Influence of audio on perception and comprehension of video sequences

A SUBJECTIVE TEST OF PERCEPTION OF AUDIOVISUAL CONTENT WHERE AUDIO QUALITY WAS CHANGED

SARA LÅNGVIK

KTH ROYAL INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE AND COMMUNICATION (CSC)

Influence of audio on perception and comprehension of video sequences

A subjective test of perception of audiovisual content where audio quality was changed

Ljudets påverkan på uppfattning av video sekvenser

Ett subjektivt test av uppfattning av audiovisuellt material där audiokvalitén ändrades

Sara Långvik
830917-6389
s.langvik@kth.se

Abstract

The importance of the role audio plays in multimodal information media is often underestimated or even overlooked. This applies to a wide range of media in everyday use- from videoconferencing systems to multi modal human-computer interfaces and to some extent even computer games. When improving audiovisual (AV) systems, the focus is generally on video. A question seldom asked, most likely because it is of a difficult nature, is how we actually perceive audiovisual media.

In this study the importance of audio was examined by exposing subjects to audiovisual video clips that varied only in audio quality. In some videos background noise from a subway train was added and was to be regarded as an external signal. The subjects were asked to rate their perception of the audio, video, audiovisual quality as well as the listening effort by rating the comprehensibility of the content. They were also required to pay attention to the content in each video and to answer questions about it.

Results show a perceived difference in video quality following the perception of audio quality, although the video quality was never altered. Results also show that a light reverb infused in the audio signal was considered to have a positive impact on comprehensibility while still being perceived as decreasing the audio quality.

The thesis has been written in collaboration with telecomunication company Ericsson and all tests and material have been recorded at Ericsson laboratories in Kista, Stockholm.

Sammanfattning

Vikten av audio inom multimodal informationsmedia är ofta underskattad. Detta gäller media som videokonferenssystem, interaktiva datainterface och i någon mån även dataspel. När förbättringar inom audiovisuella system görs, satsas det ofta mer på den visuella delen, t.ex. genom en större videoskärm eller effektivare video codec. En fråga som sällan ställs, säkerligen för att den är svår, är hur vi egentligen uppfattar media.

Denna studie ämnar att påvisa vikten av audio genom ett subjektivt test där testpersonerna betygsatte sin kvalitetsuppfattning av audio, video, audiovisuella (AV), hur lätt eller svårt dom fann det att uppfatta innehållet i en serie korta videoklipp, där endast audiokvalitén ändrats. Även bakgrundsljud från tunnelbana fanns med i ljudsignalen för vissa video- klipp och var ämnat att vara betraktat som extern audiosignal (miljö). Testpersonerna ombads betygsätta sin uppfattning av audiokvalité, videokvalité, total audiovisuell kvalité och hur lätt eller svårt det var att uppfatta innehållet i videon. Utöver det blev de ombedda att svara på några frågor angående innehållet i videon.

Resultaten visar att den uppfattade videokvalitén följer uppfattningen om audiokvalité även om videokvalitén förblev oförändrad genom hela testet. Videoklipp som innehöll ljud med tillagd efterklang blev betygssatta som sämre i audiokvalité även om uppfattningen blev betygssatt som bättre.

Detta examensarbete har utförts i samarbete med telekommunikationsföretaget Ericsson. Själva studien och det tillhörande råmaterialet har blivit inspelat och tagit plats på Ericsson laboratories i Kista, Stockholm.

Table of contents

1 Foreword

The idea of the topic of this thesis work is a result of a serious interest in audio on a technical, psycho acoustical and psychological level. The consumption of audiovisual media has increased drastically in the past decades and is increasingly consumed "on the go", meaning in public and between tasks (such as home and work), on the train, on the street and using small devices such as telephones and pads. How this effects our perception of media and even our comprehension of media content becomes thus increasingly important.

It is, in my opinion, easy to find niched and specified information on specific aspects and factors (linked to audiotechnical parameters) within both visual as well as audio media. The challenge remains in finding relevant information on human perception of these factors in an everyday environment. This is expected as it introduces many different aspects that intersect and overlap, which may influence the total experience. Tests and studies taking that many aspects and factors into account are very hard to execute because they quickly reach too great an amount of data to keep track of.

My interest in audio in this thesis is based on my experience of video often being assigned a more important role in media than audio. Examples of this can be found in videoconferencing situations, where improvements often mean a bigger screen and yet audio remains low in quality. Another example is the constant focus on visual effects and visual programming in user interfaces and reviews of games (PC, console and applications). This seems to have lead to a general assumption that audio does not matter in the same way as video. With this thesis I aim to further investigate whether this is a misconception.

This has to my fortune also been a topic of interest at telecommunication company Ericsson with whom this thesis has been put together. The thesis also serves as further research on subjective

2 Background and Previous work

In this chapter earlier work that relates to this study is summarized to provide a deeper under-standing of the measures that have been taken to conduct the study. It also reveals some of the earlier hypotheses, theories and conclusions that have helped in building this study. There are many factors and aspects that have needed to be taken into consideration throughout the study as it is one of a fundamental kind- based and built on the foundation of many a hypothesis. Lastly, there's a short conclusion of previous work in regards to this thesis.

The perceived quality of audiovisual media in videoconferencing systems as well as consumer electronics is an ongoing topic that companies within this sector continuously strive to improve. As new codecs, distribution schemes and applications are being developed, it seems that focus mainly lies within the perception of only visual information (Winkler, 2005). According to Storms and Zyda (2001) the combination of audio and video, however, plays a part in the overall perception of an audiovisual service. This applies in particular to the perceived pleasantness, intelligibility and comprehensibility of such a service, according to Oyda, Czyzewski and Kostek (2001). The role that audio has on overall perception of audiovisual quality has also been found to differ between different interactive contexts such as task based and non task based contexts. This was stated by Susini et al. (2012) while Borowiak et al. (2014) pointed out the importance of audio cues for the same purpose. Further studies on the matter of perception during multi-modal tasks have been conducted by Rimell et al. (2008) (see chapter 2.2) .

2.1 Linking Audio and Video with regards to Perception

Beerends and De Caluwe (1999) established links between certain factors within the auditory and the visual modality when it comes to overall perception of audiovisual (AV) media by studying certain changes within both video and audio that are harder to notice than others when produced simultaneously. The subjects were introduced to about 25 seconds long audiovisual commercials where the audio and/or video channel was changed. Their task was to determine their perception of audiovisual quality or only audio or video quality in the media. The changes included filtering in the frequency spectrum for the audio channel and luminance changes such as lightness for the video channel. The results showed that even though audio did in fact influence the perception of video quality, the influence of video on audio quality was higher. As the study focused on whether an overall change in video and audio was detected when displayed simultaneously, it disregarded a deeper analysis of the results. The results also applied to a change in both the visual and auditory channel and does thus not allow for conclusions to be drawn on the contributions of each channel, separately.

2.2 The influence of content and task on perception of multimedia quality

Examining the perceived quality of the two modalities (visual and hearing) separately and together has been found to be influenced by the interaction of one another. Another influence that plays an important role in quality rating is task- based performances when the task is more related to one of the two modalities. When focusing on one modality, for example audio, the perceived quality of the other modality (video) seems to follow the quality of the audio, regardless of the video's true quality (Rimell et al. 2008). Through findings like these, it is clear that the two modalities influence each other and that they both play an important role in the total perception of audiovisual media, although video is generally perceived as superior to audio.

2.3 Altering perception of Audio through Video

Further studies within the role of correlation between audio and video have mostly pursued answers to whether video quality can alter the perception of audio. In a study by M.P.Hollier, R.Voelcker (1997), various degradations were added to a video signal to measure the perception of audio within the general perception of the AV experience. Also a number of audio degradations were added although they were fewer in quantity than the amount of different video degradations. The degradation in video included two levels of: Blurring, White noise and flickering at the edges. The audio variations included: No degradation, Band filtering and Modulated Noise Reference, which is speech modulating a combination of input speech and noise, mainly used in the telecommunication industry (ITU P.810). As this was more of an overall study investigating whether video quality affects the perception of audio quality, the findings were that there are drastic differences in the perceived quality. The differences are also very drastic between small changes such as in what order the questions were asked and whether only the audio channel was activated.

2.4 Videoconferencing and Interactive media studies

B.Belmudez et al. (2009) found the separate audio and video channel impact to be of varying importance to the perceived quality in videoconferencing systems based on the focus of attention each channel requires. Many studies conducted within subjective testing of perceived quality have incorporated degradation of audio and/ or video using codecs and compression rates as parameters for change (factors of influence). The study by B.Belmudez et al. (2009) showed that subjects are less sensitive to changes in audio when their attention is dedicated or partly dedicated to performing a task. However, the perception of audio does vary based on the amount of focus on solely the audio or video channel.

2.5 Intelligibility and attention

Based on previous research by K.S.Rhebergen et al. (2008) on the role of real life background noise on the intelligibility of speech, Wong et al. (2012) found that the intelligibility of speech in noise is affected by the presence of linguistic information, the level of low frequency noise and amplitude fluctuations. The study was conducted using recorded noise from real life situations; background noise from a café, street and lower deck of a double decker bus. Results from the study indicate less intelligibility of speech for background noise with linguistic information than static noise. However, the differences are far less dramatic when the overall level of noise is dropped. This is somewhat contradictory to the Lombard effect and Lombard speech, which among other things points out the unintentional focus on an audio channel when noise is present (See the Lombard Effect, page 30).

2.6 Conclusion of previous studies in regards to this thesis

Not many studies have been conducted solely devoted to investigating the implications of audio in the overall perception of video quality. This may be because of the various possibilities of content and mediated information of a video that can range from generic music to information given by multiple speakers. As mentioned before, the level of interaction and concentration also plays a role in perception (Rimell et al. 2008) of AV- i.e. are we listening to a speaker or are we participating in a discussion?

There are several aspects to audio perception such as the previously mentioned, objectively measurable codecs, digital compression and streaming rates. However, factors connected to human subjective perception, and measurable for example with psychoacoustics methods, are rarely considered. For example, natural/realistic stimuli are rarely used in perceptual tests for the evaluation of quality of service (QoS) or quality of experience (QoE), while artificial test tone and white noise are often used.

A number of studies and research on psychoacoustic phenomena as well as perception of quality in multimedia have been conducted, although not many of them have been merged for the purpose of studying the effects they may have on one another. As the use of multimedia grows and evolves at high speed, an understanding of the effects and perception of multimedia is vital. Because of the many factors that connect and overlap each other from neurology to psychology and cognition as well as culture, it is important to continuously study the effects of multimedia as it evolves. After all, we evolve alongside multimedia.

Since the question "Can the perception of video quality be altered by changing the quality of the appurtenant audio signal?" hasn't been answered in previous work, this thesis needs to focus on answering that question before moving on to possible follow-up questions.

3 Aim

The aim of the study is, on a general level, to investigate whether audio effects the perceived quality of video. The question of the thesis is thus: ***"Can the perception of video in audio visual media be changed by changing the audio signal?"***.

Furthermore the study aims to (with the assumption that the answer to the first question is affirmative) answer follow up questions such as ***"How does the perception of video change?"***. Since small differences in the presentation of video, audio, the content of video, length of video and the order of questions asked can effect perception, this subjective test has taken into consideration many of the possible influences these changes may have, as well as the (parameter) changes tested. The factors (parameters) focused on in this thesis have been carefully chosen based on everyday usage of communication media today.

Another influence that has been looked at is the importance of comprehensibility. However, although the study does investigate this measuring the results, it does not dig deeper into the cognitive aspects of attention, learning and memory. Nor does it focus on intelligibility of speech.

4 Method

In this chapter the procedure of the test is explained. Preparations for the test and how the test was conducted is also explained. As many factors and aspects have been taken into account, a little about each one is also presented.

Abbreviations, specifications and terminology is explained throughout the chapter. The specification parts that do not explain the actual procedure, but rather explain measures taken in the procedure are marked with *. The choice of method was qualitative as the data collected was subjective feedback by voting. Even though statistics have calculated, their purpose is not to prove an already existing hypothesis and thus the method isn't entirely quantitative. Because of the large number of factors and aspects needed to be taken into consideration, the possibilities of using past research in the field as reference points to build hypotheses on, were limited.

4.1 Walkthrough of the whole test

27 participants (4 women, 23 men) of ages 20-65 were instructed to watch two sequences of audiovisual clips and rate their level of comprehension of the content as well as their perception of the video quality, audio quality and audiovisual quality. They were also requested to answer a few questions about the content of each clip. All of the subjects were invited via an internal e-mail list used for recruiting test subjects at Ericsson and were given two cinema tickets as a reward for their participation. The two sequences consisted of 12 video clips each. There was a 10 minute break between the two sequences. The duration of each video clip was about 45-50 seconds, after which participants were given 20 seconds to answer two questions about the content and 20 seconds to rate their perceived quality of the previous video. All participants reported having normal hearing and did not suffer from conditions such as colour blindness. Before the test sequence started, the participants were shown three reference videos in order to get used to and familiarise themselves with how the test was to be conducted.The order of video clips watched was randomized in a controlled way, so that all participants were introduced to all the audio degradations.

The method chosen for the study was a compilation of subjective testing and semi struc-

tured interviews. A combination of the two was judged to give the best and most reliable results as the test focused on subjective perception, which is difficult to measure in only numbers.

4.2 Content

Since both perceived quality and comprehension was to be looked at, the content needed to include spoken information in a situation that the subjects could relate to. The footage in the videos included four single speakers (two women and two men) reading six manuscripts divided in two or three parts- 14 sections in total. Four single speakers (two women and two men) speaking freely about a subject of their choice while often introducing an item connected to the subject, such as a coffee cup. To minimize the possibility of recognition and previous knowledge of the topics, the stories delivered by the speakers were all made up or personal. The gender balance in all videos was half and half. The manuscript videos were read by an equal number of women and men and the free improvisation videos were presented by an equal number of women and men. The videos were 24 in total (in spoken content) and separated into two groups dependent on the content: Manuscript videos and Free improvisation videos.

Manuscript

Four different speakers (two women and two men) were chosen to read a manuscript about various topics. The topics spoken about included:

Subject (video) 1: General information (update) on issues in a housing system   **-2 parts**

Subject (video) 2: Advertisement information on a new mobile phone   **-2 parts**

Subject (video) 3: A story about a walk in the woods   **-3 parts**

Subject (video) 4: Legal text on rules that apply for a test   **-3 parts**

Subject (video) 5: News information on a planned space trip   **-2 parts**

Subject (video) 6: Welcoming information for a participant at school   **-2 parts**

The speakers talked for about 45- 50 seconds in each part. The stories told were not unique to the

speakers as they all read the same manuscript. This was to minimize the risk of the perception of a video based on the person in the frame or preferences to male or female voices rather than the video and audio in itself.

Figure 1. A still frame of one of the free videos. The man in the video is introducing a meditation bowl.

Free improvisation

Additional speakers (two women and two men) speaking about two subjects each about freely chosen topics were recorded. All speakers introduced an object as seen in figure 1 and presented a made up story about them. Each story was unique to the speaker (none of the stories were repeated by a different person). This kind of video will be referred to as "free improvisation video" in the thesis. The objects introduced in the free improvisation videos were:

Person 1:
- A Kenyan statue of a giraffe
- A Nepalese meditation bowl (see figure 1)
- A Fair Trade textile shopping bag

Person 2:
- A marker pen

- A bottle of sparkling water

Person 3:
- A coffee cup
- An umbrella
- A calculator

Person 4:
- An audio cable

- A hair clip

Reference videos

Reference videos of the speakers were shot for the purpose of introducing the subjects to the test methodology prior to the test. The videos were also freely improvised. This way determining the perceived quality would be more consequent. The reference videos were chosen so that the range of degradations was at each extreme and one of mid quality.

All topics were presented in Swedish, as it was the native language of all speakers. All test subjects were also native Swedish speakers, in accordance to the ITU-T standard for subjective testing P.800.

4.3 Randomisation of order

A running order for each sequence was generated by first giving each original clip a unique number. Using "Random sequence generator" at www.random.org (a free website number generator) 12 different playlists were created.

For each playlist eight degredations of audio were also drawn using the same generator. The eight degradations appeared equally over all playlists.

4.4 Video

Figure 2. Sony FDR-AX1 video camera. Picture taken from http://www.sony.co.uk

Equipment

A 4K camera (Sony FDR-AX1) was used for capturing the footage of all videos. See "Resolution" page 20.

Positioning

The camera was positioned at a distance of approximately 250 cm from the speakers and zoomed in on the waist upward and about the size of a head above the scalp of the speaker, leaving room for three people in the horizontal plane. Two camera lights were lit on each side of the speaker to help the video camera interpret the light better. The background was a white screen with matte texture. Figures 3 and 4 show the video recording setup from different angles.

Figure 3. A snapshot of the equipment setup slightly before recording. The panorama picture reveals that the room has been treated with acoustic panels to minimize resonance and standing waves.

Figure 4. A closer look at the setup. The microphone is placed on the right hand side of the camera, pointing towards the middle of the screen.

