Review of Low Frame Rate Effects on Human Performance

Jessie Y. C. Chen and Jennifer E. Thropp

***Abstract*—In this paper, we conducted a comprehensive survey of the effects of different frame rates (FRs) on human performance and reviewed more than 50 studies and summarized them in the areas of psychomotor performance, perceptual performance, behavioral effects, and subjective perception. Overall, there seems to be strong support for a threshold of around 15 Hz for many tasks, including those that are psychomotor and perceptual in nature. Less impressive yet acceptable performance may be accomplished at around 10 Hz for many tasks. Subjective reactions to the quality and watchability of videos seem to support rates of 5 Hz, although videos presented at 15 Hz and above are generally more widely preferred. These generalizations regarding superior and acceptable FRs may also be subject to the effects of several moderating factors such as display characteristics, nature of the tasks, viewing condition, additional cues, and user experience.**

***Index Terms*—Frame rate (FR), human factors, human performance, video quality.**

I. Introduction

In RECENT years, there have been growing applications of robotic technologies in fields such as space exploration, search and rescue, national defense, entertainment, police special weapons and tactics operations, health care, and personal assistance [1]. These unmanned systems will extend the ranges and capabilities of their human operators' perception and action, and will have a major impact on future combat operations [2]. In order to effectively employ the unmanned assets, it is essential that the robotic operator maintain an effective perception of the remote environment through the communication channel between the human operator and the robot. However, factors such as distance, obstacles, bandwidth, and electronic jamming may pose challenges for maintaining sufficient signal strength [3]. As a result, the quality of video feeds that a robotic operator relies on for remote perception may be degraded, and the operator's performance may be compromised [4].

Common forms of video degradation caused by low bandwidth include reduced frame rate (FR) (frames per second), reduced resolution of the display (pixels per frame), and a lower gray scale (number of levels of brightness or bits per frame) [5]. The product of FR, resolution, and gray scale is bandwidth (bits per second), and it is important to determine how to tradeoff these three variables with a given bandwidth so that the operator performance can be optimized [6]. Generally, for applications in virtual environments (VEs), many researchers recommend 10 Hz (i.e., 10 frames per second or 10 fps) to be the minimum FR to avoid performance degradation [7]. However, depending on the nature of the tasks, the same degrees of slowing of FRs may have different effects on human performance. To provide a comprehensive survey of the effects of different FRs on human performance, this paper reviewed more than 50 studies and summarized them in the areas of psychomotor performance, perceptual performance, behavioral effects, and subjective perception. First, some background information about the mechanisms of FR and other related sources of temporal distortions are presented. This discussion is followed by a brief review of the major moderating factors that can impact the adverse effects of slow FR on human performance.

A. Mechanisms of FR

FR is the tempo at which the new frames of the graphical scene are computed, rendered, and displayed. Frame time (the length of time required to compute, render, and display the image), which is the inverse of FR, is the primary source of delay in desktop personal computers; in VEs, input devices such as 3-D trackers are additional sources of delay [8]. FR is limited by several computational factors, including frame time, which is in turn determined by the current computational demands and the actual image displayed [i.e., complexity or level of detail (LOD)]. Possible sources of low FRs in VEs include the computation required to process tracker data (e.g., input from peripheral devices such as mice or joysticks), computation speed resulting from the effects of user movement (e.g., use of head trackers), the graphical rendering time, and communication overhead in distributed systems. If any of these factors overwhelms the system or bandwidth limitations, the FR may be reduced. The FR however is not an entity that exists in isolation; it is also associated with other computational demands.

B. Related Source of Temporal Distortion

Bryson [9] differentiates FR from lag, which are two major sources of temporal distortion in real-time computing systems. Lag is the temporal discrepancy between the data stream from the tracker (e.g., input device such as a mouse) and the actual resulting graphic. Therefore, lag is the time delay between

Manuscript received May 8, 2006. This work was supported by the Robotics Collaboration Army Technology Objective of the U.S. Army. This paper was recommended by Associate Editor M. Dorneich.

J. Y. C. Chen is with the U.S. Army Research Laboratory, Orlando, FL 32826 USA (e-mail: jessie.chen@us.army.mil).

J. E. Thropp is with the Kennedy Space Center, United Space Alliance, Cape Canaveral, FL 32899 USA (e-mail: jennifer.thropp@usa-spaceops.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2007.904779

the user's action and its displayed result [10]. In terms of teleoperation, delay is considered to be the temporal difference between communication between local and remote sites (i.e., how long it takes for the user's command or input to produce a result in the system's action) [11]. Lag is also related to system latency (SL) and system responsiveness (SR).

Lag is created by the following factors: delays in tracker signal; delays in communication between tracker and computer; delays from computations involved in processing tracker data; and delays from graphical rendering. SL is the frame time plus the time needed for the system to collect input samples from the real world, such as tracker performance. The SR resolution however can be complicated as it is the combination of SL and the time between the completion of the user's last action and the next source of input into the display. Furthermore, Watson *et al.* [10] explain that frame time, SL, and SR vary randomly. The variation in latency is greater than the variation in frame time, and the variation in SR is the greatest, as it builds upon the variation in both frame time and SL. Thus, it is important to consider both the means and standard deviations (SDs) of these display speeds.

Watson *et al.* [10] maintain that there are several factors that cause delay in an interactive graphics system, some of which affect only frame time, while others impact only latency, and still others impact both frame time and latency [8]. However, all these potential sources impact SR. Thus, FR and lag are closely related and are often manipulated at the same time (e.g., see [12]). Bryson [9] has stated that it is difficult to isolate the effects of lag and FR from other types of VE distortions. Watson *et al.* [10] also state that both lag and FR should be considered by designers. However, there is not much research to guide these designers.

Latency and FR occur together but must be carefully manipulated in experiments in order for the effects to be properly detected. Ellis *et al.* [13] found that latency was a more reliable and stronger influence than FR on tracking performance in a VE. In a tracing task, latency was also more influential, i.e., having a multiplicative effect on response time [14]. This may be because latency can be experimentally varied over a greater range than FR. For instance, in the study by Ellis *et al.* [13], the experimental latency ranged from 80 to 480 ms, which yields a proportion of 1 : 6, whereas the FR range of 6–20 fps yields a proportion of 1 : 3.7.

C. Update Rate and Refresh Rate

The *update rate* in teleoperation or in VE applications refers to the frequency at which the image of the remote site is captured and then displayed to the remote operator (e.g., via head-mounted display (HMD) or other displays), and it depends on the bandwidth limitation [15]. The update rate is therefore the upper limit and determinant of display rate. For instance, if the FR of the HMD peaks at 100 Hz, and the update rate was 30 Hz, unique images could only be displayed at a maximum rate of 30 Hz. Usually, in VE applications, FRs are higher than the update rates, so the image would appear smooth. In some VE literature, however, the update rate and FR are sometimes used interchangeably [16].

The *refresh rate*, on the other hand, refers to the number of images presented to the eye every second (usually between 10 and 60 Hz) [17]. Basically, the refresh rate is a hardware-determined constant, while the update rate can vary widely based on scene complexity and other factors, such as available computing power for generating the images [16], [18]. The following example by Richard *et al.* [17] helps to clarify the distinction between refresh rate and FR: a display system with a refresh rate of 60 Hz and an FR of 4 Hz will present 15 consecutive identical images before there are any changes in the scene.

Watson *et al.* [7] state that both high LOD in a frame's image and adequate SR facilitate user performance, and yet make strong computational demands that often end in an LOD/SR tradeoff. The update rate is directly proportional to bandwidth and is impacted by factors such as computation and rendering speeds of the graphics system. Thus, the update rate could be increased by generating scenes with simpler components such as lines and bioptic images as opposed to polygons and stereo images [15]. Cai [19] describes the merit of context-aware displays, in which the human user's perceptual state (e.g., user's scanning behavior with context anticipation) and the actual display content are monitored in order to provide the optimum LOD for that specific image. This is an important consideration, as human visual search can be guided by content and purpose [20], and excessive LOD can be minimized when not needed in order to allocate bandwidth to other demands, such as higher FR. In terms of temporal resolution, Cai [19] found that when a video was presented at less than 4 Hz, the bandwidth usage was clearly lower than continuous streaming at approximately 17 Hz. More context-aware displays such as gaze-contingent displays will be discussed in Section II.

D. Variable Versus Constant FR

Many bandwidth studies have empirically manipulated constant FRs. Yet, the amplitude and deviation of FRs can vary in a computing system, which ultimately yields a variable FR that may be reported as a range or mean. This variability can occur in presentations with significant change in detail between scenes in the frames. Normally, in VEs, an adaptive detail management system adjusts the object polygon count from frame to frame, depending on the LOD needed, and consequently, this determines the frame update rate in order to present an approximately constant polygon count in each scene. However, when the LOD changes dramatically between frames, an adaptive detail management system may initially calculate the timing of the next frame erroneously.

Fluctuations in the actual frame times occur even when a specific FR is sought in a controlled environment. In other words, a particular FR can actually be a range of FRs, which yields a mean FR along with SD in frame time. For instance, Watson *et al.* [10] used the experimental stimuli presented for a 100-ms frame time and introduced an SD in a frame time of 20 ms, which in turn generated a range of frame times of 72–129 ms (the inverse of which corresponds to a range of FRs of 13.9–7.8 Hz). The implementation of frame time SD can replicate the reality that there are numerous mechanisms that result in a variable, rather than constant, FR in video transmission. Procedures for calculating variable FR values are presented in detail in [10].

II. Adverse Effects of Slow FR on Task Performance and Moderating Factors

Extremely low update rates can be problematic. As the FR becomes lower, the time (lag) between frames becomes longer, and the scene may appear jittery. Objects may consequently appear to move in saccades, and the human visual system has to conduct spatiotemporal interpolation to "fill" these visual gaps [17]. Arthur *et al.* [14] state that the frame update rate and the lag in receiving and processing tracker data are among the two factors that are often associated with human performance in virtual worlds. Head-coupled tracking involves tracking head movements such that when the user moves his or her head, the perspective generated on the display will mimic real-life changes in perspective from head movements. Arthur *et al.* [14] state that FR is directly related to a specific type of head-coupled display system lag in which there is a delay between receiving the eye positions and updating the display. This is the time to compute and render a scene from the user's perspective, as current tracker measurements are needed; two are needed if the display is stereoscopic.

