2007

Design For Auditory Displays: Identifying Temporal And Spatial Information Conveyance Principles

Ali Ahmad

University of Central Florida

DESIGN FOR AUDITORY DISPLAYS: IDENTIFYING TEMPORAL AND SPATIAL INFORMATION CONVEYANCE PRINCIPLES

by

ALI AHMAD
B.S. University of Jordan, 2000
M.S. University of Central Florida, 2003

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2007

Major Professor: Kay M. Stanney

ABSTRACT

Designing auditory interfaces is a challenge for current human-systems developers. This is largely due to a lack of theoretical guidance for directing how best to use sounds in today's visually-rich graphical user interfaces. This dissertation provided a framework for guiding the design of audio interfaces to enhance human-systems performance.

This doctoral research involved reviewing the literature on conveying temporal and spatial information using audio, using this knowledge to build three theoretical models to aid the design of auditory interfaces, and empirically validating select components of the models. The three models included an audio integration model that outlines an end-to-end process for adding sounds to interactive interfaces, a temporal audio model that provides a framework for guiding the timing for integration of these sounds to meet human performance objectives, and a spatial audio model that provides a framework for adding spatialization cues to interface sounds. Each model is coupled with a set of design guidelines theorized from the literature, thus combined, the developed models put forward a structured process for integrating sounds in interactive interfaces.

The developed models were subjected to a three phase validation process that included review by Subject Matter Experts (SMEs) to assess the face validity of the developed models and two empirical studies. For the SME review, which assessed the utility of the developed models and identified opportunities for improvement, a panel of three audio experts was selected to respond to a Strengths, Weaknesses, Opportunities, and Threats (SWOT) validation questionnaire. Based on the SWOT analysis, the main strengths of the models included that they provide a systematic approach to auditory display design and that they integrate a wide variety of knowledge sources in a concise manner. The main weaknesses of the models included the lack

of a structured process for amending the models with new principles, some branches were not considered parallel or completely distinct, and lack of guidance on selecting interface sounds. The main opportunity identified by the experts was the ability of the models to provide a seminal body of knowledge that can be used for building and validating auditory display designs. The main threats identified by the experts were that users may not know where to start and end with each model, the models may not provide comprehensive coverage of all uses of auditory displays, and the models may act as a restrictive influence on designers or they may be used inappropriately. Based on the SWOT analysis results, several changes were made to the models prior to the empirical studies.

Two empirical evaluation studies were conducted to test the theorized design principles derived from the revised models. The first study focused on assessing the utility of audio cues to train a temporal pacing task and the second study combined both temporal (i.e., pace) and spatial audio information, with a focus on examining integration issues. In the pace study, there were four different auditory conditions used for training pace: 1) a metronome, 2) non-spatial auditory earcons, 3) a spatialized auditory earcon, and 4) no audio cues for pace training. Sixty-eight people participated in the study. A pre- post between subjects experimental design was used, with eight training trials. The measure used for assessing pace performance was the average deviation from a predetermined desired pace. The results demonstrated that a metronome was not effective in training participants to maintain a desired pace, while, spatial and non-spatial earcons were effective strategies for pace training. Moreover, an examination of post-training performance as compared to pre-training suggested some transfer of learning. Design guidelines were extracted for integrating auditory cues for pace training tasks in virtual environments.

In the second empirical study, combined temporal (pacing) and spatial (location of entities within the environment) information were presented. There were three different spatialization conditions used: 1) high fidelity using subjective selection of a "best-fit" head related transfer function, 2) low fidelity using a generalized head-related transfer function, and 3) no spatialization. A pre- post between subjects experimental design was used, with eight training trials. The performance measures were average deviation from desired pace and time and accuracy to complete the task. The results of the second study demonstrated that temporal, non-spatial auditory cues were effective in influencing pace while other cues were present. On the other hand, spatialized auditory cues did not result in significantly faster task completion. Based on these results, a set of design guidelines was proposed that can be used to direct the integration of spatial and temporal auditory cues for supporting training tasks in virtual environments.

Taken together, the developed models and the associated guidelines provided a theoretical foundation from which to direct user-centered design of auditory interfaces.

ACKNOWLEDGMENTS

I would like to express my deepest thanks to my wonderful advisor, Dr. Kay M. Stanney, who sincerely and passionately worked with me to complete my doctoral research. Her comments and continuous review and feedback were instrumental to the successful completion of this dissertation.

I would like to show my appreciation to Dr. Christopher Geiger, Dr. Brian Goldiez, Dr. Charles Reilly, Dr. Barbara Shinn-Cunningham, and Dr. Arthur Tang for serving on my doctoral committee. Dr. Geiger's expertise in simulation and scheduling provided useful insights for ordering the presentation of audio cues. Dr. Goldiez's expertise in human-participant studies in virtual realities guarded me from falling into many of the traps associated with these studies. Dr. Reilly's diverse expertise in Industrial Engineering and Operation Research provided useful insights with respect to the logicality and progression of this dissertation. Dr. Shinn-Cunningham's extensive expertise in audio research was very instrumental to delivering quality outcomes of this work and ensuring its consistency with the current state-of-the-art in audio research. Dr. Tang's expertise in spatial visual displays provided insights on the similarities and differences between audio and vision when presenting space-related information.

I would like to express my appreciation to the Office of Naval Research (ONR) for sponsoring this work under their Small Business Technology Transfer Research (STTR) award to VRSonic, Inc. ONR provided a support umbrella that enabled the success for this work. In particular, I would like to thank Hesham Fouad and Paul Cummings from VRSonic for developing the audio environments and making their VibeStation software available for our research, Richard Schaffer and Pete Wierzbinaki from Lockheed Martin for developing the visual environments using their ManSim software, and Anna Cole from Naval Research

Laboratory (NRL) for developing the software for analyzing participant log files. NRL and Anna were very generous in loaning us computer equipment and a motion tracker to successfully run our studies. I would like to also thank the student team at the Human System's Integration lab at the University of Central Florida for their support to this work at its various stages of development, in particular, I would like to thank Roberto Champney, Brad Hill, Alex Katsaros, Christopher Lee, Diana Ovadia, Chris Reid, and Nina Schwartz.

I would like to express my appreciation to my family for their continuous support and confidence. My mother and father provided to me, my three brothers, and two sisters a fertile platform coupled with endless love and support that gave all of us the opportunity to prosper and become successful in all of our endeavors.

I would like to thank many of my friends and colleagues who showed interest in my work and reviewed portions of its development. Your encouragement and support instilled in me the confidence required to complete this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

CAVE	Cave Automatic Virtual Environment
CQB	Close Quarters Battle
HIP	Human Information Processing
HRTF	Head-Related Transfer Function
IID	Interaural Intensity Difference
ITD	Interaural Time Difference
MOUT	Military Operations in Urban Terrain
MRT	Multiple Resource Theory
NASA TLX	NASA Task Load Index
NRC	National Research Council
NRL	Naval Research Laboratory
ONR	Office of Naval Research
SME	Subject Matter Expert
SSQ	Simulator Sickness Questionnaire
SWOT	Strengths, Weaknesses, Opportunities, and Threats
VE	Virtual Environment
VIRTE	Virtual Technologies and Environments
WIMP	Windows, Icons, Menus, and Pointing devices

CHAPTER ONE: GENERAL INTRODUCTION

Humans live and interact in a sound-rich world. Sounds provide a multitude of information to humans about their environment, from ecological sounds characterizing the environment to speech, which is an intrinsic component of human-to-human communication. Despite its importance to real world interaction, audio cues have been underutilized in today's human computer interfaces, which mainly consist of visual constructs such as windows, icons, menus, and pointing devices (WIMP) system. Designing auditory interfaces is a challenge for current human-systems developers. This is largely due to a lack of theoretical guidance for directing how best to use sounds in today's visually-rich graphical user interfaces. This dissertation provides a framework for guiding the design of audio interfaces to enhance human-systems performance.

A review by the National Research Council (NRC) of Engineering Research forecasts that multimodal systems will soon have extreme graphics "with some spatial audio interfaces and haptic interfaces" and in the not too distant future, "spatial-audio effects, full-hand haptics, and olfactory displays will also be available" (National Research Council, 2000, p.25). Despite these optimistic projections, currently there is limited understanding with respect to the utility of using such multimodal technology, and there are few existing guidelines to aid in designing and implementing multimodal systems. There is a need to develop theoretical models and associated design guidelines to aid the design of auditory interfaces, which taken together can provide a structured process for integrating sounds in interactive interfaces. It is imperative to validate the theoretical design principles through empirical studies, and thus provide a foundation for a user-centered design process for auditory displays.

This dissertation uses the alternative three papers format. Chapter 1, this chapter, provides an overall introduction to this doctoral research. Chapter 2 is the first paper, which is currently under review by the *Theoretical Issues in Ergonomics* journal. Paper 1 presents the theoretical underpinnings of this work and illustrates the development of three models, including an audio integration model that outlines an end-to-end process for adding sounds to interactive interfaces, a temporal audio model that provides a framework for guiding the timing for integration of these sounds to meet human performance objectives, and a spatial audio model that provides a framework for adding spatialization cues to interface sounds. In addition, paper 1 includes the results of subject matter experts' evaluation of the developed models. Chapter 3 is the second paper, which is under review by the *Military Psychology* journal. Paper 2 presents the results of an empirical evaluation study that examines using audio cues to train a temporal pacing task. Chapter 4 is the third paper, which is under development. Paper 3 presents the results of a second empirical evaluation study, which examines using audio cues to train combined temporal and spatial tasks. Chapter 5 provides a general discussion based on the three papers combined. Chapter 6 concludes the dissertation and provides directions for future research.

Taken together, this body of research, from the theoretically-driven models to the validated design guidelines, establishes a foundation for user-centered auditory display design, thereby advancing the state-of-the-art in auditory science.

CHAPTER TWO: THEORETICAL FOUNDATIONS FOR INTEGRATING AUDIO IN INTERACTIVE INTERFACES*

This paper proposes theoretical foundations for conveying temporal (i.e., relating to time) and spatial (i.e., relating to space) information using auditory cues in interactive systems. Three theoretical models are developed to aid the design of auditory interfaces, including an audio integration model that outlines an end-to-end process for adding sounds to interactive interfaces, a temporal audio model that provides a framework for when to integrate these sounds to meet certain performance objectives, and a spatial audio model that provides a framework for adding spatialization cues to interface sounds. The models presented in this paper, which are each coupled with a set of design guidelines theorized from the literature, put forward a structured process for integrating sounds in interactive interfaces.

Introduction

Humans live and interact in a sound-rich world. Sounds provide a multitude of information to humans about their environment, from ecological sounds characterizing the environment to speech, which is an intrinsic component of human-to-human communication. Despite its importance to real world interaction, audio cues have been under utilized in today's human computer interfaces, which mainly consist of visual constructs such as windows, icons, menus, and pointing devices (WIMP) system. The WIMP paradigm has become a standard for interacting with computers. Nevertheless, it suffers from limitations in that it fails to adapt to

* Manuscript submitted to *Theoretical Issues in Ergonomics*.

individual user's capabilities and limitations and some users still find such systems difficult to master (Pew, 2003). In addition, these primarily visual interfaces may overload users with information and do not utilize human information processing (HIP) resources across multiple modalities (Wickens, 1984; 1992), thereby missing out on the advantages associated with the multiplicative effects of multi-sensory processing (Rowe, 1999). Such parallel processing, as discussed in Wickens' (1984; 1992) multiple resource theory (MRT), has been demonstrated to result in improvements in human-computer performance. Thus, it is important to consider how to effectively integrate sound into human-system interactions.

A recent review by the National Research Council of Engineering Research forecasts that multimodal systems will soon have extreme graphics "with some spatial audio interfaces and haptic interfaces" and in the not too distant future, "spatial-audio effects, full-hand haptics, and olfactory displays will also be available" (National Research Council, 2000, p.25). Despite these optimistic projections, currently there is limited understanding with respect to the utility of using such multimodal technology, and there are few existing guidelines to aid in designing and implementing multimodal systems. In terms of the utility of spatial audio, such cues are anticipated to improve performance in high stress applications, such as air craft cockpits and advanced command and control operations centers, as they are suggested to increase situational awareness (Begault, 2000). In addition, spatial audio is suggested to contribute to the sense of immersivity in virtual environments (Begault, 1994). Current advances in spatial sound technology make it possible to consider further the benefits of leveraging audio to enhance human-system performance. The current research investigates using audio to enhance temporal (i.e., relating to time-varying characteristics) and spatial (i.e., relating to location in space) information conveyance. Through the models and design guidelines presented in this paper, this

work aims to establish a foundation of user-centered auditory display design and provide a basis for user-centered evaluation of auditory displays, thereby advancing the state-of-the-art in auditory display design.

Audio Integration Model

The current work presents an audio integration model to address the end-to-end decision making process for integrating auditory cues in interactive applications (see Figure 1). This model has four steps, identifying the performance objective(s) to be met by the integration of audio cues, selection of the audio cues to be presented, and identification of the temporal and then the spatial parameters for this presentation.

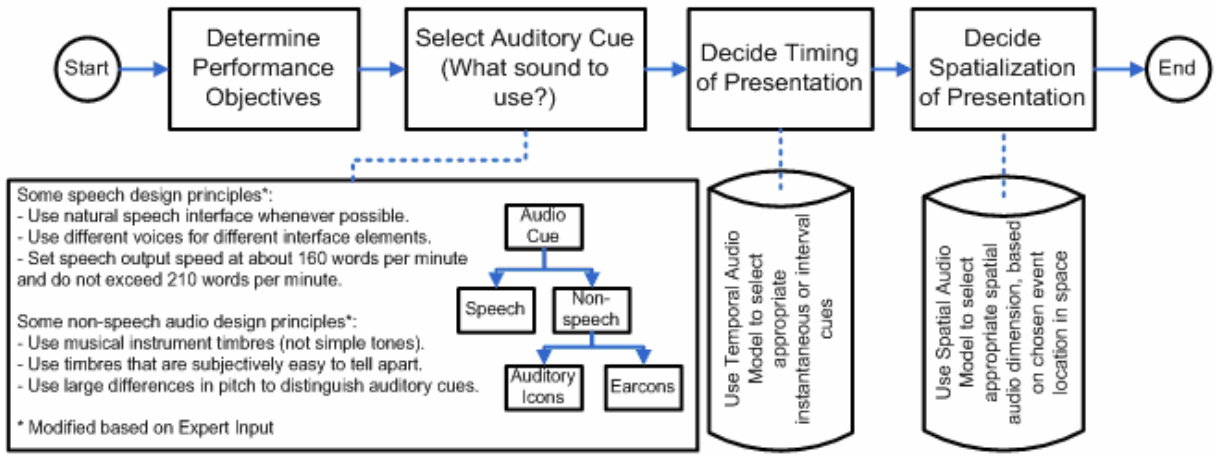

Figure 1: Audio Integration Theoretical Model

When designing audio interfaces, the audio designer first needs to decide upon a performance objective(s) - why sounds need to be added, i.e., to enhance which aspects of performance? Sounds in interactive applications can be used to influence psycho-motor (physical), affective (emotional), and cognitive (intellectual) aspects of human performance

(Bloom, 1956). Psycho-motor goals include physical movement, coordination, and use of motor-skill areas. These skills require practice and are measured in terms of speed, precision, distance, procedures, or techniques in execution. Audio has been shown to enhance psycho-motor performance by timing start and finish of goal-directed movements, marking cycle ends for rhythmic movements, and driving coordination in rhythm-modulated movements (Thaut, 2005). Affective goals deal with effect on human emotions and include feelings, values, appreciation, enthusiasms, motivations, and attitudes (Bloom, 1956). Audio has been shown to influence human emotion, for example, fast tempo, large pitch variation, sharp envelope, few harmonics, and moderate amplitude variation can drive happiness, while slow tempo, low pitch level, few harmonics, round envelope, and pitch contour down can drive sadness (Fahlenbrach, 2002). Cognitive goals deal with knowledge and include recall and recognition of facts, storage of procedural patterns, and intellectual abilities and skills (Bloom, 1956). Audio has been shown to enhance cognitive performance through reducing the amount of needed attentional resources by drawing attention via audio cues and decreasing user's cognitive workload by distributing processing across multiple sensory systems (Brewster, 1997; Brown & Boltz, 2002).

Once the target performance objective(s) is determined, the audio designer needs to select appropriate sounds for inclusion in the interface. Interface sounds can be speech or non-speech (Kramer, 1994). Non-speech sounds include auditory icons (if the sound has semantic mapping to an interface element) and earcons (Blattner et al., 1989; Brewster, 1994; 1997; Gaver, 1986). The selection of interface sounds is not a trivial task, and poor selection of sounds can severely hinder the efficacy of the interactive interface regardless of the timing and spatializing of the sound. Some general design principles for selecting speech and non-speech sounds for interactive interfaces include:

- Use natural speech interface whenever possible (Tsimhoni et al., 2001).
- Use different voices for different interface elements (ETSI, 2001).
- Set speech output speed at about 160 words per minute and do not exceed 210 words per minute (ETSI, 2001).
- Use musical instrument timbres- not simple tones (Brewster, 1994).
- Use timbres that are subjectively easy to tell apart (Brewster, 1994).
- Use large differences in pitch to distinguish auditory cues (Brewster, 1994).

Several authors have provided detailed guidelines for selecting interface sounds, which will not be duplicated in the current work, as the focus herein is on the overall process of sound integration (c.f., Barrass, 1997; Blattner et al., 1989; Brewster, 1994; 1997; 1998; 2003; ETSI, 2002; Gaver, 1986; Kramer, 1994; McGookin & Brewster, 2004; Patterson & Mayfield, 1990; Walker & Kramer, 2005).

After selecting the interface sounds to use, decisions regarding the timing of presentation and whether the sounds need to be spatialized are made, which are discussed in the following sections.

Temporal Audio Theoretical Model

In general, temporal information is used to describe events that take place at a specific instant in time (i.e., instant-based) or over a time interval (i.e., interval-based - before and after, overlaps and overlapped-by, starts and started-by, finishes and finished-by, during and contains, meets and met-by, and equal; Allen, 1984; Schreiber, 1994), as well as ordering and constraints between such events (Allen & Ferguson, 1997; Vila, 1994). Such events can be grouped based

on how predictable they are (i.e., triggered, definite, spontaneous; Allen & Ferguson, 1994).

Audio is known to be superior as compared to visual when processing such temporal information (ETSI, 2002; Kramer, 1994). Thus, events are specific occurrences in time that can be either instantaneous (i.e., instant-based) or they can span a time interval (i.e., interval-based) and often involve both (Schreiber, 1994). For example, a progress bar is a visual construct that is commonly used to convey status information about a download task. The download task has two main instantaneous events; start and finish. Each increment on the progress bar defines an instance, including start and finish; however, the download time reflects an interval-based measure. Conveying temporal information is dependent upon whether instant-based or interval-based temporal information is needed. Instant-based temporal information focuses on conveying information related to a particular point of time or a particular event. On the other hand, interval-based temporal information focuses on conveying information related to durations, rhythms, rates, and changes over time. Interval-based systems facilitate comparisons between different durations (Allen, 1983). In a very general sense, the following holds true (Allen, 1983; 1984; Allen & Ferguson, 1994; 1997; Schreiber, 1994; Vila, 1994):

- Instant-based temporal information is used to specify a point in time such as an alarm or warning.
- Instant-based temporal information is used to specify the start or finish of an activity.
- Interval-based temporal information is used to indicate status and progress.
- Interval-based temporal information is used to perform comparisons.

The temporal audio theoretical model shown in Figure 2 illustrates that audio can be used to covey instant-based and interval-based temporal information. For example, an audio format that is often used to guide rhythmic movements is a metronome, which marks exact instant-based

time increments by a regularly repeated tick (Kurtz & Lee, 2003). On the other hand, an example of interval-based audio is using sounds with varying tempos that indicate relative distance to objects (Day et al., 2004); i.e., shorter time delays between sounds as a user gets closer to target or destination. A rhythmic sound is a special case of interval-based audio that repeats at consistent time-increments within the interval (Thaut, 2005). This classification can be of importance to the audio designer as rhythmic audio can be used to train pace-controlled psychomotor actions, such as those associated with medical procedures, dancing, and sports (Boyle et al., 2002; Interactive Metronome, 2005; Kaplan, 2002; Kern et al., 1992; Libkuman et al., 2002; Wijnalda et al., 2005). When an external sensory stimulus is used to guide such rhythmic movement, audio cues generally result in the least variability from target rhythm as compared to visual or tactile cues (Chen et al., 2002; Kolers & Brewster, 1985).

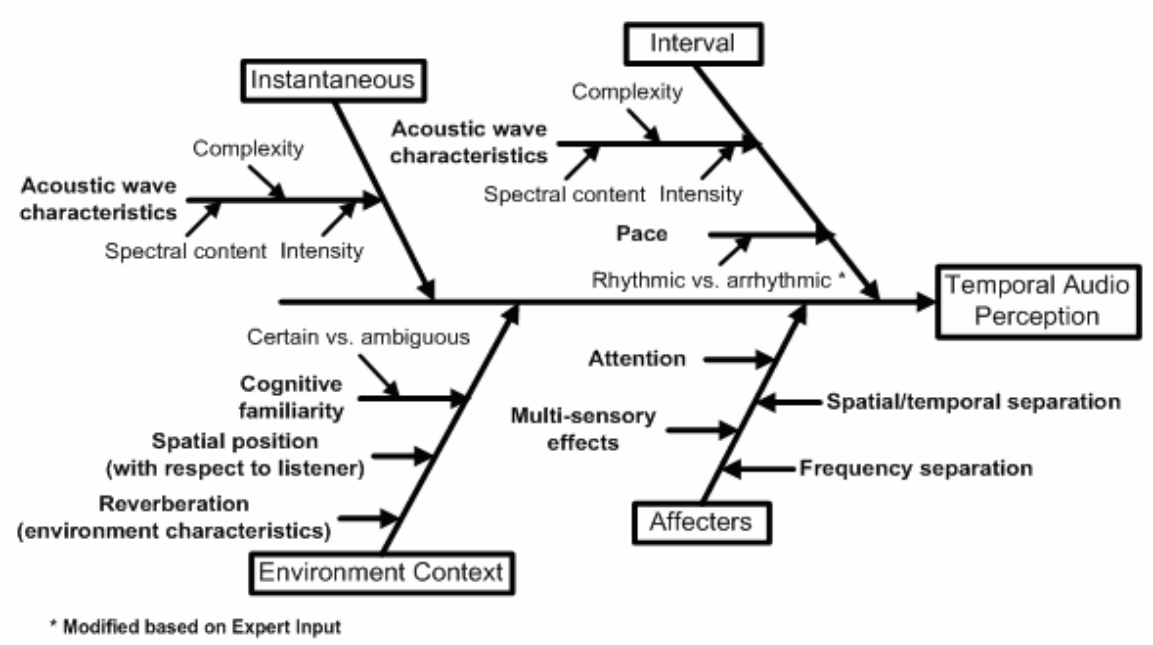

Figure 2: Temporal Audio Theoretical Model

As indicated in Figure 2, the main characteristic of an instantaneous sound is choosing the acoustic wave attributes that enable a user to hear the sound as intended by the audio designer at the right time (Plomp, 2002; Zwicker & Fastl, 1999). The acoustic wave attributes that enable listeners to hear and differentiate sounds are spectral content, intensity, and complexity (see Figure 2). The human ear works as a frequency analyzer that processes the spectral content of an acoustic wave, thus extracting useful temporal and spectral information from various frequency bands that create human perception of sound intensity and pitch (Plomp, 2002). The intensity of an acoustic wave is its amplitude or pressure level, it is commonly measured relative to a reference level, in decibels (dB) (Zwicker & Fastl, 1999). For example, a quiet forest will be about 15 dB, normal human conversations are about 70 dB, and a noisy environment is about 110 dB. The audio designer selects (or develops) sounds with appropriate intensity, spectral content and complexity to implement a particular instant-based sound (see Figure 2). In order to enable hearing the instantaneous sound, tradeoffs exist between overcoming ambient noises that might be present and producing sounds that are annoying or harmful to the listener (Walker & Kramer, 2004). In addition to the acoustic wave attributes, an interval-based sound involves the pace at which the acoustic wave repeats within a particular time interval (see Figure 2). If the acoustic wave repeats at consistent time-increments within the interval, then the sound is rhythmic, otherwise it is arrhythmic (Thaut, 2005). As aforementioned, rhythms are an important consideration for motor task performance (Karageorghis & Terry, 1997). They assume an important role in organizing music events into coherent and comprehensible events and forms. The audio designer can select acoustic wave attributes and pace for an interval-based sound to achieve different psychomotor, affective, or cognitive performance objectives.