4.4.1 Size and conversion

The footage was shot using a 4k video camera. The footage was then imported into Adobe Premiere CS6 and sized down to HD quality using the 1440 x 1080 (HD Anamorphic 1.33) standard. H.264 was chosen as video format as it is a video codec widely used in both amateur as well as professional situations.
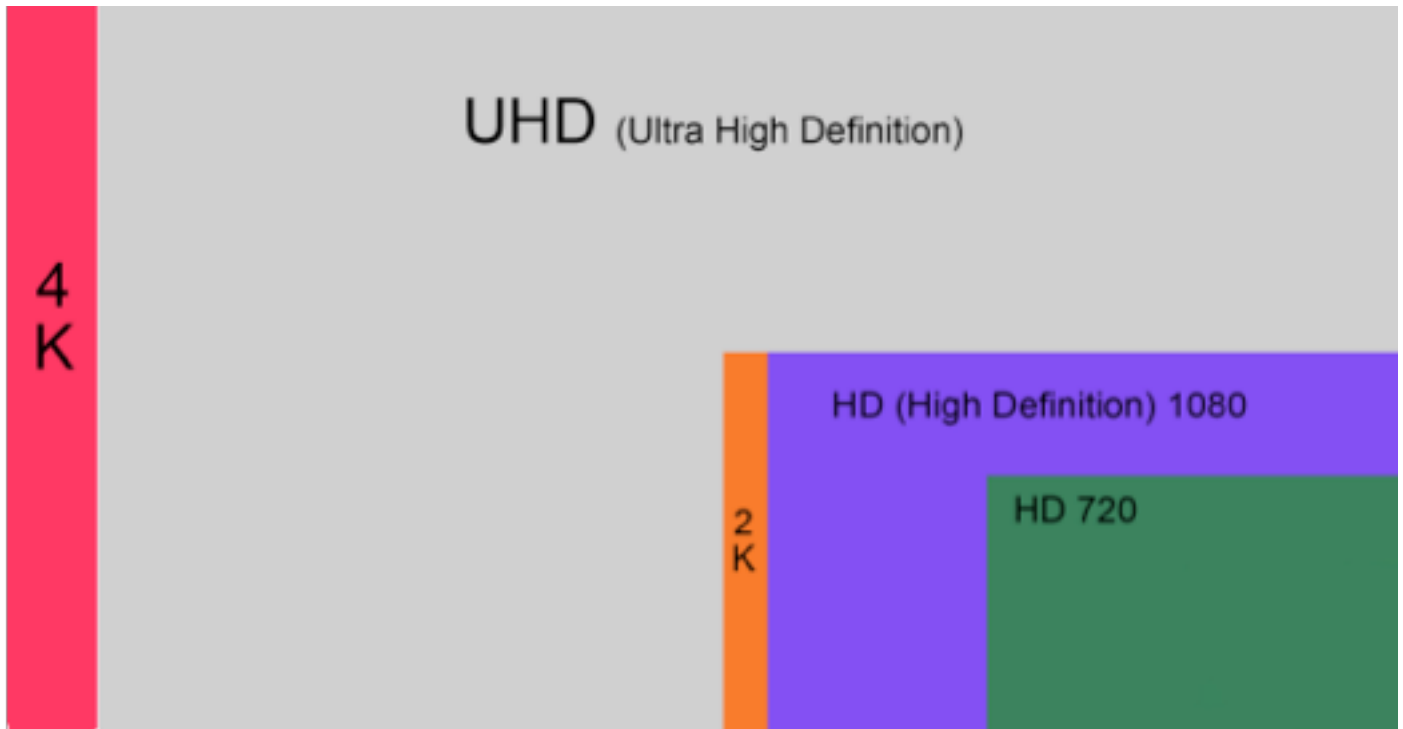

Figure 5. A table showing the difference in size on a screen. 4K is about four times the size of HD1080.

4K resolution *

The term 4K refers to the amount of pixels on the horizontal plane, which exceeds 4000 in number. This standard was created by the Digital Cinema Initiatives, (DCI), a group of joint motion picture studios establishing standards for digital video media. The definition of 4K according to the DCI is 4096x2160 pixels and is widely used by professionals in the film industry.

HD resolution *

HD stands for High Definition and is a consumer standard within video media today. The numbers that follow the abbreviation (such as HD 1080) refers to the resolution on the vertical plane, as opposed to the 4K system, where the resolution is defined on the horizontal plane. In some cases, the aspect ratio of each horizontal pixel is given as (1.33). The number refers to the value each pixel has been converted to and should be multiplied with in order to display video information correctly.

A display resolution of 1440 x 1080 will, using 1.33 aspect ratio be translated into a full 1920 x 1080 display resolution. HD resolution standards are also found in smaller sizes, such as 1280x720. These standards can be found in a high number of TV- screens and projectors as well as computer monitors today. Most TV broadcasters also allow HD.

H.264 *

H.264/MPEG-4 AVC is a block-oriented motion-compensation-based video compression standard that is able to achieve high quality video with a relatively low bit rate. Developed by VCEG (Video Coding Experts Group) together with the more familiar MPEG (Moving Picture Experts Group), H.264 is identical to MPEG-4 AVC (Advanced video coding). It is widely used for online video streaming using lossy audio compression, and is considered a standard in BluRay discs, where the audio compression is lossless.

4.4.2 Colour correction

The videos were colour corrected to ensure that none of the videos differed noticeably in colour information. Colour correction included slight adjustments of brightness, contrast and hue and saturation levels. Figure 6 shows a difference between uncorrected and corrected video.

Figure 6. Before and After. The picture on the left is a still picture from the video footage before colour correction. The picture on the right has been slightly colour adjusted and brightened.

This was done subjectively using A-B comparison between videos and checked on various monitors to minimize a preception of discolouration on monitors that offered less in colour resolution (bit-depth). This way the colour quality was kept consistent over several types of displays.

4.5 Audio

4.5.1 Technical Specifications

Microphone *

The appurtenant audio for the speakers was recorded simultaneously as the videos were shot using two different microphones. The microphone used primarily for picking up sound in the videos was a Shure KSM32, shown in figure 7. The microphone is a large diaphragm condenser microphone with directional a polar pattern (cardioid) and a sensitivity of 16mV/Pa at 1000Hz. The frequency response is quite flat except for a slight boost around 4-7kHz and below 50Hz, which was cut off in the audio processing stage, shown in figure 8 and 9.

Figure 7. Shure KSM32 Large diaphragm condenser microphone. Courtesy of Shure at www.shure.eu
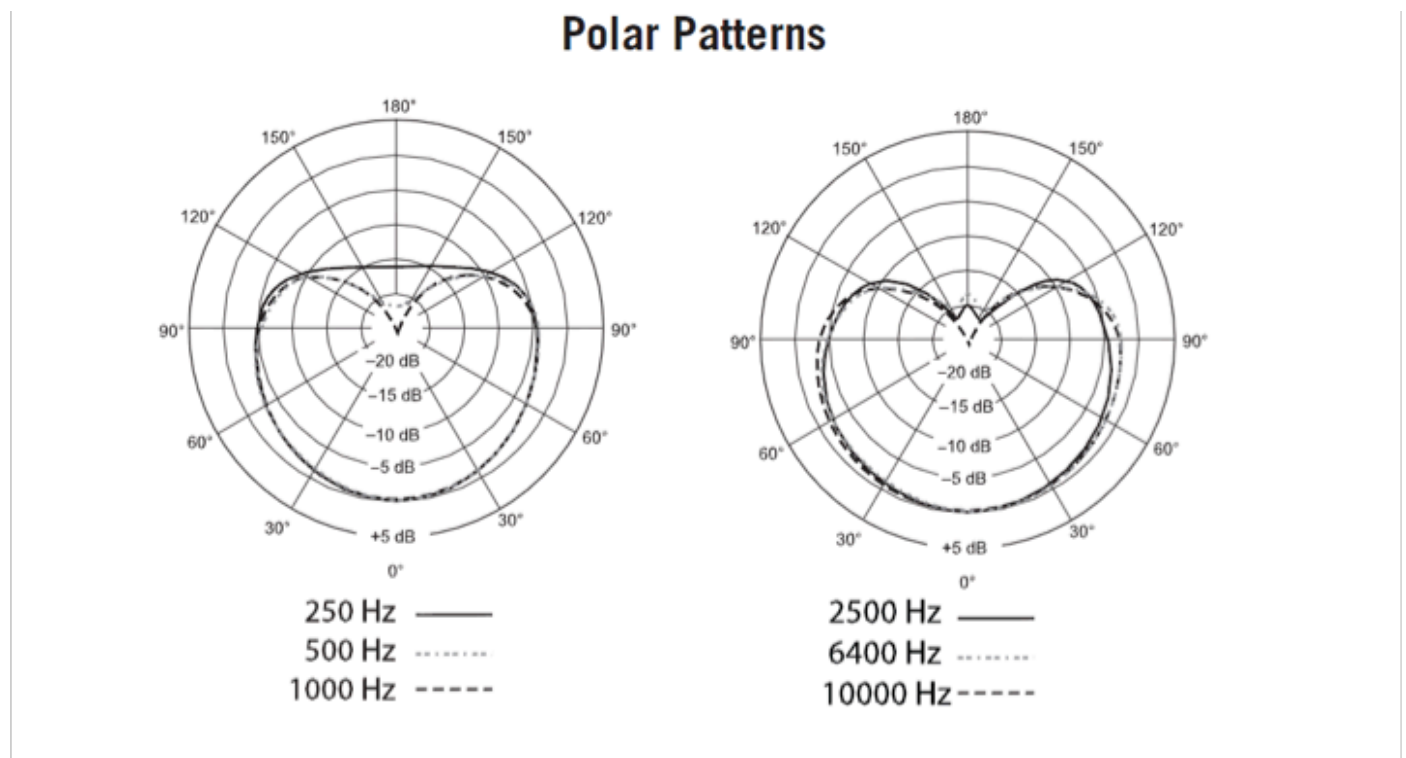

Figure 8. The polar pattern of the Shure KSM32 microphone shows its pickup sensitivity from different angles. The pickup varies sightly with frequency, being less dismissive of frequencies below low-mid (ca 400-500Hz).The picture is included in the technical specification that comes with the microphone.
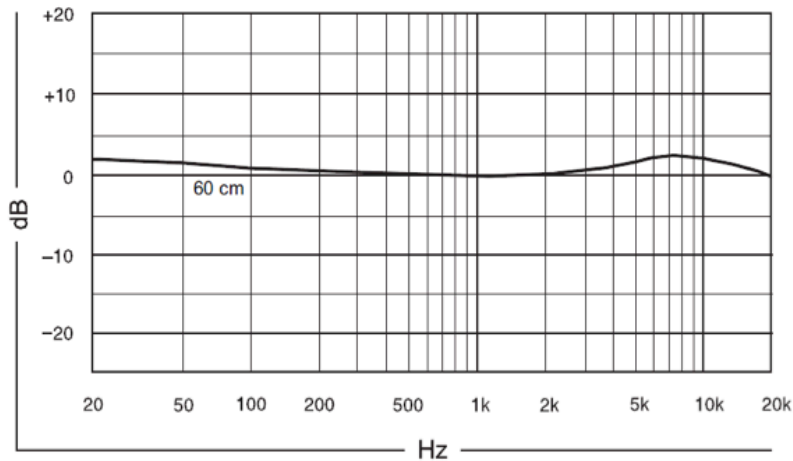
Frequency Responses

Figure 9. The frequency response of the Shure KSM32 microphone. There is a slight boost in the low frequencies below 300Hz. This does effect the frequencies that are important for the reproduction of human speech, as there is very little vital information in these frequencies. The boost starting at 3000Hz and peaking at 7000Hz is quite common for microphones suited for vocals as this area strengthens the frequency areas of consonants and sibilance (s-sounds). The area 2kHz- 4kHz is the most important for intelligibility of human speech.

The microphone was placed at a 35° angle from the speaker and 0° off axis from the middle of the recorded video frame, at a distance of approximately 30 cm. A shotgun microphone was attached to the video camera, picking up scratch sound (a term used for all audio that is used as a guide) that was later used as a reference to synchronize the externally picked up sound to the video correctly. When shooting a single person, the speaker was instructed to stand in the middle of the video frame for best audio pickup, as shown in figure 10.
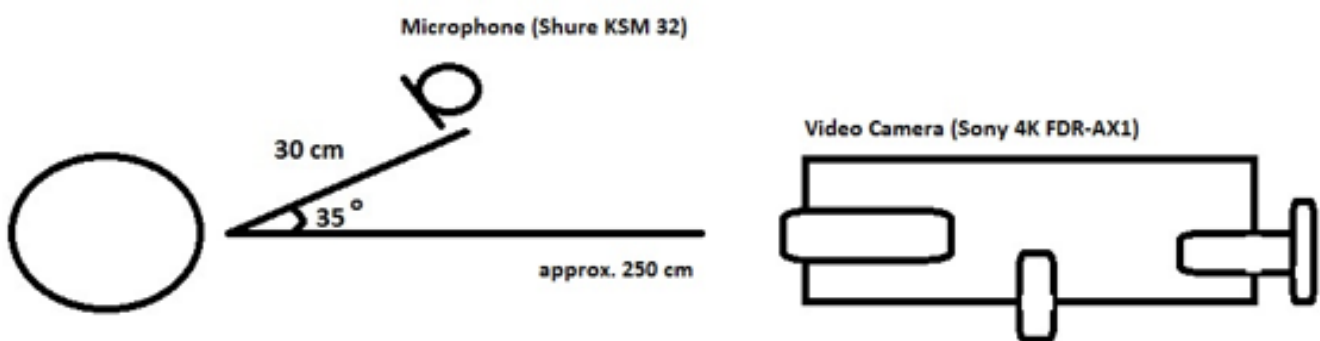

Figure 10. Positioning of pick-up equipment from above. The circle on the left represents the sound source. The picture is not scaled proportionately and shows only a sketch of sthe setup as a guide.

24

Digital Audio Workstation *

All recorded audio was imported into Avid ProTools 10 and slightly processed. The process included slight compression for better control the dynamics. Speech or vocals (sung vocals, beat boxing, rapping, yoddeling) is one of the most dynamic instruments in the World and must therefore be "tamed" a bit to better fit into the restricted frame of the digital realm. However, this process is something a human ear barely hears, but is noticed by the metering systems. The process did not take loudness into consideration, as that was a separate process (see "Choice of loudness", page 26).

4.5.2 Audio quality

The audio of the videos was recorded in 48kHz/ 16-bit PCM and loudness matched to a perceived loudness of -23dB in Adobe Audition. The recorded external noise was also loudness adjusted to

Figure 11. Logotype of EBU recommendation R-128. A majority of radio stations and other broadcast stations in Europe follow this recommendation, standardized for the purpose of keeping a constant loudness level in broadcasting, especially for commercial stations with compressed advertisement material.

Specifications of the original audio

The original audio of the videos was left at a sample frequency of 48kHz and a bit depth of 16-bit. All audio was coded in PCM (Wave).

Sample Frequency *

The sample frequency of audio refers to the frequency samples taken of an analogue signal in digital converting. The higher the frequency, the more accurate the translation to digital information within the range of audio frequencies will be. 48kHz is a standard sample frequency for visual media content such as DVD and BluRay and easily allows quality representation of the frequency spectrum of human hearing.

Bit depth *

The bit depth refers to the amount of bits used to represent the dynamics (position in binary code) of a digital audio signal. The higher the bit depth, the more detailed the dynamics of the sound. A bit depth of 16 is standard for audio CD:s and allows for a dynamic range of 96 dB.

4.5.3 Choice of Loudness

Loudness is the expression used for the perceived (subjective) audio level of a sound source. Since human perception of audio level isn't linear(perceived level is dependent on frequency), several factors have to be taken into consideration. All audio needs to be loudness matched in order to minimize the risk of biase due to difference in perceived amplitude. Loudness, measured in Loudness Units (LU) are not limited to only Sound Pressure Level (SPL). The main factors affecting the perception of loudness of a sound source are the acoustical sound pressure, the frequency range and the duration of the sound. In this test, after comparing the two, perceived loudness was chosen over -23 LUFS (although the differences where barely distinguishable).
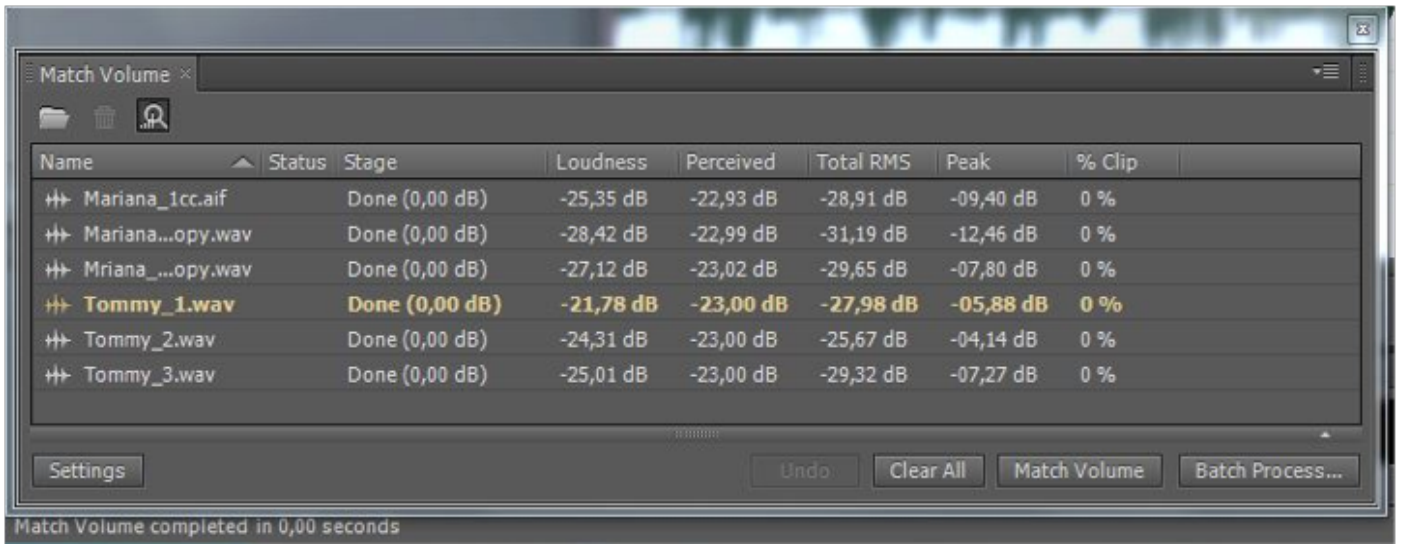
Figure 12. In the process of matching loudness in Adobe Audition. The program offers several settings for loudness. Here loudness and perceived loudness are compared next to the RMS and the peak value.

EBU R-128 *

Loudness has been actively researched and tested by the International Telecommunication Union (ITU), which has resulted in several recommendations such as the ITU-R BS.1770 in 2006 and an updated version called BS.1770-2 in 2011. The later recommendation has been adopted by the European Broadcasting Union (EBU) for standardizing of broadcast levels under the name EBU R-128. This standard obeys a Loudness level of -23LUFS (Loudness Units Full Scale).

Perceived Loudness *

In 2011 a study conducted by Begnert et al. (2011) a comparison between perceived differences of EBU R-128 metering and subjective human loudness was made. The study focused on whether differences between the R-128 and subjective human hearing could be established using natural speech, commercial speech, non-mastered music, mastered music etc. The results of the study reveal slight differences between content and perceived loudness if only at an average maximum difference of 2.8 dB.

