NASA/TM–20210017507

Advanced Multimodal Solutions for Information Presentation

Elizabeth M. Wenzel
NASA Ames Research Center

Martine Godfroy-Cooper
San Jose State University Foundation

April 2021

NASA STI Program…in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question via to help@sti.nasa.gov

- Phone the NASA STI Help Desk at (757) 864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM—20210017507

Advanced Multimodal Solutions for Information Presentation

Elizabeth M. Wenzel
NASA Ames Research Center

Martine Godfroy-Cooper
San Jose State University Foundation

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California

April 2021

Available from:

NASA STI Program
STI Support Services
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

Table of Contents

Acronyms and Definitions

2D	2 dimensional
3D	3 dimensional
AA	adaptive automation
AT	audiotactile
cm	centimeter
CNS	central nervous system
DoD	Department of Defense
DoF	degree of freedom
DRM	Design Reference Missions
DVE	Degraded Visual Environment
EVA	extravehicular activity
g	gravitational
HCI	Human-Computer Interaction
HDD	head-down display
HRP	Human Research Program
HUD	head-up display
Hz	Hertz
IOR	inhibition of return
ISO	International Standards Organization
ISS	International Space Station
JNDs	Just Noticeable Differences
LADAR	Laser Detection and Ranging
LDEMs	long duration exploration missions
MRT	Multiple Resource Theory
ms	millisecond
NASA	National Aviation and Space Administration
RADAR	Radio Detection and Ranging
RT	response time
SA	situation awareness
SOA	stimulus onset asynchrony
TBW	temporal binding window
TOJ	temporal order judgment
VA	visual-auditory
VRI	visual reorientation illusion
VT	visuotactile

Advanced Multimodal Solutions for Information Presentation

Elizabeth M. Wenzel[1] and Martine Godfroy-Cooper[2]

This document is based on a 2017 Final Report commissioned by the Human Research Program, Space Human Factors and Habitability Element, Risk of Inadequate Human-Computer Interaction (HCI).

1. HRP Risks and Gaps Addressed

The work reviewed here addresses several Gaps relevant to the Risk of Inadequate Human-Computer Interaction (HCI) including Gap HCI-03, "We need HCI guidelines (e.g., display configuration, screen-navigation) to mitigate the performance decrements and operational conditions of long duration spaceflight," and feeds into the subsequent task, "Long-duration Information Systems/Multimodal Information Validation." Under Gap HCI-06: We need guidelines to ensure crewmembers receive all of the information required to accomplish necessary tasks in a timely fashion, even when operating autonomously. It also is very relevant for Gap HCI-04, "We need to understand how emerging multi-modal and adaptive display and control technologies are best applied to the design of HCI for proposed long-duration DRM (Design Reference Missions) operations." A specific goal of the report is to help formulate the topics to be covered by a new research solicitation that includes multimodal information displays.

The report will help to close the HCI-03 and the HCI-04 gaps by assessing what is currently known about multimodal, adaptive information systems, including best practices for combining modalities for the most effective information displays. The review will provide an analysis of potential multimodal display technologies for long duration missions and, in particular, will focus on their potential role in EVA activities. Potential issues in developing interface guidelines for long duration exploration missions (LDEMs) are also briefly considered.

2. Statement of the Problem and Scope of the Report

High-workload, fast-paced, and degraded sensory environments are the likeliest candidates to benefit from multimodal information presentation. For example, during extra-vehicular activity (EVA) and telerobotic operations, the sensory restrictions associated with such a hostile environment provide a major challenge to maintaining the situation awareness (SA) required for safe operations. In particular, orientation, navigation, and collision avoidance are critical aspects of EVA tasks that need to be addressed to ensure the safety of the crew and the success of the mission. Multimodal displays hold promise to enhance situation awareness and task performance by utilizing different sensory modalities and maximizing their effectiveness based on appropriate interaction between modalities.

---

[1] NASA Ames Research Center; Moffett Field, California.
[2] San Jose State University Foundation; San Jose, California.

During EVA, the visual and auditory channels are likely to be the most utilized with tasks such as monitoring the visual environment, attending visual and auditory displays, and maintaining multichannel auditory communications. Previous studies have shown that compared to unimodal displays (spatial auditory or 2D visual), bimodal presentation of information can improve operator performance during simulated extravehicular activity on planetary surfaces for tasks as diverse as orientation, localization or docking, particularly when the visual environment is degraded or workload is increased (Wenzel, et al., 2012; 2014). Tactile displays offer a third sensory channel that may both offload information processing effort and provide a means to capture attention when urgently required. For example, recent studies suggest that including tactile cues may result in increased orientation and alerting accuracy, improved task response time and decreased workload (Jones & Sarter, 2008), as well as provide self-orientation cues in microgravity on the International Space Station (ISS) (e.g., van Erp & van Veen, 2006).

An important overall issue is that context-dependent factors like task complexity, sensory degradation, peripersonal vs. extrapersonal space operations, workload, experience level, and operator fatigue tend to vary greatly in complex real-world environments and it will be difficult to design a multimodal interface that performs well under all conditions. As a possible solution, adaptive systems have been proposed in which the information presented to the user changes as a function of task/context-dependent factors. However, this presupposes that adequate methods for detecting and/or predicting such factors are developed. Further, research in adaptive systems for aviation suggests that they can sometimes serve to increase workload and reduce situational awareness (Kaber, Perry, Segall, McClernon & Prinzel, 2006). It will be critical to develop multimodal display guidelines that include consideration of smart systems that can select the best display method for a particular context/situation.

Multimodal displays will also play an important role for long-duration information systems and will likely begin to be developed in the early phases of cislunar Gateway, and later lunar or Mars transit missions. Information systems are envisioned for LDEMs that require spacecraft with greater crew autonomy and increased dependence on computer-provided information needed to perform routine tasks, as well as time- and safety critical tasks. Such a system will require a single, common interface that is easy to learn and use and accesses key information from all relevant vehicle/habitat systems to enable task performance in both nominal and emergency conditions. Understanding of multimodal display technologies and their interactions will help to inform interface guidelines for LDEMs.

The scope of the current report is an analysis of potential multimodal display technologies for long duration missions and, in particular, will focus on their potential role in EVA activities. The review will address multimodal (combined visual, auditory and/or tactile) displays investigated by the National Aeronautics and Space Administration (NASA), industry, and Department of Defense (DoD). It also considers the need for adaptive information systems to accommodate a variety of operational contexts such as crew status (e.g., fatigue, workload level) and task environment (e.g., EVA, habitat, rover, spacecraft). Current approaches to guidelines and best practices for combining modalities for the most effective information displays are also reviewed. Potential issues in developing interface guidelines for LDEMs are briefly considered.

Olfaction is not addressed here since practical olfactory cueing systems have yet to be developed and research indicates that smell is a very limited sensory channel for information presentation. Consideration of control technologies in multimodal interfaces is also outside the scope of this report.

3. Rationale for Developing Multimodal Displays

A primary theoretical framework underlying the development of multimodal displays is Wicken's (1980; 2002) Multiple Resource Theory (MRT) which accounts for improved performance when information associated with concurrent tasks differ in their information processing stage (e.g., perceptual/cognitive vs. response), processing code (e.g., spatial vs. nonspatial information), and/or processing modality (visual, auditory, tactile). MRT asserts that sensory channels are associated with separate attentional resources, thus competition for limited processing resources will be reduced when multimodal information displays are utilized. However, association with separate attentional resources does not mean sensory channels are completely independent of each other. They may interact with each other in a variety of ways as illustrated by the work on crossmodal information processing and crossmodal attention (Spence & Driver, 2004; Ferris & Sarter, 2008). Later refinements of MRT have acknowledged more nuanced roles for multisensory channels. For example, performance improvements when offloading the visual channel may be due to reduced visual scanning requirements at the peripheral level rather than central processing mechanisms (Wickens & Liu, 1988). The phenomenon of preemption may also account for some of the effects of distributing information across sensory channels (Wickens, et al, 2003). For example, an auditory message can involuntarily divert attention away from a concurrent visual task because auditory messages are highly effective at capturing attention; they also require an immediate shift in attention because they cannot be held for long in working memory. Research generally shows that while factors like preemption may play a role, the more powerful factor in producing performance benefits is multiple attentional resources.

Antecedants of advanced multimodal displays include the development of virtual environments that require multimodal inputs to enhance the sense of immersion in a virtual world. The need for enhanced technology interfaces for a wide variety of users, including the disabled, has also been a driving factor in their development. Perhaps of most critical importance is the need for advanced displays in complex, real-world environments such aerospace, medicine and process control where data overload and attention management are primary contributors to decreased performance in human-machine systems. Such factors can be mitigated by off loading overused channels such as vision and distributing information across sensory channels in ways appropriate to particular tasks.

4. Benefits of Multimodal Displays

Multimodal displays offer a number of benefits including increased bandwidth, redundancy, disambiguation, modality appropriateness, complementarity, and substitution (Sarter, 2013; Oviatt, 2012). *Increased bandwidth*, or an increase in the amount of information that can be transmitted over a fixed time period, is the most frequently cited advantage of multimodal presentation. Another advantage, particularly cited for critical alarm systems, is *redundancy* or the presentation of the same information in more than one sensory channel, usually visual combined with auditory displays. The notion is that if visual attention is directed elsewhere, a corresponding auditory alarm will ensure that the alarm does not go undetected. *Disambiguation* occurs when information from two or more sensory channels is combined when information in one channel would be ambiguous. For example, in a military context a speech based command "approach ammunition depot alpha" may be ambiguous if more than one depot is in the area. A visual targeting marker on a corresponding shared visual map can disambiguate the correct target depot from the others. A related concept is *modality appropriateness* or the differential use of sensory channels based on whether a given channel is best suited for presenting a particular kind of information. For example, the auditory channel is often considered best for presenting short commands or alarms that need to be responded to quickly, while the visual channel is appropriate for conveying more complex information that needs to be available for longer durations (Stanney, et al., 2004). *Complementarity* refers to the presentation of related information that should be merged to form a combined percept. It is related to the "binding problem" (Roskies, 1999) or how we combine information from multiple sources to form a unified percept of an object in space or an event in time. In addition to the combination of different types of information within a sensory channel (e.g., auditory streaming effects leading to auditory object formation and scene analysis, Bregman, 1990), the simultaneous presentation of information across sensory channels can provide a richer, more accurate representation of a particular object or event (e.g., audible speech combined with a synced visual video of the moving mouth of the speaker). Finally, the notion of *substitution* is that when one sensory channel becomes temporarily or permanently unavailable or degraded, the same information can be carried by another sensory channel. For example, during EVA, astronauts typically have a restricted field of view combined with an impoverished visual environment. Auditory or tactile displays can provide supplementary information to enable navigation, exploration and task execution for repairs and construction.

5. Structural and Functional Differences Between Modalities

Over the last decades, multisensory research has demonstrated that although important differences are present between information processing in the different sensory modalities, certain stages of information processing (such as those contributing to maintaining and updating the representation of space, or those believed to direct attention toward particular aspects of external stimuli) may actually be shared between modalities (Downar, Crawley, Mikulis & Davis, 2000; Spence & Gallace, 2007).

The inherent structural and functional differences between vision, audition and touch have important implications for multisensory integration. Like vision and audition, the tactile sense provides information from stimuli external to the body, such as "*what*" and "*where*." In the "what" domain, the tactile sense carries both static and dynamic information. Static stimuli characteristics accessible to the tactile sensory system, such as shape, size and texture can also be made available to the visual system. Dynamic information, such as frequency, intensity and temporal variation can find equivalents in the visual, auditory and tactile systems. The registration of the spatial properties of a stimulus ("where") between modalities is more challenging for the central nervous system (CNS) and yet, a necessary step towards multisensory integration. Auditory, visual and somatosensory signals are represented in different neural encoding formats at the level of the cochlea, the retina, and the body, respectively. Whereas vision is tuned to spatial processing supported by a 2D retinotopic (eye-centered) spatial organization, audition is primarily tuned to frequency analysis resulting in a tonotopic map, i.e., an orderly map of frequencies along the length of the cochlea (Culler et al., 1943). As a consequence, the auditory system must derive the location of a sound on the basis of acoustic cues that arise from the geometry of the head and the ears (binaural and monaural cues, Yost, 2000). Tactile stimuli undergo a distributed coding of attributes such as pressure, vibration and temperature. Information about stimuli location is first represented relative to the skin surface in the primary sensory cortex's homunculus (Penfield & Boldrey, 1937). However, because the body is flexible, the location of touch in space crucially depends on the body's posture at the time of touch. Therefore, the external touch location must be derived by integration of skin location and postural information, i.e., proprioceptive inputs (Heed & Azañón, 2014), a process referred to as tactile remapping (Driver & Spence, 1998).

