A Precision Instrument for the Clinical Measurement of Stereoscopic Acuity

William L. Larson* **Roland Giroux****

Abstract

When tested by different methods, a patient's stereoscopic acuity is seldom the same. This raises the question as to which one gives the true stereoscopic acuity. To answer this, a programmable test for stereoscopic acuity was designed to serve as a testing standard. This is described and a comparative study with it and four other tests is reported.

Abrégé

Différents tests d'acuité stéréoscopique donnent rarement les résultats identiques même pour le même sujet. Il est donc question de savoir lequel des tests donne un résultat véridique. Ce travail décrit un nouveau test basé sur un programme d'ordinateur, qui servirait comme un test étalon. Une comparaison de ce nouveau test avec quatre tests déjà existants est incluse.

Tests to determine the quality of stereopsis can be used for the clinical evaluation of visual function.[1] This is because the perception of depth through stereopsis is best only when there is good acuity in both eyes[2,3], when fixational errors are insignificant[4] and when integration at the cerebral cortex is complete[5]. The quality of stereopsis is usually expressed in terms of the least binocular disparity which can be detected as a difference in perceived depth[6]. This is called stereoscopic acuity and for observers with perfect vision, it is found to be in a range of from 2 to 8 minutes of arc[7]. Values greater than this can usually be attributed to an imbalance of visual acuity, oculomotor imbalance or to some degree of suppression[7]. Stereoscopic acuity is unique in that it is an indicator of the overall quality of binocular vision.

At present, stereoscopic acuity is of limited clinical use. As Scott and Marsh stated, "the inability to ascertain individual differences in stereoacuity beyond a certain level limits or precludes some potential applications of the measure"[8]. This limitation came to our attention because we had the opportunity to test the same persons with a number of currently available tests. If each test were equally valid, one would expect that similar results would be found. This was seldom the case. In addition, the least disparity in the test was usually too great to test for normal stereoscopic acuity.

Superficially, most manufactured tests for stereoscopic acuity appear to be of some use. It has been our experience that tests new to us have always seemed worthwhile. With experience, our opinion has changed for the worse and in some cases we have finally concluded that a test was worthless.

One of us (Larson) undertook to study the design and construction of stereoscopic acuity tests with the objective of finding one which could be relied upon for clinical use. Ultimately, it was deemed necessary to design and construct a test based on the principles enunciated by Howard in 1919[7]. Howard's test (which is *not* the same as the Howard-Dolman) is beyond reproach from a scientific point of view but is not suitable for clinical use because it is too slow, occupies too much space and cannot detect certain anomalies of stereopsis. These defects have been overcome by changes in the test's design and by the use of automation and computer control. The result is a precision test for stereoscopic acuity which is an effective clinical test and is a standard against which other tests can be judged.

Early tests for stereoscopic acuity (developed in the 19th century) presented retinal image disparities by means of a stereoscope. Howard observed that these tests were unreliable. Therefore, he devised a test in which real objects (two parallel vertical rods) positioned in real space (at a distance of 6 meters) produced the disparities. He screened the ends of the rods from view so that all monocular distance cues were eliminated. He used pilots in the American Air Force as subjects and his results showed for the first time that the limit of stereoscopic acuity is about 2 seconds of arc. He also observed that an acuity of more than 8 seconds was associated with some sign of oculomotor or visual acuity imbalance. He concluded that perfect stereoscopic acuity was 8 seconds of arc or less.

Most tests in use today present disparities stereoscopically. We suggest that this is done more for reasons of marketability and ease of manufacture than for the production of an effective test. With the advent of vectographs, tests which previously required a stereoscope could be performed with polarized glasses. Unfortunately, the plastic film of the vectographs was not sufficiently stable to permit the presentation of the smaller disparities. The least disparity which can be maintained is not known to us but commercially available vectographs have a least disparity of 20 sec. or more. Tests printed on paper are also subject to instability due to expansion and contraction, particularly with changes in humidity.

* O.D., M. Eng., F.A.A.O., Professeur agrégé
** O.D., M. Sc., Professeur agrégé

Ecole d'Optométrie
UNIVERSITE DE MONTREAL
c.p. 6128 Succ. "A"
Montréal, Québec
Canada

One test printed on paper claims to present a disparity of 3.6 sec. If this were possible, a dimensional difference of 0.0035 mm (0.00014 inch) would have to be maintained. To say the least, this claim is unrealistic.

Howard's test was reliable because real objects in real space were used to produce the desired disparities. Because no lenses, prisms, mirrors, filters or polarizers were used, there was no possibility of them modifying the results. Because the dimensions of the apparatus were large enough to be measured easily, the accuracy of disparities could be verified with precision.

