Evaluation of the Perceived Effects of a Video Conferencing for Windows Mobile

by

Xiong Wen

Thesis in partial fulfilment of the degree of

Master in Technology in

Information and Communication Technology

Agder University

Faculty of Engineering and Science

Grimstad

Norway

June 2008

Abstract

The aim of this thesis is to test the video quality in videoconferencing from the end user's perception based on the Windows Mobile EMP platform. Video conferencing is promising service in practice and the quality of the service is not well described. The video quality may be affected by many factors and the influence brought by the EMP platform as well as the artificial control is going to be tested and evaluated.

The Windows mobile architecture in general and the Windows mobile videoconferencing framework in special are illustrated in the thesis. This gives information about how the videoconferencing work in Windows Mobile system and how the EMP platform provides a framework for videoconferencing integration. On the other hand, the effects of a bad timing, to high resource consumption, data loss, and other effects that may bring the bad influence for video quality from the end user perception are discussed in the thesis.

This paper proposes the test cases about how to test the video quality automatically and chooses one simple and feasible one as the final solution. The measurement of video quality is decided from the existing methods and some improvements have made to give a comprehensive measuring of the video quality besides the traditional signal-to-noise respect. Also, it explains what the input is, how to create the test application and what the output of the system is.

Hopefully, the implementation of the system can be carried out in later research and the work we done will give help to video conferencing service.

Preface

This thesis is done as the conclusion of two-year Master of Science program in Information and Communication Technology (ICT) at Agder University, Faculty of Engineering and Science in Grimstad, Norway. The project has lasted from January to June 2008 and the workload equals to 30 ECTS.

The project is proposed by Ericsson Company and approved by Agder University. First, I would like to thank my supervisor in the university, PH.DAndreas Häber as he has given me much help throughout the whole project especially for my report.

Also, I owe much to Knut Bakke, who is my technical supervisor in Ericson and other colleagues in the company. There are many helpers so that it is pity that I will give a list of all the names. Their experience has contributed much to the project and prevents me from the repeated work. Finally, I would like to thank Mr. Stein Bergsmark and Mrs. Sissel Andreassen for their efforts in our studies during these two years.

Table of Content

Table of Figures

1 Introduction

1.1 Background

Mobile phones, the communication devices, are getting more interesting and sophisticated with an amazing growing speed. Since the communication technologies are emerging and developing, the quality of service for mobile devices has been improved to a large extend. Different functions and usage on mobile phones are created and renewed based on the development of technologies and standards. It is expected to reach higher connection speed, better security mechanisms, and quality during the communication for both voice and video. Video conferencing service is in practice but how to test the affect of video quality automatically and to what extent it influences the quality still need to be discussed.

Videoconferencing is kind of service that make use of communication network to transmit video and audio with two or more participants. The system is composed with terminals, MCU (Multi-point Control Unit) and network. The terminals are composed with video input and output, audio input and output, codec and network interface like modem. MCU is the device to control the terminals, bandwidth, and progress of the conference when the conference is multi-point. It is responsible for transmit the audio, video and data form the one terminal to other terminals. The network can be digital network like ISDN (Integrated Services Digital Network) and IP or the traditional PSTN (Public Switch Telephone Network). With more participants, one possibility for architecture is to employs central servers for conferencing control, which may create bottlenecks for performance. What is more, the cost for maintaining the servers brings the individuals away from the service. Another alternative is the peer-to-peer architecture with a direct connection between users which is better for two organizations can be taken into account.

One significant difference of videoconferencing service from generic data service is that videoconferencing is stream service with the large throughout. Without compression technology, the network resource could be saturated by the high video rate, which leads to the congestion even crash of the network. Consequently, the core technology used in a videoconference system is digital compression of audio and video streams in real time [1]. Codec (coder or decoder) is the hardware or software to perform compression and the compression rate differs from various standards. Two popular kinds of standards for codec in videoconferencing system now are H.26x and MPEG. The digital stream of 1s and 0s resulted from codec is subdivided into labeled packets, which are then transmitted through a digital network. The use of modems in the transmission line allows the low-speed applications in Plain Old Telephone System. One good case is videotelephony.

At present, there are three umbrellas of standards for videoconferencing system: ITU (International Telecommunication Union) H.320, ITU H.323 and ITU H.324. H.320 as an old standards, because its inherit limitation and high cost, has gradually faded in today’s application. H.323, subject to IP switching network instead of circuit switching network, is based on TCP/IP

protocols and applies to implementation of voice over IP VoIP. Since the problems of QoS (Quality of Service) in IP network has not been complete solved, network vibration, conditions like confusion for packet sequence, and network congestion always happen, which will bring bad effects for videoconferencing services. H.324 is the standard for transmission over POTS (Plain Old Telephony Service), or audio telephony network [1]. For 3G mobile devices, 3GPP has published 3G-324M standard for implementation of video call.

As mentioned above, stream media service is introduced to videoconferencing technology. Stream media service refers to the continuous media using stream transmission technology like the real-time audio stream in videoconferencing and data stream which is transmitted and broadcast at random. There are three ways of streaming media for video conferencing. The first is the unicast, which means a dedicated channel between terminals with high demand for bandwidth with good flexibility. This mode is better for bi-directional interaction videoconferencing. Multicast mode takes advantage of IP multicast technology to establish the special network capable of multicasting, which makes it suitable to view videoconferencing in one direction. The users can select the item to initiate the connection by ordering program mode and perform the operation like start, stop, forwards, backwards or pause. Peer-to-peer [2] (P2P) mode arranges for media to be sent from clients to clients, and it prevents the server and its network connections from becoming a bottleneck.

Another challenge for mobiles phones services is the Operating System (OS) as the platform. The platform of mobile phones is part of “Open OS” segment, which develops fast in the recent years. Many mobile phone companies have started to establish the department for “Open OS” to provide a more fast, convenient and secure platform for customers. The services like videoconferencing will run on the operating system and quality may vary depending on the platform. There are Symbian, Linux and Windows Mobile of platforms for various serial of phones produced by different manufacturers. Windows mobile is what be used for this project and it is an operating system for mobile phones provided by Microsoft. This operating system is based on Windows CE [3] (a technology of core) with a comprehensive application suite. Microsoft has provided its own framework for various applications including Microsoft Telephony API, which is the basic framework for video conferencing service.

Ericsson Mobile Platforms (EMP) will provide platforms integrating Windows mobile with the cellular technology from EMP. The hardware and software contains all the necessary components and functions for mobile-device-related communication services.

1.2 Project Definition

The project is defined as:
“Within this project an overview of Microsoft’s video conferencing framework for Windows Mobile will be presented. Additionally, architecture for video conferencing service with 3GPP standards will be given. Effects affecting the end user perception, such as bad timing, to high resource consumption and data loss, shall be described. Furthermore, a test setup for observing and/or measuring some of the described effects by utilizing the videoconferencing framework

shall be specified and implemented. Finally, an evaluation of the videoconferencing service based on the measured effects shall be analyzed and discussed."

1.3 Problem Statements

This project is to ensure a well-performed integration of a third party video conference in a Windows Mobile platform and to discover the possible effects for video conferencing service. For this purpose, what we need to do is to find a good way to test the video conferencing service in the Windows Mobile 6.x. For example, how the connection for video conferencing in Windows Mobile platform can be successfully set up via 3G network? If the video frames are delayed for a certain period of time, what will be the result for quality of video conferencing from the user perception? The platform in use is Window Mobile EMP supported by Ericsson Company.

Although a great many of applications are provided by Windows Mobile, video conferencing is still left to mobile manufacturers to implement on the platform. The thesis is expected to present the Windows Mobile architecture in general and the video conferencing framework on the platform in special. Though a framework for the integration of the service on Windows Mobile is available, we can not guarantee the quality of the service since we can not test by applying a real application with stacks for processing of streaming on Windows Mobile system. Furthermore, the video conferencing service provided by the project is tested to a level that will ensure good quality of integration of video conference in Windows Mobile enabled EMP platform.

Another problem is that the factors that will affect the quality of video conferencing, among which there are three important ones: timing, resource consumption, and data loss. There is chance that the service will be of little value since the bad quality of the service is caused by the distortion of image and speech based on the above factors. How they affect the performance of video conferencing is according to the results based on testing so that it is possible to find out appropriate improvements to achieve better quality. There can be some expectations of the factors but it is not sure whether the results can correspond with them. If so, we can come to a conclusion based on ideal situations. If not, we can discuss why the problems happen and what can we do to minimize the bad effects.

In order to solve the main problem, four sub-problems are defined and they are described in details in the following.

How to create a test application for video conferencing on Windows Mobile?

This task is to promise a successful of set up of video call and to control the frames of the audio and video so that the effects of timing, data loss and resource consumption on video conferencing service can be detected. The codec of audio and video should conform to the corresponding protocols in 3GPP as mentioned in the background. There are standards for the coding, multiplexing, error checking and other techniques to transmit streams. The problem is that what protocols stack should be taken and whether all these techniques are necessary to utilize for processing and which mechanism I should take. Also, how to use the Microsoft framework in videoconferencing application on Windows Mobile for call control is another important issue for this project.

How to set up and implement test environment on an EMP platform?
In order to test the application, the test environment should be set up. The test environments are about where we should put the devices or interface and how to make them working. The platform is an integration of hardware based on base band circuit including DDB (Digital Base-Band), ABB (Analog Base-Band), PMU (Power Management Unit), RF (Radio Frequency) etc and API provided for developing users' own applications. But for user application, the developers need I/O devices and programming to define the specific functions. The key problem is how to integrate platform and all the devices and what are the necessary components. The environments are critical for the performance of testing and the results.

Not all the module of the environments will be involved in video telephony application and how much of the environments should be implemented has to be decided. Maybe it is not difficult to set up the test environment, but how to implement it can be more difficult and time-consuming.

How to perform the practical tests based on test application?
The real work comes down to perform our own application. To do this, the voice and images are sent and they will be received on the same side with a loop. The test application will run on Windows Mobile EMP platform with a loop back to the reference phone. The video sent and received will not be the same due to the influence of channels, platform and other respects. This will generate some results for analysis and discussion later. It is expected that the results will be exactly the same every time we execute it. But It is not always the case due to the changeable environment including the time and space, hopefully this will not affect much in the conclusion.

Do the results correspond with the expected assumptions?
As we mentioned above, timing, resource consumption, and data loss can affect the performance and effect can be estimated based on theory study. The effects should be measured with some criterions or equipments and they are left for later research. But what have been tested may have some difference from the expected ones, so a study why this happens is needed for further research. If we know the reason for this, we can make improvements for our platform and provide better quality of services.