To solve the problem of the differences in reference frames[3] between modalities, the CNS must proceed to the transformation of the information in a common, absolute spatial representation of the world, and later, read the information back out into the motor coordinates needed for each effector system ("*how*"). This process, encompassing the integrated neural representation of a "body schema" and of the space around the body (peripersonal space), is a prerequisite for the effective "piloting" of the body to avoid or manipulate objects in pursuit of behavioral goals (Popper & Eccles, 1977).

6. Multisensory Integration: Evidence from Cross-Modal Behavioral and Neuroscience Research

Both behavioral and neurophysiological studies suggest that there are extensive crossmodal links between vision, audition, and touch. In particular, the phenomena of modality expectations, modality shifting and cross-modal spatial and temporal links have been demonstrated in many psychological and neurophysiological laboratory studies of multimodal information processing (see Spence & Driver, 2004).

6.1. Attentional Effects

Modality expectations represent a top-down influence on attention allocation across sensory channels. Such expectations are formed by the observed frequency or the perceived importance of a cue in a particular modality. Behavioral, neuroimaging, and neurophysiological studies using various modality-cuing paradigms suggest that expecting a cue to appear in a particular modality leads to an enhanced readiness to detect (faster reaction times) and discriminate information (increased accuracy) in that sensory channel (e.g., Spence and Driver, 1997). Conversely, response times increase to cues presented in an unexpected modality and with modality uncertainty. The effects of modality uncertainty are less pronounced for tactile than for visual and auditory cues (Boulter, 1977).

The *modality shifting effect* is a related phenomenon in which significant performance decrements occur when attention is shifted from a highly expected modality to an infrequently used sensory channel (Spence and Driver, 1997). Thus, in general, response times will be slower to events or targets in less frequent modalities. This performance degradation is exacerbated when participants shift attention to the visual or auditory channel away from rare events presented in the tactile modality (Spence et al., 2000).

Cross-modal links in attention have been repeatedly shown to impact performance in multimodal tasks (see Spence & Driver, 2004). For example, one effect of cross-modal spatial links is an increased readiness to perceive a signal in one modality when a preceding cue in a different sensory channel originates from the same or similar location (bottom-up, exogenous attention). Such effects have been observed for all pairings of vision, audition and touch and are greatest when the bimodal stimuli occur at identical locations. Conversely, performance degrades when the two stimuli are at (substantially) different locations.

Deliberate shifts of attention (top-down, endogenous attention) to a particular location in one modality (e.g., visual or auditory) have been shown to lead to concurrent shifts in the other modality.

---

[3] The term "reference frame" is used to refer to the center of a coordinate system for representing objects, including the body itself, and relations between objects (Cohen & Andersen, 2002).

Behavioral and physiological research has shown that these crossmodal spatial links can lead to enhanced facilitation of responses when simultaneous stimuli are at the same location or to response suppression for stimuli at different locations (e.g., see Driver and Spence, 2000; Spence & Driver, 2004). Such response changes may be unintentionally induced in multimodal interface design if the location and timing of information presentation is not controlled carefully. Surprisingly, the potentially detrimental effects of crossmodal links are not considered in existing guidelines for multimodal displays.

Such cross-modal temporal and spatial links, in particular, determine the likelihood and the strength of multisensory integration.

6.2. Temporal Synchrony

The “temporal rule” states that multisensory interactions are dependent on the coincidence of the neural responses to different stimuli (albeit within a certain window). Stimuli with overlapping neural responses yield interactions, whereas those yielding asynchronous responses do not. The concept of a temporal binding window (TBW) was introduced that determines the maximum temporal separation between two sensory events to be perceived as referring to the same object. For visual and auditory stimuli such as simple flashes and beeps, the temporal window extended from 25 to 50ms and was found to be asymmetrical, i.e., the tolerance to stimulus onset asynchrony (SOA) is greater when the visual stimulus precedes the auditory stimulus, than in the opposite configuration (Keetels & Vroomen, 2005; Zampini, Guest, et al., 2005a). For audio-tactile pairs, Zampini, Brown, et al. (2005b) obtained JNDs of about 80ms, and for visual-tactile pairs, JNDs have been found on the order of 35–65ms (Keetels & Vroomen, 2008b; Spence et al., 2001). The size of the audiovisual TBW is up to 5 times larger (200ms) with audio-visual speech stimuli (van Wassenhove, Grant, & Poeppel, 2007). In other words, when the stimulus complexity increases, the sensitivity for temporal order deteriorates (Vatakis & Spence, 2006). Paired comparisons between visual, auditory and tactile stimuli with a discrimination threshold method revealed that the temporal resolution for synchrony perception was similar for visual-auditory (VA) and visuotactile (VT), while audiotactile (AT) showed superiority in terms of temporal resolution (Fujisaki & Nishida, 2009).

For VA stimuli that are repetitively presented in streams, the perception of synchrony breaks down if the temporal frequency is above ~4 Hz (Benjamin, van der Smagt, & Verstraten, 2008; Fujisaki & Nishida, 2005). Above this rate, observers are no longer able to discriminate whether the auditory and visual stimuli are synchronous and the two modality streams are perceived as being segregated with no order between them (VT and AT temporal resolutions were ~4 Hz and ~10 Hz, respectively).

To summarize, temporal lags below 20ms are usually unnoticed, probably because of hard-wired limitations on the resolution power of the individual senses. Above this limit, delays do become noticeable, in particular if stimuli: (1) have fast transient rise times; (2) are spatially separated; (3) have predictable onsets; and (4) are presented rhythmically at rates < 4Hz.

Bresciani, Dammeier & Ernst (2008) investigated the interactions between visual, tactile and auditory sensory signals for the perception of sequences of events. Sequences of flashes, taps and beeps were presented simultaneously. For each experimental session, the participants were instructed to count the number of events presented in one modality (Target) and ignore the stimuli presented in the other modalities (Background). The results showed that for the perception of sequences of events: (1) vision, touch and audition are automatically integrated; (2) the respective contribution of the three modalities to the integrated percept differs; (3) the relative contribution of each modality depends on its relative reliability (1/variability); and (4) task-irrelevant stimuli (potential distractors) have more weight when presented in two rather than one modality.

Thorough VT, AT and VAT synchronization may be an issue when time delays are introduced in the context of teleoperations. Time delays present in the control loop of human teleoperation in space can be considered another aspect of increased task workload and can have a critical impact on human performance and mission effectiveness (Sheridan, 1993; Fong, et al., 2010; Podnar, et al., 2010; Schmidt, et al., 2012). Relatively low latencies on the order of 400ms will be present during lunar telerobotic control from Lagrange points or from Mars orbit during future surface exploration and control missions (Lester & Thronson, 2011; Valinia, et al., 2012). Lester and Thronson (2011) have described the limits of a "cognitive horizon" for teleoperation in space, i.e., a latency limit of ~500ms beyond which performance degrades. However, the sensory modality and the type of task determine the size of specific latency thresholds: 100ms for haptic surface perception, 200ms for visual feedback applications (Lester & Thronson, 2011). For visual just-noticeable differences, Ellis et al., (1999a, b) reported thresholds of 15–20ms, while for points of subject equality, the threshold was 50ms. Thresholds for the perception of visual image instability, on the other hand, have been reported as 180–320msec depending on head velocity (Allison, et al., 2001; see also Adelstein, et al., 2003). Relatively little research exists for audition: in one study, Wenzel (2001) showed no performance disruption in an auditory localization task (head motion enabled) until latency increased to 250–500ms (for 3s-duration stimuli) while disruption was minimal for longer duration stimuli (8sec duration stimuli, 500 ms latency). In a 2015 study comparing the single and combined effects of a visual auditory docking task, Wenzel & Godfroy-Cooper reported that the auditory system is less sensitive to latency compared to the visual system. A similar effect was found for docking response time in the auditory condition. However, there was no facilitatory effect in the bimodal condition.

6.3. Spatial Alignment

The "spatial rule" states that multisensory interactions are dependent on the spatial alignment and/or overlap of receptive fields responsive to stimuli. That is, facilitative multisensory interactions can be observed even when stimuli are spatially misaligned in their external coordinates, provided that the responsive neurons contain overlapping representations. If these representations do not overlap, no interaction is seen, and in many cases, even response suppression is observed (i.e., inhibitory interactions). However, a perfect spatial and temporal alignment is not required for multisensory integration to occur as long as the sound and light are presented within close spatial (Stein & Meredith, 1993) and temporal proximity (Colonius & Diederich, 2012; Vroomen & Keetels, 2010). Some authors have suggested that stimuli must be presented within a ~30–40° of each other (though not necessarily within the same hemispace) for effects to occur either in the case of facilitating stimulus detection (for visual-auditory interactions: Hughes et al., 1994; Harrington & Peck, 1998; for visual-tactile interactions: Foster et al., 2002) or in the case of influencing localization judgments (for auditory-tactile interactions: Calclin et al., 2002), with wider separations failing to generate facilitating interaction effects (Stein et al., 1989). Others have found interaction effects across wider spatial disparities for tasks requiring visual intensity judgments, sometimes independent of the locus of sounds (Stein et al., 1996; Hairston et al., 2003). Others emphasize the spatial ambiguity of a stimulus over its absolute position, such as in the cases of the ventriloquism effect and shifts in attention (Bertelson, 1999; Bertelson & De Gelder, 2004; Spence & Driver, 2000; Hairston et al., 2003). Such a crossmodal shift of exogenous (or automatic/reflexive) spatial attention can occur between pairs of auditory, visual, and tactile stimuli (Spence & McDonald, 2004). The effects of a shift of crossmodal exogenous spatial attention are generally most pronounced when there is some time between the onset of the sound and the light (~100–300ms, see Berger, Henrik & Rafal, 2005) and when they originate from approximately the same spatial location (Spence & McDonald, 2004). Crossmodal exogenous spatial attention shifts result in faster response times (RTs) and higher detection sensitivity for information appearing at attended as compared to unattended locations (Santangelo & Spence, 2009). When there is more time between the onset of the sound and the light (>300ms) an inhibitory after-effect can be observed (inhibition of return [IOR]) that has been demonstrated between all possible pairings of audition, vision and touch (Klein, 2000).

The variability of these findings raises the possibility that task requirements may influence spatial limitations on multisensory interactions (Spence, 2013). Another possibility is that some varieties of multisensory interactions occur with neurons and/or brain regions with large spatially insensitive receptive fields.

6.4. Inverse Effectiveness

A third principle, called the "principle of inverse effectiveness" was formulated based on the observation that the relative increase in spike rate in multisensory neurons after multisensory stimulation was larger when the unimodal stimuli (sound or light alone) evoked only a weak response as compared to those unimodal stimuli that evoked a strong response in the neuron (Meredith & Stein, 1983; Holmes, 2009). This multisensory facilitation can be reinforced by a semantic congruence between the sensory inputs, and is susceptible to being modulated by attentional factors, instructions, or inter-individual differences. Overall, these neurophysiological principles have been useful in predicting the circumstances in which multisensory integration occurs in human behavior (see Stein & Stanford, 2008 for a review).

7. The Forgotten Dimension: Depth

Of all our senses, only vision and audition allow us to perceive information that is currently out of reach. Whereas vision enables us to determine the spatial location of something that can be seen in frontal space only, audition helps with localizing sounds in all directions. Conversely, the sense of touch is relatively constrained in terms of depth perception as one can only feel things that are within reach. Perceiving touch is thus bound to the body. One aspect that has received little to no attention in the literature on multisensory integration is how the distance between multisensory stimuli and an observer affects multisensory integration.

7.1. Peripersonal vs. Extrapersonal Space

A few studies have provided neurophysiological and neuropsychological evidence indicating that information presented at different distances from the observer, or in different regions of space, is processed differently in the brain (Previc, 1998; Graziano, Reiss, & Gross, 1999). Objects at a distance can naturally be perceived through a limited number of senses, namely vision, audition and olfaction. By contrast, objects nearer to, and therefore within reach or in contact with the body can impact upon all sensory systems, including gustation and all the sub-modalities of touch (see Craig & Rollman, 1999 for a review). There seems to be a clear distinction between the region of space directly surrounding different body parts (peripersonal space, near) and the space that is out of reach (extrapersonal space, far) in terms of multisensory interactions occurring in these regions of space (Occeli et al., 2011). Representations of peripersonal space are body-centered or body part-centered (see Làvadas, 2002 for a review), restricted to the space immediately surrounding the body (~70 cm in humans), and involve the integration of information from multiple sensory modalities (somatosensory, proprioceptive, visual and auditory. The brain's representations of the visuotactile peripersonal space can be modulated to incorporate mirror images, inanimate objects, and tools held in the hand (Austen, Soto-Faraco, & Kingstone, 2001; Pavani, Spence, & Driver, 2000). Tool use seems to “capture” extrapersonal space and results in it being incorporated into peripersonal space. The question remains whether this applies to tools that are teleoperated, such as a robotic arm or rovers on a planetary surface.

