A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI

Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum

Aalborg University, Laboratory of Computer Vision and Media Technology
Niels Jernes Vej 14, DK-9220 Aalborg, Denmark
{mnielsen,mst,tbm,eg}@cvmt.dk

Abstract. Many disciplines of multimedia and communication go towards ubiquitous computing and hand-free- and no-touch interaction with computers. Application domains in this direction involve virtual reality, augmented reality, wearable computing, and smart spaces, where gesturing is a possible method of interaction. This paper presents some important issues in choosing the set of gestures for the interface from a user-centred view such as the learning rate, ergonomics, and intuition. A procedure is proposed which includes those issues in the selection of gestures, and to test the resulting set of gestures. The procedure is tested and demonstrated on an example application with a small test group. The procedure is concluded to be useful for finding a basis for the choice of gestures. The importance of tailoring the gesture vocabulary for the user group was also shown.

1 Introduction

Much work has been conducted in the investigation and development of natural interaction interfaces, including gesture interfaces [1][2][3][4][5][6][7]. Science fiction literature and movies also dream up gesture interfaces, for example the movies Johnny Mnemonic (1995), Final Fantasy (2001), and Minority Report (2002). Furthermore, gesture interfaces are applied to solve problems with people with physical disabilities [8].

It is important to remember that a gesture interface is to be seen as complementing or an alternative to existing interface techniques, such as the old desktop paradigm. Other examples are the new alternatives to the mouse, such as ergonomic trackballs, mouse pens, and the iGesture Pad [9]. They can all navigate a Windows interface with the mouse cursor just as well or better than the mouse, while they may be more or less useless when it comes to fast computer games, such as 3D shooters and Airplane Simulators.

When developing a gesture interface, the objective should not be "to make a generic gesture interface". A gesture interface is not universally the best interface for any application. The objective is "to develop a more efficient interface" for a given application.

This can be illustrated by the following example. An interface is desired for artistic modelling of a sculpture. An artist is hired for the job. The artist may be given a mouse and a keyboard for a CAD program. The result is perfect to the smallest detail in regard to accuracy of the lines, because it is possible to set coordinates explicitly.

A. Camurri and G. Volpe (Eds.): GW 2003, LNAI 2915, pp. 409–420, 2004.

However, these tools may not be well suited to convey the creativity of the artist. If the artist is instead provided with a gesture interface in which a virtual kind of clay model can be altered by touching and squeezing it, it will not be accurate in terms of coordinates and straight lines, but it might be aesthetically closer to the artist's vision. Thus, the choice of interface is to be related to the desired outcome of the application.

Consequently, the first step is the analysis of which kind of interface is most suitable for this task. This might lead to the conclusion that a gesture interface is the type that should be developed.

The focus in this paper is the next step; to design this hand-gesture interface and to find the specific gestures that make this specific interface for this specific application most efficient.

The set of gestures in an interface is called the "gesture vocabulary". This should not be confused with a general non-verbal communication dictionary. There are several ways of labelling gestures. This paper uses the taxonomy presented by Justine Cassell [1].

Section 2 investigates the foundation of gestures and gives an understanding of the gesture domain. Section 3 proposes a procedure for finding and testing gestures for a gesture interface. Section 4 presents the executed test of the procedure. Section 5 concludes the work.

2 Choosing Gestures

This section will motivate the need for the proposed procedure. There are two approaches [22] of choosing the gestures; the technical based approach and the human based approach.

The technical approach of choosing gestures is to choose a set of gestures which is simple to recognise. It starts with an idea for a method that can recognise some certain gestures. Like an example at the authors' department, where the method counts the number of extended fingers. Once this gesture vocabulary has been established, the next step is to apply them to an application. The problem with this technical approach is that some gestures are stressing or impossible to perform to some people. Furthermore, the functionalities and the gestures are illogically mapped.

