VQEG 2001-023

IEEE G-2.1 Audio-Video Techniques Committee
G-2.1.6 Subcommittee on Video Compression Measurements

JND Visual Impairment Scale Measurement Method (JNDVIS)

Keywords
Fidelity, quality, vision, impairment, artifact, perception, subjective method, psychophysics

1 Introduction

Human subjective judgments of visual impairment of video sequences are an important means of evaluating the performance of video systems. These evaluations may be employed during the design of systems, or during operational use of a system. In either case, it is critical to understand how the rendered video will appear to the human viewer.

A further purpose of subjective measurements is to provide a database against which the performance of so-called objective impairment metrics may be compared. Objective impairment metrics are software or hardware instruments that predict human impairment judgments through mathematical analysis of the video signal. In this case, as well, it is critical that the measurements accurately and consistently reflect human sensitivity to impairment.

Still another purpose of subjective measurement is to enable the production of calibrated samples of video impairment. Such samples would be useful in testing of future video impairment metrics, and for research and educational purposes.

The purpose of this standard is to ensure that subjective measurement made for these and possibly other purposes are collected in an effective and standardized way, so that the measurements are repeatable, and the units of measurement are consistent.

The approach adopted in this standard is to measure impairment of a processed video sequence, relative to an original sequence, in units of just-noticeable-differences (JNDs). The JND is a well-established unit of measure in the study of human sensation and perception, which has also served as the foundation for image quality models (Carlson & Cohen, 1980). In simple terms, two sequences that are one JND apart are just discriminable. The method described in this standard has recently been used to measure visual impairments in twenty five video sequences (Watson & Kreslake, 2001).

2 Scope

This Standard describes a method of measuring the subjective visual impairment of digital video. The method may be applied in cases where it is desired to measure the perceived difference between a source video sequence (SRC) before and after it has been encoded by a Hypothetical Reference Circuits (HRCs). An HRC is a generic name given to a particular set of digital processing operations, such as compression and decompression by a particular codec. Examples of two specific HRCs would be CCIR-601, and MPEG MP@ML as implemented by a particular codec at a particular bit-rate.

2.1 Baseline Configuration

The method described in this standard is general and may be applied in a wide variety of applications and to a wide variety of digital video materials. We also identify a particular implementation of the method which we call the *baseline configuration*. We expect the definition of future configurations to permit testing for future applications.

In the baseline configuration, the source video (SRC) is standard definition digital video as represented by ITU-601 (ITU-R, 1995). It is useful to consider this also as an HRC, which we call hrc0. Elsewhere we provide a complete listing of the elements of the baseline configuration.

3 Comparison with Existing Standards

ITU Recommendation BT.500 provides an extensive description of several subjective methods for the assessment of television quality (ITU-R, 1998). Among these, the Double Stimulus Impairment Method (EBU), the Double Stimulus Continuous Quality Scale (DSCQS) are the most well defined and most widely used. Each of those methods relies on subjective ratings using a particular arbitrary scale (e.g. 0-100) and a range of material that is dependent upon the application. A possible drawback of these methods is the results they provide are in arbitrary units that may also depend upon the range of materials (a context effect). In the present method, observers make only a binary judgment of which of two presentations is the more impaired, and the results are expressed in just-noticeable-differences (JNDs) which are neither arbitrary or context-dependent.

