SUBJECTIVE RATING AND OBJECTIVE EVALUATION OF THE ACOUSTIC AND INDOOR CLIMATE CONDITIONS IN VIDEO CONFERENCING ROOMS

Charlotte Hauervig-Jørgensen, Cheol-Ho Jeong and Jørn Toftum
Technical University of Denmark, Kongens Lyngby, Denmark,
email: charlotte.dk10@hotmail.com
Emine Celik Christensen
Acoustics and Noise, Rambøll, Copenhagen, Denmark

Today, face-to-face meetings are frequently replaced by video conferences in order to reduce costs and carbon footprint related to travels and to increase the company efficiency. Yet, complaints about the difficulty of understanding the speech of the participants in both rooms of the video conference occur. The aim of this study is to find out the main causes of difficulties in speech communication. Correlation studies between subjective perceptions were conducted through questionnaires and objective acoustic and indoor climate parameters related to video conferencing. Based on four single-room and three combined-room measurements, it was found that the traditional measure of speech, such as the speech transmission index, was not correlated with the subjective classifications. Thus, a correlation analysis was conducted as an attempt to find the hidden factors behind the subjective perceptions, revealing the speech intelligibility during video conferencing was highly correlated to EDT, D_{50}, and MTI in the 125 Hz frequency band.
Keywords: Video Conference, Speech perception, Room acoustics, Indoor Climate

1. Introduction

A video conference room allows its users to communicate face-to-face, using body language in addition to speech, without having to be physically present in the same room. Replacing face-to-face meetings with video conferences provides a great advantage for users throughout the world with a desire to be several places at once, and positive effects on companies with a need to connect to other offices or companies, both nationally and globally. With video conferencing, costs and carbon footprint associated with the travel required for an actual face-to-face meeting is thus reduced [1]. Hence, the ultimate goal for video conference rooms is *"to replace the traditional meeting room with two or more rooms in different locations, while keeping the feeling of attending the same meeting"* [2].

However, video conferencing poses many challenges since the audio quality is not only affected by one room, but also by two rooms: the source room and receiver room, whose roles change during the meeting, depending on who is speaking. The participants should be able to communicate clearly with their colleagues in the room and on the other side of the screen, while not having to worry about their speech not being intelligible. The video conference room thus has three roles to fulfill; **a meeting room, a recording room** and **a listening room** [3].

Typically, a meeting room is intended for both regular meetings as well as video conferences. Hence, the first role of the room is very important. Secondly, the room has to work as a recording room. The speech of the participants physically present in one room is captured by microphones, which are typically placed on the table or suspended from the ceiling. The microphone not only captures the speech of the participants, but captures the acoustic characteristics of the room along with it. Hence, when the sound is played through the loudspeakers in the connected room, the listeners hear the original speech (possibly with degraded quality due to the VC system) along with the acoustical characteristics of the recording room. Therefore, a video conference room also has to fulfill its final role of a listening room, enabling the listeners to understand the speech from the connected room clearly when played through the loudspeakers.

Thus, in order to obtain the optimal acoustic conditions for a video conference, the room intended for video conferencing must fulfil all three roles at once, making video conference acoustics a very special complicated case.

1.1 Previous work

Svensson et. al. investigated what it takes to create optimal sound conditions in a video conference room [3]. In 2012, new additions and alterations were made to the Norwegian standard, NS8175. Those values will be used in the analysis and interpretation of the results of the measurements. In Denmark, the Building Regulation [2010] has no recommended values for video conference rooms.

In this study, the ideal values of acoustic parameters in video conference rooms specified in NS8175:2012 were compared to the measured values of the chosen video conference rooms at Rambøll. In addition, the measured values were compared to the findings of [4]. Gundersen suggests that for single rooms, the EDT should not exceed 0.6 seconds and the clarity level should be at least 6 dB. For combined rooms, the EDT should not exceed 0.8 seconds and the clarity level should be at least 10 dB.

1.2 Challenges

When evaluating the acoustics of any type of room, challenges will occur since objective measurements not always correspond to subjective rating. Thus, a combination of subjective and objective investigations is the only way to give a full evaluation of the acoustics, along with selecting the most convenient parameters to be evaluated.

However, there is an extra factor which may degrade the correlation between objective and subjective evaluations for video conference rooms: the counterpart's room. For example, if the counterpart's room is acoustically bad, this may significantly degrade the subjective rating of the VC room, although the VC room is actually in a good acoustic condition.

