Taylor & Francis
Taylor & Francis Group

Iterative process of design and evaluation of icons for interactive TV menu

DOKSHIN LIM†*, CAROLE BOUCHARD‡ and AMÉZIANE AOUSSAT§

†Mobile Human Interface Design Team, Samsung TN Division, Seoul, Korea
‡Product Design Laboratory, Ecole Nationale Supérieure d'Arts et Métiers (ENSAM), 75013 Paris, France
§Ecole Nationale Supérieure d'Arts et Métiers (ENSAM), 75013 Paris, France

This paper shows an iterative process of design and evaluation of icons for future interactive TV services. In doing the RNRT (French National Network of Research in Telecommunications) iTV project, we tried to generate icons easy to identify, associate and memorise for 32 categories and services of our iTV system.

Through an iterative process, the Multiple Index Approach was applied until an acceptable icon set was achieved. In addition to existing evaluation criteria such as the intuitiveness, associativeness, preference and suitability with subjective certainty of users, we emphasised the importance of the learnability measured by recall tests.

As a conclusion, we propose a methodology of icon design and evaluation for information appliances that integrate unfamiliar features with common users.

Keywords: Icon usability; User interface design; Interactive TV; Design process

1. Introduction

1.1 *RNRT iTV project*

In the iTV domain, technologies are being developed so vigorously that the definition changes every day. We adopt the definition of Gawlinski (2003) that describes the iTV as anything that lets the television viewer or viewers and the people making the television channel, program or service, engage in a dialogue.

In contrast to the technological innovation, there are few studies questioning which services are really useful from the users' perspective and how to design them to be usable by the majority of users. The goal of the RNRT (French National Research Network in Telecommunications) Interactive TV project was then set to anticipate future iTV services centred on common users.

Our target user population was common users. More specifically, we chose to satisfy the needs of 'TV Families' and 'Telly Traditionalists' in the iTV domain (Freeman and Lessiter 2003) that lie between early adopters and laggards. 'TV family' people typically live as part of young working families. Their confidence with technology may come from their experience with it in their workplace. Their considerable TV viewing hours and the fact that they are more likely than average to have all of the latest home entertainments suggest that they are TV fans living in modern families. The influence of their children is likely to be a major motivator in adopting digital television (DTV). Telly traditionalists are avid viewers and are overrepresented in all of the higher TV viewing categories. These people may be retired or semi-retired who fill their time by watching TV now that their children have flown the nest. '*TV Families*' (they represent 30% of the population in the UK) is chosen because they are not technology enthusiasts but have great interests in interactivity. '*Telly Traditionalists*' (they represent 20% of the population in the UK) is chosen because ease-of-use is a major concern when they become interested in interactive services.

The project integrated specialists in multimodal interaction: psychologists, product designers, computer programmers, and a representative of a technology provider as well as usability engineers. As a result of the project, we proposed a model of future iTV including 90 services that

*Corresponding author. Email: dos.lim@samsung.com

we consider useful for common users. The services are also organised in a usable way. All services are accessible by navigating a two-depth menu organised on the basis of the results obtained by card sorting with 18 users representing the two target categories.

The present paper shows an iterative process of icon design within the RNRT iTV project.

1.2 *Icon design for iTV*

When the menu structure was outlined, we decided to add icons to each menu item. The icon denotes any small raster image appearing in a graphic user interface (GUI) display. Well-designed icons are known to allow speed and directness with which items can be recognised (Mullet and Sano 1995).

At the very early phase of design, we integrated existing icon design guidelines such as Guidelines to Standardizers of ICT products and services in the CEN ICT domain (Gill 2005) and Multimedia User Interfaces for Interactive Systems and TV (MUSIST). We also decided to incorporate text labels to the icons because, despite the speed and directness, using only icons without labels can leave the system with a much poorer performance than when using texts only or an icon-text combination (Wiedenbeck 1999).

