Subjective Assessment of Video Quality on Audio Quality

Sondra Moyls

Mus 406A
Mr. Kirk McNally
December 17, 2013

Table of Contents

Sondra Moyls
Kirk McNally
Mus 406A
17 December 2013

Subjective Assessment of Video Quality on Audio Quality

0 Introduction

This paper seeks to explore how the quality of video influences the perception of audio quality, particularity in regards to encoding of digital media. While various studies have evaluated the transparency of audio codecs, few have considered a secondary stimulus. In a 1999 study published by the AES, participants were asked to evaluate the quality of two commercials presented as 32 separate stimuli, each with a different pairing of audio and visual quality. They found that as the video quality diminished from full resolution to a video band-limited to 0.025 MHz, participants perception of audio quality diminished by 1.2 points on a 9 point scale, and concluded that video resolution played a significant role in the evaluation of audio quality (Beerends 358). In a similar 2001 study by the TDF-C2R Broadcasting and Wireless Research Center of Télédiffusion de France, participants were asked to evaluate changes in the audio and video quality of a 30 minute news broadcast. In contrast to the 1999 study, they found that video degradation had no effect on the participants' perception of audio quality (Joly et al., 441). In both experiments, the stimuli were presented on a professional grade TV monitor and loudspeakers.

In a 2012 survey of 3000 teenagers, The Nielsen Company found that 64% listened to music through YouTube, which offers video resolutions between 144p and 1080p ("Neilson 360 Report"). As more people find music and media online, how can these resolutions impact the perceived quality of the audio, particularly for music? The following double-blind study seeks to explore how video quality can impact the audio quality of a music track using the MUSHRA

(MUlti Stimulus Hidden Reference Assessment) evaluation method, as recommended in ITU-R BS.1532-1.

1 The Experiment

1.1 Set-up

The source material consisted of a folk rock song chosen to represent a spectrum of frequencies and instruments without added effects that might skew the impression of audio quality. To create five different tests, the audio track was divided into five 20 second clips, each from a different section of the song, and paired with the corresponding video at a given resolution, either 720p, 480p, 360p, or 144p. One of the tests was presented without video. For each test, the audio was presented six times at different qualities: a wav file, and five mp3 files encoded at different bit rates (Figure 1). In total 30 stimuli were evaluated by the participants.

Test Number	**Video Quality**	**Audio Quality**
1	No Video	.wav (Reference) mp3 (256kbps) mp3(128kbps) mp3 (96kbps) mp3 (64kbps) mp3 (56kbps)
2	720p Recommended bit rate: 2500kbps	.wav (Reference) mp3 (256kbps) mp3(128kbps) mp3 (96kbps) mp3 (64kbps) mp3 (56kbps)
3	480p Recommended bit rate: 1000kbps	.wav (Reference) mp3 (256kbps) mp3(128kbps) mp3 (96kbps) mp3 (64kbps) mp3 (56kbps)
4	360p Recommended bit rate: 750kbps	.wav (Reference) mp3 (256kbps) mp3(128kbps) mp3 (96kbps) mp3 (64kbps) mp3 (56kbps)

Test Number	Video Quality	Audio Quality
5	144p	.wav (Reference) mp3 (256kbps) mp3(128kbps) mp3 (96kbps) mp3 (64kbps) mp3 (56kbps)

Figure 1. Material Presented in Each Test.

Participants were presented each of the five tests twice, in a random order (for a total of ten tests). Within each test, the audio was also presented in a random order to prevent bias or ordering effects. Participants were able to listen to each of the stimuli as many times as they wanted, and in any order. They were also given a labeled reference, which played the wav file. Comparing each file to the reference, participants evaluated their perception of the audio quality on a scale from 0 to 100 (Figure 2). Participants could only adjust the slider of the file they were currently listening to.

