NAVAL POSTGRADUATE SCHOOL
Monterey, California

DISSERTATION

AUDITORY-VISUAL CROSS-MODAL PERCEPTION PHENOMENA

by

Russell L. Storms

September 1998

Dissertation Supervisor: Michael J. Zyda

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1998	Doctoral Dissertation

4. TITLE AND SUBTITLE
AUDITORY-VISUAL CROSS-MODAL PERCEPTION PHENOMENA (U)

5. FUNDING NUMBERS

6. AUTHOR(S)
Storms, Russell L.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/ MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/ MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the United States Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The quality of realism in virtual environments is typically considered to be a function of visual and audio fidelity mutually exclusive of each other. However, the virtual environment participant, being human, is multi-modal by nature. Therefore, in order to more accurately validate the levels of auditory and visual fidelity 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 main experiments which were completely automated using HTML, Java, and JavaScript computer programming languages. Visual and auditory display quality perception were measured intramodally and intermodally by manipulating visual display pixel resolution and Gaussian white noise level and by manipulating auditory display sampling frequency and Gaussian white noise level.

Statistically significant results indicate that 1) medium or 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 2) 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 virtual environments must be a function of both auditory and visual display fidelities inclusive of each other.

14. SUBJECT TERMS
Virtual Environment, Auditory Display, Visual Display, Perception, Cross-Modal, Fidelity, Experimental Design

15. NUMBER OF PAGES
275

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

AUDITORY-VISUAL CROSS-MODAL PERCEPTION PHENOMENA

Russell L. Storms
Major, United States Army
B.S., United States Military Academy, 1986
M.S., Naval Postgraduate School, 1995

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL

September 1998

ABSTRACT

The quality of realism in virtual environments is typically considered to be a function of visual and audio fidelity mutually exclusive of each other. However, the virtual environment participant, being human, is multi-modal by nature. Therefore, in order to more accurately validate the levels of auditory and visual fidelity 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 main experiments which were completely automated using HTML, Java, and JavaScript computer programming languages. Visual and auditory display quality perception were measured intramodally and intermodally by manipulating visual display pixel resolution and Gaussian white noise level and by manipulating auditory display sampling frequency and Gaussian white noise level.

Statistically significant results indicate that 1) medium or 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 2) 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 virtual environments must be a function of both auditory and visual display fidelities inclusive of each other.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost I want to thank my dissertation committee. Professor McGhee always made sure I was doing things as expected from a Ph.D. Student. Professor Ziomek enlightened me about room acoustics for which I am eternally grateful. Don Brutzman always ensured that I was giving VRML its best chance. Rudy Darken, just three doors down the hall, was always available for me to bounce off my numerous and various ideas. Beth Wenzel's work has always inspired me and she also helped me to understand what experimental design was all about. Durand Begault, along with his terrific humor, helped me in the restructuring of my experimental design which proved to be most influential in my dissertation. Professor Mike Zyda, my dissertation supervisor, and just one door down the hall, was always available for me. Thanks Professor Zyda for your trust and support in my research endeavors. I will always be indebted to your kindness for me and my family and for your appreciation for my independent ways. Also, thanks to your wife Tyerin for her gracious hospitality.

I want to thank Elektra Records for allowing me to use musical portions of their CD for my research. I also want to thank Mr. Chuck Dachis for allowing me to use his photographs of radios in my research. Thanks to the entire NPSNET Research Group and CS Staff for all your support including: John Locke, Jimmy Liberato, Bill Cockayne, John Falby, Kent Watsen, Don McGregor, Ted Lewis, Rosalie Johnson, Mike Williams, Freddy Zyda, Jean Brennan, Shirley Oliveira, Rob Cortilla, and Walt Lundaker (who is sorely missed). Thanks to all the faculty and to my fellow Ph.D. students past and present including: Mike Holden (Dr. Chop, CS41), Eric Bachmann (CS41), Mickey Harn, Gary Stone, and Chris Eagle. Thanks to Sandra Day for her meticulous review of my dissertation. Thanks to David Pratt for letting me share his office (SP-250). A very special thanks goes out to CAPT Frank Petho, NSA Chair, for letting me use an office to conduct my experiments while Spangel Hall's electrical work was being done. Thanks to

Professor Hamming for our numerous talks. You are terribly missed. Also, thanks for all the subjects who volunteered their time to participate in my experiments.

Last, I want to thank my family. To my daughter, Janell, it is interesting that during your entire life, thus far, I have been working towards my Ph.D. You have provided me with lots of pleasurable (and sometimes not so pleasurable) distractions during the last three years. Perhaps one day, you too will be getting your Ph.D. Finally, none of this would have been possible if it were not for the love of my life, my wife Deanna. This whole Ph.D. process has perhaps been harder on her than it has been for me. Thanks for sticking it out with me. I love you. So now, lets get on with our lives and see how things are at Cana of Galilee.

DEDICATION

This Dissertation is Dedicated to the Memory of Mrs. Sherman and to the Celebration of Her Life.

I still cannot believe that Mrs. Sherman is gone. I say Mrs. Sherman and not Doris, for she will always be Mrs. Sherman to me. With the loss of Mrs. Sherman, the world has become a lesser place. I have always loved and respected Mrs. Sherman. She was always so kind and warm to me, even if Mark and I were up to no good, which I am sure happened on a few too many occasions. It saddens me greatly that I will not be able to mingle within her graciousness again, nor to be surrounded by her constant good nature towards everything. Mrs. Sherman will be surely missed, but it has been my pleasure and unbelievably great fortune to have shared some of her time on this earth and to taste her great buttermilk pancakes, which are still the best I have ever eaten.

I. INTRODUCTION

A. MOTIVATION

The fidelity requirements for virtual environments 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 ten years. 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. The quality of realism in these virtual environments is typically considered to be a function of visual and audio fidelity mutually exclusive of each other [BARF95]. Herein lies a problem: the virtual environment participant, being human, is multi-modal by nature. Thus, the quality of realism in virtual environments needs to be based on multi-modal criteria comprising all of our senses, as opposed to the current use of singular modality criteria. As such, the fidelity requirement of virtual environments must be based on multi-modal criteria comprising all of our senses. However, insufficient experimental data exists to make informed multi-modal design decisions.

B. OBJECTIVE

Because of current limitations in today's computer technology, it is impossible to render realistic information to all of our senses in real-time to the interactive virtual environment participant. However, since there have been significant advances in visual and audio display technology, it is appropriate to concentrate on the vision and audition sensory modalities. As such, the objective of this research effort correspondingly focuses on the two sensory modalities of vision and audition. In particular, the objective of this effort is to gain a better understanding of the intersensory or cross-modal effects between the auditory and visual sense modalities. 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 required for the immersed virtual environment participant.

C. SCOPE

Intersensory phenomena have been studied for many years by researchers in numerous disciplines such as: Psychoacoustics, Psychology, Physiology, Neurology, Philosophy, Musicology, Ecology, and Computer-Human Interaction, and by different organizations such as: Human Factors, Audio Engineering Society, Acoustical Society of America, Department of Defense, Artistic Community, and also the Film and Entertainment Industry. Thus, there is a large amount of intersensory research, but this knowledge is often kept within the discipline from which it was derived. Consequently, there is little cross-disciplinary transfer of intersensory knowledge. This lack of cross-disciplinary knowledge exists not only with intersensory research, but also seems to extend to many areas of academic and commercial interests. This is a pity, for there are no doubt countless examples of redundant research efforts all because of a lack of cross-disciplinary knowledge exchange. Nevertheless, in terms of modeling and simulation, the National Research Council (NRC) has recently investigated the possible collaboration opportunities between the Department of Defense and the Entertainment Industry [ZYDA97]. This collaboration is a much needed first step towards better cross-disciplinary knowledge transfer.

Computer Science in particular is severely lacking in its knowledge and use of intersensory phenomena. Therefore, it is important to note that the scope of this effort is filtered through the perspective of a computer scientist for use by other computer scientists. The results of this effort are intended to aid the computer scientist in developing better virtual worlds through appropriate use of auditory and visual display fidelities based on auditory-visual cross-modal perception phenomena. It is also important to note that the scope of this effort is not to identify absolute visual and/or audio fidelity requirements such as pixel resolution and sampling frequency respectively, but rather to identify the effects of auditory-visual cross-modal perception phenomena which can be used to justify a certain level of audio and/or visual fidelity.

D. APPROACH

The approach taken is that of the experimental psychologist. A series of experiments were designed to identify if there exists any pertinent auditory-visual cross-modal perception interactions. Specifically, one pilot study and three main experiments were conducted. Each of the three main experiments was completely automated using Hyper Text Markup Language (HTML), Java, and JavaScript [FLAN96] [LADD98]. The pilot study was also completely automated but was developed using Virtual Reality Modeling Language (VRML) [HART96] [LEAR96] [ROEH97]. All experiments were conducted at the Naval Postgraduate School (NPS) in Monterey, California. A total of 130 volunteer participants comprised from the students, faculty, staff, and guests of NPS served as subjects. Each experiment involved a 3x3 factorial within subjects design. (See [GOOD95] for a description of factorial design experiments.) The two independent variables were visual and audio display quality having three levels each consisting of low, medium, and high qualities. The visual display parameters that were manipulated were pixel resolution and Gaussian white noise level. The audio display parameters that were manipulated were sampling frequency and Gaussian white noise level. Partial counterbalancing was achieved through the technique of balanced Latin squares. (See [GOOD95] for a description of the Latin squares technique.) The basic idea of the experiments was to manipulate visual and auditory display parameters intra-modally and inter-modally and to likewise measure visual and auditory display perception intra-modally and inter-modally. During the experiments, which each lasted approximately 30 minutes, a single subject wore headphones and sat in front of a 20-inch display monitor. The task of the subject was to rate the perceived quality of audio-only, visual-only, and audio-visual displays through Likert rating scales ranging from 1 to 7. (See [GOOD95] for a description of Likert rating scales.) Thus, the dependent variables are the perception of visual display quality and the perception of auditory display quality. It is hoped that by carefully 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 audio-visual display, what affect (if any) do various audio quality levels have on the perception of visual quality and vice versa? The following are some examples:

1) Are changes in the audio and/or visual qualities of an audio-visual display perceivable and can these changes be attended to also?

2) 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?

3) 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) Does a low-quality auditory display coupled with a low-quality visual display cause a decrease/increase 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?

5) Does a high-quality auditory display coupled with a high-quality visual display cause an increase/decrease 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?

E. LIMITATIONS

Another facet of this effort was to confine all software development to the ever-evolving internet technology. The reasons for this are as follows:

1) To easily obtain software. All the software used to execute the experiments in this effort were simply downloaded. This downloaded software included: Netscape 2.0, 3.0, and 4.0 [NETS98]; Sun's Java Development Kit (JDK) 1.0, 1.1.2, 1.1.4, and 1.1.5 [SUNM98]; Silicon Graphics Inc. (SGI) CosmoPlayer VRML 2.0 beta Netscape Plugin

and VRML 2.0 Release Netscape Plugin [COSM98]; Sony's Community Place VRML 2.0 Browser [SONY98b], and Intervista's WorldView 2.0 Browser [INTE98].

2) To reduce cost. All downloaded software was free!

3) To verify the feasibility of conducting scientific experiments with HTML/Java/ JavaScript/VRML.

4) To support seamless portability and repeatability of research. The experiments outlined in this dissertation are currently being set up to be repeated at the College of Computing at Georgia Institute of Technology in Atlanta, Georgia.

5) To eventually conduct on-line auditory-visual cross-modal experiments which potentially have thousands (if not millions) of subjects/trials.

Another chosen limitation was that of hardware. To complement the ease of access and portability of all software, all the hardware used in this effort is available as commercial off-the-shelf (COTS) products. As such, no specific, hard to get, or intractably expensive piece of hardware is needed for this research effort.

F. DISSERTATION ORGANIZATION

This dissertation is organized around ten chapters, including a list of references, a bibliography, and four appendices. Chapter II discusses relevant background material including: Perception, The Senses, Audition, Vision, Attention, Gestalt Theory, Synesthesia, and Multimedia. Chapter III presents a thorough literature review covering: Virtual Environments (VE), Auditory-Visual Perceptual Organization, Auditory-Visual Art Forms and Film, Auditory-Visual Cross-Modal Matching, Visual Dominance Over Audition, Auditory-Visual Threshold Perception, and Auditory-Visual Suprathreshold Perception. Chapter IV discusses the issues relevant to the overall development of the experimental design process including: Motivation, Design Considerations, Design Selections, and Software Design. Chapter V discusses Visual Display Development, Auditory Display Development, and Auditory-Visual Display Development. Chapter VI gives a complete description of the experimental design of the initial pilot study to include: Location, Participants, Apparatus, Procedure, Results and Discussion, and

Summary and Conclusions. Chapter VII gives a complete description of the experimental design involving visual display pixel resolution manipulation of a static radio image, as well as auditory display sampling frequency manipulation of a section of music including: Location, Participants, Apparatus, Procedure, Changes from Pilot Study, Data Collection and Analysis, Results and Discussion, and Summary and Conclusions. Chapter VIII gives a complete description of the experimental design involving visual display Gaussian white noise level manipulation of a static radio image, as well as auditory display Gaussian white noise level manipulation of a section of music including: Location, Participants, Apparatus, Procedure, Results and Discussion, and Summary and Conclusions. Chapter IX gives a complete description of the experimental design involving visual display pixel resolution manipulation of a fruit-flower scene, as well as auditory display sampling frequency manipulation of a section of music including: Location, Participants, Apparatus, Procedure, Results and Discussion, and Summary and Conclusions. Chapter X presents the overall findings of this dissertation to include: Overall Results, Conclusions, Impact, Observations, Recommendations, Future Work, and Final Thoughts.

II. BACKGROUND

A. INTRODUCTION

The intent of this chapter is to give the computer scientist a high-level overview of some of the basic background knowledge which is required in order to understand this multi-disciplinary research effort. As such, the information outlined in this chapter is by no means comprehensive. Furthermore, the concepts outlined in this chapter lay the foundation for understanding the scope of this research effort. Because of the wide variety of topics covered including Perception, The Senses, Audition, Vision, Attention Theory, Gestalt Theory, Synesthesia, and Multimedia, the reader will hopefully gain a better appreciation for the interdisciplinary nature and breadth of knowledge required when conducting intersensory research.

B. PERCEPTION

1. Definition

First and foremost it is important to remember that "We can only obtain a rather one-sided idea of the development of perception if we neglect the interrelations of the different senses in creating our perceptual world" [SCHL35]. With this in mind a formal definition of perception from a psychological point of view is as follows:

> The psychology of perception, then, involves the study of the way an observer relates to his environment -- the way in which information is gathered and interpreted by an observer. This relationship is the result of a continuing process of learning, judging, interpreting, and reacting to the environment which begins at birth and continues throughout the life span of the individual. [MURC73]

From a physiological perspective, the following describes the nature of a stimulus:

> An excitation originating in any of the receptors does not remain strictly localized, but irradiates to some extent throughout the entire nervous system, thus affecting the excitatory states of all other mechanisms and consequently the sensory responses for which such excitatory states are important predisposing factors. [GILB41]

2. Stimulus

A stimulus is defined as "...any chemical or physical activator which causes a response in a receptor" [FOST68]. In total, there are only six classes of stimuli: (1) *mechanical*, (2) *thermal*, (3) *photic*, (4) *acoustic*, (5) *chemical*, and (6) *electrical*. Furthermore, an effective stimulus is one that produces a sensation, the dimensions of which are: quality, intensity, extension, duration, and like and dislike [FOST68].

Murch explains that the term *stimulus* is but half of a pair of correlated terms, the other half being *response*. As such, if we conform strictly to this correlated definition of stimulus, a circular definition enfolds. "This concept of stimulus would force us to regard the response as dependent on the object or event (stimulus) and the stimulus as dependent on the response" [MURC73]. Herman von Helmholtz tried to avoid this circular definition by introducing the concepts of *distal stimulus* (the external object or event) and *proximal stimulus* (the sensory representation of the stimulus by the nervous system) [HELM66]. However, Helmholtz's concepts of distal and proximal stimulus fall short because the circularity problem remains. "The distal stimulus gives rise to the proximal stimulus which in turn contributes to the building of a percept representative of the initial distal stimulus" [MURC73]. The distinction between distal and proximal stimuli are better explained by using the terms: *potential stimulus* and *effective stimulus* [GIBS66] [GIBS67].

> Any object or event in the environment is a *potential* stimulus. When such a potential stimulus stands in a constant relationship with a given response, it is an *effective* stimulus. Thus we are able to describe the environment independently of the responses of an observer. This is particularly important when we consider that one is often unaware of all the responses elicited by a stimulus. [MURC73]

The inherent linkage between sensation and perception can best be summed up as follows: "To sense is to respond, to perceive is to know" [MURC73].

But what happens when we are exposed to multiple stimuli? When two or more stimuli occur at the same time and/or space some very interesting perceptual phenomena arise. The cause of this phenomena can be explained as follows: "When two qualitatively different stimuli are applied to the *same* locus on the sensory surface very rapidly, rapidly enough so that the two stimuli are perceived as a single event, the perceptual qualities of the two [stimuli] merge" [MARK78]. Multiple stimuli response and sensory interaction are the crux of this dissertation. Some of the well-known and accepted intersensory theories and perspectives are presented in the next section.

C. THE SENSES

1. Classification

The concept of separate sense modalities has been around for a long time having its roots date back to the time of Aristotle (circa 384-322 B.C.) [WALK81]. Although we typically believe we have only five senses, we really have upwards of 30 or 40 senses depending on how the senses are classified. One such classification divides the senses into the following modalities: *Vision*, *Audition*, *Cutaneous Sensitivity*, *Olfaction*, *Gustation*, *Kinesthesis*, *Labyrinthine Sensitivity*, and *Organic Sensitivity*. [FOST68] Figure 1 depicts this classification of the senses along with associated sense organs, stimulus, and sensory qualities.

Modality	Sense Organ	Peripheral Nerve Endings	Cortical Nerve Projections	Normal Stimulus	Sensory Qualities
Vision	eye	rods and cones of retina	occipital lobe	photic energy	colors (red, gray)
Audition	ear	hair cells of organ of Corti	temporal lobe	acoustic energy	tones and noises
Cutaneous sensitivity	skin	specialized and free nerve endings	parietal lobe	mechanical and thermal energy	pressure pain, heat, cold
Olfaction	olfactory cleft of nostril	rods of olfactory epithelium	rhinencephalon	volatile substances	odors (fragrant, spicy)
Gustation	tongue and mouth region	taste buds of papillae	parietal lobe	soluble substances	sweet, salt, sour, bitter
Kinesthesis	muscles joints, tendons	specialized and free nerve endings	parietal lobe	mechanical energy	pressure, pain
Labyrinthine sensitivity	nonauditory labyrinth	hair cells of crista and macula	none (?); projects to the cerebellum	mechanical forces and gravity	none
Organic sensitivity	portions of gastro-intestinal tract	specialized and free nerve endings	parietal lobe	mechanical energy	pain, pressure

Figure 1. Classification of the Senses From [FOST68].

2. Sensory Interaction

In 1940, Ryan [RYAN40] conducted a thorough literature survey on sensory interaction. Based on the intersensory research investigated, the following are some of Ryan's findings:

> (1) ...it is extremely rare outside of the controlled conditions of the laboratory that even a single object is the product of operations of a single sensory system.
>
> (2) Under certain conditions it can be shown that qualities perceived by one sensory system are influenced by stimuli reaching other sense organs.
>
> (3) ...it is evident that sensory systems are part of a unified organism and by no means isolated from one another. [RYAN40]

Ryan ultimately concludes that the study of the interrelations among the senses is "...sorely in need of further investigation..." [RYAN40].

In 1941, Gilbert [GILB41] conducted another extensive literature review on intersensory facilitation and inhibition. It is interesting to note that Ryan was unaware of Gilbert's work until after Ryan's work was published, and Gilbert does not mention Ryan's efforts. Nevertheless, Gilbert makes the following conclusions concerning the effect of heteromodal (intersensory) stimulation on sensitivity to stimulus intensity:

> (1) Under conditions of momentary heteromodal stimulation (a) a sufficiently intense stimulus will momentarily reduce sensitivity in another modality, and increase it after an optimum interval (about 1/2 sec.); (b) a less intense heteromodal stimulus will momentary increase sensitivity.
>
> (2) Under conditions of prolonged stimulation, there is some evidence that the *quality* of the heteromodal stimulus may determine the direction of the effect, some stimuli acting as excitants, others as depressants. It is not clear, however, whether there is a differential effect among the various modalities.
>
> (3) The affect will be limited by the liability of the sensation affected, and individual differences in their susceptibility to heteromodal influence. [GILB41]

Upon reviewing all intersensory research (through 1941), Gilbert realized that the current view on the psychophysical aspect of intersensory interactions is lacking. Gilbert's final concluding remarks state that:

> Modern psychophysics has produced overwhelming evidence of the inadequacy of the traditional static relationship between stimulus and response, wherein each attribute of a sensory response was conceived of as determined simply by the value of a corresponding physical dimension of the "adequate" stimulus. Actual experimental evidence... has shown that the dimensions of stimulation are inter-dependent in affecting a sensory response, and that sensation may be dependent on the interaction of excitations, on mental set, physiological state of the organism, practice, and numerous other factors, all interrelated in a constant state of flux. [GILB41]

In 1947, Sherrington [SHERR47] tries to explain higher-order sensory integration as a process in which "...each sense system is served by specific receptors that project to specific sensory centers in the brain. Intersensory interaction is the concept by which multisensory stimuli of the real world (e.g., rhythm) are integrated in the brain" (summarized by [WALK81]).

In 1954, London [LOND54] presented his findings based on the extensive intersensory research conducted in the Soviet Union. Upon the review of numerous intersensory experiments, London concludes that the conditions that influence sensory interaction are best summarized as follows: 1) *Strength of accessory stimulus*, 2) *Excitatory state of sense organs*, 3) *Duration of accessory stimulation*, 4) *Termination of accessory stimulation*, 5) *Affectivity of stimulus*, 6) *Physiological state*, 7) *Diurnal variation*, 8) *Summation, repetition and cumulation of accessory effects* [LOND54] [STON68].

In reviewing London's research efforts, Stone and Pangborn findings indicate that:

> We respond to environmental stimuli through all avenues of sensory input, and, although the extent of their interrelationship is not well understood, it is generally accepted that the stimulation of one sense organ influences to some degree the sensitivity of the organs of another sense. [STON68]

Stone and Pangborn ultimately conclude that "...there exists a great need for further definitive [intersensory] studies. Quantification of individual variability in response to dual stimulation does not seem to have been investigated, nor has three-way stimulation been reported" [STON68].

In 1966, Gibson [GIBS66] [GIBS79] suggests that:

> ... perceptual systems cannot be gracefully categorized in terms of specific sensory systems, that under natural conditions many senses respond and interact to environmental stimulation, and the organism itself is initiating rather than reacting to events. This means that intersensory perception and integration are not specialized higher-order complex reactions, but are the rule for all perception. (summarized by [WALK81])

In other words, it is the particular surrounding environment which determines how our senses respond and interact. As a result, sensory interaction must be based on the complexity of natural life events and not on simple isolated systems.

In 1978, a more modern view of sensory interaction is provided by Lawrence Marks which is outlined in the excellent book, *The Unity of the Senses: Interrelations among the Modalities* [MARK78]. From a simple to a more complex perspective, Marks describes what he calls the *Five Doctrines* of sensory correspondence. Briefly, these five doctrines are outlined as follows:

> 1. *Doctrine of Equivalent Information.* ...different senses can inform us about the same features of the external world.
>
> 2. *Doctrine of Analogous Attributes and Qualities.* Despite the salience of the phenomenal differences among qualities of various sense modalities, there are a few properties held in common.
>
> 3. *Doctrine that Different Senses have Corresponding Psychophysical Properties:* ...this theory proposes that at least some of the ways the senses behave and operate on impinging stimuli are general characteristics of sensory systems, similar from vision to hearing, from touch to olfaction.
>
> 4. *Doctrine that Similar or Identical Neurophysiological Mechanisms Parallel Sensory Correspondence.* ...there is a neural analogue to each of the psychological doctrines [the first three doctrines].
>
> 5. *Doctrine of the Unity of the Senses.* ...incorporates all of the first four theories, and in which the several senses are interpreted as modalities of a general, perhaps more primitive sensitivity. [MARK78]

According to the various intersensory research studied by Marks, he believes that the dimension of *quality* appears to show the fewest similarities from modality to modality, but that *intensity* displays the strongest cross-modal similarity. However, Marks concedes that "The entire area of cross-modality comparisons of sensory quality has hardly been explored experimentally" [MARK78]. Furthermore, Marks concludes that any sensory interaction is highly stimuli dependent. As Marks explains:

> Perhaps the most crucial factor in determining the significance of any interaction is the objective relationship between the stimuli that are used. When stimuli presented to different senses bear no meaningful relation to each other, interaction often seems to be small or nonexistent. ...But meaningfully related stimuli are quite a different matter. ... Meaningful perceptual interactions...occur when concurrent information enters different sensory channels.[MARK78]

An interesting point by Marks which deserves mentioning is that:

> Similarity across the senses must necessarily be one step removed from similarity within a sense, for there is, by definition, no continuity between modalities. If the senses were truly continuous there would only be one sense. [MARK78]

In 1981, based on her research with blind and normal children, Susanna Millar [MILL81] concludes that the sense modalities are neither separate nor unitary. "They [modalities] are some of both, complementary to each other, and information can be used flexibly from different modalities" [WALK81]. A further conclusion that Millar makes is that "...we are slowly beginning to understand the interrelationships of the sense modalities. Global generalizations do not seem to hold. No one current theory seems capable of encompassing the diversity of findings" [WALK81].

In 1981, O'Connor and Hermelin [OCON81], having conducted experiments with children suffering from either specific perceptual or general cognitive handicaps, describe sensory integration through the concept of *sensory capture* as follows:

> One aspect of sensory integration can be demonstrated by the phenomenon of "sensory capture," in which conflicting input to different sense modalities is often not perceived as such. Instead, the observer seems to resolve such conflict by making one sense impression conform with another dominant one. ...Such "capture" of one sensory input by another is of interest because it suggests that there may be a degree of perceptual equivalence between various sensory information, so that the same stimulus qualities tend to be perceived in various modalities. [OCON81]

3. Neurological Perspective

Because of recent advances in technology in the field of neurology, there has been a surge in intersensory research from a neurological perspective. The reason for this much deserved neurological emphasis it that:

> ...there has been comparatively little done to understand the neural phenomena that make multisensory integration possible. The paucity of neural data about multisensory integration is due in part to different strategies researchers have used to explore the functional organization of the nervous system, and also to the inherent difficulties in conducting multisensory studies. ...For while the perceptual phenomena demonstrates that interactions among different sensory modalities are commonplace and that constancies among the modalities must exist in order to use them together effectively, there is no comparable body of literature describing the neural mechanisms that underlie them. Nevertheless, there is a good deal of information about the location in the brain where inputs from different modalities converge. [STEI93]

One place in the brain where visual, auditory, and somatosensory inputs converge is in the superior colliculus as depicted in Figure 2. Furthermore, in looking at the horizontal and vertical meridians of the different sensory representations in the superior colliculus, one can see that they are very similar in terms of a common coordinate system. Stein and

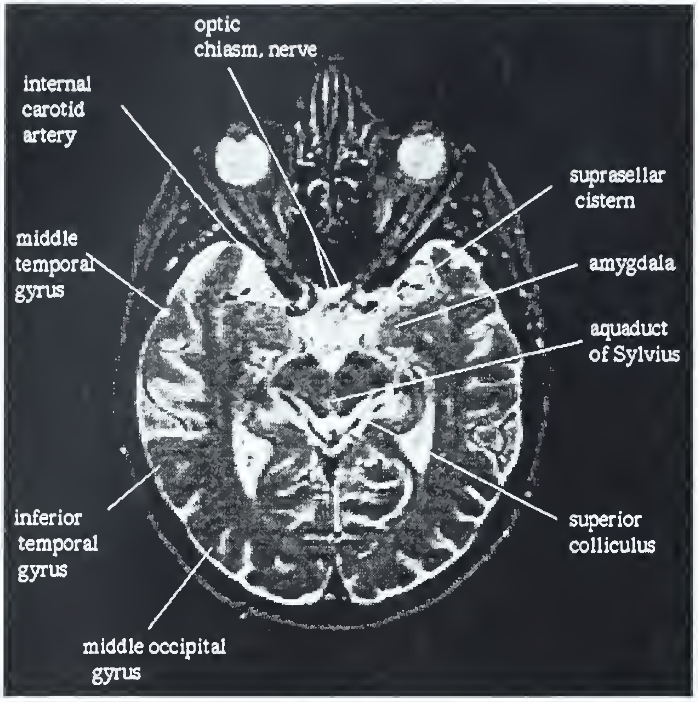


Figure 2. The Superior Colliculus From [HARV98].

Meredith conclude that this common coordinate system suggests a representation of *Multisensory Space* (see Figure 3). By examining the neurological responses of superior colliculus in various animals, primarily the cat, Stein and Meredith have found considerable evidence supporting the principles of multisensory convergence and interaction based on single neuron evoked potentials as depicted in Figure 4. Stein and Meredith believe that neurological studies in other animals are very important and lead to a better understanding of human perception. Thus, based primarily on the neurological studies of other animals, primarily cats, Stein and Meredith outline the rules in terms of space and time governing multisensory integration as based on unimodal receptive field characteristics as follows:

> Space: spatially coincident multisensory stimuli tend to produce response enhancement, whereas spatially disparate stimuli produce either depression or no interaction.

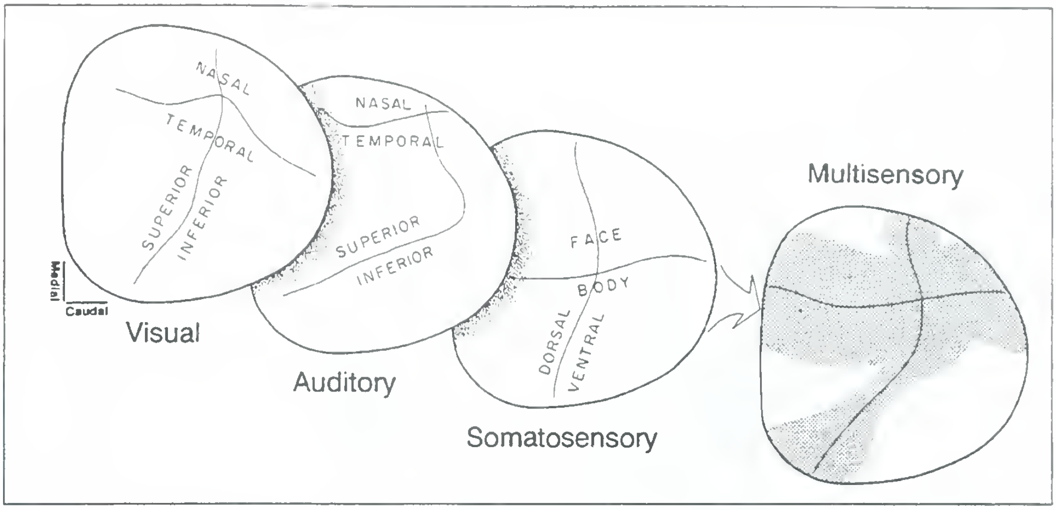


Figure 3. Common Coordinate System in the Superior Colliculus Suggesting Multisensory Space From [STEI93].

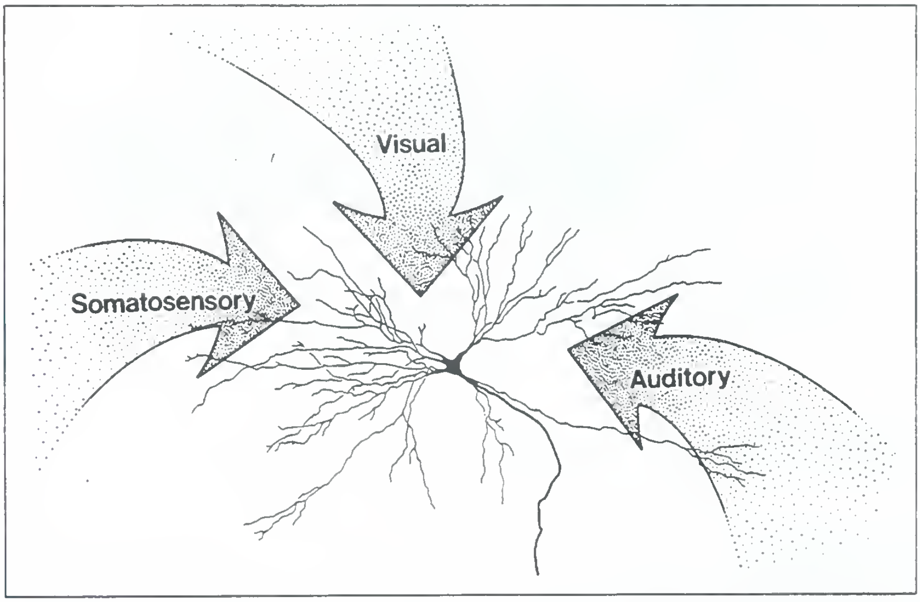


Figure 4. Convergence of Inputs from the Different Senses on a Single Neuron From [STEI93].

Time: maximal multisensory interactions are not dependent on matching the onset of two different sensory stimuli, or their latencies, but on how the activity patterns resulting from the two inputs overlap.

[Overall]...the spatial register among the receptive fields of multisensory neurons and their temporal response properties provide a neural substrate for enhancing responses to stimuli that covary in space and time and for degrading responses that are not spatially and temporally related. [STEI93]

Although they found considerable evidence supporting a neurological basis for sensory integration, Stein and Meredith conclude that: "an enormous number of challenges must be met before we understand more fully the process involved in integrating information from different sensory modalities" as seen in Figure 5.

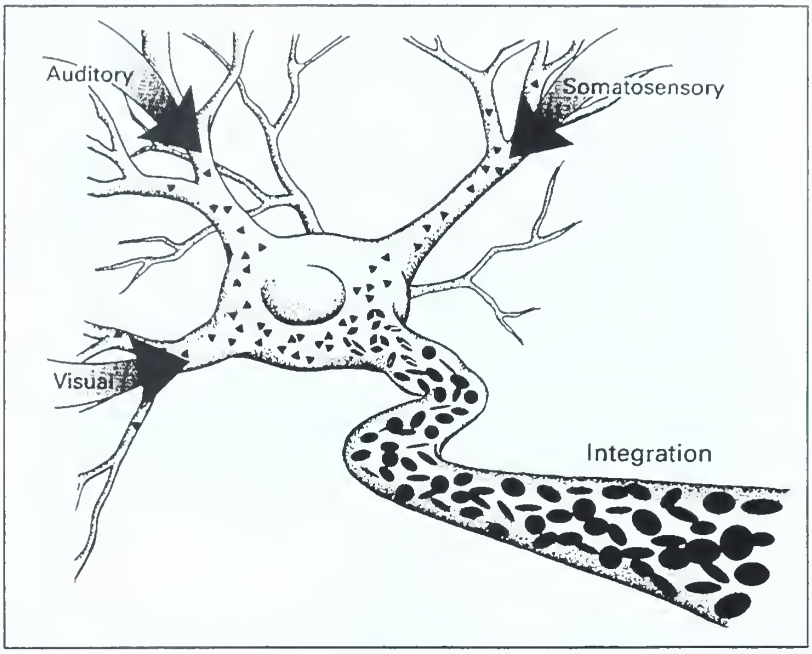


Figure 5. Neurons Synthesize Information from Different Sensory Modalities From [STEI93].

D. AUDITION

1. Definition

Before audition can be defined, we need to have an understanding of what is meant by sound. The following gives a formal definition of sound:

> Sound is the perception by humans of vibrations in some physical medium, usually air. These physical vibrations of the air are evidenced by alternating rarefractions and compressions. Man's primary sense organ for the sound stimulus is the ear. [SILB68] (see Figure 6)

The formal definition of hearing (the sense of audition) from a physiological perspective is as follows:

> Hearing is the response of an animal to sound vibrations by means of a special organ for which such vibrations are the most effective stimulus. The critical phrase here is "most effective," which means that this special organ (which we shall call an ear) is more sensitive to sound than it is to any other form of energy. All other mechanoreceptors respond to acoustic vibrations if these vibrations are strong enough and sufficiently low in frequency, but they do so crudely, requiring large amounts of energy in comparison with what they require in the stimuli that are most appropriate to them and in relation to what the ear requires within its proper frequency range. Organs in the skin (tactual and deep pressure endings) in muscles, tendons, and joints (kinesthetic endings), in the vestibular labyrinth (gravity and motion receptors), and even pain organs throughout the body can all be excited by sounds of sufficient strength. But none of these organs approaches the ear in delicacy and in the effectiveness of utilization of sounds as a means of gaining information about the outside world. [WEVE74]

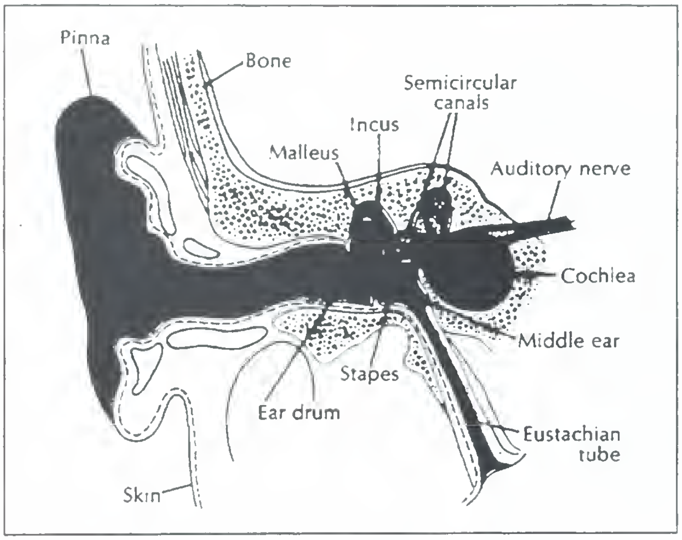


Figure 6. The Ear From [MURC73].

In other words, although the entire human body is capable of *hearing* sounds, the ear is the most sensitive to sound which in turn makes it the primary mechanism for hearing sounds.

2. Subjective Evaluation

Given that we can hear sounds, how do we rate the *quality* of sound? What is of good quality to one person may be of bad quality to another. As a result, rating the quality of sound is a subjective task based largely on the rendering capability of the equipment that is generating the task. Another aspect to the quality of sound is that of content. For example, some may like to listen to rock-and-roll where intentional distortion is often reproduced as high quality; whereas, others may think the musical *quality* of rock-and-roll is poor. Content is an important consideration when conducting sound quality tests of loudspeakers or headphones, and studies have shown that when conducting sound quality experiments "...the problem of selecting test material was evident. Relevant test material has not yet been defined. Different recording techniques influence the assessment of the sound quality" [THEI86]. Although content is important, this research effort focuses on the perception of the physical characteristics of the sound. But what physical characteristics, dimensions, attributes, etc., of sound are applicable to rate?

Zwicker and Zwicker [ZWIC91] propose that:

> The information received by our auditory system can be described most effectively in the three dimensions of specific loudness, critical-band rate, and time. The resulting three-dimensional pattern is the measure from which the assessment of sound quality can be achieved. [ZWIC91]

In experiments conducted to identify perceived sound quality of loudspeakers, Gabrielsson and Lindström had subjects rate music on a category scale from 0-10 using the following dimensions: "*Clarity*, *Fullness*, *Spaciousness*, *Brightness*, *Softness*, *Absence of Extraneous Sounds*, and *Fidelity*." [GABR85] as depicted in Figure 7.

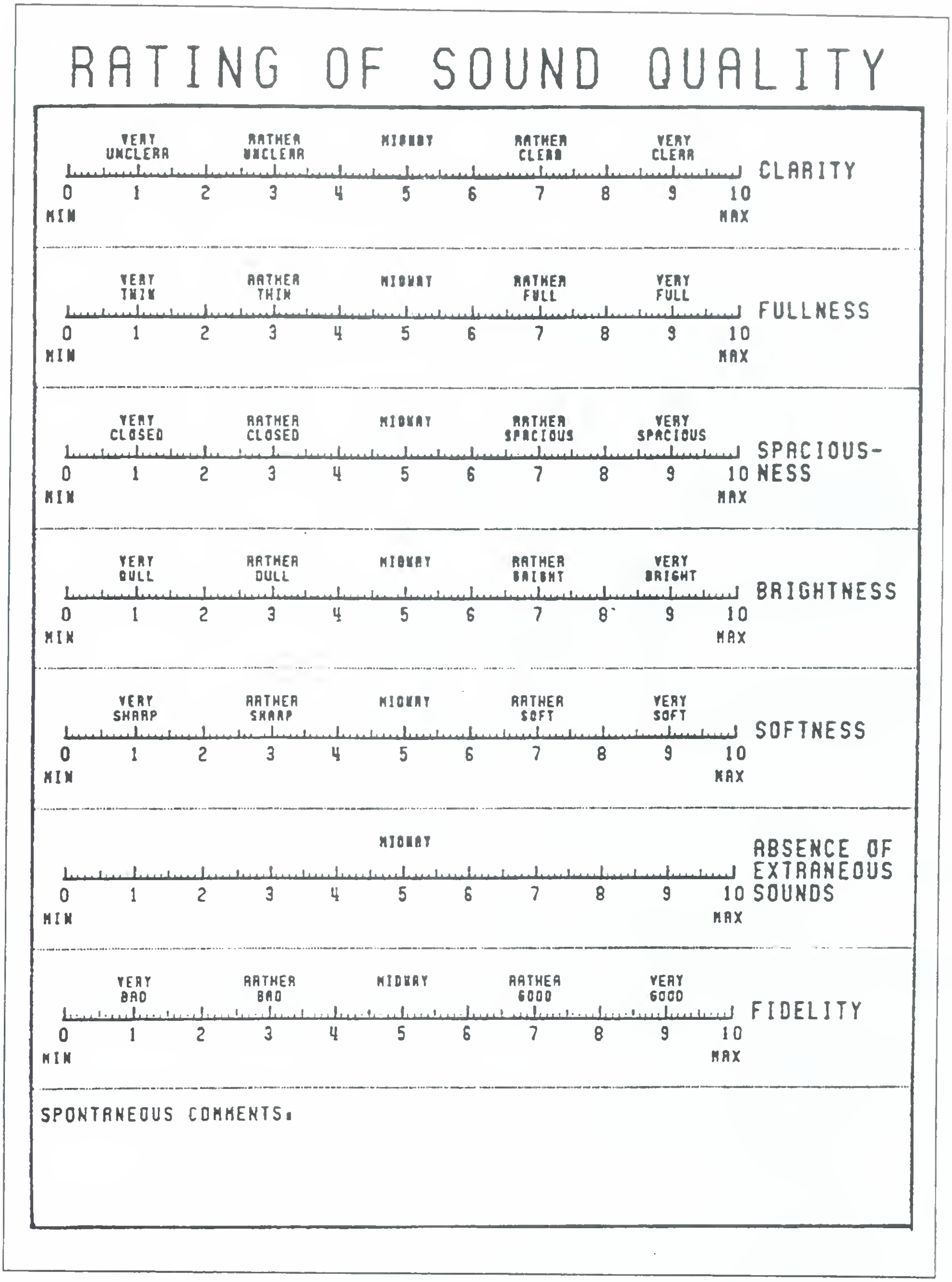


Figure 7. Sound Quality Rating Scale From [GABR85].

Based on Gabrielsson and Lindström's efforts, Toole [TOOL85] expanded the dimensions on which to rate sound quality to include a specific rating format for spatial quality as depicted in Figure 8.

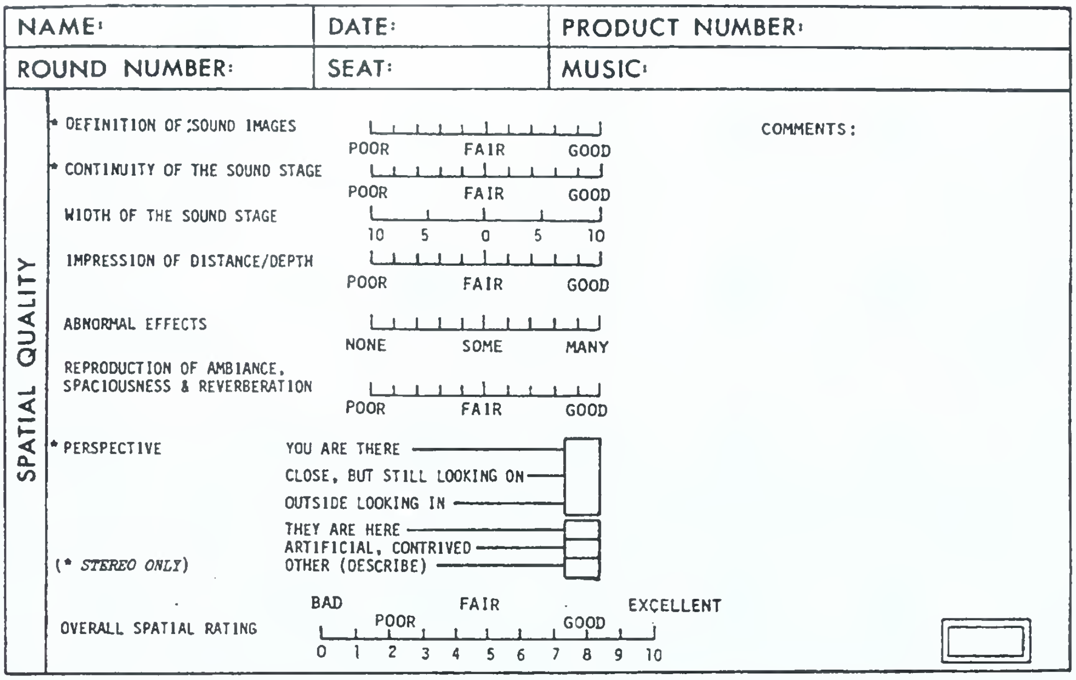


Figure 8. Spatial Quality Rating Scale From [TOOL85].

In evaluating the quality of loudspeakers using an impulsive tone-burst signal, Furmann et al. [FURM90] had subjects rate the following attributes on a scale of 0-10:

1) *Sharpness* -- The sound contains components whose mid-and high-frequency levels are too high.

2) *Pureness* -- The sound is not distorted, devoid of sounds not appearing in the signal, readable in the entire frequency range.

3) *Equalness* -- The sound retains the proportion of tónes; it is linear without expansion of tones.

4) *Clearness* -- The sound is pure and clear; different instruments and voices can be distinguished easily; onsets and transients in the music can be perceived easily.

5) *Feeling of Space* -- The reproduction is spacious; the sound is open, has width and depth, fills the room, gives the impression of the subjects presence in the space surrounded by sound. [FURM90]

In measuring subjective and objective acoustical measurements, Burkhard and Genuit [BURK92] recognize that any acoustical measurement system should yield information that relates to how humans hear. As such, Burkhard and Genuit identify the relevant parameters that are involved during the classification of a sound event by a human listener as seen in Figure 9.

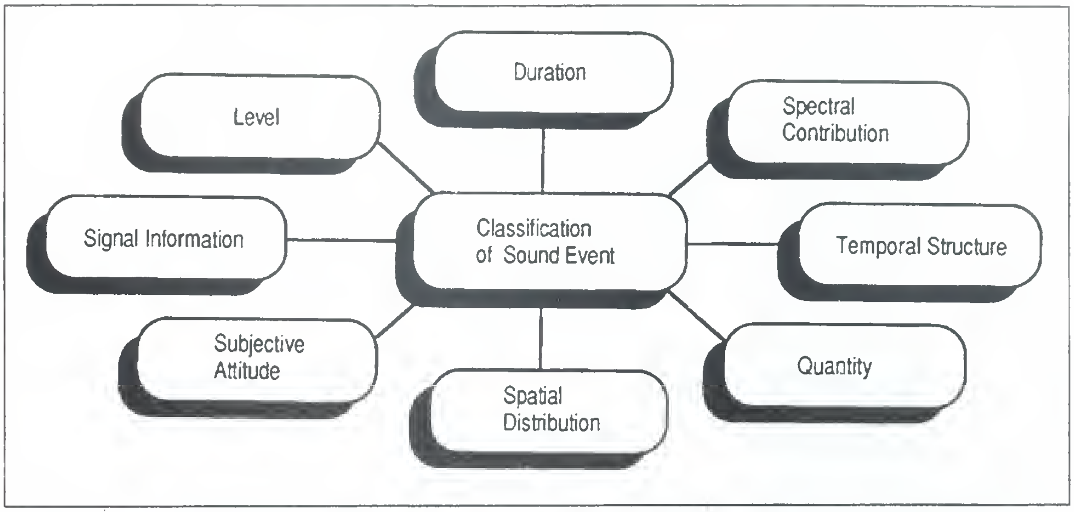


Figure 9. Parameters Relevant to Evaluation of Sound by Human Listeners From [BURK92].

In terms of spatial hearing, Blauert [BLAU97], identifies proven and hypothesized psychophysical theories corresponding to positional auditory events. These events are categorized as follows: *Basic vs. Supplemental*, *Homosensory vs. Heterosensory*, and *Fixed-position vs. Motional*. The physical processes and phenomena which make use of these psychophysical theories are outlined in Figure 10. For more insights in how humans perceive the quality of sound, see the following: [BECH90] [TOOL90] [VIEM90] [BURK92] [THUR92].

Physical phenomena and processes considered	Participating sensory organs	Usual designation	Categorization
Sound conducted through the air to one or both eardrums	Hearing (one ear suffices)	Monaural theories for air-conducted sound	B, Ho, F
Interaural differences for air-conducted sound at both eardrums	Hearing (both ears necessary)	Binaural theories for air-conducted sound	B, Ho, F
Sound conducted through the air to the eardrums and sound conducted through bone in the skull (generated by air-conducted sound)	Hearing	Bone-conduction theories	S, Ho, F
Sound conducted through the air to the eardrums and light on the retinas	Hearing, vision	Visual theories	S, He, F
Sound conducted through the air to the eardrums and to the cochlea and vestibular organ	Hearing, sense of balance	Vestibular theories	S, He, F
Sound conducted through the air to the eardrums and sound received by tactile receptors (such as the hair at the nape of the neck)	Hearing, sense of touch	Tactile theories	S, He, F
Head movements during which air-conducted sounds are modified at the eardrums	Hearing, sense of balance; receptors of tension, position, and orientation; vision	Motional theories	S, He, M

Categories: Basic (B) vs. Supplemental (S); Homosensory (Ho) vs. Heterosensory (He); Fixed-position (F) vs. Motional (M).

Figure 10. Psychophysical Theories of Spatial Hearing From [BLAU97].

E. VISION

1. Definition

A formal definition of vision is as follows.

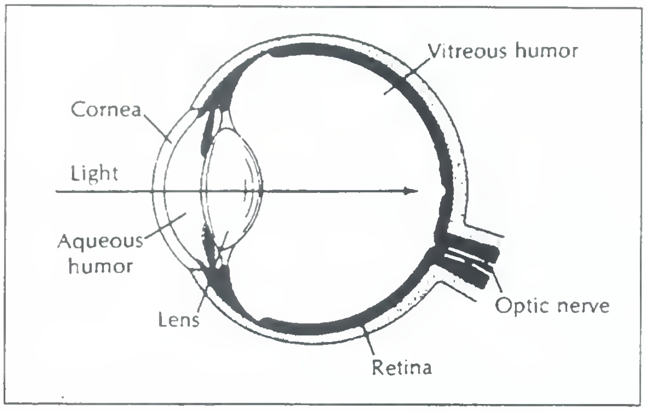


Figure 11. The Eye From [MURC73].

> Vision is a complex phenomenon consisting of several basic components. Sight from external sources is brought to a focus on the retina of the eye. Changes are produced which initiate electrical impulses. These are conducted over the optic nerve and optic tract to the brain where the visual sensation is perceived and interpreted. [MCNA68] (see Figure 11)

2. Subjective Evaluation

An approved method for the subjective evaluation of visual displays can be found in the *Method for the Subjective Assessment of the Quality of Television Pictures* published by the Geneva International Telecommunications Union [GENE86]. This publication recommends using a five-point rating scale for evaluating quality. The five points on the rating scale are as follows: 1 *Bad*, 2 *Poor*, 3 *Fair*, 4 *Good*, and 5 *Excellent*. Also, the use of non-expert observers is recommended, and the number of observers should be at least ten and preferably twenty. Also, the publication recommends that an experimental testing session should not last more than roughly 30 minutes, and that a duration of 10 seconds for visual stimuli is sufficient for still or moving sequences. Furthermore, the publication suggests that visual stimuli may be based on a randomizedblock design derived from Greco-Latin squares. (See [GOOD95] for an example of the Latin squares technique.)

After an exhaustive literature review, Padmos and Milders [PADM92] present a long list of quality criteria for simulator images. This list includes criteria based on: *Visually Perceiving the Environment*, *Physical Image Properties*, *Image Capacity*, *Appearance of Surfaces*, *Visibility and Light Effects*, and other miscellaneous features. The target simulator for this quality criteria is that of the vehicle simulator, but the criteria apply equally well to virtually any type of simulator image.

3. Visual Dominance

The current view of visual dominance can be attributed to the work of Posner et al. (see [POSN76]). Posner's efforts tried to identify why the visual modality tends to "dominate conscious judgements about the presence and location of objects" [POSN76]. Posner's general theory of visual dominance includes the following four propositions:

> Proposition 1. Visual stimuli are not as automatically alerting as stimuli in other modalities.
>
> Proposition 2. In order for a visual event to serve as an effective alerting stimulus, the subject must first process it by active attention.
>
> Proposition 3. The consequence of active attention toward any one modality is a reduction in the availability of the attentive mechanisms to input from other modalities.
>
> Proposition 4. To compensate for the low alerting capability of visual signals, subjects exhibit a general attentional bias toward the visual modality whenever they are likely to receive reliable input from that modality. This bias may not be obvious to them, but it can be viewed as a strategy of a very pervasive sort. [POSN76]

F. ATTENTION

"The essence of the concept of attention is the focusing of awareness" [DEMB79]. Our span of attention is derived from our span of perception. Perception spans the range from subliminal stimuli (unconscious awareness) to liminal stimuli (conscious awareness) as depicted in Figure 12. Using the common searchlight metaphor as depicted in Figure 12, the three main aspects of attention in perception are as follows: 1) *Selective Attention*: corresponds to the direction of the search light; 2) *Focused*

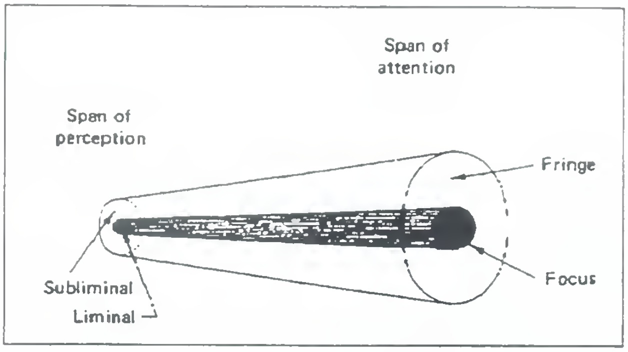


Figure 12. The Span of Attention and the Span of Perception From [DEMB79].

Attention: corresponds to the immediate center of the beam of light illuminated by the searchlight; and 3) *Divided Attention*: corresponds to both the immediate center of the beam of light and the fringe just outside the beam of light. Overall, attention plays a pivotal role in human information processing, one that not only selects information sources to process but also acts as a commodity or resource of limited availability [WICK92] (see Figure 13).

1. Selective Attention

As the searchlight metaphor explains, selective attention directs the searchlight. Thus, selective attention is concerned with the process of how, when, what, and where we actually focus on (or attend to) various and numerous stimuli. The selection process acts as sort of a filter between sensory processing and attention as depicted in Figure 14. Numerous theories over the years have tried to describe the nature of this selection process. One of the more popular theories is *Broadbent's Filter Theory* [BROA58].

a. Broadbent's Filter Theory

> Broadbent proposed that the brain contains a selective filter which chooses messages on the basis of physical characteristics toward which it is "tuned" and rejects others. The filter spares the limited-capacity system from being overloaded; complex forms of input are rejected on the basis of simple qualities, and a higher-level analysis of them need not

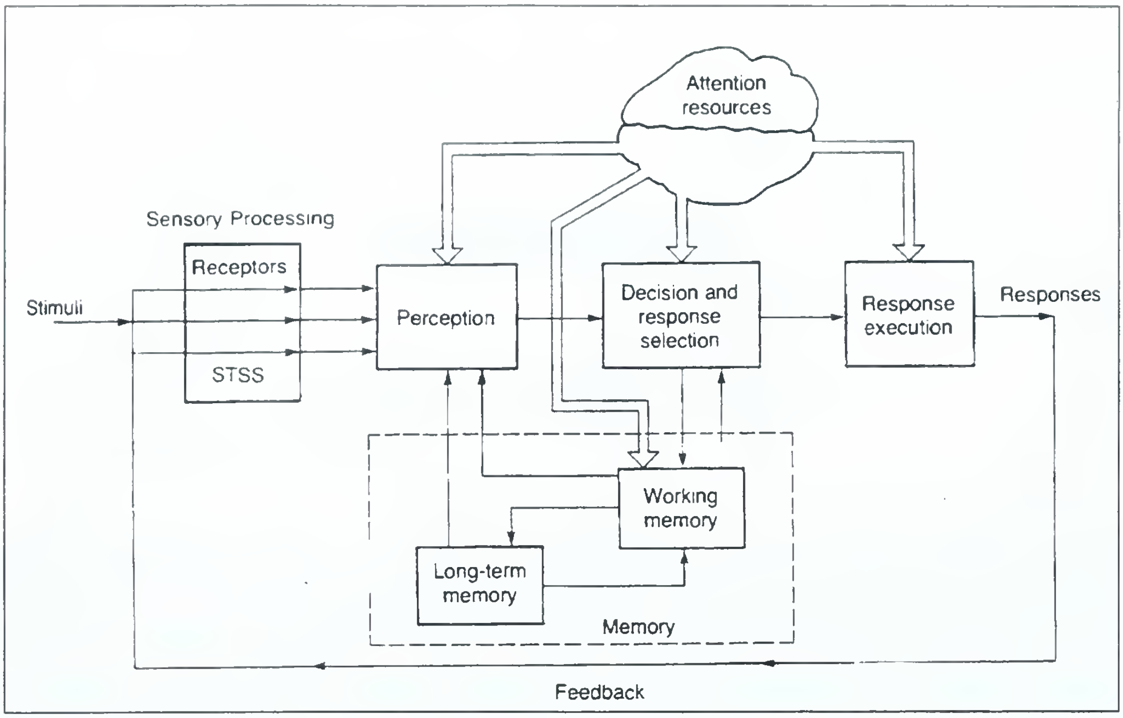


Figure 13. A Model of Human Information Processing From [WICK92].