1.4 Problem Demarcations and Premises

Generally speaking, a full delivery of video conferencing service is not demanded in this project. The key problem is design a good way of testing the service on the Windows Mobile Enabled EMP platform. This has much to do with the structure of videoconferencing system and in this project, we are supposed to take peer to peer mode and MCU is not needed as we are only to test the service between two terminals. This is because videoconferencing with multi-point are much more complex and a server for controlling the service must be taken into account. To do this, the task might be too ambitious and hard to accomplish. The network we try to connect is supposed to be WCDMA network as traditional GSM may not afford good quality for video conferencing.

As mentioned above, the purpose is to test the important effects from the end user perception in

performance. In practice, not all the possible effects will be detected based on one test application and it may be rather difficult to test some certain factors. At the beginning, the factors we will take into account are the expected timing delay, resource consumption and data loss. It is possible that only one or two of them will be discussed and it is based on the later research work. There may be other factors, but in this project we will not consider it since there are not so significant compared to the ones mentioned above.

The following gives the required hardware and software for the projects.

Hardware

EMP test board will be supported by Ericsson Company. It is composed with many chips, and I/O devices which can support various applications. It is used for design and testing to develop mobile phones. To some extent, it is confidential.

Software

Windows Mobile is the platform we use in the project and general knowledge about operating system is needed. Embedded C++ is the language used in the project.

1.5 Motivations

As known for all of us, in the recent twenty years, mobile phones industry has developed so fast that life of humans has been affected and enriched to a large extend. From the first generation of analogue mobile phones to the personal smartphones with multiple functions, various services are provided and improved for mobile users. Videoconferencing as an interesting service becomes more demanding and promising and consequently, good quality is required for this service.

At present, there is no QoS guarantee for videoconferencing service and the research for this purpose is of great value from both theoretical and practical perspective. Some concerning technologies is quite new like the standards for codec we take in this project. With the implementation of these new technologies, we can find out the benefits of using them and expose the underlying risks.

The combination of hardware and software for mobile phones has become a trend that most companies are competing for. For a long term, the mobile phone segment “OpenOS”(Open Operating System) is expected to grow substantially. Though at present, Symbian [4] developed by Nokia initially, occupies the largest part of the market (over 70%), Windows mobile is almost the one who grows fastest and supposed to obtain a considerable market share. It is expected to dominate the future market when the hardware and software are advanced, especially for the software. From the view of developers, the developing of mobile phones will not be separated but be done on the basis of same codes. The sharing of resources saves the cost of research by Microsoft and also reduces the difficulty for the third party to program the cross-platform applications. This is why we want to develop video conferencing on Windows Mobile instead of other operating systems. Also, it can be found out whether the video conferencing framework of Windows Mobile can perform well on EMP platform.

Based on the result got from the project, we can find out the factors to influence the videoconferencing service and to what extent the effect can be for the performance. This will help to improve the quality of videoconferencing service for further mobile phone development.

As a result of this project, I can have deeper understanding of videoconferencing-related technologies, and videoconferencing framework on Windows Mobile. I will appreciate the new knowledge and experience of this project and it will help me in the corresponding field. Since the project is supported by Ericsson, there is a lot of practical work to do. The practical work of testing gives me advantages for future career.

1.6 Report Outline

The thesis is structured as following:

Chapter 1 gives an introduction to the master project which is the current chapter.

Chapter 2 includes the basic theories and related technologies about videoconferencing and quality measurement. This provides good reason for the later solutions of the project.

Chapter 3 proposes the possible solution that I have thought about and decides one as the main solution.

Chapter 4 discusses the further work that we can achieve in the later research.

Chapter 5 gives the conclusion of the project based on the work.

2 Theories and Literature Review

2.1 Video Conferencing in 3GPP

3GPP, as and abbreviation for “Third Generation Partnership Project” was found as a collaboration of agreement in December 1998 [14]. The current organizational partners ARIB (Association of Radio Industries and Businesses), TTC (Telecommunication Technology Committee) in Japan, ETSI, T1 in America and TTA in Korea [14] are participating in the activities of proposing drafts and approving standards for the third generation communication system. The ITU organization accepts and evaluates the proposals as drafts of standards from organizations and associations mentioned above all over the countries and areas. The function of ITU-T is to provide global telecommunication standards by studying technical, operating and tariff questions [15].

2.1.1 Circuit Switching Mode

In communication network, there are circuit switching, packet switching, frame switching, ATM switching, IP switching and optical switching mode used by switching devices. In popular WCDMA R99 [5] network, though the core network is IP based with high data rate, QoS can not guarantee good quality. As a result, the protocol taken by video conferencing service is circuit switching oriented and I will only explain this mode in details. So called circuit switching mode, is to establish a dedicated route between two ends, and the route starts from the sender and connects one station by one station. Once the connection has been set up, it will maintain for dedicated state which can not be used by others until the communication ends. Only if disconnected after finishing communication, this route can be left for other users. At present, this mode is taken by telephony and telegraph system.

Circuit switching [6] mode refers to the process of establishing connection, and dedicating the connection to certain users until they are released. Circuit switching network contains a physical channel and supports the single connection between the two ends in the process of network connection. Traditional voice telephone service carries out circuit switching mode by PSTN (Public Switch Telephone Network) instead of voice IP.

In this network, the CS area is based on the architecture of 64kbs MSC (Mobile Switching Center) and all the video from UTRAN has to be coded, transformed and transmitted to core network by circuit switching mode. There is a 64kbit/s channel dedicated for mobile video communication service in circuit switching area of WCDMA R99 network. The terminals for video service should conform to the standards mentioned in the next section.

It is expected that 3G network is forwarding to packet switching mode but for video communication service circuit switch mode still has some advantages. Communication channels using packet switching are not dedicated to the transmission of information between the source and destination. Under the condition that data are required to be transmitted in order with invariable fast speed, circuit switching is the ideal choice. To set up the connection at the

beginning, there is some time delay; messages can be sent at real time after successful connection, and the time delay for transmission can be ignored. Once the physical channel has been set up between the terminals, the real time communication can be promised. The data is sent in order during the communication so that there is no problem about disorder. The transmission of data is liable and fast, and the data can keep it original sequence without loss. Both analogue signals and digital signals can be transmitted using circuit switching mode. As the channel is “transparent”, users do not care about the structure of transmission channels which simplifies the application for users. From the view of control, the switching devices are easier. Moreover, this mode can promise the demand for bandwidth to ensure the quality of video conferencing. Consequently, circuit switching is better for system requiring large data transmission with high quality like telephony switching, file transmission, high-speed fax and less better for burst service and error-sensitive data service.

However, without error control, the liability of data exchange is lower than packet switching mode [7]. Since the occupied bandwidth is fixed and the connection is dedicated, the efficiency of network resources is relatively low. Users have to bear high cost when renting the digital channel to transmit data. Sometimes, the channels intend to be wasted at idle time and the time for establishing and detaching channels can not compensate for the data transmission in short periods of time. All these are the disadvantages of circuit switching mode that has not been solved now.

2.1.2 Protocols for Videoconferencing

As we mentioned above, for video conferencing application, both the speeches and images are sent from the source to the destinations. The speeches and images are transmitted in the form of audio and video signals. The signals have to be coded because they can not be transferred directly in the channel because of its property. The specifications of standards of audio and video codec are made by 3GPP. There are four framework protocols for video conferencing: H.320, H.324, H.310 and H.323. For this project, the video conferencing should conform to 3G-324M [8] standards of framework constituted by 3GPP. 3G-324M is the 3GPP circuit-switched mobile video telephony standard based on ITU-T Recommendation H.324/M [9] and other international standards. As a result, the implementation of codec for audio and video also has to summit to ITU-T Recommendation H.324/M.

Since transmission using IP results in the bad quality of conversational delay-sensitive communication, the problems of IP based videoconferencing communication using 3G network still can not be solved thoroughly in a short period of time. 3G-324M is regarded as a solution to provide convinced quality for these communication services. Another important point is that, it wins the support of most 3G technologies including W-CDMA, TD-SCDMA and CDMA2000 and is welcome by both the service providers and device developers in 3G solutions.

3G-324M allows the delay-sensitive multimedia communications like videoconferencing and video on demand with 3G technologies. The first 3G real-time multimedia service based on 3G-324M appears in Japan and expands to other countries. On the other hand, 3G-324M enabled mobile terminals have been developed to provide various 3G-324M based services.

As mentioned above, 3G-324M is different from the general communication protocols based on IP. Instead, it works on the circuit switching based TDM (Time Division Multiplexing) channels, which set up on the baseband protocols between the two communication parties. The advantage of TDM is the low time-delay service without routing at every hop on IP communication paths. The low time-delay services as videoconferencing in high error bit rate environment are improved to works well in public cellular network.

3G-324M network seems like ‘backwards’ to circuit switching times instead of ‘next generation IP’. But different from IP, it can carry out conversational video calling in cellular network as well as help the service providers and device developers to provide the services requiring wide band.

From the technical view, 3G-324M is extremely similar to H.324/M, but it appoints H.263 [10] video codec as a mandatory basic bench, and recommends MPEG-4 for video codec. As for our application we will choose H.264 as video codec to obtain better quality. Annex C of H.324 defines the use of H.324 terminals in error-prone transmission environments ("also referred to elsewhere as H.324/M") [5]. Other ITU-T Recommendations in the H.324/M-Series includes the H.223 [11] multiplex, H.245 [12] control, and G.723.1 [13] audio codec [9], which is the same as H.324. But the effective transmission methods in error-prone environment will be given in 3G-324M. The following figure 1 illustrates the architecture of H.324 protocol stacks. As the codec is the most important part for video conferencing in our project, I will explain them instead of other control protocols.

