D3

N82-13673

VIDEO FRAMERATE, RESOLUTION AND GRAYSCALE TRADEOFFS FOR UNDERSEA TELEMANIPULATOR CONTROL

Vivek Ranadive
Link-a-Bit Co.
San Diego, CA

Thomas B. Sheridan
MIT
Cambridge, MA

ABSTRACT

The high costs associated with a human diver working at ocean depths greater than 100m makes the use of remotely controlled visual inspection and manipulation attractive. However, long coaxial cables from the surface to the remote teleoperator cause additional difficulties. Therefore, sound communication without a tether is desireable. This form of telemetry, however, poses severe bandwidth restrictions so that its use for image transmission is in question.

The product of Frame Rate (F) in frames per second, Resolution (R) in total pixels and grayscale in bits (G) equals the transmission baud rate in bits per second. Thus for a fixed channel capacity there are tradeoffs between F, R and G in the actual sampling of the picture for a particular manual control task - in the present case remote undersea manipulation. A manipulator was used in the MASTER/SLAVE mode to study these tradeoffs. Images were systematically degraded from 28 frames per second, 128 x 128 pixels and 16 levels (4 bits) grayscale, with various FRG combinations constructed from a real-time digitized (charge-injection) video camera.[1]

When subjects first saw the video pictures with which they had to perform remote manipulation tasks, they refused to believe that they could succeed. Much to their surprise, however, they discovered that they were able to perform with a considerably degraded picture. It was found that frame rate, resolution and grayscale could be independently reduced without preventing the operator from accomplishing his/her task. Threshold points were found beyond which degradation would prevent any successful performance. It was observed that frame rate and grayscale could be degraded considerably more than resolution before teleoperation became impossible.

Isoperformance curves (curves of constant performance) were found for two subjects for various combinations of frame rate, resolution, and grayscale. These results were found to correlate closely with isotransmission curves (curves along which the information transmission rate is the same).

A general conclusion is that a well trained operator can perform familiar remote manipulator tasks with a considerably degraded picture, down to 50 K bits/sec, well below the several m bits/sec (5 MHz) normally used for broadcast video.

INTRODUCTION

Loss of life and costs approaching $5000 per working hour for deep ocean diving to do inspection and manipulation (as part of oil, gas, mineral and military operations) have motivated the development of remotely controlled vehicles having sensors and manipulators. Usually the operator observes through closed circuit video. (Sonic imaging is under development to provide a TV-like picture when water is too turbid for a video camera to work). Normally a high band-width coaxial cable is used as a communication link. However, such a cable can be extremely heavy for the submersible to drag around and can easily get caught in platform structure, rocks, etc.

Acoustic telemetry is an attractive alternative which avoids cable entanglement - at least for the lower portion of the signal path. Even though attenuation of sound in water is proportional to the fourth power of distance, several KHz for up to one Km distance are reasonable goals to send video pictures. The problem is: what kind of "real-time" image can be sent over a severely band-limited channel? More specifically, what are the best tradeoffs between frame rate, resolution and grayscale for an operator performing master-slave remote manipulation tasks?

SIMULATION OF UNDERSEA VIDEO IMAGING

Consider a remotely controlled submersible (teleoperator) with a video camera mounted on it. Figure 1. The bottleneck in this system is the information transmission capability from teleoperator to ship (or human controller).

A transmission channel of K bits/second could have

F frames per second

$R = \ell \times \ell$ pixels/frame resolution

G bits of grayscale/pixel = $\log_2$ (number of intensity levels) per pixel

Thus, the information transmission rate is

$$F \frac{\text{frames}}{\text{sec}} \times (\ell \times \ell) \frac{\text{pixels}}{\text{frame}} \times G \frac{\text{bits}}{\text{pixel}} = K \frac{\text{bits}}{\text{sec}}$$

Normal broadcast television has

$f = 30$ frames/sec

$\ell \times \ell \simeq 512 \times 512$ pixels/frame

assume $b = 4$ bits of gray

This means that normal broadcast television requires upwards of 30,000,000 bits/second. Current acoustic technology makes it possible to transmit from 30,000 bits/second in the shorter ranges to 3,000 bits/second for longer distances, up to 1 Km. It is, therefore, imperative for operators to learn to perform with coarse, slow pictures. Given the restricted bandwidth of operation, there must be compromises between the various features contributing to making up picture quality.

