NTIA Technical Memorandum TM-11-474

Preliminary Investigation into the Impact of Audiovisual Synchronization of Impaired Audiovisual Sequences

Margaret H. Pinson
Arthur Webster
William Ingram

technical memorandum

U.S. DEPARTMENT OF COMMERCE • National Telecommunications and Information Administration

NTIA Technical Report TM-11-474

Preliminary Investigation into the Impact of Audiovisual Synchronization of Impaired Audiovisual Sequences

Margaret H. Pinson
Arthur Webster
William Ingram

U.S. DEPARTMENT OF COMMERCE

March 2011

DISCLAIMER

Certain commercial equipment and materials are identified in this report to specify adequately the technical aspects of the reported results. In no case does such identification imply recommendations or endorsement by the National Telecommunications and Information Administration, nor does it imply that the material or equipment identified is the best available for this purpose.

CONTENTS

Page

FIGURES

TABLES

PRELIMINARY INVESTIGATION INTO THE IMPACT OF AUDIOVISUAL SYNCHRONIZATION OF IMPAIRED AUDIOVISUAL SEQUENCES

Margaret H. Pinson, Arthur Webster, and William Ingram[1]

The quality perception of an audiovisual sequence is heavily influenced by the quality of the audio, the quality of the video, and the audiovisual time synchronization. The questions then arise: what is the relative importance of each factor, and can a model be devised that is generally applicable? Previous work either examined the relative influences of audio and video quality for synchronized video or investigated the quality impact of synchronization errors on unimpaired video sequences. This experiment is a first attempt to combine all three factors into a single experiment, to judge the complex interactions among individual measurements of audio and video quality and synchronization errors.

Key words: audio quality, audiovisual quality, subjective testing, synchronization, video quality

1 INTRODUCTION

The experiment described in this document seeks to combine two separate lines of research. The first line of research analyzes subjective experiments that explore the relationship between audio quality and video quality, measured separately, and the overall quality of an audiovisual experience. Each of these experiments produced a model that mapped audio quality (***a***) and video quality (***v***) to the overall audiovisual quality (***av***). These models are similar; however the terms and the values of the coefficients used in the model differ from one experiment to the next. At least thirteen previous experiments have been performed on this topic [1]-[12]. A detailed analysis in [12] indicates that only the cross term ($\boldsymbol{a} \times \boldsymbol{v}$) is needed to predict the overall audiovisual quality. It provides us with a simple and reasonably accurate model that has been tested in a wide variety of circumstances, from CIF to HDTV, from video conferencing to broadcast television, both coding only and with transmission errors, in a professional viewing/listening environment and on a PC.

The second line of research analyzes audiovisual synchronization errors. The perceptibility of synchronization errors is a critical issue, because audio and video are occasionally split and processed separately. A variety of acceptability thresholds are recommended in literature. The Advanced Television Systems Committee (ATSC) recommends very stringent control of synchronization errors, such that "the sound program should never lead the video program by more than 15 milliseconds, and should never lag the video program by more than 45 milliseconds" [13]. ITU-R Rec. BT-1359-1 states that in subjective evaluations, detectability thresholds are about -45 ms to +125 ms and acceptability thresholds are about -90 ms to +185 ms on the average, where a positive value means that the audio lags the video. An Intel study on

---

[1] The authors are with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80305.

lipsynch when audio leads video indicated that the average threshold for audiovisual synchronization detection by the average end-user was -185.19 ms [15]. ITS examined the relationship between synchronization errors (***e***) and the overall quality of an audiovisual experience (***av***) [16]. This experiment produced a graph showing the rate at which quality dropped as synchronization errors increased. All of these studies examined sync errors in isolation (e.g., original audio-video clips without coding errors).

The goal of this current experiment is to refine our understanding of how audio-video synchronization (AV-Synch) errors influence the perceived multimedia quality of service. The primary variables of this experiment are audiovisual synchronization errors (***e***), which can be positive or negative in milliseconds, audio quality (***a***), video quality (***v***), clips (***c***) and subjects (***s***). It is important to fully span the ***a*** and ***v*** quality spaces for proper model fitting. A critical constraint on this experiment was the desire to limit the number of sequences viewed by the subjects.