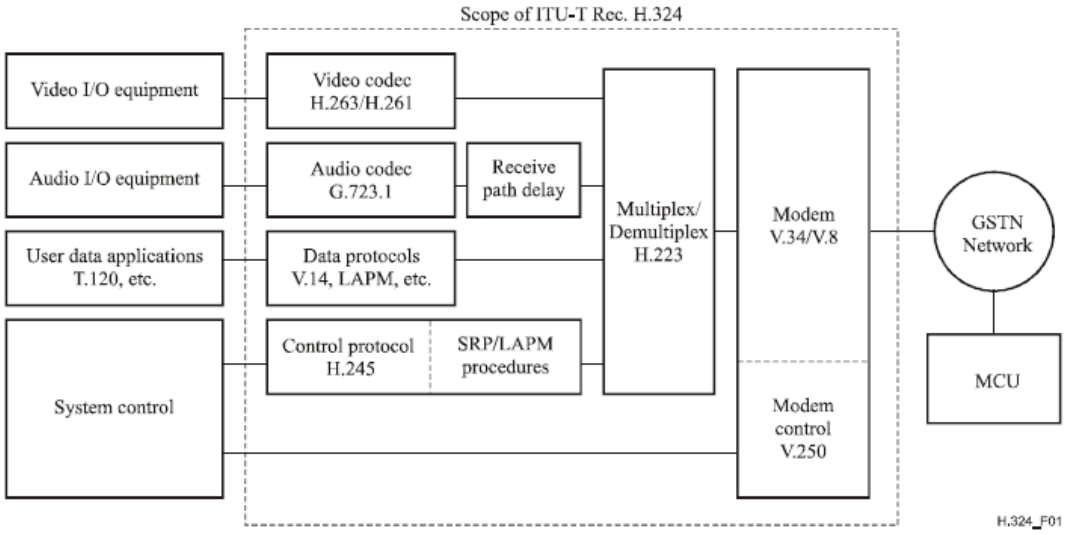

Figure 1 Architecture of H.324 [9]

2.1.3 Video Codec

Video compression is the core technology in multimedia applications. The compression standards for low bit rate video proposed by ITU-T plays an important role for the practice of video applications. The same bit rate can be lowed to a half of H.263 or MPEG-4 by H.264 (when we talk about MPEG-4, it normally refers to the part 2 of MPEG-4 standards). In other words, the maximum of signal-to-noise rate can be increased by 2db with H.264 technology.

H.264 has broad prospect for applications like real-time video communication, Internet video transmission, video streaming service, multi-point communication, video compression storage and video database. The main features of H.264 come down to three aspects. The first one is the focus on practice with mature technologies in pursuit of high coding rate and brief manifestation. Additionally, it applies the multi-layer technologies to separate the coding and channels, considering the characteristics of channels in the algorithms of source codec. Last but not the least, significant improvements have been made in key components like

Because of the strong compatibility with channels and high compression rate, H.264 is more popular in digital video communication and storage. But the good quality of H.264 is at the cost of computation. It is estimated the complexity of coding computation is triple of H.263 and decoding complexity is double of that.

2.1.4 Audio Codec (more)

ITU-I standards do not have restrictions on audio codec; only audio services based on IMT-2000 requires AMR (Adaptive Multi-Rate) [16] are required to support 3G-324M devices. G.723 is the old recommendation for audio coding, compatible with H.323. This recommendation covers the audio coding for compressing the speech or other audio signal component of multimedia services at a very low bit rate as part of the overall H.324 family of standards [13]. The highest data rate of AMR is 12.2 kbps depending on the distance of base stations, the signal interference and traffic. The advantage of AMR is that it can offer good quality of voice at best efforts in the current channel.

2.2 Microsoft Telephony API (Windows Mobile)

2.2.1 Microsoft Telephony Overview

What is telephony? It is service that integrates computers with communication devices and network. From a wide view, telephony includes both the traditional calls and the related modern communications. For traditional telephony, the devices were limited to normal stable telephone and the network was the Public Switched Telephone Network (PSTN). Since the development of computer, mobile devices and network, telephony has not only referred to what is mentioned above, but also expanded to cover applications with devices like camera and network as GSM, CDMA and Internet.

Possible telephony applications include [17]:
- Multicast multimedia IP conferencing
- Voice calls over the Internet (VoIP)
- Automatic Call Distribution (ACD) Center client and server applications
- Basic voice calls on the PSTN
- PBX-like controls such as call park and selective forwarding on a corporate phone network without the need to invest in specialized hardware

- Interactive voice response (IVR) systems
- Real-time collaboration

Microsoft Telephony architecture is shown in the figure 2 and the components are described as following.

Windows Mobile provides the universal API (Application Programming Interface) for various kinds of service and it helps developers to write their own applications on different purpose. The video conferencing application to be tested on Windows Mobile system in this project is based on the knowledge of Microsoft telephony framework. For communication-related TAPI is "Telephony Application Programming Interface".

TAPI is a suit of functions of programs related to telecommunication service. Dated to 1994, it has been developed by Microsoft, Intel and other telecommunication companies. TAPI has provided universal methods for different kinds of hardware, and information including data, voice fax, and video can be transmitted by TAPI. As a result, programs using TAPI has strong compatibility and different Windows programs can share devices.

TAPI integrates remote communication with operating system and support the traditional and IP telephony service to provide voice, data and video communication. The hardware required for the service include the audio and video card, modem, ISDN cable, ATM network and camera, etc. with these hardware, it is possible to connect to the local computer, telephone line, LAN, WAN and Internet to communicate. Besides making and receiving calls, the programs can take advantage of TAPI to provide the enhanced telephony functions, for example, incoming call display, the voice mail and video conferencing.

TAPI service providers are an abstraction layer between TAPI applications and the underlying hardware and transport protocols [18]. "Service Provider" is an interesting name for diver and a TSP (Telephony Service Provider) is a driver that enables the communications between TAPI application programs and TAPI devices. "A telephony service provider (TSP) supplies call controls and a media service provider (MSP) supplies detailed control over the media of a call. A TSP/MSP pair can implement the capabilities needed for applications ranging from basic telephony to IP multicast multimedia conferencing." [18]. When other devices are required like PBX's and voice card, TSP provided by hardware manufacturers has to be used. TSP are responsible for interpreting API functions into commands that the hardware can understand and transformed the events returned by hardware to the forms that TAPI programs can recognize. Since different communication devices varied in characteristics, different TAPI functions are supported by different TSP. One possibility is that a certain TSP can support several kinds of devices and according to the access devices, the TSP play the corresponding role. For example, if TSPI are applied to enable the calling display, and at the same time, the modem can support this application, Windows application programs can get the calling information using TAPI. With the same TSP, if the modem is not able to do it, the application can not be carried out. Otherwise without the support of TSP, modem alone can not make the application perform successfully. Consequently, if the application programs can not run as expected, one reason may be the

problems of the TSP or device or both of them.

Another important issue is the compatibility for TAPI programs and TSP. As the operating system is forward-oriented compatible, if the operating system which can only support TAPI2.0 is employed, but the application programs and TSP can only support TAPI2.1, the programs will be run on that operating system. When TSP is loaded by TAPI, TSP will automatically identify the current version of TAPI and check whether it can support this version. A TSP can only be installed in its supported environment. Just like TSP, the application programs will also verify the compatibility with TAPI version and it must find a TAPI version that is not higher that the version by operating as well as installed TSP.

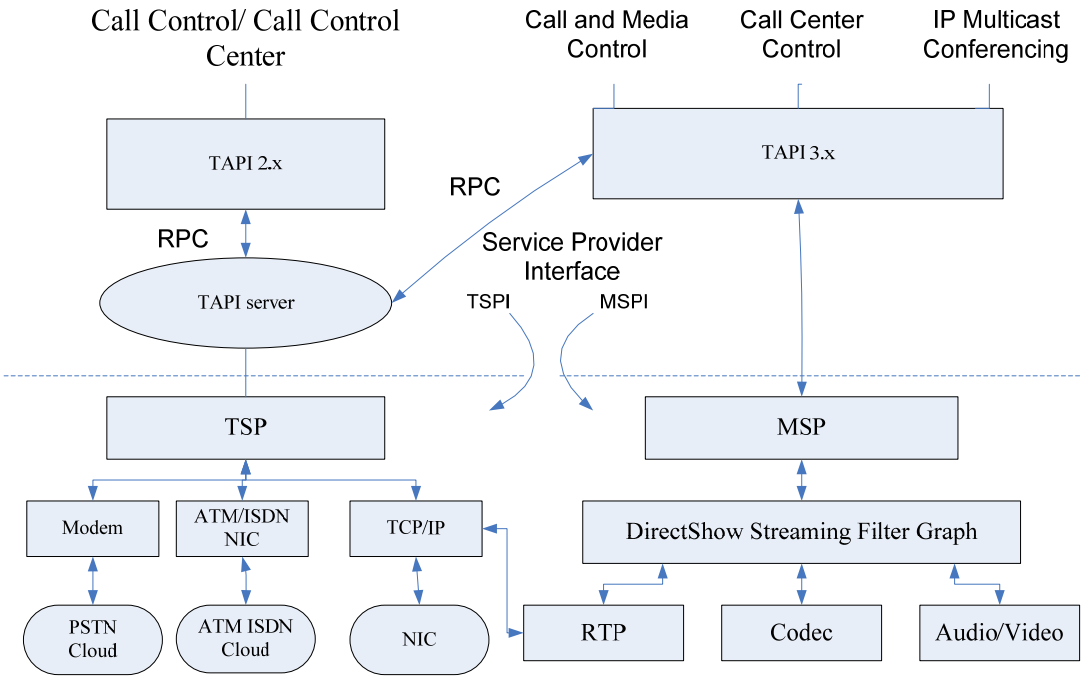

Figure 2 Microsoft Telephony Architecture

The diagram (figure 3) bellows shows the components for TAPI applications. A complete telephony provides basic services with necessary call functions, supplementary services with functions as hold and transfer, and extended services to allow the developers access the functions given indirectly by TAPI [19]. Actually, not all of them are necessary in every application program since the level of service mentioned above has to be defined first when starting the application. Either TAPI 2.x or TAPI 3.x can be selected for the application.