Common sense dictates that for some aspects of telemanipulation, framerate is most important while grayscale and resolution are not (e.g. moving a known object of good contrast while for other aspects high resolution with some grayscale is essential but framerate is not (e.g. identifying a fixed object). The literature provides little on this, although a recent literature review by Cole and Kishimoto was helpful.[2]

An experimental system called FRAG was developed to allow the experimenter to set the frame rate, resolution and grayscale of the video image. The system was a combination of commercially available components and special purpose (or home built) hardware.

The system used a General Electric TW2200 charge-injection camera. It contained not only the 16,384 pixels of the 128 x 128 camera, but all of the circuit logic necessary to perform a sequential raster scan and to generate synchronization signals. The CID arrays were fabricated as a silicon P.MOS device similar in many respects to some microprocessor and memory arrays.

The General Electric PN 2110 automation interface included analog or 8-bit digitized or thresholded binary video, power-clock signals, TTL signal level buffering and conversion and analog sweeps for CRT display presentation.

Variable framerate

FRAG was designed to continuously scan 28 frames/sec regardless of the selected sampling period since a slower speed integrated noise and saturated the picture. If the selected sampling speed was 28 frames/second then each frame could be displayed as it was sampled. For slower speeds, however, it was necessary to store frames in RAM and display each frame more than once. Thus, for a frame rate of 14 frames/sec, each frame would be displayed twice, and so on.

Variable Resolution

The GE interface had two registers of 128--x and y, corresponding to the horizontal and vertical scans. FRAG modified the counts from the x and y registers to result in the desired resolution.

Since the function of FRAG was to simulate as accurately as possible the effects of a low resolution picture transmission, it was decided to adopt a simple sampling scheme where every Nth pixel in both x and y was repeated N times. Thus, in Figure 2, a, b, c, and d would all be represented by A. This procedure required little real-time processing, and could be accomplished at fairly high speeds.

Grayscale

Four switches on FRAG produced 16, 8, 4, or 2 levels of gray by adjusting the coarseness of the digital-to-analog converters.

Example of tradeoffs

Using a cover picture of Life magazine, various combinations of resolution and grayscale are shown in Figure 3. For obvious reasons, frame rate cannot be shown in this manner.

Ranges of video adjustment

The FRAG system had the following ranges of operation:

Frame Rate: 0.109 to 28 frames/sec

Resolution: (8 x 8) to (128 x 128) pixels

Grayscale: 16 (at 2 levels) to 4 bits (16 levels)

Note that these were merely the available settings and were not necessarily usable by a human operator.

EXPERIMENT

Experimental Configuration

The teleoperator system including FRAG was set up as in Figure 4.

A modified Argonne E2 master-slave manipulator was used. It was attached to an Interdata computer which initialized the manipulator arms. This arm could move in all six degrees of freedom plus grasp. Although the manipulator was equipped with force-reflection, this feature was disabled so that the operator's only feedback came through the visual channel.

Experimental Tasks

The time required to complete a task under manual teleoperator control was expected to increase with the complexity of the task. The deterioration of the picture being used was expected to further increase the task completion time.

Tasks were selected to test manipulator performance based on the following criteria:

(i) task be representative of undersea manipulation. Such tasks include assessing damage, bolting/unbolting, connecting hoses, lifting objects, opening/closing valves, and reaching into confined spaces.

(ii) task performance be sensitive to required task accuracy.

In view of these criteria two tasks were designed for the teleoperator equipment

TAKE-OFF-NUT TASK (TON)

This task required the operator to locate a nut on a hub and then unscrew it. It was important not to drop the nut after removing it. The general method used by the subject was to grasp the nut, turn 180°, pull back to test if the nut was off, and if not release the grasp, reverse 180°, regrasp and repeat the operation. This task was representative of various "useful tasks" according to criteria (i).

OBJECT PLACEMENT TASK (123)

In accordance with criterion (ii) we chose a task which required fine positioning movements: to pick up a cork and place it sequentially in a preset sequence of three square areas on the table. Random 1-2-3 orderings were created so as to ensure the task being "closed loop", that is, to make visual feedback essential in moving between the three given squares in different trajectories.

Sound Feedback

Preliminary experiments showed that subjects made considerable use of sound feedback. For example, the sound of the manipulator colliding with the task hub became a valuable tool to determine position. In an ocean environment, such feedback probably would not be available. Accordingly, the lab airconditioner was turned on full to mask such feedback from the task.

