynamic written labels/keywords
- Dynamic written notation
- Graphic spoken discourse
- Graphic spoken labels/keywords
- Graphic spoken notation

The three dynamic graphic spoken language modalities are graphic representations of someone speaking and may be used for lip reading and acoustic language disambiguation.

6.2 Lexicon Documents

Lexicon documents define, explain and illustrate the key concepts of modality theory. There are currently 68 such documents or concepts. To mention but a few, dimensionality, information channel, interpretational scope, modality structures (icons, lists, tables, etc.), saliency, metaphor, and specificity are all lexicon document entries. Due to the heterogeneous nature of their topics, no rigid document structure has been enforced on lexicon documents. Most lexicon documents include a definition and a number of illustrations but are otherwise tailored to their specific contents. Lexicon documents are not directly accessible from the taxonomy tree (or main screen), but are reached through hypertext links from modality documents and other lexicon documents.

7. Discussion and Future Work

As remarked in the introduction, it remains an open question to what extent the current version of the workbench will need to be further re-designed in order to function as a design support tool. A key question concerns automation. The information mapping methodology assumes that practical information mapping is done in two broad iterative phases (Bernsen 1994a). In the first phase, information is collected and succinctly represented concerning the information to be represented and exchanged between user and system during task performance on the artifact to be designed. In the second phase, this information is ‘put through’ a design tool based on modality theory, which will map the collected domain and task information onto a set of input/output modalities which could optimise the interface to the artifact. The question is whether the workbench might be developed into such a tool. One possibility might be to fully automate the workbench by developing its current set of information mapping rules into the rule set of a knowledge-based system which could support interface design at any level of detail. However, a recent case study of a realistic design process (Verjans and Bernsen 1994) strongly indicates that this is not feasible. The real world of IT artifacts and their various work domains, tasks to be supported, user types, etc. is quite simply too complex and unmanageable to make such an endeavour a realistic one. At the opposite extreme, the workbench might not be automated at all but would make its information easily accessible to interface designers who would use their 'natural intelligence' to let the information constrain their design decisions. Furthermore, the workbench information should be developed down to a certain level of detail only, leaving the lower levels of interface design detail to designer craft skill, guidelines, standards, etc. The latest case study (Bernsen and Verjans 1995) suggests the existence of a natural division of labour between rule-based information mapping on the one hand, and the subsequent design of lower-level interface details on the other. In addition, this study investigates a first extension of modality theory from being solely a theory of output representational modalities to being a theory of the interaction between input and output modalities. **Our most recent work demonstrates the feasibility of building a theory of input modalities along the principles of the theory of output modalities presented in this paper. The dynamics between input and output is then viewed as feedback relationships.**

References

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

Bernsen, N.O. (1993b): Matching information and interface modalities. An example study. *Esprit Basic Research project GRACE Deliverable* 2.1.1.

Bernsen, N.O. (1994a): Modality theory in support of multimodal interface design. In *Proceedings of the AAAI Spring Symposium on Intelligent Multi-Media Multi-Modal Systems,* Stanford, March, 37-44.

Bernsen, N.O., (1994b): Foundations of multimodal representations. A taxonomy of representational modalities. *Interacting with Computers* Vol. 6, 4, 347-71.

Bernsen, N.O. (1994c): A revised generation of the taxonomy of output modalities. *Esprit Basic Research project AMODEUS-2 Working Paper* RP5-TM-WP11, 1994.

Bernsen, N.O., (1994d): 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, June, 165-179. Forthcoming in the *Focus on Computer Graphics* Series, Springer Verlag 1995.

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

Bernsen, N.O., Lu, S. and May, M. (1994): Towards a design support tool for multimodal interface design. The taxonomy workbench and theory demonstrator. *Esprit Basic Research project AMODEUS-2 Working Paper* RP5-TM-WP5.

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

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

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

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

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

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

May, M. (1993d): Taxonomy and levels of mappings in the construction of multimodal interfaces. *Workshop Notes from AI-ED '93:* World Conference on Artificial Intelligence in Education, Edinburgh, August. *Workshop on Graphical Representations, Reasoning and Communication*..

May, M. and Bernsen, N.O. (1993): The taxonomy workbench. *Esprit Basic Research project GRACE Deliverable* 2.1.8.

May, M. and Tobin, R. (1993): The taxonomy workbench. Purpose, functionality and layout. *Esprit Basic Research project GRACE Deliverable* 2.1.9.

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

Verjans, S. (1994): EuroCODE: Preliminary information mapping analysis of an AMODEUS common exemplar. *Esprit Basic Research project AMODEUS-2 Internal Report* RP5-TM-IR1.

Verjans, S. and Bernsen, N.O. (1994): PaTerm: A case study in information mapping. *Esprit Basic Research project AMODEUS-2 Working Paper* RP5-TM-WP6.

Acknowledgements. The work reported in this paper was done on grants from the Esprit Basic Research project AMODEUS-2 and the Danish Research Council for the Natural Sciences whose support is gratefully acknowledged.