The Influence of Individual Audio Impairments on Perceived Video Quality

Leslie Gaston[1], Jon Boley[2], Scott Selter[3] and Jeffrey Ratterman[3]

[1] University of Colorado Denver, Denver, CO, 80217, USA
Leslie.Gaston@ucdenver.edu

[2] LSB Audio LLC, Lafayette, IN, 47905, USA
Jon@lsbaudio.com

[3] Master of Science in Recording Arts Candidates
University of Colorado Denver, Denver, CO, 80217, USA
Scott.Selter@email.ucdenver.edu and Jeffrey.Ratterman@email.ucdenver.edu

ABSTRACT

As the audio, video and related industries work toward establishing standards for subjective measures of audio/video quality, more information is needed to understand subjective audio/video interactions. This paper reports a contribution to this effort that aims to extend previous studies, which show that audio and video quality influence each other and that some audio artifacts affect overall quality more than others. In the current study, these findings are combined in a new experiment designed to reveal how individual impairments of audio affect perceived video quality. Our results show that some audio artifacts enhance the ability to identify video artifacts, while others make discrimination more difficult.

1. INTRODUCTION

Previous studies have established that certain audio impairments are more annoying than others [1] and that audio quality does have an effect on video quality [2]. This study combines these findings in an effort to discover whether particular audio artifacts have a larger effect on perceived video quality.

For example, does band limitation of the audio make a picture seem less bright or vivid? Do "birdies" contribute to the annoyance of video pixelation?

These questions are compelling: however, each potential pairing of video and audio artifacts yields an unwieldy number of combinations to explore for a single test. In order to conduct a preliminary study of a

more manageable size, a test based on Signal Detection Theory (SDT) is used. For this test, we limited ourselves to discovering whether a given audio artifact results in a perceived change in video. With SDT, it is possible to determine with some confidence whether the test population is able to tell the difference, just guessing at the changes, or whether there are some interesting "False Alarms" that cause people to suppose that the video has changed (when in fact only the audio changed).

There are standards for subjective tests to determine audio quality, and standards for the subjective evaluation of video quality. Some subjective tests have even been done to determine the impact of audio quality on video programming [3], [4], [5]. However, there are no standards for subjective tests to measure audio/video quality. Such procedures are being developed in ITU Study Group 9 (Question 12, 2009-2012) [6].

Therefore, it is our intention that the results of this test will inform the audio and video communities at large of the potential impact of individual impairments as we move toward improving the audio/visual experience for broadcast, Internet Protocol Television (IPTV), and any application where the quality of the combined experience is of utmost importance.

2.   TEST METHODS

We would like to determine if audio impairments cause a perceived change of video quality. Our test methods depart slightly from previous methodologies. We are not determining quality on an impairment scale; rather, we seek only to find whether a given artifact causes any perceived change. (Further testing will be necessary to determine how and to what degree the picture was perceived to change.) As we will demonstrate, different audio impairments do have varying effects.

2.1.   Impairments under test

The audio impairments chosen for the test are a high frequency shelf at + 6 dB with a crossover frequency of 5 kHz, a high frequency shelf at - 6 dB with a crossover frequency of 5 kHz, and a low bit rate, MPEG-1 Layer III (MP3) encoding of 56 kbps (constant bit rate).

The types of video changes our test subjects were asked to identify during testing were: a brightness gain of +9

(which corresponds to the "Brightness and Contrast" QuickTime filter in Apple's Final Cut Pro), desaturation of -.31 to -.55 (which corresponds to a linear reduction of color (RGB) in the amount of 31 to 55 percent, based on a scale of -1 to 0, where 0 represents no change in color value), and pixelation, performed by reducing the video resolution from 720x480p anamorphic widescreen to 570x380p and 675x450p, both anamorphic 16x9, and then re-exporting as a 720x480p widescreen file. This effectively reduces the original quality by 20.83% and 6.25%, respectively.

These video settings were chosen to be above the perceptual threshold, but not detectable with 100% accuracy. By picking points between threshold and saturation on the psychometric function, our goal was to identify both increases and decreases in sensitivity.

The test sequences consist of one the following pairs being showed before or after a reference clip ($A_R V_R$):

$A_R V_R$ = Audio Reference / Video Reference
$A_n V_R$ = Audio Impaired / Video Reference
$A_R V_n$ = Audio Reference / Video Impaired
$A_n V_n$ = Audio Impaired / Video Impaired

Table 1: This table shows the four types of combinations possible in one audio/video clip.

For each audio/video clip that is presented, three audio impairments (A_1: high shelf cut [HSTrim], A_2: high shelf boost [HSGain], and A_3: low bit rate mp3 [MP3]) and three video impairments (V_1: desaturation [desat], V_2: brightness gain [bright], and V_3: pixelation [pixel]) are used. This brings the number of combinations for each clip to 16. The study is based on two clips, a movie scene (from *Batman Returns* [7]) and a live concert scene (from *Foo Fighters: Live at Wembley Stadium* [8]). Both clips were edited to 20 seconds in an effort avoid fatigue in our 26 viewing subjects.

2.2.   Test procedure

The test subject views a pair of audio/video clips, played one after the other, separated by a two-second pause of black. In each case, one is the reference and the other is either the reference again or an impaired version. After viewing the two examples, the subject is asked "Did the video change? Yes or No". The clips are

presented randomly, so either the reference or the impaired clip plays first, and each test subject sees the pairings in a different order than other subjects.