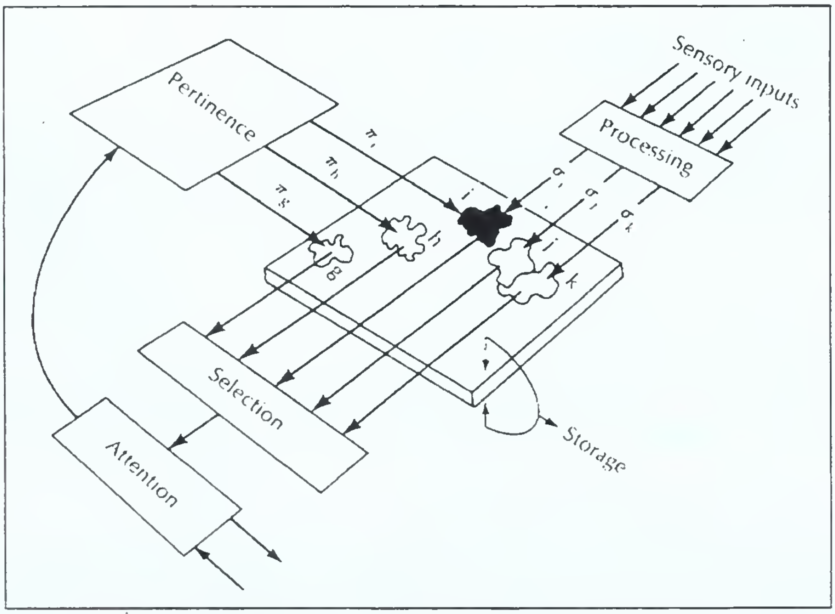


Figure 14. Selective Attention From [MURC73].

occur. ...In essence, the filter model views the selective nature of attention as resulting from restrictions in the capacity of the nervous system to process information. ...Preference is shown for novel or intense events, acoustic over visual signals, sounds of high frequency, and signals of biological importance to the organism. [DEMB79] (see Figure 15)

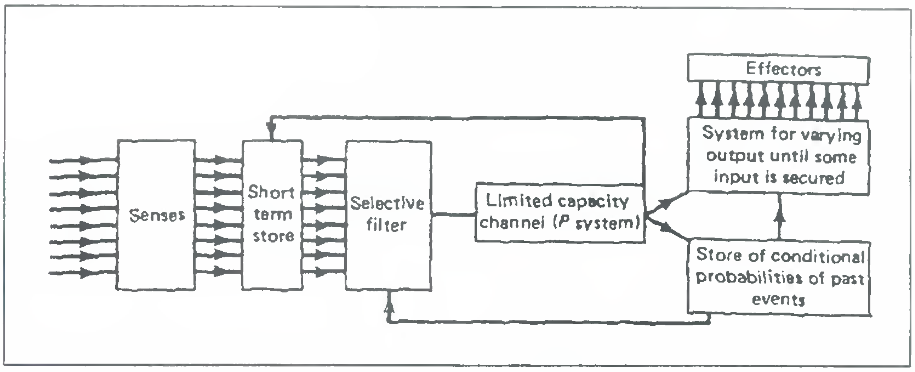


Figure 15. Information-Flow in Broadbent's Filter Theory From [DEMB79].

b. *Filter Attenuation Theory*

Although the *Filter Theory* seemed adequate, a number of studies, primarily conducted by Anne Treisman [TREI69] [TREI73], soon identified certain limitations. As a result, a modification was made to the *Filter Theory* resulting in the *Filter Attenuation Theory*.

> The essence of this modification is that filtering is not an all-or-none affair. Treisman suggested that the filter does not cut off rejected messages entirely, but instead attenuates their strength. Thus, under some conditions, the weakened signals can still contact higher-level elements of the perceptual system. [DEMB79] (see Figure 16)

c. *Response-Selection Theory*

An entirely different perspective of selection attention was formalized by Deutch and Deutch [DEUT63]. This theory, called the *Response-Selection Theory*, maintains "...that *all* mental inputs are fully analyzed perceptually and that selection takes place only when the observer responds to stimuli" [DEMB79].

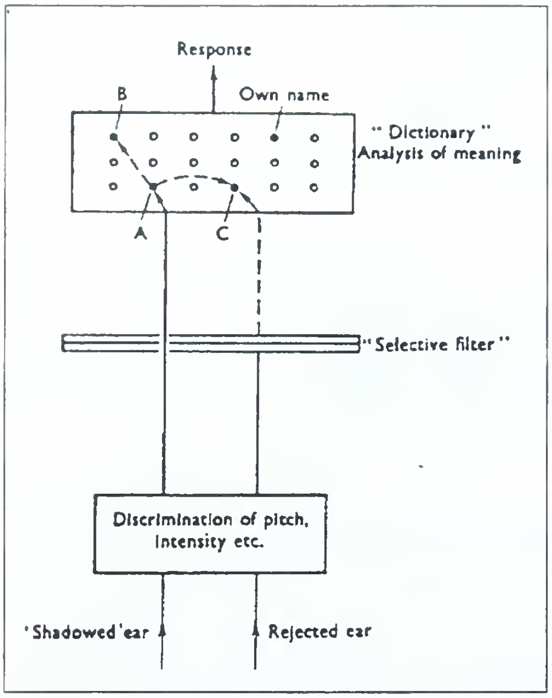


Figure 16. Information Flow in Treisman's Filter Theory From [DEMB79].

d. Hybrid Theory

Recognizing the debate over the various theories of selective attention (which continues still today), Dember [DEMB79] suggests another possible solution as follows:

> It is conceivable that our cognitive capacities are more flexible than we have been willing to assume, and that both perceptual and response selection can take place under appropriate circumstances. ...This new breed of attentional theory may very well prove of conceivable value in directing research toward a more satisfactory solution to the mystery of selection attention. [DEMB79]

2. Divided Attention

Whereas selective attention deals with our ability to direct our focus among stimuli, divided attention deals with our ability to divide our attention among stimuli or tasks. Divided attention occurs when "the task is to attend to several simultaneously active input channels or messages, responding to each as needed" [BOFF86]. Early researchers believed that it was impossible to attend to several simultaneous stimuli -- that attention was indivisible. Nowadays, divided attention is readily believed, but how we divide our attention has raised considerable debate. The issue is whether or not we process simultaneous inputs in parallel or in serial. However, the conclusions drawn from considerable research suggest that "...both modes of processing occur, depending on the task and on the circumstances," [KAHN73] and whether or not the stimuli are intramodal or intermodal. Our ability to divide our attention among various stimuli directly corresponds to our limited ability to *time-share* among these various stimuli.

3. Time-Sharing

Our ability to time-share depends on how efficient we schedule and switch between various stimuli. For example, if we are given plenty of time to complete two separate tasks, we will probably complete one task then switch to completing the other task. However, if the amount of time we are given is drastically reduced, we might have to engage in completing both tasks concurrently. Processing tasks concurrently leads to three further factors which will influence our ability to successfully complete concurrent processing. These factors are: *confusion* of the task, *cooperation* between task processes, and *competition* for task resources. [WICK92]

> *Confusion* results when elements for one task become confused with the processing of another task because of their similarity.
>
> *Cooperation* occurs when there is a high similarity of processing routines between tasks which can result in the possible integration of the two task elements into one.
>
> *Competition*, the critical element of concurrent task time-sharing, relates to the level of difficulty between the tasks -- the greater the difficulty, the greater the competition. [WICK92]

When we say that difficult tasks (stimuli) are in competition with one another, this competition refers to competing for the limited amount of total available resources needed to complete the tasks. With this in mind, there are two theories on how resources are allocated to attention: 1) *Single-Resource Theory*, and 2) *Multiple-Resource Theory*.

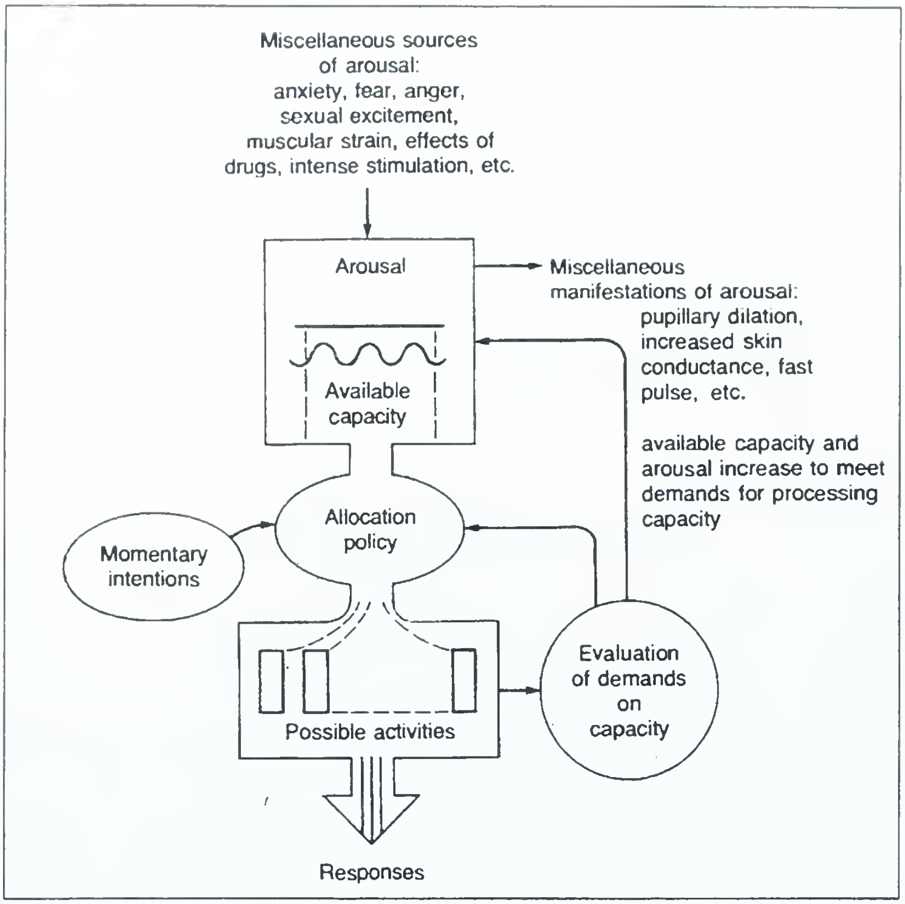


Figure 17. Single Resource Theory From [WICK92].

a. Single-Resource Theory

The Single-Resource Theory (see [KAHN73]) argues that we have one single supply of undifferentiated resources available to all tasks and mental activities. "As task demands increase either by making a given task more difficult or by imposing additional tasks, physiological arousal mechanisms produce an increase in the supply of resources" [WICK92]. The Single-Resource Theory is depicted in Figure 17. The main limitation of this theory is that it compares task difficulty within the same dimensional constraints. As such, it does not consider the structure of the task as it relates to the

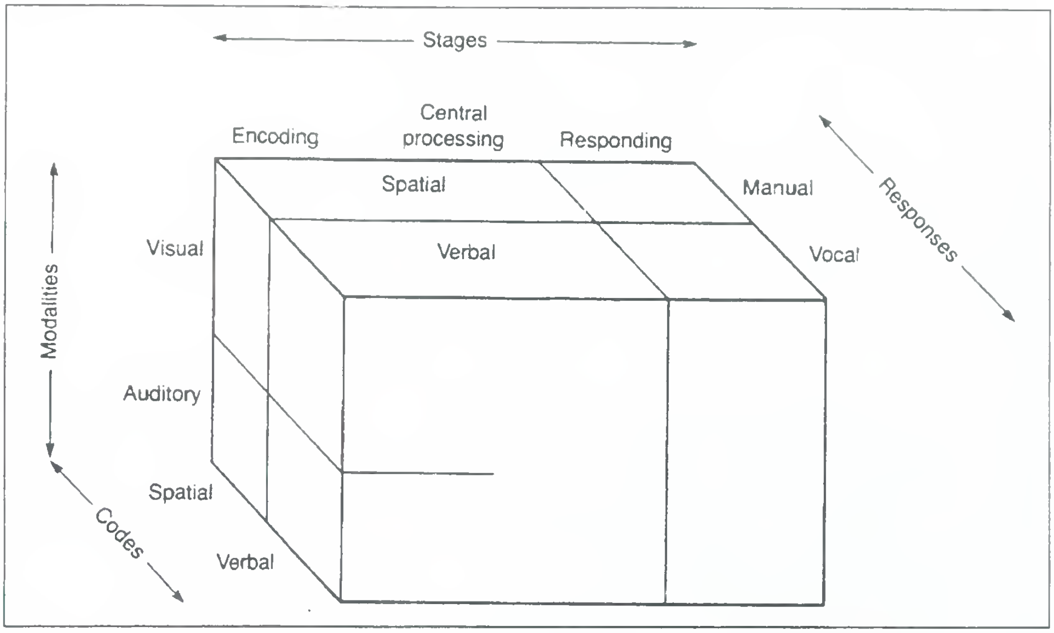


Figure 18. Multiple Resource Theory From [WICK92].

processing of the task such as its *Codes*, *Modalities*, and *Stages*. [WICK92] Correcting this limitation provides the impetus for the *Multiple-Resource Theory*.

b. Multiple-Resource Theory

The Multiple-Resource Theory stipulates that tasks are processed based on multi-dimensional constraints. These constraints involve the task's *Codes* (Spatial vs. Verbal), *Modalities* (Auditory vs. Visual), and *Stages* (Encoding, Central Processing, and Responding) as depicted in Figure 18. As such, "...people have several different capacities with resource properties. Tasks will interfere more and difficulty-performance trade-off's will be more likely to occur, if more resources are shared." [WICK92] For example, two visually dominating tasks may compete for the same resources resulting in greater interference (competition) of the two tasks. But, if one task is visually dominating and one task is aurally dominating, they may not have to compete with each other, for they utilize separate resources as depicted in Figure 18 as opposed to common resources as depicted in Figure 17.

4. Sustained Attention

Sustained attention deals with our ability to maintain focused attention over prolonged time periods. Sustained attention is commonly referred to as vigilance. During the early Cold War years (1950s through 1980s), there was an increased threat of global thermonuclear war. As such, radar operators monitored their radar scopes for potential incoming missiles for prolonged periods of time (vigilance). Because of the severe repercussions that could result if a radar and/or sonar operator missed a *bleep* on the scope, the study of vigilance became very popular (on both sides of the cold war). The results of these studies provided new insights into such theories as: Vigilance, Signal Detection, Expectancy, Arousal, and Habituation. The concept of sustained attention does not play a role in this dissertation. It is being presented to complete the discussion of attention and to clarify the issues of attention that are relevant to this research effort. During the preliminary literature review of this dissertation, much time was spent reviewing auditory-visual vigilance studies. For a listing of pertinent auditory-visual cross-modal signal detection and vigilance research, see APPENDIX B. AUDITORY-VISUAL CROSS-MODAL SIGNAL DETECTION AND VIGILANCE BIBLIOGRAPHY.

5. Cognitive Ecology Perspective

Ecology is the study of the interaction of living creatures with their environment. For ecological psychology, the focus is the relation of mind to environment. Cognitive Ecology is a new field "... a deep ecology of the mind, in which mind and environment are treated not as separate objects or topics but as codefining poles of experiences and actions" [FRIE96]. In the book, *Cognitive Ecology* [FRIE96], two qualitatively different aspects of attention are described as having: (1) a clear nucleus of focus of attention, and (2) a *fringe* to that experience. The focus of attention refers to the typical searchlight metaphor of attention. The *fringe* refers to:

> ... many types of experience, such as: (1) feelings of familiarity, (2) feelings of knowing, such as tip-of-the-tongue-experiences, (3) feelings of relation between objects

or ideas, (4) feelings of action tendency, as in intentions, (5) feelings of expectancy, (6) feelings of rightness or being on the right track. ...(7) metaknowledge of one's memory or one's abilities... [and] (8) Perhaps the most pervasive fringe feeling is that of meaningfulness, that one knows the larger context of any given moment of focal attention although that context is not part of the content of attention. [FRIE96]

There are three issues in which this *fringe* experience are relative to cognitive ecology: 1) the issue of knowledge of content, 2) the issue of capacity, and 3) the issue of agency. The second issue, that of capacity, identifies potential shortcomings of the tradition view of attention. Specifically:

> Attention is normally viewed either explicitly, or more recently implicitly, as a limited-capacity system. ...This may be because only focal attention is normally investigated. A mind that is defined literally as part of its environment (the subjective pole of attention in a subject-object field) should have much broader attentional capacities than a mind defined as separate. Many of the anomalies of attention and consciousness research, such a blind sight and the other agnosias, are cases that violate the standard limited-capacity conception. Investigation of fringe phenomena may serve to expand, or perhaps undermine, models of attentional limits. [FRIE96]

G. GESTALT THEORY

Gestalt Theory was founded by German Psychologists Max Wertheimer [WERT12], Kurt Koffka [KOFF35], and Wolfgan Kohler [KOHL40]. The basic idea of Gestalt Theory is that we perceive things wholistically as opposed to its parts. "Certainly to process information as wholistic or gestalt stimuli rather than as separate elements is an efficient thing for the organism to do -- and possibly that is the advantage of gestalt patterns" [GARN70]. As a result, to view things as whole, rather than as parts, we perceptually organize things, objects, etc. into groups. The *Gestalt Factors of Perceptual Organization* include the following:

> 1) Factor of Similarity, 2) Factor of Proximity, 3) Factor of Common Fate, 4) Factor of Objective Set, 5) Factor of Inclusiveness, 6) Factor of Good Continuation, 7) Factor of Closure, 8) Factor of Fixation, 9) Factor of Contour, and 10) Factor of Object Interdependence. [MURC73]

Gestalt Theory was developed primarily to explain how we perceptually group visual objects, but its concepts can also be applied to the other senses.

H. SYNESTHESIA

One of today's leading experts in the study of synesthesia is Richard Cytowic. He defines synesthesia as

> ...an involuntary joining in which the real information of one sense is accompanied by a perception in another sense. In addition to being involuntary, this additional perception is regarded by the synesthete as real, often outside the body, instead of imagined in the mind's eye. [CYTO89]

It is estimated that synesthesia occurs in about one in 25,000 individuals [CYTO95], so its occurrence is fairly rare. One of the most common forms of synesthesia is that of colored hearing. A synesthete experiences colored hearing when certain sounds (physical stimuli) evoke perceptions of various colors. For example, when listening to certain classical music, a synesthete might experience shades of blue and/or green. Colored hearing is the most common form of synesthesia. Another more bizarre example is that of gustatory-tactile synesthesia. In this case, the synesthete experiences (perceives) certain shapes based on various tastes (physical stimuli) (see Figure 19) In fact, because of the bizarre nature of this condition, Cytowic wrote an entire book based on the research of a man with gustatory-tactile synesthesia. See [CYTO93] for an in-depth review of gustatory-tactile synesthesia.

The concept of synesthesia dates back over two hundred years. For an exhaustive survey of all classic and contemporary synesthesia literature dating back over this interval, see [BARO96]. The validity of synesthesia, though, has suffered over the years for it is introspective in nature. However, Cytowic has helped to validate synesthesia by examining the neural substrates of synesthesia as outlined in [CYTO89] [CYTO93]. The results of Cytowic's research indicate that:

> The synesthetic experience may be a result of a fundamentally mammalian process in which the cortex briefly ceases to function in the modern manner, permitting the senses to fuse, or, rather, we should say, perceive fusion that may be there all along but that never arises to consciousness. At its essence, synesthesia may be a remnant of how early mammals perceived their world. ...Synesthesia is what we all do without knowing that we do it, whereas synesthetes do it and know that they do it. [CYTO89]

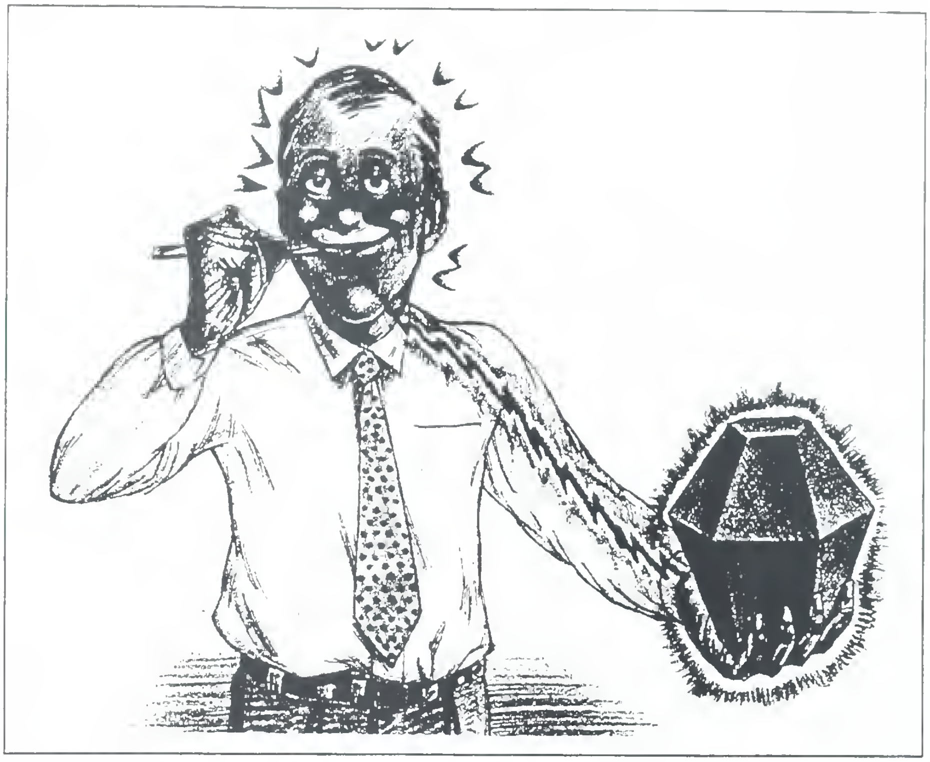


Figure 19. Tasting Shapes From [CYTO89].

I. MULTIMEDIA

"According to a recent projection, multimedia and creative technologies will represent a new market of $40 billion by the year 2000 and $65 billion by the year 2010" [GUPT97]. As such, there is indeed a market emphasis on multimedia and there are still many unanswered questions. To support the continued growth of multimedia, it must expand and develop in parallel with internet technology, not as an afterthought or as an add-on. As such,

> ... the central integrated media-systems-related issue that must be addressed during the next decade is storage, indexing, structuring, manipulating, and "discovery" of integrated multimedia information units (MIUs) that include structured data values (strings and numbers), text, images, audio, and video. The key research focus in this area centers on managing multimedia information units in the context of a highly distributed and interconnected network of information collections and repositories. Current data and knowledge management technology that addressees collections of formatted data and text

is inadequate to meet the needs of video and audio information, as well as the mixture of modalities in MIUs. [GUND97]

In [BLAT96], Blatter and Glinert express the need for a greater understanding and need for multimodal integration. They correctly recognize that "Although we have seen much progress in recent years in the use of single modalities, the general problem of designing *integrated multimodal systems* is not well understood" [BLAT96]. One of the reasons for the current lack of integrated multimodal systems is that the system designers, i.e. computer scientists, are not knowledgeable with the issues associated with multimodal concepts. Thus,

> ...the (computer) scientists who design the new interfaces and human-computer communications devices must address issues whose solutions lie outside of their discipline. Integrating modalities requires understanding how people use their various senses to perceive and interact with the world around them. Despite more than 100 years of research into these issues, much remains unknown. [BLAT96]

As a result, "Research by non-computer scientists shows that computer scientists have sometimes failed to appreciate the distinction between human and computer modalities" [BLAT96]. This explains why it is typical to judge a simulation or virtual environment by the auditory and visual technical rendering capabilities of the system (computer and displays), as opposed to how well stimulated are the auditory and visual sensory modalities of the immersed participant, i.e. an engaged human.

Brenda Laurel [LAUR93], provides numerous insights into the use of multimedia and human-computer interaction. She states that "Multiple modalities are desirable only insofar as they are appropriate to the action being represented" [LAUR93]. With an artistic background, Laurel brings a much-needed dimension to field of multimedia. With her creative experience, she correctly recognizes that an artistic touch can lead to better (smarter) multimodal integration in multimedia systems. Accordingly, Laurel states:

> But we mustn't fall prey to the notion that more is always better, or that our task is the seemingly impossible one of emulating the sensory and experimental bandwidth of the real world. Artistic selectivity is the countervailing force -- capturing what is essential in the most effective and economic way. A good line-drawn animation can sometimes do a better job of capturing the movements of a cat than a motion picture, and no photograph will ever capture the essence of light in quite the same way as the paintings of Monet. The point is that first-person sensory and cognitive elements are essential to human-

> computer activity. There is a huge difference between an elegant, selective multi-sensory representation and a representation that squashes sensory variety into a dense but monolithic glob of text. [LAUR93]

Thus, we must not assume that we always need the best possible graphics and audio. The particular application, overall sensory perception, and creative use of stimuli ought to drive fidelity requirements.

J. SUMMARY

In summary, this chapter has provided the computer scientist with a high-level overview of Perception, The Senses, Audition, Vision, Attention Theory, Gestalt Theory, Synesthesia, and Multimedia.

III. LITERATURE REVIEW

A. INTRODUCTION

This chapter presents a literature review on relevant auditory-visual cross-modal perception phenomena. Whereas the background provided in the previous chapter presents a general overview of the concepts underlying the psychological and physiological nature of auditory and visual perception, this chapter specifically focuses on VEs and auditory-visual intersensory phenomena. Using the background provided in the previous chapter, the reader can better understand the theoretical basis and overall findings of the numerous auditory-visual research endeavors outlined in this chapter.

B. VIRTUAL ENVIRONMENTS

1. Definition

The National Research Council's (NRC) *Committee on Virtual Reality Research and Development* defines VE systems with the following explanation:

> Virtual environment systems differ from other previously developed computer-centered systems in the extent to which real-time interaction is facilitated, the perceived visual space is three-dimensional rather than two-dimensional, the human-machine interface is multimodal, and the operator is immersed in the computer-generated environment. [DURL95]

But what does *virtual* mean? Ellis [ELLI96] tries to clarify the term virtual by introducing the concept of *virtualization* which is the "...process by which a viewer interprets patterned sensory impressions to represent objects in an environment other than that from which the impressions physically originate" [ELLI96]. Ellis continues to explain that virtualization applies primarily to vision and audition and that there are three levels of virtualization: *Virtual Space*, *Virtual Image*, and *Virtual Environment* as depicted in Figure 20. Furthermore, because of the diverse nature of VEs, the NRC *Committee* explains that the development of a VE requires "...a crucial need for cooperation among many disciplines, including computer science, electrical and

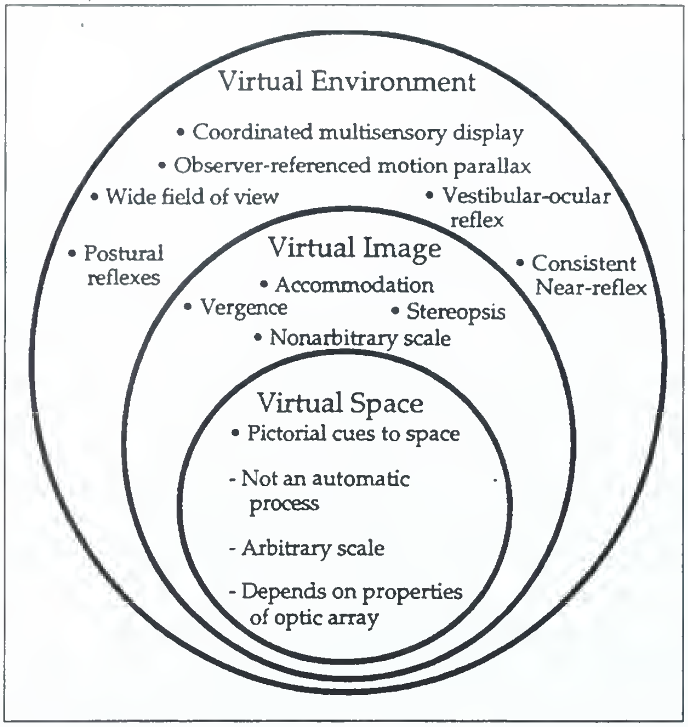


Figure 20. Levels of Virtualization From [ELLI96].

mechanical engineering, sensorimotor psychophysics, cognitive psychology, and human factors" [DURL95]. Cross-disciplinary transfer of knowledge is typically lacking, causing a potential degradation of VE development. This dissertation attempts to better facilitate cross-disciplinary transfer of knowledge and to hopefully improve VE development with respect to auditory-visual cross-modal perception considerations.

2. Multimodal Concerns

"...the development of multimodal synthetic environments is an extremely important and challenging endeavor. [It]...requires that we carefully examine our current assumptions concerning VE architectural requirements and design constraints" [DURL95]. One of the first multimodal networked VEs was that of *Networked SPIDAR* [ISHI94]. In this networked VE, participants collaborated on the design of 3D objects using visual, audio, and haptic information. The developers of *Networked SPIDAR* believed that "A networked virtual environment must support these interactions [visual,

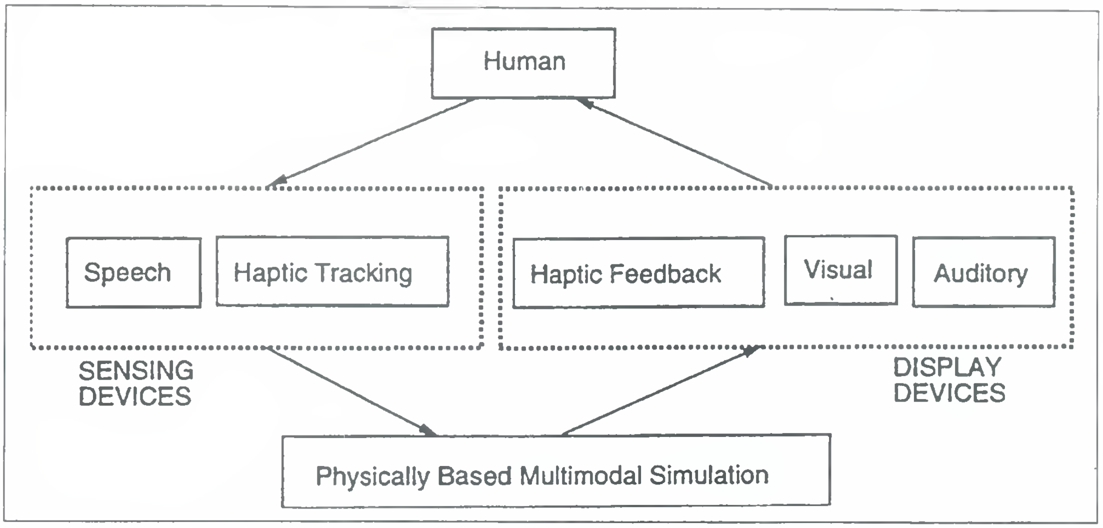


Figure 21. Multimodal Modes in Virtual Environments From [GUPT97].

audio, and haptic] without contradiction in either time or space" [ISHI94]. Gupta et al. [GUPT97] also describes experiments using multimodal environments to enhance computer-aided design (CAD). They describe the relationship of the inserted human participant to auditory, visual, and haptic feedback devices as depicted in Figure 21. However, the majority of research and development in VEs has typically focused on the sense of vision (i.e., the visual channel). Accordingly:

> To date much of the design emphasis in VE systems has been dictated by the constraints imposed by generating the visual scene. The nonvisual modalities have been relegated to special-purpose peripheral devices. ...However, many of the issues involved in the modeling and generation of acoustic and haptic images are similar to the visual domain; the implementation requirements for interacting, navigating, and communicating in a virtual world are common to all modalities. Such multimodal issues will no doubt tend to be merged into a more unitary computational system as the technology advances over time. [DURL95]

Thus, proper VE development must focus on all modalities equally. This focus on the modalities need not only concentrate on the intra-relationships but also on the inter-relationships. As the NRC *Committee* explains: "Detailed study of both intrasensory and intersensory illusions is important because, in many cases, the existence of illusions enables SE [synthetic environment] systems design to be simplified and therefore to increase its cost-effectiveness" [DURL95]. Furthermore, under the category of *Psychological Considerations* the NRC *Committee* recommends further study in "...channel-interaction effects that occur with multimodal interfaces." Some notable channel-interaction (intersensory) effects:

> ...include those on the dominance of vision over audition and haptics in cases of intermodality conflict (e.g., as evidenced in the ventriloquist effect) and on the use of auditory stimuli to improve the perception of events that are represented primarily in the visual or haptic domains (as in the use of sound effects) [DURL95].

It seems fairly obvious by this point that proper development of VEs must consider multimodal factors. Since we currently have the technology to render very high quality auditory and visual displays, the proper use of this technology must not neglect potential auditory and visual cross-modal perception phenomena. Brenda Laurel makes the point that auditory and visual cross-modal issues have always been a consideration in the art world. Now with the recent surge in the development of VE technology, the same cross-modal considerations of the Arts apply to VEs. Brenda Laurel states:

> VR has reinvigorated and recontextualized the study of human sensation and perception. While much is known about the human visual or auditory or tactile senses, relatively little is known "scientifically" about how these senses combine. Still less is known about how they combine in the context of representations, as opposed to the context of the actual world. For example, it is well known in the folklore of computer game design that high-quality audio makes people perceive visual displays to have higher resolution. It is also well-known that the converse is not true: Great graphics will not turn a PC's beeps and boops into Beethoven.The study of sensory combinatorics, that is, how vision affects audition or how the two in concert affect emotion, was almost exclusively the province of the arts until VR came on the scene. [LAUR93]

3. Fidelity Requirement

What are the fidelity requirements of a VE? First and foremost (and sometimes neglected), the intended outcomes of the particular application ought to drive the fidelity requirements. For example, the visual fidelity of a VE intended to train surgeons in open-heart surgery probably needs to be greater than the visual fidelity of a VE intended to teach children how to read. Another consideration is that of the human sensory system: the fidelity requirements of VEs need not exceed that of the human perceptual system. As such, "Knowledge of normal human resolving power on the input side, i.e., the sensory side, allows one to predict the display resolution beyond which finer resolution cannot be perceived and would therefore be wasted" [DURL95]. For example, the auditory fidelity of many VEs, in terms of frequency range, need not exceed that of the nominal range of human hearing (i.e., 20 Hz - 20 kHz). A caveat pertains here: some research indicates that our perceptual frequency range is much greater (see [OOHA91] [BOYK97]). Nevertheless, the capabilities of the human sensory system ought to drive the fidelity requirements of VEs as depicted in Figure 22.

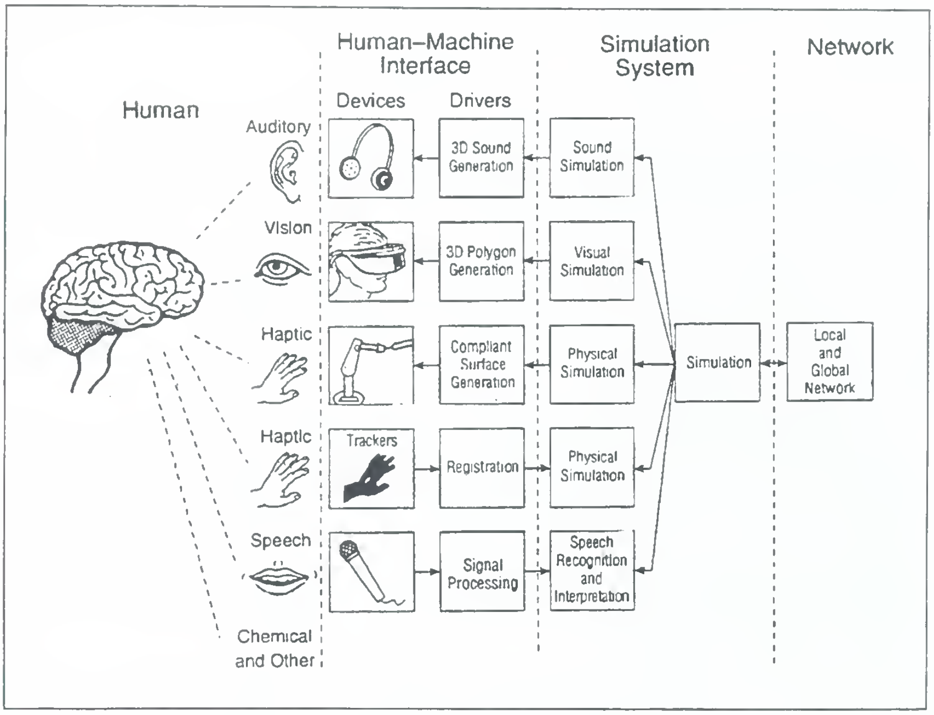


Figure 22. Computer Technology Organization for Virtual Reality From [DURL95].

Details regarding humans' ability to detect and discriminate visual, auditory, tactile, and kinesthetic information along with corresponding technical specifications of VE equipment is presented in the excellent paper by Barfield et al. [BARF95]. Barfield states that "It is important to have a thorough understanding of the capabilities of the human's sensory systems and to use this knowledge in the design of virtual worlds and in deriving technical specifications for virtual environment equipment" [BARF95].

When Barfield compares the human sensory system with technical specifications of VEs, he considers the modalities as separate entities. However, the VE participant, being human, is multimodal by nature. As a result, one very key consideration neglected in Barfield's paper is how the senses interact, and another is how this sensory interaction may or may not conflict with how the singular modality capabilities derive the specifications of VEs. The NRC *Committee* also recognizes that visual fidelity requirements are influenced by other modalities and that a greater understanding is needed in multimodal integration in hopes of answering the following unanswered questions:

> How are the required visual display system parameters affected within multimodal systems? Can visual display system requirements be relaxed in multimodal display environments? What are the perceptual effects associated with the merging of displays from different display sources? [DURL95]

One factor in considering auditory and visual fidelity requirements is that of display resolution. In a VE, the auditory and visual resolutions ought to be properly matched. As Brenda Laurel correctly states:

> ... we also sometimes expect certain kinds of patterns to occur. Although, there are many reasons for emphasizing one modality over another, we tend to expect that the modalities involved in a representation will have roughly the same "resolution." A simplistic cartoon-style animation with naturalistic character voices and environment sounds, for instance, seems out of whack. A computer game that incorporates breathtakingly high-resolution, high-speed animation but only produces little beeps seems brain-damaged. [LAUR93]

On analyzing the use of performed sound and music in VEs, Pressing [PRES97] classified sound into three categories: 1) *artistic expression*, 2) *information transfer*, and 3) *environmental sounds*. Pressing concluded that: "Across all three categories the need for further research on the psychological aspects of sound and performance in virtual environments was apparent" [PRES97]. Another fidelity consideration is that "...cartoons and caricatures, despite their drastic loss of information and fidelity, may better serve to represent the world, clarify visual relationships...and effect our thoughts...than pictures of high fidelity" [FRIE96]. Similarly, on integrating sounds and motions in VEs, "Sounds tend to affect the listener in a more subconscious and impressionistic way than visual cues" [HAHN98]. Furthermore, when considering the fidelity requirement of VEs, there are many perspectives from which to view fidelity, perhaps all of which are correct! Flach and Holden [FLAC98] outline the following definitions of fidelity from various scientific perspectives.

> 1) Newton's Way: Fidelity is derived from three-dimensional space and time (e.g., chronometric analysis).
>
> 2) Einstein's Way: Since space and time are relative to a certain frame of reference, they cannot be scientifically committed to any sense of realism; therefore, space and time cannot be used as a measure of fidelity.
>
> 3) Fechner's Way: Fidelity is defined in relation to the correspondence between the simulated world and the "real" world as measured using the ruler and clock of classical physics.
>
> 4) Helmholtz's Way: Fidelity is defined relative to the ability to simulate the biological mechanisms -- the proximal stimulus. Thus, binocular and binaural inputs might be considered essential to a high-fidelity experience of space.
>
> 5) Broadbent's Way: Information processing rate, sensitivity, bias, and stability might prove the best measures of fidelity.
>
> 6) Dewey's Way: The measure of fidelity is the degree to which the simulation captures the richness of natural couplings between perception and action.
>
> 7) Gibson's Way: With fidelity, the constraints on action take precedence over the constraints on perception, and reality of experience is defined relative to functionality, rather than to appearances. (Paraphrased from [FLAC98])

4. Presence

Presence, the sense of *being there*, has been a heavily debated topic among VE developers. There is no argument that the sense of presence within a VE is an extremely vital aspect of any VE, and that "...virtual environments that are best at simulating multiple senses are also best at evoking a feeling of presence an immersion" [ANDE97]. The debate over presence is a debate about definition and measurement. Depending on your interpretation, there can be many possible meanings of presence. For instance, a well-written book can cause one to be immersed into the intricacies of a good plot. A great live theater production or cinematic movie can also stir the senses causing a sense of *being there* -- presence. In VE applications, we typically measure presence by how well our senses (all of them) are stimulated. For "...it is both the interactivity and the quality of the rendering that results in the *immersiveness* of a virtual reality or multimedia system" [BEGA94]. Sheridan [SHERI96] makes an interesting observation that through evolution, our senses developed in order, from tactile to vision to audition, but that technology used to stimulate our senses has developed in reverse, from audition to vision to tactile as depicted in Figure 23.

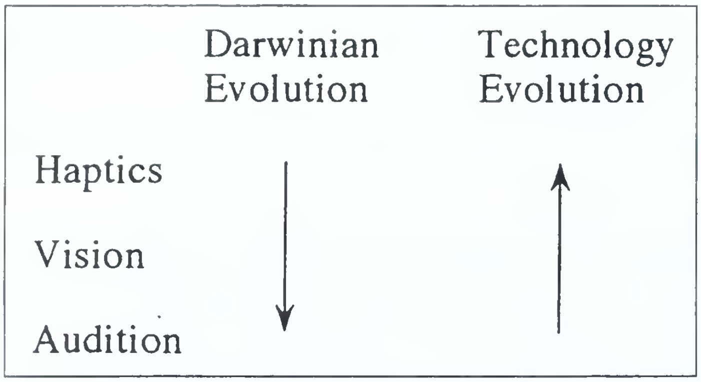


Figure 23. Darwinian Vs. Technological Evolution From [SHERI96].

In VE applications, most agree that the level of presence is directly proportional to the level of audio, visual and tactile fidelity. Accordingly, "Tight linkage between visual, kinesthetic, and auditory modalities is the key to the sense of immersion that is created by many computer games, simulations, and virtual-reality systems" [LAUR93]. As such, the level of fidelity is directly proportional to the level of presence. Thus, the level of presence must be a function of fidelity. Nevertheless, most do not agree on how to measure the level of presence. Sheridan uses the following *Three Attribute Scale of Presence* to rate the fidelity of picture, sound, and tactile images.

> 1. Virtual image resolution (pixels or taxels per frame), refresh rate (frames per second) and gray-or color-scale (bits per pixel or taxel) are too few to convey realism.
>
> 2. Virtual image fidelity is fairly realistic. Resolution (pixels or taxels per frame), refresh rate (frames per second) and gray-or color-scale (bits per pixel or taxel) are enough to convey good sense of reality.
>
> 3. Virtual image is compelling. Difficult to discriminate the virtual from the real based on any given image. [SHERI96]

Slater and Wilber [SLAT97] discuss various parameters affecting presence including the parameter of *vividness* as it relates to pictorial realism. They describe an experiment using a driving simulator in which two different levels of the pictorial realism were presented to the immersed participant. The results indicated that: "There was a significant difference in the level of reported presence between the two levels of pictorial realism, with the more realistic resulting in a higher level of reported presence" [SLAT97]. As a result of their research, Slater and Wilber introduce the *Framework for Immersive Virtual Environments* (FIVE) which shows the relationship to presence among several factors including visual, auditory, and tactile displays as depicted in Figure 24. Also, in a previous research effort [SLAT94], Slater found that a person's dominant sense may influence a person's sense of presence.

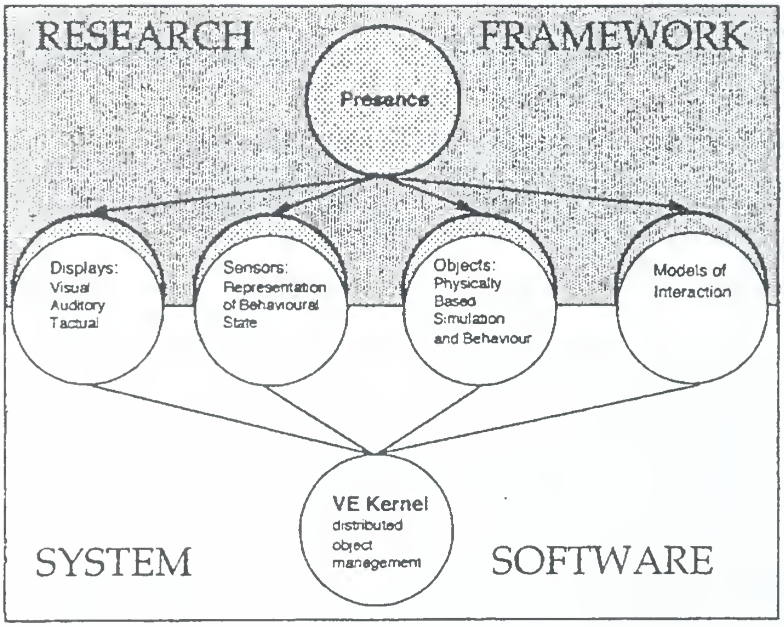


Figure 24. Framework for Immersive Virtual Environments From [SLAT97].

Hendrix [HEND94] [HEND96a] [HEND96b] conducted a number of experiments to measure the level of presence within VEs during a navigation task as function of visual and audio display parameters. In one set of experiments, the visual display parameters manipulated were: 1) presence or absence of head tracking, 2) presence or absence of stereoscopic cues, and 3) size of geometric field of view used to create the visual image projected on the visual display. In another set of experiments, the audio display parameters manipulated were: 1) presence or absence of spatialized sound, and 2) nonspatialized versus spatialized sound. The results from the experiments involving visual display parameter manipulation concluded: "...a significant positive correlation between the reported level of presence and the fidelity of the interaction between the virtual environment participant and the virtual world" [HEND96a]. The results from the experiments involving audio display parameter manipulation indicated that:

> ...the addition of spatialized sounds significantly increased the sense of presence but not the realism of the virtual environment. Despite this outcome, the addition of a spatialized sound source significantly increased the realism with which the subjects interacted with the sound source, and significantly increased the sense that sounds emanated from specific locations within the virtual environment. The results suggest that, in the context of a navigation task, while presence in virtual environments can be improved by the addition of auditory cues, the perceived realism of a virtual environment may be influenced more by changes in the visual rather than auditory display media. [HEND96b]

As such, although spatialized sounds can increase the sense of presence with in a VE, the perception of realism in a VE is still dominated by the visual modality.

C. AUDITORY-VISUAL PERCEPTUAL ORGANIZATION

1. Gestalt Theory

The perception of an auditory-visual display can be considered in terms of the Gestalt point of view. If we extend the *Gestalt Factors of Perceptual Organization* discussed earlier in GESTALT THEORY (Chapter II, Section G) from visual-only stimuli to visual and audio stimuli, the factors of *Similarity*, *Proximity*, *Fixation* and *Object Interdependence* become particularly interesting to the possible perceptual grouping of an auditory-visual display. The definitions of these (visual) factors are as follows:

> *Similarity*: If a number of elements are present in the perceptual field, those with similar characteristics will be seen as though they are grouped together.

> *Proximity*: Elements of the perceptual field located near one another will tend to be seen as a group or unit.
>
> *Fixation*: The organization of certain kinds of patterns clearly depends on where the observer fixes his attention.
>
> *Object Interdependence*: ...prevalent in the organization of complex patterns encountered in visual experience is a tendency to group objects that are functionally rather than physically similar. We frequently see objects in this way if they display some kind of interdependent relationship. [MURC73]

When a high-quality visual display is coupled with a high-quality auditory display, for the intended presentation of an audio-visual display, the factor of *Similarity* may cause a perceptual *quality* grouping of the audio-visual display. Also, through the perceptual illusion of the ventriloquism effect, the audio portion of an audio-visual display may perceptually emanate from the proximal locality of the visual display perhaps causing a perceptual grouping based on the factor of *Proximity*. When viewing any audio-visual display, the observer must, at sometime, fixate on the display which in turn might cause a perceptual grouping by the factor of *Fixation*. Furthermore, since it is typical to hear music playing on a radio, music (audio) and a radio (visual) may be perceptually grouped together through the factor of *Object Interdependence*.

2. Auditory Scene Analysis

In terms of auditory-visual interaction, Al Bregman mentions in his book, *Auditory Scene Analysis: The Perceptual Organization of Sound* that there many similarities between visual and auditory perceptual groupings. Specifically,

> ... the similarity of principles of organization in the visual and auditory modalities is that the two seem to interact to specify the nature of an event in the environment of the perceiver. This is not too surprising, since the two senses live in the same world and it is often the case that an event that is of interest can be heard as well as seen. Both senses must participate in making decisions of "how many," of "where," and of "what." [BREG90]

But as opposed to the Gestalt point of view, which focuses on the similarities among modalities, Bregman also presents an interesting ecological point of view which focuses on the differences of the modalities.

> There is a crucial difference in the way that humans use acoustic and light energy to obtain information about the world. This has to do with the dissimilarities in the ecology

of light and sound. In audition humans, unlike their relatives the bats, make use primarily of the sound-emitting rather than the sound-reflecting properties of things. They use their eyes to determine the shape and size of a car on the road by the way in which its surfaces reflect the light of the sun, but use their ears to determine the intensity of the crash by receiving the energy that is emitted when this event occurs. The shape reflects energy; the crash creates it. For humans, sound serves to supplement vision by supplying information about the nature of events, defining the "energetics" of a situation. [BREG90]

This difference between vision and audition is further evidenced through the use of echoes. In audition, we are mainly interested in the direct source of sound rather its echoes, but we can also combine direct sound and indirect sound (echoes) to establish a mixed sound which still conveys information of the direct sound but with the additional properties (i.e. reverberation) of the indirect sound. However, with vision, we are mainly concerned with the indirect image (echoes or reflections), and we are not able to combine direct and indirect images to establish a mixed visual image. Bregman suggests that it is these ecological differences which might cause "apparent violations of the principle of exclusive allocation of sensory evidence." [BREG90]

D. AUDITORY-VISUAL ART FORMS AND FILM

1. Art Forms

In terms of the Arts, Joseph Schillinger explains the correlation of visual and auditory art forms through mathematics. Schillinger believed that:

> A scientific theory of the arts must deal with the relationship that develops between works of art as they exist in their physical forms and emotional responses as they exist in their psycho-physiological form, *i.e.*, between the forms of excitors and the forms of reaction. As long as an art-form manifests itself through a physical medium, and is perceived through an organ of sensation, memory and associative orientation, it is a *measurable quantity*. Measurable quantities are subject to the laws of mathematics. Thus, analysis of esthetic form requires mathematical techniques, and the synthesis of forms (the realization of forms in an art medium) requires the *technique of engineering*. [SCHI48]

Schillinger referred to the visual art form as *Elements of Visual Kinetic Composition* and the auditory art form as *Elements of Music*. The *Elements of Visual Kinetic Composition* consisted of the following four main components:

1. Linear, plane and solid trajectories (distance, dimension, direction, form).

2. Illumination (forms and intensity of light).
3. Texture (density of matter, quality of surface).
4. General component: time. [SCHI48]

The *Elements of Music* consisted of the following five main components:

1. Frequency (pitch).
2. Intensity (relative dynamics).
3. Quality (harmonic composition).
4. Density (quantitative aggregation of sound).
5. General component: time. [SCHI48]

As such, Schillinger believed that mathematics might appropriately describe visual and auditory correlated art forms and that "The correlation of the general component in both art forms may be assigned to different proportionate relations, such as harmonic ratios, distributive powers, series of growth, etc." [SCHI48]. Some of these mathematical relations which describe art forms are depicted in Figure 25.

$$\frac{\text{quality of matter's surface}}{\text{pitch} + \text{relative dynamics}} \qquad \frac{\text{quality of matter's surface}}{\text{relative dynamics} + \text{harmonic composition}}$$

$$\frac{\text{quality of matter's surface}}{\text{harmonic composition} + \text{quantitative aggregation of sound}} \qquad \frac{\text{quality of matter's surface}}{\text{quantitative aggregation of sound} + \text{pitch}}$$

Figure 25. Combined Visual-Auditory Art Form Mathematics From [SCHI48].

Furthermore, Figure 26 depicts Schillinger's concept of the overall relationship among the components of a combined kinetic art form.

2. Film

For many years, the entertainment industry has realized the important relationship between visuals and sound. Even before sound was an integral part of film, *silent* movies were accompanied with specific music to enhance the *mood* of certain scenes. As Gary Rydstrom of Skywalker Sound explains:

> Storytelling, mood setting, character development, drama and style can all be more successfully realized by the careful collaboration of images and sounds. There is a magical level reached when picture and sound work together, a creative dimension not

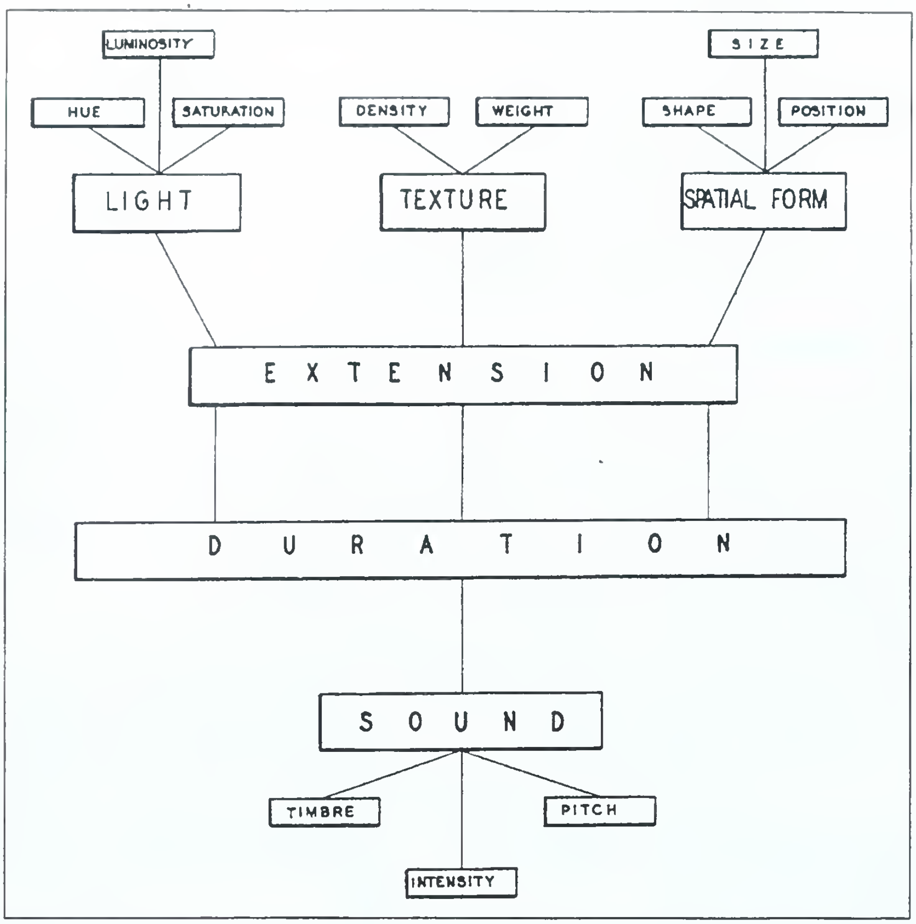


Figure 26. Components of a Combined Kinetic Art Form From [SCHI48].

> reached by either picture or sound alone. ...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. [RYDS94]

Realizing this important relationship between visuals and sound in film, Lipscomb and Kendall [LIPS90] [LIPS94] investigated the perceptual judgement of the relationship between musical and visual components in film. In their experiments, they took various motion picture sequences and manipulated their soundtracks. The motion picture sequence containing the original soundtrack along with the motion picture sequence containing various manipulated soundtracks were presented to subjects. The task of the subject was to select the soundtrack that best fit the visuals of the film. Interestingly, the results indicated that "the composer-intended musical score [the original score] was identified as the best fit by the majority of subjects for all conditions" [LIPS94]. In a related experiment, they also found significant results strongly suggesting that a musical soundtrack can in fact change the perceived *meaning* of a film presentation.

