NTNU – Trondheim
Norwegian University of
Science and Technology

Room Acoustic Conditions for Audio and Video Conferencing

Erlend Inge Gundersen

Master of Science in Electronics
Submission date: June 2012
Supervisor: Peter Svensson, IET

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

Problem description

The thesis will consider room acoustic conditions for audio and video conferencing, and try to formulate recommendations. This will be conducted by mapping the type of rooms commonly used for video conferencing, by collecting information from existing installations. From this, a representative selection spanning over the variety of room-types can be made. In this selection room acoustic measurements will be done: background noise measurements and impulse responses with appropriate measuring equipment. By processing the results, conference connections can be simulated between any of those rooms, giving a large number of room combinations. For each simulation acoustical parameters can be calculated, giving an objective rating of sound quality.

The expected result is to find relationships between the acoustical properties of two rooms (reverberation time, time/energy-ratio, noise level) and the speech quality of the video conference, described by acoustical measurements. The values of the acoustical parameters can be used to formulate acceptable measures regarding reverberation and speech clarity.

Abstract

A video conferencing situation combines the acoustical properties of two rooms. As the talker is located in a one room, the sound reaching the listener in an other room will be colored by the acoustical characteristics of both of them. This thesis aims to survey the current conditions in a selection of video conferencing rooms, by investigating several room acoustic parameters involving background noise, reverberation time, speech clarity and speech intelligibility. Convolution of recorded impulse responses enables the rooms to be combined, and the combined results to be evaluated. The evaluation of the results allow limit values for acceptable quality for video conferencing to be suggested. Limits for the highest acceptable values for the early decay time and the lowest acceptable values for speech clarity are suggested both for the single-room situation, and for the combination of rooms. The suggested values are based on specifications from building standards and relations between measured room acoustic descriptors.

Sammendrag

I en videokonferanse kombineres de akustiske egenskapene til to rom. Når taleren er plassert i ett rom, vil lyden som når lytteren i et annet rom være farget av den akustiske karakteristikken til begge rommene. Denne oppgaven har som mål å kartlegge de nåværende forholdene i videokonferanserom, ved å gjøre målinger i et utvalg av rom. Dette vil bli gjort ved å undersøke en rekke romakustiske parametre som bakgrunnsstøy, etterklangstid, taleklarhet og taleforståelighet. Ved å ta konvolusjon av målte impulsresponser kan to rom kombineres, og dermed kan egenskapene til mange ulike romkombinasjoner evalueres. Ut i fra disse resultatene er det blitt foreslått grenser for hva som er akseptable verdier for etterklangstid og taleklarhet for både enkeltrom og for kombinasjoner av rom. Forslagene er basert på spesifikasjoner fra byggestandarder, forhold mellom målte romakustiske deskriptorer og resultater av foretatte målinger.

Aknowledgements

The author wishes to thank Sigrid Louise Gundersen, his sister, for the introduction to Viju Norge AS. As a former employee at Viju, she recognized the common interest for the field of room acoustics between NTNU and Viju. Viju is a supplier of visual collaboration solutions, where video conferencing has an important place. A connection was established and the scope of the project was formed in a collaboration between supervisor Professor Peter Svensson at NTNU, Acoustician Paal Evjen at Viju and the author. Essential to the content of this thesis, Sales Director Knut Bentzen at Viju has made video conferencing rooms available for measurements at Viju's offices in Trondheim, Oslo and Stavanger, as well as rooms at other businesses with relations to Viju. The author has also been allowed to take part in a sales meeting where Viju's four office locations in Norway all participated in the same video conference. This have given a valuable insight in the use and performance of a variety of video conferencing rooms.

Many thanks to all of you, who in different ways have contributed to the making of this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The use of video conferencing has been a field of rapid growth the latest years. Video conferencing provide positive effects for a company with multiple office locations or with the need to connect to other companies. By replacing conversional meetings that require long distance travel with a video conference, it will lead to both a reduction of expenses and a gain in work efficiency by eliminating the need to travel. However, if a video conference is to be a valid alternative to a face-to-face meeting, the audio and video quality needs to be as close to a normal meeting as possible. A low resolution video-feed and muffled sparking sound may cause serious obstacles for the communication. In the worst case, by forcing the users to worry about the transmission quality, it may take the attention away from the communication of importance. The goal should be clear. To make the communication between two rooms as good as it would be if all participants were sitting around the same table. A large part of creating this illusion is by ensuring there is good sound quality and a comfortable sound environment in the room.

Rooms used for video conferencing vary in size, shape, interior, number of occupants and are meant for accurate and natural transmission of speech. It is possible to compare the desired acoustical properties with other types of rooms as meeting rooms, class rooms, listening rooms and recording rooms. A video conferencing room does in fact have to work as three rooms at the same time. Firstly, it must work as a regular meeting room for the people that are physically present. In many cases, the room is intended for use in both regular meetings and meetings with video connection, making this a important aspect. Secondly, the room has to work as a listening room. With a video connection established, the speech from the connected room is played through loudspeakers. Therefore, it is important for the room to be correctly fitted with loudspeakers and interior which is appropriate for a comfortable listening setting. Lastly, the room has to work as a recording room. The speech of the people in the room is to be captured by microphones for the transmission. Microphones are normally placed on the table, or suspended from the ceiling. Either way, the microphone captures the characteristics of the room along with the sound. As the sound is played over the loudspeaker on the listeners side, this room also affect the sound, making the listener hear the original speech along with the acoustical properties of two rooms. It is this situation, that makes video conferencing acoustics a special case.

It has long been an understanding of the acoustical conditions in single rooms, but what is the case when rooms are combined? This thesis will explore the acoustical conditions that rise when rooms are combined, and how it differs, from ordinary single room situations. The approach will be to investigate a selection of video conferencing rooms in order to survey the current conditions, as previous work on the topic is sparse. For the purpose at hand, it is important to evaluate what is going on in the speech frequencies, so the lower frequencies can mainly be disregarded. The focus will be directed towards the temporal effects of sound.

1.2 Previous work

Rooms with different purposes and uses will require focus on different acoustical properties. What parameters have an effect on the perceived quality, and what values of these are the ideal ones? For the particular case of video conferencing, there are no clear specification of parameters or values. A way to approach this will be to look to rooms with similar desired properties, such as speech intelligibility and acoustical comfort. Svensson and Nilsson have stated that optimum room acoustic comfort can be achieved by using a selection of appropriate acoustic descriptors, [1]. These are properties such as sound level, sound propagation, speech intelligibility and reverberation. As examples, the speech intelligibility and sound level should be prioritized in learning and communication environments, sound propagation should be reduced in open office landscapes, reverberation should be prioritized in rooms for live music and in hospitals the reduction of sound level is most important. Nilsson, [2], have studied the relation between room acoustic parameters in classrooms, and concludes that there is no clear relation between the reverberation time and the speech clarity. Nilsson states that his results emphasizes the importance of supplementary measures to support the reverberation time.

An important part of this thesis will be to examine the acoustical properties of combinations of two rooms. A way to do this is by convolution of room impulse responses. Svensson, Zidan and Nielsen have done a study on the properties of convolved room impulse responses , [3], which include comparison of predicted and measured results of acoustical parameters important in video conferencing situations. The methodology of convolving impulse responses will be used in the processing of measurements in this thesis.

1.3 The challenge of objective acoustical description

Everything that is heard by human ears is evaluated subjectively. Each listener has an individual perception of the acoustical properties of a sound. It is therefore an important task to find and use a set of acoustic descriptors that can assess the acoustical situation objectively. This is a challenge, being as the human hearing is one of the most intricate systems in the human body. The outer ear picks up changes in air pressure and transforms it to mechanical movement through the eardrum and the smallest bones in the human body, located in the middle ear. From here, the movement proceeds to the inner ear, where the hair cells picks up the contributions over the frequency range of the human hearing. These are in turn sent as signals through sensory nerves to the brain, and processed for the listener to understand. The chances of replicating and totally understand the effects of each step of hearing are slim. Therefore, the way to evaluate hearing subjectively can be to have a representable selection of listeners and having them subjectively evaluate presented stimuli. The results can reveal relationships between objective descriptors and the subjective perception of these. By having objective descriptors with known values for what is desirable, the room acoustic planning and design will be much simpler.

1.4 Combined acoustic quality

As previously mentioned, rooms intended for different use may have the need for focus on different descriptors and varying values for these. Some countries have building codes where acceptable and minimum requirements for reverberation time are specified. This is for types of rooms like classrooms, offices and other public areas. However, for the majority for rooms, there are no strict specifications. This leaves the task of specifying a reasonable reverberation time for the purpose of the room to the design team. This leads to a wide spread of the acoustic quality in rooms where this may be important. Naturally, the requirements will change with the intended use of a room, and for audio and video conferencing, the speech quality is paramount. As a teleconferencing situation always includes the interaction of two rooms, the need to evaluate the combined acoustical properties of both rooms rises. Furthermore, one room will more often than not, be used in combination with several different rooms, all with different acoustical properties, creating a number of acoustical situations for one single room. One room may be carefully designed for optimal speech quality, but in the combination with an acoustically terrible room, the good room will not be able to perform as intended. By doing research on existing rooms used for video conferencing, it is possible to evaluate how well the rooms perform combined with one another. From this, it may be possible to formulate recommendations for limit values of acoustical properties of rooms that makes them suitable for each other. It should be clear on what combinations that should be avoided, or what improvements to be done to the room acoustically to make it have an acceptable acoustical quality combined with most other rooms.

Ideally, in a video conferencing scenario the perception of audio and visual input should be as close to a normal meeting as possible. By having obstacles as reverberant, muffled sound and a blurry image, chances are, the conferencing system might not be as attractive to use. By making the experience as natural and close to reality as possible, the content of the meeting will be in focus, giving no reason to be distracted by the technology.

1.5 Selection of video conferencing rooms

The research in this thesis is based on the measurements done in eleven video conferencing rooms. The measurements have been done in a selection of rooms in three different cities in Norway; Trondheim, Oslo and Stavanger. This have been possible through the collaboration of Viju Norge AS. Viju has made rooms available at their office locations in Trondheim, Oslo and in their headquarters in Stavanger. Further arrangements were made by Viju allowing the use of rooms at Aker Solutions in both Oslo and Stavanger, and Progressus Management in Stavanger. Additionally, measurements were done in rooms at NTNU and SINTEF in Trondheim.

The rooms are meant to span over a variety of properties, in order to cover a representable selection of video conferencing rooms. The choice of rooms were done from the following criteria:

- Size, small to large
- Perceived acoustical conditions, poor-excellent
- Interior, properties of floor, walls and ceiling

1.6 Structure

In this thesis, a set of measurements are carefully chosen to capture the room acoustic conditions of rooms used as both sender and receiver in a video conferencing situation. This includes multiple recording points of the room impulse response and the use of both an omni-directional microphone and microphones in a dummy head. From the measured impulse responses, several acoustical parameters can be derived. Results of the acoustical conditions in both single rooms and combinations of rooms will be presented. In chapter 2, there is given an overview of the theory which is used throughout the thesis for obtaining the desired results. Chapter 3 explain the measurement procedure and how the measurements are processed. The results will be given in chapter 4. This includes a description of all the rooms in which there are performed measurements, together with the acoustical properties, floor plans and pictures for each room. Furthermore, results will be presented for the individual rooms and for the combinations of rooms. Comparisons will be made between the two situations and suggestions for recommended limits of acoustical parameters will be made both for the single-room and combined-room situation.

Chapter 2

Theory

2.1 Room impulse response

The impulse response is defined as the temporal evolution of the sound pressure observed at a point in the room as a result of the emission of a Dirac impulse at another point in the room. A room impulse response (IR) can thus be viewed as the representation of an acoustic space, like an acoustic fingerprint. By convolving the impulse response of a room with a dry[1] audio recording, the resulting output will be how the recording would sound in that room. The impulse response contains the information of how the room reflects the sound. That is, both the early reflections from big surfaces and the late reverberations of the room. Figure 2.1 shows an explanation of the different parts of an impulse response.

For the purposes in this thesis, it is desirable to investigate the acoustic behavior of two rooms combined, as the case will be with a speaker in one room and the listener in the other. This is possible by convolving the impulse response of the two rooms. This combines the acoustical properties of the two rooms, making it possible to evaluate the combined acoustic behavior.

Figure 2.2 illustrates how the IR of the sender room and the IR of the receiving room are convolved into one combined response for the both of them.

[1] A recording with little or no reverberation, could be an anechoic recording

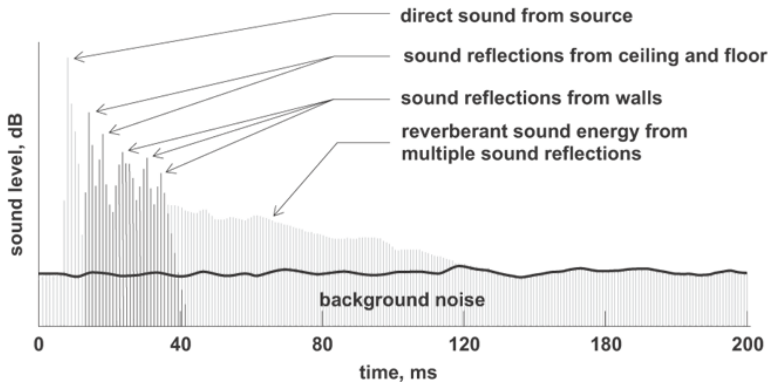


Figure 2.1: Explanation of the parts of an impulse response. Figure from [4]

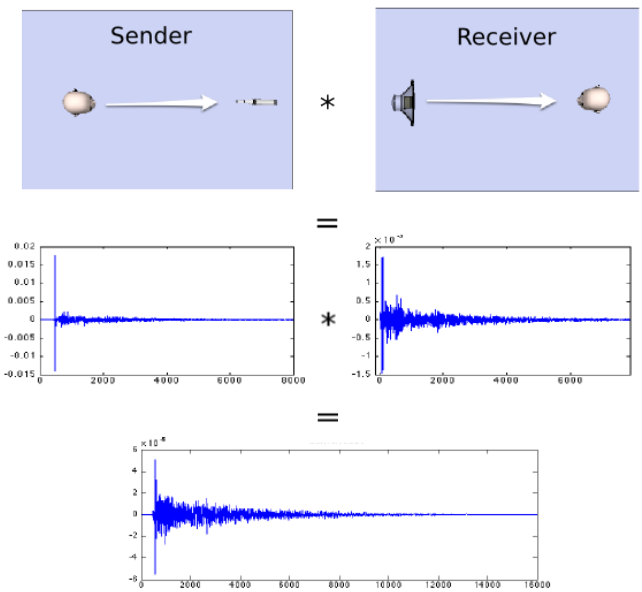


Figure 2.2: Convolution of the impulse responses of two rooms.

2.2 Subjective aspects and measured quantities

The ISO 3382-1 standard, [5], states that subjective studies of acoustical characteristics have shown that several quantities obtained from measured impulse response are correlated with particular subjective aspects regarding the acoustics. The stan-

dard emphasizes the acoustical character of auditoria and performance spaces, but several of the acoustic quantities are directly applicable for situations concerning audio and video conferencing.
There are five types of quantities, of which three relates to rooms where speech quality is most important. These three are listed in Table 2.1. The table lists which subjective listener aspect relates to which acoustic quantity. The value of the just noticeable difference (JND) is listed for each of them. The JND of speech clarity is studied by Bradley [6], and have closely related results to the values in the ISO-standard. The other two types are more important for larger performance spaces, and less important for the speech quality. This is the subjective aspects of apparent source width (ASW) and the listener envelopment (LEV). All these quantities are found to be subjectively important and can be obtained directly from post-processing of impulse responses.

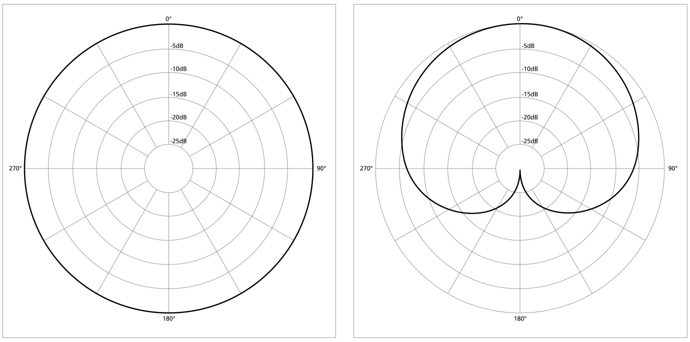
Table 2.1: Acoustic qualities according to listener aspects, partly data from ISO 3382-1

Subjective listener aspect	**Acoustic quantity**	**JND**
Subjective level of sound	Sound strength, G, in dB	1 dB
Perceived reverberance	Early decay time, EDT, in sec.	Rel. 5%
Perceived clarity of sound	Speech clarity, C_{50}, in dB	1 dB

The early decay time and speech clarity will be much used in the analysis of the rooms, and will be covered separately later in the theory section.

2.3 Microphone directivity

Microphones can have different characteristics of directivity. The most basic is the omnidirectional microphone, which records sound equally in all directions. As for directive microphones, a cardioid directivity is most common. It is named after its heart-like directivity pattern. This is a type of microphone commonly used for vocal and speech, as it rejects sound from other directions. A microphone with a cardioid characteristic has a directivity factor of Q = 3, in comparison with an omnidirectional microphone have a factor of Q = 1. Thus, the result of using a cardioid microphone, is an increase of 4.8 dB in the direct sound. The directivity patterns are shown in Figure 2.3(a) and 2.3(b).



(a) Omnidirectional (b) Cardioid

Figure 2.3: Polar plots of microphone directivity. Figures from [7]

2.4 Acoustical treatment

In a room with a lot of hard surfaces and little soft interior and furniture, the acoustics can be perceived to sound industrial-like. This differs a lot from the natural and comfortable environment that is desirable for good communication. Cisco has presented some guidelines for video conferencing acoustics [8]. The obvious solution to this is to introduce sound absorbers. Soft penetrable surfaces will absorb sound and so reduce the reverberation in the room. Heavy drapes, carpets, upholstered chairs and sofas are examples of objects with such properties. For hard uncovered walls the solution may be acoustic panels. Many variations of acoustic panels covered with fabric, wood or even metal for wall and ceiling mounting are commercially available, and designed for effective sound absorption. Some are even specially designed for rooms intended for tele- and video conferencing.

Many rooms have ceilings covered with sound absorption. This is usually in the form of an acoustical drop ceiling or absorbing plates. If there is no treatment in the ceiling, this will be a good place to start, as large areas easily can be covered. Further, having opposing walls of naked reflective surfaces is not desirable. This may cause standing waves creating flutter echo as the sound is bounced back and forth without being attenuated. Flutter echo is caused by successive, repetitive reflections, equally spaced in time, which can produce a reduction in the speech intelligibility within the room. So, if it is not possible to apply absorption on all walls, it should be placed on adjacent walls. Additionally, by spreading out the absorption it will have a better effect than clusters of absorption do.

It is strongly recommended to avoid bare concrete surfaces. Sound will almost be totally reflected, which leads to a drastically increase of the reverberation time. On the other hand, concrete works very well as a sound isolator to adjacent rooms, as it allows very little sound through. So, with the concrete surface covered, it can contribute to a reduced noise level from outside sources and an increased privacy in a room.

2.5 The room acoustic parameters

The acoustic parameters, except the background noise, will be derived from the integrated impulse response method, described in section 3.3 of the ISO 354:2003 standard, [9]. This is a method of obtaining decay curves by reverse-time integrating squared impulse responses.

2.5.1 Reverberation time (EDT, T_{20}, T_{30}):

Reverberation time is the time it takes for the sound level to drop 60 dB after a sound source is stopped. The measure is as a value in seconds. Short reverberation time suggests there is high acoustical absorption in the room. The NS-8175 standard states that regular furnished rooms will often have a reverberation time of 0.5 seconds, [10]. The measurement of reverberation time is done for several reasons. There is a strong dependance between the reverberation time and the sound pressure level from noise sources, the intelligibility of speech and the perception of privacy in a room.

