2000:02

LULEÅ TEKNISKA UNIVERSITET

DOCTORAL THESIS

# Using Multiple Performance Parameters in Testing Small Graphical Symbols

Dominic Paul T. Piamonte

**Doctoral thesis**

Institutionen för Arbetsvetenskap
Avdelningen för Industriell ergonomi

2000:02 • ISSN: 1402-1544 • ISRN: LTU-DT--00/02--SE

# ABSTRACT

Graphical symbols have become a mainstay in interfaces of most modern-day industrial machines and personal tools with their increased computing power and ever-shrinking sizes. Their use has transcended national boundaries and is seen all over the world. However, only lately have studies started to gather momentum, attempting to analyse how they are really understood by people of different backgrounds and needs, and how helpful can they really be in making man-machine interactions better. Amidst this situation, this doctoral study was conceived. The main objective of the study was to analyse the understanding of small graphical symbols (pictograms and icons) among different potential user groups around the world. Different test indices or parameters were used to examine how symbol understanding would differ or be similar across different study groups. An initial study done in the Philippines and Sweden using three sets of twenty-three videophone symbols designed in Japan, USA and England preceded the main study. The main study utilised three sets of videophone symbols and were based on studies done by the Human Factors Technical Committee (HFTC) of the European Telecommunication Standards Institute (ETSI). Subjects from Indonesia, Malaysia, Philippines, Thailand, Sri Lanka, as well as subjects from Finland, Sweden and USA participated in the study. Based on the results and learnings from the initial study, as well as considering the recommendations of the ETSI studies, more measurement parameters were used. Thus, four tests were used, namely: a) spontaneous identification or free recall, b) the cued response, c) semantic rating scales, and d) preference tests. Confidence judgements complemented the subjects' answers for the first two parts. These four tests yielded 10 test parameters that were used in comparing symbol understanding of the different country groups. These parameters were: (1-4) spontaneous identification and cued response hit rates and certainty ratings (confidence judgements); 5) misses; 6) false alarms or confusions, 7) missing values; 8) individual symbol preference, 9) symbol set preference; and 10) semantic rating scales. Results from the spontaneous identification tests in all countries revealed very poor identification of most symbols in contrast to the cued response test results. The first test simulates the users' initial encounters with the symbols. Barely recognising what the symbols meant implied the need to either redesign the symbols or to ensure adequate opportunities for familiarising and educating the prospective users with the new symbols. The subjective certainty scores helped in studying the level of confidence of the answers by the subjects. Furthermore, the studies revealed that symbols could be easily recognised (high hit rates) but also confused as representing another (wrong) function at the same time. These were the "false alarms," instances when the symbols were chosen under the wrong referent contexts. The "missing values" were likewise important since they indicated situations when respondents either did not know the answer or thought that none among the symbols were comprehensible or representative of a desired function. The preference tests pertained to aesthetics of the symbols individually and as a set. In turn, the SDT scores revealed that symbols could have different connotative meanings in relation to the functions they were intended to represent. Between countries, Asian subjects performed comparably well with the European and American subjects, but usually at the expense of more errors and confusions. Tests with the elderly revealed other important issues in symbols evaluation. For example, they failed to perform the free recall and semantic tests claiming they were either too difficult or too complicated, strongly suggesting the need to have simpler and more practical tests for the elderly. In all the results implied the gravity of the problem that can occur if graphical symbols are introduced arbitrarily without considering and involving the prospective users at the earliest design stages possible.

# TABLE OF CONTENTS

## PREFACE

The term technology generally pertains to the processes human beings use to design and utilise tools and machines to increase their control and understanding of the material environment. The word technology comes from two Greek words: *tekhne*, which refers to an art or craft, and *logia*, meaning an area of study. Thus, technology literally means the study, or science, of crafting (Encarta, 1997). It is often stated that technology is an essential condition of advanced, industrial civilisation, that technological development is supposed to lead to the improvement of quality of human life - to help us do things better, faster, safer and more efficiently. However, it is also claimed that the rate of technological change has developed its own momentum in recent centuries. Innovations now seem to appear at a rate that increases geometrically, without respect to geographical limits or political systems. These innovations tend to transform traditional cultural systems, frequently with unexpected social consequences. Specifically, the temptation to forget the human being as the centre and reason for such a development has indeed become a problem. Thus, technology can be also conceived as "both a creative and a destructive process, " (Encarta, 1997).

Ergonomics, despite its relatively young age of existence as an accepted and essential scientific discipline, has been at the forefront of putting man where he belongs amidst all the technology - the maker and controller of his work. Ergonomics involves the study of the human being's physical and psychosocial traits, capacities, and limitations. It is an applied science utilising such knowledge to match the person to his/her tasks and tools. Thus, in its broadest sense, ergonomics is indeed "the science and practice of fitting the task (technology's tools, equipment, and interfaces) to the human being" (Kroemer, 1997).

This doctoral thesis is just a humble example of a striving ergonomist's attempt to contribute to a little knowledge and insight on the ergonomics of man-machine interfaces. It is an area reflective of the current rapid pace of technological development. Indeed, the use of graphical symbols personifies the current economic and technological globalisation - they are found everywhere and are used by more and more people. However, a question still remains - are they always understood the way they are intended to be understood? This paper, in essence, attempts to answer this question.

## ACKNOWLEDGEMENTS

It is quite easy to lay personal claim to an invention or discovery because of one's hardship and perseverance. It is likewise easy to contend that knowledge can be obtained by oneself without anyone's help, if he or she perseveres hard enough in studying and doing research alone. However, *I sincerely believe* that a person's knowledge and skills encompass a collection of the experiences, knowledge, and expertise derived from countless people a person has come across during his or her lifetime. Thus, one cannot be boastful or full of pride when knowledge is gained, honed, and broadened. Instead, one should become all the more appreciative and indebted to all the people behind such pursuit of knowledge. In my case, there are countless of them. Below are those to whom I am indebted throughout my research work. This doctoral work is as much theirs as it is mine. My thanks and prayers to:

- To *Prof. Kjell Ohlsson*, my amiable and rock-solid supporter and supervisor, always with time to share and knowledge to impart
- To *Dr. John "Abey" Abeysekera*, the person who really started this journey for me way back during the M.Sc. days - the one who has never ceased to have faith and belief in me and in what we have done
- To *Prof. Houshang Shahnavaz*, my revered "sensei", always a source of wisdom and guidance
- To *Dr. Martin Böcker*, who never ceased to lend a helping hand wherever he may be
- To *My Colleagues and Classmates*, from M.Sc. days to Ph.D. days, the shared hardships, triumphs, tears and laughter
- To *All the Subjects* in the studies, for their willingness to participate and share their time and effort
- To *Ingrid and Anita*, ever present to assist and make every day's work not just a bit easier but also more rewarding
- To *My Parents*, for their boundless faith and ever-powerful prayers
- To *My Family*, we all did this together
- And, to *The Lord Almighty*, my infinite source of inner strength and inspiration

**MARAMING SALAMAT PO!**

**TACK SÅ MYCKET !**

## PAPERS PRODUCED FROM THE STUDIES

This doctoral thesis is based on the following refereed journal and conference papers as well as some unpublished data:

1. Piamonte, D.P.T., Abeysekera, J.D.A. & Ohlsson, K. (1996). A user-based evaluation of telecommunication icons: some cross-cultural issues. In L.K. Yong, L., Herman, Y.K. Leung & J. Moyes (Eds.), *Human Factors of IT: Enhancing Productivity and Quality of Life, Proceedings of the First Asia-Pacific Conference on Computer-Human Interface* (pp. 145-150). Singapore: Information Technology Institute.

2. Piamonte, D.P.T., Ohlsson, K., and Abeysekera, J.D.A. (2000). Evaluating public graphic symbols using multiple test parameters. (for publication) *International Journal on Occupational Ergonomics.*

3. Piamonte, D.P.T. (1999). The relevance of error analysis in graphical symbols evaluation. *International Journal of Occupational Safety and Ergonomics, 5(4),* 513-526.

4. Piamonte, D.P.T., Abeysekera, J.D.A. & Ohlsson, K. (1999). Testing videophone graphical symbols in Southeast Asia. In H.J. Bullinger & J. Ziegler (Eds.), *Human-Computer Interaction: Ergonomics and User Interfaces: Vol. 1* (pp. 793-797). London: Lawrence Erblaum Associates.

5. Piamonte, D.P.T., Ohlsson, K., and Abeysekera, J.D.A. (1999). An empirical evaluation of videophone symbols: an international study. In H.J. Bullinger & J. Ziegler (Eds.), *Human-Computer Interaction: Ergonomics and User Interfaces: Vol. 1 (pp. 798-801).* London: Lawrence Erblaum Associates.

6. Piamonte, D.P.T., Abeysekera, J.D.A., and Ohlsson, K. (2000). Understanding small graphical symbols: a cross-cultural study (accepted for publication). *International Journal of Industrial Ergonomics.* Amsterdam: Elsevier Science Publishers B.V.

7. Piamonte, D.P.T., Ohlsson, K., and Abeysekera, J.D.A. (2000). On the merits of using multiple indices in evaluating small graphical symbols (accepted for paper presentation). In *International Ergonomics Association Triennial Conference – IEA 2000,* San Diego, California.

# SUMMARY OF PAPER 1

The first paper was a study aimed to evaluate graphical symbols (icons) designed in Japan and Europe (industrialised countries) with subjects coming from the Philippines (an industrially developing country). The 23 icon referents and the 69 icons used by Tudor (1994) were selected. Sixty-two of the icons were designed in England. The remaining 7 icons were designed in Japan. The method was the questionnaire type of testing. Each questionnaire had two parts: recognition/subjective certainty tests, and semantic differential. Questionnaires were produced per icon family with the icons and their referents presented in random orders. The semantic differential was composed of 6 bipolar adjectives in 7 point scales also arranged in random orders for each of the icons or referents. After presenting the study's objectives and explaining the icons, the subjects first underwent the recognition and certainty tests. Icons and referents were randomly presented and to be matched correctly. After one week, the semantic differential testing was performed. The icons were again presented. The same tasks were done for the three icon families. Thus, each subject performed 3 recognition/subjective certainty tests, and 3 semantic differential tests, one for each icon family.

The results showed that for the 23 referents represented by 23 icon families, only 5 of the 23 referents had a mean recognition percentage of at least 60 percent. Further, the icons with high mean recognition percentages also exhibited high mean rating scores for the certainty tests. Rank order correlations were also at least 0.700 for these icons. Icons with low recognition means had low certainty levels and semantic scores. For the SD tests, results showed that correlations were high between the semantic scales and the percent mean of correct recognition regardless of icon families.

Compared to American subjects used in another study, the Filipino subjects fared better with the Japanese-designed icons. Such difference may be a reflection of a Philippine social condition where majority of current computerised and electronic appliances in the Philippines is imported from Japan. Americans on the other hand are obviously more akin to be exposed to products they have made. Although USA still ranks as the major trading partner of the Philippines followed by Japan, the latter ranks first in terms of the computer and telecom products marketed and used locally. Icons are also commonly found in these products. This renders the view that cultural familiarity and meaning of icons can affect their usability when applied as interface tools in new or modern technologies.

The SD results showed that SD testing could be a valid index of how messages of symbols like the icons are perceived and understood. It is simple, inexpensive but reliable. Admittedly though, there are limitations in this pilot study. A distinct group of potential users (educated and computer literate) was used as subjects. Future studies will need to utilise varying groups of potential users (students, professionals, experts, novices, etc.) across different cultures of Asia.

# SUMMARY OF PAPER 2

The next five papers, including this one used three sets of graphical symbols representing seven videophone functions or referents designed and tested in the West (Europe). Here, the respondents were from Indonesia, Malaysia, Philippines and Thailand. More tests were likewise used and were based on the ETSI studies by Böcker (1993). In the spontaneous identification or free recall test (Part One), each page contained one set of the seven videophone symbols, subjects would write what videophone function they thought was represented by each symbol. They were also asked to rate the level of certainty for each of their answers using a seven-point rating scale. In the cued response test (Part Two), the subjects first read a referent and its description then they chose one symbol from a set of seven symbols which best represented the referent in question. They were again asked to rate their certainties for their answers using the 7-point rating scales. For the last part, the subjects would choose seven symbols to represent the seven referents and one symbol set they preferred each subject group to minimise order and learning effects. In all, the test parameters studied were hit rates (mean percentages of correct identification/recognition in Parts One and Two), subjective certainty ratings, missing values, confusions, and symbol and symbol set preferences.

Overall, Set 1's symbols (recommended ETSI symbols) had the highest hit rates compared to Sets 2 and 3 in almost all of the 7 referents or functions tested. Taking between tests as another main effect, the hit rates in the cued response tests (Part Two) were significantly higher than in the spontaneous identification tests (Part One) for all the 7 referents. Symbols with high hit rates exhibited high certainty ratings as well - again, mostly symbols from Sets 1and 3. Confusions were instances when a symbol was associated to the wrong referents. The different countries were quite similar in the confusions of the symbols – easily identifying familiar symbols ("camera" and "document camera") and confusing them with other functions. Further, the symbols of Set 1 were the preferred symbols by at least 50% of the subjects for most of the referents.

Part One's results showed that graphic symbols could be quite difficult to understand or comprehend for the Asian subjects based on the very low hit rates. However, the high improvements in hit rates in Part Two showed that with proper instructions and cues, Asian subjects would not have major problems recognising graphic symbols. Certainty levels and missing values could help further qualify the levels of hit rates attained by the symbols. False alarms or confusions were demonstrated to represent the degree of confusing a symbol to another symbol presented at the same time. In all, multiple test parameters were useful in identifying not only the best symbols to use but also other potential problems in using symbols among different countries, which in turn may be helpful in considering needed changes or improvements.

## SUMMARY OF PAPER 3

If symbols are intended as parts of interfaces of devices for international use or for standardisation, different tests are usually needed. The previous study (Paper 2) had already partially shown this. This paper would concentrate on the so-called non-hit parameters. Two hundred forty university students and employees and professionals (127 males and 113 females) from small and large companies from five Asian countries participated (Indonesia, Malaysia, Philippines, Thailand and Sri Lanka). Again, the stimuli were three sets of symbols representing seven referents or functions of a videophone for a total of twenty-one symbols used in the previous paper. Both referents and symbols were based on the ETSI study headed by Böcker (1993). The questionnaires were divided into three parts: symbol identification/subjective certainty tests, symbol association/subjective certainty tests, and symbol and set preferences.

When ISO 9186 and ANSI Z535.3 comprehension levels were considered through the hit rates, most of the tested videophone symbols performed poorly. Combining country results, only seven of the 21 symbols reached ISO's required comprehension level of at least 67%. Furthermore, the wrong answers during the cued response tests were analysed using confusion matrices. Symmetric confusions usually suggest visual or conceptual similarities. Asymmetric confusions occur when subjects simply chose the wrong symbol for a given referent. Across symbol sets, "selfview" and "videophone" symbols had the highest instances of being mistaken as representing other functions. More differences were noted when the countries were compared based on symbol confusions. The results also showed that confusions could be useful indicators of the suitability of the symbol and may even complement the hit rates. For example, the videophone symbols of Sets 1 and 3 had very similar hit rates among the countries. With regards to the microphone symbols wherein subjects from the five countries had similar hit rates also, Set 3's version performed best for Indonesia, Malaysia and Sri Lanka (lower symbol confusions).

Determination of the hit rates gives the level of association of the symbol to its intended referent or function. However, when symbols appear simultaneously, the hit rates cannot reveal the dynamics of how each symbol can affect the other's level comprehension and association to its referent. Confusions are instances when symbols are selected under the wrong referent contexts. Together with hit rates, they can give an idea on a symbol's distinctness from other symbols under different referent contexts as well as the different patterns of their interactions in different groups. Other parameters also exist, which were not discussed here but could also be helpful in evaluating graphical symbols.

## SUMMARY OF PAPER 4

This paper was a summary of the study on videophone symbols using both objective and subjective parameters with respondents from the five Southeast Asian countries. It used the same methods based on ETSI studies by Böcker (1993) with some modifications (i.e. inclusion of the spontaneous identification tests).

Between tests, correct responses in the cued response tests (Part Two) were significantly higher than in the spontaneous identification tests (Part One), except for the symbols for "microphone". Furthermore, only the symbols (of Sets 1 and 3) for referents "microphone" and "videophone" garnered hit rates of at least 67% in Part One. Altogether in both tests and across all symbol sets, the highest hit rates were for "camera", "microphone" and "videophone" while the "document camera", "handsfree", "selfview" and "still picture" had the lowest hit rates in all countries. Comparing countries, there were significant differences in hit rates for symbols for "camera", document camera", "handsfree". In these symbols, Sri Lankan subjects generally had the lowest hits compared to the other four countries. No significant differences were noted when subjects were compared based on occupation (student vs. employees/professionals).

Tabulations made on the rates of confusions showed more differences between countries. For example, while Thailand had the most instances of confusing Set 1's symbol for "camera" than the other four countries, it had the lowest instances of confusing Set 1's videophone symbol for other functions. Confusions are very useful indicators of the suitability of the symbol and may even complement the hit rates. For example, with regards to the microphone symbols wherein subjects from the five countries had similar hit rates also, Set 3's version performed best for Indonesia, Malaysia and Sri Lanka (lower symbol confusions). Lastly, missing values were instances when some of the subjects gave no response or answers during the cued response tests. Between countries, Thailand had the most instances of missing values among all countries. Sri Lanka was the opposite with no missing values in the cued response test. The results in the spontaneous identification and cued response tests in testing videophone symbols showed very low rates of correct identification and even correct association of the symbols across all countries. Differences between countries lie mostly in the levels of confusing the symbols from each other and their missing values.

# SUMMARY OF PAPER 5

The previous papers dealt with the study on videophone symbols using subjects from five Southeast Asian countries. This paper would discuss the part of the research involving U.S. and Finnish elderly subjects and comparing them with the Asian results. The same materials (symbols and questionnaires) and method used in the previous papers were used in this study. Further, aside from the spontaneous identification and cued response tests, symbol and set preferences were also examine. This meant that the subjects had to choose symbols and set they thought best represent the referents individually and as a set, respectively. In all, the subjects would choose seven symbols to represent the seven referents and one symbol set they preferred most.

The still picture symbols failed to score hits above 67% in both Part One and Part Two. In contrast, the symbols for camera, microphone and videophone had high recognition rates regardless of test type and set. Regarding preferences, the US subjects preferred five of seven of Set 1's symbols. These results were similar to the Southeast Asian study. One important factor that leads subjects or potential users to their choice of symbols is easy recognition. In some parts of Northern Finland, videophone-based services for the elderly are being tested. This was one of the major reasons why the current study was also tested among a group of elderly subjects. Compared to the subjects from Southeast Asia and even the USA, the Finnish elderly failed to do Part One. Regarding Part Two, and when compared to younger subjects (from Asia and the US), the elderly subjects had much lower hit rates in all symbols of the three sets used. The trend of demographic evolution shows that the European population, in particular, is one that is becoming old. This trend will have implications on the social structure, where respect for an independent living should be taken into consideration. It should taking into account the new services and products just available or those under development, such as telealarms and teleservices to assist people at home or elsewhere. The results seemed to suggest that only by a conscious and active participation of the elderly in designing and developing modern products would their needs be truly considered and met.

As a whole, empirical tests using multiple indices are important to properly evaluate the symbols on their usability especially across different user groups. This can be exemplified by the elderly and US subject groups who chose symbol sets quite different from the Asians. Hit rate is an important parameter, but must be tested under different user scenarios to determine if the symbols indeed are understood with and without cues. At the same time hits rates alone are not enough. Subjective certainties, false alarms (confusions), preferences, and even missing values are equally useful factors needed to help make deeper analyses.

## SUMMARY OF PAPER 6

This study was part of the international project on evaluating telecommunication symbols. It would report on the parts involving European (Swedish) and US (American) subjects. The same test method used in previous papers was used in testing the same three sets of icons and pictograms representing seven referents or functions of the videophone. Each questionnaire had four parts: symbol identification tests, cued response tests, semantic differential ratings, and symbol and symbol set preferences.

Regarding identification and cued-response tests (Parts One and Two), Set 1 symbols generally had the higher hit rates in both countries. However, both countries performed poorly in the spontaneous identification tests (Part One). For example, subjects from both countries managed to attain the Organization for International Standardization's ISO 3864 (ISO, 1984) minimum correct recognition rate of 66.7% in only 5 of the 21 symbols tested. The poor results thus strongly suggest that without sufficient learning opportunities and learning aids, symbols may be harder to understand by almost anyone than previously thought.

In addition, when trying to study possible cultural influences in symbol evaluation, levels of correct identification are not enough. Compared to the American subjects, Swedish subjects tended to give lower certainty ratings especially during the spontaneous identification tests. This can suggest certain subject bias. Confusion matrices also helped to show other differences between the symbol sets and country groupings. Semantic scales have likewise been shown very useful in studying how symbols are perceived. The Swedish group as a whole gave lower mean semantic ratings compared to the American group. As mentioned above, the differences in patterns of ratings suggest that some subject bias occur between the two groups, which may be culturally-linked and can help in determining which aspects of symbol design and usage may be more helpful (ex. instructions, learning aids, etc.). Awareness of such subject bias and their implications are important on how one interprets the test results. Admittedly though, the above findings need to be analysed more deeply to be able to concretely transform such differences into practical applications of symbol design and testing.

# SUMMARY OF PAPER 7

This paper attempted to summarise the results of all the tests using the ETSI-recommended symbols for 7 videophone functions or referents and to develop some practical recommendations. The results of the tests in the different countries showed that the spontaneous identification test (Part One) had significantly lower hit rates compared with the hit rates in the cued response test (Part Two). Part One is a situation akin to using the technology the first time with minimal knowledge and instructions. Thus, it was not surprising that only familiar symbols or those that are concrete representations of the functions would have higher hit rates (ex. microphone, and camera), confirming one of the main hypotheses of this study. Further, it was interesting to note that symbols could indeed be quite difficult to understand or comprehend based on the very low hit rates in Part One. On the other hand, by using the cued-response tests several objective parameters were derived that helped reveal other possible differences between the candidate symbols and between subject groups. These can be summarised as follows. Overall, hit rates or correct symbol-referent associations are still the most important objective parameter of symbol performance. However, hit rates are not the only important parameter, especially when evaluating several sets of symbols and using different subject groups. Non-hit parameters such as false alarms or confusions and missing values can be of great importance as well. False alarms or confusing the symbol as representing the wrong referent is a very good gauge of studying the how distinct a symbol is compared to the other symbols presented at the same time. This parameter becomes more useful when used among subject groups with different cultural backgrounds (i.e. countries). The predominant types of confusions (symmetrical and asymmetrical) in each group could be used to study further how the symbols are understood and even confused by the subjects. These would then be helpful if to determine if the problems would merit (a) re-design of the symbols; (b) replacement of the symbols; (c) more explanation or emphasis on how the definition of the functions. The issue would then be whether the recommendations should focus on symbol re-design or replacements. Another would be on the use of instructional aids. Missing values are also very important since they represent situations wherein the prospective user plainly lacks the knowledge of which among the symbols represent the desired function (referent). In actual user scenarios, they can be akin to non-use or under-use of the device. If this would be one of the main problems, recommendations would be more on steps ensuring familiarity with the device and its functions. Lastly, the confidence judgements shown in subjective certainties, semantic scales and preferences were also helpful in gaining insight in subject group differences especially to certain possible subjective biases that may be culture-related. They also showed that when intended for diverse user groups, graphical symbols that are more concrete and famliar have better chances of being understood correctly (associated to the correct functions).

# 1 INTRODUCTION

## 1.1 From signs and symbols to icons: A background on graphical symbols

The ubiquitous nature and applicability of the current computer and information technology are quite palpable nowadays. Innumerable tools and equipment in schools, homes and workplaces have become computer-based. From the lowly flat irons, ovens and telephones, to different audio-video appliances and high- tech desktop and notebook computers, all of these have become equipped with more complex functions. Consequently, one needs to learn how to manipulate graphic displays and buttons to avail of the added functions. In the midst of these complicated machines and the humans' attempt to utilise them, are the graphical symbols. This introduction is a literary discourse on the roles played by graphical symbols in human-machine interactions, particularly those involving graphical user interfaces. It would discuss graphical symbols in the realms of cognitive ergonomics and human-computer interaction. It would start with some definition of terms followed by a discussion on how symbols are probably processed and understood. It would then proceed to a brief history of the evolution of graphical symbols - from being a major part of daily written languages in the past, to their current role in graphical user interfaces (GUIs) of modern-day devices.

## 1.2 Definition of terms

The advent of computer technology and the subsequent widespread use of computer-related products have given the ordinary human being the potential to perform numerous and more complex tasks under shorter periods of time. However, in order to harness the computer's power to assist humans in their work, designers and human factors specialists alike need to understand the technology while simultaneously being sensitive to human capacities and needs (Shneiderman, 1992). They need to apply human factors principles and processes in making human-computer interactions suited, or more importantly, centred to the human instead of the machine - the computer. This is one of the reasons how the Graphical User Interfaces (GUIs) came about. GUIs can be defined as the use of direct manipulation and icons or other graphical symbols on a display to interact with a computer or computer-based devices (Stramler, 1993). Horton (1994) and Böcker (1997) added that present-day GUIs in computers utilise the so-called WIMP interface (windows, icons, mouse and pointer).

In general, a symbol is any graphical character or other representation which is intended to: (a) stand for something else, (b) communicate a use for an object/structure, or (c) communicate what should or should not be done at a given time or location (Stramler, 1993). Graphical symbols usually pertain to terms like icons and pictograms or pictorial symbols (Böcker, 1993). The former refers to symbols that are simple, concrete and usually self-explanatory of the ideas, objects or functions they represent (Wood and Wood, 1987). The latter types are usually more abstract, conveying messages by analogy or symbolism. Thus, pictograms require certain learning processes to be understood. These terms would be discussed in more detail later under the cognitive aspects of symbol understanding.

Nonetheless, the term *icon* is the most popular among the terms and has become synonymous to any small visual symbol. Aside from being graphical, the term icon also refers to the non-linguistic representation of an object or action (Stramler, 1993). In addition, when one talks of computer systems and documentation and graphical interfaces in general, Horton (1994) described icons as the small pictorial symbols used on the computer menus, windows and screens. They represent the capabilities of the computer system. They can be activated and bring forth these capabilities into action. Thus, compared to the commonly encountered traffic, directional and safety signs and symbols, icons and pictograms in computers and other audio-visual devices are much smaller and usually appearing in sets. This paper deals with small graphical symbols. They would generally be used here to pertain to both icons and pictograms.

## 1.3 Some aspects of cognitive ergonomics of graphical symbols

### *1.3.1 Visual image processing*

Knowing how people process visual images is essential in understanding how graphical symbols can be designed and utilised better. Horton (1994) used two models to explain the visual image processing of symbols. The first, he called the *naïve model of visual processing*. According to this model, human vision is likened to the camera. Perception forms an interior image that is an exact copy of the scene in front of the eyes. This scene is then photographically recorded and stored in memory. Although this model is simple and intuitive, Horton claims that this view is erroneous. If this were true, objects will only be recognised if they appear exactly similar as the first time they were seen. The other model was called the *richer model of visual processing*. He cited at least five ways how it differed from the former model. Moreover, it more adequately explained how symbols could be understood and even remembered (Horton, 1994):

- *Perception filters visual inputs. Less than 1/1000 of 1 percent of the information taken in by the retina makes it through short-term memory*
- *We remember not just one visual image for the object but several*
- *In additional to visual memories, we remember the name of the object and its characteristics. All these memories are linked so that recall of one triggers others*
- *Much, perhaps most, of what we see comes not from the eyes but from the memory. We see mostly what we already know and have seen before. Visual input seems*

*designed more to trigger the correct memories than to tell us directly what something looks like; and*

- *What we currently see and what we have seen affect how we interpret visual input and what we remember. We tend to seek out details that confirm our current interpretation of a scene.*

Thus, the second model implies that visual processing is much more complicated than one would normally presume. It involves the processes of visual perception, memory (short- and long-term), that are equally affected by personal experiences in the external world.

### *1.3.2 The role of context in symbol meaning*

After the concept of visual image processing of symbols, we still need to know how symbols generate their meaning. One can easily agree that graphical symbols existing alone are of no meaning. It is when a graphical symbol is used in a particular context that it can trigger memories and associations in one's mind referred to as its meaning (Horton, 1994, Edworthy and Adams, 1996, Böcker, 1997, and Dewar, 1999). Context is the situation or scenario wherein we encounter the symbol. It involves all other things in the field of view that can add or interfere with our perception of the symbol. Horton (1994) suggested that context could be any hint that leads us toward one particular interpretation of the symbol. Contexts can thus be the adjacent symbols, related labels or texts, other windows or displays, or other objects in the environment. We cannot know how graphical symbols will be interpreted without knowing the context wherein they will be seen (Wolff and Wogalter, 1998). Further, depending on the contexts and the users, several meanings can be generated by a symbol:

$$\mathbf{Icon}_{i} + \mathbf{context}_{tj} + \mathbf{viewer}_{k} = \mathbf{meaning}_{ijk}$$

Horton (1994) illustrated this by giving the example of the symbol shown in Figure 1.

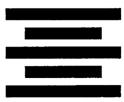


FIGURE 1.

When shown alone, a symbol's meaning depends on the user's previous associations. It could denote a battery for someone with an electrical engineering background. For someone involved with word processors, the same symbol could be interpreted as centring lines of texts. However, when shown with other related symbols, the intended and correct meaning would become clear to most of the users. In this case, its meaning was indeed the centring lines of texts (Figure 2).

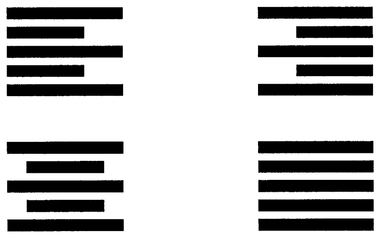


FIGURE 2.

### *1.3.3 The sign-signal-referent relationship*

The model suggested by Familant and Detwailer (1993) is another way to see how symbols are related to their intended meanings or referents (Figure 3). They suggested that information become encoded as a "signal" that can be interpreted if it is mapped to a "referent" to the real world. The signal's characteristics depend on the medium of communication. For example, the word "house" when spoken has acoustical features, and has visual features when shown as a picture. In turn, the "sign" is the relationship between the signal and its referent. This relationship is seen to be successful when a user can unequivocally refer to the signal's intended referent. Familant and Detwailer (1993) further added that signs could be "iconic" or "symbolic". They termed it iconic when the features of both the signal and the referent intersect (ex. symbol of "trash can" to represent file erasure function - "throwing away something"). It is symbolic when the sign relationship is arbitrary and has to be learned. Böcker (1997) further elaborated that visual symbols should not necessarily be pure icons nor pure pictograms or symbols It is better to understand the terms “iconic” and “pictographic or symbolic” as “endpoints” or anchors on a continuous scale. A symbol may then lie anywhere along such a scale depending on its composition.


FIGURE 3. Sign, signal, referent relationship (Familant and Detwailer, 1993).

### *1.3.4 Symbol classifications*

Together with the above models of symbol perception, symbol characteristics have also become the bases of symbol classifications. In general, symbols are categorised whether they are image-related (representational), concept-related (abstract), or arbitrary (Modley, 1966, Dreyfuss, 1972, Tudor, 1994, and Edworthy and Adams, 1996). For example, in GUIs such as those in computer-based word processing programs, the pictorial image of a printer is a direct or concrete representation of an object easily recognised.

Abstract or concept-related symbols are those where an aspect of the concept lends itself to symbolisation in a way that it evokes the whole concept itself. The symbols for store and retrieve as illustrated in Figures 4 and 5 are good examples of abstract symbols.

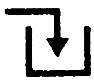


FIGURE 4.

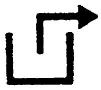


FIGURE 5.

Arbitrary symbols are symbols whose meanings are arbitrarily assigned to the graphics. For example, when the symbol of an exclamation mark or a question mark (Figures 6 and 7) is used, neither of them signifies any aspect of the real world objects. Rather, a meaning has been arbitrarily assigned to each one and one has to learn to understand what these meanings are.

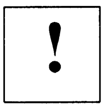


FIGURE 6.

FIGURE 7.

When arbitrary symbols are used together with other symbols such as in GUIs, their graphic images should be simple and clearly distinct from the others whose concepts are closely related or may be used under similar contexts (Edworthy and Adams, 1996).

Classifying symbols based on abstraction or representativeness is only one of several methods of classifying graphical symbols (Böcker, 1997). For example, Dewar (1999) divided the symbols into five main categories: industrial and occupational (workplace); representing methods (machines, instructions); management of public places (transportation, museums, hospitals); knowledge; and particular activities (sports). Symbols may also be classified according to the following (Gittins, 1986, Horton, 1994, and Böcker, 1997):

a) *Surface features:* form (associative vs. key), type (dynamic vs. static), colour (monochrome vs. coloured), and location (isolated vs. attached) as differentiating factors

b) *Metaphors used*: use of a unifying metaphor. This pertains to the use of knowledge acquired in a different context to quickly learn and use functions of a new system being introduced. A good example is the desktop metaphor in computer systems where the symbols from the office environment (files, folders, trashcan, telephones, etc.) are used to denote the different functions of the computer's operating system.

c) *Structural Description*: this was suggested by May, Böcker, Barnard and Green (1990) based on their application of a method of structural description developed in linguistics by Halliday (1970). Based on the latter, a structural description is a formal way of how to analyse the components of an entity and the different themes or thematic transitions that can be generated by different psychological subjects. An entity may be any task or experience such as listening to someone talk or looking at pictures. The different thematic transitions can be used as indicators of the cognitive efforts required by each subject to properly recognise the object in question.

## 1.4 The use of graphical symbols throughout the times

### *1.4.1 From paintings to printing press*

The use of signs and symbols is not new. It is also obvious that the evolution of graphical symbols is much tied to the history of human communication itself. As early as 50,000 BC, symbols appeared as paintings or carvings on caves and stone walls, with depictions of humans dating as far back as 11,000 years ago (Dewar, 1999). Among the ancient civilisations, one could easily recognise the use of symbols in the writings of the Chinese, Egyptian and African, and even the early American natives. They had used pictorial images and symbols to tell stories, myths and even to provide instructions (Horton, 1994). In their book on warning design, Edworthy and Adams (1996) cited the work of Senner (1989) wherein the latter discussed the history of written communication as beginning with the use of pictorial symbols to represent simple ideas. Soon, the simple stand-alone symbols gave way to abstract writing systems like the cuneiform and hieroglyphics. Early in this process, he pointed out that the pictorial representations and their use became so stylised, eventually giving birth to the modern alphabets and syllabaries (syllabic scripts). This development implied that those who had learned the symbols' meanings and their system of use would be the ones who could only use the written communication system. Moreover, during these early times, the prevailing technology was harnessed to facilitate the use of such written communication system. Simple technologies yielded simple devices for writing and publication. As technology advanced, so did the use of the texts or modern alphabets in communication. The advent of the printing press introduced the mass circulation of published documents. Books, newspapers, magazines and all sorts of print materials began to be published earnestly and numerously in countries touched by the technology of the printing press.

### *1.4.2 Modern technology, globalisation and the return of the graphical symbols*

The rise of modern technology certainly was not confined in the printing press. Almost all facets of human endeavours have been continually affected by modern technology. For in a way, technology itself is an expression of the human society's evolution. Thus, modern technology has also led to the growth in international travel and communications, as well as the increase in the use of computerised products by the *common tao* (person). However, this "globalisation" of travel and products have returned us to situations wherein we need to convey simple concepts quickly and efficiently, and would not require extensive learning (Edworthy and Adams, 1996). We thus find ourselves going back to the use of graphic symbols whenever it is deemed appropriate. Edworthy and Adams (1996) traced the use of symbols in public information systems, industrial safety and the present world of computer interfaces to the great success of using highway signs.

Then during the last two decades, graphical symbols as part of graphical user interfaces went into the upturn. A primary factor behind this has been the rapid development in microprocessor technology leading to the rise of computer and information technology. The first microprocessor was the Intel™ 4004, produced in 1971 (Encarta, 1997b). It was originally developed for a calculator containing 2300 transistors on a 4-bit microprocessor. It could perform only 60,000 operations per second. The first truly general-purpose microprocessor was developed in 1974. It was the 8-bit Intel™ 8080. It contained 4500 transistors and could execute 200,000 instructions per second. Modern microprocessors have much greater capacity and speed. They include those manufactured by Intel™ (Intel Pentium Series) and by the Sun Microsystems™ (UltraSparc). Apple™, IBM™ and Motorola™ have jointly developed the PowerPC microprocessors, while Digital Equipment Corporation (Digital™) has the Alpha microprocessors. All of these contain millions of transistors each (Encarta, 1997b). It must also be noted that the technology of microprocessors and integrated circuit fabrication is still changing rapidly. Currently, the most sophisticated microprocessors contain about ten million transistors. By the year 2000, advanced microprocessors are expected to contain more than 50 million transistors, and about 800 million by 2010 (Encarta, 1997b).

With the above rapid advancement in microprocessor technology and its wide-ranging applicability, a lot of human tools and devices have inevitably become smaller, multi-functional, and "computer-based," with interfaces that are equally smaller with numerous control features. Accordingly, to properly handle these have led to the development of mixed interactive dialogue graphical interfaces for personal computers and workstations such as Xerox's™ "Star" user interface, the Apple Macintosh™ operating system, and the widely-used MS-Windows™ by Microsoft™ (Böcker, 1997). They generally utilise toolbars, pull-down menus, and dialogue boxes containing either one or a few buttons, to arrays of buttons with symbols as shortcuts to complex functions (Böcker, 1997). At present, the proliferation of graphical symbols is even no longer confined to the traditional desktop and notebook PCs. They are found among the smaller and more compact PC variants that are called "palm-sized" notebooks and organisers. Even telecommunications products are not spared, as exemplified by the increasingly popular mobile phone. In these products, buttons contain symbols representing the device's additional and often more complex functions (i.e. text and voice mails, telephone directories, and personal settings).

Several reasons have been given why icons and pictograms have come to be increasingly used again. Most of these point to their advantages over text-based messages themselves. Compared to text, symbols are more distinct (Maguire, 1985), faster and easier to recognise (Collins and Lerner, 1982), and can even reduce the likelihood of errors (Lodding, 1982). Horton (1994) pointed out those graphical symbols such as icons used in computers and computer-related products make these products go global. In their paper on safety pictograms, Davies, Haines, Norris and Wilson (1998) mentioned that in the light of the emerging European market, graphical symbols have become very attractive since they have the potential to cross the language barrier. Messages may need not always be translated or explained in the 13 languages of the Union. Aside from the potential of symbols to make things usable under international settings, some studies have shown their usability among geriatric people. As one gets older, contrast sensitivity decreases. However, Kline, Ghali, Kline and Brown (1990) showed in their studies that the visibility of symbols was better than their equivalent worded signs among older subjects even under poor contrast conditions.

# 2 CURRENT PROBLEMS AND ISSUES

## 2.1 Restrictions on the advantages of using graphical symbols

The use of graphical symbols, however, is not devoid of problems. Foremost among these is that some symbols are simply not understood well (Brelsford, Wogalter and Scoggins, 1994). The advantages cited above were evident when the concepts being represented were well understood or concrete. Hence, the symbols used were representational or image-related types. The advantages diminish when the ideas become more abstract. Sanders (1998) further added that the superiority of symbols can only be attained if they contain a small amount of detail, are sufficiently distinct in shape and are unambiguous. Regarding older people, Davis et al (1998) cited the studies of Easterby and Hakiel (1981) showing that although symbols are easily recognisable, understanding their meaning is another matter, and that it is generally poorer compared to younger people. In addition, most of the studies focused on symbols occurring individually and/or bigger in sizes (i.e., informational and directional symbols in public places such as highways, airports, train stations and hospitals). In turn, GUIs typically have smaller symbols (such as those in toolbars and small display screens of portable devices).

Another problem with the use of graphical symbols pertains to the relationship between the range of symbols presented simultaneously and the contexts in which they are used. Edworthy and Adams (1996) aptly pointed this out by stating that although a symbol is well understood in a given context, this advantage will only last as long as the range of the symbols used in that context does not lead into higher chances of confusion. For example, in database or document filing programs, a folder with a certain colour or mark may indicate a certain type of file. Such differentiating features should be distinct enough to help the user avoid confusions with other file types.

## 2.2 Standardisation problems

The next set of problems regarding graphical symbols pertains to standardisation. It is easy to understand that the proliferation of graphical symbols in an equally increasing number of modern devices is a problem in itself. Compounded by the growth in number and diversity of targeted users owing to globalisation, standardisation in design and implementation of symbols is obviously desirable. Dewar (1999) pointed out that although efforts done by the International Organisation for Standardisation (ISO) has led to some internationally accepted symbols, there still remains a high variety of symbols used around the world. Symbol designs vary not only geographically but also across different applications (Olmstead, 1999). These are complicated by the fact that an acceptable and uniform level of comprehension when testing symbols is still lacking (Wolff and Wogalter, 1998, and Dewar, 1999). Comprehension standards vary from

International Organisation for Standardisation's 67 percent (Zwaga, 1989) to 85 percent of the American National Standards Institute (ANSI, 1993). Nonetheless, Dewar (1999) suggested that it is more important that acceptable levels of comprehension should consider the nature of the message and the consequences of failing to understand it.

## 2.3 Symbol evaluation issues

The tendency to generalise that symbols can be easily designed and understood by anyone is still high. If the person who creates or designs a symbol understands its meaning and can easily convince a colleague that it is a very good representation of a certain referent or function, then the rest of the general public will do so as well (Edworthy and Adams, 1996 and Rutter and Becka, 1997). Researches using naive populations have shown that there is at least 15% of the subjects who are not able to correctly understand the symbols being tested (Zwaga, 1989). Although the 85% level of correct interpretations can be considered as high, the extent of the implications of having 15% getting it wrong must always be weighed. With the current globalisation wherein technology and its products transcend national borders, the extent and levels of 'naive users' assume greater implications. Thus, the relevance of empirical testing graphical symbols for their intended uses and meanings among potentially different target user groups cannot be over-emphasised. Graphical symbols have to be tested for their intended uses and contexts among the intended users.

### *2.3.1 Current test methods*

With regards to symbol testing, various methods have been utilised to evaluate icons and pictograms. Some of the methods frequently used were the following: a) appropriateness tests, b) preference ratings tests, c) naming and matching tests d) comprehension tests, e) recognition and recall tests, and f) paired-comparison tests (Green and Pew, 1978; Mackett-Stout and Dewar, 1981; Jones, 1983; Rogers, 1986; Nolan, 1989; Zwaga, 1989; Magyar, 1990; Salasoo, 1990; Stammers, 1990; Clarke and Brown, 1992; Lin and Kreifeldt, 1992; and Lin, Kreifeldt and Chi, 1992). Webb, Sorensen and Lyons (1989) described these and other methods of evaluating symbols as involving psychophysics, scaling, recognition and memory testing, as well as statistical modelling and analysis. The type of testing employed was often dependent on the type of study being pursued. Generally, appropriateness testing was the preliminary procedure to screen several candidate designs. Matching studies determined how well the symbols performed as a set and their likelihood of being confused with each other (Nolan, 1989).

Another type of test worth mentioning is the spontaneous recall or identification test. It is also oftentimes known as the open-ended type of test. The symbols being tested are shown to the subjects, individually or simultaneously, and they are asked to state/write on spaces provided what they think each symbol mean for them. This type of test is usually regarded as the ”yardstick” of all other types of symbol recognition tests (Dewar, 1994 and 1999, and Wolff and Wogalter, 1998). This type of test is indeed very helpful and informative because it can reveal the ease of understanding of the symbol, as well as the extent of confusion and types of errors that the users make. However, one big drawback of this method is that data reduction can be time consuming (Dewar, 1999). Intended use or context of the symbols should also be clear to the subjects to avoid unnecessary or invalid confusions and errors (Wolff and Wogalter, 1998).

### *2.3.2 Use of several test parameters*

The use of the single test methods as described above has the advantage of speed, simplicity and ease in interpreting results. However, to properly evaluate and decide which among several alternative symbols would perform best, several tests are needed. Further, multiple test parameters is deemed most suitable when the candidate symbols are intended as parts of interfaces of devices for international use or for standardisation. For example, Mackett-Stout and Dewar (1981) studied different versions of public information symbols using a number of measures and deriving an "efficiency index" for each symbol. In turn, the International Standards Organisation (ISO) came out with the ISO 9186. It was a six-stage procedure for the development and testing of public information symbols (Zwaga, 1989). Its major portions were the comprehensibility judgement tests and the comprehension tests. On the other hand, the International Telecommunications Union (ITU-T) had F.910 (ITU, 1995). The Recommendation F.910 endorsed a symbol testing procedure composed of four parts. It involved the determination of need for new symbols, the creation and evaluation of the new designs, and the selection and approval.

### *2.3.3 The MIA - Multiple Index Approach*

An increasing number of icons and pictograms in most computerised personal, office and even industrial devices are small and appear in sets to represent series of functions. The Human Factors Technical Committee of the European Telecommunications Standards Institute (ETSI) evaluated videotelephony symbols using a multiple index approach (MIA) which can also be used in testing symbols for other commercial products (Böcker, 1993; ETSI, 1993). This method was also developed in response to the issue that most of the methods mentioned earlier were meant to test graphical symbols that were often large in size (traffic and directional symbols). These symbols are those usually found in public places, and appearing individually (“stand alone”). The ETSI used MIA to test different sets of candidate graphical symbols (point-to-point videophone pictograms) intended to appear together in small displays and were thus quite small themselves (ETSI, 1993). Another rational behind the MIA was the relevance of tests reflecting actual user scenarios. This view could be easily appreciated under the principle of contextualisation discussed earlier. To properly evaluate symbol understanding, it should be tested in contexts or situations it is designed for use. For example, when symbols are designed to appear together, then they should be tested as such, in sets or groups. The intended device where they would be used (i.e. videophone) together with the referents to be tested should also be properly presented and explained.

In all, using MIA would enable the tester to collect seven indices that can help in the final selection of symbols. The indices were divided into the objective and subjective. The objective indices (also known as performance indices) were hit rates, false alarm rates and missing values. The subjective indices were subjective certainty and suitability, symbol and symbol set preferences (Böcker, 1993). It was emphasised that although the test of associativeness (hit rates) was the main index of performance the other objective parameters (false alarm rate and missing values) were also essential when testing symbol comprehension among symbols appearing in sets. The subjective parameters (subjective certainty, subjective suitability and preferences) would be used to decide between two or more graphical symbols with similar performance data. This way, the best set of symbols can be decided using factors other than simple correct symbol-referent matching. Thus, it is not difficult to appreciate the relevance and applicability of using such a multi-factorial approach in evaluating graphical symbols across diverse user groups can be of great help in evaluating candidate graphical symbols for international use.

### *2.3.4 The semantic differential test*

The semantic differential technique or SDT was originally developed to measure the connotative meaning of a concept or idea (Osgood, Suci and Tannenbaum, 1957; Edwards and Porter, 1972; Henerson, Morris and Fitz-Gibbon, 1987). Its basic purpose was to analyse concepts in question based on a limited number of dimensions of meaning. This involved subjective ratings of the stimuli on a number of bipolar adjectives belonging to three or four semantic differential factors. These were evaluative, potency, activity and understandability, with the first three known as the EPA factors of Osgood (Osgood et al., 1957). Other semantic factors can actually be derived as shown by Caron, Jamieson and Dewar (1980). Concepts can also be compared with other concepts to evaluate their similarity or difference in meanings, or compare the same sets of concepts among different groups, including national and language groups (Oppenheim, 1986).

SDT has also been used to evaluate pictograms and other graphic symbols (Dewar and Ells, 1979; Caron, Jamieson, and Dewar, 1980; and Vora, Helander, Swede and Wilson, 1991). In order to knowing whether a candidate symbol is appropriate for use, it must be tested for comprehension. In their study of 20 traffic sign symbols, Dewar and Ells (1979) considered the understandability factor as representing the extent to which the subject knows (familiarity) and comprehends the object or concept in question. Potency and action in turn were related to comprehension and subjective meaning. Comprehension for each symbol was determined by computing the percentage of correct responses.

Semantic differential scores for each symbol were then calculated using mean ratings of the adjective pairs comprising the index of meaning for each factor. Through rank order correlation, they analysed the relationship between percentages of correct comprehension and semantic differential scores. They found out that the evaluative and understandability semantic factors might be useful in evaluating the users' perception and understanding of the symbols. In contrast, Vora et al. (1991) used adjective pairs commonly used by graphic designers (such as balanced - unbalanced) but had only one factor able to explain more than 60 percent of the variation. In a series of studies, Lin and his colleagues employed the principles of semantic differential testing towards proposing a tool that computer symbol designers can use in predicting information needed for evaluating and modifying them (Lin, 1992; Lin and Kreifeldt, 1992; and Lin et al. 1992). Instead of using the EPA factors of Osgood, they derived three cognitive factors, which they believe affected symbol recognition and comprehension. They tested its applicability in actual symbol evaluations where the factors they derived were able to explain about 87% of the variances. However, it would have been equally interesting had they tested the method using prospective users also. This way, symbol evaluation would have had more credence by using the users as evaluators. Still, SDT may be able to give insights on the possible psychological factors that can affect symbol recognition and comprehension. Ossner (1990) mentioned other studies showing that the results of various methods may be actually "related to the semantic features inherent in the content/proposition without, however, determining them". SDT may thus be a useful tool in graphic symbol evaluation but more studies are needed using prospective users instead, which brings us to another factor affecting symbol evaluation - the subjects or respondents used in the tests.