Experimental subjects and their training

Two subjects were used as manipulator operators, both students in engineering. It was decided to compromise in the direction of well-trained subjects rather than use more subjects who were less well trained.

After gaining familiarity with the (force-reflecting) manipulator, force feedback was removed and the subjects were asked to do the TON task with direct vision. After some practice, the subjects were asked to perform the same task using a good (conventional) video picture. Finally, after the subjects were comfortable with this, the FRAG system and its accompanying degraded quality digitized picture was used instead of the conventional high quality video picture.

To insure that the subjects made definite progress during the training sessions, their performance was continuously monitored. The subjects had the best possible picture from the digitizer during the period. Figure 5 shows the learning curves for both subjects.

At the end of 10 intensive hours with FRAG, the two subjects' learning curves leveled in comparable fashion and the subjects were considered trained. The learning data were obtained on the basis of 10 trials for each data point value for each subject.

Experimental Protocol

The experiments were ordered so that two of the three variables (frame rate, resolution, and grayscale) were kept at a constant level while one third was varied.

As a performance baseline the best possible image conditions were used. All results were then compared with respect to this case. The best possible case had the conditions:

28 frames/sec frame rate

128 x 128 pixels resolution

4 bits grayscale

During the experiment, each subject was allowed to practice freely on each new image condition until "ready." Then twelve readings (i.e. time to accomplish tasks) on each were taken and the last six readings used as data. The TON task hub and 1-2-3 task paper were periodically reoriented to prevent the task from becoming rote.

RESULTS

Best Case

The average data for the best case was, surprisingly:

TASK	SUBJECT #1 TIME	SUBJECT #2 TIME
Take-off-Nut (TON)	28 seconds	28 seconds
Squares (1-2-3)	28 seconds	28 seconds

Relative performance measures P were defined as follows:

where T_n = TON time and T_a = 1-2-3 time

best case data = 28 seconds for both tasks

{all values averaged over six measurements}

$P_n = \frac{28}{T_n} \times 100$ for performance on TON task

$P_a = \frac{28}{T_a} \times 100$ for performance on 123 task

$P = \frac{P_n + P_a}{2}$ for overall performance.

Variable Framerate Results

The grayscale and resolution were kept constant and the framerate was varied. Figure 6 provides the results. This experiment was performed with two sets of grayscale/resolution settings: 1) 128 x 128 pixels, 4 bits gray

2) 64 x 4 pixels, 2 bits gray.

From these results it was clear that framerates below 5.6 frames/second considerably degraded performance and increased variability. At low framerates subjects had to use a "move-and-wait" strategy. Even though the move-and-wait strategy was time consuming, it worked!

Variable Resolution Results

The resolution was varied while maintaining a constant framerate and grayscale. Figure 7 provides performance curves for the variable resolution case.

It was found that the subjects successfully accomplished the nut removal (TON) task at resolutions as low as 32 x 32 pixels. It was noticed that the 64 x 128 pixel case resulted in considerably better performances than did 64 x 64. This showed that total number of pixels was more important than symmetry.*

---

*Symmetry was not entirely unimportant since 64 x 64 was definitely preferable to 128 x 32!

At lower resolutions, adequate operator training was important. A well trained operator could perform the task (after the manipulator is positioned) with little visual feedback.

Variable Grayscale Results

Keeping the framerates and resolution constant, the grayscale was varied. Results are shown in Figure 8. Notice that with maximum resolution (128 x 128 pixels) and maximum frame rate (28 f/sec), grayscale can be reduced to the two bit level without affecting performance. This is not true at a lower frame rate (14 f/s) where lowering grayscale does degrade performance.

DISCUSSION AND FURTHER ANALYSIS OF RESULTS

From the data gathered for these two specific tasks, it is apparent that the three parameters F, R and G could each be degraded, keeping the other two constant, without much effect on performance, up to a certain point where performance then degraded rapidly. Under the specified conditions: F = 28 f/s, R = 128 x 128, G = 4 bits, reducing the frame rate by a factor of 4 (2 bits) affected performance by only 20%. Similarly, reducing the grayscale alone by a factor of 2 bits reduced performance by 25%. In the case of the resolution, however, two bit reduction degraded performance of the TON task by 70%, while making the 123 task impossible to accomplish. It is especially useful to consider these facts in terms of the number of bits per second to be transmitted:

FRAME* RATE	RESOLUTION	GRAYSCALE	** PERFORMANCE	# OF BITS/SEC
28 f/s	128x128	4	100%	1,835,000
7 f/s	128x128	4	80%	458,750
28 f/s	128x128	1	75%	458,750
28 f/s	32x32	4	*33%	458,750

It is clear from these data that the number of bits per second could be kept the same and yet produce different performance for different combinations of F, R & G.