Two types of reverberation time can be measured, Early Decay Time (EDT) and Reverberation Time (RT) as T_{20} or T_{30}. Both the EDT and either T_{20} or T_{30} should be calculated. The ISO 3382 standard, [11], points out that the T_{20} measurement is preferred over the T_{30}. The signal-to-noise ratio can often be a problem, as it may be difficult or impossible to achieve an evaluation range of more than 20 dB. The EDT is a reverberation time derived from the initial 10 dB of decay and is subjectively more important and related to perceived reverberation, while T_{20} and T_{30} is related to the physical properties of the room. This is because the first arrivals are psychoacoustically more important than the later reverberation range. The reverberation time T_{20} or T_{30} give information on the diffuse sound decay and are derived from the decay section of the impulse response between -5 dB and -25 dB or -35 dB below the initial level. This is illustrated in Figure 2.4. All these three measures of the decay curve can be used to find the time it takes for the sound to decrease by 60 dB, the RT60. EDT must be multiplied with a factor of 6, T_{20} with a factor of 3 and T_{30} with a factor of 2, to be comparable as RT60. Thus, if the reverberation curve is straight, the EDT, T_{20} and T_{30} will all produce the same value. This is not usually the case. Figure 2.5 illustrates how the decay curves are obtained from the impulse response using the integrated impulse response method.

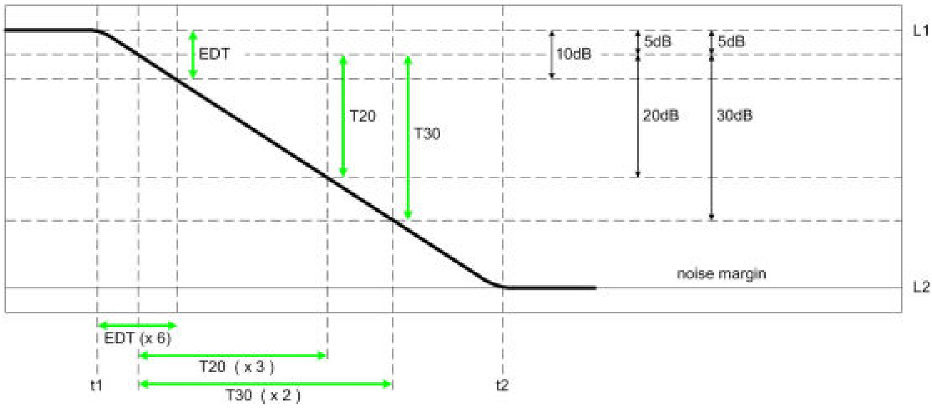


Figure 2.4: Overview of the different methods to calculate the reverberation time, from the decay curve. Figure from [12]

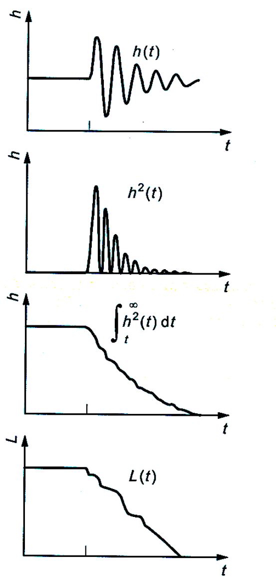


Figure 2.5: How the decay-curve is obtained from the impulse response. Figure from ISO 18233:2006, [13]

The ISO 3382-2 standard, [11], specifies methods for measurement of reverberation time in ordinary rooms. Measurements will be done according to the ISO standard and the integrated impulse response method will be used. The standard defines three different degrees of precision in which the measurements can be done. The three types are *survey*, *engineering* and *precision*. These methods are appropriate for different kind of use, and has individual requirements regarding number of source-microphone combinations and frequency resolution of the results. The engineering method is appropriate for verification of building performance with specifications of reverberation time or room absorption. The requirement is to perform measurements of the reverberation time for at least two different source-positions and at least six independent source-microphone combinations. This results in an assumed nominal accuracy better than 5 % in octave bands and better than 10 % in one-third-octave-bands. The engineering method will therefore be used in this research.

2.5.2 Speech clarity (C_{50}):

The early-to-late arriving sound energy ratio is used as a measure of speech clarity. Clarity is a measure of the ratio between early sound reflections that support speech and late reflections that interfere with speech. C_{50} is a measure of the degree to which the individual sounds stands apart from one another. For an interval of 50 ms, the balance between early- and late-arriving energy is calculated using the following equation:

$$C_{50} = 10log\frac{\int_0^{50ms} p^2(t)dt}{\int_{50ms}^{\infty} p^2(t)dt}\ dB \tag{2.1}$$

Where
p(t) is the instantaneous sound pressure of the impulse response.

For speech, the clarity will be measured using the ratio of the energy in the first 50 msec and the energy after the first 50 msec. In a very dry room with little reverberation, the speech will be very clear and the C_{50} will have a large positive value. On the other hand, in a room with much reverberation, the speech will be unclear and the C_{50} will have a relatively large negative value. For clarity in music, the first 80 msec is defined as the early sound.

2.5.3 Speech transmission index(STI):

In addition to the speech clarity parameter, the Speech Transmission Index (STI) can be used to give an objective rating of speech intelligibility. It is a measure to predict the intelligibility of speech transmitted from talker to listener by a transmission channel. In particular, the effect of changes in the acoustic properties of spaces can be assessed. The use of STI has gained international acceptance, as it has been proven useful in many situations. The STI method was introduced by Steeneken and Houtgast in the 1970s [14], and has been developed and refined since then. All current aspects of the speech transmission index is covered in the IEC 60268-16 standard, [15]. The standard states that applications of the STI includes measurement of potential speech intelligibility and communication in rooms and auditoria.

The STI give a value between 0 - 1, where 0 is not understandable and 1 is perfect intelligibility. This parameter was originally measured with the use of special modulated signals, but can also be derived by post-processing a measured room impulse response. This is called the indirect method, which should only be used for linear systems. Certain requirements must be met in order to compute the STI from the impulse response, and the background noise is usually corrected for with a separate recording. The background noise is then recorded in octave bands and included in the post-processing of the impulse response in order to calculate the STI- value of the room.

Figure 2.6: Speech transmission index

There are several limitations of the STI method, and in general the STI is a conservative approach that may underestimate intelligibility in some applications.

2.5.4 Background noise

The steady state sound pressure level in a room is what that is referred to as the background noise. The sound pressure level is measured over an interval of time and averaged, giving a single value, L_{eq}, the equivalent continuous level. Mathematically the L_{eq} is defined as:

$$L_{eq} = 10log\frac{1}{T}\int_0^T \frac{p^2(t)}{p_0^2}dt\ dBA \tag{2.2}$$

Where
p(t) is the instantaneous sound pressure of the impulse response

p_0 is zero reference sound pressure
T is the time interval for averaging

The equivalent level is given in dBA, an A-weighted decibel level. The weighting is applied because the human ear does not perceive sound levels equally over all frequencies. The A-weighting curve is shown on Figure 2.7.

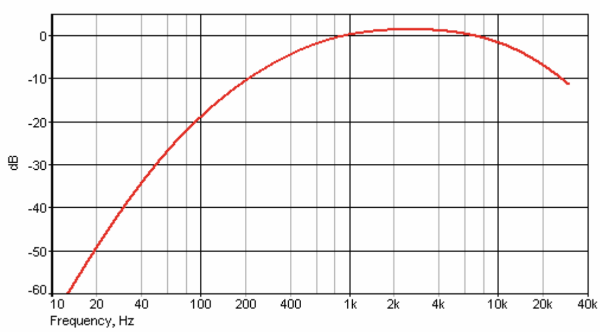


Figure 2.7: A-weighting curve. Figure from [16]

The NS-8175 standard, [10], has specifications for several types of noise measurements for different room types. Table 2.2 and 2.3 shows limit values that can be applicable for video conferencing rooms.

Table 2.2: Sound classes in buildings for educational purposes. Highest limit for indoors A-weighted equivalent sound pressure level, $L_{p,AeqT}$. (Part of table 13 in NS8175, [10]).

	Measurement	**Class A**	**Class B**	**Class C**	**Class D**
In classrooms/ meeting rooms from outside sources	$L_{p,AeqT}[dBA]$	25	28	32	35

Table 2.3: Sound classes for offices. Highest limit for indoors A-weighted equivalent sound pressure level, $L_{p,AeqT}$. (Part of table 38 in NS8175, [10]).

	Measurement	**Class A**	**Class B**	**Class C**	**Class D**
In offices, from outside sources	$L_{p,AeqT}[dBA]$	30	35	40	45

2.6 Objective parameters and perceived quality

In a perfect world there would be a direct relationship between the values of objective parameters and the perceived quality. Unfortunately, there is still a lot of how the parameters affect the quality that is left unknown. This is very much the case with the room acoustic parameters. There is little research and firm knowledge about how the quality is related to the reverberation time, and even less about quality related to speech clarity. However, arguments can be made in order to draw approximated curves of how the quality relates to these parameters.

For some types of rooms, the reverberation time is specified with values in standards, as in the norwegian standard NS-8175 [10]. The reverberation time is specified for different sound classes, from A to D. Class A is the preferred value which will give especially good sound conditions, and class C is the highest value that is still within the quality requirements of national building codes. Table 2.4 and 2.5 shows the values of each class for classrooms and meeting rooms for educational purposes and for offices and meeting rooms.

Table 2.4: Sound classes for buildings for educational purposes. Highest limit for reverberation time, T. (Part of table 11 in NS8175, [10]).

	Class A T in [s]	**Class B** T in [s]	**Class C** T in [s]	**Class D** T in [s]
Classrooms, meeting rooms	0.3	0.5	0.6	0.6

Table 2.5: Sound classes for offices. Highest limit for reverberation time, T. (Part of table 36 in NS8175, [10]).

	Class A T in [s]	**Class B** T in [s]	**Class C** T in [s]	**Class D** T in [s]
Offices, meeting rooms	0.6	0.6	0.8	0.9

There are comments to these tables saying that the limits are valid for regular middle sized rooms, and the reverberation times *should not* be significantly lower than the given values. This is because a very short reverberation time can cause the sound to have a low signal level as the room does not contribute to increase the sound level. The room will be very dry, giving little feedback to the speaker and it will create an unnatural, and perhaps uncomfortable sound environment. The values in the tables state the highest limit for the reverberation time to ensure quality, saying that a longer reverberation time will make the quality and intelligibility of speech decrease. It is quite clear, that a longer reverberation time will make reflected sound drown the direct sound, making it more difficult to understand speech content. From these arguments, there can be drawn a curve displaying how the quality relates to the reverberation time, the curve is shown in Figure 2.8.

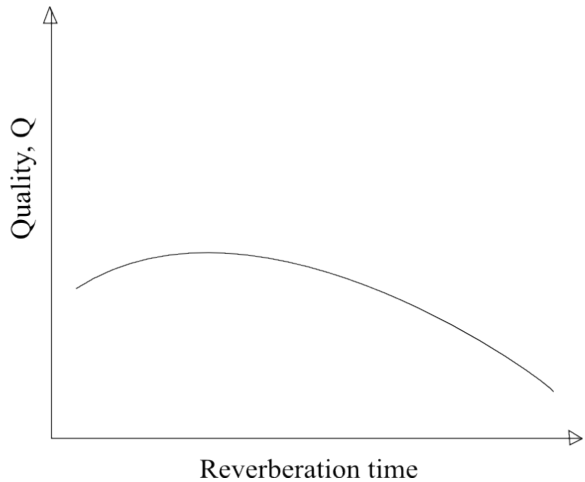


Figure 2.8: How quality is believed to vary with reverberation time

For speech clarity, the quality is known to increase with a higher value of C_{50}. This value depends on the early-to-late energy ratio of the impulse response. As the information content in speech will be in the early energy part, a high C_{50} value will give better intelligibility and better quality of the speech signal. Beyond this, there is not much precise knowledge on exactly how the relation is between the speech clarity and perceived quality. A suggestion on how the relationship may look, is shown in Figure 2.9. The suggested curve shows a slope that flattens out in the beginning and the end. It can be assumed that an initial increase from a very low value to a slightly higher value of the C50 may not affect the quality much. In the other end, the quality may not keep on increasing with higher C50, but quite possibly slope off. The quality may even decrease a little, although there is no certainty of this.

2.7 Hypothesis

This thesis aims to suggest values for objective acoustical descriptors which will give acceptable sound conditions for video conferencing connections. This will be done by surveying the current conditions in a selections of rooms. The result will

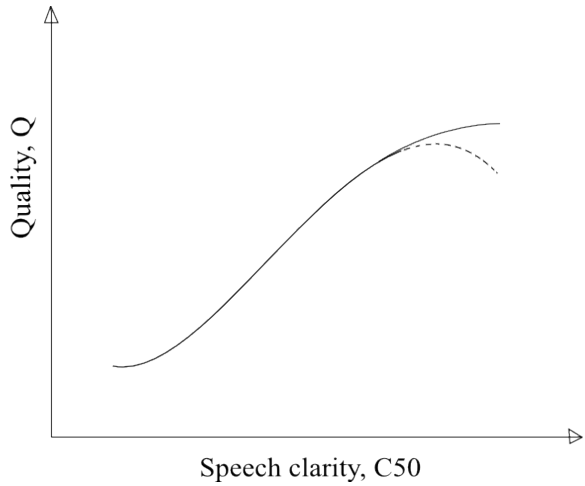


Figure 2.9: How quality is believed to vary with speech clarity

allow rooms to be divided into different categories from A to D according to their acoustic quality. The categories are based on values of early decay time and speech clarity. Category A will have the best conditions, that is, the lowest value of EDT and the highest value for C_{50}. Category B will have good sound conditions and category C will be the limit for what will be considered as acceptable. Category D will have levels where many will experience the sound quality to be uncomfortable, that communication is disturbed, and that the room is tiring to spend much time in. The categories will look somewhat like the example plot in Figure 2.10. This is just a dummy plot for explanatory purposes with randomly generated values.

A hypothesis can be stated: All rooms considered to have acceptable acoustical values, will in combination with an other acceptable room, create a combination that is considered acceptable.

Figure 2.11 shows an example plot with generated values of how the convolved values may differ from the single-room situation. The grey areas indicates the different intervals of acceptable values for the two situations.

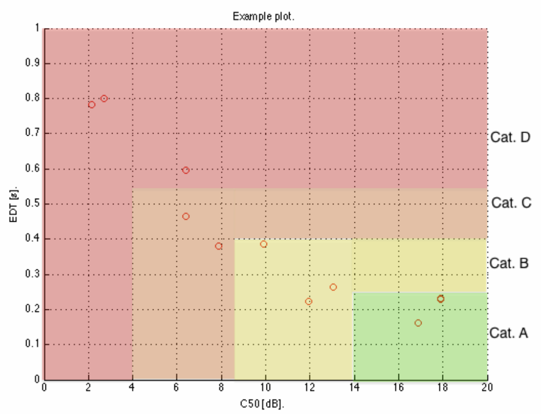


Figure 2.10: Example plot of how EDT and C50 values can be divided into categories. The values are randomly generated.

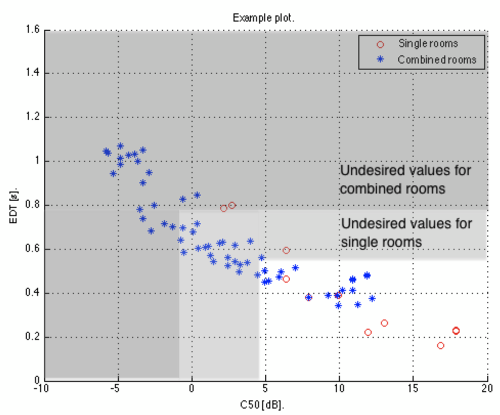


Figure 2.11: Example plot of how acceptable values can be chosen for single rooms and combined rooms. The values are randomly generated.

Chapter 3

Measurement

The same procedure for measurement of acoustic parameters will be done for all the chosen rooms. First, there will be performed one separate measurement of the background noise. The measurement is done with a Norsonic 116 sound level meter, by averaging the background noise in the room over an interval of 1 minute. Secondly, a set of impulse responses will be captured, which is the main part of the measurements. These will be post-processed, and can give several important room acoustic parameters. A specific set of impulse responses are to be measured, three types in total. The transmitting situation, IRA, is meant to simulate the response from when the speaker is present in the room, and the speech is captured by the microphone in the video conferencing setup. The receiving situation, IRB, will simulate the response when the listener is present in the room and the sound is played through the loudspeakers. To simulate a realistic situation, there will be used different microphones in these two cases. The transmission will be measured using a loudspeaker and a 1/2" measurement microphone placed on the table. In the receiving situation the same loudspeaker will be used and there will be used a dummy head with microphones in both ears, representing a human listener.

Both the IRA and IRB will be measured at various positions according to regular seating arrangements in the room. In addition to these two measurements, a third set measurement of the impulse response, IRC, will measure at more points in the room in order to derive the reverberation time with a good degree of precision, see section 2.5.1. This series of room acoustic measurements are performed in order to obtain the acoustic conditions of a representable selection of rooms used for video conferencing. The selection of rooms are made available through the cooperation of Viju Norge AS. It is desirable to perform this measurements in a wide specter of rooms. In the planning stage, it was decided that a selection of ten rooms, from from small to large room volume will be ideal. This was based on the variety of rooms in Viju's standardized VideoWorkplace solutions. The VideoWorkplace solutions are available in a small, medium and a large size with capacity of respectively 1-2, 1-6 or 1-10 persons. By having rooms of different sizes, with different reverberation time and acoustic conditions, it is possible to cover a many realistic video conferencing situations.

The use of the same measurement equipment in all rooms will ensure similar conditions, as the focus is on the properties of the room, not the video conferencing equipment. If the video conferencing setup is to be used in measurements, it may pose a challenge. As the existing system in the rooms certainly will differ, it can turn out to compromise the measurements.The number and positioning of measuring points for the IR are determined from the number of people normally present and their seating arrangement. Three speaker positions and two listener positions are chosen.

The three types of IR measurement:

- Impulse response A (IRA) : The "Speaker-to-microphone" response. This is performed with the microphone in one position, lying on the table. The loudspeaker is positioned at three different points around the table, to include different seating arrangements. Setup illustrated in Figure 3.1
- Impulse response B (IRB): The "Loudspeaker-to-listener" response. This is performed with the loudspeaker in one position, in front of the screen. The dummy head is placed at two different positions at the table. Setup illustrated in Figure 3.2
- Impulse response C (IRC): This is measurements solely done in order to get sufficient measurements for calculation of reverberation time, according to the ISO 3382-2 standard [11]. The loudspeaker is placed in one position. The microphone is placed at three different positions, intended to cover much of the space of the room. Setup illustrated in Figure 3.3

To ensure similarity in the measurement procedure, there was made a form describing the steps of performing measurements and evaluating the rooms. This form can be seen in the appendix.