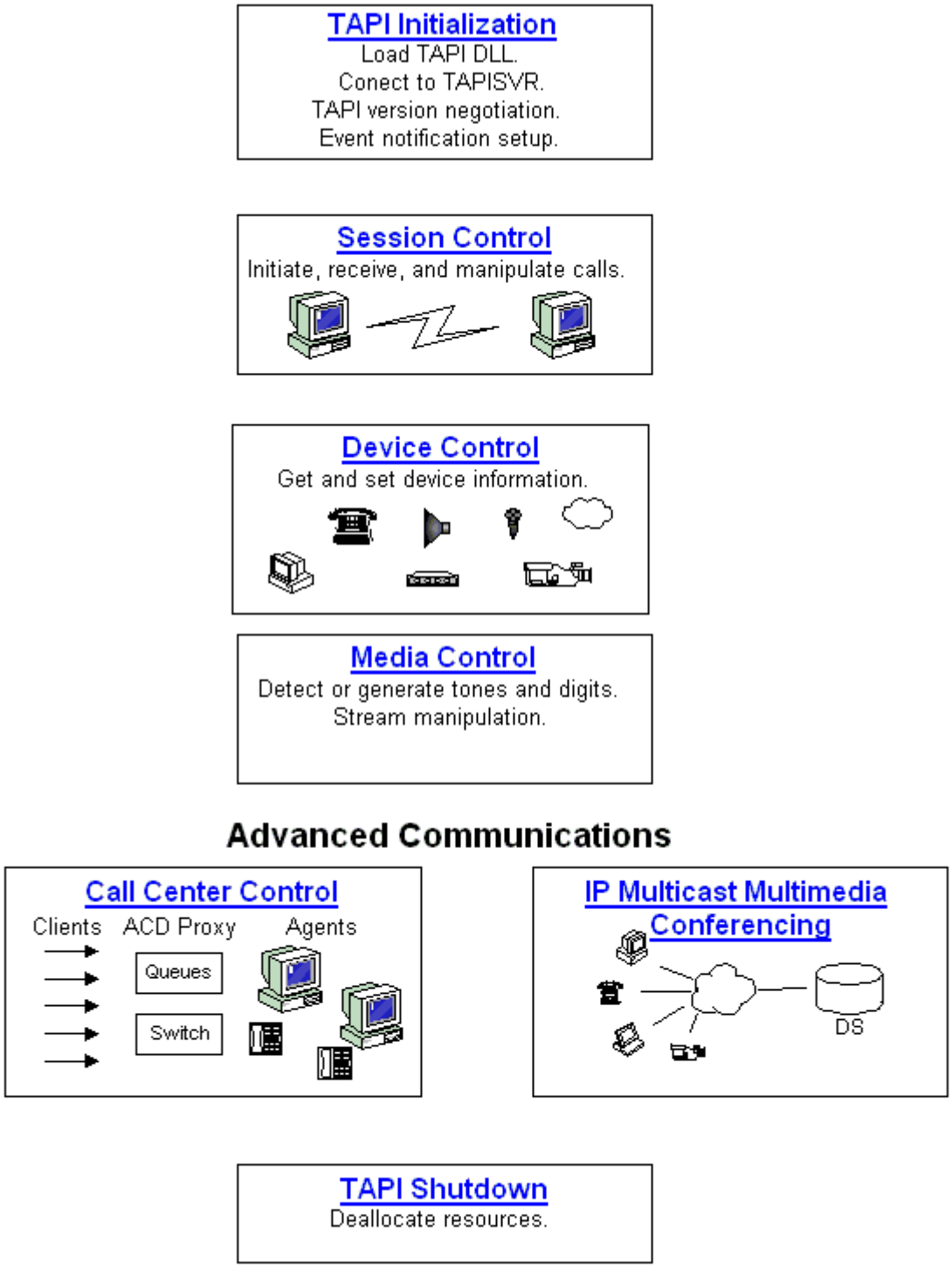

Figure 3 Components for TAPI Applications [20]

2.2.2 Introduction to TAPI 3.1

As mentioned above, there are several versions of TAPI including TAPI2.x and TAPI 3. Here I will focus on TAPI 3.1 which is the version to be used. TAPI version 3.1 is a COM-based API that

merges classic and IP telephony [msdn]. Four major components to TAPI 3.1 are COM API, TAPI server, Telephony Servive Providers (TSPs) and Media Stream Providers (MSPs), whose relationship is illustrated in figure 2.

TAPI 3.1 architecture is presented in the following figure 4 [21].

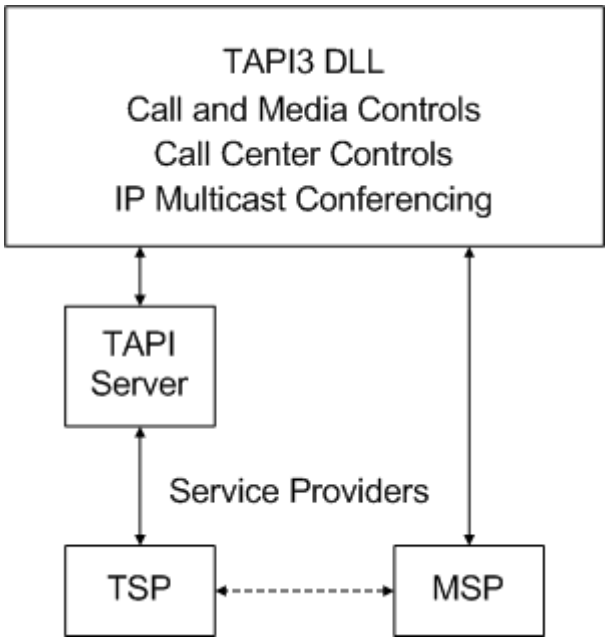

Figure 4 TAPI 3.1 Architecture

Seen from the above structure, TAPI 3.1 contains three control interfaces for developers, which are Call and Media Controls, Call Center Controls and IP Multicast Conferencing.

Call and Media Controls are composed of a suite of COM objects, interfaces and methods to establish paging among two or more computers. Five main objects are TAPI, Address, Terminal, Call and CallHub. TAPI object stands for all resources of telephony. Application based on TAPI 3.1 should create an instance of TAPI object first and initialize it. Address object defines an entity to set up and receive a call. Through this entity, application programs can identify whether the designated address can support certain type of media, list the calling related to certain address currently and transfer a calling. Terminal objects define an initiator or receiver of the stream, for instance microphone or speaker. Application programs can select appropriate Terminal to transmit streams. Connections among two or more addresses are given by Call objects. CallHub objects provide a community in multi-calling. Only with necessary priority, CallHub objects can control other participants in the process of calling.

Call Center Controls help the developer to set up the calling center with a serial of objects. Core functions like pre call number, calling queue, route management can be realized by these objects.

IP Multicast Conferencing Controls allows developers to create IP conferencing system with multiple points. It is carried out by three parts, Directory Controls for managing the conferencing

list of server, Conference Blob Controls for controlling the given conference and Multicast COM interfaces to allow application programs to obtain the multi-point transmitting address from the server.

TAPI Server process abstracts TSPI (Telephony Service Providers Interface) from TAPI 3.x and TAPI 2.x so that TSP of TAPI 2.1 can also be used by TAPI 3.x. TAPI applications load TAPI DLL into their process, and TAPISRV is the TAPI Server process, which communicate with TAPI through a private RPC interface. TAPISRV is implemented as a service process within SVCHOST [21].

TSP and MSP constitutes Service providers. They cooperate to provide the service for telephony, in which TSP serves for Call Control and MSP for Media Control service.

Telephony Service Providers (TSP) accept the calling from TAPI which has no relationship with protocols and transform them into calling related to corresponding protocols. Two bound TSP related to IP is H.323 and IP Multicast Conferencing TSP.

Media Service Providers (MSP) provides a universal interface to control various media streams of callings.

2.3 Possible Effects for Videoconferencing

The effects for video conferencing can be summarized to two aspects, which are the transmission property of the network and the contents of video conferencing. The first aspect contains the time delay, the jitter and the packet loss (frame loss). The codec for the streams are the main factors for the contents.

2.3.1 Time delay

One important aspect that affects the end user’s perception of mobile video conferencing is delay. In practice, time delay results from many factors, including the network, the terminals and the call set up, etc. As mentioned above, the time delay for transmission in circuit switching network can be ignored. Actually, the videoconferencing system based on circuit switching network has less latency of transmission because the network has provided better insurances for the time-delay mechanisms.

In videoconferencing service, time delay happens to both audio and video streams. However, the time delay for audio is much more critical than that for video. It is supposed that the voice in mobile networks in one direction (point to point) had better to be no more than 150ms [22]. The delay is caused by the time spent on coding and decoding of the audio and the time delay in the network. Users will have discontinuous feeling of voice with too much latency. Humans can not detect much difference of voice with delay time in 50ms, and if larger than 50 to 200ms, people can hear the intervals of voice but it will not bring much interference for the understanding for voice and communication. But the quality of voice will be seriously damaged under the condition that the latency is more than 200ms.

The time delay for audio in the transmissions system can be divided into One-Way-Delay and Round-Trip-Delay. The former one refers to the transmission time of voice from the sender to receiver and it indicated the transmission quality of the network, which exerts the indirect influence on the audio quality. The Round-Trip-Delay is normally the double of One-Way-Delay as the time for audio transmission from the sender to the receiver and then back. The test exactness depends on the precision of the equipment and the clock of both sides.

For the video streams, the delay between two end users is more tolerant than audio and a value of several seconds is still acceptable. But if more than several seconds, the video may not be continuous and clear, resulting in the distortion of the images in video.

Actually, the exact computation for time delay is not necessary, because of the obvious difference time delay in the variable environment. But the effects still need to be tested approximately, since the poor quality will be caused by the large time delay. One example is to open the local video and received video at one terminal with a camera targeted at a clear clock and the time delay between the two video is the rough latency.

Another problem related to time delay is synchronization (lip-sync). This problem is obvious in the system with codec. The lip-sync related delay is supposed not to be so evident without the encoding and decoding of the media streams. It is accepted that systems must maintain a constant lip-sync of within 50ms to enable natural face-to-face conversation [23]. To achieve the synchronization between audio and video streams, the audio streams have to be delayed in order to match the video streams. This is because the time for capturing and encoding a single video frame is longer than that for audio frames.

2.3.2 Jitter

Jitter is another factor that will affect the quality of both audio and video. It refers to the difference of time delay between the neighboring packets or frames. The jitter buffering is user for smoothing the difference of arriving time for audio and video packets.

The reason for jitter derives from many respects. The performance of audio and video codec, the congestion of the network and the property of network devices can lead to the jitter of audio and video. For audio, if the time delay varies from time to time, voice that people hear will change in terms of the speed and users will have bad perception. Even worse, the communication between the users will be discontinuous at random. On the other hand, the detection of jitter on video will contribute to discovering the worse tendency of video quality during the transmission.

2.3.3 Packet Loss (Frame Loss)

It may happen to every network that the packets or frames can not arrive at the destination because of the delay or arrive in disorder. In videoconferencing system, the packet loss or frame loss has the direct influence on the quality of the video. The three types of encoded video frames are Intra coded frames (I frames), Predicted frames (B frames) and Bi-directional predictive frames (P frames). It is proved that loss about any type of the frame (I, B, P frames) especially the key

frames will lead to the deterioration of the video quality in different degree without the compensation of video decoding or retransmission mechanisms. But the loss of B frames will not cause the bad effect of other frames on quality; the only influence is the lowing of bit rate.

If the received frames are delayed exceeding the threshold time, they will be discarded. In TDM and SDH/SONET network, there is no mechanism for frame retransmission but in WCDMA it is not the case. In WCDMA 99, the frame retransmission mechanism is defined in data link layer, so that the compensation for the video codec is not critical as the TDM network.