4 History

- In early 1998, the IEEE G-2.1.6 Subcommittee on Video Compression Measurements initiated a "Task Force to define a unit of measure and means of calibration for video quality analysis." This effort was lead by Leon Stanger and John Libert. The Task Force discussed various methods that might be used to derive an absolute scale of video quality, and specifically, a scale measured in JNDs. See http://grouper.ieee.org/groups/videocomp/1998g216/ls-rptd1.html for an early report from this group.
- Subcommittee G-2.1.6 Meeting - July 27, 1999 - Minneapolis MN meeting - investigation of pair-comparison methods proposed
- Subcommittee G-2.1.6 Meeting - November 1, 1999 - Fort Lauderdale, FL - new plan proposed by Andrew Watson, Ann Marie Rohaly and John Libert.
- Subcommittee G-2.1.6 Meeting - April 24-25, 2000 - Boulder, CO - Andrew Watson presented his Proposal: Measurement of a JND Scale for Video Quality, IEEE Doc. G-2.1.6/112,
- Subcommittee G-2.1.6 Meeting - August 7, 2000 - Schaumberg, IL - Andrew Watson presented his Preliminary Report on JND Measurement, IEEE Doc. G-2.1.6/116, August 7, 2000 and his Proposal on JND Measurement, IEEE Doc. G-2.1.6/117, August 7, 2000.
- Subcommittee G-2.1.6 Meeting - October 30, 2000 - San Antonio, TX - Andrew Watson presented Efficient Adaptive Estimation of Sensory Scales, IEEE Doc. G-2.1.6/120. Andrew Watson and Leon Stanger will draft wording for the standard, to present at the January meeting.
- This document describes a specific method to satisfy the needs of the Task Force.

5 Contents

6 References

ITU-R. (1998). Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures International Telecommunications Union BT.500-8.

7 Definitions

7.1 Perceptual Scale/Sensory Scale

A perceptual scale is a relationship between magnitudes along a physical dimension and their corresponding perceived intensities. Values along the perceptual scale may be expressed in units of JND.

7.2 JND

A just-noticeable difference (JND) is a unit of measure for a perceptual scale. It is defined as the distance along the scale that allows two intensities to be perceived as different. In the Thurstone Model, it corresponds to one standard deviation of the underlying variability of the perceptions. In casual terms JND refers to the physical intensity difference that is just detectable by a human observer.

7.3 Thurstone Model

A theoretical model of the sensory decision process, in which sensory responses are represented by probability densities, and decisions are predicted by statistical judgements.

7.4 Binary Pair Comparison

A subjective test in which two samples along some physical dimension are presented to an observer who must report which is greater. An example would be asking an observer which of two lights is brighter.

7.5 HRC

A particular digital processing or encoding that may be applied to a video sequence. An example would be a particular MPEG codec with specific parameters. A reference HRC (hrc-0) may be defined (such as CCIR 601), in which format the source video is encoded (see SRC).

7.6 SRC

A particular sequence of video content. This is the co-called original video, encoded in hrc-0, the reference HRC.

7.7 Condition

A combination of a particular SRC and a particular HRC. This proposal describes a method of measuring the perceptual distance between two conditions in units of JND.

8 Description of the Measurement Technique

8.1 Theory

The JNDVIS method is based upon Thurstone's "Law of Comparative judgement" (Thurstone, 1959). Thurstone proposed that physical sensory stimuli (such sounds) that vary in some physical scale (such as sound pressure) might give rise to sensory magnitudes arranged along a one-dimensional internal sensory scale (such as loudness), as pictured in Figure 1. However, the sensory magnitude varies from presentation to presentation, due to the unavoidable variability of neural systems. In one particular case (Thurstone's "Case Five"), the distributions are assumed to be Normal, with a standard deviation of 1 (as shown by the two example Gaussians). The assumption of a unit standard deviation may be made without loss of generality because the internal perceptual scale is without units. By this assumption we in effect define the scale in units of standard deviations, which we also call JNDs.

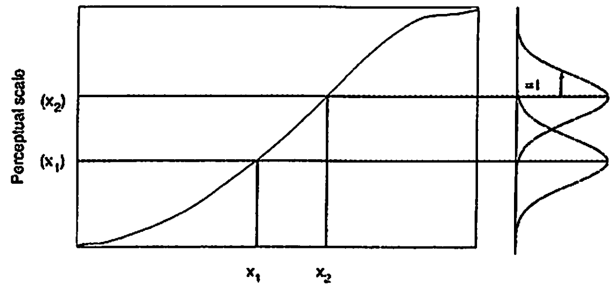

Figure 1. Thurstone Model of sensory discrimination.