The subjective questionnaire survey was performed with participants using the video conference rooms at Rambøll, but they were calling other offices and companies – not their own rooms. Hence, if the connected rooms at another office or company had poor acoustic conditions, the room at Rambøll is unable to perform as intended. If this is the case, a *"Good"* room may be evaluated as a *"Bad"* room, due to the acoustics of the connected room. Then the objective measurements of a room subjectively classified as *"Bad"* might prove to be objectively "*Good"*.

Using the obtained subjective ratings and room acoustic measures, a correlation analysis was performed in order to gain an idea of which parameters to focus on when designing the acoustic and indoor climate conditions for video conference rooms.

2. Method

2.1 Questionnaire survey

A one paper questionnaire survey with questions regarding speech intelligibility, noise, temperature, lighting and air quality was made, with the target population being the users of the video conference rooms at Rambøll. The survey questions were carefully written with the respondent's perspective in mind, in order to avoid confusion. Each question was evaluated on a 5-point scale, where 1 was "Extremely satisfied", indicated by a happy smiley, 5 was "Not at all satisfied", indicated by a sad smiley, and the numbers in-between were levels of those. Nevertheless, all questions included a response category of "*Do not know*". The responses in the *"Do not know"* category were indicated by the number *"6"* and sorted out before analyzation. The results were then sorted by room number, and the median value of each question was calculated in order to classify the room as *Good* or *Bad* in the question concerned. The classification was then evaluated and used in order to select 2 *Good Rooms* and 2 *Bad Rooms* for objective acoustic and indoor climate measurements.

2.2 Indoor climate

To measure the indoor climate conditions in the four video conference rooms chosen, the following five parameters were measured; CO_2-concentration, Air temperature, Relative air humidity, Particle number concentration and Lighting conditions, using a Vaisala transmitter connected with a HOBO UX120-006M data logger, a TSI P-Trak 8525 and a Lux-meter. The data retrieved from the measurements were processed in order to compare and evaluate the results against the recommended values of each parameter stated in the Danish Building Regulation [BR15], the Danish Working Authority A.1.2 [2008], [DS474 E [1995] and EN12464:1 [2011].

2.3 Acoustics

Impulse response measurements according to ISO 3382 have been conducted *in* the four video conference rooms, *between* a combination of those and of the video conference *system* itself. EDT, T20, and C50 are of main interest in this study.

For the *in-room* measurements, hereafter referred to as single room measurements, the Dirac software has been used with an internal exponential sine sweep (length 2.73 s and 6-8 repetitions), played over an omnidirectional loudspeaker. A microphone (Type 4192) and a NEXUS conditioning amplifier were used. In a pair of *Good Room – Good Room* and *Bad Room – Bad Room*, one of the rooms were acting as the source room, and one of them as the receiver room shown in Figure 1.

For the impulse response measurements *between* rooms, hereafter referred to as combined room measurements, an external exponential sweep signal covering 60 Hz to 6000 Hz were played over an omnidirectional loudspeaker in the source room while the video conference system was turned on, and recorded by an omnidirectional microphone connected to a hard drive recorder in the receiver room. Their impulse responses are processed via a deconvolution technique [Angelo farinas paper].

To measure the acoustic quality of the video conference system itself, the ideal scenario would have been to bring the system to an anechoic chamber for separate measurements, which would have perfectly eliminated the room acoustic influence on the impulse response. Unfortunately, it was not possible to do so. Therefore, a directive loudspeaker placed at 3 cm from the video conference table microphone generates an exponential sweep and an omnidirectional microphone was placed 3 cm from the loudspeaker in the receiver room for recording.

