Thesis Master Information Science
Human Centered Multimedia

University of Amsterdam Faculty of Science
Sheldon Pijpers - 10668578

A requirement analysis for a multi-party conferencing testbed

---

Supervisor: Marwin Schmitt

Marwin Schmitt, signature:

Dr. Frank Nack, signature:

July 10, 2014

A requirement analysis for a multi-party conferencing testbed

Sheldon Pijpers
CWI: Centrum Wiskunde & Informatica
The Netherlands
sheldon.pijpers@cwi.nl

ABSTRACT
Current videoconferencing services such as Skype and Google+Hangouts provide mechanisms for engaging in multi-party conversations. Although these services provide basic support, they lack functionalities that take into account the users roles and context. Currently, the multimedia research community is actively engaged in conducting experiments concerning Quality of Experience (QoE). This paper provides a requirement analysis for a multi-party conferencing testbed, that is designed for conducting controlled telecommunication experiments for assessing QoE. A pre-study, in the form of an online survey, investigated the experience with previous tools and identified the interest towards using the CWI tool for future studies. Requirements are derived through semi-structured interviews by looking into the experimental process and issues that stakeholders are currently facing. Results show that having the capability to pre-define the experimental conditions and manually adjust these throughout the experiment are integral aspects within the tool. Furthermore, various control possibilities to interact with the test participants are needed. Subjective assessment integration in the form of questionnaires and logging of technical conditions are important requirements to support the analysis phase. Documentation, coding support and easy customizability are crucial aspects influencing the overall tool usability. The listed requirements provide a framework for further development of QoE assessment tools in the area of telecommunication studies and, furthermore, contribute to the open-source development of the multi-party conferencing testbed.

1. INTRODUCTION
Over the last decade, we have witnessed a tremendous growth of video-mediated group communication. With the proliferation of multimedia technology and network accessibility, new dynamic solutions for multi-party conferencing have emerged. By adapting to these changes both businesses and individuals can interact with each other in a timely and cost-effective manner [6]. Current services such as Skype[1], Google+Hangouts[2] and Facetime[3] provide mechanisms for engaging in multi-party communication settings. Although these services are immensely popular in terms of their usage, they lack adaptability to conversational aspects. Therefore, in order to create a richer user experience, mechanisms are needed that take into account the varying network and communication parameters that occur during these multi-party gatherings. Currently, The International Telecommunication Union (ITU), a standardization organization from the United Nations (UN), is studying Quality Of Experience (QoE) assessment in the direction of multi-party-conferencing. ITU-T P.1301 is a standard that is currently in development and aims to provide subjective quality assessment in multiparty communication settings [7]. Recommendations however, are still lacking, and the knowledge required to support videoconferencing services that adapt to the influential factors of QoE remain scarce. Such knowledge can be obtained through subjective tests by assessing the end-users perception of varying video quality conditions. By gathering subjective feedback from each test participant individually, more insight can be gained on how to optimize the current user experience. However, in order to conduct these type of experiments, testbeds are needed that provide the level of control for monitoring and gathering both objective and subjective data in a video-mediated environment.

For a long period of time, testbeds are being used as an experimental platform to evaluate and optimize services and new technologies [1]. With the implementation of testbeds, researchers can replicate the usage or behavior of technological elements (e.g. network topology, video quality) in a safe and controlled manner. Within the domain of video-mediated communication (VMC), assessing QoE has gained large interest. Technical conditions, such as network variances, delays and resolution can be measured with Quality of Service (QoS). Measuring the effects on the user (QoE) remains an ongoing issue as influential factors such as context and user roles need to be taken into account [10]. By developing a testbed that supports the conduction of controlled extensive user trials, the goal is to gain more insight on QoE assessment

[1]http://www.skype.com
[2]http://www.google.com/+/learnmore/hangouts/
[3]http://www.apple.com/ios/facetime/

and shift towards videoconferencing systems that act upon the influencing factors of QoE.

The CWI multi-party conferencing testbed developed by Schmitt et al.[10] provides such a framework, in which researchers have control over various communication and media parameters. The main drawback of this testbed, is that it is built for in-house use and is therefore tailored to the process and experimental requirements of the CWI staff. In order to extend the knowledge in this domain, the goal is to make the tool publicly available so that other researchers can integrate this tool in their experimental studies. This papers focuses on a requirement analysis for the multi-party conferencing testbed developed by Schmitt et al. [10]. The research question that forms the basis of this study is stated as follows: "What are the requirements for researchers conducting experiments with the CWI tool?". In order to provide the desired tool support, it is important to understand the process, scope of tasks and issues that researchers are currently facing in telecommunication studies. By focusing on support throughout the general phases of an experiment, namely the design, conduction and analysis phase, a list of requirements will be documented to support the open-source development of the CWI tool. Furthermore, this research is relevant as it can enhance the development of QoE assessment videoconferencing tools.

The remainder of this paper is structured as follows. In section 2 a literature study is conducted. Afterwards, an overview of the CWI videoconferencing tool is given. Section 4 presents the methodology of this paper. On the basis of an online survey, potential stakeholders are interviewed. The results are discussed and documented in the form of requirements. The penultimate chapter discusses the contribution and limitations of this paper. Finally, a conclusion is given in which proposed future work and an overall summary of the findings are described.

2. RELATED LITERATURE

The literature study shows that various VMC experiments in the direction of QoE assessment have been conducted. Although these studies provide useful insights, they mainly fail to take into account the user roles and context.