A *pair comparison* is a test in which we ask a human observer, under defined and controlled circumstances, to report which of two physical samples (in our example, sound pressures) is greater. In Thurstone's "Case Five" above, the probability of a correct judgment in a pair comparison is a function only of the distance between the sensory magnitudes induced by the two intensities of the pair. We can therefore estimate these distances by finding which values would most likely give rise to the data in hand. The result will be a sensory scale: a function relating perceived intensity to physical intensity. The scale provides a description of the distance, in JNDs, between any two physical intensities in its domain. Mathematical details, and an example, will be given below.

For this estimation procedure to be well behaved, the pair comparison trials must be properly distributed over the intensity range. Likewise, since comparisons between intensities that are too close or too distant will be uninformative, proper selection of pairs is an important concern.

To measure the perceptual distance between two encodings of the same video, we will position the two versions at the end points of the physical scale and populate the interval between them with interpolated versions. After the measurement process is complete, we will be able to read from the scale the perceived distance between the two versions.

8.2 Apparatus

Application of this method requires an apparatus capable of random access selection and presentation of short sequences of digital video. Some general requirements of this apparatus are described here. A specific instance of the apparatus is described in an informative annex (Section 9.1).

8.2.1 Computer

The computer must be capable of interfacing to the video storage system. It must be capable of selecting a specific video file from the storage system and sending it to the video display with a delay of less than 1 second. As an example, a computer might use a fiber-channel interface to connect to a disk array subsystem, and use a Serial Digital Interface (SDI) to send digital video to an SDI-capable video monitor. The computer must interface to a keypad or switch with at least three buttons or switches for observer response input.

8.2.2 Video Storage System

The video storage system must be capable of accommodating all of the digital video sequences used in the measurement procedure. As an example, to measure the JND scale for one condition may require 21 video sequences. If each of these is encoded as CCIR 601 digital video and each is 8 seconds in duration, then each will require about 200 Mbytes of storage and the total requirement will be 4 Gbytes. In practice, it is useful to have sufficient storage for conditions corresponding to several SRCs and several HRCs. If there are 25 conditions, for example, the requirement is for over 100 Gbytes.

8.2.3 Video Display

A high quality studio monitor capable of displaying video derived from uncompressed CCIR-601 digital video.

8.2.4 Subjective Testing Facility

Testing should be conducted in a facility that conforms to ITU Rec. 500 (ITU-R, 1998).

Since the order of presentations may depend upon previous observer responses, multiple simultaneous observers cannot be used and the facility need only accommodate one observer at a time.

Viewing distance should be fixed and established by user requirements. In the baseline configuration, viewing distance is set at five picture heights (5H).

A keypad, interfaced to the computer, should be accessible by the observer during testing.

8.3 Observers

The method generally requires a large population of observers. A different observer is used for each measurement of the scale for each condition. In the baseline configuration, at least three observers are required for each condition.

8.3.1 Expertise

In the baseline configuration, observers should be drawn from the general population. Expert observers, defined as those professionally involved in evaluation of digital video quality, should not be used in the baseline configuration, but may be used in other configurations as applications warrant.

8.3.2 Color Vision

Observers should be examined using an Ishihara Test for Color-Blindness. Observers who fail the latter test are excluded from the measurement.

8.3.3 Acuity

Observers should also be screened for visual acuity of at least 20/30 in both eyes using a standard Snellen eye chart. Subjects should be tested with their normal optical correction (glasses or contact lenses), under a C-15 light. The test is administered separately for each eye.

8.4 Procedure

8.4.1 Selection of a Condition

Select an SRC *s* and a pair of HRCs: *h1* and *h2*. These define the conditions *{s,h1}* and *{s,h2}* to be tested. These two conditions correspond to particular videos which we may write as v(s,h1) and v(s,h2). In the baseline configuration, one of the HRCs is hrc-0, defined as the original source video.

8.4.2 Preparation of test sequences

Create a set of video samples through linear combination of v(s,h1) and v(s,h2). That is,

$$v(s,h1,h2,k) = (1-w[k])\ v(s,h1) + w[k]\ v(s,h2) \qquad (1$$

Where the weights w[k] vary in linear steps from 0 to 1. The linear combination is applied directly to the Y, Cb, and Cr values of the digitally encoded video. The number of steps should be at least 21 (step size = 0.05). The number of steps required is dependent upon the number of JNDs to be measured. If the condition being measured is likely to have more than 15 JNDs, then more steps should be used (e.g. 2 * JND). There is no penalty (except the time required for computation and the space required for storage) of having too many steps.

