Building technical laboratory set-ups for user-centered evaluation of real-time person to person communication

Svein Heiestad
Telenor Research – Norway
svein.heiestad@telenor.com

Bjørn Hestnes
Telenor Research – Norway
bjorn-olav.hestnes@telenor.com

Peter Brooks
Telenor Research
Norway
PBrooksF@aol.com

Trond Ulseth
Telenor Research
Norway
trond.ulseth@telenor.com

Carl Aaby
Tandberg ASA
Norway
cha@tandberg.no

Abstract

This paper offers advice on the technical set-up of laboratory facilities for data collection of users of real-time person-person communication technologies. Two rooms for test persons are connected with a network simulator that can be set to certain network modes. Equipment used is commercially available, but with patches to keep variable parameters stable (e. g. frame-rate). The end-to-end parameters are difficult to exactly determine by a human being and therefore a test method developed for measuring end-to-end technical characteristics is described.

Key words: Usability labs, audio conferencing, videoconferencing, multimedia conferencing, face-to-face, quality of service (QoS), Human Factors, fitness-for-purpose

1. Introduction

The development of services such as videoconferencing and multimedia conferencing for real-time person-to-person communication is a growing and complex area, with many technical variables associated with terminal, network and media protocol characteristics. The use of IP networks raises new problems for the quality of experience for users. Performing tests of user-user interaction requires special laboratories and technical set-ups that pose certain additional requirements than for more 'traditional' usability testing of user-system interaction. A recent EC project on the fitness-for-purpose of real-time human communication services (Brooks et al., 2003) has resulted in an evaluation methodology for real-time communication studies (Heim et al., 2003). Whereas Heim et al. focus more on the psychological design of real-time communication studies, the current paper offers recommendations for the more technical aspects of the set-up that may be less familiar to human factors experts but which should be taken into account in addition to psychological aspects.

2. Design considerations

The following design recommendations are explained in relation to the necessary set-up elements of measurement methods, laboratory space, equipment and networks:

- Know your equipment and the person to person characteristics
- Three instead of two rooms - designed for real-time communication
- The crucial characteristic of the test rooms is the sound insulation between them
- The audio and video sent to one test room is a result of a set of environmental parameters in the other test room:
    - reverberation time - room acoustics
    - ambient lightening
    - background
    - viewing position
- Communication with test rooms is different - the laboratory needs an intercom system integrated with the technical set-up
- Design for highest possible quality, then introduce degradation to reduce quality.

3. Measurement methods

"Know your equipment and the person to person characteristics"

With modern digital communication technology it is no longer possible to use simple engineering metrics to determine the technical parameters being studied. 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).

To measure end-to-end parameters between test users it is necessary to develop equipment to measure: audio delay, video delay and asynchrony. As there is no way to precisely determine the audio and video synchronization, both streams have to be measured independently. Three types of measurements should be taken: person-to-person, eye-to-eye and mouth-to-ear.

3.1 Person to person

The importance of knowing the technical parameters of test equipment cannot be stressed enough. In conducting experiments, tests are often based on comparison of different communication settings and communication services. The aim is to only vary one parameter at a time, and by measuring the technical parameters of the equipment one can make sure that this is achieved, and the necessary tuning of equipment is done. For example, if one wants to compare the user-based characteristics of audio and video telephony, it is crucial that the delay is the same. Videoconferencing equipment available today typically introduces a far greater delay than audio telephony equipment. By tuning your equipment so delay is equal, you make sure that results intended to be based on the communication media are not invalidated by delay.

3.2 Eye to eye

To measure the delay of the video streams a video generator and an oscilloscope can be used. Two codecs are set up in a call. The first has a video generator connected to the video-in port, so the image from the video generator is sent to the other codec. If the oscilloscope is connected directly to the video generator and the video-out port of the second codec, the change in image is recorded on the oscilloscope. Therefore it is possible to read the delay as the time difference between the instance when the image is received by the first codec until it is decoded by the second codec.