4.6 Audio Degradations

The audio degradations (in addition to the original audio) introduced were:

1. Wide Band (50 Hz- 7000 Hz)   Abbreviation: WB

2. Narrow Band (200 Hz- 3400 Hz)   Abbreviation: NB

3. Reverberation (Stereo)   Shortening: Rev

The reason for testing perception for NB and WB is that these are codecs widely used within telephony today, NB being telephone audio and WB being an upgraded version (HD telephony). The degradations are explained in this chapter. Each of the degradations along with the original audio was also combined with an additional audio track consisting of noise recorded in a metro train. The noise did not include any intelligible speech and was not subject to any audio degrada-

Two different kinds of codecs were chosen for the final processing of the narrow- and wide band degradations. Various ITU- standards of band pass filters can be found for simulating codec behavior of an audio signal. In this study an actual codec was chosen as it was judged to be a slightly more accurate approach to everyday use than merely a filter.

Narrow band (NB)

AMR: Adaptive Multi- Rate

The narrow band codec was released in 1999 and is optimized for coding speech. Signals ranging from 200Hz to 3400Hz are encoded at variable bit- rates ranging from 4.75 kbit/s to 12.2 kb/s and a sample frequency of 8kHz. In the case of the test a bit- rate of 12.2kb/s was used. AMR is a standard speech codec used by 3G and GSM.

"The weighting corresponds to a bandpass filtering characteristic whose mask can be found in ITU-T Recommendation P.48 [REFERENCE COMES HERE] . In this specification a sampling rate of 8000Hz is used, leaving the highest possible frequency sampled with minimum aliasing to be 4000Hz in theory (about 3500 Hz in practice).
The spectral shapes of this weighting were obtained in a round-robin series of measurements

made on analog telephones in the1970's. he average send and receive frequency response characteristics were derived from these measurements.

For loudness balance purposes this standard includes a 300-3400 Hz bandpass filter, known as the SRAEN (Systéme de Rétérence pour la détermination des Affaiblissements équivalents pour la Netteté; Reference System for determining Articulatión Ratings) filter.

This weighting has been chosen to simulate speech signals obtained from a regular handset as the P.48 IRS weighting standard models an average narrowband telephone."

(A compilation of ITU-T standards G.711, G.721, and G.728)

Wide band (WB) and ITU-T recommendations

AMR- WB: Adaptive Multi- Rate Wide Band.

The codec is an improvement of the standard telephony codec AMR as it provides a frequency range of 50-7000 Hz. It is in accordance with the ITU-T standard G.722.22. The standard uses a sampling rate of 16 000 Hz which in practice results in a bandwidth of 50-7000 Hz. Compared to the option of narrow band it provides a substantial improvement of the quality. This applies especially for applications where speech is to be heard through high quality loudspeakers e.g. for audio or video conference systems or other audiovisual telecommunication. It also introduces comfort noise, which is a synthetic background noise to fill in the silent gaps that are a result of voice activity detection. The codec is used for improved wireless communication systems such as Telia's recently launched "HD Voice".

The ITU-T Recommendation G.722 uses a sampling rate of 16 000 Hz, which in practice results in a bandwidth of 50-7000 Hz. Compared to the option of narrow band it provides a substantial improvement of the quality. This applies especially for applications where speech is to be heard through high quality loudspeakers e.g. for audio or video conference systems or other audiovisual telecommunication.

Reverb

Reverberation (reverb for shortening) is the multiple echoes that occur when sound waves bounce

off the physical boundaries of an environment (such as walls). This is a naturally occurring phenomenon that helps the auditory sense to locate and grasp a physical environment. The decision to use reverb as a degradation of sound was influenced by previous research on the impact of early reflections, reverb and anechoic sound when measuring intelligibility of speech, such as (but not limited to) the Lombard effect (see page 30). The impact of early reflections is the relevant one to this paper as it has been stated to noticeably enhance intelligibility of speech (Bradley et al. 2003). On the other hand, research by Hu et al. (2013) on the effects of early reflections used in cochlear implants yielded in opposite results. In their study the average speech intelligibility scores were 90% when presented with an anechoic signal and 70% when presented with only a short, early reflected signal with an RT60 (reverb time) of only 0.3 seconds.

This raises a question of whether early reflections improve intelligibility of speech only for signals that are not directly fed into the ear, but rather travel through a medium from a sound source to the ear, allowing for natural acoustic pressure and phase interference? In this paper, the study   introduces the subjects to an audio channel distributed through headphones, which can be considered to be a "compromise" between cochlear implants and loudspeakers.

The Lombard Effect

The Lombard Effect, named after its discoverer Étienne Lombard (Lombard, 1911), is the involuntary and mostly by oneself unnoticed change in vocal amplitude (SPL) when speaking in a noisy environment. Although often being incorrectly referred to as a "reflex", the Lombard Effect has been found to incorporate differences within cultures, thus making it a cognitive phenomenon.

In 2010 Hodoshima et al. conducted a study where they introduced their subjects to recorded messages read out in quiet, reverberant and noisy environments. Some of the recorded audio signals were also introduced with reverberation and some with white noise. Their conclusion was that the subjects found reverberant messages to be more intelligible than quiet ones. Interestingly enough, the same applied regardless of whether the simulated reverberation matched the surrounding acoustics or not. This phenomenon has been referred to as "Lombard speech".

Early reflections (Reverb)

The full band (unprocessed) audio channel that introduced maximum audio quality of 16-bit linear PCM was infused with a reverb using the standard Dverb- reverb of DAW ProTools 10. The reverb was short and had a pre-delay of 20ms and a wet mix of 24% - all in all a standard reverberation of a medium sized room.
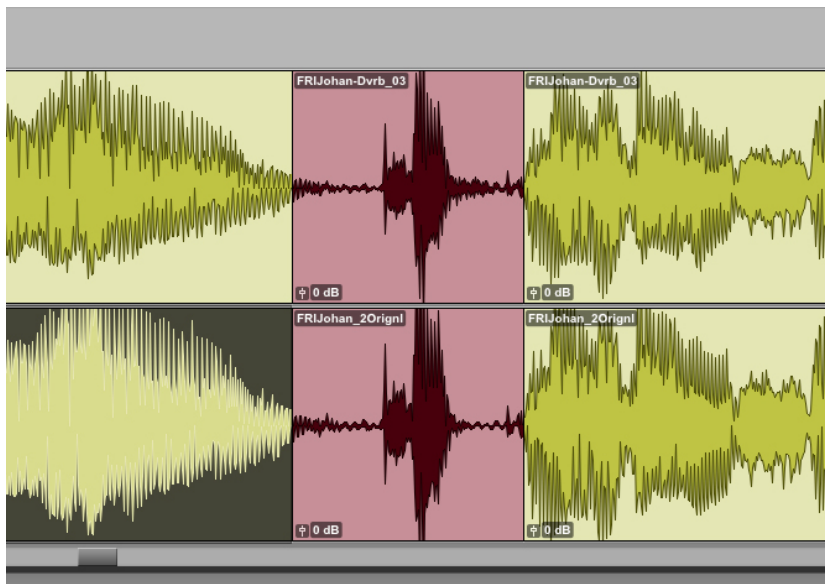

Figure 13. A snapshot of the audio wave of an unprocessed (bottom) and processed (top) audio signal in ProTools 10.The reverberated signal (to the right) has a longer tail and less detail. The audio waves are taken from a free improvisation video of a story about a trip to Nepal.
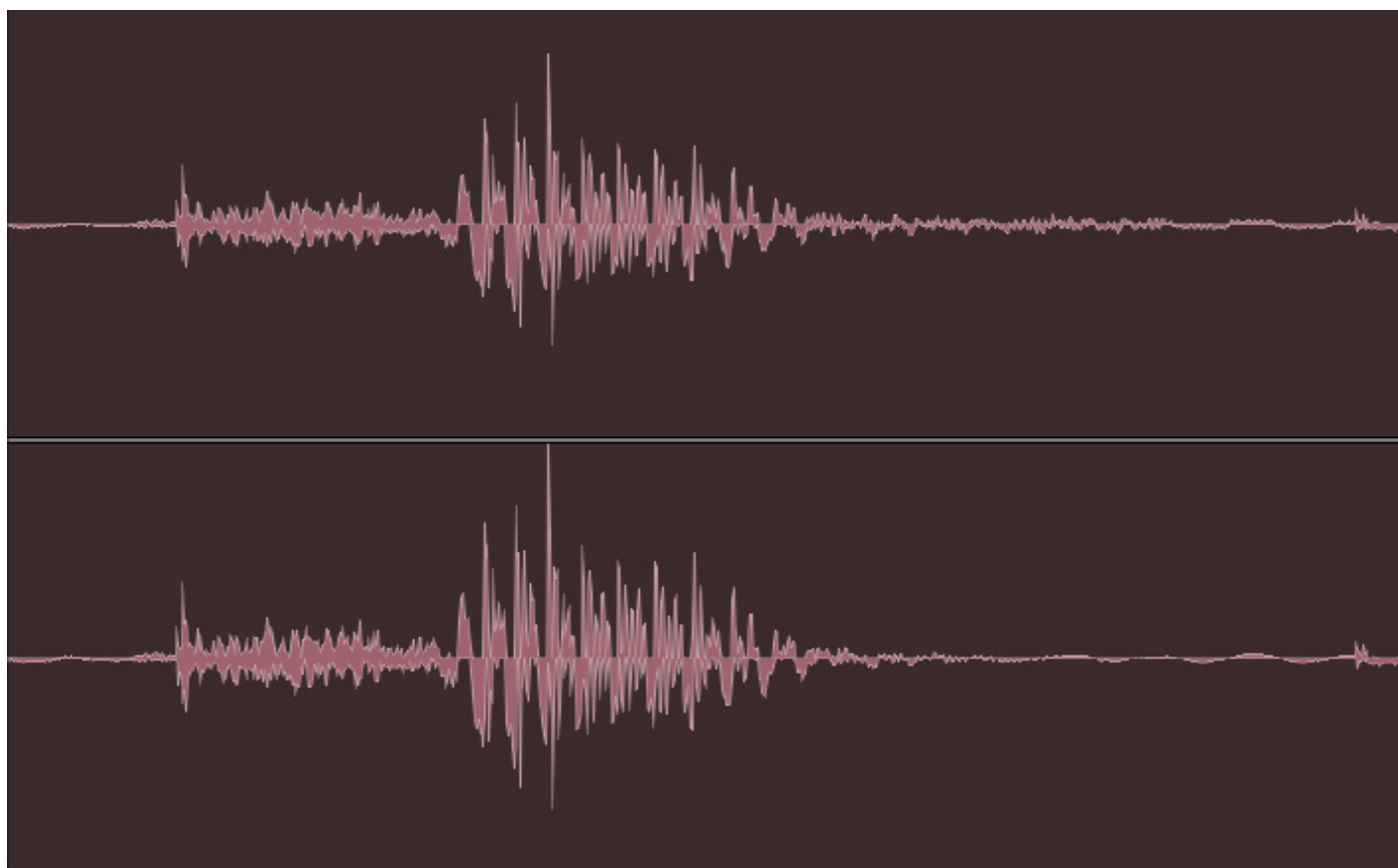

Figure 14. Zooming in on the syllable "Kat" in Katmandoo the effects of reverberation (top audio wave) become more clear. Notice how the "silent" sounds (at the end of the word) are louder than in the original (bottom audio wave). Simultaneously the peak is softer than in the original- resulting in poorer dynamic range.

4.7 Facilities and acoustics

The video footage of the speakers was recorded at a multimedia laboratory used for testing video and sound setups at Ericsson in Kista, Stockholm. The room was acoustically treated as well as sound proofed and the noise floor in sound pressure level was measured to be ca 29 dBA (SPL A-weighted).

The background of the speakers was a white canvas screen and lighting was used to adjust the appearance of the speakers to minimize any discolouring that the camera may add.

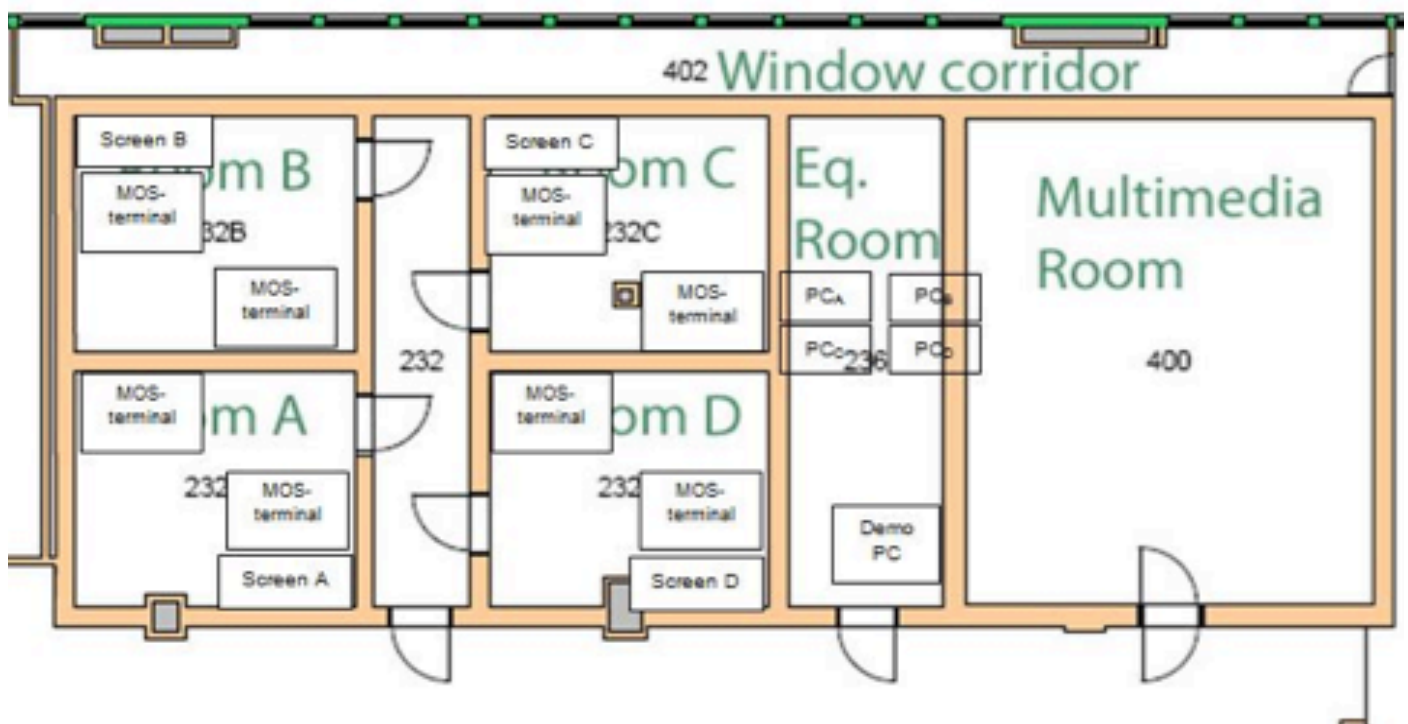Figure 15. An image of the Ericsson Multimedia Lab facilities taken from the Ericsson international technical report. The facilities are used for testing various media.

Room (space), size

The rooms were optimized for subjective testing of Audio, Video, AV, 3D audio, two-party teleconference tests (interactive, conversational) and multiparty teleconference tests. (See figure 15, Overview of Multimedia labs)

Figure 16. Panoramic view inside booth "D". The smaller screen on the table is the touch display for voting.

Temperature

The temperature was kept constant at 22°C with the help of silent air conditioning, designed to apprehend a constant temperature without making noise. During pauses and between test sessions, a more effective air conditioning was turned on.

dBA

The noise floor was measured with an SPL meter in each room and reached no more than 28 dBA. This was slightly less than in the multimedia labs where the original audio was recorded.

4.8. Running tests

The tests were not completely in accordance with MUSHRA (Multi Stimulus test with Hidden Reference and Anchor), an ITU-R recommendation for evaluation of multiple or double stimulus that has been found successful (Thoma 2012) in that the subjects were not in control of the sequence of the videos, nor could they replay any of the videos.

The setup was designed to accept four participants at a time during test. Three rooms labeled "A", "C", and "D" were equipped with one desktop PC with mouse, keyboard and screen each) and a touch panel PC:s used for recording the scores during MOSTER- testing (MOS- testing).
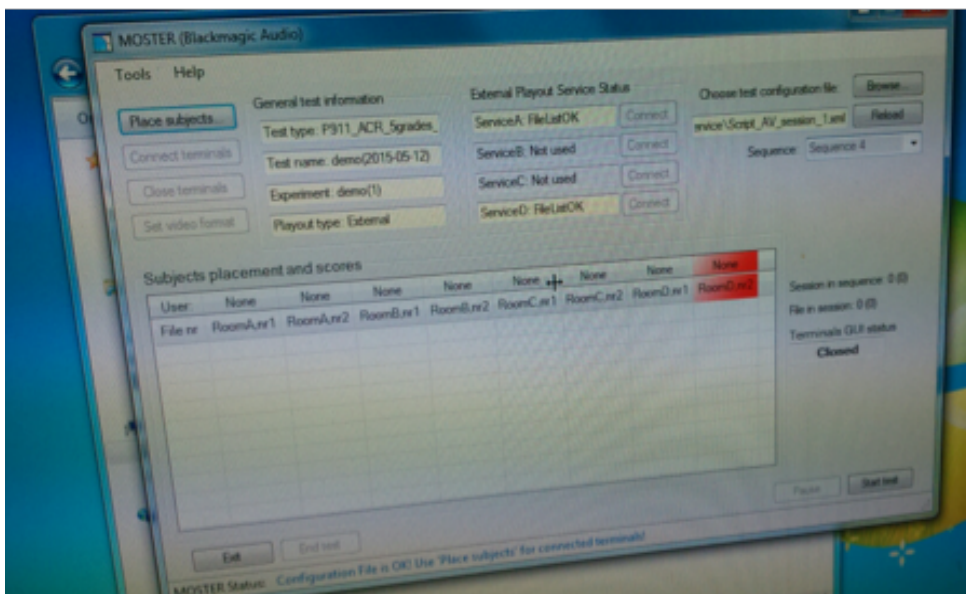

Figure 17. A screen shot of MOSTER reporting a disconnected terminal. The terminals are named after the testing booths: "A", "C" and "D". The results are reported in a column below each terminal.

The MOSTER system occupied four computers: three test computers (A, C and D) and a "master" computer control PC. How to run the test was specified in an xml- file on the control PC. The xml-file specified the size of the video display, the background colour and the colour shown between videos, the type of voting scale (buttons, A-B, resolution of scales, name of scales), the voting time, the number of videos in each sequence including the reference sequence and which playout computers to connect to (A, C and D in this case).

