Technical Report CVMT 03-01, ISSN 1601-3646, CVMT, Aalborg University, March, 2003

A PROCEDURE FOR DEVELOPING INTUITIVE AND ERGONOMIC GESTURE INTERFACES FOR MAN-MACHINE INTERACTION

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

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

ABSTRACT

Many disciplines of multimedia and communication go towards ubiquitous computing and hand free interaction with computers. Application domains in this direction involve virtual reality, augmented reality, wearable computing, and smart spaces. This paper presents two main approaches to developing and testing these interfaces using gestures. It presents the important issues in gesture communication, from a technological viewpoint as well as a user viewpoint such as the technological complexity, learning rate, ergonomics, and intuition. These issues must be taken into account when choosing the gesture vocabulary. 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 and time consuming. The importance of using theory from ergonomics is also concluded.

1. INTRODUCTION

A lot of 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 an alternative to existing interface techniques, such as the old desktop paradigm. This paper proposes that a gesture interface should be developed specifically for a given task. Good 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 gesture interface”. A gesture interface is not universally the best interface for any application. The objective is “to develop a more efficient interface” to 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. If the artist is 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 aesthetically be closer to the artist’s vision. Thus, the choice of interface is a matter of what outcome of the application is desired.

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 interface and to find the specific gestures that make this specific interface for this specific application most efficient.

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

2. CONCERNING GESTURES

This section will investigate gesturing and approaches to finding the gestures to use in an interface.

2.1. Taxonomy

The set of gestures in an interface will be called the “gesture vocabulary”. This should not be confused with a general non-verbal communication dictionary.

There are several ways of labelling gestures, e.g. from a descriptive point of view and a semantic point of view. Descriptive labelling is describing the movement of the gesture, while the semantic labels refer to what the gestures communicate and their purpose.

Descriptive labels:

Static gestures are postures, i.e. relative hand- and finger positions, not taking the movements into account. *Dynamic gestures* are movements, i.e. the hand trajectory and/or posture switching over time. Dynamics in a gesture can alter their meanings [3].
Spatio-temporal gestures are the subgroup of dynamic gestures that moves through the workspace over time.

Semantic labels:

Semantic type labels, as described by Justine Cassell [1], can be conscious or spontaneous, interactional or propositional. Conscious gestures have meaning without speech, while spontaneous gestures only have meaning in context of speech.
Emblems are conscious communicative symbols that represent words. These are interactional gestures. An example is a ring formed by the thumb and index finger. In western culture this means “O.K.” and in Japan it means “money”.
Propositional gestures are not interactional but consciously indicate places in the space around the performer and can be used to illustrate sizes or movement. Examples include “it was this big” or “put-that-there”.
More frequent are the spontaneous gestures:
Iconic gestures are illustrations of features in events and actions, or how they are carried out. This can be to depict how a handle was triggered, or looking around a corner.
Metaphoric gestures are like iconic gestures, but represent abstract depictions of non-physical form. Circling the hand to represent “the meeting went on and on” is this type.
Deictic gestures refer to the space between the narrator and the listener(s). This can be pointing towards objects or people being told about, or merely refer to movement or directions. They can also be used in a more abstract way to wave away methods in “we don’t use those”, and picking the desired methods in front of one self as in “these are what we use”. These are mainly spontaneous, but can also occur consciously, like when pointing at an object.
Beat gestures are used to emphasize words. They are highly dynamic, as they do not depict the spoken messages with postures. An example is when the speaker makes a mistake in the speech and correct one self, while doing a punch with the hand.
Cassell states that emblems and metaphoric gestures are culturally dependent, while other types are mostly universal, though some cultures use them more than others. Asian cultures use very little gesturing while Southern European cultures use a lot gesturing.
The gestures that are generally relevant for machine interaction are deictic, iconic, pro-positional, and emblems.

An important question to ask is if the cultural dependence is a problem. Conventional interfaces that are international are generally in English, but most software is available with selectable national language packages, and some nations have different keyboards. In a gesture interface this can be translated to selectable gesture sets, if it should become a problem that an emblem 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 as presented.

