JYVÄSKYLÄ STUDIES IN COMPUTING
127

Markku Häkkinen

Why Alarms Fail
A Cognitive Explanatory Model

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN COMPUTING 127

Markku Häkkinen

Why Alarms Fail

A Cognitive Explanatory Model

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Agora-rakennuksen salissa AgAud 2
joulukuun 21. päivänä 2010 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
in building Agora, Auditorium AgAud 2, on December 21, 2010 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2010

Why Alarms Fail

A Cognitive Explanatory Model

JYVÄSKYLÄ STUDIES IN COMPUTING 127

Markku Häkkinen

Why Alarms Fail

A Cognitive Explanatory Model

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2010

Editor
Seppo Puuronen
Department of Computer Science and Information Systems, University of Jyväskylä
Pekka Olsbo, Sini Tuikka
Publishing Unit, University Library of Jyväskylä

Cover pictures by Markku Häkkinen

URN:ISBN:978-951-39-4193-2
ISBN 978-951-39-4193-2 (PDF)

ISBN 978-951-39-4145-1 (nid.)
ISSN 1456-5390

Jyväskylä University Printing House, Jyväskylä 2010

ABSTRACT

Häkkinen, Markku
WHY ALARMS FAIL: A cognitive explanatory model
Jyväskylä: University of Jyväskylä, 2010, 102 p.
(Jyväskylä Studies in Computing
ISSN 1456-5390; 127)
ISBN 978-951-39-4193-2 (PDF), 978-951-39-4145-1 (nid.)
Diss.

Alarms are ubiquitous in our modern world, existing as personal alarms of convenience, such as alarm clocks and battery indicators on mobile phones, critical alarms of necessity in complex human-system interfaces, such as ground proximity warning systems in aircraft, and large scale public notification systems for disasters and emergencies. Response to alarms is a fundamental process in many species, well exhibited as a key survival mechanism in birds and mammals. However, in spite of the life safety importance of alarms, many humans reliably ignore them. The reasons for alarm failure are many, and this dissertation sets out to examine factors which impact the efficacy of alarms. Unlike the animal model, where alarms have a natural, biological salience, human alarms are implemented through a variety technical means, and often presented as abstract signals. As technology has evolved, so too have the ways in which alarms can be presented. Starting with early research on speech-based alarms and moving through an examination of developments from the field of accessible information systems, an approach to defining multimodal alarms is presented that addresses the sensory and cognitive factors that impact alarm detection and understanding.

Keywords: alarms, warnings, auditory displays, multimodal displays, attention, accessibility, human factors, usability

Author's address Markku Häkkinen
Department of Mathematical Information Technology
Mattilanniemi 2, Agora building
University of Jyväskylä
mhakkinen@acm.org

Supervisors Professor, Dr. Pertti Saariluoma
Department of Computer Science and Information Systems
University of Jyväskylä

Professor, Dr. Pekka Neittaanmäki
Department of Mathematical Information Technology
University of Jyväskylä

Professor, Dr. Heikki Lyytinen
Department of Psychology
University of Jyväskylä

Reviewers Docent, Dr. Antti Oulasvirta
Helsinki Institute for Information Technology

Professor, Dr. Matti Vartiainen
Aalto University School of Science and Technology

Opponents Professor, Dr. Jouni Kivistö-Rahnasto
Department of Industrial Management
Tampere University of Technology

Dr. Jaakko Saijonmaa
CASSIDIAN

ACKNOWLEDGEMENTS

Approximately 30 years ago, while working as a research programmer in the psychoacoustics lab at the Central Institute for the Deaf, a literature search related to the perception of short sequences of tones in *PsychAbstracts* led to the chance finding of an article in the journal *Human Factors* by Carol Simpson and Douglas Williams. Having done some work with simple voice synthesis, this article, entitled "Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings," captured my attention. I called Dr. Simpson to discuss her research and within the year I found myself beginning a replication and expansion of that research, resulting from that truly serendipitous find. Fast forward and once again the topic is alerting signals, though the context had shifted from aviation to look instead at alarms aimed at the public, and in particular, to members of the population who may find difficulty in receiving the alarm itself as a result of visual disabilities. The journey to the present point has been long, but each step along the way has been enlightening, from auditory warnings and synthetic speech, to assistive technologies and information standards, to the present synthesis that seeks to address the challenges of improving the effectiveness of alarms and warnings for all.

Dr. Helen Sullivan, my wonderful colleague, collaborator and spouse, has been of immeasurable support throughout this extended effort. Words alone are fully insufficient when it comes to offering to her my thanks and gratitude.

When I first visited the University of Jyväskylä about 11 years ago, Dr. Pekka Neittaanmäki introduced me to the vibrant, multidisciplinary research environment that is the Agora Center. Over the years, the frequency of my visits to Jyväskylä increased, and, combined with the regular encouragement of Dr. Neittaanmäki to undertake PhD studies, I found myself once again in graduate school. In this process, I was introduced to Dr. Pertti Saariluoma, whose expertise, patience, and encouragement have made this dissertation possible. My first meeting with Dr. Saariluoma was enjoyable and most memorable, covering a broad range of topics related to psychology, human factors and cognitive science. Every subsequent meeting with Dr. Saariluoma has been intense and mentally invigorating, whether in his office, at a coffee shop, or on the morning Pendolino from Tampere. As a result of my involvement in digital talking books for the blind and dyslexic, I looked forward to the opportunity to meet with Dr. Heikki Lyytinen when I first arrived at Agora, and the many hours spent with him since, in discussion and brainstorming, have been a real pleasure. I wish to sincerely thank Drs. Saariluoma, Neittaanmäki, and Lyytinen for their time, expertise, and support.

I want to thank Drs. Antti Oulasvirta and Matti Vartiainen, the reviewers of this dissertation, for their excellent comments and suggestions. Their knowledge and expertise is evident and has assisted me in improving the content of this work. Drs. Jouni Kivistö-Rahnasto and Jaakko Saijonmaa receive my sincere appreciation for agreeing to serve as opponents.

In the late stages of preparing the thesis Dr. Seppo Puuronen provided excellent editorial assistance, and I want to especially thank him for his time, effort, and quick responses. Thanks also to Pekka Olsbo for assistance in the preparation for the publication of this thesis.

Acknowledgement is also given for the support received from the SCOPE project, which has been generously funded by TEKES. A special thank you goes to my team on the iSCOPE notification research, Matti Haataja, Pauli Kettunen, and Jonne Räsänen, who are creating some brilliant and really cool mobile apps and Web services based on the multimodal alerting model. Thanks also to Dr. Veikko Hara and Markku Lauttamus for their support on the SCOPE project, and to the Department of Mathematical Information Technology and COMAS for support during the development this work.

Lastly, I wish also to express gratitude to members of my family who have been supportive throughout this process, in particular, my father, Dr. Raimo Häkkinen, my three children, Ian, Emily, and Alexander, my brother Dr. Pertti Häkkinen, and my aunt Ritva Gurovitsch.

Jyväskylä December, 2010
Markku Häkkinen

FIGURES

CONTENTS

LIST OF INCLUDED ARTICLES

I Hakkinen, M.T. & Williges, B.H. (1984). Synthesized Warning Messages: Effects of an Alerting Cue in Single- and Multiple Function Voice Synthesis Systems. *Human Factors*, 26(2), 319-330.

II Williges, B.H., Schurick, J.M., Spine, T.M., & Hakkinen, M.T. (1986). Using Speech in the Human-Computer Interface. Chapter 7. In R.W. Ehrich & R.C. Williges (Eds.), *Human-Computer Dialogue Design* (pp. 241-271). Amsterdam: Elsevier.

III DeMeglio, M., Hakkinen, M., & Kawamura, H. (2002). Accessible Interface Design: Adaptive Multimedia Information System (AMIS). In K. Miesenberger, J. Klaus, & W. Zagler (Eds.), *Computers Helping People with Special Needs.*, Springer Lecture Notes in Computer Science, Vol. 2398 (pp. 406-412). Berlin: Springer.

IV Sullivan, H.T. & Hakkinen, M. T. (2006). Disaster Preparedness for Vulnerable Populations: Determining Effective Strategies for Communicating Risk, Warning, and Response. Paper presented at the *Third Annual Magrann Research Conference on The Future of Disasters in a Globalizing World*, New Brunswick, New Jersey.

V Hakkinen, M.T. & Sullivan, H.T. (2007). Effective Communication of Warnings and Critical Information: Application of Accessible Design Methods to Auditory Warnings. In B. Van de Walle, P. Burghardt & C. Nieuwenhuis (Eds.), *Proceedings of the 4th International ISCRAM Conference* (pp. 167-171). Delft: ISCRAM.

VI Sullivan, H.T., Hakkinen, M.T., & Piechocinski, D. (2009). Improving participation, accessibility and compliance for campus wide mobile emergency alerting systems. In Löffler, J. & Klann, M. (Eds.), *Mobile Response*, Springer Lecture Notes in Computer Science, Vol. 5424 (pp. 32-40). Berlin: Springer.

VII Sullivan, H., Hakkinen, M., & DeBlois, K. (2010). Communicating critical information using mobile phones to populations with special needs**.** *International Journal of Emergency Management, 7(1), 6– 16.*

'Come here,' they sang, 'renowned Ulysses, honour to the Achaean name, and listen to our two voices. No one ever sailed past us without staying to hear the enchanting sweetness of our song – and he who listens will go on his way not only charmed, but wiser, for we know all the ills that the gods laid upon the Argives and Trojans before Troy, and can tell you everything that is going to happen over the whole world.'

Book XII, The Sirens, Scylla and Charybdis, The Cattle of the Sun. *The Odyssey*, Homer

Ring the alarum-bell! Blow, wind! come, wrack!
At least we'll die with harness on our back.

Macbeth (Act 5, Scene 5), Wm. Shakespeare

Hang a lantern aloft in the belfry arch
Of the North Church tower as a signal light,--
One if by land, and two if by sea;
And I on the opposite shore will be,
Ready to ride and spread the alarm
Through every Middlesex village and farm,
For the country folk to be up and to arm."

The midnight ride of Paul Revere, Longfellow

The Wolf, however, did truly come at last. The Shepherd-boy, now really alarmed, shouted in an agony of terror: "Pray, do come and help me; the Wolf is killing the sheep"; but no one paid any heed to his cries, nor rendered any assistance.

Aesop

"Is there any point to which you would wish to draw my attention?"
"To the curious incident of the dog in the night-time."
"The dog did nothing in the night-time."
"That was the curious incident," remarked Sherlock Holmes.

The Memoirs of Sherlock Holmes, A.C. Doyle

It beeped on the computers at work. It beeped on cellphones and Palm Pilots and beeped from the kitchen and living room and beeped a constant beep-beep-beep on the street below and down at the station house and in the tenements of Faneuil Heights and East Bucky Flats. Everything beeped these days.

Mystic River, Dennis Lehane

Ground Proximity Warning System: *Pull Up! Pull Up!*
Pilot: *(in Chinese) "What does 'Pull Up' mean?"*

Last words captured on Cockpit Voice Recorder
Transcript, China Northern Airlines Crash,
13 November 1993

1 INTRODUCTION

1.1 Alarms: Origins, Evolution, and Present Day Challenges

The raising of an alarm, and subsequent response, is a phenomenon that has been exhibited in living species for perhaps tens of millions of years (Turner, 2000). Alarm production and response is, in fact, essential to the survival of any species. Failure to respond to alarm signals can have significant, negative consequence, commensurate with the level of hazard which caused the alarm. And alarms do fail, across many contexts (e.g., Bliss, 2003; Lachman, Tatsuoka, & Bonk, 1961; Proulx, Laroche, Jaspers-Fayer, & Lavallée, 2001; Xiao, Seagull, Nieves-Khouw, Barczak, & Perkins, 2004).

The significance of alarms to humans is evident by their recurring presence in mythology and literary works through the ages. Authors such as Homer (1967), Aesop (1998), Longfellow (2000) and Doyle (1938) have described events, real and fantasized, in which alarms (or the lack thereof) have determined the fates of individuals and kingdoms.

Humans have differentiated themselves from the animal kingdom by creating objects producing sounds suitable to signal alarm, first, in the form of bells in 2000 BC China (Shen, 1987). The bell saw growing use in European churches in the 4th and 5th century AD (Gatty, 1848; Tyack, 1898). Later, electrical-mechanical devices were developed creating a variety of sounds that ideally would capture people's attention. In the world wars of the last century, alarms evolved new meanings, such as to warn of air raids, or to alert crews of warships to action.

Over time, human use of alarms has evolved, from ensuring safety of life and property, to simply, and sometimes with annoyance, alerting us to relatively mundane events; alarms now serve functions of necessity and of convenience. As a result, we today live in an alarm rich environment, surrounded by a cacophony of visual and auditory alerts from many sources, with many potential meanings and intents. In spite of being designed to capture our attention, and elicit action, alarms can now often fail to produce the expected response.

The reason for failure is not simple, and this thesis will focus on how mismatches between technology and the fundamental psychological principles of perception, attention, learning, and memory results in predictable, yet preventable, alarm failure.

In nature, we see the use of alarms as a survival mechanism. Visiting our forests, backyards and gardens, we may notice the warning call of a chickadee, signaling others of its own species to the presence of a predator. What we may not realize is the chickadee has evolved a complex threat call, encoding information on the nature of the threat, and signaling other species to respond in a cascade of birdsong and action. At night, you may hear the barking of a dog, signaling to its master the presence of an intruder, and an example of cross species alerting, in which man has used the natural sensory and alarm capabilities of the domesticated animal as a means of protection. The canary has also served humans as an alerting mechanism. In English coal mines, caged canaries acted as a biological sensor that signals alarm only when the canary's song stops, indicating that the air in the mine is no longer safe.

Since the middle ages, church bells have been rung to herald significant events and summon citizens to action. In the last century, electro-mechanical alarms appeared on ships and would signal hazard or a call to arms. Sirens became mobile, alerting pedestrians and motorists to the presence of emergency vehicles. During the World Wars, sirens were used to signal danger from enemy bombardment, and after the World War, the Cold War saw those same sirens serve as a sad wail presaging thermonuclear war. Today those same sirens can signal severe weather, Tsunami, or other hazards. Or they may signal something less serious, such as a celebration or call out for volunteer firemen. Though technology and threats have evolved over time, some alarms, such as the siren, have remained largely unchanged in acoustic characteristics for the past 70 years.

Today, alarms come in many forms and in many contexts: the simple alarm clock to awaken us, the intermittent tones of a low-battery indicator on a mobile phone, a buzzing smoke alarm in the home, the ubiquitous auto theft alarms (now part of the urban soundscape), musical sequences or vibration to signal arrival of an important text message, spoken messages to alert pilots or drivers to potential collisions, and electronic tones and announcements in several languages to signal the impending arrival of a Tsunami wave.

Though we may react viscerally to alarming sounds, such as the crack of a gunshot, a human scream, or the crying of a child, we have grown dependent upon artificial, often abstract alarms. In contrast to the effectively hard-wired responses that many species exhibit when a biologically salient, natural alarm occurs, the modern human must learn and assign their own taxonomy of meaning for the many artificial alarms we have created to serve us. The abstraction of alarms adds a requirement to learn their meaning, or to associate meaning to novel signals based upon past experience. In creating this abstraction, are we neglecting the lessons to be learned from the biologically salient alarm signals when we consider the design of modern alarm systems?

In many cases, alarm systems, in conjunction with electronic sensors, augment our own sensory capabilities, detecting hazards in our environment before we may normally sense them. In other cases, such as in public warning, alarms are triggered by government authorities in response to risks real and perceived. In all cases, alarms are sounded with the intent to arouse response. In their most critical form, alarms strive to motivate people to take action that will prevent injury or loss of life. And in many cases, alarms go unheeded, either because they are not heard, not understood, or otherwise not acted upon for reasons without simple explanation. In complex, human/machine systems, whether they are commercial airliners, industrial processes, fire alarms or civil disaster alerting services, alarm failure can have catastrophic consequences.

Why do alarms fail? To answer that, we have to understand both the nature of the alarm itself and the abilities and context of the intended recipient. Designers of alarm systems seek to reach the widest audience practical yet may rely upon design guidance that does not reflect the reality of the application. With the best of intentions, and practice, alarms may be created that, by design will not be detected or understood by significant portions of the population they are intended to reach.

1.2 Approach

Whether creating a simple public fire alarm or a critical alarm component in a complex vehicle control system, the use of empirical or evidence-based design guidance should form the foundation of a human-centered design approach. But the question arises, are existing guidelines and approaches to alarm design resulting in effective systems? Can these guidelines anticipate effective alarm designs for the rapidly evolving diversity of end-users, their contexts and devices? These are among the questions this thesis seeks to examine.

1.2.1 Developing the Explanatory Framework

The approach of this work is influenced by explanatory frameworks (Saariluoma, 2005) as a means for providing a scientific and knowledge-based approach for solving design problems. To understand the problem of alarm failure we need to examine the nature of the alarm in our modern context, and to understand how human capacities in sensation, perception and cognition can affect our ability to perceive, understand and act upon these alarms. By understanding human capabilities in alarm detection and processing, we can begin to understand what factors contribute to alarm failure, and in turn, understand what characteristics may make alarms more or less reliable. We further have to understand whether existing empirically-based guidelines can fit emerging and unique design problems, and if not, what is it that we do know that can frame an effective solution path. In current design, the practice of pulling an existing alarm model off the shelf may be analogous to trying to fit a square peg into a round hole. We'll see examples where alarms that may adhere to established guidelines nonetheless fail in their application context.

Within the explanatory framework approach, this thesis draws from research and practice in several domains to answer the research questions: human factors research focused on developing or validating design guidelines, scientific knowledge on perceptual, attentional and cognitive processes, and design practice related to developing information presentation models for individuals with sensory or cognitive impairments. A key premise that will be explored is that lessons learned from solving the challenge of information presentation for the sensory or cognitively disabled will have applicability to individuals who are not chronically or congenitally disabled, but who may be disabled (or impaired) by situation or context.

Though the social sciences have long studied response to public warning alarms and the design of the risk communication message, examination of the reasons as to why individuals may fail to perceive alarms receives little study by that community (Mileti & Sorenson, 1990). This thesis argues that most alarm research outside of specialist applications such as aviation (e.g. Patterson, 1982), and medicine (e.g. Edworthy & Hellier, 2006), typically fails to discuss the perceptual and cognitive realities of alarm detection and response, operating under an assumption that most people will receive an alarm or warning. The focus is thus on message content, social processes and influences that cannot come into play without receipt of the initial alarm signal. All of the preparedness and training that may be practiced will be of little value if alarms go unnoticed.

Creating effective alarms depends not only on the use of effective actionable, messages, but also on ensuring that the message can break through the filters which may prevent these critical signals from achieving their intended result. These filters may arise from the ambient environment (e.g., noise or competing conversations), through impairments (e.g., visual, hearing or cognitive), through limitations in our ability to process signals, through cultural or linguistic differences, or simply resulting from unfamiliarity with the alarm itself. Though there exists a significant body of research examining auditory alarms in two contexts, aviation, dating from the 1960's and medical instrumentation from the 1980's, there is little empirical research in the area of adaptive and multimodal alarms, and even less in the design of alarm systems for the general public. The emphasis in the research literature on alarm design for specialist populations, generally pilots and medical staff, is obviously of great value. However, the potential for misapplication of specialist-centric alarms to public (non-specialist) populations may be problematic.

In everyday life, we may be engaged in a variety of tasks that place demands on our attention and that limit our ability to process information received. We may be in unfamiliar locations, unable to communicate in the local language, yet need to be able to assimilate critical information when it is presented. Alarms themselves can be ambiguous, confusing, or mistaken for non-critical events. In cases where alarms are heard for the first time, we are faced, in effect, with one trial learning; alarms must be immediately recognized as a signal to act. Understanding the role these limitations play in warning failure is key, as are the approaches to overcoming, or mitigating these limitations.

Of particular focus in this work is the use of multiple modalities, composed not only of sounds designed to capture our attention, but the written and spoken word that can enhance detection, identification and convey added meaning. Such added meaning is critical in cases where familiarity with alarm signals is not likely, or possible. Multimodal delivery also serves as the only practical way to ensure that those with sensory impairments are able to receive alarm information in at least one effective modality. However, delivery of multimodal alarms can be problematic in the public context, particularly as certain modalities may require proximity to the signal source. The use of ubiquitous technologies, such as mobile devices, holds promise, in that the alarm source with literally be in the hand (or pocket) of the alarm recipient. In fact, this thesis will argue that mobile devices offer a significant practical framework for effective alarms: personalization.

The experience of information accessibility, the practice of making digital information available to those with disabilities, such as sensory or cognitive impairments, is posited as a model for making alarms and warnings more broadly effective and accessible to all. The emerging philosophy and practice of Design for All (Stephanidis, 1995) is not simply an approach and motivation for designers to address the needs of people with disabilities, but serves more beneficially as a design model which can enable creation of systems which are effective for people not simply based on ability, but also on contextual and situational limitations that impair the ability of an individual to perceive, understand and act upon an alarm at any point in their life.

I will conclude that in the absence of biologically salient alarm signals for humans, the most promising means to achieve optimal alarm detection and response is to utilize personalization of a multimodal alerting signal. Personalization, in this context involves both user selected and system directed approaches to message presentation characteristics, format and style. It is important to note that the possibility of personalization is greatly enhanced by the growing number of mobile devices in the world, numbering in 2009 upwards of 4 billion mobile phones on a planet with 6 billion people. In this context, I propose a basic alarm presentation model that can serve as a framework for further empirical research in alarm accessibility, multimodality, and personalization.

1.2.2 Summarizing the Research Questions

In this thesis three core questions are examined. First, to what extent are current guidelines and standards for alarm design applicable to the growing diversity of alarm applications? Second, are design approaches from the information accessibility context applicable to creating effective alarm presentation models? And third, can a personalization and multimodality framework for alarm presentation increase the effectiveness of alarms?

Using the explanatory framework, these questions are examined with the proposed multi-modal alarm specification model presented as a means to support implementation of guideline-based alarm designs as well as providing an extensible facility for empirical research into personalized, multimodal alarms that address a diversity of users and contexts.

1.3 Structure of the thesis

This thesis is composed of five parts comprising:

- an introduction to the problem of alarm failure
- a review of natural and artificial alarms, information accessibility and associated technologies
- the sensory and cognitive aspects of alarm response in humans and discussion of existing models
- a summary of research articles
- a discussion of alarm design based upon a model that mitigates alarm failure through support for multimodal presentation and personalization

The articles selected by the author span a broad time period, beginning with research on auditory warnings carried out from 1981 to 1983. This early empirical research is fundamental and a keystone to the author's later work. It should be noted that the original findings have been referenced in human factors design guidelines handbooks (e.g. Salvendy, 1987; MIL-STD-411F, 1999). A later article introduces the author's work on information accessibility for people with disabilities, which resulted in contributions to international standards such as ANSI-NISO Z39.86 and W3C's SMIL. The present challenges of creating effective alarms for emergency and disaster notification are introduced in the remaining articles, and through the synthesis of the earlier warning research with the authors' work in accessibility of information for those with disabilities and standards development, a model is proposed that can increase the likelihood alarms will be detected, understood and responded to.

2 ALARM MECHANISMS: NATURAL AND ARTIFICIAL

A complete understanding of alarms is not possible without a survey of their occurrence in nature and their biological salience for species survival. From this initial ethological perspective, we can proceed to understand how alarms have evolved and developed as a component of human activity.

2.1 Natural Alarms: The Drive to Survive

Birds, the modern descendant of the dinosaur, demonstrate a sophisticated alerting system that is a model of efficacy. Take, for example, the common chickadee (*Poecile atricapillus*). When a chickadee detects the presence of a predator, such as a hawk, a song is produced indicating not only the presence of the predator, but also the relative size and distance (Templeton, Greene & Davis, 2005). Other chickadees within the aural range of the alerting song, though not being an individual match for the hawk, may band together to engage in a mobbing tactic to ward off the predator. Interestingly, other species of birds may "eavesdrop" and react to these alerting songs (Hurd, 1996; Templeton & Greene, 2007). Crows may be drawn to the scene of initial alarm and mob the hawk, along with Jays, while smaller species are observed to seek shelter or hides.[1] As a result the mobbing activity and the sudden defensive disappearance of available prey, the hawk will likely move on to seek easier targets.

---

1 This author uses threat calls as a reliable indicator of the presence of predatory birds in the yard of his suburban New Jersey home. Detection of visual cues (small birds entering hides and clusters of jays and crows) and auditory cues such as the mobbing calls of jays, will reward the amateur naturalist with the sight of a red tailed hawk or American Bald Eagle. In an analogous fashion, the sound of television news helicopters hovering, stationary, over this same neighborhood serves as an alerting signal, often indicating a serious accident on an adjacent highway, or an emergency situation at a local university.

The sophistication of the chickadee alerting system was studied by Templeton and others, and similar capabilities have been reported in a variety of bird species (Marler & Slabbekoorn, 2004). Mammals also exhibit complex alerting signals, such as in the Diana monkey (*Cercopithecus diana*), which can produce a variety of vocal alerts to signify different predators, and indicating relatively sophisticated semantics within the calls themselves (Zuberbühler, 2000). Rainey, Zuberbühler & Slater (2004) reported that hornbill birds react appropriately to the Diana threat calls, indicating that alarm salience crosses species boundaries.

Auditory alerts are not the only means by which animals will signal warning. In the African welt, meerkats, wildebeest, and other animals will engage in visually distinctive behaviors that signal alarm, often in combination with auditory alarm calls. For example, the prairie dog will both issue an auditory signal and jump, the "yip-jump" signal (Fredericksen & Slobodchikoff, 2007). The combination of natural alarm signals, what is termed multimodal warning in the evolutionary ecology literature (e.g., Rowe & Guilford, 1999), is common in many species, and is a term that is seeing increasing use in technology-based warning literature (e.g., Spence & Driver, 1999; Sullivan & Hakkinen, 2006), though the concept and benefits of multimodality warnings predate the technological use of the term (e.g.,Williges, Schurick, Spine, & Hakkinen, 1986).