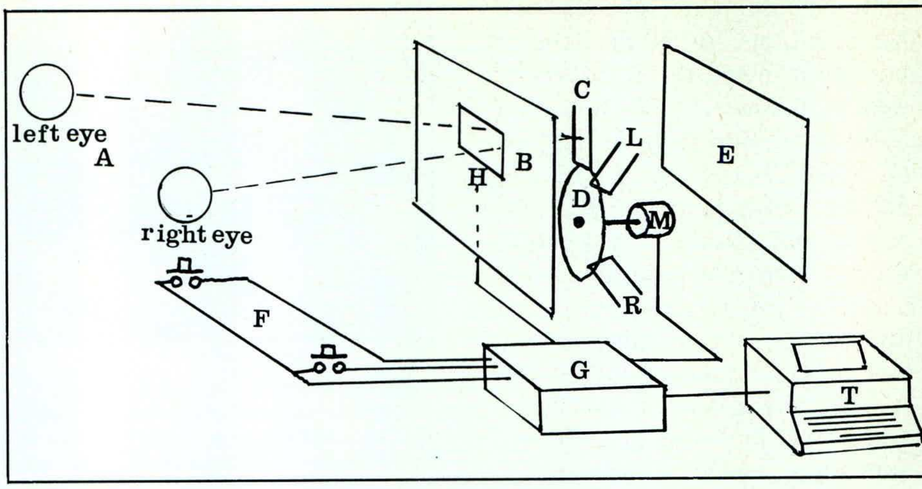

Fig. 1.
Sketch showing the elements of the automated stereoscopic acuity test. The patient (A) looks through a window in a plate (B) at a pair of pins (C) mounted on a disc (D). These are seen in silhouette against an illuminated surface (E). The pins (C) being equidistant from the observer, he presses both switches (F) at the same time. The computer (G) then closes the shutter (H) and causes the motor (M) to turn so that another pair of pins will be seen when the shutter opens. If the left pin is nearer (L) the left switch is actuated, if the right is nearer (R) the right switch is actuated. When the test is completed, the results are typed out by the terminal (T).

DESCRIPTION OF THE TEST

The pinwheel

The means of presenting the required disparities is the most important element in any test for stereoscopic acuity. Its design must meet the following criteria:

1. disparities can be changed quickly (so as not to waste time),
2. disparities can be reproduced repeatedly without excessive variation (so that the results can be precise) and
3. the instrument can be calibrated (so as to establish the amount of each disparity).

When real metal rods are used to create disparities, as in Howard's experiment, the first criterion is difficult to realize because both rods must be displaced after each display. In order to follow Howard's example and at the same time to produce a compact and quickly executed test, a new way of presenting the rods had to be devised. This took the form of a flat metal disc to whose periphery were fastened a series of pairs of pins (the rods) whose separation in depth was arranged to give the disparities needed. Please refer to the sketch in Fig. 1. The disc and pin assembly (the pinwheel) was attached, by means of a hub, to the shaft of a stepper motor. The motor was mounted within a box whose front surface (a metal plate) was provided with an opening (a rectangular window) through which only the pair of pins at the top of the disc could be seen. The motor could step to any one of 24 equally spaced positions. Because each level of disparity was to include two examples (right pin nearer than left and left pin nearer than right) and one pair was to be of zero disparity (neither nearer), only 11 disparity levels could be presented. The principal levels were powers of 2 because 2 sec is the absolute limit of stereoscopic acuity and because this provides a sequence in which levels increase by a multiple of 2. Disparities were chosen so that at 60 cm the least was 4 and the greatest 512 sec. A distance of 60 cm was found to be the least for which the smallest disparities could be produced. Because the test was for clinical use, it was deemed unnecessary to provide an example of the absolute limit (2 sec). When scientific studies require the absolute threshold, the viewing distance can be increased to 1 meter at which the smallest disparity is 1.4 sec. Details of these and other dimensions are presented in Table 1 along with the

TABLE 1

Comparative dimensions of Howard's apparatus and Larson's at two distances

		Dimensions								
		Rod				Window				Disparity (at 60 mm interocular dist.)	
	Test Dist.	Gap		Diam		Width		Height		Smallest	Largest
Apparatus	(cm)	(cm)	(min)	(cm)	(min)	(cm)	(min)	(cm)	(min)	(sec)	(sec)
Howard's	600.0	6.0	34.4	1.0	5.7	20.0	114.6	12.0	68.8	1.7	116.7
Larson's	60.0	0.3	17.0	0.08	4.5	2.5	143.0	1.1	63.0	4.0	512.0
	100.0	0.3	10.2	0.08	2.7	2.5	85.8	1.1	37.8	1.4	184.3

same dimensions for Howard's original apparatus. At 60 cm, and in brackets 1 meter, the 11 disparities were as follows: 4 (1.4), 6 (2.2), 8 (2.9), 12 (4.3), 16 (5.8), 24 (8.6), 32 (11.5), 64 (23), 128 (46), 256 (92) and 512 sec. (184). In both cases, disparities were calculated for a person with an interocular separation of 60 mm. One pair of pins was of zero disparity. The upper limit of 512 sec was chosen because poor stereoscopic acuity is better evaluated by other tests and also because the monocular cue of diameter difference becomes available at about 200 sec.