2.2. Choosing the gesture vocabulary

One of the most difficult parts is to find a feasible gesture vocabulary. It is a risk that the gesture interface is hard to remember and perform for the user [8]. Sign language is not convenient to use as the gestures are rather complicated, and there is a sign language for different languages.
It is important to limit the vocabulary. This will benefit both the users as well as the technical solution in having to learn fewer gestures.
Methods that can be used for this purpose include:

- Context dependence: Available options vary with the context of the current selection. This is like the context menu in Windows applications, where a right click spawns a menu of items that are only relevant to the selected object.
- Spacial Zones: The space around the user is divided into zones, which has their own contexts that define the functions of the gestures.

2.3. Technology based vocabulary

In this approach to finding the gesture vocabulary the focus is to make it easy for the recognition algorithm to recognise the gestures. The core of the approach is the following:

- Easy to recognise technically

An example is to define a gesture by how many fingers are stretched, see Figure 1. These gestures make the gesture vocabulary for arbitrary applications, whose functionalities are forced upon those gestures without being logic towards their semantic interpretation.

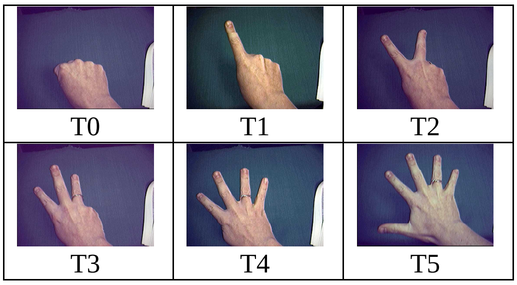

Figure 1. Technology based gestures. The author cannot perform these gestures.

Table 1 shows examples of applications, where functionalities have been assigned to the gestures.

Table 1. Imposed functionalities for demo applications:

Application Gestures	Painting	Object handling
T0	Residue/Release	Residue
T1	Paint/Select	Select
T2		Copy-paste
T3		Delete
T4-5	Menu	Release

These applications have been implemented and tested. They showed that these gestures are inappropriate for various reasons:

- Stressing
- Impossible for some people to perform
- Illogical imposed functionality

2.4. Human based gesture vocabulary

Section 2.3 showed that by choosing technology based gestures, it is relatively easy to implement the interface, but at the expense of usability. 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.

The five principles of usability are [11][12][13]:

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

These principles are important for the entire interface that the gestures are used in. This means the structure, sequencing, and feedback from the user interface, but also the gestures themselves. By sequencing is meant the steps that are required to get from one state to the goal state. An example is word processing or file management. There are two methods to move a section or file: copy-paste-delete, or cut-paste. This is an example of minimizing the steps in the sequencing. However, some people may prefer the three steps, because it provides a security by leaving the original in place, in case something goes wrong in the transaction.

Furthermore, usability engineering follows nine heuristics [13]:

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

Given these usability principles, the core of the human based approach is the following characteristics:

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

These principles facilitate Learnability and Memorability, and minimize the chance for Error. Efficiency and Coverage is mainly a task for the sequencing and structure to ensure. With a gesture-based input, it is even more necessary to simplify the dialogue and sequencing than the conventional input methods, as there will be less possible commands to give.

Furthermore, it will be convenient to keep a relaxed version of the technology-based core in mind:

- Possible for an application to recognise unambiguously

In order to simplify the dialogue, functionalities can be accessed through a menu, and feedback should be limited to yes and no queries.

2.5. Ergonomics and biomechanics

The gestures in section 2.3 were found stressing to perform. This is why it is relevant to look into the ergonomics and biomechanics of gesturing to ensure that a physically stressing gesture is avoided. This theory will be used in the selection and instruction of gestures.
In this section the biomechanics in the hand [14][15] is described. Figure 2 and Figure 3 shows the terms used in this section.

