FOUNDATIONS OF MULTIMODAL REPRESENTATIONS

A Taxonomy of Pure Modalities

Niels Ole Bernsen, CCI, Risø National Laboratory and Roskilde University

Summary: Advances in information technologies are producing a very large number of possible interface modality combinations which are potentially useful for the expression of information in human-computer interaction. However, a principled basis for analysing arbitrary modality types and combinations as to their capabilities of information representation is still lacking. The paper presents an approach to the analysis of modality types and their combinations and takes some steps towards its implementation, departing from a taxonomy of pure generic modalities of representation. A small number of key properties appear sufficient for creating a taxonomy of generic modalities which is both relatively simple, robust, intuitively plausible, and reasonably exclusive and exhaustive. These properties are: analogue and non-analogue representations; arbitrary and non-arbitrary representations; static and dynamic representations; linguistic and non-linguistic representations; different media of representation; and modality structure.

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

1. Introduction

This paper presents a principled approach to the analysis of unimodal and multimodal representations for usability engineering purposes. The work forms part of the ESPRIT Basic Research project GRACE which ultimately aims at providing a sound theoretical basis for usability engineering in the domain of multimodal representations. Whereas the enabling technologies for multimodal (including virtual reality) representation are growing rapidly, there is a lack of theoretical understanding of the principles which should be observed in mapping information from some task domain into presentations at the human-computer interface in a way which optimises the usability and naturalness of the interface, given the specific purposes of the artifact. To achieve at least part of this understanding, it appears, the following objectives should be pursued, listed in order of increasing complexity:

- to establish sound foundations for creating a taxonomy and a set of related categorisations of the generic modalities which go into the creation of multimodal output representations for human-computer interaction (HCI);
- to establish sound foundations for describing and analysing any particular type of unimodal or multimodal output representation relevant to HCI;

- to use the above taxonomy and related descriptive apparatus in creating a conceptual framework for describing interactive computer interfaces;
- to apply the results of the steps above to the analysis of the problems of information-mapping and information-transformation between work/task domains and human-computer interfaces in information systems design;
- to use, if possible, results of the work described above in building design tools for the support of usability engineering.

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

This paper addresses the first and second objectives above in proposing a taxonomy of generic, pure representational modalities (Sects. 2 and 3). The taxonomy builds on a small number of key properties which all express important dimensions of modality description and are briefly analysed in Sects. 4 and 5. The analysis leads to a characterisation of pure generic modalities (Sect. 6) and a discussion of the exclusiveness and exhaustiveness of the taxonomy (Sect. 7). Sect. 8 discusses the issue of modality structure. The distinction between external and internal representations is crucial to the analysis of particular types of unimodal or multimodal representation and is discussed in Sect. 9. The concluding discussion (Sect. 10) argues that the taxonomy enables the specification of a principled approach to the analysis of any particular modality type or combination of modality types falling within the scope of this paper.

A key consideration in any attempt to address the issues dealt with in this paper is the following. There are literally thousands of possible and potentially useful combinations of interface representation modality types. Only approaches which manage to identify a limited set of fundamental properties relevant to the analysis of arbitrary representational modality types would seem to stand a chance of handling on a principled basis such a large number of different modality types and their combinations.

2. On Representation

The term *multimodal representation* designates combinations of two or more pure or unimodal representational modalities for HCI purposes. Such representations are *external* to the human cognitive system. We are not here dealing with internal cognitive representations (see Sect. 9 below, however).

External representations are considered as representations by the human cognitive system and are, as far as we are concerned, produced by data structures in computers and other items of information technology. The concept of an external representation implies a distinction between *what is represented* and its *representation*. In general, there is a one-to-many mapping relationship between what is represented and its possible representations. One and the same object, situation, event, process, set of data, procedural instruction, etc., can be represented in many different ways. Some of these representations may be better than others for given task goals, given also the information to be represented and the nature of human cognition. Moreover, in many cases representations do not allow a simple and universal decoding of what they represent but require additional knowledge for this to be possible with any degree of certainty or confidence. In general, there is a one-to-many mapping relationship between a representation and what it may represent. The additional knowledge which is needed for non-arbitrary interpretation is knowledge of the *mapping principles* between representation and what is represented. Foreign spoken languages, unknown to us, are examples in point but so are many examples of graphical and other representations. The relationship between representations and what is to be represented is shown in Diagramme 1. Diagramme 2 in Sect. 9 below provides a less simplified representation.

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

Diagramme 1. Representation requires mapping principles. Diagramme 1 is multimodal and is composed of written natural language and graphical picture icons.

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

3. A Taxonomy of Pure (or Uncombined) Representational Modalities

The question to be addressed in this section is the following: what are the generic types of representational modalities in their pure or uncombined forms? For HCI purposes, combined representational forms are often more interesting because of the increase in expressive power that comes from combining different modalities. However, combined representational forms are combined from something, namely pure representational forms. I shall argue that if we want to adopt a principled approach to the analysis of modality combinations, we have to start by analysing the pure forms of modalities.