2 EXPERIMENT DESIGN

2.1 Design Overview

Experimental design is always a tradeoff between sampling and organizing the various design variables to maximize the useful information obtained from the experiment and performing the minimum amount of work. Designs that achieve balance between the various experimental variables produce more accurate results since systematic errors are reduced by averaging. This experiment, ITS AV-Synch 2010, has the following number of samples for each design variable:

e: Nine different delay levels were chosen after examining results in [13] and [16]. These delays were expected to provide for both positive and negative shifts, a large delay impairment, a moderate delay impairment, and a small delay impairment, in addition to the no AV-Synch error case. The delays in milliseconds are shown in Table 1. Positive values mean that the audio lags behind the video and negative values mean that the audio leads the video.

Table 1. Millisecond Delays for Level of Impairment.

Name	**E1**	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**	**E9**
Delay	0	-180	315	-45	-360	180	90	-90	495

v: Three different video quality levels—high quality (or original), medium quality, and low quality. Three levels are deemed the absolute minimum to span the quality space (there will be some finer variability of quality about each of these three points due to content variability).

a: Three different audio quality levels—high quality (or original), medium quality, and low quality. Three levels are deemed the absolute minimum to span the quality space (there will be some finer variability of quality about each of these three points due to content variability). The audio quality levels were chosen to perceptually match the video quality levels, ***v***.

c: Nine different clips, all of which have audio that is inherently related to the video (so subjects can assess the effect of AV-Synch errors). Different kinds of AV-Synch are desirable (e.g., voice, percussion). The clip Heavy Bag and its two lowest A and V quality levels together with its highest (original) quality level were included from the ITS 2010 experiment described in [12], to enable comparisons between these data sets. Each clip was 15 seconds long.

s: Of the 28 total subjects, each viewing all impairments, 25 were paid subjects and three were summer interns employed by ITS. Since the total number of combinations (***e*** × ***v*** × ***a*** × ***c***) is 729, which is too large for a single experiment, one third of the full matrix was used (i.e., 243 combinations) so that each subject's participation was anticipated to be under 2.5 hours including training and breaks. In practice, all of the subjects finished their viewing session in slightly under two hours.

Figure 1 shows the one third sampling. A numeral one (1) in Figure 1 indicates inclusion of this (***e*** × ***v*** × ***a*** × ***c***) combination, and a numeral zero (0) indicates exclusion. At the macro level, this figure presents the audio quality levels (***a***) in three large rows and the video quality levels (***v***) in three large columns. Within each of these nine squares are the nine delay levels (***e***) in rows and the nine clips (***c***) in columns. Thus, each square shows all (***e*** × ***c***) combinations for one audio level / video level pair. The one third sub-sampling was performed systematically, as follows:

- Clips are grouped by threes, and included or excluded together. These groups were chosen randomly.
- Delay levels are grouped by threes, and included or excluded together. These groups were chosen to include one negative delay, one positive delay, and one delay near zero.
- Each audio level/video level pair includes all nine clips.
- Each audio level/video level pair includes all nine delays.
- Each audio level includes all clip/delay combinations (with varying video levels).
- Each video level includes all clip/delay combinations (with varying audio levels).

This sub-sampling emphasizes inclusion of a wide range of clips and delays. The drawback is that comparisons between audio levels and video levels are made more difficult (e.g., using ANOVA).

This experiment used the same monitor, speakers, Blu-Ray test control, sound isolation booth and subjective test method as that previously described for the ITS 2010 experiment in [12].

2.2 HRCs

To maintain overlap with the ITS 2010 experiment, the audio and video levels for the Heavy Bag sequence were selected to be the same in ITS AV-Synch 2010. For the other sequences, level 1 was chosen to contain the original video and audio. Levels 2 and 3 were chosen by a panel of experts to match their opinion of "medium quality" and "low quality." The video coders were limited to MPEG-2 and H.264; and the audio coders were limited to AAC, MP3 and T-ref. The use of the audio codec T-ref [17], [18] was needed to produce low quality audio samples that did not contain objectionable aliasing artifacts.

	Video Level High (1) C1 to C9	Video Level Medium (2)	Video Level Low (3)
Audio Level High (1) E1 to E9	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
Audio Level Medium (2)	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
Audio Level Low (3)	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	0 0 0 0 0 0 1 1 1	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1
	1 1 1 0 0 0 0 0 0	0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 1

Figure 1. ITS AV-Synch 2010 sampling matrix indicating the one-third of sequences to be examined.

Each audio HRC and video HRC was selected individually. The levels chosen were jointly considered and modified, such that the experimenters perceived all medium level audio and video samples as being approximately fair quality, and all low level audio and video samples as being approximately bad quality. Audio compression rates ranged from 8 to 64 Kbits/s and video compression rates ranged from 192 to 1000 Mbits/s.

