Aperture size requirements

John Seymour, 7/21/2006

1. Abstract

George Santayana[1] said that "those who cannot remember the past are condemned to repeat it".

Various standards document [1, 2, 4, 5, and 6] contain tables that give recommendations for aperture size when measuring halftones at various screen rulings. Reconciling small differences between the tables in the various documents has been a recurring issue for the relevant standards committees. From a practical standpoint, the differences are trifling.

This paper lays out the computations that drive the tables as a guide to the standards committees. It is intended that this document be the final word on aperture size, at least until something else is written on the topic.

2. The Source

2.1. Sigg's original recipe

The recommendations for aperture size are all based on a paper from Franz Sigg [3]. In his paper, he pointed out a source of error in measuring halftone areas. The density read by a densitometer depends on the number of halftone dots in the aperture. In turn, the number of halftone dots within the aperture of a densitometer depends upon the positioning of the aperture with respect to the halftone dots.

His paper gives the following formula that predicts the maximum error due to repositioning in the computation of percent dot area. The inputs to his formula are the screen ruling of the sample being measured and the diameter of the circular aperture of the densitometer.

$$\varepsilon_{\mathrm{p}} = \frac{K}{\sqrt{(D \cdot S)^3}} \qquad (1)$$

Where

ε_p is the maximum error due to positioning,

K is a constant, being 900 if the screen ruling is in lines per cm, and 3650 if the screen ruling is in lines per inch,

D is the aperture diameter in mm, and

S is the screen ruling in lines per unit.

Since I am not scared of taking numbers to weird powers, I prefer the following

[1] At least I think it was Santayana. I am not sure I remember for sure. I will have to look it up again.

$$\varepsilon_{\mathrm{p}} = K(DS)^{-1.5} \tag{1b}$$

As an example, if the diameter of the aperture is 2 mm and the halftone ruling is 133 lines per inch, then the maximum error is 0.84%. This means that a 50% patch may read anywhere from 49.16% to 50.84%[2], depending upon the positioning of the aperture.

One source of a tiny discrepancy is that the constants do not exactly agree with each other. If we take $K = 900$ to be the exact number, then the corresponding constant for English units is $K = 3643.28$ for the metric constant. If 3650 is taken as exact, then the metric constant is $K = 901.66$. (To convert from metric constant to English, you multiply by $2.54^{1.5} \approx 4.04809$.)

From a practical standpoint, the difference is inconsequential. Furthermore, the numbers were arrived at strictly by curve fitting so neither constant is exact anyway. Since standards people need to be pedantic nitpickers, we need to decide which constant is exact, and then base all subsequent calculations on that number. I *would* vote to go with the metric constant for two reasons. First, if the English constant is used, then the equation is a silly mixing of an English screen ruling and a metric aperture size. Second, the number 900 is so much easier.

2.2. Formulation based on dot count

I said I *would* vote on metric, but I have another consideration. Sigg's formula is a bit cumbersome in that it requires two inputs. The real issue is to get enough halftone dots in the aperture. As such, a more elegant form of equation 1 would not be based on screen ruling and aperture size, but rather on the total number of dots in the aperture.

Such a formulation is "cleaner" in that it is unit-less and requires only one input. Also, the formula is more readily applicable to gravure, where the halftone dots are not laid out on a rectangular lattice, so that a halftone pattern has two screen rulings, depending on the angle that the screen ruling is measured.

If the diameter of the aperture is *D* mm, then the area is $\frac{\pi}{4}D^2$ mm^2. The area of each halftone dot, or rather, the area of the space allocated for each halftone dot is $\frac{100}{S^2}$ mm^2.

We can determine the number of halftone dots in this area, *T*, by dividing the area of the aperture by the area allocated to each dot:

$$T = \frac{\frac{\pi}{4}D^2 mm^2}{\frac{100}{S^2} mm^2} = \frac{\pi}{400}D^2 S^2 \tag{2}$$

We can solve this for the product DS:

[2] Note that this is actually an error in percentage *points* and not an error as a percentage of the answer.

$$DS = \left(\frac{400T}{\pi} \right)^{0.5} \tag{3}$$

This is then substituted in Equation 1b

$$\varepsilon_{\mathrm{p}} = 900 \left(\left(\frac{400T}{\pi} \right)^{0.5} \right)^{-1.5} = 900 \left(\frac{400T}{\pi} \right)^{-0.75} = \frac{900 \left(\frac{\pi}{400} \right)^{0.75}}{T^{0.75}} \tag{4}$$

$$\varepsilon_p = \frac{9}{2} \sqrt{5} \left(\frac{\pi}{T} \right)^{3/4} \tag{5}$$

$$\varepsilon_p \approx 23.74433 T^{-0.75} \tag{5b}$$

I recommend that equation 5b by taken as the equation on which to base standards for aperture size based on screen ruling.