Furthermore, in order to avoid the deficiencies of the current practice of icon design such as 'the designer-orientedness, the beauty-orientedness and the monolithism' (Chen 2003), we mainly focused on the integration of the user-centred design process. User-centred design (ISO 13407) (Shneiderman 1998) involves users as much as possible in the process. It is then highly interactive and involves much testing and revisions. These activities are iterated until the objectives are satisfied.

Our iterative process of design and evaluation involved users in three steps until a set of wholly new icons replacing precedented ones was sufficiently small (see figure 1).

2. Generation of first icon set

A designer in the multidisciplinary iTV project team was charged with the generation of 32 icons for main and sub-menu items adapted to our target user population.

When popular metaphors or standards exist already, we adopted the existing idea and adjusted it to the iTV use context. We referred to existing standards (European Telecommunications Standards Institute (ETSI 1993 and 1994) (MUSIST) and universal public signs (Yang *et al.* 1980) (see figure 2). We tried not to copy icons that are familiar to frequent PC users. However, for services mainly known for their functionality in the PC environment, we adopted existing icons for PC environment (e.g. 'Print' and 'Help' functions and 'Mail'; see figure 2).
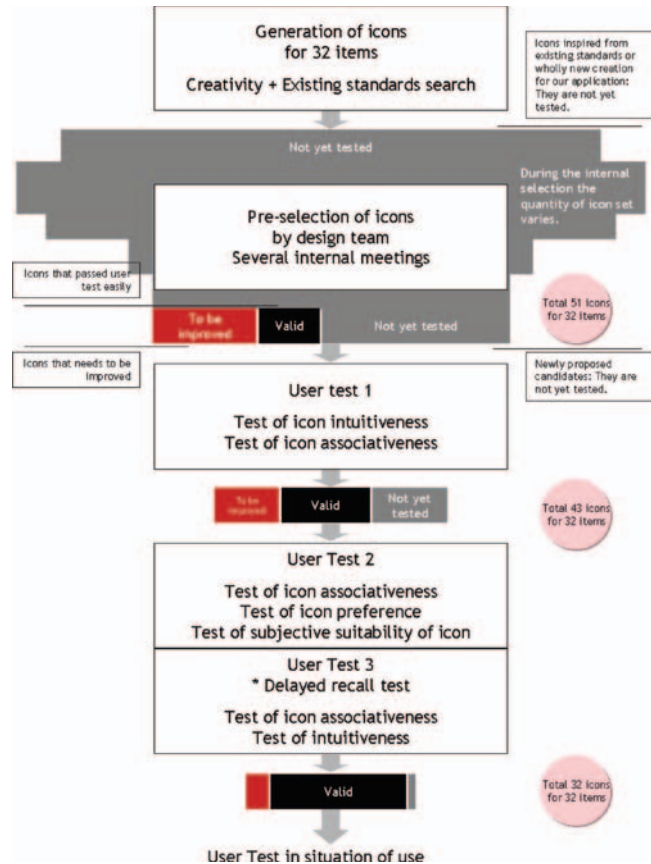

Figure 1. Iterative process of design and evaluation of icons for iTV services.

For services such as 'Going out', 'Cyber sightseeing' or 'Multimedia encyclopedia' for which there isn't yet well-known metaphors, rapid brainstorming sessions were organised involving project team members. Ideas expressed in words and pictures were converged into picture metaphors on the basis of which the icons were designed.

Icons proposed in these ways (search for existing standards, creativity) were at first gone through an internal preselection.

3. Preselection of icons by the design team

The designer tried to propose several alternatives for each item. The first icon set was presented to our multidisciplinary design team and passed by a brief preselection process. All project team members participated by votes. The ease of identification, familiarity, user friendliness and attractiveness were major criteria used at this stage.

The quantity of the icon set went through ups and downs. Finally, 51 icons (among 32 referents, 19 had two alternatives) were internally selected. The selected icons went through the 3-step user tests described in sections 4 and 5.