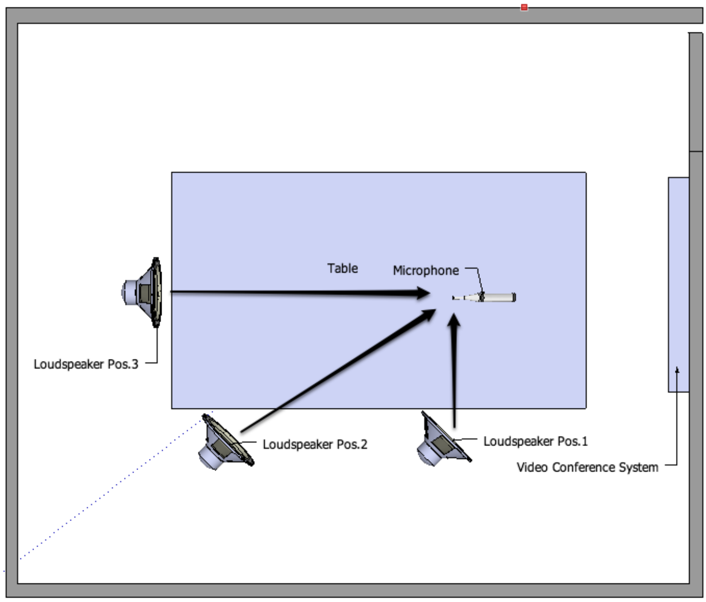


Figure 3.1: Sketch of an IRA measurement setup

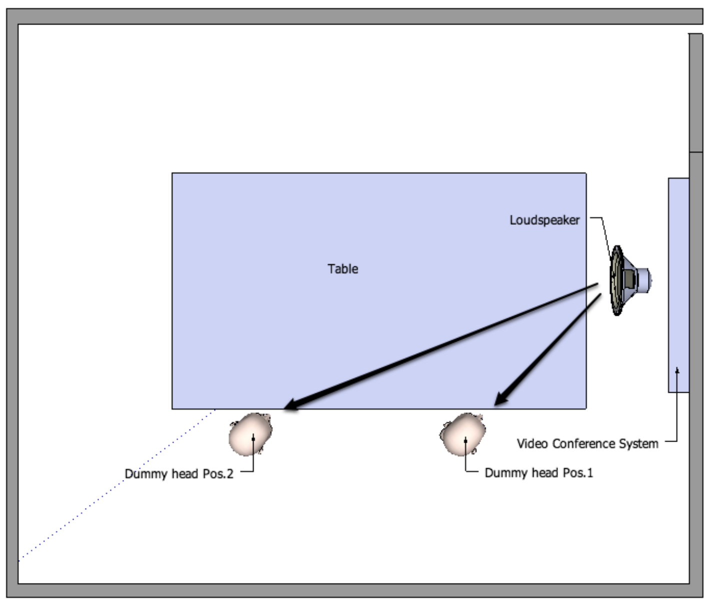


Figure 3.2: Sketch of an IRB measurement setup

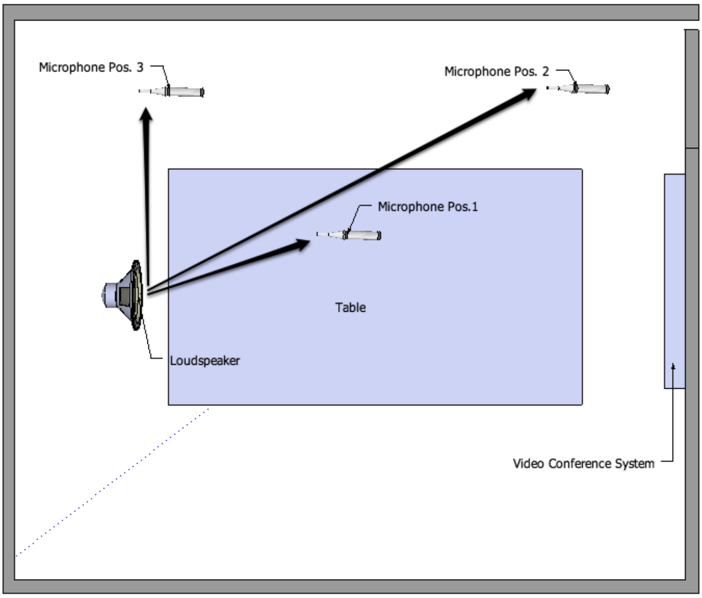


Figure 3.3: Sketch of IRC measurement setup

3.1 Acoustic parameters

Reverberation time (T_{20}), early decay time (EDT), speech clarity (C_{50}) and speech transmission index (STI), can all in different ways be obtained from post-processing of impulse response measurements. Some of the parameters require special setup for the IR measurements. The reverberation times T_{20} and EDT needs measurements of six different microphone-loudspeaker positions in order to achieve the desired degree of precision. The most accurate measurement of the speech transmission index requires that the noise in all octave bands are accounted for, in order to produce the most correct value of speech intelligibility. The STI values are calculated by using the impulse response through the indirect method of measuring STI, described in the IEC 60268-16 standard, [15]. The calculation of the STI in this thesis will be done without correction for background noise and speech level influence, only by considering the reflections and decay.

3.2 Recordings with dummy head

Recordings of the impulse response using the dummy head are done with the intension of representing the ears of a listener. The dummy head has one microphone in each ear in order to capture a bi-neural recording. With the shape of the head and two recording channels, the recording contains a simple "head related transfer function" (HRTF). The HRTF is like a frequency filter, which dampens and attenuates different parts of the frequency spectrum. Every person has a different HRTF, which is the result of the shape of the head, hair and shoulders.

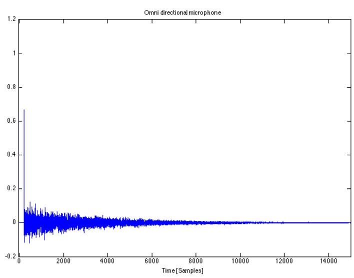
Using the dummy head for recording impulse responses gives the possibility to produce a bi-neural auralization of a room (or a combination of two rooms). By convolving an anechoic speech signal with the recorded IR of a room, the speech recording will get the reflection and reverberation properties of the room. Listening to it will give the impression that the speech signal is performed in that room. In this way, the same speech signal can be convolved with several different impulse responses and give the impression of being recorded everywhere from outside in a forest to inside a large empty warehouse.

The dummy head is important for capturing the direction of the sound. The recordings of the right and left channels of the dummy head are convolved individually and presented to the right and left ear of the listener. So, in addition to perceived spaciousness, the auralization allows directional localization of the sound source.

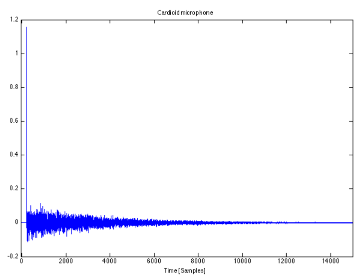
The dummy head was tested to see if there were a difference between the left and right microphone. A higher level was recorded in the right ear. This must be compensated for, should the recordings be used for auralizations. Auralizations will not be a part of this thesis, but the results of the test on level differences of the dummy head can be found in the appendix, section B.1.

3.3 Microphone directivity

The measurements are performed with the use of an omni-directional microphone. The microphone commonly used for recording of speech in a video conferencing situation is a cardioid microphone. It is therefore interesting to see how the results of these two microphone types may differ. This is done by evaluating computer-generated impulse responses in Matlab with different directivity factors. The impulse responses are made with an ideally exponential decay and specified with values of 0.5 s for the reverberation time, room volume of 55 m^3 and source-receiver distance of 1.5 m. Aside from the obvious difference in the direct sound, the two IRs in Figure 3.4(a) and 3.4(b) have very similar characteristics. The contribution of the direct sound does not seem to affect the results of the calculation of the parameters much. The speech clarity, which can be believed to benefit from a directive microphone characteristic, shows a difference less than what is regarded as the just noticeable difference.



(a) Omnidirectional



(b) Cardioid

Figure 3.4: Impulse response of different microphone directivity

3.4 Processing of measurements

After performing the measurements, there is a considerable amount of processing and computation that has to be done. As much of the results will be based on the acoustical parameters of convolved impulse responses, it is appropriate and most efficient to perform the convolution and the calculation of the acoustical parameters in the numerical computing environment; Matlab. The measurements are originally captured using the PC software WinMLS. To ensure the same conditions as in WinMLS, the filter coefficients for the octave band filtering used by WinMLS was exported to be used for the Matlab computation.

The results obtained from the computations in Matlab has been compared to the results obtained from the same computations in WinMLS. The speech clarity, C_{50}, and the early decay time, EDT, shows a very satisfying correlation between the two methods of computation. Therefore, the processing of these results in Matlab will give accurate results. The reverberation time, T_{20}, seems to have a somewhat poorer match for some of the rooms, especially in the lower frequencies.

The overview shows that the results are very similar from the 500 Hz octave band and higher. Below this level the results are spreading. The fact that the 63 and 125 Hz octave bands does not deliver valuable data is no big concern. It is from the 250 Hz octave band and upwards the most important frequencies for speech are found. The cause of the somewhat spreading results may partly come from low signal-to-noise ratio, which leads to questionable data in the WinMLS computation. An other important factor for deviation between the two, is the built-in automatic truncation of the impulse response. A window of this option from WinMLS is included in Figure 3.5. The automatic truncation option detects the relevant part of the impulse response, cropping off the redundant information. The part of the impulse response that have good information is from the top at the start, including the decaying slope down to the noise floor. In Matlab, the IR is truncated manually based on the longest reverberation time in the room. This difference in IR truncation seems to cause some differences in the calculated values for T_{20}.

In the following plots, Figure 3.6 and 3.7, the comparison of the results from computation of EDT and C_{50} in Matlab and WinMLS is presented. It shows the results obtained from the impulse response alone and when it is convolved with itself. The convolution is done in Matlab, and the computation of the acoustical parameters are done both in WinMLS and Matlab for comparison. This is firstly done for the sake of comparison of results from Matlab and WinMLS. Secondly, it is done to show the effect a convolution of two rooms has. It can clearly be seen that the room alone has got a significantly lower reverberation time at all frequencies. A similar plot is done for the speech clarity, showing that the Matlab computations give the same results as WinMLS.

In the end, the very high correlation between the EDT and C_{50} values indicates the good validity of the Matlab calculation. These parameters will lay the base for much of the results of the convolved rooms. The lack of consistently correlation in the calculation of T_{20} is somewhat unfortunate. However, as the EDT will be used as the measure for the reverberation time, the subjective evaluation of the reverberation time will be more in focus.

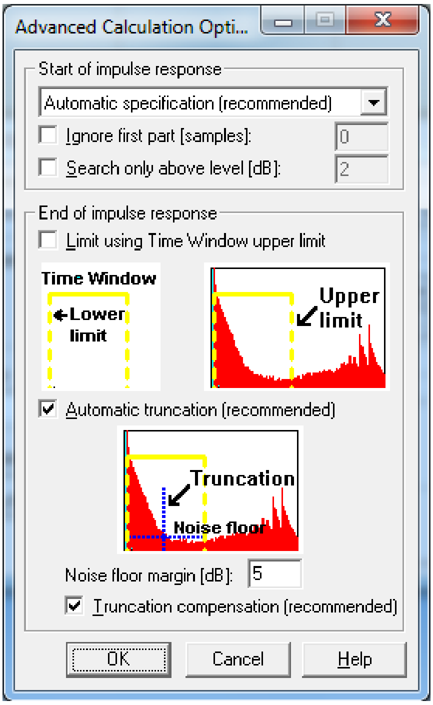


Figure 3.5: Impulse response truncation alternatives in WinMLS.

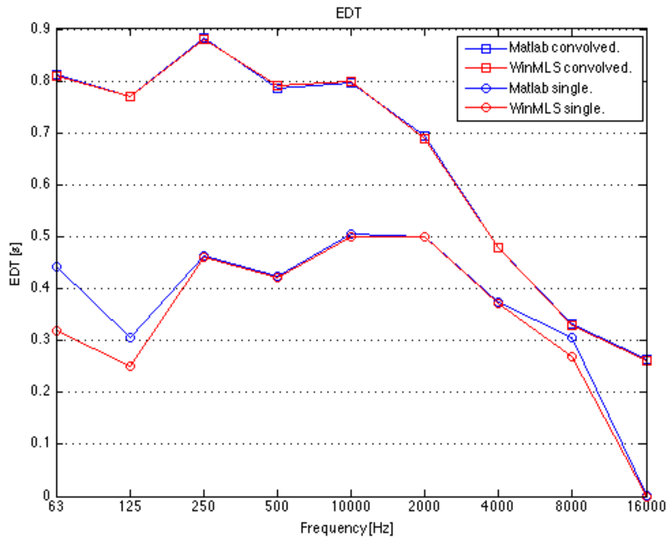


Figure 3.6: EDT computed in Matlab and WinMLS. For single room and the room convolved with itself.

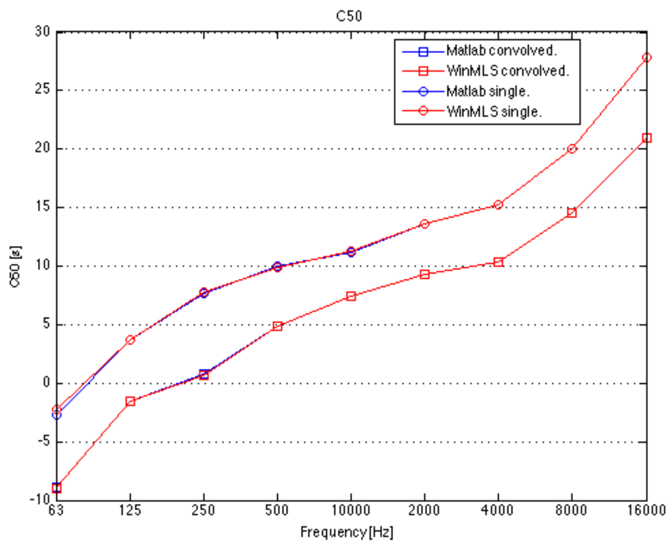


Figure 3.7: C_{50} computed in Matlab and WinMLS. For single room and the room convolved with itself.

Chapter 4

Results

The results of the room acoustic measurements will first be given by a description of the different rooms in terms of room volume, interior properties and other aspects that may be of interest. The relationship between properties of the single-room situation will be presented and the results will be compared to the results of the combination of rooms.

4.1 Room descriptions

Acoustic measurements are done in eleven rooms. Each room is described with a floor plan, pictures and single value representation of its acoustical parameters. The single value for the reverberation time and speech clarity is obtained from measurements by evaluating the arithmetic mean of the 500 Hz and the 1 kHz octave band values.

The sketches of the rooms floor plan shows an approximately layout of the interior of the room, not in accurate scale. There is performed three different types of measurements with individual microphone- loudspeaker setup as explained in chapter 3. This is indicated by using symbols of three different colors. Together with the sketch is a legend explaining the different measuring points.

Room 1

This room is a multi purpose room. It is used as a small lecture room, meeting room, lunch room and for video conferencing. The room is equipped with a video conferencing unit, white boards, projector, screen, bookshelves and a large table. One wall consists mostly of windows, with light curtains. The second wall has two whiteboards and a projector screen on a concrete back wall. Bookshelves are covering roughly 25% of the total wall area, behaving as a good sound absorbent. The remaining wall area is covered with wooden panels. The floor is hard linoleum and the ceiling consists mostly of wooden acoustical treatment. The room has the capacity of approximately 10 occupants for video conferencing purposes. The video conferencing system is combined with the screen, loudspeaker and camera in the same unit. The microphone is placed on the table.

Table 4.1: Overview, Room 1

Room 1	**NTNU, Acoustic Studio B245**
Volume:	$5.8m * 5.7m * 2.8m = 92.9m^3$
Background noise, L_{eq}:	31.4 dBA
Reverberation time, T_{20}:	0.49 s
Early decay time, EDT:	0.47 s
Speech clarity, C_{50}:	7.86 dB
STI:	0.81 (Excellent)

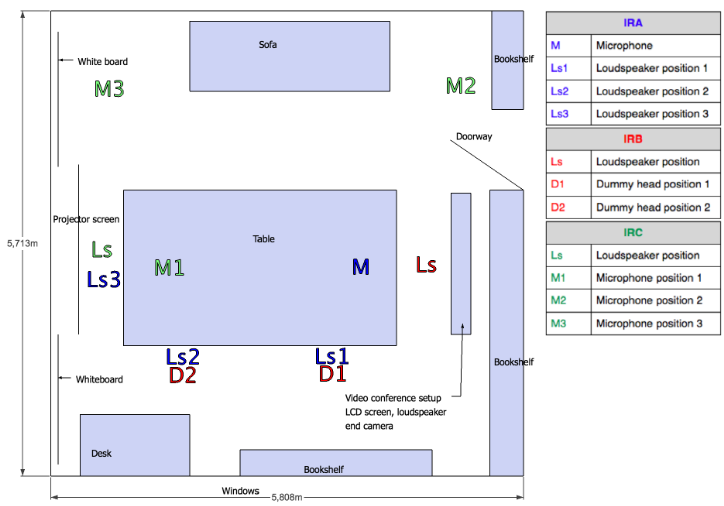


Figure 4.1: Sketch of Room 1 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.2: Room 1

Room 2

This is a room for meetings, presentations and video conferencing. One side wall is mainly windows covered by blinds and the other side wall is somewhat irregularly shaped with a lowered ceiling, see the picture in 4.4(b). The back wall is covered by sound absorbing plastic plates. Most of the ceiling consists of an acoustical drop ceiling and the entire floor is covered by a thin carpet. This is the room with the largest volume in the selection, and it has the capacity of approximately 17 seated occupants. The video conferencing system consists of a projector and screen, two table microphones and two loudspeakers, one on each side of the screen.

Table 4.2: Overview, Room 2

Room 2	SINTEF, Lerkendal, Conference room 105
Volume:	$6.95m * 5.78m * 2.90m = 116.5m^3$
Background noise, L_{eq}:	40.5 dBA
Reverberation time, T_{20}:	0.55 s
Early decay time, EDT:	0.51 s
Speech clarity, C_{50}:	8.4 dB
STI:	0.80 (Excellent)

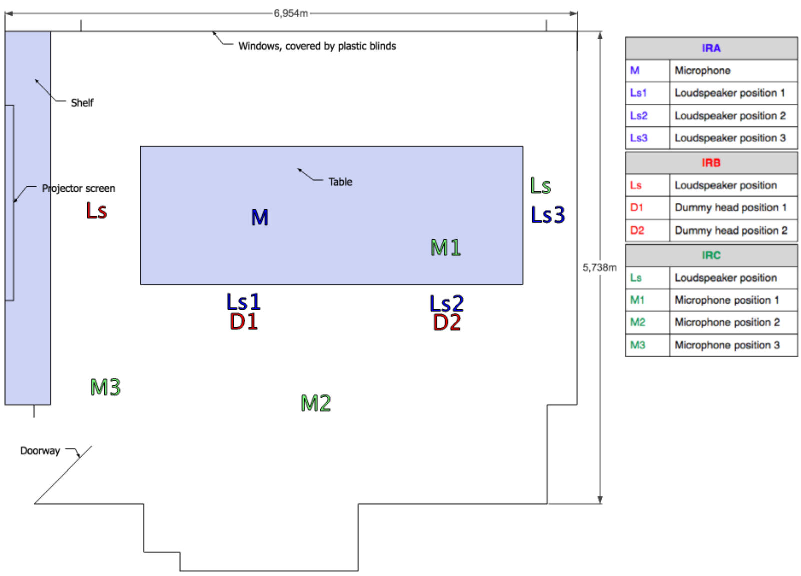


Figure 4.3: Sketch of Room 2 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.4: Room 2

Room 3

This room is used for video conferencing and as a meeting room. The two opposing side walls both consists of windows covered by blinds. The ceiling is an acoustical suspended ceiling with 3 large ventilation screens, and the floor is carpeted. The room have the capacity of 10 occupants. The video conferencing system consists of two projectors and two screens. Three microphones are suspended from the ceiling above the table. The loudspeakers are also suspended from the ceiling, with three in the front and two in the back of the room. The loudspeakers in the back of the room are there to make sure that the sound reaches all participants in the room. However, a person seated close to a loudspeaker may experience that the sound is only heard from that single loudspeaker. A collapse of sound to the rear loudspeaker is not fortunate, as it creates a separation of the visual communication on the screen in the front, and the speech communication now heard coming from behind. This may be avoided by playing a signal with some delay or damping through the rear loudspeakers.

