This paper was presented at the IS&N Conference '95 in Heraklion, Greece. Full Reference:

Condon, C. & Keuneke, S. (1995)
Counting the Costs and Benefits of Metaphor
in: Clarke, A.; Campolargo, M; Karatzas, N. (Eds.)
Bringing Telecommunication Services to the people - IS&N '95
Springer, Berlin

Counting the Costs and Benefits of Metaphor

Chris Condon & Stephan Keuneke

Brunel University Centre for Research in Information Environments
Bremer Institut für Betriebstechnik und angewandte Arbeitswissenschaft
Chris.Condon@brunel.ac.uk keu@biba.uni-bremen.de

Abstract. It has been demonstrated that the use of suitable metaphors in the user service interface can have a dramatic effect on the way in which the user perceives the services, depending on the category of metaphor chosen. An earlier paper by the authors postulated that interactional metaphors might lead to greater long term usage of services, as they direct the user to thinking about what the services are for, rather than thinking about the services themselves. To test this hypothesis, a model of usage was built, showing the likely impact of using different types of metaphor. The model showed that usage levels over time varied according to metaphor category. These usage patterns were then imposed on existing techno-economic models from specific industry sectors. Interactional metaphors led to the highest long term usage in most industry sectors, although the construction industry showed higher usage with spatial metaphors. In all sectors, inappropriate choice of metaphor would be sufficient to destroy the economic advanatages of advanced communications services.

1 Introduction

Experiments within the MITS project have demonstrated that the use of suitable metaphors in the user service interface can have a dramatic effect on the way in which the user perceives the services and whether they are likely to be used. Two factors were identified as critical to this. The first factor concerns the mapping between the vehicle (the real world artefact) and the service and is discussed by Anderson et al. [AND94] The second factor concerns the categorisation of metaphors and the manner in which the choice of metaphor category affects what the user is aware of.[CON94]

It is this second factor which is considered in this paper which seeks to assess its impact on the actual usage of new services. The changes in the user's perception are qualitative, rather than quantitative but, to understand their impact on usage, they must be translated into quantitative measurements. An earlier paper by the authors had postulated that interactional metaphors might be more successful, as they direct the user to thinking about what the services are for, rather than thinking about the services themselves.[CON94] To test this hypothesis, a model of usage was built, using CRIMP (CRoss IMPact analysis tool),[KRA94] showing the likely impacts of the different types of metaphor.

2 The CRIMP Model

2.1 Cross Impact Analysis

The CRIMP tool models the impacts of trends on other trends and on themselves. Once a model is built, the software steps through the changes in trends in each time frame. Depending on the confidence associated with the predicted impacts, a randomising factor is added to the model and the run-through is repeated a hundred or more times. Actual values are not used by the model as each factor is normalised into a range from -8 to +8 with the initial value set to 0. This makes the tool valuable for modelling subjective values, as in the model described below. The underlying principles and mathematics of CRIMP are described elsewhere.[GOR68], [KAN72], [HEL77], [DUI95]

2.1 Factors

Four factors were identified as having a potential impact on usage trends. The first of these summarises the various elements which lead to resistance to new technologies. It would be fatal for a company to assume that a new service is used as much and in the way it is supposed to from the day it is installed. Users have to be trained and old working habits have to be overcome, sometime even by those who do not directly work with the new service. There is evidence that even systems that are well accepted after installation show a decay of use in the long run.[HUT93] This factor had already been identified by URSA as a potential inhibitor:

> "Human factor problems in the form of psychological resistance are often associated with process re-engineering or company re-organisation as it may be perceived by part of the management and the labour force as a threat to their position or to the control they exert. The old jobs consisted of specialists who did one task. The new case handlers perform a variety of tasks. Therefore people working on case handing process teams will find their work far different from the repetitious performance of one task to which they were accustomed."[SIN94]

At an earlier stage of the MITS project a number of experiments were carried out which identified critical factors in the choice and design of interface metaphors. Of these, the BIBA demonstrator showed a *qualitative* difference between the impact of metaphors on the use of advanced telecommunications services. A system of categorisation of metaphors based on three axes had been proposed:

Spatial

Activity-based

Interactional

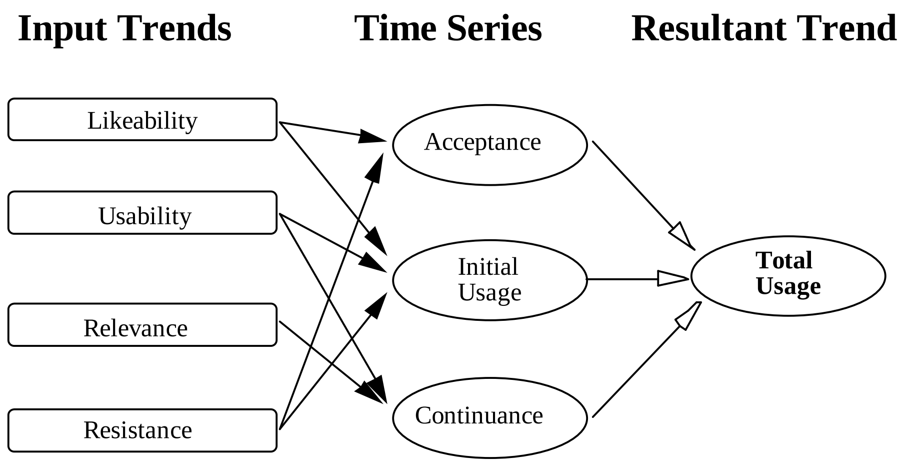
Particular metaphors usually embody varying aspects of each of these concerns and can be positioned in a three dimensional space with respect to these three axes. The classification, originally based on an extensive revision of the interface metaphor types identified by Hutchins,[HUT89] is further described in Anderson et al.[AND94] It should be noted that the classification deals with the underlying metaphor, not the medium in which it is presented. For example, spatial metaphors can be presented in verbal form, as in some text-based adventure games, "You are in a room. There are doors on your left and on your right. Stairs lead down." Experiments demonstrated that metaphors typifying these three axes influence the user in more than just *how* the user sees the functionality available; they also influence *what* the user sees.[CON94]

Activity-based interface metaphors can focus on differing levels of generality. For example, collaborative systems can be designed in terms of metaphors for specific tasks, such as project management, or can be generalised, such as 'agents'.[LAU90] Activity-based metaphors turn the user's attention to the functionality of the services offered and, by making the user more aware of the functionality at their disposal, tend to support the general <u>usability</u> of the system.

The spatial aspect of the metaphor is often considered as a means of providing a location for tools and methods of communicating and working. However, these spaces can also be more or less explicitly defined. So, it is possible to utilise stereotypical aspects of particular places in the design (e.g. libraries) or, more commonly, general properties of spaces (e.g. rooms).[CON90],[CON92] Spatial metaphors emphasise the interface itself, turning the user's attention towards the presentation of the services. By providing interfaces which attract the user, spatial metaphors encourage initial <u>likeability</u> and, it is suggested, are particularly good for new or novice users.

Concerning the interactional aspect, interfaces can support particular forms of communication (e.g. conventional e-mail), or provide less explicit spaces or opportunities for interaction through artefacts such as forms.[HAM91a] Interactional metaphors tend to turn the user's attention towards the activities which the services support and therefore the <u>relevance</u> to the users tasks, providing longer term motivation. Although this might appear to point to interactional metaphors as most suitable in the long term, arguments have been presented that a spatial metaphor such as the room is more suitable for real-time interaction in which the interaction itself is clear, whereas an interactional metaphor such as the form is better in supporting non real-time cooperation in which the interaction is less obvious.[HAM91b]

This gives us four trends for the model which impact on a fifth factor, <u>Usage</u>. The factors do not have an even effect on usage over time and are therefore filter through time series (see below). The interactions between the trends and time series are shown in the diagram below.



