Project Number:	IST-1999-11577
Project Title:	Fitness-for-Purpose of Person-Person Communication Technologies
Short Title:	Eye-2-Eye
Deliverable Type:	Public

Volume 2 of 2: The Evaluation Methodology as a stand-alone resource

CEC Deliverable Number: IST11577/SEF/DIS/DS/Pub/013/b1

Contractual Date of Delivery to the CEC: 28 February 2003

Actual Date of Delivery to the CEC: 19 March 2003

Title of Deliverable: Fitness for Purpose Evaluation Methodology

Workpackage contributing to the Deliverable: WP 1

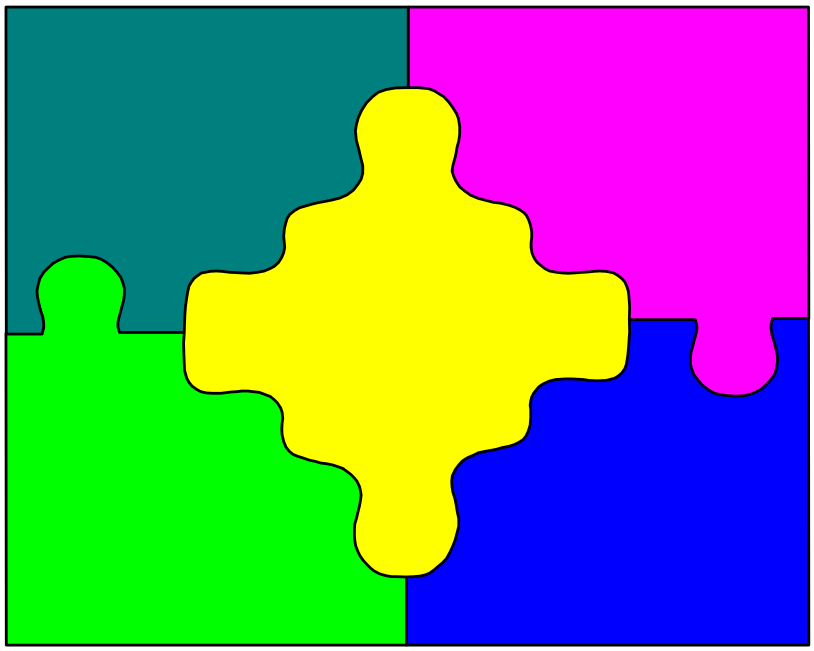
Nature of the Deliverable: Report

Author(s):

SINTEF	Jan Heim, Peter Brooks, Asbjørn Følstad, Trond Schliemann
University of Nottingham	Claire O'Malley, Patrick Brundell
Telenor	Bjørn Hestnes, Svein Heiestad, Trond Ulseth
IvD	Han Frowein, Paul Devoldere
TANDBERG	Carl Aaby

Abstract: This report describes the process and development of the Eye-2-Eye Evaluation Methodology (EM) (Volume 1) and provides the actual EM as a stand-alone resource (Volume 2). The process included extracting relevant key data and information from the Eye-2-Eye empirical studies, feedback from external workshops with increasingly mature versions of the EM, and performing internal and external case studies. The EM is a collection of: guidance on research design, checklists, data collection tools, forms and templates, and reference data for designing and performing empirical studies on real-time person-person communication.

Keyword list:
Telephony, audio conferencing, avatar telephony, videoconferencing, multimedia conferencing, face-to-face, quality of service, fitness-for-purpose, Human Factors, baseline experiment, laboratory experiment, field study, focus group study, evaluation methodology, toolkit



Fitness for Purpose

Evaluation Methodology

A toolkit from the Eye-2-Eye Consortium

EC IST Project IST-1999-11577
Main Contact: Jan Heim, SINTEF

Jan.Heim@sintef.no

Content

1. Introduction

1.1 The Evaluation Methodology - What is it?

The Evaluation Methodology (EM) is a collection of:

- Guidance on research design
- Data collection instruments
- Checklists
- Forms and templates
- Reference data from the Eye-2-Eye project

..specially tailored for real-time person-person communication.

This document is a result of the Eye-2-Eye project (IST-1999-11577). The project has applied laboratory and field studies in order to assess the "fitness-for-purpose" of different communicative media and services. Fitness-for-purpose is formally defined as:

> *"The correct balance between technological performance and human performance, such that the interaction is both sufficient and beneficial for person-person communication and consistent with human expectations from face-to-face communication."* (Brooks, P. et al, 1999).

A practical approach is to ask: *What kind of media and services is best suited for a given communication activity?* The Eye-to-Eye project has developed a collection of Fitness-for-Purpose Guidelines, based on the results from laboratory and field studies, as well as a Cost-Benefit Analysis tool[1]. These tools would generally be a good starting point for answering a question as posed above.

As empirical studies are time consuming and quite expensive, they should be considered only when convinced that other approaches are not feasible for finding the required answer. Alternative approaches include focus group studies and interviews with target users (described in section 3).

If it is decided to do an empirical study, one should decide if the proper approach is *laboratory or field* studies. Other methodological approaches (like user-surveys, case studies etc) are not covered in this document. Rationales for making these general decisions are described in section 2.

1.2 Who is this document for?

The audience for the EM is developers and researchers within the field of communication technology. Some prior experience in conducting behavioural studies is an advantage; this document might bee seen as an "add on" to standard textbooks in the field (e.g. Wiklund, 1994). The general approach is based on the "Human factors" tradition and social and experimental psychology. This document focuses on the

[1] Both tools will be publicly available during March 2003

issues that are unique for studies where two persons are communicating by means of communication technology. It is not a general-purpose methodology for conducting behavioural studies.

1.3 When do you need this document?

You should use this document:

- if you want to conduct empirical communication studies
- if your main focus is person-to-person communication,
- if you need advice and guidance for designing and carrying out communication studies.
- if your main focus is usability of the user interface of communication technology involving two participants communicating in real-time

1.4 When do you not need this document?

You should look for information elsewhere:

- if your problem can be studied without doing empirical studies,
- if your problem can be studied with only one participant at the time, including "man-machine" user interface studies
- if you need advise on statistical issues,
- if you need advice on conducting surveys.

1.5 EM's application area, general benefits and advantages

The type of situation for which this document is designed, is when two persons are communicating by means of some communication technology. Although this toolkit has been developed for dyadic communicating, it is likely that several of the tools presented here may be applied with minor modifications to situations were more than two persons are communicating[2].

The main purpose of this document is that the potential user should not have to "invent the wheel" over again. There are forms, templates and guides included that are intended to be used without much tailoring to the specific project. This should provide means for keeping the budget as low as possible, and to ensure higher quality of the process. In many cases the results will also have a higher value because they can be compared with the provided reference data describing results from similar investigations.

The tools described here can support projects involved both in laboratory and field studies. The term "laboratory" does not necessarily means a fully equipped space with all technical laboratory facilities. Often two offices, with the necessary connections for communication between them, are sufficient. Field studies are usually less concrete and are more diverse than laboratory studies. Therefore the material provided

[2] A selection of tools from this document has in November 2002 been modified and applied to the IST project "Virtue", with the main modification of addressing 3-way communication

for laboratory studies are much more of the "how to do it" nature than the hints and recommendations given on field studies.

Performing empirical studies normally demands substantial resources in terms of labour and equipment. It is therefore advised that a budget is worked out in advance and that a cost-benefit perspective is applied when deciding whether to do empirical studies or not. A team rather than a single person should in most cases carry out the study. A multi-disciplinary team is a great advantage, especially if it has both technical competence and competence within human factors or neighbouring disciplines.

1.6 How to use this document.

If you are in the very beginning of your project it is recommended to read through all of chapter 2. It may help to take the necessary steps towards defining the problem, designing the study, defining the user population, defining the tasks, and to decide on test conditions and what measures to aim for in the study.

After having decided what to measure in the study, chapter 3 will provide some general advice on the use of *different data collection methods*. Some different qualitative methods are described as possible supplements to quantitative measures.

Chapter 4 describes and gives advice on necessary *technical equipment and laboratory facilities* for conducting laboratory studies.

When the high-level decisions are made, it is important to *plan the data collection process* in detail. Unless one is highly skilled in this field, there is no room for improvising after the actual data collection has started. Chapter 4 focuses on the concrete aspects of doing studies of person-to-person communication technology. Here you will find advice and guidance for preparing a laboratory study.

Chapter 6 gives a high level description and practical advice for *planning and conducting field studies*. Conducting field studies is a rather complex matter, so it is recommended to use this chapter only as a supplement to more comprehensive material - for example (Bickman & Rog, 1998)[3].

The description and advice on how to perform a laboratory study are found in chapter 4. Chapter 7 presents, in addition, the actual *data collection instruments* used in the Eye-to-Eye project, along with a set of *reference data* for comparison with own findings.

The *tasks and scenarios* needed to perform the study must in most cases be designed specifically for each project. The Appendix gives examples of tasks that can be used, and may serve as examples on how to write tasks and scenarios.

1.7 Related material

This document has no intention of being an extensive handbook on designing empirical communication studies. There already exist several excellent web sites,

[3] Note that the term "Field studies" used in this document does not correspond to Bickman & Rog's term "Field Work" , but rather to their concept of "Case studies"

handbooks and tool-collections covering relevant methods and evaluation techniques. The Eye-to-Eye Evaluation Methodology is designed to be complementary to these existing materials by specifically addressing person-person relations in real-time communication

Related reference material includes:

- ETSI EG 201 472 (2000) "Usability evaluation for the design of telecommunication systems, services and terminals"

This is an extensive documentation of standardised methodology for performing usability evaluations. It builds on ISO standards 13407 (Human-Centred Design Process), and reviews these standards and their application for telecommunication issues.

It presents a wide scope of relevant methods and techniques in detail, and will supply necessary background knowledge for the EM to a large extent.

- ITU-T P.910 (1996) "Subjective video quality assessment methods for multimedia applications"

This ITU Recommendation documents non-interactive subjective assessment methods for evaluating the quality of video images. The document describes a laboratory set-up to product test sequences and carry out subjective assessment, characteristics of test sequences, test methods and experimental designs and analysis of data. Particular consideration is given to the use of different grade quality rating scales, ranging from 5-grade to 11-grade scales.

The Recommendation is intended for applications such as videotelephony, videoconferencing and storage and retrieval when it is considered appropriate to exclude real-time interaction in the test task. All of the Eye-2-Eye evaluations are based on real-time interaction between two test participants.

- ITU-T P.920 (1996) "Interactive test methods for audio-visual communications"

This ITU Recommendation proposes interactive evaluation methods for quantifying the impact of coding artefacts and transmission delay on point-to-point or multipoint audio-visual communications. The document describes the methods and experimental design and appropriate questionnaires. The methodology is based on conversation opinion tests and a number of conversation tasks are proposed which are relatively simple and efficient in obtaining ratings of quality. The different tasks are relatively artificial laboratory-based procedures designed primarily for measuring the effects on ratings of quality of speech delay, audio-visual delay and synchronisation between audio and video signals.

The Eye-2-Eye approach differs primarily in that the tasks are less artificial and designed to produce data on person-perception, social presence, task outcome, communicative behaviour and cost-benefit.

- ITU-T P.910 (1996). "Subjective video quality assessment methods for multimedia applications"

Focuses on audiovisual quality measurements of non-interactive test persons. Therefore has limited application to real-time person-person communication evaluation unless it is considered valid that the test person is not interactive with another person. Provides recommendations on different subjective scales, such as both 9 and 11-point numerical quality scales.

- ETNA Project Report (2001) ”New techniques for assessing audio and video quality in real-time interactive communications” IHM-HCI 2001, Lille, France.

The ETNA project produced a taxonomy of real-time multimedia tasks and applications and used a combination of field trials and controlled experiments to investigate the audio and video quality requirements for a number of these tasks. The report discusses limitations of the ITU Recommended Scales and suggests approaches to collecting opinion, task and physiological measure from users. It provides an approach to assessing communication tasks and contexts and an Evaluation Checklist.

The ETNA taxonomy and classification of methodology complements the Eye-2-Eye Evaluation Toolkit and provides additional background material on choice of methods.

- USINACTS (1998): "Usability in the Information Society"

Human factors advise produced as a freely available CD-ROM and website http://www.cordis.lu/infowin/acts/rus/projects/ac224.htm. Offers a tutorial on "How to design user-friendly products and services in the Information Society". It complements the Eye-2-Eye EM by being mainly concerned with person-device/service interaction rather than person-person communication. It was an important source for ETSI EG 201 472.

- ACTS II Guidelines G4 (1996) "Organization of advanced communication services trials with residential users". Brussels: European Commission. ACTS Research Program

Developed during the EC 4th Framework Programme, this EC guideline gives advice on field trials with communication services. This extends to more examples of field studies, particularly tests of service concept and real use in uncontrolled situations.

- www.UsabilityNet.org

This web-site presents a lot of resources for professional usability work. It includes an easy-to-understand presentation of most known methods and tools for user testing and

evaluation, as well as a set of case studies and related materials, exemplifying the use and effects of using usability methods.

The web-site also gives an overview of courses, conferences and events, books, design guidelines and other things related to usability work in the laboratory and field.

The web-site is excellent as easy accessible reference materials for non-experts in human factors work.

2. Designing real-time communication studies

Performing empirical studies is expensive. They normally need to run for quite a time and substantial material resources and manpower are needed to complete a study. It is therefore necessary to be quite clear on the needs for the study.

The first step is to make some explicit statements about what the problem is. What do I need to know? Then one should look for alternative ways to find the knowledge. Are there guidelines, standards, or scientific papers that can provide what you need? Only if there is not, one should consider doing empirical studies.

2.1 Defining the problem

In the context of studying communication technology, a problem definition will typically have the form of:

What is the **<appropriate> <communication service>** for a given **<communicative situation>**?

A further definition of the central terms in this statement is necessary:

appropriate: In this case, the term relates to the "fitness-for-purpose" for the situation of interest. Typically it will have the form of "What is best", "What is sufficient" , "Is it possible" , "What is acceptable" etc. Appropriateness in this context will be in relation to a dependent measure in the study; i.e. what is measured.

communication service: This is defined in terms of the communicative media and their technical configurations. Unless there is available a set of data that is directly comparable to the data in the present study, more than one technology should be assessed. It may be meaningful to study only one technology if the purpose is to get feedback on the design in order to improve a solution. However, if the interest is in whether something is "better" or "more suitable" than something else, one needs a comparison. (See section 3 for examples of tools and typical outcome)

communicative situation: The situation may be described in terms of the users goals and intentions. For an actual study, the situations should be described as tasks and contexts users engage in (see 2.2). The more general the description, the more difficult the process of defining a task will be.

In general we are interested in what people do when they communicate, what they prefer, their choices and how they perceive each other. Section 2.4 treats this question in detail.

Examples:
What is sufficient bandwidth for video telephony when two people need to discuss a budget and it is critical that they reach an agreement within a time limit?

Is ordinary desktop phone more efficient than instant text messages for exchanging factual information between work colleagues?

What is acceptable sound quality for mobile applications when giving remote advice in noisy environments?

2.2 Defining the tasks and scenarios

For doing empirical studies, it is necessary to describe the central concepts in the problem statement as concrete as possible. It is necessary to decide exactly how the concepts will be reflected in the empirical study in terms of the communication technology, the tasks and the measurements to be applied.

One of the most important issues is the <situation>. The tasks and scenarios that are chosen to study should be one or more examples of the <situation> that are defined in the problem statement. A task may in general be identified by the purpose of the conversation. The situational context, in which the task is embedded, is also an important aspect of the task definition[4].

It is important that the tasks and scenarios under study is *representative* for all the possible manifestations of the <situation>. In doing laboratory studies, (see section 4) the challenge is to recreate or simulate most of the important aspects of the general <situation> mentioned in the problem statement. In order to design representative tasks, it might be a good idea to assess the "target task" according to some salient task dimensions. McGrath (1993) offers a frequently sited classification system. Another approach is to let the "problem space" of the project define the salient feature of the tasks.

Some of the task dimensions that might be salient in communication studies are:

- Degree of familiarity between the interlocutors
- Whether the communication as such is the goal in itself, or if there is some explicit result to be achieved
- Formality of the situation
- Whether the interaction involves exchange of factual information
- The extent of exchange of opinions and attitudes
- Whether the task has only one or several plausible outcomes
- Symmetry: Whether the roles of the interlocutors are the same or different
- The degree of co-operation necessary to solve the task
- Whether trust, deception or persuasion are essential aspects of the communication
- The costs of face-to-face meetings, in terms of resources, efforts and personal investment
- The physical context; mobile, indoors - outdoors, at home, at work, alone or with others etc.
- The number of simultaneous interlocutors
- The number of potential interlocutors.

A reasonable strategy for defining tasks is then first to describe the "real life" situations that the technology shall support. Then to analyse what is the salient dimensions of the situations, and then to construct feasible tasks that reflects the "real life" situations as close as possible with respect to the task dimensions.

[4] In this document the term 'task' will be used also for activities like small talk, chatting and activities without a clear extrinsic goal

Be aware that there are no "general purpose" tasks, although there might be "general purpose" communication technology. A task that is used for the purpose of studying communication must be *explicit* in terms of the instruction given to the participants. There must be no doubt about the purpose of the interaction (unless uncertainty is a salient task dimension), and there must be an explicit procedure for collecting the results of the interaction.

Examples of tasks used in the Eye-2-Eye project are described, some with detailed participant instructions, in the appendices at the end of this document.

2.3 Define the user population and sample of participants

The <communicative situation> part of the problem formulation will be the starting point for defining the user population. It is important to be clear about for whom the communicative service should be designed. Is it for example the "general population" or "business managers"? This has to be clearly stated because the choice of participants for the study depends on this statement.

Although one might be rather ambitious about the range of the "target group" for a product, it is usually very hard to come up with a sample of participants that is "representative" in a statistical sense. A representative sample should be a true "small-scale" model of the population with respect to the distribution of all the important properties of the users[5]. A few essential principles apply:

- first decide the group of interest in the population
- then find essential characteristics that affect your task
- establish quotas of users to have representativeness of them in your user group

A practical and much used strategy, both in laboratory and field studies, is to identify a group of participants that is both interesting and available. This could for example be students at a university, staff members of a co-operating company, clients of a health- and social care centre and so on. Then one has to assess on a rational basis how representative the results of the study are for populations not actually represented in the sample. Although this is a practical approach it is clearly deviating from the standard methodological ideal of "representativity". It is therefore important to be explicit on the following questions:

- Who is the target population for the communicative service?
- How is the sample of participants defined?
- How to define the population that the participants are representative for?
- What is the difference between the target population and the population actually studied?