Table 4.3: Overview, Room 3

Room 3	**Viju, Trondheim, TRD1**
Volume:	$7.2m * 4.2m * 2.70m = 81.6m^3$
Background noise, L_{eq}:	35.0 dBA
Reverberation time, T_{20}:	0.41 s
Early decay time, EDT:	0.34 s
Speech clarity, C_{50}:	12.1 dB
STI:	0.87 (Excellent)

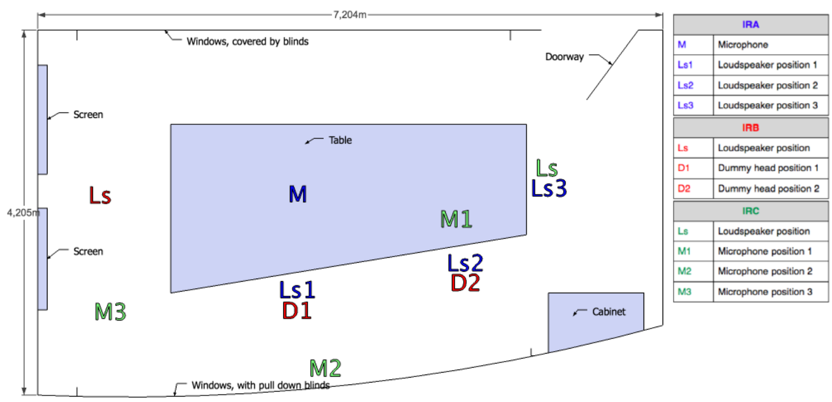


Figure 4.5: Sketch of Room 3 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.6: Room 3

Room 4

This is a small room, with video conferencing as the main usage. The intimate atmosphere can create a feeling of close and direct connection between the two parties of the video conference. The front and back wall are plain gypsum walls. One side wall is mainly windows and the other consists of a smaller window and the door. The room is somewhat oddly shaped with a narrow space extending along the windowed side wall, see the floor plan in Figure 4.7. The room has the capacity for three occupants. The video conferencing system has two 40-inch screens with the loudspeaker under and camera mounted above, and a microphone placed at the table.

Table 4.4: Overview, Room 4

Room 4	Viju, Trondheim, TRD2
Volume:	$3.6m * 2.5m * 2.70m = 24.3m^3$
Background noise, L_{eq}:	35.6 dBA
Reverberation time, T_{20}:	0.41 s
Early decay time, EDT:	0.24 s
Speech clarity, C_{50}:	14.2 dB
STI:	0.88 (Excellent)

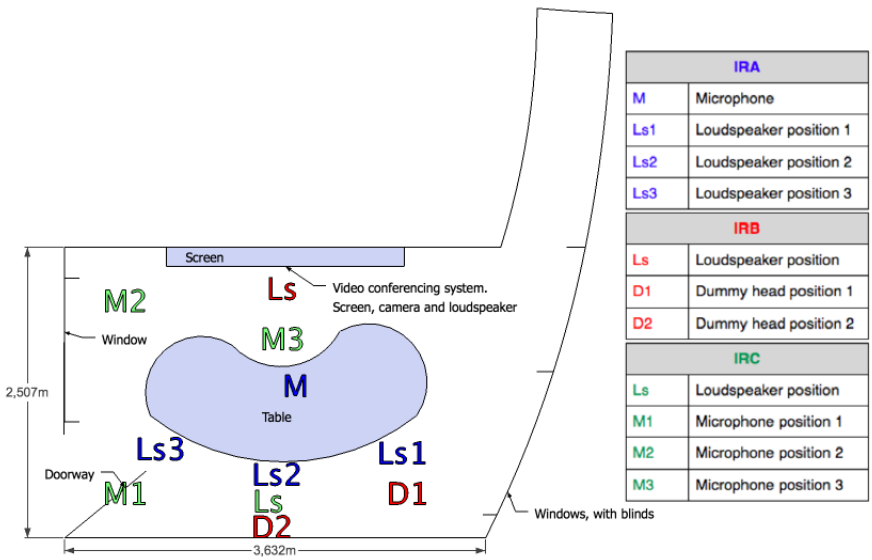


Figure 4.7: Sketch of Room 4 with the different microphone loudspeaker setups.

Figure 4.8: Room 4

Room 5 and Room 6

This room is measured both with acoustical curtains folded to the side (Room 5) and with the curtains covering two of the glass walls (Room 6). The curtains are especially designed for acoustical purposes, and contributes to create a more intimate atmosphere by improved acoustic quality and visual privacy in the room. Of the other two walls, one is mainly windows and the last is a plain gypsum wall with two 50-inch screens. The screens are part of a combined video conferencing unit, with loudspeakers below and the camera on top of the screens. Two microphones are attached to the ceiling. The ceiling is suspended and the floor is hard wood. The room is dedicated for video conferencing, with the capacity of 5 occupants.

Table 4.5: Overview, Room 5

Room 5	**Viju, Fornebu, OSL1, without curtains**
Volume:	$4.7m * 4.5m * 2.8m = 59.2m^3$
Background noise, L_{eq}:	44.4 dBA
Reverberation time, T_{20}:	0.38 s
Early decay time, EDT:	0.25 s
Speech clarity, C_{50}:	14.7 dB
STI:	0.87 (Excellent)

Table 4.6: Overview, Room 6

Room 6	**Viju, Fornebu, OSL1, with curtains**
Volume:	$4.7m * 4.5m * 2.8m = 59.2m^3$
Background noise, L_{eq}:	44.4 dBA
Reverberation time, T_{20}:	0.29 s
Early decay time, EDT:	0.18 s
Speech clarity, C_{50}:	16.7 dB
STI:	0.90 (Excellent)

It must be mentioned that there was much noise from the ventilation when the measurements were performed. The room had just before been fully occupied, making the ventilation noise very noticeable. The background noise was measured to 44.4 dBA. In a similarly constructed room in the same building (Room 7), the background noise was measured to 30.9 dBA. This indicates that the contribution from the ventilation system adds about 14 dBA to the regular noise level.

The results displayed in Table 4.5 and 4.6 shows an improvement in the values for

T_{20}, EDT, C_{50} and STI by covering two walls with curtains. This room stands out as being one of best in the selection of rooms in the survey. With the curtains it takes an extra leap, and the room clearly has the best values for intelligibility and clarity of speech, reverberation time and sound decay. This is even with the extensive background noise.

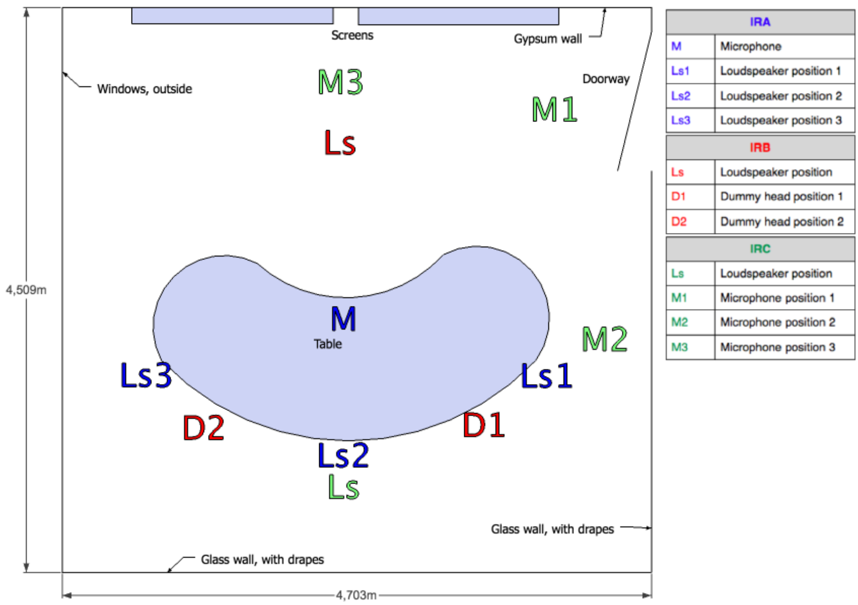


Figure 4.9: Sketch of Room 5 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.10: Room 5

Room 7

As the previous room, this also have two walls of glass, one of windows and the last wall made of gypsum, with the possibility of covering two glass walls with curtains. The projector screen and two white boards covers much of the gypsum wall. The room is furnished with a table and chairs. The video conferencing system is set up with two microphones attached to the ceiling and loudspeakers built into the ceiling. The room has a capacity of approximately 11 occupants. Although being a rather large room with mostly glass walls, the reverberation time is kept at a good level.

Table 4.7: Overview, Room 7

Room 7	**Viju, Fornebu, OSL6, Board Room**
Volume:	$7.1m * 5.7m * 2.8m = 113.3m^3$
Background noise, L_{eq}:	30.9 dBA
Reverberation time, T_{20}:	0.53 s
Early decay time, EDT:	0.34 s
Speech clarity, C_{50}:	13.1 dB
STI:	0.84 (Excellent)

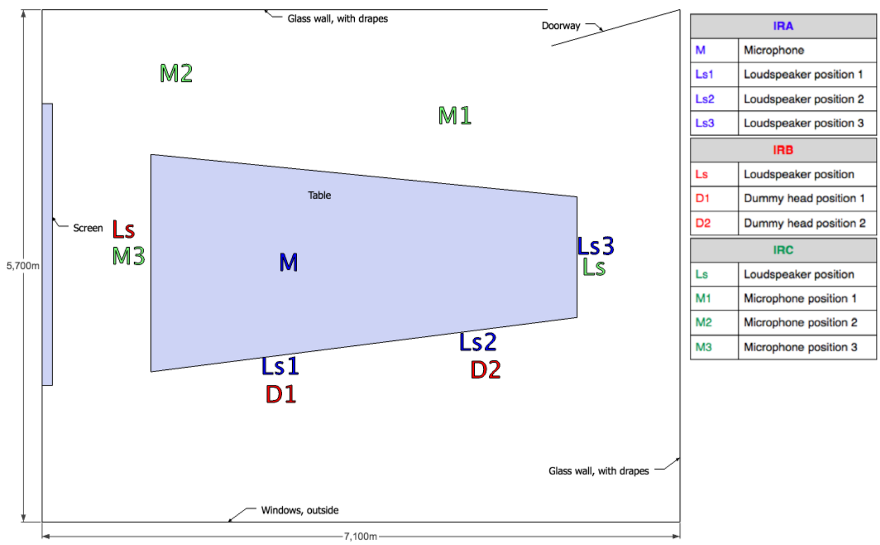


Figure 4.11: Sketch of Room 7 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.12: Room 7

Room 8

This room is perceived as having uncomfortable sound conditions related to noise and room reverberation. Three of the walls consists of plain gypsum, the last is a glass wall with a wooden door. The floor is carpeted and the ceiling consists of a suspended ceiling in the middle with opening around on all sides. These openings reveals surfaces of concrete. Two loudspeakers are hanging high up on the front wall. With this placement, the loudspeakers are close to the concrete surfaces. This can be unfortunate as it will create very noticeable reflection of the direct sound, and may be a part of the reason for why the room is perceived as being uncomfortable. Two microphones are placed on the table, the camera is mounted on top of the 50" screen standing in the front corner of the room.

Table 4.8: Overview, Room 8

Room 8	**Aker Solutions, Fornebu, Attitude**
Volume:	$7.5m * 4.4m * 2.7m = 89.1m^3$
Background noise, L_{eq}:	38.2 dBA
Reverberation time, T_{20}:	0.70 s
Early decay time, EDT:	0.61 s
Speech clarity, C_{50}:	8.0 dB
STI:	0.78 (Excellent)

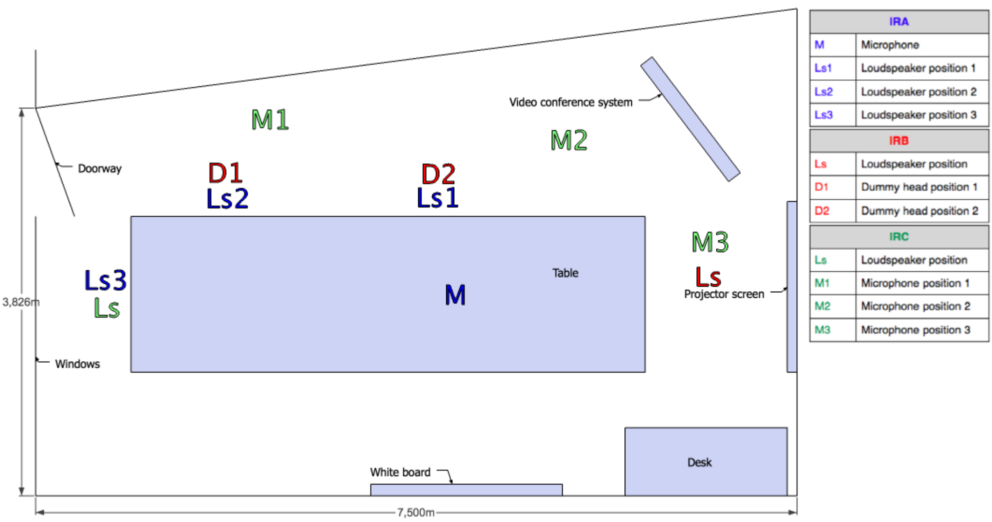
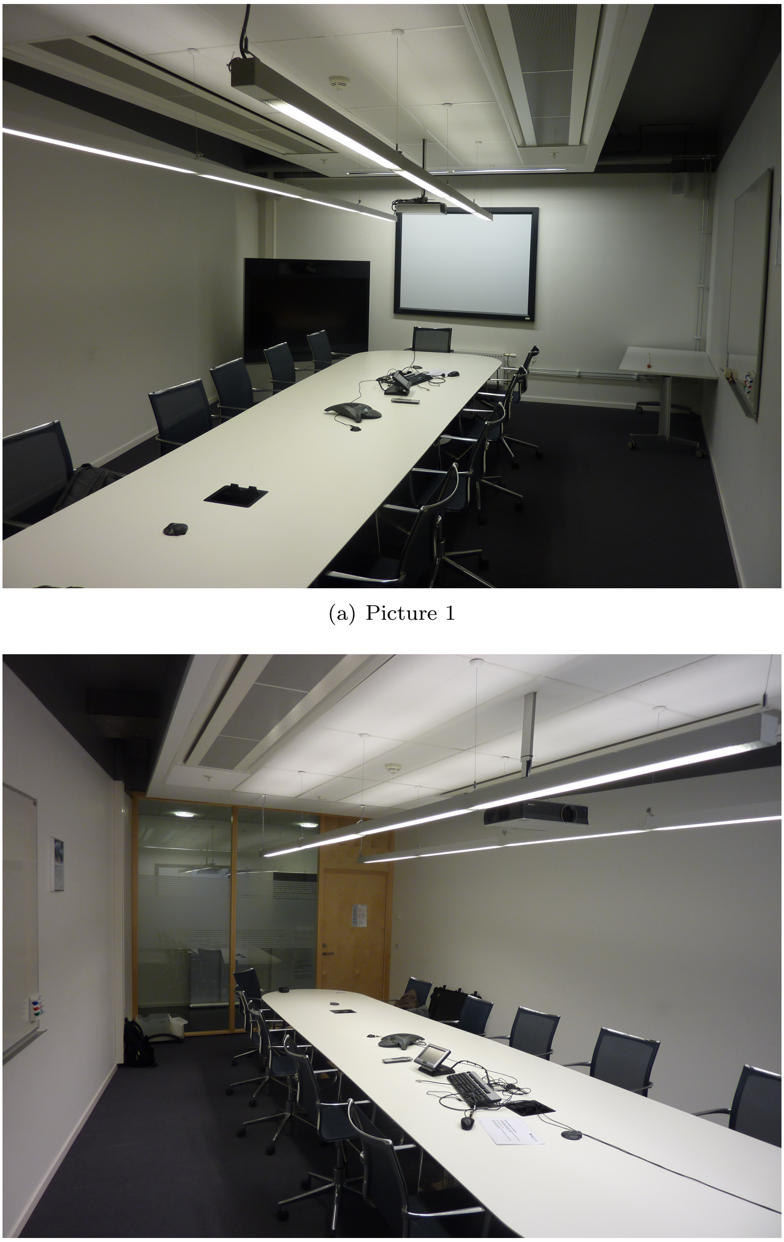


Figure 4.13: Sketch of Room 8 with the different microphone loudspeaker setups.



(a) Picture 1

(b) Picture 2

Figure 4.14: Room 8

Room 9

This room has a suspended acoustical ceiling and a carpeted floor. The two side walls are mainly windows, covered with blinds. The front and back wall are plain gypsum. Two 50" screens are hanging on the front wall, with one loudspeaker at each side. There is one single microphone suspended from the ceiling just in front of the table. This is a dedicated video conferencing room with capacity of 4-5 occupants.

Table 4.9: Overview, Room 9

Room 9	Viju, Stavanger, SVG2
Volume:	$4.6m * 4.5m * 2.6m = 53.8m^3$
Background noise, L_{eq}:	36.1 dBA
Reverberation time, T_{20}:	0.40 s
Early decay time, EDT:	0.30 s
Speech clarity, C_{50}:	14.4 dB
STI:	0.87 (Excellent)

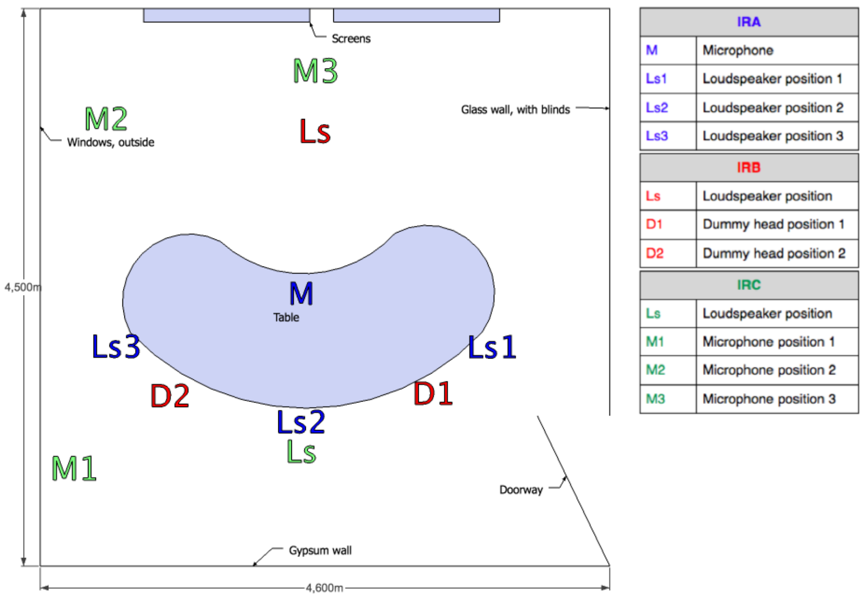


Figure 4.15: Sketch of Room 9 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.16: Room 9

Room 10

This room have some obvious signs of acoustical treatment. The walls are concrete, but have large parts covered by thick wall carpets and curtains. The curtains were folded together during the measurements, as shown in the pictures. The floor is covered with a thick carpet, with apparent good sound absorption qualities. There is a lowered acoustical ceiling covering most of the overhead area. A microphone is suspended from the ceiling, and the loudspeaker is integrated in the 40” screen on the wall. The room has a capacity of 7-8 occupants for video conferencing.

Table 4.10: Overview, Room 10

Room 10	**Aker Solutions, Stavanger, M7D**
Volume:	$4.7m * 5.95m * 2.7m = 75.5m^3$
Background noise, L_{eq}:	30.8 dBA
Reverberation time, T_{20}:	0.34 s
Early decay time, EDT:	0.29 s
Speech clarity, C_{50}:	13.2 dB
STI:	0.87 (Excellent)

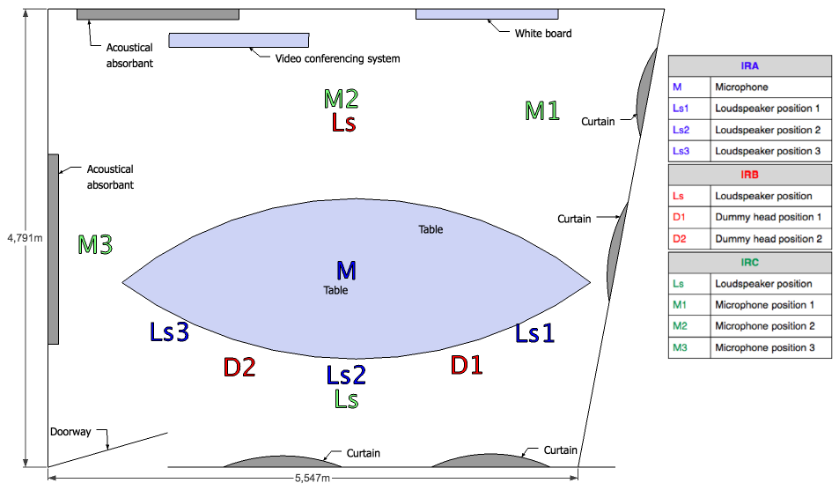


Figure 4.17: Sketch of Room 10 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.18: Room 10

Room 11

This room is used as a meeting room with video conferencing possibilities, and has the capacity for as much as 15 occupants. The acoustic quality in this room is perceived as being poor, and strenuous to spend much time in. The hard wooden floor, glass and concrete walls and the windows contribute to this fact. The ceiling is partly covered by acoustical plates, but there are also areas of reflective surfaces. In the survey, this is the room that clearly stands out with the poorest values for all the acoustical parameters, except for the background noise. The room has two microphones suspended from the ceiling and one placed on the table. The loudspeakers are placed in the top corners of the front wall where there is a projector screen and a 40" screen.