3. The Recommendations

3.1. Required number of dots

All of the recommendations in the standards are based on a simple premise: The minimum aperture size is the smallest aperture that guarantees an error in computation of dot area that is less than 1%. The recommended aperture size is the smallest aperture that guarantees an error less than 0.5%.

Based on equation 5, the following equation will compute the required number of halftones dots *T* for any particular specification for positional error.

$$T = \pi \left(\frac{164025}{16} \right)^{\frac{1}{3}} \left(\frac{1}{\varepsilon_p} \right)^{\frac{4}{3}} \tag{6}$$

$$T \approx \left(\frac{23.74433}{\varepsilon_p} \right)^{\frac{4}{3}} \tag{6b}$$

The following table is based on equation 6b:

Error	T
0.1	1470.32
0.2	583.50
0.5	171.97
1.0	68.25
2.0	27.08
3.0	15.77

Thus, the number of dots required for errors less than 0.5% and 1% are 171.97 and 68.25, respectively.

3.2. Required aperture diameter

Given a required number of dots, we can compute the aperture size required for a given halftone dot count using the following formula:

$$d = \sqrt{\frac{4T}{\pi}} \tag{7}$$

Where

d is the width of the required circular aperture, measured in halftone dot spacings.

This formula is derived by first noting that the number of dots is essentially a measurement of area. The circle with that requisite area is then determined.

We can then substitute equation 6 in for T in equation 7 to get the requisite circular aperture width as a function of the

$$d = \sqrt{\frac{4\pi\left(\frac{164025}{16}\right)^{\frac{1}{3}}\left(\frac{1}{\varepsilon_p}\right)^{\frac{4}{3}}}{\pi}} = \sqrt{\left(\frac{64\times164025}{16}\right)^{\frac{1}{3}}}\left(\frac{1}{\varepsilon_p}\right)^{\frac{2}{3}} = 3(30)^{\frac{1}{3}}\left(\frac{1}{\varepsilon_p}\right)^{\frac{2}{3}} \tag{8}$$

$$d \approx 9.3216975\left(\frac{1}{\varepsilon_p}\right)^{\frac{2}{3}} \tag{8b}$$

Below is an expansion of the previous table, with values for d (required aperture width) included.

Error	T	d
0.1	1470.32	43.267
0.2	583.50	27.257
0.5	171.97	14.797
1.0	68.25	9.322
2.0	27.08	5.872
3.0	15.77	4.481

Thus, the required widths for 0.5% and 1% are 14.797 halftone dots and 9.322, respectively.

3.3. Noncircular apertures

A traditional densitometer has a circular aperture. Since densitometers with non-circular apertures (such as video densitometers) are now available, it is important to extend the recommendations to include non-circular apertures.

The assumption is made that the positioning error is strictly a function of the area of the aperture, independent of the shape of the aperture. Given this assumption, one can use equation T to compute the requisite number of halftone dots, and then dimension the aperture accordingly.

This assumption is not strictly true, but it is a close approximation. If an instrument with a round aperture is rotated about the center of the aperture, there is no change in the number of halftone dots that are seen. As a square aperture is rotated, however, the number of halftone dots will go up and down. This is a variation that is in addition to what Sigg had looked at. In the extreme, one can envision an aperture that is considerably narrower than the screen ruling so that, regardless of its length, it will always see the maximum (black-to-white) variation on a highlight patch

From this simple analysis, I would guess that a rectangular aperture might have slightly more positional error than a round aperture of the same area.

3.4. My recommended recommendations

The following table is based on the preceding equations from this section. I recommend that this table be used in future standards.

LPI	LPC	$D_{\min}$	D_{rec}	$A_{\min}$	A_{rec}
65	25.6	3.6	5.8	10.4	26.3
85	33.5	2.8	4.4	6.1	15.4
100	39.4	2.4	3.8	4.4	11.1
120	47.2	2.0	3.1	3.1	7.7
133	52.4	1.8	2.8	2.5	6.3
150	59.1	1.6	2.5	2.0	4.9
175	68.9	1.4	2.1	1.4	3.6
200	78.7	1.2	1.9	1.1	2.8

The first column (LPI) is taken to be exact numbers. I have anglocentrically chosen to base the table on English units.

The second column (LPC) is computed from the first column by dividing through by 2.54.