Several environmental factors and affecters can change how a user perceives the timing of a particular sound (see Figure 2). The environment context factors include listener's familiarity with the sound (i.e., is the sound certain, known to the listener, or is it ambiguous?), spatial position of the sound source with respect to the listener, and other characteristics of the listening environment such as reverberation (Walker & Kramer, 2004).

Factors that can affect how users perceive a particular sound and may influence their judgments regarding multiple concurrent sound streams include spatial and temporal separation between sounds, frequency separation, and the presence of other concurrent sensory stimuli, such as visual (see Figure 2; Bregman, 1990). For example, a listener will tend to associate a sound event with a concurrent visual event, which is known as the ventriloquist effect (Alais & Burr, 2004). Listeners' attention can also drive how a user perceives the timing of a particular sound (Jones, 2004). Assessing the user's environment and tasks with respect to familiarity with sound, spatial position of a sound source within the application, reverberation, and attention is important when the audio designer is implementing instant and interval-based sounds for an interactive application. For example, when integrating two sounds that are instant-based, the audio designer will need to ensure that the user will not mistake them to be a single interval-based sound and that they will not be masked by other sensory stimuli (e.g., visual dominance), and to do so, selecting spatial and/or temporal separation of the sounds can drive differentiability. For such integration, general guiding principles include (Bregman, 1990):

- Use sounds with different frequency ranges to drive differentiation between multiple sound streams.
- Use sounds with different spatial locations to drive differentiation between multiple sound streams.

The conceptual framework presented in Figure 2 identifies the general factors that must be considered when designing instant and interval-based temporal audio cues in an interactive application. Table 1 presents design principles that can be used to guide this design process while meeting certain psychomotor, affective, and cognitive performance objectives.

Table 1: Temporal Information Conveyance Design Principles

Performance Objective	Audio Type	Cue	Design Principles
Psychomotor	Instant	Certain	• Use sounds to time start/finish of goal-directed movements such as throwing a ball (Thaut, 2005). • Use sounds to mark start/end points for rhythmic movements such controlling finger tapping (Thaut, 2005).
	Interval	Rhythmic	• Use rhythmic music to synchronize physical activity in pace setting, and matching tasks (Wijnalda et al., 2005). • Use rhythmic audio to influencerhythmic physical movement (Kolers & Brewster, 1985; Repp, 2006).
		Ar-rhythmic	• Use arrhythmic audio to enhance physical activities such as workouts (Karageorghis & Terry, 1997). • Use arrhythmic audio in rhythm-modulated movements to increase or decrease user's pace* (Thaut et al., 2004).
Affective	Interval	Ar-rhythmic	• Use upbeat music to lessen anger, and depression rather than negative music (slow) (Karageorghis & Terry, 1997). • Use to drive happiness or sadness, by controlling tempo, pitch, and harmonics (Fahlenbrach, 2002).
Cognitive	Instant	Certain	• Use sound to enhance memory performance in terms of object identification (Davis et al., 1999). • Use earcons to reduce user's mental workload and reduce error recovery time (Brewster, 1997). • Use sound to capture attention (Frauenberger et al., 2005).
	Interval	Rhythmic	• Use coherent sounds to convey information that requires predictability to reduce the amount of attentional resources needed* (Brown & Boltz, 2002). • Use rhythmic audio as a temporal ordering mechanism to facilitate remembering (Thaut, 2005).
		Ar-rhythmic	• Use arrhythmic audio with varying tempos to provide feedback regarding position or distance* (Day et al., 2004). • Use arrhythmic audio to convey cause of dysfunction or for urgent situations (alarm design)* (Guillaume et al., 2002).

*: Modified based on Expert Input.

Psychomotor Performance Design Guidelines

There are three types of temporally-related psychomotor performance objectives; these are goal-directed, rhythmic, and rhythm-modulated movements (Thaut, 2005). Goal-directed movements have a specific target, such as throwing a ball a certain distance in minimum time, or swinging a golf club. Rhythmic movements repeat at a constant rate such as tapping fingers or drawing circles. Rhythm-modulated movements repeat over time but at either increasing or decreasing rates, such as speeding up during exercise. Sounds can be used to influence all three movement types. For goal directed movements, the audio designer can use instant-based sounds to time start and finish of a movement (Thaut, 2005). Rhythmic movements can be controlled using both instant and interval-based sounds. The audio designer can use instant-based sounds to mark cycle ends for rhythmic movements (Thaut, 2005). Also, the audio designer can use interval-based sounds as a guide for rhythmic movements by providing sounds that repeat at consistent pace. People generally move in synch (Repp & Penel, 2004) and with less variability (Kolers & Brewster, 1985) with an auditory-modulated rhythm as compared to a visually-modulated rhythm. Once people synchronize their movements with that of auditory tones, they generally can maintain the pattern without the audio being played (Kolers & Brewster, 1985).

Rhythm-modulated movements can be controlled using interval-based sounds. In general, rhythm-modulation with audio cues involves using sounds that repeat at increasing or decreasing pace as a guide for physical movements. These movements are common in rehabilitation studies that involve gait, where typically a music tempo is initially chosen that accommodates an individual's gait capabilities and then the tempo is increased incrementally as gait performance improves (Thaut et al., 2004). Repp (2006) provided an example of rhythm

modulation, when he noted that people who initially tapped their fingers at their own pace were influenced when exposed to rhythmic audio, and resynchronized their pace to that of the audio when the tempo difference was less than 10%.

Music is an audio format that is commonly used when performing psychomotor activities during exercise, which can involve both rhythmic and rhythm-modulated movements. Music is found to enhance physical activities in several ways, first, music can divert performer's attention away from physical stress and fatigue, second, music can enhance psychomotor arousal by acting as a stimulant before exercise and as a sedative during exercise, and third, music can enable performers to synchronize their physical rhythm to that of musical rhythm (Karageorghis & Terry, 1997). Music can be used to support user performance in three modes; these are pace-fixing (i.e., playing music at a constant tempo to enable synchronization), pace-matching (i.e., playing music at a tempo that matches user's pace), and pace-influencing (i.e., playing music at varying tempo to influence a user to slow-up or slow-down) (Wijnalda et al., 2005).

In summary, audio can be used to influence the temporal aspects of psychomotor tasks in the following ways:

- Instant-based sounds can be used to time start and finish of goal-directed movements (Thaut, 2005).
- Instant-based sounds can be used to time cycle ends of rhythmic movements (Kolers & Brewster, 1985; Thaut, 2005).
- Interval-based sounds can be used to drive synchronization of physical movements (Thaut, 2005).
- Interval-based sounds can be used to influence rhythm-modulated movements to increase or decrease a user's pace (Thaut et al., 2004; Wijnalda et al., 2005).

- Music can be used to synchronize physical rhythm to that of musical rhythm (Karageorghis & Terry, 1997).
- Music can be used to influence user's pace either to speed up or to slow down (Wijnalda et al., 2005).

Affective Performance Design Guidelines

Affective performance objectives may include influencing a listener's happiness, sadness, or fear, among other emotions (Fahlenbrach, 2002; Karageorghis & Terry, 1997). In order to drive such affective emotional experiences, the audio designer can vary the intensity, rhythm (tempo), and form of an interval-based sound (Fahlenbrach, 2002). For example, fast tempos and high pitches tend to evoke positive pleasant emotions, whereas slower tempos with lower pitches evoke more negative somber emotions. Listening to upbeat music generally results in positive moods and listening to slower music generally results in negative moods (Karageorghis & Terry, 1997). A high intensity instant-based sound can scare a near-by person, and theatre interval-based sounds drive suspension and excitement (Begault, 2000). It is important to note that sound's affect on human emotions depends on socio-cultural codes and taste influences (Fahlenbrach, 2002). Audio can be used to influence affective tasks in the following ways (Fahlenbrach, 2002):

- Use fast tempo, large pitch variation, sharp envelope, few harmonics, and moderate amplitude variation to drive happiness.
- Use slow tempo, low pitch level, few harmonics, round envelope, and pitch contour down to drive sadness.

- Use many harmonics, fast tempo, high pitch level, round envelope, and pitch contour up to drive potency.

Cognitive Performance Design Guidelines

Cognitive performance objectives may include capturing user's attention (such as alerting a user to system malfunctions), decreasing user's workload, enhancing information exchange between user and system, and providing feedback to the user (Brewster, 1997; Day et al., 2004; Frauenberger et al., 2005; Guillaume et al., 2002).

Auditory alarms can be instant- or interval-based. Instant-based alarms present information regarding the nature of triggered events, whereas interval-based alarms are used to present information regarding the nature and urgency of triggered events (Guillaume et al., 2002), for example, high urgency can be expressed by fast pace, variable high pitch, irregular harmonics, and fast onset ramp, while low urgency can be expressed by slow pace, descending pitch, regular harmonics, and slow onset ramp. Listeners tend to perceive sounds with faster vibrato and low frequency filtering as more important (Hakkila & Rankainen, 2003).

Integrating instant-based sounds into graphical user interfaces can decrease user's workload (Brewster, 1997). Both synthesized speech messages (communicating numerical values and words) and rhythmical musical tones can be recognized successfully (Rigas et al., 2001). Brown and Boltz (2002) indicate that coherent interval-based sounds exhibit a high degree of internal predictability, which is common in conversational speech and western music, and can reduce the amount of needed attentional resources. Rhythms with their internal cyclic periodic nature create anticipation and predictability (Thaut, 2005). The temporal ordering

resulting from rhythm organizes time and hence using rhythmic interval-based sounds can make remembering easier since events may be patterned over time.

When used in virtual environments, sounds can enhance the sense of presence (or "being there", as reported by users), by providing "natural" ambient environment cues, and the recall and recognition of visual objects and their spatial locations, by providing feedback cues that can be utilized by users to remember environment characteristics (Davis et al., 1999; Dinh et al., 1999). Increased usage of instant and interval-based sounds may also allow for increased system-to-user information transfer, device/system miniaturization, increased user mobility, increased accessibility for people with disabilities, and enhanced navigation (Frauenberger et al., 2005). For example, when audio is used as a navigation aid, varying tempos can suggest relative distance to objects (Day et al., 2004); i.e., shorter time delays between interval-based sounds can be implemented, as a user gets closer to a target or destination. This can be useful in driver assistance systems to aid in lane keeping, blind spot monitoring, and collision avoidance (Day et al., 2004).

Audio can be used to influence the cognitive aspects of tasks in the following ways:

- Instant- and interval-based sounds can be used to provide feedback (Guillaume et al., 2002).
- Instant- and interval-based sounds can be used as an alarm to alert the user to system malfunctions (Hakkila & Rankainen, 2003).
- Interval-based sounds can be used to reduce user's workload by dividing needed processing across vision and audition (Brewster, 1997).
- Interval-based sounds can be used to reduce load on user's memory by providing structured ordering of time and enhanced predictability (Thaut, 2005).

- Instant- and interval-based sounds can be used to enhance exchange of information between system and user by utilizing more communication channels (e.g., vision and audition; Frauenberger et al., 2005).

Spatial Audio Theoretical Model

Once the decisions on what sounds to use and when to use them in a particular interactive application have been made (see Figure 1), the audio designer might opt to augment some of the sounds with additional cues that enable listeners to hear the sounds coming from a particular point in space. A conceptual distance-based model inspired by Cutting & Vishton's (1995) spaces model is herein proposed to describe treatment of spatialization cues at varying distances from a listener. Cutting & Vishton focused on visual depth perception and presented a framework to segment the space around an observer into three circular, egocentric regions. These regions are personal space, action space, and vista space. Personal space is the area immediately surrounding an observer; it extends to slightly beyond the arm's reach. Action space is within the individual's accessible area of action; the observer can interact with various objects within this space and it allows for quick action. Vista space is beyond 30 m, where only monocular and static information are available. In the current work, this conceptualization is extended to sound perception and the space around a listener. Specifically, for spatial sound perception, three regions are defined, these are personal space (i.e., inside a listener's head; ~<10 cm), proximal (i.e., nearby) space (~10 cm- 1 m), and distal space (beyond 1 m). All distances are defined from the center of a listener's head. The personal space dimensions are defined based on the size of an average human head (about 16 cm in diameter). In proximal space, the

spherical nature of sound waves reaching the listener makes interaural intensity difference (IID) cues and monaural spectral cues dependent on the source's distance from the listener. In distal space, the sound waves reaching a listener are planar in nature, and are less sensitive to the effects of head size and pinnae structure. Figure 3 shows the proposed segmentation for auditory spaces.

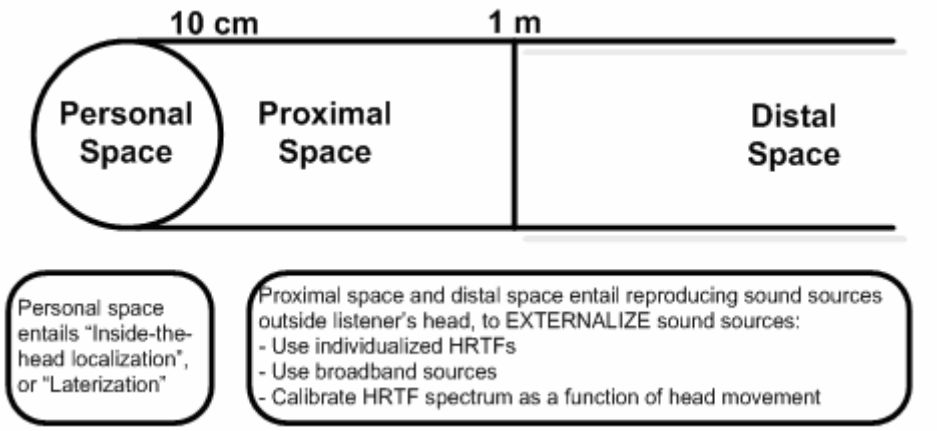

Figure 3: Proposed Spaces' Definition

In addition to distance from the listener (see Figure 3), there are several factors that influence sound perception: sound location, environmental context, and affecters (see Figure 4). Humans perceive sound location in three dimensions; azimuth, elevation, and distance (see Figure 4). The most important cues for localizing a sound source's angular position (azimuth) are interaural time and intensity differences (ITD and IID). Interaural cues are based on the relative differences between wave fronts at the two ears on the horizontal plane (Blauert, 1983). Due to the nature of IID and ITD, cones of confusion are created; these refer to points anywhere on a conical surface extending out from the ear (Duda, 1997). When perceiving azimuth, additional angular position localization cues take place due to head and source movement (Begault, 1994). The head and source movements result in dynamic spectral modifications to the

acoustic signals reaching a listener's ears, which improve localization ability and reduce front-back ambiguity (Wightman & Kistler, 1999).

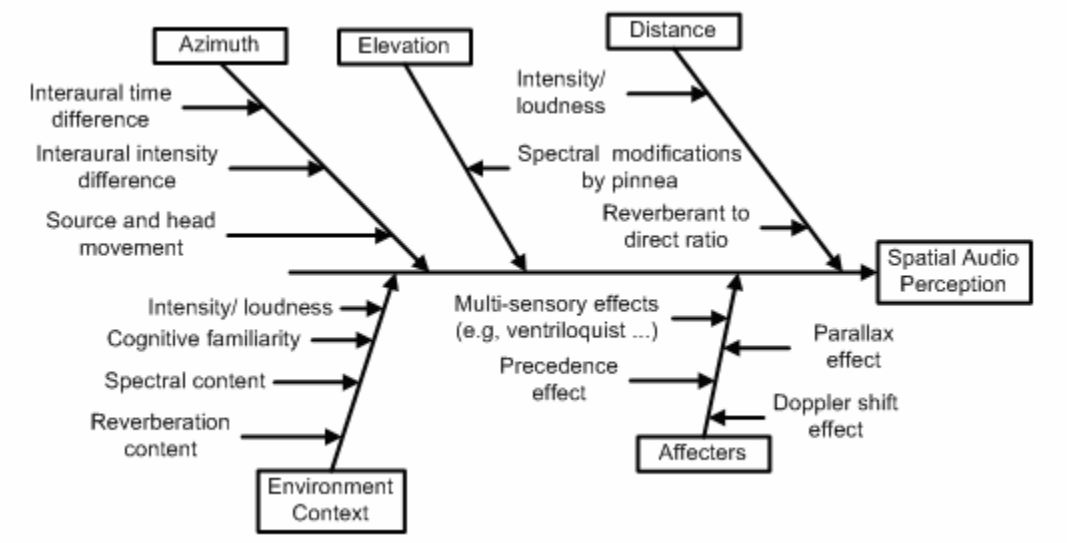

Figure 4: Spatial Audio Perception

With regard to perceiving the elevation of a sound source (see Figure 4), the human pinnae provide spectral modifications to acoustic signals that aid in elevation judgment with respect to the median plane (Hebrank & Wright, 1974). The spectral modifications resulting from pinnae folds produce a unique set of micro-time delays, resonances, and diffractions that translate into a unique descriptor for each sound source position in the median plane (Begault, 2000).

With regard to perceiving the distance of a sound source (see Figure 4), the intensity of a sound source is the most prominent distance cue in anechoic environments (or with familiar sounds) (McGregor et al., 1985; Middlebrooks & Green, 1991). The intensity of sound is inversely proportional to the squared distance from the sound source (Begault, 2000). In reverberant environments, the ratio of reflected to direct sound plays an important role for

distance perception (Blauert, 1983). This ratio of reflected to direct sound creates perceptual differences in the sound quality that depends on source distance (Middlebrooks & Green, 1991).

Perceiving the environmental context of a sound requires integrating loudness, cognitive familiarity cues, spectral content, and reverberation content (see Figure 4) (Begault, 1994; 2000). The loudness of a sound source perceived by a listener provides information regarding the reverberation characteristics of the environment, as well as the location of the various sound sources with respect to the listener (Middlebrooks & Green, 1991). The presence of different familiar sounds enables the listeners to judge their environment (Begault, 1994). The spectral content of sound reaching a listener is dependent on the environment spatial layout, absorption characteristics, and the presence of various obstacles within the environment, which result in different reflection and diffraction patterns (Blauert, 1983). The reverberant content of sound reaching a listener characterizes the spatial dimensions on the listening environment and its reflective characteristics (Blauert, 1983; Middlebrooks & Green, 1991).

Several affecters (see Figure 4) are known to influence human perception of a sound source; these include multi-sensory, precedence, parallax, and Doppler-shift effects (Begault, 1994; 1999). The presence of concurrent multi-sensory visual or haptic stimuli can affect listener's judgment regarding sound source location (Driver & Spence, 2000). For example, a ventriloquist effect explains the correlation of apparent location of an auditory event with a concurrently occurring visual event (Alais & Burr, 2004). The precedence effect describes the fact that humans tend to localize a sound source using the first information available to them. If two sounds are played at the same time (or within 15 ms), humans will tend to assume a single location depending on which signal got to them first (Wallach et al., 1949). This effect explains the localization of sounds in reverberant environments. The auditory parallax effect describes

the differences in interaural cues that take place depending on sound distance and may lead to better distance judgment accuracy for sources to the side compared to sources straight ahead (Holt & Thurlow, 1969). The Doppler-shift effect explains changes in pitch as a moving sound source passes by a listener (Neuhoff & McBeath, 1996). When a sound source is moving towards the listener, the sound waves propagate in the same direction as the source, but as soon as it crosses over the listener's position, the propagation becomes opposite to the source direction, which results in a sudden drop in the sound pitch perceived by the listener.

Once integrated, the proposed spaces' definition and spatial sound perception variables provide a conceptual framework to present design principles for integrating spatialization to sound sources (see Figure 5 and Table 2).

Personal Space Design Guidelines

When designing audio cues to be used without externalization (i.e., inside the head; in personal space), only laterality needs to be considered since sound sources will be perceived as falling on the axis connecting the listener's two ears (Jeffress & Taylor, 1961). To enhance laterality, the following design considerations for personal space should be considered:

- Manipulating the interaural cues (time and intensity differences) moves sound on the intracranial axis connecting the listener's ears (Blauert, 1983).
- Frequencies between 1500 Hz and 5000 Hz should be avoided to enhance interaural cues (Mills, 1972).
- Echoes and reverberation effects should be excluded to enhance interaural cues (Shinn-Cunningham, 2001).