The system

To be effective and of general use, tests for stereoscopic acuity must ensure that success cannot be obtained by any means other than depth perception through stereopsis. If a test is too simple, it can be memorised. If it does not include repetitions, success may be had by lucky guessing. A test is defeated if correct answers can be found by means other than stereopsis. The Titmus STEREO-TESTS which includes a 4 dot test is familiar to all optometrists and has all three of these defects. No optometrist, for example, can use this test on himself because he knows the answers already. The nearer of 4 dots is to be identified. Therefore the probability of success by guessing is 1 in 4. It has already been reported that monocular cues are made available by the eccentric position of the nearer dot[9]. The effectiveness of this test is therefore questionable.

The pinwheel was incorporated into a system which is shown in Fig. 2. The pinwheel and motor are within the box seen to the right, the patient's head is immobilized and positioned by the head rest and responses are obtained by two switches recessed in the table-top (the left one being visible in the figure). Seen on the table is a model used to demonstrate the test procedure. Not shown is the computer which runs the test nor the terminal which records the results. In other configurations, the computer may be contained within the pinwheel box and the terminal attached to its top surface.

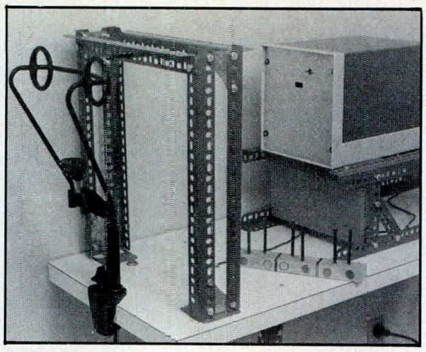

Fig. 2.
The test system. The head rest is to the left. The pinwheel is inside the box to the right. The small dark rectangle on the front face of the box is the window through which the pins are seen in silhouette against an illuminated background. Only the left push button can be seen. The metal bridge supports a refractor when needed. The demonstration model is seen on the table.

All test procedures were programmed. Each program is a list of instructions which in its entirely describes fully and invariably a given test procedure. The results obtained with the system should therefore be free of operator errors. A number of different procedures have been programmed to emulate established tests and to create new ones. Any or all of these programs can be stored in the computer's permanent memory so as to be ready for immediate use.

The program

Programming details are beyond the scope of this description but certain of them are of special interest because they ensure that errors do not contaminate the data. The position of the pinwheel is checked after every measurement to ensure that the disparity seen was the one intended. A shutter closes the window as soon as the switch or switches have been actuated and remains closed until the disc is in the next position and has stopped vibrating. The shutter is always closed for the same time so that there is no clue to the amount by which the disparity has been changed. The time between the opening of the shutter and the actuation of the switches is recorded and if it is less than the typical reaction time the response is ignored and a message is displayed on the terminal to warn the operator that the person may be guessing.

Howard's original test procedure was programmed for use as a reference against which other procedures could be compared. This was found to be unsuitable for clinical use because it took too long to perform and was unreliable when stereopsis was abnormal. In it, the threshold of stereopsis was taken as the disparity at which the nearer rod was identified correctly in 75% of the presentations. This required 20 presentations to be made at each disparity otherwise calculations could not be made to within 5%.

A clinical test was devised and programmed in two versions; one for screening and the other for a detailed study. In the complete test, the approximate acuity was located quickly (starting from 64 sec) and then the threshold region was thoroughly investigated. Presentations at each disparity level were made in groups of three: left pin nearer (here identified as left), right pin nearer (right) and neither pin nearer (equal). The order of presentation was chosen at random so that the probability of success by guessing was 1 in 3 for each presentation. The probability of guessing all three presentations correctly was 1 in 27. In the detailed study, three repetitions of the group of three were made at a given disparity. The probability of success by guessing was thereby reduced to 1 in 19,683. When there was no penalty for answering "equal" incorrectly this probability was reduced to 1 in 729. If this were too low, the test could be repeated to give a probability of 1 in 531,441.