Delay in the TV monitors is believed to be in the range of a few milliseconds and so for most tests is not large enough to require being determined exactly. As the camera is an analogue device the delay is very small and should have no significant impact on the total delay of the system.

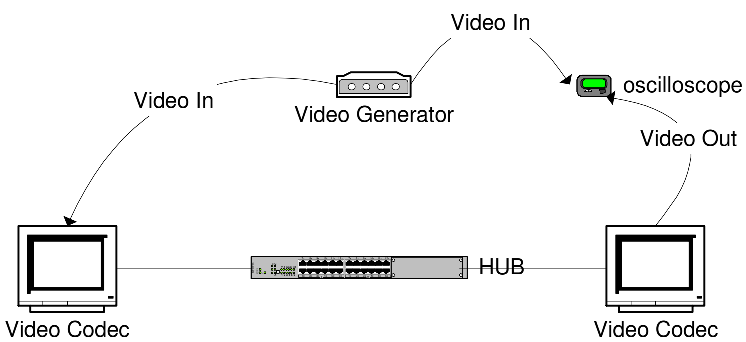

Figure 1. Diagram of measurement network

Measurement equipment tests have showed delay to vary by about 50ms, depending on the change of the image. A small change in the image resulted in a delay about 50 ms less than the delay of a complete image change. This is expected, as a new image would force the codec to encode an I-frame[1], and with the smaller change in the image the codec encoded a P-frame[2]. In real life the video delay would vary, and this is not believed to be apparent to the user.

3.3 Mouth to ear

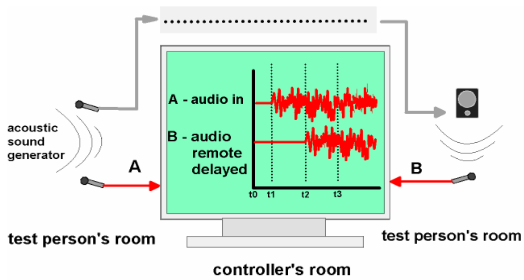

Figure 2. Audio delay measurement

Audio delay between the mouth of one of the test persons and the ear of the other can be measured with two microphones and an oscilloscope. An acoustical pulse is generated from a speaker’s position 1m from the microphone on the encoder side. The microphone is also connected to the oscilloscope. At the decoder side the microphone was also connected to the oscilloscope and placed at a listener's position (1 m from the loudspeaker). When measuring systems with low delay (5-10ms)

[1] A frame coded as a separate image, not depending on previous frames.
[2] A frame coded as changes from the last I-frame.

we have found that most of the delay is a result of the distance between the user's mouth to the microphone and from the loudspeaker to the ear. However, some voice coding algorithms may introduce a significant delay.

4. Laboratory space

"You need three instead of two rooms"

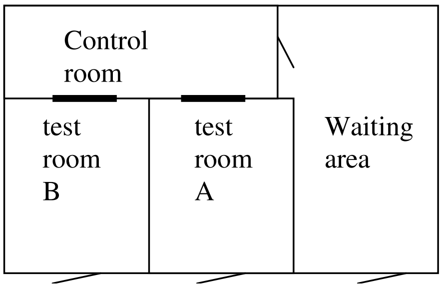

Figure 3. Example of a lab space

A 'traditional' usability laboratory consists of a test room and an observation room. For person-person communication the situation is different. There are two users, which require two separate test rooms, and it should 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.

4.1 Test rooms

"The crucial characteristic of the test rooms is sound insulation"

Figur 4. Test and control room at the Telenor lab

The rooms should be sound proof to the degree that natural sound will not be transmitted from one test room to the other. This is essential to support the experience of persons communicating from remote distance. The one-way mirror, ventilation system, the door and the bushing of cables very often turn out to be the weak points. Sound insulation between test rooms should be at least 35 dB within the frequency range from 100 Hz to 8000 Hz.

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 tests, ambient *lighting* is also important. 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 with good contrast to the user.