### *2.3.5 Prospective users as subjects*

Another critical factor affecting symbol evaluation is the type of subjects involved in the studies. Take the ordinary situation wherein prospective users are choosing between different models of a computer-related product. Each model has the same functions as the others but represented by different types of icons or pictograms (one icon type or family, one model). Thus, as each user proceeds in matching the functions with the symbols, he/she gets to encounter how usable the product is for him. After trying the functions of a product based on the symbols, the user gets to know if he did it right or not. This is usually the part where users give comments that the "keys" or the symbols or icons are difficult to recognise, or not compatible with the functions (referents). Thus, graphical symbols as interfaces in computer-based systems play a critical role in their usability (and appeal) to the users. Users must be able to see and use symbols as appropriate, meaningful, memorable and learnable in order for them to fully utilise the products. However, icons and pictograms are generally designed and developed with 'western users' in mind but targeted for international use. Moreover, the subjects used in some of the traditional methods discussed earlier could hardly even be described as representing typical users, such as industrial design students, professionals, and even communication employees (Tudor, 1994).

### *2.3.6 Culture and symbol evaluation*

#### *2.3.6.1 Need for more studies on culture*

Finally, users' involvement become more crucial when users' backgrounds are vastly different as in eastern vs. western users, where one culture develops the products and their symbols, and another culture/s use/s them. Keesing (1974) defined culture as something that also influences the "publicly available symbolic forms through which people experience and express meaning". Although the influence of culture on technology has been recognised, empirical studies are still inadequate (Abeysekera, Shahnavaz, and Chapman, 1990, Sukaviriya and Moran, 1990, Shahnavaz, 1991). Regarding symbol evaluation itself, only recently has cross-cultural studies started to gain attention (Lin, 1999, and Olmstead, 1999). Studies focusing on third world countries, such as those in Asia, are still very few (Plocher, Garg, and Chestnut, 1999).

In his studies on culture, technology and ergonomics, Shahnavaz (1998) revealed that pictorial signs and symbols for public use are very much related to culture. He suggested that the best way to consider cultural aspects in the design is to adapt a user-centred design approach. The users' needs and backgrounds have to be reflected in the design and evaluation of the candidate symbols. Böcker (1993) also stressed the need to explain and test the functions (represented by the symbols) in manners reflecting actual or typical user scenarios. Tudor (1994), in using subjects who were potential customers as well, pointed out the relevance of involving actual prospective users, including novices, of the products where the symbols are intended to be used. For example, if symbols are intended for users of diverse cultural backgrounds, the choice of subjects must reflect this. Nothing exemplifies this better than Southeast Asia, or even Asia as a whole

#### *2.3.6.2 A brief background on Southeast Asia and the ASEAN*

Asia is the largest of the world's seven continents. Together with the outlying islands, it covers an area estimated to be 44,936,000-sq. km, or about one-third of the world's total land area. The Asian people account for three-fifths of the world's population. In the early 1990s, Asia was estimated to have had more than 3.2 billion people living in the different countries (Microsoft Encarta, 1997). Based on culture, the Asian continent may also be divided into two: that which is Asian in culture (East Asia, Southeast Asia, and South Asia) and that which is not (Asia of the former Soviet Union, and Southwest Asia). Of particular interest is Southeast Asia. It is a region comprising the Indochinese and Malay peninsulas and several nearby island groups. Southeast Asia is bordered on the north by China; on the east by the South Pacific Ocean; on the south by the Indian Ocean; and on the west by the Indian Ocean, the Bay of Bengal, and the Indian subcontinent. Southeast Asia includes the countries of Brunei, Myanmar (formerly known as Burma), Cambodia (Kâmpùchéa), Indonesia, Laos, Malaysia, the Philippines, Singapore, Thailand, and Vietnam.

The Association of Southeast Asian Nations (ASEAN) is a regional alliance of seven independent countries of Southeast Asia. ASEAN was founded in Bangkok in August 1967 by Indonesia, Malaysia, Philippines, Singapore, and Thailand (Microsoft Encarta, 1997). Brunei joined after attaining independence in 1984. Vietnam was admitted as the first Communist member in 1995. Its principal objectives were to accelerate economic growth and promote regional peace and stability. A joint forum with Japan was established in 1977, and a cooperation agreement with the European Community was signed in 1980. In January 1992, ASEAN members agreed to establish a free-trade area and to cut tariffs on non-agricultural goods over a 15-year period beginning in 1993. Since then, and except for the Asian economic crisis in 1996-1997, Southeast Asia has been experiencing rapid economic and industrial growth. With a fast growing economy coupled with a huge potential market based on a population much bigger than Western Europe or Northern America, the influx of western technology and its products are unavoidable.

*2.3.6.3 The need to test and use graphical symbols in a shrinking world*

The graphical symbols only typify the current globalisation of technology and its products. They have grown in number and types. They are being used across an equally increasing number of applications and products. This is ably exemplified by the videophone as discussed in this study. The videophone's forerunner is the ordinary telephone that facilitates verbal communication between two or more parties separated by some distance. Through technological developments, the videophone not only relays voice or audio inputs. It can even send video and data between communicating parties. For additional functions under limited space, symbols as interface are thus very suitable. However, although graphical symbols have the potential to ease the introduction and use of complex devices such as the videophone, poorly designed symbols can cause the opposite effect. If they are targeted for a wide spectrum of user groups typified by the people of Asia (different age, language and of course, cultural background), an empirical evaluative method is essential. It must consider the diversity of the prospective users (subjects) and the multi-factorial nature of symbol processing and understanding as discussed above.

# 3 RATIONALE AND OBJECTIVES OF THE PRESENT STUDY

The above discourse on the nature, impact, and issues of graphical symbols can be summarised into a few inter-related factors and comprise the rationale behind the current study.

1. The rapid pace in the development of the microprocessor has led to the equally rapid development of the computer and information technology.

2. A concurrent globalisation of the economy has likewise led to both the technology and its products to transcend national borders.

3. The above two factors are reflected by the proliferation worldwide of computer-based industrial and personal tools that are getting smaller but with increasing functions and complexity.

4. These have also led to the increase in use and importance of graphical symbols in interfaces (icons and pictograms) to make the products understandable and usable. Compared to the traditional directional and traffic signs and symbols, graphical symbols in GUIs are smaller and appear in sets or groups.

5. Thus, symbol testing is very important to help determine which among the several alternative symbols will be understood easily and correctly by the intended users around the world.

6. When symbols are intended for international use, using multiple performance or test parameters is essential. In this thesis, performance or test parameters pertain to both objective and subjective parameters of testing (objective and subjective data).

7. At present however, cross-cultural studies dealing with symbol evaluation, especially in Asia, are still meagre in number and far in between.

Thus, the main objective of this study was to analyse the understanding of small graphical symbols (pictograms and icons) among different potential user groups from eastern (Asian) countries like Indonesia, Malaysia, Philippines, Thailand and Sri Lanka. For comparative purposes, subjects from Finland, Sweden and USA were utilised. Different performance parameters derived from several tests were used to examine how symbol understanding would differ or be similar across different study groups. Based on the results, the study also aimed to formulate general recommendations regarding designing and testing small graphical symbols across different cultural groups.

# 4 HYPOTHESES

The hypotheses tested were the following:

(a) Subject groups (countries) would have some significant differences in interpreting the videophone symbols and their intended referents or functions based on different test parameters (hit rates, confusions, missing values, preferences and semantic scaling).

(b) Specifically, Asian subjects would perform poorly compared to European subjects in identifying videophone symbols.

(c) Videophone symbols that are familiar, simple and concrete would perform better than symbols that are new, complex and abstract.

# 5 METHODOLOGY

## 5.1 Initial Study: Philippine-Swedish Study

### *5.1.1 Respondents*

This study was based on a pilot study as discussed in Paper 1. One hundred subjects (50 each from the Philippines and Sweden) were involved in the initial study. The Filipino subjects had a mean age of 30.1 years (Range = 21-42 years, S.D.=4.60). The Swedish subjects' mean age was 26.6 years (Range = 21-49 years, S.D.=5.21). In both groups, the subjects were university students and office employees of universities as well as small- to medium-sized companies (25-100 employees/company). All subjects had at least two years experience using computers or computer-related devices at work and at home.

### *5.1.2 Materials and equipment*

In the form of booklets, several graphical symbols representing certain videophone referents or functions were tested. Twenty-three (23) referents and 69 videophone symbols used by Tudor (1994) were selected (Figures 8 and 9). All the referents and videophone symbols were based on Paper 1.

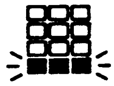
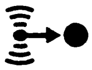
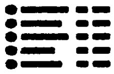
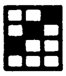
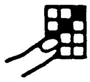
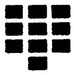
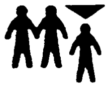
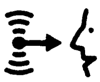
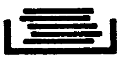
| REFERENTS | ABSTRACT SYMBOLS | CONCRETE SYMBOLS | PROPOSED SYMBOLS |
|---|---|---|---|
| 1. Achieve Dial Tone |  |  |  |
| 2. Answer Ringing Call |  |  |  |
| 3. Call Log |  |  |  |
| 4. Conference |  |  |  |
| 5. Dialpad |  |  |  |
| 6. Drop |  |  |  |
| 7. Help Specific |  |  |  |
| 8. Help System |  |  |  |
| 9. HFAI |  |  |  |
| 10. Hold |  |  |  |
| 11. Message |  |  |  |
| 12. Music On Hold |  |  |  |

FIGURE 8. The three sets of videophone graphical symbols used in the initial study, continued next page (from Tudor, 1994).

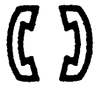
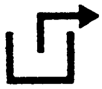
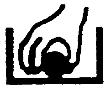
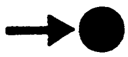
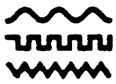
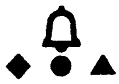
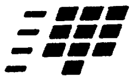
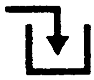
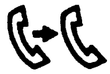
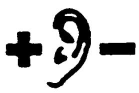
| REFERENTS | ABSTRACT SYMBOLS | CONCRETE SYMBOLS | PROPOSED SYMBOLS |
|---|---|---|---|
| 13. Mute |  |  |  |
| 14. Notes |  |  |  |
| 15. Phone Call Active |  |  |  |
| 16. Retrieve |  |  |  |
| 17. Ringer Select |  |  |  |
| 18. Speakerphone |  |  |  |
| 19. Speed Dial |  |  |  |
| 20. Store |  |  |  |
| 21. Switch Hook Control |  |  |  |
| 22. Transfer Call |  |  |  |
| 23. Volume |  |  |  |

FIGURE 9. The three sets of videophone graphical symbols used in the initial study, continued from previous page (Tudor, 1994).

The symbols were grouped into three types (sets): abstract, concrete, and proposed based on their general attributes. Sixty-two of the videophone symbols were designed in England. The remaining seven were designed in Japan. These were the proposed symbols for ”Volume”, ”Achieve Dial Tone”, ”Hold”, ”Music on Hold”, ”Mute”, ”Store”, and ”Speakerphone”. The questionnaire had two parts: recognition/subjective certainty tests, and semantic differential ratings using six bipolar scales (Simple-Complex, Concrete-Abstract, Meaningful-Meaningless, Familiar-Strange, Related-Unrelated, and Sharp-Dull). These bipolar adjectives were selected based on studies by Dewar and Ells (1979), Caron, Jamieson, and Dewar (1980), and Vora, Helander, Swede and Wilson (1991). Questionnaires were produced per symbol set with the symbols and their referents presented in random orders.

### *5.1.3 Procedure*

Three orderings of the videophone symbol sets were generated and randomly assigned to the subjects during the recognition/subjective certainty task. The order of the referents was likewise randomised when presented. In the recognition tests, the videophone symbols and referents were presented and to be matched correctly. For each recognition test, confidence judgements on the answers were indicated using a 7-point scale. After one week, the semantic differential testing was performed. The symbols were again presented. This time, they were to rate each symbol using the six bipolar adjectives. The same tasks were done for the three symbol families. Thus, each subject performed 69 recognition/subjective certainty tests, and 69 semantic differential tests (6 bipolar scales each), for the three symbol sets.

Likewise, in the semantic differential tests, the symbol sets and referents were also presented in random order among the subjects. The response variables in SDT were in the form of ratings using 7-point scales anchored in six bipolar adjectives. The subjects were asked to rate each symbol using the six bipolar scales. Spearman's rank order correlation was used to study the relationship between hit rates and semantic differential scores of the different videophone symbols.

## 5.2 Main Study: Southeast Asia, Europe and USA (Papers 2 to 7)

In the design of the methodology for the main study, several factors were considered. It took into account the recommendations by Tudor (1994) in using prospective users as subjects and considering typical user scenarios for the tests. It also adopted the MIA steps used by Böcker (1993). Specifically, since cultural diversity was a prime factor (Shahnavaz, 1998) because the videophone symbols were also intended for international use, subjects from different countries were targeted. This also bolstered the consideration of using multiple parameters of testing symbol understanding across such diverse user (subject) groups.

### 5.2.1 Respondents

Three hundred fifty-seven subjects from Southeast Asia, Finland, Sweden and USA were involved in the main study (187 males and 170 females). The overall age range was 17 - 52 years old, with a mean of 25 (S.D. = 6.8). Table 1 shows the gender and age distribution per country. Two hundred forty of them were from five Asian countries (forty-eight subjects per country). These countries were Indonesia, Malaysia, Philippines, Thailand and Sri Lanka. They were students, employees and professionals (doctors, lawyers and engineers), all of whom were computer-literate or had at least experienced using computers or computer-related products for the past 2-5 years at the time of the study. At the time of the study, none of the subjects had experienced using a videophone.

Furthermore, forty-eight of the subjects were from Sweden, forty from USA and 29 from Finland. The latter (Finnish subjects) were a special case involving elderly subjects (Paper 6). The former two pertained to subjects who were students and employees in Sweden and the US respectively (Paper 5), and were similar in demographic traits to the Asian subjects (except of course, for nationality or country).

TABLE 1. Gender and Age distribution, in years, per country, n = 48 per country, except U.S.A. where n = 40.

| Subject's Country | Sex | N | Minimum | Maximum | Mean | Std. Dev. |
|---|---|---|---|---|---|---|
| Indonesia | Male | 26 | 17 | 38 | 30 | 7.2 |
| | Female | 22 | 17 | 42 | 24 | 6.0 |
| Malaysia | Male | 24 | 21 | 38 | 26 | 5.4 |
| | Female | 24 | 21 | 33 | 25 | 3.2 |
| Philippines | Male | 25 | 18 | 44 | 25 | 8.0 |
| | Female | 23 | 18 | 36 | 25 | 6.1 |
| Thailand | Male | 24 | 17 | 52 | 26 | 8.2 |
| | Female | 24 | 17 | 34 | 24 | 5.5 |
| Sri Lanka | Male | 28 | 18 | 32 | 23 | 3.2 |
| | Female | 20 | 19 | 40 | 26 | 5.8 |
| Sweden | Male | 24 | 19 | 33 | 23 | 4.7 |
| | Female | 24 | 20 | 47 | 29 | 9.4 |
| USA | Male | 21 | 20 | 46 | 28 | 7.8 |
| | Female | 19 | 19 | 39 | 24 | 5.2 |

### 5.2.2 *Materials and equipment*

In Papers 2 to 7, three sets of videophone symbols gathered from studies by the European Telecommunications Standards Institute (ETSI) were tested in the form of questionnaires (Figure 10). Set 1 was the target set and contained the symbols adopted by ETSI as standard. The two other sets (Sets 2 and 3) acted as distracter sets. Each symbol set contained seven symbols representing seven videophone functions (referents). The seven videophone functions tested were the following (Böcker, 1993; ETSI, 1993):

| | |
|---|---|
| a) Videophone / Telephone: | upgrading / downgrading the call |
| b) Camera on / off: | turning on and off picture transmission |
| c) Microphone on / off: | turning on and off sound transmission |
| d) Selfview on / off: | turning on and off selfview function |
| e) Document Camera on / off: | switching between document and person camera |
| f) Still Picture on / off: | turning on and off screen freeze |
| g) Handsfree on / off: | switching between handset and handsfree modes |

### 5.2.3 *Procedure*

Testing of the subjects were done in universities, hotels and offices located in the cities of the countries involved in the studies. These were:

1. Metro-Manila, Philippines
2. Bangkok, Thailand
3. Bandung, Indonesia
4. Sarawak, Malaysia
5. Peradeniya, Sri Lanka
6. Luleå, Sweden
7. Oulu, Finland
8. San Francisco, California, and Florida, USA

Both questionnaires and oral instructions were made available in English for subjects from Malaysia, Philippines, Sweden and USA. For the other countries, they were also made available in the country's mother tongue (native languages).

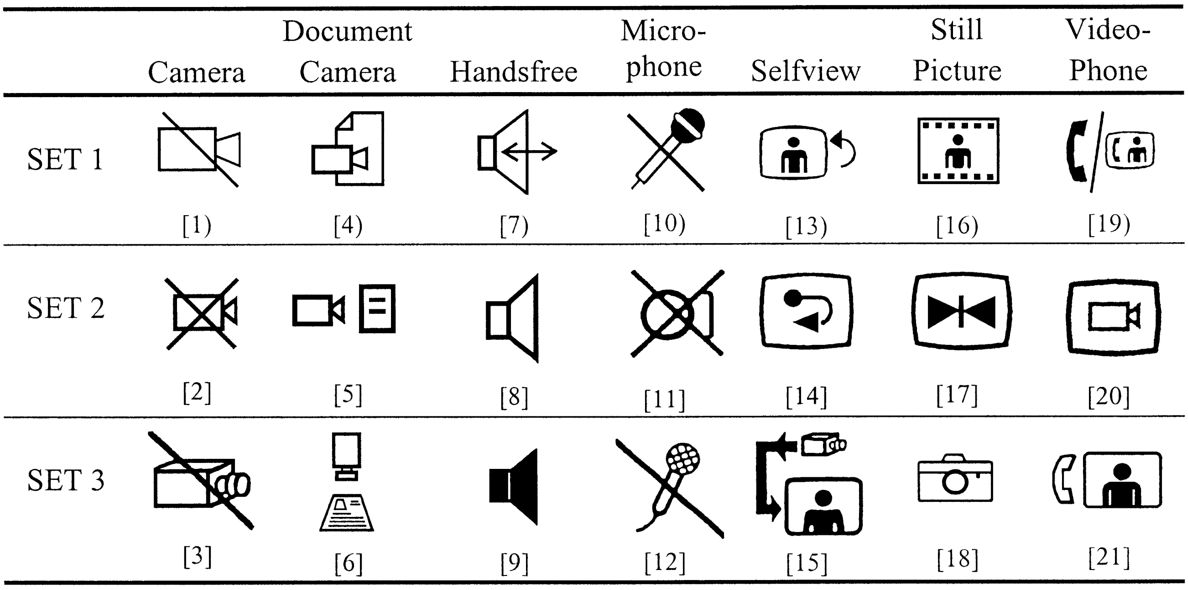



FIGURE 10. Graphical symbols used in the main studies (Papers 2 to 7) as based on Böcker (1993) for the European Telecommunications Standards Institute (ETSI, 1993).

Tests were performed in small groups in each country, lasting for about 45-60 minutes. Subjects were randomly given one of three versions of the questionnaires. The subjects were shown an illustration of a videophone. Its general functions were then discussed. Afterwards, instructions were given on how to go about the different test parts. Questions were entertained prior to the administration of the tests. Emphasis was given on avoiding omissions in trials in order to get back to them later. The order of the tests would also be strictly followed; that is, Part One followed by Tests Two, Three, and Four.

In the spontaneous identification test (Part One), each page contained one set of seven videophone symbols rendering three pages for the three symbol sets. On spaces provided for, they would write what videophone function they thought was represented by each symbol. They were also asked to rate the level of certainty for each of their answers using a seven-point rating scale (from ”very certain” to ”very uncertain”). In the cued response test (Part Two), the subjects first read a referent (function) and its description. Then they had to choose one symbol from a set of seven symbols which they thought best represented the referent in question. Each page contained one referent-description and one set of symbols. They were also asked to rate their certainties for their answers using the 7-point rating scales. There were seven videophone referents tested on three sets of symbols rendering a total of twenty-one trials in Part Two.

In the third test, five semantic bipolar scales similar to the first study were used to evaluate each symbol together with its correct referent, for a total of twenty-one SDTs. In Part Four, the three symbols (one for each set) together with the referent they represented were shown. The subjects had to choose one symbol they thought best represent that referent. Next, with the three symbols sets presented together, they had to choose the set they prefer most. In all, the subjects would choose seven symbols to represent the seven referents and one symbol set they preferred most.

### *5.2.4 Experimental design*

The whole main study was a 2 x 7 x 3 (Occupation x Country x Symbol Set) repeated measures design with 2-between and 1-within factors set-up. Between factors were those of occupation (students and employees/professionals) and country (Indonesia, Malaysia, Philippines, Thailand and Sri Lanka). The within factor consisted mainly of the symbol sets (Sets 1, 2 and 3). The videophone symbols were based on the studies by Böcker (1993; ETSI 1993), with Set 1 as the proposed set of symbols for seven basic functions to be found in a videophone. In all four tests, the order of the symbol sets was counterbalanced across all subject groups. Three orderings of the symbol presentation and referents were generated and randomly distributed in each subject group.

The response variables from the four tests were the following:

(a) Hits and Misses (mean percentages of correct and incorrect identification and recognition in Parts One and Two, using ISO 9186 of 67% correct recognition as a general guide of acceptable level)

(b) Confusions (false alarms) for each videophone symbol (total numbers and frequency distributions)

(c) Missing values (frequencies of No Answers in Tests One and Two)

(d) Subjective certainty ratings using 7-point scales for answers in Tests One and Two (frequency distributions and mean ratings)

(e) Semantic differential ratings of all symbols in the three sets using five 7-point scales each anchored in bipolar adjectives (frequency distributions and mean ratings, Part Three)

(f) Symbol and symbol set preferences (percent distribution, Part Four)

# 6 RESULTS AND DISCUSSION

## 6.1 Initial Study: Philippine-Swedish Study

Table 2 summarises the results for the recognition-matching tasks. Concerning the referents and both countries in general, only 4 of the 23 referents had symbols with at least 67% mean recognition in all three sets ("Conference", "Dialpad", "Music on Hold", and "Volume"). All the other icons had rates less than 67% regardless of which icon set they belonged. Again, since the hit rates (proportion of answers considered as 'correct') were basically binomial in distribution, arcsine transformations were performed in order to meet assumptions of normality and that multifactor ANOVAs would be valid.

### *6.1.1 Main effects and interactions*

Between sets, the Proposed Icons had the most number of high (at least 67%) recognition rates (Phil.: 11, Sweden: 8), followed by the Concrete Icons (Phil: 9, Sweden: 8), and Abstract Icons last (Phil.: 6, Sweden: 4). Between these icons, the results of multifactor ANOVA with *Set* as main effect showed that for the referents "Answer Ringing Call", "Message", "Mute", "Retrieve", "Ringer Select", "Speed Dial", and "Transfer Call", the Concrete and/or Proposed icons had significantly higher recognition rates compared to the Abstract versions ($F(2, 6)$ and $p<0.05$).

When the *Country* factor was taken as the other main effect, the Filipino subjects scored significantly higher recognition rates than the Swedish subjects for the icons representing "Answer Ringing Call", "Retrieve Call", "Speed Dial" and "Transfer Call". However, except for the latter case, these rates were generally below the 67% correct recognition of ISO 9186. For interaction effects between country and icon set, in both subject groups recognition rates were generally significantly higher with either the Concrete or Proposed icons ($F(2, 6)$, $p<0.05$) for their symbols for "Retrieve Call", "Speed Dial" and "Transfer Call".

### *6.1.2 Certainty ratings*

The countries were likewise compared regarding certainty ratings to their answers in the recognition tasks using Mann-Whitney tests (Appendix 1). For the abstract set, the following icons showed significant differences in certainty rating between the two countries: "Answer Ringing Call", "Conference Call", "Dialpad", "Drop", "Help System", and "Speed Dial". For the Concrete Set, these were icons for "Dialpad", "Help Specific", "HFAI", "Hold", "Music on Hold", "Mute", Notes, "Ringer Select", "Store", "Switch Hook Control", "Transfer", and "Volume". For the Proposed Set, significant differences in certainty rating scales were noted for "Dialpad", "Help Specific", "Help System", and "Phone Call Active".

TABLE 2. Study 1: Recognition rates in percent (first rows, bold-faced: above 67%) and certainty ratings (second rows) for the 23 Referents by subjects from Philippines and Sweden (N=100). Note: Phil. = Philippines, Swed. = Sweden.

| | **Abstract Icons** | | **Concrete Icons** | | **Proposed Icons** | |
|---|---|---|---|---|---|---|
| **Referents** | PHIL. | SWED. | PHIL. | SWED. | PHIL. | SWED. |
| 1. Achieve Dial Tone | 24 | 20 | 26 | 38 | 22 | 26 |
| | 2.72 | 2.64 | 3.68 | 3.74 | 3.06 | 3.20 |
| 2. Answer Ringing Call | 12 | 8 | 60 | 16 | 58 | 34 |
| | 3.62 | 2.56 | 4.58 | 4.20 | 4.38 | 4.02 |
| 3. Call Log | 50 | 14 | 26 | 22 | 42 | 52 |
| | 3.40 | 3.36 | 3.36 | 3.02 | 2.80 | 3.24 |
| 4. Conference | **78** | **94** | **96** | **96** | **96** | **92** |
| | 4.82 | 5.60 | 5.56 | 5.18 | 5.60 | 5.60 |
| 5. Dialpad | **86** | **82** | **86** | **90** | **80** | **92** |
| | 4.78 | 5.42 | 5.02 | 6.06 | 5.22 | 5.82 |
| 6. Drop | 34 | 28 | 38 | 26 | 34 | 26 |
| | 3.46 | 2.78 | 3.80 | 3.72 | 3.98 | 3.50 |
| 7. Help Specific | 20 | 42 | 64 | 36 | 38 | 46 |
| | 2.32 | 2.94 | 4.22 | 2.28 | 3.74 | 4.88 |
| 8. Help System | 34 | 14 | 20 | 46 | 34 | 32 |
| | 2.38 | 3.42 | 3.70 | 3.52 | 4.06 | 5.12 |
| 9. HFAI | 24 | 44 | 30 | 06 | 38 | 20 |
| | 2.68 | 3.40 | 2.66 | 4.02 | 2.80 | 2.40 |
| 10. Hold | 34 | 58 | 48 | 60 | **80** | 58 |
| | 2.88 | 3.12 | 4.48 | 3.28 | 4.78 | 3.96 |
| 11. Message | 36 | 24 | 20 | 10 | **90** | **74** |
| | 4.60 | 4.12 | 4.20 | 4.68 | 5.08 | 4.52 |
| 12. Music On Hold | **96** | **92** | **92** | **90** | **88** | **92** |
| | 5.34 | 5.54 | 6.24 | 5.34 | 5.78 | 5.32 |
| 13. Mute | **68** | 62 | 60 | **72** | 46 | 56 |
| | 4.14 | 4.06 | 4.12 | 5.10 | 3.70 | 4.22 |
| 14. Notes | 46 | 60 | 58 | 38 | **78** | **76** |
| | 3.90 | 3.38 | 4.96 | 6.06 | 4.98 | 5.02 |
| 15. Phone Call Active | 46 | 26 | 34 | 52 | 46 | 44 |
| | 3.50 | 3.16 | 3.74 | 3.90 | 4.32 | 3.32 |
| 16. Retrieve Call | **70** | 22 | 44 | 16 | 10 | 8 |
| | 4.16 | 3.32 | 3.64 | 3.12 | 2.62 | 2.34 |
| 17. Ringer Select | 44 | 56 | **86** | **90** | **84** | **86** |
| | 3.44 | 3.60 | 5.96 | 5.14 | 5.36 | 5.56 |
| 18. Speed Dial | 52 | 42 | **74** | 34 | **70** | 62 |
| | 4.14 | 3.22 | 4.50 | 3.74 | 3.60 | 4.08 |
| 19. Speakerphone | 62 | **72** | **90** | **80** | **72** | **84** |
| | 5.60 | 4.78 | 5.44 | 5.14 | 4.98 | 4.64 |
| 20. Store | 46 | 50 | 38 | 30 | 52 | 24 |
| | 3.58 | 2.92 | 4.00 | 2.92 | 2.82 | 3.20 |
| 21. Switch Hook Control | 28 | 48 | 20 | 22 | 44 | 20 |
| | 2.64 | 2.22 | 3.46 | 2.54 | 3.32 | 3.14 |
| 22. Transfer Call | 42 | 28 | **76** | **70** | **90** | 30 |
| | 3.34 | 3.30 | 5.30 | 3.84 | 5.48 | 3.30 |
| 23. Volume | **90** | **96** | **80** | **94** | **80** | **88** |
| | 5.50 | 5.42 | 5.06 | 6.06 | 5.62 | 5.68 |

Among the four referents with at least 67% correct recognition in all sets, and except for the concrete version of "Music on Hold", the Swedish subjects exhibited significantly higher ratings than their Filipino counterparts. But when comparing icons for referents with less than 67% recognition and where the two countries significantly differ in their recognition ratings, the Filipinos had the higher certainty ratings for "Answer Ringing Call", "Ringer Select", "Speed Dial", and "Transfer Call". Swedish subjects had significantly higher ratings only for the concrete icon for "Mute". There were no differences between their ratings of icons "Message" and "Retrieve".

### *6.1.3 Semantic Differential Scales*

Tables 3 to 6 contained the different levels of correlation between the icons of the three icon sets and the semantic scales for each country.

TABLE 3. Correlations (rank order) between semantic differential scores and percentage correct identification (N=100); *p < 0.01.

| | **Abstract Icons** | | **Concrete Icons** | | **Proposed Icons** | |
|---|---|---|---|---|---|---|
| **Referents** | PHIL. | SWED. | PHIL. | SWED. | PHIL. | SWED. |
| 1. Achieve Dial Tone | | | | | | |
| Concreteness | .419* | .409* | .389* | .308 | .375* | .187 |
| Familiarity | .274 | .078 | .379* | .181 | .535* | .292 |
| Meaningfulness | .348 | .011 | .365* | .004 | .544* | .102 |
| Relatedness | .217 | .203 | .224 | .137 | .465* | .049 |
| Sharpness | .337 | .277 | .480* | .068 | .548* | .144 |
| Simplicity | .353 | .185 | .365* | .009 | .308 | .377* |
| 2. Answer Ringing Call | | | | | | |
| Concreteness | .109 | .210 | .159 | .236 | .498* | .076 |
| Familiarity | .318 | .215 | .064 | .249 | .441* | .126 |
| Meaningfulness | .333 | .083 | .158 | .213 | .443* | .170 |
| Relatedness | .323 | .145 | .168 | .147 | .490* | .108 |
| Sharpness | .295 | .155 | .267 | .089 | .580* | .041 |
| Simplicity | .136 | .477* | .083 | .055 | .581* | .013 |
| 3. Call Log | | | | | | |
| Concreteness | .330 | .366* | .019 | .460* | .023 | .376* |
| Familiarity | .028 | .088 | .206 | .281 | .063 | .504* |
| Meaningfulness | .258 | .127 | .106 | .582* | .042 | .171 |
| Relatedness | .246 | .002 | .046 | .363* | .042 | .182 |
| Sharpness | .085 | .039 | .064 | .513* | .059 | .261 |
| Simplicity | .221 | .162 | .115 | .574* | .097 | .230 |
| 4. Conference | | | | | | |
| Concreteness | .336 | .108 | .225 | .273 | .198 | .026 |
| Familiarity | .388* | .356 | .158 | .319 | .151 | .174 |
| Meaningfulness | .350 | .170 | .064 | .082 | .104 | .008 |
| Relatedness | .432* | .393* | .284 | .111 | .206 | .019 |
| Sharpness | .467 | .065 | .243 | .347 | .135 | .102 |
| Simplicity | .465* | .330 | .214 | .272 | .194 | .000 |

TABLE 4. Correlations (rank order) between semantic differential scores and percentage correct identification (N=100); *p < 0.01, *continued.*

| Referents | Abstract Icons PHIL. | Abstract Icons SWED. | Concrete Icons PHIL. | Concrete Icons SWED. | Proposed Icons PHIL. | Proposed Icons SWED. |
|---|---|---|---|---|---|---|
| 5. Dialpad | | | | | | |
| Concreteness | .359 | .007 | .346 | .106 | .272 | .014 |
| Familiarity | .044 | .249 | .326 | .153 | .388* | .038 |
| Meaningfulness | .114 | .202 | .371* | .115 | .388* | .273 |
| Relatedness | .233 | .198 | .407* | .118 | .385* | .107 |
| Sharpness | .288 | .107 | .268 | .005 | .274 | .110 |
| Simplicity | .211 | .280 | .471* | .112 | .543* | .080 |
| 6. Drop | | | | | | |
| Concreteness | .536* | .330 | .695* | .348 | .604* | .193 |
| Familiarity | .702* | .210 | .418* | .387* | .619* | .363* |
| Meaningfulness | .602* | .448* | .425* | .358 | .732* | .429* |
| Relatedness | .659* | .433* | .469* | .310 | .746* | .418* |
| Sharpness | .315* | .251 | .253* | .310 | .698* | .483* |
| Simplicity | .598* | .233 | .752* | .193 | .735* | .341 |
| 7. Help Specific | | | | | | |
| Concreteness | .472* | .463* | .313 | .153 | .298 | .236 |
| Familiarity | .478* | .440* | .593* | .312 | .211 | .175 |
| Meaningfulness | .367* | .090 | .411* | .128 | .007 | .120 |
| Relatedness | .324 | .149 | .610* | .279 | .214 | .049 |
| Sharpness | .159 | .150 | .270 | .342 | .230 | .190 |
| Simplicity | .324 | .543* | .196 | .133 | .079 | .075 |
| 8. Help System | | | | | | |
| Concreteness | .250 | .000 | .443* | .128 | .231 | .028 |
| Familiarity | .129 | .152 | .451* | .171 | .182 | .074 |
| Meaningfulness | .049 | .078 | .622* | .228 | .191 | .075 |
| Relatedness | .080 | .246 | .554* | .172 | .230 | .018 |
| Sharpness | .052 | .131 | .429* | .222 | .184 | .598* |
| Simplicity | .277 | .233 | .539* | .061 | .083 | .366* |
| 9. HFAI | | | | | | |
| Concreteness | .037 | .250 | .204 | .422* | .142 | .393* |
| Familiarity | .507* | .033 | .125 | .354 | .430* | .212 |
| Meaningfulness | .404* | .237 | .135 | .292 | .552* | .165 |
| Relatedness | .479* | .132 | .081 | .403* | .547* | .397* |
| Sharpness | 434* | .265 | .252 | .357 | .368* | .124 |
| Simplicity | .502* | .098 | .108 | .402* | .459* | .167 |
| 10. Hold | | | | | | |
| Concreteness | .197 | .421* | .315 | .205 | .269 | .029 |
| Familiarity | .280 | .276 | .304 | .403* | .397* | .493* |
| Meaningfulness | .269 | .009 | .414* | .198 | .257 | .153 |
| Relatedness | .266 | .104 | .390* | .349 | .223 | .341 |
| Sharpness | .166 | .014 | .308 | .383* | .201 | .039 |
| Simplicity | .264 | .209 | .365* | .377* | .420* | .093 |
| 11. Message | | | | | | |
| Concreteness | .091 | .116 | .110 | .404* | .084 | .193 |
| Familiarity | .290 | .191 | .009 | .285 | .132 | .315 |
| Meaningfulness | .141 | .183 | .106 | .380* | .145 | .150 |
| Relatedness | .263 | .313 | .062 | .80* | .222 | .162 |
| Sharpness | .143 | .539* | .085 | .342 | .094 | .076 |
| Simplicity | .162 | .075 | .046 | .279 | .212 | .167 |

TABLE 5. Correlations (rank order) between semantic differential scores and percentage correct identification (N=100); *p < 0.01, *continued.*

| Referents | Abstract Icons | | Concrete Icons | | Proposed Icons | |
|---|---|---|---|---|---|---|
| | PHIL. | SWED. | PHIL. | SWED. | PHIL. | SWED. |
| 12. Music On Hold | | | | | | |
| Concreteness | .272 | .166 | .325 | .437* | .300 | .039 |
| Familiarity | .287 | .080 | .312 | .071 | .455* | .057 |
| Meaningfulness | .303 | .320 | .304 | .445* | .227 | .008 |
| Relatedness | .295 | .115 | .304 | .397* | .284 | .117 |
| Sharpness | .246 | .260 | .164 | .357 | .306 | .081 |
| Simplicity | .269 | .300 | .307 | .455* | .391* | .055 |
| 13. Mute | | | | | | |
| Concreteness | .432* | .319 | .568* | .284 | .301 | .317 |
| Familiarity | .181 | .085 | .590* | .351 | .389* | .366* |
| Meaningfulness | .445* | .388* | .462* | .369* | .408* | .121 |
| Relatedness | .392* | .345 | .523* | .339 | .443* | .387* |
| Sharpness | .370* | .322 | .635* | .527* | .377* | .251 |
| Simplicity | .307 | .072 | .428* | .098 | .451* | .267 |
| 14. Notes | | | | | | |
| Concreteness | .260 | .097 | .117 | .009 | .289 | .091 |
| Familiarity | .221 | .080 | .005 | .246 | .219 | .183 |
| Meaningfulness | .113 | .423* | .008 | .140 | .219 | .023 |
| Relatedness | .260 | .155 | .036 | .036 | .253 | .000 |
| Sharpness | .045 | .065 | .178 | .177 | .191 | .289 |
| Simplicity | .221 | .037 | .151 | .041 | .270 | .015 |
| 15. Phone Call Active | | | | | | |
| Concreteness | .447* | .542* | .152 | .246 | .397* | .099 |
| Familiarity | .388* | .400* | .232 | .289 | .227 | .121 |
| Meaningfulness | .493* | .538* | .371* | .179 | .427* | .277 |
| Relatedness | .557* | .375* | .274 | .418* | .163 | .170 |
| Sharpness | .467* | .240 | .205 | .213 | .271 | .168 |
| Simplicity | .178 | .175 | .207 | .454* | .230 | .102 |
| 16. Retrieve Call | | | | | | |
| Concreteness | .435* | .386* | .083 | .151 | .034 | .038 |
| Familiarity | .449* | .321 | .270 | .201 | .041 | .103 |
| Meaningfulness | .563* | .171 | .205 | .362* | .031 | .457* |
| Relatedness | .275 | .261 | .272 | .052 | .097 | .269 |
| Sharpness | .359 | .437* | .425* | .108 | .002 | .016 |
| Simplicity | .441* | .283 | .239 | .353 | .211 | .378* |
| 17. Ringer Select | | | | | | |
| Concreteness | .406* | .072 | .019 | .250 | .440* | .249 |
| Familiarity | .617* | .110 | .165 | .177 | .421* | .089 |
| Meaningfulness | .523* | .305 | .068 | .075 | .288 | .067 |
| Relatedness | .734* | .128 | .135 | .065 | .364* | .189 |
| Sharpness | .323 | .276 | .057 | .233 | .317 | .200 |
| Simplicity | .472* | .070 | .014 | .161 | .198 | .222 |

TABLE 6. Correlations (rank order) between semantic differential scores and percentage correct identification (N=100); *p < 0.01, *continued.*

| Referents | Abstract Icons PHIL. | Abstract Icons SWED. | Concrete Icons PHIL. | Concrete Icons SWED. | Proposed Icons PHIL. | Proposed Icons SWED. |
|---|---|---|---|---|---|---|
| 18. Speed Dial | | | | | | |
| Concreteness | .448* | .026 | .470* | .092 | .304 | .112 |
| Familiarity | .240 | .043 | .714* | .016 | .112 | .031 |
| Meaningfulness | .772* | .222 | .525* | .035 | .295 | .161 |
| Relatedness | .614* | .092 | .380* | .072 | .333 | .224 |
| Sharpness | .574* | .057 | .580* | .188 | .349 | .268 |
| Simplicity | .612* | .015 | .452* | .109 | .189 | .032 |
| 19. Speakerphone | | | | | | |
| Concreteness | .627* | .406* | .141 | .212 | .352 | .307 |
| Familiarity | .613* | .376* | .155 | .344 | .403* | .189 |
| Meaningfulness | .788* | .395* | .149 | .229 | .336 | .349 |
| Relatedness | .805* | .387* | .208 | .243 | .453* | .357 |
| Sharpness | .652* | .420* | .141 | .229 | .398* | .194 |
| Simplicity | .786* | .272 | .211 | .258 | .565* | .256 |
| 20. Store | | | | | | |
| Concreteness | .060 | .213 | .258 | .356 | .241 | .073 |
| Familiarity | .032 | .017 | .340 | .483* | .413* | .193 |
| Meaningfulness | .112 | .146 | .228 | .204 | .378* | .206 |
| Relatedness | .096 | .003 | .295 | .256 | .421* | .050 |
| Sharpness | .028 | .059 | .065 | .139 | .032 | .075 |
| Simplicity | .110 | .037 | .481* | .103 | .240 | .129 |
| 21. Switch Hook Control | | | | | | |
| Concreteness | .249 | .043 | .251 | .026 | .397* | .283 |
| Familiarity | .411* | .153 | .197 | .129 | .496* | .109 |
| Meaningfulness | .456* | .161 | .311 | .268 | .701* | .535* |
| Relatedness | .515* | .030 | .293 | .031 | .634* | .188 |
| Sharpness | .496* | .164 | .330 | .171 | .428* | .227 |
| Simplicity | .334 | .045 | .379 | .122 | .459* | .579* |
| 22. Transfer Call | | | | | | |
| Concreteness | .268 | .193 | .205 | .441* | .225 | .141 |
| Familiarity | .185 | .184 | .159 | .162 | .091 | .030 |
| Meaningfulness | .172 | .412* | .209 | .347 | .155 | .120 |
| Relatedness | .094 | .433* | .177 | .015 | .261 | .397* |
| Sharpness | .335 | .164 | .034 | .171 | .256 | .250 |
| Simplicity | .013 | .299 | .231 | .351 | .318 | .060 |
| 23. Volume | | | | | | |
| Concreteness | .347 | .260 | .213 | .117 | .389* | .328 |
| Familiarity | .148 | .197 | .274 | .079 | .627* | .280 |
| Meaningfulness | .111 | .200 | .289 | .064 | .518* | .449* |
| Relatedness | .412* | .249 | .321 | .007 | .528* | .384* |
| Sharpness | .025 | .249 | .233 | .044 | .485* | .391* |
| Simplicity | .110 | .185 | .308 | .080 | .584* | .397* |

When the countries were compared in their semantic scales of the icons, the Filipinos gave consistently higher ratings on almost all scales compared to the Swedish subjects (Appendix 2). Further, the semantic scales for concreteness, simplicity and familiarity had the most number of significant correlations to the rate of correct recognition of the icons.

The results of the initial study revealed a few but very important findings. First, the subjects poorly recognized majority of the icons than expected. This may mean that icons as interface per se without any other visual cue, aid or even prior practice to their use is not always helpful, as shown in an earlier study (Paper 1). They need to be tested thoroughly to be sure that they are easily and correctly recognized. Second, although very few differences were noted between countries regarding their recognition rates of the icons, significant differences were present in their certainty and semantic ratings. Filipinos tend to give higher ratings compared to the Swedes who generally were more conservative and gave lower ratings. Filipinos, like most Asians tend to give higher ratings than westerners. This could represent a kind of cultural bias on evaluation tasks and need to be studied further. Nonetheless, it turned out that concreteness and simplicity of the icons were equally, if not more important. These results supported other studies pointing to the relevance concreteness of the design and the function in their recognition. It is easier for subjects to recognize icons or symbols that represent concrete icons compared to abstract ones.

Lastly, regarding culture and compared to studies done by Tudor (1994) using the same icon sets among America subjects, the Filipino subjects (followed by Swedish subjects) had the highest recognition scores for most of the icons designed in Japan. This could be a reflection of the Philippine socio-economic condition where majority of current computerised and electronic appliances is imported from Japan. Icons or other graphical symbols are commonly found in these products. Americans on the other hand are obviously more akin to be exposed to products they have made. Although USA still ranks as the major trading partner of the Philippines followed by Japan, the latter ranks first in terms of the computer and telecom products marketed and used locally. Thus, this renders the view that cultural familiarity and meaning of icons can affect their usability when applied as interface tools in new or modern technologies. Admittedly, there are limitations in this pilot study. First, a distinct group of potential users (educated and computer literate) was used as subjects.

The tests used in all these studies were cued-response tests. The target functions were presented and explained and subjects were asked to match the appropriate icons. Dewar (1994) has pointed out that matching tests or multiple tests could greatly influence results of symbol recognition tests. He considered the spontaneous identification or "open-ended" tests as the "main standard" of measuring symbol understanding and thus must be the basis of comparing other symbol test results. Still, most of the icons in the two studies had very low recognition scores (<67%). To further study the observations made regarding the possible role of culture as well as other possible group differences in recognising symbols, similar and other types of tests would be needed using different cultural groups. Consequently, subjects from Asia, Europe and America were then targeted for the next series of experiments.

## 6.2 Main Study: Southeast Asia, Sweden, USA, Finland (Papers 2 to 7)

The next series of tests involved 240 subjects from 5 Asian countries, 48 subjects from Sweden, 40 subjects from the United States, and 29 subjects from Finland (Total number of subjects: 357). Compared to the earlier study, the main study used multiple test parameters to evaluate three sets of graphical symbols representing the same device (videophone) but for different referents. The different parameters were hit rates for both spontaneous identification and cued response tests, subjective certainties, false alarms or confusions, missing values, preferences, and semantic differential ratings.

### *6.2.1 Exposure to computer-related devices*

The main study was first done in the five Southeast Asian countries (Papers 2 to 4) followed by tests in Sweden, Finland and the USA (Papers 5 to 7). Data on subjects' exposure to computer-related devices as well as opinions regarding technology were also obtained. Tables 7 to 9 contain data on the use of computer and/or computer-related devices (i.e. modern audio-video and telecommunication devices) among the subjects. Except for subjects from Sri Lanka, almost all respondents (>90%) have used computer-related devices for the past 2-5 years.

TABLE 7. Percent of subjects using computer or computer-related devices (n = 48 per country, except U.S.A. where n = 40).

| Country | | Count | Percent |
|---|---|---|---|
| Indonesia | Yes | 48 | 100.0 |
| Malaysia | Yes | 48 | 100.0 |
| Philippines | Yes | 47 | 97.9 |
| | No | 1 | 2.1 |
| | Total | 48 | 100.0 |
| Thailand | Yes | 48 | 100.0 |
| Sri Lanka | Yes | 40 | 83.3 |
| | No | 8 | 16.7 |
| | Total | 48 | 100.0 |
| Sweden | Yes | 48 | 100.0 |
| USA | Yes | 40 | 100.0 |

Concerning actual use of computer, at the time of the study at least 68.8% of all subjects per country use computers from occasionally (at least once a week) to often (at least once a day) (Figure 11). Indonesian subjects followed by Sri Lankan subjects use the computer least often, about 68.8% and 72.9% respectively. All other country groups had more than 90% of the subjects using computers at least once a week.