The test session

A typical test session began by entering the patient's (or subject's) name and interocular distance at the terminal. The person was then shown the two push button switches and how to use them to indicate relative positions of the pins. Push left button for left, right button for right and both at the same time for equal. The head rest was adjusted to place the eyes at the same height as the window and to restrain the head from sideways movements (to obviate monocular parallax). If the patient were presbyoic, lenses were provided to ensure clear vision. This was done with a trial frame or else a phoropter head which replaced the head rest (it attached to the metal bridge shown in Fig. 2). Before the test began, zero disparity was displayed in the window. This ensured that the patient knew what equal looked like. The operator started the test from the terminal whereupon the window was closed for 2.4 sec. When it reopened, the patient indicated the nearer pin by means of the switches. The window then closed and reopened to reveal another pair of pins. This sequence was repeated until the test was either failed or the acuity found. In either case, the results were typed out at the terminal. The time required to perform the complete test was usually less than 3 minutes.

Trial 1

Person "A" 's name

Date = 801209
P.D. = 64MM
Dist = 0.6M

Howard's Test, 20 repetitions

Disp. = 4 sec (2 $\log_2$sec)

90% Correct, errors: left 0, right 2

Response time: left = 1.2, right = 1.6 sec.

Trial 2

Person "B" 's name

Date = 810114
P.D. = 60MM
Dist = 1M

Disparity	- - - nearest pins - - -		
Sec ($\log_2$sec)	left	right	equal
4 (2)	3/3	3/3	3/3
3 (1.6)	1/2	0/2	1/2
2 (1)	1/2	0/2	1/2
Equal 10	2	0	8
Time (2)	3.7	2.0	2.7

Fig. 3.
Two examples of results as typed at the terminal. Left: Howard's test. Right: Complete clinical test. Refer to text for a complete description.

The results

Results have been presented variously depending on the test. Two examples are shown in Fig. 3 "Trial 1" is in the format of Howard's test with 20 repetitions at each disparity level. A reminder of this is given on the fourth line. Pin disparities are given in seconds and in brackets as the logarithm to the base 2. The 6th line shows the percentage of correct answers and which pin was nearer when the errors were made. As shown, 90% of the responses were correct (18/20) and both errors were made when the right pin was nearer. The average response time after the window opened is shown on line 7. This was slightly less (by 0.4 sec) when the left pin was nearer. An interpretation of these data is that the person had an excellent stereoscopic acuity (4 sec is the test's lower limit at 0.6 meters) and was able to detect the nearer pin quickly. Detecting the right pin nearer was the most difficult decision because both errors occured then. The "right" response time, being longer, confirms this.

"Trial 2" shows the format for the complete clinical test. This was performed at a distance of 1 meter. The table presents the score at and below the perfect level (vide infra). The first column shows the disparity. The remaining columns show the score depending on which pin was nearest. The denominator of the fraction gives the number of presentations while the numerator gives the number of correct identifications. The perfect level is the least disparity at which all left and right identifications were made correctly; in this case 4 sec. The following lines show that only two presentations were made of each type. When the right pin was the nearer, no responses were correct (0/2). The equal line shows that "equal" was presented 10 times and that it was identified twice as left but never as right. The bottom line shows the average response time for each kind of presentation at the perfect level.

These examples show ways in which information can be compressed to assist in the interpration of the results.

DISCUSSION

This system used mechanical, electronic and computer technology to simplify and expedite stereoscopic acuity measurements. A clinician could learn to operate it in less than an hour. Patients have

TABLE 2

Stereo-acuity of 10 subjects by 5 methods

Subject no.	Stereo-acuity (sec of arc) Test A	Test B	Test C	Test D	Test E
1	4	11	85	20	60
2	8	7.2	39	20	120
3	12	15	55	20	30
4	12	18	140	20	30
5	12	36	210	20	240
6	16	3.6	65	20	30
7	16	7.2	140	20	120
8	24	3.6	140	20	30
9	24	40	440	20	120
10	64	3.6	140	20	30

found the test easy to comprehend and execute. So far, the youngest person to use it was 4 years old.

It has been demonstrated that performance cannot be improved by guessing. One person was confident that he could succeed, at least to some extent, using monocular vision alone. To make things easy for him, Howard's procedure was used (so that there would be two choices instead of three) and he was told that the nearer pin should look wider than the farther one. He failed the test at the 512 sec level with only 65% correct; a score which he might have equalled with his eyes shut. He had good stereoscopic acuity when tested binocularly.