The subjects' answers to "Did the video change?" are categorized as follows:

Correct Rejection: The answer to "Did the video change" should be "No". If the subject answers "No", then that is a "correct rejection" for the pairing, where:

- $(A_R V_R$ paired with $A_R V_R$, written $A_R V_R$ / $A_R V_R)$: No Change
- $(A_n V_R$ / $A_R V_R)$: Audio Change

False Alarm: The answer should be "No". If the subject answers "Yes", then that is a "false alarm", where:

- $(A_R V_R$ / $A_R V_R)$: No Change
- $(A_n V_R$ / $A_R V_R)$: Audio Change

Miss: The answer should be "Yes". If the subject answers "No", then that is a "miss", where:

- $(A_R V_n$ / $A_R V_R)$: Video Change
- $(A_n V_n$ / $A_R V_R)$: Audio-Video Change

Hit: The answer should be "Yes". If the subject answers "Yes", then that is a "hit", where:

- $(A_R V_n$ / $A_R V_R)$: Video Change
- $(A_n V_n$ / $A_R V_R)$: Audio-Video Change

If it is true that an audio change for a given artifact can cause a perceived video change, then we should see a significant change in the number of hits relative to false alarms. However, other pairings are considered as interesting as well (see section 5, "Results").

2.3.   Subjective Testing Standards

There are standards for subjective testing for audio and separate standards for video. However, we are not testing audio quality. Furthermore, in the early stages of this test, we played the audio from full-range, nearfield speakers as the video played on a television screen. Our team felt that this "divorced" the audio from the picture, and would allow our subjects to mentally separate the audio experience from the viewing experience with ease. Therefore, we decided to use the speakers that are built-in to the set. Although we cannot prove whether the average television viewer

listens on separate loudspeakers, we felt that for the purposes of this test, a more integrated audio/visual experience was desirable.

For the video, we adopted the standard viewing distance articulated in ITU Recommendation Rec. ITU-R BT.500-11 [9]. For picture calibration, we used the Digital Video Essentials (DVE) Test and Demonstration Materials Disc.

Viewers were seated 3.5 feet from the television, which closely represents the *3H* (3 times the screen height, in our case 13 inches) preferred viewing distance set forth by BT.500-11.

The playback volume was set in order to get an average of 78 dB SPL, according to the television mixing standard [10]; see section 2.5, "Critical Material", for further details.

2.4.   Equipment List

We used a 26-inch, 16x9 Toshiba LCD Television (26LV610U) for our viewing test. For playback of the sequences, we used Final Cut Pro (v.6.0.6) through a Canopus ADVC-100, which converts the digital video and audio signals to analog. The analog connections were made to the "Video 1" input of the television.

2.5.   Critical Material

A scene from *Batman Returns* was chosen for its very dark picture. Some Hollywood movies have a very dark look, which is often seen slightly degraded on television when the films are broadcast. The clip we chose also features music and a loud explosion. The audio ranges from 55 dB SPL at its quietest to 80 dB SPL during the explosion.

The *Foo Fighters* clip features a brightly lit, indoor scene of live, "rock-n-roll" concert music with an average of 78 dB SPL. It was chosen to be representative of music programming, which offers some unique audio challenges (annoyances) when broadcast with low bit rate encoding, especially when a hi-hat cymbal is present (as in this clip).

2.6. Test Procedure and Interface

Test subjects watched 32 different sequences:

01. Batman reference ($A_R V_R$)
02. Batman-Bright ($A_R V_2$)
03. Batman-BrightHSGain ($A_2 V_2$)
04. Batman-BrightHSTrim ($A_1 V_2$)
05. Batman-BrightMP3 ($A_3 V_2$)
06. Batman-Desat ($A_R V_1$)
07. Batman-DesatHSGain ($A_2 V_1$)
08. Batman-DesatHSTrim ($A_1 V_1$)
19. Batman-DesatMP3 ($A_3 V_1$)
10. Batman-HSGain ($A_2 V_R$)
11. Batman-HSTrim ($A_1 V_R$)
12. Batman-MP3 ($A_3 V_R$)
13. Batman-Pixel ($A_R V_3$)
14. Batman-PixelHSGain ($A_2 V_3$)
15. Batman-PixelHSTrim ($A_1 V_3$)
16. Batman-PixelMP3 ($A_3 V_3$)
17. Foo reference ($A_R V_R$)
18. Foo-Bright ($A_R V_2$)
19. Foo-BrightHSGain ($A_2 V_2$)
20. Foo-BrightHSTrim ($A_1 V_2$)
21. Foo-BrightMP3 ($A_3 V_2$)
22. Foo-Desat ($A_R V_1$)
23. Foo-DesatHSGain ($A_2 V_1$)
24. Foo-DesatHSTrim ($A_1 V_1$)
25. Foo-DesatMP3 ($A_3 V_1$)
26. Foo-HSGain ($A_2 V_R$)
27. Foo-HSTrim ($A_1 V_R$)
28. Foo-MP3 ($A_3 V_R$)
29. Foo-Pixel ($A_R V_3$)
30. Foo-PixelHSGain ($A_2 V_3$)
31. Foo-PixelHSTrim ($A_1 V_3$)
32. Foo-PixelMP3 ($A_3 V_3$)

Table 2: File name of sequence and combination of impairments ($A_x V_x$)

The order of the sequence was randomized for each viewer. In addition to these 32 clips, two extra instances of the reference files for the *Batman* and *Foo Fighters* clips were mixed in with the other impairments. (This meant the viewer had to rate 36 items, although only 32 files were used). The viewer would only rate 12 items at a time to prevent fatigue, taking breaks of 10 – 15 minutes.