In the baseline configuration, one of the HRCs is the source video, written hrc0. This leads to the simpler expression

$$v(s,h,k) = (1-w[k])\ v(s,hrc0) + w[k]\ v(s,h) \qquad (2$$

In the baseline configuration, 21 steps are used, ranging from 0 to 1 in steps of 0.05. Thus 21 videos must be created before the start of an experimental session.

8.4.3 Presentation of one pair (a single trial)

A single trial consist of the following events (see Figure 2). A pair of intensities indexed by k_1 and k_2 are selected. The videos corresponding to these can be written $v(s,h_1,h_2,k_1)$ and $v(s,h_1,h_2,k_2)$, or in the baseline configuration, $v(s,h,k_1)$ and $v(s,h,k_2)$. An ordering, either $\{k_1,k_2\}$ or $\{k_2,k_1\}$, is selected randomly, Based on this ordering, the first of the two videos is retrieved from the storage system and presented on the video display. After the presentation, there is a pause of 1 second, followed by presentation of the second video. After completion of the second video, the computer, via the keypad, accepts a response from the observer that is 1 or 2, depending upon whether they saw the larger artifact in time interval 1 or 2.

The duration of both videos will be fixed and equal but dependent upon the requirements of the user. An unlimited time is allowed fro the response. In the baseline configuration, the duration is seven seconds.

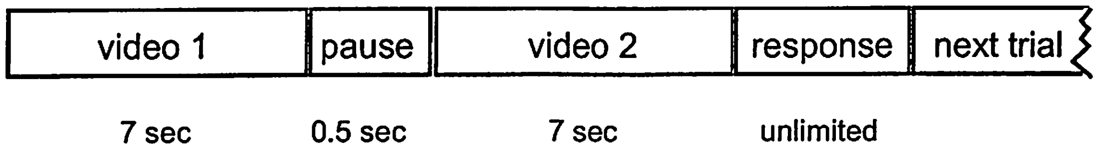

Figure 2 Event sequence for one trial.

8.4.4 Observer Response

Following the second video, the observer identifies the interval containing the larger impairment by pressing a key on the keypad corresponding to "first interval," "second interval," or "repeat." The observer is instructed to use the repeat key only in exceptional circumstances, such as involuntary eye closure during a presentation. If the repeat key is pressed, the trial is repeated, but with the order of intensities randomized.

8.4.5 Sequencing of trials

The sequencing of trials, and determination of when to terminate the measurement, are controlled by the EASE procedure (Appendix 11.2).

8.4.6 Blocks of Trials

Trials for a single condition should be completed before another condition is measured. Trials for one condition should be grouped into blocks of no more than 15 minutes duration. In the baseline configuration each block is 32 trials. Since in the baseline configuration each video is 7 seconds, and a complete trial lasts about 18 seconds, so 32 rials will take about ten minutes. A rest break of 5 minutes is provided after each block.

8.4.7 Practice

Before the start of data collection for a new observer, a practice block of at least 4 trials should be completed. This block should use an SRC that is not the SRC to be viewed by that observer.

8.4.8 Instructions to the Observer

Before the start of the experiment, the experimenter should read to the observer a set of defined instructions. These instructions should advise the observer of what to expect on each trial, of how long a block will last, and of what to do when a block is complete. They should instruct the observer in how to respond on each trial. The observer should be advised that the two presentations may appear identical, in which case they should make their best guess. They should also inform the observer regarding any special aspects of the testing facility. We provide example instructions in an informative annex (Section 10).

8.4.9 Trials/Condition and Time/Condition

Data collection for one condition terminates when three conditions are met:

1. A block is complete,
2. A minimum number of trials (min_trials) is complete,
3. A minimum number of trials/JND (trials_jnd) is complete, where the estimate of JND is obtained as described below.

In the baseline configuration, min_trials=32 and trials_jnd=20.