Various testbeds have been reviewed that provide similar approaches in evaluating QoE assessment. However, the focus of these testbeds is often mainly based on one particular aspect (e.g. network effects) or is applied in a different context (e.g. mobile context). Studies providing VMC testbed analyses from the user experience perspective have however not been discovered.

Due to the various set-ups/tools used throughout the experiments, comparisons between results are likely to be less reliable. A common testbed that can be used by a community of stakeholders, might provide an easier solution for conducting experiments and comparing future results in a more reliable manner.

2.1 Video-mediated communication

Along with the growth of text-based platforms a new paradigm of video communication has become easily accessible on a wide range of interactive systems. Hardware advancements in the form of built-in cameras providing high quality video and audio have become a commodity in notebooks, smartphones and tablets. These technological improvements have shown to provide a more flexible conversion into the field of video-mediated communication. Current services such as Skype and Facetime allow cross-platform accessibility, making it easier for users to engage in video face-to-face communication. [6]. By integrating mechanisms that provide multi-party conversations, video-mediated group communication is becoming increasingly popular.

By using these video-mediated platforms, a social shift is made towards a more interactive and cost-effective way of communicating. As the overall benefits of video-mediated group communication are gradually becoming more and more important, knowledge is required to shift towards VMC systems that take the influential factors of QoE into account [10].

2.2 Quality of Experience (QoE) research

For a long period of time, Quality of Service (QoS) has been the common framework to assess the performance of systems. With the emergence of multimodal systems and sensory modalities, the field of user-centered interaction has grown extensively. In order to design and build systems that provide the desired user experience, human-centric evaluation methodologies are needed [12]. Only relying on QoS as a measurement of success has proven to be unreliable, as QoS mainly deals with the evaluation of a service from a system perspective. By combing both human and system perspectives into one theoretical framework, Wu et al.[12] aim to assess the correlation between QoS and QoE. A similar framework has been proposed by Geerts et al.[4], who present a QoE model by mapping the technical and user aspects into one QoE framework. By integrating both domains into a multidisciplinary approach Geerts et al. aim to provide a useful framework for measuring Quality of Experience in future studies. The videoconferencing tool developed by Schmitt et al. [10] elaborates on the QoE frameworks proposed by Wu et al.[12] and Geerts et al.[4] by presenting a similar model applicable within a VMC setting for controlled experiments.

Within the domain of VMC, various QoE experiments have been conducted. In previous research carried out by Tam et al. [11] a dyadic experiment investigated the effects of network and computational delays. Results provide insight into feasible delay conditions for two-way interaction within VMC settings. Other research conducted by Geerts et al. [5] report various synchronization requirements for collaborative video watching. Subjective tests in the direction of multi-party conferencing have been conducted by Berndtsson et al. [2], to gain more insight on how to optimize the current user experience and shift towards standard quality evaluation test methodologies. In order to enhance the current field of videoconferencing, subjective quality assessment methods are needed. These test methodologies need to be formally agreed on, so that other researchers can evaluate telecommunication experiments in a similar manner.

ITU-T has recently started to look into QoE assessment in the context of multi-party communication [7]. Currently, ITU-T has a recommendation series for quality assessment in both audio (P.8xx Series) and audiovisual (P.9xx Series) contexts. A recommendation series concerning multi-party communication (P13xx series) is in development. Table 1 presents an overview of specific ITU-T standards for interactive test methods.

Table 1: ITU Recommendations

P.805:	Subjective evaluation of conversational quality
P.920:	Interactive test methods for audiovisual communications
P.1301:	Subjective quality evaluation of audio and audiovisual multiparty telemeetings

2.3 Testbeds

A testbed framework for evaluating QoE in a multi-dimensional approach is proposed by de Moor et al. [3]. By conducting mobile field trials in which various QoS conditions can be monitored on different dimensions and levels (e.g. network, context), the goal is to provide an integrated QoE framework applicable within mobile contexts. Within the testbed three entities are integrated to measure technical, contextual and user assessment aspects.

The QoE-Lab, presented by Mehmood et al. [8] focuses on a testbed that investigates the effects of varying wireless network conditions throughout mobile computing contexts. The testbed makes use of a wireless network emulator to simulate realistic complex network settings (e.g. by manipulating packets/network routing) that occur in everyday life. Various scenarios (under varying networking conditions) are evaluated to gain insight in how this effects the user experience. Evaluations with both objective and subjective quality test methods are conducted to gain insight in the correlation between the various network conditions and the user-perceived QoE. Overall, the goal is to gain more insight on how to seamlessly optimize the mobility across different wireless technologies by manipulating conditions at the network level.

The CWI multi-party conferencing testbed developed by Schmitt et al. [10] focuses on the end-user's perception of audiovisual conditions, independently from how it might arise technically. Through a media processing pipeline, various technical aspects can be implemented and monitored to gain insight on how this effects the users perception of the audiovisual quality. With this testbed the influence of both network and media conditions can be assessed in a video-mediated environment. Thus, the testbed assess QoE by manipulating conditions at the application level.

3. SYSTEM

This section provides an overview of the three main components of the multi-party conferencing testbed applied throughout the design, conduction and analysis phase of the experiment. The goal of this tool is to gain more insight in QoE assessment by supporting a framework in which varying conditions can be monitored in a controllable environment.

3.1 Video client

The video client for multiparty conferencing depicted in Figure 1 provides an example of how the client is configured during the conduction of an experiment. As shown, the client presents an overview of the test participants that are currently active in the experiment. The current task that the test participants have to discuss is integrated in the lower left corner of the interface.