2.7. Test Population

Students and faculty of the University of Colorado Denver's College of Arts and Media were invited to participate in the experiment. Majors in the College include Visual Arts, Music, Theater, Video, and Recording Arts. Of the 26 subjects, 20 were majors in Recording Arts. Five of our subjects were female. Sixteen subjects were between the age of 20-29, five were between 30-40, two were between 40-50, and three were between 50-65 years old.

2.8. Training

For the training session, subjects were seated in front of the LCD television. The full training script appears in the Appendix. A "practice clip" from the movie *Earth* [11] was shown of a woodsy, nature scene with ducklings and their mother. First, viewers watched the reference clip. Next, they were shown the same clip with a brightness gain filter added. Viewers were then asked to articulate any difference they perceived. Responses about the affected clip included terms such as "bright" and "washy".

When subjects were shown the reference next to the desaturation clip, comments included terms such as "dull" and "pale". Finally, terms such as "blurred" and "out of focus" were used to compare the reference clip to the resolution-affected clip.

With these comments, the research team was satisfied that our test subjects could perceive a difference between a reference clip and one with a change in brightness, saturation, or resolution (pixelation).

Test subjects were not trained on recognizing audio artifacts, because the goal of the test was simply to see if these artifacts had an impact on what viewers visually perceived. Unlike an audio quality test, this experiment is not concerned with recognizing or judging audio quality, only with whether changes in audio affect a perceived change in video.

2.9. Participant comments

There were many interesting comments made by the test subjects that may shed light on some of the data and analysis presented later in Sections 3, 4, and 5.

Concerning performance over the course of the test, one participant observed, "it was easier to remain focused

and objective at the beginning of the test" and "I felt like my answers were probably stronger at the beginning." (This is opposite of what the results showed, as subjects' scores improved in measureable ways, and does not present a bias as the clips were randomized). See Section 5.3.

Although our subjects were not advised to look for "clues" or "reference points", it is natural to assume that each person would develop their own way of recognizing differences during the course of the test. Here are some things the participants said:

- "The bad mp3 audio was the most detrimental to the viewing experience."
- "I could totally tell when the *Foo Fighters* was distorted or gray washed."
- "I found myself looking for reference points."
- "It seemed to me that the *Foo Fighters* clips had less of these reference points."

We also received feedback that strengthens the validity of the answers we pulled from a largely homogeneous group of people. Subjects brought differing opinions about the way they perceived the changes between the two test videos:

- "The *Batman* clips were easier to tell apart than the *Foo Fighters* ones" and likewise, "The *Foo Fighters* sequences were a little harder than the *Batman* ones."
- "I could not decipher much of a change between most *Batman* clips."

Other observations from our subjects included:

- "(I) was intrigued at how the sound affected my visual perceptions."
- "It was tough to tell if the video had changed or if it was just the audio."
- "Sometimes I thought that the video quality had gone down just a little bit, but I couldn't tell if it was just my brain playing tricks on me."

3.   RESULTS

A preliminary look at the results confirms that not only is there is a link between audio impairments and a perception of video quality, but also confirms that not all audio impairments have the same effect.

Because two different video clips were used (*Batman Returns* and the *Foo Fighters* concert*),* the data is separated accordingly:

Percent Correct: *Batman*		Video			
		Ref (V_R)	Desat (V_1)	Bright (V_2)	Pixel (V_3)
Audio	Ref (A_R)	86%	73%	99%	69%
	HSTrim (A_1)	96%	69%	88%	77%
	HSGain (A_2)	92%	77%	96%	73%
	MP3 (A_3)	69%	62%	99%	77%

Table 3: Percent of correct answers to "Did the video change? Yes or No" for the *Batman* clip.

Percent Correct: *Foo Fighters*		Video			
		Ref (V_R)	Desat (V_1)	Bright (V_2)	Pixel (V_3)
Audio	Ref (A_R)	83%	99%	92%	69%
	HSTrim (A_1)	77%	99%	88%	73%
	HSGain (A_2)	69%	99%	77%	65%
	MP3 (A_3)	77%	96%	81%	81%

Table 4: Percent of correct answers to "Did the video change? Yes or No" for the *Foo Fighters* clip.