Cross Impacts in the MITS model

This results in the following matrix defined for the CRIMP model:

	Likeability	Usability	Relevance	Resistance	Usage
Likeability				-0.1	S
Usability					S
Relevance				-0.5	S
Resistance				0.2	-2
Usage					1
PoorL	0				
PoorU		0			
PoorR			0		

The five trends defined above are shown across the top of the matrix as trends which are impacted **on**. These impacts come **from** the eight factors shown in the side column. These consist of the same five trends together with three actions: PoorL (poor likeability), PoorU (poor usability) and PoorR (poor relevance). These actions denote poor implementation of the interface in each of the three categories, but in the *a priori* model their impact is set to zero. Generally, the following is given as a guidance to the scale of cross impacts:[DUI94]

0.1 –> Small Impact

0.5 –> Medium Impact

1.0 –> Large Impact

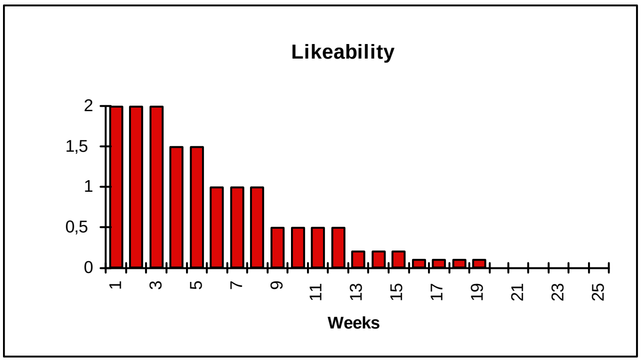
2.0 –> Very Large Impact

Thus, it can be seen that likeability is defined as leading to a small reduction in resistance, and relevance a medium one. However, resistance tends to build on itself, with a small-medium sized impact. All the trends also directly impact on usage. In the cases of likeability, usability and relevance, these impacts change over time. The 'S's indicate time series which are described below. Resistance has a constant, very large, negative impact on usage, while usage has a strong, positive impact on itself (as more people use telecommunications services, their usefulness increases). Although the scale of these impacts are based on 'common sense' rather than empirical evidence, run-throughs of the model with varying impacts showed no significant difference in final results.

2.2 Time Series

In the cases of resistance and usage, the impact on usage is constant. For the other three trends, the manner in which they impact on usage was set as a time series with an emphasis on one of three time series:- the *acceptance* of the services by the user; the *initial usage* of the services; the longer-term *continuance* of use.

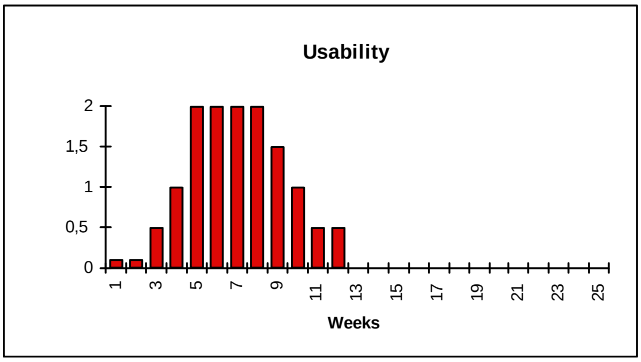
Service Acceptance. It is common to see some enthusiasts taking to new services immediately, with others taking a more cautious attitude. Although acceptance will be affected by all three trends, the dominant factor at the interface at this stage is likeability of the interface. A sufficiently likeable system will attract short term interest, even from users to whom it has no relevance and even if its usability is poor, c.f. the attraction of early VR (Virtual Reality) demonstrations. This can be expressed as a very strong initial impact in favour of using the system which gradually fades away:



Service Acceptance Time Series

All time series are shown over a period of 25 weeks, after which it is assumed that usage will settle down at a relatively constant level. The scale of impacts is the same as that given in the main CRIMP matrix, i.e. from 0.1 (weak) to 2.0 (very strong).

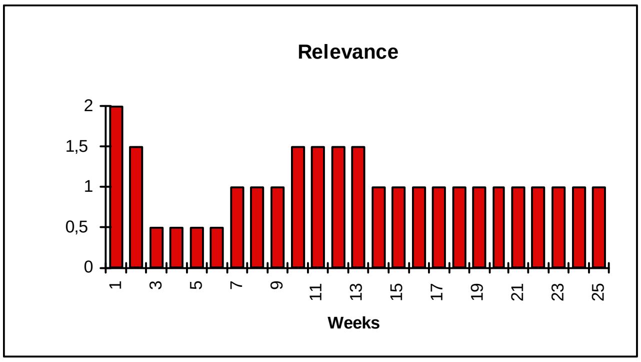
Initial Usage. Poor usability may lead to a lack of usage even though the initial likeability of the interface attracted the user's attention and the system is perceived as having high relevance to the user's tasks. Even with poor usability, some users will master the services. Thus, the impact of poor usability will not be immediate but will mainly be in the form of a certain percentage of new users giving up the system within a relatively short time, as shown in the graph below:



Initial Usage Time Series

Service Continuance. As with likeability, the relevance of the services will affect the initial willingness of users to take up the service but it will also affect the users' long-term motivation. People who have started to use systems which have been designed to meet their known needs and have been implemented with high usability may still drift away from their use as poor motivation leads to a decreasing awareness of the relevance of the system to their work. This will affect existing users more strongly than those who have yet to start using the services and both will therefore encourage a fall-off of usage among existing users. It should be noted that it is not the actual relevance of the system which matters: it is the relevance perceived by the user

The impact of the perceived relevance of the system will initially be very strong. As factors such as usability involve the user more in issues of 'how' rather than 'why' to use the system, the effect will fade away. Having mastered the usage of the system, the effects of motivation then come more clearly into play and will continue into the longer term:



Service Continuance Time Series

In all cases, the exact formulæ for the graphs cannot be known. However, this does not affect the validity of the model as long as the **relative** time frames are understood and there is some idea of the **relative** impact of the different factors: as explained above, CRIMP works with qualitative data, not just quantitative. For example, the shape of the 'Service Acceptance' curve above will remain the same whether a service is taken up by 10% or 90% of potential users within a given time frame: all that changes will be the gradation of the time axis. It may be noted that relevance shows a greater overall impact: this was taken into account when defining the actions described below.

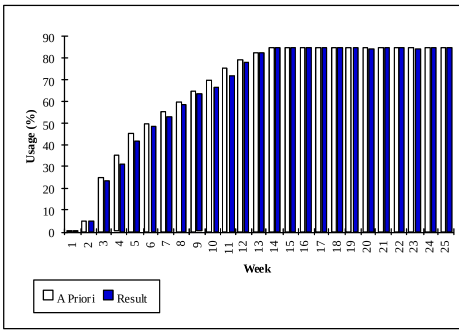
2.3 Trends

The four trends of likeability, usability, relevance and resistance were all given constant values throughout the time series of 50 on a scale of 0-100. This is an arbitrary figure representing 'typical' values for these factors (the actual figure for a constant trend does not affect the CRIMP model, only changes to the figure will have an impact). For the overall usage it is expected that, typically, it will take time for users to learn to use the services but that usage will steadily climb, with 50% of potential users active within six weeks. After this, usage will continue to grow to a maximum of 85% – a figure of 100% never being achievable, due to factors such as equipment out of service, users on leave, etc. Running the model with this assumption showed a close correlation with the *a priori* figures, indicating a robust, consistent model.

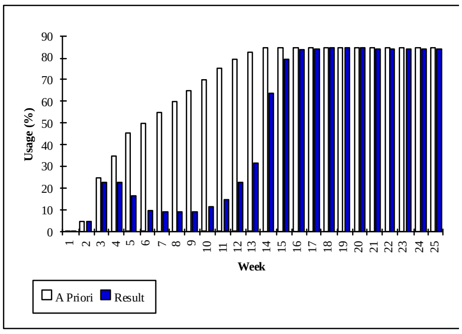
2.4 Results

The constant values for the three interface factors assumed 'typical' interfaces, i.e. just good enough to allow the expected take-up of services. These could, in turn, be affected by poor interface design in any one of three areas (the PoorL, PoorU and PoorR factors). Factors which have a single effect on a model are known in the CRIMP methodology as *actions*.