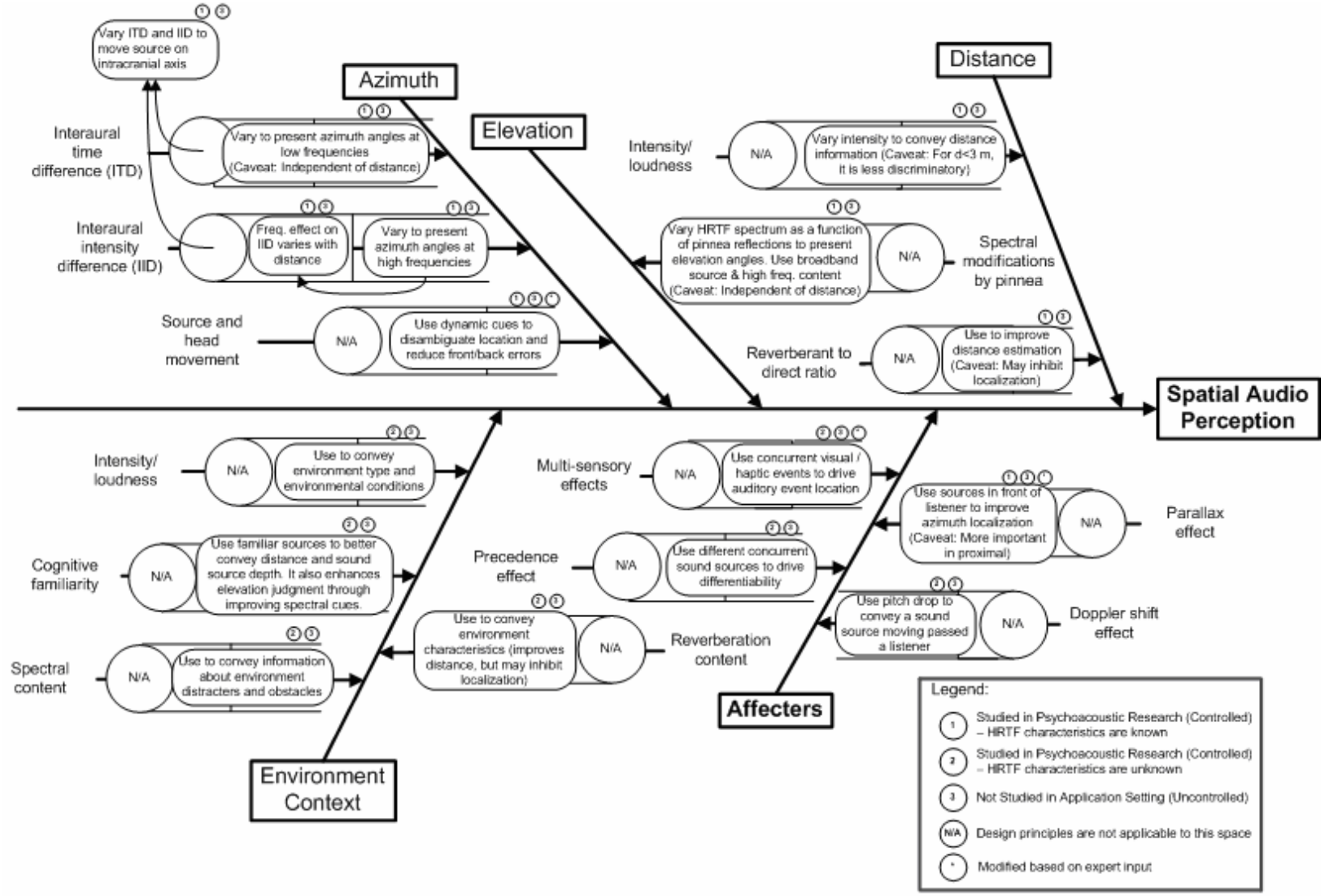

Figure 5: Spatial Audio Theoretical Model

Table 2: Spatial Information Conveyance Design Principles

Effect	Audio Cue	Space	Design Principles/Notes
Azimuth	Interaural time difference (ITD) [1, 3] → Dominate at low frequency	Personal	Use ITD to move source on intracranial axis.
		Proximal	Use ITD to present source azimuth location at low frequencies (independent of distance).
		Distal	
	Interaural intensity difference (IID) [1, 3] → Dominate at high frequency as wavelength of sound signal is small compared to head dimensions.	Personal	Use IID to move source image on intracranial axis.
		Proximal	Use IID to present source azimuth location at high frequencies. The frequency effect increases as sound distance goes from 1m-10 cm. Note: ILD increases dramatically as a lateral sound source approaches listener's head.
		Distal	Use IID to present source azimuth location at high frequencies.
	Source and head movement [1, 3]	Proximal	Use dynamic cues resulting from source and head movement to disambiguate location and reduce front/back errors. Note: Percent of front/back errors is greater for distances < 50cm. Note: Small changes in position don't produce perceptible changes in the HRTF*.
		Distal	Use dynamic cues resulting from source and head movement to disambiguate location and reduce front/back errors. Note: Small changes in position don't produce perceptible changes in the HRTF*.
Elevation	Spectral modifications by pinnae [1, 3]	Proximal	Use broadband sources and high frequency content and vary audio presentation as a function of pinnae reflection to present elevation angels. Note: High frequency content varies with elevation, not with distance in near field.
		Distal	Use broadband sources and high frequency content and vary audio presentation as a function of pinnae reflection to present elevation angels. Note: Independent of distance.
Distance	Intensity/ loudness [1, 3] → Can be lost in noisy operational environments.	Proximal	Use intensity/loudness cues to convey distance information. Note: HRTFs change substantially with distance for sources between 10cm and 1m. Note: For distances less than 1m, low frequency content dominates distance perception. Note: For distances less than 1m, humans use ITD and IID to determine distance. Note: Distance judgment is better for lateral sources.
		Distal	Use intensity/loudness cues to convey distance information. Note: Fore distances < 3m, these cues are less discriminatory. Note: For distances > 1m, HRTFs vary with location (not distance).
	Reverberant-to-direct ratio [1, 3] → Can be lost in noisy operational environments.	Proximal	Use reverberant-to-direct ratio to improve distance estimation. Note: May inhibit localization and impede speech intelligibility. Note: The reverberant to direct ratio increases as the source distance decreases from 1m-10cm.
		Distal	Use reverberant-to-direct ratio to improve distance estimation. Note: May inhibit localization and impede speech intelligibility. Note: Independent of distance.

Effect	Audio Cue	Space	Design Principles/Notes
Environment Context	Intensity/loudness [2, 3]	Proximal Distal	Use to convey environment characteristics (e.g., absorption, composition, and presence of obstacles) and environmental conditions (e.g., temperature, and distance). Note: The intensity of a sound reaching a listener depends on distance, and environment type, characteristics, and presence of obstacles.
	Cognitive familiarity [2, 3]	Proximal Distal	Use familiar sound sources to better convey sound source distance and depth information. Note: Enhances elevation judgment by improving spectral cues.
	Spectral content [2, 3]	Proximal Distal	Use to convey information about environment distracters and obstacles, which result in modifying the spectral content reaching the listener.
	Reverberation content [2, 3]	Proximal Distal	Use to convey information about environment characteristics (e.g., absorption, composition, and presence of obstacles). Note: May inhibit localization.
Other	Multi-sensory effects [2, 3]	Proximal Distal	Use concurrent visual/haptic events to drive auditory event location judgment*.
	Precedence effect [2, 3]	Proximal Distal	Use different concurrent sound sources (in terms of frequency, or complexity) to drive differentiability between competing auditory streams.
	Parallax effect [1, 3]	Proximal	Use sound sources in front of the listener to improve azimuth localization. Note: Azimuth localization accuracy degrades for sources that are on lateral location with respect to listener, this effect is more apparent for sources near head because of head shadow effect*.
		Distal	Use sound sources in front of the listener to improve azimuth localization. Note: Azimuth localization accuracy degrades for sources that are on lateral location with respect to listener.
	Doppler shift effect [2, 3]	Proximal Distal	Use pitch drop to convey a sound source moving by the listener.

1: Studied in Psychoacoustic Research (Controlled) – HRTF characteristics are known.
2: Studied in Psychoacoustic Research (Controlled) – HRTF characteristics are unknown.
3: Not studied in Application Setting (Uncontrolled).
Space: Personal: <10 cm; Proximal: (10cm – 1m); and Distal: (> 1m)
*: Modified based on Expert Input.

Proximal and Distal Spaces Design Guidelines

When designing sounds to be used with externalization (i.e., outside the head; in proximal and distal spaces), using individualized head-related transfer functions (HRTFs), broadband sound signals, and calibrating the HRTF as a function of head movement enhance spatial sound perception (Begault, 1994). Audio designers can use interaural time and intensity

difference cues to present sound azimuth information in both proximal and distal spaces (see Figure 5) (Blauert, 1983). It is recommended that the audio designer uses time difference cues when working with low frequencies and intensity difference cues when working with high frequencies (Blauert, 1983; Middlebrooks & Green, 1991; Mills, 1972). It is important to note that, in proximal space, the frequency effect increases as the sound source decreases from 1 m to 10 cm, and that the intensity differences increase dramatically as a lateral sound source approaches the listener's head (Brungart & Robinowitz, 1999; Brungart et al., 1999). In order to disambiguate sound source location, and reduce front-back errors, the audio designer can utilize the dynamic cues resulting from source and head movement (Brungart, 1999a). To generate perceptible changes in the HRTF, gross source and head movements might be required (Wenzel, 1992; Wenzel et al., 1993). The percent of front-back errors is greater for sources that are located less than 50 cm from the listener (Brungart et al., 1999). Audio designers can modify the HRTF spectral content to mimic pinnae reflections to present sound elevation information in both proximal and distal spaces (Brungart & Robinowitz, 1999). The pinnae reflections are independent of distance. In order to present information pertaining to sound source distance, the audio designer can use intensity/loudness and reverberant-to-direct cues in both proximal and distal spaces (Blauert, 1983; Middlebrooks & Green, 1991). It is important to note that, in personal space, just noticeable differences (JND) in distance (defined in terms of percent of distance) decrease with decreasing distance (i.e., as go from 1 m down to 10 cm). Also, the reverberant to direct ratio increases as the source distance decreases from 1 m to 10 cm (Brungart & Robinowitz, 1999). For distances less than 1 m, the low frequency content of a sound signal dominates distance perception, and humans can use interaural time differences to judge distance from the sound source (Brungart, 1999a; 2001). Further, in this space, distance estimation

accuracy is best for sources to the side of the listener and worst for sources anywhere in the median plane (Brungart, 1999b; Brungart & Rabinowitz, 1999). In the distal space (i.e., beyond 1 m), the HRTFs vary only with location, not with distance (Brungart & Rabinowitz, 1999). For distances less than 3 m, intensity/loudness cues are less discriminatory, and hence distance perception is best for sources beyond 3 m (Strybel & Perrott, 1984). Tradeoffs exist when the audio designer attempts to use reverberant-to-direct ratio as a distance cue as they may inhibit localization and impede speech intelligibility (Shinn-Cunningham, 2004). In addition to conveying distance information, the audio designer can use intensity/loudness cues to convey information pertaining to the environment surrounding the listener in both proximal and distal spaces. The sound characteristics reaching a listener is dependent on distance, environment characteristics (e.g., absorption, composition, and presence of obstacles), and environmental conditions (such as temperature) (Begault, 1994; Walker & Kramer, 2004). In terms of sound source characteristics, humans judge source distance and depth more accurately for familiar sounds (Gardner, 1969). In addition, humans have better judgment of source elevation for familiar sounds due to the improvement in spectral content cues (Duda, 1997). The spectral content reaching a listener is modified as a result of passing around environment objects and obstacles (Begault, 1994). Also, in real world listening environments, sounds reaching a listener include both direct content and reverberant content that results from the reflections from walls and other environment objects, this reverberation content can be used to convey information pertaining to environment characteristics such as absorption, composition, and presence of obstacles (Begault, 1994; Blauert, 1983; Middlebrooks & Green, 1991). The audio designer should be aware that tradeoffs exist when adding the additional reverberation content to sounds to make it more realistic as they tend to inhibit the ability to localize sound sources accurately

(Shinn-Cunningham, 2004). When presenting sound sources in proximal and distal spaces, the audio designer can utilize concurrent visual and haptic event to drive a listener's judgment to sound source location (Driver & Spence, 2000). In order to enable a listener to differentiate between different auditory streams, the audio designer needs to vary the frequency and/or complexity of the different concurrent sounds sources (Bregman, 1990). When possible, the audio designer should use sources directly in front of the listener in order to improve azimuth localization in both proximal and distal spaces, but more importantly for sources closer to the listener (Brungart, 1999b; Brungart & Rabinowitz, 1999), since azimuth localization accuracy is worse for lateral sources with respect to the listener. To convey information pertaining to moving sound sources in proximal and distal space, the audio designer should utilize the pitch drop of a sound source moving by a listener (Neuhoff & McBeath, 1996).

Taken together, the following are design principles for conveying spatial information in the proximal and distal spaces (see Figure 5):

- Use interaural time and intensity difference cues to present sound azimuth information (Blauert, 1983).
- Use intensity/loudness and reverberant-to-direct to present sound distance (Blauert, 1983; Middlebrooks & Green, 1991).
    - In proximal space, interaural intensity difference cues can also be used (Brungart, 1999a; 2001).
- Use spectral modifications in the HRTF to represent sound elevation (Brungart & Robinowitz, 1999).
- Utilize dynamic cues resulting from source and head movement to disambiguate sound source location and reduce front-back errors (Brungart, 1999a).

- Use sounds in front of the listener, if possible, in order to improve azimuth localization (Brungart & Rabinowitz, 1999).
    - This is more important in the proximal space (Brungart, 1999b).
- Use reverberant content to add realism to sounds sources (Begault, 1994; Walker & Kramer, 2004).
    - Note that that addition of reverberation content may impede localization (Shinn-Cunningham, 2004).

Theoretical Models Validation

The current work presented three theoretical models for integrating sounds in interactive applications. In order to assess the utility of the developed models, and identify opportunities for improvement, a panel of audio experts was selected to respond to a validation questionnaire. The panel of experts included; 1) a research scientist who specializes in near field auditory localization and auditory display design, 2) an entrepreneur who focuses on spatial audio integration in real world and virtual reality training systems, and 3) an associate professor who specializes in auditory attention, binaural and spatial hearing, auditory scene analysis, effects of reverberation on perception, and physiologically-based models of spatial auditory processing.

Each panel member was provided with a copy of the models along with a two-page summary describing the developed models and was asked to provide a candid Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis of the developed models. In addition, each member responded to a set of questions pertaining to their satisfaction level (on a Likert 5-

point scale, with 1: Strongly disagree, 3: Neutral and 5: Strongly agree) for each model with respect to whether or not:

- The model is well-thought through and is at the right level of detail.
- The model provides audio designers a useful tool for integrating auditory cues in their applications.
- The model is clearly displayed graphically, in terms of text and presentation.
- The model is consistent with previous literature on auditory research.
- The model logic is easy to follow.
- The model does not contain terms that are unknown to an audio designer.
- The model provides a unique way for synthesizing prior auditory knowledge.
- The model can be used as a part of a training tool for audio designers.

Table 3 presents the experts' SWOT responses. The main strengths of the models include that they provide a systematic approach to auditory display design and that they integrate a wide variety of knowledge sources in a concise manner. The main weaknesses of the models include the lack of a structured process for amending the models with new principles, some branches were not considered parallel or completely distinct, and lack of guidance on selecting interface sounds. The main opportunity identified by the experts was in the ability of the models to provide a seminal body of knowledge that can be used for building and validating auditory display design. The main threats identified by the experts were that users may not know where to start and end with each model, the models may not provide comprehensive coverage of all uses of auditory displays, and the models may act as a restrictive influence on designers or they may be used inappropriately.

Table 3: Expert SWOT Analysis

Strengths	**Weaknesses**
(E1) These models provide, for the first time, the possibility of a systematic approach to auditory display design. (E2) Concise coverage of most relevant cues influencing sound perception. (E3) Integration of wide variety of sources of knowledge. (E3) Organization is generally useful. (E3) Immense amount of information is packed into a very concise summary.	(E1) There should be a systematic approach developed that outlines structured approach for amending the models with new principles. (E2) Guidance is very weak for how to choose sounds, other than mentioning speech and the use of musical instruments rather than timbre. (E3) Organizational framework not completely consistent, in that some "branches" are not parallel or completely distinct. (E3) Not always clear if word "cue" means an acoustic attribute (the normal meaning of the word) or a physical sound source. (E3) Rationales should be given for each recommendation to enable users to weight how to compromise, if compromise is necessary.
Opportunities	**Threats**
(E1) These models, once developed fully, will constitute a seminal body of principals that can be used to develop and validate auditory display design. (E1) The models can provide a structured framework upon which other principals in auditory display design can be developed, cataloged and related to existing work. (E3) Give more intuition into the recommendations.	(E1) The models may never provide a comprehensive coverage of all uses of auditory display and, if adopted, may either act as a restrictive influence or be used inappropriately for applications that do not fit into the intended uses. (E3) Users may not know where to start or what each point means.

E1 – E3: Expert assignment based on response time.

Table 4 provides the descriptive statistics for experts' responses on satisfaction questionnaires. The average responses ranged from 3.0 to 4.7. Figure 6 graphically illustrates average expert responses for each model. There are only three average responses below 4, but still greater than 3, which are clarity of graphics of temporal model, and logicality of spatial and temporal models. Appendix A graphically displays individual expert responses for each model.

Table 4: Statistics for Expert Responses on Theoretical Models Validation Questionnaires

Model	Dimension	Mean	StDev	Minimum	Median	Maximum
Audio Integration	Can be used for training	4.0	0.0	4.0	4.0	4.0
	Consistency with Literature	4.7	0.6	4.0	5.0	5.0
	Graphics are clear	4.7	0.6	4.0	5.0	5.0
	Logical	3.7	1.2	3.0	3.0	5.0
	No unknown terms	4.0	1.0	3.0	4.0	5.0
	Right level of detail	4.0	0.0	4.0	4.0	4.0
	Uniqueness	4.0	1.0	3.0	4.0	5.0
	Usefulness	4.0	1.7	2.0	5.0	5.0
Temporal Audio	Can be used for training	4.7	0.6	4.0	5.0	5.0
	Consistency with Literature	4.3	0.6	4.0	4.0	5.0
	Graphics are clear	3.0	1.0	2.0	3.0	4.0
	Logical	3.0	1.0	2.0	3.0	4.0
	No unknown terms	4.3	0.6	4.0	4.0	5.0
	Right level of detail	4.0	1.0	3.0	4.0	5.0
	Uniqueness	4.3	1.2	3.0	5.0	5.0
	Usefulness	4.0	1.0	3.0	4.0	5.0
Spatial Audio	Can be used for training	4.3	1.2	3.0	5.0	5.0
	Consistency with Literature	4.3	0.6	4.0	4.0	5.0
	Graphics are clear	4.0	1.0	3.0	4.0	5.0
	Logical	3.3	0.6	3.0	3.0	4.0
	No unknown terms	4.0	1.0	3.0	4.0	5.0
	Right level of detail	4.3	0.6	4.0	4.0	5.0
	Uniqueness	4.3	1.2	3.0	5.0	5.0
	Usefulness	4.0	1.0	3.0	4.0	5.0

Based on input from the audio experts, the following changes have been made to the models:

- Sample interface sound selection guidelines are now included on the audio integration model (see asterisks in Figure 1).
- Modifications to the various branches of the spatial audio theoretical model, for example, initially only ventriloquist effect was included in the affecters section, and now this is extended to include other multi-sensory effects (see asterisks in Figure 5).

- The wording of the design principles associated with the temporal and spatial theoretical models is improved for clarity (see asterisks in Tables 1 and 2).
- The wording of the temporal and spatial audio theoretical model is enhanced for clarity (see asterisks in Figures 2 and 5).
- Thorough coverage for both spatial and temporal theoretical models' dimensions is given in this paper.

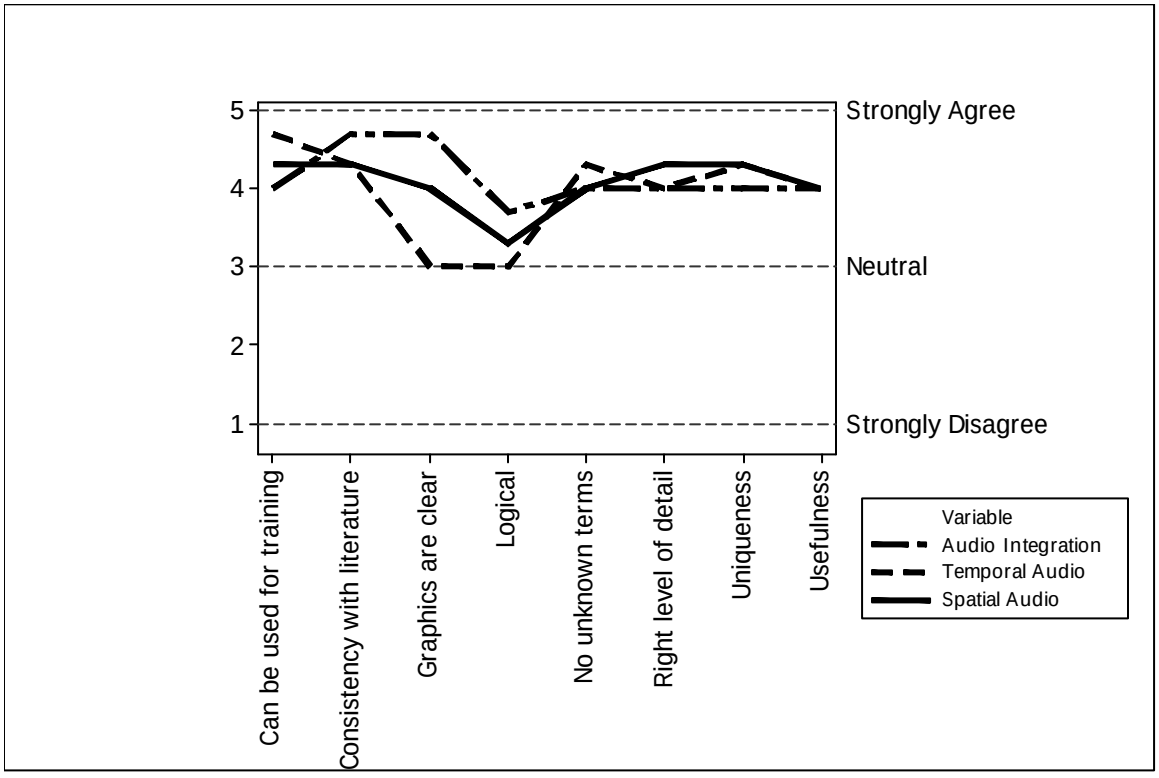

Figure 6: Average Expert Responses to Validation Questionnaires

Discussion and Conclusions

The set of models presented in this paper put forward a foundational framework for integrating sound cues in interactive application. Previous research has attempted to develop comprehensive frameworks for spatial visualization design (Buagajska, 2003) and perceiving layout (Cutting & Vishton, 1995). Nevertheless, there are currently no foundational theoretical models for integrating auditory information.

Three models were presented in this paper, including: 1) an audio integration model that addresses the end-to-end decision making process for integrating auditory cues in interactive applications, 2) a temporal audio theoretical model that addresses considerations pertaining to the timing of presenting auditory cues to achieve certain performance objectives, and 3) a spatial audio theoretical model that addresses the spatialization of sounds. Taken together, the developed models equip audio designers with a body of knowledge with respect to integrating sounds in interactive applications. Nevertheless, as noted by experts, the models may never provide a comprehensive coverage of all uses of auditory display and, if adopted, may either act as a restrictive influence or be used inappropriately for applications that do not fit into the intended uses, and audio designers may not know where to start or what each point means. Despite these limitations, as noted by experts, the developed models provide concise coverage of most relevant cues influencing sound perception. In addition, the models provide a vehicle for cataloguing and conveying design principles for auditory displays.

Future Work

The models presented in this paper put forward for the first time a classification framework for auditory information integration within interactive applications. Several future research directions emanate from the present work, these include:

- Evaluating alternative presentation approaches to make the models more intuitive.
- Building the knowledge base presented in the models in a training tool format for audio designers.
- Developing a systematic approach that outlines how to amend the models with new principles.
- Developing theoretical models for selecting various interface speech and non-speech sounds.

Acknowledgements

This material is based upon work supported in part by the Office of Naval Research (ONR) under its Virtual Technologies and Environments (VIRTE) program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views or the endorsement of ONR. The authors would like to thank Dr. Douglas Brungart, Dr. Hesham Fouad, and Dr. Barbara Shinn-Cunningham who kindly provided feedback on the theoretical models presented in this paper.

CHAPTER THREE: TRAINING PACE IN VIRTUTAL REALITY TRAINING SYSTEMS[†]

This study presents an experimental evaluation of the utility of using auditory cues to train temporal tasks (e.g., pace setting) in virtual reality training systems. There were four different auditory cues used for training pace: 1) a metronome, 2) non-spatial auditory earcons, 3) a spatialized auditory earcon, and 4) no audio. Sixty-eight people participated in the study. A pre- post between subjects experimental design was used, with eight training trials. The measure used for assessing pace performance was the average deviation from a predetermined desired pace. The results demonstrated that a metronome was not effective in training participants to maintain a desired pace, while, spatial and non-spatial earcons were effective strategies for pace training. Moreover, an examination of post-training performance as compared to pre-training suggests some transfer of learning. Design guidelines were extracted for integrating auditory cues for pace training tasks in virtual environments.

