Russell L. Storms
storms@airmics.gatech.edu
Army Research Laboratory
Georgia Institute of Technology
Atlanta, Georgia 30332-3142

Michael J. Zyda
zyda@acm.org
Naval Postgraduate School
Monterey, California 93943-5118

Interactions in Perceived Quality of Auditory-Visual Displays

Abstract

The quality of realism in virtual environments (VEs) is typically considered to be a function of visual and audio fidelity mutually exclusive of each other. However, the VE participant, being human, is multimodal by nature. Therefore, in order to validate more accurately the levels of auditory and visual fidelity that are required in a virtual environment, a better understanding is needed of the intersensory or cross-modal effects between the auditory and visual sense modalities. To identify whether any pertinent auditory-visual cross-modal perception phenomena exist, 108 subjects participated in three experiments which were completely automated using HTML, Java, and JavaScript programming languages. Visual and auditory display quality perceptions were measured intra- and intermodally by manipulating the pixel resolution of the visual display and Gaussian white noise level, and by manipulating the sampling frequency of the auditory display and Gaussian white noise level. Statistically significant results indicate that high-quality auditory displays coupled with high-quality visual displays increase the quality perception of the visual displays relative to the evaluation of the visual display alone, and that low-quality auditory displays coupled with high-quality visual displays decrease the quality perception of the auditory displays relative to the evaluation of the auditory display alone. These findings strongly suggest that the quality of realism in VEs must be a function of both auditory and visual display fidelities inclusive of each other.

1 Motivation

1.1 Motivation

The fidelity requirements for VEs have traditionally focused on the singular modality of vision. As a result, in an attempt to render visual displays as close as possible to the fidelity of the human visual system, the fidelity of visual display systems has increased dramatically in the last decade. Likewise, as a result of better audio technology, there has been a recent surge of emphasis on the fidelity requirements concerning the singular modality of audition. As a result, the fidelity of auditory display systems has increased dramatically in the last five years. These rapid advances in visual and auditory display technologies have helped to create increasingly realistic virtual environments. Their quality of realism is typically considered to be a function of visual and auditory fidelity mutually exclusive of each other as presented in Barfield et al. (1995), but herein lies a problem: the virtual environment participant, being human, is multimodal by nature. Thus, the quality of realism in virtual environments needs to be based on multimodal criteria that comprise all of our senses, as opposed to the current use of singular modality criteria. As such, the fidelity

Presence, *Vol. 9, No. 6, December 2000,* 557–580

requirement of virtual environments must similarly also be based on multimodal criteria that comprise all of our senses. However, insufficient experimental data exists to make informed multimodal design decisions.

1.2 Objective

Because of current limitations in today's computer technology, it is impossible to render to the interactive VE participant realistic information to all senses in real time. However, due to the significant advances in visual and auditory display technology, it is appropriate to concentrate on the vision and audition sensory modalities. As such, the objective of this effort correspondingly focuses on the two sensory modalities of vision and audition. By gaining a better understanding of auditory-visual cross-modal effects, system designers can more accurately verify and validate the levels of auditory and visual fidelity that are required for the immersed VE participant.

1.3 Scope

The results of this effort are intended to aid the VE, simulations, and gaming developer in creating better virtual worlds, simulations, games, and the like through an appropriate use of auditory and visual display fidelities that are based on auditory-visual cross-modal perception phenomena. It is important to note that the scope of this effort is not to identify absolute visual and/or auditory fidelity requirements (such as pixel resolution and sampling frequency, respectively), but rather to identify the effects of auditory-visual cross-modal perception phenomena that can be used to justify a certain level of auditory and/or visual fidelity.

1.4 Approach

To identify whether relevant auditory-visual cross-modal perception phenomena exist, the approach taken is that of the experimental psychologist. A series of three experiments investigates the existence of pertinent auditory-visual cross-modal perception interactions. Each experiment is completely automated using HTML, Java, and JavaScript (Flanagan, 1996; Ladd & O'Donnell, 1998). All experiments are conducted at the Naval Postgraduate School (NPS) in Monterey, California. A total of 108 volunteer participants—comprising students, faculty, staff, and guests of NPS—served as subjects. Each experiment involves a 3×3 factorial within-subjects design. The two independent variables are visual and auditory display quality having three levels, each consisting of low, medium, and high qualities. The visual display parameters manipulated are pixel resolution and Gaussian white noise level, and the auditory display parameters manipulated are sampling frequency and Gaussian white noise level. Partial counterbalancing is achieved through the technique of balanced Latin squares. The basic idea of the experiments is to manipulate visual and auditory display parameters intramodally and intermodally, and likewise to measure visual and auditory display perception intramodally and intermodally. During the experiments, each of which lasts approximately thirty minutes, a single subject wears headphones and sits in front of a 20 in. display monitor. The subject's task is to rate the perceived quality of auditory-only, visual-only, and combined auditory-visual displays through Likert rating scales ranging from 1 (low) to 7 (high). Thus, the dependent variables are the perception of visual display quality and the perception of auditory display quality. It is hoped that, by varying the fidelity of both auditory and visual displays, it will be possible to measure auditory-visual cross-modal perception interactions. Specifically, this effort aims to answer the following question: in an auditory-visual display, what effect (if any) does auditory quality have on the perception of visual quality and vice versa? Specifically:

1. Does a high-quality auditory display coupled with a low-quality visual display cause a decrease/increase in the perception of audio quality and/or an increase/decrease in the perception of visual quality relative to established baseline conditions derived from auditory-only and visual-only quality perception evaluations?
2. Does a low-quality auditory display coupled with a high-quality visual display cause an increase/de-

crease in the perception of audio quality and/or a decrease/increase in the perception of visual quality relative to established baseline conditions derived from auditory-only and visual-only quality perception evaluations?

2 Visual and Auditory Display Development

2.1 Introduction

The visual display selected for this study is a radio, and the auditory display is a selection of music. The rationale for choosing a radio and music is based on the eventual coupling of the auditory and visual displays to form a combined auditory-visual display. Based on psychological factors such as Gestalt perceptual grouping theory (Wertheimer, 1912; Koffka, 1935; Kohler, 1940; Garner, 1970; Murch, 1973) and visual dominance and the ventriloquism effect (Howard & Templeton, 1966; Pick, Warren, & Hay, 1969; Bermant & Welch, 1976; Radeau & Bertelson, 1976; Warren, Welch, & McCarthy, 1981; Ragot, Cave, & Fano, 1988), an auditory-visual display consisting of a radio and music might be perceptually grouped together, thereby producing a more tightly coupled display. In a higher cognitive sense, we are likely to associate music (audio) with a radio (visual).

2.2 Visual-Display Development

To obtain the visual image of a radio, a photograph of a radio was taken from the book *Radios by Hallicrafters with Price Guide* (Chuck Dachis, 1995). This radio image is then digitized using a flatbed scanner at 600×600 pixel resolution as depicted in Figure 1. This particular radio is chosen because it contains various features including letters and numbers, smooth and rough surfaces, straight and curved lines, patterns (on the speaker), and reflections. The reason for having numerous features is to provide test subjects with a wide variety of cues from which to make their quality ratings.

Using the original scanned image at 600 pixels/inch, Adobe Photoshop is then used to make various copies with degraded pixel resolutions but all having the same dimensions, the size of which nearly fills the display area of a 20 in. computer monitor. Approximately thirty images of the radio, ranging from 200 to 600 pixels/inch, are produced. The next step involves establishing levels of pixel resolution that are noticeably different, but not just-noticeably-different or obviously different. The goal is to establish low-, medium-, and high-quality visual displays for use in the main experiment.

The basic idea is to create changes in pixel resolution that the subject can distinguish, but only with some effort. This process of establishing the noticeable levels of pixel resolution is very time consuming. Preliminary subjects are presented seven images of the radio with varying levels of pixel resolution (using the same graphics accelerator and computer monitor chosen for the experiment, as described later). A subject is then asked to arrange (if possible) the images in ascending or descending order of quality. After repeating this process with fifteen subjects, a consensus is finally reached ultimately determining the low-, medium-, and high-quality visual displays of the radio to be used in the three main experiments of this study. Numerous factors can affect the final rendering of the visual display such as: computer monitor specifications, computer monitor desk size (user selected resolution), video/graphics accelerator specifications, and software application graphics rendering capabilities. Nevertheless, a relative quality ordering of the visual displays is established, for the intent of this research effort is to focus on the perceptual effects of various quality visual displays, and not on the absolute levels of pixel resolution that determine these various quality displays. It is important to note that even the high-quality visual display has some, albeit slight, degradation of pixel resolution. The reason for this is based on the design of the experiment, the goal of which is to have three noticeably different quality displays based on pixel resolution, and not to have one display with absolutely no perceivable pixel-resolution degradation and two displays which do have pixel-resolution degradation. If this were the case, the unwanted issue of absence or presence of noticeable pixel resolution is introduced. As such, subjects might be compar-

Figure 1. *Radio image (Dachis, 1995).*

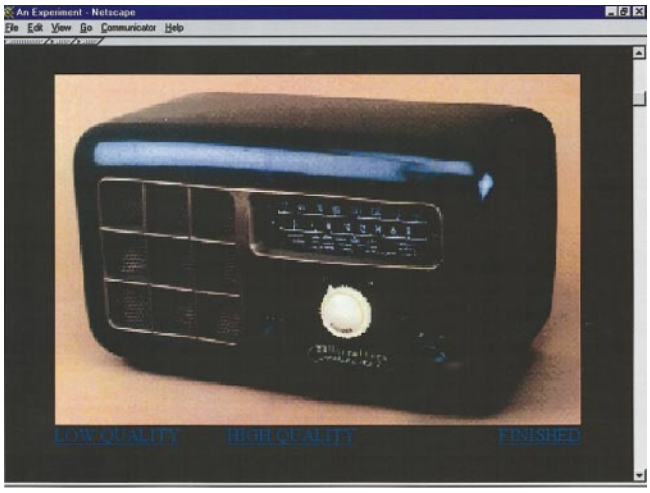


Figure 2. *Experiment 1: low-quality visual display familiarization.*

ing the one display with no perceivable pixel-resolution degradation to the two displays which do have pixel-resolution degradation. Thus, to ensure that subjects are making quality ratings based only on degree of pixel resolution (and not absence or presence), the high-quality display must also have a small amount of perceivable pixel-resolution degradation. To establish the low-, medium-, and high-quality visual displays for use in the second experiment (described later), the same process is repeated, using the original scanned image of the radio at 600 pixels/inch but with varying degrees of Gaussian noise levels.