7.2. Visuotactile and Audiotactile Interactions in Frontal Peripersonal Space

The majority of studies of multisensory interactions in the depth plane have focused on visuotactile and audiotactile interactions in frontal peripersonal space (Zampini, et al., 2007). In such studies, tactile stimuli are often delivered to the hands while the other parts of the skin surface are largely ignored. Neurophysiological evidence appears to support the notion that space is divided into several distinct regions and suggests that both audiotactile and visuotactile interactions are more pronounced in the space directly surrounding the body. In a 2012 study, Canzoneri et al. demonstrated that participants responded more rapidly to the tactile targets when the simulated sound location was situated close to their hand as compared to when it appeared further away. Similar results have been reported with vision and touch (Gray & Tan, 2002).

The specific spatio-temporal properties of dynamic stimuli can also influence the strength of multisensory interactions. When visual stimuli approach the body at a certain speed, predictions concerning the location of a tactile stimulus in 3D space can be made (e.g., when and where contact with the body is expected), thus enhancing the speed with which stimuli at the expected time and location are processed relative to other times and locations. Similar results have been observed in a study of visuotactile interactions, where responses to tactile targets were fastest to tactile stimuli presented at the same time at which contact with the face was expected, based on the speed of a reaching movement (Kandula et al., 2015).

Given that tactile perception is inherently bound to the body, the stronger multisensory interactions involving touch in peripersonal space require stimuli from different modalities to not only be aligned at any particular depth, but also specifically to be aligned with the body (or extensions of the body in the case of tool-use).

Canzoneri et al.’s results (2012) suggest that the presentation of sensory information from slightly beyond the border of peripersonal space does not result in the sudden absence of audiotactile interactions, but rather shows a more gradual decline in performance, as expressed by a gradual increase in RTs. This effect appeared to be the most pronounced for sounds that are perceived to be approaching the participants and less for sounds that appear to recede from the participant. Drawing a distinction between the space that is reachable and that which is not would appear to be useful in terms of the possible (motor) interactions with the environment and the perception of stimuli that are somehow relevant in terms of their proximity to the body of the observer. The existence of two different spatial representations (dichotomous vs. continuous) has been supported by the results of several studies (Caggiano et al., 2009; Makin et al., 2008). A study of sound localization in depth (Canzoneri et al., 2012) indicates that participants perceived the location of simulated approaching or receding unimodal stimuli in terms of a spatial continuum in depth. During audiotactile stimulation, however, RTs showed a somewhat steeper decrease at a certain distance from the observer, indicative of there being some kind of boundary in peripersonal space. Such a border was not observed when audiovisual stimuli were used (Teneggi et al., 2013), underlining the idea that multisensory interactions that involve tactile stimulation may display a dichotomous border for peripersonal space, perhaps due to the asymmetric nature of tactile perception.

Unlike tactile perception, auditory and visual perception are not constrained by the distance at which stimuli are presented. Consequently, audiovisual interactions do not show the same asymmetric effects as audiotactile and visuotactile interactions. Thus, multisensory interactions involving touch may be especially pronounced close to the body, given that spatial alignment in depth will always require that all unimodal component stimuli be presented in peripersonal space in alignment with the body.

7.3. Rear Peripersonal and Extrapersonal Space

The area of space behind the observer, which cannot be seen directly, provides an interesting opportunity to investigate the interactions taking place between auditory and somatosensory information. The distinction between peripersonal and extrapersonal space can also be made when it comes to the space behind the observer. The results of neurophysiological (Graziano & al., 1999) and neuropsychological studies (Farnè & Làvadas, 2002) have suggested that audiotactile spatial interactions may be more prevalent in near rear space than in frontal space. Kitagawa et al. (2005) have similarly shown that audiotactile interactions were greater when the auditory stimuli were presented close to (i.e., 20 cm) rather than far from (i.e., 70 cm) the back of the participant's head. Farnè and Làvadas (2002) have argued that this is because people will typically perceive an object's approach from behind by means of auditory cues (see also Kitagawa & Spence, 2006). It will be particularly important for future research to:

- assess whether spatial influences on audiotactile interactions are dependent on the particular body surface stimulated
- investigate the impact of spatial distance between auditory and somatosensory stimuli, particularly within rear space
- determine to what extent auditory-somatosensory interactions are coded in terms of the egocentric and/or external spatial position of the stimulated body surface

7.4. Body Posture and Sensory Deprivation

Tactile location has been demonstrated to be represented both in skin-based anatomical and external reference frames in a variety of tasks (Heed & Azañón, 2014; Overvliet, Azañón & Soto-Faraco, 2011; Harrar & Harris, 2010). Spatial transformations in touch are impaired by limb-crossing but are consistent with tactile location being recoded rapidly and efficiently, followed by integration of skin-based and external reference information to specify the reach target (Brandes & Heed, 2015). A 2016 study (Noel et al.) investigating the relative contribution of the different sensory systems under different body postures in a tactile temporal order judgment (TOJ), showed that mechanisms governing the alignment between somatotopic and external reference frames extend beyond those imposed by body posture to include spatial features conveyed by the auditory and visual modalities, with a heavier weighting of auditory than visual spatial information. More specifically, auditory and auditory-visual deprivation exacerbated the difference in tactile temporal acuity between uncrossed to crossed leg postures, an effect not seen for visual-only deprivation. Furthermore, the effects under combined visual-auditory deprivation were greater than those seen for auditory deprivation. This result is of particular interest in the context of space operations where sensory cues are altered or missing.

8. Applied Studies of Multimodal Displays

Although direct behavioral inferences from neuronal data can be problematic, general insights into multisensory integration provided by neurophysiological studies, such as the spatial and temporal rules, have generally been confirmed in applied behavioral studies. However, it is also clear that behavioral integration is complex and depends on factors such as the type and reliability of each unisensory stimulus involved, the nature of the task being performed, and the level of operator workload (Gray, Spence, Ho & Tan, 2013). Such factors may determine whether the combined performance effects of multisensory integration are additive, superadditive, or no better than unisensory performance. If stimuli are chosen that violate human perceptual and attentional systems (e.g., intersensory conflict due to incongruent spatial locations, temporal asynchrony, or even semantic incongruency; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004), performance may actually be subadditive (worse than unisensory performance) by setting up conditions that confuse or overload the user and result in multisensory suppression.

Although many multimodal display studies have been conducted in recent years, they have often been developed in a somewhat trial and error manner without consideration of the basic mechanisms of human multisensory integration and cross-modal attention (Sarter, 2006). Further, performance measures are not always directly compared among the possible unimodal, bimodal, and/or trimodal displays in a given study, making clear inferences about relative multisensory benefits problematic. Here we present representative studies of multisensory integration performance in bimodal and trimodal displays, particularly for attentional orienting and warnings, that have been investigated for applications in automobiles, military combat vehicles, aviation, and medical equipment.

8.1. Automobile Displays

Multimodal automobile interfaces warning drivers of events such as the source and spatial location of potential collisions have been under development for a number of years. Research in this area has demonstrated inconsistent results regarding whether multimodal displays produce better performance compared to unimodal displays. For example, in a driving simulator, Lee, McGehee, Brown and Marshall (2006) investigated the effectiveness of an interface designed to reengage attention during adaptive cruise control when a driver may need to assume control in a potential emergency braking situation. They compared performance for four unimodal displays: a visual icon depicting a collision, an auditory warning tone presented via a dashboard speaker, two different tactile signals (seat vibration or brake pulse), and a trimodal combination of the unimodal displays. Their results showed no significant differences in braking RTs between the multimodal and unimodal signals. Spence and Ho (2008) later noted that there was a large spatial separation between the different signals with visual and auditory signals at approximately eye level and the tactile displays at seat and foot levels in the Lee et al. (2006) study. Ho, Reed and Spence (2007) compared audio, tactile and bimodal audio-tactile signals in a collision avoidance display with frontal locations that were more spatially congruent, i.e., a car horn presented via a speaker on the dashboard and a tactile signal presented in the middle of the driver's stomach. The results showed that braking RTs were significantly shorter for the bimodal signals compared to the best unimodal signal. Similar multimodal benefits have been demonstrated in other driver warning systems when auditory and visual stimuli are spatially congruent (Spence & Ho, 2008; van Erp & van Veen, 2004). Such data support the idea that multimodal effects may vary considerably depending upon factors like the presence or absence of spatial congruency.

8.2. Military and Aviation Applications

Multimodal interfaces have also been developed for military ground combat vehicles and aircraft for applications such as threat alerting. Oskarsson, Eriksson and Carlander (2012) compared the effectiveness of unimodal, bimodal, and trimodal threat warning displays in a simulated combat vehicle where the task was to turn toward a threat as fast and accurately as possible. The display components consisted of virtual 3D sounds presented via headphones, directional tactile signals delivered via one of 12 tactors located on the operator's belt, and either a visual head-down display (HDD) or head-up display (HUD). The allocentric HDD visual display used moving lines to indicate the direction of threat and was placed at waist height approximately 30º left of the center. The exocentric HUD used a pulsing arrow that indicated both threat direction and angular distance and was superimposed at eye level on the out-the-window screen mounted in front of the operator. Performance was assessed in terms of accuracy in localizing the threat and mean RT to orient to it. Experiment 1, which utilized the visual HDD, showed that the trimodal display produced better performance compared to the best of the unimodal and bimodal displays. Interestingly, multisensory facilitation occurred with this interface even though there was some physical spatial incongruence between the signals, i.e., the auditory and tactile signals were presented from around the body while the visual signal was always presented in front of the operator. It may be that subadditive integration did not occur because the different signals were associated with different types of attentional orienting. The tactile and auditory cues were exogenous (an automatic/reflexive capture of attention by peripheral cues) while visual cues were endogenous (voluntary shifting of spatial attention). In experiment 2, Oskarsson et al. (2012) utilized the HUD so that all three cues were spatially congruent. Three displays for cueing threat direction were compared: the HUD with 3D audio, tactile with 3D audio, and HUD, 3D audio, and tactile belt combined into a trimodal display. Performance was significantly better for the trimodal HUD display compared to the bimodal displays. The trimodal HUD was also better than the trimodal display with the HDD. Similar benefits for multimodal directional alerting systems have also been observed in the context of military aviation under conditions such as high acceleration in simulated aerial combat (van Erp, Eriksson, Levin, Carlander, Veltman, & Vos, 2007) and in the presence of helicopter noise (Brill, J. C., Lawson, B. D., & Rupert, 2015; McGrath, et al., 2004).

In another aviation context, Ngo, Pierce & Spence (2012) compared unimodal and multimodal performance in a simulated air traffic control scenario in which participants had to monitor and control aircraft. The task goals were to ensure that the aircraft landed or exited at the correct altitude, speed, and direction and maintained a safe separation from all other aircraft and boundaries. The performance measures recorded included enroute time, handoff delay, and conflict resolution delay. The standard baseline visual condition (the aircraft in conflict was highlighted in red) was compared to experimental conditions in which the visual cue was accompanied by a temporally sychronous auditory cue (500 Hz tone), a vibrotactile cue (tactors at each side of the waist), or a combined audiotactile cue. The results showed that performance was significantly better when auditory or audiotactile warning signals were presented. However, performance with the unimodal tactile warning signal did not improve performance compared to the unimodal visual alert.

8.3. Medical Applications

A relatively new area for multimodal display research is in medical applications. Ferris and Sarter (2011) recently investigated multimodal interfaces for attentional orienting and task management in the context of a surgical procedure to enhance patient monitoring by anesthesiologists. They compared performance with the commonly used auditory and visual warning systems to conditions in which tactile displays were added. In a surgery simulation, visual (popup messages on a patient screen in front of the anesthesiologist) and auditory signals (alarm sounds and an auditory sonification, i.e., a periodic signal that conveyed heart rate and blood oxygenation information) were combined with three types of vibrotactile signals delivered via a vibrotactile vest. In addition to a tactile alarm, two tactile displays were presented that provided continuous information concerning the patient's status. Several performance measures were collected for two tasks, physiological monitoring and anesthesia induction, and combined in a multitask performance score. Unlike the simple alerting function of many of the previously described multimodal interfaces, this system was designed to both inform the operator about the nature of a critical event as well as orient attention to a particular spatial location. For example, lung volume information was presented via signals to the anesthesiologist's back while blood pressure information was presented via the arm. For all tactile displays, the combined trimodal interface produced superior performance compared to the typical auditory-visual interface alone. However, it is not clear that this enhancement was due to multisensory facilitation since performance was not measured for the tactile system alone.

8.4. Effects of Workload and Experience Level

A number of studies have indicated that although multisensory cues may not be more effective than unimodal cues under conditions of low perceptual workload (Lavie, 2005), they may still retain their capacity to capture an operator's spatial attention under high workload when unimodal cues are no longer effective (Spence, 2010). Similarly, the multimodal interfaces described here may or may not function as well in real operating environments where workload can be substantially higher than can be produced in a laboratory or simulator. Few multimodal interface studies involving tactile cues have systematically varied operator workload. In one such study, Mohebbi, Gray & Tan (2009) reported that the effectiveness of unimodal auditory and tactile collision warnings varied substantially with workload in a driving simulator. Auditory warnings that were highly effective under low workload (just driving) were totally ineffective under moderate workload conditions (simple mobile phone conversation while driving). Tactile signals, on the other hand, were effective under both low and moderate workload conditions but significantly less effective in a high workload condition (complex phone conversation).