TABLE 8. Percentage distribution of computer use (n = 48 per country, except U.S.A. where n = 40).

| | | Rarely | Occasionally | Often | Not Applicable | Total |
|---|---|---|---|---|---|---|
| Indonesia | Count | 15.0 | 12.0 | 21.0 | | 48.0 |
| | % | 31.3 | 25.0 | 43.8 | | 100.0 |
| Malaysia | Count | 2.0 | 8.0 | 38.0 | | 48.0 |
| | % | 4.2 | 16.7 | 79.2 | | 100.0 |
| Philippines | Count | 2.0 | 20.0 | 25.0 | 1.0 | 48.0 |
| | % | 4.2 | 41.7 | 52.1 | 2.1 | 100.0 |
| Thailand | Count | 3.0 | 17.0 | 28.0 | | 48.0 |
| | % | 6.3 | 35.4 | 58.3 | | 100.0 |
| Sri Lanka | Count | 5.0 | 23.0 | 12.0 | 8.0 | 48.0 |
| | % | 10.4 | 47.9 | 25.0 | 16.7 | 100.0 |
| Sweden | Count | 2.0 | 14.0 | 32.0 | | 48.0 |
| | % | 4.2 | 29.2 | 66.7 | | 100.0 |
| USA | Count | 2.0 | 16.0 | 22.0 | | 40.0 |
| | % | 5.0 | 40.0 | 55.0 | | 100.0 |

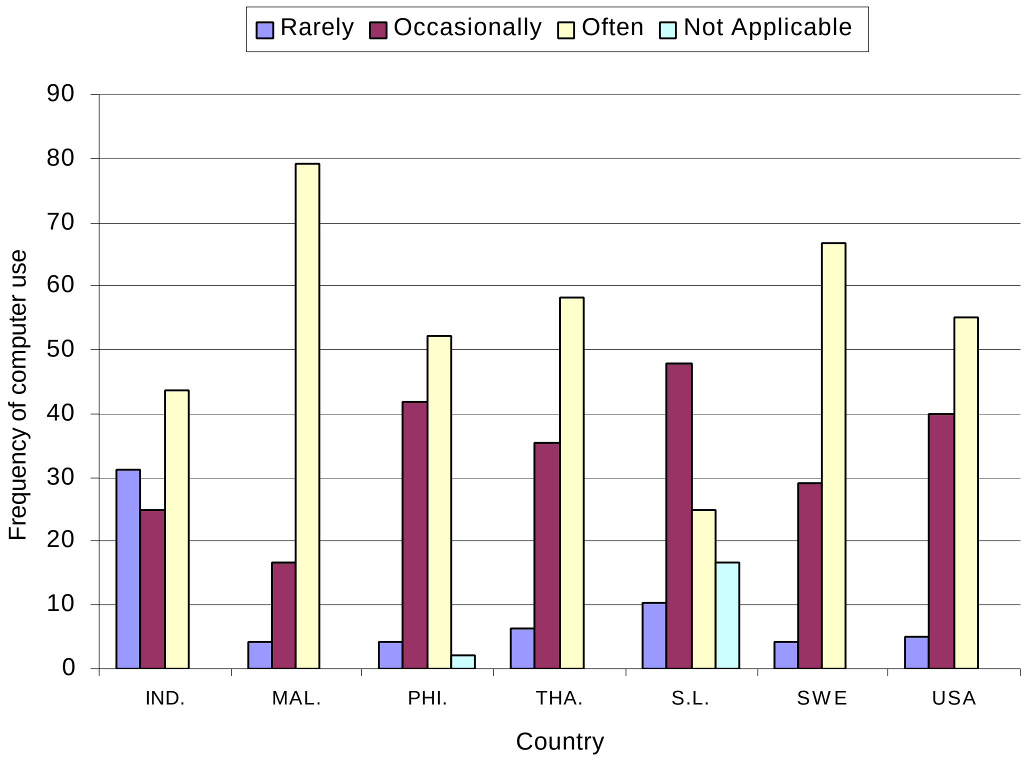



FIGURE 11. Percentage distribution of computer use per country in percent (n = 48 per country, except U.S.A. where n = 40).

Regarding the type of computer interface used by the subjects, all country groups use both keyboard and windows-based interfaces (Table 9) with the latter (windows) as the predominant one being used.

TABLE 9. Computer interface used (n = 48 per country, except U.S.A. where n = 40).

| Country | | Keyboard -based | Graphics/ Windows-based | Both | Not Applicable | Total |
|---|---|---|---|---|---|---|
| Indonesia | Count | 6.0 | 6.0 | 36.0 | | 48.0 |
| | % | 12.5 | 12.5 | 75.0 | | 100.0 |
| Malaysia | Count | 2.0 | 4.0 | 42.0 | | 48.0 |
| | % | 4.2 | 8.3 | 87.5 | | 100.0 |
| Philippines | Count | 7.0 | 8.0 | 32.0 | 1.0 | 48.0 |
| | % | 14.6 | 16.7 | 66.7 | 2.1 | 100.0 |
| Thailand | Count | 6.0 | 4.0 | 38.0 | | 48.0 |
| | % | 12.5 | 8.3 | 79.2 | | 100.0 |
| Sri Lanka | Count | 18.0 | 4.0 | 18.0 | 8.0 | 48.0 |
| | % | 37.5 | 8.3 | 37.5 | 16.7 | 100.0 |
| Sweden | Count | | 14.0 | 34.0 | | 48.0 |
| | % | | 29.2 | 70.8 | | 100.0 |
| USA | Count | 2.0 | 6.0 | 32.0 | | 40.0 |
| | % | 5.0 | 15.0 | 80.0 | | 100.0 |

### *6.2.2 Attitudes towards technology*

Three basic rating scales were used to compare the subject groups' attitudes towards technology (ETSI, 1993). Table 10 and Figures 12-14 contain the results of the ratings. Majority (>50%) of the subjects in each country think that technology gives more advantages and makes life easier in general. However, most Swedish subjects agreed that some aspects of technological progress are worrying compared to most Asian and U.S. subjects who thought otherwise.

TABLE 10. Mean ratings of attitudes towards technology (n = 48 per country, except U.S.A. where n = 40).

| Country | | Mean | S.D. |
|---|---|---|---|
| Indonesia | More Advantages | 3.2 | 1.1 |
| | Easier Life | 3.8 | 1.0 |
| | Technology Worrying | 3.3 | 1.1 |
| Malaysia | More Advantages | 4.0 | 1.0 |
| | Easier Life | 4.3 | 1.0 |
| | Technology Worrying | 3.9 | 1.0 |
| Philippines | More Advantages | 3.8 | 0.9 |
| | Easier Life | 4.3 | 0.9 |
| | Technology Worrying | 3.8 | 0.9 |
| Thailand | More Advantages | 3.9 | 1.0 |
| | Easier Life | 4.3 | 1.0 |
| | Technology Worrying | 3.4 | 1.2 |
| Sri Lanka | More Advantages | 3.6 | 1.3 |
| | Easier Life | 4.4 | 0.8 |
| | Technology Worrying | 3.8 | 1.2 |
| Sweden | More Advantages | 3.8 | 0.8 |
| | Easier Life | 3.5 | 0.9 |
| | Technology Worrying | 3.8 | 1.0 |
| USA | More Advantages | 3.8 | 1.1 |
| | Easier Life | 4.1 | 1.2 |
| | Technology Worrying | 3.4 | 1.3 |

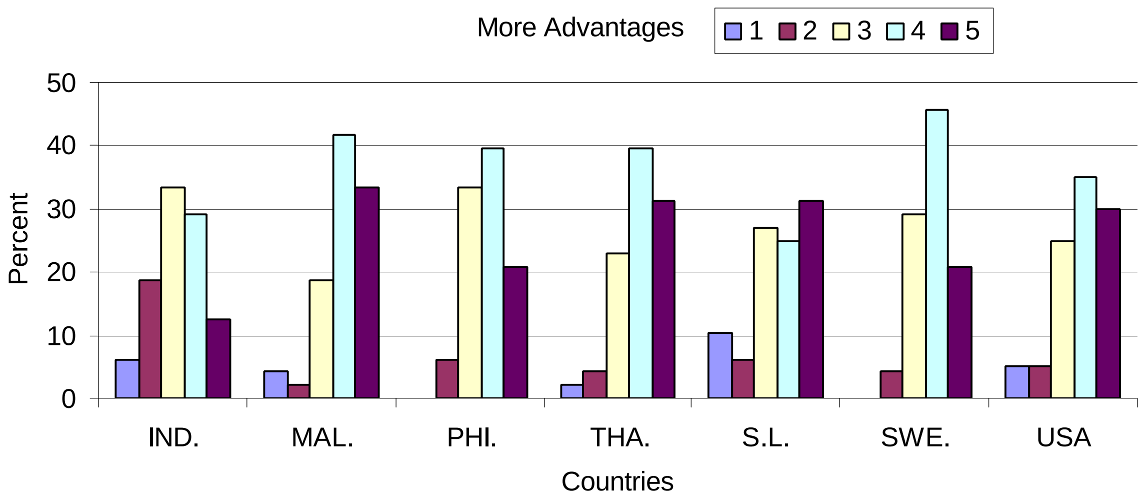



FIGURE 12. Percentage distribution of each country's ratings on technology as having more advantages than disadvantages (1 = do not agree at all, to 5 = totally agree).

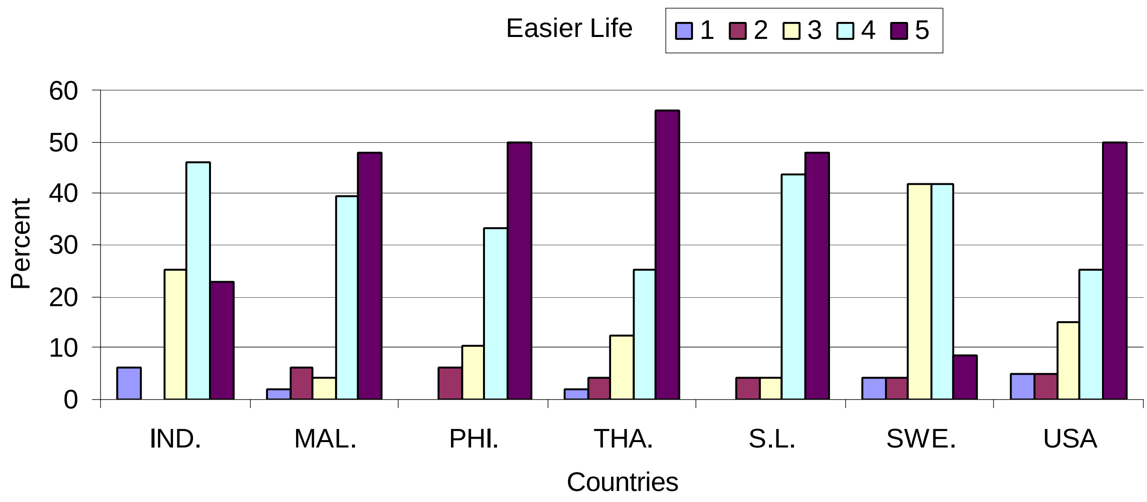



FIGURE 13. Percentage distribution of each country's ratings on technology making life easier (1 = do not agree at all, to 5 = totally agree).

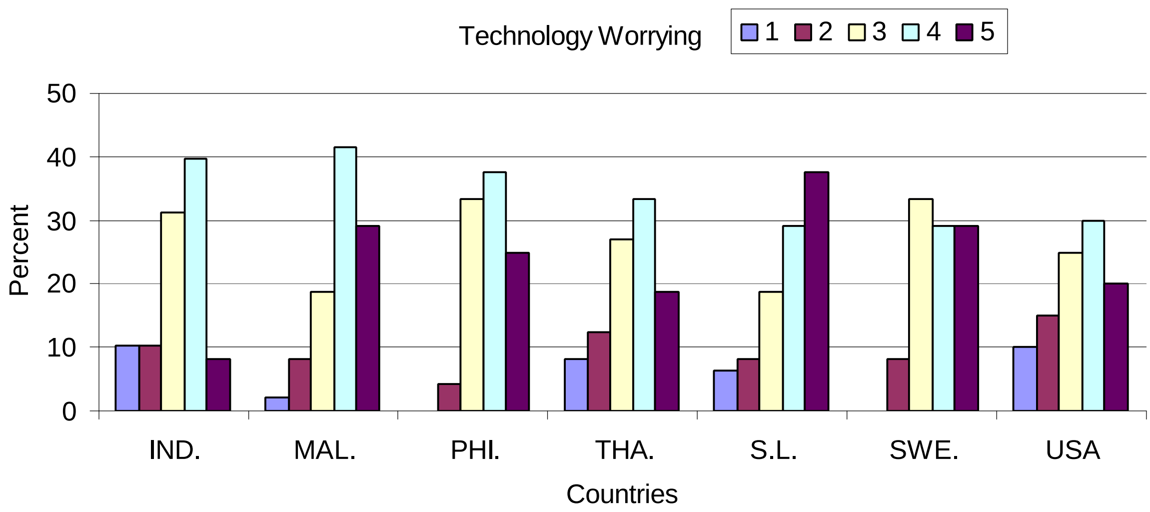



FIGURE 14. Percentage distribution of each country's ratings that current technology is worrying (1 = do not agree at all, to 5 = totally agree).

### *6.2.3 Hit rates*

#### *6.2.3.1 Descriptives*

Hit rates in Part One referred to instances when subjects wrote the word or words that correctly state or described the function or referent of each particular symbol. In turn, Part Two's hit rates were instances when subjects associated the referent in question to the correct videophone symbol among a set of other videophone symbols presented. Tables 11 and 12 contain the results of the hit rates for Part One (Spontaneous identification tests) and Part Two (Cued-response tests) respectively. Based on the mean hit rates alone in both Parts One and Two, the graphical symbols of Set 1 performed best individually and as a set. However, Part One's hit rates were generally lower compared to Part Two, mainly because the former was the harder type of test.

When the 67% of ISO 9186 was used as another basis of performance, very few of the symbols, especially in Part One met this criterion. Between individual symbols, very few reached the minimum 67% criterion. Only Set 1's symbols for "camera" (68.2%) and "microphone" (77.4%), and Set 3's "microphone" (77.4%) had overall mean hit rates above 67% in Part One. However, between sets, none of the 3 sets reached 67% (overall mean for each set in Part One was 40.5%, 16.2%, and 36.5% respectively, Table 11).

TABLE 11. Part One hit rates (%) per country (n = 48 per country, except USA, where n = 40).

| **SET 1** | Ind. | Mal. | Phi. | Thai. | S.L. | Swe. | U.S.A. | *Means* |
|---|---|---|---|---|---|---|---|---|
| [1] Camera 1 | 62.5 | 60.4 | 81.3 | 72.9 | 37.5 | 87.5 | 75.0 | *68.2* |
| [4] Doc. Camera 1 | 31.3 | 25.0 | 35.4 | 20.8 | 4.2 | 58.3 | 25.0 | *28.6* |
| [7] Handsfree 1 | 31.3 | 16.7 | 25.0 | 29.2 | 6.3 | 12.5 | 5.0 | *18.0* |
| [10] Microphone 1 | 77.1 | 75.0 | 83.3 | 85.4 | 60.4 | 70.8 | 90.0 | *77.4* |
| [13] Selfview 1 | 18.8 | 6.3 | 4.2 | 27.1 | 14.6 | 4.2 | 10.0 | *12.2* |
| [16] Still Picture 1 | 25.0 | 18.8 | 20.8 | 31.3 | 12.5 | 12.5 | 10.0 | *18.7* |
| [19] Videophone 1 | 75.0 | 45.8 | 64.6 | 75.0 | 37.5 | 54.2 | 70.0 | *60.3* |
| *Means* | *45.9* | *35.4* | *44.9* | *48.8* | *24.7* | *42.9* | *40.7* | ***40.5*** |
| **SET 2** | | | | | | | | |
| [2] Camera 2 | 43.8 | 52.1 | 58.3 | 52.1 | 37.5 | 87.5 | 80.0 | *58.8* |
| [5] Doc. Camera 2 | 25.0 | 14.6 | 22.9 | 16.7 | 4.2 | 50.0 | 10.0 | *20.5* |
| [8] Handsfree 2 | 6.3 | 10.4 | 0.0 | 14.6 | 6.3 | 12.5 | 0.0 | *7.2* |
| [11] Microphone 2 | 8.3 | 10.4 | 10.4 | 12.5 | 16.7 | 12.5 | 10.0 | *11.5* |
| [14] Selfview 2 | 6.3 | 4.2 | 2.1 | 8.3 | 2.1 | 0.0 | 0.0 | *3.3* |
| [17] Still Picture 2 | 12.5 | 6.3 | 4.2 | 16.7 | 2.1 | 4.2 | 10.0 | *8.0* |
| [20] Videophone 2 | 10.4 | 2.1 | 2.1 | 6.3 | 6.3 | 0.0 | 0.0 | *3.9* |
| *Means* | *16.1* | *14.3* | *14.3* | *18.2* | *10.7* | *23.8* | *15.7* | ***16.2*** |
| **SET 3** | | | | | | | | |
| [3] Camera 3 | 60.4 | 56.3 | 60.4 | 50.0 | 41.7 | 75.0 | 80.0 | *60.5* |
| [6] Doc. Camera 3 | 12.5 | 20.8 | 25.0 | 22.9 | 12.5 | 41.7 | 20.0 | *22.2* |
| [9] Handsfree 3 | 6.3 | 18.8 | 2.1 | 22.9 | 6.3 | 8.3 | 0.0 | *9.2* |
| [12] Microphone 3 | 70.8 | 81.3 | 81.3 | 77.1 | 70.8 | 70.8 | 90.0 | *77.4* |
| [15] Selfview 3 | 10.4 | 8.3 | 8.3 | 16.7 | 14.6 | 12.5 | 10.0 | *11.5* |
| [18] Still Picture 3 | 10.4 | 20.8 | 27.1 | 18.8 | 8.3 | 16.7 | 20.0 | *17.4* |
| [21] Videophone 3 | 66.7 | 43.8 | 60.4 | 60.4 | 54.2 | 58.3 | 55.0 | *57.0* |
| *Means* | *33.9* | *35.7* | *37.8* | *38.4* | *29.8* | *40.5* | *39.3* | ***36.5*** |

*Legend: Ind.=Indonesia, Mal.=Malaysia, Phi.=Philippines, Thai.=Thailand, S.L.=Sri Lanka, Swe.=Sweden, U.S.A.=United States of America.*

TABLE 12. Part Two hit rates (%) per country (n = 48 per country, except USA, where n = 40).

| SET 1 | Ind. | Mal. | Phi. | Thai. | S.L. | Swe. | U.S.A. | *Means* |
|---|---|---|---|---|---|---|---|---|
| [1] Camera 1 | 72.9 | 70.8 | 83.3 | 83.3 | 41.7 | 75.0 | 80.0 | *72.4* |
| [4] Doc. Camera 1 | 91.7 | 91.7 | 81.3 | 87.5 | 75.0 | 95.8 | 90.0 | *87.6* |
| [7] Handsfree 1 | 39.6 | 37.5 | 35.4 | 56.3 | 18.8 | 50.0 | 50.0 | *41.1* |
| [10] Microphone 1 | 83.3 | 87.5 | 89.6 | 89.6 | 66.7 | 75.0 | 95.0 | *83.8* |
| [13] Selfview 1 | 60.4 | 72.9 | 81.3 | 60.4 | 39.6 | 70.8 | 95.0 | *68.6* |
| [16] Still Picture 1 | 43.8 | 77.1 | 66.7 | 50.0 | 37.5 | 62.5 | 65.0 | *57.5* |
| [19] Videophone 1 | 87.5 | 93.8 | 93.8 | 91.7 | 81.3 | 95.8 | 95.0 | *91.3* |
| *Means* | *68.5* | *75.9* | *75.9* | *74.1* | *51.5* | *75.0* | *81.4* | ***71.8*** |
| **SET 2** | | | | | | | | |
| [2] Camera 2 | 54.2 | 62.5 | 62.5 | 62.5 | 41.7 | 70.8 | 70.0 | *60.6* |
| [5] Doc. Camera 2 | 79.2 | 77.1 | 68.8 | 79.2 | 35.4 | 91.7 | 90.0 | *74.5* |
| [8] Handsfree 2 | 41.7 | 27.1 | 31.3 | 45.8 | 14.6 | 66.7 | 40.0 | *38.2* |
| [11] Microphone 2 | 20.8 | 33.3 | 39.6 | 35.4 | 29.2 | 33.3 | 55.0 | *35.2* |
| [14] Selfview 2 | 16.7 | 20.8 | 31.3 | 31.3 | 20.8 | 45.8 | 50.0 | *31.0* |
| [17] Still Picture 2 | 25.0 | 41.7 | 25.0 | 52.1 | 47.9 | 66.7 | 50.0 | *44.1* |
| [20] Videophone 2 | 35.4 | 16.7 | 31.3 | 33.3 | 12.5 | 20.8 | 30.0 | *25.7* |
| *Means* | *39.0* | *39.9* | *41.4* | *48.5* | *28.9* | *56.5* | *55.0* | ***44.2*** |
| **SET 3** | | | | | | | | |
| [3] Camera 3 | 68.8 | 64.6 | 70.8 | 60.4 | 54.2 | 66.7 | 60.0 | *63.6* |
| [6] Doc. Camera 3 | 52.1 | 72.9 | 68.8 | 77.1 | 45.8 | 75.0 | 80.0 | *67.4* |
| [9] Handsfree 3 | 35.4 | 37.5 | 33.3 | 47.9 | 25.0 | 58.3 | 35.0 | *38.9* |
| [12] Microphone 3 | 77.1 | 85.4 | 85.4 | 89.6 | 75.0 | 87.3 | 90.0 | *84.3* |
| [15] Selfview 3 | 64.6 | 62.5 | 79.2 | 68.8 | 39.6 | 91.7 | 80.0 | *69.5* |
| [18] Still Picture 3 | 27.1 | 35.4 | 45.8 | 35.4 | 8.3 | 70.8 | 40.0 | *37.5* |
| [21] Videophone 3 | 85.4 | 87.5 | 79.2 | 77.1 | 75.0 | 87.5 | 85.0 | *82.4* |
| *Means* | *58.6* | *63.7* | *66.1* | *65.2* | *46.1* | *76.8* | *67.1* | ***63.4*** |

*Legend: Ind.=Indonesia, Mal.=Malaysia, Phi.=Philippines, Thai.=Thailand, S.L.=Sri Lanka, Swe.=Sweden, U.S.A.=United States of America.*

In Part Two, there were more symbols with hit rates at least 67%. Set 1, aside from being the only set with an overall mean hit rate above 67% (71,8%), had the most number of symbols meeting this criterion (5 of 7 symbols). These were Set 1's symbols for "camera" (72.4%), "document camera" (87.6%), "microphone" (83.8%), "selfview" (68.6%), and "videophone" (91.3%). Set 3 followed with 4 of its 7 symbols having hit rates at least 67%. These were "document camera" (67.4%), "microphone" (84.3%), "selfview" (69.5%), and "videophone" (82.4%). Set 2 performed the worst both as a set and with its individual symbols. Only its symbol for "document camera" (74.5%) had an overall mean hit rate above 67%.

When studying the hit rates between country groups, all 7 countries fared very poorly in all 3 sets during Part One, with all of them having mean hit rates below 50% (Table 11). In Part Two, the results were comparatively better. Only one country group (Sri Lanka) had a mean hit rate below 67% in Set 1. However, for

Set 3, only Sweden and USA had mean hit rates that were above the 67% criterion. Again, none of the country groups performed well when using Set 2 (all below the 67% criterion).

*6.2.3.2 Statistical Analyses*

The patterns observed in the descriptives were generally confirmed by statistical analyses. For example, regarding tests (Part One versus Part Two) as a *within* factor, hit rates in the cued response tests (Part Two) were significantly higher (at $F(1, 321)$, $p<0.05$) than in the spontaneous identification tests (Part One). Between symbol sets, Set 1 had significantly higher hits in both tests for all referents except for "handsfree" and "selfview" ($F(2, 642)$, $p<0.05$). On the other hand, Set 2 garnered the lowest hits in both Parts One and Two of the test.

*Between* countries, there were significant differences ($F(4, 321)$, $p<0.05$) in hit rates for symbols for "camera", document camera", "handsfree", "selfview", and "videophone". In these symbols, Sri Lankan subjects generally had the lowest hits compared to the other six countries. Focusing on the Swedish and American subjects, some similarities and differences were likewise noted. For example, in the identification tests (Parts One and Two), Set 1 symbols generally had the higher hit rates in both countries. However, both countries performed poorly in the spontaneous identification tests (Part One). For example, subjects from both countries managed to attain the Organization for International Standardization's minimum correct recognition rate of 67% in only five of the 21 symbols tested. Compared to the Southeast Asian studies using the same symbol sets, although the Swedish and American subjects fared quite better than their Asian counterparts especially in Part One, the scores were still mostly below the critical 67%. As discussed earlier, spontaneous identification tests simulate, albeit, in a limited manner, the 'first-time encounter' of an interface. The poor results thus strongly suggest that without sufficient learning opportunities and learning aids, symbols may be harder to understand by even subjects or targeted users from western countries.

### *6.2.4 Confusions and missing values*

Confusions or false alarms were instances when a symbol was associated to the wrong referents. They were derived during the cued response test in Part Two (Tables 13 to 19). For example, in the Indonesian group during the cued-response test and testing Set 1 symbols, when the referent camera was presented, 10% of the subjects wrongly selected symbol for document camera.

Tables 13 to 19 clearly showed that Set 2 had the most instances of false alarms and Set 1 the least. Across symbol sets, "selfview" and "videophone" symbols had the highest number of false alarms.

TABLE 13. Confusion Matrices and Missing Values, Indonesia, n = 48.

| **Symbol Selected** | **Referent Presented Set 1** | | | | | | | **Set 2** | | | | | | | **Set 3** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | - | 2 | 2 | | 2 | 5 | 3 | - | 3 | 2 | | 3 | 2 | 7 | - | 5 | 4 | 1 | 1 | 2 | 2 |
| **2**. Doc. Camera | 5 | - | | | | 4 | 1 | 7 | - | | | 7 | 6 | 4 | | - | 1 | | 2 | 7 | |
| **3**. Handsfree | | | - | 7 | | | | | | - | 18 | 1 | | 2 | | 1 | - | 7 | 3 | | 1 |
| **4**. Microphone | | | 13 | - | | | | 1 | 2 | 7 | - | 1 | | | | | 6 | - | | 1 | |
| **5**. Selfview | 1 | | 1 | | - | 17 | 2 | 1 | 2 | 1 | 8 | - | 4 | 3 | 6 | 6 | 3 | | - | 10 | 4 |
| **6**. Still Picture | 2 | 1 | | | 8 | - | | 3 | | 7 | 3 | 8 | - | 7 | 7 | 6 | | | 1 | - | |
| **7**. Videophone | 2 | | 10 | 1 | 5 | 1 | - | 10 | 3 | 4 | 2 | 8 | 21 | - | 2 | 4 | 16 | 2 | 10 | 15 | - |
| *Missing Values* | *3* | *4* | | *3* | | *1* | | | | *7* | *7* | *12* | *3* | *8* | | *1* | *1* | *1* | | | |

TABLE 14. Confusion Matrices and Missing Values, Malaysia, n = 48.

| **Symbol Selected** | **Referent Presented Set 1** | | | | | | | **Set 2** | | | | | | | **Set 3** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | - | 2 | 2 | | 1 | 1 | 2 | - | 6 | 3 | 3 | 5 | 2 | 4 | - | 6 | 6 | | | 2 | 1 |
| **2**. Doc. Camera | 7 | - | 1 | | | 1 | | 2 | - | 1 | 1 | 8 | 7 | 10 | 1 | - | 1 | | | 3 | 2 |
| **3**. Handsfree | 1 | | - | 4 | 1 | 1 | | 1 | | - | 15 | 2 | | 4 | 1 | | - | 5 | 1 | 1 | |
| **4**. Microphone | | | 12 | - | | | | 3 | 1 | 15 | - | 1 | 1 | 6 | | | 4 | - | | | 1 |
| **5**. Selfview | 3 | 2 | 5 | | - | 8 | 1 | | | 9 | 11 | - | 3 | 12 | 6 | 6 | 1 | | - | 17 | 2 |
| **6**. Still Picture | 1 | | 3 | | 7 | - | | 2 | 2 | 2 | 1 | 4 | - | 4 | 9 | 1 | | | | - | |
| **7**. Videophone | 2 | | 7 | 2 | 4 | | - | 10 | 2 | 4 | | 18 | 15 | - | | | 17 | 2 | 17 | 8 | - |
| *Missing Values* | | | | | | | | | | 1 | 1 | | | | | | 1 | | | | |

TABLE 15. Confusion Matrices and Missing Values, Philippines, n = 48.

| **Symbol Selected** | **Referent Presented Set 1** | | | | | | | **Set 2** | | | | | | | **Set 3** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | - | 5 | | | | 1 | 1 | - | 4 | | | 2 | 5 | 2 | - | 4 | 2 | | 1 | 4 | 2 |
| **2**. Doc. Camera | 6 | - | 4 | | | 3 | 1 | 4 | - | 2 | | 1 | 7 | 5 | 3 | - | 1 | | | 2 | 5 |
| **3**. Handsfree | | 1 | - | 4 | 2 | | | | 1 | - | 16 | 3 | | 5 | 1 | | - | 3 | | | |
| **4**. Microphone | | | 8 | - | | | | 1 | 2 | 11 | - | | 2 | 2 | | | 12 | - | | | |
| **5**. Selfview | 1 | | 4 | | - | 11 | 1 | 3 | | 5 | 10 | - | 10 | 12 | 4 | 5 | 3 | 2 | - | 17 | 3 |
| **6**. Still Picture | 1 | 2 | 3 | | 2 | - | | 3 | 1 | 7 | 3 | 8 | - | 7 | 5 | 5 | 3 | | 1 | - | |
| **7**. Videophone | | 1 | 12 | 1 | 5 | 1 | - | 6 | 7 | 7 | | 18 | 12 | - | 1 | 1 | 11 | 1 | 8 | 3 | - |
| *Missing Values* | | | | | | | | 1 | | 1 | | 1 | | | | | | 1 | | | |

TABLE 16. Confusion Matrices and Missing Values, Thailand, n = 48.

| Symbol | Referent Presented | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Set 1 | | | | | | | Set 2 | | | | | | | Set 3 | | | | | | |
| **Selected** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | - | 5 | 3 | 1 | 4 | 4 | 3 | - | 7 | 1 | | 3 | 1 | 6 | - | 9 | 2 | | 3 | 7 | 6 |
| **2**. Doc. Camera | 4 | - | | | 1 | 3 | 1 | 5 | - | 1 | | 4 | 4 | 3 | | - | | | 1 | | 2 |
| **3**. Handsfree | | | - | 4 | | | | | | - | 18 | | 2 | 1 | 1 | | - | 4 | | 2 | |
| **4**. Microphone | | | 8 | - | | | | | | 16 | - | 2 | | 6 | | | 11 | - | | | |
| **5**. Selfview | | | 3 | | - | 15 | | 1 | 1 | 3 | 7 | - | 3 | 4 | 4 | | 2 | | - | 16 | 3 |
| **6**. Still Picture | 2 | 1 | 1 | | 9 | - | | 3 | 2 | 1 | | 1 | - | 10 | 12 | 2 | 1 | | | - | |
| **7**. Videophone | 2 | | 4 | | 5 | 2 | - | 9 | | 2 | 3 | 20 | 13 | - | 1 | | 8 | 1 | 11 | 6 | - |
| *Missing Values* | | | 2 | | | | | | | 2 | 3 | 3 | | 2 | 1 | | 1 | | | | |

TABLE 17. Confusion Matrices and Missing Values, Sri Lanka, n = 48.

| Symbol | Referent Presented | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Set 1 | | | | | | | Set 2 | | | | | | | Set 3 | | | | | | |
| **Selected** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | - | 4 | 6 | 1 | 2 | 3 | 1 | - | 9 | 2 | | 4 | 2 | 4 | - | 15 | 4 | 3 | 3 | 2 | 1 |
| **2**. Doc. Camera | 5 | - | 4 | 1 | 4 | 3 | 4 | 9 | - | 5 | 5 | 5 | 8 | 19 | 1 | - | 6 | 2 | 7 | 7 | 7 |
| **3**. Handsfree | 4 | 2 | - | 8 | 2 | 2 | 1 | | 2 | - | 9 | 1 | 2 | 2 | 1 | 2 | - | 2 | 3 | 2 | 1 |
| **4**. Microphone | | 1 | 10 | - | | | | 3 | 8 | 10 | - | 4 | 1 | 4 | | | 5 | - | 1 | | 1 |
| **5**. Selfview | 8 | 1 | 7 | 1 | - | 21 | 3 | 4 | | 12 | 15 | - | 1 | 8 | 6 | 4 | 7 | 5 | - | 20 | 2 |
| **6**. Still Picture | 4 | 2 | 2 | 1 | 15 | - | | 3 | 3 | 9 | 3 | 9 | - | 5 | 10 | 3 | | | | - | |
| **7**. Videophone | 7 | 2 | 10 | 4 | 6 | 1 | - | 9 | 9 | 3 | 2 | 15 | 11 | - | 4 | 2 | 14 | 0 | 15 | 12 | - |
| *Missing Values* | | | | | | | | | | | | | | | | | | | | | |

TABLE 18. Confusion Matrices and Missing Values, Sweden, n = 48.

| Symbol | Referent Presented | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Set 1 | | | | | | | Set 2 | | | | | | | Set 3 | | | | | | |
| **Selected** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | - | 2 | | | | | 2 | - | 2 | | | 4 | | 8 | - | 8 | 6 | | | | 2 |
| **2**. Doc. Camera | 3 | - | 3 | | 3 | | | 1 | - | | | | 6 | 8 | | - | 2 | | | 4 | 2 |
| **3**. Handsfree | | | - | 7 | | | | | | - | 17 | | | | 2 | | - | 6 | | | |
| **4**. Microphone | | | 9 | - | | | | | | 8 | - | 3 | | 8 | | | 8 | - | | | |
| **5**. Selfview | | | 3 | 1 | - | 13 | | | 2 | 2 | 8 | - | 4 | 2 | 4 | | | | - | 10 | 2 |
| **6**. Still Picture | 3 | | | | 11 | - | | 3 | | 2 | 5 | 4 | - | 8 | 10 | 2 | | | | - | |
| **7**. Videophone | 6 | | 9 | 4 | | 5 | - | 10 | | 2 | | 13 | 6 | - | | | 2 | | 4 | | - |
| *Missing Values* | | | | | | | | | | 2 | 2 | 2 | | 4 | | 2 | 2 | | | | |

TABLE 19. Confusion Matrices and Missing Values, U.S.A., n = 40.

| Symbol Presented | Referent Presented Set 1: 1 | 2 | 3 | 4 | 5 | 6 | 7 | Set 2: 1 | 2 | 3 | 4 | 5 | 6 | 7 | Set 3: 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1**. Camera | - | 2 | 3 | | | 3 | 2 | - | | 4 | | 4 | | 2 | - | 4 | 5 | | 2 | | 1 |
| **2**. Doc. Camera | | - | 2 | | | 7 | | | - | | | 3 | 8 | 9 | | - | | | | 9 | |
| **3**. Handsfree | 3 | | - | 2 | | | | | | - | 6 | 3 | | 2 | | | - | 4 | | 2 | |
| **4**. Microphone | | | 7 | - | | | | 2 | | 8 | - | | | 2 | | | 13 | - | | | |
| **5**. Selfview | 2 | 2 | 3 | | - | 4 | | | 2 | | 2 | - | | 7 | 2 | 2 | 2 | | - | 11 | 5 |
| **6**. Still Picture | | | | | 1 | - | | | | 8 | 8 | 3 | - | 4 | 10 | 2 | | | | - | |
| **7**. Videophone | 3 | | 3 | | 1 | | - | 10 | 2 | 4 | | 7 | 10 | - | 4 | | 6 | | 6 | 2 | - |
| *Missing Values* | | | 2 | | | | | | | | 2 | | 2 | 2 | | | | | | | |

TABLE 20. Missing Values or no answers in percent (%) in Cued Response Tests, per symbol set (n = 48 per country).

| Countries | CAMERA 1 | 2 | 3 | DOC. CAMERA 1 | 2 | 3 | HANDS-FREE 1 | 2 | 3 | MICRO-PHONE 1 | 2 | 3 | SELF-VIEW 1 | 2 | 3 | STILL PICTURE 1 | 2 | 3 | VIDEO-PHONE 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Indonesia | 6 | – | – | 2 | – | 2 | 6 | 15 | 2 | – | 15 | 2 | 8 | 23 | – | – | 6 | – | – | 17 | – |
| Malaysia | – | – | – | – | – | – | – | 2 | – | – | – | – | – | – | – | – | – | – | – | – | – |
| Philippines | – | 2 | – | – | – | – | – | 2 | – | – | – | 2 | – | 2 | – | – | – | – | – | – | – |
| Thailand | – | – | 2 | – | – | – | 4 | 2 | 2 | – | 6 | – | – | 6 | – | – | – | – | – | 4 | – |
| Sri Lanka | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| Sweden | – | – | – | – | – | 4 | – | 4 | 4 | – | 4 | – | – | 4 | – | – | – | – | – | 8 | – |
| U.S.A. | – | – | – | – | – | – | 5 | – | – | – | 5 | – | – | – | – | – | 5 | – | – | 5 | – |

When symbols appear simultaneously or as a set and the main task is to pick out which among these symbols correspond to a desired function or object, there are other possible outcomes aside from a hit. These are the so-called non-hit outcomes or parameters such as misses, confusions and missing values (failure to respond or no answer). The wrong answers during the cued response tests were analysed using confusion matrices (Tables 13 to 19). In this particular part of the study, tabulations were made on the number of instances of confusions – the instance wherein each symbol was selected in the context of a different referent. The results showed that in the seven countries studied, the symbols of Sets 1 and 3 had the highest hit rates (with the former usually with higher hit rates than the latter) with hits above ISO ≥ 67%. The results of the confusion matrices could be used to study which among the symbols in these two sets are more distinct, i.e., less confused with the other symbols in the same set. For example, if hit rates would be the sole basis of performance, then Set 1 would be deemed the best set due to the higher mean hit rates of its symbols. However, since the symbols are designed to appear simultaneously as a set, distinctness from the other symbols is important. In other words, each symbol should be manifest and not easily confused with the other symbols. When the symbols were compared between sets across all countries (Tables 21 to 23), Set 1 had the least number of confusions for "only" 4 of its seven symbols ("camera", "selfview", "still picture", and "videophone"). Set 3's symbols had the least number of confusions for "document camera" and "microphone", making them equally suitable for these referents since their hit rates were also even above ISO's 67% level. In addition, studying instances of false alarms or confusions can also help in two ways. It can help decide which among symbols with similar hit rates are better (the one with lesser false alarms, or less like to be confused with the other symbols), and they can help identify different confusion patterns between user (subject groups) and thus different means of solving the problems of wrong symbol associations as discussed below.

Confusions can be classified into two types: symmetric and asymmetric (Nolan, 1989). The former occurs when subjects chose symbol *x* when presented with referent *y*, and symbol *y* when presented with referent *x*. Symmetric confusions usually suggest visual or conceptual similarities. These were exemplified by the "selfview"/"still picture" (visual similarity) and ""handsfree"/"microphone" symbols of Set 1 (conceptual similarity). Another case of symmetric confusion due to visual similarities would be that of the "camera"/"document camera " symbols of Set 3. These symmetric confusions were present in almost all countries. Thus, the problem may not be "culturally-linked" but a design problem – that the symbols concerned were visually too similar. Moreover, according to basic ergonomic design principles, these symbols would then be prime considerations for re-design or replacement to make them more distinct from the other.

Asymmetric confusions occur when subjects simply chose the wrong symbol for a given referent. For example, Table 13 showed that Set 1's symbol for "document camera" (1) was wrongly selected by respondents from Indonesia when the referents "camera" (5), "still picture" (4) and "videophone" (1) was presented. Between sets, Set 2 clearly had the most number of confusions above this level and Set 1 the least in all seven countries. Across symbol sets, "selfview" and "videophone" symbols had the highest instances of being mistaken as representing other functions. More differences were noted when the countries were compared based on symbol confusions. For example, while Thailand had the most instances of confusing Set 1's symbol for "camera" than the other four countries, it had the lowest instances of confusing Set 1's videophone symbol for other functions. In asymmetric confusions, the problem may lie in the vagueness or too much generality of the symbol/s in question with regards to the other referents. Thus, they can easily be associated with several referents. The results above showed the countries with varying patterns of asymmetric confusions. These could be useful in deciding which symbol (or referent) in each country may need more explanations and tests to avoid vagueness and misconceptions of the functions being represented.

Regarding the two western subject groups, again, Set 1 had the least instances of significant confusions in both countries. However in Set 1, Swedish subjects tend to have more problems confusing the handsfree function as represented by the symbols for videophone (9) and microphone (9). In turn, the Americans confused the handsfree more with the microphone function (7) (Tables 18 and 19). The Swedish group showed symmetrical confusions between the selfview and still picture symbols of Set 1 (11, 13). These were not observed with the US subjects.

When confusions are concentrated to one or a few referents, they also represent the degree of confusing a symbol to another symbol presented at the same time. For example, Set 1's microphone symbol was often associated with the referent handsfree (Indonesia and Malaysia). Such situations could mean that the symbols were be so similar to each other that a re-design is on order to make them more distinct.

TABLE 21. Total number of confusions or false alarms per country in Set 1

| | Countries | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Set 1 | Ind. | Mal. | Phi. | Thai. | S.L. | Swe. | USA | Total |
| Camera | 14 | 8 | 7 | 20 | 17 | 4 | 10 | 80 |
| Doc. Camera | 10 | 9 | 14 | 9 | 21 | 9 | 9 | 81 |
| Handsfree | 7 | 7 | 7 | 4 | 19 | 7 | 5 | 56 |
| Microphone | 13 | 12 | 8 | 8 | 11 | 9 | 7 | 68 |
| Selfview | 21 | 19 | 17 | 18 | 41 | 17 | 4 | 137 |
| Still Picture | 11 | 11 | 8 | 13 | 24 | 14 | 1 | 82 |
| Videophone | 19 | 15 | 20 | 13 | 30 | 24 | 7 | 128 |
| Total | 95 | 81 | 81 | 85 | 163 | 84 | 43 | 632 |

TABLE 22. Total number of confusions or false alarms per country in Set 2.

| | Countries | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Set 2 | Ind. | Mal. | Phi. | Tha. | S.L. | Swe. | USA | Total |
| Camera | 17 | 23 | 13 | 18 | 21 | 14 | 10 | 116 |
| Doc. Camera | 24 | 29 | 19 | 17 | 51 | 15 | 20 | 175 |
| Handsfree | 21 | 22 | 25 | 21 | 16 | 17 | 11 | 133 |
| Microphone | 11 | 27 | 18 | 24 | 30 | 19 | 12 | 141 |
| Selfview | 19 | 35 | 40 | 19 | 40 | 18 | 11 | 182 |
| Still Picture | 28 | 15 | 29 | 17 | 32 | 22 | 4 | 147 |
| Videophone | 48 | 49 | 50 | 47 | 49 | 31 | 33 | 307 |
| Total | 168 | 200 | 194 | 163 | 239 | 136 | 101 | 1201 |

TABLE 23. Total number of confusions or false alarms per country in Set 3

| | Countries | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Set 3 | Ind. | Mal. | Phi. | Tha. | S.L. | Swe. | USA | Total |
| Camera | 15 | 15 | 13 | 27 | 28 | 16 | 12 | 126 |
| Doc. Camera | 10 | 7 | 11 | 3 | 30 | 8 | 9 | 78 |
| Handsfree | 12 | 8 | 4 | 7 | 11 | 8 | 6 | 56 |
| Microphone | 7 | 5 | 12 | 11 | 7 | 8 | 0 | 50 |
| Selfview | 29 | 32 | 34 | 25 | 44 | 16 | 16 | 196 |
| Still Picture | 14 | 10 | 14 | 15 | 13 | 12 | 12 | 90 |
| Videophone | 49 | 44 | 25 | 27 | 47 | 6 | 18 | 216 |
| Total | 136 | 121 | 113 | 115 | 180 | 74 | 73 | 812 |

Lastly, missing values were instances when some of the subjects gave no response or answers during the cued response tests. Tables 13 to 19 illustrates the distribution of the missing values across countries for each set. Table 20 summarises the total missing values under each set for each country. There were no major differences between countries for missing values in Part One. Between sets, however, Set 2 had the highest instances of missing values. Between countries, Indonesia, Thailand and Sweden had the most instances of missing values among the seven countries. Sri Lanka was the opposite with no missing values in the cued response test. Furthermore, although Sri Lanka had the lowest hits for some of the symbols in both tests, it had the lowest missing values, especially when compared to Indonesia and Thailand. This could suggest a type of response bias and/or differences in understanding test procedures not reflected in the other parameters.

When arrayed with the other non-hit parameters or errors, missing values are also very important since they represent situations wherein the prospective user plainly lacks the knowledge of which among the symbols represent the desired function (referent). In actual user scenarios, they can be akin to non-use or under-use of the device.

### *6.2.5 Combining the objective performance parameters*

As discussed earlier, it would be helpful to simultaneously consider the different objective parameters to be able to compare the videophone symbols more. Table 24 shows how the videophone symbols performed against each other (different symbols from the 3 sets for each referent) based on the three objective parameters that could be derived from Part Two. The best performing videophone symbols were then determined by combining these three objective parameters (Table 25). The results showed that in majority of the referents and among most countries, the symbols belonging to Set 1 performed best. That is, symbols from Set 1 were correctly associated to the intended function or referent, least confused to represent another referent, and least number of no answers (possibility of not being understood). Understandably, for some countries symbols from Set 3 were equally good (e.g., Set 3's document camera, handsfree, microphone, still picture, and videophone). This could be noted for the countries of Indonesia, Malaysia, Sri Lanka and Sweden.

TABLE 24. Videophone symbols' performances per country group based on the three objective parameters in Part Two. For each referent, the symbol with more ✓ signifies better performance (i.e. higher hit rates, lower number of false alarms, lower number of missing values).

| | | **Indonesia** | | | **Malaysia** | | | **Philippines** | | | **Thailand** | | | **Sri Lanka** | | | **Sweden** | | | **USA** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Referent | **Set** | Hit | FA | MV | Hit | FA | MV | Hit | FA | MV | Hit | FA | MV | Hit | FA | MV | Hit | FA | MV | Hit | FA | MV |
| Camera | **1** | ✓✓ | ✓✓ | | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓ | ✓ | | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ |
| | **2** | | | ✓ | | | ✓ | | ✓ | | | ✓✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| | **3** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ |
| Doc. Camera | **1** | ✓✓ | ✓✓ | | ✓✓ | ✓ | ✓ | ✓✓ | ✓ | ✓ | ✓✓ | ✓ | ✓ | ✓ | ✓✓ | ✓ | ✓✓ | ✓ | ✓ | ✓✓ | ✓✓ | ✓ |
| | **2** | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓✓ | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | ✓✓ | | ✓ |
| | **3** | | ✓✓ | | ✓ | ✓✓ | ✓ | ✓ | | ✓ | ✓ | ✓✓ | ✓ | | ✓ | ✓ | ✓ | ✓✓ | | ✓ | ✓✓ | ✓ |
| Handsfree | **1** | | ✓✓ | | | ✓✓ | ✓ | | ✓ | ✓ | ✓ | ✓✓ | | | | ✓ | ✓ | ✓✓ | ✓ | ✓ | ✓✓ | |
| | **2** | | | | | | | | | | | | | | ✓ | ✓ | ✓✓ | | | | | ✓ |
| | **3** | | | | | ✓✓ | ✓ | | ✓✓ | ✓ | | ✓ | | | ✓✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ |
| Microphone | **1** | ✓✓ | | ✓ | ✓✓ | ✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓✓ | ✓✓ | ✓ |
| | **2** | | ✓ | | | | ✓ | | | ✓ | | | | | | ✓ | | | | | | |
| | **3** | ✓ | ✓✓ | | ✓ | ✓✓ | ✓ | ✓ | ✓ | | ✓✓ | ✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓ | ✓ | ✓ |
| Selfview | **1** | ✓ | ✓ | | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓ | ✓✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓✓ | ✓✓ | ✓ |
| | **2** | | ✓✓ | | | | ✓ | | | | | ✓✓ | | | ✓✓ | ✓ | | | | | ✓ | ✓ |
| | **3** | ✓✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓✓ | | ✓ | | | ✓ | ✓✓ | ✓✓ | ✓ | ✓ | | ✓ |
| Still Picture | **1** | | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓ | ✓✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| | **2** | | | | | | ✓ | | | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | ✓ | ✓ | |
| | **3** | | ✓ | ✓ | | ✓✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | | | ✓ |
| Videophone | **1** | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓✓ | ✓ | ✓✓ | ✓ | ✓ | ✓✓ | ✓✓ | ✓ |
| | **2** | | | | | | ✓ | | | ✓ | | | | | | ✓ | | | | | | |
| | **3** | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓✓ | ✓ | ✓ | ✓ | ✓ |

*Legend: Ind.=Indonesia, Mal.=Malaysia, Phi.=Philippines, Thai.=Thailand, S.L.=Sri Lanka, Swe.=Sweden, U.S.A.=United States of America.*

TABLE 25. Best videophone symbols when combining hit rates, false alarms (confusions) and missing values (no answers), per country. Numbers indicate Symbol Set.

| | Best performing symbols for each referent tested using the three symbol sets (Sets 1 ,2 and 3) | | | | | | |
|---|---|---|---|---|---|---|---|
| Referent | Ind. | Mal. | Phi. | Thai. | S.L. | Swe. | USA |
| Camera | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Doc. Camera | 1 | 1 & 3 | 1 | 1 & 3 | 1 | 1 | 1 |
| Handsfree | 1 | 1 & 3 | 1 & 2 | 1 | 3 | 1 | 1 |
| Microphone | 1 & 3 | 1 & 3 | 3 | 1 | 3 | 1 | 1 |
| Selfview | 3 | 1 | 1 | 1 | 2 | 1 | 1 |
| Still Picture | 1 | 1 | 1 | 1 | 3 | 3 | 1 |
| Videophone | 1 | 1 | 1 | 1 | 1 | 3 | 1 |

*Legend: Ind.=Indonesia, Mal.=Malaysia, Phi.=Philippines, Thai.=Thailand, S.L.=Sri Lanka, Swe.=Sweden, U.S.A.=United States of America.*

### *6.2.6 Subjective certainties*

Subjective certainties (Tables 26 and 27) reflected the subjects' degree of confidence or confusion with their answers and could thus help in weighing the hit rates between symbols and symbol sets. They also helped reveal a possible cultural bias (observed in one country group) that is important when interpreting data wherein subjective factors can play an important role.