Existing icons, public signs, symbols	Intended Meaning	Icon iTV	Referent & Intended Meaning
	ETSI standard Videophone / Telephone: In videophony, for switching between videophone (sound and picture) and telephone (sound only modes).	Visiophone	Videophone
	Public sign Museum	Visite touristique	Cyber sightseeing
	Windows Print	Imprimer	Print
	Windows Help	Aide	Help
	Windows e-mail	Messagerie	Mail
	Web www.amazon.com Shopping cart	Boutique achat	Home shopping

Figure 2. Example of icons using existing standards.

4. User testing method

The Multiple Index Approach (MIA) had been developed, tested and employed in the context of a European Telecommunications Standards Institute (ETSI) study on pictograms for basic videophone functions (see also Böcker 1993). Piamonte *et al.* (1999) and Piamonte (2000) found that this method had been found to be suitable as a general testing method for pictograms from all areas. Our test method relied mainly on this MIA.

The main purpose of evaluation was to collect data to help select the best-suited icons from a number of icon proposals (for some items, two candidates were proposed). Three tests (User tests 1, 2 and 3) were designed. Each

step aimed to measure a composition of the following criteria:

1. Intuitiveness: Icon intuitiveness is a measure employed in the work of Nielson and Sano (1994) and Nielsen (1995). Our icons will normally be presented with text labels. However, it would be ideal if the intended meaning of an icon could be recognised at a glance even when it is not associated with a text label. In the icon intuitiveness test, participants were shown icons one by one without their labels and were asked to describe what each icon was supposed to represent.
2. Associativeness: One referent (name and description of an item) is presented on a questionnaire. A group

of icons is presented on the TV screen. The participant's task is to choose the appropriate icon for the referent in question. His choice can be analysed in terms of the hit rate, false alarms and missing values. In addition, subjective certainty is required for each rating. The following is the description of each index.

- Hit rate: Score of correct associations between a referent and icon.
- False alarms: The number of times a wrong icon is associated with a given referent.
- Missing values: The number of instances a participant does not answer a question presumably because he does not know the answer.
- Subjective certainty: This index indicates how certain the participants feel in their identification and association. For example, if the user is extremely uncertain about a choice on a test of icon associativeness, he may decide, in a real situation, not to use it at all.

3. Preference: The participants indicate which candidate icon for one referent best represents the referent in question. Calculating the p-value of Student's t-test will show if there exists a significant difference between two candidates.

4. Subjective suitability: The subjective suitability indicates how certain the user is in the given association in question of an icon and a referent. The multiple scale dimension analysis will allow dissociating icons in groups, depending on their mean of subjective suitability. The groups with low suitability should seriously be redesigned.

Each test needed to use different orders of presenting icons to counterbalance learning effects. Icons were displayed on the TV screen (Sony 70 cm in diagonal standard cathode screen). The distance between the television screen and the users' faces was set to 3 m, within the range Gawlinski recommended (2003).[1]

Each step needed a different composition of the above selection criteria (see figure 1). Participants who fell in the two target categories were recruited in ENSAM.[2]

5. Results

5.1 *User test 1 (intuitiveness/associativeness)*

Four students (22–26 years old) participated in the test.[3] We showed icons, one by one, without their labels.

Participants were asked to state their best guess as to what the icon was supposed to represent. This test assessed the degree to which the graphic chosen for the icon represented the intended concept. Subjective uncertainty enabled convictions on making decisions.

Some icons passed the test easily, with most users guessing the intended meaning or at least guessing something that was very close and would not be misleading in the context of the full system. For example, only one user explicitly used the words 'See and Listen (in French "Voir et écouter")' to describe our 'Watch/Listen ("Regarder/écouter")' icon, but descriptions like 'Audiovisual ("Audiovisuel")' and 'Image and sound ("Image et son")' actually showed that the general idea was understood. In cases like this, we were satisfied that users would understand the icon when it was combined with its label in the full system. So we did not feel a need to change it, especially given that we had planned several additional user tests that would reveal any hidden problems. For these icons, only minor adjustments were made (see figure 3).