Correlation of Performance and Display Bit Rate

For the purpose of studying further the various trade-offs, "isoperformance curves" were constructed for combinations of F, R & G along which the performance was (almost) the same. Isotransmission curves (of constant information transmission rate) were superposed. These curves are shown in Figures 9, 10, and 11 respectively for RG with constant F, GF with constant R and RF with constant G.

The comparison shows, remarkably, that telemanipulation performance correlates very closely with bits per second of the display!

* Data combined for two subjects
** TON Performance, P not defined

Noise Problems at Low Framerates

It was noticed that at low framerates there was more noise in the picture. As a result of this, the operator's task was additionally complicated with a slower picture. From examination of the FRAG system, it was clear that the noise problems were not because of the charge-injection camera. It was therefore unclear whether:

(i) the noise was actually present in the slow images, or only

(ii) there appeared to be more noise in the slow pictures

The eye is often thought of as a low pass filter. This is in fact the reason why pictures (television, movies...) are shown not faster than 30 frames/second.

The fastest sampling rate of the FRAG was 28 f/s. Each of these frames was made up of signal and noise. In each consecutive frame the signal was the same* but the noise changed. At the higher frequencies the visual nervous system averaged the noise. At lower frame rates, however, each frame was displayed several times; here the frequency of the noise was now much lower. For example, at 28 f/s the noise had a frequency of 28 hz. At 1 f/s, the frequency was 1 Hz. Owing to the low-pass nature of the eye, the signal to noise ratio increased with frame rate.

Mathematical Model

Assume that the human eye averages over a period of τ seconds.

Let n = number of frames/τ seconds
F_n = nth frame in a particular τ period
S_n = signal in nth frame, all 1's white
W_n = noise in nth frame, all 0's black

$F_n = S_n + W_n(\sigma_n^2)$, where W_n^2 = variance of noise. For the τ seconds period the variance will be σ_n^2/n since there were assumed to be n frames/τ seconds. Signal to Noise Ratio (SNR) $= \dfrac{\text{variance of signal}}{\text{variance of noise}}$

as $\dfrac{\sigma_n^2}{n}$ decreases, SNR increases

and $\dfrac{\sigma_n^2}{n}$ decreases as n increases

Thus, for a higher n (i.e. for a higher frame rate) there will be a higher SNR

* For a stationary picture.

so that there appears to be less noise in the picture. In other words, at slow frame rates, the perceptual system "forgets" the information between frames. At high frame rates the signal is constant but the noise changes from frame to frame. When successive frames are averaged (integrated) the signal will appear to show more distinctively.

CONCLUSIONS

A first conclusion is that trained human operators could (much to their own disbelief) perform fairly complicated remote manipulation tasks with a coarse, intermittent digitized picture requiring as little as 50,000 bits/second.

Further, for a picture at 128x128 resolution, 28 f/s frame rate and 4 bits of grayscale, each of these three parameters could be decreased considerably individually, without preventing the operator from accomplishing the task.

In the range of operation (up to R = 128x128, F = 28 f/x, G = 4 bits) frame rate and grayscale could be degraded by greater factors than resolution before making task accomplishment impossible.

For the given tasks and manipulator, "threshold points" existed for all three parameters:

For FRAME RATE: 3 f/s when resolution = 128x128, grayscale = 4 bits

For RESOLUTION: 64x64 when frame rate = 28 f/s, grayscale = 4 bits

For GRAYSCALE: 1 bit

Any further degradation of a parameter beyond these points (while holding the other two constant) resulted in degradation in performance such that the task could not be completed.

It was also observed that lowering the sampling rate created more problems than making the display slower. There appeared to be more noise in images at low frame rates. This was believed to be because of the low pass nature of the visual nervous system.

A final general observation is that as F, R and G were varied, performance tended to correlate very well with bits per second in the picture.

REFERENCES

1. Ranadive, V., Video Resolution, Frame Rate and Grayscale Tradeoffs under Limited Bandwidth for Undersea Teleoperation, MIT SM Thesis, Sept. 1979.