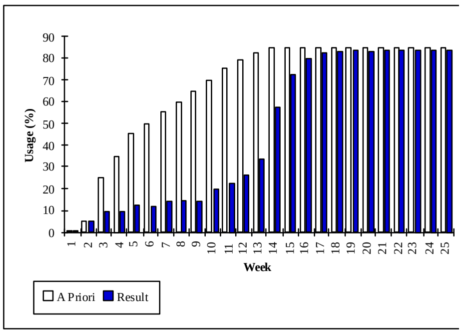
The model was run, in turn, with a negative impact for each of the actions on the relevant trend, e.g. with a cross impact of -10 of *PoorL* on *likeability*. To compensate for the greater impact of the *relevance* time series, the cross impact for this was -5, i.e. a 10% reduction on the 'standard' interface, whereas the other trends were treated to a 20% reduction. By running each in turn, the following usage profiles were obtained:



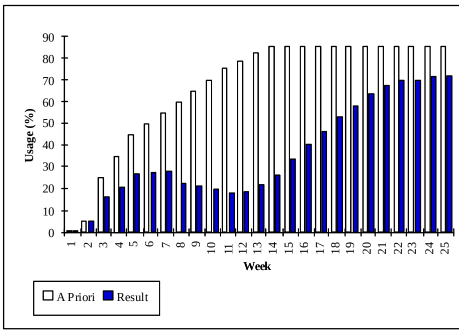
Initial values



Reduced Usability



Reduced Likeability



Reduced Relevance

Impact of Actions on Usage over time

The output shows that, in all cases, the impact on overall usage is much greater than the 10-20% changes made to the interface factors. It also demonstrates that reductions in usability and likeability have immediate effects which are overcome over time, whereas a reduction in perceived relevance has a smaller but longer term impact.

3 Industry Sector Models

To fully understand these changes in usage, it was then necessary to examine their impact within specific industry sectors. RACE project URSA had carried out a number of such studies which were used to provide the *a priori* values. These resulted from a systematic and in-depth study of the innovative usage of advanced communications in many economic sectors. The two objectives of this study were:

- to identify the key applications on which innovative demand for advanced communications is likely to be based in the European economy
- to describe and quantify the benefits generated by these applications

It is difficult to measure the impact of advanced communications on company productivity and other aspects of competitive advantage, as advanced services are still in the development phase and as usage conditions in pilot experiments cannot often be compared with real commercial usage conditions. URSA therefore adopted an in-depth and simulation based case-study approach to measuring user benefits, complemented by an empirical survey. This approach involved the following steps:

- Identify key applications for sectors into which economic activity is aggregated at the EU level.
- Reconstruct key value generating processes of representative companies in the different sectors.
- Simulate impacts of the identified applications on the outcome of the value generating processes in each company.
- Check validity of simulation results in an empirical survey on company acceptance of the identified applications. A summary of simulation results was presented to 120 companies distributed across the sectors, and their feedback was collected through interviews and questionnaires.

Benefits in terms of productivity gains together with expected penetration rates allowed for the calculation of an ECU equivalent of the total impact of the identified applications on each sector. Data on turnover, employment and number of companies was taken from European Commission surveys.[EUR93]

4 Impact of Metaphor Choice on the Sector Models

4.1 Method

The MITS usage patterns shown above were then put into selected models from the URSA project, chosen to accord with the pilots within the MITS project. Although the MITS model appears to show a dominant effect for the relevance factor, this is not necessarily the case when this data is combined with other effects. For example, likeability has a much greater impact in the very short term which could, in some cases, be more important than the longer term effect of relevance.

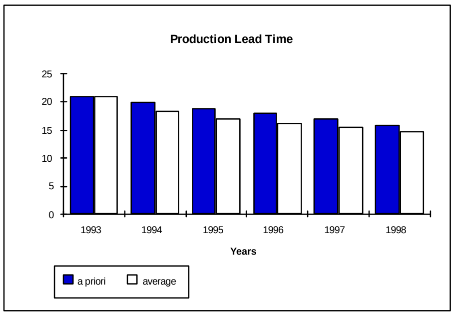
Two types of action were therefore defined:

• **Delay**: the dramatic impact of poor likeability and usability in the early stages can be summed up as a delay in take-up.

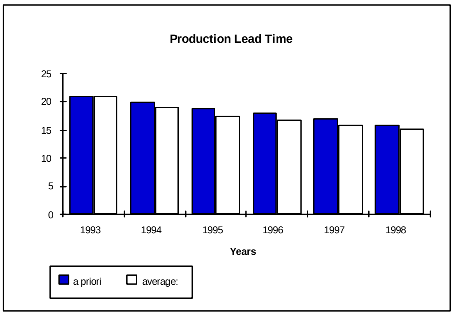
• **Reduction**: although not as dramatic in its initial impact, poor relevance leads to a small longer-term reduction in usage.