In other cases, users did not guess the correct meaning of an icon exactly but we still decided to keep it. 'Learn ("Apprendre")' was one such example, because we thought that it could be easily accepted when it is presented with the text label. The underlying problem was that sub-categories were not coherent to the main category icon. So, we reused the 'teacher' image to all sub-categories of the main menu 'Learn' (see figure 3).

With respect to the 'Going out ("Sorties")', users did not recognise the intended meaning. They identified the icon as 'Assisting a person at home' and 'Home automation appliance' (two subjects out of four had no idea). At this stage, the project team generated a wholly different concept – an icon using a chick coming out of the egg. Major change was also made to the icon of the main menu 'Practical life' ("Vie pratique"). A thumb recognised as 'Nice', 'OK' and 'More information' did not seem appropriate. It was redesigned as a Swiss knife (see figure 3).

The intuitiveness is the principal factor that defines the usability of an icon in an absolute condition. However, in real context of use, the correlation between coexisting icons should also be considered. According to the hit rate and false alarms, we were able to make decisions on some icon sets. For example, we discovered poor associativeness of three icons in the main menu 'Communicate': 'Visiophone', 'Dialogue on line' and 'Discussion forum' (all produced 3 errors out of 4). They were redesigned using ordinary smiling heads and empty smiling heads at the same time (see figure 3) because one of the distinguishing factors among the three services is if the correspondent is anonymous or not.

At the end of the User test 1, we obtained a set of 43 icons. Eighteen icons could already be considered to be able to pass easily. Fifteen wholly new icons were added either

---

[1] The distance between television viewers' faces and the television screen is typically 2–3.5 m (Gawlinski 2003).
[2] Ecole Nationale Supérieure d'Arts et Métiers.
[3] This type of test can be done by a small number of users, typically five (Nielsen 1995).

Intended Meaning	Before user test 1	Motif of change	After user test 1
Watch / Listen *Regarder / écouter*		Simplify contour lines	
Learn to cook *Cours interactif de cuisine*		Use consistent concept in the same category	
Learn language *Cours interactif de langue*		Use consistent concept in the same category	
Learn yoga *Cours interactif de yoga*		Use consistent concept in the same category	
Going out *Sortie*		Change concept itself	
Practical life *Vie pratique*		Propose an alternative in using wholly different concept	
Communicate *Communiquer*		Use consistent concept in the same category	
Visiophone *Visiophone*		Use consistent concept to differenciate	
Dialogue *Dialogue en direct*		Use consistent concept to differenciate	
Discussion Forum *Forum de discussion*		Use consistent concept to differenciate	

Figure 3. Example of changes made from the results of User test 1.

as new alternatives or replacements. Ten icons seemed to need improvements (see figure 4) but they were retained within the set in the absence of clearer icons with the hope that they would be easier to understand with their labels.

5.2 *User test 2 (icon associativeness/icon preference/ icon suitability)*

At this stage, the laborious icon intuitiveness test was replaced by the icon suitability test at the end. Instead of asking users to state what icons inspired them, referents were given with associated icons for users to evaluate the suitability among four choices (very bad, bad, good and very good).

An icon preference test was added in order to choose better icons when there were more than two alternatives for one referent (we had 10 referents having more than two alternatives).

Thirteen subjects (students or staff of the New Product Design Laboratory) whose age varied from 22 to 59

participated in the test. Tests were performed in a group of three or four subjects at a time.

By integrating the associativeness and the subjective certainty (see figure 5) and by using p-values of the significant difference in the mean of subjective preference, we were able to eliminate seven alternatives among 10 couples of icons.