In examining natural alarms in the animal kingdom, one can raise the question of nature vs. nurture: are threat calls and responses to threats innately hardwired or learned? In research with meerkats, Hollen and Manser (2007) observed that young meerkats reacted to threat calls from the adults, prior to developing any external reference to the actual source of the threat. Studies of physiological responses in infant chimpanzees (Bernston, et al., 2004) demonstrated threat related responses when presented with sounds of chimpanzee screams. What is important to note in these animal examples is that alerting and response mechanisms are likely to be innately present in most species, a mechanism for survival of the species itself. Of further note is that the initiation of a threat call, in response to certain stimuli, such as the shape of a hawk flying overhead, will trigger an alarm that is salient not only to members of the same species but to members of other species.

Mimicry across species may also play into the cascade effect of alerts, but in all cases, it must be noted, that the successful production and response to alerting calls is crucial to species survival. Birds, for example, which do not respond to a threat call, would be naturally selected out of a population, an effective application of Darwinian natural selection. Failure to respond, as we shall see in the coming sections, can have fatal consequences across species, irrespective of brain size or technological advancement.

In the context of cross species alerting, the common dog often serves as a natural alarm system for humans. The barking of a dog is a signal that, depending upon time and location, can signal passersby, or a significant threat such as an intruder. In his book, *Before the Dawn,* Wade (2006) describes the development of early human civilization in what is now central Asia. Domesticated dogs, genetic derivatives of the wolf, became what could be called the first alerting system utilized by humans. The use of dogs to watch for threats and signal their presence by barking allowed humans to remove themselves from the role of the constantly vigilant guard. Vigilance, a high demand task, was thus transferred to the dog, allowing the human to assume other, more productive tasks. As can be attested today, the barking of a dog still arouses a response, with the level of response determined by a variety of factors (e.g., familiarity with the animal, fear of dogs, etc.).

Humans naturally respond to a variety of stimuli that can evoke an alarm response. Research has examined parental reaction to the cries of infants, which can produce arousal, alertness and other physiological responses (e.g., Boukydis & Burgess, 1982). Hard science is perhaps not needed to explore this question; simply ask any parent to comment as to the effectiveness of a baby's scream in producing a response, such as feeding, or diaper change.

2.2 Artificial (or Abstract) Alarms

It can be assumed that early humans utilized their vocal abilities to generate screams and threat calls, much as we may do today, in response to a threat and serving to alert others within the social group to danger. As human civilization continued its development, the rise of technology would influence the creation of what we would term to artificial alarms. In the Bronze age, a period in which a metal could be forged and shaped to perform a specific tool or object led to the creation of the bell, first seen in China in 1700 BC (von Falkenhausen, 1994). There is little available evidence that would indicate the early use of those bells specifically for alarming. The significant point is that the creation of the bell, and perhaps before that, wooden instruments to produce sound, introduced the possibility of artificial alarm signals. The transition began, from naturally produced, species salient alarms to those that were artificially generated and required abstraction of signal to threat. Without training, would abstract alarms be able to evoke responses with the apparent success of salient natural signals?

Bells would develop a key role as an early alerting signal, most closely associated with their appearance within Buddhist temples within Asia in 200 BC (Rostoker, Bronson & Dvorak, 1984) and then Church structures in Italy around 400 AD (Tyack, 1898). The ringing of a bell could summon residents to a central point for religious services or public meetings. Bells saw early association with fire services, notably by Benjamin Franklin and the Union Fire Company in Philadelphia in 1750 (Carp, 2001), as a means to sound an alarm and summon volunteers to assist in putting out the fire. With the advent of electrification, bells could be sounded remotely and the first electrical fire alarm systems appeared in 1852 (IEEE, 2004). Steam power in the 1800's also introduced another form of alarm, the steam whistle. Locomotives would utilize the steam whistle as a warning signal, intended to warn those who may be on the tracks ahead of the approaching train. Mechanical warning sirens began to appear around 1900[2], with application to vehicles and outdoor warning signals. A variety of other alarm types emerged, including electrical buzzers and klaxon horns for automobiles in 1908.

All of these alarm mechanisms generated sounds that in themselves conveyed no direct meaning as to the nature of the threat or emergency. Knowledge of appropriate response could come only from educational programs, observation, or associative learning. Abstraction is thus a key feature of alarm design. Meaningful, effective salience could only come from close pairing of the threat (e.g., fire) with associated warning. The biological salience of natural threat calls was thus replaced by abstract, artificial signals which had no immediate relevance to the human listener.

In the case of London, in the Second World War, the air raid siren developed salience. The threat of air raids were real, as was the devastation and loss of life. Exploding bombs were not an abstract concept, and the pairing of the siren with the threat of air attack proved an effective means to encourage movement to shelter. A number of studies on the human response to the air raids exist, such as Schmideberg (1942) and Jones, Woolven, Durodie & Wessely (2004), indicating to some degree the effectiveness of the shelter and warning systems, and the resilience of the population in adapting to the wartime situation.

In contrast, within the United States during the Second World War, air raids did not take place, with minor exception, the 1942, hysteria inspired "air raid" on Los Angeles and rare balloon raids by the Japanese (Conley, 1968), though an infrastructure of sirens and air raid wardens were widely implemented. Of note are a series of articles from the Acoustical Society of America in 1942[3] regarding the design of public warning sirens and their placement within urban settings. Problems with outdoor warning sirens were studied and efforts taken to identify appropriate sound levels (up to 135 dB) and placement (e.g., Volkmann & Graham, 1942).

Air raid warning sirens remained in place after the war, taking on new roles even within the emerging nuclear threat environment of the Cold War. Many fire and emergency service stations found the presence and availability of the siren a useful tool to summon volunteers, much as in the previously mentioned model of Franklin in 1750 (Carp, 2001). Even today, communities in the United States utilize an air raid siren on a regular basis to summon emergency service volunteers or in some cases, to mark celebratory events. Such sirens also form part of the public warning system for severe weather, such as tornadoes in the United States. In countries at risk for Tsunami, sirens have been used for warning since the 1950's, with many news sirens installed in the months after the December 2004 Indian Ocean disaster.[4]

---

2 Sentry Siren Inc. began production of outdoor sirens in 1905.

3 The Journal of the Acoustical Society of America, Volume 14, Issue 1.

4 The author visited Phuket, Thailand in May 2005 to observe installation of Tsunami warning sirens, meet with the deputy mayor of Patong and collect data on public response to the sirens.

A significant problem arises from the use of a single warning mechanism for multiple purposes. The nature of the siren signal is alerting, but the recipient may not readily determine the reason for its activation. Severe weather events, possibly induced by climate change, now appear in regions that previously may never have experienced, for example, a tornado. A severe weather siren in such a case may be assumed, based on prior experience by residents, as indicating a call for volunteers to fight a fire, or even a regularly ignored test of civil defense readiness, rather than a tornado warning. As early as 1960 (Lachman, et al., 1961) reported confusion as to meaning of a siren used to warn of the impending arrival of a Tsunami in Hilo, Hawaii. If a greater number of people had understood the meaning of the siren, lives may have been saved through timely evacuation. Forman (1963) discussed the lack of response to sirens during the cold war, attributing it to a resignation as to the futility of trying to survive nuclear war. Gregg, et al (2007) reported that in present day Hawaii, significant numbers of the population do not understand the meaning of the Tsunami sirens. In Switzerland up to 50% of the Swiss population reportedly do not understand the meaning of emergency sirens[5] and Sullivan and Hakkinen (2006) summarize a number of reports on inappropriate or non-existent public response to recent Tsunami warnings.

2.3 The Use of Speech to Improve Warning Systems

Early mechanical and electrical alarms could indicate the existence of a hazard or threat, but not provide further direction. During the Second World War, "speaker" cars were often used to provide verbal direction and updates to a population, such as those used during the London Blitz. During the cold war, citizens in the United States were advised to tune to a Conelrad radio station for further instructions once the warning sirens had been activated (Frank, 1962). But such two stage processes added complexity and delay, assuming that the first alarm had in fact been heard and correctly identified (the siren), and then requiring the appropriate response to be recalled (tune to a specific radio channel). The government of the United Kingdom[6] has undertaken a public awareness campaign to direct individuals, at the sound of a siren (or obvious critical incident), to "Go In. Stay In. Tune In." The success of such efforts, are difficult to predict, and prior cold war experience may presage public response today.

A closer coupling of spoken information to the initial alarm signal could improve identification and response to the warning. The use of speech as a modality for information display has grown significantly in the past 30 years, seeing early application in aural warning systems within aircraft cockpits (e.g. Simpson & Williams, 1975), complex computer-based tasks (Hakkinen & Wil-

---

5 Reported to this author by Swiss government officials at IDRC 2008 in Davos.

6 UK Cabinet Office.
http://www.cabinetoffice.gov.uk/ukresilience/nscwip/goinstayintunein.aspx

liges, 1984) and in assistive technologies for people with visual impairments (Thatcher,1994; Hakkinen & DeWitt, 1997). Automotive application of speech displays[7] occurred first in the early1980's and today there is widespread and growing application of computer generated speech in automotive navigation systems, hand held devices, educational toys, and telephony-based services.

Sound, by its very nature, has significant value in serving as an alerting signal for humans in complex human-machine systems, notably those characterized as hands-busy/eyes-busy environments (Salvendy, 1987). For pilots of modern aircraft, aural and speech displays allow attention to be focused outside the cockpit during critical flight phases rather than on scanning visual instruments or warning annunciation panels. In less complex environments, such as mobile devices in which displays may be small or otherwise unreadable due to the operating environment, a speech display can provide an effective means to convey information to the user (Arons & Mynatt, 1994).

The use of speech as a component of alarm systems has been growing, though the majority of research, design guidelines, and applications have evolved from the context of cockpit warning systems and medical devices.

Speech and auditory displays in general, allow information to be conveyed relatively quickly, no matter where the visual attention of the human system operator is focused. However, reception and understanding of speech can be degraded by many factors, including masking by ambient environmental sounds and noise, unfamiliar or poor speech articulation, unfamiliar accents, sensory or cognitive impairment, competing conversations, or attentional demands upon the listener in the same or alternate modalities. In the following sections we will explore some of these issues in more detail.

2.4 Speech Displays and Safety

One vital application of speech technology is in serving to improve system safety. The Human Engineering Design Criteria standard MIL-STD-1472F (USDOD, 1999) states:

> "Speech displays may be used where mobility is necessary or where the user's eyes are busy. They should announce discrete events, not continuous status information. They should not be used if display use frequency is high, if simultaneous display of multiple messages is required, if messages are long, if messages include information that must be memorized, or if messages include a series of instructions that must be remembered." (ibid, p. 52)

The standard goes on to describe speech as:

---

7 The introduction of speech warnings in the early 1980's led to the first comedian as automobile, delivering the punch line to the query, "When is a door not a door?" The car dutifully replies via a synthesized or recorded voice, when a vehicle door is open, "Your door is ajar."

> "Most effective for rapid (but not automatic) communication of complex, multi-dimensional information. Meaning [is] intrinsic in signal and context when standardized. Minimum of new learning required" (ibid, p. 43).

A related standard, the Aircrew Station Alerting Systems standard MIL-STD-411F (USDOS, 1997), states:

> "Verbal warning signals shall be audible signals in verbal form indicating the existence of a hazardous or imminent catastrophic condition requiring immediate action and shall only be used to complement red warning or other critical visual signals. The verbal warning signals shall be presented at levels that will ensure operator reception under noise conditions in the specific aircraft." (ibid, p. 14)

The Handbook of Human Factors (Salvendy, 1987) summarizes (p. 550) general guidelines for use of speech displays, including, "for listeners without special training in coded signals," based on Deathridge (1972), "situations of stress which might cause the operator to forget the meaning of coded signals", "spoken information should be highly reliable," based on Simpson (1983), and Williges and Williges (1982), and "use speech rather than non-speech to minimize information processing requirements by eliminating the need for decoding non-speech signals."

In practice, we see that speech displays can serve to alert pilots and vehicle operators to hazardous situations and can provide guidance to avoid catastrophe. Aircraft accidents in which cockpit warning systems are activated yet fail to positively affect the outcome are well documented, and dramatic; often, among the final words heard on cockpit voice recorder tapes will be that of the *Ground Proximity Warning System*, offering forth repeated utterances such as *"Terrain, Terrain. Pull Up! Pull Up!"* (e.g., Flight Safety Foundation, 2006). Though speech warnings have undoubtedly prevented many aviation accidents, successful use is harder to quantify as it may not always be reported, though resources such as the US National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System (NASA, 2006) can provide clues (Bliss, et al, 1999). One successful measure of speech-based warnings is the reduction in mid-air collisions due to TCAS (Traffic Collision Avoidance Systems). TCAS alerts, in the form of spoken directives to advise pilots of proximity to other aircraft and to take evasive action, occur frequently in crowded airspace and successfully minimize the chance for mid-air collision, but such systems are not without room for improvement (e.g., Bliss, et al, 1999; Pritchett & Hansman, 1997).

In a modern cockpit the pilot may be called upon to identify and follow unexpected spoken alerts about terrain closure or conflicting traffic within the context of ongoing conversations between crew members, air traffic control communications, and "normal" system generated announcements of flight parameters. Loukopoulos, Dismukes, and Barshi (2009) describe the cockpit environment, and specifically, the high demands placed on flight crews during each phase of flight. Cockpit voice recording transcripts, analyzed by accident investigators, often detail the calm voice of the aural warning system presenting multiple advisories and alerts while conversation continues between members of a flight crew, unaware of the severity of their situation. A USAF multimedia reconstruction of a C-5 crash in Dover, Delaware, in April, 2006, widely available on the internet[8], effectively highlights the events leading to the accident using the actual cockpit audio and the state of flight displays and controls. With highly trained flight crews, the expectation should be that alarms can occur, and if they do occur, safety of the plane and crew is jeopardized if timely action is not taken. Reality indicates that even with highly trained aviators, alarms fail.

As an alternative to speech displays, auditory icons consisting of non-speech sounds have been seen as a viable and efficient means to convey a finite set of warnings or alerts (e.g., Brewster, et al., 1995; Graham, 1999; Edworthy, 1994). To be effective, associative learning of the semantics of individual auditory icons is required, which can limit their usefulness for more general alerting contexts in which extensive training is not possible. Experimental studies show that responses can be learned and retained with a well designed set of auditory icons (Stephan, et al., 2006) and the use of "natural sounds" (Ulfvengren, 2003) shows promise. However, there are examples when poorly designed aural warnings can lead to catastrophe, and pilot surveys (Peryer, et al., 2005) suggest possible changes.

One example of design problems with non-speech auditory warnings is demonstrated in the 2005 Helios 737 crash near Athens, Greece, in which the pilots wrongly interpreted the meaning of a warning signal (AAISB, 2006). The events that unfold dramatically demonstrate how poor alarm design, impaired cognition, and language differences catastrophically interacted. Reason (1990) has described a model of accident causation in which a series of factors line up to provide the opportunity for the failure or accident to occur. Often called the "Swiss Cheese" model (Reason, 2000), any one "slice" by itself, representing an error or other failure in design, management, perception, or action, may not pose a significant risk. But when one slice is combined with other slices (representing a chain of events), the probability that at least one opening on each slice lines up with an opening on adjacent slices can lead to a pathway through all slices. This alignment creates the opportunity for failure to occur and can seal the fate of passengers and crew. The Reason model is not unique to aviation and can be applied to situations that involved interaction between humans and technology. The disastrous response to Hurricane Katrina in 2005 is riddled with aligning holes, particularly with regard to vulnerable populations in New Orleans (SNAKE, 2005).

In the case of Helios Flight 522 (AAISB, 2006), this sequence or chain began first when the maintenance crew inadvertently left the cabin pressure outflow valves in manual mode rather than the normal automatic mode, requiring the flight crew to manually close the valves after takeoff to ensure cabin pressure will be maintained. Second, the pilot and co-pilot were mismatched in terms of language skills and reportedly did not have strong skills in their com-

---

8 YouTube: C5 Galaxy Crash at Dover
http://www.youtube.com/watch?v=fI5xTmmPbsY

mon language of English. Third, unaware of the improperly set outflow valve switch, the pilots did not close the outflow valves. Fourth, and most germane to our discussion, an auditory warning sounded as the aircraft was climbing to cruising altitude. What happens next argues strongly for great care in the design of auditory warnings and a need for designers to think outside of narrowly defined functions. The non-speech warning that sounded could indicate two possible events, dependent upon flight mode: cabin pressurization altitude alert, or improper takeoff configuration. Erroneously, the crew mistook this warning as a faulty takeoff configuration alert and began to debug the situation, oblivious to the fact that the available oxygen was rapidly decreasing. The flight crew may have violated other procedures at this point, such as failing to use oxygen masks. Thus, their behavior may have been adversely affected by both poor communication in English and diminished cognitive function due to hypoxia. As the aircraft was on autopilot, it continued climbing to a programmed altitude of 34,000 feet, with passengers and crew soon unconscious. The aircraft eventually crashed as it ran out of fuel near Athens, Greece.

Why the designers of the aural warning system of the 737 chose to use a non-speech warning signal rather than a spoken message, and then assigned the same signal to two different functions is not known. A non-exhaustive survey of the Aviation Safety Reporting System (NASA, 2006) indicates several pilot submitted reports of the potential confusion arising from this particular signal. In hindsight and particularly in a situation where diminished cognitive function due to hypoxia may lead to further confusion with ambiguous warnings, this seems a clear case where unique, multimodal and semantically rich warnings would be ideal. Assuming that the designers of the 737 warnings system utilized available human factors guidelines, one must ask whether the designers evaluated the warning system as a whole throughout all flight phases, or simply evaluated each warning subsystem independently and only in the expected flight phases. The AAISB (2006) calls for improved training for flight crews, and in August 2010, the US Federal Aviation Administration (FAA) issued a proposed Airworthiness Directive[9] for the 737, calling for changes in crew training and the installation of independent cabin altitude and takeoff warning indicator lights that function in parallel with the auditory tone.

The experience with highly trained professional pilots should serve as a cautionary tale for designers of any alarming system. When considering auditory warnings for the general public, findings that a seemingly obvious fire alarm can be confusing serves as caution to the wider use of non-speech auditory warning signals (Proulx, et al, 2001). This potential for confusion is not new and has been seen and reported in the context of outdoor public warning for disasters. Lachman, et al (1961) studied human behavior during the Hawaiian Tsunami of 1960 and found that significant numbers of the local population either did not hear the warning siren or failed to grasp its meaning. More recently, reports from Japan and the United States describe continuing failure on the

---

9 AD Notice of Proposed Rulemaking, Federal Register: August 11, 2010, Page 48620-48623. Docket No. FAA-2010-0761; Directorate Identifier 2010-NM-069-AD.

part of citizens to respond appropriately to Tsunami warnings, even after the tragic events of the South Asian Tsunami in 2004 should still be fresh in the public memory (Sullivan & Hakkinen, 2006). After tests of a new Tsunami warning siren in Phuket, Thailand, which incorporated both a multi-tone alerting signal and a spoken warning presented in several languages, some members of the public reported confusion as to what the signals meant. A unique challenge described by Sullivan and Hakkinen is the strength of human curiosity, as some members of the public proceed toward the beach at the sound of a Tsunami warning or upon seeing the natural environmental cues (e.g., receding water that exposes the sea bottom) of the impending tidal wave. Anecdotal reports that curiosity seekers and surfers will go to the beach when Tsunami warnings are issued are common.

Many factors can influence the efficacy of warning signals. Without significant training, low occurrence, but life critical alarms may be better presented as a combination of well designed and evaluated alerting signals and speech messages. The longstanding design guideline regarding the use of an alerting cue (speech or non-speech) prior to presentation of critical messages (e.g., MIL-STD-411F, 1999) appears to remain valid though unfortunately often applied in an inconsistent manner by designers.

With the addition of speech displays in automobiles, particularly those which provide guidance on vehicle navigation, new questions and concerns can arise (e.g., Zajicek & Jonsson, 2005; Bliss & Acton, 2003). Pilots are highly trained, yet may miss, ignore (for a variety of reasons including questions regarding the validity of the signal), or misunderstand the auditory display of information. As speech displays continue to move into the general population, the rigorous training regimen afforded pilots is not practical and thus the resulting systems must be intuitive and usable with little or no training. Directives to a driver such as *"Turn right in 50 meters"* or *"Traffic congestion ahead, slow down"* may seem simple in nature but when the presentation is factored in with ambient noise, driving task load (night time, weather, traffic), on-going conversation (with fellow passengers or on a mobile phone), the human capability to process and act upon a speech message may be impacted, as is seen to occur in the cockpit environment (Loukopoulos, Dismukes, & Barshi, 2009). Little data seems available on the actual operational accuracy of in-car navigation systems or even their potential impact on vehicle accidents, in contrast to the more controlled laboratory evaluations of those systems. There is clear evidence that mobile phone usage can result in road accidents (Strayer, et al, 2003) and some research suggests that speech-based applications, such as access to email, can have negative impact on driver performance (Lai, Wood & Considine, 2000). If speech-based applications are to be applied to creating warning systems for a broader audience, as is already the case in some mass notification systems, what are the implications for understandability and actionability of these messages delivered to people in everyday settings?

2.5 Can Human Factors Design Guidelines Help?

Over time, the technology to implement auditory displays has evolved significantly and influenced the usability and user experience with this form of information delivery. For designers of systems deciding to incorporate speech displays, design guidelines vary in age and general applicability, when they exist at all. A review of existing design guidelines and research suggests that fundamental questions remain in the design guidance for effective use of speech in auditory displays and warnings.

Guidelines are part of the toolset engineers and designers use to build effective, usable and safe products and systems. When developing human-system interfaces, a number of sources exist to aid the designer in the creation of effective visual and auditory displays, interaction styles, and physical input or control mechanisms. These sources include The Handbook of Human Factors (Salvendy, 1987), US Department of Defense Standard 1472F (USDOD, 1999), ISO User Interface Standards (ISO, 2006), and the NASA Man-Vehicle Interface Guidelines (NASA, 1995).

In some areas, design guidelines are relatively clear and established and become entrenched as common design practice. Ideally such guidelines will have a solid, empirical basis. However, guidelines should never be treated as absolute design rules as the context of their initial development may not be generalized to wider application. Designers who choose to adopt "industry standard practices" rather than undertake formalized evaluation of the human factors issues in their given application context may thus cause "best practice" to evolve into new or mutated de facto standards or design guidance without empirical foundation in their applied contexts.

In many cases, designers may find no applicable guidelines for a given design problem. Human Factors Engineering and its related disciplines (e.g., usability engineering, user psychology, et al) in such cases can offer insight and methodologies in understanding the abilities and limitations of the user/operator in context, and for the evaluation and refinement of user interfaces through informal and formal iterative testing. Usability evaluation or testing, though recommended, is not always economically feasible or may be limited by the availability of sufficient expertise in the industrial context. As an alternative, designers may intuit new devices or interaction techniques based upon what is known, combined with creative problem solving and common (or uncommon) sense. In an ideal world, usability testing would seek to validate the design in the context of the specific application, but such "research" may not in all likelihood produce generalizable guidelines. Further, the complexity of some systems, arguably an ecosystem of technological components from multiple vendors, may make in vivo, operational testing of individual system elements impractical if not impossible (e.g., modern aircraft cockpits can consist of avionics components from multiple vendors and, in the worst case, lack adherence to a common user interface or human factors principles). This situation is analogous to the concepts of unit testing and system testing in software engineering practice. A 2006 report from the US National Transportation Safety Board (NTSB, 2006) pointed out that aircraft certification, a rigorous engineering and flight evaluation process, does not apply that same rigor to certifying the pilot/aircraft interface. In light of the fact that the majority of aviation accidents are attributed to human error (Boeing, 2006), the lack of rigorous evaluation of the human-system component of aircraft is troubling. If we then view this problem in the context of other warning systems, we are left with a similar conclusion, in that there appear to be no up to date standards or certification for public alarm design that takes into account the availability of newer technologies.

In cases where laboratory evaluations of user interface designs are conducted, one can legitimately question whether the simulation environment can capture the full reality, and concomitant stress, distractions, fatigue, and environmental conditions that can affect the efficacy of a warning system. Simpson (2007) has pointed out the problem in evaluating military cockpit warning signals in the lab and then finding the operational usage of this same signals to result in poor performance. For example, can laboratory simulations model the "perfect storm" (or Reason's "Swiss Cheese" cascade) of the unique circumstances of a tornado that rapidly changes course, ready to strike an urban area without sufficient lead-time for warning, heavy rains filtering the public warning sirens, power outages limiting access to radio and television, and overloaded mobile phone circuits limiting communication. Approximation is possible in the laboratory, but pressure, fear and stress can be difficult to simulate while remaining within the boundaries of ethical research. It is just these difficult to model situations in which errors or omissions in design are compounded. The developers and users of guidelines need to be aware of the consequences of their decisions and endeavor to anticipate the worst case error scenarios particularly in warning system design.

The rate of change for guidelines publications does not appear to keep pace with technological developments. Sometimes established guidelines will be based on the technology of the period in which they were developed and the applicability of the guideline may be called into question when considered in the context of newer technological developments that have arisen in the interim. This *guideline lag* may lead to systems being built today which are based upon guidelines suited to technologies far less capable than what is currently available commercially. Further, designers familiar with the most current technologies may discredit guidelines which refer to older technologies, in spite of the fact that the underlying principles may be equally applicable, though some interpretation and extrapolation by the designer may be required. Mobile phone alerts, for example are being widely implemented on university campuses in the United States following the shooting incident at Virginia Tech, yet there appear to be no clear guidelines, standards, support for those with sensory disabilities or plans for interoperability between products from different vendors (Sullivan, Hakkinen & Piechocinski, 2009). Such developments are driven by perceived need, ease of implementation, and commercial opportunism, and less by a careful analysis of the full problem.