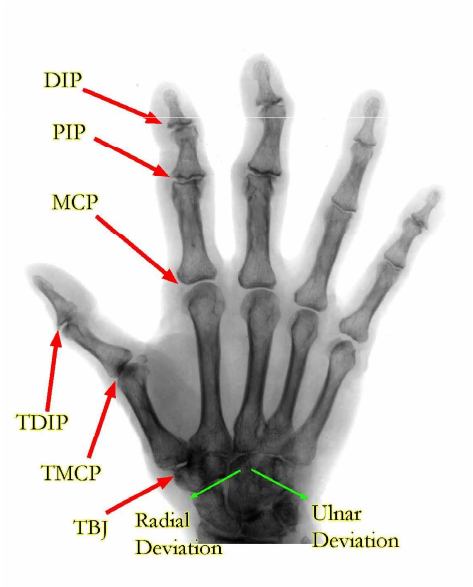

Figure 2: X-ray of right hand. Joint names and wrist movement.

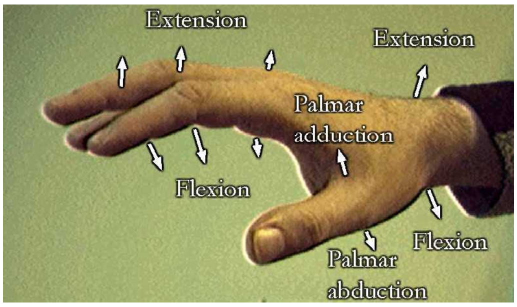

Figure 3: Hand from the side. Finger- and wrist motion.

Table 2 lists the ranges of motion for the joints [16] for the average hand in degrees, where zero degrees in all joint angles gives a stretched hand. The wrist extension/flexion of 70/75 thus mean that the wrist can extend 70 degrees upwards, and flex 75 degrees downwards.
Hyperextension is to extend the joint farther than naturally by external force.
Adduction is to move the body part towards the central axis, which in the hand is between middle and ring fingers. This is like gathering the fingers. Abduction is to move the body part away from the central axis. This is like spreading the fingers. Pronation is to rotate the wrist around the forearm. If the neutral position faces the palm sideways, pronation faces the palm downwards, and supination faces it upwards.

Table 2: Range of motion for joints in the hand and wrist. Degrees. (*1) More than one joint. (*2) Rotation is done with the entire forearm.

Motion Joint	ExtensionFlexion	Hyper Extension	Adduction Abduction
MCP	90	0-45	30
PIP	100	0	0
DIP	80	0	0
TMCP	90	10	0
TDIP	80	15	0
		Palmar Adduction Abduction	**Radial Adduction Abduction**
TBJ		Contact/45	Contact/60
		Pronation/ Supination *2	**Radial/ Ulnar**
Wrist *1	70/75	70/85	20/35

These numbers are the static constraints for gesture postures. However, there are also dynamic constraints, which are inter-finger- and intra-finger constraints.
Intra-finger constraints are the dependencies between the joints in the fingers. Inter-finger constraints are the dependencies between the postures of the neighbouring fingers.

Furthermore, finger postures affect pressure in the carpal tunnel [17], most severely with metacarpophalangeal joint angles (MCP) at zero degrees (fingers extended), and least at 45 degrees.

The main principles in ergonomics include [18][19][20]:

- Avoid outer positions
- Avoid repetition
- Relax muscles.
- Relaxed neutral position is in the middle between outer positions
- Avoid staying in static position
- Avoid internal and external force on joints and stopping body fluids

Given these aspects, a gesture interface is by nature ergonomically superior to physical handheld devices, which introduces external force. Considering and comparing biomechanics in gestures, it is the internal forces and posture angles that are interesting.

The TV control gesture in [2] is a tight fist. Feedback is provided to navigate a pointer on the screen. The dialogue is very simple, but tests showed that the gesture was very tiring. The ergonomics in the gesture supports this conclusion: The neutral position is for all joints in the middle between outer positions. This means that a fist is a forced position by the muscles in the hand. Furthermore, the hand is raised to head height, or higher, which places the shoulder in or close to an outer position of external rotation (rotated away from body), with the weight of the arm and hand on it. On the positive side, the wrists are kept straight while operating the TV.