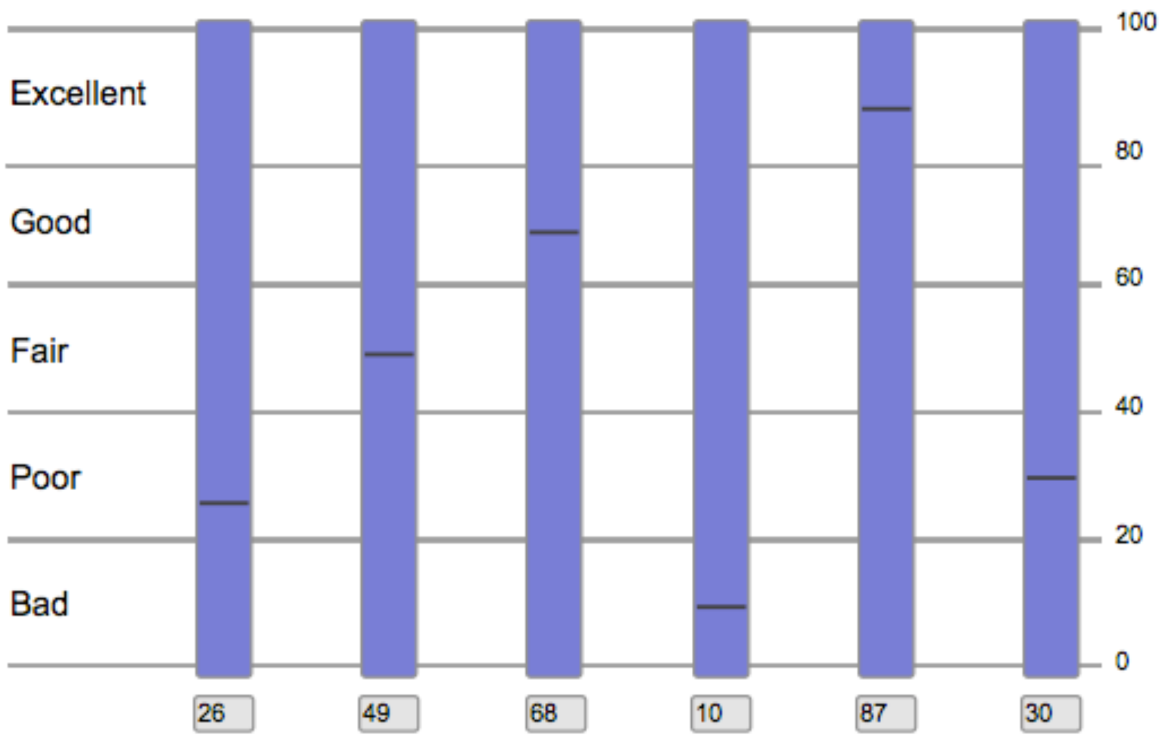

Figure 2. MUSHA Subjective Evaluation Scale

For each test, the video was presented to the left of the sliders in 640x360 resolution (Figure 3). The test was performed on a 13-inch MacBook Pro (mid 2012) with Intel HD graphics and the highest brightness setting. All tests were conducted in the same room, which contained no natural light and moderate level lighting. Audio was presented in stereo through headphones (Sennheiser HD 280 Pro). Participants were asked not to alter the volume of the files.

Figure 3. Test Screen Setup

1.2 Software Used

The testing interface uses a modified version of the MUSHRA Max/MSP patch originally developed by Christopher Hummersone at the University of Surrey, Australia. These modifications include the use of jitter objects so that the patch can process video files instead of audio files. Instead of looping the files continuously and muting the files that aren't selected, the modified patch starts and stops the jitter window each time a stimulus is selected. The first test starts as soon as the reference (R) is selected, and for all following tests, the reference starts immediately.

Video files were obtained from YouTube using ClipConverter to save the files in the various resolutions available. The wav file was ripped from CD and saved as mp3s at various bit-rates using LAME (http://lame.sourceforge.net/) for Audacity. Audio files and videos were synced using Logic Pro 9 at 48kHz using the highest render setting.

1.3 Participants

Ten participants between the ages of 19 and 32 and from a musical and non-musical background took part in this study. Each participant was briefed on the goal of the study and the functions of the test software. The average time to complete the test was 15 minutes.

2 Results

2.1 Mean Data

For each participant, each of the five tests were presented twice. The values assigned were then averaged for each participant. For all ten sets of data, the mean score for each stimulus was calculated (Figure 4).