2. Cole, R.E. and B.H. Kishimoto, Remote Operator Performance Using Bandwidth-Limited TV Displays: A Review and Proposal, Naval Ocean Systems Center Rep. NOSC-TD-379, August, 1981.

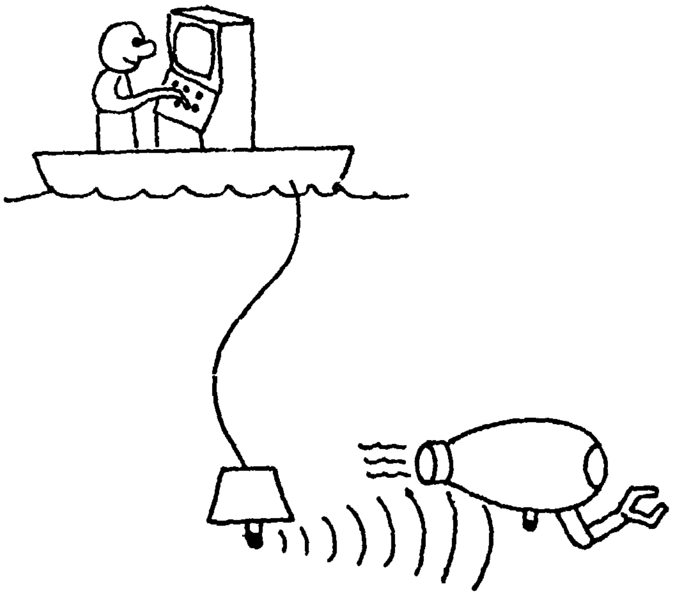

Figure 1. Undersea telemanipulator with acoustic telemetry

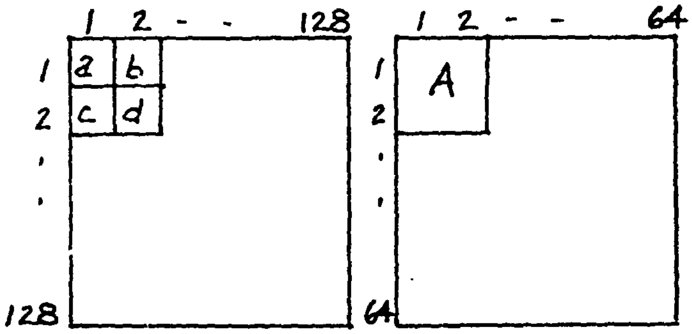

Figure 2. Resolution averaging