Data on mean subjective certainty ratings indicate how the tested symbols compare to each other in terms of the level of confidence the subjects felt when choosing them. Confidence judgements (mean subjective certainty ratings) were mostly higher in Sets 1 and 3 compared to Set 2 in both Parts One and Two (Tables 26 and 27), with those in Part Two slightly higher than Part One. Confidence judgements for the symbols for "microphone" and "videophone" were either very high (Sets 1 and 3) or very low (Set 2) in both Parts One and Two. The above patterns were quite expected. First, the symbol-referent identification task in Part One was generally more difficult than the referent-symbol association task in Part Two. Hence, the former tended to have lower ratings (subjects had lower levels of certainty in their answers) than the latter. Second, Set 2's symbols are mostly abstract compared to the graphical symbols of Sets 1 and 3. This contributed further to the difficulty in identifying or matching the target referents with Set 2's symbols.

Between country groups, it was also interesting to note some differences in trends with regards to hit rates and confidence judgements (subjective certainties). Swedish subjects as a group tended to be more conservative (i.e. lower mean ratings) in Part One compared to the six other country groups. Sri Lanka was the opposite. Although it had the lowest mean hit rates in Parts One and Two, it exhibited higher subjective ratings than most of the other countries.

TABLE 26. Part One results on Confidence Judgements (Mean Subjective Certainty Ratings) per country (n = 48 per country, except USA, where n = 40)

| **SET 1** | Ind. | Mal. | Phi. | Tha. | S.L. | Swe. | USA. | *Means* |
|---|---|---|---|---|---|---|---|---|
| [1] Camera 1 | 4.5 | 5.3 | 5.0 | 5.4 | 4.7 | 5.2 | 5.7 | *5.1* |
| [4] Doc. Camera 1 | 3.8 | 3.4 | 3.5 | 3.0 | 3.3 | 3.5 | 3.8 | *3.3* |
| [7] Handsfree 1 | 4.6 | 4.6 | 4.8 | 4.2 | 5.5 | 4.6 | 4.1 | *4.6* |
| [10] Microphone 1 | 5.2 | 5.4 | 5.6 | 5.8 | 5.8 | 5.4 | 5.4 | *5.5* |
| [13] Selfview 1 | 4.2 | 4.1 | 4.3 | 3.8 | 5.1 | 2.0 | 4.3 | *3.9* |
| [16] Still Picture 1 | 4.8 | 4.1 | 4.4 | 4.9 | 4.8 | 3.0 | 5.1 | *4.4* |
| [19] Videophone 1 | 5.7 | 5.1 | 5.6 | 5.3 | 6.8 | 4.3 | 4.9 | *5.1* |
| *Means* | *4.6* | *4.5* | *4.6* | *4.6* | *4.8* | *4.0* | *4.7* | ***4.6*** |
| **SET 2** | | | | | | | | |
| [2] Camera 2 | 4.1 | 4.8 | 4.4 | 4.6 | 4.6 | 4.2 | 5.1 | *4.5* |
| [5] Doc. Camera 2 | 3.4 | 3.5 | 3.2 | 2.9 | 3.8 | 3.8 | 3.9 | *3.4* |
| [8] Handsfree 2 | 5.0 | 4.9 | 4.7 | 5.0 | 4.6 | 5.3 | 5.3 | *4.9* |
| [11] Microphone 2 | 2.2 | 2.0 | 2.1 | 2.6 | 3.3 | 1.5 | 2.5 | *2.0* |
| [14] Selfview 2 | 2.6 | 3.6 | 2.8 | 2.7 | 2.9 | 2.4 | 3.4 | *2.7* |
| [17] Still Picture 2 | 3.3 | 3.8 | 3.0 | 3.5 | 3.8 | 2.1 | 4.0 | *3.2* |
| [20] Videophone 2 | 4.0 | 4.2 | 3.9 | 3.8 | 4.1 | 3.3 | 4.7 | *3.9* |
| *Means* | *3.3* | *3.6* | *3.3* | *3.5* | *3.7* | *3.2* | *4.1* | ***3.5*** |
| **SET 3** | | | | | | | | |
| [3] Camera 3 | 4.0 | 4.4 | 5.1 | 4.7 | 4.4 | 4.6 | 5.2 | *4.6* |
| [6] Doc. Camera 3 | 4.0 | 3.5 | 3.9 | 4.3 | 4.0 | 3.5 | 3.8 | *3.7* |
| [9] Handsfree 3 | 5.2 | 5.0 | 5.1 | 4.8 | 4.3 | 5.3 | 5.5 | *5.0* |
| [12] Microphone 3 | 5.4 | 5.5 | 5.8 | 5.8 | 5.2 | 5.0 | 5.7 | *5.3* |
| [15] Selfview 3 | 4.7 | 4.4 | 4.7 | 4.1 | 4.5 | 3.4 | 4.3 | *4.2* |
| [18] Still Picture 3 | 5.6 | 5.0 | 5.3 | 5.3 | 5.0 | 4.4 | 5.6 | *5.2* |
| [21] Videophone 3 | 5.3 | 4.8 | 5.5 | 5.1 | 4.8 | 4.3 | 5.7 | *5.1* |
| *Means* | *4.8* | *4.6* | *5.0* | *4.8* | *4.5* | *4.3* | *5.1* | ***4.7*** |

*Legend: Ind.=Indonesia, Mal.=Malaysia, Phi.=Philippines, Tha.=Thailand, S.L.=Sri Lanka, Swe.=Sweden, U.S.A.=United States of America.*

TABLE 27. Part Two results on Confidence Judgements (Mean Subjective Certainty Ratings) per country (n = 48 per country, except USA, where n = 40)

| **SET 1** | Ind. | Mal. | Phi | Tha. | S.L. | Swe. | U.S.A. | *Means* |
|---|---|---|---|---|---|---|---|---|
| [1] Camera 1 | 4.2 | 4.5 | 4.4 | 5.1 | 5.3 | 4.9 | 3.1 | *4.5* |
| [4] Doc. Camera 1 | 4.6 | 4.6 | 4.2 | 4.7 | 4.2 | 5.3 | 3.3 | *4.4* |
| [7] Handsfree 1 | 4.1 | 3.7 | 4.1 | 4.1 | 4.6 | 4.4 | 3.2 | *4.0* |
| [10] Microphone 1 | 5.3 | 5.8 | 6.1 | 6.0 | 5.1 | 5.2 | 5.7 | *5.6* |
| [13] Selfview 1 | 4.0 | 4.3 | 3.8 | 4.6 | 4.8 | 4.4 | 3.3 | *4.2* |
| [16] Still Picture 1 | 4.3 | 4.6 | 4.4 | 3.8 | 5.1 | 4.6 | 3.0 | *4.3* |
| [19] Videophone 1 | 5.7 | 5.9 | 6.2 | 5.8 | 5.6 | 6.0 | 6.3 | *5.9* |
| *Means* | *4.6* | *4.8* | *4.8* | *4.9* | *5.0* | *5.0* | *4.0* | ***4.7*** |
| **SET 2** | | | | | | | | |
| [2] Camera 2 | 3.8 | 4.2 | 4.3 | 4.4 | 4.4 | 4.3 | 3.6 | *4.2* |
| [5] Doc. Camera 2 | 4.3 | 4.4 | 4.6 | 4.6 | 4.4 | 4.9 | 3.8 | *4.4* |
| [8] Handsfree 2 | 3.1 | 3.6 | 3.1 | 3.4 | 4.1 | 3.3 | 3.4 | *3.4* |
| [11] Microphone 2 | 3.7 | 3.2 | 3.3 | 3.7 | 4.2 | 3.7 | 4.3 | *3.7* |
| [14] Selfview 2 | 2.6 | 3.5 | 3.7 | 3.1 | 4.0 | 2.6 | 3.4 | *3.2* |
| [17] Still Picture 2 | 3.9 | 4.1 | 3.9 | 3.9 | 4.5 | 3.1 | 3.0 | *3.8* |
| [20] Videophone 2 | 3.2 | 3.1 | 3.0 | 3.3 | 3.8 | 2.3 | 4.0 | *3.2* |
| *Means* | *3.4* | *3.7* | *3.7* | *3.7* | *4.2* | *3.5* | *3.6* | ***3.7*** |
| **SET 3** | | | | | | | | |
| [3] Camera 3 | 4.7 | 4.6 | 4.6 | 5 | 5.6 | 5.2 | 4.0 | *4.8* |
| [6] Doc. Camera 3 | 4.1 | 4.3 | 4.8 | 4.6 | 4.7 | 4.3 | 4.2 | *4.4* |
| [9] Handsfree 3 | 4.0 | 3.7 | 3.8 | 3.5 | 4.3 | 4.3 | 3.5 | *3.9* |
| [12] Microphone 3 | 5.1 | 5.2 | 5.3 | 5.5 | 4.7 | 5.7 | 6.1 | *5.4* |
| [15] Selfview 3 | 4.4 | 4.1 | 4.5 | 4.9 | 4.2 | 4.6 | 4.2 | *4.4* |
| [18] Still Picture 3 | 4.4 | 3.8 | 4.6 | 4.3 | 4.6 | 5.3 | 4.0 | *4.4* |
| [21] Videophone 3 | 5.1 | 6.1 | 5.9 | 5.1 | 4.8 | 4.9 | 5.7 | *5.4* |
| *Means* | *4.5* | *4.5* | *4.8* | *4.7* | *4.7* | *4.9* | *4.5* | ***4.7*** |

*Legend: Ind.=Indonesia, Mal.=Malaysia, Phi.=Philippines, Tha.=Thailand, S.L.=Sri Lanka, Swe.=Sweden, U.S.A.=United States of America.*

### *6.2.7 Symbol and set preferences*

Preferences generally indicate the aesthetics; the subjects' personal judgement that symbols and set best represented the referents in question (Böcker, 1993 and Mackett-Stout and Dewar, 1981). Aesthetic judgements as reflected in preference tests can be helpful when confronted with choosing between symbols representing the same function whose performance parameters were nearly equal.

Regarding preferences (Tables 28 and 29), at least 50 percent of the subjects preferred five of the seven symbols of Set 1 compared to those of symbols of Sets 2 and 3. Between sets, Set 1 was clearly preferred among the three sets (58.3% to 68.8%) among the Asian countries. However, regarding set preferences, the half of the US subjects, preferred Set 3 and 45% preferred Set 1. Half of the Swedish subjects preferred Set 1 and almost 40% preferred Set 3. Thus, a big portion of the western subjects seemed to prefer Set 3 aside from Set 1.

TABLE 28. Symbol preferences in (%) with all countries combined (N = 240).

| | Symbol Preferences | | |
|---|---|---|---|
| | Set 1 | Set 2 | Set 3 |
| Camera | 26.3 | 37.5 | 35.8 |
| Doc. Camera | 50.8 | 19.6 | 29.6 |
| Handsfree | 52.9 | 20.4 | 22.5 |
| Microphone | 52.1 | 8.3 | 39.2 |
| Selfview | 44.2 | 7.5 | 48.3 |
| Still Picture | 55.0 | 25.8 | 18.8 |
| Videophone | 67.9 | 4.2 | 27.9 |

TABLE 29. Symbol set preferences in (%) per country.

| | Set Preferences | | |
|---|---|---|---|
| | Set 1 | Set 2 | Set 3 |
| Indonesia | 68.8 | 4.1 | 27.1 |
| Malaysia | 58.3 | 6.3 | 35.4 |
| Philippines | 68.8 | 0.0 | 31.2 |
| Thailand | 64.6 | 6.2 | 29.2 |
| Sri Lanka | 62.5 | 2.1 | 35.4 |
| Sweden | 50.0 | 0.0 | 37.5 |
| U.S.A. | 45.0 | 0.0 | 50.0 |
| *Means* | 59.7 | 2.7 | 35.1 |

### *6.2.8 Semantic Differential Tests*

Table 30 summarises the results of the semantic differential tests for the three symbol sets. Generally, symbols representing familiar or common functions had high semantic ratings (ex. camera and microphone) across all sets. Among the sets, Set 2 had the lower semantic ratings. Set 2's low hit rates and semantic scores confirmed one of the hypotheses; that is, abstract symbols would tend to have poor comprehension and rating. Novel or complex functions in turn had lower semantic ratings (ex. handsfree, selfview) and could indicate again the need for more learning time. Furthermore, frequency distributions of the ratings can also help analyse the relationships between the functions and the symbols representing them. Figures 11-17 (Appendix 3) contain the precent distribution for the 21 symbols depicting the seven-videophone functions. Positive skewness (high frequency distributions towards positive side, i.e. ratings of 5 to 7) were characteristic for symbols with high hits (at least 67%) and certainty ratings in both Part One and Part Two.

For example, Sets 1 and 3's symbols for microphone and videophone had hit rates of sixty percent or more in both Parts One and Two. Correspondingly, their SDT scores were highly skewed on the positive side. These strongly suggest the subjects' high levels of confidence and understanding of these symbols' intended meanings. The symbols can thus be expected to perform satisfactorily when used by the subjects. On the other hand, negatively skewed and platykuric (low humped or scattered) distributions can suggest lower levels of semantic or connotative concepts of the symbols in relation to the functions. These can be seen from the frequency distributions of symbols for handsfree (Sets 1-3), microphone (Set 2), still picture (Sets 2-3) and videophone (Set 2). These suggest the need to re-design the symbols. On the other hand, the referents or functions themselves may also need to be learned or explained further. This latter view in a way also gains support by the observations made by Rutter and Becka (1997). In a usability testing of cordless telephones, their subjects were tested on their comprehension of labels (functions) currently available in such telephones. Very few of these functions were understood by the majority of their subjects (Rutter and Becka, 1997). Thus, when confronted with using new devices, part of the problem may lie on possibility that the subjects do not even understand some of the functions themselves.

TABLE 30. Mean ratings of each semantic scale, per set, per country (n = 48 per country, except U.S.A. where n = 40).

| **Semantic Scales** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Set 1 | Ind. | Mal. | Phi. | Tha. | S.L. | Swe. | USA |
| Meaningful | 5.4 | 5.3 | 5.3 | 5.4 | 5.0 | 5.4 | 5.8 |
| Concrete | 5.4 | 5.3 | 5.4 | 5.4 | 4.3 | 5.3 | 5.7 |
| Familiar | 5.3 | 5.2 | 5.3 | 5.3 | 4.5 | 5.2 | 5.6 |
| Simple | 5.8 | 5.5 | 5.4 | 5.4 | 5.0 | 5.5 | 5.8 |
| Clear | 5.5 | 5.2 | 5.3 | 5.3 | 4.9 | 5.2 | 5.6 |
| Set 2 | | | | | | | |
| Meaning | 4.0 | 3.9 | 3.4 | 3.7 | 4.1 | 4.0 | 4.3 |
| Concrete | 3.9 | 3.8 | 3.4 | 3.6 | 3.6 | 3.7 | 4.3 |
| Familiar | 3.9 | 3.8 | 3.4 | 3.6 | 3.8 | 3.6 | 4.2 |
| Simple | 4.7 | 4.5 | 3.6 | 4.0 | 4.1 | 4.1 | 4.4 |
| Clear | 3.9 | 3.7 | 3.2 | 3.6 | 4.1 | 3.6 | 4.0 |
| Set 3 | | | | | | | |
| Meaning | 5.4 | 5.3 | 5.2 | 5.3 | 4.8 | 5.4 | 6.0 |
| Concrete | 5.4 | 5.3 | 5.1 | 5.3 | 4.2 | 5.2 | 5.9 |
| Familiar | 5.4 | 5.2 | 5.1 | 5.2 | 4.5 | 5.1 | 5.9 |
| Simple | 5.6 | 5.5 | 5.3 | 5.4 | 4.8 | 5.2 | 5.8 |
| Clear | 5.4 | 5.2 | 5.1 | 5.2 | 4.9 | 5.1 | 5.7 |

Regarding the western subjects, again, the Swedish group as a whole gave lower mean semantic ratings compared to the American group (Table 17). Although, there is high correlation between the hit rates and all the semantic scales, both the Swedish and US subjects had highest ratings for the symbols'meaningfulness and simplicity, with the latter group putting also prime importance to concreteness.

### *6.2.9 The Finnish Elderly (Paper 5)*

The multiple index tests used in the Asian, Swedish and US studies were similarly carried out with 29 elderly subjects (14 women and 15 men). The women were 61 to 81 years old (mean age: 69.6 years) and the men 60 to 88 years old (mean age: 70.9 years). The mean age for the whole group was 70.3 years. Compared to the subjects from Southeast Asia and even the USA, the Finnish elderly failed to do parts One and Three. Majority of the subjects claimed that the tasks were either too difficult or too complicated. Nonetheless, Tables 31 and 32 contained the results for Part Two while Tables 32 and 33 contain the results for Part Four. Compared to younger subjects (from Asia and US), the elderly subjects had much lower hit rates in all symbols of the 3 sets used.

In Part Two, the subjects also expressed the level of certainty to their answers using a response scale from "1" (very certain) to "7" (very uncertain). Table 32 shows the computed mean certainty level and standard deviation for each icon in different icon family.

TABLE 31. The results for cued-response tests (Part Two), Finnish elderly as subjects, N=29.

| | **Set 1** | | **Set 2** | | **Set 3** | |
|---|---|---|---|---|---|---|
| **Referents** | **n** | **%** | **n** | **%** | **n** | **%** |
| Camera | 10 | 34 | 3 | 11 | 8 | 30 |
| Document camera | 14 | 48 | 10 | 36 | 7 | 26 |
| Handsfree | 6 | 21 | 6 | 21 | 3 | 11 |
| Microphone | 13 | 45 | 7 | 25 | 17 | 63 |
| Selfview | 7 | 24 | 5 | 18 | 8 | 30 |
| Still Picture | 4 | 14 | 4 | 14 | 2 | 7 |
| Videophone | 15 | 52 | 5 | 18 | 14 | 52 |

TABLE 32. Certainty ratings for the cued-response tests (Part Two) using Finnish elderly, N=29.

| | **Set 1** | | **Set 2** | | **Set 3** | |
|---|---|---|---|---|---|---|
| **Referents** | **Mean** | **S. D.** | **Mean** | **S. D.** | **Mean** | **S. D.** |
| Camera | 4.4 | 1.9 | 4.6 | 1.9 | 4.1 | 1.8 |
| Document camera | 4.1 | 1.9 | 4.5 | 1.7 | 4.2 | 2.0 |
| Handsfree | 4.5 | 2.0 | 5.0 | 1.7 | 4.7 | 1.6 |
| Microphone | 4.2 | 2.0 | 4.8 | 1.8 | 4.0 | 1.8 |
| Selfview | 4.2 | 1.9 | 4.5 | 2.0 | 4.3 | 1.9 |
| Still Picture | 4.1 | 2.3 | 4.3 | 1.9 | 3.9 | 1.8 |
| Videophone | 4.1 | 2.2 | 5.0 | 1.5 | 4.1 | 1.9 |

TABLE 33. The results for Part Four using Finnish elderly subjects, N=29.

| **Referents** | **Set 1** | **%** | **Set 2** | **%** | **Set 3** | **%** | **None** | **%** |
|---|---|---|---|---|---|---|---|---|
| Camera | 6 | 21 | 17 | **61** | 4 | 14 | 1 | 4 |
| Document camera | 9 | 32 | 6 | 21 | 12 | **43** | 1 | 4 |
| Handsfree | 12 | **41** | 9 | 31 | 5 | 17 | 3 | 10 |
| Microphone | 3 | 11 | 8 | 29 | 16 | **57** | 1 | 4 |
| Selfview | 12 | **43** | 9 | 32 | 6 | 21 | 1 | 4 |
| Still Picture | 5 | 18 | 10 | 36 | 12 | **43** | 1 | 4 |
| Videophone | 10 | 36 | | 0 | 17 | **61** | 1 | 4 |
| **Family** | | 0 | 11 | 39 | 16 | **57** | 1 | 4 |

In Part Four, the subjects chose for each of the seven videophone referents the symbol or icon that he/she preferred most. The subjects also had to choose from the three icon families the icon family he/she preferred most. Table 33 presents the number and percentages of the subjects, who liked the icon best and the number of subjects who did not answer to the question. It was quite evident that the elderly subjects preferred the most of the symbols of Set 3 and even preferred Set 3 as a whole. These were in stark contrast to subjects from Southeast Asia and USA, who preferred Set 1 and its symbols, the proposed standard set by ETSI.

Part Four also contained questions regarding attitudes towards technology. The results showed that a majority of the subjects, both men and women agreed either totally or almost totally to the view that the advantages of technological progress outweigh the disadvantages (Table 34). 91% of the men and 54 % of the women thought that the progress in technology totally or almost totally makes life a lot easier (Table 35). The progress in technology was thought to make life more difficult by about 13 % of the subjects.

TABLE 34. Frequency distribution for the question: Advantages of technological progress outweigh the disadvantages.

| | | **N** | **%** | **Men** | **%** | **Women** | **%** |
|---|---|---|---|---|---|---|---|
| **Do not agree at all** | **1** | 1 | 4 | | | 1 | 8 |
| | **2** | | | | | | |
| | **3** | 2 | 8 | | | 2 | 15 |
| | **4** | 12 | 50 | 5 | 45 | 7 | 54 |
| **Totally agree** | **5** | 11 | 38 | 6 | 55 | 5 | 23 |

However, about sixty-seven percent (67%) of the subjects (54 % of the men and 77 % of the women) were totally or almost totally worried about some aspects of today's technological progress (Table 36). Eighteen percent of the men were not worried at all. These findings showed that elderly people are not entirely hostile to modern technology as commonly believed. What may worry them, more probably, is the usability of modern products, so life can indeed be better or easier for them.

TABLE 35. Frequency distribution for the question: Progress in technology makes life a lot easier:

| | | N | % | Men | % | Women | % |
|---|---|---|---|---|---|---|---|
| **Do not agree at all** | **1** | 3 | 13 | 1 | 9 | 2 | 15 |
| | **2** | 1 | 4 | | | 1 | 8 |
| | **3** | 3 | 13 | | | 3 | 23 |
| | **4** | 7 | 29 | 3 | 27 | 4 | 31 |
| **Totally agree** | **5** | 10 | 42 | 7 | 64 | 3 | 23 |

TABLE 36. Frequency distribution for the question: Some aspects of today's technological progress are worrying.

| | | N | % | Men | % | Women | % |
|---|---|---|---|---|---|---|---|
| **Do not agree at all** | **1** | 2 | 8 | 2 | 18 | | |
| | **2** | 1 | 4 | | | 1 | 8 |
| | **3** | 5 | 21 | 3 | 27 | 2 | 15 |
| | **4** | 10 | 42 | 4 | 36 | 6 | 46 |
| **Totally agree** | **5** | 6 | 25 | 2 | 18 | 4 | 31 |

The trend of demographic evolution shows that the European population, in particular, is one that is becoming old. This trend will have implications on the social structure, where respect for an independent living should be taken into consideration, taking into account the new services and products just available or those under development, such as tele-alarms and teleservices to assist people at home or elsewhere.

There can be many pleasures and satisfaction to seniority, but there are also negative physical, cognitive, and social consequences of ageing. Understanding the human factors of ageing can lead to computer designs that will facilitate access by the elderly. The benefits to the elderly include practical needs for writing, accounting, and the full range of computer tools, plus the satisfactions of education, entertainment, social interaction, communication and challenge. In this case, symbols were intended to improve the usability of a product (i.e. videophone). However, the results showed that the opposite might occur among the elderly. Other benefits include increased access of the society to the elderly for their experience, increased participation of the elderly in society through communication networks, and improved chances for productive employment of the elderly. (Schneiderman, 1992). The results in this case study emphasise these points as well. Only by a conscious and active participation of the elderly in designing and developing modern products will their needs be truly met.

### *6.2.10 General discussion*

The results of the tests in the different countries showed that the spontaneous identification test (Part One) had significantly lower hit rates compared with the hit rates in the cued response test (Part Two). Part One and together with Part Two simulate (though in a limited manner) real videophone call scenarios wherein the symbols are placed on the terminal with one function to be performed at a time. Part One is a situation akin to using the technology the first time with minimal knowledge and instructions. Thus, it was not surprising that only familiar symbols or those that are concrete representations of the functions would have higher hit rates (ex. microphone, and camera), confirming one of the main hypotheses of this study. Further, it was interesting to note that symbols could indeed be quite difficult to understand or comprehend based on the very low hit rates in Part One.

The cued-response tests were similar to situations when using the videophone the first few times but with more information and instructions given. These findings give further credence to the importance of providing the prospective users all the possible cues and learning aids when attempting to use a new product or device the first few times (Lund, 1997, Shahnavaz, 1998 and Wolff and Wogalter, 1998). By using the cued-response tests several objective parameters were derived that helped reveal other possible differences between the candidate symbols and between subject groups. These can be summarised as follows:

a) Overall, the main studies showed that hit rates or correct symbol-referent associations are still the most important objective parameter of symbol performance. In this aspect, Set 1's symbols performed best. Six of its 7 symbols were correctly and easily associated to their intended functions or referents for all subject groups.

b) However, hit rates are not the only important parameter, especially when evaluating several sets of symbols and using different subject groups. Non-hit parameters such as false alarms or confusions and missing values can be of great importance as well. For example, Set 3's symbols fared comparatively well in being associated to the intended functions, attaining hits as high as (if not higher than) Set 1. In these cases, the non-hit parameters such as the false alarms and missing values were very helpful in determining which symbol performed better among the subject groups.

c) False alarms or confusing the symbol as representing the wrong referent is a very good gauge of studying how distinct a symbol is compared to the other symbols presented at the same time. This parameter becomes more useful when used among subject groups with different cultural backgrounds (i.e. countries). The predominant types of confusions (symmetrical and asymmetrical) in each group could be used to study further how the symbols are understood and even confused by the subjects. These would then be helpful to determine if the problems would merit (a) re-design of the symbols; (b) replacement of the symbols; (c) more explanation or emphasis on how the definition of the functions. In short, would the recommendations focus on symbol re-design or replacements, or on learning or instructional aids to make sure that the device and its functions are well understood?

d) Missing values are also very important since they represent situations wherein the prospective user plainly lacks the knowledge of which among the symbols represent the desired function (referent). In actual user scenarios, they can be akin to non-use or under-use of the device. If this would be one of the main problems, recommendations would be more on steps ensuring familiarity with the device and its functions.

In turn, the subjective parameters such as subjective certainties, semantic scales and preferences were also helpful in gaining insight in subject group differences especially to certain possible subjective biases that may be culture-related. They also showed that when intended for diverse user groups, graphical symbols that are more concrete and famliar have better chances of being understood correctly (associated to the correct functions).

The above results also merited further discussion on the current issues of using spontaneous identification and cued-response tests when testing symbols, which was actually one of the minor objectives of the study. Wolff and Wogalter (1998) contended that the cognitive processes that people usually perform when encountering graphical symbols in the real world are more closely mirrored by the processes involved in open-ended tests than cued-response tests. The evaluation of symbols in open-ended tests (like the spontaneous identification test) has also been regarded as less-likely to produce constrained and distorted answers compared to the cued-response type of tests (Neisser, 1987). These contentions are valid and will hold when "stand-alone" symbols are the ones in question (ex. traffic signs and symbols, road signs, and most directional or informational symbols in public places and work sites). But when symbols are designed to appear together or as sets, results of open-ended tests can be quite difficult as experienced in the studies. Although, they were helpful in revealing the reality that symbols could still be very poorly recognised or understood by a large number of people, eliciting clear-cut information could be difficult. Criteria need to be established for the kinds of answers that would be counted as correct. This was one of the realisations in the study – that there could always be some subjective judgement of correctness of the participants' responses. This is even magnified considering the different languages or mother tongues of the subject groups tested. Thus, reliability assessment would be better if there would be more than one judge something that was also pointed out by Dewar (1994) and Wolff and Wogalter (1998). Understandably, the downside then would be the extra time and resources needed when more "judges" are needed. Further, the study also revealed the difficulty of performing open-ended tests with elderly subjects. This appeared to be a difficult task for this subject group compared to the cued-response test.

The cued-response tests as adopted from Böcker (1993) had a more straight-forward manner of quantifying correct responses as well as the errors. Together with the manner of presenting each referent and one symbol set at a time, they more closely mimic the projected user-scenario of how the videophone symbols would be used in real life – that is, situations wherein the users have some knowledge of the device at hand and some intended functions to be represented by the symbols. On the other hand, spontaneous identification tests, when properly done, simulate "first-time use" scenarios, wherein both device and symbols can be quite new and unfamiliar. Correct symbol-referent associations (hit rates) for each test type can thus be quite different but have their own uses and relevance in symbol evaluation.

Lastly, in this research, the different tests were designed to measure initial (a priori) meaning and appropriateness, ease of learning and remembering, as well as the probability of confusion with other designs. The methods utilised were largely based on the multiple index approach (MIA) developed by Böcker with the ETSI (Böcker, 1993) and with some modifications. An open-ended test (spontaneous identification) and semantic differential scales were also employed. The latter was in consideration of the recommendation that it remains as an important method of measuring symbol understanding. The issue between using open-ended and cued-response tests was already discussed in detail earlier. Nonetheless, when Böcker developed the MIA and performed the tests, one of the main objectives was to decide which among several candidate graphical symbols would best represent seven particular videophone functions. The principles and logic of the MIA were discussed together with the results they obtained. In turn, present research acknowledges the appropriateness of the MIA as the method when testing graphical symbols intended to appear together in an interface. Using this approach of multiple test parameters helped not only in identifying best performing symbols, that is, which among the different symbols were understood correctly as they were intended to be. They also aided in studying how different subject groups (country groups) could differ in their understanding of symbols and how such differences could help in determining the interventions needed to make graphical symbols better interfaces.

# 7 CONCLUSIONS AND RECOMMENDATIONS

Based on all the studies, the aims of the project were realised. Different tests were utilised in order to evaluate candidate telecommunication symbols developed in the west using subjects from eastern (Asian) and western countries. A quasi-experimental approach utilising multiple test parameters was adopted in testing different sets of graphical symbols for the videophone (icons from CIAJ of Japan, IDEO of London, both from studies by Tudor (1994), and ETSI of Europe). The following conclusions could then be drawn. First, majority of the icons tested was poorly recognised by all subject groups especially when based on the ISO 9186 67% recognition rate on spontaneous identification tests as basis. The icons or symbols above this level were mostly concrete representations of their referents and/or quite familiar to the users already (CIAJ and ETSI symbols).

Second, although Asian subjects generally had comparable recognition rates to western subjects, they had more instances of confusions. Third, empirical tests using multiple indices are important to properly evaluate the symbols on their usability especially across different user groups. The evaluation method should reflect expected usage scenarios to determine the suitability of each symbol alone, and together with the other symbols as well. Hit rate remains an important parameter, but how it will be derived has to be considered closely. For example, cued responses and spontaneous identification are useful to determine if the symbols indeed are understood with and without/or least of cues respectively. It must be remembered though, that when using open-ended tests such as the spontaneous identification test, the method on how to accurately interpret and judge correct answers have to be carefully considered. At the same time hits rates alone are not enough. False alarms, missing values, subjective certainties and preferences can also be very useful factors in order to make deeper analyses. They enable the tester to see other often subtle but important differences (i.e. subjective biases, confusing symbols to other referents and symbols) on how users perceive and understand symbols. Using semantic differential scales could also help in understanding how symbols are perceived in relation to the functions they are supposed to represent. Taken altogether, they are highly beneficial not only in properly interpreting test results of different user groups, but also in formulating instructions and other aids in learning to use new products faster and more satisfactorily.

Modern society has come back to use graphical symbols, especially as an essential part of the interface between humans and their increasingly complex machines. Specifically, they are a primary GUI component, helping put order and sense to the complexities of the human-computer systems so abundant nowadays. They offer numerous advantages over text-based interfaces, but with their own disadvantages of course, both of which need to be weighed when contemplating the use of symbols. However, as discussed above, designing symbols that can be understood by a majority of the target users is not a simple task. Part of the complexity is how people perceive symbols and render meaning to it. Understanding something about the human visual image processing can assist in gaining a few insights of how they are understood. But knowledge and awareness of the possible different contexts in which symbols can be used or perceived, and the different levels of user experience and expectations can equally (and greatly) affect how the symbols in question are perceived as they are intended to be. These issues have led us to the critical role played by symbol testing and evaluation. Several methods are available depending on one's objective, but in almost all cases user involvement is of prime consideration.

Evaluating graphical symbols is complicated and the use of multiple indices can be of great help in understanding how symbols are recognised, understood and learned by different user groups. Still, there are a few limitations in this study that need to be addressed and are thus recommended towards better designing and testing of symbols:

a) Since the studies were mainly based on pencil and paper tasks, more insights may be obtained if:
    - the symbols are tested using actual interfaces (computer screens, dialpads, etc.) and under more varied scenarios (e.g. with and without help options, 'normal vs. highly-stressed situations', cross-applications)
    - the functions themselves need to be tested for their comprehension among the subjects or prospective users

b) The semantic differential testing results were interesting but discussed here in still quite a limited manner and needs deeper analyses.

c) In a more ergonomics approach, users and experts should both be considered in all possible stages of design and testing. One simple but concrete follow-up is to design an iterative type of method to generate graphical symbols. Briefly, such method will involve a process of eliciting from the subjects their own ideas of how the symbols should look like, identifying the design factors behind them that can then be used by ergonomics experts and designers to contrive alternative symbols.

d) An often-expressed issue in graphical symbols use is the applicability and potential benefit of using the same graphical symbols for the same functions across different devices or equipment. It is highly recommended to do more studies to validate this view not only across varying devices but also across varying country or cultural groups.

# 8 REFERENCES


Abeysekera, J.D.A., Shahnavaz, H., & Chapman, L.J. (1990). Ergonomics in developing countries. In *Advances in Industrial Ergonomics and Safety II (pp. 771–778)*. London: Taylor and Francis.

American National Standards Institute (ANSI). (1993). *Accredited standard on safety colors, signs symbols, labels and tags, Z535.1-5*. Washington DC: National Electrical Manufacturers Association.

Brelsford, J.W., Wogalter, M.S., & Scoggins, J.A. (1994). Enhancing comprehension and retention of safety-related pictorials. In *Proceedings of the Human Factors Ergonomics Society 38th Annual Meeting (pp. 836-840).*

Böcker, M. (1993). A multiple index approach for the evaluation of pictograms. In: *Proceedings of the 14th International Symposium on Human Factors in Telecommunications (pp. 73-85).* Darmstadt, Germany, May 11 - 14, Heidelberg: R. v. Decker's Verlag, G. Schenck GmbH.

Böcker, M. (1997). Symbol design, In: *A Focus Seminar in Symbol Design and Evaluation: Forum for Corporate Usability within Ericsson*. Stockholm: Ericsson AB.

Caron, J.P., Jamieson, D.G., & Dewar, R.E. (1980). Evaluating pictographs using semantic differential and classification techniques. *Ergonomics, 23*, 137-146.

Clarke, A.M., & Brown, S. (1992). Pictograms in multipoint videotelephony: an international study. In E.J. Lovesey (Ed.), *Contemporary Ergonomics 1992 (pp. 494-499).* London: Taylor and Francis.

Collins, B.L., & Lerner, N.D. (1982). Assessment of fire safety symbols. *Human Factors, 24(1)*, 75-84.

Davies, S., Haines, H., Norris, B., & Wilson, J. (1998). Safety pictograms: are they getting the message across? *Applied Ergonomics, 29(1)*, 15-23.

Dewar, R. (1994). Design and evaluation of graphic symbols. In *Proceedings of Public Graphics (pp. 24.1-24.18).* Utrecht, Netherlands: University of Utrecht, Department of Psychonomics.

Dewar, R. (1999). Design and evaluation of public information symbols. In H.J.G. Zwaga, T. Boersema, & H.C.M. Hoonhout (Eds.), *Visual Information for Everyday Use - Design and Research Perspectives (pp. 285-303).* London: Taylor and Francis.

Dewar, R.E., & Ells, J.G. (1979). The semantic differential as an index of traffic sign perception and comprehension, *Human Factors, 19(2)*, 183-189.

Dreyfuss, H. (1972). *Symbol sourcebook*. New York: McGraw-Hill.

Easterby, R.S., & Hakiel, S.R. (1981). Field testing of consumer safety signs: the comprehension of pictorially presented messages, *Applied Ergonomics, 12(3),* 143-152.

Edwards, A.L., & Porter, B.C. (1972). *The affective domain.* Washington DC: National Special Media Institutes-Gryphon House.

Edworthy, J., & Adams, A. (1996). *Warning design - A research prospective*. London: Taylor and Francis.

Encarta Encyclopedia, (1997a). Asia. *Microsoft™ Encarta 1997 Encyclopedia CD Version.*

Encarta Encyclopedia, (1997a). Technology. *Microsoft™ Encarta 1997 Encyclopedia CD Version.*

Encarta Encyclopedia, (1997b). Microprocessor. *Microsoft™ Encarta 1997 Encyclopedia CD Version*.

European Telecommunications Standards Institute (ETSI). (1993). *Technical Report - Human Factors (HF), Results of an evaluation study of pictograms for point-to-point videotelephony*, Sophia Antipolis Cedex: ETSI.

Eysenck, M.W., & Keane, M.T. (1999). *Cognitive Psychology* (3rd ed.). East Sussex: Psychology Press.

Familant, M.E., & Detwailer, M.C. (1993). Iconic reference: evolving perspectives and organizing framework. *International Journal of Man-Machine Studies*, *38*, 705-728.

Gittins, D. (1986). Icon-based human computer interaction. *International Journal of Man-Machine Studies*, *24*, 519-543.

Green, P., & Pew, R.W. (1978). Evaluating pictographic symbols: an automotive application. *Human Factors*, *20*, 103-114.

Halliday, M.A.K. (1970). Language structure and language function. In J. Lyons (Ed.), *New Horizons in Linguistics.* Middlesex: Penguin.

Henerson, M.E., Morris, L.L., & Fitz-Gibbon, C.T. (1987). *How to Measure Attitudes.* Newbury Park: Sage Publications.

Horton, W. (1994). *The Icon Book - Visual symbols for computer systems and documentation.* New York: Wiley and Sons.

International Telecommunications Union (ITU). (1995). *Recommendation F.910 (02/95) - Procedures for designing, evaluating and selecting symbols, pictograms and icons.* Geneva: ITU.

Jones, S. (1983). Stereotipy in pictograms of abstract concepts. *Ergonomics, 26(6),* 605-611.

Keesing, R.M. (1974). Theories of culture. *Annual Review of Anthropology, 3,* 73-79.

Kline, T.J.B. Ghali, L.M., Kline, D.W., & Brown, S. (1990). Visibility distance of highway signs among young, middle-aged, and older observers: Icons are better than text, *Human Factors, 32,* 609-619.

Kroemer, K. (1997). *Fitting the Task to the Human* (5th ed.). London: Taylor and Francis, ISBN 0-7484-0664-4.

Lin, R. (1992). An application of the semantic differential to icon design, In *Proceedings of the Human Factors Society 36th Annual Meeting, Atlanta, Georgia, October 12 - 16, 1992, Vol. 1, (pp. 336-340)*. Santa Monica: Human Factors Society.

Lin, R., & Kreifeldt, J.G. (1992). Understanding the image functions for icon design, In: Innovations for Interactions. In *Proceedings of the Human Factors Society 36th Annual Meeting, Atlanta, Georgia, October 12 - 16, 1992, Vol. 1, (pp. 341-345).* Santa Monica: Human Factors Society.

Lin, R., Kreifeldt, J.G., & Chi, C.F. (1992). A study of evaluation design sufficiency for iconic interface design: the design perspective. In Lovesey, E.J. (Ed.), *Contemporary Ergonomics 1992, (pp. 376-384)*. London: Taylor and Francis.

Lodding, K.N. (1982). Iconic interfacing. *IEEE Computer Graphics and Applications, 3,* 11-20.

Lund, A.M.L., (19979. Expert ratings of usability maxims. *Ergonomics in Design, 5(3),* 15-20.

Mackett-Stout, J., & Dewar, R. (1981). Evaluation of symbolic public information signs. In *Proceedings of Human Factors Society Conference, Vol. 23, No. 2, (pp. 139-151).*

Maguire, M.C. (1985). A review of human factors guidelines and techniques for the design of graphical human-computer interfaces. *Computers and Graphics, 9(3),* 221-235.

Magyar, R.L. (1990). Assessing icon appropriateness and icon discriminability with a paired comparison testing procedure, In *Countdown to the 21st Century. Proceedings of the Human Factors Society 34th Annual Meeting, Orlando, Florida, October 8 - 12, 1990, Vol. 2,* (*pp. 1204-1208).* Santa Monica: Human Factors Society.

May, J., Böcker, M., Barnard, P.J., & Green, A.J.K. (1990). Characterising structural and dynamic aspects of the interpretation of visual interface objects, In *Proceedings of Esprit '90.* Dordrecht: Kluwer Academic Publishers.

Modley, R. (1966). Graphic symbols for world-wide communication. In Kepes, G. (Ed.), *Sign, Image and Symbol, (pp. 108 - 125).* London: Studio Vista.

Neisser, U. (1987). The present and the past. In *Second International Conference on the Practical Aspects of Memory*. Swansea, Wales.

Nolan, P.R. (1989). Designing screen icons: ranking and matching studies. In *Proceedings of the Human Factors Society 33rd Annual Meeting, (pp. 380-384)*, Santa Monica: Human Factors.

Olmstead, W.T. (1999). The usability of symbols for health care facilities: the effects of culture, gender and age. In Zwaga, H.J.G., Boersema, T., & Hoonhout, H.C.M. (Eds.), *Visual Information for Everyday Use - Design and Research Perspectives, (pp. 315-320).* London: Taylor and Francis.

Oppenheim, A.N. (1983). *Questionnaire Design and Attitude Measurement*. London: Heinemann.

Osgood, C.E., Suci, G., & Tannenbaum, P. (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press.

Ossner, J. (1990). Transnational symbols - The rule of pictograms and models in the learning process. In Nielsen, J. (ed.), *Designing Interfaces For International Use, (pp. 11 - 38)*, Amsterdam: Elsevier.

Plocher, T.A., Garg, C., & Chestnut, J. (1999). Connecting culture, user characteristics and user interface design, In Bullinger, H. & Ziegler, J. (Eds.), *Advances in Human Factors/Ergonomics, Proceedings of the 1999 Human-Computer Interaction Conference: Ergonomics and User Interfaces*, (pp. 803-807). München, Germany, August 22-26, 1999.

Rogers, Y. (1986). Evaluating the meaningfulness of icon sets to represent command operations, In *Proceedings of Human Computer Interaction 86, (pp. 586 - 603)*. London: British Computer Society Ed.

Rutter, B.G., & Becka, A.M. (1997). Evolution of a cordless phone. *Ergonomics in Design, 5(3)*, 9-14.

Salasoo, A. (1990). Towards usable icon sets: a case study from telecommunications engineering, In *Countdown to the 21st Century. Proceedings of the Human Factors Society 34th Annual Meeting, Orlando, Florida, October 8 - 12, 1990, Vol. 1*, (*pp. 203-207).* Santa Monica: Human Factors Society.

Sanders, A.F. (1998). Information processing and design. In *Chapter 29: Ergonomics, Work System Design, Encyclopedia of Occupational Health and Safety, (pp. 29.71-29.75).* Geneva: ILO.

Senner, W.M. (1989). *The Origins of Writing.* Lincoln: University of Nebraska Press.

Shahnavaz, H. (1991). *Transfer of Technology to Industrially Developing Countries and Human Factors Considerations, Center for Ergonomics of Developing Countries*, Luleå: Luleå University Press, Sweden.

Shahnavaz, H. (1998). Cultural differences, In: *Chapter 29: Ergonomics, Designing for Everyone, Encyclopedia of Occupational Health and Safety, (pp. 29.79-29.83).* Geneva: ILO.

Shneiderman, B. (1992). *Designing the User Interface*. New York: Addison-Wesley Publishing Company.

Stammers, R.B. (1990). Judged appropriateness of icons as a predictor of identification performance, In Lovesey, E.J. (Ed.), *Contemporary Ergonomics 1990, (pp. 371 - 376).* London: Taylor and Francis.

Stramler, J.H. Jr. (1993). *The Dictionary for Human Factors/Ergonomics*. London: CRC Press.

Sukaviriya, P., & Moran, L. (1990). User interface for Asia. In Nielsen, J. (Ed.), *Designing Interfaces For International Use, (pp. 189-218).* Amsterdam: Elsevier.

Tudor, L. (1994). The generation and testing of candidate telecommunications icons, *International Conference in Human-Computer Interaction Proceedings, Vol. 11, St. Petersburg, Russia, August 2-6, (pp. 62-68).*

Van Dam, (1984). Computer software for graphics, *Scientific American, 251*, 102-113.

Vora, P., Helander, M., Swede, H., & Wilson, J., (1991). Developing guidelines for symbol design: a comparison of methodologies, *In Proceedings of Interface 1991, (pp. 6-11).*

Webb, J.M., Sorenson, P.F., & Lyons, N.P. (1989). An empirical approach to the evaluation of icons, *SIGCHI Bulletin, 21(1),* 87-90.

Wolff, J.S., & Wogalter, M.S., (1998). Comprehension of pictorial symbols: effects of context and test method, *Human Factors, June 1998*, 173-186.

Wood, W.T., & Wood, S.K. (1987). Icons in Everyday Life, In *Social Ergonomics and Stress Aspects of Work with Computers, (pp. 97-104).*

Zwaga, H.L.G. (1989). Predicting the public's comprehension of graphic symbols, In Megaw, E.D. (Ed.), *Contemporary Ergonomics 1989, (pp. 431-436).* London: Taylor and Francis.

# PAPERS PRODUCED FROM THE STUDIES

Proceedings of the
First Asia Pacific Conference on
Computer Human Interaction

APCHI '96
Singapore 25-28 June 1996

**Edited by**
**Dr. Lim Kee Yong**
**Linda Herman**
**Dr. Ying K. Leung**
**Dr. Jackie Moyes**

**PAPER 1**

# A User-based Evaluation of Telecommunication Icons: Some Cross-Cultural Issues

# *A User-based Evaluation of Telecommunication Icons: Some Cross-Cultural Issues*

**Paul Piamonte, M.D.; John D.A. Abeysekera, Ph.D.; Kjell Ohlsson, Ph.D.**
*Industrial Ergonomics Division*
*Dept. of Human Work Sciences, Luleå University*
*Luleå, SWEDEN S-971 87*
*+46-0920-91407, email:John.Abeysekera@arb.luth.se*

**Abstract**
Telecommunication and computer icons are often designed and developed in western countries. These may create problems among potential users in eastern countries. Simple and reliable methods are needed when such icons are used in Third World countries. This paper presents a pilot study utilising semantic differentials in evaluating candidate icons designed in western countries using subjects from an Asian country. The data indicate that icon recognition and perception may vary in an Asian setting. Semantic differential scores correlated highly with the evaluation results and may be useful in similar cross-cultural studies of symbols as interface tools.

Keywords
Icons, culture, semantic differential testing, information technology, user-based evaluation.

## 1. Introduction

Icons can be defined as “symbols, glyphs or pictographs used to present a concept, idea or object,” (Wood and Wood, 1987). In human-computer interfaces, they help introduce the technology (product) and its functions. They can speed up access to desired operations, and promote a sense of use. Under the present economic globalization wherein the designers can be far removed and different from the potential users, the opposite effects can occur when the icons are unfamiliar, difficult to understand, incomprehensible, or unappealing. Of particular interest is the Asia-Pacific region. Driven by an ever-improving economy, there is a concomitant increase in the access and use of computer and other telecom products. Researches are needed to design and develop simple, inexpensive but reliable methods to evaluate icons applicable in such countries of vastly different cultural backgrounds.

Numerous methods have already been tried in evaluating icons and other graphic designs (Lin and Kreifeldt, 1992). The semantic differential (SD) test developed by Osgood, Suci and Tannenbaum (1957) was a method of measuring the psychological meaning of concepts or objects by direct ratings of the stimuli using scales anchored on the extremes by bipolar adjectives. Kerlinger (1973) cited several studies done on the meanings of objects, personalities, pictures and abstract concepts The use of SD in evaluating graphic symbols have been carried out as well (Lin and Kreifeldt, 1992). Icon characteristics such as concreteness, familiarity and learnability were discussed. However, aside from the methodology, the use of actual potential users as subjects is equally important in studying the possible cultural implications of using symbols such as icons in the interfaces. Tudor (1994) emphasised the importance of explaining and testing the icons in manners reflective not only of typical user scenarios but also of the users themselves. In the Asian setting most, if not all, of the icons for use in computerised products are designed in industrialised countries but targeted for use by nationals from varying cultures. This pilot study aims to evaluate icons designed in Japan and Europe (industrialised countries) using the semantic differential approach with subjects coming from the Philippines (an industrially developing country).