	No Video	**720p**	**480p**	**360p**	**144p**
.wav	78.6	70.5	74.5	69.6	76.1
.mp3, 256kbps	81.8	71.2	74.7	77.9	81.4
.mp3, 128kbps	78.6	83.1	76.5	81.5	69.9
.mp3, 96kbps	64.6	71	73.4	68.4	73.7
.mp3, 64kbps	41.1	35.8	32.3	33.7	29.2
.mp3, 56kbps	23.1	23.4	22.7	12.1	10.7

Figure 4. Mean Scores Given to Each Stimuli

Figure 5 shows the mean subjective rating for each audio file. For all video files, the 64kbps and 56kbps audio files received significantly lower subjective ratings that for the other four audio files. When no video was presented and when a low resolution, 144p, video was presented participants showed the highest preference for the 256kbps mp3 file and the lowest for the 56kbps file. When a 720p, 480p, or 380p video was presented, participants showed the highest preference for the 128kbps mp3 and lowest for the 56kbps file. In no situation was the wav file preferred.

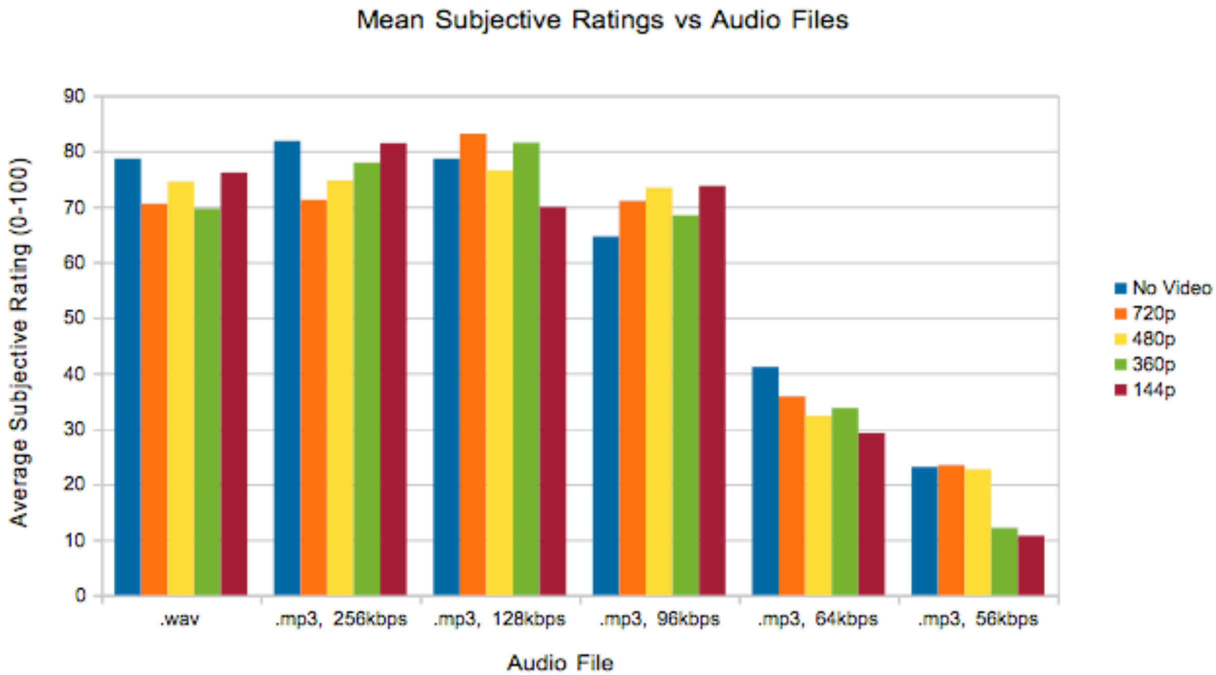

Figure 5. Subjective Ratings of Audio Files for Each Set of Videos

The smallest difference in subjective rating was for the 480p video, differing by only 1.3 points (out of 100) for the 4 highest quality sound files. When no video was presented, the subjective rating differed the most, by 12 points. Both the 64kbps and the 56kbps audio files were given lower ratings as the resolution of the video decreased, dropping by 10.9 and 12.4 points respectively.

2.2 Evaluation

Overall, the results show a change in the subjective rating depending on the video file presented.

Many online audio/video streaming services such as YouTube, Blip TV, and Vimeo use 128kbps as the standard audio bit rate. The preference given to the 128kbps file for 720p, 480p, and 360p resolutions may suggest a preference for a sound that is most commonly associated with streaming video.