Table 4.11: Overview, Room 11

Room 11	**Progressus, Stavanger, Board Room**
Volume:	$8.4m * 4.4m * 2.7m = 99.8m^3$
Background noise, L_{eq}:	40.8 dBA
Reverberation time, T_{20}:	0.97 s
Early decay time, EDT:	0.92 s
Speech clarity, C_{50}:	3.33 dB
STI:	0.67 (Good)

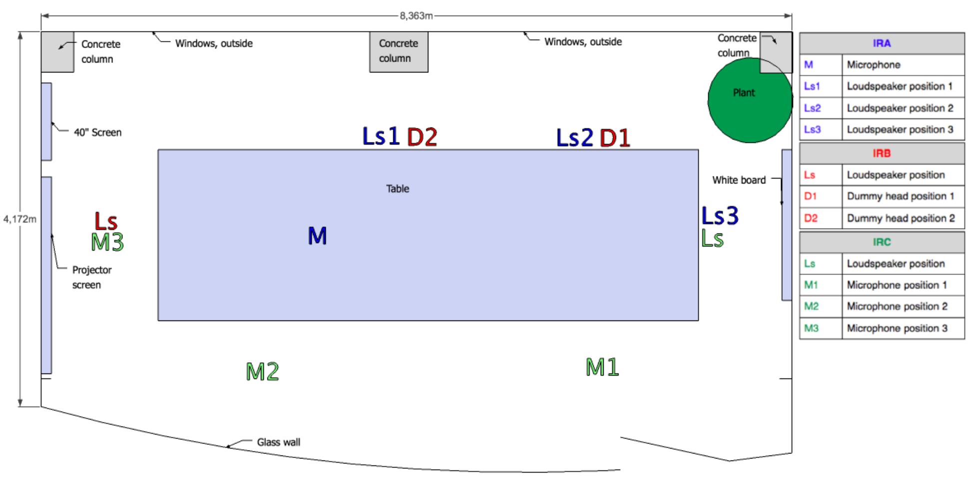


Figure 4.19: Sketch of Room 11 with the different microphone loudspeaker setups.

(a) Picture 1

(b) Picture 2

Figure 4.20: Room 11

4.2 Individual rooms

4.2.1 Early decay time and speech clarity

The early decay time and the speech clarity are two important measures in mapping the acoustical conditions of video conferencing rooms. Both the EDT and the C_{50} vary with frequency, and are calculated for the octave bands from 63-16000 Hz. To represent the rooms with one single value for each parameter, the 500 and 1000 Hz octave band values are averaged.

Figure 4.21 and 4.22 shows the EDT and C_{50} value of the eleven measured rooms. From these bar diagrams, it is easy to see that there is a relation between the two parameters. It is almost possible to turn the EDT plot on its head and place it on top of the C_{50} diagram. When the EDT is plotted versus the C_{50}, the best results are at the bottom right and the poorest at the top left. There can be observed tendencies of grouping of some rooms.

One room stands out as better than the rest, having the lowest reverberation time and the highest speech clarity. This is the room with two walls covered by curtains especially designed for acoustical performance. On the other side, one room clearly stands out as being the poorest. This room have one large wall of glass, with the opposing wall being windows and structural pillars made of concrete. The room have a very long reverberation time, and also a low value for speech clarity.

Table 4.12 shows the suggested limits for the quality categories. Figure 4.23 illustrates the same in a plot of EDT versus C_{50}. The choices for the categorization are discussed in section 5.2.

Table 4.12: Suggested quality categories

Category	**EDT value [s]**	**C50 value [dBA]**
Cat.A	0 – 0.30	> 12
Cat.B	0.30 – 0.45	9.0 – 12.0
Cat.C	0.45 – 0.60	6.0 – 9.0
Cat.D	> 0.60	≤ 6.0

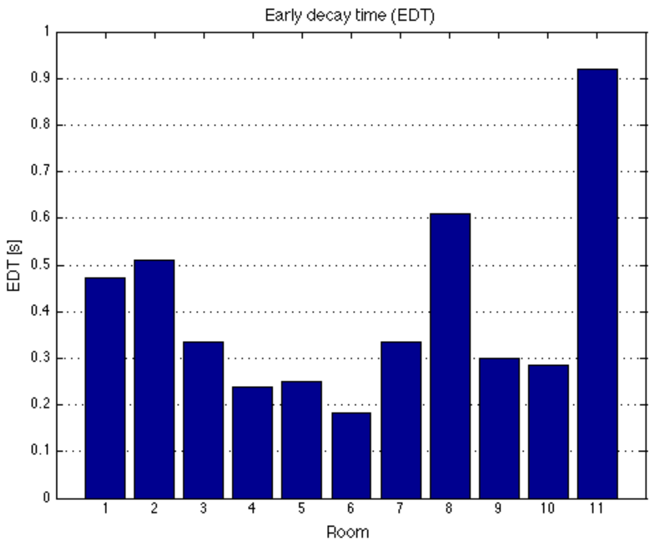


Figure 4.21: Early decay time for the selection of rooms

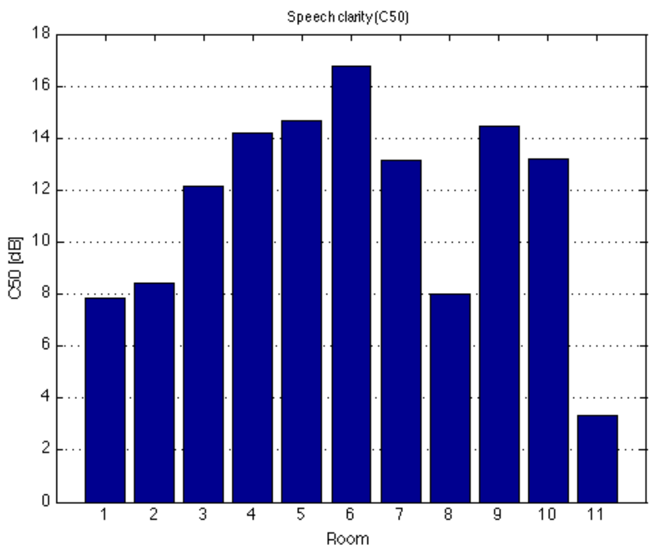


Figure 4.22: Speech clarity for the selection of rooms

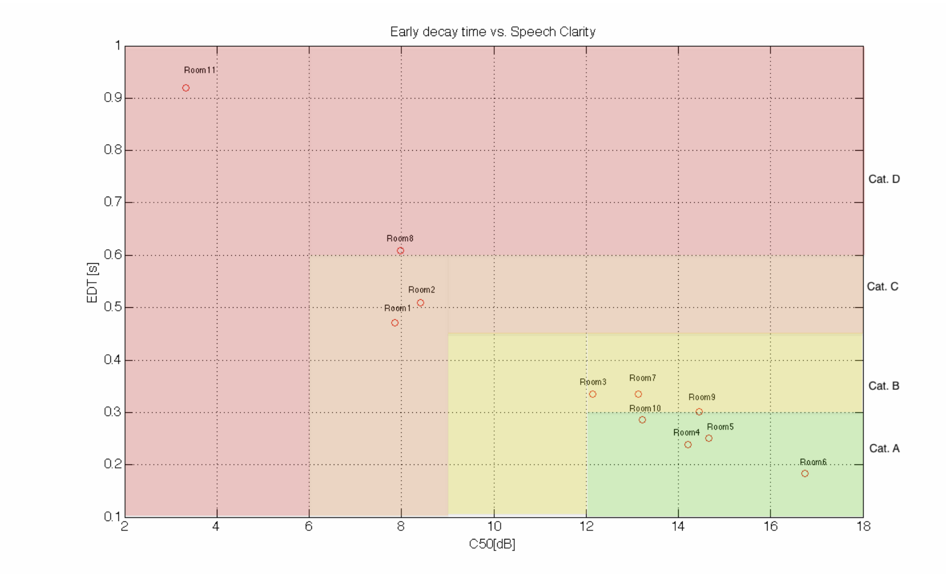


Figure 4.23: EDT plotted versus C50 with suggested quality categories indicated

4.2.2 Background noise

The background noise was measured in all rooms, averaged over an interval of one minute. This was done with doors and windows closed, and with all the videoconferencing equipment turned on. The measurement in room 5 and 6 were done with much ventilation noise, as the room had previously been fully occupied and the ventilation had kicked in. The levels of the background noise varies from 30.8 dBA, as the lowest, to 44.4 dBA, being the highest. The noise levels are plotted in Figure 4.24 with the sound classes for noise in offices from NS-8175 indicated.

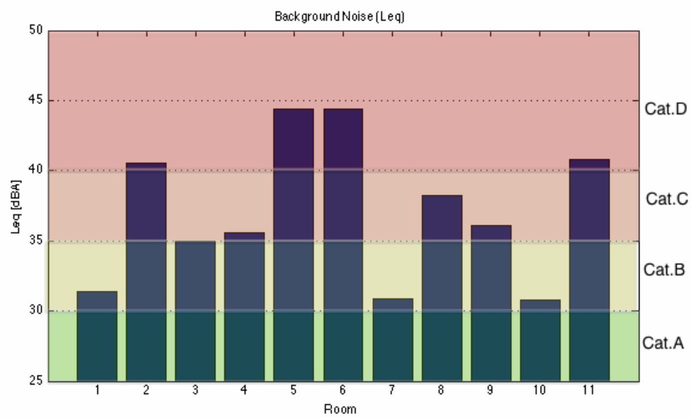


Figure 4.24: Background noise in the selection of rooms, sound categories from Table 2.3 indicated

4.2.3 Speech transmission index versus speech clarity

The speech transmission index and the speech clarity both give a measure for the intelligibility of speech. Figure 4.25 shows the values of the STI plotted against the C_{50} values for each of the eleven rooms. As expected, there is an obvious relationship showing that a higher STI gives a higher value for speech clarity.

The majority of rooms are rated as being "Excellent" by the STI. Only one room is rated as "Good". A curve of the third order is added as a good fit for the data points.

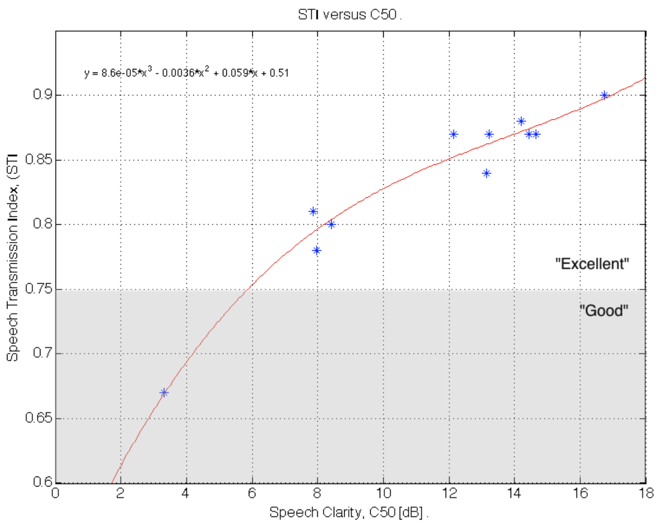


Figure 4.25: The relationship between STI (without noise) and C50

4.3 Combination of rooms

4.3.1 Early decay time

The early decay time for all the combinations of rooms are calculated after performing convolution of the impulse responses in Matlab. The EDT of all room combinations are displayed in a histogram in Figure 4.26. The single room values are also included for comparison. Each of the eleven rooms are convolved with all the other rooms, including itself. This gives a total of 121 room combinations. Figure 4.27 shows a graphic representation of the EDT of all combinations. The figure shows eleven bars for each room, which is all its combinations with the other rooms. Every other room has red and blue bars, to easily separate them. There are some trends that can be seen, and from the very first glance, all combinations with room 8 and 11 have a very high EDT.

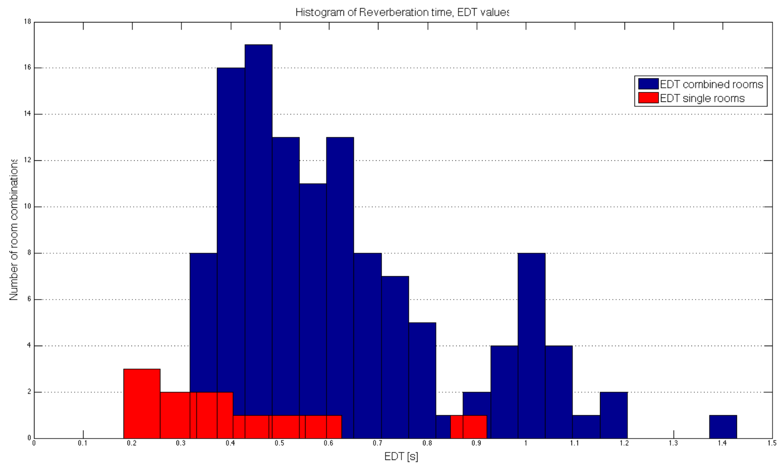


Figure 4.26: Histogram of EDT values, single rooms and combined rooms

Figure 4.28 shows how the EDT is distributed over the share of rooms. The curves indicates the percent-wise share of rooms that fall within a given value of the EDT. The blue line is for single rooms and the red line for the combined rooms. Indicated on the figure is the point for $EDT = 0.5s$. It can be seen that 73% of the single rooms are within this value. On the other hand, only 45% of the combined rooms have values below this point. This shows that when both the transmitting and receiving room are considered together, the parameter value will degrade significantly. This is not necessarily a bad property, it is merely the effect

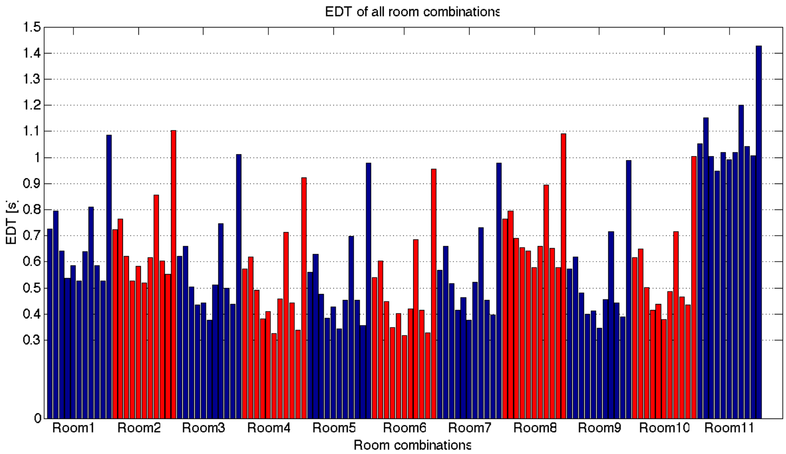


Figure 4.27: EDT of all room combinations

that rises when considering the combination of two rooms. There are no standard values for preferred parameter values of combined rooms yet. From the figure, it is observed that the main part, approximately 80%, of the single rooms have EDT values below 0.6 s. To span the similar share of the combined rooms, the EDT values increase up to 0.8 s.

4.3.2 Speech clarity

The speech clarity data is represented by averaging the measurements of different loudspeaker-microphone positions, as it is done for the reverberation time. Figure 4.29 shows the C_{50} values of the single rooms and the combinations of rooms in a histogram. The C_{50} values for all room combinations are displayed in Figure 4.30.

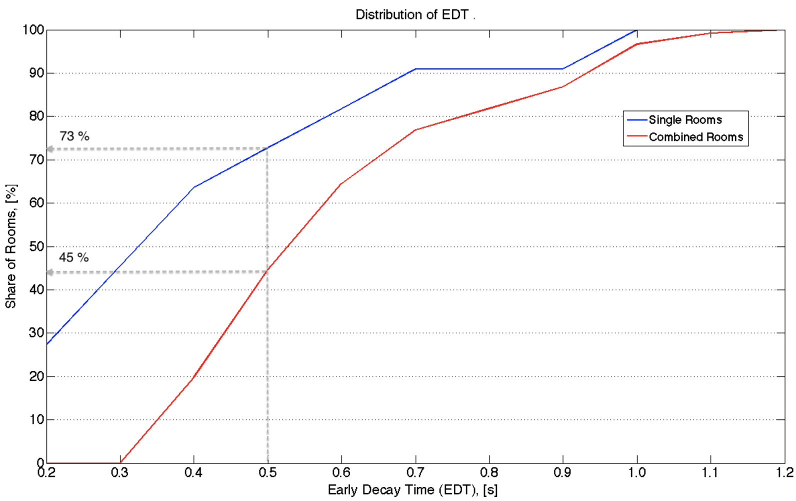


Figure 4.28: Percent-wise distribution of EDT in single rooms, and for combined rooms

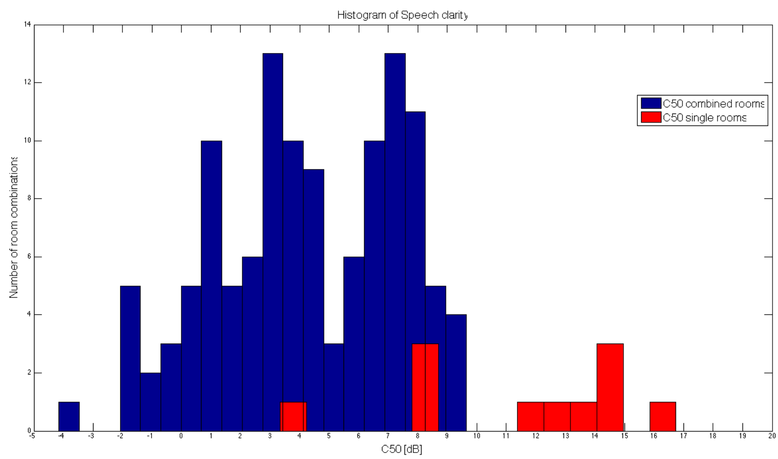


Figure 4.29: Histogram of C_{50} values, single rooms and combined rooms

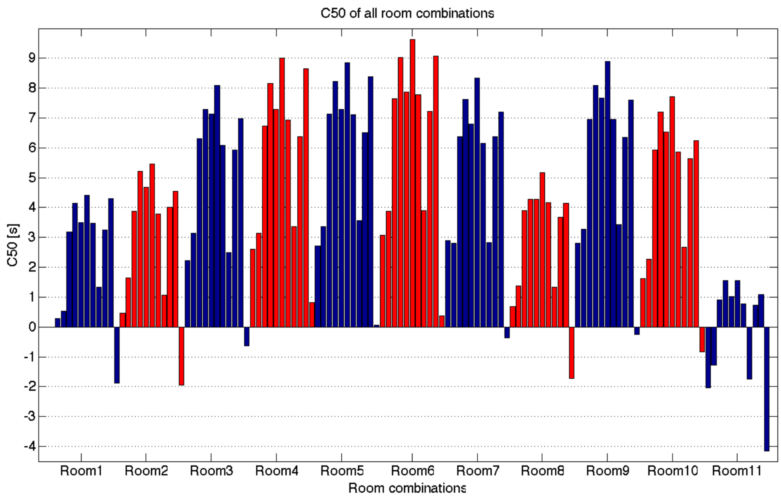


Figure 4.30: C_{50} of all room combinations

4.3.3 Early decay time versus speech clarity

To get an overview of the how the single room values differ from the combined values, all rooms are shown convolved with themselves in Figure 4.31. The convolution leads to a clear shift in the data points. It can be observed that a convolution leads to an average increase of 0.22 seconds for the early decay time. The speech clarity suffers an average drop of 7.0 dB in the convolution.