Since the frames can be tagged in order, it is possible to see whether the received the frames are in order. The disorder of frames will cause the distortion of voice and images. It is interesting to see how the frame order affects the video quality and what their relationship is. The reset of frames is necessary to make sure the video are organized in the correct order according the sequence number of the frame header.

2.3.4 Codec

The quality of video can be affected by the codec according to the different compression algorithms. It is regarded that H.264 has the highest compression efficiency followed by MPEG-2 and H.263 respectively.

Under the same type of codec, the higher streaming bit rate will lead to better video quality. Actually, any codec is costly; the detailed information can be lost after encoding and compression. Generally speaking, the video quality after decoding is in the direct ratio of coding bit rate.

2.4 Evaluation methods for videoconferencing quality

2.4.1 Subjective and Objective Video Quality Assessment

To evaluate the video communication system, the quality of video is required to shown to the observers. It is supposed to be a tough task to measure the video quality and also not precise because of the various factors that affects the measuring. The video quality is born to be a subjective, which makes the requirements for correct and exact of results even more difficult to be obtained. The video quality for an observer or user depends on the task itself, for example, the passive watching of a DVD film, the attendance of video conferencing or trying to recognize a person from the video scene. The objective assessment of video quality can present exact repeatable outcomes but there is not any objective measurement can simulate the subjective visual perception of human beings.

Concluded from the above statements, there are two kinds of measuring methods for video quality, which are subjective assessment and objective assessment. Subjective assessment for video quality is more reliable, but it is too complex to be carried out. The computation speed of objective assessment is much higher with repeatable testing results and even can be used in real-time detection. However, the results from objective assessment should also conform to those derived from the subjective assessment.

In subject assessment [24], the reference video is broadcasted to observers and the marks given by them based on their subjective impression should be recorded. All the marks will be gathered to calculate the MOS (Mean Opinion Score) as the result. That is to say, the quality is assessed from the subjective visual perception of humans. The subjective perception for the video is determined by eyes and brain, as the complicated interaction of human visual system for different elements. The perception for video is affected by the spatial fidelity and temporal fidelity. The spatial fidelity counts for the fact that whether human can clearly view every section of the video scene wherever the obvious distortion exists. The temporal fidelity is about the smoothness of motion. However, the opinion of the observers for the video quality tends to have much to do with the observing environment, the mood of the observers and the interactive programs between the video and observers. The users for a certain task should pay attention to the corresponding part of the video. The concept of good quality between evaluating the video and watching a film is quite different. For instance, the video quality from the perception of an observer of user under good environment is better, which does not rely on the quality of video itself. Other influences includes the visual focus, from where the observer or user look at the video instead of looking at all the respects of the video at the same time, and “the newest effect” which refers that humans always notice the change caused by the updated contents other than the old. All these are what make the evaluation difficult to carry out as mentioned above. For the kind of measurement, ITU has defined two standards for this type of measurement in ITU-R BT.500-7. Those are DSCQS (Double Stimulus Continuous Quality Scale) and SSCQR (Signal Stimulus Continuous Quality Scale).

The complexity and cost of subjective assessment hesitates people to apply for video quality evaluation and objective assessment [25] becomes an alternative to solve this problem. Without the help of ender user, the objective video quality measurement can be done by programs or software which process the video signal to produce the quality. Also, it can provide real time quality monitoring for video applications [26]. Objective assessment is recommended by ITU-R VQEG. VQEG has given two simple parameters PSNR (Peak Signal-to-Noise) and MSE (Mean Square Error). Furthermore, many models of video quality are based on the perception of human eyes. Three typical types of objective video assessment metrics of objective methods are full reference, reduced reference and no reference. The full reference type is a method that compares the video source and the processed while reduced reference is about the comparison of a reduced video source to t full result. Obviously, no reference type is conducted without any reference signal. One example is to estimate the distortion directly from the processed signal and the disadvantage is the uncertainty for the visible distortion without any reference.

Actually, the most popular applied type for testing the video quality is the full reference model. The following figure 5 gives the model of full reference quality evaluation method of video.

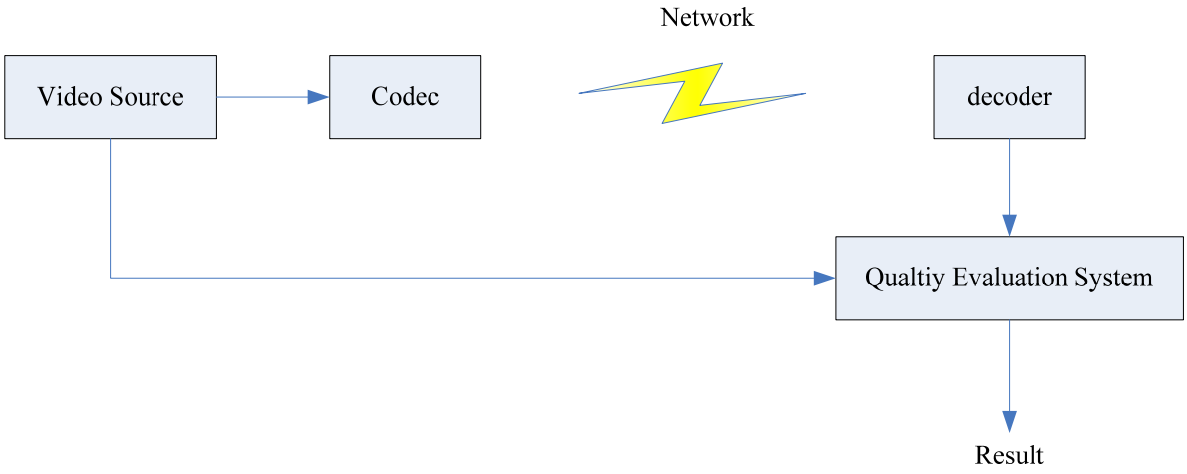

Figure 5 Full-reference Video Quality Measurement

The input of the model is the both the reference signals and depressed signal to be detected, and the output is the difference between them. The whole process involves the pixel-level processing, the temporal and spatial alignment for the input and output streams. Furthermore, it takes the human perception for video into account to make it accord with results from subjective assessment.

2.4.2 Metrics for Objective Video Quality Evaluation

The metric for evaluating the video quality can be divided into mathematical-based type and model-based type. From the literal view, the former type can give a computable result based on the mathematical formula, and PSNR belongs to this type. The model-based metrics are more complex since they depend on the model of Human Visual System instead of mathematical computation. At the same time, this feature determines good relationship between the model-based metrics as objective assessment for video quality and subjective assessment. As showing below, JND, SSIM, JND, VQM, DVQ, PDM, MPQM, and PVQM are model-based type.

A wide range of full reference algorithms have been developed including: MPQM (Moving Pictures Quality Metric - 1996) from EPFL in Switzerland, the US Government NTIA ITS lab's VQM (Video Quality Metric – 1999), Sarnoff's JND (Just Noticeable Differences), and Wang's SSIM ('Structural SIMilarity) [28].

- PSNR (Peak Signal-to-Noise Rate)
- SSIM (Structural Similarity Index Metric)
- JND(PQR) (JND- Just Noticeable Difference, PQR-Picture Quality Ratings)
- VQM (Video Quality Model)
- DVQ (Digital Video Quality)
- PDM (Perceptual Distortion Metric)
- MPQM (Moving Picture Quality Metric)

- PVQM (Perceptual Video Quality Measure)

2.4.3 PSNR

PSNR is the most popular metric for objective assessment of video quality because it is simple and fast to obtain based on the input data. As a full-reference metric, it describes the video quality from logarithmic level. Generally speaking, the higher value of PSNR implies the better video quality for the same tested video. It can be calculated at frame level to show that how close the frames of processed video streams are similar to those of the source frames. Also, it can show the quality from the perspective of overview with the maximum value based on the maximum value of the luminance part of video signals.

However, since the metric only consider the luminance component of video streams, without caring of chrominance component, which is also an important factor for user perception, it seems too monotonous to reflect the detected video quality. Furthermore, PSNR does not perform well to show the subjective assessment of video quality. That is to say, video with enjoyable subjective human perception does not necessarily have high PSNR value. Under this condition, it is possible that video may have comparatively low value, but with the good feeling of clearness caused by the sensitive section of human perception.

The following formula shows how to calculate the value of PSNR.
PSNR(db) = 10log(10)(2^n-1)^2/MSE
in which MSE means the Mean Square Error, and n is the number of bit for frames.

For a video stream which has N * M pixels with m-bit depth, the MSE can be described from the frame level or the holistic video stream.
From the frame level, the MSE can be calculated as:

$$MSE = \frac{1}{NM.K}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{m=1}^{M}\left[x(i,j,k) - \bar{x}(i,j,k)\right]^2$$

where K is the number of frames in the video stream; x(i,j,k) and $\bar{x}$(i,j,k) are the pixel luminance value in the (i,j) location in the kth frame for the original and processes stream respectively [28].

For the holistic video stream, MSE can be:

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[(f(i,j) - F(i,j)\right]^2}{M \cdot N}$$

in which, f(i,j) is the maximum value of luminance of original video stream at pixel (i, j), F(i, j) is that of processed one [29].

2.4.4 SSIM

SSIM is another approach for video quality assessment illustrated in reference [30]. This metric compares the structure of referenced signal and the processed signal as “a measurement of

deviations in luminance, contrast and structure" [31]. The value of SSIM index is between 0 and 1, in which a value of 0 means no correlation with the original, and 1 means the full match between the original and the processed video as an ideal situation. Consequently, it shows the "perceived structural information loss" [30] rather than the perceived errors. Compared to PSNR, the relationship between SSIM and subjective assessment of video quality is closer, being proved by [30]. Furthermore, the SSIM index can be applied to the Y, Cb, Cr color components independently and combined to a single quality measurement with weighted summation. More comprehensively, this approach takes both the luminance and chrominance of video into account to obtain an overall quality for the whole video stream. However, comparatively, SSIM can not work well to evaluate the degraded video that is seriously distorted.

The resulting new measure is named SSIM between signals x and y:

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{(2\,\mu_x\,\mu_y + C_1)(2\,\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where are respectively the mean of x, the mean of y, the variance of x, the variance of y, and the covariance of x and y. C1 and C2 are the constant.

The quality index with Y, Cb and Cr components of the j-th sampling window in the i-th video frame is given as following:

$$\mathrm{SSIM}_{ij} = W_Y\,\mathrm{SSIM}_{ij}^{Y} + W_{Cb}\,\mathrm{SSIM}_{ij}^{Cb} + W_{Cr}\,\mathrm{SSIM}_{ij}^{Cr}$$

The frame level quality can described by combining the above values as:

$$\mathrm{Q}_i = \frac{\sum_{j=1}^{R_s} w_{ij}\,\mathrm{SSIM}_{ij}}{\sum_{j=1}^{R_s} w_{ij}}\,,$$

where Qi denotes the quality index measure of the i-th frame in the video sequence, and wij is the weighting value given to the j-th sampling window in the i-th frame.