An important issue related to workload is whether the semantic content and delivery method of the warning should change as a function of operator workload. In the anesthesiology study by Ferris and Sarter (2011), multimodal displays that provided continuous tactile information concerning the patient's health resulted in better performance than warning signals that were only active when the situation became critical, presumably because the continuous signals allowed the operator to anticipate critical situations and respond more rapidly. However, these results were only observed under low operator workload. Under high workload, the continuous signals tended to be ignored, resulting in better performance with discrete warnings. To our knowledge, applied studies directly assessing whether an operator can utilize semantic information from auditory or haptic icons under high workload have yet to be conducted.

Another topic that has not been well studied is how the level of experience or training of the operator may affect performance with multimodal displays. Typically, participants in lab and simulator studies are given limited training prior to collection of data. However, training to asymptotic performance would likely be too cumbersome to be practical in conducting experiments. Further, if a given display requires much training it is less likely that it will be adopted as a good solution. In addition, unlike real world scenarios with unpredictable events, lab studies must necessarily test multimodal displays under simplified and controlled conditions in order to obtain results that can be statistically analyzed. It is unclear how well people can respond to many different warning signals that may not occur with any degree of regularity.

It is also unclear whether highly experienced operators who have developed complex mental representations of their operating environment would benefit from many of the relatively simple multimodal displays like those developed so far. Similar to pilots subjected to multiple alarms during an emergency situation with poorly designed cockpit displays, experienced users may simply find them annoying and turn them off. One experimental approach would be to directly compare performance for novice and expert participants, such as pilots, astronauts or doctors, although this is not always a practical possibility for many researchers.

8.5. Relevance to Space Applications

While many of the results discussed above will be relevant to space applications, conditions specific to space environments such as degraded sensory cues and the impact of micro- and macro-gravity conditions will need to be carefully considered when designing displays for spacecraft systems, EVA missions, and astronaut support. Compared to the earth environment, such factors may alter the relative reliability of individual sensory systems, the manner in which they integrate, and the way in which factors like workload impact performance. Although some conclusions may be drawn from simulation studies or buoyancy tank and centrifuge studies mimicking altered gravity conditions, investigation in a true space environment will remain the best test of display effectiveness.

9. Current Standards and Guidelines

Sarter (2013) discusses five questions that need to be addressed when designing multimodal information displays for particular tasks or types of information:

1. Is a multimodal display even desirable or required?
2. Which modalities should be selected for the design?
3. How should modalities be mapped to tasks and types of information?
4. How should modalities be combined, synchronized and integrated?
5. When and how should multimodal information presentation adapt to changing task contexts and circumstances?

Surprisingly little consistent guidance that directly addresses multimodal displays is currently available in the form of standards and guidelines to aid in making these decisions. Such standards and guidelines vary considerably in terms of their specific focus and level of abstraction. For example, some are essentially unimodal, addressing the properties and preferable uses of the individual sensory channels (MIL-STD-1472G, Stanney, et al., 2004, pp. 234–243). They focus on the appropriateness and most effective implementation of a particular modality for a given type of information, task, interaction, and/or environment. They may also focus on a very specific type of display for a very specific task such as multimodal warning signals for driver-vehicle interfaces (Campbell, et al., 2016). Other more general guidelines are concerned with the effective combination and integration of sensory channels (Stanney, et al., 2004, p. 249, Giang, et al., 2010, p. 74). These typically are based primarily on research using bimodal information presentation and few, if any, directly address trimodal (or beyond) information integration. Finally, some guidelines are high-level design principles that can apply independent of modality, such as complementarity, consistency, and redundancy of information presentation either within or across sensory channels (Reeves, et al., 2004; Sutcliffe, 2003).

9.1. The Selection of Unimodal vs. Multimodal Displays

Before investing in the development of a multimodal display system, a display designer must decide whether multiple modalities are warranted or if a unimodal display is sufficient or even preferable for a given task environment. As noted earlier, there are a number of rationales proposed for including multiple modalities in the design of an interface. These include the attention and information processing advantages evidenced by research on multiple resource theory. Multimodal displays can also support the functional benefits of increased bandwidth, redundancy, disambiguation, modality appropriateness, complementarity, and substitution. Some authors also suggest that multimodal information presentation should be supported to account for differences in user preferences, needs, and abilities (e.g., ETSI, 2003; ISO 14195-3, 2002; Sutcliffe, 2003).

However, utilizing a multimodal rather than a unimodal display can involve information processing trade-offs. In complex task/data environments such as space exploration missions, the use of multiple sensory channels may be necessary to support time and resource sharing and to compensate for the sensory degradations present in the space environment. In less demanding environments where tasks are simpler and self-paced, it may be better to limit the interface design to one or two modalities in order to avoid the potential costs of a multimodal design that may include increased interface management and monitoring demands on the user.

Given that a multimodal design is chosen, the appropriate and required sensory channels must be specified. Such decisions are determined in part by environmental factors. For example, high levels of noise may limit the use of the auditory channel while degraded visual conditions may require the use of auditory or tactile displays. The selection of modalities and the choice of media also must consider the types of tasks and information that the user needs to handle (Stanney, et al., 2004; ISO 14195-3, 2002; Oviatt, 2012).

9.2. Assigning Sensory Modalities to Tasks and Types of Information

A critical step in designing multimodal interfaces is to decide which of the modalities should be assigned to the kinds of information to be presented and the tasks to be accomplished in a given task environment. Stanney, et al. (2004) provides detailed summaries and tables of guidelines for each of the sensory modalities of vision, audition, and haptics (tactile, kinesthetic) and groups their perceptual properties according to functional categories such as alerts and warnings, localization and orienting, and representation of complex spatial or conceptual relationships. For example, the detailed sensitivity of foveal vision can accommodate stimuli of long duration. It allows for delayed and prolonged attending and is appropriate for presenting complex graphics or other large amounts of detailed information which may need to be referred to at a later point in time. Peripheral vision, on the other hand, is effective at detecting transient information such as motion, luminance changes, and the appearance of new objects, but it does not support the recognition of objects or details. Consequently, it can perform an alerting or orienting function in a display design to attract and direct user's attention to changing information at a particular spatial location within the field of view.

The auditory channel differs from vision along several important dimensions. It is omnidirectional, allowing information to be detected from any location in space, independent of head and body orientation, and in parallel with information presented through other sensory channels. Auditory information presentation is also transient, although this is somewhat compensated by a longer short-term memory storage compared to vision (Wickens, 1992). Since the auditory channel is always "on," an auditory display can function as an omnidirectional alerting system whose purpose is "to point the eyes" or to direct attention to events not in the frontal visual field.

Considerable research has been devoted to using speech output in auditory displays and a number of guidelines have been developed for this medium alone. Michaelis and Wiggins (1982) proposed that speech output should be used for simple short messages that will not be referred to later. They also recommended speech generation when messages deal with events in time and require an immediate response, or when the visual channel is overloaded. As in space missions, speech displays may be particularly appropriate if the task environment has bright or poor illumination, involves severe vibrations and high g-forces, or if it requires that the user can move around freely (e.g., on the ISS or proposed habitat modules for deep space gateway or transport missions). Speech interfaces may also help support task sharing. Cohen and Oviatt (1994) suggested that spoken interaction with machines is preferable when limited keyboard and/or screen space is available, when the user is disabled and the use of other modalities is difficult or impossible, when natural language interaction is preferred, and when the user's hands or eyes are busy. However, cognitive limitations can significantly limit such task sharing. Research has shown that displays such as a speech-based in-vehicle email system (Lee et al., 2001) or simultaneous cellphone conversations (Strayer et al., 2003) demand attention and impair driver performance due to resource competition.

Another form of auditory information presentation that has received significant attention is non-speech audio. Simple non-speech signals can be useful when the higher level cognitive processing required for speech interpretation is not possible or desirable, e.g., alerts or warnings in emergency situations. Other non-speech audio used to represent a specific event, object, function, or action. Earcons are structured, often musically-based, short sounds (e.g., the sound your computer makes when booting up). Auditory icons are brief sounds that takes advantage of the user's prior knowledge and natural associations between sounds and their consequences (e.g., the sound of something thrown into a trash can when deleting a computer file). A more complex, dynamic form of non-speech audio is sonification, i.e., ‘‘the transformation of data relations into perceived relations in an acoustic non-speech signal for the purposes of facilitating communication or interpretation’’ (Kramer, 1993). A recent example of the use of sonification combined with spatialized sound was developed for the United States Army program on Degraded Visual Environment (DVE) mitigation. The study evaluated the use of 3D auditory cueing for obstacle avoidance and detecting drift trajectories in helicopter brownout conditions, and as a replacement for or compliment to LADAR/RADAR imagery (Godfroy, Miller, Szoboszlay, & Wenzel (2017).

In contrast to vision and audition, touch is a generally underutilized channel for information presentation (see Wenzel & Godfroy-Cooper, 2016). Touch refers to the sensations that result from mechanical, thermal, chemical, or electrical stimulation of the skin. Cues presented via the tactile channel are transient in nature. Like vision and hearing, touch allows for the concurrent presentation and extraction of several stimulus dimensions, e.g., frequency and amplitude in the case of vibrotactile cues. However, touch is unique among all the senses in that it is the only modality that is capable of both simultaneously sensing and acting upon the environment.

In the normal use of touch, known as active touch, we manipulate the objects we are sensing. Active touch is assumed to be predominant in naturalistic environments where an observer is moving around and thus actively seeks information about objects in the world. Support for active touch is essential in highly interactive displays such as virtual environments. Passive touch, when information is presented to the skin in the absence of observer movements, is atypical of normal tactile perception and leads the person to focus on sensations on the body surface (Gibson, 1962). Passive touch cues have generally been utilized in tactile displays. A variety of tactile displays have been developed to aid spatial orientation and navigation in situations in which the human operator can become disoriented. Circumstances leading to disorientation may include an absence of stable reference frames, such as when flying through clouds or flying under high G-load conditions

(Rupert, et al., 1996; Rupert, 2000; van Veen & van Erp, 2003), working in microgravity environments in space (Rochlis & Newman, 2000; Traylor & Tan, 2002; van Erp & van Veen, 2003; 2006; van Erp, van Veen, & Ruijsendaal, 2008), or navigation in unfamiliar terrain (Gilson, Redden, & Elliott, 2007; Jones et al., 2006; Lindeman, Sibert, Lathan, & Vice, 2004). In such displays, vibrotactile actuators are used to present information about the intended direction of an operator or vehicle, the pitch and roll of an aircraft, and/or the location of way points in the environment.

As noted earlier, olfaction is minimally addressed here since practical olfactory cueing systems have yet to be developed and research regarding odor identification, training, vocabulary, and odor intensity indicates that smell is a limited sensory channel that may be used to determine the presence of an odor but should not be counted on to identify a specific stimulus (e.g., see Mulgund, Stokes, Turieo & Devine, 2002; Sarter, 2013). Sarter (2013) summarizes guidelines proposed by Kaye (2001) for olfactory displays that reflect the limitations of this sensory channel: (1) olfactory displays should use smell changes rather than intensities since humans can readily distinguish qualitative odor characteristics but not quantitative aspects such as intensity; (2) olfactory interfaces are best used to represent slowly changing data or medium duration events since odors linger and may merge when they overlap; (3) olfactory displays need to ensure that nausea or allergic reactions are not induced unless the intention is to warn of a dangerous situation (e.g., the unpleasant smell of a gas leak); and (4) olfactory displays are best suited to representing ambient information about the background or status of a system or environment (again, the sulfurous odor representing a gas leak, or a pleasant odor indicating all is well).

9.3. The Integration of Modalities

As discussed in the previous section on Multisensory Integration, the spatial and temporal combination and synchronization of sensory channels is a critical determinant of an integrated perceptual experience and must be considered when assigning various modalities to tasks and types of information (Sarter, 2006; 2013). To date, few guidelines provide specific and/or consistent agreement on how to approach the issue of integration. Some guidelines suggest that the combination of media should be minimized and used only as necessary to convey information (ISO, 2002). The basis for this recommendation is unclear and leaves the interface designer without much help on deciding when or whether to use a multimodal display at all. Other multimedia design principles under ISO (2002) reflect general principles of good interface design, calling for modality combinations that support thematic congruence, manageable information loading, complementary viewpoints, consistency, and redundancy (see also Sutcliffe, 2003).