Introduction

Auditory cues have been used to train spatial knowledge in virtual training systems from personal guidance systems for the visually impaired (Loomis et al., 1998) to Close Quarters Battle for Military Operations in Urban Terrain (MOUT CQB) (Jones et al., 2005). Nevertheless, there is an under explored opportunity for using auditory cues to present temporal information in such environments. This paper focuses on exploring the utility of auditory cues

[†] Manuscript submitted to *Military Psychology*.

for training temporal tasks in virtual reality training systems. Temporal events or tasks are specific occurrences in time that can be either instantaneous (i.e., instant-based) or can span a time interval (i.e., interval-based) and often involve both (Schreiber, 1994). For example, a progress bar is a visual construct that is commonly used to convey status information about a download task. The download task has two main instantaneous events; start and finish. Each increment on the progress bar also defines an instance; however, the overall download time reflects an interval-based measure. Audio is known to be superior when compared to visual when processing such temporal information (ETSI, 2002; Kramer, 1994, Repp & Penel, 2002).

When training temporal information, the appropriate training strategy will be dependent upon whether instant-based or interval-based temporal information is involved in the target training task. An example for using audio to train an instant-based temporal task is tapping a person's finger on a surface; the finger tapping can be controlled by playing a metronome sound that repeats every second, and asking the person to match their tapping speed with the metronome (Kurtz & Lee, 2003). Certain conditions, if met, enhance coordination accuracy and stability of such temporal performance, such as using synchronous and/or alternating tapping with fingers, and using simple harmonic ratios with pacing metronomes (Kurtz & Lee, 2003).

In addition, motor learning can take place by extending perceptual learning (Meegan et al., 2000), which deals with performing tasks related to the use of the senses, such as discriminating temporal intervals denoted by brief auditory stimuli. This is beneficial to this work where pace training took place in a perceptual fashion, where participants acquired pace skills solely through a virtual training system without actual motor performance, and these would in turn need to be extended into motor learning when using the acquired pace skills in real world tasks.

The work to date on auditory pacing strategies has largely been done in the physical world. This paper explores the efficacy of using auditory information to guide participants in controlling their pace to a predetermined desired value in a virtual training system. This study also examines short-term transfer of learning of the predetermined desired pace.

Background Literature

Real world applications for using audio to guide rhythmic movements are common in medicine, dancing, and sports (Boyle et al., 2002; Kaplan, 2002; Kern, et al., 1992; Libkuman et al., 2002; Wijnalda et al., 2005). When an external sensory stimulus is used to guide such rhythmic movement, audio cues generally result in the least variability from target rhythm as compared to visual or tactile cues (Chen et al., 2002; Kolers & Brewster, 1985). Once people synchronize their movements with that of auditory tones, they generally can maintain the pattern without the audio being played (Kolers & Brewster, 1985).

There are three types of temporally-related movements; these are goal-directed, rhythmic, and rhythm-modulated (Thaut, 2005). Goal-directed movements have a specific target, such as throwing a ball a certain distance in minimum time, or swinging a golf club. Rhythmic movements repeat at a constant rate such as tapping fingers or drawing circles. Rhythm-modulated movements repeat over time but at either increasing or decreasing rates, such as speeding up during exercise. Audio can be used to influence all three-movement types. For goal directed movements, instant-based sounds can be used to time start and finish of a movement (Thaut, 2005). Rhythmic movements can be controlled using both instant and interval-based sounds. Instant-based sounds can be used to mark cycle ends for rhythmic movements (Thaut,

2005). Also, interval-based sounds can be used as a guide for rhythmic movements by providing sounds that repeat at a consistent pace. Controlling pace or rhythmic movements in general, is contingent upon selecting an appropriate audio format for a specific time-related application. For example, an audio format that is often used to guide rhythmic movements is a metronome, which marks exact instant-based time increments by a regularly repeated tick (Kurtz & Lee, 2003). On the other hand, an example of interval-based audio is using sounds with varying tempos that indicate relative distance to objects (Day et al., 2004); i.e., shorter time delays between sounds as a user gets closer to a target or destination.

Rhythm-modulated movements can be controlled using interval-based sounds. In general, rhythm-modulation with audio cues involves using sounds that repeat at increasing or decreasing pace as a guide for physical movements. These movements are common in rehabilitation studies that involve gait, where typically a music tempo is initially chosen that accommodates an individual's gait capabilities and then the tempo is increased incrementally as gait performance improves (Thaut et al., 2004). Repp (2006) provided an example of rhythm modulation, when he noted that people who initially tapped their fingers at their own pace were influenced when exposed to rhythmic audio, and resynchronized their pace to that of the audio when the tempo difference was less than 10%.

Metronome use in behavioral ecology studies dates to the late 60's (Wiens et al., 1969); researchers have generally used metronomes to monitor specific individual's activities over time, or to time individual actions. For example, metronomes (or feedback earcons) have been used in training Cardiopulmonary Resuscitation (CPR) (Boyle et al., 2002; Kern et al., 1992). Metronomes have also been used to train rhythm-driven arm movements. For example, Thaut (2005) used a metronome to synchronize movement frequency and found less spatial and

temporal variability in metronome-driven trials as compared to self-paced trials. Thaut (2005) also used a metronome stimulus with three different sections (beats per minute) to demonstrate that participants were able to re-synchronize with a changing metronome within 50 ms. An interactive metronome, which plays feedback earcons, has been used to train consistent and correct timing on tasks (Bartscherer & Dole, 2005). For example, Libkuman et al. (2002) discussed an experiment where an interactive metronome was used for golf training. A constant metronome was played to guide participants in different tasks that they were required to perform (e.g., continuously moving hands in a circle and clapping them when they reached a particular point in the circle; a metronome was used to indicate at which point they should clap). In addition, earcons provided feedback regarding whether movements were late (low pitched tone in the left ear), early (high pitched tone in the right ear) or on time (+/- 15 ms of beat; high pitched tone presented in both ears at the same time). An experiment was performed in which pre- and post- evaluations of accuracy of golf shots from people who used this training device to train timing was compared to those who used other forms of golf training. Results suggested that training in timing and pacing of fine motor activities using audio as guidance can transfer to more complex tasks such as golf swings. This study is of interest since it shows that although the metronome-based training was not directly associated with the actual golf training, it still gave participants a grasp over pace control that was later useful when performing actual golf swings. There has been no use for spatialized feedback earcons or metronomes in pace setting to the best of the authors' knowledge.

Taken together, these studies demonstrate, as Kaplan (2002) suggested, that metronomes or interactive earcons can be used:

- To produce steady clicks to indicate desired beat.

- At the beginning of an exercise to establish the right tempo that corresponds to a beat and at the end to check whether the tempo stayed the same.
- To gradually become comfortable with a faster or slower tempo.
- To train consistent and correct timing on various tasks.

Beyond metronomes, music can be used to enhance physical activities such as exercise, which can involve both rhythmic and rhythm-modulated movements (Karageorghis & Terry, 1997). Music has been found to enhance physical activities in several ways. First, music can divert performer's attention away from physical stress and fatigue; second, music can enhance psychomotor arousal by acting as a stimulant before exercise and as a sedative during exercise; and third, music can enable performers to synchronize their physical rhythm to that of a musical rhythm (Karageorghis & Terry, 1997). Music can be used to support user performance in three modes; these are pace-fixing (i.e., playing music at a constant tempo to enable synchronization), pace-matching (i.e., playing music at a tempo that matches user's pace), and pace-influencing (i.e., playing music at varying tempo to influence a user to slow-up or slow-down) (Wijnalda et al., 2005). Taken together, these studies demonstrate that music can be used for pace setting in the following ways:

- To aid performers in synchronizing their physical rhythm to that of musical rhythm.
- To influence performers either to speed up or slow down during exercise.

Research Hypothesis

Design principles for conveying temporal information in virtual training environments can be theorized based on the above review; these include:

1. Metronomes can be used to train a consistent pace.
2. Earcons/metronomes can be used to gradually influence (increase/decrease) traversal pace.
3. Metronome-based (or audio-based in general) pace setting can be used to train paces that become internalized and can be used when a metronome is removed.

Method

In combat situations, soldiers might be required to follow a particular walking pace set by the group leader (USArmy, 2003). This study utilized a virtual reality training system to perceptually train participants on a pace setting task. Figure 7 provides a conceptual framework for the study. As depicted in Figure 7, each participant performed pre-training, post-training, and a set of eight training sessions. The training sessions included audio cues for training pace, as discussed below. Participants' performance was assessed by:

- Comparing performance on training session 8 (audio cues present and after learning the cues) to pre-training to assess the utility of using audio to cue pace.
- Comparing performance on post-training (no audio cues present) to pre-training to assess pace internalization.

Participants

A total of sixty-eight participants (mean age = 19.9 years; s.d. = 3.6 years; 36 females and 32 males) participated in this study. Participants were randomly assigned to one of four different treatment conditions. All participants reported normal or corrected to normal vision

and normal hearing. All participants were right-handed. Participants were recruited through a university-based subject pool, and they voluntarily agreed to participate in the experiment for class credit.

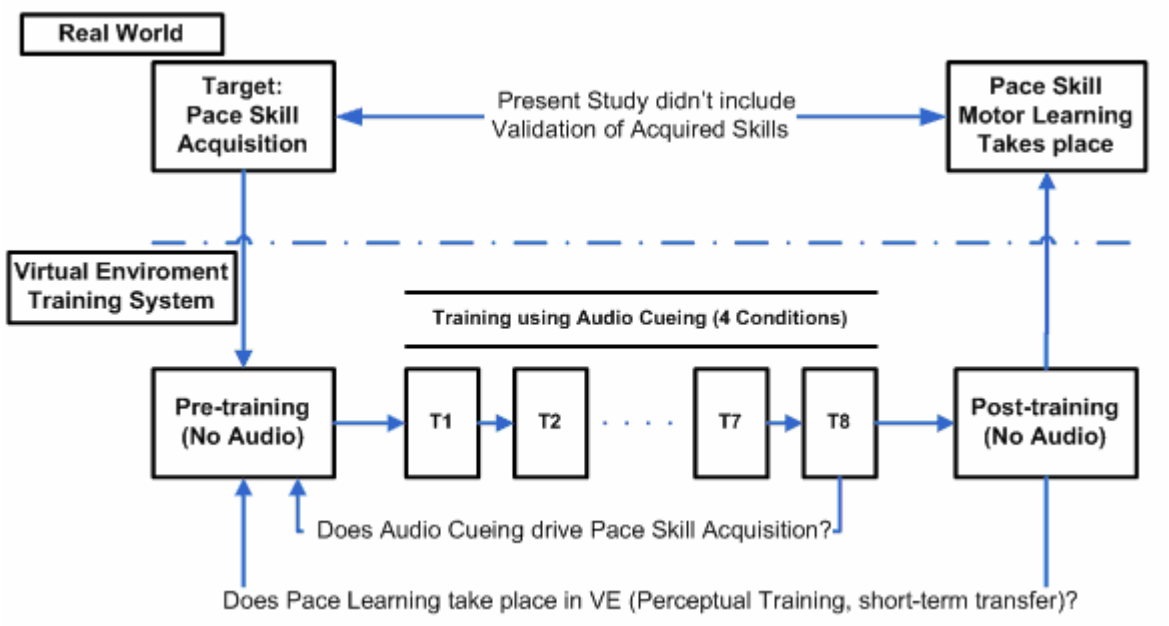

Figure 7: Pace Training Model

Apparatus

The experimental setup consisted of two dual-processor Dell Precision computers. One computer was used to generate graphics for a training task with ManSim software and the other was used to produce audio cues using ViBeStation software. Participants interacted with the training task through an immersive Virtual Research Systems V8 Head Mounted Display (HMD). Audio cues were presented using Sennheiser headphones built into the V8 HMD. Head tracking was done using an Intersense InertiaCube tracker. Participants navigated through the task through a standard Saitek game controller.

Several questionnaires were used in this study including, the verbalizer-visualizer (Richardson, 1977), immersive tendencies (Witmer & Singer, 1988), presence (Witmer & Singer, 1988), NASA Task Load Index (TLX) workload (Hart & Staveland, 1988), and the simulator sickness (SSQ; Kennedy et al., 1993) questionnaires. The verbalizer-visualizer questionnaire is scored on a scale from 1-15; as the participant's score on the questionnaire goes higher, the more indication that the participant is a visualizer. The immersive tendencies questionnaire has 39 questions on a scale from 1-7. As the participant's average score goes higher, the more indication that the participant has a tendency to get immersed. The presence questionnaire has 37 questions on a scale from 1-7. As the participant's average score goes higher, the higher the participant's reported sense of presence. The NASA-TLX consists of six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. For each scale, individuals rate the demands imposed by the task as well as each scale's contribution to the total workload, the latter of which is calculated by summing the product of each scale's rating and weight. The SSQ has participants report the degree to which they experience a set of symptoms as one of "None," "Slight," "Moderate," or "Severe," which are then combined into a total sickness score.

Virtual Environment

The virtual environment (VE) was designed to mimic a room clearing exercise and included a 15 room building to be cleared. Ten different variations of the environment at comparable task difficulty were created to be used for training and testing. Each environment

contained 5 open doors, 8 enemy entities, 4 friendly entities, and 4 mouse holes. See Appendix B for a screen shot of the VE used in this study.

Figure 8 shows an example environment layout. Four different versions of each environment layout were created based on the various audio conditions evaluated (see experimental design, below). All environments had "user's foot steps" and "gun shots" sounds implemented. One environment variation was randomly selected for pre- and post-testing. The pre- and post-testing environment did not include auditory cues for pace training. Pre- and post-testing were used to assess the perceptual transfer of training within virtual environment task performance when auditory cues are removed, see Figure 7. In addition, the training environments had audio implementations depending on the experimental condition;

- No audio for pacing: No audio cues for pace training were present, i.e., only "user's foot steps" and "gun shots" sounds were implemented.
- Metronome: A metronome sound with three settings (slow: 2 m before open door; medium: no open door or mouse holes; fast: 2 m before and after mouse holes) was implemented depending on the location of the participant in the environment.
- Non-spatial earcons: Two diotic metaphoric audio cues were selected to guide participants in setting their pace. If the participant was traversing the hallway at the "correct" pace, no audio was played. If the participant needed to slow down (2 m before an open door), a drum sound was played. If the participant needed to speed up (2 m before and after a mouse hole), a flute sound was played. The rate of the played audio was proportional to the deviation from the predetermined desired pace.
- Spatialized earcons: A spatialized (front or back) metaphoric audio cue was selected (flute) to guide participant in setting their pace. If the participant was traversing the

hallway at the "correct" pace, no audio was played. If the participant needed to slow down (2 m before an open door), the audio was played in front of the user. If the participant needed to speed up (2 m before and after a mouse hole), the audio was played from the back. The rate of the played audio was proportional to the deviation from the predetermined desired pace.

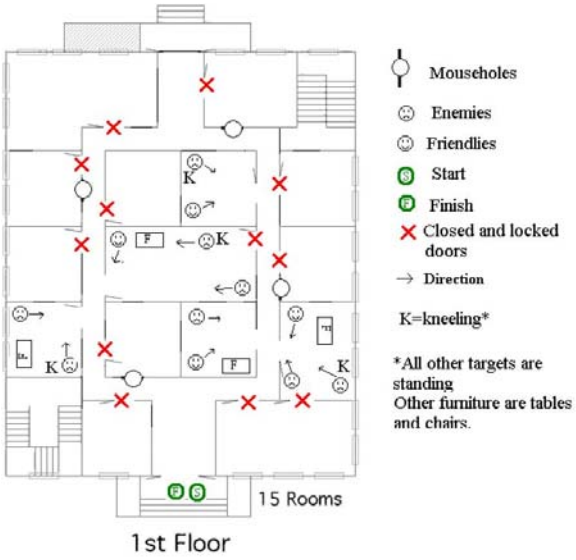

Figure 8: Virtual Environment Layout Example

Tasks

Each participant performed a series of Close Quarters Battle for Military Operations in Urban Terrain (CQB for MOUT) activities in a virtual environment. There were two primary tasks that the participants were expected to complete. The first task was to maintain a consistent pace while traversing the environment hallways. Three predetermined paces were selected based on the environment configuration and dimensions, these were a slow pace when approaching open doors, a fast pace when passing mouse holes on walls, and a medium pace when no open

doors nor mouse holes were present. The second task was to enter and clear all open rooms and engage all hostile and non-hostile units located therein. To engage units, participants had to point their weapon towards a unit, and then use the correct button on the game controller to identify them as either friendly or foe (i.e., left controller button was used to clear friendly units, and right controller button was used to fire upon foe units). When moving through the environment, turning was controlled through head movement. Locomotion (i.e., stepping forward and back) was controlled using the game controller. All participants wore the HMD and headphones during pre-training test, all training sessions, and post-training test.

Experimental Design

This study utilized a pre-post between subjects one-factor ANOVA design. The one factor was audio condition, with four different levels: 1) a metronome, 2) cueing using non-spatial audio (using one audio cue to indicate the need to speed up and another audio cue to indicate the need to slow down), 3) cueing using 3D audio (audio was played either in front [to push them backward] or to the back [to push them forward of the listener to regulate pace); and 4) a no audio control. Performance was assessed using the average deviation from the predetermined desired traversal pace for the hallway traversal task, and time and accuracy for the room clearing task. Both audio cueing effects on pace setting (comparing last training to the pre-training) and near term transfer of pace setting skills (comparing post-training to the pre-training) were evaluated. Performance on the room-clearing task was assessed using percent hits over fire and average task completion time. In addition, workload, presence, and simulator sickness were assessed.

Procedure

Before the start of a test session, participants completed an informed consent, demographics questionnaire, immersive tendencies questionnaire and verbalizer-visualizer questionnaire. After that, participants completed task training involving interaction with the training environment and a pre-recorded animation that illustrated the predetermined traversal paces. Once the training was done, participants completed a pre-training test, where they were required to complete their tasks with no audio cueing. Then, participants were randomly assigned to one of the four different audio conditions. Additional training was given based on the assigned condition. Participants in an audio condition watched a presentation illustrating the earcons that they would experience, while participants in the no-audio condition were reminded verbally on their task objectives and the need for maintaining a consistent pace. Participants then completed eight training sessions based on the assigned experimental condition. The order of the training sessions was randomized for each participant. After the training sessions, participants completed a post-training test, which used the same environment and settings as the pre-training test. In addition, participants completed the simulator sickness questionnaire and NASA TLX workload index after pre-testing, training session number 1, training session number 8, and post-testing. Once testing was completed, participants completed a presence questionnaire. Finally, the participants were provided with a written and oral debrief about the pace training experiment.

Results

Performance data in terms of participants' location, speed, and time were logged through the experimentation software. Although 68 participants completed the study, only 53 had their

data correctly logged via software, and hence were included in performance data analysis. Participant log files were processed for average traversal paces and average deviations from predetermined desired pace. Subjective questionnaire data were manually input and all 68 participants' data were available. Statistical significance was assessed using an alpha of 0.1 for pace performance data because of the nature of audio equipment and implementation used, which may have induced some variation. For the subjective questionnaire data, an alpha of 0.05 was used.

Pace Performance Results

Table 5 provides descriptive statistics on average traversal pace, average deviation from predetermined desired pace, percent hits over fire and average time to complete each task for pre-training, training session number 8, and post-training. These results included performance on both pace and room clearing tasks. Nevertheless, since the objective of this study was pace training, more attention will be given to pace task performance.

Figures 9, 10, and 11 depict the average paces for hall traversal, entry danger areas, and mouse hole danger areas. Table 6 presents ANOVA p-values for average pace and average deviation from predetermined desired pace for entry danger areas, mouse hole danger areas, and hall traversal.

An ANOVA was used to compare the average deviations from the predetermined desired pace among the various audio conditions. ANOVA results show significance for both training session number 8 compared to pre-training ($p < .002$) and post-training compared to pre-training ($p < .06$).

Table 5: Statistics for Pace Performance

Dependent Variable	Audio Condition	Pre-Training	Training 8	Post-Training
Average Pace for Hall, Mouse Hole, and Door Entry Areas (m/s)	**No Audio**	0.381 (0.095)	0.431 (0.094)	0.433 (0.087)
	Metronome	0.375 (0.093)	0.4426 (0.0939)	0.407 (0.099)
	Non-spatial	0.398 (0.088)	0.395 (0.052)	0.362 (0.094)
	Spatial	0.367 (0.104)	0.3784 (0.072)	0.345 (0.054)
Average Deviation from Desired Pace (m/s)	**No Audio**	0.11438 (0.036)	0.1254 (0.058)	0.109 (0.062)
	Metronome	0.117 (0.042)	0.123 (0.059)	0.111 (0.061)
	Non-spatial	0.129 (0.047)	0.101 (0.046)	0.119 (0.049)
	Spatial	0.106 (0.038)	0.0894 (0.045)	0.094 (0.033)
Percent Hit Over Fire	**No Audio**	90.16 (10.43)	94.62 (7.08)	95.46 (5.76)
	Metronome	95.44 (11.41)	95.71 (8.52)	98.41 (4.03)
	Non-spatial	91.62 (11.58)	95.88 (8.46)	90.38 (11.36)
	Spatial	93.96 (9.14)	95.39 (12.57)	94.72 (8.76)
Average Time to Complete Task (ms)	**No Audio**	249089 (52471)	223846 (51413)	209721 (42083)
	Metronome	265369 (54494)	200095 (67130)	243808 (44833)
	Non-spatial	267570 (60322)	208650 (26968)	240376 (44088)
	Spatial	264393 (65162)	227838 (37764)	245615 (41862)

Table 6: ANOVA Comparisons- p-values

Dependent Variable	Training 8 – Pre-training	Post-training – Pre-training
Average Pace (Door Entry)	0.426	0.561
Average Pace (Mouse Hole)	0.117	0.001*
Average Pace (Hall)	0.452	0.206
Average Dev (Door Entry)	0.291	0.522
Average Dev (Mouse Hole)	0.962	0.020*
Average Dev (Hall)	0.002*	0.055*

* Significant

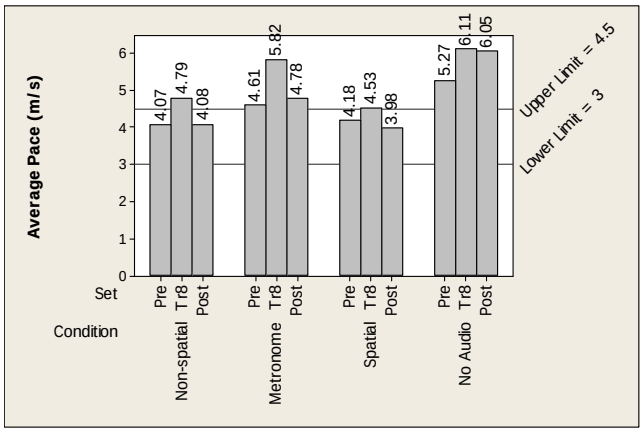

Figure 9: Average Pace for Hall Traversal

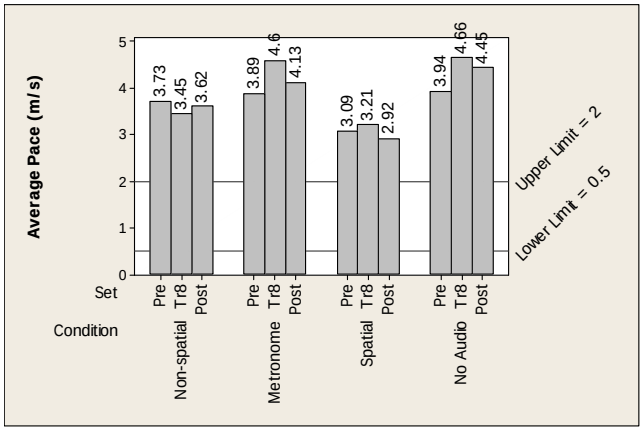

Figure 10: Average Pace for Entry Danger Areas (Before Open Doors)

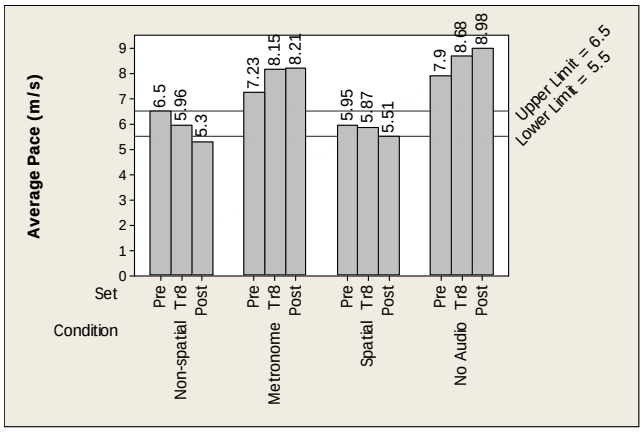

Figure 11: Average Pace for Mouse Hole Danger Areas

Tukey's post-hoc analysis indicated that the average deviation from predetermined desired pace for the no audio condition was significantly different from the non-spatial auditory cueing condition ($p<.004$) and the spatial auditory cueing condition ($p<.02$). On average, the deviation from predetermined desired pace for the no audio condition was greater than that for the non-spatial audio by 0.055 m/s (s.d. = 0.015) and greater than that for the spatial audio by 0.046 m/s (s.d.= 0.015). These results suggest that spatial and non-spatial audio appear to be more effective strategies for pace setting for hall traversal and in the vicinity of mouse holes than no audio. Also, the graphs indicate that the metronome was not an effective strategy for pace setting. The average deviation from predetermined desired pace for the metronome cueing condition was 0.014 m/s (s.d. = 0.015) greater than that for the no audio condition, which was not significant ($p>.77$).