Finally in the third level, the overall quality of the entire video stream is given by

$$\mathrm{Q} = \frac{\sum_{i=1}^{F} W_i\,\mathrm{Q}_i}{\sum_{i=1}^{F} W_i}\,,$$

where F is the number of frames and Wi is the weighting value assigned to the i-th frame.

2.4.5 JND(PQR)

JND is a psychological concept as a threshold of difference that can be detected by humans. "The Sarnoff JND Vision Model is a method of predicting the perceptual ratings that human subjects will assign to a degraded color-image sequence relative to its non-degraded counterpart"[32]. Based on JND, which refers to "Just Noticeable Differences", the extent of differences between the reference video and processed video can be detected. The following figure gives the architecture for JND model.

The front end processing transform the video input signals to the separated luminance and chrominance quantities. For luminance and chrominance signals for each, the corresponding JND map luma and chro are generated based on both the reference and the processed signal via the respective processing stage as the output. For the JND maps, it gives a more comprehensive view of the video quality and severity of the artifacts, while signal JND value output works well in rating the distortions for the tested sequence from the perception of observers.

The luminance and chrominance JND maps can be reduced to a single parameter, as the PQRs (Picture Quality Ratings) for each map. The value of the parameter is assigned to 90 percent of the value [32] derived from the JND values for all pixels above a threshold. For the whole field being processed, the PQR is produces with a linear combination of sum and maximum luminance and chrominance PQR values.

JND metric has provided different kinds of criterion for objective video quality assessment. These criterions contain three necessary aspects for dynamic and complicated moving sequences: spatial, temporal and full-color analysis. The human visual system mode in JDN processing makes the results have little relationship with compressing process and its related side effects, which makes senses in multi-compressing system. Besides full measurement for system, the combination of each part of JND models provides an important index for the performance of the whole video system.

2.4.5 VQM

Video Quality Model (VQM) is proposed by the Institute for Telecommunication Science as a full reference model. Proved to have a high relationship with subjective video quality assessment [33], ANSI has made it as an objective video quality standard. VQM is computed as a linear combination of parameters including the Y luminance and Cb, Cr chrominance components for measuring the effects of blurring, distortion, noise and jerky. Additionally, VQM can be modeled on several criterions, including television, videoconferencing, general, developer and PSNR with low computational complexity. However, to use VQM, people has to pay for the patent owned by ITS and NTIA.

2.4.6 DVQ

The DVQ (Digital Video Quality) metric is an attempt to incorporate many aspects of human visual sensitivity in a simple image processing algorithm [34]. The filtering operations during the processing of video for implementation may be complex and time consuming and DVQ adopts Discrete Cosine Transform (DCT) which is available for many applications, for the decomposition into spatial channels to simply the computation. The differences between the reference and test video sequence will be converted to video quality to give the perceptual error over various dimensions. Also, the employment of DVQ involves the patent owned by NASA (the National Aeronautics and Space Administration) Ames Research Center.

2.4.7 PDM

In 1999, S.Winkler proposed a Perceptual Distortion Metric (PDM) [SW] based on a spatial and temporal model of human visual system.

The color space conversion module transform the luminance information Y and chrominance information Cr and Cb into R, G, B values using a linear function. Then the three resulting components are subjected to its perceptual decomposition respectively, including the spatial and temporal mechanisms. Through the contrast gain control, all the sensor differences are gathered to investigate the visibility of distortions and to compute the difference values in detection and pooling stage.

2.4.8 MPQM

Moving Picture Quality Metric MPQM is also a full-reference metric based on HVS model. That is to say, both the original and the processed version of video stream is required as input, even though the original may not be the well-known source. This metric has taken into account the impacts of compressing, transmission and depressing in the whole process of video communication instead of simply depending on the network risks to deduce the video quality.

MPQM model are sensitive to errors and it performs well in evaluating video quality with visible errors. Two important factors that effect the human perception are covered by MPQM model: contrast sensitivity and visual masking. Only contrast of the object exceeds the threshold, the object can be recognized from the perception of human eyes. The masking is caused by the interference of intra signals. Furthermore, the evaluation is done at pixel level with combination of signal in multiple channels.

2.4.9 PVQM

The Perceptual Video Quality Measure (PVQM)[35] as developed by KPN/Swisscom CT and it uses the same approach in measuring video quality as the Perceptual Speech Quality Measure (PSQM, ITU-T Rec. P.861[26]) for speech quality measurement. This model deals with spatial, temporal distortions, and spatio-temporally localized distortions of video. Before the measurement, a spatio-temporal-luminance alignment is taken into the algorithm to get well-matched frames of the two compared video streams.

The edginess of the luminance Y is computed by calculating the local gradient of the luminance signal (using a Sobel like spatial filtering) in each frame and then averaging the results over space and time. For the chrominance indicator, it is caculated as a weighted average over the color error of both the Cb and Cr components of the video frames. The temporal decorrelation is defined as the value of absolute difference of 1 and the correlation between the current and previous frames. In the last step the different indicators are mapped onto a single quality indicator, using a simple multiple linear regression.

Besides these metrics, there is some software to test the video quality and the patent is owned by the company. Consequently, the use of the software will consume the users a considerable sum of money and increase the cost of the product.

3 Solution

This part will give the solution for the project and it helps for the implementation and testing. First, I will give two test cases about how to test the video quality based on the platform provided by Ericsson and the first one is what I will focus on. Then more information will be described as the input, the test application and output.

3.1 Test scenarios

The test cases will be given out for video conferencing about how to test the effects of time delay, data loss for audio and video frames during the transmission. With the help of my supervisor, we defined two test cases to explore the mystery of video conferencing application. The first test case is easier to implement since there are already some test applications on video conference for the reference phone. More complex, the second test is suggested as an alternative but it is clearer or more direct to find out the effects in video conferencing service since the frames being sent are subject to control according to the demand. It is supposed that the continued solution is based on the first test case and the second test case is proposed for further research.

Test case 1

The test case (shown in figure 6) is to perform real test by making a video call with a smart phone. We will call the devices (U370) from the phone and loop it back to the phone. U370 is the EMP platform with Windows Mobile system, including the hardware and software. We can compare what we send with what we receive to see to observe the quality of video conferencing. If the voice and picture we send and receive do not have much difference from our perception, then it is supposed that the quality is good without much bad influence of time delay and data loss.

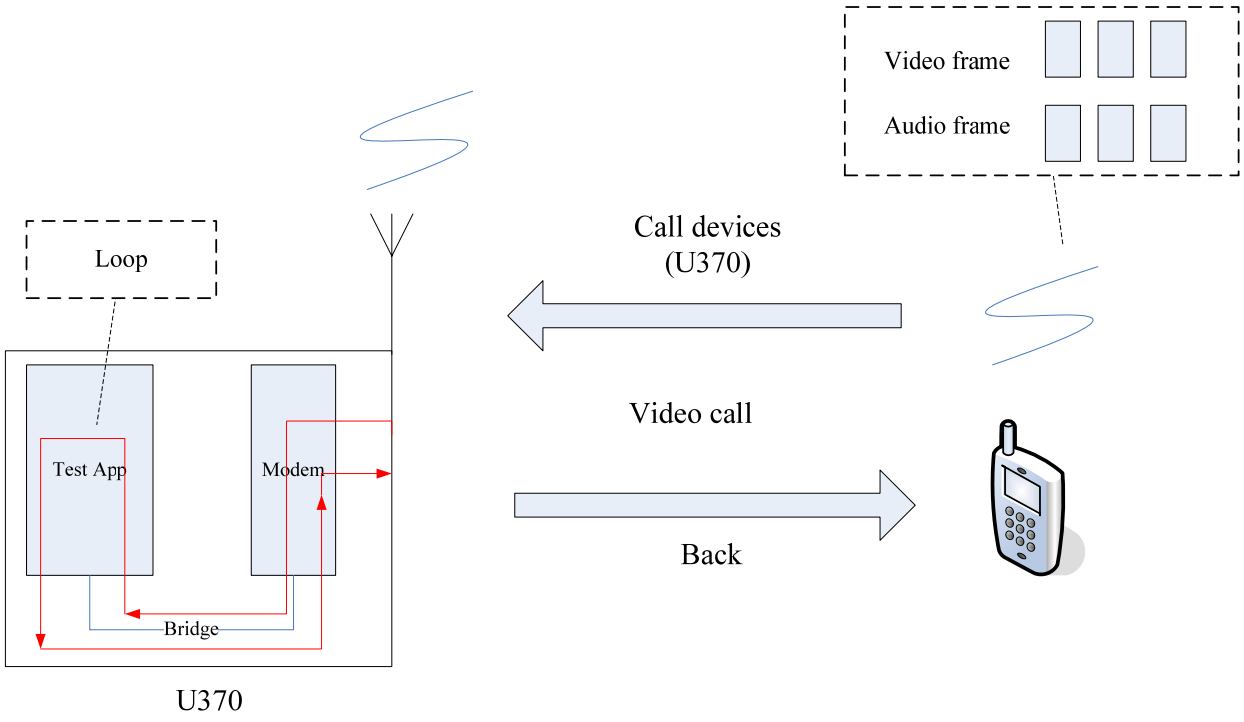

Figure 6 Test Case 1

Test case 2

This test case is more about theoretical research on effects mentioned above with all the devices and software. The figure 7 shows how they connect and works together. In this scenario, we have two U370 test platform, one is for sending the frames and the other for receiving. The frames are transmitted in WCDMA network. On the sending side, some video and audio frames can be generated with application and both are sent with certain rules. For example, when a 'black' video frame is sent a 'black' audio frame is all transmitted at the same time. On the other side, we can receive the frames through antenna from the sender and try to analyze the differences. The first effect is the time delay, and we can see that if the audio and video frames are synchronized. That is, when we get a 'black' video frame, can we also get a 'black' audio frame? If not, it means that there is some time delay during the transmission. For data loss, we can make signs of each frame we send and see whether we can receive all these frames.