If a multimodal design is chosen, other guidelines propose that modality combinations should be based on user preferences, needs, and abilities (e.g., ETSI, 2003; ISO, 2002, Sutcliffe, 2003). This recommendation is problematic because it shifts the responsibility from the designer to the user who is then allowed or required to manage the interface. Users typically don't have the knowledge or time to determine and implement appropriate and effective modality combinations in advance, or on the fly during ongoing operations. Basing multimodal design on user preferences is especially untenable in collaborative work environments where modality choices and combinations tend to be based on group dynamics (rather than individual preferences). Such interfaces also evolve over time as team members adopt strategies that are introduced successfully by one person and tend to be chosen for the purposes of disambiguation, common ground, and error recovery (Ho & Sarter, 2004). As Andre and Wickens (1995) point out, users often ''want what's not best for them.'' These considerations lead to the conclusion that, at minimum, the multimodal interface designer should limit the modality combinations available to the user based on what is known about crossmodal constraints on attention and perception (Spence & Driver, 2004).

Reeves et al. (2004) propose that modalities should be integrated in a manner compatible not only with user preferences but also with context and system functionality (see also Sutcliffe, 2003). One guideline for effective modality combinations they propose is to "Ensure system output modalities are well synchronized temporally (for example, map-based display and spoken directions, or virtual display and non-speech audio)." Many current systems often use a modality sequence of an auditory alert followed by visual presentation of relevant information. For example, route guidance systems for cars use an auditory signal to notify the driver of an upcoming turn, and a visual display then provides more detailed information about the turn (Mollenhauer et al., 1997). An example of multimodal guidelines focused on a specific type of display for a specific application context are those delineated by Campbell, et al. (2016, p. 4–4) for the design of multimodal warning messages for driver-vehicle interfaces.

9.4. Adaptive Multimodal Information Displays

There is widespread agreement in the literature that fixed assignments of modalities to tasks or types of information are not desirable or even possible. Instead, multimodal interfaces need to be flexible and take into consideration possible changes in the needs, abilities and experience level of the user, the types of tasks being performed, the task environment, and the level of workload (Reeves et al., 2004; Jameson & Gajos, 2012). For example, since workload tends to vary greatly in complex real-world environments, it will be difficult to design a multimodal interface that performs well under all conditions. As a possible solution, Ferris and Sarter (2011) proposed adaptive systems in which the information presented to the user changes as a function of workload. However, this presupposes that adequate methods of detecting and/or predicting operator workload are developed. While interface adaptivity may reduce the need for interface management by the user, it also has the potential to lead to confusion when interface settings are automatically adjusted, especially when such changes occur in the context of high workload and fast-paced operations (Woods, 1993). Research in adaptive systems for aviation suggests that they can serve to increase workload and reduce situational awareness (Kaber, Perry, Segall, McClernon & Prinzel, 2006). As a result, a user may decide to turn a well-intentioned multimodal interface into a more tractable unimodal one or even turn the adaptive function off if allowed. The much-maligned adaptive menus of Microsoft Office 2003 are a well-known example of how implementation of such a display can go wrong.

Whether the flexibility of a multimodal interface should take the form of system-controlled adaptivity and/or user-controlled adaptability remains a matter of considerable debate. Reeves et al. (2004) suggest that a user profile could be captured and determine interface settings. Buisine & Martin (2003) propose to adapt multimodal system settings to observed user preferences for a given task and leave the modality choice to the user only when no preference evolves. These rather simple approaches to adaptive displays based on user preference are not likely to be sufficient in the context of complex task environments such as LDEMs crew members will require at least some degree of automated support.

One could think of adaptive displays as one aspect of the development of autonomous information systems. In a review of augmented cognition for HCI, Hale, Stanney & Schmorrow (2012) describe adaptive displays as "adaption of the presentation" including modality augmentation and modification of information type (e.g., verbal vs. spatial). Kaber (2013) also addresses adaptive automation (AA) and discusses the parameters of *what*, *how*, *when*, and *who* that form the basis of experimental or actual adaptive system design. *What* is automated can have a significant effect on human performance; AA can reduce workload and improve task performance when applied to information acquisition and task execution but less so for decision making. *How* refers to how task responsibility is divided between the human and the system. There is a paucity of research in this area and designers must take care to consider the capabilities of both the human and the system to ensure that operators are given a set of coherent and integrated tasks to perform. The question of *when* has to do with what automation triggering mechanisms will invoke dynamic function allocation (as opposed to static automation or manual control). These mechanisms may range from physiological measures, to behavioral/performance measures, or a hybrid combination of both. Research is still needed to assess their reliability and sensitivity to changing task demands, especially in a broad range of operational settings. The design parameter *who* refers to whether the system (adaptive automation) or the user (adaptable automation) has control over dynamic function allocation. Research suggests that adaptive automation is often preferable since human control adds cognitive overhead beyond the primary tasks they are performing. However, it is critical to human-computer interaction and performance that "automation etiquette" be observed, e.g., warning operators of mode transitions and minimizing interruptions to the operators' primary tasks. Similar design considerations are addressed in the guidelines for adaptive displays delineated by Giang, et al. (2010).

9.5. Supplementary Material

In order to provide a more concrete sense of the range and nature of currently available standards and guidelines for multimodal display design, an appendix has been included with excerpts from selected documents and articles.

MIL-STD-1472G (2012) is an updated version of a widely used US military standard (DoD Design Criteria Standard: Human Engineering) that provides a few standards for combining modalities in information displays. Typically, these have to do with combining visual and auditory information and emphasize the primacy of visual information. Except for 3-D audio displays, newer technologies such as tactile displays are not addressed.

ISO 14915-3 (2002) is an international standard that provides a rather comprehensive set of multimedia design principles that calls for media combinations that support thematic congruence, manageable information loading, complementary viewpoints, consistency, and redundancy (see also Sutcliffe, 2003). However, in considering integration of multimodal information, it tends to focus on the impact of *media* selection per se rather than modality selection; for example, combining still pictures with moving images, or nonrealistic audio (e.g., tone sequences) with speech. Annex C of the Standard briefly addresses cross-modality media.

Stanney, et al. (2004) is an academic review article that provides useful tables of design guidelines for the individual visual, auditory and haptic modalities. It also proposes design guidelines for multimodal integration based on consideration of evidence from the behavioral and, to some extent, the neurophysiological literature. Similarly, Reeves, et al. (2004) is a brief, though often cited as comprehensive, review article that discusses six categories of guidelines and represents a preliminary effort to establish general principles for multimodal interaction design.

Giang, et al. (2010) is a lengthy, comprehensive document published by Defence Research and Development Canada that addresses a variety of topics related to the design of multimodal interfaces. In chapter 5, Crossmodal Attention, the authors list a number of proposed design guidelines based on evidence from crossmodal attention research (p. 74). Giang, et al. (2010) also propose design considerations for Intelligent Adaptive Interfaces that include multimodal displays, although they indicate that this research area is not well-developed and concrete guidelines are yet to be established (p. 77–79).

10. Conclusions and Research Needs

10.1. Multimodal Standards and Guidelines

The literature on multimodal displays and their underlying basis in human crossmodal perception and neurophysiology is large and varied. It ranges from controlled lab studies, often using simplified stimuli and with relatively small effect sizes, to applied studies in a wide variety of task domains with varying degrees of operational realism. Frequently, only two modalities are studied at a time. Thus, it is not surprising that there are not a lot of well-established and consistent standards and guidelines.

Woods et al. (1994) suggest that guidelines should be ''a synthesis or abstraction of current knowledge and a stimulus to the growth of knowledge.'' However, some current guidelines for multimodal interface design seem to be written without considering the existing knowledge base on human crossmodal perception and neurophysiology and its performance implications. In fact, they may propose to leave modality choices up to the user who is unlikely to have sufficient knowledge about multimodal information processing, and under critical task conditions, unable to devote the time and effort to make such choices without compromising performance on primary tasks. Many guidelines also do not include justifications to help a designer understand why a recommendation is made and under what circumstances it will be appropriate to apply it. Also, they may not be specific to the design of multimodal interfaces, instead reiterating general design guidelines such as the need for consistency or redundancy. Other guidelines focus on the choice of individual sensory channels for given tasks and contexts, or they describe high-level design objectives, such as the need to support synchronization of sensory channels or adaptation of displays based on user/task demands. Examples of effective multimodal information presentation may be provided but generally very little concrete direction is given on how to move from the stated guidelines to actual implementation of a multimodal display design based on these rules.

The shortcomings of current multimodal guidelines are due, in part, to the fact that a considerable number of research questions remain to be addressed. These include investigating performance with truly multimodal, rather than various bimodal combinations, the effects of stress and workload on multimodal information processing, and the impact of training and experience on display effectiveness. Overall, individual multimodal system designs will need to be validated to make sure they are performing as intended using relevant task-specific conditions. Their success or failure will likely be driven primarily by individual task requirements.

Similar issues are true for adaptive displays except that perhaps even less is known about important issues and general guidelines are largely unavailable. Critical issues include what triggering measures should be used to change a display, how reliable they are, what criterion or threshold points are viable for each measure, and how they depend on the task constraints. Similar to automation systems in general, other issues include whether the system or the user has responsibility for deciding what and when adaptation occurs and how to provide the user with adequate situational awareness about display changes.

10.2. Multimodal Displays in Space Environments

NASA has identified a number of potentially significant biomedical risks that may limit deep space missions, including missions to the Moon and Mars. Among them, and relevant for the present report, are changes in eye-head-hand control, gaze function, postural and/or locomotor ability and perception (Bloomberg et al. 2015; Reschke, et al, 2016; Sipes, et al. 2016; Nicogossian, et al., 2016). Evidence from space flight research has demonstrated that the function of each of these subsystems is altered by removing gravity, a fundamental orientation reference for vestibular, proprioceptive and haptic receptors used in the control of orientation, posture, navigation and coordination of movements. The decrement in sensation, and their consequences for perception and action, affects all the senses, although hearing function under microgravity has received little attention (Persterer, et al., 1993).

Space activities are also highly constrained by the affordances of the environment. For example, in flight, the crew may experience a visual reorientation illusion (VRI) induced by the architectural symmetries of the cabin interior that typically defines multiple "visual vertical" directions, usually separated by 90º. When VRI occurs, crews lose their sense of direction with respect to the entire vehicle and reach or look in the wrong direction for remembered objects. Distortions of the visual space have also been reported and may influence astronauts' ability to accurately perform cognitive and sensorimotor tasks, such as those involved in robotic operations. In space, the environment is not structured with a gravitational reference and a visual horizon so perspective is irrelevant and astronauts perceive heights and depths of objects as taller and shallower on orbit, respectively (Clement et al., 2013). On surface, low visibility conditions and non-earth gravitational forces will likely impact inertial navigation. EVAs in space are uniquely defined by the loss of information normally provided by the auditory and somatosensory systems, and a restricted field of view. In this hostile environment, astronauts may not be able to keep a six degree of freedom (DoF) SA, in particular regarding potentially life-threatening objects approaching at high speed (extrapersonal space). Another critical aspect of EVAs in space is the absence of tactile feedback associated with operations, whether directly guided by hand or mediated by tool use (peripersonal space).

Therefore, different types of risks will be associated with different types of activities: rendezvous/docking and remote manipulator system operations, piloted landing, vehicle egress and extravehicular activities, and rover operations. Although the decrement in sensations under microgravity, and their consequences for perception and action, is relatively well documented at the level of each sensory system individually, less is known about interactions between the senses. Furthermore, the role of tactile displays for space activities has been primarily investigated as a substitutive device for orientation and navigation within the spaceship.

The potential roles for each modality (e.g., redundancy, complementarity or substitution) need to be addressed in context (task and environment specific) and in interaction with the other senses. The first important aspect here is that there is a significant decrement in sensation associated with space operations. As a consequence, the normal contribution (weight) of each sensory modality to multimodal perception experienced on Earth will not be relevant in space, since the reliability of the different senses will change. In other words, the usual dominance of the visual cues in multisensory perception may be challenged, and the role of auditory and tactile cues may increase. It is, thus, fundamental to determine the role of each sensory modality, alone and in combination, for very specific tasks where multisensory information presentation would lead to performance and safety improvement. Last, the combined presentation of "real" and "virtual" information may also affect the way sensory modalities are integrated, and this aspect should be investigated.

Multimodal displays will also play an important role for information systems developed for long duration missions. Information systems developed for LDEMs will require that spacecraft provide greater crew autonomy and support increased dependence on flexible computer-provided information needed to perform routine tasks, as well as time- and safety critical tasks. Such a system will require a single, common interface that is easy to learn and use and accesses key information from all relevant vehicle/habitat systems to enable task performance in both nominal and emergency conditions. Understanding of multimodal display technologies and their interactions will help to inform interface guidelines for LDEMs.

11. References

Adelstein, B. D., Lee, T. G. & Ellis, S. R. (2003) Head Tracking Latency in Virtual Environments: Psychophysics and a Model. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting,* pp. 2083–2087.

Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U., & Zacher, J. E. (2001) Tolerance of temporal delay in virtual environments. *Proceedings of Virtual Reality*, pp. 247–253.