Subjective Questionnaires' Results

The participants' visual/verbal abilities were assessed using the visualizer-verbalizer questionnaire. The average score on the questionnaire was 6.75 (s.d.= 1.68); 7 participants had a score above 8 (visualizers), 8 participants had a score of 8 (neutral), and 44 participants had a score below 8 (verbalizers). The participants' immersive tendencies were assessed using the immersive tendencies questionnaire. The average score on the immersive tendencies questionnaire was 4.25 (s.d. = 0.53).

After completing all testing and training tasks, the participants' subjective sense of presence was assessed using the presence questionnaire. The average score on the presence questionnaire was 4.43 (s.d. = 0.55). The Pearson Correlation between immersive tendencies

and presence questionnaires average scores was 0.3 ($p < .013$). A one-way ANOVA was used to compare presence scores among the various auditory conditions. No significance was detected ($p > 0.12$; power > 42.54%), which indicates that presence scores did not significantly depend on auditory condition.

The average participants' total sickness scores on the SSQ were:

- Pre-training → 18.37 (s.d. = 22.27).
- Training session number 1 → 19.09 (s.d. = 21.37).
- Training session number 8 → 30.86 (s.d. = 33.28).
- Post-training → 27.83 (s.d. = 33.69).

These value suggest that participants experienced various degrees of simulator sickness throughout the experiment; however, no significant effects were found for audio condition on average total sickness scores ($p > 0.05$; for pre-training, training session 1, training session 8, and post-training).

In addition, participants completed the NASA TLX questionnaire to assess subjective workload. The total participant scores on the NASA TLX questionnaire were:

- Pre-Training → 19.48 (s.d. = 10.41).
- Training session number 1 → 24.35 (s.d. = 11.52).
- Training session number 8 → 24.27 (s.d. = 11.36).
- Post-Training → 22.37 (s.d. = 11.97).

Paired t-tests were used to compare the average workload scores. The results show that the total TLX score was statistically higher for training number 1 ($p<.0001$), training number 8 ($p<.0001$) and post-training ($p<.017$) compared to pre-training. The Kruskal-Wallis test was used to compare median total TLX scores across audio conditions, since the scores were not

normal. The results showed no significant difference between post-training and pre-training medians, but, the training number 8 medians were statistically different among the various auditory conditions ($p<.017$). Table 7 shows the NASA TLX total score medians.

Table 7: NASA TLX Total Score Medians

Audio Condition	N	Training 8 - Pre-training	Post-training – Pre-Training
No Audio	17	-1.50	-3.00
Metronome	17	3.50	2.50
Non-spatial	17	5.00	3.00
Spatial	17	7.00	4.25

Discussion

The present study examined the utility of using audio to train pace in an interactive virtual training system for MOUT CQB. In particular, three auditory conditions were compared to a no audio control group to assess the utility of using auditory cues to set pace and whether short term transfer of training for pace skill can take place.

The results of the experiment revealed that using a metronome was not effective in training participants to maintain a desired pace. This is at odds with past literature that suggests metronomes can be used for guiding rhythmic movements in the real world (c.f., Boyle et al., 2002; Kern et al., 1992, Kurtz & Lee, 2003). Thus, there may be a fundamental difference between physical and virtual worlds that hinders the use of a metronome. This might be due to the difficulties of metronome implementation with respect to footstep sound implementation. When trained with the metronome, participants had to match two sounds; the footsteps sound to the metronome, which may have been difficult due to volume differences between the metronome and footstep sounds. Future virtual metronome implementation should consider the

relationship between metronome loudness level and the other sounds present, such as footsteps, and consider training first with just one of the sounds (e.g., the metronome) and then introducing the sound that is to be synched to the metronome. Although in real world applications, a metronome is a very effective strategy for pacing setting, this study has shown that there might be some difficulties in integrating a metronome in interactive VE training systems. These findings fail to validate the first design principle that metronomes can be used to train a consistent pace in virtual environments.

On the other hand, the results of the present experiment provide preliminary validation for using earcons to influence (increase/decrease) traversal pace gradually. Specifically, auditory cueing, as indicated by deviation from predetermined desired pace for the difference between training session number 8 and pre-training, has shown to be an effective strategy for setting desired pace. Both the non-spatial and spatial audio conditions resulted in less deviation from the predetermined desired pace as compared to the no audio condition and these deviations were statistically significant. This supports the literature related to using auditory cues to drive pace setting (c.f., Karageorghis & Terry, 1997; Thaut, 2005).

The results also suggest that audio-based pace setting in virtual environments may become internalized and thus can be used when the audio is removed. Specifically, examining performance on post-training compared to pre-training, indicates that the average deviation from desired pace in hallways was significant when audio cueing was present (i.e., training session number 8, $p<.002$), as well as when it was taken away upon post-training (i.e., post-training, $p<.055$), suggesting transfer of learning, or the ability of participants to maintain pace setting when the audio cueing was removed. This extends the literature describing the internalization of pace skills. Past literature demonstrated that people maintain their rhythm after removing a

synchronizing auditory cue, and when people learn timing on fine motor tasks (like drawing circles or clapping hands) using metronomes and feedback earcons, the learned skills transfer to more complex skills, such as golf swings (c.f., Kolers & Brewster, 1985; Libkuman et al., 2002). Previous research dealt with training pace using auditory cues in the real world, whereas in this study, the pace training findings are extended to a virtual world and show that internalization is attainable in virtual environment training systems. The evidence supporting internalization is very encouraging, as it suggests that virtual environments may be effective trainers for temporal skills that may be difficult to train in the real world due to limited access or potential danger (e.g., emergency procedures, military operations, etc.).

An interesting finding from this study was that both non-spatial and spatial auditory cueing resulted in comparable pace setting performance. The non-spatial audio condition used two different earcons, while the spatial audio condition used a front-back spatialization of a single earcon. Both audio conditions had a similar implementation of earcons' variation based on deviation from desired pace, where the rate of played audio was proportional to the deviation from the predetermined desired pace. The implication of this finding could be that if auditory earcons are to be used for pace training in a virtual environment, the choice of implementing spatialized or non-spatialized earcons can be a design decision, as both appear effective. If the environment contains many spatialized cues, then using non-spatialized cues might be a better option. Nevertheless, if there are several different earcons in the environment, spatialized earcons could be a better option, as they can simplify the number of earcons used.

Participants' subjective assessments indicate that the training environment produced various degrees of sickness as indicated by total sickness score greater than 7.48 (Stanney et al., 2002); however, these symptoms due not appear related to the audio cues presented as there were

no significant differences in total sickness scores due to audio condition. When assessing participants' subjective feelings of presence, no significant differences were found based on audio condition, which can imply that earcons and metronome conditions did not differ from the no audio condition in terms of achieving a sense of immersiveness. One possible explanation for the lack of significance in reported subjective feelings of presence may be due to possible network delays between the tracker, visual computer and audio computer. This warrants further investigation, as any slight delay may indicate that audio implementations were not perceived as realistic by participants, thus further work on the ecological validity of audio cues for pace training is warranted. The results of the subjective workload assessment indicate that audio conditions may increase perceived workload as indicated by comparing workloads between training session number 8 and pre-training and that the training environment imposed some sickness risks to participants. The implication of this increased workload is that using audio to train pace in a VE may impose additional workload for trainees, and hence designers need to balance this increase in workload vs. the desired benefit of pace training. The implication of the reported simulator sickness is that care should be exercised when using virtual training simulators to ensure participants are in good health before they leave the simulator facility. As there were no performance losses associated with this increase in perceived workload and simulator sickness, and in fact, pace performance was enhanced with audio cues, it would appear that if effectively implemented the benefits of such audio cueing should outweigh the costs.

The results of the study presented in this paper provide design principles for integrating auditory cues in interactive environments to train temporal tasks (e.g., setting pace), these include:

- Spatialized and non-spatialized auditory earcons may be effective strategies for pace training in interactive training systems.
- Audio designers should consider using spatialized earcons when there are several different earcons in an environment (adds one less earcon).
- Audio designers should consider using non-spatialized earcons when several aspects of the environment are spatialized using audio (reduces the complexity of the environment).
- Using auditory cues for pace training can increase user's workload, and hence designers need to exercise care when adding these cues.
- It may be difficult to utilize strategies that have proven effective in the real world in interactive applications. For example, metronomes are an effective real world option for pace setting. But due to the complexity of metronome implementation, such techniques may not prove as effective for pace setting in virtual environments.

Conclusions

This paper presents the results of an experimental evaluation on the utility of using auditory cues to train temporal tasks in interactive virtual environments. Study participants completed pre and post-training tests and eight training sessions where different audio conditions were used to train pace. Auditory cueing using both spatialized and non-spatialized earcons was found to be an effective strategy for pace setting in such environments. Moreover, there was an indication that participants were able to internalize pace training, thus indicating transfer of learning. The results of the present study provide preliminary evidence that using audio to train temporal tasks can be extended to virtual training systems. Nevertheless further research is

needed to assess; 1) when it is best to use spatialized vs. non-spatialized earcons for presenting pace setting cues, 2) the utility for using audio to train concurrent spatial and temporal tasks in virtual environments, and 3) how to best design auditory for these training systems to enhance their ecological validity.

Acknowledgements

This material is based upon work supported in part by the Office of Naval Research under its Virtual Technologies and Environments (VIRTE) program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views or the endorsement of ONR. Parts of this work were previously presented at the 5th Annual Auditory Perception Cognition and Action Meeting (APCAM 2006) and as a poster at the Society of Hispanic Professional Engineers Eastern Technical and Career Conference (SHPE ETCC).

CHAPTER FOUR: SPATIAL AND TEMPORAL INFORMATION INTEGRATION USING AUDIO

This study presents an experimental evaluation of the utility of using auditory cues to train spatial (e.g., location of enemy and friendly units) and temporal (e.g., pace setting) tasks in virtual reality training systems. To train spatial tasks, three different spatial audio fidelity levels were used: 1) non-spatial, 2) a default HRTF, and 3) a "best-fit" HRTF. Non-spatial auditory earcons (i.e., temporal auditory cues) were presented to all participants to train traversal pace. Thirty people participated in the study. Participants were engaged in a series of Close Quarters Battle for Military Operations in Urban Terrain activities in a virtual environment, where they were to clear a series of rooms while travelling at a target pace. A pre- post between subjects experimental design was used, with eight training trials. The measure used for spatial task performance was time to complete each task, and the measure used for temporal task performance was the average deviation from a predetermined desired pace. The results demonstrated that temporal auditory cues were effective in influencing pace while other cues were present. On the other hand, spatialized auditory cues did not result in significantly faster task completion. Based on these results, a set of design guidelines was proposed that can be used to direct the integration of spatial and temporal auditory cues for supporting training tasks in virtual environments. The results of this study should be of interest to systems' designers and researchers who are examining the utility of integrating audio cues into human-system interfaces with the objective of enhancing human performance and improving training effectiveness.

Introduction

Next generation training systems are transitioning from primary visual-audio experiences to truly multimodal engagements. Such multimodal systems are expected to have extreme graphics "with some spatial audio interfaces and haptic interfaces" and eventually, "spatial-audio effects, full-hand haptics, and olfactory displays will also be available" (National Research Council, 2000, p.25). Despite these projections, currently there is a limited understanding with respect to the utility of using such multimodal technology to enhance training, and there are few existing guidelines to aid in designing and implementing multimodal systems (for some notable exceptions see ETSI, 2002; Stanney et al., 2004).

Of specific interest to the current study, advances in spatial audio technology make it possible to leverage audio cues to enhance multimodal human-systems performance. When implemented effectively, spatial audio cues are expected to improve human performance in high stress applications, such as aircraft cockpits and advanced command and control operations centers, as they are suggested to increase situational awareness and direct attention (Begault, 2000). For example, spatialized audio cues have been used to train spatial knowledge in virtual training systems from personal guidance systems for the visually impaired (Loomis et al., 1998) to Close Quarters Battle for Military Operations in Urban Terrain (CQB for MOUT) (Jones et al., 2005). In general, these previous studies have found that spatial displays enable participants to complete their tasks up to one and a half times faster than non-spatial displays. Auditory cues have also been used to train temporal knowledge, such as rhythm (Thaut, 2005) and traversal pace (Ahmad et al., under review b). These studies, among others, have shown that audio cues

generally result in the least variability from target rhythm as compared to visual or tactile cues and once rhythm is established, it can be maintained without the audio being played (Chen et al., 2002; Kolers & Brewster, 1985).

The current study explores the effectiveness of audio cues to train spatial and temporal tasks in combination. Specifically, this paper explores the efficacy of using auditory cues to guide participants in controlling their pace to a predetermined desired value, and presenting spatial audio information to indicate presence of enemy and friendly entities, with the objective of the latter being to enable faster task completion. The validated design guidelines presented in this paper aim to establish a baseline of user-centered audio design science, thereby advancing the state-of-the-art in auditory display design.

Background Literature

Sounds provide substantial information to humans about their surroundings, from ecological sounds characterizing the environment to speech, which is the foundation of human-to-human communication (Walker & Kramer, 2004). Despite its recognized importance, audio is under-utilized in current human-system interfaces, which are primarily visual - mainly consisting of constructs such as windows, icons, menus, and pointing devices (WIMP; Pew, 2003). Yet audio has the potential to enhance all components of human task performance (i.e., psychomotor, affective, and cognitive; Bloom, 1956).

The temporal dimensions of audio can be used to enhance different aspects of human performance. For example, instant-based sounds can be used to time start and finish of goal-directed movements (Thaut, 2005), or control rhythmic movements by marking their cycle ends.

Interval-based sounds that repeat at a consistent pace can also be used as a guide for rhythmic movements. The preceding examples illustrate the utility of audio to enhance psychomotor task performance. In addition, audio can be used to influence a listener's affective states (Fahlenbrach, 2002; Karageorghis & Terry, 1997). For example, fast tempos and high pitches tend to evoke positive pleasant emotions, whereas slower tempos with lower pitches evoke negative somber emotions. Finally, audio can be used to drive cognitive task performance, such as by capturing user's attention (e.g., alerting a user to system malfunctions), decreasing user's workload, enhancing information exchange between user and system, and providing feedback to the user (Brewster, 1997; Day et al., 2004; Frauenberger et al., 2005; Guillaume et al., 2002). Such uses of audio to guide temporal task performance may be effective in enhancing training in combat situations where soldiers are required to carry out timely clearing of hostage situations, during which members of a 4-man team must follow the pace set by the group leader (US Army, 2003).

The spatial dimensions of audio can also be used to enhance human performance. Spatial audio has been used to enhance psychomotor performance, such as in target acquisition (Billinghurst et al., 1998) and target localization (Tannen et al., 2004) tasks. Spatialized audio can also influence one's affective state, such as through the use of spatialized sounds to enhance the emotional experience in movies (Baumgartner et al., 2006) and virtual environments (VEs) (Kim et al., 2004). Spatialized audio has also been shown to enhance cognitive performance, such as Wenzel's (1992) use of spatialized audio in cockpit displays to represent the bearing of targets to inform cognitive tasks such as path (re)planning and Jones' et al. (2005) use of spatialized audio to enhance search and detection tasks. Such uses of audio to guide spatial task

performance may also be effective in enhancing room clearing task performance, during which members must locate enemy and friendly units (US Army, 2003).

In using audio to enhance temporal and spatial tasks in combination, several issues arise, specifically - which interface sounds should be spatialized and which should not, and how to add spatialization cues to interface sounds. In terms of which interface sounds should be spatialized, as the human listener is only adept in spatializing three sounds, on average, at one time (Sulzen, 2001), it is herein suggested that when combining spatial and temporal auditory cues only the spatial auditory cues be spatialized and that the latter be limited to three *concurrent* spatialized cues. Our previous research has suggested that either spatial or non-spatial auditory earcons can be effective in training traversal pace (Ahmad et al., under review b), which is corroborated by other studies that have investigated the utility of audio cueing for training temporal pacing tasks (c.f., Karageorghis & Terry, 1997; Thaut, 2005). On the other hand, when training spatial tasks, the use of spatialization has been demonstrated to be key to enhancing performance. For example, Edwards et al. (2004) unsuccessfully attempted to use non-spatialized audio in an effort to enhance performance on spatial tasks (i.e., disassembling a system of parts, replacing a part, and then re-assembling the system; within an immersive VE). The implementation of audio cues did not result in improvements in terms of reducing task completion time or number of collisions. The authors' explanation was based on users' comments, which "clearly indicated that force feedback cues presented potentially more useful information than audio." As aforementioned, Loomis et al. (1998), Jones et al. (2005), and several others (Apostolos et al., 1992; Mulgund et al., 2002; Nelson et al., 2001) have demonstrated that spatialized audio can lead to substantial gains in performance time, among other benefits such as more natural

interaction and reduced workload. Thus, the first theorized design principle for combining spatial and temporal audio cues is:

> *When combining spatial and temporal audio cues, spatialize the auditory cues used to enhance spatial tasks – limiting this to three concurrent spatialized cues - and use non-spatialized cues to enhance temporal tasks.*

In terms of how to add spatialization cues to interface sounds, lessons learned from interruption management can prove useful (c.f., McFarlane, 2002; McFarlane & Latorella, 2002). When audio cues are being used to convey temporal information (such as pace; an ongoing task), and it is required to "phase in" spatial audio information (pertaining to locations of friendly and enemy units; an interruption task), the timing and content of this information is important. Tradeoffs exist between using audio cues to capture attention to a pending interruption task, and the preemptive nature of audio cues, which may worsen the performance on the ongoing task (Wickens et al., 2005), while improving faster switching to the interruption task (Ho et al., 2004). The resulting effect of interruption on task completion time is unclear and seems to depend on task complexity; i.e., generally with no effect on simple cognitive tasks, but a slowing down of complex tasks (Burmistrov & Leonova, 2003). The selection of the most suitable strategy depends on the nature of ongoing and interruption tasks, timing of the interruption, clarity of the interruption, perceived importance of the interruption, and user's workload prior to the interruption (Bailey et al., 2001; Banbury et al., 2003; Gillie & Broadbent, 1989). Successful interruption management entails; 1) varying the nature of ongoing and interruption tasks (e.g., from visual to auditory or from non-spatial to spatial audio), 2) using an interruption task of a simpler nature than the ongoing task, 3) providing partial information pertaining to the interruption task (e.g., nature, urgency and cognitive requirements), 4) giving the user some

control over when to switch to the interruption task, and 5) providing cues to the user with respect to the ongoing task, when it is resumed (e.g., place markers to indicate when and at what stage of the ongoing task switching to the interruption task should take place) (Bailey & Konstan, 2006; Gillie & Broadbent, 1989; Ho et al., 2004; Sawhney & Schmandt, 2000). This implies that when a user is engaged in a non-spatial audio task, to improve performance an interrupting audio task should be spatial, and these gains should improve with better audio spatialization fidelity (Bregman, 1990; Sawhney & Schmandt, 2000). Audio spatialization fidelity refers to how representative is a listener's perception of a virtual source compared to real world perception, the fidelity of which can be enhanced through the use of head-related transfer functions (HRTFs) that describe the changes in sound signal spectrum resulting from a listener's anatomy (Wenzel et al., 1993). HRTFs enable at least three levels of spatialization fidelity; 1) Generalized or default HRTFs, which are based on dummy measurements, 2) Individualized HRTFs, which require individualized measurement using tiny microphones placed in a listener's ears, and 3) "best-fit" HRTFs, which entails personalizing an HRTF from a database (for example the CIPIC HRTF database; Algazi et al., 2001). "Best-fit" HRTFs enable customizing the spatialization for a listener without going through the time-consuming invasive process of HRTF measurement. In terms of fidelity, generalized HRTFs are at one end of the spectrum, while individualized HRTFs are at the other end, and "best-fit" HRFFs are somewhere in the middle depending on the effectiveness of the personalization procedure. If selected properly, "best-fit" HRTFs are expected to enhance localization performance and provide close to "natural" spatial sound experiences (Wenzel et al., 1993; Wightman & Kistler, 1989). Hence, when a user is engaged in a non-spatial task, generalized HRTFs are expected to result in the least gains in performance, while individualized or properly selected "best-fit" HRTF are

expected to result in the highest gains in performance. In addition, with respect to an auditory-based interruption, more exposure to the interruption task should improve its clarity and result in better performance (Banbury et al., 2003).

Thus, additional theorized design principles for combining spatial and temporal audio cues include:

1) Use higher audio spatialization fidelity to improve overall performance in terms of faster task completion.

2) Use cognitively simple spatial and temporal tasks to enhance performance during interruption.

3) Enable the user to control when to switch between tasks.

4) Provide information to the user with respect to the nature of a pending interruption.

Research Hypothesis

Design principles for conveying concurrent temporal and spatial information in virtual reality training environments can be theorized based on the above review; these include:

1. Non-spatial audio cues will be effective in guiding participants to control their pace.
2. Spatialized audio cues will result in faster overall task completion time when compared to non-spatial audio cues, and the improvement will depend on spatialization fidelity. Specifically;
    - Generalized (default) HRTFs will result in faster completion time than non-spatial audio cues.