The total number of trials or minutes required to measure a condition will depend upon the duration of each video and the number of JNDs spanned by the condition. For example, in the baseline configuration each video is 7 seconds, and each trial is around 18 seconds. If a condition spans 16 JNDs, it will require about 16 20 = 320 trials, or about 10 blocks, which with breaks amounts to 150 minutes or 2.5 hours. A condition spanning 1 JND, on the other hand, will require only 10 minutes.

8.4.10 Estimation of JNDs

After completion of a sufficient number of trials, the sensory scale in units of JND can be estimated for that condition using the methods described in the Appendix (11.1). In the baseline configuration the measurement for each condition is repeated on at least three observers. The mean of the three observers is taken as the final estimate of JND for the condition. Standard deviation or standard error of the mean should also be reported.

8.5 Parameters

Here we provide a table of parameters related to the JNDVIS method, along with values defined for the baseline configuration.

Parameter	Baseline Confuguration	Comments
Viewing distance	5 H (H=picture height)	
Sequence duration	7 seconds	
Observer type	Non-expert	Screened for acuity and color vision
Reference format (hrc-0)	ITU-601	525 or 625 line
trials_block	32	Trials/block
min_trials	32	The mimumum for a condition.
trials_jnd	20	Trials/jnd
Replications/condition	3	

Table 1 Parameters of Baseline Configuration.

9 Informative Annex: An implementation of JNDVIS

In this annex we describe a particular implementation of the JNDVIS method. It is understood that other systems for carrying out the method are possible.

9.1 Apparatus

The apparatus used in this implementation is diagrammed in Figure 3. The apparatus consists of a computer, a disk array, a video monitor, fibrechannel interface, and SDI interface, a joystick, and an audio speaker system.

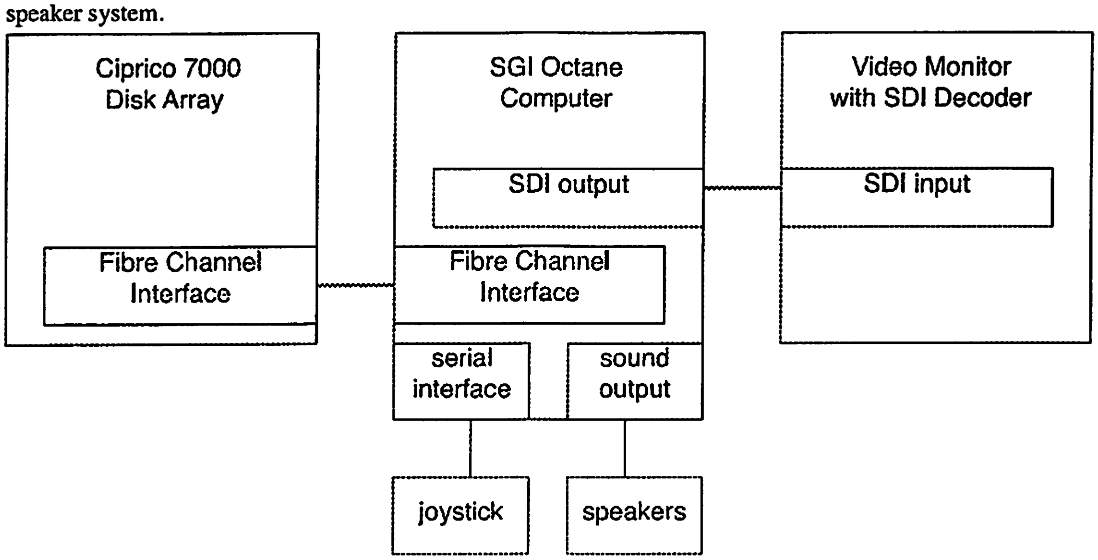

Figure 3 Appratus for JNDVIS method.

9.1.1 Computer and Interface

The computer is a Silicon Graphics (SGI) Octane computer Model No: CMNB015ANF250 with a 250 MHZ MIPS R10000 Processor and main memory of 384 Mbytes, running under operating system IRIX64 Release 6.4. The computer connects to a Ciprico 7000 disk array subsystem via an SGI Fibre Channel interface. The computer is also equipped with a Serial Digital Interface (SDI) system. Under software control, individual digital files corresponding to particular video sequences can be selected, read from the disk array, and sent from the computer to the video monitor via the SDI interface.