We also found that with respect to the main category 'My garden ("Mon jardin")' missing values were high (6 and 7 out of 13 for two alternatives); because of the imaginative label, subjects strongly tended to think of gardening-related TV programs or shops. We renamed this category as 'Perso' and redesigned it as a drawer that would let people think that one can keep personal data in it.

Nine subjects (out of 13) chose the icon of 'Dialogue' when asked to find the icon for 'Discussion forum'. The category 'Communicate' leaves users to learn the difference between 'Visiophone', 'Dialogue' and 'Discussion forum'.

5.3 *User test 3 *Delayed session (icon associativeness/ icon intuitiveness)*

Nine subjects who had participated in the User test 2 were gathered again two weeks later. We asked them to

Intended Meaning	Before user test 1
My Garden (*Mon jardin*): This main category is kind of one's "secret garden" to stock one's personnal data (video, photo, text, ...) that not anyone can access to	
On demand (*A la demande*): This subcategory leads to display menus of on-demand services such as "Films on demand", "Music on demand", etc.	
TV program guide (*Guide TV*): This service allows users to consult interactive TV programs	
Live (*En direct*): Display a variety of information related to the program on	
Video supervision (*Vidéo surveillance*): This service allows TV viewers to supervise the house (Who knock the door, my son in the swimming pool, ...)	

Figure 4. Example of icons proved not easy to identify and to associate on User test 1.
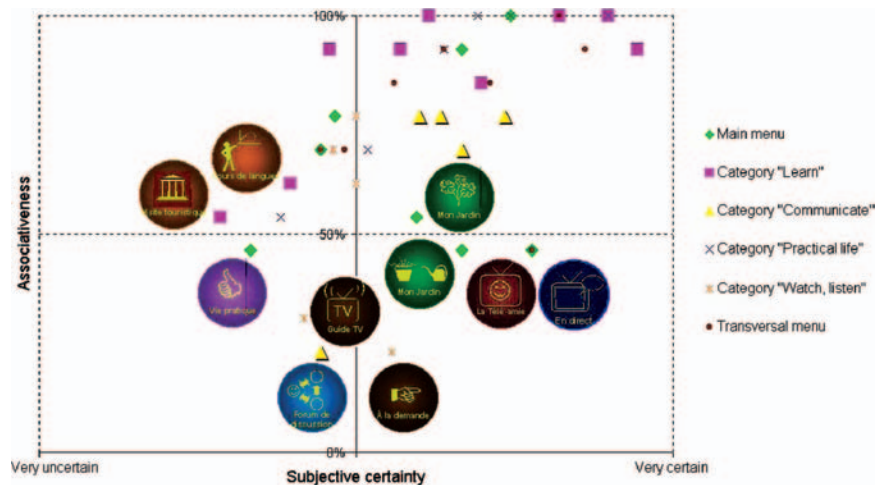

Figure 5. Associativeness and subjective certainty on User test 2.

associate two referents on each category of the icons. They were also asked to describe what would happen if they chose that item. The aim was to see if users were able to recall well the attribution of icons and services contents.

For icons such as 'Friendly TV ("TV amie")', 'Discussion Forum ("Forum de discussion")', 'On demand ("A la demande")' and 'Live ("En direct")', which had shown poorer intuitiveness and associativeness during previous tests, we found that most of the users associated them with their correct label as well as we intended (at least eight correct answers out of nine for all categories).

Nevertheless, by only seeing icons and labels, they had difficulty in remembering the exact functions the items were designed to do from the previous tests.

At the end of the User tests 2 and 3, we finally achieved 32 icons to use for a working prototype of iTV services.

Two icons were new propositions, not yet tested. One icon in the category 'Communicate' did not seem easily associable to its referent 'Discussion forum', but we continued to use it, hoping that by learning through several uses, users could memorise it. We could say that the other 29 icons were comprehensible and suitably represented the related items. The final icon set ready to be integrated into the working prototype is shown in figure 6.