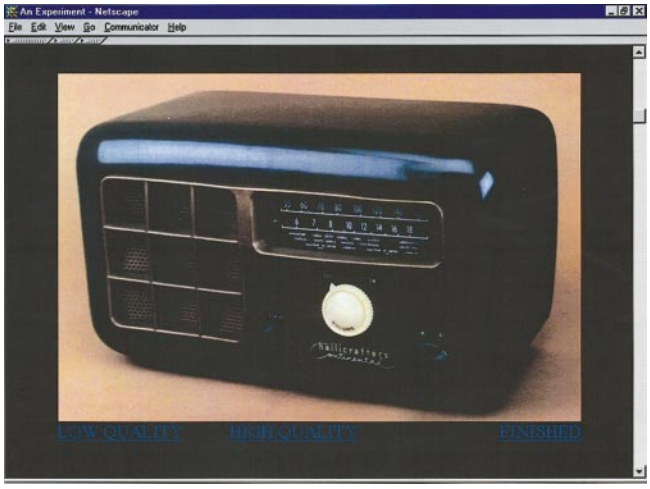


Figure 3. *Experiment 1: high-quality visual display familiarization.*

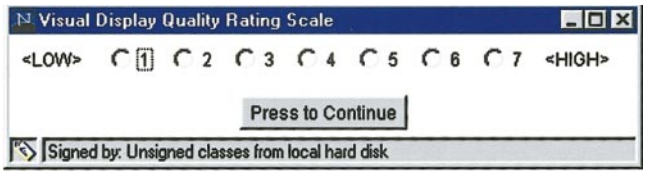


Figure 4. *Experiment 1: visual display quality rating scale.*

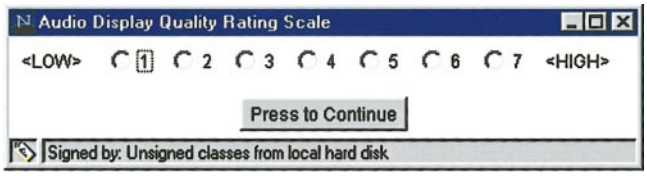


Figure 5. *Experiment 1: auditory display quality rating scale.*

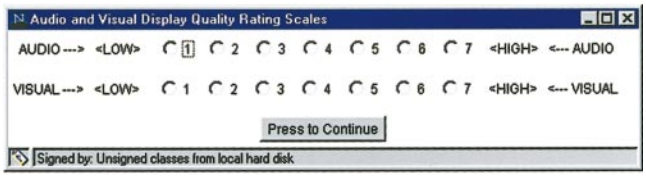


Figure 6. *Experiment 1: combined auditory-visual rating scale.*

2.3 Auditory-Display Development

In constructing the auditory displays for this experiment, the only consideration was the choice of musical content. Because one of the quality parameters to be manipulated in this study is sampling frequency, a conscious decision is made not to include vocals (speech). The rea-

son for this is that the frequency range of speech is much less than that of typical musical instruments. For example, if the sampling frequency of music containing vocals is altered, the noticeable effect will be greater with the musical instruments than with the vocals. As such, if subjects focus on the vocals (which is fairly common), they might not be aware of any changes to the musical instruments. Therefore, choosing music without vocals eliminates the possibility of subjects focusing on the nonperceivable speech qualities. In terms of the type of music to use, choices considered were jazz, pop, rock, alternative, and classical. The consideration here is that, if a subject were familiar with the music, the subject might have some preconceived expectations or might make unwanted comparisons from a previous listening experience to the auditory display that is to be evaluated. To reduce the chance that subjects might have previously heard the music, an obscure portion of alternative music is selected. The music is taken from the song "A Forest" from the CD *Mixed Up* by a group called The Cure. (The use of the music is by courtesy of Elektra Entertainment Group, a division of Warner Communications, Inc.)

Using the *Mixed Up* CD, a twenty-second selection of "A Forest" is recorded into Sonic Foundary's *SoundForge* (1998) at 44.1 kHz (sampling frequency). The portion of music selected contains cymbals (among other instruments), resulting in a very wide frequency range of sound. *SoundForge* is then used to reproduce the 44.1 kHz, 20 sec. musical selection at numerous sampling frequencies ranging from 4 kHz to 44.1 kHz. Similar to creating the visual displays, the next step involves establishing sampling frequencies that are noticeably different, but not just-noticeably-different or obviously different. The goal is to establish low-, medium-, and high-quality auditory displays for use in the experiment, and the basic idea is to create changes in sampling rate that the subject could distinguish, but only with some effort. This process of establishing noticeable sampling frequencies is very time consuming. Preliminary subjects are presented seven music selections with varying sampling frequencies (using the same audio card and headphones chosen for the experiment, as described later). These subjects are then asked to arrange (if possible) the musical selections in ascending or descending order of quality. After repeating this process with fifteen preliminary subjects, a consensus was finally reached that ultimately determined the low-, medium-, and high-quality auditory displays of music to be used in the three main experiments of this study. A consensus also established a constant volume (loudness) setting for the auditory displays. Just as with the visual displays, a relative quality ordering is established with the auditory displays, because the intent of this research effort is to focus on the perceptual effects of various quality auditory displays, and not on the absolute sampling frequencies that determine these various quality displays. It is interesting to note that the high-quality auditory display, unlike the high-quality visual display, did not have to be slightly degraded to avoid the absence-or-presence degradation issue that was a concern with the visual displays. This is because our eyes are accustomed to a certain fidelity (quality), but our ears are not as discerning—which was readily apparent during the process of selecting the three auditory display qualities. When evaluating the various selections, not one subject could distinguish between 44.1 kHz or 22.05 kHz, which could be attributed to the various factors involved in the final rendering of the auditory display such as: how the original sound is produced, audio card specifications, rendering format (headphones, speakers, monophonic, stereo, spatialized, and so on), and rendering format specifications. Nevertheless, in terms of the higher visual and auditory qualities in this study, the ears were not as discerning when evaluating sampling frequency as the eyes were at evaluating pixel resolution or Gaussian noise levels.

2.4 Auditory-Visual Display Development

After establishing the visual and auditory displays, the next step is to develop the combined auditory-visual displays. The considerations here are determining how long to render the displays and synchronizing the rendering of both auditory and visual displays. To eliminate any potential confounds, the amount of time that a subject is given to view or hear the displays when presented separately must be the same amount of time given to view/hear the combined auditory-visual displays. During the process of establishing both the auditory and visual low-, medium-, and high-quality displays, subjects

are asked if they need more or less time to view or hear the appropriate displays. Based on a consensus, eight seconds is chosen for both displays. Interestingly, some subjects at first thought they needed more time (approximately twenty seconds), but, when given more time, the subjects realized that they were changing their minds too often about the quality and that, when it came time to rate the quality of the display, they forgot what they were thinking. The subjects then requested a shorter time duration. In a related experiment conducted to measure the scene-dependent quality variations in digitally coded television pictures (Aldridge et al., 1995), subjects were asked to assess distortions introduced by MPEG-2 coding. MPEG-2 sequences of ten and thirty seconds were used. One of the findings of this experiment was that the 30 sec. sequences were too long, because they exceeded the duration of human working memory (WM). WM duration is only approximately twenty seconds, and the rate of decay in WM is dependent on the amount of information (Peterson & Peterson, 1959; Wickens, 1992). Thus, the 8 sec. display duration chosen for this experiment is within VM constraints. Based on the earlier conducted preliminary subject consensus and human WM constraints, then, all displays during the three experiments, whether presented separately or in combination, are presented to the subject for eight seconds.

3 Data Analysis

In this experiment, all the quality ratings made by the subjects are considered ordinal data. The reason for this is that the quality ratings are derived from rating scales that are used to rank the quality perception of the displays on a scale of 1 (lowest) to 7 (highest). Furthermore, because this research does not assume a certain underlying distribution of the data, a nonparametric data analysis method is utilized. Specifically, a one-sample sign test is used to compare the number of observations above and below a certain hypothesized value, which in this case is zero. To answer the questions outlined earlier supporting the goal of this experiment, the one-sample sign test is used to investigate the following null hypotheses:

1. The difference between the visual-only quality rating of a combined auditory-visual display and the baseline rating for the visual-only quality display is zero.
2. The difference between the auditory-only quality rating of a combined auditory-visual display and the baseline rating for the auditory-only quality display is zero.
3. The difference between the visual quality rating of a combined auditory-visual display when also rating the auditory display and the baseline rating for the visual-only quality display is zero.
4. The difference between the auditory quality rating of a combined auditory-visual display when also rating the visual display and the baseline rating for the auditory-only quality display is zero.

Specifically, a one-sample sign test is used to compare the number of observations above and below the difference in the baseline ratings for the auditory-only and visual-only quality displays and

- the visual-only quality rating of a combined auditory-visual display,
- the auditory-only quality rating of a combined auditory-visual display,
- the visual quality rating of a combined auditory-visual display when also rating the auditory display, and
- the auditory quality rating of a combined auditory-visual display when also rating the visual display.

The data analysis derived from the one-sample sign test forms the foundation from which all major findings in this research effort are derived. All significant findings of this research effort are set at an alpha level of 0.05. In other words, the degree of confidence supporting all experimental findings is at the 0.05 level.

4 Experiment 1: Static Resolution

4.1 Introduction

Experiment 1: Static Resolution investigates the perceptual effects from manipulating visual display pixel

resolution and auditory display sampling frequency. The visual display consists of a static image of the aforementioned radio, and the auditory display is a selection of music. The goal of this experiment is to answer the following questions:

1. Does a high-quality auditory display coupled with a low-quality visual display cause a decrease/ increase in the perception of audio quality and/or an increase/decrease in the perception of visual quality relative to established baseline conditions derived from auditory-only and visual-only quality perception evaluations?
2. Does a low-quality auditory display coupled with a high-quality visual display cause an increase/decrease in the perception of audio quality and/or a decrease/increase in the perception of visual quality relative to established baseline conditions derived from auditory-only and visual-only quality perception evaluations?

4.2 Location

All sessions of Experiment 1: Static Resolution is conducted in an isolated room under the same ambient conditions. Before each session, the following conditions prevail.

- All nonessential electronic equipment is turned off.
- Telephones are unplugged.
- Windows are closed and covered with blackout cloth.
- All overhead lights are turned off.
- A 60 W incandescent desk lamp is turned on behind the computer monitor to eliminate any glare.
- The entry door to the room is closed.
- A *Do Not Disturb* sign is placed on the outside of the door.
- The subject is asked to turn off any audible pagers, mobile phones, and/or watches.