The computer connects to a joystick (BG Systems Joystick Model No: JF3-1-MP-CH-MP) via a serial interface (BG Systems CerealBox Analog to Digital Interface Model No: LV824-E) for observer response input. The joystick is equipped with left and right thumb switches and a trigger, which are associated with responses "first interval" (left switch), second interval" (right switch) and "repeat" (trigger).

9.1.2 Video Storage System

The system used here is a Ciprico Model 7000 FibreStore Disk Array, Model FS2, which uses 9 physical drives read in parallel to achieve throughput of up to 200 Mbytes/sec The total capacity of the system is over 400 Gbytes.

9.1.3 Video Display

This system uses a Sony Trinitron Model BVM-20E1U monitor equipped with a SDI decoder interface.

10 Informative Annex: Instructions to the Observer

The following are a set of example instructions to the observer.

"Welcome. Thank you for taking part in our experiment. Today's experiment will be divided into a number of blocks, each taking about fifteen minutes. After each block, we provide a break of several minutes. During each block, you will be presented with a series of 32 trials. Each trial consists of two successive presentations of a short video clip. The two clips differ in video quality - your job is to pick the one with the lower quality."

(Pointing out the buttons)

"You make your selection by pressing one of the two buttons at the top of the joystick: the left button if the first presentation is of lower quality, the right button if the second presentation is the worst of the two. The computer will tell you whether you are right or wrong. Don't worry if you get it wrong. The computer is programmed to make the task harder and harder until it finds where you start to error. The computer will also tell you when the session is complete. Do you have any questions so far?"

"A few more things before we get started. We need you at a set viewing distance, so at all times you must keep your back to the back of the chair. Don't lean forward when viewing the clips. And please do not attempt to move the chair."

"Finally, you may see two clips that look exactly the same. They may both be of high quality or they both might have defects. In either case, try to make your best guess as to which one seems worse. Other clips will be more obvious when trying to discern the worse of the two. If, for some reason, you don't see one of the two clips, you may press the red trigger button on the back of your joystick with your index finger. This trigger will replay the last pair. Only use this in case of 'emergencies.' Never press it if you've seen both clips- make your best guess."

"Do you have any further questions before we start the practice trials?"

(At this point the observer completes the practice trials. A digitized voice message indicates when the practice session is complete)

"You will hear a message similar to that after every session. It will usually say "This block is complete, time for a break." At that point, get up and stretch, and come get me to reset the computer. I will be at my desk if you run into any problems or have any further questions. Are you ready to get started?"

11 Appendices

11.1 Estimation of Sensory Scale from Binary Pair Comparison Data

Consider a set of binary pair-comparison data collected over a physical interval x. Each trial record is of the form

$$r = \{x_1, x_2, d\} \qquad (3$$

where x1 and x2 are the physical intensities compared, and d is the decision which is either correct (d=1) if x2 is judged larger or incorrect (d=0) otherwise.

A physical intensity x evokes a mean sensory response Ψ(x) on the sensory scale Ψ. From trial to trial, the sensory response is variable with unit standard deviation. The probability of each outcome is thus

$$\begin{aligned} P(r) &= C\left(\frac{\psi(x_2)-\psi(x_1)}{\sqrt{2}}\right) \quad if\ d=1 \\ &= 1 - C\left(\frac{\psi(x_2)-\psi(x_1)}{\sqrt{2}}\right) \quad if\ d=0 \end{aligned} \qquad (4$$

where C is the cumulative standard normal distribution.

The likelihood of a complete data set is thus given by

$$L = \prod_j^J P(r_j) \qquad (5$$

Using standard optimization techniques, we can then find the parameters that maximize the likelihood function L. In practice, it is often easier (and equivalent) to maximize the log of L. We consider two possible types of estimate: sampled and functional.