2.3 Audiovisual Sequences

This experiment contains nine audiovisual sequences. Each original sequence contains video that, when visually inspected by an expert, appears to be of good or better quality. These video sequences were carefully chosen to span a wide range of coding difficulty and a variety of visual characteristics (e.g., scrolling text, fast motion, rapid scene cuts, and random motion). As is common with broadcast television, both the audio and the video were edited to produce these video sequences. Sometimes audiovisual synchronization clues exist, and other times audio was recorded separately from the video and so no audiovisual synchronization clues exist. For example, in clip Mall Music, the person shown in Figure 2 continues to talk while a montage of footage depicting the outdoor mall is displayed. Occasionally, audio from that montage is mixed into clip Mall Music's audio track, and this provides additional audiovisual synchronization clues. All clips have at least a few seconds of visibly synched audio and video.

Eight of the nine clips contain human speech and the ninth one features a marching band. One track contains both speech and music. Three of the clips feature sporting events that include rapid movement.

Figure 2 shows a sample frame from each sequence (original only), along with a description of the audio and video. These audiovisual sequences can be viewed in the Consumer Digital Video Library (www.cdvl.org) and downloaded royalty-free for research and development purposes.

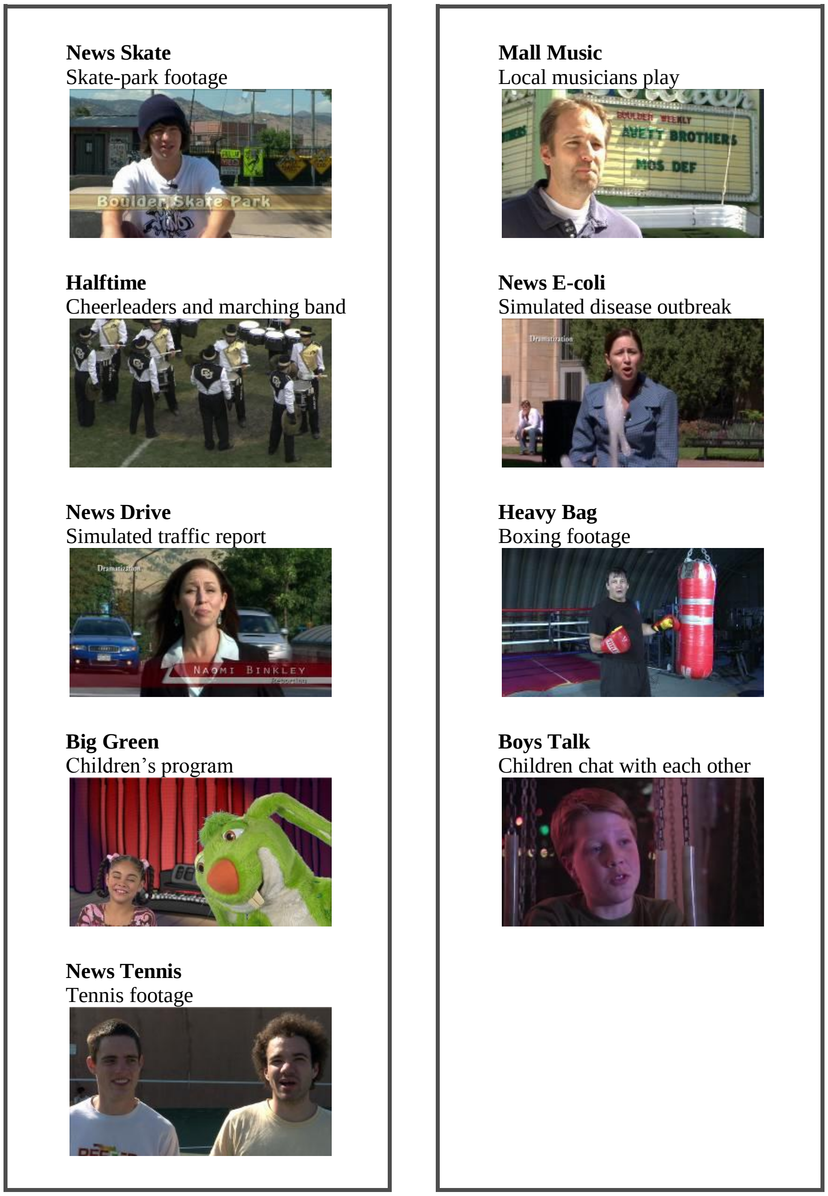

Figure 2. Sample frame from each video sequence displayed below a brief description of the audio.