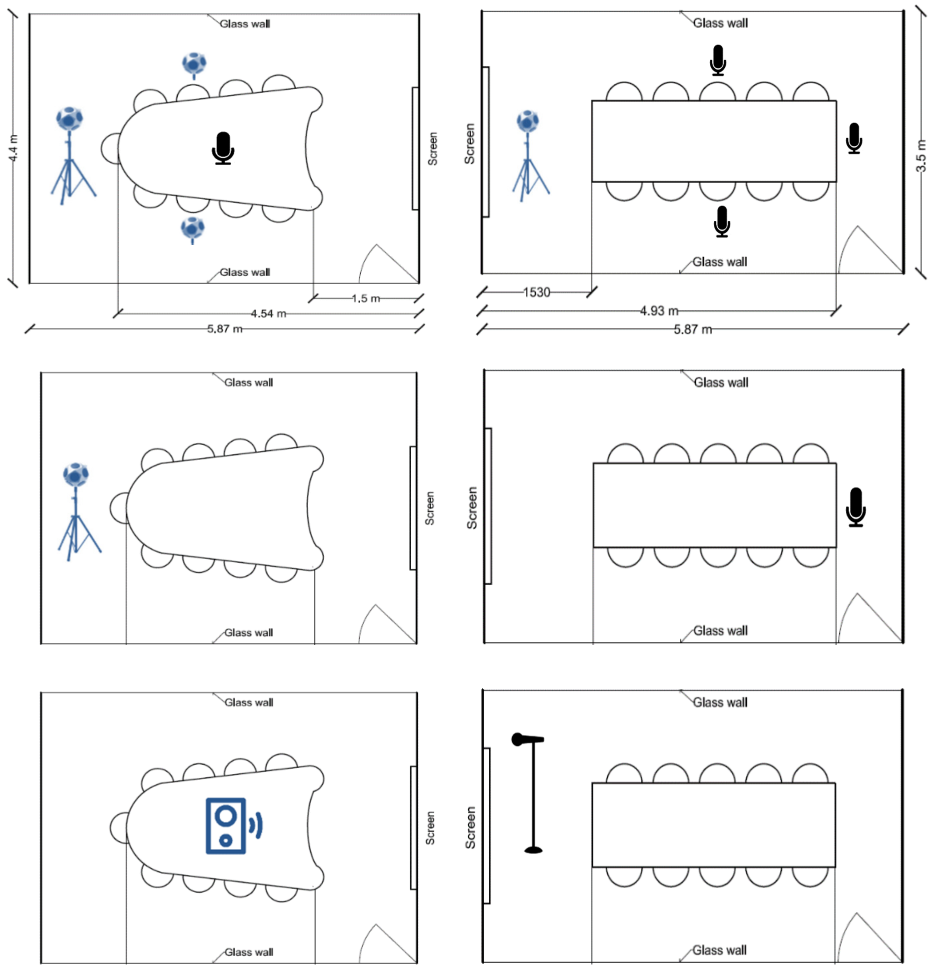

Figure 1: Measurement setup for *single rooms*, *combined rooms* and the *system*

3. Results

3.1 One paper questionnaire

The respondents seemed to be satisfied with the speech intelligibility when having a conversation with their colleagues *in* the video conference rooms. However, the results indicated that the respondents were less satisfied with the speech intelligibility while speaking with their colleagues *through* the video conference system, as expected. Important subjective scores are found in Fig. 6 [regression figure]. More information can be found in [5].

Even though it is unknown which type of room the respondents have spoken to while conducting the survey, and whether these rooms have affected the subjective evaluation of the rooms at Rambøll, a selection of video conference rooms that differ slightly in size and position were chosen for objective investigations based on their ratings in the questionnaire results. The subjectively rated *Bad Rooms* and *Good Rooms* are referred to as B1, B2, G1 and G2 respectively.

3.2 Indoor climate

The levels of CO_2-concentration were acceptable in all rooms, as well as the particle number concentration. Due to the equipment load and high occupancy frequency, the temperature conditions in all rooms were unacceptable. Some rooms provided necessary lighting, while other rooms had deficient lighting.

3.3 Acoustics – Single rooms

Figure 2 below present a comparison of EDT and C_{50} between rooms respectively, with the red line representing the recommended values in [4]. Note that for single rooms, the EDT should not exceed 0.6 seconds and the clarity level should be at least 6 dB. Figure 2 indicates that all rooms met the recommended values stated by Gundersen, except for B2 in the 125 Hz and 250 Hz band. However, an unexpected tendency was the one of room B1, which had the shortest EDT in Figure 2 and thus the highest clarity level, despite of its subjective classification as Bad. Likewise, room G1 does not appear to act as expected according to its subjective classification as Good, due to its clarity level being lower than the level of room B1. In terms of reverberation time, it is found that all rooms exceeded the recommended reverberation times of both class C (0.52 s) and D (0.42 s) in NS8175, with a few exceptions in the mid and high frequencies in room G2 in Figure 3. Hence, there seems to be no clear correlation between the subjective classifications and the objective parameters EDT, C_{50} and T_{20}, as well as STI.