4.3 Participants

Thirty-six volunteer participants (eighteen female and eighteen male, composed of students, faculty, staff, and guests of NPS) served as subjects. Based on the preliminary findings of a previously conducted pilot study, the number of male and female subjects in this experiment is balanced. The average age of the subjects is 36.5 years, ranging in age from 15 to 63. (Two female subjects did not give their age.) All subjects are required to have 20/20 or corrected-to-20/20 vision and normal hearing. Because the experiment did not involve precise measurements of pixel resolution or sampling frequency, a vision and hearing test was not needed. Before conducting the experiment, each subject was asked, as part of a voluntary consent form, if he or she meets the vision and hearing requirements.

4.4 Apparatus

The main hardware platform of the experiment was a Pentium 200 MHz (MMX) personal computer with 64 MB of main memory running Microsoft Windows *95*. The auditory displays are generated by a Sound Blaster 64 AWE Gold audio card and rendered via Sennheiser HD 540 reference II headphones. The visual displays are generated by a Diamond Multimedia Viper V330 128-bit graphics accelerator card and rendered via a Sony Multiscan 20 in. sfII computer monitor (set at 800×600 resolution). The entire automated experiment is contained within a Netscape Communicator 4.05 HTML browser window using JavaScript to render the visual-only, auditory-only, and combined auditory-visual displays. Java pop-up windows, developed using Sun's Java Development Kit (JDK) 1.1.5, are used to collect subject responses.

4.5 Procedure

The experiment involves a 3×3 factorial within-subjects design. The two independent variables are visual and audio display quality; the two dependent variables are the corresponding quality perception of the auditory and visual displays. The three levels of the visual quality independent variable consist of low-, medium-, and high-quality visual displays of the radio image depicted earlier having resolutions of 350, 450, and 550 pixels/inch, respectively. The three levels of the audi-

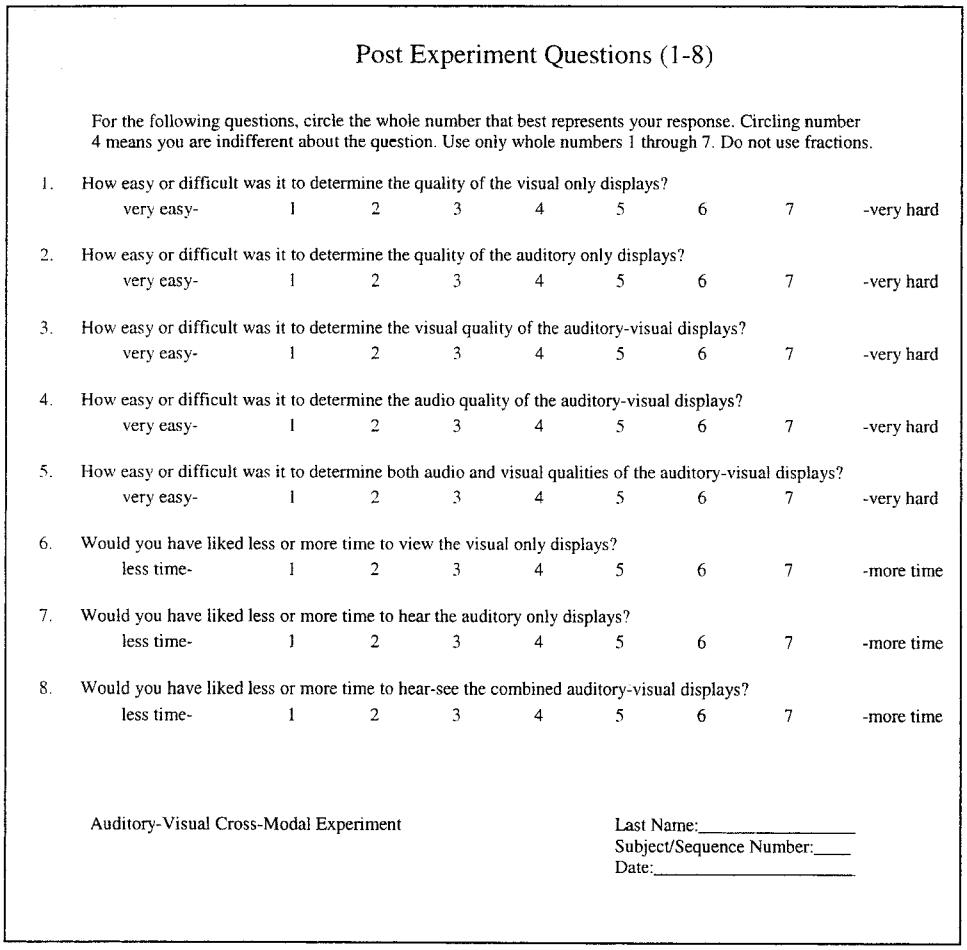
tory quality independent variable consist of low-, medium-, and high-quality auditory displays of the same music selection presented monophonically, having sampling rates of 11 kHz, 23 kHz, and 35 kHz, respectively. The visual display parameters manipulated are pixel resolution, and the auditory display parameters manipulated are sampling frequency. During the experiment, which lasts approximately thirty minutes, each subject wears headphones and sits in front of a 20 in. computer display monitor. The task of the subject is to rate the perceived quality of auditory-only, visual-only, and combined auditory-visual displays via Likert rating scales ranging from 1 (low) to 7 (high).

After reading a brief experimental overview and signing a voluntary consent form, the subject is seated in a chair facing the computer monitor. The subject is instructed to adjust the seat height and/or monitor orientation to that which is most comfortable and which represents his/her typical computer monitor placement. Although a standard viewing position/orientation is much desired in experimental design, the focus of this experiment is not on precision, but rather perception. Accordingly, the idea is for the subject to be relaxed, comfortable, and in his/her typical viewing position/orientation. Nevertheless, no subject sat closer than approximately one foot or farther than approximately three feet from the computer monitor. The subjects are instructed on how to wear and fit the headphones and also how to adjust the volume, if necessary. To maintain identical testing conditions, it was hoped that no one would need to adjust the headset volume, and no one did.

Once the subject is seated and wearing the headphones, an automated computer program contained within an HTML browser window instructs the subject to enter some personal data via the keyboard. These personal data are used to create a unique data file to collect the specific subject's data for the remainder of the experiment. This is the only time that the keyboard is utilized; for the remainder of the experiment, only the mouse is needed. The automated experiment continues by presenting the subject with a series of instructions giving a full explanation of what is and is not required of the subject. The visual-only, auditory-only, and combined auditory-visual displays are rendered via JavaScript, and Java pop-up windows collect subject responses.

As the automated experiment continues, the subject is first presented a familiarization process that includes a series of instructions, visual and auditory displays, and rating scales in order to ensure that the headphones are working properly, familiarize the subject with how the visual displays will be presented on the computer monitor, and familiarize the subject with what the rating scales look like, how they will appear and disappear automatically, and how to use them. After this familiarization process, the next task is for the subject to memorize the quality differences between the lowest- and highest-quality visual displays. During this memorization process, the subject calibrates himself to the maximum possible quality range spanned by the low- and high-quality extremes. The low- and high-quality extremes correspond to the low and high behavior anchors, respectively. During this process, the subject has direct control in viewing the low- and high-quality displays simply by clicking on either the *LOW QUALITY* or *HIGH QUALITY* hypertext link. Figure 2 depicts the appearance of the low-quality visual display, having 250 pixels/inch, and figure 3 depicts the appearance of the high-quality visual display, having 600 pixels/inch. Note that the actual pixel resolution experienced by the subject can be viewed only on the actual 20 in. computer monitor utilized in the experiment. However, the low- and high-quality displays depicted in figure 2 and figure 3 are fairly good representations of the quality difference between the actual displays used in the experiment. When the subject is ready to begin rating the visual displays, he or she clicks on the *FINISHED* hypertext link. After another set of instructions (and when the subject is ready to begin making quality ratings), a visual display is then rendered for eight seconds, after which it automatically disappears, and a Java pop-up window automatically appears to facilitate rating the visual display as depicted in figure 4. The subject rates a total of nine visual-only displays: three of each quality (low, medium, and high), presented in random order.

After rating the visual-only displays, the subject uses the same process, as with the visual displays, to memorize the quality differences between the lowest- and highest-quality auditory displays. The lowest- and highest-quality auditory displays correspond to 8 kHz and



Post Experiment Questions (1-8)

For the following questions, circle the whole number that best represents your response. Circling number 4 means you are indifferent about the question. Use only whole numbers 1 through 7. Do not use fractions.

1. How easy or difficult was it to determine the quality of the visual only displays?
   very easy- 1 2 3 4 5 6 7 -very hard
2. How easy or difficult was it to determine the quality of the auditory only displays?
   very easy- 1 2 3 4 5 6 7 -very hard
3. How easy or difficult was it to determine the visual quality of the auditory-visual displays?
   very easy- 1 2 3 4 5 6 7 -very hard
4. How easy or difficult was it to determine the audio quality of the auditory-visual displays?
   very easy- 1 2 3 4 5 6 7 -very hard
5. How easy or difficult was it to determine both audio and visual qualities of the auditory-visual displays?
   very easy- 1 2 3 4 5 6 7 -very hard
6. Would you have liked less or more time to view the visual only displays?
   less time- 1 2 3 4 5 6 7 -more time
7. Would you have liked less or more time to hear the auditory only displays?
   less time- 1 2 3 4 5 6 7 -more time
8. Would you have liked less or more time to hear-see the combined auditory-visual displays?
   less time- 1 2 3 4 5 6 7 -more time