E. AUDITORY-VISUAL CROSS-MODAL MATCHING

Cross-modal matching is using information obtained through one sensory modality to make a judgment about an equivalent stimulus from another modality. Lawrence Marks has been studying auditory-visual cross-modal matching over the last twenty-five years. He has conducted several experiments which suggest a strong auditory-visual cross-modal matching among brightness, pitch, and loudness. In 1974 [MARK74], he had subjects match pure tones to the brightness of gray surfaces. His results indicated that most subjects matched increasing auditory pitch to increasing visual brightness. Marks further concludes that his findings "...mimic those of synesthesia..." [MARK74] (see SYNESTHESIA, Chapter II, Section H). In 1982 [MARK82], Marks conducted a series of four expériments in which subjects used scales of loudness, pitch, and brightness to evaluate the meanings of various auditory-visual synesthetic metaphors such as: *sound of sunset*, *murmur of dawn*, and *bright whisper* to name a few. He found that loudness and pitch expressed themselves metaphorically as greater brightness, and likewise, that brightness expressed itself metaphorically as greater loudness and as higher pitch. This series of experiments led Marks to believe that:

> The ways that people evaluate synesthetic metaphors emulate the characteristics of synesthetic perception, thereby suggesting that synesthesia in perception and synesthesia in language both may emulate from the same source -- from a phenomenological similarity in the makeup of sensory experiences of different modalities. [MARK82]

Marks has also conducted experiments involving auditory-visual cross-modal perception of intensity [MARK86], auditory-visual cross-modal similarities in *speeded discrimination* [MARK87], and additional experiments concerning auditory-visual cross-modal similarities with pitch, loudness, and brightness [MARK89]. The results of these experiments are similar to his earlier experiments and provide more evidence to support strong auditory-visual cross-modal matching among pitch, loudness, and brightness. In terms of cross-modal matching, one might conclude from Marks' findings that our senses are integrated somehow. However, Stein and Meredith offer a different point of view based on a neurological perspective:

> While cross-modal matching is clearly an intersensory phenomenon, and may involve multisensory neurons, one could make the case that it has little to do with the integration of inputs from different modalities per se, and that multisensory areas of the brain need not play any special role in this process. The judgments of equivalence across modalities could depend on the individual inputs being held in the central nervous system in modality-specific form, so that they are independent of one another but still may be accessed by another neural pool. [STEI93]

F. VISUAL DOMINANCE OVER AUDITION

1. Ventriloquism Effect

A well-known auditory-visual intersensory phenomenon is that of the *Ventriloquism Effect* (see [HOWA66]). As the name implies, this phenomenon refers to the illusion created by a skilled ventriloquist when we think we hear the dummy talking, when in fact we are actually hearing the altered voice of the ventriloquist. Not only do we hear the dummy talking but we actually think the sounds of the dummy are emanating from the dummy's mouth and not from the ventriloquist even though we know that the dummy cannot really talk as depicted in Figure 27.This effect demonstrates the strong spatial coupling that occurs between the auditory and visual senses, and as a result has been the topic of much research (see [HOWA66] [PICK69] [BERM76] [RADE76] [WARR81] [RAGO88] [STEI93]). One reason why the ventriloquism effect occurs is that the visual sense is usually the dominant sense as discussed earlier in Visual Dominance (Chapter II, Section E). As a result, "...unless there are dramatic differences in the intensities of different stimuli, the visual effect on the information generated in most other sensory systems is greater than their effect on visual perception" [STEI93]. Therefore:

> ...if visual stimuli are appearing at the same frequency and providing information of the same general type or importance as auditory or proprioceptive stimuli, biases toward

Figure 27. The Ventriloquist From [STEI93].

the visual source at the expense of the other two [auditory and proprioceptive] will be expected [WICK92].

2. Experimental Results Supporting the Ventriloquism Effect

Radeau and Bertelson [RADE76] conducted an experiment on the effect of a textured visual field on modality dominance during the ventriloquism effect. The results indicated that "...visual texture affects the degree of auditory capture of vision, but not the degree of visual capture of audition..." [RADE76]. Bermant and Welch [BERM76] investigated the effect of degree of separation of an audio-visual stimulus and eye position upon the spatial interaction of the ventriloquism effect. One of the more interesting results of this study was that "...the ventriloquism effect is not dependent on the use of a visual source which has been experimentally associated with the production of sounds" [BERM76]. The role of auditory-visual *compellingness* in the ventriloquism effect was studied by Warren et al.[WARR81] where it was found that given a highly compelling stimulus situation, "...subjects showed a very high visual bias of audition, a significant auditory bias of vision, and a sum of bias effects that indicated that their perception was fully consonant with the assumption of a single perceptual event" [WARR81]. Ragot et al. [RAGO88] explored auditory and visual ventriloquism reciprocal effects. Their findings suggested that "...visual dominance appears when attention is divided between visual and auditory modalities, but seems to be absent...when the subjects are asked to attend to one modality while knowing the other" [RAGO88]. Knudsen and Brainard [KNUD95] present neurological evidence from studying the optic tectum (also referred to as the superior colliculus). This evidence explains the ventriloquism effect supporting visual dominance over audition. They conclude that:

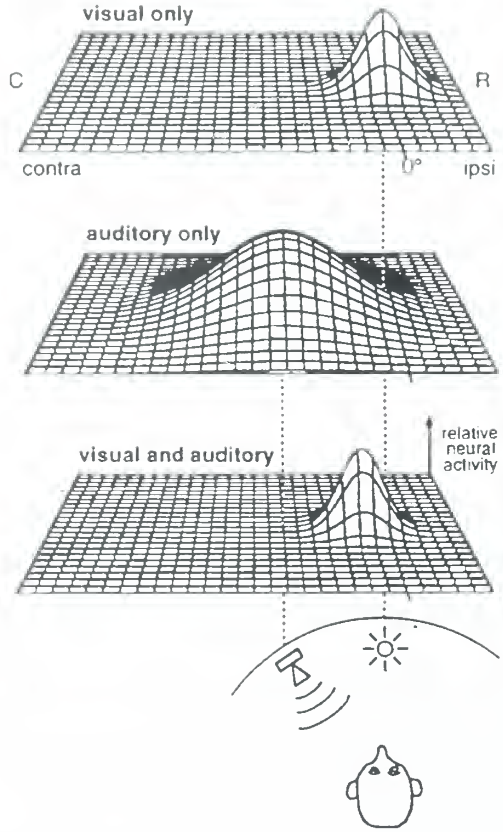
> The angular [spatial] distance that can separate visual and auditory stimuli and still result in facilitatory interactions in tectal neurons depends on the sizes of their visual and auditory receptive fields. Because visual receptive fields are consistently smaller than auditory receptive fields,...bimodal tectal neurons are more sensitive to displacements of a visual stimulus from its optimal location than to displacements of an auditory stimulus. As a consequence, the site in the bimodal tectal map that is activated by visual and auditory stimuli should be more sensitive to the location of the visual stimulus than to the location of the auditory stimulus. [KNUD95]

Knudsen and Brainard believe that the behavioral correlates of this neurological evidence support increased sensitivity and localization activity when stimuli contain both visual and auditory components. Figure 28 depicts the hypothetical neural representations on the tectal surface that occur with spatially separate auditory and visual stimuli.

3. Auditory-Visual Divided Attention Experimental Findings

During signal detection (temporal in nature and typically associated with sustained attention or vigilance), the auditory channel proves dominant over the visual channel, which is why warning signals are typically produced with auditory devices. (see APPENDIX B. AUDITORY-VISUAL CROSS-MODAL SIGNAL DETECTION AND VIGILANCE BIBLIOGRAPHY.) However, in most other areas, our visual sense dominates the hearing sense as can be seen from the following experimental findings.

In 1954, the United States Air Force released an extensive technical report which compared the visual and auditory senses as channels for data presentation during cockpit crew coordination [HENN54]. As mentioned in this report:



Hypothetical neural representations of spatially separate visual and auditory stimuli (*bottom*), schematically illustrated on a plane representing the tectal surface. The relative activity of different tectal loci is indicated by the relative height above the plane. Neurons located outside of the zones of excited neurons are inhibited (not shown) by the stimulus. Top: A frontal visual stimulus results in a sharp peak of activity centered in the rostral (R) tectum. Middle: An auditory stimulus located more peripherally results in a peak of activity centered further caudal (C) in the tectum. The peak is broader because auditory receptive fields are much larger than visual receptive fields. Bottom: The combination of visual and auditory stimuli results in a single peak of activity located between the peaks for the unimodal stimuli but biased towards the location at which the visual stimulus was represented.

Figure 28. Hypothetical Neural Representation of Auditory and Visual Stimuli on the Tectal Surface From [KNUD95].

> The evidence seems to indicate that when a person is required to divide his attention or to shift back and forth between two tasks, one visually controlled, the other aurally controlled, either task can be made a "priority" task at the expense of the other. Sense channel as such does not determine this priority.

One of conclusions of this report indicated that there was little experimental evidence comparing audition and vision as channels for data presentation. The Air Force found that

"The majority of the studies have been concerned with receptor processes and sensory thresholds rather than with perceptual phenomena" [HENN54]. Ultimately, the Air Force recognized:

> ...the many practical difficulties that have stood in the way of directly comparing these two sense modalities [audition and vision] in the experimental laboratory. It has not thus far been possible to establish common dimensions along which to locate comparable visual and auditory stimuli. Furthermore, different psychophysical procedures must frequently be employed in comparing the two modalities (largely because of the temporal-sequential character of auditory stimuli). As a consequence, it is not possible to compare directly auditory and visual judgments with broad generality and high degree of practicability. [HENN54]

Francis Colavita [COLI74] describes a series of experiments exploring *sensory dominance* in which subjects responded to suprathreshold auditory and visual stimuli. The auditory stimuli consisted of tones and the visual stimuli consisted of light flashes. The stimuli were randomly presented as auditory-only, visual-only, and combined auditory-visual. The subject's task was to identify which stimuli occurred. When subjects were presented with the combined auditory-visual stimuli, the subjects typically only responded that a visual light flash occurred, and usually did not even notice that an auditory stimuli (tone) was present. Thus, in this task, the findings suggest visual dominance over the auditory sense.

In a study investigating the perceived duration of auditory and visual intervals, Behar et al. [BEHA74], found that auditory intervals (white noise) were consistently judged to be about 20% longer than visual intervals (light from a neon glow-lamp) of the same duration. This finding "...calls attention to the contribution of peripheral variables and indicates that they must not be ignored in accounting for psychophysical judgments" [BEHA74].

Burrows and Solomon [BURR75] conducted an experiment investigating the ability to scan auditory and visual information in parallel. Subjects were presented with pairs of letters, one being a visually presented letter and the other being an aurally presented letter. The pairs of letters were presented simultaneously or sequentially. The subjects' efficiency of memory retrieval was measured in both conditions: 1) simultaneously presented letters or 2) sequentially presented letters. Their results indicated that:

> Parallel scanning is possible with a simultaneous presentation but not with sequential presentation. In retrospect, this is not surprising. The simultaneous condition provides the opportunity for two, modality specific, continuous records of the auditory and visual stimuli, unbroken by switches to another modality. In the sequential condition, the record for each modality must contain "dead time" whenever a switch to the other mode of presentation takes place. [BURR75]

Egeth and Sager [EGET77] explored the locus of visual dominance over audition in which subjects responded to suprathreshold stimuli consisting of an audio-only tone, a visual-only light flash, and a combined auditory-visual tone-light flash. Their findings suggest that:

> ...sensory or perceptual processing of the [auditory] tone is not affected by the light, i.e., that visual dominance is nonsensory in locus and depends on the relevance of the [visual] light stimulus. This interpretation was reinforced by other findings which showed that the degree of visual dominance was sensitive to the probability of light, tone, and light-plus-tone trials and to instructions to attend to a specific modality, but was not sensitive to the intensity of the light. [EGET77]

Jones and Kabanoff [JONE75] conducted an experiment to determine if eye movements are a factor in auditory localization. Jones and Kabanoff based this research on the hypothesis that "...intersensory effects depend upon anatomical linkages of the different sensory areas via the motor cortex, which may serve to integrate neural activity by sampling the state of the different sensory receptors" [JONE75]. They found that auditory localization accuracy is increased if the subject moves his eyes in the direction of the intended target. Their findings suggest that "...voluntary eye movement rather than a visual map is likely to provide the framework for spatial judgments" [JONE75].

McGurk and MacDonald [MCGU76] investigated the effect of seeing certain lip movements associated with hearing contradictory speech sounds. Subjects were presented auditory-only speech sounds and mismatched auditory-visual (speech-lip movements) combinations. Their results were remarkable. During the combined auditory-visual mismatches, most subjects were convinced they were hearing what they were seeing (lip movements), when in fact the lip movements were not the correct lip movements for the associated speech sound that they were hearing. Furthermore, even if one has prior knowledge of the auditory-visual mismatches, it does not preclude one from being convinced they were hearing what they were seeing (incorrectly). The results of this experiment were so strong that it is commonly referred to as the *McGurk Effect*. It is interesting to note that "...the sight of lip movement actually modifies activity in the auditory cortex. By whatever mechanisms the visual cue actually enhances the processing of auditory inputs, it is the functional equivalent of altering the signal-to-noise ratio of the auditory stimulus by 15-20 decibels..." [STEI93].

Rosenblum and Fowler [ROSE91] investigated if loudness judgements of speech are more closely related to the visual degree of exerted vocal effort than to the actual emitted acoustical properties of intensity. As in the McGurk Effect, subjects were presented conflicting audio-visual stimuli. Their findings suggest that when making loudness judgements of speech, the visual cues of vocal effort significantly outweigh the cues provided by the appropriate levels of acoustic intensity.

Massaro and Warner [MASS77] conducted an experiment which investigated divided attention between auditory and visual perception. In their experiment, subjects were asked to recognize test tones and test letters under selective and divided attention. They concluded that "...the degree of capacity limitations and attentional control during visual and auditory perception is small but significant" [MASS77].

Hanson [HANS81] conducted an experiment to investigate if common processing of semantic, phonological, and physical systems were involved during reading and listening. Subjects were simultaneously presented two words, one visually and one aurally, but were instructed to attend to only one modality and to make responses based on that attended modality. Her results indicated that the unattended words had an influence on semantic and phonological decisions, but had no influence on the physical task. (In the physical task, the visual words were presented in either small or capital letters and the aural words were presented in either a male or female voice.) Hanson concludes that the written and spoken words "share semantic and phonological processing but have separate modality-specific codes that operate on information prior to the convergence of information from visual and auditory inputs" [HANS81].

G. AUDITORY-VISUAL THRESHOLD PERCEPTION

The body of evidence presented thus far clearly indicates that under certain conditions, auditory-visual perceptual phenomena do exist. In fact, most auditory-visual research has focused on threshold levels, absolute sensitivity, or just-noticeable-differences (JND). Gilbert [GILB41] and Ryan [RYAN40] independently conducted exhaustive literature surveys covering these topics and a summary of their findings was presented earlier in Sensory Interaction (Chapter II, Section C). Additional evidence supporting auditory-visual perceptual phenomena from threshold level stimuli can be found in the following references: [SERR35] [PRAT36] [LOND54] [THOM58] [LOVE70]. Nevertheless, for a better understanding of this type of research, the findings of two experiments are presented showing auditory-visual perceptual phenomena from threshold-level stimuli.

An example of the research reviewed by Gilbert and Ryan is that of Kravkov [KRAV36], one of the early pioneers in the area of intersensory experimentation. Kravkov's experiment investigated the influence of sound upon the light and color sensitivity of the eye. In this experiment three female subjects were presented an auditory stimulus consisting of a 2100 Hz tone at 100 decibels for a duration of about 10 minutes. During these 10 minutes, measurements were made of color and light sensitivity. The results are as follows:

> 1. The rod sensibility of the eye decreases under the influence of simultaneous sound.
>
> 2. The colour sensibility of the eye changes differently under the influence of sound, according to the wavelength of the stimulating light. ...Whereas the colour sensibility for green rises during the acoustic stimulation the colour sensibility for orange-red decreases. [KRAV36]

In 1952, Gregg and Brogden [GREG52] conducted an experiment on the effect of simultaneous visual stimulation on absolute auditory sensitivity. In their experiment subjects were presented an auditory tone along with an auxiliary light source. Their results indicate that when subjects were asked to report the presence of a visual light source along with an auditory tone, the light stimulus decreased subject sensitivity to a 1000 Hz tone. However, when subjects were only required to report the presence of an auditory tone, the light stimulus increased sensitivity to the auditory tone.

H. AUDITORY-VISUAL SUPRATHRESHOLD PERCEPTION

This section presents the motivation and findings of those experiments in which suprathreshold auditory stimuli influenced visual perceptual quality, fidelity, or resolution; and/or suprathreshold visual stimuli influenced auditory perceptual quality, fidelity, or resolution. These experimental findings are of primary interest and directly support the motivation for this dissertation.

1. Motivation

When one talks about the using both audio and visual displays for some kind of simulation, game, VE, etc., some people will say that the use of high quality sound positively influences their perception of the visual images. For example, Brenda Laurel states that: "...in the game business we discovered that really high-quality audio will actually make people tell you that the games have better pictures, but really good pictures will not make audio sound better; in fact, they make audio sound worse" [TIER93]. Why is this? The reason is probably because simulations, games, VEs, etc., all started out as having only visuals, and then added sounds later. The addition of the sounds, then, adds to the overall perception of the experience. As a result, the visuals appear better. It is also interesting to note that the reverse is usually never reported, that the use of high-quality visual images positively influences perception of auditory displays. Why is this? Again, the answer is probably because we are used to games based on the visual displays. However, if games started out as audio only and then added visuals later, then perhaps, the addition of high-quality visual displays might positively influence subject perception of the visual images. Unfortunately, few examples exist to help analyze this hypothesis.

As described earlier in Sensory Interaction (Chapter II, Section C), there are various theories about sensory interaction. In terms of auditory-visual sensory interaction, in particular, studies of infants have revealed evidence that there exists a:

> ...spatially organized, functional relation between auditory and oculomotor systems from birth. This coordination may be enhanced by intrinsic spatial properties of the visual system that act to ensure auditory and visual colocation. Such a functional relation might in turn facilitate the detection of intermodal equivalence, since sounds are usually accompanied by sights. [BUTT81]

Stein and Meredith theorize that "combinations of, for example, visual and auditory cues can enhance one another and can also eliminate any ambiguity that might occur when cues from only one modality are available" [STEI93]. Murch believes that "under many conditions the encoding of strictly visual material or strictly auditory material involves the use of short-term storage of both systems" [MURC73]. Since auditory and visual displays can influence each other, then as Durand Begault suggests, "...another solution for improving the immersivity and perceived quality of a visual display and the virtual simulation in general is to focus on other perceptual senses -- in particular, sound" [BEGA94]. For example, Negroponte recounts the following story of designing military tank simulators:

> In the design of military tank trainers, considerable effort was made to have the highest achievable display quality (at almost any cost), so that looking at the display was as close to looking out the window of a tank as possible. Fine. Only after painstaking endeavors to keep increasing the number of scan lines did the designers think to introduce an inexpensive motion platform that vibrated a little. By further including some additional sensory effects -- tank motor and trend sounds -- so much realism was achieved that the designers were then able to reduce the number of scan lines; they nonetheless exceeded the requirement that the system look and feel real. [NEGR95]

However, the empirical evidence supporting how auditory and visual displays can influence the quality perception of each other is lacking. One reason for the lack of empirical evidence is that "...the first problem in comparing vision and hearing is of specifying perceptually relevant dimensions for both modalities, a problem which still resists truly satisfactory solution" [JONE81]. Nevertheless, after an exhaustive literature review, the following experiments present the only findings in which auditory displays influenced the quality perception of visual displays or visual displays influenced the quality perception of auditory displays.

2. Experimental Results

W. Russell Neuman [NEUM90] [NEUM91] conducted an experiment to measure the effect of changes in audio quality on visual perception on High-Definition Television (HDTV). The experimental design was to keep the quality of the visual stimuli constant, while only manipulating the auditory stimuli. The auditory conditions were as follows: low fidelity (very low-quality speaker system) vs. high fidelity (very high-quality speaker system); monaural vs. stereo; and three types of television programming: sports, situation comedy, and action-adventure. Subjects were presented a short video clip along with one of the auditory conditions. The subjects were then asked to rate 1) their liking, 2) their level of interest, 3) their psychological involvement in the programming, 4) picture quality, and 5) audio quality. Their results indicated that subjects "...had a difficult time distinguishing mono from stereo and even low-fidelity from high-fidelity sound. ...[and] video with better quality and stereo sound were consistently rated as more likable, interesting, and involving" [NEUM91]. Perhaps the most interesting finding was that a few subjects perceived an increase in visual quality when coupled with better audio even though the visual quality remained constant throughout the experiment. This finding, however, was not statistically significant and it only occurred in one of the three presented types of television programming.

Iwamiya [IWAM92] investigated 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. The factors used to evaluate the impression of both audio and visual images were: *tightness*, *evaluation*, *brightness*, *uniqueness*, and *cleanness*. "These factors are considered to be the intermodalities between auditory and visual processing" [IWAM92]. Iwamiya found that the factors of brightness, tightness, and cleanness of the auditory images enhanced the perception of brightness, tightness, and cleanness of the visual images. Iwamiya concludes that: "The better the matching of sound and image, the higher the evaluation of auditory and visual impression. This kind of *synagetic* interaction is controlled by the feedback loop from the total integrated impression of auditory in visual information." [IWAM92]

Hollier and Voelcker [HOLL97] conducted an experiment investigating the influence of video quality on audio perception. Thirty-two subjects watched video clips 10 seconds in duration with supporting audio (speech) commentaries. In total there were eight video quality variations and four audio quality variations. Their results indicated that 1) when no video was present, the perceived audio quality was always worse than if video was present, and 2) although only small differences were noted, a decrease in video quality corresponded to a decrease in perceived audio quality. They ultimately propose an algorithmic approach for the proper development of an auditory-visual cross-modal perceptual model depicted in Figure 29. In their final discussion of the experiment,

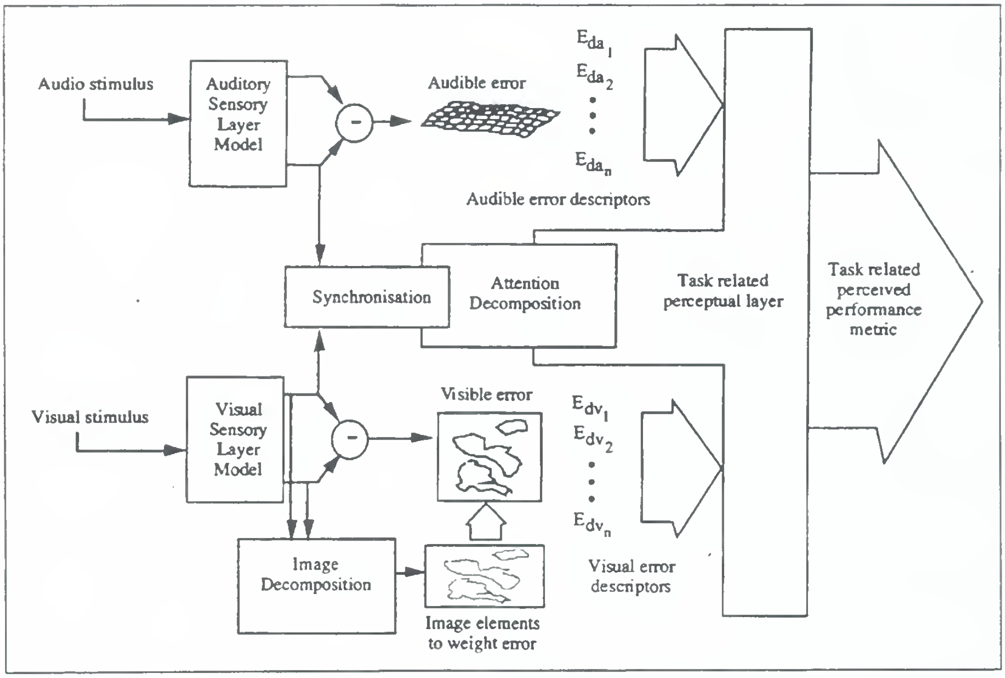


Figure 29. Auditory-Visual Perceptual Model From [HOLL97].

Hollier and Voelcker state that "for a majority of applications both in the communications and entertainment industry separate evaluation of audio or video quality is likely to become of limited value" [HOLL97].

Two companion papers by Woszczyk et al. [WOSZ95] and Bech et al. [BECH95] discuss the design and results of an experimental procedure examining the interaction between the auditory and visual modalities in the context of a home theater system. Their approach acknowledges that "...experiments involving both modalities require a novel approach that recognizes domains of cooperative interaction between the senses" [WOSZ95]. With the growing interest and development of virtual reality systems, Woszczyk identifies the need for testing the interaction of audio and visual displays in order to bring about "substantial improvements in the integration of various audio and video parts of these [virtual reality] systems, and thereby provide important perceptual benefits that enhance [the] audio-visual experience of the viewers" [WOSZ95]. The testing of audio-visual interaction is critical because "Auditory and visual channels work both independently and in mutual cooperation on both cognitive and sensory levels of perception," [WOSZ95]. In order to study the interaction between the audio and visual sensory modalities "it is necessary to focus on the total experience and not on the two modalities individually" [BECH95], which supports Woszczyk et al.'s observations that "The matching of auditory and visual data triggers perceptual synergy between modalities and promotes intermodal fusion" [WOSZ95]. In their experiments, subjects assessed audio-visual reproductions using the subjective dimensions of *action*, *space*, *mood*, and *motion* while asking specific questions focusing on *quality*, *magnitude*, *degree of involvement*, and *audio-visual balance*. Quality was defined as: *distinctness*, *clarity*, and *detail of the impression*. One of their findings, of particular interest is that both visual and audio perceived quality increased with increasing screen size. To further explore auditory-visual interaction, Bech conducted two more experiments to investigate the influence of stereophonic (audio) width on the perceived quality of an audio-visual presentation using multichannel surround sound systems. During the experiments, the subjects were asked to evaluate the quality (fidelity) of the spatial information contained in audio-visual reproductions. The results indicate that "the quality of [perceived] spatial reproduction increases linearly with an increase in the stereophonic [audio] width" [BECH97].

Hugonnet [HUGO97] presents what he considers to be a new concept of spatial coherence between sound and picture in stereophonic TV production. "From a cultural and historical point of view, our perception of sound corresponding to image has remained monophonic" [HUGO97]. As such, Hugonnet describes methods of production and post-production to achieve spatial coherence of stereo sound with various TV content including: talk shows with two people, talk shows with more than two people, concerts, sports, and drama. He found that when people are first exposed to stereo sound when watching TV, people found the relation between visual and auditory images strange and not very comfortable. However, once people became accustomed to the stereo sound, if they were re-exposed to mono sound, they perceived the quality of the mono sound to be of lower sound quality. Hugonnet concludes by recognizing the importance of auditory-visual interaction and states: "It is up to us to bring about a radical change in audiovisual perception, where sound will gain its right place, on a par with the visual image" [HUGO97].

I. SUMMARY

In summary, this chapter has provided an overview of Virtual Environments, Auditory-Visual Perceptual Organization, Auditory-Visual Art Forms and Film, Auditory-Visual Cross-Modal Matching, Visual Dominance over Audition, Auditory-Visual Threshold Perception, and Auditory-Visual Suprathreshold Perception.

IV. EXPERIMENTAL DESIGN OVERVIEW

A. INTRODUCTION

This chapter describes the motivation and initial considerations that led to the development of the experimental design used to gather empirical evidence supporting suprathreshold auditory-visual cross-modal quality perception phenomena. The various considerations outlined in this chapter were instrumental in developing the experimental design of the pilot study which ultimately led to the three main experiments forming the foundation of this dissertation. The experimental design details of the pilot study and three main experiments are described in greater detail in the next four chapters. Thus, the intent of this chapter is not to focus on details, but rather to provide an overview of the choices that were considered during the initial experimental design development.

B. MOTIVATION

Based on the findings from the exhaustive background and literature review outlined in the previous two chapters, the following are some key observations:

1) There is neurological and physiological evidence supporting auditory-visual cross-modal perception phenomena.

2) There is psychological and psychophysical evidence supporting auditory-visual cross-modal perception phenomena.

3) There is empirical evidence supporting the ability to divide attention between audition and vision.

4) There is empirical evidence suggesting that sound can influence the perceived mood of motion pictures.

5) There is empirical evidence supporting auditory-visual cross-modal perception phenomena concerning increased sensitivity/acuity in audition and/or vision.

6) There is a need to enhance multimedia and VE development through better understanding of auditory-visual cross-modal perception phenomena.

7) There is a lack of empirical evidence supporting auditory-visual cross-modal perception phenomena in which suprathreshold auditory stimuli influenced visual perceptual quality and suprathreshold visual stimuli influenced auditory perceptual quality.

Based on these key observations, which stem from wide-ranging interdisciplinary research, there is a need for empirical evidence supporting suprathreshold auditory-visual cross-modal quality perception phenomena. The ultimate goal of this dissertation answers the following question: In an audio-visual display, what affect (if any) do various audio quality levels have on the perception of visual quality and various visual quality levels have on the perception of auditory quality? The following are some specific derivations of this question:

1) Are changes in the audio and/or visual qualities of an audio-visual display perceivable and can these changes be attended to also?

2) 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?

3) 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) Does a low-quality auditory display coupled with a low-quality visual display cause a decrease/increase 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?

5) Does a high-quality auditory display coupled with a high-quality visual display cause an increase/decrease 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?

In order to answer these questions concerning auditory-visual perceptual phenomena, the approach taken was to conduct an experiment to facilitate measuring responses to various auditory-visual suprathreshold stimuli. The overall design of the experiment consists of three main portions: 1) visual-only displays, 2) auditory-only displays, and 3) combined auditory-visual displays. During the visual-only portion, subjects are presented visual displays and are then asked to rate their visual quality. During the auditory-only portion, subjects are presented auditory displays and are then asked to rate their auditory quality. During the combined auditory-visual portion, subjects are presented combined auditory-visual displays, and are then asked to rate the quality of both the auditory portion and visual portion of the combined auditory-visual display. The goal is to compare the subject's quality ratings made during the visual-only and auditory-only portions with the subject's visual and auditory quality ratings made during the combined auditory-visual portion. The results of this comparison are analyzed to answer the questions of interest, and as such are the quintessential contribution of this dissertation. The initial design considerations of this experiment are now presented.

C. DESIGN CONSIDERATIONS

1. Software and Hardware

The first key consideration in the experimental design is that the experiment be automated. The goal is to create a computer program that can render visual-only, auditory-only, and combined auditory-visual displays while also capturing the required responses of the subject. An automated experiment is chosen since it helps to produce identical testing conditions, thereby reducing any potential confounds (i.e., confounding factors) that might arise through human error. Keeping in mind the self-imposed limitations described earlier in LIMITATIONS (Chapter I, Section E), the software chosen for the experiment consisted of HTML, Java, JavaScript, and VRML (all freely downloadable). The basic idea is to have the entire experiment contained within an HTML browser window as depicted in Figure 30. The visual-only, auditory-only, and combined auditory-visual displays could then be rendered via JavaScript and/or VRML

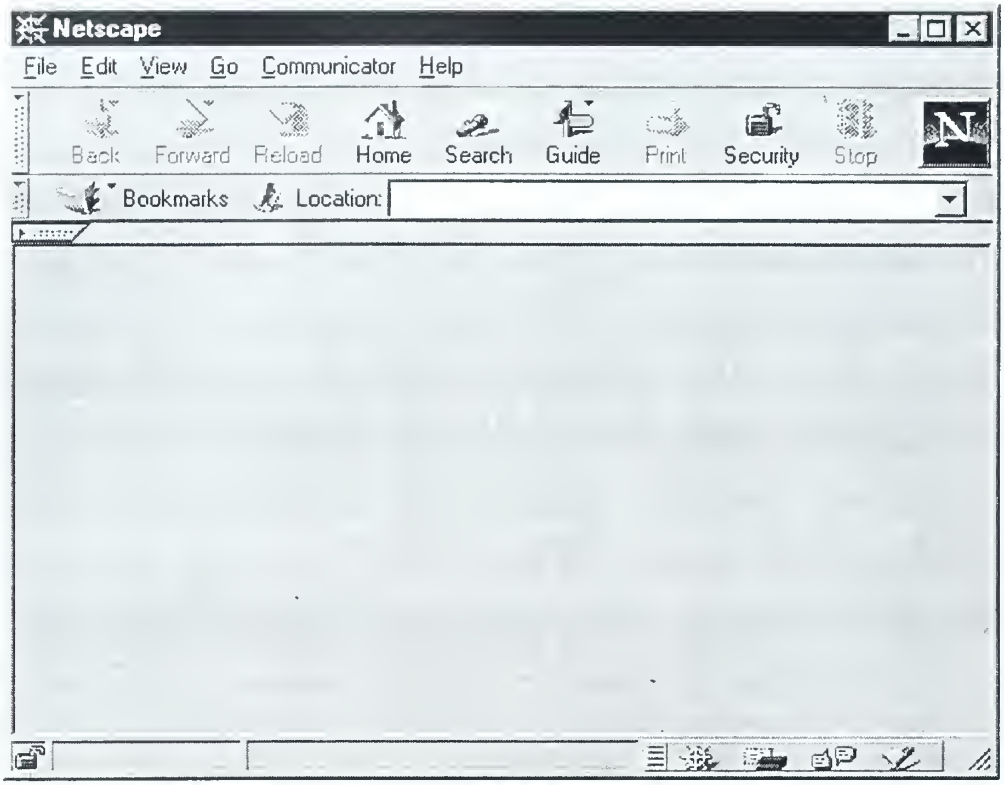


Figure 30. Netscape HTML Browser Window.

within the main HTML window. The subjects' responses are then obtained with rating scales using Java pop-up windows as depicted in Figure 31. Furthermore, based on the

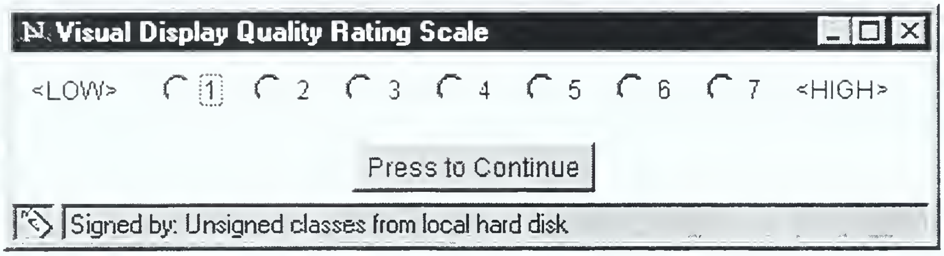


Figure 31. Java Pop-up Visual Display Rating Scale.

software utilized, and keeping in mind the limitations of this dissertation, a personal computer (PC) was used for all experiments. The specifics of the software and hardware used are explained in greater detail during the description of the pilot study and the three main experiments in subsequent chapters.

2. Visual Displays

Important considerations in the development of this experiment include choosing the rendering, type/content, and quality manipulation parameters of the visual displays. The possible rendering choices of the visual displays considered were: 17-inch computer monitor, 20/21-inch computer monitor, 28-inch computer monitor, large screen TV, and triple large-screen TVs. Because of fidelity considerations and amount of available controlled laboratory space, the TVs were not utilized. The high cost of the 28-inch monitor precluded its use, and the 17-inch monitor proved to be too small. As a result, a 20-inch computer monitor was selected to render all the visual displays.

Choosing the type and content of the visual display was perhaps the most difficult task during the development of the experiment. Possible types of visual displays considered included: static (still image) or dynamic (motion video, user controlled navigation in 2D space, or user controlled navigation in 3D space). To reduce the excessive computational requirements of motion video, to reduce frame rate synchronization errors with associated auditory displays, and to reduce user-computer interaction training and variations associated with user controlled navigation, static images were chosen as the display type. Once the decision was made to use static visual displays, the next difficult task was to choose the content. After considering numerous possibilities, two visual displays were chosen: 1) a radio and 2) scene depicting a bowl of fruit and flowers. Figure 32 and Figure 33 depict (in color) the radio and fruit-flower scene respectively. The rationale for the choice of content of these displays will be explained in greater detail during the description of the pilot study and three main experiments in subsequent chapters.

Once the choice of rendering and type/content of the visual displays were determined, the quality-manipulation parameters were selected. Since the results of this

Figure 32. Color Visual Display of Radio.

Figure 33. Color Visual Display of Fruit-Flower Scene.

research effort will hopefully benefit multimedia and VE development, pixel resolution and noise level were chosen as the quality parameters to be manipulated. Selecting pixel resolution is perhaps the most prevalent decision in creating visual scenes for any VE. Increasing pixel resolution corresponds to an increase in realism at the expense of 1) an increase in rendering time, 2) an increase in storage requirements, and 3) an increase in download time (if networked). Thus, the VE developer must carefully consider the amount of required pixel resolution. Noise level, the other parameter, was chosen based on similar considerations as pixel resolution when one considers quality levels of MPEG video. High-quality MPEG video has a greater signal-to-noise ratio than low-quality MPEG video. Thus, a lower-quality visual image will have a greater noise level than that of a higher quality image. Another factor for using noise level was based on the visual display's eventual coupling with an auditory display which is explained in the next section. A final consideration in the choice of visual displays was the ability to produce the various required quality levels. For example, if a potential quality metric cannot be produced due to software or hardware constraints, then that quality metric is not feasible. Since Adobe *Photoshop* [ADOB98] was utilized, its capabilities provided the limits of possible quality parameter manipulation. As such, all the visual displays used throughout all the experiments were developed using Adobe *Photoshop*.

3. Auditory Displays

Equally important considerations in the development of this experiment were choosing the fidelity, rendering, content, and quality manipulation parameters of the auditory displays. The possible fidelity choices of the auditory displays considered were: monophonic, stereophonic, and spatialized. The rendering possibilities of the auditory displays considered were: headphones, left and right small-computer speakers, left and right high-fidelity speakers, quad configuration of high-fidelity speakers, and surround-sound configuration of high-fidelity speakers. In order to minimize any potential experimental confounds due to varying room acoustics, headphones were chosen to render the auditory displays. Similarly, to minimize any unforeseen confounds from using stereophonic or spatialized sound, monophonic fidelity was chosen for all auditory displays. Another factor for choosing monophonic audio fidelity was due to the static nature of the visual displays. Once the decision was made to use monophonic auditory displays, the next difficult task was to choose the content. After numerous possibilities, a music sound was chosen as the content of the auditory displays. The rationale for using music as the content of the auditory displays will be explained in greater detail during the description of the pilot study and three main experiments in subsequent chapters. Once the choice of fidelity, rendering and content of the auditory displays were determined, the quality manipulation parameters were selected.

As stated earlier, since the results of this research effort will hopefully benefit multimedia and VE development, sampling frequency and noise level were chosen as the quality parameters to be manipulated. The choice of sampling frequency is similar to that of pixel resolution. Increasing sampling frequency corresponds to an increase in realism at the expense of 1) an increase in rendering time, 2) an increase in storage requirements, and 3) an increase in download time (if networked). Thus, the VE developer must carefully consider sampling frequencies. Noise level, the other parameter, was chosen because signal-to-noise ratio is another common quality metric of audio. The amount of noise level, specifically Gaussian noise, was also chosen because of the eventual coupling of auditory to visual displays with varying noise levels. As such, the level of Gaussian noise becomes a common quality metric between both auditory and visual displays as will be explained in greater detail during the description of the main experiments in the subsequent chapters. As with the visual displays, a final consideration in the choice of auditory displays was the ability to produce the various required quality levels. For example, if a potential quality metric cannot be produced due to software or hardware constraints, then that quality metric is not feasible. Since Sonic Foundary's *Sound Forge* software [SONI98] was utilized, its capabilities provided the limits of possible quality parameter manipulation. As such, all the auditory displays used throughout all the experiments were developed using *Sound Forge*.

4. Location and Subjects

The location for conducting all experiments was at the Naval Postgraduate School (NPS) in Monterey, California. To limit external environmental noises and to control distractions, all experiments were conducted within an isolated room (office) in which the experimenter had total control of audio and visual conditions. As such, scheduling conflicts typically associated with the main laboratory were eliminated, which greatly facilitated the process of running experiment sessions. Furthermore, since all experiments were conducted at NPS, the NPS student body provided an excellent source of engaged and attentive volunteer subjects.

5. Data Analysis

Another important consideration in the experimental design was that of the eventual data analysis process. The important factor was that the data collection format had to mesh with the data analysis process. As such, a considerable amount of time was spent deciding how to analyze the resulting data even before the data was collected. Accordingly, the chosen method of data analysis helped to derive the format of data collection. Since *StatView* [SASI98] software was chosen to do the statistical analysis of the experimental results, the data collection process was in turn automated to facilitate the ease of importing data into *StatView*.

D. DESIGN SELECTIONS

Based on the motivation and initial design considerations, a pilot study was designed to investigate the perceptual effects from manipulating visual display pixel resolution and auditory display sampling frequency. The visual display consisted of the aforementioned radio, and the auditory display was a selection music. The entire automated experiment was contained within an HTML browser window using VRML to render the visual-only, auditory-only, and combined auditory-visual displays, and using Java pop-up windows to collect subject responses. The details of the experimental design are outlined in Chapter VI. The lessons learned from this pilot study were instrumental in designing the three main experiments of this dissertation as follows: 1) Experiment 1: Static Resolution, 2) Experiment 2: Static Noise, and 3) Experiment 3: Static Resolution NonAlphanumeric. Each experiment was fully automated and contained within an HTML browser window using JavaScript to render the visual-only, auditory-only, and combined auditory-visual displays, and using Java pop-up windows to collect subject responses.

As its name implies, Experiment 1: Static Resolution is designed to investigate the perceptual effects from manipulating visual (static as opposed to dynamic) display pixel resolution and auditory display sampling frequency. The visual display consisted of the aforementioned radio, and the auditory display was a selection music. The details of the experimental design are outlined in Chapter VII.

Experiment 2: Static Noise is designed to investigate the perceptual effects from manipulating visual (static) display Gaussian noise level and auditory display Gaussian noise level. The visual display consisted of the aforementioned radio, and the auditory display was a selection music. The details of the experimental design are outlined in Chapter VIII.

Experiment 3: Static Resolution NonAlphanumeric is designed to investigate the perceptual effects from manipulating visual (static) display pixel resolution and auditory display sampling frequency. The visual display consisted of the aforementioned fruit-flower scene, and the auditory display was a selection music. The details of the experimental design are outlined in Chapter IX.

E. SOFTWARE DESIGN

In order to better understand the type of computer programming used to develop the main experimental design, a brief overview of the software design and development is now provided.

1. Overview

All software used in the development of the main experimental design is custom-designed and encapsulated into an HTML file. For each main experiment, a total of nine HTML files are developed. Each HTML file corresponds to the predetermined randomized sequence of appropriate auditory-only, visual-only, and combined auditory-visual stimuli. This randomization is based on the Latin square technique (see [GOOD95] for a description of the Latin squares technique). As such, to initiate an experiment testing session, the appropriate HTML file is simply executed. In an effort to minimize delays in rendering any of the auditory or visual stimuli, all auditory and visual displays (files) were pre-loaded into memory as the HTML file is being executed for the first time.

2. Development

The development of the overall software design of the main experiment was divided into three main components: 1) displaying instructions, 2) auditory and visual display rendering, and 3) user input.

a. Displaying Instructions

Since the experiment is to be automated, the user (subject) is presented with numerous sets of instructions. The wording of the various sets of instructions was fine-tuned throughout the pilot study in order to eliminate any possible ambiguities. All the various sets of instructions were written as separate Java *applets* which were simply embedded into the main HTML code. As such, all nine HTML files shared the same Java instruction applets. Thus, if any one set of instructions needed to be rewritten for clarity, only that one set of instructions had to be rewritten and recompiled, as opposed to rewriting the instructions in all nine HTML files. An example of the Java programming code used to produce one set of instructions is depicted in Figure 34.

```
import netscape.javascript.*;
import java.applet.*;
import java.awt.*;
import java.awt.event.*;
public class InstructionsAudioVisual extends Applet implements WindowListener,
                                        ActionListener {
  private Button EnterButton;
  private Panel EnterPanel;
  private TextArea Text;
  public JSObject win;
  public void init() {
    Text = new TextArea("\n", 9, 67, 3);
    Text.append(" (1) You will now be rating the VISUAL quality of a combined audio-visual display.\n");
    Text.append("\n");
    Text.append(" (2) A total of 9 audio-visual displays will be presented randomly.\n");
    Text.append("\n");
    Text.append(" (3) Each audio-visual display will be presented for 8 seconds.\n");
    Text.append("\n");
    Text.append(" (4) After which, you will be prompted ONLY for your VISUAL rating.\n");
    Text.append("\n");
    EnterPanel = new Panel();
    EnterPanel.setLayout(new FlowLayout(FlowLayout.CENTER));
    EnterButton = new Button("Press to Continue");
    EnterButton.addActionListener(this);
    EnterPanel.add(EnterButton);
    GridBagLayout gridbag = new GridBagLayout();
    GridBagConstraints c = new GridBagConstraints();
    setFont(new Font("Helvetica", Font.PLAIN, 14));
    setLayout(gridbag);
    c.fill = GridBagConstraints.BOTH;
    c.gridwidth = GridBagConstraints.REMAINDER; //end row
    gridbag.setConstraints(Text, c);
    add(Text);
    c.gridwidth = GridBagConstraints.REMAINDER; //end row
    gridbag.setConstraints(EnterPanel, c);
    add(EnterPanel);
    c.gridwidth = GridBagConstraints.REMAINDER; //end row
  } //end
  public void windowClosed(WindowEvent event) {
  }
  public void windowDeiconified(WindowEvent event) {
  }
  public void windowIconified(WindowEvent event) {
  }
  public void windowActivated(WindowEvent event) {
  }
  public void windowDeactivated(WindowEvent event) {
  }
  public void windowOpened(WindowEvent event) {
  }
  public void windowClosing(WindowEvent event) {
    System.gc();
  }
  public void actionPerformed(ActionEvent event) {
    Object source = event.getSource();
    if (source == EnterButton) {
      win = JSObject.getWindow(this);
      win.eval("audioVisualWrite()");
      win.eval("goToAudioVisualDisplays()");
      System.gc();
    } // end if
  } //end actionPerformed
} // end Applet
```

Figure 34. Example of Java Applet used to Render Instructions.

b. Auditory and Visual Display Rendering

All auditory and visual displays were rendered via JavaScript function calls within the main embedded HTML file. Figure 35 depicts a portion of the JavaScript programming code used to render three combined auditory-visual displays. Specifically, 1) *function HLC()* is used to render a combined high-quality auditory and low-quality visual display; 2) *function HMC()* is used to render a combined high-quality auditory and medium-quality visual display; and 3) *function HHC()* is used to render a combined high-quality auditory and high-quality visual display.

```
...
...
function HLC() {
  highWrite();
  lowWrite();
  document.highSound.play(false);
  document.images["RenderDisplays"].src = lowVisual;
  goToCombinedDisplays();
}

function HMC() {
  highWrite();
  medWrite();
  document.images["RenderDisplays"].src = medVisual;
  document.highSound.play(false);
  goToCombinedDisplays();
}

function HHC() {
  highWrite();
  highWrite();
  document.images["RenderDisplays"].src = highVisual;
  document.highSound.play(false);
  goToCombinedDisplays();
}
...
...
```

Figure 35. Example of JavaScript Function Calls.

c. User Input

All user input is accomplished via Java *Frames* which contain the appropriate rating scales.A *Frame* is basically a window which can be made to appear and disappear (i.e., a pop-up window). Figure 36 depicts a portion of the Java programming code used to render a visual-only rating scale.

```
...
...
public class RatingScalesVisualAndRatingsTest extends Frame implements WindowListener,
                                    ActionListener
                                    {
  private ShowRatingScalesVisualAndRatingsTest thisScale;
  public final static String TITLE = "Visual Display Quality Rating Scale";
  Checkbox oneV,twoV,threeV,fourV,fiveV,sixV,sevenV;
  Button EnterButton;
  private Panel VisualPanel, EnterPanel;
  public RatingScalesVisualAndRatingsTest(ShowRatingScalesVisualAndRatingsTest owner) {
    super(TITLE);
    Panel VisualPanel = new Panel();
    VisualPanel.setLayout(new FlowLayout(FlowLayout.CENTER));
    VisualPanel.add(new Label(" <LOW>"));
    CheckboxGroup VisualGroup = new CheckboxGroup();
    oneV = new Checkbox("1", VisualGroup, false);
    VisualPanel.add(oneV);
    twoV = new Checkbox("2", VisualGroup, false);
    VisualPanel.add(twoV);
    threeV = new Checkbox("3", VisualGroup, false);
    VisualPanel.add(threeV);
    fourV = new Checkbox("4", VisualGroup, false);
    VisualPanel.add(fourV);
    fiveV = new Checkbox("5", VisualGroup, false);
    VisualPanel.add(fiveV);
    sixV = new Checkbox("6", VisualGroup, false);
    VisualPanel.add(sixV);
    sevenV = new Checkbox("7", VisualGroup, false);
    VisualPanel.add(sevenV);
    VisualPanel.add(new Label("<HIGH>"));
    EnterPanel = new Panel();
    EnterPanel.setLayout(new FlowLayout(FlowLayout.CENTER));
    EnterButton = new Button("Press to Continue");
    EnterButton.addActionListener(this);
    EnterPanel.add(EnterButton);
    setLayout(new GridLayout(2, 1, 1, 3));
    add(VisualPanel);
    add(EnterPanel);
    pack();
    setLocation(180,220);
    addWindowListener(this);
    thisScale = owner;
  } //end
  public void windowClosed(WindowEvent event) {
  }
...
...
  public void windowClosing(WindowEvent event) {
    dispose();
    System.gc();
  }
  public void actionPerformed(ActionEvent event) {
    Object source = event.getSource();
    if (source == EnterButton) {
      thisScale.myReturn();
      dispose();
      System.gc();
    } // end if
  } //end actionPerformed
} // end Frame
```

Figure 36. Example of Java Frame used to Render Rating Scales.

F. SUMMARY

In summary, this chapter has provided an overview of the overall experimental design process of this research effort to include its motivation, design considerations, eventual design selections, and overall software design.

V. VISUAL AND AUDITORY DISPLAY DEVELOPMENT

A. INTRODUCTION

Given that the pilot study is designed to investigate the perceptual effects from manipulating visual-display pixel resolution and auditory-display sampling frequency, the required associated visual and auditory displays need to be created. The visual display selected for the pilot study is a radio (Chapter IV, Figure 32), 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 1) psychological factors such as Gestalt perceptual grouping theory and the Ventriloquism Effect, and 2) neurological evidence supporting auditory-visual sensory interaction, an auditory-visual display consisting of a radio and music might be perceptually grouped together thereby producing a more tightly coupled display. Furthermore, in a higher cognitive sense, we are likely to associate music (audio) with a radio (visual). The ultimate goal is for the combined auditory-visual display to be experienced as a single entity, and not as separate auditory and visual displays. The following describes the development process of the visual, auditory, and combined auditory-visual displays used in the pilot study. This development process was instrumental in the eventual experimental design of the three main experiments.

B. VISUAL-DISPLAY DEVELOPMENT

To obtain the visual image of a radio, various techniques were utilized. First, a digital camera was used to take pictures of a radio in various settings (i.e. indoors and outdoors). However, the lighting and shadowing of these digital photos proved too difficult to manage properly. To eliminate lighting and shadowing problems, the next method involved using a flatbed scanner. The radio was simply placed on the scanner, while the scanner recorded the image of the radio. This method actually produced fairly good images, but there were still minor lighting and shadowing problems. Ultimately, a photograph of a radio was taken from the book *Radios by hallicrafters with Price Guide* by Chuck Dachis [DACH95]. This book contains many professionally photographed radios. After deliberating over the many pictures, a particular radio was finally chosen. This radio image was then digitized using a flatbed scanner at 600 x 600 pixel resolution. The color version of this radio is depicted earlier in Chapter IV, Figure 32. Since the visual displays of this experiment only involve the manipulation of pixel resolution, the overall color content (impression) of the image does not change much when changing pixel resolution. As a result, for the remaining discussion of this radio, all figures will be presented in black and white. However, it is important to emphasize that during the experiment, the visual displays of the radio were all presented in color. The black and white version of this radio at 600 x 600 pixel resolution is presented in Figure 37. This

Figure 37. Visual Display of Radio at 600 pixels/inch.

particular radio was chosen because it contained many various features including: letters and numbers, smooth and rough surfaces, strait and curved lines, patterns (on the speaker), and reflections. The basis for having numerous features is to provide test subjects with a wide variety of cues from which to make their quality ratings. Incidentally, in an effort to avoid any potential copyright infringements, Chuck Dachis, the author of the book was contacted by telephone for the purpose of obtaining permission to use the photograph of the radio. Chuck Dachis gave his permission to use any photograph necessary for the experiments, and was very pleased that his photographic efforts were being used in scientific research.

Using the original scanned image at 600 pixels/inch, Adobe *Photoshop* [ADOB98] was then used to make various copies with degraded pixel resolutions all having the same dimensions, the size of which nearly fills up the display area of a 20-inch computer monitor. Approximately 30 images of the radio ranging from 200 to 600 pixels/inch were produced. The next step involved establishing levels of pixel resolution that were noticeably different, but not just-noticeably-different or obviously different. The goal was to establish low-, medium-, and high-quality visual displays for use in the experiment. An example that is obviously different is asking a subject to compare the quality between Figure 37 with Figure 38. As one can see, the difference is obvious, resulting in an inconsequential response from the subject. An example that is perhaps just-noticeably-different, is asking a subject to compare the quality between Figure 37 and Figure 39. In this case, it is fairly difficult to distinguish the quality difference between the two radios. 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 was very time consuming. Preliminary subjects were presented (using the same graphics accelerator and computer monitor chosen for the experiment as described later) about six or seven images of the radio with varying levels of pixel resolution. A subject would then be asked to arrange (if possible) the images in ascending or descending order of quality. After repeating this process with about 15 subjects, a consensus was finally reached which ultimately determined the low-, medium-

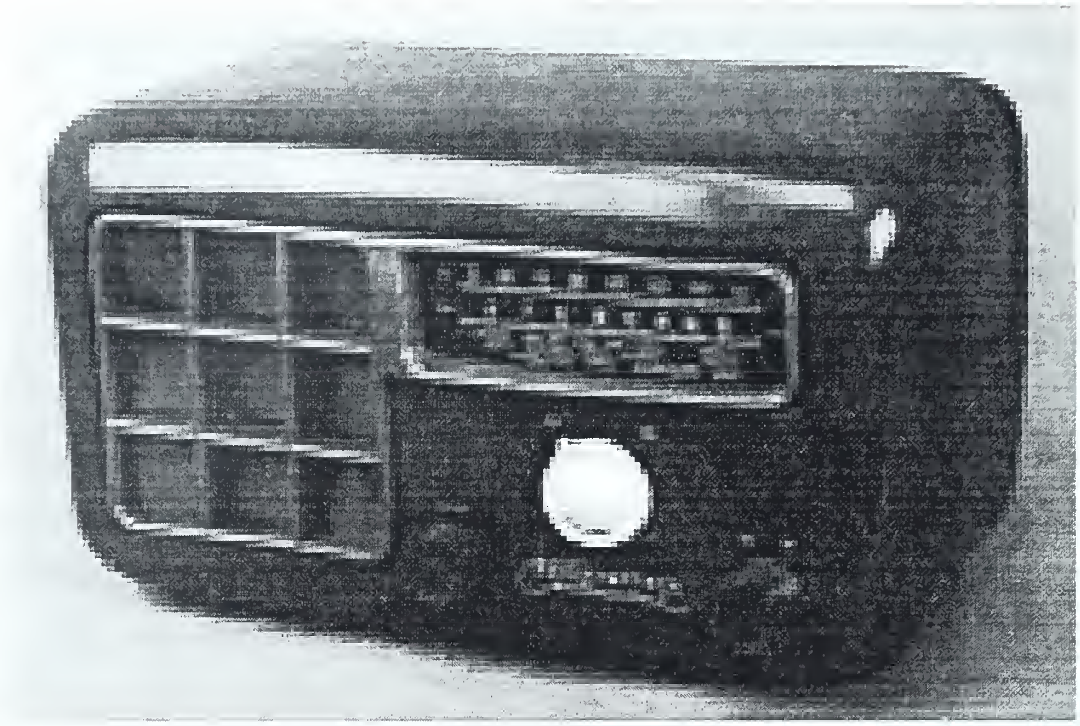


Figure 38. Obviously Different Poor-Quality Visual Display of Radio.

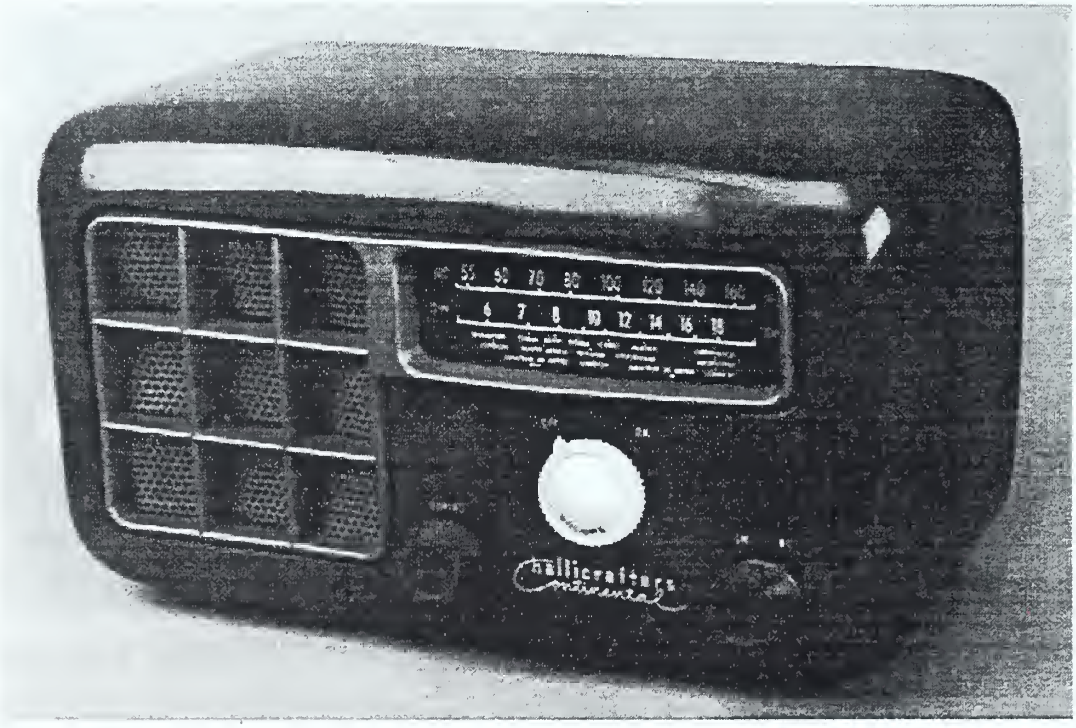


Figure 39. Just-Noticeably-Different High-Quality Visual Display of Radio.