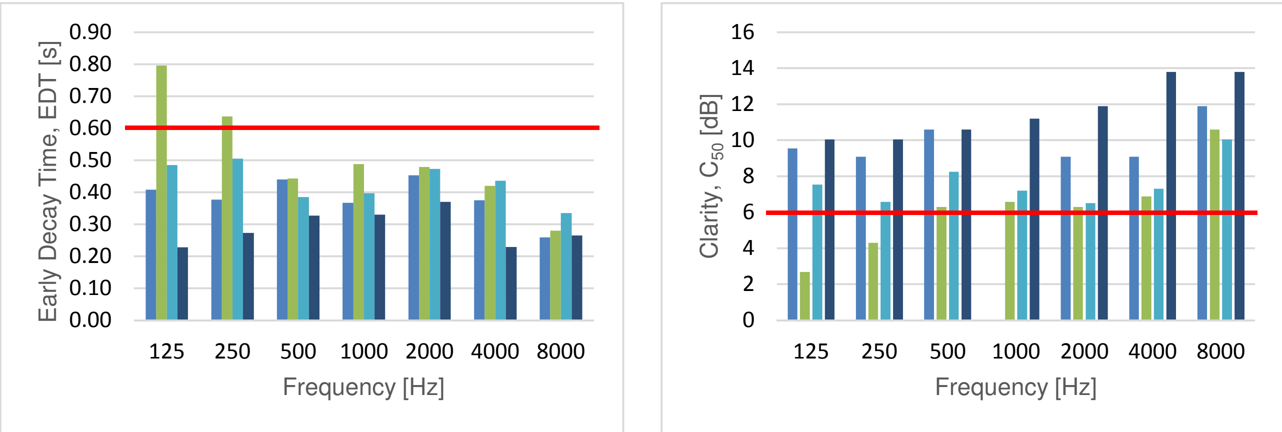

Figure 2: Early decay time and clarity in single rooms. B1 (1st blue), B2 (green), G1 (2nd blue), G2 (3rd blue).

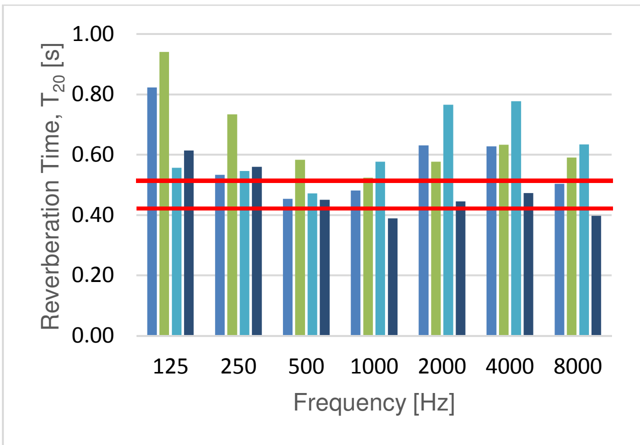

Figure 3: Reverberation time variation between single rooms. The same colour code as in Fig 2.

3.4 Acoustics – Combined rooms

Figure 4 present a comparison of EDT and C_{50} between rooms respectively, with the red line representing the recommended values in [4].