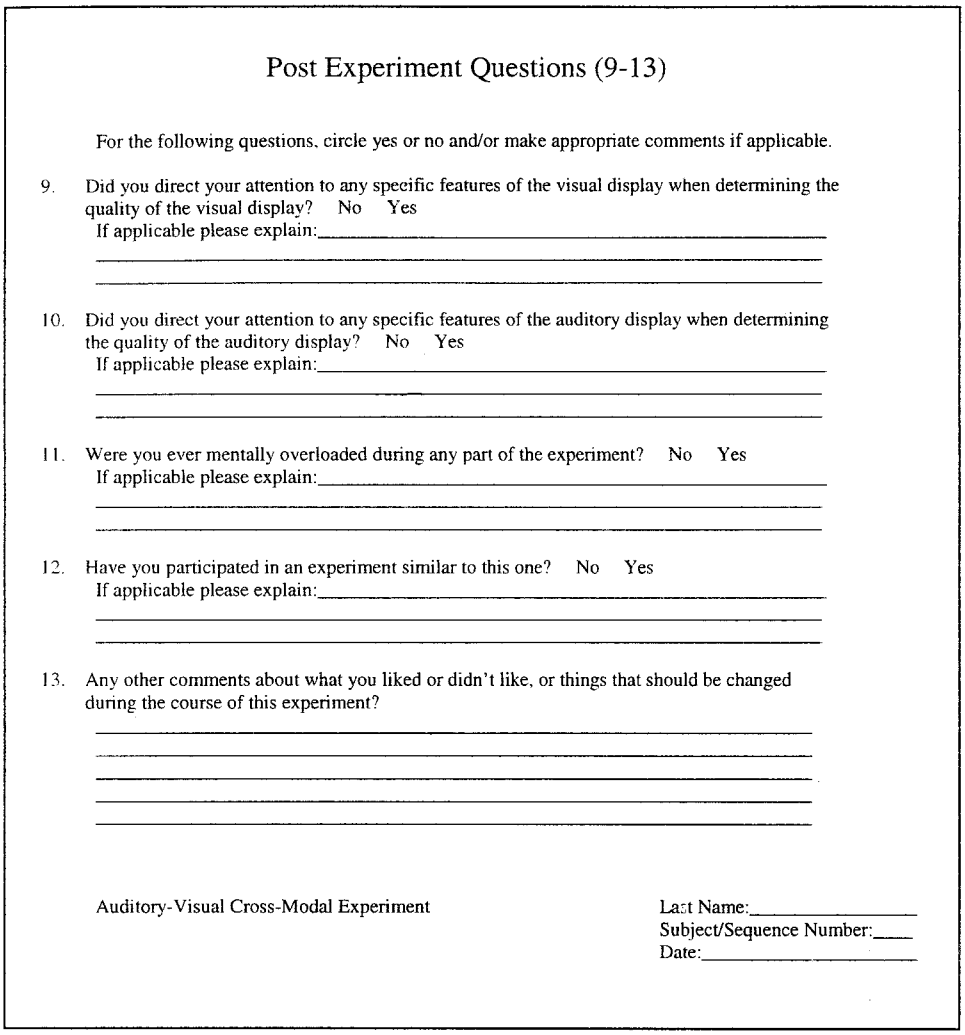
Auditory-Visual Cross-Modal Experiment

Last Name:_____
Subject/Sequence Number:____
Date:_____

Figure 7. *Postexperiment questions 1 through 8.*

44.1 kHz, respectively. The subject uses the exact same process as used with the visual displays to rate nine auditory-only displays (three of each quality presented in random order) by using the auditory rating scale as depicted in figure 5.

After rating the auditory displays, the subject is presented with instructions and then rates only the visual quality of nine combined auditory-visual displays. (The nine permutations of the auditory and visual qualities are partially counterbalanced through the Latin squares technique.) The subject is next presented with instructions and then rates only the auditory quality of nine combined auditory-visual displays. Finally, the subject is presented with instructions and then rates both the visual and auditory displays of eighteen combined auditory-visual displays. After each of the eighteen combined auditory-visual displays is presented (the nine permutations of the auditory and visual qualities are partially counterbalanced through the Latin squares technique, and then presented in reverse order for a total of eighteen combined auditory-visual ratings), the subject rates both the auditory and visual displays using the combined auditory-visual rating scale depicted in figure 6. After the subject has completed rating all of the displays, the automated portion of the experiment terminates. The subject then completes a brief postexperiment survey, consisting of thirteen questions (as depicted in figure 7 and figure 8.)



Post Experiment Questions (9-13)

For the following questions, circle yes or no and/or make appropriate comments if applicable.

9. Did you direct your attention to any specific features of the visual display when determining the quality of the visual display? No Yes
If applicable please explain:_____

10. Did you direct your attention to any specific features of the auditory display when determining the quality of the auditory display? No Yes
If applicable please explain:_____

11. Were you ever mentally overloaded during any part of the experiment? No Yes
If applicable please explain:_____

12. Have you participated in an experiment similar to this one? No Yes
If applicable please explain:_____

13. Any other comments about what you liked or didn't like, or things that should be changed during the course of this experiment?
_____

Auditory-Visual Cross-Modal Experiment

Last Name:_____
Subject/Sequence Number:____
Date:_____

Figure 8. *Postexperiment questions 9 through 13.*

After completing the postexperiment questions, the subject is allowed to ask any overall questions about the experiment. The experiment is then terminated, and the subject is free to go.

4.6 Experimental Validity

The first and most important consideration is whether the quality of the visual-only and auditory-only displays developed for this experiment are rank-ordered by the subjects according to their intended rankings. If this were not the case, the validity of the experiment would be jeopardized. However, in looking at figure 9, one can see that the overall quality ratings of the visual-only displays are properly rank-ordered by the subjects according to this experiment's intended low-, medium-, and high-quality rankings. Likewise, in looking at figure 10, one can see that the overall quality ratings of the auditory-only displays are properly rank-ordered by the subjects according to this experiment's intended low-, medium-, and high-quality rankings. Given that the data regarding quality of all displays are properly rank-ordered, data analysis with respect to the null hypotheses can continue.

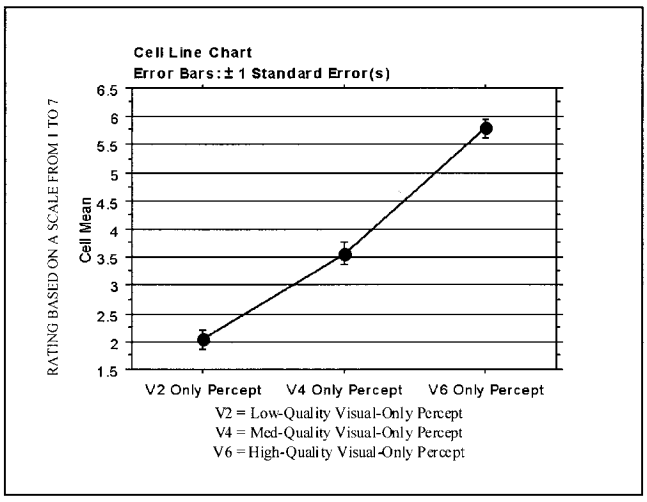


Figure 9. *Experiment 1: visual-only quality percept ratings.*

4.7 Findings

In terms of the first null hypothesis, when presented a combined high-quality visual and high-quality auditory display and asked only to rate the quality of the visual display, a statistically significant finding at the 0.0161 level (a *p*-value of 0.0161) suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

In terms of the second null hypothesis, when presented a combined low-quality auditory and high-quality visual display and when asked only to rate the quality of the auditory display, a statistically significant finding at the 0.0002 level strongly suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display.

In terms of the third null hypothesis, there are no significant findings at the 0.05 level. However, it is worth mentioning that, when presented a combined high-quality visual display coupled with either a medium- or high-quality auditory display and asked to rate both auditory and visual displays, the results at the 0.10 level suggest that the quality perception of the high-quality visual display is increased.

In terms of the fourth null hypothesis, when presented a combined low-quality auditory and high-quality visual display and when asked to rate both auditory and visual displays, a statistically significant finding at the 0.0107 level suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display. Also, when presented a combined high-quality auditory and low-quality visual display and asked to rate both auditory and visual displays, a statistically significant finding at the 0.0241 level suggests that the quality perception of a high-quality auditory display is increased when coupled with a low-quality visual display.

In terms of the postexperiment questions, the results indicate that determining the quality of both auditory and visual displays of a combined auditory-visual display proved to be more difficult than determining the quality of either auditory or visual display presented either alone or in combination. Furthermore, the results indicate that eight seconds is an adequate amount of time to rate the visual-only and auditory displays, but that slightly more than eight seconds is desired when rating the combined auditory-visual displays. Finally, the remaining questions of the postexperiment survey reveal that 31 of the 36 subjects (86.1%) focused on alphanumerics to determine the quality of the visual displays, and that 20 of the 36 subjects (55.5%) felt that they were mentally overloaded when having to rate both auditory and visual displays simultaneously.

4.8 Conclusions

Overall, the findings suggest that—whether asked specifically to attend to both auditory and visual modalities or asked to attend to only one modality—similar and dissimilar cross-modal auditory-visual perception phenomena exist. These findings suggest that when manipulating visual display pixel resolution and auditory display sampling frequency:

- When attending only to the visual modality or attending to both auditory and visual modalities, a high-quality visual display coupled with a high-quality auditory display causes an increase in the perception of visual display quality relative to established baseline conditions derived from visual-only quality perception evaluations.
- When attending only to the auditory modality or

attending to both auditory and visual modalities, a low-quality auditory display coupled with a high-quality visual display causes a decrease in the perception of auditory display quality relative to established baseline conditions derived from auditory-only quality perception evaluations.

- When attending to both auditory and visual modalities, a high-quality auditory display coupled with a low-quality visual display causes an increase in the perception of auditory display quality relative to established baseline conditions derived from auditory-only quality perception evaluations.

However, would the same findings hold true when manipulating other quality parameters? The second experiment investigates whether manipulating visual display Gaussian white noise level and auditory display Gaussian white noise level produce the same results.

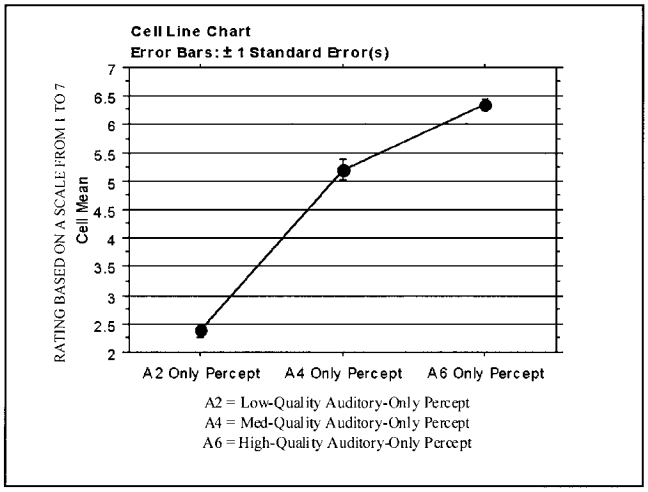


Figure 10. *Experiment 1: auditory-only quality percept ratings.*

5 Experiment 2: Static Noise

5.1 Introduction

Experiment 2: Static Noise investigates the perceptual effects from manipulating visual display Gaussian noise level and auditory display Gaussian noise level. The visual display consists of a static image of a radio, and the auditory display is a selection of music.

5.2 Location

All testing sessions of Static Noise are conducted in a similar isolated room under the same ambient conditions as outlined earlier in the first experiment (Static Resolution).