The gestures in Figure 1 are stressing because they do not follow the inter-finger constraints of the hand. While the limitations of the constraints are different for each person, the user must use more or less force to position the fingers for the system to recognise the correct posture. Furthermore, the recognition is vulnerable to how stretched the fingers are. This means that the users must stretch them to outer positions.

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 an investigation into the interfaces of known applications is conducted to identify the needed functionalities.

In section 2.4 a set of human based gestures was found with the given principle in mind. However, a formal experiment must be conducted in order to include the following:

- Generalisation of intuitive mapping of functionality and gestures
- Semantic interpretation of gestures
- Cultural differences
- Natural unconscious gesturing in human-human communication
- Physical stress of gestures

It is important to remember that a gesture set must be tailored for the purpose of the application, and for the user group of the application. A good gesture set should not be imposed on any arbitrary application.

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 by writing scenarios for testees. They will be taken through scenarios where 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 2.4 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 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 topdown presents gestures and finds which functions are logically matched with those. Below are some outlines of the approaches:

Subconscious, bottom up: Create scenarios of communicating between people. Record the transactions on video and analyse the subconscious gesturing that occur.

Conscious, bottom-up: Ask for each functionality which gesture would be used, or make a guided drawing test. A testee gets a random drawing project and guides the

operator with gestures what to draw. Objects to be drawn are on the table.
Top-Down: Ask each testee what a given gesture could mean, or do a drawing test: The operator guides the testee with gesturing what to put down on the paper using a predefined gesture vocabulary. This is useful for testing a gesture vocabulary.

Another tool that is needed is a benchmark to measure the goodness of a gesture by the principles that are valued in the human-based approach.

3.1. Procedure to find 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.

Step A. Find the functions.

Find the functions that the gestures will have to access. Keep in mind the user interface in existing similar applications on standard interfaces (e.g. if the new application is an architectural design application, take a look at 3D Studio, MicroStation, CAD, etc.).
Keep the gestures vocabulary to a minimum, e.g. with use of context menus or spacial zones (one gesture activates different things depending on context).

Step B. Find logical gestures

Find the logical gestures that represent the functions found in step 1. This is done through experiments with people by taking them through scenarios under camera surveillance where they communicate the same messages, which they would communicate to the computer, to the "operator" (i.e. the person who conducts the experiment).
It is important that the scenarios take the testees away from normal technical thinking, especially when conducting the tests on technically minded people. Otherwise, it is a 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, 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.

Step C. Process the obtained data

The recorded data is processed 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 2.5 should be taken into account in the selection of gestures:

- Evaluate internal force caused by posture
  - o Deviation from neutral position
  - o Outer limits
  - o 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.

Step D. Benchmark the chosen gesture vocabulary

The final step is the test the resulting gesture vocabulary. If there are more vocabularies these can only be compared if there is a standard test available.
The following is to be tested in the benchmark:

Attribute	Tested in test no.
Semantic interpretation	1
Generalisation	1
Intuitivity	1, 2
Memory, learning rate	2
Stress	3

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 = wrong guesses divided by number of gestures

Test 2: Memory

Give the gesture vocabulary to the testee, who will them try the gestures to make sure they are understood.
Present a slideshow of functions in a swift pace, 2 seconds per function. The testee must perform them correctly. Continue until they are all correct. 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.
Score = number of restarts.

Test 3: Stress

This is a subjective evaluation of ergonomics.
Present list with a sequence of 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 room for comments to illuminate if it was certain gestures that gave stress.
Use the 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 is done in praxis.

4. HUMAN-BASED EXPERIMENT

This section demonstrates and tests how to use the procedure and theory that is presented in sections 2.4 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.