Similarly, we should start by analysing pure *generic* modalities before considering the huge number of actual or possible representational *types* and their combinations. The modalities to be described shortly are (pure) generic modalities because each of them have a number of different (and equally pure) modality types subsumed under them.
Such different types of one and the same generic modality have different properties and different capabilities of representing information which will have to be accounted for eventually. However, doing so at this stage would run the risk of confusing the basic issues involved. The pure generic modalities are presented in Table 1.

1. Spoken language (natural or otherwise) including single words and letters (spoken language icons).
Well-Known Types: Spoken letters, words, numerals, other spoken language related sounds, text, lists.
Characteristics: Non-analogue, non-arbitrary, dynamic.
Medium: Sound qualities/audition.

2. Written language (natural or otherwise) including single words and letters (written language icons).
Well-Known Types: Written letters, words, numerals, other written language related signs, text, lists, tables, musical notation.
Characteristics: Non-analogue, non-arbitrary, static (normally).
Medium: Visual and graphical qualities/vision.

3. Real-world sound representations including single sounds (sound icons).
Well-Known Types: Single sounds, sound sequences, music?
Characteristics: Analogue, non-arbitrary, dynamic.
Medium: Sound qualities/audition.

4. Arbitrary sound representations including single sounds (sound icons).
Well-Known Types: Single sounds, sound sequences.
Characteristics: Non-analogue, arbitrary, dynamic.
Medium: Sound qualities/audition.

5. Diagrammatic pictures including graphical picture icons (2D and 3D).
Well-Known Types: Diagrams, pure maps, sequences of such, lists, tables.
Characteristics: Analogue, non-arbitrary, static, [prototypical category].
Medium: Visual and graphical qualities/vision.

6. Non-diagrammatic real-world representations or pictures including non-diagrammatic picture icons.
Well-Known Types: Photographs, naturalistic drawings, sequences of such, lists, tables.
Characteristics: Analogue, non-arbitrary, static, [prototypical category].
Medium: Visual and graphical qualities/vision.

7. Arbitrary diagrammatic forms and sequences of such (points, 1D, 2D and 3D geometrical forms).
Well-Known Types: Points, lines, boxes, circles, volumes, etc., sequences of such, lists, tables.
Characteristics: Non-analogue, arbitrary, static.
Medium: Visual and graphical qualities/vision.

8. Animated diagrammatic pictures including animated icons (1D, 2D and 3D).
Well-Known Types: Animations, sequences of such.

Characteristics: Analogue, non-arbitrary, dynamic [prototypical category].
Medium: Visual and graphical qualities/vision.

9. Dynamic real-world representations including dynamic picture icons.
Well-Known Types: Films, videos, sequences of such.
Characteristics: Analogue, non-arbitrary, dynamic, [prototypical category].
Medium: Visual and graphical qualities/vision.

10. Animated arbitrary diagrammatic forms including animated icons.
Well-Known Types: Points, lines, boxes, circles, volumes, etc., sequences of such.
Characteristics: Non-analogue, arbitrary, dynamic.
Medium: Visual and graphical qualities/vision.

11. Graphs including graph icons.
Well-Known Types: A graph space containing 1D, 2D or 3D geometrical forms.
Characteristics: Non-analogue, non-arbitrary, static or dynamic [patterns in data].
Medium: Visual and graphical qualities/vision.

12. Real-world touch representations including touch icons.
Well-Known Types: Single touch representations, touch sequences.
Characteristics: Analogue, non-arbitrary, dynamic.
Medium: Tactile and kinaesthetic qualities/touch.

13. Arbitrary touch representations including touch icons.
Well-Known Types: Touch signals of differents sorts.
Characteristics: Non-analogue, arbitrary, dynamic.
Medium: Tactile and kinaesthetic qualities/touch.

14. Touch language (natural or otherwise) including single words and letters (touch language icons).
Well-Known Types: Touch letters, words, numerals, other touch language related signs, text, lists, tables.
Characteristics: Non-analogue, non-arbitrary, dynamic.
Medium: Tactile and kinaesthetic qualities/touch.

Table 1. The pure generic modalities. Square brackets indicate properties discussed in the text but not included in Table 2 below which provides a structured view of the taxonomy.