2.4 Dependent Measures

There are several methods for studying how people perceive different aspects of a communication services, for example the sound quality, the clarity of the image, the smoothness of the motion and so on. The ETNA project (http://www-mice.cs.ucl.ac.uk/multimedia/projects/etna/) has conducted research on establishing

[5] General books in statistics can advise on different means for drawing a representative sample from a defined population

audio-visual quality requirements for a range of real-time multimedia applications. There is, however, little applicable knowledge in the human factors area as to the specific audio-visual requirements for different kinds of communicative tasks. The perceptual boundaries for audio and video quality in multimedia communications have not been systematically investigated, and are likely to differ according to a specific task and the characteristics of the group of users undertaking this task. The present set of methods and techniques will primarily focus on users' attitudes and behaviour in communication situations, the reader is referred to the ETNA project (see reference above), as well as other related materials (1.7) if the study of perceptual qualities is the main interest.

Within general Social Psychology it is widely recognised that the relationship between people's attitudes and their actual behaviour towards the same issues are often very weak. What people say they will do, or say they prefer, is not always a good predictor of what they actually will do in a given situation. That is why one has to be specific about what type of user response is actually being measured. If the perspective is a marketing one, it would probably be of interest to study people's preferences, if the perspective is business management one would probably be more interested in users' task performance, efficiency, etc.

When assessing communication technology, two of the most important questions is what users prefer and what they actually use. The questions on preference may be general ("Would you like to have a Videophone rather than an ordinary phone"?) or directed towards a specific task or situation ("Would you use an SMS-message to inform your parents that you are staying the night with a friend")? In both cases the questions are hypothetical, and may not give appropriate answers to the question of the kind of technology or service that would actually be purchased or used in a given situation. Even when people are asked very specific questions of hypothetical situations they tend to answer in accordance with their general views ("Seeing the person is always better than audio only"), rather than in accordance with their actual behaviour. (See Schliemann et.al (2001) for an empirical demonstration of this point).

When assessing well-known technology, respondents can give adequate answers regarding their actual use, ("Approximately how many mobile calls did you make yesterday"?) but this is not an option with communication services that are completely new. However, after having tried out a new service in a more or less real life context, users can express what they feel are the advantages and drawbacks with the service for this particular situation. This is referred to as a *subjective assessment*. Measuring the actual *task performance* (for example how effectively a given problem was solved) and the measurement of *communicative behaviour* (length of utterances, number of interruptions etc.) is, in this context, referred to as *objective performance*[6].

It is a general finding that non-trivial performance differences between media and medium conditions are rather scarce in the research literature (Heim, J. et. al., 2001), and they may be hard to establish empirically. For example a large number of participants may be needed in order to establish small, but significant effects. It is therefore recommended to consider the inclusion of subjective assessment in studies

[6] This differs slightly from the definition given in the ETNA project and should not be confused with assessment of objective technological parameters as bandwidth, frame rate, etc.

of communication technology, as they are rather easy and cheap to collect and are often more sensitive to the medium than actual task outcome. Chapter 7 presents a selected set of data collection instruments for subjective measures, used in the Eye-to-Eye project, along with a set of reference data that could be used for comparison and validation of own findings.

So, the best source of information to support decisions in the field of communicative studies is based on knowledge of what people actually do. Laboratory studies are well suited for assessing user behaviour, but if possible one should also study behaviour in "real life", for example by logging frequency of use (see Følstad, A. et. al. (2002) for an example). The ideal approach for assessing new communication media with respect to users' preference and actual use is to do real-life field studies of large numbers of users over long time periods. Due to limited resources, this is usually not an option.

For all kind of measurements there will always be the important issue of generalisability - how far may an empirical finding be generalised to other contexts of use or other user groups? A possible way to address this question is to conduct focus groups after concluding on your empirical studies. The main results are presented and discussed in a comparable user group, in order to verify or extend the empirical findings.

2.4.1Task outcome

This is usually the main measure of user behaviour in a laboratory setting. It presupposes that the users are given an explicit task, and that the instruction for the task is embedded in a scenario or a "cover story" to help the users to engage in the activities. There are different kinds of measures:

- How well was the problem solved?
  - expressed as a percentage ("Task solved to a 75% degree") or
  - How many task sub-goals were achieved?
- How many errors were encountered?
- How long time did it take?
- Task specific measures (degree of trust, co-operation, antagonism etc.)
- How is the outcome distributed between the two interlocutors?

Composite measures of these may give indications of for example the efficiency and error rates with which the task was solved.

If comparison between two or more technical solutions is the main interest, task completion and time may be very useful. Be aware, however, that measuring time in some cases may result in trivial findings; we know for example that it takes longer time to write than to talk.

While time and error measures may be common to most tasks, the other measures must be designed as an integral component of the specific task under study. When task outcome is measured, an explicit procedure is required for registration of the users' behaviour. Be aware that the sensitivity of the dependent measure may be of crucial importance to the successes of the study. If the measure were dichotomous, for

example, a rough estimate of collaboration vs. non-collaboration one would expect that many pairs of participants must be studied in order for a systematic tendency to reveal itself. Usually it is better to have measures that aim to determine the degree of the dependent variable, in this case the degree of co-operation.

2.4.2Medium choice

In real life and field studies (section 6) medium choice may be studied by logging frequency of use. This is best done by some non-intrusive technical device, but it is also possible to ask the users to tabulate manually all incoming and outgoing calls. Medium choice studies are difficult to carry out in a laboratory setting. Since the measure is one of frequency it would either require that the same users participate over a long period of time or one would have to involve several users. A few choices in favour of one or another technical solution would probably not be sufficient to make any important decisions.

A typical set-up for studying medium choice is to let users have available two or more communication services, and then let them decide which service to use. One should be aware that such choices would be influenced by several factors. One important factor is cost. In some contexts (e.g. young people communicating with friends) cost may be the main determinant, while in a business context other factors may be more important.

A careful examination of the < communication technology > part of the problem statement is necessary in order to know which factors should be kept constant, and which should be varied between the possible choices. for example, if the main interest is in screen resolution, and the participants may choose between high and low resolution, it must be decided whether the participants should also experience the natural trade-offs involving signal delay and costs.

The overhead of setting up the communication (involving physical movements to an appropriate location, and the user interface of the service technology) is a very important factor when users choose a medium. Therefore one should always decide whether set-up costs should vary between the services (using a “typical” set-up configuration), or being held constant (the same amount of effort involved with all services).

Communicative pattern

In real-life and field studies it might be of interest to know who talks to whom? One might then compare this information with information on medium choice (see above). This might again shed light on the feasibility of the medium for different communicative situations.

2.4.3Communicative behaviour

With respect to objective measures of performance it is just as important to measure effects on participants’ communicative behaviour as it is to measure task outcome. There are two main reasons for this: One is that if there are changes in task outcomes as a result of using media of different characteristics, then it is important to know how and why tasks were affected, especially if one intends to make improvements to the technology. For example, participants may do “worse” in terms of task outcome

because the technology made it difficult for them to manage their turn-taking (e.g., because of delays), or because they mis-heard what the other said. The second reason is that there may be no difference in task outcome but users may have a more difficult time communicating to perform the task and they may have to change their communication strategies. Although this may not affect their task outcomes it may reduce their willingness to use the technology in future.

Analysing communicative behaviour can range from fairly straightforward measures such as numbers of words, turns and interruptions, to more complex measures such as the functions of particular utterances (e.g., certain types of questions) in achieving goals.

In order to provide an objective record of the conversation, normally the dialogue is first transcribed. This may be more or less detailed depending on what kinds of measures are required. For example, for some purposes it may be important to know the duration of pauses in speech, or the literal transcription of particular sounds that are not part of words (as in attempts to interrupt). There are standard conventions in fields such as psycholinguistics for coding linguistic and paralinguistic expressions, or the researcher may wish to develop his or her own conventions. Often, the addition codes to categorise certain spoken or non-verbal events in a dialogue requires a second pass once the transcript has been generated.

2.4.4Preference

In preference studies the users usually assess more than one communication service or more than one version of the same technology. Typically, they are presented with one or more scenarios, and then given one or several rating scales to assess how suitable, usable, enjoyable, adequate etc. the different communication services are in the situation.

If the study is based on hypothetical scenarios, preference studies may be a good indication of respondents' *general* preferences, and an indication of what is seen as commonly preferred in the population. Even a few respondents' answers can be a good reflection of the common opinion in the general population. Even if preference studies with hypothetical scenarios has shown to be not very sensitive to group differences (Schliemann, T. et al., 2001), such a study may be appropriate for indicating users' general attitudes towards a new service.

One would expect that the users' preferences are much more precise if they have actually used the service in a particular situation over some time, rather than to assess services used in hypothetical situations. This may be accomplished in field studies (section 6). However, if more than one service is available at the same time in a field study, observation of choice behaviour should be used instead of, or in addition to, preference ratings.

2.4.5Subjective assessments

Even though users' claims and speculations may be unreliable and not in accordance with what they would actually do in a given situation, "listen to the user" is often the easiest and most appropriate way to get information about what they feel and think about a particular question. Since it is difficult to get an overview of how several users have responded in an open-ended question, it is often more efficient to ask users

to fill in forms or make ratings on rating scales - a so-called 'structured approach'. Another advantage with this latter approach is that it allows for comparisons across different studies.

Examples:

The 'Person Perception questionnaire' (see 7.5) will measure one person's opinion and attitudes towards the communication partner after having performed a communication task together.

The 'Social presence questionnaire' (see 7.7) measures different aspects of the perceived presence of the communication partner in dyadic communication context.

2.5 Choosing the right paradigm: Experiments or Field studies

There are several factors that determine what kind of study to conduct.

Preferably, the nature of the problem should guide the choice of experiment or field study. The <situation> part of the problem statement might guide the choice of approach. Is it easily to build a laboratory environment that resembles the environment in real life? Are the tasks that can be given users in the lab, representative for the tasks in real life?

Most studies will in practice be dependent on the available resources as one important factor. Are the necessary space, equipment and participants for doing a laboratory study available? Is there an adequate pool of users that have the appropriate equipment in order to do field studies? What kind of personnel and economical resources are available for the study?

These are the types of questions that will influence the choice of a laboratory or field study.

In general, an "ill-defined" task is best studied in a field setting, while a "well-defined" task is suitable for laboratory studies. An ill-defined task is one that has no one "correct" solution, and may often better be described as an "activity" rather than a "task". A well-defined task has an identifiable correct solution, against which one can measure actual user performance. The number of errors that users make before reaching a correct solution or the time needed to reach the solution are adequate measures for well defined tasks.

Most tasks fall somewhere in between these two descriptions. Often, communication situations motivated by social or emotional issues will be of an "ill-defined" nature. If one is studying the nature of grandparent-grandchildren communication, it would probably be best done in a field study rather than a laboratory. If one is studying joint problem solving of simple tasks, it may be appropriate to do a laboratory study. If the problems are complex and normally need to be solved over several days, a field study is most appropriate.

A field setting may give information about what kind of technology that is used for talking to different persons. This may be difficult to simulate in a laboratory, as it usually only will support the studying of participants in pairs.

A laboratory study requires full control of the test context and test parameters. This provides a high internal validity to the results. On the other hand, a laboratory study with detailed control of "all possible" parameters, will have low external validity, as "natural" context variations are eliminated. A field study will, on the contrary, give high external validity and low internal validity.

The following table summarises the different characteristics of a field study vs. a laboratory study, with respect to task type, objective measures and subjective measures.

	Field studies	**Laboratory studies**
Task	Ill-defined Complex	Well-defined Simple
Validity	High external Low internal	Low external High internal
Objective measures	Medium choice frequency Communicative pattern Duration	Task Outcome Efficiency and Effectiveness Communicative behaviour
Subjective measures	Medium Preference Person Perception Experienced Costs and Benefits Social presence	

Table 1: Characteristics of laboratory - vs. field studies

2.6 Important issues when designing with several measures and conditions

The different technologies under study will normally define the conditions in the study. If one is interested in the interaction between technologies and tasks (what technology is suitable for which task?) different tasks must be included. The different tasks thus represent the different conditions.

In choice and preference studies each participant should be exposed to all the technologies in the study. Their choices and/or preferences are the dependent variables in the study.

It is often very efficient to let the same participants respond to several questionnaires and rating scales in the same session. There is, however, a possibility that the sheer amount of questions may make data less valid, since the participants may get confused, tired or bored. When there are many dependent measures involved in a study, the possibility that some comparisons turn out to be "statistically significant" become higher, even if there are no "substantial" relations in the data. This is because there are random fluctuations in all measurements of human behaviour, so some measures will correlate every now and then just because of pure chance.

If the design involves different tasks and different technologies, one has a choice of letting each subject participate in one, some or all combinations of tasks and technologies. Unless there are experts on experimental design and appropriate inferential statistics in the evaluation team, one should carefully consider the following issues:

- Do not try to merge all interesting parameters into the same study. If several conditions are of interest it is better to perform a series of small independent studies, than one giant multi-factor study. A simple study only needs to have two conditions; the experimental condition and the control condition (e.g. the new technology and the traditional technology).

- The simpler you keep the design (fewer conditions in the same study), the simpler statistical treatment is required, and the easier interpretation of data and results.

- If the same participant engages in more than one condition, there may be carry-over effects from one condition to the other with respect to learning, warm-up and fatigue. Therefore it is easier to interpret the results from studies where one participant only engages in one condition (between-subject design). If each participant engages in several conditions (within-subject design) one has to rule out possible carry-over effects by systematic variation of the order that the participants engage in the different conditions. This is called repeated measurement designs, and requires statistical analysis by means of repeated measures ANOVA - which is more complex than for a simpler designs.

- If it is decided to have several (more than three) dependent measures in the study, it is advised to formulate some assumptions about what results shall be obtained before the data analysis. Then one can see whether the predictions were confirmed or not. If there are findings that are statistically significant but not predicted, it is advised to reformulate the basic assumptions and then try to reproduce the findings in a new study.

3. Collecting Qualitative data

3.1 Qualitative vs. Quantitative methods

The main difference between qualitative vs. quantitative methods, is the "richness" with which you approach your data collection. Qualitative methods generally provide a broad approach and potential rich information, but with less control, while quantitative methods provide accurate results, but with only a narrow focus on the problem of interest. Qualitative and quantitative methods refer to methodologies that produce qualitative and quantitative data respectively.

Quantitative data is sometimes referred to as "countable" data - i.e. empirical data of a discrete nature, which are easily compared to each other in magnitude (frequency, duration etc.). Qualitative data provides more context information and a more holistic view, typically collected through less structured methods like interviews, observations etc.

Deciding to collect qualitative or quantitative data might depend on several issues:

- the nature of the research problem
- when in the development process data collection is done
- availability of resources.

Early in the research process the research problem will often be of a type that can not be defined in distinct and objective terms. A good start might be to collect as much relevant information about the topic as possible without respect to a well-defined methodology. For example, aiming at a better understanding of peoples' attitude towards a new communication service could call for depth interviews of typical potential users on their opinions or attitudes, rather than ask for ratings on a set of attributes on a scale. Concept testing would typically be better done with qualitative methods.

Early in the development process of a new technology or service, it can often be useful to elicit people's general opinions and attitudes towards a given technology or service, or an organisation's general usage of the present communication services. In such cases qualitative methods are suitable to establish an 'overview' of the situation.

Availability of resources could also be the reason for the choice of data collection approach. A full-scale experimental approach is very time and resource consuming, while for instance a focus group can elicit a lot of information from a group of people in quite a short time.

3.2 Focus groups

A focus group is an informal assembly of users whose opinions are requested about a specific topic. The goal is to elicit perceptions, feelings, attitudes, and the ideas of the participants about the topic. A focus group will typically produce a pool of qualitative data.

A focus group usually involves 6-10 participants, in addition to one or two moderators. A focus group with few (2-3) participants can also be useful, but then the process will resemble more that of a focused interview or group interview. The moderator needs to have some appropriate training before taking on the responsibility of leading a focus group.

A vehicle for eliciting interesting discussions in focus groups is to let the participants try out and "play" with the communication technology under study. If the communication technology is not known in advance, hands-on demonstrations might be a good substitute for actual usage over time.

Focus groups may be very instrumental in giving information about different user groups' thoughts and attitudes about new and future communication media. One should carefully consider whether one wants to use homogeneous or mixed groups. Both the expected group dynamics and the question of representative samples need to be considered before deciding on the composition of the groups.

See http://www.usabilitynet.org/tools/focusgroups.htm or ETSI EG 201 472 (2000) for further details and references.

3.3 Interviews

The interview is a method for discovering facts and opinions held by potential users of the system or service being investigated. It is usually done by having one interviewer speaking to one informant at a time. Interviews need to be carefully analysed and targeted to ensure they make their appropriate impact. This will usually need more resources than doing the actual interviews themselves.

A 'structured interview' is based on the interviewer following a well planned schedule, with a list of questions that requires an answer by the interviewee, and with well planned prompts and explanations of these prompts in case the interviewee does not understand.

An 'open interview' would be more convenient for collecting general information on a certain topic. If the interviewer does not know exactly what information to look for (e.g. in early phases of requirement collection for a new product) but still wants the informant to share their general knowledge on a certain topic, an open interview would be appropriate. Planning an open interview would typically mean deciding the topic to talk about, and the interviewer will prompt concrete questions as the interview goes on. In this kind of interviews it is important that the interviewer has adequate training and a good general background knowledge of the topic.

As with focus groups, a realistic demonstration and trying out of the technology may be necessary if the user is not familiar with the technology in advance.

See http://www.usabilitynet.org/tools/interviews.htm or ETSI EG 201 472 (2000) for further details and references.

3.4 Questionnaires

Questionnaires are administered to the informants in order to get their individual response to a well-defined set of questions. The questionnaire would typically be

prepared as a 'very structured interview form', only with no interviewer present. This requires that the questionnaire is well founded around the actual research problem, and 'piloted' in advance, in order to make sure that the instructions and the questions posed are easily understood by the participants.

A questionnaire must be related to a certain context known to the informants. When designing a communication study, a questionnaire will typically be administered after a test trial - addressing different aspects of the user's experiences during the test.

Questionnaires could also be administered 'remotely' to a large group of respondents - either as common surveys with paper forms to fill in, or as web-surveys. This way you will reach a lot more people, but there are no opportunities to check that the respondents have understood the context of the questionnaire. Depending on the nature of the topic and the use of incentives to encourage people to answer, your answer rate could be very low, and hence your sample of respondents will not necessarily be representative to the user-group being addressed.

Broad surveys might be suitable to assess communication technology that has been on the market for a while.

See http://www.usabilitynet.org/tools/surveys.htm or ETSI EG 201 472 (2000) for further details and references.

Subjective Assessment Tools

An important success criterion of a service is how it affects the user's subjective experience (Section 2.4.5). Data on this can be accessed through focus groups and interviews, but would normally be collected with standardised questionnaires administered to a larger number of users. Performing user tests in a laboratory setting, is a good opportunity to collect subjective data by administering a questionnaire to the participants after having done their tasks. The questionnaire should then be related to the task performed by the participants, and it must be obvious to the participants how the questions relate to the task they have just performed.