5.4 *User test in the context of real use using a working prototype*

The limit of the above experiments is that we were not able to test icons in the real context of use because at that moment, our mock-up was not working. We could not see if users given real use tasks like 'Save or visualise your

Figure 6. Thirty-two icons of interactive TV services.

personal multimedia data' chose the intended icon 'My garden' (modified as 'Perso' on User test 2).

A final usability test followed in which the icons were shown as a part of the full user interface of the global system. Twenty-four adults and adolescents (six over 55 years old, 12 between 22 and 40 years old and six under 17 years old) participated in the test. Users were given scenarios of use and asked to 'think aloud' when they used the system to perform given tasks. The aim of the test was multiple: the usability of categories and hierarchy of services, using PDA remote control, contents of services as well as the usability of icons. The tested working prototype looked like figure 7.

We observed that at first users paid more attention to the text labels than the icons to use the system. For example, the main menu 'Perso' reminded 'Personalisation' or 'Personal configuration' rather than 'Personal data'. The icon of a drawer didn't orient users to the intended meaning. This main menu was finally renamed as 'My data ("Mes données")'.

When at the end of the test we asked users to recall icons, they were not able to draw exact concepts. Apart from some adolescents, subjects showed a tendency to draw

Figure 7. Screen of television and PDA remote control of the working prototype.

geometric forms (e.g. diagonal line, rectangle), but not precise concepts (e.g. Swiss knife, drawer) of an icon.

Did icons contribute to the ease-of-use of our iTV service system? We did not find the answer to this through the test because users depended more on text labels than on icons. Some exceptions were observed when users were navigating on the two categories: 'Watch/Listen' and 'Communicate'. In the former category, three unfamiliar items – 'Diffused programmes', 'On-demand' and 'Mediatheque' – needed users to figure out what they were for. The latter ('Communicate') has three services that caused confusion: 'Visiophone', 'Discussion Forum' and 'Dialogue'. With respect to these items, text labels were not sufficient to understand the contents of the items when users paid attention to icons. We observed that users tried to figure out the contents by analysing the differences in icon design.

Our experience led us to think that the major role of icons when users are not familiar with the system may be to bring a user-friendly environment of use. When users are familiar with the system after successive uses, we expect them to depend more and more on icons than on text labels, due to the speed and directness with which recognition and identification of icons take place. In that case, we can consider icon-only interfaces for expert users.

6. Conclusion

We used the Multiple Index Approach (MIA) for the iterative processes of the design and evaluation of icons structured in the menu. We showed that the choice of indexes should be varied throughout iterations, depending on the evolution of icon sets.

The initial user test (User test 1) needed most of the icon intuitiveness test and then the associativeness test. Spontaneous icon identification may seem laborious, but it would be better to do it at least once in the early phase of iteration, even with a small number of users, in order to get productive assessment from users. For further iterations, the spontaneous identification may be replaced by a simple closed questionnaire of subjective suitability. The associativeness test is well adapted for a navigation system structured in a hierarchical menu.

User test 2 was composed of the icon associativeness, preference and suitability tests. At this stage, by using the preference test, it was possible to eliminate some less powerful alternatives.

User test 3 aimed to verify, for suspicious icons, if they could be well associated and well memorised after some delay. With respect to today's information appliances, we emphasise the importance of the learning effect through time. Common users encounter more and more new products in everyday life. New features which do not have names that instantly suggest their function need to

be learned and memorised. In this context, even if a new feature (icon, contents, etc.) shows poor performance at first because it isn't familiar to users. However, if it can be easily learned and accepted, it can be considered as a well-designed one. The importance lies on the rate of improvement, not on the performance at first glance.

The main contribution of the present work is that we improved the multiple index approach of icon testing by adding a recall test session. We deliberately conducted the recall test two weeks after the first usage test. Further research should define optimal length of delay to perform such kind of recall test.