4.2 Electrical Engineering

The chosen application examined in the electrical engineering model was that of EDI, in this case referring to the exchange of information with customers and suppliers, including videoconferencing as well as more conventional EDI. This closely corresponds with the services being examined in the PTT Telecom and Nokia pilots of MITS. URSA's CRIMP model shows the introduction of these services as an action having a direct impact on production lead time. This was therefore taken as an indicator. The results of the URSA model (unmodified) showed a reduction in lead time of 1.5 days as advanced services were implemented. Reductions in likeability reduced this only in the first year, but poor relevance reduced the long term improvement by one third (from 1.5 days saved to 1 day).



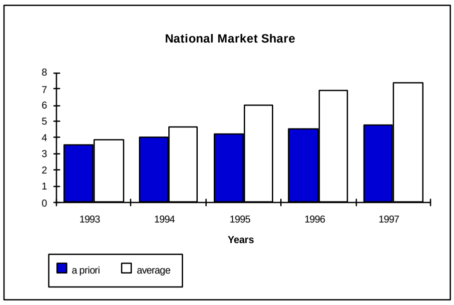
URSA's results

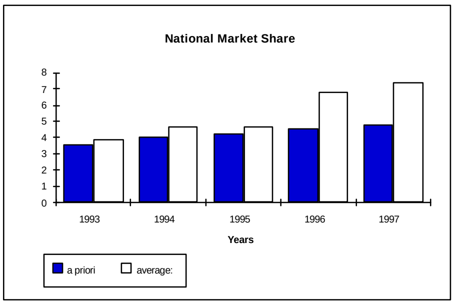


Impact of poor relevance

4.3 Construction

For construction, the chosen application was more dramatic and more obviously linked with metaphor: the virtual meeting room. This application is one of the central pilots within the BRICC project, using similar service interfaces to those in the BIBA/Vero pilot of MITS. URSA related this application to its potential impact on the national market share of a large construction company, in particular the ability of such services to allow the company to penetrate into parts of the sector in which it currently has little presence. In this case, poor likeability had a greater effect.



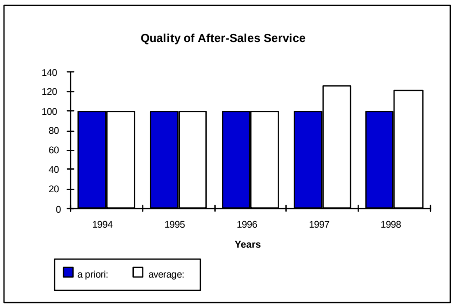


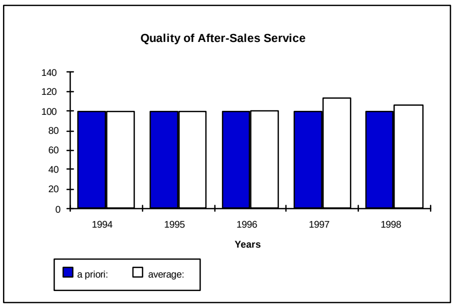
URSA's Results **Impact of Poor likeability**

In this case, the impact of the new services is predicted as dramatic and steadily increasing, both in absolute and proportional terms. By the final year of the model, the potential market share for the company becomes 7.45% rather than 4.8%. The action continues to have a significant impact for some time after its introduction. Although the potential market share in the final year appears to be little reduced, it is likely that the reductions in the previous years would jeopardise this.

4.4 Transport Equipment

For transport equipment, the application chosen (as in the MITS pilot by LUTCHI and the Automobile Association) was that of remote delivery of expertise for maintenance and interference rectification. As an additional service which could be provided to customers, this was not related by URSA directly to market indices of the types used in the other models, but to a general index of quality of after-sales services, on a scale from 0-200 with 100 representing a 'typical' quality level which can currently be maintained. In this case, poor relevance proved to have a much stronger impact:





URSA's Results **Impact of Poor relevance**

As a more advanced service which will not be available for some time, its impact does not come into effect until 1997. The impact of this service is therefore not as strong as those of the applications examined in the other sectors. In this case, the figures for 1997 are hardly changed, but by 1998 the impact of the service has faded away almost completely, back to a value of 107.