4. Technical set-up for Laboratory studies

A critical factor for successful data collection is the technical and environmental context in which data shall be collected. For field studies, the context is partly given by the actual field setting, while in laboratory studies, the context is defined by laboratory facilities and available equipment.

This chapter focuses on the design and technical set-up of laboratory facilities for data collection in person-person communication. It is meant to be a high-level guide to the planning of your technical facilities. Each laboratory will have to be designed specifically to fit the local facilities and ambitions.

4.1 Laboratory Space

A ”traditional” usability laboratory will as a minimum consist of a test room and a observation room. The test room is where the test person, or user, performs the task to be investigated (often with a test leader present), while the observation room is where observers can monitor the activity in the test room - either via camera and monitors, or through a one-way mirror.

For person-person communication the situation is different. There are two users, which require two separate test rooms, and it must be possible to observe both rooms from one observation point. To estimate a proper *size* for test rooms designed for audiovisual communication, aspects to be considered are acoustics, monitor size and viewing distance.. Whether the rooms shall be designed for communications between two individuals or between groups is also crucial.

The laboratory space can either be built completely from new, or existing facilities could be made into a suitable space. In Eye-2-Eye, both approaches were used. At SINTEF in Oslo the laboratory was built from ground in an open room of approximately 6x6,5 m. Two separate test rooms were built, with one-way mirrors facing the control room. At the University of Nottingham the lab space was designed around already existing office facilities. The control room was placed between the two testrooms, with one-way mirrors on both sides. Figures 1 and 2 below show the general design of the laboratory space at SINTEF and Nottingham.

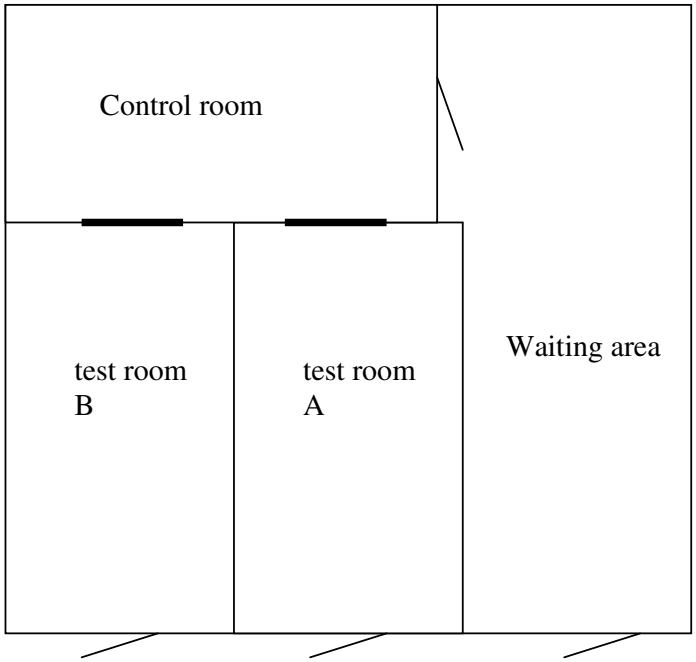


Figure 1: Laboratory design 1 at SINTEF

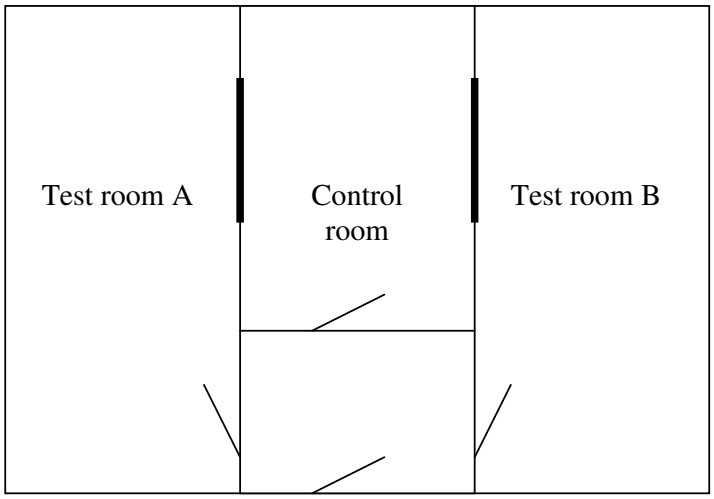


Figure 2: Laboratory design 2 at University of Nottingham

4.1.1Test rooms

The crucial characteristics of the test rooms is *sound insulation*. Generally the rooms must be insulated to avoid unnecessary noise from adjacent rooms. Specifically, the rooms have to be sound proof to the degree that natural sound will not be transmitted from one test room to the other. The one way mirror, ventilation system, the door and the bushing of cables very often turn out to be the weak points. This is essential to support the experience of the two persons communicating with each other from remote distance.

The sound insulation between the test rooms was specified to be at least 35 dB within the frequency range from 100 Hz to 8000 Hz (Heim et. al., 2000).

Acoustical room design is important to avoid a sound described as 'echoey', reverberant, or 'bottom of the barrel'.
If the room is designed for video, ambient *lighting* is also important. Only artificial, diffuse light above and in front of the participants is recommended. Avoid direct and strong reflected light in the camera's field of view. The background for the camera should be a non-reflective surface with no pattern and should provide good contrast to the user.

Another important factor is the *furnishing of the rooms*. The idea is to create a natural atmosphere for the test setting. If the test is focussing on a work context, provide a desk and general office artefact that resembles an office setting. If the test setting is domestic, provide furniture that resembles a home environment. Users should not be "reminded" more than necessary that they are in a laboratory setting. Green plants and pictures on the wall will contribute to a more "non-clinical" atmosphere.

Ventilation is also a critical factor - especially if the test rooms are small. Insufficient ventilation is likely to influence the test results if participants become tired, for example, and lose concentration. Make sure that the ventilation system does not cause noise in the room.

In-room equipment (e.g. PC's) will cause noise and also extra heating to the room. If possible, the noisy and heat-generating equipment should be placed outside the test rooms. At the SINTEF lab facilities, all computers were placed in the control room, with cable connections to the test rooms for plugging in mouse, keyboard and monitor.

4.1.2Control room

The test is controlled and monitored from the control room. This is were you want to keep the noisy equipment and all the apparatus that needs to be controlled or managed during the test.

You must be able to *monitor the user* from the control room - i.e. you need a full view of both users and their relevant interaction with whatever technology they are testing. A direct view via one-way mirrors is not absolutely necessary, but "nice-to-have". Normally, both sound and image should be transmitted from one test room to the other via the control room. The sound and image signals are extracted in the control room for recording purposes, and/or for viewing on a monitor.

The test leaders will need *immediate access* to all the necessary equipment for controlling and monitoring the test situation from their position in the test room. This includes connection set-up between test rooms, monitors, recording equipment, and an external phone. Separate lighting control for each room is also important. Ideally, it should be possible to operate all the necessary equipment without having to leave your seat.

The test leader will also need immediate access to the test persons in each room to give general information or individual instructions. In a real-time person-to-person set-up, ordinary intercom or telephone systems will cause problems. It will not be possible to give separate instructions to each of the test persons during the test. A push-to-talk system should therefore be integrated in the technical set-up.

Often, the control room will work as an *observation area* as well. Assigned test observers, as well as external observers could monitor the activity in the test rooms, either through the one-way mirrors, or via monitors.

Ventilation is also a critical factor in the control room. During the test period, the rooms could be in use for the whole day by the same persons. It is important too design a ventilation system with enough capacity. Remember to calculate for all the planned equipment in this room.

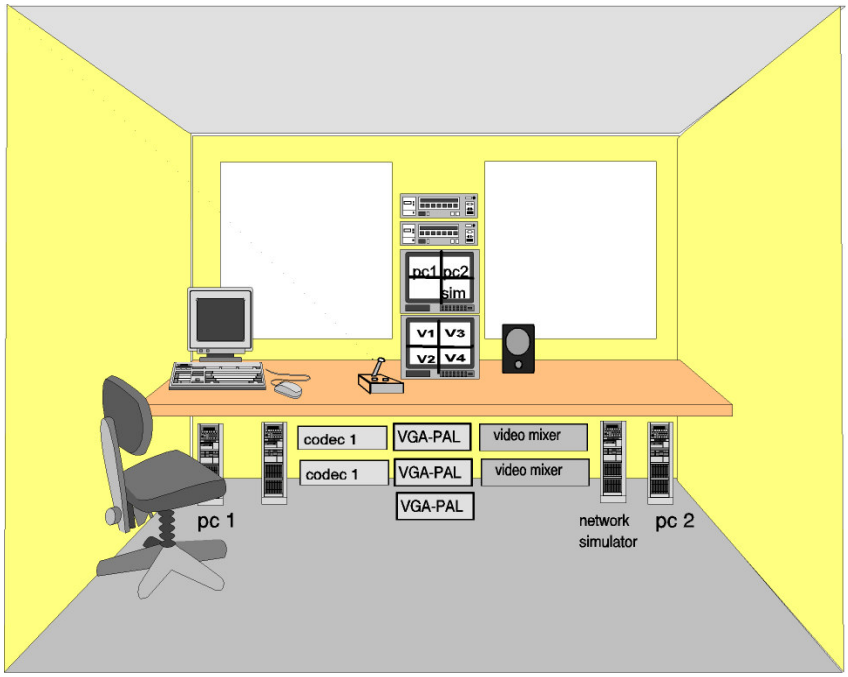


Figure 3: : Example of lay-out of the control room (lab design 1)

4.1.3Waiting area

The waiting area, or *welcoming area*, is primarily for creating a comfortable atmosphere for the test participants before the actual test starts. A couple of comfortable chairs and a table helps create a relaxed setting to the users for initial introduction to the test.

A main issue when testing person-person communication is *whether the two participants should meet each other* before the test. It could be the case that the user test involves evaluating "first time interaction", and hence a main point would be to avoid the participants seeing each other before actual tele-connection is made. If this is the case, make sure that participants shown around in the waiting area are not able

to see the other participant in the test room via monitors or one-way mirrors. There must be access to the test rooms without having to pass through the waiting area.

Any equipment and furniture in the waiting area should be as *little intrusive as possible* to the test participant. Often test participants are a bit nervous about the whole situation, and the initial few minutes of welcoming and introduction is the test leader's chance to reduce this anxiety. It is normally a good idea to offer test participants a drink in this area before commencing the actual test.

After the test, a *debrief* would normally be a part of the session. If debriefing is done with both test persons together, they could be brought back to the waiting area.

The waiting area could also serve as *observation room* during the test. This requires a monitor to be set up in the room, conveying the sound and image from both test rooms.

4.2 Equipment

What equipment is needed in the laboratory depends on the planned test scenarios. Person-person communication requires, as a minimum, a network connection between the two test rooms, and options for controlling these connections from the control-room.

Depending on the defined test conditions, the "technical environment" could either be controlled and kept constant, or be the actual variables of interest for the test. It is important to understand which technical parameters actually influence the communicative situation, and knowing how to manipulate or control these in a proper way.

See (Heim et al, 2000) for a review of the state-of-the-art of real-time person-person communication technologies.

Terminal equipment
A user, or a group of users, has two input devices that communicate with the corresponding output devices at the remote end:

- a *microphone* sends information to the remote *loudspeaker*,
- a *camera* sends information to the remote *screen*.

The microphone reduces the bandwidth, introduces distortion and mixes sound with the room's (or environment's) acoustic characteristics. Noise cancelling microphones may be advantageous in combination with echo cancellers(see section 4.2.2).

The camera reduces resolution, zooms in and shows parts of the site. A narrow angle will give more pixels per person than a wider angle for a group. In a CIF-mode (Common Intermediate Format which is 352x288 pixels) a one person communication will give about 60% of the pixels in a head-and-torso view. In a 3-group view, only 1/9 of the pixels will be available per user. That gives 7% or effectively 90x75 pixels per person in today's highest space resolution.

In addition, the screen reduces viewability due to false light (e.g., sunlight) on the screen's surface that results in low contrast and a loudspeaker introduces acoustic distortion in its mechanical parts, reduces the bandwidth and mixes the sound with the room's acoustic characteristics.

Non-terminal equipment
Between the input and output devices there can be a number of components:

PANs (Personal Area Networks): This could be a cable-based Firewire, USB or a wireless Bluetooth connection/network. The quality can be reduced due to delay of information, loss of information and loss of connection.

Encoder: This has the function in the codec to encode the information from an input device and compresses it so it is made ready for transmission on the network. QoS for audio is influenced when the information is compressed with the introduction of distortion and delay. For video there will be a reduction in the number of frames, number of pixels in the horizontal and vertical direction and the introduction of delay and distortion.

Wireless access to networks (for mobile networks): Quality is reduced because of delay, loss of information and loss of connection.

LANs (Local Area Network): Can introduce loss of information and a longer delay of information.

Routers: Introduce a longer delay of information and may introduce loss of information (depending on traffic conditions).

Tunnels, Firewalls and Encryption: Introduce delay.

Domains: Can reduce quality because of different QoS classes and models.

WAN: In addition to the LANs, Routers and the Domains, a delay is introduced because electrons require time to be transported between sites.

Decoder: This has the function in the codec to decode the information from the network, extract it and make it available for the output devices. Quality can be affected as described above for the encoder.

All these reductions in quality can be categorised into three groups:
- Network characteristics
- Codec characteristics
- Environment characteristics.

These are explained below by considering in turn terminal equipment, codec and echo canceller, network characteristics and the need for recording equipment.

4.2.1Terminal equipment (camera, microphone, loudspeaker and screen)

When communication over video or audio, the sound and image properties in the communication link is a crucial factor, whether kept constant or varied. It is therefore important that these can be monitored and/or manipulated in a proper way.

The equipment chosen must support the highest quality expected in your series of experiments, i.e. the camera chosen must support the highest space resolution of interest to the test, in order to compare it's quality with lower resolution conditions.

Viewing distance and camera position
The optimal viewing distance is a function of several parameters. The major parameter is the screen dimensions. ITU-T Recommendation P.910 recommends that when testing an audio-visual system, the viewing distance should be in the range 4H to 8H where H is the height of the screen. The recommendation indicates that the preferred viewing distance increases when visual quality is degraded. In other words, short viewing distances are more demanding than long viewing distances.

Another aspect to be considered is eye-contact or the parallax difference. ETSI ETR 297 recommends that the difference between the camera axis and the image display's eye-level axis does not exceed 8° at the closest point of the preferred viewing range/distance. See (Heim et.al, 2000) for a full discussion of these aspects.

General recommendations for video-based communication are:

- The camera shall be placed centrally on the top of the monitor.
- Viewing distances in the range of 4H to 6H (H is the screen height of the monitor) to avoid parallax difference in eye-contact and keeping the test-person's eyes not more than 1/3 H from the top of the screen (see Figure 4).

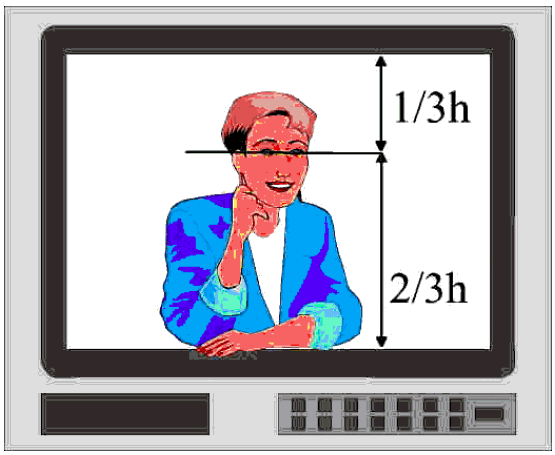


Figure 4 A template to secure parallax-free eye-contact

The choice of settings are of course dependant on the *purpose of your test*. If the purpose is to test a new service of some kind, your technical set-up must resemble that of the service you want to test, including the viewing conditions and controls provided by the service. If the purpose is purely testing quality-of-service, further recommendations can be made:

- The camera should have automatic iris-control

- The camera should have auto focus
- No self-view during the tests (to obtain the closest comparison between videoconferencing and other communication services or face-to-face)
- The user shall not have any control over the equipment during the tests.

4.2.2Codec and Echo canceller

One codec is connected to each user terminal. The codec consists of a coder, for coding the outgoing signal, and a decoder, for decoding the incoming signals.

The following parameters will be influenced by the codec characteristics:

- Delay
- Audio/video sync
- Lip synchronisation
- Audio immediate
- Video space resolution
- Video time resolution.

When using a loudspeaker sound system, an *echo canceller* is required. When sound is transmitted from one room to the other, the sound from the loudspeaker will be fed into the microphone in the same room, and transmitted back to the sender again. The echo canceller compares sound waves and cancels the “returned sound”. This is necessary to avoid feedback in the sound system, and to avoid annoying echo of own speech.

An echo canceller is normally built into the codec, but not necessarily. Echo cancellers with certain characteristics can be connected separately to the codec, if required.

The choice of a codec (with or without echo canceller) again depends on the purpose of your test, and the required transmission characteristics of the equipment.

4.2.3Network characteristics

Network characteristics are dependent on the actual network considered (e.g. PSTN; ISDN, IP network or mobile network). Network QoS control mechanisms may also be an issue to consider.

In most cases it is not possible to control the network characteristics when carrying out field studies. However, the network characteristics should be tested or monitored. The parameters to test or monitor depends on the actual network, but usually delay will be one of these.

In laboratory studies the network characteristics can be controlled.

If the test condition addresses comparison of different services, or single attributes of a network connection, it will normally be sufficient to manipulate these parameters separately. But if the test condition is to be a more complex interaction between several network parameters (delay, packet loss, image frequency, transfer protocols

etc), a *network simulator* is recommended. Network simulators are available or can be built for any type of network.

As an example the solution used in the Eye-2-Eye project is described in the following. The network simulator is connected between the two communication terminals, as illustrated in Figure 5. It is used to secure control and reproducibility of the network parameter variations.