An important question to ask is if the cultural dependence is a problem. Conventional international interfaces are generally in English, but most software is available with selectable national language packages, and some nations have different keyboard layouts. In a gesture interface this can be translated into selectable gesture sets, if it should become a problem that an emblem [1] is illogic to another culture. Furthermore, if a culturally dependent gesture is used, this does not necessarily mean that it is utterly illogic for other cultures to learn them if presented with them.

2.1 Human Based Gesture Vocabulary

The human based gesture approach investigates the people who are going to use the interface. Wizard-of-Oz experiments have proven valuable in the development of gestures [10]. The experiments simulate the response of the system by having a person respond to the user commands. This approach tests a developed interface.

Gesture research [1] shows that there is no such thing as a universal gesture vocabulary, so a good gesture vocabulary may only match one specific application and user group. Developers need to make it easy for the users to use their systems. Otherwise, the users will choose the work of the competitors. The human based principles should make the gestures:

- Easy to perform and remember
- Intuitive
- Metaphorically and iconically logical towards functionality
- Ergonomic; not physically stressing when used often

This means that the gesture vocabulary must be tailored for the specific task. Of course, it would still be wise to keep in mind that they should be possible to recognize unambiguously by the system.

In order to achieve these principles it is necessary to take usability theory, and biomechanics/ergonomics. See [22] for further details. The work of Jakob Nielsen [23] on usability gives a good overview. Research on usability focuses on five main principles [11][12][13]:

Learnability	The time and effort required to reach a specific level of use performance.
Efficiency	Steady-state performance of expert users.
Memorability	Ease of system intermittently for casual users.
Errors	Error rate for minor and catastrophic errors.
Coverage	The amount of operators discovered vs. the total operators.

Furthermore, Nielsen suggests a set of heuristics [13] which facilitates these principles:

1. Use simple and natural dialogue
2. Speak the user's language
3. Minimize user memory load
4. Be consistent
5. Provide feedback
6. Provide clearly marked exits
7. Provide shortcuts
8. Provide good error messages
9. Prevent errors

Biomechanics and ergonomics [14][15][18][19][20] tell about constraints in postures and the usage of the gestures, such as avoiding to stay in static positions, and moving joints too far from their neutral positions. There are constraining relationships between individual joints within fingers and between neighbouring fingers. These constraints are different in people.

Wrist movement can have adverse effects on comfort as well as the external- and internal force. Experiments have shown that the entire body posture is important to be aware of, even when dealing with hand gestures. This is because it affects the elbow and wrist postures, which again affects the force needed to set the finger postures [17].

The ergonomics show that it is important to make the recognition algorithms tolerant to de-stressing movements, which allows the user to avoid staying fixed in e.g. a static “residue” or “pointing” gesture. Tolerance for deviations in gestures is desirable when implementing gesture interfaces, also because of varying hand shapes and posture performance.

3 Approach to Finding Gestures

Section 2 outlined the importance of choosing a logical and ergonomic gesture vocabulary and the parameters that can be tuned to achieve that. This section presents a procedure to find an appropriate gesture vocabulary for a given application.

In order to ensure intuitive and logical mapping, the first step is an investigation into the interfaces of known applications is conducted to identify the needed functionalities.

In the development of interfaces scenarios have proven valuable [21] to define the context, functionalities, and investigate the user and problem domain.

A tool in this approach is to examine human-to-human non-verbal communication in these scenarios by preparing scripts. Testees will be taken through the scenarios such that they will communicate the same things to a person as they would communicate to the computer application.

Points 1 and 2 of the nine usability heuristics from section 2.1 support the view that the gestures must be chosen by looking at natural gesturing, but also show that the testees must be part of the designated user group.

Two approaches of this investigation are at hand; bottom-up and top-down. Bottom-up takes functions and finds matching gestures, while the top-down presents gestures and finds which functions are logically matched with those. Another tool that is needed is a benchmark to measure the goodness of gestures by the principles that are valued in the human-based approach.

3.1 Procedure to Finding Gestures

This section describes the proposed procedure and benchmark when developing a gesture interface. In section 4 an example of execution will be shown in order to test and improve the procedure, as the development of such a procedure and benchmark is an iterative process.