One focus of this section is to take a closer look at a specific area of guidelines, pertaining to the use of speech for information delivery. Based on the alerting nature of auditory signals, the use of speech displays for warning messages has been the focus of significant human factors research for over 40 years. Arons and Mynatt (1994) have cited a lack of applicable guidelines for creating auditory interfaces, primarily in the Computer-Human Interaction context. Stanton (1994) collected papers on several facets of auditory alarms and published it as *The Human Factors of Alarm Design*, and Stanton and Edworthy (1999) have collaborated to release an updated version, *The Human Factors of Auditory Warnings*. Both works highlight key issues in auditory warning design and application and make recommendations for further research. However, despite the existence of both guidelines and empirical research in this area, applications of auditory warnings, as we have seen, appear problematic.

Research to be presented in a later section will present the findings that primary task performance, as well as detection and transcription of warning messages, can be degraded when extensive use is made of speech displays in a human/computer task. The performance degradation observed suggests that using speech output for multiple functions may not be appropriate in all situations. Further research has been indicated to explore the effective design of speech presentation in a multi-function environment to determine whether presentation features can be manipulated to promote detection of and attention to key messages without degrading task performance.

2.6 Guidelines and Public Warning

As spoken alarms have found application in typical commercial or military aircraft cockpits, they have also found use in automobiles, medical equipment, and in some building alarm systems. If designers of public warning systems are to rely upon design guidance originally derived from contexts such as aviation or patient monitoring systems (Stanton & Edworthy, 1999), is there any expectation that low probability alarms, received by diverse members of the public who are not trained in their meaning, will detect, understand, or comply with the protective actions dictated by the warning? Recent surveys of warning research (e.g., Wogalter, et al, 2001; Laughery, 2006) focus on printed warnings, though they mention in passing the existing literature base in auditory warnings and the describe the potential for multimodal warnings that combine sound and image. In examining specific recommendations for speech warnings, we see mention of urgency in the voice and message contents, but no mention of an alerting cue or the recommended acoustic characteristics (e.g. Wogalter, et al 2001; Salvendy, 2006). Yet, we find the alerting cue widely used in practice for a variety of message types (e.g. informational messages onboard trains, public announcements).

The guidelines emerging from the public warning community tend to focus more on generalities than directly implementable design guidance[10]. Leaving design to interpretation also limits the opportunity for standardization. Whereas aviation systems have a high level standardization, public warning technical standards are only in the early stages, such as Common Alerting Protocol (CAP)[11], a technical standard for the delivery of warning information, and CMAS, the Commercial Mobile Alerting System.[12]

2.7 Speech Displays and their Role in Enabling Information Accessibility

When we think of disabilities, we generally focus on the traditional sensory or physical impairments that limit an individual's ability to perceive external stimuli or to undertake physical mobility. Disabilities, of course, cover a broad range of sensory, communicative, physical, and cognitive impairments that limit functional interaction with our physical and social environment. The traditional definition of disabilities can be expanded to include those who are situationally disabled, either through illness or temporary injury, or through functional impairment due to linguistic differences, stress, fatigue, or other factors. Effective design strategies exist for limiting the functional impairments brought about by a disability, primarily through two approaches, augmenting mainstream products with support assistive technologies (Thatcher, 1994), and through universally designed products which are directly accessible to individuals with a disability (Stephanidis, 1995; Hakkinen & DeWitt, 1997).

For people with visual or print disabilities, access to common information such as newspapers and books is limited. Those who are blind have no unassisted access to print material. Those with low vision may be able to read with the assistance of magnifying aids, but often at rate far slower than someone without a visual impairment. Dyslexics and others with learning or cognitive disabilities may be able to see the printed page but not understand the words and sentences present. Prior to the advent of information technology, the visually and print impaired relied upon services such as Braille documents or talking books for the blind, which were essentially full or abridged versions of a publication narrated and recorded onto magnetic tape (DAISY, 2006b).

With the advent of information technology, new challenges, and eventually, opportunities were revealed. The emergence of speech synthesizers would, however, open the door to access to computer-based, digital information by those with visual disabilities.

---

10 One example: WGBH Access Alerts http://ncam.wgbh.org/invent_build/analog/alerts

11 CAP standard is found at http://www.oasis-emergency.org/cap

12 http://www.fcc.gov/pshs/services/cmas.html

One of the earliest applications of text to speech technology was in the development of assistive technologies for the blind (Klatt, 1987). CRT terminals and early personal computers could be adapted, through specialized hardware and software, to transmit the characters within a display's text buffer to a speech synthesizer. This class of assistive technology came to be known as *screen readers*. As personal computer operating systems were transformed from character-based interfaces into Graphical User Interfaces (GUIs), users of screen readers were faced with new challenges (Thatcher, 1994).

Early screen reading systems relied upon access to the textual representation of the application as displayed on screen. Typically, semantic information about what was displayed was not available directly from the application itself, aside from what which could be inferred by screen position or color and usually through the intervention of a sighted assistant. As applications moved to a GUI format, screen reading software could infer some semantics by recognizing dialog objects such as text boxes or menus, but remained largely in the dark about many graphic screen elements. By the early 1990's screen reading software were able to display some application semantics through changes in speech qualities of the synthesizer, such as indicating different screen elements with male or female voices. Today most screen readers have adopted an approach by which they communicate directly with standard, system provided accessibility interfaces to determine application behavior, function, and state rather than attempting to determine this information simply from screen presentation (Blenkhorn & Evans, 2001).

As mobile phones have become small sized personal computers, the same accessibility challenges have redeveloped. In this instance though, the relative simplicity of the mobile phone user interface has facilitated the development of mobile screen readers. However, following the same model as the desktop, the mobile screen readers are an extra cost item that often raises the cost of ownership beyond the means of many who would benefit from the technology.

With the rapid growth of the World Wide Web, further challenges were faced by visually disabled users and screen reading software due to the different ways in which the early Web browsing software rendered HTML content. One successful approach that bypassed the traditional screen reading model was proposed by Hakkinen and DeWitt (1997) in the form of the non-visual browser. This approach viewed HTML as a structured document with embedded semantics that could be used generate and provide navigation into a synthesized audio rendering of the document. The World Wide Web Consortium (W3C) created the Web Accessibility Initiative in 1997 (W3C, 2006b) to lead the effort of adding accessibility features to all W3C Web technologies. The addition of accessibility features would facilitate conversion of Web content into alternate forms such as synthetic speech or Braille. Subsequent developments within W3C would result in the enhancement of the ability to generate synthetic speech from mark-up and utilize digitized speech within Web-content. These developments included the Aural Styles of CSS2 (W3C, 1998), the Speech Synthesis Mark-up Language (W3C, 2004), the Synchronized Multimedia Integration Language (W3C, 1998 & 2005), and the Voice Browser activity (W3C, 1998) which resulted in an open version of VoiceXML.

Concurrent with the development of the Web, Talking Book libraries around the world were faced with the need to transition from the previously mentioned magnetic tape-based talking books to a new digital format. Building upon work in non-visual browsing, synchronized multimedia, and structured document mark-up, the United States Library of Congress and an International Consortium of Talking Book libraries, known as DAISY, the Digital Accessible Information SYstem (DAISY, 2006a), began efforts in 1997 to define a worldwide standard for new generation digital talking books. The first DAISY standard (DAISY, 1998) was published in 1998 and supported structured digital audio books with synchronized text presentation and navigation. A second, more robust standard, ANSI-NISO Z39.86 (DAISY, 2002), was published in 2002 and supported both digital audio and synthetic speech versions of publications. Key features or concepts of the Z39.86 standard have been incorporated into mainstream standards such as Open eBook from the International Document Publishing Forum (IDPF, 2006), and the Consumer Electronics Association's new Audio Book standard (CEA, 2006).

It should be noted that designing for accessibility has historically translated into benefits to those who are not disabled. The common example of this occurrence is the curb cut (DOJ, 2007). People who use wheel chairs would face difficulties when trying to negotiate the transition between sidewalk and the street. Curb cuts were first proposed in 1970 as a means to incorporate ramp-like features at pedestrian crossing points. This feature was of immediate value to those who use wheel chairs but was also soon discovered as a usability improvement for parents who transport their children in baby carriages, delivery persons who use wheeled carts to carry heavy packages, and travelers using wheeled suitcases. Another example of this sharing of beneficial design features is captioning of television programs for the deaf (Ball, 1972), which is used now in many contexts, from persons learning a new language to environments, such as restaurants or other establishments where ambient noise levels may make listening to a common television impossible. With the advent of the Web accessibility efforts, the term "electronic curb cut" came into common use (Bergman & Johnson, 1995), and was particularly appropriate given the frequent use of the term "information superhighway." Many of the accessibility improvements made to the World Wide Web, for example, also have found applicability in areas as diverse as mobile Web browsing and Web-based voice services. It would seem clear that lessons can be learned from the realm of designing for disabilities, particularly from the development of interfaces that meet the requirements of those with visual impairments.

Because the Web and information standards play such a significant role in the public face of government services and in the underlying communications frameworks, the attention to accessibility in the standards processes shows a path to the development of disaster warning and crisis management systems that support the needs of people with disabilities (Harkins, Strauss & Vanderheiden, 2005; Sullivan & Hakkinen, 2006).

2.8 Types and Characteristics of Machine-Generated Speech

If we proceed with the premise that the published guidelines available to systems designers, particularly in the area of speech and auditory display lag well behind the capabilities of current technologies, this author would be remiss to not review briefly the key characteristics of speech technology and describe its present state of development with regard to speech displays. implications for broader application in warning systems in that speech displays are widely incorporated.

2.8.1 Speech Synthesis

Klatt (1987) presents a comprehensive view of the development of speech synthesis. Of particular interest to us are the development of synthesizers designed for control by computer software and systems. The Votrax ML-1 saw early use in the late 1970's and early 1980's for research in speech warning messages (Simpson & Hart, 1977; Simpson & Williams, 1980; Hakkinen & Williges, 1982, 1984). The early synthesizers such as the ML-1 were based upon phoneme synthesis. All speech stimuli (or messages) were required to be coded by hand using a phonetic alphabet and the results tested for intelligibility with target users. Often, iterative hand tuning of the phonemic instructions was required for optimal results.[13]

Following 1980, there was a rapid development of text to speech synthesizers, which effectively provided front-end processing to a back-end phoneme-based synthesizer. This front-end essentially would take text as input and perform a lookup of matching phoneme sequences for the words and numbers contained within the text string. As text to speech developed, these front-ends would incorporate additional rules to handle abbreviations, numbers and commonly occurring numeric formats.

The quality of speech synthesis throughout the 1980's and mid 1990's, though improved, could be easily distinguished from human speech. In most instances, listeners unfamiliar with synthetic speech expressed some level of difficulty in understanding what was being spoken by the synthesizer. Poor synthesis quality of consonants could lead to easy confusion by the listener, particularly in high noise environments (e.g., "flaps" vs. "slats" within the aviation context). Research by Pisoni and his colleagues (Luce, et al, 1983) suggested that synthetic speech required more attention and processing than natural speech, which could adversely affect task performance. More recent research

13 In the Hakkinen and Williges (1984) study, all speech messages were hand coded and tuned using the Votrax ML-1 phoneme set.

has suggests this may remain an issue (Lai, Wood & Considine , 2000), as is the mixed use of synthetic and natural speech in the same application (Gong and Lai, 2003).

The advent of diphone based synthesis, in which the generated speech is actually composed of small fragments of pre-recorded human speech resulted in a significant improvement of speech quality. Though more natural sounding, the speech still had non-human prosody and rhythm and would frequently mispronounce common words, though this could be easily corrected through an exceptions dictionary and improved sentence processing rules. In cases where synthesized speech is used for critical of public applications, hand tuning of the messages and use of unit selection is common (Black, 2002). As of 2010, numerous high quality, diphone based synthesizers are available commercially, though open source synthesizers continue to lag in quality. Synthesizers for mobile devices are now commonplace, with Nokia offering some Symbian-based phones with many North American, European[14], and Asian lanuages. Apple has introduced screen reading technology, VoiceOver[15], in both the iPod music player, iPhone, and iPad[16] making all three devices usable by people who are blind. The added benefit for users of these devices engaged in eyes-busy tasks is evidenced by the marketing of the iPod Shuffle, which does not have a visual display and relies on the VoiceOver spoken interface. Google's Android [17]platform for mobile devices incorporates support for speech synthesis and for screen reading technology for use by the visually impaired.

The one constant characteristic of synthetic speech remains the relatively easy perception by the listener that the speaker is non-human. In the context of synthesized alarm message, this is not necessarily a negative characteristic. From the earliest demonstrations of synthetic speech, the descriptive labels of robot-like, monotone and expressionless were common. Though today's voices sound less robotic and more natural, they still maintain a monotone, if not boring, style of presentation. When synthetic speech is used for information of short duration, such qualities may be less noticeable, but for longer duration content, such as email, documents, or books, it may be more problematic. It can be said that persons with visual disabilities have been more tolerant of the poor quality of speech synthesis and have learned to adapt, motivated by the need to access information that would otherwise not be available.

When comparing speech synthesis to narration by a human, it could be expected that there would be a clear preference for the use of human speech whenever possible. However, it is not uncommon to hear the sentiment expressed among those with visual disabilities that natural speech is only preferred for presentation of leisure reading material whereas synthetic speech is

---

14 http://europe.nokia.com/support/download-software/text-to-speech

15 http://www.apple.com/accessibility/voiceover/

16 Apple, iPod, iPhone, iPad, and VoiceOver are trademarks of Apple Computer.

17 http://accessibleandroid.blogspot.com/

preferred for professional publications and text books.[18] This may seem contra intuitive in that educational materials, particularly in the sciences, math, and engineering can pose significant problems for most synthesizers. If one factors in the research findings (e.g. Delogu, Conte & Sementina, 1998; Luce, Feustel & Pisoni, 1983) that suggest auditory processing requirements for synthetic speech may be greater than for natural speech, there could be significant implications for those who use synthetic speech for learning and knowledge acquisition. There is an inadequate amount of research on this issue and it remains an area in need of further investigation, particularly in a comparative examination of the use of general purpose speech synthesizers, high-quality (constrained vocabulary) synthesizers, and pre-recorded natural speech for both informational (or instructive) and critical warning messages.

Development of Web languages to facilitate improved speech synthesis began in the mid-1990's, primarily as a result of the emerging need of people with visual impairments to access the World Wide Web. The pwWebSpeak browser (Hakkinen, 1996) incorporated a speech styling framework, similar to what would become the AuralCSS properties in Cascading Style Sheets 2.0 (W3C, 1998). Many of the concepts in AuralCSS were first suggested by TV Raman in 1994 in Aster (Audio System For Technical Readings)[19]. Aster was designed to convert LaTeX documents into audio form using speech synthesis, including complex mathematics. Though AuralCSS was integrated in CSS, there has been little visible commercial or open source implementation of the aural aspects of the CSS standard. The Speech Synthesis Markup Language, SSML, (W3C, 2004), developed as an effort to provide a unified mechanism for hinting and controlling speech synthesizers at a time when a variety of vendor specific solutions had been developed. In contrast to AuralCSS, SSML has seen wider adoption and usage.

One remaining challenge in speech synthesis technology is the lack of implementations for many of the world's languages. This is a particular problem for people with visual impairments in developing countries who do not have the ability to read textual material in their local language using assistive technologies. It should be noted that speech synthesis offers significant technical and economic advantages over the alternative of refreshable Braille display hardware, whose total cost may exceed the price of a Linux-based personal computer able to support speech synthesis by 6 times. For the developing countries, speech synthesis is the only economically viable way to provide access to digital information for people with visual disabilities (as well as those who may be illiterate).

---

18 The author interviewed visually impaired users of speech synthesis while engaged in the development of the DAISY talking book standards. Empirical studies examining the reported TTS and natural speech preferences have not been conducted.

19 http://emacspeak.sourceforge.net/

2.8.2 Digitized Speech

Historically, recording of sounds for later playback began in the late 1800's and by 1900 the phonograph was becoming a recognized appliance for listening to music. The core technology evolved from simple wax covered drums and wire recorders to the phonograph disc. Other inventors discovered the concept of magnetic recording and by the 1930's the tape recorder was in production. These technologies were all analog-based, meaning that the audio signals were transformed into a continuous electrical waveform stored either via etchings on physical media or electrically onto magnetic tape.

The practical implementation of digital recording of sounds emerged in the 1970's with commercial and research initiatives. Driven primarily by the music industry, recording of digital masters of musical performances became common by the late 1970's, though the delivery to consumers remained in the form of magnetic tape or vinyl records until the advent of the Compact Disc in the 1980's. Digital audio recording found significant use in psychoacoustics research, leading to development of specialized hardware (Cox, et al, 1975) and software (Hakkinen & Engebretson, 1978) for the recording, editing, and programmed playback of digital sounds. One limitation in these early applications of digital audio were the storage requirements of the digitized sounds themselves; however, as memory costs dropped and capacity rose, the use of digitally recorded audio became widespread. One of the earlier perceived advantages of text to speech technology was that the memory footprint of a typical speech synthesizer was relatively small when compared to the storage requirements and costs of digital audio recordings of human speech. Today, mobile audio devices, such as the ubiquitous iPod contain tens of gigabytes of storage and can contain hundreds of hours of spoken audio or music. It should also be noted that as speech synthesis quality improved and moved beyond algorithmically generated speech sounds and toward databases of actual human speech diphones, the effective footprint of the synthesizer has grown to occupy sizes reaching 1 gigabyte.

Today, designers of systems requiring speech display can choose between using synthesized or natural speech, dictated by the needs of the application. For applications with relatively fixed or static information, or messages that can be easily constructed from a set of well matched pre-recorded fragments, digitized natural speech is a viable solution, and one that may gain more consumer acceptance when compared to even the best quality speech synthesizer. For applications which render arbitrary textual information into speech the only practical solution is speech synthesis, though in the context of emergency notifications and alarms, the vocabularies required to define many warning and protective action messages may actually comprise a small vocabulary, and thus allowing the potential use of limited domain, high quality synthesis.

2.9 Alarms of Convenience, Alarms of Necessity

In our discussion of alarm and how they may be designed and implemented, we need to differentiate between alarms that provide critical, life safety information with those that only serve to assist us in our daily living, such as an alarm clock, an egg timer, a signal indicating new email has arrived, or a phone ringer. I term these latter alarms of convenience. In contrast, alarms of necessity are those that signal imminent risk or harm, such as a smoke alarm, fire alarm, warning siren, or heart rate alarm. As a result of poor selection of alarm signals, it can be difficult to distinguish alarms from one another.

A typical kitchen may have a standalone kitchen timer, a microwave oven, conventional oven, and other devices which may generate one or more alarms, signaling end of a time interval, end of a cooking interval, or that an oven has reached a certain temperature. These are convenience alarms, and, generally are comprised of a single or repeating beep. In a typical home, you may additionally find other alarms of convenience, such as a washing cycle complete on a clothes washer, or door bell. In the home most, if not all alarms of convenience are auditory. In general, there has been little innovation regarding design of convenience alarms, with the exception of the alarm clock. Though clock radios have long had the ability to awaken individuals from sleep with either the radio or a buzzer, the customization of the simple alarm clock has begun to evolve through the use of digital technology. The Apple iPod, for example allows some customization of the alarm sound from a variety of predefined signals (see FIGURE 1. Alarm Clock preferences on Apple's iPod.)

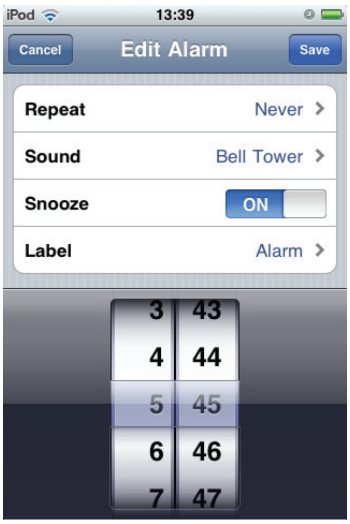

FIGURE 1. Alarm Clock preferences on Apple's iPod.

Beyond setting the alarm sound, mobile phones such as iPhone also allow specifying whether a vibration pattern should occur simultaneously with the alarm sound. Alarm clocks for those who are deaf have included vibration capabilities as well as the ability to trigger a flashing light. Recognition that an alarm clock may not always waken the individual sufficiently to actually arise from the bed (e.g., the snooze button syndrome) led to the development of the "Clocky," a class project at MIT by Gauri Nanda.[20] Clocky will leap off a bed side table is the alarm is not responded to, and will roll about the floor until the alarm is shut off, forcing the individual to rise from the bed and chase down the clock.

Alarms of necessity have shown less innovation, sometimes constrained by specific engineering codes that require fire alarm designers, for example, to adhere to strict design guidelines. The adaptation of traditional fire alarms for those with hearing impairments is one example of limited innovation, through the addition of visual indicators, such as flashing strobe lights (see FIGURE 2.).

FIGURE 2. Audio/Visual Fire Alarm incorporating strobe light and buzzer.

2.10 Chapter Summary

In this section, I have described alarms from both a natural and artificial perspective, and then built upon the technology of artificial alarm delivery. Contrasts between speech and non-speech alarms are presented, along with failures that have resulted from inconsistent or poor alarm design.

I next introduced the topic of information accessibility, and in particular an examination of technical solutions and standards for enabling access to information by people with sensory disabilities. The disaster warning communi-

20 http://alumni.media.mit.edu/~nanda/index2.html

ty has recognized the importance of multimodal delivery, more specifically in the context of multiple channels of message delivery (e.g., sirens, radio, TV, and telephone). In contrast, the accessibility community recognizes the importance of multiple modality within a given channel of delivery to ensure population with sensory or cognitive limitations have a higher likelihood of receiving an alarm (Hakkinen, 2007; Brooks, 2006; Sullivan, et al, 2008). These two views of multimodal warning contrast the need to reach members of a community at risk who exhibit differences in how they acquire information from an external source with the need to enable the acquired information to be perceived and understood by the recipient irrespective of their abilities. Information displays including speech and multimodal presentation will be an integral part of effective and accessible warning systems.

The developments that have led to information accessibility standards, enabling of accessibility within technical standards such as HTML and XML, the inclusion of digitized and synthetic speech capabilities across a range of devices, and the enabling of accessibility features within many personal communications devices such as mobile phones sets the stage for later discussion of personalized alarms.

In the next chapter we begin to define the psychology of human response to alarms. Alarm response is, in practice, a chain of sensory, perceptual, and cognitive processes.

3 HUMAN RESPONSE TO ALARMS

What are the underlying mechanisms that humans (and other species) rely upon in detecting and responding to alarms? And more importantly, how do these mechanisms play a role in our failure to respond to alarms. Our examination will undertake a hierarchical approach, mimicking the pathway from the sensory level to the higher cognitive processes that lead, ideally, to successful alarm response. At each level of the pathway, we will discuss the idea of filters, both those that allow a significant signal to pass to the next stage of processing, and those filters that are efficient in halting the progress of a significant signal. Some of the filters operate within the spectrum of the physical characteristics of the signal itself, later stage filters operate at the semantic level, and final stage filters operate on the associative knowledge level.

3.1 The Alarm Processing Model

In simple form, we might classify the alarm processing model as *Detect*, *Perceive*, *Recognize*, and *Respond*, as shown in Figure 3. The model is far from simple, however. From the perspective of psychological science, the processing model is in effect a layering of processes and knowledge that will determine whether an alarm will be successfully acted upon. Further, the model and processes are themselves influenced by internal and external factors, including the ability and state of the alarm receiver and the environmental factors that may influence all aspects of the processing model.

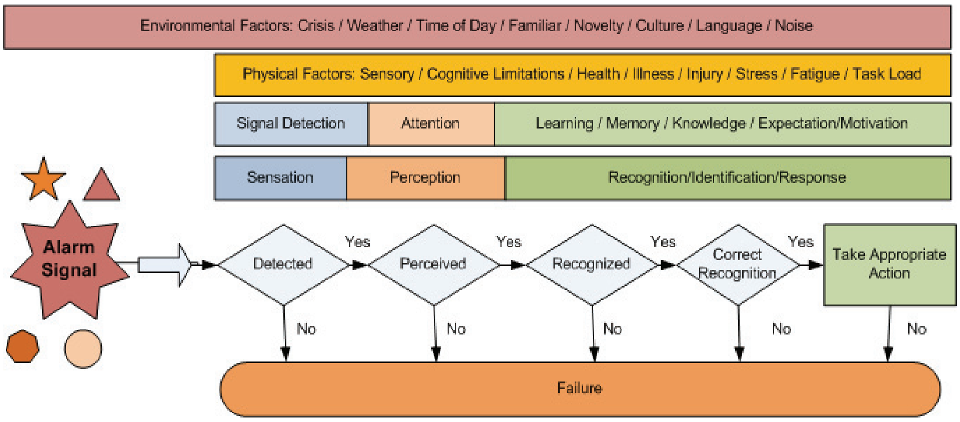

FIGURE 3. General Alarm Processing Model

3.1.1 Detection

When an alarm is generated, it emits information in the form of acoustic, light, and/or kinetic waves. If the information is within proximity of the human receiver, and the frequency, intensity and the duration of the informative event is sufficient to be detected by one or more of the human senses, we have arrived at the first stage of alarm processing, detection. Thus, for signals to be detected, they must be within range of the sensory receiver, distinguish themselves from the background noise existent in that modality in order to become an event or signal of interest.

