Definition of thresholds for stereoscopic depth

Manfred Fahle, Sigrid Henke-Fahle, John Harris

Abstract
In the laboratory, thresholds for stereoscopic depth perception are usually determined by asking observers to discriminate between a stimulus with a given depth offset and its mirror image. Threshold is most often defined as the disparity difference that yields 75% or 83% correct responses. Disparities used for clinical tests of stereopsis are much higher. Here it is argued that, among other factors, this is because of the fact that clinical tests usually require the detection of a depth difference (offset versus no offset), rather than the discrimination between two directions of depth difference (in front versus behind). From a formal comparison of the two tasks, the data show that discrimination, or classification is easier by at least a factor of 2 than detection. The contribution of variations of the threshold criterion and learning to the differences between stereoeucuity as measured in laboratory and clinic is also discussed. These differences are relevant to the design of tests for clinical use.

(Br J Ophthalmol 1994; 78: 572-576)

In a number of perceptual laboratory tasks, often called hyperacuity, detection thresholds fall below a photoreceptor diameter, as opposed to two point resolution where best performance is slightly above a receptor diameter. Thresholds are usually measured by asking observers to determine which of two stimuli is nearer in a stereoscopic test or to discriminate between an offset to the right versus offset to the left in a vernier or bisection task. Stereoscopic thresholds are often measured by presenting, in random order, two stimuli at slightly different distances from the observer. The binocular disparity leading to 75% or 83% correct responses regardless of which of the stimuli is closer to the observer is considered as the observer’s threshold for stereoscopic detection and is usually below 10 arc seconds, at least in trained observers.

In clinical tests of stereopsis, on the other hand, the smallest disparities used are generally much higher than 10 arc seconds. Several likely reasons for the apparent discrepancy between laboratory and clinical stereoscopic performance can be found in the literature, including, in the clinic, the lack of training of the observers and the adoption, in the tests, of a threshold criterion of almost 100% correct responses. However, one important difference between the two kinds of test is the nature of the subject’s task. In the typical clinical test, this is the detection of a non-zero binocular disparity, whereas in the usual laboratory experiment two disparities of opposite sign must be discriminated, or classified. In this paper, we show that the difference in the psychophysical tasks can account for a factor of 2 of the discrepancy between clinical and laboratory findings. In the discussion, we estimate how large the contributions of training and threshold criteria may be.

In previous studies using other tasks, whether detection of a stimulus was better than discrimination between stimuli was found to depend on the nature of the task. In the present investigation, we have compared thresholds for the classification with those for the detection of stereoscopic depth differences, in the same observers, and with similar stimuli and psychophysical procedures. In the first of these tests, observers have to indicate the direction of a slant in depth, rather than just to indicate whether a slant in depth is present or not, as in the second kind of test. While the former seems to be a more difficult task than the latter - observers must discriminate the direction of slant - it might actually be the other way round. Consider the problem of detecting whether or not a horizontal stereoscopically presented line is tilted away from the frontoparallel plane (cf Fig 1B with Fig 1C), compared with that of discriminating whether the left or the right hand end of the line is nearer to the observer (cf Fig 1A with Fig 1B). It is obvious that, for a particular slant away from frontoparallel, the difference in orientation between two lines slanted in opposite directions away from frontoparallel is larger (by a factor of 2) than that between a slanted line and a frontoparallel line. If there is no special visual mechanism which only detects departures from frontoparallel (but not their sign), so that the same mechanisms have to be used in the detection and classification tasks, then thresholds in the detection task should be higher (by a factor of 2) than those for the classification task.

Materials and methods
Stereoacuities were presented on two oscilloscope screens (Tektronix 604A, with greenish P31 phosphor) under microcomputer control via 16 bit digital to analogue converters. Two dots (diameter approximately 0.7 arc minutes, separated by 10 minutes of arc when disparity was zero) with appropriate disparities were produced on each monitor and the ray paths were superimposed optically by use of a beam splitting pellicle and polarising filters in front of both monitors, and matched filters in front of the eyes which assured that each eye saw only one pair of points with variable binocular disparity (cf Fahle and Westheimer). All stimuli were presented for either 300 or 1000 ms on the oscilloscope screens. Luminance of the stimuli was typically 150 cd/m² on a surround of 2 cd/m², and so contrast C0 = A/L/L was 76.