## 2. Methodology

### *2.1 Subjects and Materials*

Thirty subjects (15 males and 15 females) aged 22 to 45 years from Metro-Manila were randomly selected to participate. All were office employees with 2 to 5 years experience working with computers. The 23 icon referents and the 69 icons used by Tudor (1994) were selected. The icons were grouped into three types: abstract, concrete and proposed based on their general attributes (Table 1). Sixty-two of the icons were designed in England. The remaining 7 icons were designed in Japan (Table 1, Proposed Icons = “Volume”, “Achieve Dial Tone”, “Hold”, “Music on Hold”, “Mute”, “Store”, and “Speakerphone”). The questionnaire had two parts: recognition/subjective certainty tests, and semantic differential. Questionnaires were produced per icon family with the icons and their referents presented in random orders. The semantic differential was composed of 6 bipolar adjectives in 7 point scales also arranged in random orders for each of the icons or referents. The first page contained a standard set of instructions.

### *2.2 Procedure*

After presenting the study’s objectives and explaining the icons, the subjects first underwent the recognition and certainty tests. Icons and referents were randomly presented and to be matched correctly. For each recognition test, certainty levels on the answers were indicated by using a 7 point scale. After one week, the semantic differential testing was performed. The icons were again presented. This time, they were to rate each icon based on the 6 bipolar adjectives. The same tasks were done for the three icon families. Thus, each subject performed 3 recognition/subjective certainty tests, and 3 semantic differential tests, one for each icon family.

## 3. Results and Discussion

Tables 1 and 2 give the summary of the results of recognition/certainty tests and the semantic differential testing. For the 23 referents represented by 23 icon families, only 5 of the 23 referents had a mean recognition percentage of at least 60% in all 3 icon families (Conference, Dialpad, Music On Hold, Speakerphone, Volume). As a family, the abstract icons had the highest recognition means for the

| | **ABSTRACT ICONS** | | **CONCRETE ICONS** | | **PROPOSED ICONS** | |
|---|---|---|---|---|---|---|
| **REFERENTS** | % Mean Recognition | Mean Ratings | % Mean Recognition | Mean Ratings | % Mean Recognition | Mean Ratings |
| 1. Achieve Dial Tone | 27.6% | 3.38 | 32.1% | 3.44 | 23.3% | 3.14 |
| 2. Answer Ringing Call | 10.0% | 4.33 | 60.0% | 4.89 | 56.7% | 4.76 |
| 3. Call Log | 50.0% | 3.93 | 26.7% | 4.63 | 46.7% | 2.97 |
| 4. Conference | 80.0% | 5.00 | 93.3% | 5.57 | 93.3% | 5.54 |
| 5. Dialpad | 83.3% | 4.96 | 83.3% | 5.28 | 80.0% | 5.54 |
| 6. Drop | 37.9% | 4.18 | 37.9% | 4.45 | 33.3% | 4.70 |
| 7. Help Specific | 21.4% | 2.33 | 71.4% | 4.70 | 42.9% | 4.67 |
| 8. Help System | 36.7% | 2.09 | 20.0% | 2.83 | 33.3% | 4.40 |
| 9. HFAI | 25.0% | 4.00 | 34.5% | 3.00 | 39.3% | 4.09 |
| 10. Hold | 35.7% | 3.30 | 50.0% | 5.60 | 73.3% | 5.32 |
| 11. Message | 36.7% | 4.18 | 20.0% | 3.83 | 89.7% | 5.23 |
| 12. Music On Hold | 93.3% | 5.39 | 90.0% | 6.11 | 90.0% | 6.04 |
| 13. Mute | 70.0% | 4.86 | 56.7% | 4.88 | 40.0% | 4.83 |
| 14. Notes | 44.8% | 4.31 | 58.6% | 5.41 | 76.7% | 5.17 |
| 15. Phone Call Active | 51.7% | 3.87 | 33.3% | 4.60 | 50.0% | 4.67 |
| 16. Retrieve | 72.4% | 4.71 | 50.0% | 4.27 | 10.7% | 4.00 |
| 17. Ringer Select | 43.3% | 3.85 | 86.7% | 6.23 | 83.3% | 5.52 |
| 18. Speed Dial | 50.0% | 4.80 | 70.0% | 4.95 | 75.9% | 4.36 |
| 19. Speakerphone | 63.3% | 5.63 | 86.7% | 5.81 | 70.0% | 5.19 |
| 20. Store | 43.3% | 4.31 | 43.3% | 3.77 | 56.7% | 3.35 |
| 21. Switch Hook Control | 27.6% | 3.13 | 23.3% | 4.00 | 43.3% | 3.92 |
| 22. Transfer | 43.3% | 3.77 | 76.7% | 5.74 | 83.3% | 5.56 |
| 23. Volume | 90.0% | 5.48 | 80.0% | 5.46 | 79.3% | 6.09 |

Table 1. Mean Recognition and Ratings Tests for the 23 Referents (N=30).

referent for "Mute." Concrete icons were solely high for the referents "Help Specific," and "Answer Ringing Call." The Proposed icon groups had the highest recognition means for referents "Hold," "Message," and "Notes." Both concrete and proposed icon families had high mean scores for referents "Ringer Select," "Speed Dial," and "Transfer." Overall mean recognition percentages between the 3 icon groups were 5%0 to 59% with no significant differences among them ($p<.05$). Similarly, the icons with high mean recognition percentages also exhibited high mean rating scores for the certainty tests. Rank order correlations were also at least 0.700 for these icons. Icons with low recognition means had low certainty levels and semantic scores. For the SD tests, results showed that correlations were high between the semantic scales and the percent mean of correct recognition regardless of icon families.

| *Semantic Scales* | *Abstract Icons* | *Concrete Icons* | *Proposed Icons* |
|---|---|---|---|
| Concrete-Abstract | 0.707* | 0.887* | 0.888* |
| Familiar-Strange | 0.673* | 0.871* | 0.867* |
| Meaningful-Meaningless | 0.656* | 0.886* | 0.883 |
| Related-Unrelated | 0.668* | 0.859* | 0.714* |
| Sharp-Dull | 0.667* | 0.878* | 0.833* |
| Simple-Complex | 0.716* | 0.879* | 0.896* |

Table 2. Correlations (rank order) between Semantic Differential Scores and Percentage Correct Recognition (N=30); *$p < 0.01$.

A few insights can be gleaned from the results. Culture can be defined as something which also influences the publicly available symbolic forms through which people experience, express and understand meanings (Keesing, 1974). Of the 7 icons designed in Japan, 5 had high mean percent of correct recognition. These results differed from those of Tudor (1994). Using American subjects the 7 icons designed in Japan had varying levels of mean values for correct recognition. Such difference may be a reflection of a Philippine social condition where majority of current computerised and electronic appliances in the Philippines are imported from Japan. Americans on the other hand are obviously more akin to be exposed to products they have made. Although USA still ranks as the major trading partner of the Philippines followed by Japan, the latter ranks first in terms of the computer and telecom products marketed and used locally. Icons are also commonly found in these products. These render the view that cultural familiarity and meaning of icons can affect their usability when applied as interface tools in new or modern technologies.

The SD results showed that SD testing can be a valid index of how messages of symbols like the icons are perceived and understood. It is simple, inexpensive but reliable. Admittedly though, there are limitations in this pilot study. A distinct group of potential users (educated and computer literate) was used as subjects. Future studies will need to utilise varying groups of potential users (students, professionals, experts, novices, etc.) across different cultures of Asia. It will be quite useful and interesting to study how icons in HCI's are recognised and understood in these countries as compared to their counterparts in Europe and America.

## References


Keesing, R M (1974), "Theories of Culture", *Annual Review of Anthropology,* vol 3, pp. 73-79.

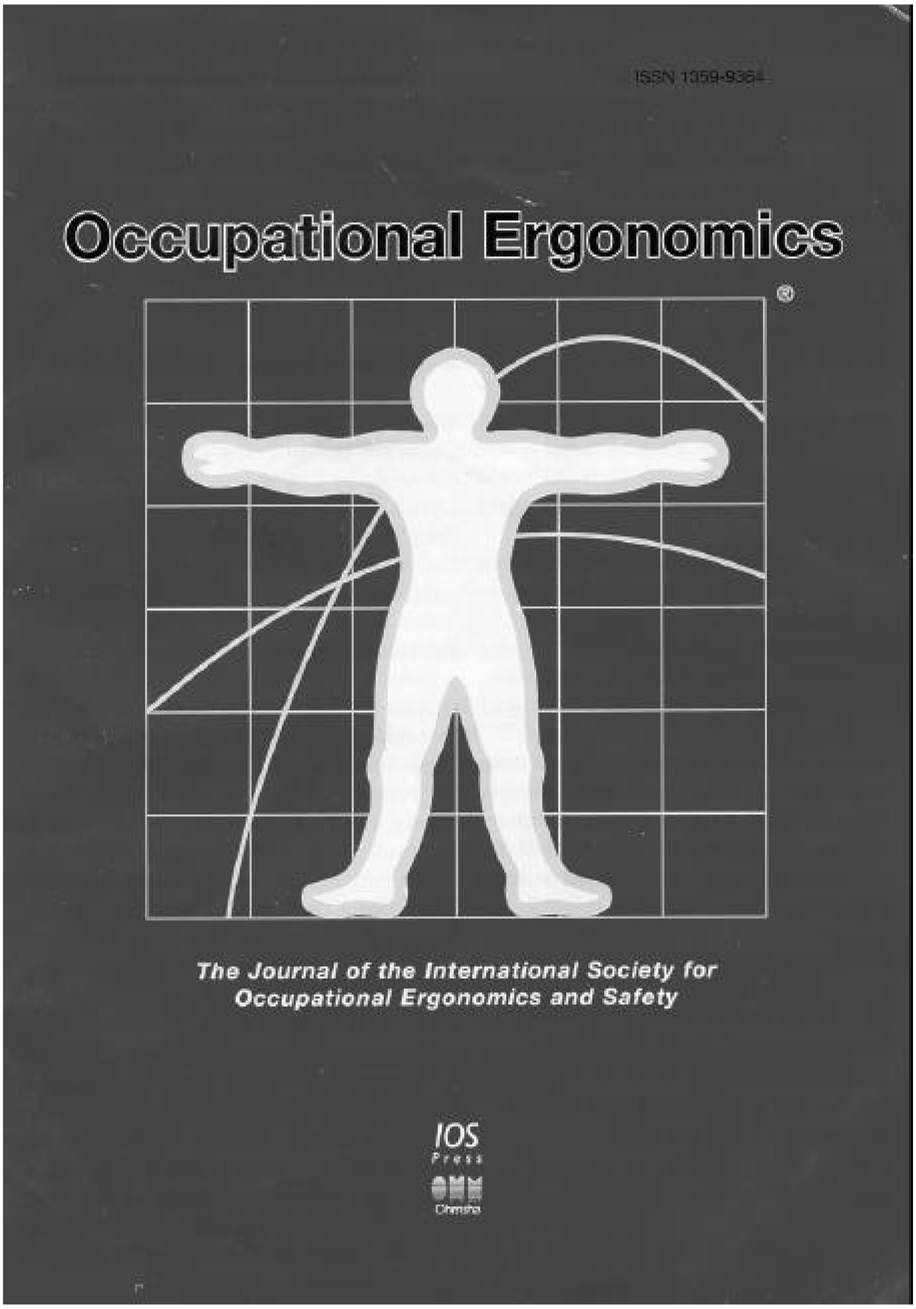
Kerlinger, F N (1973), *Foundations of Behavioral Research,* Holt-Saunders International.

Lin, R, and Kreifeldt, J G (1992), "Understanding the Image Functions for Icon Design", Human Factors Society 36th Annual Meeting, Human Factors Society (Sta. Monica), pp. 341-345.

Osgood, C E, Suci, G, and Tannenbaum, P (1957), *The Measurement of Meaning,* University of Illinois Press.

Tudor, L G. (1994), "The Generation and Testing of Candidate Telecommunicat-ions Icons", *AT & T Bell Research Papers,* pp. 62-68.

Wood, W T, and Wood, S K (1987), "Icons in Everyday Life", *Social Ergonomics and Stress Aspects of Work with Computers,* pp. 97-104.




## PAPER 2

# EVALUATING PUBLIC GRAPHIC SYMBOLS USING MULTIPLE TEST PARAMETERS

# EVALUATING PUBLIC GRAPHIC SYMBOLS USING MULTIPLE TEST PARAMETERS

DOMINIC PAUL T. PIAMONTE[1], KJELL OHLSSON[2], and JOHN ABEYSEKERA[1]
[1]Industrial Ergonomics Division, Human Work Sciences Department,
Luleå University of Technology, S-97187 Luleå, Sweden
[2]Engineering Psychology Division, Human Work Sciences Department,
Luleå University of Technology, S-97187 Luleå, Sweden
Paul.Piamonte@arb.luth.se, Kjell.Ohlsson@arb.luth.se
John.Abeysekera@arb.luth.se
*Phone: +46 920 72169, Fax: +46 920 91030*

***Abstract***
Graphic symbols require less space, are visually more distinct, easy to learn and can help reduce the likelihood of errors when properly designed. With the current globalization and advent of information technology, diverse user groups increasingly use them across the world. However, generating and testing graphic symbols in these groups can be intricate. When a user chooses which among the present symbols will bring about the target function or information, several factors come into play. Different tests may thus be helpful in determining which symbol or symbol sets are best. They also need to be tested based on the contexts of their intended use. This study discusses the results of using multiple test methods in evaluating videophone symbols among groups of different Asian nationalities. Correct associations, confusions and preferences are some of the important factors to be considered when determining the graphic symbols to use.

## 1. INTRODUCTION

Graphic symbols (icons and pictograms) have become one of the strategies to introduce technologies and their products across user groups of different cultures and languages. Icons generally refer to symbols that are simple, concrete and usually self-explanatory of the ideas, objects or functions they represent [30]. Pictograms are usually more abstract, conveying messages by analogy or symbolism and thus require certain learning processes to be understood [3]. Graphic symbols have become attractive as product interfaces since they require less space and are non-text dependent. They can even be visually more distinct than texts making them very useful in work groups of varying cultures and languages. In their paper on safety pictograms, Davies, Haines, Norris, and Wilson [5] mentioned that in the light of the emerging European market, graphic symbols have become very attractive since they have the potential to cross the language barrier. Messages may need not always be translated or explained in the 13 languages of the Union. Regarding ergonomics of computer-based work, graphic symbols can facilitate access to certain functions. Studies also suggest that when properly designed and tested, they have the potential to help expedite work in emergency and highly stressful settings such as rescue and medical (clinic and hospital) emergencies.

The use of graphic symbols, however, is not devoid of problems. Foremost among these is that some symbols are simply not understood well [2]. The above advantages were cited as evident when the concepts being represented were well understood or concrete, and not when ideas become more abstract. Sanders [21] further added that the superiority of symbols could only be attained if they contain a small amount of detail are sufficiently distinct in shape and are unambiguous. Regarding older people, Davis et al [5] also cited the studies of Easterby and Hakiel [7] showing that although symbols were easily recognizable, understanding their meaning was generally poorer compared to younger people.

## 2. EVALUATING GRAPHIC SYMBOLS

Numerous evaluation studies have been done in search of what, where and how symbols can best be used. A review of these studies reveals at least two critical elements affecting evaluation results – the methods used and the subjects involved.

### 2.1. Current methodologies

Various methods have been utilized to evaluate icons and pictograms. Some of the methods frequently used were a) appropriateness tests, b) preference ratings tests, c) naming and matching tests d) comprehension tests, e) recognition and recall tests, and f) paired-comparison tests [4, 8, 9, 11, 12, 15, 17, 18, 19, 20, 24 and 31]. Webb, Sorensen and Lyons [28] described these and other methods of evaluating symbols as involving psychophysics, scaling, recognition and memory testing, as well as statistical modeling and analysis. Generally, appropriateness testing was the preliminary procedure to screen several candidate designs. Matching studies determined how well the symbols performed as a set and their likelihood of being confused with each other [18].

However, owing to the complexity of how symbols are understood and learned, no single testing method is usually sufficient when determining which symbols to use especially among diverse user groups. More than one measure needs to be used, and the relevance of each specific criterion depends on the purpose of the symbol and where it will be used [6]. In a method developed for the European Telecommunications Standards Institute (ETSI), Böcker [3] utilized what he called the Multiple Index Approach (MIA) for empirical evaluation of graphic symbols. The method consisted of three-part questionnaires evaluating seven indices deemed important when testing graphic symbols designed to appear simultaneously as a set (i.e. graphic symbols appearing on dialpads and videophone terminals). He pointed out that such approach was essential in order to reflect typical usage scenarios of these symbols. For example, testing candidate videophone symbols should reflect the real-life situation wherein a prospective user had to choose which among the simultaneously visible symbols would bring about a specific function being sought (ex. switching on the camera function of the videophone). The test simulated this by using cued response tests. It presented one symbol set and one referent (function) at a time and subjects had to select which symbol best represented the referent. The possible outcomes would then either be a hit (choosing the right symbol), or a miss (wrong symbol). Although the cued response test was appropriate in this manner, it would have been more interesting if the spontaneous identification or free recall test was also used. The free recall test is often regarded as the main standard of measuring symbol comprehension and often used to compare results using other test methods [6, 29].

The ETSI studies pointed out that although the test on proper recognition of the right symbols to their referents (hits) was the main index of performance, other parameters had to be noted as well. The other non-hit possibilities would be confusion of symbols to one another, failure to choose any symbol at all (no choice or missing values), subjective certainties and representativeness, as well as symbol and symbol set preference [3].

Lastly, testing graphic symbols needs proper contextualization as discussed by Wolff and Wogalter [29]. In real-life, symbols do not appear independently of the environment where they are used. The contexts or situations of how they are presented often contain hints or cues that limit the possibilities of what the symbols could be. Thus, there is a need for adequate information about the context on how the symbols will actually appear [3, 29]. In Böcker's MIA, the introduction of the study and description (illustration) of the basic components and functions of a videophone prior to the tests somehow helped the subjects to contextualize what the graphic symbols were all about.

### 2.2. Subject groups

The other critical factor affecting symbol evaluation is the type of subjects involved in the studies. Prospective users' involvement becomes more crucial when their backgrounds are vastly different as in eastern vs. western users, where one culture develops the products and their symbols, and another culture/s use/s them. Keesing [10] defined culture as something that also influences the "publicly available symbolic forms through which people experience and express meaning." Although the influence of culture on technology has been recognized, empirical studies are still inadequate [1, 22, 25]. In his studies on culture, technology and ergonomics, Shahnavaz [23] revealed that pictorial signs and symbols for public use are very much related to culture. He suggested that the best way to consider cultural aspects in the design is to adapt a user-centered design approach. The users' needs and backgrounds have to be reflected in the design and evaluation of the candidate symbols. However, icons and pictograms are generally designed and developed with "western users" in mind but targeted for international use. To compound these, the subjects used in some of the methods mentioned above could hardly even be described as representing typical users (industrial design students, professionals, and even communication employees) [26]. Obviously, all these matters need to be looked at and studied more to determine which among the symbols and their sets are best or most appropriate. Considering that most graphic symbols and their devices are targeted for worldwide use – among consumers of varying cultures, the challenge lies in developing test methods that are inexpensive, yet easy to conduct and comprehensive enough to cover the above factors.

## 3. OBJECTIVES OF THE STUDY

In view of the above, the present study aims to analyze the merits of simple pencil and paper tasks with multiple test parameters in evaluating graphic symbols among subjects of diverse nationalities. In the design of the methods, some critical factors based on the literature survey were considered. Specifically, Böcker's MIA was adopted but a spontaneous identification test was added for better comparison of the results. Since the graphic symbols studied were designed and tested in the West (Europe), subjects from the East (Southeast Asia) were used to help shed some light if culture indeed can influence symbol comprehension.

## 4. METHODS AND MATERIALS

### 4.1. Symbols used in the study

In the form of booklets, several graphic symbols representing videophone referents or functions were tested. Each symbol was approximately 1.2 x 1.2 cm in size. The study tested three sets of seven videophone symbols representing seven videophone functions or referents for a total of twenty-one symbols (Figure 1).

All the symbols were based on the ETSI study with one of these sets (Set 1) recommended by ETSI as the standard. The seven-videophone functions tested were the following [3]:

| | |
|---|---|
| a) Videophone / Telephone: | upgrading / downgrading the call |
| b) Camera on / off: | turning on and off picture transmission |
| c) Microphone on / off: | turning on and off sound transmission |
| d) Selfview on / off: | turning on and off selfview function |
| e) Document Camera on / off: | switching between document and person camera |
| f) Still Picture on / off: | turning on and off screen freeze |
| g) Handsfree on / off: | switching between handset and handsfree modes |

### 4.2. Subjects

One hundred ninety-two (99 males and 93 females) university students and employees of small and large companies from four Southeast Asian countries acted as subjects in the study. There were 48 subjects each from Indonesia, Malaysia, Philippines and Thailand. The mean age was 25.4 years (SD = 6.4), ranging from 17 to 52 years old. All were computer-literate or had at least experienced using computers or computer-related products for the past 2-5 years.

### 4.3. Procedure

The tests were in the form of questionnaires. They were done in small groups in each country, lasting for about 45-60 minutes. Although all the subjects could read and speak English, translations of the questionnaires were also provided in each country's native language. Subjects were randomly given one of three versions of the questionnaires. They were shown an illustration of a videophone. Its main parts and functions were then carefully discussed without revealing the actual referents to be tested. For example, terms such as "selfview" and "document camera" were not stated directly. Rather they were told that by pushing certain keys, one could see oneself or view documents. Afterwards, instructions were given on how to go about the different test parts. Questions were entertained prior to the administration of the tests. Emphasis was given on avoiding omissions in trials in order to get back to them later on. The order of the tests would also be strictly followed; that is, Part One followed by Parts Two and Three.

In the spontaneous identification or free recall test (Part One), each page contained one set of the seven videophone symbols, for a total of three pages for the three symbol sets. On spaces provided for, they would write what videophone function they thought was represented by each symbol. They were also asked to rate the level of certainty for each of their answers using a seven-point rating scale (from "very certain" to "very uncertain").

In the cued response test (Part Two), the subjects first read a referent and its description. Then they had to choose one symbol from a set of seven symbols which they thought best represented the referent in question. Each page contained one referent-description and one set of symbols. They were also asked to rate their certainties for their answers using the 7-point rating scales. There were seven videophone referents tested on three sets of symbols rendering a total of twenty-one trials in Part Two.

For the last part, three symbols (one for each set) together with the referent they represented were shown. The subjects had to choose one symbol they thought best represent that referent. Next, with the three symbols sets presented together, they had to choose the set they preferred most. In all, the subjects would choose seven symbols to represent the seven referents and one symbol set they preferred most.

### 4.4. Experimental Design

The study utilized a 2 x 5 x 3 (Occupation x Country x Symbol Set) repeated measures design with 2 between and 2 within factors. Between factors were those of occupation (students and employees) and country (Indonesia, Malaysia, Philippines, Thailand and Sri Lanka). The within factor consisted of the symbol sets (Sets 1, 2 and 3) and test types (spontaneous identification and cued response tests). The results for the subject groups by occupation in each country were collapsed since the focus in this part of the study was response differences between countries for different test parameters. In all three parts of the tests, the order of the symbol sets was counterbalanced across all subject groups. Three orderings of the symbol presentation and referents were generated and randomly distributed in each subject group to minimize order and learning effects. The test parameters studied were the following:

a) Hit rates (mean percentages of correct identification/recognition in Parts One and Two)

b) Subjective certainty ratings using 7-point scales for answers in Parts One and Two

c) Missing values (No answers in Parts One and Two)

d) Confusions for each videophone symbol (total numbers and frequency distributions)

e) Symbol and symbol set preferences (percent distribution, Part Four)

## 5. RESULTS

### 5.1. Hit rates and subjective certainties

In the spontaneous identification or free recall test, a hit (correct answer) occurs when the subject correctly identifies the symbol by writing its referent or function. For the cued response test, a hit pertains to correctly choosing the symbol representing the referent being asked. A one-between (country) and two-within (symbol set and test type) repeated measure ANOVAs were performed to compare the results in the two tests. Post hoc tests and pairwise multiple comparisons using Tukey's HSD were done for the significant main effects (symbol set, test type, and country at $p < 0.05$).

In testing symbols intended for public use, the Organization for International Standardization's ISO 9186 recommended at least 67% comprehension of the symbols [5]. This standard was intended more for symbols or signs appearing individually rather than as a group. Due to lack of a more suitable guideline, ISO 9186 was nonetheless adapted to roughly analyze and compare the results of the identification and recognition tests. Figures 2a-h contain the hit rates in the spontaneous identification (free recall) and the cued-response tests respectively for the three symbol sets.

a. *Between symbol sets.* Overall, Set 1's symbols had the highest hit rates compared to Sets 2 and 3 in almost all of the 7 referents or functions tested. The ANOVAs showed Set 1 symbols with significantly higher hit rates than Set 2 symbols in all seven referents, ($F(2, 376)$, $p < 0.05$). They were also significantly higher than Set 3 symbols except for "microphone", and "selfview". The hit rates of Sets 2 and 3 were not significantly different from each other except, for "microphone", $F(2, 376) = 419.72$, $p < 0.001$, "selfview", $F(2, 376) = 67.61$, $p < 0.001$, and "videophone", ($F(2, 376) = 347.91$, $p < 0.001$, where Set 3's hit rates were higher. Although Set 1 had the highest hit rates, only 3 of its symbols had hits above the 67% criteria of ISO 9186 ("camera" – 73.4%, "microphone" – 83.8% and "videophone" – 78.4%). Set 3 had 2 of its symbols above 67% ("microphone" – 81.0% and "videophone" – 70.0%). All of Set 2's symbols had hits below 67%.

b. *Between tests*. Part One was a much harder task compared to Part Two as proven in the results. Taking between tests as a main effect, the hit rates in the cued response tests (Part Two) were significantly higher than in the spontaneous identification tests (Part One) for all the 7 referents. However, when considering ISO 9186 in Part One, only the symbol of Sets 1 and 3 for "microphone" met the 67% correct criterion. In Part Two, the hit rates of Sets 1 and 3 for "camera", "microphone" and "videophone" and Sets 1 and 2's symbols for "document camera" had hits beyond 67%.

c. *Between countries.* Significant differences on hit rates between countries were noted only for the "handsfree" ($F(3, 188) = 3.40$, $p < 0.019$), and "videophone" ($F(3, 188) = 2.92$, $p < 0.035$). In the "handsfree" symbols, subjects from Thailand had significantly higher hit rates than the three other countries. The "videophone" symbols showed Malaysia having lower hit rate compared to Indonesia and Thailand. However, if the 67% criterion was considered, only Thailand had a hit rate above this level ("camera" – 69.4%).

d. *Subjective certainties.* Subjective certainty ratings for both Parts One and Two are also shown on Figures 2a-h (line graphs). The ratings were basically consistent with the identification and cued response data. Symbols with high hit rates exhibited high certainty ratings as well - again, mostly symbols from Sets 1and 3. Spearman's rank-order correlation was also applied between the parameters hit rates, subjective certainties and symbol preferences (Table 1). Correlation between hits and certainties is highest for Set 1's "camera", "handsfree", "microphone" and "videophone". There were more significant correlation between hit rates and subjective certainties than hit rates and preferences.

### 5.2. Confusion matrices

Confusions were instances when a symbol was associated to the wrong referents. They are another way of analyzing incorrect answers or misses. In this study, focus would be on those derived during the cued response test (Part Two). Based on the confusion matrices on Tables 2 and as summarized on Table 3, Set 2 had the highest instances of confusions and Set 1 having had the least confusions across all countries. This was expected since the set (Set 2) with the most misses would normally have the most confusions. Confusion instances are critical especially when the symbols are presented as a set or simultaneously. They indicate which among the symbols appear similar and thus would merit a re-design. Looking at all the three sets, the symbols for "videophone" and "selfview" had the highest instances of being confused with other functions. The symbols with the least instances of being confused with other referents were "handsfree" (Sets 1 and 3), "camera" (Set 2), and "document camera" (Set 3). The different countries were quite similar in the confusions of the symbols – easily identifying familiar symbols ("camera" and "document camera") and confusing them with other functions. On the other hand, the symbols for "selfview" and "videophone" were commonly confused with other referents since they were too general to specify

### 5.3. Missing values and preferences

Missing values refer to situations when subjects did not write any answer in Parts One and Two (Table 4). Part One, being the harder task, had more missing values than Part Two. Purely abstract symbols also tend to be identified least as shown by the results in Set 2 having the highest missing values for both Parts. Specifically, the symbols for "selfview", and "microphone" of Set 2 had the most number of missing values owing to their abstractness. There were no major differences between countries for missing values in Part One. Though in Part Two, Indonesia had missing values a bit higher compared to Malaysia, the Philippines and Thailand.

Tables 5 and 6 contain the mean symbol and symbol set preferences of all countries. The symbols of Set 1 was the preferred symbol by at least 50% (significant at $p<0.05$) of the subjects for document camera, handsfree, still picture, and videophone. For the other symbols, the subjects preferred Set 2’s “camera” and Set 3’s “microphone”. Between sets, about sixty-five percent of the subjects preferred Set 1.

### 5.5. Combined results

By considering the results of both the hit rates and subjective certainties in Parts One and Two, as well as confusions and preferences, Set 1’s symbols performed best in five of the seven referents tested. Exceptions were in symbols for “microphone”, and “selfview”, where symbols of Sets 1 and 3 performed equally better. If the results of the confusion matrices and preferences were added, Set 1’s “selfview” and Set 3’s “microphone” would be the best symbols for these referents. Set 2’s symbols were mostly abstract and ambiguous, garnering the lowest scores in most of the test parameters used in the study.

It is important to note also that in Part One all countries performed poorly, registering hit rates of at least 67% in only up to 3 of the 7 referents, mostly with Set 1's symbols (Figures 2a-h). These improved to 4-6 symbols with hit rates of at least 67% in Part Two. Between countries, Thailand had significantly higher hits for "handsfree", "microphone" (with the Philippines), "still picture" (with Malaysia), and "videophone" (with Indonesia). Although Malaysia and the Philippines had hits of 67% or more in only 1-2 of the symbols in Part One, they registered the biggest improvements in Part Two, having 5-6 symbols with at least 67% hit rates (Sets 1 and 3).

## 6. DISCUSSION

### 6.1. Hit rates and subjective certainties

Graphic symbols have become a popular part of interfaces of modern devices. They help make the products appear more attractive and useful with minimal space demands. But to the uninitiated user, graphic symbols can also easily make him confused, wasting unnecessary time and effort trying to make sense of everything. When a user chooses which among the present symbols will bring about the target function or information, several factors come into play. Different tests may thus be helpful in determining which symbol or symbol sets are best. This study examined some of these factors affecting symbol understanding using simple and inexpensive tests.

The most common gauge of symbol performance is the hit rate. The hit rate is the correct association between a referent and a symbol (videophone symbols in this case). In this study, hit rates were generated using two types of tests – the open-ended or spontaneous identification test and the cued response test. The former is usually designed to measure initial (a priori) meaning while the latter tests the associativeness of a symbol to its intended meaning or referent. Parts One and Two simulate (though in a limited manner) real videophone call scenarios wherein the symbols are placed on the terminal with one function to be performed at a time. Part One is a situation akin to using the technology the first time with minimal knowledge and instructions. Part One's results showed that graphic symbols could be quite difficult to understand or comprehend for the Asian subjects based on the very low hit rates. None of the subjects had ever used a videophone and this finding demonstrated that new users could indeed easily experience difficulties in using new technologies with graphic symbols as the sole interface and no other cues (learning aids, help functions) readily present. The higher hit rates of Part Two (cued responses) validated this point. With proper and simple verbal and written instructions, improvements in symbol recognition could usually be achieved [14, 22]. Performing Parts One and Two are essential in determining which symbols to retain, re-design or discard. Improving performance from Part One to Part Two shows that the symbols can be acceptable, but proper cues and learning aids are very important to retain recognition and understanding. Very low hit rates in both tests (such as those in Set 2) denote that the symbols are poorly understood and are not suitable for use by diverse groups.

Subjective certainties indicate the subjects' level of sureness that their choices of symbols to particular referents are right. High certainty levels of Set 1's "microphone" and "videophone" aptly showed these patterns. Furthermore, the hit rates generally had significant positive correlation with their subjective certainties (Figures 2a-h). Since there were more significant correlation's between hit rates and subjective certainties than hit rates and preferences, these may suggest preference as more a reflection of aesthetics and personal choice and may not always reflect correct understanding of the symbols.

### 6.2. Symbol distinctiveness – the confusion factor

As mentioned earlier, using hit rates alone cannot reveal the dynamics between symbols, especially when they are designed to appear together as a set. Confusion between symbols or instances when a symbol is associated to the wrong referents do occur and are thus also important. When symbols for the same referent have similar hit rates confusion matrices are helpful to determine which is better and more distinct. In the case of the symbols for "selfview", although Sets 1 and 3 appear to have similar hit rates, the former seemed to be better by having much lesser instances of confusion. The same was true for the videophone symbols of Sets 1 and 3. They had similar hit rates but the videophone of Set 1 seemed to be more distinct. Confusing one symbol to another may even be more hazardous than a miss since the former represents how often a symbol is mistakenly chosen under different referent conditions thereby decreasing its usability [3]. When confusions are concentrated to one or a few referents, they also represent the degree of confusing a symbol to another symbol presented at the same time. For example, Set 1's "microphone" was mostly associated with the referent "handsfree" by Indonesian and Malaysian subjects. Such situations could mean that the symbols might be so similar to each other that a re-design is in order to make them more distinct from one another.

### 6.3. Missing values and preferences

Missing values is another important way to gauge the quality of a symbol [3]. They are situations wherein the subjects did not know the answer or no idea of a symbol's meaning. In actual videophone usage scenario, the user could not identify which symbol represented the desired action. This factor, although seldom used, is still very helpful in identifying symbols subjects find difficult to understand and are easily derived from tests such as those in Parts One and Two. Preferences represent the subjects' personal judgement which of the symbols and sets best represent the referents in question (Tables 5 and 6). Table 6 shows clearly that Set 1 was the preferred set with five of its seven symbols significantly preferred over their counterparts. Since preference represent aesthetics (symbol's appeal to a subject) it can be very helpful when deciding symbols which have very similar results in all the other performance measures.

### 6.4. The cultural factor

One's nationality is a common indicator of one's culture. This study attempted to study if subjects from Southeast Asian countries of diverse cultures (nationality) differ in their recognition and understanding of graphic symbols. The results above showed that differences *as well as* similarities occurred between the Asian subject groups studied. For the latter, all of them performed very poorly in Part One, having high hits only for "microphone" and "camera" (Set 1 and 3). This could be explained by the fact that these symbols are very familiar and concrete representations of very familiar functions. The high improvements in hit rates in Part Two likewise showed that with proper instructions and cues, Asian subjects would not have major problems recognizing graphic symbols. Differences lay more when the symbols were probably new, unfamiliar or abstract such as the symbols for "handsfree", wherein Thailand had higher hits compared to the other countries, and "videophone" with

Indonesia and Thailand performing better than Malaysia. The missing value factor also helped qualify the misses or wrong answers. Among the 4 countries, Indonesia had the highest number of no answers with Thailand and the Philippines having the least. But compared to the European study, the Asian subjects had lesser instances of no answer [3]. This factor is interesting and requires further study since missing values could represent not only the inability to recognize the symbols ("did not know", "no idea") but also the level of conservatism when answering tests.

## 7. CONCLUSIONS

In this study, ETSI's recommended symbols (Set 1) turned to be the best and most preferred by the Asian subjects but may be poorly understood if they are introduced without adequate instructions and practice. Multiple test parameters were very valuable in determining initially which among the symbols will be readily understood and which will require more learning or even a re-design. They need not be expensive and could be done economically and conveniently using simple paper and pencil tasks.

Among the test parameters, the hit rate factor is still very important but insufficient, and by itself is too general an index. Certainty levels and missing values (no answers) could help further qualify the levels of hit rates attained by the symbols. The extent a symbol is wrongly associated to other referents (confusion) is likewise important, especially if together with other symbols they are designed to appear simultaneously as a set. And when symbols are intended for international use, using multiple test parameters become more useful in identifying not only the best symbols to use but also the changes or improvements needed by the different user groups.

## 8. ACKNOWLDGEMENTS

Acknowledgements are sincerely extended to the following: to Mr. Martin Böcker for graciously sharing his thoughts and opinions and all possible assistance regarding their ETSI studies on icons and pictograms; and to the Swedish Telecommunications Research Board (KFB) for the financial and logistical supports; to the ASEAN Ergonomics Society (Indonesia, Malaysia, Philippines and Thailand) for helping facilitate the conduction of the tests in these countries.

## 9. REFERENCES

[1] J. D. A. Abeysekera, H. Shahnavaz and L. J. Chapman, Ergonomics in developing countries, *Advances in Industrial Ergonomics and Safety II*, Taylor and Francis, London, 1990, pp. 771-778.

[2] J. W. Brelsford, M. S. Wogalter and J. A. Scoggins, Enhancing comprehension and retention of safety-related pictorials, *Proceedings of the Human Factors Ergonomics Society 38$^{th}$ Annual Meeting*, 1994, pp. 836-840.

[3] M. Böcker, A multiple index approach for the evaluation of pictograms, *Proceedings of the 14th International Symposium on Human Factors in Telecommunications, Darmstadt, Germany*, R. v. Decker's Verlag, G. Schenck GmbH, Heidelberg, 1993, pp. 73-85.

[4] A. M. Clarke and S. Brown, Pictograms in multipoint videotelephony: an international study, *Contemporary Ergonomics 1992*, Lovesey, E. J. (ed.), Taylor and Francis, London, 1992, pp. 494-499.

[5] S. Davies, H. Haines, B. Norris and J. Wilson, Safety pictograms: are they getting the message across? *Applied Ergonomics*, **29** (1998), 15-23.

[6] R. Dewar, Design and evaluation of graphic symbols, *Proceedings of Public Graphics, Utrecht, Netherlands*, University of Utrecht, Department of Psychonomics, 1994, pp. 24.1-24.18.

[7] R. S. Easterby and S. R. Hakiel, Field testing of consumer safety signs: the comprehension of pictorially presented messages, *Applied Ergonomics*, **12** (1981), 143-152.

[8] P. Green and R. W. Pew, Evaluating pictographic symbols: an automotive application, *Human Factors* **20** (1978), 103-114.

[9] S. Jones, Stereotipy in pictograms of abstract concepts, *Ergonomics*, **26** (1983), 605-611.

[10] R. M. Keesing, Theories of culture, *Annual Review of Anthropology*, **3** (1974), 73-79.

[11] R. Lin and J. G. Kreifeldt, Understanding the image functions for icon design, *Innovations for Interactions. Proceedings of the Human Factors Society 36th Annual Meeting, Atlanta, Georgia*, Human Factors Society, Santa Monica, 1992, pp. 341-345.

[12] R. Lin, J. G. Kreifeldt and C. F. Chi, A study of evaluation design sufficiency for iconic interface design: the design perspective, *Contemporary Ergonomics 1992*, Lovesey, E. J. (ed.), Taylor and Francis, London, 1992, pp. 376 - 384.

[13] K. N. Lodding, Iconic interfacing, *IEEE Computer Graphics and Applications*, **3** (1982), 11-20.

[14] A. M. L. Lund, Expert ratings of usability maxims, *Ergonomics in Design*, **5** (1997), 15-20.

[15] J. Mackett-Stout and R. Dewar, Evaluation of symbolic public information signs, *In Proceedings of Human Factors Society Conference*, 1981, pp. 139-151.

[16] M. C. Maguire, A review of human factors guidelines and techniques for the design of graphic human-computer interfaces, *Computers & Graphics* **9** (1985), 221-235.

[17] R. L. Magyar, Assessing icon appropriateness and icon discriminability with a paired comparison testing procedure, *In Proceedings of the Human Factors Society 34th Annual Meeting*, 1990, pp. 1204-1208.

[18] P. R. Nolan, Designing screen icons: ranking and matching studies, *Proceedings of the Human Factors Society 33rd Annual Meeting*, Human Factors, Santa Monica, 1989, pp. 380-384.

[19] Y. Rogers, Evaluating the meaningfulness of icon sets to represent command operations, *Proceedings of Human Computer Interaction '86*, British Computer Society Ed., 1986, pp. 586-603.

[20] A. Salasoo, Towards usable icon sets: a case study from telecommunications engineering, In: *Countdown to the 21st Century. Proceedings of the Human Factors Society 34th Annual Meeting*, Orlando, Florida, October 8 - 12, Vol. 1, Santa Monica: Human Factors Society, 1990, pp. 203 - 207.

[21] A. F. Sanders, Information processing and design, Chapter 29: Ergonomics, Work System Design, *Encyclopaedia of Occupational Health and Safety*, ILO, Geneva, 1998, pp. 29.71-29.75.

[22] H. Shahnavaz, Transfer of Technology to Industrially Developing Countries and Human Factors Considerations, Center for Ergonomics of Developing Countries, Luleå: Luleå University Press, Sweden, 1991.

[23] H. Shahnavaz, Cultural differences, Chapter 29: Ergonomics, Designing for Everyone, *Encyclopaedia of Occupational Health and Safety*, ILO, Geneva, 1998, pp. 29.79-29.83.

[24] R. B. Stammers, Judged appropriateness of icons as a predictor of identification performance, *Contemporary Ergonomics 1990*, Lovesey, E. J. (ed.), Taylor and Francis, London, 1990, pp. 371-376.

[25] P. Sukaviriya and L. Moran, User interface for Asia, *Designing Interfaces For International Use*, Nielsen, J. (ed.), Elsevier, Amsterdam, 1990, pp. 189-218.

[26] L. Tudor, The generation and testing of candidate telecommunications icons, *International Conference in Human-Computer Interaction Proceedings*, St. Petersburg, Russia, August 2-6, 1994, pp. 62-68.

[27] P. Vora, M. Helander, H. Swede and J. Wilson, Developing guidelines for symbol design a comparison of methodologies, *In Proceedings of Interface*, 1991, pp. 6-11.

[28] J. M. Webb, P. F. Sorenson and N. P. Lyons, An empirical approach to the evaluation of icons, *SIGCHI Bulletin*, **21**(1989), 87-90.

[29] J. S. Wolff and M. S. Wogalter, Comprehension of pictorial symbols: effects of contexts and test method, *Human Factors*, **40** (1998), 173–186.

[30] W. T. Wood and S. K. Wood, Icons in Everyday Life, *Social Ergonomics and Stress Aspects of Work with Computers*, 1987, 97-104.

[31] H. L. G. Zwaga, Predicting the public's comprehension of graphic symbols, *In Contemporary Ergonomics 1989*, E. D. Megaw, ed., Taylor and Francis, London, 1989, pp. 431-436.

TABLE 1. Correlation coefficients (Spearman's rho) of hit rates with subjective certainties and preferences.

| | Symbol Set | Subjective Certainties | Preferences |
|---|---|---|---|
| Camera | 1 | 0.281* | 0.162* |
| | 2 | 0.215* | 0.057 |
| | 3 | 0.267* | 0.041 |
| Document Camera | 1 | 0.135 | 0.162* |
| | 2 | 0.171* | 0.003 |
| | 3 | 0.007 | 0.004 |
| Handsfree | 1 | 0.260* | 0.034 |
| | 2 | 0.238* | 0.030 |
| | 3 | 0.105 | 0.005 |
| Microphone | 1 | 0.207* | 0.033 |
| | 2 | 0.069 | 0.017 |
| | 3 | 0.149* | 0.011 |
| Selfview | 1 | 0.039 | 0.073 |
| | 2 | 0.299* | 0.024 |
| | 3 | 0.181* | 0.101 |
| Still Picture | 1 | 0.211* | 0.035 |
| | 2 | 0.034 | 0.291* |
| | 3 | 0.251* | 0.190* |
| Videophone | 1 | 0.297* | 0.077 |
| | 2 | 0.138 | 0.047 |
| | 3 | 0.151* | 0.084 |

***correlation with hit rate is significant at P<0.05.***

Table 2. Confusion matrices per country (n = 48 per country) with hits (highlighted boxes), misses (columns) and confusions (rows) for the three symbol sets. First column in each set contains the seven videophone symbols and the second rows the seven referents or functions.