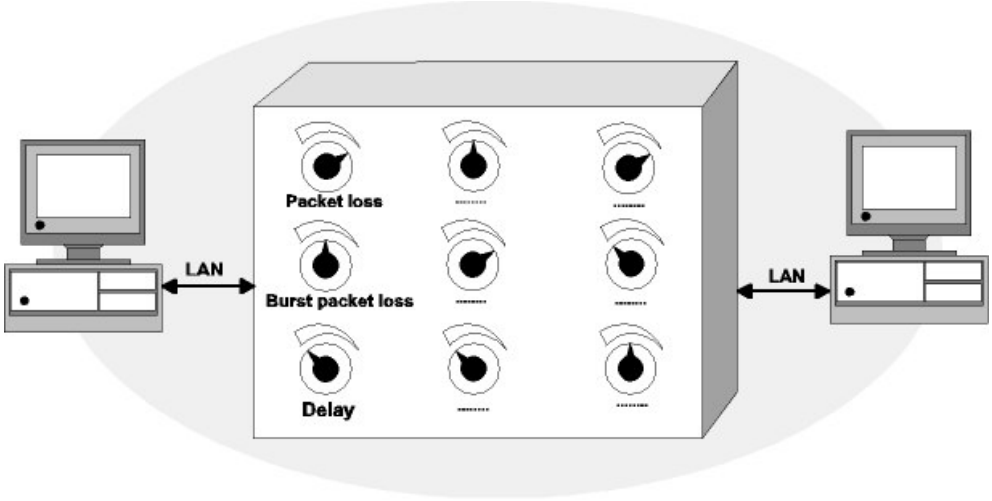


Figure 5: Network simulator

Technical Parameter Measurement
With modern digital communication technology it is no longer possible to exactly determine the technical network parameters by using simple engineering metrics. Furthermore, some of the relevant degradation effects may be generated in a specific part of a communications link, others are the sum of effects generated in several parts. It is therefore necessary to measure user-based person-to-person characteristics (e.g., audio delay between a test person's mouth and the other person's ear) as well as characteristics for individual elements of a device or a connection (e.g., the terminal jitter buffer). See (Heim et. al., 2001) for detailed descriptions of measurement equipment used in the Eye-to-Eye project.

4.2.4Recording equipment

When testing person-person communication, it will often be required to record the interaction between the partners. The actual information stream between the two terminals can be recorded from the codec, or really any point between the two terminals with a line output. This is normally the same signal(s) as will be shown on the monitors in the observation room(s).

Depending on the test setting, it could be relevant to record other aspects too - e.g. the user interaction with the terminal, or a screen dump from the computers when communication via text chat. In order to collect all the necessary information in one recording, a video-splitter is required. A four-way splitter will bring up to four video-streams together on the same screen.

Figure 6 shows how a complete laboratory set-up for audio and video communication can be connected, including recording equipment.

Terms used in the figure, that are not mentioned in the glossary list (section 9) are explained as follows:

- V1-V4 = Video camera (or video signal) 1 - 4
- Chat-MM= Chat and MultiMedia
- AX= Audio switch
- VX = Video switch

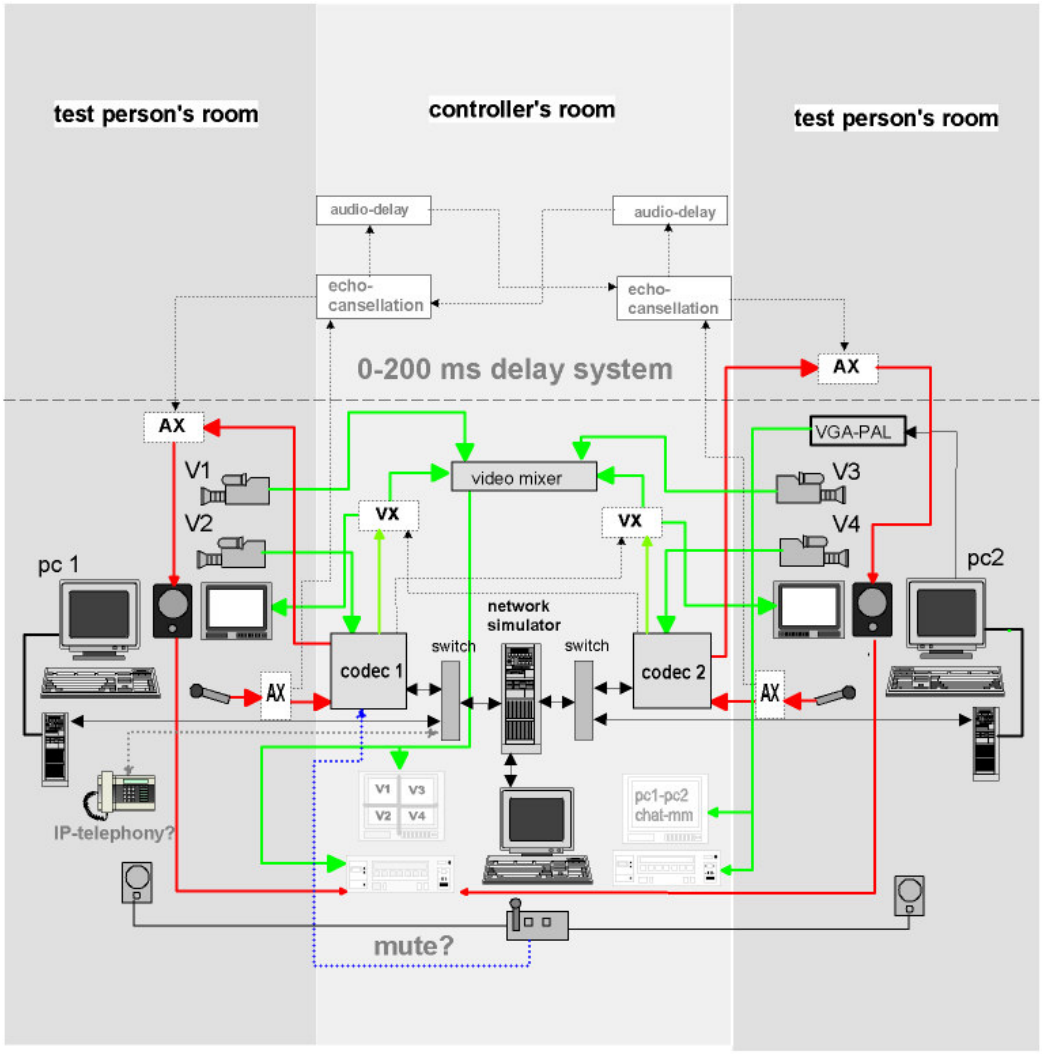


Figure 6: Schematic lay-out of equipment at the Eye-2-Eye laboratory

4.3 Fixed environmental condition

Unless the environment itself is subject for investigation, there is a range of environmental parameters that should be kept constant in order to provide optimal and equal conditions for all participants. The list below shows some of the parameters and values used in the Eye-2-Eye laboratory studies.

Parameter	Value
Lighting Conditions	100 % artificial
	Diffuse (between 500 lux and 1000 lux) (ETSI ETR 297)
	White light (100 % of the basic colours red, blue and green (RGB))
	Direct and strong reflected light in the camera's or the user's field of view avoided (ETSI ETR 297)
	Direct lighting onto the screen surface avoided
	Light source placed above the participant
Background	No patterns
	Non-reflective
	Providing good contrast to the user (ETSI ETR 297)
Acoustics	Room Acoustics: Ratio critical distance/actual distance (2 achieved by acoustical room design and microphone/loudspeaker position
	Room Noise: A-weighted equivalent level of ambient noise, LAeq < 40 dB(A)
	Sound insulation between the test rooms: (35 dB within the frequency range from 100 Hz to 8000 Hz
Audio	SLR (between A and B) 8 dB (1 dB (ITU-T Recommendation P. 79)
	RLR (between C and D) 2 dB (1 dB (ITU-T Recommendation P. 79)
	Image frequencies of sine-wave signals in the frequency band 9 kHz up to 15 kHz produced at the digital interface < reference level obtained at 1 kHz by at least 25 dB (both the test signals and the reference signal shall be applied acoustically on the microphone)
Viewing Distance	Range 4H to 6H (where H is screen height)
Camera Position	Placed centrally on the top of the monitor
	Field of view:
	from waist to top of head of participant
	eyes of participant approximately at 1/3-2/3 horizontal division of monitor
Camera Parameters	Automatic iris-control
	Autofocus
	Colour correction in accordance with test chart
Self View	No Self View
User Control	No User Control

Table 2: Environmental parameters kept constant in the Eye-to-Eye laboratory studies

5. Recommendations and Templates for Conducting Laboratory Studies

This chapter contains a selection of recommendations and checklists for planning and performing a communication study in the laboratory. Section 7 provides examples of actual data collection instruments as well as reference data for comparison of your results.

This chapter consists of:
1. Checklist for conducting Laboratory Experiments
2. Recommendations for planning the test
3. Legal and ethical issues
4. Confidentiality and informed consent
5. Checklist for one experimental run
6. Pre- and post-test interviews
7. Data registration
8. Data analysis
9. Practical considerations and requirements.

5.1 Checklist for conducting Laboratory Studies

Sub-task	**Section number in EM document**
Define problem	2.1
Define Tasks and Scenarios (high level)	2.2
Define User population, sample size and recruitment procedure	2.3
Define Technical and physical context	2.5
Decide Test conditions	2.4 / 2.6
Decide Dependent measures	2.4
MILESTONE A: Preparations and Design	
Draft plan for test process	5.2
Produce first version of test materials	4 / 7
Recruit participants to pilot study	2.3
Pilot testing	-
MILESTONE B: Draft test plan and Piloting	
Revise test materials	-
Plan in detail the experimental phase	4
Recruit participants to main study	2.3
Conduct experiments	-
Register raw data	5.7
MILESTONE C: Final test plan and Main data collection	
Analyse data	5.8
Conclude	-
Present results	-
MILESTONE D: Analysis and Conclusions	

5.2 Recommendations for planning the test

Planning the test is really important in order to get the outcome as intended. A laboratory-based user test involves a lot of activities before, during and after the actual data collection. Involving external users also complicates the total picture.

- Designing the test depends on making the right top-level decisions regarding research focus, user tasks and user groups.
- Tools and tasks should be piloted and refined in several iterations. Are they understandable to the user? Does the data collection instruments provide the information actually needed?
- The success of the actual data collection depends on the pool of users, and that what population they actually represent is explicitly stated. Recruiting and motivating users for participation is a crucial issue. Ensure that all participants are recruited in the same way, a written procedure should describe what to do and what to say during recruitment.
- Explicit procedures for allocating participants to experimental conditions must be worked out.
- Testing with several users probably requires more than one test leader, which requires that all test leaders have been equally trained and co-ordinated. The goal is that all users are given the exact same treatment (besides the experimental variations) throughout the test.
- A set of all material to be used in the experiment and a description of how it is applied during a test session is needed. This includes instructions, questionnaires and task forms, and possible other material to be used in the test.
- A description of how to operate the apparatus in the study must be worked out so that all settings can be done prior to the participants' arrival.

5.3 Legal and ethical issues

Be aware that both formal and informal rules exist related to involving users in your laboratory tests:

- National legislation regulate the use of person-identifiable information
- Person-identifiable information should only be collected for the purpose of future contact with the participants if that is necessary. These records should be deleted when the project terminates.
- Offer to present or submit information about project results after the investigation.
- Everybody in contact with the participants should be briefed about the necessity of treating participants with dignity and respect.
- Young persons (under 18) need parent's permission to participate.

Codes of conduct and ethical principles vary from country to country, and national (and international) regulations should be investigated during the planning process. The American Psychological Association (1992a) provides an extensive document on ethical principles and code of conduct. See http://www.apa.org/ethics/code.html

5.4 Confidentiality and informed consent

- During recruitment inform participants ...
  - about the general purpose of the research (without giving away information that can influence on the result of the study)
  - if videotape is used for documentation
  - that full confidentiality are guaranteed
  - practical information (time, place, eventual payment, etc)
- During pre-experiment instruction …
  - further information about the nature of the experiment
  - that participation is voluntary, and they can withdraw at any time
  - that electronically stored data will not contain information that can identify the participant
  - that individual data can be deleted on the participant's request
- Get written consent about participation, videotaping and usage of videotapes for demonstration of project activities
- During post-experiment interview and de-briefing
  - Ask participants not to inform other potential users about the experiment.

5.5 Checklist for one experimental run

A checklist for one experimental run is needed. Describe step-by-step instructions to the experimenter. This includes the main interactions between participant and experimenter, as well as the experimenter's duties before, during and after the each trial. Here is an example:

Experimental procedure - generic example

Prepare for new participant. Load videotape, set apparatus and prepare all relevant test materials
Welcome participants, show them around the localities, and introduce them to other personnel they will interact with
Inform about the purpose of the experiment
Conduct pre-experiment questionnaire.
Read instructions.
"Any questions?" – answer
Start videotape.
Start trial.
If applicable; give within-trial instructions.
End trial.
Stop videotape.
Conduct post-experiment questionnaire
Dismiss participants, pay.
Archive videotapes and questionnaires.

5.6 Pre- and post-test interviews

For different reasons, it can be useful to collect some general information about your participants ahead of the actual test trial (*pre-interview*). This will typically be some demographic data that could possibly be relevant to your study.

If standardised measurement tools (either public or your "own") are used, it can be interesting to archive the data for future purpose of reference. In this respect it is always useful to have a set of demographic information on each participant to link to the actual questionnaire data. The pool of reference data will grow with each experiment, and might give interesting information on variations associated with different demographics.

A pre-interview should be very brief and simple, and not challenge the users mind at all. It should only take a few minutes. A typical pre-interview form could consist of the following information:

- Participant number
- Name of study and experiment condition
- Age
- Gender
- Formal education level
- Work experience (present occupation, number of years)
- General experience with use of computers
- Frequency of computer-use per week

Depending on the research question of interest, it could be interesting to collect some more specific information on the participants' attitudes towards or skills within a certain area. The Eye-to-Eye project collected information on the users technology experience as a part of the pre-interview session (see section 7.3)

After finishing the test trial, a *debriefing* is advised in order to give the participants a "smooth exit" from a situation that sometimes is experienced as stressful or uncomfortable. A debriefing would typically consist of:

- letting participants meet each other face-to-face (if they haven't already met)
- informal chatting about the test situation, the purpose of the test, etc, and answering any questions the participants might have
- if appropriate, showing participants the results of their performance
- informing the participants that they should not tell other potential participants about the test

A *post-interview* is a way of collecting additional information to the actual test data, if this is found necessary. This is typically very useful in a *pilot-phase*, in order to get feedback on the procedures used, the clarity of the instructions etc. A post-interview can also generally be used to collect qualitative data on other aspects of the test situation, i.e. users comfort/discomfort with the situation, their general view on the communication service used etc.

If a post-interview is conducted, the debriefing process will normally be part of this interview. The post-interview should also be quite brief and of short duration.

5.7 Data registration

There are several ways in which data may be registered in a laboratory study.

Data may be based on observers' notes during the session or during playback of videotape after the session. The more structured the registration process is, the less demands are laid on the observer's skills and experience. When observation of the participants' behaviour is the basis for the data in the study, it is advised to provide the observers with a set of pre-defined observational categories. These should be prepared on paper forms, to be ticked off each time the particular behaviour is observed. Usually the time for the observation is also noted. An event recorder or event recording software may be useful if there are several observation categories or if they must be recorded at a high speed.

Data may also be collected by letting the participants fill in forms and questionnaires. Usually, there will then be a need for a conversion of the paper material to an electronic medium for further data analysis.

- Before and during test:
  - Data collection procedures should be piloted, if possible allowing for several iterations
  - All test papers should be labelled (version control)
  - During test, all papers must be labelled with participant's number

- After test:
  - Paper and videotape shall be archived immediately after test
  - Provide electronic registration forms with same lay-out as paper materials used in the test
  - Electronic registration forms should be piloted before testing starts
  - Ensure frequent back-up of registered data
  - Make sure that test data can not be paired with person identifiable information.

Automatic logging of user behaviour may be useful, if there is a clear relation between the problem under investigation and the kind of data provided by automatic logging. If the problem relates to call frequency, duration and/or "who-talks-to-whom", automatic logging may be feasible. Beaware that one would usually need some software to process the logs in order to get meaningful information. Logging the content of the conversation may also be useful for acquiring qualitative information.

5.8 Data analysis

Most studies require that data be analysed to some extent, usually by means of statistical data analysis. Unless there are expertise in statistical analysis within the project group, there is a need that somebody in the team looks into this issue. A very good starting point is ETSI EG 201 472 (2000) (see section 1.7)

A special consideration for communication experiments is the way one looks at each pair and each participant. In most cases the "pair" is the natural unit of data analysis.

So if there are 50 participants in the study, there will only be 25 units for the data analysis. This is a natural thing when the dependent measure is, for example, time needed to solve a problem together, or number of times participant A interrupts participant B. When the measure is the same for both participants, for example the degree to which one found the medium suitable for the task, it may be tempting to enter all single participants into the analysis as independent observations. In most cases this will not be feasible, because the observations may not be regarded as "statistically independent". What one participant thinks of the question may be influenced by the other participant's attitude. In such cases one should either combine the observations as the "pairs mean score" (or something like that), or one could treat the two participants as different aspects of the pair's response - "Participant A's attitude, and Participant B's attitude".

5.9 Practical considerations and requirements

Finally in this chapter, a list of practical issues to consider when performing laboratory studies is:

- An experimental session should not take more than about 90 minutes
- There should be a suitable incentive for participation
- All personnel that come in contact with the participants, like receptionists and colleagues in the vicinity of the lab, should be informed about the event
- A "Do not disturb" sign on the laboratory door may be useful, but must be removed when the laboratory is not in actual use
- There must be a suitable place to wait for subjects that arrive in advance of the appointment
- Participants should not have to wait for more than 5 minutes if they arrive at the agreed time
- Depending on the length of the session, participants should be offered refreshments like tea, coffee or soft drinks.

6. Studying communication technology use in the field

The term "field study" covers most research activities aimed at studying peoples' activities in a realistic, as opposed to an artificial, context. Although a field study might introduce "artificial " activities like being interviewed or trying out new technology, the main characteristic is that the user is using the technology in a "real" setting, for example at home or at work. The issue of deciding between conducting a laboratory or a field study is discussed in chap 2.5.

This chapter discusses the special considerations that should be taken for field studies of person-person communication services. For more general information about conducting studies in the field, the reader is referred to general books on methods in social psychology, for example Bickman and Rog (1998).

Some of the recommendations and templates presented in section 5 apply to the design of field studies as well, especially sections 5.3, 5.4 and 5.6. However, planning a field study is far less straight-forward as is the case for a controlled laboratory study, as much of the planning will have to be adapted to the actual real-life context in question. Therefore, this chapter does not provide the same level of details in the advice and recommendations for conducting the study.

6.1 Critical issues in designing the study

When planning the study, consider the following:

- **Participants**
  The participants should be representative for the target population defined in the problem statement. This requirement is discussed in 2.3. One should be aware that the requirement of having a "statistically representative sample" is extremely difficult in field studies. However, if the project is to introduce or assess a technology in a specific organisation, this requirement does not necessarily apply, sometimes the sample can be regarded as *being* the population.

  The required number of participants depends on the method for data collection and analysis. Data collection that involves in-depth interviews or direct observation may be conducted with as little as with 5-15 participants. Data collection only through log-data and surveys should involve more participants.

  It is important to assess the potential communication patterns between the participants in advance. If there are 10 participants, each of them might speak to all the nine others, so there are potential 45 dyads in the study. Since each conversation might be initiated from each side of the dyad, there is a total of 90 different kinds of calls. The planning of the study might take this into consideration with respect to the amount of effort involved in analysing all potential dyads.