For each sequence that was to be played out a play order text file was created, pointing to a number that served as a file ID for each video in a video catalogue listing all videos including reference videos. The actual media files were all stored in identical folders on each of the playout computers (A, C and D). The Control PC recorded the score of the votes placed on the touch displays and created an Excel document for each session.

4.9 Instructions and testing

Before the participants stepped into the test booths they were given instructions on an A4 paper with a picture, showing the setup of the test, after which a thorough, oral presentation of the procedure was given. The participants were then all introduced to the test booths and given a recap of the instructions by showing how the equipment works and introducing the paper questionnaires. They were instructed to rate the media as if it was an everyday situation of conferencing or other media telepresence (i.e. NOT as if they were sitting in a Dolby THX certified cinema). They were also informed of the occasionally occurring background noise, which they were to regard as normal external disturbances- NOT as noise leaking into the actual audio recording.

Stora skärmen

1. Video visas på stora skärmen. Efter videon visas en fråga i taget i 10 sekunder. (2 frågor per video)
2. Kryssa i ditt svar på svarsblanketten.
3. Rösta på röstningsterminalen. Tryck PÅ slidern- inte bredvid. Du har 20 sekunder på dig att rösta.

Efter att röstningstiden gått ut visas följande video.
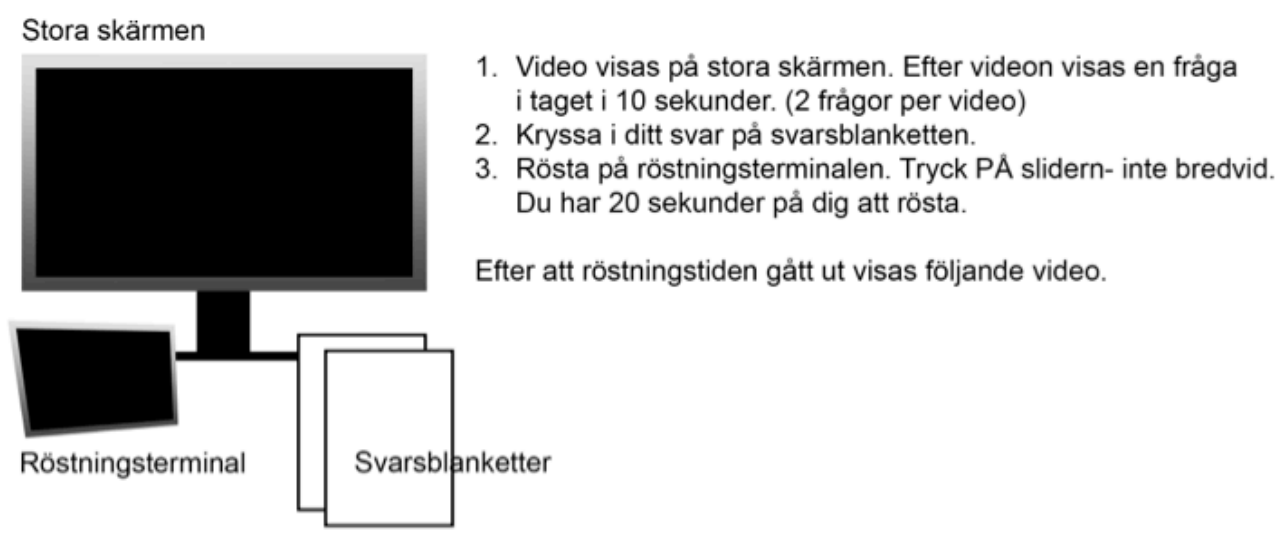
Röstningsterminal   Svarsblanketter

Figure 18. This image was printed on the instructions along with a walkthrough of how the test was to be performed. The text in Swedish is a three-step guide in which order the parts of the tests are being run.

Paper questionnaire

The questionnaires were six A4 papers stapled together, numbering each question and listing two questions per page. The papers were printed on only one side to minimize hassle when turning pages. The participants were also reminded that the number on each question only appears on the paper and is not displayed in the video itself.

The questions were not written on the papers, only multiple choice check boxes of three

possibilities: A, B and C. The font used for the multiple choice answer boxes was "Cambria", size 14.

Screen size

The screens used for displaying the videos on in each test booth were 20″ HP LP 2065. They could be adjustable on a vertical level to best fit the visual field of the participant.

Touch screen

The touch panel PC was an IEI PPC AFL12A. The screen reacted to both touch as well as to a small metal pen that ensured a higher accuracy rate of hitting small targets.

Headphones

All subjects were given a pair of Sennheiser HD 280 headphones for the test. The band of the headphones was adjustable for each person' s comfort. None of the subjects were able to change the volume.

Pre test

During the pre test the subjects were given the opportunity to perform the test as it was to be carried out including voting and answering questions. After this they all had the chance to make comments, ask questions or report errors or other disturbances. The pre test consisted of 3 videos of varying content and quality. The references included one of poor audio quality with background noise, one of medium audio quality and one of high audio quality with background noise.

Figure 19. Sennheiser HD 280 headphones used for the test. The band of the headphones are adjustable for each person's own comfort. The picture is taken at Ericsson labs and modeled by an acoustic dummy head designed for recording binaural audio signals.

Media content

At the end of each clip a black screen displaying a question about the content of the video was introduced for 30 seconds. The black colour was chosen because it did not frame the questions which also had a black background. This was to eliminate any chance of subjects finding the text to appear smaller if it was framed.

Voting procedure

When each video had played out, the subjects were instructed to vote on four categories regarding their perception and impression of the video. Voting was done using a small touch screen terminal.
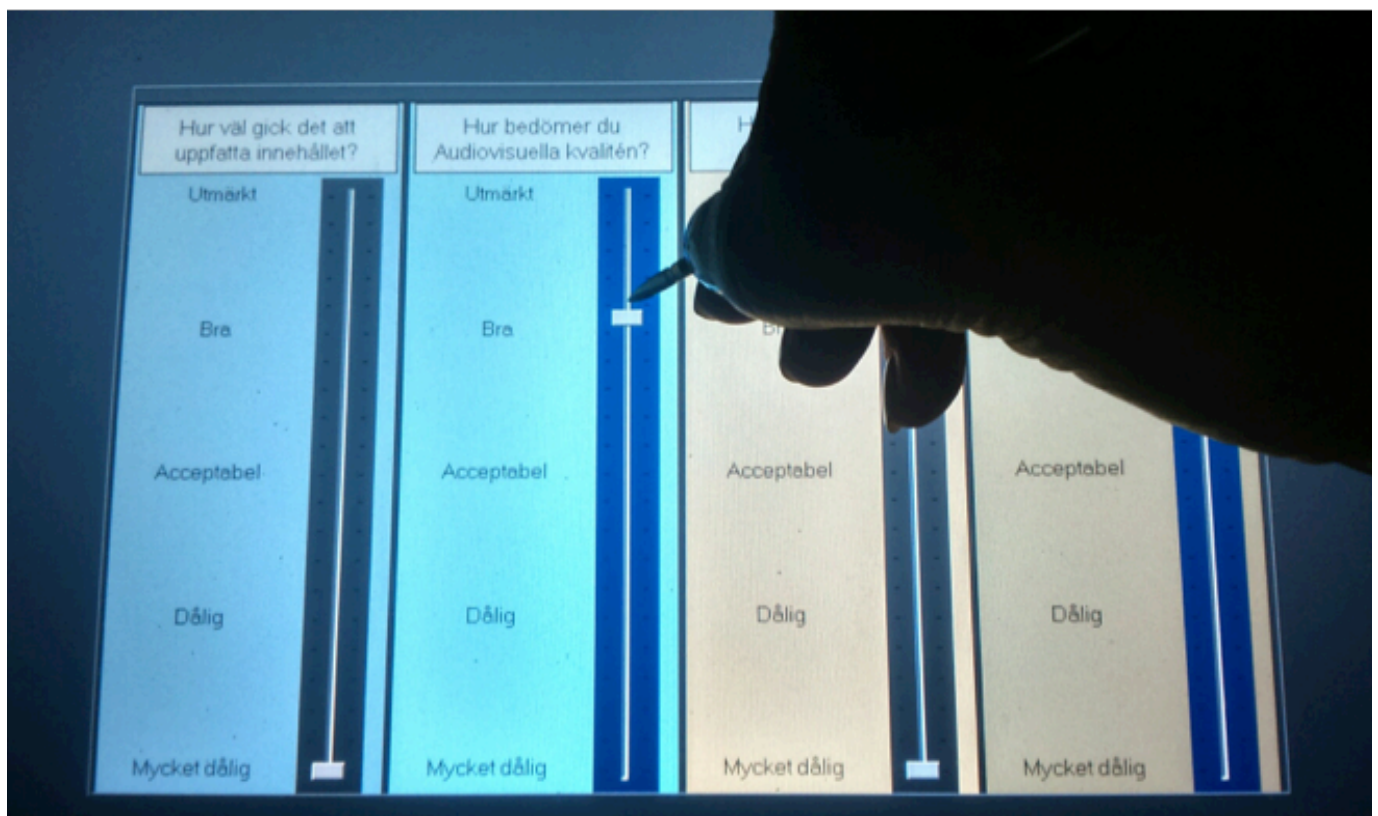

Figure 20. The voting terminal was available for 20 seconds after the last question had been displayed. The sliders were continuous and responded to both touch of finger and pen.

The categories voted for were:

1. How well did you comprehend the content?

2. How did you perceive the overall audio-visual quality?

3. How did you perceive the audio quality?

4. How did you perceive the video quality?

The voting scale was in accordance with the ITU recommendation P.800: "Methods for subjective determination of transmission quality" and was presented as a continuous scale:

Bad- Poor- Fair- Good- Excellent

"B.4.5 Opinion scales recommended by the ITU-T Various five-point category-judgement scales may be used for different purposes. The layout and wording of opinion scales, as seen by subjects in experiments, is very important, and should follow the standard arrived at through years of experience. The following opinion scales are those most frequently used for ITU-T applications and equivalent wording should be used depending on language which might result in small variations to the original English text:
a) Listening-quality scale Quality of the speech Score

Excellent 5 Good 4 Fair 3 Poor 2 Bad 1"

(ITU-T recommendation P.800: "Methods for subjective determination of transmission quality)

The quantity evaluated from the scores (mean listening-quality opinion score, or simply mean opinion score) is represented by "MOS" (see chapter 4.8 "Running tests").
Because the tests were run in Swedish, the translation of the scale was:

Mycket Dålig- Dålig- Acceptabel-Bra- Utmärkt

The questions were formed to match the voting scale grammatically.

Questions on content

Multiple choice questions limited to a maximum of three possibilities were introduced at the end of each video clip. The subjects marked their answers on a paper sheet, checking multiple choice boxes for A, B or C. Some questions were limited to two possibilities of answers (a and b). All questions referred to what had been said and/or seen in the video. The possible answers were categorized as:

- **One right answer**
- **One wrong but likely answer (a word that may have occurred in the video clip)**
- **One wrong answer**

Example1. A question from the manuscript videos "What is the main goal of the space project?" offers a) as a right answer, b) as a likely answer, as "China" was mentioned at some point in a different context in the video and c) as a wrong answer.

a)   **To build a permanent base on the moon**
b)   **To keep China from building a base on the moon**
c)   **To spend money**

Example2. Same setup as in example 1, but the question is referring to the visual content of the video.
Question 2: What kind of glasses did the person wear?

a)   **Square**
b)   **Round**
c)   **No glasses**

The questions served two purposes:
- To keep the subjects from paying too much attention to monitoring video/ audio quality by itself and keep the viewing as natural to an everyday situation as possible.
- To monitor both intelligibility and comprehensibility of the content, as it's more important to

understand content in an information video than merely seeing good quality. Each question was displayed for 10 seconds at the end of the video, one at a time. The subjects filled in a paper sheet showing only the multiple choice answers.

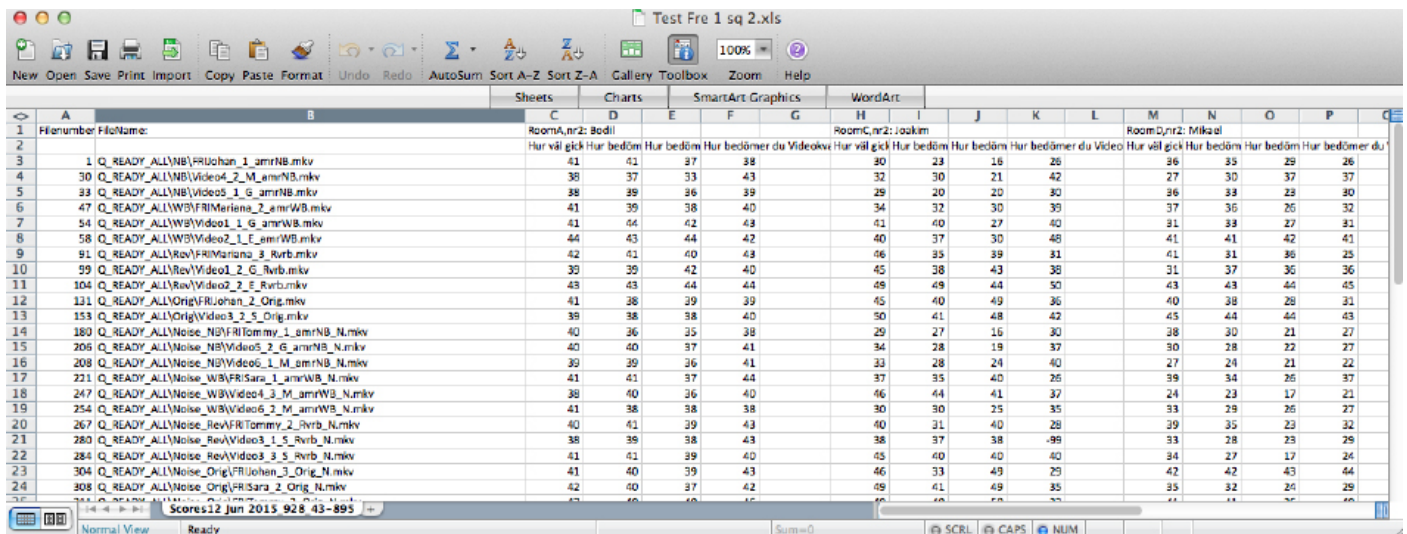| Filenumber	FileName:	Hur väl gick	Hur bedöm	Hur bedöm	Hur bedömer du	Videokvi	Hur väl gick	Hur bedöm	Hur bedöm	Hur bedömer du Video	Hur väl gick	Hur bedöm	Hur bedöm	Hur bedömer du
		RoomA,nr2: Bodil					RoomC,nr2: Joakim				RoomD,nr2: Mikael			
1	Q_READY_ALL\NB\FRIJohan_1_amrNB.mkv	41	41	37	38		30	23	16	28	36	35	29	26
30	Q_READY_ALL\NB\Video4_2_M_amrNB.mkv	38	37	33	43		32	30	21	42	27	30	37	37
33	Q_READY_ALL\NB\Video5_1_G_amrNB.mkv	38	39	36	39		29	20	20	30	36	33	23	30
47	Q_READY_ALL\WB\FRIMariene_2_amrWB.mkv	41	39	38	40		34	32	30	39	37	36	26	32
54	Q_READY_ALL\WB\Video1_1_G_amrWB.mkv	41	44	42	43		41	40	27	40	31	33	27	31
58	Q_READY_ALL\WB\Video2_1_E_amrWB.mkv	44	43	44	42		40	37	30	48	41	41	42	41
91	Q_READY_ALL\Rev\FRIMariana_3_Rvrb.mkv	42	41	40	43		46	35	39	31	41	31	38	23
99	Q_READY_ALL\Rev\Video1_2_G_Rvrb.mkv	39	39	42	40		45	38	43	38	31	37	36	36
104	Q_READY_ALL\Rev\Video2_2_E_Rvrb.mkv	43	43	44	44		49	49	44	50	43	43	44	45
131	Q_READY_ALL\Orig\FRIJohan_2_Orig.mkv	41	38	39	39		45	40	49	36	40	38	28	31
153	Q_READY_ALL\Orig\Video3_2_5_Orig.mkv	39	38	38	40		50	41	48	42	45	44	44	43
180	Q_READY_ALL\Noise_NB\FRITommy_1_amrNB_N.mkv	40	36	35	38		29	27	16	30	38	30	21	27
206	Q_READY_ALL\Noise_NB\Video5_2_G_amrNB_N.mkv	40	40	37	41		34	28	19	37	30	28	22	27
208	Q_READY_ALL\Noise_NB\Video6_1_M_amrNB_N.mkv	39	39	36	41		33	28	24	40	27	24	21	22
221	Q_READY_ALL\Noise_WB\FRISara_1_amrWB_N.mkv	41	41	37	44		37	35	40	26	39	34	26	37
247	Q_READY_ALL\Noise_WB\Video4_3_M_amrWB_N.mkv	38	40	36	40		46	44	41	37	24	23	17	21
254	Q_READY_ALL\Noise_WB\Video6_2_M_amrWB_N.mkv	41	38	38	38		30	30	25	35	33	29	26	27
267	Q_READY_ALL\Noise_Rev\FRITommy_2_Rvrb_N.mkv	40	41	39	43		40	31	40	28	39	35	23	32
280	Q_READY_ALL\Noise_Rev\Video3_1_S_Rvrb_N.mkv	38	39	38	43		38	37	38	-99	33	28	23	29
284	Q_READY_ALL\Noise_Rev\Video3_3_S_Rvrb_N.mkv	41	41	39	40		45	40	40	40	34	27	17	24
304	Q_READY_ALL\Noise_Orig\FRIJohan_3_Orig_N.mkv	41	40	39	43		46	33	49	29	42	42	43	44
308	Q_READY_ALL\Noise_Orig\FRISara_2_Orig_N.mkv	42	40	37	42		49	41	49	35	35	32	24	29

Figure 21. An example of an Excel sheet produced after a session. The file name on the left indicates which media file has been played and the results of each subjects' voting is displayed under the corresponding voting column (i.e. which category they were rating) and separated by booth (person).

Group discussions

After each test session, the group of test subjects were gathered into a quiet room to discuss their perception of the test. Each sessions was recorded with permission of the subjects. The discussion was a semi- structured interview based on the following questions:

- Do you have an AV interest, such as owning a home theatre system?
- Did you find anything to be of difficulty in the test?
- Was the pace at which you watched the videos and answered questions etc. good?
- How did you experience the background noise?
- Was the rating system easy to use? Did the rating system make sense to you?
- Were the pre-test videos sufficient in number?
- Did you experience difference in quality in the audio or video?
- Did you find it hard to concentrate on both measuring quality and answering ques tions?