128 x 128

64x64

32x32

16x16

Figure 3. The same picture at various combinations of resolution and grayscale

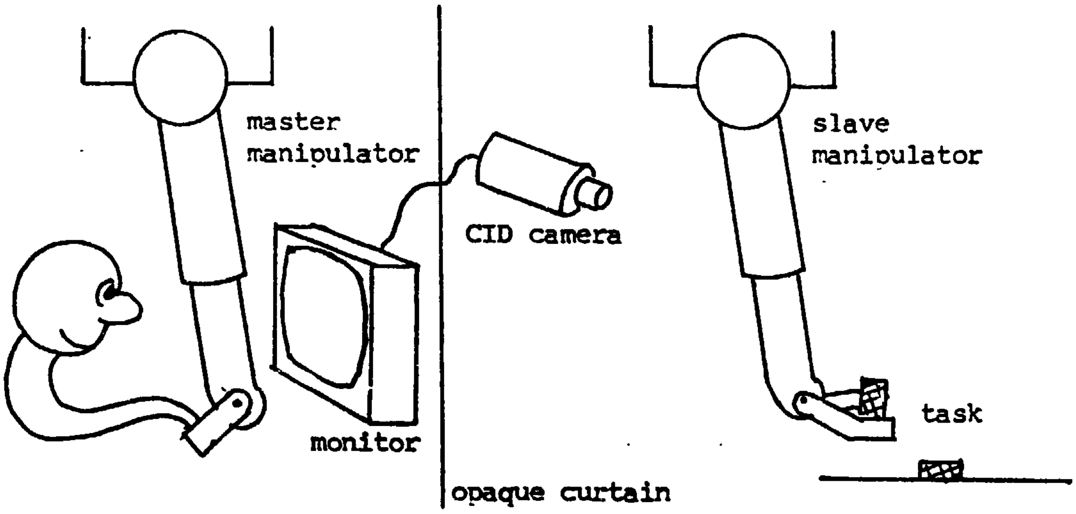

Figure 4. Experimental configuration

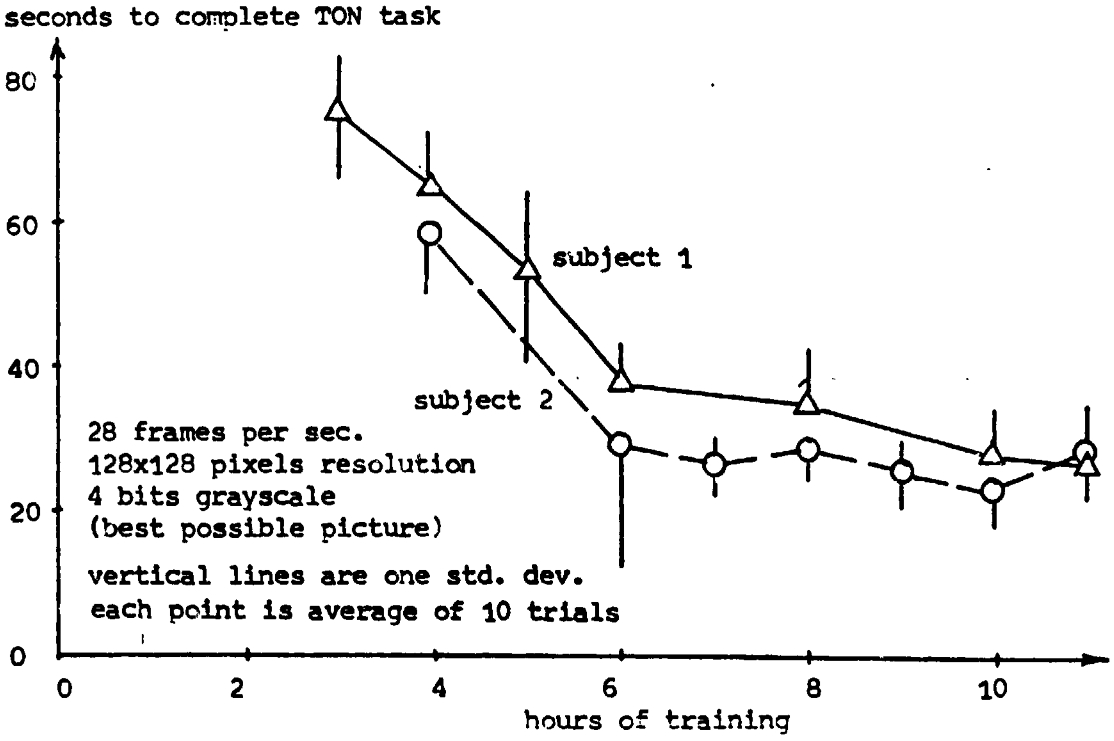

Figure 5. Learning curves for two subjects on TON task

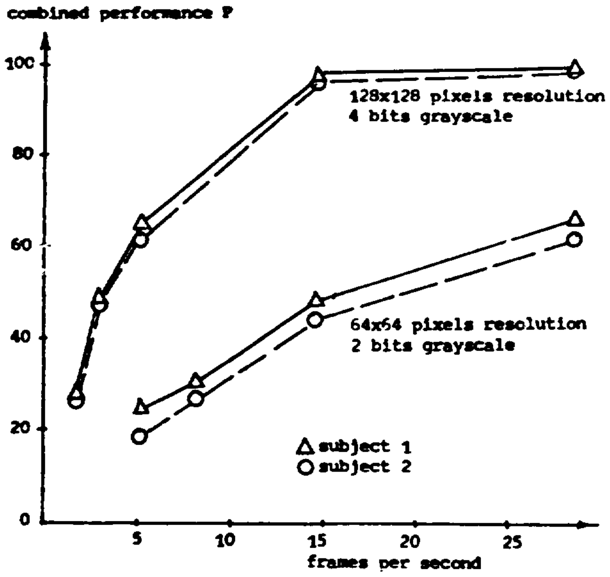

Figure 6. Effect of framerate