Table 1 obviously raises a large number of questions. It should be said from the outset that this table might just as well have included a slightly smaller number of cells as well as a somewhat larger number of cells without any change to the principles on which it is based. Fewer cells (in fact, 12) might have resulted from merging cells 5 and 6 and cells 8 and 9. More (in fact, 18) cells might have resulted from splitting cell 2 into a static and a dynamic cell, creating a 'diagrammatic' version of cell 3; creating both a static and a dynamic version of cell 11; and creating a 'diagrammatic' version of cell 12. While this can be done easily, there are a number of rather different reasons, none of which seem very deep, why it has not been done in creating Table 1. Briefly, they are as follows: 'Mere' prototypical differences are also important (cells 5 and 6 and cells 8 and 9). Dynamical written language is not that common or important (cell 2). The non-existence of prototypical differences is also important (no 'diagrammatic' version of cell 3, no 'diagrammatic' version of cell 12). And graphs, even though it is quite important that they can be static as well as dynamic, basically have to be multimodal to be useful. It is a matter of personal preference if, at this stage, one wants a completely principled table of pure modalities, in some sense, or if one prefers one tempered by considerations such as the above. Sound 'diagrammes', for instance, are actually being used in HCI and might have an important future ahead of them. The question of music and some other difficult cases of modality types will be briefly discussed later. Table 4 below presents the complete set of pure generic modalities discussed in this paper, i.e., the 18 modalities just indicated plus 3 additional ones which are all forms of analogue natural language.

4. Icons

It appears from the taxonomy that *icons* can be created from any generic modality. The term 'icon', therefore, does not, strictly speaking, designate a modality. Rather, icons are defined by their singularity and their representational function. An icon is chosen to symbolise something in a particular context. And given their singularity, icons are almost always semantically ambiguous as to what they symbolise. Context may significantly help in disambiguating an icon but its ambiguous character is independent of whether the icon is analogue or not, non-arbitrary or not, static or dynamic, linguistic or non-linguistic, and is also medium-independent (see below). Its ambiguity is primarily due to its singularity. If icons are neither generic modalities nor constitute a type subsumed under one particular generic modality, what are they? It is proposed to view icons as a particular type of *modality structure*. I shall return to this topic in Sect. 8 below.

5. The Properties of Analogueness, Arbitrariness, Static/Dynamic and Media, and Common Sense

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

5.1 Analogue and non-analogue representational modalities

The distinction between analogue and non-analogue (external) representations is quite important as well as being intuitively obvious in most cases. It designates the difference between representations, in whatever modality, which represent through recognisable topological similarity with what they represent and representations which represent through conventional pairing between representation and what is represented. As long as we focus only on external representations (including touch) and do not consider the nature of internal cognitive representations, this distinction is clear in most cases. In practice, however, the distinction sometimes can be difficult to draw primarily because of the existence of *levels of abstraction* in analogue representation, whether the representation be a sound, a piece of graphics such as an ordinary diagramme or a tactile/kinaesthetic one. A highly abstract representation may have so few recognisable similarities with what it represents that it may just as well, arguably, be considered a non-analogue representation. The less recognisable similarity there is between what is represented and its representation, the more we may have to rely on additional knowledge of the *representational conventions* used in order to decode particular representations. In the limit, where we find, e.g., natural language and arbitrarily chosen icons, we have to rely exclusively on representational conventions.

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

Categorising our 14 pure modalities according to the analogue/non-analogue distinction generates a classification of external representations which is orthogonal to the taxonomy of pure generic modalities:

Analogue external representations
3. Real-world sound representations.
5. Diagrammatic pictures.
6. Non-diagrammatic real-world representations or pictures.
8. Animated diagrammatic pictures.
9. Dynamic real-world representations.
12. Real-world touch representations.

Non-analogue external representations
1. Spoken language.
2. Written language.
4. Arbitrary sound representations.
7. Arbitrary diagrammatic forms.

10. Animated arbitrary diagrammatic forms.
11. Graphs.
13. Arbitrary touch representations.
14. Touch language.

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

5.2 Arbitrary and non-arbitrary representational modalities

The distinction between non-arbitrary and arbitrary pure generic modalities marks the difference between external representations which, in order to perform their representational function, rely on an already existing system of meaning and representations which do not. The reason why this distinction tends to be overlooked is that, in most cases, it coincides with the distinction between analogue and non-analogue representations. There are four exceptions, however:

Non-arbitrary, non-analogue representations
1. Spoken language.
2. Written language.
11. Graphs.
14. Touch language.

That spoken, written and touch language constitute exceptions in the above sense is straightforward. That the graph category constitutes an exception is a consequence of the quite different fact that graphs are based on organised data. These data constitute the already existing system of meaning from which graphs are constructed using conventional mapping principles. From another point of view, the existence of these exceptions means that as many as 10 out of the 14 pure generic modalities exploit already existing systems of meaning. The only pure generic modalities which do not do so are the expressedly arbitrary graphical diagrammatic forms, sound representations and touch representations. It seems obvious that, *ceteris paribus*, exploiting already existing systems of meaning is an advantage in usability engineering as in the external representation of information in general. Unfortunately, this can be done in many different ways for a given design or other representational purpose, not all of which are appropriate.