3.1.1 Step A. Find the Functions

Find the functions needed by the application and which the gestures will have to communicate. Keep in mind the user interface of existing similar applications of standard interfaces (e.g. if the new application is an architectural design application, 3D Studio, MicroStation, CAD, etc. would be relevant).

Keep the gesture vocabulary at a minimum, e.g. with use of context menus or spatial zones (the same gesture activates different things depending on context).

3.1.2 Step B. User Tests – Collect Gestures from User Domain

The goal is to find the gestures that represent the functions found in step A. This is done through experiments with people by taking them through scenarios under camera surveillance where they communicate the above functions, which they would otherwise communicate to the computer, e.g. to the "operator" (i.e. the person who conducts the experiment) or another testee.

It is important to design the experiments on a way that the testees use gestures in a natural way, especially when conducting the tests with technically minded people. Otherwise, there is the risk that they will still think in terms of interfaces and algorithms. If it is desired to write a scenario with a technical interface aspect, it can be performed as a Wizard-of-Oz experiment [3][10], which tests not only the gesturing, but also the design of the entire interface, including the feedback from the system and the sequencing in the interfacing.

The number of people that are required for this investigation depends on how broad the user group is, and how diverse the results of the test are.

3.1.3 Step C. Analysis – Extract Gesture Vocabulary

The video recorded data is evaluated to extract the gestures that the testees used in their interaction.

Note and capture the frames with the commonly used gestures, and note how consistently the different testees use them. Note if they are used only as static postures or if the dynamics play an important part in the interpretation of the gesture.

The theory in section 2 should be taken into account in the selection of gestures:

- Evaluate internal force caused by posture
  - Deviation from neutral position
  - Outer limits
  - Forces from inter-joint relations
- Evaluate frequency and duration of that gesture
- Consider effect on wrist from wrist and finger posture

See section 4 step C how this is done in praxis. Furthermore, note that the selection should not be strictly limited to the recordings. It is meant to be an inspiration.

3.1.4 Step D. Test – Benchmark the Chosen Gesture Vocabulary

The final step is to test the resulting gesture vocabulary. This might lead to changes in the gesture vocabulary. The test has three parts. The following is to be tested in the benchmark: Semantic interpretation, generalisation, intuitivity, memory, learning rate, and stress. The lowest score is best.

Test 1: Guess the Function

Give the testee a list of functions. Present the gestures and ask the person to guess the functions. Gestures that depend on context must be presented in context. Score = errors divided by number of gestures

Test 2: Memory

Give the gesture vocabulary to the testee, who will then try the gestures to make sure they are understood.

Present a slideshow of names of functions in a swift pace, 2 seconds per function. The testee must perform them correctly while the name is displayed. The order should be logical towards sequences in the application. Restart the slideshow at every mistake, and show the gesture vocabulary to the testee between each retry. Continue until they are all correct. Score = number of restarts.

Test 3: Stress
This is a subjective evaluation of ergonomics. Present the list with a sequence of the gestures. The testee must perform the sequence X times, where X times the size of gesture vocabulary equals 200. Between each gesture go back to neutral hand position Note how stressing they are. Make space for comments to illuminate if it was certain gestures that gave stress.

Use the following score list for each gesture and overall for the sequence: 1) No problem. 2) Mildly Tiring/Stressing. 3) Tiring/Stressing. 4) Very annoying. 5) Impossible

The benchmark can be used to compare two gesture vocabularies, but test 2 is only comparable if the vocabularies are of the same size. If testing a single vocabulary, reasonable success criteria must be stated. These aims depend on the gesture vocabulary at hand. See section 4 step D how this can be done in praxis. It can also be used to compare different user profiles and test cultural dependences.

4 Human-Based Experiment

This section demonstrates and tests how to use the procedure, and the theory that is presented in sections 2.1 through 3.1.