5.3 Participants

The subjects were 36 volunteer participants (27 male, 9 female) comprising students, faculty, staff, and guests of NPS. Based on the limited gender findings of the first experiment (Static Resolution), the number of male and female subjects in this experiment is not balanced. The average age of the subjects is 36.1 years, ranging in age from 19 to 54. As with the previous experiment, all subjects are required to have 20/20 or corrected-to-20/20 vision and normal hearing.

5.4 Apparatus

The apparatus used in this experiment is identical to that of the first experiment (Static Resolution).

5.5 Procedure

Except for a few changes that will be discussed, the procedure of this experiment is identical to that of the first experiment (Static Resolution). The experiment involves a 3×3 factorial within-subjects design. The two independent variables are visual and audio display quality, and the two dependent variables are the corresponding quality perception of the auditory and visual displays. The development process of the visual displays is identical to that of the first experiment, except that Gaussian white noise levels are manipulated with Adobe Photoshop as opposed to pixel resolution. The three levels of the visual quality independent variable consist of low-, medium-, and high-quality visual displays of the same radio image used in the first experiment, but having added Gaussian noise level amounts of 24, 18, and

12, respectively. The number corresponding to the amount of Gaussian noise is a relative number based on a scale of 1 to 999 that is used in Adobe Photoshop. Likewise, the development process of the auditory displays is identical to that of the first experiment, except that Gaussian noise levels of the same music selection used in the first experiment at 44.1 kHz are manipulated with Sonic Foundary's SoundForge as opposed to sampling frequency. The resulting three levels of the auditory quality independent variable consist of low-, medium-, and high-quality auditory displays of the same music selection presented monophonically at 44.1 kHz and having mixed-in Gaussian noise level amounts of 31%, 23%, and 15%, respectively. Thus, both the visual and auditory display parameters manipulated are Gaussian noise level.

The lowest- and highest-quality auditory displays in which the subjects are supposed to memorize during the self-calibration phase correspond to the music selection at 44.1 kHz, having mixed-in Gaussian noise level amounts of 45% and 10%, respectively. The lowest- and highest-quality visual displays in which the subjects are supposed to memorize during the self-calibration phase are depicted in figure 11 and figure 12, respectively (see p. 570). The low-quality visual display has an added Gaussian noise level amount of 45, whereas the high-quality visual display has an added Gaussian noise level amount of 10. Besides the different auditory and visual stimuli utilized, the procedure continues exactly as in the previous experiment except for minor changes in the readability of instructions, an increase in the number of visual-only and auditory-only quality ratings, and a decrease from eighteen to nine combined auditory-visual ratings during the final portion of the experiment. These changes are now discussed.

Based on the subjects' comments on the previous experiment, the readability of the instructions is enhanced by adding more white space. The content of the instructions is not changed.

To establish a stronger confidence in the baseline ratings for the visual-only and auditory-only displays, the number of quality ratings made during the visual-only and auditory-only portions is increased from nine to twelve. However, to conform with the data analyses of the previous experiment, the first three ratings, consisting of one low-, medium-, and high-quality, are disregarded. The idea is to allow the subject, unknowingly, to see/hear the three quality levels one time before having to make a rating. The baseline ratings are still based on an average of three quality ratings to conform with the data analyses of the first experiment. The only result is an increase in the confidence of the baseline ratings and not an increase of the number of stimuli used to average the baseline ratings.

The final portion of the experiment is also changed based on subjects' comments from the first experiment. Subjects felt that rating eighteen combined auditory-visual displays is somewhat long and tiresome. As a result, the number of combined auditory-visual display ratings during the remaining experiments is decreased from eighteen to nine in an effort to maintain a higher level of subject interest. Accordingly, data analyses from the first experiment consider only the first nine of the eighteen total combined auditory-visual display ratings.

Again, other than the above-mentioned changes, the procedure of this experiment is identical to that of the previous experiment. Therefore, the same data analyses are used to examine the results.

5.6 Experimental Validity

As in the first experiment, the most important consideration is whether the quality of the visual and auditory displays developed for this experiment are rank-ordered by the subjects according to their intended rankings. If this were not the case, the validity of the experiment would be jeopardized. However, in looking at figure 13, one can see that the overall quality ratings of the visual-only displays are properly rank-ordered by the subjects according to this experiment's intended low-, medium-, and high-quality rankings. Likewise, in looking at figure 14, one can see that the overall quality ratings of the auditory-only displays are properly rank-ordered by the subjects according to this experiment's intended low-, medium-, and high-quality rankings. Given that the data regarding quality of all displays are properly rank-ordered, data analysis with respect to the null hypotheses can continue.

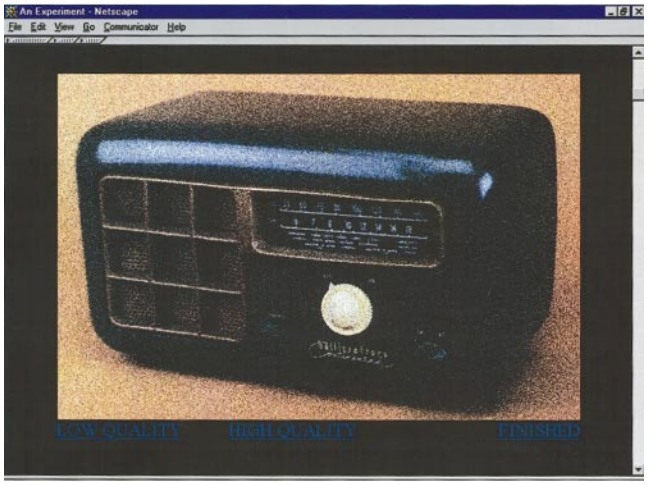


Figure 11. *Experiment 2: low-quality visual display familiarization.*

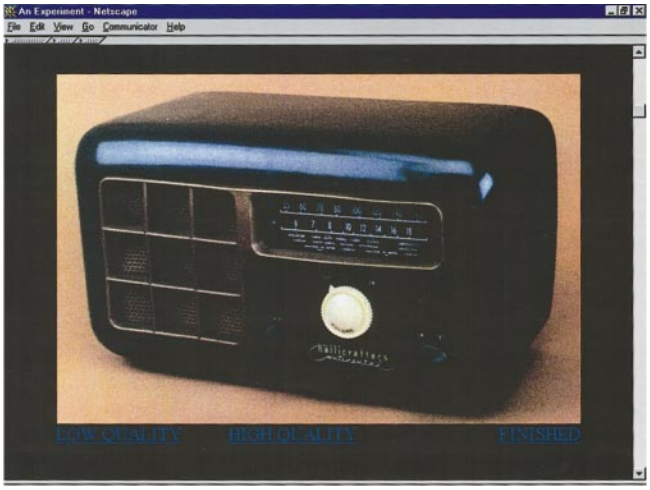


Figure 12. *Experiment 2: high-quality visual display familiarization.*

5.7 Findings

In terms of the first null hypothesis, none of the quality combinations have statistically significant findings. In terms of the second null hypothesis, when presented a combined low-quality auditory and high-quality visual display and asked only to rate the quality of the auditory display, a statistically significant finding at the 0.0290 level suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display. Furthermore, when presented a combined high-quality auditory and high-

Figure 15. *Fruit flower scene (an Adobe Photoshop 4.0 sample image).*

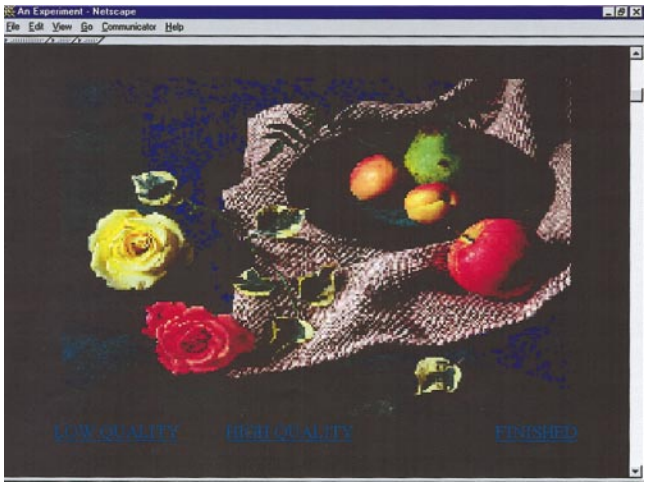


Figure 16. *Experiment 3: low-quality visual display familiarization.*

Figure 17. *Experiment 3: high-quality visual display familiarization.*

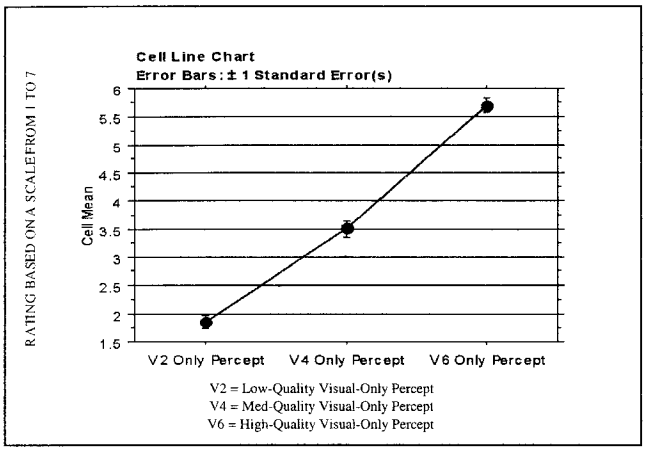


Figure 13. *Experiment 2: visual-only quality percept ratings.*

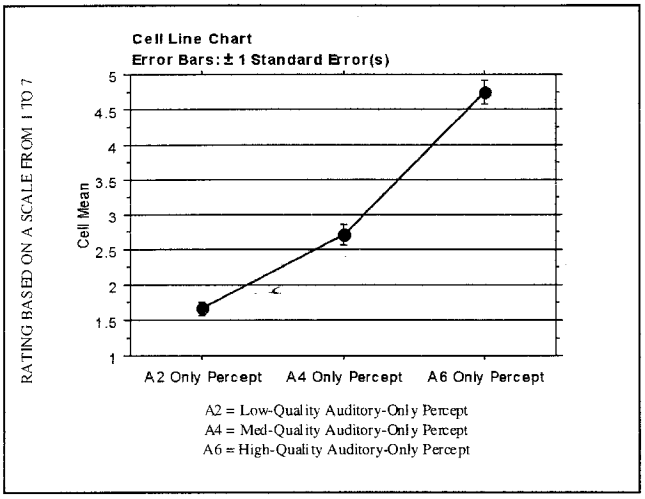


Figure 14. *Experiment 2: auditory-only quality percept ratings.*

quality visual display and asked only to rate the quality of the auditory display, a statistically significant finding at the 0.0243 level suggests that the quality perception of a high-quality auditory display is increased when coupled with a high-quality visual display.