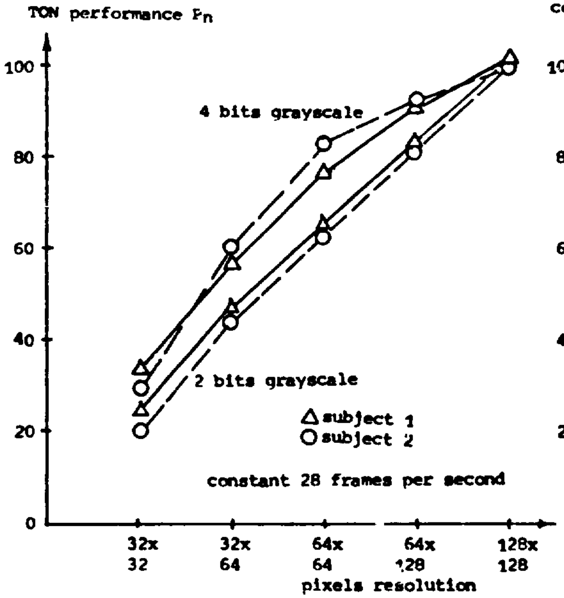

Figure 7. Effect of resolution

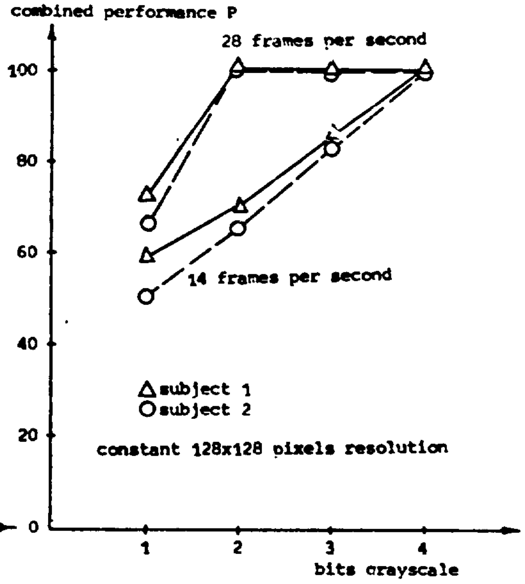

Figure 8. Effect of grayscale

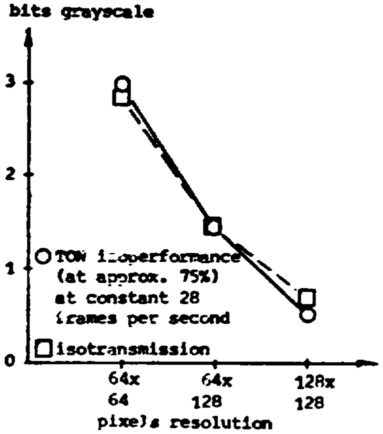

Figure 9. Isoperformance

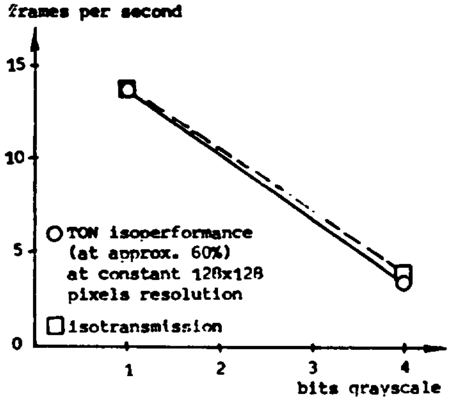

Figure 10. Isoperformance

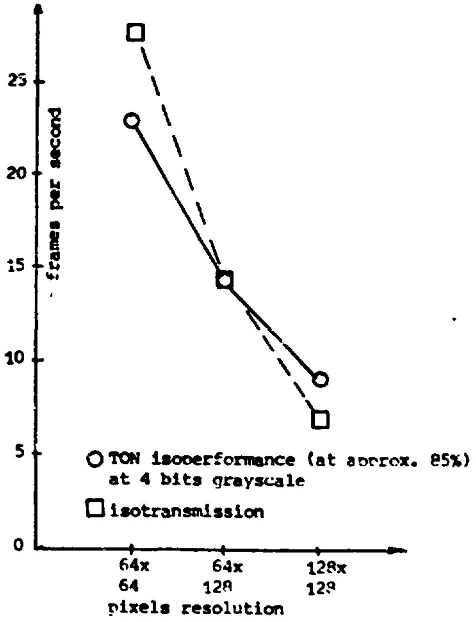

Figure 11. Isoperformance