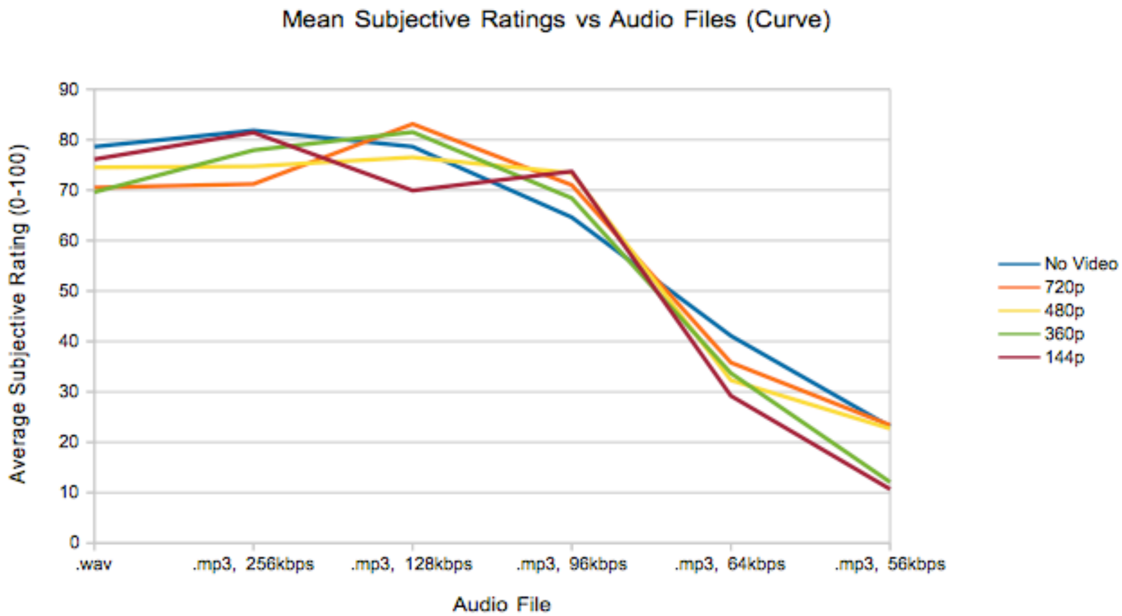

Figure 6. Subjective Ratings of Audio Files - Curve

Removing the 128kbps file from the results, it's easier to see a general curve in the evaluation of the audio quality for each video resolution (Figure 7). For the reference wav file and the 256kbps mp3, the highest ratings were given when there was no video and when the 144p video was presented. This seems to make sense, as the 144p resolution is extremely low, and may be easy to disregard when it is so far detached from the audio quality; however, the subjective quality of the 96kbps audio file was highest when the 144p was presented, and lowest when there was no video. The reverse is true for the 64kbps audio file, and for the 56kbps file. This suggests that for lower quality audio, the quality of the video can directly impact the preserved quality, either making it appear better or worse. Further experimentation with bit rates between 128kbps and 64kbps may reveal a threshold where picture quality boosts or reduces the subjective quality, as is seen in the results at 96kbps.