- "Best-fit" HRTFs will result in faster completion time than either non-spatial audio cues or generalized (default) HRTFs.

3. When combining temporal and spatial information, the best gains in performance will result from using simple tasks, allowing user control over switching between tasks, and supplying information with respect to pending tasks.

Method

This study utilized a virtual reality training system to: 1) perceptually train participants on a pace setting task, and 2) assist participants in locating enemies and friendlies using spatial audio. Figure 12 provides a conceptual framework for the study. As depicted in Figure 12, each participant performed pre-training, post-training, and a set of eight training sessions. Performance was assessed by:

- Comparing performance on training session 8 (audio cues present) to pre-training (no training audio cues present) to assess audio cueing efficacy.
- Comparing performance on post-training (no audio cues present) to pre-training (no training audio cues present) to assess training internalization.

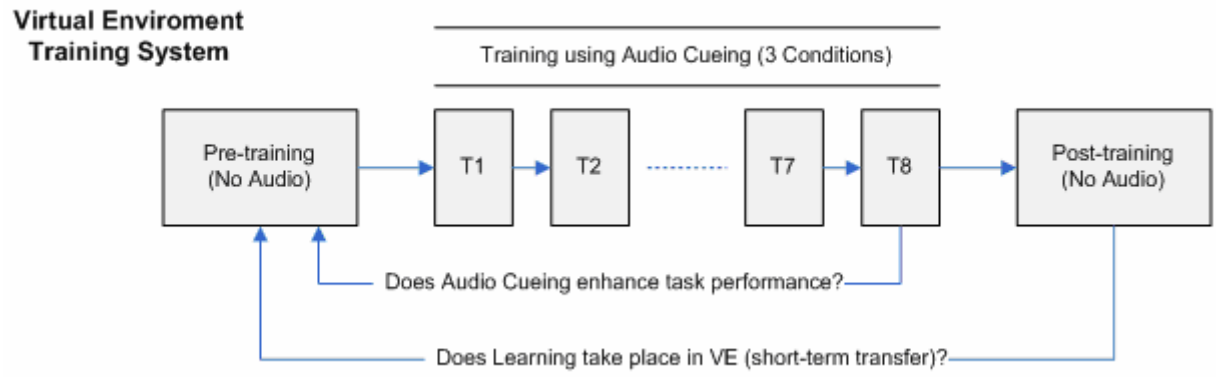

Figure 12: Integration Study Training Model

Participants

Thirty students from the University of Central Florida (mean age = 21.47 years; s.d. = 3.1 years; 6 females and 24 males) participated in this study. Participants were randomly assigned to one of three different treatment conditions. All participants reported normal hearing. Twenty-four participants were right-handed, four were left-handed (including 1 female), and two males were ambidextrous. Participants voluntarily agreed to participate in the experiment for class credit.

Apparatus

The experimental setup consisted of one-dual-processor Dell Dimension 9200 computer and one-Pentium 4 Dell Dimension 8500 computer. The Dimension 8500 was used to generate graphics for a training task with ManSim software and the Dimension 9200 was used to produce audio cues using ViBeStation software. Participants interacted with the training task through an

eMagin 3DVisor Head Mounted Display (HMD). Audio cues were presented using Sennheiser headphones. Head tracking was done using an Intersense InertiaCube tracker. Participants navigated through the task through a standard Saitek game controller.

The participants assigned to a "best-fit HRTF" condition (see Experimental Design, below), used a profiler tool based on the approach suggested by Seeber and Fastl (2003). The tool allows users to choose five candidate HRTFs and audition each of them while changing the sound source position to one of eight predetermined locations around their head. Figure 13 demonstrates the "best-fit" HRTF profiler tool.

Figure 13: "Best-fit" HRTF Profiler Tool

Several questionnaires were used in this study including, the verbalizer-visualizer (Richardson, 1977), immersive tendencies (Witmer & Singer, 1988), presence (Witmer & Singer, 1988), NASA Task Load Index (TLX) workload (Hart & Staveland, 1988), and the simulator sickness (SSQ; Kennedy et al., 1993) questionnaires. The verbalizer-visualizer questionnaire is scored on a scale from 1-15; as the participant's score on the questionnaire goes higher, the more indication that the participant is a visualizer. The immersive tendencies questionnaire has 39 questions on a scale from 1-7. As the participant's average score goes

higher, the more indication that the participant has a tendency to become immersed. The presence questionnaire has 37 questions on a scale from 1-7. As the participant's average score goes up, the higher the participant's reported sense of presence. The NASA-TLX consists of six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. For each scale, individuals rate the demands imposed by the task, as well as each scale's contribution to total workload, the latter of which is calculated by summing the product of each scale's rating and weight. Total workload has possible scores ranging from zero to 60. The SSQ has participants report the degree to which they experience a set of symptoms as one of "None," "Slight," "Moderate," or "Severe," which are then combined into a Total Sickness score. The SSQ has possible scores ranging from zero to 235. In addition, the participants completed two spatial aptitude tests: Map Planning and Cube Comparison (Ekstrom et al., 1976).

Virtual Environment

The VE was designed to mimic a room clearing exercise and included a 15 room building to be cleared. Ten different variations of the environment at comparable task difficulty were created to be used for training and testing. One variation was used for task familiarization, one for pre-training and post-training tests, and one each for eight training sessions; the assignment of the training environments order was randomized among participants. See Appendix B for a screen shot of the VE used in this study. Each environment contained eight enemy entities, four friendly entities, four mouse holes, and six M16 objects to collect. The number of enemy and friendly units in each room was limited to three to match a human listener's spatial auditory capacity (Sulzen, 2001).

In a previous work by the authors (Ahmad et al., under review a), three theoretical models were developed to guide the addition of audio cues to interactive applications such as VE training systems. These models established the foundations of the present study as follows:

- The audio integration model aided in specifying the desired training performance objectives, which were: 1) Cognitive, information related to the rooms to clear and presence of friendly and enemy units, and 2) Psychomotor, information related to pace setting.
- The temporal audio model provided guidance on when to add sound within the VE, more specifically, for the cognitive performance objective, instantaneous sounds were implemented to indicate the presence of various enemy and friendly units, and for the psychomotor performance-objective, arrhythmic interval-based sounds were added to indicate the need to speed up or slow down.
- The spatial audio model provided specification on how to add spatialized cues to selected interface sounds, which were implemented using HRTFs. In the present study, both generalized and "best-fit" HRTFs were used to add spatialization cues to the instantaneous sounds related to presence of enemy and friendly units.

Figure 14 shows an example environment layout. Audio cues were used for enemy and friendly voices, pace setting, and M16 objects. The audio cues-enabled environments were then presented using different spatialization fidelities (see experimental design, below). Pace setting used two diotic metaphoric audio cues to guide participants in setting their pace. If the participant was traversing the hallway at the "correct" pace, no audio was played. If the participant needed to slow down, a drum sound was played. If the participant needed to speed up, a flute sound was played. The rate of the played audio was proportional to the deviation

from the predetermined desired pace. All environments had "user's foot steps" and "gun shot" sounds implemented. One environment variation was randomly selected for pre and post-testing. The pre and post-testing environment had only non-training audio cues, which included "user's foot steps" and "gun shot" sounds. Pre- and post-testing were used to assess the transfer of training within VE task performance when auditory cues were removed (see Figure 12). The eight training environments had different spatial audio fidelity implementations depending on the experimental conditions, which were randomly assigned to each participant;

- Non-spatial: No audio cues for training were spatialized, i.e., all sounds were played equally loud at both ears and at the same time. The audio cues implemented were enemy and friendly voices, pace-setting sounds, and M16 objects sounds. In addition, "user's foot steps" and "gun shot" sounds were implemented.
- Spatial using "default": Enemy & friendly voices and M16 object sounds were spatialized using dummy HRTF measurements, i.e., additional spatial cues were added to the sounds to create a virtual source image at a particular point in space. The pace-setting sounds, "user's foot steps" and "gun shot" sounds were not spatialized.
- Spatial using "best-fit HRTF": Enemy and friendly voices and M16 object sounds were spatialized using a subjectively selected "best-fit" HRTF, i.e., additional cues were added to the sounds to create a virtual source image at a particular point in space. The pace-setting sounds, "user's foot steps" and "gun shot" sounds were not spatialized.

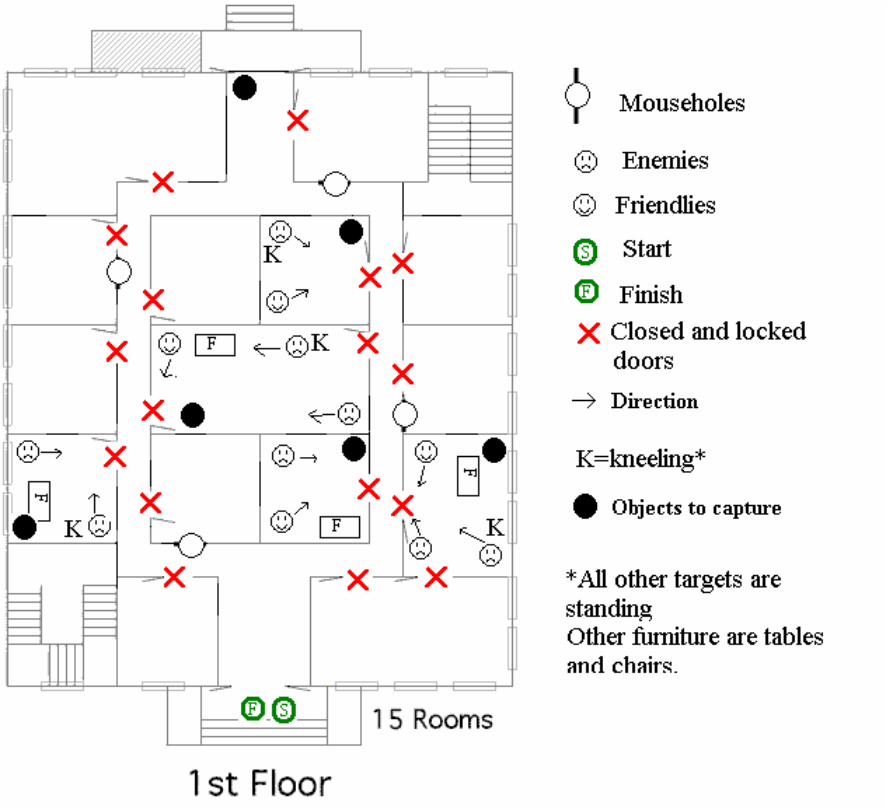

Figure 14: Virtual Environment Layout Example

Tasks

Each participant performed a series of Close Quarters Battle for Military Operations in Urban Terrain (CQB for MOUT) activities in a virtual environment. There were two primary tasks that the participants were expected to complete. The first task was to open, enter and clear rooms and engage all enemy and friendly units located therein. To engage units, participants had to point their weapon towards a unit, and then use the correct button on the game controller to identify them as either friendly or foe (i.e., left controller button was used to clear friendly units, and right controller button was used to fire upon foe units). The participants were also required

to collect a set of M16 objects scattered throughout the environment at random locations (using the left controller button). The second task was to maintain a consistent pace while traversing the environment hallways. Two predetermined paces were selected based on the environment configuration and dimensions, these were a fast pace when passing mouse holes on walls (1.2 – 1.35 m/s), and a medium pace (0.4 – 1.0 m/s) when neither doors nor mouse holes were present. When moving through the environment, turning was controlled through head movements. Locomotion (i.e., stepping forward and back) was controlled using the game controller. The tasks were of simple cognitive nature (i.e., a maximum of 3 units to clear in each room and separating the pacing task from the spatial task by not placing any friendly or enemy units to be cleared in the hallways). In addition, the following were considered to facilitate testing the proposed design guidelines:

- The participant controlled when to switch between the temporal and spatial tasks by deciding when to: 1) open and enter rooms, and 2) exit the rooms to the hallway.
- Partial information pertaining to the pending spatial task was provided in terms of audio cues at different spatialization fidelity levels that indicated how many units to clear in each room (i.e., each friendly or enemy unit had an associated audio event associated with its location in space).
- Pace setting cues were present at all times when the participant was walking through the hallways.

Experimental Design

This study utilized a pre-post between subjects one-factor ANOVA design. The one factor was audio condition, with three different levels: 1) non-spatial audio cues, 2) cueing using spatial audio (default HRTF), and 3) cueing using spatial audio ("best-fit" HRTF). Performance was assessed using the total time required to complete the simulation to assess spatial task performance and the average deviation from the predetermined desired traversal pace for the temporal pacing task. Both audio cueing effects on task performance (comparing last training to the pre-training test) and near term transfer of skills (comparing post-training to the pre-training) were evaluated. In addition, workload, presence, and simulator sickness were assessed.

Procedure

Before the start of a test session, participants completed an informed consent, demographics questionnaire, immersive tendencies questionnaire and verbalizer-visualizer questionnaire. In addition, participants completed the cube comparison and map planning aptitude tests. After that, participants completed task familiarization. In the first part of the task familiarization, participants reviewed a PowerPoint presentation that explained the goals of the MOUT task, how to use the controller, and an illustration of the different audio cues used in the environments (pre-training test, post-training test, and eight training sessions, see Figure 12). During the second part of the task familiarization, participants completed hands-on training and experienced how to navigate through the environment and how to clear enemy and friendly units. Once the familiarization was done, participants completed a pre-test, where they were required to complete clear a room while controlling their pace with no audio cueing (only footsteps and gun

shot sounds were present). Then, participants were randomly assigned to one of the three different spatial audio conditions under which they completed eight training sessions. Each training session involved completing the MOUT room clearing task while controlling pace. The order of the training sessions was randomized for each participant. After the training sessions, participants completed a post-test, which used the same VE and settings as the pre-test (again, only footstep and gun shot sounds were present). All participants wore the HMD and headphones during pre-training, all training sessions, and post-training. In addition, participants completed the SSQ and NASA TLX index after pre-testing, training session number 1, training session number 8, and post-testing. Once testing was complete, participants responded to the presence questionnaire. Finally, participants were provided with a written and oral debrief about the integration experiment.

Results

Performance data in terms of participants' location, speed, and time were logged through the experimentation software. Participant log files were processed for total time to complete the tasks and average deviation from predetermined desired pace. Subjective questionnaire data were manually input.

Time and Pace Performance Results

Table 8 provides descriptive statistics on the total time to complete tasks and average deviation from predetermined desired pace.

Table 8: Time and Pace Descriptive Statistics

Trial	Condition	Average Time (min)	Average Pace (m/s)	Average Deviation (m/s)
Pre-test	Best-fit	4.81	1.04	0.32
		(1.12)	(0.18)	(0.07)
Pre-test	Default	5.39	0.91	0.27
		(1.47)	(0.07)	(0.03)
Pre-test	Non-Spatial	5.10	0.94	0.27)
		(2.74)	(0.18)	(0.06)
Training 8	Best-fit	3.55	0.95	0.22
		(1.00)	(0.12)	(0.07)
Training 8	Default	3.52	0.99	0.23
		(0.75)	(0.17)	(0.11)
Training 8	Non-Spatial	3.78	0.92	0.21
		(0.80)	(0.12)	(0.10)
Post-test	Best-fit	3.30	1.00	0.27
		(0.70)	(0.13)	(0.07)
Post-test	Default	3.05	1.06	0.29
		(0.93)	(0.15)	(0.08)
Post-test	Non-Spatial	3.32	0.98	0.25
		(1.02)	(0.19)	(0.12)

Figures 15 and 16 depict the average pace and time to complete tasks, respectively. A repeated measures ANOVA was used to compare the average deviations from the predetermined desired pace and average performance time among the various audio conditions. ANOVA results show significance for the within subject variable training session for the deviation from predetermined desired pace ($p < 0.01$) and time to complete tasks ($p < 0.001$). It also shows that the between subject spatial fidelity level variable had no significant effects on the deviation from predetermined desired pace nor the time to complete tasks (both $p > 0.9$). There were no significant interaction effects between training session and spatial fidelity level ($p > 0.8$). The effect sizes for the time to complete tasks were; 1) Default HRTF compared to non-spatial = -0.315, 2) “best-fit” HRTF compared to non-spatial = -0.273, and 3 “best-fit” HRTF compared to Default HRTF = 0.042. All these indicate a low influence of the spatial fidelity level on task

completion time. Since spatial fidelity did not show overall significance, no post-hoc analysis was performed for this variable.

On average, the deviation from predetermined desired pace for training session number 8 (where audio pacing cues were present) was significantly less than that for pre-test ($p < 0.05$) by 0.065 (s.d.= 0.023) or 22.73% less on average, and less than that of post-test ($p < 0.05$) by 0.038 (s.d.=0.016) or 14.67% less on average. These results suggest that non-spatial audio cues appear to be an effective strategy for pace setting. In addition, the time to complete tasks for; 1) training session number 8 was significantly less than that of pre-test ($p < 0.01$) by 1.65 (s.d. = 0.29) or 31.41 % less on average, 2) post-test was less than that of pre-test ($p < 0.01$) by 1.98 (s.d. = 0.29) or 37.76% less on average, and 3) post-test was less than of training session 8 ($p < 0.01$) by 0.33 (s.d. = 0.11) or 9.25% less on average. These results suggest that there may be significant learning effects with respect to task completion times. There were no significant correlations between time to complete tasks and scores on the Cube Comparison and Map Planning spatial aptitude tests ($p > 0.5$ for training session number 8 and post-test, and $p > 0.085$ for pre-test), see Table 9.

Table 9: Correlation between Time to Complete Tasks and Score on Spatial Aptitude Tests

Time to complete	Pre-Test	Training Trial 8	Post-Test
Cube Comparison Test Score	-0.358	-0.111	-0.067
Map Planning Test Score	-0.361	-0.07	-0.021

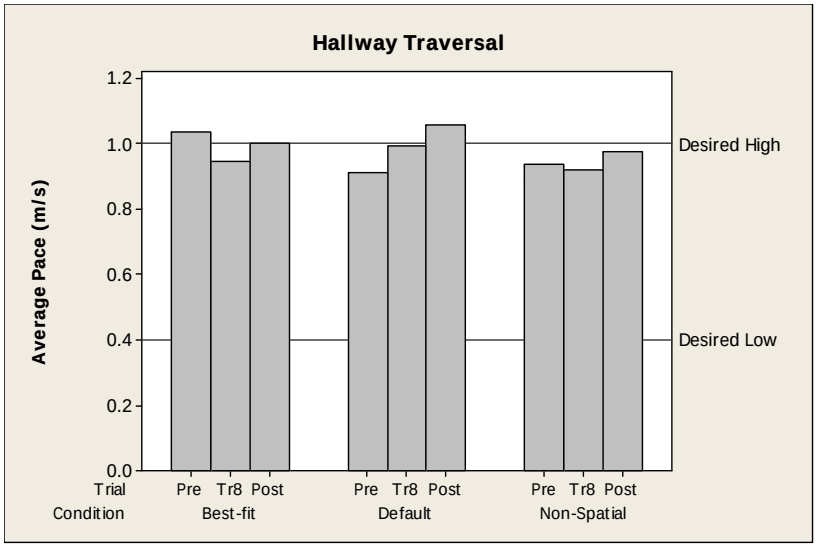

Figure 15: Average Traversal Pace

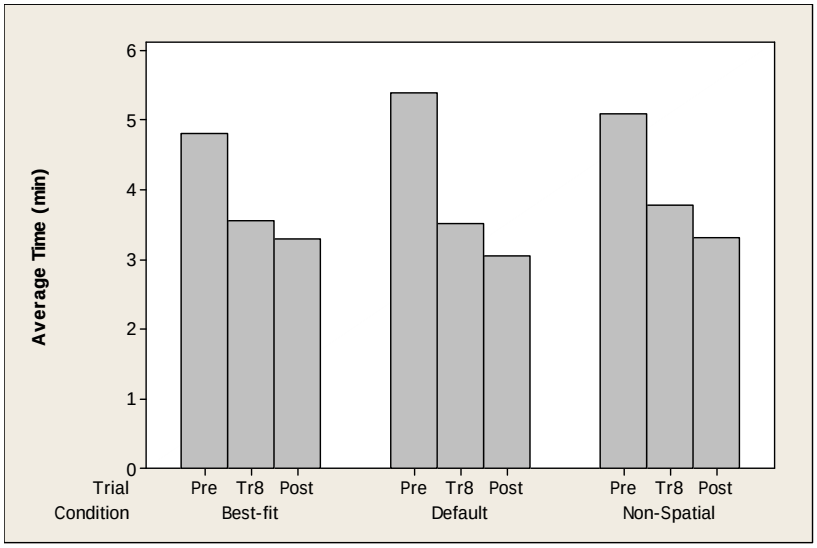

Figure 16: Average Time to Complete Tasks

Subjective Questionnaires' Results

The average score on the Cube Comparison test was 17.23 (s.d. = 13.40) and on the Map Planning test was 24.57 (s.d. = 7.62). The average score on the Visualizer-Verbalizer

questionnaire was 7.37 (s.d.= 2.04); 9 participants had a score above 8 (visualizers), 7 participants had a score of 8 (neutral), and 14 participants had a score below 8 (verbalizers). The average score on the immersive tendencies questionnaire was 4.60 (s.d. = 0.65). The average score on the presence questionnaire was 4.43 (s.d. = 0.69). The Pearson Correlation between the immersive tendencies and presence questionnaires average scores was 0.3 ($p < .05$). A one-way ANOVA was used to compare presence scores among the various audio spatialization conditions, which showed significance ($p < 0.02$). Tukey's post-hoc comparison showed that non-spatial audio resulted in higher presence scores than spatialization using the default HRTF, which was statistically significant (mean difference = 0.85, s.d. = 0.27, $p < 0.02$). One possible explanation for this finding may be possible network delays between the tracker, visual computer and audio computer. This warrants further investigation, as any slight delay may indicate that the audio implementations were not perceived as realistic by participants. All other differences were not significant.

The average participants' total sickness scores on the SSQ were:

- Pre- training test → 26.80 (s.d. = 32.77).
- Training session number 1 → 32.91 (s.d. = 35.31).
- Training session number 8 → 43.63 (s.d. = 46.80).
- Post-training test → 39.77 (s.d. = 46.82).

These values suggest that participants experienced a relatively high level of simulator sickness (about the 95 percentile, as reported by Stanney, Graeber, & Kennedy, 2005) throughout the experiment; however, no significant effects were found for audio condition on average total sickness scores ($p > 0.05$; for pre-test, training session 1, training session 8, and post-test). The implication of the reported simulator sickness is that care should be exercised when using virtual

training simulators to ensure participants are in good health before they leave the simulator facility.

In addition, participants completed the NASA TLX questionnaire to assess subjective workload. The participant total scores on the NASA TLX questionnaire were:

- Pre-training test → 25.92 (s.d. = 8.31).
- Training session number 1 → 27.08 (s.d. = 8.93).
- Training session number 8 → 22.88 (s.d. = 11.24).
- Post-training test → 19.28 (s.d. = 12.48).

