Effective Communication of Warnings and Critical Information: Application of Accessible Design Methods to Auditory Warnings

Markku T. Häkkinen
Department of Computer Science & Information Systems, Agora Human Technologies Center
University of Jyväskylä, Jyväskylä, Finland
mhakkine@jyu.fi

Helen T. Sullivan
Department of Psychology
Rider University
Lawrenceville, New Jersey, USA
hsullivan@rider.edu

ABSTRACT

When a system initiates an auditory warning or alert, detection and correct identification of the information by the human recipient can be influenced by a variety of factors. Examples from aviation and public warning demonstrate instances where messages are ignored, not understood or misinterpreted. The reasons why messages may fail can stem from the design of the message itself, environmental conditions, and sensory or cognitive impairments. Based upon experience from several contexts and from the development of assistive technology for people with disabilities, promising design approaches are being explored in research on warning system design. The importance of multimodal warnings, selection of speech type, and internationalization are discussed.

Keywords

Warning, auditory displays, speech displays, speech synthesis, multimodal displays, accessibility.

INTRODUCTION

The use of speech as a modality for information display has grown significantly in the past 30 years, from use in aural warning systems within aircraft cockpits and in assistive technologies for people with visual impairments to today's widespread application in 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. 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. In public warning systems, such as Tsunami alerting towers in Japan and Thailand, spoken instructions augment the non-speech alerting siren.

Speech and auditory displays in general, allow information to be conveyed relatively quickly, no matter where the visual attention of the human 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 developing information systems for people with disabilities, accessible design approaches are used to enable access to those with sensory limitations such as visual impairments or deafness. Though there is growing interest in the challenges faced by those with learning and cognitive impairments, there remains a lack of empirical design guidance. The importance of research into the needs of people with disabilities in public warning has been previously discussed (Brooks, 2006; Sullivan & Hakkinen, 2006).

PUBLIC WARNING

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. Gregg, et al (2007), in a follow-up study in Hawaii found that only 13% of the local population understood the meaning of the warning siren. Reports from Japan and the United States describe continuing failure on the 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. The proposed use of SMS delivered messages poses further challenges to reception of the message, based upon environmental conditions (bright sun), failure to detect the SMS alert (e.g., phone in silent or standby mode), or visual impairment.

RESEARCH & GUIDELINES FOR AUDITORY ALARMS AND SPEECH DISPLAYS

One focus of this work 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 Human Computer 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, operational applications of auditory warnings, as we have seen, appear problematic.

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 (non-speech) alerting cue prior to presentation of critical messages (e.g., MIL-STD-411F) appears to remain valid though unfortunately often applied in an inconsistent manner by designers.

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) show promise. However, there are examples when poorly designed aural warnings can lead to catastrophe, and surveys of aircraft pilots (Peryer, et al, 2005) suggest possible changes.

LESSONS FROM ACCESSIBLE INFORMATION

The move to provide accessibility to electronic information, notably in terms of access to the World Wide Web (W3C, 2006) and to electronic books (Hakkinen & Kerscher, 1998; DAISY, 2006), led to several key developments. The use of structured information formats (e.g., HTML and XML) allowed for information to be presented in multiple modalities (visual, auditory and tactile/Braille), allowed for navigation of the content in a semantically rich manner, and allowed for adaptive styling of the content during rendering. Talking book software can allow multimodal access to structured information using synthetic and natural speech rendering, textual display (with changes in text style, contrast and size), and using synchronized presentation in Braille. Based upon the Synchronized Multimedia Integration Language (W3C, 2005), talking books also allow for adaptive rendering through the use of author, user, and system selected controls. A single talking book may contain multiple languages, with the language of rendering determined by user or system settings. In addition, the talking books can contain information that is selectively displayed based upon the requirements of the user. A common example is that of footnotes, which may optionally be rendered in context of their occurrence in the text. The accessibility concept of providing parallel and selective 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. Williges, et al (1986) and Selcon, et al (1995) have shown the potential value of redundancy of information presentation forms in the context of warnings. For example, augmenting auditory displays with visual rendering would allow reference or review of the information, concurrent with or after the initial presentation. Selective rendering of message content can allow for adaptive delivery of standard messages based upon the receivers' language, experience, environmental conditions or disability.