Within each of these tables, Signal Detection Theory provides us further delineation between correct and incorrect guesses, categorized by "hits", "misses", "correct rejections", and "false alarms" as described in Section 2.2:

Hit = % correct for all V_n stimuli
Miss = (100 - % correct) for V_n
Correct Reject = % correct for V_R
False Alarm = (100 - % correct) for V_R

In the two figures below, the percentage of instances where subjects correctly identified the reference audio and video can be seen by lining up the Audio Artifact "Ref" column with the dotted, open-circle line indicating the reference video. For example, in Fig. 1,

subjects correctly identified the A_RV_R / A_RV_R pairing of *Batman* 86% of the time. In Figure 2, subjects correctly identified the A_RV_R / A_RV_R pairing of *Foo Fighters* 83% of the time.
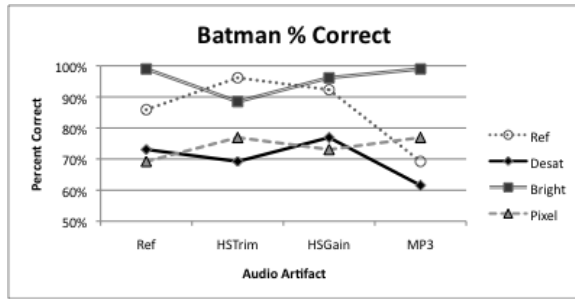

Figure 1: Results for *Batman* clips. Subjects were able to identify pairings of two reference videos (A_RV_R / A_RV_R) 86% of the time. However, subjects were only able to correctly identify a pairing of a reference video with one that had desaturation applied to the video and MP3 audio (A_RV_R / A_3V_1) 62% of the time.
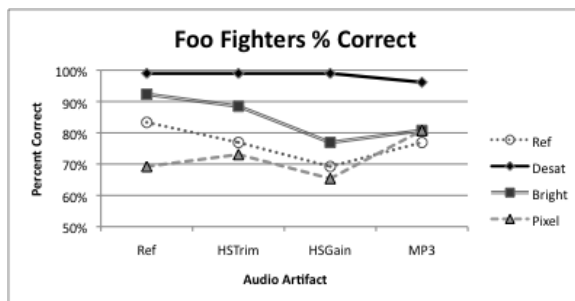

Figure 2: Results for *Foo Fighters* clips. Subjects were able to identify pairings of two reference videos (A_RV_R / A_RV_R) 83% of the time. However, subjects were only able to correctly identify a pairing of a reference video with one that had pixelation applied to the video and audio with high shelf gain (A_RV_R / A_2V_3) 65% of the time. Subjects were almost 100% correct when identifying all clips with a desaturation change, except when MP3 audio accompanied the picture.

4.   STATISTICAL ANALYSIS

Based on a binomial distribution, we obtained our desired significance level of 0.05 (p<0.04) for 31 of our 32 stimuli (26 subjects rating each sequence). In other words, the probability of randomly getting our percent correct data points or better was <0.04, except for the Batman "DesatMP3" stimulus for which this probability was 0.084. Initially, this might appear to suggest that

the subjects may have been guessing when presented with the Batman "DesatMP3" stimulus (i.e. unable to discriminate between Batman DesatMP3 and Batman Reference).

However, when analyzed in more detail using Signal Detection Theory, we see that the sensitivity metric (d') provides more insight. (See Figures 3 and 4.) The following graphs illustrate the bias values and d' values. The d' value is found based on "hit" (H) and "false alarm" (F) rates. If it is assumed that they correspond to points on a Gaussian distribution, d' can be calculated such that:

$$d' = z(H) - z(F) \tag{1}$$

where $z(x) =$ is the z-score, or the point at which the standard normal distribution is equal to the value specified. [12]
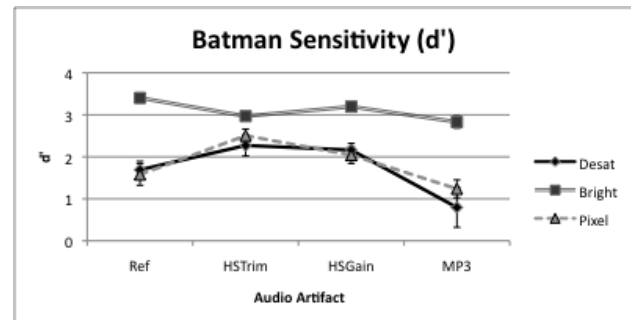

Figure 3: *d'* values for *Batman* video and associated audio artifacts, along with 95% confidence intervals.
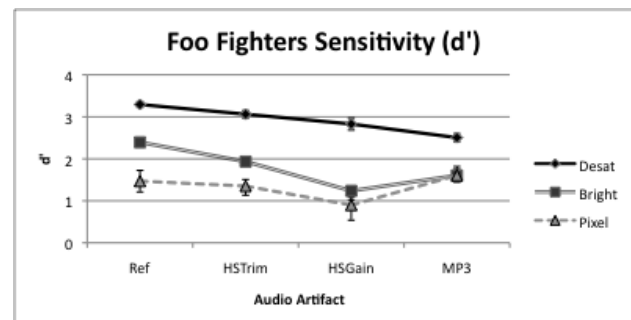

Figure 4: *d'* values for *Foo Fighters* video and associated audio artifacts, along with 95% confidence intervals.

Note, for example, that although the confidence interval for the Batman "DesatMP3" clip is larger than the others (±0.47), the confidence interval using *d'* data is

below the confidence interval for the Batman "DesatRef" clip, indicating that the reduction in sensitivity is in fact significant. Note also that, because the confidence interval does not cross zero, the statistics suggest that subjects were still able to discriminate between Batman "DesatMP3" and "BatmanRef".