The stimulus conditions are shown schematically in Figure 1. In the first experiments,
observers made a judgment about a single pair of stereoscopically fused dots. In the classification task, observers had to indicate whether (as though viewed from above) the imaginary line joining the dots was slanted in depth to the right or to the left (later referred to as the right/left task). In the detection task, observers had to indicate whether the dots were at the same distance or at different distances (the yes/no task – cf Fig 1A–C). Two variations of this second task were used. In the first, the slant in depth was always in the same direction (‘yes/no to 2 AFC’). In the second, slants to the right or left had to be discriminated from frontoparallel (‘yes/no 3 AFC’ – but the direction of offset did not have to be indicated, so the discrimination was still between frontoparallel and slanted).

The experiments were repeated with a classic, simultaneous presentation, two alternative forced choice procedure. Two pairs of stereoscopically presented dots, separated vertically, were shown simultaneously side by side (Fig 1D–F). The central vertical line, which appeared between the two pairs of dots, was presented on only one monitor, and so was perceived monocularly. The reason for this line was to make it harder for the subject to compare the disparities between (rather than within) dot pairs, but without providing a zero disparity reference. It is known that insertion of a single point between two other points of differing disparity makes discrimination of their disparities much harder. 10

In the classification task, the observers had to judge which of the two pairs of dots was slanted so that the lower dot of the pair appeared closer to the observer (‘yes/no to 2 AFC’) or which pair was slanted as opposed to frontoparallel (‘yes/no 3 AFC’) (Fig 1D). In the detection task, they had to indicate whether the right or the left pair of dots was slanted in depth while the other pair was frontoparallel (Fig 1E, F).

In the experiments using both single and simultaneous presentations, thresholds were measured by using a method of constant stimuli (MCS) and auditory feedback about the correctness of responses was given throughout the experiment. Within each experimental run, successive stimuli could have one of three different offsets, chosen to bracket the estimated 75% correct threshold. The offsets were interleaved within a run, and their order of presentation was randomly chosen. After the run, the 75% correct threshold was estimated by fitting a psychometric function to the data. Each reported threshold is based upon at least 200 responses per observer.

Seven observers took part in the experiments. Some basic data about their age, sex, and visual status are given in Table 1. On the usual clinical criteria, they would be considered to have normal vision.

**Results**

The results from the single presentation experiments, shown in Figure 2, demonstrate that thresholds are higher, by approximately a factor of two, for the detection of stereoscopic depth difference between two points than for the classification of direction of the depth difference. The same is true for two stereoscopic stimuli (consisting of two dots each) presented side by side in a simultaneous 2 AFC task where observers had to indicate (a) detection: which of two simultaneously presented pairs of dots was offset in depth (versus frontoparallel) and (b) discrimination: which pair was slanted such that the lower point was closer to the observer (Fig 3).

The mean ratios and standard errors between the results shown in Figures 2 and 3 are as follows:

<table>
<thead>
<tr>
<th>Stimulus description</th>
<th>Presentation times (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(YIN 2 AFC)(R/L)</td>
<td>(YIN 3 AFC)(R/L)</td>
</tr>
<tr>
<td>One stimulus pair</td>
<td></td>
</tr>
<tr>
<td>2:60 (0:56)</td>
<td>3:95 (0:68)</td>
</tr>
<tr>
<td>2:29 (0:51)</td>
<td>3:32 (0:43)</td>
</tr>
<tr>
<td>2:45 (0:37)</td>
<td>3:50 (0:39)</td>
</tr>
<tr>
<td>Two simultaneous stimuli pairs</td>
<td></td>
</tr>
<tr>
<td>2:62 (0:5)</td>
<td>4:94 (0:95)</td>
</tr>
<tr>
<td>1:84 (0:22)</td>
<td>7:15 (2:39)</td>
</tr>
<tr>
<td>2:23 (0:29)</td>
<td>6:04 (1:28)</td>
</tr>
</tbody>
</table>