The separation performed between the analogue/non-analogue distinction, on the one hand, and the arbitrary/non-arbitrary distinction, on the other, does seem quite important. It shows why, e.g., natural language can compete successfully with analogue graphics for many interface representational purposes. Despite being non-analogue considered as a form of external representation, natural language does build on an already existing system of meaning. And the separation between the analogue/non-analogue and arbitrary/non-arbitrary distinctions demonstrates that explanations of why, e.g., natural language representational modalities are in some cases inferior, and in others superior, to analogue graphical modalities cannot simply be provided through appeal to the analogue/non-analogue distinction. One has to look deeper than that (see below). Furthermore, the distinction between arbitrary and non-arbitrary representational modalities is the one to consider when analysing the basic differences between representations which are, and representations which are not, based on already existing systems of meaning.

Music is a difficult case. It does seem to be based, in some sense, on an already existing system of meaning but might, arguably, belong to a pure modality category of its own. One possibility is to re-categorise music as being non-analogue, non-arbitrary and dynamic just like spoken language, and then stress the difference between musical 'meaning' and linguistic 'meaning'.

5.3 Static and dynamic representational modalities

This is another important distinction because the differences between static and dynamic external representations have profound implications for their usability in specific task domain contexts. What is dynamic changes through time. This distinction marks, i.a., the obvious differences between:

Static pure generic modalities
5. Diagrammatic pictures.
6. Non-diagrammatic real-world representations or pictures.
7. Arbitrary diagrammatic forms.

and
Dynamic pure generic modalities
8. Animated diagrammatic pictures.
9. Dynamic real-world representations.
10. Animated arbitrary diagrammatic forms.

Written language is sometimes presented dynamically and graphs can be static as well as dynamic (see Table 4 below). The sound medium is inherently dynamic. The medium of touch appears to be inherently dynamic because of its close relationship with kinaesthesis. However, it is quite possible that finer distinctions will ultimately have to be made in this latter domain.

5.4 Representational modalities in different media

A fourth classification which is orthogonal to the taxonomy of pure modalities is the one between different media of representation. The 14 generic modalities identified above are related to three different *media,* i.e., sets of perceptual qualities and the corresponding sensory equipment needed for perceiving them, namely:

Visual and graphical qualities/vision
2. Written language.
5. Diagrammatic pictures.
6. Non-diagrammatic real-world representations or pictures.
7. Arbitrary diagrammatic forms.
8. Animated diagrammatic pictures.
9. Dynamic real-world representations.
10. Animated arbitrary diagrammatic forms.
11. Graphs.

Sound qualities/audition
1. Spoken language.
3. Real-world sound representations.
4. Arbitrary sound representations.

Tactile and kinaesthetic qualities/touch
12. Real-world touch representations.
13. Arbitrary touch representations.
14. Touch language.

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

If, in other words, we choose a given (pure) modality for the representation of information, this modality inherits a specific medium of expression whose different generic modalities of representation share a number of perceptual qualities which can be manipulated for representational purposes. This makes it possible to use the concept of *information channels* for the analysis of types and instances of representational modalities and modality combinations. A channel of information is a perceptual aspect of some medium which can be used to carry information. If, for instance, differently numbered but otherwise identical iconic ships are being used to express positions of ships on a screen map, then different colouring of the ships can be used to express additional information about them. Colour, therefore, is an example of an information channel (Hovy and Arens 1990, Bernsen 1992).

Evidently, there are other media of expression than the three media considered in this paper and the taxonomy might eventually have to be expanded to include them. So far, (output) media of expression such as machine gesture, smell and taste are outside the scope of the taxonomy.

5.5 A common sense classification

A categorisation which is close to common sense is the following. The 14 modalities can be divided into the categories:

a. Language (natural or otherwise)
1. Spoken language.
2. Written language.
14. Touch language.

b. Pictures of something
5. Diagrammatic pictures.
6. Non-diagrammatic real-world representations or pictures.
8. Animated diagrammatic pictures.
9. Dynamic real-world representations.

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

4. Arbitrary sound representations.
7. Arbitrary diagrammatic forms.
10. Animated arbitrary diagrammatic forms.
13. Arbitrary touch representations.

d. Non-visual 'pictures' or analogue representations
3. Real-world sound representations.
12. Real-world touch representations.

e. Graphs
11. Graphs.

The categories (a)-(c) are familiar in their own right. Category (d) corresponds to category (b), only covering different media. Language does not seem to have a label for category (d) but makes it tempting to use the term 'picture' analogously in characterising (d). Common sense may not have a position on graphs. Graphs have been placed in a separate category above but might perhaps just as well have been placed in category (b) as far as common sense goes. What is special about graphs is that, on the one hand, they are useless as external representations without an accompanying explanation of their mapping principles. This is also true of arbitrary pure modalities. On the other hand, graphs represent structured data which in their turn represent the world, and graphs do have structural similarities with the data they represent (Bernsen 1992). Graphs are therefore in a very specific sense 'in between' analogue and non-analogue representations without this fact acting as a threat to the clarity of the distinction.