In terms of the third null hypothesis, there are no significant findings at the 0.05 level. However, it is worth mentioning that, when presented a combined high-quality visual display coupled with a low-quality auditory display and asked to rate both auditory and visual displays, the results at the 0.10 level suggest that the quality perception of the high-quality visual display is increased.

In terms of the fourth null hypothesis, when presented a combined medium-quality auditory and medium-quality visual display and asked to rate both auditory and visual displays, a statistically significant finding at the 0.0029 level suggests that the quality perception of a medium-quality auditory display is increased when coupled with a medium-quality visual display. Furthermore, when presented a combined high-quality auditory and high-quality visual display and asked to rate both auditory and visual displays, a statistically significant finding at the 0.0294 level suggests that the quality perception of a high-quality auditory display is increased when coupled with a high-quality visual display.

In terms of the postexperiment questions, the results indicate that eight seconds is an adequate amount of time to rate the visual-only and auditory displays, but that slightly more than eight seconds is desired when rating the combined auditory-visual displays. Furthermore, 29 of the 36 subjects (80.1%) focused on alphanumerics to determine the quality of the visual displays, and only seven of the 36 subjects (19.4%) felt that they were mentally overloaded when having to rate both auditory and visual displays simultaneously.

5.8 Conclusions

Overall, the findings suggest that, whether asked to attend specifically to both auditory and visual modalities or asked to attend only to one modality, similar and dissimilar cross-modal auditory-visual perception phenomena exist. These findings suggest that, when manipulating both visual and auditory display Gaussian noise level:

- When attending only to the auditory modality, a low-quality auditory display coupled with a high-quality visual display causes a decrease in the perception of auditory quality relative to established baseline conditions derived from auditory-only quality perception evaluations.
- When attending only to the auditory modality or attending to both auditory and visual modalities, a high-quality auditory display coupled with a high-quality visual display causes an increase in the perception of

visual quality relative to established baseline conditions derived from visual-only quality perception evaluations.
- When attending to both auditory and visual modalities, a medium-quality auditory display coupled with a medium-quality visual display causes an increase in the perception of auditory quality relative to established baseline conditions derived from auditory-only quality perception evaluations.

Thus far, the first two experiments have used a somewhat perceptually tight coupling of radio and music to represent the visual and auditory displays. However, might the same findings hold true if the auditory and visual displays are not semantically associated with each other? The next section describes the final experiment of this research effort, which investigates the answer to this question.

6 Experiment 3: Static Resolution Nonalphanumeric

6.1 Introduction

Experiment 3: Static Resolution NonAlphanumeric is designed to investigate the perceptual effects from manipulating visual-display pixel resolution and auditory display sampling frequency. The visual display consists of a fruit-flower scene (an Adobe Photoshop 4.0 sample image), depicted in figure 15 (see p. 570), and the auditory display is a selection of music.

6.2 Location

The location and ambient conditions for this experiment are identical to that of the previous experiment (Static Noise).

6.3 Participants

The subjects were 36 volunteer participants (14 male, 22 female) comprising students, faculty, staff, and guests of NPS. Again, based on the limited gender findings of the first two experiments, the number of male and female subjects in this experiment is not balanced. The average age of the subjects is 35.5 years, ranging in age from 11 to 59. (Two female subjects did not give their age.) As with the previous experiment, all subjects are required to have 20/20 or corrected-to-20/20 vision and normal hearing.

6.4 Apparatus

The apparatus used in this experiment is identical to that of the first two experiments (Static Resolution and Static Noise).

6.5 Procedure

The procedure of this experiment is identical to that of the previous experiment (Static Noise). The three levels of the visual quality independent variable consist of low-, medium-, and high-quality visual displays of the fruit-flower scene depicted earlier, having resolutions of 34, 50, and 66 pixels/inch, respectively. Another key aspect for using the fruit-flower scene is that it has no alphanumerics (hence the name of this experiment). In the previous two experiments, 60 out of 72 subjects (83.3%) focused on alphanumerics when determining the quality of the visual displays. As a result, another goal of this experiment is to investigate whether a lack of alphanumeric features has any affect on the overall ability of the subjects to determine the quality of the visual displays. The three levels of the auditory quality independent variable consist of low-, medium-, and high-quality auditory displays of the same music selection presented monophonically, having sampling rates of 11 kHz, 19 kHz, and 35 kHz, respectively. The visual display parameters manipulated are pixel resolution, and the auditory display parameters manipulated are sampling frequency.

The lowest- and highest-quality auditory displays in which the subjects are supposed to memorize during the self-calibration phase correspond to the music selection at 8 kHz and 44.1 kHz, respectively. The lowest- and highest-quality visual displays in which the subjects are supposed to memorize during the self-calibration phase are depicted in figure 16 and figure 17, respectively (see p. 570). The low-quality visual display has a

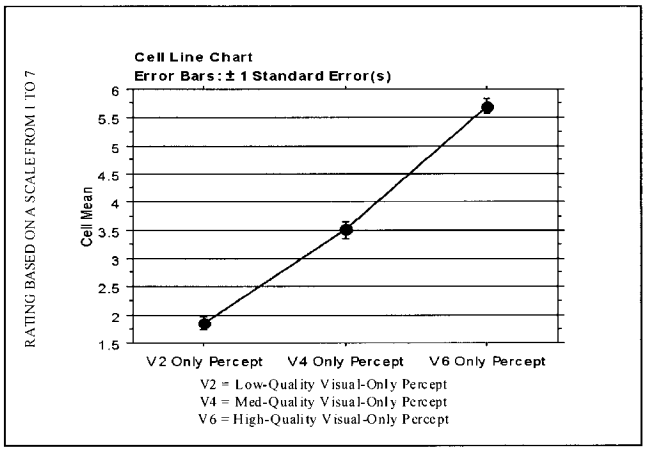


Figure 18. *Experiment 3: visual-only quality percept ratings.*

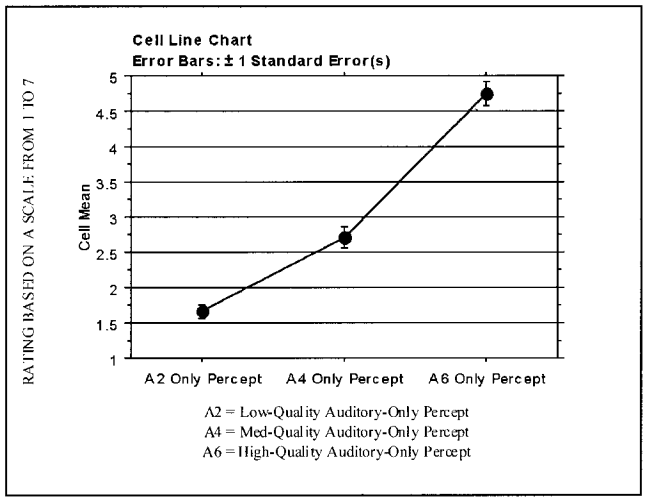


Figure 19. *Experiment 3: auditory-only quality percept ratings.*

resolution of 28 pixels/inch, whereas the high-quality visual display has a resolution of 72 pixels/inch. Besides the different auditory and visual stimuli utilized, the procedure continues exactly as in the second experiment. As a result, the same data analyses are used to examine the results.

6.6 Experimental Validity

As with the previous experiments, the most important consideration is whether the quality of the visual and auditory displays developed for this experiment are rank-ordered by the subjects according to their intended rankings. If this were not the case, the validity of the experiment would be jeopardized. However, in looking at figure 18, one can see that the overall quality ratings of the visual displays are properly rank-ordered by the subjects according to this experiment's intended low-, medium- and high-quality rankings. Thus, a lack of alphanumeric features had no effect on the overall ability of the subjects to determine the quality of the visual displays. Likewise, in looking at figure 19, one can see that the overall quality ratings of the auditory displays are properly rank-ordered by the subjects according to this experiment's intended low-, medium-, and high-quality rankings. Given that the data regarding quality of all displays are properly rank-ordered, data analysis with respect to the null hypotheses can continue.

6.7 Findings

In terms of the first null hypothesis,

- when presented a combined high-quality visual and medium-quality auditory display and asked only to rate the quality of the visual display, a statistically significant finding at the 0.0201 level suggests that the quality perception of a high-quality visual display is increased when coupled with a medium-quality auditory display, and
- when presented a combined high-quality visual and high-quality auditory display and asked only to rate the quality of the visual display, a statistically significant finding at the 0.0161 level suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

In terms of the second null hypothesis, there are no statistically significant findings in any of the quality combinations.

In terms of the third null hypothesis, when presented a combined high-quality visual and high-quality audi-

tory display and asked only to rate both auditory and visual displays, a statistically significant finding at the 0.0125 level suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

In terms of the fourth null hypothesis, the results suggest that, when presented a combined medium-quality auditory and low-quality visual display and asked to rate both auditory and visual displays, a statistically significant finding at the 0.0351 level suggests that the quality perception of a medium-quality auditory display is decreased when coupled with a low-quality visual display.

In terms of the postexperiment questions, the results indicate that determining the quality of both auditory and visual displays of a combined auditory-visual display proved to be more difficult than determining the quality of either auditory or visual display presented either alone or in combination. Furthermore, the results indicate that eight seconds is an adequate amount of time to rate the visual-only and auditory displays, but that slightly more than eight seconds is desired when rating the combined auditory-visual displays. Finally, only nine of the 36 subjects (25.0%) felt that they were mentally overloaded when having to rate both auditory and visual displays simultaneously.

6.8 Conclusions

Overall, the findings suggest that, whether asked specifically to attend to both auditory and visual modalities or asked to attend only to one modality, similar and dissimilar cross-modal auditory-visual perception phenomena exist. These findings suggest that, when manipulating visual display pixel resolution and auditory display sampling frequency and:

- when attending only to the visual modality, a high-quality visual display coupled with a medium-quality auditory display causes an increase in the perception of visual quality relative to established baseline conditions derived from visual-only quality perception evaluations;
- when attending only to the visual modality or attending to both auditory and visual modalities, a high-quality visual display coupled with a high-quality auditory display causes an increase in the perception of visual quality relative to established baseline conditions derived from visual-only quality perception evaluations; and
- when attending to both auditory and visual modalities, a medium-quality auditory display coupled with a low-quality visual display causes a decrease in the perception of auditory quality relative to established baseline conditions derived from auditory-only quality perception evaluations.

Therefore, even though the auditory and visual displays are less perceptually tightly coupled auditory-visual displays than in the first two experiments, the results indicate that the effects of auditory-visual cross-modal perception phenomena persist.