In addition to sensitivity, the bias for each response set was calculated according to the equation:

$$c = [z(H) + z(F)] / 2 \qquad (2)$$

Figures 5 and 6 show the calculated values for the bias. A positive value indicates a bias toward answering "yes", while a negative value indicates a bias toward answering "no".
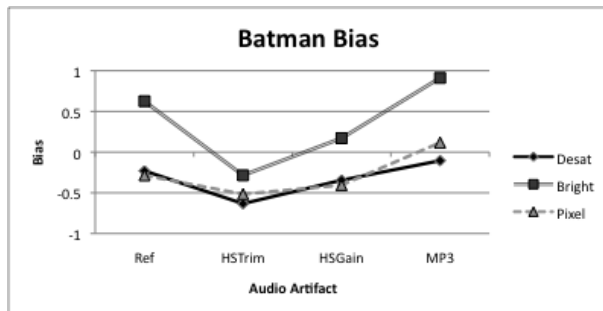

Figure 5: Bias values for *Batman* video and associated audio artifacts.
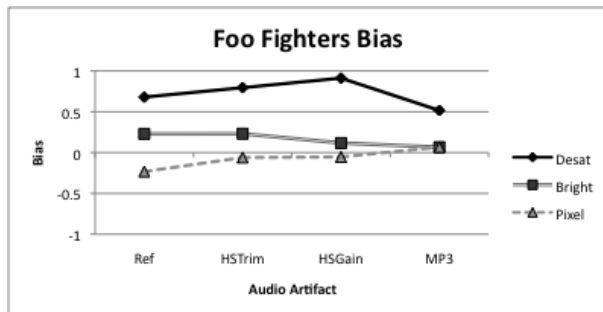

Figure 6: Bias values for *Foo Fighters* video and associated audio artifacts.

For each video clip with an audio artifact, d' was compared to d' for the same clip with the reference audio track. The resulting Δd' (i.e., Δd' = d'$[A_n V_n]$-d'$[A_R V_n]$) then represents the change in sensitivity corresponding to a particular audio artifact. For example, a positive Δd' indicates that the presence of a

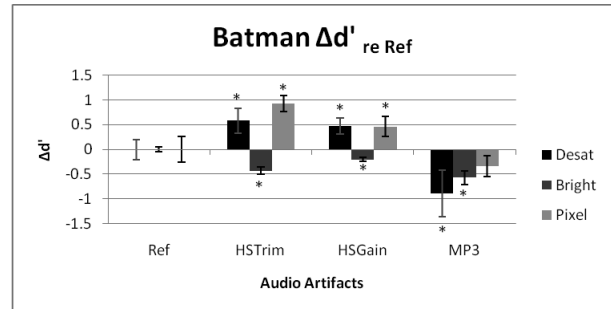particular audio artifact is associated with an increase in sensitivity, or easier discrimination.

Figure 7: Δd' values for *Batman* video and associated audio artifacts, along with 95% confidence intervals
(* indicates a significant difference at α=0.05)
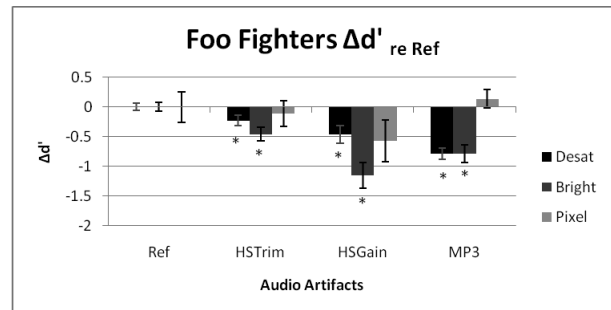
Figure 8: Δ d' values for *Foo Fighters* video and associated audio artifacts, along with 95% confidence intervals.
(* indicates a significant difference at α=0.05)

4.1. Interpretation of statistical analysis

Based on the statistical analysis, our findings indicate that:

- Desaturation and pixelation were easier to spot in the Batman clip with HSTrim and HSGain.

- Desaturation was harder to spot in both clips with MP3.

- Brightness was harder to spot in both clips with any audio artifact.

5.   FURTHER OBSERVATIONS OF DATA

In addition, we observe the following:

- Perhaps one of the most interesting results to arise from this study is the correlation between MP3-quality audio and resolution-affected video. The *Foo Fighters* "PixelMP3" (A_3V_3) results in a correct rate that is 12% higher than that of the *Foo Fighters* Pixel ($A_R V_3$), 81% versus 69%. As with the *Foo Fighters* clip, the *Batman* "PixelMP3" (A_3V_3) results in a correct percentage that is 8% higher than that of *Batman* Pixel ($A_R V_3$): 77% versus 69%. Of interest in the *Batman* "PixelMP3" clip is that fact that the answers for both *Batman* "Pixel" (A_RV_3) and *Batman* "MP3" (A_3V_R) have the same correct percentage of 69%, while the clip that combines the two artifacts results in a correct rate that is higher than each artifact individually.   This evidence supports the conclusion that MP3-quality audio heightens the perception of pixelation in video. More succinctly, it "heightens" the perception, but only in the sense that it creates a bias (note that sensitivity did not improve). In this case, it appears that mp3 actually lowered subjects' "criterion" for identifying a change.