11.1.1 Sampled Estimate

In the sampled estimate, the parameters estimated are the actual sample values of the scale $\Psi(x_k)$ at each of the *K* physical intensities x_k , *k*=1,...,*K* used in the measurement. We assume only that scale increases monotonically with intensity. It is more efficient to actually estimate the K-1 differences $\Delta_k=\Psi(x_{k+1})-\Psi(x_k)$, and to require that these are positive. Any standard optimization software may be used to maximize the leiklihood with respect to these *K*-1 parameters.

11.1.2 Functional Estimate

In the parametric method, we estimate a small number of parameters of a function. The function used here is

$$M(1-T)^{-P}\,\mathrm{Max}(0, x-T)^P \qquad (6$$

This is a power function with an exponent of *P* a value of *M* when *x*=1, and a threshold of *T*.

Examples of both functional and sampledestimates are shown in Figure 4. The red curve shows the functional estimate, with parameters $M = 6.23, P = 0.811$, and $T = 0.398$. Note that the curve is flat until the threshold is reached, whereupon it begins to rise until it reaches the maximum value of *M*.

The sampled estimate is given by the black points, whose horizontal coordinates correspond to the intensities used in the measurement, and whose vertical coordinates are the estimated corresponding JND values. The maximum value is $M_N = 6.37$, quite close to the functional estimate *M*.

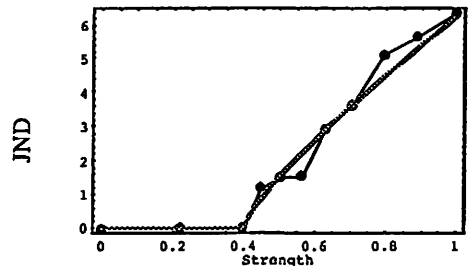

Figure 4. Functional (red) and Intensity estimates of sensory scale.

11.2 EASE Method

In the preceding section we showed how to estimate the sensory scale from binary pair-comparison data. Here we describe a method for selecting which pairs to compare. This method is called EASE (Efficient Adaptive Scale Estimation).

The challenge is that only certain pairs will provide informative binary comparisons. If the intensities are too close the observers decisions will be random, while if they are too far apart, the observer will always be correct. The most informative pairs are those about 1 JND apart. However, in advance of the measurement, we do not know which pairs are 1 JND apart. We solve this conundrum with an iterative adaptive method, as pictured in Figure 5.

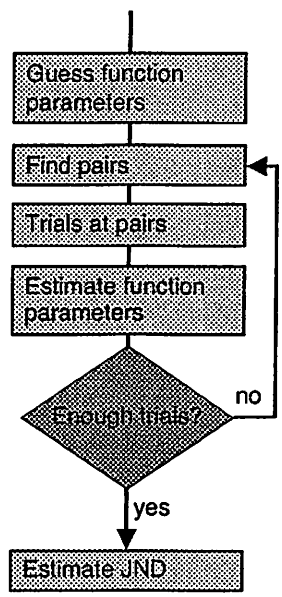

Figure 5. Flow diagram for the EASE method.

The EASE method begins by assuming a particular functional form for the scale function (Equation 6) and guessing at particular parameters M, T, and P. This assumed function is shown by the black points in Figure 6. These figures show a simulation of the EASE method, in which the true sensory scale is known (red curve).Based on the guessed function, a set of pairs that span the intensity domain and that are about 1 JND apart are selected (colored line segments). Trials are conducted at each of these pairs. Using all the data collected, a new estimate of the function parameters is obtained. If enough trials have been collected, the procedure terminates. Otherwise, it selects new pairs based on the new function estimate and continues. After termination, final estimates of function parameters are obtained. There are two important processes within the EASE method. The first is estimation of the scale, the second is selection of best pairs based on that scale.