The test case for testing the approach is an architectural design application. The application enables the user to place virtual 3D objects on the table, moving them, and changing style settings on them. The key functions of a simple design interface are: Activate menu, select, select all, insert, move, scale, rotate, delete, yes/confirm, no/undo, and copy-paste. In order to cover conscious and subconscious top-down investigation three scenarios are chosen for the user tests, which are outlined below:

- ***Scenario I: The Pub Visit***
  - Noisy environment
  - Verbal communication prevented
  - Spontaneous and forced gesturing
- ***Scenario II: Planning of Furnishing***
  - Collaboration through verbal communication
  - Spontaneous gesturing
- ***Scenario III: Simulated Furnishing***
  - Directing through gestures
  - Conscious gesturing

Scenario I simulated a pub visit, where similar messages are being conveyed to the bartender in a noisy environment, where gesturing is necessary. The remaining scenarios were based on the principle that a floor plan is to be furnished with templates of furniture. In the first scenario two testees are verbally discussing how they want to furnish the room. In those discussions they will use some spontaneous gestures, mainly deictic and propositional. Scenario 3 is similar to a wizard of oz experiment, where a testee directs an operator through gestures alone, being aware of which functions are available. Paper and scissors are available for the operator to simulate scaling and copying. The testees used in the scenarios were engineers.

Fig. 1. Scenario II-III from stationary camera (left) and head-mounted camera (right).

The camera setup for all scenarios is that a stationary camera, records the testee including audio. The testee is equipped with a head-mounted camera on headphones, see Figure 1. The number of testees is limited to 5-7 in this test of the procedure.

4.1 Analysis – Extract Gesture Vocabulary

Once the scenarios were done, the gestures used were observed and counted for each function.

The gesture variation and frequency were noted. It was also noted whether the information in the gesture is static and dynamic. Sometimes the information in a dynamic gesture is actually static. Then the gestures were evaluated in terms of ergonomics, functionality and recognition. The gestures found in the video material should be seen as an inspiration, but the choice is not limited to those. The result for move will be discussed below.

Table 1 shows the result of the move gesture from scenario III. The full results are available in [22]. The gestures resulting from the analysis are presented in Figure 2.

Table 1. Results – Move. S = Static, D = Dynamic, N = Frequency of particular gesture.

Function	F	Gesture	S/D	N
Move	12	1. "Put-that-there"	D	4
		2. Wave it in the desired direction, stop with confirm or palm up like "halt".	D	3
		3. Wave it in the direction, rely on "snap to grid" to stop movement.	D	5

Move Gesture: There were variations of *move* gestures. In the pub scenario they were characterized by the fact that there were two bottles that needed to be moved aside, hence, the "open gate" and "birdie flapping wings" gestures. These can be broken down to waving the hand one way for one bottle and then the other way for the other bottle. This leaves only two versions of the move gesture: "put-that-there" and "wave