CURRENT RESEARCH

Research is currently underway by the authors to examine several issues in the design of effective warnings. With the continuing evolution of speech synthesis technologies and reductions in memory costs, designers of warning systems have the option to use synthetic and natural speech displays in creating speech-based warning systems. Previous research has discussed the alerting quality of synthetic speech in some contexts (Simpson & Williams, 1980; Hakkinen & Williges, 1984) and the ability of synthetic speech to convey urgency (Edworthy et al, 2003). However, these studies generally do not take into account the rapid evolution speech synthesis quality (or naturalness). Thus one goal of this research is to better understand the effects of speech characteristics on the detection and understanding of critical information display. Though synthetic speech which maintains non-human qualities may be better at getting the attention of a listener, particularly when there may be competing human conversations, lack of familiarity and poor intelligibility can limit immediate understanding of the synthesized message. This is a particular problem for low occurrence alarms where training or familiarization is not practical. In contrast, more natural sounding synthetic speech (e.g., concatenative or diphone-based synthesis) or pre-recorded natural speech messages may have higher intelligibility but can require additional speech and non-speech cues to enhance the alerting effectiveness of the message. Messaging approaches in which synthetic and natural speech may be mixed to allow delivery of both fixed and dynamic information can pose further problems in terms of task performance (Gong & Lai, 2003). The current research is seeking, through structured comparisons of detection and identification of simulated warning messages, to gain a better empirical understanding on the appropriate use of synthesized and natural speech for effective public warning.

The adaptation of critical information delivery to ensure that a message or alert is received accurately is a second theme of this research. In particular we are examining how messages, authored in a standard format, can be selectively adapted for rendering based on the users' ambient environment and personal characteristics. One of the questions being explored is whether standard alerting formats, such as the Common Alerting Protocol, or CAP (OASIS, 2006), contain sufficient information for effective, multimodal rendering. Using the example of the CAP message, specific element types within the standard convey severity and location information and are combined with free form descriptive text content. In the "all hazards" context, the CAP `description` element can represent human readable information in a variety of written forms from different agencies. In our examination of real world samples from government information sources (e.g., the United States Government's Geological Survey and Weather Service CAP feeds), the free form text can pose challenges in rendering in alternate formats such as synthetic speech or small displays. Such free form text lacks the structural semantics that can aid in rendering and understandability of this critical information. Ideally, considering any human renderable information in a critical message as a triplet of media elements that includes textual, image, and auditory equivalents (augmented with internationalization attributes) would enable receiving devices to select a rendering (single media element or combined elements) appropriate to the user and context. Standardized semantics would further facilitate personalization of the delivered message through application of presentation styling (e.g., application of local style sheets for large print with logical reformatting/reflow, high contrast colors, etc.). This concept is similar to what is seen in the application model of digital talking books, and would require extending standards like CAP to incorporate enhanced content, and for the agencies generating alerts and warnings to support these enhancements.

We anticipate that our research will contribute in the areas of proposed guidelines for speech-based delivery of critical information, both in terms of identifying required elements in the underlying information formats and in the selection of appropriate presentation characteristics. In the context of global, all hazards warning, guidelines should have an empirical foundation and support for application in multi-lingual and multi-cultural environments. To that end, research that supports the development of design guidelines should be replicated across languages and cultural settings to validate their general applicability. This is a particular concern due to the great variability in speech synthesizer quality across languages and the likely variation in audio generation capabilities in devices used world-wide. A by-product of our current work will be methods and tools for the evaluation of natural and synthetic speech for warning systems, and a reusable audio messaging system prototype.