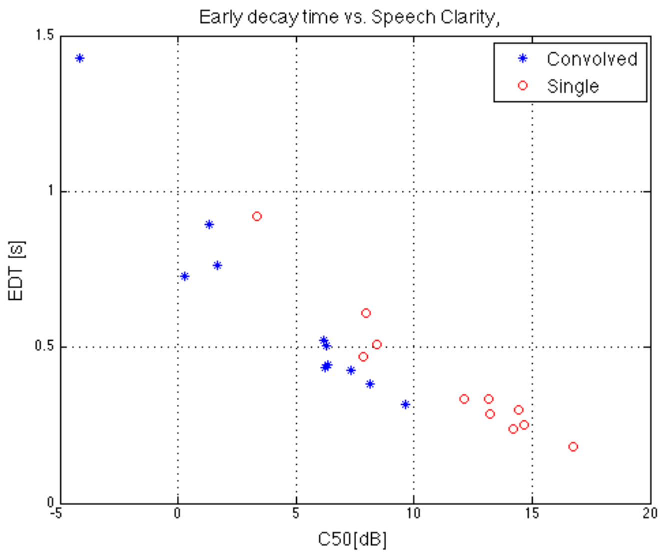


Figure 4.31: Results of single rooms, and convolution of the room with itself.

Figure 4.32 shows the relationship between the early decay time and the speech clarity for all room combinations and for single rooms. It can be seen clearly, that a lower reverberation time leads to better speech clarity. Most of the single rooms, displayed as red circles, have higher values for speech clarity than the highest of the combined rooms. An interesting observation is that single rooms have better speech clarity than combined rooms with the same reverberation time. The grey areas indicates which values that are considered to be acceptable for single rooms and for combination of rooms. For the single-room situation the white area is considered acceptable, and both grey areas to have undesirable values. The combination of rooms, only the area of the darkest shade of grey is considered undesirable. The choice of these values is discussed in section 5.3.

It is interesting to know how different room types perform together. The best combination has an EDT of 0.31 seconds and a C_{50} of 9.6 dB. The worst combi-

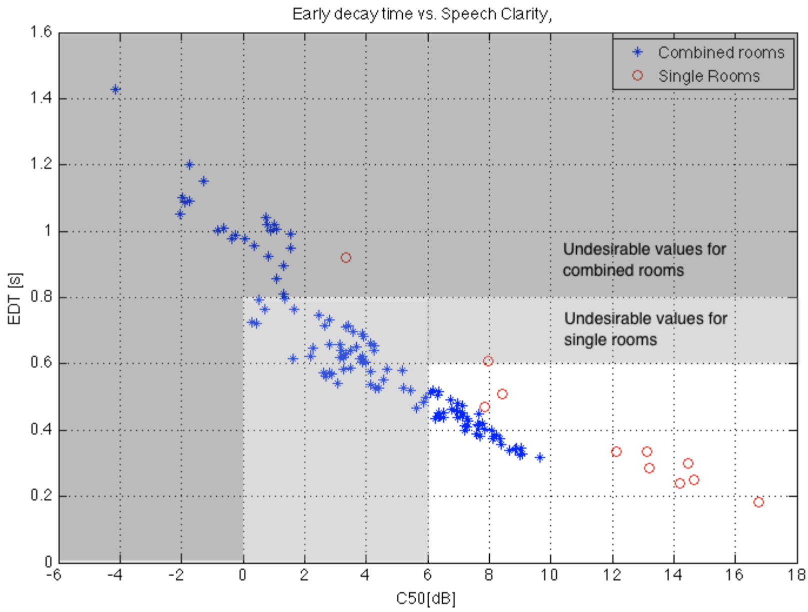
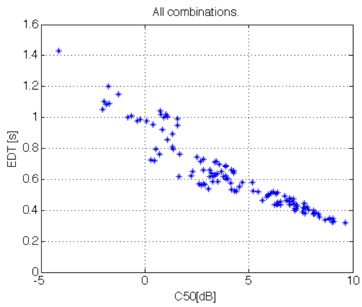
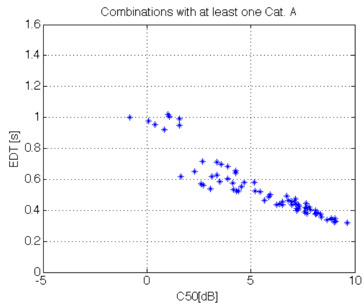


Figure 4.32: Suggested limits for EDT and C50 values

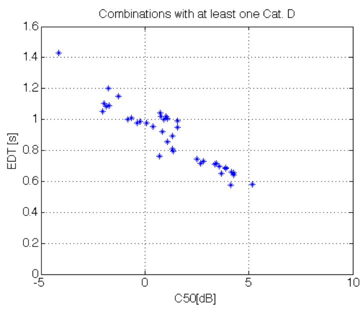
nation gives an EDT of 1.43 seconds and C_{50} of -4.2 dB. If the best room is to be combined with the worst room, the EDT is 0.99 seconds and the C_{50} is 1.6 dB. Different selections of room combinations are plotted in Figures 4.33(a)-4.33(e) and will be discused in section 5.3



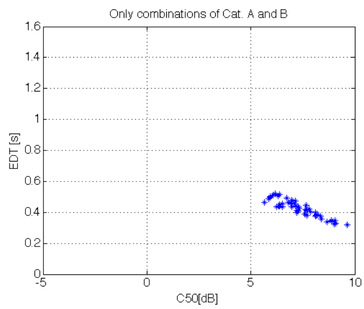
(a) All room combinations



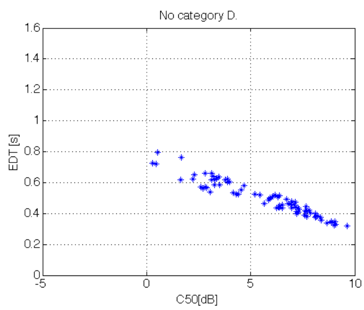
(b) Combinations with at least one room of category A



(c) Combinations with at least one room of category D



(d) Combinations of categories A and B



(e) No category D

Figure 4.33: Combinations

Chapter 5

Discussion

5.1 Method of measurement

This thesis focuses on the room acoustic conditions of the video conferencing rooms, not the equipment itself. The properties of the video conferencing equipment is therefore not regarded in this research. Measurements will be done in the same way with the same measurement equipment in all rooms. Through post-processing of the measurements, the acoustic conditions can be described, both in single rooms and for any combination of two rooms.

Different microphones are used when the room is in the sender and the receiving situation. For recording the IR of the sender room, an omni directional 1/2"-microphone is used. As for recording the IR in the receiving room, microphones in a dummy head captures the rooms response. The frequency response for the dummy head and the measurement microphone are clearly different. While the microphone have a fairly flat response, this is not the case for the dummy head. Although in the interval between 300Hz and 2000Hz, they are fairly correlated. On frequencies lower and higher than this, they differ significantly. This does not necessarily cause a problem. In fact, this is part of the reason why the dummy head is used, as the frequency response of our hearing is not flat either.

An alternative to the chosen method could be to measure the room combination directly. It could be done by having the source located in one room, and recording measurements in a different receiving room. This way the transmission channel would be included. However, the results would heavily rely on the quality of the recording/playback equipment. Many rooms are custom made and the types of microphones, loudspeakers and processing units differs much from room to room.

These differences would have made it challenging to get comparable results for the actual room acoustic conditions.

5.2 Individual rooms

The results obtained from the room acoustic measurements for each room are intended as a survey of the current conditions in video conferencing rooms. There are no standard parameter values or design specifications made especially for this type of rooms. By combining the results with specifications of similar types of rooms, there can be defined different quality categories within ranges of the measured parameters. From the plot of the C_{50} and the EDT of each room, there is an obvious relation. A short early decay time gives high speech clarity, and when the early decay time is long, the speech clarity goes down. Values for these two parameters will therefore be used together in the segmentation of categories.

An overview of the acoustical conditions in all eleven rooms are presented in the results chapter. The reverberation time, T_{20}, varies from 0.29 up to 0.97 seconds. This is a significant spread, which can help in getting insight in what values that most definitely should be avoided. As the T_{20} values represent the reverberation properties of the room, the early decay time, gives a more subjectively representation of how the sound decay is perceived inside a room. People have the ability to suppress redundant sound content below a reasonable level. The EDT measures how fast the first 10 dB of a sound decays, indicating how long the sound will uphold a level so high it may be disturbing. If the level is sustained, the reverberation will interfere with the following speech and diminish its intelligibility. A relative change in EDT of 5 %, is regarded as the just noticeable difference, [5]. As an example, Room 5 and 6 can be compared. This is the same room, with and without acoustical curtains covering two of the walls. The EDT is relatively 28% shorter with the curtains. On other words, a change that is very noticeable. The two rooms with the most similar value for EDT, have the values 0.30 s and 0.29s. This makes the shorter one only 3.3% lower, which will not perceived to be different.

The values for T_{20} are higher than the values for EDT in all the measured rooms. If the decay curves obtained from the impulse responses were perfectly linearly decreasing, the value for these two parameters would be exactly the same. However, there is a tendency of a drop in the early part of the decay curves before it evens out for a steady decay. The drop makes sure that the decay from 0 to -10 dB of the steady state sound level leads to a shorter value for the reverberation time than the drop from -5 to -25 dB, which is used for the calculation of the T_{20} value.

As mentioned, the value for speech clarity is seen to be strongly related to the reverberation time. This is well known, as the values for speech clarity are obtained by computing the ratio between the early and late energy in an impulse response. A short reverberation time will result in little energy in the late part of the impulse response, giving a high C_{50} value. Thus, short reverberation time yields high speech clarity. The value of the just noticeable difference for C_{50} is 1 dB. The variation of 13.4 dB in C_{50} values between the rooms will then be very noticeable. Of the eleven rooms, five of them have a high speech clarity with values within a range where it is hard to notice a difference between them. Another group of three rooms have similar C_{50} values around 8 dB. This gives two groups of rooms, one with good and one with intermediate speech clarity. In addition to these two groups, one room stands out with a significantly higher and one with a lower value. This shows that video conferencing rooms have a wide and noticeable variation in speech clarity. Still there seems to be certain C_{50} levels that are more common than others. Of the selection of rooms surveyed here, these levels are around 14 dB and around 8 dB.

A measure of the intelligibility of speech is, in addition to speech clarity, given by the speech transmission index. The rooms considered in this research are of a reasonable small size, the largest having a room volume of 116 m^3. This indicates that there should be a relatively good speech intelligibility regardless of the design of the room. This is shown to be the case, as only one of the rooms are rated as "Good" and all others as "Excellent". Thus, acceptable rooms should have a STI rating as "Excellent", and have STI values preferably above 0.85. Results show that the levels of the STI follows the values for the speech clarity, this is shown in Figure 4.25. The figure include a curve of best-fit third-order polynomial to the data points. The STI clearly express a quality rating of the speech intelligibility in a room through its results. This can be transferred to the C_{50}, as there is a lack of quality measures for the speech clarity. From the fitted curve, it can be seen that the STI quality rating have a transition just below the C_{50} value of 6 dB. This may indicate that a value of 6 dB represents a lower limit for what can be regarded as excellent speech clarity.

The relation of the STI and C_{50} is a good match with the results in the study of relationship among measures of speech intelligibility in rooms by Bradley, [18]. Bradley also introduced another predictor for intelligibility of speech in rooms, [19]. Similar to the C_{50}, it uses the ratio concept. This predictor uses the useful-to-detrimental sound ratio, and is named U_{80}. This predictor combines a measure of the room acoustics and the speech-to-noise ratios in a single quantity. Useful-to-detrimental sound ratios are the ratio of the useful early arriving sound to the combination of the later arriving speech sounds and the ambient noise in the room. The U_{80} is an alternative to the STI and according to Bradley the preferred predictor of speech intelligibility. Nevertheless, this thesis has chosen the STI as the measure for speech intelligibility as it is a widely established predictor with clearly stated intervals of quality.

An important aspect of the acoustic quality of a room is the noise level. From the measurements of the background noise, four of the eleven rooms have noise levels considered to be disturbing and that are prone to cause fatigue with extended use. Two of these four rooms are known to usually have a significantly lower noise level, as the measurement was taken with the ventilation system going on full power. The noise measurement does in fact show an interesting observation. The rooms with the clearly highest background noise levels, are considered the best when evaluated by the EDT and C_{50} values. So, the noise level does not seem to have any relationship with the other parameters. In fact, both the room with the best and worst values for EDT and C_{50} have noise level in class D, according to NS 8175, section 10.5. Even though the background noise at these levels does not affect the other acoustic measurements performed here, it is a factor that will cause fatigue with exposure over time. Thus, the background noise level is an important factor towards the total sound quality of a room.

It is suggested for the rooms be divided into four categories, according to the value of their acoustical parameters. As the values of EDT and C_{50} shows to have a firm relationship, the categories include both values. The categories are determined based on the limit values for meeting rooms in offices and in educational rooms from NS 8175, [10]. These values are given in Table 2.4 and 2.5. Together with these values, the measurement results are also considered for the choice of intervals. Some observations and choices are made; An EDT within 0.6 seconds is acceptable. Within this limit, all results for the STI are considered "Excellent". A higher value for the EDT than this is not recommended. The values for the categories require a somewhat shorter reverberation time than what is given for meeting rooms in offices, given in Table 2.5. There are two reasons for this. First, the EDT is the chosen quantity of measurement for the reverberation time for the combined rooms. The EDT has a lower value than the T_{20} in all results, thus the categories are adjusted accordingly. Secondly, for the room to perform as a video conferencing room, the acoustical requirements are higher than for a regular meeting room. It is expected of the room to have the properties for being a good listening- and recording room in addition to meet the requirements of meeting rooms. Table 4.12 shows the interval values for the suggested categories. The lowest acceptable value for the C_{50} is 6 dB. The reasoning for choosing this value is from the evaluation of the relationship between C_{50} and STI. By looking at Figure 4.25, it can be seen that the C_{50} value in the transition between "Good" and "Excellent" is approximately 6 dB. From this, only values higher than 6 dB for the speech clarity is considered acceptable for the single-room situation.

5.3 Combination of rooms

To evaluate the combination of paired rooms, a set of impulse responses from the sender room is convolved with impulse responses of the receiver room. This combines the room acoustic properties of both rooms, which gives information of the quality of the total video conferencing transmission. In the results, the relation between values for single rooms and combined rooms are presented and compared to the results of the single rooms. These results show that there is a significant difference in the values of single rooms and for when they are combined.

- The mean value of the early decay time for the single rooms is $EDT_{single} = 0.40s$, while the mean value of the combined rooms is $EDT_{comb} = 0.63s$. That is a difference of 0.23 seconds, making it an increase of 58% for the EDT of combined rooms.

- The mean value of the speech clarity for the single rooms is $C_{50_single} = 11.5dB$, while the mean value of the combined rooms is $C_{50_{comb}} = 4.2dB$. That is a difference of 7.3 dB, making it an decrease of 63% for the C_{50} of combined rooms.

The acoustical parameters for speech clarity and reverberation time, C_{50} and EDT, are important in assessing sound conditions in a room. These two parameters have been seen to have a relationship, but at the same time there is such a spread in the data, indicating that both measures are necessary. It can be observed that even where there only is a small variation in the EDT, it has a considerable impact on the C_{50} and the other way around. The most extreme example is where there is no change in the C_{50}, which is approximately at 2 dB, the EDT changes from 0.6 - 1.0 seconds. This is a massive change in the reverberation time that has no effect on the speech clarity. Figure 4.31 show the shift in the values that is caused by the convolution. The single rooms are compared with a convolution of the room with itself. This situation is created to show the effect the convolution has on the values of EDT and C_{50}. There is a trend that rooms with higher value of the EDT, have a larger increase through the convolution than the rooms with shorter EDT. This is not the case with the speech clarity, which keeps a steady decrease regardless of the value it has in the single room. When all room combinations are considered, the selection of rooms provide a total of 121 room combinations. In Figure 4.32 all these combinations are plotted together with the values for single rooms. Further, five different selections of combined rooms are shown in Figure 4.33(a)- 4.33(e). Some observations can be made:

- The data points for all combinations span over EDT values from 0.31– 1.42 seconds and C50 values from -4.2– 9.6 dB.

- When at least one of the four rooms in category A is part of a combination, most results with reverberation time longer than 1.0 second are eliminated. Still, this is not sufficient to secure an acceptable level for the combined speech quality.
- Where at least one room is category D, most values are considered too poor for video conferencing.
- The combinations of categories A and B shows the range of values that are considered to give very good conditions for a video conference with high speech intelligibility and good sound quality. These values are also within the suggested limits acceptable quality for single rooms.
- Finally, when no room of category D is included, all combinations are considered to have a reasonably good quality of their acoustical parameters to make it suitable for a video conference situation. Though, the combination of two rooms in category C are just within what is acceptable.

As stated in the hypothesis in section 2.7, combinations of acceptable rooms are considered to give acceptable values. This is achieved when rooms of category D are removed from the selection.

5.4 Audio equipment in video conferencing rooms

Microphone and loudspeaker placement

The most common placement of the microphones are on the table or from the ceiling. From the ceiling the microphones can either be suspended down and directed towards the seating area, or mounted directly on the ceiling itself.

The loudspeaker setup and placement varies a lot. The most basic case is with a fully integrated video conferencing setup, where the loudspeaker, the screen and the camera is all part of one unit. Other setups include multiple LCD screens or projector screens and thus more individual placement of the loudspeakers. The most common and basic placement is the two channel setup, with one loudspeaker on each side of the screen. In bigger rooms it may be desirable to have additional channels further back in the room, to make sure the sound covers all listener positions in the room equally.

The choices for microphone and loudspeaker placement is individual for each room. The size of the room and expected number and placement of participants are important factors to take into consideration. Some rooms can be designed for special purposes, which requires carefully precision in the caption and reproduction of sound. This can as an example be purposes of consulting with off-site experts, allowing them to get the full, most correct grasp of any situation. This can be in cases of medical examinations or simulation of industrial processes. In situations when life, health or big sums of money are at stake, the quality is paramount.

Directivity of microphone

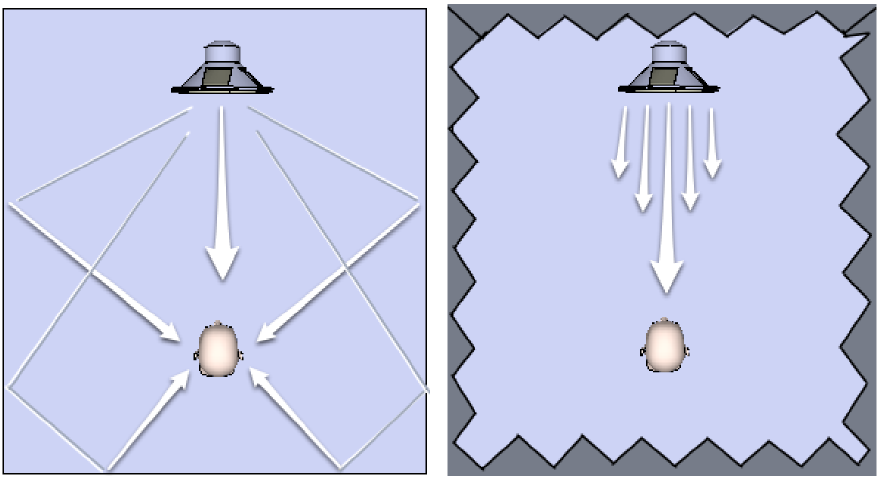
In a video conferencing situation, it is desired to have the microphone capture people speaking. Generally, any sound other than speech can be considered to be a disturbance. Hence, microphones with a directivity characteristic are often used. The cardioid microphone have a theoretic gain of 4.8 dB compared to an omni-directional. Beyond the difference in the direct sound, it is shown that the generated IRs of the omni- and cardioid directional microphone are quite similar at a distance of 1.5 m. Thus, the calculation of the early energy used for the speech clarity is not affected much by the directivity of the measurement microphone. Neither is the reverberation time, as the sound decay in a similar manner for both situations. However, by placing the cardioid microphone close to the talker it will give a bigger advantage. The value of this advantage must be considered towards what placement is practical and assures equal conditions for all participants in the conference connection.