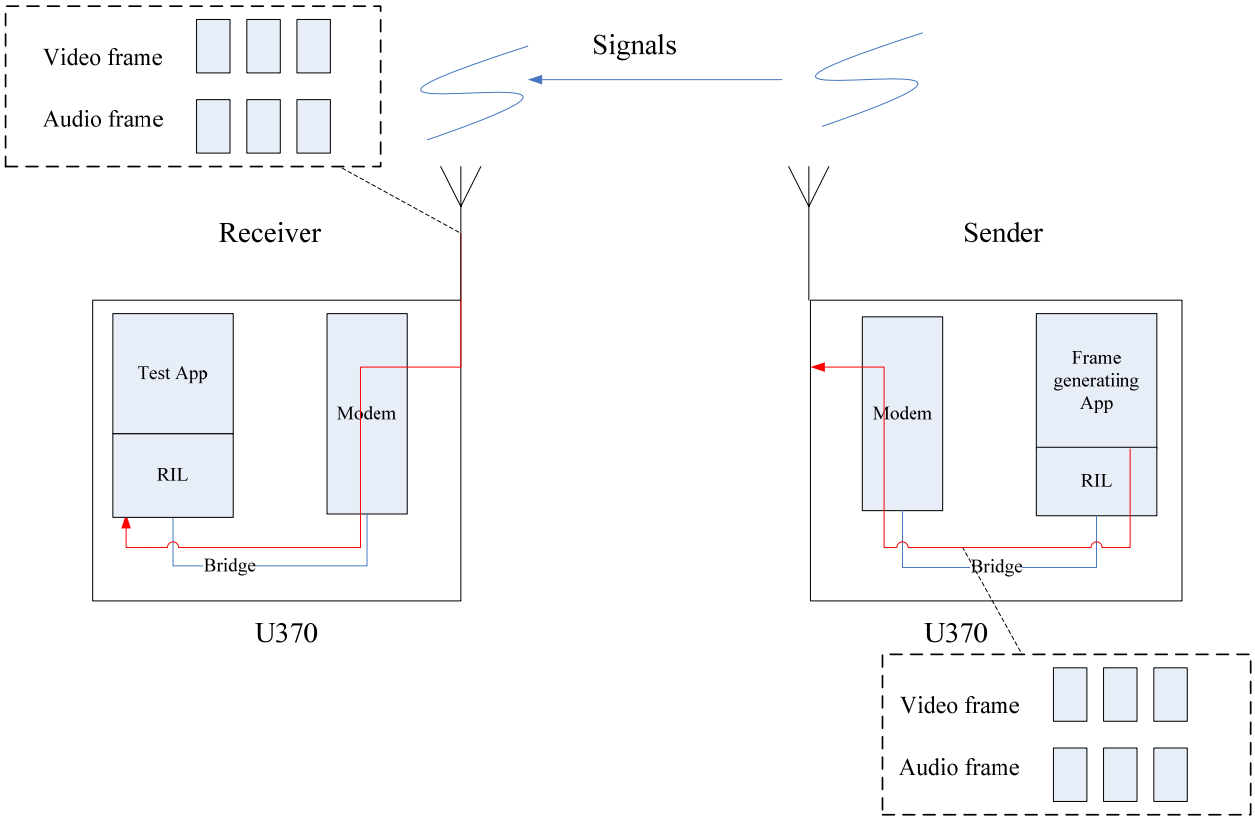

Figure 7 Test Case 2

Since the second test case is suggested as an advanced alternative, the first test case is the focus that we are going to test for the quality of video conference.

3.2 Input

As mentioned above, the full-reference objective video quality measurement is taken in our project, that is to say, there are two video input adopted, the video sent and the video received for comparison. The video format varies ranging from the RGB basic format to other kinds of coded ones, the video format in our test case will be given as below.

For the video to be sent during the video call, the video is first caught by the camera of the reference phone. This video before any transformation is encoded in a YUV format. The YUV format caught by the camera can be tuned to 4:2:0 and 4:2:2, the ratio is supposed to be 4:2:2 for the quality measured. However, the YUV video is not compressed. As a result, both the phone and the U370 equipment can not afford the transmission of large data. On the other hand, the video need large space for the storage. To solve this problem, the filter should be applied to get the compressed video which is good for transmission. H.264 video format is supposed to be used as the pre-condition of project and the purpose of this filter is to transform the YUV video format into H.264 video format. The whole process is described in the following figure 8.

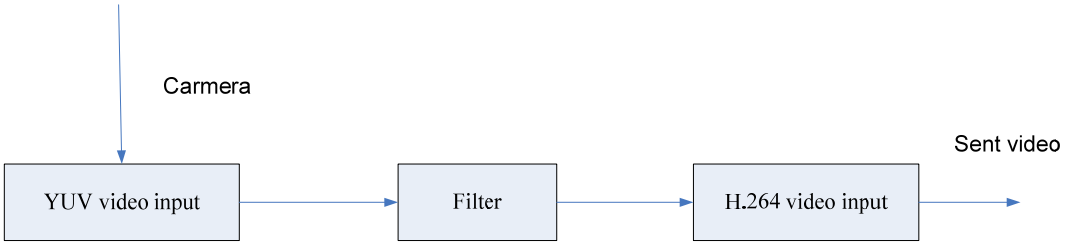

Figure 8 the Sent Video Input

Also, we need the received video (figure 9) as the input for measurement. The video we received is still coded in H.264 format since we did not do any transformation of the video in the loop back during the transmission of the video. In order to recover the raw video format, the defilter is applied to carry the transformation.

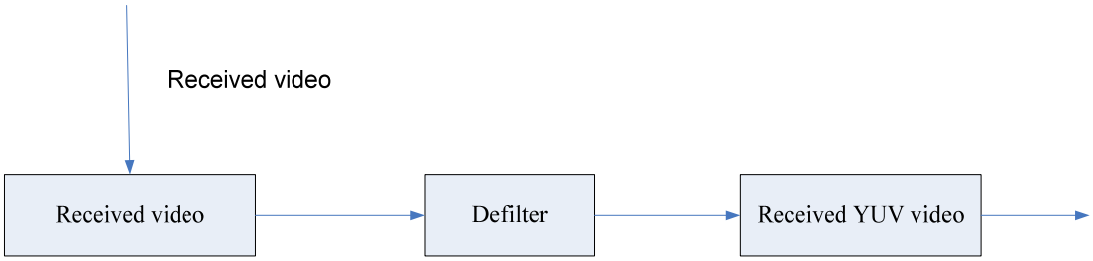

Figure 9 the Received Video Input

These two video now are obtained for the measurement.

3.3 Control test application

As specified above, the loop happens in the test application to control the processing of the video streams, including the time delay, the frame loss and other factors. It contains two aspects to process the video, the control of the video call and the processing of the video. The following figure 10 shows how this test application works.

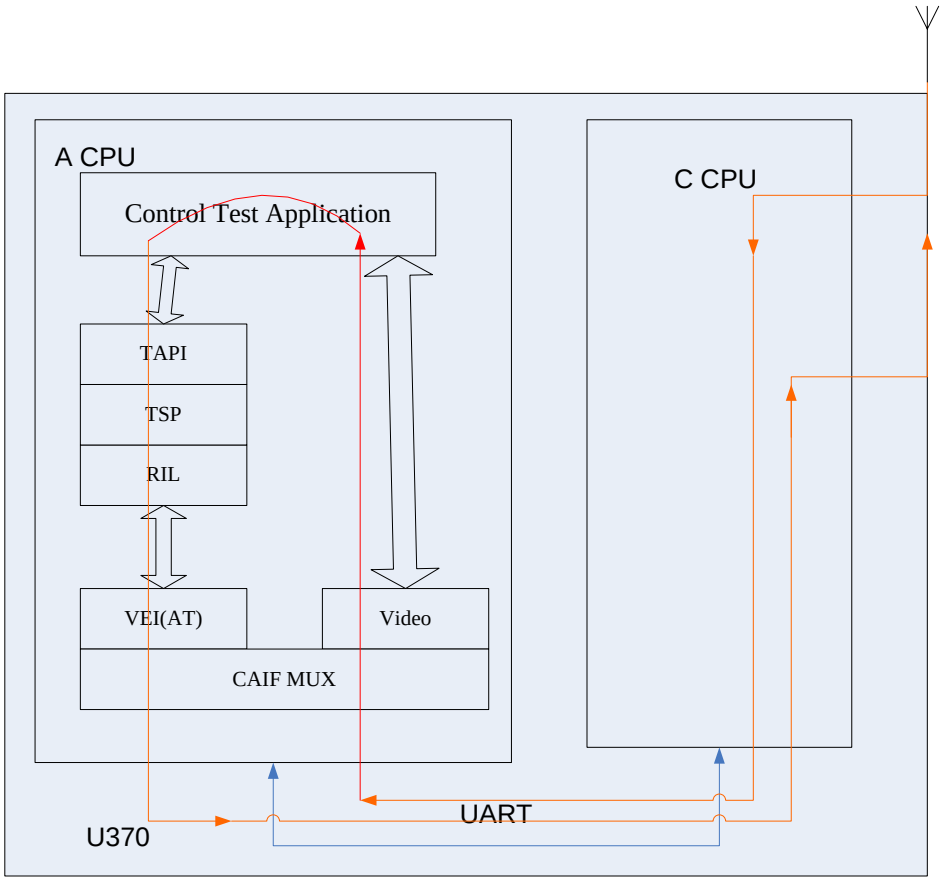

Figure 10 Control Test Application

The U370 platform is what we use in our project and it is developed by Ericsson. This platform contains two CPU, A-CPU and C-CPU, and these two CPUs are connected by the bridge which is called UART. The control test application is created in A-CPU, and it should work with the hardware through the interface. CAIF MUX is the basic interface between the software and hardware. CAIF MUX has six ports for respective purpose, and in our application, we need VEI (AT) port for control and the video port for the processing of video. CAIF MUX can not work directly with the test application, to connect with CAIF MUX, TAPI, TSP and RIL are necessary as the interface between Windows Mobile OS and the lower layer. How these interface work is specified in the State-of-the-Art. The video stream is sent out to the WCDMA network through the modem of the platform. Thus, we do not need TCP and IP protocols in this project to transmit the video stream.

For the video, the video streams are received at the video port and sent to the application for further processing. We take two ways respectively to do the video processing. One is to send out what have been received without any artificial change of the video frames. This way is just a loop back of video streams and used to test the effects of the platform itself. As a result, the control test application is only the control of the video call and the forwarding of the video frames without any transformation.

The other is to control the frames including the time delay and frame loss. First, delay the time to

prolong the transmission time of video. The time for video can be several seconds. In the test application, we set three different lengths respectively as following.
Time delay = 0.5s
Time delay = 1s
Time delay = 5s

To test the influence of frame loss, we can lose or insert certain frames at random. To lose frames, we can discard 5, 10, 30 continuous frames to see the corresponding effects of video quality on the side of reference phone. Another method is to Insert 5, 10, 30 black video frames at random time respectively.