- One of the largest examples of the impact of an audio impairment on video quality can be seen by looking at three *Foo Fighters* sequences: "BrightHSGain" (A_2V_2), "HSGain" (A_2V_R), and "Bright" (A_RV_2).   There was a 15% decrease (92% to 77%) in the number of correct answers from *Foo Fighters* "Bright" to "BrightHSGain". The same outcome can also be seen with the *Batman* clips of the same nature, to a lesser degree, likely due to the dark image present in the *Batman* sequence. These two outcomes indicate that a gain in audio high frequencies may mask the perception of changes in video brightness. Looking at delta d', we see the same huge effect for the Foo Fighters clip. The effect was smaller for the Batman clip, but still statistically significant.

- An interesting result in this study shows that both high shelf audio boosts and cuts have a masking effect on video brightness perception.

In the *Batman* clip, "Bright" (A_RV_2) was answered correctly 100% of the time, but "BrightHSTrim" (A_1V_2) had a number of correct answers (89%), which is 11% lower. In the *Foo Fighters* clip, "BrightHSTrim" (A_1V_2) has a correct rate of 88%, four percent lower than the 92% correct rate of "Bright" (ARV2). (Note that sensitivity also dropped by 0.43-0.46 for these clips.)

- Interestingly, the data show that MP3-quality audio masks a perceived change in video brightness in the *Foo Fighters* clip.   When MP3-quality audio is introduced with video brightness, "BrightMP3" (A_3V_2), the correct answer rate drops from 92% to 81% (d' dropped by 0.79). Meanwhile, when looking at percent correct, it appears that MP3-quality audio in the *Batman* clip had no impact on perceived video brightness.   In fact, the sensitivity drops (similar to *Foo Fighters*), but the bias increases, thus making the percent correct look like there was no effect.

- As with MP3 audio's "masking effect" on video brightness in the *Foo Fighters* clip, MP3 audio has a masking effect on both the *Batman* and *Foo Fighters* clips in terms of desaturation perception. The *Batman* "DesatMP3" (A_3V_1) clip shows an 11% decrease in the correct rate over *Batman* Desat clip (A_RV_1), 62% over 73%. (Sensitivity dropped by 0.9.) In the *Foo Fighters* clip, the same artifact pairings from Batman result in a similar outcome but on a smaller masking scale – a four percent drop in the correct between *Foo Fighters* – Desaturation (A_RV_1) and DesatMP3 (A_3V_1). For the Foo Fighters clip, sensitivity dropped by 0.79.

- Like the effects of the high shelf boost and cut audio artifacts on perceived video brightness, the addition of high shelf boosts and cuts to *Batman's* desaturated video results in heightening and masking effects, respectively.

For a full list of the artifact pairing analyses see the Appendix.

5.3.   Other data observations

In analyzing the breakdowns of each session of 12 sequences, averaging across the 26 subjects reveals that subjects improved in the number of correct answers as the test progressed. After averaging the first of 12 across the 26 subjects the correct rate is 82%. For the second set of 12 the average is 85% and after the third set the average is 86%. Furthermore, there were five perfect scores in first set and seven for each of the second and third sets. This improvement over time does not present a bias as the clips were randomized, with no set of 12 ever containing the same order or the same clip pairings.

6.   CONCLUSION

Our experiment has shown that when comparing two videos one after the other, and when asked whether the videos were the same or different, viewers' responses were influenced by the presence of different audio impairments. Furthermore, different audio impairments caused a different number of correct responses, which in turn depended on the video impairment that was present.

These results must be considered important when striving toward standards related to the subjective evaluation of audio/visual material, especially when the programming will be degraded by any of the artifacts included here.

7.   FUTURE WORK

We have only tried a few impairments, and based on the results of this test we are certain that other artifacts would show interesting results as well: different bit rates, lip-sync error, and distortion in audio, as well as different resolutions, edge detection, and shadow delineation problems for video.

Because the combinations of impairments to test would be so numerous, it would be advisable for future research to focus on those that are most problematic for television and Internet broadcast.

8.   ACKNOWLEDGEMENTS

9.   APPENDIX

9.1. Scripts

Training Script: "Thank you for joining us today. We are going to be watching video clips, and we'd like for you to tell us whether they are the same or different. You'll mark down your answer on a sheet of paper. But first, we'll look at a video and talk about the differences you perceive in this short clip".

Session script: "In these sessions you'll simply be asked whether the video changed or not. If you think the two clips are different, mark "yes". If you think they are the same, mark "no". You can only view each pair once."

9.2. Test Sheet

Name: _____
Age: _____ Gender: _____

Session 1 of 3

1. Did the video change? ☐ Yes ☐ No
2. Did the video change? ☐ Yes ☐ No
3. Did the video change? ☐ Yes ☐ No
4. Did the video change? ☐ Yes ☐ No
5. Did the video change? ☐ Yes ☐ No
6. Did the video change? ☐ Yes ☐ No
7. Did the video change? ☐ Yes ☐ No
8. Did the video change? ☐ Yes ☐ No
9. Did the video change? ☐ Yes ☐ No
10. Did the video change? ☐ Yes ☐ No
11. Did the video change? ☐ Yes ☐ No
12. Did the video change? ☐ Yes ☐ No

9.3. Full Raw Data Results (Batman)

Batman - Pixelation and MP3: Batman-Pixel ($A_R V_3$) and Batman-MP3 ($A_3 V_R$) clips have the same correct rate by test subjects of 69%. When the artifacts are paired together in Batman-PixelMP3 ($A_3 V_3$), the false alarm rate falls and the percent of correct answers rises to 77%. The rise of eight percent in correct answers could indicate that MP3-quality audio heightens the

effect of pixelation in this case, making it more noticeable to subjects. However, the change in sensitivity is not significant.