3 ANALYSIS

The audiovisual subjective data was analyzed against the three main variables of interest: audio quality, video quality, and differential delay. Table 2 shows the accuracy of various types of linear regression and multiple linear regression models that can be built from these components. In these models, $|\boldsymbol{d}|$ is the absolute value of the delay and ρ is Pearson correlation.

Table 2. Accuracy of Models with Different Parameters

Model	**Model Weights**	**ρ**
$\hat{y} = \alpha + \beta\, \boldsymbol{a}$	$\hat{y} = 1.618 + 0.379\, \boldsymbol{a}$	0.43
$\hat{y} = \alpha + \gamma\, \boldsymbol{v}$	$\hat{y} = 1.514 + 0.386\, \boldsymbol{v}$	0.49
$\hat{y} = \alpha + \delta\, \|\boldsymbol{d}\|$	$\hat{y} = 3.102 + -0.0017\, \|\boldsymbol{d}\|$	0.33
$\hat{y} = \alpha + \beta\, \boldsymbol{a} + \gamma\, \boldsymbol{v}$	$\hat{y} = 0.386 + 0.373\, \boldsymbol{a} + 0.382\, \boldsymbol{v}$	0.65
$\hat{y} = \alpha + \mu\, (\boldsymbol{a} \times \boldsymbol{v})$	$\hat{y} = 1.622 + 0.115\, (\boldsymbol{a} \times \boldsymbol{v})$	0.65
$\hat{y} = \alpha + \mu\, (\boldsymbol{a} \times \boldsymbol{v}) + \delta\, \|\boldsymbol{d}\|$	$\hat{y} = 1.946 + 0.115\, (\boldsymbol{a} \times \boldsymbol{v}) + -0.0017\, \|\boldsymbol{d}\|$	0.73

Caution should be used when interpreting the rest of this analysis and the figures that follow. The accuracy is somewhat limited by this experiment's design (i.e., due to the sub-sampling of clips by three, a different set of conditions was used to compute each mean error).

An experiment conducted by ITS in 2009 [11] (which we will call ITS AV-Synch 2009) indicated different weights for the positive and negative delays. This can be seen in Figure 3, reprinted from [16]. However, ITS AV-Synch 2010 indicates that in the presence of audio impairments and video impairments, the positive and negative delays impact quality approximately equally. This can be seen in Figure 4. This is why the equations in Table 2 have a single delay term, $|\boldsymbol{d}|$, instead of separate terms for positive and negative delay, as we would expect from Figure 3.

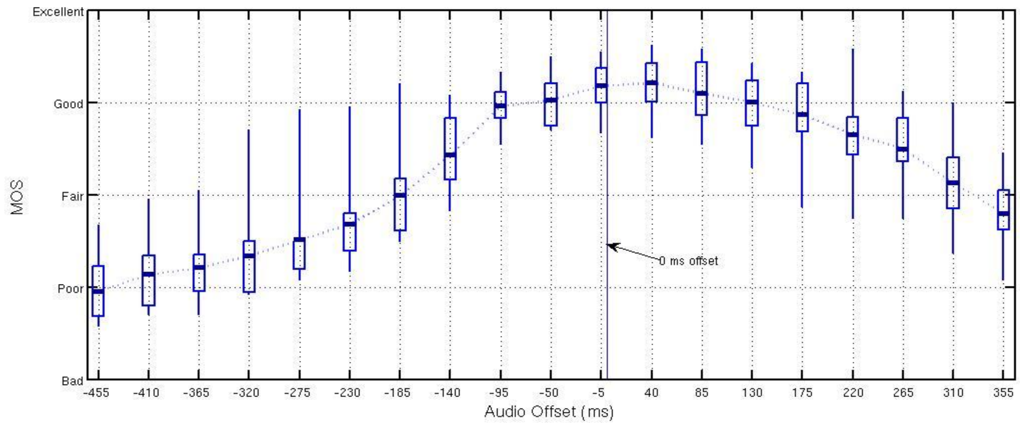

Figure 3. MOS as a function of audio offset, from ITS AV-Synch 2009, which explored the impact of audiovisual synchronization errors on source audio and source video.