It may also be noted that the common sense categorisation maps directly into the analogue (b+d)/non-analogue (a+c+e) distinction above and may therefore be considered simply a more differentiated version of that distinction. Apart from the special case of graphs, the list of categories (a)-(d) is probably the one which is closest to our standard intuitions about the domain of investigation. In particular, categories (a)-(c) are well-known and (d) is easily understood as an extension of (b) into two media different from that of vision and visual qualities.

The fact that the classification just considered is close to common sense does not make it theoretically irrelevant. On the contrary, this is the classification we have to turn to in order to analyse the basic differences between, e.g., natural language modalities and analogue graphics, sound and touch modalities as external representations of information (see Bernsen 1993).

6. What Is a Pure (Generic) Modality?

Having presented a taxonomy of pure generic modalities and some orthogonal classifications of these above, the following operational definition of a pure generic modality comes out straightforwardly. A pure generic modality is characterised by a specific *medium of expression* and what may be termed a *profile* constituted by its characteristics as selected from the following list of binary opposites: analogue/non-analogue, arbitrary/non-arbitrary, static/ dynamic, linguistic/non-linguistic. Given this list, our 14 pure generic modalities all have different medium/profile characteristics, except for the following two pairs:

(a) 5. Diagrammatic pictures.
6. Non-diagrammatic real-world representations or pictures.

(b) 8. Animated diagrammatic pictures.
9. Dynamic real-world representations.

The pairs (a) and (b) are analogous to each other, the difference between them being their static and dynamic character, respectively. There does not seem to be any principled way of distinguishing between external representations that are 'really' like what they represent and external representations that are less like what they represent because of leaving out some aspects of what they represent (cf. Twyman 1979). In other words, there seems to be a *continuum of representation* between fully naturalistic external representations and more or less abstract and schematic representations. However, we do appear to have robust intuitions to the effect that analogue diagrammes are different from 'real pictures' and these intuitions seem to be equally valid in the dynamic domain. If this is true, then it is likely that the only way of making the diagrammatic/non-diagrammatic distinction is through the use of *prototypical instances.* There are prototypical 'real pictures' such as ordinary photographs, and there are prototypical analogue diagrammes such as diagrammes of house layouts, engine parts or traffic accidents. Similarly, there are prototypical dynamic real-world representations such as videos and there are prototypical animated diagrammatic representations such as "virtual reality" computer games using sound and graphics and many scientific visualisations. If that is true, then the fact that principled distinctions are impossible within each of the pairs (a) and (b) is rooted in facts of the matter and is not a consequence of lack of important characteristics in the proposed taxonomy.

Incidentally, the taxonomy ignores the difference between analogue external representations which have a 'real original' which they more or less faithfully represent and analogue external representations which in some sense might have had a real original but just happen not to have one, for instance because the real entitity is about to be built as a result of ongoing work on CAD analogue screen representations. Such distinctions belong to the level of analysis of atomic types (see Sect. 8 below).

The taxonomy is presented in Table 2. Note that, in this table, shading acts as an extra information channel. Being arbitrary, the various types of shading have to be explained in the note to the table. Since the top row and left-hand column of Table 2 contain word icons, these have had to be disambiguated in the text above. Note finally that Table 2 provides a much clearer overview of the taxonomy that did Table 1. This is due to the combined abstract diagrammatic and written natural language properties of Table 2.

	Analogue	Non-analogue	Arbitrary	Non-arbitrary	Static	Dynamic

Spoken language		X		X		X
Written language		X		X	X	(X)
Real sound	X			X		X
Arbitrary sound		X	X			X
Diagram pictures	X			X	X	
Non-diagramme pictures	X			X	X	
Arbitrary diagrams		X	X		X	
Animated diagramme pictures	X			X		X
Dynamic pictures	X			X		X
Animated arbitrary diagrams		X	X			X
Graphs		X		X	X	X
Real touch	X			X		X
Arbitrary touch		X	X			X
Touch language		X		X		X

Table 2. Taxonomy of pure generic external modalities. No shading indicates the medium of visual and graphical qualities/vision. Light shading indicates the medium of sound qualities/audition. Darker shading indicates the medium of tactile and kinaesthetic qualities/touch.

7. Exclusiveness and Exhaustiveness of the Taxonomy

Questions pertaining to the *exclusiveness* of the taxonomy have been discussed at various occasions above. The conclusion is that the taxonomy is generally exclusive, but that exclusiveness is compromised in at least two types of cases. One is that the distinction between analogue and non-analogue modalities in the same medium and either being both static or both dynamic, is sometimes difficult to draw (cf. Sect. 5.1 above). Another is that some modality distinctions in the graphical medium are based on prototypicality and hence, by definition, so to speak, are not exclusive in any (other) principled sense (Sect. 6). It remains an open question to what extent prototypicality will prove to be relevant to a more detailed understanding of modalities in other media of expression.