The ratios between the thresholds obtained in the two alternative yes/no tasks and the right/left
Table 1  Some characteristics of the seven observers in the present study

<table>
<thead>
<tr>
<th>Age</th>
<th>Sex</th>
<th>Refraction R</th>
<th>Refraction L</th>
<th>Binocular function*</th>
<th>Acuity†</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>F</td>
<td>-5.5 sph</td>
<td>-5.5 sph</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
<tr>
<td>51</td>
<td>F</td>
<td>-1.0 sph</td>
<td>-1.0 sph</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
<tr>
<td>32</td>
<td>F</td>
<td>0</td>
<td>0</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
<tr>
<td>41</td>
<td>F</td>
<td>-1.25 sph</td>
<td>-1.75 sph</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
<tr>
<td>SHF</td>
<td>M</td>
<td>+0.75 sph</td>
<td>+0.75 sph</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
<tr>
<td>MF</td>
<td>M</td>
<td>+0.75 sph</td>
<td>+0.75 sph</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
<tr>
<td>NF</td>
<td>F</td>
<td>0</td>
<td>0</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
<tr>
<td>UW</td>
<td>F</td>
<td>-6.5 sph</td>
<td>-5.0 sph</td>
<td>Rings 1-9</td>
<td>25/20</td>
</tr>
</tbody>
</table>

*Titmus stereo test. †With optical correction: Landolt rings.

Discussion

We first consider why the measured thresholds in the various tasks should have this relation, from a functional point of view, and then turn to the implications for a comparison of laboratory and clinical tests. The results suggest that, for the cases where there are two possible stimuli (right/left and yes/no 2 AFC), a certain minimal difference (d) between the stimuli is required for the resolution of stereoscopic offsets, and it is of little importance whether both stimuli are offset by d/2 (though in opposite directions), or only one is offset by d while the other stimulus is frontoparallel. This is to say that thresholds are not determined by a mechanism which detects deviations from the frontoparallel plane. If such a detector exists, its sensitivity is clearly lower than that of the mechanisms determining thresholds: at threshold, there is nothing special about the frontoparallel plane!

The results indicate that, to a first approximation, the difference between the stimuli is important, not their absolute values. The discrimination between a frontoparallel and a slanted stereoscopic stimulus leads to a threshold that is twice as large as that required for discrimination between slant to the left versus to the right. Thresholds for the detection of a slant are higher if the slant may be to the left or to the right ('3 AFC') than if slant is always in the same direction (for example, right; '2 AFC'). This corresponds well to the subjective impression of the observers: in the 2 AFC case, in which the subject knows that the discrimination is between 'frontoparallel' and 'slanted to the right', all stimuli that appear frontoparallel or slanted to the left are classified as frontoparallel, and only the ones appearing to be slanted to the right are classified as slanted. This strategy obviously cannot be used in the '3 AFC' task, where slants to both sides are possible. The additional uncertainty about the stimulus leads to higher thresholds.

Although our main concern in this study was the relative performance in two stereocuity tasks, rather than the absolute level of stereocuity, it is clear from Figures 2 and 3 that there were large differences between the observers. In particular, experiments, the performance of individual observers varied by a factor of up to about 14. This variability occurred despite the fact that all observers had 'normal' monocular acuity and stereoscopic vision (see Table 1). Such variability is just as likely in the population assessed by clinical tests, but less likely in the population assessed by laboratory tests, which often consists of the experimenter(s) and a few highly trained colleagues. Since laboratory thresholds are likely to be close to the absolute limit of the stereoscopic system, increased variability in the clinic can only come from the inclusion of individuals with poor stereocuity (compared with laboratory measurements). Thus this variability is probably one factor underlying the