Batman - Pixelation and HSTrim: With Batman-Pixel (A_RV_3) the correct rate is 69%, and with Batman-HSTrim (A_1V_R) the correct rate is 96%. When the artifacts are paired in Batman-PixelHSTrim (A_1V_3), the correct rate rises to 77% from the Batman-Pixel $(A_R V_3)$ correct rate of 69%. (Sensitivity also improved.) This comparison indicates a high shelf audio cut has a heightening effect on resolution-affected video in this case.

Batman - Pixelation and HSGain: The results are the same as "Pixelation and HSTrim," but to a lesser degree. Batman-Pixel (A_RV_3)'s correct rate is 69% and Batman-HSGain (A_2V_R)'s correct rate is 92%. When paired, Batman-PixelHSGain (A_2V_3), the correct rate is 73%, slightly higher than the Batman-Pixel $(A_R V_3)$ rate. Similarly, sensitivity increased, though not as much as with Pixelation and HSTrim. This indicates that a high shelf boost in the audio heightens the effect pixelation.

Batman - Desaturation and MP3: Batman-Desat (A_RV_1) yields a 73% correct rate and Batman-MP3 (A_3V_R) yields a 69% correct rate. When the artifacts are paired in Batman-DesatMP3 (A_3V_1), the correct rate is 62%. This 11% decrease from Batman-Desat (A_RV_1) indicates MP3-quality audio masks the Desaturation effect. In this case, sensitivity was also reduced significantly.

Batman - Desaturation and HSTrim: Batman-HSTrim (A_1V_R) has a correct rate of 96%. When paired with desaturated video in Batman-DesatHSTrim (A_1V_1), the correct rate is 69%, a four percent reduction from Batman-Desat (A_RV_1). A high shelf cut in audio would appear to have a slight compensation effect on perceived video saturation. However, the sensitivity is increased, indicating that HSTrim actually enhances discrimination of desaturation effects.

Batman - Desaturation and HSGain: Batman-HSGain (A_2V_R) has a correct rate of 92%. When paired with the desaturated video, Batman-DesatHSGain (A_2V_1) has a correct rate of 77%. This is a four percent increase in percent correct over Batman-Desat (A_RV_1). Similarly, sensitivity was also increased. This indicates that a high shelf audio boost has a heightening effect on perceived video saturation.

Batman - Brightness and MP3: Batman-Bright (A_RV_2)'s correct rate is 100% while Batman-MP3 (A_3V_R)'s correct rate is 69%. When combined, Batman-BrightMP3 (A_3V_2), the correct rate remains at 100%. This might indicate that MP3-quality sound does not affect the perceived video quality in this specific audio/video sequence. However, we see that sensitivity was reduced, indicating that the MP3 audio actually made brightness discrimination more difficult.

Batman - Brightness and HSTrim: Batman-HSTrim (A_1V_R) has a correct rate of 96% and when paired with the brightness video in Batman-BrightHSTrim (A_1V_2), the correct becomes 89%, 11% lower than that of Batman-Bright (A_RV_2). (Sensitivity was also reduced.) This indicates that high shelf audio cuts compensate for the addition of brightness in video.

Batman - Brightness and HSGain: Batman-HSGain (A_2V_R) has a correct rate of 92%. Combined, Batman-BrightHSGain (A_2V_2) yields a correct rate of 96%, a four percent drop from the correct rate of Batman-Bright (A_RV_2). (Sensitivity was also slightly reduced.) This indicates a slight compensation effect of high shelf audio boosts on perceived video brightness.

9.4. Full Raw Data Results (Foo Fighters)

Foo Fighters - Pixelation and MP3: The Foo-Pixel (A_RV_3) correct rate is 69%, while the correct rate of Foo-MP3 (A_3V_R) is 77%. When combined, Foo-PixelMP3 (A_3V_3)'s correct rate is 81%, 12% greater than Foo-Pixel $(A_R V_3)$. Because the correct rate of Foo-PixelMP3 (A_3V_3) is greater than both Foo-Pixel $(A_R V_3)$ and Foo-MP3 (A_3V_R), it is a strong indication that MP3-quality audio heightens the pixelation effect. However, the sensitivity metric did not significantly change, indicating that criterion effects are responsible for this change in perception.

Foo Fighters - Pixelation and HSTrim: Foo-HSTrim (A_1V_R) has a correct rate of 77%. When the artifact is combined with Foo-Pixel $(A_R V_3)$, Foo-PixelHSTrim (A_1V_3)'s correct rate is 73%. This indicates that a high shelf audio cut may slightly heighten the pixelation effect, but the change in sensitivity is not significant.

Foo Fighters - Pixelation and HSGain: Both Foo-HSGain (A_2V_R) and Foo-Pixel $(A_R V_3)$ have a correct rate of 69%. When paired, Foo-PixelHSGain (A_2V_3)'s correct rate is 63%. The six percent difference, and the corresponding drop in sensitivity, indicate that a high shelf audio boost exacerbates perception of pixelation.