4.1. Step A. Find the functions

In this step the functions that are to be accessed by gestures are found.
The presence of a menu severely limits the need for various gestures, because all functions can be accessed through a menu. Of course, in a standard window based application, the user can choose to use the menu as well as various shortcuts, which the usability heuristics also recommend.
The test will therefore also identify possible gestures as shortcuts to key functions, as advised by point 6 of the nine heuristics in section 2.4.
The key functions of a simple design interface are:

1. Activate menu
2. Select
3. Select all
4. Insert
5. Move
6. Scale
7. Rotate
8. Delete
9. Yes/confirm
10. No/undo
11. Copy-Paste

4.2. Step B. Find logical gestures

Prepare scenarios that implement the types of messages needed for the application.
In order to cover conscious and subconscious top-down investigation, three scenarios are chosen:

- The Pub
- Planning of furnishing
- Simulated furnishing

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. The number of testees is limited to 5-7 because this is a test of the procedure.

Scenario A: The Pub

This is written as a mixture of conscious and subconscious bottom-up gesture study. The testee will consciously try to communicate with gestures knowing the recipient cannot hear well, due to the noise level in the pub. The scenario removes the testee completely from technical thinking.
This scenario tests the gesture functions: Menu, Select, No/Undo, “Copy”, Yes/Confirm, Scale/Modify attribute, Select all, Move.

Script summary: The bartender cannot hear anything, because of the loud music. The testee asks for the menu card, but there is no menu card. The testee sees that there is a bottle behind the others and wants the bartender to move the bottles aside to see what it is.
Testee wants to order one Highland Park whisky.
The bartender misunderstands and points at Cragganmore whisky. Testee tells no, and bartender points at Highland Park, and testee confirms. Bartender asks for how much. When testee gets the whisky, (s)he decides for two instead, and bartender gives one more. Afterwards the testee decides to get one of all of them.

The testees are divided into pairs and are given two different customer scripts. When one testee is the customer the other is the bartender and vice versa.
The two scripts together contain all the functions of the script summary.

The setup is a small counter and a display behind it with whiskies , see Figure 4.

Figure 4. Scenario A from stationary camera (left) and head-mounted camera (right).

Scenario B: Planning of furnishing, subconscious study

Furnishing a room on paper. There are paper templates on the table of objects to be placed in the room.
This scenario is closer to the actual application than the first scenario.
This is natural communication. Discuss the furnishing and placement of door/windows on the draft living room.
The testee gets no script to follow, because his communication should be natural.
Two testees can discuss the furnishing at a time, and one of them is chosen to be the template master.

Scenario C: Simulated furnishing, conscious study

Like scenario B but the testee is conscious about the functionalities that he can use and use gestures to communicate. It is a kind of Wizard-of-Oz experiment, except that there is no computer between the operator and testee. See setup in Figure 5.
The testee can only do gesturing to design the room. The operator must not know what the testee intends. The testee will be shown a list of all functions that must be used. The list is hidden, but can be called for by “activate menu”.
To simulate copies and scaling, a blank piece of paper, a pen, and scissors must be present.

Figure 5. Scenario B-C from stationary camera (left) and head-mounted camera (right). Testee is doing a rotation gesture.

4.3. Step C. Process the data

For each function, the various gestures are found. The each gesture, it is important to note whether the information lies in the static posture alone or in the dynamics. Furthermore, the frequency of the gestures is evaluated. Gestures found in the pub scenario are described in Table 3.

Table 3. Scenario A - The Pub - results. F = Frequency, S/D = Static/Dynamic, N = Frequency of particular gesture. N is filtered from repetition if misunderstood or due to impatience.