The third column ($D_{\min}$) is based on the number 9.322 as computed in 3.2. This number represents the minimum aperture size (in units of dot spacing) required to achieve an error of less than 1%. The equation is

$$D_{\min} = 10 \times \frac{9.322}{LPC} \tag{9}$$

Similarly, the fourth column is based on the number 14.797 from the same section.

$$D_{rec} = 10 \times \frac{14.797}{LPC} \tag{10}$$

Note that third and fourth columns are based on the actual computed values of the LPC, and not on the value in the table, which was rounded to one decimal place.

The fifth and sixth columns are computed from the third and fourth columns (once again, actual and not rounded). The computation is merely a computation of the area of the rounded aperture from the minimum radius column.

$$A_{\min} = D_{\min}^2 \frac{\pi}{4} \tag{11}$$

$$A_{rec} = D_{rec}^2 \frac{\pi}{4} \tag{12}$$

4. What do the standards say?

The tables in the standards disagree because of round-off errors at different points for the calculations.

4.1. ISO/CD 13655 (WG8N0011)

WG8N0011 (Section D.1.3) says:

> *For a periodic screen and a single measurement, a rule of thumb says that at least 79 halftone dots need to be within the sampling aperture although 177 are better.*

These numbers are somewhat different than the minimum dot counts for 1% and 0.5% error that were derived in 3.1 (68.25 and 171.97). The numbers are clearly not just rounded to the nearest digit, but they are close enough to suggest that the intent was for the "at least" number of halftone dots to ensure less than 1% error, and for the "better" number of halftone dots to ensure an error less than 0.5%.

Why is there a discrepancy between 79 and 68.25? WG8N0011 goes on to say:

> *This corresponds to the rule that for circular apertures the diameter ought to be at least ten times, better 15 times, greater than the screen width.*

The numbers 79 and 177 can be computed directly from the numbers 10 and 15. A circular aperture of diameter *d* will have an area of $d^2 \frac{\pi}{4}$ halftone dots, so that apertures of diameter 10 and 15 correspond to $10^2 \frac{\pi}{4} \approx 78.54$ halftone dots and $15^2 \frac{\pi}{4} \approx 176.71$. Rounding these numbers to the nearest whole number gives 79 and 177.

But, where do the aperture diameters 10 and 15 come from? The numbers 10 and 15 are themselves rounded from 9.322 and 14.797 halftone dots, as computed in 3.2. I am guessing that 9.322 had been rounded to 10 mainly because 10 is a very "handy" number for a rule of thumb. After all, there are ten fingers and thumbs on the two hands, at least for most people.

The following table also appears in WG8N0011:

LPI	LPC	D_{min}	D_{rec}	A_{min}	A_{rec}
65	25.6	3.9	5.9	12.0	27.0
85	33.5	3.0	4.5	7.0	15.8
100	39.4	2.5	3.8	5.1	11.4
120	47.2	2.1	3.2	3.5	7.9
133	52.4	1.9	2.9	2.9	6.4
150	59.1	1.7	2.5	2.3	5.1
175	68.9	1.5	2.2	1.7	3.7
200	78.7	1.3	1.9	1.3	2.9

I have gone through all the possibilities of how the numbers in this chart were calculated, and suggest this as the most plausible sequence of calculations.

First, the metric screen frequencies in this chart are to be taken as precise. The screen frequencies in English units are derived from the metric.

The minimum area, based on requiring 79 and 177 halftone dots, is readily computed. The minimum and recommended areas are calculated as

$$A_{\min} = 79 \times \left(\frac{10}{S}\right)^2 \text{, and} \tag{13}$$

$$A_{rec} = 177 \times \left(\frac{10}{S}\right)^2 \tag{14}$$

Dividing ten by S converts from lines per cm to mm per line. Squaring this gives the area of a single halftone dot. Since 79 or 177 of these halftone dots are required, we multiply by 79 or 177 to get the area required.

Using this calculation and rounding the results to one decimal place gives numbers that all agree with the chart in WG8N0011, with one exception. For a line screen of 25.6 lines per cm, the entry is 12.0, but the computed result is 12.054. Perhaps the number was initially rounded to two decimal places, and then this rounded number was in turn rounded to one decimal place[3].