An auditory alarm tone that is presented at a volume level below the ambient noise in a crowded rail station risks being undetected by the intended recipients. A visual alarm signal, out of the field of view of the human receiver, risks being undetected, unless reflected light from the source may capture the attention of the viewer. The tactile, vibration alarm of a mobile device in the pocket of a motorcycle rider on a country road risks being undetected due to the competing, inherent vibration induced by the road surface. Other factors can mask the presence of a signal: a sensory impairment such as blindness or deafness or self selected sensory filters such as hearing protectors or MP3 player ear-buds. Environmental events, such as heavy rain or wind can mask the sound of outdoor sirens (Lachman, et al., 1961).

3.1.2 Perception

Assuming the alarm signal has been detected, perceiving the signal as an event that should receive attention and processing is the next step. Attention plays a significant role in the process of orienting to the presence of a potential signal, detecting signals for focused, or focal, processing, and then maintaining vigilance, or a state of heightened alert, for subsequent alarm signals (see Figure 4).

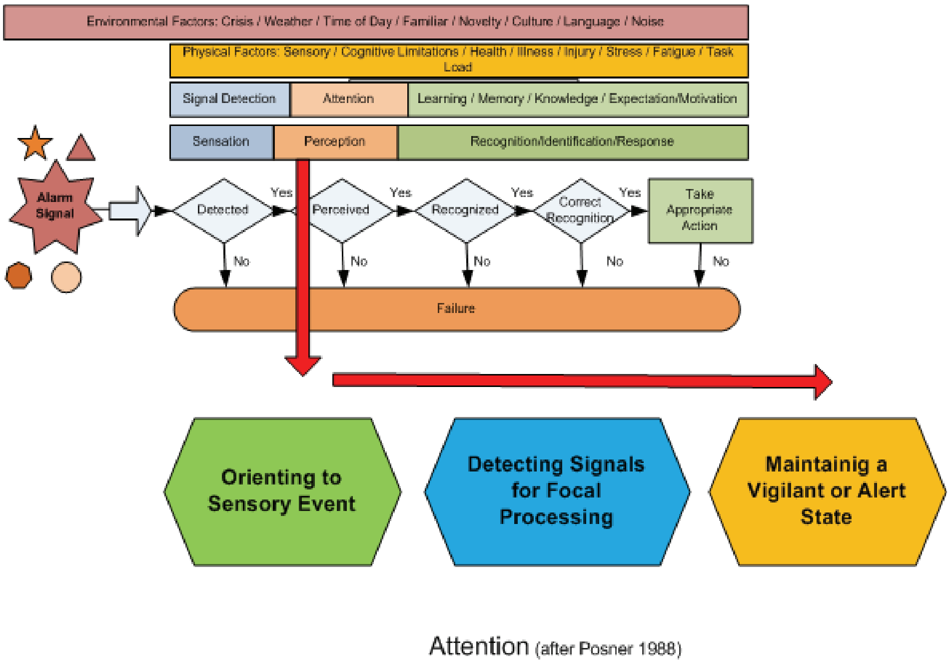

FIGURE 4. Posner's Attention model overlaid on the early stages of alarm processing.

If attention cannot be maintained, the alarm can be lost amongst other signals, competing information, or noise.

3.1.3 Recognition

With attention now on the alarm signal, existing knowledge is used to attempt recognition of the signal. At this stage, prior experience or learning comes into play to determine whether the signal is something we already understand. If the signal is recognized as an alarm, either through direct identification of a previously learned signal, or through generalization, inference or association, the question remains whether this recognition is correct.

3.1.4 Response

If the recognition is correct, the human receiver can either use information conveyed as part of the alarm signal, such as spoken direction for reaching safety, to take protective action. Alternately, if the human receiver is familiar with emergency protective actions, then the alarm signal serves as a trigger for activating the appropriate response. If the alarm signal is not recognized, or incorrectly recognized, the human receiver may ignore the alarm signal and shift attention to another task, or undertake incorrect or inappropriate actions.

3.2 Factors that Impair Alarm Processing

Throughout the stages of this model, physical and environmental factors can impair the processing of an alarm. As mentioned previously, sensory or cognitive disabilities can impact detection, perception and recognition of an alarm. Cognitive impairments, such as memory or language disorders can result in failure to recognize or understand an alarm and also inhibit recall of appropriate actions. Fatigue, acute illness, stress, or workload can all further impair the ability of the human receiver to process an alarm.

Environmental factors further may impact alarm processing. The occurrence of an actual alarm is unpredictable, and by its nature, is unexpected. Weather, time of day, perceived likelihood of occurrence can all impact the human receiver's ability to process an alarm. In unfamiliar locations, such as might be expected by a tourist in a foreign country, language, culture, environmental cues and ambient "noise," expressed in a visual and auditory clutter of unfamiliar sights, signage, text, and language can negatively impact the human receiver's ability to detect and process an alarm.

The above model can be compared to other warning response models that have been developed (e.g, Wogalter). As we will see in later sections, current models focus little on the early stages of the processing model. Without understanding the significance of the early stage, warnings may never go beyond the role of background noise in an environment that grows richer in visual and audio signals competing for our attention.

3.3 The Orienting Reflex

Early research by Russian physiologists, notably Ivan Pavlov (1927), identified the Orienting Reflex (OR) as a key process in evoking response to changes in the environment or novel stimuli. As described by Sokolov (1963):

> It is this reflex which brings about the immediate response in man and animals to the slightest changes in the world around them, so that they immediately orientate their appropriate receptor-organ in accordance with the perceptible quality in the agent bringing about the change, making full investigation of it. The biological significance of this reflex is obvious. If the animal were not provided with such a reflex its life would hang at every moment by a thread.

The OR would appear fundamental to the human ability to react to the presence of an alarm signal. Though there has been some controversy regarding whether the foundation of the orienting response is stimulus significance or stimulus novelty (e.g., O'Gorman, 1979), the fundamental role of the OR in alarm response is clear.

3.4 Habituation

Any discussion of the Orienting Reflex (or alarm response) is correctly followed by the introduction of habituation. In the most basic definition, habituation is the reduction and eventual disappearance of a response to a stimulus as a result of repeated stimulus presentation. Simple stimulus response pairings, such as a startle response to a loud noise, can easily demonstrate habituation through repeated presentation of the noise. Repeated, false, fire alarms can result in habituation, with tragic consequences when the alarm is no longer false (e.g., the Seton Hall University dormitory fire in 2003). The basic concept of habituation manifests itself as one of the significant components of alarm failure. The repurposing of alarms, such as a common signal for severe weather, national security alerts, or volunteer fire alarms can also lead to a form of habituation, given the low probability of severe weather or national security problems, and the higher probability of a local fire or accident. The assumption is that any alarm heard is intended for the most likely event, and can be ignored as it is not a threat to personal safety.

3.5 Signal Detection

A common process for all humans is to attempt to detect a signal of interest from a background of noise. Signal Detection Theory was first described by Tanner and Swets (1954) and arose from the study of radar and sonar operators, engaged in detecting targets accurately from a background of auditory or visual noise. The Signal Detection model can be described in terms of correct detections and false alarms, as depicted in a Hit/Miss table (see FIGURE 5. Signal Detection Hit Miss Table).

	Not Detected	Detected
Signal Present	Miss	Correct
Signal Not Present	Correct	False Alarm

FIGURE 5. Signal Detection Hit Miss Table

Though arising from a specialist user context, the SDT paradigm can be operated in any human context, and its role as a foundational component in generalized alarm response appears overlooked. Traditional SDT experiments have used visual or auditory presentation of signals in a controlled laboratory environment. But now assume a real world scenario, a crowded Intercity train travelling to Helsinki. In that context, some percentage of mobile phones will have the default Nokia ringtone. If your phone, too, has that default ringtone, then whenever a mobile rings in you vicinity within the train car, you may first assume it is yours, you may attempt to localize the source (your pocket or purse), and then if you believe the phone to be yours, you reach for it and answer, only to discover it was not your phone. From an SDT perspective, the signal was there, it simply wasn't yours. In public spaces, where there are multiple personal devices, each associated with a user of the device, identical signals make the signal detection task a more challenging problem. The common feature of ringtone personalization can mitigate the challenge, assuming the user of the phone is able to locate and operate the ringtone customization feature.

3.6 Attention and Vigilance

The ability to detect alarms implies that we have sufficient attention available to determine that a significant signal or stimulus is present. For example, a chickadee is likely to be scanning for potential cues indicating the presence of a predator, though such scanning is built into other life sustaining activity patterns. Specific cues, such as a shadow moving against the surface of the ground, can trigger the appropriate threat call. Maintaining sustained attention for extended periods is a cognitive function termed vigilance. Wickens (1984) defines vigilance as: "… a state of readiness to detect and respond to certain specified small changes occurring at random time intervals in the environment."

As humans, attention and vigilance are fundamental processes, though active vigilance is a specialized function, with periods of heightened attention that may not be sustainable for long periods without loss of sensitivity to the target of interest. Vigilance tasks are generally those that require active monitoring of a process or space, involving visual and /or auditory signals, and as might be expected, task performance can degrade over time. The countless individuals serving as security screeners at airports are engaged in a vigilance task, continually scanning an x-ray image for the presence of suspicious objects or weapons and research suggests that errors in detection may be higher as the probability of a target being present is lower. Ideally, for the airport security screeners, some level of actual (or training) targets could improve overall detection chances.

If the task of the professional engaged in vigilance is difficult, what then for the untrained, casual participant in an environment in which low probability alarms may be present? Add further that unlike the vigilance task of the professional, the ambient environment of our participant may contain a variety of visual and auditory distractors, from which our non-professional will have to detect a relevant signal. We live in a stimulus rich environment, and must learn to function with high levels of distraction. Assume, further, an environment with many ongoing voice sources or conversations, and that the warning designer followed human factors guidelines that warnings be announced in an unemotional, monotone voice. What are the chances for signal detection, when compared to an amplified voice, with emotion, exclaiming "The building is on fire! Evacuate now!" Fortunately, research by Edworthy, et al (1999) has explored the use of urgency and signal words for auditory warnings. Edworthy and her colleagues have found that urgency in spoken messages can increase detection and reduce response time.

Early research in auditory attention (Cherry, 1953; Broadbent, 1958) focused on the ability of a listener to identify and follow a single conversation in the presence of other ongoing speech activity. This was described as the "cocktail party" effect, and allows us to focus on a single voice within an ongoing sea of conversation. Broadbent (1958), in studies of dichotic listening, proposed the idea of the early filter model. In this model stimuli not deemed of interest are discarded before reaching our awareness. Only stimuli of interest, for example, a specific voice, is passed along for further processing and interpretation. Late stage filter models, as proposed by Deutsch and Deutsch (1963) and Norman (1968) theorized that all stimuli are processed and only filtered once semantic information is available to guide selection. Treisman has proposed models of selective (Treisman, 1969) and divided attention (Treisman & Davies, 1973), identifying that performance in multi-channel attention is better between modalities rather than within a single modality. Wickens (1984) discusses the interaction of divided attention, time sharing, workload and task type (verbal/spatial) and the resulting effects on task performance of mixed modality interaction. The clear guidance in alarm design and interaction is to understand how the modality of interaction can impact alarm receipt.

Attention clearly plays a role in our ability to detect and process information, and in turn our ability to respond to alarms. Models of attention can explain, at least in part, why alerts sometimes miss their intended audience.

3.7 Motivation to Respond

Failure to respond to a warning or alert can be due to many of the factors mentioned earlier, but once an alerting signal is detected and semantically identified as a warning, there may still be a failure to act. For example, when voice warning systems were first developed for aircraft, pilots may find that they trusted their own instincts or skills more than they trusted a disembodied voice giving guidance and potentially perceiving this warning system as second guessing their piloting skills. False alarms, though less common today, have led to pilots who turn off warning systems, in their assumption that the alarms generated may be in error. Complacency and lack of discipline within the cockpit can lead to warnings or status announcements that are ignored or entirely missed by members of the flight crew in spite of crew resource management training. In the final analysis, as in the case of the recent C-5 accident, "human error" is often to blame when warning and displays systems functioned normally (US Air Force, 2006). Though human error in the cockpit can be minimized through good design, training, cockpit resource management and a culture of safety, the specific problem of complacency may require more assertive measures on the part of cockpit warning systems.

We can also question the motivation to prepare for and develop awareness of low probability hazards. Tourists in Phuket, Thailand on the morning of 26 December 2004 had no expectation that Tsunamis posed a risk to their personal safety. Those closest to the epicenter of the undersea earthquake in the Andaman Sea had minutes to respond to the threat. Since there were no organized warning systems in place, only those who recognized the natural warnings signs were able to warn others and flee to safety (Sullivan & Hakkinen, 2006). On a visit to Phuket in May of 2005, a newly installed Tsunami warning system was being tested, with tourists expressing either confusion or a lack of interest in the system. A political cartoon in the local newspaper at the time depicted tourists being motivated to evacuate to high ground with the promise of free beer.[21] It seems that personal safety was not sufficient motivation to interrupt a holiday with participation in an evacuation drill. Is it a question of poor warning design that does not effectively state the risk in an effective manner or simply a case of limited motivation by those who believe a Tsunami will not occur again? Or is it a sense of resignation, as suggested by Forman (1963), in that in the case of a Tsunami, there is no safety. False, or inaccurate Tsunami warnings can increase the potential for future mistrust of such warnings (Davis & Izadkhah, 2008). With transient populations such as tourists, refugees, and migrant workers, the challenge is to develop effective warnings that require little or nothing in the way of formal training. Addams-Moring (2007), for example, has suggested mechanisms to improve trust of mobile device delivered disaster alerts, and Sullivan, et al (2009) have examined the issue of trust in campus notification systems. If we assume that some percentage of any population will either not prepare or not be familiar with the risk and warning signals, it is essential that any alarm system convey, in a trusted manner and clear manner, as much information as possible to convey the nature of the hazard and the protective action required.

3.8 Sensory and Cognitive Impairments

Failure to detect and respond to warnings can result from chronic or temporary impairments in our sensory systems. Those who are blind or have impaired vision will not respond to visual warnings, while those who are deaf or have impaired hearing will not respond to auditory alarms. For example, a deaf tourist in Phuket will have no direct cue that a Tsunami warning siren has been triggered, other than visual cues if other beach goers respond to the alarm. The impact of Dyslexia and other reading disabilities are often overlooked in context of information displays (Sullivan & Hakkinen, 2007). Complex graphics may well convey detailed instructions as to evacuation routes and safety points, but can be a significant challenge to those with visual and reading impairments.

Aging members of society can exhibit sensory, cognitive and physical limitations that can impair their ability to detect, understand and respond to warnings.

Stress should not be ignored as a contributor to impaired cognition, particularly in times of crisis. Bourne and Yaroush (2003) provide an excellent summary of research related to stress, cognition and human performance. The

---

21 Phukette Gazette, 7 May, 2005, pg. 18.

dynamic nature of an emergency, the need to make rapid decisions, potentially of a life critical nature for the individual, or those in the individual's care, environmental conditions such as smoke, dust, or chaos can all contribute to stressors that impact cognition and ability to respond.

As has been discussed in the prior chapter, assistive technologies for those with disabilities have tended to focus on the specific needs of people with sensory impairments. Converting written messages into spoken form is achievable and ensuring that audio and video messages contain textual captions done regularly in broadcast television and cinema. Less well defined are effective approaches to communication for people with cognitive impairments. Approaches such as the use of photographs to replace textual menus on mobile devices is one example (Vanderheiden, et al 2004) for aging people with memory impairments, while the easy to read movement in Sweden[22] is developing print and online materials that are easily understood. Then challenge is transferring what works for people with disabilities into the development of public warning technologies.

3.9 Warning /Alarm Response Models

Several models have been proposed that can be used to describe how alarms are responded to in the context of risk communication and safety. In general these models have been developed for describing response to visual warnings, but in some cases include reference to auditory and multimodal warning. All models embody elements described thus far in terms of human response to alarms, and in essence are highly similar. With the exception of the work by Wogalter (Conzola & Wogalter, 2001), the models appear to exclude explicit pathways for alarm failure.

3.9.1 Wogalter's C-HIP

Wogalter and his colleagues have developed the Communication-Human Information Processing (C-HIP) model for warning and risk communication (Conzola & Wogalter, 2001).

[22] Centrum för lättläst. http://www.lattlast.se/

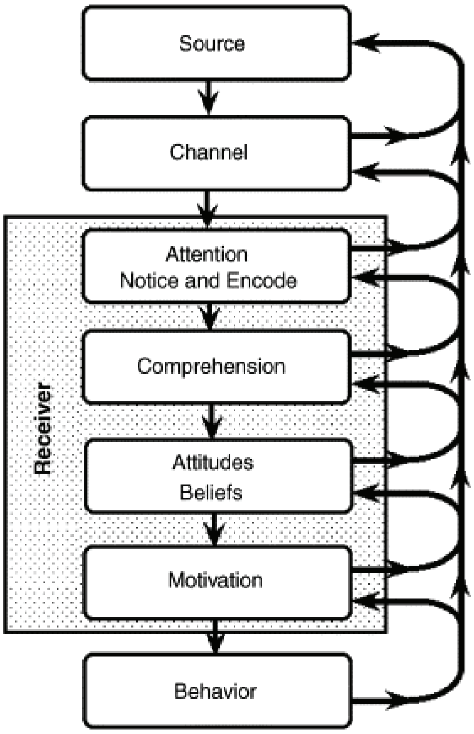

FIGURE 6. Wogalter's C-HIP model, from Conzola & Wogalter, 2001.

In Wogalter's model, the receiver (human), is presented with a message containing risk or other critical information. The source of the information may be a government agency or news media and the channel conveys how the information is delivered (e.g., siren, radio broadcast, signage, warning label). The entry point of the message to the human is gated by the "Attention Notice and Encode" block. If the message fails to enter the receiver or the user fails to attend to the message, no further processing is possible and the receiver may seek/receive information from other channels. C-HIP shares similarities with this author's model in Figure 3. In the original presentation of C-HIP, descriptions of the factors that can impair alarm receipt are not addressed. Subsequently, Wogalter's chapter in Salvendy (2006), revises the C-HIP model to include factors limiting potential receipt of a warning, such as blindness and deafness, splits attention into two parts by adding "maintaining attention," and makes reference to environmental factors that can impact warning receipt and comprehension. The C-HIP model serves as a framework for a comprehensive set of guidelines in Wogalter, et al (2002) and Salvendy (2006). Much of Wogalter's guidelines come from work on visual product warning labels, with specific detail on visual layout, use of symbols, and typography. Auditory warning guidelines are more limited, in that reference is made to the work of Edworthy, et al. (2003) on the use of urgency in the creation of non-speech and speech warnings.

Further, speech is cited as an effective means of communication, yet there is no discussion of synthetic vs. natural speech messages. The issue of synthetic speech quality vs. natural speech is significant in auditory warning design, based upon existing guidelines, but one that has not seen significant empirical research in the past decade. For example, the recommendation based on Simpson and Williams (1980) and Hakkinen and Williges (1984) that synthesized warning messages may not require an alerting cue because the unique vocal qualities of synthetic speech serves as its own alerting function. However, the quality of synthetic speech production has improved dramatically since the 1980's and the quality of the synthesized utterance is much closer to human speech, thereby losing any alerting quality.

3.9.2 The relationship of Human Error models to alarm failure

Any failure by a human to respond to an alarm may be termed an error. Reason (2000) has defined the Generic Error Modeling System (GEMS) as a way to describe error occurrence from a cognitive processing perspective. GEMS focuses on the errors that occur when an individual fails to carry out a required action, either by performing a task or plan incorrectly, or choosing the wrong plan carryout.

In GEMS, actions are defined as:

- Skills-based action -- Routine actions performed without significant cognitive effort
- Rules-based action -- Actions based on a set of rules or procedural steps
- Knowledge-based action -- The application of existing knowledge (experience) to perform a new/novel action

GEMS identifies three types of error that can occur in the context of the actions:

- Slips – Errors in which the user's intent is correct but actions are not carried out correctly
- Lapses – Errors in which the user forgets a critical step or loses track of steps in a procedure
- Mistakes - Errors, termed planning failures, in which the user carries out an inappropriate set of actions

In the context of alarms, we can associate expected responses to various categories of alarms using the GEMS action model. All of these actions imply some form of training or experience, as would be appropriate to learning of responses to various forms of hazards signaled by an alarm. For errors to occur in alarm response, the alarm must first be detected. The fundamental failure to respond to an alarm signal after detection, through failure to perceive and attend to the signal, or through incorrect recognition and action selection, can be characterized as a mistake in the GEMS model. Once the alarm has been recognized and an action response initiated, we can then anticipate the possibility of both slips and lapses. For the work, of this thesis, we are generally concerned with the detection and successful recognition of the alarm; slips and lapses, though of interest, generally occur at the later, response stage.

3.9.3 Importance of Ambient Environmental Factors in Models

Like most disaster drills, which take place with advance planning, generally in the daytime, and during good weather, alarm models and warning guidelines do not appear to account for environmental factors that may adversely affect the performance of the human component in the "real world" situation. Emergencies and disasters operate on a random basis, occurring during any time of year, day, and weather. Storms and torrential downpours, common in some locations also prone to natural disasters can mask the sound of warning sirens, and disrupt electrical service, hampering television and radio communications and make evacuation or emergency response difficult.[23] Models which do not fully take into account environmental factors can lead to warning systems that may not function effectively in challenging environments, which may be more typically found in disaster prone areas.

3.9.4 Importance of Individual Differences in Models

The concept of Individual Differences, in its classic sense, means that differences are understood to exist across individuals in dimensions such as intelligence and personality. In the context of disability and accessibility, individual differences can manifest themselves in terms of sensory, cognitive or learning impairments. Warning models in the literature have paid little specific attention to populations with special needs. While some mention is directed to the information needs of populations such as the blind and deaf (e.g., Wogalter's chapter in Salvendy, 2006), effective strategies for including such populations in mainstream warning have been limited or non-existent. In reviewing the literature, no reference is made to the guidelines for information accessibility, such as those developed by the World Wide Web Consortium (W3C, 2007), which, for example could provide guidance to warning system designers. Efforts to raise awareness and identify the information requirements of special needs populations have been undertaken by Brooks (2006), Sullivan and Hakkinen (2006) and a field trial of accessible mobile alerts was undertaken by the Wireless Rehabilitation Engineering Research Center at Georgia Tech.[24]

---

23 On 23 October 2004, this author was in Japan during Typhoon Tokage, and experienced the Niigita earthquake which occurred contemporaneously. Though the quake epicenter was distant from the core typhoon, ground movement was felt in my location, and Tsunami warnings were issued with the resulting wave eventually posing little threat.

24 Wireless RERC at Georgia Tech. http://www.wirelessrerc.org

3.9.5 Old Media Models vs. New Media influence in Models

Media consumption models have changed dramatically with the advent of the internet. A growing segment of the global population[25] no longer relies upon traditional print or broadcast media for their news. Most discussion of public warning and risk communication (e.g. Mileti & Sorenson, 1990) place significant emphasis on the use of newspapers and radio/TV to convey preparedness and emergency information. Today, we find that many individuals do not consume traditional media, and may, at times, be disconnected from local or national media which may convey broadcast alerts. These individuals may be connected with friends, co-workers, or social networks, often via a mobile device. Authorities and media must adapt to mobile device delivery and social media, and recognize that individuals may not be connected to any traditional alarm delivery channel. This concept of disconnection must also be recognized in developing regions of the world prone to disasters. During the Mumbai terror attacks of November 2008, reports of individuals on the ground in Mumbai received updates and news via Twitter, SMS and phone calls from friends and colleagues outside of India, who were watching the events on cable news networks.[26] Effective and timely alarms may not only come from authorities, but from members of an individual's social network (Palen & Liu, 2007).

---

25 Pew Media Survey 2009. http://pewresearch.org/pubs/1151/state-of-the-news-media-2009

26 Forbes. Mumbai: Twitters Moment. http://www.forbes.com/2008/11/28/mumbai-twitter-sms-tech-internet-cx_bc_kn_1128mumbai.html

4 SUMMARY OF PAPERS

In arriving at this point, we have now set the stage for a discussion of papers that can contribute to an effective and human centered approach to effective warnings. We have discussed the problem of alarm failure, examined the nature of alarms from both a naturalistic and technological viewpoint, we have examined the technologies associated with warnings and begun an exploration of the approaches that can make information accessible to people with disabilities. We next examined key aspects of how humans process information about alarms, from initial detection through several models of alarm processing.

Now we will examine seven papers from four thematic areas:

1. Speech-based and Multimodal Warnings
2. Accessibility Standards and Multimodal Information for People with Disabilities
3. Communicating Disaster Preparedness and Warning Information to People with Disabilities
4. Use of Mobile Phones for Emergency Notification

In the discussion of these papers, we will introduce the concept of adaptive presentation of audio information, based on the influences of the assistive technology domain, as a way to potentially address the problems arising from the application of speech displays in domains where human capabilities are challenged by the operational contexts.

4.1 Speech and Multimodal Warnings

In the timeframe of 1980, designers of human-system interfaces were able to begin considering the use of synthesized speech as a means of conveying information to the users of their systems. Turning to the human factors handbooks of that period, such as MIL-STD-1472 (USDOD, 1999) and Van Cott and Kinkade (1972) guidelines regarding speech displays were based upon natural speech. A key guideline in those hand books would state that if a speech message is used to convey a warning, it should be preceded in time by an alerting cue at 1000 Hertz, with duration of 500 milliseconds, and with 500 milliseconds of silence to avoid forward masking of the following message.