- **Service technology**
  The technology to be investigated may be an existing service already used by the participants, or it may be introduced for the first time. The motivation of the users to participate in a study is often the excitement by given the opportunity to try out new technology. If the service is under development, it is therefore of paramount importance that it is properly de-bugged before the participants start using it.

Frequent technical errors are de-motivating for the users and invalidating the data that are being collected.

- **Methods for data collection**
  It is important to understanding the users' tasks and motivation for communication in advance of collecting data. In order to establish a good rapport with the informants, a basic knowledge of the users' daily task and communicative activities is necessary.

  Data collection could be conducted through qualitative methods such as in-depth interviews, observations and focus groups, or quantitative approaches like questionnaires and logging. Even if data collection is mainly conducted through quantitative methods, it is recommended to use a certain amount of qualitative methods both before the data collection (in developing and piloting the data-collection tool) and after the data collection (in validating the results).

  In field studies it is natural to measures communication service choices and preferences. The question of "task outcome", i.e. how well the communicative tasks were performed, may often only be treated indirectly, as questions about how suitable, how effective etc the different media are for different kind of tasks. Such data are then based on the subjective assessment of effectiveness etc. As pointed out elsewhere (Schliemann, Heim, Asting & Følstad, 2001) the distinction between task-outcome measured by objective means and subjective assessments is important, and the two measures may often not coincide.

- **Cross-sectional vs. longitudinal study**
  If the study involves the introduction of services that is new to the participants, it is probably necessary to stretch out the data collection in time (longitudinal data collection). It may be useful to collect baseline data before the introduction of the communication service, right after the introduction of each new service, and when the participants have familiarised themselves with the service. If the study involves actual use of a service that is familiar and in use by the participants, it may be possible with one single data collection (cross -sectional study).

- **Control conditions**
  Although a field study may not reach the same degree of rigorous control as a laboratory study, it is still advised to include control conditions. This might include a control group that matches the actual users as closely as possible. It may also include measuring the communication before and after introduction of a new service. The identification of those users that are engaged in the different tasks and activities is therefore very important for the allocation of participants to study vs. control conditions. This is normally referred to as a quasi-experimental approach (e.g., Cook & Campbell, 1979).

  Registering pre-treatment communication frequencies by logging telephony calls and e-mail usage are easy and relatively cheap in organisations with traditional switchboards. This may be very useful, both for composition of experimental and control groups, and for establishing a pre-introduction baseline.

When the design of the study is set consider the following:

- *Recruit participants:* Make sure by personal contact with the participants that they really belong to the target population. Then, the schedule of the data collection should be presented along with an indication of the workload and any disruption to be expected by the participant. Make the participation seem worthwhile - remember that the participants should be offered something in return for their participation.

- *Develop and pilot equipment and data collection methods:* All equipment and data collection needs to be thoroughly piloted before it is introduced to the participants. If possible other representatives of the target population should be used in the piloting process.

6.2 Introducing the service

If the study involves the introduction of a new service, the participants should get sufficient training at the point of introduction. Do not trust the participants to read provided written material, if possible the training should be conducted one-to-one. The participants might be given pre-defined exercises in order to make them start using the new service. Provide an easy way for the participants to contact resource persons for trouble-shooting, and make sure to call on the participants a short time after introduction in order to assure that they actually have started to use the equipment.

6.3 Logging and interpreting frequency and duration

A good data source for the use of a particular service is the log of actual use of the service. An automatically generated log may provide information on:

- who is calling
- who is receiving
- what communication service is being used
- time
- duration.

If, in addition, one wants information on what the service is being used *for* a manual log is recommended (where the participants logs their instances of use).

Logging the *content* of the conversation raises serious ethical questions, and is normally not recommended. However, with careful planning and by obtaining the users' informed consent, it might be an option. Normally this would also involve some kind of permission from the authorities according to national legislation.

One should be aware that automatic logging of usage often produces a huge amount of data. One would therefore need some kind of software to systematise the data before it is analysed by traditional methods.

One approach to analysing log-data is to begin with displaying the data graphically. Visual inspection will often reveal interesting trends or differences between the different communication services that might be further followed up.

6.4 Interviews and questionnaires

To gain in-depth knowledge of the user's particular use and experience with a service, in-depth or semi-structured interviews are recommended. It is recommended to conduct interviews with 5-15 persons, depending on the variation within the target population. The interviewer should have a general idea of what kind of information is required from the interviewee, but not necessarily the categories in which the answers might fall. Rather, the interviews should elicit the relevant user categories for this particular combination of user population and communication service.

When there is a reasonable amount of knowledge about what categories and labels that are relevant for the user population, quantitative feedback may be gathered. The content and problem areas addressed in the interviews may be further refined into questionnaire items. A questionnaire is a good way to gather information from a higher number of participants. Also the responses to a questionnaire may serve as a validation of the results produced by the interviews.

In the final analyses, the interview and questionnaire data may serve as a source of information on how the users have experienced the service and tasks for which they have used it. This analysis will be complementary to a log analysis.

7. Data collection Instruments and Reference data

This chapter consists of selected data collection instruments used in the Eye-to-Eye project, as well as reference data with descriptive statistics from the same data collection processes. Each "issue" consist of:

- a short introduction to the application of the instrument
- the actual instrument, including instructions for use, and
- reference data, if they exist.

The idea is that the reader and potential user of the tools could either copy the actual page(s) from the paper-version for use in their own test situation, or copy the content electronically from the document in order to integrate it into their complete set of test materials.

This chapter consists of:

1) Form for Participants consent
2) Pre-Interview
3) Technology Experience Questionnaire
4) Technology Experience Reference data
5) Person Perception Questionnaire
6) Person Perception Reference data
7) Social Presence Questionnaire
8) Social Presence Reference data
9) Communicative Behaviour data collection process
10) Communicative Behaviour Reference data

7.1 Form for participants consent

As the participants are being observed and their behaviour usually is recorded by means of audio and/or video, there is a need to ensure that they agree to the use of this kind of information in the project and potential future applications. Here is a template that can be used:

Video Consent Form

You will be audio- and video taped during completion of this task. These tapes are strictly confidential and will be kept to be used only for research purposes. They will not be associated with your name at any time.

You will not be asked to give any personal information in the course of the experiment. If you wish, you are free to leave at any point during the experiment.

I, (please print your name) ……. consent to participate in the experiment and consent to the use of audio and video recording and these recordings being kept for future research purposes.

Signed:……

Date: ……..

7.2 Pre interview

Pre interview should be conducted before the start of the actual experiment session. It will normally be done in connection with presenting and introducing the test context, and serves as a way of "registering" the participant, and collecting some general information about the user.

The pre-interview can be very brief, as exemplified below, or done in connection with more structured collection of relevant background information, as presented in the next section.

Pre-Interview

P#: ☐

IDnr. ☐ ***Age*** ☐ *yrs* ***Gender*** ☐ *M* ☐ *F*
Education
☐ *Primary School* ☐ *College* ☐ *University* *Subject:* ☐
Work
☐ *Number of years in present occupation* *Occupation/Nature of work* ☐
Computer experience ☐ *How many years has the participant used a computer?* ☐ *How many hours, in average, does the user use a computer during a week?*
Done by: *Date:*

7.3 Technology experience Questionnaire

Results from user interaction with communication technology might depend on the users former experience with similar or related technology. It is therefore important to measure the participants' prior knowledge of communication technology that might have an effect on their performance in a test situation.

The Eye-to-Eye project developed and used a general Technology experience Questionnaire that was administered to every participant in the Baseline studies and Laboratory studies, prior to actual user tests. The questionnaire presents a set of technologies, and asks the user to tick off how often they use the different technologies (see next page).

The categories are of a qualitative nature, but results have been converted to discrete values from 1 to 6, referring to the six categories from left to right. Because of the ordinal character of the data collected, the reference data (section 7.4) provides median values and max-min values for the data collected, rather than means.

Technology experience Questionnaire

For each technology, please check of how often you have used it. If none of the categories correspond to your actual usage, please check the one that is closest.

	Have no knowledge of the technology	Have never tried	Have tried	Use monthly (once or more per month)	Use weekly (once or more per week)	Use daily (once or more per day)
Ordinary telephone						
Mobile phone						
SMS messages using mobile phone						
Email (e-mail)						
Telephone over internet (using a PC to make a voice call to another PC or telephone)						
Video over internet (one or more people at each end, using a PC and suitable software)						
Videophone (one person at each end, using special videophone equipment)						
Videoconference (use of special equipment for videoconferencing with more than one person at each end)						
Instant messaging - ”Chat” (two persons that at the same time are sending text messages to each other by means of a PC)						
Chat-room on the Internet - ”Chat-room” (where one can use text to have a conversation to one or more persons at the same time)						
Text-based “virtual worlds” (where several people can interact by means of text in an artificial reality, often a sort of a game)						
Graphical-based “virtual worlds” (where several people can interact by means of graphics in an artificial reality, often a sort of a game)						

7.4 Technology Experience reference data

Reference data for the Technology experience Questionnaire, based on data from 184 students - age 18-47 (mean 24). Normal distribution of the variables can not be assumed.

	Median	Min	Max
Ordinary telephone	6	3	6
Mobile phone	6	2	6
SMS messages using mobile phone	6	2	6
Email (e-mail)	6	2	6
Telephone over internet	2	1	5
Video over internet	2	1	4
Videophone	2	1	3
Videoconference	2	1	3
Instant messaging - ”Chat”	3	1	6
Chat room on the Internet	3	1	6
Text-based “virtual worlds”	2	1	6
Graphical-based “virtual worlds”	2	1	6

Reference data for the Technology experience Questionnaire, based on data from four different user groups from Baseline study 1.

User group	Student N=17			Young N=12			Senior citizen N=11			Business professionals N=12		
	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max
Ordinary telephone	6	4	6	6	6	6	6	6	6	6	5	6
Mobile phone	6	4	6	6	3	6	5	2	6	6	4	6
SMS messages using mobile phone	6	4	6	6	2	6	2	2	3	5	2	6
Email (e-mail)	5	4	6	5	2	5	3	2	6	6	6	6
Telephone over internet	2	2	6	2	2	5	2	2	6	2.5	2	4
Video over internet	2	2	5	2	2	3	2	2	5	3	2	3
Videophone	2	2	2	2	2	3	2	2	3	2	2	3
Videoconference	2	2	3	2	1	2	2	2	3	3	2	4
Instant messaging - ”Chat”	3	2	4	3	2	5	2	2	5	3	2	5
Chat room on the Internet	3	1	3	4	2	5	2	2	2	2	2	4
Text-based “virtual worlds”	2	1	5	2	1	3	2	2	2	2	2	3
Graphical-based “virtual worlds”	2	1	5	2	1	5	2	2	2	2	2	3

7.5 Person perception questionnaire

The service used to conduct communication may have an effect on the user's subjective perception of their communication partner. To investigate the extent of this effect for a particular service, data could be collected by means of the 'Person Perception Questionnaire'

This questionnaire collects data on people'ds first impression of their partner after having communicated over a given task, based on the fact that communication partners don't know each other in advance. Participants are asked to judge their partner on a set of person characteristics, on bipolar scales describing opposite dimensions like kind-unkind, open-reserved, honest-dishonest.

Full presentation of the rationale and background for developing the questionnaire is described in Schliemann et. al. (2001).

A modified version of the questionnaire was developed for use in some of the experiments. This version consists of the six person characteristics from the original questionnaire that showed to be main ones differentiating between users. The assessment of the other person is also done simpler, by omitting the questions about the strength of the impression. The modified version also uses a 10 point scale, rather than a 7 point scale as in the original, in order to be consistent with other instruments used in the Eye-2-eye laboratory studies.

The modified version of the person perception questionnaire is shorter and quicker to fill in, and could be used in tests where person perception is not one of the main issues.

Person perception questionnaire

Do you know the other person who participated in the experiment?

Yes ☐ No ☐ Don't know ☐

Try to describe the person you just met according to the characteristics and scales below. A brief interaction may not seem like a sufficient basis for doing this. However, people tend to form a first impression of other people even after a very short time. It is this subjective and intuitive impression you should try to express.

What is important is not whether the characteristic is right or wrong, but that you express how **you** perceive the person.

The questionnaire will be treated confidentially. The other person will not get to know your answers.

For each scale below there are **two answers**.

Firstly, put a cross in one of the seven squares

- A cross in the square to the far left indicates that you consider that the person possesses much more the characteristic to the left of the scale.
- A cross in the square to the far right indicates that you consider that the person possesses much more the characteristic to the right of the scale
- A cross in the underlined square in the middle of the scale indicates that you consider that the person possesses an equal amount of both characteristics.

It is important that the scale describes the quality of your impression, and not how strong or weak this impression is. **A cross in the middle of the scale does not indicate doubt or "don't know"**.

Please pay careful attention to the words at each end of the scale so that you don't make a mistake in your rating.

Secondly, put a cross against one of the three statements below the scale

- "I have a vague impression of this aspect of the person's character" indicates that you have a weak feeling or sensation about the other person when it comes to this aspect.
- "I have a strong impression of this aspect of the person's character" indicates that you have a quite clear and strong intuition about the person according to this aspect.
- "I find it impossible to rate this aspect of the person's character" indicates that you don't know or have absolutely no idea how to rate this aspect.

If you choose "I find it impossible to rate this aspect of the person's character", you should not put a cross in any of the squares for the scale describing this aspect.

If anything is unclear to you, please ask the test leader.

Formal ☐ ☐ ☐ ☐ ☐ ☐ ☐ Informal

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Domineering ☐ ☐ ☐ ☐ ☐ ☐ ☐ Submissive

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Kind ☐ ☐ ☐ ☐ ☐ ☐ ☐ Unkind

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Reserved ☐ ☐ ☐ ☐ ☐ ☐ ☐ Open

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Intelligent ☐ ☐ ☐ ☐ ☐ ☐ ☐ Unintelligent

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Insensitive ☐ ☐ ☐ ☐ ☐ ☐ ☐ Sensitive

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Trustworthy ☐ ☐ ☐ ☐ ☐ ☐ ☐ Untrustworthy

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Cooperative ☐ ☐ ☐ ☐ ☐ ☐ ☐ Competitive

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ Fair

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Warm ☐ ☐ ☐ ☐ ☐ ☐ ☐ Cold

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Dishonest ☐ ☐ ☐ ☐ ☐ ☐ ☐ Honest

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Friendly ☐ ☐ ☐ ☐ ☐ ☐ ☐ Unfriendly

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Boring ☐ ☐ ☐ ☐ ☐ ☐ ☐ Interesting

I have a vague impression of this aspect of the person's character ( )
I have a strong impression of this aspect of the person's character ( )
I find it impossible to rate this aspect of the person's character ( )

Person perception Questionnaire - modified version

We want you to characterise the person you have been communicating with according to the attributes listed in the scales below. A short meeting may not seem a sufficient basis for doing this, however, we do get an impression of the people we meet even after a brief meeting such as you have just experienced. It is not important whether or not your characterisation is right or wrong, but it is important that you express your impression of the other person.

1. I would describe the other person as

formal	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	informal
	0	1	2	3	4	5	6	7	8	9	10	

2. I would describe the other person as

co-operative	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	competitive
	0	1	2	3	4	5	6	7	8	9	10	

3. I would describe the other person as

warm	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	cold
	0	1	2	3	4	5	6	7	8	9	10	

4. I would describe the other person as

friendly	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	unfriendly
	0	1	2	3	4	5	6	7	8	9	10	

5. I would describe the other person as

trustworthy	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	untrustworthy
	0	1	2	3	4	5	6	7	8	9	10	

6. I would describe the other person as

unfair	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	fair
	0	1	2	3	4	5	6	7	8	9	10	

7.6 Person perception reference data

Data are based on the 'full version' of the Person Perception questionnaire.

The reference values are based on data from 184 students, age 18-47 (mean 24). Normal distribution of the variables can not be assumed. Several "don't apply" or "can't say" responses are to be expected.

	Mean	Standard deviation
Formal - Informal	4,45	1,50
Domineering - Submissive	3,84	1,07
Kind - Unkind	2,26	0,93
Reserved - Open	5,13	1,37
Intelligent - Unintelligent	2,51	1,10
Insensitive - Sensitive	5,02	1,06
Trustworthy - Untrustworthy	2,31	0,97
Co-operative - Competitive	2,16	1,29
Unfair- Fair	5,70	0,92
Warm - Cold	2,92	1,11
Dishonest - Honest	5,81	0,91
Friendly - Unfriendly	1,93	0,95
Boring - Interesting	4,66	1,48

7.7 Social presence questionnaire

The service used to conduct communication may have an effect on the user's subjective perception of the social presence that they felt during the communication. The term "social presence " or "telepresence" has been conceived of as a quality of the communication media, a synthesis of spacial and reciprocal communicative cues such as the capacity to transmit visual non-verbal cues such as facial expressions, gaze direction awareness, turn-taking cues and the apparent distance and "realness" of others (Short, Williams and Christie, 1976, Muhlbach et al, 1995). To investigate the extent of this effect for a particular service, data could be collected by means of the 'Social Presence Questionnaire'. A complete presentation of the rationale and background for developing the questionnaire is described in O'Malley et. al. (2002).

Participants are asked to indicate the degree to which they agree or disagree with twelve statements that address various aspects of social presence.

The questionnaire items used to assess the social presence of the media are taken from earlier work (Short, Williams and Christie, 1974; Mulhbach et al, 1995), modified in some cases and with new items. They are intended to measure the following various factors that influence social presence (number in parenthesis refer to the item number in the questionnaire presented on the next page):

- Mutual attentiveness and responsiveness (1)
- Turn-taking cues (2)
- Feedback (3)
- Informational non-verbal cues (4,5)
- Spatial presence (6, 7, 9, 10)
- "Realness" (8, 11)
- Emotional contact (12)

Social presence Questionnaire

The statements below treat your experience of the communication you were engaged in with the other person in the experiment.