The purpose of the discussion was to find out a bit about how the participants experienced the test and possibly find factors to support and/or contradict the results. The results of the discussion are written in Chapter 5.4 "Results of group discussions" (page 54).

5 Results

In this chapter, the results of the test conducted are presented. The analysis of the results can be found in the chapter "Discussion", where some of the results are also shown. Charts shown in this chapter are referred to in "Discussion". Each chart has a short description for clarification. Below is a short summary of the test.

30 persons of 250 invited were accepted to participate in the test, of which 27 showed up (4 women and 23 men). The participants were all adults between ages 20 and 65. All reported having normal hearing and eye sight, which was a criteria. People wearing glasses were accepted as participants. Anyone reporting any kind of colour blindness was not accepted. All participants were native speakers of Swedish. Each participant was given written instructions prior to the test as well as an oral introduction to the test with the chance to ask questions [see APPENDIX]. A pre- test was conducted as training for the participants, before the actual test session began, allowing them first acquaint themselves with the test procedure. The test session included one of each extreme of the video clips as well as one of mid- quality (one with narrow band audio, one with original audio and one with wide band audio). The participants were free to dim the lights to their liking although the lights were set as half dimmed by default. They also had the freedom to choose the viewing distance to the screen although most sat at a 50 cm distance to the video display. A maximum of three people participated in a session at a time, resulting in 10 sessions over the course of three days (3, 3 and 4). The 648 videos were organized as follows:

Free videos (improvisational)

Manuscripts with heavy/more complicated text

All manuscript videos (read from text displayed)

The subjects in manuscript videos 1,2,4 and 5 were considered to have difficult text by (from now on referred to as "heavy" text) the participants and was regarded as different than video 3 and 6. Figures 35a and 35b show a difference in accuracy rate and average "Comprehension" rate between videos 1,2,4,5 and videos 3 & 6. The number of questions answered in total was 1296. The questions were organized into 432 visual questions (questions about visual content) and 864 comprehension questions in total. Each sequence of videos included 24 videos in total (48 ques-

tions). One question in eight was a question about visual content. The rest of the questions (40 in total) referred to the spoken content.

5.1 Rating

In this chapter all abbreviations and shortenings will be used for audio degradations as follows:

NB	=	Narrow Band
WB	=	Wide Band
Rev	=	Reverb
Orig	=	Original
(n)	=	noise

The scale used for rating the videos ranged from 10- 50 (10 corresponding to the lowest score and 50 to the highest score) counting two decimals. The mean value of each category was calculated. The Audiovisual category is referred to as AV. "Uppfattning" has been translated (from Swedish) to "Comprehension" in this table, though most are in Swedish and rephrased in t in the description.

FRI total	Comprehension	AV	Audio	Video	Amount
NB	40.23	35.54	30.67	36.1	39
NB(n)	39.5	35.21	31.03	38.06	34
WB	42.88	39.41	38.56	38.12	34
WB(n)	41.51	38.16	34.62	39.32	37
Rev	45.35	42.19	42.89	40.35	37
Rev(n)	42	39.11	37.53	39.63	38
Orig	44.81	42.71	43.14	41.05	21
Orig(n)	43.17	39.5	38.93	39.07	30
Total average	42.43	38.98	37.17	38.96	
ST DEV	2.04928864	2.709983856	4.780476478	1.539292509	

Figure 22 . The table shows average rating of the free improvisation videos within degradation categories. Videos with noise in the background are marked with (n). The average of each category is shown in the vertical column "total average".

Average of Rating

On average Comprehension was rated higher than any of the other categories. Variation between same ratings in the same category was calculated using a standard deviation formula. Significance was thus calculated as follows:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

The square root of the sum of

the squared differences between each data point (vote) and the mean average of all data points.

The category with the lowest mean score was Audio. However, it was also the category with the highest fluctuation in the ratings (4,78). The category with the lowest fluctuation in rating was Video, followed by Comprehension (2,05) and AV (2,71) (see figure 24). Videos were separated by content as many subjects reported perceiving noticeable differences between some manuscript videos (video 1, 2, 4 and 5) and free improvisation videos.

Statistics of the overall rating of in each category and each degrading were calculated. The significance was calculated by dividing the squared differences of the sum of the value of the votes by the mean average of all the votes. The full confidence values are shown in tables 28a and 28b.
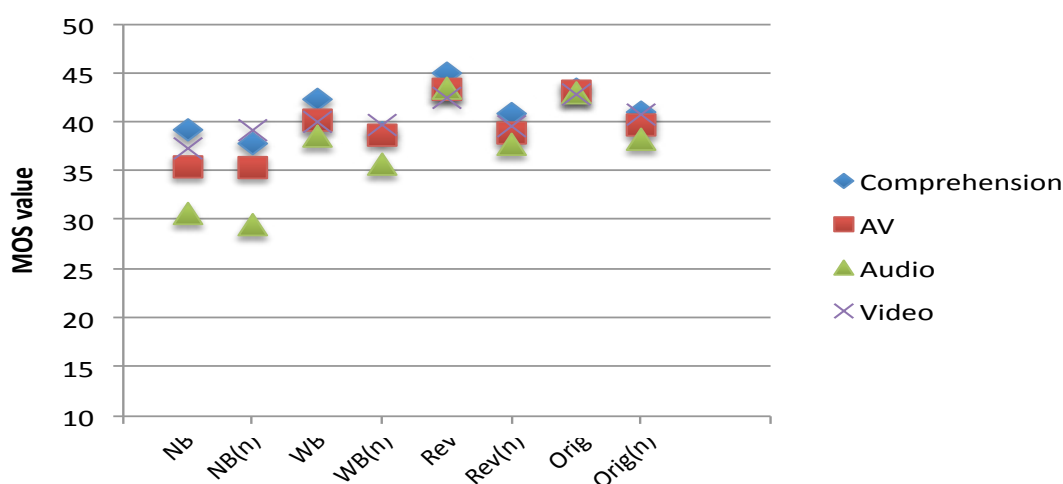
Figure 23 shows the rating of all versions of audio degrading within all of the four categories. The full scale is shown as participants rarely found any quality below acceptable.

5.2 Noise Versus No Noise

In general, videos with noisy backgrounds received a lower score than ones without noisy background in all categories and degradations, except for manuscript videos' AV with Narrow Band, where the Narrow Band clean signal received the average score 35,35 while Narrow Band with

noisy background received 35,58 (avg. difference 0,23 for manuscript videos and 0,33 for free videos).
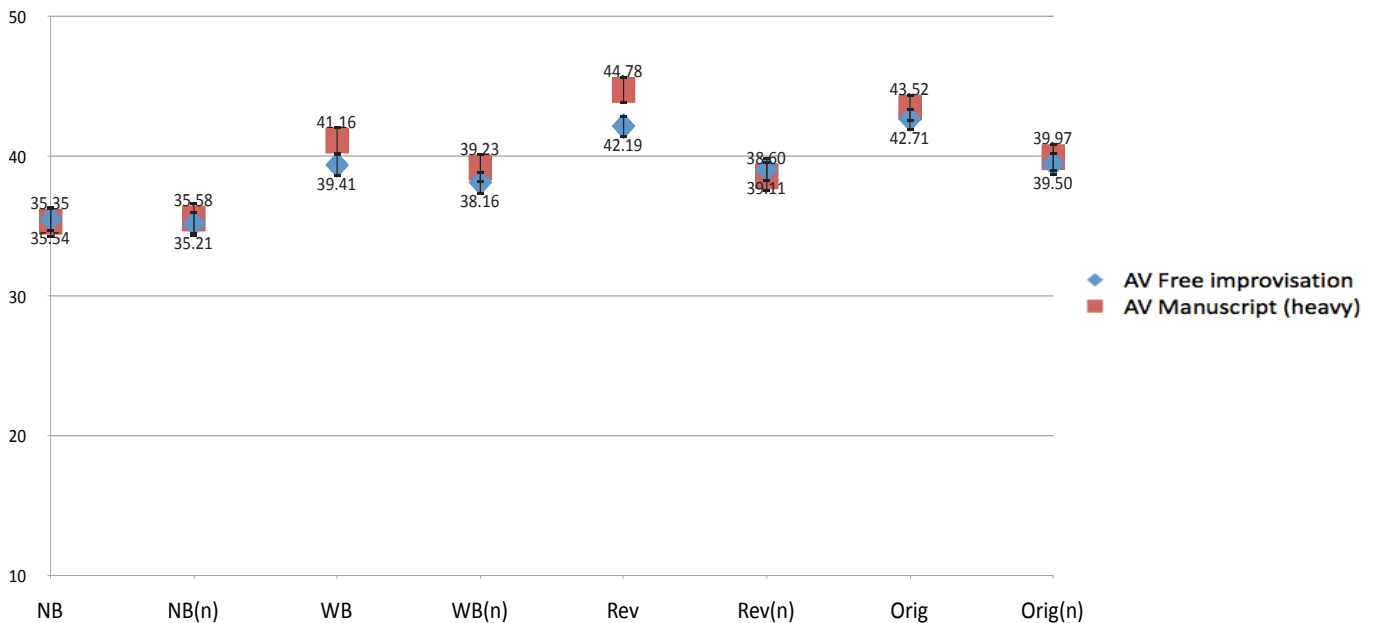
Figure 25. The average AV quality score was higher for manuscript videos in general, except for Rev(n), where the free videos received a 0.51 higher score than the manuscript, and clean NB, where the free videos received a 0.19 higher score than the manuscript videos.
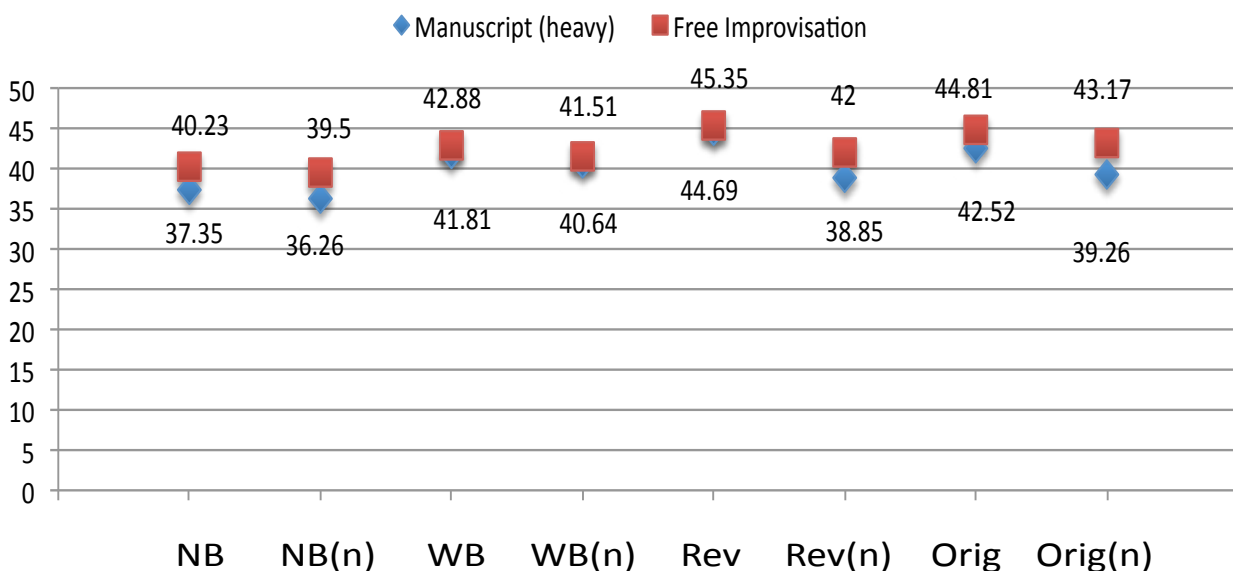

Figure 26. The Free videos received a higher score in all categories than the manuscript videos 1,2, 4 and 5. Here video no. 3 and 6 were left out as the subjects neither considered them manuscripted nor truly improvised.

Comprehension Manuscript (heavy) Video
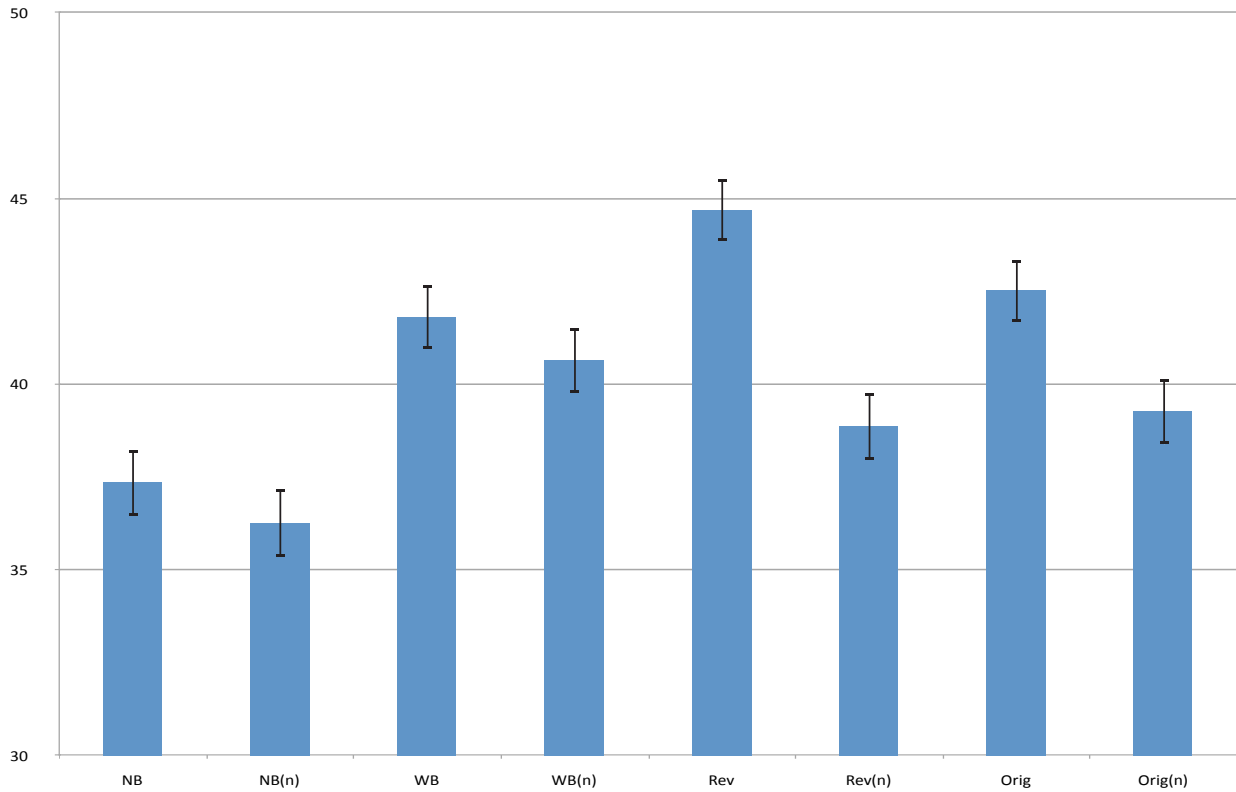
Figure 27a. shows Confidence rated for free improvisation videos.

Comprehension Free Improvisation Video
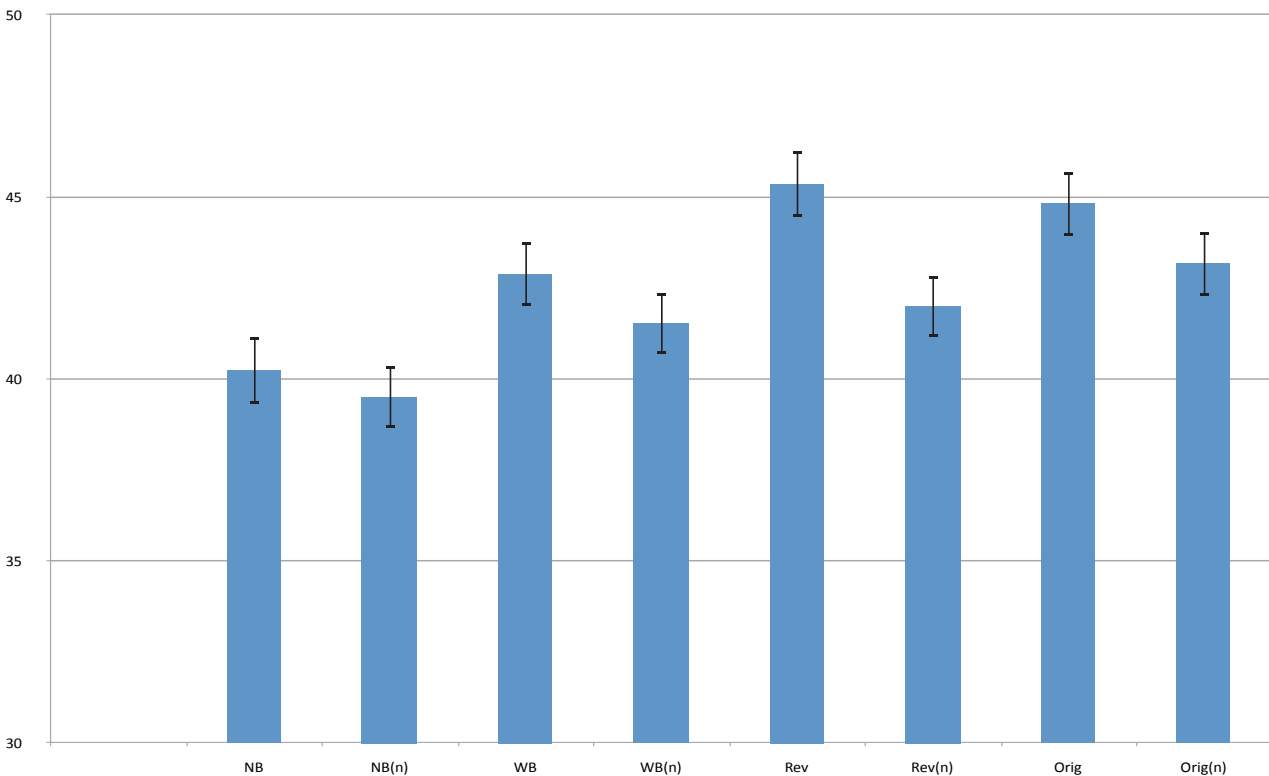
Figure 27b. shows Confidence rated for manuscript videos with heavy content.