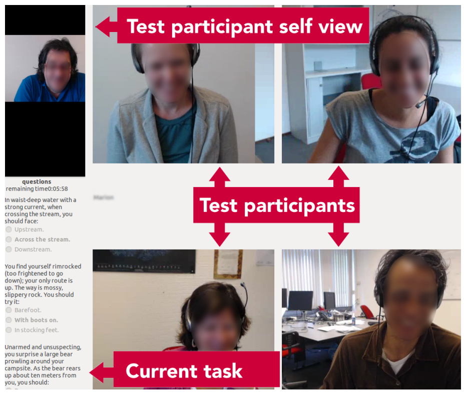

Figure 1: Video client

3.2 Observer control client

The observer control client shown in Figure 2 provides a GUI for the experiment conductor. Within this client the experiment conductor can dynamically join the conversation, see the status of the participants and set and execute the experiment procedure. Furthermore the experimental design (e.g. task set-up, manipulation of parameters) can be implemented and manually adjusted within this client.

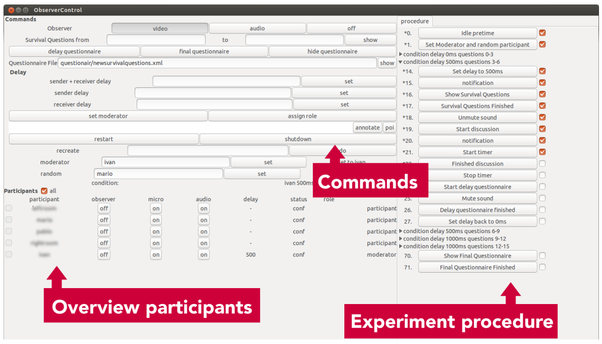

Figure 2: Observer control client

3.3 Tool for post-processing experimental data

An example of the tool for post-processing the results is presented in Figure 3. With this tool sessions of conducted experiments can be viewed and analyzed. Various types of data scripts (e.g. speech pattern data, questionnaire data) can be processed and exported and used for further analysis. The tool provides an overview of the labeled speech pattern of each participant. Furthermore, the color denotes the type of identified speech activity. The experiment analyst also has the possibility to manually tag and categorize speech data.

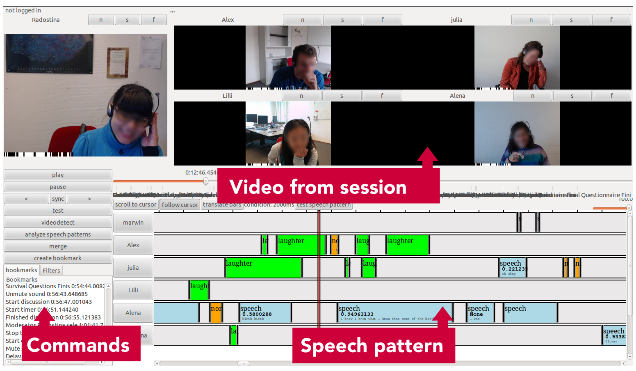

Figure 3: Speech pattern analysis tool

4. METHODOLOGY

A pre-study in the form of an online survey was designed and conducted to gain insight in how researchers have experienced previous used tools during telecommunication experiments. Furthermore, the interest towards using the CWI tool for future studies is investigated. Based on the survey results, participants were selected and online semi-structured interviews were carried out. The gathered qualitative data is analyzed and documented in the form of requirements.

4.1 Online survey

The sample population of the online survey contained 10 (N=10) European researchers from various companies and research institutions (e.g. BT[4], Deutsche Telekom[5]), ranging from (PhD) researchers to (assistant) professors that all have familiarity with conducting telecommunication experiments in previous studies. The goal of the survey was to evaluate how the population has experienced previous used tools and to assess if there is an interest in using the CWI tool for future studies. The survey contained a set of 6 closed questions, 4 open questions and 2 multiple answer questions with an optional 'input' function. As the population has never worked with the CWI tool before, a description of the tool components was integrated in the survey. Within the survey a set of questions was used to investigate the experience with tools used in previous telecommunication studies. The closed questions containing a varying five point Likert-scale were analyzed by computing the mode, median and frequency using SPSS. As each Likert question was treated as unique and stand-alone, the questions were analyzed as Likert-type items on an ordinal measurement scale. A one-sample t-test was performed to determine significance levels at $P \leq 0.05$. The open questions were used as a qualitative measure to identify aspects that influence the participants in using the CWI tool for future studies. Furthermore, similarities and associations were analyzed to gain more insight on the importance of particular aspects. Lastly, input for suggestions and recommendations were integrated to define certain issues that have been either overlooked or not yet identified within the survey.

4.2 Semi-structured interviews

Based on the results of the online survey, semi-structured interviews were designed and conducted. Data was gathered from 7 (N=7) one-to-one interviews that were all recorded and transcribed. Within this sample, data is derived from 5 interviewees that have conducted the online survey. The remaining 2 interviewees have been recommended outside the scope of the survey results. The semi-structured interviews were used as a qualitative measure to gain insight in support issues throughout the design, conduction and analysis phase in previous telecommunication experiments. In order to identify significant patterns and prioritize requirements, the following actions were performed on the qualitative data:

1. Identify support issues
2. Identify repetitiveness of support issues
3. Prioritize and categorize issues

The interview format contained 10 questions with additional follow-ups used to gain more insight in particular topics. Requirements were documented according to each (transitional) phase of the experiment. Furthermore, general requirements that have been identified were documented and ranked based on importance.