Liu *et al.* [15] state that 3-D tracking and "pick-and-place" tasks are two skills that are often required for telemanipulation. Reddy [21] explains that both poor FR and large delay can degrade the human–system interaction, often in terms of impaired depth perception on the visual display. Thus, the use of low FRs can have adverse effects on tele-manipulation, especially as far as depth perception is concerned. However, there are several moderating factors that might impact the adverse effects of low FRs. In this section, we briefly review five of those moderators: task dependency; viewing condition dependency; display luminance and the number of gray-scale levels; auditory cues in addition to visual cues; and end-user characteristics.

A. Task Dependency

Johnson and Caird [22], who investigated the impact of FR on sign language gesture recognition, remind readers that the required FR for the performance of a task can depend on the nature of the task itself. If animation must be conveyed, then 20 Hz may be required, considering that the human visual systems takes 50 ms to process a frame [23]. However, conveying a basic understanding of a sequence may require fewer frames.

It has been reported that 10 Hz is the minimum threshold for performance in immersive VEs (see [7], [9], [10], and [24]). However, higher rates of 10–15 Hz are required for acceptable performance [25]. For architectural walkthrough tasks, various minimum FRs have been recommended, including 6 Hz [26], 7 Hz [27], and 10 Hz [9], [25], [28].

Watson *et al.* [8] differentiate between open- and closed-loop tasks, each of which is affected by FR. In a closed-loop task, such as tracking, continuous feedback is provided. Thus, the result of the previous action feeds into the next action. Continuous feedback is therefore required for the operator to execute the next move. Automobile driving is considered to be a closed-loop tracking task because the driver is required to maintain the position of the vehicle within the specified lanes and to follow a prescribed pathway. In a tracking task, delay increases the time that passes until feedback is received by the user, and as a result, performance can degrade. Difficult tracking tasks are especially impaired by delays.

Open-loop tasks, on the other hand, have little or no feedback available to the operator. For example, placement task performance is dependent on predictive planning and is not guided by adjustments to feedback, as are closed-loop tracking tasks. However, placement tasks end in a closed-loop phase, where finer adjustments are made to place the object in its final position, i.e., a small target space.

B. Viewing Condition Dependency

Parkhurst and Niebur [29] demonstrate the human observer's different needs of visual detail as determined by different viewing conditions. In a *velocity-based* LOD graphics-rendering technique, objects remain stationary but appear as a blur in the observer's visual field if he or she turns his or her head. Therefore, the blurred object does not need to have a high LOD. Thus, LOD is reduced as the object moves across the visual field but then increases again as the observer slows and stops his or her viewpoint rotation. In this rendering method, LOD is linearly related to the rotational velocity of the observer. This method also keeps the system's computational load roughly constant.

In *distance-based* LOD manipulation (i.e., real-time viewpoint-dependent simplification), objects that are farther away from the viewer do not need to be rendered at high complexity. This reduction in detail then allows more resources to be allocated to the generation of higher FRs, as per the LOD–FR tradeoff. There are other times when the human visual system does not naturally process a high LOD, such as when objects are in the periphery of the visual field [30] and when rapidly moving objects appear as blurred [31], [32].

Resolutions can also be manipulated such that the user's area of interest is displayed in a higher resolution while the peripheral area is in a lower resolution [33], [34]. By reducing the resolution in the area not fixated by the user, the remaining computational resources can be used for increasing the FR. These variable-resolution displays are usually gaze contingent, but there are also the so-called "region of interest" displays that do not rely on real-time eye tracking [33]. Parkhurst *et al.* [35] suggested that high resolution in a foveal region of 5° viewing angle might be sufficient for visual search tasks. However, there are speed–accuracy tradeoffs in terms of user performance that the system designers need to consider (e.g., a user tends to perform search tasks less carefully and therefore more rapidly when the high-resolution central region is larger than 5° in radius) [35]. Additionally, gaze-contingent displays can affect the users' eye movements, their perceived quality of the video, and their task performance [33]. Two detailed reviews on gaze-contingent displays can be found in [33] and [34].

Another consideration is pixel size, which can interact with FR to influence the accuracy of perception of motion in depth. Although the object may be moving at a constant velocity, this velocity appears to accelerate or decelerate, depending on whether it is located near or far from the viewer. According to Pfautz [36], the motion of small objects near the line of sight and far from the viewer is likely to be limited by pixel size. However, the motion of large objects far from the line of sight and near the viewer is likely to be limited by FR. If the FR and pixel size are correct, then the viewer can detect the perspective depth.

There may also be differential effects for stereoscopic and monoscopic displays (see [14] and [37]). In monoscopic displays, the image is presented to only one eye; in some settings, this can be done to avoid retinal image disparity. In stereoscopic displays, both eyes receive an image, but the image displayed to each eye is slightly different because of the position of the eyes relative to each other. Lion [37] found that stereo performance was significantly better than mono performance in a manual tracking task, and also observed an interaction wherein the greatest performance occurred in stereoscopic displays with high FRs. This was confirmed by Richard *et al.* [17], who demonstrated that the tracking task completion time was lower overall for the stereo condition.

There may also be effects for the use of head-coupled imagery on visual performance (see [37]). In restricted head coupling, participants rest their chins against a chin rest to control for the effects of parallax. This is another viewing condition variable that may be a moderator in the relationship between FR and human performance if both are simultaneously manipulated.

C. Display Luminance and the Number of Gray-Scale Levels

The number of gray-scale levels (gl) is a characteristic of the video monitor and is usually measured in bits. Whiting *et al.* [38] demonstrated a linear relationship between gray-scale level and video display luminance: in the range of 80–180 gl, there is a slope of 0.209 $(\mathrm{cd/m^2})/\mathrm{gl}$. Whiting *et al.* were able to generate a mean luminance of 6.2 $\mathrm{cd/m^2}$ (1.8 ft-lm), which corresponds to a gray-scale level of 128. The gray-scale level can also serve as a source of image noise. Whiting *et al.* [38] manipulated a pixel noise SD of 5, 10, 14, and 20 gl for static images and 20 gl for dynamic noise. Also, the image contrast can be defined as the gray-scale level increment of the target. Thus, the gray-scale level could be a characteristic of visual display quality that could impact target detection performance independently as well as interactively with FR (also see [39]).

D. Auditory Cues in Addition to Visual Cues

It is also necessary to consider the influence of the auditory components of visual presentations. In some settings, audition is thought to play a larger role in temporal perception than vision (e.g., see [40]). For instance, musical tempo can influence the human perception of time-in-passing and display FR (see [41] and [42]). Speech reading ability, which is rooted in target recognition, is significantly more accurate in audio–visual conditions than in an audio-only mode [39], [43]. This effect may be a moderator that should be distinguished from possible FR manipulation effects.

Audio–visual skew, or asynchrony between the video and the audio aspects of the image sequence, can also affect recognition. It can be caused by low FRs that present outdated visual information that does not match the continuous real-time delivery of sound. While this mismatch can typically impair performance, it can be helpful in some instances of low FR. For example, an asynchrony in which the audio lags the visual component of the presentation can actually aid performance beyond that which can be normally perceived at 30 Hz [44]. The direction of audio–visual skew can be a moderator worth considering. Therefore, it is important to examine the full range of effects generated by an interaction among audition, vision, and FR.

E. End-User Characteristics

Although many aspects of the system and experimental methods can influence the relationship between FR and performance, it is also important to consider the individual differences in humans who use these systems. One possible moderator is the level of experience the operator has with computers and virtual reality (VR). Those who have been exposed to these systems are more familiar with the different viewing and psychomotor effects that can result from temporal distortions that would otherwise not be present in the everyday real-world human–machine experience. Therefore, the performance degradations associated with slow FR tend to be less pronounced for experienced operators [9]. Based on their data, Liu *et al.* [15] suggest a minimum update rate of 2 Hz for experienced users and 10 Hz for inexperienced users of a teleoperation tracking task. Adaptation to poor temporal fidelity has also been demonstrated. Increasing the operators' exposure to low temporal fidelity can aid performance, and training is one way to accomplish adaptation (see [22]).

The following sections discuss the adverse effects of slow FR in further detail, and the discussions are presented in the areas of psychomotor performance, perceptual performance, behavioral effects, and subjective perception.

III. Psychomotor Performance

A. Placement Performance

According to Fitts' law, the time required to acquire a target is a function of: 1) the physical distance to the target; and 2) the physical size of the target [45]. The interaction between low FR on Fitts' law was assessed by Bryson [9, Experiment 2]. This placement task involved moving a cursor from an initial position to a final position in a computer system. The performance means showed that increased lag and low FR both have similarly detrimental effects on placement performance. More specifically, lags and frame times greater than 250 ms (< 4 Hz) dramatically increased the difficulty level of the task. However, the effect of great lag and low FR was less pronounced in the participant who was experienced with VEs with poor SR.

Similarly, Liu *et al.* [15] found that operators who were experienced with HMDs were able to perform the pick-and-place tasks at update rates as low as 10 Hz but had significantly more errors below 2 Hz. Those who were inexperienced demonstrated impairments once the update rate dropped below 30 Hz, which suggests that the experience level can moderate the relationship between FR and performance. Liu *et al.* also observed that the task completion time was longer at update rates below 10 Hz.

Ware and Balakrishnan [12] assessed the effects of low FR and lag in a target acquisition in VR display. Reaction time degradations were the most dramatic up to 5 Hz but stabilized at around 10 Hz, beyond which performance improvements are minimal. Thus, it may not be worth the extra cost to increase FRs beyond 10 Hz for this type of task and setting.

While Ware and Balakrishnan [12] did not find much benefit for target acquisition performance when the FRs were over 10 Hz, Claypool *et al.* [46] showed that by increasing the FR from 15 to 30 Hz, the participants' first-person-shooter (FPS) video game scores (i.e., target acquisition performance) significantly improved. It was also found that participants had a very difficult time acquiring targets when the FRs were at 7 Hz or lower. In fact, when the FRs were increased from 3 to 60 Hz, the participants' scores resulted in a sevenfold performance improvement.

One way to enhance placement task performance at low FR is to provide force feedback. For example, Massimino and Sheridan [47] showed that an FR as low as 5 Hz might be adequate for placement tasks if force feedback was available.

Meehan *et al.* [48] assessed the placement performance in VEs; however, they focused on the physiological reactions imposed by various FRs rather than task performance. Therefore, this study is going to be discussed in Section V.