Simpson and Williams (1980) called into question the need for an alerting cue, and found no available empirical evidence to support the guideline. They went on to report a study in which a phoneme-based synthesizer was used to present six different warning messages within a flight simulator cockpit. The synthesized warning messages were presented with and without alerting cues in the context of a simulated flight with normal voice communications taking place. Their findings showed no significant difference in detection or accuracy of response to either cued or un-cued warnings. What they did observe, however, was that the presence of the alerting cue served to delay response to the warning by a duration approximately equivalent to the cue (500 milliseconds) plus the silent interval to limit forward masking (500 milliseconds). In the context of high speed flight, response delays of 1 second could have significant consequences. Simpson and Williams went on to recommend that the alerting cue appeared to serve no useful purpose, and that the unique quality of the synthesized speech message served as its own alerting function.

4.1.1 Speaker Uniqueness and Cues

In the Simpson and Williams (1980) study, synthesized speech was used only for warning messages, assuring that any synthesized voice heard would be one conveying a warning. What would happen if synthetic speech was also used for lower priority or informational messages, thereby losing the quality of uniqueness for warning messages? In article I, Hakkinen and Williges (1984) explored this question within the context of a simulated air traffic control environment, incorporating a multi-function speech synthesis system. Subjects in the Hakkinen and Williges study played the role of an aircraft carrier-based air traffic controller, marshalling arriving aircraft into the approach pattern and guiding them, ideally, to an eventual landing.

The Hakkinen and Williges study had two goals, to replicate the findings of Simpson and Williams within a single function synthetic speech warning system, and then go on to explore whether the Simpson and Williams finding would also be applicable to a multiple function speech synthesis system. A phoneme-based speech synthesizer was used to present emergency messages and aircraft status, as well as controller requested information. Workload in the task was varied by altering the number of aircraft under control of the subject. The findings by Hakkinen and Williges within the single function speech environment indeed confirmed that the alerting cue only served to increase the response time to the warning by the approximate length of the cue complex itself and did not significantly add to the user response time to the message.

In the multiple function environment, where synthetic speech was used for both informational and warning messages, no significant difference in response time was found between warnings with or without the alerting cue. Accuracy of warning message detection in the multiple function environment was significantly lower for the no cue condition, but still relatively high at 95.9%.

Hakkinen and Williges concluded that if synthetic speech were used only for warning messages, the alerting cue was unnecessary. However, if synthetic speech were used in a multiple function environment, an alerting cue may be necessary to ensure a high level of responding. They went on to recommend further research to examine whether alerting cues would be needed if digitized natural speech alerts were used.

In this study, Hakkinen and Williges also reported that in the multiple function condition, primary task performance (successfully guiding aircraft to a landing on the carrier) was degraded along with a secondary task of message transcription accuracy. They hypothesized that the task performance, at least in the context of the air traffic controller simulation, could be affected by the amount of information presented in the auditory channel. Further research was recommended and subsequently carried out (Williges, et al, 1986) to explore ways to mitigate task performance degradation in multiple function speech displays.

4.1.2 Visual Redundancy (or Multimodal Presentation)

In the study described in article II, visually redundant presentation of speech messages were examined to determine if an opportunity to review the content of the message visually could mitigate an initial failure to fully comprehend the spoken message and improve accuracy of detection and transcription, particularly when the message may be interrupting an ongoing message. They compared speech, visual, and speech/visual redundant presentation in the same task environment as the first study and found that the speech and speech/visual redundant presentation modes significantly improved message transcription accuracy to 99%, compared to 95% for the visual mode alone. They concluded that speech displays can be effective in tasks with a high visual workload and that visual redundancy of the information presented in speech is beneficial in that it can provide a visual confirmation to improve accuracy of the information transfer process.

4.1.3 Conclusion

Articles I and II are useful in setting up the discussion of the research question examining to what extent current guidelines and standards for alarm design may be applicable to the growing diversity of alarm applications. Though the alerting cue recommendations are robust in the context in which they were studied, it is unclear if the guidelines hold fast for current day systems using state of the art speech synthesis or digitized recordings of human speech. Edworthy, et al (2003) have studied speech qualities and the use of warning signal words with natural and synthetic speech, and has suggested that perceived "urgency" of the spoken warning may be useful. Replication of the original studies with current day synthetic speech and natural speech would lend some confidence to the original findings or suggest that the current guidelines be revised. However, the general finding is that a unique voice, used for warning messages, can improve alarm detection accuracy and time. When a non-unique voice is used, alarms should be prefixed with alerting signals to improve detection accuracy. In all cases, warning accuracy can be improved with the addition of a simultaneous, redundant visual presentation of the message, what we term, multimodal presentation. Though accuracy is improved, responses are lengthened in this multimodal presentation.

The fact that most research related to voice warnings has been conducted using English language task environments and stimuli is troubling. It would be valuable to replicate such studies in multiple languages to explore cultural and linguistic differences and the potential impact on the design of speech warnings.

4.2 Structured Audio and Adaptive Presentation

In order to set the stage for article III by DeMeglio, Hakkinen and Kawamura (2002), and to understand how the standards-based multimodal foundation would nfluence the subsequent articles, we need to review the origins of the technology and standards that the work is based upon.

One of the significant developments resulting from the move toward digitized and synthetic speech was the recognition that these display formats permitted a richer form of both presentation and interaction than was possible with pre-recorded audio content. Traditional magnetic tape recordings afforded only linear navigation, measured by either a "distance" counter or by careful navigation using fast forward or rewind. Such a model of interaction would be unsuitable for navigation of rich multimedia, yet it should be noted that most first and second generation software media players only afforded this level of interface to digital audio.

In this section we are going to focus on two key and largely integrated efforts toward improving access to the World Wide Web and Talking Book.

4.2.1 Non-visual access to the Web

With the advent of the World Wide Web, challenges were faced by visually disabled users and their screen reading software due to the different ways in which the early Web browsing software rendered HTML content. One successful approach that bypassed the traditional screen reading model was proposed by Hakkinen (1996) and Hakkinen & DeWitt (1997) in the form of the nonvisual browser named pwWebSpeak. This approach viewed HTML as a structured document with embedded semantics that could be used to generate and provide navigation into a synthesized audio stream. The non-visual browser thus allowed visually impaired users to interact with structured Web-content in a purely auditory manner.

In advance of what would eventually become the aural properties of Cascading Style Sheets, the pwWebSpeak browser described by Hakkinen (1996) allowed auditory styling of HTML elements and content through changes in synthesizer speech qualities and the insertion of non-speech sounds to signal different elements, such as headings, list items, or hyperlinks. For example, hyperlinks could be read in a synthetic voice, preceded by a synthesized word label (such as "Link"), or prefixed by a digitized sound recording of a chime or a person announcing "link." The concept of signaling the presence of significant information within a stream of information using non-speech cues was influenced somewhat by the author's earlier research with alerting cues and multi-function speech synthesis systems described in articles I and II.

With the creation of the Web Accessibility Initiative by the W3C in 1997, efforts were undertaken to formalize recommendations for accessibility features in Web User Agents (e.g., Web browsers), the HTML language, and in authoring tools (W3C, 1999; W3C, 2000; W3C, 2002). A number of features incorporated in the W3C User Agent guidelines (W3C, 2002) were first demonstrated in pwWebSpeak.

4.2.2 Digital Talking Books

During 1996, a project was initiated to develop a replacement for traditional books on tape. The concept of synchronized multimedia, that of linking an audio narration with a structured document containing the text of the narration was initially implemented in the pwWebSpeak browser using a standard digital audio plug-in working in conjunction with time-based positioning highlighting of elements in the HTML document structure. This approach produced a prototype reading system that could allow the visually impaired reader to listen to a pre-recorded digital audio book or read an electronic text using synthetic speech and utilize the same consistent interface for reading control. In addition, dyslexic readers would be able to listen to the audio narration while reading a synchronized, visual presentation of the corresponding text.

Concurrent with this activity, the Swedish Talking Book Library had been developing a Digital Audio Book, called DAISY 1.0, which utilized the detection of intervals of silence in the audio narration to provide what was called phrase level navigation (Hansson, et al 1994). The resulting audio file format was non-standard, proprietary and though it did provide for a table of contents and section based navigation, it did not utilize or expose the underlying text or semantics of the document being narrated. During early 1997, and at the request of the major talking book libraries, these two efforts were combined and led to the development of an open, non-propriety format based upon structured text (HTML), a standard audio file (e.g., MP3), the Synchronized Multimedia Integration Language (SMIL), and a novel mechanism, the navigation control center, for providing navigation through the structured multimedia document (DAISY, 2005).

One of the usability advantages seen in the pwWebSpeak browser was the ability to bind keyboard navigation of the document to structural elements within the HTML content. Simple key strokes allowed for "jumping" between heading elements, lists and list items, tables, and any structural element meaningful within the context of the current document. This concept would also be meaningful to the digital talking book and provide a level of usability not possible in the traditional linear audio or phrase-based models. This approach, called Structured Audio Navigation, was implemented in the DAISY 2.0 standard (DAISY, 1998) and described by Hakkinen and Kerscher (1998). It should be noted that the concept of structured audio has been described by others, notably Hindus, et al (1993) and Arons (1991), but these earlier implementations relied generally on acoustic characteristics rather than structure based upon document semantics.

With the development of ANSI-NISO Z39.86, a refinement and extension of the navigation model was developed (Hakkinen and Kerscher, 2000). In the Z39.86 standard, a move was made to a new set of XML document formats, of which, the NCX was defined as the navigation control application. The NCX consisted of a series of hierarchical nodes that referenced navigation targets within a digital publication. For the purposes of maximum accessibility, each node in the navigation structure consisted of a textual label, an auditory label, and a graphic label, what can be called the *accessibility triplet*. It was hypothesized that in the full scope of potential applications, talking books may be created for people with linguistic or cognitive limitations that used iconic labeling of structure and content rather than textual. In addition, each accessibility triplet incorporated the standard XML language attribute, allowing multiple human languages to be supported within a single navigation structure.

The hierarchical table of contents navigation of the NCX, which generally mirrored the structure of the source publication, could be augmented with what were called Navigation Lists. These lists offered alternative paths through a publication that could be defined either by the author or the producer. It was further hypothesized that in academic contexts, educators may make use of the facility to collect individual readings from within a larger publication.

A key decision in the design of the digital talking book standards was the adoption of the W3C's Synchronized Multimedia Integration Language, or SMIL (W3C, 2005). In the typical talking book model, only limited subsets of SMIL capabilities were used. The modular design of SMIL 2.0 allowed for the creation of a lightweight,[27] defined subset, based approximately on what is called SMIL-Basic for use in talking books, and named the DTB Profile. Because SMIL was designed for rich, multimedia applications which could consist of

27 Implementing the full SMIL specification in a small handheld device was seen as burdensome and unnecessary, leading to the selection of the "lightweight" SMIL subset which would provide the necessary media objects, control and synchronization.

separate audio, video, image, and text tracks, with extensive control features for presentation, layout, timing, and synchronization, the door was left open for talking books to eventually make use of these more advanced features.

The use of digital talking book navigation features in other forms of multimedia was described by Hakkinen, et al (2002) via the proposed gNCX (or generalized Navigation Control language). It was envisioned that hierarchical navigation could be defined for self contained as well as composite documents, even when such navigation was not originally defined by the content authors. An example was presented in which a general purpose disaster preparedness guide could be created using the best available Web multimedia content from a variety of sources. An end-user could be presented with essentially a new "publication" consisting of component documents re-purposed from different sources (e.g., a Tsunami preparedness chapter from the Hawaii Office of Emergency Preparedness, an Earthquake preparedness chapter from the Tokyo Metropolitan Government, and a Hurricane preparedness chapter from the State of Florida).

4.2.3 Adaptive Presentation of Information

And with the stage now set, we arrive at our second review paper. By 2001, there were only a handful of commercially available DAISY playback systems (both hardware and software), and none made full use of the underlying capabilities of the content. Further, there was little variation in the user interfaces, which in itself was not inherently a poor situation from the user perspective. These interfaces were specifically tailored to the needs of the visually impaired user community and based upon requirements developed as part of the standards development process (e.g., EBU, 1996, LOC, 1998). It should be noted, though, that there was little innovation in the interfaces beyond the original prototype players of 1998.

A project was initiated in 2001 to develop a prototype DAISY playback system that would allow exploration and prototyping of different user interface styles for Digital Talking Book content. The resulting software prototype, the Adaptive Multimedia Information System, or AMIS is described in article III (DeMeglio, Hakkinen, and Kawamura, 2002). AMIS incorporated several facilities that would offer the opportunity to explore alternative interfaces to the talking book. The first of these features was the concept of "skins", which had been popularized in several PC-based media players (e.g., WinAmp)[28] and Web browsers, including Google Chrome[29] and Mozilla Firefox[30]. In AMIS, skins could be used to explore different user interface designs, and exposing (or hiding) control mechanisms. The second of these features was the concept of input and output plugins. The input plugins would allow for different hardware and software input and control mechanisms including touch screens, game control-

28 http://www.winamp.com
29 http://www.google.com/chrome
30 http://www.mozilla.org/firefox

lers, specialized buttons, sip/puff switches, scanning menus, and speech recognition. On the output side, plugins could be used to direct output to refreshable Braille displays or displays showing magnified (enlarged) text. A variety of prototype plugins were created and informally evaluated in Japan, Europe, and the United States by users with visual, physical/mobility, learning, and cognitive impairments. The software was subsequently localized to a number of languages including English, Japanese, Hindi, and Thai.

The open source development approach of AMIS, in conjunction with SMIL-based tools, was seen as offering a core platform for research in speech-based information displays and to investigate synchronized presentation of information in multiple forms within and across modalities, dynamic adaptation of content presentation, and selectability of speech type (digitized, natural speech or synthetic speech). AMIS was subsequently used as a research platform for developing accessible disaster preparedness materials in Japan.

4.2.4 Conclusion

In forming the second research question, we asked whether design approaches from the information accessibility context are applicable to creating effective alarm presentation models. The approach taken in developing both accessible information standards and the software systems to allow accessible interaction with that content is presented as an effective model for development of adaptable and accessible warnings described in the following articles. Specific techniques used in the talking book standards, such as the accessibility triplet, hold promise for inclusion in the emerging standards for public warning information, such as CAP. The failures in warning systems, in which alarms are not responded to may be mitigated through adaptive and accessible information interfaces, and in particular, through the exploitation of the declarative and programmatic use of SMIL content presentation and interaction capabilities. The utilization of XML languages as means for encoding alarm structure and information, and specifically, applying the accessibility triplet model of the NCX to ensure every information unit of a warning has multiple modalities available, can offer a distinct advantage in alarm design. Combining this approach with the timing and synchronization capabilities of SMIL, alarms may be created that are formally defined with a structural, semantic, and timing model that facilitates alarms that will be identified and responded to, through repetition and requirements for user confirmation of receipt.

4.3 Communicating Disaster Preparedness and Warning Information to People with Disabilities

In article IV, Sullivan & Hakkinen (2006) describe how information accessibility concepts may be applied to disaster preparedness. Specifically, an examination of the challenges faced by people with disabilities during disasters was undertaken. Looking at the 9/11 World Trade Center attack, the Indian Ocean Tsunami, and Hurricane Katrina, the need for effective communication of preparedness information and warnings for all members of a community, including those with disabilities is essential.

4.3.1 People with Disabilities during Disasters

People with disabilities are often the most vulnerable during a disaster or emergency (SNAKE, 2005). Drawing upon the examples from recent disasters the authors describe how the specific information needs of those with disabilities have not been incorporated into preparedness and warning systems, or emergency response plans. In many cases, people with disabilities have been completely excluded from emergency response planning, and only high profile events, such as 9/11, have served to highlight the problem, through dramatic and compelling examples.

The authors provide examples of complex information, such as evacuation maps, that pose significant challenges for understanding by individuals in a population who may have cognitive or sensory disabilities, or do not speak the local language. Alarm signals, such as sirens or fire alarms, designed solely for those who can hear, further typify the kinds of communication challenges faced by the deaf.

Information accessibility standards have been developed for electronic books and Web-based information, and the authors present the case that these standards should be applied to the creation of disaster preparedness materials and warning systems. Using the model of DAISY talking books, the authors describe a prototype, accessible multimedia preparedness guide developed by and for a community of individuals with cognitive impairments in Hokkaido, Japan. The community lives a coastal town prone to both earthquakes and tsunamis, and rapid evacuation to safety zones is vital to ensure survival when a tsunami warning is issued. The guide, based upon the multimodal presentation model of DAISY, incorporates text, image, and audio description of earthquake safety tips and evacuation routes. The PC-based delivery model of the guide is problematic for some users, both from user interface complexity perspective and from cost of the PCs themselves. Looking beyond PCs, the authors propose the use of low cost consumer multimedia devices to deliver accessible preparedness information, and present a prototype, interactive evacuation guide running on hand-held gaming device.

4.3.2 Accessible Information Standards for Public Warning

In article V, Hakkinen and Sullivan (2007), examine the application of accessible information standards to the specific problem of public warning. Describing the development of accessible Web content and Digital Talking Books, the authors discuss how information can be rendered in combinations of synthetic or natural speech, in Braille, and visually, with styling of the presentation to include speech and non-speech cues to indicate significance of information.

The adaptive presentation of critical information is presented that would allow for dynamic application of message presentation characteristics, levels of details, cueing, language and redundancy to ensure accurate detection, understanding and timely response. To support diverse population varying in culture and language, the importance of incorporating the principles of software internationalization and localization in information standards are discussed. Finally, user profiles that can facilitate adaptation of critical information to specific user needs are highlighted as critical to effective message delivery.

4.3.3 Conclusion

In reinforcement of the conclusion for article III, supporting the point raised by research question two that design approaches from the information accessibility context are applicable to creating effective alarm presentation models, the authors propose that accessible multimedia standards be used to convey emergency preparedness information and present examples of a prototype accessible, multimedia for disaster preparedness. A key point in both article IV and V is that an accessible approach, using multimedia, can also benefit those who are not disabled, or those who are situationally disabled by illness, age, cultural or language differences.

4.4 Use of Mobile Phones for Emergency Notification

The use of mobile phones as a communication device has grown dramatically, to the point where approximately one half of the world population now has a mobile phone.[31] The mobile phone is a personal communication device, serving as a platform for social networking and interpersonal communication by voice and SMS. It also enables customization, whether by selection of a particular model or style of handset, to onscreen themes, and custom, user selected ring tones and message indicators. Mobile phones also serve as a lifeline, during times of ill health, crisis, or disaster. Beginning with 9/11 the value of mobile phone during disasters became quickly apparent.

4.4.1 Emergency Notification on University Campuses

In the United States, a variety of incidents involving school shootings led to the widespread adoption of emergency notification systems on university campuses. In article VI, Sullivan, Hakkinen and Piechocinski (2009) and article VII, Sullivan, Hakkinen and DeBlois (2010), the authors examined the problem of creating effective emergency notification systems on a university campus and specifically looked at the challenges faced by those individuals with sensory and cognitive disabilities. The primary challenge observed by the authors is that emer-

[31] ITU. http://www.itu.int/ITU-D/ict/statistics/ict/index.html

gency notification systems are designed in a "one-size fits all" approach, meaning the underlying assumption is that notification recipients can see, hear, and understand the language and meaning of a notification. In both papers, the authors surveyed a representative sample of the university population to determine technology adoption and utilization, as well as to explore questions such as trust of notification sources. As most present notification systems in use on university campuses are commercial products designed for general purpose notification of communities or businesses, it likely significant numbers of the campus population may have information presentation needs outside the product designer's expectations, including students with learning disabilities and international students whose first language is not the local language. In surveys of university communities in the US, between 10 and 20 percent of the campus population will have an identified disability (ranging from 10 to 20 percent of the campus population). In addition, many campuses have significant numbers of international students, whose native language may differ from that of the local university population. For example, at the university in New Jersey studied by the authors, between 4 and 5 percent of the student body is comprised of international students.

Many universities and colleges in the United States receive federal funding and as such, must comply with federal regulations pertaining to accessibility and disability rights. Further most vendors of emergency notification systems market their products to schools, businesses, local and federal governments. All products and services offered to the federal government are required to comply with the federal Section 508[32] ICT accessibility guidelines, yet during the development of Article VI and VII, the author surveyed vendor Web sites, and did not locate references to either Section 508 or to any accessibility features of the notification products.

4.4.2 Technology Adoption

In both articles, survey findings indicate a high (99%) adoption rate of mobile phones by the student population of the university in the study. Students also reported use of personalization and customization on their mobile phones and altering such features as ringtones and silent modes to accommodate to avoid class room interruptions. When it came to adoption of the campus emergency notification service, results were less unanimous, with approximately 62% of students utilizing the service in 2008, and rising to 81% in 2009. In an unexpected finding, 100% of the students surveyed who reported having a disability had signed up for the notification service, while only 71% of the international students.

[32] http://www.section508.gov
[33] Vendor names were kept confidential and not published.

4.4.3 Multi-Modal Alerts

In examining the diverse information needs of the campus population, the authors defined a mapping of a standard text alert into the various modalities required by a diverse population of users with disabilities (see FIGURE 7. Mapping of SMS/Text alerts to support sensory and cognitive disabilities.). In this mapping, the authors proposed a multimodal presentation model in which the alert would be transformed into a variety of presentation formats, including graphic symbols, animations, audio, tactile (vibration and/or Braille). The problem, however, is the present focus of emergency notification systems utilizing SMS or e-mail do not provide sufficient information to enable the adaption of the message into multiple modalities. The Common Alerting Protocol (CAP) only recently audio as an optional presentation component of the CAP message.

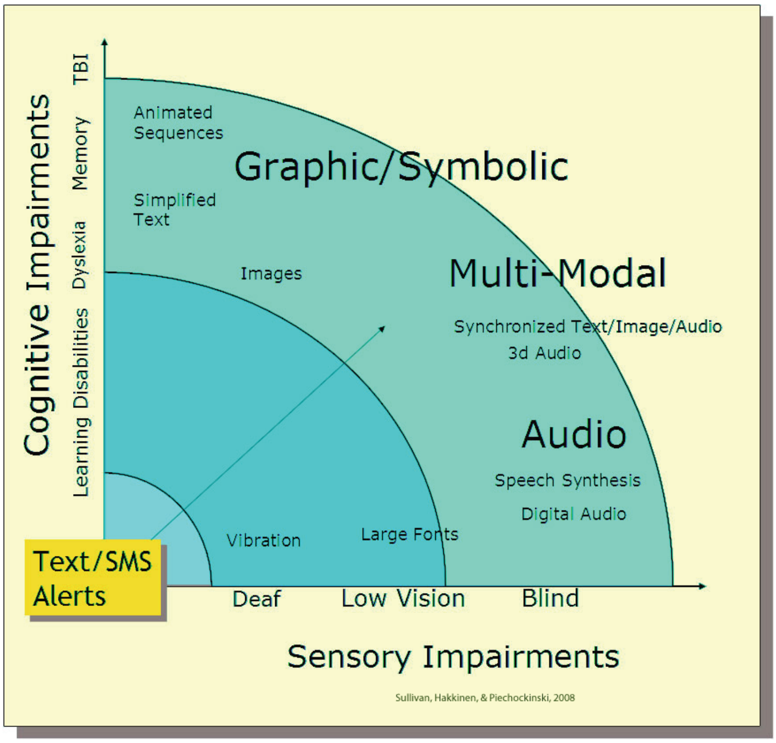

FIGURE 7. Mapping of SMS/Text alerts to support sensory and cognitive disabilities.

The authors describe how multimodal alerting can be supported via a mobile phone application designed to receive enhanced CAP messages and then transform the message into appropriate modalities based upon a user profile.

4.4.4 Conclusion

One of the key points in these paper is that the current standards for emergency notifications, such as the common alerting protocol (or CAP), do not provide sufficient structure or meta data for supporting the multi-modal transformation. Further, there is insufficient empirical data to fully define the most effective message formats within a modality for specific types of disabilities (e.g., cognitive disabilities), though we can infer promising approaches from the well established accessibility standards such as the W3C Web Accessibility Initiative's Web Content Accessibility Guidelines 2.0 and the User Agent Accessibility Guidelines 1.0. From this foundation and understanding the requirements and user preferences for emergency notification on mobile phones, an answer to the final research question, whether design approaches from the information accessibility context are applicable to creating effective alarm presentation models can be addressed.

4.5 Author's Contribution to Collaborative Research

The research presented in articles I and II was independently designed, developed and conducted by the author under the supervision of Beverly Williges at Virginia Tech. The initial study, reported in article I, was designed to replicate the findings of Simpson and Williams (1980) and then to examine the impact of using speech for more than just alerting functions. The following study, reported in article II, built on the earlier findings and examined the use of both visual and speech modalities for message presentation. In support of both studies, the author developed task specific software extensions to the GENIE experimental platform (Lindquist, Fainter & Hakkinen 1985) to control the speech-based and visual interfaces. Article II was a chapter in a book on Human Computer Interaction, and the research described covered findings on both speech output and speech recognition. The author was a member of the research group which conducted studies on speech interfaces, but was not directly involved in the speech recognition studies reported in the chapter.

The work presented in article III describes the Adaptive Multimedia Information System, based on design concepts from the author's non-visual browser research (Hakkinen & DeWitt 1997) and his work on the ANSI NISO Z39.86 Digital Talking Book Standard, in which he was a key technical contributor. The system described was implemented by Marisa DeMeglio under the technical direction of the author, who also developed portions of the user interface. The third author, Hiroshi Kawamura, was the project sponsor.

The work presented in article IV was a collaborative effort between the two authors, with the research and writing responsibilities were shared. Field research in Thailand and Japan, and development of software prototypes, was carried out by the author.

For article V, the lead author on this paper had primary responsibility for the work, including research, writing, and presentation, with the emphasis on accessibility standards and human factors aspects of warning design.

The research described in article VI and VII comprised survey research and analysis of campus and community-based notification systems. The author shared overall research, survey design and writing responsibilities with Helen Sullivan, while the third author, an undergraduate student, participated in survey design, data collection, and analysis.