[4]http://www.home.bt.com
[5]http://www.laboratories.telekom.com

Table 2: Experience previous used tools

Tool research topics	Very Poor (1)	Poor (2)	Fair (3)	Good (4)	Very Good (5)	Mode	Median	P
Q5. Meeting the experimental requirements	0 (0%)	1 (10%)	4 (40%)	4 (40%)	1 (10%)	3/4	3.5	<0.001
Q6. Usability experience	0 (0%)	3 (30%)	4 (40%)	2 (20%)	1 (10%)	3	3	<0.001
Q7. Support experimental phases	1 (10%)	2 (20%)	5 (50%)	1 (10%)	1 (10%)	3	3	<0.001

5. RESULTS

A pre-study in the form of an online survey was designed and conducted to gain insight in how researchers have experienced previous used tools. Furthermore, the online survey identified if there is an interest in using the CWI tool for future studies. The content of the questionnaire that led to the responses can be found in Appendix A. The design of the questionnaire was setup according to the guidelines and principles stated by Robson [9].

Figure 4: Experience in conducting telecommunication studies

5.1 Online Survey Results

In Figure 4 the sample population's overall experience in conducting telecommunication studies is depicted. As can be seen 60% (N=6) of the population states that their experience in conducting these type of studies is good/very good, whereas 30% (N=3) states to have a fair experience with the conduction of telecommunication experiments.

5.1.1 Experience tools

Various close-ended questions were used to determine how the sample population has experienced previous used tools during telecommunication experiments. Table 2 outlines the significant results used to analyze the current situation. As shown, the majority of the population states that the current tools provide either a fair/good rating in terms of meeting the experimental requirements. Furthermore, 30% (N=3) of the sample states that the usability of the tools was experienced poor, whereas 40% (N=4) states that they have a fair usability experience with the tool(s). Lastly, 30% (N=3) states that the tool support throughout the general phases of the experiment is considered poor/very poor. Only 20% (N=2) experiences this support as good/very good. These results conclude that both support and usability are important issues that should be taken into account. A cross-tabulation was conducted to investigate possible correlations between the tool experience topics. However, no significant patterns were identified.

5.1.2 Data collection

A multiple answer question was provided which was used to gain insight in the types of data that are usually collected within telecommunication studies. As shown in Table 3, the majority of the respondents collect questionnaire data. Furthermore, it has been identified that 30% (N=3) of the participants have stated to collect video, speech and questionnaire data during previous studies. Other types of data that were identified consisted of QoS metrics, delay data and task execution times.

Table 3: Data collection telecommunication experiments

Data collection type	Frequency
Questionnaires	8
Video data	5
Speech data	4
Other	2

5.1.3 Interest on the tool from CWI

As depicted in Figure 5, the interest on using the CWI videoconferencing tool is high. The values "likely/extremely likely" are consistently greater compared to those who stated neutral, unlikely, very unlikely. This percentage of 70% (N=7) towards likely/extremely likely denotes there is a high interest in using the CWI videoconferencing tool for future telecommunication studies.

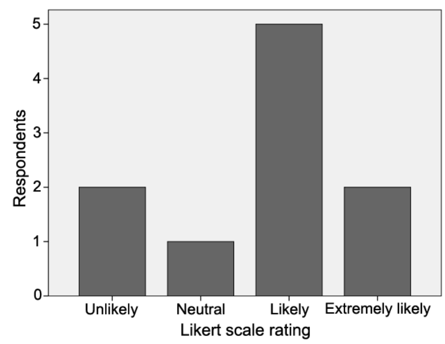

Figure 5: Interest CWI tool

5.1.4 Influential aspects CWI tool usage

Among the results of the open ended questions, important aspects that could influence future tool usage were analyzed. Within the population, a large portion of the respondents stated that usability and high degree of control are important factors. Furthermore, data type collection/integration, stable performance and easy customizability are factors that are identified to have great influence on the eventual usage of the tool.

5.2 Semi-structured interviews

Semi-structured interviews (Appendix B) were used as a qualitative measure to gain insight in the process throughout the experimental phases and functionalities required to further the development of the CWI tool. Questions were structured according to the three general phases of an experiment (design, conduction, analysis). Within the format of the interview, the following aspects were taken into account:
- Experimental process during design, conduction and analysis of the experiment
- Tool(s) support within these phases
- Issues within these phases

The rationale for applying semi-structured interviews is that they provide a flexible format in which various topics and follow-ups can be discussed for collecting reliable qualitative data [9].

5.3 Semi-structured interview results

This section provides the results derived from the semi-structured interviews. The results are documented in the form of requirements and are structured according to its priority (requirement number 1 denotes the highest priority). From the interviews, requirements were derived on a level that influences the overall tool usage. These requirements provide insight into general matters concerning tool support.

5.3.1 General requirements

In Table 4 the list of general requirements is documented. Within this list, documentation and coding support have been identified 4 times among the interviewees. Problems faced were mainly based on the implementation of external tool elements. As documentation and coding support in this area were lacking, a lot of time and effort was needed to understand the various tool components. Similar to the previous stated requirement, time alignment of media channels has been mentioned 4 times. Issues identified were mainly related to incorrectly time-aligned sessions, as this influences the overall measurements throughout the analysis phase. The other 4 requirements enlisted in the general section are all derived from single interviewees but are considered as useful aspects to take into account.