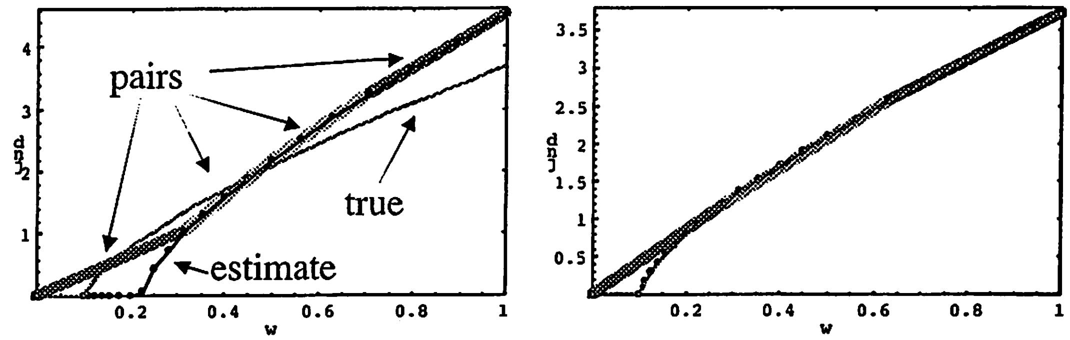

Figure 6. Illustration of the EASE method. The first panel shows the initial state, the second panel shows the state after 128 trials.

11.2.1 Scale Estimation

Estimation of the sensory scale is accomplished using the functional estimate method outline above in Section 11.1.2. However, due to the need for rapid estimation, direct search may be used as the optimization method. This is accomplished by defining a reasonable range for each parameter (*M,T,P*) and subdividing each range into N steps (*N*=10 is usually sufficient). Then the log likelihood is evaluated at each of the N^3 locations in the parameter space, and the location yielding the maximum liklihood is selected. For additional speed, it is possible to fix the parameter *P*=1.

11.2.2 Parameter Ranges

In the direct search method described above, it is necessary to select ranges for the parameters M, P, and T. The parameter T has a natural range of 0 to 1. The parameter P can be given a range from 0.5 to 2.0. The parameter M does not have any fixed natural range, since it depends upon the particular condition being measured. In the EASE procedure, it is set to 4 times the current estimate of M.

11.2.3 Pair Selection

Once a sensory scale function is specified, it is possible to select pairs that are about 1 JND apart. It is also important that the pairs include the end points of the intensity domain. The algorithm we adopt to achieve this is as follows. Let x_i I=1,...,I be the set of possible intensities. Let Ψ (x) be the specified scale, First we compute the set of values $\Psi(x_i)$. The largest of these is $\Psi(x_I)$. We compute

$$\text{steps} = \text{Max}[1, \text{Round}[\Psi(x_I).\,/\text{jndstep}]] \tag{7}$$

where *steps* is the number of steps (pairs) that can fit within the intensity domain, and *jndstep* is a parameter of the procedure specifying how far apart in JNDs each pair should be (e.g. jndstep=1). Then we compute

$$\text{psistep} = [\Psi(x_I).\,/\text{steps} \tag{8}$$

where *psistep* is the approximate size of each step in JNDs.

We then compute the set of Ψ values that span the range and are 1 JND apart, i.e.

$$\Psi_{\sigma\tau\varepsilon\pi}=\text{step psistep, step} = 0,\ldots,\text{steps} \tag{9}$$

Next we find the values x_i which yield $\Psi(x_i)$ that are closest to the desired sample points $\Psi_{\sigma\tau\varepsilon\pi}$.

These sample points are then partitioned into a sequence of adjacent pairs. If the resulting list of pairs is empty, the single pair {x1, xI} is returned.

11.2.4 Termination

The ease procedure terminates when enough trials have been collected. The number of trials required is dependent upon the number of JNDs (*M*) spanned by the condition under test. This quantity is estimated with each iteration of the EASE method. We recommend a minimum of 32 trials/JND. It is also recommended to establish an absolute minimum number of trials (e.g. 32) so that unstable early estimates of M do not cause premature termination.

12 Bibliography

Carlson, C. R., & Cohen, R. W. (1980). A simple psychophysical model for predicting the visibility of displayed information. *Proceedings of the Society for Information Display, 21*(3), 229-245.

ITU-R. (1995). Recommendation ITU-R BT.601-5, Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios.

ITU-R. (1998). *Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures* (BT.500-8): International Telecommunications Union.

Thurstone, L. L. (1959). *The Measurement of Values*. Chicago: University of Chicago Press.

Watson, A. B., & Kreslake, L. (2001). Measurement of visual impairment scales for digital video. *Proceedings of the SPIE, 4299*.