4.5 Conclusions

The models all demonstrate that factors associated with metaphor usage and extending beyond conventional usability can be critical in supporting the potential for advanced telecommunications services. In many cases, the most critical factor that designers of new services must consider is the relevance of the service, not just the actual relevance to the user's tasks, but also the manner in which the interface makes the user aware of the relevance. In some cases, likeability can also be a critical factor, as in the example from construction.

These categories are not exclusive, nor should the designer concentrate on a single factor to the exclusion of others. For example, the 'virtual meeting room' used in the construction sector example could be designed to emphasise the 'room', the **spatial** aspect, or to emphasise the 'meeting', the **interactional** aspect, and ideally should project both aspects for their separate usefulness. In this case, the **spatial** aspect could be used to present the intial impression of the system with a high the likeability factor, whereas **interactional** aspects could be supported more strongly as the users begin to use the services, emphasising their relevance.

References

[AND94] Anderson, B; Smyth M; Knott, R.P; Bergan, J; Bergan, M. & Alty, J.L. "Minimising Conceptual Baggage: Making Choices about Metaphor", *proc* HCI'94, Cambridge University Press, 1994

[CON90] Condon, Chris "Networked Cooperative Work: Usability Issues of MILAN", *proc* Telematics '90, BIBA, Bremen, Germany, 3-5 Dec 1990

[CON92] Condon, Chris, "Multimedia & CSCW in Manufacturing: Working with MILAN" *proc* Conferencia ESPRIT: CIME & Competividad Empressarial, Bilbao, 1992

[CON94] Condon, Chris & Keuneke, Stephan "Metaphors and Layers of Signification: The Consequences for Advanced User Service Interfaces" in "Towards a Pan-European Telecommunications Service Infrastructure - IS&N '94", ed H-J Kugler, A Mullery, N Niebert, Springer-Verlag, 1994, pp 75-88

[DUI94] Duin, Heiko & Kiessling, Silvia "CRIMP for Windows 3.1 User Manual", BIBA, 1994

[DUI95] Duin, H. "Object-Oriented Scenario Management for Simulation Models" To be presented on: IMACS European Simulation Meeting on Simulation Tools and Applications, 28-30 August, 1995, Gyor, Hungary

[EUR93] Eurostat, "Enterprises in Europe", Second Report, Commission of the European Communities

[GOR68] Gordon, T. J. , Hayward, H. "Initial Experiments with the Cross Impact Matrix Method of Forecasting", Futures, December 1968, S. 100-116

[HAM91a] Hämmäinen, Heikki "Form-based Approach to Distributed Co-operative Work" PhD Thesis, Department of Computer Science, Helsinki University of Technology, 1991

[HAM91b] Hämmäinen, Heikki & Condon, Chris "Form and Room: Metaphors for Groupware", *proc* COCS '91, ACM Conference on Organizational Computer Systems, Atlanta, Georgia, 5-8 Nov 1991

[HEL77] Helmer, O. "Problems in Futures Research: Delphi and Causal Cross-Impact Analysis", Futures, February 1977, S. 17-31

[HUT89] Hutchins, E. "Metaphors for Interface Design" in Taylor, M.M, Neél, F. & Bouwhuis, D.G. (eds) "The Structure of Multimodal Dialogues, Elsevier: North Holand, 1989, pp11-28

[HUT93] Hutchinson, Chris & Rosenberg, Duska "Cooperation and Conflict in Knowledge-Intensive Computer Supported Cooperative Work" in "CSCW: Cooperation or Conflict?" Springer-Verlag, London 1993

[KAN72] Kane, J. "A Primer for a New Cross-Impact Language - KSIM", Technological Forecasting and Social Changes 4, 1972, S. 129-142

[KRA94] Krauth, J. , Hamacher, B. , Duin, H. "A Set of Tools for Rough Enterprise Modelling", *Proc* CISS (First Joint Conference of Simulation Societies), 1994

[LAU90] Laurel, Brenda "Interface Agents: Metaphors with Character" in B. Laurel (ed.) The Art of Human-Computer Interface Design Addison-Wesley 1990

[SIN94] Sinnigen, Marita; Cools, Jan-Pieter & Sampera, Jaume "URSA Deliverable 5" RACE Project R2091, 1994