, and high-quality visual displays of the radio to be used in the experiment. Resolutions of 425 pixels/inch, 450 pixels/inch, and 500 pixels/inch were selected as the low-, medium-, and high-quality visual displays respectively to be used in the pilot study. In general, however, the actual (absolute) pixel resolution is not important, for there are numerous factors which affect the final rendering of the visual display such as: 1) computer monitor specifications, 2) computer monitor desk size (resolution), 3) video/graphics accelerator specifications, and 4) software application graphics-rendering capabilities. An example of this last factor, in terms of the pilot study, relates to the capability of rendering textured images via the *CosmoPlayer* VRML Plugin [COSM98] to *Netscape Communicator* [NETS98]. Since the visual displays were represented as textured images in *CosmoPlayer*, the displays had to be further processed (filtered) by *CosmoPlayer*. This resulted in noticeably degraded quality in the visual displays. This fact was well known ahead of time and was incorporated into the initial development of the low-, medium-, and high-quality visual displays. As a result, the only way to actually visualize the correct representations of the low-, medium-, and high-quality displays selected, is to view them through *CosmoPlayer*. However, because the pilot study implementation was eventually abandoned, it is not possible to adequately depict the visual displays as figures to view in this dissertation. Nevertheless, the important thing is that a relative quality ordering of the visual displays was 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 also important to note that even the high-quality visual display, has some, albeit slight, pixel resolution degradation. The reason for this is based on the design of the experiment. The goal 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 comparing the one display with no perceivable pixel resolution degradation to the two displays which do have pixel resolution degradation. Thus, in order to ensure that subjects are making quality ratings based only on degree of pixel resolution (not absence or presence), the high-quality display must also have a small amount of perceivable pixel resolution degradation.

C. AUDITORY-DISPLAY DEVELOPMENT

Constructing the auditory displays was much easier than constructing the visual displays, since music can be obtained easily from any compact disc (CD). The only consideration was the musical content. Since the quality parameter to be manipulated in the pilot study is sampling frequency, a conscious decision was made not to include vocals (speech). The reason for this is because 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 focused 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 is 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. As such, to reduce the chance that subjects might have previously heard the music, an obscure portion of alternative music was selected. Another consideration in choosing the music was that the experimenter (myself) would have to listen to this piece of music for perhaps hundreds of times. So, the particular music selected was also very much liked by the experimenter (me). The music was taken from a song called *A Forest* from the CD *Mixed up* by The Cure which was produced by Elektra Entertainment Group, a division of Warner Communications Inc. In order to avoid any potential copyright infringements, a letter was written to Elektra

Records requesting to use portions of *A Forest* for scientific research. Elektra replied with an official letter granting permission to use portions of *A Forest* as long as a courtesy credit is given (see Figure 40). Thus, in accordance with Elektra's stipulation, portions of *A Forest* by The Cure, courtesy of Elektra Entertainment Group, are used in the conduct of this experiment. (Thanks Elektra.)

Using the *Mixed up* CD by The Cure, a 20 second selection of *The Forest* was recorded into Sonic Foundary's *SoundForge* [SONI98] at 44.1 kHz (sampling frequency). The portion of music selected contained cymbals (among other instruments) resulting in a very wide frequency range of sound. *SoundForge* was then used to reproduce the 44.1 kHz 20-second musical selection at numerous sampling frequencies ranging from 4 kHz to 44.1 kHz. Similar to creating the visual displays, the next step involved establishing sampling frequencies that were noticeably different, but not just-noticeably-different or obviously different. The goal was to establish low-, medium-, and high-quality auditory displays for use in the experiment. 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 was again very time consuming. Preliminary subjects were presented (using the same audio card and headphones chosen for the experiment as described later) about six or seven music selections with varying sampling frequencies. These subjects were then asked to arrange (if possible) the musical selections in ascending or descending order of quality. After repeating this process with about 15 preliminary subjects, a consensus was finally reached which ultimately determined the low-, medium-, and high-quality auditory displays of music to be used in the experiment. Sampling rates of 11 kHz, 17 kHz, and 44.1 kHz were selected as the low-, medium-, and high-quality auditory displays respectively for use in the pilot study. A consensus also established a constant volume setting for the auditory displays. Again, it is important to remember that the actual (absolute) sampling frequency is not important, for there are numerous factors which

February 13, 1998

<u>VIA FAX 408-656-2814</u>

Russell Storm
Major, US Army
Dept. of Computer Science
Naval Post Graduate School
Monterey, California 93943

Re: "The Cure/"A Forest"

Gentleperson:

This will confirm that Elektra Entertainment Group, a division of Warner Communications Inc. ("Elektra") has no objection to your use of portions of the master recording "A Forest" (the "Master") performed by The Cure ("Artist") solely for the purposes of a scientific experiment in connection with your dissertation as described in the attached facsimile dated January 30, 1998. You shall not distribute any copies of the Master.

You acknowledge that as between you and Elektra, Elektra is the exclusive owner of all rights in and to the Master for the United States and Canada, and that you will not use the Master for any purpose other than that described above. You will be responsible for obtaining any other required consents and making all required payments, and you indemnify Elektra from any claims by third parties in connection with the foregoing.

You will provide a courtesy credit as follows: "A Forest" by The Cure courtesy of "Elektra Entertainment Group".

Please confirm you acceptance of the foregoing by signing in the space below and returning this letter back to us. Your use of the Master shall constitute such acceptance.

75 Rockefeller Plaza New York NY 10019 tel 212 275 4000 fax 212 275 4001
...ctra Records Sire Records Asylum Records

affect the final rendering of any auditory display such as: 1) how the original sound was produced, 2) audio card specifications, 3) rendering type (i.e., headphones or speakers), and 4) rendering type specifications. Nevertheless, as with the visual displays, the important thing is that a relative quality ordering of the auditory displays was established, for 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 need to be slightly degraded in order to avoid the absence or presence degradation issue which was a concern with the visual displays. The reason for this is that our eyes are accustomed to a certain fidelity (quality), but our ears are not as discerning. This was readily apparent during the process of selecting the three auditory display qualities. When evaluating the various selections, not one subject could not 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, as discussed earlier. Nevertheless, in terms of the higher qualities, the ears were not as discerning when evaluating sampling frequency as the eyes were at evaluating pixel resolution.

D. AUDITORY-VISUAL DISPLAY DEVELOPMENT

After establishing the visual and auditory displays, the next step was to develop the combined auditory-visual displays. The consideration here is 1) determining how long to render the displays, and 2) synchronizing the rendering of both auditory and visual displays. In order to eliminate any potential confounds, the amount of time 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 were asked if they needed more or less time to view or hear the appropriate displays. Based on a consensus, seven seconds was chosen for both displays.

Interestingly, some subjects at first thought they needed more time (around 20 seconds), but when given more time, the subjects realized that they were changing their minds too often about the quality, and 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, subjects were asked to assess distortions introduced by Motion Picture Expert Group-2 (MPEG) coding (see [MPEG98]). MPEG-2 sequences of 10 and 30 seconds length were used. One of the findings of this experiment was that the 30 second sequences were too long. This finding supports previous evidence of the length of human working memory (WM).

> There is evidence to suggest that WM has a duration of about 20 s and that the rate of decay in WM is dependent on the amount of information presented, as it has a limited capacity. Both of these facets of memory can be seen as important in the results, in that the end of the sequences are more accessible to memory recall (the recency effect) and may bias the subjects overall rating. [PETE59] [WICK92] [ALDR95]

Although the displays in the pilot study and main experiments are static, as opposed to motion video, the same concept of human WM applies. Therefore, based on subject consensus and human WM theory, all displays for the pilot study, whether presented separately or in combination, are presented to the subject for seven seconds. Having now established all required displays, the main design of the pilot study was ready to be developed.

E. SUMMARY

In summary, this chapter has provided an overview of the selection and development process of the auditory-only, visual-only, and combined auditory-visual displays utilized in the experimental design of this research effort.

VI. PILOT STUDY

A. INTRODUCTION

The pilot study played a crucial role in this research effort. The lessons learned from the pilot study were essential to the development and use of appropriate auditory and visual displays and to the overall design of the three main experiments forming the foundation of this dissertation.

B. LOCATION

All experiment sessions of the pilot study were conducted in the same isolated room under the same ambient conditions. The dimensions of the room were approximately 10 feet x 20 feet. Before each session, 1) all nonessential electronic equipment was turned off, 2) telephones were unplugged, 3) windows were closed and covered with blackout cloth, 4) the main overhead lights were turned off, 5) a 60 watt incandescent desk lamp was turned on behind the computer monitor to eliminate any glare, 6) the door to the room was closed, 7) a *Do Not Disturb Sign* was placed on the outside of the door, and 8) the subject was asked to turn off any audible pagers, mobile phones, and/or watches. This last condition was only implemented by accident, after a subject's beeper sounded during an experiment session.

C. PARTICIPANTS

A total of 22 volunteer participants (6 Female, 16 Male) comprised from the students, faculty, staff, and guests of NPS served as subjects ranging in age from 28 to 62. All subjects were required to have 20/20 or corrected 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 were not needed. Nevertheless, before conducting the experiment, each subject was asked, as part of a voluntary consent form, if he or she met the vision and hearing requirements.

D. APPARATUS

A Pentium 166 MHz personal computer with 64 MBytes main memory running *Microsoft Windows NT 4.0* served as the main hardware platform of the pilot study. The low-, medium-, and high-quality auditory displays, described earlier, were generated by a *Sound Blaster 16 PnP* audio card [CREA98] and rendered via *Sennheiser HD 540 reference II* headphones [SENN98]. The low-, medium-, and high-quality visual displays, described earlier, were generated by an Elsa *Gloria-8* graphics accelerator card [ELSA98] and rendered via a Sony *Multiscan 20 inch sf II* computer monitor [SONY98a] set at 800 x 600 resolution. The entire automated experiment was contained within a *Netscape Communicator 4.05* HTML browser window [NETS98] using *CosmoPlayer 2.0 VRML* plug-in [COSM98] to render the visual-only, auditory-only, and combined auditory-visual displays, and using Java pop-up windows developed using JDK 1.1.5 (Java Development Kit) [SUNM98] to collect subject responses.

E. PROCEDURE

The experiment involved a 3x3 factorial within subjects design. The two independent variables were visual and audio display quality. The two dependent variables were the corresponding quality perception of the auditory and visual displays. The three levels of the visual quality independent variable consisted of low-, medium-, and high-quality visual displays of the radio image depicted earlier in Chapter IV, Figure 32 having resolutions of 425 pixels/inch, 450 pixels/inch, and 500 pixels/inch respectively. The three levels of the auditory quality independent variable consisted of low-, medium-, and high-quality auditory displays of the same music selection having sampling rates of 11 kHz, 17 kHz, and 44.1 kHz respectively. As such, the visual display parameters manipulated were pixel resolution, and the auditory display parameters manipulated were sampling frequency. During each experiment, which lasts approximately 30 minutes, each subject wears headphones and sits in front of a 20-inch computer display monitor. The task of the subject was to rate the perceived quality of audio-only, visual-only, and audio-visual displays via rating scales as either low-, medium-, or high-quality.

After reading a brief experimental overview and signing a voluntary consent form, the subject was seated in a chair facing the computer monitor. The subject was instructed to adjust the seat height and/or monitor orientation to that which was most comfortable and which represented their typical computer monitor viewing habit. Although a standard viewing position/orientation is much desired in experimental design, the focus of this experiment was not on precision, but rather perception. Accordingly, the idea was for subjects to be 1) relaxed, 2) comfortable, 3) and in their typical viewing position/orientation. Nevertheless, no subject sat closer that about one foot or further than about three feet from the monitor. The subjects were instructed on how to wear and fit the headphones, and also how to adjust the volume if necessary. In order to maintain identical testing conditions, it was hoped that no one would need to adjust the previously established headset volume. If a subject did adjust the headset volume, that subject's data would not be included in the final data analysis. However, no subject needed to adjust the headset volume.

Once the subject was seated and wearing the headphones, an automated computer program contained within an HTML browser window instructed the subject to enter some personal data information as depicted in Figure 41. This personal data was used to create a unique data file to collect the specific subject's data for the remainder of the experiment. The file created is a *.csv* (comma separated variable) file which can easily be imported into *Microsoft Excel*. This was the only time for which the keyboard was utilized. For the remainder of the experiment, only the mouse was needed. The automated experiment continues by presenting the subject with a series of instructions giving full explanation of what is and is not required of the subject. The visual-only, auditory-only, and combined auditory-visual displays were rendered via VRML, and Java pop-up

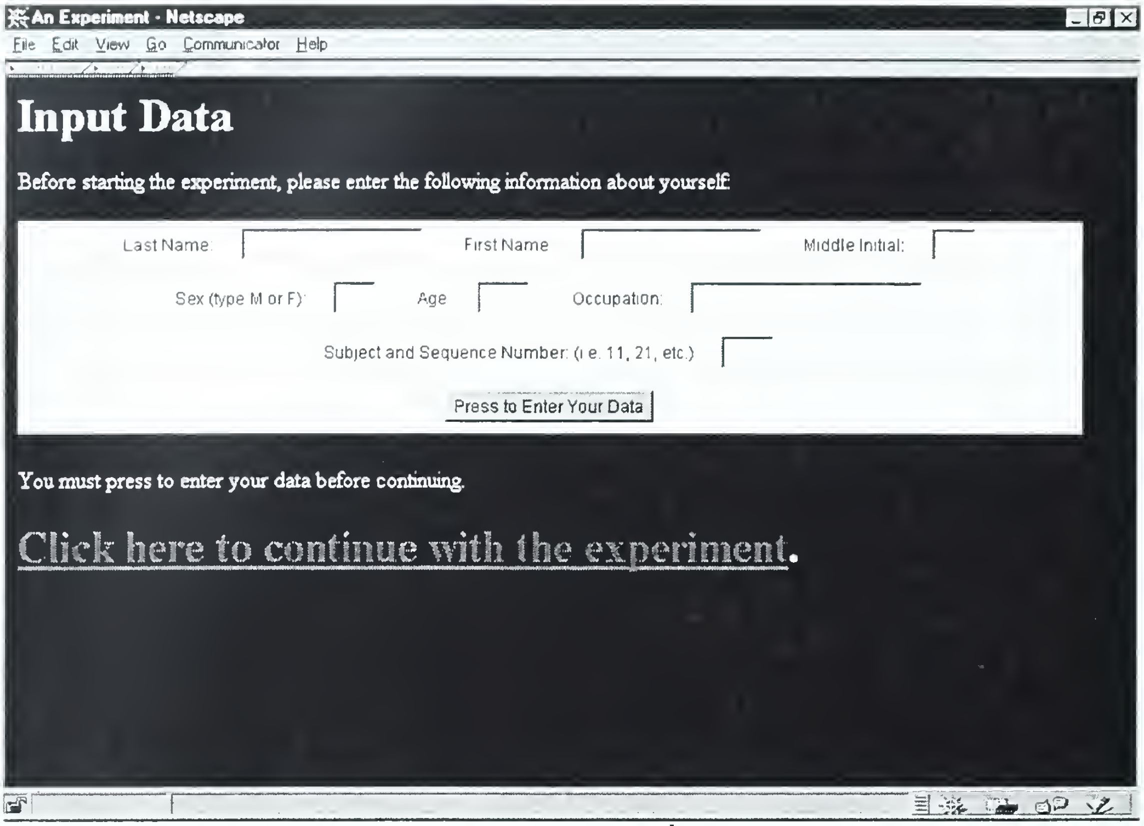


Figure 41. Pilot Study: Initial Data Input Screen.

windows collected subject responses. The primary reason for using VRML is for the eventual goal of manipulating auditory and visual displays in 3D scenes. Even though only static visual displays are currently used, the idea was to develop the foundation of the experiment using VRML to facilitate an easy transition to full 3D scenes. Other considerations for using VRML are as follows 1) it is freely downloadable, 2) it is easy to use, 3) it has a very short learning curve, and 4) it is new technology worth investigating.

As the automated experiment continues, the first set of instructions presented to the subject is depicted in Figure 42. The idea is for the subject to memorize the quality differences among the three displays. The same process was repeated again to give the subject yet another chance to review and memorize the three quality levels. Next, the subject is instructed how to rate the visual-only displays as depicted in Figure 43. After

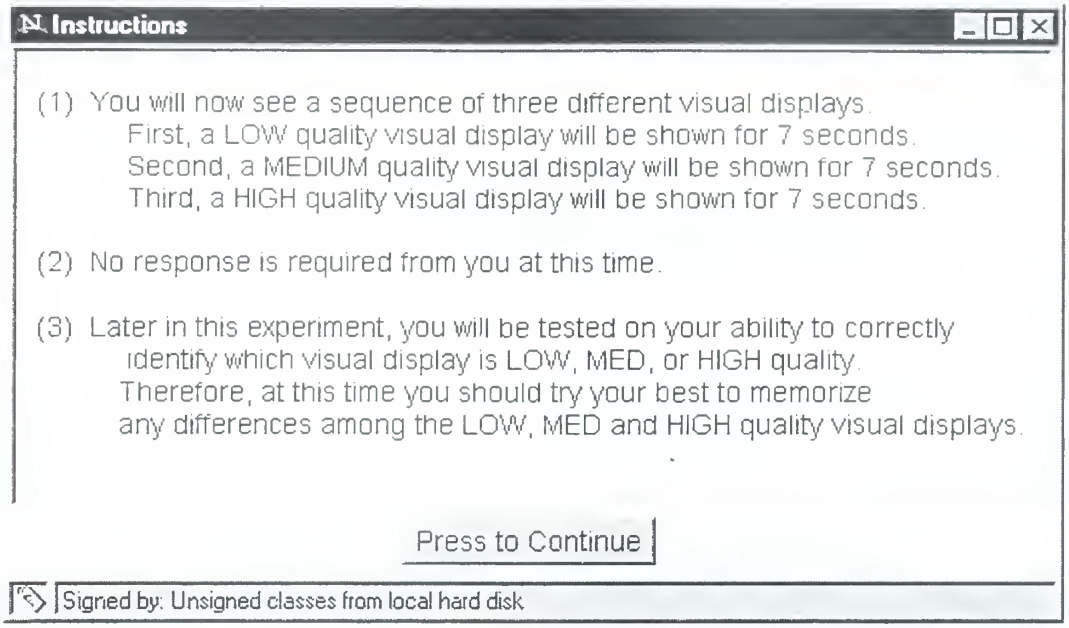


Figure 42. Pilot Study: Visual-Only Familiarization Instructions.

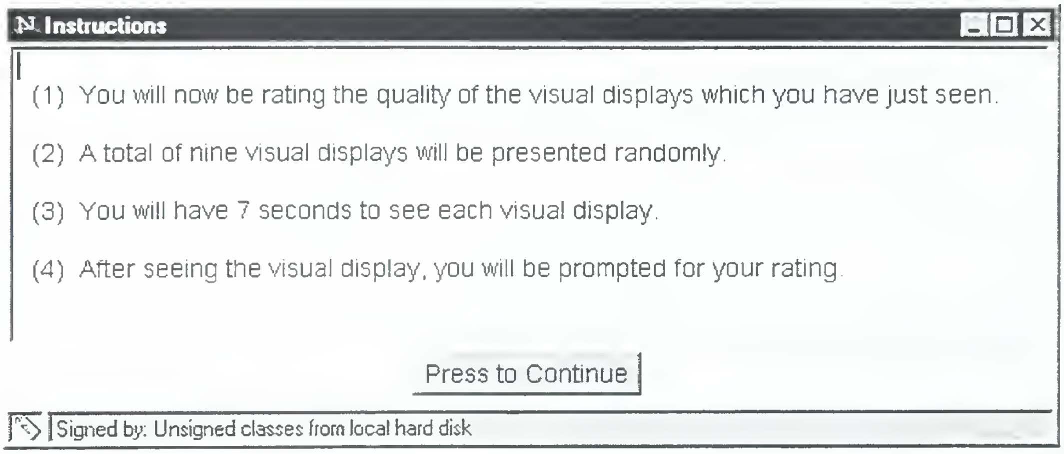


Figure 43. Pilot Study: Visual-Only Rating Instructions.

the seven seconds for which each visual display is rendered, the visual display automatically disappears, and a Java pop-up window automatically appears to facilitate the visual display rating as depicted in Figure 44. 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 exact same process to rate nine

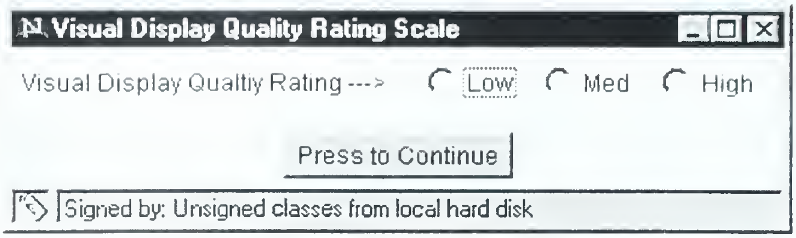


Figure 44. Pilot Study: Visual Display Rating Scale.

auditory-only displays (three of each quality presented in random order) by using the auditory rating scales as depicted in Figure 45. After rating the auditory displays, the

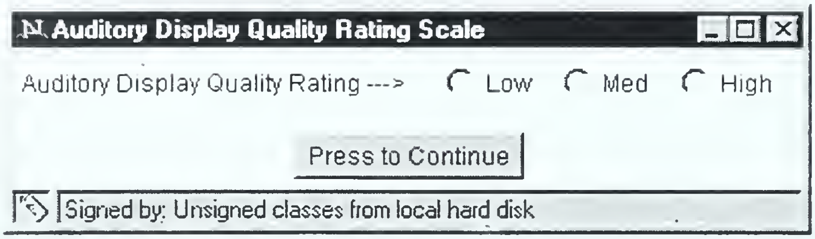


Figure 45. Pilot Study: Auditory Display Rating Scale.

subject is presented with instructions on how to rate the combined auditory-visual displays as depicted in Figure 46. After each of the 18 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 18 combined auditory-visual ratings), the subject rates both the auditory and visual displays using the combined auditory-visual rating scale depicted in Figure 47. After the subject has completed rating all of the displays, the automated portion of the experiment terminates. The subject is then asked to complete a brief post-experiment survey consisting of 13 questions as depicted in Figure 48 and Figure 49. After completing the post-experiment 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.

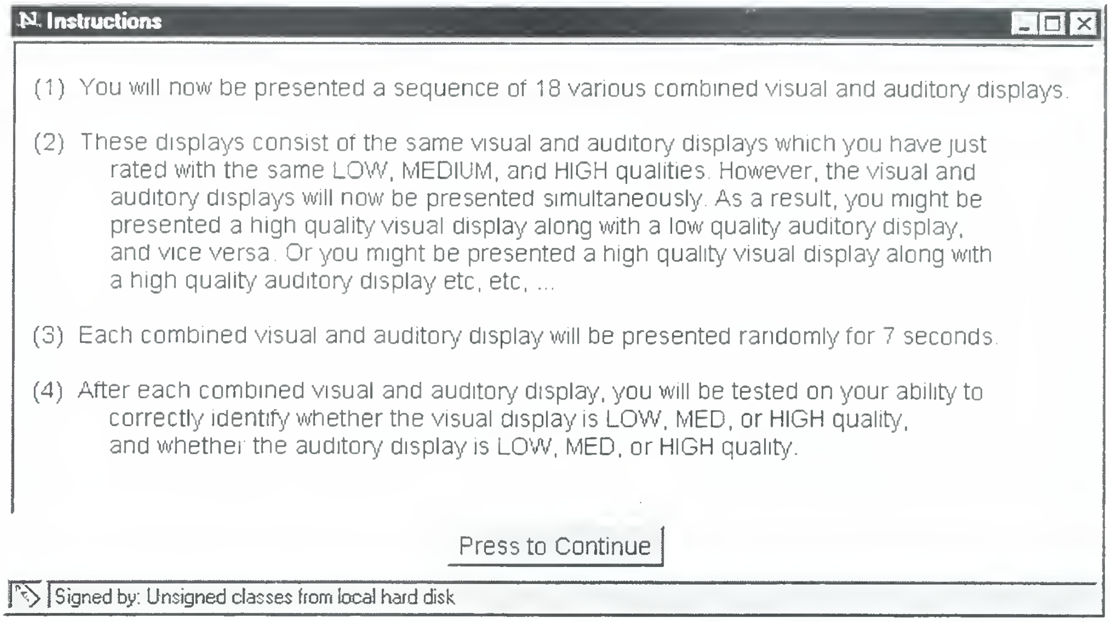


Figure 46. Pilot Study: Combined Auditory-Visual Rating Instructions.

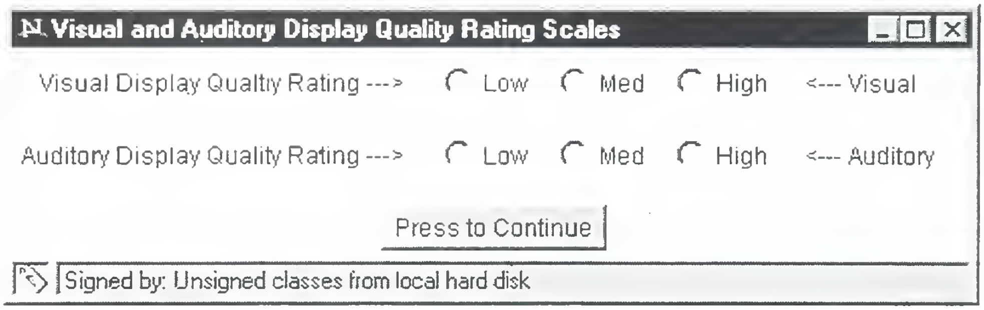


Figure 47. Pilot Study: Combined Auditory-Visual Rating Scale.

F. RESULTS AND DISCUSSION

The results of the pilot study proved invaluable and led to a completely redesigned experiment. Software and hardware problems, procedural problems, as well as validating some experimental design criteria were identified and are discussed below.

Post Experiment Questions

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 quality of the auditory-visual displays?

   very easy- 1 2 3 4 5 6 7 -very hard

4. 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

5. 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

6. Would you have liked less or more time to hear-see the auditory-visual displays?

   less time- 1 2 3 4 5 6 7 -more time

7. Time wise, was the overall experiment too short or too long?

   too short- 1 2 3 4 5 6 7 -too long

8. Was the experiment mentally exhausting or not?

   not very- 1 2 3 4 5 6 7 -yes very

Auditory-Visual Cross-Modal Experiment (Phase 1) 5

Last Name:________________

Subject and Sequence Number:________

Date:________________

Figure 48. Pilot Study: Post-Experiment Questions 1 - 8.

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 (Phase 1) 6

Last Name:____________________

Subject and Sequence Number:__________

Date:____________________

Figure 49. Pilot Study: Post-Experiment Questions 9 - 13.

1. Software and Hardware Problems

Perhaps the biggest problem of the pilot study was that the software and hardware utilized proved to be unstable. A computer hardware problem, which was never isolated, caused four complete system crashes, resulting in the need to completely reload *Windows NT* and all experiment software applications. This hardware problem caused the loss of valuable time of the subject as well as the experimenter not to mention the loss of the irreplaceable collected data. Furthermore, the *Windows NT* operating system crashed on numerous occasions during pilot study development and also during experiment sessions, again causing a considerable loss of valuable time and data. The use of VRML also caused unpredictable system crashes. This problem seemed to occur during Java-VRML intercommunication, and was evident by receiving the *Microsoft Visual C++ Runtime Library* error number *R6025: Pure Virtual Function Call.* Having tried numerous possible fixes, this unpredictable error remained. Another problem associated with VRML was synchronizing the combined auditory-visual displays. The reason for this is because the synchronization was based on the specifications of the particular audio and video hardware utilized. As a result, the synchronization of the displays could only be done through trial and error which was very time consuming. Furthermore, this limits the portability aspect of the experiment which is turn severely precludes the possibility of conducting future on-line experiments. Ultimately, because of the unreliable nature of the software and hardware, the pilot study was terminated before collecting the required number of data points to warrant proper data analysis. However, the results of the 13 subjects who successfully completed the experiment without any system crashes suggest that further examination of auditory-visual cross-modal perception phenomena is warranted. These results are discussed later.

2. Procedural Problems

Identifying experimental design procedural errors was another very important contribution of this pilot study. The main procedural errors identified were: visibility of

Netscape's status window, rating scales default setting, time delay between ratings, narrow range of rating scales, and memorization versus perception measurement.

a. Netscape Status Window

After asking one of the test subjects about the difficulty of the experiment, the subject said that it was not too hard to rate the quality of the displays, for he was simply looking at Netscape's status window while the displays were being loaded. He figured correctly, that the larger the file size, the better the quality. Thus, he simply looked at the status window, as opposed to the displays, resulting in very accurate responses. The immediate correction to this problem was to cover the status bar with a piece of black cloth. Ultimately it was discovered that the key sequence *ctrl-alt-s* toggles the appearance of Netscape's status window.

b. Rating Scales Default Setting

Unbeknownst to the subject, the subject's response time to rate the various displays was being measured. Upon analyzing the response time data, the response time to rate the medium-quality for the auditory-only, visual-only, and combined auditory-visual displays were significantly lower than that of the high- or low-quality displays. In analyzing why this might be, it became apparent that the reason was because the medium-quality choice was the default radio button setting on all the rating scales as depicted in Figure 50. As a result, if the subject were to make a medium-quality choice,

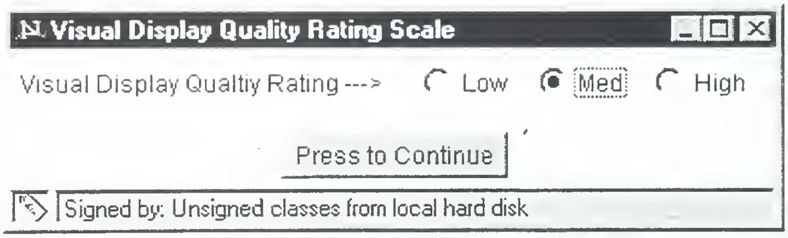


Figure 50. Pilot Study: Default Visual Quality Rating Scale.

the subject need only click the *Press to Continue* button on the rating scale. For the low-and high-quality choices, the subject had to select the appropriate radio button and then click the *Press to Continue* button on the rating scale which takes longer time. This problem was corrected by removing the medium-quality default choice as depicted earlier in Figure 44.

c. Time Delay Between Ratings

Because of how VRML was implemented in the experimental design, there was a noticeable time delay associated with the loading and unloading of the VRML Plug-in to Netscape. Many subjects complained that this time delay caused them to lose perspective on the relative quality ordering of the displays. Subjects wanted a faster turn-around time between quality ratings. A possible correction to this problem is to redesign VRML's use so that its plug-in is only loaded once at experiment start-up. However, compounded with the previous problems associated with VRML, the main experiments were redesigned without 3D VRML, resulting in 2D HTML displays.

d. Narrow Range of Rating Scales

Because of the experimental design, the range of the rating scales is small having only three possible values: low, med, high. This small range introduces unwanted floor and ceiling effects. For example, if a high-quality rating is not selected, for whatever reason, the only possible choices remaining are medium- and low-quality. Likewise, if a low-quality rating is not selected, for whatever reason, the only possible choices remaining are medium- and high-quality. As a result, this three-choice rating scale introduces unwanted floor and ceiling effects which in turn reduces the ability to properly measure any degrees of perceptual effects caused by the various quality displays. In terms of the goal of this research effort, using a three-choice rating scale severely hampers supporting data analyses. The correction to this problem is addressed later.

e. Memorization Versus Perception Measurement

The biggest procedural error was in the overall experimental design. This error stems from the basis by which subjects make their quality ratings. The question is one of measurement. Given that the task of a subject was to memorize the three auditory and visual display qualities, subjects responses were more likely based on their ability to memorize the given quality differences as opposed to perceiving potential changes in display qualities. Thus, the experiment becomes more of a matching problem as opposed to measuring perceptual phenomena. Because of this potential error, the experiment was completely redesigned as described in the next chapter.

3. Validated Design Criteria

Several positive outcomes resulted from the pilot study. In analyzing the post-experiment surveys, a seven-second duration of visual-only, auditory-only, and combined auditory-visual displays proved desirable and adequate. The subjects' approval also validated the overall length of the experiment, which typically lasted around 30 minutes. Furthermore, the responses of the subjects also suggested that with some effort, all the displays were noticeably different. This finding was very important for it validated the subjective relative quality ordering of the displays, which in turn validated the technique used to develop the various quality levels of the displays.

G. SUMMARY AND CONCLUSIONS

Because of the many experimental procedure errors identified during the pilot study, a valid data analysis of the results is not possible nor desired. Nevertheless, a few points are worth mentioning. In terms of memorization (the matching problem), the subjects were better able to correctly identify the quality levels of the visual-only and auditory-only displays, as opposed to correctly identifying the quality levels of the visual and auditory displays when presented in combination. Some subjects were better than others at identifying correct quality levels. In *post-hoc* analyses, there also appeared to be gender differences in identifying correct quality levels as well as differences in response times. Overall, the results of the pilot study indicate that there are differences in the subjects' ability to correctly match auditory-only, visual-only, and combined auditory-visual displays, and that gender may play a factor in correctly identifying the various displays. In the final analysis, the results of the pilot study greatly facilitated a new and improved experimental design ultimately supporting the goal of this research effort to investigate auditory-visual cross-modal perception phenomena.

VII. EXPERIMENT 1: STATIC RESOLUTION

A. 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 a radio depicted earlier in Chapter IV, Figure 32, and the auditory display is a selection of music. Specifically, 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?

3) Does a low-quality auditory display coupled with a low-quality visual display cause a decrease/increase 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) Does a high-quality auditory display coupled with a high-quality visual display cause an increase/decrease 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?

B. LOCATION

All sessions of Experiment 1: Static Resolution were conducted in the same isolated room under the same ambient conditions. The dimensions of the room were approximately 10 feet x 20 feet. Before each session, 1) all nonessential electronic equipment was turned off, 2) telephones were unplugged, 3) windows were closed and covered with blackout cloth, 4) the main overhead lights were turned off, 5) a 60 watt incandescent desk lamp was turned on behind the computer monitor to eliminate any glare, 6) the door to the room was closed, 7) a *Do Not Disturb Sign* was placed on the outside of the door, and 8) the subject was asked to turn off any audible pagers, mobile phones, and/or watches.

C. PARTICIPANTS

A total of 36 volunteer participants (18 Female, 18 Male) comprised from the students, faculty, staff, and guests of NPS served as subjects. Based on the preliminary findings of the 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 were required to have 20/20 or corrected 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 were not needed. Before conducting the experiment, each subject was asked, as part of a voluntary consent form, if he or she met the vision and hearing requirements.

D. APPARATUS

A Pentium 200 MHz (MMX) personal computer with 64 MBytes main memory running *Microsoft Windows 95* served as the main hardware platform of the experiment. The auditory displays are generated by a *Sound Blaster 64 AWE Gold* audio card [CREA98] and rendered via *Sennheiser HD 540 reference II* headphones [SENN98]. The visual displays are generated by a Diamond Multimedia *Viper V330* 128 bit graphics accelerator card [DIAM98] and rendered via a Sony *Multiscan 20-inch sf.II* computer monitor [SONY98a] set at 800 x 600 resolution. The entire automated experiment is contained within a *Netscape Communicator 4.05* HTML browser window [NETS98] using *JavaScript* to render the visual-only, auditory-only, and combined auditory-visual displays. Java pop-up windows, developed using JDK 1.1.5 [SUNM98], were used to collect subject responses.

E. PROCEDURE

The experiment involved a 3x3 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 in Chapter IV, Figure 32 having resolutions of 350 pixels/inch, 450 pixels/inch, and 550 pixels/inch, respectively. 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, 23 kHz, and 35 kHz, respectively. As such, the visual display parameters manipulated are pixel resolution, and the auditory display parameters manipulated are sampling frequency. During the experiment which lasts approximately 30 minutes, each subject wears headphones and sits in front of a 20-inch computer display monitor. The task of the subject is to rate the perceived quality of auditory-only, visual-only, and 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 was most comfortable and which represents their typical computer monitor viewing habit. 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 was for subjects to be 1) relaxed, 2) comfortable, 3) and in their typical viewing position/orientation. Nevertheless, no subject sat closer that about one foot or further than about 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. In order to maintain identical testing conditions, it was hoped that no one would need to adjust the headset volume. No subject needed to adjust the headset volume.

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 information as depicted in Figure 51. (Note that Netscape's status window

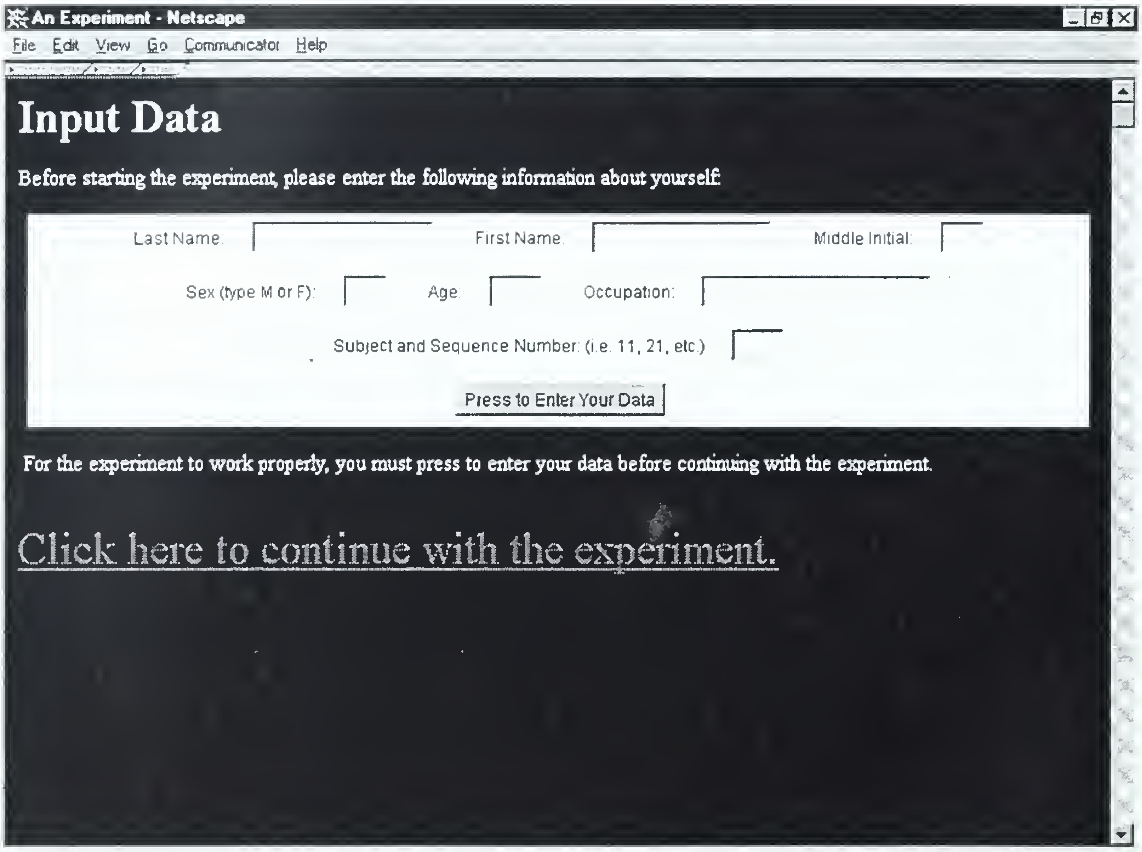


Figure 51. Experiment 1: Data Input Screen.

is not visible at the bottom of the screen as compared with that of the pilot study depicted earlier in Chapter VI, Figure 41.) This personal data is used to create a unique data file to collect the specific subject's data for the remainder of the experiment. The file created is a *.csv* (comma separated variable) file which can easily be imported into *Microsoft Excel*. This is the only time for which the keyboard was utilized. For the remainder of the experiment, only the mouse is needed. The automated experiment continues by

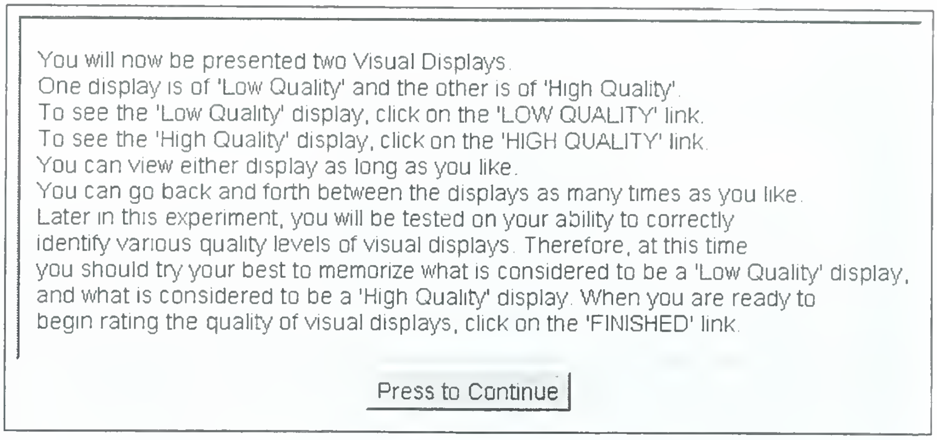


Figure 52. Experiment 1: Visual Display Instructions.

presenting the subject with a series of instructions giving 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 collects subject responses.

As the automated experiment continues, the subject is first presented with a series of instructions, displays, and rating scales in order to 1) ensure the headphones are working properly, 2) familiarize the subject with how the visual displays will be presented on the computer monitor, and 3) 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 first set of instructions presented to the subject is depicted in Figure 52. The idea is for the subject to memorize the quality differences between the lowest and highest quality visual displays. As a result, the subject calibrates himself or herself to the maximum possible quality range spanned by the low- and high-quality extremes. 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 53 depicts the appearance of the low-quality visual display having 250 pixels/inch and Figure 54 depicts the appearance of the high-quality visual display having 600 pixels/inch. Note, that the original displays were depicted in

Figure 53. Experiment 1: Low-Quality Visual Display Familiarization.

Figure 54. Experiment 1: High-Quality Visual Display Familiarization.

color, and that the actual pixel resolution experienced by the subject can only be viewed on the actual 20 inch computer monitor. However, the low- and high-quality displays depicted in Figure 53 and Figure 54 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. The subject is then presented with the instructions depicted in Figure 55. When ready, each

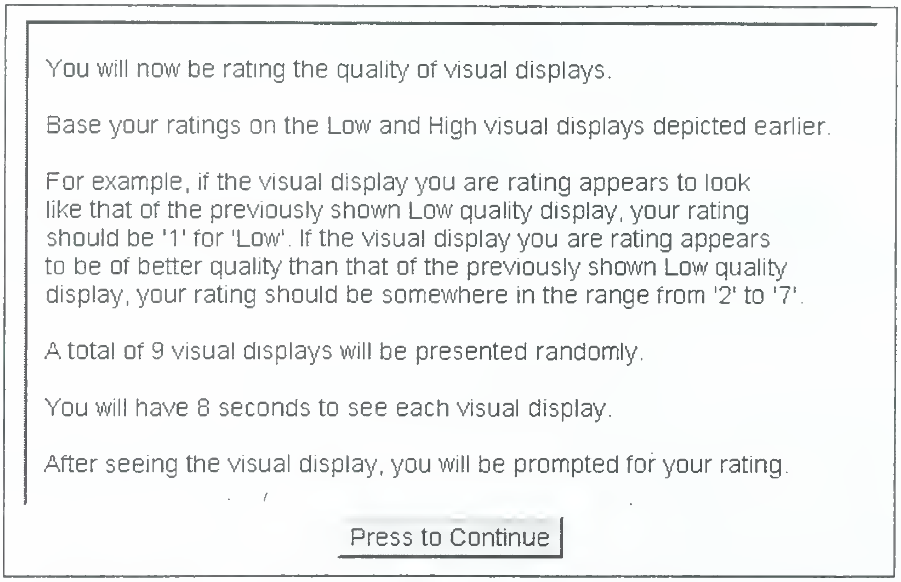


Figure 55. Experiment 1: Visual Display Rating Instructions.

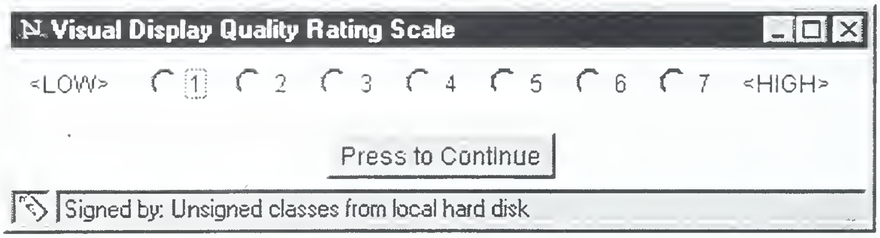


Figure 56. Experiment 1: Visual Display Quality Rating Scale.

visual display is 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 56. 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 corresponded to 8 kHz and 44.1 kHz respectively. The subject uses the exact same process, as with the visual displays, to rate nine auditory-only displays (three of each quality presented in random order) by using the auditory rating scales as depicted in Figure 57. After rating the auditory displays, the subject is

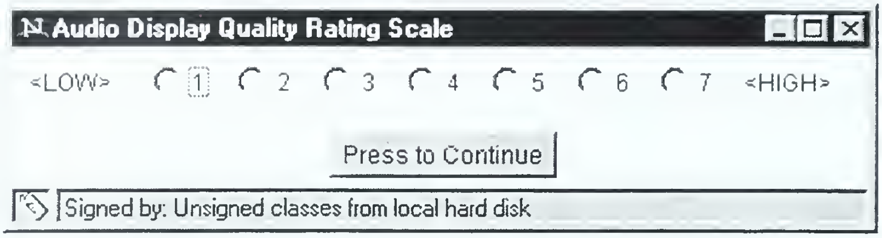


Figure 57. Experiment 1: Auditory Display Quality Rating Scale.

presented with instructions on rating 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) as depicted in Figure 58. The subject is then presented with instructions on rating only the auditory quality of nine combined auditory-visual displays (the nine permutations of the auditory and visual qualities are partially counterbalanced through the Latin squares technique) as depicted in Figure 59. Finally, the subject is presented with instructions on rating 18 combined auditory-visual displays as depicted in Figure 60. After each of the 18 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 18 combined auditory-visual ratings), the subject rates both the auditory and visual displays using the combined auditory-visual rating scale depicted in Figure 61. After the subject has completed rating all of the displays, the automated portion of the experiment terminates. The subject is then asked to complete a brief post-experiment survey consisting of 13 questions. This survey is identical to the

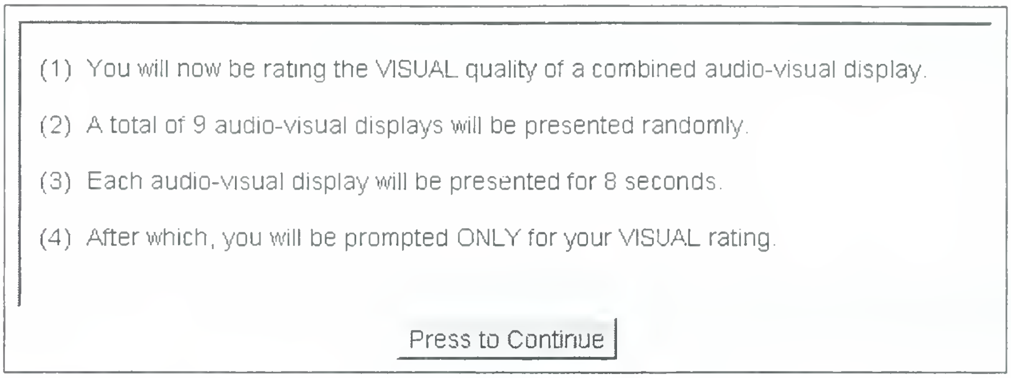


Figure 58. Experiment 1: Visual-Only Rating Instructions When Given A Combined Auditory-Visual Display.

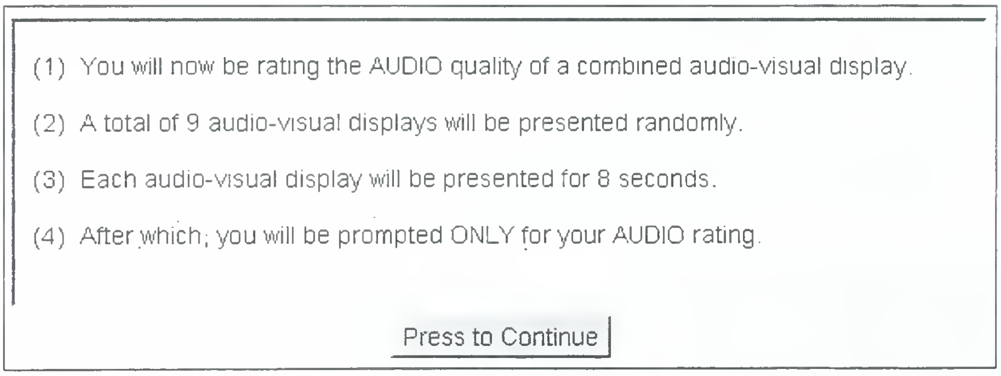


Figure 59. Experiment 1: Auditory-Only Rating Instructions When Given A Combined Auditory-Visual Display.

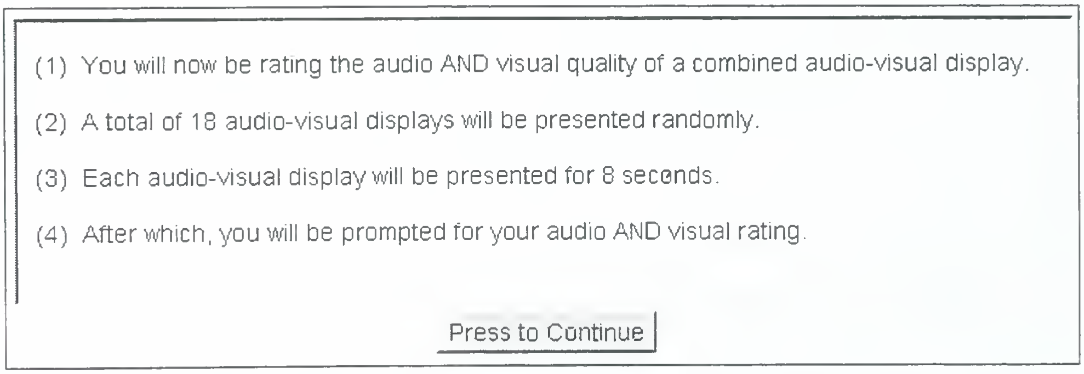


Figure 60. Experiment 1: Combined Auditory-Visual Rating Instructions.

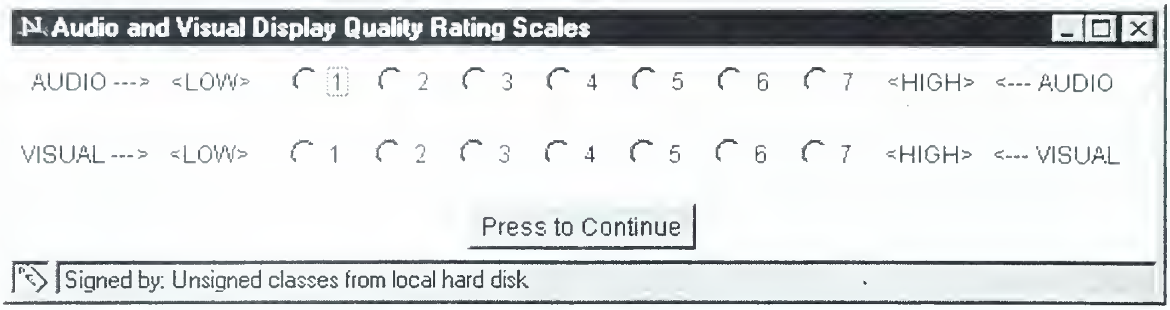


Figure 61. Experiment 1: Combined Auditory-Visual Rating Scale.

one used in the pilot study as depicted earlier in Chapter VI, Figure 48 and Figure 49. After completing the post-experiment 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.

F. CHANGES FROM PILOT STUDY

The following discussion describes how the results from the pilot study were implemented in the redesign of this experiment and how these implemented results affected the overall execution of the main experiment.

1. Software and Hardware Functionality

Switching to a new hardware platform proved to be extremely reliable and never exhibited any problems. Switching to *Microsoft Windows95* also proved to be very reliable since the operating system never once crashed. Eliminating the use of VRML also eliminated the system crashes associated with the *Microsoft Visual C++ Runtime Library* error number *R6025: Pure Virtual Function Call.* Furthermore, by using *JavaScript* as opposed to VRML, the combined auditory-visual displays were automatically synchronized when being rendered. This eliminated the trial and error process associated with VRML ultimately saving a lot of time and effort during the development of the main experiment, and thereby better supporting the portability aspect of the experiment for the eventual goal of conducting future on-line experiments.

2. Procedural Changes

a. Netscape Status Window

The use of the black cloth to cover Netscape's *Status Window* on the computer monitor was negated by learning the ability to use the key sequence *ctrl-alt-s* to toggle the on and off the *Status Window.* This not only increased the professionalism of the experiment, but also, albeit small, increased the size of the viewing display area.

b. Rating Scales Default Setting

By eliminating any default setting on the rating scales, the subject's response time measurement became uniform across all possible ratings, thereby allowing proper data analysis of response time.

c. Time Delay Between Ratings

By eliminating the use of VRML, the time required to load and unload the VRML Plug-in was likewise negated. As a result, through the use of JavaScript, there was practically no perceivable time delay between ratings. Given that the time between ratings was now instantaneous, the overall amount of time to complete the experiment was significantly reduced. This facilitated adding additional data collection aspects to the experimental design, while not increasing the overall duration of the experiment. As with the pilot study, subjects completed the experiment in about 30 minutes.

d. Range of Rating Scales

Given that the range of all rating scales was increased from three to seven choices, the floor and ceiling effects were significantly reduced if not altogether eliminated. This increased range provides the ability to properly measure any potential degrees of perceptual effects caused by the various quality displays.

e. Elimination of the Matching Problem

The matching (memorization) problem of the pilot study was eliminated by not requiring the subjects to memorize the three low, medium, and high display qualities. In this experiment, the subject is only required to memorize the lowest and highest possible quality extremes. During the rating process, the subject is never reexposed to the lowest and highest quality displays. Furthermore, the subject is not aware of how many quality levels are actually being presented. Since there are seven possible choices on the rating scales, not three, the subject can only guess that there may be upwards of seven possible quality levels for both the auditory and visual displays. By only requiring the subject to memorize the lowest and highest possible quality extremes, each subject, in essence, self-calibrates himself or herself, when rating the quality displays that fall between the given lowest and highest qualities. In fact, unbeknownst to the subject, only three quality levels: low, medium, and high, are presented. Thus, when rating the various auditory and visual displays, the rating process becomes purely subjective (perceptual) and not based on memorizing the exact quality level of a particular display.

f. Duration of Displays

During the pilot study, all displays were rendered for seven seconds, however, in this experiment all displays were rendered for eight seconds. The reason for increasing the length of the displays by one second had to do with the auditory display development for the follow-on experiment, Experiment 2: Static Noise. In this experiment, which is described in the next chapter, Gaussian white noise level is the manipulated auditory display parameter. As such, a one half second fade-in and fade-out of Gaussian white noise was added to the auditory display to negate the abrupt onset of the rendered Gaussian white noise which is somewhat shocking and startling if unexpected. This startling effect might cause subjects to become uneasy or unnerved.

Thus, to maintain consistency of display duration among all experiments, all displays among the experiments were rendered for eight seconds.

G. DATA COLLECTION AND ANALYSIS

Before the results of the experiment are discussed, it is important to understand the nature of the data collected and the chosen method of data analysis.

1. Data Collection

To better understand the method of data analysis, it is first necessary to understand the method of data collection. The idea of the experiment was to first capture the subject's quality perception of the visual-only and auditory-only displays. During this initial portion of the experiment, subjects rate nine displays consisting of three low, three medium, and three high qualities presented in random order. The average rated value for each quality display establishes the subject's baseline quality rating for each low-, medium-, and high-quality display. This baseline quality rating can then be compared to other all future quality ratings.

During the next portion of the experiment, subjects rate only the visual display quality of a combined auditory-visual display. The subject is presented nine combined auditory-visual displays corresponding to the nine permutations formed by the three auditory and three visual display qualities. The ordering of these nine displays is partially counterbalanced through the Latin squares technique. As such, the subject again rates the three low, three medium, and three high qualities of the visual displays. The average rated value for each quality display establishes the subject's visual quality rating for each low-, medium-, and high-quality display when presented in combination with the three quality levels of the auditory displays.

During the next portion of the experiment, subjects rate only the auditory display quality of a combined auditory-visual display. The subject is presented nine combined auditory-visual displays corresponding to the nine permutations formed by the three auditory and three visual display qualities. The ordering of these nine displays is again partially counterbalanced through the Latin squares technique. As such, the subject again rates the three low, three medium, and three high qualities of the auditory displays. The average rated value for each quality display establishes the subject's auditory quality rating for each low-, medium-, and high-quality display when presented in combination with the three quality levels of the visual displays.

During the final portion of the experiment, subjects rate both the auditory and visual display qualities of a combined auditory-visual display. The subject is presented 18 combined auditory-visual displays corresponding to 1) the nine permutations formed by the three auditory and three visual display qualities and 2) the reversal of the nine permutations formed by the three auditory and three visual display qualities all of which is again partially counterbalanced through the Latin squares technique. As such, the subject rates, yet again, the three low, three medium, and three high qualities of the visual displays and the auditory displays. The average rated value for each quality display establishes the subject's visual and auditory quality rating for each low-, medium-, and high-quality display when having to rate both visual and auditory displays simultaneously. However, to conform with the next two experiments, only the first nine of the 18 combined auditory-visual displays are utilized during data analysis.