A repeated measures ANOVA was used to compare the average workload scores. The results revealed that the total TLX scores were statistically dependent on trial order ($p<0.001$), but they did not depend on audio condition ($p>0.4$). The interaction between trial order and audio condition was not significant ($p>0.2$). Specifically, training session 8 had an 11.71% lower total TLX score, on average, compared to pre-test (mean difference = 3.03, s.d. = 1.35, $p < 0.05$), post-test had a 25.60% lower total TLX score, on average, compared to pre-test (mean difference = 6.63, s.d. = 1.77, $p< 0.01$), training session 8 had a 15.51% lower total TLX score, on average, compared to training session 1 (mean difference = 4.20, s.d. = 1.57, $p< 0.02$), post-test had a 28.80% lower total TLX score, on average, compared to training session 1 (mean difference = 7.80, s.d. = 1.84, $p< 0.001$), and post-test had a 15.73% lower total TLX score, on average, compared to training session 8 (mean difference = 3.60, s.d. = 1.20, $p< 0.01$).

Discussion

The present study examined the utility of using audio to train spatial and temporal pacing tasks in an interactive virtual training system for MOUT CQB. In particular, three audio spatialization fidelity conditions were compared to assess the utility of using spatial auditory cues to enhance spatial task performance, while using non-spatial audio cues to guide trainees in maintaining a set pace.

Research Hypothesis One: Non-spatial audio cues will be effective in guiding participants to control their pace.

The results of the experiment revealed that non-spatial auditory cues were effective in guiding participants in controlling their pace, even with the presence of spatial guidance sounds. This is in agreement with past literature that suggests both non-spatial and spatial auditory cues can be used for pace setting in virtual training systems (c.f., Ahmad et al., under review b, which extends Karageorghis & Terry, 1997; Thaut, 2005). The present study sheds further light on the robustness of non-spatial audio cues to support pace training, as they were found to be effective even in the presence of potentially distracting auditory cues. The implications of this finding include; 1) the decisions related to temporal and spatial information presentation using audio are potentially a sequential process, whereby the timing of cues is selected and then the spatial nature of those cues is determined, which provides support to the theoretical models developed in Ahmad et al. (under review a), and 2) supports the extension of interruption management theories on maintaining performance on an ongoing temporal task after interruption by a spatial task, if sequenced properly (Gillie & Broadbent, 1989; Ho et al., 2004). The results of the present study provide support to Ahmad's et al. (under review a) proposed process for

integrating audio in interactive application, which starts with making decisions pertaining to timing audio cue presentation based on desired performance objectives and then choosing to add spatialization cues if required in a particular application. In terms of interruption management, the present study showed that the performance on an ongoing pacing task could be maintained even when several spatial interruptions took place.

Research Hypothesis Two: Spatialized audio cues will result in faster overall task completion time when compared to non-spatial audio cues, and the improvement will depend on spatialization fidelity.

The present study did not provide support for the utility of spatial auditory cues to influence faster task completion, which is at odds with Loomis et al. (1998). Although, this is an unexpected finding, other studies conducted in similar environments, such as Jones et al. (2005) and Milham (2005), did not show significant spatial auditory effects. This may be due to the complexities involved with spatializing auditory signals, or to the limitations of the technology used to present these cues. Specifically, it was expected that using different spatialization fidelities, if produced effectively to distinguish the interrupting spatial task from the ongoing temporal task (i.e., with effective task separation), and would result in improved overall task performance (Burmistrov & Leonova, 2003; Gillie & Broadbent, 1989). The current implementation was not found to be effective, which may be due to the participants' tendency to enter each room along the route instead of utilizing the spatial audio cues that provided guidance on the presence of friendly and enemy units in each room. Some possible explanations for the participant's behavior are; 1) The participants may not have maintained urgency with regard to fast completion of tasks, 2) The environment layout, which is a set of sequential rooms may have encouraged room-by-room search, and 3) The participants may not have been able to deduce

from the spatial audio cues enough information with respect to the location of the various units. If participants experienced an inability to effectively utilize the spatial audio cues, this may suggest technological limitations with regard to using HRTFs to produce virtual sound source images. This was further supported by participants' subjective feeling of presence, where non-spatial audio resulted in higher presence scores as compared to spatialization using default HRTFs, which could be as indicated earlier due to possible network delays between computer systems. Further investigation is warranted to clarify the effectiveness of using spatial audio cues to enhance performance and presence in virtual training systems.

Research Hypothesis Three: When combining temporal and spatial information, the best gains in performance will result from using simple tasks, allowing user control over switching between tasks, and supplying information with respect to pending tasks.

The current study provided partial support to hypothesis three, as participants were able to maintain their performance on the pacing task while being interrupted by the spatial room-clearing task. Nevertheless, the study did not show improvement in task completion times, as discussed above. Moreover, there was no significant interaction effect found between the different spatialization strategies and the performance on the pacing. The resulting non-significant interaction between spatial and temporal task performance provides support to the models developed in Ahmad et al. (under review a). In these models, the integration of audio cues follows a sequential process, i.e., first the decisions pertaining to the timing of audio cues are made, then the decisions pertaining to spatialization are made, which assumes no interaction between these decisions. The results support such a process. The present VE design utilized simple tasks (e.g., only three units to clear in each room), allowed the user to decide when to switch between tasks (spatial and temporal tasks were separated by closed doors), and provided

information about pending interruption tasks (i.e., audio cues at different spatialization were used to indicate the presence of various friendly and enemy units). Despite these implementations, only partial support was provided to hypothesis three, which extends the theories related to the dominance of auditory cues when it comes to temporal processing (c.f., ETSI, 2002). This finding implies that, under the current state of technology, audio cues alone might not be sufficient to present spatial information within VE training systems, hence possibly augmenting the presentation with additional visual and haptic cues may produce better training outcomes (Begault, 2000).

The current study revealed significant learning effects in terms of task completion time, with participants performing tasks faster over time. This may indicate that the training environments used were simplistic to reveal significant spatial auditory cueing effects. The spatial auditory cues were expected to guide participants to which rooms they should enter (i.e., opening the doors only for the rooms that contained friendly and enemy units and passing by rooms that were empty). During the study, however, participants may have ignored these spatial audio cues and cleared the MOUT environment room-by-room. This may have allowed participants to gain experience as they did the training tasks, thereby leading to performance improvements across trials. The workload results suggest that perceived workload was less over training trials, which may further indicate a learning effect across trials. This learning effect may have been exacerbated by the environment designs, which used the same environment layout and location of rooms, while distributing enemy and friendly units differently (the latter changed the nature of the audio cues from one environment to the next). It is recommended that future studies vary both the environment layout and nature of auditory cues to address this learning effect.

There were two findings with respect to perceived workload; 1) The workload decreased over time spent on trials, and 2) The workload was higher for the audio-enabled training environments, and that increase in workload did not depend on the spatialization fidelity level implemented within these environments. The decrease in perceived workload over training trials may also have contributed to the maintained performance on the temporal pacing task as indicated by Bailey et al. (2001). The reduction in perceived workload could be due to the participants getting used to the environment, which, based on Bailey et al. (2001), can create more opportunity for the user to switch to pending interruption tasks without incurring losses in performance. There are two additional implications of these workload findings; first, the training environments, which contained audio cues resulted in higher workload than the pre and post-test environments, which did not contain these cues. This implies that tradeoffs may exist between achieving desired training benefits via audio cues vs. the increase in trainees workload, which is consistent with the workload results reported in Ahmad et al. (under review b). Secondly, this increase in workload did not depend on the types of audio implementation (within the environment used in the present study). This implies that it is important to assess the potential impact on workload when deciding which audio cues to incorporate in a training system regardless of the type of cues used.

An interesting finding from this study was that there were no significant correlations between the time to complete tasks, which was used to measure spatial task performance, and scores on the Cube Comparison and Map Planning spatial visual aptitude tests. As these tests measure the ability to maintain spatial orientation with objects in space and the speed of scanning a spatial field, respectively (Ekstrom et al., 1976), one would have expected higher scores on these tests to correlate with the performance on a spatial task guided by spatialized

audio cues. This suggests the need for developing new aptitude tests that may correlate better with spatial auditory systems, as the visual-spatial tests do not seem to generalize to spatial audio.

There are several implications to this study with regard to integrating auditory cues in interactive environments to train spatial and temporal tasks, these include:

- When combining auditory cues to enhance performance on spatial and temporal tasks, non-spatial auditory cues can be used effectively to enhance traversal pace.
- The integration of auditory cues can follow a sequential process; where first, decisions pertaining to timing of cues are made, followed by decisions pertaining to the spatialization of cues.
- Using audio alone may not result in improvements in spatial task performance and hence may benefit from supplementation with visual and haptic cues.
- When designing VE training systems, it might be beneficial to use a complex environment to guard against task learning and focus on the desired training objectives.

In addition, there are two other insights that can be gleaned from this study:

- An effort should be made to ensure the wellness of users of VE training system before they leave the simulator facility, as they may experience substantial simulator sickness during exposure.
- There is a need to develop new aptitude tests that may correlate better with spatial auditory systems.
- When using audio cues in a virtual training system, consider the tradeoffs that exist between achieving training outcomes vs. increased participants' workload that be associated with the use of audio cues.

Conclusions

This paper presents the results of an experimental evaluation on the utility of using auditory cues to train spatial and temporal tasks in interactive virtual environments. Study participants completed pre- and post- tests and eight training sessions where different audio spatialization fidelities were used to train spatial tasks while maintaining a consistent pace. Auditory cueing was found to be effective for pace setting, but no significant findings were found for task completion times or for the interaction between spatial and temporal training. This demonstrates the need for further exploration on how to best integrate both spatial and non-spatial auditory cues in virtual training systems to enhance training outcomes. The overarching finding of the present study is that selection of spatial and temporal information presentation schemes using audio can follow a sequential process within VE training systems. This finding can be of importance to system designers and human-system integration researchers who are investigating how to integrate additional modalities such as audio in the next generation training systems.

Acknowledgements

This material is based upon work supported in part by the Office of Naval Research under its Virtual Technologies and Environments (VIRTE) program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views or the endorsement of ONR.

CHAPTER FIVE: GENERAL DISCUSSION

This dissertation aimed to establish foundations for a user-centered auditory design science. The overarching aim of this work is to put forward a framework for integrating sound cues in interactive applications, coupled with a set of empirically validated design principles. The research was broken into; 1) building comprehensive theoretical frameworks for integrating audio cues in interactive applications, 2) conducting a first empirical study on the utility of audio cues to train pace in a virtual reality training system, and 3) conducting a second empirical study to evaluate using audio to present both temporal and spatial information.

Until recently, audio cues have been under-utilized in interactive applications. This could be due to the lack of theoretical guidance of how to best integrate such cues. The present research established three theoretically-driven models for adding audio cues to interactive applications that include; 1) an audio integration model that addresses the end-to-end decision making process for integrating auditory cues in interactive applications (Chapter 2, Figure 1), 2) a temporal audio theoretical model that addresses considerations pertaining to the timing of presenting auditory cues (Chapter 2, Figure 2), and 3) a spatial audio theoretical model that addresses the spatialization of sounds (Chapter 2, Figure 5). The face validity of the developed models was assessed using an SME SWOT analysis, the results of which are displayed in Chapter 2, Table 3. These models bring the audio design science closer to previous developments in spatial visualization design (Buagajska, 2003) and perceiving layout (Cutting & Vishton, 1995). The empirical evaluation studies focused on validating; 1) the sequential decision making process, where first the timing of audio cues presentation is selected, then making the decisions pertaining to spatializing these cues (hence, temporal and spatial

information conveyance decisions can be made using a sequential process), and 2) the potential advantages of the temporal aspects of audio to train psychomotor tasks in VE training systems.

The first empirical study was based on the temporal audio model (Chapter 2, Figure 2) and focused on training temporal pacing tasks. The results of the first study provided preliminary evidence on utility of audio cues in training traversal pace in virtual reality training systems. The results of this study revealed that using a metronome was not effective in training participants to maintain a desired pace, which is at odds with past literature that suggests metronomes can be used for guiding rhythmic movements in the real world (c.f., Kern et al., 1992; Kurtz & Lee, 2003). This implies that strategies that were effective for real world training may not extend well to virtual reality based training systems. On the other hand, using earcons to influence (increase/decrease) traversal pace gradually was supported. Both the non-spatial and spatial audio conditions resulted in less deviation from a predetermined desired pace as compared to the no audio condition and these deviations were statistically significant. This supports the literature related to using audio cues to drive pace setting (c.f., Karageorghis & Terry, 1997; Thaut, 2005). Also, the results of the first study suggested that audio-based pace setting in VEs may become internalized, hence can be used when the audio is removed. This extends the literature describing the internalization of pace skills to VE training systems (c.f., Kolers & Brewster, 1985; Libkuman et al., 2002); previous research had shown that when audio cues are used to establish a psychomotor rhythm in the real world, people were able to maintain their rhythm even once the audio cues were removed. The current study provides evidence that this internalization of pace holds true when training pace in virtual environments as well. Future studies should evaluate if pace, once trained in a VE, can be transferred to the real world. The

findings of the first study show support to the utility of using the temporal dimensions of audio (Chapter 2, Figure 2) to achieve psychomotor performance objectives (Chapter 2, Table 1).

The second empirical study examined using audio to train both temporal pacing and spatial room-clearing tasks concurrently (Chapter 2, Figures 2 and 5). The results of the second study revealed that non-spatial auditory cues were effective in guiding participants in controlling their pace, even with the presence of spatial guidance sounds. This is in agreement with past literature that suggests both non-spatial and spatial auditory cues can be used for pace setting in virtual training systems (Chapter 3, which extends Karageorghis & Terry, 1997; Thaut, 2005 both of which focused on real world training). The second study also demonstrated that audio cues were successful in setting pace in a VE training system even while other sounds were present. This supports the theoretical models developed in Chapter 2, Figure 2 in terms of temporal information presentation using audio.

The second study did not provide support for the utility of spatial auditory cues to influence faster task completion, which is at odds with Loomis et al. (1998). Although, this is an unexpected finding, other studies conducted in similar environments, such as Milham (2005), did not show significant spatial auditory effects. Some possible explanations for the unexpected findings are; 1) The participants may not have maintained urgency with regard to fast completion of tasks, 2) The environment layout, which was a set of sequential rooms, may have encouraged room-by-room search rather than reliance on the audio cues, and 3) The participants may not have been able to deduce from the spatial audio cues enough information with respect to the location of the various units, thus suggesting the need for enhanced HRTFs.

There also were no significant interaction effects found between the different spatialization strategies and the performance on the pacing task, this may indicate that decisions

regarding the auditory presentation of spatial and temporal information can be made using a sequential process within VE training systems. The lack of a significant interaction effect provides support to the sequential process for audio cue integration illustrated in Chapter 2, Figure 1, where first decisions related to timing are made and then decisions related to spatialization are made.

Both empirical evaluation studies resulted in a set of validated design principles that can be added to Chapter 2, Table 1. These design principles include:

- Audio cues can be used to influence traversal pace in VE training systems, and this holds true when a) using both spatial and non-spatial earcons, and b) when other spatial and non-spatial cues are present.
- Using audio alone may not result in improvement in spatial (e.g., location of units in a VE) task performance, and hence may benefit from supplementation with visual and haptic cues.

The limitations of this dissertation are particular to the findings from the empirical evaluation studies, where 1) undergraduate student participants were involved, 2) no appreciable time was allowed between training sessions and post-training test to evaluate retention of acquired skills, and 3) possible network delays existed between the computer systems used in the studies. The nature of study participants may have biased the study findings, as undergraduate students may not have perceived task criticality in the same manner as other individuals, such as real world soldiers, with respect to maintaining pace and fast completion of tasks (USArmy, 2003). Also, since there was no appreciable time between training and testing, there is no means to assess the retention of task training within the current data, which is important for evaluating the effectiveness of training. Possible network delays may have resulted in participants

perceiving spatial audio cues as being unrealistic and hence not achieving “true” immersiveness. Despite these limitations, the present studies showed significant effects for audio cueing for pace setting in desktop VE systems, and hence provide initial guidance for future studies that should address the limitations with respect to; 1) using more representative trainees, 2) evaluating retention of task training by allowing appreciable time between training and testing, and 3) ensuring that network delays are below the levels that can result in unrealistic spatial audio perception by participants.

The results and discussion of the empirical studies provide practical guidance on how best to integrate audio cues in virtual training systems, and potential tradeoffs that may exist, thus establishing foundations of a much needed user-centered audio design science.

CHAPTER SIX: CONCLUSIONS AND FUTURE WORK

This research resulted in building comprehensive theoretical models for audio integration in interactive systems, and conducted two studies to empirically evaluate selected aspects of the developed models. The theoretical models provide for the first time a process for audio integration, and a cataloging system for temporal and spatial audio design principles. The empirical results from the studies support the inclusion of audio cues in virtual reality training environments to train traversal pace. The results of the empirical evaluation provide additional validated design guidelines that further augment the theoretical models. The results of this work support the research needs of the National Research Council (2000, p.25), as the developed models provide guidance on how to integrate spatial and temporal audio cues within multimodal systems.

Several possible research directions are identified based on the results of the work presented in this dissertation; these include:

- Using the spatial audio theoretical model to drive the specification and design of HRTFs.
- Conducting further empirical evaluation of the utility of using audio cues to provide guidance with respect to spatial location of entities within VEs.
- Developing the temporal model further in terms of providing more details on how to influence multiple human performance objective using audio cues.
- Further expanding on techniques for effectively integrating audio to support both spatial and temporal tasks when performed concurrently.

LIST OF REFERENCES

Ahmad, A., Stanney, K.M., & Fouad, H. (under review a, Dissertation Chapter 2). Theoretical foundations for integrating sound in interactive interfaces: Identifying temporal and spatial information conveyance principles. *Theoretical Issues in Ergonomics Science*.

Ahmad, A., Stanney, K.M., & Fouad, H. (under review b, Dissertation Chapter 3). Temporal audio applications: Using audio to train pace in a virtual environment training system. *Military Psychology*.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, pp. 257-262

Algazi, V.R., Duda, R.O., Thompson, D.M., & Avendano, C. (2001). *The CIPIC HRTF Database*. Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24. Available on line: http://interface.cipic.ucdavis.edu/data/doc/CIPIC_HRTF_Database.pdf, accessed January, 28, 2006.

Allen, J.F. (1983). Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, *26*(11), pp. 832-843.

Allen, J.F. (1984). Towards a general theory of action and time. *Artificial Intelligence*, *23*(2), pp. 123-154.

Allen, J.F., & Ferguson, G. (1994). Actions and events in interval temporal logic. *Journal of Logic and Computation*, *4*, pp. 531–579.

Allen, J.F., & Ferguson, G. (1997). Actions and events in interval temporal logic. In O. Stock (Ed.), *Spatial and Temporal Reasoning* (pp. 205-245). Kluwer Academic Publishers, Dordrecht, Netherlands.

Apostolos, M., Zak, H., Das, H., & Schenker, P.S. (1992). *Multisensory feedback in advanced teleoperations: Benefits of auditory cues*. Proceedings of the Sensor Fusion V, (pp. 98-105). : SPIE.

Bailey B.P., Konstan J.A., & Carlis J.V. (2001). The effects of interruptions on task performance, annoyance, and anxiety in the user interface, in: M. Hirose (Ed.) *Human-Computer Interaction - INTERACT 2001 Conference Proceedings.* Amsterdam: IOS Press, 593-601.

Bailey B.P., & Konstan J.A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, 22 (4), 685-708.

Banbury S., Fricker L., Tremblay S., & Emery L. (2003). Using auditory streaming to reduce disruption to serial memory by extraneous auditory warnings. *Journal of Experimental Psychology: Applied*, 9 (1), 12-22.

Baumgartner, T., Lutz, K., Schmidt, C.F., & Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. *Brain Research*, 1075, 151-164.

Barrass, S. (1997). *Auditory Information Design*. Unpublished PhD Thesis. The Australian National University, Canberra, Australia. Available online at http://thesis.anu.edu.au/public/adt-ANU20010702.150218/, accessed January 9, 2006.

Bartscherer, M.L., & Dole, R.L. (2005). Interactive Metronome Training for a 9-year-old boy with attention and motor coordination difficulties. *Physiotherapy Theory and Practice 21*(4), 257-269.

Begault, D.R. (1994). *3D Sound for Virtual Reality and Multimedia*. Academic Press, Inc., Cambridge, MA.

Begault, D.R. (1999). *Auditory and non-auditory factors that potentially influence virtual acoustic imagery*. Proceeding of AES 16th International Conference on Spatial Sound Reproduction, pp. 13-26, (Rovaniemi, Finland), April 10-12.

Begault, D.R. (2000). *3D Sound for Virtual Reality and Multimedia*. NASA/TM-2000-209606. Moffett Field, Calif.: National Aeronautics and Space Administration, Ames Research Center; Hanover, MD.

Billinghurst, M., Bowskill, J., Dyer, N., & Morphett, J. (1998). Spatial information displays on a wearable computer. *Computer Graphics and Applications, IEEE, 18*(6), 24 -31.

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction, 4*(1), pp. 11-44.

Blauret, J. (1983). *Spatial hearing: The psychophysics of human sound localization*. Translated by J.S. Allen. MIT Press, Cambridge, Mass.

Bloom, B.S. (Ed., 1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Susan Fauer Company, Inc.

Boyle, A. J., Wilson, A.M., Connelly, K., McGuigan L., Wilson, J., & Whitbourn, R. (2002). Improvement in timing and effectiveness of external cardiac compressions with a new non-invasive device: the CPR-Ezy. *Resuscitation, 54*(1), 63-7.

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press: Cambridge, Mass.

Brewster, S.A. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces*. PhD Thesis, University of York, UK.

Brewster, S.A. (1997). Using non-speech sound to overcome information overload. *Displays, Special Issue on Multi-Media Displays, 17*, 179-189.

Brewster, S.A. (1998). Using non-speech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction, 5*(3), 224-259.

Brewster, S.A. (2003). Non-speech auditory output. In J.A. Jacko and A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. 220-239). Mahwah, NJ: Lawrence Erlbaum.

Brown, S.W., & Boltz, M.G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance, 28*(3), pp. 600–615.

Brungart, D.S., & Robinowitz, W.M. (1999). Auditory localization of nearby sources. Head-related transfer functions. *Journal of the Acoustical Society of America, 106*(3), pp. 1465 - 1479.

Brungart, D.S., Durlach, N.I., & Robinowitz, W.M. (1999). Auditory localization of nearby sources. II. Localization of a broadband source. *Journal of the Acoustical Society of America, 106*(4), pp. 1956 - 1968.

Brungart, D.S. (1999a). Auditory localization of nearby sources III: Stimulus effects. *Journal of the Acoustical Society of America, 106*(6), pp. 3589 - 3602.

Brungart, D.S. (1999b). *Auditory parallax effects in the HRTF for nearby sources*. Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz: New York, Oct 11-20.

Brungart, D.S. (2001). Preliminary model of auditory distance perception for nearby sources. In S. Greenberg & M. Slaney (Eds.), C*omputational Models of Auditory Function* (pp. 83-95). IOS Press.

Buagaska, M. (2003). *Classification Model for Visual Spatial Design Guidelines in the Digital Domain*. 2nd Workshop on Software and Usability Cross-Pollination: The Role of Usability Patterns. September 1-2, Zürich, Switzerland. Available online: http://wwwswt.informatik.uni-rostock.de/deutsch/Interact/06Bugajska_2003.pdf, accessed October, 1, 2006.

Burmistrov I., & Leonova A. (2003). Do interrupted users work faster or slower? The micro-analysis of computerized text editing task, in: J. Jacko & C. Stephanidis (Eds.) *Human-Computer Interaction: Theory and Practice (Part I) - Proceedings of HCI International 2003, Vol. 1*, Mahwah: Lawrence Erlbaum Associates, 621-625.