Function	F	Gesture	S / D	N
Drink card	8	1. Iconic for a square	D	4
		2. Iconic for folding out paper vertical	D	2
		3. Iconic for folding out paper horizontal	D	1
		4. Iconic for writing on paper	D	1
Select	20	1. Pointing with index finger	S	15
		2. Wave at object while pointing between object and oneself	S	4
		3. Wave with backside of hand between object and oneself	D	1
Select all	11	1. Wave sideways pointing at all	D	3
		2. As (1) but show 3 fingers first	D	1
		3. Iconic for “cover it all”, like smearing sun block in the back of another person	D D	1 1
		4. Metaphoric for “what? I don’t understand”	D	2
		5. Wave towards oneself	D	3
		6. Select pointing at each bottle in a row		
Move	11	1. Birdie waving wings and pointing	D	1
		2. Circlish wave	D	1
		3. Iconic for “open gate”	D	1
		4. Wave hand, move the direction of the palm	D	5
		5. Wave hand, move the direction of the backside	D	1
		6. Wave finger	D	2
Double size	5	1. Point at bottle, wave to oneself (more, more, gimme, gimme)	D	1
		2. Iconic for “grow (taller)”	D	1
		3. Iconic for “pour two times”	D	1
		4. Two fingers like “two”	D	1
		5. Two layered fingers, emblem for a “double drink”.	S	1
Yes/ confirm	7	1. Emblem: Thumbs-up	S	2
		2. Point upwards as in “one”	S	2
		3. Wave up	D	1
		4. Facial: Nod	D	2
No/ stop	4	1. Emblem “Halt”	S	1
		2. As (1) but with waving	S	2
		3. Facial: Shake head	D	1
Copy/ One more	8	1. Point at glass	S	1
		2. Wave towards oneself	D	2
		3. Two fingers	S	4
		4. Iconic for “pour two times”	D	1

In scenario B the testees did not abide by the rule that only one person should control the templates. It is not relevant to extract gestures as means of delivering messages in this scenario. The gesturing found are propositional and deictic gestures and direct interaction of the objects. The results will be discussed in relation to those of scenario C, which are described in Table 4. The most important result of scenario B was that the testees were prepared for scenario C and got accustomed with the templates and the system.

Table 4. Scenario C - The Living Room - results. F = Frequency, S/D = Static/Dynamic, N = Frequency of particular gesture. N is filtered from repetition if misunderstood or due to impatience.

Function	F	Gesture	S/D	N
Activate menu	14	1. Wave palm upwards like "lift it up"	D	4
		2. Wave index finger upwards like "lift it up"	D	2
		3. Flip hand forward, palm faced up towards self on the table	S	3
		4. Point forward in the air as if pressing button	S	3
		5. Point distant outside work space	S	2
Select	59	1. Pointing with index finger	S	59
Select all	4	1. Circle palm over entire workspace	D	1
		2. Iconic for "pile it all together"	D	1
		3. Wave spread extended hand over workspace	D	2
Insert	37	1. Point at object in template zone and point at workspace where to put it "put-that-here"	D	37
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
Scale	4	1. Both palms squeeze together	D	1
		2. Index and thumb indicate old size and moving to new size.	D	2
		3. Hands shape square, and indicates growth.	D	1
Rotate	14	1. Iconic for grabbing a ball and rotate	D	2
		2. Rotating index and thumb as if manually rotating object.	D	8
		3. As (2) but using both index fingers	D	1
		4. Pointing and rotating whole hand around finger	D	1
		5. Point and move in circles	D	1
		6. Flip. Point with index and thumb and rotate around wrist	D	1
Delete	8	1. Point at template zone	S	6
		2. Crossing both index fingers in a fencing movement	D	1
		3. Wave hand over object	D	1
Yes/ confirm	10	1. Thumb up	S	9
		2. Point upwards	S	1
No/undo	5	1. Wave up-pointing index finger, like "naughty, naughty"	D	3
		2. Wave palm	D	2
Copy-Paste	9	1. Two fingers held up, point at new location	S	3
		2. Multiple copies by showing how many, like (1).	S	3
		3. Select on the menu	S	3

4.3.1. *Discussion and extraction of gestures*

The discussion of each gesture to be extracted follows. The gestures resulting from the discussion are found in Figure 6.

1-2. *Menu*. The link between the drink card and *menu* failed in terms of the gestures used to ask for the menu. Only one testee saw the menu in Scenario C as a kind of sheet, like in the pub scenario. The general consensus was that the menu "popped up", hence illustrating this popping up into the air above the workspace. The interface can be implemented with a menu zone above the workspace. However, in some applications with e.g. 3D objects there may not be space for this. Those applications would benefit from a static menu gesture such as the one where the palm is iconic for a sheet of paper.