The required areas ($A_{\min}$ and A_{rec}) could be converted into diameter specifications by using equation 7. Alternately, a new formula could have been developed, by substituting the value for A from equation 13 or 14 into equation 7:

$$D_{\min} = \sqrt{\frac{4A_{\min}}{\pi}} = \sqrt{\frac{4 \times 79 \times \left(\frac{10}{S}\right)^2}{\pi}} = \frac{20}{S}\sqrt{\frac{79}{\pi}} \text{, and} \tag{15}$$

$$D_{rec} = \sqrt{\frac{4A_{rec}}{\pi}} = \sqrt{\frac{4 \times 177 \times \left(\frac{10}{S}\right)^2}{\pi}} = \frac{20}{S}\sqrt{\frac{177}{\pi}} \tag{16}$$

4.2. ISO 13655, first edition, 10-1996, and CGATS.5-1993

Quoted from Annex F:

> *Another factor that should be considered in the selection of instrument aperture when measuring half-tone images is the relationship to screen ruling. Table F.1 shows the minimum recommended aperture size as a function of common screen ruling.*

[3] Dave McDowell has since corrected me. Perfect agreement can be made to the 13655 draft table table if we start from the 10/15 rule rather than from the 79/177 rule.

Table F.1 — Minimum recommended aperture size

Nominal screen frequency		Sampling aperture minimum size
lines/cm	lines/in	mm
26	65	3,5
33	85	3,0
39	100	3,0
47	120	2,0
52	133	2,0
59	150	2,0
79	200	1,5
118	300	1,0

These aperture sizes are obviously rounded to the nearest 0.5 mm, which is a very reasonable thing to do. On the other hand, I have not been able to determine how these numbers were calculated. I have compared them against the 10/15 rule and the 79/177 rule, and have looked at metric and at English. None of the combinations agree completely with the table.

This verbiage and table also appear in CGATS.5-1993.

4.3. ISO 14981

ISO 14981 says that

> *The diameter of a circular sampling aperture should be not less than 15 times the screen width; it shall be not less than 10 times the screen width that corresponds to the lower limit for the screen ruling stated by the vendor, see 4.7. The area of non-circular sampling apertures shall not be smaller than that required for circular sampling apertures.*

The 10/15 rule was discussed earlier. There is no other mention of aperture size and screen ruling.

4.4. CGATS.5-2003

The table below is a comparison between ISO/CD 13655 (WG8N0011) and CGATS.5-2003. The differences are bolded in the CGATS.5 table.

		ISO/CD 13655 (WG8N0011)				CGATS.5-2003			
LPI	LPC	$d_{\min}$	d_{rec}	$A_{\min}$	A_{rec}	$d_{\min}$	d_{rec}	$A_{\min}$	A_{rec}
65	25.6	3.9	5.9	12.0	27.0	**3.8**	**5.7**	**11.3**	**25.5**
85	33.5	3.0	4.5	7.0	15.8	3.0	4.5	**7.1**	**15.9**
100	39.4	2.5	3.8	5.1	11.4	**2.6**	**3.9**	**5.3**	**11.9**
120	47.2	2.1	3.2	3.5	7.9	2.1	3.2	3.5	**7.8**
133	52.4	1.9	2.9	2.9	6.4	1.9	2.9	**2.8**	6.4
150	59.1	1.7	2.5	2.3	5.1	1.7	**2.6**	2.3	5.1
175	68.9	1.5	2.2	1.7	3.7	**1.4**	**2.1**	**1.5**	**3.5**
200	78.7	1.3	1.9	1.3	2.9	1.3	**2.0**	1.3	**3.0**

Clearly, the tables were calculated a little differently. Why the difference?

Whereas the WGN0011 table was based on a minimum dot count of 77 and 179, CGATS.5 is based on the minimum widths of 10 and 15 screen rulings, using metric units. Or at least the minimum radius column is. The other three columns are all based on the rounded values in the minimum radius column.

The recommended radius column is simply 1.5 (that is, 15 divided by 10) times the rounded minimum radius column.

The column for minimum of non-round apertures is also based on the rounded minimum diameter column, with the use of equation 11. The final column (recommended area of non-round aperture) is also computed directly from the rounded values of the minimum radius column.

5. Bibliography

[1] ANSI-CGATS.5-2003, *Graphic technology – spectral measurement and colorimetric computation for the graphics arts*, *Graphic technology – spectral measurement and colorimetric computation for the graphics arts*

[2] ISO/CD 13655 (ISO TC 130/JWG 8 N0011),

[3] Sigg, Franz, *Errors in Measuring Halftone Dot Area*, Journal of Applied Photographic Engineering 9: 27 - 32 (1983)

[4] ISO 14981, *Graphics technology – Process control – Optical, geometrical and metrological requirements for reflection densitometers for graphic arts use*

[5] ISO 13655, first edition, 10-1996

[6] CGATS.5-1993