5 CONCLUSION: AN ADAPTIVE, MULTIMODAL ALARM MODEL

Alarm design, to date, has generally taken an approach that implies a one size fits all recipients model. Implementation of alarms has traditionally been based upon human factors design guidelines of varying age and validity in today's alarm rich environment. Research has generally looked at the alerting characteristics of warnings for specialist populations, and not at alarms for the general population. Research has also produced a variety of alarm or warning models, generally based upon a visual foundation. All such models follow a relatively similar structure in which alarm failure at the early, initial perceptual stages and multimodal presentation modes are not actively considered. We should approach the model of alarm response at the outset with the certainty that some channels may be filtered, internally via sensory or cognitive impairments (either chronic, acute, or situational) or externally via environmental factors (such as noise or smoke). Though some recent aircraft alarms have exhibited adaptation of the warning signal (increasing intensity until a response is made), adaptation is rarely seen in other contexts. In the simplest case, we see alarm clocks and kitchen timers with alarm design following a convention that has gone largely unchanged for decades. Rather than seeing a progressive model of human centered design, humans have had to adapt to the alarm with mixed results. We have adapted to their use, and from this adaptation, we often learn to not respond.

Personalization has been a growing trend in mobile devices. Today mobile phones can be customized with personalized ring tones and caller identification images. Phones may be switched to profiles appropriate to their usage context, in a meeting, outdoors, etc. Specific callers may have unique ring tones or images to signify a family member or friend is calling. These convenience features that personalize the phone can also support the needs of people with cognitive impairments by providing visually salient cues rather than abstract signals or textual labels. Those who are deaf utilize mobiles, not for voice communication, but for text based communication, and the deaf necessarily utilize the vibration mode to signal incoming messages. Vibration mode is obviously of use to those who cannot hear, or by those in noisy environments or restricted by context (e.g., in a meeting or library). Some mobile phones incorporate text to speech synthesis to read textual information to the phone's user, such as contacts lists and the content of text messages. Text to speech benefits those who may be blind or visually impaired, as well as those in traditional eyes-busy environments (e.g., driving an automobile). All of these personalization features can function to improve the user's ability to detect and respond to incoming communications, while also meeting the requirements of information accessibility for those with disabilities.

Why is information accessibility a good model for general information and interaction design, and in particular, for warning and critical information? Equivalent, alternative presentations of information is a fundamental basis of accessibility, as encoded in national and international accessibility standards such as the US Section 508 law and the W3C Web Content Accessibility Guidelines. Any information delivery system that embodies accessibility will permit multiple modalities to be supported, as well as personalization. Multimodal presentation is not, however, solely beneficial to individuals with disabilities.

Providing parallel renderings of information may have value in adapting critical information presentation to a given operational and user context to achieve a higher likelihood of detection and response. Selcon, et al (1995) has shown the potential value of redundancy of information presentation forms in the context of visual warnings and Liu (2001) has explored multimodal displays in automobiles that can be less distracting to the driver. In situations where attentional demands lessen the chance for detection, multiple modalities can mitigate some of the deficit, and we have seen the value of multimodal presentation in alarm research. The challenge will be to develop and refine a robust model that takes into account user context and situational, attentional and information processing impairments that may raise barriers to critical information receipt, understanding and response. Such a model could help define rules for adaptive presentation of critical information, one that can dynamically apply appropriate message characteristics, levels of details, cueing, language and redundancy to ensure accurate detection, understanding and timely response.

Forcing an individual to adapt to poor alarm design sets the stage for alarm failure, knowing that we can't alter the fundamental processes in sensation, attention, and cognition. Creating alarm systems that are adaptive, and personalized to the context of the individual can be the best solution. We can't recreate, for humans, the animal model of biological salience through hard-wired, natural alarm signals. However, we can allow individuals to select and customize alarm and warning signals on their mobile devices, to provide the most salience for signals of critical importance. This signal salience approach can carry over to personal systems within homes and vehicles, but not without the creation of profiling tools that would allow personalization to move with the user, thereby allowing family members or visitors to utilize their own personal alarm profile.[34] A key aspect of personal profiles is that privacy should be maintained

Our surveys of mobile phone use clearly indicate that personalization is a viable and active, element of how people customize and adapt their technology to personal needs and requirements. An excellent example of the potential personalized alarms is the work of Smith, et al (2006) in which smoke alarms that incorporated the voice of a parent as the alerting signal were more effective at waking sleeping children than the standard buzzer alarm. Any alarm system that will support customization of the cue and message shows promise for improving user responding.

5.1 Defining a Multi-Modal Alarm Specification Language

To achieve personalization, we must also focus on standards to ensure that personalization models are consistently supported across platforms and devices. To that end, we can take the lessons learned from the research summarized in the prior chapters, and the practical developments of Web and information accessibility standards, to propose a language for alarm specification, what I call the Multi-Modal Alarm Specification Language (MMASL). In basic form, MMASL is an XML application for specifying the critical information content and presentation style for any form of alert or alarm. Borrowing from the previously described work on DAISY and SMIL, MMASL supports creation of alerts which consist of multimodal cues and message content. The alarm is broken down into a sequence of elements, the cue and message. Each element incorporates the "accessibility triplet" model developed in the DAISY NCX. In this approach, information content in the alarm can be specified in one or more modalities, including textual, audio, tactile or graphic. In cases where a modality is not specified, such as when only a textual message is provided, the mobile device may transform textual information into an audio form using text to speech synthesis.

As with any well-designed XML application MMASL natively supports encoding of information in multiple languages, facilitating internationalization and localization of alarms. The use of language attributes allows specific textual content of an alarm to be tagged with the national language of the message content. As mobile phones generally allow user selection of an operating language, the user of a phone enabled with MMASL support would only receive the portion of an alarm in their selected language, should the alarm incorporate multiple languages (such as might be sent by public authorities in culturally/linguistically diverse locations).

34 The National (now Global) Platform for Inclusive Interfaces has proposed a cloud-based system for personal user interface profiles. http://npii.org/programs/private-preference-permission-system

To increase the likelihood that users will receive and read an alarm on their mobile, MMASL incorporates concepts from SMIL which allows alarm elements to be repeated, for example, until either a repetition count has been reached or a user has confirmed that the element has been read (or heard). A user can indicate confirmation by activating, for example, a "Continue" or "More Info" button. It should be noted that with advanced sensor capabilities present in today's mobile devices (accelerometers, proximity sensors, GPS, speech recognition, cameras, etc), it may be possible to use multiple active and passive inputs to detect, for example, that the user has oriented toward the device's display screen as a mechanism to terminate the alerting cue and begin message presentation.

5.1.1 The MMASL User Agent

The MMASL client application should be viewed as a user agent[35] for alarm and notification messages. As such, the MMASL client should follow the guidelines defined by the W3C User Agent Accessibility Guidelines (W3C UAAG, 2010).[36] In user agents implementing a Document Object Model (or DOM), scripting may be used to control interactive behavior of a loaded XML application. For example, based upon user preferences, a repeating alarm sequence in MMASL may utilize CSS aural properties[37] to control volume. With each repetition, the volume of the audio presentation may be increased programmatically to favor alarm detection by the end user.

A common feature of many user agents is the ability of the user to define personal settings for common features. An MMASL client would provide options for user specification of default alarm behaviors and characteristics. As is the case in the Web accessibility context,

An argument in favor of utilizing XML applications like MMASL in mobile devices is the widespread adoption MMS (the Multimedia Message Service), which includes a subset of SMIL, and the adoption of the ePub format in eBook devices. ePub is based on DAISY text structuring and incorporates the NCX for table of content presentation and navigation. Though ePub itself is not suitable for alerting messages, as the intent is quite different, the fact that multiple handheld devices support the XML format, and the NCX, suggests that the technical limitations will not hinder XML-based alarm user agent implementation.

5.1.2 The Structure of MMASL

To better understand MMASL we can begin by looking at the basic time line and elements of an alarm, shown in FIGURE 8. MMASL Alarm Structure

---

35 Definition of User Agent: A user agent is any software that retrieves, renders and facilitates end user interaction with Web content or protocols.

36 The author is an invited expert to W3C User Agent Accessibility working group and is a contributor to the creation of UAAG 2.0, the latest version of the guidelines.

37 http://www.w3.org/TR/CSS21/aural.html

Alarms consist of a message conveying relevant information, which may be preceded by an alerting signal, or cue. In our examination of alarm research, we have found recommendations that an alerting cue be used to signal the onset of an alarm. We have also found support that when the auditory characteristics of the message presentation are unique, they can serve the alerting function, by focusing attention on the message, and no cue is required. Any alarm model must support cued and un-cued messages, based on the characteristics of the message, and this capability is a basis of MMASL.

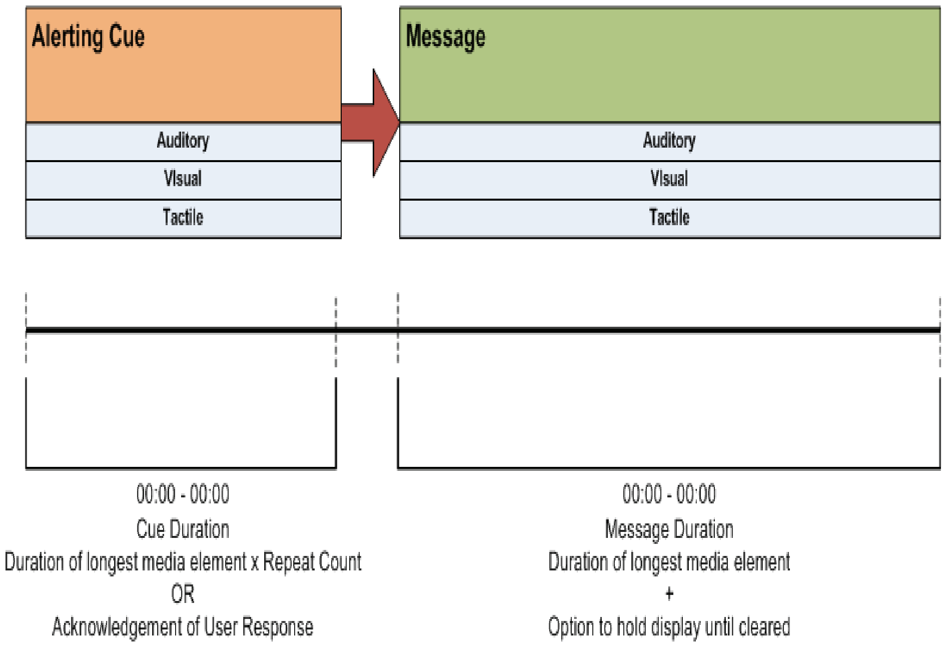

FIGURE 8. MMASL Alarm Structure

An MMASL XML alarm consists of a top-level container element <mmasl>, which itself contains one or more <alarm> elements.

<alarm> elements would specify a language attribute to indicate the language of the alarm. There may be, for example, multiple alarm elements within an <mmasl> corresponding to the different languages required for an alarm's audience. The language attribute is also supported within individual elements of <mmasl> to allow individual messages to contain multilingual components.

Alarm cues and messages may not be immediately detected, or the user's attention may only shift focus once portions of an alarm have already been presented. As potential solution to the problem of ephemeral alarms, MMASL incorporates the capability repeat both alarm and message components, and to control sequenced presentation based upon user response. To support this, the contents of the alarm element can take on the behavior of SMIL media objects and containers. Utilizing two timing related attributes from SMIL, *repeatCount* and *end*, both cue and message can be repeated a finite number of iterations or indefinitely. In both cases, repetition can be terminated by a user response.

The <alarm> element can be considered to be the equivalent of the SMIL <seq>, or media sequence element, in which a series of media elements are played in sequence, from first to last (e.g., when the implicit or specified duration of the first media element is reached, the next media element in the sequence is played). The two top level media elements in the alarm are the cue and the message. The intent behind the incorporation of these two elements is based on the premise that an alerting cue can facilitate detection and recognition of the alarm in some contexts (e.g Hakkinen & Williges, 1984) and has been recommended as a design guideline for alarms containing speech. Once the cue has been presented, it is followed by the message. Both the cue and the message elements are structured as accessibility triplets, containing auditory, visual, and tactile versions of the cue and message.

The overall structure of the alarm language is shown in the graphical representation of the XML schema for MMASL in FIGURE 9. The structure of the MMASL Schema (for the full schema definition, see appendix A). As fully specified, MMASL allows for creation of complex, multimodal alarms incorporating both time-based and interactive control information presentation. Based on the context and purpose of the alarm, it may be presented with or without a cue, with one or more parallel and sequential channels (modalities) of information, and allows specification of an external event, such as a user confirmation action (e.g., button click) or system state changes (e.g., end of condition that triggered alarm) to control alarm presentation.

The MMASL external events can be particularly useful in allowing repetition of an alerting cue to capture attention, followed by a user action (button press) that terminates the cue and presents the message text. Alternately, an alarm signal may be triggered due to a dangerous system state, for example in industrial process control, and repeated until the alarm recipient (system operator) brings the system back to safe operation.

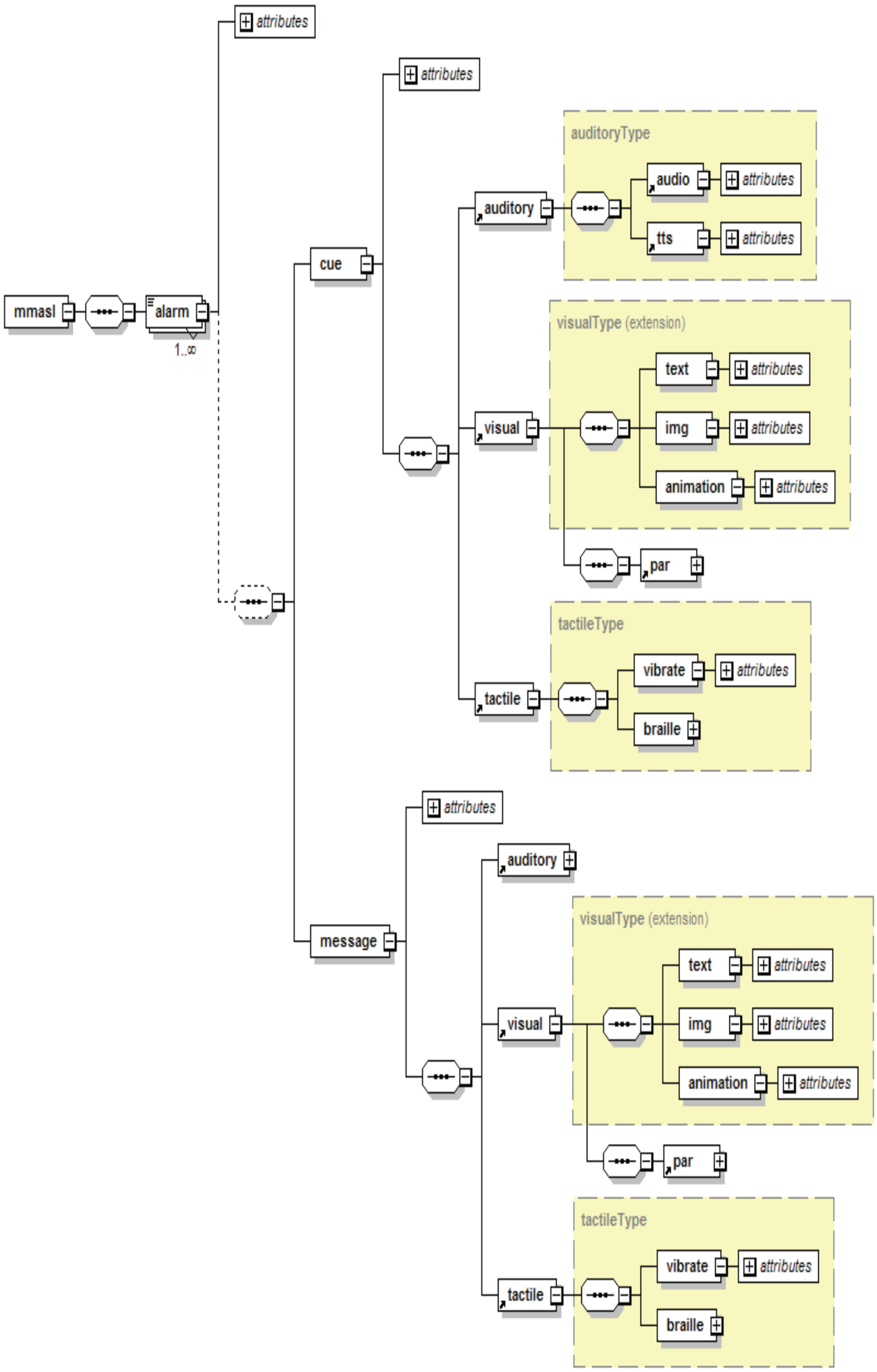

FIGURE 9. The structure of the MMASL Schema

The following represents a simple MMASL alert:

```
<mmasl>
<alarm lang="en-us">
  <cue repeatCount="indefinite" end="btnOK.click">
   <auditory>
  <audio src="alerttone.mp3"/>
  </auditory>
  <visual>
    <par>
      <text>Fire Alarm!</text>
      <img src="alertgraphic.png" alt="Fire Alarm"/>
   </par>
  </visual>
  <tactile>
     <vibrate src="alertpattern.mid"/>
  </tactile>
 </cue>
 <message end="btnOK.click">
   <auditory>
     <audio src="chembldgfireevac.mp3"/>
   </auditory>
  <visual>
      <text id="T1">
        A fire is reported in the chemistry building. Evacuate the building now.
      </text>
  </visual>
  <tactile/>
 </message>
</alarm>
</mmasl>
```

In the example, the cue combines auditory, visual and tactile components that are repeated until the user confirms receipt of the alarm by touching the device screen. Once the cue is terminated, the defined message is presented, in this case including both a visual, text message and a pre-recorded audio message. To achieve the repetition and wait for user response, the <cue> element uses two attributes, *repeatCount* and *end*.

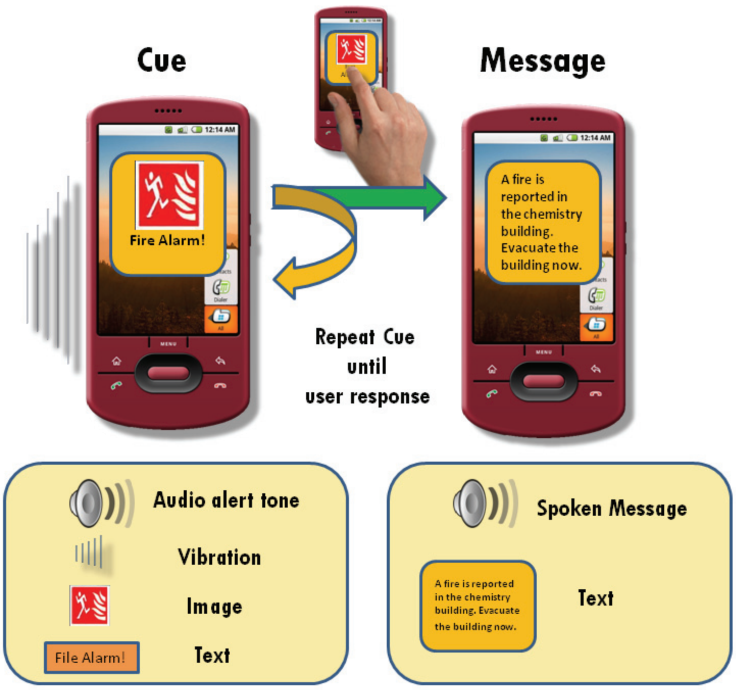

FIGURE 10. Simple Alarm Example using MMASL

5.1.3 MMASL Examples

The ability of the MMASL to support the requirements of various alerts can be shown in the following examples.

The Hakkinen and Williges (1984) study utilized a combination of hardware and software to generate an alarm consisting of an 500 ms alerting cue, 500 ms of silence, followed by synthesized warning message. In MMASL, this would be represented as:

```
<mmasl>
<alarm lang="en-us">
  <cue dur="1s">
    <auditory>
   <audio src="500Hz500ms.mp3"/>
   </auditory>
  </cue>
  <message>
    <auditory>
      <tts>Left engine failure</tts>
   </auditory>
  </message>
</alarm>
</mmasl>
```

In the example above, the duration of the cue is 1 second, during which a 500 ms alerting tone would be played, followed by 500 ms of silence, presented via a single MP3 audio file. This cue would then be followed by a synthesized audio message “Left engine failure.” Modifying the MMASL above to support the variations studied in Williges, Schurick, Spine, & Hakkinen (1985), for example, to evaluate visual and auditory messages without cue, would result in:

```
<mmasl>
<alarm lang="en-us">
 <message>
   <auditory>
     <tts>Left engine failure</tts>
  </auditory>
  <visual>
     <text>Left engine failure</text>
 </message>
</alarm>
</mmasl>
```

The addition of the visual element provides a redundant modality of presentation, while the cue element has been removed.

In the next example, we have a tsunami alert presented in four languages. Such an alert may be broadcast by a digital audio siren located on a popular tourist beach in Thailand, or on mobile phones able to process MMASL.

```
<mmasl>
<alarm repeatCount=”3”>
  <cue repeatCount="2" >
   <auditory>
  <audio src="tsunami-screech.mp3"/>
  </auditory>
 </cue>
 <message >
   <auditory>
     <tts lang=”th”>สึนามิใกล้, อพยพไปยังพื้นดินสูง </tts>
     <tts lang=”en”>Tsunami approaching, evacuate to high ground</tts>
     <tts lang=”jp”>津波近づいて、高台に避難</tts>
     <tts lang=”fr”> Tsunami, évacuer à la masse élevée </tts>
  </auditory>
  <visual>
    <image=”tsunamiEvac.png”/>
  </visual>
  </message>
</alarm>
</mmasl>
```

In the example above, the alert is repeated three times, with an initial alerting cue presented twice. Following the cue presentation, the evacuation message is presented via text to speech synthesis. If the MMASL alert is being presented by an outdoor digital siren, each of the TTS messages would be played in sequence, first in Thai, then English, Japanese, and French. In the case of a mobile phone, with a local language selected by phone user, only the message matching the user’s local language will be presented. In cases where the user’s preferred local language is not one of the choices, a fallback is to present a graphic, language neutral, tsunami evacuation symbol.

Another variant of the multilingual alarm provides a possible solution to the problem exhibited in the following transcript of the last words from a cockpit voice recorder recovered from a China Northern Airlines crash on November 13, 1993 in Urmqui (Aeronautics Committee, 1998; Baron, 2005):

> Ground Proximity Warning System: *Pull Up! Pull Up!*
> Pilot: *(in Chinese) "What does 'Pull Up' mean?"*[38]

Historically, and in practice, the international language of government regulated commercial aviation has been English. With the rapid expansion of air travel in Asia and other parts of the world, experience will take time to build among flight crews, and English language skills will likely not be uniformly high (Tajima, 2004). Understanding the nature of modern, developing civil aviation, it is important to recognize the variability in language skills among flight crews; stacking the deck against those crews by presenting time critical alerts in English only may result in more incidents such as the one indicated in our transcript above. A pragmatic and realistic recognition that the cockpit warning systems should support and adapt to the abilities of flight crew is long overdue, though adopting such an approach is challenging in view of the certification and regulatory processes for aviation systems. Assuming an aircrew alerting system based upon MMASL had been available on the above referenced Chinese airliner, the following alarm would perhaps have produced a different outcome:

```
<mmasl>
<alarm repeatCount="indefinite" end="alarmCancel">
  <cue>
   <auditory>
    <audio src="whoopwhoop.mp3"/>
   </auditory>
   </cue>
  <message >
   <auditory>
      <audio src="pull-up-en.mp3" lang="en"/>
      <audio src="pull-up-cn.mp3" lang="zh-cn"/>
   </auditory>
   <visual>
       <text lang="en">Pull Up!</text>
       <text lang="zh-cn"> 往上拉</text>
   </visual>
  </message>
</alarm>
</mmasl>
```

38 John Sampson, Editor, *Air Safety Week*, provided the author with additional background information and sources on this incident.

In the MMASL example above, the “pull up” alarm, presented in audio form as the non-speech alerting cue “Whoop-Whoop” followed by the spoken command “Pull Up!”(meaning climb immediately to gain altitude) is repeated until the aircraft returns to safe flight conditions. MMASL has been used to augment the message with both English and Chinese versions of the “Pull Up” command, which would play in sequence:

“Whoop-Whoop… Pull Up! 往上拉!”

The example has also provided a visual, text version of the message in both languages, which could optionally be presented on the pilot’s heads up display (HUD). The benefit of this approach to a non-native speaker of English, who may be fatigued after a lengthy day of multiple flight segments and operating potentially in difficult weather conditions, should be obvious. The addition of familiar language can aid users in recognizing the message and performing the correct action, whether in a cockpit, or responding to an emergency notification received while travelling in a foreign country.

MMASL can allow for user preference, or a user profile, to determine how different types of message content should be presented (e.g. user prefers visual symbols over speech, when available, or wants all messages read with female speech synthesizer). Using the SMIL model of content selection, user preferences can offer the possibility of allowing the user to select from a variety of presentation options in order to provide a meaningful, readily recognizable signal to identify critical and non-critical alarms.

5.2 Contributions

In beginning this thesis, I set out three research questions: First, to what extent are current guidelines and standards for alarm design applicable to the growing diversity of alarm applications? Second, are design approaches from the information accessibility context applicable to creating effective alarm presentation models? And third, can a personalization and multimodality framework for alarm presentation increase the effectiveness of alarms?