We briefly explained the final test with a working prototype during which we once again tested the icons. We found that the major role of icons was rather to make the system more user-friendly than to make it easier to use the system, because users didn't pay particular attention to icons. It seemed that users relied more on text labels than on icons to figure out items. We also found that users began to pay attention to icons in order to figure out the service contents by analysing icons when the text labels were not sufficient.

In this paper, we showed a way of iterative design and evaluation of icons. User tests of icons may begin very early in the process, even when any working prototype is not yet ready (generally when the designer begins to design icons, the working prototype doesn't exist yet). When major corrections were made in early phases, further development became simpler (e.g. less charge for the programmer). However, a final test of icons in the real context of use with a working prototype is also necessary.

Acknowledgements

Support

References

Böcker, M., 1993, A multiple index approach for the evaluation of pictograms. *Proceedings of the 14th International Symposium on Human Factors in Telecommunications*, **1,** pp. 73–85, Heidelberg, Germany.

Chen, P., 2003, Toward a methodology of graphical icon design. *Proceedings of the IEEE Fifth International Symposium on Multimedia Software Engineering*, ISMSE'03.

ETSI, 1993, Human Factors: The Multiple Index Approach (MIA) for the evaluation of pictograms. TC-HF ETR 070 Reference: DE/HF-1010B, European Telecommunications Standards Institute.

ETSI, 1994, Human Factors: Pictograms for point-to-point videophony. TC-HF ETS 300 375 Reference: DE/HF-01010-A, European Telecommunications Standards Institute.

Freeman, J. and Lessiter, J., 2003, Easy TV 2002, Independent Television Commission (ITC) Research Report.

Galwinkski, M., 2003, *Interactive Television Production* (London: Focal Press).

ISO 13407 Human centred design processes for interactive systems. Available online at: www.usabilitynet.org/tools/13407stds.htm (accessed 18 March 2005).

Mullet, K. and Sano, D., 1995, *Designing visual interfaces* (California: Prentice Hall PTR).

Musist (Multimedia User Interfaces For Interactive Systems and TV) GSM, Stuttgart. Available online at: www.gsm.de/musist/mstyle.htm (accessed 31 March 2004).

Nielsen, J., 1993, *Usability Engineering* (New York: Academic Press).

Nielsen, J., 1995, Icon Usability. Available online at: www.useit.com/papers/sun/icons.html (accessed 18 March 2005).

Nielsen, J. and Sano, D., 1994, Design of SunWeb – Sun Microsystems' Intranet. Available online at: www.useit.com/papers/sunweb/ (accessed 18 March 2005).

Piamonte, D.P.T., 2000, 'Using Multiple Performance Parameters in Testing Small Graphical Symbols'. Unpublished doctoral dissertation, Institutionen för Arbetsvetenskap Avdelningen för Industriell ergonomi, Luleå, Sweden.

Piamonte, D.P.T., Ohlsson, K. and Abeysekera, J., 1999, *Final project report*: *Across the seas – A user-based evaluation of candidate telecommunication icons*. (Sponsored by the Communications Research Board of Sweden, KFB) (Luleå: Luleå University Pess).

Shneiderman, B., 1998, *Designing the User Interface: Strategies for Human-Computer Interaction*, 3rd edition (Reading, MA: Addison-Wesley).

Gill, J., 2005, Pictograms, icons and symbols. Tiresias Guidelines to Standardizers of ICT products and services in the CEN ICT domain. Available online at: www.tiresias.org/guidelines/pictograms.htm (accessed 18 March 2005).

Wiedenbeck, S., 1999, The use of icons and labels in an end user application program: an empirical study of learning and retention. *Behaviour & Information Technology*, **18,** pp. 68–82.

Yang, L., Kumekawa, M. and Hotta, A., 1980, A study of universal design for public signs. *The Journal of the Asian Design International Conference*, **1.** Available on CD-ROM.