5.3.2 Design phase

As shown in Table 5, predefining the technical conditions per trial/subject is the most crucial aspect within the design phase of the experiment (identified 5 times among the interviewees). Currently, various stakeholders make use of external scripts and complex configurations to pre-define the experimental conditions. Preliminary test integration and defining the test subjects are both aspects that have been identified 4 times. Having the possibility to conduct a preliminary test has shown various benefits (e.g. testing the experimental conditions, acquaint test participants with tool equipment). Lastly, specifying the communication levels and integrating source material have both been identified 3 times during interview analysis. As stakeholders can have various experimental goals, the tool should provide the possibility to define the communication level(s) beforehand (e.g. audio only experiment).

Table 4: General requirements

G-1. Documentation and coding support A user manual should be created to enhance the understanding of the various tool components. Furthermore, the source code should be well supported with understandable and clear comments.
G-2. Time alignment of media channels During experiment conduction, the media (audio/video) channels should be time aligned (e.g. not influenced by adapting conditions). Incorrect time alignment will have influence on the recording and analysis of the data streams (e.g. objective measurements with turn durations).
G-3. Cross-platform integration Provide support for cross-platform (e.g. tablet, smartphone) experiments.
G-4. Integration of real-time chat Integrate a real-time chat channel that is synchronized with other applicable media streams.
G-5. Codec parameters support The tool should provide video and audio codec support for different scenarios and experiments.
G-6. Support tool customization The tool should be split up into various modularities to support the integration/customizability of (external) elements. Dividing the tool into modularities reduces the complexity of elements that are fixed, integrated or dependent on other elements within the framework of the tool.

5.3.3 Conduction phase

Table 6 outlines the identified requirements for tool support during experiment conduction. From the sample population (N=7) 5 interviewees have stated that the real-time adjustment of technical conditions is an important aspect during experiment conduction. The interview analysis has shown that a variety of issues have been faced during the adjustment of technical condi-

Table 5: Requirements tailored to the design phase

D-1. Pre-defining the experimental conditions per trial/subject The experiment conductor(s) can easily predefine independent test variables (e.g. delay) based on the setup of the tasks and test subjects.
D-2. Preliminary test integration A preliminary test is used to familiarize the test subjects with the test equipment and experimental procedure. Furthermore the experiment conductor can test technical conditions as a preliminary step before experiment conduction.
D-3. Integration of test subjects into experimental framework The tool should provide a framework for easily integrating the desired amount of test subjects participating throughout the overall experiment session. Furthermore the test conductor(s) should have the ability to label each test subject.
D-4. Specifying the communication levels The experiment conductor specifies the desired type of communication levels applicable for experiment conduction (e.g.,audio, video, audio/video).
D-5. Integration of source material The implementation of external source material is useful for experiments that make use of images, video fragments and other types of media data. Furthermore it should be possible to integrate media streams separately (e.g. the audio and video stream of a fragment).

Table 6: Requirements tailored to the conduction phase

C-1. Real-time manual adjustment of technical conditions Changing technical conditions (e.g. changing levels of delay, packet-loss) should be flexible and easily configurable during experiment conduction.
C-2. Interaction flexibility throughout experiment conduction A control panel should provide an overview of the test participants. It should be possible for the test conductor to manipulate the interaction possibilities among the test participants and to have interaction control with either one or all test participants when needed.
C-3. Dynamically adaptable streaming During experiment conduction the tool should be dynamically adaptable to the implemented technical conditions and network variances. It should provide the flexibility to switch through adaptive test conditions throughout the experiment.
C-4. Real-time annotation during experiment conduction Ability to add audio/video annotations during experiment conduction.
C-5. Crash issues support If the tool malfunctions during experiment conduction it should be able to restart the experiment from a certain task interval. This requires a logging of the experiment process.

tions (e.g. noises appearing, restarting conversation set-up, complex technical configuration). Similar to the first requirement, 5 interviewees have discussed issues regarding the interaction flexibility with either the test conductor or between the test participants. Having the ability to easily communicate with a selection of test participants has shown to be cumbersome, as tools were often developed to communicate with either none or all test participants. Subsequently, 3 interviewees have stated that the tool should be dynamically adaptable to the test conditions throughout the experiment. As additional requirements (based on the input of 2 interviewees), crash issues and real-time annotation are considered important aspects.

5.3.4 Analysis phase

The requirements enlisted in Table 7 are based on findings identified within the analysis phase. All the interviewees (N=7) have stated that subjective assessment in the form of questionnaires is an essential component within telecommunication experiments. Various issues regarding the flexibility of the questionnaire design have been identified (e.g. integrating various input types). Furthermore, a lot of interviewees stated that the manual handling of questionnaire data took up a lot of time. With the possibility to integrate questionnaires, results can be exported and viewed in an easier manner. The logging of technical conditions and recording of speech/video data have both been identified 4 times among the interviewees. Issues faced were mainly based on the lack of monitoring of the technical conditions throughout the experiment. An additional requirement (derived from 1 interviewee) is based on the analysis of real-time annotations inserted during experiment conduction. With this functionality the experiment analyst should be able to easily track and view inserted annotations.