Watson *et al.* [10] examined placement performance in the context of variable FRs using the sinusoidal method to calculate the SD in frame time. Average placement and grab times significantly improved when the mean FR increased from 10 to 20 Hz. In a later study, Watson *et al.* [7] again assessed the effects of FR means and SD on grasping and placement tasks. This time, the SD was calculated using the frame-latency manipulation method, and they were absolute values of FRs (e.g., 4 Hz). The grasp time and the number of grasp attempts were significantly greater at the 9-Hz mean FR and when the SD of SR (SDSR) was 4 Hz. Overall, the grasping performance was impaired the most when the mean FRs were low and the SD was high. Based on these results, Watson *et al.* [7] speculated that improving the mean FR above 17 Hz may not necessarily increase the grasping performance; to test this, a second study was conducted using a set of higher FRs. FR SD was based on percentages of FRs (e.g., 5.6% of 17 Hz). It was found that increasing FR from 17 to 25 Hz did not significantly improve the grasping performance, as hypothesized. Improved placement performance occurred at higher mean FRs and lower SD, as was also found in Experiment 1. Overall, these results suggest that there may be a grasping performance threshold at higher FRs (such as at 17 Hz), and that grasping performance may be compromised only at lower mean FRs (such as 9 and 13 Hz).

Methods of calculating SD in frame time have also been compared. In a third experiment, Watson *et al.* [7] tested the hypothesis that SDSR control via absolute values was inferior to SDSR control via percentages of mean FR. The SD was manipulated as absolute values of FRs (e.g., 4 Hz). Results showed that there was a significant increase in grasp time only when the FR was at the lowest (17 Hz) and SDSR was highest (7.56 Hz). A reduction in the SR range occurred at higher FRs, and therefore, absolute FR control did not yield many significant effects. Thus, Watson *et al.* suspected that SR control based on absolute FR values may be inferior to SR control based on percentages of mean FRs.

In summary, results from studies that have manipulated a constant FR indicate that 10 Hz and less can be detrimental to placement performance in a VE. Specifically, it has been shown that FRs less than 4–5 Hz seem to dramatically increase the difficulty level of the task [9], [12]. However, the variable FR approach may be more accurate when the FRs are low, since SD and fluctuations of FRs are important considerations [10]. FR thresholds for ideal performance in variable FR configurations are likely similar to constant FR thresholds. Watson *et al.* [10] found that grasping performance was compromised only at lower mean FR such as 9 and 13 Hz. For improved placement performance, the average FR should increase to at least 17.5 Hz [7] or 20 Hz [10].

B. *Tracking Performance*

Tracking tasks require continuous manual control and include most forms of vehicle control and computer-based tasks that require continuous cursor/joystick positioning. Bryson [9, Experiment 1] assessed the tracking performance at various FRs and found that in general the normalized error was linearly dependent on frame time. Furthermore, the slope of this linear relationship was dependent upon the frequency of the targets moving in a sinusoidal path. Thus, longer frame times (i.e., lower FRs) at higher frequencies of target motion produced the greatest tracking error. In consideration of these results and those from Experiment 2 (see Section III-A), Bryson [9] concluded that frame time effects are congruent with those of delay in tracking tasks.

Vehicle control is considered to be a tracking task, since the driver must maintain position of the vehicle within the lanes following a specific pathway [8]. In a study on the teleoperation of unmanned ground vehicles (UGVs), McGovern [49] did not find driving performance degradation when the image update rates were lowered from 30 to 7.5 Hz. On the other hand, Van Erp and Padmos [4] reported that when the update rate dropped below 5 Hz in an indirect driving task, the participants' lateral control significantly degraded. Day [50] also assessed the effects of low FRs on a remote vehicle driving task. Mean task completion times and mean time in error increased as the FR decreased, and the performance for FRs 1–4 Hz was significantly lower than the control condition, i.e., 25 Hz. Not surprisingly, posttask questionnaires indicated that 50% of the participants felt they would have performed tasks significantly better with a higher-quality video.

Teleoperation-based tracking performance is of particular importance in FR studies as communication problems can arise from bandwidth limitation [15]. The UGV operator faces possible sources of signal degradation in the forms of physical distance from the UGV, obstacles, and electronic jamming [3].

Successful teleoperation therefore relies on tracking performance in terms of navigation and target tracking. Liu *et al.* [15] found that, similar to their placement task performance previously discussed, operators who were experienced with HMDs were able to perform tracking tasks at update rates as low as 10 Hz but had significantly more errors below 2 Hz. Inexperienced operators, on the other hand, showed a performance degradation almost immediately when the FRs started to decrease. French *et al.* [3] found that low FRs (i.e., 2 and 4 Hz) yielded significantly longer navigation times and higher cognitive, visual, and subjective workload ratings, but it did not affect the target identification accuracy or situational awareness. French *et al.* [3] concluded that 16 Hz is ideal for UGV teleoperation. However, to obtain effective task performance, a minimum of 8 Hz is necessary.

Lampton *et al.* [51] observed that FR degradations affect the participants' tracking performance in a VE. Although the FRs were not systematically manipulated, Lampton *et al.* did find that the slow FRs (i.e., approximately 5 Hz) made the 3-D tracking tasks very difficult. It was reported that participants were only able to keep the cursor on the moving target less than 9% of the time.

Ellis *et al.* [13] also examined tracking performance in VEs at various FRs and degrees of latency. Decreased FR (e.g., 6 Hz versus 20 Hz) significantly increased the root-mean-square error, increased the perceived control, and decreased feelings of stability. However, the participants' judgment of realism was not affected by the various FR and latency values.

Display format and viewing characteristics can also interact with FR to impact manual tracking performance. Lion [37] considered the differential effects for stereoscopic and monoscopic displays, as well as the use of head-coupled imagery. The 33-Hz FR yielded better manual tracking performance than the 22-Hz FR. The cursor was also kept closer to the target 21% more of the time in the 33-Hz condition as opposed to the 22-Hz condition. There was also a significant interaction between stereo displays and FR, with better performance in the stereo 33-Hz condition.

Arthur *et al.* [14] also manipulated the effects of stereo versus mono displays as well as the use of head-coupled imagery. Error rates in a 3-D tree-tracing task were measured. The response time increased along with increased lag and the lower FR of 10 Hz. Regression equations indicated that lag likely contributed more to performance effects than did reduction in FR.

An interaction between viewing condition and FR was also found by Richard *et al.* [17]. The tracking time was fastest and stable in the range of 14–28 Hz in the mono condition but remained stable at FRs as low as 7 Hz in the stereo condition. At low FR (i.e., below 7 Hz), the stereo vision increased performance by 50% over the mono condition. Additionally, when the FRs were low (i.e., 1–7 Hz), it took participants ten trials in the mono condition to reach performance asymptote. By contrast, in the stereo condition, the mean capture time was lower overall, and the curve appeared mostly flat from the earliest trials. Thus, less learning/adaptation took place at 1–7 Hz in the stereo condition due to the fact that it is more similar to natural viewing.

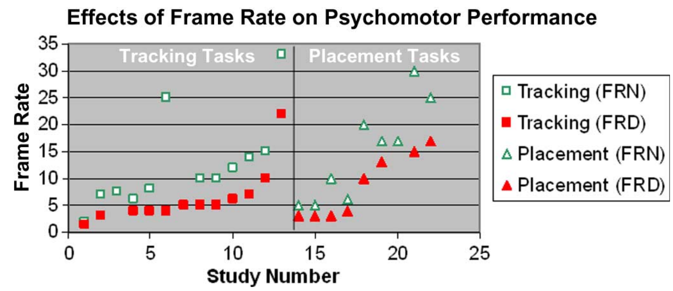

Fig. 1. Summary of FR effects on psychomotor performance. Note: The x-axis denotes the number the study was designated. The list of studies with their corresponding numbers is provided in Table I. Filled red squares and triangles signify the FR level where human performance starts to significantly degrade. Unfilled green squares and triangles signify the lowest FR levels where participants perform normally.

In summary, higher FRs such as 20 and 33 Hz yielded better manual tracking performance than lower FRs such as 6 Hz and below (see [13] and [37]). The detrimental effect of low FR is especially notable when the target moved at higher frequencies [9, Experiment 1].

Teleoperation may require higher FRs. French *et al.* [3] suggested that 16 Hz is ideal for tracking in UGV teleoperation and that 8 Hz is an absolute minimum requirement. Additionally, Day [50] demonstrated that remote vehicle control was significantly better at 25 Hz than at 4 Hz.

Acceptable tracking performance was found to be possible at very low FRs. Rates above 10 Hz may not be necessary for some applications; thus, systems with smaller bandwidth capability may still be effective. The minimal satisfactory FR was 2–10 Hz for teleoperation for experienced operators [15]. Finally, target acquisition in a 3-D environment seems to benefit most when the FR is increased to 10 Hz [12].

There may also be a significant interaction between stereoscopic display and FR, with better performance in stereo and at higher FRs. It therefore seems that the combination of high FRs and stereo displays more closely mimics the real-world visual experience, thereby requiring less user adaptation and learning [17]. This suggests that overall, if lower refresh rates are used, stereoscopic displays may compensate to boost operator performance. Based on these results, Lion [37] proposes the importance of giving design priority to the development of systems with stereoscopic displays and higher FR.

A summary of these research findings is presented graphically in Fig. 1 and is also provided in more detail in Table I.