Foo Fighters - Desaturation and MP3: The correct rate of Foo Fighters – Desaturation (A_RV_1) is 100%. The correct rate of Foo-MP3 (A_3V_R) is 77%. Combined, Foo-DesatMP3 (A_3V_1)'s correct rate of 96% shows a drop of four percent from Foo Fighters – Desaturation (A_RV_1). This indicates a slight masking effect of MP3-quality audio on perceived video saturation. This is confirmed by the significant drop in sensitivity.

Foo Fighters - Desaturation and HSTrim/HSGain: At the conclusion of this test it was found the desaturation effect of the Foo clip was too far removed from the threshold detection level to perform an analysis based on percent correct. (However, it can be noted that some individuals taking the test marked their answers faster when the desaturated video came second to the reference clip versus the slow answer time when the reference clip came second.) When analyzed via Signal Detection Theory, we see that HSTrim results in a slight reduction in sensitivity and HSGain may reduce it even further (though not significantly).

Foo Fighters - Brightness and MP3: The correct rate of Foo-Bright (A_RV_2) is 92%. The correct rate of Foo-MP3 (A_3V_R) is 77%. When combined, Foo-BrightMP3 (A_3V_2) shows a correct rate of 81%, which is higher than Foo-Bright (A_RV_2). (Also note that the sensitivity is reduced when the audio is MP3-encoded.) This indicates that MP3-quality audio masks some of the effect of the video brightness filter.

Foo Fighters - Brightness and HSTrim: Foo-HSTrim (A_1V_R) has a correct rate of 77%. When the artifact is combined with video brightness in Foo-BrightHSTrim (A_1V_2), the correct rate is 89%, a four percent drop in picture change detection. (Sensitivity is also reduced.) This indicates that a high shelf audio cut compensates for the brightness picture change.

Foo Fighters - Brightness and HSGain: Foo-HSGain (A_2V_R) has a correct rate of 69%. When combined with the brightness-affected video, Foo-BrightHSGain (A_2V_2) yields a correct rate of 77%. This is a 15% reduction in the detection of boosted video brightness. (Similarly, sensitivity was reduced significantly.) This indicates that a high shelf audio boost compensates for the elevation in brightness.

10. SCORES

Subject	% Correct Answers
Subject 1	89
Subject 2	86
Subject 3	92
Subject 4	86
Subject 5	64
Subject 6	75
Subject 7	67
Subject 8	83
Subject 9	86
Subject 10	86
Subject 11	97
Subject 12	72
Subject 13	81
Subject 14	83
Subject 15	69
Subject 16	81
Subject 17	92
Subject 18	83
Subject 19	78
Subject 20	83
Subject 21	86
Subject 22	92
Subject 23	83
Subject 24	86
Subject 25	78
Subject 26	92

Table A-1: Scores from individual subjects.

11. REFERENCES

[1] Marins, P., Rumsey, F., Zielinski, S., "The Relationship between Selected Artifacts and Basic Audio Quality in Perceptual Audio Codecs", AES 120th Convention, May, 2006, Paris, France, Paper #6745 .

[2] Beerends, J., De Caluwe, Frank. "The Influence of Video Quality on Perceived Audio Quality and Vice Versa", *Journal of the Audio Engineering Society,* Vol. 47, No. 5 pp. 355-362, May, 1999.

[3] IBID, pp. 355-362.

[4] Joly, A., Montard, N., Buttin, M., "Audio-visual quality and interactions between television audio and video," Sixth International, Symposium on Signal Processing and its Applications, 2001, vol.2, pp. 438-441.

[5] Garcia, M.N., Raake, A., "Impairment-factor-based audio-visual quality model for IPTV." Proceedings from the International Workshop on Quality of Multimedia Experience, pp.1-6, 29-31, July 2009.

[6] "Question 12/9 – Objective and subjective methods for evaluating perceptual audiovisual quality in multimedia services within the terms of Study Group 9", International Telecommunications Union Telecommunications Standard Sector, 5 Nov. 2009, <http://www.itu.int/ITU-T/studygroups/com09/sg9-q12.html>, March 10, 2010.

[7] *Batman Returns.* Dir. Tim Burton. Perf. Michael Keaton, Michelle Pfeiffer, Danny DeVito. Warner Bros., 1992. Blu-Ray. Chapter 18, 0:58:52-0:59:14.

[8] *Foo Fighters Live at Wembley Stadium.* RCA, 2008. Blu-ray. Chapter 5, 0:19:34-0:19:55

[9] ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union Radiocommunications, (1974–2002).

[10] "ATSC Implementation Subcommittee Finding: Multichannel Audio Program Delivery and Metadata Considerations (Pre-emission)", Advanced Television Systems Committee Document IS-318, 10 September 2004, p. 6.

[11] *Earth.* Dir. Alastair Fothergill and Mark Linfield. Perf. James Earl Jones. Disney Nature, 2009. Blu-ray. Chapter 3, 0:21:44 – 0:21:55.

[12] MacMillan, N.A., C.D. Creelman. Detection Theory, A User's Guide. Second Edition. MahWah, New Jersey: Erlbaum. 2005.