Figure 27a (top) and 27b (bottom) A closer look at the results of rated "Comprehensibility" reveal the confidence value of each category. A general opinion of the participants was that the quality never reached unacceptable and didn't vary as much as they had expected. This is reflected by the slim variation of points given by most participants. Figuree 28a and 28b show the confidence values as numbers.

Confidence Free Imp total	Comprehension	AV	Audio	Video
NB	0.63506945930457	0.8978024	1.7106401	0.5028263
NB(n)	0.64315984227028	0.8978024	1.682823	0.4894152
WB	0.61976433166272	0.850516	1.5199436	0.4894152
WB(n)	0.62727690088338	0.8616343	1.6068666	0.4830999
Rev	0.59874922826498	0.8195777	1.4457548	0.4770229
Rev(n)	0.61976433166272	0.850516	1.5403465	0.4830999
Orig	0.60551497101934	0.8195777	1.4288448	0.4711697
Orig(n)	0.61251537665855	0.850516	1.5199436	0.4830999

Figure 28a. Table of confidence rated for free improvisation videos. The table shows the full numbers. At the end the numbers were rounded off to two decimals.

Confidence Manuscript videos (heavy)	Comprehension	AV	Audio	Video
NB	0.9005130757021	1.1191923	1.9522686	0.6434048
NB(n)	0.9129345324874	1.1191923	2.02079	0.6353114
WB	0.8554585217794	1.0340625	1.7346362	0.6200004
WB(n)	0.8660857431844	1.0602455	1.7821693	0.6353114
Rev	0.8257803552053	0.9981881	1.6120328	0.5989773
Rev(n)	0.8885852235567	1.0741055	1.7346362	0.6434048
Orig	0.8452131399533	1.0097282	1.6499681	0.6057456
Orig(n)	0.8771191273927	1.0602455	1.7579209	0.6200004

Figure 28b. Table of confidence rated for manuscript videos with heavy content. The table shows the full numbers. At the end the numbers were rounded off to two decimals.
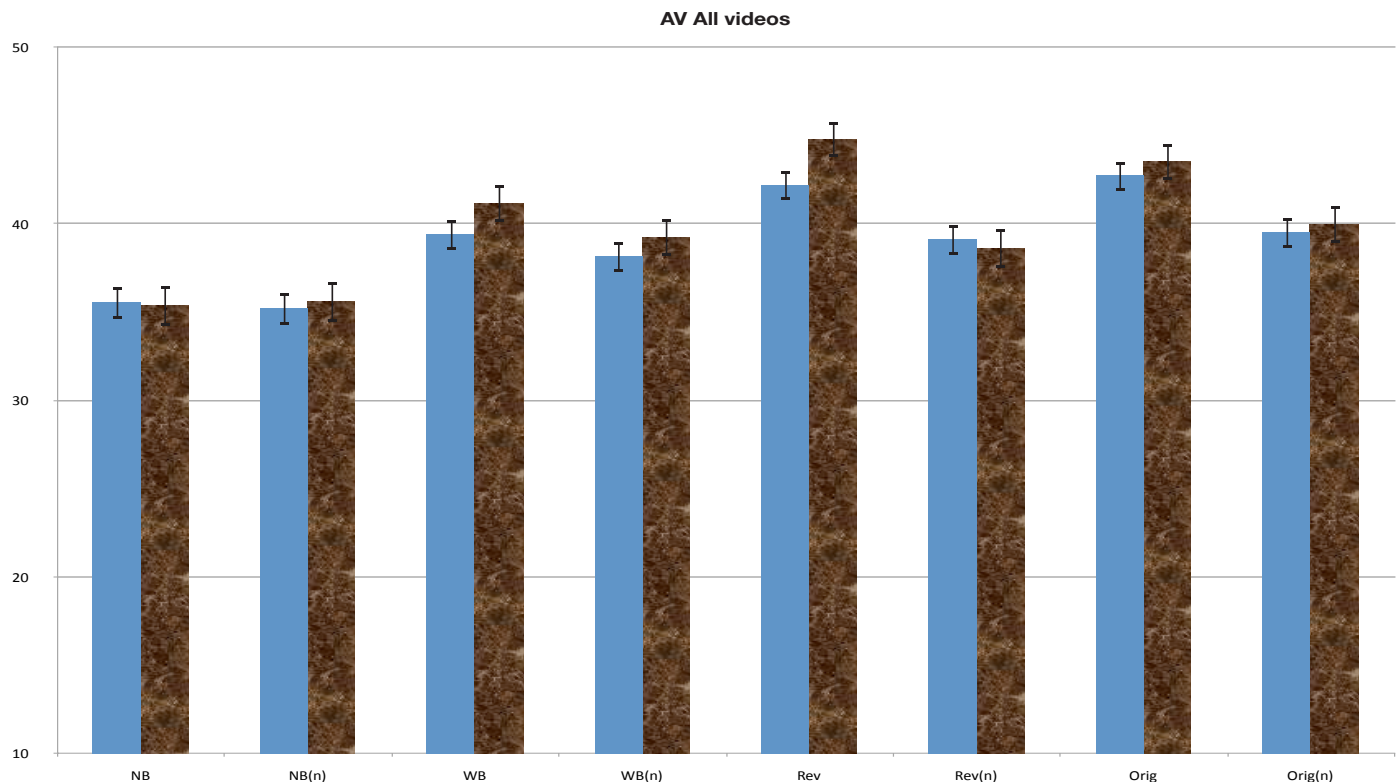
Figure 29. Confidence levels of AV ratings for all videos. The variation is greater in general for the manuscript videos with heavy content than for the free improvisation videos.

Subjects experienced less difference between Narrow Band videos with and without background noise compared to noisy and "clean" videos with higher audio quality. The biggest difference experienced was between noise and no noise in the AV quality rating of manuscript videos with reverb (avg. difference 6,18).

The perceived video quality was experienced as worse in a noisy environment when comparing the same kind of audio degradation. No significant difference was found in the subjects' rating between videos containing female speakers and videos containing male speakers when calculating the statistical significance (using the standard diviation formula $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$) of the two and so the taking gender into consideration in further statistics was discarded. Table 31 shows a quick comparison between the perception of videos with similar content containing female and male speakers.However, there were clear differences in comprehension rating between types of video content. Manuscript videos constantly received a lower average score on comprehension than the free improvisational videos. However, the AV quality rating for manuscript videos remained at a higher score than the free videos, except for NB.
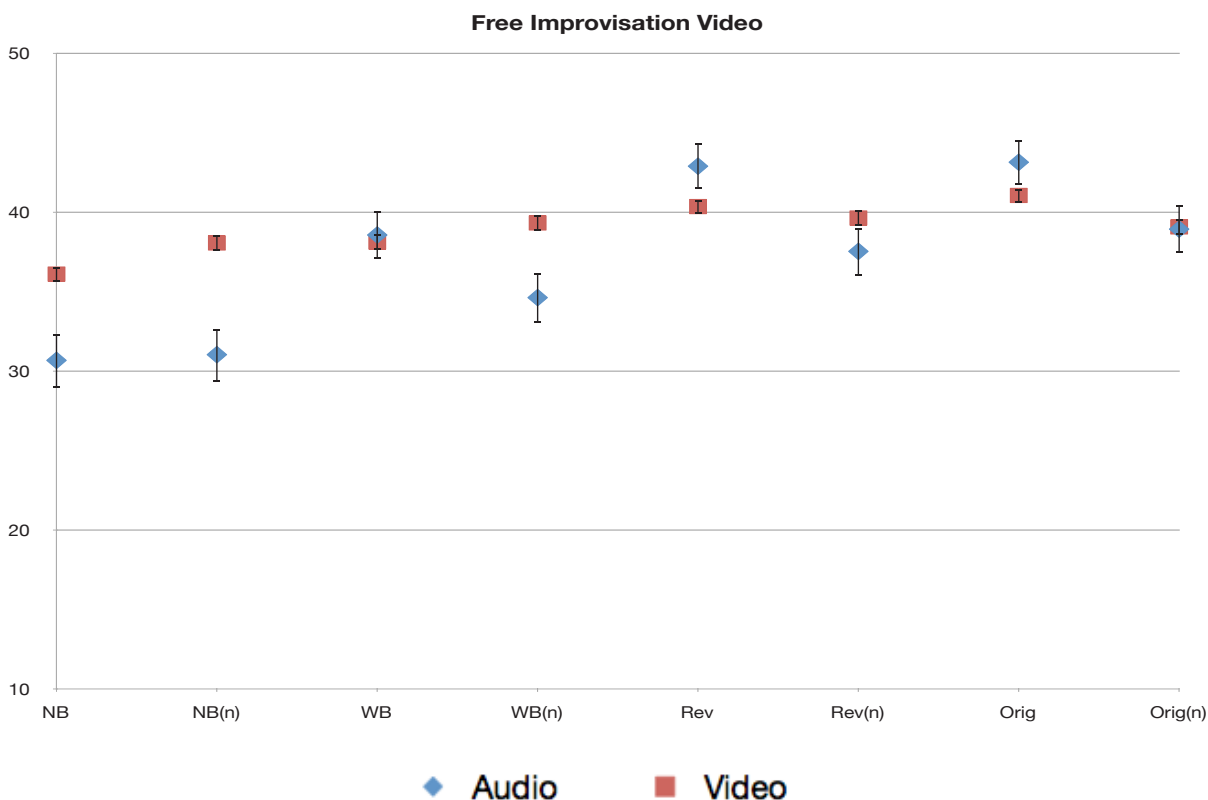
Figure 30a. This chart shows the results of the audio and video (separately) quality ratings for the free improvisation videos in closer detail. There is a great difference between the variety of perceived audio quality and video quality. As most participants claimed, they perceived very little difference in video quality.

Manuscript (heavy) Video

◆ Audio   ■ Video
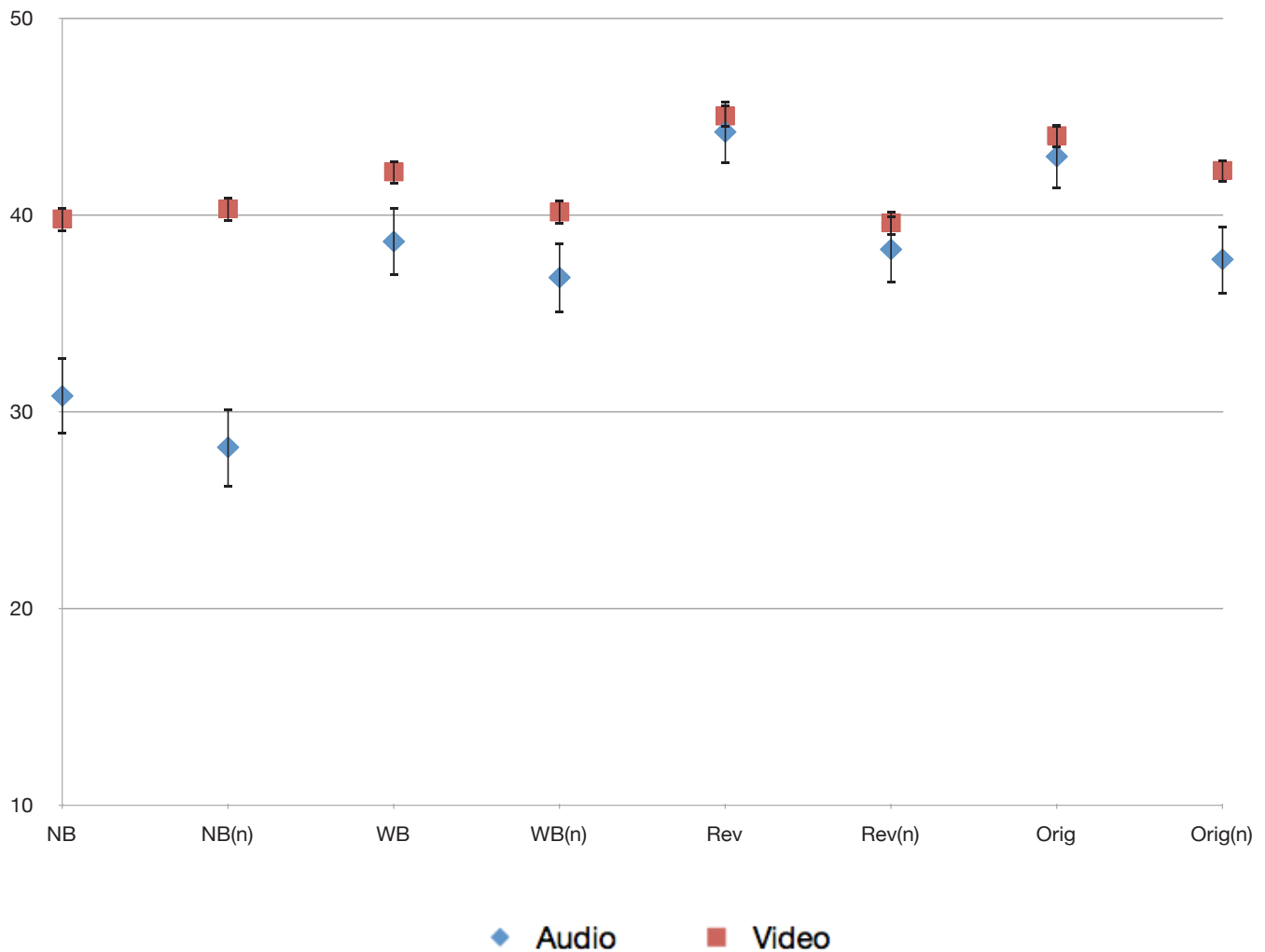
Figure 30b. The chart shows the results of audio and video (separately) ratings of the manuscript videos in closer detail. When comparing Figure 30a and Figure 30b, there is a slight difference of variation in perceived video quality. The free improvisation videos contained more movement of the subject in the camera than the manuscript videos, which test subjects reported to have generated a slight difference in their perception of video quality. This claim was discussed in groups afterwards and divided opinions between subjects.

Video 1

	Male	Female
Avg Comprehension	39.1	39.8
Avg AV	39	39.8
StDev AV	6.10488	5.956268
StDev Comprehension	7.568746	7.035285

Video 4

	Male	Female
Avg Comprehension	37.7	39.6
Avg AV	36.2	40.71
STDev AV	8.02171	7.259856
STDev Comprehension	7.40393	8.554777

Table 31. A table of videos 1 and 4 shows the standard deviation between the scores of female and male speakers for videos with similar content, within categories "Comprehension" and "AV quality". Both videos are manuscript videos. Video 1 was seen 38 times with a male speaker and 18 times with a female speaker. Video 4 was seen 21 times with a male speaker and 55 times with a female speaker. Other conditions (degrading and noise) was not considered in this calculation.

5.3 Results of Questions

Only one person left one question unanswered during the test. All questions answered were thus 1295 in total. There were 50 different questions in total, of which each person was introduced to 48. This was because some of the manuscript videos showed the same content although the questions asked about it were different. Many of the subjects reported the content being of great importance for comprehension and attention given to each video, so the videos were categorized as "Free Improvisation videos" and "Manuscript (heavy content) videos" (which included manuscript videos 1, 2, 4 and 5).
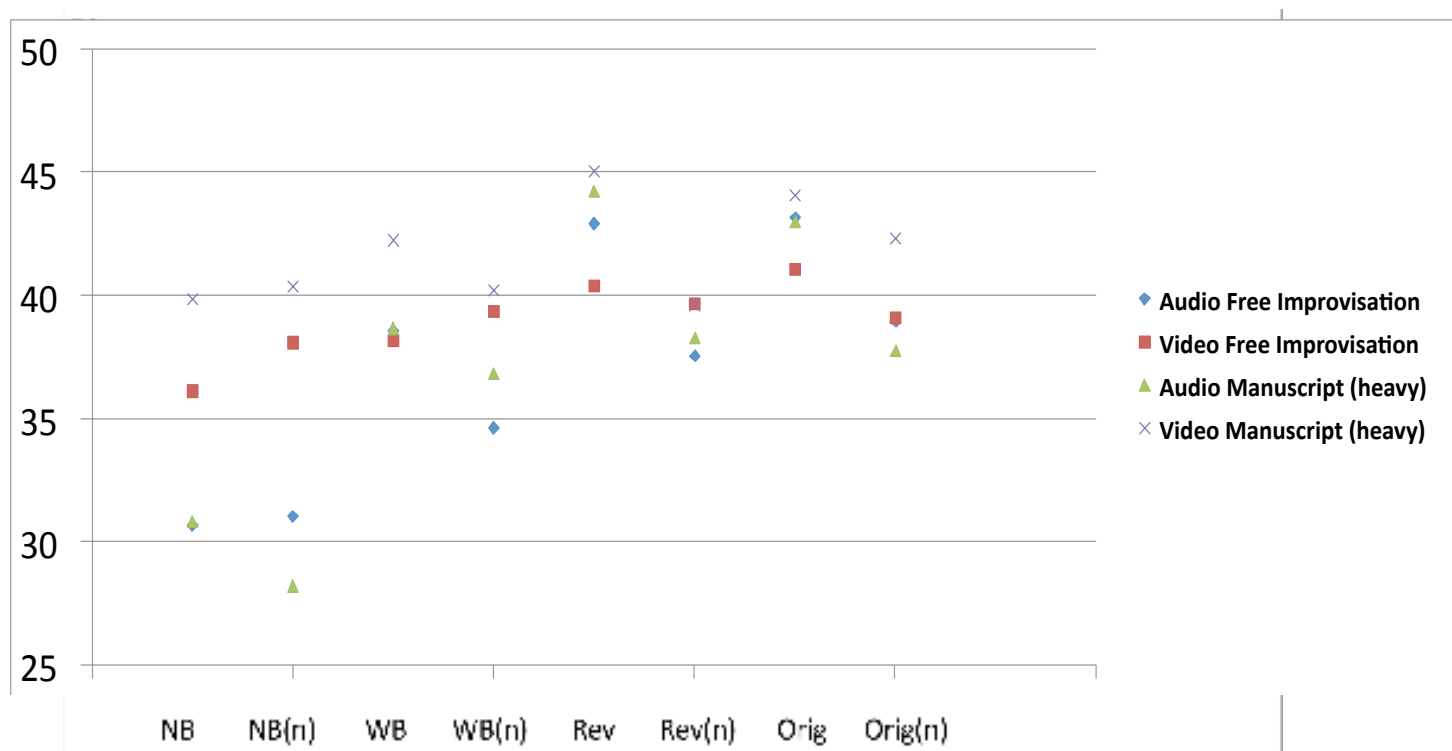
Figure 32. The figure shows an overview of how the audio and video quality separately was perceived for manuscript videos and free improvisation videos. The perceived video quality follows audio quality although it remained unchanged.