INDONESIA

| | Set 1 | | | | | | | Set 2 | | | | | | | Set 3 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HF | SV | SP | C | VP | D | M | H | SV | SP | C | VP | D | M | HF | SV | SP | C | VP | D | M |
| Handsfree | 19 | | | | | | 7 | 20 | 1 | | | 2 | | 18 | 17 | 3 | | | 1 | 1 | 7 |
| Selfview | 1 | 29 | 17 | 1 | 2 | | | 1 | 8 | 4 | 1 | 3 | 2 | 8 | 3 | 31 | 10 | 6 | 4 | 6 | |
| Still Picture | | 8 | 21 | 2 | | 1 | | 7 | 8 | 12 | 3 | 7 | | 3 | | 1 | 13 | 7 | | 6 | |
| Camera | 2 | 2 | 5 | 35 | 3 | 2 | | 2 | 3 | 2 | 26 | 7 | 3 | | 4 | 1 | 2 | 33 | 2 | 5 | 1 |
| Videophone | 10 | 5 | 1 | 2 | 42 | | 1 | 4 | 8 | 21 | 10 | 17 | 3 | 2 | 16 | 10 | 15 | 2 | 41 | 4 | 2 |
| Doc. Camera | | | 4 | 5 | 1 | 44 | | | 7 | 6 | 7 | 4 | 38 | | 1 | 2 | 7 | | | 25 | |
| Microphone | 13 | | | | | | 40 | 7 | 1 | | 1 | | 2 | 10 | 6 | | 1 | | | | 37 |

MALAYSIA

| | HF | SV | SP | C | VP | D | M | H | SV | SP | C | VP | D | M | HF | SV | SP | C | VP | D | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Handsfree | 18 | 1 | 1 | 1 | | | 4 | 13 | 2 | | 1 | 4 | | 15 | 18 | 1 | 1 | 1 | | | 5 |
| Selfview | 5 | 35 | 8 | 3 | 1 | 2 | | 9 | 10 | 3 | | 12 | | 11 | 1 | 30 | 17 | 6 | 2 | 6 | |
| Still Picture | 3 | 7 | 37 | 1 | | | | 2 | 4 | 20 | 2 | 4 | 2 | 1 | | | 17 | 9 | | 1 | |
| Camera | 2 | 1 | 1 | 34 | 2 | 2 | | 3 | 5 | 2 | 30 | 4 | 6 | 3 | 6 | | 2 | 31 | 1 | 6 | |
| Videophone | 7 | 4 | | 2 | 45 | | 2 | 4 | 18 | 15 | 10 | 8 | 2 | | 17 | 17 | 8 | | 42 | | 2 |
| Doc. Camera | 1 | | 1 | 7 | | 44 | | 1 | 8 | 7 | 2 | 10 | 37 | 1 | 1 | | 3 | 1 | 2 | 35 | |
| Microphone | 12 | | | | | | 42 | 15 | 1 | 1 | 3 | 6 | 1 | 16 | 4 | | | | 1 | | 41 |

PHILIPPINES

| | HF | SV | SP | C | VP | D | M | H | SV | SP | C | VP | D | M | HF | SV | SP | C | VP | D | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Handsfree | 17 | 2 | | | | 1 | 4 | 15 | 3 | | | 5 | 1 | 16 | 16 | | | 1 | | | 3 |
| Selfview | 4 | 39 | 11 | 1 | 1 | | | 5 | 15 | 10 | 3 | 12 | | 10 | 3 | 38 | 17 | 4 | 3 | 5 | 2 |
| Still Picture | 3 | 2 | 32 | 1 | | 2 | | 7 | 8 | 12 | 3 | 7 | 1 | 3 | 3 | 1 | 22 | 5 | | 5 | |
| Camera | | | 1 | 40 | 1 | 5 | | | 2 | 5 | 30 | 2 | 4 | | 2 | 1 | 4 | 34 | 2 | 4 | |
| Videophone | 12 | 5 | 1 | | 45 | 1 | 1 | 7 | 18 | 12 | 6 | 15 | 7 | | 11 | 8 | 3 | 1 | 38 | 1 | 1 |
| Doc. Camera | 4 | | 3 | 6 | 1 | 39 | | 2 | 1 | 7 | 4 | 5 | 33 | | 1 | | 2 | 3 | 5 | 33 | |
| Microphone | 8 | | | | | | 43 | 11 | | 2 | 1 | 2 | 2 | 19 | 12 | | | | | | 41 |

THAILAND

| | HF | SV | SP | C | VP | D | M | H | SV | SP | C | VP | D | M | HF | SV | SP | C | VP | D | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Handsfree | 27 | | | | | | 4 | 22 | | 2 | | 1 | | 18 | 23 | | 2 | 1 | | | 4 |
| Selfview | 3 | 29 | 15 | | | | | 3 | 15 | 3 | 1 | 4 | 1 | 7 | 2 | 33 | 16 | 4 | 3 | | |
| Still Picture | 1 | 9 | 24 | 2 | | 1 | | 1 | 1 | 25 | 3 | 10 | 2 | | 1 | | 17 | 12 | | 2 | |
| Camera | 3 | 4 | 4 | 40 | 3 | 5 | 1 | 1 | 3 | 1 | 30 | 6 | 7 | | 2 | 3 | 7 | 29 | 6 | 9 | |
| Videophone | 4 | 5 | 2 | 2 | 44 | | | 2 | 20 | 13 | 9 | 16 | | 3 | 8 | 11 | 6 | 1 | 37 | | 1 |
| Doc. Camera | | 1 | 3 | 4 | 1 | 42 | | 1 | 4 | 4 | 5 | 3 | 38 | | | 1 | | | 2 | 37 | |
| Microphone | 8 | | | | | | 43 | 16 | 2 | | | 6 | | 17 | 11 | | | | | | 43 |

*Note: HF =Handsfree, SV = Selfview, SP = Still Picture; CA = Camera; VP = Videophone; DC = Document Camera; MP = Microphone.*

Table 3. Total number of confusions per country in Part Two (Cued response tests).

| | Set 1 | | | | Set 2 | | | | Set 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ind. | Mal. | Phi. | Tha. | Ind. | Mal. | Phi. | Tha. | Ind. | Mal. | Phi. | Tha. |
| Camera | 14 | 8 | 7 | 20 | 17 | 23 | 13 | 18 | 15 | 15 | 13 | 27 |
| Doc. Camera | 10 | 9 | 14 | 9 | 24 | 29 | 19 | 17 | 10 | 7 | 11 | 3 |
| Handsfree | 7 | 7 | 7 | 4 | 21 | 22 | 25 | 21 | 12 | 8 | 4 | 7 |
| Microphone | 13 | 12 | 8 | 8 | 11 | 27 | 18 | 24 | 7 | 5 | 12 | 11 |
| Selfview | 21 | 19 | 17 | 18 | 19 | 35 | 40 | 19 | 29 | 32 | 34 | 25 |
| Still Picture | 11 | 11 | 8 | 13 | 28 | 15 | 29 | 17 | 14 | 10 | 14 | 15 |
| Videophone | 19 | 15 | 20 | 13 | 48 | 49 | 50 | 47 | 49 | 44 | 25 | 27 |

Table 4. Missing Values per country (Note: Ind.=Indonesia, Mal.=Malaysia, Phi.=Philippines, Tha.=Thailand. The first column in each country denotes missing values in Part One and the second column for missing values in Part Two.)

| | SET 1 | | | | | | | | SET 2 | | | | | | | | SET 3 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | IND | | MAL | | PHI | | THA | | IND | | MAL | | PHI | | THA | | IND | | MAL | | PHI | | THA | |
| Camera | 5 | 3 | 1 | 0 | 3 | 0 | 2 | 0 | 8 | 0 | 3 | 0 | 8 | 1 | 4 | 0 | 6 | 0 | 5 | 0 | 3 | 0 | 4 | 1 |
| Doc. Camera | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 8 | 0 | 7 | 0 | 9 | 0 | 7 | 1 | 13 | 0 | 6 | 0 | 4 | 0 |
| Hands-free | 5 | 3 | 2 | 0 | 3 | 0 | 3 | 2 | 8 | 7 | 2 | 1 | 6 | 1 | 2 | 2 | 5 | 1 | 1 | 0 | 3 | 0 | 4 | 1 |
| Micro-phone | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 7 | 21 | 0 | 20 | 0 | 13 | 3 | 2 | 1 | 1 | 0 | 0 | 1 | 2 | 0 |
| Selfview | 5 | 4 | 3 | 0 | 8 | 0 | 1 | 0 | 19 | 11 | 10 | 0 | 12 | 1 | 10 | 3 | 4 | 0 | 7 | 0 | 5 | 0 | 7 | 0 |
| Still Picture | 2 | 0 | 4 | 0 | 5 | 0 | 2 | 0 | 13 | 3 | 10 | 0 | 10 | 0 | 8 | 0 | 2 | 0 | 1 | 0 | 3 | 0 | 2 | 0 |
| Video-phone | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 10 | 8 | 6 | 0 | 4 | 0 | 2 | 2 | 1 | 0 | 3 | 0 | 1 | 0 | 2 | 0 |
| Total | 19 | 11 | 11 | 0 | 20 | 0 | 9 | 2 | 90 | 36 | 60 | 1 | 67 | 3 | 48 | 10 | 27 | 3 | 31 | 0 | 21 | 0 | 25 | 2 |

TABLE 5. Symbol preferences in (%) with all countries combined (N = 240).

| | Symbol Preferences | | |
|---|---|---|---|
| | Set 1 | Set 2 | Set 3 |
| Camera | 30.2 | 35.9 | 33.3 |
| Doc. Camera | 52.6 | 19.8 | 27.6 |
| Handsfree | 52.1 | 19.3 | 23.4 |
| Microphone | 38.5 | 0.5 | 54.7 |
| Selfview | 48.4 | 8.3 | 43.2 |
| Still Picture | 56.8 | 22.4 | 20.3 |
| Videophone | 68.2 | 5.2 | 26.6 |

TABLE 6. Symbol set preferences in (%) per country.

| | Set Preferences | | |
|---|---|---|---|
| | Set 1 | Set 2 | Set 3 |
| Indonesia | 68.8 | 4.1 | 27.1 |
| Malaysia | 58.3 | 6.3 | 35.4 |
| Philippines | 68.8 | 0.0 | 31.2 |
| Thailand | 64.6 | 6.2 | 29.2 |
| *Means* | *65.1* | *4.2* | *30.7* |

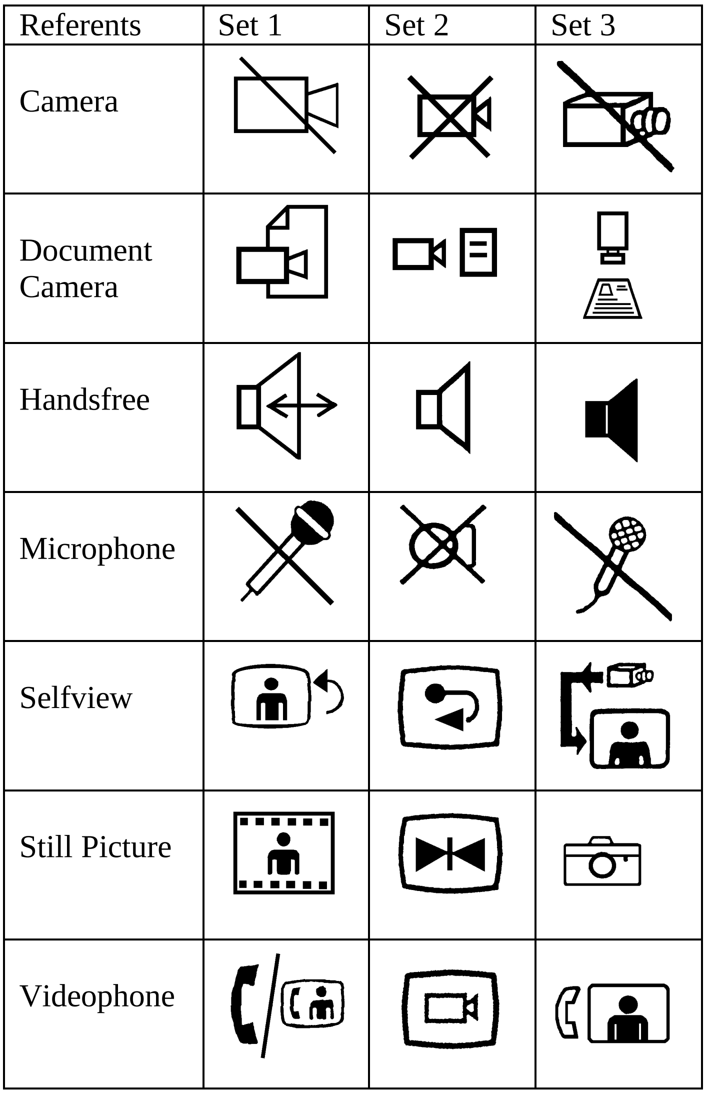



Figure 1. The seven referents and the three sets of videophone symbols used in the study.

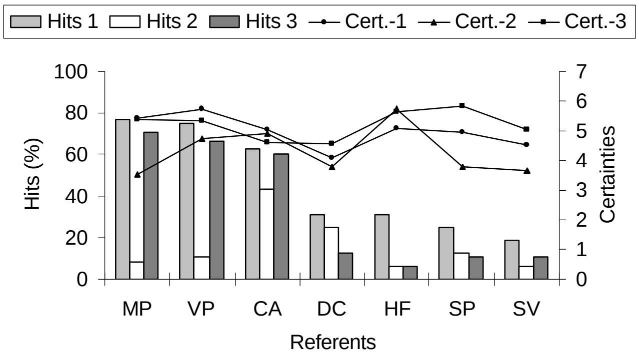



Figure 2a. Part One, Indonesia

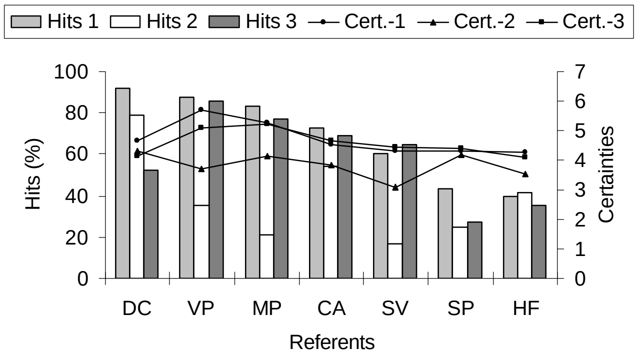



Figure 2b. Indonesia, Indonesia

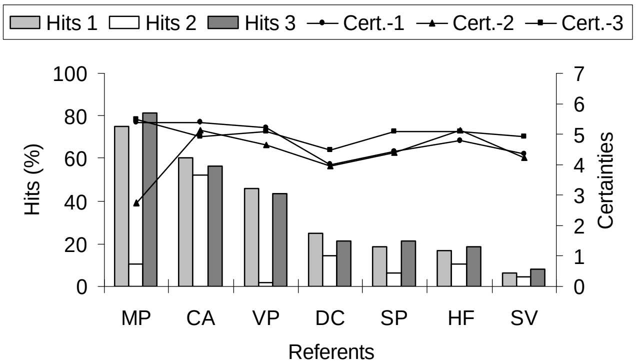



Figure 2c. Part One, Malaysia

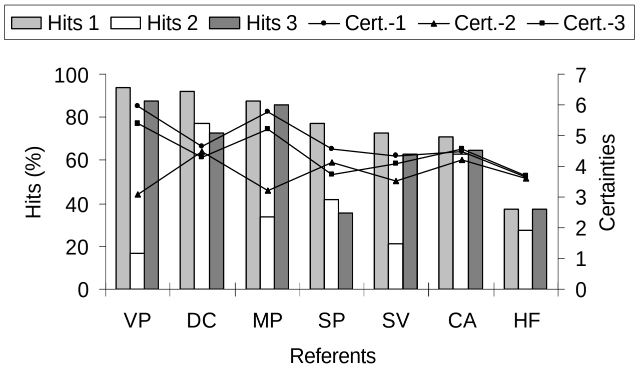



Figure 2d. Part Two, Malaysia

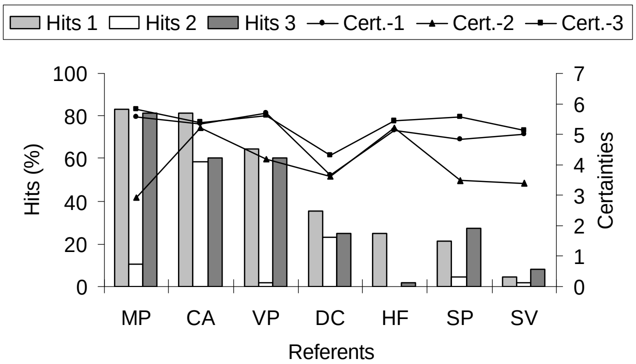



Figure 2e. Part One, Philippines

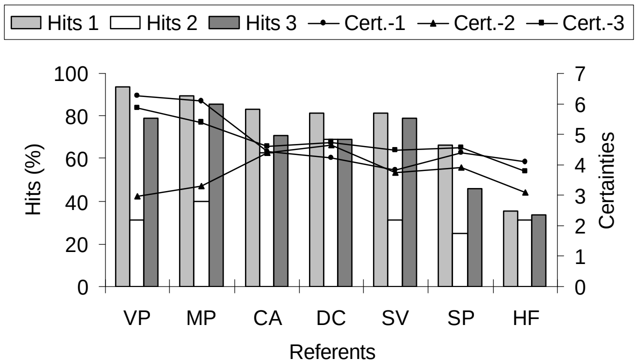



Figure 2f. Part Two, Philippines

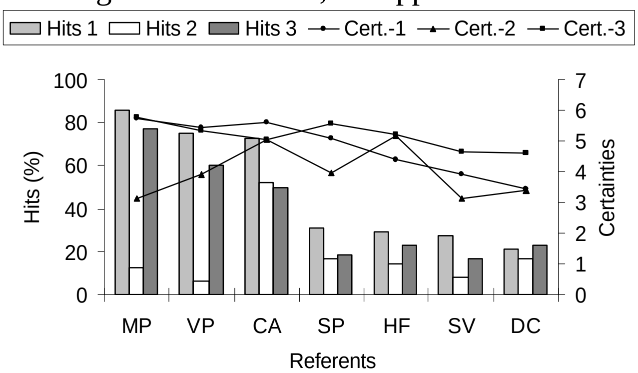



Figure 2g. Part One, Thailand

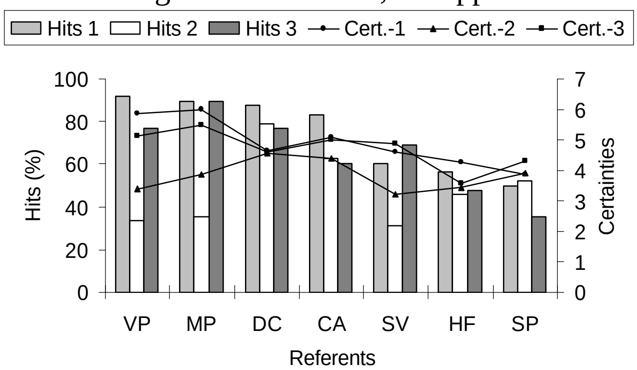



Figure 2h. Part Two, Thailand

Figures 2a-h. Parts One and Two Hit rates and Subjective Certainties for the 3 symbol sets. Bar graphs represent the Hit rates (%) while line graphs represent Subjective Certainties (n = 240. CA=Camera, DC=Document Camera, HF=Handsfree, MP=Microphone, SV=Selfview, SP=Still Picture, VP=Videophone).


**PAPER 3**

# THE RELEVANCE OF ERROR ANALYSIS IN GRAPHICAL SYMBOLS EVALUATION

# THE RELEVANCE OF ERROR ANALYSIS IN GRAPHICAL SYMBOLS EVALUATION

Dominic Paul T. Piamonte
Industrial Ergonomics Division
Human Work Sciences Department
Luleå University of Technology, 971 87 Luleå, Sweden

## ABSTRACT

In an increasing number of modern tools and devices, small graphical symbols appear simultaneously in sets as parts of the human-machine interfaces. The presence of each symbol can influence the other's recognisability and correct association to its intended referents. Thus, aside from correct associations, it is equally important to perform certain error analysis of the wrong answers, misses and confusions, and even lack of answers. This research aimed to show how such error analyses could be valuable in evaluating graphical symbols especially across potentially different user groups. The study tested three sets of icons representing seven videophone functions. The methods involved parameters such as hits, confusions, missing values and misses. The association tests showed similar hit rates of most symbols across majority of the subject groups. However, exploring the error patterns helped detect differences in the graphical symbols' performances between subject groups, which otherwise seemed to have similar levels of recognition. These are very valuable not only in determining the symbols to be retained, replaced or re-designed, but also in formulating instructions and other aids in learning to use new products faster and more satisfactorily.

## 1. INTRODUCTION

### 1.1. Advent of Graphical Symbols

Under the current globalisation, it is an increasing challenge for ergonomists to ensure the usability, comfort and safety of new technologies and their products across varied user groups. Graphical symbols (icons and pictograms) play an important part in achieving this goal. From their traditional use of appearing singly or in isolation to denote locations or services, they have started to appear simultaneously in sets, as actual parts of the human-machine interfaces of modern devices. They are now also used to present the concepts, ideas or objects, and functions of the technology and products where they are used. When properly designed, tested and taught, graphical symbols can help introduce new technologies and products to different user groups across the world regardless of culture or language groups.

### 1.2. Methods in Evaluating Graphical Symbols

As in all products designed for public use, graphical symbols need to be evaluated which among them are best suited across all intended user groups. The test methods to be employed are often dependent on the kind of study being pursued. Different methods have been used in evaluating symbols such as icons and pictograms (Magyar, 1990; Nolan, 1989; Webb, Sorenson & Lyons, 1989, and Vora, Helander, Swede & Wilson, 1991). If symbols are intended as parts of interfaces of devices for international use or for standardisation, different tests are usually needed. For example, the International Standards Organisation (ISO) has come out with the ISO 9186. It is a six-stage procedure for the development and testing of public information symbols (Zwaga, 1989). Its major portions are the comprehensibility judgement tests, comprehension test with a suggested comprehension level of at least 67%, and matching tests. The American National Standards Institute likewise has the ANSI Z535.3 (ANSI, 1991) describing the methods of evaluating the comprehensibility of graphical symbols. It utilises four categories namely correct answer, wrong answer, critical confusion and no answer. It further recommends a comprehension level of at least 85%. On the other hand, the International Telecommunications Union (ITU-T) has F.910. Recommendation F.910 endorses a symbol testing procedure composed of four parts. It involves the determination of need for new symbols, the creation and evaluation of the new designs, and the selection and approval (ITU-T, 1995).

Another important aspect of symbols evaluation is the need to explain and test the functions (represented by the symbols) in manners reflecting typical user scenarios (Tudor, 1994). As mentioned earlier, numerous modern computer-based devices contain small graphical symbols simultaneously appearing as groups as part of the general control interface. The most appropriate way to evaluate the symbols in such cases is to present each candidate set of symbols against one target referent at a time. For example, in a typical user scenario a prospective user intends to use a device or equipment and is confronted with controls represented by a set of graphical symbols. He/she then has to choose which among these symbols correspond to his desired function. One such method was developed by Böcker (1993) with the Human Factors Technical Committee of the European Telecommunications Standards Institute (ETSI). They evaluated videotelephony symbols using a multiple index approach (MIA) which can also be used in testing symbols for other commercial products. The MIA tested several sets of candidate videophone symbols by presenting each set to the subject with only one referent at a time. This approach enabled the tester to collect seven indices that could help in the final selection of symbols. The indices were hit rates, false alarm rates, missing values, subjective certainty and suitability, symbol and symbol set preferences (Böcker, 1993).

### 1.3. Importance of Analysing Errors

The hit rate in MIA reflected the levels of correct associations of the symbols to their intended referents, and was the main index of performance. However, it has been emphasised that hit rates alone cannot be the sole basis of determining how well a symbol is understood especially if it appears together with other symbols. Analysing the patterns or characteristics of errors such as misses, false alarms or confusions with other symbols, and even missing values are equally important (Böcker, 1993 and ETSI, 1993). Misses are instances of selecting the wrong symbol in the context of a particular function being sought. Their patterns of distribution can give some knowledge if the referents themselves are conceptually clear to the users. False alarms are instances of a symbol being wrongly selected under different non-corresponding functions or referents. Missing values are instances wherein no response or no answers were given. A hit or correct association is just one of the possible outcomes in symbol testing. Misses and other forms of errors are the others that need to be closely studied as well. These considerations become more valuable when faced with potentially different user groups.

This paper aimed to show how such analyses of errors, i.e. their distribution pattern and other characteristics, could be valuable in evaluating graphical symbols especially across potentially different user groups. The study was part of an international project in evaluating telecommunication icons using multiple inter-related test parameters. Some of these are the errors or so-called "non-hit parameters", which are the foci of this paper. As stated earlier, graphical symbols always have the potential of widespread usability among different user groups, but they still need to be tested when the groups in question are as diverse as in Asia. This paper focused on some countries in Southeast Asia - a region characterised by an immense diversity in language and culture. Modern products and the symbols they contain are often designed and tested in the west but targeted for worldwide use. Thus, a region such as the Southeast presents a big challenge to the designers of such products to make them usable and thus, commercially viable as well. It is then essential to evaluate products and their interfaces with these potential users in mind. It is hypothesised that simple recognition tests eliciting hit rates may fail to exhibit differences on how symbols are understood by different user groups. Other parameters, as discussed above, would be utilised to detect other possible differences in symbol recognition by different user groups. In the design of the study, the MIA by Böcker (1993) as well as Tudor's (1994) comments and recommendations were greatly considered.

## 2. METHODS

### 2.1. Subjects

Two hundred forty university students and employees and professionals (127 males and 113 females) from small and large companies from five Asian countries participated in this study. These countries were Indonesia, Malaysia, Philippines, Thailand and Sri Lanka. There were forty-eight subjects per country with equal numbers of students and employees, all of whom were computer-literate or had at least experienced using computers or computer-related products for the past 2-5 years. At the time of the study, none of the subjects ever experienced using a videophone. The mean age was 25.1 years (SD = 6.1), ranging from 17 - 52 years old. These participants came from the major cities of these countries and the tests conducted in their schools, offices and homes.

### 2.2. Materials

The stimuli were three sets of symbols representing seven referents or functions of a videophone for a total of twenty-one symbols (Figure 1). The referents were "camera", "document camera", "handsfree", "microphone", "selfview", "still picture" and "videophone." Both referents and symbols were based on the ETSI study headed by Böcker (1993). The sets were grouped based on the outcome of the ETSI tests. Set 1 had the best performance based on the MIA tests (and was recommended by ETSI as the standard), followed by Set 3, and Set 2 as last. These three sets were chosen also based on their symbols' general attributes such concreteness or abstraction, another factor being studied by the project. The questionnaires were divided into three parts: symbol identification/subjective certainty tests (Part I), symbol association/subjective certainty tests (Part II), and symbol and set preferences (Parts III and IV). Three versions of the questionnaires with different orders of the symbols and sets were generated and used in the study. This paper would deal on the results of Part II.

### 2.3. Procedure

The tests were done in small groups in each country in their schools, homes or offices, lasting for about 45-60 minutes. The subjects were randomly given one of the three versions of the questionnaires. The versions differed only in symbol and referent orderings. An illustration of a videophone was shown and its general functions were then discussed. Afterwards, instructions were given on how to go about the different test parts. Questions were entertained prior to the conduction of the tests. Emphasis was given on avoiding skips in trials in order to get back to them later. The order of the tests would also be strictly followed: i.e., Part I followed by Parts II, then Parts III and IV.

In the cued response test (Part II), the subjects first read a referent and its description. Then they were asked to select, by putting a mark such as circle or x, one symbol from a set of seven symbols they thought best represented the referent in question. Each page contained one referent description and one set of symbols. Subjective certainties for their answers using the 7-point rating scales (from ”very certain” to ”very uncertain”) were likewise done. There were seven videophone referents tested on three sets of symbols rendering a total of twenty-one trials.

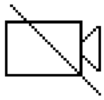
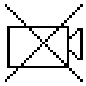
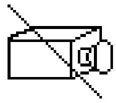
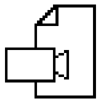
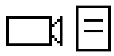
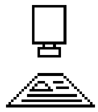
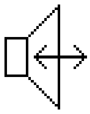
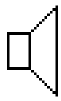
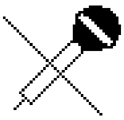
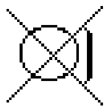
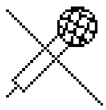
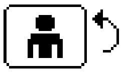
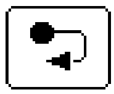
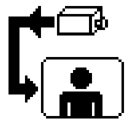
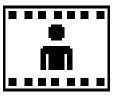
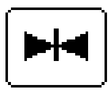
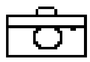
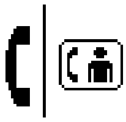
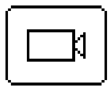
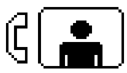
| **Referents** | **Set 1** | **Set 2** | **Set 3** |
|---|---|---|---|
| **1. Camera** |  |  |  |
| **2. Document Camera** |  |  |  |
| **3. Handsfree** |  |  |  |
| **4. Microphone** |  |  |  |
| **5. Selfview** |  |  |  |
| **6. Still Picture** |  |  |  |
| **7. Videophone** |  |  |  |

**Figure 1. The videophone referents and the three sets of symbols used in the study (Böcker, 1993).**

## 3. RESULTS AND DISCUSSION

### 3.1. Correct Associations and Subjective Certainties

Percentages of correct associations (hit rates) and subjective certainties in the cued-response tests are shown on Table 1. Correct association resulted when a subject correctly matched the referent in question to its right symbol. Correct responses were labelled as "1" (hits) and wrong responses as "0" (misses). Missing data or "no responses" were initially treated as "0" also during this stage of the analysis. Thus, the response data were binomial percentages and arcsine transformations are warranted to ensure that assumptions in normality were met. Such transformations of percentages eliminate subjects as a variable leaving only the independent variables and lowering the degrees of freedom. A repeated measures factorial design (multi-factor ANOVA) was then used to determine significant differences in the hit rates. Games-Howell post hoc tests were done to analyse significant main effects.

**TABLE 1. Hit Rates in the Cued-Response Tests and Subjective Certainties among the three sets of graphical symbols representing seven videophone functions, n = 48 per country.**

| **HITS** | **Camera** | | | **Doc. Camera** | | | **Handsfree** | | | **Micro-phone** | | | **Selfview** | | | **Still Picture** | | | **Video-phone** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Countries** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** |
| Indonesia | 73 | 54 | 69 | 92 | 79 | 52 | 40 | 42 | 35 | 83 | 21 | 77 | 60 | 17 | 65 | 44 | 25 | 27 | 88 | 35 | 85 |
| Malaysia | 71 | 63 | 65 | 92 | 77 | 73 | 38 | 27 | 38 | 88 | 33 | 85 | 73 | 21 | 63 | 77 | 42 | 35 | 94 | 17 | 88 |
| Philippines | 83 | 63 | 71 | 81 | 69 | 69 | 35 | 31 | 33 | 90 | 40 | 85 | 81 | 31 | 79 | 67 | 25 | 46 | 94 | 31 | 79 |
| Thailand | 83 | 63 | 60 | 88 | 79 | 77 | 56 | 46 | 48 | 90 | 35 | 90 | 60 | 31 | 69 | 50 | 52 | 35 | 92 | 33 | 77 |
| Sri Lanka | 42 | 42 | 54 | 75 | 35 | 46 | 19 | 15 | 25 | 67 | 29 | 75 | 40 | 21 | 40 | 38 | 48 | 8 | 81 | 13 | 75 |
| **Means** | 71 | 57 | 64 | 86 | 68 | 64 | 38 | 33 | 36 | 84 | 32 | 83 | 63 | 25 | 64 | 55 | 39 | 31 | 90 | 26 | 81 |

| **CERTAINTIES** | **Camera** | | | **Doc. Camera** | | | **Handsfree** | | | **Micro-phone** | | | **Selfview** | | | **Still Picture** | | | **Video-phone** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Countries** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** |
| Indonesia | 4.2 | 3.8 | 4.7 | 4.6 | 4.3 | 4.1 | 4.1 | 3.1 | 4.0 | 5.3 | 3.7 | 5.1 | 4.0 | 2.6 | 4.4 | 4.3 | 3.9 | 4.4 | 5.7 | 3.2 | 5.1 |
| Malaysia | 4.5 | 4.2 | 4.6 | 4.6 | 4.4 | 4.3 | 3.7 | 3.6 | 3.7 | 5.8 | 3.2 | 5.2 | 4.3 | 3.5 | 4.1 | 4.6 | 4.1 | 3.8 | 5.9 | 3.1 | 6.1 |
| Philippines | 4.4 | 4.3 | 4.6 | 4.2 | 4.6 | 4.8 | 4.1 | 3.1 | 3.8 | 6.1 | 3.3 | 5.3 | 3.8 | 3.7 | 4.5 | 4.4 | 3.9 | 4.6 | 6.2 | 3.0 | 5.9 |
| Thailand | 5.1 | 4.4 | 5.0 | 4.7 | 4.6 | 4.6 | 4.1 | 3.4 | 3.5 | 6.0 | 3.7 | 5.5 | 4.6 | 3.1 | 4.9 | 3.8 | 3.9 | 4.3 | 5.8 | 3.3 | 5.1 |
| Sri Lanka | 5.3 | 4.4 | 5.6 | 4.2 | 4.4 | 4.7 | 4.6 | 4.1 | 4.3 | 5.1 | 4.2 | 4.7 | 4.8 | 4.0 | 4.2 | 5.1 | 4.5 | 4.6 | 5.6 | 3.8 | 4.8 |
| **Means** | 4.9 | 4.6 | 5.5 | 4.7 | 4.9 | 5.1 | 4.3 | 3.9 | 4.5 | 5.9 | 4.0 | 5.8 | 4.5 | 3.8 | 5.0 | 4.6 | 4.5 | 4.9 | 6.0 | 3.7 | 6.0 |

The hypotheses tested in this part were if hit rates would differ between sets and between countries. The data on Table 1 showed that between the three symbol sets (main effect) and except for the "handsfree" symbols, Set 1 symbols had significantly higher hit rates than Sets 2 and 3 in the seven referents tested ($F(2, 30)$, $p < 0.05$). These results were expected partly because Sets 1 and 3 symbols were more concrete and representative of the referents they represented. With country as another main effect, Indonesia, Malaysia, Philippines, and Thailand had comparably similar levels of hit rates. Only Sri Lanka showed significantly lower hit rates compared to the other countries in the symbols for "camera", "document camera", and "handsfree" ($F(4, 15)$, $p < 0.05$) – results that would be discussed in detail later. Regarding the certainty ratings, they were generally higher for Sets 1 and 3 compared to Set 2. The certainty scores of the symbols for microphone and videophone were either very high (Sets 1 and 3) or very low (Set 2). However, Kruskal-Wallis tests showed only a few significant differences between countries regarding their certainty ratings (Table 2).

Overall, when ISO 9186 and ANSI Z535.3 comprehension levels were considered through hit rates, most of the tested videophone symbols performed poorly. Combining country results, only seven of the 21 symbols reached ISO's required comprehension level of at least 67%. These were Set 1 symbols for "camera" (71%), "document camera" (86%), "microphone" (84%), and "videophone" (90%); Set 2 symbol for "document camera" (68%); and Set 3 symbols for "microphone" (83%), and "videophone" (81%). If ANSI Z535.3 was the basis, only two of the symbols would've met its 85% comprehension level (Set 1 symbols for "document camera", 86%, and "videophone", 90%). Thus, by looking merely at the hit rates, the following could be stated. Very few of the symbols performed well regardless of country, and most of these were from Set 1. The symbols most poorly understood were basically abstract in their general attributes (Set 2). Moreover, when comparing countries, only Sri Lanka seemed to differ significantly compared to the other countries. Notwithstanding that some other cultural differences also exist and except for Sri Lanka, the four other countries lie fairly close to each other, with standards of living and education quite similar as well. This can thus partially explain the results but admittedly will need more studies to be confirmed. However, by analysing the errors or non-hit parameters of the test procedures, more differences among the symbols and countries were discovered as discussed below.

**TABLE 2. Kruskal-Wallis Tests for Subjective Certainties for each symbol (*n* = 48 per country, *df* =4, *p*<0.05).**

| Symbol Sets | Chi-Square | Asymp. Sig. |
|---|---|---|
| **Set 1** | | |
| Camera | 12.12 | .017 |
| Document Camera | 1.51 | .826 |
| Handsfree | 4.95 | .292 |
| Microphone | 6.96 | .138 |
| Selfview | 8.15 | .086 |
| Still Picture | 12.87 | .012 |
| Videophone | 1.03 | .906 |
| **Set 2** | | |
| Camera | 3.57 | .468 |
| Document Camera | 6.71 | .348 |
| Handsfree | 7.26 | .292 |
| Microphone | 5.08 | .279 |
| Selfview | 16.51 | .002 |
| Still Picture | 3.95 | .413 |
| Videophone | 3.78 | .437 |
| **Set 3** | | |
| Camera | 10.35 | .035 |
| Document Camera | 3.87 | .424 |
| Handsfree | 4.01 | .405 |
| Microphone | 2.10 | .718 |
| Selfview | 3.25 | .517 |
| Still Picture | 5.83 | .212 |
| Videophone | 7.47 | .113 |

### 3.2. Confusion Matrices and Misses

When symbols appear simultaneously or as a set and the main task is to pick out which among these symbols correspond to a desired function or object, there are other possible outcomes aside from a hit. These are the so-called non-hit outcomes or parameters such as misses, confusions and missing values (failure to respond or no answer). The wrong answers during the cued response tests were analysed using confusion matrices (Tables 4 to 8). In this particular part of the study, tabulations were made on the rates of confusions – the rate wherein each symbol was selected in the context of a different referent. Confusions can be classified into two types: symmetric and asymmetric (Nolan, 1989). The former occurs when subjects chose symbol *x* when presented with referent *y*, and symbol *y* when presented with referent *x*. Symmetric confusions usually suggest visual or conceptual similarities. These were exemplified by the “selfview”/“still picture” (visual similarity) and “”handsfree”/”microphone” symbols of Set 1 (conceptual similarity). Another case of symmetric confusion due to visual similarities would be that of the “camera”/”document camera “ symbols of Set 3. These symmetric confusions were present in all countries. Thus, the problem may not be “culturally-linked” but a design problem – that the symbols concerned were visually too similar. Moreover, according to basic ergonomic design principles, these symbols would then be prime considerations for re-design or replacement to make them more distinct from the other.

**TABLE 3. Confusion Matrices, Indonesia, n = 48.**

| | **Symbol Selected** | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Referent** | **Set 1** | | | | | | | **Set 2** | | | | | | | **Set 3** | | | | | | |
| **Presented** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | – | 10 | | | 2 | 4 | 4 | – | 15 | | 2 | 2 | 6 | 21 | – | | | | 13 | 15 | 4 |
| **2**. Doc. Camera | 4 | – | | | | 2 | | 6 | – | | 4 | 4 | | 6 | 10 | – | 2 | | 13 | 13 | 8 |
| **3**. Handsfree | 4 | | – | 27 | 2 | | 21 | 4 | | – | 15 | 2 | 15 | 8 | 8 | 2 | – | 13 | 6 | | 33 |
| **4**. Microphone | | | 15 | – | | | 2 | | | 38 | – | 17 | 6 | 4 | 2 | | 15 | – | | | 4 |
| **5**. Selfview | 4 | | | | – | 17 | 10 | 6 | 15 | 2 | 2 | – | 17 | 17 | 2 | 4 | 6 | | – | 2 | 21 |
| **6**. Still Picture | 10 | 8 | | | 35 | – | 2 | 4 | 13 | | | 8 | – | 44 | 4 | 15 | | 2 | 21 | – | 31 |
| **7**. Videophone | 6 | 2 | | | 4 | | – | 15 | 8 | 4 | | 6 | 15 | – | 4 | | 2 | | 8 | | – |

Note: Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

**TABLE 4. Confusion Matrices, Malaysia, n = 48.**

| Referent Presented | Symbol Selected: Set 1 | | | | | | | Set 2 | | | | | | | Set 3 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | – | 15 | 2 | | 6 | 2 | 4 | – | 4 | 2 | 6 | | 4 | 21 | – | 2 | 2 | | 13 | 19 | |
| **2**. Doc. Camera | 4 | – | 0 | | 4 | | | 13 | – | | 2 | | 4 | 4 | 13 | – | | | 13 | 2 | |
| **3**. Handsfree | 4 | 2 | – | 25 | 10 | 6 | 15 | 6 | 2 | – | 31 | 19 | 4 | 8 | 13 | 2 | – | 8 | 2 | | 35 |
| **4**. Microphone | | | 8 | – | | | 4 | 6 | 2 | 31 | – | 23 | 2 | | | | 10 | – | | | 4 |
| **5**. Selfview | 2 | | 2 | | – | 15 | 8 | 10 | 17 | 4 | 2 | – | 8 | 38 | | | 2 | | – | | 35 |
| **6**. Still Picture | 2 | 2 | 2 | | 17 | – | | 4 | 15 | | 2 | 6 | – | 31 | 4 | 6 | 2 | | 35 | – | 17 |
| **7**. Videophone | 4 | | | | 2 | | – | 8 | 21 | 8 | 13 | 25 | 8 | – | 2 | 4 | | 2 | 4 | | – |

Note: Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

**TABLE 5. Confusion Matrices, Philippines, n = 48.**

| Referent Presented | Symbol Selected: Set 1 | | | | | | | Set 2 | | | | | | | Set 3 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | – | 13 | | | 2 | 2 | | – | 8 | | 2 | 6 | 6 | 13 | – | 6 | 2 | | 8 | 10 | 2 |
| **2**. Doc. Camera | 10 | – | 2 | | | 4 | 2 | 8 | – | 2 | 4 | | 2 | 15 | 8 | – | | | 10 | 10 | 2 |
| **3**. Handsfree | | 8 | – | 17 | 8 | 6 | 25 | | 4 | – | 23 | 10 | 15 | 15 | 4 | 2 | – | 25 | 6 | 6 | 23 |
| **4**. Microphone | | | 8 | – | | | 2 | | | 33 | – | 21 | 6 | | | | 6 | – | 4 | | 2 |
| **5**. Selfview | | | 4 | | – | 4 | 10 | 4 | 2 | 6 | | – | 17 | 38 | 2 | | | | – | 2 | 17 |
| **6**. Still Picture | 2 | 6 | | | 23 | – | 2 | 10 | 15 | | 4 | 21 | – | 25 | 8 | 4 | | | 35 | – | 6 |
| **7**. Videophone | 2 | 2 | | | 2 | | – | 4 | 10 | 10 | 4 | 25 | 15 | – | 4 | 10 | | | 6 | | – |

Note: Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

**TABLE 6. Confusion Matrices, Thailand, n = 48.**

| Referent Presented | Symbol Selected | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Set 1** | | | | | | | **Set 2** | | | | | | | **Set 3** | | | | | | |
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1.** Camera | – | 8 | | | | 4 | 4 | – | 10 | | | 2 | 6 | 19 | – | | 2 | | 8 | 25 | 2 |
| **2.** Doc. Camera | 10 | – | | | | 2 | | 15 | – | | | 2 | 4 | | 19 | – | | | | 4 | |
| **3.** Handsfree | 6 | | – | 17 | 6 | 2 | 8 | 2 | 2 | – | 33 | 6 | 2 | 4 | 4 | | – | 23 | 4 | 2 | 17 |
| **4.** Microphone | 2 | | 8 | – | | | | | | 38 | – | 15 | | 6 | | | 8 | – | | | 2 |
| **5.** Selfview | 8 | 2 | | | – | 19 | 10 | 6 | 8 | | 4 | – | 2 | 42 | 6 | 2 | | | – | | 23 |
| **6.** Still Picture | 8 | 6 | | | 31 | – | 4 | 2 | 8 | 4 | | 6 | – | 27 | 15 | | 4 | | 33 | – | 13 |
| **7.** Videophone | 6 | 2 | | | | | – | 13 | 6 | 2 | 13 | 8 | 21 | – | 13 | 4 | | | 6 | | – |

Note: Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

**TABLE 7. Confusion Matrices, Sri Lanka, n = 48.**

| Referent Presented | Symbol Selected | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Set 1** | | | | | | | **Set 2** | | | | | | | **Set 3** | | | | | | |
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1**. Camera | – | 10 | 8 | | 17 | 8 | 15 | – | 19 | | 6 | 8 | 6 | 19 | – | 2 | 2 | | 13 | 21 | 8 |
| **2**. Doc. Camera | 8 | – | 4 | 2 | 2 | 4 | 4 | 19 | – | 4 | 17 | | 6 | 19 | 31 | – | 4 | | 8 | 6 | 4 |
| **3**. Handsfree | 13 | 8 | – | 21 | 15 | 4 | 21 | 4 | 10 | – | 21 | 25 | 19 | 6 | 8 | 13 | – | 10 | 15 | | 29 |
| **4**. Microphone | 2 | 2 | 17 | – | 2 | 2 | 8 | | 10 | 19 | – | 31 | 6 | 4 | 6 | 4 | 4 | – | 10 | | |
| **5**. Selfview | 4 | 8 | 4 | | – | 31 | 13 | 8 | 10 | 2 | 8 | – | 19 | 31 | 6 | 15 | 6 | 2 | – | | 31 |
| **6**. Still Picture | 6 | 6 | 4 | | 44 | – | 2 | 4 | 17 | 4 | 2 | 2 | – | 23 | 4 | 15 | 4 | | 42 | – | 25 |
| **7**. Videophone | 2 | 8 | 2 | | 6 | | – | 8 | 40 | 4 | 8 | 17 | 10 | – | 2 | 15 | 2 | 2 | 4 | | – |

Note: Figures by column represent confusion in percentages, when the symbol in each column was wrongly associated to the videophone referent (function) on first column. Figures by row represent distribution of misses per referent. Referents were presented one at a time, and symbols one set at a time per referent.

Asymmetric confusions occur when subjects simply chose the wrong symbol for a given referent. For example, Table 4 showed that Set 1's symbol for "document camera" (1) was wrongly selected by respondents from Indonesia when the referents "camera" (10.4%), "still picture" (8.3%) and "videophone" (2.1%) was presented. Between sets, Set 2 clearly had the most number of confusions above this level and Set 1 the least in all five countries.

Across symbol sets, “selfview” and “videophone” symbols had the highest instances of being mistaken as representing other functions. More differences were noted when the countries were compared based on symbol confusions. For example, while Thailand had the most instances of confusing Set 1’s symbol for “camera” than the other four countries, it had the lowest instances of confusing Set 1’s videophone symbol for other functions. In asymmetric confusions, the problem may lie in the vagueness or too much generality of the symbol/s in question with regards to the other referents. Thus, they can easily be associated with several referents. The results above showed the countries with varying patterns of asymmetric confusions. These could be useful in deciding which symbol (or referent) in each country may need more explanations and tests to avoid vagueness and misconceptions of the functions being represented.

Confusions are also very useful indicators of the suitability of the symbol and may even complement the hit rates. For example, the videophone symbols of Sets 1 and 3 had very similar hit rates among the countries. Nevertheless, when symbol confusion was also considered, Set 1’s version was better because of its lesser number of confusions. With regards to the microphone symbols wherein subjects from the five countries had similar hit rates also, Set 3’s version performed best for Indonesia, Malaysia and Sri Lanka (lower symbol confusions). In the case of the Philippines and Thailand, Set 1’s microphone would be the better choice.

Misses are the opposite of hits. These occur when the subject selected the wrong symbol for the referent being presented. Tables 3 to 7 illustrate the misses (in percent) in rows in accordance to the referents it was wrongly associated. The essential thing with misses is that its pattern of distribution under each referent can give an idea of the understandability of the referent or function itself. For example, the referent handsfree was associated with almost all the symbols of in all sets in quite high levels. This could suggest that the concept of “handsfree” might not be entirely clear to most of the subjects in all countries.

### 3.3. Missing Values

Missing values were instances when some of the subjects gave no response or answers during the cued response tests. Table 8 illustrates the distribution of the missing values across countries for each set. Between sets, Set 2 had the highest instances of missing values. Between countries, Thailand had the most instances of missing values among all countries. Its missing values ranged from 2% to 23% across all 21 symbols. Sri Lanka was the opposite with no missing values in the cued response test. When arrayed with the other non-hit parameters or errors, they are also very important since they represent situations wherein the prospective user plainly lacks the knowledge of which among the symbols represent the desired function (referent). In actual user scenarios, they can be akin to non-use or under-use of the device.

**TABLE 8. Missing Values or no answers given in Cued Response Tests per set (n = 48 per country).**

| | Camera | | | Doc. Camera | | | Hands-free | | | Micro-phone | | | Selfview | | | Still Picture | | | Video-phone | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Countries** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** |
| Indonesia | 6 | – | – | 2 | – | 2 | 6 | 15 | 2 | – | 15 | 2 | 8 | 23 | – | – | 6 | – | – | 17 | – |
| Malaysia | – | – | – | – | – | – | – | 2 | – | – | – | – | – | – | – | – | – | – | – | – | – |
| Philippines | – | 2 | – | – | – | – | – | 2 | – | – | – | 2 | – | 2 | – | – | – | – | – | – | – |
| Thailand | – | – | 2 | – | – | – | 4 | 2 | 2 | – | 6 | – | – | 6 | – | – | – | – | – | 4 | – |
| Sri Lanka | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |

## 4. CONCLUSIONS

In this study, three sets of videophone symbols designed and tested in the west were tested among subject groups from the east (Southeast Asia). And when symbols are intended as parts of interfaces of devices for international use, different tests are usually needed. Usually, the different parts are designed to measure initial (a priori) meaning and appropriateness, ease of learning and remembering, as well as the probability of confusion with other symbols. The cued response test simulates (though in a limited manner) real videophone call scenarios wherein the symbols are placed on the terminal with one function to be performed at a time. It depicts situations using the videophone the first few times with bits of information and instructions given. Determination of the hit rates gives the level of association of the symbol to its intended referent or function. However, when symbols appear simultaneously, the hit rates cannot reveal the dynamics of how each symbol can affect the other's level comprehension and association to its referent. Thus, analyses of the errors or non-hit parameters are crucial. Confusions are instances when symbols are selected under the wrong referent contexts. Together with hit rates, they can give an idea on a symbol's distinctness from other symbols under different referent contexts as well as the different patterns of their interactions in different groups. Subjective certainties are also important. They reflect the subjects' degree of confidence or confusion with their answers. Other parameters also exist, which are not discussed but are also helpful in evaluating graphical symbols. For example, preferences generally indicate the aesthetics. They are useful especially when considering which among symbols with similar hit rates are more appealing for the subjects. Lastly, this study was just one part of an international project. It attempted to demonstrate how multiple indices, including those that analyse non-hits can be helpful in evaluating symbols across different user groups. Different patterns in errors between countries were shown, but admittedly, deeper analyses are still needed to convert such observations into more practical guides towards designing symbols across diverse user groups – issues that would be addressed by the succeeding parts of the project.

Altogether, the results of this study supports the view that analyses of errors or non-hit parameters such as confusions, misses, and missing values are invaluable in properly evaluating graphical symbols when they appear simultaneously or in groups. They enable the tester to see other often subtle but important differences on how different users perceive and understand symbols. These are important not only in properly interpreting test results of different user groups but also in formulating instructions and other aids in learning to use new products faster and more satisfactorily.

**Acknowledgements**

- To Mr. Martin Böcker for his invaluable support in acquiring the videophone symbols used in this study and in sharing his thoughts and expertise throughout.
- To *KFB* (Swedish Communications Board) for the financial and other logistical assistance.

## REFERENCES

American National Standards Institute (ANSI). (1991). *Accredited standard on safety colors, signs symbols, labels and tags (Standard No. ANSI Z535.1-5).* Washington, DC, USA: National Electrical Manufacturers Association.

Böcker, M. (1993). A multiple index approach for the evaluation of pictograms. In *Proceedings of the 14th International Symposium on Human Factors in Telecommunications, Darmstadt, Germany* (pp. 73-85). Heidelberg, Germany: R. v. Decker's Verlag, G. Schenck GmbH.

European Telecommunications Standards Institute (ETSI). (1993). *Technical Report - Human Factors (HF): Results of an evaluation study of pictograms for point-to-point videotelephony*. Sophia Antipolis, France: ETSI.

International Telecommunications Union (ITU). (1995) *Recommendation F.910 (02/95) - Procedures for designing, evaluating and selecting symbols, pictograms and icons*. Geneva: ITU.

Magyar, R. L. (1990). Assessing icon appropriateness and icon discriminability with a paired comparison testing procedure. In *Countdown to the 21st Century. Proceedings of the Human Factors Society 34th Annual Meeting, Orlando, Florida* (pp. 1204-1208). Santa Monica, CA, USA: Human Factors Society.

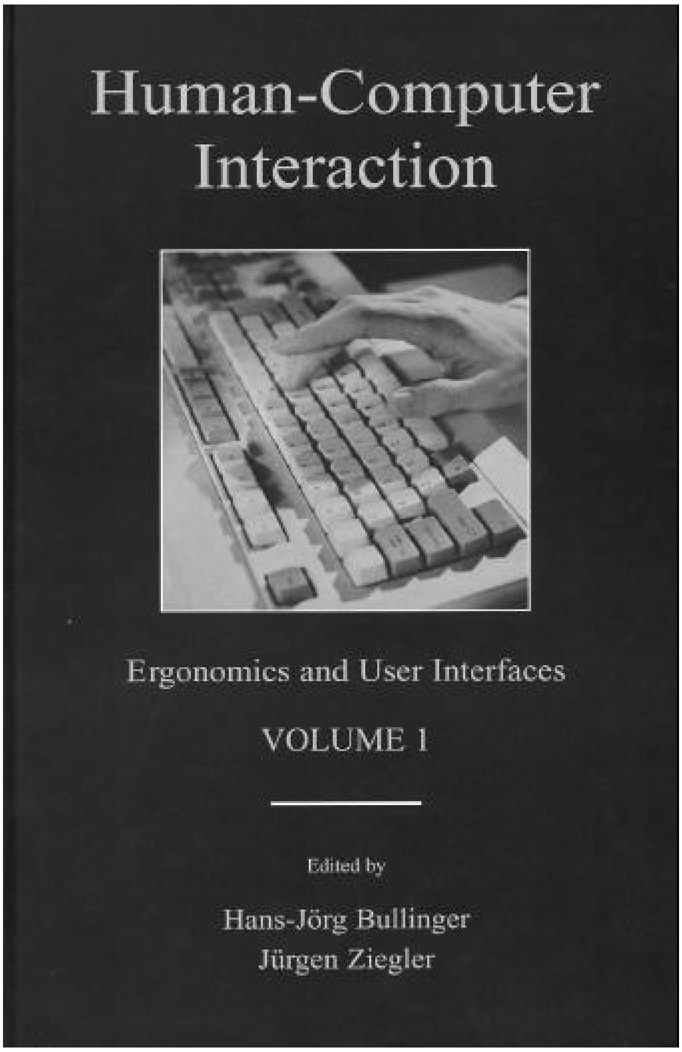
Nolan, P. R. (1989). Designing screen icons: ranking and matching studies. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 380-384). Santa Monica, CA, USA: Human Factors Society.