The response time, the time to rate each display, was also collected. However, the subject was not aware of this fact. A conscious decision was made not to inform the subject, to avoid the possibility of the subject thinking that the faster the response, the better the score as in some kind of race. The idea is to keep the subject as relaxed as possible so that the subject's decisions are based purely on perception, and not on time (speed) related factors.

2. Data Analysis

As in any experiment, proper/valid data analysis is critical. The first step towards a valid data analysis involves understanding and identifying the type of data collected such as nominal, ordinal, interval, and continuous. In this experiment, all the quality ratings collected are considered ordinal data. The reason for this is that the quality ratings are derived from rating scales which are used to rank the quality perception of the displays by giving a rating on a scale of 1 (lowest) to 7 (highest). To be contrasted with interval data, the difference in quality between the low and medium displays is not necessarily the same difference in quality between the medium- and high-quality displays. This is a very important point, which must be considered when selecting the proper data analysis method.

The underlying distribution of the data is another very important factor in deciding how to analyze the data. Parametric data analysis can be used when assuming a certain underlying distribution of the data. Nonparametrics are used to test hypotheses about data from which the underlying distribution of data is not assumed. Thus, 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 used to compare the number of observations above and below a certain hypothesized value, which in this case is zero as described below. As such, 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 a) the visual-only quality rating of a combined auditory-visual display, and b) the baseline rating for the visual-only quality display is zero.

2) The difference between a) the auditory-only quality rating of a combined auditory-visual display, and b) the baseline rating for the auditory-only quality display is zero.

3) The difference between a) the visual quality rating of a combined auditory-visual display when also rating the auditory display, and b) the baseline rating for the visual-only quality display is zero.

4) The difference between a) the auditory quality rating of a combined auditory-visual display when also rating the visual display, and b) 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 1) the visual-only quality rating of a combined auditory-visual display, 2) the auditory-only quality rating of a combined auditory-visual display, 3) the visual quality rating of a combined auditory-visual display when also rating the auditory display, and 4) 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 .05. In other words, the degree of confidence supporting all experimental findings is at the .05 level. As such, only P-values at the .05 level will be reported as significant. This P-value is the probability of making a Type I Error. In other words, the P-value is the probability of rejecting the null hypothesis when in fact the null hypothesis is true. As such, the smaller the P-value, the greater the confidence in rejecting the null hypothesis which in turn supports the alternative hypothesis (see [GOOD95] for more discussion on alpha level, null hypothesis, alternative hypothesis, and Type I Error).

H. RESULTS AND DISCUSSION

The overall results of this experiment suggest significant auditory-visual cross-modal perception phenomena relevant to VE and multimedia developers. The major findings of this experiment are now discussed.

1. Validity

The first and most important consideration is whether the quality of the visual and auditory displays developed for this experiment are rank ordered by the subjects

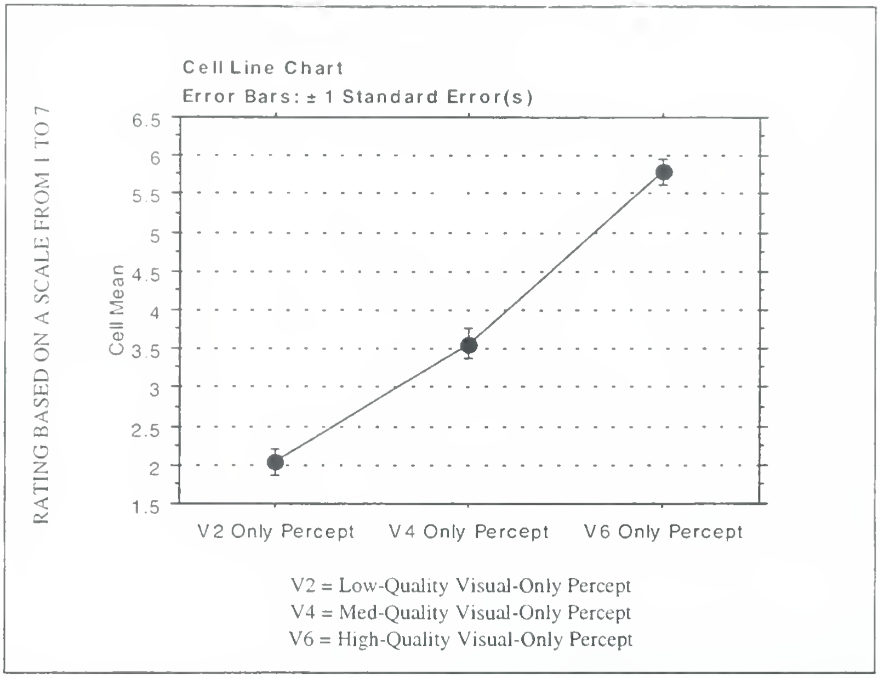


Figure 62. Experiment 1: Visual-Only Quality Percept Ratings.

according to their intended rankings. If this were not the case, the validity of the experiment would be jeopardized. However, in looking at Figure 62, 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. Likewise, in looking at Figure 63, 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 hypotheses can continue.

2. Findings

Figure 64 represents the results of all one sample sign tests based on the first null hypothesis which states: the difference between a) the visual-only quality rating of a combined auditory-visual display, and b) the baseline rating for the visual-only quality

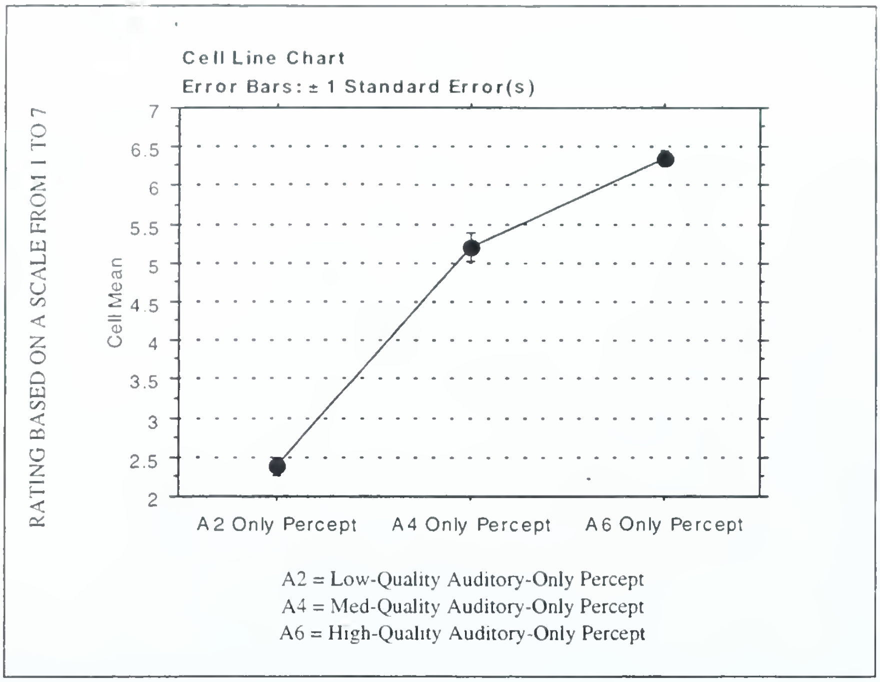


Figure 63. Experiment 1: Auditory-Only Quality Percept Ratings.

display is zero. As one can see from the results, when presented a combined high-quality visual and high-quality auditory display, when only asked to rate the quality of the visual display, a statistically significant finding at the .0161 level (a P-value of .0161) suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

Figure 65 represents the results of all one sample sign tests based on the second null hypothesis which states: the difference between a) the auditory-only quality rating of a combined auditory-visual display, and b) the baseline rating for the auditory-only quality display is zero. As one can see from the results, when presented a combined low-quality auditory and high-quality visual display, when only asked to rate the quality of the auditory display, a statistically significant finding at the .0002 level strongly suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display.

One-Sample Sign Test for V2 A2 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	10
# Obs < Hyp Value	17
# Obs = Hyp Value	9
P-Value	2478

One-Sample Sign Test for V2 A4 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	14
# Obs < Hyp Value	16
# Obs = Hyp Value	6
P-Value	8555

One-Sample Sign Test for V2 A6 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	15
# Obs < Hyp Value	13
# Obs = Hyp Value	8
P-Value	8506

One-Sample Sign Test for V4 A2 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	15
# Obs. < Hyp Value	18
# Obs = Hyp Value	3
P-Value	7283

One-Sample Sign Test for V4 A4 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	14
# Obs < Hyp. Value	16
# Obs = Hyp Value	6
P-Value	8555

One-Sample Sign Test for V4 A6 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	15
# Obs < Hyp Value	15
# Obs = Hyp Value	6
P-Value	> 9999

One-Sample Sign Test for V6 A2 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	16
# Obs < Hyp Value	16
# Obs = Hyp Value	4
P-Value	> 9999

One-Sample Sign Test for V6 A4 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	18
# Obs. < Hyp Value	11
# Obs = Hyp Value	7
P-Value	2649

One-Sample Sign Test for V6 A6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp Value	22
# Obs < Hyp Value	8
# Obs = Hyp Value	6
P-Value	.0161

V2A2 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV = High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV = High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV = High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 64. Experiment 1: One Sample Sign Tests for Visual-Only Quality Percept of Combined Auditory-Visual Displays.

Figure 66 represents the results of all one sample sign tests based on the third null hypothesis which states: the difference between a) the visual quality rating of a combined auditory-visual display when also rating the auditory display, and b) the baseline rating for the visual-only quality display is zero. As one can see from the results, there are no significant findings at the .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, when asked to rate both auditory and visual displays, the results at the .10 level suggest that the quality perception of the high-quality visual display is increased.

One-Sample Sign Test for A2 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs < Hyp. Value	16
# Obs. = Hyp. Value	4
P-Value	>.9999

One-Sample Sign Test for A2 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	12
# Obs. < Hyp Value	20
# Obs. = Hyp. Value	4
P-Value	.2153

One-Sample Sign Test for A2 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	5
# Obs. < Hyp Value	26
# Obs = Hyp. Value	5
P-Value	.0002

One-Sample Sign Test for A4 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	4
P-Value	>.9999

One-Sample Sign Test for A4 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	12
# Obs. < Hyp. Value	19
# Obs. = Hyp. Value	5
P-Value	.2810

One-Sample Sign Test for A4 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs. < Hyp. Value	12
# Obs. = Hyp. Value	6
P-Value	.3616

One-Sample Sign Test for A6 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	6
P-Value	.8555

One-Sample Sign Test for A6 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs. < Hyp. Value	12
# Obs. = Hyp. Value	6
P-Value	.3616

One-Sample Sign Test for A6 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	15
# Obs. < Hyp. Value	11
# Obs. = Hyp. Value	10
P-Value	.5572

A2V2 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 65. Experiment 1: One Sample Sign Tests for Auditory-Only Quality Percept of Combined Auditory-Visual Displays.

Figure 67 represents the results of all one sample sign tests based on the fourth null hypothesis which states: the difference between a) the auditory quality rating of a combined auditory-visual display when also rating the visual display, and b) the baseline rating for the auditory-only quality display is zero. The results suggest that: 1) when presented a combined low-quality auditory and high-quality visual display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0107 level suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display, and 2) when presented a combined high-quality auditory and low-quality visual display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0241 level suggests that the

One-Sample Sign Test for	V2 A2 CAV Difl	V2 A4 CAV Diff	V2 A6 CAV Diff	V4 A2 CAV Diff	V4 A4 CAV Diff	V4 A6 CAV Diff	V6 A2 CAV Diff	V6 A4 CAV Diff	V6 A6 CAV Diff
Hypothesized Value	0	0	0	0	0	0	0	0	0
# Obs > Hyp. Value	12	12	12	11	18	17	18	20	19
# Obs < Hyp. Value	15	15	16	20	9	15	12	10	9
# Obs = Hyp. Value	9	9	8	5	9	4	6	6	8
P-Value	7011	.7011	5716	.1496	.1221	.8601	3616	.0987	.0872

V2A2 CAV = Low-Quality Visual Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 CAV = Low-Quality Visual Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 CAV = Low-Quality Visual Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 CAV = Med-Quality Visual Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 CAV = Med-Quality Visual Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 CAV = Med-Quality Visual Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 CAV = High-Quality Visual Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 CAV = High-Quality Visual Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 CAV = High-Quality Visual Percept of Combined High-Visual and High-Auditory Quality Display

Figure 66. Experiment 1: One Sample Sign Tests for Visual Quality Percept When Also Rating the Auditory Display of Combined Auditory-Visual Displays.

quality perception of a high-quality auditory display is increased when coupled with a low-quality visual display.

In terms of response times, Figure 68 represents the average visual quality rating response times of a combined auditory-visual display, when only asked to rate the quality of the visual display. Figure 69 represents the average auditory quality rating response times of a combined auditory-visual display, when only asked to rate the quality of the auditory display. Figure 70 represents the average combined auditory and visual quality

One-Sample Sign Test for A2 V2 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	17
# Obs. < Hyp. Value	15
# Obs. = Hyp. Value	4
P-Value	8601

One-Sample Sign Test for A2 V4 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs. < Hyp. Value	19
# Obs. = Hyp. Value	1
P-Value	.7359

One-Sample Sign Test for A2 V6 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	8
# Obs. < Hyp. Value	23
# Obs. = Hyp. Value	5
P-Value	.0107

One-Sample Sign Test for A4 V2 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs. < Hyp. Value	14
# Obs. = Hyp. Value	6
P-Value	.8555

One-Sample Sign Test for A4 V4 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	12
# Obs. < Hyp. Value	20
# Obs. = Hyp. Value	4
P-Value	.2153

One-Sample Sign Test for A4 V6 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	4
P-Value	>.9999

One-Sample Sign Test for A6 V2 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	8
# Obs. = Hyp. Value	7
P-Value	.0241

One-Sample Sign Test for A6 V4 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs. < Hyp. Value	13
# Obs. = Hyp. Value	7
P-Value	.7111

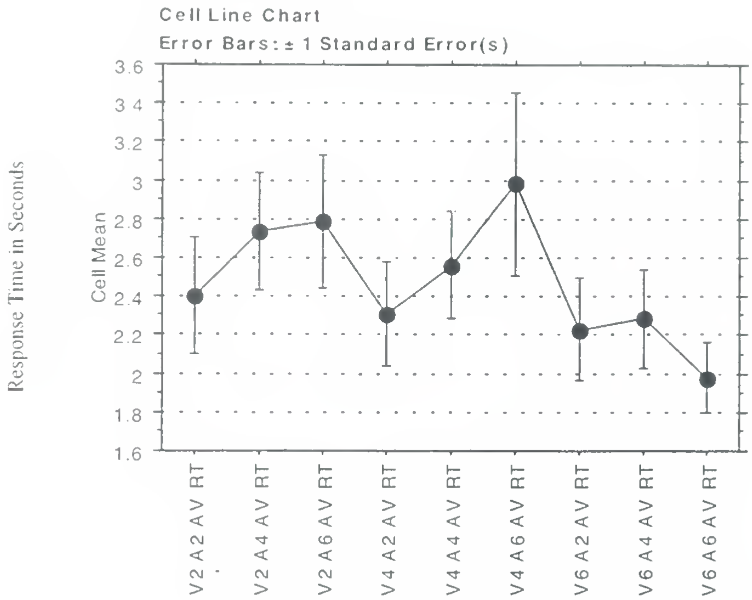
One-Sample Sign Test for A6 V6 CAV Diff

Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	6
P-Value	.8555

A2V2 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 CAV = High-Quality Auditory Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 CAV = High-Quality Auditory Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 CAV = High-Quality Auditory Percept of Combined High-Auditory and High-Visual Quality Display

Figure 67. Experiment 1: One Sample Sign Tests for Auditory Quality Percept When Also Rating the Visual Display of Combined Auditory-Visual Displays.

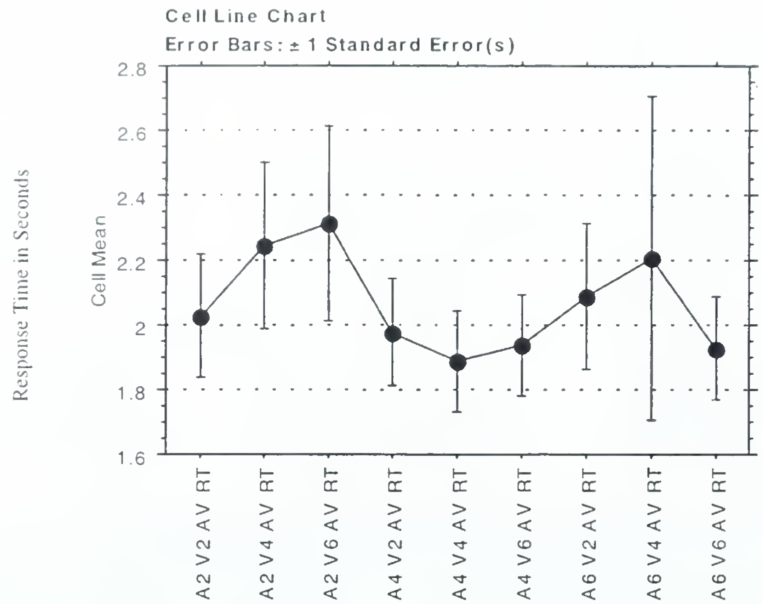


V2A2 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 68. Experiment 1: Visual-Only Quality Rating Response Times of a Combined Auditory-Visual Display.

rating response times of a combined auditory-visual display, when asked to rate both the auditory and visual displays.

In looking at the results of the response times, one can see various trends based on a particular auditory-visual quality combination. However, several factors limit the ability to correctly analyze these temporal results in any statistically valid manner. These factors are discussed in the last chapter. Nevertheless, one key observation is worth mentioning. Nevertheless, the response time to rate the visual-only display of a combined auditory-visual display exhibited the only occasion in the entire experiment where gender seems to be a factor. In looking at Figure 71, it is apparent in every condition, that females need

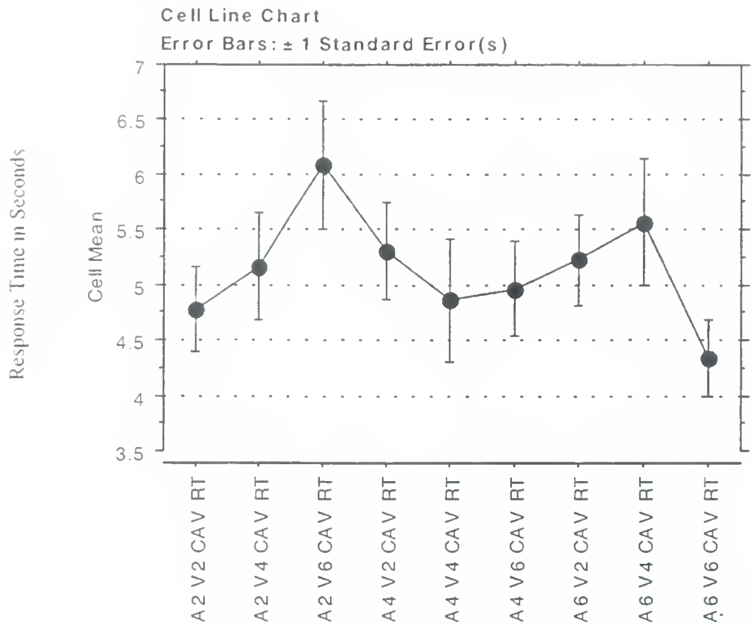


A2V2 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 69. Experiment 1: Auditory-Only Quality Rating Response Times of a Combined Auditory-Visual Display.

more time than males to rate the visual displays. The reason for this is not known, but does suggest that it might be harder for females to filter out the auditory information while trying to attend only to the visual display. Another reason might be a result of the competitive nature of males. Specifically, males might have been more prone to answer as quickly as possible; whereas, females simply took as much time as they felt they needed.

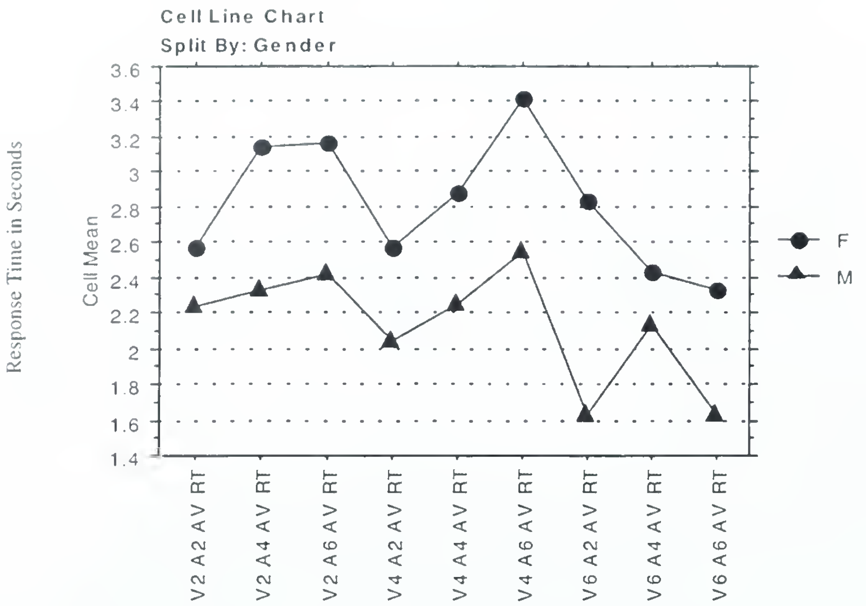
In terms of the post-experiment questions, Figure 72 represents the subject's opinion on 1) how easy or difficult it was to determine the quality of the various displays, and 2) if less or more time was needed to adequately rate the various displays. Keeping in



A2V2 CAV RT = Time to Rate Both Low-Auditory and Low-Visual Quality Displays of a Combined Display
A2V4 CAV RT = Time to Rate Both Low-Auditory and Med-Visual Quality Displays of a Combined Display
A2V6 CAV RT = Time to Rate Both Low-Auditory and High-Visual Quality Displays of a Combined Display
A4V2 CAV RT = Time to Rate Both Med-Auditory and Low-Visual Quality Displays of a Combined Display
A4V4 CAV RT = Time to Rate Both Med-Auditory and Med-Visual Quality Displays of a Combined Display
A4V6 CAV RT = Time to Rate Both Med-Auditory and High-Visual Quality Displays of a Combined Display
A6V2 CAV RT = Time to Rate Both High-Auditory and Low-Visual Quality Displays of a Combined Display
A6V4 CAV RT = Time to Rate Both High-Auditory and Med-Visual Quality Displays of a Combined Display
A6V6 CAV RT = Time to Rate Both High-Auditory and High-Visual Quality Displays of a Combined Display

Figure 70. Experiment 1: Response Times of Both Auditory and Visual Displays of a Combined Auditory-Visual Display.

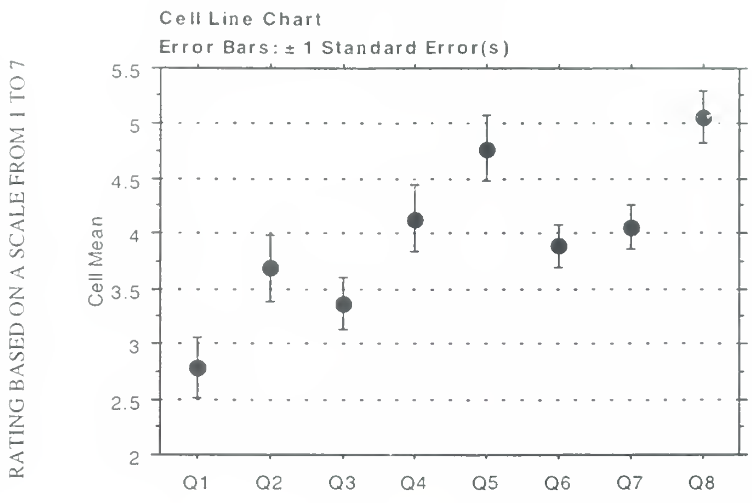
mind that subjects used a Likert rating scale ranging from 1 to 7 (4 being neutral) to rate their opinions, the results indicate that determining the quality of both auditory and visual displays of a combined auditory-visual display proved to be more difficult than determining the quality of either auditory or visual display presented either alone or in combination. Furthermore, the results indicate that eight seconds was an adequate amount of time to rate the visual-only and auditory displays, but that slightly more than eight seconds was desired when rating the combined auditory-visual displays.



V2A2 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 71. Experiment 1: Comparison of Male and Female Response Times When Rating a Visual-Only Display of a Combined Auditory-Visual Display.

Finally, the remaining questions of the post-experiment 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. Some very interesting observations were also observed concerning the descriptions subjects used to determine the quality of the various displays. These observations are outlined in the final chapter.



Q1 = How easy or difficult was is to determine the quality of the visual-only displays?
Q2 = How easy or difficult was is to determine the quality of the auditory-only displays?
Q3 = How easy or difficult was is to determine the visual quality of the auditory-visual displays?
Q4 = How easy or difficult was is to determine the auditory quality of the auditory-visual displays?
Q5 = How easy or difficult was to determine both the auditory and visual qualities of the auditory-visual displays?
Q6 = Would you have liked less or more time to view the visual-only displays?
Q7 = Would you have liked less or more time to hear the auditory-only displays?
Q8 = Would you have liked less or more time to hear-view the combined auditory-visual displays?

Figure 72. Experiment 1: Post-Experiment Questions 1 - 8.

I. SUMMARY AND CONCLUSIONS

Overall the findings suggest that whether asked to specifically attend to both auditory and visual modalities, or asked to attend to only one modality, both 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:

1) 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.

2) 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.

3) 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? As such, the next chapter investigates whether manipulating visual display Gaussian white noise level and auditory display Gaussian white noise level produce the same results.

VIII. EXPERIMENT 2: STATIC NOISE

A. 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 depicted in Chapter IV, Figure 32, and the auditory display is a selection of music. As in the previous experiment, 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?

3) Does a low-quality auditory display coupled with a low-quality visual display cause a decrease/increase 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) Does a high-quality auditory display coupled with a high-quality visual display cause an increase/decrease 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?

B. LOCATION

Because the building containing the room of the first experiment was undergoing electrical rewiring resulting in many power outages, the location of this experiment was moved to a different building. Nevertheless, all testing sessions of Experiment 2: Static Noise were conducted in a similar isolated room under the same ambient conditions. The dimensions of the room were slightly smaller than that of the first experiment at approximately 10 feet x 10 feet. Before each session, 1) all nonessential electronic equipment was turned off, 2) telephones were unplugged, 3) windows were closed and covered with blackout cloth, 4) the main overhead lights were turned off, 5) a 60 watt incandescent desk lamp was turned on behind the computer monitor to eliminate any glare, 6) the door to the room was closed, 7) a *Do Not Disturb Sign* was placed on the outside of the door, and 8) the subject was asked to turn off any audible pagers, mobile phones, and/or watches.

C. PARTICIPANTS

A total of 36 volunteer participants (27 Male, 9 Female) comprised from the students, faculty, staff, and guests of NPS served as subjects. Based on the limited gender findings of the first experiment (Experiment 1: 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 were required to have 20/20 or corrected 20/20 vision and normal hearing. Because the experiment did not involve precise measurements of Gaussian noise levels, a vision and hearing test were not needed. Before conducting the experiment, each subject was asked, as part of a voluntary consent form, if he or she met the vision and hearing requirements.

D. APPARATUS

The apparatus used in this experiment is identical to that of Experiment 1: Static Resolution. See Chapter VII, Section D.

E. PROCEDURE

Except for a few changes which will be discussed, the procedure of this experiment is identical to that of the first experiment, Experiment 1: Static Resolution. The experiment involved a 3x3 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 development process of the visual displays was identical to that of the first experiment, except that Gaussian white noise levels were manipulated with Adobe *Photoshop* [ADOB98] 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 radio image depicted in Chapter IV, Figure 32, 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 was identical to that of the first experiment, except that Gaussian noise levels of the original music selection at 44.1 kHz, were manipulated with Sonic Foundary's *SoundForge* [SONI98] 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 having mixed in Gaussian noise level amounts of 31 percent, 23 percent, and 15 percent, respectively. As such, both the visual and auditory display parameters manipulated are Gaussian noise level. During the experiment, which lasts approximately 30 minutes, each subject wears headphones and sits in front of a 20-inch computer display monitor. The task of the subject is to rate the perceived quality of audio only, visual-only, and audio-visual displays via Likert rating scales ranging from 1 (low) to 7 (high).

The lowest- and highest-quality auditory displays in which the subjects were supposed to memorize during the self-calibration phase corresponded to the music selection at 44.1 kHz, having mixed in Gaussian noise level amounts of 45 percent and 10 percent, respectively. The lowest- and highest-quality visual displays in which the subjects were supposed to memorize during the self-calibration phase are depicted in Figure 73 and Figure 74, respectively. 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. Again, it is important to remember that the original

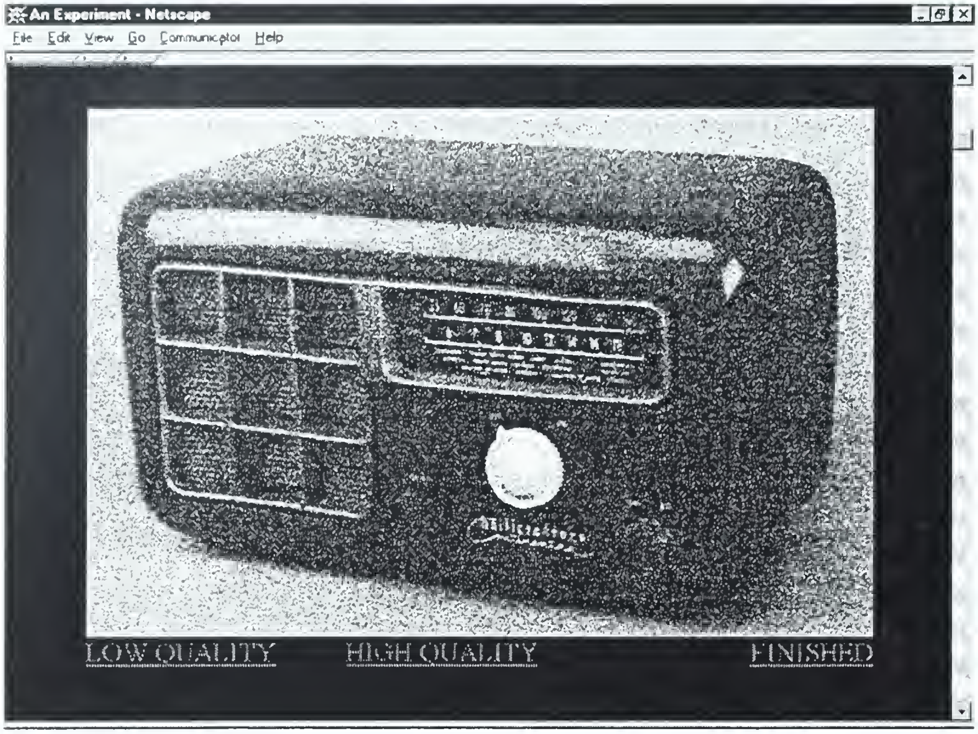


Figure 73. Experiment 2: Low-Quality Visual Display Familiarization.

Figure 74. Experiment 2: High-Quality Visual Display Familiarization.

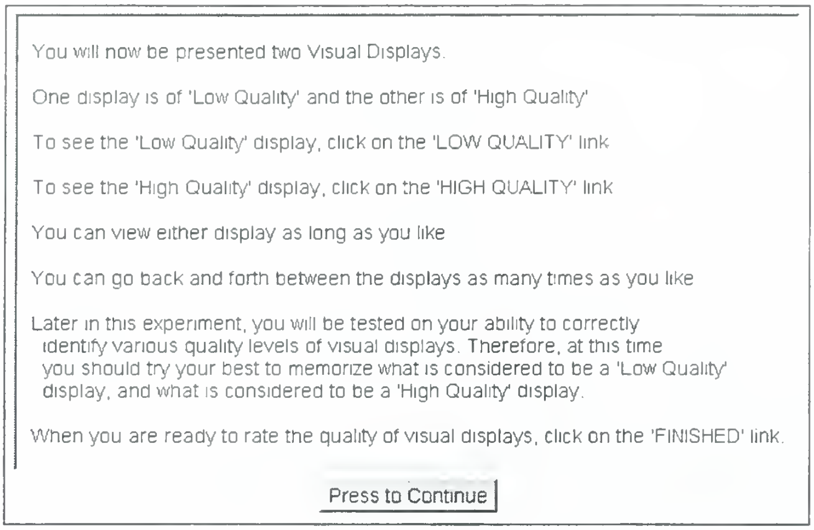


Figure 75. Experiment 2: Visual Display Instructions.

displays were depicted in color, and that the actual Gaussian noise level experienced by the subject can only be viewed on the actual 20-inch computer monitor. However, the low- and high-quality displays depicted in Figure 73 and Figure 74 are fairly good representations of the quality difference between the actual displays used in the experiment. Besides the different auditory and visual stimuli utilized, the procedure continues exactly as in the previous experiment except for 1) minor changes in the readability of instructions, 2) an increase in the number of visual-only and auditory-only quality ratings, and 3) a decrease from 18 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 was enhanced by adding more *white space*. An example of this is comparing the instructions from the previous experiment as depicted in Chapter VII, Figure 52 with the revised instructions as depicted in Figure 75. Note that the content of the instructions was not changed only the readability was enhanced through increased use of *white space*.

In order 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 was increased from 9 to 12. However, to conform with the data analysis of the previous experiment, the first three ratings, consisting of one low-, medium-, and high-quality were disregarded. The idea was to allow the subject, unknowingly, to see/hear the three quality levels one time before having to make a rating. The baseline ratings were still based on an average of three quality ratings to conform with the data analysis of the previous, and 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 was also changed based on subjects' comments from the previous experiment. Subjects felt that rating 18 combined auditory-visual displays was somewhat long and tiresome. As a result, the number of combined auditory-visual display ratings during the final portion of the experiment was decreased from 18 to 9 in an effort to maintain a higher level of subject interest.

Again, other than the above mentioned changes, the procedure of this experiment is identical to that of the previous experiment. As a result, the same data collection factors and data analysis are used to examine the results.

F. RESULTS AND DISCUSSION

As with the previous experiment, the overall results of this experiment suggest significant auditory-visual cross-modal perception phenomena relevant to VE and multimedia developers. The major findings of this experiment are now discussed.

1. Validity

The first and 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

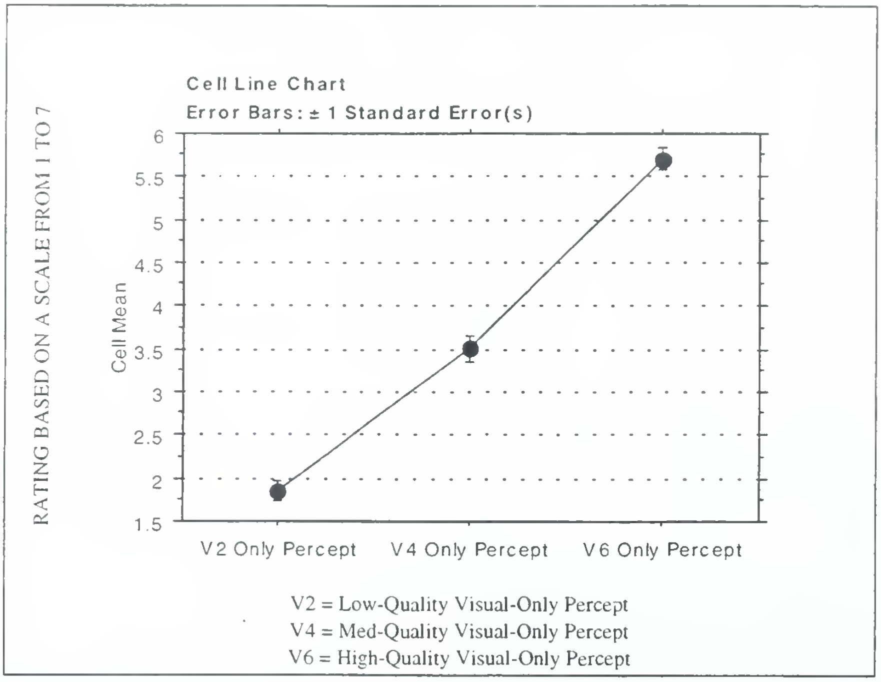


Figure 76. Experiment 2: Visual-Only Quality Percept Ratings.

experiment would be jeopardized. However, in looking at Figure 76, 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. Likewise, in looking at Figure 77, 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 hypotheses can continue.

2. Findings

Figure 78 represents the results of all one sample sign tests based on the first null hypothesis which states: the difference between a) the visual-only quality rating of a combined auditory-visual display, and b) the baseline rating for the visual-only quality

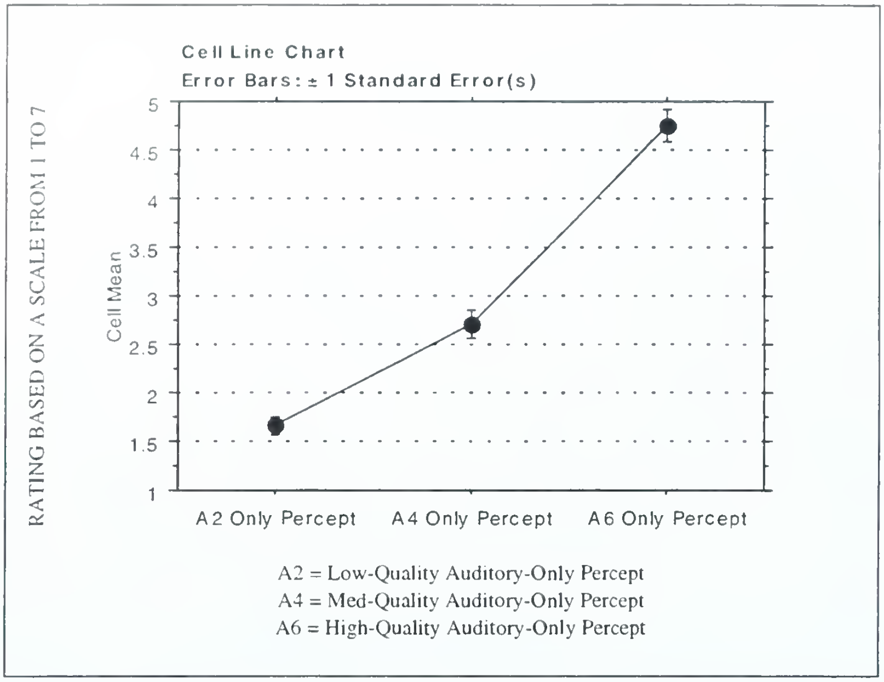


Figure 77. Experiment 2: Auditory-Only Quality Percept Ratings.

display is zero. As one can see from the results, there are no statistically significant findings in any of the quality combinations.

Figure 79 represents the results of all one sample sign tests based on the second null hypothesis which states: the difference between a) the auditory-only quality rating of a combined auditory-visual display, and b) the baseline rating for the auditory-only quality display is zero. As one can see from the results, 1) when presented a combined low-quality auditory and high-quality visual display, when only asked to rate the quality of the auditory display, a statistically significant finding at the .0290 level suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display, and 2) when presented a combined high-quality auditory and high-quality visual display, when only asked to rate the quality of the auditory display, a statistically significant finding at the .0243 level suggests that the quality perception of a

One-Sample Sign Test for V2 A2 AV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	11
# Obs < Hyp. Value	17
# Obs = Hyp. Value	8
P-Value	3449

One-Sample Sign Test for V2 A4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs < Hyp. Value	17
# Obs = Hyp. Value	3
P-Value	>.9999

One-Sample Sign Test for V2 A6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp Value	17
# Obs. < Hyp. Value	13
# Obs = Hyp Value	6
P-Value	.5847

One-Sample Sign Test for V4 A2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	19
# Obs. < Hyp. Value	15
# Obs. = Hyp. Value	2
P-Value	.6076

One-Sample Sign Test for V4 A4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs < Hyp. Value	17
# Obs = Hyp. Value	1
P-Value	>.9999

One-Sample Sign Test for V4 A6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs. < Hyp Value	15
# Obs. = Hyp. Value	3
P-Value	.7283

One-Sample Sign Test for V6 A2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	17
# Obs. < Hyp. Value	11
# Obs. = Hyp. Value	8
P-Value	.3449

One-Sample Sign Test for V6 A4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs. < Hyp. Value	14
# Obs. = Hyp. Value	4
P-Value	5966

One-Sample Sign Test for V6 A6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	12.
# Obs. = Hyp. Value	3
P-Value	.1628

V2A2 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV = High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV = High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV = High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 78. Experiment 2: One Sample Sign Tests for Visual-Only Quality Percept of Combined Auditory-Visual Displays.

high-quality auditory display is increased when coupled with a high-quality visual display.

Figure 80 represents the results of all one sample sign tests based on the third null hypothesis which states: the difference between a) the visual quality rating of a combined auditory-visual display when also rating the auditory display, and b) the baseline rating for the visual-only quality display is zero. As one can see from the results, there are no significant findings at the .05 level. However it is worth mentioning that there are three findings at the .10 level which one can see from the figure.

Figure 81 represents the results of all one sample sign tests based on the fourth null hypothesis which states: the difference between a) the auditory quality rating of a

One-Sample Sign Test for A2 V2 AV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	8
# Obs. < Hyp. Value	18
# Obs. = Hyp Value	10
P-Value	.0755

One-Sample Sign Test for A2 V4 AV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	9
# Obs < Hyp. Value	17
# Obs = Hyp Value	10
P-Value	1686

One-Sample Sign Test for A2 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	7
# Obs < Hyp. Value	19
# Obs. = Hyp. Value	10
P-Value	.0290

One-Sample Sign Test for A4 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	17
# Obs. < Hyp. Value	14
# Obs. = Hyp. Value	5
P-Value	.7201

One-Sample Sign Test for A4 V4 AV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	22
# Obs. < Hyp. Value	11
# Obs = Hyp. Value	3
P-Value	.0801

One-Sample Sign Test for A4 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	12
# Obs. = Hyp. Value	3
P-Value	.1628

One-Sample Sign Test for A6 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	19
# Obs. < Hyp. Value	13
# Obs. = Hyp. Value	4
P-Value	3771

One-Sample Sign Test for A6 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs. < Hyp. Value	13
# Obs. = Hyp. Value	5
P-Value	.4731

One-Sample Sign Test for A6 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	24
# Obs. < Hyp. Value	10
# Obs. = Hyp. Value	2
P-Value	.0243

A2V2 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 79. Experiment 2: One Sample Sign Tests for Auditory-Only Quality Percept of Combined Auditory-Visual Displays.

combined auditory-visual display when also rating the visual display, and b) the baseline rating for the auditory-only quality display is zero. The results suggest that: 1) when presented a combined medium-quality auditory and medium-quality visual display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0029 level suggests that the quality perception of a medium-quality auditory display is increased when coupled with a medium-quality visual display, and 2) when presented a combined high-quality auditory and high-quality visual display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0294 level suggests that the quality perception of a high-quality auditory display is increased when coupled with a high-quality visual display.

One-Sample Sign Test for V2 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	10
# Obs. < Hyp. Value	20
# Obs. = Hyp. Value	6
P-Value	.0987

One-Sample Sign Test for V2 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	20
# Obs. < Hyp. Value	14
# Obs. = Hyp. Value	2
P-Value	3915

One-Sample Sign Test for V2 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	15
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	5
P-Value	>.9999

One-Sample Sign Test for V4 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	19
# Obs. < Hyp. Value	15
# Obs. = Hyp. Value	2
P-Value	.6076

One-Sample Sign Test for V4 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	23
# Obs. < Hyp. Value	11
# Obs. = Hyp. Value	2
P-Value	.0576

One-Sample Sign Test for V4 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	12
# Obs. = Hyp. Value	3
P-Value	.1628

One-Sample Sign Test for V6 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	22
# Obs. < Hyp. Value	11
# Obs. = Hyp. Value	3
P-Value	.0801

One-Sample Sign Test for V6 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs. < Hyp. Value	11
# Obs. = Hyp. Value	7
P-Value	.2649

One-Sample Sign Test for V6 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	17
# Obs. < Hyp. Value	13
# Obs. = Hyp. Value	6
P-Value	.5847

V2A2 CAV = Low-Quality Visual Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 CAV = Low-Quality Visual Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 CAV = Low-Quality Visual Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 CAV = Med-Quality Visual Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 CAV = Med-Quality Visual Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 CAV = Med-Quality Visual Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 CAV = High-Quality Visual Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 CAV = High-Quality Visual Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 CAV = High-Quality Visual Percept of Combined High-Visual and High-Auditory Quality Display

Figure 80. Experiment 2: One Sample Sign Tests for Visual Quality Percept When Also Rating the Auditory Display of Combined Auditory-Visual Displays.

In terms of response times, Figure 82 represents the average visual quality rating response times of a combined auditory-visual display, when only asked to rate the quality of the visual display. Figure 83 represents the average auditory quality rating response times of a combined auditory-visual display, when only asked to rate the quality of the auditory display. Figure 84 represents the average combined auditory and visual quality rating response times of a combined auditory-visual display, when asked to rate both the

One-Sample Sign Test for A2 V2 CAV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	13
# Obs < Hyp. Value	12
# Obs = Hyp. Value	11
P-Value	>.9999

One-Sample Sign Test for A2 V4 CAV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	13
# Obs < Hyp. Value	14
# Obs. = Hyp. Value	9
P-Value	>.9999

One-Sample Sign Test for A2 V6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	15
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	5
P-Value	>.9999

One-Sample Sign Test for A4 V2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	18
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	2
P-Value	.8642

One-Sample Sign Test for A4 V4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	26
# Obs. < Hyp. Value	8
# Obs. = Hyp. Value	2
P-Value	.0029

One-Sample Sign Test for A4 V6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	14
# Obs. = Hyp. Value	1
P-Value	.3105

One-Sample Sign Test for A6 V2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	12
# Obs. = Hyp. Value	3
P-Value	.1628

One-Sample Sign Test for A6 V4 CAV Diff
Hypothesized Value: 0

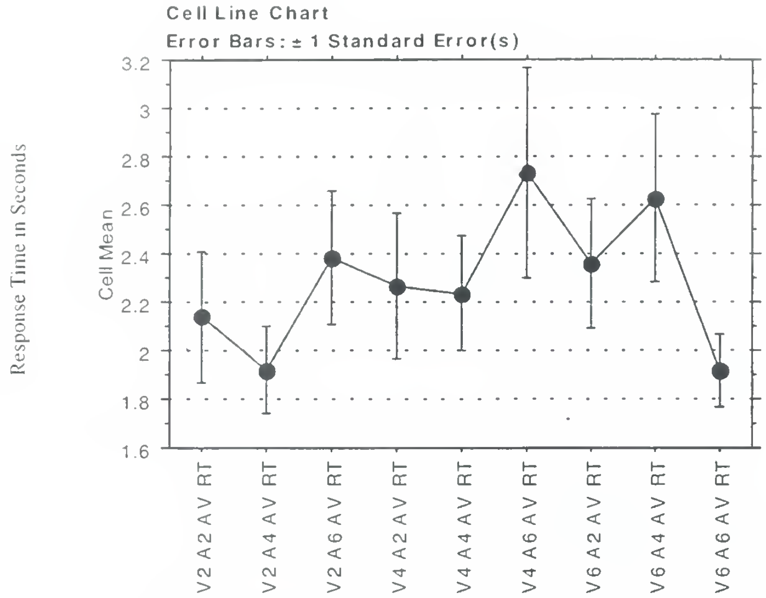
# Obs. > Hyp. Value	16
# Obs. < Hyp. Value	18
# Obs. = Hyp. Value	2
P-Value	.8642

One-Sample Sign Test for A6 V6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	22
# Obs. < Hyp. Value	9
# Obs. = Hyp. Value	5
P-Value	.0294

A2V2 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 CAV = High-Quality Auditory Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 CAV = High-Quality Auditory Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 CAV = High-Quality Auditory Percept of Combined High-Auditory and High-Visual Quality Display

Figure 81. Experiment 2: One Sample Sign Tests for Auditory Quality Percept When Also Rating the Visual Display of Combined Auditory-Visual Displays.

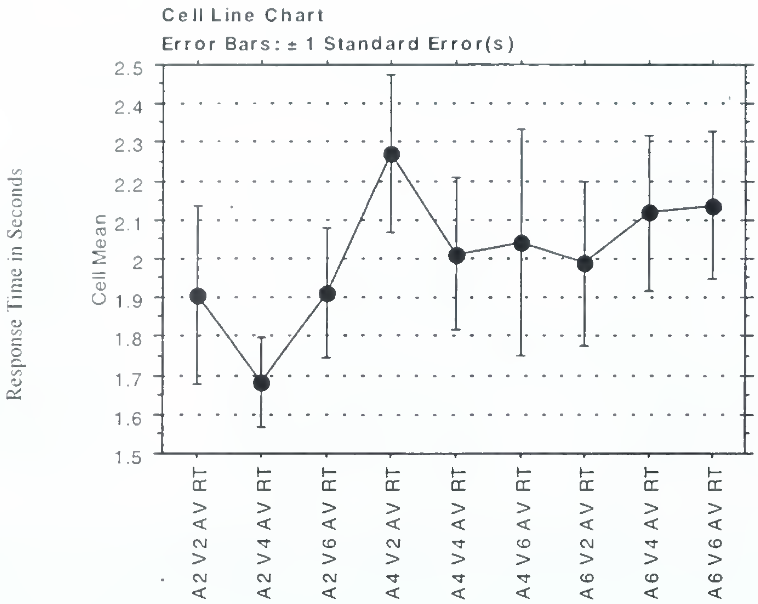


V2A2 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 82. Experiment 2: Visual-Only Quality Rating Response Times of a Combined Auditory-Visual Display.

auditory and visual displays. In looking at the results of the response times, one can see various trends based on a particular auditory-visual quality combination. However, several factors limit the ability to correctly analyze these temporal results in any statistically valid manner. These factors are discussed in the last chapter.

In terms of the post-experiment questions, Figure 85 represents the subject's opinion on 1) how easy or difficult it was to determine the quality of the various displays, and 2) if less or more time was needed to adequately rate the various displays. Keeping in mind that subjects used a Likert rating scale ranging from 1 to 7 (4 being neutral) to rate their opinions, the results indicate that determining the quality of both auditory and visual



A2V2 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 83. Experiment 2: Auditory-Only Quality Rating Response Times of a Combined Auditory-Visual Display.

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 was an adequate amount of time to rate the visual-only and auditory displays, but that slightly more than eight seconds was desired when rating the combined auditory-visual displays.

Finally, the remaining questions of the post-experiment survey reveal that 29 of the 36 subjects (80.1%) focused on alphanumerics to determine the quality of the visual displays, and that only 7 of the 36 subjects (19.4%) felt that they were mentally overloaded when having to rate both auditory and visual displays simultaneously. As in

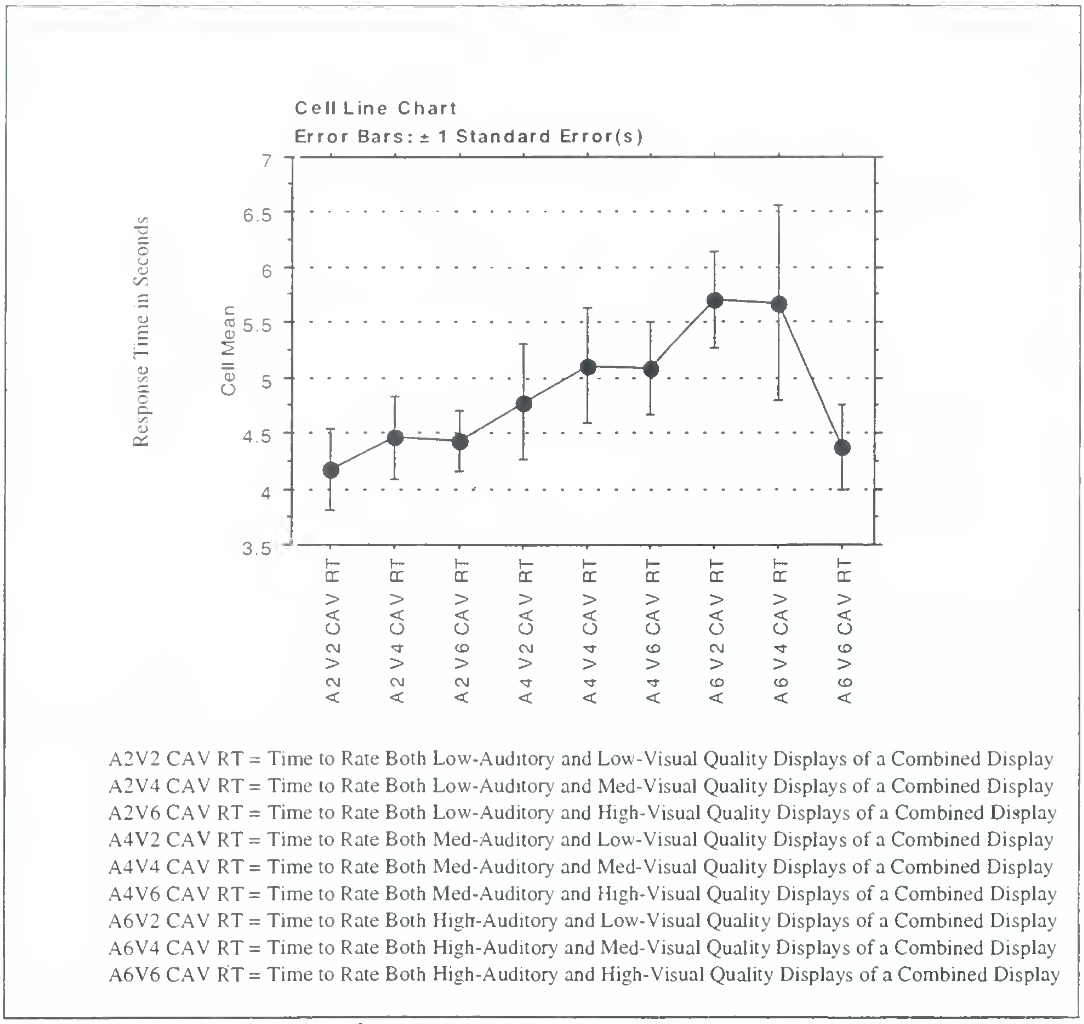
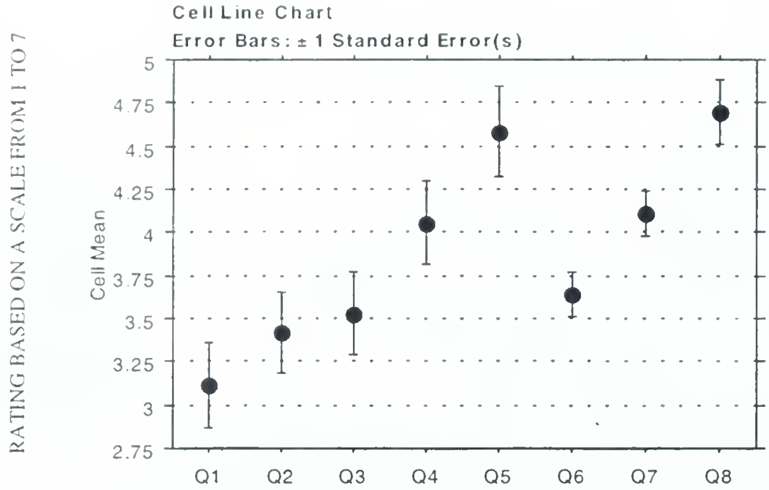


Figure 84. Experiment 2: Response Times of Both Auditory and Visual Displays of a Combined Auditory-Visual Display.

the previous experiment, some very interesting observations were also observed concerning the descriptions that the subjects used to determine the quality of the various displays. These observations are outlined in the final chapter.

G. SUMMARY AND CONCLUSIONS

Overall the findings suggest that whether asked to specifically attend to both auditory and visual modalities, or asked to attend to only one modality, both similar and



Q1 = How easy or difficult was is to determine the quality of the visual-only displays?
Q2 = How easy or difficult was is to determine the quality of the auditory-only displays?
Q3 = How easy or difficult was is to determine the visual quality of the auditory-visual displays?
Q4 = How easy or difficult was is to determine the auditory quality of the auditory-visual displays?
Q5 = How easy or difficult was to determine both the auditory and visual qualities of the auditory-visual displays?
Q6 = Would you have liked less or more time to view the visual-only displays?
Q7 = Would you have liked less or more time to hear the auditory-only displays?
Q8 = Would you have liked less or more time to hear-view the combined auditory-visual displays?

Figure 85. Experiment 2: Post-Experiment Questions 1 - 8.

dissimilar cross-modal auditory-visual perception phenomena exist. These findings suggest that when manipulating both visual and auditory display Gaussian noise level:

1) 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.

2) 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.

3) 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 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 were not semantically associated with each other? The next chapter describes the final experiment of this research effort which investigates the answer to this question.

IX. EXPERIMENT 3: STATIC RESOLUTION NONALPHANUMERIC

A. 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 the aforementioned fruit-flower scene depicted in Chapter IV, Figure 33 and the auditory display is a selection of music. As in the previous experiments, the goal of this experiment is to investigate 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?

3) Does a low-quality auditory display coupled with a low-quality visual display cause a decrease/increase 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) Does a high-quality auditory display coupled with a high-quality visual display cause an increase/decrease 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?

B. LOCATION

The location and ambient conditions for this experiment were identical to that of the previous experiment, Experiment 2: Static Noise. See Chapter VIII, Section B.

C. PARTICIPANTS

A total of 36 volunteer participants (14 Male, 22 Female) comprised from the students, faculty, staff, and guests of NPS served as subjects. 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 were required to have 20/20 or corrected 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 were not needed. Before conducting the experiment, each subject was asked, as part of a voluntary consent form, if he or she met the vision and hearing requirements.

D. APPARATUS

The apparatus used in this experiment is identical to that of the first two experiments: Experiment 1: Static Resolution and Experiment 2: Static Noise. See Chapter VII, Section D.