In answering the first question, we can begin by recognizing that warnings do fail to evoke the expected response in many cases. While significant research effort has been undertaken to develop empirically-based design guidelines for specialized populations, the applicability of these guidelines to more general alerting contexts is not without problems. What may work well in the lab, under controlled circumstances, may fail once applied in real world settings. It should be noted that even in the target specialist populations, such as aviation, alarm failures continue to occur. In our examination of design guidelines, we might conclude that context is everything. From the early research on speech-based warnings described in articles I and II we have seen how an established human factors guideline stating that alerting cues should precede voice warnings is questioned when synthetic speech is used as the source of the warning. The advantages of synthetic speech results from the unique voice quality of the speech, in effect serving as its own alerting function (e.g. if you hear a synthetic voice in the cockpit, it must be a warning) and the effect was readily demonstrated in both cockpit and complex human computer interactions scenarios. However, if synthetic speech is used for more than just alerts and warnings, which could be expected in complex systems, the alerting cue is once again shown to be useful to distinguish the critical messages from the non-critical. If a designer should decide to utilize synthetic speech only for warnings, the continual evolution of synthesis technology may make past guidelines moot. Improved synthesizer technology has resulted in voice characteristics that are becoming difficult to discern from natural, human speech, leading once again to the likelihood that the alerting cue should be reinstated if voice warnings (either natural sounding synthesis or digitally recorded human speech) are used in environments where other speech activity is presented.

If we take these findings as a lesson for other warning related applications, it is clear that a reliance on guidelines in isolation from the underlying empirical data and context used to generate them, can lead to systems which may not provide optimal solutions to the problem of effective warning performance.

To answer the second question, we have examined how information can be effectively conveyed, even in the most challenging circumstances, such as in the development of information accessibility for people with disabilities. Information accessibility guidelines, such as ANSI-NISO Z39.86 and the work of the W3C are being incorporated in national and international information and communications technology standards and legislation (e.g. US Section 508). Article III described a specific, standards-based solution for information accessibility that has been used subsequently to create accessible disaster preparedness materials. However, we must also recognize, as discussed in articles IV and V, that much of the public disaster and emergency notification community has failed to effectively address the information needs of vulnerable groups, be they people with disabilities or those who may be linguistically or culturally isolated from the local population. Though progress has been made in some areas of information accessibility, significant work remains to achieve emergency notification and warning systems that reach all at risk from disaster.

And to answer the third question, we have recognized the rapid emergence of the mobile communications device, enabling a personalized channel for receiving many forms of information, and have examined the use of mobile phones in campus emergency notification systems, as discussed in articles VI and VII. The article also indicates that such applications are not without problems, but hold promise for creating effective and widely available platforms for delivering critical information. Using our existing knowledge of the warning guideline research, the personalization capabilities of the mobile phone, work in information standards, the foundational knowledge human capacity in sensation, attention and cognition, combined with a model of how alarm processing is affected by internal and external factors in humans, a new, general model for alarm presentation has been proposed, providing a flexible, multimodal framework for research and development in a variety of contexts. The alarm presentation model takes a standards-based approach to defining alarms that incorporates a content and presentation model which enables creation of effective, personalized alarms. MMASL thus provides a framework for exploring the possibilities of alarm personalization and improving accessibility of alarms for those with sensory, cognitive, or situational disabilities.

It is hoped that this thesis will inspire further research to improve alarms and warning systems that place an emphasis on a human-centered approach that recognizes the variety of abilities and challenges present in the population of potential alarm recipients.

YHTEENVETO (FINNISH SUMMARY)

Hälytykset ovat läsnä kaikkialla nykymaailmassa. Kaikki ihmiset ovat jatkuvasti tekemisissä erilaisten henkilökohtaisten hälytysjärjestelmien kuten herätyskellot tai akun tilaa ilmaisevien merkkivalojen kanssa. Erityisen kriittisiä ovat monimutkaisten järjestelmien tiloista kertovat hälytykset kuten lentokoneen GPWS ja katastrofeja koskevat kuulutukset.

Hälytykset ovat myös evolutiivisesti tärkeitä esimerkiksi linnuille ja nisäkkäille. Turvallisuuden tarpeesta huolimatta monet ihmiset jättävät hälytykset huomiotta.

Tämä väitöskirja keskustelee tekijöistä, jotka vaikuttavat hälytysten tehokkuuteen. Toisin kuin eläinmaailmassa, jossa hälytykset perustuvat luonnollisiin signaaleihin ihmisten käyttämät hälytykset toteutetaan erilaisten teknisten keinojen avulla. Usein esitetään abstrakteja signaaleja. Tämä tekee ihmisten hälytyksistä erityislaatuisia biologisessa maailmassa. Nykyisen tekniikan kehittyneisyys tekee mahdolliseksi uusien hälytysmenetelmien kehittämisen. Erityisesti puheeseen perustuvat hälytykset tarjoavat mielenkiintoisia mahdollisuuksia kuten myös erilaiset multimodaaliset järjestelmät. Tässä työssä esitetään kognitiiviseen psykologiaan perustuva malli hälytyksiin liittyvistä ihmisen ja teknologian vuorovaikutusprosesseista.

REFERENCES

AAISB (2006). *Aircraft Accident Report Helios Airways Flight HCY522 Boeing 737-31S at Grammatiko, Hellas on 14 August 2005.* Athens: Hellenic Republic Ministry of Transport & Communications Air Accident Investigation & Aviation Safety Board.

Aeronautics Committee (1998). *Report on Aviation Safety*. New York: Association of the Bar of the City of New York.

Addams-Moring, R. (2007). How can I trust this warning? Mobile early warning as a usability challenge. In H. Sullivan & M. Hakkinen, (Eds.), *Proceedings of the Workshop on The Role of Psychology in Disaster Mitigation: Meeting the Human Challenges of Risk Communication, Preparedness, and Response to Disaster Through a Scientific Approach.* Retrieved from: http://www.psydrr.org/index.php?page=2007-workshop-proceedings.

Aesop (1998). (R. Temple & O. Temple, trans.). London: Penguin Books.

Arons, B. (1991). Hyperspeech: Navigating in speech-only hypermedia. In *Proceedings of the third annual ACM conference on Hypertext*. New York: ACM.

Arons, B., & Mynatt, E. (1994). The Future of Speech and Audio in the User Interface. *ACM SIGCHI Bulletin*, 26(4), 44-47.

Ball, J.E.D. (1972). *Television and the Deaf*. Washington, DC: Public Broadcasting Service.

Baron, R. (2005). *Barriers to Effective Communication: Implications for the Cockpit (Guest Editorial)*. AirlineSafety.com. Retrieved from: http://www.airlinesafety.com/editorials/BarriersToCommunication.htm

Bergman, E. & Johnson, E. (1995). Toward Accessible Human-Computer Interaction. In J. Nielsen (Ed.), *Advances in Human Computer Interaction 5* (pp.87-114). Norwood, NJ: Albex Publishing.

Berntson, G.G., Boysen, S.T., Bauer, H.R.,Torello, M.S. (1989). Conspecific screams and laughter: Cardiac and behavioral reactions of infant chimpanzees. *Developmental Psychobiology*, 22(8), 1098-2302.

Black, A.W. (2002). Perfect Synthesis for all of the people all of the time. In *Proceedings of the 2002 IEEE Workshop on Speech Synthesis*. New York: IEEE.

Blenkhorn, P. & Evans, G. (2001). The Architecture of a Windows Screen Reader. In C. Marincek, C. Buhler, H. Knops, & R. Andrich (Eds.), *Assistive Technology - Added Value to the Quality of Life* (pp. 119-123). Amsterdam: IOS Press.

Bliss, James P.(2003). An investigation of alarm related accidents and incidents in aviation. *International Journal of Aviation Psychology*, 13(3), 249-268.

Bliss, J.P., & Acton, S.A. (2003). Alarm mistrust in automobiles: how collision alarm reliability affects driving. *Applied Ergonomics*, 34, 499-500.

Boeing (2006). *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations, 1959-2005*. Seattle: Boeing Commercial Aircraft Company.

Boukydis, C. F. Z. & Burgess, R. L. (1982). Adult physiological response to infant cries: Effects of temperament of infant, parental status, and gender. *Child Development*, 53, 1291-1298.

Bourne, L.E. & Yaroush, R.A. (2003). Stress and Cognition: A cognitive psychological perspective. Unpublished manuscript.

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1995). Experimentally derived guidelines for the creation of earcons. In *Proceedings of BCS-HCI '95*, Volume 2, 155-159. Berlin: Springer.

Broadbent, D.E. (1958). *Perception and communications.* London: Pergamon Press.

Brooks, M. Challenges for Warning Populations with Sensory Disabilities. In B. Van de Walle and M. Turoff, (Eds.), *Proceedings of the 3rd International ISCRAM Conference*, Newark: ISCRAM.

Carp, B.L. (2001). Fire of Liberty: Firefighters, Urban Voluntary Culture, and the Revolutionary Movement. *William and Mary Quarterly, 3rd Series*, 58(4), 781-818.

CEA(2006). *CEA 2003C - Digital Audiobook File Format and Player Requirements.* Consumer Electronics Assocation. Retrieved from: http://www.ce.org/Standards/browseByCommittee_2646.asp

Cherry, C. (1953). Some experiments on the reception of speech with one and with two ears. *Journal of the Acoustical Society of America*, 25, 975-979.

Conley, C.W. (1968). The Great Japanese Balloon Offensive. *Air University Review*, January-February 1968.

Conzola, V. C., & Wogalter, M. S. (2001). A communication–human information processing (C–HIP) approach to warning effectiveness in the workplace. *Journal of Risk Research*, 4, 309-322.

Cox, J.R., Engebretson, A.M., Garfield, S., Miller, J.D., Hakkinen, M.T., Monsen, R., Sachs, R., Scott, B., & Vemula, R. (1975). Speech and hearing computer system. In *Biomedical Computer Laboratory Annual Progress Report No. 11*. St. Louis: Washington University School of Medicine.

DAISY (1998). *DAISY 2.0 Recommendation.* The DAISY Consortium.

DAISY (2002). *ANSI-NISO Z39.86-2002 - Digital Talking Book Standard.* US Library of Congress and The DAISY Consortium. Retrieved from: http://www.daisy.org/z3986

DAISY (2006a). *Digital Accessible Information System*. The DAISY Consortium. Retrieved from: http://www.daisy.org

DAISY (2006b). *DAISY History.* The DAISY Consortium. Retrieved from: http://www.daisy.org/about_us/history.asp

Davis,I. & Izadkhah, Y.O. (2008). Tsunami early warning system (EWS) and its integration within the chain of seismic safety. *Disaster Prevention and Management*, 17(2), 281-291.

Deathridge, B.H. (1972). Auditory and other sensory forms of information presentation. In *Van Cott, H.P., & Kincade, R.G., (eds). Human Engineering Guide to Equipment Design (revised edition).* Washington, DC: US Government Printing Office.

Delogu, C., Conte, S., & Sementina, C (1998) Cognitive factors in the evaluation of synthetic speech. *Speech Communication*, 24(2),153-168.

DeMeglio, M., Hakkinen, M., & Kawamura, H. (2002). Accessible Interface Design: Adaptive Multimedia Information System (AMIS). In K. Miesenberger, J. Klaus, & W. Zagler (Eds.), *Computers Helping People with Special Needs.*, Springer Lecture Notes in Computer Science, Vol. 2398 (pp. 406-412). Berlin: Springer.

Deutsch, J.A., & Deutsch, D. (1963). Attention: Some Theoretical Considerations. *Psychological Review*, 70(1), 80-90.

DOJ (2007). *ADA Best Practices Tool Kite for State and Local Governments.* Washington, DC: Department of Justice.

Doyle, A.C. (1938). *The Complete Sherlock Homes.* Garden City, NY: Garden City Publishing.

EBU (1996). Reaching Forward to the 21st Century - Next Generation of Talking Books. European Blind Union.

Edworthy, J. (1994). The design and implementation of non-verbal auditory warnings. *Applied Ergonomics*, 25(4), 202-10.

Edworthy, J. & Hellier, E. (2006). Alarms and human behaviour: implications for medical alarms. British Journal of Anaesthesia, 97(1), 12-17.

Edworthy, J., Heiller, E., Walters, K., Clift-Matthews, W., & Crowther, M. (2003). Acoustic, Semantic and Phonetic Influences in Spoken Warning Signal Words. *Applied Cognitive Psychology*, 17: 915-933.

Flight Safety Foundation (2006). CVR/FDR Transcripts. Aviation Safety Network, a Service of the Flight Safety Foundation. Retrieved from: http://aviation-safety.net/investigation/cvr/transcripts/

Forman, R.E. Resignation as a Collective Behavior Response. *The American Journal of Sociology*, 69(3), 285-290.

Frank, Pat (1962). *How to Survive the H Bomb and Why.* New York: Lippincott.

Frederiksen, J.K. & Slobodchikoff, C.N. (2007). Referential specificity in alarm calls of the black-tailed prairie dog. *Ethology, Ecology & Evolution,* 19, 87-99.

Gatty, A. (1848). The Bell: Its origin, history, and uses. London: George Bell.

Gong, L., & Lai, J. (2003). To mix or not to mix synthetic speech and human speech? Contrasting impact on judge-related task performance versus self-rated performance and attitudinal responses. *International Journal of Speech Technology*, 6, 121-131.

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics*, 42, 9, 1233 - 1248.

Gregg, C.E., Houghton, B.F., Paton, D., Johnston, D.M., Swanson, D.A., & Yanagi, B.S. (2007). Tsunami Warnings: Understanding in Hawai'i. *Natural Hazards*, 40(1), 71-87.

Hakkinen, M.T. & Engebretson, A.M. (1978). PARAPET (Programmable Audio Recording And Playback Editing Tool): A general utility program for RAP-III. In *Central Institute for the Deaf Research Periodic Progress Report No. 22*. St. Louis: Central Institute for the Deaf.

Hakkinen, M.T. & Williges, B.H. (1982). Synthesized voice warning messages: Effects of alerting cues and message environment. Paper presented at the *26th Annual Meeting of the Human Factors Society*, Seattle, WA.

Hakkinen, M.T. & Williges, B.H. (1984). Synthesized Warning Messages: Effects of an Alerting Cue in Single- and Multiple Function Voice Synthesis Systems. *Human Factors*, 26(2), 1984, 319-330.

Hakkinen, M.T. (1996) Non-Visual Browsing. Presentation at the *WWW5 Workshop on Web Accessibility*, Paris, France.

Hakkinen, M.T., and DeWitt, (1997). An Accessible Web Browser: The Application of First Order Design Principles in pwWebSpeak. Presentation at the *CSUN Conference on Technologies and Disabilities.* Los Angeles, CA.

Hakkinen, M.T. (1998). Audio Browsing. Presentation at the *First W3C Voice Browsing Workshop*, Boston, MA.

Hakkinen, M.T. & Kerscher, G. (1998). Structured Audio: Using Document Structure to Navigate Audio Information. Presentation at the *CSUN Conference on Technologies and Disabilities*. Los Angeles, CA.

Hakkinen, M.T. & Kerscher, G. (2000). Applying a Navigation Layer to Digital Talking Books: SMIL, XML, and NCX. Presentation at the *WWW9 Multimedia Workshop.* Amsterdam.

Hakkinen, M., DeMeglio, M., & Kawamura, H. (2002). gNCX: Structure-based Navigation to Enhance Usability of Digital Media. Poster presentation at *WWW 2002*, Honolulu.

Hakkinen, M.T. (2007). Human Factors Issues in the Design of Multimodal Public Warning Systems. In *Proceedings of the International Disaster Reduction Conference*, Harbin, China. Davos: IDRC.

Hakkinen, M.T. & Sullivan, H.T. (2007). Reaching Vulnerable Populations: Importance of Accessibility in the Design of Preparedness and Warning Systems. In *Proceedings of the 2nd International ISCRAM CHINA Workshop,* Harbin, China.

Hakkinen, M.T. & Sullivan, H.T. (2007). Effective Communication of Warnings and Critical Information: Application of Accessible Design Methods to Auditory Warnings. In B. Van de Walle, P. Burghardt & C. Nieuwenhuis (Eds.), *Proceedings of the 4th International ISCRAM Conference* (pp. 167-171). Delft: ISCRAM.

Hansson, K., Sönnebo, L., & Lindholm, J. (1994). Digital Talking Books - A Report from a Practical, Ongoing Project. In *Proceedings of ICCHP 1994:* 262-268.

Harkins, J., Strauss, K.P., & Vanderheiden, G. (2005). Emergency notification and communication. In: *State of the Science Conference Proceedings Research and Policy Recommendations*. Washington, DC: Gallaudet University.

Hindus, D., Schmandt, C., and Horner, C. (1993). Capturing, structuring, and representing ubiquitous audio. ACM Trans. Inf. Syst. 11, 4, 376-400.

Ho, C. & Spence, C. (2005). Assessing the Effectiveness of Various Auditory Cues in Capturing a Driver's Visual Attention. *Journal of Experimental Psychology: Applied*. 11(3), 157.

Homer (1967). *The Odyssey of Homer.* (R. Lattimore, trans.). New York: Harper and Rowe Publishers.

Hollen, L.I., & Manser, M.B. (2007). Motivation before meaning: motivational information encoded in meerkat alarm calls develops earlier than referential information. *American Naturalist*, 169:758–767.

Hurd, C.R. (1996). Interspecific attraction to the mobbing calls of black capped chickadees (*Parus atricapillus*). *Behav Ecol Sociobio*, 38:287-292.

IDPF (2006). *Open eBook Publication Structure Specification 1.2.* International Document Publishing Forum. Retrieved from: http://www.idpf.org/oebps/oebps1.2/index.htm

IEEE (2004). *Milestones:Electric Fire Alarm System, 1852*. Retrieved from: http://www.ieeeghn.org/wiki/index.php/Milestones:Electric_Fire_Alarm_System,_1852

ISO (2006). *Catalog of ISO Standards*. International Standards Organization, Switzerland. Retrieved from: http://www.iso.ch

Jones, E., Woolven, R., Durodie,B., & Wessely, S. (2004). Civilian Morale During the Second World War: Responses to Air Raids Re-examined. *Social History of Medicine*, 17(3), 463-479.

Klatt, D. H. (1987) Review of text-to-speech conversion for English, *Journal of the Acoustical Society of America*, 82: 737 - 793.

Lachman, R., Tatsuoka, M., and Bonk, W.J. (1961). "Human behavior during the tsunami of May, 1960." Science, 1961, 133, 1405-1409.

Lai, J., Wood, D., & Considine, M. (2000). The effect of task conditions on the comprehensibility of synthetic speech. In *Proceedings of CHI 2000*. New York: ACM.

Laughery, K.R. (2006). Safety Communications: Warnings. *Applied Ergonomics*, 37(4), 467-478.

Lindquist, T. E., Fainter, R. G., and Hakkinen, M. T. (1985). GENIE: A Modifiable Computer-Based Task for Experiments in Human-Computer Interaction. *International Journal of Man-Machine Studies*, 23(4), 391-406

Liu, Y.C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers performance in advanced traveler information systems. *Ergonomics,* 44(4), 425-442.

Longfellow, H.W. (2000). *Henry Wadsworth Longfellow: Poems and Other Writings.* Des Moines, IA: The Library of America.

Loukopoulos, L.D., Dismukes, R.K., & Barshi, I. (2009). The Multitasking Myth. Burlington: Ashgate Publishing.

LOC (1998). *Digital Talking Books: Planning for the Future.* National Library Service for the Blind and Physically Handicapped. Washington, DC: US Library of Congress.

Luce, P.A., Feustel, T.C., & Pisoni, D.B. (1983). Capacity demands in short term memory for synthetic and natural speech. *Human Factors,* 25.

Marler, P. & Slabbekoorn, H. (editors) (2004). *Nature's Music: The Science of Birdsong*. New York: Academic Press..

Mileti, D. & Sorensen, J. (1990). *Communication of emergency public warnings.* ORNL-6609, Oak Ridge: Oak Ridge National Laboratory.

NASA (1995). *Man-Systems Integration Standards Revision B, July 1995.* Washington, DC: National Aeronautics and Space Administration.

NASA (2006). *Aviation Safety Reporting System.* National Aeronautics and Space Administration. Retrieved from: http://akama.arc.nasa.gov/ASRSDBOnline/QueryWizard_Begin.aspx

NCTS (2000). *Effective Disaster Warnings: Report by the Working Group on Natural Disaster Information Systems, Subcommittee on Natural Disaster Reduction.* Washington, DC: National Science and Technology Council.

Norman, D.A. (1968). Toward a theory of memory and attention. *Psychological Review,* 75(6), 522-536.

NTSB (2006). *Safety Report on the Treatment of Safety-Critical Systems in Transport Airplanes.* Number SR-06/02. Washington, DC: National Transportation Safety Board.

O'Gorman, J.G. (1979). The orienting reflex: Novelty or significance detector? *Psychophysiology*, 16(3).

Palen, L. & Liu, S.B. (2007). Citizen Communications in Crisis: Anticipating a Future of ICT-Supported Public Participation. In *Proceedings of ACM CHI 2007.* New York: ACM.

Patterson, R. D. (1982). Guidelines for auditory warning systems on civil aircraft, Report No. 82017. London: Civil Aviation Authority.

Pavlov, I.P. (1927). *Conditioned Reflexes.* (G.V. Anrep, trans.). Mineola, NY: Dover Publications.

Peryer, G., Noyes, J., Pleydell-Pearce, K., & Lieven, N. (2005). Auditory Alert Characteristics: A Survey of Pilot Views. *International Journal of Aviation Psychology*, 15(3), 233-250

Pritchett, A.R. & Hansman, R.J. (1997). Pilot non-conformance to alerting system commands during closely spaced parallel approaches," In *Proceedings of Digital Avionics Systems Conference, 1997.* AIAA/IEEE, Volume 2.

Proulx, G., Laroche, C., Jaspers-Fayer, F. & Lavallée, R.(2001). Fire Alarm Signal Recognition. *Canadian Institute for Research in Construction, Internal Report No. 828*.

Rainey, H.J., Zuberbuhler, K. & Slater, P.J.B. (2004). Hornbills can distinguish between primate alarm calls. *Proc. R. Soc. Lond. B,* 271, 755–759

Reason, J. (1990). *Human Error*. Cambride University Press. Cambridge.

Reason, J. (2000). Human Errors: Models and Management. *Western Journal of Medicine*, 172(6): 393–396.

Rostoker, W., Bronson, B., & Dvorak, J. (1984). The Cast-Iron Bells of China. *Technology and Culture*, 25(4), 750-767.

Saariluoma, P. (2005). Explanatory frameworks for interaction design. In: Pirhonen, A., Isomäki, H., Roast C. & Saariluoma, P. (Eds.), *Future interaction design*. London: Springer

Salvendy, G. (1987). *Handbook of Human Factors.* New York: John Wiley & Sons.

Salvendy, G. (ed) (2006). *Handbook of Human Factors and ergonomics*, 3rd edition. Hoboken: Wiley.

Schmideberg, M. (1942). Some Observations on Individual Reactions to Air Raids. *Int. J. Psycho-Anal.*, 23:146-176.

Selcon, S.J., Taylor, R.M., & McKenna, F.P. (1995). Integrating Multiple Information Sources: Using redundancy in the design of warnings. *Ergonomics*, 38(11), 2362-2370.

Shen, S. (1987). Acoustics of Ancient Chinese Bells. *Scientific American*, 256(4).

Simpson, C. A., and Hart, S. G. (1977). Required attention for synthesized speech perception for three levels of linguistic redundancy. Paper presented at the *93rd meeting of the Acoustical Society of America*. New York: ASA.

Simpson, C. A., and Williams, D. H. (1980). Response time effects of alerting tone and semantic context for synthesized voice warnings. *Human Factors*, 22, 319-330.

Simpson, C.A. (1983). Advanced Technology – new fixes or new problems? Verbal communication in the aviation system. Paper presented at *Beyond pilot error: a symposium of scientific focus.* Washington, DC: Air Line Pilots Association, Washington.

Simpson, C.A. (2007). Doing science on auditory display design in the cockpit: Merging laboratory rigor and the aircraft cockpit environment. In *Proceedings of the 13th International Conference on Auditory Display.* Montreal. 139-142.

Smith, G.A., Splaingard, M., Hayes, J.R., Xiang, H. (2006). Comparison of a Personalized Parent Voice Smoke Alarm With a Conventional Residential Tone Smoke Alarm for Awakening Children. *PEDIATRICS*, 118(4), 1623-1632.

SNAKE (2005). *Report on Special Needs Assessment for Katrina Evacuees (SNAKE) Project.* Washington, DC: National Organization on Disability.

Sokolov, E.N. (1963). Higher Nervous Functions: The orienting reflex. *Annual Review of Physiology*, 25, 545-580.

Spence, C., & Driver, J. (1999). A new approach to the design of multimodal warning signals. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics: Vol. 4. Job design, product design and human-computer interaction* (pp. 455–461). Hampshire, England: Ashgate Publishing.

Stanton, N.A. (1994). *Human Factors of Alarm Design*. London: Taylor and Francis Publishers.

Stanton, N.A. & Edworthy, J. (1999). *Human Factors in Auditory Warnings.* Aldershot: Ashgate Publishers.

Stephan, K.L., Smith, S.E., Martin, R.L., Parker, S.P.A., and McNally, K.I. Learning and rentention of associations between auditory icons and denotative references: Implications for the design of auditory warnings. *Human Factors*, 48(2). 288-299.

Stephanidis, C. (1995). Towards User Interfaces for All: Some Critical Issues. In *Symbiosis of Human and Artifact (edited by Anzai, Ogawa, and Mori)*. Amsterdam: Elsevier.

Strayer, D.L., Drews, F.A., & Johnston, W.A. (2003). Cell Phone-induced Failures of Visual Attention During Simulated Driving. *Journal of Experimental Psychology: Applied*, 9(1), 23-32.

Sullivan, H.T. & Hakkinen, M. T. (2006). Disaster Preparedness for Vulnerable Populations: Determining Effective Strategies for Communicating Risk, Warning, and Response. Paper presented at the *Third Annual Magrann Research Conference on The Future of Disasters in a Globalizing World*. New Brunswick, New Jersey.