Ventilation is also a critical factor - especially if the test rooms are small. 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. If possible, the noisy and heat-generating equipment should be placed outside the test rooms, e.g. in the control room.

4.2 Control room

"Communication with test rooms is different"

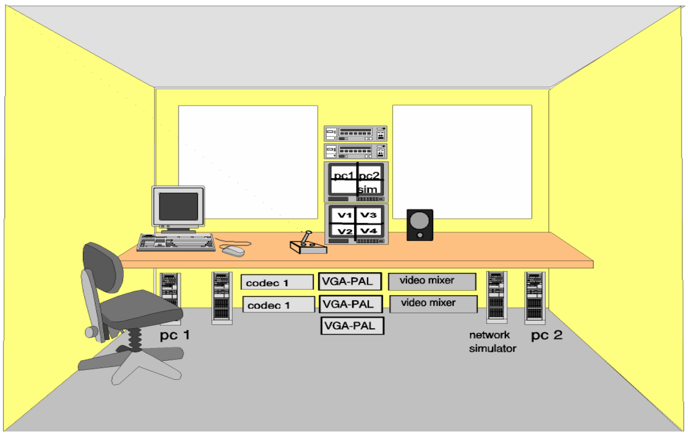

Figure 5. Control room lay-out

You should 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 recommended. 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. Separate lighting control for each room is also important. 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.

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 to design a ventilation system with enough capacity. Remember to calculate for all the planned equipment in this room.

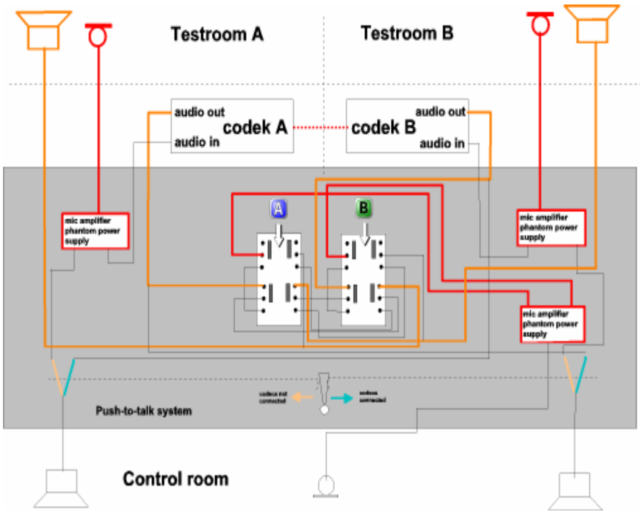

Figure 6. Example of an integrated intercom system

5. Equipment

The equipment 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 test conditions, parameters of the "technical environment" can both be variables to control and the actual independent variables of interest for the test.

5.1 Terminal equipment

The equipment chosen should support the highest quality expected in your series of experiments; e.g. the camera chosen should support the highest space resolution of interest to the test, in order to compare its quality with lower resolution conditions.

Viewing distance and camera position is important. General recommendations for video-based communication are:

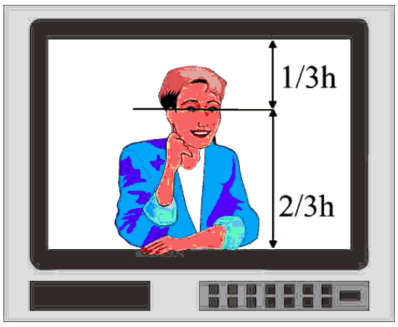

Figure 7. A template to secure parallax-free eye contact

- The camera should 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) avoid parallax difference in eye-contact and keep the test-person's eyes not more than 1/3 H from the top of the screen (figure 7).

The choices of settings are of course dependent 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
- The user should not have any control over the equipment during the tests.

5.2 Codec 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 synchronisation, audio immediate, video space resolution and video time resolution.

When using a loudspeaking 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 an echo annoyance.