E. PROCEDURE

The procedure of this experiment is identical to that of the previous experiment, Experiment 2: Static Noise. The experiment involved a 3x3 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 fruit-flower scene depicted earlier in Chapter IV, Figure 33 having resolutions of 34 pixels/inch, 50 pixels/inch, 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 such, 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. As such, the visual display parameters manipulated are pixel resolution, and the auditory display parameters manipulated are sampling frequency. During the experiment which lasts approximately 30 minutes, each subject wears headphones and sits in front of a 20-inch computer display monitor. The task of the subject is to rate the perceived quality of auditory-only, visual-only, and auditory-visual displays via Likert rating scales ranging from 1 (low) to 7 (high).

The lowest and highest quality auditory displays in which the subjects were supposed to memorize during the self-calibration phase corresponded to the music selection at 8 kHz and 44.1 kHz respectively. The lowest and highest quality visual displays in which the subjects were supposed to memorize during the self-calibration phase are depicted in Figure 86 and Figure 87 respectively. The low-quality visual display has a resolution of 28 pixels/inch; whereas the high-quality visual display has a resolution of 72 pixels/inch. Again, it is important to remember that the original displays were depicted in color, and that the actual pixel resolution experienced by the subject can only be viewed on the actual 20 inch computer monitor. However, the low- and high-quality displays depicted in Figure 86 and Figure 87 are fairly good representations of the quality difference between the actual displays used in the experiment. Besides the different auditory and visual stimuli utilized, the procedure continues exactly as in the previous experiment. As a result, the same data collection factors and data analysis are used to examine the results.

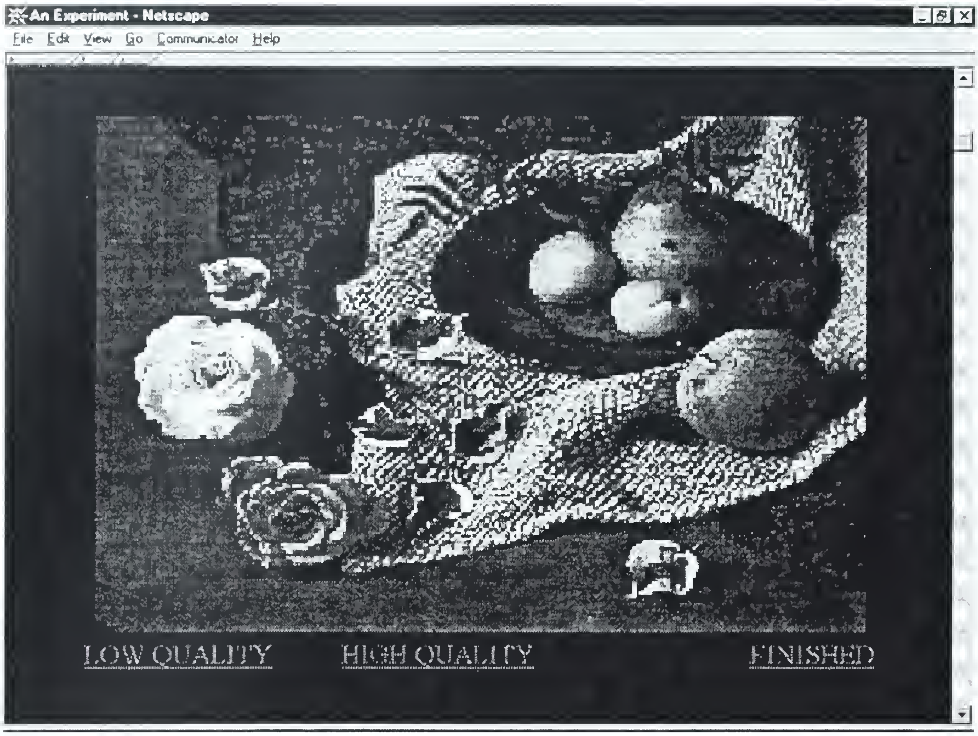


Figure 86. Experiment 3: Low-Quality Visual Display Familiarization.

Figure 87. Experiment 3: High-Quality Visual Display Familiarization.

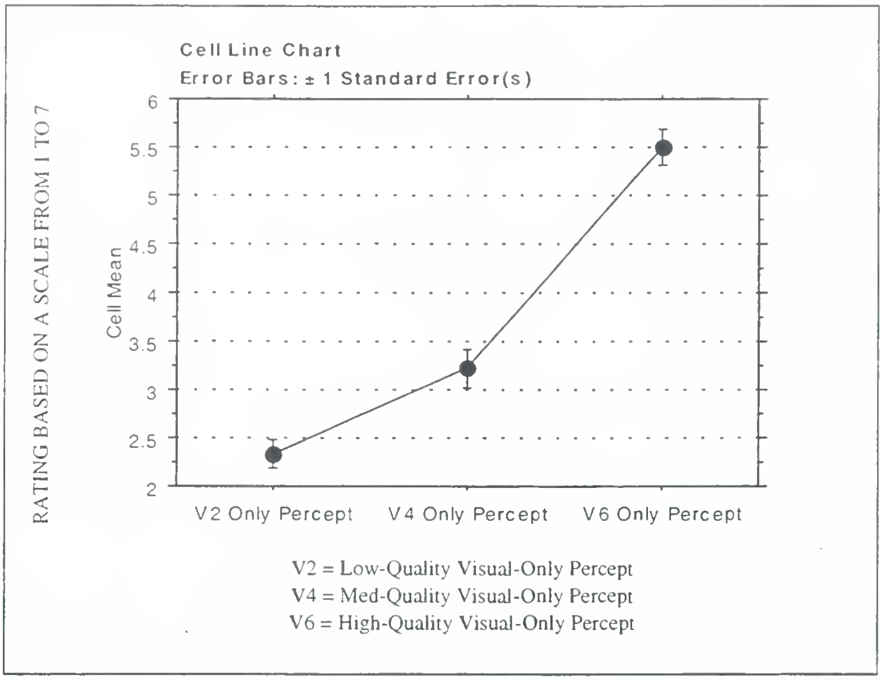


Figure 88. Experiment 3: Visual-Only Quality Percept Ratings.

F. RESULTS AND DISCUSSION

As with the previous experiment, the overall results of this experiment suggest significant auditory-visual cross-modal perception phenomena relevant to VE and multimedia developers. The major findings of this experiment are now discussed.

1. Validity

As with the previous experiments, the first and 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 88, 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. As such, a lack of alphanumeric features has no affect on the overall

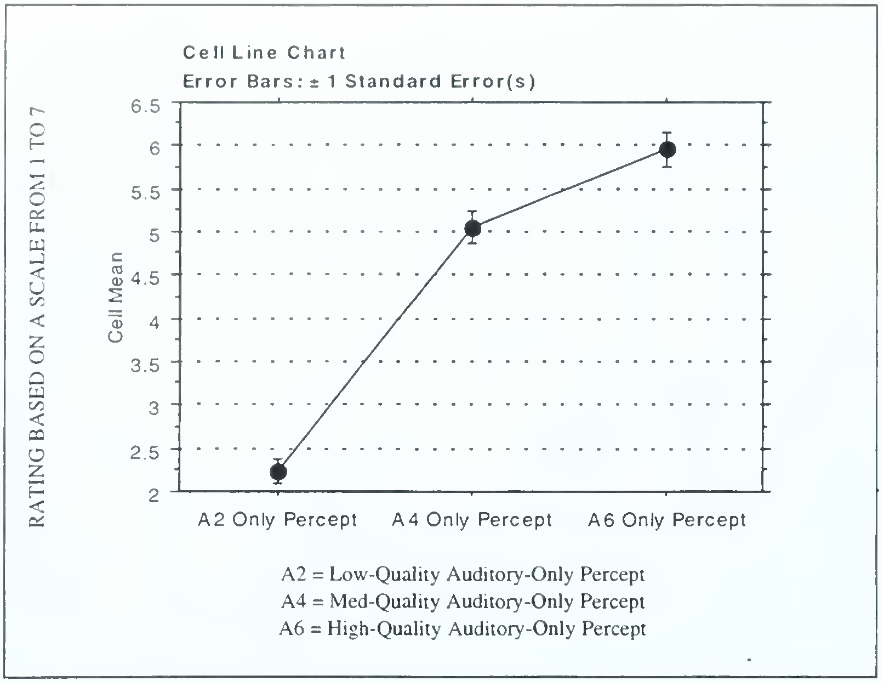


Figure 89. Experiment 3: Auditory-Only Quality Percept Ratings.

ability of the subjects to determine the quality of the visual displays. Likewise, in looking at Figure 89, 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 hypotheses can continue.

2. Findings

Figure 90 represents the results of all one sample sign tests based on the first null hypothesis which states: the difference between a) the visual-only quality rating of a combined auditory-visual display, and b) the baseline rating for the visual-only quality display is zero. As one can see from the results, 1) when presented a combined high-quality visual and medium-quality auditory display, when only asked to rate the quality of the visual display, a statistically significant finding at the .0201 level suggests that the quality perception of a high-quality visual display is increased when coupled with a

One-Sample Sign Test for V2 A2 AV Diff		One-Sample Sign Test for V2 A4 AV Diff		One-Sample Sign Test for V2 A6 AV Diff	
Hypothesized Value: 0		Hypothesized Value: 0		Hypothesized Value: 0	
# Obs. > Hyp Value	14	# Obs. > Hyp. Value	13	# Obs. > Hyp Value	17
# Obs < Hyp Value	16	# Obs. < Hyp. Value	15	# Obs. < Hyp Value	13
# Obs. = Hyp. Value	6	# Obs. = Hyp Value	8	# Obs. = Hyp Value	6
P-Value	.8555	P-Value	.8506	P-Value	.5847

One-Sample Sign Test for V4 A2 AV Diff		One-Sample Sign Test for V4 A4 AV Diff		One-Sample Sign Test for V4 A6 AV Diff	
Hypothesized Value: 0		Hypothesized Value: 0		Hypothesized Value: 0	
# Obs. > Hyp. Value	18	# Obs. > Hyp Value	11	# Obs. > Hyp Value	18
# Obs. < Hyp. Value	12	# Obs. < Hyp Value	17	# Obs. < Hyp Value	13
# Obs = Hyp. Value	6	# Obs. = Hyp Value	8	# Obs. = Hyp. Value	5
P-Value	.3616	P-Value	.3449	P-Value	.4731

One-Sample Sign Test for V6 A2 AV Diff		One-Sample Sign Test for V6 A4 AV Diff		One-Sample Sign Test for V6 A6 AV Diff	
Hypothesized Value: 0		Hypothesized Value: 0		Hypothesized Value: 0	
# Obs. > Hyp. Value	20	# Obs. > Hyp. Value	23	# Obs. > Hyp. Value	22
# Obs. < Hyp. Value	9	# Obs. < Hyp. Value	9	# Obs. < Hyp Value	8
# Obs. = Hyp. Value	7	# Obs. = Hyp Value	4	# Obs. = Hyp Value	6
P-Value	.0614	P-Value	.0201	P-Value	.0161

V2A2 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV = High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV = High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV = High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 90. Experiment 3: One Sample Sign Tests for Visual-Only Quality Percept of Combined Auditory-Visual Displays.

medium-quality auditory display, and 2) when presented a combined high-quality visual and high-quality auditory display, when only asked to rate the quality of the visual display, a statistically significant finding at the .0161 level suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

Figure 91 represents the results of all one sample sign tests based on the second null hypothesis which states: the difference between a) the auditory-only quality rating of a combined auditory-visual display, and b) the baseline rating for the auditory-only quality display is zero. As one can see from the results, there are no statistically significant findings in any of the quality combinations.

One-Sample Sign Test for A2 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	15
# Obs. < Hyp. Value	11
# Obs. = Hyp. Value	10
P-Value	.5572

One-Sample Sign Test for A2 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	11
# Obs. < Hyp. Value	15
# Obs. = Hyp. Value	10
P-Value	.5572

One-Sample Sign Test for A2 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	17
# Obs. = Hyp. Value	5
P-Value	.7201

One-Sample Sign Test for A4 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	14
# Obs. = Hyp. Value	8
P-Value	>.9999

One-Sample Sign Test for A4 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	13
# Obs. < Hyp. Value	19
# Obs. = Hyp. Value	4
P-Value	.3771

One-Sample Sign Test for A4 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	17
# Obs. = Hyp. Value	5
P-Value	.7201

One-Sample Sign Test for A6 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	13
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	7
P-Value	.7111

One-Sample Sign Test for A6 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	13
# Obs. < Hyp. Value	15
# Obs. = Hyp. Value	8
P-Value	.8506

One-Sample Sign Test for A6 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	12
# Obs. < Hyp. Value	13
# Obs. = Hyp. Value	11
P-Value	>.9999

A2V2 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 91. Experiment 3: One Sample Sign Tests for Auditory-Only Quality Percept of Combined Auditory-Visual Displays.

Figure 92 represents the results of all one sample sign tests based on the third null hypothesis which states: the difference between a) the visual quality rating of a combined auditory-visual display when also rating the auditory display, and b) the baseline rating for the visual-only quality display is zero. As one can see from the results, when presented a combined high-quality visual and high-quality auditory display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0125 level suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

Figure 93 represents the results of all one sample sign tests based on the fourth null hypothesis which states: the difference between a) the auditory quality rating of a

One-Sample Sign Test for V2 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	18
# Obs. = Hyp. Value	4
P-Value	.5966

One-Sample Sign Test for V2 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	16
# Obs. < Hyp. Value	13
# Obs. = Hyp. Value	7
P-Value	7111

One-Sample Sign Test for V2 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	17
# Obs. = Hyp. Value	5
P-Value	7201

One-Sample Sign Test for V4 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	12
# Obs. < Hyp. Value	17
# Obs. = Hyp. Value	7
P-Value	.4583

One-Sample Sign Test for V4 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	12
# Obs. < Hyp. Value	19
# Obs. = Hyp. Value	5
P-Value	.2810

One-Sample Sign Test for V4 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	14
# Obs. < Hyp. Value	18
# Obs. = Hyp. Value	4
P-Value	.5966

One-Sample Sign Test for V6 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	19
# Obs. < Hyp. Value	12
# Obs. = Hyp. Value	5
P-Value	.2810

One-Sample Sign Test for V6 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	10
# Obs. = Hyp. Value	5
P-Value	.0708

One-Sample Sign Test for V6 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	21
# Obs. < Hyp. Value	7
# Obs. = Hyp. Value	8
P-Value	.0125

V2A2 CAV = Low-Quality Visual Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 CAV = Low-Quality Visual Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 CAV = Low-Quality Visual Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 CAV = Med-Quality Visual Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 CAV = Med-Quality Visual Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 CAV = Med-Quality Visual Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 CAV = High-Quality Visual Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 CAV = High-Quality Visual Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 CAV = High-Quality Visual Percept of Combined High-Visual and High-Auditory Quality Display

Figure 92. Experiment 3: One Sample Sign Tests for Visual Quality Percept When Also Rating the Auditory Display of Combined Auditory-Visual Displays.

combined auditory-visual display when also rating the visual display, and b) the baseline rating for the auditory-only quality display is zero. The results suggest that when presented a combined medium-quality auditory and low-quality visual display, when asked to rate both auditory and visual displays, a statistically significant finding at the .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 response times, Figure 94 represents the average visual quality rating response times of a combined auditory-visual display, when only asked to rate the quality of the visual display. Figure 95 represents the average auditory quality rating response

One-Sample Sign Test for A2 V2 CAV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	16
# Obs. < Hyp. Value	14
# Obs = Hyp. Value	6
P-Value	.8555

One-Sample Sign Test for A2 V4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp Value	21
# Obs. < Hyp. Value	10
# Obs. = Hyp. Value	5
P-Value	.0708

One-Sample Sign Test for A2 V6 CAV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	13
# Obs. < Hyp. Value	16
# Obs. = Hyp. Value	7
P-Value	.7111

One-Sample Sign Test for A4 V2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	10
# Obs. < Hyp. Value	23
# Obs. = Hyp. Value	3
P-Value	.0351

One-Sample Sign Test for A4 V4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	12
# Obs. < Hyp. Value	22
# Obs. = Hyp. Value	2
P-Value	.1214

One-Sample Sign Test for A4 V6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	15
# Obs. < Hyp. Value	17
# Obs. = Hyp. Value	4
P-Value	.8601

One-Sample Sign Test for A6 V2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	11
# Obs. < Hyp. Value	18
# Obs. = Hyp. Value	7
P-Value	.2649

One-Sample Sign Test for A6 V4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	11
# Obs. < Hyp. Value	18
# Obs. = Hyp. Value	7
P-Value	.2649

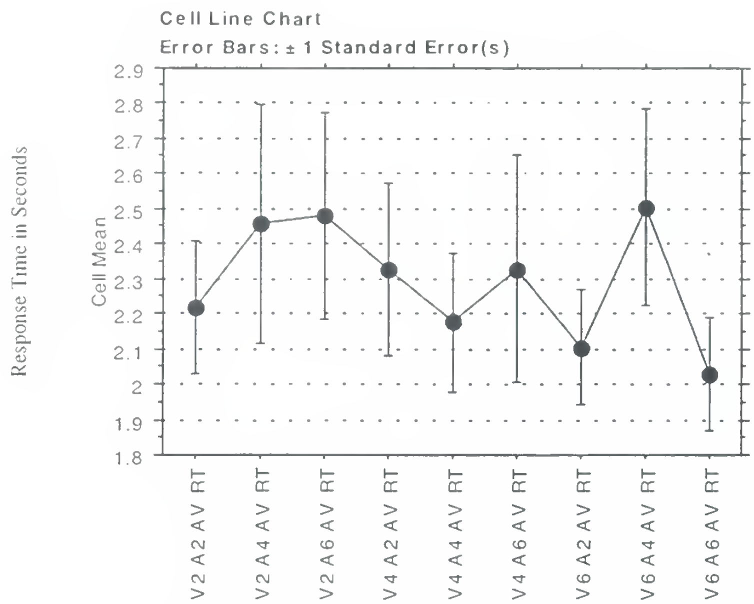
One-Sample Sign Test for A6 V6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	10
# Obs. < Hyp. Value	19
# Obs. = Hyp. Value	7
P-Value	.1360

A2V2 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 CAV = High-Quality Auditory Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 CAV = High-Quality Auditory Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 CAV = High-Quality Auditory Percept of Combined High-Auditory and High-Visual Quality Display

Figure 93. Experiment 3: One Sample Sign Tests for Auditory Quality Percept When Also Rating the Visual Display of Combined Auditory-Visual Displays.

times of a combined auditory-visual display, when only asked to rate the quality of the auditory display. Figure 96 represents the average combined auditory and visual quality rating response times of a combined auditory-visual display, when asked to rate both the

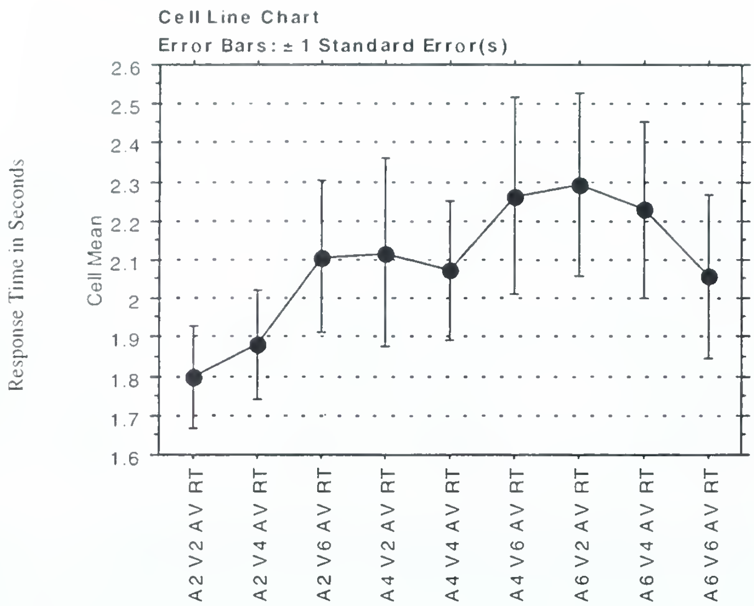


V2A2 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 94. Experiment 3: Visual-Only Quality Rating Response Times of a Combined Auditory-Visual Display.

auditory and visual displays. In looking at the results of the response times, one can see various trends based on a particular auditory-visual quality combination. However, several factors limit the ability to correctly analyze these temporal results in any statistically valid manner. These factors are discussed in the last chapter.

In terms of the post-experiment questions, Figure 97 represents the subject's opinion on 1) how easy or difficult it was to determine the quality of the various displays, and 2) if less or more time was needed to adequately rate the various displays. Keeping in mind that subjects used a Likert rating scale ranging from 1 to 7 (4 being neutral) to rate their opinions, the results indicate that determining the quality of both auditory and visual

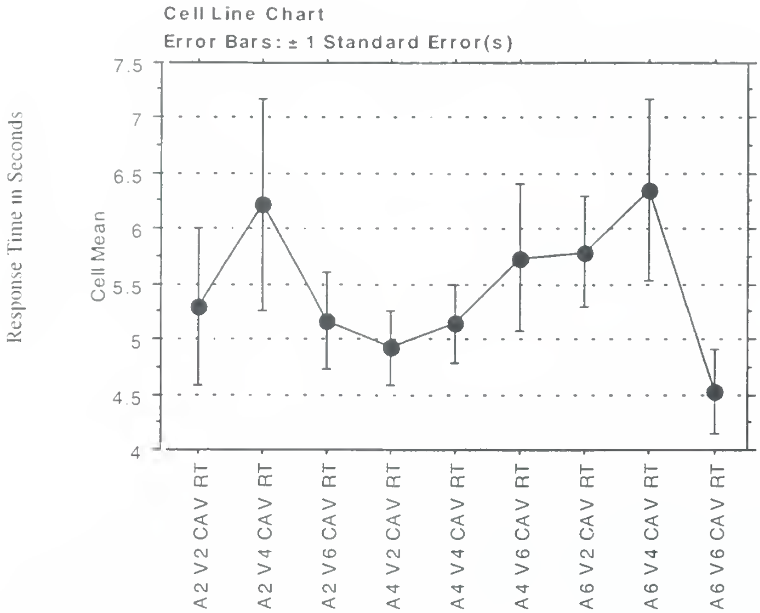


A2V2 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 95. Experiment 3: Auditory-Only Quality Rating Response Times of a Combined Auditory-Visual Display.

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 was an adequate amount of time to rate the visual-only and auditory displays, but that slightly more than eight seconds was desired when rating the combined auditory-visual displays.

Finally, the remaining questions of the post-experiment survey reveal that only 9 of the 36 subjects (25.0%) felt that they were mentally overloaded when having to rate both auditory and visual displays simultaneously. As in the previous experiment, some very interesting observations were also observed concerning the descriptions that the



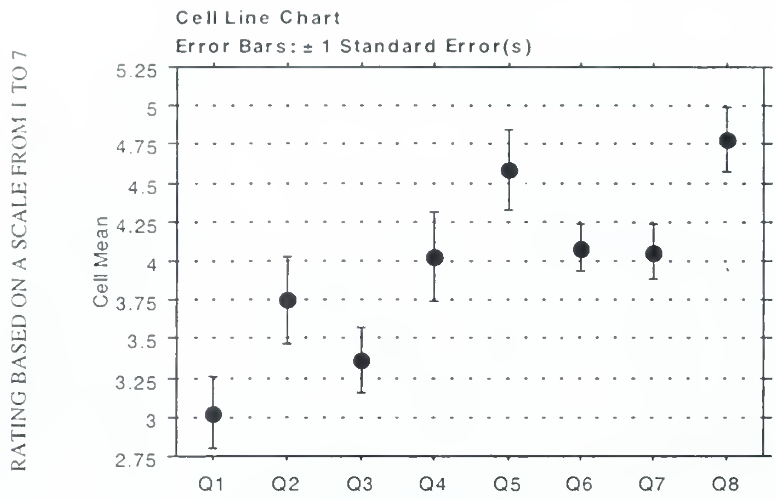
A2V2 CAV RT = Time to Rate Both Low-Auditory and Low-Visual Quality Displays of a Combined Display
A2V4 CAV RT = Time to Rate Both Low-Auditory and Med-Visual Quality Displays of a Combined Display
A2V6 CAV RT = Time to Rate Both Low-Auditory and High-Visual Quality Displays of a Combined Display
A4V2 CAV RT = Time to Rate Both Med-Auditory and Low-Visual Quality Displays of a Combined Display
A4V4 CAV RT = Time to Rate Both Med-Auditory and Med-Visual Quality Displays of a Combined Display
A4V6 CAV RT = Time to Rate Both Med-Auditory and High-Visual Quality Displays of a Combined Display
A6V2 CAV RT = Time to Rate Both High-Auditory and Low-Visual Quality Displays of a Combined Display
A6V4 CAV RT = Time to Rate Both High-Auditory and Med-Visual Quality Displays of a Combined Display
A6V6 CAV RT = Time to Rate Both High-Auditory and High-Visual Quality Displays of a Combined Display

Figure 96. Experiment 3: Response Times of Both Auditory and Visual Displays of a Combined Auditory-Visual Display.

subjects used to determine the quality of the various displays. These observations are outlined in the final chapter.

G. SUMMARY AND CONCLUSIONS

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



Q1 = How easy or difficult was is to determine the quality of the visual-only displays?
Q2 = How easy or difficult was is to determine the quality of the auditory-only displays?
Q3 = How easy or difficult was is to determine the visual quality of the auditory-visual displays?
Q4 = How easy or difficult was is to determine the auditory quality of the auditory-visual displays?
Q5 = How easy or difficult was to determine both the auditory and visual qualities of the auditory-visual displays?
Q6 = Would you have liked less or more time to view the visual-only displays?
Q7 = Would you have liked less or more time to hear the auditory-only displays?
Q8 = Would you have liked less or more time to hear-view the combined auditory-visual displays?

Figure 97. Experiment 3: Post-Experiment Questions 1 - 8.

suggest that when manipulating visual display pixel resolution and auditory display sampling frequency:

1) 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.

2) 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.

3) 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 were not perceptually tightly coupled auditory-visual displays as in the first two experiment, the results indicate that the effects of auditory-visual cross-modal perception phenomena persist. The next chapter presents an overview of the combined results from all three experiments.

X. SUMMARY AND CONCLUSIONS

A. INTRODUCTION

This chapter represents the culmination of two and a half years of research and development in support of evidence concerning auditory-visual cross-modal perception phenomena. The overall results, conclusions, impact, observations, recommendations, future work, and final thoughts are presented.

B. OVERALL RESULTS

Because all collected data were derived from identical experimental conditions based on the same low-, medium-, and high-quality ordering of the auditory and visual stimuli, combining datasets from all three experiments is justified in order to consider overall results. As such, the following are the overall results from combining the datasets from all three experiments.

1. Participants

Overall a total of 108 volunteer participants (59 Male, 49 Female) comprised from the students, faculty, staff, and guests of NPS served as subjects. The overall average age of the subjects is 36.1 years ranging in age from 11 to 63 (four female subjects did not give their age). All subjects were required to have 20/20 or corrected 20/20 vision and normal hearing. As such, before conducting the experiment, each subject was asked, as part of a voluntary consent form, if he or she met the vision and hearing requirements.

2. Validity

Again, the first and most important consideration is whether the overall quality of the visual and auditory displays are rank ordered by the subjects according to their intended rankings. In looking at Figure 98, one can see that the overall quality ratings of

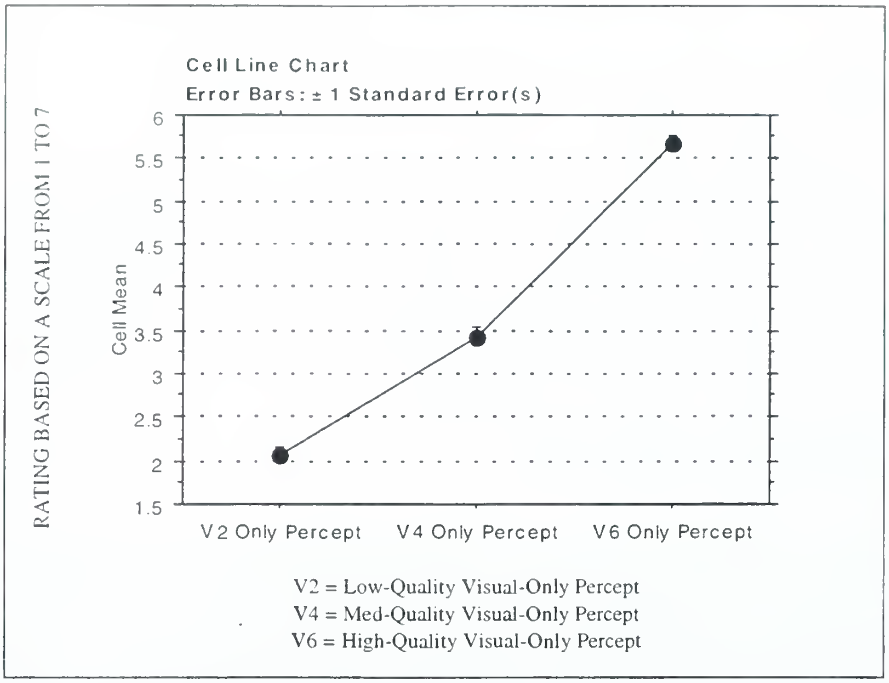


Figure 98. Combined Data: Visual-Only Quality Percept Ratings.

the visual displays are properly rank ordered by the subjects. Likewise, in looking at Figure 99, one can see that the overall quality ratings of the auditory displays are properly rank ordered by the subjects. Given that the data regarding quality of all displays are properly rank ordered, data analysis with respect to the hypotheses can continue.

3. Overall Findings

Figure 100 represents the results of all one sample sign tests based on the first null hypothesis which states: the difference between a) the visual-only quality rating of a combined auditory-visual display, and b) the baseline rating for the visual-only quality display is zero. As one can see from the results, 1) when presented a combined high-quality visual and medium-quality auditory display, when only asked to rate the quality of the visual display, a statistically significant finding at the .0124 level suggests that the quality perception of a high-quality visual display is increased when coupled with a medium-quality auditory display, and 2) when presented a combined high-quality visual

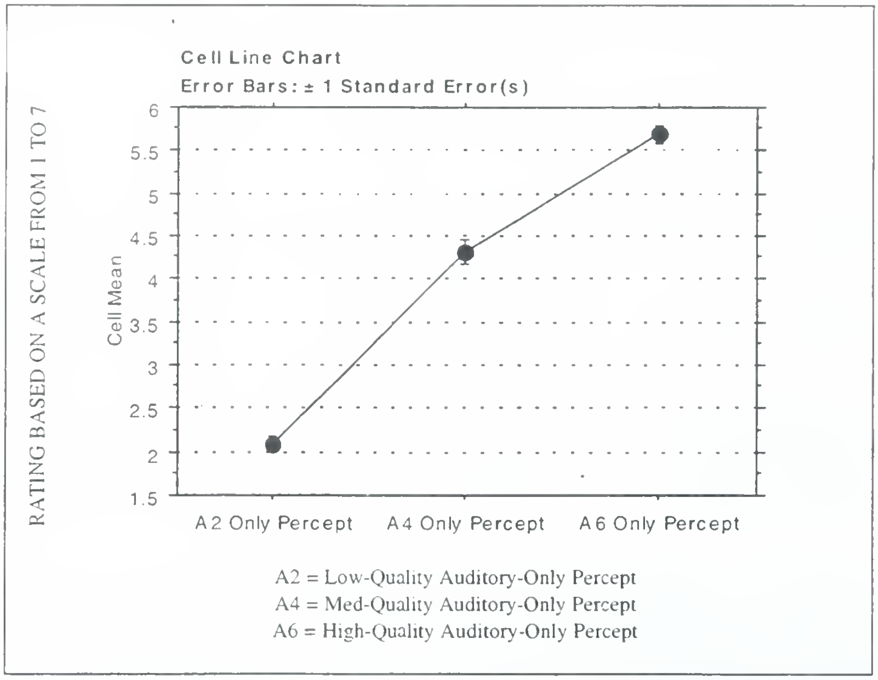


Figure 99. Combined Data: Auditory-Only Quality Percept Ratings.

and high-quality auditory display, when only asked to rate the quality of the visual display, a statistically significant finding at the .0002 level strongly suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

Figure 101 represents the results of all one sample sign tests based on the second null hypothesis which states: the difference between a) the auditory-only quality rating of a combined auditory-visual display, and b) the baseline rating for the auditory-only quality display is zero. As one can see from the results, 1) when presented a combined low-quality auditory and medium-quality visual display, when only asked to rate the quality of the auditory display, a statistically significant finding at the .0375 level suggests that the quality perception of a low-quality auditory display is decreased when coupled with a medium-quality visual display, and 2) when presented a combined low-quality auditory and high-quality visual display, when only asked to rate the quality of

One-Sample Sign Test for V2 A2 AV Diff
Hypothesized Value: 0

# Obs > Hyp Value	35
# Obs < Hyp Value	50
# Obs = Hyp Value	23
P-Value	.1284

One-Sample Sign Test for V2 A4 AV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	43
# Obs < Hyp. Value	48
# Obs. = Hyp Value	17
P-Value	6752

One-Sample Sign Test for V2 A6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	49
# Obs. < Hyp. Value	39
# Obs. = Hyp Value	20
P-Value	.3374

One-Sample Sign Test for V4 A2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	52
# Obs. < Hyp. Value	45
# Obs. = Hyp. Value	11
P-Value	.5426

One-Sample Sign Test for V4 A4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	43
# Obs. < Hyp. Value	50
# Obs. = Hyp. Value	15
P-Value	.5341

One-Sample Sign Test for V4 A6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	51
# Obs. < Hyp. Value	43
# Obs. = Hyp. Value	14
P-Value	.4705

One-Sample Sign Test for V6 A2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	53
# Obs. < Hyp. Value	36
# Obs. = Hyp. Value	19
P-Value	.0893

One-Sample Sign Test for V6 A4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	59
# Obs. < Hyp. Value	34
# Obs. = Hyp. Value	15
P-Value	.0124

One-Sample Sign Test for V6 A6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	65
# Obs. < Hyp. Value	28
# Obs. = Hyp. Value	15
P-Value	.0002

V2A2 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV = Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV = Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV = High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV = High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV = High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 100. Combined Data: One Sample Sign Tests for Visual-Only Quality Percept of Combined Auditory-Visual Displays.

the auditory display, a statistically significant finding at the .0002 level strongly suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display.

Figure 102 represents the results of all one sample sign tests based on the third null hypothesis which states: the difference between a) the visual quality rating of a combined auditory-visual display when also rating the auditory display, and b) the baseline rating for the visual-only quality display is zero. As one can see from the results, 1) when presented a combined high-quality visual and low-quality auditory display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0172 level suggests that the quality perception of a high-quality visual display is

One-Sample Sign Test for A2 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	39
# Obs. < Hyp. Value	45
# Obs. = Hyp. Value	24
P-Value	.5856

One-Sample Sign Test for A2 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	32
# Obs. < Hyp. Value	52
# Obs. = Hyp. Value	24
P-Value	.0375

One-Sample Sign Test for A2 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	26
# Obs. < Hyp. Value	62
# Obs. = Hyp. Value	20
P-Value	.0002

One-Sample Sign Test for A4 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	47
# Obs. < Hyp. Value	44
# Obs. = Hyp. Value	17
P-Value	.8341

One-Sample Sign Test for A4 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	47
# Obs. < Hyp. Value	49
# Obs. = Hyp. Value	12
P-Value	.9188

One-Sample Sign Test for A4 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	53
# Obs. < Hyp. Value	41
# Obs. = Hyp. Value	14
P-Value	.2564

One-Sample Sign Test for A6 V2 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	46
# Obs. < Hyp. Value	45
# Obs. = Hyp. Value	17
P-Value	>.9999

One-Sample Sign Test for A6 V4 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	49
# Obs. < Hyp. Value	40
# Obs. = Hyp. Value	19
P-Value	.3966

One-Sample Sign Test for A6 V6 AV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	51
# Obs. < Hyp. Value	34
# Obs. = Hyp. Value	23
P-Value	.0821

A2V2 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV = Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV = Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV = High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 101. Combined Data: One Sample Sign Tests for Auditory-Only Quality Percept of Combined Auditory-Visual Displays.

increased when coupled with a low-quality auditory display, and 2) when presented a combined high-quality visual and medium-quality auditory display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0042 level strongly suggests that the quality perception of a high-quality visual display is increased when coupled with a medium-quality auditory display, and 3) when presented a combined high-quality visual and high-quality auditory display, when asked to rate both auditory and visual displays, a statistically significant finding at the .0034 level strongly suggests that the quality perception of a high-quality visual display is increased when coupled with a high-quality auditory display.

One-Sample Sign Test for V2 A2 CAV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	36
# Obs < Hyp. Value	53
# Obs = Hyp. Value	19
P-Value	.0893

One-Sample Sign Test for V2 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	48
# Obs. < Hyp. Value	42
# Obs = Hyp. Value	18
P-Value	.5984

One-Sample Sign Test for V2 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	41
# Obs. < Hyp. Value	49
# Obs. = Hyp. Value	18
P-Value	.4608

One-Sample Sign Test for V4 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	42
# Obs. < Hyp. Value	52
# Obs. = Hyp. Value	14
P-Value	.3533

One-Sample Sign Test for V4 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	53
# Obs. < Hyp. Value	39
# Obs. = Hyp. Value	16
P-Value	.1750

One-Sample Sign Test for V4 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	52
# Obs. < Hyp. Value	45
# Obs. = Hyp. Value	11
P-Value	.5426

One-Sample Sign Test for V6 A2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	59
# Obs. < Hyp. Value	35
# Obs. = Hyp. Value	14
P-Value	.0172

One-Sample Sign Test for V6 A4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	59
# Obs. < Hyp. Value	31
# Obs. = Hyp. Value	18
P-Value	.0042

One-Sample Sign Test for V6 A6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	57
# Obs. < Hyp. Value	29
# Obs. = Hyp. Value	22
P-Value	.0034

V2A2 CAV = Low-Quality Visual Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 CAV = Low-Quality Visual Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 CAV = Low-Quality Visual Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 CAV = Med-Quality Visual Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 CAV = Med-Quality Visual Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 CAV = Med-Quality Visual Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 CAV = High-Quality Visual Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 CAV = High-Quality Visual Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 CAV = High-Quality Visual Percept of Combined High-Visual and High-Auditory Quality Display

Figure 102. Combined Data: One Sample Sign Tests for Visual Quality Percept When Also Rating the Auditory Display of Combined Auditory-Visual Displays.

Figure 103 represents the results of all one sample sign tests based on the fourth null hypothesis which states: the difference between a) the auditory quality rating of a combined auditory-visual display when also rating the visual display, and b) the baseline rating for the auditory-only quality display is zero. The results suggest that there are no statistically significant findings in any of the quality combinations. However, it is worth mentioning that when presented a combined low-quality auditory and high-quality visual display, when asked to rate both auditory and visual displays, the results at the .0586 level suggests that the quality perception of a low-quality auditory display is decreased when coupled with a high-quality visual display.

One-Sample Sign Test for A2 V2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	46
# Obs. < Hyp. Value	41
# Obs. = Hyp. Value	21
P-Value	6683

One-Sample Sign Test for A2 V4 CAV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	50
# Obs. < Hyp. Value	43
# Obs = Hyp. Value	15
P-Value	.5341

One-Sample Sign Test for A2 V6 CAV Diff
Hypothesized Value: 0

# Obs > Hyp. Value	36
# Obs. < Hyp. Value	55
# Obs. = Hyp Value	17
P-Value	.0586

One-Sample Sign Test for A4 V2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp Value	44
# Obs. < Hyp. Value	53
# Obs. = Hyp. Value	11
P-Value	4168

One-Sample Sign Test for A4 V4 CAV Diff
Hypothesized Value: 0

# Obs > Hyp Value	50
# Obs. < Hyp. Value	50
# Obs. = Hyp. Value	8
P-Value	>.9999

One-Sample Sign Test for A4 V6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	52
# Obs. < Hyp. Value	47
# Obs. = Hyp. Value	9
P-Value	.6879

One-Sample Sign Test for A6 V2 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	53
# Obs. < Hyp. Value	38
# Obs. = Hyp. Value	17
P-Value	1418

One-Sample Sign Test for A6 V4 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	43
# Obs. < Hyp. Value	49
# Obs = Hyp. Value	16
P-Value	.6024

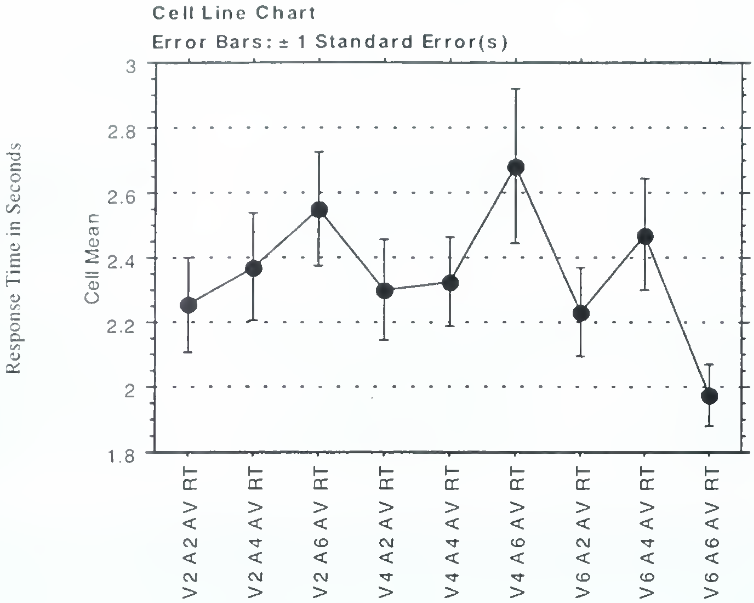
One-Sample Sign Test for A6 V6 CAV Diff
Hypothesized Value: 0

# Obs. > Hyp. Value	46
# Obs. < Hyp. Value	44
# Obs. = Hyp. Value	18
P-Value	.9161

A2V2 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 CAV = Low-Quality Auditory Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 CAV = Med-Quality Auditory Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 CAV = High-Quality Auditory Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 CAV = High-Quality Auditory Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 CAV = High-Quality Auditory Percept of Combined High-Auditory and High-Visual Quality Display

Figure 103. Combined Data: One Sample Sign Tests for Auditory Quality Percept When Also Rating the Visual Display of Combined Auditory-Visual Displays.

In terms of response times, Figure 104 represents the overall average visual quality rating response times of a combined auditory-visual display, when only asked to rate the quality of the visual display. Figure 105 represents the overall average auditory quality rating response times of a combined auditory-visual display, when only asked to rate the quality of the auditory display. Figure 106 represents the overall average combined auditory and visual quality rating response times of a combined auditory-visual

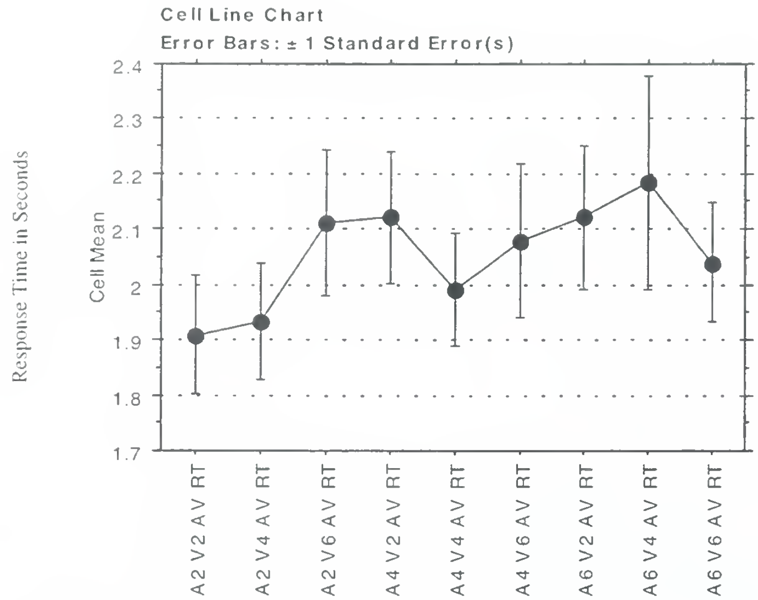


V2A2 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Low-Auditory Quality Display
V2A4 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and Med-Auditory Quality Display
V2A6 AV RT = Time to Rate Low-Quality Visual-Only Percept of Combined Low-Visual and High-Auditory Quality Display
V4A2 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Low-Auditory Quality Display
V4A4 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and Med-Auditory Quality Display
V4A6 AV RT = Time to Rate Med-Quality Visual-Only Percept of Combined Med-Visual and High-Auditory Quality Display
V6A2 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Low-Auditory Quality Display
V6A4 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and Med-Auditory Quality Display
V6A6 AV RT = Time to Rate High-Quality Visual-Only Percept of Combined High-Visual and High-Auditory Quality Display

Figure 104. Combined Data: Visual-Only Quality Rating Response Times of a Combined Auditory-Visual Display

display, when asked to rate both the auditory and visual displays. Again, in looking at the overall results of the response times, one can see various trends, however, several factors limit the ability to correctly analyze these temporal results in any statistically valid manner. These factors are discussed in the OBSERVATIONS section below.

In terms of the post-experiment questions, Figure 107 represents the overall subject's opinion on 1) how easy or difficult it was to determine the quality of the various displays, and 2) if less or more time was needed to adequately rate the various displays. Keeping in mind that subjects used a Likert rating scale ranging from 1 to 7 (4 being neutral) to rate their opinions, the overall results indicate that determining the quality of

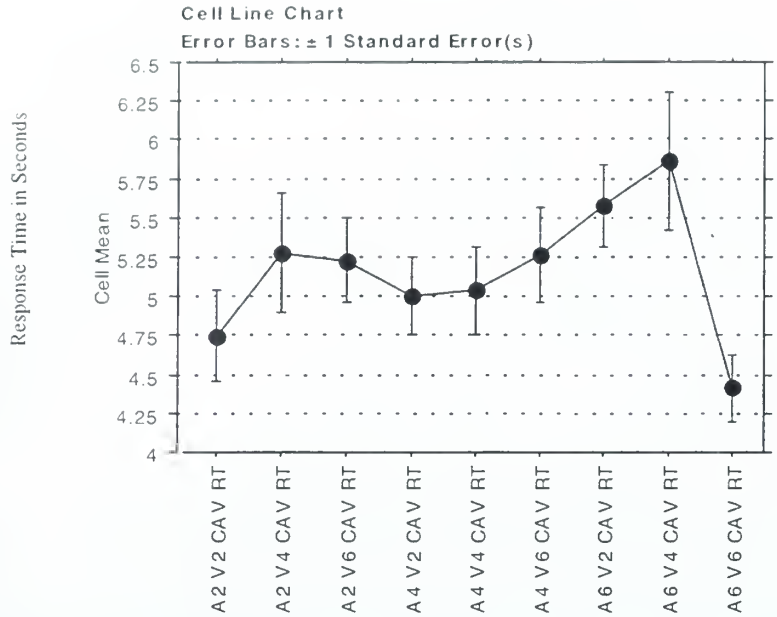


A2V2 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Low-Visual Quality Display
A2V4 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and Med-Visual Quality Display
A2V6 AV RT = Time to Rate Low-Quality Auditory-Only Percept of Combined Low-Auditory and High-Visual Quality Display
A4V2 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Low-Visual Quality Display
A4V4 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and Med-Visual Quality Display
A4V6 AV RT = Time to Rate Med-Quality Auditory-Only Percept of Combined Med-Auditory and High-Visual Quality Display
A6V2 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Low-Visual Quality Display
A6V4 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and Med-Visual Quality Display
A6V6 AV RT = Time to Rate High-Quality Auditory-Only Percept of Combined High-Auditory and High-Visual Quality Display

Figure 105. Combined Data: Auditory-Only Quality Rating Response Times of a Combined Auditory-Visual Display.

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 was an adequate amount of time overall to rate the visual-only and auditory displays, but that slightly more than eight seconds was desired when rating the combined auditory-visual displays.

Finally, the remaining questions of the post-experiment survey reveal that 60 out of 72 subjects (83.3%), focused on alphanumerics when determining the quality of the visual displays (only applicable in the first two experiments) and that 36 of the 108



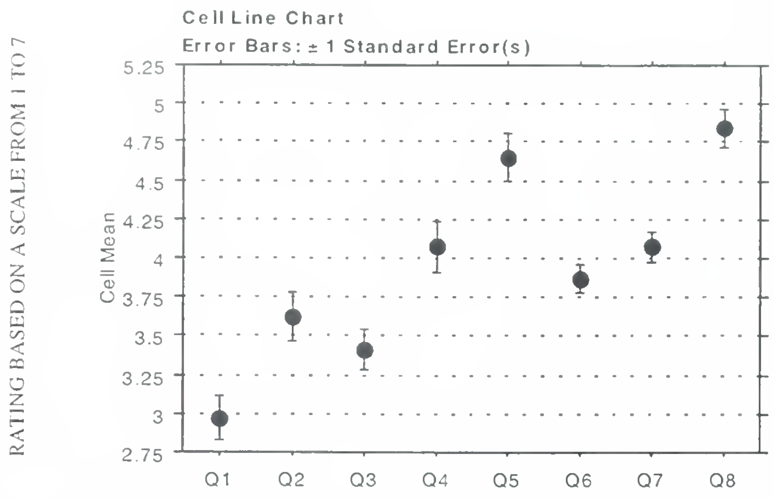
A2V2 CAV RT = Time to Rate Both Low-Auditory and Low-Visual Quality Displays of a Combined Display
A2V4 CAV RT = Time to Rate Both Low-Auditory and Med-Visual Quality Displays of a Combined Display
A2V6 CAV RT = Time to Rate Both Low-Auditory and High-Visual Quality Displays of a Combined Display
A4V2 CAV RT = Time to Rate Both Med-Auditory and Low-Visual Quality Displays of a Combined Display
A4V4 CAV RT = Time to Rate Both Med-Auditory and Med-Visual Quality Displays of a Combined Display
A4V6 CAV RT = Time to Rate Both Med-Auditory and High-Visual Quality Displays of a Combined Display
A6V2 CAV RT = Time to Rate Both High-Auditory and Low-Visual Quality Displays of a Combined Display
A6V4 CAV RT = Time to Rate Both High-Auditory and Med-Visual Quality Displays of a Combined Display
A6V6 CAV RT = Time to Rate Both High-Auditory and High-Visual Quality Displays of a Combined Display

Figure 106. Combined Data: Response Times of Both Auditory and Visual Displays of a Combined Auditory-Visual Display.

subjects (33.3%) felt that they were mentally overloaded when having to rate both auditory and visual displays simultaneously.

C. OVERALL CONCLUSIONS

The goal of this research has been achieved. By varying the quality (fidelity) of both auditory and visual displays, it has been possible to measure auditory-visual cross-modal perception phenomena. The overall conclusions suggest that 1) whether asked to specifically attend to both auditory and visual modalities or asked to attend to only one



Q1 = How easy or difficult was is to determine the quality of the visual-only displays?
Q2 = How easy or difficult was is to determine the quality of the auditory-only displays?
Q3 = How easy or difficult was is to determine the visual quality of the auditory-visual displays?
Q4 = How easy or difficult was is to determine the auditory quality of the auditory-visual displays?
Q5 = How easy or difficult was to determine both the auditory and visual qualities of the auditory-visual displays?
Q6 = Would you have liked less or more time to view the visual-only displays?
Q7 = Would you have liked less or more time to hear the auditory-only displays?
Q8 = Would you have liked less or more time to hear-view the combined auditory-visual displays?

Figure 107. Combined Data: Post-Experiment Questions 1 - 8.

modality, 2) whether manipulating visual display pixel resolution or Gaussian noise level, 3) whether manipulating auditory display sampling frequency or Gaussian noise level, or 4) whether an auditory-visual display is tightly or loosely coupled, cross-modal auditory-visual perception phenomena exist. Overall, these findings strongly suggest:

1) When attending only to the visual modality, a high-quality visual display coupled with either a medium- or 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.

2) When attending only to the auditory modality, a low-quality auditory display coupled with either a medium- or 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.

3) When attending to both auditory and visual modalities, a high-quality visual display coupled with a low-, medium-, or 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.

Another finding worth mentioning, which is just slightly above the level of statistical significance set for this research, is that when 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 quality relative to established baseline conditions derived from auditory-only quality perception evaluations.

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 audio can influence the quality perception of video, and that video can influence the quality perception of audio. The results also indicate that although we can divide our attention between audition and vision, we are not consciously aware of potentially significant intermodality effects.

D. IMPACT

Because of the multi-disciplinary nature of this research effort, the impact of the overall findings are far reaching having both theoretical and commercial implications.

1. Theoretical Impact

The theoretical impact of the findings in this study are diverse. The following describes the impact on Sensory Interaction, Visual Dominance, Divided Attention, and Time-sharing.

a. Sensory Interaction

Because the overall findings indicate that auditory quality can influence visual quality perception and vice versa, some sort of sensory interaction must be taking place. These findings support the many conclusions outlined earlier in Chapter II, Section C. For example, these findings support the early intersensory research conclusions of both Ryan [RYAN40] and Gilbert [GILB41]. Also, O'Connor and Hermelin [OCON81] would argue that these findings support the concept of *sensory capture*. But how this sensory interaction occurs is still not known. Stein and Meredith [STEI93] might conclude that this interaction could be taking place at the neurological level based on single multi-modal neurons as depicted earlier in Figure 4 and Figure 5. However, Gibson [GIBS66] [GIBS79] might argue that this sensory interaction is based on the complexity of natural life events.

b. Visual Dominance

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 either a medium- or 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 discussed earlier in Chapter II, Section E and Chapter III, Section F. 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.

c. Divided Attention

The overall findings of this research indicate that humans can effectively divide their attention between the auditory and visual sensory modalities. This ability to divide one's attention between the auditory and visual sensory modalities supports the various attention theories discussed earlier in Chapter II, Section F.

d. Time-Sharing

Although this research supports the ability to divide attention among the auditory and visual sensory modalities, the time-sharing question remains: do we process these simultaneous auditory and visual stimuli in parallel or in serial? If the overall results indicate that we process simultaneous auditory and visual stimuli in serial, this would lend support the Single-Resource Theory discussed earlier in Chapter II, Section F. If the overall results indicate that we process simultaneous auditory and visual stimuli in parallel, this would lend support the Multiple-Resource Theory discussed earlier in Chapter II, Section F. Since 33.3% of all subjects felt that they were mentally overloaded when having to rate both auditory and visual displays simultaneously, one might conclude that these particular subjects did not have adequate time to simultaneously rate both auditory and visual displays in a serial manner and therefore had to process the simultaneous auditory and visual displays in parallel, which was mentally overloading. If this were true, this would lend support to the Multiple-Resource Theory. However, it is important to note that in this research effort, no assumptions can be made as to how the subjects processed the simultaneous auditory and visual stimuli. Consequently, no time-sharing conclusions can be made from the overall results of this research effort.

2. Commercial Impact

The commercial impact of the findings in this study are diverse. For example, one of the overall findings of this research effort suggests that when attending only to the visual modality, a high-quality visual display coupled with either a medium- or high-quality auditory display causes an increase in the overall visual quality perception of an auditory-visual display. Thus, suppose the fictitious company, *ACME Cyber Art*, sells contemporary paintings via the internet. *ACME Cyber Art's* current web-based advertising only depicts photographs of the various paintings from which prospective customers can purchase on-line. *ACME Cyber Art*, however, wants to increase its sales. One possible strategy to increase sales, is to simply add medium- or high-quality music to their web page while prospective customers are looking at the various artworks. As such, the perceptual visual quality of the various artworks might increase relative to itself, thereby possibly increasing the probability that the customer will make a purchase.

Another finding of this research effort suggests that when attending only to the auditory modality, a low-quality auditory display coupled with either a medium- or high-quality visual display causes a decrease in the overall auditory quality perception of an auditory-visual display. Thus, suppose the next GRAMMY Awards were partially decided via internet-based votes. As such, music fans would point their web browser to the GRAMMY Awards web site to cast their votes. This GRAMMY web site would contain high-quality visual images of the various nominated musical talents. By clicking on the visual image of a particular musical talent, one could hear a short 15 second audio clip of the nominated song. In an effort to 1) decrease rendering time, 2) decrease storage requirements, and 3) decrease download time, suppose the GRAMMY web site designers decreased the sampling frequency of the audio clips from 44.1 kHz to 10 kHz. As a result, to the surprise of the GRAMMY web 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 low-, medium- or 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 3D scene depicting a typical family living room. The current virtual living room contains a TV and stereo system which 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, causing an unwanted increase in the visual rendering time of the scene, the VE developer adds 1) high-quality music to the stereo system, and 2) 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 impacted by this research effort. Overall, the findings of this research effort are indeed important which can greatly benefit the gaming business, multimedia industry, entertainment industry, VE community, and also the Internet industry.

E. OBSERVATIONS

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

1. Response Time Measurement

After observing 130 subjects throughout the course of the various experiments, the use of the rating scales to collect subject responses times is perhaps invalid. The reason for this stems from the physical layout of the rating scales and the functionality of the mouse. Since the rating scales consist of one or two horizontal set(s) of radio buttons, the distance between the *Push to Continue* button and choice number *one* is further than the distance between the *Push to Continue* button and choice number *four*. As a result, it will always take a longer time to select, for example, choice numbers *one* and *seven* as opposed to choice number *four*. To alleviate this problem, all response times need to be normalized to establish a common time metric among all choices. This normalization process is achieved through *Fitts's Law* which states that "...the time to move the hand to a target depends only on the relative precision required, that is, the ratio between the target's distance and its size" [CARD83] (see [WICK92] for more information on Fitts's Law). Nevertheless, Fitts's Law was not considered in this research effort.

In terms of the combined rating scale, some subjects complained that the visual scale should have been on the top whereas others preferred the current format with the auditory scale on top. The functionally of the mouse and mouse pad also have an undetermined effect on response time. Some subjects complained that the mouse would occasionally stick or slide improperly, while others did not experience any problems. Some subjects would keep their hands on the mouse the entire time, while 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 position of the rating scale even before the scale appeared. Furthermore, adept computer users are much more efficient at using the mouse as opposed to some one using the mouse's *point-and-click* paradigm for the first time. Some subjects who were accustomed trackball users felt uncomfortable using the mouse.With all the preceding observations, the use of the rating scales in all three experiments to capture response time ought to be considered invalid. Therefore, as stated earlier, any statistical analysis of the results of the response times must keep in mind the aforementioned observations.