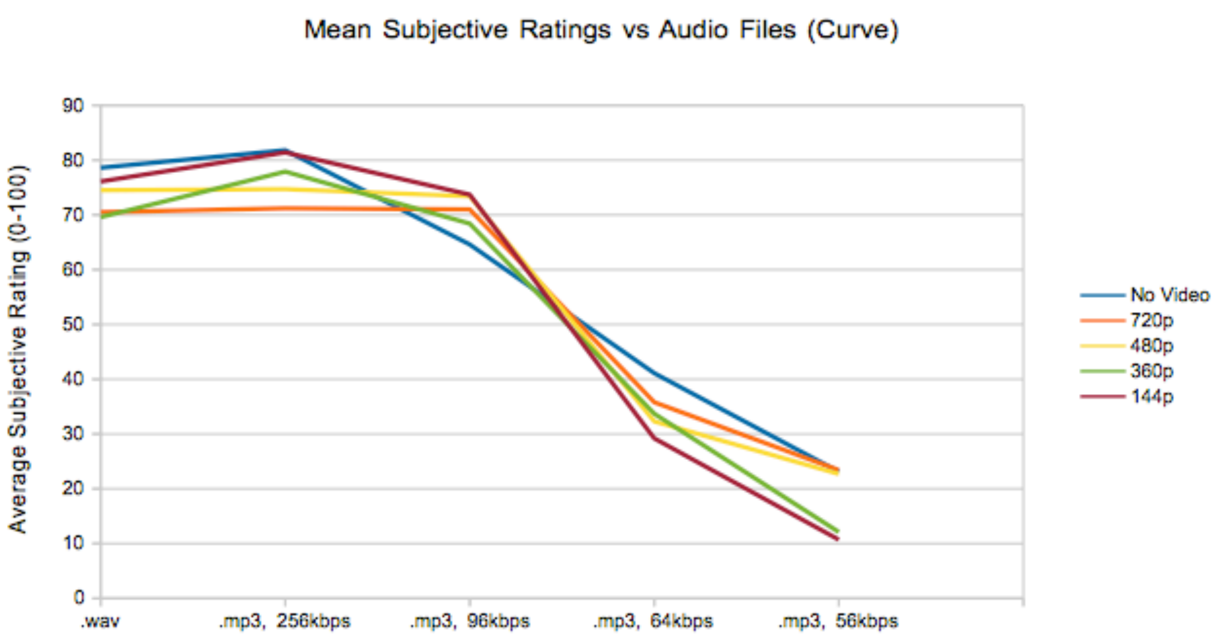

Figure 7. Subjective Ratings of Audio Files (without 128kbps)

3 Future Considerations

Future investigations into the effects of video quality on subjective audio quality could benefit from a larger and more diverse group of participants. For this study, a different section of the song and music video were used for each test, and it's possible some of the tests were easier or more difficult to evaluate than others, depending on the auditory information in the clip presented. Future tests may benefit from multiple genres of music being tested, or different types of videos.

The size of the video resolution could be altered in future tests. Pixelation caused by jitter should be investigated if Max/MSP is to be used as a platform.

This test used LAME to encode mp3s, but the current standard encoder for streaming is ACC-LC.

4 Conclusion

Video quality does appear to impact the perceived quality of audio. For standard internet video resolutions (720p, 480p, 360p), preference is given to 128kbps audio file. For high quality audio files, the subjective quality is highest when no video is present, or when the video is very low quality. For low quality audio files, the subjective quality is generally must lower, and degrades with the quality of the picture. The reversed subjective ratings for video with audio encoded at 96kbps should be investigated further.

5 References

"Advanced encoding settings." YouTube Help. N.p., n.d. Web. 16 Dec. 2013. <https://support.google.com/youtube/answer/1722171?hl=en>.

ITU-R BS.1532-1 "Method for the subjective assessment of intermediate quality level of coding systems", 2001-2003

Beerends, John G. and Frank E. De Caluwe. "Influence of Video Quality on Perceived Audio and Vice Versa" Journal of the Audio Engineering Society Vol. 47 No.5 (1999).

Bosi, Marina, and Richard E. Goldberg. *Introduction to digital audio coding and standards.* Boston: Kluwer Academic Publishers, 2003.

Ekeroot, Jonas, Jan Berg, and Arne Nykänen. "Selection of audio stimuli for listening tests" Presented at the Audio Engineering Society 130th Convention, London, UK. (2011). Paper number 8445.

Hummersone, Christopher. *MUSHRA Test GUI.* Computer software. *Institute of Sound Recording: Software and Digital Resources.* University of Surrey (2011). Web. 12 December 2013. <http://iosr.surrey.ac.uk/software/index.php>

Link, Martin. *Internet Audio Quality and the Mushra Test.* Audio Delivery The Changing Home Experience - Audio Engineering Society 17th UK Conference (2002): 91-96

"Music Discovery Still Dominated by Radio, Says Nielsen Music 360 Report." Web. 16 Dec. 2013. <http://www.nielsen.com/us/en/press-room/2012/music-discovery-still-dominated- by-radio--says-nielsen-music-360.html>.

Joly, Alexandre, Nathalie Montard, and Marcel Buttin. "Audio-Visual Quality and Interactions Between Television Audio and Video" Signal Processing and its Applications, Sixth International, Symposium Vol. 2 (2001): 438 - 441.