6. DISCUSSION

The results derived from the semi-structured interviews provide insight in requirements tailored to enhance the development of VMC testbeds in the area of subjective testing. The literature study has shown that current QoE testbeds mainly focus on evaluating the effects of one particular condition (e.g. network variances) or are applied in a different context (e.g. a mobile computing context). VMC testbeds that provide the desired level of control for assessing QoE have not been identified. An important requirement that

Table 7: Requirements tailored to the analysis phase

A-1. Subjective assessment integration The integration of questionnaires should be supported within the experiment (during conduction or afterwards). Furthermore the questionnaire design should be flexible (allowing various types of data input) and should be exportable in a presentable format (e.g. Excel).
A-2. Logging of technical conditions The tool should provide logging capabilities of the manipulated technical conditions. Furthermore it should be able to identify and monitor persisting degradations during the playout (e.g. network variances). This is particularly useful for mapping different data types (e.g. comparing speech/video data with MOS scores).
A-3. Recording of speech/video data Provide recorded sessions labeled to each participant. Furthermore it should be possible to gain insight in particular conversational aspects (e.g. type of speech data, turns taken, turn duration) and have the ability to manually label sound insections.
A-4. Analysis of real-time annotations inserted during experiment conduction After experiment conduction the test conductor(s) can easily identify the inserted annotations for further analysis.

is essential within the design phase of the experiment, is the ability to pre-define the experimental conditions per trial/subject. After setting the study goal, the experiment conductor should have a framework to easily integrate the experimental design into the set-up of the testbed. In order to provide optimal flexibility, the testbed should also provide the ability to manually adjust conditions throughout the experiment. Furthermore, the test conductor should have the possibility to monitor the experiment session and interact with the test participants when needed. If for instance, the experiment conductor aims to provide support to a single (or selection of) test participant(s), it should be feasible to only interact with the chosen selection. This is particularly useful for minimizing the overall distraction among the other test participants during experiment conduction. As all the interviewees have stated that subjective assessment is an important requirement, the testbed should provide functionalities for designing and integrating questionnaires into the experiment procedure. The results should be easily exportable in a presentable format (e.g. Excel), so that the experiment analyst can easily review and compare the results with objective data (e.g. logging of technical conditions) afterwards. A common issue that has influenced the overall usage of tools in previous studies is based on documentation and coding support. Although a large scale of adjustment is already possible within the current testbed, it should be easily feasible to customize and integrate certain aspects within the tool. As researchers conducting experiments in this field mainly focus on investigating unidentified aspects and new technologies, the testbed should provide support so that it can be easily modified and extended when needed. Therefore, a structured user manual and source code support are important factors that will most likely influence the adaptation of the CWI tool in future studies.

A comparison is conducted to gain insight in the requirements that need to be considered when developing the next version of the CWI tool. From the list of requirements documented in the general phase, requirement G-1 (Documentation and coding support) needs to be looked into. As the tool was initially built for in-house use, documentation for external use and source code support were not taken into account. Requirement G-2 (Time alignment of media channels) is mainly dependent of the quality of GStream-er[6]. Requirement G-3 (Cross-platform integration) is in its current state mainly designed to run on Ubuntu[7]. Cross-platform accessibility is mainly dependent on the usage of Python, GTK and GStreamer. However, due to the prior focus of the tool it is currently built to operate on Linux. Thus, changes are needed in order to conduct cross-platform experiments. Real-time chat (requirement G-4) is currently integrated over xmpp. Requirement G-5 (Codec support) is currently integrated via GStreamer, which supports open source codec libraries (e.g. libav and x264) for implementing various codecs. The last requirement within the general phase, G-6 (Support tool modularity), mainly depends on the degree of customizability that is needed to fit the experiment goal of the test conductor. Within the current tool framework GStreamer is concerned with the media processing. Furthermore, a range of design patterns are applied for separating various concerns. Experiment customizability is within itself a part of the tool that provides flexibility during experiment design and conduction (e.g. task set-up, control pane etc.).

Within the design phase requirements D-1 (Pre-defining the experimental conditions per trial subject), D-2 (Preliminary test integration), D-3 (Integration of test subjects into the experimental framework) and D-4 (Specifying the communication levels) are all integrated within the current tool framework. In order to adapt these functionalities to the experiment set-up, they can be adjusted through configuration scripts or set manually during experiment conduction. The last requirement of the design phase, D-5 (Integration of source material) is partially integrated with Ambulant[8].

[6]http://gstreamer.freedesktop.org

[7]http://www.ubuntu.com

[8]http://www.ambulantplayer.org

From the requirements tailored to the conduction phase, requirement C-1 (Real-time manual adjustment of technical conditions) is mainly dependent on the type of test conditions the experiment conductor is evaluating (e.g. delay). As GStreamer provides support for adjusting various parameters such as delay or video quantization, it is mainly dependent on the type of conditions the conductor aims to adjust (e.g. changing the codec within a running pipeline is not possible). With the implementation of the observer control client, requirement C-2 (Interaction flexibility throughout experiment conduction) is fully integrated within the current tool framework. The integration of requirement C-3 (Dynamically adaptable streaming) is fully supported by adapting to the integrated test conditions. Real-time annotation during experiment conduction (requirement C-4) is currently integrated by providing functionalities for adding small annotations and setting bookmarks during experiment conduction. Lastly, the tool currently provides basic support for handling crash issues (C-5).

From the requirements enlisted in the design phase, requirement A-1 (Subjective assessment integration), A-2 (Logging of technical conditions) and A-4 (Analysis of real-time annotations inserted during experiment conduction) are all fully integrated within the tool. Requirement A-2 (Logging of technical conditions) is supported within the tool for integrated technical parameters and contains log data gathered from GStreamer. Furthermore, a framework for logging (time aligned to the media) is integrated.