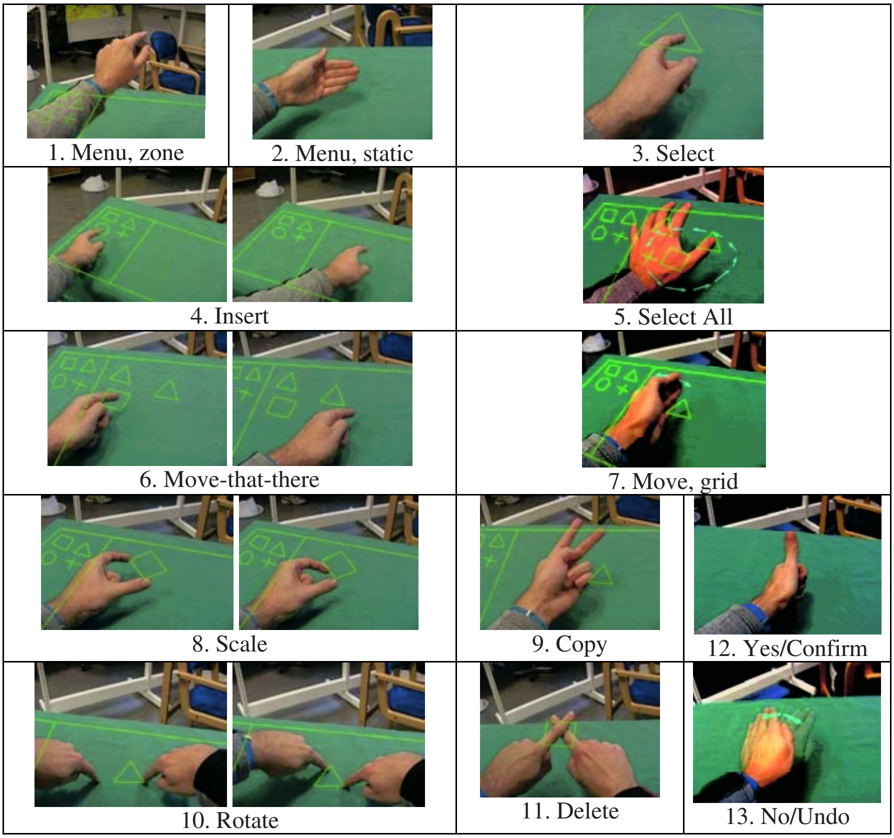

Fig. 2. Resulting gestures. The small light arrows indicate movement.

in the desired direction". Notable is that the first was used mainly when moving an object far, while the latter was used to move it slightly to a wall or the middle between two walls. It seems natural to implement both, and the waving gesture moves one tick at a time in a "snap to grid"-feature. The gesture is not mistaken for "select all" or "no/undo", because the palm faces sideways, instead of downwards.

4.2 Step D. Benchmark

The benchmark was done with 3 test groups; 8 engineers, 5 architects, and 11 engineer students.

In benchmark test 1 the "menu, using zone" is removed, because its context is hard to illustrate without doing a Wizard-of-Oz experiment. This leaves 12 gestures to be guessed, and it is expected that people will get up to 2-3 errors: The menu gesture is not very logical and the no/undo and delete gestures are interchangeable.

For test 2 it is a rather big vocabulary to learn, and the slideshow is performed at a speed of 2 seconds. It is expected that people will need 3 retries.

The stress in test 3 is expected to be mildly tiring, because it is laborious to perform 200 static and dynamic gestures in a row without resting. The results of the

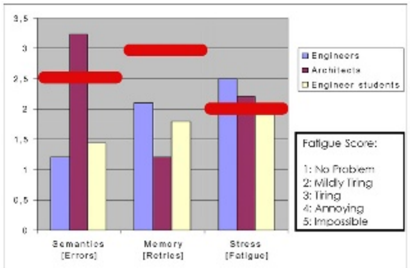

Fig. 3. Benchmark results. Comparison between 3 test groups. The horizontal lines denote the expected results. It shows that the engineer groups scored better in the semantic test, while architects scored better in the memory test.

benchmark are found in figure 3. It is interesting that the two engineer groups performed similarly, and notably different from the architect group.

In test 1 the problems for the engineer groups were mainly those that were expected, but "Move, using grid" and "no/undo" got mixed up as well. The architects did not perform well and explained that they could not follow the logic in the iconic symbolism. This supports the notion of cultural dependence and that the user groups must be chosen carefully.

Test 2 showed surprisingly that it was generally easier to remember this large vocabulary, especially for the architects. Their spatial intelligence and practice with visual objects seem to give them an advantage.

Half of the testees found the gesturing to be more tiring than expected. The mean is just between mildly tiring and tiring. The question is where the subjective threshold is between the categories. Individual stress assessments are presented in table 2.

Table 2. Individual Gesture Stress scores. Highlighted results are commented in the text.