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. However, if one chooses to use an external echo canceller on a codec with a built-in echo canceller, the internal echo canceller has to be turned *off*. Echo cancellers may introduce delay from a few ms up to 100 ms. Echo cancellers adapt automatically to the room characteristics and need a few seconds to optimize performance. It is important to know exactly how the echo canceller works to achieve full control of the audio system parameters. The performance of the echo canceller (i.e. how well the echo is cancelled) needs to be verified. The choice of codec (with or without echo canceller) again depends on the purpose of your test, and the required transmission characteristics of the equipment.

5.3 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. Some have already been introduced above (e.g., viewing distance) and the list below shows examples of other parameters and recommended values.

Parameter	Value
Lighting	100 % artificial
Conditions	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)

Table 1: Constant Environmental parameters

6. Network

6.1 Network 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 an IP network the main network parameters influencing the perceived quality are: bandwidth, packet size, delay, delay jitter, packet loss, burst packet loss and sequencing.

6.2 Terminal characteristics

When manipulating the network characteristics, one also needs to know how the terminal responds to these manipulations. For example, if the jitter buffer is dynamic, the delay is increased if you introduce packet loss, and the terminal may have an in-built packet loss mechanism (e.g. the IPLR[3] algorithm) that is used to reduce awareness of packet loss.

6.3 Network simulator

"From best quality to lower quality"

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

[3] Reference: Patent pending US provisional 60/423,393

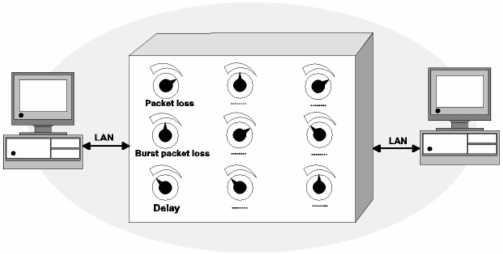

Figure 8 Network simulator

normally be sufficient to manipulate these parameters separately. But a *network simulator* is recommended if the test condition is to be a more complex interaction between several network parameters (delay, packet loss, image frequency, transfer protocols, etc). Network simulators are available or can be built for any type of network to secure control and reproducibility of the network parameter variations. The network simulator will never improve the network performance or quality, only reductions will be possible.

7. Conclusion

For usability experts used to working with traditional usability laboratories there are some very important differences to bear in mind when planning and designing test laboratories and technical set-ups for real-time person to person communication. In addition to the traditional environmental attributes, knowledge of the characteristics of the technical equipment can ensure that user-test results are a consequence of the designed variables rather than otherwise unknown characteristics of your technical set-up.

Acknowledgements

The basis of this work was performed as part of the EC project Eye-2-Eye – Fitness-for-purpose of Person-Person Communication Technologies (IST-1999-11577). The authors have benefited substantially from the whole project group: Asbjørn Følstad, Jan Heim, Trond Schliemann, Jan Håvard Skjetne (SINTEF); Han Frowein (Instituut voor Doven); Carl Aaby (TANDBERG); Patrick Brundell, Claire O'Malley, (University of Nottingham); external peer reviewers Pedro Concejero (Telefónica Investigación y Desarrollo) and Manfred Laube (Alcatel SEL)

Gelöscht: ¶

References

Brooks, P., Schliemann, T., Hestnes, B., Frowein, H., Aaby, C., O'Malley, C. (2003). Final Report Project IST-1999-11577 Eye-2-Eye: Fitness-for-Purpose of Person-Person Communication Technologies. (Available from www.eye-to-eye.org)

ETSI ETR 297: Human Factors (HF); Human Factors in Videotelephony.

Heim, J., Brooks, P., Følstad, A., Schliemann, T., Hestnes, B., Heiestad, S., Ulseth, T., Frowein, H., Devoldere, P., Aaby, C., O'Malley, C., Brundell, P., (2003). *Fitness-for-Purpose Evaluation Methodology*. . (Available from www.eye-to-eye.org)

Patent pending US provisional 60/423,393

ITU-T Recommendation P.79 (1993): Calculation of loudness ratings for telephone sets.