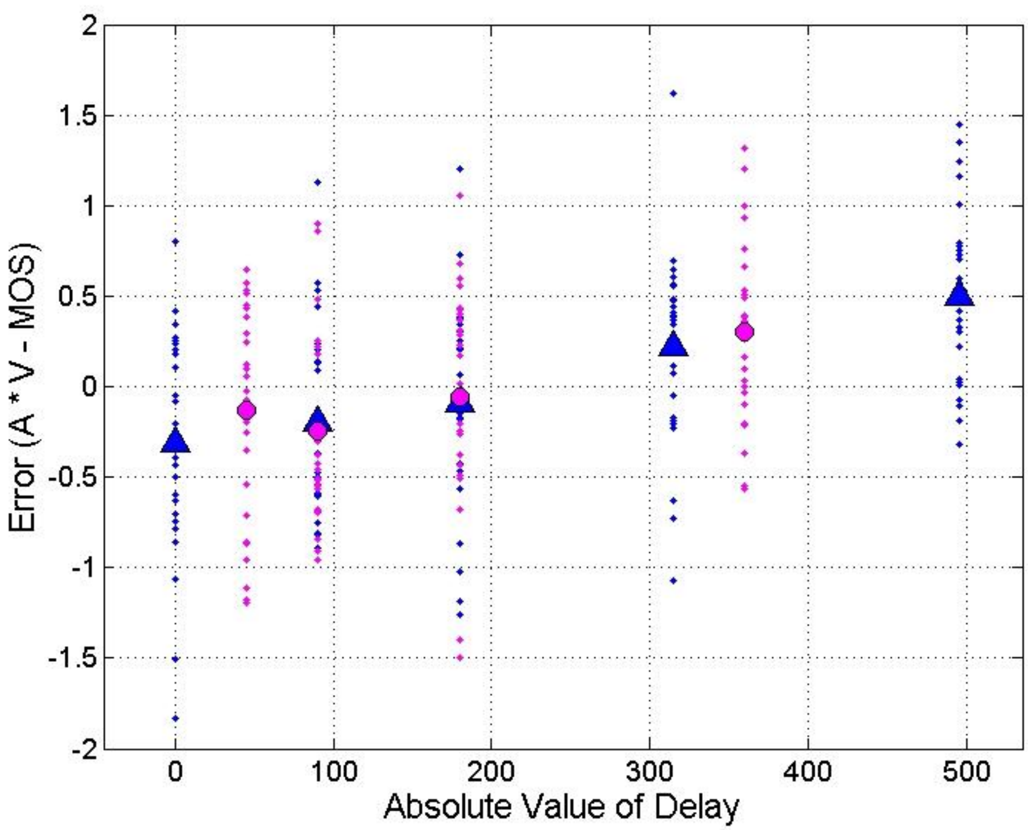

Figure 4. Absolute value of delay shifts plotted against the error remaining after predicted MOS using the multiplicative model $\hat{y} = \alpha + \mu\ (a \times v)$. Positive delays are plotted in blue, with the mean plotted as a large blue triangle; and negative delays are plotted in pink, with the mean plotted as a pink circle.

The three obvious differences between these two experiments are (1) the range of audio quality: source audio for ITS AV-Synch 2009 versus audio quality ranging from excellent to bad for ITS AV-Synch 2010, (2) the range of video quality: source video versus video quality ranging from excellent to bad, and (3) the video resolutions: CIF versus HDTV. However, it is possible that the difference in the weight given to positive and negative delays stems from the source content. Seven of the nine scenes used in ITS AV-Synch 2009 experiment showed a close-up of a face viewed face-on and spanning about half of the screen vertically for the entire sequence, with nothing of particular interest occurring in the background. The remaining two scenes showed a close-up of a face from the side, and a boxing scene somewhat similar to Heavy Bag. Thus, perhaps perceiving synchronization was a particularly easy task. By contrast, ITS AV-Synch 2010 was intended to span a more typical range of television content. Only Boys Talk and News Drive show a close-up of a face throughout the entire sequence, and in News Drive the cars driving past in the background create additional stimuli and impairments that might distract viewers from synchronization clues.

Figure 5 shows the same plot as Figure 4 but for each scene separately. These plots indicate scene-dependent influences on the delay variable. Boys Talk and Big Green show a steadily increasing linear impact of delay on MOS, spanning a range of approximately 1.0 MOS. Mall Music’s plot is nearly flat, spanning around 0.25 MOS. News E-coli is the only scene where the direction of the synchronization error might be important (note that negative delays mean audio leads video, and positive delays mean that video leads audio). News Tennis appears to have a step function, where synchronization errors below 200 ms are ignored, while synchronization errors greater than 300ms appear to be equally objectionable.