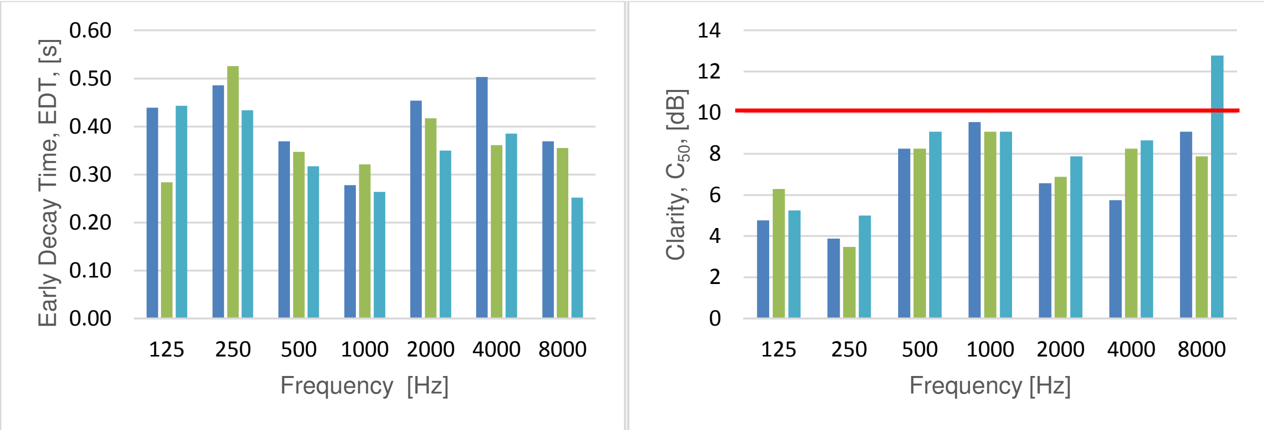

Figure 4: EDT and clarity between combined rooms. B1-B2 (1st blue), G1-G2 (green), B1-G1 (light blue).

Figure 4 shows that the EDT did not exceed the recommended value in any of the combinations in any frequency band. In terms of clarity in Figure 4, the recommended clarity level was only met in the 8 kHz frequency band in the *Bad – Good* combination. Again, there seems to be no clear correlation between the subjective ratings of the single rooms and the objective measurements.

However, there appears to be a clear tendency in the 125 Hz frequency band, where the subjective ratings of the single rooms actually do correspond to the objective results of their combinations. This tendency also appears in Figure 5 below of the modulation transfer indices (MTI), where the MTI is highest in the *Good– Good* combination and lowest in the *Bad – Bad* combination. Yet, this tendency disappears in the reverberation time variation, as seen in Figure 5. Thus, the reverberation time did not appear to be useful in evaluating the acoustic quality of speech in a VC room.

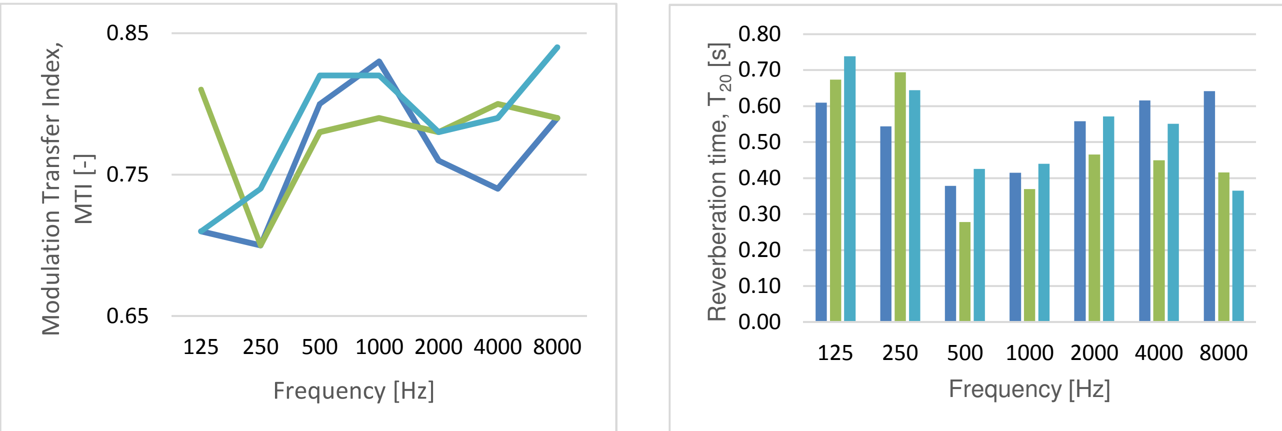

Figure 5: MTI and T20 between combined rooms. B1-B2 (1st blue), G1-G2 (green), B1-G1 (light blue).

4. Correlation Analysis

The Pearson correlation coefficient was computed in MATLAB, using subjective ratings and objective measures. A p-value of 0.05 was selected to indicate a significant correlation.

A correlation analysis showed that "speech in room" was correlated with $T_{20,250\text{ Hz}}$ and $T_{20,500\text{ Hz}}$, and $D_{50,125\text{ Hz}}$ at a significant level of 5%. "Speech in room" was also correlated with $EDT_{125\text{ Hz}}$, $C_{50,125\text{ Hz}}$, $MTI_{125\text{ Hz}}$, $MTI_{500\text{ Hz}}$ at a significant level of 10%. Figure 6 shows two regression lines between the subjective rating of speech in rooms and C_{50} and MTI in the 125 Hz band.