2. Synesthesia Encounter

After discussing the experiment with one of the female subjects, she said that sometimes she experienced various shades of colors when listening to classical music. She was not aware of all the research that has been done concerning synesthesia. It was very interesting to discuss synesthesia with someone who actually experiences synesthesia.

3. Subjects Description and Use of the Stimuli

Perhaps the most interesting observations were gathered from the post-experiment 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 simply amazing. This 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. The following is an excerpted compilation of the items focused on by the subjects and also the terms used to describe what they focused on when determining visual and auditory displays quality.

a. Experiment 1: Static Resolution

Visual Display Quality Terms:

fonts, lines at edge, patterns, straight lines, text, control knobs, frame around frequency window, matrix on speaker pattern, numbers on frequency scale, name on radio, top left edge of radio, the" on" and "off" labels, the word "hallicrafters" on the radio, outside edges of radio, lower speaker line, the lines going through the image, dial, anti-aliasing, legibility of characters, the word "turning," the number "12," the upper right-hand portion of the radio, the white dots on speaker pattern, contrast of radio to background, pieces of dirt on top of radio, highlights, grill, letters, blurring of letters and numbers, ridges on dial, inconsistencies of corners and the line along the backside of the radio, the word "continental" on the radio, reflecting light, white knob.

Auditory Display Quality Terms:

sense of remoteness, cymbals, the cymbals crash, compressed versus open, frequencies, low sounded muddy and didn't sustain, treble, guitar, highs versus lows, opening highs, high was more clear, high hat on drums, frequency range, dynamic range, the presence of the closer sound appeared to be of better quality, low was muffled and high was more treble, the counter point of low frequency organ line, the keyboard resonance was more dynamic in the highs than in the lows, high sounded tinny and low quality had more base, base/treble, more base in high and less base in low, high was painful and low was not painful, qualities seemed reversed, low sounded farther back and high sounded farther forward, the first note, drum sound, low quality was more pleasing, high was more irritating, low was more damped than high, the low quality sounded muted, snare drums, low sounded better, clearness of music, low had less volume, high was more broad sounding, bass was high, the poor music reminded me of music in a can, the good music was a definite stereo sound.

Combined Auditory-Visual Display Quality Terms:

It was hard to believe that the older radio could play the newer alternative music, reversal of auditory and visual qualities.

b. Experiment 2: Static Noise

Visual Display Quality Terms:

small print above lower right and left dials, words under frequency scale, numbers on frequency scale, granularity quality of background, the "on" and *"off" switch, name of radio, judge readability of alphanumerics, granularity of edges, brightness of white knob, better resolution means better quality, right side of radio, letters above the knobs, the word "continental," mesh in speaker, reflection on front top, darkness of black, clarity of dial numbers, the amount of brownish distortion in black finish of radio, contrasts between light and dark, glare in front right top quadrant of radio, shine on top, shadows, light reflection, lower right-hand-corner, background static, sharpness of "on" "off" knob, grille holes, outlay of radio, looked at dots all over, fuzziness of the grid lines on the speaker, corners, graininess of picture, textures, haze on top and haze on reflection, bottom left of whole image.*

Auditory Display Quality Terms:

piano accompaniment in the background, general level of static, clarity of bass, clearer is higher quality, the louder static was low quality and the lower static was the higher quality, differentiate the amount of static present, loudness of static versus loudness of audio signal, hiss level, bass tones, the crispness of the music, the frequency pitch of the static background noise, amount of snow/ interference, white noise level, amount of feedback, scratchiness, the frequency of static, level of noise, percent of volume taken up by noise, the loudness of the background rain, treble.

Combined Auditory-Visual Display Quality Terms:

sometimes reversed auditory and visual qualities.

c. Experiment 3: Static Resolution Nonalphanumeric

Visual Display Quality Terms:

pixellation on lower leaf, outline of apple and fruit on the plate, upper edge of apple, right side of leaf on table, bottom edge of red rose, flowers, carpet, texture, shadowing, fruit skin, the roses, peach, pear, looking for continuous lines, clarity of black spot on pear, weave of cloth, rose petals, smoothness of apples, the overall colors, the brighter the better the quality, blade of grass in lower left corner, curved edges and color blends, the contrast with the yellow and red roses, looked at cleaner images, pink rose petals, hard edges, the pixels.

Auditory Display Quality Terms:

high-end tenor quality, high frequencies, low quality sounded as though it was played in a box, mushing sound for low quality, more pinging for high quality, tone increased with high quality sound, low quality has a deeper tone, high was tinny, the low was hollow sounding, the high was sharper, the chimes sounded muted and the high was full and loud, high quality had higher notes, bass was muffled and high had crisp cymbals, more bass means better quality, range of tones, muffling of resonance, equality of left and right ears, hissing or lack thereof in the background, low end fidelity and range of sound, things I could not express, tonal quality, clearness of bass, the higher pitched instrument coming through clearer, one is clear, the other is distant, the guitar in the back, loudness of the shower, brush strokes for the cymbals, the peaks, the more the instruments the more the quality.

Combined Auditory-Visual Display Quality Terms:

The bowl of fruit does not mix well with the choice of music. The choice of music should have been classical, reversal of audio and visual qualities, drumbeat and treble, the more the bass the better the quality,

4. Reversals

A very common response from the subjects was that they sometimes felt they may have reversed the rating of auditory and visual qualities. This auditory-visual dyslexia may be attributed to some of the findings concerning auditory-visual cross-modal perception.

5. 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 were utilized. Their astonishment is probably attributed to the number of choices on the rating scales (seven). Thus, subjects may have been anticipating seven levels of quality, and as a result conformed (perceptually) to accepting seven quality levels.

F. RECOMMENDATIONS

1. Recruiting Subjects

The recruiting of volunteer subjects took much longer time to accomplish than originally planned. One should anticipate allocating more time to recruit subjects than the total amount of time to actually test subjects.

2. Statistical Analysis Package

Because the statistical analysis software package was chosen well in advance of collecting data, as well as mastering its use, the data analysis portion was accomplished with much greater ease.

3. Hardware and Software Platform

Because of the immense amount of time and data lost due to hardware and software related issues during the experimental design phase of this research effort, it is crucial to insure the reliability and usability of all chosen hardware and software as early as possible in the design phase.

4. Downloaded Software

The use of all the freely downloaded software used in this effort greatly facilitated the software development of the main experiments, since the experimenter merely has to download the software and start developing. There is no need to waste time venturing out to the computer software store. Furthermore, since the software is free, precious research funding can be used for other things such as hardware.

5. *Photoshop* and *SoundForge*

This research would not have been possible without the software to create the various visual and auditory displays. Adobe *Photoshop* [ADOB98] and Sonic Foundary's *SoundForge* [SONI98] proved to be outstanding software packages and their use is highly recommended.

6. Visual Dominance

It is interesting to note, that because this dissertation is a written document, only the visual stimuli can be presented to the reader which is evident by the numerous figures. The auditory stimuli can only be imagined. Thus, the reader has a much better understanding of the visual stimuli, but not the auditory stimuli. Is this not another example of visual dominance?

G. FUTURE WORK

1. Choice of Quality Parameters and Stimuli

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

One possible scenario using a VE might first include the process of having subjects watch a virtual person (in 3D 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 (radio). As a result, if the same experiments outlined in this dissertation were then conducted after this initial process, the overall results might indicate an increase in statistically significant auditory-visual cross-modal perception phenomena.

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 [HOLL97] as depicted earlier in Figure 29). 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, etc. 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.

3. Intersensory Research

The exhaustive literature review and results of this research effort make it clear that in order 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 investigating auditory-visual-haptic interactions.

4. On-line Experiments

Because of the potential to easily acquire many (perhaps thousands) subjects, the use of on-line experiments can greatly facilitate scientific research. As such, all the experiments contained in this research effort can be used on-line. However, on-line experiments make it difficult to control the conditions of the experiment (i.e., hardware specifications, proper subject participation, environmental conditions, etc.). Being able to control the conditions is vital when conducting experiments. Nevertheless, a first attempt has been made towards conducting on-line experiments which can hopefully be used toward future on-line research.

H. FINAL THOUGHTS

It is hoped that this dissertation will help to bridge the current multi-disciplinary gap among multimedia and VE developers. Furthermore, this dissertation is intended to become the key reference that researchers need to read before attempting to evaluate multi-modal perceptual effects in combined auditory and visual displays.

LIST OF REFERENCES

[ADOB98] Adobe, *Photoshop*, Image Manipulation Software Application, WWW URL, as of July 15, 1998, *http://www.adobe.com/prodindex/photoshop/mail.html*

[ALDR95] Aldridge, R., Davidoff, J., Ghanbari, M., Hands, D., and Pearson, D., "Measurement of scene-dependent quality variations in digitally coded television pictures," *IEE Proc.-Vis. Image Signal Process.*, Vol. 142, No. 3, June 1995, pp. 149-154.

[ANDE97] Anderson, David B., and Casey, Michael A., "The sound dimension," *IEEE Spectrum*, March 1997, pp. 46-51.

[BARF95] Barfield, Woodrow, Hendrix, Claudia, Bjorneseth, Ove, Kaczmarek, Kurt A., and Lotens, Wouter, "Comparison of Human Sensory Capabilities with Technical Specifications of Virtual Environment Equipment," *Presence*, Vol. 4, No. 4, Fall 1995, pp. 329-356.

[BARO96] Baron-Cohen, Simon, and Harrison, John E., (Eds.), *Synaesthesia: Classic and Contemporary Readings*, Blackwell Publishers, 1996.

[BECH90] Bech, Søren, "Listening Tests on Loudspeakers: A Discussion of Experimental Procedures and Evaluation of the Response Data," *Proceedings of the Audio Engineering Society 8th International Conference*, Washington, D.C., 1990.

[BECH95] Bech, Søren, Hansen, Villey, and Woszczyk, Wieslaw, "Interaction Between Audio-Visual Factors in a Home Theater System: Experimental Results," Preprint No. 4096, presented at *The 99th Audio Engineering Society Convention*, New York, New York, October 6-9, 1995.

[BECH97] Bech, S, "The Influence of Stereophonic Width on the Perceived Quality of an Audio-Visual Presentation Using a Multichannel Sound System," Preprint No. 4432, presented at *The 102nd Audio Engineering Society Convention*, March 22-25, 1997.

[BEHA74] Behar, Isaac, and Bevan, William, "The Perceived Duration of Auditory and Visual Intervals: Cross-Modal Comparison and Interaction," *American Journal of Psychology*, Vol. 74, 1961, pp. 17-26.

[BEGA94] Begault, Durand R., *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Inc., Cambridge, Massachusetts, 1994.

[BERM76] Bermant, Robert I., and Welch, Robert B., "Effect of Degree of Separation of Visual-Auditory and Eye Position upon Spatial Interaction of Vision and Audition," *Perceptual and Motor Skills*, Vol. 43, 1976, pp. 487-493.

[BLAT96] Blattner, Meera M., and Glinert, Ephraim P., "Multimodal Integration," *IEEE Multimedia*, Winter 1996, pp. 14-24.

[BLAU97] Blauert, Jens, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Revised Edition, The MIT Press, Cambridge, Massachusetts, ISBN: 0-262-02413-6, 1997.

[BOFF86] Boff, Kenneth R., Kaufman, Lloyd, and Thomas, James P., (Eds.) "Divided Attention," *Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance*, John Wiley and Sons, New York, 1986, pp. 26-16 through 26-23.

[BOYK97] Boyk, James, *There's Life Above 20 Kilohertz! A Survey of Musical Instrument Spectra to 102.4 KHz*, Music Lab, California Institute of Technology, Pasadena, California, 1997. (http://www.cco.caltech.edu/~boyk/spectra/spectra.htm)

[BREG90] Bregman, Albert S., *Auditory Scene Analysis*, MIT Press, Cambridge, Massachusetts, 1990.

[BROA58] Broadbent, D. E., *Perception and Communication*, Pergamon, Oxford, 1958.

[BURK92] Burkhard, Mahlon, and Genuit, Klaus, "Merging Subjective and Objective Acoustical Measurements," *Proceedings of the Audio Engineering Society 11th International Conference*, Portland, Oregon, 1992.

[BURR75] Burrows, David, and Solomon, Barry A., "Parallel scanning of auditory and visual information," *Memory & Cognition*, Vol. 3, No. 4, 1975, pp. 416-420.

[BUTT81] Butterworth, George, "The Origins of Auditory-Visual Perception and Proprioception in Human Development," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.37-70.

[CARD83] Card, Stuart K., Moran, Thomas P., and Newell, Allen, *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 1983.

[CHRI94] Christel, Michael G., "The Role of Visual Fidelity in Computer-Based Instruction," *Human-Computer Interaction*, Vol. 9, 1994, pp. 183-223.

[COLI74] Colavita, Francis B., "Human sensory dominance," *Perception & Psychophysics*, Vol. 16, 1974, pp. 409-412.

[COSM98] Cosmo Player, VRML Rendering Plugin Software, WWW URL, as of July 15, 1998, *http://www.cosmosoftware.com/download/*

[CREA98] Creative Labs Inc., *Sound Blaster*, Computer Audio Hardware, WWW URL, as of July 15, 1998, *http://www.soundblaster.com*

[CYTO89] Cytowic, Richard, *Synesthesia: A Union of the Senses*, Springer-Verlag, New York, 1989.

[CYTO93] Cytowic, Richard, *The Man Who Tasted Shapes*, G. P. Putman's and Sons, New York, 1993.

[CYTO95] Cytowic, Richard E., "Synesthesia: Phenomenology And Neuropsychology. A Review of Current Knowledge," *PSYCH*, Vol. 2, No. 10, July 1995.

[DACH95] Dachis, Chuck, *Radios by hallicrafters with Price Guide*, Schiffer Publishing, Ltd., Atglen, Pennsylvania, 1995.

[DEMB79] Dember, William N., and Warm, Joel S., *Psychology of Perception*, 2nd Ed., Holt, Rinehart, and Winston, New York, 1979.

[DEUT63] Deutsch, J. A., and Deutsch, D., "Attention: Some theoretical considerations," *Psychological Review*, Vol. 70, 1963, pp. 19-26.

[DIAM98] Diamond Multimedia, Computer Multimedia Hardware, WWW URL, as of July 15, 1998, *http://www.diamondmm.com*

[DURL95] Durlach, Nathaniel I., and Mavor, Anne S., (Eds.) *Virtual Reality: Scientific and Technological Challenges*, National Research Council, National Academy Press, Washington, D.C., 1995.

[EGET77] Egeth, Howard E., and Sager, Lawrence C., "On the locus of visual dominance," *Perception & Psychophysics*, Vol. 22, No. 1, 1977, pp. 77-86.

[ELLI96] Ellis, Stephen R., "Presence of Mind: A Reaction to Thomas Sheridan's 'Further Musings on the Psychophysics of Presence'," *Presence*, Vol. 5, No. 2, Spring 1996, pp. 247-259.

[ELSA98] ELSA, Computer Multimedia Hardware, WWW URL, as of July 15, 1998, *http://www.elsa.com/*

[FLAC98] Flach, John M., and Holden, John G., "The Reality of Experience: Gibson's Way," *Presence*, Vol. 7, No. 1, February 1998, pp. 90-95.

[FLAN96] Flanagan, David, *Java in a Nutshell: A Desktop Quick Reference for Java Programmers*, O'Reilly & Associates, Inc., Sebastopol, California, 1996.

[FOST68] Foster, Dean and Danker, W. H., "The Nature of Stimuli," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, 1968, pp. 7-10.

[FRIE96] Friedman, Morton P., and Carterette, Edward C., *Cognitive Ecology*, Academic Press, January 1996.

[FURM90] Furmann, Anna, Hojan, Edward, Hiewiarowicz, and Perz, Piotr, "On the Correlation between the Subjective Evaluation of Sound and the Objective Evaluation of Acoustic Parameters for a Selected Source," *Journal of the Audio Engineering Society*, Vol. 38, No. 11, November 1990, pp. 837-844.

[GABR85] Gabrielsson, Alf and Lindström, Björn, "Perceived Sound Quality of High-Fidelity Loudspeakers," *Journal of the Audio Engineering Society*, Vol. 33, No. 1/2, January/February, 1985, pp. 33-53.

[GARN70] Garner, W. R., "The Stimulus in Information Processing," *American Psychologist*, Vol. 25, 1970, pp. 350-358.

[GENE86] Geneva: International Telecommunication Union, "Method for the Subjective Assessment of the Quality of Television Pictures. Recommendation 500-3," *Recommendations and Reports of the CCIR*, Section 11D: Picture Quality and the Parameters Affecting It, 1986.

[GIBS66] Gibson, J. J., *The Senses Considered as Perceptual Systems*, Houghton Mifflin, Boston, 1966.

[GIBS67] Gibson, J. J., "On the proper meaning of the term stimulus," *Psychological Review*, Vol. 74, 1967, pp. 533-534.

[GIBS79] Gibson, J. J., *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, 1979.

[GILB41] Gilbert, G. M., "Inter-Sensory Facilitation and Inhibition," *Journal of General Psychology*, Vol. 24, 1941, pp. 381-407.

[GOOD95] Goodwin, C. James, *Research in Psychology: Methods and Design*, John Wiley & Sons Inc., New York, 1995.

[GREG52] Gregg, Lee W., and Brogden, W. J., "The Effect of Simultaneous Visual Stimulation on Absolute Auditory Sensitivity," *Journal of Experimental Psychology*, Vol. 43, 1952, pp. 179-186.

[GUND97] Gunderson, Martin, a non-titled position paper in *Modeling and Simulation: Linking Entertainment & Defense*, Zyda, Michael and Sheehan, Jerry, (Eds.), National Academy Press, Washington, D.C., September 1997.

[GUPT97] Gupta, Rakesh, Sheridan, Thomas, and Whitney, Daniel, "Experiments Using Multimodal Virtual Environments in Design for Assembly Analysis," *Presence*, Vol. 6, No. 3, June 1997, pp. 318-338.

[HAHN98] Hahn, James K, Fouad, Hesham, Gritz, Larry, and Lee, Jong Won, "Integrating Sounds and Motions in Virtual Environments," *Presence*, Vol. 7, No. 1, February 1998, pp. 67-77.

[HANS81] Hanson, Vicki L., "Processing of written and spoken words: Evidence for common coding," *Memory & Cognition*, Vol. 9, No. 1, 1981, pp. 93-100.

[HARV98] Harvard Medical School, Superior Colliculus Image, WWW URL, as of July 30, 1998, *http://www.med.harvard.edu/AANLIB/cases/caseM/mr1_t/023.html*

[HART96] Hartman, Jed and Wernecke, Josie, *The VRML 2.0 Handbook: Building Moving Worlds on the Web*, Addison Wesley, Reading, Massachusetts, 1996.

[HELM66] Helmholtz, Hermon L. F. von, *Handbuch der Physiologischen Optik*, Voss, Hamburg and Leipzig, 1866.

[HEND94] Hendrix, Claudia, *Exploratory studies on the sense of presence in virtual environments as a function of visual and auditory display parameters*, unpublished Master's Thesis, Sensory Engineering Laboratory, Department of Industrial Engineering, University of Washington, 1994.

[HEND96a] Hendrix, Claudia, and Barfield, Woodrow, "Presence within Virtual Environments as a Function of Visual Display Parameters," *Presence*, Vol. 5, No. 3, Summer 1996, pp. 274-289.

[HEND96b] Hendrix, Claudia, and Barfield, Woodrow, "The Sense of Presence within Auditory Virtual Environments," *Presence*, Vol. 5, No. 3, Summer 1996, pp. 290-301.

[HENN54] Henneman, Richard H., and Long, Eugene R., *A Comparison of the Visual and Auditory Senses as Channels for Data Presentation*, Technical Report No. 54-363, USAF, Wright Air Development Center, Dayton, Ohio, 1954.

[HOLL97] Hollier, M. P., and Voelcker, R., "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, 1997.

[HOWA66] Howard, I. P., and Templeton, W. B., *Human Spatial Orientation*, Wiley, New York, 1966.

[HUGO97] Hugonnet, Christian, "A New Concept of Spatial Coherence Between Sound and Picture in Stereophonic (and Surrounding Sound) TV Production," Preprint No. 4539, presented at the *103rd Audio Engineering Society Convention* New York, New York, September 26-29, 1997.

[INTE98] Intervista, *WorldView*, VRML Rendering Plugin Software, WWW URL, as of July 15, 1998, *http://www.intervista.com*

[ISHI94] Ishii, Masahiro, Nakata, Masanori, and Sato, Makoto, "Networked SPIDAR: A Networked Virtual Environment with Visual, Auditory, and Haptic Interactions," *Presence*, Vol. 3, No. 4, Fall 1994, pp. 351-359.

[IWAM92] Iwamiya, S., "Interaction between Auditory and Visual Processing when Listening to Music via Audio-Visual Media," *Second International Conference on Music Perception and Cognition*, UCLA, Society for Music Perception and Cognition, Los Angeles, California, February, 1992.

[JONE75] Jones, Bill, and Kabanoff, Boris, "Eye movements in auditory space perception," *Perception & Psychophysics*, Vol. 17, No. 3, 1975, pp. 241-245.

[JONE81] Jones, Bill, "The Developmental Significance of Cross-Modal Matching," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.109-136.

[KAHN73] Kahneman, Daniel, *Attention and Effort*, Prentice-Hall, Inc., New Jersey, 1973, pp. 136-155.

[KNUD95] Knudsen, E. I., and Brainard, M. S., "Creating a Unified Representation of Visual and Auditory Space in the Brain," *Annual Review of Neuroscience*, Vol. 18, 1995, pp.19-43.

[KOFF35] Koffka, Kurt, *Principles of Gestalt Psychology*, Harcourt, Brace, and World, New York, 1935.

[KOHL40] Kohler, Wolfgan, *Dynamics in Psychology*, Liveright, New York, 1940.

[KRAV36] Kravkov, S. V., "The Influence of Sound upon the Light and Color Sensibility of the Eye," *Acta Opthalmologica Scandinavica*, Vol. 14, 1936, pp. 348-360.

[LADD98] Ladd, Eric, and O'Donnell, Jim, et al., *Using HTML 4.0, Java 1.1, and JavaScript 1.2*, 2nd Edition, Que Corporation, 1998.

[LEAR96] Lea, Rodger, Matsuda, Kouichi, and Miyashita, Ken, *JAVA for 3D and VRML Worlds*, New Riders Publishing, Indianapolis, Indiana, 1996.

[LAUR93] Laurel, Brenda, *Computers as Theatre*, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1993.

[LIPS90] Lipscomb, Scott D., *Perceptual Judgement of the Symbiosis between Musical and Visual Components in Film*, Master's Thesis, University of California, Los Angeles, California, 1990.

[LIPS94] Lipscomb, Scott D., and Kendall, Roger. A., "Perceptual Judgement of the Relationship between Musical and Visual Components in Film," *Psychomusicology*, Vol. 13, Spring/Fall, 1994 pp.60-98.

[LOND54] London, Ivan D., "Research on Sensory Interaction in the Soviet Union," *Psychological Bulletin*, Vol. 51, No. 6, 1954, pp. 531-568.

[LOVE70] Loveless, N. E., Brebner, J., and Hamilton, P., "Bisensory Presentation of Information," *Psychological Bulletin*, Vol. 73, No. 3, March 1970, pp. 161-199.

[MARK74] Marks, Lawrence E., "On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness," *American Journal of Psychology*, Vol. 87, No. 1-2, 1974, pp. 173-188.

[MARK78] Marks, Lawrence E., *The Unity of the Senses: Interrelations among the Modalities*, Academic Press, New York, 1978.

[MARK82] Marks, Lawrence E., "Bright Sneezes and Dark Coughs, Loud Sunlight and Soft Moonlight," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 8, No. 2, 1982, pp. 177-193.

[MARK86] Marks, Lawrence E., Szczesiul, Rosemary, and Ohlott, Patricia, "On the Cross-Modal Perception of Intensity," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 12, No. 4, 1986, pp. 517-534.

[MARK87] Marks, Lawrence E., "On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 13, No. 3, 1987, pp. 384-394.

[MARK89] Marks, Lawrence E., "On Cross-Modal Similarity: The Perceptual Structure of Pitch, Loudness, and Brightness," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 15, No. 3, 1989, pp. 586-602.

[MASS77] Massaro, Dominic W., and Warner, David S., "Dividing attention between auditory and visual perception," *Perception & Psychophysics*, Vol. 21, No. 6, 1977, pp. 569-574.

[MCGU76] McGurk, Harry, and MacDonald, John, "Hearing Lips and Seeing Voices," *Nature*, Vol. 264, December 23/30 1976, pp. 746-748.

[MCNA68] McNamara, B. P., "Vision," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1968, pp. 19-23.

[MILL81] Millar, Susanna, "Crossmodal and Intersensory Perception and the Blind," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.281-314.

[MPEG98] Motion Picture Expert Group (MPEG), MPEG Format Specifications, WWW URL, as of July 22, 1998, *http://www.mpeg.org*

[MURC73] Murch, Gerald M., *Visual and Auditory Perception*, Bobbs-Merrill Company, Inc., Indianapolis, 1973.

[NEGR95] Negroponte, Nicholas, *being digital*, Alfred A. Knopf, New York, 1995.

[NETS98] Netscape, Web Browser Software, WWW URL, as of July 15, 1998, *http://home.netscape.com/computing/download/*

[NEUM90] Neuman, W. Russell, *Beyond HDTV: Exploring Subjective Responses to very High Definition Television*, MIT Media Library, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1990.

[NEUM91] Neuman, W., Crigler, A., and Bove, V. M., "Television Sound and Viewer Perceptions," *Proceedings of the Audio Engineering Society 9th International Conference*, Vol. 1/2, February, 1991, pp. 101-104.

[OCON81] O'Connor, N., and Hermelin, B., "Coding Strategies of Normal and Handicapped Children," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.315-343.

[OOHA91] Oohashi, Tsutomi, Nishina, Emi, Kawai, Norie, Fuwamoto, Yoshitaka, and Imai, Hiroshi, "High-Frequency Sound Above the Audible Range Affects Brain Electric Activity and Sound Perception," Preprint No. 3207, presented at the *91st Audio Engineering Society Convention* New York, New York, 1997.

[PADM92] Padmos, Pieter, and Milders, Maarten V., "Quality Criteria for Simulator Images: A Literature Review," *Human Factors*, Vol. 34, No. 6, 1992, pp. 727-748.

[PETE59] Peterson, L. R., and Peterson, M.J., "Short-term memory retention of individual items," *Journal of Experimental Psychology*, Vol. 58, 1959, pp. 193-198.

[PICK69] Pick, H. L. Jr., Warren, D. H., and Hay, J. C., "Sensory Conflict in Judgements of Spatial Direction," *Perception and Psychophysics*, Vol. 6, 1969, pp. 203-205.

[POSN76] Posner, Michael I., Nissen, Mary Jo, and Klein, Raymond M., "Visual Dominance: An Information-Processing Account of Its Origins and Significance," *Psychological Review*, Vol. 83, No. 2, 1976, pp. 157-171.

[PRAT36] Pratt, Carroll C., "Interaction Across Modalities: I. Successive Stimulation," *The Journal of Psychology*, Vol. 2, 1936, pp. 287-294.

[PRES97] Pressing, Jeff, "Some Perspectives on Performed Sound and Music in Virtual Environments," *Presence*, Vol. 6, No. 4, August 1997, pp. 482-503.

[RADE76] Radeau, Monique, and Bertelson, Paul, "The effect of a textured visual field on modality dominance in a ventriloquism situation," *Perception & Psychophysics*, Vol. 20, No. 4, 1976, pp. 227-235.

[RAGO88] Ragot, Richard, Cave, Christian, and Fano, Michel, "Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation," *Bulletin of the Psychonomic Society*, Vol. 26, No. 4, 1988, pp. 350-352.

[ROEH97] Roehl, Bernie, Couch, Justin, Reed-Ballreich, Cindy, Rohaly, Tim, and Brown, Geoff, *Late Night VRML 2.0 with Java*, Ziff-Davis Press, Emeryville, California, 1997.

[ROSE91] Rosenblum, Lawrence D., and Fowler, Carol A., "Audiovisual Investigation of the Loudness-Effort Effect for Speech and Nonspeech Events," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 17, No. 4, 1991, pp. 976-985.

[RYAN40] Ryan, T. A., "Interactions of the Sensory Systems in Perception," *Psychological Bulletin*, Vol. 37, 1940, pp. 659-698.

[RYDS94] Rydstrom, Gary, "Film Sound: How It's Done in the Real World," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

[SASI98] SAS Institute, *StatView*, Statistical Analysis Software, WWW URL, as of July 15, 1998, *http://www.statview.com/*

[SCHI48] Schillinger, Joseph, *The Mathematical Basis of the Arts*, Philosophical Library, New York, 1948.

[SCHL35] Schiller, Paul, "Interrelation of Different Senses in Perception," *British Journal of Psychology*, Vol. 25, 1935, pp. 465-469.

[SENN98] Sennheiser, Audio Rendering Equipment, WWW URL, as of July 15, 1998, *http://www.sennheiser.com*

[SERR35] Serrat, William D., and Karwoski, Theodore, "An Investigation of the Effect of Auditory Stimulation on Visual Sensitivity," *Journal of Experimental Psychology*, Vol. 19, 1936, pp. 604-611.

[SHERI96] Sheridan, Thomas B., "Further Musings on the Psychophysics of Presence," *Presence*, Vol. 5, No. 2, Spring 1996, pp. 241-246.

[SHERR47] Sherrington, C. S., *The Integrative Action of the Nervous System*, Yale University Press, New Haven, 1947.

[SILB68] Silbiger, H. R., "Hearing," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1968, pp. 24-29.

[SLAT94] Slater, Mel, Usoh, Martin, and Steed, Anthony, "Depth of Presence in Virtual Environments," *Presence*, Vol. 3, No. 2, Spring 1994, pp. 130-144.

[SLAT97] Slater, Mel, and Wilber, Sylvia, "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments," *Presence*, Vol. 6, No. 6, December 1997, pp. 603-616.

[SONI98] Sonic Foundary, *Sound Forge*, Computer Sound File Manipulation Software, WWW URL, as of July 15, 1998, *http://www.soundforge.com/*

[SONY98a] Sony, Computer Monitors, WWW URL, as of July 15, 1998, *http://www.sony.com/*

[SONY98b] Sony, Community Place, VRML Browser Software, WWW URL, as of July 15, 1998, *http://vs.spiw.com/vs/what1.html*

[STEI93] Stein, Barry E., and Meredith, M. Alex, *The Merging of the Senses*, The MIT Press, Cambridge, Massachusetts, 1993.

[STON68] Stone, Herbert and Pangborn, R. M., "Intercorrelation of the Senses," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1968, pp 30-46.

[SUNM98] Sun Microsystems, Inc., Java Programming Language, WWW URL as of July 15, 1998, *http://www.java.sun.com/*

[THEI86] Theile, Günther, "On the Standardization of the Frequency Response of High-Quality Studio Headphones," *Journal of the Audio Engineering Society,* Vol. 34, No. 12, December 1986, pp. 956-969.

[THOM58] Thompson, Richard F., Voss, James F., and Brogden, W. J., "Effect of Brightness of Simultaneous Visual Stimulation on Absolute Auditory Sensitivity," *Journal of Experimental Psychology*, Vol. 55, No. 1, 1958, pp. 45-50.

[THUR92] Thurmond, Bob, "Measurement and Perception Quality in Sound Systems," *Proceedings of the Audio Engineering Society 11th International Conference*, Portland, Oregon, 1992.

[TIER93] Tierney, John, "Jung in Motion, Virtually and Other Computer Fuzz," *The New York Times*, September 16, 1993, pp. C1 and C9.

[TOOL85] Toole, Floyd E., "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *Journal of the Audio Engineering Society,* Vol. 33, No. 1/2, January/February 1985, pp. 20-32.

[TOOL90] Toole, Floyd E., "Identifying and Controlling the Variables," *Proceedings of the Audio Engineering Society 8th International Conference*, Washington, D.C., 1990.

[TREI69] Treisman, Anne M., "Strategies and models of selective attention," *Psychological Review,* Vol. 76, 1969, pp. 282-299.

[TREI73] Treisman, Anne M., and Davies, Alison, "Divided Attention to Ear and Eye," *Attention and Performance IV,* Kornblum, Sylvan (Ed.), Academic Press, New York, 1973, pp, 101-117.

[VIEM90] Viemeister, Neil, "An Overview of Psychoacoustics and Auditory Perception," *Proceedings of the Audio Engineering Society 8th International Conference*, Washington, D.C., 1990.

[WALK81] Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) *Intersensory Perception and Sensory Integration*, Plenum Press, New York, 1981.

[WARR81] Warren, David H., Welch, Robert B., and McCarthy, Timothy J., "The role of visual-auditory 'compellingness' in the ventriloquism effect: Implications for transitivity among the spatial senses," *Perception & Psychophysics*, Vol. 30, No. 6, 1981, pp. 557-564.

[WERT12] Wertheimer, Max, Experimentelle Studien über das Sehen von Bewegungen, *Zeitschrift für Psychologie*, Vol. 61, pp.161-265, 1912.

[WEVE74] Wever, Ernest Glen, "The Evolution of Vertebrate Hearing," *Handbook of Sensory Physiology*, Vol. V/1, Auditory System, (Eds.) Keidel, Wolf D. and Neff, William D., Springer-Verlag, New York, 1974, pp. 423-454.

[WICK92] Wickens, Christopher D., *Engineering Psychology and Human Performance*, 2nd Ed., Harper Collins Publishers Inc., 1992.

[WOSZ95] Woszczyk, Wieslaw, Bech, Søren, and Hansen, Villey, "Interaction Between Audio-Visual Factors in a Home Theater System: Definition of Subjective Attributes," Preprint No. 4133, presented at *The 99th Audio Engineering Society Convention*, New York, New York, October 6-9, 1995.

[ZWIC91] Zwicker, Eberhard and Zwicker, U. Tilmann, "Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System," *Journal of the Audio Engineering Society*, Vol. 39, No. 3, March 1991, pp. 115-126.

[ZYDA97] Zyda, Michael and Sheehan, Jerry, (Eds.), *Modeling and Simulation: Linking Entertainment & Defense*, National Academy Press, Washington D.C., September 1997.

BIBLIOGRAPHY

The following is the complete list of all references, cited and not cited, that were utilized in the research and development of this dissertation.

Adobe, *Photoshop*, Image Manipulation Software Application, WWW URL, as of July 15, 1998, *http://www.adobe.com/prodindex/photoshop/mail.html*

Aldridge, R., Davidoff, J., Ghanbari, M., Hands, D., and Pearson, D., "Measurement of scene-dependent quality variations in digitally coded television pictures," *IEE Proc.-Vis. Image Signal Process.*, Vol. 142, No. 3, June 1995, pp. 149-154.

Anderson, David B., and Casey, Michael A., "The sound dimension," *IEEE Spectrum*, March 1997, pp. 46-51.

Barfield, Woodrow, Hendrix, Claudia, Bjorneseth, Ove, Kaczmarek, Kurt A., and Lotens, Wouter, "Comparison of Human Sensory Capabilities with Technical Specifications of Virtual Environment Equipment," *Presence*, Vol. 4, No. 4, Fall 1995, pp. 329-356.

Baron-Cohen, Simon, and Harrison, John E., (Eds.), *Synaesthesia: Classic and Contemporary Readings*, Blackwell Publishers, 1996.

Bech, Søren, "Listening Tests on Loudspeakers: A Discussion of Experimental Procedures and Evaluation of the Response Data," *Proceedings of the Audio Engineering Society 8th International Conference*, Washington, D.C., 1990.

Bech, Søren, Hansen, Villey, and Woszczyk, Wieslaw, "Interaction Between Audio-Visual Factors in a Home Theater System: Experimental Results," Preprint No. 4096, presented at *The 99th Audio Engineering Society Convention*, New York, New York, October 6-9, 1995.

Bech, S, "The Influence of Stereophonic Width on the Perceived Quality of an Audio-Visual Presentation Using a Multichannel Sound System," Preprint No. 4432, presented at *The 102nd Audio Engineering Society Convention*, March 22-25, 1997.

Behar, Isaac, and Bevan, William, "The Perceived Duration of Auditory and Visual Intervals: Cross-Modal Comparison and Interaction," *American Journal of Psychology*, Vol. 74, 1961, pp. 17-26.

Begault, Durand R., *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Inc., Cambridge, Massachusetts, 1994.

Bermant, Robert I., and Welch, Robert B., "Effect of Degree of Separation of Visual-Auditory and Eye Position upon Spatial Interaction of Vision and Audition," *Perceptual and Motor Skills*, Vol. 43, 1976, pp. 487-493.

Bernstein, Ira H., Rose, Robert, and Ashe, Victor M., "Energy Integration in Intersensory Facilitation," *Journal of Experimental Psychology*, Vol. 86, No. 2, 1970, pp. 196-203.

Bevan, William, and Pritchard, Joan Faye, "The Effect of Visual Intensities upon the Judgements of Loudness," *American Journal of Psychology*, Vol. 77, 1964, pp. 93-98.

Blattner, Meera M., and Glinert, Ephraim P., "Multimodal Integration," *IEEE Multimedia*, Winter 1996, pp. 14-24.

Blauert, Jens, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Revised Edition, The MIT Press, Cambridge, Massachusetts, ISBN: 0-262-02413-6, 1997.

Boff, Kenneth R., Kaufman, Lloyd, and Thomas, James P., (Eds.) "Divided Attention," *Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance*, John Wiley and Sons, New York, 1986, pp. 26-16 through 26-23.

Bregman, Albert S., *Auditory Scene Analysis*, MIT Press, Cambridge, Massachusetts, 1990.

Broadbent, D. E., *Perception and Communication*, Pergamon, Oxford, 1958.

Burkhard, Mahlon, and Genuit, Klaus, "Merging Subjective and Objective Acoustical Measurements," *Proceedings of the Audio Engineering Society 11th International Conference*, Portland, Oregon, 1992.

Burrows, David, and Solomon, Barry A., "Parallel scanning of auditory and visual information," *Memory & Cognition*, Vol. 3, No. 4, 1975, pp. 416-420.

Bush, Karen M., "Stimulus equivalence and cross-modal transfer," *The Psychological Record*," Vol. 43, 1993, pp. 567-584.

Butterworth, George, "The Origins of Auditory-Visual Perception and Proprioception in Human Development," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.37-70.

Card, Stuart K., Moran, Thomas P., and Newell, Allen, *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 1983.

Christel, Michael G., "The Role of Visual Fidelity in Computer-Based Instruction," *Human-Computer Interaction*, Vol. 9, 1994, pp. 183-223.

Churchland, Patricia S., and Sejnowski, Terrence J., *The Computational Brain*, The MIT Press, Cambridge, Massachusetts, 1992.

Colavita, Francis B., "Human sensory dominance," *Perception & Psychophysics*, Vol. 16, 1974, pp. 409-412.

Colavita, Francis B., Tomko, Rosemary, and Weisberg, Daniel, "Visual prepotency and eye orientation," *Bulletin of the Psychonomic Society*, Vol. 8, No. 1, 1976, pp. 25-26.

Cosmo Player, VRML Rendering Plugin Software, WWW URL, as of July 15, 1998, *http://www.cosmosoftware.com/download/*

Creative Labs Inc., *Sound Blaster*, WWW URL, as of July 15, 1998, *http://www.soundblaster.com*

The Cure, "A Forest (Tree Mix)," *mixed up*, a compact disc (CD) distributed by Elektra Entertainment, a division of Warner Communications Inc., 1990.

Cytowic, Richard, *Synesthesia: A Union of the Senses*, Springer-Verlag, New York, 1989.

Cytowic, Richard, *The Man Who Tasted Shapes*, G. P. Putman's and Sons, New York, 1993.

Cytowic, Richard E., "Synesthesia: Phenomenology And Neuropsychology. A Review of Current Knowledge," *PSYCH*, Vol. 2, No. 10, July 1995.

Dachis, Chuck, *Radios by hallicrafters with Price Guide*, Schiffer Publishing, Ltd., Atglen, Pennsylvania, 1995.

Dember, William N., and Warm, Joel S., *Psychology of Perception*, 2nd Ed., Holt, Rinehart, and Winston, New York, 1979.

Deutsch, J. A., and Deutsch, D., "Attention: Some theoretical considerations," *Psychological Review*, Vol. 70, 1963, pp. 19-26.

Diamond Multimedia, Computer Multimedia Hardware, WWW URL, as of July 15, 1998, *http://www.diamondmm.com*

Durlach, Nathaniel I., and Mavor, Anne S., (Eds.) *Virtual Reality: Scientific and Technological Challenges*, National Research Council, National Academy Press, Washington, D.C., 1995.

Efron, Robert, "The Minimum Duration of a Perception, *Neuropsychologia*, Vol. 8, 1970, pp. 57-63.

Egeth, Howard E., and Sager, Lawrence C., "On the locus of visual dominance," *Perception & Psychophysics*, Vol. 22, No. 1, 1977, pp. 77-86.

Ellis, Stephen R., "Presence of Mind: A Reaction to Thomas Sheridan's 'Further Musings on the Psychophysics of Presence'," *Presence*, Vol. 5, No. 2, Spring 1996, pp. 247-259.

ELSA, Computer Multimedia Hardware, WWW URL, as of July 15, 1998, *http://www.elsa.com/*

Flach, John M., and Holden, John G., "The Reality of Experience: Gibson's Way," *Presence*, Vol. 7, No. 1, February 1998, pp. 90-95.

Flanagan, David, *Java in a Nutshell: A Desktop Quick Reference for Java Programmers*, O'Reilly & Associates, Inc., Sebastopol, California, 1996.

Foster, Dean and Danker, W. H., "The Nature of Stimuli," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1968, pp. 7-10.

Friedman, Alinda, Polson, Martha Campbell, Daffoe, Cameron G., and Gaskill, Sarah J., "Dividing Attention Within and Between Hemispheres: Testing a Multiple Resources Approach to Limited-Capacity Information Processing," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 8, No. 5., 1982, pp. 625-650.

Friedman, Morton P., and Carterette, Edward C., *Cognitive Ecology*, Academic Press, January 1996.

Furmann, Anna, Hojan, Edward, Hiewiarowicz, and Perz, Piotr, "On the Correlation between the Subjective Evaluation of Sound and the Objective Evaluation of Acoustic Parameters for a Selected Source," *Journal of the Audio Engineering Society*, Vol. 38, No. 11, November 1990, pp. 837-844.

Gabrielsson, Alf and Lindström, Björn, "Perceived Sound Quality of High-Fidelity Loudspeakers," *Journal of the Audio Engineering Society,* Vol. 33, No. 1/2, January/February, 1985, pp. 33-53.

Garner, W. R., "The Stimulus in Information Processing," *American Psychologist*, Vol. 25, 1970, pp. 350-358.

Geneva: International Telecommunication Union, "Method for the Subjective Assessment of the Quality of Television Pictures. Recommendation 500-3," *Recommendations and Reports of the CCIR*, Section 11D: Picture Quality and the Parameters Affecting It, 1986.

Gibson, J. J., "On the proper meaning of the term stimulus," *Psychological Review*, Vol. 74, 1967, pp. 533-534.

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

Gibson, J. J., *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, 1979.

Gilbert, G. M., "Inter-Sensory Facilitation and Inhibition," *Journal of General Psychology*, Vol. 24, 1941, pp. 381-407.

Goodwin, C. James, *Research in Psychology: Methods and Design*, John Wiley & Sons Inc., New York, 1995.

Gregg, Lee W., and Brogden, W. J., "The Effect of Simultaneous Visual Stimulation on Absolute Auditory Sensitivity," *Journal of Experimental Psychology*, Vol. 43, 1952, pp. 179-186.

Gunderson, Martin, a non-titled position paper in *Modeling and Simulation: Linking Entertainment & Defense*, Zyda, Michael and Sheehan, Jerry, (Eds.), National Academy Press, Washington, D.C., September 1997.

Gupta, Rakesh, Sheridan, Thomas, and Whitney, Daniel, "Experiments Using Multimodal Virtual Environments in Design for Assembly Analysis," *Presence*, Vol. 6, No. 3, June 1997, pp. 318-338.

Hahn, James K, Fouad, Hesham, Gritz, Larry, and Lee, Jong Won, "Integrating Sounds and Motions in Virtual Environments," *Presence*, Vol. 7, No. 1, February 1998, pp. 67-77.

Hanson, Vicki L., "Processing of written and spoken words: Evidence for common coding," *Memory & Cognition*, Vol. 9, No. 1, 1981, pp. 93-100.

Hartman, Jed and Wernecke, Josie, *The VRML 2.0 Handbook: Building Moving Worlds on the Web*, Addison Wesley, Reading, Massachusetts, 1996.

Harvard Medical School, Superior Colliculus Image, WWW URL, as of July 30, 1998, *http://www.med.harvard.edu/AANLIB/cases/caseM/mr1_t/023.html*

Helmholtz, Hermon L. F. von, *Handbuch der Physiologischen Optik*, Voss, Hamburg and Leipzig, 1866.

Hendrix, Claudia, *Exploratory studies on the sense of presence in virtual environments as a function of visual and auditory display parameters*, unpublished Master's Thesis, Sensory Engineering Laboratory, Department of Industrial Engineering, University of Washington, 1994.

Hendrix, Claudia, and Barfield, Woodrow, "Presence within Virtual Environments as a Function of Visual Display Parameters," *Presence*, Vol. 5, No. 3, Summer 1996, pp. 274-289.

Hendrix, Claudia, and Barfield, Woodrow, "The Sense of Presence within Auditory Virtual Environments," *Presence*, Vol. 5, No. 3, Summer 1996, pp. 290-301.

Henneman, Richard H., and Long, Eugene R., *A Comparison of the Visual and Auditory Senses as Channels for Data Presentation*, Technical Report No. 54-363, USAF, Wright Air Development Center, Dayton, Ohio, 1954.

Hollier, M. P., and Voelcker, R., "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, 1997.

Howard, I. P., and Templeton, W. B., *Human Spatial Orientation*, Wiley, New York, 1966.

Hughes, Howard C., Reuter-Lorenz, Patricia A., Nozawa, Goerge, and Fendrich, Robert, "Visual-Auditory Interactions in Sensorimotor Processing: Saccades Versus Manual Responses," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 20, No. 1., 1994, pp. 131-153.

Hugonnet, Christian, "A New Concept of Spatial Coherence Between Sound and Picture in Stereophonic (and Surrounding Sound) TV Production," Preprint No. 4539, presented at the *103rd Audio Engineering Society Convention*, New York, New York, September 26-29, 1997.

Intervista, *WorldView*, VRML Rendering Plugin Software, WWW URL, as of July 15, 1998, *http://www.intervista.com*

Ishii, Masahiro, Nakata, Masanori, and Sato, Makoto, "Networked SPIDAR: A Networked Virtual Environment with Visual, Auditory, and Haptic Interactions," *Presence*, Vol. 3, No. 4, Fall 1994, pp. 351-359.

Iwamiya, S., "Interaction between Auditory and Visual Processing when Listening to Music via Audio-Visual Media," *Second International Conference on Music Perception and Cognition*, UCLA, Society for Music Perception and Cognition, Los Angeles, California, February, 1992.

Jaquish, Gail, "Intra-individual variability in divergent thinking in response to audio, visual, and tactile stimuli," *British Journal of Psychology*, Vol. 74, 1983, pp. 467-472.

Jones, Bill, and Kabanoff, Boris, "Eye movements in auditory space perception," *Perception & Psychophysics*, Vol. 17, No. 3, 1975, pp. 241-245.

Jones, Bill, "The Developmental Significance of Cross-Modal Matching," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.109-136.

Kaeseler, Preben, "Designing Interaction between Auditory and Visual Stimuli: Case stories from the Virtual LEGO-life," Preprint No. 4589, presented at the *103rd Audio Engineering Society Convention* New York, New York, September 26-29, 1997.

Kahneman, Daniel, *Attention and Effort*, Prentice-Hall, Inc., New Jersey, 1973, pp. 136-155.

Knudsen, E. I., and Brainard, M. S., "Creating a Unified Representation of Visual and Auditory Space in the Brain," *Annual Review of Neuroscience*, Vol. 18, 1995, pp.19-43.

Koffka, Kurt, *Principles of Gestalt Psychology*, Harcourt, Brace, and World, New York, 1935.

Kohler, Wolfgan, *Dynamics in Psychology*, Liveright, New York, 1940.

Komiyama, Setsu, "Subjective Evaluation of Angular Displacement between Picture and Sound Directions for HDTV Sound Systems," *Journal of the Audio Engineering Society*, Vol. 37, No. 4, April 1989, pp. 210-214.

Kravkov, S. V., "The Influence of Sound upon the Light and Color Sensibility of the Eye," *Acta Opthalmologica Scandinavica*, Vol. 14, 1936, pp. 348-360.

Ladd, Eric, and O'Donnell, Jim, et al., *Using HTML 4.0, Java 1.1, and JavaScript 1.2*, 2nd Edition, Que Corporation, 1998.

Laurel, Brenda, *Computers as Theatre*, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1993.

Lea, Rodger, Matsuda, Kouichi, and Miyashita, Ken, *JAVA for 3D and VRML Worlds*, New Riders Publishing, Indianapolis, Indiana, 1996.

Lewkowicz, David J., "Perception of Auditory-Visual Temporal Synchrony in Human Infants," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 22, No. 5, 1996, pp. 1094-1106.

Lipscomb, Scott D., *Perceptual Judgement of the Symbiosis between Musical and Visual Components in Film*, Master's Thesis, University of California, Los Angeles, California, 1990.

Lipscomb, Scott D., and Kendall, Roger. A., "Perceptual Judgement of the Relationship between Musical and Visual Components in Film," *Psychomusicology*, Vol. 13, Spring/Fall, 1994 pp.60-98.

London, Ivan D., "Research on Sensory Interaction in the Soviet Union," *Psychological Bulletin*, Vol. 51, No. 6, 1954, pp. 531-568.

Loveless, N. E., Brebner, J., and Hamilton, P., "Bisensory Presentation of Information," *Psychological Bulletin*, Vol. 73, No. 3, March 1970, pp. 161-199.

Marks, Lawrence E., "On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness," *American Journal of Psychology*, Vol. 87, No. 1-2, 1974, pp. 173-188.

Marks, Lawrence E., *The Unity of the Senses: Interrelations among the Modalities*, Academic Press, New York, 1978.

Marks, Lawrence E., "Multimodal Perception," *Handbook of Perception, Perceptual Coding*, Vol. VIII, Carterette, Edward C., and Friedman, Morton P., (Eds.), Academic Press, New York, 1978, pp. 321-339.

Marks, Lawrence E., "Bright Sneezes and Dark Coughs, Loud Sunlight and Soft Moonlight," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 8, No. 2, 1982, pp. 177-193.

Marks, Lawrence E., Szczesiul, Rosemary, and Ohlott, Patricia, "On the Cross-Modal Perception of Intensity," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 12, No. 4, 1986, pp. 517-534.

Marks, Lawrence E., "On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 13, No. 3, 1987, pp. 384-394.

Marks, Lawrence E., "On Cross-Modal Similarity: The Perceptual Structure of Pitch, Loudness, and Brightness," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 15, No. 3, 1989, pp. 586-602.

Massaro, Dominic W., and Warner, David S., "Dividing attention between auditory and visual perception," *Perception & Psychophysics*, Vol. 21, No. 6, 1977, pp. 569-574.

McGurk, Harry, and MacDonald, John, "Hearing Lips and Seeing Voices," *Nature*, Vol. 264, December 23/30 1976, pp. 746-748.

McNamara, B. P., "Vision," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1968, pp. 19-23.

Millar, Susanna, "Crossmodal and Intersensory Perception and the Blind," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.109-136.

Miner, Nadine, Gillespie, Brent, and Caudell, Thomas, "Examining the Influence of Audio and Visual Stimuli on a Haptic Interface," in the *Proceedings of IMAGE 96*, Scottsdale, Arizona, June 23-27, 1996.

Motion Picture Expert Group (MPEG), MPEG Format Specifications, WWW URL, as of July 22, 1998, *http://www.mpeg.org*

Murch, Gerald M., *Visual and Auditory Perception*, Bobbs-Merrill Company, Inc., Indianapolis, 1973.

Negroponte, Nicholas, *being digital*, Alfred A. Knopf, New York, 1995.

Neuman, W. Russell, *Beyond HDTV: Exploring Subjective Responses to very High Definition Television*, MIT Media Library, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1990.

Neuman, W., Crigler, A., and Bove, V. M., "Television Sound and Viewer Perceptions," *Proceedings of the Audio Engineering Society 9th International Conference*, Vol. 1/2, February, 1991, pp. 101-104.

Netscape, Web Browser Software, WWW URL, as of July 15, 1998, *http://home.netscape.com/computing/download/*

O'Conaill, Bird, Whittaker, Steve, and Wilber, Sylvia, "Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication," *Human-Computer Interaction*, Vol. 8, 1993, pp. 389-428.

O'Connor, N., and Hermelin, B., "Coding Strategies of Normal and Handicapped Children," *Intersensory Perception and Sensory Integration*, Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) Plenum Press, New York, 1981, pp.315-343.

O'Leary, Ann, and Rhodes, Gillian, "Cross-modal effects on visual and auditory object perception," *Perception & Psychophysics*, Vol. 35, No. 6, 1984, pp. 565-569.

Otto, Norman C., "Listening Test Methods for Automotive Sound Quality," Preprint No. 4586, presented at the *103rd Audio Engineering Society Convention* New York, New York, September 26-29, 1997.

Padmos, Pieter, and Milders, Maarten V., "Quality Criteria for Simulator Images: A Literature Review," *Human Factors*, Vol. 34, No. 6, 1992, pp. 727-748.

Peterson, L. R., and Peterson, M.J., "Short-term memory retention of individual items," *Journal of Experimental Psychology*, Vol. 58, 1959, pp. 193-198.

Pick, H. L. Jr., Warren, D. H., and Hay, J. C., "Sensory Conflict in Judgements of Spatial Direction," *Perception and Psychophysics*, Vol. 6, 1969, pp. 203-205.

Posner, Michael I., Nissen, Mary Jo, and Klein, Raymond M., "Visual Dominance: An Information-Processing Account of Its Origins and Significance," *Psychological Review*, Vol. 83, No. 2, 1976, pp. 157-171.

Pratt, Carroll C., "Interaction Across Modalities: I. Successive Stimulation," *The Journal of Psychology*, Vol. 2, 1936, pp. 287-294.

Precoda, Kristin, and Meng, Teresa H., "Subjective Audio Testing Methodology and Human Performance Factors," Preprint No. 4585, presented at the *103rd Audio Engineering Society Convention* New York, New York, September 26-29, 1997.

Pressing, Jeff, "Some Perspectives on Performed Sound and Music in Virtual Environments," *Presence*, Vol. 6, No. 4, August 1997, pp. 482-503.

Radeau, Monique, and Bertelson, Paul, "The effect of a textured visual field on modality dominance in a ventriloquism situation," *Perception & Psychophysics*, Vol. 20, No. 4, 1976, pp. 227-235.

Ragot, Richard, Cave, Christian, and Fano, Michel, "Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation," *Bulletin of the Psychonomic Society*, Vol. 26, No. 4, 1988, pp. 350-352.

Regan, D., and Spekreijse, "Auditory-visual interactions and the correspondence between perceived auditory space and perceived visual space," *Perception*, Vol. 6, 1977, pp.133-138.

Roehl, Bernie, Couch, Justin, Reed-Ballreich, Cindy, Rohaly, Tim, and Brown, Geoff, *Late Night VRML 2.0 with Java*, Ziff-Davis Press, Emeryville, California, 1997.

Rosenblum, Lawrence D., and Fowler, Carol A., "Audiovisual Investigation of the Loudness-Effort Effect for Speech and Nonspeech Events," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 17, No. 4, 1991, pp. 976-985.

Ryan, T. A., "Interactions of the Sensory Systems in Perception," *Psychological Bulletin*, Vol. 37, 1940, pp. 659-698.

Rydstrom, Gary, "Film Sound: How It's Done in the Real World," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

SAS Institute, *StatView*, Statistical Analysis Software, WWW URL, as of July 15, 1998, *http://www.statview.com/*

Schiller, Paul, "Interrelation of Different Senses in Perception," *British Journal of Psychology*, Vol. 25, 1935, pp. 465-469.

Schillinger, Joseph, *The Mathematical Basis of the Arts*, Philosophical Library, New York, 1948.

Sennheiser, Audio Rendering Equipment, WWW URL, as of July 15, 1998, *http://www.sennheiser.com*

Serrat, William D., and Karwoski, Theodore, "An Investigation of the Effect of Auditory Stimulation on Visual Sensitivity," *Journal of Experimental Psychology*, Vol. 19, 1936, pp. 604-611.

Shaw, Edgar A. G., "The External Ear," *Handbook of Sensory Physiology*, Vol. V/1, Auditory System, (Eds.) Keidel, Wolf D., and Neff, William D., Springer-Verlag, New York, 1974, pp. 455-490.

Sheridan, Thomas B., "Further Musings on the Psychophysics of Presence," *Presence*, Vol. 5, No. 2, Spring 1996, pp. 241-246.

Sherrington, C. S., *The Integrative Action of the Nervous System*, Yale University Press, New Haven, 1947.

Silbiger, H. R., "Hearing," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1968, pp. 24-29.

Slater, Mel, Usoh, Martin, and Steed, Anthony, "Depth of Presence in Virtual Environments," *Presence*, Vol. 3, No. 2, Spring 1994, pp. 130-144.

Slater, Mel, and Wilber, Sylvia, "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments," *Presence*, Vol. 6, No. 6, December 1997, pp. 603-616.