Gesture	Avg.Score	Variance
Rotate	***1.88***	***1.27***
No/undo	***2.13***	***1.27***
Select	1.13	0.13
Copy	***2.13***	***2.41***
Move, grid	***1.75***	***0.79***
Scale	1.25	0.21
Menu	1.38	0.27
Select All	1.5	0.57
Delete	1.5	0.57
Yes/Confirm	1.5	0.21
Move, Insert	***1.63***	***1.13***

Several testees found it laborious to use two hands in general. For rotation this is a worthwhile trade off for the precise rotation control given by using both hands like this. For deletion it is deliberate in order to prevent spontaneous deletion of objects.

The gesture no/undo was given either 1 or 4 by the testees. The stress on those who gave it 4 came from producing the movement with radial- and ulnar deviation in the wrist with the forearm muscles. The correct method is to do the motion with the entire forearm and elbow. It may be necessary to convert the gesture into a static gesture; to simply hold the palm out.

Concerning copy: Stress comes from doing selection with a pronated forearm. The result is that the hand must be supinated rather far, and the testees who found it stressing did it as far as the motion limit. “Move, using grid” is similarly dependent of flexing/extending the wrist or using the entire forearm, or just the fingers’ TCP joints to do the waving.

Move and Insert was found laborious because it was in two steps and can be a long movement. Obviously, this depends on how far the object is to be moved. It is especially stressing when the hand has to cross the body axis. The distance of the gesturing from the body also has impact on the physical fatigue.

The stress benchmark clearly proves the importance of clear instruction how to perform the gestures, and to be aware of the entire body posture when using gesture interfaces.

5 Conclusions

A procedure to find gestures based on human factors is presented and tested. While the traditional technology-based approach is technically solvable, it leads to an awkward gesture vocabulary without intuitive mapping towards functionality, and a system which works under strictly pre-defined conditions.

The human-based approach and procedure described in section 3 was tested and lead to an easy-to-use gesture vocabulary. It was fast for the testees to learn and remember it. However, it is rather time consuming, and the scenarios must be carefully written. This time is well spent, if it is a matter of the future users using the application or preferring another application, because it is too stressing or slow to use.

It was shown in the test that user profile is relevant when choosing a gesture vocabulary, because the gestures extracted from engineers performed best with engineers.

The experiments also revealed that gesturing concerns more than just a hand posture or movement. It affects a greater part of the body, biomechanically speaking. It is important to analyse and instruct the user in the execution of the gestures and the posture of the entire upper body.

Usability heuristic 4, to be consistent, is a tricky when dealing with gesturing. There are no platform standards to be consistent with. Being consistent towards nonverbal communication is desirable, but most gesturing is spontaneous. This means that people are not aware of the gestures they use. Test scenario III showed an interesting aspect, though. Despite the fact that the testees chose their gestures consciously, they were not consistent in their choices. The same person would use 2-3 different gestures for the same instruction. That gives the developer of a gesture interface some degree of flexibility in choosing the gesture for a function.

5.1 Future Work

Further testing and tweaking is necessary, increasing the number of testees for better generalization and cultural coverage. Benchmarking a single gesture vocabulary requires goals, which can be hard to predefine. The benchmark needs to be tested by comparing different gesture vocabularies.

The computer vision recognition of the human-based gesture vocabulary is hard to solve technically, and the question stands how this will be solved. With the aid of vast corpora the technology may be driven towards robust solutions. The use of shared resources and data sets to encourage the development of complex processing and recognition systems has been very successful in the speech analysis and recognition field. Applying this in gesture recognition might help to produce robust solutions, which is the aim of the current FG-Net project.

Acknowledgements

This research is in part funded by the ARTHUR project under the European Commission's IST program (IST-2000-28559) and the FG-Net project under the European Commission's IST program (IST-2000-26434). This support is gratefully acknowledged. The authors would also like to thank the participants of the experiments.