IV. Perceptual Performance

A. Target Recognition Performance

1) Constant FR Effects on Target Recognition: Some researchers examined the effects of slow FR on target recognition performance in the military context. For example, French *et al.* [3] investigated the robotic operators' target identification performance using a UGV. Their simulated degradation of video feed from the UGV included 2, 4, 8, and 16 Hz. They found that FRs as slow as 2 Hz did not significantly affect their performance, including target identification and situational awareness. In another study, Chen *et al.* [52] examined

TABLE I
SUMMARY OF FR EFFECTS ON PSYCHOMOTOR PERFORMANCE

#	Author(s) and Ref. #	# of Sub	FR Manipulations	Task Type	Measures
1	Liu et al. [15]	2	0.1, 0.25, 0.5, 1, ***1.5***, **2**, 3, 5, 10 & 30 Hz	Tracking (experienced subjects)	Accuracy & Speed
2	Richard et al. [17]	42	1, 2, ***3***, **7**, 14, 28 Hz	Tracking & Grasping (Stereo condition)	Accuracy & Speed
3	McGovern [49]	NR	**7.5**, 30 Hz	Tracking	Vehicle Control
4	Bryson [9, Exp. 1]	2	2, 3, ***4***, **6**, 10, 20, 30, 60 Hz	Tracking	Accuracy
5	French et al. [3]	24	2, ***4***, **8**, and 16 Hz	Tracking	Accuracy & Speed
6	Day [50]	30	1, 2, 3, ***4***, and **25** Hz	Tracking	Accuracy & Speed
7	Lampton et al. [51]	24	***5*** Hz	3-D Tracking	Accuracy
8	Van Erp & Padmos [4]	8	3, ***5***, **10**, 30 Hz	Tracking	Vehicle Lateral Control
9	Liu et al. [15]	3	0.1, 0.25, 0.5, 1, 1.5, 2, 3, ***5***, **10** & 30 Hz	Tracking (inexperienced subjects)	Accuracy & Speed
10	Ellis et al. [13]	10	***6***, **12**, and 20 Hz	Tracking	Accuracy & Perceived control & stability
11	Richard et al. [17]	42	1, 2, 3, ***7***, **14**, 28 Hz	Tracking & Grasping (Mono condition)	Accuracy & Speed
12	Arthur et al. [14]	10	***10***, **15**, 30 Hz	3-D tree-tracing	Accuracy & Speed
13	Lion [37]	20	***22*** and **33** Hz	Tracking	Accuracy
14	Ware & Balakrishnan [12]	12	0.666, 1, 2, ***3***, **5**, 10, 15, 60 Hz	Target acquisition in a 3D environment	Speed
15	Massimino & Sheridan [47]	6	***3***, **5**, 30 Hz	Placement	Accuracy & Speed
16	Liu et al. [15]	5	0.25, 1, 1.5, 2, ***3***, **10** & 30 Hz	Placement	Accuracy & Speed
17	Bryson [9, Exp. 2]	2	2, 3, ***4***, **6**, 10, 20, 30, 60 Hz	Placement	Accuracy & Speed
18	Watson et al. [10]	10	***10*** and **20** Hz	Placement	Accuracy & Speed
19	Watson et al. [7, Exp. 1]	11	9, ***13***, and **17** Hz	Placement & Grasping	Accuracy & Speed
20	Watson et al. [7, Exp. 3]	10	mean FRs **17**, 33, & 41 Hz, each with SDs of 0.50, 3.78, & 7.56 Hz	Placement & Grasping	Accuracy & Speed
21	Claypool et al. [46]	60	3, 7, 15, 30, and 60 Hz	Target acquisition (first-person-shooter video game score)	Accuracy
22	Watson et al. [7, Exp. 2]	12	***17***, **25**, and 33 Hz	Placement & Grasping	Accuracy & Speed

Note: FR values in ***Bold-Italics*** signify the FR levels where human performance starts to significantly degrade. FR values in **Bold** signify the lowest FR levels where participants perform normally. Numbers of research participants are reported in the third column (NR= Not reported).

the robotic operator performance in a simulated military environment. They manipulated the FRs (5 versus 30 Hz) of the sensor video feed of the simulated unmanned aerial vehicle (UAV) and UGV. The participants' target detection performance was not significantly degraded as the FRs dropped from 30 to 5 Hz.

Cai [19] reported that human figure recognition and tracking in a video was found to be a log function of FR, and a major improvement in this performance occurs by increasing the FR from 0.05 Hz to 1 Hz. Thus, the minimal satisfactory FR was 1 Hz in normal viewing conditions. Performance increases somewhat between 1 and 7 Hz, and then seems to stabilize from 7 Hz until the maximum tested value of 17 Hz.

Sign language interpretation is a pattern recognition task that may depend on temporal resolution. Johnson and Caird [22] found that a low FR can impair the recognition of American Sign Language gestures. Although there was no significant main effect for FR, means indicated a trend for impaired performance and increased error rate at lower FRs (e.g., 1 and 5 Hz). However, over the course of additional trials, there was an improvement in performance. Training may therefore be useful when the bandwidth is limited.

Perception of the visual cues associated with speech and lip reading may also be considered a form of target recognition. Visual information can aid interpersonal communication for those with impaired hearing by coupling the visual cues of speech reading or "lip reading" with the limited auditory perception that remains. FR is thought to impact specific aspects of lip reading, including the apparent duration of visual speech components such as mouth movements [53] and dynamic components such as syllable timing [54].

Adding audio cues to visual presentation can aid target recognition. Frowein *et al.* [43] assessed speech reception target recognition in the domain of video telephony. Audio–visual recordings of common sentences were presented in both audio–visual and audio-only modes. Participants were asked to repeat each sentence immediately after the presentation. At 5–6 Hz, the performance in the audio–visual condition was significantly better than the audio condition, and at 15 Hz, the added performance improvement stabilized, with no further performance benefits for the 30-Hz condition. Thus, there may be an interaction between presentation mode and FR in speech perception.

Vitkovitch and Barber [39] hypothesized that the perception of the speaker's facial movement was important to help listeners perceive speech. In this study, participants listened to two simultaneous recordings of a message and shadowed (i.e., repeated) the words of the target message. The recordings were in either audio-only or audio–visual format. Having a visual display did improve performance over an audio-only display. Even the audio–visual video presented at the lowest FR (8.3 Hz) yielded improved performance over the audio-only condition; this is consistent with the findings of Frowein *et al.* [43]. Across all FRs, performance improved when the FR was increased. However, there was no significant performance benefit from increasing the FR from 16.7 to 25 Hz, which suggests that 16.7 Hz may be adequate for video transmission for recognition tasks of this nature; this is also consistent with the findings of Frowein *et al.* [43].

In another study by Vitkovitch and Barber [55, Experiment 1], participants lip-read number sequences in the range of 1–30 in a visual-only condition. By turning off the sound, Vitkovitch and Barber were able to determine that the detrimental effect of low FR is not attributable to audio–visual asynchrony. The gray-scale level of the image was also manipulated. Performance accuracy significantly improved when the FRs were increased from 8.3 to 25 Hz and also from 8.3 to 16.7 Hz. There was no significant gray scale and FR interaction, which suggests that both may be independent, and yet additive, sources of visual cues in lip-reading ability. In a second experiment, Vitkovitch and Barber sought to assess whether repeated exposures to stimuli of low FR and limited gray scales can aid the participants' lip-reading ability. Both digital and analog formats as well as gray-scale level were compared. There was no improvement in accuracy over time for those in the low FR (8.3 Hz). This may indicate that the high perceptual demand of the low FR was too great, and thus, increased experience was not enough to allow participants to compensate for the visual degradation. The main effects of both FR and gl indicate that performance was higher at 25 Hz than at 8.3 Hz as well as 16 gl compared to 8 gl. The results of both experiments show that FRs higher than 8.3 Hz are necessary for lip reading, with 16.7 Hz being adequate.

Williams *et al.* [56] investigated the effects of FR and the recognition of visemes, or visual phonemes [57], which are the visual speech cues such as mouth movements and teeth positions that accompany speech sounds. Generally, viseme recognition decreased along with decreasing FR. Often, viseme recognition dramatically dropped when the FR was reduced from 5 to 2 Hz.

The visual distortion caused by low FR can result in an asynchrony with the audio components of the presentation. Knoche *et al.* [44] manipulated both FR and audio–visual skew upon the ability of the participants to identify the middle consonants of nonsense words. Results showed that lower FRs worsen the effects of negative audio skew (in which audio leads visual). Conversely, positive skews (in which visual leads audio) can improve consonant comprehension. At FRs between 10 and 15 Hz, audio lags between $+127$ and $+167$ ms can improve comprehension beyond that which normally occurs at 30 Hz. This may be because longer frame times allowed participants to "close the gap" between audio and visual stimuli. Additionally, the authors suspect that there is a ± 120-ms limit on the ability for humans to integrate audio–visual information (any more or less suggests great asynchrony and possible failed integration). Whereas audio and video can be perceived in synch with each other anywhere between ± 120 ms for 30 Hz, this window changes to -54 to $+186$ ms at 10 Hz. The implications of these results are that in a "noisy" environment, a video at 10 Hz will be effective if audio lags visual by $+120$ to $+170$ ms.

Constant FR investigations have also been applied to dynamic medical displays, such as X-ray fluoroscopy, cardiac cineangiography, real-time 2-D ultrasound, rapid-sequence nuclear magnetic resonance imaging, radioiosotope ventriculography, and ultrafast computer tomography. There is also a degree of "noise" and contrast involved in these displays that can further complicate the detection of low contrast signals. Whiting *et al.* [38] examined the effects of FR on signal detection in a radiological application. Higher FRs aided signal detection performance, specifically as the FR progressed from 0 Hz (static image) to approximately 4 Hz. It then improved modestly until the FR increased 16 Hz, then mildly improved at higher FRs thereafter. Although 100–200 ms was previously suggested as the maximum persistence of visual memory, the results of the current study suggest a 1500-ms decay time in this task.

2) Variable FR Effects on Target Recognition: Variable FR configurations have also been tested in target recognition tasks. Another approach to the target detection environment is the use of a variable FR system. The velocity-based LOD-rendering technique is one such approach to determining FR. It can yield faster graphics-rendering times when a high LOD is not necessary, and this in turn can improve the visual search performance [29].

Parkhurst and Niebur [29] employed the velocity-based LOD graphics-rendering technique to present target objects in a virtual room. FRs were either held constant at 6 Hz or variable so that they were allowed to increase with decreasing LOD as the observer turned his or her head to search for targets. Target detection accuracy did not change because of LOD changes. Reaction times increased with decreasing LOD in the constant FR condition, but in the variable FR condition, reaction times decreased when the LOD decreased, likely because the extra bandwidth was allocated toward generating a higher FR. Overall, there was support for the velocity-based LOD-rendering technique, since faster rendering times associated with decreased LOD can aid visual search performance.

To summarize the research on FRs and target recognition, continuous FR presentation in a purely visual signal detection task demonstrates an apparent 16-Hz threshold for adequate performance. Increasing the FR beyond this level improved the performance only mildly [38].

In the context of military target recognition, French *et al.* [3] reported that FRs as slow as 2 Hz did not significantly affect the robotic operators' target identification performance (as compared to conditions wherein the FRs were 4, 8, and 16 Hz). Similarly, Chen *et al.* [52] did not observe significant target detection performance degradation as the FRs dropped from 30 to 5 Hz. The minimal satisfactory FR for tracking a moving human figure (e.g., as in surveillance video monitoring) has been found to be as low as 1 Hz [19].