	Analogue	Non-analogue	Arbitrary	Non-arbitrary	Static	Dynamic

Spoken language	- f	X	- d	X	- n	X
Written language	- f	X	- d	X	X	(X)
Real sound	X	- d	- d	X	- n	X
Arbitrary sound	- d	X	X	- d	- n	X
Diagram pictures	X	- d	- d	X	X	- d
Non-diagramme pictures	X	- d	- d	X	X	- d
Arbitrary diagrams	- d	X	X	- d	X	- d
Animated diagramme pictures	X	- d	- d	X	- d	X
Dynamic pictures	X	- d	- d	X	- d	X
Animated arbitrary diagrams	- d	X	X	- d	- d	X
Graphs	- d	X	- d	X	X	X
Real touch	X	- d	- d	X	- d	X
Arbitrary touch	- d	X	X	- d	- d	X
Touch language	- f	X	- d	X	- d	X

Table 3. This table shows that there seem to be strict limits to the existence of pure modalities in addition to the ones already identified. Cells that were empty in Table 2 have been filled with one of the following labels: **-f** means that it is simply a contingent matter of fact that a cell is empty or near-empty; **-d** means that it is a matter of definition that a cell is empty; and **-n** means that it is a fact of nature that a cell is empty.

The question whether the taxonomy is *exhaustive* is the question whether there might exist other pure generic modalities than those listed in Tables 1 and 2 above. In one sense this is evidently the case since we have ignored media of expression such as gesture, taste and smell and their corresponding perceptual qualities. Let us consider here a second sense of the question, namely, whether there are or might be other pure generic modalities in the three media we are considering. The answer is that this is only possible to a very limited extent. Why this is so begins to become apparent from Table 3.

In the cases in which it is truly a matter of definition or a fact of nature that a cell is empty nothing could possibly change the situation. In Table 3, the large majority of cases belong to one of these two categories. Contingent factual falsehood (**-f**) is only apparent in the case of language where we just might have had more onomatopoetica in our languages or more written or touch languages using hieroglyphs or other diagrammatic picture icons. It is not to be expected that information systems artifact design is going to change this situation very much (although some interfaces use so many graphical picture icons that they come close to re-inventing the hieroglyphs). As indicated above, it may be questioned whether static touch modalities should be added.

One intuition, for what it is worth, is that touch should be generally categorised as being dynamic because of the intimate connection between touch sensations and movements of the body surface of the person doing the touching. However, we can of course receive many different touch sensations without movement, such as electrical current, heat, passive contact with objects, etc. I must confess to not having a clear answer to propose on this question, nor to have any clear idea of its significance. Similarly, as noted above, music is a difficult case for which no convincing solution has been proposed in this paper.

If the information provided in Table 3 is correct, the taxonomy of pure modalities is close to being exhaustive, given the limited number of media it addresses. As said in Sect. 3 above, we might have reduced or increased the number of pure generic modalities in a number of simple and well-defined respects. Otherwise, there seems to be little opportunity for creating additional pure modalities within the media addressed. This means that, at the level of descriptive generality adopted, we have a reasonably robust taxonomy of the pure generic modalities of external representation which, either in their unimodal forms or in combination with other modalities, go into the building of human-computer interfaces to constitute multimodal and virtual reality representations.

Table 4 shows the full set of permutations on the taxonomy, including all the pure generic modalities considered in this paper. To reduce the size of the table from 48 to 24 rows, linguistic modalities have been indicated in boldface numbers. The table shows the exclusivity of the taxonomy, except for one point. Consider the idea of a dynamic (non-analogue) touch language in the sense in which non-analogue written language can be represented dynamically (cf. Sect. 3 above). Dynamic touch language would have to belong to row x of Table 4 which therefore would contain two linguistic modalities, thus disproving the exclusivity of the taxonomy. And if we refrain from doing this, the exhaustiveness of the taxonomy has been disproved. Even worse, once dynamic touch representations have been included in the linguistic case, we seem to have to consider including the distinction between static and dynamic touch representations in general, thus adding *static* real touch, arbitrary touch, diagrammatic touch and analogue touch language to the taxonomy. A total of 5 such static touch modalities have been inserted into the otherwise 'forbidden' zones of the taxonomy.