3.4 Quality test application

Considering the two sides of the coin and our case of testing the video quality, we decided to employ the PSNR and the PVQM metrics to explore the comprehensive aspects of the video quality. The algorithm is described in section 2.4.2. The details of how PVQM works is illustrated by [35], and I will not explain it in many details here. But what I want to mention is the main steps used in PVQM.

It is known that the video should be analyzed and measured to obtain the key parameters which can indicate whether the video is good or bad quality. However before that, the video input including the raw video and tested video should be exerted the spatial-temporal and luminance-chrominance alignment for extract the feature of video to be analyzed.

To determine the quality of video, three key indicators of PVQM are calculated based on the extracted feature. There are the edginess of the luminance, the temporal decorrelation indicator, and the chrominance indicator. These three indicators can be computed at each frame level and the concrete algorithm is specified in [35] so that I will not give more explanations. One important point I want to mention is that all these three indicators are the result of the comparison of two video input, the video to be sent and the video to be received.

The video we are trying to test as the stream level is composed of frames. The question is how to test the video quality at the stream level since we got the result to measure the quality of each video frame. Recommended by [35], all of the three frame indicators use Lebesgue-7 weight to aggregate the frame indicators of their own. This is the first choice we made to test the video quality.

The alternative solution is to compare each frame indicator frame by frame. A threshold for each three aspects of indicator can be set to decide whether the quality of the corresponding frame can be accepted. If the indicator of the frame is within the threshold, it is supposed that this frame can be tolerant for the standard of the corresponding indicator (edginess of the luminance, the temporal decorrelation indicator, and the chrominance indicator). Finally, the percentage of the tolerant frame over the whole tested video can be calculated for each of the three key indicators. Since it is hard to decide the quality based on the percentage, we can assume that if the video of

80 percent for one of the three key indicators which manifest one aspect of the video quality, can be tolerant, it is said that this aspect of the video quality is good, otherwise the quality is bad. As said above, this method is not the optimal to be recommended since it is more complex and computational than the first one. But, since it compares the indicator of three aspects at each frame, the results are more precise and comprehensive.

Since the quality is affected by many aspects including the network, the platform, the video codec, the artificial control, we intended to test the video quality in several different conditions as specified below. (As mentioned in the State-of-the-art, the effects during the transmission in the network can be ignored.)

Condition 1: Test the video quality affected by the platform with the codec
As mentioned in the second section, the video can be affected by the codec. In order to test the influence, the input of the sent video and the received video is the coded H.264 format. The test application to control the video in platform U370 is just the loop back without any artificial control.

Condition 2: Test the video quality affected by the platform only without the influence of the codec
In this condition, we input the YUV formatted video to measure the quality. To test the influence of the platform only, we can measure the coded H.264

Condition 3: Test the video quality under the artificial control
For this condition, we select to open the artificial control in the control test application. Since the time delay and the frame loss can be controlled in different criterion, the video quality can be tested under each criterion.

However, the effects of the platform are also taken into account in this condition because we use the same platform. Fortunately, the influence of the platform can be ignored compared with the artificial control. But to exclude the effects of the codec, we try to input the same video format in condition 2.

3.5 Output

Specified in former section, we use PSNR and PVQM together to measure the video quality. The output is thus not single number but several parameters to indicate several respects of the video quality. PSNR is the traditional algorithm and can be calculated as a full reference metric. As described in section 2.4.3, the larger the result, the better the quality.

The other output depends on the measurement of PVQM at stream level. As we mentioned, if we take the recommended algorithms, the three indicators can be combined to a single result DMOS. As the same as PSNR, the higher the value, the worse the image quality.

As we said before, the second method is more complex and each of the three key indicators is

shown independently. They specifies the three respects of video quality, consequently, the video quality is evaluated comprehensively. We hope that this method also works well but we are not sure whether it is feasible before we perform the test on the platform.

5 Discussion

The research we done is just a beginning as there is more work can be carried out later. First, we only use the first test case to measure the video quality and the second test case as the alternative is not looked deep into. If we can carry the second test case, both sending and receiving side can be controlled and even the format of frame can fixed as the frames are generated by the test application. The advantage is that we find out the video quality from the simplest video frame to the more complex video stream.

As I discussed in section 3.4, I proposed the alternative methods of measuring the video quality by comparing frame by frame. This is a new way that has not been researched at present and further experiments and simulation should be carried out to show the performance of this method. On the other hand, not all the effects are mentioned here, like the bandwidth. The problem is that I can not cover all the respects in the test application and I did not find a good way to test the comprehensive respects. However, all these have the value of more research.

Another point is that we do not integrate all the equipment to set up the testing environment, and more practical work worth to be done. For example, the data of video streams are encapsulated before the transmission, then how to get the useful data excluded from the headers is a significant issue. Based on the practical work and programming, we can get more data to convince readers of our research. It is hoped that we can get more support from the company later to continue the work.

4 Conclusion

As we know, many researchers are working at testing the video quality and they have proposed many different methods to explore the mystery. Since the main factors which can affect the video quality have been found out, people are more interested in to what extent these factors affect the video quality and how we can control these factors.

In our project, we have singled out several conditions to test the possible factors including the codec and the platform. As the platform is developed by Ericsson, the performance of the video service can show the capability of the platform and give help for developing the platform. Furthermore, we control the video during transmission artificially to observe the corresponding consequence of video quality. This method does good to knowing more about the effects of video quality.

The research done for the project paves the road for further practice as we only propose the solution for how to test the video quality.

Reference

[1] Videoconferencing, http://en.wikipedia.org/wiki/Videoconferencing, December 20, 2007

[2] Lisa McGarthwaite, “Client-Server versus Peer-to-Peer Architecture: Comparisons for Streaming Video”, available from http://cs.winona.edu/CSConference/2005proceedings/lisa.pdf,

[3] Windows Mobile, http://www.microsoft.com/Windowsmobile/default.mspx, December 28, 2007

[4] Sonera MediaLab, “Symbian Application Development White Paper”, January 27, 2003

[5] Qualcomm Incorported, Commonalities between CDMA 2000 and WCDMA, October, 2006, http://www.qualcomm.com/common/documents/white_papers/Commonalities_CDMA2000_WCDMA_wp.pdf

[6] Circuit Switching, http://en.wikipedia.org/wiki/Circuit_switching, January 20, 2008

[7] Harte, Lawrence, Introduction to Mobile Data: Circuit Switched, Packet Switched, Mobitex, CDPD, GPRS, EVDO and Cellular Packet Data

[8] 3G TS 26.911 v3.1.0:*3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Codec(s) for Circuit Switched Multimedia Telephony Service Terminal Implementor’s Guide*, October 1999

[9] ITU-T Recommendation H.324: "Terminal for low bit rate multimedia communication"

[10] ITU-T Recommendation H.263: "Video coding for low bit rate multimedia communication"

[11] ITU-T Recommendation H.223: "Multiplexing protocol for low bit rate multimedia communication"

[12] ITU-T Recommendation H.245: "Control protocol for multimedia communication"

[13] ITU-T Recommendation G.723.1: "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s"

[14] 3GPP, http://www.3gpp.org/About/about.htm, December, 18, 2007

[15] ITU-T, http://www.itu.int/ITU-T/info/index.html, December, 18, 2007

[16] 3G-324M, http://en.wikipedia.org/wiki/3G-324M, December 18, 2007

[17] Telephony Application Programming Interface Version 3.1, http://msdn.microsoft.com/en-us/library/ms734215(VS.85).aspx, January 20, 2008

[18] TAPI Service Providers, http://msdn.microsoft.com/en-us/library/ms725513(VS.85).aspx January 22, 2008

[19] Telephony, http://msdn.microsoft.com/en-us/library/aa454282.aspx, January 28, 2008

[20] TAPI Applications, http://msdn.microsoft.com/en-us/library/ms734223(VS.85).aspx, January 28, 2008

[21] TAPI 3.1 Overview, http://msdn.microsoft.com/en-us/library/ms734214(VS.85).aspx, February 10, 2008

[22] Time delay for voice, http://www.ikn.tuwien.ac.at/ftw-a1/services.htm, February 15, 2008

[23] Lee Ellison, Video Telephony Services over 3G Networks, Nikkei Electrics Asia, December 2005, http://techon.nikkeibp.co.jp/article/HONSHI/20061122/124194/, February 15, 2008

[24] Zhou Wang, Bovik, A.C., Ligang Lu, "Why is image quality assessment so difficult?", Acoustics, Speech, and Signal Processing, 2002

[25] Olsson, S, Stroppiana, M, Baina, J, "Objective methods for assessment of video quality : state of the art" , IEEE Transactions on Broadcasting. Vol. 43, no. 4, pp. 487-495. Dec. 1997

[26] Srinath Loganathan, "Adaptation of a Perceptual Video Quality Measure to Low Bitrate Multimedia Applications", July 2005, http://www.bth.se/fou/cuppsats.nsf/all/fc50ce1b54ae3900c12570810040f530/$file/MEE05_19.pdf, February 28, 2008

[27] The Status of Objective Metrics, http://www.videoclarity.com/WPObjectiveStatus.html, March 2, 2008

[28] Neelima Singh, "Performance Analysis for Objective Methods of Video Quality Assessment", October 18, 2005

[29] Yubing Wang, "Survey of Objective Video Quality Measurement", February, 2006

[30]Zhou Wang, Ligang Lu and Alan C. Bovik, "Video Quality Assessment Based on Structural Distortion Measurement, Signal Processing, Image Communication", VOL.19, No. 2, PP. 121-132, February 2004

[31] David Chih-Che Lin, Paul M. Chou, "Objective Human Visual System Based Video Quality

Assessment Metric for Low bit-rate Video Communication System”

[32] Dr. Jeffrey Lubin, “Sarnoff JND Vision Model”, Sarnoff Corporation, August 5, 1997

[33] Stephen Wolf and Margaret H. Pinson, “Low Bandwidth Reduced Reference Video Quality Monitoring System”, Institute for Telecommunication Sciences (ITS) National Telecommunications and Information Administration (NTIA)

[34] A.B.Watson, J. Hu, and J. F. III. McGowan, “Digital Video Quality Metric Based on Human Vision,” J. Electronic Imaging, vol. 10, no. 1, 2001, pp. 20–29.

[35] Hekstra A.P.1; Beerends J.G.; Ledermann D.; de Caluwe F.E.; Kohler S.; Koenen R.H.; Rihs S.; Ehrsam M.; Schlauss D., “PVQM - A perceptual video quality measure”, Signal Processing: Image Communication, 2002

[36] ITU-T Rec. P.861, “Objective Quality Measurement of. Telephone-band Speech Codecs,”