5.5 The effect of recorded reverberation

In the interpretation of the acoustical parameters for combined rooms, certain questions rise. Can the results of the combined rooms be compared directly with the standard single room results? Is the same quality expected over a video connection as it is in a regular meeting room?

Consider two situations in two different rooms. One room is a reverberant room where sound is played to a listener through a loudspeaker, Figure 5.1(a). The other room is a very dry room where the reverberation from the first room is added to the sound and played to a listener through a loudspeaker, Figure 5.1(b). The reverberation time and sound will now basically be the same. Will these two situations sound equally good, or will the artificially added reverberation diminish the sound quality?

An article written by Gilford for BBC, [17], explains some of the aspects of this situation. Sound is recorded through a microphone on a single channel. This creates a monaural chain, providing only one channel between the speaker and listener. As regular hearing persons are accustomed to hear speech with two ears which together supply information about the direction and the position of the source, this



(a) Reverberant Room (b) Dry Room

Figure 5.1: Two situations of perceiving reverberation

creates a problem. The binaural human hearing provides an automatic mechanism for partially rejecting sound other than that coming from the direction of the source to which one is listening. By having only one channel of information and having both the direct speech signal and the corresponding reverberation coming from the same direction, the rejection mechanism will be inhibited. This will lead to a loss of speech intelligibility as one is conscious of the reverberant sound and the background noise to a larger extent. This can indicate that listening to an acoustically poor room in a good room will sound worse than the other way around. The bottom line is that if the transmitting room is perceived as having poor acoustical qualities, that room is certain to be perceived as so by the receiving room, regardless of the acoustical quality of the receiving room.

If there is a significant difference in perceived quality from the two different situations, the results of reverberation time in single rooms and in combined rooms should not be compared directly. In this case, the recommendation for the combined room values would be stricter in order to achieve the same conditions as a single-room situation.

This discussion enters the field of psychoacoustics. The results in this thesis have not considered any such difference in perceived reverberation. A straight forward comparison have been presented, which in the least give a good view of the objective differences in the two cases.

Chapter 6

Conclusion

The acoustic quality of a room with speech as the main usage have determined by parameters related to reverberance, background noise, speech clarity and speech intelligibility. These parameter have been measured in a selection of video conferencing rooms. Eleven rooms was chosen, and a special set of measurements were performed. The intension was to survey the current conditions in this type of rooms and investigate the acoustical properties when two rooms are combined. From the results, and specifications of limit values for noise and reverberation time given in the current building standard [10], recommendations for values for the combined parameters have been given.

There is a variety of rooms that are used for video conferencing. Some rooms are special designed for video conferencing as the sole usage, while other rooms are meant for multiple purposes, with video conferencing just being one of them. This spread is reflected by the results. Although there is a spread, nine of the eleven rooms are within the acceptable limit for reverberation time specified for rooms for educational purposes. This means that the rooms are found suitable for undisturbed communication and work that require concentration, in terms of the reverberation time. The results show a clear trend, indicating that a shorter reverberation time will give a higher speech clarity. Even though this relationship is clear, there are variations which expresses the need for the use of the speech clarity as an acoustical descriptor together with the reverberation time.

The STI is a good predictor of speech intelligibility and it focuses on giving a clear quality rating of the results. It is concluded the all rooms used for video conferencing should have a high value for the STI and be rated as "Excellent". The results indicate that the background noise does not have an impact on any of the other parameters. As the noise level is a very important factor in assessing the total sound quality and acoustic comfort of a room, it is very important to include this parameter in measurements and make improvements if necessary. The noise levels that have been measures is not regarded as being destructive of intelligibility and communication, but high levels are tiring and will over time give a decrease in concentration and efficiency.

Based on values specified in standards and the results obtained through experiments, parameter values for the early decay time and the speech clarity are suggested. For single rooms, the highest value for EDT is suggested to not exceed 0.6 seconds. A reasonable minimum limit for the C_{50} is suggested at 6 dB. The limits are bound to be of a different character for the combined rooms. First, by looking at the suggested categories for the single rooms, all rooms which are within the acceptable limits, combined with each other, should be an acceptable combination. This is stated in the hypothesis in section 2.7 and has been the working theory. As a result, the highest EDT value for combination of rooms is suggested to not exceed 0.8 seconds. A reasonable minimum limit for the C_{50} is suggested at 0 dB.

The EDT and C_{50} are chosen as descriptors for the combinations of rooms, and provide valuable acoustic information. However, the values are seemingly unaffected by both changes in background noise levels and microphone directivity. This indicates that even with similar EDT and C_{50}, the sound conditions in two rooms can be perceived as being different. Thus, in order to achieve a full overview of how the acoustic conditions in a room affect the quality of the video conference, parameters as background noise and also the properties of the audio equipment in the rooms should be considered.

Bibliography

[1] C. Svensson and E. Nilsson. *Optimum Room Acoustic ComfortTM (RACTM) can be achieved by using a selection of appropriate acoustic descriptors*, in Proceedings of Euronoise 2008, Acoustic 08 Paris.

[2] Erling Nilsson. *Room acoustic measures for classrooms*, Internoise 2010, 13-16 June 2010 Lisbon, Portugal.

[3] U. Peter Svensson, Hassan EL- Banna Zidan, Johan L. Nielsen. *Properties of convolved room impulse responses*, 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. October 16-19, 2011, New Paltz, NY.

[4] Gary W. Siebein, "Architectural acoustics," in AccessScience, McGraw-Hill Companies, 2008, http://www.accessscience.com

[5] ISO 3382-1:2009(E), Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces

[6] J.S. Bradley, R. Reich, S.G. Norcross. *A just noticeable difference in* C_{50} *for speech*, Applied Acoustics 58 (1999) 99-108

[7] http://en.wikipedia.org/wiki/Microphone. The page last modified on 29 May 2012 at 00:01.

[8] Cisco, *Guidelines for video conferencing room acoustics.* D14377.01. 08.2010

[9] ISO 354:2003, Acoustics – Measurement of sound absorption in a reverberation room.

[10] Norwegian Standard NS 8175:2008. Acoustic conditions in buildings – Sound Classification of various types of buildings.

[11] ISO 3382-2:2008, Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms

[12] http://www.variableakustik.eu/24167.html. Date: 14.05.2012

[13] ISO 18233:2006. Acoustics – Applications of new measurement methods in building and room acoustics.

[14] H.J.M. Steeneken and T. Houtgast. *A physical method for measuring speech-transmission quality*, Journal of the Acoustical Society of America 67(1), Jan 1980

[15] IEC 60268-16 Edition 4.0, 2011. Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index.

[16] Rod Elliott. *A-Weighting Filter For Audio Measurements.* http://sound.westhost.com/project17.htm

[17] C.L.S Gilford, British Broadcasting Corporation. *The Acoustic Design of Talks Studios and Listening Rooms*, Journal of the Audio Engineering Society, Vol. 27, No.1/2, 1979 January/February

[18] J.S. Bradley. *Relationships among Measures of Speech Intelligibility in Rooms*, Journal of the Audio Engineering Society, Vol. 46, No.5, 1998 May

[19] J.S. Bradley. *Predictors of Speech Intelligibility in Rooms*, Journal of the Acoustical Society of America, Vol. 80, No.3, September1986

Appendix A

Matlab-code

A.1 Loading IRs and calculation of single room values

```

%% Load Speaker and Listener IRs
clear all
close all

Sp = 3 ;                    % Number of speaker positions
Lis = 2;                    % Listening positions
Rooms = 11;                 % Total number of rooms
fs = 48000;                 % Sampling frequency
% Maximum reverberation time for all rooms:
T_R1 = 0.52;
T_R2 = 0.55;
T_R3 = 0.45;
T_R4 = 0.46;
T_R5 = 0.47;
T_R6 = 0.37;
T_R7 = 0.65;
T_R8 = 0.84;
T_R9 = 0.55;
T_R10= 0.55;
T_R11= 1.11;

for R = 1:Rooms

    eval(sprintf(['n_cut = T_R' num2str(R) '*fs*0.60;'])); % End sample value of
          IR windowing

    for i = 1:Sp
    % IRA "Speaker-to microphone"
            IRstr = strcat('Measurements/Room', num2str(R), '/IRA_R', num2str(R)
                 , '_Sp',num2str(i),'_M1.wmb');

```

```
            eval(sprintf(['Room' num2str(R) '_IRA' num2str(i) ' = loadimp(IRstr)
                ;']));
    %Windowing
            eval(sprintf(['[y x] = max(abs(Room' num2str(R) '_IRA' num2str(i) ')
                ) ;']));
            eval(sprintf(['Room' num2str(R) '_IRA' num2str(i) ' = Room' num2str(
                R) '_IRA' num2str(i) '(1:round(x+n_cut));']));

    % IRC Additional measuring points for reverberation time
            IRCstr = strcat('Measurements/Room', num2str(R), '/IRC_R', num2str(R
                ), '_Sp3_M',num2str(i+1),'.wmb');

            eval(sprintf(['Room' num2str(R) '_IRC' num2str(i) ' = loadimp(IRCstr
                );']));
    %%Windowing
            eval(sprintf(['[yc xc] = max(abs(Room' num2str(R) '_IRC' num2str(i)
                ')) ;']));
            eval(sprintf(['Room' num2str(R) '_IRC' num2str(i) ' = Room' num2str(
                R) '_IRC' num2str(i) '(1:round(xc+n_cut));']));

    % Calculating room acoustic parameters for single rooms
    IRname  = ['Room' num2str(R) '_IRA' num2str(i)];
    T30name = ['T30_Room' num2str(R) '_IRA' num2str(i)];
    T20name = ['T20_Room' num2str(R) '_IRA' num2str(i)];
    C50name = ['C50_Room' num2str(R) '_IRA' num2str(i)];
    Gname   = ['G_Room' num2str(R) '_IRA' num2str(i)];
    EDTname   = ['EDT_Room' num2str(R) '_IRA' num2str(i)];

    eval(sprintf(['[' T30name ',' T20name ',' EDTname ',' C50name ',' Gname ']
        = calcParam(' IRname ',' num2str(fs) ' );']));

    IRname_IRC  = ['Room' num2str(R) '_IRC' num2str(i)];
    T30name_IRC = ['T30_Room' num2str(R) '_IRC' num2str(i)];
    T20name_IRC = ['T20_Room' num2str(R) '_IRC' num2str(i)];
    EDTname_IRC   = ['EDT_Room' num2str(R) '_IRC' num2str(i)];

    eval(sprintf(['[' T30name_IRC ',' T20name_IRC ',' EDTname_IRC ']  =
        calcParam(' IRname_IRC ',' num2str(fs) ' );']));

    % Calculating mean value of parameters
        if i == 1
            % C50
            eval(sprintf(['C50_R' num2str(R) '_sum = C50_Room' num2str(R) '_IRA'
                 num2str(i) ';']));
            % G
            eval(sprintf(['G_R' num2str(R) '_sum = G_Room' num2str(R) '_IRA'
                num2str(i) ';']));

            % T30
            eval(sprintf(['T30_R' num2str(R) '_sumA = T30_Room' num2str(R) '_IRA
                ' num2str(i) ';']));
            eval(sprintf(['T30_R' num2str(R) '_sumC = T30_Room' num2str(R) '_IRC
                ' num2str(i) ';']));

            % T20
            eval(sprintf(['T20_R' num2str(R) '_sumA = T20_Room' num2str(R) '_IRA
                ' num2str(i) ';']));
            eval(sprintf(['T20_R' num2str(R) '_sumC = T20_Room' num2str(R) '_IRC
                ' num2str(i) ';']));

            % EDT
            eval(sprintf(['EDT_R' num2str(R) '_sumA = EDT_Room' num2str(R) '_IRA
                ' num2str(i) ';']));
```

```
            eval(sprintf(['EDT_R' num2str(R) '_sumC = EDT_Room' num2str(R) '_IRC' num2str(i) ';']));

        elseif i == 2
            % C50
            eval(sprintf(['C50_R' num2str(R) '_sum = C50_R' num2str(R) '_sum + C50_Room' num2str(R) '_IRA' num2str(i) ';']));
            % G
            eval(sprintf(['G_R' num2str(R) '_sum = G_R' num2str(R) '_sum + G_Room' num2str(R) '_IRA' num2str(i) ';']));

            % T30
            eval(sprintf(['T30_R' num2str(R) '_sumA = T30_R' num2str(R) '_sumA + T30_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['T30_R' num2str(R) '_sumC = T30_R' num2str(R) '_sumC + T30_Room' num2str(R) '_IRC' num2str(i) ';']));
            % T20
            eval(sprintf(['T20_R' num2str(R) '_sumA = T20_R' num2str(R) '_sumA + T20_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['T20_R' num2str(R) '_sumC = T20_R' num2str(R) '_sumC + T20_Room' num2str(R) '_IRC' num2str(i) ';']));
            % EDT
            eval(sprintf(['EDT_R' num2str(R) '_sumA = EDT_R' num2str(R) '_sumA + EDT_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['EDT_R' num2str(R) '_sumC = EDT_R' num2str(R) '_sumC + EDT_Room' num2str(R) '_IRC' num2str(i) ';']));

        elseif i == 3
            % C50
            eval(sprintf(['C50_R' num2str(R) '_sum = C50_R' num2str(R) '_sum + C50_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['C50_R' num2str(R) '_mean = C50_R' num2str(R) '_sum ./ ' num2str(i) ' ;']));
            % G
            eval(sprintf(['G_R' num2str(R) '_sum = G_R' num2str(R) '_sum + G_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['G_R' num2str(R) '_mean = G_R' num2str(R) '_sum ./ ' num2str(i) ' ;']));

            % T30
            eval(sprintf(['T30_R' num2str(R) '_sumA = T30_R' num2str(R) '_sumA + T30_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['T30_R' num2str(R) '_sumC = T30_R' num2str(R) '_sumC + T30_Room' num2str(R) '_IRC' num2str(i) ';']));

            eval(sprintf(['T30_R' num2str(R) '_mean = (T30_R' num2str(R) '_sumA + T30_R' num2str(R) '_sumC)./ ' num2str(2*i) ' ;']));

            % T20
            eval(sprintf(['T20_R' num2str(R) '_sumA = T20_R' num2str(R) '_sumA + T20_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['T20_R' num2str(R) '_sumC = T20_R' num2str(R) '_sumC + T20_Room' num2str(R) '_IRC' num2str(i) ';']));

            eval(sprintf(['T20_R' num2str(R) '_mean = (T20_R' num2str(R) '_sumA + T20_R' num2str(R) '_sumC)./ ' num2str(2*i) ' ;']));

            % EDT
            eval(sprintf(['EDT_R' num2str(R) '_sumA = EDT_R' num2str(R) '_sumA + EDT_Room' num2str(R) '_IRA' num2str(i) ';']));
            eval(sprintf(['EDT_R' num2str(R) '_sumC = EDT_R' num2str(R) '_sumC + EDT_Room' num2str(R) '_IRC' num2str(i) ';']));

```

```
            eval(sprintf(['EDT_R' num2str(R) '_mean = (EDT_R' num2str(R) '_sumA
                + EDT_R' num2str(R) '_sumC)./ ' num2str(2*i) ' ;']));
        end
    end
end
```

A.2 Convolution of IRs and calculation for combined rooms

```
%% Load Speaker and Listener IRs
% First: - Load all impulse responses

clear all
close all

Sp = 3 ;                % Number of speaker positions
Lis = 2;                % Listening positions
Rooms = 11;              % Total number of rooms
fs = 48000;              % Sampling frequency
%Maximum reverberation time for all rooms
T_R1 = 0.52;
T_R2 = 0.55;
T_R3 = 0.45;
T_R4 = 0.46;
T_R5 = 0.47;
T_R6 = 0.37;
T_R7 = 0.65;
T_R8 = 0.84;
T_R9 = 0.55;
T_R10= 0.55;
T_R11= 1.11;

for R = 1:Rooms

    eval(sprintf(['n_cut = T_R' num2str(R) '*fs*0.60;'])); % End sample value of
        IR windowing

    for i = 1:Sp
% IRA "Speaker-to microphone"
            IRstr = strcat('Measurements/Room', num2str(R), '/IRA_R', num2str(R)
                , '_Sp',num2str(i),'_M1.wmb');

            eval(sprintf(['Room' num2str(R) '_IRA' num2str(i) ' = loadimp(IRstr)
                ;']));
    %Windowing
            eval(sprintf(['[y x] = max(abs(Room' num2str(R) '_IRA' num2str(i) ')
                ) ;']));
            eval(sprintf(['Room' num2str(R) '_IRA' num2str(i) ' = Room' num2str(
                R) '_IRA' num2str(i) '(1:round(x+n_cut));']));

% IRC Additional measuring points for reverberation time
            IRCstr = strcat('Measurements/Room', num2str(R), '/IRC_R', num2str(R
                ), '_Sp3_M',num2str(i+1),'.wmb');

```

```
            eval(sprintf(['Room' num2str(R) '_IRC' num2str(i) ' = loadimp(IRCstr
                );']));
    %%Windowing
            eval(sprintf(['[yc xc] = max(abs(Room' num2str(R) '_IRC' num2str(i)
                ')) ;']));
            eval(sprintf(['Room' num2str(R) '_IRC' num2str(i) ' = Room' num2str(
                R) '_IRC' num2str(i) '(1:round(xc+n_cut));']));
    end

% IRB "Loudspeaker-to-listener"

%Left microphone channel
    for i = 1:Lis

            IRstr_L = strcat('Measurements/Room', num2str(R), '/IRB_R', num2str(
                R), '_Ls1_Lis',num2str(i),'_Ch1.wmb');

            eval(sprintf(['Room' num2str(R) '_IRB_L' num2str(i) ' = loadimp(
                IRstr_L);']));
        %%Windowing
            eval(sprintf(['[y_L x_L] = max(abs(Room' num2str(R) '_IRB_L' num2str
                (i) ')) ;']));
            eval(sprintf(['Room' num2str(R) '_IRB_L' num2str(i) ' = Room'
                num2str(R) '_IRB_L' num2str(i) '(x_L-100:round(x_L+n_cut));']))
                ;

    end

%Right microphone channel
    for i = 1:Lis

            IRstr_R = strcat('Measurements/Room',num2str(R), '/IRB_R', num2str(R
                ), '_Ls1_Lis',num2str(i),'_Ch2.wmb');

            eval(sprintf(['Room' num2str(R) '_IRB_R' num2str(i) ' = loadimp(
                IRstr_R);']));
        %Windowing
            eval(sprintf(['[y_R x_R] = max(abs(Room' num2str(R) '_IRB_R' num2str
                (i) ')) ;']));
            eval(sprintf(['Room' num2str(R) '_IRB_R' num2str(i) ' = Room'
                num2str(R) '_IRB_R' num2str(i) '(x_R-100:round(x_R+n_cut));']))
                ;
    end
end

%% Folding IRs
% Second: - Folding IRs of all combinations of Speaker and Listener positions
%         - Take mean value of acoustical parameters of multiple IRs

% Have to Fold one room at the time. That is, room 1 with all others, then
% room 2 with all others ect.

% Enter which room that is to be folded with all others:
Room = 1;

   for R2 = 1:Rooms
    b = 1; % Counter
       for i = 1:Sp

            for j = 1:Lis

                eval(sprintf(['Sp' num2str(i) '_Room' num2str(Room) '_Lis'
                    num2str(j) '_Room' num2str(R2) '_L = contwo(Room' num2str(
```

```
                    Room) '_IRA' num2str(i) ',Room' num2str(R2) '_IRB_L'
                    num2str(j) ');']));

                % Room acoustic calculations T20, T30, C50, G, STI
    IRname  = ['Sp' num2str(i) '_Room' num2str(Room) '_Lis' num2str(j) '_Room'
        num2str(R2) '_L'];
    T30name = ['T30_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b)];
    T20name = ['T20_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b)];
    C50name = ['C50_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b)];
    Gname   = ['G_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b)];
    EDTname = ['EDT_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b)];

                eval(sprintf(['[' T30name ',' T20name ',' EDTname ',' C50name ',
                     ' Gname ']  = calcParam(' IRname ',' num2str(fs) ' );']));
                if b == 1
                % C50 sum
                eval(sprintf(['C50_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     = C50_Room' num2str(Room) '_Room' num2str(R2) '_No'
                    num2str(b) ';']));
                % T20 sum
                eval(sprintf(['T20_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     = T20_Room' num2str(Room) '_Room' num2str(R2) '_No'
                    num2str(b) ';']));
                % T30 sum
                eval(sprintf(['T30_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     = T30_Room' num2str(Room) '_Room' num2str(R2) '_No'
                    num2str(b) ';']));
                % G sum
                eval(sprintf(['G_Room' num2str(Room) '_Room' num2str(R2) '_sum =
                      G_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b
                    ) ';']));
                % EDT sum
                eval(sprintf(['EDT_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     =  EDT_Room' num2str(Room) '_Room' num2str(R2) '_No'
                    num2str(b) ';']));

                else

                % C50 sum
                eval(sprintf(['C50_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     = C50_Room' num2str(Room) '_Room' num2str(R2) '_sum +
                    C50_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b
                    ) ';']));
                % T20 sum
                eval(sprintf(['T20_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     = T20_Room' num2str(Room) '_Room' num2str(R2) '_sum +
                    T20_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b
                    ) ';']));
                % T30 sum
                eval(sprintf(['T30_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     = T30_Room' num2str(Room) '_Room' num2str(R2) '_sum +
                    T30_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b
                    ) ';']));
                % G sum
                eval(sprintf(['G_Room' num2str(Room) '_Room' num2str(R2) '_sum =
                     G_Room' num2str(Room) '_Room' num2str(R2) '_sum + G_Room'
                    num2str(Room) '_Room' num2str(R2) '_No' num2str(b) ';']));
                % EDT sum
                eval(sprintf(['EDT_Room' num2str(Room) '_Room' num2str(R2) '_sum
                     = EDT_Room' num2str(Room) '_Room' num2str(R2) '_sum +
                    EDT_Room' num2str(Room) '_Room' num2str(R2) '_No' num2str(b
                    ) ';']));