Tudor, L. G. (1994). The generation and testing of candidate telecommunications icons, In Blumenthal, B., Gornostaev, J., and Unger, C. (Eds.), *Proceedings of the East-West International Conference in HCI, St. Petersburg, Russia* (pp. 11-62). Moscow, Russia: International Centre for Scientific and Technical Information (ICSTI).

Vora, P., Helander, M., Swede, H., & Wilson, J. (1991). Developing guidelines for symbol design: a comparison of methodologies. In *Proceedings of Interface 1991* (pp. 6-11). Dayton, OH, USA: University of Dayton.

Webb, J. M., Sorenson, P. F., & Lyons, N. P., 1989, An empirical approach to the evaluation of icons. *SIGCHI Bulletin*, *21*, 87-90.

Zwaga, H. L. G. (1989). Predicting the public's comprehension of graphic symbols. In Megaw, E. D. (Ed.), *Contemporary Ergonomics 1989* (pp. 431-436). London, UK: Taylor & Francis.




# PAPER 4

# TESTING VIDEOPHONE GRAPHICAL SYMBOLS IN SOUTHEAST ASIA

# TESTING VIDEOPHONE GRAPHICAL SYMBOLS IN SOUTHEAST ASIA

**Piamonte, D. P. T.[1], Abeysekera, J. D. A.[1] and Ohlsson, K.[2]**
[1]Division of Industrial Ergonomics, [2]Division of Engineering Psychology, Department of Human Work Sciences, Luleå University of Technology, Sweden

## Introduction

Southeast Asia is a region characterized by diversity in language and culture. Although it is slowly recovering from the recent economic crisis, it still remains as one of the most dynamic regions in the world. Economic trades and technology transfers from the west still abound in most of its countries. One palpable sign is the continued rise in the use of western products. With these also came the use of graphical symbols (icons and pictograms) as a common mode of user interface of the products. Easy learnability, less space requirements and being non-text dependent are just few of the reasons why graphical symbols have become a major means of presenting information and functions especially in travel, communication and consumer products (Zwaga, 1989 and Horton, 1994). In order to elicit the above benefits of using graphical symbols, they must be tested across their intended user groups. Symbols are often designed and tested in the west but targeted for international use. Thus, the extent of their usability and acceptability among user groups with different languages and cultures from those of the west can be questionable. Studies on how icons and pictograms in Asia are still sparse. This paper aimed to study how symbols designed and tested in Europe fared among subjects from different countries in Southeast Asia. It was part of an international study on graphic symbols involving countries from Southeast Asia (Piamonte, 1998). Specifically, videotelephone symbols designed and tested by the European Telecommunication Standards Institute (ETSI) (Böcker, 1993) were evaluated using subjects from five Southeast Asian countries.

## Methods and Materials

Three sets of videophone symbols were derived studies by Böcker (1993) for the European Telecommunications Standards Institute (ETSI), including the set (Set 1) eventually proposed by ETSI as the standard (Figure 1). The symbols were tested in the form of questionnaires. Each symbol set contained seven symbols representing seven videophone functions.

Testing of the subjects were done around Metro-Manila, Philippines, Bangkok, Thailand, Bandung, Indonesia, Sarawak, Malaysia, and Sri Lanka, and in were and in small groups in each country, lasting for about 45-60 minutes. Subjects were randomly given one of three versions of the questionnaires. The subjects were shown an illustration of a videophone. Its general functions were then discussed. Afterwards, instructions were given on how to go about the different test parts. Questions were entertained prior to the administration of the tests. Emphasis was given on avoiding omissions in trials in order to get back to them later on.

In the spontaneous identification test (Part One), each page contained one set of the seven videophone symbols, for a total of three pages for the three symbol sets. On spaces provided for, they would write what videophone function they thought was represented by each symbol. They were also asked to rate the level of certainty for each of their answers using a seven-point rating scale (from " very certain" to "very uncertain"). Part Two was a cued response test patterned after a method developed by Böcker (1993) in testing simultaneously appearing symbols. The subjects first read a referent and its description. Then they had to choose one symbol from a set of seven symbols which they thought best represented the referent in question. Each page contained one referent-description and one set of symbols. They were also asked to rate their certainties for their answers using the 7-point rating scales. There were seven videophone referents tested on three sets of symbols rendering a total of twenty-one trials in Part Two.

## Results and Discussion

Two hundred forty-eight subjects (127 males and 113 females) from the five Asian countries acted as respondents. The mean age was 25.12 years (SD=6.08, Range=17-52 years). There were forty-eight subjects per country with equal numbers of students and employees. In both Parts One and Two, correct responses resulted when a subject correctly identified (Parts One) or matched (Part Two) the referent in question to its right symbol. Correct responses were labeled as "1" (hits) and wrong responses as "0" (misses). Missing data or "no responses" were initially treated as also "0" during this stage of the analysis.

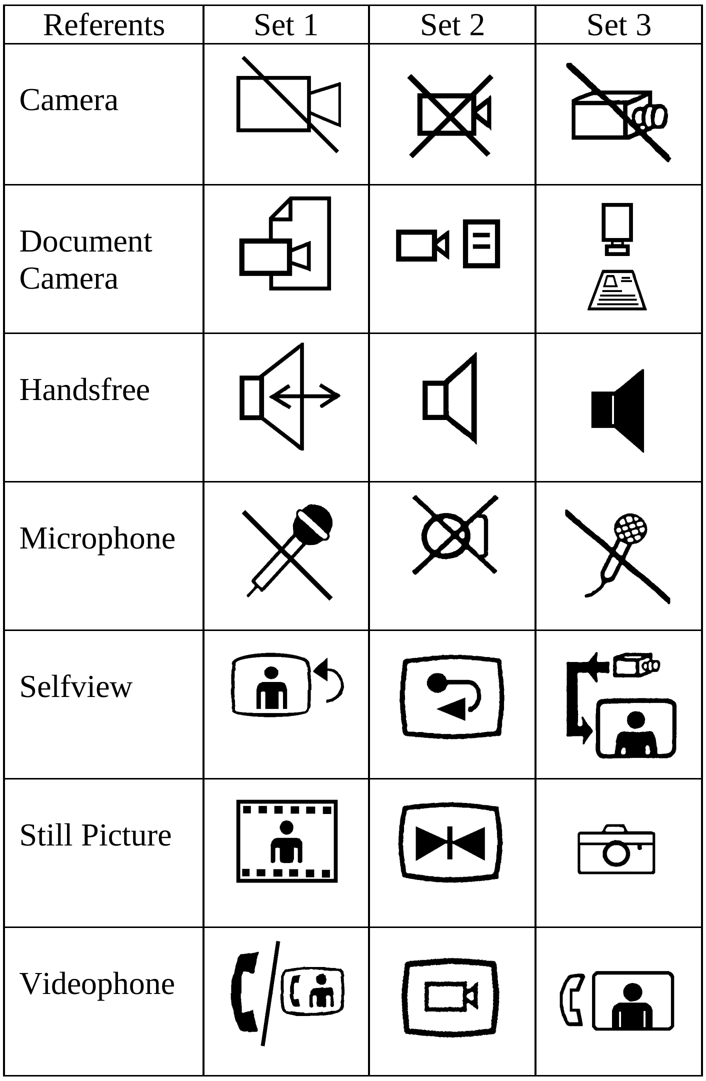



FIGURE 1. Graphical symbols used in the present study as based on the studies by Böcker (1993) for the European Telecommunications Standards Institute.

The response data were thus basically binomial percentages and arcsine transformations are warranted to ensure that assumptions in normality were met. Such transformations of percentages eliminate subjects as a variable leaving only the independent variables and lowering the degrees of freedom. A repeated measures factorial design (multi-factor ANOVA) with Tukey's honestly significant differences as post hoc tests were done to determine any significant differences in the hit rates based on test type, subject group and symbol set. Kruskal-Wallis tests were likewise done to compare certainty ratings.

Between tests, correct responses in the cued response tests (Part Two) were significantly higher (at $F(1, 111)$, $p<0.05$) than in the spontaneous identification tests (Part One), except for the symbols for "microphone". Between symbol *sets*, Set 1 had significantly higher hits in both tests for all referents except for "document camera" and "handsfree" ($F(2, 111)$, $p<0.05$). On the other hand, Set 2 garnered the lowest hits in both Parts One and Two of the test. Furthermore, only the symbols (of Sets 1 and 3) for referents "microphone" and "videophone" garnered hit rates of at least 67% in Part One. Altogether in both tests and across all symbol sets, the highest hit rates were for "camera", microphone" and "videophone" while the "document camera", "handsfree", "selfview" and "still picture" had the lowest hit rates in all countries. Comparing countries, there were significant differences ($F(4, 111)$, $p<0.05$) in hit rates for symbols for "camera", document camera", "handsfree". In these symbols, Sri Lankan subjects generally had the lowest hits compared to the other four countries. Certainty ratings were mostly higher for Sets 1 and 3 compared to Set 2 in both Part One and Part Two. Certainty ratings of the symbols for microphone and videophone were either very high (Sets 1 and 3) or very low (Set 2) in both Parts One and Two. No significant differences were noted when subjects were compared based on occupation (student vs. employees/professionals).

Confusions can be classified into two types: symmetric and asymmetric (Nolan, 1989). The former occurs when subjects chose symbol *x* when presented with referent *y*, and symbol *y* when presented with referent *x*. Symmetric confusions usually suggest visual or conceptual similarities. Asymmetric confusions occur when subjects simply chose the wrong symbol for a given referent. Tabulations made on the rates of confusions showed more differences between countries. For example, while Thailand had the most instances of confusing Set 1's symbol for "camera" than the other four countries, it had the lowest instances of confusing Set 1's videophone symbol for other functions. Confusions are very useful indicators of the suitability of the symbol and may even complement the hit rates. For example, the videophone symbols of Sets 1 and 3 had very similar hit rates among the countries. Nevertheless, when symbol confusion was also considered, Set 1's version was better because of its lesser number of confusions. With regards to the microphone symbols wherein subjects from the five countries had similar hit rates also, Set 3's version performed best for Indonesia, Malaysia and Sri Lanka (lower symbol confusions). In the case of the Philippines and Thailand subjects, Set 1's microphone would be the better choice. Lastly, missing values were instances when some of the subjects gave no response or answers during the cued response tests. Between countries, Thailand had the most instances of missing values among all countries. Its missing values ranged from 2% to 23% across all 21 symbols. Sri Lanka was the opposite with no missing values in the cued response test. Missing values or subjects giving no answers as a parameter is a very good gauge of situations wherein the subjects did not know the answer because he/she couldn't understand the referent or the symbols presented. In actual user scenarios, this would be akin to the user not knowing which control would cause the target function (Böcker ,1993).

Altogether, some both similarities and differences were noted regarding symbol understanding among Southeast Asian subjects. The results in the spontaneous identification and cued response tests in testing videophone symbols showed very low rates of correct identification and even correct association of the symbols across all countries.

Differences between countries lie mostly in the levels of confusing the symbols from each other and their missing values. Both results showed the potential difficulties encountered by people from this region of recognizing and using these symbols. However, it is still believed that graphical symbols undeniably are still very useful. When conditions warrant that they be used alone or as a primary interface, user testing, adequate information, training as well as other help procedures become very important.

## References

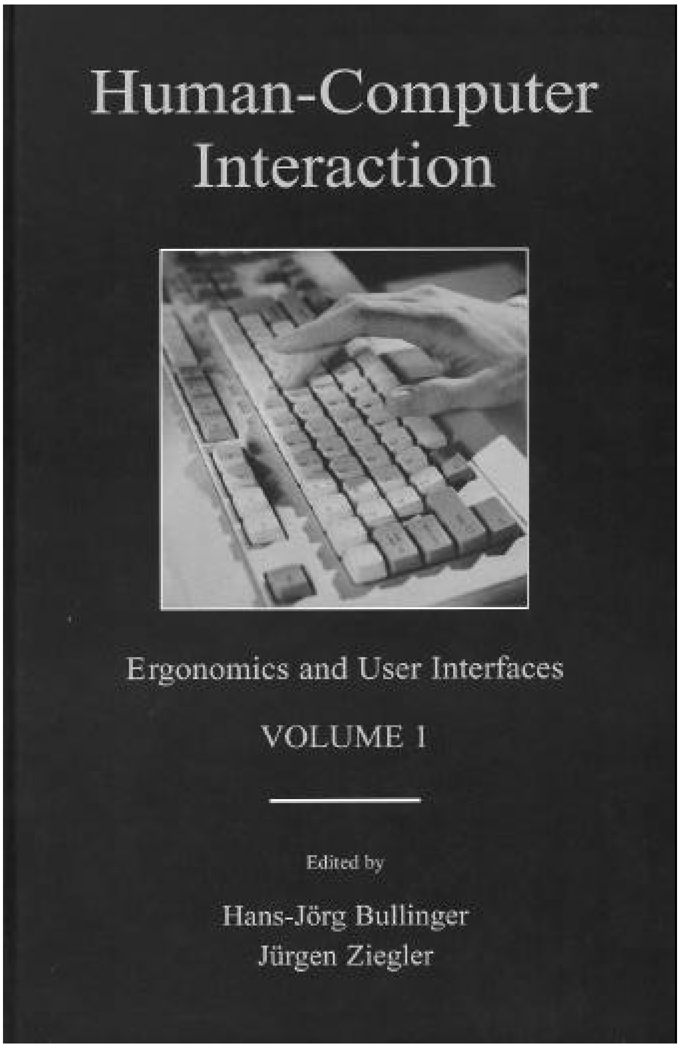

Böcker, M. (1993). A multiple index approach for the evaluation of pictograms, In: *Proceedings of the 14th International Symposium on Human Factors in Telecommunications*, Darmstadt, Germany, May 11 - 14, Heidelberg: R. v. Decker's Verlag, G. Schenck GmbH, pp. 73-85.

Horton, W. (1994). *The Icon Book – Visual symbols for computer systems and documentation*, New York: Wiley and Sons.

Nolan, P. R. (1989). Designing screen icons: ranking and matching studies. In *Proceedings of the Human Factors Society 33rd Annual Meeting*, Santa Monica: Human Factors, pp. 380-384.

Piamonte, D. P. T. (1998) *A User-based Evaluation of Candidate Telecommunication Icons*, Industrial Ergonomics Division and Engineering Psychology Division, Human Work Sciences Department, Luleå University of Technology-Sweden, ISSN: 1402-1757, ISRN:LTU-LIC, 98/12-SE, Luleå: Luleå University of Technology.

Zwaga, H. L. G. (1989), Predicting the public's comprehension of graphic symbols. In Megaw, E. D. (Ed.): *Contemporary Ergonomics*, pp. 431-436.




# PAPER 5

# AN EMPIRICAL EVALUATION OF VIDEOPHONE SYMBOLS: AN INTERNATIONAL STUDY

# AN EMPIRICAL EVALUATION OF VIDEOPHONE SYMBOLS: AN INTERNATIONAL STUDY

**Piamonte, D. P. T.[1], Ohlsson, K.[2], and Abeysekera, J. D. A.[1]**
[1]Division of Industrial Ergonomics, [2]Division of Engineering Psychology, Department of Human Work Sciences, Luleå University of Technology, Sweden

## Introduction

The widespread use of graphical symbols in the light of the current economic globalisation is undeniable. Symbols such as icons or pictograms as interfaces in modern technological devices have become very common. Although generally designed in western countries, their use can be found worldwide from computers to audio-visual appliances and other similar devices. Their use offers several potential advantages. They require minimum space, are easy and quick to recognise, and are non-text dependent. The use of graphical symbols, however, is not devoid of problems. Foremost among these is that some symbols are simply not understood well (Brelsford, Wogalter and Scoggins, 1994). The above advantages were cited as evident when the concepts being represented were well understood or concrete, and not when ideas become more abstract. Regarding older people, Davis, Haines, Norris and Wilson (1998) cited the studies of Easterby and Hakiel (1981) showing that although symbols are easily recognizable, understanding their meaning are generally poorer compared to younger people. One basic principle of effective UI design is to involve the potential users in the design and evaluation stages especially when the user groups are quite diverse. Thus, empirical tests are important to properly evaluate the symbols on their usability especially across different cultures.

This study was part of an international project on evaluating telecommunication symbols (Piamonte, Ohlsson and Abeysekera, 1999b). One of the major objectives was to test different graphical symbols (in this case, those of the videophone), designed and tested in Western Europe using subjects from different subject groups from Asia and the United States. The elderly as a special group was also included. The previous paper dealt with the study using subjects from five Southeast Asian countries (Piamonte, Abeysekera, and Ohlsson, 1999a). This paper would discuss the results involving U.S. and Finnish elderly subjects and comparing them with the Asian results as well.

## Methodology

The materials (symbols and questionnaires) used in the previous paper were also used in this study (Piamonte *et al*, 1999a and b). However, aside from the spontaneous identification and cued response tests, symbol and set preferences were also tested. Except for the spontaneous identification tests (Part One), the whole method was based on the multiple index approach (MIA) developed by Böcker (1993) with the European Telecommunications Standards Institute (ETSI).

In the spontaneous identification test (Part One), each page contained one set of the seven videophone symbols, for a total of three pages for the three symbol sets. On spaces provided for, they would write what videophone function they thought was represented by each symbol. They were also asked to rate the level of certainty for each of their answers using a seven-point rating scale (from "very certain" to "very uncertain"). In the cued response test (Part Two), the subjects first read a referent and its description. Then they had to choose one symbol from a set of seven symbols which they thought best represented the referent in question. Each page contained one referent-description and one set of symbols. They were also asked to rate their certainties for their answers using the 7-point rating scales. There were seven videophone referents tested on three sets of symbols rendering a total of twenty-one trials in Part Two. In the last part, the three symbols (one for each set) together with the referent they represented were shown. The subjects had to choose one symbol they thought best represent that referent. Next, with the three symbols sets presented together, they had to choose the set they prefer most. In all, the subjects would choose seven symbols to represent the seven referents and one symbol set they preferred most.

## Results and Discussion

For purposes of extended comparisons, subjects from the United States were included, using the same method in the Southeast Asian studies (Piamonte *et al*, 1999a). The nineteen subjects were students and employees in the University of Central Florida. The age range was 19-43 years with a mean of 22.6 years (SD=5.24). All of them have used computers and computer-related products at home and at work, with 78.9% (n = 15) using both graphics-based and DOS-based programs. Similar to the previous study, the US subjects performed poorly in Part One, even failing to identify the handsfree symbols in all 3 sets. The still picture symbols failed to score hits above 67% in both Part One and Part Two. In contrast, the symbols for camera, microphone and videophone had high recognition rates regardless of test type and set. Set 1 had the most number of symbols with recognition rates above 67% and Set 2 with the least. Certainty ratings generally were higher when the hits were higher as well.

Aside from correct identification and associations, other parameters were also measured. These were confusion, missing values, symbol and set preferences. Regarding preferences, the US subjects preferred five of seven of Set 1's symbols. These results were similar to the Southeast Asian study. However, they differ from the latter group by equally preferring Sets 1 and 3 as a group. One important factor that leads subjects or potential users to their choice of symbols is easy recognition. Set 3's hit rates were generally comparable to Set 1. Hence, it was not surprising for preference to be similar as well.

In some parts of Northern Finland, videophone-based services for the elderly are being tested. This was one of the major reasons why the current study was also tested among a group of elderly subjects. The respondents in the tests were twenty-nine (29) subjects from Finland (14 women and 15 men). The women were 61 to 81 years old (mean age: 69.6 years) and the men 60 to 88 years old (mean age: 70.9 years). The mean age for the whole group was 70.3 years. Compared to the subjects from Southeast Asia and even the USA, the Finnish elderly failed to do Part One. Majority of the subjects claimed that the tasks were either too difficult or too complicated. When compared to younger subjects (from Asia and the US), the elderly subjects had much lower hit rates in all symbols of the three sets used. Lastly, in contrast to the other groups also, the preference tests revealed the elderly favoring Set 3 to Sets 1 and 2. They commented that Set 3's symbols were more concrete and easily recognizable than the others.

The trend of demographic evolution shows that the European population, in particular, is one that is becoming old. This trend will have implications on the social structure, where respect for an independent living should be taken into consideration. It should taking into account the new services and products just available or those under development, such as telealarms and teleservices to assist people at home or elsewhere. Understanding the human factors of aging can lead to computer designs that will facilitate access by the elderly. Benefits include increased access of the society to the elderly for their experience, increased participation of the elderly in society through communication networks, and improved chances for productive employment of the elderly. (Shneiderman, 1992). In this case, the symbols were intended to improve the usability of a product (i.e. videophone). However, the results showed that the opposite might occur among the elderly. Only by a conscious and active participation of the elderly in designing and developing modern products will their needs be truly met.

As a whole, empirical tests using multiple indices are important to properly evaluate the symbols on their usability especially across different user groups. The evaluation method should also simulate actual usage scenarios to determine the suitability of each symbol alone, but together with the other symbols as well. This can be exemplified by the elderly and US subject groups who chose symbol sets quite different from the Asians. Hit rate is an important parameter, but must be tested under different user scenarios (spontaneous identification and cued responses) to determine if the symbols indeed are understood with and without cues. At the same time hits rates alone are not enough. Subjective certainties, false alarms (confusions), preferences, and even missing values are equally useful factors needed to make a deeper analysis. They enable the tester to see other often subtle but important differences (i.e. subjective biases, confusing symbols to other referents and symbols) on how users perceive and understand symbols.

*Acknowledgements*

- To Mr. Martin Böcker, of Siemen's-Nixdorf and ETSI Human Factors Group, for his generosity not only in sharing his resources on icons and pictograms, but more so for unselfishly sharing his insights and ideas about the study.
- To the National Communications Research Board of Sweden (*Kommunikations Forsknings Beredningen, KFB*) for the financial and logistical support

## References

Brelsford, J. W., Wogalter, M. S., and Scoggins, J. A., (1994). Enhancing comprehension and retention of safety-related pictorials, In *Proceedings of the Human Factors Ergonomics Society 38th Annual Meeting*, pp. 836-840.

Böcker, M. (1993). A multiple index approach for the evaluation of pictograms, In: *Proceedings of the 14th International Symposium on Human Factors in Telecommunications*, Darmstadt, Germany, May 11-14, Heidelberg: R. v. Decker's Verlag, G. Schenck GmbH, pp. 73-85.

Davies, S., Haines, H., Norris, B., and Wilson, J., (1998). Safety pictograms: are they getting the message across? *Applied Ergonomics*, 29, 1, 15-23.

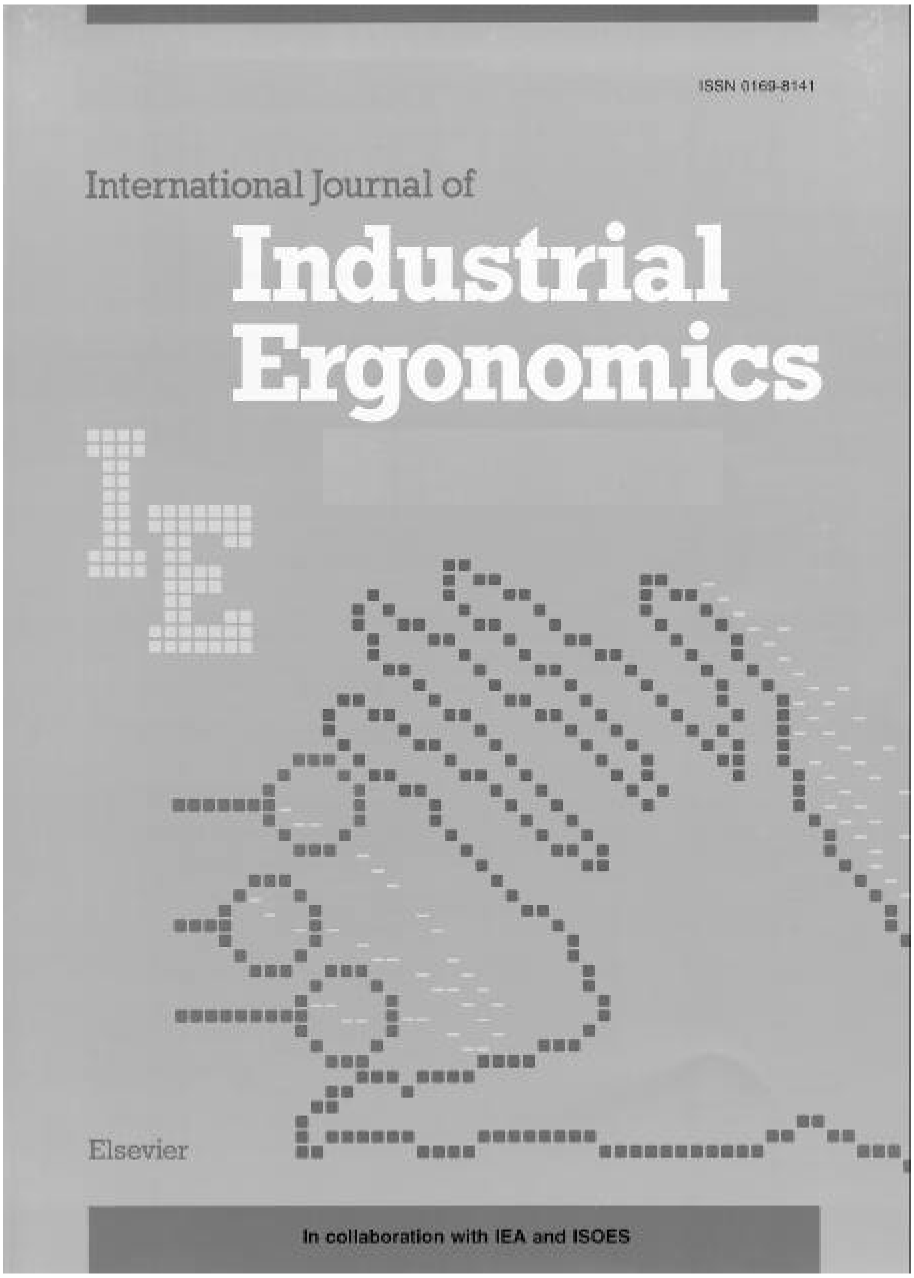
Easterby, R. S., and Hakiel, S. R., (1981). Field testing of consumer safety signs: the comprehension of pictorially presented messages. *Applied Ergonomics*, 12, 3, 143-152.

Horton, W. (1994). *The Icon Book – Visual symbols for computer systems and documentation*, New York: Wiley and Sons.

Piamonte, D. P. T., Abeysekera, J. D. A., and Ohlsson, K. (1999a). Testing videophone graphical symbols in Southeast Asia. In *Proc. 8th Int. Conference on Human-Computer Interaction* (HCI International '99, Munich, Germany, August 22-27, 1999).

Piamonte, D. P. T., Ohlsson, K., Abeysekera, J. D. A. (1999b). *Final Project Report: Across the seas-Auser - based evaluation of candidate telecommunication icons (Sponsored by the Communications Research Board of Sweden, KFB)*. Luleå: Luleå University Press (in press).

Shneiderman, B. (1992). *Designing the user interface – Strategies for effective human-computer interaction (2nd edition)*. New York: Addison-Wesley.




## PAPER 6

# Understanding Small Graphical Symbols: A Cross-Cultural Study

# Understanding Small Graphical Symbols: A Cross-Cultural Study

**D. Paul T. Piamonte[1], John D. A. Abeysekera[1] and Kjell Ohlsson[2]**
[1]Division of Industrial Ergonomics, Dept. of Human Work Sciences, Luleå University of Technology, 971 87 Luleå, Sweden
[2]Division of Human Work Sciences, Linköping University, Linköping, Sweden

## Abstract

Graphical symbols such as icons or pictograms as interfaces in modern technological devices have become quite common. Although generally designed in western countries, their use can be found worldwide from audio-visual appliances to computers and computer-related devices. A basic ergonomic tenet is to involve the potential users in the design and evaluation stages especially when the user groups are quite diverse. This study was part of an international project on evaluating telecommunication symbols. One of the major objectives was to test different graphical symbols (of the videophone), designed, and tested in Western Europe using subject groups from Asia, Europe, and the United States. This paper would discuss the major portions of the tests involving U.S. and Swedish subjects. Performance indices used were hits, certainties, confusions, and semantic differential ratings. They were useful in analysing how the symbols were recognised, confused, and perceived by different subject groups. They also helped detect differences between groups which otherwise seemed to have similar test results. Implications of the use of such multiple performance indices in symbol design and testing among diverse groups would also be discussed.

*Keywords:* Graphical symbols, icons, pictograms, cross-cultural studies, evaluation, semantic differential testing

## 1 Introduction

The ubiquitous nature and applicability of the current computer and information technology can be felt worldwide. Innumerable tools and equipment in schools, homes, and workplaces have become computer-based. From the lowly flat irons, ovens and telephones, to different audio-video appliances and high-tech desktop and notebook computers, all of these have become equipped with more complex functions. Consequently, one needs to learn how to manipulate graphic displays and buttons to avail of these added functions. In the midst of these complicated machines and the humans attempting to utilise are the graphical symbols. In general, a symbol is any graphical character or other representation. It is intended to (a) stand for something else, (b) communicate a use for an object/structure, or (c) communicate what should or should not be done at a given time or location (Stramler 1993). Graphical symbols, in turn, usually pertain to terms like *icons*, and *pictograms* or *pictorial symbols* (Böcker, 1993). The former refers to symbols that are simple, concrete and usually self-explanatory of the ideas, objects, or functions they represent (Wood and Wood 1987). The latter are usually more abstract, conveying messages by analogy or symbolism. Thus, pictograms require certain learning processes to be understood.

In addition, when one talks of computer systems and documentation, Horton (1994) described icons as the small pictorial symbols used on the computer menus, windows, and screens. They represent the capabilities of the computer system. They can be activated and bring forth these capabilities into action. By using powerful microprocessors or microchips, almost any device can now even be equipped with additional (and usually more complex) functions. Graphical symbols are thus ideal as interface. They use very little space, are non-text dependent but can convey lots of information (Maguire, 1985). They help the products go global. Thus, the rapid proliferation of complicated and multifunctional devices and their symbols is complemented by the cultural diversity of the intended users. The important question then, is whether symbols are really understood the way they should be across groups of varying cultures. Only during the last few years has research gained ground on this field of cultural differences in user perception (Resnick et al. 1997, Lin 1999). This study was part of an international project on evaluating telecommunication symbols. One of the major objectives was to test different graphical symbols designed and tested in Western Europe using subjects from different subject groups from Asia, Europe and the United States. This paper would report on the parts involving European (Swedish) and US (American) subjects.

## Method

Eighty-eight university students and employees of small and large companies from Sweden (48, i.e. 24 males and 24 females) and the United States (40, i.e. 21 males and 19 females) participated in the study. The test method used three sets of icons and pictograms representing seven referents or functions of the videotelephone (Figure 1). Both referents and pictograms were from studies of the European Telecommunications Standards Institute (ETSI) with one of the sets (Set 1) recommended by ETSI as the standard. The test was in the form of questionnaires. Each questionnaire had four parts: symbol identification tests, cued response tests, semantic differential ratings, and symbol and symbol set preferences.

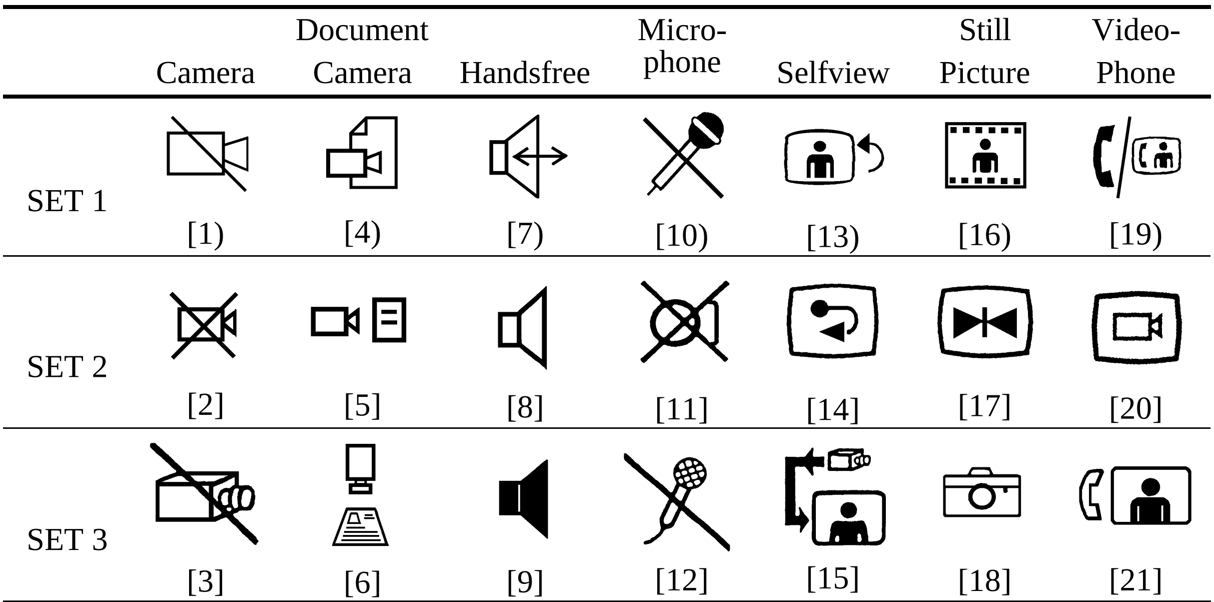


FIGURE 1. The seven videophone referents and the three sets of symbols used in the study.

The subjects were randomly given one of three versions of the questionnaires. An illustration of a videophone was shown and its general functions were then discussed. Afterwards, instructions were given on how to go about the different test parts. Questions were entertained prior to the conduction of the tests. Emphasis was given on avoiding omissions in trials in order to get back to them later. The order of the tests would also be strictly followed; i.e. Part One followed by Parts Two, Three, then Four. In the spontaneous identification test (Part One), each page contained one set of videophone symbols arranged in random order. A subject's task was to write on spaces provided what videophone function he/she thought each symbol represented. The same tasks were to be repeated for the two other symbol sets. In the cued response test Part Two), the subjects were to read a referent and its description. They then chose one symbol from a set of seven symbols they thought best represented the referent in question. Each page contained one referent description and one set of symbols. In both Parts One and Two, the subjects were also asked to rate their certainties for their answers using the 7-point rating scales. In the third part, five semantic scales were used to evaluate each symbol together with its correct referent, for a total of twenty-one SDTs. In Part Four, the three symbols (one for each set) together with the referent they represented were shown. The subjects had to choose one symbol they thought best represent that referent. Next, with the three symbols sets presented together, they had to choose the set they prefer most. In all, the subjects would choose seven symbols to represent the seven referents and one symbol set they preferred most.

## Results and discussion

Table 1 contains the means of the correct recognition scores (hits) and the means certainty ratings of the Swedish and US subjects for Parts One and Two using the three symbol sets. Regarding identification tests (Parts One and Two), Set 1 symbols generally had the higher hit rates in both countries. However, both countries performed poorly in the spontaneous identification tests (Part One). For example, subjects from both countries managed to attain the Organization for International Standardization's ISO 3864 (ISO, 1984) minimum correct recognition rate of 66.7% in only *5 of the 21 symbols* tested. Compared to some earlier studies using the same symbol sets, although the Swedish and American subjects fared quite better than their Asian counterparts (Piamonte 1998 and 1999) especially in Part One, the scores were still mostly below the critical 66.7%. Spontaneous identification tests simulate, albeit, in a limited manner, the 'first-time encounter' of an interface. The poor results thus strongly suggest that without sufficient learning opportunities and learning aids, symbols may be harder to understand by almost anyone than previously thought.

However, when trying to study possible cultural influences in symbol evaluation and understanding, levels of correct identification are not enough. Other parameters are essential (Nolan 1989, Lin 1999). This was shown by the results of the other parameters such as certainties (Table 1), confusions (Tables 2-4) and semantic differential scales (Table 5). Compared to the American subjects, Swedish subjects tended to give lower certainty ratings especially during the spontaneous identification tests. This can suggest certain subject bias. Confusion matrices showed some differences between the symbol sets and country groupings. Again, Set 1 had the least number of significant confusions in both countries. However in Set 1, Swedish subjects tend to have more problems confusing the handsfree function as represented by the symbols for videophone (9) and microphone (9), while Americans confused the former more with the microphone function (7) (Table 2). In addition, Swedish group showed symmetrical confusions between the selfview and still picture symbols of Set 1 (11, 13). These were not observed with the US subjects.

Confusions are very important in design considerations. They show which among the symbols need to be changed or re-designed or which among the functions are metaphorically related that training needs to be emphasised during the early part of using the symbols.

Semantic scales have been shown to be very useful in studying how symbols are perceived (Lin 1999). The scales used in this study were based mostly on the studies by Lin (1999) and Dewar (1977). In the semantic scales, again, the Swedish group as a whole gave lower mean semantic ratings compared to the American group. Although, there is high correlation between the hit rates and all the semantic scales, both the Swedish and US subjects had highest ratings for the symbols' meaningfulness and simplicity, with the latter group putting also prime importance to concreteness.

As mentioned above, the differences in patterns of ratings suggest that some subject bias occur between the two groups, which may be culturally-linked and can help in determining which aspects of symbol design and usage may be more helpful (ex. instructions, learning aids, etc.). Awareness of such subject bias and their implications are important on how one interprets the test results. For example, the Swedish subjects tended to be more conservative on their ratings compared to the Americans. However, this does not necessarily mean a lower level of correct symbol perception (as the hits rate results showed). Admittedly though, the above findings need to be analysed more deeply to be able to concretely transform such differences into practical applications of symbol design and testing.

## Acknowledgements


- To Mr. Martin Böcker for his invaluable support in acquiring the videophone symbols used in this study and in sharing his thoughts and expertise throughout.
- To Prof. Michael Kaplan of Central University of Florida, for sharing his precious time doing parts of the tests in USA.
- To *KFB* (*KommunikationsForskningsBeredningen* - Swedish Communications Board) for the financial and other logistical assistance.


## References


Böcker, M., 1993. A multiple index approach for the evaluation of pictograms. Proceedings of the 14th International Symposium on Human Factors in Telecommunications. R. v. Decker's Verlag, G. Schenck GmbH, Heidelberg, pp. 73-85.

Böcker, M., 1997. Symbol design. In A Focus Seminar in Symbol Design and Evaluation: Forum for Corporate Usability within Ericsson. Stockholm, Sweden.

Dewar, R.E. and Ells, J.G., 1977. The semantic differential as an index of traffic sign perception and comprehension. Human Factors, 19 (2): 183-189.

Horton, W., 1994. The Icon Book - Visual Symbols for Computer Systems and Documentation. John Wiley and Sons, New York, 417 pp.

International Standards Organization (ISO), 1984. International standard for safety colours and safety signs: ISO 3864. Geneva, Switzerland.

Lin, R., 1999. Cultural differences in icon recognition. In: G. Salvendy, M. Smith and R. Koubek (Eds.), Design of Computing Systems Volume I; Proceedings of the 8th International Conference on Human-Computer Interaction,. Amsterdam: Elsevier, pp. 725-729.

Maguire, M.C., 1985. A review of human factors guidelines and techniques for the design of graphical human-computer interfaces. Computers and Graphics, 9 (3): 221-235.

Nolan, P.R., 1989. Designing screen icons: ranking and matching studies. In Proceedings of the Human Factors Society 33rd Annual Meeting, Human Factors, Santa Monica, pp. 380-384.

Piamonte, D.P.T., 1998. A User-based Evaluation of Candidate Telecommunication Icons, Licentiate Thesis. ISSN: 1402-1757, ISRN: LTU – LIC -- 98/12 -- SE Divisions of Industrial Ergonomics and Engineering Psychology, Department of Human Work Sciences, Luleå University of Technology, Luleå, 90 pp.

Piamonte, D.P.T., 2000. The relevance of error analysis in graphical symbols evaluation. International Journal of Occupational Safety and Ergonomics (in press).

Resnick, M.L., Zanotti, A. and Jacko, A., 1997. Cultural differences in the perception of responsibility for child safety. In Proceedings of the Human Factors Society 39th Annual Meeting, Human Factors, Sta. Monica, pp. 1015-1019.

Stramler, J.H. Jr., 1993. The Dictionary for Human Factors/Ergonomics. CRC Press, London, 413 pp.

Wood, W.T. and Wood, S.K., 1987. Icons in everyday life. Social Ergonomics and Stress Aspects of Work with Computers, 97-104.

TABLE 1. Means correct recognition scores (hits) and certainty ratings for Parts One and Two, Swedish and US subjects (Sweden, n=48, USA, n=40).

| | PART ONE (Spontaneous Identification) | | | | PART TWO (Cued-Response) | | | |
|---|---|---|---|---|---|---|---|---|
| | SWEDEN | | USA | | SWEDEN | | USA | |
| | Hits | Cert. | Hits | Cert. | Hits | Cert. | Hits | Cert. |
| [1] Camera 1 | 87.5 | 5.2 | 75.0 | 5.7 | 75.0 | 4.9 | 80.0 | 3.1 |
| [2] Camera 2 | 87.5 | 4.2 | 80.0 | 5.1 | 70.8 | 4.3 | 70.0 | 3.6 |
| [3] Camera 3 | 75.0 | 4.6 | 80.0 | 5.5 | 66.7 | 5.2 | 60.0 | 4.0 |
| [4] Doc. Camera 1 | 58.3 | 3.5 | 25.0 | 4.2 | 95.8 | 5.3 | 90.0 | 3.3 |
| [5] Doc. Camera 2 | 50.0 | 4.0 | 10.0 | 4.3 | 91.7 | 4.9 | 90.0 | 3.8 |
| [6] Doc. Camera 3 | 41.7 | 4.0 | 20.0 | 4.4 | 75.0 | 4.4 | 80.0 | 4.2 |
| [7] Handsfree 1 | 12.5 | 4.8 | 5.0 | 4.3 | 50.0 | 4.4 | 50.0 | 3.4 |
| [8] Handsfree 2 | 12.5 | 5.5 | 0.0 | 5.3 | 66.7 | 3.5 | 40.0 | 3.4 |
| [9] Handsfree 3 | 8.3 | 5.3 | 0.0 | 5.8 | 58.3 | 4.5 | 35.0 | 3.5 |
| [10] Microphone 1 | 70.8 | 5.4 | 90.0 | 5.4 | 75.0 | 5.2 | 95.0 | 5.7 |
| [11] Microphone 2 | 12.5 | 2.9 | 10.0 | 3.5 | 33.3 | 3.9 | 55.0 | 4.5 |
| [12] Microphone 3 | 70.8 | 5.0 | 90.0 | 5.7 | 87.5 | 5.7 | 90.0 | 6.1 |
| [13] Selfview 1 | 4.2 | 2.2 | 10.0 | 4.5 | 70.8 | 4.4 | 95.0 | 3.3 |
| [14] Selfview 2 | 0.0 | 2.9 | 0.0 | 3.8 | 45.8 | 2.7 | 50.0 | 3.4 |
| [15] Selfview 3 | 12.5 | 3.7 | 10.0 | 4.7 | 91.7 | 4.6 | 80.0 | 4.2 |
| [16] Still Picture 1 | 12.5 | 3.0 | 10.0 | 5.1 | 62.5 | 4.6 | 65.0 | 3.0 |
| [17] Still Picture 2 | 4.2 | 2.8 | 10.0 | 4.6 | 66.7 | 3.1 | 50.0 | 3.2 |
| [18] Still Picture 3 | 16.7 | 4.8 | 20.0 | 5.6 | 70.8 | 5.3 | 40.0 | 4.0 |
| [19] Videophone 1 | 54.2 | 4.3 | 70.0 | 4.9 | 95.8 | 6.0 | 95.0 | 6.3 |
| [20] Videophone 2 | 0.0 | 3.8 | 0.0 | 4.9 | 20.8 | 2.5 | 30.0 | 4.2 |
| [21] Videophone 3 | 58.3 | 4.3 | 55.0 | 5.7 | 87.5 | 4.9 | 85.0 | 5.7 |
| *Overall Means* | *35.7* | *4.1* | *31.9* | *4.9* | *69.4* | *4.5* | *67.9* | *4.1* |

TABLE 2. Part Two confusion matrices for Set 1, per country.

| | SWEDEN | | | | | | | | USA | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SET 1 | HF | SE | SP | CA | VP | DC | MP | SET 1 | HF | SE | SP | CA | VP | DC | MP |
| Handsfree | - | | | | | | 7 | Handsfree | - | | | 3 | | | 2 |
| Selfview | 3 | - | 13 | | | | 1 | Selfview | 3 | - | 4 | 2 | | 2 | |
| Still Picture | | 11 | - | 3 | | | | Still Picture | | 1 | - | | | | |
| Camera | | | | - | 2 | 2 | | Camera | 3 | | 3 | - | 2 | 2 | |
| Videophone | 9 | | 5 | 6 | - | | 4 | Videophone | 3 | 1 | | 3 | - | | |
| Doc. Camera | 3 | 3 | | 3 | | - | | Doc. Camera | 2 | | 7 | | | - | |
| Microphone | 9 | | | | | | - | Microphone | 7 | | | | | | - |
| *Missing* | | | | | | | | *Missing* | *2* | | | | | | |

TABLE 3. Part Two confusion matrices for Set 2, per country.

| | SWEDEN | | | | | | | | USA | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SET 2 | HF | SE | SP | CA | VP | DC | MP | SET 2 | HF | SE | SP | CA | VP | DC | MP |
| Handsfree | - | | | | | | 17 | Handsfree | - | 3 | | | 2 | | 6 |
| Selfview | 2 | - | 4 | | 2 | 2 | 8 | Selfview | | - | | | 7 | 2 | 2 |
| Still Picture | 2 | 4 | - | 3 | 8 | | 5 | Still Picture | 8 | 3 | - | | 4 | | 8 |
| Camera | | 4 | | - | 8 | 2 | | Camera | 4 | 4 | | - | 2 | | |
| Videophone | 2 | 13 | 6 | 10 | - | | | Videophone | 4 | 7 | 10 | 10 | - | 2 | |
| Doc. Camera | | | 6 | 1 | 8 | - | | Doc. Camera | | 3 | 8 | | 9 | - | |
| Microphone | 8 | 3 | | | 8 | | - | Microphone | 8 | | | 2 | 2 | | - |
| *Missing* | *2* | *2* | | | *4* | | *2* | *Missing* | | | *2* | | *2* | | *2* |

TABLE 4. Part Two confusion matrices for Set 3, per country.

| | SWEDEN | | | | | | | | USA | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SET 3 | HF | SE | SP | CA | VP | DC | MP | SET 3 | HF | SE | SP | CA | VP | DC | MP |
| Handsfree | - | | | 2 | | | 6 | Handsfree | - | | 2 | | | | 4 |
| Selfview | | - | 10 | 4 | 2 | | | Selfview | 2 | - | 11 | 2 | 5 | 2 | |
| Still Picture | | | - | 10 | | 2 | | Still Picture | | | - | 10 | | 2 | |
| Camera | 6 | | | - | 2 | 8 | | Camera | 5 | 2 | | - | 1 | 4 | |
| Videophone | 2 | 4 | | | - | | | Videophone | 6 | 6 | 2 | 4 | - | | |
| Doc. Camera | 2 | | 4 | | 2 | - | | Doc. Camera | | | 9 | | | - | |
| Microphone | 8 | | | | | | - | Microphone | 13 | | | | | | - |
| *Missing* | *2* | | | | | *2* | | *Missing* | | | | | | | |

TABLE 5. Mean semantic scale ratings of the 21 graphical symbols, per country (SWE - Sweden, n=48, USA, n=40)

| | Meaningful | | Concrete | | Familiar | | Simple | | Clear | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SWE | USA | SWE | USA | SWE | USA | SWE | USA | SWE | USA |
| [1] Camera 1 | 6.3 | 6.2 | 6.3 | 6.0 | 6.3 | 6.2 | 6.5 | 6.0 | 6.3 | 5.9 |
| [2] Camera 2 | 6.0 | 5.9 | 6.2 | 5.7 | 5.9 | 5.9 | 6.4 | 6.3 | 6.2 | 5.9 |
| [3] Camera 3 | 6.0 | 6.5 | 5.9 | 6.4 | 5.6 | 6.6 | 6.2 | 6.4 | 6.0 | 6.1 |
| [4] Doc. Camera 1 | 5.3 | 6.0 | 5.3 | 6.1 | 5.2 | 5.9 | 5.6 | 6.1 | 5.2 | 6.2 |
| [5] Doc. Camera 2 | 5.7 | 6.2 | 5.6 | 6.3 | 5.1 | 5.9 | 5.7 | 6.2 | 5.7 | 6.0 |
| [6] Doc. Camera 3 | 4.0 | 5.9 | 3.8 | 5.7 | 3.7 | 6.0 | 3.5 | 5.8 | 3.4 | 5.8 |
| [7] Handsfree 1 | 3.8 | 3.7 | 3.5 | 3.5 | 3.3 | 3.7 | 3.9 | 3.8 | 3.0 | 3.4 |
| [8] Handsfree 2 | 4.4 | 4.1 | 4.0 | 3.7 | 3.9 | 3.6 | 4.3 | 4.2 | 3.5 | 3.2 |
| [9] Handsfree 3 | 4.6 | 4.5 | 4.0 | 4.2 | 4.3 | 3.8 | 4.5 | 4.2 | 4.0 | 3.7 |
| [10] Microphone 1 | 5.8 | 6.5 | 6.0 | 6.3 | 5.9 | 6.4 | 6.2 | 6.4 | 6.1 | 6.4 |
| [11] Microphone 2 | 2.0 | 2.0 | 1.9 | 2.4 | 1.8 | 2.0 | 2.3 | 2.3 | 1.8 | 2.0 |
| [12] Microphone 3 | 6.0 | 6.6 | 6.0 | 6.5 | 5.9 | 6.6 | 5.9 | 6.5 | 6.0 | 6.4 |
| [13] Selfview 1 | 5.3 | 5.7 | 4.9 | 5.5 | 4.5 | 5.3 | 5.2 | 5.7 | 4.9 | 5.7 |
| [14] Selfview 2 | 2.8 | 3.3 | 2.0 | 3.3 | 1.8 | 3.4 | 2.3 | 3.4 | 2.1 | 3.1 |
| [15] Selfview 3 | 5.6 | 6.2 | 5.5 | 6.3 | 5.0 | 5.9 | 5.3 | 5.9 | 5.3 | 6.2 |
| [16] Still Picture 1 | 5.0 | 6.2 | 4.4 | 6.1 | 4.9 | 5.5 | 4.6 | 5.8 | 4.5 | 5.3 |
| [17] Still Picture 2 | 3.8 | 4.3 | 3.0 | 4.3 | 3.3 | 4.5 | 3.3 | 4.4 | 2.8 | 4.4 |
| [18] Still Picture 3 | 5.4 | 6.1 | 4.9 | 5.7 | 5.2 | 6.1 | 5.0 | 5.5 | 4.8 | 5.3 |
| [19] Videophone 1 | 6.5 | 6.3 | 6.5 | 6.6 | 6.4 | 6.7 | 6.3 | 6.7 | 6.5 | 6.8 |
| [20] Videophone 2 | 3.3 | 4.6 | 3.3 | 4.4 | 3.2 | 4.0 | 4.2 | 4.4 | 2.8 | 3.9 |
| [21] Videophone 3 | 6.4 | 6.4 | 5.8 | 6.5 | 5.9 | 6.5 | 6.1 | 6.2 | 6.0 | 6.4 |
| *Overall Means* | *5.0* | *5.4* | *4.7* | *5.3* | *4.6* | *5.2* | *4.9* | *5.3* | *4.6* | *5.1* |

## PAPER 7

# ON THE MERITS OF USING MULTIPLE INDICES IN EVALUATING SMALL GRAPHICAL SYMBOLS

# ON THE MERITS OF USING MULTIPLE INDICES
## IN EVALUATING SMALL GRAPHICAL SYMBOLS


Dominic Paul T. Piamonte
Division of Industrial Ergonomics
Lulea University of Technology
Sweden

Kjell Ohlsson
Division of Industrial Ergonomics
Linköping University
Sweden

John Abeysekera
Division of Industrial Ergonomics
Lulea University of Technology
Sweden



## ABSTRACT

This paper summarizes an international study that aimed to evaluate candidate telecommunication graphical symbols (icons and pictograms) developed in the west across different cultural groups by means of tests producing multiple indices or parameters of performance. Prospective users from eastern (Asian) and western countries were used as subjects. Several tests were performed utilizing videophone symbols based on studies done by the Human Factors Technical Committee (HFTC) of the European Telecommunication Standards Institute (ETSI). The ETSI-recommended symbols for 7 videophone functions or referents were tested using more than 300 subjects from Indonesia, Malaysia, Philippines, Thailand, Sri Lanka, Finland, Sweden and the USA. Two other sets of 7 symbols each from the ETSI study were also tested. The tests used were spontaneous identification, the cued response, and the preference tests. Confidence judgement (subjective certainty ratings) complemented the subjects' answers for the first tests. Semantic differential scaling tests (SDT) were also done as added evaluatory tool. Results from spontaneous identification tests revealed very poor identification of most of the symbols in contrast to the cued response test results. Barely recognizing what the symbols meant strongly suggested the need to either redesign the symbols or to ensure adequate opportunities for familiarizing and educating the prospective users with the new symbols. The subjective certainty scores helped in studying the level of confidence of the answers by the subjects. Furthermore, the studies revealed that symbols could be easily recognized (high hit rates) but also confused as representing another (wrong) function at the same time. The "missing values" were also important since they indicated situations when respondents either did not know the answer or thought that none among the symbols were comprehensible or representative of a desired function. The preference tests pertained to aesthetics of the symbols individually and as a set. In turn, the SDT scores revealed that symbols could have different connotative meanings in relation to the functions they were intended to represent. Overall, Asian subjects performed comparably well with the European and American subjects, preferring the same set of videophone symbols, but usually at the expense of more errors and confusions. Thus, the studies showed that using multiple indices helped reveal subtle but potentially important differences in the results between different cultural groups.