With regard to speech reading, Frowein *et al.* [43] concluded that the FR is a critical contributor, and that FRs up to 15 Hz can improve this performance. Comparably, Vitkovitch and Barber [39], [55] concluded that 16.7 Hz may be adequate, and lower rates such as 8.3–12.5 Hz may result in a great loss of dynamic visual information and audio–visual mismatch, thereby causing major performance decrements. Conversely, Williams *et al.* [56] considered the low rate of 5 Hz to be the minimum threshold for continuous speech reading. Another important aspect of speech reading is the potential for audio accompaniment, which can be skewed at lower visual FRs. For instance, at 10 Hz, the audio presentation should lag the visual presentation by at least $+40$ ms [44].

Finally, in sign language tasks, observers may be able to adapt to low FRs. If there is enough opportunity to learn these gestures, low multimedia rates of 1 and 5 Hz may be adequate [22].

B. Perception and Psychophysics

The effects of slow FRs on human perception have been examined in a wide variety of environments. For example, in a study conducted in a VE, Piantanida *et al.* (as cited in [21]) found that the participants' depth and egomotion perception degraded when the FRs dropped. On the other hand, Van Erp and Padmos [4] assessed the effects of slow update rates of the viewing system for indirect driving and observed that an update rate of as low as 3 Hz did not appear to affect the participants' speed perception.

Augmented reality (AR), in which computer graphics are overlaid onto the physical environment through transparent HMDs, has also been used to study the effects of slow FRs. Effective AR depends in part on the mechanisms of human visual perception, such as depth cues and motion detection. Lai and Duh [58] assessed the role of different FRs on the perception of visualization dynamic information via the height change in bar graphs presented in an HMD. Users can visualize changes in dynamic information displays by tracking the motion of the stimulus (e.g., fluctuating bar graph height). Users can also anticipate the future motion of the stimulus change as a function of the implied velocity of the stimulus (i.e., representational momentum; see [59]). Thus, the display's ability to convey motion can impact the perception of dynamic information representation. Participants judged the percentage change in a dynamic 3-D bar chart (e.g., the rate at which the height of a bar graph was increasing). Interestingly, the greatest judgment inaccuracies were made at the fastest FR, i.e., 160 Hz. It was found that participants overestimated the percentage at lower speeds and underestimated at higher speeds. This may be because in the slower speeds, participants use cognitive resistance (see [60]) in later stages of the presentation to reduce the degree of representational momentum. It is also important to consider the typical human cognitive processing speed of 170 ms [23], since changes that occur below this threshold may not be accurately perceived. This may explain the performance improvements for the longer frame times (i.e., slower FRs).

Mastoropoulou and Chalmers [41] hypothesized that music could be used as a distractor to affect the user's perception of time enough to downplay slower rendering times. Participants with computer graphics experience were tested. Animated clips were compared at different FRs, and each compared pair had a difference of 4 Hz (16 versus 20 Hz or 12 versus 16 Hz). The two types of music that accompanied the animations were "slow tempo and relaxing" and "fast tempo and exciting." Frequency analyses showed that slow tempo music decreased the perceived scene velocity. Contrary to the hypothesis, playing fast tempo music did not convince observers that the clip was shorter and progressed at a higher temporal rate. The lack of support for the hypothesis could be because a difference of 4 Hz was not enough for participants to perceive a velocity difference when the two different FRs were compared.

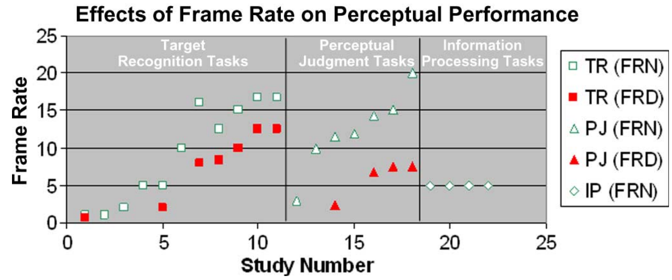

Fig. 2. Summary of FR effects on perceptual performance. Note: The x-axis denotes the number the study was designated. The list of studies with their corresponding numbers is provided in the Table II. Filled red squares and triangles signify the FR levels where human performance starts to significantly degrade. Unfilled green squares, triangles, and circles signify the lowest FR levels where participants perform normally.

Pfautz [36, Experiment E] was interested in the impact of spatiotemporal sampling on the perception of constant-velocity motion. There was an interaction between FR and stimulus velocity, so that higher velocities at higher FRs (e.g., 15 and 60 Hz versus 5 and 7.5 Hz) yielded more accurate response distances. In a second related experiment, Pfautz [36, Experiment F] assessed the effects of viewpoint manipulation in an air traffic control collision task. The task involved making a time-to-contact judgment for two moving aircraft on a screen. The higher FR, i.e., 20 Hz, provided greater judgment accuracy than 7.5 Hz. However, this accuracy was impaired when participants were allowed to alter their viewpoint to suit their preferences.

Reddy [21, Experiment 1] assessed the performance in a virtual heading task while manipulating FR at either 2.3 Hz (considered excessively low) or 11.5 Hz (considered to be moderately acceptable). In the heading task, the subject was passively navigated through the VE with his/her direction of fixation oriented differently from the direction of movement, and the task was to determine whether she/he was moving toward the left or right of the fixation point. The higher FR of 11.5 Hz had clear performance benefits in terms of both accuracy and task completion time. The same task was again executed in a second experiment (see [21, Experiment 2]. Higher FRs were manipulated in this experiment, but this time using 6.7 and 14.2 Hz. As in the first experiment, similar trends were found, such that accuracy and reaction time were better in the higher FR condition. More precise angle discriminations were also possible at this higher FR.

In summary, Van Erp and Padmos [4] assessed the effects of slow update rates of the viewing system for indirect driving and observed that an update rate as low as 3 Hz did not appear to affect the participants' speed perception. Piantanida *et al.* (as cited in [21]), on the other hand, found that the participants' depth and egomotion perception degraded when the FRs dropped.

The perception of constant-velocity motion seems to be best served by higher FRs such as 15, 20, and 60 Hz [36, Experiments E and F]. However, there is an upper limit to the ideal FR. Excessively rapid FRs (e.g., 160 Hz) can impair the apparent rate of change in dynamic informational displays [58]. On the other hand, differences of 4 Hz at modest FRs (e.g.,

TABLE II
SUMMARY OF FR EFFECTS ON PERCEPTUAL PERFORMANCE

#	Author(s) and Ref. #	# of Sub	FR Manipulations	Task Type	Measures
1	Cai [19]	NR	Range from ***0.5***, **1**, to 17 Hz	Tracking moving human figure	Accuracy
2	Johnson & Caird [22]	48	**1**, 5, 15, or 30 Hz	Target recognition & Info processing	Learning
3	French et al. [3]	24	**2**, 4, 8, and 16 Hz	Target identification	Accuracy
4	Chen et al. [52]	15	**5** and 30 Hz	Target detection	Accuracy
5	Williams et al. [56]	30	***2***, **5**, 10, 15, and 30 Hz	Target recognition-speech reading	Accuracy
6	Knoche et al. [44]	20	**10**, 15, 30 Hz	Target recognition	Accuracy
7	Whiting et al. [38]	5	0, 2, 4, ***8***, **16**, & 32 Hz	Target detection	Accuracy
8	Vitkovitch & Barber [55, Exp.2]	60	***8.3***, **12.5**, and 25 Hz	Target recognition-lip reading	Accuracy
9	Frowein et al. [43]	16	5, 6, 7.5, ***10***, **15**, and 30 Hz	Target recognition & Info processing	Verbal information recall
10	Vitkovitch & Barber [55, Exp.1]	16	8.3, ***12.5***, **16.7**, 25 Hz	Target recognition-lip reading	Accuracy
11	Vitkovitch & Barber [39]	52	8.3, ***12.5***, **16.7**, 25 Hz	Target recognition-lip reading	Accuracy
12	Van Erp & Padmos [4]	8	**3**, 5, 10, 30 Hz	Perception-Speed & distance estimation	Accuracy
13	Lai & Duh [58]	24	**10**, 40, 80, 120, 160 Hz.	Judgment of visual percentage change	Accuracy
14	Reddy [21, Exp. 1]	20	***2.3*** Hz and **11.5** Hz	Directional (orientation) judgment	Accuracy & reaction time in heading task
15	Mastoropoulou & Chalmers [41]	48	16 vs. 20 Hz or **12** vs. 16 Hz	Viewing animated clips	Temporal perception
16	Reddy [21, Exp. 2]	5	***6.7*** Hz and **14.2** Hz	Directional (orientation) judgment	Accuracy & reaction time in heading task
17	Pfautz [36, Exp. E]	6	5, ***7.5***, **15**, 60 Hz	Time-to-contact judgment	Accuracy
18	Pfautz [36, Exp. F]	4	***7.5*** Hz vs. **20** Hz	Time-to-contact judgment	Accuracy
19	Ghinea & Thomas [76]	NR	**5**, 15, 25 Hz	Info processing	Info assimilation
20	Gulliver & Ghinea [62]	36	**5**, 15, and 25 Hz	Info processing	Eye path movement
21	Gulliver & Ghinea [70]-[72]	36	**5**, 15, and 25 Hz	Info processing	Info assimilation & eye path movement
22	Gulliver, et al. [73]	48	**5**, 15, 25 Hz	Info processing	Info assimilation
	Parkhurst & Niebur [29]	6	Constant at 6 Hz or varied (from 9 to 31.5 Hz)	Target recognition	Accuracy & reaction time

in the range of 12 to 20 Hz) may be undetected by viewers, even if they are experienced with computer graphics [41].

In heading performance, there is a strong performance benefit when the FR increases to about 10–15 Hz; however, performance improvements taper off afterward at higher rates. Thus, it seems that 15 Hz might be considered an approximate minimum threshold for VEs. FRs below 10 Hz will yield sharp performance degradations in terms of both response time and heading accuracy. Heading tasks have many parallels to other VE applications such as driving and flight simulators, and so the results can extend to these domains as well [21].

A summary of these research findings is presented graphically in Fig. 2 and is also provided in more detail in Table II.