```

```
                    end
                  b = b + 1;
                end
            end

            % C50 mean
             eval(sprintf(['C50_Room' num2str(Room) '_Room' num2str(R2) '_mean =
                  C50_Room' num2str(Room) '_Room' num2str(R2) '_sum ./' num2str(i*j)
                  ';']));
            % T20 mean
             eval(sprintf(['T20_Room' num2str(Room) '_Room' num2str(R2) '_mean =
                  T20_Room' num2str(Room) '_Room' num2str(R2) '_sum ./' num2str(i*j)
                  ';']));
            % T30 mean
             eval(sprintf(['T30_Room' num2str(Room) '_Room' num2str(R2) '_mean =
                  T30_Room' num2str(Room) '_Room' num2str(R2) '_sum ./' num2str(i*j)
                  ';']));
            % G mean
             eval(sprintf(['G_Room' num2str(Room) '_Room' num2str(R2) '_mean = G_Room
                  ' num2str(Room) '_Room' num2str(R2) '_sum ./' num2str(i*j) ';']));
            % EDT mean
             eval(sprintf(['EDT_Room' num2str(Room) '_Room' num2str(R2) '_mean =
                  EDT_Room' num2str(Room) '_Room' num2str(R2) '_sum ./' num2str(i*j)
                  ';']));

        end
```

A.3 The calcParam-function

```


function [T30, T20, EDT, C50, G] = calcParam(irIN,fs)

% Input: convolved IR of combinations of rooms
%              or IR of single room

% 20120418 PS Modification: scaled octave band filters to give the same
%             passband amplitude. Also mod. G calculation to compensate for
%             + 3 dB per octave.

% Split up early and late part of IR (For C50)

fs = 48000;
irSq = irIN.^2;
[npeak,ninit] = findinit(irSq,0.001);
n50 = ninit + round(0.050*fs);
irearly = irIN(ninit:n50);
irlate = irIN(n50+1:end);
irtot = irIN(ninit:end);

noctbands = 9; % Number of octave bands

Bbp = zeros(noctbands,96000);

% Filter coefficients

```

```
load Filter_coeff/Oct63Hz.txt; Bbp(1,:) = Oct63Hz(:,2);
load Filter_coeff/Oct125Hz.txt;Bbp(2,:) = 2*Oct125Hz(:,2);
load Filter_coeff/Oct250Hz.txt;Bbp(3,:) = 4*Oct250Hz(:,2);
load Filter_coeff/Oct500Hz.txt;Bbp(4,:) = 8*Oct500Hz(:,2);
load Filter_coeff/Oct1000Hz.txt;Bbp(5,:) = 16*Oct1000Hz(:,2);
load Filter_coeff/Oct2000Hz.txt;Bbp(6,:) = 32*Oct2000Hz(:,2);
load Filter_coeff/Oct4000Hz.txt;Bbp(7,:) = 64/1.0353*Oct4000Hz(:,2);
load Filter_coeff/Oct8000Hz.txt;Bbp(8,:) = 128/1.0353*Oct8000Hz(:,2);
load Filter_coeff/Oct16000Hz.txt;Bbp(9,:) = 256/1.5019*Oct16000Hz(:,2);

% Create empty vectors for parameter values
T30 = zeros(noctbands,1);
T20 = zeros(noctbands,1);
EDT = zeros(noctbands,1);
C50 = zeros(noctbands,1);
G = zeros(noctbands,1);

% Calculates the parameter value for each octave band. (63Hz-16kHz)
    for i = 1:noctbands

        irearlyf = contwo(irearly,Bbp(i,:)); %Bbp: filter coeffs
        irlatef = contwo(irlate,Bbp(i,:)); %Bbp: filter coeffs
        irtotf = contwo(irtot,Bbp(i,:)); %Bbp: filter coeffs

        ncut = round(length(irIN)*1.1);
        irearlyf = irearlyf(1:ncut);
        irlatef = irlatef(1:ncut);
        irtotf = irtotf(1:ncut);

       L =  backint(irtotf.^2);

    % Calc T30
        T30(i) = calT60auto(L,fs,-5,-35);
    % Calc T20
        T20(i) = calT60auto(L,fs,-5,-25);
    % Calc EDT
        EDT(i) = calT60auto(L,fs,0,-10);
    % Calc C50
        C50(i) = 10*log10(sum(irearlyf.^2)/sum(irlatef.^2));
    % Calc G
        G(i) = 10*log10(sum(irtotf.^2)) - 3*i;

    end
end
```

A.4 The contwo-function

Function made by Professor Peter Svensson

```
function iout = contwo(i1,i2);
% contwo.m convolves two real-valued time signal using FFT.
% They don't need to be of the same length, and the output
% will have the length = length(i1) + length(i2) - 1.
%
% Peter Svensson 970902
```

```
%
%   iout = contwo(i1,i2);

% 'Starting contwo!'
[n1,slask] = size(i1);
if n1 < slask,
   i1 = i1.';
   n1 = slask;
end
[n2,slask] = size(i2);
if n2 < slask,
   i2 = i2.';
   n2 = slask;
end

nfft = 2*2^(ceil(log(max([n1 n2]))/log(2)));
% disp(['Used fft size is ',int2str(nfft)])
iout = real(ifft( fft(i1,nfft).*fft(i2,nfft) ));
iout = iout(1:n1+n2-1);
```

A.4.1 The findinit-function

Function made by Professor Peter Svensson

```
function [npeak,ninit] = findinit(ir2,reltrig);
% findinit finds the direct sound of an impulse response.
% It gives the index with the peakvalue, "npeak"  and
% the first index before this with a value that is
% greater than the peakvalue*reltrig, "ninit".
% "reltrig" could typically be 0.001 or smaller, if the
% impulse response is measured and has some noise.
%
% Input parameters:
%   ir2     A squared impulse response
% reltrig   The relative trigger value that the direct
%           sound will detected by.
% Output parameters:
% npeak   The index value with the peak of the impulse response
% ninit   The index value with the direct sound
%
% Peter Svensson 020115 (svensson@iet.ntnu.no)
%
% [npeak,ninit] = findinit(ir2,reltrig);

if any(ir2<0),
  error('The squared impulse response contains values that are smaller than zero
      . Are you sure the impulse response is squared?')
end
peak = max(ir2);
trigvalue = peak*reltrig;
npeak = find(ir2 == peak);
initvect = find(ir2(1:npeak) > trigvalue);
%i = 1;
%while initvect(i+2) > initvect(i) + 2,
%     i = i+1;
%end
```

```
%ninit = initvect(i);
ninit = initvect(1);
```

A.4.2 The backint-function

Function made by Professor Peter Svensson

```
function Lout=backint(ir2,Trunccomp,Plots);
% backint backward integrates the squared impulse
% response and returns the level, in dB.
%
% Input parameters:
%   ir2      A squared impulse response (filtered or not)
% Trunccomp (optional) If Trunccomp = 'yes', then a
%           compensation for the truncated energy is made.
% Plots    (optional) If Plots = 'Plots' then results will
%        be plotted.
%
% Output parameter:
% Lout     The level of the result, i.e. 10*log10( x )
%        where x is the backwards integrated ir2. Lout
%        is normalized, i.e., it starts at 0 dB.
%
% Note that if the impulse response ends with zeros, these
% will be cut away before the backward integration.
%
% Uses the function linfit
%
% Peter Svensson 020115 (svensson@tele.ntnu.no)
%
% Lout = backint(ir2,Trunccomp);

if nargin == 2,
   if Trunccomp(1) == 'y' | Trunccomp(1) == 'Y',
      Trunccomp = 'yes';
      Plots = 'xxxxx';
   elseif Trunccomp(1) == 'p' | Trunccomp(1) == 'P',
      Trunccomp = 'no ';
      Plots = 'plots';
   else,
      Trunccomp = 'no ';
      Plots = 'xxxxx';
   end
elseif nargin == 3,
   if Trunccomp(1) == 'y' | Trunccomp(1) == 'Y',
      Trunccomp = 'yes';
   else,
      Trunccomp = 'no ';
   end
   if Plots(1) == 'p' | Plots(1) == 'P',
      Plots = 'plots'
   else,
      Plots = 'xxxxx';
   end
else,
   Trunccomp = 'no ';
```

```
   Plots = 'xxxxx';
end

%----------------------------------------------
% First cut away zeros at the end

lastofnonzeros = length(ir2);

if ir2(lastofnonzeros)==0,
   zvec = find(ir2==0);
   nz = length(zvec);
   if nz>1,
      vec2 = find(diff(zvec)~=1);
      n2 = length(vec2);
      if n2>0,
         ind2 = vec2(n2);
         lastofnonzeros = zvec(ind2+1)-1;
      else,
         lastofnonzeros = zvec(1)-1;
      end
   elseif nz == 1,
      lastofnonzeros = lastofnonzeros-1;
   end
   ir2 = ir2(1:lastofnonzeros);
end

%----------------------------------------------
% Now eventual zeros at the end are cut away but add a very
% small number for possible zeros in-between. The log function
% causes trouble otherwise, for truncation compensation on non-
% integrated data.

ir2 = ir2+1e-300;
l = length(ir2);

if Trunccomp == 'yes',
   disp(['     Compensating truncation'])
   sampint = 10;
   if l < 1000,
      sampint = 5;
   end
   ir2sam = ir2(10:sampint:l);
   l2 = length(ir2sam);
   ir2sam = filter(ones(20,1),1,ir2sam)/20;
   if Plots == 'plots',
      hold off; plot(log(ir2sam)); pause
   end
   [A,B,r2] = linfit(log(ir2sam),sampint);
   straightline = A+B*sampint*[0:length(ir2sam)-1];
   if Plots == 'plots',
      hold; plot(straightline,'g'); pause
   end
   restenergy = -exp(B*l+A)/B;
   Trunclev = 10*log10(restenergy/sum(ir2));
   disp(['     Truncated energy is',num2str(Trunclev),' dB'])
   clear ir2sam
   ir2 = [ir2;restenergy];
   l = l+1;
end
Lout = cumsum(ir2(l:-1:1));
clear ir2
Lout = Lout(l:-1:1);
Lout = 10*log10(Lout);
```

```
Lout = Lout-Lout(1);
hold off
%if Plots == 'plots',
%   plot(Lout)
%end
```

A.4.3 The calT60auto-function

Function made by Professor Peter Svensson

```

function [T60,r2T60,Lstart,Lstop] = calT60auto(Lin,fs,Lstart,Lend,plotoption);
%
%   This function makes a linear regression of the vector Lin between
%   two points specified by the user by two level values: Lstart and Lend.
%   In this range, the T60 and correlation
%   value r2T60 are calculated by the use of 'fs'. 'Lin' must contain
%   the level values of a backward integrated impulse response.
%
%   If plotoption = 1, then a plot will be shown.
%
% [T60,r2T60,Lstart,Lstop] = calT60auto(Lin,fs,Lstart, Lend);

% Mod 940203: if eg. -35 dB is out of the range, zeros are given as results
% Mod 960207: Plot the curve and let the user choose intergration points.
%             Plot also the fitted line
% Mod 960221: Print out numerical values
% Mod 960417: Print out that two points should be marked
% Mod 981104: Clarify text printout
% Mod. 091008: Make it automatic

if nargin < 5
    plotoption = 0;
end

deltat = 1/fs;

ivsegment = find(Lin<Lstart & Lin > Lend);
segment = Lin(ivsegment);
l = length(segment);

if l > 0,
   [A,B,r2T60] = linfit(segment,1/fs);
   T60 = -60/B;
   Lstart = segment(1);
   Lstop = segment(l);
    ivec = ivsegment;
    if plotoption == 1
    figure(1)
    plot([0:length(Lin)-1],Lin, ivec, A-B*ivsegment(1)/fs + B*ivec/fs,'g',
         ivsegment, A-B*ivsegment(1)/fs + B*ivsegment/fs,'r' )
  axis([0 length(Lin) -50 0])
   maxxval = length(ivec);
   text(maxxval*0.85,-5,[ 'T60 = ',num2str(T60),' s'])
   text(maxxval*0.85,-10,['r2 = ',num2str(r2T60)])
   text(maxxval*0.85,-15,['Levelrange = ',num2str(Lstart-Lstop),' dB'])
```

```
  text(length(Lin)/10,-45,'Press CR to continue')
   pause
    end
else,
   A = [0]; B = [0]; r2T60 = [0];
   Lstart = 0; Lstop = 0;
end
```

A.4.4 The linfit-function

Function made by Professor Peter Svensson

```
function [A,B,r2] = linfit(segment,deltax);
% linfit makes a line fit to a number of y-values that have equally
% spread x-values. The algorithm assumes  the first x-value to be 0.
%
% Input parameters:
% segment   A vector of y-values
% deltax    The step size along the x-axis (e.g. 1/fs)
%
% Output parameters:
% A,B      Line coefficients:   y = A + B*x
% r2       The squared correlation coefficient
%
% Peter Svensson 981112 (svensson@tele.ntnu.no)
%
% [A,B,r2] = linfit(segment,deltax);

% 971001 Took care of row oriented input vector.
% 981112 Small improvements to speed up calculations.

segment = segment(:);
l = length(segment);

sumx = (l-1)*l/2;
sumxx = (l-1)*l*(2*l-1)/6;
sumy = sum(segment);
sumxy = sum([0:l-1].'.*segment);
sumyy = sum(segment.^2);

B = (l*sumxy - sumx*sumy)/(l*sumxx - sumx*sumx);
A = (sumy-B*sumx)/l;
r2 = B*(l*sumxy-sumx*sumy)/(l*sumyy-sumy*sumy);
B = B/deltax;
```

A.4.5 The loadimp-function

The loadimp-function is used for importing impulse responses from WinMLS to Matlab and copyrighted by Morset Sound Development 1998-2003.

Appendix B

Miscellaneous

B.1 Level differences of dummy head

To examine the levels of right and left microphones, their levels are measured with a headset. The impulse response of the dummy head is found by using open headphones over the ears. A window of 1024 samples were applied to eliminate redundant information. By averaging the levels from 100-5000 Hz a value of difference between the ears can be found. This shows that there is a difference between the right and the left ear. The right ear have a higher level, with the difference of 2-2.5 dB.

Table B.1: Dummy head, levels

	Left headphone	**Right headphone**
Left microphone	94.2 dB	93.3 dB
Right microphone	96.3 dB	95.7 dB
Difference	2.1 dB	2.4 dB

All recorded impulse responses are stored, and will be freely available for use if interest rises and further work involving auralization is intended.

B.2 Measurement form

Measurement of room acoustic parameters

For Master Thesis at NTNU.
By Erlend Gundersen

1. **Location, Date**:

2. **Room details**:
   a. Take picture

   b. Size (L x W x H):

L [m]	W [m]	H [m]	V [m^3]

   c. Sketch (Shape and positions of interior):

   d. Interior
      i. Roof (Wood, metal, acoustically treated, etc.):

      ii. Floor (Wood, thick carpet, thin carpet, linoleum, etc.):

      iii. Walls (Wood, glass, concrete, curtains, absorbents, etc.):
      Wall 1:____
      Wall 2:____
      Wall 3:____
      Wall 4:____

3. **Measurements**

3.1 **Noise level measurement**

- NorSonic sound level meter. Perform L(eq) measurement, average 1 minute
Note: Have all video conferencing equipment turned on

3.2 A **Impulse response measurement «Speaker to microphone» (IRA)**

- Calibrate microphone

Sound card settings (D-audio):
- Phantom power: Left, (No software mixer)
- Output gain: -30.00 dB, Input gain: 48 dB

Equipment:
- Genelec loudspeaker + stand (Cables: Jack-jack 3-pointed, power cord)
- 1/2'' measurement mic (XLR-cable) and microphone stand
- Sound card w/usb-cable
- Calibrator

Choose positions to measure according to actual room setup

Height of loudspeaker:	

Name file: «IRA_R1_Sp1_M1» (for Room1, Speaker pos 1, Microphone pos 1)

Notes:

3.2 C **Additional impulse response measurements (IRC)**

Additional IR measurement for calculation of reverberation time.
Equipment: Same as for measuring IRA.
- Number of IRA and IRC measurements combined should be 6, with at least two different positions for the source.
  (Example: 3 mic positions for 2 source pos. **OR** 2 mic positions for 3 source pos.)
- Draw positions of source and receiver in the sketch.

Notes:

3.2 B **Impulse response measurement «Loudspeaker to listener» (IRB)**

Sound card settings (Daudio):
- Phantom power: Left and right channel
- Output gain: -30.00 dB, Input gain: 48 dB

Equipment:
- Genelec loudspeaker + stand (Cables: Jack-jack 3-pointed, power cord)
- Neumann dummy head + stand (Cables: 2x XLR-cables)

Height of loudspeaker:	
Height of dummy head	

Name file: «IRB_R1_Ls1_Lis1_Ch1» and «IRB_R1_Ls1_Lis1_Ch2» for left and right ear.
(for Room1, Loudspeaker pos 1, Listener pos 1 left/right)

Notes:

Sketches of measurement setup

IRA: **«Speaker-to-microphone»**

IRC: **«Additional IR for reverberation time»**

IRB: **«Loudspeaker-to-listener»**

B.3 Equipment list

WinMLS software
Loudspeaker - Genelec, Model 1029A
Microphone - Bruel og Kjaer, type 4165, 1/2” free-field
Dummy head - Neumann, Type KU 81i
Sound level meter - Norsonic 116(with preamplifier N-1201 and microphone Bruel and Kjaer, type 4190)
Sound card - D-audio USB Audio Reference Preamplifier
Calibrator - Bruel og Kjaer 1kHz, 94dB