Indicate by circling one of the numbers on each scale, whether you agree or disagree in the following statements

1. I knew when the other person was paying attention

agree ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ disagree
0 1 2 3 4 5 6 7 8 9 10

2. We both seemed to know when we could take our turn to speak

agree ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ disagree
0 1 2 3 4 5 6 7 8 9 10

3. I knew when I had been understood

agree ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ disagree
0 1 2 3 4 5 6 7 8 9 10

4. I didn't know how the other person was feeling

agree ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ disagree

0 1 2 3 4 5 6 7 8 9 10

5. One does not get a good enough idea of how people at the other end are reacting

agree ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ disagree
0 1 2 3 4 5 6 7 8 9 10

6. The other person seemed very far away

agree ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ disagree
0 1 2 3 4 5 6 7 8 9 10

7. The sound seemed to come from the other person

agree	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	disagree
	0	1	2	3	4	5	6	7	8	9	10	

8. The other person did not seem real

agree	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	disagree
	0	1	2	3	4	5	6	7	8	9	10	

9. I had the sense of “being in the same room” as the other person

agree	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	disagree
	0	1	2	3	4	5	6	7	8	9	10	

10. The other person looked unnaturally small

agree	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	disagree
	0	1	2	3	4	5	6	7	8	9	10	

11. It was just like a face-to-face meeting

agree	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	disagree
	0	1	2	3	4	5	6	7	8	9	10	

12. I had a good impression of personal contact with the other person

agree	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	disagree
	0	1	2	3	4	5	6	7	8	9	10	

7.8 Social Presence Reference data

The reference values are based on data from students and staff at the university of Nottingham, UK age 18-48 (mean 25). Normal distribution of the variables can not be assumed. Several "don't apply" or "can't say" responses are to be expected.

Scores for items 1-3, 6-7 and 9-12 are reversed in order that, for clarity, a higher score for any item indicates "greater" social presence. This is necessary because with the described questionnaire a higher score indicates disagreement with the statement. However, on some items the statement is worded negatively so a reversal of the score is not necessary.

Media Condition	description	Screen size	Screen resolution	Video delay	Audio delay
A	Synchronous	29"	CIF	200ms	200ms
B	Delayed	29"	CIF	650ms	650ms
C	Asynchronous	29"	CIF	400ms	200ms
D	Small screen CIF	3.5"	CIF	650ms	650ms
E	Small screen QCIF	3.5"	QCIF+Filter	650ms	650ms
F	Small screen avatar	3.5"	N/A	1000ms	1000ms

Technical condition A:

Questionnaire Item	N	Mean	Standard Deviation
1-attention	38	8.03	1.90
2-turns	38	6.26	2.59
3-understood	38	7.53	1.93
4-feeling	38	5.11	2.47
5-reacting	38	6.13	2.54
6-far away	38	4.29	2.91
7-sound	38	7.50	2.53
8-real	38	7.50	2.39
9-same room	38	4.26	2.88
10-small	38	6.32	3.35
11-face to face	38	3.97	2.91
12- Personal contact	38	5.47	2.72

Technical condition B:

Questionnaire Item	N	Mean	Standard Deviation
1-attention	21	8.33	2.03
2-turns	21	5.67	2.24
3-understood	21	6.71	2.28
4-feeling	21	5.05	2.58
5-reacting	21	6.19	2.36
6-far away	21	4.14	3.20

7-sound	21	7.76	2.14
8-real	21	7.86	1.80
9-same room	21	5.52	2.96
10-small	21	6.62	3.04
11-face to face	21	4.71	3.16
12- Personal contact	21	6.05	2.16

Technical condition C:

Questionnaire Item	N	Mean	Standard Deviation
1-attention	23	8.48	1.12
2-turns	23	7.22	2.05
3-understood	23	7.65	1.64
4-feeling	23	5.74	2.91
5-reacting	23	7.30	2.12
6-far away	23	3.30	2.30
7-sound	23	6.57	2.87
8-real	23	7.52	2.37
9-same room	23	5.00	2.71
10-small	23	2.57	2.31
11-face to face	23	4.48	2.69
12- Personal contact	23	6.09	2.66

Technical condition D:

Questionnaire Item	N	Mean	Standard Deviation
1-attention	28	8.14	1.48
2-turns	28	5.82	2.58
3-understood	28	6.82	2.71
4-feeling	28	5.71	2.02
5-reacting	28	5.61	2.44
6-far away	28	4.04	2.44
7-sound	28	6.86	2.49
8-real	28	8.25	1.73
9-same room	28	4.18	2.75
10-small	28	6.54	3.31
11-face to face	28	3.54	2.65
12- Personal contact	28	5.21	2.38

Technical condition E:

Questionnaire Item	N	Mean	Standard Deviation
1-attention	29	7.70	2.34
2-turns	29	5.90	2.86
3-understood	29	6.90	2.32
4-feeling	29	5.00	2.73

5-reacting	29	5.80	2.82
6-far away	29	4.17	3.39
7-sound	29	7.20	2.58
8-real	29	7.60	3.08
9-same room	29	4.10	2.75
10-small	29	6.17	3.15
11-face to face	29	3.33	3.03
12- Personal contact	29	5.30	2.73

Technical condition F:

Questionnaire Item	N	Mean	Standard Deviation
1-attention	16	3.63	2.99
2-turns	16	6.94	3.09
3-understood	16	5.94	3.17
4-feeling	16	3.38	2.73
5-reacting	16	3.25	2.32
6-far away	16	4.19	2.90
7-sound	16	6.38	2.90
8-real	16	6.63	1.96
9-same room	16	3.38	2.50
10-small	16	5.50	2.83
11-face to face	16	1.69	2.15
12- Personal contact	16	4.06	2.62

7.9 Communicative behaviour data collection

The service used to conduct communication may have an effect on communicative behaviour. Measuring such behaviour may indicate whether variation in media or technical parameters have changed the style of communication, or made it more like face to face communication. For example, a long audio delay may disrupt turn taking behaviour leading to more fomality in the communication. This might be shown by longer and fewer turns which minimises problematic turn taking and fewer interruptions because people are monitoring their own behaviour. Standard procedure to collect such data is to transcribe dialogues. This enables a simple analysis of structural elements of the dialogue such as the number of words, turns and interruptions, or complex and time consuming discourse and conversational analyses. An example of a fairly straightforward transcript is given below. This is taken from a dialogue where participants have to give each other information about a route on a map when they don't have visual access to the map:

In this scheme, each line of text corresponds to one speaker's turn at speaking. In order to identify who is speaking the turn should be preceded by a code such as "A" or "B" (in this case, A = instruction giver, B = instruction follower).

In cases where speakers interrupt each other, or where there is overlapping speech, a left-angled bracket (<) signifies the beginning of the utterance in which the interruption or overlapping speech occurred. Then the exact point at which there was an interruption is indicated with a backslash (/), followed by the interrupting speech. Finally, the end of the turn in which there was an interruption is signalled by a right angled bracket (>).

Paralinguistic features such as backchannels (e.g., "uhuh", or "mhm") can be signalled with codes like "mhm".

The following examples should make it clear:

A mhm well basically youre just going up towards the top of the screen
B okay
A <but move slightly first to the right and then slightly to the left until youre about/
B i dont understand ive gone up at an angle and now ive moved across in a straight line to the right now where do you want me to be you want me to go up>
A no you shouldnt have moved across to the right
B but you told me to
…….
A and then go left until youre level with the bottom hand corner of the cattle cattle stockade
B okay and im about a centimetre away from the edge of the screen maybe less
A slightly more yeah
B okay
A you just past it then just go a tiny bit past it
B mhm
A now go carry on to the left but go up
B how high up ?

Such simple codes enable counting of events such as number of turns, turns, length of turn (words per turn), interruptions, interruption rate, backchannels etc. It is a relatively straightforward programming task to design software to recognise the codes for behaviours of interest and produce statistics from a text transcription file.

However, it may also be important to capture the content of what is said. In this case researchers usually develop a coding scheme to capture the function of certain kinds of utterance. For example, in analysing dialogues from the map task given above, researchers have used what's called Conversational Games Analysis (see Kowtko, Isard & Doherty-Sneddon, 1991; Boyle et al., 1994; Doherty-Sneddon et al., 1997).

7.10 Communicative behaviour reference data

The reference values are based on data from students and staff at the university of Nottingham, UK age 18-48 (mean 24). Normal distribution of the variables can not be assumed.

Media Condition	description	Screen size	Screen resolution	Video delay	Audio delay
A	Synchronous	29”	CIF	200ms	200ms
B	Delayed	29”	CIF	650ms	650ms
C	Asynchronous	29”	CIF	400ms	200ms
D	Small screen CIF	3.5”	CIF	650ms	650ms
E	Small screen QCIF	3.5”	QCIF+Filter	650ms	650ms
F	Small screen avatar	3.5”	N/A	1000ms	1000ms
G	Text-chat	17”	N/A	N/A	N/A
H	Audio conference	N/A	N/A	N/A	< 40ms
I	Face to face	N/A	N/A	N/A	N/A
J	High quality video	17”	4CIF	40ms	40ms
K	Analogue video	10”	PAL	<40ms	<40ms
L	Large screen avatar	17”	N/A	1000ms	1000ms

Data from participants engaged in the Map task (Boyle, Anderson and Newlands, 1984)

		H-Audio conference	I-Face-to-Face	K-AnalogueVideo	A-Synch	C-Asynch
N		18	16	14	24	16
Turns	Mean	148	110	108	118.0	127.2
	Stdev	*83*	*60.5*	*43*	*51.5*	
	Median	*122*	*83.5*	*115*	*104.5*	
Words	Mean	1124	740.5	752	1077.9	1118.8
	Stdev	*575*	*424.5*	*306*	*546.3*	
	Median	*950*	*576.5*	*734*	*995.5*	
Giver words	Mean	736	492.6	485	334.4	299.2
	Stdev	*317*	*256*	*219*	*278.4*	
	Median	*728*	*412*	*525*	*287.0*	
Follower words	Mean	388	248	268	741.9	818.9
	Stdev	*304*	*193*	*130*	*296.5*	
	Median	*278*	*154*	*287*	*780.5*	
Words/turn	Mean	7.75	6.7	7.0	9.0	8.9
	Stdev	*1.16*	*0.7*	*1.3*	*1.7*	*1.5*
	Median	*7.9*	*6.8*	*6.4*	*9.0*	*8.8*
Giver words/turn	Mean	10.5	9.1	9.0	5.1	4.6
	Stdev	*2.5*	*1.5*	*2.6*	*2.0*	*1.2*
	Median	*10.8*	*8.8*	*8.9*	*4.7*	*5.0*
Follower words/turn	Mean	4.9	4.2	5.0	12.9	13.1
	Stdev	*1.57*	*1.4*	*2.0*	*2.9*	*3.1*
	Median	*4.7*	*4*	*4.4*	*12.5*	*12.4*
Interruptions	Mean	11.6	9.2	5.4	16.0	7.4
	Stdev	*11.3*	*6.7*	*4.1*	*12.3*	*5.4*
	Median	*8.5*	*7.0*	*4.0*	*15.5*	*6.5*
Giver interruptions	Mean				8.0	3.7
	Stdev				*6.1*	*2.7*
	Median				*7.5*	*3.7*
Follower interruptions	Mean				8.0	3.8
	Stdev				*6.3*	*2.7*
	Median				*8*	*3.8*

Data from "Acquiring a Company" task (Valley et al, 1998)

		D-Small screen CIF	E-Small screen QCIF	F-Small screen Avatar	B-Delayed	A-Synchronous
N		13	16	13	11	10
Turns	Mean	61.5	71.7	31.9	65.9	60.3
	Stdev	*32.8*	*31.5*	*8.6*	*26.3*	*26.7*
	Median	*49.0*	*61.0*	*31.0*	*69.0*	*56.0*
Words	Mean	1031.6	1115.1	802.5	875.3	1064.0
	Stdev	*585.4*	*382.8*	*265.2*	*483.2*	*364.0*
	Median	*910.0*	*1138.0*	*863.0*	*769.0*	*1083.5*
Buyer words	Mean	542.7	535.1	343.0	438.8	508.0
	Stdev	*412.8*	*239.2*	*163.4*	*286.6*	*194.7*
	Median	*321.0*	*518.0*	*380.0*	*368.0*	*516.5*
Seller words	Mean	488.2	579.7	459.5	435.2	553.2
	Stdev	*288.1*	*247.9*	*154.8*	*233.5*	*237.0*
	Median	*465.0*	*578.0*	*514.0*	*446.0*	*558.0*
Words/turn	Mean	17.4	16.3	25.1	12.6	17.8
	Stdev	*8.4*	*3.9*	*6.1*	*3.5*	*5.5*
	Median	*15.3*	*14.9*	*23.9*	*11.8*	*16.5*
Buyer wds/trn	Mean	17.7	15.8	21.2	13.2	18.8
	Stdev	*12.6*	*5.2*	*8.2*	*6.8*	*7.4*
	Median	*13.8*	*16.3*	*22.6*	*11.8*	*17.5*
Seller wds/tn	Mean	17.0	16.8	29.0	12.1	20.0
	Stdev	*9.5*	*6.0*	*8.3*	*3.0*	*7.3*
	Median	*16.8*	*17.4*	*28.6*	*11.1*	*20.1*
Interruptions	Mean	16.3	15.7	-	28.4	15.0
	Stdev	*15.9*	*13.7*		*27.1*	*15.2*
	Median	*14.0*	*12.0*	-	*19.0*	*12.0*
Buyer intrs	Mean	8.2	7.7	-	14.2	7.7
	Stdev	*7.9*	*6.7*		*13.3*	*7.8*
	Median	*7.0*	*6.0*	-	*10.0*	*6.5*
Seller intrs	Mean	8.2	7.9	-	14.2	7.3
	Stdev	*8.0*	*7.0*		*13.8*	*7.5*
	Median	*7.0*	*6.0*	-	*9.0*	*5.5*

		L-Large screen Avatar	I-Face-to-face	H-Audio Conferencing	J-High Quality Video	G-Text-chat
N		12	17	21	24	16
Turns	Mean	33.7	67.1	66.57	85.5	38.81
	Stdev	*14.4*	*37.2*	*52.88*	*41.34*	*22.51*
	Median	*36.5*	*58.0*	*49.00*	*70.00*	*30.50*
Words	Mean	974.1	976.5	1325.52	1394.79	627.69
	Stdev	*510.9*	*608.1*	*1430.90*	*1076.36*	*390.98*
	Median	*924.0*	*973.0*	*710.00*	*897.50*	*564.00*
Buyer words	Mean	420.8	476.8	627.29	614.13	313.13
	Stdev	*266.0*	*307.4*	*703.26*	*443.01*	*228.86*
	Median	*313.5*	*326.0*	*365.00*	*458.00*	*287.50*
Seller words	Mean	553.3	497.17	698.19	779.25	314.44
	Stdev	*337.6*	*349.4*	*820.07*	*692.67*	*187.84*
	Median	*464.0*	*367.0*	*365.0*	*486.50*	*277.50*
Words/turn	Mean	29.6	14.43	17.30	15.56	16.76
	Stdev	*13.4*	*4.7*	*8.06*	*7.67*	*7.50*
	Median	*26.3*	*14.3*	*14.73*	*13.26*	*14.69*
Buyer wds/trn	Mean	24.5	14.4	16.65	13.83	16.85
	Stdev	*11.2*	*5.3*	*11.13*	*5.78*	*8.48*
	Median	*21.8*	*12.3*	*14.54*	*13.23*	*14.48*
Seller wds/tn	Mean	34.8	14.3	17.95	17.18	16.56
	Stdev	*19.4*	*6.6*	*9.88*	*11.22*	*8.20*
	Median	*28.4*	*12.7*	*17.10*	*13.20*	*14.99*
Interruptions	Mean	-	5.8	5.67	26.92	10.38
	Stdev		*6.8*	*7.07*	*27.01*	*11.42*
	Median	-	*2.0*	*2.00*	*16.00*	*5.50*
Buyer intrs	Mean	-	2.9	2.62	13.50	5.06
	Stdev		*3.4*	*3.40*	*13.41*	*5.60*
	Median	-	*1.0*	*1.00*	*8.00*	*2.50*
Seller intrs	Mean	-	2.8	2.57	13.42	5.31
	Stdev		*3.3*	*3.41*	*13.62*	*5.83*
	Median	-	*1.0*	*1.00*	*8.00*	*3.00*

8. References

Bickman & D. J. Rog (Eds). Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications. P101-126

Boyle, E., Anderson, A. & Newlands, A. (1994) The effects of eye contact on dialogue and performance in a co-operative problem solving task. *Language and Speech*, *37*, 1-20.

Brooks Peter et. al (1999). *Final report.* AC314 Vis-a-Vis: Fitness-for-Purpose of Videotelephony in Face-to-Face situations, CEC deliverable A0314/NSSL/PB/DR/P/005/b1

Cook, T.D. and D. Campbell (1979) Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand-McNally.

Følstad, A. et. al. (2002). *Results of Field Experiments of Communication Media.* IST Project 1999-11577 Eye-2-Eye: Fitness-for-Purpose of Person-Person Communication Technologies, CEC Deliverable IST11577/SEF/DIS/DS/Pub/004/b1

Heim, J. et al. (2000). *Initial Verification of Real-time Communication Requirements.* IST Project 1999-11577 Eye-2-Eye: Fitness-for-Purpose of Person-Person Communication Technologies, CEC Deliverable IST11577/SEF/DIS/DS/5FP/001/b1, July 2000.

Heim, J. et. al. (2001). *Final verification of Real-time Communication Requirements.* IST Project 1999-11577. Eye-2-Eye: Fitness-for-purpose of Person-Person Communication Technologies, CEC Deliverable IST11577/SEF/DIS/DS/Pub/001/a1, October 2001

McGrath, J. E. A typology of tasks .In: Baecker, R. M., Readings in groupware and computer-supported cooperative work, assisting human-human collaboration, pp. 165-168 Published by: Morgan Kaufmann Publishers, Inc., 1993 ISBN: 1-55860-241-0

Mulbach, L., Bocker, M. & Prussog, A. (1995). *Telepresence in videocommunications: A study on stereoscopy and individual eye contact.* Human Factors, 37(2), 290-305.

O'Malley, C. et al (2002). *Results of Laboratory Experiments of Communication Media.* IST Project 1999-11577 Eye-2-Eye: Fitness-for-Purpose of Person-Person Communication Technologies, CEC Deliverable IST11577/UON/SOP/DS/Pub/003/b1, Dec 2002.

Schliemann, T. et al. (2001). *Results of Baseline Communication Experiments.* IST Project 1999-11577 Eye-2-Eye: Fitness-for-Purpose of Person-Person Communication Technologies, CEC Deliverable IST11577/SEF/DIS/DS/5FP/002/b1, July 2001.

Schliemann, T., Asting, T., Følstad, A., Heim, J. (2001). *Medium Preferences and Medium Effects in Person-Person Communication.* ACM CHI'02 Proceedings, pp. 710-711

Short, J., Williams, E. & Christie, B. (1976) *The Social Psychology of Telecommunications*. Wiley.

Valley, K. L., Moag, J., & Bazerman, M. H. (1998). 'A matter of trust': Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior & Organization*, *34*, 211-238).