V. BEHAVIORAL EFFECTS

Eye movements are considered important in understanding visual perception, attention, and cognitive processes. The eye will focus on particularly informative areas of a display [58]; thus, poor visual quality may in fact fail to attract eye fixation. Quality of service (QoS) refers to the goodness or degree to which the multimedia presentation is considered tolerable in terms of viewing experience and is partially determined by FR, which in turn is attributable to bandwidth limitation implications.

Gulliver and Ghinea [62] investigated changes in eye movements as influenced by low FRs. It was hypothesized that jerkier and disjointed video presentations from low FRs would be more perceptually intolerable and annoying, thereby causing more variability in eye paths. Interestingly, no correlation was found between FR and eye path movements. Thus, Gulliver and Ghinea suggest that lower FRs (e.g., 5 Hz) do not influence the flow of informational assimilation.

It has also been shown that individuals are able to conduct video-mediated conversations with the use of hand motions at FRs as low as 1 Hz [63] and one frame change every 5 s, although subjective measures indicated that participants thought this extremely low FR was ineffective [64]. Thus, despite noticeably low temporal quality, these types of visual cues can retain their effectiveness.

Extremely slow FRs and their associated visual lag may cause cue conflict (i.e., discrepancy between visual and vestibular systems), which has been shown to induce cyber/simulator sickness [16], [65]. Lower FRs can also impair feelings of presence, cause balance loss, and increase heart rate [48]. Meehan *et al.* [48] assessed the placement performance in VEs and found that the 10-Hz condition caused many participants to lose a sense of balance, causing anomalous dependent measure scores that consequently required readjustment. The change in heart rate was significantly higher in the 10 Hz than in the 15-Hz condition, which indicates increased stress at the lower FR. Additionally, the reported behavioral presence significantly increased as the FR increased from 15 to 20 Hz.

VI. SUBJECTIVE ASSESSMENT

Varying FRs may influence the subjective ratings of video quality and viewing enjoyment. This may be of importance to service providers who need to know the minimum level of quality that their customers find acceptable, as well as viewers who need to be able to recognize the subject matter and observe smoother movements [66].

QoS involves spatial (intraframe) characteristics (e.g., picture resolution and color depth), temporal (interframe) characteristics (e.g., FR), and spatiotemporal integration (e.g., as auditory effects that are in synchrony with the visual cues)

[67]. It is related to the image quality/FR tradeoff in bandwidth allocation, in which FRs can be reduced in order to improve the spatial aspects of the video. According to Steinmetz [68], people generally perceive the video and the audio as "in sync" if the "skews" (i.e., time differences between the video and audio output) are under 80 ms. Therefore, a video with a 12.5-Hz or higher FR should appear to be in sync with the audio.

"Watchability" is a characteristic of the video that includes the viewer's acceptance of the video's audio signals, the continuity of visual messages, lip movement, speech synchronization, and the relationship between the audio and visual message components [67]. Lowering FRs has been shown to reduce the watchability of video clips, such that 15 Hz might be just acceptable, 10 Hz is much less acceptable, and 5 Hz is very unacceptable [67]. High-action videos (e.g., sports clips) on a desktop display with a resolution of 352 × 288 pixels have yielded similar satisfaction ratings across FRs of 25, 15, and 5 Hz [69], and in fact have been considered acceptable 80% of the time when FRs were only 6 Hz [66]. When viewed on a palmtop display (sized 176 × 144), a low FR was found to be less acceptable than on a desktop display, such that the critical lowest acceptable FR value was 12 Hz, and 6 Hz was considered acceptable only 50% of the time.

Gulliver and Ghinea [70]–[72] found that the degradation of perceived video quality and satisfaction was not noticeable when the FR was reduced from 25 to 15 Hz, but it was noticeable when the FR was reduced to 5 Hz. Thus, presentations made at 15 and 25 Hz appear perceptually comparable to the participants. Similarly, Claypool *et al.* [46] showed that FPS video game players did not perceive the quality of the pictures as significantly lower when the FR was reduced to 15 Hz from 30 or 60 Hz. Other studies, on the other hand, have found that reductions in FR (i.e., from 25 to 15 and then to 5 Hz) often do not impair viewing enjoyment; however, there is a reduction in perceived quality [73]. Some techniques have been proposed to enhance the perceived quality for lower-FR video. For example, by using a variable FR streaming algorithm to transmit hand movements without delay in a video-mediated conversation, it was shown that an FR reduction from 30 to 1 frame changes every 5 s had minimal effect on engagement and enjoyment [64].

The relationship between FR and perceived quality may however be context dependent. Masry and Hemami [74] found that the perceived quality was higher for lower FRs (such as 10 and 15 Hz) in higher-action sequences, but it was found unacceptable in clips involving less motion; this indicates a possible preference specific to motion types. In a similar vein, video clip subject matter may result in different levels of perceptual quality and user satisfaction, independent of FR [70]. For example, across all FRs (5, 15, and 25 Hz), pop music videos had higher subjective quality and satisfaction ratings than a chorus video.

Quality of perception (QoP) is the perceptual experience of the user, including the enjoyment of watching the video and the ability to analyze, synthesize, and assimilate the information presented (see [62], [69]–[72], and [75]). Thus, QoP involves the viewer's information processing. The multimedia can often be greatly degraded before perception, information

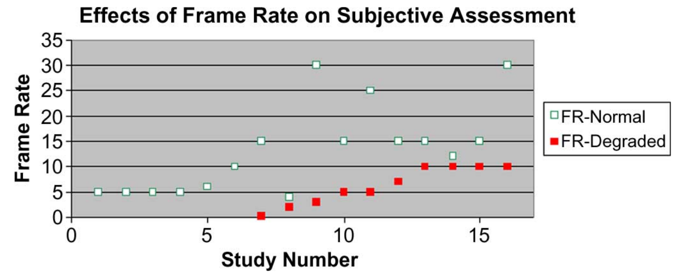

Fig. 3. Summary of FR effects on subjective assessment. Note: The x-axis denotes the number the study was designated. The list of studies with their corresponding numbers is provided in the Table III. Filled red squares signify the FR levels where participants perceive as degraded. Unfilled green squares signify the lowest FR levels where participants perceive as normal.

assimilation, and learning are impaired. Ghinea and Thomas [76] demonstrated that viewers could integrate visual details to learn and then correctly answer questions about the content of videos even as the FR was reduced from 25 to 5 Hz. Often, lower FRs (5 Hz) resulted in more correct answers about what had been viewed, possibly because there was more time to view frames before they changed (i.e., at 25 Hz, each frame is visible for 40 ms, but at 5 Hz, a frame is visible for 200 ms). The robustness of information assimilation despite low FR has been documented in other studies (see [62], [69]–[73], [75], and [76]).

According to Ghinea and Chen [75], cognitive styles such as field dependence (i.e., degree to which a person's perception of information is influenced by the contextual field) and field independence were not moderators in the relationship between FR and QoP. Furthermore, eye-path movement, which is associated with attentional and cognitive processes, did not differ across FRs of 5, 15, or 25 Hz [72].

van Breda *et al.* [77] reported that UAV operators assessed their effort in viewing the simulated sensor feed as higher when the FR for the camera image was low (i.e., 3 Hz) compared to normal FRs. Wilson and Sasse [78] have also shown that low FRs can cause the viewers physiological strain despite the fact that most of them did not report noticing a difference between videos presented at 5 and 25 Hz. Exposure to the 5-Hz condition resulted in greater experiences of stress, which is characterized by increased galvanic skin response, increased heart rate, and decreased blood volume pulse. The results of this study suggest that it may not be sufficient to measure subject reactions alone, since physiological effects can also occur. This partially corroborates the findings of Meehan *et al.* [48], in which 10 Hz caused large increases in heart rate compared to higher FRs.

A summary of these research findings is presented graphically in Fig. 3 and is also provided in more detail in Table III.

VII. Conclusion

In this paper, we have reviewed more than 50 studies and summarized them in the areas of psychomotor performance, perceptual performance, behavioral effects, and subjective perception. Generally, psychomotor performance improves at higher FRs and lower SDSR. In placement tasks, 4 Hz may be too low for VE performance, and 10 Hz may

TABLE III
SUMMARY OF FR EFFECTS ON SUBJECTIVE ASSESSMENT

#	Author(s) and Ref. #	# of Sub	FR Manipulations	Task Type	Measures
1	Ghinea & Chen [75]	132	**5**, 15, 25 Hz	Info assimilation	QoP
2	Ghinea & Thomas [69]	30	**5**, 15, 25 Hz	Info assimilation	QoP
3	Gulliver et al. [73]	48	**5**, 15, 25 Hz	Info assimilation	QoP
4	Wilson & Sasse [78]	24	**5** and 25 Hz	Viewing video	Subjective quality rating
5	McCarthy et al. [66, Exp. 1]	41	**6**, 10, 12, 15, 18, 20, 24 Hz	Viewing video on desktop display	Subjective visual acceptability & Proportion of eye movements
6	Masry & Hemami [74]	19	**10**, 30 Hz	Viewing motion clips (higher action sequences)	Perceived quality
7	Chen [64]	32	***0.2*** & **15** Hz, variable	Conversational behavior	Perceived engagement/ enjoyment
8	French et al. [3]	24	***2***, **4**, 8, and 16 Hz	Teleoperating ground robots	Perceived workload
9	Van Breda et al. [77]	NR	***3***, **10- 30** Hz	Viewing 2-D vs. 3-D maps	Subjective effort rating
10	Gulliver & Ghinea [70]-[72]	36	***5***, **15**, and 25 Hz	Info assimilation	QoP
11	Wilson & Sasse [78]	24	***5*** and **25** Hz	Viewing video	Physiological measure (stress)
12	Claypool et al. [46]	60	3, 7, 15, 30, and 60 Hz	First-person shooter video game	Perceived quality of video
13	Apeteker et al. [67]	60	5, ***10***, **15** Hz	Viewing video clips	Watchability & QoS
14	McCarthy et al. [66, Exp. 2]	37	6, ***10***, **12**, 15, 18, 20, 24 Hz	Viewing video on palmtop display	Subjective visual acceptability
15	Meehan et al. [48]	33	***10***, **15**, 20, and 30 Hz	Placement task	Physiological & Subjective presence
16	Masry & Hemami [74]	19	***10***, 30 Hz	Viewing motion clips (lower motion clips)	Perceived quality

disrupt human balance. Experimental results have suggested a minimum threshold of 17.5 Hz for successful placement performance [10]. Tracing performance may also require more than 10 Hz.