Sonic Foundary, *Sound Forge*, Computer Sound File Manipulation Software, WWW URL, as of July 15, 1998, *http://www.soundforge.com/*

Sony, Computer Monitors, WWW URL, as of July 15, 1998, *http://www.sony.com/*

Sony, Community Place, VRML Browser Software, WWW URL, as of July 15, 1998, *http://vs.spiw.com/vs/what1.html*

Spence, Charles, and Driver, Jon, "Audiovisual Links in Endogenous Covert Spatial Attention," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 22, No. 4, 1996, pp. 1005-1030.

Sporer, Thomas, *Objective Audio Signal Evaluation -- Applied Psychoacoustics for Modeling the Perceived Quality of Digital Audio*, Preprint No. 4512, presented at the *103rd Audio Engineering Society Convention* New York, New York, September 26-29, 1997.

Stein, Barry E., and Meredith, M. Alex, *The Merging of the Senses*, The MIT Press, Cambridge, Massachusetts, 1993.

Stone, Herbert and Pangborn, R. M., "Intercorrelation of the Senses," *Basic Principles of Sensory Evaluation*, ASTM Special Technical Publication No. 433, American Society for Testing and Materials, Philadelphia, Pennsylvania, 1968, pp 30-46.

Sun Microsystems, Inc., Java Programming Language, WWW URL, *http://www.java.sun.com/*

Tanner, Theodore C. Jr., *Psychoacoustic Criteria for Auditioning Virtual Imaging Systems*, Preprint No. 4568, presented at the *103rd Audio Engineering Society Convention* New York, New York, September 26-29, 1997.

Theile, Günther, "On the Standardization of the Frequency Response of High-Quality Studio Headphones," *Journal of the Audio Engineering Society,* Vol. 34, No. 12, December 1986, pp. 956-969.

Thompson, Richard F., Voss, James F., and Brogden, W. J., "Effect of Brightness of Simultaneous Visual Stimulation on Absolute Auditory Sensitivity," *Journal of Experimental Psychology*, Vol. 55, No. 1, 1958, pp. 45-50.

Thurmond, Bob, "Measurement and Perception Quality in Sound Systems," *Proceedings of the Audio Engineering Society 11th International Conference*, Portland, Oregon, 1992.

Tierney, John, "Jung in Motion, Virtually and Other Computer Fuzz," *The New York Times*, September 16, 1993, pp. C1 and C9.

Toole, Floyd E., "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *Journal of the Audio Engineering Society,* Vol. 33, No. 1/2, January/February 1985, pp. 20-32.

Toole, Floyd E., "Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 1," *Journal of the Audio Engineering Society,* Vol. 34, No. 4, April 1986, pp. 227-235.

Toole, Floyd E., "Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 2," *Journal of the Audio Engineering Society,* Vol. 34, No. 5, May 1986, pp. 323-348.

Toole, Floyd E., "Identifying and Controlling the Variables," *Proceedings of the Audio Engineering Society 8th International Conference*, Washington, D.C., 1990.

Treisman, Anne M., "Strategies and models of selective attention," *Psychological Review,* Vol. 76, 1969, pp. 282-299.

Treisman, Anne M., and Davies, Alison, "Divided Attention to Ear and Eye," *Attention and Performance IV*, Kornblum, Sylvan (Ed.), Academic Press, New York, 1973, pp, 101-117.

Treisman, Anne M., and Gelade, Garry, "A Feature-Integration Theory of Attention," *Cognitive Psychology*, Vol 12., 1980, pp. 97-136.

Vernon, P. E., "Auditory Perception. I. The Gestalt Approach," *British Journal of Psychology*, Vol. 25, 1934, pp. 123-139.

Viemeister, Neil, "An Overview of Psychoacoustics and Auditory Perception," *Proceedings of the Audio Engineering Society 8th International Conference*, Washington, D.C., 1990.

Walk, Richard D., and Pick, Herbert L. Jr., (Eds.) *Intersensory Perception and Sensory Integration*, Plenum Press, New York, 1981.

Warren, David H., Welch, Robert B., and McCarthy, Timothy J., "The role of visual-auditory 'compellingness' in the ventriloquism effect: Implications for transitivity among the spatial senses," *Perception & Psychophysics*, Vol. 30, No. 6, 1981, pp. 557-564.

Wertheimer, Max, Experimentelle Studien über das Sehen von Bewegungen, *Zeitschrift für Psychologie*, Vol. 61, pp.161-265, 1912.

Wever, Ernest Glen, "The Evolution of Vertebrate Hearing," *Handbook of Sensory Physiology*, Vol. V/1, Auditory System, (Eds.) Keidel, Wolf D. and Neff, William D., Springer-Verlag, New York, 1974, pp. 423-454.

Wickens, Christopher D., and Liu, Yili, "Codes and Modalities in Multiple Resources: A Success and a Qualification," *Human Factors*, Vol. 30, No. 5, 1988, pp. 599-616.

Wickens, Christopher D., *Engineering Psychology and Human Performance*, 2nd Ed., Harper Collins Publishers Inc., 1992.

Woszczyk, Wieslaw, Bech, Søren, and Hansen, Villey, "Interaction Between Audio-Visual Factors in a Home Theater System: Definition of Subjective Attributes," Preprint No. 4133, presented at *The 99th Audio Engineering Society Convention*, New York, New York, October 6-9, 1995.

Yakovlev, P. A., "The Influence of Acoustic Stimuli Upon the Limits of Visual Fields for Different Colors," *Journal of the Optical Society of America*, Vol. 28, August, 1938, pp. 286-289.

Yoshikawa, Shokichiro, Noge, Satoru, Funaki, Yasuo, Inoue, Takashi, Sawaguchi, Masaki, Kurozumi, Koichi, and Yamada, Norio, "Monitor Levels and Quality Evaluation of HDTV 3-1 Multichannel Sound," Preprint No. 3723, presented at *The 95th Audio Engineering Society Convention*, New York, New York, October 7-10, 1993.

Zwicker, Eberhard and Zwicker, U. Tilmann, "Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System," *Journal of the Audio Engineering Society*, Vol. 39, No. 3, March 1991, pp. 115-126.

Zyda, Michael and Sheehan, Jerry, (Eds.), *Modeling and Simulation: Linking Entertainment & Defense*, National Academy Press, Washington, D.C., September 1997.

APPENDIX A. LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
CAD	Computer-Aided Design
CD	Compact Disc
CSV	Comma Separated Variable (file format)
COTS	Commercial Off-The-Shelf
FIVE	Framework for Immersive Virtual Environments
HDTV	High-Definition Television
HTML	HyperText Markup Language
JDK	Java Development Kit
JND	Just-Noticeable-Difference
MIUs	Multimedia Information Units
MMX	Multimedia Extensions
MPEG	Motion Picture Expert Group
NPS	Naval Postgraduate School
NRC	National Research Council
PC	Personal Computer
SGI	Silicon Graphics Inc.
VE	Virtual Environment
WM	Working Memory
VRML	Virtual Reality Modeling Language

APPENDIX B. AUDITORY-VISUAL CROSS-MODAL SIGNAL DETECTION AND VIGILANCE BIBLIOGRAPHY

This appendix lists references encountered during the preliminary literature review. These references pertain primarily to studies investigating auditory-visual cross-modal effects in signal detection and vigilance. Since these topics are peripheral to the primary dissertation topic, these references are not included in the main body of the dissertation, but are nevertheless included to provide further insights and observations of auditory-visual intersensory phenomena.

Baker, Robert S., Ware, J. Robert, and Sipowicz, Raymond R., "Vigilance: A Comparison in Auditory, Visual, and Combined Audio-Visual Tasks," *Canadian Journal of Psychology*, Vol. 16, 1962, pp. 192-198.

Banks, William P., Roberts, David, and Ciranni, Michael, "Negative Priming in Auditory Attention," *Journal of Experimental Psychology*, Vol. 21, No. 6, 1995, pp. 1354-1361.

Behar, Isaac, and Bevan, William, "The Perceived Duration of Auditory and Visual Intervals: Cross-Modal Comparison and Interaction," *American Journal of Psychology*, Vol. 74, 1961, pp. 17-26.

Bernstein, Ira H., Clark, Mary H., and Edelstein, Barry, "Effects on an Auditory Signal on Visual Reaction Time," *Journal of Experimental Psychology*, Vol. 80, No. 3, 1969, pp. 567-659.

Brown, A. E., and Hopkins, H. K., "Interaction of the Auditory and Visual Sensory Modalities," *Journal of the Acoustical Society of America*, Vol. 41, No. 1, 1967, pp. 1-6.

Buckner, Donald N., and McGrath, James J., "A comparison of performances on single and dual sensory mode vigilance tasks," in Buckner, Donald N., and McGrath, James J. (Eds), *Vigilance: A symposium*, McGraw-Hill, New York, 1963, pp. 53-69.

Colquhoun, W. Peter, "Evaluation of Auditory, Visual, and Dual-Mode Displays of Prolonged Sonar Monitoring in Repeated Sessions," *Human Factors*, Vol. 17, No. 5, 1975, pp. 425-437.

Corcoran, D.W.J., and Weening, D.L., "On the Combination of Evidence from the Eye and Ear," *Ergonomics*, Vol. 12, No. 3, 1969, pp. 383-394.

Doll, Theodore J., and Hanna, Thomas E., "Enhanced Detection with Bimodal Sonar Displays," *Human Factors*, Vol. 31, No. 5, 1989, pp. 539-550.

Fidell, Sanford, "Sensory Function in Multimodal Signal Detection," *Journal of the Acoustical Society of America*, Vol. 47, No. 4, 1970, pp. 1009-1015.

Gruber, Alin, "Sensory Alternation and Performance in a Vigilance Task," *Human Factors*, Vol. 6, February 1964, pp. 3-12.

Gunn, Walter J., and Loeb, Michel, "Correlation of Performance in Detecting Visual and Auditory Signals," *American Journal of Psychology*, Vol. 80, 1967, pp. 236-242.

Hatfield, Jimmy L., and Loeb, Michel, "Sense mode and coupling in a vigilance task," *Perception and & Psychophysics*, Vol. 4, No. 1, 1968, pp. 29-36.

Hatfield, Jimmy L., and Soderquist, David R., "Coupling Effects and Performance in Vigilance Tasks," *Human Factors*, Vol. 12, No. 4, 1970, pp. 351-359.

Kobus, D.A., Russotti, J., Schlichting, C., Haskell, G., Carpenter, S., and Wojtowicz, J., "Multimodal Detection and Recognition Performance of Sonar Operators," *Human Factors*, Vol. 28, No. 1, 1986, pp. 23-29.

Miller, Jeff, "Channel Interaction and the Redundant-Targets Effect in Bimodal Divided Attention," *Journal of Experimental Psychology*, Vol. 17, No. 1, 1991, pp. 160-169.

Nickerson, Raymond S., "Intersensory Facilitation of Reaction Time: Energy Summation or Preparation Enhancement," *Psychological Review*, Vol. 80, No. 6, 1973, pp. 489-509.

Osborn, William C., Sheldon, Richard W., and Baker, Robert A., "Vigilance Performance under Conditions of Redundant and Nonredundant Signal Presentation," *Journal of Applied Psychology*, Vol. 47, No. 2, 1963, pp. 130-134.

APPENDIX C. SOUND LOCALIZATION, 3D SOUND, AND VIRTUAL ENVIRONMENT BIBLIOGRAPHY

This appendix lists additional references encountered during the preliminary literature review. These references pertain primarily to studies investigating sound localization, 3D sound, and virtual environments. Since these topics are peripheral to the primary dissertation topic, these references are not included in the main body of the dissertation, but are nevertheless included to provide further insights and observations on the perception and use of sound in virtual environments.

Alsaks, Y. A., and Sayers, S. A., "Three Dimensional Sound Simulation using DSP Techniques," *Proceedings of IEEE SOUTHEASTCON '92*, conference date 12-15 April 1992, Birmingham, Al., IEEE, Vol. 1, pp. 234-237.

Anderson, David B., Barrus, John W., Howard, John H., Rich, Charles, Shen, Chia, and Waters, Richard C., "Building Multiuser Interactive Multimedia Environments at MERL," *IEEE Multimedia*, Winter 1995, pp. 77-82.

Aoki, Shigeaki, Cohen, Michael, and Koizumi, Nobuo, "Design and Control of Shared Conferencing Environments for Audio Telecommunication Using Individually Measured HRTFs," *Presence*, Vol. 3, No. 1, Winter 1994, pp. 60-72.

Aoki, Shigeaki, Miyata, Hiroyuki, and Sugiyama, Kiyoshi, "Stereo Reproduction with Good Localization over a Wide Listening Area," *Journal of the Audio Engineering Society*, Vol. 38, No. 6, June 1990, pp. 433-439.

Ashmead, Daniel H., Davis, DeFord L., and Northington, Anna, "Contribution of Listener's Approaching Motion to Auditory Distance Perception," *Journal of Experimental Psychology*, Vol. 21, No. 2, 1995, pp. 239-256.

Apple Computer Inc., *Audio Interchange File Format AIFF-C*, Draft, August 26, 1991.

Axen, Ulrike, "Traversing Alpha Shapes for Processing the Geometrical Data into Sound," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

Ballou, Glen, (Ed.) *Handbook for Sound Engineers: The New Audio Cyclopedia*, 2nd Ed. Howard W. Sams & Company, Carmel, Indiana, 1991.

Bargar, Robin, "Realtime Considerations," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

Bargar, Robin, and Das, Sumit, "Sound for Virtual Immersive Environments," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

Begault, Durand R. and Wenzel, Elizabeth M., "Techniques and Applications for Binaural Sound Manipulation in Human-Machine Interfaces," *NASA Technical Memorandum 102279*, August 1990. Also found later in the *International Journal of Aviation Psychology*, Vol. 2, 1992, pp. 1-22.

Begault, Durand R. and Wenzel, Elizabeth M., *Technical Aspects of a Demonstration Tape for Three-Dimensional Sound Displays*, NASA Technical Memorandum 102826, NASA-Ames Research Center, Moffett Field, California, October 1990.

Begault, Durand R., "Challenges to the Successful Implementation of 3-D Sound," *Journal of the Audio Engineering Society*, Vol. 39, No. 11, November 1991, pp. 864-870.

Begault, Durand R., "Preferred Sound Intensity Increase for Sensation of Half Distance," *Perceptual and Motor Skills*, Vol. 72, 1991, pp. 1019-1029.

Begault, Durand R., "Binaural Auralization and Perceptual Veridicality," presented at *The 93rd AES Convention*, San Francisco, California, October 1-4, 1992.

Begault, Durand R., "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems," *Journal of the Audio Engineering Society*, Vol. 40, No. 11, November 1992, pp. 895-904.

Begault, Durand R. and Wenzel, Elizabeth M., "Headphone Localization of Speech," *Human Factors*, Vol. 35, No. 2, 1993, pp. 361-376.

Begault, Durand R., "Head-up Auditory Displays for Traffic Collision Avoidance System Advisories: A Preliminary Investigation," *Human Factors*, Vol. 35, No. 4, 1993, pp. 707-717.

Begault, Durand R., *Call Sign Intelligibility Improvement Using a Spatial Auditory Display*, NASA Technical Memorandum 104014, NASA-Ames Research Center, Moffett Field, California, April 1993.

Begault, Durand R., and Erbe, Tom, "Multichannel Spatial Auditory Display for Speech Communications," presented at *The 95th AES Convention 1993*, October 7-10, 1993.

Begault, Durand R., and Pittman, Marc T., *3-D Audio Versus Head Down TCAS Displays*, NASA Contractor Report 177636, Contract NCC-2-327, NASA, March 1994. (Also submitted to *International Journal of Aviation Psychology*)

Begault, Durand R., Wenzel, Elizabeth M., Shrum, Richard, and Miller, Joel, "A Virtual Audio Guidance and Alert System for Commercial Aircraft Operations," *The Proceedings of International Conference on Auditory Display (ICAD) 96*, Palo Alto, California, November 4-6, 1996.

Begault, Durand R., *The Sonic CD-ROM for Desktop Audio Production: An Electronic Guide to Producing Computer Audio for Multimedia*, Academic Press, Inc., Cambridge, Massachusetts, 1996.

Begault, Durand R., and Wenzel, Elizabeth M., *3-D Audio Traffic Alert and Collision Avoidance System*, NASA Ames Research Center, Moffett Field, California, 1997. Available at *http://vision.arc.nasa.gov/AFH/Brief/Auditory.S.T./3-D.A.T.html*

Bennett, John C. and Edeko Frederik O., "A New Approach to the Assessment of Stereophonic Sound System Performance," *Journal of the Audio Engineering Society,* Vol. 33, No. 5, May, 1985, pp. 314-321.

Bohn, Dennis A., "Environmental Effects on the Speed of Sound," *Journal of the Audio Engineering Society*, Vol. 36, No. 4, April 1988, pp. 223-231.

Bosi, Marina, *A Real-Time System for Spatial Distribution of Sound*, Center for Computer Research in Music and Acoustics, Department of Music Report No. STAN-M-66, Stanford University, Stanford, California, August 1990.

Brandenburg, Karlheinz, and Bosi, Marina, "Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding," *Journal of the Audio Engineering Society*, Vol. 45, No. 1/2, January/February 1997, pp. 4-21.

Brown, Marc H., and Hershberger, John, "Color and Sound in Algorithm Animation," *IEEE Computer*, December 1992, pp. 52-63.

Bronkhorst, Adelbert W., "Localization of real and virtual sound sources," *Journal of the Acoustical Society of America*, Vol. 98, No. 5, Pt. 1, November 1995, pp. 2542-2553.

Bronkhorst, Adelbert W., Veltman, J. A. (Hans), van Breda, Leo, "Application of a Three-Dimensional Auditory Display in a Flight Task," *Human Factors*, Vol. 38, No. 1, 1996, pp. 23-33.

Brungart, Douglas S., "Distance Simulation in Virtual Audio Displays," in *Proceedings of the IEEE 1993 National Aerospace and Electronics Conference. NAECON 1993*, Dayton, Ohio, Vol. 2, May 24-28, 1993, pp. 612-617.

Burgess, David A., *Real-Time Audio Spatialization with Inexpensive Hardware*, Graphics Visualization and Usability Center, Georgia Institute of Technology, October, 1992.

Burov, V. A., Gurinovich, O. V., and Tagunov, E. Y., "Reconstruction of the Spatial Distribution of the Nonlinearity Parameter and Sound Velocity in Acoustic Nonlinear Tomography," *Acoustical Physics*, Vol. 40, No. 6, 1994, pp. 816-823.

Calhoun, Gloria. L., Valencia, German, and Furness, Thomas. A. III, "Three-Dimensional Auditory Cue Simulation for Crew Station Design/Evaluation," in *Proceedings of the Human Factors Society--31st Annual Meeting*, Santa Monica California, 1987, pp. 1398-1402.

Calhoun, Gloria. L., Janson, W. P., and Valencia, G., "Effectiveness of Three-Dimensional Auditory Directional Cues," in *Proceedings of the Human Factors Society--32st Annual Meeting*, Santa Monica California, 1988, pp. 68-72.

Carlile, Simon, and Wardman, Daniel, "Masking produced by broadband noise presented in virtual auditory space," *Journal of the Acoustical Society of America*, Vol. 100, No. 6, December 1996, pp. 3761-3768.

Chen, Jiashu, Van Veen, Barry D., and Hecox, Kurt E., "A spatial feature extraction and regularization model for the head-related transfer function," *Journal of the Acoustical Society of America*, Vol. 97, No. 1, January 1995, pp. 439-452.

Cherry, E. Colin, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *Journal of the Acoustical Society of America*, Vol. 25, No. 5, September 1953, pp. 975-979.

Chowning, John M., *The Simulation of Moving Sound Sources*, An Audio Engineering Society Preprint, Preprint No. 726 (M-3), Presented at the 38th Convention May 4-7, 1970.

Chowning, John and Sheeline, C., *Auditory Distance Perception Under Natural Sounding Conditions*, Report No. STAN-M-12, Department of Music, Center for Computer Research in Musics and Acoustics (CCRMA), Stanford University, California, November, 1982.

Clifton, Rachel K., Freyman, Richard L., Litovsky, Ruth Y., and McCall, Daniel, "Listeners' expectations about echoes can rise or lower echo threshold," *Journal of the Acoustical Society of America*, Vol. 95, No. 3, March 1994, pp. 1525-1533.

Cohen, Elizabeth A., "Technologies for Three-Dimensional Sound Presentation Issues in Subjective Evaluation of the Spatial Image," April 1997. Available at *http://carbon.cudenver.edu/aes/tech/TECH3D.HTML*

Coleman, Paul D., "Failure to Localize the Source of an Unfamiliar Sound," *Journal of the Acoustical Society of America*, Vol. 34, No. 3, march 1962, pp. 345-346.

Cornell, Gary, and Horstmann, Cay S., *Core JAVA*, SunSoft Press, Mountain View, California, 1996.

Czyzewski, Andrzej., "A Method of Artificial Reverberation Quality Testing," *Journal of the Audio Engineering Society,* Vol. 38, No. 3, March, 1990, pp. 129-141.

Dahl, L., *NPSNET: Aural Cues For Virtual World Immersion*, Master of Computer Science Thesis, Naval Postgraduate School, Monterey, California, September, 1992.

Davis, Mark F., "Loudspeaker Systems with Optimized Wide-Listening-Area Imaging," *Journal of the Audio Engineering Society*, Vol. 35, No. 11, November 1987, pp. 888-896.

Divenyi, Pierre L., and Oliver, Susan K., "Resolution of steady-state sounds in simulated auditory space," *Journal of the Acoustical Society of America*, Vol. 85, No. 5, May 1989, pp. 2042-2052.

Doll, Theodore J., Hanna, Thomas E., and Russotti, Joseph S., "Masking in Three-Dimensional Auditory Displays," *Human Factors*, Vol. 34, No. 3, 1992, pp. 255-265.

Doll, Theodore J., and Hanna, Thomas E., "Spatial and Spectral Release from Masking in Three-Dimensional Auditory Displays," *Human Factors*, Vol. 37, No. 2, 1995, pp. 341-355.

Doan, Tu T., "Understanding MIDI," *IEEE Potentials*, Vol. 13, February 1994, pp. 10-11.

Duda, R., "3-D Sound Perception," presented during the *CCRMA Summer Workshop: Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design*, Stanford University, Stanford, California, June 26 - July 8, 1995.

Durlach, N. I., and Braida L. D., "Intensity Perception. I. Preliminary Theory of Intensity Resolution," *Journal of the Acoustical Society of America*, Vol. 46, No. 2 (Part 2), March 1969, pp. 372-383.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburen, H. S. and Wenzel, E. M., "On the Externalization of Auditory Images," *Presence*, Vol. 1, No. 2, Spring 1992, pp. 251-257.

Elen, Richard, "Ambisonic mixing - an introduction," *Studio Sound*, September 1983. Available at: *http://www.york.ac.uk/inst/mustech/3d_audio/elen/ambimix.htm*

Ericson, M., D'Angelo, W., Scarborough, E., Rodgers, S., Amburn, P., and Ruck, D., "Applications of Virtual Audio," in *Proceedings of the IEEE 1993 National Aerospace and Electronics Conference. NAECON 1993*, Dayton, Ohio, Vol. 2, May 24-28, 1993, pp. 604-611.

Filipanits Jr., Frank, *Design and Implementation of an Auralization System with a Spectrum-Based Temporal Processing Optimization*, unpublished Master's Thesis, University of Miami, Florida, May 1994. Available at: *http://alumni.caltech.edu/~franko/thesis/thesis.html*

Fowler, Barry, "P300 as a Measure of Workload during a Simulated Aircraft landing Task," *Human Factors*, Vol. 36, No. 4, 1994, pp. 670-683.

Freyman, Richard L., Zurek, Patrick M., Balakrishnan, Uma, and Chiang, Yuan-Chuan, "Onset dominance in lateralization," *Journal of the Acoustical Society of America*, Vol. 101, No. 3, March 1997, pp. 1649-1659.

Fu, Ping, "Stepping Into Alpha Shapes," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

Gardner, Bill and Martin, Keith, *HRTF Measurements of a KEMAR Dummy-Head Microphone*, MIT Media Lab Perceptual Computing - Technical Report #280, MIT Media Lab, Massachusetts, May 1994.

Garinther, Georges R., and Anderson, B. Wayne, "Enhanced Armor using the Vehicular Intercommunication System," *Army RD & A*, September-October 1996, pp. 33-35.

Gaver, William W., *Synthesizing Auditory Icons*, Rank Xerox Cambridge EuroPARC, a preprint of a paper submitted to INTERCHI'93, 1993.

Gerzon, Michael A., "Periphony: With-Height Sound Reproduction," *Journal of the Audio Engineering Society*, Vol. 21, No. 1, January/February 1973, pp. 2-10.

Giguere, Christian, and Abel, Sharon M., "Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay," *Journal of the Acoustical Society of America*, Vol. 94, No. 3, Pt. 1, August 1993, pp. 769-776.

Glasgal, Ralph, and Yates, Keith, *Ambiophonics: Beyond Surround Sound to Virtual Sonic Reality*, Ambiophonics Institute, Northvale, NJ, 1995.

Good, Michael, D., and Gilkey, Robert H., "Sound localization in noise: The effect of signal-to-noise ratio," *Journal of the Acoustical Society of America*, Vol. 99, No. 2, February 1996, pp. 1108-1117.

Hagsand, Olof, "Interactive Multiuser VEs in the DIVE System," *IEEE Multimedia*, Spring 1996, pp. 30-39.

Hahn, James K., Hesham, Fouad, Gritz, Larry, and Lee, Jong W., "Integrating Sounds in Virtual Environments," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

Hartmann, William Morris, Rakerd, Brad, "Localization of sound in rooms IV: The Franssen effect," *Journal of the Acoustical Society of America*; Vol. 86, No. 4, October 1989, pp. 1366-1373.

Hartmann, William Morris, and Rakerd, Brad, "Auditory spectral discrimination and the localization of clicks in the sagittal plane," *Journal of the Acoustical Society of America*, Vol. 94, No. 4, October 1993, pp. 2083-2092.

Hartmann, William M., and Wittenberg, Andrew, "On the externalization of sound images," *Journal of the Acoustical Society of America*, Vol. 99, No. 6, June 1996, pp. 3678-3688.

Heller, Rachelle S., and Martin, C. Dianne, "A Media Taxonomy," *IEEE Multimedia*, Winter 1995, pp. 36-45.

Holt, Robert E., and Thurlow, Willard R., "Subject Orientation and Judgment of Distance of a Sound Source," *Journal of the Acoustical Society of America*, Vol. 46, No. 6 (Part 2), 1969, pp. 1584-1585.

International MIDI Association, *1.0 MIDI Specification*, 1983.

Kang, George S., and Heide, David A., "Canned Speech for Tactical Voice Message Systems," presented at *The 1992 Tactical Communication Conference*, Fort Wayne, Indiana, April 28-30, 1992.

Karr, Clark R., Reece, Douglas, and Franceschini, Robert, "Synthetic soldiers," *IEEE Spectrum*, March 1997, pp. 39-45.

Kennedy, Robert S., Berbaum, Kevin S., Collyer, Stanley C., May, James G, and Dunlap, William, P., "Spatial Requirements for Visual Simulation of Aircraft at Real-World Distances," *Human Factors*, Vol. 30, No.2, 1988, pp. 153-161.

Kidd, Jr., Gerald, Mason, Christine R., and Rohtla, Tanya L., "Binaural advantage for sound pattern identification," *Journal of the Acoustical Society of America*, Vol. 98, No. 4, October 1995, pp. 1977-1986.

Kim, Youngmoo, *Sound Localization in the Median Plane*, Music 151 Final Project, Stanford University, Stanford, California, December 15, 1993.

Kistler, Doris J., and Wightman, Frederic L., "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *Journal of the Acoustical Society of America*, Vol. 91, No. 3, March 1992, pp. 1637-1647.

Konishi, Masakazu, "Listening with Two Ears," *Scientific American*, April 1993, pp. 66-73.

Konrad, Christopher M., Kramer, Arthur F., Watson, Stephen E., and Weber, Timothy A., "A Comparison of Sequential and Spatial Displays in a Complex Monitoring Task," *Human Factors*, Vol. 38, No. 3, 1996, pp. 464-483.

Kozhevnikova, I. K., and Samokhin, V. F., "Sound Sources of a Tail-Rotor Helicopter," *Acoustical Physics*, Vol. 40, No. 6, 1994, pp. 852-858.

Lakatos, Stephen, *Temporal Constraints on Apparent Motion in Auditory Space*, Center for Computer Research in Music and Acoustics, Department of Music Report No. STAN-M-74, Stanford University, Stanford, California, November 1991.

Lapsley, Phil, Bier, Jeff, Shoham, Amit, and Lee, Edward A., *DSP Processor Fundamentals: Architectures and Features*, Berkeley Design Technologies, Inc, 1996.

Lehnert, H. and Blauert, J., "Virtual Auditory Environment," *91 ICAR. Fifth International Conference on Advanced Robotics. Robots in Unstructured Environments*, June 19-22, 1991,Vol. 1, IEEE, New York, New York, pp. 211-216.

Levergood, Thomas M., Payne, Andrew C., Gettys, James, Treese, G. Winfield, and Stewart, Lawrence C., *AudioFile: A Network-Transparent System for Distributed Audio Applications*, Technical Report Series, CRL 93/8, Digital Equipment Corporation, Cambridge Research Lab, Cambridge, Massachusetts, June 11, 1993.

Litovsky, Ruth Y., and Clifton, Rachel K., "Use of sound-pressure level in auditory distance discrimination by 6-month-old infants and adults," *Journal of the Acoustical Society of America*, Vol. 92, No. 2, Pt. 1, August 1992, pp. 794-802.

Litovsky, Ruth and Macmillan, Neil A., "Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli," *Journal of the Acoustical Society of America*, Vol. 96, No. 2, Pt. 1, August 1994, pp. 752-758.

Loomis, Jack M., Hebert, Chick, and Cicinelli, Joseph G., "Active localization of virtual sounds," *Journal of the Acoustical Society of America*, Vol. 88, No. 4, October 1990, pp. 1757-1764.

Lytle, Wayne, "Music Animation," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

Makous, James C., and Middlebrooks, John C., "Two-dimensional sound localization by human listeners," *Journal of the Acoustical Society of America*, Vol. 87, No. 5, May 1990, pp. 2188-2200.

Malham, D.G., "3-D sound for virtual reality systems using Ambisonic techniques," presented at the *VR93 Conference*, London, England, April 1993. Available at: *http://www.york.ac.uk/inst/mustech/3d_audio/vr93papr.htm*

Marks, Lawrence E., "Contextual Processing of Multidimensional and Unidimensional Auditory Stimuli," *Journal of Experimental Psychology*, Vol. 19, No. 2, 1993, pp. 227-249.

Marks, Lawrence E., "'Recalibrating' the Auditory System: The Perception of Loudness," *Journal of Experimental Psychology*, Vol. 20, No. 2, 1994, pp. 382-396.

Martens, William, *Spatial Image Formation in Binocular Vision and Binaural Hearing*, paper presented at the 3D Media Technology Conference, Montreal, Canada, June 1, 1989.

Martens, William, *Demystifying Spatial Audio*, Ono-Sendai Corporation, San Francisco, California, 1992.

Martins, William, "Spatial Sound at SIGGRAPH: Is it 3D?," *CyberEdge Journal*, September/October, 1995. Available at: *http://www.cyberedge.com/6i3.html*

McEachern, Robert, "How the Ear Really Works," *Proceedings of the IEEE-SP International Symposium Time-Frequency and Time-Scale Analysis*, conference date October 4-6, 1992,Victoria, BC, Canada, pp. 437-440.

McMillen, Keith, Wessel, David L., and Wright, Matthew, "The ZIPI Music Parameter Description Language," *Computer Music Journal*, Vol. 18, Winter, 1994.

McMillen, Keith, "ZIPI: Origins and Motivations," *Computer Music Journal*, Vol. 18, Winter 1994.

McMillen, Keith, Simon, David, and Wright, Matthew, "A Summary of the ZIPI Network," *Computer Music Journal*, Vol. 18, Winter 1994.

Middlebrooks, John C., "Narrow-band sound localization related to external ear acoustics," *Journal of the Acoustical Society of America*, Vol. 92, No. 5, November 1992, pp. 2607-2624.

Miner, Nadine, and Caudell, Thomas, "Computational Requirements and Synchronization Issues of Virtual Acoustic Displays," submitted to *Presence*, April 1997.

Moog, Bob, "MIDI: Musical Instrument Digital Interface," *Journal of the Audio Engineering Society*, Vol. 34, No. 5, May 1986, pp. 394-404.

Moorer, James. A., "About This Reverberation Business," *Computer Music Journal*, Vol. 3, No. 2, 1979, pp. 13-28.

Mulligan, B. E., Mulligan, M. J., and Stonecypher, J. F., "Critical Band in Binaural Detection," *Journal of the Acoustical Society of America*, Vol. 41, No. 1, 1967, pp. 7-12.

Munshi, Anees S., "Equalization of Room Acoustics," *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech and Signal Processing*, Conference Date, March 23-26, 1992, Vol 2, IEEE, 1992, pp. 217-220.

Neuhoff, John G., and McBeath, Michael K., "The Doppler Illusion: The Influence of Dynamic Intensity Change on Perceived Pitch," *Journal of Experimental Psychology*, Vol. 22, No. 4, 1996, pp. 970-985.

O'Donnell, Bob, "What is MIDI, Anyway?," *Electronic Musician*, January, 1991, pp. 74-76.

Pan, Davis, "A Tutorial on MPEG/Audio Compression," *IEEE Multimedia*, Summer 1995, pp. 60-74.

Perceptronics, *SIMNET - M1 Sound System Interface Protocol*, August 18, 1986.

Perrott, David R., Marlborough, Kent, Merrill, Paul, and Strybel, Thomas, "Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate," *Journal of the Acoustical Society of America*, Vol. 85, No. 1, January 1989, pp. 282-288.

Perrott, David R., and Saberi, Kourosh, "Minimum audible angle thresholds for sources varying in both elevation and azimuth," *Journal of the Acoustical Society of America*, Vol. 87, No. 4, April 1990, pp. 1728-1731.

Perrott, David R., Sadralodabai, Toktam, Saberi, Kourosh, and Strybel, Thomas Z., "Aurally Aided Visual Search in the Central Visual Field: Effects of Visual Load and Visual Enhancement of the Target," *Human Factors*, Vol. 33, No. 4, 1991, pp. 389-400.

Perrott, David R., Costantino, Brian, and Cisneros, John, "Auditory and visual localization performance in a sequential discrimination task," *Journal of the Acoustical Society of America*, Vol. 93, No. 4, Pt. 1, April 1993, pp. 2134-2138.

Perrott, David R., Cisneros, John, McKinley, Richard L., and D'Angelo, William, "Aurally Aided Visual Search under Virtual and Free-Field listening Conditions," *Human Factors*, Vol. 38, No. 4, 1996, pp. 702-715.

Plenge, G., "On the differences between localization and lateralization," *Journal of the Acoustical Society of America*, Vol. 56, No. 3, September 1974, pp. 944-951.

Pralong, Daniele, Carlile, Simon, "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," *Journal of the Acoustical Society of America*, Vol. 100, No. 6, December 1996, pp. 3785-3793.

Pratt, Jay, and Abrams, Richard A., "Inhibition of Return to Successively Cued Spatial Locations," *Journal of Experimental Psychology*, Vol. 21, No. 6, 1995, pp. 1343-1353.

Proakis, John G., and Manolakis, Dimitris G., *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd Ed., Prentice Hall, Upper Saddle River, New Jersey, 1996.

Ranga, E., "A Three Speaker Stereo Sound System," presented at the conference *IEE Colloquium on 'Vehicle Audio Systems' (Digest No. 183)*, London, United Kingdom, December 6, 1991, pp. 3/1-3/2.

Rayleigh, Lord Strutt J., "On Our Perception of Sound Direction," *Philosophical Magazine*, Vol. 13, pp. 214-232, 1907.

Reichbach, Jonathan D., and Kemmerer, Richard A., "SoundWorks: An Object-Oriented Distributed System for Digital Sound," *IEEE Computer*, March 1992, pp. 25-37.

Ricard, Gilbert L., and Meirs, Susan L., "Intelligibility and Localization of Speech from Virtual Directions," *Human Factors*, Vol. 36, No. 1, 1994, pp. 120-128.

Robinson, Christopher P., and Eberts, Ray E., "Comparison of Speech and Pictorial Displays in a Cockpit Environment," *Human Factors*, Vol. 29, No. 1, 1987. pp. 31-44.

Roesli, John, *Free-Field Spatialized Aural Cues for Synthetic Environments*, Master of Computer Science Thesis, Naval Postgraduate School, Monterey, California, September, 1994.

Rossing, Thomas D., *The Science of Sound*, 2nd Ed., Addison-Wesley, Reading Massachusetts, 1990.

Saberi, Kourosh, and Perrott, David R., "Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate," *Journal of the Acoustical Society of America*, Vol. 87, No. 4, April 1990, pp. 1732-1737.

Saberi, Kourosh, and Perrott, David R., "Minimum audible movement angles as a function of sound source trajectory," *Journal of the Acoustical Society of America*, Vol. 88, No. 6, December 1990, pp. 2639-2644.

Salava, Tomas, "Acoustic Load and Transfer Functions in Rooms at Low Frequencies," *Journal of the Audio Engineering Society*, Vol. 36, No. 10, October 1988, pp. 763-775.

Salava, Tomas, "Low-Frequency Performance of Listening Rooms for Steady-State and Transient Signals," *Journal of the Audio Engineering Society*, Vol. 39, No. 11, November 1991, pp. 853-863.

Schroeder, M. R., "Digital Simulation of Sound Transmission in Reverberant Spaces," *Journal of the Acoustical Society of America*, Vol. 47, No. 2 (Part 1), 1970, pp. 424-431.

Schroeder, Manfred. R., "Statistical Parameters of the Frequency Response Curves of Large Rooms," *Journal of the Audio Engineering Society,* Vol. 35, No. 5, May 1987, pp. 299-306.

Schroeder, Manfred R., "Normal Frequency and Excitation Statistics in Rooms: Model Experiments with Electric Waves," *Journal of the Audio Engineering Society,* Vol. 35, No. 5, May 1987, pp. 307-316.

Sellen, Abigail J., "Remote Conversations: The Effects of Mediating Talk With Technology," *Human-Computer Interaction*, Vol. 19, 1995, pp. 401-444.

Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K., "Cross-frequency interactions in the precedence effect," *Journal of the Acoustical Society of America*, Vol. 98, No. 1, July 1995, pp. 164-171.

Silicon Graphics, "Adding Attitude to Your Application with Audio," *Pipeline*, Silicon Graphics, Vol. 4, No. 3, May/June 1993.

Smith, Julius O., and Abel, Jonathan S., "Closed-Form Least-Squares Source Location Estimation from Range-Difference Measurements," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-35, No. 12, December 1987, pp. 1661-1669.

Sorkin, Robert D., Wightman, Frederic L., Kistler, Doris S., and Elvers, Greg C., "An Exploratory Study of the Use of Movement-Correlated Cues in an Auditory Head-Up Display," *Human Factors*, Vol. 31, No. 2, 1989, pp. 161-166.

Storms, Russell, and Roesli, John T., *NPSNET-PAS: A Networked Real-Time Polyphonic Free-Field Audio Spatializer*, NPSNET Research Group, Naval Postgraduate School, Monterey, California, November 1994.

Storms, Russell, *Headphones Versus Free-Field Systems for Generating Three-Dimensional Sound in Virtual Environments*, NPSNET Research Group, Naval Postgraduate School, Monterey, California, January 1995.

Storms, Russell, *Notes Relating to 3D Sound*, from the CCRMA Summer Workshops 1995, NPSNET Research Group, Naval Postgraduate School, Monterey, California, July 1995.

Storms, Russell L., *NPSNET-3D Sound Server: An Effective Use of the Auditory Channel*, Master's Thesis, Naval Postgraduate School, Monterey, California, September 1995.

Storms, Russell, Biggs, Lloyd, Cockayne, William, Barham, Paul, Falby, John, Brutzman, Don, and Zyda, Michael, "The Auralization and Acoustics Laboratory," *Proceedings of the International Conference on Auditory Displays (ICAD)*, Palo Alto, California, November 1996.

Strybel, Thomas Z., Manligas, Carol L, and Perrott, David R., "Minimum Audible Movement Angle as a Function of the Azimuth and Elevation of the Source," *Human Factors*, Vol. 34, No. 3, 1992, pp. 267-275.

Strybel, Thomas Z., and Neale, Wayne, "The effect of burst duration, interstimulus onset interval, and loudspeaker arrangement on auditory apparent motion in the free field," *Journal of the Acoustical Society of America*, Vol. 96, No. 6, December 1995, pp. 3463-3475.

Takala, Tapio and Hahn, James, "Sound Rendering," *Computer Graphics*, Vol. 26, No. 2, July 1992, pp. 211-220.

Takala, Tapio, Hahn, James K., Gritz, Larry, Geigel, Joe, and Lee, Jong W., "Using Physically-Based Models and Genetic Algorithms for Functional Composition of Sound Signals, Synchronized to Animated Motion," *Proceedings of ICMC93 (International Computer Music Conference)*, September 10-15, 1993, Tokyo, Japan.

Takala, Tapio and Hahn, James, "Sound Rendering," Course Number 12. Sound Synchronization and Synthesis for Computer Animation and VR, presented at *SIGGRAPH '94*, Orlando, Florida, 1994.

Theile, Günther, "On the Naturalness of Two-Channel Stereo Sound," *Journal of the Audio Engineering Society*, Vol. 39, No. 10, October 1991, pp. 761-767.

Tonnesen, Cindy and Steinmetz, Joe, *3D Sound Synthesis*, January 1995, available at: *http://www.cs.umd.edu/projects/hcil/eve.restore/eve-articles/I.B.1.3DSoundSynthesis.html*

Tyler, Dolores M., Waag, Wayne L., and Halcomb, Charles G., "Monitoring Performance Across Sense Modes: An Individual Differences Approach," *Human Factors*, Vol. 14, No. 6, 1972, pp. 539-547.

Vernon, P. E., "Auditory Perception. II. The Evolutionary Approach," *British Journal of Psychology*, Vol. 25, 1935, pp. 265-283.

Verschuur, D. J., Kaizer, A. J., Druyvesteyn, W. F., and De Vries, D., "Wigner Representation of Loudspeaker Responses in a Living Room," *Journal of the Audio Engineering Society*, Vol. 36, No. 4, April 1988, pp. 203-212.

Vreuls, Donald, and Obermayer, Richard W., "Human-System Performance Measurement in Training Simulators," *Human Factors*, Vol. 27, No. 3, 1985, pp. 241-250.

Wagenaars, W. M., "Localization of Sound in a Room with Reflecting Walls," *Journal of the Audio Engineering Society*, Vol. 38., No. 3, March 1990, pp. 99-110.

Watkins, William H., and Feehrer, Carl E., "Acoustic Facilitation of Visual Detection," *Journal of Experimental Psychology*, Vol. 70, No. 3, 1965, pp. 332-333.

Wenzel, Elizabeth M., Wightman, Frederic, Kistler, Doris, and Foster, Scott H., "Acoustic origins of individual differences in sound localization," *Journal of the Acoustical Society of America*, Vol. 84, Suppl. 1, Fall 1988, p. S79.

Wenzel, Elizabeth M., and Foster, Scott H., "Realtime Digital Synthesis of Virtual Acoustic Environments," *Computer Graphics*, Vol. 24, No. 2, March 1990, pp. 139-140.

Wenzel, Elizabeth. M., *Three-Dimensional Virtual Acoustic Displays*, NASA Technical Memorandum 103835, July 1991.

Wenzel, Elizabeth M., "Localization in Virtual Acoustic Displays," *Presence*, Vol. 1, No. 1, Winter 1992, pp. 80-107.

Wenzel, Elizabeth M., Arruda, Marianne, Kistler, Doris. J., and Wightman, Frederic. L., "Localization using nonindividualized head-related transfer functions," *Journal of the Acoustical Society of America*, Vol. 94, No. 1, July 1993, pp. 111-123.

Wenzel, Elizabeth M., and Begault, Durand R., *Localization in Reflective Environments*, NASA Ames Research Center, Moffett Field, California, 1997. Available at *http://vision.arc.nasa.gov/AFH/Brief/Auditory.S.T./Localization.R.html*

Wenzel, Elizabeth M., and Begault, Durand R., *Measurement of Personalized HRTFs*, NASA Ames Research Center, Moffett Field, California, 1997. Available at *http://vision.arc.nasa.gov/AFH/Brief/Auditory.S.T./Measurement.P.html*

Wenzel, Elizabeth M., and Begault, Durand R., *The Role of Dynamic Information in Virtual Acoustic Displays*, NASA Ames Research Center, Moffett Field, California, 1997. Available at *http://vision.arc.nasa.gov/AFH/Brief/Auditory.S.T./The.Role.of.D.html*

Wenzel, Elizabeth M., and Begault, Durand R., *Terminal Area Productivity (TAP) Program -- Taxi Navigation and Situation Awareness (T-NASA) System: 3-D Audio Ground Collision Avoidance System (GCAS) & Navigation System*, NASA Ames Research Center, Moffett Field, California, 1997. Available at *http://vision.arc.nasa.gov/AFH/Brief/Auditory.S.T./Terminal.A.html*

Wheeler, Andrew, Ellinger, Joshua, and Glicker, Steven, *The Design and Implementation of an Experimental Virtual Acoustic Display*, Applied Research Laboratories and the Electrical and Computer Engineering Department, The University of Texas at Austin, GR-EM-93-1, February 14, 1993.

Wiener, Francis M., and Ross, Douglas A. "The Pressure Distribution in the Auditory Canal in a Progressive Sound Field," *Journal of the Acoustical Society of America*, Vol. 18, No. 2, October 1946, pp. 401-408.

Wightman, Frederic L. and Kistler, Doris J., "Headphone Simulation of Free-field Listening I: Stimulus Synthesis," *Journal of the Acoustical Society of America*, Vol. 85, No. 2, February 1989, pp. 858-867.

Wightman, Frederic L. and Kistler, Doris J., "Headphone Simulation of Free-field Listening II: Psychophysical Validation," *Journal of the Acoustical Society of America*, Vol. 85, No. 2, February 1989, pp. 868-878.

Wightman, Frederic L., and Kistler, Doris J., "Monaural sound localization revisited," *Journal of the Acoustical Society of America*, Vol. 101, No. 2, February 1997, pp. 1050-1063.

Wright, Donald, Hebrank, John H., and Wilson, Blake, "Pinna reflections as cues for localization," *Journal of the Acoustical Society of America*, Vol. 56, No. 3, September 1974, pp. 957-962.

Wright, Matthew, "Answers to Frequently Asked Questions About ZIPI," *Computer Music Journal*, Vol. 18, Winter 1994.

Wright, Matthew, "Examples of ZIPI Applications," *Computer Music Journal*, Vol. 18, Winter 1994.

Yoshikawa, Shokichiro, Noge, Satoru, Yamamoto, Takeo, and Saito, Keishi, *Does High Sampling Freqiency Improve Perceptual Time-Axis Resolution of Digital Audio Signal?*, An Audio Engineering Society Preprint, Preprint No. 4562 (I-3), Presented at the 103rd Convention September 26-29, 1997.

Zakarauskas, Pierre, and Cynader, Max S., "A computational theory of spectral cue localization," *Journal of the Acoustical Society of America*, Vol. 94, No. 3, Pt. 1, September 1993, pp. 1323-1331.

Ziomek, Lawrence J., *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing,* CRC Press, Boca Raton, Florida, 1995.

Zyda, M., Pratt, D., Falby, J., Lombardo, C. and Kelleher, K., "The Software Required for the Computer Generation of Virtual Environments," *Presence*, Vol. 2, No. 2, Spring 1993, pp. 130-140.

Zyda, M., Pratt, D., Falby, J., Barham, P. and Kelleher, K., "NPSNET and the Naval Postgraduate School Graphics and Video Laboratory," *Presence*, Vol. 2, No. 3, Summer 1993, pp. 244-258.

APPENDIX D. INTERNET RESOURCES

The first section of this appendix contains the URL's of some research institutions which are currently doing research in various aspects of sound. The second section contains the URL's of various sound related commercial products.

Auditory Perception Lab, Dept. of Psychology, University of California, Berkeley: *http://ear.berkeley.edu/auditory_lab/*

Center for Computer Research in Music and Acoustics (CCRMA), Dept. of Music, Stanford University: *http://ccrma-www.stanford.edu/Welcome.html*

Center for Experimental Music and Intermedia (CEMI), University of North Texas: *http://www.scs.unt.edu/cemi/cemi.htm*

Center for New Music and Audio Technologies (CNMAT), University of California, Berkeley: *http://www.cnmat.berkeley.edu/*

Center for Research in Computing and the Arts (CRCA), University of California, San Diego: *http://crca-www.ucsd.edu*

Center for Research in Electronic Art Technology (CREATE), Dept. of Music, University of California, Santa Barbara: *http://www.ccmrc.ucsb.edu/*

Center for Studies in Music Technology (CSMT), Yale University: *http://www.music.yale.edu:/*

Dipartimento di Ingegneria Industriale, University of Parma, Angelo Farina: *http://pcfarina.eng.unipr.it/*

Faculty of Music, McGill University, Montréal: *http://www.music.mcgill.ca/*

Graphics, Visualization, and Usability Center, Georgia Tech: *http://www.cc.gatech.edu/gvu/multimedia/*

Harvard Computer Music Center, Harvard University: *http://www-mario.harvard.edu*

Hearing Development Research Laboratory (HDRL), Waisman Center, University of Wisconsin: *http://www.waisman.wisc.edu/hdrl/*

Human Interface Technology Lab (HIT LAB), University of Washington: *http://www.hitl.washington.edu/*

Human Research and Engineering Directorate (HRED), Army Research Laboratory: *http://www.arl.mil/ARL-Directorates/HRED/hred.html*

Image Synthesis Group, Dept. of Computer Science, Trinity College, Dublin: *http://vangogh.cs.tcd.ie*

Institut de Recherche et Coordination Acoustique/Musique (IRCAM), Institute for Acoustic/Music Research: *http://www.ircam.fr*

Interval Research Corporation, Palo Alto, California: *http://www.interval.com*

Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology (HUT): *http://www.hut.fi/HUT/Acoustics/index.html*

Machine Listening Group, MIT Media Lab, Massachusetts Institute of Technology: *http://sound.media.mit.edu/*

National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign: http://www.ncsa.uiuc.edu/

NASA Ames Research Center, Moffett Field, California: *http://www.arc.nasa.gov/*

NAVE Research Group, Dept. of Computer Science, University of Colorado at Boulder: *http://www.cs.colorado.edu/~cboyd/*

Norwegian network for Technology, Acoustics and Music (NoTAM), University of Oslo: *http://www.notam.uio.no/index-e.html*

Parmly Hearing Institute, Loyola University Chicago: *http://parmly-2.ls.luc.edu/parmly/*

Princeton Sound Kitchen, Princeton University: *http://www.music.princeton.edu:80/PSK/*

SCCP Virtual Reality SOUND, University of Aizu: *http://www-ci.u-aizu.ac.jp/VirtualReality/WWW/sound.html*

Sound Localization Research, San Jose University: *http://www-engr.sjsu.edu/~duda/Duda.Research.html*

Visual Systems Laboratory, University of Central Florida: *http://www.vsl.ist.ucf.edu/*

The WORLDSONG Project: *http://www.hyperreal.com/~mpesce/worldsong.html*

York University Music Technology Group, The University of York: *http://www.york.ac.uk/inst/mustech/3d_audio/ambison.htm*

This section contains the URL's of various sound related commercial products.

AdB International Corporation: *http://www.adbdigital.com/*

Aureal Semiconductor: *http://www.aureal.com*

The Binaural Source: *http://www.btown.com/binaural.html*

CATT: *http://www.netg.se/~catt/*

Chromatic Research: *http://www.chromatic.com/*

Circle Surround: *http://www.surround.net/*

Creative Labs: *http://www.creaf.com/*

Crystal River Engineering: *http://www.cre.com/index.html*

DirectSound Xtra: *http://www.directxtras.com/ds_home.htm*

Dolby Laboratories: *http://www.dolby.com/*

E-mu Systems Inc.: *http://www.emu.com/*

Ensoniq Corporation: *http://www.ensoniq.com/*

Firsthand: *http://www.firsthand.com/*

HeadRoom: *http://headroom.headphone.com/*

Headspace: *http://www.headspace.com*

HoonTech: *http://www.hoontech.co.kr/hoontech_eng.html*

Lake DSP: *http://www.lakedsp.com/*

Level Control Systems: *http://www.lcsaudio.com/lcs.html*

Lexicon: *http://www.lexicon.com/*

MIDI Home Page: *http://www.eeb.ele.tue.nl/midi/index.html*

MIDI Manufacturers Association: *http://www2.midi.org/mma/*

Muscle Fish: *http://www.musclefish.com/*

NuReality: *http://www.nureality.com/*

Paradigm Simulation Inc.: *http://www.paradigmsim.com/*

Pyramid Systems: *http://tmgweb.com/psi/*

Qsound: *http://www.qsound.ca/*

RealAudio: *http://www.real.com/*

Reality by Design, Inc.: *http://www.rbd.com/*

Realistic Sound Experience (RSX) Technology: *http://www.intel.com/ial/rsx/*

Roland Sound Space: *http://www.rolandcorp.com/products/PA/RSS-10.html*

SENSE8: *http://www.sense8.com/*

Sound Retrieval System (SRS): *http://www.srslabs.com/*

Sony IMAX Theatre: *http://www.spe.sony.com/Pictures/sonytheatres/imax/imaxtech.html*

Spatializer Audio Laboratories: *http://www.catalog.com/cgibin/var/3dstereo/index.html*

Symbolic Sound Corporation: *http://www.SymbolicSound.com/*

THX: *http://www.thx.com/*

Tucker-Davis Technologies Inc.: *http://tdt-quikki.com/*

Unofficial SGI Audio Apps List: *http://reality.sgi.com/employees/cook/audio.apps/*

Virtual Audio Imager (VAI): *http://www.purestereo.com/brown.html*

Visual Synthesis Incorporated (VSI): *http://www.vsicorp.com/*

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, CA 93943-5101

3. Chairman, Code CS 2
   Computer Science Department
   Naval Postgraduate School
   Monterey, CA 93943-5000

4. Dr. Michael J. Zyda, Code CS/Zk 5
   Computer Science Department
   Naval Postgraduate School
   Monterey, CA 93943-5118

5. Dr. Robert B. McGhee, Code CS/Mz 1
   Computer Science Department
   Naval Postgraduate School
   Monterey, CA 93943-5118

6. Dr. Rudolph P. Darken, Code CS/Dr 1
   Computer Science Department
   Naval Postgraduate School
   Monterey, CA 93943-5118

7. Dr. Don Brutzman, Code UW/Br 1
   MOVES Academic Group
   Naval Postgraduate School
   Monterey, CA 93943-5118

8. Dr. Lawrence J. Ziomek, Code EC/Zm 1
   Electrical and Computer Engineering Department
   Naval Postgraduate School
   Monterey, CA 93943-5121

9. Dr. Elizabeth M. Wenzel 1
   NASA-Ames Research Center
   Moffet Field, CA 94035-1000

10. Dr. Durand R. Begault 1
    NASA-Ames Research Center
    Moffet Field, CA 94035-1000

11. CDR Michael J. Holden, Code CS/Hm 1
Computer Science Department
Naval Postgraduate School
Monterey, CA 93943-5118

12. Eric Bachmann, Code CS 1
Computer Science Department
Naval Postgraduate School
Monterey, CA 93943-5118

13. CAPT Frank C. Petho, Code NS/Pe 1
National Security Affairs Department
Naval Postgraduate School
Monterey, CA 93943-5218

14. Mr. Chuck Dachis 1
4500 Russell Dr.
Austin, TX 78745

15. Dr. George R. Talbott 1
4031 Charter Oak Drive
Orange, CA 92667

16. Mr. Fred Sherman 1
1 Stepping Stone Lane
Kings Point, NY 11024

17. Alan Silvestri 1
72 Fern Canyon Road
Carmel, CA 93923

18. Brenda Laurel 1
Interval Research Corporation
1801 Page Mill Road, Building C
Palo Alto, CA 94304

19. Jim Ballas 1
NCARAI
Naval Research Laboratory
Code 5513
Washington DC 20375-5337

20. Robin Bargar 1
National Center for Supercomputing Applications Beckman Institute
405 S. Mathews
Urbana, IL 61801

21. Meera Blattner 1
Department of Applied Science
University of California, Davis
One Shields Avenue, Hertz Hall
Davis, CA 95616

22. Capt. Jay Kistler, USN 1
N6M
2000 Navy Pentagon
Room 4C445
Washington, DC 20350-2000

23. George Phillips 1
CNO, N6M1
2000 Navy Pentagon
Room 4C445
Washington, DC 20350-2000

24. Dr. Mike Macedonia 1
Chief Scientist and Technical Director
US Army STRICOM
12350 Research Parkway
Orlando, FL 32826-3276

25. National Simulation Center (NSC) 1
ATTN:ATZL-NSC (Jerry Ham)
410 Kearney Avenue --- Building 45
Fort Leavenworth, KS 66027-1306

26. Director 1
Office of Science & Innovation
OSI, MCCDC
3300 Russell Road
Quantico, VA 22134-5021

27. Capt. Dennis McBride, USN 1
Office of Naval Research (341)
800 No. Quincy Street
Arlington, VA 22217-5660

28. Col. Crash Konwin, USAF 1
DMSO
1901 N. Beauregard St.
Suite 504
Alexandria, VA 22311

29. Sid Kissen 1
National Security Agency
Attn: S312
9800 Savage Road
Fort George G. Meade, MD 20755

30. Nancy Spurlin 1
STRICOM
ATTN: AMSTI-RFE
12350 Research Parkway
Orlando, FL 32826

31. Norma L. Chandler 1
19325 SW 344th Street
Homestead, FL 33034

32. George L. Storms... 1
14465 SW 292 Street
Leisure City, FL 33033

33. Dr. Russell L. Storms 1
404 Cana of Galilee Court
Tucker, GA 30084