## INTRODUCTION

Several reasons have been given why icons and pictograms have come to be increasingly used again. Most of these point to their advantages over text-based messages themselves. Compared to text, symbols are more distinct, faster and easier to recognise (Collins and Lerner, 1982), and can even reduce the likelihood of errors (Lodding, 1982). Horton (1994) pointed out those graphical symbols such as icons used in computers and computer-related products make these products go global.

However, owing to the complexity of how symbols are understood and learned, no single testing method is sufficient when determining which symbols to use. Although the use of single test methods has the advantage of speed, simplicity and ease in interpreting results, multiple tests parameters are more suitable when the candidate symbols are intended as parts of interfaces of devices for international use or for standardisation. More than one measure needs to be used and the relevance of each specific criterion depends on the purpose of the symbol and where it will be used (Dewar, 1999). Furthermore, Böcker (1993) contended that the most realistic evaluation method is one that tries to represent actual usage scenarios. For example, take the case of using videophone symbols. A prospective user is faced with the task of choosing which among the simultaneously visible symbols will bring about a specific function being sought (ex. switching on the camera function of the videophone). The possible outcomes of either his/her action will be a hit (choosing the right symbol), or a miss (wrong symbol). Other non-hit possibilities are false alarms and even failing to choose any symbol at all - no choice or missing values). This study aims to evaluate videophone symbols using multiple performance parameters and discuss the merits of using such an approach.

## METHODOLOGY

Three hundred fifty-seven subjects from Southeast Asia, Finland, Sweden and USA were involved in the study (187 males and 170 females, Papers 2 to 7). The overall age range was 17 - 52 years old, with a mean of 25 (S.D. = 6.8). Two hundred forty of them were from five Asian countries (forty-eight subjects per country). These countries were Indonesia, Malaysia, Philippines, Thailand and Sri Lanka. They were students, employees and professionals (doctors, lawyers and engineers), all of whom were computer-literate or had at least experienced using computers or computer-related products for the past 2-5 years at the time of the study. At the time of the study, none of the subjects had experienced using a videophone. Three sets of icons and pictograms representing seven referents or functions of the videotelephone for a total of twenty-one symbols (Figure 1) were used. Both referents and pictograms were based on the ETSI study (Böcker, 1993). One of these sets (Set 1) was the ETSI-recommended standard videophone symbols for the seven referents. The test was in the form of questionnaires. Each questionnaire had four parts: symbol identification-confidence judgement tests, cued response-confidence judgement tests, semantic differential ratings, symbol and symbol set preferences. The subjects were randomly given one of three versions of the questionnaires. An illustration of a videophone was shown and its general functions were then discussed. Afterwards, instructions were given on how to go about the different test parts. Questions were entertained prior to the conduction of the tests. Emphasis was given on avoiding skips in trials in order to get back to them later. The tests used were spontaneous identification, the cued response, and symbol and set preference tests. Confidence judgement (subjective certainty ratings) complemented the subjects' answers for the first tests. Semantic differential scaling tests (SDT) were also done as added evaluatory tool.

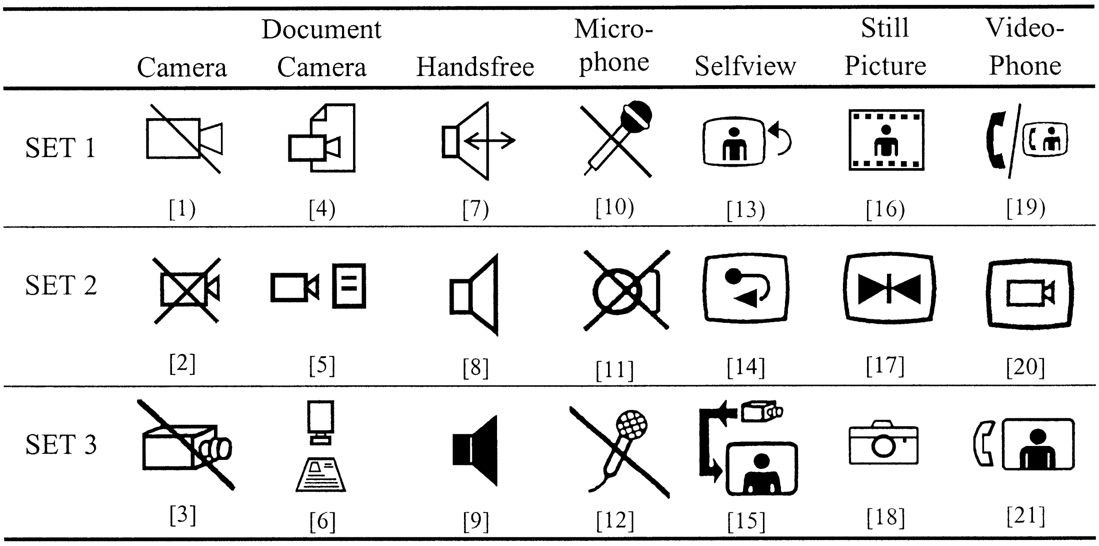



Figure 1. Graphical symbols used in the tests based on Böcker (1993) for the European Telecommunications Standards Institute (ETSI, 1993).

## RESULTS AND DISCUSSION

The results of the tests in the different countries showed that the spontaneous identification test (Part One) had significantly lower hit rates compared with the hit rates in the cued response test (Part Two). Part One and together with Part Two simulate (though in a limited manner) real videophone call scenarios wherein the symbols are placed on the terminal with one function to be performed at a time. Part One is a situation akin to using the technology the first time with minimal knowledge and instructions. Thus, it was not surprising that only familiar symbols or those that are concrete representations of the functions would have higher hit rates (ex. microphone, and camera), confirming one of the main hypotheses of this study. It was interesting to note that symbols could indeed be quite difficult to understand or comprehend based on the very low hit rates in Part One. Compared to the Southeast Asian countries, although the Swedish and American subjects fared quite better than their Asian counterparts especially in Part One, the scores were still mostly below the critical 67% of ISO. As discussed earlier, the spontaneous identification tests simulate, albeit, in a limited manner, the ‘first-time encounter’ of an interface. The poor results thus strongly suggest that without sufficient learning opportunities and learning aids, symbols may be harder to understand by even subjects or targeted users from western countries. Further, the spontaneous identification test itself was quite difficult for the elderly subjects with almost everyone failing to perform the test claiming it was too difficult or complicated.

The cued-response tests were similar to situations when using the videophone the first few times but with more information and instructions given. These findings give further credence to the importance of providing the prospective users all the possible cues and learning aids when attempting to use a new product or device the first few times (Lund, 1997, Shahnavaz, 1998 and Wolff and Wogalter, 1998). By using the cued-response tests several objective parameters were derived that helped reveal other possible differences between the candidate symbols and between subject groups. These can be summarised as follows:

Overall, the studies showed that hit rates or correct symbol-referent associations are still the most important objective parameter of symbol performance. In this aspect, Set 1’s symbols performed best. Six of its 7 symbols were correctly and easily associated to their intended functions or referents for all subject groups. However, hit rates are not the only important parameter, especially when evaluating several sets of symbols and using different subject groups. Non-hit parameters such as false alarms or confusions and missing values can be of great importance as well. For example, Set 3's symbols fared comparatively well in being associated to the intended functions, attaining hits as high as (if not higher than) Set 1. In these cases, the non-hit parameters such as the false alarms and missing values were very helpful in determining which symbol performed better among the subject groups.

False alarms or confusing the symbol as representing the wrong referent is a very good gauge of studying how distinct a symbol is compared to the other symbols presented at the same time. This parameter becomes more useful when used among subject groups with different cultural backgrounds (i.e. countries). The predominant types of confusions (symmetrical and asymmetrical) in each group could be used to study further how the symbols are understood and even confused by the subjects. These would then be helpful to determine if the problems would merit (a) re-design of the symbols; (b) replacement of the symbols; and/or (c) more explanation or emphasis on how the definition of the functions. In short, would the recommendations focus on symbol re-design or replacements, or will it be more learning opportunities or merely providing instructional aids to make sure that the device and its functions are well understood?

Missing values are also very important since they represent situations wherein the prospective user plainly lacks the knowledge of which among the symbols represent the desired function (referent). They can be akin to non-use or under-use of the device. If this would be one of the main problems, recommendations would be more on steps ensuring familiarity with the device and its functions.

In turn, the subjective parameters such as confidence judgements (subjective certainties), semantic scales and preferences were also helpful in gaining insight in subject group differences especially to certain possible subjective biases that may be culture-related. For example, between country groups, Swedish subjects as a group tended to be more conservative (i.e. lower mean ratings) in Part One compared to the six other country groups. Sri Lanka was the opposite. Although it had the lowest mean hit rates in the spontaneous identification and cued response tests, it exhibited higher subjective ratings than most of the other countries.

## CONCLUDING REMARKS

The methods utilised were largely based on the multiple index approach (MIA) developed by Böcker with the ETSI (Böcker, 1993) and with some modifications. The present research acknowledges the appropriateness of the MIA as the method when testing graphical symbols intended to appear together in an interface. Overall, Asian subjects performed comparably well with the European and American subjects, preferring the same set of videophone symbols, but usually at the expense of more errors and confusions. Using multiple indices helped reveal subtle but potentially important differences in the results between different cultural groups. This approach helped not only in identifying the best performing symbols, that is, which among the different symbols were understood correctly as they were intended to be. They also aided in studying how different subject groups (country groups) could differ in their understanding of symbols and how such differences could help in determining the interventions needed to make graphical symbols better interfaces.

## REFERENCES


Böcker, M. (1993). A multiple index approach for the evaluation of pictograms. In: *Proceedings of the 14th International Symposium on Human Factors in Telecommunications (pp. 73-85).* Darmstadt, Germany, May 11 - 14, Heidelberg: R. v. Decker's Verlag, G. Schenck GmbH.

Collins, B.L., & Lerner, N.D. (1982). Assessment of fire safety symbols. *Human Factors, 24(1),* 75-84.

Dewar, R. (1999). Design and evaluation of public information symbols. In H.J.G. Zwaga, T. Boersema, & H.C.M. Hoonhout (Eds.), *Visual Information for Everyday Use - Design and Research Perspectives (pp. 285-303).* London: Taylor and Francis.

Horton, W. (1994). *The Icon Book - Visual symbols for computer systems and documentation.* New York: Wiley and Sons.

Lodding, K.N. (1982). Iconic interfacing. *IEEE Computer Graphics and Applications, 3,* 11-20.

Mackett-Stout, J., & Dewar, R. (1981). Evaluation of symbolic public information signs. In *Proceedings of Human Factors Society Conference, Vol. 23, No. 2, (pp. 139-151).*

Mackett-Stout, J., & Dewar, R. (1981). Evaluation of symbolic public information signs. In *Proceedings of Human Factors Society Conference, Vol. 23, No. 2, (pp. 139-151).*

Zwaga, H.L.G. (1989). Predicting the public's comprehension of graphic symbols, In Megaw, E.D. (Ed.), *Contemporary Ergonomics 1989, (pp. 431-436).* London: Taylor and Francis.

# APPENDICES

APPENDIX 1. Mann-Whitney Test for Subjective Certainty Ratings results between countries, Philippine-Swedish Study), *significant difference at $p < 0.05$.

| Referents | Abstract Symbols | Concrete Symbols | Proposed Symbols |
|---|---|---|---|
| 1. Achieve Dial Tone | 1216.0 | 1210.5 | 1235.5 |
| 2. Answer Ringing Call | *738.0 | 1209.0 | 1204.5 |
| 3. Call Log | 1233.5 | 1093.5 | 1073.5 |
| 4. Conference | *918.0 | 1159.0 | 1214.0 |
| 5. Dialpad | *969.5 | *879.5 | *957.0 |
| 6. Drop | *971.0 | 1214.5 | 1091.5 |
| 7. Help Specific | 1026.0 | *584.5 | *819.0 |
| 8. Help System | *956.5 | 1169.5 | *830.5 |
| 9. HFAI | 1012.5 | *788.0 | 995.5 |
| 10. Hold | 1211.0 | *803.5 | *900.5 |
| 11. Message | 1131.5 | 1060.5 | 1145.5 |
| 12. Music On Hold | 1241.0 | *958.0 | 1072.5 |
| 13. Mute | 1182.5 | *805.0 | 1068.5 |
| 14. Notes | 1045.0 | *798.0 | 1213.0 |
| 15. Phone Call Active | 1129.0 | 1191.5 | *852.5 |
| 16. Retrieve Call | 994.0 | 1004.0 | 994.5 |
| 17. Ringer Select | 1232.0 | *943.0 | 1178.0 |
| 18. Speakerphone | *936.0 | 1238.0 | 1120.5 |
| 19. Speed Dial | *935.5 | 975.0 | 1110.5 |
| 20. Store | 1021.5 | *836.0 | 1116.0 |
| 21. Switch Hook Control | 1052.0 | *859.5 | 1124.0 |
| 22. Transfer Call | 1129.5 | *724.0 | *569.0 |
| 23. Volume | 1210.5 | *872.5 | 1207.0 |

APPENDIX 2. Mean Semantic Ratings of the videophone symbols in the Philippine (PHI) - Swedish (SWE) Study, n = 50 per country.

| | Abstract | | Concrete | | Proposed | |
|---|---|---|---|---|---|---|
| **REFERENTS** | PHI | SWE | PHI. | SWE | PHI. | SWE |
| **Achieve Dial Tone** | | | | | | |
| Concreteness | 3.7 | 2.3 | 4.2 | 3.6 | 3.5 | 3.4 |
| Familiarity | 4.0 | 2.1 | 4.1 | 3.4 | 3.9 | 2.9 |
| Meaningfulness | 4.1 | 2.6 | 4.5 | 4.1 | 3.8 | 3.1 |
| Relatedness | 4.4 | 2.4 | 4.4 | 3.6 | 4.0 | 3.1 |
| Sharpness | 3.9 | 2.7 | 4.2 | 3.8 | 3.7 | 2.9 |
| Simplicity | 3.9 | 2.5 | 4.5 | 3.7 | 3.7 | 3.2 |
| **Answer Ringing** | | | | | | |
| Concreteness | 3.1 | 2.6 | 4.9 | 5.3 | 5.6 | 4.8 |
| Familiarity | 3.4 | 2.4 | 5.4 | 4.9 | 5.2 | 4.3 |
| Meaningfulness | 4.1 | 2.9 | 5.6 | 5.4 | 5.6 | 4.3 |
| Relatedness | 4.1 | 2.8 | 5.7 | 5.7 | 5.8 | 4.8 |
| Sharpness | 3.5 | 2.7 | 5.1 | 4.4 | 5.4 | 3.8 |
| Simplicity | 3.4 | 2.9 | 5.3 | 5.5 | 5.6 | 4.8 |
| **Call Log** | | | | | | |
| Concreteness | 3.4 | 2.0 | 4.5 | 4.0 | 3.4 | 2.1 |
| Familiarity | 3.0 | 2.3 | 4.7 | 3.4 | 3.0 | 1.8 |
| Meaningfulness | 3.6 | 2.3 | 4.7 | 3.9 | 3.5 | 2.2 |
| Relatedness | 3.6 | 2.8 | 5.1 | 3.7 | 3.7 | 2.6 |
| Sharpness | 3.2 | 2.3 | 4.7 | 3.5 | 3.4 | 2.4 |
| Simplicity | 3.5 | 2.4 | 4.3 | 3.4 | 3.5 | 3.2 |
| **Conference** | | | | | | |
| Concreteness | 5.4 | 4.2 | 6.2 | 5.1 | 5.9 | 5.4 |
| Familiarity | 5.2 | 4.5 | 6.1 | 4.9 | 5.8 | 5.8 |
| Meaningfulness | 5.3 | 5.3 | 6.1 | 5.2 | 5.9 | 5.6 |
| Relatedness | 5.3 | 5.0 | 6.0 | 5.3 | 6.0 | 5.6 |
| Sharpness | 4.9 | 5.1 | 5.8 | 4.3 | 6.0 | 4.8 |
| Simplicity | 5.2 | 4.6 | 6.1 | 5.2 | 6.1 | 5.6 |

| | Abstract | | Concrete | | Proposed | |
|---|---|---|---|---|---|---|
| **REFERENTS** | PHI | SWE | PHI. | SWE | PHI | SWE |
| **Help Specific** | | | | | | |
| Concreteness | 1.9 | 2.9 | 5.7 | 4.1 | 4.7 | 5.8 |
| Familiarity | 2.2 | 3.4 | 5.5 | 4.3 | 5.2 | 6.4 |
| Meaningfulness | 2.3 | 4.0 | 5.8 | 4.7 | 5.2 | 6.4 |
| Relatedness | 2.2 | 4.1 | 5.6 | 4.6 | 5.1 | 6.4 |
| Sharpness | 2.2 | 3.7 | 5.4 | 4.4 | 4.9 | 5.5 |
| Simplicity | 2.1 | 3.7 | 5.7 | 4.0 | 5.2 | 6.6 |
| **Help System** | | | | | | |
| Concreteness | 1.7 | 2.1 | 2.6 | 3.1 | 3.5 | 5.7 |
| Familiarity | 2.0 | 2.3 | 2.7 | 3.0 | 4.1 | 5.5 |
| Meaningfulness | 2.7 | 3.2 | 3.0 | 3.4 | 4.2 | 5.8 |
| Relatedness | 2.8 | 2.5 | 3.3 | 3.6 | 4.0 | 6.1 |
| Sharpness | 2.3 | 3.1 | 2.7 | 3.3 | 4.2 | 5.3 |
| Simplicity | 1.8 | 2.2 | 2.6 | 3.0 | 4.3 | 6.0 |
| **Message** | | | | | | |
| Concreteness | 5.1 | 4.0 | 3.1 | 3.5 | 6.2 | 6.2 |
| Familiarity | 4.9 | 4.3 | 3.7 | 3.4 | 6.2 | 6.3 |
| Meaningfulness | 5.2 | 4.2 | 3.9 | 3.4 | 6.2 | 6.1 |
| Relatedness | 5.1 | 3.9 | 3.2 | 3.3 | 6.2 | 6.2 |
| Sharpness | 5.0 | 3.7 | 3.0 | 3.3 | 6.1 | 5.8 |
| Simplicity | 5.4 | 4.5 | 3.5 | 3.5 | 6.2 | 6.5 |
| **Music on Hold** | | | | | | |
| Concreteness | 5.2 | 3.1 | 6.0 | 5.3 | 6.0 | 4.1 |
| Familiarity | 5.2 | 3.0 | 5.8 | 5.0 | 6.5 | 4.3 |
| Meaningfulness | 5.4 | 3.5 | 6.0 | 5.4 | 6.3 | 4.9 |
| Relatedness | 5.3 | 3.4 | 6.1 | 5.6 | 6.4 | 5.0 |
| Sharpness | 5.0 | 3.2 | 5.7 | 5.1 | 5.9 | 3.9 |
| Simplicity | 5.2 | 2.6 | 6.2 | 5.1 | 6.3 | 5.1 |

APPENDIX 2 (continued). Mean Semantic Ratings of the videophone symbols in the Philippine (PHI) - Swedish (SWE) Study, n = 50 per country.

| | **Abstract** | | **Concrete** | | **Proposed** | |
|---|---|---|---|---|---|---|
| **REFERENTS** | PHI | SWE | PHI. | SWE | PHI. | SWE |
| **Dialpad** | | | | | | |
| Concreteness | 5.2 | 4.9 | 5.9 | 6.1 | 5.7 | 5.7 |
| Familiarity | 5.3 | 4.9 | 5.7 | 6.2 | 5.5 | 5.7 |
| Meaningfulness | 5.5 | 4.8 | 5.9 | 6.0 | 5.6 | 5.7 |
| Relatedness | 5.7 | 5.0 | 5.9 | 6.1 | 5.8 | 5.6 |
| Sharpness | 5.2 | 4.1 | 5.6 | 5.6 | 5.4 | 5.0 |
| Simplicity | 5.5 | 4.7 | 6.0 | 6.0 | 5.7 | 5.7 |
| **Drop Call** | | | | | | |
| Concreteness | 2.8 | 3.1 | 4.6 | 3.2 | 3.8 | 2.6 |
| Familiarity | 2.9 | 2.5 | 4.1 | 3.2 | 4.0 | 2.9 |
| Meaningfulness | 3.2 | 3.5 | 4.4 | 3.6 | 4.4 | 3.2 |
| Relatedness | 3.4 | 3.2 | 4.5 | 3.4 | 4.3 | 3.0 |
| Sharpness | 3.2 | 3.4 | 4.3 | 3.4 | 3.9 | 3.0 |
| Simplicity | 3.0 | 2.8 | 4.8 | 3.2 | 3.9 | 2.6 |
| **HFAI** | | | | | | |
| Concreteness | 3.0 | 3.0 | 3.5 | 2.2 | 3.3 | 3.7 |
| Familiarity | 3.3 | 3.0 | 3.3 | 1.7 | 3.0 | 3.4 |
| Meaningfulness | 3.9 | 4.0 | 3.6 | 2.5 | 3.7 | 2.9 |
| Relatedness | 3.8 | 3.7 | 3.9 | 2.4 | 3.9 | 3.3 |
| Sharpness | 3.5 | 3.7 | 3.5 | 2.2 | 3.5 | 2.4 |
| Simplicity | 3.4 | 3.5 | 3.8 | 2.3 | 3.7 | 3.6 |
| **Hold** | | | | | | |
| Concreteness | 2.9 | 2.3 | 5.2 | 3.5 | 5.4 | 4.0 |
| Familiarity | 3.0 | 2.5 | 5.1 | 3.5 | 5.1 | 4.3 |
| Meaningfulness | 3.4 | 2.4 | 5.4 | 4.1 | 5.6 | 4.5 |
| Relatedness | 3.4 | 2.5 | 5.7 | 3.9 | 5.6 | 4.8 |
| Sharpness | 3.2 | 2.5 | 5.2 | 3.1 | 5.1 | 3.5 |
| Simplicity | 3.1 | 2.6 | 5.3 | 3.7 | 5.5 | 4.5 |
| **Ringer Select** | | | | | | |
| Concreteness | 3.0 | 2.5 | 6.4 | 6.0 | 6.0 | 4.6 |
| Familiarity | 3.5 | 2.3 | 6.4 | 6.1 | 5.6 | 4.5 |
| Meaningfulness | 3.7 | 3.1 | 6.3 | 6.3 | 5.6 | 4.8 |
| Relatedness | 3.9 | 2.9 | 6.4 | 6.3 | 5.8 | 4.3 |
| Sharpness | 3.9 | 2.9 | 6.3 | 5.6 | 5.5 | 4.4 |
| Simplicity | 3.5 | 2.4 | 6.5 | 5.9 | 5.9 | 4.6 |
| **Speed Dial** | | | | | | |
| Concreteness | 4.5 | 3.2 | 5.1 | 3.1 | 4.2 | 3.1 |
| Familiarity | 4.2 | 3.1 | 4.9 | 3.1 | 4.2 | 3.8 |
| Meaningfulness | 4.9 | 3.8 | 5.1 | 4.1 | 4.7 | 3.7 |
| Relatedness | 5.0 | 3.9 | 5.1 | 4.1 | 4.6 | 3.9 |
| Sharpness | 4.5 | 3.8 | 5.0 | 3.3 | 4.7 | 3.5 |
| Simplicity | 4.5 | 3.0 | 4.9 | 3.7 | 4.5 | 3.3 |
| **Speakerphone** | | | | | | |
| Concreteness | 5.2 | 5.3 | 6.0 | 5.9 | 4.9 | 4.9 |
| Familiarity | 5.1 | 5.4 | 5.9 | 5.4 | 4.6 | 4.9 |
| Meaningfulness | 5.1 | 5.2 | 5.9 | 5.8 | 5.5 | 4.6 |
| Relatedness | 5.2 | 5.4 | 6.2 | 5.8 | 5.3 | 5.0 |
| Sharpness | 5.2 | 5.3 | 6.0 | 5.6 | 4.9 | 5.1 |
| Simplicity | 5.2 | 5.2 | 6.1 | 5.6 | 4.9 | 4.7 |
| **STORE** | | | | | | |
| Concreteness | 4.4 | 4.8 | 3.6 | 3.1 | 4.0 | 3.6 |
| Familiarity | 4.3 | 4.5 | 3.8 | 2.8 | 3.9 | 3.6 |
| Meaningfulness | 4.7 | 4.8 | 4.1 | 3.1 | 4.4 | 3.8 |
| Relatedness | 4.7 | 4.5 | 4.1 | 3.0 | 4.3 | 3.7 |
| Sharpness | 4.1 | 4.4 | 3.7 | 3.2 | 4.0 | 3.3 |
| Simplicity | 4.6 | 4.5 | 4.0 | 3.2 | 4.0 | 3.8 |

| | **Abstract** | | **Concrete** | | **Proposed** | |
|---|---|---|---|---|---|---|
| **REFERENTS** | PHI | SWE | PHI. | SWE | PHI | SWE |
| **Mute** | | | | | | |
| Concreteness | 5.5 | 5.6 | 4.9 | 4.8 | 4.5 | 4.7 |
| Familiarity | 5.2 | 5.6 | 5.3 | 4.7 | 4.5 | 4.6 |
| Meaningfulness | 5.6 | 5.7 | 5.5 | 4.8 | 4.6 | 4.2 |
| Relatedness | 5.6 | 5.5 | 5.5 | 4.7 | 4.6 | 4.5 |
| Sharpness | 5.4 | 5.5 | 5.1 | 4.8 | 4.7 | 4.5 |
| Simplicity | 5.4 | 5.7 | 5.4 | 4.6 | 4.5 | 4.9 |
| **Notes** | | | | | | |
| Concreteness | 5.5 | 4.1 | 6.3 | 5.6 | 6.1 | 5.6 |
| Familiarity | 5.5 | 4.4 | 6.0 | 5.3 | 6.0 | 5.3 |
| Meaningfulness | 5.5 | 4.6 | 6.2 | 5.6 | 6.1 | 5.6 |
| Relatedness | 5.8 | 4.6 | 6.2 | 5.5 | 6.1 | 5.7 |
| Sharpness | 5.5 | 4.4 | 5.8 | 5.3 | 5.9 | 5.0 |
| Simplicity | 5.9 | 4.5 | 6.2 | 4.9 | 6.2 | 5.6 |
| **Phone Call Active** | | | | | | |
| Concreteness | 2.6 | 2.7 | 4.2 | 3.3 | 4.9 | 4.0 |
| Familiarity | 2.9 | 2.3 | 4.1 | 3.2 | 4.7 | 3.9 |
| Meaningfulness | 3.0 | 3.8 | 4.1 | 4.1 | 5.1 | 4.0 |
| Relatedness | 3.2 | 3.7 | 4.4 | 4.0 | 5.5 | 4.2 |
| Sharpness | 2.8 | 3.6 | 4.2 | 3.2 | 4.7 | 4.4 |
| Simplicity | 3.6 | 4.4 | 4.2 | 3.6 | 5.2 | 4.6 |
| **Retrieve** | | | | | | |
| Concreteness | 4.1 | 4.0 | 4.0 | 3.0 | 2.5 | 1.9 |
| Familiarity | 4.2 | 3.6 | 4.4 | 2.8 | 2.5 | 1.6 |
| Meaningfulness | 4.7 | 4.6 | 4.3 | 3.2 | 3.3 | 1.9 |
| Relatedness | 4.4 | 4.4 | 4.0 | 3.0 | 2.9 | 2.2 |
| Sharpness | 4.0 | 4.3 | 4.1 | 2.8 | 2.8 | 2.2 |
| Simplicity | 4.1 | 4.2 | 4.2 | 2.6 | 2.9 | 2.3 |
| **Switch Hook** | | | | | | |
| Concreteness | 2.8 | 2.9 | 3.9 | 3.3 | 4.3 | 2.8 |
| Familiarity | 2.9 | 2.8 | 3.8 | 3.4 | 4.1 | 2.3 |
| Meaningfulness | 3.5 | 2.7 | 4.2 | 3.6 | 4.6 | 2.9 |
| Relatedness | 3.2 | 3.1 | 4.3 | 3.7 | 4.5 | 2.8 |
| Sharpness | 3.4 | 2.5 | 4.0 | 3.6 | 3.8 | 2.3 |
| Simplicity | 3.1 | 3.5 | 4.1 | 3.7 | 4.4 | 2.9 |
| **Transfer Call** | | | | | | |
| Concreteness | 3.3 | 3.0 | 5.7 | 5.2 | 6.0 | 5.1 |
| Familiarity | 3.2 | 2.5 | 5.9 | 4.3 | 5.7 | 4.5 |
| Meaningfulness | 3.5 | 3.3 | 5.8 | 4.6 | 6.1 | 4.9 |
| Relatedness | 3.8 | 3.3 | 6.1 | 4.9 | 6.3 | 5.0 |
| Sharpness | 3.6 | 2.9 | 5.7 | 4.6 | 5.7 | 4.1 |
| Simplicity | 3.3 | 2.9 | 6.1 | 4.9 | 6.2 | 5.0 |
| **Volume** | | | | | | |
| Concreteness | 5.8 | 5.9 | 5.5 | 6.4 | 4.9 | 6.4 |
| Familiarity | 6.4 | 6.0 | 5.6 | 6.4 | 5.9 | 6.3 |
| Meaningfulness | 6.3 | 6.2 | 5.7 | 6.5 | 5.4 | 6.0 |
| Relatedness | 6.0 | 6.1 | 5.7 | 6.7 | 5.5 | 6.2 |
| Sharpness | 6.2 | 5.5 | 5.3 | 6.2 | 5.0 | 5.9 |
| Simplicity | 6.5 | 6.1 | 5.6 | 6.5 | 5.6 | 6.1 |

APPENDIX 3. Figures 11 to 17 illustrating Percentage Distributions of the Semantic Differential Scores in the Main Study.

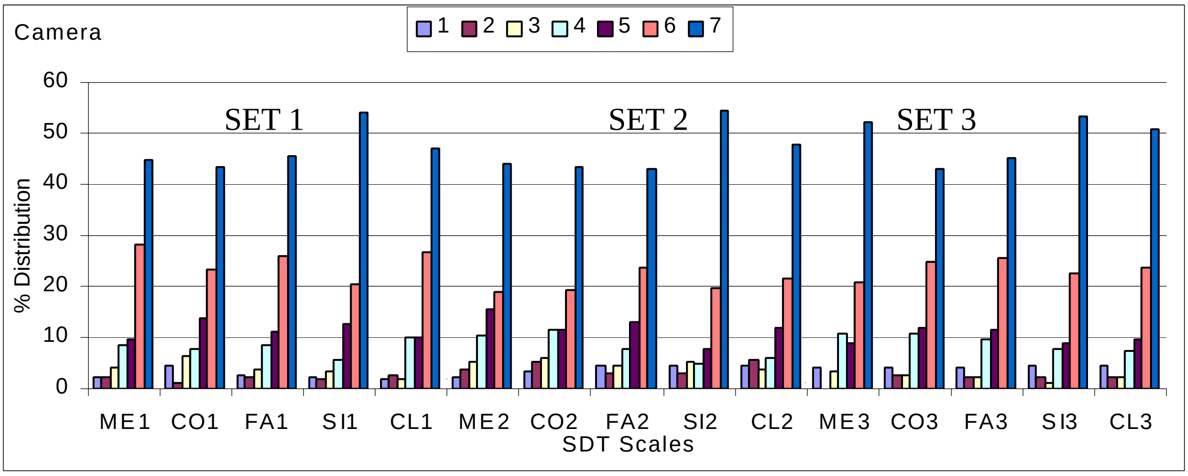



FIGURE 11. Percentage distributions of the semantic differential scores for camera (N = 328). Legend: ME=Meaningful, CO=Concreteness, FA=Familiarity, SI=Simplicity, CL=Clarity.

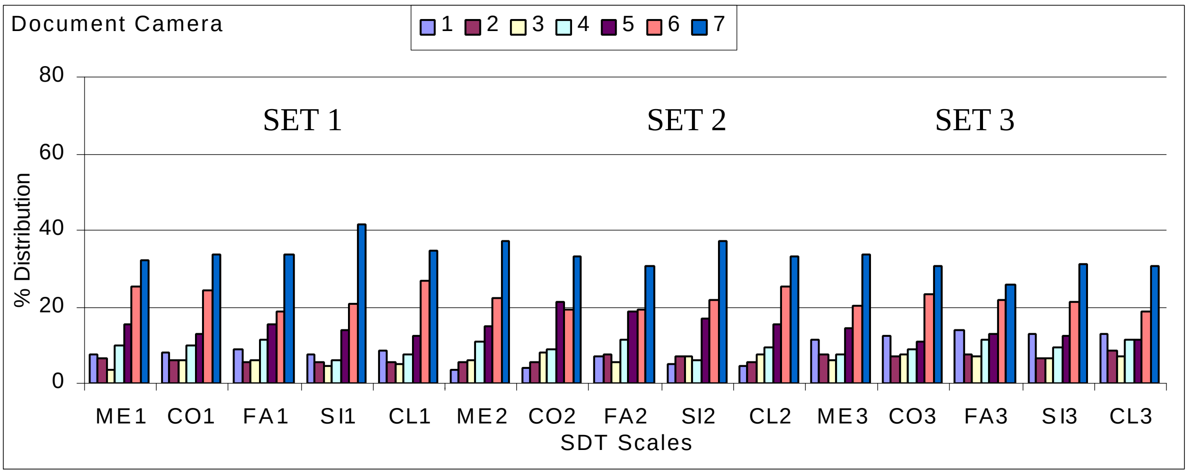



FIGURE 12. Percentage distributions of the semantic differential scores for document camera (N = 328). Legend: ME=Meaningful, CO=Concreteness, FA=Familiarity, SI=Simplicity, CL=Clarity.

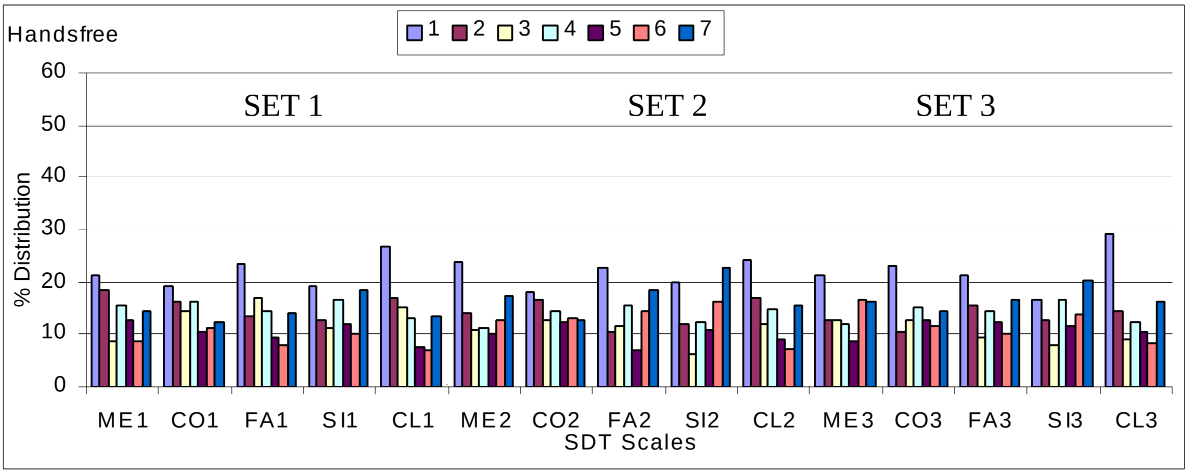



FIGURE 13. Percentage distributions of the semantic differential scores for handsfree (N = 328). Legend: ME=Meaningful, CO=Concreteness, FA=Familiarity, SI=Simplicity, CL=Clarity.

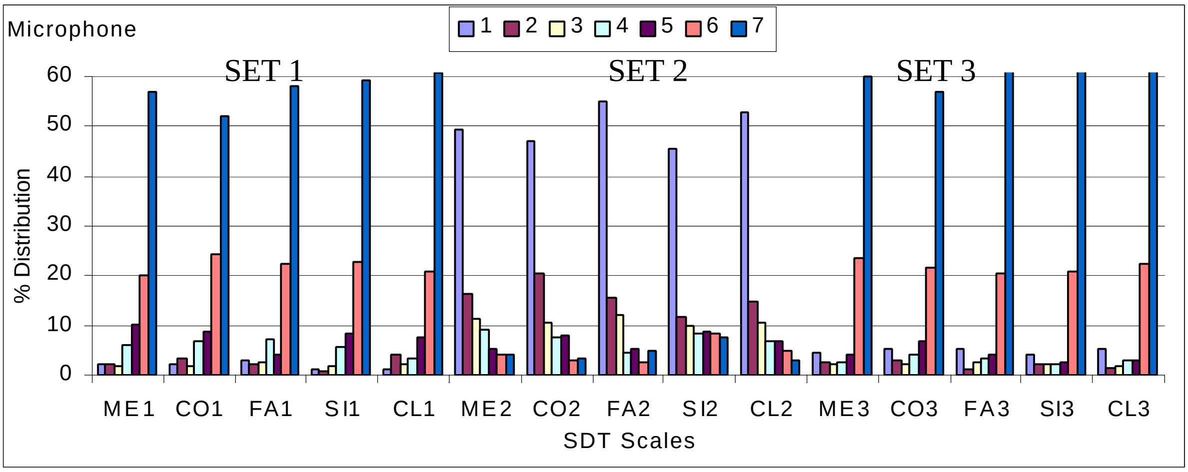



FIGURE 14. Percentage distributions of the semantic differential scores for microphone (N = 328). Legend: ME=Meaningful, CO=Concreteness, FA=Familiarity, SI=Simplicity, CL=Clarity.

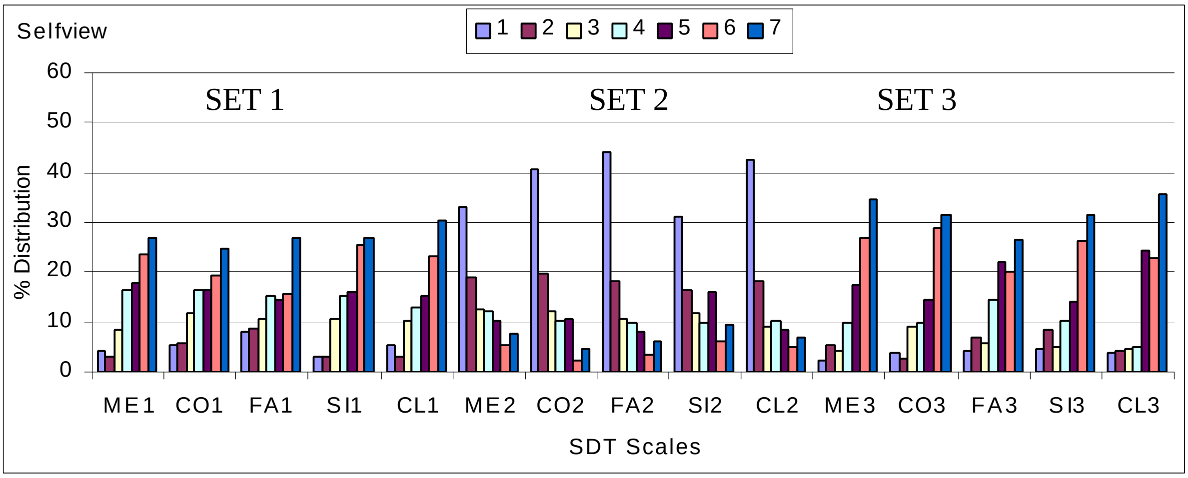



FIGURE 15. Percentage distributions of the semantic differential scores for selfview (N = 328). Legend: ME=Meaningful, CO=Concreteness, FA=Familiarity, SI=Simplicity, CL=Clarity.

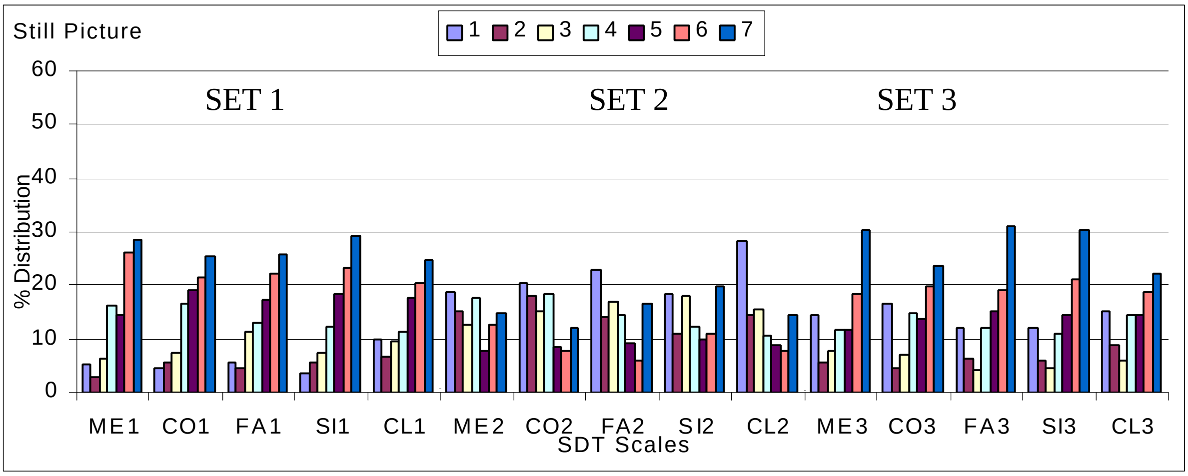



FIGURE 16. Percentage distributions of the semantic differential scores for still picture (N = 328). Legend: ME=Meaningful, CO=Concreteness, FA=Familiarity, SI=Simplicity, CL=Clarity.

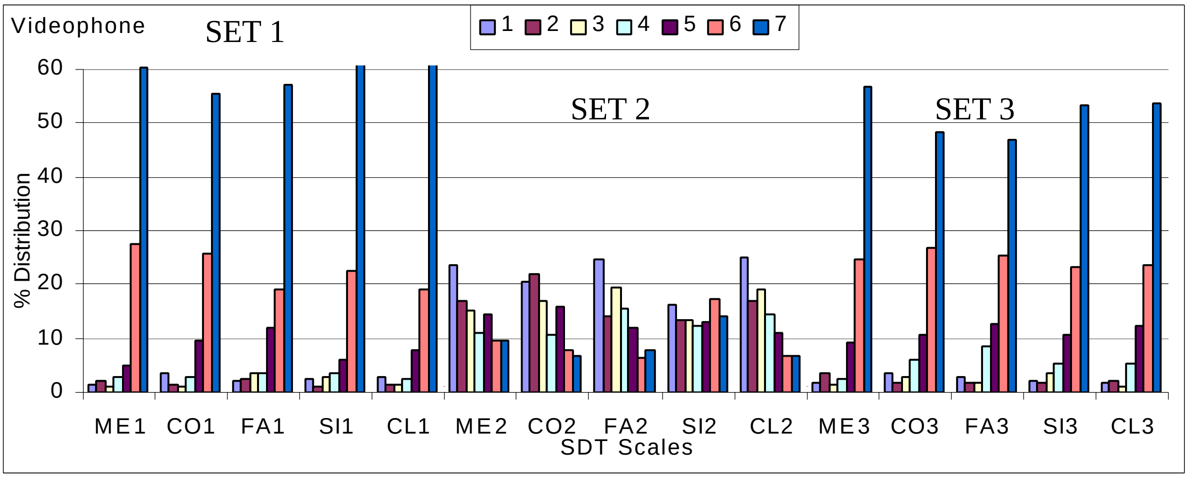



FIGURE 17. Percentage distributions of the semantic differential scores for videophone (N = 328). Legend: ME=Meaningful, CO=Concreteness, FA=Familiarity, SI=Simplicity, CL=Clarity.