The comparison shows that the current state of the tool mainly lacks support from requirements G-1, G-3 and G-6 enlisted in the general section. Within the current tool framework requirement D-1, D-2, D-3 and D-4 are designed to be configured over scripts, and thus require coding. Furthermore, requirement D-5 requires additional coding in order to provide the identified support. Requirement C-5 currently provides basic support (e.g. clients can be restarted), but does however require additional coding in order to optimize this issue. Within the current tool framework a large scope of adjustment is already possible. This shows that a large emphasis should be put on the overall understanding of the tool components (e.g. supporting source code and documentation). Various tool concerns are currently separated with interfaces and layers, providing easier customization possibilities. These aspects, however, need to be formalized and documented so that other stakeholders can easily modify and adapt to it.

As each question within the online survey was treated independent and unique, significance levels were determined on an ordinal measurement scale. Although the results showed to be significant, possible influential factors, such as gender and age, have not been taken into account. These aspects were however not considered relevant for the study goal of this research. Another limitation of the methodologies applied in the requirement elicitation study is related to the sample size. Due to the small amount of respondents derived from the online questionnaires, the number of possible one-to-one interviews were limited. The small sample size could have been resolved by expanding the distribution of questionnaire invitations. Nevertheless, the sample population contained researchers from major companies and research institutions (e.g. Deutsche Telekom, BT and Ericsson) who are mostly active within notable telecommunication communities (e.g. ITU-T). Furthermore, the semi-structured interviews showed repetitive answers and patterns which indicated that main key requirements were identified. By integrating these requirements, a shift can be made towards a more standardized tool for conducting telecommunication experiments. This shows that an open-source approach can be beneficial as it provides a framework that can be used by a wide range of stakeholders. Furthermore, comparisons between future experiments will most likely be more reliable if they are conducted with the same tool.

The stability of the requirements is another issue to take into consideration. As research and technical developments are rapidly accelerating, requirements are most likely to change over time. Within the domain of this research, the derived QoE tool requirements mainly focus on ongoing user experience issues for future telecommunication settings. As results are derived from a community of researchers that all have experience in QoE assessment, the requirements are assumed to be more robust to future developments. However, in order to tackle this temporal dimension, a community-based approach could provide a useful method for supporting the evolving needs and goals of telecommunication users.

7. CONCLUSION AND FUTURE WORK

This paper presents a requirement analysis for a multiparty conferencing testbed, which is designed for conducting telecommunication experiments in the direction of QoE assessment. On the basis of semi-structured interviews, qualitative results were analyzed and documented in the form of requirements. In order to provide optimal tool support, the requirements are prioritized and categorized according to the design, conduction and analysis phase of the experiment. Furthermore, a set of common issues are identified which provide insight into general requirements concerning tool support.

From the requirements tailored to the design phase, it has been identified that pre-defining the independent test variables over the experiment tasks and test subjects is a crucial functionality which the tool should provide. Futhermore, manual adjustment of technical conditions and having interaction flexibility with the test participants, have shown to be important requirements during experiment conduction. Subjective assessment integration in the form of questionnaires and logging of technical conditions are important requirements that are considered essential within the analysis phase. Aspects influencing the overall tool usability are mainly focused on documentation, coding support and the degree of customizability the tool provides.

The comparison shows that various requirements are not implemented within the current framework of the tool. Thus, by integrating these requirements, a shift can be made towards a more flexible and accessible tool tailored to the needs and goals of potential stakeholders.

Future work directions should focus on the technical feasibility and implementation of the documented requirements. Afterwards, a usability study can be conducted with representative users to evaluate aspects concerning user experience.

Overall, the contributions of this paper are twofold. First, it presents a framework for further development of QoE assessment tools in the area of telecommunication studies. Second, it provides a specific focus on requirements that are needed to support the open-source development of the CWI multi-party conferencing tool.

8. ACKNOWLEDGEMENTS

This work was supported by the CWI, the national research institute for Mathematics and Computer Science (www.cwi.nl). I would like to thank Pablo Cesar for granting me the opportunity to conduct this thesis research at the CWI. A special thanks go to my supervisor Marwin Schmitt for his ongoing support and supervision throughout the learning process of this master thesis. I thank Frank Nack for assessing this paper and providing guidance and support throughout my Master program. Last but not least, I would like to thank my parents for their endless love and support.

9. REFERENCES

[1] P. Ballon, J. Pierson, and S. Delaere. Test and experimentation platforms for broadband innovation: Examining european practice. In *Conference Proceedings of 16th European Regional Conference by the International Telecommunications Society (ITS), Porto, Portugal*, pages 4–6, 2005.

[2] G. Berndtsson, M. Folkesson, and V. Kulyk. Subjective quality assessment of video conferences and telemeetings. In *Packet Video Workshop (PV), 2012 19th International*, pages 25–30. IEEE, 2012.

[3] K. De Moor, I. Ketyko, W. Joseph, T. Deryckere, L. De Marez, L. Martens, and G. Verleye. Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting. *Mobile Networks and Applications*, 15(3):378–391, 2010.

[4] D. Geerts, K. De Moor, I. Ketyko, A. Jacobs, J. Van den Bergh, W. Joseph, L. Martens, and L. De Marez. Linking an integrated framework with appropriate methods for measuring qoe. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 158–163. IEEE, 2010.

[5] D. Geerts, I. Vaishnavi, R. Mekuria, O. Van Deventer, and P. Cesar. Are we in sync?: synchronization requirements for watching online video together. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 311–314. ACM, 2011.