Tasks are close to, but somewhat higher than, a factor of 2. If three stimulus configurations are presented, the ratios between the thresholds obtained in those conditions and the right/left tasks are even higher. In a further series of experiments, not reported here, measurements were made with an adaptive staircase procedure (PEST), rather than the method of constant stimuli. PEST adjusts the stimulus on each presentation to find the disparity at which the subject performs at some specified level, such as 75% correct. Very similar results, regarding both the thresholds themselves as well as their ratios, were found with that method, suggesting that the present results are robust.
Definition of thresholds for stereoscopic depth

Figure 3. Stereoscopic thresholds and ratios as in Figure 2, but for a simultaneous two alternative forced choice task, as indicated in Figure 1 D-F.

higher average thresholds measured in the clinic. In tests used for clinical assessment of patients, the smallest disparities tested are around 1 minute of arc, such as 40 seconds of arc for ring number 9 in the Titmus test, between 15 and 480 seconds of arc in the TNO test, but higher (550 seconds) in the Lang test. These values exceed the typical laboratory stereocuity (which we assume to be 10 seconds) by a factor of 4 (Titmus test), of between 1-5 and 48 (TNO test), and of 55 (Lang test). Various factors contribute to this difference. Firstly, thresholds in the laboratory are usually defined as the stimulus which gives 75% or 83% correct responses, while it is expected that a clinical test can be solved by all observers with normal stereocuity (that is, performance will be close to 100%). The 75% or 83% correct thresholds from the laboratory are usually estimated by fitting a psychometric function (often the normal probability integral) to a set of experimental data, in which percentage correct detection or discrimination has been measured as a function of binocular disparity. The increase in thresholds which would be found by selecting the disparity at which performance just reaches 100% in the present study would be exactly a factor of two. This is because, for an ideal psychometric function, the 75% correct point lies halfway between the 50% and 100% correct points. Secondly, there is a fair amount of learning involved in the extremely low thresholds reported in the literature to untrained observers, even with normal vision, will perform worse than experienced observers. This has been measured both for stereoscopic thresholds and for vernier acuity thresholds. For example, in a stereocuity task, in which thresholds were repeatedly measured over 4000 presentations, acuity in one subject improved by a factor of about 4, and in another by a factor of less than 2. Similarly, in a group of six initially naive subjects with no known visual impairments, vernier acuity was measured in 40 successive blocks, each of 240 stimulus presentations. The mean vernier threshold fell from about 16 arc seconds in block 1 to about 6 arc seconds in block 40. These changes in threshold, taken across both studies, correspond to an improvement by a factor of about 2.5. In many laboratory studies, the subjects (often the authors themselves) will have undergone more intensive training than this. Thirdly, all the above mentioned tests require detection of a (depth) difference, not ‘just’ an indication of its sign. As we have shown, given the usual definition of stereoacuity, laboratory tests may overestimate performance, when compared with clinical tests. For comparison purposes most published hyperacuity thresholds would have to be multiplied by a factor of two.

Taken together, differences in these three factors, threshold criterion, amount of learning, and type of task, can account for a factor of around 10 difference in the estimates of stereoacuity from clinic and laboratory. Our estimate of the combined effect of these factors may be conservative: for example, increasing disparity in the task used here until performance was just perfect would lead to an increase of threshold by a factor of 2. If one’s criterion for a successful clinical test is that failure implies certain abnormality, then this factor would be larger. However, whatever the contributions of threshold criterion and learning to the difference between stereocuity thresholds in laboratory and clinic, this study shows that, for procedural reasons, the difference is smaller by a factor of 2 than one might think at first sight.

We wish to thank Mrs U Weber for help during the experiments, Mrs H Welker for secretarial help, the observers U Weber, H Welker, EDeLuca, A Hildingen, and N Fahlbe for their patience. Supported by grants from the Deutsche Forschungsgemeinschaft (Fa 1195/2 and Ez 19/1–9) and the British Council/DAAD.