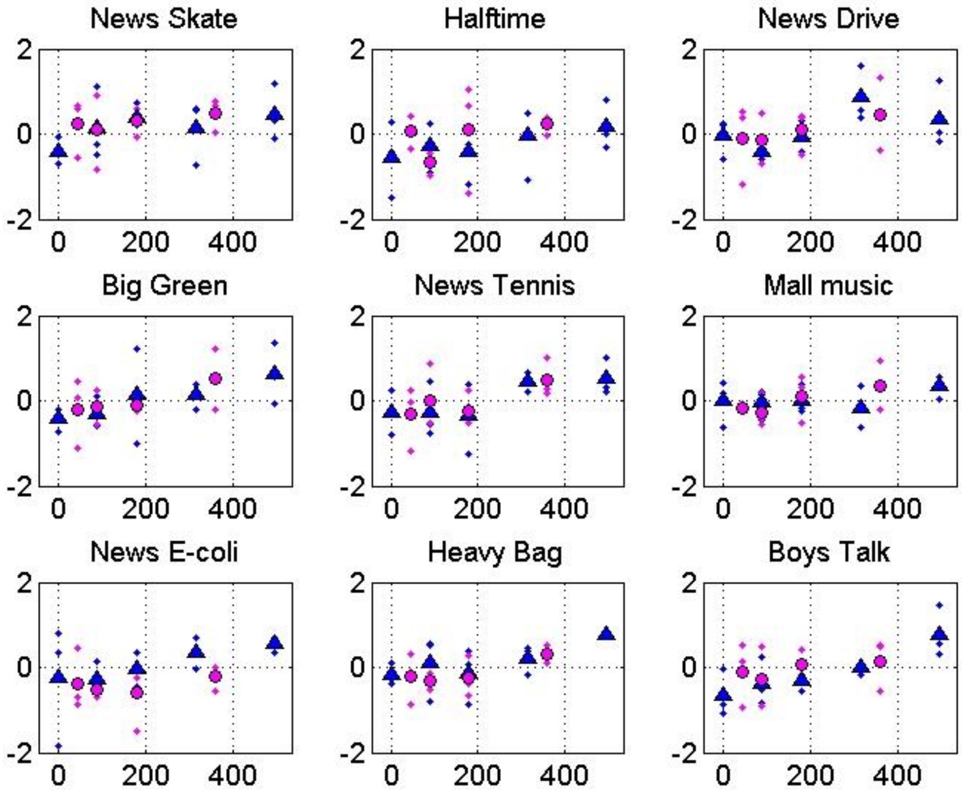

Figure 5. Same information as seen in Figure 4, but separated out by scene. Axis labels are omitted to improve visibility. X-axis is the absolute value of delay; y-axis plots the multiplicative model's error: $\hat{y}$ - MOS.

Figure 6 shows the same type of plot as Figure 4 but separated by audio level and video level. For example, audio level 1 with video level 1 is the upper-left corner, and depicts the best quality level for each. This subplot shows the strongest decrease in MOS (on the y-axis) as the synchronization error increases (on the x-axis). The worst audio level (3) is on the bottom, and the worst video level (3) is along the right. These subplots show a more subdued decrease in MOS as the synchronization error increases.