The "general acoustics" rating is significantly correlated with the background noise at 5% significant level and $T_{20,125\text{ Hz}}$ at 10% significant level.

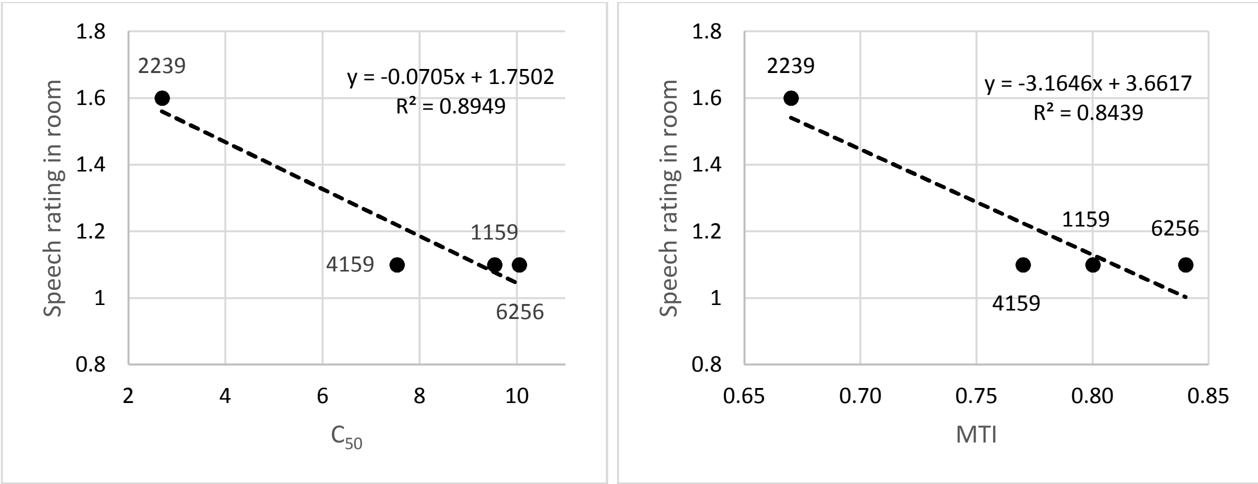

Figure 6: Speech in rooms vs. C_{50} and MTI_{125Hz}

5. Conclusion

This study tries to correlate the subjective ratings of chosen VC rooms with the traditional acoustic measures, EDT, T_{20}, D_{50}, and MTI. It was concluded that the speech in room is highly correlated with EDT, D_{50}, and MTI, and the general rating is highly correlated to the background noise level and T_{20}. The subjective rating of speech intelligibility is quite well explained in the MTI in the 125 Hz band, which gives an indication of the fact that the low frequency modulation characteristic of the room is likely to affect the speech perception during the VC communication. STI is not correlated with any of the subjective ratings, and it is therefore most likely that STI is not a good parameter to capture acoustic problems in VC rooms.

Currently, the subjective rating is collected without knowing the counterpart's room acoustic condition, which prevents us from drawing a clear conclusion out of the study. In the future, controlled systematic room acoustic simulations or experiments including both rooms and the VC system itself, together with listening tests, are needed to draw more clear conclusions of the interaction between the two rooms and the effects of the VC system on the speech communication.

REFERENCES

1 Coudriet, G.A. & Babich, J.E. (2011): *"Effective Design of Audio/Video Conference Rooms"*, a paper presented at Noise-Con July 2011, INCE 25th Annual Conference, Portland, OR.

2 Berg, T., Svensson, U.P., Taraldsen, G. & Tronstad, T.V. (2012): *"Influence of microphone directivity and room acoustical factors on the perceived speech quality in audio conferencing"*, a paper presented at EURONOISE 2012, Prague, European Acoustics Association.

3 Torheim, N.K & Lie, E. (2012): *"Video conferencing with improved sound"*, Article from The Research Council of Norway, Science Nordic, February 24 2012.

4 Gundersen, E. I. (2012): "Room Acoustic Conditions for Audio and Video Conferencing", Master Thesis from NTNU-Trondheim, Norwegian University of Science and Technology

5 Hauervig-J., C. (2016): "Subjective and Objective Evaluation of the Acoustic and Indoor climate Conditions in Video Conference Rooms", Bachelor Thesis from DTU, Technical University of Denmark