References

1. Justine Cassell, "A Framework For Gesture Generation and Interpretation" in Cipolla, R. and Pentland, A. (eds.), Computer Vision in Human-Machine Interaction, pp. 191-215. New York: Cambridge University Press, 1998
2. W.T.Freeman and C.D.Weissman, "Television Control By Hand Gestures" from IEEE Intl. Wksp on Automatic Face and Gesture Recognition, June 1995.
3. C. Hummels, P.J. Stapers, "Meaningful Gestures for Human Computer Interaction: Beyond Hand Postures", Proceedings of the 3rd International Conference on Automatic Face &Gesture Recognition (FG'98), Nara, Japan, April 14-16. IEEE Computer Society Press, Los Alamitos, CA. 591-596, 1998.
4. T. Moeslund, M. Stoerring, E. Granum, "A Natural Interface to a Virtual Environment through Computer Vision-estimated Pointing Gestures", In I. Wachsmuth and T. Sowa (Eds.),GW2001 Springer LNAI 2298, pages 59-63, 2002.
5. Patrizio Paggio and Bradley Music, "Linguistic Interaction in Staging – A Language Engineering View" in L. Qvortrup (ed.) Virtual Interaction: Interaction in/with Virtual Inhabited 3D Worlds, 2000
6. S. Steininger, B. Lindemann, T. Paetzold, "Labeling of Gestures in SmartKom – The Coding System" in I. Wachsmuth and T. Sowa (eds.) GW 2001, LNAI 2298, pp 215-227, 2002
7. Norbert A. Streitz et al, "Roomware: Towards the Next Generation of Human-Computer Interaction Based on an Integrated Design of Real and Virtual Worlds", German National Research Center for Information Technology, Integrated Publication and Information Systems Institute, Germany, 2001
8. Simeon Keates, Peter Robinson, "The Use of Gestures in Multimodal Input", University of Cambridge, Proceedings of ACM SIGCAPH ASSETS 98 35-42., 1998
9. Fingerworks, "iGesture Pad", http://www.fingerworks.com/igesture.html
10. Nicole Beringer, "Evoking Gestures in SmartKom – Design of the Graphical User Interface" in I. Wachsmuth and T. Sowa (eds.) GW 2001, LNAI 2298, pp 228-240, 2002
11. Marcello Federico, "Usability Evaluation of a Spoken Data-Entry Interface", ITC-Irst Centro per la Ricera Scientifica e Technologica, 1999
12. Nigel Bevan, Ian Curson, "Methods for Measuring Usability", Proceedings of the sixth IFIP conference on human-computer interaction, Sydney, Australia, 1997.
13. Jakob Nielsen, "The Usability Engineering Life Cycle", IEEE, 1992

14. J. Lin, Ying Wu, T.S.Huang, “Modeling the Constraints of Human Hand Motion”, Proc. 5th Annual Federated Laboratory Symposium(ARL2001), Maryland, 2001.
15. Jintae Lee and Tosiyasu Kunjii, “Model-Based Analysis of Hand Posture”, University of Aizu, IEEE, 1995
16. Charles Eaton MD, “Electronic Textbook on Hand Surgery”, http://www.eatonhand.com/, 1997
17. Keir, Bach, Rempel, “Effects of Finger Posture on Carpal Tunnel Pressure During Wrist Motion”, Division of Occupational Medicine, U.C.S.F., 1998
18. Chris Grant, “Ten Things You Should Know about Hand and Wrist Pain”, F-One Ergonomics, Ann Arbor, Michigan.
19. G. Shaw, A. Hedge, “The Effect of Keyboard and Mouse Placement on Shoulder Muscle Activity and Wrist posture”, CU Ergo, Cornell University.
20. A. Hedge, T. M. Muss, M. Barrero, “Comparative Study of Two Computer Mouse Designs”, Cornell University, 1999
21. Wolfgang Dzida and Regine Freitag, “Making Use of Scenarios for Validating Analysis and Design”, IEEE, 1998
22. Michael Nielsen, Thomas Moeslund, Moritz Störring, Erik Granum, “A Procedure For Developing Intuitive And Ergonomic Gesture Interfaces For Man-Machine Interaction”, Technical Report CVMT 03-01, ISSN 1601-3646, http://www.cvmt.dk/~mnielsen, 2003
23. Jakob Nielsen, http://www.useit.com/