CONCLUSION

With the development of accessible Web content and Digital Talking Books, it became possible to render information 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 structure, and the selective hiding (and revealing) of information (e.g., footnotes). In reviewing the research and context of speech warning systems, it seems probable that some of the key features developed for the Web and Talking Book domain may have conceptual if not practical applicability in the design of effective speech displays for critical and non-critical information.

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. Further, applying principles of software internationalization and localization, in conjunction with user profiles, can allow speech displays to be easily adapted for optimal information transfer for a world where multiple languages abound.

REFERENCES

1. Arons, B., & Mynatt, E. (1994). The Future of Speech and Audio in the User Interface. *ACM SIGCHI Bulletin*, 26(4), 44-47.
2. Brewster, S.A., Wright, P.C., and Edwards, A.D.N (1995). Experimentally Derived Guidelines for the Creation of Earcons. In the *Proceedings of the Human Computer Interaction Conference 95*.
3. Brooks, M. (2006). *Proceedings of the 3rd International ISCRAM Conference (B. Van de Walle and M. Turoff, eds.)*, Newark, NJ (USA), May 2006
4. DAISY (2002). ANSI-NISO Z39.86-2002. Digital Talking Book Standard. *US Library of Congress and The DAISY Consortium*. Retrieved from the internet: http://www.daisy.org/z3986
5. Edworthy, J., Heiller, E., Walters, K., Clift-Matthews, W., and Crowther, M. (2003). Acoustic, Semantic and Phonetic Influences in Spoken Warning Signal Words. *Applied Cognitive Psychology*, 17: 915-933.
6. Gong, L. and Lai, J. (2003). To Mix or Not to Mix Synthetic Speech and Human Speech? Contrasting Impact on Judge-Rated Task Performance versus Self-rated Performance and Attitudinal Responses. *International Journal of Speech Technology*, 6(2), 123-131.
7. Gregg, C.E., Houghton, B.F., Paton, D., Johnston, D.M., Swanson, D.A., and Yanagi, B.S. (2007). Tsunami Warnings: Understanding in Hawai'i. *Natural Hazards*, 40(1), 71-87.
8. 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.
9. 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, March 1998.
10. Lachman, R., Tatsuoka, M., and Bonk, W.J. (1961). "Human behavior during the tsunami of May, 1960." Science, 1961, 133, 1405-1409.
11. OASIS (2004). Common Alerting Protocol v1.0. Oasis Standard 200402. Retrieved from the Web: http://www.oasis-open.org/committees/download.php/6334/oasis-200402-cap-core-1.0.pdf
12. 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
13. Proulx, G., Laroche, C., Jaspers-Fayer, F. & Lavallée, R. *Fire Alarm Signal Recognition. Canadian Institute for Research in Construction, Internal Report No. 828*, June, 2001.
14. 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.
15. 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.
16. Stanton, N.A. (1994) Human Factors of Alarm Design. Taylor and Francis Publishers: London.
17. Stanton, N.A. & Edworthy, J. (1999) Human Factors in Auditory Warnings. Ashgate Publishers: Aldershot.
18. 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, April, 2006.

19. Ulfvengren, P. (2003) Design of Natural Warning Sounds in Human-Machine Systems. Doctoral Thesis, KTH Royal Institute of Technology, Stockholm.
20. W3C (2005). Synchronized Multimedia Integration Language (SMIL) 2.1. World Wide Web Consortium (W3C). Retrieved from the internet: http://www.w3.org/TR/2005/REC-SMIL2-20051213/
21. W3C (2007). Web Accessibility Initiative (WAI). World Wide Web Consortium (W3C). Retrieved from the internet: http://www.w3.org/WAI.
22. Williges, B.H., J.M. Schurick, T.M. Spine & M.T. Hakkinen (1986) Using speech in the human-computer interface, in Ehrich, R.W. & R.C. Williges (eds), *Advances in Human Factors/Ergonomics: 2 Human-Computer Dialogue Design*, 1986, Elsevier Science Publishers B.V.