For teleoperation, higher FRs such as 16 Hz are suggested to aid in navigation and target tracking. A rate of approximately 10 Hz may suffice for HMD-facilitated tracking. This lower required FR will be useful for freeing up bandwidth for other uses, such as rendering more graphically detailed images.

Speech reading is an often-studied form of target recognition. Generally, 13 Hz may be too low for speech reading from video transmission, rather at least 15 Hz may be necessary. However, with appropriate learning opportunity, rates around 5 Hz (and possibly less) may be sufficient. As with video speech reading, target detection in a visual radiological task may be significantly more successful at 16 Hz. Perceptual VE task performances seem to be best served by FRs above 15 Hz. This includes perceiving motion of a constant velocity and heading performance.

Individuals seem to be able to gather content-based information about videos that are viewed at low FRs such as 5 Hz. While this FR is very low, it may benefit information assimilation because each frame remains on the screen for a relatively longer duration compared to frames that are presented at higher rates. This provides viewers with more time to observe each frame. Eye movement studies support these findings; the lack of effect for FR upon eye movements indicates there are no significant changes in attentional and cognitive processes.

In terms of watchability and satisfaction, FRs of 5 Hz can appear "unacceptable" and of lower visual quality. Still, they are often considered to be enjoyable to watch despite the temporal distortion. However, FR reductions from 25 to 15 Hz are often perceptually similar in appearance to viewers.

Also, FRs at 10 Hz and below have been shown to cause stress in terms of physiological reactions in addition to general performance decrements.

Overall, there seems to be strong support for a threshold of around 15 Hz for many tasks, including those that are psychomotor and perceptual in nature. Only Lion [37] and Watson *et al.* [7, Experiment 2] (out of the 56 studies reviewed) showed that significant performance impairment persisted at FRs above 15 Hz. Less impressive yet acceptable performance may be accomplished at around 10 Hz for many tasks. However, it is important to consider that individuals may be more susceptible to stress symptoms at FRs at and below 10 Hz. Information assimilation seems to be far more impervious to the detrimental effects of FR, since participants can observe and retain much of the information from clips presented at rates as low as 5 Hz. Subjective reactions to quality and watchability of videos seem to support rates of 5 Hz, although videos presented at 15 Hz and above are generally more widely preferred.

These generalizations regarding superior and acceptable FRs may also be subject to the effects of several moderators. The majority of the studies reviewed have involved other independent variables and equipment characteristics that could be contributing to or detracting from the proposed effects of various FRs.

The use of variable FR manipulation has been recommended over constant FRs because of its proposed accuracy advantage [10]. There are many different ways to determine the SD in laboratories; however, some techniques may be more accurate in terms of how FR truly varies in real-world systems (e.g., SD based on percentages of mean FRs).

Stereoscopic displays will likely enable a performance advantage over monoscopic displays, and when possible, they should be used to assist operators in compensating for lower FR presentations. If the displays are stereoscopic, FRs as low as 7 Hz may be adequate, but monoscopic displays may only be useful for FRs as low as 14 Hz [17].

User experience may also be an important moderator. Individuals with HMD use experience were able to tolerate FRs as low as 10 Hz in tracking, whereas inexperienced participants needed 30 Hz [15]. This difference in tolerable FR is quite large, and so it is clear that the demographics should be measured and controlled when the temporal distortion is manipulated.

REFERENCES

[1] J. Y. C. Chen, E. C. Haas, K. Pillalamarri, and C. N. Jacobson, "Human robot interface: Issues in operator performance, interface design, and technologies," U.S. Army Res. Lab., Aberdeen Proving Ground, Aberdeen, MD, Tech. Rep. ARL-TR-3834, 2006.

[2] T. Oron-Gilad, J. Y. C. Chen, and P. A. Hancock, "Remotely operated vehicles (ROVs) from the top-down and the bottom-up," in *The Human Factors of Remotely Operated Vehicles*, N. Cooke, H. Pringle, H. Pedersen, and O. Connor, Eds. New York: Elsevier, 2005.

[3] J. French, T. G. Ghirardelli, and J. Swoboda, "The effect of bandwidth on operator control of an unmanned ground vehicle," in *Proc. I/ITSEC*, Orlando, FL, 2003.

[4] J. B. F. Van Erp and P. Padmos, "Image parameters for driving with indirect viewing systems," *Ergonomics*, vol. 46, no. 15, pp. 1471–1499, Dec. 2003.

[5] A. Rastogi, "Design of an interface for teleoperation in unstructured environments using augmented reality displays," M.S. thesis, Univ. Toronto, Toronto, ON, Canada, 1996.

[6] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press, 1992.

[7] B. Watson, N. Walker, W. Ribarsky, and V. Spaulding, "Effects of variation in system responsiveness on user performance in virtual environments," *Hum. Factors*, vol. 40, no. 3, pp. 403–414, 1998.

[8] B. Watson, N. Walker, P. Woytiuk, and W. Ribarsky, "Maintaining usability during 3-D placement despite delay," in *Proc. IEEE VR Conf.*, 2003, pp. 133–140.

[9] S. Bryson, "Effects of lag and frame rate on various tracking tasks," in *Proc. SPIE, Stereoscopic Displays Appl. IV*, Bellingham, WA, 1993, vol. 1925, pp. 155–166.

[10] B. Watson, V. Spaulding, N. Walker, and W. Ribarsky, "Evaluation of the effects of frame rate variation on VR task performance," in *Proc. IEEE Virtual Reality Annu. Symp.*, 1997, pp. 38–44.

[11] G. Tharp, A. Liu, L. French, S. Lai, and L. Stark, "Timing considerations of helmet mounted display performance," in *Proc. SPIE Conf. Human Vis., Visual Process., Digital Displays*, San Jose, CA, 1992, vol. 1666, pp. 570–576.

[12] C. Ware and R. Balakrishnan, "Reaching for objects in VR displays: Lag and frame rate," *ACM Trans. Comput.-Hum. Interact.*, vol. 1, no. 4, pp. 331–357, Dec. 1994.

[13] S. R. Ellis, B. D. Adelstein, S. Baumeler, G. J. Jense, and R. H. Jacoby, "Sensor spatial distortion, visual latency, and update rate effects on 3-D tracking in virtual environments," in *Proc. VR*, Houston, TX, 1999, pp. 218–221.

[14] K. W. Arthur, K. S. Booth, and C. Ware, "Evaluating 3-D task performance for fish tank virtual worlds," *ACM Trans. Inf. Syst.*, vol. 11, no. 3, pp. 239–265, Jul. 1993.

[15] A. Liu, G. Tharp, L. French, S. Lai, and L. Stark, "Some of what one needs to know about using head-mounted displays to improve teleoperator performance," *IEEE Trans. Robot. Autom.*, vol. 9, no. 5, pp. 638–648, Oct. 1993.

[16] E. M. Kolasinski, "Simulator sickness in virtual environments," U.S. Army Res. Inst. Behav. and Soc. Sci., Alexandria, VA, Tech. Rep. 1027, 1995.

[17] P. Richard, G. Birebent, P. Coiffet, G. Burdea, D. Gomez, and N. Langrana, "Effect of frame rate and force feedback on virtual object manipulation," *Presence*, vol. 5, no. 1, pp. 95–108, 1996.

[18] R. Pausch, T. Crea, and M. Conway, "A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness," *Presence*, vol. 1, no. 3, pp. 344–363, 1992.

[19] Y. Cai, "Minimalism context-aware displays," *Cyberpsychol. Behav.*, vol. 7, no. 6, pp. 635–644, 2004.

[20] A. L. Yarbus, *Eye Movements During Perception of Complex Objects*. New York: Plenum, 1967.

[21] M. Reddy, "The effects of low frame rate on a measure for user performance in virtual environments," Dept. Comput. Sci., Univ. Edinburgh, Edinburgh, U.K., Tech. Rep. ECS-CSG-36-97, 1997.

[22] B. F. Johnson and J. K. Caird, "The effect of frame rate and video information redundancy on the perceptual learning of American sign language gestures," in *Proc. Conf. Human Factors Comput. Syst.*, Vancouver, BC, Canada, 1996, pp. 121–122.

[23] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human–Computer Interaction*. Hillsdale, NJ: Erlbaum, 1983.

[24] I. S. MacKenzie and C. Ware, "Lag as a determinant of human performance in interactive systems," in *Proc. INTERCHI Conf. Human Factors Comput. Syst.*, 1993, pp. 488–493.

[25] M. D. McKenna and D. Zeltzer, "Three dimensional visual display systems for virtual environments," *Presence*, vol. 1, no. 4, pp. 421–458, 1992.

[26] J. M. Airey, J. H. Rohlf, and F. P. Brooks, "Towards image realism with interactive update rates in complex virtual building environments," in *Proc. Symp. Interactive 3-D Graph.*, 1990, pp. 41–50.

[27] R. Pausch, "Virtual reality on five dollars a day," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 1991, pp. 265–269.

[28] S. K. Card, G. G. Robertson, and J. D. MacKinlay, "The information visualizer, an information workspace," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 1991, pp. 181–188.

[29] D. Parkhurst and E. Niebur, *Evaluating velocity-based level of detail rendering of virtual environments using visual search*, 2001. [Online]. Available: http://cnslab.mb.jhu.edu/pubs/Parkhurst_Niebur01b.pdf

[30] V. Virsu and J. Rovamo, "Visual resolution, contrast sensitivity, and the cortical magnification factor," *Exp. Brain Res.*, vol. 37, no. 3, pp. 475–494, Nov. 1979.

[31] B. J. Murphy, "Pattern thresholds for moving and stationary gratings," *Vis. Res.*, vol. 18, no. 5, pp. 521–530, 1978.

[32] D. C. Burr and J. Ross, "Contrast sensitivity at high velocities," *Vis. Res.*, vol. 22, no. 4, pp. 479–484, 1982.

[33] D. J. Parkhurst and E. Niebur, "Variable-resolution displays: A theoretical, practical, and behavioral evaluation," *Hum. Factors*, vol. 44, no. 4, pp. 611–629, 2002.

[34] E. M. Reingold, L. C. Loschky, G. W. McConkie, and D. M. Stampe, "Gaze-contingent multiresolutional displays: An integrative review," *Hum. Factors*, vol. 45, no. 2, pp. 307–328, 2003.

[35] D. J. Parkhurst, E. Culurciello, and E. Niebur, "Evaluating variable resolution displays with visual search: Task performance and eye movements," in *Proc. ACM Eye Tracking Res. Appl. Symp.*, 2000, pp. 105–109.