7 Overall Summary and Observations

Overall, these results provide the empirical evidence to support what most people in the gaming business, multimedia industry, entertainment industry, and VE community have suspected all along: that auditory displays can influence the quality perception of visual displays, and that visual displays can influence the quality perception of auditory displays. (For a more in-depth review of the findings presented in this study, see Storms (1998).) The results also indicate that, although we can divide our attention between audition and vision, we are not consciously aware of potentially significant intersensory effects.

7.1 Theoretical Impact

One of the overall findings of this research effort suggests that, when attending only to the auditory modality, a low-quality auditory display coupled with a high-quality visual display causes a decrease in the perception of auditory quality. The reason for degrading

the perception of the auditory quality might be based on the concept of visual dominance. Perhaps at some higher cognitive level, the higher-quality visual display is being compared with the lower-quality auditory display. This unconscious comparison might cause one to perceive that the auditory quality is worse than it actually is because of the dominating nature of the visual modality.

Overall, the results of this effort complement the findings of previously conducted intersensory research focusing on suprathreshold auditory-visual stimuli. In a study concerning auditory fidelity of high-definition television (HDTV), Neuman (1990) and Neuman, Crigler, and Bove (1991) found that subjects perceived an increase in visual quality when coupled with better audio. While investigating the effect of visual information on the impression of sound and the effect of auditory information on the impression of visual images when listening to music via audio-visual media, Iwamiya (1992) concluded that the factors of brightness, tightness, and cleanness of the auditory images enhanced the perception of brightness, tightness, and cleanness of the visual images. Hollier and Voelcker (1997) investigated the influence of video quality on audio perception and found that, when no video was present, the perceived audio quality was always worse than if video were present, and also that a decrease in video quality corresponded to a decrease in perceived audio quality. Woszczyk, Bech, and Hansen (1995), Bech, Hansen, and Woszczyk (1995), and Bech (1997) investigated the interaction between the auditory and visual modalities in the context of a home theater system. These researchers acknowledge that "experiments involving both modalities [audition and vision] require a novel approach that recognizes domains of cooperative interaction between the senses." They found that both visual and audio perceived quality increases with screen size, and that "[perceived] quality of spatial reproduction increases linearly with an increase in stereophonic width." Furthermore, Hugonnet (1997) found that, when people are first exposed to stereo sound when watching TV, most people find the relationship between visual and auditory images strange and not very comfortable. However, once people have become accustomed to stereo sound, if they are reexposed to mono sound, they perceive the mono sound to be of lower quality. Furthermore, these findings complement the research of Lipscomb (1990) and Lipscomb and Kendall (1994) that suggested that a musical soundtrack can in fact change the perceived meaning of an audio-visual film presentation. Likewise, in terms of filmmaking, Rydstrom (1994) explains that "when approached creatively, the combination of sound and image can bring something to vivid life, clarify the intent of the work, and make the whole experience more memorable."

The results of this study also provide new insights on previously conducted auditory-visual intersensory experiments focusing on threshold levels, absolute sensitivity, and just-noticeable-differences (JND) (Kravkov, 1936; Pratt, 1936; Serrat & Karwoski, 1936; Gilbert, 1941; Ryan, 1940; Gregg & Brogden, 1952; London, 1954; Thompson, Voss, & Brogden, 1958; Loveless, Brebner, & Hamilton, 1970). Exactly how this sensory interaction occurs is still not known. Schillinger (1948) could explain the correlation of visual and auditory information via mathematics. O'Connor and Hermelin (1981) would argue that the findings of this research effort support the concept of sensory capture. Marks (1974, 1978, 1982, 1987, 1989) and Marks, Szczesiul, and Ohlott (1986) might argue that the findings of sensory interaction provide more evidence of auditory-visual cross-modal matching. These findings also support Bregman's concept of auditory scene analysis (1990) in that "both senses must participate in making decisions of "how many," "where," and of "what."' Stein and Meredith (1993) might conclude that sensory interaction is taking place at the neurological level, based on single multimodal neurons. However, Gibson (1966, 1979) might argue that this sensory interaction is based on the complexity of natural life events. Cytowic (1989, 1995) and Baron-Cohen and Harrison (1996) might argue that auditory-visual cross-modal perception phenomena is related to synesthesia.

7.2 Commercial Impact

The findings in this study have diverse commercial impact. For example, one of the overall findings of this

effort suggests that, when attending only to the visual modality, a high-quality visual display coupled with a high-quality auditory display causes an increase in the overall visual quality perception of an auditory-visual display. Thus, suppose that the fictitious company, ACME Cyber Art, sells contemporary paintings via the Internet. ACME Cyber Art's current Web-based advertising depicts only photographs of the various paintings that prospective customers can purchase online. ACME Cyber Art, however, wants to increase its sales. One possible strategy to increase sales is to add high-quality music to their Web page while prospective customers are looking at the various artworks. By adding music, the perceptual visual quality of the various artworks might increase relative to itself, thereby possibly increasing the probability that the customer makes a purchase.

Another finding of this research suggests that when, attending only to the auditory modality, a low-quality auditory display coupled with a high-quality visual display causes a decrease in the overall auditory quality perception of an auditory-visual display. Thus, suppose the next GRAMMY Awards are partially decided via Internet-based votes. To cast their votes, music fans would point their Web browser to the GRAMMY Awards Web site, which contains high-quality visual images of the various nominated musicians and signers. By clicking on the image of a particular person or musical group, one could hear a short, eight-second audio clip of the nominated song. In an effort to decrease rendering time, storage requirements, and download time, suppose the designers of the GRAMMY Web site decreased the sampling frequency of the audio clips from 44.1 kHz to 10 kHz. As a result, to the surprise of the site designers, most fans complained that the quality of the audio clips was very poor, making it impossible to cast their votes properly. Consequently, the Internet-based voting of the GRAMMY Awards might be a huge failure.

Another finding of this research effort suggests that, when attending to both auditory and visual modalities, a high-quality visual display coupled with a high-quality auditory display causes an increase in the overall visual quality perception of an auditory-visual display. Thus, suppose a VE developer has been tasked to increase the realism (and perhaps presence) of a 3-D scene depicting a typical family living room. The current virtual living room contains a TV and stereo system that is rendered using high-quality visual graphics. However, the living room scene does not have any associated sounds. Instead of increasing the pixel resolution of the living room scene and causing an unwanted increase in the visual rendering time of the scene, the VE developer adds high-quality music to the stereo system, and an MPEG video sequence containing high-quality audio to the TV display. As a result, the perceptual visual quality of the scene ought to increase by simply adding the associated auditory displays without the need to manipulate any of the visual displays.

These preceding examples highlight just some of the numerous possibilities of this research effort. Overall, the findings are indeed important in ways that can greatly benefit the gaming business, multimedia industry, entertainment industry, VE community, and also the Internet industry.

7.3 Observations

The following section describes some of the overall informal observations noted during the conduct of the main experiments. No formal data analyses are performed on the observations, which are merely presented to provide the reader with additional peripheral insights on the overall findings of this research effort.

7.3.1 Mouse. Although response time was measured, it was not analyzed. Nevertheless, the functionality of the mouse and mouse pad also has an undetermined effect on response time. Some subjects complained that the mouse would occasionally stick or slide improperly, while others did not report any problems. Some subjects would keep their hands on the mouse the entire time, and others would place their hands in their laps and then grab the mouse when it was time to make a response. On a side note, some subjects used the mouse/cursor to read all the instructions and also to point at salient quality features. Some subjects would also slide their cursor to the relative quality posi-

tion of the rating scale even before the scale appeared. Furthermore, adept computer users are much more efficient at using the mouse as opposed to someone using the mouse's point-and-click paradigm for the first time. Some subjects who are accustomed trackball users felt uncomfortable using the mouse.

7.3.2 Subjects' Description and Use of the Stimuli. Perhaps the most-interesting observations are gathered from the postexperiment questions which asked the subjects if they focused on any particular features when determining quality and, if so, to describe those features. The diverse responses are amazing, and the diversity stems from the various backgrounds of the subjects. For example, in describing a straight line on the radio, a computer graphics programmer might use the term *aliasing*, whereas the novice might use the term *jaggedness*. Also, some subjects felt that it was easier to determine the auditory and visual qualities simultaneously because they could use the stimulus in one modality to support their quality decision in the other modality.

7.3.3 Reversals. A very common response from the subjects is that they sometimes felt that they might have reversed the rating of auditory and visual qualities. This auditory-visual dyslexia may be attributed to some of the overall findings of this research effort.

7.3.4 Recognizable Quality Levels. Upon completion of the experiment, some subjects were astonished when they were told that only three levels of auditory and visual stimuli are utilized. Their astonishment is probably attributed to the number of choices on the rating scales (seven). Thus, subjects may have been anticipating seven quality levels and, as a result, conformed (perceptually) to the seven choices on the rating scales. For future experiments, the use of visual-analogue scales (which permit essentially an indefinite number of responses on a line scale with defined endpoints) might prove more useful.

8 Future Work

8.1 Choice of Quality Parameters and Stimuli

Because pixel resolution, Gaussian noise level, and sampling frequency are the only quality parameters that were manipulated, the use of other quality metrics is warranted. Furthermore, the effects from using various other stimuli, such as motion video and 3-D VEs are also needed. A greater scope of potential auditory-visual perception phenomena can thereby be investigated.

One possible scenario using a VE might first include the process of having subjects watch a virtual person (in 3-D space) place a radio (playing music) on a table. After this initial process of watching the virtual radio being placed (dynamically) on the virtual table, subjects might perceive a stronger perceptual grouping between the radio (visual) and music (audio) through increased temporal and spatial synchronization, thereby decreasing the cognitive distance between the radio (visual) and music (audio). As a result, if the same experiments outlined in this study are conducted after this initial process, the overall findings might indicate an increase in statistically significant auditory-visual cross-modal perception phenomena.

8.2 Auditory-Visual Quantitative Perceptual Model

Given that auditory-visual cross-modal perception phenomena exist, the next logical step is to incorporate these overall findings into some type of useful auditory-visual quantitative perceptual model similar to that proposed by Hollier and Voelcker (1997), as depicted in figure 20. This model can then be used to derive appropriate (quantitative) levels of auditory and visual fidelity for use by developers in the gaming business, multimedia industry, entertainment industry, VE community, and the Internet industry. For example, given a certain application, this auditory-visual quantitative perceptual model could help to derive the appropriate levels and specific amounts of visual display pixel resolution and auditory display sampling frequency as a function of visual-only, auditory-only, and/or combined auditory-visual media.