[6] A. Kappas and N. C. Krämer. *Face-to-face communication over the Internet: emotions in a web of culture, language, and technology.* Cambridge University Press, 2011.

[7] P. Le Callet, S. Möller, and A. Perkis. Qualinet white paper on definitions of quality of experience, version 1.1, june 3, 2012. *Lausanne: European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, 2012.

[8] M. A. Mehmood, A. Wundsam, S. Uhlig, D. Levin, N. Sarrar, and A. Feldmann. Qoe-lab: Towards evaluating quality of experience for future internet conditions. In *Testbeds and Research Infrastructure. Development of Networks and Communities*, pages 286–301. Springer, 2012.

[9] C. Robson. *Real world research: a resource for users of social research methods in applied settings.* Wiley Chichester, 2011.

[10] M. Schmitt, S. Gunkel, P. Cesar, and P. Hughes. A qoe testbed for socially-aware video-mediated group communication. In *Proceedings of the 2nd international workshop on Socially-aware multimedia*, pages 37–42. ACM, 2013.

[11] J. Tam, E. Carter, S. Kiesler, and J. Hodgins. Video increases the perception of naturalness during remote interactions with latency. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 2045–2050. ACM, 2012.

[12] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang. Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 481–490. ACM, 2009.

A. QUESTIONNAIRE FORMAT

Question	Type	Form
Q1. What is your occupation?	Open-ended	Text input
Q2. Are you familiar with conducting studies in a telecommunication setting? (if no, skip to question 9)	Close-ended	1 Yes 2 No
Q3. How would you rate your experience in conducting these type of studies?	Close-ended	1 Very poor 2 Poor 3 Fair 4 Good 5 Very Good
Q4. What type of tool(s) did you use to conduct these studies?	Close-ended *Multiple choice	1 Self-developed from scratch 2 Off-the-shelf software 3 Modified software 4 Other
Q5. How well did the tool(s) you used for these studies meet your experimental requirements?	Close-ended	1 Very poor 2 Poor 3 Fair 4 Good 5 Very Good
Q6. How was your experience with these tool(s) in terms of usability?	Close-ended	1 Very poor 2 Poor 3 Fair 4 Good 5 Very Good
Q7. How did you experience the support of the tools you have used throughout the general phases of the experiment?	Close-ended	1 Very poor 2 Poor 3 Fair 4 Good 5 Very Good
Q8. What kind of data do you usually collect within these types of studies?	Close-ended *Multiple choice	1 Questionnaires 2 Video data 3 Speech data 4 Other
Q9. How likely is your interest in using the CWI tool for conducting controlled experiments in a multiparty conference setting?	Close-ended	1 Extremely unlikely 2 Unlikely 3 Neutral 4 Likely 5 Extremely likely
Q10. What aspects could influence you in using the tool?	Open-ended	Text input
Q11. What type of studies would you conduct with this tool?	Open-ended	Text input
Q12. Are there any further suggestions/recommendations?	Open-ended	Text input

B. SEMI-STRUCTURED INTERVIEW FORMAT

Topic	**Question**

Introduction	Thank you for being willing to take part in (a follow up) interview (to the previous survey).
	- Inform interviewee of the interview structure
	- Name and position at CWI
	- Explain purpose and nature of study
	    - Requirement analysis for video conferencing tool
	    - Conduct telecommunication studies in a controlled environment
	    - Three components explanation
	        - Video client
	        - Observer control client
	        - Tool for data analysis
	- Goal: Make it open source and flexible so that other researchers can benefit from it. Gain understanding in their goals and needs.
	
	Questions are based on your workflow in previous telecommunication studies.
	
	Before we get started I can assure you that your answers will remain completely anonymous but I do ask you if I can record this session for further analysis.
	    (if no, only take notes)

Background information	Q1. Can I first ask you what your occupation is? a. Tasks/responsibilities b. Experience/familiarity telecommunication studies

Design/planning	In order to gain more insight in your workflow I would like to walk through your process of conducting experiments.
	
	Q2. During previous telecommunication studies, after setting the study goal of the experiment, which practical steps do you usually follow during the planning/design phase of the experiment?
	    (Probe on experimental set-up)
	
	Q3. How have you experienced the tool(s) support during this phase?
	a. Setting up experiment
	b. Experimental factor integration
	c. Ease of use
	    (Probe on issues and solutions)

Conduction	Q4. After the experimental design phase, conducting the experiment is the next step. Could you shed some light on how you have conducted experiments in previous studies?
	a. Process
	b. Experience flexibility
	    1. Adjusting parameters/technical conditions
	    2. Modify task set-up
	    3. Implementing new technologies

	Q5. During experiment conduction having control of the process (e.g. jump in when needed, changing delays etc.) is an important aspect of the tool. Are there any problems that you are currently facing in terms of tool control? a. Difficulties adjusting settings and controlling the experiment b. Interaction with test participants c. Other aspects Q6. How could this control be optimized in future studies (Ask in relation to all sources identified in 5.)? *(Probe on possible solutions)*
Analysis	Q7. What type(s) of data do you usually collect during the experiment conduction? *(Combination of data types possible)* Q8. How have you experienced the process from experiment conduction to data analysis (Based on data collection type)? a. Data collection Q9. How does the tool provide support in analyzing the experiment? a. Accessibility b. Interaction with data c. Analysis of data
Additional remarks	Q10. Thank you very much for your help and giving up your time. Is there any aspect or topic that has not been covered in this interview?