[36] J. D. Pfautz, "Depth perception in computer graphics," Univ. Cambridge Comput. Lab., Cambridge, U.K., Tech. Rep. UCAM-CL-TR-546, 2002.

[37] D. M. Lion, "Three dimensional manual tracking using a head-tracked stereoscopic display," Hum. Interface Technol. Lab., Seattle, WA, Tech. Rep. TH-93-7, 1993.

[38] J. Whiting, D. Honig, E. Carterette, and N. Eigler, "Observer performance in dynamic displays: Effect of frame rate on visual signal detection in noisy images," in *Proc. SPIE: Human Vis., Visual Process., Digital Display II*, 1991, vol. 1453, pp. 165–176.

[39] M. Vitkovitch and P. Barber, "Effect of video frame rate on subjects' ability to shadow one of two competing verbal passages," *J. Speech Hear. Res.*, vol. 37, no. 5, pp. 1204–1211, Oct. 1994.

[40] R. B. Welch, L. D. Duttonhurt, and D. H. Warren, "Contributions of audition and vision to temporal rate perception," *Percept. Psychophysiol.*, vol. 39, no. 4, pp. 294–300, 1986.

[41] G. Mastoropoulou and A. Chalmers, "The effect of music on the perception of display rate and duration of animated sequences: An experimental study," in *Proc. Theory Practice Comput. Graph.*, 2004, pp. 128–134.

[42] G. Mastoropoulou, K. Debattista, A. Chalmers, and T. Troscianco, "The influence of sound effects on the perceived smoothness of rendered animations," in *Proc. 2nd Symp. Appl. Perception Graph. Vis.*, 2005, pp. 9–15.

[43] H. W. Frowein, G. F. Smoorenburg, L. Pyters, and D. Schinkel, "Improved speech recognition through videotelephony: Experiments with the hard of hearing," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 4, pp. 611–616, May 1991.

[44] H. Knoche, H. de Meer, and D. Kirsh, "Compensating for low frame rates," in *Proc. CHI*, Portland, OR, 2005, pp. 1553–1556. Late Breaking Results: Posters.

[45] P. Fitts, "The information capacity of the human motor systems in controlling the amplitude of movement," *J. Exp. Psychol.*, vol. 47, no. 6, pp. 381–391, 1954.

[46] M. Claypool, K. Claypool, and F. Damaa, "The effects of frame rate and resolution on users playing first person shooter games," in *Proc. ACM/SPIE MMCN Conf.*, San Jose, CA, 2006.

[47] M. J. Massimino and T. B. Sheridan, "Teleoperator performance with varying force and visual feedback," *Hum. Factors*, vol. 36, no. 1, pp. 145–157, Mar. 1994.

[48] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, "Physiological measures of presence in stressful virtual environments," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 645–652, Jul. 2002.

[49] D. E. McGovern, "Experience and results in teleoperation of land vehicles," in *Pictorial Communication in Virtual and Real Environments*, S. Ellis, M. Kaiser, and A. Grunwald, Eds. London, U.K.: Taylor & Francis, 1991, pp. 182–195.

[50] P. N. Day, "Cognitive effects of delayed visual feedback in real-time operator control loops," in "Unpublished First Year Report," Heriot Watt Univ., Edinburgh, U.K., 1999. [Online]. Available: http://www.macs.hw.ac.uk/isel/phil/pubs/pnd_phd_Year1Report.pdf

[51] D. R. Lampton, B. W. Knerr, S. L. Goldberg, J. P. Bliss, J. M. Moshell, and B. S. Blau, "The virtual environment performance assessment battery (VEPAB): Development and evaluation," *Presence: Teleop. Virtual Environ.*, vol. 3, no. 2, pp. 145–157, 1994.

[52] J. Y. C. Chen, P. J. Durlach, J. A. Sloan, and L. D. Bowens, "Robotic operator performance in simulated reconnaissance missions," U.S. Army Res. Lab., Aberdeen Proving Ground, Aberdeen, MD, ARL Tech. Rep. ARL-TR-3628, 2005.

[53] D. W. Massaro, J. Beskow, M. M. Cohen, C. L. Fry, and T. Rodriguez, "Picture my voice: Audio to visual speech synthesis using artificial neural networks," in *Proc. Int. Conf. AVSP*, Santa Cruz, CA, 1999, pp. 133–138.

[54] A. Q. Summerfield, "Use of visual information for phonetic perception," *Phonetica*, vol. 36, no. 4/5, pp. 314–331, 1979.

[55] M. Vitkovitch and P. Barber, "Visible speech as a function of image quality: Effects of display parameters on lipreading ability," *Appl. Cogn. Psychol.*, vol. 10, no. 2, pp. 121–140, 1996.

[56] J. J. Williams, J. C. Rutledge, D. C. Garstecki, and A. K. Katsaggelos, "Frame rate and viseme analysis for multimedia applications," *J. VLSI Signal Process. Syst.*, vol. 23, no. 1/2, pp. 7–23, 1997.

[57] C. Fisher, "Confusion among visually perceived consonants," *J. Speech Hear. Res.*, vol. 11, pp. 796–804, 1968.

[58] W. Y. Lai and H. B. L. Duh, "Effects of frame rate for visualization of dynamic quantitative information in a head-mounted display," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2004, pp. 6485–6490.

[59] J. J. Freyd and R. A. Finke, "A velocity effect for representational momentum," *Bull. Psychon. Soc.*, vol. 23, pp. 443–446, 1985.

[60] R. A. Finke and G. C. W. Shyi, "Mental extrapolation and representational momentum for complex implied motions," *J. Exp. Psychol.: Learn., Memory, Cogn.*, vol. 14, no. 1, pp. 112–120, Jan. 1988.

[61] J. B. Peiz, R. Canosa, and J. Babcock, "Extended tasks elicit complex eye movement patterns," in *Proc. Eye Tracking Res. Appl. Symp.*, Palm Beach Gardens, FL, 2000, pp. 37–43.

[62] S. R. Gulliver and G. Ghinea, "An eye opener: Low frame rates do not affect fixations," in *Proc. IEEE Conf. Multimedia Expo*, 2004, pp. 205–208.

[63] M. Chen, "Achieving floor control with a low-bandwidth gesture-sensitive videoconferencing system," in *Proc. ACM Multimedia*, 2002, pp. 476–483.

[64] M. Chen, "Conveying conversational cues through video," Ph.D. dissertation, Stanford Univ., Palo Alto, CA, 2003.

[65] K. M. Stanney, R. R. Mourant, and R. S. Kennedy, "Human factors issues in virtual environments: A review of the literature," *Presence*, vol. 7, no. 4, pp. 327–351, 1998.

[66] J. D. McCarthy, M. A. Sasse, and D. Miras, "Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamlined video," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Vienna, Austria, 2004, pp. 535–542.

[67] R. Apteker, J. Fisher, V. Kisimov, and H. Neishlos, "Video acceptability and frame rate," *IEEE Multimedia*, vol. 2, no. 3, pp. 32–40, Autumn/Fall 1995.

[68] R. Steinmetz, "Human perception of jitter and media synchronization," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 1, pp. 61–72, Jan. 1996.

[69] G. Ghinea and J. P. Thomas, "QoS impact on user perception and understanding of multimedia video clips," in *Proc. ACM Multimedia*, 1998, pp. 49–54.

[70] S. R. Gulliver and G. Ghinea, "Changing frame rate, changing satisfaction?" in *Proc. IEEE Multimedia Expo*, 2004, pp. 177–180.

[71] S. R. Gulliver and G. Ghinea, "Region of interest displays: Addressing a perceptual problem?" in *Proc. IEEE 6th Int. Symp. Multimedia Softw. Eng.*, Dec. 13–15, 2004, pp. 2–9.

[72] S. R. Gulliver and G. Ghinea, "Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 34, no. 4, pp. 472–482, Jul. 2004.

[73] S. R. Gulliver, T. Serif, and G. Ghinea, "Persuasive and standalone computing: The perceptual effects of variable multimedia quality," *Int. J. Hum.-Comput. Stud.*, vol. 60, pp. 640–665, 2004.

[74] M. Masry and S. S. Hemami, "An analysis of subjective quality in low bit rate video," in *Proc. ICIP*, 2001, pp. 465–468.

[75] G. Ghinea and S. Y. Chen, "The impact of cognitive styles on perceptual distributed multimedia quality," *Brit. J. Educ. Technol.*, vol. 34, no. 4, pp. 393–406, Sep. 2003.

[76] G. Ghinea and J. P. Thomas, "Information assimilation over the Internet: An initial study," in *Proc. ISECON*, 2000, pp. 17–22.

[77] L. van Breda, C. Jansen, and H. Veltman, "Supervising UAVs: Improving operator performance by optimizing the human factors," in *Proc. 11th Annu. HCI Int. Conf.*, Las Vegas, NV, 2005.

[78] G. M. Wilson and M. A. Sasse, "The head or the heart? Measuring the impact of media quality," in *Proc. CHI*, 2000, pp. 1–6.

Jessie Y. C. Chen received the B.A. degree in linguistics from the National Tsing-Hua University, Hsin-Chu, Taiwan, R.O.C., in 1987, the M.A. degree in communication studies from the University of Michigan, Ann Arbor, in 1989, and the Ph.D. degree in applied experimental and human factors psychology from the University of Central Florida, Orlando, in 2000.

Prior to joining the U.S. Army Research Laboratory (ARL), Orlando, she was a Post-Doctoral Fellow with the U.S. Army Research Institute for Behavioral and Social Sciences. She is currently a Research Psychologist with the ARL. Her research interests include human–robot interaction, robotics controller performance issues, and user interface designs for robotics control.

Jennifer E. Thropp received the B.S. degree in psychology, the M.S. degree in modeling and simulation, and the Ph.D. degree in applied/experimental human factors psychology from the University of Central Florida, Orlando, in 1998, 2004, and 2006, respectively.

She was a Post-Doctoral Researcher with the MURI Laboratory, University of Central Florida. She was also an Intern with the U.S. Army Research Laboratory, Orlando. She is currently a Human Factors Specialist with United Space Alliance, Kennedy Space Center, Cape Canaveral, FL. Her research interests include human factors in space operations, psychophysics, and human–automation interaction.

Dr. Thropp is a member of the Human Factors and Ergonomics Society.