Chen, Y., Repp, B.H., & Patel, A.D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, 21, 515-532.

Cutting, J.E., & Vishton, P.M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Handbook of perception and cognition, Vol. 5: Perception of space and motion*, pp. 69 – 117, San Diego, CA: Academic Press.

Davis, E.T., Scott, K., Pair, J., Hodges, L.F., & Oliverio, J. (1999). *Can audio enhance visual perception and performance in a virtual environment?* Proceedings of the Human Factors and Ergonomics Society (HFES) 43rd Annual Meeting, pp. 1197-1201.

Day, R., Holland, S., Bowers, D., & Dil, A. (2004). Using Spatial Audio in Minimal Attention Interfaces: Towards an Effective Audio GPS Navigation System. Technical Report. Available online: http://computing-reports.open.ac.uk/index.php/content/download/150/887/file/TR2004_08.pdf, accessed October 5, 2006.

Dinh, H.Q., Walker, N., Hodges, L.F., Chang Song, & Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Proceedings of IEEE Virtual Reality*, March 11-17, pp. 222-228.

Driver, J., & Spence, C. (2000). Multi-sensory perception: Beyond modularity and convergence. *Current Biology, 10*(20), R731-R735.

Duda, R.O. (1997). Elevation Dependence of the Interaural Transfer Function. In R.H. Gilkey & T.R. Anderson (Eds.). *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Edwards, G.W., Barfield, W., Nussbaum, M.A. (2004). The use of force feedback and auditory cues for performance of an assembly task in an immersive virtual environment. *Virtual Reality, 7*(2), 112 – 119.

Ekstrom, R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). *Kit of Factor-Referenced Cognitive Tests*. Educational Testing Service (ETS), Princton, New Jersey.

European Telecommunications Standards Institute [ETSI] (2002). *Human factors: guidelines on the multimodality of icons, symbols, and pictograms* (Rep. No. ETSI EG 202 048 v 1.1.1). Sophia Antipolis Cedex, France.

Fahlenbrach, K. (2002). Feeling sounds- emotional aspects of music videos. *IGEL-Conference Pécs*: (1-10). Germany: Martin-Luther-Universität Halle-Wittenberg.

Frauenberger, C., Putz, V., Holdrich, R., Stockman, T. (2005). *Interaction Patterns for Auditory User Interfaces*. Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland.

Gardner, M.B. (1969). Distance Estimation of 0 Degrees or apparent 0 Degree-oriented Speech Signals in Anechoic Space. *Journal of the Acoustical Society of America, 54*, pp. 47-53.

Gaver, W. (1986). Auditory icons: Using sound in computer interfaces. *Human Computer Interaction, 2*(2), 167-177.

Gillie T. & Broadbent D. (1989). What makes interruptions disruptive? A study of length, similarity and complexity, *Psychological Research*, 50 (4), 243-250.

Guillaume, A., Drake, C., Rivenez, M., Pellieux, L., Chastres, V. (2002). Perception of urgency and alarm design, *Proceedings of the 8th International Conference on Auditory Display (ICAD)*, Kyoto, Japan, pp. 357-361.

Hakkila, J., & Rankainen, S. (2003). Dynamic auditory cues for event importance level. *Proceedings of the 2003 International Conference on Auditory Display (ICAD)*, Boston, MA, USA, July 6-9.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX. In P. Hancock & N. Meshkati. (Eds). *Human Mental Workload*. Amsterdam: North-Holland.

Hebrank, J., & Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. *Journal of the Acoustical Society of America, 56*(6), pp. 1829-1834.

Ho C.-Y., Nikolic M.I., Waters M.J. & Sarter N.B. (2004). Not now! Supporting interruption management by indicating the modality and urgency of pending tasks, *Human Factors*, 46 (3), 399-409.

Holt, R.E., & Thurlow, W.R. (1969). Subjective orientation and judgment of distance of a sound source. *Journal of the Acoustical Society of America, 46*(6B), pp. 1584-1585.

Interactive Metronome, available online: http://www.interactivemetronome.com/im/index.asp, accessed December, 12, 2005.

Jeffress, L.A., & Taylor, R.W. (1961). Lateralization vs. localization. *Journal of the Acoustical Society of America, 33*(4), pp. 482-483.

Jones, D., Stanney, K.M., & Fouad, H. (2005). *An Optimized Spatial Audio System for Virtual Training Simulations: Design and Evaluation.* The 11th Internal Conference on Auditory Displays (ICAD), Ireland, July 6-9.

Jones, M.R. (2004). Attention and Timing. In J.G. Neuhoff (Ed.), *Ecological Psychoacoustics*, San Diego, CA: Elsevier Academic Press.

Kaplan, R. (2002). *Rhythmic training for dancers*, Human Kinetics, Champaign, IL.

Karageorghis, C.I., & Terry, P.C. (1997). The Psychophysical Effects of Music in Sport and Exercise: A Review. *Journal of Sport Behavior, 20*(1), pp. 54-68.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3,* 203–220.

Kern, K.B., Sanders, A.B., Raife, J., Milander, M.M., Otto, C.W., & Ewy, G.A. (1992). A study of chest compression rates during cardiopulmonary resuscitation in humans: the importance of rate-directed chest compressions. *Archives of Internal Medicine, 152*(1), pp. 145-149.

Kim, H-S., DiGiacomo, T., Egges, A., Lyard, E., Garchrry, S., & Magnenat-Thalmann, N. (2004). *Believable virtual environment: Sensory and perceptual believability*. Retrieved May 17, 2007 from http://www.miralab.unige.ch/papers/329.pdf.

Kolers, P.A., & Brewster, J.M. (1985). Rhythms and responses. *Journal of Experimental Psychology: Human Perception and Performance, 11*(2), pp. 150-167.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory Display* (pp. 1-77). Reading, MA: Addison-Wesley.

Kurtz, S., & Lee, T.D. (2003). Part and whole perceptual-motor practice of a polyrhythm. *Neuroscience Letters, 338*, pp. 205–208.

Libkuman, T.M., Otani, H., & Steger, N. (2002). Training in timing improves accuracy in Golf. *Journal of General Psychology, 129*(1), pp. 77-96. Available online: http://www.findarticles.com/p/articles/mi_m2405/is_1_129/ai_86431660/pg_6, accessed December 21, 2005.

Loomis, J.M., Golledge, R.G., & Klatzky, R.L. (1998). Navigation system for the blind: Auditory display modes and guidance. *Presence: Teleoperators & Virtual Envimnments*, 7, 193-203.

McFarlane, D.C. (2002). Comparison of four primary interruption methods for coordinating interruption of people in human-computer interaction. *Human Computer Interaction, 17*, pp. 63-139.

McFarlane, D.C. & Latorella, K.A (2002). The scope and importance of human interruption in human-computer interaction. *Human Computer Interaction, 17*, pp. 1-61.

McGookin, D.K. & Brewster, S.A. (2004). Understanding concurrent earcons: Applying auditory scene analysis to concurrent earcon recognition. *ACM Transactions on Applied Perception, 1*(2), pp. 130-155.

McGregor, P., Horn, A.G., & Todd, M.A. (1985). Are familiar sounds ranged more accurately? *Perceptual and Motor Skills, 61*, 1082.

Meegan, D.V., Aslin, R.N., & Jacobs, R.A. (2000). Motor timing learned without motor training. *Nature Neuroscience, 3*. Available online: http://www.bcs.rochester.edu/people/robbie/meeganaslinjacobs.nn00.pdf , accessed October 1, 2006.

Middlebrooks, J.C., & Green, D.M. (1991). Sound localization by human listeners. *Annual Review of Psychology, 42*, pp. 135 - 159.

Milham, L.M (2005). Investigation the Effects of 3-D Spatialized Auditory Cues on the Development of Situational Awareness in Teams. PhD Dissertation. University of Central Florida, Orlando, Florida.

Mills, W. (1972). Auditory localization. In J.V. Tobias (Ed.), *Foundations of Modern Auditory Theory* (pp. 301-348). New York: Academic Press.

Mulgund, S., Stokes, J., Turieo, M., & Devine, M. (2002). *Human/machine interface modalities for soldier systems technologies.* Final Technical Report, Natick, Massachusetts.

National Research Council (2000). *Design in the New Millennium: Advanced Engineering Environments: Phase 2*. National Academy of Engineering, National Academies Press, Washington, D.C.

Nelson, W., Bolia, R., & Tripp, L. (2001). Auditory localization under sustained +G acceleration. *Human Factors, 43*(2), 299-309.

Neuhoff, J.G., & McBeath, M.K. (1996). The Doppler illusion: The influence of dynamic Intensity change on perceived pitch. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), pp. 970-985.

Patterson, R.D., & Mayfield, T.F. (1990). Auditory warning sounds in the work environment [and Discussion]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 327*(1241), Human Factors in Hazardous Situations. (Apr. 12, 1990), pp. 485-492.

Pew, R.W. (2003). Evolution of human-computer interaction: From Memex to Bluetooth and beyond. In J.A. Jacko & A. Sears (Eds.), *Human-Computer Interaction Handbook*, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Plomp, R.(2002). *The intelligent ear: On the nature of sound and perception.* Mahwah, NJ : Lawrence Erlbaum Associates.

Repp, B.H., & Penel, A. (2002). Auditory Dominance in Temporal Processing: New Evidence From Synchronization With Simultaneous Visual and Auditory Sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1085-1099.

Repp, B.H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research. 68*, 252-257.

Repp, B.H. (2006). Does an auditory distractor sequence affect self-paced tapping? *Acta Psychol (Amst), 121*(1), pp. 81-107.

Richardson, A. (1977). Verbalizer-visualizer: A cognitive style dimension. *Journal of Mental Imagery*, 1(1), 109-126.

Rigas, D., Memery, D., & Yu, H. (2001). Experiments in using structured musical sounds, synthesized stimuli to communicate information: Is there a case for integration and synergy. *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processin*g, Hong Kong, pp. 465- 468.

Rowe, C. (1999). Receiver psychology and the evolution of multi-component signals. *Animal Behaviour, 58*, pp. 921-931, available online: http://www.staff.ncl.ac.uk/candy.rowe/Rowe99.pdf, accessed January 24, 2006.

Sawhney N. & Schmandt C. (2000). Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments, *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7 (3), 353-383.

Schreiber, F.A. (1994). Is Time a Real Time? An Overview of Time Ontology in Informatics. In W.A. Halang, & A.D. Stoyenko (Eds.), *Real Time Computing*, Springer Verlag NATO-ASI Vol. F127, pp. 283-307.

Seeber, B.U., & Fastl, H. (2003). *Subjective Selection of Non-Individual Head-Related Transfer Function*. Proc. 2003 International Conference on Auditory Display, pp. 259-262, Boston University, Bost, MA, July 6-9, 2003.

Shinn-Cunningham, B.G. (2001). Localizing sounds in rooms. Position Paper. Campfire: Acoustic Rendering for Virtual Environments. AC*M SIGGRAPH and Eurographics Campfire*. Snowbird, Utah, May 26-29.

Shinn-Cunningham, B.G. (2004). The Perceptual Consequences of Creating a Realistic, Reverberant 3-D Audio Display. *Proceedings of the International Congress on Acoustics*, Kyoto, Japan, 4-9 April.

Stanney, K.M., Graeber, D.A., & Kennedy, R.S. (2005). Virtual environment usage protocols. In W. Karwowski (ed.), *Handbook of Standards and Guidelines in Ergonomics and Human Factors* (pp. 381-398). Mahwah, NJ: Lawrence Erlbaum.

Stanney, K.M., Kingdon, K., Graeber, D., & Kennedy, R.S. (2002). Human performance in immersive virtual environments: Effects of duration, user control, and scene complexity. *Human Performance*, 15(4), pp. 339-366.

Stanney, K.M., Samman, S., Reeves, L., Hale, K., Buff, W., Bowers, C., Goldiez, B., Nicholson, D., & Lackey, S. (2004). A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations. *International Journal of Human-Computer Interaction*, 17(2), 229-257.

Strybel, T.Z., & Perrott, D.R. (1984). Discrimination of relative distance in the auditory modality: the success and failure of the loudness discrimination hypothesis. *Journal of the Acoustical Society of America, 76*, pp. 318 - 320.

Sulzen, J. (2001). Modality-based Working Memory. Project Report. School of Education, Stanford University, Stanford, CA. Available online: http://ldt.stanford.edu/~jsulzen/james-sulzen-portfolio/high/index.html

Tannen, R.S., Nelson, W.T., Bolia, R.S., Warm, J.S., & Dember, W.N. (2004). Evaluating adaptive multisensory displays for target localization in a flight task. *International Journal of Aviation Psychology*, 14(3), 297-312.

Thaut, M.H., Mcintoshn, G.C., Rice, R.R., Miller, R.A., Rathbun, J., & Brault, J.M. (2004). Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Movement Disorders, 11*(2), pp. 193 – 200.

Thaut, M.H. (2005). *Rhythm, music, and the brain.* Routledge, Taylor and Frances Group, New York: NY.

Tsimhoni, O., Green, P., & Lai, J. (2001). Listening to natural and synthesized speech while driving: Effects on user performance. *International Journal of Speech Technology, 4*(2), pp. 155-169.

USArmy - Department of Army (2003). Urban Operations, Field Manual (FM3-06). Available online: http://www.globalsecurity.org/military/library/policy/army/fm/3-06/, accessed November 28, 2006.

Vila, L. (1994). A survey on temporal reasoning in artificial intelligence. *AI Communications, 7*, 4-28.

Walker, B.N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: hearing, grouping, and meaning making. In J.G. Neuhoff (Ed.), *Ecological Psychoacoustics*, San Diego, CA: Elsevier Academic Press.

Walker, B.N., & Kramer, G. (2005). Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception, 2*(4), pp. 407-412.

Wallach, H., Newman, E.B., Rosenzweig, M.R. (1949). The Precedence Effect in Sound Localization. *American Journal of Psychology, 62*(3), pp. 315-336.

Wenzel, E.M. (1992). Localization in Virtual Acoustic Displays, *Presence, 1*(1), pp. 80-107.

Wenzel, E.M., Arruda, M., Kistler, D.J., & Wightman, F.L. (1993). Localization using Non-individualized Head-related Transfer Functions. *Journal of the Acoustical Society of America, 94*, pp. 111-123.

Wickens, C.D. (1984). Processing Resources in Attention. In R. Parasurraman & R. Davies (Eds.), *Varieties of attention*. New York: Academic Press.

Wickens, C.D. (1992). *Engineering psychology and human performance (2nd Ed.).* New York: Harper Collins.

Wickens C.D., Dixon S.R., & Seppelt B. (2005). *Auditory preemption versus multiple resources: Who wins in interruption management?* Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (HFES'05), Santa Monica: HFES, 463-467.

Wiens, J.A., Martin, S.G., Holthaus, W.R., & Iwen, F.A. (1970). Metronome timing in behavioral ecology studies. *Ecology, 51*(2), pp. 350-352.

Wightman, F. & Kistler, D. (1989). Headphone simulation of free-field listening. II: Psychophysical validation. Journal *of the Acoustical Society of America, 85*(2) 868-878.

Wightman, F.L., & Kistler, D.S. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America, 105*(5), pp. 2841-2853.

Wijnalda, G., Pauws, S., & Vigoli, F. (2005). A personalized music system for motivation in sport performance. *Pervasive Computing*, pp. 26-32.

Witmer, B.G., & Singer, M.J. (1998). Measuring presence in virtual environments: A presence puestionnaire. *Presence, 7*(3), pp. 3-8.

Zwicker, E. & Fastl, H. (1999). *Psychoacoustics: Facts and models, 2nd Edition.* Germany: Springer-Verlag.

APPENDIX A: RESPONSES TO VALIDATION QUESTIONNAIRE

Appendix A graphically displays individual expert responses to validation questionnaire (see Chapter 2). Figures A.1, A.2, and A.3 are based on expert responses for Audio Integration Model (Chapter 2, Figure 1), Temporal Audio Theoretical Model (Chapter 2, Figure 2), and Spatial Audio Theoretical Model (Chapter 2, Figure 5), respectively.

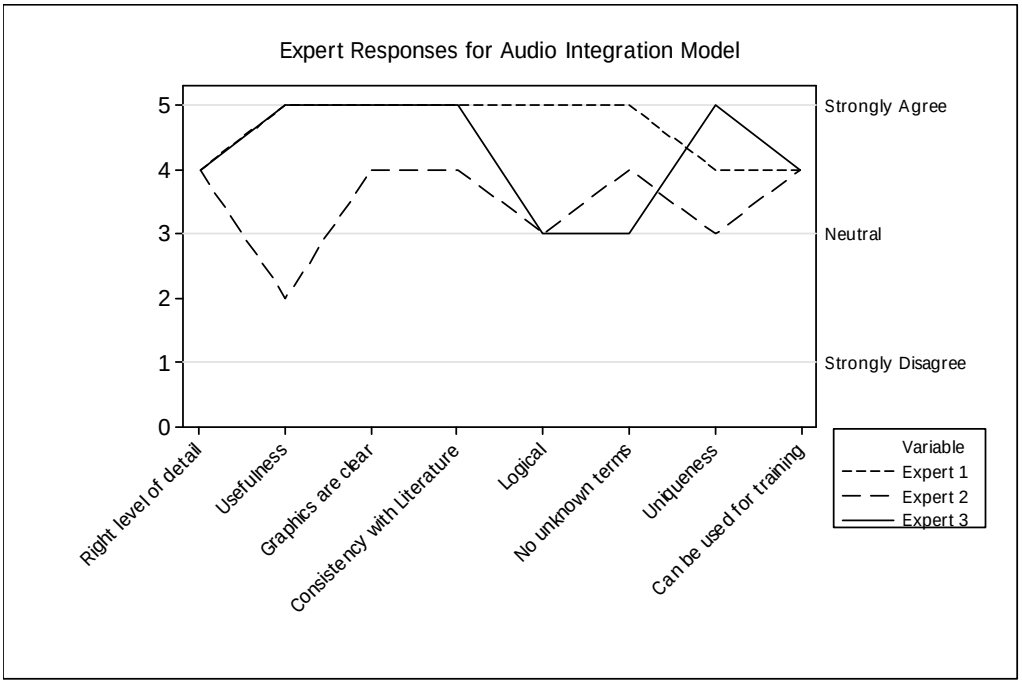

Figure A.1: Expert Responses for Audio Integration Model

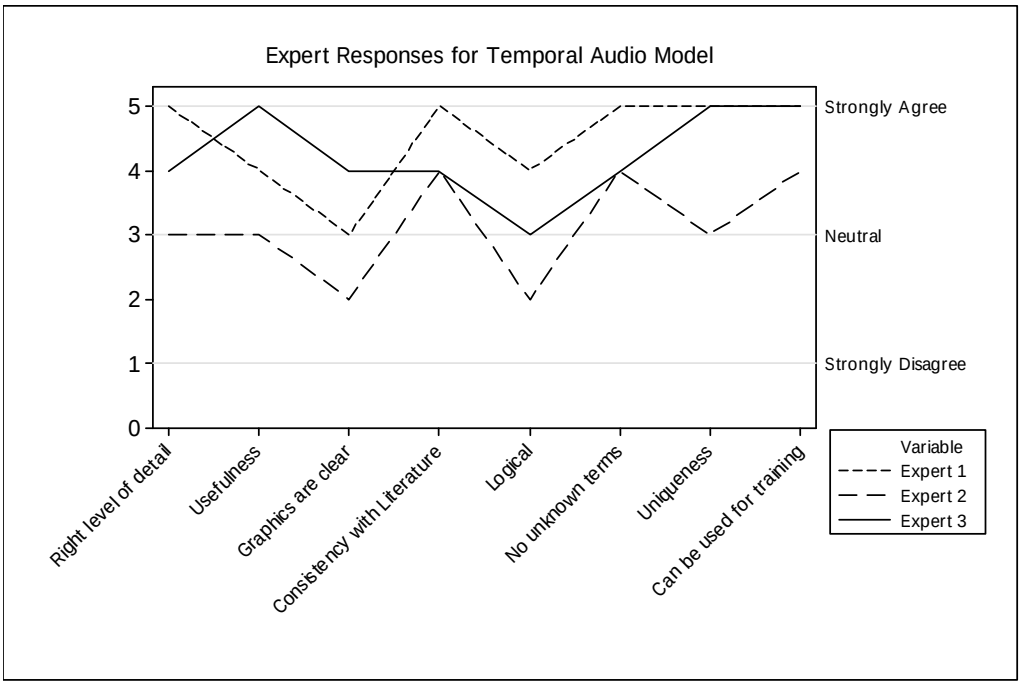

Figure A.2: Expert Responses for Temporal Audio Theoretical Model

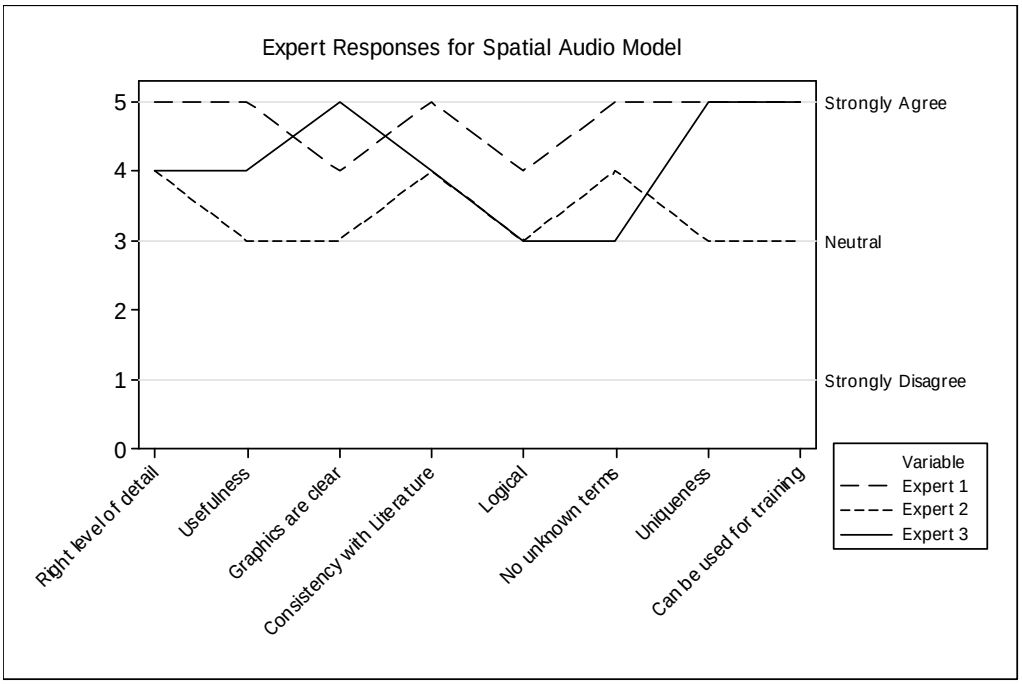

Figure A.3: Expert Responses for Spatial Audio Theoretical Model

APPENDIX B: VE USED IN EMPIRICAL STUDIES

Appendix B provides a graphical illustration for the VE used in the pace empirical validation study (Chapter 3) and integration empirical validation study (Chapter 4). Figure B.1 provides a screen shot of the VE.

Figure B.1: Screen Shot of VE used in Empirical Validation