This was expected and is supported by the findings in the results for comprehension. Figure 32 shows that the perceived video quality follows the perceived audio quality for both types of videos, as expected. This is the case for all the audio degradations in the free improvisation videos. Interestingly enough, the same clarity is not displayed in the manuscript videos. The lower the audio quality in this category of videos, the less similarity between the perceived video- and audio quality is found. The perceived video quality doesn't start to follow the perceived audio quality until the audio quality is higher. However, the perceived audio quality seems to be less dependent of the content in the videos.

Video FRI	Wrong answers		Accuracy
NB	4	28	85.70%
NB(n)	3	26	88.50%
WB	1	26	96.20%
WB(n)	7	26	73.10%
Rev	4	28	85.70%
Rev(n)	4	28	85.70%
Orig	1	16	93.80%
Orig(n)	4	22	81.80%

Table 33a. Accuracy percentage of the "Free improvisation" videos. The left column lists the audio degradation. The second column from the left lists the number of wrong answers given by participants, followed by the total amount of questions.

Video Manus	Wrong answers		Accuracy
NB	15	32	53.10%
NB(n)	15	34	55.90%
WB	18	34	47.10%
WB(n)	8	34	76.50%
Rev	8	32	75.00%
Rev(n)	18	32	43.80%
Orig	9	42	78.60%
Orig(n)	19	40	52.50%

Table 33b. Accuracy percentage of the Manuscript videos with heavy content.

The tables drawn in figures 33a and 33b show the amount of wrong answers, total amount of questions asked and and an accuracy percentage for both categories, manuscript and free improvisation videos. In the free videos are taken into consideration in the tables, except for the reference videos.

An accuracy percentage was calculated by dividing the number of right answers per question with the total number of answers. The percent shows the amount of correct answers. The accuracy rate was calculated for each audio degrading. The average accuracy percentage for the free videos was 86.31% and for the manuscript videos 60.31%, leaving the latter with a lower score.

Visual questions	wrong	total	
All	60	432	86.00%
Manuscript (heavy)	40	164	76%
Free Improvisation	21	268	92%
All Non-visual questions	76	822	91%

Table 34. The table shows the accuracy percentage of questions regarding visual content. For a deeper interpretation manuscript (heavy) videos have been separated from free improvisation videos. In many cases only one of two questions on a video were about visual content.

The accuracy also decreased when the audio quality increased. In the case of all NB videos, the accuracy increased when noise was present in the videos. This was also the case for manuscript videos in WB. , however, when noise was introduced in the same category the accuracy increased. When audio quality increased (original and reverberated), the accuracy decreased drastically when noise was present in the manuscript videos. In the free videos, accuracy changes weren't as dramatic.

A table was also drawn to measure whether the occurrence of questions about visual content was effecting the accuracy rate in comparison to the questions about what was said in the video. The change was found to be of little importance and thus not taken into consideration further.

Average of Comprehension of Manuscript videos

Comprehension videos 1,2,4 & 5	Comprehension Videos 3 & 6
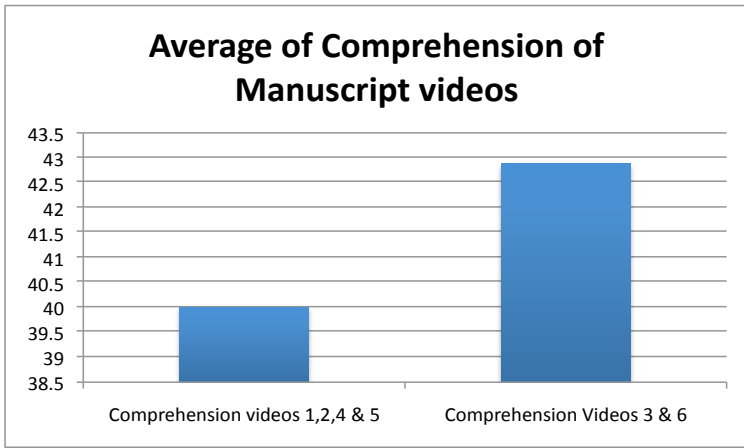
Figures 35a. A comparison has been made of the rated comprehension between different types of manuscript videos. The column on the left shows manuscript videos with heavy content and the column on the right shows manuscript videos with easy content.

Accuracy rate of Manuscript videos

Comprehension videos 1,2,4 & 5	Comprehension Videos 3 & 6
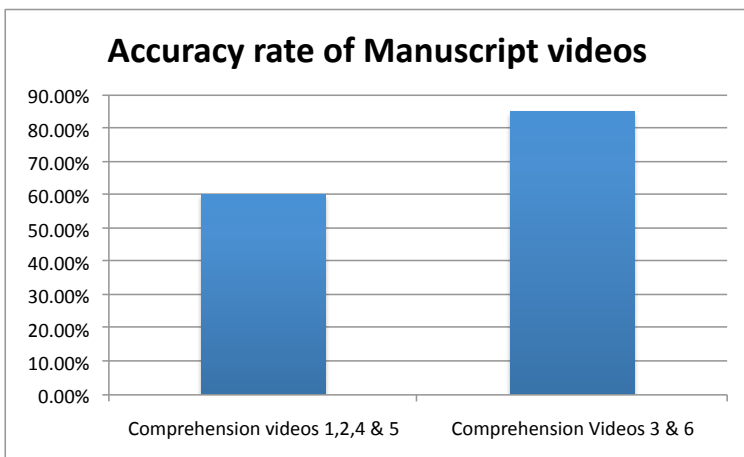
Figures 35b. The table shows a difference in accuracy between manuscript videos with heavy content (left) and and manuscript videos with easy content (right).

5.4 Results of group discussions

All participants agreed on the number of videos in the pre-test being sufficient. Many of them were, however, of the opinion that the pre-test videos did not vary enough in quality. The reference point given at the beginning of the introduction was to regard the videos according to an ordinary video conferencing situation. This was found to be a very good reference by a majority of the participants. They also found that being told how to relate to the background noise was important for their judgment of quality. Many also reported finding it difficult to rate anything as bad, because they thought all videos were at least acceptable. This does concur with the results of the rating.

Only a few of participants reported to have a small interest in audio, but all reported to use this kind of multimedia at least on a weekly basis. A clear majority were of the opinion that testing comprehension together with quality was a good idea as they could relate to it. Most of the participants had participated in audio, visual and audiovisual tests before, which they claimed felt more monotone in comparison to this test.

There was a clear difference of opinion between participants on the matter of comprehension and difficulty to concentrate on the content while rating quality. Some found answering questions very difficult and others had no problem concentrating on both. Participants were then asked whether they think they may have leaned towards one or the other (quality or comprehension) and rated the other similarly without much thought. Only a handful thought this may be possible. A majority of the participants thought the questions on visual content were difficult and led to them paying less attention to the spoken content in the videos although results show that the accuracy rate of the visual questions didn't differ from the other questions. All participants found the manuscript videos to be more challenging than the free improvisational ones.

6 Discussion

6.1 Analysis of results

The variation in audio ratings was the highest throughout the test. This was expected as the audio quality did vary. The least variety was found in the rating of perception of video, which was also expected as the video didn't vary in quality. Although subjects did mentioned that they found content where the speakers moved to be of less video quality, the rating of video quality followed the rating of audio quality, which supports any hypothesis about the two modalities effecting one another.

In all categories NB was rated having poorer video quality than videos with better audio quality, as shown in figure 32. Videos with noise were rated lower in quality than videos without noise within the same degradation. What's interesting is that figures 36a and 36b show that in general, comprehension was rated lower for videos with noise although the accuracy rate was higher in all the NB(n) than in the NB ones, as well as WB(n) compared to WB for the manuscript videos. The results of the WB accuracy for the manuscript videos were unexpected as the accuracy increased drastically when noise was introduced.
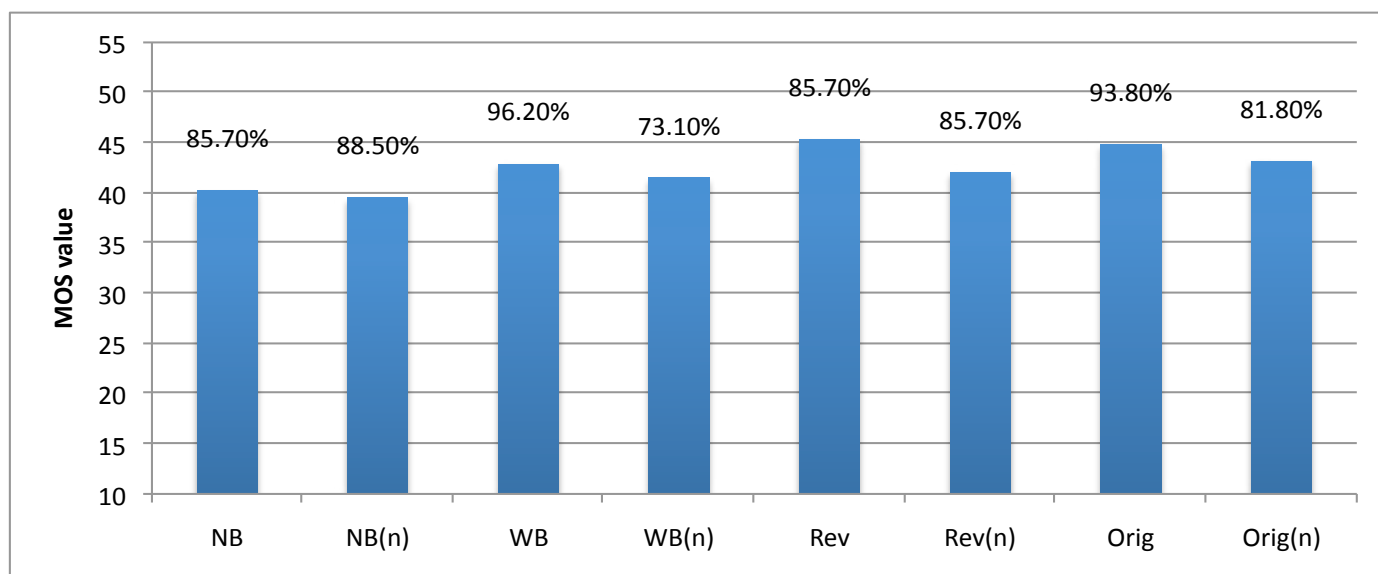
Figure 36a. Accuracy percentage shown on top of mean value of comprehension for votes in each degradation category for the Free Improvisation videos.

Although the reason for this remains unclear, one explanation could be the difficulty of the questions asked. Subjects did find some questions harder than others, although because this was arbitrary, establishing a pattern of this was hard. The more difficult questions may have occurred by chance in the WB degradation category.

Another hypothesis is that this is an indication of the Lombard speech effect- that we make more of an effort to concentrate when there are distractions around, such as noise, especially when the sound source is of poorer quality. As the audio quality is increased, the videos where noise was present had an accuracy rating that is less than the ones without.
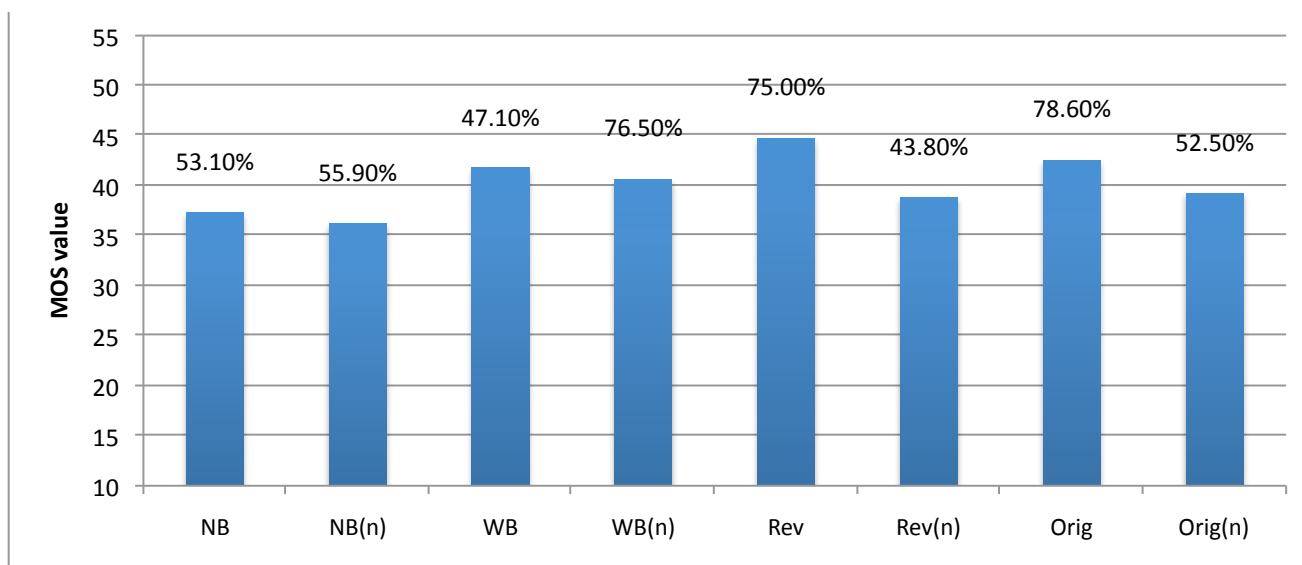
Figures
36b. The table shows the perceived comprehension for each degradation category for manuscript videos with heavy content. The Accuracy percentage is shown on top of each column.
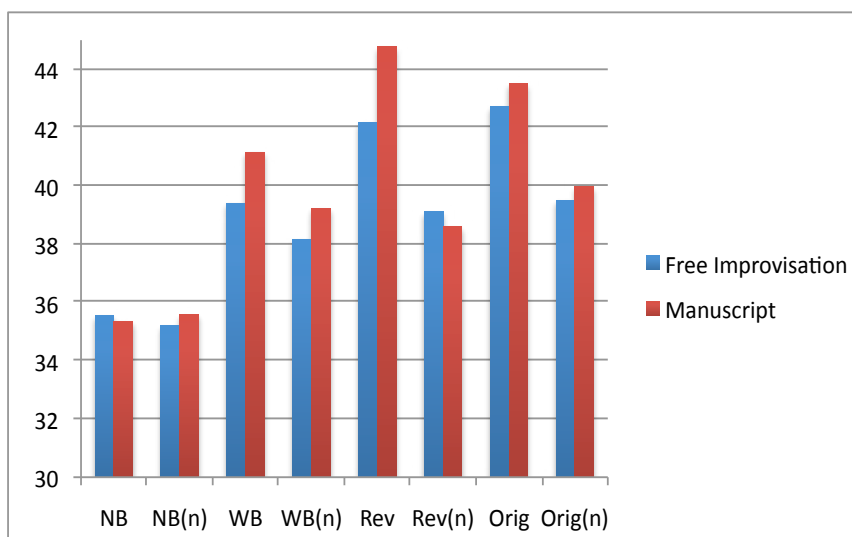
Figure 37. Perceived audiovisual quality for all videos. Manuscript (heavy) videos with reverb are considered to have higher AV quality compared to the free improvisation videos. However, the opposite occurs for reverb with background noise.

The question of effort does rise when taking into consideration that most of the participants thought the manuscript videos (heavy) were boring and therefore more challenging to follow. The free improvisation videos were easier to follow as well as the videos with higher audio quality and thus the effort needed to put into watching the free videos may have been estimated to be less than for watching poorer audio quality and manuscript videos. This may have led to a lack of attention when it came to the "easier videos" where participants may have estimated less of a need to focus.

A closer look at the differences between manuscript and free improvisation videos in figure 37 reveals that the overall AV is rated higher for the manuscript videos in all degradations except in NB and Rev(n). The reason for this general difference in rating is unclear. A hypothesis for this may be that the manuscript videos feel more serious as they contain more information and the presentation is more formal. Thus the perception of audiovisual quality also seems to increase.
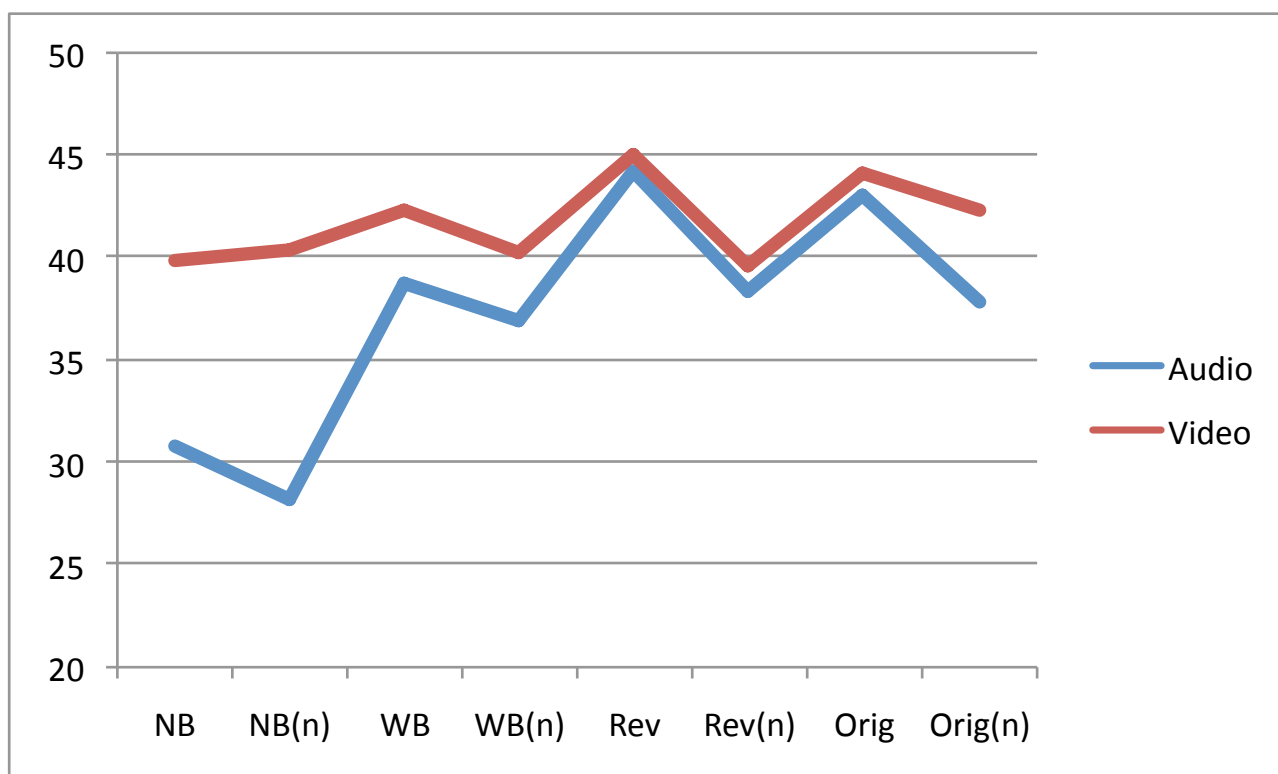
Figure 38. The audio- and video quality of the category "Manuscript (heavy) videos". The perception of video quality seems to fluctuate in accordance to the fluctuation of perceived audio quality.