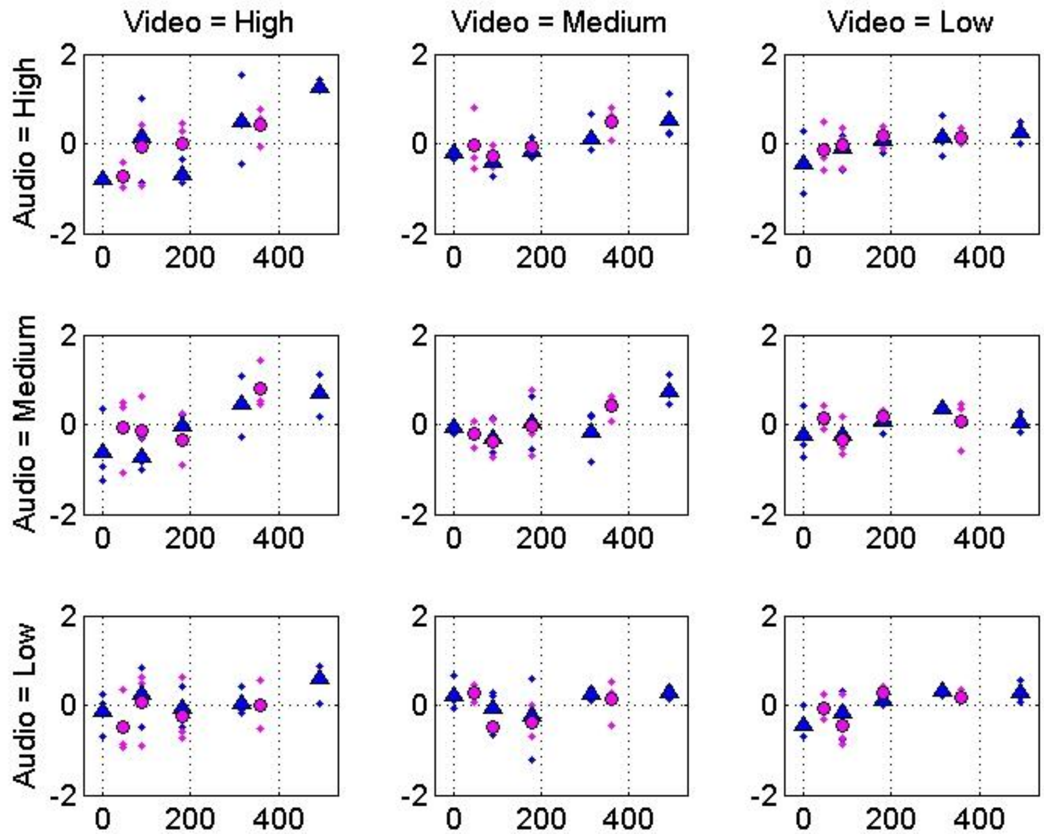

Figure 6. Same plots as seen in Figure 4, but separated out by audio impairment level and video impairment level. Impairment levels are 1=high (the original), 2=medium and 3=low. X-axis is absolute value of delay; y-axis plots the multiplicative model's error: $\hat{y}$ - MOS.

These patterns and the relatively low correlation of the best model seen in Table 2 indicates that scene-dependent differences in perceptual response to audiovisual synchronization errors (from Figure 5) and the audio and video quality interactions (from Figure 6) require additional study. Additional subjective testing would be needed to identify scene characteristics responsible for the variation in these MOS responses, and to establish reliable patterns for the quality interactions.

4 ACKNOWLEDGMENTS

The ITS experiments presented herein were only possible due to video quality expertise from Stephen Wolf, audio quality expertise from Stephen Voran and Andrew Catellier, and computer programming from Scott Hanes.

5 REFERENCES

[1] ANSI-Accredited Committee T1 Contribution, T1A1.5/93-104, "Report on an experimental combined audio/video subjective test method," Jul., 1993, Bellcore, USA.

[2] ANSI-Accredited Committee T1 Contribution, T1A1.5/94-141, "Report on extension of combined audio/video quality model," Jul., 1993, Bellcore, USA.

[3] ANSI-Accredited Committee T1 Contribution, T1A1.5/94-124, "Combined A/V model with Multiple Audio and Video Impairments," Apr., 1995, Bellcore, USA.

[4] C. Jones and D. Atkinson, "Development of opinion-based audiovisual quality models for desktop video-teleconferencing," *Record of the 6th IEEE International Workshop on Quality of Service*, May 1998, Napa, California. Available at www.its.bldrdoc.gov/n3/video/documents.htm.

[5] ITU-T Contribution COM12-61-E "Study of the influence of experimental context on the relationship between audio, video, and audiovisual subjective qualities," Noel Chateau, France Telecom/CNET, Sep. 1998.

[6] ITU-T Contribution COM 12-19-E, "Relations between audio, video, and audiovisual quality," KPN Research, The Netherlands, Feb. 1998.

[7] J. G. Beerends and F. E. de Caluwe: "The influence of video quality on perceived audio quality and vice versa." *J. Audio Eng. Soc.*, vol. 47, no.5, pp. 355–362, May 1999.

[8] D. S. Hands, "A basic multimedia quality model," *IEEE Trans. Multimedia*, vol. 6, no. 6, pp. 806-816, Dec. 2004.

[9] S.Winkler and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content." *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 973–980, Oct. 2006.

[10] M. N. Garcia, A. Raake, "Impairment-factor based audio-visual quality model for IPTV," *International Workshop on Quality of Multimedia Experience (QoMEx)*, Jul. 2009.

[11] M. McFarland, M. Pinson, C. Ford, A. Webster, W. Ingram, S. Hanes and K. Anderson, "Relating Audio and Video Quality Using CIF Video," NTIA Technical Memorandum TM-10-472, Sep. 2010.

[12] M. H. Pinson, W. Ingram and A. Webster, "Audiovisual Quality Component Analysis," *IEEE Signal Processing Special Issue on Multimedia Quality Assessment*, submitted for publication.

[13] "ATSC implementation subcommittee finding: relative timing of sound and vision for broadcast operations," Doc. IS-191, Jun. 2003.