		an	-an	ar	-ar	stat	dyn	gra	sou	tou
a		x		x		x		x		
b		x		x		x			x	
c		x		x		x				x
d		x		x			x	x		
e		x		x			x		x	
f		x		x			x			x
g	5/6,**16**	x			x	x		x		
h		x			x	x			x	
i	12/19, **20**	x			x	x				x
j	8/9	x			x		x	x		
k	3/18,**15**	x			x		x		x	
l	12/19,**20**	x			x		x			x

m	7		x	x		x		x		
n			x	x		x			x	
o	13		x	x		x				x
p	10		x	x			x	x		
q	4		x	x			x		x	
r	13		x	x			x			x
s	**2**,11a		x		x	x		x		
t			x		x	x			x	
u	14		x		x	x				x
v	11b,**17**		x		x		x	x		
w	**1**		x		x		x		x	
x	**14**		x		x		x			x

Table 4. The full set of permutations on the taxonomy. The 12 rows in dark shading are necessarily empty, except for the problem of touch. Rows a-f are empty because analogue representations cannot be arbitrary. Rows h-j, n-o and t-u are empty because of the dynamic character of sound and touch representations (see text, however). The table clearly shows how the remaining 12 rows contain all the (21) pure generic modalities discussed in this paper. ***/*** between two numbered modalities indicates that the difference between them is based on prototypes. Numbered modalities in boldface are linguistic modalities. The modalities are:

1 = Spoken language.
2 = Written language.
3 = Real sound.
4 = Arbitrary sound.
5 = Diagram pictures.
6 = Non-diagramme pictures.
7 = Arbitrary diagrammes.
8 = Animated diagramme pictures.
9 = Dynamic pictures.
10 = Animated arbitrary diagrammes.
11a = Static graphs.
11b = Dynamic graphs.
12 = Real touch.
13 = Arbitrary touch.
14 = Touch language.
15 = Analogue spoken language (onomatopoetica).
16 = Analogue written language (hieroglyphs).
17 = Dynamic written natural language.
18 = Diagrammatic sound.
19 = Diagrammatic touch.
20 = Analogue touch language.

8. Modality Structure

It was suggested in Sect. 4 above that icons constitute a modality structure which can be found anywhere among the pure generic modalities. Let us reconsider the well-known types of each of the pure generic modalities from Table 1 above. It is clearly important to the understanding of the expressive potential of modalities to analyse in depth the different types subsumed under our 14 generic modalities. While this is outside the scope of this paper, I want to mention an interesting observation suggested by the omnipresence of icons in the taxonomy. This suggests that modality structures may exist across all the distinctions basic to the taxonomy. Intuitively, one might perhaps have assumed that even if the list of pure generic modalities is finite and reasonably exclusive and exhaustive, the number of modality types subsumed under it would form a huge and poorly structured class whose individual members would have to be analysed more or less independently of each other. The case of icons suggests that this might not be the case. If one follows this lead, it turns out that several other structural types are found across the taxonomy. Lists, for instance, can be made up of icons, words, text, pictures, animations, touch qualities and so on. Analogue diagrammes can be created across the media. Topological maps can be created not only in diagrammatic graphics but also in sound and touch. Abstracting the 'well-known types' from Table 1 above and applying this idea, we obtain a much more structured view of the pure modality types (see Table 5).

One interesting point about Table 5 is that the domain of each pure generic modality has now been split into two different subsets of types. The first subset contains the *atoms of representation* characteristic of the modality. These will have to be analysed in their own right. The second subset contains the *modality structures* or structuring principles which may be applied to the atoms of a particular generic modality. These modality structures are limited in number and each cut across many different generic modalities. A second point of interest is that the list of types no longer contains only well-known types. Some less well-known types have been added through using the structuring principles generatively. I shall return to these two points in the conclusion.

1. Spoken letters, words, numerals, other spoken language related sounds, text (= spoken word **sequences**), **lists**, **icons**.

2. Written letters, words, numerals, other written language related signs, text (= written word **sequences**), **lists**, **tables**, musical notation, **icons**.

3. Single real-world sounds, sound **sequences**, **lists**, sound **diagrammes**, **maps**, **icons**, music?

4. Single arbitrary sounds, sound **sequences**, abstract sound **diagrammes**, **lists**, **icons**.

5. Diagrammatic pictures (analogue **diagrammes**), **maps**, **sequences** of such, **lists**, **table**s, **icons**.

6. Photographs, naturalistic drawings, **sequences** of such, aereal **maps**, **lists**, **tables**, **icons**.

7. Points, lines, boxes, circles, volumes, etc., **sequences** of such, abstract **diagrammes**, **lists**, **tables**, **icons**.

8. Diagrammatic animations (animated **diagrammes**), **sequences** of such, **maps**, **lists**, **tables**, **icons**.

9. Films, videos, **sequences** of such, aereal **maps**, **lists**, **tables**, **icons**.

10. Points, lines, boxes, circles, volumes, etc., **sequences** of such, abstract dynamic **diagrammes, lists**, **tables**, **icons**.