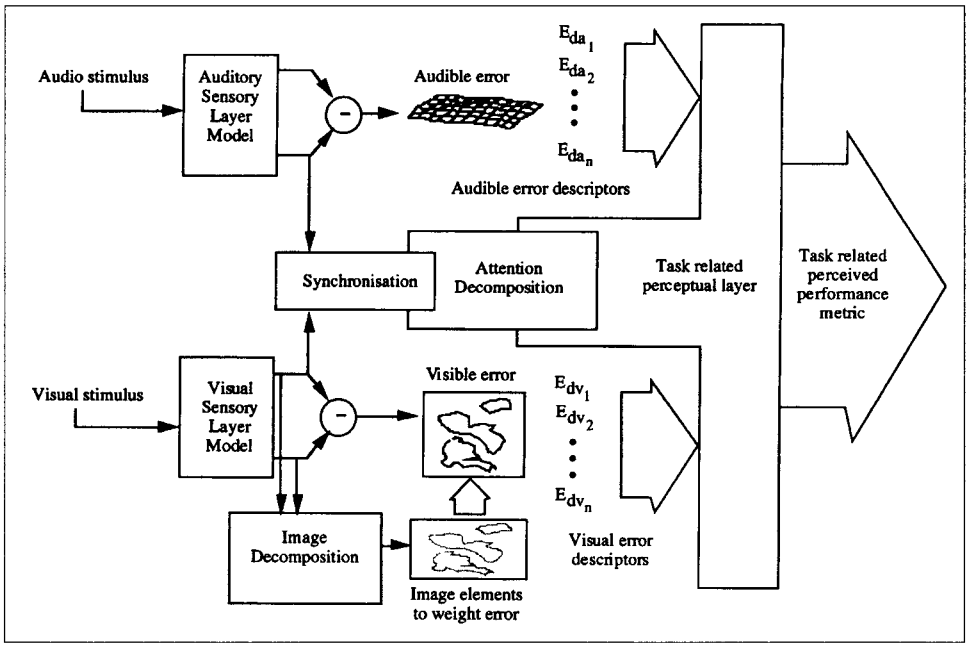


Figure 20. *Auditory-visual perceptual model from Hollier and Hoelcker (1997).*

8.3 Intersensory Research

The results of this research effort make it clear that, to better understand the proper use of multisensory stimuli, more research emphasis needs to be placed on investigating intersensory phenomena. This increased emphasis need not be limited to auditory-visual interactions, but ought to include the investigation of auditory-visual-haptic interactions.

8.4 Online Experiments

Because of the potential to easily acquire many subjects (perhaps thousands), the use of online experiments can greatly facilitate scientific research. In light of this, all the experiments contained in this research can be used online. However, online experiments make it difficult to control the conditions of the experiment (hardware specifications, proper subject participation, environmental conditions, and the like). The ability to control conditions is vital when conducting experiments. Nevertheless, by embedding all experiments in this study within an HTML browser, an attempt has been made towards conducting future online auditory-visual experimental research.

Acknowledgments

This research effort is funded by a large number of U.S. government agencies, including U.S. Army Research Laboratory (ARL); U.S. Army Simulation, Training, and Instrumentation Command (STRICOM); Defense Advanced Research Projects Agency (DARPA); Defense Modeling and Simulation Office (DMSO); and the Office of Naval Research (ONR).

References

Aldridge, R., Davidoff, J., Ghanbari, M., Hands, D., & Pearson, D. (1995). Measurement of scene-dependent quality variations in digitally coded television pictures. *IEE Proc.-Vis. Image Signal Process*, *142*(3), 149–154.

Barfield, W., Hendrix, C., Bjorneseth, O., Kaczmarek, K. A., & Lotens, W. (1995). Comparison of human sensory capa-

bilities with technical specifications of virtual environment equipment. *Presence: Teleoperators and Virtual Environments*, *4*(4), 329–356.

Baron-Cohen, S., & Harrison, J. E. (Eds.). (1996). *Synaesthesia: Classic and Contemporary Readings*. Oxford: Blackwell Publishers.

Bech, S., Hansen, V., & Woszczyk, W. (1995, October 6–9). *Interaction between audio-visual factors in a home theater system: experimental results* (Preprint No. 4096). Paper presented at The 99th Audio Engineering Society Convention, New York.

Bech, S. (1997, March 22–25). *The influence of stereophonic width on the perceived quality of an audio-visual presentation using a multichannel sound system* (Preprint No. 4432). Paper presented at The 102nd Audio Engineering Society Convention, Munich.

Bermant, R. I., & Welch, R. B. (1976). Effect of degree of separation of visual-auditory and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills*, *43*, 487–493.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Cytowic, R. E. (1989). *Synesthesia: A Union of the Senses*. New York: Springer-Verlag.

———. (1995). Synesthesia: Phenomenology and neuropsychology: A review of current knowledge. *PSYCH*, *2*(10), http://psyche.cs.monash.edu.au/v2/psyche-2-10-cytowic.html.

Dachis, C. (1995). *Radios by Hallicrafters with Price Guide*. Atglen, PA: Schiffer Publishing, Ltd.

Flanagan, D. (1996). *Java in a Nutshell: A Desktop Quick Reference for Java Programmers*. Sebastopol, CA: O'Reilly & Associates, Inc.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, *25*, 350–358.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

———. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gilbert, G. M. (1941). Inter-sensory facilitation and inhibition. *Journal of General Psychology*, *24*, 381–407.

Gregg, L. W., & Brogden, W. J. (1952). The effect of simultaneous visual stimulation on absolute auditory sensitivity. *Journal of Experimental Psychology*, *43*, 179–186.

Hollier, M. P., & Voelcker, R. (1997). Objective Performance Assessment: Video Quality as an Influence on Audio Perception. Preprint No. 4590. Presented at the 103rd Audio Engineering Society Convention. New York, New York, September 26–29.

Howard, I. P., & Templeton, W. B. (1996). *Human Spatial Orientation*. New York: Wiley.

Hugonnet, C. (1997, September 26–29). *A new concept of spatial coherence between sound and picture in stereophonic (and surrounding sound) TV production* (Preprint No. 4539). Paper presented at The 103rd Audio Engineering Society Convention, New York.

Iwamiya, S. (1992). The interaction between auditory and visual processing when listening to music via audio-visual media. *JASJ*, *48*(3), 146–153.

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt, Brace, and World.

Kohler, W. (1940). *Dynamics in Psychology*. New York: Liveright.

Kravkov, S. V. (1936). The influence of sound upon the light and color sensibility of the eye. *Acta Opthalmologica Scandinavica*, *14*, 348–360.

Ladd, E., & O'Donnell, J., et al. (1998). *Using HTML 4.0, Java 1.1, and JavaScript 1.2. 2nd Edition*, Que Corporation.

Lipscomb, S. D. (1990). *Perceptual judgment of the symbiosis between musical and visual components in film*. Unpublished master's thesis, University of California, Los Angeles, CA.

Lipscomb, S. D., & Kendall, R. A. (1994). Perceptual judgment of the relationship between musical and visual components in film. *Psychomusicology*, *13*(Spring/Fall), 60–98.

London, I. D. (1954). Research on sensory interaction in the Soviet Union. *Psychological Bulletin*, *51*(6), 531–568.

Loveless, N. E., Brebner, J., & Hamilton, P. (1970). Bisensory presentation of information. *Psychological Bulletin*, *73*(3), 161–199.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *American Journal of Psychology*, *87*(1-2), 173–188.

———. (1978). *The Unity of the Senses: Interrelations among the Modalities*. New York: Academic Press.

———. (1982). Bright sneezes and dark coughs, loud sunlight and soft moonlight. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 177–193.

———. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394.

——— (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Ex-*

perimental Psychology: Human Perception and Performance, *15*(3), 586–602.

Marks, L. E., Szczesiul, R. & Ohlott, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 517–534.

Murch, G. M. (1973). *Visual and Auditory Perception*. Indianapolis, IN: Bobbs-Merrill Company, Inc.

Neuman, W. R. (1990). *Beyond HDTV: Exploring subjective responses to very high definition television*. MIT Media Library. Cambridge, MA: Massachusetts Institute of Technology.

Neuman, W., Crigler, A., & Bove, V. M. (1991). Television sound and viewer perceptions. *Proceedings of the Audio Engineering Society 9th International Conference*, *1*(2), 101–104.

O'Connor, N., & Hermelin, B. (1981). Coding strategies of normal and handicapped children. In Walk, R. D., & Pick, H. L. Jr. (Eds.), *Intersensory Perception and Sensory Integration* (pp. 315–343). New York: Plenum Press.

Peterson, L. R., & Peterson, M. J. (1959). Short-term memory retention of individual items. *Journal of Experimental Psychology*, *58*, 193–198.

Pick, H. L. Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception and Psychophysics*, *6*, 203–205.

Pratt, C. C. (1936). Interaction across modalities: I. Successive stimulation. *The Journal of Psychology*, *2*, 287–294.

Radeau, M., & Bertelson, P. (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics*, *20*(4), 227–235.

Ragot, R., Cave, C., & Fano, M. (1988). Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation. *Bulletin of the Psychonomic Society*, *26*(4), 350–352.

Ryan, T. A. (1940). Interactions of the sensory systems in perception. *Psychological Bulletin*, *37*, 659–698.

Rydstrom, G. (1994). Film sound: How it's done in the real world. Course Number 12: Sound Synchronization and Synthesis for Computer Animation and VR. Paper presented at SIGGRAPH '94, Orlando, Florida.

Schillinger, J. (1948). *The Mathematical Basis of the Arts*. New York: Philosophical Library.

Serrat, W. D., & Karwoski, T. (1936). An Investigation of the effect of auditory stimulation on visual sensitivity. *Journal of Experimental Psychology*, *19*, 604–611.

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge, MA: The MIT Press.

Storms, R. L. (1998). Auditory-visual cross-modal perception phenomena. Unpublished doctoral dissertation, Naval Postgraduate School, Monterey, California.

Thompson, R. F., Voss, J. F., & Brogden, W. J. (1958). Effect of brightness of simultaneous visual stimulation on absolute auditory sensitivity. J*ournal of Experimental Psychology*, *55*(1), 45–50.

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory 'compellingness' in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*(6), 557–564.

Wertheimer, M. (1912). Experimentelle studien über das sehen von bewegungen. *Zeitschrift für Psychologie*, *61*, 161–265.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: Harper Collins Publishers, Inc.

Woszczyk, W., Bech, S., & Hansen, V. (1995, October 6–9). *Interaction between audio-visual factors in a home theater system: definition of subjective attributes* (Preprint No. 4133). Paper presented at The 99th Audio Engineering Society Convention, New York.