[14] ITU-T Rec. BT.1359-1, "Relative timing of sound and vision for broadcasting," Geneva, 1998.

[15] A. C. Younkin and P. J. Corriveau, “Determining the Amount of Audio-Video Synchronization Errors Perceptible to the Average End-User,” *IEEE Transactions on Broadcasting*, vol. 54, no. 3, Sep. 2008.

[16] C. Ford, M. McFarland, W. Ingram, S. Hanes, M. Pinson, A. Webster and K. Anderson, “Multimedia Synchronization Study,” NTIA Technical Memorandum TM-10-464, Oct. 2009. Available at www.its.bldrdoc.gov/pub/pubs.php.

[17] B. Cotton, “New reference condition for very low bit rate voice coder evaluation,” CCITT SGXII Contribution D.108, Sep. 1991.

[18] S. Voran, “Observations on the t-reference condition for speech coder evaluation,” CCIT SGXII Contribution SQ 13.92, Feb. 1992. Available at www.its.bldrdoc.gov/audio.

FORM **NTIA-29**
(4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION NO. TM-11-474	2. Government Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE Preliminary Investigation into the Impact of Audiovisual Synchronization of Impaired Audiovisual Sequences		5. Publication Date March 2011 6. Performing Organization Code ITS.
7. AUTHOR(S) Margaret H. Pinson, Arthur Webster, William Ingram		9. Project/Task/Work Unit No. 3139011-300
8. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Telecommunication Sciences National Telecommunications & Information Administration U.S. Department of Commerce 325 Broadway Boulder, CO 80305		10. Contract/Grant Number.
11. Sponsoring Organization Name and Address National Telecommunications & Information Administration Herbert C. Hoover Building 14th & Constitution Ave., NW Washington, DC 20230		12. Type of Report and Period Covered NTIA Technical Memorandum
14. SUPPLEMENTARY NOTES		

15. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

The quality perception of an audiovisual sequence is heavily influenced by the quality of the audio, the quality of the video, and the audiovisual time synchronization. The questions then arise: what is the relative importance of each factor, and can a model be devised that is generally applicable? Previous work either examined the relative influences of audio and video quality for synchronized video or investigated the quality impact of synchronization errors on unimpaired video sequences. This experiment is a first attempt to combine all three factors into a single experiment, to judge the complex interactions among individual measurements of audio and video quality and synchronization errors.

16. Key Words (Alphabetical order, separated by semicolons)

audio quality, audiovisual quality, subjective testing, synchronization, video quality

17. AVAILABILITY STATEMENT ☒ UNLIMITED. ☐ FOR OFFICIAL DISTRIBUTION.	18. Security Class. (This report) Unclassified 19. Security Class. (This page) Unclassified	20. Number of pages 22 21. Price:

NTIA FORMAL PUBLICATION SERIES

NTIA MONOGRAPH (MG)

A scholarly, professionally oriented publication dealing with state-of-the-art research or an authoritative treatment of a broad area. Expected to have long-lasting value.

NTIA SPECIAL PUBLICATION (SP)

Conference proceedings, bibliographies, selected speeches, course and instructional materials, directories, and major studies mandated by Congress.

NTIA REPORT (TR)

Important contributions to existing knowledge of less breadth than a monograph, such as results of completed projects and major activities. Subsets of this series include:

NTIA RESTRICTED REPORT (RR)

Contributions that are limited in distribution because of national security classification or Departmental constraints.

NTIA CONTRACTOR REPORT (CR)

Information generated under an NTIA contract or grant, written by the contractor, and considered an important contribution to existing knowledge.

JOINT NTIA/OTHER-AGENCY REPORT (JR)

This report receives both local NTIA and other agency review. Both agencies' logos and report series numbering appear on the cover.

NTIA SOFTWARE & DATA PRODUCTS (SD)

Software such as programs, test data, and sound/video files. This series can be used to transfer technology to U.S. industry.

NTIA HANDBOOK (HB)

Information pertaining to technical procedures, reference and data guides, and formal user's manuals that are expected to be pertinent for a long time.

NTIA TECHNICAL MEMORANDUM (TM)

Technical information typically of less breadth than an NTIA Report. The series includes data, preliminary project results, and information for a specific, limited audience.

For information about NTIA publications, contact the NTIA/ITS Technical Publications Office at 325 Broadway, Boulder, CO, 80305 Tel. (303) 497-3572 or e-mail info@its.bldrdoc.gov.

This report is for sale by the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161,Tel. (800) 553-6847.