As already mentioned, the same person can demonstrate different acuities when tested with different commercially available tests. Table 2 shows how the stereoscopic acuities of ten subjects differed when measured by 5 different tests; here identified by the letters A to E. Test A was the complete test decribed above. Test B and C were stereopairs, printed on paper backed with cardboard, viewed through a Brewster stereoscope. Both were by the same manufacturer. Test D was a random dot vectograph. Test E was a random dot red-green anaglyph printed on thick paper. Manufacturer's instructions were followed faithfully in each case. The results are arranged in order of stereoscopic acuity with respect to test A. It is seen that none of the other tests are in the same order. Test D is unique in attributing its best stereoscopic acuity to everyone. On the other hand, tests C and E often find very poor acuities. In tests A and B the range of numbers is similar but only 30% had similar acuities. Subject 10 showed the greatest dissimilarity which was 64 sec by test A and 3.6 sec by test B.

Which of these tests gave the best estimate of stereoscopic acuity? We suggest test A because it was performed without optical aids and used real objects whose disparity had been confirmed by calibration. It offered protection against guessing because the probability of success by chance was 1 in 729. On the other hand, the probability of success by chance in test D was at best only 1 in 4. We suggest these results substantiate our claim that this test sets a standard against which the performance of others can be judged.

Including zero (equal) among the disparity choices is not a usual practice. While it is of minor importance when testing persons with normal binocular vision it is an effective means for revealing anomalies of binocular perception.

The usefulness of zero disparity and of three possible responses is illustrated by a subject whose acuity was 2 sec with the automated version of Howard's test. We were surprised to find that he could do no better than 64 sec with the complete clinical test. Repeated measurements confirmed this. How could his stereoscopic acuity be 2 sec by one test and 64 sec by another? The subject provided the answer to this question himself. He explained that he could *always* tell when the left pin was nearer but at disparities of less than 64 sec he could *never* tell when right was nearer. In Howard's test, if left were not seen to be nearer he replied right and was always correct. When there were two alternatives, right and equal, this strategy was no longer successful. His stereopsis was evidently abnormal. Howard's procedure could not detect this but ours could.

Stereoscopic acuity is a threshold measurement which is customarily expressed in terms of probability; see Ogles discussion in Vol. 4 of Davson's The Eye.[10] This is appropriate for scientific studies but we are of the opinion that it is unsuitable for clinical use. Instead, we prefer to define stereoscopic acuity as the least disparity at which no errors are made. This is also a threshold because it is the least disparity beyond which the specified condition cannot be attained. To avoid confusion between this and the usual definition we have called it the perfect level.

For clinical purposes, stereoscopic acuity is the perfect level. When depth perception is required for a particular occupation this is so because mistakes will be made without it. The appropriate threshold is therefore the perfect level because the scientific definition includes the probability of errors. Most clinical tests for stereoscopic acuity identify this as the least disparity at which no error is made. Therefore, the perfect level has already been accepted as the clinical norm. Nevertheless, there is information to be gained by an examination of the failure mode below the perfect level. The "clinical procedure" provides for this by reporting results obtained at disparities of less than the perfect level and by showing how zero disparity was seen.

Acknowledgement

This study was supported by Le fonds de développement de la recherche, Université de Montréal.

References

1. Pearlman, J.T. Stereoscopic vision testing: a Review. *Amer. Orthoptic Journal* 19: 78-86, 1969.
2. Levy, N.S., and Glick, E.B. Stereoscopic perception and Snellen visual acuity. *Amer. J. Ophthal.* 78: 722-724, 1974.
3. Westheimer, G., and McKee, S.P. Stereoscopic acuity with defocused and spatially filtered retinal images. *J. Opt. Soc. Am.* 70: 772-778, 1980.
4. de Decker, W. and Scheffel, T. Simultaneous investigation of fixation disparity, stereothreshold and horopter balance. *Albrecht v. Graefes Arch. klin. exp. Ophthal.* 207: 97-109, 1978.
5. Ronne, G. The physiological basis of sensory fusion. *Acta Ophthalmologica* 34: 1-26, 1956.
6. Graham, C.H. (Ed.) Vision and Visual Perception. New York: John Wiley, 1965, pp. 525-528.
7. Howard, H.J. A test for the judgement of distance. *Amer. J. Ophthalmol.* 2: 656-675, 1919.
8. Scott, W.E. and Mash, J. Stereoacuity in normal individuals. *Annals of Ophthal.* 6: 99-101, 1974.
9. Cooper, J. and Warshowsky, J. Lateral displacement as a response cue in the Titmus Stereo Test. *Amer. J. Optom. & Physiol. Optics* 54: 537-541, 1977.
10. Davson, H. (Ed.) The Eye. New York: Academic Press, 1962, pp. 286-288.