14 Fendick M, Westheimer G. Effects of practice and the
History of ophthalmology

Couching for cataract – the suttiah’s tale

Operations for depressing the lens and curing cataract were described in AD 37 by Celsus, who compiled a medical encyclopaedia during Tiberius’s reign. The technique persisted unchanged, and at the turn of this century it was observed by British ophthalmologists working in India. They were not impressed. R H Elliot describes the method as ‘crude and filthy’, and Drake-Brockman decries it in terms which would make a suttiah’s ears burn (suttiah being the Hindu term for oculist).

The suttiah travelled constantly, conducting his daily business roughly as follows: patients (with eye disease from cataract, infection, trauma, or any cause whatsoever) would trustingly settle cross legged in the dust for ‘removal of their cataract’. Requesting a bowl of water, the oculist took the lancet out of his instrument pouch, casually twisted a bit of cotton around his finger, and wrapped it round the blade, leaving about four millimetres free. His copper instruments would look markedly rough and unfinished to us, but as they were handed down from his father, the suttiah prized them highly.

He would make no attempt to wash in the water, it was for another purpose entirely. Having assured everyone that miraculous cures were imminent, he would instruct a relative to squat behind the patient with his arms firmly around the patient’s body. Holding his lancet, the suttiah would wipe sweat from his own eyes and the patient’s and begin, erecting the lid and steadying the globe with one finger. Additional relatives were probably recruited at this stage if the patient resisted. Choosing a point close to the cornea, in the lower outer quadrant, the suttiah jabbed his lancet into the eye until penetration was checked by the twist of cotton. It was then withdrawn. Looking down, the suttiah then picked up his needle. This was inserted into the vitreous and moved around vigorously in order to depress the lens.

Undoubtedly, some suttiahs showed slicker hand movement than others. The same scene with a truly clumsy one is painful to imagine, but as the trade was handed down from father to son rather than by skill it must have happened. The needle was then stilled and the patient asked if he could see. If not, the motions were continued (perhaps more feverishly). Presumably the patient was often told that the return of sight would take some time (that is, after the suttiah had left town). After withdrawal, the needle was put into the water, and the suttiah fished out and displayed the tiny bit of membrane (which he had thrown in beforehand) saying ‘here is the membrane that has destroyed your sight!’ Payment was then required. If there was no money the suttiah entered the patient’s dwelling, taking food, cooking pots, or cloth to the value of his fee. Many suttiahs gave special rates for old people, and in these cases covered the patient’s head with a cloth and performed the operation in darkness. Removing the cloth, the patient was asked whether he could see (a prime example of a leading question) and the suttiah left hastily.

What actually happened within the eye during all this was studied by Elliot who, during a long and vitriolic campaign against couching, got hold of 54 specimens of operated globes. Although forward dislocation of the lens occurred occasionally, backward dislocation was the rule. ‘Couched’ lenses were found floating in the vitreous, hopelessly entangled in exudate within it, and fixed to the iris or ciliary body by masses of scar tissue. In many cases the retina was detached – Elliot attributes this to severe septic infection within the globe.

Concluding that only 20% of operations yielded a discernible increase in vision, Elliot presented the globes to the the Royal College of Surgeons, where unfortunately they were destroyed by wartime bombing in 1941. Whether Elliot ever got hold of an actual suttiah is not stated. Undoubtedly, they and their lore were very elusive. Drake-Brockman spent 22 years unsuccessfully trying to get a set of instruments from a suttiah. When his nephew finally procured a set (presumably by foul play) in 1895, Drake-Brockman presented them to the Ophthalmological Society with glee. Elliot presented a set to the Royal College of Surgeons in 1922. They are there today. Back in 1895, it was agreed that they looked ‘extremely crude and damaging’, and that the suttiahs’ trade was causing widespread destruction of sight. It is interesting to speculate whether in any Indian villages, the suttiah still visits, plying instruments which – remembering the strict father-son rule – may date from before the eighteenth century.

FIONA ROMAN