Also, although the overall AV and Comprehension was rated highest for all Rev videos- even higher than the Orig videos, the same did not apply when noise was present. This supports the findings in the studies by Hodoshima et al. (2010) – where early reflections were considered to increase the attention span, but contradicts the statement about the same applying even in noisy

environments. On the other hand, the noisy environment in the study conducted for this thesis was artificial, as it was only played back in the headphones and the environment very contradictory to the noise in terms of location. These factors may have had an impact on the overall perception of the media.

The main question of the thesis is whether changes in audio can alter the perception in video quality. This is shown quite clearly to be affirmative in figure 38, which shows subjects' perception of audio and video quality in the category Manuscript (heavy). Although the perception of video quality seems to follow that of audio quality (as video quality actually remained unchangedthroughout the test), this pattern is dependent on the content of the video presented.

This is supported by figure 39, which shows subjects' perception of audio and video quality in the category Free improvisation videos. Although the perception of audio and video quality follows a more similar pattern to one another when the audio quality is higher, there is a notable difference in this pattern the more audio quality is degraded.
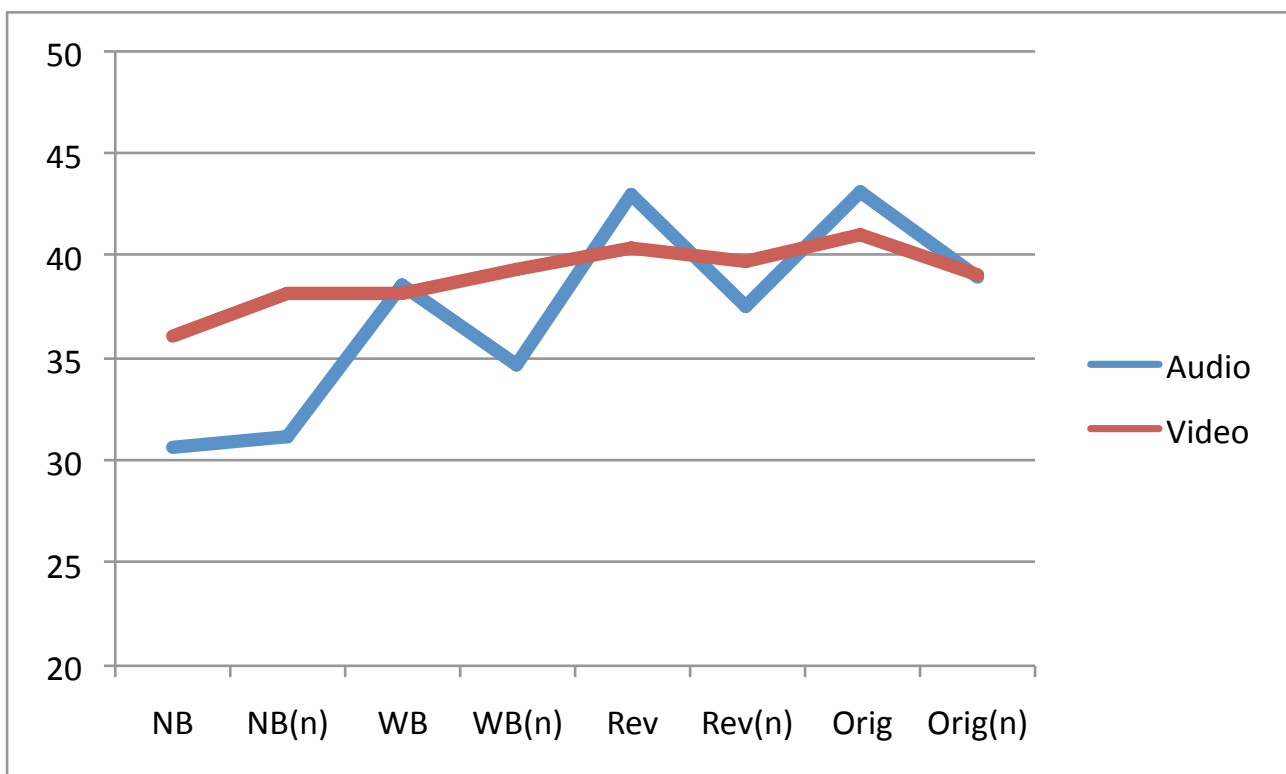
Figure 39. The perceived audio and video quality of the "Free Improvisation videos". The perception of video remains more stable than in the "Free improvisation videos". The better the audio quality, the more the perceived video quality follows that of the audio quality.

This can also be seen in the Manuscript (heavy) videos, but the differences are less dramatic.

It is interesting to note that the two categories are rated so differently. It seems that subjects have perceived the audio quality of the Manuscipt (heavy) videos as poorer when the video has been presented with the poorest quality audio along with background noise. This has to some extent been "forgiven" in the Free Improvisation videos category. A reason for this may be the heaviness of the content in the manuscript category. As subjects stated in the group discussions this category felt much harder to comprehend. Therefore any distractions such as noise or bad audio may have made them lose concentration. Subjects also pointed out that they perceived a difference in video quality between Free improvisation videos and Manuscript videos. As the speakers moved a bit more in the Free improvisation videos, the video quality was perceived as a worse than in the Manuscript videos.

6.2 Conclusion

The question of the thesis is "Can the perceived quality and comprehension of video be effected through audio?". Answering this question has been possible only on a surface level and thus leaves doors open for future investigation on a deeper level.

The conclusion of this study is that the role of audio is extremely important for the majority of purposes we use multimedia and interactive media. The importance varies with content and presentation.

As the study has included many factors based on parameter changes that we are subject to in our daily lives, such as background noise, many hypotheses worth looking into further have emerged. One of these is whether the comprehension and perception of quality of audio are separated in the presence of a reverb. Another one is whether and how comprehension increases when there is noise present in a video with poor audio quality. Most of the findings were expected, such as the perception of video quality following the perception of audio quality, and videos with noise being perceived as having worse audio quality than ones without. What was less expected was that the audio quality was perceived as worse when a reverb was present than when the audio was not degraded, however, simultaneously the comprehension was perceived as better for the reverb videos than the videos with non-degraded audio. This was not supported by the accuracy

rate, which was slightly lower for the reverb videos than for the original videos.

It is worth mentioning that some questions asked about the content in the videos may have been found to be more difficult than others and thus effected the accuracy. Because the difficulty of a question is highly subjective and therefore impossible to measure, only speculations about the reasons for these unexpected results can be made.

Some of the findings have been unexpected, such as the difference in accuracy between WB and WB(n), where WB(n) had a higher accuracy. What has caused these results in accuracy is yet a mystery, but could be the result of questions that have been found hard to answer. In the future, this might be a matter worth looking further in to.

When measuring correlations between parameter changes, multivariate data analysis such as ANOVA (Analysis of Variance) and MANOVA (Multirate Analysis of Variance) is usually performed. This is a difficult process that involves advanced mathematical calculations, it requires dedication and time or a powerful computer program to do the calculations. The limits of the timeframe given for the thesis did not allow this. Therefore the outcome of the results may have been more credible had multivariable data analysis been applied. As this test has been a pioneer of its kind, in the future, it can be run with a larger quantity of people and the results analyzed using MANOVA to ensure a more reliable outcome.

The study has also functioned as a "test run" for Ericsson on how to perform subjective tests. Although research has been done in creating guidelines and standards for subjective tests, very few, if any, have taken comprehension and listening effort into consideration. Subjective tests are hard to set up as there are many aspects and factors to take into consideration. Furthermore test subjects may even perceive audiovisual media differently based on small factors changes such as distance to the screen (Berndtsson et al, 2012). This thesis work is one of the first of its kind and will hopefully in the future only be a small part of a large quantity of quality research in similar areas.

A problem in today's media is the slow development of the relationship between human and equipment. Because media is developed at a faster pace than human perception, the importance of understanding and testing human response to it is increasing. This aspect is sometimes neglected in media usage and media development, but it is important as media is consumed in an increasingly broad variety of circumstances and environments as well as using different equipment. It may

be because of the high number of factors that subjective testing needs to take into consideration that increases the chance of a few ones falling between the cracks.

As audiovisual media is increasingly being consumed in public where distractions such as noise and visual stimuli are louder than in a quiet home environment popular formats of media consumption have changed to better suit phones, pads and other portable media devices. How we interpret, perceive and/or comprehend media in such conditions is worth looking into. After all- the most important thing when mediating information of any kind, whether it be through audiovisual media or in some other way, is to get it right at the receiving end.

7 Acknowledgements

8 References

Beerends, J.G. & De Caluwe, F.E. (1999) "The Influence of Video Quality on Perceived Audio Quality and Vice Versa" , Audio Engr Soc 47:355-362 7

Begnert, F., Ekman,H. and Berg, J. (2011): "Difference between the EBU R-128 meter recommendation and human subjective loudness perception", Convention Paper 8489, 2011, AES].

Berndtsson, G., Folkesson, M. & Kulyk, V. (2012): "Subjective quality assessment of video conferences and telemeetings", Packet International workshop 2012 19th international, Munich, Pages 25-30

Belmudez, B., Moeller, S., Lewcio,B., Raake, A. & Mehmood, A. (2009) "Audio and Video channel impact on perceived audio-visual quality in different interactive contexts", Multimedia Signal Processing, 2009. MMSP '09. IEEE International Workshop , Rio de Janeiro, Pages 1-5

Borowiak, A., Reiter, U. & Svensson, U.P. (2014) "Audio Quality Preferences Measured Under Different presentation conditions", Journal of The Audio Engineering Society. vol. 62 (4)

Bradley, J.S., Sato, H. & Picard, M.(2003) "On the importance of early reflections for speech in rooms", J Acoust Soc Am. 2003 Jun;113(6):3233-44

European Broadcast Union, Recommendation R-128 (2014), "Loudness normalisation and permitted maximum level of audio signals", found at https://tech.ebu.ch/docs/r/r128.pdf

European Broadcast Union (2011) "Ten things you need to know about R-128", [online] available at: https://tech.ebu.ch/docs/events/ibc11-ebutechnical/presentations/ibc11_10things_r128.pdf

Garnier, M. & Henrich, N. (2013) ""Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?"" , Computer Speech & Language Volume 28, Issue 2, March 2014, Pages 580–597
Hodgetts, P. (2008) "The HD survival Handbook", Intelligent Assistance Inc., chapter p.19-25

Hodoshima, N., Takayuki A. & Kurisu, K.(2010), "Intelligibility of speech spoken in noise and reverberation", International congress of acoustics 2010, Australia, vol 1. Pages 3632-3636

Hollier, M.P. & Voelcker, R. (1997) "Objective Performance Assessment: Video Quality as an influence on audio perception", AES convention 103, 1997, UK

Hu,Y. & Kokkinakis, K. (2013) "Effects of early and late reflections on intelligibility of reverberated

speech by cochlear implant listeners", J. Acoust. Soc. Am. 135, EL22 (2014)

Internal Organization for Standardization: "Acoustics- Normal equal-loudness-level contours" (2003) [Online] Available at: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail. htm?csnumber=34222

International Telecommunication Union (ITU-T) Recommendation G.722.2 (2003) [Online] Available at: https://www.itu.int/rec/T-REC-G.722.2-200307-I/en

International Telecommunication Union (ITU-T) Recommendation H-320 (2004): "Narrow-band visual telephone systems and terminal equipment " [Online] Available at: http://www.itu.int/itu-t/recommendations/rec.aspx?rec=H.320

Lavandier, M & Culling, J.F., (2007) "Speech segregation in rooms: Monaural, binaural and interacting effects of reverberation on target and interferer", Acoustical Society of America 2008, Pages 2237–2248

Oyda,P., Czyzewski, A. & Kostek, B. (2001) "Determination of influence of visual cues on perception of spacial sound",110th Audio Eng. Soc. Convention, Amsterdam, Preprint No.5311

Rimell, A. N., Mansfield, N. J. & Hands, D.(2008) ""The influence of content, task and sensory interaction on multimedia quality perception"", Ergonomics
Vol. 51, No. 2, February 2008, Pages 85–97

Rhebergen, K.S., Versfeld, N.J., Dreschler, W.A. (2008): "Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing" , Ear Hear. 2008 Apr;29(2):169-75. doi: 10.1097/AUD.0b013e31816476d

Skovenborg, E. & Nielsen, S.H.(2004) "Evaluation of Designs for Loudness-Matching Experiments", Int. Conf. "Subjective and Objective Assessment of Sound" (SOAS), 1-3 September, 2004, Poznan, Poland

Storms,R.I. & Zyda, M.J. (2001) "Interactions in perceived quality of auditory- visual displays" Presence (Volume:9 , Issue: 6), Pages 557 - 580

Susini, P., Misdariis, N., Lemaitre, G. & Houix, O. (2012) "Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback", J Multimodal User Interfaces
DOI 10.1007/s12193-011-0086-0

Suzuki, Y. & Takeshima, H. (2004) "Equal-loudness-level contours for pure tones (acoustical society of America), J. Acoust. Soc. Am. 116, 918

Thoma, H. (2012) "A system for subjective evaluation of audio, video and audiovisual quality using MUSHRA and SAMVIQ methods, Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on, Pages 31 - 32

Van Hurkman, A. (2014) "Color correction handbook: Professional Techniques for Video and Cine-

ma", 2nd ed., Parson Education: Peachpit Press, ISBN-13: 978-0-321-92966-2

Wikipedia (2015) "Narrowband" [Online] Available at: https://en.wikipedia.org/wiki/Narrowband

Wikipedia (2015) "Wideband Audio" [Online] Available at: https://en.wikipedia.org/wiki/Wideband_audio

Wikipedia (2015)"4K resolution" [Online] Available at: https://en.wikipedia.org/wiki/4K_resolution

Winkler, S. (2005) "Digital Video Quality- Vision models and metrics", John Wiley & Sons, Ltd, England, ISBN 0-470-02404-6

Wong, L.L.N., Ng, E.H.N. and Soli, S.D.(2012): "Characterization of speech understanding in various types of noise", Journal of the Acoustical Society of America, 2012, v. 132 n. 4, Pages 2642-2651

Zollinger, S.A. & Brumm,H. (2011) "The Lombard effect", Current Biology, Vol.21, Issue 16, Pages R614-R615 p.R614-R615

APPENDIX

Instruktioner för testet

Testet är ca 40 minuter långt och består av två sessioner med en längre paus utanför testrummet. Du kommer att få titta på ett antal videosekvenser (12 st. per session), besvara frågor om innehållet och bedöma hur du upplever kvalitén i videosekvensen. I dessa videosekvenser pratar olika personer (en för var video) i ca 45–50 sekunder. Innehållet varierar mellan vardagsinformation och fri monolog. Vissa videosekvenser innehåller även inspelat bakgrundsljud.

Efter varje videosekvens får du svara på några frågor angående innehållet. Pappret du får framför dig är avsett för att kryssa i ditt svarsalternativ för frågorna i videon. Antalet svarsalternativ på pappret behöver inte motsvara svarsalternativen i videon.

Stora skärmen

1. Video visas på stora skärmen. Efter videon visas en fråga i taget i 10 sekunder. (2 frågor per video)
2. Kryssa i ditt svar på svarsblanketten.
3. Rösta på röstningsterminalen. Tryck PÅ slidern- inte bredvid. Du har 20 sekunder på dig att rösta.

Efter att röstningstiden gått ut visas nästa videosekvens.

Röstningsterminal   Svarsblanketter

När du röstar skall du markera ditt svar på touch-displayen framför dig (den lilla skärmen).
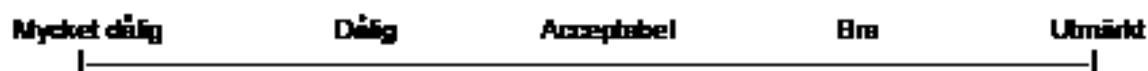Dessa frågor kommer du att få rösta på:

Fråga 1: Hur väl gick det att uppfatta innehållet?
Fråga 2: Hur bedömer du audiovisuella kvalitén?
Fråga 3: Hur bedömer du audiokvalitén?
Fråga 4: Hur bedömer du videokvalitén?

Din bedömning ger du genom att använda följande skala:

Mycket dålig	Dålig	Acceptabel	Bra	Utmärkt

Betygsslidern är från start alltid inställd med lägsta möjliga betyg. Även om du anser att videokvaliteten motsvarar det lägsta betyget måste du klicka på slidern för att betyget ska registreras. När du klickar på slidern ska den ändra färg. Kontrollera att den gör det! Nästa videosekvens spelas upp när röstningstiden (20 sekunder) gått ut.

Innan första sessionen startar kommer 3 videosekvenser att spelas upp, så att du vänjer dig vid testsituationen. Efter det finns det möjlighet att ställa frågor innan själva testet börjar.

När testsessionerna är slut blir du informerad om detta via röstningsterminalen.

Efter hela testet kommer du att få besvara några frågor om hur du upplevde testet.

Var vänlig diskutera inte dina åsikter om kvalitén under testet med andra testdeltagare!

Tack för din medverkan!

VIDEO 1

FRÅGA 1

- [ ] A
- [ ] B
- [ ] C

FRÅGA 2

- [ ] A
- [ ] B
- [ ] C

VIDEO 2

FRÅGA 1

- [ ] A
- [ ] B
- [ ] C

FRÅGA 2

- [ ] A
- [ ] B
- [ ] C