3. *Selection* was an obvious deictic pointing gesture in both scenarios, with a few other examples in the pub scenario. The pointing gesture is either static or moving along the line of sight between the testee and the object. However, the information is the same, whether it is static or dynamic. The biomechanical aspect is important to note here because if the palm is faced down in this pointing gesture, then the wrist is pronated to its outer position. If a user uses the application for an hour or so, this will lead to excessive stress. Furthermore, the pronation will affect the range of motion available for further finger postures. Therefore, the wrist should be kept neutral, with the right palm facing left, and vice versa.

4. *Inserting* objects was made simple by using a separate template zone and workspace zone. All testees took advantage of the template zone when inserting furniture with a "put-that-there" gesture. This usage of zones can be applied to the interface of the architectural application as well.

5. *Select All*. There are many variations of *select all*, but they are very similar. There are one handed and two handed versions, and they are all dynamic, illustrating some kind of covering piling up or claiming everything is one's own. To ease the recognition algorithm a two handed gesture can be chosen, but if a one handed gesture is chosen, an open palm with spread fingers, circling once over the work space is a simple solution. As this function is not frequently used, it is accepted that the wrists can be pronated almost to the outer position, so that the palm faces downwards.

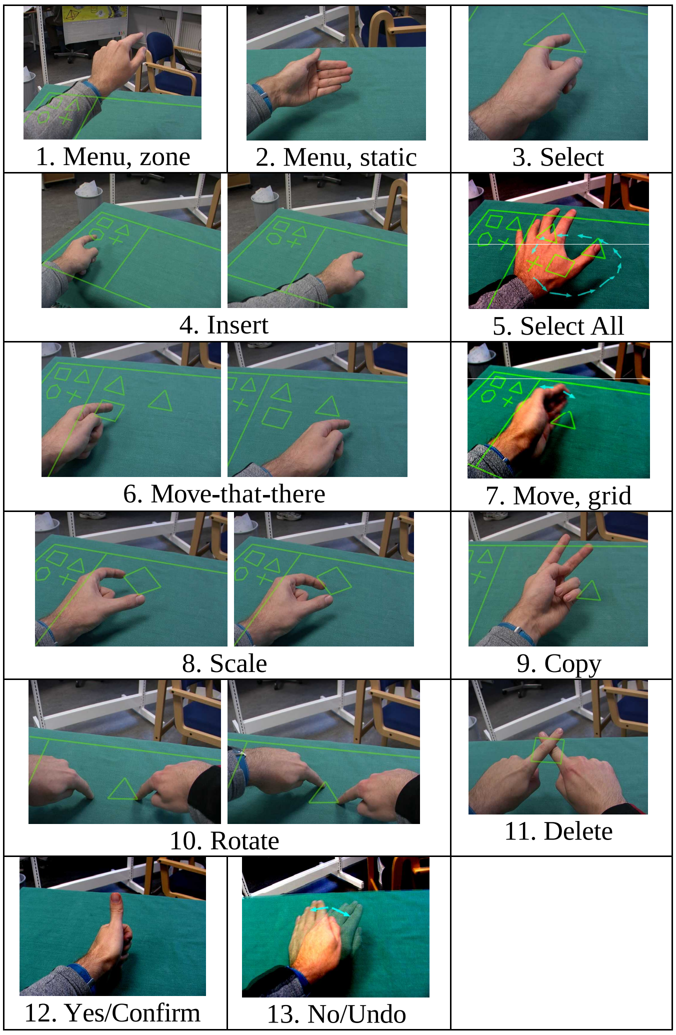

Figure 6. Resulting gestures. The small light arrows indicate movement.

6-7 *Move*. 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 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 palm is not mistaken for “select all” or “no/undo”, because the palm faces sideways, instead of downwards.

8. *Scale* gestures depended on context, especially in the pub scenario. In the furnishing there were some useful gestures: Two people used the index finger and thumb as a measuring device to indicate the shrinkage or growth of the objects. Two people chose to use two hands to show the growth or shrinkage, but did not show the desired size and thus relied on fixed size steps. Therefore, the one handed gesture is chosen.