11. A graph space containing 1D, 2D or 3D geometrical forms, **lists**, **tables**, **icons**.

12. Single real-world touch representations, touch **sequences**, **lists**, touch **diagrammes**, **maps**, **icons**.
13. Touch signals of differents sorts, touch **sequences**, **lists**, **tables**, **icons**.
14. Touch letters, words, numerals, other touch language related signs, text (= touch word **sequences**), **lists**, **tables**, **icons**.

Table 5. Generic modality types organised according to modality structures.

9. External and Internal Representations

We have been considering only external representations in this paper. If, on the other hand, we attempt to go inside the cognitive system to consider the nature of *internal* representations, then the analogue/non-analogue distinction may not have much relevance any more. The reason is apparent from the distinction between arbitrary and non-arbitrary external representations in Sect. 5.2 above. It turned out that there are more categories of external representations which exploit already existing systems of meaning than there are analogue external representations. A different way of expressing this is to point out that internal representations, in order to serve their purpose of representing the world, may themselves to a large extent be analogue representations (in some sense) which build on material that has ultimately been derived from perceptual input to the human cognitive system. This may also be true of representations to which the words of natural language have been conventionally attached (see Bernsen 1993). Internal representations, therefore, constitute a domain of research very different from the domain of external representations addressed above. It would perhaps have been easier if we did not have to enter this domain when analysing the foundations of modalities and multimodal representation. This cannot be avoided, however, since we cannot avoid issues concerning what, e.g., natural language is good at representing, or what graphics is good at representing, what various combinations of natural language and graphics are good at representing, or indeed what many different combinations of modalities are good or bad at representing. It is not possible to provide relevant explanations of such issues without discussing the nature of the internal representations linked to, e.g., natural language and graphics. The reason is both simple and compelling:

Considered purely as an external representation, a written natural language sequence on a screen would merely clutter up the screen without contributing anything else. What makes the sequence potentially useful for the representation of information is the fact that users *understand* the language used, i.e., that they have access to the system of meaning on which the language is based. This means that they are able to form appropriate internal representations of what some written word sequence represents. These internal representations are not identical to the external (written) representations which cause or evoke them. If they were, then another process of interpretation would have to take place, and so on *ad infinitum.* So if we want to, e.g., optimally combine natural language and graphics for representing something on a screen, we cannot avoid considering the nature of the internal representations which are likely to be evoked in users by the natural language we consider using and by the graphics we might use. Nor can we avoid considering the cognitive processes operating on internal representations, the various cognitive limitations on these processes, effects of users' background knowledge on their understanding of mapping principles as well as on their interpretation of the specific types of external representation used, and so on. The same is true of many other modality combinations.

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

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

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

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

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

10. Concluding Discussion

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

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

If the taxonomy of pure generic modalities is anything to go by, there is a huge number of existing and possible combinations of modality types. A tiny fraction of these are well-known and have been rather extensively analysed in the literature, such as (static) analogue graphical/written natural language maps, (static) graphical/written natural language analogue or abstract diagrammes, or graphical/written natural language graphs. However, there are literally thousands of possible modality combinations. For instance, no one is currently able to exclude the unfamiliar prospect that some combination of written natural language tables and datagloves might some day achieve prototypical status and a name in language because of having become popular for the performance of some prominent task category; or, some combination of types belonging to every one of our 14 (or 21 or 26!) pure generic modalities might soon become involved in advanced virtual reality representations of, say, flight decks. There is no clear sense at this point in undertaking a detailed analysis of each and every such actual or possible modality type combination. Ignoring the scale of such an undertaking it would not be feasible without sound foundations. Rather, the only viable solution seems to be to establish foundations which enable a principled scientific analysis of *any given* modality combination *once* it is considered for analysis.

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

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

(2) Analyse the atoms of each pure generic modality starting from their characterisation through the taxonomy.

(3) Analyse the modality structures which cut across the boundaries imposed by the different categorisations of the taxonomy, i.e., the 14 generic modalities, the binary opposites and the media.

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

References

Bernsen, N.O.: Matching Information and Interface Modalities. An Example Study.*Working Papers in Cognitive Science* WPCS-92-1, Centre of Cognitive Science, Roskilde University 1992. *Esprit Basic Research project GRACE Working Paper* 1992.

Bernsen, N.O.: Specificity and Focus. Two Complementary Aspects of Analogue Graphics and Natural Language. *Esprit Basic Research project GRACE Working Paper* 1993 (submitted).

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.

Twyman, M.: A Schema for the Study of Graphic Language (Tutorial Paper). In Kolers, Wrolstad and Bouna (Eds.): *Processing of Visible Language*, 1979.

Acknowledgements: The work described in this paper was carried out under Esprit Basic Research project 6296 GRACE whose support is gratefully acknowledged. Discussions with Michael May have been of great help during the development of ideas. Kenneth Holmquist has provided valuable comments on an earlier draft.