Wiklund, Michael (ed.)(1994). *Usability in Practice.* London: Academic Press

ETSI EG 201 472 (2000) "Usability evaluation for the design of telecommunication systems, services and terminals"

ITU-T P.910 (1996) "Subjective video quality assessment methods for multimedia applications"

ITU-T 9.920 (1996) "Interactive test methods for audio-visual communications"

ETNA Project Report (2001) "New techniques for assessing audio and video quality in real-time interactive communications" IHM-HCI 2001, Lille, France. (http://www-mice.cs.ucl.ac.uk/multimedia/projects/etna/)

www.UsabilityNet.org

9. Glossary of main Eye-2-Eye terminology and concepts[7]

Acceptable price: The price that end-users are willing to pay for a particular communication service or for improved quality of service. The price of telephony (equipment as well as service) should be used as a benchmark when asking (potential) end-users about acceptable price; and the latter should be expressed as a percentage of the price of telephony, e.g. 50% (half the price of telephony), 300% (three times the price of telephony), etc.

Asynchrony: When audio and video information that leaves one communicating party at the same time is received by the other communicating party at different times (e.g., typically the audio information arrives before the video information in an asynchronous situation)

Audio telephony: An 'ordinary' telephone service as distinct from Audio conferencing

Audio conferencing: A telephone service that does not rely on amplification of the voice signal in very close proximity to the recipient's ear

Avatar telephony: A service for transmitting voice signals in real-time over a telecommunication network in combination with a graphical (human) representation of the speaker

Benefits: Benefits to the end-users from using a particular communication service (e.g. savings of travel time and costs, achievement of task goals, ease of use, easy accessibility to the called party, increased communication quality and effectiveness, etc.).

Communication activity: What the end-users (want to) do with a communication service (e.g. social chatting, buying or selling shares, conducting a job interview, etc.).

Communication media: Types of information with which humans communicate. Examples are text, audio, moving image (video, moving graphics) and still image.

Communication service: A service that is provided via a telecommunication network. Examples are ordinary telephony, email, videoconferencing, avatar telephony, audio conferencing.

Communication situation: The combination of task, motive, content and user (group) characteristics.

Communicative behaviour: End-user behaviour while using a communication service, including turn taking, interruptions, verbal and non-verbal back-channels and gaze.

Conference: used as follows within the scope of Eye-2-Eye: (a) From a technical orientation a point-to-point connection (i.e., there were no studies of multi-point connection); From a service orientation it is always person (or group)-to-person (or group) communication.

Costs: Costs that the end-user has to pay for using a particular communication service. These include not only financial costs but also subjective costs; e.g. the user may see loss of privacy as one of the costs to pay for having a videophone.

Duplex: A mode of operation by which information can be transmitted in both directions simultaneously between two points.

Dyadic: (Distance) communication between two people

Effectiveness (ISO 9241 definition): The accuracy and completeness with which specified users can achieve specified goals in particular environments.

[7] This is a general list for the Eye-2-Eye project as a whole and is not restricted specifically to this document.

Efficiency (ISO 9241 definition): The resources expended in relation to the accuracy and completeness of goals achieved.
End-users: The people who use a communication service for person-to-person communication.
End-users: The people who use the communication service*(s)*.
Fitness-for-Purpose: The correct balance between technological performance and human performance, such that the interaction is both sufficient and beneficial for person-person communication and consistent with human expectations from face-to-face communication.
Frame rate: The frequency by which a full video frame is updated, sometimes called video temporal resolution or image frequency.
Group: (Distance) communication between three or more people.
Half-duplex: A mode of operation where, at a given instance, only one of the two correspondent information streams is transmitted.
Interpersonal perception. The extent to which the perception of the other person's attributes (how likeable, intelligent, friendly etc.) is positive or negative.
Media effects: The effect a particular communication medium has on an end-users task outcome, communicative behaviour, attitudes and beliefs.
Media preferences: The subjective assessment by users or user groups of when a given communication medium is preferred over another.
Multimedia conferencing: A service for transmitting voice, video and data signals in real-time over a telecommunication network
Multi-point: Distance communication between three or more locations
Packet loss: A loss of one packet that can be described using a certain statistical model
Pilot study: A small "pre-stage study" done with a few participants when the draft test materials are produced. This is done to evaluate the materials, logistics and potential outcome of a draft test plan. Results of a pilot study will normally lead to refinement of the test plan, and in some occasions to discarding the whole test concept.
Point-to-Point: Distance communication between two locations
Quality of service: Those aspects of the service which are assumed to affect the degree of satisfaction of the user of the service (e.g. the number of frames per second in videoconferencing, the auditory bandwidth in audio conferencing).
Resolution: A term denoting the degree of detail which can be created by a particular visual display system
Satisfaction (ISO 9241 definition): The comfort and acceptability of the work system to its users and other people affected by its use.
Target audience: The people or organisations who are going to use the fitness-for-purpose guidelines, the cost-benefit analysis tool and/or the fitness-for-purpose evaluation toolkit.
Task elements: Features of *tasks* that can be expected to vary (e.g., extrinsic-intrinsic origin, symmetrical-asymmetrical balance, originator-recipient role, ego involvement level, information dependency, sociability level)
Task goal: The aim or object towards which the communication is directed. It is what end-users want to do with the *communication technology* (e.g. social chatting, buying or selling shares, conducting a job interview, etc.).
Task outcome: The extent to which task performance dependent on the medium

Task: What users of *communicative technology* actually do in order to accomplish some *task goal*. In experiments tasks may be described to the participants or they are embedded in scenarios as a part of a *situation*.
Usability (ISO 9241 definition): The *effectiveness*, *efficiency*, and *satisfaction* with which specified users achieve specified goals in particular environments.
User groups: End-users who with respect to their usage of communication service may be grouped together (e.g. business executives, university students, grandparents, deaf people, etc.).
Videoconferencing: A service for transmitting voice and video signals in real-time over a telecommunication network
Videotelephony: See Videoconferencing.
Willingness to pay: An end-users willingness to pay in financial terms for a given communication service in a given situation.

Appendices: Tasks and procedures used in the Eye-2-Eye project

A number of tasks have been used in the Baseline and Laboratory studies of the Eye-to-Eye project. Some of the tasks have been developed especially for the purpose of the project, and other tasks have been adopted from previous research.

The tasks that have been adapted from previous research are:

- The "Map task" (Boyle et.al., 1994)
- The "Acquiring a Company" game (Valley et al, 1998)

The tasks that have been developed specially for the project are:

- Consensus-making through web-evaluation

Participants were shown a series of six web-pages, and were asked to assess the web pages on Content and Appearance. Participants were discussing the web-pages between them before giving an individual rating.

- Willingness-to-lie:

A simple "bluffing-game" with cards. Participants were given a "prepared" deck of cards, and played one by one card against each other. A total of 20 cards were played, and the higher card in each round won a sum of money. Participants had always the option to lie about the card they were possessing. Calling and revealing a bluff, would cause the bluffer to loose money.

- Negotiation:

A simple negotiation task was developed. A baker and a merchant will share costs and profit on a shipping of goods from the "far east". They both have the same list of goods to choose from (with a predefined, common price), but the expected profit from each type of goods varies between the two negotiators. The task aim was to agree upon a selection of goods to be shipped with the boat. The "hidden" task characteristics is the fact that most goods have a high vs. low profit for the baker and the merchant respectively, while some goods have a moderately high profit for both. Another factor was that the merchant had a slightly better "profit-matrix" than the baker, and hence a better chance of high profit.

- Remote Inspection task

In this task, the aim is to make a number of connections between different pins on a circuit board in order to repair it.

The following appendices present examples of tasks and corresponding task instructions used in the Eye-to-Eye project. These examples show a set of possible tasks to use in communication studies in the laboratory, and the instructions included show the necessary level of instructions to be given to the participants.

The actual instructions and solving of the laboratory task is only a part of what is going on during a laboratory test session. The context of which the laboratory task is solved in, as well as the pre-work and post-work done by the test leader, depends on the overall approach of laboratory study. Which additional data to be collected (pre- and post-task questionnaires, different subjective measures, etc.) also depend on the overall approach for the study.

Appendix 1 describes a full version of the test procedure used for conducting experiment 4 in the Eye-to-Eye Baseline studies, involving “the map-task” (Boyle et. al., 1994). This section serves as a full-scale example of how to plan and perform a laboratory study in detail, including task instructions given to the participants. The next appendices refer only to the tasks and corresponding task instructions used, and these tasks should consequently be included in a complete test procedure context similar to the one described in appendix 1.

Appendix 1: Test procedure for Baseline study 3 - The map task

The following is the full description of test procedure used for conducting laboratory study 4 in the Eye-to-Eye Baseline studies. It is a detailed description of every step in the process of performing one user test with two participants communicating over video-link[8], including the wording of the instructions and information given by the test leader to the participants.

When more than one test leader is involved in the tests, such documentation is crucial, in order to ensure equal treatment of all the participants in the study.

Use of letters A and B throughout the test procedure refers to test participants A and B, room A and B, and documents prepared exclusively for participants A and B.

Preparations

- Have ready the following test materials
  - Experiment 4 Checklist
  - Pre-interview form(2x)
  - Video consent form (2x)
  - Instructions B
  - Map 1-B
  - Instructions A
  - Map 1-A
  - Map 2-B
  - Map 2-A
  - Person Perception questionnaire (2x)
  - Payment receipt (2x)

- Prepare test rooms and waiting area
  - Prepare test rooms according to technical checklist[9] for the video condition
  - Tidy all the rooms to which the test participants have access
  - Provide pens and "waiting literature" in both test rooms
  - Coffee and cold drinks in the waiting area

- Prepare technical equipment
  - Prepare, label and insert new videotape, ready for recording (technical checklist)
  - Check that the video set-up works as required (technical checklist)
  - Perform calibration of echo canceller (technical checklist)
  - Note context- and test information in the Experiment 4 Checklist

[8] When technical settings vary with the different test conditions, it might be necessary with separate test procedures for the different conditions.

[9] "Technical checklist" is a separate checklist, describing set-up instructions and all necessary parameter settings for making the whole set of technical equipment work as required - one list for each technical test condition.

Pre-interviewing the test participants

- Receive participant B at the backdoor[10]
  - Be there in time before the test person arrives
  - Welcome participant B
  - Show B around the lab area, and introduce other scientific personnel
  - Offer coffee/cold drinks

- Take B to room B (bring pre-interview and video accept forms)
  - Read pre-instructions:
    It is very much appreciated that you could find the opportunity to participate in this test. The test you are about to take part in is part of a project where we look at how communication between two persons works with the use of different media. You will communicate with another person by means of a video-phone. You will sit here, and the image of your communication partner will be seen there (pointing). You will be given a task to solve together with the other person.

    You are here as a volunteer test person. Should you for any reason during the test feel uncomfortable about the situation, you are free to discontinue at any time. I will also remind you that we are not testing your capability or skills in interacting with the technology, but rather how suitable the technology is for communicating.

    The whole test will take about an hour, including filling in some questionnaires. We prefer the test to be done in one go, so please prepare yourself for this.

    Any questions?
    Ok, before we start, I would like to register some information about you.

  - Pre-interview B
  - Fill in video acceptance form for B
  - Prepare B for a bit of waiting, and refer to the ”waiting literature”

- Wait for A to arrive at the front door

- Receive participant A at the front door
  - Welcome participant A
  - Show B around the lab area, and introduce other scientific personnel. (N.B: Make sure A doesn’t see B)
  - Offer coffee/cold drinks

- Take A to room A (bring pre-interview and video accept forms)
  - Read pre-instructions:
    As above
  - Pre-interview A
  - Fill in video acceptance form for A
  - Refer to the ”waiting literature”

[10] Participants were arriving at separate entrances, in order to not meet each other before the test

Task 1

- Enter room B
  - Ask B to turn off any cellular phone
  - Distribute map 1-B, instructions B, and read instructions 1-B:
    Your map was drawn by an explorer to provide a route to some treasure buried at the finish point. The other person in the experiment has a map but there is no route drawn on it.

    Your task is to explain the route to the other person, as quickly and accurately as possible. However, the two maps were drawn by different explorers and some of the landmarks on the maps may slightly differ.

    After you have completed this task, you will solve a similar task, but with the roles swapped.

    Any questions?

  - If there are any questions, repeat the relevant part of the instructions - without elaborating on it.
  - Ask B to wait for a short while, and inform that the test will start when you call on the intercom.

- Enter room A
  - Ask A to turn off any cellular phone
  - Distribute map 1-A, instructions A, and read instructions 1-A:
    Your map was drawn by an explorer to provide a route to some treasure buried at the finish point. However, there is no route drawn on your map, only land marks.

    The other person in the experiment has a map with a route drawn on it. However, the two maps were drawn by different explorers, and some of the landmarks on the maps may slightly differ.

    Your task is, with the help of the other person, to draw the route onto your map as quickly and accurately as possible.

    After you have completed this task, you will solve a similar task, but with the roles swapped.

    Any questions?

  - If there are any questions, repeat the relevant part of the instructions - without elaborating on it.
  - Ask A to wait for a short while, and inform that the test will start when you call on the intercom.

- Test participants solve task 1
  - start video recorder
  - connect the video-link
  - over intercom: Tell participants to start
  - start stop watch
  - [Test participants do the task]
  - When finished, stop the watch and register used time in the form
  - Inform over intercom that you will come shortly with new tasks
  - disconnect the video-link
  - stop video recorder

Task 2

- Enter room B
  - Collect Map 1-B
  - Distribute map 2-B, and read instruction 2-B:
    You have now received a new map. Roles have been swapped, so this time there is no route on your map. Your task is, with the help of the other person, to draw the route onto your map as quickly and accurately as possible.

    Also this time, some of the landmarks on the maps may slightly differ.

    Any questions?

  - If there are any questions, repeat the relevant part of the instruction - without elaborating on it.
  - Ask B to wait for a short while, and inform that the test will start when you call on the intercom.

- Enter room A
  - Collect map 1-A
  - Distribute map 2-A, and read instructions 2-A:
    You have now received a new map. Roles have been swapped, so this time your map has a route drawn on it. Your task is to explain the route to the other person, as quickly and accurately as possible.

    Also this time, some of the landmarks on the maps may slightly differ.

    Any questions?

  - If there are any questions, repeat the relevant part of the instructions - without elaborating on it.
  - Ask A to wait for a short while, and inform that the test will start when you call on the intercom.

- Test participants solve task 2
  - start video recorder
  - connect the video-link
  - over intercom: Tell participants to start
  - start stop watch
  - [Test participants do the task]
  - When finished, stop the watch and register used time in the form
  - Inform over intercom that the task is over, and ask them to wait a few minutes.
  - disconnect the video-link
  - stop video recorder

Person Perception questionnaire

- Enter room B
  - collect Map 2-B and Instruction-B
  - Distribute person perception questionnaire form, and read the instructions:
    As described in the questionnaire

    Any questions?

- Enter room A
  - collect Map 2-A and Instruction-A
  - Distribute person perception questionnaire form, and read the instructions:
    As described in the questionnaire

    Any questions?

Closing up

- Collect Person Perception forms, and bring both participants to the waiting area
- Debriefing. You are allowed to tell:
  - the aim of the testing is to investigate the use of different services for different purposes or tasks
  - briefly present the different services involved
  - explain that there are different tasks, but don't get into details
- If participants want, show them their routes drawn on the maps
- Payment and receipts
- Register participants e-mail address, in case they want results sent to them
- Ask the participants not to tell other colleagues/students about the test
- Thank you and goodbye

After the test

- Eject the videotape, label and archive
- Fill in the Experiment 4 Checklist
- Collect all test papers and staple them together
- Prepare the rooms for next test session

Appendix 2: Test Materials - The Map task

Instruction - the follower

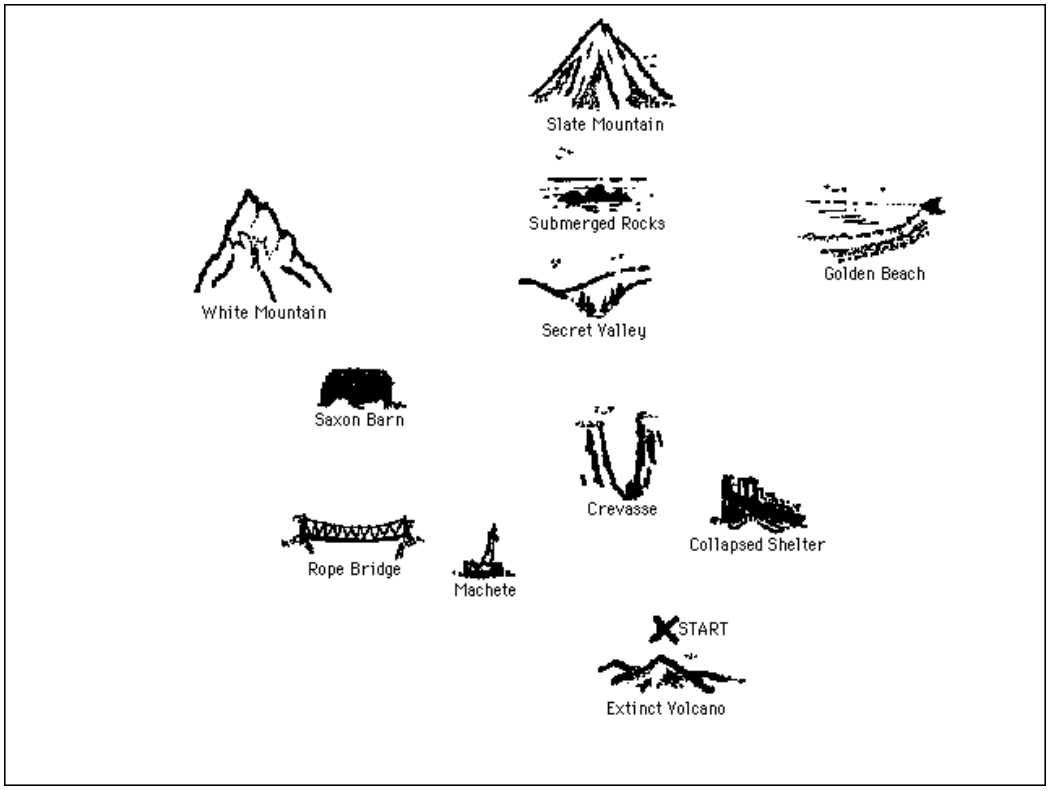
The map you have in front of you was drawn by an explorer to show the way to a treasure buried at the end of the route. There is, however, no route on the map - only landmarks.

The other person in the experiment has a map of the same area, with a route drawn on it. The two maps were, however, drawn by two different explorers, so some of the landmarks might be different.

Your task is, with help from the other person, to draw the route onto your map, as quickly and accurate as possible. The route doesn't have to be a solid line, but can, if preferred, be a dashed line.

When the task is solved: Turn the sheet around, to show that you're finished.