Andre, A.D. & Wickens, C.D. (1995). When users want what's not best for them. *Ergonomics in Design*, pp. 10–14.

Austen, E. L., Soto-Faraco, S., Pinel, J. P. J., & Kingstone, A. F. (2001). Virtual body effect: factors influencing visual-tactile integration. *Psychonomic Society*, *6*(2).

Barber, D., Howard, T. M., & Walter, M. R. (2016, May). A multimodal interface for real-time soldier-robot teaming. In *Proceedings of the SPIE* (Vol. 9837).

Benjamin, J. S., van der Smagt, M. J., & Verstraten, F. A. (2008). Matching auditory and visual signals: Is sensory modality just another feature? *Perception, 37,* 848–858.

Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General*, *134*(2), 207.

Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. *Advances in psychology*, *129*, 347–362.

Bertelson, P. & De Gelder, B. (2004). The Psychology of Multimodal Perception. In Spence, C., & Driver, J. (Eds.). *Crossmodal Space and Crossmodal Attention* (pp. 141–177). Oxford University Press.

Bloomberg, J. J., Reschke, M. F., Clement, G. R., Mulavara, A. P., & Taylor, L. C. (2015). Risk of Impaired Control of Spacecraft/Associated Systems and Decreased Mobility Due to Vestibular/Sensorimotor Alterations Associated with Space flight. NASA HRP HHC Report.

Brandes, J., & Heed, T. (2015). Reach Trajectories Characterize Tactile Localization for Sensorimotor Decision Making. *The Journal of Neuroscience*, *35*(40), 13648–13658.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* The MIT Press, Cambridge, MA.

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain research bulletin*, *75*(6), 753–760.

Brill, J. C., Lawson, B. D., & Rupert, A. H. (2015). Audiotactile aids for improving pilot situation awareness. In *Proceedings of the International Symposium for Aviation Psychology*, May 4-7, Dayton, OH, 13–18.

Buisine, S. & Martin, J.-C. (2003). Design principles for cooperation between modalities in bi-directional multimodal interfaces. In: *Proceedings of the CHI 2003 workshop on ''Principles for multimodal user interface design,''* Ft. Lauderdale, FL.

Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., & Casile, A. (2009). Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *Science*, *324*(5925), 403–406.

Campbell, J. L., Brown. J. L., Graving, J. S., Richard, C. M., Lichty, M. G., Sanquist, T., Bacon L. P. Woods, R., Li, H., Williams, D. N. & Morgan, J. L. (2016, December). *Human factors design guidance for driver-vehicle interfaces* (Report No. DOT HS 812 360). Washington, DC: National Highway Traffic Safety Administration.

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PloS one*, *7*(9), e44306.

Clément, G., Skinner, A., & Lathan, C. (2013). Distance and size perception in astronauts during long-duration spaceflight. *Life*, *3*(4), 524–537.

Cohen, P.R. & Oviatt, S.L. (1994). The role of voice in human-machine communication. In: Roe, D., Wilpon, J. (Eds.), *Voice Communication Between Humans and Machines*. National Academy of Sciences Press, Washington, DC, pp. 34–75.

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews: Neuroscience*, *3*(7), 553–562.

Colonius, H., & Diederich, A. (2012). Focused attention vs. crossmodal signals paradigm: deriving predictions from the time-window-of-integration model. *Frontiers in integrative neuroscience*, *6*(62).

Craig, J. C., & Rollman, G. B. (1999). Somesthesis. *Annual Review of Psychology*, *50*(1), 305–331.

Culler, E., Coakley, J. D., Lowy, K., & Gross, N. (1943). A revised frequency-map of the guinea-pig cochlea. *The American Journal of Psychology*, *56*(4), 475–500.

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature neuroscience*, *3*(3), 277–283.

Driver, J. & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in cognitive sciences*, *2*(7), 254–262.

Driver, J., & Spence, C. (2004). The Psychology of Multimodal Perception. In Spence, C. & Driver, J. (Eds.). *Crossmodal Space and Crossmodal Attention* (pp. 179–220). Oxford University Press.

Ellis, S.R., Young, M.J., Ehrlich, S.M., & Adelstein, B.D. (1999a) Discrimination of changes of rendering latency during voluntary hand movement. *Proceedings, 43rd Annual Meeting HFES,* pp. 1182–1186.

Ellis, S.R., Young, M.J., Adelstein, B.D., & Ehrlich, S.M. (1999b) Discrimination of changes in latency during head movement. *Proceedings, Computer Human Interfaces,* pp. 1129–1133.

ETSI (European Telecommunications Standards Institute) (2002). *Human factors (HF); Guidelines on the multimodality of icons, symbols and pictograms. (ETSI EG 202 048)*. Sophia Antipolis, France: Author.

ETSI (European Telecommunications Standards Institute) (2003). *Multimodal interaction, communication and navigation guidelines.* (Report ETSI EG 202 191 V1.1.1 (2003–08). ETSI, Sophia Antipolis, France.

Farnè, A., & Làdavas, E. (2002). Auditory peripersonal space in humans. *Journal of Cognitive Neuroscience*, *14*(7), 1030–1043.

Ferris, T. K., & Sarter, N. B. (2008). Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*, *50*(1), 17–26.

Ferris, T. K., & Sarter, N. (2011). Continuously informing vibrotactile displays in support of attention management and multitasking in anesthesiology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(6), 600–611.

Fong, T., Abercromby, A., Bualat, M. G., Deans, M. C., Hodges, K. V., Hurtado Jr., J., Landis, R., Lee, P., & Schreckenghost, D. (2010) Assessment of robotic recon for human exploration of the Moon. *Acta Astronautica,* Vol. 67, pp. 9–10.

Foster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*, **143**, 480–487.

Fujisaki, W., & Nishida, S. Y. (2005). Temporal frequency characteristics of synchrony–asynchrony discrimination of audio-visual signals. *Experimental Brain Research*, *166*(3-4), 455–464.

Fujisaki, W., & Nishida, S. Y. (2009). Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception. *Experimental brain research*, *198*(2-3), 245–259.

Giang, W., Santhakumaran, S., Masnavi, E., Glussich, D., Kline, J., Chui, F., Burns, C., Histon, J. & Zelek, J. (2010). *Multimodal Interfaces: Literature Review of Ecological Interface Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces.* Defence R&D Canada Contract Report DRDC Toronto CR 2010-051

Gibson, J.J. (1962). Observations on active touch. *Psychological Review*. 69, 477–490.

Gilson, R. D., Redden, E. S., & Elliott, L. R. (2007). *Remote tactile displays for future soldiers*. (Tech. Rep. ARL-SR-0152.) Aberdeen Proving Ground, MD: Army Research Laboratory.

Godfroy, M., Miller, J. D., Szoboszlay, Z. & Wenzel, E. M. (2017). 3D-Sonification for Obstacle Avoidance in Brownout Conditions. American Helicopter Society 73rd Annual Forum & Technology Display, May 9–11, 2017, Fort Worth, TX.

Gray, R., & Tan, H. Z. (2002). Dynamic and predictive links between touch and vision. *Experimental Brain Research*, *145*(1), 50–55.

Gray, R., Spence, C., Ho, C. & Tan, H. Z. (2013). Efficient multimodal cuing of spatial attention. *Proceedings of the IEEE,* 101(9), 2113–2122.

Graziano, M. S., Reiss, L. A., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature*, *397*(6718), 428–430.

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of cognitive neuroscience*, *15*(1), 20–29.

Hale, K. S., Stanney, K. M. & Schmorrow, D. D. (2012). Augmenting Cognition in HCI: Twenty-First Century Adaptive System Science and Technology. In J. A. Jacko (Ed.). *The Human-Computer Interaction Handbook: Fundamemtals, Evolving Technologies, and Emerging Applications* (pp. 1343–1358). CRC Press.

Harrar, V., & Harris, L. R. (2010). Touch used to guide action is partially coded in a visual reference frame. *Experimental brain research*, *203*(3), 615–620.

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research*, *122*(2), 247–252.

Heed, T., & Azañón, E. (2014). Using time to investigate space: a review of tactile temporal order judgments as a window onto spatial processing in touch. *Front. Psychol*, *5*, 76.

Ho, C.-Y. & Sarter, N.B. (2004). Supporting synchronous distributed communication and coordination through multimodal information exchange. In: *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, New Orleans, LS.

Ho, C., Reed, N. & Spence, C. (2007). 'Multisensory in-car warning signals for collision avoidance. *Human Factors*, 49, 1107–1114.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain topography*, *21*(3-4), 168–176.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 131.

Jameson, A. & Gajos, K. Z. (2012). Systems That Adapt to Their Users. In J. A. Jacko (Ed.). *The Human-Computer Interaction Handbook: Fundamemtals, Evolving Technologies, and Emerging Applications* (pp. 431–456). CRC Press.

Jones, L. A., Lockyer, B., & Piateski, E. (2006). Tactile display and vibrotactile recognition on the torso. *Advanced Robotics*, 20, 1359–1374.

Jones, L. A. & Sarter, N. B. (2008). Tactile Displays: Guidance for Their Design and Application, *Human Factors*, 50(1), 90–110.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, *36*(5), 447–462.

Kaber, D. B. (2013). Adaptive Automation. In J. D. Lee, A. Kirlik & M. J. Dainoff (Eds.). *The Oxford Handbook of Cognitive Engineering* (pp. 594-609). Oxford University Press.

Kaye, J. J. (2004). Making Scents: aromatic output for HCI. *Interactions*, *11*(1), 48–61.

Kandula, M., Hofman, D., & Dijkerman, H. C. (2015). Visuo-tactile interactions are dependent on the predictive value of the visual stimulus.*Neuropsychologia*, *70*, 358–366.

Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research*, *167*(4), 635–640.

Keetels, M., & Vroomen, J. (2008). Tactile—visual temporal ventriloquism: No effect of spatial disparity. *Perception & Psychophysics*, *70*(5), 765–771.

Kitagawa, N., Zampini, M., & Spence, C. (2005). Audiotactile interactions in near and far space. *Experimental Brain Research*, *166*(3-4), 528–537.

Kitagawa, N. & Spence, C. (2006). Audiotactile multisensory interactions in human information processing. *The Japanese Psychological Association*. *48*(3), pp. 158–173

Klein, R. M. (2000). Inhibition of return. *Trends in cognitive sciences*, *4*(4), 138-147.

Kramer, G. (1993). *Auditory display: Sonification, audification, and auditory interfaces.* Perseus Publishing.

Làvadas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends in cognitive sciences*, *6*(1), 17–22.

Laurienti, P. J., Kraft, R. A. Maldjian, J. A., Burdette, J. H. & M. T. Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research,* 158, 405–414.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75–82.

Lee, J. D., Caven, B., Haake, S., Brown, T.L. (2001). Speech-based interaction with in-vehicle computers: the effects of speech-based e-mail on drivers' attention to the road. Human Factors 43 (4), 631–640.

Lee, J. D., McGehee, D. V., Brown, T. L. & Marshall, D. (2006). Effects of adaptive cruise control and alert modality on driver performance. Transp. Res. Record, 1980, 49–56.

Lester, D. & Thronson, H. (2011) Low-Latency Lunar Surface Telerobotics from Earth-Moon Libration Points. *Proceedings of the AIAA SPACE 2011 Conference & Exposition,* Long Beach, California, September 27–29.

Lindeman, R. W., Sibert, J. L., Lathan, C. E., & Vice, J. M. (2004). The design and deployment of a wearable vibrotactile feedback system. In *Proceedings of the Eighth International Symposium on Wearable Computers* (pp. 56–59). Los Alamitos, CA: IEEE Computer Society.

MacGrath, B. J., Estrada, A., Braithwaite, M. G., Raj, A. K., & Rupert, A. H. (2004). Tactile situation awareness system flight demonstration final report (USAARL Rep. 2004-10). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Makin, T. R., Holmes, N. P., & Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behavioural brain research*, *191*(1), 1–10.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391.

Michaelis, P.R. & Wiggins, R.H. (1982). A human factors engineer's introduction to speech synthesizers. In: Badre, A., Shneiderman, B. (Eds.), *Directions in Human–Computer Interaction.* Ablex, Norwood, NJ, pp. 149–178.

MIL-STD-1472G, 11 January 2012, Department Of Defense Design Criteria Standard: Human Engineering (Superseding MIL-STD-1472F, 23 August 1999).

Mohebbi, R., Gray, R., & Tan, H. Z. (2009). Driver reaction time to tactile and auditory rear-end collision warnings while talking on a cell phone. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(1), 102-110.

Mollenhauer, M.A., Hulse, M.C., Dingus, T.A., Jahns, S.K., Carney, C. (1997). Design decision aids and human factors guidelines for ATIS displays. In Noy, Y.I. (Ed.), *Ergonomics and Safety of Intelligent Driver Interfaces*. LEA, Mahwah, NJ, pp. 23–61.