9. *Copy*. There was also consensus in the *copy* gestures. The testees illustrated that they wanted more than one. Most testees preferred to do a multiple copy function, by showing how many copies they wanted, instead of copying one at a time. However, it is still the same gesture. The “two” gesture is therefore chosen. Biomechanics need consideration in this gesture as well. The common gesture, where the palm faces away from the user, has a pronated forearm, which in a natural environment is not a problem, because the entire forearm is lifted up, and they are not used very often. However, in this interface, the hand should be kept down in the workspace while doing the gesture, and the frequency is considerate. If the gesture is performed with a pronated forearm, there is a risk that the wrist is further extended, and the middle finger constrains the ring finger from flexing. Therefore, it is important that the gesture uses a supinated forearm, which leads the palm to face the user.

10. *Rotation* is also found in one-handed and two-handed versions. They generally show the rotation with either the thumb and index finger, two index fingers, or the entire hand. As the thumb an index finger is a lot like the chosen scale gesture, and using it for scaling is easier to implement, because the angle of view is better for the recognition, the two-handed version is chosen for rotation. That way it is easy to control the desired angle. Furthermore, one hand can show the point of the rotation, while the other rotates around that finger.

11. *Deleting* objects are mostly done using the template zone as a dump zone with “put-that-there” gestures. This is a dangerous implementation of deletion, because if a user selects an object and then decides to insert a new object, then the selected object disappears. It is important that the chosen gesture does not occur unwittingly. The alternatives are wave hand over object, but this is reserved for the undo gesture. The index finger fencing is too stressing, but if it is converted to a static gesture, the result is a crossing the fingers over the object. It can also be followed by an “are you sure”-query, as it should not occur on a regular basis.

12. *Yes/confirm.* All testees used thumb-up to confirm.

13. *No/undo* was mainly done by hand or finger waving. In order to avoid confusion between indecisive

selection finger pointing with undo, the extended hand waving is chosen. In order to spare the wrist extension, this gesture should not be as uptight as "halt!" gesturing with the wrist extended almost to the outer position, but with a relaxed wrist position.

4.4. Step D. Benchmark

The benchmark was done with 8 testees, 7 software engineers and one occupational therapist.

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 benchmark are found in Table 5.

Table 5. Benchmark scores. Range shows the observed minimum and maximum scores.

	Average Score	Range	Variance
Test 1 – Semantics	0.10	0.0-0.33	0.02
Test 2 – Memory	2.13	0.0-6.0	3.55
Test 3 – Stress	2.5	1.0-3.0	0.57

In test 1 one testee got 3 errors and one got 4 errors. The problems were mainly those that were expected, but "Move, using grid" and "no/undo" got mixed up as well.

Test 2 showed surprisingly that it was generally easier to remember this large vocabulary. One extreme result was 6 retries. This testee was very affected by the time pressure and started laughing during the test. Calculating mean and variance without this outlier gives a mean of 1.57 and a variance of 1.29. This is very satisfactory.

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 found in Table 6.

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.

Table 6. Individual Gesture Stress scores. Highlighted results are commented in the text.

Gesture	Average Stress 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***

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. DISCUSSION

Two approaches, technology-based and human-based, to developing gesture interfaces are presented.

The technology-based approach is technically solvable, but 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 very time consuming, and the scenarios must be carefully written.

The 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.

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.

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 aims for the results, 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 how this will be solved stands.

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 and in the image analysis field in the specific cases where it has been applied. This is the aim of the current FG-Net project.

6. 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.

Thanks to the participants of the experiments.

7. 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] M. Stoerring, E. Granum, T. Moeslund, “A Natural Interface to a Virtual Environment through Computer Vision-estimated Pointing Gestures”, *Workshop on Gesture and Sign Language based HCI, London, UK*, 2001

[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 Univeristy, 1999

[21] Wolfgang Dzida and Regine Freitag, “Making Use of Scenarios for Validating Analysis and Design”, *IEEE*, 1998