Sullivan, H.T., Hakkinen, M.T., Piechocinski, D., and Merritt, T.R. (2008). Emergency Notifications in Mobile Contexts: Effective Approaches for Universal Design. In F. Fiedrich and B. Van de Walle, (Eds.), *Proceedings of the 5th International ISCRAM Conference.* Delft: ISCRAM.

Sullivan, H., Hakkinen, M., & DeBlois, K. (2010). Communicating critical information using mobile phones to populations with special needs. *International Journal of Emergency Management*, 7(1), 6–16.

Sullivan, H.T., Hakkinen, M.T., & Piechocinski, D. (2009). Improving participation, accessibility and compliance for campus wide mobile emergency alerting systems. In Löffler, J. & Klann, M. (Eds.), *Mobile Response*, Springer Lecture Notes in Computer Science, Vol. 5424 (pp. 32-40). Berlin: Springer.

Tanner, W.P., & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61(6).

Tajima, A. (2004). Fatal miscommunication: English in aviation safety. *World Englishes*. 23(3), 451–470.

Templeton, C.N., Greene, E. & Davis, K. (2005). Allometry of Alarm Calls: Back-Capped Chickadees Encode Information About Predator Size. *Science*, 301, 1934-1937.

Templeton, C.N. & Greene, E. (2007). Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. *Proceedings of the National Academy of the Sciences of the United States.* 104(13), 5479-5482.

Thatcher, J. (1994). Screen reader/2: access to OS/2 and the graphical user interface. In *Proceedings of the First Annual ACM Conference on Assistive Technologies*. 39-46. New York: ACM.

Treisman, A. (1969). Strategies of and models of selective attention. *Psychological Review*, 76, 282-299.

Treisman, A., & Davies, A. (1973). Divided attention to Eye and Ear. In *S. Kornblum, (ed), Attention and Performance IV*. New York: Academic Press.

Turner, D. (2000). The Functions of Fossils: Inference and Explanation in Functional Morphology. *Stud. Hist. Phil. Biol. & Biomed. Sci.*, 31(1), 193–212.

Tyack, G.S. (1898). *A Book about Bells.* White Fish: Kessinger Publishing.

Ulfvengren, P. (2003). *Design of Natural Warning Sounds in Human-Machine Systems.* Doctoral Thesis. Stockholm: KTH Royal Institute of Technology.

US Air Force (2006). C-5 accident investigation board complete. Air Force Link. Retrieved from: http://www.af.mil/news/story.asp?storyID=123021742

USDOD (1997). *AIRCREW STATION ALERTING SYSTEMS. MIL-STD-411F.* Washington, DC: United States Department of Defense.

USDOD (1999). *Human Engineering. MIL-STD-1472F.* Washington, DC: United States Department of Defense Design.

Van Cott, H.P., & Kincade, R.G., eds. (1972). *Human Engineering Guide to Equipment Design (revised edition)*. Washington, DC: US Government Printing Office.

Vanderheiden, G.C., Nelson, R.K., Yan, L., & Sesto, M.E. (2004). Strategies for Mainstream Cellular Phone Use by Individuals with Moderate to Severe Cognitive Impairments. In *Proceedings of the Human Factors and Ergonomics Society Conference 2004*, 937-940. Santa Monica: HFES.

Volkmann, J.E. & Graham, M.L. (1942). A Survey on Air Raid Alarm Signals. *J. Acoust. Soc. Am*., 14(1), 1-9.

Von Falkenhausen, L. (1994). *Suspended Music: Chime-Bells in the Culture of Bronze Age China.* Berkeley: University of California Press.

Wade, N. (2006). *Before the Dawn*. New York: Penguin Press.

W3C (1996). *Cascading Styles Sheets 1.0.* World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/CSS1

W3C (1997). *Cascading Styles Sheets 2.0*. World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/REC-CSS2

W3C (1998). *Synchronized Multimedia Integration Language (SMIL) 1.0*. World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/REC-SMIL

W3C (1999). *Web Content Accessibility Guidelines 1.0.* World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/WAI-WEBCONTENT

W3C (2000). *Authoring Tools Accessibility Guidelines 1.0*. World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/ATAG10/

W3C (2002). *User Agent Accessibility Guidelines 1.0*. World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/UAAG10/

W3C (2004). *Speech Synthesis Markup Language (SSML) 1.0*. World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/speech-synthesis/

W3C (2005). *Synchronized Multimedia Integration Language (SMIL) 2.1*. World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/2005/REC-SMIL2-20051213/

W3C (2006a). *Cascading Styles Sheets 2.1.* World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/TR/CSS21

W3C (2007). *Web Accessibility Initiative (WAI)*. World Wide Web Consortium (W3C). Retrieved from: http://www.w3.org/WAI.

Wickens, C.D. (1984). *Engineering Psychology and Human Performance.* Columbus: Charles E. Merrill Publishing Co..

Williges, B.H., & Williges, R.C. (1982). Structuring human/computer dialogue using speech technology. In *Proceedings of the workshop on standardization for speech I/O technology*. Gaithersburg, MD: National Bureau of Standards.

Williges, B.H., Schurick, J.M., Spine, T.M., and Hakkinen, M.T. (1986). Using Speech in the Human-Computer Interface. Chapter 7 in *Human-Computer Dialogue Design (eds. R.W. Ehrich & R.C. Williges)*, Amsterdam: Elsevier.

Y. Xiao, F. J. Seagull, F. Nieves-Khouw, N. Barczak, & S.Perkins. (2004). Organizational-historical analysis of the failure-torespond-to-alarms problems. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 34(6).

Zajicek, M., & Jonsson, I-M. (2005). Evaluation and Context for In-car Speech Systems for Older Adults. In *Proceedings of CLIHC'05*, Cuernavaca, Mexico.

Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour*, 59(5), 917-927.

Appendix A. MMASL Schema Documentation

Schema **mmasl.xsd**

URL: http://www.talkinginterfaces.org/mmasl/schema/mmasl10.xsd


```
<?xml version="1.0" encoding="utf-8"?>
<!--MMASL - MultiModal Alarm Specification Language
Version: 1.0.0
Author: Markku T. Häkkinen - University of Jyväskylä
Date: 12 June 2010
Revised: 5 October 2010 - comments and attribute edits

-->
<!--Description: MMASL is an XML application for specifying alarms and alerts
in a manner that supports presentation and interaction rules that enhance the
efficicacy of the message.-->
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified" attributeFormDefault="unqualified"
version="1.0">
  <xs:element name="mmasl">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="alarm" maxOccurs="unbounded">
          <xs:complexType mixed="true">
            <xs:sequence minOccurs="0">
              <xs:element name="cue">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element ref="auditory" />
                    <xs:element ref="visual" />
                    <xs:element ref="tactile" />
                  </xs:sequence>
                  <xs:attribute ref="src" use="optional" />
                  <xs:attribute ref="lang" use="optional" />
                  <xs:attribute ref="dur" use="optional" />
                  <xs:attribute ref="repeatCount" use="optional" />
                  <xs:attribute ref="repeatDur" use="optional" />
                  <xs:attribute ref="end" use="optional" />
                </xs:complexType>
```

```
        </xs:element>
        <xs:element name="message">
          <xs:complexType>
            <xs:sequence>
              <xs:element ref="auditory" />
              <xs:element ref="visual" />
              <xs:element ref="tactile" />
            </xs:sequence>
            <xs:attribute ref="src" use="optional" />
            <xs:attribute ref="lang" use="optional" />
            <xs:attribute ref="dur" use="optional" />
            <xs:attribute ref="repeatCount" use="optional" />
            <xs:attribute ref="repeatDur" use="optional" />
            <xs:attribute ref="end" use="optional" />
          </xs:complexType>
        </xs:element>
      </xs:sequence>
      <xs:attribute name="id" />
      <xs:attribute name="capURI" type="xs:anyURI">
        <xs:annotation>
          <xs:documentation>URI to originating CAP message if
          used to generate alarm</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute ref="lang" />
    </xs:complexType>
  </xs:element>
 </xs:sequence>
 </xs:complexType>
</xs:element>
<xs:attribute name="lang" type="xs:language">
  <xs:annotation>
    <xs:documentation>xml:lang of element content from ISO
    639-1</xs:documentation>
  </xs:annotation>
</xs:attribute>
<xs:attribute name="src" type="xs:anyURI">
  <xs:annotation>
    <xs:documentation>URI to media or text
    resource</xs:documentation>
```

```
  </xs:annotation>
</xs:attribute>
<xs:attribute name="id" type="xs:ID">
  <xs:annotation>
    <xs:documentation>xml:id unique identifier for
    element</xs:documentation>
  </xs:annotation>
</xs:attribute>
<xs:attribute name="repeatCount" type="xs:string">
  <xs:annotation>
    <xs:documentation>inherited from SMIL. Specifies the number
    of iterations of the simple duration. It can have the
    following attribute values: numeric value This is a (base 10)
    "floating point" numeric value that specifies the number of
    iterations. It can include partial iterations expressed as
    fraction values. A fractional value describes a portion of
    the simple duration. Values must be greater than 0.
    "indefinite" The element is defined to repeat indefinitely
    (subject to the constraints of the parent time
    container).</xs:documentation>
  </xs:annotation>
</xs:attribute>
<xs:attribute name="repeatDur" type="xs:string">
  <xs:annotation>
    <xs:documentation>inherited from SMIL. Specifiies Specifies
    the total duration for repeat. It can have the following
    attribute values: Clock-value Specifies the duration in
    element active time to repeat the simple duration.
    "indefinite" The element is defined to repeat indefinitely
    (subject to the constraints of the parent time
    container).</xs:documentation>
  </xs:annotation>
</xs:attribute>
<xs:attribute name="end" type="xs:string">
  <xs:annotation>
    <xs:documentation>inherited from SMIL. The end attribute
    allows the author to constrain the active duration by
    specifying an end value using a simple offset, a time base,
    an event-base, a syncbase, or DOM methods calls. Subset of
    SMIL values, time, accesskey, or DOM
```

```
      event.</xs:documentation>
    </xs:annotation>
  </xs:attribute>
  <xs:attribute name="dur" type="xs:string">
    <xs:annotation>
      <xs:documentation>inherited from SMIL. Specifies the simple
      duration. The attribute value can be: Clock-value - Specifies
      the length of the simple duration, measured in element active
      time. Value must be greater than 0.</xs:documentation>
    </xs:annotation>
  </xs:attribute>
  <xs:attribute name="class" type="xs:string">
    <xs:annotation>
      <xs:documentation>The class attribute assigns a class name or
      a set of class names to an element. Any number of elements
      may be assigned the same class name or names. Multiple class
      names must be separated by white space
      characters.</xs:documentation>
    </xs:annotation>
  </xs:attribute>
  <xs:element name="auditory" type="auditoryType">
    <xs:annotation>
      <xs:documentation>Auditory information container. Elements
      within include audio for audio file playback (e.g., mp3, ogg)
      and tts for text to speech output.</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="tactile" type="tactileType">
    <xs:annotation>
      <xs:documentation>Tactile information container. Elements
      include vibration for mobile device vibration and braille for
      text to be rendered by a Braille display</xs:documentation>
    </xs:annotation>
  </xs:element>
  <xs:element name="visual">
    <xs:annotation>
      <xs:documentation>Visual information container. Elements
      include text, img, and animation.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
```

```
    <xs:complexContent>
      <xs:extension base="visualType">
        <xs:sequence>
          <xs:element ref="par" />
        </xs:sequence>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="seq" type="visualType">
  <xs:annotation>
    <xs:documentation>inherited from SMIL. seq allows the author
    to explicitly defined sequential presentation of media
    elements. seq is default for auditory and
    visual.</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:element name="par" type="visualType">
  <xs:annotation>
    <xs:documentation>inherited from SMIL. par allows the author
    to explicitly define media which should be played in
    parallel. Can be used in the visual element to simultaneously
    display text, image, and or animation</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:complexType name="visualType">
  <xs:annotation>
    <xs:documentation>defines the elements valid for visual
    media</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="text">
      <xs:annotation>
        <xs:documentation>inherited from SMIL. references an
        external text file or contains text to be rendered
        visually.</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:attribute ref="src" />
        <xs:attribute ref="lang" />
```

```
      <xs:attribute ref="class" />
      <xs:attribute ref="dur" />
      <xs:attribute ref="end" />
    </xs:complexType>
  </xs:element>
  <xs:element name="img">
    <xs:annotation>
      <xs:documentation>inherited from SMIL. refernces an
      external image file.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attribute ref="src" use="required" />
      <xs:attribute name="alt" type="xs:string" use="required">
        <xs:annotation>
          <xs:documentation>Alternate text equivalent of image
          as per W3C WCAG 2.0</xs:documentation>
        </xs:annotation>
      </xs:attribute>
      <xs:attribute ref="dur" />
      <xs:attribute ref="end" />
    </xs:complexType>
  </xs:element>
  <xs:element name="animation">
    <xs:annotation>
      <xs:documentation>inherited from SMIL. references and
      external animation resource. Types to be
      defined.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:attribute ref="src" />
      <xs:attribute ref="lang" />
      <xs:attribute ref="class" />
      <xs:attribute ref="dur" />
      <xs:attribute ref="end" />
    </xs:complexType>
  </xs:element>
 </xs:sequence>
</xs:complexType>
<xs:element name="audio">
 <xs:annotation>
```

```
    <xs:documentation>inherited from SMIL. defines the audio
    media element.</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:attribute ref="src" />
    <xs:attribute ref="lang" />
    <xs:attribute ref="class" />
  </xs:complexType>
</xs:element>
<xs:element name="tts">
  <xs:annotation>
    <xs:documentation>defines the Text To Speech media
    element.</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:attribute ref="src" />
    <xs:attribute ref="lang" />
    <xs:attribute ref="class" />
  </xs:complexType>
</xs:element>
<xs:complexType name="auditoryType">
  <xs:annotation>
    <xs:documentation>defines the elements valid for auditory
    media</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element ref="audio" />
    <xs:element ref="tts" />
  </xs:sequence>
</xs:complexType>
<xs:complexType name="tactileType">
  <xs:annotation>
    <xs:documentation>defines the elements valid for tactile
    media</xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="vibrate">
      <xs:annotation>
        <xs:documentation>references a vibration pattern
        file.</xs:documentation>
```

```
      </xs:annotation>
      <xs:complexType>
        <xs:attribute ref="src" use="required" />
      </xs:complexType>
    </xs:element>
    <xs:element name="braille">
      <xs:annotation>
        <xs:documentation>references an external braille source
        file or is a container for Braille
        text.</xs:documentation>
      </xs:annotation>
      <xs:complexType>
        <xs:attribute ref="src" />
        <xs:attribute ref="class" />
      </xs:complexType>
    </xs:element>
  </xs:sequence>
 </xs:complexType>
</xs:schema>
```

J Y V Ä S K Y L Ä S T U D I E S I N C O M P U T I N G

1 ROPPONEN, JANNE, Software risk management - foundations, principles and empirical findings. 273 p. Yhteenveto 1 p. 1999.

2 KUZMIN, DMITRI, Numerical simulation of reactive bubbly flows. 110 p. Yhteenveto 1 p. 1999.

3 KARSTEN, HELENA, Weaving tapestry: collaborative information technology and organisational change. 266 p. Yhteenveto 3 p. 2000.

4 KOSKINEN, JUSSI, Automated transient hypertext support for software maintenance. 98 p. (250 p.) Yhteenveto 1 p. 2000.

5 RISTANIEMI, TAPANI, Synchronization and blind signal processing in CDMA systems. - Synkronointi ja sokea signaalinkäsittely CDMA järjestelmässä. 112 p. Yhteenveto 1 p. 2000.

6 LAITINEN, MIKA, Mathematical modelling of conductive-radiative heat transfer. 20 p. (108 p.) Yhteenveto 1 p. 2000.

7 KOSKINEN, MINNA, Process metamodelling. Conceptual foundations and application. 213 p. Yhteenveto 1 p. 2000.

8 SMOLIANSKI, ANTON, Numerical modeling of two-fluid interfacial flows. 109 p. Yhteenveto 1 p. 2001.

9 NAHAR, NAZMUN, Information technology supported technology transfer process. A multi-site case study of high-tech enterprises. 377 p. Yhteenveto 3 p. 2001.

10 FOMIN, VLADISLAV V., The process of standard making. The case of cellular mobile telephony. - Standardin kehittämisen prosessi. Tapaustutkimus solukkoverkkoon perustuvasta matkapuhelintekniikasta. 107 p. (208 p.) Yhteenveto 1 p. 2001.

11 PÄIVÄRINTA, TERO, A genre-based approach to developing electronic document management in the organization. 190 p. Yhteenveto 1 p. 2001.

12 HÄKKINEN, ERKKI, Design, implementation and evaluation of neural data analysis environment. 229 p. Yhteenveto 1 p. 2001.

13 HIRVONEN, KULLERVO, Towards better employment using adaptive control of labour costs of an enterprise. 118 p. Yhteenveto 4 p. 2001.

14 MAJAVA, KIRSI, Optimization-based techniques for image restoration. 27 p. (142 p.) Yhteenveto 1 p. 2001.

15 SAARINEN, KARI, Near infra-red measurement based control system for thermo-mechanical refiners. 84 p. (186 p.) Yhteenveto 1 p. 2001.

16 FORSELL, MARKO, Improving component reuse in software development. 169 p. Yhteenveto 1 p. 2002.

17 VIRTANEN, PAULI, Neuro-fuzzy expert systems in financial and control engineering. 245 p. Yhteenveto 1 p. 2002.

18 KOVALAINEN, MIKKO, Computer mediated organizational memory for process control. Moving CSCW research from an idea to a product. 57 p. (146 p.) Yhteenveto 4 p. 2002.

19 HÄMÄLÄINEN, TIMO, Broadband network quality of service and pricing. 140 p. Yhteenveto 1 p. 2002.

20 MARTIKAINEN, JANNE, Efficient solvers for discretized elliptic vector-valued problems. 25 p. (109 p.) Yhteenveto 1 p. 2002.

21 MURSU, ANJA, Information systems development in developing countries. Risk management and sustainability analysis in Nigerian software companies. 296 p. Yhteenveto 3 p. 2002.

22 SELEZNYOV, ALEXANDR, An anomaly intrusion detection system based on intelligent user recognition. 186 p. Yhteenveto 3 p. 2002.

23 LENSU, ANSSI, Computationally intelligent methods for qualitative data analysis. 57 p. (180 p.) Yhteenveto 1 p. 2002.

24 RYABOV, VLADIMIR, Handling imperfect temporal relations. 75 p. (145 p.) Yhteenveto 2 p. 2002.

25 TSYMBAL, ALEXEY, Dynamic integration of data mining methods in knowledge discovery systems. 69 p. (170 p.) Yhteenveto 2 p. 2002.

26 AKIMOV, VLADIMIR, Domain decomposition methods for the problems with boundary layers. 30 p. (84 p.). Yhteenveto 1 p. 2002.

27 SEYUKOVA-RIVKIND, LUDMILA, Mathematical and numerical analysis of boundary value problems for fluid flow. 30 p. (126 p.) Yhteenveto 1 p. 2002.

28 HÄMÄLÄINEN, SEPPO, WCDMA Radio network performance. 235 p. Yhteenveto 2 p. 2003.

29 PEKKOLA, SAMULI, Multiple media in group work. Emphasising individual users in distributed and real-time CSCW systems. 210 p. Yhteenveto 2 p. 2003.

30 MARKKULA, JOUNI, Geographic personal data, its privacy protection and prospects in a location-based service environment. 109 p. Yhteenveto 2 p. 2003.

31 HONKARANTA, ANNE, From genres to content analysis. Experiences from four case organizations. 90 p. (154 p.) Yhteenveto 1 p. 2003.

32 RAITAMÄKI, JOUNI, An approach to linguistic pattern recognition using fuzzy systems. 169 p. Yhteenveto 1 p. 2003.

33 SAALASTI, SAMI, Neural networks for heart rate time series analysis. 192 p. Yhteenveto 5 p. 2003.

34 NIEMELÄ, MARKETTA, Visual search in graphical interfaces: a user psychological approach. 61 p. (148 p.) Yhteenveto 1 p. 2003.

35 YOU, YU, Situation Awareness on the world wide web. 171 p. Yhteenveto 2 p. 2004.

36 TAATILA, VESA, The concept of organizational competence – A foundational analysis. - Perusteanalyysi organisaation kompetenssin käsitteestä. 111 p. Yhteenveto 2 p. 2004.

37 Lyytikäinen, Virpi, Contextual and structural metadata in enterprise document management. - Konteksti- ja rakennemetatieto organisaation dokumenttien hallinnassa. 73 p. (143 p.) Yhteenveto 1 p. 2004.
38 Kaario, Kimmo, Resource allocation and load balancing mechanisms for providing quality of service in the Internet. 171 p. Yhteenveto 1 p. 2004.
39 Zhang, Zheying, Model component reuse. Conceptual foundations and application in the metamodeling-based systems analysis and design environment. 76 p. (214 p.) Yhteenveto 1 p. 2004.
40 Haarala, Marjo, Large-scale nonsmooth optimization variable metric bundle method with limited memory. 107 p. Yhteenveto 1 p. 2004.
41 Kalvine, Viktor, Scattering and point spectra for elliptic systems in domains with cylindrical ends. 82 p. 2004.
42 Dementieva, Maria, Regularization in multistage cooperative games. 78 p. 2004.
43 Maaranen, Heikki, On heuristic hybrid methods and structured point sets in global continuous optimization. 42 p. (168 p.) Yhteenveto 1 p. 2004.
44 Frolov, Maxim, Reliable control over approximation errors by functional type a posteriori estimates. 39 p. (112 p.) 2004.
45 Zhang, Jian, Qos- and revenue-aware resource allocation mechanisms in multiclass IP networks. 85 p. (224 p.) 2004.
46 Kujala, janne, On computation in statistical models with a psychophysical application. 40 p. (104 p.) 2004.,
47 Solbakov, Viatcheslav, Application of mathematical modeling for water environment problems. 66 p. (118 p.) 2004.
48 Hirvonen, Ari P., Enterprise architecture planning in practice. The Perspectives of information and communication technology service provider and end-user. 44 p. (135 p.) Yhteenveto 2 p. 2005.
49 Vartiainen, Tero, Moral conflicts in a project course in information systems education. 320 p. Yhteenveto 1p. 2005.
50 Huotari, Jouni, Integrating graphical information system models with visualization techniques. - Graafisten tietojärjestelmäkuvausten integrointi visualisointitekniikoilla. 56 p. (157 p.) Yhteenveto 1p. 2005.
51 Wallenius, Eero R., Control and management of multi-access wireless networks. 91 p. (192 p.) Yhteenveto 3 p. 2005.
52 Leppänen, Mauri, An ontological framework and a methodical skeleton for method engineering – A contextual approach. 702 p. Yhteenveto 2 p. 2005.
53 Matyukevich, Sergey, The nonstationary Maxwell system in domains with edges and conical points. 131 p. Yhteenveto 1 p. 2005.
54 Sayenko, Alexander, Adaptive scheduling for the QoS supported networks. 120 p. (217 p.) 2005.
55 Kurjenniemi, Janne, A study of TD-CDMA and WCDMA radio network enhancements. 144 p. (230 p.) Yhteenveto 1 p. 2005.
56 Pechenizkiy, Mykola, Feature extraction for supervised learning in knowledge discovery systems. 86 p. (174 p.) Yhteenveto 2 p. 2005.
57 Ikonen, Samuli, Efficient numerical methods for pricing American options. 43 p. (155 p.) Yhteenveto 1 p. 2005.
58 Kärkkäinen, Kari, Shape sensitivity analysis for numerical solution of free boundary problems. 83 p. (119 p.) Yhteenveto 1 p. 2005.
59 Helfenstein, Sacha, Transfer. Review, reconstruction, and resolution. 114 p. (206 p.) Yhteenveto 2 p. 2005.
60 Nevala, Kalevi, Content-based design engineering thinking. In the search for approach. 64 p. (126 p.) Yhteenveto 1 p. 2005.
61 Katasonov, Artem, Dependability aspects in the development and provision of location-based services. 157 p. Yhteenveto 1 p. 2006.
62 Sarkkinen, Jarmo, Design as discourse: Representation, representational practice, and social practice. 86 p. (189 p.) Yhteenveto 1 p. 2006.
63 Äyrämö, Sami, Knowledge mining using robust clustering. 296 p. Yhteenveto 1 p. 2006.
64 Ifinedo, Princely Emili, Enterprise resource planning systems success assessment: An integrative framework. 133 p. (366 p.) Yhteenveto 3 p. 2006.
65 Viinikainen, Ari, Quality of service and pricingin future multiple service class networks. 61 p. (196 p.) Yhteenveto 1 p. 2006.
66 Wu, Rui, Methods for space-time parameter estimation in DS-CDMA arrays. 73 p. (121 p.) 2006.
67 Parkkola, Hanna, Designing ICT for mothers. User psychological approach. – Tieto- ja viestintätekniikoiden suunnittelu äideille. Käyttäjäpsykologinen näkökulma. 77 p. (173 p.) Yhteenveto 3 p. 2006.
68 Hakanen, Jussi, On potential of interactive multiobjective optimization in chemical process design. 75 p. (160 p.) Yhteenveto 2 p. 2006.
69 Puttonen, Jani, Mobility management in wireless networks. 112 p. (215 p.) Yhteenveto 1 p. 2006.
70 Luostarinen, Kari, Resource , management methods for QoS supported networks. 60 p. (131 p.) 2006.
71 Turchyn, Pavlo, Adaptive meshes in computer graphics and model-based simulation. 27 p. (79 p.) Yhteenveto 1 p.
72 Zhovtobryukh, Dmytro, Context-aware web service composition. 290 p. Yhteenveto 2 p. 2006.