Mulgund, S. Stokes, J., Turieo, M. & Devine, M. (2002). Human/Machine Interface Modalities for Soldier Systems Technologies. Report to U.S. Army Natick Soldier Center SBCCOM, Reference 71950-00.

Ngo, M. K., Pierce, R. & Spence, C. (2012). Utilizing multisensory cues to facilitate air traffic management. *Human Factors,* 54, 1093–1103.

Nicogossian, A. E., Williams, R. S., Huntoon, C. L., Doarn, C. R., Polk, J. D., & Schneider, V. S. (2016). *Space Physiology and Medicine: From Evidence to Practice*. Springer.

Noel, J. P., Łukowska, M., Wallace, M., & Serino, A. (2016). Multisensory simultaneity judgment and proximity to the body. *Journal of vision*, *16*(3), 21–21.

Oskarsson, P.-E., Eriksson, L. & Carlander, O. (2012). Enhanced perception and performance by multimodal threat cueing in a simulated combat vehicle. *Human Factors*, 54, 122–137.

Overvliet, K. E., Azañón, E., & Soto-Faraco, S. (2011). Somatosensory saccades reveal the timing of tactile spatial remapping. *Neuropsychologia*, *49*(11), 3046–3052.

Oviatt, S. (2012). Multimodal Interfaces. In J. A. Jacko (Ed.). *The Human-Computer Interaction Handbook: Fundamemtals, Evolving Technologies, and Emerging Applications* (pp. 405-429). CRC Press.

Pavani, F., Spence, C., & Driver, J. (2000). Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychological science*, *11*(5), 353–359.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain: A journal of neurology*.

Persterer, A., Opitz, M., Koppensteiner, C., Nefjodova, M., Müller, C., & Berger, M. (1993). Audimir: Directional hearing at microgravity. *Journal of the Audio Engineering Society*, *41*(4), 239–247.

Podnar, G. Dolan, J. & Elfes, A. (2010) Telesupervised Robotic Systems and the Human Exploration of Mars. *Journal of Cosmology*, Vol. 12, pp. 4058–4067.

Popper, K. R., & Eccles, J. C. (1977). The self and its brain. *Springer International*.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological bulletin*, *124*(2), 123.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S. Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B. Martin, J. C., McTear, M., Raman, T. V., Stanney, K. M., Su, H. & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM, 47*(1), pp. 57–59.

Reschke, M. F., Clement, G., Thorson, S. L., Harm, D. L., Mader, T. H., Dudley, A. M., Wood, S. J., Bloomberg, J. J., Mulavara, A. P., Gibson, C. R. & Willams, D. R. (2016). Neurology. In Nicogossian, A. E., Williams, R. S., Huntoon, C. L., Doarn, C. R., Polk, J. D., & Schneider, V. S. (Eds.). *Space Physiology and Medicine: From Evidence to Practice* (pp. 245–282). Springer.

Rochlis, J. L., & Newman, D. J. (2000). A tactile display for International Space Station (ISS) extravehicular activity (EVA). *Aviation, Space, and Environmental Medicine*, 7, 571–578.

Roskies, A. L. (1999). The binding problem. *Neuron*, 24, 7–9.

Rupert, A.H., et al. (1996). Tactile Interface to Improve Situation Awareness. *Aviation, Space, and Environmental Medicine.* 67(A10): p. 53.

Rupert, A. H. (2000). An instrumentation solution for reducing spatial disorientation mishaps: A more “natural” approach to maintaining spatial orientation. *IEEE Engineering in Medicine and Biology Magazine,* March/April, 71–80.

Santangelo, V., & Spence, C. (2009). Crossmodal exogenous orienting improves the accuracy of temporal order judgments. *Experimental brain research*, *194*(4), 577–586.

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, 36, 439–445.

Sarter, N. B. (2013). Multimodal displays: conceptual basis, design guidance, and research needs. In J. D. Lee, A. Kirlik & M. J. Dainoff (Eds.). *The Oxford Handbook of Cognitive Engineering* (pp. 556–565). Oxford University Press.

Schmidt, G. R., Landis, G. A & Oleson, S. R. (2012) HERRO Missions to Mars and Venus Using Telerobotic Surface Exploration from Orbit. *43rd Lunar and Planetary Science Conference.*

Sheridan, T. (1993) Space Teleoperation Through Time Delay: Review and Prognosis." *IEEE Transactions of Robotics and Automation*, Vol. 9, No. 5, pp. 592–605.

Sipes, W. E., Polk, J. D., Beven, G. & Shepanek, M. (2016). Behavioral Health and Performance. In Nicogossian, A. E., Williams, R. S., Huntoon, C. L., Doarn, C. R., Polk, J. D., & Schneider, V. S. (Eds.). *Space Physiology and Medicine: From Evidence to Practice* (pp. 367–389). Springer.

Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*(1), 182–200.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, *1296*(1), 31–49.

Spence, C. & Driver, J. (1997). Crossmodal links in attention between audition, vision, and touch: Implications for interface design. International Journal of Cognitive Ergonomics 1 (4), 351–373.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport*,*11*(9), 2057–2061.

Spence, C., & Gallace, A. (2007). Recent developments in the study of tactile attention. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *61*(3), 196.

Spence, C. &. Ho, C. (2008). Tactile and multisensory spatial warning signals for drivers. *IEEE Transactions on Haptics,* 1(2), 121–129, Jul.–Dec. 2008.

Spence, C., & McDonald, J. (2004). *Of the Exogenous Spatial Orienting of Attention*. *The handbook of multisensory processes*, 3.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, *130*(4), 799.

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

Stanney, K. M. & Cohn, J. V. (2012). Human-Computer Interaction in Aerospace. In J. A. Jacko (Ed.). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications* (pp. 643–667). CRC Press.

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *Cognitive Neuroscience, Journal of*, *1*(1), 12–24.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Cognitive Neuroscience, Journal of*, *8*(6), 497–506.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266.

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.

Sutcliffe, A., 2003. Multimedia and virtual reality: designing multisensory user interfaces. LEA, Mahwah, NJ.

Teneggi, C., Canzoneri, E., Di Pellegrino, G., & Serino, A. (2013). Social modulation of peripersonal space boundaries. *Current biology*, *23*(5), 406–411.

Traylor, R., & Tan, H. Z. (2002). Development of a wearable haptic display for situation awareness in altered-gravity environment: Some initial findings. In *10th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperation Systems* (pp. 159–164). Los Alamitos, CA: IEEE Computer Society.

Valinia, A., Garvin, J. B., Vondrak, R., Thronson, H., Lester, D., Schmidt, G., Fong, T., Wilcox, B., Sellers, P. & White, N. (2012) Low-Latency Telerobotics from Mars Orbit: The Case for Synergy Between Science and Human Exploration, *First Exploration Telerobotics Symposium,* NASA Goddard Space Flight Center.

van Erp, J. B., Eriksson, L., Levin, B., Carlander, O., Veltman, J. A., & Vos, W. K. (2007). Tactile cueing effects on performance in simulated aerial combat with high acceleration. *Aviation, Space, and Environmental Medicine*, *78*(12), 1128–1134.

van Erp, J. B. F. & Self, B. P. (Eds.) (2008). *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments.* NATO RTO Technical Report, TR-HFM-122.

van Erp, J. B. F., van Veen, A. H. C. (2003). A Multi-purpose Tactile Vest for Astronauts in the International Space Station. *Proceedings of EuroHaptics*, Dublin, Ireland, Trinity College, 405–408.

van Erp, J. B. F. & van Veen, H., (2004). Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behavior,* 7, 247–256.

van Erp, J. B. F., van Veen, A. H. C. (2006). Touch down: The effect of artificial touch cues on orientation in microgravity. *Neuroscience Letters*, 404, 78–82

van Erp, J. B. F., van Veen, A. H. C., Jansen, C., & Dobbins, T. (2005). Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception*, 2, 106–117.

van Erp, J.B.F., Van Veen, H.A.H.C. & Ruijsendaal, M. (2007). More than a feeling: bringing touch into astronauts' spatial orientation. *Microgravity Science and Technology*. 19 (5), 108–112.

van Veen, H. A. H. C., & van Erp, J. B. F. (2000). Tactile information presentation in the cockpit. In S. Brewster & R. Murray-Smith (Eds.), *Haptic HCI 2000* (pp. 174–181). Berlin: Springer Verlag.

van Veen, H. A. H. C., & van Erp, J. B. F. (2003). Providing directional information with tactile torso displays. Eurohaptics, 471–474.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607.

Wenzel, E. M. (2001). Effect of increasing system latency on localization of virtual sounds with short and long duration*. Proceedings of the International Conference on Auditory Display, ICAD 2001,* Espoo, Finland, July 29-August 1, pp. 185–190.

Wenzel, E. M., Godfroy, M. & Miller, J. D. (2012) Prototype Spatial Auditory Display for Remote Planetary Exploration. *Proceedings of the 133rd Convention of the Audio Engineering Society,* San Francisco, California, October 26–29.

Wenzel, E. M., Godfroy, M. & Miller, Joel D. (2014). “Spatial auditory displays: Substitution and complementarity to visual displays.” *Proceedings of the International Conference on Auditory Display,* New York, NY, June 22-25, 2014.

Wenzel, E. M. & Godfroy-Cooper, M. (2015). Multimodal augmented displays for surface telerobotic missions. NASA HRP-SHFH Report.

Wenzel, E. M. & Godfroy-Cooper, M. (September, 2016). *The Role of Tactile Cueing in Multimodal Displays: Application in Complex Task Environments for Space Exploration.* Report to the Human Research Program, Space Human Factors and Habitability Element, Risk of Inadequate Human-Computer Interaction (HCI).

Wickens, C.D. & Liu, Y. (1988). Codes and modalities in multiple resources: a success and a qualification. *Human Factors*, *30*(5), pp. 599–616.

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

Wickens, C.D., Goh, J., Helleborg, J. Horrey, W. J. & Tailleur, D. A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors*. *45*(3), pp. 360–380.

Woods, D.D. (1993). The price of flexibility. In: *Proceedings of the First International Conference on Intelligent User Interfaces.* Orlando, FL, pp. 19–25.

Woods, D.D., Johannesen, L. & Potter, S. (1994). The sophistry of guidelines: Revisiting recipes for color use in human-computer interface design. In: *Proceedings of the 36th Annual Meeting of the Human Factors and Ergonomics Society*.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871–884.

Yost, W.A. (2000). *Fundamentals of hearing*, Academic Press, New York.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005a). Audio-visual simultaneity judgments. *Perception & psychophysics*, *67*(3), 531–544.

Zampini, M., Brown, T., Shore, D. I., Maravita, A., Röder, B., & Spence, C. (2005b). Audiotactile temporal order judgments. *Acta psychologica*, *118*(3), 277–291.

Zampini, M., Torresan, D., Spence, C., & Murray, M. M. (2007). Auditory–somatosensory multisensory interactions in front and rear space. *Neuropsychologia*, *45*(8), 1869–1877.

Appendix: Excerpts of Selected Multimodal Display Guidelines

MIL-STD-1472G, 11 January 2012, Department of Defense Design Criteria Standard: Human Engineering (Superseding MIL-STD-1472F, 23 August 1999). This updated version of a widely used military standard provides a few standards for combining modalities in information displays. Typically, these have to do with combining visual and auditory information and emphasize the primacy of visual information. Except for 3-D audio displays, newer technologies such as tactile displays are not addressed.

5.3.1.2.6 Relation to visual displays. When used in conjunction with visual displays, audio warning devices shall be supplementary or supportive and shall be used to alert and direct user attention to the appropriate visual display.

5.3.1.5.7 b. Redundant visual warning. All nonverbal aural annunciations shall be accompanied by a visual annunciation which defines the condition. In a cockpit, this may be an illuminated display. In the case of a warning horn on a backing vehicle, the vehicle's backward motion provides adequate redundancy.

5.3.1.7 Audio displays as part of the user interface. Audio displays may be used as part of the user-computer interface, where (a) the common mode of visual display is restricted by overburdening or user mobility needs and it is desirable to cue, alert, or warn the user, or (b) the user shall be provided feedback after control actuation, data entry, or completion of timing cycles and sequences. For other requirements, see 5.3.1.1 and 5.3.1.2. For frequency, see 5.3.1.3.3. For audibility of audio displays, see 5.3.1.4.1.

a. Supportive function. Audio signals used in conjunction with visual displays shall be supplementary to the visual signals. Audio signals shall be used to alert and direct the user's attention to the appropriate visual display.

5.3.1.7.1 3-D audio displays. 3-D audio displays shall meet the following:

a. Use. 3-D audio displays or multiple voice communications may be used in an environment with numerous and important spatial cues or where an user is likely to be highly tasked visually (e.g., fighter cockpits) to enhance situation awareness, segregate multiple channels, or rapidly redirect the user's vision.

b. Presentation format. For most applications, 3-D audio displays shall present data as discrete sound sources located at a constant distance at various azimuths and elevations.

c. Angular separation. Angular separation between discrete sounds shall be not less than 15 degrees in the horizontal plane and not less than 30 degrees in the vertical plane.

d. Binaural versus monaural. 3-D audio cues shall be presented binaurally.