Map - The follower



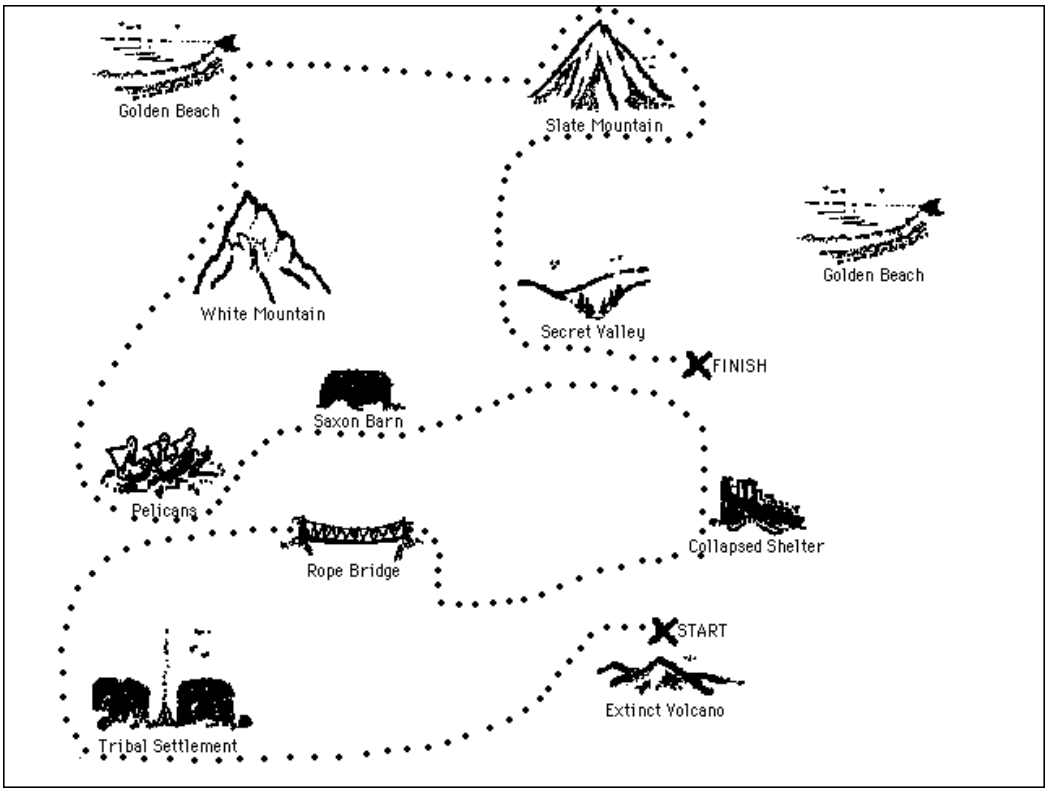
Instruction - The giver

The map you have in front of you was drawn by an explorer to show the way to a treasure buried at the end of the route. The other person in the experiment has a map of the same area, but there's no route drawn onto the map.

Your task is to explain the route to the other person, as quickly and accurate as possible. The two maps were, however, drawn by two different explorers, so some of the landmarks might be different.

When the task is solved: Turn the sheet around, to show that you're finished.

Map - The Giver



Appendix 3: Task Materials - Acquiring a Company

Instructions – buyer

You represent Company A (the acquirer), which is considering acquiring Company T (the target). A mutually acceptable price must be negotiated for the acquisition to take place. You are going to meet a representative from Company T to determine if you can reach an agreement that would be acceptable for you and Company T.

You are unsure about how high a price you are willing to pay, and the main complication is the following: The value of Company T depends directly on the outcome of a major oil exploration project, and no one outside the firm has any information on the results of this exploration effort. If the project was a total failure, the company under Company T's management could be worth as little as £0/share. But if the project was a total success, the value of the company under Company T's management could be as high as £100/share. In fact, the value of the firm is equally likely to be any amount between £0 and £100 (per share). Thus, 0, 1, 2, … , 98, 99 and £100 (per share) are all equally likely. However, Company T may know the actual, true value of the firm.

By all estimates, the company will be worth considerably more in the hands of Company A than under Company T management. In fact, whatever the ultimate value under Company T management, *the company will be worth 50% more under the management of Company A than under Company T.* If the company was worth £20/share under Company T's management, the value under Company A would £30/share. If the company was worth £50/share under Company T's management, the value under Company A would be £75/share. Similarly, if the company was worth £80/share under Company T's management, the value under Company A would be £120/share.

The board of directors of Company A has asked you to try to negotiate the profitable purchase of Company T from a representative of Company T. From all indications, Company T would be happy to be acquired by Company A, *provided it is at a profitable price*.

Thus, you (Company A) *will not know the results of the exploration project before negotiating with Company T. However, Company T is likely to have more information about their true value. In addition, Company T is expected to prefer any agreement that provides them with a price higher than their true value.*

As the representative of Company A, you want to negotiate to maximise the interests of Company A. If you acquire Company T, your success in this negotiation will be evaluated on the amount by which the value of the firm *under Company A's management* is greater than the amount you pay for the company.

You may only agree on a fixed Pound price per share. No other conditions or terms may be added to the agreement. You may reveal any or all of the information provided. But, you may not show the representative of Company T this sheet of paper.

Payment Details – Company A

Company A's profit is the value of the company under Company A's management, minus the purchase price:

Company A's profit = (Value under Company A's management – Purchase price)

– Example: If the value of the company to them (company T) is £60, after purchase, the value to you (company A) is £90. If you buy it for £70, you will make a profit of (£90-£70) = £20

– If the value of the company to them (company T) is £10, after purchase, the value to you (company A) is £15. If you buy it for £50, you will make a loss of (£15-£50) = £35

– Do not pay more for the company than it is worth to you!

Instructions – seller

You represent Company T (the target). Company A (the acquirer) is considering acquiring your company if a mutually acceptable price can be negotiated. You are going to meet a representative from Company A to determine if you can reach an agreement that would acceptable for you and Company A.

Company A has some information about your company. They know that Company T has been engaged in a highly risky oil exploration project. They realise that the value of the company depends directly on the outcome of your recent exploration project. No-one outside the Company T (including Company A) has any information on the results of this exploration effort. Going into the project, experts inside and outside the company knew that if the project was a total failure, the company under Company T's management could be worth as little as £0/share. But if the project was a total success, the value of the company under Company T's management could be as high as £100/share. In fact, all values between 0 and £100 (per share) were viewed as equally likely. However, the results of the exploration project are now known. You know that the company is worth £40 under Company T's management. Company A does not have this information.

By all estimates. The company will be worth considerably more in the hands of Company A than under Company T's management. However, you do not know how much more the company will be worth if owned by Company A.

The board of directors of Company T has asked you to try to negotiate the profitable sale of Company T with the representative of Company A. As the representative of Company T, you want to negotiate to maximise the interests of Company T. Your success in this negotiation can be assessed by how much more than £40 (per share) you can get Company A to pay for the company. Obviously, you would not accept any offer below £40

You may only agree on a fixed Pound price per share. No other conditions or terms may be added to the agreement. You may reveal any or all of the information provided. But, you may not show the representative of Company A this sheet of paper.

Payment details – Company T

Company T's profit is the sale price minus the value of the company under T's management:

Company T's profit = (sale Price – Value prior to purchase)

– Example: You are Company T and the value of your company to you, before a sale, is £40. If you sell it for £100, you will make a profit of (100-40) =£60

– Do not sell the company for less than it is worth to you!

Appendix 4: Test Materials - Consensus-making through web-evaluation

Test Instruction

In this task, we are interested in your assessment of some selected web-sites. No special knowledge about Internet is required to do this test.

We are interested in your immediate impression of the web-site appearance. You will be shown different web front pages on the PC-screen. You shall not click on the links on the pages, only look at them. The web-sites will be presented one by one, and you shall discuss each web-site with the other test participant, to better be able to make up your mind about what you think of the web-sites.

You will have approx. one minute to discuss each site. Afterwards, please indicate your assessment of the web-site on the assessment form.

The web sites shall be assessed on two scales: Content and Appearance. It is of course impossible to say something complete about the site content and appearance based on a single front-page, but we are only interested in your immediate impression of the front-page.

Meaning that the discussion should be concentrated around the following issues:

- How appealing is the appearance of the web-site?
- How exiting or interesting does the content of the site seem to be?

After the discussion, you shall give your individual assessment of the web-site on the form you have received. You must not tell each other what you have marked off, and your assessment doesn't have to be similar to the other persons assessment.

A total of six web-pages shall be assessed.

Web evaluation form

Indicate your evaluation of each web-site by the scales below. There are two scales for each site to be evaluated.

site 1

	very disagreable			neutral			very agreeable
Design	❍	❍	❍	❍	❍	❍	❍

	very disagreable			neutral			very agreeable
Content	❍	❍	❍	❍	❍	❍	❍

site 2

	very disagreable			neutral			very agreeable
Design	❍	❍	❍	❍	❍	❍	❍

	very disagreable			neutral			very agreeable
Content	❍	❍	❍	❍	❍	❍	❍

site 3

	very disagreable			neutral			very agreeable
Design	❍	❍	❍	❍	❍	❍	❍

	very disagreable			neutral			very agreeable
Content	❍	❍	❍	❍	❍	❍	❍

site 4

	very disagreable			neutral			very agreeable
Design	❍	❍	❍	❍	❍	❍	❍

	very disagreable			neutral			very agreeable
Content	❍	❍	❍	❍	❍	❍	❍

site 5

	very disagreable			neutral			very agreeable
Design	❍	❍	❍	❍	❍	❍	❍

	very disagreable			neutral			very agreeable
Content	❍	❍	❍	❍	❍	❍	❍

site 6

	very disagreable			neutral			very agreeable
Design	❍	❍	❍	❍	❍	❍	❍

	very disagreable			neutral			very agreeable
Content	❍	❍	❍	❍	❍	❍	❍

Appendix 5: Test Materials - Willingness to Lie

Test Instructions

In this task, you will participate in a simple card game where you can win money. You will receive a part of a pack of cards, and the same will your opponent. The cards must not be shuffled, but stay in the same order as they are given to you.

The game runs over several rounds. One round consist of both of you presenting one card each. Each round starts with the test leader asking you to pick up the top card from the deck. You may look at the card, but not show it to your opponent.

After having looked at the card, you shall tell each other which card you have. You can choose whether to lie or tell the truth. E.g. if you pick up a 5 of Clubs, you can tell the truth (5 of clubs) or lie (10 of spades). You declare your card one at the time, and the order between you is changed for each round.
The player who presents the higher card wins! (unless he is caught in lying)

Ace is the highest card. If you both present cards of the same value, the round is annuled. You don't have to show your card to your opponent after finishing the round (unless you are accused of lying).

You both start the game with 7 pounds. The one who looses the round pays the other 1 pound.

If you think your opponent lies about his card, you can accuse him of lying. The accusation must be made immediately after your opponent has declared his card. When one of the players is accused of lying, the round is terminated, and the accused player must show his card (tell the real value of the card).

If you are right in your accusation about lying, your opponent will have to pay you 2 pounds. If you are wrong (i.e. your opponent told the truth about his card), you have to pay your opponent 2 pounds.

So, you can lie to make your opponent believe you have a higher card than you actually have, but the punishment for being caught is hard. Equally, the punishment is hard for accusing your opponent of lying when he actually tells the truth.

After each round (either played or terminated by an accusation) the test leader will update you on the pay-offs and "financial standings".

A total of 20 rounds are played. Payment is done after the test is finished.

Test leader's score-sheet

	A			B			Accumulated winnings	
	A's card	***A says***	*Blame* / *Expose*	***B's card***	***B says***	*Blame* / *Expose*	***A's money (start: £7)***	***B's money (start: £7)***
1 (A)	H-6		B ❑	C-9		B ❑		
			E ❑			E ❑	=	=
2 (B)	S-2		B ❑	H-7		B ❑		
			E ❑			E ❑	=	=
3 (A)	C-12		B ❑	C-3		B ❑		
			E ❑			E ❑	=	=
4 (B)	S-8		B ❑	D-13		B ❑		
			E ❑			E ❑	=	=
5 (A)	D-9		B ❑	D-6		B ❑		
			E ❑			E ❑	=	=
6 (B)	H-10		B ❑	S-7		B ❑		
			E ❑			E ❑	=	=
7 (A)	C-7		B ❑	H-8		B ❑		
			E ❑			E ❑	=	=
8 (B)	S-10		B ❑	C-8		B ❑		
			E ❑			E ❑	=	=
9 (A)	D-8		B ❑	C-10		B ❑		
			E ❑			E ❑	=	=
10 (B)	D-7		B ❑	S-6		B ❑		
			E ❑			E ❑	=	=

	A			B			Accumulated winnings	
	A's card	***A says***	*Blame* *Expose*	***B's card***	***B says***	*Blame* *Expose*	***A's money (start: ??)***	***B's money (start: ??)***
11 (A)	D-9		B ❑	C-6		B ❑		
			E ❑			E ❑	=	=
12 (B)	C-7		B ❑	H-2		B ❑		
			E ❑			E ❑	=	=
13 (A)	D-3		B ❑	D-12		B ❑		
			E ❑			E ❑	=	=
14 (B)	S-13		B ❑	H-8		B ❑		
			E ❑			E ❑	=	=
15 (A)	S-6		B ❑	S-9		B ❑		
			E ❑			E ❑	=	=
16 (B)	H-7		B ❑	C-10		B ❑		
			E ❑			E ❑	=	=
17 (A)	C-8		B ❑	D-7		B ❑		
			E ❑			E ❑	=	=
18 (B)	D-8		B ❑	H-10		B ❑		
			E ❑			E ❑	=	=
19 (A)	D-10		B ❑	S-8		B ❑		
			E ❑			E ❑	=	=
20 (B)	H-6		B ❑	S-7		B ❑		
			E ❑			E ❑	=	=

Total sscores

	A	B	Total
Number of lies			
Number of blames (B)			
Number of exposings (E)			

Appendix 6: Test Materials - Negotiation

Instruction - the baker

A baker and a merchant have got an offer to make a joint investment: They will hire a cargo ship to carry goods from the far east. When the ship arrives, they will split the goods in two equal parts. They have agreed to pay half the total cost each, but they have not yet agreed about what the ship shall carry.

In this situation, you are the baker. The other person in the experiment is the merchant. The cargo shall contain five of the goods in the list below. When the cargo ship arrives, the goods will be split this way: Each of you will get half of each type of goods.

You as a baker, have a high profit on some of the goods, and a low profit on other goods. Your total profit will be the sum of profits for each type of goods that you agree to bring home.

Your task is to negotiate a choice of goods that gives you the highest possible profit, and at the same time is acceptable to the merchant. You can make notes on a piece of paper while you discuss.

Goods	**Your profit for each type of goods**
Flour	900
Cinnamon	900
Sugar	800
Pots and pans	650
Coal	650
Coffee	500
Butter	500
Raisins	500
Almonds	500
Pepper	400
Candies	400
Furs	200
Soft soap	0
Christmas tree decorations	0

Selected goods:

1:
2:
3:
4:
5:

Instruction - the merchant

A baker and a merchant have got an offer to make a joint investment: They will hire a cargo ship to carry goods from the far east. When the ship arrives, they will split the goods in two equal parts. They have agreed to pay half the total cost each, but they have not yet agreed about what the ship shall carry.

In this situation, you are the merchant. The other person in the experiment is the baker. The cargo shall contain five of the goods in the list below. When the cargo ship arrives, the goods will be split this way: Each of you will get half of each type of goods.

You as a merchant, have a high profit on some of the goods, and a low profit on other goods. Your total profit will be the sum of profits for each type of goods that you agree to bring home.

Your task is to negotiate a choice of goods that gives you the highest possible profit, and at the same time is acceptable to the baker. You can make notes on a piece of paper while you discuss.

Goods	**Your profit for each type of goods**
Soft soap	900
Christmas tree decorations	900
Sugar	800
Pepper	700
Candies	700
Coffee	500
Butter	500
Raisins	500
Almonds	500
Pots and pans	450
Coal	450
Furs	200
Flour	0
Cinnamon	0

Selected goods:

1:
2:
3:
4:
5:

Appendix 7: Test Materials - Remote inspection task

Instruction - Expert

In this task, the aim is to make a number of connections between different pins on a circuit board in order to repair it.

One person (the "novice") is seated in front of the circuit board and has to make the actual physical connections between the pins using wires with crocodile clips at the ends. The other person (the "expert") is seated in another room and has a set of pictures indicating which pins are to be connected together. There is an audio link between the 2 rooms, like a handsfree telephone, and the novice has a camera which they can use to show the expert various views of the circuit board so that they can find the pins identified in the pictures.

You will play the role of the expert, and will help the novice at the circuit board find the correct pins to connect together. You should try to work co-operatively with the expert to find out which pins need to be connected together. You will be able to see various views of the circuit board as shown to you by the novice with the video camera.

Over the course of the task, the image quality you see will change due to fluctuating network parameters. This is a normal part of the experiment and is not a fault with the equipment.

There are 8 pairs of pins to connect, and you should work through the pairs in the order that they are given to the expert. Try to be as sure as you possibly can be about each connection that you make – do not try to correct mistakes that you make if you have already moved on to another pair of pins. However, if you find that you have used a pin earlier that you decide was connected in error, you can disconnect it and use for a new connection, but DO NOT GO BACK and try to find the correct connection for the earlier pin.

The experiment ends when you have connected all 8 pairs of pins. You can use whatever strategy you want to try to match the pins with the pictures that the expert has. If you have any questions during the task, please feel free to say so and the experimenter will come in and try to help.

If more than 45 minutes elapses before the task is completed, we will stop the experiment.

Instruction - Novice

In this task, the aim is to make a number of connections between different pins on a circuit board in order to repair it.

One person (the "novice") is seated in front of the circuit board and has to make the actual physical connections between the pins using wires with crocodile clips at the ends. The other person (the "expert") is seated in another room and has a set of pictures indicating which pins are to be connected together. There is an audio link between the 2 rooms, like a handsfree telephone, and the novice has a camera which they can use to show the expert various views of the circuit board so that they can find the pins identified in the pictures.

You will play the role of the novice, and will make the connections on the circuit board with the wires provided. You should try to work cooperatively with the expert to find out which pins need to be connected together. You can use the camera to show the board to the expert.

Over the course of the task, the image quality seen by the expert will change due to fluctuating network parameters. This is a normal part of the experiment and is not a fault with the equipment. You will not see the effects of these fluctuations on your screen, but you may get complaints from your expert if the image deteriorates too badly.

There are 8 pairs of pins to connect, and you should work through the pairs in the order that they are given to the expert. Try to be as sure as you possibly can be about each connection that you make – do not try to correct mistakes that you make if you have already moved on to another pair of pins. However, if you find that you have used a pin earlier that you decide was connected in error, you can disconnect it and use for a new connection, but DO NOT GO BACK and try to find the correct connection for the earlier pin.

The experiment ends when you have connected all 8 pairs of pins. You can use whatever strategy you want to try to match the pins with the pictures that the expert has. If you have any questions during the task, please feel free to say so and the experimenter will come in and try to help.

If more than 45 minutes elapses before the task is completed, we will stop the experiment.