ISO 14915-3 (2002) Software ergonomics for multimedia user interfaces — Part 3: Media selection and combination. This rather comprehensive set of multimedia design principles that calls for media combinations that support thematic congruence, manageable information loading, complementary viewpoints, consistency, and redundancy (see also Sutcliffe, 2003). However, in considering integration of multimodal information, it tends to focus on the impact of media selection per se rather than modality selection; for example, combining still pictures with moving images, or nonrealistic audio (e. g., tone sequences) with speech. Annex C of the standard briefly addresses cross-modality media:

Examples of media-combination patterns

C.1 General
This annex describes commonly used combinations with guidance on the design issues raised by each pattern. Media can be combined concurrently or sequentially to achieve a communication goal.
C.2 Moving image, sound, speech
Moving image with a voice track and sound is a familiar combination from film. This combination can be used effectively for explaining visual information via the voice channel, with sounds to give additional information about the image.
EXAMPLE A video of a fire-evacuation procedure with speech instructions and the sound of a real fire incident to dramatize the presentation.
C.3 Still image, speech, text
Photographs or diagrams can be augmented with text captions or presented separately with explanatory text. Speech can be used to draw attention to specific parts of the image and reference important parts of the text. In this combination, important information can be conveyed by text, and speech used as a supplementary commentary.
EXAMPLE In a tourist-information kiosk, a photograph of a landscape is described in speech with text captions to point out important locations of interest.
C.4 Still image, speech, sound
In this combination, the audio channels can clash, so care should be taken that speech and sound are suitably integrated. Sound can be used to convey information related to the image with a voice commentary.
EXAMPLE A photograph of a bird is accompanied by bird song and a speech clip introducing the species.
C.5 Two (or more) still images, speech, text
Images can be compared or linked together in a sequence using text captions and speech to draw the user’s attention to important information.
EXAMPLE Two similar flowers are shown and the user’s attention is drawn to differences by text captions and a voice commentary.
C.6 Two (or more) texts, speech
Different texts can be compared or linked together using speech to draw the user’s attention to important words or phrases. This combination can be used when single words or phrases need to be found; however, only one text can be read at a time, so if the complete text has to be assimilated, this combination should not be used.
EXAMPLE A modern Greek text is displayed alongside an ancient Greek text with the same subject matter, so that differences in the written language can be evaluated.

Stanney, K., Samman, S., Reeves, L., Hale, K., Buff, W., Bowers, C., Goldiez, B., Nicholson, D. & Lackey, S. (2004). A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations. *International Journal of Human-Computer Interaction*. 17(2), pp. 229–257. This paper provides useful tables of design guidelines for the individual visual, auditory and haptic modalities. It also proposes design guidelines for multimodal integration based on consideration of evidence from the behavioral and, to some extent, the neurophysiological literature.

Stanney, et al. (2004) Table 5: Preliminary Cross-Modal Integration Rules

- Temporal and spatial coincidence: When seeking a large amount of perceptual integration of multimodal stimuli and neural coactivation (e.g., for enhanced reaction time [RT], augmented perception, enhanced memory), ensure different modalities are close temporally and spatially (Driver & Spence, 2000).
- Working memory (WM) capacity enhancement: When seeking WM capacity enhancements, direct sensory stimuli to a multitude of sensory modalities while avoiding extensive cross-encoding among visual and auditory percepts into linguistic terms (Baddeley, 1990, 2000; Barnard, 1999; Schneider, 1999; Sulzen, 2001).
- Intersensory facilitation effect (IFE): When seeking enhanced target acquisition, introduce a preceding sensory accessory in an alternative modality than the primary percept (Heuermann & Colonius, 2001).
- Congruency effectiveness: When seeking enhanced RT via redundant-signals or neural coactivation, employ congruent combinations of cross-modal percepts as opposed to incongruent combinations (Dyson & Quinlan, 2002).
- Spatial orientation augmentation: When seeking enhanced spatial orientation, ensure multimodal spatial senses yield congruent information (Freedman & Rekosh, 1968), such as by coupling head movements with visual scene updating or auditory localization.
- Magnitude or inverse effectiveness: If a system has less than optimal displays (e.g., low resolution CRT), then couple with additional modality displays in order to garner the multiplicative effect of crossmodal integration (Meredith & Stein, 1986).

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S. Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B. Martin, J. C., McTear, M., Raman, T. V., Stanney, K. M., Su, H. & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM, 47*(1), pp. 57–59. This article discusses six categories of guidelines and represents a preliminary effort to establish general principles for multimodal interaction design (excerpted summary below).

Requirements Specification.
Design for broadest range of users and contexts of use. Designers should become familiar with users' psychological characteristics (for example, cognitive abilities, motivation), level of experience, domain and task characteristics, cultural background, as well as their physical attributes (for example, age, vision, hearing).

Address privacy and security issues. Users should be recognized by an interface only according to their explicit preference and not be remembered by default. In situations where users wish to maintain privacy by avoiding speech input or output, multimodal interfaces that use speech should also provide a non-speech mode to prevent others from overhearing private conversations.

Designing Multimodal Input and Output.
Maximize human cognitive and physical abilities. Designers need to determine how to support intuitive, streamlined interactions based on users' human information processing abilities (including attention, working memory, and decision making) for example:

- Avoid unnecessarily presenting information in two different modalities in cases where the user must simultaneously attend to both sources to comprehend the material being presented; such redundancy can increase cognitive load at the cost of learning the material.
- Maximize the advantages of each modality to reduce user's memory load in certain tasks and situations, as illustrated by these modality combinations:

Integrate modalities in a manner compatible with user preferences, context, and system functionality. Additional modalities should be added to the system only if they improve satisfaction, efficiency, or other aspects of performance for a given user and context.

- Match output to acceptable user input style (for example, if the user is constrained by a set grammar, do not design a virtual agent to use unconstrained natural language);
- Use multimodal cues to improve collaborative speech (for example, a virtual agent's gaze direction or gesture can guide user turn-taking);
- Ensure system output modalities are well synchronized temporally (for example, map-based display and spoken directions, or virtual display and non-speech audio);
- Ensure the current system interaction state is shared across modalities and that appropriate information is displayed in order to support: users in choosing alternative interaction modalities, multidevice and distributed interaction, system capture of a user's interaction history.

Adaptivity. Multimodal interfaces should adapt to the needs and abilities of different users, as well as different contexts of use. Dynamic adaptivity enables the interface to degrade gracefully by leveraging complementary and supplementary modalities according to changes in task and context. Individual differences (for example, age, preferences, skill, sensory or motor impairment) can be captured in a user profile and used to determine interface settings.

Consistency. Presentation and prompts should share common features as much as possible and should refer to a common task including using the same terminology across modalities. Additional guidelines include providing consistent:

- System output independent of varying input modalities (for example, the same keyword provides identical results whether user searches by typing or speaking).
- Interactions of combined modalities across applications (for example, consistently enable shortcuts).
- System-initiated or user-initiated state switching (for example, mode changing), by ensuring the user's interaction choices are seamlessly detected and that the system appropriately provides feedback when it initiates a modality change.

Feedback. Users should be aware of their current connectivity and know which modalities are available to them. They should be made aware of alternative interaction options without being overloaded by lengthy instructions that distract from the task. Also, confirm system interpretations of whole user input after fusion has taken place, rather than for each modality in isolation.

Error Prevention/Handling. User errors can be minimized and error handling improved by providing clearly marked exits from a task, modality, or the entire system, and by easily allowing users to undo a previous action or command. To further prevent users from guessing at functionality and making mistakes, designers should provide concise and effective help in the form of task-relevant and easily accessible assistance. Some specific examples include:

- Integrate complementary modalities in order to improve overall robustness during multimodal fusion, thereby enabling the strengths of each to overcome weaknesses in others.
- Give users control over modality selection, so they can use a less error-prone modality for given lexical content.
- If an error occurs, permit users to switch to a different modality.
- Incorporate modalities capable of conveying rich semantic information, rather than just pointing or selection.
- Fuse information from multiple heterogeneous sources of information (that is, cast a broad "information net").
- Develop multimodal processing techniques that target brief, or otherwise ambiguous information, and are designed to retain information.

Giang, et al. (2010) ***Multimodal Interfaces: Literature Review of Ecological Interface Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces.*** **Defence R&D Canada Contract Report, DRDC Toronto CR 2010-051, (pp. 269).** This comprehensive document addresses a variety of topics related to the design of multimodal interfaces. In chapter 5, Crossmodal Attention, the authors list a number of proposed design guidelines based on evidence from crossmodal attention research (p. 74).

“For multimodal interface designers, there are many factors to consider for the optimal communication of information. The study of multimodal perception, integration and application to interface design not a widely understood field, despite the fact that abundant research exists. However, a number of design guidelines have surface from this review on crossmodal attention:

- Stimuli should be placed in the peripersonal space for maximum effectiveness.
- Auditory and tactile stimuli are best for presenting warning signals.
- Although there is no direct method of prevention for operator confusion by multimodal integration, reference to past experiments and simulation using Bayesian modeling can be useful tools in preventing conflict situations.
- Using multimodal cues may be beneficial in directing the operator’s attention to a single location accurately, but the use of multiple senses may slow response time. Thus, the use of multimodal cues depends on the important of accuracy versus response time.
- The use of two sensory modalities is useful when in-parallel processing (attention is spatially divided) is required.
- Low speed stress and load stress result in higher operator performance, thus designers should keep the number of channels and the rate of change of signal presentation low.
- Attention is focused more on channels that update frequently and thus more important parameters should be displayed at a higher frequency.
- Steps should be taken to reduce complacency, which occurs with higher reliable sources. Higher reliable sources should thus be displayed using warning signals instead of monitoring tasks for maximum effectiveness.

Giang, et al. (2010) also propose design considerations for **Intelligent Adaptive Interfaces** that include multimodal displays, although they indicate that this research area is not well-developed and concrete guidelines are yet to be established (p. 77–79).

“Attempts for literature review of adaptation guidelines to multimodal displays have been conducted; however, concrete guidelines for this area have yet to be established. Nonetheless, basic guidelines have been provided. In order to convey information adaptively in a multimodal way, the following six considerations must be taken into account:

- Choice of the information that is to be conveyed (“content selection”).
- Selection of modalities through which the information will be conveyed (“modality allocation”).
- Selection of the format in which the modalities will be used to present that information (“modality realization”).
- Determinations of mechanism(s) that are used combine the modalities (“modality combination”).
- Evaluating the effect of environmental and cognitive factors on user’s perceptual integration (“situated multimodality”).
- Analysis of performance of the human user in the interface (“task analysis”) (Tripathi, 2008).

The interface’s ability to collect personal data (individual differences) allows the display to present information adaptively and in accordance with the situation and the user’s needs and preferences. For example, if the interface detects that an operator is experiencing an overload in the visual modality, the interface could adapt and present information through another modality such as tactile or audition. Hou et al. (2007b) pointed out that it is absolutely vital for the operator’s states and intentions to be clear to the interface; thus, it would be helpful for the interface to indicate its perception of the operator’s states, intentions and mission goals. Additional basic multimodal guidelines are as follows:

- Maximize advantages of each modality to reduce user’s memory load in certain task and situations.
- Integrate compatible modalities in context with user preferences and system functionality, for example, allow gestures to augment or replace speech input in noisy environments.
- Avoid presenting information in different modalities unnecessarily in cases where the user must attend both sources to comprehend the material being presented. This can cause an increase cognitive load at the cost of learning material” (Reeves et al., 2004).
- Selection options for preferred presentations via different modalities should be available.
- Users should be able to adjust in terms of scalability individual modalities. For example, features within individual modalities such as display contrast should be able to adjust in accordance to the environment and the user’s preferences.
- Schneider-Hufschmidt, Groh, Perrin, Hine, & Furner (2003) said that information content should be designed appropriately to provide constant multimodal presentation and be stored in “delivery-independent form” so that translations of information in different modalities are consistent. This statement contradicts the first guideline provided within this list. The first statement appears to be a more intuitive guideline since attempting to present information “delivery-independent form,” may result in

downplaying modality specialization. It is important that the development of multimodal adaptive interfaces select modality usage optimally.

- Modality selection in the design stage should be determined by two factors: appropriateness and availability in relation to various factors such as urgency, purpose, information importance, and processing code along with each modality being assigned a rank order value of 0–1 depicting its desirability level; 0 being the least desirable and 1 being the most desirable (Hameed & Sarter, 2009). It is not specified why the authors suggested a ranking system from 0–1 for desirability level; however, another ranking system (e.g. 1–5) could also be employed as long as each ranking level is clearly defined and consists of specific criteria. This allows the interface to take on the responsibility of DRDC Toronto CR 2010-051 79 automation in accordance with the user's preferences and needs by collaborating authority amongst the user and the interface."